Using Ant Colony Optimization (ACO) on Kinetic Modeling of the Acetoin Production in *Lactococcus Lactis C7*

Nor Farhah Binti Saidin[1], Chuii Khim Chong[1], Yee Wen Choon[1], Lian En Chai[1],
Safaai Deris[1], Rosli M. Illias[2], Mohd Shahir Shamsir[3], and Mohd Saberi Mohamad[1,*]

[1] Artificial Intelligence and Bioinformatics ResearchGroup, Faculty of ComputerScience and InformationSystems, Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia
`{nfarhah3,ckchong2,ywchoon2,lechai2}@live.utm.my`,
`{safaai,saberi}@utm.my`
[2] Department of Bioprocess Engineering, Faculty of Chemical Engineering,
Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia
`r-rosli@utm.my`
[3] Department of Biological Sciences, Faculty of Biosciences and Bioengineering,
Universiti Teknologi Malaysia, Skudai, 81310 Johor, Malaysia
`shahir@fbb.utm.my`

Abstract. Ant colony optimization (ACO) is a population-based meta-heuristics that can be used to find approximate solutions based on ant's behavior patterns for solving difficult combinatorial optimization problems. To apply Ant colony optimization (ACO), the artificial ants incrementally build solutions by moving on the graph as their results. The solution construction process is stochastic and is biased by a pheromone model. It is a set of parameters associated with graph components which either nodes or edges whose values are modified at runtime by the ants. Best value of kinetic parameters from the experimental data can be obtained by implementing the ant colony optimization (ACO). Model development that can represent biochemical systems is one of the hallmarks of systems biology. Scientists have been gathering data from actual experiments but there is a lack in computer models that can be used by scientists in analyzing the various biochemical systems more effective. However it is also time consuming and expensive and rarely produce large and accurate data sets when carried out in wet lab. Parameter estimation is used to adjust the model to reproduce the experimental results in the best possible way for a set of experimental data. *Lactococcus Lactis C7* will be used as the dataset and functions as the benchmark dataset for the experiments. The results were gathered by conducting some steps in SBToolBox. The result and discussions sections which include the comparison on the performances in term of computational time, average of error rate, standard deviation and production of graph. ACO shows a better result compared to the other algorithms like Simulated Annealing (SA) and Simplex algorithms. The method used in this research also can be used for other datasets as well.

---

[*] Corresponding author.

A.S. Sidhu & S.K. Dhillon (Eds.): *Adv. in Biomedical Infrastructure 2013*, SCI 477, pp. 25–35.
DOI: 10.1007/978-3-642-37137-0_5

Keywords: Ant Colony Optimization (ACO), meta-heuristics, Lactic Acid Bacteria, Ant Behavior.

1 Introduction

For Biological investigations, there are several good reasons for inserting or adding kinetic modeling to improve skills. There are a lot of information and data contained in dynamic biological data than can be extracted by using simple procedures. If the collection of data that has just been collected contains an unknown answer, then an investment in computing can give a competitive advantage both in new knowledge and in productive experimental design.

Performing a lab experiment is costly as lab experimentation requires a substantial budget. Besides that, it rarely produces accurate data sets as because of the limit of accuracy of the measuring instruments. It is possible to calculate the limits of accuracy of different measuring instruments. Errors which are not quantifiable also occur in experiments and actual effect cannot be calculated.

Manually following the tradition, sometimes the experimental data is accepted as true, and theories have been modified to make the next prediction match the data. Meanwhile in diagram, the pathway leading from a mismatch to a new hypothesis is label as "imagination." This is purely human activity and computer helps nothing here. However, that modeling still helps because it enables human to know with precision what your imagination must accomplish. Finally, the advantage of the model for the design of informative experiments can be used.

