RESEARCH ARTICLE

A Newton Cooperative Genetic Algorithm Method for *In Silico* Optimization of Metabolic Pathway Production

Mohd Arfian Ismail[1,2◉], Safaai Deris[2◉]*, Mohd Saberi Mohamad[2‡], Afnizanfaizal Abdullah[2*‡]

1 Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, Gambang, Kuantan, Malaysia, **2** Artificial Intelligence and Bioinformatics Group, Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia

◉ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* safaai@utm.my (SD); afnizanfaizal@utm.my (AA)

Abstract

This paper presents an *in silico* optimization method of metabolic pathway production. The metabolic pathway can be represented by a mathematical model known as the generalized mass action model, which leads to a complex nonlinear equations system. The optimization process becomes difficult when steady state and the constraints of the components in the metabolic pathway are involved. To deal with this situation, this paper presents an *in silico* optimization method, namely the Newton Cooperative Genetic Algorithm (NCGA). The NCGA used Newton method in dealing with the metabolic pathway, and then integrated genetic algorithm and cooperative co-evolutionary algorithm. The proposed method was experimentally applied on the benchmark metabolic pathways, and the results showed that the NCGA achieved better results compared to the existing methods.

Introduction

Recently, computational system biology has gained attention from many researchers and become an important research area. The computational system biology has contributed to the understanding on the process of the complex biology where it enables the biological process to act as a system. It can be achieved by utilizing the techniques and knowledge in the molecular biology and genetics where it allows living cell to be manipulated as a real factory and gives insights to researchers into ways to improve cell production. One way to improve cell production is by *in silico* optimization of metabolic pathway production. Metabolic pathway can be defined as a series of chemical reactions that occur within the microorganism cell. The computational system biology has enable the metabolic pathway to be represented by mathematical model. Due to that, it is possible to perform the optimization process using computer simulation (*in silico* optimization).

The *in silico* optimization of metabolic pathway production can be seen as the search for a set of components (chemical reactions) for maximizing production rate. It works by altering and tuning the value of the components in the metabolic pathway in order to find the maximum production rate. Similar to the production, the total of the component concentrations involved also needs to be considered. When minimum value of component concentrations are involved, the production cost can be reduced [1, 2]. Recently, many successful works have been published on the *in silico* optimization of metabolic pathway production. Most of these works used continuous optimization approach, such as linear programming method [2–4] and geometric programming method [5, 6]. These approaches usually involve equality and inequality constraints. They also require the definition of the decision variable that requires expert knowledge. Moreover, the search process in the continuous optimization approach totally depends on the decision variable, and this can lead to convergence problem if the decision variable is not accurate [7].

In contrast to the continuous optimization approach, the combinatorial optimization approach works by finding the optimal solution from a finite set of objects. The method functions by using a stochastic operator on a pool of candidate solutions for the optimization problem and makes it more efficient and robust [8]. Due to its stochastic nature, the method does not require any assumption regarding the structure of the model and the definition of the decision variable. This is because the stochastic operator uses random method to determine the search direction, which makes it more robust. For this reason, the present study uses the combinatorial optimization approach for the *in silico* optimization of metabolic pathway production. In this optimization process, metabolic pathway can be viewed as a nonlinear equations system. This is because the metabolic pathway can be described as mathematical model. There are many methods that can be used in dealing with the nonlinear equations system, including numerical, evolutionary and swarm intelligence methods. The Newton method, which is a numerical method, is the most popular method used [9, 10]

Newton method is an iterative method used to find an optimum point to real-valued roots. The utilization of the Newton method for the *in silico* optimization of metabolic pathway production is a good choice because of the fast convergence speed of the Newton method [11]. In this work, Newton method views metabolic pathway as a nonlinear equations system. Applying only the Newton method is not sufficient for the optimization process because the variables in the nonlinear equations system need to be tuned. This is because all the metabolic pathway components are represented by many variables in the nonlinear equations system. Therefore, a method is needed to represent and tune the variables. To overcome this problem, genetic algorithm (GA) is applied. This approach can be performed by representing the variables in the nonlinear equations system as a chromosome. Then, the chromosome is evolved and tuned. However, several issues occur when this method is applied on a complex metabolic pathway that involves many metabolic pathway components. This can cause complex representation of the solution as many metabolic pathway components need to be represented by GA. In addition, evaluating the solution is time consuming. Therefore, a method needs to be embodied into the GA to simplify the representation of the solution. It has been found that applying the cooperative co-evolutionary algorithm (CCA) is the best choice, where the CCA method works by dividing the representation of the solution into multiple sub-solutions [12–14].

In this paper, an improved method, namely the Newton Cooperative Genetic Algorithm (NCGA), was utilized for the *in silico* optimization of metabolic pathway production. This method uses the Newton method to deal with metabolic pathway. GA is then used in the optimization process and CCA is embodied into the GA to decompose the solution into multiple sub-solutions. Furthermore, this study introduced concepts for representing the solution, namely NCGA representation which includes sub-chromosome representation and

cooperative chromosome representation. The sub-chromosome is part of the cooperative chromosome, while the cooperative chromosome is a complete solution. The NCGA representation concept ensures that the NCGA is able to increase the production. In addition, this study also introduced a two-level evaluation of the solution, namely sub-chromosome evaluation and cooperative chromosome evaluation where this concept minimizes the total of the component concentrations involved. In the following section, the theory of the metabolic pathway is discussed, where the modelling of the metabolic pathway and the optimization problem are described. This is followed by a discussion of the proposed method and two case studies of the optimization of ethanol production in *Saccharomyces cerevisiae* (*S. cerevisiae*) pathway and the optimization of *tryptophan* (*trp*) biosynthesis in *Escherichia coli* (*E. Coli*) pathway. Finally, the results are presented and discussed before a brief conclusion is made.

Modeling of metabolic pathway

In this study, the mathematical model used to represent the metabolic pathway is generalized mass action (GMA) models. The representation of GMA has the following form:

$$\frac{dX}{dt} = sv(x) \tag{1}$$

where s represents the stoichiometric matrix of the system and $v(x)$ denotes the vector that contains reaction rate. There are two types of reactions, which are dependent and independent variables. The dependent variable usually represents the metabolite concentrations, while the independent variable is the enzyme activity. These variables are usually in the form of nonlinear functions. The reaction rate $v(x)$ can be represented by using the power-law function, which has the following form [6, 15]:

$$v_i = \gamma_i \prod_j x_j^{f_{ij}} \tag{2}$$

In this representation, coefficient γ_i is denoted as the rate constant for v_i and coefficient f_{ij} is the kinetic order. These two coeffcients are derived from the Taylor series in the logarithmic space around a steady state [6, 16]. They can be defined as follows:

$$\gamma_i = |v_i|_0 \tag{3}$$

$$f_{ij} = \left| \frac{\delta v_i}{\delta x_j} \frac{x_j}{v_j} \right| \tag{4}$$

where the subscript 0 refers to the value at the steady state condition.

Problem statement for optimization

The *in silico* optimization of metabolic pathway production is always constrained by steady state condition. In the steady state condition, all the variables in the metabolic pathway are in static values. This condition forces all GMA models to be equal to 0, and thus produces models as follows:

$$\frac{dX_n}{dt} = \begin{bmatrix} sv(x)_1, & \dots & sv(x)_n \end{bmatrix} = 0 \tag{5}$$

This leads to the solving of a nonlinear equations system, which can be defined as follows [1, 9, 10, 17]:

$$f(x) = [f(x)_1, f(x)_2, \ldots, f(x)_n] = 0 \tag{6}$$

where $x = (x_1, x_2 \cdots x_n)$ denotes n equations and n variables, while $f_1, f_2 \cdots f_n$ are the nonlinear functions. In order to solve a nonlinear equations system, all the equations $f(x)_1, f(x)_2 \cdots f(x)_n$ must be equal to 0. This situation is similar to the *in silico* optimization of metabolic pathway production. Therefore, the *in silico* optimization of metabolic pathway production can be considered as a method for solving a nonlinear equations system.

