Producing Succinic Acid in Yeast using A Hybrid of Differential Evolution and Flux Balance Analysis

Ana Haziqah A.Rashid¹, Yee Wen Choon¹, Mohd Saberi Mohamad¹, Lian En Chai¹, Chuii Khim Chong¹, Safaai Deris¹ and Rosli M Illias²

¹Artificial Intelligence and BioinformaticsResearchGroup, Faculty of Computing, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia ²Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia

Abstract

In the field of metabolic engineering, one of the primary goals is to maximize the production of a desired substance. However, to identify the set of gene deletions that will result to the desired production goals is difficult. This is due to the complexity of the regulatory cellular and metabolic network and also lack of good modeling and optimization tools. In this study, the optimization algorithm from previous works was implemented to identify the gene knockout on the result. The previous works faced the problem of inability to provide single run with two goals to maximize the biomass and the desired products. Besides that, the previous work also showed long computational time. In this study, a hybrid of Differential Evolution and Flux Balance Analysis (DEFBA) is proposed to solve the long computational time problem and provide an optimal set of gene knockout with high yield of the desired product. The case study in this research involved the production of succinic acid (also called as succinate) in yeast Saccharomyces cerevisiae. The results from this experiment included the list of knockout genes and the growth rate after the deletion. DEFBA had shown better results compared to the other methods. The identified list suggested gene modifications over several pathways which may be useful in solving challenging genetic engineering problems.

Keywords: Differential Evolution, Flux Balance Analysis, Gene Deletions, Succinic Acid, Yeast

1. Introduction

Saccharomyces cerevisiae is a robust, well-established industrial production organism which exhibits very good growth characteristics, a broad substrate spectrum and an extraordinarily high acid and osmotolerance. The high tolerance towards acidity is a major advantage over bacterial succinic acid production hosts, since it uses neutralization salts dispensable, which facilitates the downstream process. Most specialty and commodity applications of succinic acid require a free acid form rather than a salt form. Other than that, *Saccharomyces cerevisiae* is genome-sequenced; genetically and physiologically well characterized, and best as a tool for genetic optimization. These features made yeast particularly suitable for the biotechnological production of succinic acid.

The purpose of this research is to develop a computational approach to analyze the production of succinic acid in yeast *Saccharomyces cerevisiae*. Previously, Maranas *et al.*, (2003) developed the first rational modeling frameworks (named OptKnock) for introducing gene knockout leading to the overproduction of a desired metabolite.

OptKnock identifies a set of gene (reaction) deletions to maximize the flux of a desired metabolite without affecting the internal flux distribution such that growth is optimized. OptKnock is implementing by using mixed integer linear programming (MILP) to formulate a bi-level linear optimization that is very promising to find the global optimal solution [1]. OptGene is an extended approach of OptKnock which formulates the in silico design problem by using Genetic Algorithm (GA). These meta-heuristic methods are capable in resulting near-optimal solutions with reasonable computation time, furthermore the objective function that can be optimized is flexible [2]. However, these methods still face some problems. In this research, we compared and focused on the Evolutionary Algorithm (EAs) and OptGene. Evolutionary Algorithm (EAs) that is referred to the combination of SA and flux balance analysis (FBA) faces the premature convergence into a local extremes problem, which may result from adverse configuration and not yield (a point near) the global extremes [3]. For OptGene, the problem is to set a prior number of genes to be deleted from an individual so as to obtain a desired phenotype [2]. In this research, we used gene deletions to obtain high yield of succinic acid in Saccharomyces cerevisiae. Since Differential Evolution (DE) algorithm is a stochastic and non-deterministic approach for solving polynomial fitting problems [4], we proposed DEFBA in this research to solve the problems from the previous methods. In this research, we focused on enhancing the production of succinic acid in yeast by using DEFBA.

2. Method

Differential Evolution

Differential evolution (DE) is a stochastic optimization method to minimize the objective function that can model the problem's objectives while incorporating constrains. Each design variable in DE is represented in the chromosome by a real number. DE algorithm is simple and requires only three control parameters which are weight factor (F), crossover rates (CR), and population size (NP). In this study, the initial population was randomly generated by uniformly distributed random numbers using the upper and lower limitation of each design variable. After that, the objective function values of all the individuals of population were calculated. This step was done to find out the best individual *x*best, G of current generation, in which G is the index of generation. Bao *et al.* (2010) stated that there are three main steps of DE, which are mutation, crossover, and selection [5]. These steps were performed sequentially and were repeated during the optimization cycle. Figure 1 shows the original basic step of DE. Figure 3 shows the implementation of the fitness function, FBA, to produce DEFBA algorithm.

