International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

A Hybrid DELGA Algorithm for Protein Ligand Docking

E. Pavithra Vishalini, D. Ramyachitra, P. Lakshmi

Abstract*: Protein-ligand docking is a computational molecular modeling method that is used in drug design to predict the optimal binding pose between the ligand and receptor. AutoDock is an open-source freeware program used to predict docking poses. It uses LGA) Lamarckian genetic algorithm to enumerate the binding energy. In this research work, we proposed an approach of hybrid Differential evolution base Lamarckian genetic (DELGA) algorithm to calculate the lowest binding energy. The experiment conducted to compute the 65 molecular instances, the results exposed that our approach predicts lowest docking energy with minimum root mean square deviation (RMSD) in comparison to the LGA, SA and PSO algorithms.

Keywords: Molecular Docking, Optimization algorithms, Binding Energy, Evolutionary algorithms.

I. INTRODUCTION

Large molecules composed of one or more chains of 20 amino acids in a specific order. Protein an essential compound on skin, muscles, bones and the whole. Proteins are mandatory for the function, and structure, of cells in the body, organs, and tissues. Each protein has unique functions. An ion, molecule, or molecular group that binds to other molecules to form a larger complex [31,32]. Molecular docking is computational modeling that predicts the 3D structure of any complex used in bioinformatics in the process of drug design to save laboratory experiment time [19]. Molecular docking evaluates different possible solution [1,7]. An optimization algorithm used, it spot docking conformations with the least binding energy [13,27]. Hence, the objective to use proposed an optimization procedure to generate the best binding state characterized by least bind energy with different metaheuristics [20]. AutoDock is a one of the bioinformatics computational program used to predict the interaction of ligands with macromolecular targets depending upon binding properties of ligand and target [25]. It uses many techniques like the Simulated Annealing, Monte Carlo, and a hybrid local search GA Lamarckian Genetic Algorithm [28]. Using the advantages of both DE and LGA, we proposed a differential evolution based genetic algorithm DELGA. It performs a global search over the large conformational space of a ligand using DE and simultaneously LGA maintains a good local solution to optimize the binding pose of the ligand. Our search strategy uses two metaheuristic algorithms in the dimensionality search process with the least bind energy.This technique is different from the above-mentioned docking methods that were used.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author
E. Pavithra Vishalini *, Department of Computer Science, Bharathiar University, Coimbatore. Email: pavishalini@gmail.com
D. Ramyachitra, Department of Computer Science, Bharathiar University, Coimbatore. Email: jaichitra1@yahoo.co.in
P. Lakshmi, Department of Computer Science, Bharathiar University, Coimbatore. Email: visalaks@gmail.com

II. METAHEURISTICS

Metaheuristics techniques applied in a challenge to optimize the best solutions to overcome the complexity and dimensionality problems occur when finding a solution to search space in structural bioinformatics. Recently, many metaheuristics have been developed for the Molecular Docking problem [21]. The frequent search method used in computational molecular docking is Genetic Algorithms with local search Lamarckian evolution (LGA), Differential Evolution (DE), and Particle Swarm Optimization (PSO). These methods were applied to find the minimized binding energy of the protein-ligand complex by changing the ligand's orientation.

A. Simulated Annealing (SA)

Simulated annealing performs a global search and local search varies on high and low temperature respectively. Simulated annealing is fundamentally chosen as the random move instead of picking the best move also referred to as hill climbing. SA applied a random move in local search by a normal distribution and in the genetic algorithm by a Cauchy distribution[14,29].

B. Lamarckian genetic algorithm (LGA)

The hybrid of the Genetic algorithm with local search (LS) identified as Lamarckian genetic algorithm (LGA). In LGA, the energy is calculated from the ligand's coordinates, which together form its phenotype [22]. Basically, in GA under the mapping phase, it transforms a molecule's genotypic state variables into the corresponding set of phenotypic atomic coordinates [30]. While in LGA the genotype variables need not be inverted back LGA.[12,16]The procedure for the LGA algorithm presented in Fig 1.

