A hierarchical thermodynamic imperative drives the evolution of self-replicative life systems towards increased complexity

Sergio Menendez

Institute of Quantitative Biology, Biochemistry and Biotechnology; School of Biological Sciences; The University of Edinburgh

Correspondence: smenendez@hotmail.co.uk; sergio.menendez@ed.ac.uk

Abstract

From a basic thermodynamic point of view life structures can be viewed as dissipative open systems capable of self replication. Energy flowing from the external environment into the system allows growth of its self replicative entities with a concomitant decrease in internal entropy (complexity) and an increase of the overall entropy in the universe, thus observing the second law of thermodynamics. However, efforts to derive general thermodynamic models of life systems have been hampered by the lack of precise equations for far from equilibrium systems subjected to arbitrarily time varying external driving fields (the external energy input), as these systems operate in a non-linear response regime that is difficult to model using classical thermodynamics. Recent theoretical advances, applying time reversal symmetry and coarse grained state transitions, have provided helpful semi-quantitative insights into the thermodynamic constraints that bind the behaviour of far from equilibrium life systems. Setting some additional fundamental constraints based on empirical observations allows us to apply this theoretical framework to gain a further semi-quantitative insight on the thermodynamic boundaries and evolution of self replicative life systems. This analysis suggests that complex self replicative life systems follow a thermodynamic hierarchical organisation based on increasing accessible levels of usable energy (work), which in turn drive an exponential punctuated growth of the system's complexity, stored as internal energy and internal entropy. This growth has historically not been limited by the total energy available from the external driving field for the earth life system, but by the internal system's adaptability needed to access higher levels of usable energy. Therefore, in the absence of external perturbations, the emergence of an initial self replicative dissipative structure capable of variation that enables access to higher energy levels is sufficient to drive the system's growth perpetually towards increased complexity across time and space. Furthermore, the self-replicative system would adopt a hierarchical organisation with all permitted energy niches evolving to be optimally occupied in order to dissipate the work input from the external drive and further adapting as higher energy levels are accessed. This model is consistent with current empirical observation of life systems across both time and space and explains from a thermodynamic point of view the evolutionary patterns of complex life systems on earth. We propose that predictions from this model can be further corroborated in a variety of artificially closed systems and that they are supported by experimental observations of complex ecological systems across the thermodynamic hierarchy.

Introduction

A consensus definition of what features define a life system remains controversial (Schrödinger, 1967). However, among the basic requirements for life the capability of a system to self replicate and therefore propagate itself over space and time seems to be accepted as a minimal essential feature. Self-replicative (SR) live systems appeared on earth around 4.2 billion years ago. Current hypotheses speculate that polymeric assemblies (such as RNA) emerged under suitable constrained environments as the original SR systems, however the precise mechanisms and environment required for this event remain elusive (Attwater et al., 2010; Nakashima et al., 2018; Walker, 2017). From a biochemical point of view the process of abiogenesis could be viewed simplistically as a reaction/s where a number of basic components (reactants) organise

themselves into more complex assemblies (product/s) that acquire the emergent property of self-replication (SRn). Whether an external energy source was needed to catalyse this initial step, or simple thermal fluctuations within the equilibrium frame of an existing reversible reaction were sufficient to acquire a higher energy level assembly capable of SRn remains unclear (England, 2015). Nevertheless, in order for there to be further growth, the original SR assemblies would have required to be able to tap into an external energy source to drive future self-replication and escape further from equilibrium (Duim & Otto, 2017; Prigogine & Nicolis, 1971; Ragazzon & Prins, 2018). The nature of this original accessible energy source and the initial equilibrium conditions remain unknown, though several hypotheses have been proposed (Walker, 2017). Thus, from this thermodynamic perspective, live structures subjected to an arbitrarily time varying external driving field (external energy source) can be viewed as dissipative open systems capable of self-replication (Prigogine & Nicolis, 1971; Ragazzon & Prins, 2018). Several basic “close to or at equilibrium” models of live structures under arbitrarily time varying external driving fields have been proposed (Kleidon et al., 2010; Prigogine & Nicolis, 1971; Spanner, 1953). However, the statistical behaviour of far from equilibrium life systems cannot be modelled based on the properties of their individual microstates at one moment in time alone and, thus, these models are limited in the real world. Furthermore, efforts to derive a more general thermodynamic theory of life systems have been hampered by the lack of precise equations for far from equilibrium systems subjected to arbitrarily time varying external driving fields, as these systems operate in a non-linear response regime that is difficult to model using classical statistical thermodynamics (Crooks, 1999; England, 2013). Recent theoretical advances focusing on comparing dynamical trajectories, rather than describing local microstates, have made further progress. By applying time reversal symmetry and coarse grained state transitions concepts, these studies have provided helpful semi-quantitative insights into the thermodynamic constrains that bind the behaviour of such self replicative systems (Bartolotta et al., 2016; Crooks, 1999; England, 2013; Jarzynski, 2011; Pinero & Sole, 2018). The great advantage of these models is that they are derived under general premises and hold even when an arbitrarily time varying external field drives the system in far from equilibrium conditions. Thus, this theoretical framework, generally termed Extended/Bayesian Second Law (ESL) of thermodynamics, allows us to study the directionality of systems subjected to an external driving field that induces changes in the system over time. Complex life biological systems behave in this manner as they are always adapting to changes in their environment that drive their development.

Using these models it is possible to evaluate minimum boundary thermodynamic values that constrain a set of arbitrarily selected initial and final coarse grained states in relation to their trajectory probabilities, equation 1 (eq. 1). In other words, the reversibility of a reaction between two coarse grained states is limited by thermodynamic boundaries determined by the amount of work available to the system and the changes in internal energy and internal entropy between the states (and vice versa). Applying this model to empirically irreversible reactions where the complexity of the system increases in the form of both internal energy (increase) and internal entropy (decrease) shows that these changes in complexity must be accompanied by work made available to the system of equal or higher value to a minimal boundary value set by the irreversibility of the reaction (probability of the forward over the reverse trajectory). Thus, for inherently irreversible reactions such as the SRn of a life biological system, a minimal boundary value of available external work is needed to drive changes in the complexity of the system. This observation had intuitively already been appreciated in the study of complex ecological systems and several efforts to develop a conceptual framework for its thermodynamic analysis have been historically put forward. This has led to the development of concepts such as Minimum Entropy Production, Maximum Entropy Production, Maximum Exergy Production, etc (Jorgensen, 2005; Jorgensen et al., 2007; Kleidon et al., 2010; Meysman & Bruers, 2010; Silow & Mokry, 2010; Spanner, 1953). Furthermore, field studies in complex ecological systems designed to calculate experimental values for thermodynamic-based ecological goal functions have been carried out (Fath et al., 2001; Holdaway et al., 2010; Jorgensen, 2007; Meysman & Bruers, 2010). However, up until now, a satisfactory theoretical framework to incorporate these concepts into a coherent model that can be supported by experimental observations has been lacking, mainly because complex ecological systems operate at the far from equilibrium conditions referred to above and using classical thermodynamics analysis of microstates cannot model these conditions. However,

we have now at our disposal a semi-quantitative thermodynamic theoretical framework that can be used to model the trajectory of life complex systems across time and space. On this work we propose a small set of empirically self evident axioms that allow us to further expand our insight into the thermodynamic behaviour of life complex systems modelled as SR entities using the Extended/Bayesian Second Law of thermodynamics.

Semiquantitative analysis model and methodology

As explained above, the Extended/Bayesian Second Law (ESL) sets minimum reaction thermodynamic boundary values for arbitrary initial and final coarse grained states 1 to 2 (CGS1 to CGS2) based on the probabilities of the forward and time reversed symmetry reactions:

$$\langle W \rangle_{1\to 2} - \langle \Delta E_i \rangle_{1\to 2} + T\Delta S_i \geq K_B T ln(\frac{\pi(1 \to 2)}{\pi(2 \to 1)}) \qquad \text{(Eq. 1)}$$

Or (in this manuscript concerning self replicators: SR)

$$\langle W \rangle_{1\to 2} - \langle \Delta E_i \rangle_{1\to 2} + T\Delta S_i \geq K_B T ln(\frac{P_{SR}}{\overline{P_{SR}}}) \qquad \text{(Eq. 1.1)}$$

Where $\langle W \rangle_{1\to 2}$ represent the average work entering the system from the external driving field, $\langle \Delta E_i \rangle_{1\to 2}$ the average variation in internal energy of the system, $T\Delta S_i$ the change in internal entropy of the system and ($\frac{P_{SR}}{\overline{P_{SR}}}$) the probability of the forward self-replication reaction divided by the probability of the time reversed symmetry reaction. T stand for temperature and K_B for the Boltzmann constant.

This equation was derived under very general conditions and can be applied to a range of settings, including far from equilibrium driven systems (Bartolotta et al., 2016; England, 2013, 2015; Pinero & Sole, 2018). Importantly, the initial and final coarse grained states that define the terms of the equation can be selected arbitrarily to fit our desired conditions. Furthermore, as previously justified (England, 2013), when applying this equation to the process of self-replication (SRn), we can intuitively assume that the probability of the forward self-replication reaction ($P_{1\to 2}$ or P_{SR}) is much higher than the reverse reaction probability ($P_{2\to 1}$ or $\overline{P_{SR}}$), where the SRn process "undoes" itself following a time reversal trajectory. In fact, it has been suggested that the process of SRn is irreversible from a practical point of view (England, 2013; Saadat et al., 2020). Thus, the RHS term of eq. 1.1, $K_B T ln(\frac{P_{SR}}{\overline{P_{SR}}})$, can be assumed to always have a positive value for an established SR system undergoing SRn/growth (figure 1).