Model Pre-Processing

To have only useful data for the research, model preprocessing must be done to have accurate results. As stated before, this research would using yeast *Saccharomyces cerevisiae* as a model. Here, we used 'yeast_4.05.xml' for the data. This model also needed to undergo the model preprocessing steps. This pre-processing step functioned to remove all the dead end metabolites and reaction from the model. Besides that, it also functioned to remove all the reactions that would not be used (max and min are < tol), and find the minimal bounds for the flux through each reaction and provide the result for flux variability analysis (maxes, mins). Next, we performed single GeneDeletion function to make single gene deletion analysis using FBA, MOMA or linearMOMA; but in this case, we used FBA. Lastly, we deleted the zero (0)

value in the model. After pre-processing had been done, the number of metabolites was reduced from 1461 to 768 and the numbers of reactions had been reduced from 1865 to 1246.

After going through the model pre-processing, it was analyzed using DEFBA algorithm to get the gene deletions. In the next section, we provide a brief explanation on the core steps in Differential Evolution.

Mutation

In this research, each reaction in the metabolic model can be associated with one or more genes in the genome and each of those genes is represented by a binary variable indicating the absence (0) or the presence (1) of the gene in the reaction, and thus a set of these variables forms a chromosome representing a particular mutant that lacks some metabolic reactions when compared with the wild type. For each individual vector x_i , G in the population, mutation operation is used to generate mutated vectors in DE. Equation (1) below shows the calculation of mutation operation:

 $v_{i},G+1 = x_{best},G + F(x_{r1},G-x_{r2},G), i=1,2,3.....NP$ (1)

In this equation, vector indices r1 and r2 are different population index and randomly selected. After selecting two vectors, x_{r1} ,G and x_{r2} , G will then be used as differential variation for the mutation. Therefore, the best solution of current generation is vector x_{best} ,G while v_{is} G+1 is the best target vector and mutation vector of current generation. Weight factor F is the real value between 0 to 1 and it controls the amplification of the differential variation between the two random vectors. There are different mutation mechanisms available for DE which may be applied in the optimization search process (Wu *et al*, 2008). However, in this research we applied DE/rand/1/bin since /bin likes to have a slightly larger CR than /exp.

Crossover

In crossover, the trial vector $u_i,G+1$ is generated by choosing some parts of mutation vector, $v_i,G+1$ and other parts come from the target vector x_i,G . In Figure 2, Cr represents the crossover probability and j is the design variable component number. If random number R is larger than Cr value, the component of mutation vector, $v_i, G+1$, will be chosen into the trial vector. Otherwise, the component of target vector is selected into the trial vectors (Wu *et al*, 2008). The mutation and crossover operators are used to diversify the search area of optimization problems.

Selection

All solutions in the population have the same chance of being selected as parents without dependence on their fitness value. The child produced after the mutation and crossover operations will then be evaluated. Then, the performance of the child vector and its parent will be compared and the better one will be selected. If the parent was still better, it would be retained in the population. After that, the optimization loop of DE will be run iteratively until the stop criteria are met. There are three stop criteria used in the program. One of the stop criterions is the maximum number of optimization generation. The second criterion is the maximum number of consecutive generations when there is no better global optimum found in the whole process. Therefore, if the improvement of objective function between two consecutive generations is less than the threshold set by program, it will be considered as the fitting convergence requirement. The last stop criterion will be conformed if the accumulated number of generations fitted convergence requirement is greater than maximum counter set by the program [6].

International Journal of Bio-Science and Bio-Technology Vol.5, No.6 (2013)



Figure 1. The Flowchart of Basic Differential Evolution



Figure 2. The Crossover Operator [6]

A Hybrid of Differential Evolution and Flux Balance Analysis

As mentioned before, Flux Balance Analysis (FBA) would be used to replace and calculate the scoring fitness in Differential Evolution (DE). Figure 3 shows the flowchart of DEFBA. The fitness computation process was carried out for each site visited by DE through FBA. Cellular growth is defined as the objective function Z, vector \mathbf{c} is used to select a linear combination of metabolic fluxes to include in the objective function, \mathbf{v} is the flux map and *i* is the index variable (1, 2, 3, ..., n).

Maximize Z, where

$$Z = \sum_{i} c_i v_i = \mathbf{c} \cdot \mathbf{v}$$

where c = a vector that defines the weights for of each flux.

