Syst Synth Biol (2015) 9:S39–S42
DOI 10.1007/s11693-015-9170-1

RESEARCH ARTICLE

Innovative techniques to discover novel antimalarials

Santanu Hati[1] · **Sudeepto Bhattacharya**[2] · **Subhabrata Sen**[1]

Received: 6 April 2015 / Accepted: 10 April 2015 / Published online: 19 April 2015

Abstract Malaria a global pandemic has engulfed nearly 0.63 million people globally. It is high time that a cure for malaria is required to stop its ever increasing menace. Our commentary discusses the advent and contribution of genetic algorithm (GA) in the drug discovery efforts towards developing cure for malaria. GAs are computational models of Darwinian evolution, ideally capture and mimic the principles of genetic variation and natural selection to evolve good solutions through multiple iterations on the space of all possible candidate solutions, called the search space, to a given optimization problem. Herein we will discuss the applications, advantages, disadvantages and future directions of GA with respect to malaria.

Keywords Malaria · Genetic algorithm · *P. Falciparum* · Metaheuristics

Introduction

Due to the continual intensification of the resistance of the malaria parasite to antimalarial drugs and the malarial mosquito to insecticides, efforts to adequately control the malaria situation in many tropical countries are coming under pressure (Bloland 2001). This, in conjunction with a projected climate change substantially increases malaria risks in the days to come (Patz et al. 2005). Malaria menace causes high mortality especially in Africa and other underdeveloped countries. Another potential problem involves unavailability of vaccination (the most potent remedy). Though malaria deaths in 2010 were reduced by a third of the estimate from 2000, mainly because of the widespread usage of insecticide-treated nets and artemisinin-based combination therapies (Howitt et al. 2012) again now, it is on a rise due to reduced sensitivity to artemisinin therapy, in accordance with the emergence of parasite resistance to all partner drugs, and also due to its inherent paucity in further identifying appropriate small molecules for therapeutic purposes. Hence novel strategies involving both pharmaceutical and nonpharmaceuticals approaches are required to identify robust long term solutions to eradicate malaria.

In this respect metaheuristics play an important role. Several novel metaheuristics viz. genetic algorithm (GA), ant colony optimization (ACO) (Dorigo and Stützle 2004), polar swarm optimization (PSO) (Eberhart 1996) and etc. are used in understanding and eradicating malaria at various levels and facets. Some of these metaheuristics have been successful to some extent to tackle malaria with promises of further improvement and some of them failed. Herein we discuss few diverse case studies ranging from in silico design of antimalarial small molecules to tracking and mapping malaria under climate change scenarios where genetic algorithm has been successfully implemented.

Genetic algorithms comprise a class of optimization routines that are essentially population-based search metaheuristics. Initially developed by Holland for addressing problems of adaptive systems in domains that are characterized by both, enormous search spaces and objective functions with nonlinearities (multiple local optima), discontinuities, high dimensionality and noise. GA provides a

✉ Subhabrata Sen
subhabrata.sen@snu.edu.in

[1] Department of Chemistry, School of Natural Science, Shiv Nadar University, Tehsil Dadri, Chithera, Gautam Budhnagar 201314, UP, India

[2] Department of Mathematics, School of Natural Science, Shiv Nadar University, Tehsil Dadri, Village Chithera, Gautam Budhnagar 201314, UP, India

highly efficient search procedure to effectively address the problem of optimization in such 'difficult' domains (Holland 1975; Goldberg and Holland 1988; Goldberg 1989; Mitchell 2004; Reeves and Rowe 2002).

GA are a large class of routines, where each member of the class is characterized by the following criteria: (1) a nonzero population of well-defined structures agents (structures), (2) action of the agents in an environment, (3) evaluation of the performance of each of the agents using a fitness score, (4) creating new populations using this score as the input and ranking the agents according to their respective scores, (5) selecting the agents with better performance scores, (6) modifying the selected agents through stochastic genetic operators.