In order to further take advantage of these properties we can start by defining a simplistic model of SRn emergence, from a set of n reactants (R_n) that combine themselves in a suitable fashion to generate a self-replicative entity SR_m (figure 2). As addressed in the introduction, many unknowns still remain regarding the initial stage of the reaction where this SR entity emerges de novo (abiogenesis phase, figure 2). Fortunately, even without detailed knowledge of this initial reaction step/s we can extract some useful information from this model. Defining a bona fide SR system as inherently capable of growth when driven by an external field under suitable conditions (thus fulfilling the minimal requirements of life as discussed above), it then follows that the probability of SRn, P_{SR}, must be intrinsically bigger than the reverse SRn probability plus the probability of system decay, $\overline{P_{SR}} + P_{SRdecay}$, under suitable conditions allowing system growth (figure 1). If we further assume that $P_{SRdecay}$ is substantially bigger than $\overline{P_{SR}}$, this is the likelihood of system decay

is much greater than that of the system following the time reversal trajectory of SR, then we can simply assume that $\overline{P_{SR}} + P_{SRdecay} \approx P_{SRdecay}$. Introducing this into eq. 1.1 we obtain a Constrained Extended/Bayesian Second Law (CESL) equation (eq. 2; figure 1 and 2):

$$\langle W \rangle_{1\to2} - \langle \Delta E_i \rangle_{1\to2} + T\Delta S_i \gtrsim K_B T ln(\frac{P_{SR}}{P_{SRdecay}}) \qquad \text{(Eq. 2)}$$

a)

Extended/bayesian second law (ESL)

$$\langle W \rangle_{1\to2} - \langle \Delta E_i \rangle_{1\to2} + T\Delta S_i \geq K_B T ln(\frac{P_{SRm}}{\overline{P_{SRm}}})$$

ESL minimum boundary for SR systems

$$(\frac{P_{SRm}}{\overline{P_{SRm}}}) > 0$$

b)

Constrained extended/bayesian second law (CESL)

Constraints :

$$\overline{P_{SRm}} \ll P_{SRmdecay}$$

$$\overline{P_{SRm}} + P_{SRmdecay} \approx P_{SRmdecay}$$

$$\langle W \rangle_{1\to2} - \langle \Delta E_i \rangle_{1\to2} + T\Delta S_i \gtrsim K_B T ln(\frac{P_{SRm}}{P_{SRmdecay}})$$

Constrained lower minimum boundary

$$(\frac{P_{SRm}}{\overline{P_{SRm}}}) \gtrsim (\frac{P_{SRm}}{P_{SRmdecay}})$$

Figure 1

*Figure 1: ESL **a)** and CESL **b)** equations with their associated minimum boundaries and constraints used in their derivation.*

We can easily see that for our system to grow, P_{SR} needs to be bigger than $P_{SRdecay}$ and therefore, the RHS of Eq.2 would still be positive under this set of constraints, which resets a “new” constrained “lower” minimum boundary for the LHS of the ESL Eq. (Eq. 2 and figure 1).

In fact, it could be argued that from a quantum mechanics probabilistic point of view (Schrödinger, 1967) a bona fide SR system must be able to grow under a suitable set on conditions and this constraint (for $P_{SR} \geq P_{SRdecay}$ under favourable conditions) should be considered an inherent property of SR biological systems. Furthermore, this constraint is consistent with our empirical observation of SR systems in nature at all levels, where system growth is observed when a suitable set of conditions is present, from isolated bacterial growth (such as in a bio-incubator) to a mature complex ecosystem being perpetuated over time (Desmond-Le Quemener & Bouchez, 2014; Jorgensen, 2007; Kleerebezem & Van Loosdrecht, 2010; Kleidon, 2009; Saadat et al., 2020).

From this initial constrained model (eq. 2 and figure 2) it is then simple to see that under a suitable external driving field providing an unlimited supply of energy and given sufficient reactants, we would observe an exponential growing curve for our novel SR system over time modelled by

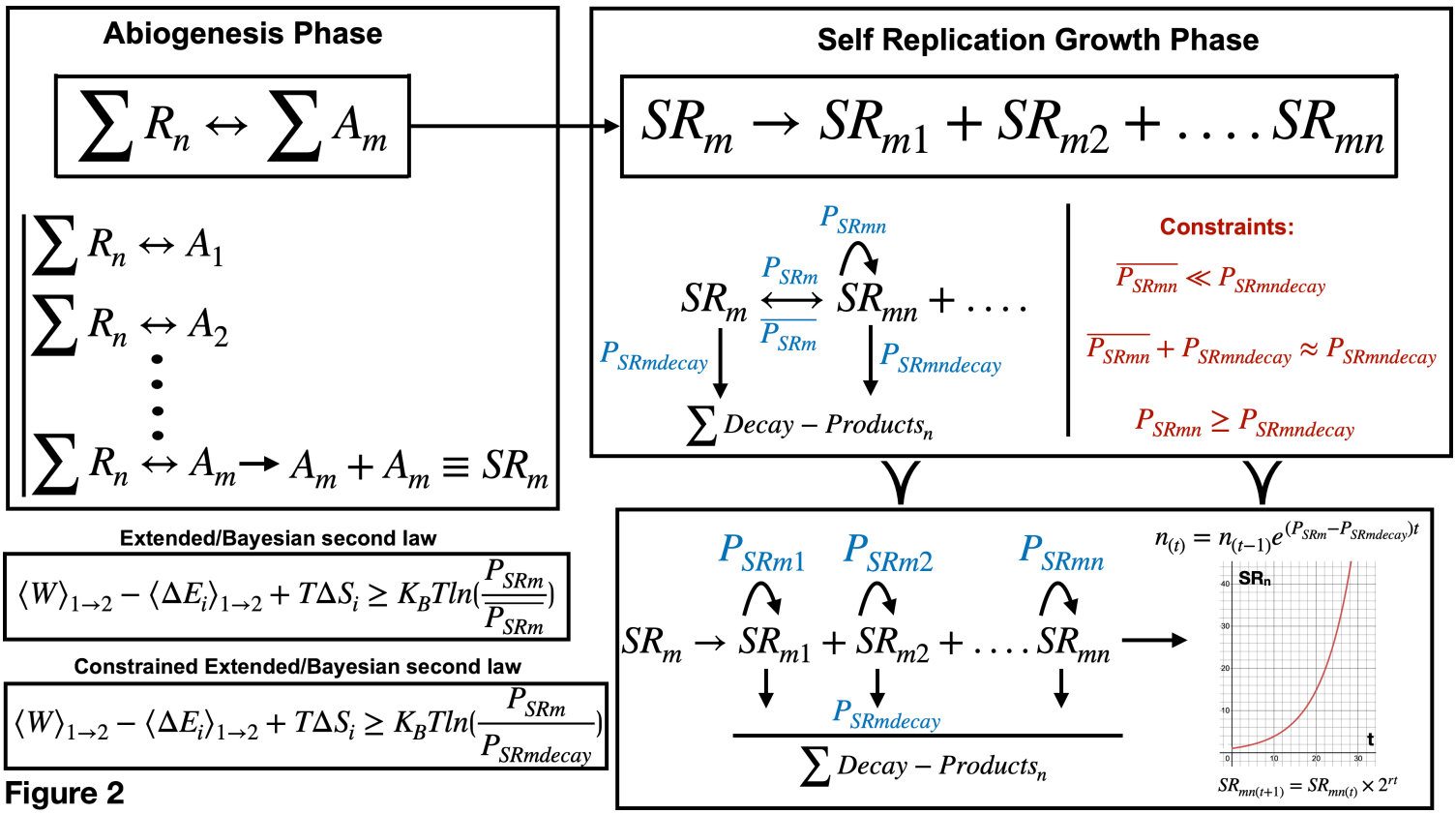


Figure 2

Figure 2: Model of original self replicator (SR_m) emergence from non-self replicative reactants (R_n) (abiogenesis) among the possible reaction assemblies (A_m) and subsequent SR growth modelled under the constraints set by the CESL equation.

$n_{(t)} = n_{(t-1)} e^{(P_{SRm} - P_{SRmdecay})t}$, as already described in England, 2013. This is again consistent with our experimental observation of equivalent life systems. For example, bacterial growth under a suitable environment progresses exponentially while resources (space, food, no competition, etc) are unlimited (figure 2, SR growth phase). Furthermore, bacterial thermodynamic growth models independently yield similar equations to the empirical Monot exponential growth equation, supporting the notion that thermodynamics underlies the growth behaviour of life systems as shown in figure 2 and modelled by $SR_{mn(t+1)} = SR_{mn(t)} \times 2^{rt}$ from (Zeng & Yang, 2020).

If we further analyse this in thermodynamic terms, we can see that if the supply of usable work (W_{usable}) is unlimited and there are not intrinsic growth limiting factors that affect ($\frac{P_{SR}}{P_{SRdecay}}$) (ie: internal factors such as durability vs decay or systemic factors such as competition or lack of resources that impede growth) our system would proceed to increase its internal energy and/or reduce its internal entropy exponentially according to the boundary limits set by eq. 2 under constrained conditions (and eq. 1 under general conditions). It is then self-evident that as long as a bona fide SR system has access to external usable work from the drive it would exponentially increase its complexity (by reducing its internal entropy and increasing its internal energy) in a manner limited only by these inherent thermodynamic constraints. We have now a thermodynamic semi-quantitative model that provides a more rigorous validation to our intuitive understanding of the growth behaviour of complex live SR systems: if “resources” (available work and reactants) and “competition” (inherent restraints within the system) are not limiting then the system would grow in terms of biomass (internal energy) and complexity (entropy) across space and time until some limiting factor/s appears and sets a maximum thermodynamic boundary to this growth,

leading to a constrained pseudo-equilibrium where all available work is used in maintaining the system's complexity far from equilibrium, with no net growth for $P_{SR} = P_{SRdecay}$, leading to

$$\langle \frac{W}{T} \rangle_{1\to 2} \approx \langle \frac{\Delta E_i}{T} \rangle_{1\to 2} - \Delta S_i \qquad \text{(see figure 3).}$$

The above model provides some useful insights, but it remains limited under real world conditions, as it does not account for the evident growth limitations that life systems operate under, nor does it account for variation of the SR assemblies (or other emergent system properties) to allow for evolution as observed in nature. In order to shed further light on the behaviour of SR life systems at varying levels of complexity (from a single bacterium to whole ecosystems) and model system directionality over space and time we need to take advantage of the capability to arbitrarily set the defining initial and final coarse grained system states in the CESL equation:

First of all, we need to acknowledge that in complex biological systems the constrained probability of the system $\frac{P_{SRm}}{P_{SRdecay}}$ may not always be positive and lead to exponential growth as considered so far in our constrained system (figure 2). Within the context of CESL, $\frac{P_{SRm}}{P_{SRdecay}}$ can take (see figure 3):

- positive value leading to exponential growth of the system and an increase in complexity due to an increase in internal energy and/or a decrease in internal entropy ($-\Delta E_i + \Delta S_i$);
- a negative value leading to exponential decay of the system and a decrease in complexity due to a decrease in internal energy and/or an increase in internal entropy ($-\Delta E_i + \Delta S_i$);
- be equal to 1, leading to a constrained driven pseudo-equilibrium where the available W is used in maintaining the system complexity ($E_i + S_i$) far from the thermodynamic equilibrium understood in the classical sense.

