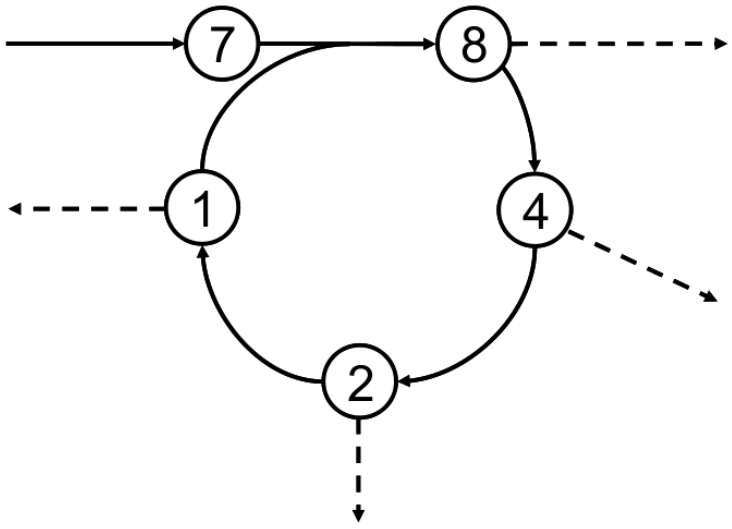
Emergence of realistic metabolic networks from artificial chemistry

Maksim Sipos

Term paper for Fall 2009 Emergent States of Matter

December 14, 2009



Abstract

Complex metabolic networks characterize the life of every cell. In this paper we review how such metabolic networks can emerge from a model that applies optimization pressure to artificial chemistries. This model can lead to metabolic networks that share many of the features of real networks, such as cycles and hierarchies.

Notice: Cover figure reproduced from [1].

Metabolism is a collection of chemical reactions that imbue cell matter with life. All life forms use chemical reactions to grow, consume resources and procreate. For example, metabolic reactions allow photosynthetic organisms to consume water, energy and carbon dioxide to produce organic compounds. Similarly, metabolic reactions allow organisms to store energy in organic molecules and release that energy when it's needed. Metabolism, interplaying with genetics and molecular cellular machinery, is a big part of the complexity of the living cell. Understanding of metabolism leads to a better understanding of the cell, and furthermore its evolution.

Some aspects of metabolism are interesting from a physics perspective. Metabolic reactions in cells connect reactants in a variety of complex ways. They can form pathways, cycles and complex networks. For instance, metabolic networks are some of the most famous examples of scale-free networks in nature [2] [3]. Scale-free networks are networks in which the distribution of node degrees is a power law (node degree is the number of node's neighbors). These networks are associated with having "hubs", that is nodes in the network that have high deegrees. These hubs make the network a "small world", in the sense that the distance between any two nodes in the network is much smaller than that in a random network of the same number of nodes and edges.

Evidently, there is a lot of complexity in the metabolic reaction networks. It is observed in nature that while different species have differing metabolic networks, some core metabolic pathways and cycles are practically universal across all life [4]. Since there is many alternative ways to combine reactants and products in metabolic reactions, a question arises: What chooses one particular metabolic pathway over its alternatives?

There is evidence that metabolic reactions observed in nature are optimal. In 1981 Baldwin and Krebs reported that the efficiency of the ubiquitous citric acid cycle is more than twice that of feasible alternatives [5]. This serves as evidence that metabolic reactions may undergo evolution towards becoming more efficient. Efficiency here refers to the amount of product of a reaction. In this paper, we will investigate how such evolutionary optimization pressure can lead to emergence of metabolic pathways that seem to share many of the features of pathways found in nature.

We should note that in this paper we are interested in looking at opti-

mality of metabolic networks from the perspective of simplified models of chemical reactions. Though many metabolic reactions and cycles are universal, metabolic rates depend on various factors, for example on organism size, in a way that we will not discuss in this paper (see [6] for a discussion on another very interesting subject to physicists, the allometric scaling in nature).

Optimization pressure has been used on detailed kinetics models of ATP and NADH producing cycles[7]. It successfully recovered some stoichiometric features of the citric acid cycle and glycolysis. Another chemical optimization method can be executed on many different alternative pathways and can sometimes predict the experimentally determined pathway [8].

However, in this paper we decide to focus on a coarse grained model of an artificial chemistry due to Riehl *et al* [1]. This model's results include interesting topological patterns that are similar to those observed in experimental data. Emergence of such patterns from a simple model is interesting, especially since this model sheds most of the detail of chemical models.

Arithmetic simplicity of metabolic networks

In this model, a simple artificial chemistry [9] of chain-like compounds is considered. All monomers in the chain are assumed to be the same and the longest possible chain is of length N. All reactions in this system form a network R_N. An example of such a network is given in Figure 1. The reactions are modeled to be of the form

$$a_i + a_j \leftrightarrow a_k \tag{1}$$

where $i + j = k$. This reaction (going left to right) models connecting two chains of length i and j into a chain of length k. The inverse reaction (going right to left) corresponds to breaking a long chain into two parts.

