NET works after all? Engineering robustness through diversity

Mondeel, Thierry D.G.A.; Astrologo, Stefania; Zhang, Yanfei; Westerhoff, Hans V.

published in
IFAC-PapersOnLine
2018

DOI (link to publisher)
10.1016/j.ifacol.2018.09.007

document version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-19 (2018) 128–137

NET works after all? Engineering robustness through diversity

Thierry D.G.A. Mondeel*, Stefania Astrologo*, Yanfei Zhang*, Hans V. Westerhoff *,,*****

** Synthetic Systems Biology and Nuclear Organization, SILS, University of Amsterdam, NL, EU.*
*** Molecular Cell Physiology, Amsterdam Institute for Molecules, Medicines and Systems, Free University, Amsterdam, NL, EU*
**** Systems Biology, School of Chemical Engineering and Analytical Science, University of Manchester, UK, EU.*

Abstract: Classical Thermodynamics restricts engineering. These restrictions are independent of mechanism and kinetics, and thereby inescapable. Forgetting these restrictions can lead to over-optimistic designs for making bio-plastics from waste, and to erroneous ideas on early or new Life on this or other planets. This can be rectified by putting the thermodynamics in place. Is every biochemical network design feasible, provided one puts classical thermodynamics in place? Or, are there other, ill-recognized, generic restrictions to bioengineering?

For a while a Non-Equilibrium Thermodynamics (NET) has been trying to discover behaviors of dynamical systems away from equilibrium that are completely independent of kinetics and mechanism and thereby not engineerable. The principles discovered were of limited use to bioengineering however.

We here show that processes away from equilibrium must indeed depend on kinetics and mechanism, but, importantly, not on all kinetic and mechanistic details: There are limitations to what the engineering of mechanisms and kinetics can achieve. It is of course better to recognize what is impossible before trying to engineer the impossible. Importantly, the new NET methodology also shows that system properties that are possible, can be engineered only in certain ways.

The new NET methodology also enables to understand, and perhaps engineer towards, a performance that, by adjusting the network, remains optimal when conditions are changing. Using our in silico discovery tool, we show that this may indeed occur in the Archeon *S. solfataricus*. What we call 'variomatic' gear shifting is a way that some cells may use to self-engineer their ways to maximal growth rates in environments that lack robust resources, such as in environments with fluctuating oxygen levels.

Population heterogeneity is another mechanism that can increase the robustness of a cell factory. We discuss a NET principle that suggests ways in which one can engineer the cells' diversity. Transcription burst size rather than kinetics should be modulated for making a diverse population perform much better than its average.

Keywords: Non Equilibrium Thermodynamics, robustness, free energy, variomatic gear shifting, Bursting, Fano factor

1. INTRODUCTION

Molecules are composed of neutrons, protons and electrons. Due to the electric charges of the latter two elementary particles, different molecules, i.e. different constellations of the particles have different energies. Energies also differ between different dynamic conformations of the molecules. Usually there is an equilibration between these conformations such that a molecule can be characterized by an average energy. The important corollary is that the impact of the molecule on the performance of any system of interest can be described in terms of that average energy rather than in terms of the impacts of all the molecules of the same identity but in the different conformations. With the myriads of individual molecules in living organisms with millions each of conformational states, the phenomenon that impact can be described in terms of averages is essential for both bioscience and bioengineering. Not even the fastest computer will ever be able to compute the behaviors of all the individual molecules of a living cell, first because its capacity is too small, and second because information is lacking on the initial state of all the individual molecules: we would not even know where to start computing.

But let us face it, we are not interested in the behavior of every individual molecule of a living cell. We are usually interested in the behavior of populations of cells that perform a certain function either in the sense of biotechnology or in the sense of pathophysiology. Accordingly, understanding the behavior of populations of cells as a function of the *average* properties of all the molecules of a given identity within them, is close enough to what we really want. With the added acknowledgement that the molecules of the various identities

10.1016/j.ifacol.2018.09.007

are engaged in dynamic networking that through nonlinear interactions gives rise to new functional properties, this understanding is the ambition of systems biology. The ambition to predict and engineer towards a useful behavior is then the ambition of systems bioengineering.

Molecular mechanics or molecular dynamics is the discipline that studies the dynamic behavior of individual molecules. Statistical mechanics deals with the statistical properties of ensembles of such molecules. It argues in terms of probabilities and probability distributions. Whenever the average behavior of an ensemble of molecules can be described in terms of averaged properties such as averages, variances and skewness, the discipline becomes statistical thermodynamics, and when averages suffice, generalized thermodynamics is the discipline in charge. What is commonly called 'kinetics' discusses reaction rates in terms of ensemble averaged concentrations or, if more sophisticated, in terms of activities. In this sense it is a branch of generalized thermodynamics. In practice kinetics also has an empirical or a quasi-probabilistic basis and in its extrapolations it is not necessarily weary of limitations imposed by thermodynamic principles.

Equilibrium thermodynamics champions at least two such principles. One, which it shares with (quantum) mechanics, is the law of conservation of energy (U). Energy U can be brought into a system through heat import, by doing work on the system, or by importing substances with high energy content (Westerhoff and Van Dam, 1987). It cannot be produced or annihilated (dissipated) however. The other is the law of entropy production, which states that entropy (S) can only be produced and not consumed or destroyed. Entropy is the logarithm of the number of realizations of a system at any given energy. The second law of thermodynamics basically rewords a probabilistic law i.e. *ceteris paribus* (i.e. in splendid isolation) a system will move from a state with lower probability to a state with higher probability, whenever such movement is possible, and not in the opposite direction. Movement in this opposite direction would destroy entropy. Ordered states usually have a smaller multiplicity and hence a higher probability and entropy, than chaotic states of a system at the same energy. Hence this second law maintains that systems in splendid isolation cannot become more ordered and the first law states that they cannot grow from low to high energy content. The paradox that the development of an adult organism from a fertilized egg should then be impossible, is resolved by acknowledging that such developing living systems must be open, in order to import energy and to export more entropy than the order (negative entropy) it creates internally. For open 'metabolic' systems the two laws of thermodynamics reduce to the requirement that 'metabolic' (approximately equal to Gibbs') free energy can only be dissipated and must in fact be dissipated to maintain and proliferate the living state (Westerhoff and Van Dam, 1987). At equilibrium the free energy differences of all reactions that are possible should equal zero. Autonomous reactions, i.e. reactions not coupled to any other processes cannot run uphill in terms of the free energy. This second law of thermodynamics is general, i.e. independent of mechanism. No enzyme or network mechanism can be engineered so that it circumvents this limitation: equilibrium thermodynamics has the strength that it is completely independent of mechanism.