Besides the steady state condition, the constraint of the components in the metabolic pathway also needs to be considered. This constraint exists in order to ensure the concentrations of the components remain within the boundaries to maintain the survival of the microorganism cell [1, 6]. Thus, the *in silico* optimization of metabolic pathway production involves improving the pathway production and at the same time tries to reduce the component concentration involved. The optimization problem can be written as follows:

$$max \, F_1 \, (v) \tag{7}$$

$$min \, F_2 \left(\sum_{j=1}^{n} x_j \right) \tag{8}$$

where objective function F_1 is the production (maximum function) and objective function, while F_2 is the total of the component concentrations involved.

## Newton cooperative genetic algorithm

In this section, NCGA method is discussed. The NCGA is used for the *in silico* optimization of metabolic pathway production. The proposed method treats the metabolic pathway as a nonlinear equations system. In dealing with a nonlinear equations system, the Newton method is applied, and then GA is used to represent the variables in the system. This is necessary in order to tune the variables to search for the optimum set. GA loses its effectiveness when it is applied in a complex metabolic pathway as the representation of the solution becomes complex and difficult to evaluate. Hence, CCA is applied to overcome this particular problem.

The present study introduces a modified GA, which includes the NGCA representation concepts and the two-level fitness evaluation. Fig 1 shows the simplified flow chart of the NCGA. In the figure, the concepts of the NCGA representation and the two-level fitness evaluation are highlighted by the dashed lines. The detailed flow of the algorithm is as follows:

Step 1. Generate the initial N sub-chromosome in N sub-population. Each sub-chromosome represents each variable in the nonlinear equations system. The number of N sub-populations depends on the number of variables in the nonlinear equations system. The example of representation of all variables in nonlinear equations system by sub-chromosomes can be viewed in Fig 2. The sub-chromosome is in binary representation.

Step 2. Evaluate the sub-chromosomes. In this step, a representative of the sub-chromosome from all the sub-populations is selected based on the fitness value, where the sub-solution with the lowest fitness value is selected first. This is done to ensure that the representatives from all the sub-populations will combine with each other in order to minimize the total component concentrations involved. This step is referred to as the sub-chromosome evaluation.

Step 3. Form the cooperative chromosome. After all the representatives are selected from their sub-populations, they are combined with each other to form a complete solution known
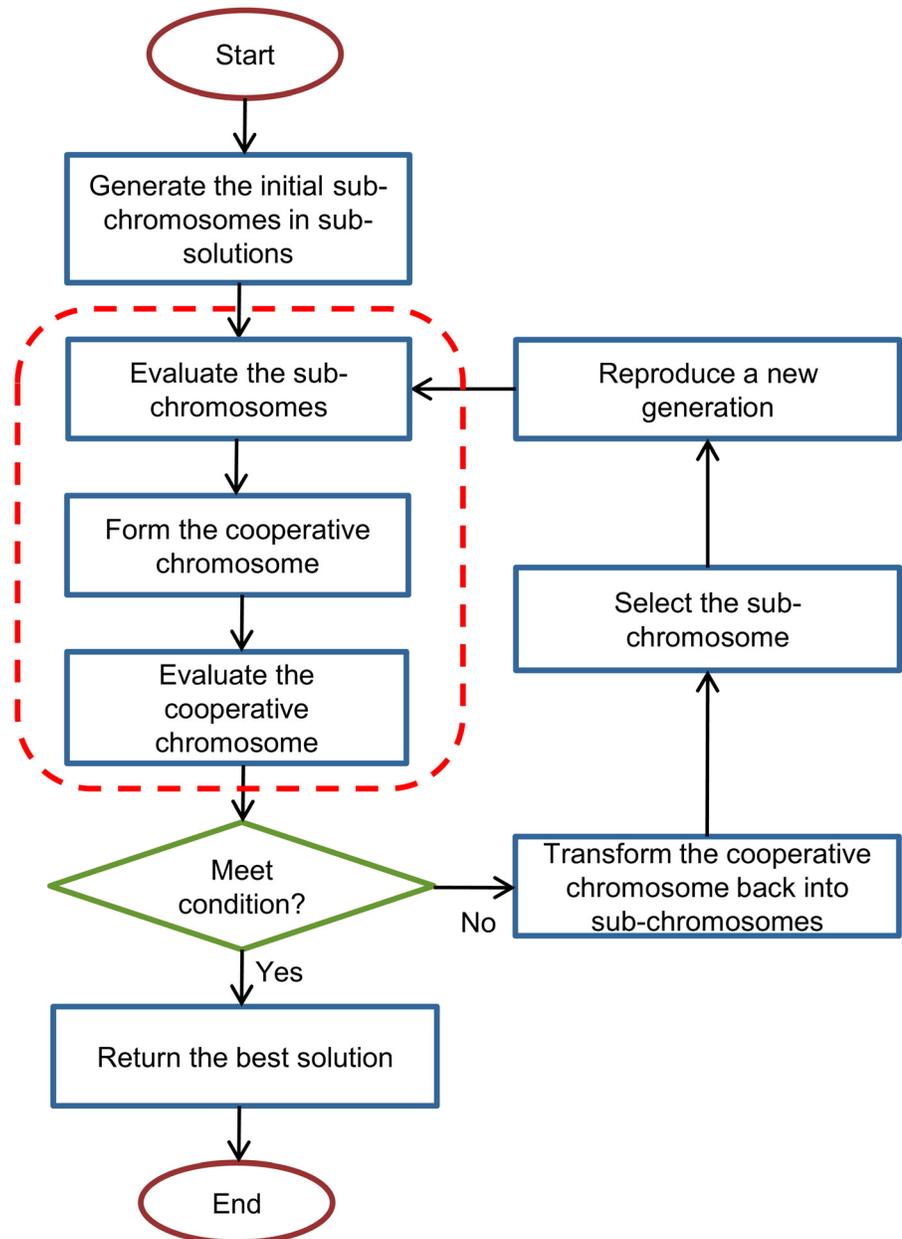
Fig 1. The flow chart of the NCGA. The NCGA starts with representation of the solution by GA and CCA. Then the Newton method is applied in order to evaluate the solution.
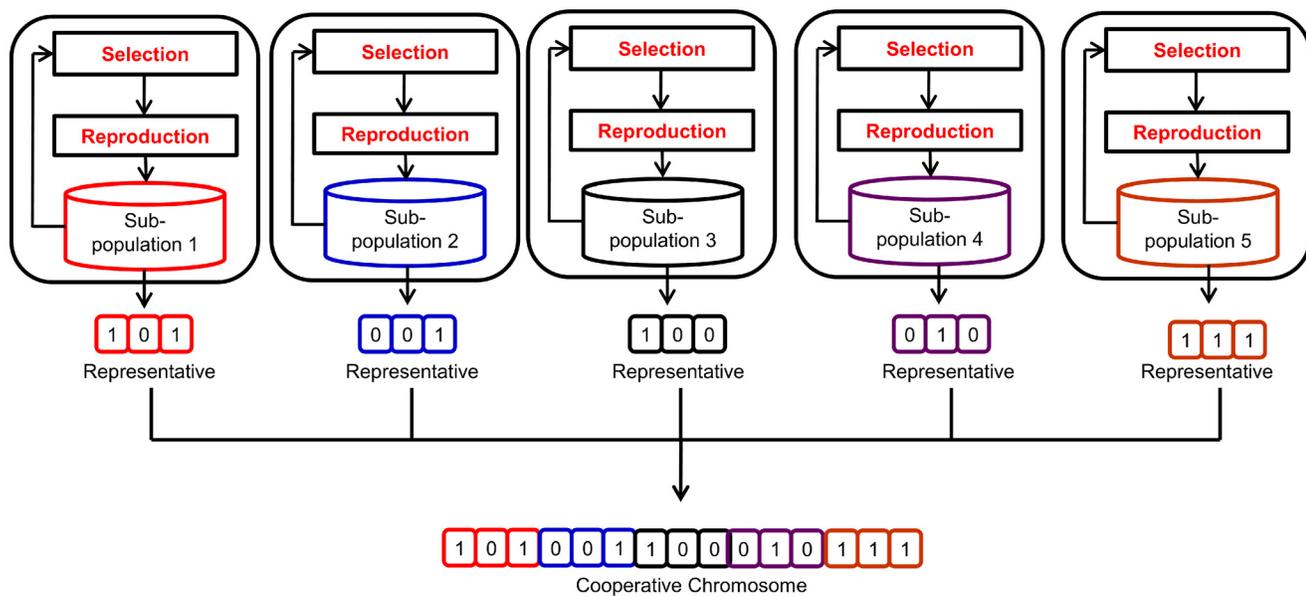
doi:10.1371/journal.pone.0126199.g001