Constrained extended/bayesian second law (CESL)

$$\langle W \rangle_{1\to 2} - \langle \Delta E_i \rangle_{1\to 2} + T\Delta S_i \gtrsim K_B T ln(\frac{P_{SRm}}{P_{SRmdecay}})$$

Figure 3

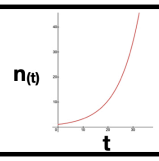
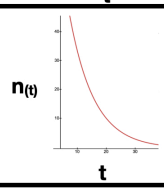
Constrained lower minimum boundary	$\langle W \rangle_{1\to 2}$	$-\langle \Delta E_i \rangle_{1\to 2}$	$T\Delta S_i$	$n_{(t)} = n_{(t-1)} e^{(P_{SRm} - P_{SRmdecay})t}$
$(\frac{P_{SRm}}{P_{SRmdecay}}) > 0$	+	-	-	
$(\frac{P_{SRm}}{P_{SRmdecay}}) < 0$	-	+	+	
$P_{SRm} = P_{SRmdecay}$ $(\frac{P_{SRm}}{P_{SRmdecay}}) = 1$	$\langle \frac{W}{T} \rangle_{1\to 2} - \langle \frac{\Delta E_i}{T} \rangle_{1\to 2} + \Delta S_i \approx 0$ **Constrained pseudo-equilibrium (no net growth) for:** $\langle \frac{W}{T} \rangle_{1\to 2} \approx \langle \frac{\Delta E_i}{T} \rangle_{1\to 2} - \Delta S_i$			

Figure 3: Possible values of $\langle W \rangle_{1\to 2}$, $\langle \Delta E_i \rangle_{1\to 2}$ and $T\Delta S_i$ during growth, decay or pseudo-equilibrium evolution of a system with their associated minimum boundaries as constrained by the CESL equation.

For the purpose of further analysis, we can define the initial and final coarse grained states to include all individual reactants and SR assemblies within the initial and final life system respectively in a hierarchical configuration. For example, in the case of a single bacterium replication the initial coarse grained state would be composed of the bacterium plus all the reactants needed to facilitate SRn and the final state composed by the two resulting bacteria plus any generated waste products (figure 4) (England, 2013). Likewise, in a more complex system transition, such as the maturation process of an ecosystem from grassland to forest, all the necessary reactants would be included into the initial coarse grained state together with all the individual SR entities present (the grassland ecosystem as a whole in this case, including bacteria, decomposing organisms, plants, animals, etc) and the final state would also include any unused reactants, waste products and all the final SR entities present (the forest ecosystem as a whole including all components as before) (figure 4). The novelty here is that we are taking the whole system as defining initial and final coarse grained states to include the trajectories of all particles involved in the process, and considering the individual SR entities (bacteria, grass, tree, etc) as organised in a hierarchical configuration of increasing complexity levels corresponding to the historical development (growth) of the system. Therefore, we are no longer exclusively defining the reaction by SR constraints understood in a classical sense (such as replication of one bacterium into two bacteria), as the initial and final coarse grained state SR entities may differ greatly due to the hierarchical progression of the system (in the above example during the transition from grassland to forest clearly a new layer of system complexity is added). Importantly however, the irreversibility of the self-replication process of individual SR entities is still the driving force of the system as a whole as per our model described above (figure 2) and we are simply expanding the initial and final states beyond the SR reactions of individual entities within the system in an integrative manner that includes all particles within the system over space and time, while still imposing this irreversibility to the coarse grained state transition as a whole. This overall irreversibility restriction to the reaction is clearly as intuitively justified for the SRn of a single bacterium (England, 2013) as it is for the “SRn” (as a coarse grained state transition as considered here) of a complex system of grassland maturing into forest ecosystem. If anything it could be argued that a forest ecosystem is less likely to spontaneously revert to grassland following the time reverse trajectory with an even lower probability due to its higher complexity (Jorgensen, 2007). Importantly, considering the system in this manner allows us to include variation of the individual SR entities leading to adaptations and to the evolution of the overall SR system within the constraints of the CESL model. Thus, the ($\frac{P_{SR}}{P_{SRdecay}}$) term refers to the “replication” (CGS1 to CGS2 transition) of the whole hierarchical complex system over time taking into account changes to the individual SR entities that allow for changes leading to evolution (for example the progression from grassland to forest ecosystem as described above). Furthermore, as we have stablished that the system’s total components N_{total} (n reactants, N_r; n waste products, N_w; and n SR entities, SR_m) must remain constant between CGS1 and CGS2, with:

$$N_{total} = N_r + N_w + SR_m \qquad \text{and}$$

$$N_{total1} = N_{total2} = (N_{r2} - N_{r1}) + (N_{w2} - N_{w1}) + (SR_{m2} - SR_{m1})$$

this allows us to evaluate changes in the extrinsic terms of CESL in a semi quantitative manner as N_{total} remains constant from CGS1 to CGS2.

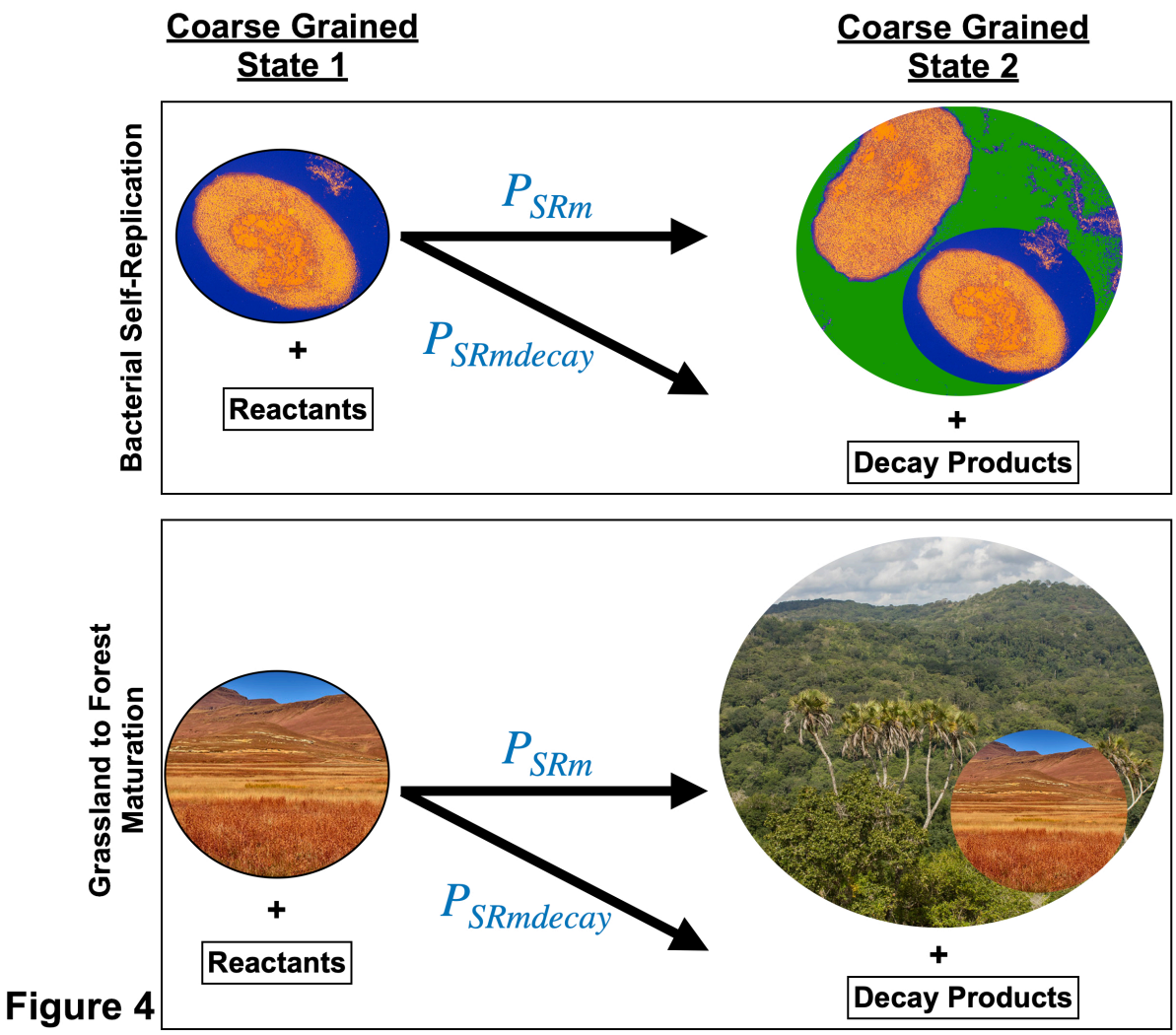


Figure 4

Figure 4: Model of the hierarchical evolution of a complex life system over time, taking into account changes to the systems boundaries to allow for variation leading to novel adaptations during the transition from a Coarse Grained State 1 to a Coarse Grained State 2 and driven by the growth of the self replicator (SR) entities contained within the system at the different stages. As an example of two such systems the self-replication of a bacterium into two bacteria (top panel) and the self-replication (growth) driven maturation of a grassland into forest ecosystem (lower panel) are represented. Of note, the initial SR entities contained in the system at Coarse Grained State 1 are incorporated hierarchically within the system at Coarse Grained State 2.

