



Enhanced-Hybrid-Age Layered Population Structure (E-Hybrid-ALPS): A Genetic Algorithm with Adaptive Crossover for Molecular Docking Studies of Drug Discovery Process

Sudha Ramachandra¹, Vinay Chavan²

¹Assistant Professor, Department of Computer Science, Santaji Mahavidyalaya, Nagpur-440015, Maharashtra, India; ²Associate Professor, Department of Computer Science, Seth Kesarimal Porwal College, Kamptee, Nagpur, Maharashtra, India.

ABSTRACT

Objectives: Age Layered Population Structure (ALPS) which introduces time labels into a traditional Genetic Algorithm (GA) is a novel search metaheuristic in overcoming premature convergence. There are two models of ALPS namely generational and steady-state with their own merits and demerits. Present work has been taken up to devise a search algorithm E-Hybrid-ALPS with the combined concepts and advantages of both the models.

Methodology: E-Hybrid-ALPS not only combined the concepts and advantages of both the models but also considered weak individual solutions to the mating pool and adaptively applied the crossover operator.

Search algorithm, a component of the molecular docking tool plays a vital role in the success of molecular docking used in drug discovery. Hence, E-Hybrid-ALPS has been implemented as a search algorithm for molecular docking. The execution was carried out with two receptor-ligand combinations namely receptor CYP2C8 and ligand Chloroquine, a therapeutic option in the treatment of Corona Virus Disease (COVID-19) and also a drug used in the treatment of Malaria and receptor CYP2B6 and ligand Cyclophosphamide a drug used in the treatment of cancer.

Results: E-Hybrid-ALPS generates poses of the ligand in the active site of the receptor, calculates the binding energy of each pose and outputs the pose with the lowest binding energy. The performance was evaluated by comparing it with the widely used molecular docking tools AutoDock and AutoDockVina which employ Lamarckian GA as a search algorithm. Lowest binding energy found by E-Hybrid-ALPS was significantly low as compared to the lowest binding energy found by AutoDock and AutoDockVina

Conclusion: E-Hybrid-ALPS which generates a ligand/drug pose with the lowest binding energy can be implemented as a search algorithm for AutoDock molecular docking tool. This helps the drug discoverer in designing a drug with a better binding affinity as lower binding energies indicate higher binding affinity.

Key Words: Age Layered Population Structure, Generational, Steady State, Weak Individuals, Adaptive Crossover, Molecular Docking, Pose, Binding Energy, Binding Affinity, Drug Potency

General Terms: Algorithm, Breeding, Offspring, Diversity

INTRODUCTION

Genetic algorithm (GA), is a random search technique introduced by John Holland based on the concept of natural evolution. GAs start with a set of randomly generated initial solutions known as population and each solution called chromosome or individual is constituted by various parameters known as genes that are encoded into the chromosome using various encoding techniques. The goodness of each solution

is evaluated by fitness function which is problem-dependent. Based on the fitness value, chromosomes are selected for breeding by various selection methods and selected chromosomes called parents are mated by applying a crossover operator to generate offspring. The set of newly generated offsprings replace parents according to some replacement policy. The cycle of selection, crossover and replacement is iterated till an optimum solution to the given problem evolves or some terminating criteria is reached^{1,2}.

Corresponding Author:

Dr. Sudha Ramachandra, Assistant Professor, Department of Computer Science, Santaji Mahavidyalaya, Nagpur-440015, Maharashtra, India; Email: sudha_ramachandra@yahoo.com

ISSN: 2231-2196 (Print)

ISSN: 0975-5241 (Online)

Received: 03.06.2020

Revised: 28.06.2020

Accepted: 15.07.2020

Published: 08.08.2020

Premature convergence, where iterations of evolutions do not find any significant better solutions is a major problem of GA³. Over the years many approaches have been tried to circumvent the problem of premature convergence and recent research has tried to address the problem by Age Layered Population Structure (ALPS) which uses the notion of age. In ALPS, age is assigned to each individual indicating the duration of its existence in the population. Competition in breeding among members of the population is restricted by dividing the entire population into groups of individuals of the same age called as layer or age layer and an upper age limit will be set to each layer. Each layer undergoes the process of selection, crossover and replacement separately. Different systems are used to set and increment the age of the individuals. Individuals whose age exceed the maximum age limit of the layer are moved to the next higher layer and each age layer evolves independently of others. Periodically, random individuals are added to the youngest layer which helps in maintaining the diversity of the population as they propagate up the layer.

Two versions of ALPS namely generational ALPS and steady-state ALPS are in existence. In generational ALPS, each cycle of selection, crossover and replacement is called as generation and age of an individual is counted in terms of generations⁴ whereas in steady-state ALPS age of an individual is assigned in terms of fitness evaluations⁵. In generational ALPS, 'n' crossovers and replacements are done in an iteration where the value of 'n' depends on crossover probability and all the individuals that did not participate in the cycle of selection, crossover and replacement are eliminated from the population while in steady-state ALPS, two individuals according to selection criteria will be selected in an iteration from the population of layer undergoing reproduction, cross over operator is applied and replacement is done according to a replacement policy. In steady-state models, individuals not participating in the selection and crossover are not removed. Generational ALPS has the advantage of doing crossovers and replacements as per crossover probability in an iteration whereas steady-state ALPS has the advantage of not eliminating the individuals which did not take part in recombination^{4,5}.

The balance between exploration and exploitation is essential for the good behaviour of any GA⁶. Selection methods aim at exploiting the best individuals and crossover methods explore the search space. A good selection method should not only exploit highly fit individuals but should also exploit good parameters of low fit individuals to arrive at the best possible solution.

