Virtual Microbes: a model for studying eco-evolutionary dynamics in microbial evolution

Thomas D Cuypers

March 12, 2018

Abstract

Microbial evolution is governed by the interactions between many different organisational levels and timescales. Metabolic function, regulatory interactions and ecosystem interactions together determine the reproductive potential, and survival, of individuals. In addition, the direction of evolution is influenced by selection for robustness and evolvability. Together these levels form the substrate of evolution, ultimately shaping the the eco-evolutionary dynamics of microbes. To study the interactions between levels of selection on evolution and ecology we need to study so called *non-supervised* models. These models define low level entities and process at a fairly fine level of detail, allowing higher level patterns and functionality to emerge from from interactions at the lower levels

by are capable of bringing about the interplay between multiple biological structures, levels of selection, and timescales in evolving populations.

We developed Virtual Microbes to study long term eco-evolutionary dynamics of microbes. Virtual microbes exchange metabolites with their environment, and evolve to synthesise building blocks for cell growth. They express genes (transporters, enzymes, and transcription factors) from a spatially structured genome, which evolves by large and small scale mutations.

Introduction

All diversity and complexity of biological organisms has arisen "from so a simple a beginning". Evolution is the single most powerful explanation for understanding all this complexity and diversity in our biosphere. Species we observe today have evolved many complex organisational levels, e.g. genes, genomes, individuals, groups, ecosystems etc. How these levels evolved, and how they in turn in u- ence the evolutionary processes, is a challenge we face today. Earlier research has shown that this interplay between levels of organisation can lead to counter-intuitive changes in the direction of selection [refs on reversal of selection levels; volvox, early death]. New levels of organisation and selection can emerge from ecosystems interactions, which can in turn lead to the interlocking of ecological and evolutionary time scales. [van der laan...]. Moreover, evolution both shapes and is shaped by all the aforementioned levels and timescales in biology, which may manifest as the evolution of evolvability [ref evolvability, ref crombach]. These features of biological systems make it exceedingly challenging to study evolutionary dynamics at both short and long timescales.

Important insights have come from bioinformatic approaches that reconstruct phylogeny and ancestral gene contents. Intriguingly, these studies show a two-faced appearance of genome evolution. On the one hand we observe strong genome size dynamics, where lineage specific gene family gains and losses drive innovation and diversification [ref Francino 2005, ref Anantharaman 2007, ref Ames 2010, ref Koonin 2002], while on the other hand highly regular scaling laws of genomic features appear to persist across domains of life [Koonin, 2011]. Inspired by these observations we previously developed the Virtual Cell model [Cuypers and Hogeweg, 2012] to study emergent patterns in genome evolution. In this *in silico* model, cells adapt to a continually changing environment by regulating their internal homeostasis of a single resource and its downstream energy carrier molecule. Virtual Cells evolve through gene level point mutations as well as genome level duplications and deletion of stretches of neighbouring genes. Within this simple, yet plausible set of biological functions, we demonstrated generic evolutionary patterns such as genome inflation and streamlining [Cuypers and Hogeweg, 2012], fixation of whole genome duplications [Cuypers and Hogeweg, 2014], and the biased retention of duplicated Transcription Factors (TFs) [Cuypers and Hogeweg, 2014]. This previous work demonstrates the exploratory power of what we call *non-supervised* modelling, *i.e.* implementing the basic processes that play an important role in a biological system, and studying the evolutionary patterns as well as the structure and behaviour of the evolved organisms.

The Virtual Cell model does however have a number of limitations if we wish to study the emergence and implications of diversity. For example, why are some species able to utilize many different resources, while others species share or partition available resources? When and why does a cross-feeding interaction between species emerge? Do functionally similar species or ecosystems evolve if we redo evolution under the same environmental resource conditions? How does prior evolution towards many or fluctuating resources influence adaptation to a single constant resource and *vice versa*? Because the Virtual Cell only includes a single resource, and does not include cell-cell and environmental

interactions, we need to extend the model.