Parameter estimation task is well known as an optimization problem that minimizes an objective function in measuring a generalized gap between the experimental data and model predictions [1]. Usually, researchers will use the Euclidean distance which is commonly referred to as least-squares error criterion. There are two objective functions commonly used which are concentration error based objective function and slope error based objective function. The concentration error based objective function is a straightforward method which calculates the sum of squared distances between the metabolite measurements and the predictions. Meanwhile, the slope error based objective function employs the decoupling techniques and uses the slope information for evaluating fitness of the function. The most effective methods for parameter estimation from time series data can be classified into gradient based methods, stochastic search algorithms and other algorithms.

Two main approaches, deterministic estimation and stochastic estimation are the two common approaches in parameter estimation. Deterministic optimisation can performs effectively in small scale linear optimisation problem. In contrast, the stochastic optimisation performance well with nonlinear problem but with no fixed setting for the control parameters; this effects the accuracy of the estimation [2]. In other words, when most of the interactions in system biology fail to be clearly studied and show its attribute, the implementaion of stochastic optimisation is very useful in such condition [3].

Global optimization method is one of the ways in parameter estimation. It needs high computational time complexity as well but they can find better values and the time taken is not as much as experiments [4]. Ant Colony Optimization (ACO) algorithm is one of the global optimization methods that are appropriate to be used for the parameter estimation from experimental results. According to Alonso *et al.* [5], the ACO algorithm produced better results on many NP-hard problems compared to other evolutionary algorithms.

There are previous works that involved other algorithms such as Genetic Algorithm (GA) and Simulated Annealing (SA) on parameter estimation. According to Kikuchi et al. [6], using conventional simple genetic algorithm (SGA), inferring parameter values of small network but with very limited number of parameters, and the convergence rate is very low. There are two problems of SGA which are early convergence in the fast stage of search and evolutionary stagnation in the last stage.

For the SA, it is physically inspired method. The global and local search of the SA depends on the temperature. Temperature will decrease automatically, switching from global to local search. According to Gonzales et al. [7], he adapted the SA from the S-system parameter estimation from time series data.

This paper focuses on the parameter estimation in lactococcus lactis C7 by estimating the parameter values of pyruvate metabolism by using Ant Colony Optimization (ACO) algorithm. This is because there are no researches that have been conducted on parameter estimation using ACO. The production that will be focused on is Acetoin production. There are also other substrates that can influence the production of acetoin such as acetolactate (aclac).

2 Materials and Methods

In this section, we describe the details of the proposed ACO into SBToolbox in MATLAB to estimate the parameter value. The dataset used is lactococcus lactis C7. The parameter value for acetoin production is the main target to be estimated. From the previous work, the algorithms that are commonly used for parameter estimation are Genetic Algorithm (GA), Simulated Annealing (SA) and Simplex. However, for this dataset, there are no experiments ever conducted for the implementation of ACO in parameter estimation. Thus, we propose ACO to estimate the parameter in this research.

2.1 Datasets

This research uses lactococcus lactis C7 as its dataset. The dataset was obtained from the BioModel database which was in SBML format (http://biomodels.caltech.edu/BIOMD0000000017). In order to make it used in MATLAB, the SBML file (.xml) of the dataset needed to be converted into a format that could be read. This dataset was obtained from a study conducted by Marcel et al [8] to understand the regulations of pyruvate metabolism using a model based on measured kinetic parameters. This model was simulated by estimating the kinetic parameter values. The initial values of the kinetic parameters for experimental and simulated algorithms is shown in Table 1 for acetoin. There were 12 kinetics parameter to be estimated for acetoin.

Table 1. The Initial Value for experimental and simulated algorithm based on kinetic parameters of acetoin production

Kinetic parameters	Experi-mental Value	Simulated Value		
		ACO	SA	Simplex
R9_V_9	106	152.1	201.0231	152.1
R9_Kaclac_9	10	11.3	14.8018	11.3
R9_Kacet_9	100	80.4	122.5118	80.4
R10_V_10	200	495	483.9463	495
R10_Kacet_10	5	0.5	5.4955	0.5
R11_V_11	105	87.1	78.9765	87.1
R11_Keq_11	1400	1113.5	854.3025	1113.5
R11_Kacet_11	0.06	0.1	0.00852	0.1
R11_Knad_11	0.02	0	0.305	0
R11_Kbut_11	2.6	2.4	1.8348	2.4
R11_Knadh_11	0.16	0.2	0.1539	0.2
R14_k_14	0.0003	0	0.0004	0