Crossover operation in any GA is responsible for producing the offspring to explore a much wider area of search space. Crossover operators which differ in their efficiency are selected based on encoding schemes of the genes and the type

of GA application. The performance and the behaviour of the GA can be greatly improved by adaptively using more than one crossover operator^{7,8,9}.

Owing to all these, authors in the present work have devised a search optimization algorithm named as E-Hybrid-ALPS synergizing the concepts and advantages of generational and steady-state models of ALPS with adaptive crossovers.

Drug discovery is an elaborate and challenging task that can be expedited by using computational methods. Drug on entering the human body elicits the therapeutic effects by interacting with the protein at the active site and this interaction is known as drug binding. During various stages of the drug discovery process, drug-protein interactions are modelled and studied by employing fast and economically favorable molecular docking methods using molecular docking tools. Drug binding can occur at various positions, orientations and conformations of the ligand in the active site of the protein known as a pose and the strength of binding interactions is evaluated by computing the binding energy which is a pointer to the binding affinity of the drug. Lower binding energies indicate higher binding affinity and higher binding energies indicate lower binding affinity. Binding affinity is also inversely proportional to drug potency which is the amount of drug required to produce a therapeutic effect^{10,11,12}. Hence, binding affinity and the pose with the lowest binding energy known as best binding pose are very important information for the drug discovery process.

The principal job of molecular docking is to generate various poses of the ligand in the active site of the protein, finding the binding energy of each pose and identifying the pose with the lowest binding energy which is carried out by a search algorithm, one of the two main components of molecular docking tool. Therefore, it is very important to have an efficient search algorithm in a molecular docking tool. Following the pose generation, the scoring function of the molecular docking tool scores a pose that is near to the experimental finding^{13,14,15}.

Most of the drugs that enter human body binds to an enzyme of the family of enzymes called Cytochrome or CYP enzymes. Identifying the ligand or drug and the receptor or protein or enzyme to which the drug can bind are crucial for any molecular docking study.

Corona Virus Disease-19(COVID-19) is a highly infectious and a pandemic. Presently, there are no drugs targeted to this disease. Recently, the World Health Organization (WHO) has proposed various therapeutic option for this disease. Chloroquine a drug which is metabolized by CYP2C8 (receptor) used in the treatment of Malaria has been proposed as one of the therapeutic options for COVID-19^{16,17,18}.

Cancer is a very serious, death-causing disease threatening the human race. It is a large group of disease which can start

in any organ or tissue of the human body. Cyclophosphamide, a drug metabolized by CYP2B6 (receptor) is one of the many drugs used to treat cancer^{19,20,21,22}.

Considering all these, in the present work, the devised algorithm E-Hybrid-ALPS has been employed as a search algorithm for molecular docking of Chloroquine and CYP2C8 as one ligand-receptor combination and Cyclophosphamide and CYP2B6 as another. AutoDock 4 and AutoDockVina are two freely available and widely used molecular docking tools²³. Hence the performance of the devised algorithm has been evaluated by making a comparison against these two tools in terms of lowest binding energy. Same authors in their previous work²⁴ have devised a search algorithm for molecular docking studies combining the concepts of generational and steady state versions of ALPS with a single crossover. In the present work, adaptive crossovers have been used to improve the algorithmic performance and added to this changes have also been incorporated that has contributed significantly in improving the performance of the algorithm. Hence the name of the algorithm in the present work is Enhanced-Hybrid-ALPS or E-Hybrid-ALPS.

E-HYBRID-ALPS MODEL

E-Hybrid-ALPS sets out with a random initial population where the problem specific genes are encoded into the chromosome by real value encoding system. As the devised algorithm is based on the concept of ALPS, individuals are assigned age and competition among individuals in breeding are restricted by grouping individuals of the same age into layers and each of which has an upper age limit. Individuals of the initial population are assigned an age of one and belong to layer zero which is the first layer. The Fibonacci ageing scheme as shown in Table 1 is used to set an upper limit for each age layer.

Table 1: Fibonacci Aging Scheme

Layer number	Max age limit
0	1
1	2
2	3
3	5
4	8

Fitness of each solution is given by an objective function which is problem-dependent. In E-Hybrid-ALPS, each layer separately undergoes reproduction with a restriction that a layer undergoing reproduction can select parents to the mating pool (which consists of individuals which may undergo crossover) from its individuals and also from the individuals belonging to the immediate younger layer but can only re-

place the individuals of its layer. Heeding to this, individuals from both the layers combined are sorted according to fitness values and by maintaining the selection pressure, individuals from the sorted group are selected twice to the mating pool once from a high fit group of individuals and latter from the high-low fit group of individuals. Considering low fit individuals to the mating pool is the novelty of the algorithm. Subsequently, two parents are chosen randomly from the mating pool for crossbreeding. Before undergoing crossover, both the parents are checked gene-wise for equality. If any genetical equality is found among parents, one of the parents will be mutated.

Each of the well-known crossovers for real encoding namely Heuristic, Arithmetic and Binary Simulated Crossover are applied gene wise to generate offsprings. These crossovers are said to generate offsprings with higher diversity⁸. In the Arithmetic crossover, genes of parental chromosomes are linearly combined to generate offsprings according to

$$\text{Offspring1} = \alpha * \text{parent1} + (1 - \alpha) * \text{parent2}$$

$$\text{Offspring2} = (1 - \alpha) * \text{parent1} + \alpha * \text{parent2}$$

Where α is any value between 0 and 1 and plays an important role in deciding dominant parents.