as the cooperative chromosome. This concept, referred to as the cooperative chromosome representation, is pictured in Fig 3.

Step 4. Evaluate the cooperative chromosome. In this step, the cooperative chromosome is tested by the Newton method. A condition at this stage is whether or not termination has occurred. The termination condition can occur in two conditions; when the maximum number of generations has been reached, and when the component concentration constraint is in the optimum range. If the termination condition is achieved, then the process skips directly to Step 8.

$$f(x) = x_1 + x_2 - x_3$$
$$f(x) = x_4 + x_5$$
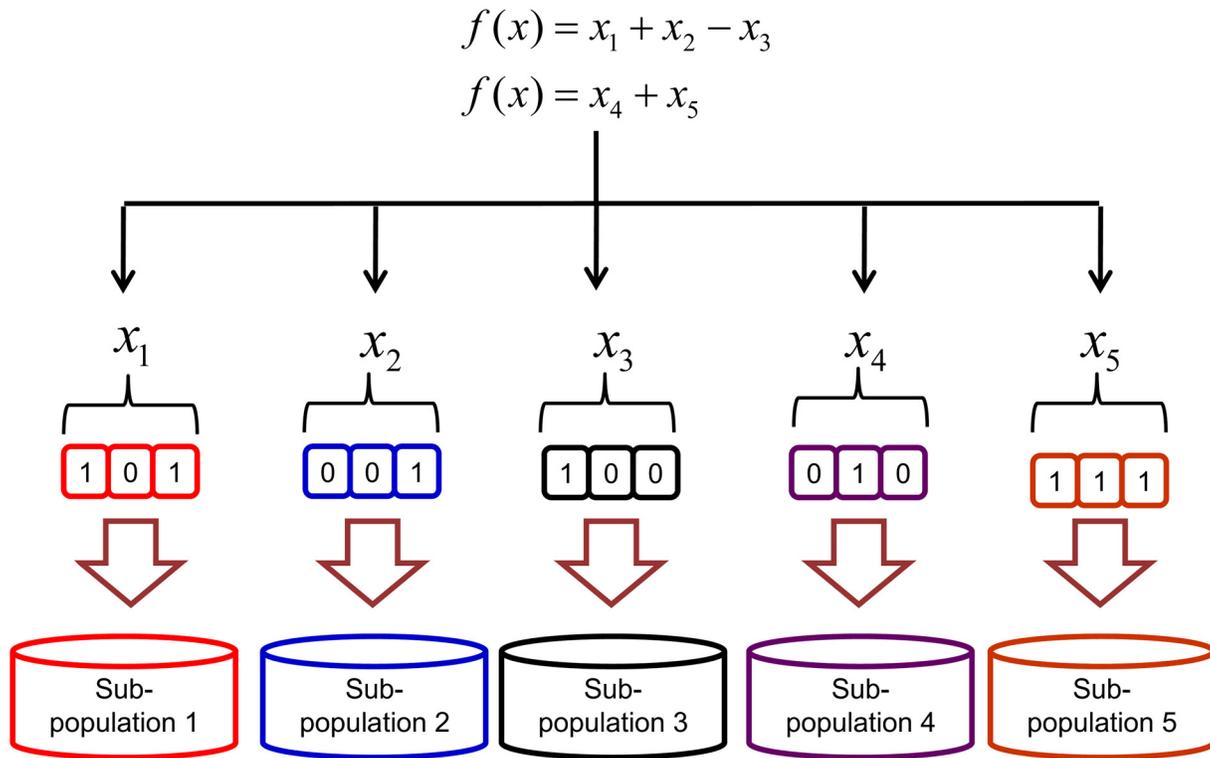
Fig 2. The representation of nonlinear equations system variables by sub-chromosomes. Each variable in nonlinear equations system represented by sub-chromosome and evolve in their own sub-population.

doi:10.1371/journal.pone.0126199.g002

Fig 3. Cooperative chromosome representation. A representative was selected from all sub-population in order to form the cooperative chromosome.

doi:10.1371/journal.pone.0126199.g003

Step 5. Transform the cooperative chromosome back into sub-chromosomes. In this step, the cooperative chromosome is decomposed back into sub-solutions with all the sub-chromosomes are going back into their own sub-population. The purpose of this step is to select sub-chromosomes in order to perform the reproduction process.

Step 6. Select the sub-chromosome. The selection process is based on the fitness values of the sub-chromosome, whereby the sub-chromosome with the lowest fitness value is selected first. This is intended to minimize the total of the component concentrations involved by making all sub-solutions that have lowest fitness value combine with each other. The process is performed until the last sub-chromosome has been selected.

Step 7. Reproduce a new generation. Crossover and mutation procedures are applied on the selected sub-chromosomes. This is performed in order to produce a new generation that has better quality compared to the previous generation. Then, the new generation goes back to Step 2.

Step 8. Return the best solution. This step decodes the cooperative chromosome into the solution. Fig 4 shows the NCGA algorithm in pseudocode format.

Case studies

To prove the effectiveness of the NCGA, a program was used and tested on the metabolic pathway benchmark, namely the optimization of ethanol production in *S. cerevisiae* and the optimization of *trp* production in *E. coli* pathway. The program written in Java language that is known as jMetal [18] was used. Besides that, JAMA version 1.0.3 [19] was used to handle the nonlinear equations system and integrate with the jMetal program. The jMetal program can be downloaded from http://jmetal.sourceforge.net/index.html while JAMA version 1.0.3 from http://math.nist.gov/javanumerics/jama/.