Thus, we can now consider the evolution of the complexity of the system ($-\Delta E_i + \Delta S_i$) within this context. In a hierarchical biological system growing from a simpler coarse grained state 1 (CGS1) (i.e. grassland) to a more complex coarse grained state 2 (CGS2) (i.e. forest) we would expect this combined term to decrease, as the system increases its E_i by accumulating biomass and decreases its S_i as it becomes more complex in terms of the number of component connections encompassed within the system (of note biomass has a higher intrinsic E_i and lower S_i than the inorganic reactants it comes from) (Jorgensen, 2007). As all reactants (and derived waste products) are accounted within the CGS1 to CGS2 transition in a hierarchical manner as described above we can calculate the directionality of change for the system's ($-\Delta E_i + \Delta S_i$) from CGS1 to CGS2. This approach has been intuitively used in the ecological sciences field (Jorgensen, 2007; Kleidon et al., 2010), but we can now describe it formally in terms of precise thermodynamic

equations. Thus this is a valuable approach as it allows us to study the term $(-\Delta E_i + \Delta S_i)$ in the CESL model in a semiquantitative manner within the context of hierarchical biological systems even if we lack exact values for these terms.Therefore, it would be helpful to be able to empirically evaluate the individual terms of the CESL eq., at least in a semi quantitative manner that allows to explore the directionality of the reaction. To do so we can consider the following analysis of the individual CESL terms:

The term $-\langle \Delta E_i \rangle_{1\to 2}$ represents the change in internal energy from a CGS1 to a CGS2. We can simply approximate the internal energy of a system to the chemical energy resulting from the addition of the chemical potential of n components of the systems (Jorgensen, 2007; Wu et al., 2017). It is challenging to calculate an exact value for the internal energy of a complex system as it is technically difficult to measure the chemical potential of all its components. However, several directly proportional indicators have been used to calculate approximate values for a system's internal energy (biomass, exergy, etc). By using these indicators we should able to assess the value of $-\langle \Delta E_i \rangle_{1\to 2}$ in a semiquantitative manner that allows us to at least assign directionally to this change. We will focus hereafter on measurements of biomass, as these have been widely used and there is a wealth of data available. However, other directly proportional indicators such as exergy should provide equivalent results regarding the directionality of change in $-\langle \Delta E_i \rangle_{1\to 2}$ (Fath et al., 2001). Biomass is clearly directly related to internal energy of a system resulting from its chemical potential energy and therefore we can assume that:

$$\langle \Delta E_i \rangle_{1\to 2} \simeq (\mu_2 N_2 - \mu_1 N_1) \simeq \alpha(\Delta biomass)$$

where u_1 and u_2 are the chemical potentials of the system, N_i the number of chemical compounds and α a specific constant that relates the biomass of a system to its chemical potential energy. If we assume that α remains constant for CGS1 and CGS2 we can then assume that an increase in total biomass from CGS1 to CGS2 would result in a directly proportional increase in its internal energy.

A similar analysis regarding the variation in internal entropy of the system,ΔS_i, can also be put forward. The internal entropy term in CESL refers to the complexity of the system from a classical statistical mechanics point of view, thus relating the microstates probability distribution to the macroscopic system properties (in other words relating to the number n of components within the system and how these n elements are interconnected to one another). Measuring precisely the entropy changes in a complex system over time remains challenging. However, as above, we can calculate changes in entropy in a semiquantitative manner that allows to at least explore the directionality of the reaction. For this we can take advantage of several ecological complexity orientators that are directly correlated to the internal entropy in a system such as ascendency, average mutual information (AMI), etc, and generally giving a measure of how well energy flows are constrained within the system network and therefore reflecting the overall constrained network complexity as compared to the remaining degrees of freedom available to the system. This approach has been used extensively within the network analysis field in ecology to evaluate system complexity and thus we can take advantage of previous experimental work. Importantly, recent work seeking to understand the complementarity of these orientators as ecological goal functions has shown that they are mutually consistent suggesting a common pattern in the network organisation of complex life systems (Fath et al., 2001). Thus measurement of Ascendency, AMI or simply the number of total biological species within the system, would yield equivalent semiquantitative results for the purpose of our analysis.Thus we can these measurements as a proxy indicator for the entropy evolution of a complex system, at least in a semiquantitative way. As these indicators are inversely correlated with ΔS_i an increase in Ascendency, AMI or total number off species from CGS1 to CGS2 would imply a decrease in S_i (or increase in complexity) as described here.

Finally we can focus on the term reflecting the available work (energy) entering the system, $\langle W \rangle_{1\to 2}$ from the external drive. Within the frame of the CESL model for the earth system we can assume that maximum work value limits are dependent of the capability of the system to input usable work from the external drive and this sets an upper limit for system growth and increase complexity ($-\Delta E_i + \Delta S_i$) for a given ($\frac{P_{SR}}{P_{SRdecay}}$) value. In other words, not all the energy available from the universe can enter the system due to its intrinsic limitations in terms of processing energy from the external drive. This can be clearly seen if we compared a non-photosynthetic system, where only chemical energy can be accessed as usable work, and a photosynthetic system, where a new energy level access directly from solar radiation can be used as work. We have then that:

$$\langle W \rangle_{1\to 2} = \langle W \rangle_{usable} \leqq \langle W \rangle_{universe}$$

and

$$\langle W \rangle_{usable} + \langle W \rangle_{unused} = \langle W \rangle_{universe}$$

We can clearly see that for the system to be able to overcome its upper bound complexity threshold set by CESL it must then be able to access higher levels of usable work/energy from the universe, as defined by the minimum boundaries set by CESL for irreversible SR systems (figure 1 and 3). This can be achieved through hierarchical variation in the properties of the system in a relevant evolutionary timescale and multiple examples of this emergence of new levels of usable work underpinned the evolution of life on hearth (such as in the evolution of photosynthetic SR organisms exemplified above). Furthermore, the more work that can be inputted into the system from the universe the lower the unused or reflected work $\langle W \rangle_{unused}$ would be, and this can be seen as a lower albedo complexity increases in life systems. The example of grassland vs forest ecosystem is appropriate here again, with a more complex forest ecosystem generating a lower albedo that a grassland ecosystem, which in turn also generates a lower albedo than an even simpler dessert ecosystem (Holdaway et al., 2010; Schneider et al.,1994). The amount of usable work available can be more readily measured for life systems and we can take advantage of this in our analysis. Multiple studies relating the capacity of complex ecosystems in terms of their usable work/reflected work (and accompanying indicators such as emergy) have been performed (Jorgensen, 2007) and we can use this data in our analysis as seen below.

Results

We can move forward from here by applying these set of empirically derived constrains to the analysis of complex live biological systems using the CESL equation. We can begin by analysing data derived from complex ecological systems where detailed measurements of relevant parameters allows us to analyse in a semi-quantitative manner changes in the CESL equation terms.

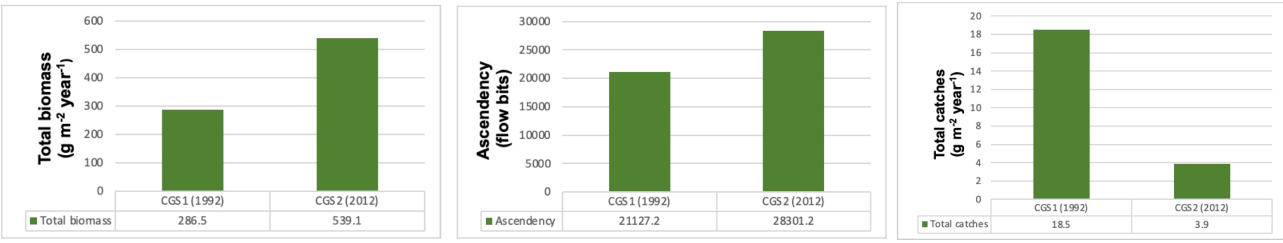
In González et al., 2016, the authors study changes in the Tongoy Bay ecosystem using trophic network analysis. Tongoy Bay ecosystem was monitored over time from an initial CGS1 (1992), prior to which the ecosystem had suffered from anthropogenic over-exploitation, to a CGS2 (2012) where this anthropogenic disturbance had over time been ameliorated. The authors study changes in ecosystem structure and function using trophic network analysis to evaluate changes in macroscopic indexes related to the ecosystem's health. Among these indexes we have data for biomass and ascendancy changes that we can input as described above into the CESL equation. Upon amelioration of the anthropogenic disturbance of the ecosystem, the authors describe an increase in the overall health, the total biomass and the interconnectivity (complexity) of the ecosystem by using several ecological indicators, showing that the system has experience a period of general recovery and growth from CGS1 to CGS2. Specifically, Ascendency (and therefore $-S_i$) and biomass (and therefore E_i) increased during this period (figure 5a). We can

also assume that ($\frac{P_{SR}}{P_{SRdecay}}$) takes a positive value (figure 3) from CGS1 to CGS2, concomitant to a decrease in anthropogenic over-exploitation (this is reflected in reduced catch numbers on this period that would have resulted in an increased overall survival), allowing for growth and establishing a positive thermodynamic minimum boundary according to CESL for the system. It then follows that the amount of work used by the system must have also increased in order to compensate for the increased in $-S_i$, E_i and positive ($\frac{P_{SR}}{P_{SRdecay}}$), following the thermodynamic constrains set by CESL (figure 3). As no fundamental changes in the available usable work, $\langle W \rangle_{usable}$, input into the system from the external drive are observed (such as the emergence of new properties like the photosynthesis example described above), therefore we must conclude that this increase in work uptake must be inherent to the reported growth of SR entities capable of utilising work within it. This can be intuitively expected as higher biomass should be able to process more work from the external driving field. Thus, under the optics of CESL analysis we can conclude that the Tongoy Bay ecosystem had been kept under growth constraining conditions, anthropogenic over-exploitation reducing ($\frac{P_{SR}}{P_{SRdecay}}$), resulting in a pseudo-equilibrium upon which the total amount of work available to the system was not fully utilised due to this growth limitation. Upon removal of this limitation the system is then able to grow and uptake more work with a concomitant increase its overall complexity ($\Delta E_i + \Delta S_i$).