The next step in this model is to construct an optimal path from an input reactant a_i with flux ν_{in} to product a_j with that $\nu_{out} = \nu_{in} j/i$. The pathway is essentially considered optimal if it has the fewest steps (intermediate reactions). Pathways are required to be in steady state and no other products are allowed other than a_j (hence the j/i term in the output flux).

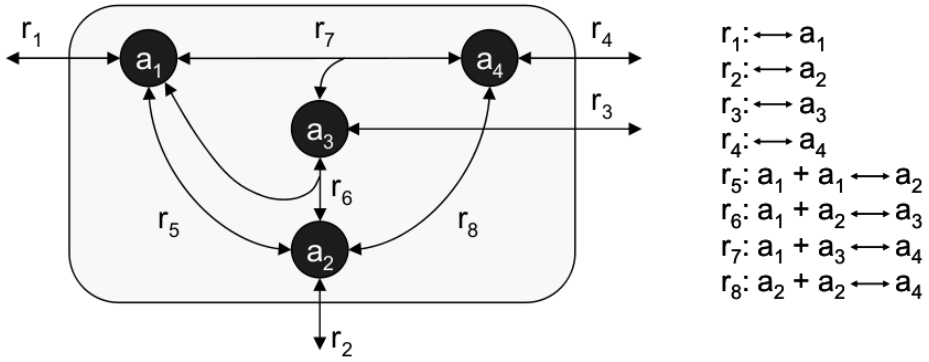


Figure 1: Example of an R_N network of reactions for $N = 4$. *Left:* Compounds are nodes in the network, and the reactions are edges that connect reactant and product compounds together. *Right:* List of reactions corresponding to R_4. This figure was adapted from [1].

We briefly review the general method used to find such optimal paths. Define the $n \times m$ matrix S where n is the number of different types of compounds (i.e. N) and m is the number of reactions. The matrix element S_{ij} then is equal to the number of molecules of type i that react in reaction j. S_{ij} has positive sign if i is a product, or negative if it is a reactant. If the flux of the reaction is denoted by ν_j, then a steady state is defined by having

$$\sum_{j=1}^{n} S_{ij}\nu_j = 0 \tag{2}$$

Then one constrains that $\nu_{out} = \nu_{in} j/i$ and looks for a solution that minimizes the number of non-zero fluxes ν_j. This can be written as a linear optimization problem and can be solved using a linear programming technique. Riehl *et al* also identify two other heuristic methods that can help with discovering optimal paths for large reaction networks.

The optimal paths found fall into three different categories. First there are paths that are sequences of connection reactions. It is intuitive to see,

for example, that the optimal reaction $a_1 \rightarrow a_8$ will proceed by

$$a_1 + a_1 \rightarrow a_2, \quad a_2 + a_2 \rightarrow a_4, \quad a_4 + a_4 \rightarrow a_8. \tag{3}$$

Then there are "addition-subtraction" pathways, in which there are both connection (addition) and disconnection (subtraction) reactions. Examples of such paths are given in Figure 2.

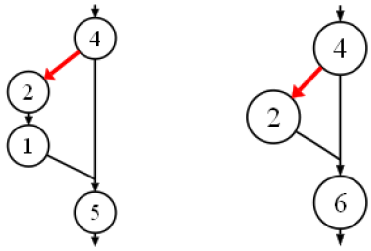


Figure 2: Examples of pathways that include both breaking and connecting reactions. This figure was adapted from [1].

Most interesting is the emergence of autocatalytic cycles for some i and j. Autocatalytic cycles are those in which the existence (flux) of the product encourages (catalyzes) a faster production. The example of a autocatalytic cycle is given in Figure 3. Another example is the cycle displayed on the cover page of this paper which is found to be a module that exists in reactions $a_7 \rightarrow a_i$ for $i \in \{1, 2, 4, 8, \ldots\}$.

Interestingly, Riehl *et al* find a relationship between the length of optimum paths and a problem in cryptography. This can be used to derive an analytical estimate of the length of optimal paths $a_i \rightarrow a_j$

$$L(i, j) \approx \log_2 \frac{i}{\gcd(i, j)} + \log_2 \frac{j}{\gcd(i, j)} \tag{4}$$

To compare the patterns observed above to experimental data, Riehl *et al* map their artificial chemistry onto the chemistry of compounds with a carbon backbone. In that mapping, a cleavage of a molecule with 6 carbons into two molecules with three carbons is mapped to $a_6 \rightarrow a_3 + a_3$. They proceed to

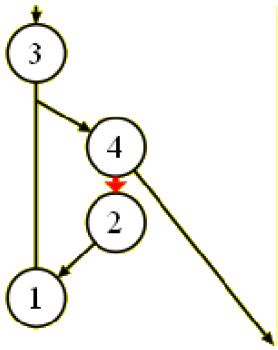


Figure 3: An example of a autocatalytic cycle $a_3 \rightarrow a_4$. Once some amount of product a_4 is created, it pays off to break some of it up into a_1 that will directly combine with a_3 to create more product a_4. This figure was adapted from [1].

compare the results of their model with average reactions observed in nature. In particular, they compare the lengths of metabolic pathways in their model, in the experiment and that given by Equation 4. As you can see in Figure 4, they find a pretty good correspondence.