Case study 1: Optimization of ethanol production in *S. cerevisiae* pathway

Modelling framework. The GMA model derived from the ethanol production by *S. cerevisiae* was used as the case study in order to observe the performance of NCGA. The GMA model derived from the *S. cerevisiae* pathway was suspended in a cell culture at pH 4.5 as recommended by Galazzo and Bailey [20]. This pathway has been applied in many studies over the years. Fig 5 depicts a schematic representation of the *S. cerevisiae* pathway. The GMA model of this pathway is described as follows:

$$
\begin{aligned}
\frac{dX_1}{dt} &= V_{in} - V_{HK} \\
\frac{dX_2}{dt} &= V_{HK} - V_{PFK} - V_{Carb} \\
\frac{dX_3}{dt} &= V_{PFK} - V_{GAPD} - 0.5V_{Gro} \\
\frac{dX_4}{dt} &= 2V_{GAPD} - V_{PK} \\
\frac{dX_5}{dt} &= 2V_{GAPD} + V_{PK} - V_{HK} - V_{Carb} - V_{PFK} - V_{ATPase}
\end{aligned}
\tag{9}
$$

Input: Initial sub-chromosomes that represent the variable in nonlinear equations system

Output   : A complete set of optimum solution, cc^g

1.   **Begin** NCGA
2. $g = 0$
3. $p_k^g = \{p_1^g, p_2^g, p_3^g, \ldots, p_k^g\}$ // p is sub-population
4.   **For** p_k^g
5. $i_{jk}^g = \{i_{1k}^g, i_{2k}^g, i_{3k}^g, \ldots, i_{jk}^g\}$ // i is sub-chromosome
6.   **For** i_{jk}^g
7.   *evaluate* i_{jk}^g
8.   **End For**
9. $rep\_i_{jk}^g \leftarrow$ *choose* i_{jk}^g // rep_i is representative from sub-population
10. $cc^g \leftarrow$ *perform by combine all* $rep\_i_{jk}^g$ // cc is cooperative chromosome
11.   **End For**
12.   *evaluate* cc^g
13.   **While** *not meet-condition*
14. $g = g + 1$
15.   **For** p_k^g
16.   **For** i_{jk}^g
17.   *select* i_{jk}^{g-1}
18. $i_{jk}^g \leftarrow$ *reproduce from* i_{jk}^{g-1}
19.   *evaluate* i_{jk}^g
20.   **End For**
21. $rep\_i_{jk}^g \leftarrow$ *choose* i_{jk}^g
22. $cc^g \leftarrow$ *perform by combine all* $rep\_i_{jk}^g$
23.   **End For**
24.   *evaluate* cc^g
25.   **End While**
26.   *return the best solution*
27. **End** NCGA

Fig 4. The pseudocode of the Newton cooperative genetic algorithm.

doi:10.1371/journal.pone.0126199.g004
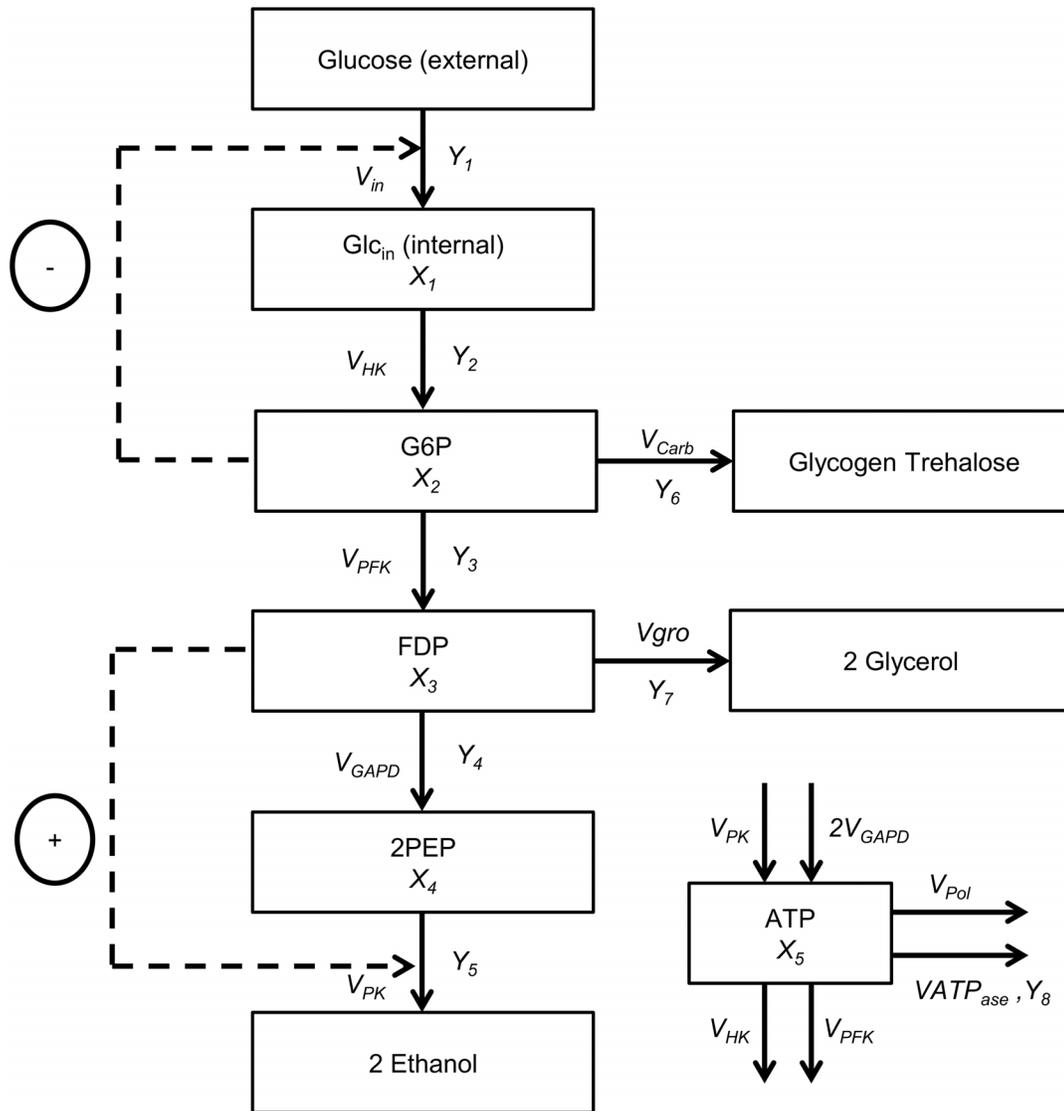
Fig 5. Anaerobic fermentation pathway in *S. cerevisiae*.

The variables and the corresponding steady state are summarized in [Table 1]. At the initial steady state, all the fluxes in the model were formulated in the following form:

$$V_{in} = 0.8122X_2^{-0.2344}Y_1$$
$$V_{HK} = 2.8632X_1^{0.7464}X_5^{0.0243}Y_2$$
$$V_{PFK} = 0.5232X_2^{0.7318}X_5^{-0.3941}Y_3$$
$$V_{Carb} = 8.904 \times 10^{-4}X_2^{8.6107}Y_7$$
$$V_{GAPD} = 7.6092 \times 10^{-2}X_3^{0.6159}X_5^{0.1308}Y_4$$
$$V_{Gro} = 9.272 \times 10^{-2}X_3^{0.05}X_4^{0.533}X_5^{-0.0822}Y_8$$
$$V_{PK} = 9.471 \times 10^{-2}X_3^{0.05}X_4^{0.533}X_5^{-0.0822}Y_5$$
$$V_{ATPase} = X_5Y_6$$

(10)

Table 1. Summary of metabolites and enzymes in case study 1.