An underlying and often overlooked limitation is however that the validity of this thermodynamics itself and therewith the validity of its second law, depends on the *proviso* mentioned above that dynamic behavior can be described in terms of average concentrations. In this paper we shall effectively demonstrate that if that proviso is not met, the second law *per se* may not be valid for some important biological systems.

In its first incarnation, non-equilibrium thermodynamics (NET) dealt with the paradox how Gibbs energy could be dissipated yet increase at the same time. We shall envisage processes in terms of a flux J, positive when proceeding in the forward direction, and a ΔG equal to the Gibbs energy of the products minus the Gibbs energy of the reactants. The two processes are growth (or 'anabolism') and catabolism, as indicated by subscripts 'a' and 'c', respectively.

First assuming that there is only growth, the rate of Gibbs energy dissipation should equal (Westerhoff and Van Dam, 1987)

$$\Phi \stackrel{\text{def}}{\equiv} -\frac{\mathrm{d}_i G}{\mathrm{dt}} = J_a \cdot -\Delta G_a > 0$$

$\frac{\mathrm{d}_i G}{\mathrm{dt}}$ is an incomplete differential equal to the Gibbs energy increase due to Gibbs energy production, which has to be negative according to the second law of thermodynamics (Nicolis and Prigogine, 1977). Because ΔG_a is (usually, though not always; Westerhoff and Van Dam, 1987) positive, J_a must be negative, implying that growth of a microorganism, or analogously, the production of value added compounds, should be impossible (in fact negative, corresponding to death) according to this equation alone. This leads to the paradox that Life cannot exist (or persist) although it does.

The resolution to this paradox that is practiced emphatically by living systems is that the thermodynamically uphill and thereby 'forbidden' anabolic process is coupled to a thermodynamically downhill process, often called 'catabolism' (referred to by subscript 'c'), at positive flux J_c that dissipates more free energy than that the anabolic process consumes when J_a is positive. Consequently, in total the Gibbs energy is then dissipated at a positive rate $-\frac{dG_i}{dt}$:

$$\Phi = J_a \cdot -\Delta G_a + J_c \cdot -\Delta G_c > 0$$

under the condition that

$$J_c \cdot -\Delta G_c > J_a \cdot \Delta G_a$$

The 'coupling' does not occur automatically however. It requires some coupling mechanism by which the anabolic flux is pushed in the negative direction (i.e. towards biosynthesis) by the thermodynamic driving force provided by the free energy of catabolism. Thereby the anabolic flux becomes a function of both free energy differences. Accepting that this may be so for both fluxes, expanding both functions as Taylor series around equilibrium, using that at

zero free energy differences the fluxes must be zero (equilibrium), and neglecting all higher than first order terms, the coupling can be described by a negative 'phenomenological coefficient' L_{ac} in the phenomenological flow-force relations:

$$J_a = L_{aa} \cdot -\Delta G_a + L_{ac} \cdot -\Delta G_c$$

$$J_c = L_{ca} \cdot -\Delta G_a + L_{cc} \cdot -\Delta G_c$$

, where ΔG_a, $-\Delta G_c$, L_{aa}, L_{cc}, L_{ca}, L_{ac} and J_c are all positive, so that also J_a becomes positive, reflecting actual biomass synthesis against the free energy difference of biomass synthesis (see below).

These equations describe deterministic behavior of the average system. The molecular world is more variable than this deterministic behavior however. It is subject to quasi-random reaction events that may transiently violate the second law of thermodynamics. Such transitions must be followed by transitions that return the overall behavior of the system to the deterministic behavior.

The underlying reason for this is that although systems tend to move from a low to a higher probability state, this probabilistic law is itself subject to stochasticity: systems can transiently move to a less probable state; the law of movement towards higher probability is only true on average. Both the deterministic behavior and the behavior that is in transient violation of deterministic behavior are due to the same processes of rapid energy exchange between molecules and their environments that occurs at temperatures above zero Kelvin. Where there are processes there are fluctuations and this is true for living systems as well. In this sense the beautiful figures in cell-biology textbooks fall short of the reality that is much more chaotic. Only the average could behave in accordance with the diagrams.

At some time zero, any real system may be in any state of any probability (Keizer, 1987). With time it will then, on average, move to more probable states and as time progresses it will be 'caught' by an environment of states that are highly probable, i.e., by a so-called 'attractor'. These attractors need not be the most probable states but they should be situated at a hill of high probability surrounded on most sides by states of lower probability. Stable steady states are such attractors. In such steady states all concentrations are often said to be independent of time, but in reality they are not precisely so: They fluctuate and are thereby varying with time. It is their time average over some limited time span that is independent of time. Both thermodynamics and kinetics deal with such time averaged fluctuating concentrations and they may do this even outside steady states, as fluctuations are often much faster than the times characteristic of the evolution of the system.

When systems are not yet close to an attractor, there are often great differences between individual systems. For reasons suggested above, this makes such conditions unattractive and not useful for scientific analysis or engineering. Engineering only one out of every one million cells in a population in terms of producing something useful is not usually relevant for bioengineering, because the corresponding productivity will be low. Existing methods of kinetics and non-equilibrium thermodynamics therefore only address systems that are already in densely-populated attractor states or attractor trajectories. Because of the closeness of those systems and because of the property that they tend to remain close to the attractor, the fluctuations around the attractor state are regular (Keizer, 1987).

The fact that most observable and relevant systems are stably in attractor states, leads to an important law: the system under considerations is stable towards all actual fluctuations. Refining the required stability to stability in the sense of Lyapunov, this means that after any possible fluctuation, the system will in average return to a state that is infinitesimally close to the attractor itself (Westerhoff and Van Dam, 1987). According to the fluctuation-dissipation principle, the return to the attractor state after the fluctuation follows deterministic behavior; or perhaps rather vice versa: the deterministic behavior follows the same path as the response to a fluctuation.