Figure 1 shows the reactions included in the model to describe the distribution of carbon from pyruvate in L. lactis. Numbers in circles indicate the following enzymes or steps: 1,'lumped' glycolysis; 2, LDH; 3, pyruvate dehydrogenase; 4, phosphotransacetylase; 5, acetate kinase; 6, acetaldehyde dehydrogenase; 7, alcohol dehydrogenase; 8, ALS; 9, acetolactate decarboxylase; 10, acetoin efflux; 11, acetoin dehydrogenase; 12, ATPase; 13, NOX; 14, non-enzymic acetolactate decarboxylation; 15, pyruvate formate lyase, which is considered not to be active under aerobic conditions; 16, chemical conversion to diacetyl, not included in the model. Substrates and products that were clamped in the model are indicated in italics.

2.2 Ant Colony Optimization (ACO)

This algorithm is inspired by the behavior of ants during the colony searching for the shortest path. The pheromones deposited by ants attract other ants which then will increase the pheromones. ACO is a probabilistic technique which solves the computational problems that have ability to reduce finding good path through nodes in graph. According to Zuniga *et al.* [9], he adapted ACO for S-system models by other nodes that were connected to it. They called the algorithm as discrete ACO. Meanwhile, the continuous ACO enhanced the aggregation pheromone system (eAPS) for parameter

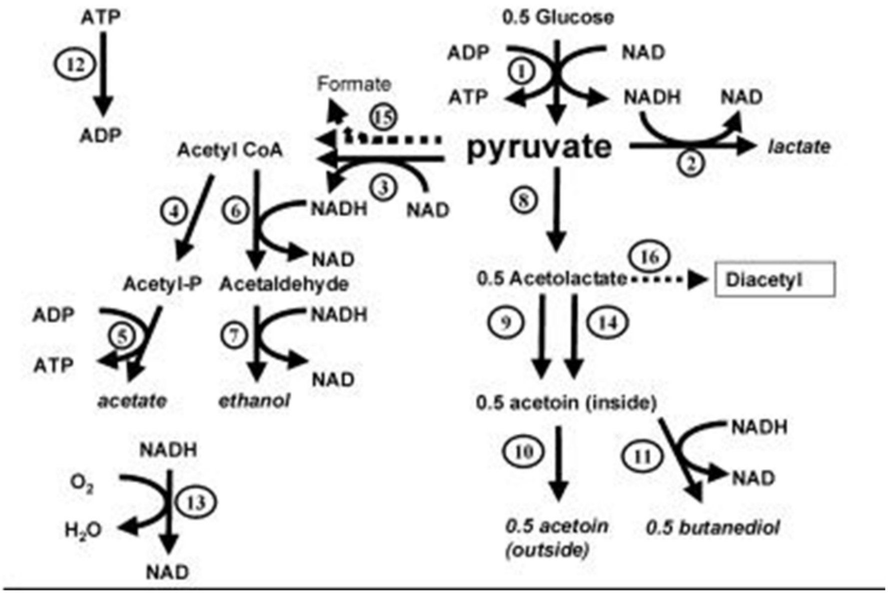

Fig. 1. Pyruvate metabolisms in *lactococcus lactis C7* (Marcel *et. al.,* 2002)

task involving S-system. Figure 2 shows the foraging of ants, Figure 3 shows the pseudocode of Ant Colony Optimization (ACO), and Figure 3 shows variable used in ACO's pseudocode.

