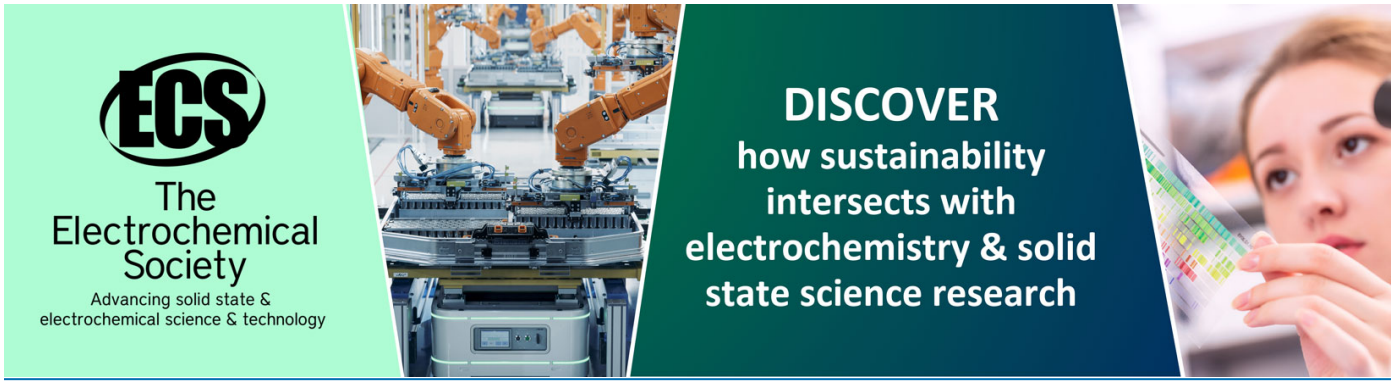
# Molecular Docking of Alkaloid Compound SA2014 towards Cyclin D1 Protein in Cancer using Firefly Algorithm





# Molecular Docking of Alkaloid Compound SA2014 towards Cyclin D1 Protein in Cancer using Firefly Algorithm

**Zulfa Afiq Fikriya[1], Mohammad Isa Irawan[1] and Awik Puji Dyah Nurhayati[2]**

[1]Department of Mathematics, Institut Teknologi Sepuluh Nopember, Indonesia
[2]Department of Biology, Institut Teknologi Sepuluh Nopember, Indonesia

zulfa.afiq@gmail.com, mii@its.ac.id, awiknurhayati@gmail.com

**Abstract**. Molecular docking or ligand binding in proteins is a developing field of computing. Molecular docking can be used to find the most appropriate interaction pattern between protein receptors and ligands and become the basis for the drug discovery and design based structures. The development of efficient docking methods and algorithms will be very useful in drug discovery simulation. Firefly algorithm is one of the method that can be used for molecular docking simulations. Firefly algorithm is used to find the optimal conformation of proteins and ligands so that the binding energy of the whole system is minimized. In this research, protein-ligand complexes from the Protein Data Bank (PDB) were used to test the performance of the algorithm. The results show that the firefly algorithm can be used to solve molecular docking. Then this algorithm is used to solve molecular docking of alkaloid compounds SA2014 from Cinachyrella anomala sea sponges towards cyclin D1 protein in cancer. The results show that the SA2014 ligand affinity for cyclin D1 protein was higher than doxorubicin (a type of chemotherapy drug) so that the SA2014 compound have a great potential as an anticancer.

## 1. Introduction

One of the emerging technology fields now is biological computing. Biological computing is a field of science that focuses on compilation of a mathematical model in completing and analyzing biological sequence problem. Biological computing or known as bioinformatics is a combination of biology and computing that uses applications from computational and analytical tools to capture and interpret biological data. Molecular docking is one of the developing field in bioinformatics. Molecular docking aims to mimic the interaction of a ligand molecule with a protein targeted at in-vitro tests [1].