The equilibrium state is such an attractor. In the above described example, equilibrium is where both free energy differences are equal to zero. Considering fluctuations in the free energy of catabolism only, the excess free-energy dissipation *subsequent* to the fluctuation is:

$$\delta\Phi = \delta J_c \cdot \delta(-\Delta G_c) = L_{cc} \cdot (\delta\Delta G_c)^2$$

This process subsequent to the fluctuation must be one of returning to the attractor as a deterministic process and must therefore have a positive free energy dissipation, i.e.:

$$0 < \Phi + \delta\Phi$$

Because the fluctuation started at equilibrium where free energy dissipation Φ was zero, the excess free energy dissipation must be positive:

$$0 < \delta\Phi$$

Accordingly, the second law of thermodynamics implies that the phenomenological coefficient L_{cc} must be positive.

The catabolic reaction may consist of the breakdown of a substrate *S* to a product *P*, so that:

$$\Delta G_c = \mu_P - \mu_S$$

μ_S and μ_P refer to the chemical potential (partial molar free energy) of catabolic substrate *S* and catabolic product *P* respectively. The above phenomenological equations state that around equilibrium the catabolic flux is proportional to this free energy drop of catabolism. This implies that the dependence of the flux on the chemical potential of the substrate is equal to minus its dependence on the chemical potential of the product. Assuming that this is not the case, would lead to the equation:

$$J_c = L_{ccP} \cdot \mu_P - L_{ccS} \cdot \mu_S$$

In an open network, *S* is produced by some other reaction and *P* consumed. Hence, *S* and *P* can fluctuate independently and also in such a way that their difference is not affected by the fluctuation. For such an equal fluctuation in the chemical potentials of the substrate and the product one finds:

$$\delta J_c = (L_{ccP} - L_{ccS}) \cdot \delta\mu_S$$

Because ΔG_c has remained equal to zero, there is no driving force for a catabolic flux in either direction, so that J_c must remain zero. Hence for a non-zero $\delta\mu_S$:

$$0 = \delta J_c = (L_{ccP} - L_{ccS}) \cdot \delta\mu_S$$

So that:

$$L_{ccS} = L_{ccP} \stackrel{\text{def}}{=} L_{cc}$$

This confirms that one can write the two fluxes as linear functions of the free energy differences rather than of the individual chemical potentials. But is also an example of Onsager's reciprocity relations.

The anabolic reaction may consist of the conversion of substrate for anabolism, *A*, into biomass *B*, so that:

$$\Delta G_a = \mu_B - \mu_A$$

The fully coupled reaction is then:

$$S + A \rightleftharpoons P + B$$

With fluxes:

$$J_c = J_a = L \cdot (-\Delta G_{tot}) = L \cdot (-\Delta G_c - \Delta G_a)$$

For the same reasons as above we can write the flux as a linear function of the overall free energy difference rather than of its individual components. Hence:

$$L_{ca} = \frac{\partial J_c}{\partial \Delta G_a} = -L = \frac{\partial J_a}{\partial \Delta G_c} = L_{ac}$$

which, more than the above, is known as Onsager reciprocity. If there are additional catabolic and anabolic processes that are not coupled to anabolism and catabolism, respectively, then these should be added to the equations. This will not affect the Onsager reciprocity (Westerhoff and Van Dam, 1987).

Onsager reciprocity is valid close to equilibrium. Further away from equilibrium the proof for it breaks down, as:

$$\begin{aligned} 0 < \Phi + \delta\Phi &= (J_c + \delta J_c) \cdot (\Delta G_c + \delta\Delta G_c) \\ &= J_c \cdot \Delta G_c + \delta J_c \cdot \Delta G_c + J_c \cdot \delta\Delta G_c + \delta J_c \\ &\cdot \delta\Delta G_c \\ &= J_c \cdot \Delta G_c + L_{cc} \cdot \Delta G_c \cdot \delta\Delta G_c \\ &+ J_c \cdot \delta\Delta G_c + L_{cc} \cdot (\delta\Delta G_c)^2 \end{aligned}$$

In this case there is no reason for L_{cc} to equal zero, as the leading term $J_c \cdot \Delta G_c$ already guarantees a positive free energy dissipation. Indeed, Onsager reciprocity has been shown to be absent in actual cases of mitochondrial oxidative phosphorylation, where the phosphorylation flux hardly depended on the free energy of respiration whilst the respiratory flux was greatly reduced by increased phosphorylation potential (Westerhoff and Van Dam, 1987). Yet the coupling, i.e. the positivity of L_{ca} and of L_{ac} persists qualitatively.

Returning to near equilibrium conditions, Onsager reciprocity is also an example of non-equilibrium thermodynamics (NET), in the sense that it (i) addresses systems that are not at equilibrium, (ii) describes the system in fewer than the total number of independent variables (i.e. ΔG_c rather than the individual chemical potentials with as corollary Onsager reciprocity) and (iii) omits some mechanistic detail such as the precise way the coupling (positivity of L_{ca}) is achieved. Yet this NET differs from equilibrium thermodynamics in that it admits some mechanistic detail, i.e. the phenomenon and extent of coupling; the statement that there must be a mechanism making $L_{ac} > 0$.

The degree of coupling is quantified by the ratio of the cross coefficient L_{ac} to the straight coefficients and has been defined, for the near equilibrium steady states, as:

$$q \stackrel{\text{def}}{=} \frac{L_{ca}}{\sqrt{L_{cc} \cdot L_{aa}}}$$

This degree of coupling q has to lie between 0 and 1 for the free energy dissipation to be always positive (Westerhoff and Van Dam, 1987). With Z defined as:

$$Z = \sqrt{\frac{L_{aa}}{L_{cc}}}$$

the so-called flow (J)-force ($-\Delta G$) relations can be rephrased as:

$$\frac{J_a}{L_{cc} \cdot -\Delta G_c} = q \cdot Z - Z^2 \cdot \frac{\Delta G_a}{-\Delta G_c}$$

$$\frac{J_c}{L_{cc} \cdot -\Delta G_c} = 1 - q \cdot Z \cdot \frac{\Delta G_a}{-\Delta G_c}$$