Of a multitude of evolutionary search routines that are current in scholarly literature and get labelled as genetic algorithm, we shall remain confined to the application of Simple Genetic Algorithm (SGA) as a tool to study the problem set in this paper (Vose 2007).

SGA is a triple $\Gamma(\Omega, \Im, g)$, where the components are as below: Ω is the search space comprising agents represented as binary strings which are the candidate solutions for a given optimization problem. $\Im$ is an exogenously defined fitness function and$g$is the search heuristic acting on a nonzero population of the candidate strings.

The fitness function $\Im$ is an injective map from Ω to $\Re$, and defines the environment for the evolutionary scheme. It evaluates each string $x_i, i = 1, 2, \ldots, |\Omega|$ and declares a fitness score. The heuristic g comprises three stochastic operators: the selection operator ψ, the crossover operator χ and the mutation operator μ.

The operator ψ maps the simplex Λ representing the population of agents at a given generation P_t to the search space Ω:

$$\psi: \Lambda \rightarrow \Omega$$

Being a non-explorative operator, the selection operator does not generate any new string in the population. The crossover operator χ acts on a pair of elements of the search space,

$$\chi: \Omega \times \Omega \rightarrow \Omega$$

$$(x, y) \mapsto \chi(x, y) x, y \in \Omega$$

The two elements $x, y \in \Omega$ on which χ acts are the parent strings, yielding two offspring strings as a result of the crossover operation. The crossover point is a randomly selected bit position from the interval $[0, l-1]$. The offspring thereby 'inherits' blocks of loci from both the parents, giving rise to the exchange of information between trial solutions. Traditionally, the value of probability of action of the operator χ is significant, ranging between 0.6 and 0.8.

The mutation operator μ acts on one single string x and changes the binary character at a locus on the string to obtain a different string:

$$\mu : \Omega \rightarrow \Omega$$

$$x_i \rightarrow x_j$$

The probability of action of μ is generally taken to be small, of the order of 10^{-3}.

The two operators χ and μ act independently of one another on the population of strings, producing the mixing Θ of the strings. The heuristic g effectively is then the composition of the selection Ψ and the mixing Θ: $g = \Theta \circ \Psi$.

The net effect of the heuristic is the creation of new strings in the search space with a spectrum of fitness values, resulting in a very efficient sampling plan. One may thus expect subsets of the search space Ω containing strings of similar profile and sharing a particular set of fitness score (schema) to emerge and evolve dynamically.

Modern day drug discovery utilize in silico techniques to find the ideal drug molecule through virtual screening of large number of compound libraries (virtual) designed with suitable reagents (chemical reactants used to prepare molecules). Number of compounds in these virtual combinatorial libraries are usually huge (in the order of millions!!!) and increases with the increase in the number of reagents. However the aforementioned library size and the related computational complexity by its very nature pose a serious challenge for an efficient optimization of the search mechanism to identify high performance molecules within searchable limits. To solve this problem and to make the in silico technique more robust genetic algorithm is often used. It is noteworthy that use of genetic algorithm is not only confined to in silico techniques but encompasses various applications in drug discovery which involve generation of better solutions. GA have emerged as one of the most favourable applications essentially because of this characteristics to mimic natural selection. To evolve good solutions, there must exist a measure associated with each individual candidate member of the search space, in order to distinguish the good candidate solutions from the bad ones on the basis of this measure. This required measure is provided by assigning a fitness function, which is often derived from the objective function to be optimized, and is used in successive genetic iterations (Mitchell 2004). In all applications of GAs, the goal is to find individual solutions from the search space that have the optimum fitness value.

Herein we discuss few case studies involving GA towards eradicating malaria. We primarily focus on ligand receptor docking studies, QSAR studies towards novel antimalarial small molecules and also application of GA in ecological mapping of various malaria vectors.