Here we introduce the Virtual Microbe, a model that builds upon the structure of the Virtual Cell model in order to study the aforementioned topics in a similar *non-supervised* way. In this model, Virtual Microbes (VMicrobes) grow and die naturally by metabolising a multitude of resources according to the rules of predefined artificial chemistry. The microbes furthermore have a strong interplay with their spacial environment by means of passive diffusion or active transport across the membrane, and lysing of dead cells.

Results and discussion

Model overview

Virtual Microbes (VMicrobes) is an individual based model that integrates metabolism, gene regulation and mutation with ecological interactions through reproductive competition and metabolite exchange. VMicrobes compete for resources and the opportunity to reproduce in a limited space. They shape their local environment by actively transporting and converting metabolites as well as by passive metabolite diffusion over their membranes. The local buildup and depletion of metabolites is dissipated through diffusion along environmental concentration gradients.

Reproductive success of VMicrobes depends on their ability to grow and divide (Figure 1A). To do so, VMicrobes are required to synthesise essential building blocks and energy. Typically, building blocks nor energy metabolites appear natively in the environment, compelling VMicrobes to synthesise them from environmentally available precursors by expressing appropriate metabolic pathways. Building blocks (indicated as the purple metabolite in Figure 1) are consumed for cell growth and for protein expression. Energy is necessary for active transport and for the catalysis of various anabolic reactions. By actively taking up and exporting metabolites and by leaking metabolites through their membranes, VMicrobes alter their environment, leading to indirect interactions (Figure 1B). One reason for VMicrobes to actively export metabolites is the effect of a toxic buildup of internal metabolite concentrations.
Upon cell death, VMicrobes also release metabolites into the environment.

The genes required to synthesise building blocks and energy metabolites, are organised in a structured genome (Figure 1C) that can undergo large-scale indels, as well as point mutations. Besides transporters and metabolic enzymes, these structured genomes also encode transcription factors (TFs), that can bind to the upstream operator sequences of genes, therewith regulating expression (as indicated in Figure 1C). This regulatory effect can furthermore be enhanced or suppressed by binding to a ligand molecule. Multiple TFs with varying binding motifs may bind to the operator, each exerting their own effect, which is a function of the inherent binding strength, the particular regulatory effect and concentration of the TF.

The complete set of enzymatic and transport reactions that are possible during a simulation (Figure 1D) are determined at the time of initialisation