The heuristic crossover uses the fitness value of the parents to find the direction of the search space. Offsprings are calculated as

$$\text{Offspring1} = \text{Best Parent} + \alpha(\text{Best Parent} - \text{Worst Parent})$$

$$\text{Offspring2} = \text{Best Parent}$$

Where the value of α is between 0 and 1

The idea behind Simulated Binary Crossover is to imitate the properties of single-point crossover used in binary encoding. An important property is that the average value of the parents is equal to the average of offsprings. Offsprings are computed as follows.

$$\text{Offspring1} = 0.5 * ((\text{parent1}.x + \text{parent2}.x) - \beta * (\text{parent2}.x - \text{parent1}.x));$$

$$\text{Offspring2} = 0.5 * ((\text{parent1}.x + \text{parent2}.x) + \beta * (\text{parent2}.x - \text{parent1}.x));$$

Where β is any random number.

In this crossover, if the value of β is set to 1, offsprings which are equal to parents will be generated while if β is set to less than 1 offsprings closer to each other than their parents will be generated and if β is set to a value greater than 1, offsprings farther apart than their parents will be generated.

In E-Hybrid-ALPS as in the case of steady-state ALPS, age is counted in terms of evaluations⁵ and after crossover, the age of the parents is incremented according to the formula as shown in Figure 1.

age = 1 + (evalscurrent - evalscreated) / initial population size

where,

evalscurrent indicates the number of binding energy evaluations done so far

evalscreated indicates the number of binding energy evaluations done at the time of the creation of the individual

Figure 1: Formula to Calculate Age.

Offspring gets the age of the parent whose age is maximum among the two and the best offsprings replace parents according to deterministic replacement policy²⁵.

In E-Hybrid-ALPS crossovers are applied adaptively meaning each of the crossover namely Arithmetic, Heuristic and Simulated Binary crossover is applied to the same parents at the same time and the best offspring are added to the population.

As mating pool comprises of individuals once from the high fit group only and latter from high-low fit groups, crossovers and replacements are done separately to both the groups in the same run maintaining the selection pressure which is another newness of the algorithm. E-Hybrid-ALPS employs the concept of generational ALPS by repeating the entire process of selection, crossover and replacement ‘n’ times by maintaining the crossover probability and the concept of steady-state ALPS by not throwing out the individuals which did not take part in reproduction.

After performing reproduction in all the evolved layers, individuals whose age exceed the maximum age limit of the layer are moved to the next higher layer and in this way, various age layers of the population are evolved. Individuals which are moved can replace weak individuals of the next higher layer provided weak individuals have not been recently moved otherwise individuals will be simply added.

After a certain number of evaluations, new random individuals replaces the individuals of the zeroeth layer and the process of reproduction and movement of individuals are repeated for all the evolved layers starting from the zeroeth layer. The entire process of reproduction, movement of individuals and zeroeth layer replacement are all iterated for ‘n’ number of fitness evaluations which is problem dependent and after which final optimized value will be displayed. Diagrammatic representation of the E-Hybrid-ALPS model is as displayed below in Figure 2.

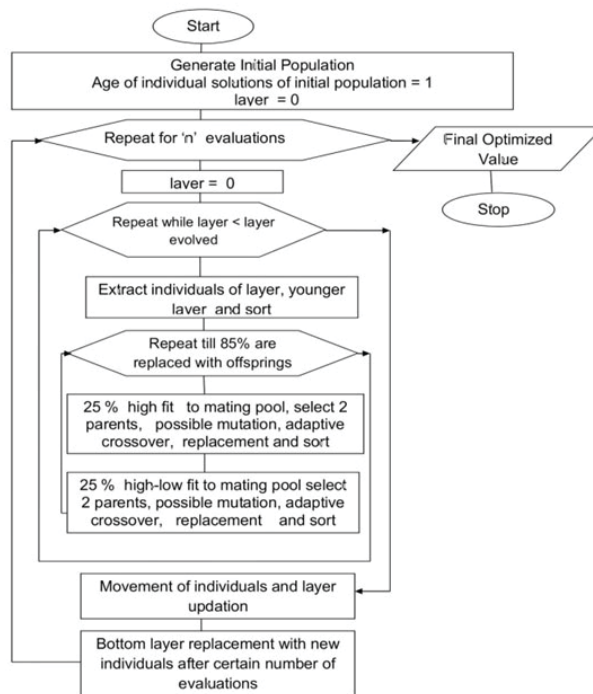


Figure 2: E-Hybrid-ALPS Model.

TOOLS AND METHODS FOR IMPLEMENTING E-HYBRID-ALPS IN MOLECULAR DOCKING

In molecular docking, the 3D structure of the ligand is placed in the active site of protein in various positions, orientations and conformations and each such placement is called a pose. In the present work, E-Hybrid-ALPS has been implemented as a search algorithm for molecular docking studies with CYP2C8 and Chloroquine as one receptor-ligand combination and CYP2B6 and Cyclophosphamide as another. 3D structures of the receptors and ligands are an essential requirement for molecular docking studies. In the present work, 3D structures of the receptors namely CYP2C8 and CYP2B6 are constructed by Homology Modeling²⁶ using Swiss Homology Modeling server^{27, 28,29,30,31,32} by providing protein sequences of each of the receptor which is available in UniProt database³³ as input. 3D structures of the ligands Chloroquine and Cyclophosphamide are downloaded from Chemspider³⁴. To accelerate the process of docking, AutoDock tool employs a grid based docking method by using a grid map for each type of atom present in the ligand along with electrostatic, desolvation and hydrogen bond energies and the same concept has been employed in E-Hybrid-ALPS. A grid as shown in Figure 3, is a 3D lattice with equally spaced grid points centred around the active site of the receptor and each of the grid point stores the potential energy of a probe atom which is due to all atoms of the receptor. AutoGrid tool has been used in the current research to generate grid maps^{35,36}.