2.3 System Biology ToolBox (SBToolBox)

MATLAB is a programming environment for algorithm development, data analysis, visualization, and numerical computation. MATLAB can solve technical computing problems faster than by using traditional programming languages, such as C, C++, and FORTRAN. The toolbox used is the Systems Biology Toolbox for MATLAB which offers systems biologists open and user extensible environment, in which to explore ideas, prototype, share new algorithms, and build applications for the analysis and simulation of biological systems. It features a wide range of functions. The Ant Colony Optimization (ACO) will be integrated into the SBToolBox(MATLAB) to obtain the best value of kinetic parameters from the experimental data.

2.4 COPASI

COPASI is one of the softwares used for simulation and modeling. Simulation and modeling have become a standard approach to understand complex biochemical processes. The parameter values gathered from the implementation of the algorithm in MATLAB will be used in this software to get the data needed for the evaluation of the metabolites; which is acetoin production that needs to be analyzed. In this research, COPASI was used in the process of evaluation of the algorithms performances.

```
Init pheromone $\tau_{ij}$ ;
repeat for all ants i: construct solution(i);
        for all ants i: global pheromone update(i);
        for all ants edges: evaporate pheromone;
                                        $(\tau_{i-j} := (1-\rho)\ \tau_{i-j})$

construct_solution(i):
init ant;
while not yet a solution:
      expand the solution by one edge probabilistically according
      to the pheromone;
                                        $(\tau_{\rho i-j} / \Sigma_{\rho i-j'} \tau_{\rho i-j'};)$

global_pheromone_update(i):
for all edges in the solution:
      increase the pheromone according to the quality;
              $(\Delta\tau_{j-j'} :=$ 1/length of the path stored)
```

Fig. 2. Pseudocode of Ant Colony Optimization (ACO)

$(V=\{0,\ldots,N\},\ E=\{i \rightarrow j\})$ = directed acyclic graph

N = food source

τ_{ij} = initial pheromone

ρ = pheromone deposited

Fig. 3. Variables used in ACO's pseudocode

2.5 Performance Measurement

For the evaluation, four performance measurements were used to evaluate the performance of Ant Colony Optimization (ACO) algorithm. The evaluation was conducted based on the time series data that had been generated. The time series data contain measurement result, y, and simulated results yi for the algorithm. The first performance measurement used was the average of the error rate of the algorithm. The average of error rate is calculated by using the following equations:

$$e = \sum_{i=1}^{N} (y - yi) \tag{1}$$

$$A = \frac{e}{N} \tag{2}$$

The algorithm has a better performance if the average of error rate is lower. The second performance measurement used was by calculating the standard deviation for 50 runs. The performance of the measurement was calculated using the equation below. If the values of the standard deviations are closer to zero, the performance of the algorithm is more accurate.

$$STD = \sqrt{\frac{e}{N}} \tag{3}$$

Production graph were also constructed. The production graph would be the third performance measurement to be used. The performances would be evaluated by comparing the simulated line and the experimental line. The closer the simulated line to the experimental line, the performances of the algorithm is more accurate and better. Last but not least, the performance measurement used was computational time. The shortest time taken for a run would be the best algorithm for the computational time.

3 Results

This section discusses on the performance of the algorithms by measuring and evaluating the standard deviation, average error rate, time and the accuracy of the algorithm. To clearly see the performance of the algorithms, a comparison between algorithms had been made, comparing the performance of Simulated Annealing (SA) and ACO. This research focuses on the acetoin production in lactococcus lactis C7. Acetoin is very important in the fermentation and production of food products. Acetoin is used as a food flavoring (in baked goods) and as a fragrance. It is also important in giving butter its characteristic taste. There are other substrates/metabolites that influence the production of acetoin such as acetolactate (aclac). For this paper, the metabolite that we focused on to estimate the parameter values is acetolactate (aclac) and acetoin production itself.

3.1 Acetoin Production

Figure 4 shows the graph production of acetoin in the pyruvate metabolism. The output graph shows four lines which represent experimental, simulated ACO, simulated simplex and simulated SA.