Psuedocode of LGA algorithm.
INITIALIZE the initial binding population P(x);
EVALUATE the binding energy at each position in P(x);
while(terminal condition) do
Begin
t = t + 1;
/* Mutation*/
Apply mutation at each bind position in M(x) and endorse the resulting position according the probability Pm;
/* Crossover*/
Select two position and apply crossover Pc until L(x) reached population size
/*Lamarckian Evolution*/
Apply p-displacement on each position in L(x)
and form the population C(x);
EVALUATE: the energy of each position in C(x);
end
end

Fig. 1.Pseudocode of LGA algorithms.

C. Particle Swarm optimization (PSO)

Particle Swarm optimization is a swarm intelligence evolutionary approach was based on simulated behavior of birds searching the nearest food solution. Here the bird or particles are considered as an individual solution in the search space[11,23]. At every iteration, it evaluates the fitness value. The fitness value of the Docking problem is considered to be the least binding energy of the protein-ligand complex [3]. In PSO there are two variables Pbest and Guest which is the personal best position and global best position respectively. The particle changes its velocity when moving the position in search space . After updating the velocity and position at each iteration the fitness energy should be calculated. At each generation, the Pbest and Gbest variable values get updated[8]. The procedure for the PSO algorithm presented in Fig 2.

Psuedocode of PSO algorithm.
Initialize the initial protein ligand binding position xij
Initialize confidence of position c1, c2, velocity vij is the speed of changing position particle and r1 and r2 are vector (0,1).
Evaluate fitness value F(x) is the binding energy using eq. Ebind = Einter + Eintra
Calculate Pbest value while
number of iteration condition fail do
calculate velocity and orientation of ligand i
Evaluate fitness value for new position F(x)
If F(xi)pbest,i) /* *Pbest is the best local new position of ligand* */
then pbest,i = xi
end if
If F(xi) gbest,i) /* *Gbest is the global best position of ligand* */
then gbest,i = xi
end if
End while

Fig. 2.Pseudocode of PSO algorithms.

D. Proposed Method

In DE the evolutionary procedure the binding position is initialized and evaluate the fitness energy value for every position until the terminal condition meets an offspring is created [17]. By comparing the offspring and parent the best one is replaced as a parent [24]. A hybrid Differential evolution based Lamarckian Genetic algorithm called DELGA is proposed in this research. For each population, the position is evaluated by using LGA in a local search. And DE performs for a global component [15]. For each new position, the fitness energy calculated [26]. The fitness energy with minimum value is updated as a parent . The procedure for the DELGA algorithm presented in Fig 3. The protein-ligand binding free energy ΔG is given by:

$$\Delta G = \sum_i w_i^{vdW} E_i^{vdW} + \sum_i w_i^{ele} E_i^{ele} + c \qquad \ldots (2.1)$$

where E_i^{vdw} is the the van der Waals energy among protein and ligand,
E_i^{ele} is the electrostatic interaction energy among protein I residue and ligand,
c is a constant.
w_i^{vdW} and w_i^{ele} are parameters to be resolute to reproduce the experimental data.

rmsd the objective which is the deviation of ligand known position and computational docked position distance.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \min \; r_2^{ij}} \qquad \ldots (2.2)$$

where N is the maximum count of atom, r^{ij} is the intermolecular distance between the atom with the same element data type as i in structure.

Psuedocode of DELGA algorithm.
INITIALIZE the initial binding population P(x);
EVALUATE the binding energy at each position in P(x);
while(terminal condition) do
Begin
t = t + 1;
/* Mutation*/
Apply mutation at each bind position in M(x) and endorse the resulting position according the probability Pm;
/* Crossover*/
Select two position and apply crossover Pc until L(x) reached population size
/*Lamarckian Evolution*/
Apply p-displacement on each position in L(x)
and form the population C(x);
/ * mutation*/
For each position xi pick three randomly different orientation x1, x2, x3
Compute the new position
Evaluate objective function F(xm) using eq. Ebind = Einter + Eintra
/ * Crossover*/
Crossing the mutant dimension i , j individual randomly
Evaluate objective function F(xc) using eq. Ebind = Einter + Eintra
/ * Selection*/
If F(xm) < F(xc)
F(xi) = F(xm)
Else
F(xi) = F(xc)
EVALUATE: the energy of each position in C(t);
Perform replacement of least evaluated bind energy.
End
Output the optimal solution
end

Fig. 3.Pseudocode Of Delga Algorithms.