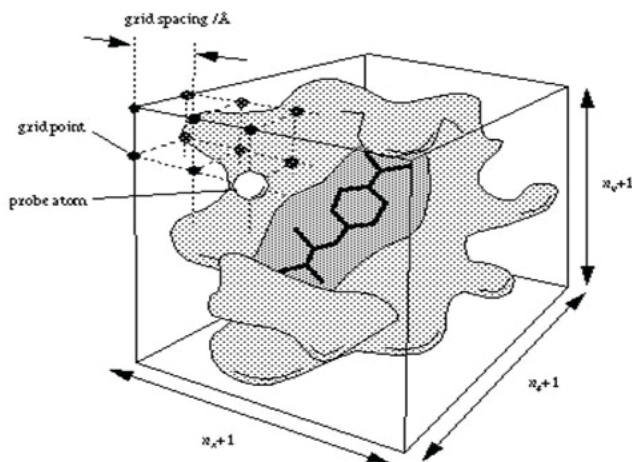


Figure 3: Grid Maps³⁷

After this, an initial random population of 50 individuals is generated where there are three translational genes representing x , y , z coordinates of the ligand within the grid dimensions, four genes representing the orientation of the ligand with random values between -180 to $+180$ degrees and torsional genes depending on the number of torsional angles of the ligand. Ligand Chloroquine has 8 torsional genes while Cyclophosphamide has 5 torsional genes. All the genes are encoded into the chromosome by real encoding system. Individuals of the initial population are assigned an age of one and belong to the layer zero which is the first layer and the population has a total five layers.

Fitness of each individual is given by an objective function which calculates the binding energy of a particular ligand pose as the sum of intermolecular interaction energy and intra molecular torsional energy. E-Hybrid-ALPS computes the intermolecular energy of a particular ligand pose or configuration by trilinear interpolation using the grid maps produced by AutoGrid as look up table. The total intermolecular interaction energy of a ligand pose is calculated as the sum of the intermolecular interaction energy between receptor and for each type of atom in the ligand along with hydrogen bonding, electrostatic and desolvation potentials^{37,38} and intramolecular torsional energy which is obtained from AutoDock tool.

Individuals are selected twice to the mating pool according to the selection methods described above with a selection pressure of 0.5³⁹ and each of the crossovers namely Arithmetic with an α value of 0.5, Heuristic crossover with an α value of 0.9⁴⁰ and Simulated Binary crossover with a β value of 1.2⁴¹ is applied by maintaining the crossover probability at 0.85⁴². To identify the best performing crossover operator, the algorithm was executed 30 times first separately with each of the crossovers and then simultaneously with all the crossovers 30 times with the same settings. Better results

were found with adaptive crossovers. After crossover, age of parents is incremented according to the formula in Figure 1 and offsprings get the maximum age of parents. The lowest binding energy of all the offsprings are calculated and the two best offsprings are added to the layer undergoing reproduction. Following this, individuals whose age exceed the maximum age according to Table 1 are moved to next higher layer. During this, the weak individuals which were existing for more than 25 energy evaluations in the higher layer were replaced. After every 50 energy evaluations, individuals of the zeroth layer were replaced with new 50 individuals. Following this, starting from the zeroth layer, all the evolved layers undergo reproduction. The whole process was repeated for 1,00,000 energy evaluations after which the lowest binding energy was noted. To check for the consistency, the algorithm was executed 30 times with each of the receptor-ligand combination and with same settings

To determine the performance efficiency of E-Hybrid-ALPS, docking studies were also carried out with AutoDock 4 and AutoDock Vina^{36,37} separately. AutoDock calculations were performed in several steps. As a first step coordinate files of receptor were prepared by adding Hydrogens, merging non-polar bonds and adding charges and ligand files were prepared by setting the number of torsions. It follows rapid grid-based energy evaluation method and requires grid map files which were generated by AutoGrid tool for each type of atom in the ligand along with electrostatic and desolvation potential. Lamarckian Genetic Algorithm (LGA) which is a combination of GA and local search is employed as a search algorithm. After setting GA parameters to the same value as that of E-Hybrid-ALPS, binding energy of a particular ligand pose was evaluated using the grid energies and finally, lowest binding energy was obtained. Dockings were done 30 times with each of the receptor-ligand configurations and binding energy in each case was noted.

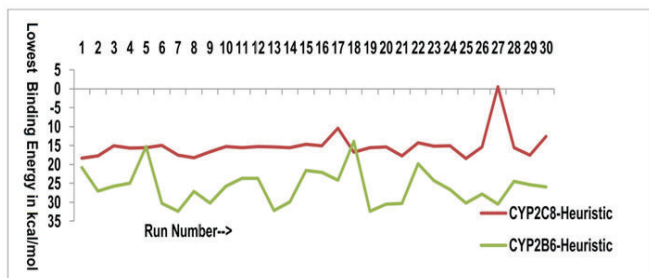
AutoDock Vina also requires receptor and ligand coordinate files which were prepared in the same way as that of AutoDock. Grid dimensions were given as input to AutoDock Vina which also uses a Lamarckian Genetic algorithm as a search algorithm. After setting GA parameters to the same value as of E-Hybrid-ALPS, dockings were done 30 times with each of the receptor-ligand combinations and lowest binding energy in each case was noted. The results of each of the investigation are discussed under Results and discussion.

