16th European Symposium on Computer Aided Process Engineering
and 9th International Symposium on Process Systems Engineering
W. Marquardt, C. Pantelides (Editors)

Systems Biology and the Silicon Cell: Order out of chaos

Hans V. Westerhoff

Manchester Center for Integrative Systems Biology, The University of Manchester, UK, EU
and
Molecular Cell Physiology, FALW, Free University, BioCenter Amsterdam, Amsterdam, The Netherlands, EU

Abstract

This paper gives an overview of the thermodynamics and kinetics background to the silicon cell (SiC!) approach. SiC! makes precise mathematical models of components of systems inclusive of their interaction properties. It then puts these component models together into a computer program and integrates the behavior. For metabolic pathways, SiC! takes the ensembles of enzyme molecules as the components. It takes the ensemble averaged metabolite concentrations as the dependent variables these components work on. We show how this approach depends on principles of non equilibrium thermodynamics and kinetics.
Metabolic control analysis is an early and characteristic approach to systems biology. Using silicon cells one can do this control analysis *in silico*. Also this analysis also has a number of theoretical foundations, which are again close to those of non equilibrium thermodynamics. We propose that Metabolic Control Analysis is in fact the extension from equilibrium thermodynamics to non equilibrium systems that so many searched for in the second half of the previous century.

1. Non-equilibrium thermodynamics

In 1931 (1,2) Onsager published two seminal papers. They revealed that there should be a remarkable symmetry in cross-cause effects relationships in coupled processes. To obtain the symmetry property, coupled processes have to be described in a certain way, a way that has since been called non equilibrium thermodynamics (3). Describing each process in terms of a driving force equal to the free energy difference across that process, and a flow equal to the steady state rate of the process, the cross dependence of the two processes on the two forces had to be equal in the limit to equilibrium.
The proof given was based on kinetics or a probabilistic version thereof, and therewith married mass-action kinetics with thermodynamics. Yet, it depended on the generic rather than the specific aspects of the kinetics and was therewith mechanism independent. Because this was also true for equilibrium thermodynamics, this mechanism independence was long thereafter considered an essential property, also of non equilibrium thermodynamics.

This non equilibrium thermodynamics (NET) was often formulated as a systems of linear equations relating all steady state fluxes in the system to all thermodynamic forces through proportionality relations, for which the matrix of proportionality constants then had to be symmetrical for the Onsager reciprocity relations to be satisfied. Because Biology tends to look at functional processes that

involve a number of coupled molecular processes, many biologists and biophysicists were attracted to this non equilibrium thermodynamics (4).

Non equilibrium thermodynamics was also useful to Biology because it helped resolve the Schroedinger paradox (5). This paradox held the development of the order and structure of well developed biological organisms out of unordered food supplies, to be in conflict with the second law of thermodynamics. The usual formulation of this law in physics is that entropy can only increase, never decrease, where entropy s a measure of chaos. Non equilibrium thermodynamics then served to resolve this paradox, by reformulating that really the entropy production needed to be positive; by exporting lots of entropy, an organism could actually increase its order (3, 5).

The entropy production function then became object of additional searches for general thermodynamic principles. Prigogine and coworkers showed that in the limit to equilibrium entropy production should be minimal at steady states (3). The minimum was with respect to variation of the independently variable thermodynamic forces. Entropy production was not minimal with respect to systems parameters (6), but again there was little interest in those systems parameters as they would carry mechanism specific information. These derivatives were thought not to lead to general results therefore.

Understanding the coupling between processes in bioenergetics was an area where NET had some additional useful contributions. It enabled the definition of a coefficient that could quantify the degree of coupling between distinct biochemical processes (7). Defining this coefficient increased the awareness that coupling would not have to be complete, and that uncoupling or slippage should be a possibility. Up to that time and also subsequently, the unfounded notion that biological systems were necessarily ideal and therefore would not waste any free energy, made biologists only consider networks where coupling would be complete. Here the emergence of the chemiosmotic coupling mechanism was important. In this mechanism a membrane that was likely to have at least some passive permeability for protons was supposed to sustain the relevant free energy intermediate, i.e. the transmembrane electrochemical potential difference for protons (8). This mechanism was one of the early examples of systems biology, where only through the integration of at least two completely different types of processes (i.e. transport and chemistry) free energy could be transduced, between two chemical processes.