The logarithmic structure evident in Equation 3 allows reactions $a_i \rightarrow a_j$ for large differences $j - i$ to proceed in few steps. Because of this, such reactions should be more common overall. The authors predict that a theoretical distribution of reactions utilizations should be a power law of slope -1. Their simulation gives them a power law of slope -1.1 whereas experimental data produces a power law of slope -0.89.

Discussion

We have analyzed a model of metabolic pathway patterns which is pretty successful, despite its apparent simplicity. However, the experimental evidence could be better if one could quantify the realism of the model reaction network topologies, beyond just seeing patterns like logarithmic cascades, cycles and modules. For this purpose there are many network topology estimators.

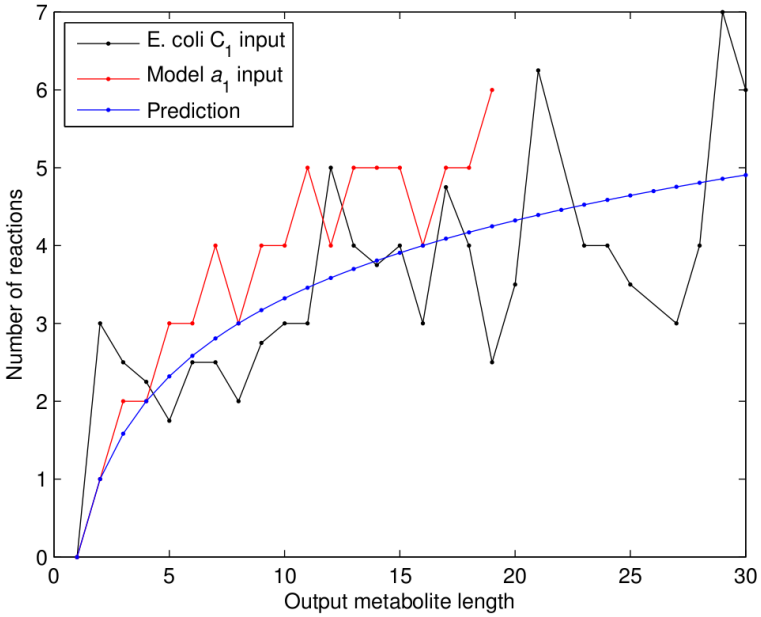


Figure 4: Comparison of model and experimental reactions $a_1 \rightarrow a_x$ (x is on the horizontal axis). This figure was adapted from [1].

Another of the weaknesses of the current models based on optimization is that they always look at metabolic pathways "locally", that is they optimize each pathway separately and then look at resultant metabolic pathway network. It would be interesting to see if metabolic networks are influenced by "global" effects of separate pathways interacting with each other. In other words, what would a minimal "holistic" model of a metabolic reaction network look like?

Ultimately, a good understanding of optimization pressures that drive metabolic pathway formation is an important goal. It can lead to a better understanding of evolution and the origin of life. Furthermore, an understanding of how to "control" metabolic pathways has practical uses: it can lead to progress in biology and medicine.

References

[1] William J. Riehl, Paul L. Krapivsky, Sidney Redner, and Daniel Segrè. Arithmetic simplicity beneath metabolic network architecture. Arxiv 0910.4943v1, 2009.

[2] Reiko Tanaka. Scale-rich metabolic networks. *Physical Review Letters*, 94(16):168101, 2005.

[3] D A Fell and A Wagner. The small world of metabolism. *Nat. Biotechnol.*, 18(11):1121–2, 2000.

[4] Eric Smith and Harold J Morowitz. Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. U.S.A.*, 101(36):13168–73, 2004.

[5] Jack E. Baldwin and Hans Krebs. The evolution of metabolic cycles. *Nature*, 291:381–382, 1981.

[6] Geoffrey B. West and James H. Brown. Life's universal scaling laws. *Physics Today*, September:36–42, 2004.

[7] O Ebenhöh and R Heinrich. Evolutionary optimization of metabolic pathways. theoretical reconstruction of the stoichiometry of atp and nadh producing systems. *Bull. Math. Biol.*, 63(1):21–55, 2001.

[8] John E Beasley and Francisco J Planes. Recovering metabolic pathways via optimization. *Bioinformatics*, 23(1):92–8, 2007.

[9] P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries-a review. *Artificial Life*, 7(3):225–276, 2001.