RESULTS AND DISCUSSIONS

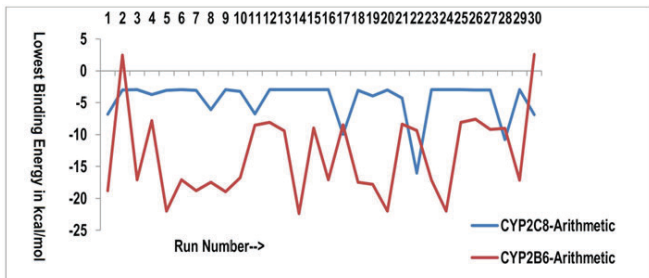
The main objective of the research was to develop a genetic algorithm for search optimization by amalgamating the concepts and advantages of generational and steady state ALPS and the effectiveness of which is determined by implementing it as a search algorithm for molecular docking studies of

the drug discovery process with CYP2C8 and Chloroquine as one receptor and ligand combination and CYP2B6 and Cyclophosphamide as another.

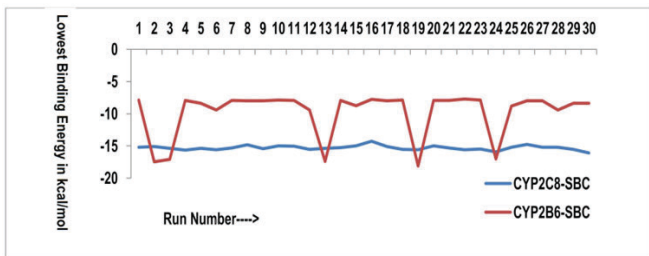
Crossover operators have a role in maintaining a balance between exploration and exploitation and traditional GA uses one crossover operator depending on the type of encoding used. E-Hybrid-ALPS employs real encoding and Heuristic, Arithmetic and Simulated Binary crossovers are found to be frequently used for real encoding system. To identify a better performing crossover operator, an investigation was done by applying each of the Heuristic, Arithmetic and Simulated Binary crossover separately and adaptively and executing the algorithm 30 times with each of the receptor-ligand combinations. Lowest binding energy in each case was noted and the pattern of variation in lowest binding energies across iterations are displayed below.



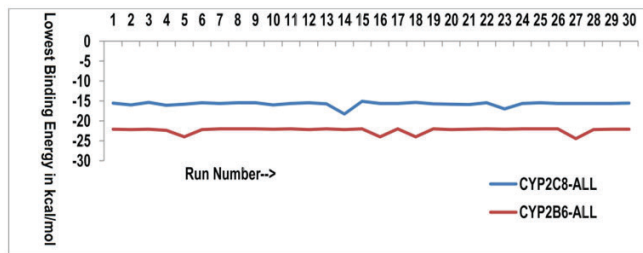
Graph 1: Variations in Lowest Binding Energy with Heuristic Crossover.



Graph 2: Variations in Lowest Binding Energy with Arithmetic Crossover.



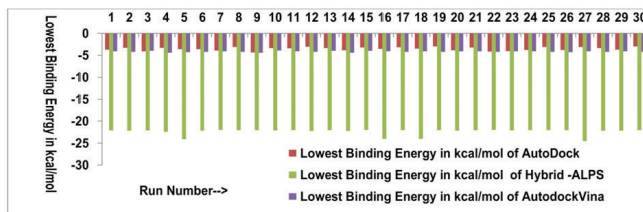
Graph 3: Variations in Lowest Binding Energy with Simulated Binary Crossover(SBC).



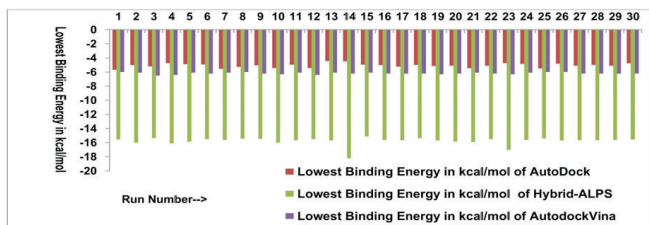
Graph 4: Variations in Lowest Binding Energy with Adaptive Crossover.

It is evident from the Graph-1, Graph-2, and Graph-3 that though each of the crossovers is successful in finding the lowest binding energy for each of the receptor-ligand combinations, the results are not consistent throughout iterations. Graph-4 depicts moderately consistent behaviour of lowest binding energy across iterations as compared to Graph-1, Graph-2 and Graph-3. Hence, it can be indisputably said that the applicability of all three crossovers namely Arithmetic, Heuristic and Simulated Binary crossover simultaneously to the same population and selecting the best performing crossover improves the performance and behaviour of a GA. Following this, a decision was taken to adaptively apply crossovers to E-Hybrid-ALPS and adding the best offsprings to the population.

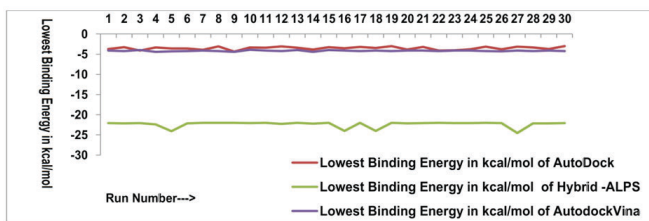
Further to this, various input parameters like initial population size, crossover probability and number of energy evaluations were set to the same value with E-Hybrid-ALPS, AutoDock and AutoDockVina. Along with these parameters, E-Hybrid-ALPS also required number of layers to be evolved as input which was set to five. With these settings, E-Hybrid-ALPS, AutoDock and AutoDockVina were executed 30 times with each of the receptor-ligand combination namely CYP2C8 and Chloroquine and CYP2B6 and Cyclophosphamide. Lowest binding energy in each case was noted. Along with the lowest binding energy, E-Hybrid-ALPS also outputs number of poses with negative energy values and the number of poses with the same energy values. All the findings are thoroughly graphed and discussed below.