III. METHODS FOR DOCKING EXPERIMENT

A. Software

AutoDock: It is open-source molecular modeling software for protein-ligand docking, which calculates the pose of ligand bind to the target or protein grid area [18]. A grid area is the physical description of the docking site. Least predicted binding energy gives better ligand affinity. In this work, all dockings were performed using Autodock4.2.6[12].

MGLTools: The proteins and ligands preprocessed using the MGLTools.

Python: Python 2.5 was used as the interpreter programming language for developing the code.

B. Data preparation

The preprocessed protein and ligand structures were saved in PDBQT file format using AutoDock version 4.2.6. Autodock Tools used to preprocess the ligand and protein used for docking, setting up the grid parameter file(GPF) and a Docking parameter file (PDF) for docking. Energy grid maps were calculated using AutoGrid grid size has a length cube edge 22.5 ˚A was applied

C. Parameterization

Representation of ligand in AutoDock, the solution is encoded by seven parameters 7 + or variables. The first three are x,y and z refer the position of ligand translation; whereas the orientation of ligand are represented by the next four variables; remaining + mentor values refer to the torsion angles needed to describe the ligand pose. We chose 65 protein-ligand complexes from the PDB database with the parameter of 27,000 iterations, the initial population is 50 individual for all five algorithms. The crossover rate and scalar number are 0.70 and 0.80 respectively for both DE and DELGA procedures. For LGA and the proposed DELGA, the mutation rate is 0.02, crossover rate 0.70. the number of iterations 200. The other parameters provided by the default setting were the same as in AutoDock[12].

IV. RESULTS AND DISCUSSION

For each complex, fifteen best values resolved by every tested algorithm are utilized as experimental data. The test results of the comparison of algorithms are shown in Table 4.1. The resultant table indicates that the hybrid DE based LGA performs best than the other tested procedure for finding the lowest binding energy in the docking problem.

The performance metrics used to analyze the problem are binding energy and RMSD for benchmark complex dataset [10]. In this research LGA with a random walk for the mutation, the operation was used to perform local exploration of promising regions, while a DE process was used to perform global exploration throughout the entire search process. After performing the docking, the solution conformations were thoroughly compared to find similarities and were gathered accordingly. Achieved results of flexible docking are presented. Comparison of the least docking energy with corresponding RMSD value of DELGA, PSO, LGA, SA, and DE represented in Fig 5.

A. Energy analysis

Optimization algorithms are directed only by energy function, and it's search ability can be assessed in terms of docked energy. Molecular docking simulation approaches trust on the calculation of least binding energy of the predicted complex, nearer it is to the native state, low binding energy specifies the enhanced result. Table 4.1 shows the results for the best docking energy and corresponding root mean square deviation (RMSD) value of 65 docking runs using four algorithms. Fig 5. shows the comparison of the binding energy values with respect to five various algorithms Here, the RMSD is a measure of comparison between the experimental ligand position in the residue and the experimented result of docking ligand position. As observed from Fig 5, the DELGA algorithm shows the maximum number of finest solutions achieving least binding energy in 54 beyond 65 molecules, followed by the PSO with 7, and SA with 4.

B. RMSD analysis

The RMSD values attained by DELGA were lower than those obtained by LGA. As shown in fig 4.2, in terms of the corresponding RMSD, the DELGA algorithm also shows the maximum number of finest solutions, attaining the least RMSD value in 45 beyond 73 molecules, followed by the LGA with 6, SA with 3, and PSO with 4 and DE with 7. This also validates the DELGA algorithm is appropriate for conduct high dimensional complex docking problems.

V. CONCLUSION

Protein-ligand docking is a major computational phase in drug discovery studies. As high conformational search space, the complexity and dimensionality are the major issues in protein-ligand docking. To overcome such issues in this research, two metaheuristic algorithms utilized in the search process. To summarize, DELGA appears well suited for more accurate and fast evaluation for high dimensional docking. PSO and SA perform well in finding energy but are not efficient for RMSD which is the deviation of the experimented and actual value of orientation accuracy estimation. LGA is actually well match able for low flexibility evaluation of ligand. Based on experiment results, it is determined that the DELGA algorithm provides an another method for a molecular docking tool.