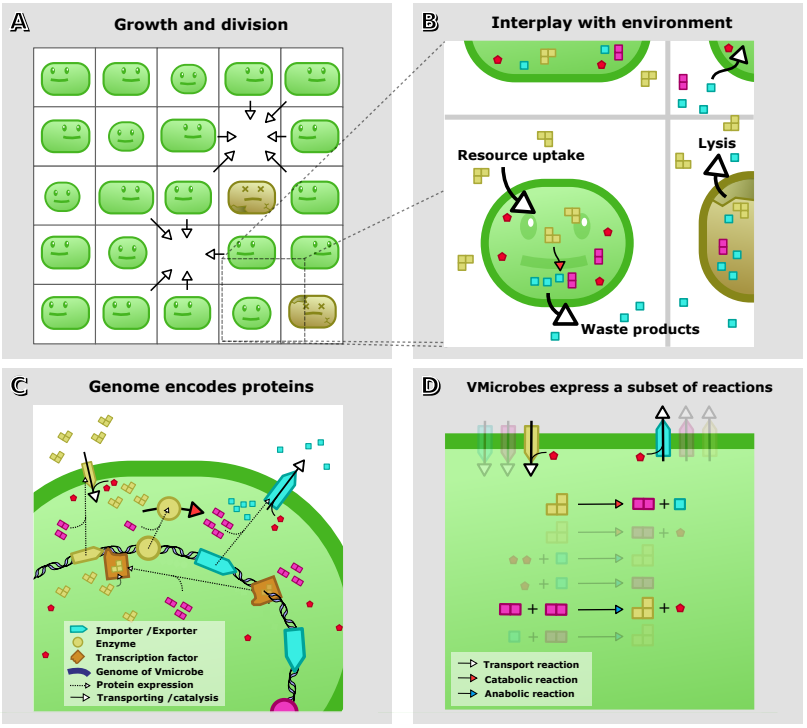

Figure 1: **Overview of the VirtualMicrobe**.
A) VMicrobes live on a grid, and compete for empty space by growing and dividing. Cells below a certain volume do not (yet) compete.
B) By taking up resources, exporting waste products, and lysing into the environment, VMicrobes have a strong interplay with their environment.
C) Genes of VMicrobes are encoded as a "pearl-on-a-string"-genome, resulting in the expression of transporter proteins, enzymatic proteins and transcription factors. Using these proteins, the VMicrobes have to synthesize the necessary compounds to express proteins and increase the cell volume. Transcription factors can modulate gene expression by binding operator sequences and sensing metabolite concentrations.
D) At the start of the simulation, the set of all possible reactions is either procedurally generated, or user defined. The VMicrobes do not necessarily express all these reactions, but can discover proteins to catalyse any of these reactions.

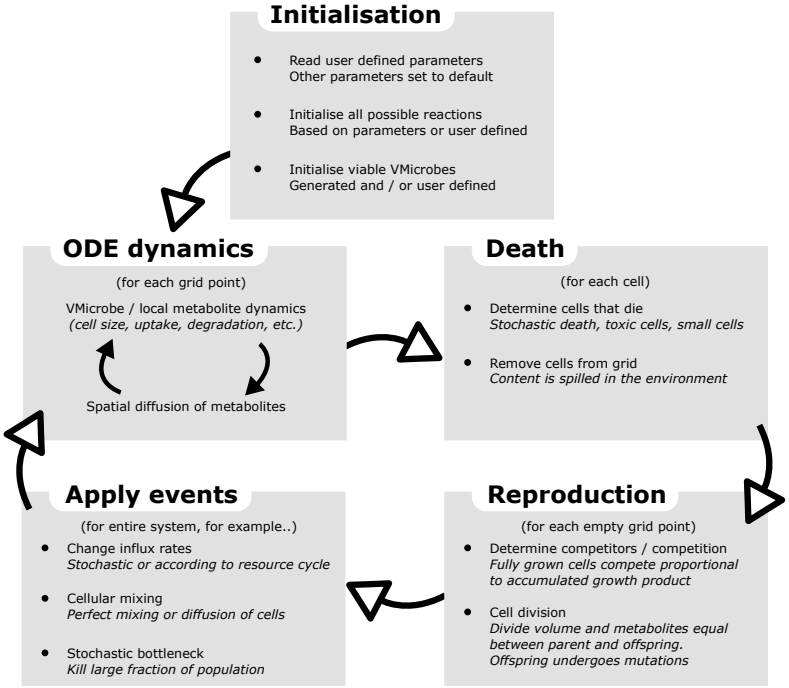

Figure 2: **Simulation phases of Virtual Microbes**.
After initialisation, the simulation process runs through a sequence of 4 phases. It runs for a specified number of update steps, or until the population goes extinct.

(see methods section **Metabolic universe construction**). This "metabolic universe" is either constructed according to flexible parametrisation options or constructed by the user. Furthermore, enzyme promiscuity can be incorporated by letting enzymes act on sets of metabolites instead of particular metabolites (see methods section **Metabolism**). After these reactions are defined, the simulation will initiate with VMicrobes that have the minimal genes to start growing, and start simulating.