International Journal of Bio-Science and Bio-Technology Vol.5, No.6 (2013)



Figure 3. The Flowchart of DEFBA

3. Results and Discussion

In this research, the *Saccharomyces cerevisiae* dataset was used to test the operation of DEFBA. The glucose uptake rate was fixed to 10 mmol/gDW/hr while a set non-growth was associated with the maintenance of 7.6 mmol ATP/gDW/hr. The results obtained were compared to Opt Gene by Patil *et al.* [2] and Simmulated Annealing with Flux Balance Analysis (SA+FBA) by Rocha *et al.* [3]. To test the accuracy of DEFBA algorithm, we had validated our list of gene knockout with the database like PathCase and yeastgenome.org. Every deleted reaction was checked with the database and journal that were related to the reaction to make sure the deletions that had been done were correct. In this paper, we are unable to compare the computational performance with previous methods as it is not reported in the original papers.

Table 1 shows the comparison of different methods for producing succinate. As shown in the results, this method produced better results than the previous works. DEFBA suggested the removal of five reactions from the network results in succinate growth rate reaching 1.7023 which is better than other previous works. OptGene [2] also suggested five reactions to be deleted, but it did not achieve high growth rate; 0.39 compared to DEFBA. For SA with FBA [3], it showed the lowest growth rate; 0.05398 and a very large number of deletions to be made to enhance the production of succinate.

 Table 1. Result Comparison of DEFBA with OptGene and SA+FBA for

 Producing Succinic Acid in Yeast

Methods	Growth Rate	List of Knockout	
	(1/hr)		
DEFBA	1.7023	SDH1, ADH1, GDH2, PDA1, OAC1	
OptGene	0.39	SDH-complex, ZWF1, PDC6, U133, U221	
(Patil et al, 2005)			
SA + FBA	0.05398	PGI1_1, PGI1_2, FBP1, PDC6, ADH4, SDH3_2,	
(Rocha et al,		AAH1_1, URH1_1, U30_, MET3, ALD4_2,	
2008)		GSH1,U103_, YER053C, CTP1_1	

Table 2 shows three of the identified gene knockout strategies (*i.e.*, mutants A, B, and C). In this research, potential genes which can be removed had been identified. Each mutant produced a set of gene deletions that was identified by using the algorithm. Table 2 shows, we can see that mutant B obtained the highest production yield of succinate, followed by mutant

C and mutant A. This amount of production yield depends on the gene knockout. More detail and explanation of the list of gene knockout are discussed below. For the production of succinate, alcohol dehydrogenase (ADH1) inhibited the reduction of acetaldehyde to ethanol. Therefore, the acetaldehyde can be fully used in the glycoxylate to produce succinate without bothering to be used to produce ethanol [7]. As we can see from the table, ADH1 mostly involved as the gene to be knocked out because it was in the list of knockout of each mutant. This shows that ADH1 plays an important role in enhancing the production of succinate. Mutant B and C showed that succinate dehydrogenase, SDH1 was also one of the genes to be knocked out. SDH1 was involved in the TCA cycle which converted succinate to fumarate. To enhance the production of succinate, we must reduce the utilization of succinate. Therefore, the deletion of SDH1 would lead to the production of succinate [8]. OAC 1 which encoded the 2-isopropylmalate transport was involved in the deletion of the three mutants. The OCA1 functioned as the mitochondrial inner membrane transporter, which transported oxaloacetate, sulfate, thiosulfate, and isopropyl malate; which are members of the mitochondrial carrier family [9]. TCA cycle occurred in the mitochondria and glycoxylate cycle occurred in the cytosol. Therefore, when OAC1 had been deleted, the oxaloacetate and isopropyl malate remained in the mitochondria and helps the TCA cycle. This activated the TCA cycle and produced succinate in the reductive pathways. Deletion of SER3 in mutant A blocked the synthesis of L-Serine via 3-Phospho-D-glycerate, which increased the demand on glycine production via glyoxylate. Overall, it led to a further increase in the flux through ICL1, ensuring a higher flux towards succinate [10]. Mutant B showed PDA1 which encoded pyruvate dehydrogenase as one of the list knockouts. PDA1 catalyzed the direct oxidative decarboxylation of pyruvate to acetyl-CoA and was regulated by glucose. Mutant C knocked out the PDC1 which encoded pyruvate decarboxylases, which then increased the conversion of pyruvate into oxaloacetic acid via PYCI. ERG10 which encoded acetyl-CoA Cacetyltransferase was a cytosolic enzyme that functioned to transfer an acetyl group from one acetyl-CoA molecule to another. The knockout of ERG10 thus reduced the usage of acetyl-CoA, which could be used in the cycle to produce succinate.