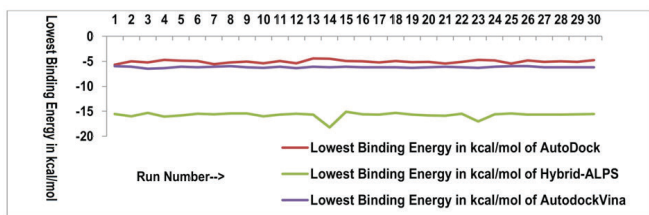
Graph 5: Lowest Binding Energy for CYP2B6 and Cyclophosphamide.



Graph 6: Lowest Binding Energy for CYP2C8 and Chloroquine.



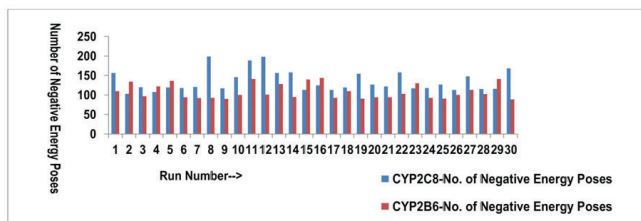
Graph 7: Variations in Lowest Binding Energy of CYP2B6 and Cyclophosphamide.



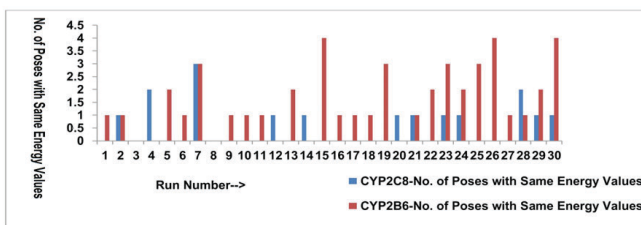
Graph 8: Variations in Lowest Binding Energy of CYP2C8 and Chloroquine.

From the Graph-5 and Graph-6, it is evident that lowest binding energy was found with E-Hybrid-ALPS in all the 30 runs for each of the receptor and ligand combinations as compared to that of AutoDock and AutoDockVina and the algorithm was successful in generating most stable pose of docking as lowest binding energy indicates most stable pose. E-Hybrid-ALPS selects parents to the mating pool once from high fit individuals having lower energy values and then from high and low fit individuals for the reproduction of each layer maintaining the crossover probability of 0.85 and the selection pressure probability of 0.5. Selecting parents from low fit individuals having higher energy values have helped in not losing the good genes of worst individuals. The results obtained from 30 runs for each of the receptor-ligand combinations has confirmed that selecting parents from high fit first and latter from high-low fit individuals for the mating pool was a good choice. Applying multiple crossovers to the same parents at the same time and selecting the best offsprings aided in maintaining population diversity as each crossover operator moves in different directions. E-Hybrid-ALPS has combined the advantages of both generational and

steady-state ALPS. As in Generational ALPS, ‘n’ number of crossovers and replacements are performed on each layer undergoing reproduction with a crossover probability of 0.85 helping in obtaining fitter and fitter individuals and as in steady-state ALPS, individuals not taking part in reproduction are not thrown out resulting in not losing good individuals. All these novel ideas incorporated in E-Hybrid-ALPS have facilitated in attaining lowest binding energy. Apart from displaying lowest binding energy, E-Hybrid-ALPS also outputs no of poses with negative binding energy values and number of poses with same binding energy values as displayed below.



Graph 9: Number of Negative Energy Poses.



Graph 10: Number of Poses with Same Energy Values.

It is apparent from Graph-9 that E-Hybrid-ALPS has obtained a large number of poses, with negative energy values with each of the receptor-ligand combination. In molecular docking, to arrive at the final binding affinity in terms of lowest binding energy, poses generated by the search algorithm have to be scored by a scoring function. The scoring function scores a pose which is near to experimentally determined pose. There are remote chances that pose with the lowest binding energy found by E-Hybrid-ALPS may not be preferred by scoring function. In such cases, if a large number of poses with negative energy values are available, it can help the scoring function in scoring a pose with the lowest binding energy which is near to experimentally determined pose. A large number of -ve energy poses generated by E-Hybrid-ALPS can aid the scoring function in scoring a better pose.

The Graph-10 shows negligible number of poses with the same energy values with each of the receptor-ligand combination. In E-Hybrid-ALPS, mutation is performed before

crossover in case parents undergoing crossover have the same genes. This has contributed E-Hybrid-ALPS in attaining the negligible number poses with same energy values corroborating the fact that E-Hybrid-ALPS has surmounted the problem of premature convergence which is prevalent in other GAs.