In a separate study a manganese mine contaminated wasteland that had undergone extensive plant biomass and species number depletion as compared to equivalent uncontaminated land was studied (Wu et al., 2017, Wu et al., 2018). Land at this contaminated wasteland showed high metal concentrations and low nitrogen and phosphorus levels. In an effort to establish a strategy for the ecological restoration of such ecosystems, experimental plots were modified by planting suitable wood plant species treated with enriched manure containing metal resistant bacteria or a chemical fertiliser only (containing equivalent levels of nitrogen, phosphorus and potassium) as a control. Plots were then monitored for 5 years (2010-2015). Total biomass, total number of plant species and total Gibbs free energy calculations for the plots were performed over the study period. These measurements showed a striking restoration of the ecosystem under the enriched manure treatment condition compared to the control (figure 5b). As above we can used CESL to model the changes in work and overall complexity in the system during this period using available data. As shown in figure 5b the total biomass and Gibbs free energy representing the overall E_i of the system markedly increased over the experiment timeframe. Similarly, there was a drastic increased in the ecosystem complexity ($-S_i$) as shown by the increase in total plant species found compared to the chemical fertiliser internal control and approaching numbers similar to uncontaminated land. By artificially introducing new SR entities capable of growth within the system (wood plant species treated with enriched manure containing metal resistant bacteria) we have a situation similar from the point of view of CESL to the model shown in figure 2 relating to the emergence of the first SR entities during abiogenesis on hearth. We transition from a CGS1 with limited or null growth (in this case due to soil toxicity and poor nutrient availability limiting growth of existing SR entities) to a more complex CGS2 due to artificial introduction in the system of novel SR entities capable of undergoing rapid growth thanks to their intrinsic properties (resistance to metals and available nutrients provided with the enriched manure). As expected these new SR entities drive growth of the system and lead to an increase in total E_i and a decrease in S_i concomitant with an increased in the total $\langle W \rangle_{usable}$ used by the system (as above due to the increase in biomass capable of taking work from the external drive). Thus, by reducing metal toxicity and providing additional nutrient resources that lead respectively to a positive ($\frac{P_{SR}}{P_{SRdecay}}$) and allow an increase in E_i we see an accompanying increase in $-S_i$ and $\langle W \rangle_{usable}$ uptake inherent to the system's growth.

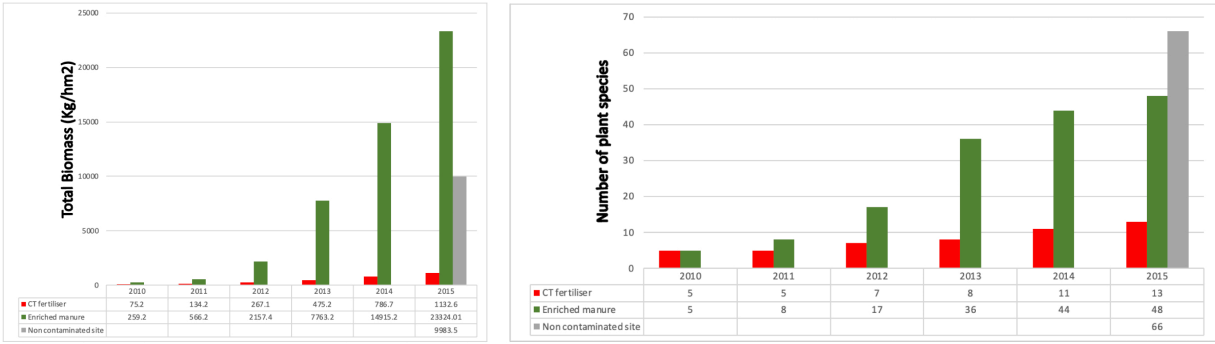
Interestingly, we can see here in an controlled experimental set-up that under the light of CESL the systems progresses to higher levels of complexity as changes within its structure allow it to further exploit improved growth conditions in a positive feed-back loop (further reduction of toxic metal in the soil and enrichment of nutrient availability) driven by the initial appearance of the novel SR entities, in an analogous manner to the process of complex ecological systems maturation during relevant historical evolutionary periods.

a)



Constrained lower minimum boundary	$\langle W \rangle_{1\to2}$	$-\langle \Delta E_i \rangle_{1\to2}$	$T\Delta S_i$	$n_{(t)} = n_{(t-1)} e^{(P_{SRm} - P_{SRmdecay})t}$
$(\frac{P_{SRm}}{P_{SRmdecay}}) > 0$	$\uparrow$	252.6	-6904.0	

b)



Constrained lower minimum boundary	$\langle W \rangle_{1\to2}$	$-\langle \Delta E_i \rangle_{1\to2}$	$T\Delta S_i$	$n_{(t)} = n_{(t-1)} e^{(P_{SRm} - P_{SRmdecay})t}$
$(\frac{P_{SRm}}{P_{SRmdecay}}) > 0$	$\uparrow$	$C: 1057.4$ $M: 23064.8$	$C: -8$ $M: -46$	

Figure 5

Figure 5: Studying complex ecosystems by using trophic network analysis to evaluate changes in macroscopic indexes related to the ecosystem's thermodynamic evolution as viewed under the CESL equation . In a) the evolution of the Tongoy Bay ecosystem was studied from CGS1 (year 1992) to CGS2 (year 2012) and changes in total biomass (top left graph), ascendency (top middle graph) and total catch (top right graph) were measured and used as directly proportional indicators for $\langle \Delta E_i \rangle_{1\to2}$, $T\Delta S_i$ *and* $(\frac{P_{SR}}{P_{SRdecay}})$ *respectively, within the terms of the CESL equation and the corresponding system growth equation (bottom panel), as justified in main text above. In b) the evolution of the ecological restoration at a manganese mine contaminated wasteland ecosystem was studied from CGS1 (year 2010) to CGS2 (year 2015) and changes in total biomass (top left graph) and number of plant species (top right graph) were measured and used as directly proportional indicators for* $\langle \Delta E_i \rangle_{1\to2}$ *and* $T\Delta S_i$ *respectively, within the terms of theCESL equation and the corresponding system growth equation (bottom panel), as justified in main text above. A restoration strategy where experimental plots were modified by planting suitable wood plant species treated with enriched manure containing metal resistant bacteria was used (green columns). An equivalent control fertiliser only (red columns) and a control non-contaminated site (grey column) are also included in the analysis.*

In order to further move forward from here we can design a thought experiment where we analyse the thermodynamic progression of a hypothetical SR biological system from evolutionary relevant coarse grained states 1 (CGS1) to 2 (CGS2) in terms of the CESL equation.

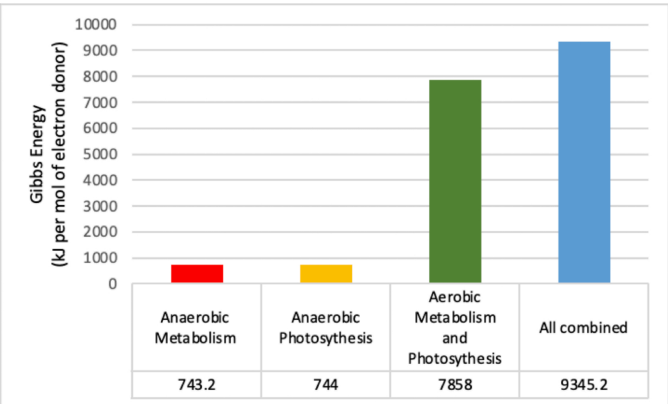
If we define our CGS1 and CGS2 to be evolutionary relevant states for live SR systems during early earth history, we can study how changes in usable work levels relate to changes in complexity, $(-\Delta E_i + \Delta S_i)$, and growth, $(\frac{P_{SR}}{P_{SRdecay}})$, over evolutionary relevant time frames.

