PAPER • OPEN ACCESS

Bat Algorithm for Solving Molecular Docking of Alkaloid Compound SA2014 Towards Cyclin D1 Protein in Cancer

View the article online for updates and enhancements.

Bat Algorithm for Solving Molecular Docking of Alkaloid Compound SA2014 Towards Cyclin D1 Protein in Cancer

Fedric Fernando[1], Mohammad Isa Irawan[1], Arif Fadlan[2]

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

fedricfernando43@gmail.com,mii@its.ac.id, afadlan@chem.its.ac.id

Abstract. Bioinformatics is an interdisciplinary field that combines biology, computer science, information engineering, mathematics, and statistic to analyze and interpret biological data Recently, bioinformatics had an important role in drug discovery. One of the steps for drug discovery is molecular docking. Molecular docking mimics the interaction between ligand and the target protein for in-vitro testing. Solving molecular docking problem isn't an easy task, because molecular docking involves many degrees of freedom. A lot of methods had been applied for this problem, one of them is artificial intelligence. This research will study the usage of bat algorithm in solving the molecular docking problem. Bat algorithm is an algorithm tha uses the characteristic of the bats in searching prey. The research will solve the molecular docking of alkaloid compound SA2014 towards cyclin D1 protein in cancer. Alkaloid compound SA2014 is isolated from marine sponge *Cinachyrella anomala*. The objective function for thi problem will be to minimize the binding energy, the lower energy means the bound of protein and ligand will be stronger. We use root mean squared deviation (RMSD) of protein structures to check the validation of bat algorithm. For validation, we used structure 3ptb and 2cpp. The validation shows that the bat algorithm is a valid method to solve the molecular docking problem because of the RMSD is not over 2Å and the free binding energy is negative. For docking SA2014 towards cyclin D1, bat algorithm shows a negative value -2.217.

1. Introduction

Biological computation, better known as bioinformatics, is a combination of biology and computation that uses application from computing tool and analysis to capture and interpret biological data Computation methods in biology have an important role in drug discovery in recent years [1]. Molecula docking is one of the methods to help in drug discovery. Molecular docking is a structure-based drug discovery that most often used because of its ability to predict the conformation of ligand and its target with high accuracy [2].

Molecular docking mimics the interaction of the ligand with the target protein in the in-vitro test [3] Molecular docking is a hard problem because it involves many degrees of freedom, so the developmen of efficient methods and algorithms will be very useful in the design of new drugs [4]. Here are some artificial intelligence that have been used in molecular docking, i.e. extreme learning machine [5]t random forest [6], particle swarm optimization [7], genetic algorithm, differential evolution [8], and harmony search [9].

This paper uses bat algorithm for solving molecular docking problem. Bat algorithm uses characteristic of the bat in searching for preys. The case for this paper is molecular docking of alkaloid compound SA2014 from *Cinachyrella anomala* towards cyclin D1 protein.