CONCLUSION

E-Hybrid-ALPS which has been devised and implemented as a search algorithm for molecular docking studies of CY-P2C8 and Chloroquine and CYP2B6 and Cyclophosphamide has the combined advantage of both generational and steady-state models of ALPS in doing crossovers, replacements according to crossover probability and in not eliminating the individuals that did not take part in reproduction and this has resulted in fitter and fitter individuals. Individuals to the mating pool are selected twice first from high fit and latter from high-low fit groups of individuals maintaining the selection pressure and performing the crossover adaptively on individuals of both the groups separately lead to a diverse population and fitter individuals. Duplication in parental genes is avoided by performing mutation before crossover that has facilitated in overcoming premature convergence. Benefits of incorporating these novel ideas into the devised algorithm E-Hybrid-ALPS have been demonstrated in getting the lowest binding energy in molecular docking consistently with each of the receptor-ligand combinations. Also, a negligible number of poses with the same energy values have confirmed that the algorithm does not lead to premature convergence. Lowest binding energy obtained from E-Hybrid-ALPS is significantly lower than that of lowest binding energy obtained from largely used molecular docking tools namely AutoDock 4 and AutoDockVina which uses Lamarckian Genetic algorithm. The most stable pose of docking is always indicated by the lowest binding energy and binding energy is also a pointer to the binding affinity of the drug. Hence by implementing E-Hybrid-ALPS which generates a ligand/drug pose with the lowest binding energy as search algorithm of molecular docking tool AutoDock can help the drug discoverer in designing a drug with a better binding affinity as lower binding energies indicate higher binding affinity and binding affinity is inversely proportional to drug potency which is a marker for the therapeutic potential of the drug. Also, E-hybrid-ALPS is not limited to the chosen receptor-ligand combination. It can be executed with any of the receptor-ligand combinations whose structures can be constructed or downloaded. Consolidating all these, it can be endorsed that by integrating E-Hybrid-ALPS as search algorithm for AutoDock, AutoDockVina or any other grid-based molecular docking tool may facilitate in improving the drug discovery process by bringing out an efficient drug benefiting the society at large.

FURTHER WORK

E-Hybrid-ALPS is not limited to molecular docking studies. It can be implemented in other areas of research including but not limited to machine learning by changing the objective function. Authors are currently working on applying E-Hybrid-ALPS to estimate the reproductive number of infectious diseases

ACKNOWLEDGEMENT

Authors sincerely thank Dr T. Shantharaju, Venn Therapeutics, the USA for his useful inputs. Our thanks to Dr Priya D. Wanjari, Principal, Santaji Mahavidyalaya, Nagpur and Dr Satish Sharma, HOD Electronics, RTMNU, Nagpur for their constant support and encouragement throughout this work. The authors are highly grateful to all the authors/ editors/ publishers of all the articles, journals and books from where the inputs for this algorithm development have been taken.

Funding: Not Applicable

Conflicts of Interests: Both the authors declare that they have no conflicting interests

REFERENCES

1. David E. Goldberg. Genetic Algorithms. India: Pearson Education; 2006.
2. Isaac R Cassar, Nathan D Titus, Warren M Grill. An Improved Genetic Algorithm for Designing Optimal Temporal Patterns of Neural Stimulation. *Journal of Neural Engineering* 2017; 14(6).
3. Adam Slowik, Halina Kwasnicka. Evolutionary algorithms and their applications to engineering problems. *Neural Computing and Applications* 2020 March 16.
4. Gregory S Hornby. ALPS: The Age-Layered Population Structure for Reducing the Problem of Premature Convergence. In: GECCO'06; 2006 July 8–12; Seattle, Washington, USA.
5. Gregory S Hornby. Steady-State ALPS for Real-Valued Problems. In: GECCO'09; 2009 July 8–12; Montreal, Quebec, Canada.
6. Ehtasham-ul Haq, Ishfaq Ahmad, Abid Hussain, Ibrahim M. Al-manjahie. A Novel Selection Approach for Genetic Algorithms for Global Optimization of Multimodal Continuous Functions. *Computational Intelligence and Neuroscience* 2019 December 05.
7. Ahmad Hassan at, Esra Alkafaween. On enhancing genetic algorithms using new crossovers. *International Journal of Computer Applications in Technology* 2017; 55(3): 202-212.
8. G Pavi, T V Geetha. A Survey on Crossover Operators. *ACM Computing Surveys* 2016; 49(4).
9. Imtiaz Ali Korejo, F A Jokhio Zain UIAbdin Khuhro, Naveed Channa. An Adaptive Crossover Operator for Genetic Algorithms to Solve the Optimization Problems. *Sindh Univ. Res. Jour. (Sci. Ser.)* 2013 June 02; 45 (2) : 333-340.
10. M Amaral, D Kokh, J Bomke, A Wegener, H Buchstaller, H Eggenweiler, et al. Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nature Communications* 2017; 22.

