An Ansatz for a Theory of Living Systems

Dominique Chu
School of Computing
The University of Kent
Canterbury, Kent, CT2 7NF, UK
Email: d.f.chu@kent.ac.uk

David J. Barnes
School of Computing
The University of Kent
Canterbury, Kent, CT2 7NF, UK
Email: d.j.barnes@kent.ac.uk

***Abstract*—Mathematical and computational modelling is now firmly established as an important tool in life sciences. However, it is not yet established as a source of biological knowledge on an equal footing with experimentation. Here we argue that there are now substantial opportunities for a new theoretical biology that not only benefits from experimental research but also influences this research. We argue that such a new theoretical biology would need to find the appropriate level of abstraction and combine aspects of Artificial Life and theoretical biophysics. We also provide an outline of a research agenda that could lead to experimentally testable insights into biological systems.**

I. Introduction

Mathematical and computational modelling is now firmly established as an important tool in the life sciences. Bioinformatics, for example, is an almost exclusively computational endeavour, but at the same time indispensable for modern life-science research. Similarly, systems biology relies heavily on computational support for data storage and analysis and model inference. Moreover, even for conventional experimental biology it is now becoming ever more common to accompany experimental data with a theoretical model to explain the results.

Despite this advance of formal methods in biology, biological knowledge generation is still heavily dominated by experimental methods. Computational and mathematical modelling is more often than not seen as an add-on or as a service rather than as a fundamental part of knowledge generation. Theoretical biology does not have the same relationship to and importance for experimental biology as theoretical physics does for experimental physics. A symptom of this is that the "Journal of Theoretical Biology" is rarely read or referred to by experimentalists.

One reason for this is clearly of historical nature. Until very recently, biology has been an almost exclusively experimental science. Consequently, biologists tend not to have a strong background in mathematics, unlike, for example, experimental physicists who do understand theoretical results. It is therefore difficult in biology to establish a true dialogue between experimentalists and theoreticians. Moreover, predictions arising from theoretical biology are not as good as they are in theoretical physics; often it is not even clear how to test them experimentally. There are also many important areas of biology that are perhaps forever beyond theoretical prediction. For example, it will likely never be possible to predict the DNA sequence of *E. coli* from first principles. Understanding the diversity of life requires experimental methods.

Many details of biological systems are accidental and not amenable to an explanation from a theory, yet there are also many general features of biological systems that reflect very specific optimisation strategies and, as such, can be explained theoretically. Within evolutionary biology and game theory there is a rich mathematical tradition explaining behavioural strategies [1]. Beyond that, in the context of the structure of the living system, there probably are overall design principles that unify the diversity of life. These principles will not be sufficient to predict details of living systems, but they can provide an evolutionary rationale for observed specifics. In turn this rationale can be used to separate the accidental from the fundamental.

There have been a number of attempts to uncover deep insights into the design principles of biological systems based on first principles. Some approaches rely on very abstract models of biological organisation. Take, as an example, Robert Rosen's (M,R) systems which are a highly idealised organisational model of living systems [2]. Another related approach is autopoiesis [3], [4]. The core idea of both of these theoretical attempts was that life is fundamentally about organisation rather than about a particular material (i.e., carbon-based biochemistry). Inspired by the same idea, there were then various attempts to capture essential organisational aspects of life using artificial chemistries [5]–[9].

Another research direction that was promising as a basis for a theoretical biology is complexity theory, especially the idea of the universality classes [10], [11]. Applied to biology this led to the discovery of scaling laws in biological systems [12]. More recently, network biology provided interesting new insights, especially the discovery of over-represented local connection patterns in biological networks such as genetic regulatory networks, termed *network motifs*. It has been possible to assign a function to those motifs [13] and hence to understand some of the underlying principles of biological systems. All the above mentioned research directions were taken note of in various research communities, but remained marginal within empirical biology. This is regrettable because formal science has very much to contribute to empirical biology and the need for theoretical understanding is arguably increasing. Development in biotechnology is accelerating and promises new and more data coming forward. This is often

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.156

taken as an indication that novel methods in data science are needed to interpret this data, which is true. However, more importantly, this creates the possibility of generating hypotheses about design principles of biological systems and to test specific conjectures in data first, without the need to spend much time and effort in the lab on an untested theory.