2. Materials and Methods

2.1. Protein and ligand

We used cyclin D1 as the protein and SA2014 as the ligand. Cyclin D1 is one of the positive regulators in the cell cycle. Cyclin D isn't expressed periodically, but will always be synthesized as long there's a growth factor. In the cell cycle, cyclin D1 not only has a role in G1 phase. In G2 phase, the level of cyclin D1 tends to increase, and in S phase, the level of cyclin D1 tends to decrease [10]. Figure 1 shows the form of cyclin D1.
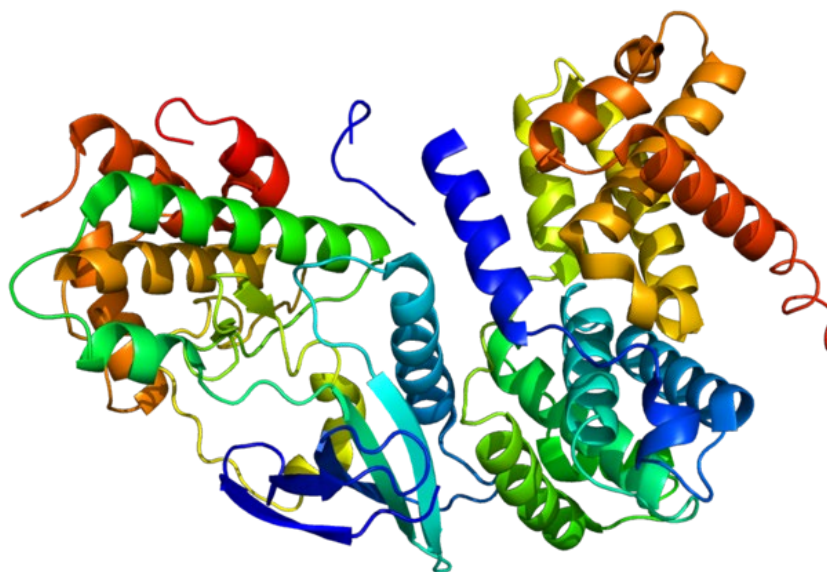

Figure 1. Cyclin D1[11].

SA2014 is derivate from a group of *Cinachyramine* and belongs to alkaloid compound groups. Alkaloid is a compound that has nitrogen and usually has a cycle form. This compound is isolated from a sea anemone *Cinachyrella anomala*. The formulation is $C_{10}H_{13}N_3O$ with structure name 1,4,9-triazatricyclo[7,3,1,0]trideca-3,5(13),10-trien-8-ol [12]. Figure 2 shows the visualization of the alkaloid compound SA2014.
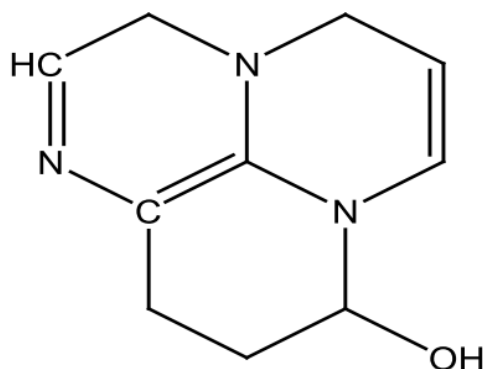

Figure 2. Structure of SA2014 compound.

2.2. Bat algorithm

Bat algorithm (BA) was introduced by Xin-She Yang in 2010. Bat algorithm is an evolutionary algorithm that mimics bat's characteristic in searching for prey. Here are some rules for BA that defined [13][14]:

- Every bat uses echolocation. Echolocation is an ability to find prey using high-frequency pulse that produced by the bat. When the pulse arrived at the prey, the pulse will come back to the bat, so that bat will know the distance of the prey. Figure 3 shows the echolocation [15].
- Every bat flies randomly with velocity v_i in position x_i, with frequency f_i, amplitude λ_i, loudness A_i. They can adjust the amplitude (or frequency) automatically from the emitted waves with emission rate $r \in [0,1]$, depending on the distance from the target.
- Even though the loudness is variated, they are assumed to be variate from a big positive number A_0 to a small constant A_{min}
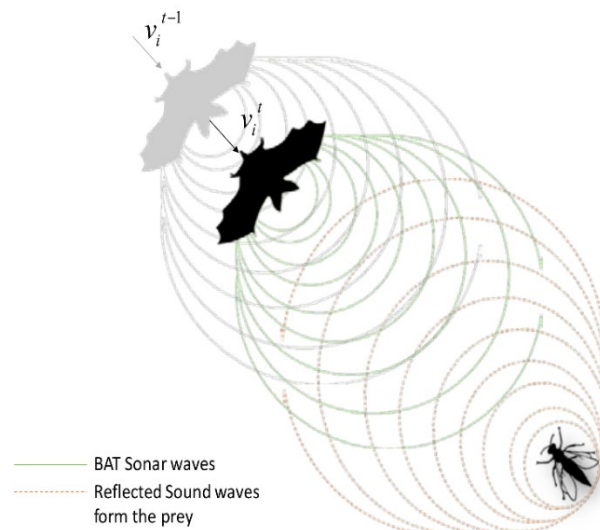

Figure 3. Echolocation.