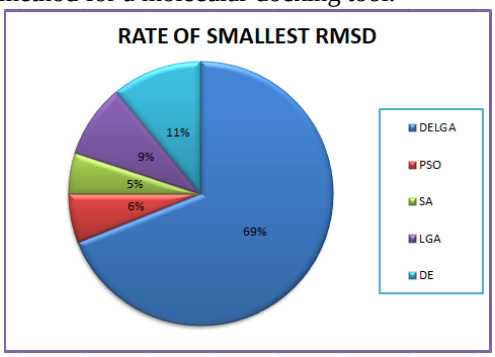

Fig. 4.Comparison of the RMSD values with respect to five various algorithms for all complexes.

Fig. 5.TABLE I. Comparison Of The Lowest Docking Energy And The Corresponding Rmsd Value Of DELGA, PSO, LGA, SA, And DE

REFERENCES

1. Vaclav., Krel., Michal., Pavel., 2016. “Exponential repulsion improves structural predictability of molecular docking”. J. Comput. Chem. 37, pp.2485–2494.
2. X Xu,Z Ma, R Duan, X Zou, 2019. “Predicting protein-ligand binding modes for CELPP and GC3;workflows and insight”. J. Comput aided mol desgn 33,pp.364-374.
3. Fan, Feng, Cheng, Sun,Ge-Fei,Guang, 2019. “ACID: a free tool for drug repurposing using consensus inverse docking stratergy”, J.Chem 11,73.
4. Mohammad Rahimian, 2019. “Measuring efficiency in DEA by differential evolution algorithm”. Annals of optimization theory and practice 2,pp.19-26.
5. Emrah Atilgan, Jianjun Hu, 2011, “Improving protein docking using sustainable genetic algorithms”.Int J Comp Inoform Sys Ind Manag 3,pp.248-255.
6. HM Berman, Bourne, Westbrook, Christine, 2003, “The protein databank”. Protein Structure, pp.394–410.
7. Daniel, Israel, William, 2019, “Computation of protein-ligand binding free energies using quantum mechanical bespoke force fields”. Med Chem Comm.
8. James Kennedy, Rui Mendes , 2002, “Population structure and particle swarm performance” Vol. 2, Proc 2002 Congress on Evolutionary Computation CEC'02 (Cat. No.02TH8600), pp. 1671–1676.
9. Yoshifumi Fukunishi, Yoshiaki Mikami, Haruki Nakamura, 2003, “The filling potential method: A method for estimating the free energy surface for protein-ligand docking”. J. Physical. Chemistry. B, pp.13201-13210.
10. Geraldo, Alessandro, 2019, “Comparitive Analysis of electrostatic models for ligand Docking”. BioRxiv,577643.
11. Namasivayam, Gunther, 2007, “PSO@AUTODOCK: A fast flexible molecular docking program based on swarm intelligence”. Chem. Biol. Drug Des. 70, pp.475–484.
12. Maxwell, Renee, Alan, Venkataraman, Edward, 2005, “Comparision of automated docking programs as virtual screening tools”.J.MedChem, pp.962-976.
13. Jin Li, Ailing Fu, Le Zhang “An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking”. Interdisciplinary Sciences: Computational Life Sciences June 2019, Volume 11, Issue 2, 2019, pp 320–328.
14. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science 220(4598), 671–680 (1983)B. Ramkumar, H.M. Kittur, and P. M. Kannan, “ASIC implementation of a modified faster carry-save adder,” Eur. J. Sci. Res., vol. 42, no. 1, 2010, pp.53–58.
15. Sergio, Antonio, Coimbra,Neves, Martin, Moorthy, 2013, “Protein-ligand docking in the new millennium-a retrospective of 10 years in the field”. Current Medicinal Chemistry, pp.2296-2314.
16. Baniamerian, Bashiri,Reza, 2019, “Modified variable neighbourhood search and genetic algorithm for profitable heterogeneous vehicle routing problem with cross-docking”. App Soft Comp.pp.441-460.
17. Devi , Sathya , Coumar 2015, “Evolutionary algorithms for de novo drug design a survey”. App Soft Comp 27,pp.543–552
18. “Autodock homepage,” http://autodock.scripps.edu/, (Last accessed: 2011.05.07).
19. C. Lipinski,F.Lombardo,B.Dominy,andP.Feeney,“Ex- perimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.” Advanced Drug Delivery Reviews, vol.23,pp.3–25, 1997.