Molecular docking algorithm seek to predict the bound conformations of ligand and protein. More spesifically, given two molecules of known three-dimensional structure, ligand and protein, is it possible to determine their three-dimensional structure when combined together (complexed). Docking algorithms using energy-based scoring function seek to identify the energetically most favorable ligand conformation when bound to the protein molecule. The general hypothesis is that lower energy scores represent better protein-ligand bindings compared to higher energy values. The docking problem is therefore an optimization problem where the task is to find the conformation with the lowest energy [2].

Molecular docking has proven to be very effective in studies of ligand protein interactions. Docking is a difficult problem by involving many degrees of freedom, so the development of methods and efficient docking algorithms will be very useful in the design of new drugs [3]. The optimization algorithm is very helpful in docking to get the drug design in a simulation. Artificial intelligence methods have been applied to molecular docking problems, including genetic algorithm, differential

evolution, particle swarm optimization [4], harmony search algorithm [5], ant colony optimization [6] and machine learning methods, namely extreme learning machine [7].

Molecular docking is a very important tool in the discovery and design of drugs based on structure. One of them is the design of cancer drugs. Cancer is a disease caused by abnormal growth of cells in the body's tissues that turn into cancer cells. These cancer cells can spread to other parts of the body so that they can cause death. Anti-cancer drugs that do not cause side effects are generally obtained from nature. Research to get potential candidates for new anti-cancer drugs is still very much needed. The design of cancer drugs is needed as the initial selection stage and to find out good treatment through interactions between proteins and protein-ligand.

In this research we study the use of firefly algorithm to solve molecular docking. Firefly algorithm is an algorithm that is based on the patterns and behavior of light from firefly flocks. Firefly algorithm is used to find optimal protein and ligan conformation so that the free energy of the overall system is minimized. The case to be taken is molecular docking of SA2014 alkaloid compounds from the sea sponge Cinachyrella anomala towards cyclin D1 protein in cancer.

## 2. Materials and Methods

### *2.1. Molecular Docking*

The aim of molecular docking is to achieve optimal conformation for proteins and ligands and the relative orientation between proteins and ligands so that the free energy of the whole system is minimized. The computational process looks for ligands that match both geometrically and energy to the binding site of this protein called molecular docking. Molecular docking helps in studying drugs / ligands and receptor / protein interactions by identifying suitable active sites in proteins, obtaining the best geometry from the receptor-ligand complex, and calculating the interaction energy from different ligands to design more effective ligands.

The main objective in molecular docking is to find an optimized conformation between the ligand and the protein that results a minimum binding energy. The interaction between ligand and protein can be described by an objective function determined according to three components representing degrees of freedom: (1) the translation of ligand molecule, involving three axes values (x, y, z) in Cartesian coordinate space; (2) the ligand orientation, modeled as a four quaternion variables including the angle slope (w); and (3) the flexibilities, represented by the free rotation of torsion (dihedral angles) of the ligand [4]. Thus, each solution for molecular docking problem is a vector that consists of n+7 variables.

The success of docking depends on two factors: the scoring function and the search method being used to find the solution. Scoring function is used to calculate the affinity of a protein-ligand complex and to sort the compound rank. The low value of free binding energy indicates that conformation formed is stable, while the high value of free binding energy indicates the unstability of the complex formed. The search algorithm is used to detemine the most stable conformation (docking pose) of the protein-ligand complex [8]. The ligand functional groups will interact with the amino acid residues of protein and form intermolecular bonds. The strength of this bond is calculated and ranked with the scoring function.

### *2.2. Firefly Algorithm*

Firefly algorithm (FA) was first developed by Xin She Yang in late 2007 and 2008 at Cambridge University, which was based on the flashing patterns and behaviour of fireflies [9]. In essence, FA works based on the following rules:

- Fireflies are unisex, which is every firefly will be interested in other fireflies regardless of their sex.
- The attractiveness is proportional to the light intensity. The light intensity will decrease as their distance increases. For any two flashing fireflies, the less brighter one will move towards the brighter one. If there is no firefly brighter than the others, it will move randomly.
- The light intensity of a firefly is determined by the objective function from the problem given.

As a firefly's attractiveness is proportional to the light intensity seen by adjacent fireflies, we can now define the variation of attractiveness $\beta$ with the distance $r$ by

$$\beta = \beta_0 e^{-\gamma r^2} \quad (1)$$

where $\beta_0$ is the attractiveness at $r = 0$.