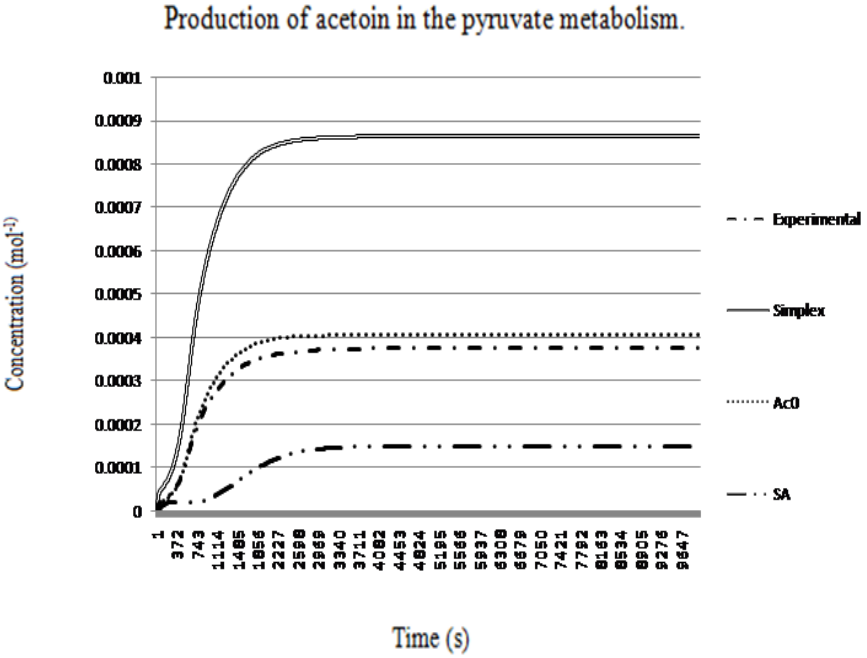

Fig. 4. Production of acetoin in pyruvate metabolism

Table 2 shows the average error rate of ACO and Simplex. The average error rate of ACO is 0.1033E-08, SA is 0.31598E-08 and the average error rate for Simplex is 2.1482E-08. For the standard deviation, the result of standard deviation using ACO is 0.7743E-08, SA is 4.3543E-08 and standard deviation for Simplex is 111.9990E-08. For measurement in term of time, the ACO computational time is 239.8957, SA is 240.7595 and Simplex is 1460.0173.

Table 2. 2 Comparison of average error rate and standard deviation using ACO, SA and Simplex for the production of acetoin in pyruvate metabolism

Algorithms / Measurements	ACO	SA	Simplex
Average Error Rate	0.1033E-08	0.31598 E-08	2.1482E-08
Standard Deviation	0.7743E-08	4.3543E-08	111.9990E-08
CPU Time (seconds)	239.8957	240.7595	1460.0173

Note: Shaded column represents the best results.

4 Discussion

Figure 4 shows the graph production of acetoin in the pyruvate metabolism. The output graph shows four lines which represent experimental, simulated ACO, simulated simplex and simulated SA. The simulated line of ACO is closer to the experimental line if compared to simulated SA and simulated Simplex. This shows that the performance of ACO in parameter estimation is better than SA and Simplex.

Table 2 shows the average error rate of ACO and Simplex. The average error rate of ACO is 0.1033E-08, SA is 0.31598E-08 and the average error rate for Simplex is 2.1482E-08. From the average error rate, it shows that ACO has the smallest values of average error rate if compared to Simplex and SA. A low average error rate means the performance of the algorithm is good because the errors involved are few.

For the standard deviation, the algorithm is good in performance and consistent if it has a standard deviation that is close to zero. In Table 2, the result of standard deviation using ACO is 0.7743E-08, SA is 4.3543E-08 and standard deviation for Simplex is 111.9990E-08. The value of standard deviation of ACO is the closest to zero if compared to the standard deviation of Simplex and SA. This shows that performance of ACO is good, consistent and better than others. For measurement in term of time, the ACO computational time is 239.8957, SA is 240.7595 and Simplex is 1460.0173. ACO shows a better result than Simplex and SA by having a slight difference with SA and extreme difference with Simplex.