As we can see from Table 2, the growth rates among three mutants were much similar to each other. In this research, 50 runs were carried out individually to look for the accuracy of this method. To define the accuracy, we had calculated the standard deviation of growth rate for this research. The standard deviation we obtained from this research was 0.1435. This value is categorized as low standard deviation. A low standard deviation indicates that the data points tend to be very close to the mean.

Table 2. Result of Different Knockout Strategies for SuccinateProduction

Mutants	Growth	Rate	Production	yield	List of knockout
	(1/hr)		(mmol/gDW/hr)		
А	1.7021		5.1226		ADH1, GAD1, ERG10, SER3 OD1C2
В	1.7023		8.2097		SDH1, ADH1, GDH2, PDA1, OAC1
С	1.7023		6.9244		SDH1, ADH1, PDC1, ERG10, OAC1

In this research, we studied the production of succinic acid in yeast by using *in silico* method. The *in silico* was done because of the limitation that might occur when using *in vivo* method. Heerde *et al* (1978) studied the metabolism of the anaerobic formation of succinic acid by *Saccharomyces cerevisiae* In this research, we included a comparison of *in vivo* and *in silico* methods. According to Heerde *et al.* (1978), Succinic acid can be formed up to 2g/l in all product fermentations [11]. In *Saccharomyces cerevisiae*, normally,

the range of succinic acid formed is 0.2 until 1.7g/l. The growth rate obtained from this research was 1.7023mmol. After converting the 1.7023mmol into g/l, we obtained 0.20087g/l which is in the range of the in vivo succinic acid production.

4. Conclusion and Future Works

In short, we can conclude that DEFBA is an efficient algorithm to predict the optimal sets of gene deletion in order to maximize the production of succinate. This is based on DE algorithm which is a population based algorithm like the genetic algorithms using similar operators; crossover, mutation and selection. The main difference in constructing better solutions is that genetic algorithms rely on crossover while DE relies on mutation operation, in which this main operation is based on the differences of randomly sampled pairs of solutions in the population. FBA approach is based on a steady state approximation to concentrations of the internal metabolites, used as a fitness function, which reduces the corresponding mass balances into a set of linear homogeneous equations. In this work, we propose DEFBA to predict optimal sets of gene deletions to maximize the production of desired phenotype. DEFBA, which is based on DE, is an efficient algorithm in constructing better solutions, where in this paper, its performance is improved in identifying optimal gene knockout strategies. Hence, it is a useful tool in metabolic engineering. With regard to further improve the performance of DEFBA, we are developing optimization algorithms in a single run to achieve two goals, for example, maximizing the biomass and the desired product. Lastly, DE is a computationally efficient and natural tool for applying gene knockout strategies to enhance the succinate. For future research, we suggest researches on lactate and malate production in yeast Saccharomyces cerevisiae. It will be interesting because all of these metabolites are actually related to one another. Other than that, an effort to study hybrid other optimization also could be interesting. From hybrid to the optimization method, we can determine whether the optimization method is best to operate alone or best if combined with another optimization method.

Acknowledgements

We would like to thank Malaysian Ministry of Higher Education for supporting this research by granting an Exploratory Research Grant Scheme (Grant number: R.J130000.7807.4L096). This research is also funded by an e-science research grant (Grant number: 06-01-06-SF1029) from Malaysian Ministry of Science, Technology and Innovation. We would also like to thank Universiti Teknologi Malaysia for supporting this research by granting a GUP research grant (Grant number: Q.J130000.2507.04H16).

References

- A. P. Burgard, P. Pharkya and C. D. Maranas, "OptKnock: A bilevel programmingframework for identifying gene knockout strategies for microbial strains optimization. Biotechnol", Bioeng, vol. 84, (2003), pp. 647– 657
- [2] K. R. Patil, I. Rocha, J. Förster and J. Nielsen, "Evolutionary programming as a platform for in silico metabolic engineering", BMC Bioinformatics, vol. 6, no. 308, (2005).
- [3] M. Rocha, P. Maia, R. Mendes, J. P. Pinto, E. C. Ferreira, J. Nielsen, K. R. Patil and I. Rocha, "Natural computation meta-heuristics for the in silico optimization of microbial strains", BMC Bioinformatics, vol. 9, no. 499, (2008).