The solution from each iteration can be written as:

$$f_i = f_{min} + (f_{max} - f_{min})\beta \tag{1}$$

$$v_i^{t+1} = v_i^t + (x_i^t - x_*)f_i \tag{2}$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{3}$$

with:
- $\beta \in [0,1]$, is a random value from a uniform distribution
- x_* is the location of the optimal solution at that current time, compared to the every bat in the current time/iteration
- $\lambda_i f_i$ produced velocity increment, f_i (or λ_i) is used to adjust the velocity change, and to fix other factor λ_i (or f_i), depending on what the problem is.
- Adding a random walk using:

$$x_{new} = x_{old} + \epsilon A^t \tag{4}$$

where:
ϵ = random vector from $[-1,1]$
A^t = average loudness at time t
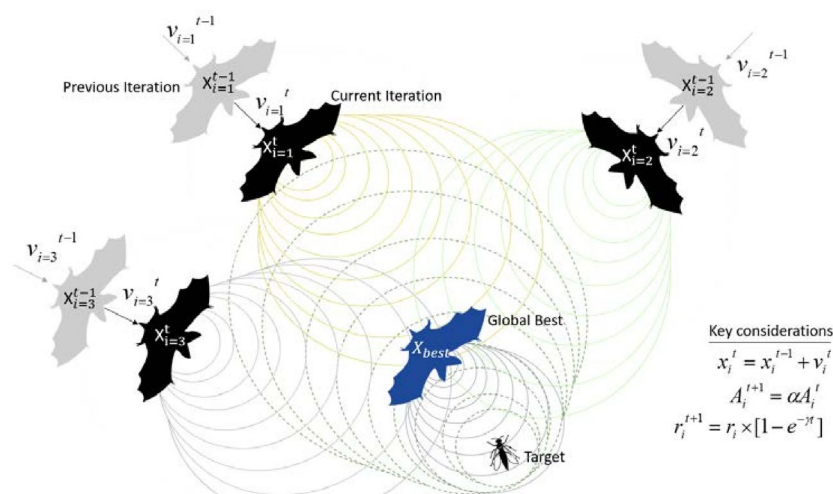Figure 4 shows the mathematical representation of the BA

Figure 4. Mathematical representation of BA[15].

Generally, loudness will tend to decrease if the bat is closer to the prey, and the pulse emission will tend to increase. Therefore:

$$A_i^{t+1} = \alpha A_i^t, r_i^{t+1} = r_i^0(1 - \exp(-yt)),$$
(5)

with:
$0 < \alpha < 1$ and $y > 0$

3. Implementation

3.1. Objective Function

To combine molecular docking and BA, we need some kind of objective function that can define what good docking is. We can use free binding energy as the objective function. From free binding energy, we can determine if a docking process is happening or not. Without energy from outside, a system will tend to go to its lowest free binding energy [16]. A positive value of free binding energy shows no bonding, while a negative value shows there is a bonding. The bigger negative value of free binding energy, the bigger energy is needed to break the bonding.

Free binding energy can be calculated by reducing the energy while bounded with energy when unbounded and then add with torsional energy when the ligand changed from unbound to bound [17]. Therefore, the formula is:

$$\Delta G = \Delta G_{bound} - \Delta G_{unbound} + N_{tor}$$
(6)

where:
ΔG : free energy
ΔG_{bound} : mechanical energy when compounds are bounded
$\Delta G_{unbound}$: mechanical energy when compounds are unbounded
N_{tor} : number of ligand's torsional point

Autodock uses a grid-based approach to approximate the energy calculations used by the energy function. During the evaluation of a candidate conformation, the grids are used as look-up tables storing the value used in the calculation, thus making overall docking very fast [18].