These equations show that provided the degree of coupling is close to 1, Z is on the order of 1 or 2 and $\frac{\Delta G_a}{-\Delta G_c} < 1/2$, anabolism is substantial as compared to catabolism, although decreasing linearly with the back pressure exerted by its own free energy. Catabolism should be subject to a similar reduction at high free energy of anabolism. Growth comes to a halt at a 'static head' free energy of anabolism of:

$$\left(\frac{\Delta G_a}{-\Delta G_c}\right)_{static\ head} = \frac{q}{Z}$$

Catabolism then still continues unless there is complete coupling ($q = 1$). Yield is given by:

$$Y \stackrel{\text{def}}{=} \frac{J_a}{J_c} = Z \cdot \frac{q - Z \cdot \frac{\Delta G_a}{-\Delta G_c}}{1 - q \cdot Z \cdot \frac{\Delta G_a}{-\Delta G_c}}$$

This equation explains why Z is called the phenomenological stoichiometry: at full coupling or at very low free energy of anabolism, it equals the ratio of growth rate to catabolic flux, i.e. the yield $Y \approx Z$. The equation also implies that, unless q is close to 1, at low free energy of anabolism the yield should decrease almost linearly with increasing free energy of anabolism. Because the denominator in this equation then goes to zero, at higher such free energies of anabolism, this decrease should become progressively stronger. The thermodynamic efficiency, equal to the product of the yield with the ratio of free energy differences, thereby exhibits an optimum in its variation with the free energy of anabolism:

$$\eta \stackrel{\text{def}}{=} \frac{J_a \cdot \Delta G_a}{J_c \cdot -\Delta G_c} = Z \cdot \frac{q \cdot \frac{\Delta G_a}{-\Delta G_c} - Z \cdot \left(\frac{\Delta G_a}{-\Delta G_c}\right)^2}{1 - q \cdot Z \cdot \frac{\Delta G_a}{-\Delta G_c}}$$

Anabolic flux, yield and efficiency should all increase with tighter coupling. After billions of years of evolution one might therefore expect the degree of coupling to equal 1 meaning that coupling would be complete. In reality coupling is less than complete. The reason is of interest to bioengineering, as q might be a parameter to use in engineering towards better productivity. There are at least three feasible explanations for this lack of complete coupling. The one referring to physical-chemical limitations to stability, is perhaps most pertinent for the many cases where ion-gradient dependent free energy transduction is involved: it may be impossible to make membranes fully tight with respect to ion leakage. A second referring to other free energy dissipating processes that are essential to maintain the living state, so-called maintenance processes, is also feasible. Perhaps a more intriguing explanation was developed by (Stucki, 1980): incomplete coupling might itself be optimal. Determining for each degree of coupling the anabolic free energy optimal for achieving maximum thermodynamic efficiency, and then plotting anabolic flux or yield (all normalized in some way by Z) for varying degrees of coupling as a function of the free energy of anabolism, optima were found at incomplete coupling. The values found for free energy of anabolism, degree of coupling and efficiency did make sense for mitochondrial oxidative phosphorylation (Stucki, 1980) and microbial growth (Westerhoff and Van Dam, 1987). The success of this theory was surprising because its computations used the above proportional relationships between fluxes and free energy differences, as well as Onsager symmetry, whilst the systems addressed were too far from equilibrium for the proofs of these properties to persist. Anyway the concept that living systems may adjust the degree of coupling in order to attain optimality, and the idea that there could be more objective functions than growth rate or growth yield, is one that will inspire us below, when we consider adjustment of the phenomenological stoichiometry Z.

As explicated above, thermodynamics is a necessary simplification of the reality where all individual systems and molecules behave differently. It thrives on systems rapidly moving to the vicinity of attractor states and then evolving further or staying put, such as in steady states. Still, for systems that are identical in terms of external and internal parameters and in (the environment of) such an attractor state, the molecule numbers of any given molecular species differ between those systems. For any single such system at steady state, the molecule number of that species fluctuates with time. When the attractor is the equilibrium state, the molecule number differences follow a Poisson distribution around the average $\overline{N}$, i.e.

$$P(N = n) = e^{-\overline{N}} \cdot \frac{(\overline{N})^n}{n!}$$

The variance of a Poisson distribution equals the average, implying that the 'relative noise', defined as the standard deviation in the molecule number relative to the average, decreases with the square root of the average. This in turn means that for molecule numbers in excess of 10 000 the molecule numbers vary between cells or with time by less than 1 %. Since the chemical potential is linearly related with the logarithm of the molecule number, the noise in the chemical potential is also small and approximately equal to the relative noise:

$$\sigma\left(\frac{\mu}{RT}\right) \approx \frac{1}{\overline{N}} \cdot \sigma(N) = \frac{1}{\sqrt{\overline{N}}}$$

In living cells molecule numbers do exceed 10 000, with the exception of genes and some mRNAs. Since it is the proteins and metabolites that matter directly for function, this is thought to rationalize the use of thermodynamics and kinetics. Below we shall discuss an exception to this that appears to be of increasing importance to biology and may prove important for bioengineering.

Above we showed that the stability of the equilibrium state is guaranteed if for any state that is attainable by a possible fluctuation, the excess free energy dissipation is positive, i.e. that the Gibbs free energy dissipation in the equilibrium state is minimal:

$$\delta\Phi > 0$$

For attractors near equilibrium where Onsager reciprocity holds, one finds for the excess free energy dissipation (Nicolis and Prigogine, 1977):

$$\delta\Phi = \delta\Phi_G + \delta\Phi_J$$

With:

$$\begin{aligned}\delta\Phi_G \stackrel{\text{def}}{=} & J_c \cdot \delta(-\Delta G_c) + J_a \cdot \delta(-\Delta G_a) \\ & = L_{cc} \cdot \Delta G_c \cdot \delta\Delta G_c + L_{ca} \cdot \Delta G_a \cdot \delta\Delta G_c \\ & + L_{aa} \cdot \Delta G_a \cdot \delta\Delta G_a + L_{ac} \cdot \Delta G_c \cdot \delta\Delta G_a\end{aligned}$$

$$\begin{aligned}\delta\Phi_J \stackrel{\text{def}}{=} & \delta J_c \cdot -\Delta G_c + \delta J_a \cdot -\Delta G_a \\ & = L_{cc} \cdot \Delta G_c \cdot \delta\Delta G_c + L_{ca} \cdot \Delta G_c \cdot \delta\Delta G_a \\ & + L_{aa} \cdot \Delta G_a \cdot \delta\Delta G_a + L_{ac} \cdot \Delta G_a \cdot \delta\Delta G_c\end{aligned}$$