Metabolite	Acronym	Symbol	Initial Steady State
Glucose (internal)	Glc_{in}	X_1	0.0345
Glucose-6-phosphate	$G6P$	X_2	1.0110
Fructose-1,6-phosphate	FDP	X_3	9.1440
Phosphoenolpyruvate	PEP	X_4	0.0095
Adenosine triphosphate	ATP	X_5	1.1278
Enzyme	**Acronym**	**Symbol**	**Initial Steady State**
Glucose transport	V_{in}	Y_1	19.70
Hexokinase	V_{HK}	Y_2	68.50
Phosphofructo-1-kinase	V_{PFK}	Y_3	31.70
Glyceraldehyde dehydrogenase	V_{GAPD}	Y_4	49.90
Pyruvate kinase	V_{PK}	Y_5	3440.00
ATPase	V_{ATPase}	Y_8	25.10
Polysaccharide biosynthesis	V_{Carb}	Y_6	14.31
Polyol biosynthesis	V_{Gro}	Y_7	203.00

doi:10.1371/journal.pone.0126199.t001

Optimization problem. The performance of the method proposed in this work can be assessed by the rate of ethanol production given by the flux of pyruvate kinase, V_{PK}. In addition, the total of the component concentrations involved must be considered. Thus, the optimization problem of this case study can be defined as follows:

$$max\, F_1(v) = V_{PK} \tag{11}$$

$$min\, F_2 = \sum_{j=1}^{5} X_j + \sum_{j=6}^{6} Y_j \tag{12}$$

The optimization problem was subjected to the steady state condition, where all the GMA models were equal to 0:

$$
\begin{aligned}
V_{in} - V_{HK} &= 0 \\
V_{HK} - V_{PFK} - V_{Carb} &= 0 \\
V_{PFK} - V_{GAPD} - 0.5 V_{Gro} &= 0 \\
2V_{GAPD} - V_{PK} &= 0 \\
2V_{GAPD} + V_{PK} - V_{HK} - V_{Carb} - V_{PFK} - V_{ATPase} &= 0
\end{aligned} \tag{13}
$$

The constraint of the component concentrations involved was also considered. This is to ensure that the microorganism is still workable. In this case study, the components were categorized into two groups, which are metabolite and enzyme. The metabolites ($X_1 - X_5$) were set approximately 20% from their steady state values, which were in the range of 0.8–1.2. For the enzymes, not all enzymes were tuned as only $Y_1 - Y_6$ were involved. The values for enzymes were set in the range of 0–50 [1, 21].

Case study 2: Optimization of *trp* biosynthesis in *E. Coli* pathway

Modelling framework. In this pathway, NCGA was used to optimize the *trp* production. This pathway is introduced by Xiu *et al.* and detailed description can be found in [22]. This
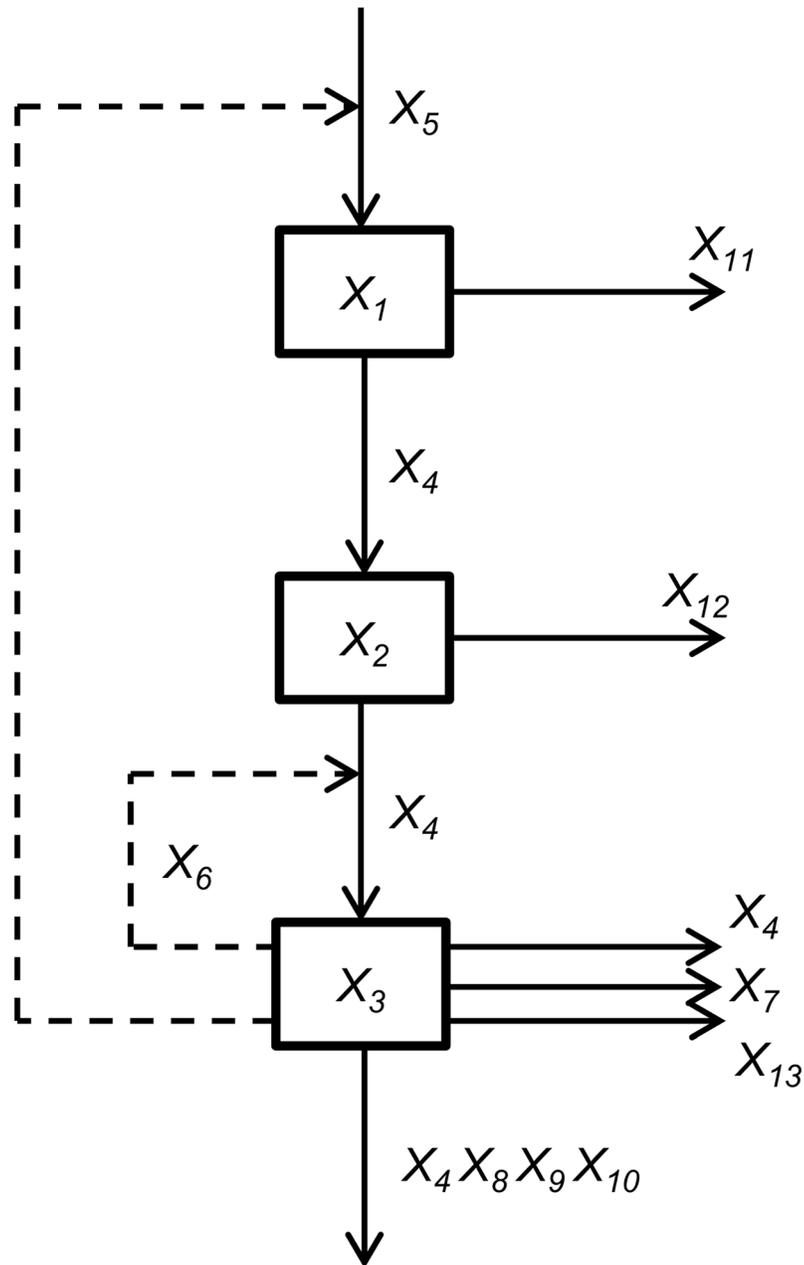
Fig 6. _Tryptophan_ biosynthesis in _E. coli_ pathway.

pathway is depicted in Fig 6 and the detail is given in Table 2. This pathway has the following GMA model:

$$
\begin{aligned}
\frac{dX_1}{dt} &= V_{11} - V_{12} \\
\frac{dX_2}{dt} &= V_{21} - V_{22} \\
\frac{dX_3}{dt} &= V_{31} - V_{32} - V_{33} - V_{34}
\end{aligned}
\qquad (14)
$$

Table 2. Details of component concentrations in case study 2.

Variable	Initial Steady State
X_1	0.184654
X_2	7.986756
X_3	1418.931944
X_4	0.00312
X_5	5
X_6	2283
X_7	0.011
X_8	430
X_9	7.5
X_{10}	0.005
X_{11}	0.9
X_{12}	0.02
X_{13}	0

doi:10.1371/journal.pone.0126199.t002

where X_1 is the mRNA concentration, X_2 is the enzyme concentration and X_3 is the *trp* concentration. In steady state condition, all components in this pathway (variables in Table 2) have the following rate:

$$
\begin{aligned}
V_{11} &= 0.6403 X_3^{-5.87 \times 10^{-4}} X_5^{-0.8332} \\
V_{12} &= 1.0233 X_1 X_4^{0.0035} X_{11}^{0.9965} \\
V_{21} &= X_1 \\
V_{22} &= 1.4854 X_2 X_4^{-0.1349} X_{12}^{0.8651} \\
V_{31} &= 0.5534 X_2 X_3^{-0.5573} X_6^{0.5573} \\
V_{32} &= X_3 X_4 \\
V_{33} &= 0.9942 X_3^{7.0426 \times 10^{-4}} X_7 \\
V_{34} &= 0.8925 X_3^{3.5 \times 10^{-6}} X_4^{0.9760} X_8 X_9^{-0.0240} X_{10}^{-3.5 \times 10^{-6}}
\end{aligned}
\tag{15}
$$

Optimization problem. The first objective function was to maximize the *trp* production given by the reaction of V_{34} [23]. In addition, the second objective function was to minimize the total concentrations involved. Thus, the optimization problem of this case study can be defined as follows:

$$
max\, F_1 = V_{34} \tag{16}
$$

$$
min\, F_2 = \sum_{j=1}^{6} X_j + X_8 \tag{17}
$$

Similar to case study 1, the optimization problem was subjected to the steady state condition. Hence, this produced GMA models with the following form:

$$
\begin{aligned}
V_{11} - V_{12} &= 0 \\
V_{21} - V_{22} &= 0 \\
V_{31} - V_{32} - V_{33} - V_{34} &= 0
\end{aligned}
\tag{18}
$$

In this case study, not all of the components were tuned. Only seven components were tuned, which are $X_1 - X_6$ and X_8. For $X_1 - X_3$, the component concentration constraints were in the range of 0.8–1.2; for X_4, the component concentration constraints were in the range of 0–0.00624; for X_5, the component concentration constraints were in the range of 5–10; for X_6, the component concentration constraints were in the range of 500–5000 and for X_8, the component concentration constraints were in the range of 0–1000 [4–6].