One promising area is quantitative design principles of biological networks; for example, gene regulatory networks. Regarding the topological properties of these networks significant progress has already been made (c.f., network motifs). However, fairly little research exists to try to understand the numerical values of networks, i.e., the numbers on the arrows in the network diagrams. It is understandable why. These numbers often relate to kinetic constants, reaction rates and the speed of biochemical processes. These quantities are hard to measure and, when they are measured, the measurements have a very large associated error. Modellers tend to regard them as a nuisance and rarely consider them than as a genuine source of general knowledge. They need to be known to model the system, but they have no value in and of themselves.

Taking genetic regulatory networks as an example. Any given network topology has a very large number of possible parameterisations. Many of those will be equivalent in the sense that they would lead to essentially the same behaviour, but they may also encode different strategies. One of the features that genetic networks encode is when and under what conditions a resource is produced by the cell. As such, the regulatory network is encoding the "economic plan" of the cell. If we start to understand the principles of how cells function then we could be in a position to predict kinetic parameters, or at least equivalence classes of kinetic parameters. This, in turn, could then be tested in experiment.

We think that understanding these quantitative properties of networks can provide deep insights into the design principles of biological systems. Kinetic parameters have evolved and, as such, should be assumed to be optimised to a particular set of adaptive pressures. Given this, they contain important information about the ecological role of the organism and its overall life strategy. Moreover, given the accelerating rate of data generation, it is only a matter of time before kinetic rate constants will be available on a large scale and can be analysed.

In this article we will use the example of resource allocation in cells to illustrate the potential value of quantitative network descriptions. Specifically, we consider how, in a series of simple models of cells, resource allocation impacts on growth. In all of these models we assume that there is the choice between either allocating nutrient to maintenance of the uptake/metabolic machinery or directly channelling it into growth/reproduction. We show that, while the numerical specifics differ depending on how one chooses the model, the overall results indicate the same principle in all cases: There is an optimal allocation of resource to growth relative to maintenance. The models presented here are somewhat abstract, although they do represent some of the essential features of real regulatory networks. Nevertheless, the presentation here serves to highlight how such models could be used to generate general predictions that can be tested experimentally in specific organisms.

The models presented here will focus on the trade-off between the need to grow and reproduce and the requirement of the cells to produce the machinery that enables this growth. For a cell it is not useful to have nutrient available when this nutrient cannot be efficiently converted into usable energy and growth. However, an efficient metabolic machinery is not very useful if it entails high maintenance costs or has extensive space requirements. If all nutrient is used for metabolism, the cell will never grow or reproduce. At the same time, if all nutrient is channeled into growth, then none is left to acquire the necessary nutrient, and the cell will die. This suggests that somewhere there is an optimal allocation of resources that maximises growth. Given a particular network, this allocation strategy is determined by the parameterisation of the network.

In biological systems the question of resource allocation appears frequently. A well known example is translation. The decoding of the mRNA in cells is performed by ribosomes. These are molecular machines that are large and very expensive to make. Calculations of the energy requirements in Baker's yeast [14, see SI] have shown that ribosome synthesis takes up a considerable part of the overall energy budget of the cell. Moreover, translation has been shown to be limited by ribosome availability. Presumably the same is true for most other unicellular organisms.

Once one takes into account the cost of the machinery, this then leads to interesting dynamics. If the cell stepped up protein production, say, in order to be able to take up more nutrient from the environment or to metabolise faster, then this would require extra ribosomes, and lead to altogether higher cost. If all energy is used in order to produce non-ribosomal protein, then there would be no machinery to actually produce the proteins. If on the other hand everything is used for ribosomes, then it would be impossible for the cell to take up and metabolise nutrient. This suggests again that there is an optimum allocation of nutrient to ribosomal genes that optimises the speed of protein synthesis. Formally, this trade-off is very similar to the reproduction-metabolism trade-off we explore here.

II. Mathematical model

In order to understand this trade-off in more detail, we devised a minimal mathematical model of growth processes. This model is a simple representation of nutrient uptake in a single-celled organism, such as a bacterium. We consider a process whereby some external nutrient N is taken up using porins P. We denote the internalised nutrient by E. The internalised nutrient is then channeled either towards more porin, with a rate constant of k_3, or towards growth, with a rate constant of k_2. This model can be summarised by the following differential equations

$$\begin{aligned} \dot{E} &= k_1 N P - (k_2 + k_3) E \\ \dot{P} &= k_3 E - k_4 P \end{aligned} \tag{1}$$