The simulation process runs for a specified number of update steps, or until the population goes extinct. Each update step consists of 4 sequential computational phases (displayed in Figure 2). In the first phase the environmental and internal microbial dynamics defined by sets of ordinary differential equations (ODEs) are run in parallel, interleaved with spatial diffusion of metabolites, for a fixed amount of time. The second phase starts after molecular dynamics are temporarily paused. Microbes die randomly with a probability that is the base death rate increased with a penalty incurred from toxic levels of internal metabolites, and are cleared from the grid after spilling their internal metabolite contents. In the third phase, Vmicrobes that have reached a minimum division volume compete to reproduce in empty spaces, with their competitive strength depending on accumulated growth product. Reproducing cells divide their volume equally between parent and offspring and offspring undergo mutations before being placed on the grid. In the fourth, final phase environmental changes may be applied, including varying metabolite influx rates and microbe spatial diffusion or complete mixing.

The complete sets of metabolic species and the enzymatic and transport reactions that are possible during a simulation are determined at the time of initialization according to flexible parametrization options. An important parameter is the number of Molecule Classes that exist in the *metabolic universe* of the simulation. Molecule Classes group metabolites into similarly behaving, related molecular species. More specifically, substrates and products of enzymatic and transport reactions are defined at the level of Molecule Classes. Genes encoding a particular metabolic reaction can thus catalyse a set of reactions where metabolic species within the same Molecule Class can replace each other within the reaction scheme. The rationale for this structure of the model is that we can represent substrate promiscuity of enzymes and allows to study its role in the evolution of gene dosage and divergence. This promiscuity also holds for TFs regarding their ligand molecules. Proteins can evolve their specificity for any of the species within a Molecule Class through mutations of species specific K parameters. Finally, if a Molecule Class consists of multiple species then a set of enzymatic reactions exist to convert the species in the MC into each other. These reactions resemble the addition or removal of small chemical groups, to yield closely related molecular species.

Methods

This section describes the construction of the metabolic universe and the spatial grid and the initialisation of Vmicrobes. and detailed life history in terms of competition, reproduction, death and mutations. The source code of the model

is available for download at `https://bitbucket.org/thocu/virtualmicrobes`, including instructions for installation and running a simple evolutionary simulation.

Environment

The environment is constructed according to a set of parameters that specify the properties of the metabolic universe and the spatially explicit environment in which Vmicrobes evolve. A high flexibility in the experimental setup poses some challenges to finding relevant, viable conditions for Vmicrobe evolution. As in real organisms, their biological functions are strongly connected and hence create various dependencies between parameter values. A straightforward example of this is the interaction between biomass production rate, resource influx and death rate of microbes, that must be adjusted to prevent rapid population extinction. For the model to generate interesting eco-evolutionary dynamics parameters need some degree of tuning.

metabolic universe construction The metabolic universe is the a priori set of all molecular species and the reactions between them. Individual Virtual Microbes can perform subsets of these reactions if the express associated enzymes in the presence of the substrates of the reactions. What follows is a description of the parameters and rules that are used to construct the metabolic universe for a specific simulation. First, the number of *resource* Molecule Classes and the number (fixed or drawn within a range) of molecule species per class is chosen. Separate *energy* MCs are defined that are used as energy currency in several cellular functions. Each MC is assigned an *energetic value* that will be used to prevent spontaneous mass creation (even though reactions are not strictly mass balanced). Next, the reactions over the set of MCs are defined. Molecules in non-energy MCs can be transported over the cell membrane when energy is spent. In addition, all molecules within a single resource MC can be converted into each other, consuming energy. A set of catabolic reactions is constructed as follows: a resource MC is chosen at random as the substrate and then a combination of reaction products is chosen from resource MCs and energy MCs that have a combined energetic value less then or equal to that of the chosen substrate. Additional constraints can be set on the number of MCs that can be produced per reaction and on the maximum yield in energy molecules so that breaking up a large molecule will require a sequence of enzymatic steps in which energy is produced gradually. When repeating the procedure to create the specified number of catabolic reactions, a maximum pathway convergence parameter limits the number of occurrences of any MC in the rhs of reactions. In similar fashion anabolic reactions are generated by finding substrate combinations that match the energetic value of a larger reaction product. Anabolic reaction can include energy MCs as substrates. The rules detailed above are designed to facilitate the random generation of viable environments that also allow for interesting evolution of Vmicrobe metabolism.