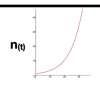
For a way forward let us assume that in our SR system usable work can be accessed from anaerobic chemical reactions, anaerobic photosynthesis, aerobic photosynthesis and aerobic chemical reactions (figure 6a). We can define our CGS1 as a system where usable work only enters the system via anaerobic bacteria capable of anaerobic chemical reactions. This results in a total amount of usable work at CGS of level 1, or W_{usable_1}, under anaerobic conditions such as those present on earth until approximately 2.5 billion years ago with low oxygen levels $(pO_2 < 0.1\,\%\,PAL)$ (*PAL*: present atmospheric level) (figure 6b and 6c). Let us further assume that our SR system in CGS1 undergoes intrinsic changes in an evolutionary relevant time frame that result on the emergence of novel properties allowing the system to access work directly from sunlight due to the appearance of first anaerobic and second aerobic photosynthesis, leading to a global oxygenation event (GOE) and a rise in oxygen levels to at least $pO_2 : 1 - 10\,\%\,PAL$. This in turn allows for aerobic chemical reactions and the complete emergence of aerobic bacteria. This final state corresponds to CGS2. It is important to remember here that we have defined CGS as organised in a hierarchical configuration of increasing complexity levels corresponding to the historical development of the system. Thus, CGS2 also includes all previous relevant modes of accessing work (anaerobic chemical reactions plus anaerobic Photosynthesis, aerobic Photosynthesis and aerobic chemical reactions) that in combination define the usable work at CGS level 2, or W_{usable_2}. Based in the available energy from these modes of accessing work we can calculate total W_{usable} per reaction (kJ per mol of electron donor) at CGS level 1 and 2 (figure 6a and 6b). As expected, $W_{usable_2} > W_{usable_1}$, as CGS2 includes the available work at CGS1

a)

Anaerobic Metabolism	**Ethanol.** $CH_3CH_2OH + 0.85SO_4^{2-} \rightarrow 0.65CH_3COO^- + 0.7HCO_3^- + 0.85HS^- + H_2O + 0.5H^+$	**Lactate.** $CH_3CHOHCOO^- + 0.5SO_4^{2-} \rightarrow CH_3COO^- + HCO_3^- + 0.5HS^- + 0.5H^+$	**Propionate.** $CH_3CH_2COO^- + 1.75SO_4^{2-} \rightarrow 3HCO_3^- + 1.75HS^- + 0.25H^+$	**Acetate.** $CH_3COO^- + SO_4^{2-} \rightarrow 2HCO_3^- + HS^-$	**Hidrogen.** $H_2(aq) + 0.25SO_4^{2-} + 0.25H^+ \rightarrow 0.25HS^- + H_2O$	**Glucose.** $C_6H_{12}O_6 + 2.4H_2O \rightarrow 0.7CH_3CH_2CH_2COO^- + 0.6CH_3COO^- + 2HCO_3^- + 2.6H_2(aq) + 3.3H^+$	**Anae. Photo.** $6CO_2 + 12H_2O \longrightarrow C_6H_{12}O_6 + 6H_2O + 6O_2$
Gibbs E (kJ per mol electron donor)	95.7	100	118	64.7	38.8	326	744
Aerobic Metabolism	**Ethanol.** CH3CH2OH + 3O2(aq) → 2HCO3- + H2O + 2H+	**Lactate.** CH3CHOHCOO- + 3O2(aq) → 3HCO3- + 2H++	**Propionate.** CH3CH2COO- + 3.5O2(aq) → 3HCO3- + 2H+	**Acetate.** CH3COO- + 2O2(aq) → 2HCO3- + H+		**Glucose.** $C_6H_{12}O_6 + 4O_2 \rightarrow 6HCO_3^- + 6H^+$	**Aerob. Photo.** 6CO2 +12H2O → C6H12O6 +6H2O+6O2
Gibbs E (kJ per mol electron donor)	1308	1333	1487	847		2883	2883

b)



Constrained lower minimum boundary	$\langle W \rangle_{1\rightarrow 2}$	$-\langle \Delta E_i \rangle_{1\rightarrow 2}$	$T\Delta S_i$	$n_{(t)} = n_{(t-1)} e^{(P_{SRm} - P_{SRmdecay})t}$
$(\frac{P_{SRm}}{P_{SRmdecay}}) > 0$	$\Delta 10X$	–	–	

c)

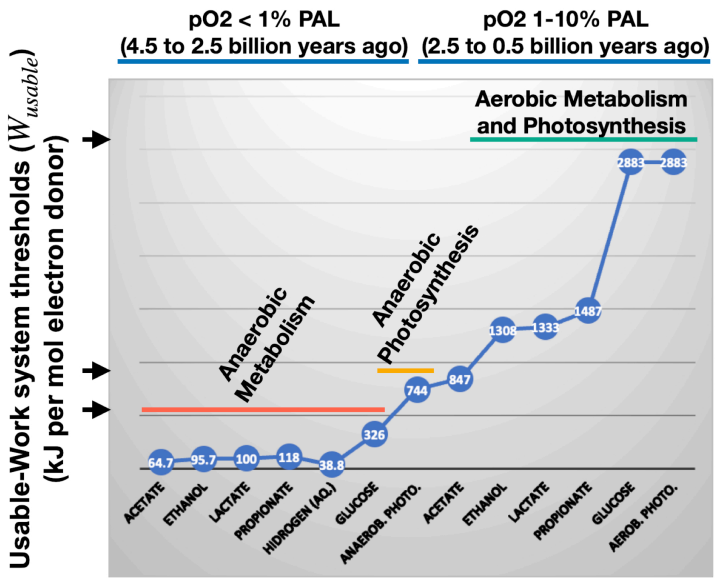


Figure 6

*Figure 6: A thought experiment where we analyse the hypothetical thermodynamic progression in terms of the CESL equation of a complex prokaryotic biological system from evolutionary relevant to earth's early life history CGS1 to CGS2. In **a)** the free energy (kJ per mol of electron donor) of evolutionary relevant prokaryotic aerobic, anaerobic (respiratory and/or fermentative) metabolism of glucose, ethanol, formate, acetate, lactate, propionate and H_2 and aerobic and anaerobic photosynthesis reactions are shown. In **b)** the free energy values for combined anaerobic and/or aerobic reactions using data form a) as described is calculated. For example, the anaerobic metabolism column in red adds the value all the possible non-photosynthetic anaerobic reactions included in a) as a representation of the possible relevant anaerobic modes of energy access. Combined values for anaerobic photosynthesis in yellow, aerobic metabolism and photosynthesis in green and total combined values in blue are calculated equivalently. In **c)** changes in usable-work system thresholds (i.e. total amount of W_{usable} that can be accessed via the free energy generated from anaerobic and/or aerobic reactions) relating to prokaryotic evolution and correlated levels of atmospheric oxygen during early history of life on earth are represented.*

(anaerobic chemical reactions) plus the work accessible by the new work modes (anaerobic Photosynthesis, aerobic Photosynthesis and aerobic chemical reactions) (figure 6a, 6b and 6c) (Heijn et al., 1992; Roden et al., 2011; McCarty, 2007) . We have thus moved from W_{usable_1} to W_{usable_2} following a relevant evolutionary time frame of events (figure 6c) (Catling et al., 2005; Sessions et al., 2009; Soo et al., 2017; Yamamoto et al., 2011). We do not know the total exact values for W_{usable_1} and W_{usable_2}, as we lack historical data regarding exact biomass changes from CGS1 to CGS2. However, it is safe to assume that with the advent of photosynthesis organisms would have colonised new ecosystems and the overall biomass would have been greatly increased during the equivalent evolutionary historical period (from the origin of anaerobic to photosynthetic/aerobic prokaryotes). Importantly, calculations based on historical records are in agreement with this assumption (Battistuzzi et al., 2004; Catling et al., 2005; Payne et al., 2009).

Therefore, we can safely assume that:

1) $W_{usable_2} > W_{usable_1}$ (to which data from figure 6b and 6c is an underestimate as it only relates to intrinsic access to usable work per reaction type, but it doesn't account for total biomass increases) and thus $\langle W \rangle_{1\rightarrow 2} > 0$.

2) We can also safely assume that during this evolutionary relevant period $K_B T ln \frac{P_{SRm}}{P_{SRm}} \gg 0$ and $K_B T ln(\frac{P_{SRm}}{P_{SRmdecay}}) > 0$, as we have multiple evidence of new organism types colonising newly available ecosystems due the novel evolutionary solutions that appeared in this time with an inherent positive growth to our system (Catling et al., 2005; Hamilton et al., 2016; Hsia et al., 2013; Payne et al., 2009).

It then follows that an increase in the total complexity of the system ($-\Delta E_i + \Delta S_i$) from CGS1 to CGS2 would be permitted under the macroscopic constraints imposed by CESL. This is in agreement with empirical observations of this period. As already explained above, ΔE_i should increase in line with the expected increase in total biomass from CGS1 to CGS2. Likewise, ΔS_i should decrease accordingly with the complexity of the system as more levels are added to its hierarchical structure. This is self-evident from the evolutionary history followed in hearth during the period that we have chosen to model. In CGS1 we only have anaerobic prokaryotes ecosystems, limited to local chemical gradients that allow for anaerobic chemical reactions maintaining the system (figure 7). These could be loosely modelled after modern geothermal vent ecosystems (Yamamoto et al., 2011). Under this pseudo-equilibrium state a maximum work value limit,

dependent of the capability of the system to input usable work from the external drive, sets an upper limit for system growth and increase complexity according to CESL (figure 7).

During an initial pseudo-equilibrium state with no net growth $(\frac{P_{SRm}}{P_{SRmdecay}}) = 1$, CGS1 cycles itself in a maintenance loop with $\langle\frac{W}{T}\rangle_{1\to2} \approx \langle\frac{\Delta E_i}{T}\rangle_{1\to2} - \Delta S_i$ and thus most of the usable work is used in maintaining the system overall complexity $(-\Delta E_i + \Delta S_i)$ far from equilibrium (pseudo-closed geothermal vent like ecosystems) . In order for the system undergo changes in complexity it must be able to access more usable work from the universe. This can be achieved via the development of novel intrinsic adaptations that allow to access a higher level of work, such as being able to perform photosynthesis or aerobic respiration (figure 7). Of note here the individual SR entities (prokaryotes in this simplistic model) may not change their intrinsic complexity substantially themselves (an anaerobic prokaryote vs an evolutionary related aerobic photosynthetic prokaryote), however the complexity reflected in CESL refers to the overall system in a hierarchical manner, which clearly increases its complexity greatly in both terms: its internal energy and its internal entropy $(-\Delta E_i + \Delta S_i)$, comparing locally isolated anaerobic bacterial communities vs aerobic photosynthetic communities widespread throughout the planetary surface (figure 7) (Battistuzzi et al., 2004; Hamilton et al., 2016; Hsia et al., 2013; Payne et al., 2009). Thus, the system can input a higher level of usable work that can be used in a phase of exponential growth to offset the cost of its increase complexity according to the minimum boundary imposed by CESL (figure 7).