The movement of a firefly $i$ is attracted to another more attractive (brighter) firefly $j$ is determined by

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2}\left(x_j^t - x_i^t\right) + \alpha_t \epsilon_i^t \quad (2)$$

where the second term is due to the attraction. The third term is randomization with $\alpha_t$ being the randomization parameter, and $\epsilon_i^t$ is a vector of random numbers at time t [9].

### *2.3. Alkaloid SA2014*

The SA2014 compound is a derivative of the cinachyramine group and belongs to the alkaloid group. Alkaloids are compounds that have an element of nitrogen and are usually cyclic. This compound was isolated from Cinachyrella anomala sea sponges. These sea sponges belong to the Demospongiae class and many are found in Kukup Beach, Special Region of Yogyakarta. Isolation on the sponge is done by cutting the sponge body into smaller parts (2-3 mm) and macerating with ethanol. Furthermore, isolation and elucidation of the compound using a thin layer chromatograph (TLC) were carried out. The results of the isolation are long crystals and have a melting point of 121℃. The compound isolates that have been obtained are then identified by collecting spectroscopic data through FT-IR, 1H-NMR spectra, 13C-NMR, two-dimensional NMR spectra. Based on the test it is known that the compound has the formula C10H13N3O with the structure name 1,4,9-triazatricyclo [7, 3, 1, 0] trideca-3, 5 (13), 10-trien-8-ol (SA2014) [10].

Nurhayati et al. [11] have been studied about docking the SA2014 alkaloid compounds and doxorubicin towards the P53 protein in breast cancer. The result showed that the SA2014 compound had the ability as an anticancer compound against T47D breast cancer through leucine amino acid interactions and phenylalanine.

### *2.4. Cyclin D1 Protein*

The cell cycle in eukaryotic cells can be divided into four phases, namely $G_1$ (Gap 1), S (Synthesis), $G_2$ (Gap 2), and M (Mitosis). During $G_1$ phase, the cell continues to grow and make preparations for DNA synthesis. The cell do the DNA synthesis and chromosome replication in the S phase. In the $G_2$ phase, cells that have replicated the chromosome will duplicate the whole other cellular components. There may be additional cell growth during $G_2$. The final preparations for the mitosis phase must be completed before the cell is able to enter the first stage of mitosis. Mitosis phase consists of four sub phases, namely prophase, metaphase, anaphase, and telophase. Under certain conditions, cells do not divide and leave $G_1$ phase into $G_0$ phase. Cells in $G_0$ phase are often called resting/silent.

Cyclin D protein is one of the positive regulators in the cell cycle. Cyclin D consists of three types namely cyclin D1, D2, and D3. In the cell cycle, cyclin D1 does not only play a role during the $G_1$ phase. In the $G_2$ phase the level of cyclin D1 tends to increase, whereas in phase S the level of cyclin D1 tends to be low. The activity of cyclin D1 in the $G_2$ phase depends on the proliferative signal which will induce cyclin D1. In this phase the cell will determine the point to continue its proliferation, when there is no cyclin D1, the cell cycle will enter the resting phase. In the S phase the presence of cyclin D1 must be suppressed because cyclin D1 has the ability to inhibit DNA synthesis in terms of its ability to bind to important regulators of DNA synthesis, PCNA [12].

Cyclin D1 is known to correlate with early cancer symptoms and the risk of tumor development and metastasis. The cyclin D1 gene, CCND1 has an amplification of 20% and there is excessive expression of cyclin D1 protein in cancer. The results of cyclin overexpression can be induced from oncogenic signals or from mutations in the cyclin gene. This results in excessive growth of cancer cells. The cyclin D1 gene is located on chromosome 11q13 where on this chromosome the genome usually experiences amplification in human carcinoma, including cancer [13].