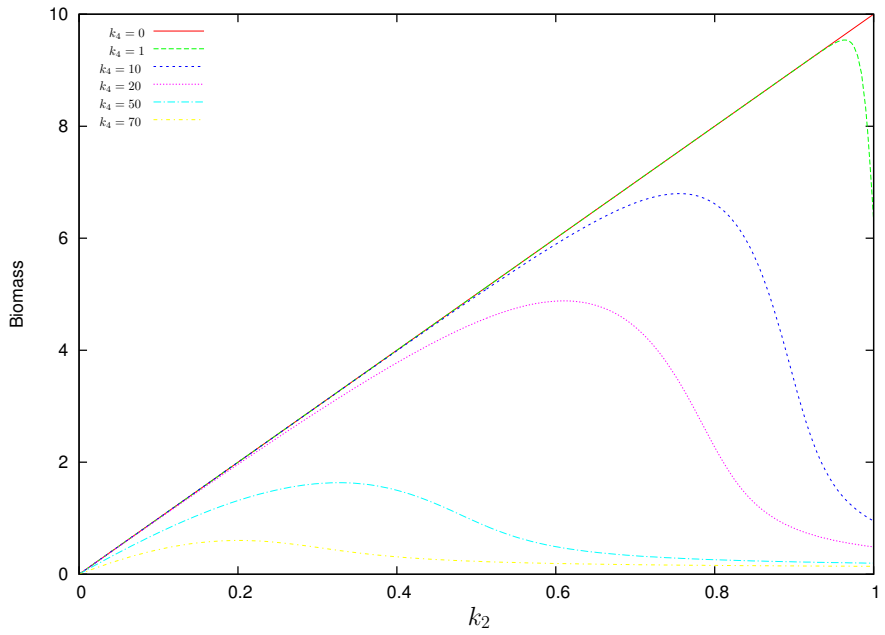

Fig. 1. Simulation of the equation system 6 for different values of the porin breakdown rate k_4. It shows the final value of the biomass (i.e., for $t \to \infty$).

To simplify the analysis without affecting generality, we set $k_1 = k_4 = 1$ and we assume that $k_3 = 1 - k_2$. We can then solve the system of equations analytically for the initial conditions $E(0) = 1, E(0) = 0$ and obtain:

$$\begin{aligned} P(t) &= \frac{1}{2}\left(e^{(a-1)t} + e^{-(a+1)t}\right) \\ E(t) &= \frac{a}{2(k_2-1)}\left(e^{(a-1)t} + e^{-(a+1)t}\right) \end{aligned} \tag{2}$$

here a is given by

$$a = \sqrt{N - Nk_2}$$

The value of a is real as long as $k_2 \geq 1$. $P(t)$ will grow exponentially in time as long as $k_2 < \frac{N-1}{N}$. With this analytical solution in hand, we can then compute the flux of the system toward biomass:

$$F_{\text{bm}} = k_2 \cdot E(t) \tag{3}$$

When written out, the k_2 terms of the flux can be seen to combine to

$$\frac{k_2}{\sqrt{k_2 - 1}}$$

Hence, in the limit $k_2 \to 0$ the flux goes to zero. In the opposite limit of $k_2 \to 1$ the flux approaches:

$$\lim_{k_2 \to 1} F_{\text{bm}} = \frac{nt}{\mathrm{e}^t} \tag{4}$$

This function tends to zero rapidly in time. Hence, in both limits the flux to biomass vanishes. If, on the other hand we set $k_2 = 0.5$, then we obtain

$$E(t)|_{k_2=\frac{1}{2}} = \frac{\sqrt{2n}}{4}\left(\mathrm{e}^{\frac{t}{2}\left(-2+\sqrt{2n}\right)} - \mathrm{e}^{-\frac{t}{2}(2+\sqrt{2n})}\right) \tag{5}$$

For large t the second term in the parenthesis goes to zero, but the first term grows exponentially. Hence, altogether, there is positive flux to biomass for $k = 1/2$ and hence there is a value $0 < k_2^* < 1$ that leads to maximum flux. This means that k_2^* is an optimal parameterisation with respect to biomass production.

III. A numerical example

We can expand the model somewhat to include depletion of the nutrient with a rate of k_0. We then arrive at the differential equation model containing equations 1 plus a depletion equation that represents that external nutrient is used up

$$\dot{N} = k_1 N P \tag{6}$$

This system is not solved as easily as the previous system and we revert to numerical solutions instead.