3.2. Bat algorithm for molecular docking

The position of the bat will show the solution to the given problem. In this case, we need a representation that can shows the result from molecular docking. We represent a sequence of values, with the 3 first values are ligand translation location (variable x , y , z), 4 values after that are the quaternion location (variable x , y , z , w), and the rests are rotation angle of each ligand point that're rotatable.
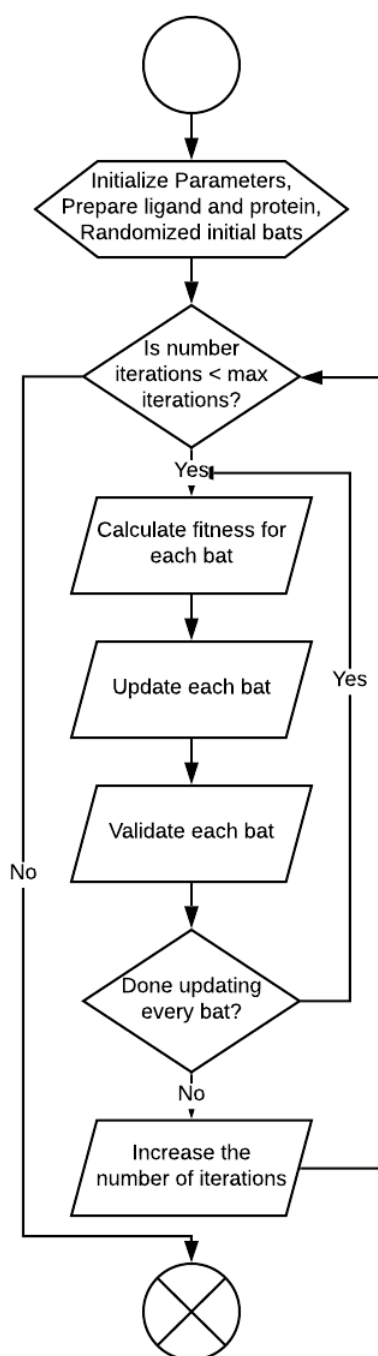

Figure 5. Bat algorithm for molecular docking.

For variable x, y, z from translation, we used Å (angstrom) or 10^{-10} meter. For variable x,y,z from quaternion, the parameters lie within [-1,1]. While for w and rotation angle of ligand point, we use degree, with value variates from $[-180,180]$.

We use the algorithm in Figure 5 in this paper. Started from initializing the parameters, prepare ligand and protein, and randomize initial bats. For prepare ligand and protein, we use data from protein data bank (PDB) and prepare it in software Chimera 1.13.1 [19] and AutoDock 4.2 [17]. We stored the value into the database as the look-up table as defined before. In bats initialization, we randomized the bats between the boundary value from AutoDock.

After we prepared everything, for each iteration, and for each bat. We calculate fitness, update, and validate the bats. In updating part, we do equation (1) – (3) for each bat, and then do some random walk for several bats that have pulse emission lower than a random number. In the validation part, we check every bat. First is to validate the solution, is the new solution is in the boundary or not, if not, we randomize that out-of-bound value to a new one. Second, we consider to update the bat or not. If a random value is lower than the loudness and the new fitness is better than the old one, we calculate new loudness and pulse first and then update the bat [13].

3.3. The parameters for bat algorithm
Table 1 shows the parameters which are used for BA.

Table 1. Parameters of BA

Parameter	Value
Population size	20
Number of iterations	1000
Loudness (A_0)	0.5
Pulse emission (r_0)	0.5
Frequency minimum (f_{min})	0
Frequency maximum (f_{max})	2
α	1
γ	0

We used α and γ as defined above to makes loudness and pulse emission become a constant. Loudness and pulse emission in bat algorithm can be a constant value [13].