## 3. Implementation

### *3.1. Objective Function*

The objective function or scoring function in molecular docking is to minimize the total free binding energy of the ligand-protein complex based on the following equation [4].

$$\Delta G = \left(V_{bonded}^{L-L} - V_{unbonded}^{L-L}\right) + \left(V_{bonded}^{R-R} - V_{unbonded}^{R-R}\right) + \left(V_{bonded}^{R-L} - V_{unbonded}^{R-L} + \Delta G_{conf}\right) \quad (3)$$

$$V = W_{vdw} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i\right) e^{\left(-\frac{r_{ij}^2}{2\sigma^2}\right)} \quad (4)$$

$$\Delta G_{conf} = W_{conf} N_{tors} \quad (5)$$

In calculating this objective function, a grid map obtained from Autodock is used. To reduce the overall runtime of the docking simmulation, AutoDock uses a grid-based approach to approximate the energy calculations used by the energy function. In this method, the active site of protein is embedded in a 3D grid map and at each point of the grid map, electrostatic interactions energy, desolvation, hydrogen bonds, and van der waals have been calculated and stored. Thus, the energy value for all points that are in the grid map can be calculated using trilinear interpolation [14].

### *3.2. Firefly Algorithm for Molecular Docking*

To solve molecular docking problem, every fireflies has three components representing the ligand translation, four components representing the ligand orientation and the other components representing the ligand conformation. The translational components are the x, y, z reference atom coordinates, the orientational components are a quaternion constituted by a unit vector and one orientational angle. The conformational components are the ligand dihedral angles (one component to each dihedral angle).

For each firefly in the initial population, the value of translation variables x, y, z are random number between the minimum and maximum extents of the grid maps. The orientation were given a random quaternion, consisting of a random unit vector and a random angle between $[-\pi, \pi]$. The flexible torsion angles (if any) were given a random value in the range of $[-\pi, \pi]$. The pseudo code of firefly algorithm is shown in Figure 1.

## 4. Results and Discussions

### *4.1. Validation*

The results of molecular docking are evaluated based on the value of the energy produced. The smaller energy value obtained indicates that the more stable the bond that occurs between proteins and ligands. Evaluation of the success of the docking method can also be done by finding the root mean square deviation (RMSD) value, by comparing the ligand reference with the results of the docking that has been done. Molecular docking results can be accepted if the RMSD docking results compared to the reference have a value of less than 2 Å.

```
Objective function f(x), x = (x_1, ..., x_d)^T
Initialize firefly population x_i (i = 1, 2, ..., n)
The light intensity I_i on x_i is determined by f(x_i)
Determine the light absorption coefficient γ

while (t <MaxGeneration)
for i = 1 : n
        for j = 1 : n
        if (I_i < I_j), move firefly i to j; end if
                Change the attractiveness at distance r with exp[−γr]
                Evaluate new solutions and update light intensity
        end for j
end for i
Sort firefly's ranking and find the best solution g*
end while
```

**Figure 1.** Pseudo code of firefly algorithm for molecular docking

For validation, the results of docking obtained with firefly algorithm were compared with the results of docking obtained with genetic algorithm in AutoDock. The experiment was carried out three times for each algorithm. The firefly algorithm parameters used are shown in table 1, while for genetic algorithm uses the default parameters in AutoDock.

**Table 1.** Parameters of firefly algorithm

| Parameter | Value |
|---|---|
| Population size | 30 |
| Number of iterations | 1000 |
| $\beta_0$ | 1 |
| $\alpha$ | 1 |
| $\gamma$ | 1 |

Protein-ligand complexes consisting of a protein and a ligand was used to test the algorithm. All of the tested complexes were selected from the Protein Data Bank (PDB). The performance of firefly algorithm and genetic algorithm of AutoDock were compared regarding the ligand-protein complexes. Table 2 and Table 3 shows the results of these experiments.

**Table 2.** Comparison between FA and GA in terms of energy scores

| PDB code | Energy (Mean) | |
|---|---|---|
| | FA | GA |
| 2cpp | -4.52 | -4.82 |
| 3ptb | -4.50 | -4.57 |

**Table 3.** Comparison between FA and GA in terms of RMSD values

| PDB code | RMSD (Mean) | |
|---|---|---|
| | FA | GA |
| 2cpp | 0.79 | 0.96 |
| 3ptb | 1.21 | 1.65 |

Table 2 shows a comparison of the docking results for two ligand-protein complexes with FA and GA based on energy scores. The average energy scores obtained with GA is smaller than FA for the both complexes. But in table 3, it can be seen that docking with FA gives smaller RMSD values for both complexes.