Numerical solutions of the system (see fig 1) confirm that there is an optimal allocation of nutrient with respect to total biomass production as long as $k_4 > 0$. As the degradation value k_4 increases the maximum becomes shallower and the total amount of biomass that can be produced becomes lower.

IV. Model of nutrient growth

The above two models are very abstract and only outline the behaviour of a real system. The question is now whether or not the same sort of behaviour holds for more complex models. In order to understand this we extended the model further by adding expression of the porin. In order to produce a porin two things need to happen. Firstly, the operon for the porin needs to be activated. Secondly, in order to be able to express a protein, the simulated cell requires energy. This energy must in turn be produced from nutrient that is taken up.

Substrate	Product	Reaction rate
N	E	$k_N \frac{N}{N+K_N} P\, bm$
E	E_0	$k_E E$
$p + E_0$	$p + P$	$\left(\text{leak} + k_P \frac{E^H}{E^H + K_P^H} E_0\right) bm$
$\{P, E\}$	$\emptyset$	$d_{\{P,E\}}$
E_0	bm	$k_2\, bm \frac{E_0}{E_0 + K_g}$

TABLE I

THE NETWORK TOPOLOGY FORMULATED AS A SYSTEM OF CHEMICAL EQUATIONS.

In this particular case the activation function of the nutrient is modelled on the operon-derepression motif that is a common feature of bacterial uptake systems. The idea is that the operon with the permease and metabolic genes is repressed under normal circumstances. Once nutrient enters the cell it will interact with the repressor and release it from the binding site. This then enables the expression of the porin.

The model is summarised in table I. For clarity of presentation we do not present the full system of differential equations, however, the table contains all terms that are contained in the model. The uptake of external nutrient N is described in the first row and depends on the amount of N via a saturating function and depends linearly on the number of porins and the biomass. This latter dependence is necessary in order to simulate exponential growth. Once nutrient is taken up into the cell, it is then converted into internal energy E_0 (second row). Expression of the porin P from gene p is activated by the presence of the nutrient E via a Hill function with Hill exponent $H = 2$ but it also requires internal energy E_0 to proceed. Note that in addition to the regulated term, P is also expressed with a certain leak rate independent of the presence of N. This leak rate is important in order to induce the system when N first appears. Expression of P is also proportional to bm. The fourth line describes breakdown of porins and the internal nutrient. Note that the value d for both is set equal to the growth rate; hence breakdown is linked to dilution in an exponentially growing population. Finally, the last row in table I describes how internal energy E_0 is converted into biomass. The conversion rate is a saturating function of E_0 and depends linearly on the growth rate constant k_2. This constant regulates the amount of nutrient that is converted into growth.

Similar to the differential equation model we stipulated that at the beginning of the simulation there are 1000 units of nutrient, which are then consumed by the cell and converted into biomass or porins. The model contains a number of parameters and the behaviour of the model is crucially dependent on how these parameters are chosen. A manual exploration of the parameter space would be difficult and time consuming. We therefore used a genetic algorithm to evolve parameter sets that maximise the production of biomass. Fitness was the biomass produced after all nutrient was used up. We evolved the system for 5000 generations. Each repetition of the GA resulted in numerically somewhat different solutions (i.e., parameter sets), but each time these solutions were very efficient, in the sense that they converted almost all nutrient

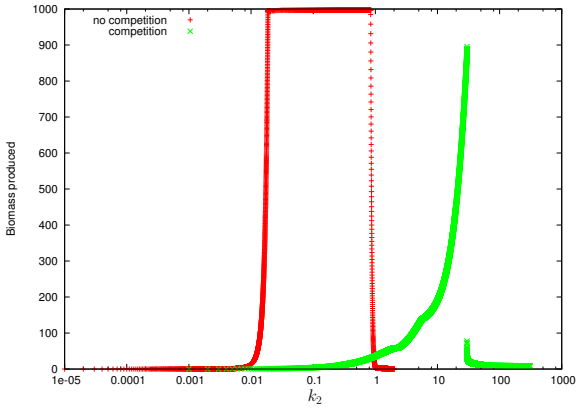

Fig. 2. The biomass produced as a function of the growth parameter k_2 in the model from table I. The left-hand (red) line uses parameters that have evolved to maximise growth yield. The right-hand (green) line has evolved to maximise growth rate. The maximum of the green line is extremely sharp and it drops from $\approx$ 900 to $\approx$ 100 within 0.001 units of k_2.

that was on offer into biomass with very little energy being used for porin production. This is clear from figure 2 which shows that at the maximum the evolved solutions achieved a fitness of almost 1000. An ideally efficient solution converts 1 unit of external nutrient into 1 unit of growth.

