Evolutionary algorithm for metabolic pathways synthesis

M.F. Gerard[1], G. Stegmayer[1] and D.H. Milone[1]

sinc(i), FICH–UNL/CONICET,
Ciudad Universitaria UNL, 4to piso FICH, (S3000) Santa Fe, Argentina.
mgerard@sinc.unl.edu.ar

Several methods to automatically search for metabolic pathways among compounds have been developed in last years. They are mainly based on classical search algorithms, such as breadth-first and depth-first search, and the A* algorithm. They first model data as an appropiate graph, and then perform the search of a pathway between compounds. Although they have served as a first attempt to address the problem, these methods have at least one of the following limitations: i) they do not consider all the substrates for the reactions (feasibility of reactions is not guaranteed); ii) they cannot model and build branched metabolic pathways; iii) they only search pathways between two compounds; iv) previously synthesized compounds in the reactions chain are not taken into account to select a new reaction (just the last one). Frequently, it leads to find solutions without biological sense.

In [2] we propose a new algorithm based on evolutionary computation for searching branched metabolic pathways, that links simultaneously several compounds through feasible reactions. This proposal, named evolutionary metabolic synthesizer (EvoMS), uses the concept of expanded set of compounds (ESC) to model metabolic pathways. Given a set of available compounds and a feasible reaction from them, the ESC is built by adding the products of the reaction to the current set of compounds. In this way, it is possible to expand the set of available compounds in order to allow a higher number of reactions to take place from the new set of compounds. Thus, given a set of compounds to relate and the initial conditions for the search, the goal is to find a metabolic pathway that uses one of those compounds (initial substrate) for synthesizing the other ones (final products).

Fig. 1 shows an outline of the proposed algorithm. As it can be seen, EvoMS uses linear structure of chromosomes to encode a sequence of reactions that model a metabolic pathway, where each gene is a reaction. Starting from a set of candidate networks, metabolic pathways are combined together to produce new potential solutions. The searching process is guided by an objetive function, which takes values in $[0, 1]$ range, that assesses four aspects of metabolic pathways: feasibility (by mean of ESC and the initial set of available compounds), the proportion of specified compounds which are in the network, proportion of nonrepeated reactions, and proportion of final products linked to the initial substrate. In order to introduce random variations into the networks, a novel mutation operator that is able to insert or erase genes from a chromosome is proposed.

sinc(i) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
M. Gerard, G. Stegmayer & D. H. Milone; "Evolutionary algorithm for metabolic pathways synthesis"
XVII Simposio Argentino de Inteligencia Artificial - 45 Jornadas Argentinas de Informática, pp. 136-138, 2016.

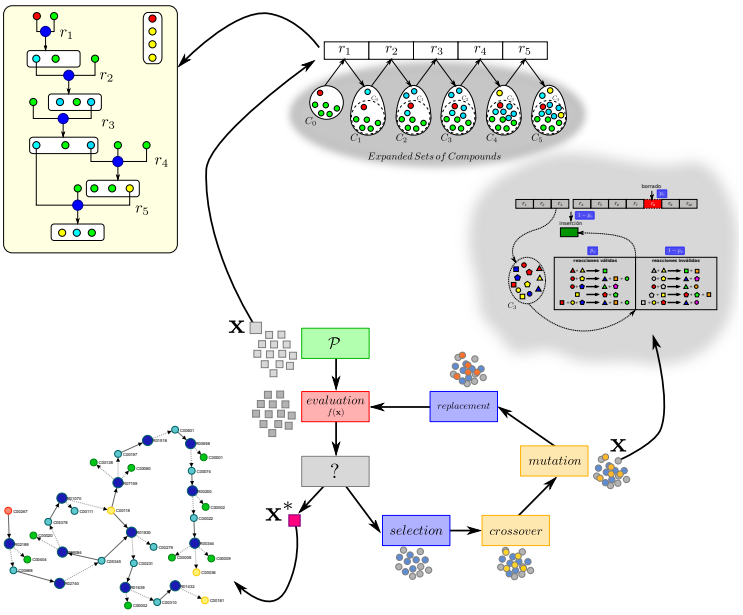

Fig. 1. Outline of the proposed algorithm. Initial substrate is in red, final products, are in yellow and available compounds are in green. Products of reactions which are not in those sets are in light blue. Reactions are in dark blue. In the upper left corner there is an example for an initial metabolic network, which only produces two of the three specified final products (yellow circles in the pathway). Each chromosome encodes a metabolic pathway, where every gene is a reaction. In the bottom left corner there is an example for a solution that relates the three final products specified.