Results and discussion

Results and discussion Many parameter settings were used while performing the experiments. Table 3 summarizes the parameter settings used when the best result was obtained. For the Newton method, fixed parameters were used; the maximum number of iterations was fixed to 50 and the tolerance was set to 10^{-6}.

In case study 1, the full results obtained by the NCGA are presented in Table 4. The table gave the best result achieved, the average of 100 experiments, the standard deviation and the comparison with other works. For the best ethanol production, the NCGA was able to increase the ethanol production up to 52.91 times from its initial steady state value. Several other works were compared with the NCGA in order to assess its performance. As seen in the table, the NCGA produced the highest amount of ethanol compared to the others. Besides that, the NCGA was able to reduce the total amount of the component concentrations involved, with the total concentration was only 294.80.

Table 3. Summary of parameter settings in producing the best result.

| Parameter	Case Study 1	Case Study 2
Number of sub-populations	11	7
Number of individuals in sub-populations	150	150
Maximum generation	300	350
Crossover point	2.0	1.0
Mutation rate	0.3	0.2

doi:10.1371/journal.pone.0126199.t003

Table 4. Best solution obtained using the NCGA in case study 1.

Variable	Best Solution	Average of 100 experiments	Standard deviation	Rodriguez-Acosta et al. [21]	Xu [6]	Ismail et al. [24]
X_1	1.1121	0.9940	0.1112	1.14	1.102	1.11
X_2	1.0301	1.0000	0.1136	1.05	1.046	1.03
X_3	1.1874	1.0057	0.1281	1.15	1.141	1.13
X_4	1.1707	1.1273	0.0392	1.17	1.171	1.18
X_5	0.9116	0.9839	0.1129	1.12	1.113	1.14
Y_1	49.73	49.9811	0.0271	49.97	50	49.99
Y_2	45.8097	45.0758	0.1806	44.77	45.953	45.83
Y_3	48.8933	49.9056	0.1039	49.89	50	49.92
Y_4	48.13	47.3806	0.1273	47.26	47.772	47.97
Y_5	47.85	49.3487	0.5269	48	48.366	48.30
Y_6	48.9782	49.7864	0.1262	49.75	50	49.79
F_1	52.91	52.74	0.0133	52.0843	52.5118	52.57
F_2	294.80	295.19	0.3274	295.27	297.664	297.384

doi:10.1371/journal.pone.0126199.t004

Table 5. Best solution obtained using the NCGA in case study 2.

Variable	Best Solution	Average of 100 experiments	Standard deviation	Marin-Sanguino et al. [5]	Vera et al. [4]	Xu [6]	Ismail et al. [24]
X_1	0.9053	1.0741	0.0995	1.1900	1.2000	1.2000	1.1100
X_2	0.8302	1.1091	0.0794	1.1480	1.1500	1.1150	1.1140
X_3	0.8000	0.8000	$1.56e^{-15}$	0.8000	0.8000	0.8000	0.8000
X_4	0.0054	0.0054	$1.14e^{-5}$	000041	0.0040	0.0054	0.00538
X_5	4.0172	4.4736	0.3071	4.0000	4.0000	4.0110	4.7540
X_6	5000	5000	0	5000	5000	5000	5000
X_8	1000	1000	0	1000	1000	1000	1000
F_1	3.9759	3.9614	0.0032	3.0620	3.0620	3.9460	3.9570
F_2	6006.5581	6006.9621	0.4075	6007.1412	6007.1540	6007.1314	6007.7814

doi:10.1371/journal.pone.0126199.t005

The complete result achieved using NCGA in case study 2 is given by Table 5. The table summarizes the best result achieved, the average of 100 experiments, the standard deviation and the comparison with other works. For the best solution, NCGA was able to produce *trp* up to 3.9759 times the initial steady state value. When the comparison was made with other works, the result produced by NCGA was the highest. For the total component concentrations involved, NCGA was able to reduce the total amount of the component concentrations involved to 6006.5581, where it was the lowest value compared to other works.

In evaluating the concept of the NCGA representation, it was compared with the single chromosome representation (traditional GA). Several experiments were conducted in evaluating this concept and the parameter settings as indicated in Table 3 were used. Fig 7 and Fig 8 give the comparison of the results. From both figures, it can be observed that all the production results that utilized the NCGA representation concept were higher compared to the results that only used single chromosome representation. This is because each variable in a nonlinear equations system was represented by multiple sub-chromosomes. In contrast to that, a single representation of chromosome represents all the variables of nonlinear equations system into a single chromosome. As the concept of NCGA representation allowed each variable in the nonlinear equations system to be represented by many sub-chromosomes and reproduced in their own sub-population, this made all variables were tuned in order to produce the optimum result. This did not happen in the single representation of the solution, as there were possibilities that not all variables were tuned because all variables were grouped together into only a single
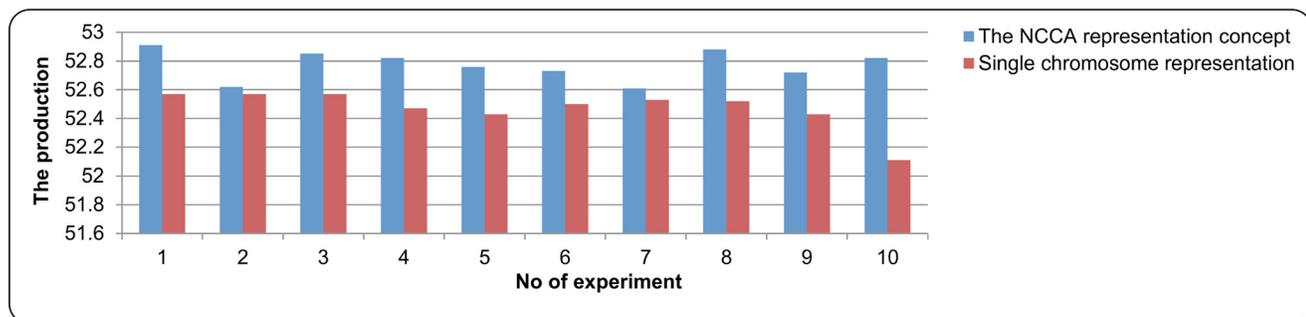
Fig 7. The comparison results of the NCGA representation concept in case study 1.
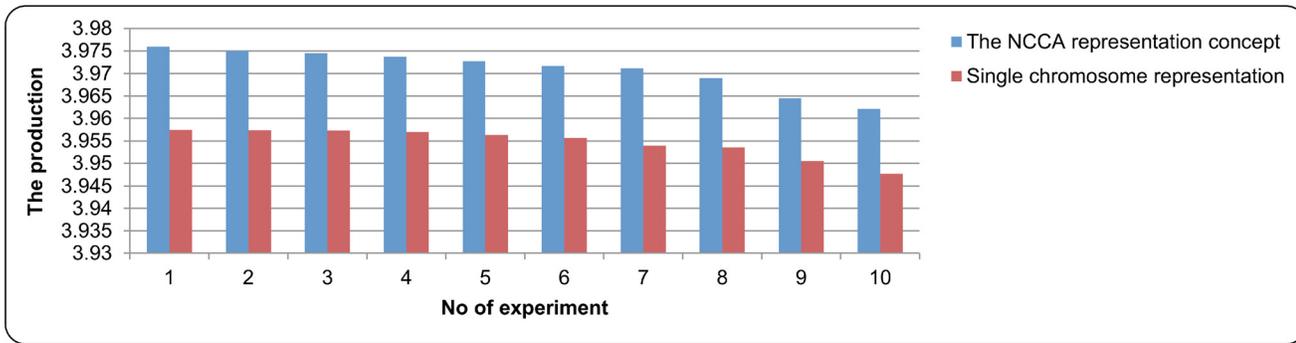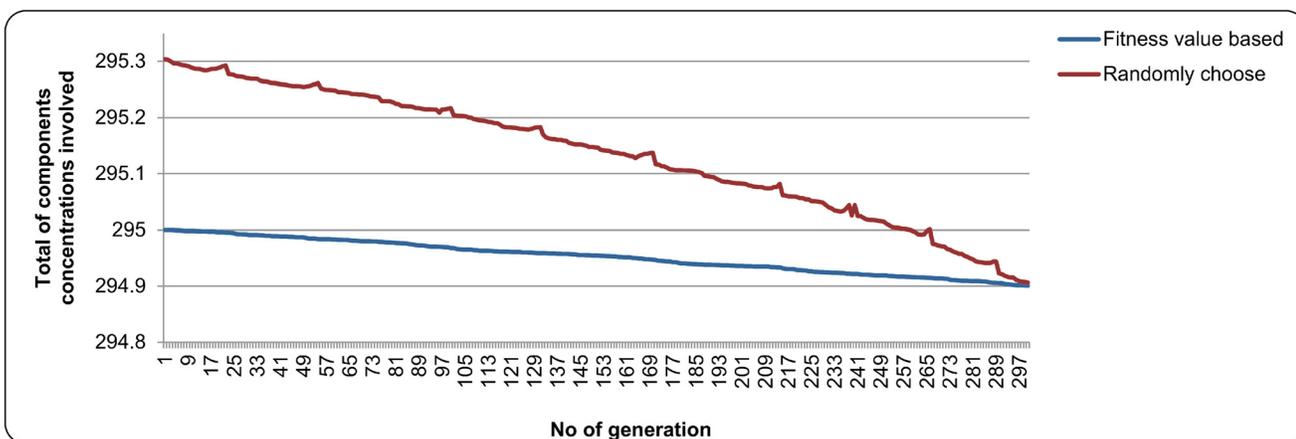
doi:10.1371/journal.pone.0126199.g007