If Onsager reciprocity applies, as approximately in steady states that are close to equilibrium, the two terms of the free energy dissipation are equal, so that:

$$\delta\Phi = 2 \cdot \delta\Phi_J$$

Away from equilibrium the free energy dissipation is always positive, also when the system responds to a fluctuation. For any fluctuation, that response, which is the δJ_c, to any fluctuation $\delta\Delta G_c$ should bring the system back to equilibrium if the system was stable. Consequently:

$$0 < \delta\Phi = \delta^1\Phi + \delta^2\Phi$$

Here we developed the change in the dissipation function into first and second order terms. For simplicity we here further consider the case in which the free energy of anabolism is zero and not fluctuating, so that we need to consider one process only. Then:

$$\delta^1\Phi = J_c \cdot \delta - \Delta G_c + \delta J_c \cdot -\Delta G_c = 2 \cdot \delta J_c \cdot -\Delta G_c = 0$$

$$\delta^2\Phi = 2 \cdot \delta J_c \cdot \delta - \Delta G_c = 2 \cdot L_{cc} \cdot (\delta\Delta G_c)^2$$

Because at equilibrium, the dissipation function equals zero. This implies that L_{cc} must be positive. For steady states away from but still close to equilibrium, the second order term remains positive.

We now consider a non-equilibrium steady state that is close enough to equilibrium for the flux and free energy difference to be smaller than what could arise out of large but realistic fluctuations out of the equilibrium state. Then the second order term would continue to dominate over the first order term and the change in free energy dissipation caused by the fluctuation should remain positive.

Here the stability of the state and hence the property that after a fluctuation the deterministic behavior should return the system to its initial state was used to prove the free energy dissipation should be minimal so that any fluctuation should bring the system to a steady state with higher free energy dissipation. Prigogine proposed to invert the argument and use this minimal free energy dissipation as a stability criterion (Nicolis and Prigogine, 1977). For long Prigogine and co-workers sought to show that this same criterion or some extension thereof also applied to steady states further away from equilibrium than the Onsager domain. To some, such a mechanism-independent stability criterion, became effectively synonymous with non-equilibrium thermodynamics.

In this paper we shall identify a non-equilibrium thermodynamics that is valid outside the Onsager domain, inclusive of stability criteria. We shall also develop this NET approach such as to be able to deal with biological systems which adjust stoichiometries of reactions (e.g. *Z*) rather than the degree of coupling. We shall identify a 'variomatic' strategy, i.e. one in which an organism continuously optimizes the stoichiometries. With this we shall show how, useful NET approaches can be developed by allowing a limited amount of mechanistic detail to enter the considerations. In addition, we shall examine the limits of the thermodynamic approach of describing average system behavior in terms of average molecule numbers or chemical potentials. We shall do this in the light of the frequent observation that in isogenic populations of cells, molecule numbers are not distributed narrowly. This phenomenon may cause rather substantial clashes with predictions made by traditional thermodynamic and kinetic approaches. We shall propose a new type of NET that begins to deal with heterogeneous cell biochemistry.

2. RESULTS

2.1 Stability criteria

Stable metabolic and related networks exhibit relationships between so-called elasticity coefficients and control coefficients (Westerhoff and Chen, 1984). The most relevant example here is the concentration control connectivity law

$$\sum_{i=1}^{m} C_i^{X_j} \cdot \varepsilon_{X_k}^i = -\delta_k^j$$

The left hand-side describes the summation over all *m* reaction processes, of the multiplication of the control exercised by process *i* over the concentration (chemical potential) of metabolite X_j with the elasticity of process i with respect to the concentration of metabolite X_k. The right–hand side of the equation is minus the Kronecker delta, which equals 0 for $j \neq k$ and 1 for $j = k$. Elasticity coefficients are local derivatives of the logarithm of any process rate with respect to the logarithm of the chemical potential of any freely fluctuating metabolite. They contain the summaries of the kinetic details of the processes that suffice to determine the control of the network. They harbor some kinetic detail but not all. The control coefficients are the dependences of the logarithm of the chemical potentials at steady state on the logarithm of any of the process activities. Logarithms are here natural ('*ln*'), and chemical potentials are normalized by *RT*. When emphasizing the thermodynamic nature of this law even further, the term 'concentration of' is replaced by 'chemical potential of'. The law has been proven by using the requirement that the deterministic response of the system to a fluctuation in the chemical potential of X_k alone, must be such that that chemical potential returns on average to its initial value, whilst all other chemical potentials remain unchanged (Westerhoff and Chen, 1984). Inversing the argument, we here propose that this connectivity property is the stability criterion for non-equilibrium steady states, also for those beyond the Onsager domain.

2.2 The variomatic strategy

When plotting the anabolic flux as a function of the free energy of anabolism (taken relative to the free energy of catabolism, i.e. $\frac{\Delta G_a}{-\Delta G_c}$) for various values of the phenomenological stoichiometry *Z*, one obtains (Fig. 1) a family of downward straight lines, running from (0, *qZ*) to $(\frac{q}{Z}, 0)$. At lower values of the free energy of anabolism, higher phenomenological stoichiometries lead to higher anabolic fluxes, but at more challenging free energies of anabolism, the systems with lower values of *Z* lead to faster anabolism. This is akin the effect of shifting to lower gear when driving a car up a steeper and steeper mountain road.

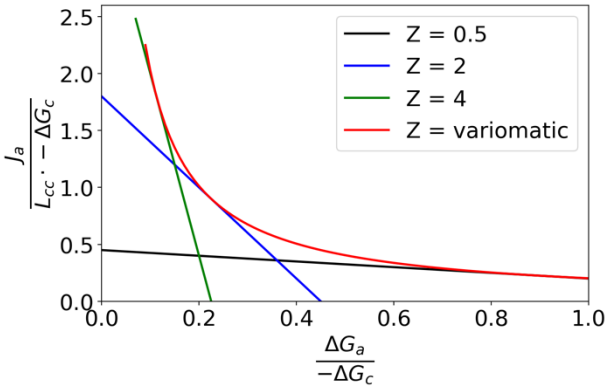

Fig. 1. Normalized anabolic flux versus ratio of free energy differences. Straight lines: Anabolic flux J_a (normalized by $L_{cc} \cdot -\Delta G_c$) as a function of the free energy of anabolism (normalized by the free energy of catabolism, i.e. $\frac{\Delta G_a}{-\Delta G_c}$), for a

degree of coupling q=0.9 at various values (i.e. 0.5, 2 and 4) of the phenomenological stoichiometry *Z*. The blue curve connects the states produced by variomatic gear shifting.