Main feature of EvoMS is the ability of finding networks that simultaneously relate several compounds. This is really important because most of metabolic pathways in nature are rather complex networks of interacting reactions among several compounds. To measure the network branching of a metabolic pathway it must be taken into account that a compound may lead to (that is, be substrate of) one or more reactions. Thus, the number of reactions that can be performed from each compound provides a simple way to calculate it. Then, the branching factor was calculated here as the ratio between the sum of the number of reactions that employ each substrates and the total number of substrates (free available compounds must not be taken into account). Accordingly, a branched pathway should have a ratio higher than 1.0.

Table 1 shows results of 20 runs[1] for EvoMS[2] versus methods based on classical search algorithms such as breadth-first search (BFS) and depth-first search

[1] Runs were performed on a single computer with an Intel i7 CPU and 8 parallel threads.

[2] Each generation takes about 140 s.

sinc(i) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
M. Gerard, G. Stegmayer & D. H. Milone: "Evolutionary algorithm for metabolic pathways synthesis"
XVII Simposio Argentino de Inteligencia Artificial - 45 Jornadas Argentinas de Informática, pp. 136-138, 2016.

Table 1. Comparisons of four algorithms for metabolic pathways searching. Search were performed for pathways between *histidine* and *serine*.

	Generations		Pathway size		Branching	
	med	max	med	max	ave	max
BFS	–	–	5	5	1.00	1.00
DFS	–	–	100	100	1.00	1.00
EAMP	23	305	9	19	1.00	1.00
EvoMS	23	518	6	9	1.06	1.22

(DFS), and also with an evolutionary algorithm for searching linear metabolic pathways (EAMP)[3] [1]. For a fair comparison, a simple linear case was used, relating two compounds. As it could be expected, BFS found the shortest paths (5 reactions) while DFS, the longest ones (100 reactions, the maximum allowed in these runs). In both cases, only linear pathways were found, as reflected by the 1.00 value in the branching factor. EAMP found linear pathways with more reactions. Regarding EvoMS, it could also find pathways with a variable number of reactions, offering alternative mechanisms for relating compounds. It should be noted that the minimum number of reactions to relate several compounds is not known in advance. Intuitively, it could be expected that pathways requiring a few reactions to link compounds of interest would be more specific than those containing a lot of them. However, larger solutions can provide more information to understand the biological process, and therefore be more interesting from the application point of view. EvoMS achieved an average 1.06 branching because this case could be solved with a linear pathway. In spite of this simplification, it should be noted that EvoMS was able to find a pathway with branching factor of 1.22. Clearly, EvoMS provides a simple method for synthesizing metabolic pathways, where solutions are networks of feasible reactions, linking specified compounds through branched connections.

The full source code for EvoMS algorithm is available for free academic use at http://sourceforge.net/projects/sourcesinc/files/evoms/. A web-interface to run EvoMS is available online at http://fich.unl.edu.ar/sinc/web-demo/evoms/. Main inputs, outputs and features are explained in [3].

References

1. Gerard, M., Stegmayer, G., Milone, D.: An evolutionary approach for searching metabolic pathways. Computers in Biology and Medicine 43, 1704–1712 (2013)
2. Gerard, M.F., Stegmayer, G., Milone, D.H.: Evolutionary algorithm for metabolic pathways synthesis. Biosystems 144, 55–67 (2016)
3. Gerard, M., Stegmayer, G., Milone, D.: EvoMS: an evolutionary tool to find de novo metabolic pathways. BioSystems 134, 43–47 (2015)

[3] This previous work provides further comparisons between three methods for linear pathways.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
M. Gerard, G. Stegmayer & D. H. Milone: "Evolutionary algorithm for metabolic pathways synthesis"
XVII Simposio Argentino de Inteligencia Artificial - 45 Jornadas Argentinas de Informática, pp. 136-138, 2016.