Spatial structure The environment is structured as a grid of equivoluminous sites that can contain up to a maximum amount of individuals (default is 1). Boundary conditions can be specified separately for the horizontal and vertical sides to be either fixed or wrapped as well as the presence of (permeable)

parameter	description	simulation values
fraction-influx	the fraction of metabolites with positive influx	1.
nr-resource-classes	number of resource Molecule Classes	8
nr-energy-classes	number of energy Molecule Classes	1
mol-per-res-class	number of species per resource Molecule Class	1
mol-per-ene-class	number of species per energy Molecule Class	1
res-energy-range	range of energy values of resource molecules	$[2, 10]$
ene-energy-range	range of energy values of energy molecules	$[1, 1]$
nr-cat-reactions	(desired) number of catabolic reactions	20
nr-ana-reactions	(desired) number of anabolic reactions	20
cat-path-convergence	max times metabolite is product in catabolic reaction	6
ana-path-convergence	max times metabolite is product in anabolic reaction	6
nr-building-blocks	number of metabolites designated as building blocks	2
diffusion-constant	spatial diffusion constant of metabolites	0.05
fluctuate-frequency	(average) time interval for influx rate changes	fixed or 200
influx-range	range of possible influx rates	$10^{Uniform(-5,-2)}$
small-mol-degradation	degradation rate of molecules	10^{-3}
neighbourhoods	grid neighbourhoods for different interaction modes	compete:Moore9 HGT:Neumann13
cell-shrink-rate	constant rate of cell shrinking/ volume maintenance	
nr-building-blocks	number of designated building block metabolites	

Table 1: Important initialization parameters and default values

barriers to Vmicrobes or metabolites throughout the grid. In addition, sub-environments can be superimposed on the grid to simulate multiple habitats with different metabolite composition and influx rates. Influx rates may vary between metabolite species and over time. When an influx range is specified influx fluctuates stochastically and takes new values $i_{new} = b^{Uniform(l,u)}$. Influx for energy molecules and Vmicrobe building block metabolites is 0 by default as these are assumed to be exclusively produced in intracellular enzymatic reactions. Metabolites degrade at a constant rate and diffuse on the spatial grid as well as over cell membranes, resulting in concentration gradients.

Most often, Vmicrobes interact with their local neighbours. For different modes of interaction it is possible to have interaction neighbourhoods with different shapes. For example, the default neighbourhood for reproductive competition is Moore9 while HGT event may occur within the larger von Neumann13 neighbourhood. Vmicrobes can diffuse to neighbouring grid points as well as being randomly redistributed over the grid at some regular interval, although both procedures come at the computational cost of redefining the local system of ODEs. At the start of a simulation a newly initialized population of Vmicrobes is placed randomly on the grid. Vmicrobe genome initialization is described in the next section.

Genome initialization

Vmicrobes can grow to reach their minimum division volume by producing designated *building block* metabolites or by acquiring them, via active or passive transport, from the environment. Part of the production has to be spent on maintenance of the cell volume, requiring Vmicrobes to maintain a minimum level of production to prevent them from shrinking below the minimum cell size. To achieve this, they need adapt their metabolism both to the intrinsic environmental resource conditions and the changing metabolic activity of their competitors.

Vmicrobes encode proteins on one or more chromosomes with a spatial (linear or circular) layout. Each protein coding gene has its own promoter and

operator site. Operator binding sites are encoded as bit strings that can be recognized by TFs with a matching binding motif. Genomes are initialized with a specified number of TFs, metabolic and transport genes. In order to initialize individuals with a viable metabolism, metabolic genes are selected according to these stochastic rules:

1. Enzymes that can produce the building blocks and energy metabolites are added first
2. Any further enzymatic steps should connect their product metabolites to the existing downstream metabolic network
3. Transport reaction are first selected that match metabolites with positive environmental influx rates
4. Further transporters are prioritized for connecting to internal metabolism
5. TFs are selected to have ligands that are part of the internal metabolism of the cell

Genes are initialized with randomized kinetic parameters and added to the genome in random spatial order. Once Vmicrobes are initialized and placed on the grid, the population starts evolving by simulating the life histories of all individuals in parallel.