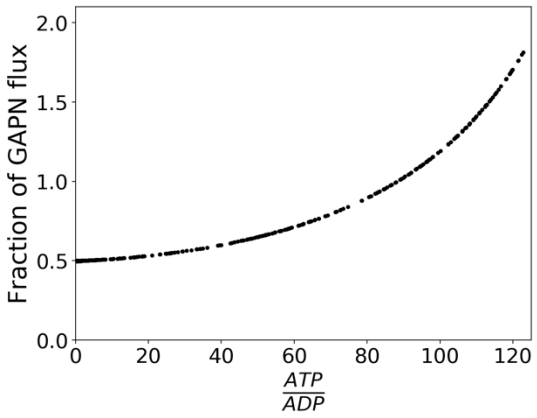

Fig. 2. The fraction of the flux from GAP to pyruvate that runs through the GAPN pathway rather than through the GAPDH + PGK pathway. Calculated for the *in silico* model of Zhang et al. (2017). GAPDH = glyceraldehyde 3 phosphate dehydrogenase. PGK = phosphoglycerate kinase. GAP = glyceraldehyde 3 phosphate. The GAPN activity was changed to 0.165 mM/min, the rate constant of the first order ATP hydrolysis reaction was modulated and the ATP/ADP ratio and steady state fluxes calculated.

Does such gear shifting happen in Biology as well? Using a genome wide metabolic map, Mondeel et al (2016) showed that some acetogenic bacteria such as *Cl. Ljungdahlii*, have electron transfer networks that allow for 24 different routes with together more than 10 different ATP/acetate ratios. Thereby, the potential for gear shifting is there, but we have not yet been able to demonstrate that these organisms do actually switch between these different routes. Awaiting options to demonstrate gear shifting experimentally, we turned to possibly the next best option: we tested gear shifting in an *in silico* replica model of the lower half of the glycolytic pathway in the Archeon *S. sulfataricus* (Zhang et al., 2016). We varied the ATP/ADP ratio by varying the unimolecular rate constant of ATP hydrolysis and calculated the fluxes both through the GAPDH + PGK route (the pathway with an ATP/pyruvate ratio of 2) and through the GAPN (the non-phosphorylating glyceraldehyde 3-phosphate dehydrogenase which this organism hosts as well) route (the pathway with an ATP/pyruvate ratio of 1). Fig. 2 shows that indeed the fraction of the flux between GAP and pyruvate that runs through the low stoichiometry pathway increases with increasing ATP/ADP ratio, hence with increasing free energy of ATP hydrolysis: *in silico* the organism shifts to lower gear when the thermodynamics gets tough.

A variomatic strategy could optimize the gear shifting so that always the highest anabolic flux is attained at every free energy of anabolism. Equating the derivative of the anabolic flux with respect to the stoichiometry, to zero, one finds the optimal phenomenological stoichiometry for every free energy of anabolism:

$$Z_{optimal} = \frac{q \cdot -\Delta G_c}{2 \cdot \Delta G_a}$$

This confirms that with increasing slope (ΔG_a) it is better to shift to lower gear, i.e. to pathways in the metabolic network that have a reduced stoichiometry. It also shows that if more input free energy is applied, it is better to operate at higher gear. Inserting this expression into the anabolic flow-force relationship:

$$\frac{J_a}{L_{cc} \cdot -\Delta G_c} = \frac{q^2}{4} \cdot \frac{-\Delta G_c}{\Delta G_a}$$

This variomatic curve is shown as the hyperbolic decrease with anabolic free energy in Fig. 1. The ratio of the anabolic flux of the fixed stoichiometry network to that of the variomatic network is:

$$\frac{J_{a,fixed\ Z}}{J_{a,variomatic}} = 4 \cdot \frac{\frac{\Delta G_a}{-\Delta G_c} \cdot \left(q \cdot Z - Z^2 \cdot \frac{\Delta G_a}{-\Delta G_c}\right)}{q^2}$$

which is always smaller than 1 except for when Z precisely corresponds to the optimal gear setting at that anabolic free energy (see above). Variomatic gear shifting in anabolism should be beneficial for anabolic flux, especially under conditions of variable free energies.

Recently we noted that in the energy and carbon metabolism networks of acetogenic bacteria various routes produce ATP and fix carbon at different stoichiometries (Mondeel et al., 2016). We proposed that those bacteria may engage in gear shifting by adjusting the gene expression of the various metabolic routes to the free energy of ATP or any other relevant anabolic free energy. We now propose to examine whether these bacteria come close to the optimal variomatic strategy. We also propose that by engineering such a variomatic gene expression strategy into other organisms, one may enhance their robustness and perhaps bioproductivity.

2.3 NET for disperse systems

There are now many observations showing that Biology may not much keep to the desire of thermodynamics and kinetics of keeping differences in molecule numbers between individual cells small relative to those numbers themselves. Where the phenomenon is not due to genetic heterogeneity and persists at steady state, it has been attributed to noise. Normally however, noise in molecule numbers is expected to be small because distributions are expected to be Poissonian (see above). Where could such large noise come from and could it be so large as to frustrate known thermodynamic and kinetic approaches? We shall consider the simple metabolic pathway where molecule X (such as an mRNA) is synthesized at a zero order rate constant $k_{bursting}$ and degraded at a first order rate constant $k_{degradation}$. What is special, is the phenomenon that the synthesis occurs in batches of size b; it is bursting. We assume that degradation occurs only in bursts of 1. When applied to a system of cells that initially each have different molecule numbers *N* of X, this leads to a distribution of molecule numbers of X that becomes steady after a while. Using the master equation approach (Van

Kampen, 1976) the average and variance in the molecule number become:

$$\overline{N} = \frac{k_{bursting} \cdot b}{k_{degradation}}$$

$$F \stackrel{\text{def}}{=} \frac{\sigma^2}{\overline{N}} = \frac{b+1}{2}$$

For a network with linear non-bursty kinetics (b=1), one obtains a Poisson distribution (F=1) at steady state. If reactions are bursting, non-Poisson distributions are attained. The equation also has an interesting corollary: the ratio of variance to average which is equal to the Fano factor F, is independent of kinetic detail other than the burst size. Here again one finds a type of thermodynamic simplification at steady state. Fig. 3 shows how in a numerical simulation that used a Gillespie algorithm, it is the Fano factor that is exclusively independent of the kinetic constants.