4. Results and Discussions

4.1. Validation
To check whenever the BA is able to be used in molecular docking problem or not, we tried some crystallized proteins from PDB, remove the ligand and redocking it again using BA. The first structure is 3ptb [20]. 3ptb is a code for crystal structure with protein β-Trypsin and ligand benzamidine. Figure 6 and 7 show the result of molecular docking of structure 3ptb using BA. Figure 6 shows the best fitness (free binding energy) for each iteration and Figure 7 shows the RMSD of the solution with the best fitness. This trial shows a negative free binding energy and a low RMSD value (< 2Å), with the best solution is [-1.357, 13.947, 16.920, 0.784, -0.4773, 0.0736, 49.375, 52.167]. To be considered as a valid solution, the RMSD value must be lower than 2Å and the free binding energy must have a negative value [21]. Therefore, the result in this trial can be a valid prediction of molecular docking of structure 3ptb.

Next one is structure 2cpp [22]. 2cpp is a code for crystal structure with protein Cytochrome P-450$_{cam}$ and ligand camphor. Figure 8 and 9 show the result of molecular docking of structure 2cpp using BA. Figure 8 shows the fitness and Figure 9 shows the RMSD. The trial shows a negative free binding energy

and a low RMSD value, with the best solution: [46.801, 43.310, 14.094, 0.308, 0.554, -0.968, -61.165]. Therefore, the result in this trial can be a valid prediction of molecular docking of structure 2cpp.
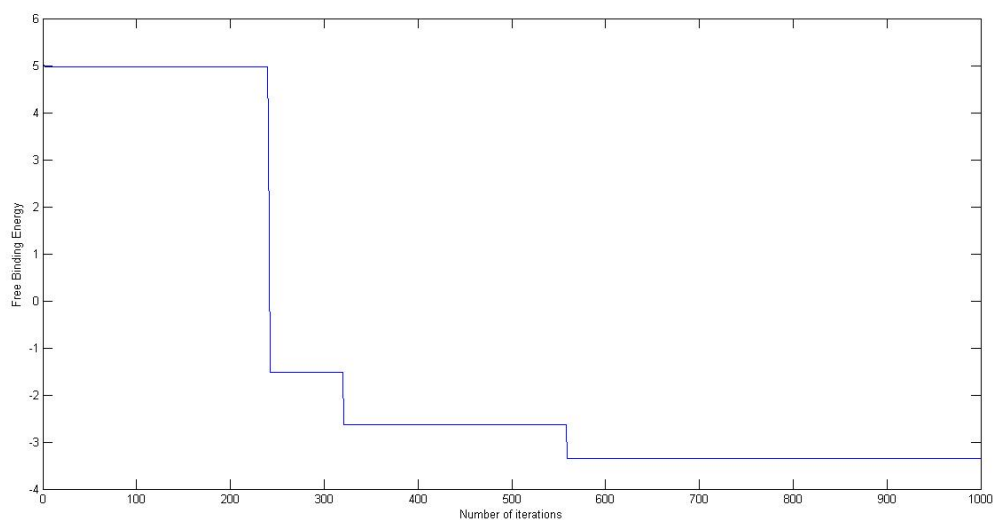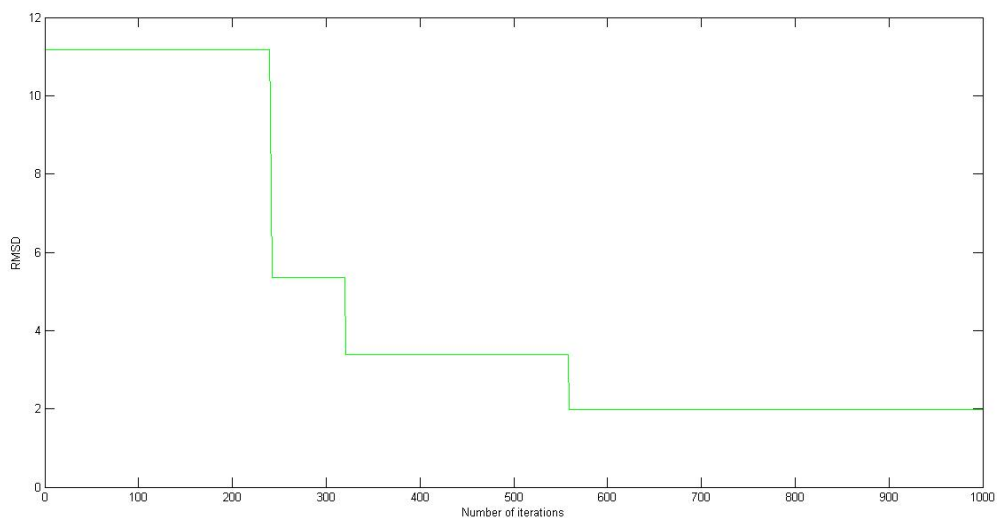