From all the results that have been discussed, ACO has the best performances. This is contributed by the advantages of ACO in finding good solution, enabling it to perform better for parameter estimation. This is because it builds a solution using local solutions, by keeping good solutions in memory. A completed tour is analyzed for the optimality by calculating the strength of the trails of pheromone. Each ant leaves a trail of pheromones when it explores the solution landscape. This trail is meant to guide other ants. More pheromone on trail increases the probability of trail being followed and it will be repeated until most ants select the same tour on every run/cycle (convergence to solution). ACO updates the pheromone only for the (local or global) best ants. Besides that, the collective interaction through indirect communication, called stigmergy can also be influential to lead to good solutions.

5 Conclusion and Future Work

As conclusion, the performance of ACO in parameter estimation is quite good and better than Simplex and SA after being implemented into SBToolbox in MATLAB. The comparison graph shows that the simulated line of ACO is the closest to experimental line compared to other algorithms. Besides that, the result generated by ACO is more consistent because the standard deviation value is the closest to zero compared to Simplex and SA and the computational time is also the shortest compared to others. The attempt to develop algorithms inspired by observing ant behavior has the ability to find good paths based on probabilistic strengths of pheromones and has become the field of ant colony optimization (ACO). Thus, this makes ACO able to perform better compared to other algorithms in term of computational time and accuracy. Therefore, it proves that ACO can solve problems in a cheaper way without using wet lab experiments and with short computational time. Besides that, ACO, which is a global optimization method, is appropriate to be used for the parameter estimation from experimental results because of the performances shown. For future work, we can increase the number of program run in MATLAB to achieve more accurate results for the performance of the algorithm.

Acknowledgments. We would like to thank Malaysian Ministry of Science, Technology and Innovation for supporting this research by an e-science research grant (Grant number: 06-01-06-SF1029). This research is also supported by UTM GUP research grant that was sponsored by Universiti Teknologi Malaysia.

Competing Interests

The authors declare that they have no competing interests.

References

1. Maksat, A., Yves, F.N., Jaap, A.K., Joke, G.B.: Systems biology: parameter estimation for biochemical models. FEBS Journal 276, 886–902 (2009)
2. Lei, M.U.: Parameter Estimation Methods for Biological Systems. Master Thesis, The University of Saskatchewan, Canada (2010)
3. Liberti, L., Kucherenko, S.: Comparison of deterministic and stochastic approaches to global optimisation. International Transactions in Operational Research 12(3), 263–285 (2005)

4. Kahng, D.S., Lee, D.: Clustering Parameter Values for Differential Equation Models of Biological Pathway. In: Proceedings of the Second International Symposium on Optimization and Systems Biology (OSB 2008), Lijiang, China, October 31-November 3, pp. 265–270 (2008)
5. Alonso, S., Cordón, O., Viana, I.F.D., Herrera, F.: Integrating Evolutionary Computation Components in Ant Colony Optimization Evolutionary Algorithms: An Experimental Study, pp. 148–180 (2004)
6. Kikuchi, M., Hajime, K., Kobayashi, S.: Moving target detection from infrared images using genetic algorithms. Systems and Computers in Japan 34(7), 76–86 (2003)
7. Gonzalez, O.R., Kuper, C., Jung, K., Naval, P.C.J., Mendoza, E.: Parameter estimation using simulated annealing for S-system models of biochemical networks. Bioinformatics 23(4), 480–486 (2007)
8. Marcel, H.N.H., Marjo, J.C.S., Dirk, E.M., Jeroen, H., Michiel, K., Iris, I.V.S., Roger, B., Hans, V.W., Jacky, L.S.: Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology 148, 1003–1013 (2002)
9. Zuñiga, P.C., Pasia, J., Adorna, H., del Rosario, R.C.H., Naval, P.: An ant colony optimization algorithm for parameter estimation and network inference problems in S-system models. In: Proceedings of International Conference on Molecular Systems Biology 2008 (ICMSB 2008), Manila, Philippines, pp. 105–106 (2008)