Life history

As described above, the growing or shrinking of Vmicrobe cell volume, determines whether they may divide or die from lack of nutrients. Additional causes of death are the buildup of toxic metabolite concentrations and natural death.

production Vmicrobes convert building blocks to a production value P by a simple function

$$\frac{dP}{dt} = \prod_{m \in B} [m] - d \cdot P$$

, with B the set of building block metabolites defined for the individual and d a low degradation rate on the production value. Different species may require different sets of essential building blocks for growth. As well as being consumed for cell growth and maintenance and protein production, P determines strength with which individuals compete to reproduce.

cell volume growth We assume that if Vmicrobes grow without being able to divide due to lack of space, they approach a maximum cell size V_{max} and that there is a continuous turnover d of the cell volume at steady state. Volume then changes as

$$\frac{dV}{dt} = g \cdot V \cdot \frac{1 - V}{V_{max}} \cdot P_{scaled} - d \cdot V$$

where P_{scaled} is the current production value of the cell, scaled by the time adjusted population wide production rate P_{pop} as:

$$P_{scaled} = \frac{P^s}{P^s + {P_{pop}}^s}$$

in which s is a constant determining the shape of the scaling function.

toxicity and death All life forms must maintain internal molecule concentrations within permissible bounds for their survival. To address this premise, we define a toxic effect function for virtual microbes. Metabolites can become toxic to microbes when the internal concentrations reach a toxicity threshold. A cumulative toxic effect is computed over the current life time τ of a Vmicrobe as

$$e_{tox} = \sum_{m \in M} \int_{t=0}^{\tau} f(m,t)dt$$

for all internal molecules M, with

$$f(m,t) = max(0, \frac{[m]_t - tox_m}{tox_m})$$

the toxic effect function for the concentration of molecule m at time t with toxicity threshold tox_m. This toxic effect increases the death rate d of microbes starting at the intrinsic death rate r

$$d = \frac{e_{tox}}{s + e_{tox}} \cdot (1 - r) + r$$

where s scales the toxic effect. Vmicrobes that survive after an update cycle retain the toxic level they accumulated so far.

reproduction Reproduction can happen in empty spaces on the grid. Competition is local between neighbouring Vmicrobes. Reproduction is determined by fitness proportional selection where fitness is given by production value. The roulette wheel includes a *no-event* term, representing the chance of no reproduction event happening. Because many different factors of the particular environment determine the theoretical maximum production value that can be reached it becomes highly impractical to determine the size of the no-event a priori. Moreover, because the average production in the population typically increases several orders of magnitude during the course of evolution, a fixed no-event would soon be rendered inconsequential. We therefore determine it dynamically, using a sliding time window on the population median production. The value of the no-event term is then determined as the maximum of the sliding window and the currently stored value. This has the effect that selection pressure will remain relatively constant over large evolutionary increases in the population production, and can be compared to evolution under chemostat conditions. It is important to note, however, that heterogeneous spatial conditions can result in very different growth conditions for Vmicrobes at different grid locations.

mutation When a Vmicrobe is selected to reproduce, it divides and the offspring inherits a copy of the parent's genome. The genome is subject to various types of mutations. Chromosomes can undergo fission, fusion, deletion and duplication. Also, stretches of genes can be inverted, deleted or duplicated, or trans-located to a different chromosome. A random genome position is chosen start a stretch mutation and stretch lengths are exponentially distributed. At a single gene level, all evolvable parameters can be independently mutated to new values (see Table 2). A new value of a continuous parameter is drawn as