In a recent example Al-Safi and Alshaikhli used GA (with Autodock 4.2, popular docking simulation software) in identifying most promising drug candidate (among 20 drugs) for malaria through ligand receptor docking studies (Al-Safi and Alshaikhli 2012). This case study assess the application of GA, in ligand-receptor docking experiment. Herein an antimalarial small molecule was docked on the protein *falcipan-2*, which plays a significant role in the life span of *P. Falciprum* (Parasite *P. Falciparum* (the parasite that causes malaria)) (Breman et al. 2004). The crystal structure of the *falcipan 2* was obtained from the protein databank (.pdb), with 3D coordinates as 2GHU.pdb. The drugs were downloaded from Pubchem chemical database. Autodock 4.2 applied Lamarckian Genetic Algorithm to optimize the binding of the ligand (the drug molecule) to the protein. Chromosomes are represented as vectors of real numbers where each vector component represents a gene. The individual chromosomes have genes representing ligand translation, ligand orientation and for conformation. Additionally a particular ligand–protein complex orientation is unique and is an "individual" in its GA. In general the chromosomes are represented as vector of real numbers. The table below below (Table 1) depicts the GA chromosomes in the autodock. The translational genes are represented by T_x, T_y and T_z and indicates the centre of the ligand in 3-D space, the quaternary genes represented by Q_x, Q_y, Q_z and Q_w represents the angle of n-rotatable bonds in the ligands.

The best molecule is selected on the basis of the total binding energy between the ligand and the protein, where lower energy indicates stronger binding. After docking of 2GHU.pdb against twenty antimalarial compounds Dioncopeltine A was identified as the best binder with lowest binding energy of 8.6 kcal/mole. The med-chem descriptors of Dioncopeltine A adheres to lipinsky's rule of 5 [hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), molecular weight, LogP and polar surface area (PSA), of drug-able compounds] thereby making Dioncopeltin A as the most potential candidate against malaria in this study.

This is case study, illustrates effectiveness of GA metaheuristics in designing appropriate drug molecules. This successful endeavor surely opens a window of opportunity to apply this among a wider variety of compounds which are not only drugs but also novel compounds.

In another example, Behesti et al. described developing a QSAR model for predicting antimalarial activity of urea derivatives using genetic algorithm-multiple regression (Behesti et al. 2012). There were a set of 68 urea compounds containing alkyl-benzopyrazine substitution (**1**) and aryl indoline substitution (**2**) (Fig. 1). As the name suggests QSAR or quantitative structure activity relationship model, develops a correlation between a molecule's 2D or 3D structural descriptors and its biological activity against a specific (for example in this case, malaria) or general therapeutic areas. The usefulness of this correlation model resides on it effective prediction of biological activity of any molecules based described by the same set of descriptors.

In order to develop an appropriate QSAR model it is extremely important to generate a robust set of molecular descriptors which truly represents the molecule's characteristics. GA plays a key role in this step. As in the case of several other genetic algorithm triggered operations, here too the GA cycle is repeated to ensure generation of satisfactory descriptor set. In this case an ideal group of molecular descriptors identified by GA were constitutional, topological, geometrical, electrostatic and quantum mechanical. Once the model is developed, the QSAR was performed on available 68 IC_{50} oral data based on these molecular descriptors. In QSAR analysis, the quality of a model is adjudged by its fitness and predictability accordingly this model was validated and the results indicated that it was robust and the predictive ability was consistent and dependable to design novel antimalarial compounds.