20. J. D.J.Blaney,“A good ligand is hard to find :Automated docking methods.” Perspect. DrugDisc.Des., vol.1,pp.301– 319, 1993.
21. Uehara, S.;Fujimoto,K.J.;Tanaka,S.Protein-liganddockingusingfitnesslearning -basedartificial bee colony with proximity stimuli. Phys. Chem.Chem.Phys. 2015, 17, 16412–16417.
22. Guan, B.;Zhang,C.;Zhao,Y.HIGA:ARunningHistoryInformationGuidedGen eticAlgorithmfor Protein–Ligand Docking. Molecules 2017, 22, 2233.[
23. Teodorovi´c, D.:TransportModelingByMulti-AgentSystems:ASwarmIntellgenceAp proach, Transport.Plan.Technol. 26(4), 289–312(2003)
24. Boyer, D.O., Martnez, C.H., Pedrajas, N.G.: Crossover Operator for Evolutionary Algorithms Based on Population Features. http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/ortizbo- yer05a-html/Ortiz-Boyer.html.
25. Yue, Yang, 2019, “Bayesian active learning for optimization and uncertainty quantification in protein docking”. Arxiv 1902.00067.
26. Xiongwu, Bernard, 2019, “Protein-Protein Docking using map objects”, Mol Docking & Mol Dynamics.
27. I.B. Beloglazova, O.S. Plekhanova, E.V. Katkova, K.D. Rysenkova, D.V. Stambol’skii, V.B. Sulimov, V.A. Tkachuk, Molecular Modeling as a New Approach to the Development of Urokinase Inhibitors . Bulletin of experimental biology and medicine 158 (2015) 700-704, http://dx.doi.org/10.1007/s10517-015-2839-3 .
28. Oshiro Kuntz, 1995, “ Flexible ligand docking using a genetic algorithm” J. Comput-Aided Mol. Design 9,pp.113-130,
29. Plewczynski , Lazniewski, Grotthuss , Rychlewski , Ginalski, 2011, “VoteDock: consensus docking method for prediction of protein–ligand interactions”. J.Com.Chem32(4),pp.568–581.
30. Morris , Goodsell , Halliday , Huey , Hart , Belew , Olson, 2015, “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function”. J Com Chem 19(14),pp1639–1662
31. Stierand, Maab, Rarey, 2006. “Molecular complexes at a glance: automated generation of two-dimensional complex diagrams”. Bioinformatics 22, pp.1710-1716.
32. Jin Li, Ailing Fu, Le Zhang, 2019, “An overview of Scoring function used for protein-ligand interactions in molecular docking”, Int Sci:Comp Life Sci,pp.1-9.

AUTHORS PROFILE

Pavithra Vishalini obtained M.Sc degree in Navarasam Arts and Science College, Bharathiar University, Tamilnadu, India in 2015 and pursuing M.Phil in CS, Bharathiar University, Coimbatore. Published a research Paper in National Conference.

Dr.D.Ramyachitra Assistant professor at Bharathiar University, Coimbatore. She has more than 50 National, International Conferences, Journal Publications in multi aspects such as SCI, SCIE, Science Direct, Web of Science, Scopus, and UGC Care respectively. She is one of a member of CSI. She presented and participated in various workshops, seminars, etc., she has done an UGC-Minor Research Project, 2009-2011, Title: An efficient scheduling strategy for protein sequence analysis on the grid. She had research experience with more than 10 years of research experience and 19 years of teaching experience. She guided M.Phil and Ph.D. Research Scholars.

P.Lakshmi pursuing a Ph.D. in Bharathiar University, Published papers in National and International Conferences. She has two years of teaching experience. Attended Seminars and Workshops.