- [4] R. Storn and K. Price, "Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces", Journal of Global Optimization, vol. 11, (**1997**), pp. 341–359.
- [5] J. Bao, Y. Chen and J. Yu, "A regeneratable dynamic differential evolution algorithm for neural networks with integer weights", SpringerLink, vol. 11, no. 12, (2010), pp. 939-947.
- [6] C. Wu and K. Tseng, "Topology Optimization of Structure Using Differential Evolution", J. Syst. Cybern. Inform, 2008 - iiisci.org, (2008).
- [7] V. Leskovaca, S. Trivićb, G. Wohlfahrte, J. Kandračd and D. Peričin, "Glucose oxidase from Aspergillus niger: the mechanism of action with molecular oxygen, quinones, and one-electron acceptors", The International Journal of Biochemistry & Cell Biology, vol. 37, no. 4, pp. 731-750.
- [8] K. S. Oyedotun and B. D. Lemire, "The Quaternary Structure of the Saccharomyces cerevisiae Succinate Dehydrogenase Homology Modeling, Cofactor Docking, and Molecular Dynamics Simulation Studies", The Journal of Biological Chemistry, vol. 279, (2003), pp. 9424-9431.
- [9] G. Palmieri, P. Giardina, C. Bianco, B. Fontanella and G. Sannia, "Copper Induction of Laccase Isoenzymes in the Ligninolytic Fungus Pleurotus ostreatus", Appl. Environ. Microbiol, vol. 66, no. 3, (2000), pp. 920-924.
- [10] A. J. A. van Maris, D. A. Abbott, E. Bellissimi, J. van den Brink, M. Kuyper, M. A. H. Luttik, H. Wouter Wisselink, W. Alexander Scheffers, J. P. van Dijken and J. T. Pronk, "Alcoholic fermentation of carbon sources in biomass hydrolysates by Saccharomyces cerevisiae: current status", Antonie van Leeuwenhoek, vol. 90, no. 4, (2006), pp. 391-418.
- [11] E. Heerde and F. Radler, "Metabolism of the anaerobic formation of succinic acid by Saccharomyces cerevisiae", Archives of Microbiology, vol. 117, no. 3, (1978), pp. 269-276.

Authors



Ana Haziqah A.Rashid, received her BSc degree in Computer Science (Bioinformatics) from Universiti Teknologi Malaysia in 2012. She has interests in computational methods such as differential evolution algorithm, flux balance analysis, minimization of metabolic adjustment, and genetic algorithm.



Yee Wen Choon, is a PhD student of Computer Science (Bioinformatics) in Universiti Teknologi Malaysia. Her research interests include evolutionary algorithms, metabolic engineering, and programming. She has published 2 international referred publication.



Mohd Saberi Mohamad, was born in Malaysia, on February 03, 1980. He received the BSc and MSc degrees in Computer Science both from Universiti Teknologi Malaysia, in 2002 and 2005, respectively. He received the PhD degree in Intelligent Systems for Bioinformatics from Osaka Prefecture University in 2010. He is now a senior lecturer at Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, Malaysia. His journal papers in the field of bioinformatics using computational intelligence approaches have been published with the total number is more than 20 papers. He has interests in computational methods such as particle swarm optimizers, hybrid approaches, genetic algorithms, support vector machines, and neural networks.



Lian En Chai, is a postgraduate student at the Artificial Intelligence and Bioinformatics Research Group, Faculty of Computer Science and Information System, Universiti Teknologi Malaysia. His current research interests involve modelling gene networks using statistical methods, including DBN, as well as analysis of cDNA microarray gene expression data.



Chuii Khim Chong, is a PhD student of Computer Science (Bioinformatics) in Universiti Teknologi Malaysia. Her research interests include evolutionary algorithms, database, and programming. She has published 2 international referred publications.



Safaai Deris, (M'98) was born in Malaysia on August 13, 1955. He is a Professor of Artificial Intelligence and Software Engineering at Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia. He received the M. E. degree in Industrial Engineering, and the Doctor of Engineering degree in Computer and System Sciences, both from Osaka Prefecture University, Japan, in 1989 and 1997, respectively. His recent academic interests include the application of intelligent techniques in scheduling and bioinformatics. He is also a reviewer for several refereed journals.



Rosli Md. Illias, was born in Malaysia on December 24, 1968. He is a Professor of Department of Bioprocess Engineering at Faculty of Chemical Engineering, Universiti Teknologi Malaysia. He received the BSc degress in Microbiology from Universiti Kebangsaan Malaysia, Malaysia, in 1992. He received the PhD degree in Molecular Biology from Edinburgh University, Scotland, in 1997. His recent academic interests include Genetic engineering, Enzyme Biotechnology and Molecular Biology.

International Journal of Bio-Science and Bio-Technology Vol.5, No.6 (2013)