The bursting leads to increased relative noise:

$$\frac{\sigma}{\overline{N}} = \sqrt{\frac{F}{\overline{N}}} = \sqrt{\frac{1 + \frac{k_{degradation}}{k_{bursting}} \cdot \overline{N}}{2 \cdot \overline{N}}} \approx_{for\ large\ burst\ size} \sqrt{\frac{k_{degradation}}{2 \cdot k_{bursting}}}$$

The implication is that, contrary to the more classical case of Fig. 3 where the bursting frequency was modulated, for large burst sizes the relative noise does not decrease with increasing average molecule number when the latter increase is achieved by increasing the burst size rather than the burst frequency. In that case the variance is predicted to increase with the square of the average molecule number and this has indeed been observed experimentally for some mRNAs.

Are there any implications of this type of dispersion for predictions of functional behavior? We shall here discuss the simple example where function depends quadratically on the molecule number of a molecular species. In terms of average molecule number, the deterministic rate equation for this function is:

$$v_{deterministic} = k \cdot (\overline{N})^2$$

The stochastic rate equation reads as follows:

$$\overline{v_{stochastic}} = k \cdot \overline{N \cdot (N-1)}$$

$$\frac{\overline{v_{stochastic}}}{k \cdot (\overline{N})^2} = 1 + \frac{\sigma^2}{(\overline{N})^2} - \frac{1}{\overline{N}} = v_{deterministic} \cdot \left(1 + \frac{F-1}{\overline{N}}\right) = v_{deterministic} \cdot \left(1 + \frac{b-1}{2 \cdot \overline{N}}\right)$$

This shows that the prediction by the average of the stochastic rate equation is equal to the prediction by the deterministic rate equation if the burst size is equal to 1 (i.e. F = 1), or if the burst size is much smaller than the average molecule number.

If the average molecule number is close to half the burst size or even smaller, then the deterministic rate equations and thereby classical kinetics fail horribly by underestimating the reaction rate. This pertains to the case of highly rare bursting, where there is sufficient time in between bursts for the molecule number to return to zero. We note that in this case the production rate of an anabolic commodity could be much higher than expected on the basis of classical kinetics.

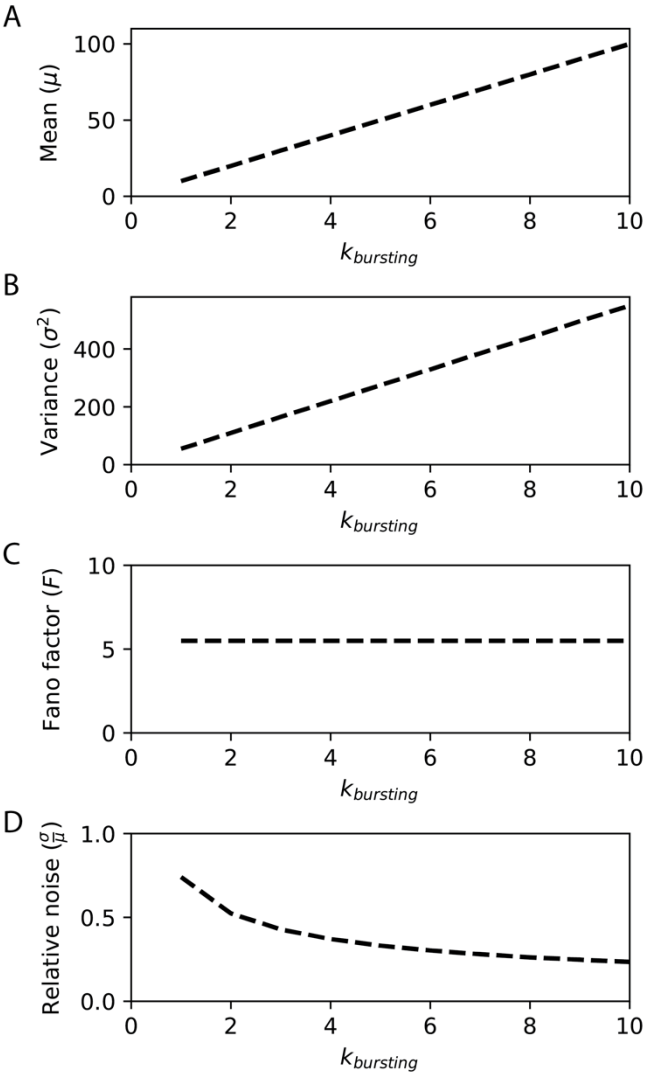

Fig. 3. Various statistical measures of the dispersion of the molecule number of an intermediate in the two-step metabolic network discussed in the text, as function of increased bursting frequency at constant burst size. Results were obtained by a computation using a Gillespie algorithm. The rate constant of bursting was modulated at a constant burst size of 10 and a degradation rate constant of 1. The four panel report variation of the following quantities (A) mean, (B) variance, (C) Fano factor, (D) (relative) noise, with respect to the varying rate constant of bursting.