Figure 6. The free binding energy of structure 3ptb.
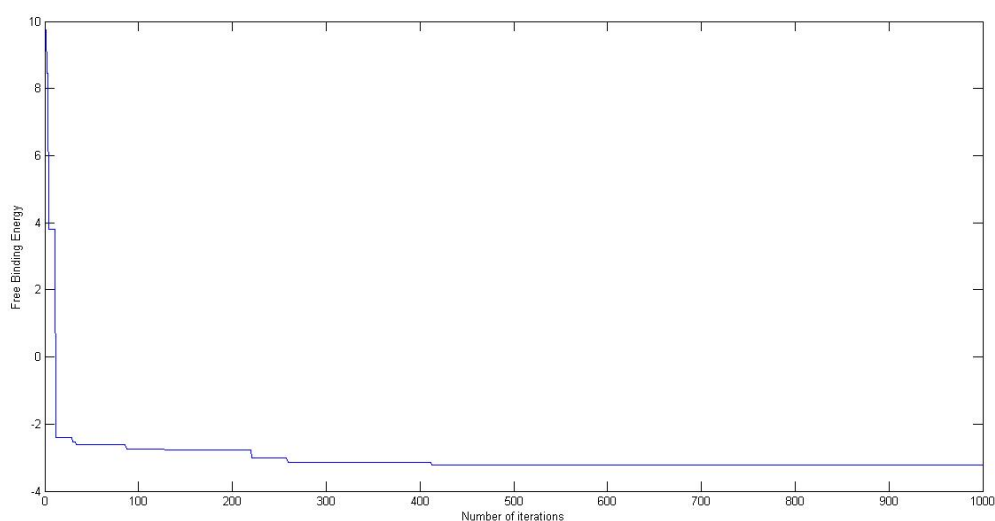

Figure 7. RMSD of structure 3ptb.

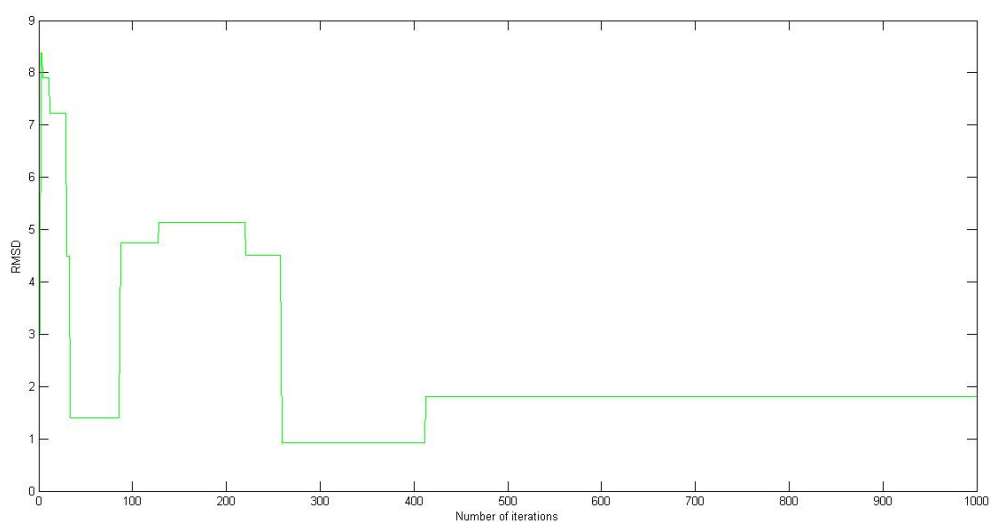**Figure 8.** The free binding energy of structure 2cpp.