Further consideration of the degree of coupling in terms of how its magnitude could contribute to the partly coupled process being optimal for certain functions, led to the conclusion that neither the degree of coupling nor the thermodynamic efficiency needed to be maximal for a number of relevant output functions to be optimal (9). Indeed it was calculated that many biological processes, including microbial growth (10) were highly inefficient, where some of the observed efficiencies could be understood in terms of the system being optimal with respect to both growth rate and power production in terms of biomass.

Non equilibrium thermodynamics continued to be successful in non biological sciences where it helped explain cross-correlations between different types of phenomena, such as heat conductance and volume flow. Notwithstanding its apparent ability to function as an early systems biology approach being able to integrate multitudes of processes in its symmetrical linear equations, NET did not develop much further however. The reason was that much of what had been accomplished was valid only for processes that were less than a couple of kJoules per mole displaced from equilibrium. Biological reality is that the free energy of hydrolysis of ATP exceeds 40 kJ/mol, and the dissipation of free energy in many processes exceeds 10 kJ/mol (10).

Therewith none of the proofs of the above principles derived by non equilibrium thermodynamics holds for many realistic biological systems and indeed there is some evidence that the relations themselves do not hold either (10).

Rottenberg (11) and subsequently we (12, 10) then retraced some steps of NET and realized that one could translate well-accepted kinetic relationships into non equilibrium flow-force relationships. This led to the discovery that there was a basis for the linear flow-force relations often postulated for non equilibrium thermodynamics. That linearity was likely to be at a range away from equilibrium that was most relevant for the regulation of processes. However, in that range there needed be no Onsager reciprocity (10), continuing to take away the basis of the validity of the minimum entropy production principle (6).

Importantly, here the paradigm was left that by definition non equilibrium thermodynamics should be devoid of mechanisms; the coefficients relating flows and forces were expressed into enzyme kinetic properties. And, using this new, 'Mosaic Non Equilibrium Thermodynamics (MNET)', the systemic implications for failing mechanisms of coupling could be predicted (10). A systems Biology approach, relating important systems function to molecular action and properties, had been born, *avant la lettre*.

Paradoxically, another, in fact older, branch of non equilibrium thermodynamics thrived on the non-linearities in and amongst the processes in biology, and certainly on the substantial distance of many biological systems from equilibrium. The self organization addressed by this type of non equilibrium thermodynamics cannot occur in the Onsager domain where flow-force relations are symmetrical (3, 13). The resolution of the Schrödinger paradox described above merely stated that export of entropy could resolve that paradox, but it had not yet been clarified how that entropy export would be coupled to the entropy decrease held characteristic of developmental biology. Mechanisms were sought for pattern formation from initially symmetrical conditions, and found, e.g. by Turing (14, 3, 15). Symmetry breaking in time was also found to occur in chemical reaction schemes and held as model for the cell cycle in living organisms. Further developments included the discovery and analysis of sets of equations that could generate even more complex phenomena such as aperiodic self-excitation and deterministic chaos (16). These analyses brought home the message that for some of these phenomena to occur quite special parameter values were needed. This reinforced the question whether indeed in biological reality those parameter values would reign, or if alternatively completely different mechanisms might be responsible for the observed complex phenomena to occur.

In the mechanisms proposed by the fields of non equilibrium thermodynamics and nonlinear dynamics, there was frequently another limitation, i.e. lack of robustness. Symmetry breaking could occur but the precise version of the asymmetry (e.g. left-right versus right-left) depended on fluctuations and would therefore be random. Yet the observation that our right foot is usually on our right-hand side is quite convincing in showing that actual developmental biology is more robust than this. The argument then became that instead of a fluctuation, a well-controlled external condition would set the symmetry breaking in motion, now reproducibly.