In addition to evolving parameter sets for yield, we also evolved sets that optimise growth rate. This was achieved by an iterative extension to the GA as follows: First we evolved a set of parameters optimised for yield, as described above. Next, we fixed this solution and re-ran the GA. In this second iteration the evolving solutions had to directly compete for nutrient with the fixed solution that had evolved previously. In this second iteration we found that the evolving solution came to dominate the fixed solution. Subsequently, we repeated this procedure, with the evolving solution of the second iteration now being fixed in the third iteration and so on. We did this altogether 8 times. We found that later solutions tended to outcompete previous solutions, in the sense that they achieved a higher share of the nutrients or even prevented growth of the previous solution altogether. The amount of improvement over previous solutions decreased with each iteration.

Analysis of the solutions that co-evolved in this way showed that the co-evolutionary strategy to outcompete the fixed solution was to take up nutrient faster, thus preventing the fixed solution from growing. We found that, after a certain number of iterations, evolving solutions failed to outcompete the fixed solution. This is to be expected because parameter values were restricted (arbitrarily) to vary within the range $[0, 15]$; consequently, there was an upper limit on the speed with which nutrient could be taken up.

In order to understand how the biomass production depends on the amount of nutrient used for growth, we chose two evolved solutions. One solution optimised for yield and one optimised for growth rate. Figure 2 compares the results. The maximum biomass attainable in the solutions here is 1000. Interestingly, as the parameter k_2 was varied, the yield-optimised

solution was either close to the maximum — converting all nutrient to biomass — or, for very large or very small k_2, it showed barely any growth at all. The transition between those two regimes is very sudden. Fitness change within the regime is negligible. Hence, substantial fitness change is limited to the transition between regimes.

The picture is even more extreme for the growth-rate optimised solution. While the yield-optimised solution had a broad range of parameter values for which biomass production was essentially constant and optimal, the yield-optimised solution has a sharply focussed maximum. It has a gentle slope leading to the maximum from low values of k_2, but drops almost instantaneously from the maximum to very poor solutions of about 100 biomass. The rate-optimised solution did not evolve to the maximum biomass production. Optimal k_2 is around 28 which is outside the allowed range of the parameter.

Note that for values of k_2 the growth-rate optimised solution produced less biomass than the yield-optimised solution. However, also note that in direct competition the rate-optimised solution would outcompete the yield-optimised one.

V. Stochastic model

The final model we present is identical to the model in table I, but was simulated using stochastic simulation methods. Rather than using differential equations, it was simulated using a Gillespie-style algorithm [15]. Unlike differential equation models, stochastic simulations admit only discrete units of particles. In particular, the minimum amount of a particle that can exist is 1. An implication of this is that models of the type we consider here can suffer "deadlock states." If the cell has no nutrient left and it has no porins available to take up nutrient, then it will die. Even though there may be a large amount of nutrient in the environment, without the pre-requisite resources, the cell is not able to utilise those. The equivalent of such a deadlock state in the differential equation model would be the steady state characterised by $E(t) = P(t) = 0$. However, this state is unstable and small fluctuations away from it will lead to jump starting the cell processes. In particular, this entails that a differential equation model will never get into a deadlock state from a running state for as long as there is nutrient available. Stochastic models, on the other hand, can fluctuate into a deadlock state.

Stochastic simulation algorithms, such as the Gillespie algorithm, execute one reaction at a time, deducting and adding molecules to the pool as stipulated by the reaction. The main task of the algorithm is to choose (i) the next reaction to execute and (ii) the time since the previous reaction. This choice is made probabilistically based on the reaction kinetics and the number of particles of each chemical species in the system. An exact algorithm, such as the Gillespie algorithm, behaves like the equivalent differential equation model in the limit of very large particle numbers. For small numbers of particles the behaviour may be very different from the differential equation model and be dominated by noise.

The deterministic model described above did not have a concept of cell division. While, technically, it is possible to implement dividing cells in differential equations, the resulting model would be extremely complicated and hard to maintain. Hence, in the deterministic model the biomass captures both the actual biomass and cell number (while not strictly differentiating between them). Stochastic models are much more suitable to implement cell division. Here we stipulated that a cell, once it has acquired more than a threshold amount of biomass, can divide. In the case of division, porins and internal energy in the dividing cell are randomly divided between mother and daughter cell.