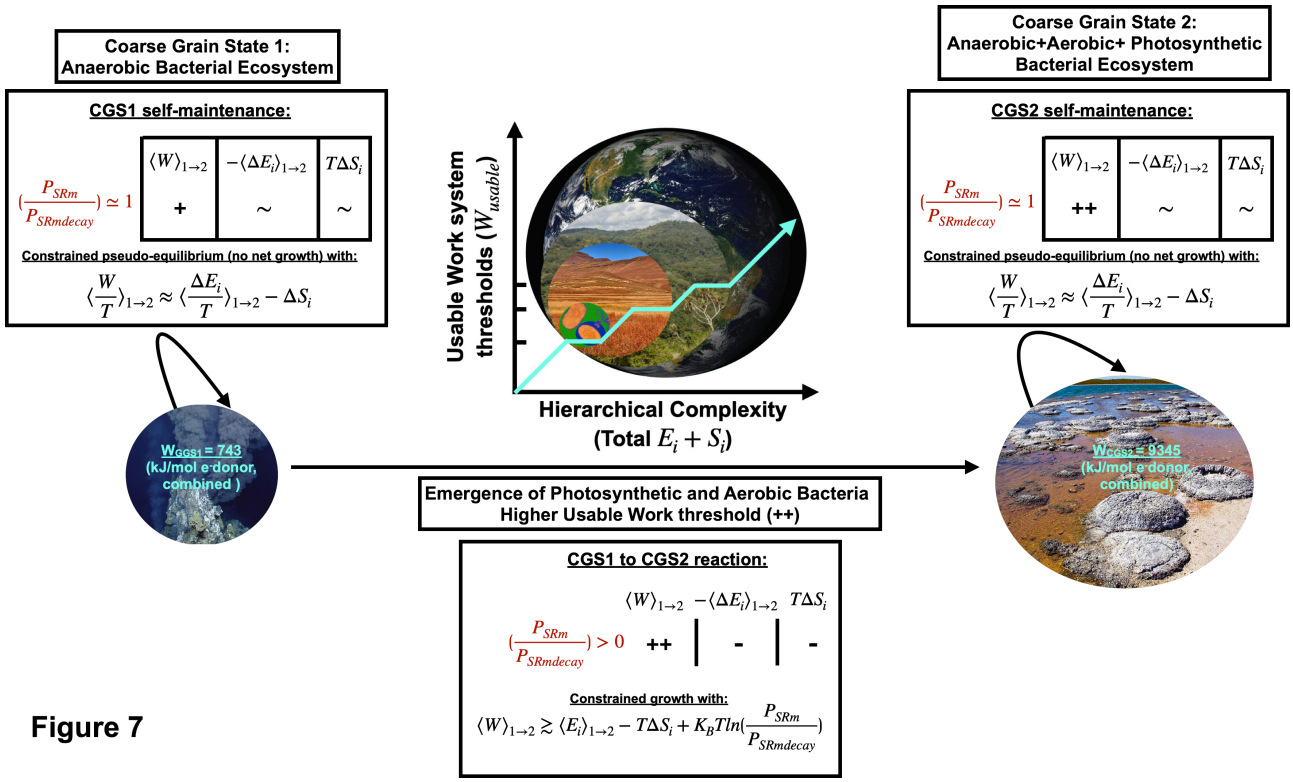


Figure 7

Figure 7: A complete model of the thermodynamic progression of a complex prokaryotic biological system from evolutionary relevant CGS1 (anaerobic) to CGS2 (aerobic and anaerobic). The system would adopt a thermodynamic hierarchical organisation with increasing levels of complexity over the modelled evolution time as higher levels of energy (higher usable work thresholds) are accessed, according to minimum boundary constraints set by the CESL equation.

Discussion

We have shown how using data from complex ecosystem modelling we can apply the CESL equation to study in a semi-quantitative manner thermodynamic changes that underpin the system's evolution. Likewise, we have shown how this analysis can be extended to model relevant evolutionary ecological timeframes of life on earth in a semi-quantitative manner, even in the absence of accurate experimental data. These results agree with current experimental observations (González et al., 2016; Holdaway et al., 2010; Schneider et al.,1994; Wu et al., 2017; Wu et al., 2018) and with our intuitive understanding of the organisation and evolution of life systems on earth (Jorgensen, 2007; Kleidon, 2009; Prigogine & Nicolis, 1971) and lead us to propose that self-replicative life systems are evolutionary and spatially organised in hierarchical thermodynamic levels of incremental work usage inputted from the external driving field (figure 7). Importantly, we have precise equations that allow us to rigorously explore this model from a semi-quantitative point of view as described in the CESL equation.

The CESL constrained model provides valuable insights regarding the initial emergence of SR systems as well. As we have seen this model supports that, in the presence of an external driving field and under suitable conditions allowing growth, SR entities would self-replicate growing exponentially according to the constrains set by Eq. 1.1 (figure 2). It has already been hypothesised that from a theoretical thermodynamic analysis point of view biological systems are constrained to evolve favouring selection of trajectories that lead to configurations that maximise positive values of work; in other words, trajectories leading to configurations of the system that are more capable of inputting the work available to them from the external drive (England, 2015; Prigogine & Nicolis, 1971). Importantly, our approach allows us to explore this hypothesis by arbitrarily setting the initial and final coarse grained states to study. Thus, we can define the initial state (CGS1) as the set of n reactants and the final state (CGS2) as the emergent m assemblies, among which systems capable of SR would emerge during abiogenesis (figure 2). Many trajectories are theoretically possible from the initial set of reactants leading to emergent systems capable of SR. However, maintaining entropy changes constant (ie: the novel m emergent assemblies are all of equivalent complexity) and under non limiting resources conditions, assembly configurations more capable of maximising positive values of work, (ie more capable of using the work driven into the system from the external drive) would be able to maximise their exponential growth (maximise ($\frac{P_{SR}}{P_{SRdecay}}$)) (figure 2 and eq. 3) (England, 2015; Perunov et al., 2016) and quickly dominate the configuration landscape. As during abiogenesis the initial de novo emergent SR systems are unlikely to have encountered strong competition environments and/or limiting resources this analysis may provide an approximation to the conditions under which life emerged on earth. Conceptually, similar experiments studying early exponential growth stages of easily monitored de novo system, such as inoculation of one bacterium in a suitable growth media or the clonal selection of rapidly growing tumour cells from a heterogeneous population seem to support this hypothesis (Greaves et al., 2012). However, it is challenging to faithfully replicate these conditions artificially and thus further work remains needed to validate this hypothesis experimentally.

From a general theoretical point of view it could be argued that the fundamental postulates of statistical mechanics should be reinterpreted in reverse for far from equilibrium self-replicative life systems under an external driving field. We propose that the external driving field imposes work-dissipation evolutionary constrains on the system with boundaries set by the total usable work available to the system, or in other words, the work that can access or be loaded into the system's assemblies due to their intrinsic molecular arrangements. Thus, under this constrain the system evolves to adopt higher work dissipation macrostates within the system's permissible microstate configurations and this is reflected in a decrease of the total entropy in the system ($-\Delta E_i + \Delta S_i$). As self-replicative growth and variation are intrinsic features of bona fide life SR systems (as long as the external drive inputs/loads work into the system), this constraint would drive the evolution of the system to follow an increase in its complexity path under stable growth conditions. Therefore, in our hierarchical thermodynamic model the work from the external driving field loaded into the

system can be viewed as an evolutionary pressure selecting for higher work dissipative system arrangements. These in turn will further follow evolutionary trajectories that maximise positive values of work and thus constraining the system to evolve to higher levels of work dissipation/use as its intrinsic features change over evolutionary relevant time frames. Importantly, this prediction based on thermodynamics analysis agrees with all our empirical observations of the evolution of self-replicative life systems on earth (England, 2015; Jorgensen, 2007; Kleidon, 2009; Perunov et al., 2016; Schneider et al.,1994), implying that the basic organisation of self-replicative life systems is constrained by these basic thermodynamic principles. Thus, Fisher's remark "Natural selection is a mechanism for generating an exceedingly high degree of improbability" could be flipped upside down and reinterpreted as: "Generating an exceedingly high degree of improbability is an inevitable consequence of a hierarchical thermodynamic natural selection mechanism".

Finally, we have shown from a practical point of view how complex artificially closed systems (such as the Tongoy Bay ecosystem) and relevant evolutionary ecological time frames of life on earth can be modelled using a thermodynamic semiquantitative analysis. Interestingly we can show how punctuated evolutionary stages of life on earth correspond to incremental levels of work loaded into the system by the external drive with a corresponding increase in system's complexity (figure 6 and 7). We observe a clear trend towards global increased complexity of the earth-life system, in agreement with current theories (Karnani et al., 2009; Walker, 2017), and proposed that in the absence of external perturbations the directionality of this path (increased complexity and work dissipation) is set by the thermodynamic evolutionary constraints discussed in this manuscript. We propose that this type of analysis would be useful in understanding the directionality of complex biological systems across time and space and that experimental validations using artificially designed models could be easily implemented.