$$v_{new} = max(min(v \cdot b^{Uniform(l,u)}, v_{max}), v_{min})$$

Parameter	Gene Types	value range in simulation
Promoter Strength	Enzyme, Transporter, TF	$[0.01, 8]$
K-bind-substrates	Enzyme, Transporter	$[0.01, 8]$
K-bind-energy	Transporter	$[0.01, 8]$
K-bind-ligand	TF	$[0.01, 8]$
K-bind-operator	TF	$[0.01, 8]$
V-max	Enzyme, Transporter	$[0.01, 8]$
effect-bound	TF	$[0.01, 8]$
effect-apo	TF	$[0.01, 8]$
ligand	TF	$M \in MoleculeClasses$
exporting	Transporter	[True,False]
sense-external	TF	[True,False]
binding-motif	TF	bit flip at random position
operator-sequence	Enzyme, Transporter, TF	bit flip at random position

Table 2: parameter mutational ranges

, where b, l, and u determine mutational effect and v_{min} and v_{min} are the upper and lower bound of parameter values. Operator and TF binding motif sequences are mutated by flipping a bit in randomly chosen positions. TFs can mutate to recognize a different MC as ligand. Finally rare mutations can cause a TF to change between sensing internal and external ligand concentrations, and transporters to change their directions of transport.

Metabolism

Metabolic enzymes catalyse reactions of the general form

$$R_0 + R_1 + \ldots \longrightarrow P_0 + \ldots$$

converting reactant MCs $\{R_i, \ldots\}$ to products $\{P_j, \ldots\}$. The rate of catalysis, v is calculated with standard Michaelis-Menten kinetics as

$$v = v_{max_{\mathcal{E}}} \cdot [\mathcal{E}] \cdot \frac{\prod_{R \in \mathcal{R}} [R]}{\prod_{R \in \mathcal{R}} ([R] + K_R)}$$

where $[\mathcal{E}]$ is the concentration of the enzyme catalysing the reaction and $\mathcal{R}$ the set of all reactant MCs.

The general forms of the equation describe reactions between MCs. Each MC consists of a set of chemically related molecule species, each named with an index (e.g. A_0 is the first species in the A MC). To model the promiscuous activity of enzymes that catalyse the general reaction between MCs, we generate a set of equations for all the catalysed reactions between species. This is done by mapping species in the n^{th} reactant MC on the lhs by their index to species at the same index in the n^{th} product MC on the rhs, if it exists. Thus we get the following reactions for molecules species:

$$A_i + B_j + \ldots \longrightarrow C_i + D_j + \ldots$$

where index i ranges over the minimum number of species in the A and C MCs and j ranges over the minimum number of species in the B and D MCs. Any species with a higher index in a reactant or product MC will not react or be produced, respectively. If the number of reactant MCs in the rhs is not equal to that in the lhs, any unmatched MCs will have reactions for all their species. Thus, for reactions of with the following form

$$A_i + B_j \longrightarrow C_i$$

where the number of substrate terms is higher than the number of product terms j ranges over all the species in MC B. This also holds for reactions of the form

$$A_i \longrightarrow C_i + D_j$$

but now reactions exist for all the species in D.

Each of the parallel reactions can have different efficiencies, owing to the different specificities K_r for each reactant species encoded by enzymes. An evolutionary process may optimize enzyme affinities for a subset of species in each reactant Molecule Class, giving rise to a specialized enzyme. Because all affinities evolve independently, it would be possible for multiple parallel reactions to evolve to high reaction rates. To retain the property of a maximum enzyme efficiency we rescale the rates of all parallel reactions if the sum of their rates exceeds the enzyme encoded maximum catalysis rate ($v_{max_\mathcal{E}}$).

transport Transporters $\mathcal{T}$ catalyse the transport of substrate S over the cell membrane by consuming energy molecules E. Transport rate is given by

$$v = v_{max_\mathcal{T}} \cdot [\mathcal{T}] \cdot \frac{[S] \cdot [E]}{([S] + K_S) \cdot ([E] + K_E)}$$

where Ks are Michaelis-Menten constants of the substrate binding to the transporter. Depending on the direction of transport (importing or exporting) S is either the external or the internal concentration of the substrate.