The differences between the stochastic and deterministic models entail that parameter values are not compatible. In particular, a solution that is perfectly viable leading to substantial growth in the deterministic model may be prone to deadlock in the stochastic case, or simply perform much worse than the deterministic simulation would suggest. Hence, in order to explore the parameter space of the stochastic models, we evolved new solutions.

The fact that stochastic models support division allows us to implement implicit fitness evolution. Rather than specifying a fitness function (which was the biomass generated in the previous case), it is now possible to let solutions compete among each other (c.f. [16]). The idea of the approach is to start the simulation with a number of random solutions that are competing for the same resource, growing and dividing. Upon each division individuals may be mutated, i.e., a random parameter of the solution will be adjusted by a small amount; parameters are constrained to remain in the interval $[0, 15]$. In this case it is not necessary to revert to iterative evolutionary rounds. Instead, the heterogeneity in the population provided sufficient adaptive pressure to obtain competitive solutions.

In figure 3 we have taken one such solution and simulated it in the absence of competition for many different values of the parameter k_2. Again, there is clearly an optimal resource allocation that maximises yield. Note that the amount of biomass produced on any one run depends on the random seed, but the amount of variation from this factor is not equal across the parameter space. For low values of k_2 the variation between simulations was small. Interestingly, around the optimal value the system becomes unstable. Even in the vicinity of the optimal value there were a few runs that resulted in deadlock, i.e., very low fitness. This is indicated by the data points close to the x-axis. The variability of the simulation increases as k_2 increases and for $k_2 > 500$ there is an area where the level of biomass produced is highly variable between runs. Such variations are expected in stochastic variables. However, the salient point for our purpose here is that there is a resource allocation that leads to optimal growth.

VI. Methods

We used Maple 16 software to numerically solve the differential equation models. The artificial evolution of the deterministic model was implemented as a Perl script that creates and modifies Maple files. The stochastic models were implemented using an in-house implementation of the Gibson-Bruck algorithm [17]. This implementation also contained

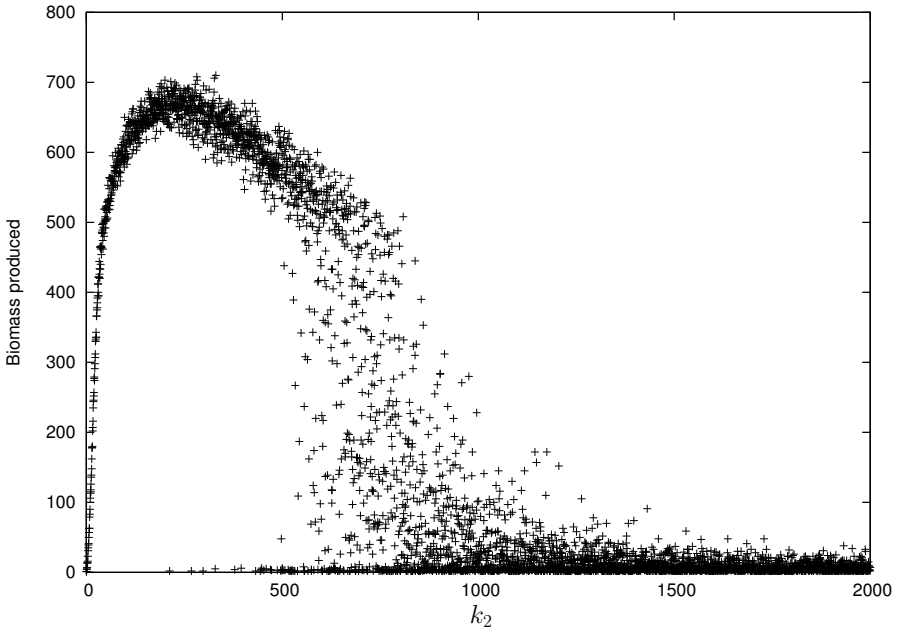

Fig. 3. The amount of biomass produced depending on resource allocation in a stochastic model.

the evolutionary model. The software will be described in a forthcoming publication [manuscript in preparation].

The genetic algorithm used a fitness-proportional selection scheme with mutation and crossover (both with a rate of 0.2 per generation). We also employed elitism conserving the best solution from each generation and a mutated version of the best solution. We found that solutions are not sensitive to variations of these parameters.