11. Pedro Torres, Ana Sodero, Paula Jofily, Floriano Silva-Jr. Key Topics in Molecular Docking for Drug Design. *International Journal of Molecular Sciences* 2019 September 15; 20(18): 4574.
12. Mohammed Saji, Salahudeen, Prasad S Nishtala. An overview of pharmacodynamic modelling, ligand-binding approach and its application in clinical practice. *Saudi Pharm J.* 2017 Feb; 25(2): 165–175.
13. S M Behera¹, R K Mohanta¹, S K Sahu, M Banerjee, L Mohanta. Molecular Docking: A Review. *World Journal of Pharmaceutical Research* 2013 December; 3(1): 236-257.
14. Rohan R Narkhede, Rameshwar S Cheke, Jaya P Ambhore, Sachin D Shinde. The Molecular Docking Study of Potential Drug Candidates Showing Anti-COVID-19 Activity by Exploring of Therapeutic Targets of SARS-CoV-2. *Eurasian Journal of Medicine and Oncology* 2020 April 29; 4(3): 185-195.
15. Emrah Atilgan, Jianjun Hu. Improving Protein Docking Using Sustainable Genetic Algorithms. *International Journal of Computer Information Systems and Industrial Management Applications* 2011; 3: 248-255.
16. Kim KA, Park JY, Lee JS, Lim S. Cytochrome P450 2C8 and CYP3A4/5 are involved in chloroquine metabolism in human liver microsomes. *Arch Pharm Res.* 2003 August; 26(8): 631-637.
17. Gao J, Tian Z, Yang X. Breakthrough: Chloroquine phosphate has shown apparent efficacy in the treatment of COVID-19 associated pneumonia in clinical studies. *Biosci Trends* 2020 March 16; 14(1): 72-73
18. Centres for Disease Control and Prevention. Information for Clinicians on Investigational Therapeutics for Patients with COVID-19. Available from: <https://www.cdc.gov/coronavirus/2019-ncov/hcp/therapeutic-options.html> [Accessed 26 April 2020].
19. World Health Organization. Cancer. Available from: https://www.who.int/health-topics/cancer#tab=tab_1 [Accessed 26 April 2020].
20. Cancer Research, UK. Cyclophosphamide. Available from : <https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/cancer-drugs/drugs/cyclophosphamide> [Accessed 26 April 2020].
21. Johnson GG, Lin K, Cox TF, Oates M, Sibson DR, Eccles R, et al. CYP2B6*6 is an independent determinant of inferior response to fludarabine plus cyclophosphamide in chronic lymphocytic leukaemia. *Blood* 2013 December 19; 122(26): 4253-4258.
22. Ibrahim El-Serafi, Parvaneh Afsharian, Ali Moshfegh, Moustapha Hassan, Ylva Terelius. Cytochrome P450 Oxidoreductase Influences CYP2B6 Activity in Cyclophosphamide Bioactivation. *Plos One* 2015 November 06.
23. Nguyen NT, Nguyen TH, Pham TNH, Huy NT, Bay MV, Pham MQ, et al. AutodockVina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity. *Journal of Chemical Information and Modeling* 2020 January 27; 60(1): 204-211.
24. Sudha Ramachandra, Vinay Chavan. A Genetic Algorithm for Conformation Search Optimization in Molecular Docking. *International Journal of Computer Science and Information Technologies* 2015; 6 (6): 5547-5551.
25. Jorge Bustos, Victor Adrian Jimenez, and Adrian Will. A comparison of different types of Niching Genetic Algorithms for variable selection in solar radiation estimation. Available from : <https://arxiv.org/pdf/2002.06036.pdf> [Accessed 30 April 2020].
26. Biozentrum. Swiss-Model. Available from : <https://swissmodel.expasy.org/> [Accessed 30 April 2020].
27. H Jalily Hasani, K Barakat. Homology Modeling: An Overview of Fundamentals and Tools. *International Review on Modelling and Simulations* 2017; 10(2).
28. Andrew Waterhouse, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* 2018; 46(w1): W296–W303.
29. Stefan Bienert, Andrew Waterhouse, Tjaart A P de Beer, Gerardo Tauriello, Gabriel Studer, Lorenza Bordoli et al. The SWISS-MODEL Repository-New Features and Functionality. *Nucleic Acids Research* 2017;45(D1): Pages D313–D319
30. Nicolas Guex, Manuel C. Peitsch, Torsten Schwede. Automated comparative protein structure modelling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* 2009 June 10; 30: S162-S173.
31. Gabriel Studer, Christine Rempfer, Andrew M Waterhouse, Rafal Gumienny, Juergen Haas, Torsten Schwede. QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics* 2020 March 15; 36(6): 1765–1771.
32. Martino Bertoni, Florian Kiefer, Marco Biasini, Lorenza Bordoli, Torsten Schwede. Modelling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports* 2017; 7.
33. UniProt. UniProtKB - P10632 (CP2C8_HUMAN). Available from: <https://www.uniprot.org/uniprot/P10632> [Accessed 3 May 2020].
34. ChemSpider. Chloroquine. Available from: <http://www.chemspider.com/Chemical-Structure.2618.html> [Accessed 3 May 2020].
35. Ruth Huey, Garrett M. Morris. Using AutoDock with AutoDockTools: A Tutorial. Available from: http://autodock.scripps.edu/faqs-help/tutorial/using-autodock-with-autodocktools/UsingAutoDockWithADT_v2e.pdf [Accessed 5 May 2020].
36. Grid Maps. Available from: http://www.csb.yale.edu/user-guides/datamanip/autodock/html/Using_AutoDock_305.9.html [Accessed 5 May 2020].
37. Zulfa Afiq Fikriya , Mohammad Isa Irawan, A wiki Puji Dyah Nurhayati. Molecular Docking of Alkaloid Compound SA2014 towards Cyclin D1 Protein in Cancer using Firefly Algorithm. *J. Phys.: Conf. Ser.* 1366 012090 2019.
38. Trilinear Interpolation. Available from: https://en.wikipedia.org/wiki/Trilinear_interpolation
39. Khalid Jebari, Mohammed Mediafire. Selection Methods for Genetic Algorithms. *Int. J. Emerg. Sci* 2013 December ; 3(4) : 333-344.
40. Mhd. Furqan, Hartono, Erianto Ongko, Muhammad Ikhsan. Performance of Arithmetic Crossover and Heuristic Crossover in Genetic Algorithm Based on Alpha Parameter. *IOSR Journal of Computer Engineering* 2017 September; 19(5): 31-36.
41. Eyal Wirsansky. *Hands-On Genetic Algorithms with Python*. Birmingham, UK: Packt Publishing Ltd; 2020.
42. Changxi Ma, Pengfei Liu. Intersection signal timing optimization considering the travel safety of the elderly. *Advances in Mechanical Engineering* 2019; 11(12): 1–8.