For transporters we assume that any species in the energy Molecule Class can be consumed to import any species of the substrate Molecule Class being transported. Thus, the set of parallel reactions is constructed by taking the product of substrate species and energy species as the lhs of the equations. Naturally, transported species map onto the same species on the opposite side of the membrane.

In addition to enzyme kinetics, metabolite influx, passive membrane diffusion, and degradation have a straightforward implementation.

transcription regulation The rates at which genes are transcribed and translated are a function of their sequence as well as the concentrations of binding TFs and their molecular ligands. The intrinsic basal expression rate of a gene is encoded by a strength parameter in a gene's promoter region. This basal expression rate can be modulated by TFs that bind to an operator sequence associated with the gene. Binding sites and TF binding motifs are modelled as bit-strings and matching depends on a certain fraction of sequence complementarity. If a minimum complementarity is chosen < 1 a match may occur anywhere within the full length of the operator binding sequence and the TF binding motif. The maximum fraction of complementarity achieved between matching sequences linearly scales the strength with which a TF binds the target gene. In addition to binding strength following from sequence complementarity, TFs encode an intrinsic binding affinity for promoters K_b, representing the structural stability of the TF-DNA binding complex.

TFs can, themselves, be bound to small ligand molecules, altering the regulatory effect they exert on downstream genes. These effects are encoded by parameters eff_{bound} and eff_{apo} for the ligand-bound and ligand-free state of the

TF, respectively, and evolve independently. Analogous to the modeling of enzyme interactions with small molecules, TF ligands are specified at the Molecule Class level and binding affinities for separate molecule species within a class evolve independently. Ligand binding to TFs is assumed to be a fast process, relative to enzymatic and transcription-translation dynamics, and modeled at quasi steady state. We determine the fraction of TF that is not bound by any of its ligands L:

$$W_{apo} = \prod_{l \in L} (1 - \frac{[l]}{[l] + K_l})$$

The fraction of time that a TF τ in a particular state σ (bound or apo) is bound to a particular operator o:

$$V_o = \frac{[\tau_\sigma] \cdot c_{\tau o} \cdot K_{b_\tau}}{1 + \sum_{\sigma \in \mathcal{S}} \sum_{\tau_\sigma \in \mathcal{T}} [\tau_\sigma] \cdot c_{\tau o} \cdot K_{b_\tau}}$$

depends on the inherent binding affinity K_{b_τ} as well as the sequence complementarity score $c_{\tau o}$ between the tf binding motif and the operator sequence [cite Neyfahk]. The binding polynomial in the denominator is the partition function of all TFs $\mathcal{T}$ in any of the states $\mathcal{S}$ that can bind the operator. Note that small declines in the concentration of free TFs due to binding to operators are neglected.

Now, the operator mediated regulation function for any gene is given by

$$Reg = \sum V_i \cdot E_i$$

with V_i the fraction of time that the operator is either unbound or bound by a TF in either ligand bound or unbound state and E_i the regulatory effect of that state (1 if unbound or eff_{bound} or eff_{apo} when bound by a ligand bound or ligand free TF, respectively). Finally, protein concentrations $[\mathcal{P}]$ are governed by the function:

$$\frac{d[\mathcal{P}]}{dt} = Pr \cdot Reg \cdot P_{scaled} \cdot E_{scaled} - degr \cdot [\mathcal{P}]$$

where Pr is the evolvable parameter promoter strength and $degr$ a fixed protein degradation rate.

References

T. D. Cuypers and P. Hogeweg. Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. *Genome biology and evolution*, 4(3):212–229, 2012.

T. D. Cuypers and P. Hogeweg. A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLoS Comput Biol*, 10(4):e1003547, 2014.

E. V. Koonin. Are there laws of genome evolution? *PLoS Comput Biol*, 7(8): e1002173, 2011.