The requirement of such an external ordering factor was in line with the more general observation that the structures of living cells do not arise completely anew in every generation: the replication of DNA is semi-conservative, the plasma membrane of newborns cells are pinched off parts of the plasma membrane of their mother cells, and most of their proteins have been and are being made by ribosomes inherited from the

mother cell. A view, in which biological structure was nothing but a perpetration of a complex dynamic structure that once got into existence, became an option. Meanwhile molecular biology found more and more factors that co-determine molecular biology, and important predictions of the simplest versions of the self-organization theory of segmental organization in Drosophila turned out to be wrong: proteins alternating their expression between segments were not directed by precisely the same promoter elements in the stripe in which they were expressed (17). Self-organization may still play a partial role in developmental biology, but it will be a partial role only. These developments have taught us that the attractiveness of a concept such as self organization should not lead to the non-critical implicit assumption that a process is self-organized. Even though self-organization may be the simplest mechanism for pattern formation in early development, that by itself has no value; there is no place for Occam's razor in Biology. Critical experimental testing is required, probably through detailed modeling and checking whether the predictions made by the model for experimentally determined actual parameters values, are in actual agreement with the behavior of the system. Likewise, hypotheses that developmental processes are due to pre-specification will need to be so concrete as to be testable, or falsifiable in Popper's sense (18).

2. Silicon cells

The suggestion that hypotheses in Biology should be testable and indeed be tested would seem to be superfluous. Would any biologists accept that her/his science should not adhere to the criteria devised for the natural sciences? On the other hand Biology is a complex science and this has had the effect that at the truly biological level, few theories have actually been testable. Because of the complexity and nonlinearity of the networks in biology, the behavior of their components is a strong function of the molecules around them. Accordingly, failure of a set of molecules to act precisely as predicted by a theory, could always be attributed to the presence of an as yet unidentified additional factor, somewhat altering the mechanisms that would otherwise work as proposed. Accordingly many biologists working at the physiological level, are satisfied with theories that allow for exceptions even when if these are not made explicit. Other biologists took the opposite stance. They decided that if at the physiological level theories could not be falsified, they should refrain from working at that level and turn to model systems that were completely controlled, notably *in vitro* systems with purified molecules. There the hard scientific criteria could be met in principle.

Genomics has altered the situation. Now, living systems such as some unicellular organisms, are completely characterizable in terms of the sequence of all their genes, and the concentrations of all mRNAs, proteins and (soon) metabolites. These concentrations can also be manipulated, enabling a large number of independent experimental tests. The physiologist can no longer propose that failure of the system to behave according to his hypothesis is due to an unidentified molecules; if there is such a failure, he should either reject the hypothesis or identify the perturbing molecule and extend his model to incorporate that molecule. The molecular biologist need no longer refrain from studying the actual functioning of his molecules, in the intact system or suitable models thereof. This new interface between molecular biology and physiology is called Systems Biology (19).

Systems Biology focuses on the functional properties that arise in the interactions between the components of biological systems (20). The cell cycling and the self-organization discussed above are examples: none of their molecules cycles or forms spatial patterns in the absence of interaction with the other molecules. Systems biology also realizes that it should reach beyond mathematical biology in that it should not devise models that might explain biological function: it should devise models that do explain those phenomena, for the parameter values that are real.

The silicon cell program (cf. www.siliconcell.net) is an epitome of this systems biology (21). It puts together the actual kinetic and interaction properties of the components of the biological system into a computer replica and then uses a computer program to calculate the system's behavior of that replica. If that behavior corresponds with the experimentally observed functional behavior of the biological system, then the mechanisms present in the replica should be the explanations of the emergent functional behavior.

Other than what their name may suggest, these silicon cells do not yet correspond to replica of entire cells. They correspond to replica of hopefully sufficiently autonomous parts of (pathways in) living cells to be testable. The strictness with which they adhere to the principle of the silicon cell that all component properties should have been determined experimentally, is also variable, but this is to improve in the future. The models in the silicon-cell program have gone through the quality control of international journals, some of which collaborate explicitly with the program.

3. At what level should one pitch the silicon cell?

When pronouncing to make precise models of functioning systems of the living cell in terms of their components, it is not immediately obvious what the components should be. The silicon cell focuses on the whole cell as the ultimate system but begins with the limited focus of pathways in those cells as the systems. The components are the catalysts in the pathway, mostly the proteins, and the 'micro-molecules' ('metabolites') through which they communicate. This does not completely specify yet the level of modelling however. To pitch the right level, both siliconcell and systems biology learn from non equilibrium thermodynamics and kinetics.