As mentioned earlier, applications of GA is not only restricted to drug discovery but encompasses any problems that are searching real solutions. Such a problem and its potential solution is discussed in the following example (Tonnag et al. 2010). Female *Anopheles* mosquito the most potent and active vector of malaria are sensitive to climate changes and have a tendency to shift locations in search of favorable condition for proliferation. This is a point of major concern for tackling malaria. Tonnang et al. have developed a model to predict and map the redistribution of female *Anopheles* mosquito under climatic changes, based on prior data on geographical range of vectors and malaria distribution. Genetic algorithm for rule set prediction model was applied to predict the geographical and ecological distribution of three species of *Anopheles* mosquito, *A. gambie*, *A. arabiensis* and *A. quadriannulatus*. Accordingly a model was developed through calibration of CLIMEX parameters. The model estimated possible

Table 1 Representation of GA chromosomes in Autodock 4.2

T_x	T_y	T_z	Q_x	Q_y	Q_z	Q_w	R_1	R_2

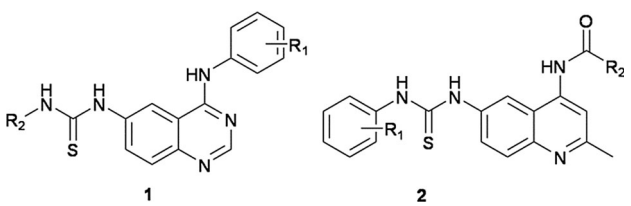

Fig. 1 Motifs for QSAR modelling

geographical distribution and seasonal influx of the species w.r.t. various climatic factors viz. temperature, rainfall, relative humidity and etc. that defines the living environment for *Anopheles* mosquitoes. The accumulated data were then analyzed and the distribution of the mosquitoes was mapped. The study indicated that these mosquitoes dwell into south and east of Africa rather than delocalizing into different climatic environment. This study helped in early planning and developing sustainable measures for climate change and the adaptation of malaria vector control in Africa. This proof of concept can be applied to various *anopheles* infested regions in the world and thereby creating preventive measures to stop their redistribution.

These examples of GA indicate opportunities to alter and improve the process drug development to generate cure via augmenting the existing methods. They are still in there nascent stage and requires substantial polishing to reach to the level where they can be used for the good of mankind. These studies are successful proof of concepts and can be taken forward to the next step. GA based in silico technique has shown remarkable improvement in robustness and accuracy. Mankind has already been cornered in his battle with malaria. Conventional techniques have failed to address the problem. Hence out of the box strategies like GA are our hope to turn the tides in our favor.

References

Al-Safi HM, Alshaikhli IFT (2012) Rational drug design using genetic algorithm: case of malaria disease. J Emerg Trends Comput Inf Sci 3(7):1093–1102

Behesti A, Pourbasheer E, Nekoei M, Vahdani S (2012) QSAR modelling of antimalarial activity of urea derivatives using genetic algorithm-multiple linear regressions. J Saudi Chem Soc. doi:10.1016/j.jscs.2012.07.019

Bloland P (2001) Drug resistance in malaria. World Health Organization, Geneva

Breman JG, Alilio MS, Mills A (2004) Conquering the intolerable burden of malaria: what's new, what's needed: a summary. Am J Trop Med Hyg 71S:1–15

Dorigo M, Stützle T (2004) Ant colony optimization. MIT Press, Cambridge

Eberhart R (1996) Computational intelligence PC tools, Academic Press Professionals (APP), San Diego, California, USA, pp 185–196

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison Wesley, Boston

Goldberg DE, Holland JH (1988) Special issue on genetic algorithms. Mach Learn 3(2–3):95–99

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor

Howitt et al (2012) Technologies for global health. Lancet 380(9840):507–535

Mitchell M (2004) An introduction to genetic algorithms. Prentice-Hall of India, New Delhi

Patz JA, Campbell-Lendrum D, Holloway T, Foley J (2005) Impact of regional climate change on human health. Nature 438:310–317

Reeves CR, Rowe JE (2002) Genetic algorithms—principles and perspectives: a guide to GA theory. Kluwer Academic Publishers, Berlin

Tonnag HEZ, Kangalawe YM, Yanda PZ (2010) Predicting and mapping malaria under climate change scenarios: the potential redistribution of malaria vectors in Africa. Malar J 9:101–111

Vose M (2007) The simple genetic algorithm. Prentice-Hall of India, India