Fig 8. The comparison results of the NCGA representation concept in case study 2.

Fig 9. The result of two-fitness evaluation concept in case study 1.

representation. As a conclusion, applying the concept of sub-chromosome and cooperative chromosome representation enables the improvement of the production.

In validating the two-level evaluation concept that was introduced in this study, several experiments were conducted where the representatives from all sub-populations were selected randomly based on their fitness value. The experiments used the parameter settings in Table 3 and the number of maximum generation was fixed to 300. The graph in Fig 9 and Fig 10 depict the results obtained in case study 1 and case study 2 respectively. In both figures, it was found that the representatives selected based on their fitness value were able to minimize the total of the component concentrations involved compared to the representatives that were selected randomly. Furthermore, the total of the component concentrations involved for the representatives that were selected based on the fitness value in Fig 9 and Fig 10 decreased slightly from the 1st generation to the 300th generation. For the total of the component concentrations involved of the randomly selected representatives, it was found that the total of component concentrations involved in the next generation sometimes were higher than the previous generation. This might be due to the random element in selecting the representatives to form cooperative chromosome. This could happen when a sub-chromosome in the current generation with higher fitness value compared to others in their sub-population was selected
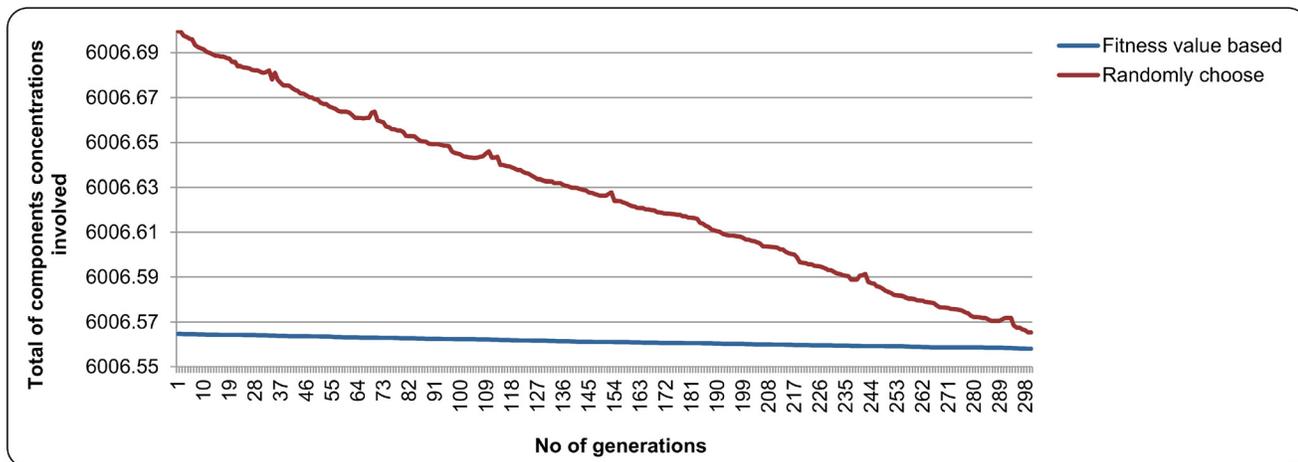
Fig 10. The result of two-fitness evaluation concept in case study 2.

randomly as a representative, and then combined with other representatives from other sub-populations. This made the total of the component concentrations involved became higher in the next generation. In conclusion, the two-level evaluation concept that was introduced in this study was able to reduce the total of components concentrations involved and allow the NCGA to perform the optimization process smoothly.

Several experiments were carried out in order to investigate the reliability of the results obtained by the NCGA. The reliability of the results can be assessed by the average results that are given in Table 4 for case study 1 and Table 5 for case study 2. For case study 1, it could be observed that all component concentrations involved were in their optimum range, thus proved that the NCGA was able to produce reliable results. In addition, the ethanol production rate was higher and the total of the component concentrations involved was the lowest compared to the previous works. Based on this observation, it can be concluded that the NCGA demonstrated the reliability in handling the optimization problem in case study 1. In case study 2, it was found that all component concentrations involved were placed in their optimal range, therefore confirmed that the NCGA was capable of producing reliable result. Besides that, the *trp* production rate was higher and the NCGA was able to reduce more for the total of the component concentrations involved compared to previous works. Based on this finding, the NCGA has shown its ability in the optimization of case study 2.

Besides comparing with previous works, the production of NCGA was also being compared with the Newton method with traditional GA (single chromosome representation). 100 experiments were performed and the best result, the average and the standard deviation were recorded. For the results that were produced by NCGA in case study 1 and case study 2 are given in Table 4 and Table 5 5 respectively, while the result that was produced by the single chromosome representation is as follows; for case study 1, the best solution result for the ethanol production rate is 52.57 and the total of the component concentrations involved is 297.38, the average for the ethanol production rate is 52.40 and the total of the component concentrations involved is 296.65, and the standard deviation for the ethanol production rate is 0.0826 and the total of the component concentrations involved is 0.3944; for case study 2, the best result for the *trp* production rate is 3.9570 and the total of the component concentrations involved is 6007.7884, the average for the *trp* production rate is 3.9510 and the total of the component concentrations involved is 6007.9811, and the standard deviation for the *trp* production rate is

Table 6. The comparison of computation time.