One could take the point of view that a precise replica model of what happens in a metabolic pathway in the living cell should consider each individual molecule explicitly in terms of its position, state, appearance and disappearance, and these as functions of time. However, the complexity accompanying such a point of view is unmanageable. Let us consider just 20 types of molecule such as ATP in the living cell. At concentrations of approximately 10 mM these would each number 6 million molecules per *E. coli* cell. Supposing that each of these molecules could be in either of two states and each at any of 500 locations, then the entire systems would have some $1000^{20000000}$ possible states. Modeling how such a system proceeds its biased random walk through these states is not only impossibly time consuming, but it is also useless in terms of the detailed information it would give. We are simply not interested in the particular behavior of such a system; we would not even know whether it corresponds to a particular experimental system we are studying, because we could not know in what precise state that system is. Inevitably we are interested in trends of behavior; in behavior that is reproducible between incarnations of a system of interest, which may all be different in terms of their precise microscopic states but are expected to behave similarly macroscopically. Lack of direct interest is however insufficient

reason not to engage in detailed modeling. The average trend of behavior of a system might depend crucially on the detailed processes and in this case one would need to model in complete detail to obtain the average behavior of interest (22).

What then is the level of detail at which we need to enter the components into the silicon cell? Here statistical mechanics and kinetics have shown the way. Consider the simplest possible reaction *i.e.* the degradation of a molecule A. The probability that the degradation occurs at some location in the cell should be proportional to the probability to find molecule A at that location and the rate should again be proportional to that probability:

$$\bar{v} = \overline{k' \cdot P(1 \mid N = n)} = k \cdot \bar{n}$$

Here P(1|N=n) represents the probability to find a molecule A within a given small time interval at the location of interest, if the total number N of molecules A equals n. For the average rate of the process this implies:

$$\bar{v} = \overline{k \cdot n} = k \cdot \bar{n}$$

which corresponds to the deterministic rate equation for this situation. With this the average behavior of the system is described in terms of the ensemble average concentration (if one also divides by the volume) of molecules of the same type. If the mixing in the system is much faster than the reactions, then that ensemble averaged concentration is the same for the entire cell (or compartment thereof) and this leads to an enormous simplification. Now the state of the system can be described by only 40 state variables, i.e. the ensemble averaged concentrations of the 20 molecules in their two internal states.

The situation becomes more complicated in essence whenever the kinetics is nonlinear. We here take quadratic kinetics as the example:

$$\bar{v} = k' \cdot P(1 \mid N = n) \cdot P(1 \mid N = n-1) = \overline{k \cdot n \cdot (n-1)} = k \cdot (\bar{n})^2 \cdot \left(1 + \left(\frac{\sigma^2 - \bar{n}}{(\bar{n})^2}\right)\right)$$

where σ^2 is the variance in the particle number. This equation shows that only under certain conditions the deterministic rate equation is followed. One is the case where the variance equals the mean, which occurs when the particle number follows a Poisson distribution. Poisson distributions occur in cases with unit stoichiometries (10) and should not be expected to be standard in biological systems. In most systems the variance may not be equal to the average number of particles, but is nevertheless of the same order of magnitude (10). Then:

$$\bar{v} = k \cdot (\bar{n})^2 \cdot \left(1 \pm O\left(\frac{1}{\sqrt{\bar{n}}}\right)\right)$$

This leads to the second condition, which should apply more frequently: deterministic kinetics applies whenever the number of particles exceeds 100.

The above argumentation is classical. Yet we repeat it here for two reasons. First, one now often encounters research programs where modeling is done stochastically rather than by using the deterministic equations, but without rationalization of why the former approach is chosen. At least one should ensure that the particle number is low or the distribution is vastly different from Poisson. Second, a

number of cases have been noted meanwhile where variances have indeed been much larger than the average particle number and some of these cases carry a truly biological signature. An exemplary case is that of the expression of a gene through mRNA to the protein level. Because a single mRNA readily leads to the synthesis of hundreds of protein molecules, a variance in mRNA close to the mean number of mRNAs may translate into variance and average also being similar at the protein level. Then kinetics at the protein level (or for similar reasons at the metabolic level) may differ highly from that predicted by the deterministic rate equation. Ehrenberg and colleagues have been working of highly relevant such cases of extremely large variance (23).

A highly relevant case is of course dictated by the digital encoding of genetic information, allowing one or two copies of a gene per cell only. When mutations occur in the haplotype, the variance is of the order of magnitude of the mean. Then nondeterministic kinetics should be used, not only at the DNA level but also at the mRNA, protein and metabolic levels, when the mutating gene is an enzyme.