The implicit evolution used a random initial population of 50 solutions. These divided as the population grew. During each division event a solution suffered a mutation with a rate of 0.2. When nutrient ran out the simulation was stopped and the population was pruned to 50 by randomly removing cells. The system was then re-started, pruned and so forth, altogether 5000 times. The population was also spatially structured. This means that 32 evolution runs were performed in parallel. At each round there was a small probability for solutions to migrate from one population to another one. Hence, the 32 simulations were weakly interacting. The evolution was halted after 5000 rounds and the fittest solution from all populations was selected as the final solution.

VII. Discussion

The series of models presented here suggests that there is an optimal allocation of resources in a cell. The particular shape of the optimality curve depends on the model but the overall phenomenon is stable for a wide range of model assumptions, although all the models presented here remain very strong idealisations relative to real cells. At the same time, we argue that the stability of the phenomenon and the plausibility argument given above lends credibility to the conjecture that real cells have a similar optimal allocation strategy.

Our results above only outline a potential research program, highlighting a research question that can be addressed in general models, while at the same time deriving results that are relevant for a wider range of biological systems. In this particular case the research challenge is to understand what the optimal allocation strategy depends on in real cells. Unlike the fairly constrained model here, reals cells have a large number of evolutionarily "moving parts." The size of a cell, for example, is under selection pressure. The larger the cell the more space it has available for uptake, but at the same time the more nutrient it needs to consume before it divides. If we want to understand optimal allocation strategies in real cells, then we likely need to take into account the cell size at division as a relevant parameter. The question is whether or not there still exists a single optimum or whether the problem then transforms into a trade-off relationship without a single optimum.

A further refinement of the model we presented here would be to take into account the effects of varying environments. Cells that are adaptable to different conditions need to carry the metabolic cost of maintaining complex regulatory networks [18]. At least in the realm of single-celled organisms, regulation provides the flexibility to exploit a wide variety of environments. Most of all, it allows the cell to maintain a presence in unstable environments. However, in each of the possible environments the cell is worse off than more specialised cells that do not have to carry the burden of complex regulation.

This is an instance of the "generalist-specialist" trade-off in evolutionary biology [19], but playing out at the level of biochemistry. As such this well known evolutionary dilemma can be directly linked to the metabolic cost of maintaining

complex regulatory networks. This cost, on the other hand, is connected to the theory of stochastic computers, i.e., Bennett's "Brownian computers" [20]. Biochemical systems are instances of such Brownian computers. One of the key properties of these systems is that a certain computational speed requires a minimal energetic investment, thus linking energy usage of regulation to speed of regulation in cells [21].

Research questions of the type outlined above have been addressed before [22]. Yet, typically the investigation and the results have remained isolated within specific disciplinary traditions. In particular, physicists are often interested in fundamental limitations of living systems. And indeed, work coming out of physics has led to important insights into the underlying principles of living systems. Yet, these disciplinary approaches that are common at the moment must be pushed out further and be connected to the daily work of experimental biologists. Doing this requires a novel approach.

Combining formal methods from theoretical physics with a more abstract approach similar to early research on artificial chemistries may be useful here. The key idea of the research on design principles was that the key to understanding life is the organisational structure of living systems, not the specifics of individual instances of life. This insight is still a valuable one, as long as it is combined with a reasonable representation of the constraints that real living systems are subject to. A limited abstraction process could then liberate the modeller from the specifics that life science research has produced today. At the same time, it would potentially lead to valuable new insights that can then be validated using this new data (first) and lead to prediction of experimental outcomes (later on).

VIII. Conclusion

Here we have presented some results of models of nutrient allocation in organisms using a number of different models with differing assumptions. All the models predict that there is an optimal allocation of nutrient, although precisely where this optimum is and the shape of the optimality curve depends on the specifics of the model assumption.

We have discussed these results as a possible approach for a more integrated approach of formal modelling in biology. Formal modelling is, at present not a primary source of biological knowledge generation, but often of secondary importance. However, we think that biology with an upgraded input from theory would lead to important new insights that are currently not possible.

References

[1] M. Nowak, *Evolutionary Dynamics: exploring the equations of life.* Harvard University Press, 2006.

[2] R. Rosen, *Life Itself.* New York: Columbia University Press, 1991.

[3] B. McMullin and D. Groß, "Towards the implementation of evolving autopoietic artificial agents," in *Proceedings of the 6th European Conference on Advances in Artificial Life*. Springer-Verlag, 2001, pp. 440–443.