References

Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A., & Holliger, P. (2010). Ice as a protocellular medium for RNA replication. *Nature Communications*, *1*. https://doi.org/ARTN 7610.1038/ncomms1076

Battistuzzi, F. U., Feijao, A., & Hedges, S. B. (2004). A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *Bmc Evolutionary Biology*, *4*. https://doi.org/Artn 4410.1186/1471-2148-4-44

Bartolotta, A., Carroll, S. M., Leichenauer, S., & Pollack, J. (2016). Bayesian second law of thermodynamics. *Phys Rev E*, *94*(2-1), 022102. https://doi.org/10.1103/PhysRevE.94.022102

Catling, D. C., Glein, C. R., Zahnle, K. J., & Mckay, C. P. (2005). Why O-2 is required by complex life on habitable planets and the concept of planetary "oxygenation time". *Astrobiology*, *5*(3), 415-438. https://doi.org/DOI 10.1089/ast.2005.5.415

Crooks, G. E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Physical Review E*, *60*(3), 2721-2726. https://doi.org/DOI 10.1103/PhysRevE.60.2721

Desmond-Le Quemener, E., & Bouchez, T. (2014). A thermodynamic theory of microbial growth. *Isme Journal*, *8*(8), 1747-1751. https://doi.org/10.1038/ismej.2014.7

Duim, H., & Otto, S. (2017). Towards open-ended evolution in self-replicating molecular systems. *Beilstein J Org Chem*, *13*, 1189-1203. https://doi.org/10.3762/bjoc.13.118

England, J. L. (2013). Statistical physics of self-replication. *J Chem Phys*, *139*(12), 121923. https://doi.org/10.1063/1.4818538

England, J. L. (2015). Dissipative adaptation in driven self-assembly. *Nat Nanotechnol*, *10*(11), 919-923. https://doi.org/10.1038/nnano.2015.250

Fath, B. D., Patten, B. C., & Choi, J. S. (2001). Complementarity of ecological goal functions. *J Theor Biol*, *208*(4), 493-506. https://doi.org/10.1006/jtbi.2000.2234

Gonzalez, J., Ortiz, M., Rodriguez-Zaragoza, F., & Ulanowicz, R. E. (2016). Assessment of long-term changes of ecosystem indexes in Tongoy Bay (SE Pacific coast): Based on trophic network analysis. *Ecological Indicators*, *69*, 390-399. https://doi.org/10.1016/j.ecolind.2016.04.019

Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, *481*(7381), 306-313. https://doi.org/10.1038/nature10762

Hamilton, T. L., Bryant, D. A., & Macalady, J. L. (2016). The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. *Environmental Microbiology*, *18*(2), 325-340. https://doi.org/10.1111/1462-2920.13118

Heijnen, J. J., & Van Dijken, J. P. (1992). In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnol Bioeng*, *39*(8), 833-858. https://doi.org/10.1002/bit.260390806

Holdaway, R. J., Sparrow, A. D., & Coomes, D. A. (2010). Trends in entropy production during ecosystem development in the Amazon Basin. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *365*(1545), 1437-1447. https://doi.org/10.1098/rstb.2009.0298

Hsia, C. C. W., Schmitz, A., Lambertz, M., Perry, S. F., & Maina, J. N. (2013). Evolution of Air Breathing: Oxygen Homeostasis and the Transitions from Water to Land and Sky. *Comprehensive Physiology*, *3*(2), 849-915. https://doi.org/10.1002/cphy.c120003

Jorgensen, S. E. (2007). *A new ecology : systems perspective*. Elsevier. http://www.ezproxy.is.ed.ac.uk/login?url=https://www.sciencedirect.com/science/book/9780444531605

Jarzynski, C. (2011). Equalities and Inequalities: Irreversibility and the Second Law of Thermodynamics at the Nanoscale. *Annual Review of Condensed Matter Physics, Vol 2*, *2*, 329-351. https://doi.org/10.1146/annurev-conmatphys-062910-140506

Jorgensen, S. E., Ladegaard, N., Debeljak, M., & Marques, J. C. (2005). Calculations of exergy for organisms. *Ecological Modelling*, *185*(2-4), 165-175. https://doi.org/10.1016/j.ecolmodel.2004.11.020

Kleerebezem, R., & Van Loosdrecht, M. C. M. (2010). A Generalized Method for Thermodynamic State Analysis of Environmental Systems. *Critical Reviews in Environmental Science and Technology*, *40*(1), 1-54. https://doi.org/Pii 91837523310.1080/10643380802000974

Kleidon, A. (2009). Nonequilibrium thermodynamics and maximum entropy production in the Earth system. *Naturwissenschaften*, *96*(6), 653-677. https://doi.org/10.1007/s00114-009-0509-x

Kleidon, A., Malhi, Y., & Cox, P. M. (2010). Maximum entropy production in environmental and ecological systems. *Philos Trans R Soc Lond B Biol Sci*, *365*(1545), 1297-1302. https://doi.org/10.1098/rstb.2010.0018

Landauer, R. (1961). Irreversibility and Heat Generation in the Computing Process. *Ibm Journal of Research and Development*, *5*(3), 183-191. https://doi.org/DOI 10.1147/rd.53.0183

McCarty, P. L. (2007). Thermodynamic electron equivalents model for bacterial yield prediction: modifications and comparative evaluations. *Biotechnol Bioeng*, *97*(2), 377-388. https://doi.org/10.1002/bit.21250

Meysman, F. J. R., & Bruers, S. (2010). Ecosystem functioning and maximum entropy production: a quantitative test of hypotheses. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *365*(1545), 1405-1416. https://doi.org/10.1098/rstb.2009.0300

Nakashima, S., Kebukawa, Y., Kitadai, N., Igisu, M., & Matsuoka, N. (2018). Geochemistry and the Origin of Life: From Extraterrestrial Processes, Chemical Evolution on Earth, Fossilized Life's

Records, to Natures of the Extant Life. *Life-Basel*, *8*(4). https://doi.org/ARTN 3910.3390/life8040039

Payne, J. L., Boyer, A. G., Brown, J. H., Finnegan, S., Kowalewski, M., Krause, R. A., Lyons, S. K., McClain, C. R., McShea, D. W., Novack-Gottshall, P. M., Smith, F. A., Stempien, J. A., & Wang, S. C. (2009). Two-phase increase in the maximum size of life over 3.5 billion years reflects biological innovation and environmental opportunity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(1), 24-27. https://doi.org/10.1073/pnas.0806314106

Perunov, N., Marsland, R. A., & England, J. L. (2016). Statistical Physics of Adaptation. *Physical Review X*, *6*(2). https://doi.org/ARTN 02103610.1103/PhysRevX.6.021036

Pinero, J., & Sole, R. (2018). Nonequilibrium Entropic Bounds for Darwinian Replicators. *Entropy (Basel)*, *20*(2). https://doi.org/10.3390/e20020098

Prigogine, I., & Nicolis, G. (1971). Biological order, structure and instabilities. *Q Rev Biophys*, *4*(2), 107-148. https://doi.org/10.1017/s0033583500000615

Karnani, M., & Annila, A. (2009). Gaia again. *Biosystems*, *95*(1), 82-87. https://doi.org/10.1016/j.biosystems.2008.07.003

Ragazzon, G., & Prins, L. J. (2018). Energy consumption in chemical fuel-driven self-assembly. *Nat Nanotechnol*, *13*(10), 882-889. https://doi.org/10.1038/s41565-018-0250-8

Roden, E. E., & Jin, Q. (2011). Thermodynamics of microbial growth coupled to metabolism of glucose, ethanol, short-chain organic acids, and hydrogen. *Appl Environ Microbiol*, *77*(5), 1907-1909. https://doi.org/10.1128/AEM.02425-10

Saadat, N. P., Nies, T., Rousset, Y., & Ebenhoh, O. (2020). Thermodynamic Limits and Optimality of Microbial Growth. *Entropy (Basel)*, *22*(3). https://doi.org/10.3390/e22030277

Schrödinger, E. (1967). *What is life? the physical aspect of the living cell & Mind and matter*. University P.

Schneider, E.D., Kay, J.J., 1994. Life as a manifestation of the second law of thermo- dynamics. Math. Comp. Model. 19, 25–48.

Sessions, A. L., Doughty, D. M., Welander, P. V., Summons, R. E., & Newman, D. K. (2009). The continuing puzzle of the great oxidation event. *Curr Biol*, *19*(14), R567-574. https://doi.org/10.1016/j.cub.2009.05.054

Silow, E. A., & Mokry, A. V. (2010). Exergy as a Tool for Ecosystem Health Assessment. *Entropy*, *12*(4), 902-925. https://doi.org/10.3390/e12040902

Soo, R. M., Hemp, J., Parks, D. H., Fischer, W. W., & Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, *355*(6332), 1436-1440. https://doi.org/10.1126/science.aal3794

Spanner, D. C. (1953). Biological Systems and the Principle of Minimum Entropy Production. *Nature*, *172*(4389), 1094-1095. https://doi.org/DOI 10.1038/1721094a0

Ulanowicz, R. E. (2001). Information theory in ecology. *Comput Chem*, *25*(4), 393-399. https://doi.org/10.1016/s0097-8485(01)00073-0

Walker, S. I. (2017). Origins of life: a problem for physics, a key issues review. *Rep Prog Phys*, *80*(9), 092601. https://doi.org/10.1088/1361-6633/aa7804

Wu, Z. J., Wu, X. F., Yang, Z. H., & Ouyang, L. N. (2017). A simple thermodynamic model for evaluating the ecological restoration effect on a manganese tailing wasteland. *Ecological Modelling*, *346*, 20-29. https://doi.org/10.1016/j.ecolmodel.2016.12.008

Wu, Z. J., Wu, X. F., Yang, Z. H., & Ouyang, L. N. (2018). Internal energy ratios as ecological indicators for description of the phytoremediation process on a manganese tailing site. *Ecological Modelling*, *374*, 14-21. https://doi.org/10.1016/j.ecolmodel.2018.02.009

Yamamoto, M., & Takai, K. (2011). Sulfur metabolisms in epsilon- and gamma-proteobacteria in deep-sea hydrothermal fields. *Front Microbiol*, *2*, 192. https://doi.org/10.3389/fmicb.2011.00192

Zeng, H., & Yang, A. (2020). Bridging substrate intake kinetics and bacterial growth phenotypes with flux balance analysis incorporating proteome allocation. *Sci Rep*, *10*(1), 4283. https://doi.org/10.1038/s41598-020-61174-0