In case a population of cells is genetically homogeneous, and the number of mRNA molecules encoding the enzymes is large or quasi-Poisson distributed, cellular processes will follow deterministic kinetics. It is these cases that the silicon cell approach has been limiting itself to until now (21). Hence the silicon cell approach describes the processes in the cell as processes that are carried out by enzymes the activity of which can be described by their ensemble-averaged activity, and their ensemble averaged kinetic properties which depend on the ensemble averaged concentrations of micromolecules ('metabolites'). This limiting case is the same as the one proposed by non equilibrium thermodynamics (15) for the description in terms of average concentrations, or in fact chemical potentials:

$$\mu = \mu^{0'} + R \cdot T \cdot \ln\overline{\left(\frac{n}{V}\right)}$$

Here the rate equations becomes

$$\bar{v} = k \cdot e^{2\cdot\left(\mu-\mu^{0'}\right)/RT}$$

where the number 2 refers to the case of quadratic kinetics, and should be replaced by 1 in the case of linear kinetics. Deterministic kinetics and non-equilibrium thermodynamics that is not restricted to the near equilibrium domain are really two expressions of the same thing.

The advantage of the deterministic kinetics/non-equilibrium thermodynamic approach is the tremendous simplification. For the 20 types of molecules that can each occur in two states and at 500 locations in the cell, the number of state variables is now 20 000, which although large is no longer unmanageably large. In practice a further simplification is possible provided the situation is that of a reasonable homogeneous space, diffusion being much more rapid than reactions, or the enzymes being distributed homogeneously over space. Then only 40 state variables suffice. The ensemble-averaged concentrations or the corresponding chemical potentials, correspond to the functions of state of thermodynamics, adding to energy content and volume for isothermal, isobaric systems (10).

Indeed, at this moment all silicon cells are spatially homogeneous within well-defined compartments and the following simple description is used (www.siliconcell.net). For each process that occurs in a cellular compartment, one formulates what it actually does. This is the transformation of molecules of one chemical nature to

molecules of a different chemical nature, as indicated by the reaction equation. In the case of transport, the molecules can be of the same chemical nature but in different explicit compartments. The reaction equation can also be encoded as a set (vector) of stoichiometries (positive for products and negative for substrates). This vector becomes a column of the stoichiometry matrix N, of which each column corresponds to a process in the cell. Often the stoichiometry matrix is formulated from the alternative point of view of the time dependence of the metabolite concentrations as a balance between of all the process rates. The end result is the same, but the former method is more in keeping with the silicon cell philosophy that the process should be independent of any knowledge about the system.

Processes are not only characterized by what they do, but also by the rate at which they do it, and by the dependence of that rate on the state of the system, i.e. the concentrations of the molecules. For each process therefore an enzyme kinetic rate equation is formulated, which is typically of the form:

$$v = g(e) \cdot f(S, X, Y, K_{eq}, k_{cat}, K_S, K_P, K_X, ...)$$

Where g(e) is often a mere proportionality, indicating that the rate is proportional to the concentration of the catalyst. Often g(e) and k_{cat} are combined into the single parameter V_{max}. Usually, X and Y are variable metabolite concentrations, corresponding to functions of state of the system, as discussed above.

When the list of all processes in the system has been compiled with heir stoichiometric and rate equations, the lists of the arguments of the functions in the rate equations contain two types of properties. The one type is that of the variables. These also occur in the lists of molecules produced or consumed by the processes. For these variable properties balance equations are then written using the expression:

$$\frac{dX}{dt} = N \cdot v$$

Where N is again the stoichiometry matrix, v is a vector of the process rates, and X is a vector of all the concentration variables. The other type of properties in the lists of the arguments of the rate equations is called parameters. The parameters are not altered by actions of the processes in the system studied, but set by external conditions or by properties that cannot be changed by the systems (e.g. the Michaelis constants of the enzymes, and sometimes the pathway substrate and product, S).