[4] F. Varela, *Principles of Biological Autonomy*, ser. The North Holland series in general systems research. North Holland, New York, 1979.

[5] P. Speroni Di Fenizio and P. Dittrich, "Artificial chemistry's global dynamics. movement in the lattice of organisation," *The Journal of Three Dimensional Images*, vol. 16, no. 4, pp. 160–163, 2002.

[6] P. Dittrich, J. Ziegler, and W. Banzhaf, "Artificial Chemistries—A Review," *Artificial Life*, vol. 7, pp. 225–275, 2001.

[7] H. Sayama, "Swarm chemistry," *Artificial Life*, vol. 15, no. 1, pp. 105–114, 2009.

[8] T. Hutton, "Evolvable self-reproducing cells in a two-dimensional artificial chemistry," *Artificial life*, vol. 13, no. 1, pp. 11–30, 2007.

[9] G. Benkö, C. Flamm, and P. Stadler, "Generic properties of chemical networks: Artificial chemistry based on graph rewriting," in *Advances in artificial life*. Springer, 2003, pp. 10–19.

[10] H. Stanley, L. Amaral, S. Buldyrev, A. Goldberger, S. Havlin, B. Hyman, H. Leschhorn, P. Maass, H. Makse, C.-K. Peng *et al.*, "Scaling and universality in living systems," *Fractals*, vol. 4, no. 03, pp. 427–451, 1996.

[11] H. Stanley, L. Amaral, S. Buldyrev, A. Goldberger, S. Havlin, H. Leschhorn, P. Maass, H. Makse, C.-K. Peng, M. A. Salinger *et al.*, "Scaling and universality in animate and inanimate systems," *Physica A: Statistical Mechanics and its Applications*, vol. 231, no. 1, pp. 20–48, 1996.

[12] G. West, J. Brown, and B. Enquist, "A general model for the origin of allometric scaling laws in biology," *Science*, vol. 276, no. 5309, pp. 122–126, 1997.

[13] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, 2006.

[14] D. Chu and T. von der Haar, "The architecture of eukaryotic translation." *Nucleic Acids Research*, vol. 40, no. 20, pp. 10 098–10 106, Nov 2012. [Online]. Available: http://dx.doi.org/10.1093/nar/gks825

[15] D. Gillespie, "Exact stochastic simulation of coupled chemical reactions," *The Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[16] D. Chu, "Modes of evolution in a parasite-host interaction: Disentangling factors determining the evolution of regulated fimbriation in e. coli," *Biosystems*, vol. 95, no. 1, pp. 67–74, 2008.

[17] M. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *Journal of Physical Chemistry*, vol. 104, pp. 1876–1889, 2000.

[18] D. Chu, N. Zabet, and B. Mitavskiy, "Models of transcription factor binding: Sensitivity of activation functions to model assumptions," *Journal of Theoretical Biology*, vol. 257, no. 3, pp. 419–429, 2009.

[19] A. New, B. Cerulus, S. Govers, G. Perez-Samper, B. Zhu, S. Boogmans, J. Xavier, and K. Verstrepen, "Different levels of catabolite repression optimize growth in stable and variable environments." *PLoS Biol*, vol. 12, no. 1, p. e1001764, Jan 2014. [Online]. Available: http://dx.doi.org/10.1371/journal.pbio.1001764

[20] C. Bennett, "On the Nature and Origin of Complexity in Discrete, Homogeneous, Locally-Interacting Systems," *Foundations of Physics*, vol. 16, no. 6, pp. 585–592, 1986.

[21] T. Kalisky, E. Dekel, and U. Alon, "Cost-benefit theory and optimal design of gene regulation functions," *Physical Biology*, vol. 4, no. 4, p. 229, 2007. [Online]. Available: http://stacks.iop.org/1478-3975/4/i=4/a=001

[22] D. Molenaar, R. van Berlo, D. de Ridder, and B. Teusink, "Shifts in growth strategies reflect tradeoffs in cellular economics." *Molecular Systems Biology*, vol. 5, p. 323, 2009. [Online]. Available: http://dx.doi.org/10.1038/msb.2009.82

[23] D. Chu, E. Kazana, N. Bellanger, T. Singh, M. Tuite, and T. von der Haar, "Translation elongation can control translation initiation on eukaryotic mrnas." *EMBO Journal*, vol. 33, no. 1, pp. 21–34, Jan 2014. [Online]. Available: http://dx.doi.org/10.1002/embj.201385651