Method	Case study 1 (s)	Case study 2 (s)
NCGA	80.40	40.45
Newton method with traditional GA	80.25	40.31

doi:10.1371/journal.pone.0126199.t006

0.0049 and the total of the component concentrations involved is 0.4872. It can be observed that all the production result produced by NCGA was higher compared to the result that was produced by traditional GA in terms of the best solution and the average for all case studies. For the total of the component concentrations involved results (the best solution and the average), NCGA was able to reduce more compared to the Newton method with traditional GA. For the standard deviation, NCGA was able to achieve smaller value of standard deviation compared to the Newton method with traditional GA. This shows that the NCGA is able to find the similar result. This is because with smaller value of the standard deviation points out the consistency and precision of the NCGA.

Other than the production and the total of the component concentrations involved, the computation time (in second) in performing experiment also needs to take into account. Table 6 gives the comparison of the computation time taken in performing experiment for NCGA and Newton method with traditional GA. Generally, the result shows that the NCGA requires more computation time compared to the Newton method with traditional GA. This might be due to the concept of the NCGA representation and the two-level evaluation concept where NCGA needs more time to perform these two processes in order to obtain good results compared to the Newton method with traditional GA.

Conclusion

In this paper, an improved method for *in silico* optimization of metabolic pathway production known as NCGA has been proposed. The NCGA was developed to improve the metabolic pathway production and at the same time minimize the total concentration of the components involved. The NCGA comprises a combination of Newton method, GA and CCA. The Newton method deals with the metabolic pathway and the GA is employed in the optimization process, with the GA representing the components in the metabolic pathway as a solution. However, as explained in this paper, the representation of the solution becomes complex when it is applied in a complex metabolic pathway. To overcome this situation, the use of the CCA was proposed in order to simplify the representation of the solution by decomposing the solution into multiple sub-solutions. In addition, this study introduced two concepts to make the NCGA perform well, namely the concept of representing the solution and a two-level evaluation of the solution. The proposed method was applied in two case studies, and it showed better results than those reported in previous studies.

Acknowledgments

We would also like to express our gratitude to the editor and anonymous reviewers who reviewed this paper.

Author Contributions

References

1. Link H, Vera J, Weuster-Botz D, Darias NT, Franco-Lara E. Multi-objective steady state optimization of biochemical reaction networks using a constrained genetic algorithm. Computers and Chemical Engineering. 2008; 32(8):1707–1713. doi: 10.1016/j.compchemeng.2007.08.009

2. Xu G. Bi-objective optimization of biochemical systems by linear programming. Applied Mathematics and Computation. 2012; 218(14):7562–7572. doi: 10.1016/j.amc.2012.01.023

3. Marin-Sanguino A, Torres NV. Optimization of biochemical systems by linear programming and general mass action model representations. Mathematical Biosciences. 2003; 184(2):187–200. doi: 10.1016/S0025-5564(03)00046-4 PMID: 12832147

4. Vera J, Gonzalez-Alcon C, Marin-Sanguino A, Torres N. Optimization of biochemical systems through mathematical programming: Methods and applications. Computers & Operations Research. 2010; 37 (8):1427–1438. doi: 10.1016/j.cor.2009.02.021

5. Marin-Sanguino A, Voit EO, Gonzalez-Alcon C, Torres NV. Optimization of biotechnological systems through geometric programming. Theoretical Biology and MedicalModelling. 2007; 4:38–54. doi: 10.1186/1742-4682-4-38

6. Xu G. Steady-state optimization of biochemical systems through geometric programming. European Journal of Operational Reseach. 2013; 225(1):12–20. doi: 10.1016/j.ejor.2012.07.026

7. Mariano AP, Costa CBB, de Angelis DdF, Filho FM, Atala DIP, Maciel MRW, et al. Optimization strategies based on sequential quadratic programming applied for a fermentation process for butanol production. Applied Biochemistry and Biotechnology. 2009; 159(2):366–381. doi: 10.1007/s12010-008-8450-6

8. Balsa-Canto E, Banga JR, Egea JA, Fernandez-Villaverde A, Hijas-Liste GM. Global Optimization in Systems Biology: Stochastic Methods and Their Applications. In: Goryanin II, Goryachev AB, editors. Advances in Systems Biology. vol. 736. Springer New York; 2012. p. 409–424.

9. Grosan C, Abraham A. A New Approach for Solving Nonlinear Equations Systems. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on. 2008; 38(3):698–714. doi: 10.1109/TSMCA.2008.918599

10. Luo YZ, Tang GJ, Zhou LN. Hybrid approach for solving systems of nonlinear equations using chaos optimization and quasi-Newton method. Applied Soft Computing. 2008 Mar; 8(2):1068–1073. doi: 10.1016/j.asoc.2007.05.013

11. Ma W, Bai ZJ. A regularized directional derivative-based Newton method for inverse singular value problems. Inverse Problems. 2012; 28(12). doi: 10.1088/0266-5611/28/12/125001

12. Gu J, Gu M, Cao C, Gu X. A novel competitive co-evolutionary quantum genetic algorithm for stochastic job shop scheduling problem. Computers & Operations Research. 2010 May; 37(5):927–937. doi: 10.1016/j.cor.2009.07.002

13. Ismail MA, Asmuni H, Othman MR. The fuzzy cooperative genetic algorithm (FCoGA): The optimisation of a fuzzy model through incorporation of a cooperative coevolutionary method. Journal of Computing. 2011; 3(11):81–90.

14. Yu Y, Xinjie Y. Cooperative Coevolutionary Genetic Algorithm for Digital IIR Filter Design. IEEE Transactions on Industrial Electronics. 2007 Jun; 54(3):1311–1318. doi: 10.1109/TIE.2007.893063

15. Voit EO. Biochemical Systems Theory: A Review. ISRN Biomathematics. 2013; 2013. doi: 10.1155/2013/897658

16. Savageau MA. Biochemical systems analysis: a study of function and design in molecular biology. Reading, Massachusetts: Addison–Wesley; 1976.

17. Babaei M. A general approach to approximate solutions of nonlinear differential equations using particle swarm optimization. Applied Soft Computing. 2013 Jul; 13(7):3354–3365. doi: 10.1016/j.asoc.2013.02.005

18. Durillo JJ, Nebro AJ. jMetal: A Java framework for multi-objective optimization. Advances in Engineering Software. 2011 Oct; 42(10):760–771. doi: 10.1016/j.advengsoft.2011.05.014

19. Hicklin J, Moler C, Webb P, Boisvert RF, Miller B, Pozo R, et al. JAMA: A Java Matrix Package; 2012. Available from: http://math.nist.gov/javanumerics/jama/.

20. Galazzo JL, Bailey JE. Fermentation pathway kinetics and metabolic flux control in suspended and immobilized Saccharomyces cerevisiae. Enzyme and Microbial Technology. 1990; 12(3):162–172. doi: 10.1016/0141-0229(90)90033-M

21. Rodriguez-Acosta F, Regalado CM, Torres NV. Non-linear optimization of biotechnological processes by stochastic algorithms: Application to the maximization of the production rate of ethanol, glycerol and carbohydrates by Saccharomyces cerevisiae. Journal of Biotechnology. 1999; 65(1):15–28. doi: 10.1016/S0168-1656(98)00178-3

22. Xiu ZL, Zeng AP, Deckwer WD. Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. Journal of Biotechnology. 1997 Oct; 58(2):125–140. doi: 10.1016/S0168-1656(97)00143-0

23. Marin-Sanguino A, Torres NV. Optimization of tryptophan production in bacteria. Design of a strategy for genetic manipulation of the tryptophan operon for tryptophan flux maximization. Biotechnology Progress. 2000; 16(2):133–145. doi: 10.1021/bp990144l PMID: 10753437

24. Ismail MA, Deris S, Mohamad MS, Abdullah A. A Hybrid of Newton Method and Genetic Algorithm for Constrained Optimization method of the Production of Metabolic Pathway. Life Science Journal. 2014; 11(9s):409–414.