*4.2. Molecular docking of SA2014 towards cyclin D1 protein*

In this research, a test was conducted with SA2014, an alkaloid class anticancer compound to determine its interaction with cyclin D1 protein. This SA2014 compound has been investigated as an anticancer in vitro. In this research also used positive control in the form of commercial drugs, namely doxorubicin. A molecular docking procedure is used as a reference to determine the best orientation of a compound against other compounds. Molecular dockinng simulation test produces a conformation of values / scores on SA2014 compounds and doxorubicin compounds. The results of the docking simulation of SA2014 compound and doxorubicin are shown in Table 4.

**Table 4.** Docking results of SA2014 compounds and doxorubicin towards cyclin D1 protein

| Ligand | Docking score |
|---|---|
| SA2014 | -2.98 |
| Doxorubicin | -1.66 |

Molecular docking results is a score that describes the total energy of a protein-ligand binding. The lower score of a docking result means the lower energy used to bind and show the interactions between protein-ligand complexes which are more stable so that they are more potent. A compound can be said to be more potent than other compounds by comparing the results of both docking scores. The results of the docking score between SA2014 alkaloid compounds and doxorubicin with cyclin D1 protein have different results. The score between SA2014 ligands and cyclin D1 protein is lower than the score between doxorubicin and cyclin D1 protein. This can be interpreted that the quality of the SA2014 ligand against cyclin D1 protein is higher than doxorubicin.

## 5. Conclusions

The results of the experiment show that firefly algorithm can be applied to solve molecular docking. Firefly algorithm is able to provide better results compared to genetic algorithm based on its RMSD value. Based on the results of the docking simulation, the SA2014 compound has a score of -2.98, while the doxorubicin has a score of -1.66. This can be interpreted that the quality of the SA2014 ligand against cyclin D1 protein is higher than doxorubicin, so that the SA2014 compound has great potential as an anticancer.

**Acknowledgements**

The author would like to thank the Institut Teknologi Sepuluh Nopember (ITS) for the scholarships given to take the master's program in the ITS mathematics department. The author also thank the

ministry of research, technology and higher education for the master research funding that has been given.

**References**

[1] Gane P J and Dean P M 2000 *Current Opinion in Structural Biology* **10** 4
[2] Thomsen R 2003 *Conference: Evolutionary Computation* **4**
[3] de Magalhães C S, Barbosa H J C and Dardenne L E 2004 *Genet. Mol. Biol.* **27** 4
[4] López-Camacho E, García Godoy M J, García-Nieto J, Nebro A J and Aldana-Montes J F 2015 *Appl. Soft Comput. J* **28** 379-393
[5] Reddy B R and Reddy L P 2013 *International Journal of Innovative Technology and Research* **1** p 615-616
[6] Korb O, Stutzle T and Exner T E 2007 *Swarm Intel* **1** p 115-134
[7] Mahdiyah U, Imah E M and Irawan M I 2016 *Contemp. Eng. Sci.* **9** 16
[8] Funkhouser T 2007 *Protein-Ligand Docking Methods* (New Jersey: Princeton University)
[9] Yang X S and He X 2013 *Int. J. Swarm Intelligence* **1** 1
[10] Nurhayati A P D, Pratiwi R, Wahyuono S, Fadlan A and Syamsudin 2014 *J. Adv. Bot. Zool.* **2** 1
[11] Nurhayati A P D, Santoso N, Setiawan E and Lianingsih F 2017 *International Journal of Drug Discovery* 8 615-616
[12] Yang K, Hitomi M and Stacey D W 2006 *Cell Division* **1**
[13] Giordano A and Normanno N 2009 *Breast Cancer in The Post-Genomic Era* (USA: Humana Press)
[14] López-Camacho E, García Godoy M J, García-Nieto J, Nebro A J, and Aldana-Montes J F 2015 *Appl. Soft Comput. J.* **28** 379-393