Figure 9. RMSD of structure 2cpp.

4.2. Prediction of molecular docking of SA2014 towards cyclin D1

Figure 8 shows the free binding energy of the docking of SA2014 to cyclin D1. The lowest free binding energy score is -2.2165177, with the corresponding solution: [8.006, -1.349, 32.966, -0.805, -0.0294, 0.336, -103.510]
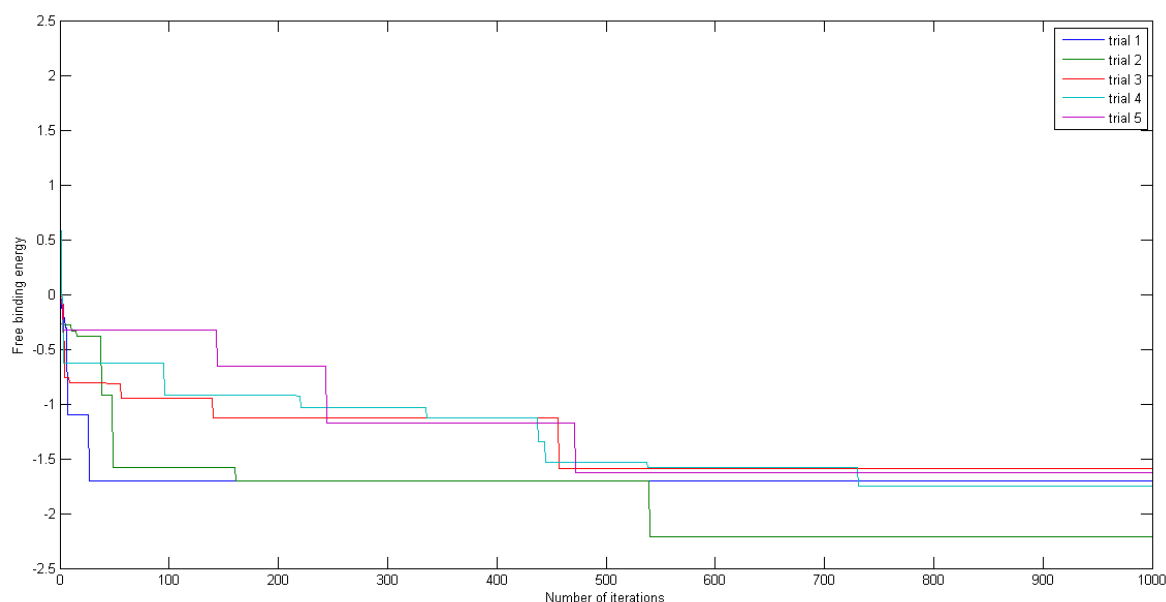
Figure 10. The free binding energy of docking of SA2014 towards cyclin D1.

4.3. Discussion

The molecular docking of structure 3ptb and 2cpp showed negative energy and a small number of RMSD. From these two docking trials, we can conclude that the prediction is quite accurate due to the fact that both predictions show negative value for fitness and less than 2Å. For the trial for 2cpp, we can see an up and down in RMSD. The smaller RMSD means that the prediction is more similar to the real structure. Even though the RMSD is smaller, doesn't mean that the free binding energy is smaller, like in Figure 9.