This could even lead to a violation of the second law of thermodynamics if the latter is formulated in terms of

averages. To illustrate this, we consider the reaction where two molecules of substrate condense to one molecule of product (2*S <=> P) following mass action kinetics. The stochastics does not affect the proportional dependence of the average reverse reaction rate on the concentration of product. Defining true equilibrium as equal average forward ($k_f \cdot \overline{N(N-1)}$) and average reverse reaction rates ($k_r \cdot \bar{P}$) we obtain for the Gibbs energy difference classically defined in terms of averages:

$$\frac{\Delta G_{eq,averages}}{R \cdot T} \stackrel{\text{def}}{\equiv} ln\left(\frac{k_r \cdot (\bar{P})_{eq}}{k_f \cdot \left((\bar{N})_{eq}{}^2\right)}\right)$$
$$= ln\left(\frac{k_r \cdot (\bar{P})_{eq}}{k_f \cdot \overline{N_{eq} \cdot (N_{eq}-1)}}\right) + ln\left(\frac{\overline{N_{eq} \cdot (N_{eq}-1)}}{(\bar{N}_{eq})^2}\right)$$
$$= ln\left(\frac{v_r}{v_f}\right) + ln\left(\frac{\overline{N_{eq} \cdot (N_{eq}-1)}}{(\bar{N}_{eq})^2}\right)$$
$$= ln\left(\frac{\overline{N_{eq} \cdot (N_{eq}-1)}}{(\bar{N}_{eq})^2}\right)$$
$$= ln\left(1 + \frac{F-1}{\bar{N}}\right) > 0 \; for \; F > 1$$

Where N refers to the number of molecules of S and P to the number of molecules of P. $\bar{S}$ and $\bar{P}$ refer to the average molecule number of S and P respectively, divided by the Volume. This equation shows that for a Fano factor in excess of 1 such as may arise from bursting, the apparent free energy of the reaction as defined in terms of the average concentrations of the substrate and product, will exceed 0, even though there is no second reaction driving the conversion of 2S to P. This contradicts the second law of thermodynamics. At equilibrium the average product concentration is predicted to be higher than in the classical case:

$$\frac{(\bar{P})_{eq}}{\left((\bar{N})_{eq}{}^2\right)} = \frac{k_f}{k_r} \cdot \left(1 + \frac{F-1}{\bar{N}}\right) = K_{eq} \cdot \left(1 + \frac{F-1}{\bar{N}}\right)$$

The excess free energy in the substrate derives from the non-Poisson nature of its probability distribution, which must itself be maintained by a non-equilibrium process (Westerhoff and Kamp, 1983). This does not detract however from the issue that the second law of thermodynamics expressed in terms of averages is violated in this case.

We propose to define the true ΔG as follows:

$$\frac{\Delta G_{eq,statistical}}{R \cdot T} \stackrel{\text{def}}{\equiv} ln\left(\frac{v_r}{v_f}\right) = ln\left(\frac{k_r \cdot (\bar{P})_{eq}}{k_f \cdot \overline{N_{eq} \cdot (N_{eq}-1)}}\right)$$

and this $\Delta G_{eq,statistical}$ does equal zero at equilibrium.

The average stochastic rate equation in terms of average molecule numbers and Fano factor that we developed above, should be one in a larger set that should apply to dispersed biological production systems. This set of new kinetic and thermodynamic rate expressions, in terms of averages and variances and perhaps higher order moments of the probability distributions, now needs to be developed for Biology.

3. DISCUSSION

In this paper we have revived non-equilibrium thermodynamics, now in view of a number of relevant recent discoveries in Biology. One has been that of metabolic networks that enable various alternative routes and a concomitant variation of stoichiometry at which a desired commodity (such as ATP, biomass or a metabolic product) is produced. We found that variation of the stoichiometry, akin to gear shifting, should enable an organism to produce its commodities at higher rates. For one particular condition of anabolic free energy, an organism may have set its network routing such as to achieve the optimal stoichiometry. When conditions are varying such that also the anabolic free energy is affected, e.g. in case of nitrogen starvation setting in, metabolic rerouting such that the stoichiometry shifts, may be advantageous. If such variation in conditions occurs frequently, a continuously varying stoichiometry in accordance with the variomatic principle developed here, might be best: We showed that there should be an optimal mode of rerouting flux through the metabolic network, corresponding to varying the stoichiometry continuously, such that anabolic flux should always be maximal. We call this the 'variomatic' mode. Reroutings of this type have been observed in various cell types, but it remains to be examined whether they effect this optimally.

Cell populations have been observed to be highly heterogeneous, with numbers of molecules of important biomolecules being much more diverse than in accordance with classical statistical thermodynamics. We have shown that bursting reactions may be at the origin of this. When burst sizes remain low as compared to average molecule numbers, kinetic and thermodynamic behavior of the cell populations should remain close to normal. When high molecule numbers are essentially achieved by high burst sizes however, the situation becomes different dramatically, with classical kinetics and thermodynamics failing.

In principle the phenomena of bursting and cell-cell heterogeneity in biological systems may bring such systems outside the realm of classical understanding. In this paper we have set steps towards new types of non-equilibrium thermodynamics and kinetics that are able to deal with the new realm. These steps should be useful in that they continue to enable to describe systems in terms of fewer than all their kinetic and stochastic details, i.e. in terms of thermodynamic averages and variances.

REFERENCES

Keizer, J. (1987). Statistical Thermodynamics of Nonequilibrium Processes, Springer, Berlin.

Mondeel, Thierry D.G.A., Rehman, S., Zhang, Y., Verma, M., Dürre, P., Barberis, M. and Westerhoff, H.V. (2016). Maps for when the living gets tough: Manoeuvring through a hostile energy landscape. *IFAC papers on line* 49, 364–370.

Nicolis, G. and Prigogine, I. (1977). *Self Organisation in Non Equilibrium Systems*. Wiley, New York.

Stucki, J.W. (1980). The Thermodynamic-Buffer Enzymes. *European Journal of Biochemistry 109*, 257–267.

Van Kampen, N. G. (1976). *Advances in Chemical Physics* 34, 245.

Westerhoff, H. V, & Chen, Y. D. (1984). How do enzyme activities control metabolite concentrations? An additional theorem in the theory of metabolic control. *European Journal of Biochemistry 142*, 425–30.

Westerhoff, H.V. and Kamp, F. (1987). Maxwell's demons in channeled metabolism: Paradoxes and their resolution. In*:* Welch, GR and Clegg, J.S. (eds.) *The Organization of Cell Metabolism*, pp. 339-355. Springer, Boston, MA.

Westerhoff, H.V. and Van Dam, K. (1987). Thermodynamics and Control of Biological Free-Energy Transduction. Elsevier, Amsterdam.

Zhang, Y., Kouril, T., Snoep, J.L., Siebers, B., Barberis, M. and Westerhoff, H.V. (2017). The peculiar glycolytic pathway in hyperthermophylic Archaea: Understanding Its whims by experimentation *in silico*. *International Journal of Molecular Science 18*, 876.