From Figure 6 to Figure 10, we can see that the free binding energy quite rarely changes. That's because of the BA might have some flaws in these trials. BA doesn't have a step likes mutation in genetic algorithm, which makes BA will only searching a new solution from a random walk. Even after we added a rule where a new random value to replace the current value if the current value is out of bounds, we still hardly get a new solution. Other than that, if the new fitness is not greater than before, BA will not update the solution, that's why random number will not easily appear in BA.

5. Conclusions

We conclude that bat algorithm is able to solve molecular docking. The BA for molecular docking structure 3ptb and 2cpp showed negative free binding energy and RMSD lower than 2Å, For a better result, we suggest to add more iterations and try a new set of parameter.

We predict the location of ligand SA2014 using bat algorithm. The predicted location is [8.006, -1.349, 32.966, -0.805, -0.0294, 0.336, -103.510]. SA2014 has no torsional point, so the solution only has a translational location and quaternion location.

BA itself doesn't have a step which generates a new solution easily, we suggest to combine BA with another algorithm in order to create a better method for molecular docking.

References

[1]   W. L. Jorgensen 2004 *Science* **303** 5665
[2]   L. G. Ferreira, R. N. Dos Santos, G. Oliva, and A. D. Andricopulo 2015 *Molecules* **20** 7
[3]   P. J. Gane and P. M. Dean 2000 *Curr. Opin. Struct. Biol.* **10** 4
[4]   C. S. de Magalhães, H. J. C. Barbosa, and L. E. Dardenne 2004 *Genet. Mol. Biol.* **27** 4
[5]   U. Mahdiyah, E. M. Imah, and M. I. Irawan 2016 *Contemp. Eng. Sci.* **9** 16
[6]   P. J. Ballester and J. B. O. Mitchell 2010 *Bioinformatics* **26** 9
[7]   Y. Liu, W. Li, and R. Ma 2012 *Int. J. Biomath.* **5** 5
[8]   E. López-Camacho, M. J. García Godoy, J. García-Nieto, A. J. Nebro, and J. F. Aldana-
         Montes 2015 *Appl. Soft Comput. J.* **28** 379–393
[9]   D. R. Tobergte and S. Curtis 2013 *J. Innov. Technol. Res.* **1** 6
[10]   K. Yang, M. Hitomi, and D. W. Stacey 2006 *Cell Div.* **1** p 1–8
[11]   P. J. Day *et al.*, 2009 *Proc. Natl. Acad. Sci.* **106** 11
[12]   A. P. . Nurhayati, R. Pratiwi, S. Wahyuono, A. Fadlan, and Syamsudin 2014 *J. Adv. Bot.
         Zool.* **2** 1
[13]   X. S. Yang 2010 *Studies in Computational Intelligence* **284** p 65–74
[14]   X. S. Yang and A. H. Gandomi 2012 *Eng. Comput. (Swansea, Wales)* **29** 5
[15]   M. Seyedmahmoudian *et al.* 2018 *Sustainability* **10** 5
[16]   V. Gapsys, S. Michielssens, J. H. Peters, B. L. de Groot, and H. Leonov 2015 Calculation of
         free binding energies *Molecular Modeling of Proteins* (New York: Springer
         Science+Business Media) p 173–209
[17]   G. M. Morris *et al.* 2009 *J. Comput. Chem.* **30** 16
[18]   D. S. Goodsell and A. J. Olson 1990 *Proteins Struct. Funct. Bioinforma* **8** 3
[19]   E. F. Pettersen *et al.* 2004 UCSF Chimera — A Visualization System for Exploratory Research
         and Analysis **25** p 1605–1612
[20]   M. Marquart, J. Walter, J. Deisenhofer, W. Bode, and R. Huber 1983 *Acta Crystallogr. Sect. B*
         39 4
[21]   O. Trott and A. J. Olson 2010 *J. Comput. Chem.* **31** 455–461
[22]   T. L. Poulos, B. C. Finzel, and A. J. Howard 1987 *J. Mol. Biol.* **195** 3