This is almost (see below) all the biologist/biochemist formulating a siliconcell does: characterize the components of the system. The computer program does the rest, which importantly includes the computation of the system behavior. It integrates the set of differential equations:

$$\frac{dX}{dt} = N \cdot diag(g(e)) \cdot f(S, X, Y, K_{eq}, k_{cat}, K_S, K_P, K_X, ...)$$

Where the biologists still has to specify the initial conditions. The specification of these is actually something that requires some knowledge about the system, but not knowledge on how and why it behaves. Alternatively, one is interested in the steady state and asks the computer to solve equations for time independence of the metabolite concentrations. These two options are available for just a click on the siliconcell model base of live models: http://www.jjj.bio.vu.nl A third option of this silicon cell live 'modelbase' calculates control coefficients, *i.e.* the dependence of steady state properties on all the process activities. Through the world-wide web anyone can now engage in *in silico* experimentation with refereed silicon-cell models of pathways of living organisms.

With this the silicon cell comes to predictions and descriptions of systems behavior at the functional state, *i.e.* away from equilibrium, in terms of its thermodynamic properties, i.e. chemical potentials or ensemble-averaged concentrations, and ensemble averaged fluxes. This was one of the aims of non equilibrium thermodynamics.

4. New non equilibrium thermodynamics?

Yet, what the siliconcell delivers may not be quite recognized as non-equilibrium thermodynamics, as it is always rather specific information which depends on the precise magnitudes of the kinetic parameters. No principles, no laws of general validity are produced by the siliconcell. Indeed, precisely because it aims at generating a computer replica of the real pathways, no reduction of complexity, no generalizations are produced without further activities. In addition the output is formulated in terms of fluxes and concentrations rather than in terms of chemical potentials.

For a while it seemed that perhaps biological systems lack general principles other than the ones valid close to equilibrium and discussed above. Few if any of the properties and principles valid near equilibrium could be extrapolated successfully to systems displaced from equilibrium to the extent that regular biological systems are.

In the late sixties of the previous century in Edinburgh (24) and Berlin (25) a new way of looking at biological systems came about, partly inspired by genetics and partly by silicon cell type of modeling of metabolism. The new approach was called metabolic control analysis (MCA). Until 1987 (10), little reference to a link between MCA and non equilibrium thermodynamics was made, even though the latter discipline was still in development.

Yet, even though this was not agreed on by its original progenitors, metabolic control analysis led to new laws and principles for biological systems, and especially for networks of biochemical reactions. We shall here discuss the summation laws of metabolic control analysis from this perspective. To do this we shall first retrace our steps and recall the derivation of the Gibbs equation and the Gibbs-Duhem equations of equilibrium thermodynamics. We first recall the balance equation for ordinary energy U, which reads as follows (15, 10):

$$dU = d_e U = d_e Q + d_e W + \sum_{j=1}^{n} \mu_j \cdot d_e n_j$$

Subscript e refers to exchange of the system with its environment. The first law of thermodynamics has been used here so as to require that no energy U can be produced or consumed. Accordingly energy in the system can only increase by the addition of heat, work or chemicals from the outside, where the latter carry a partial molar energy equal to their chemical potential. The addition of heat is equal to the reversible addition of entropy (exclusive of the entropy carried by the molecules) and volume (exclusive of the volume increase due to the addition of the molecules):

$$dU = T \cdot d_e S - P \cdot d_e V + \sum_{j=1}^{n} \mu_j \cdot d_e n_j$$

Assuming that the system is at equilibrium no entropy is produced. Because then also chemical reactions inside the system are absent, or their total contribution equals zero, internal volume changes are absent, this equation becomes the Gibbs-Duhem equation:

$$dU = T \cdot dS - P \cdot dV + \sum_{j=1}^{n} \mu_j \cdot dn_j$$

This equation shows that energy U can be calculated from its initial value and the changes in entropy, volume and molecule numbers, provided that temperature, pressure and chemical potentials are known. We shall here consider isothermal, isobaric systems, freely exchanging matter with their environment which has constant chemical potentials for all molecules. T, P and chemical potential are intensive properties and energy, entropy, volume and molecule numbers are extensive properties, meaning that the latter do not change, and the latter do change proportionally with the size of the system. In other words, when changing the size of the system by the factor λ:

$$U(\lambda^0 \cdot T, \lambda^1 \cdot S, \lambda^0 \cdot P, \lambda^1 \cdot V, \lambda^0 \cdot \mu, \lambda^1 \cdot n) = \lambda \cdot U(T, S, P, V, \mu, n)$$

Or in other words energy U is a homogenous function of order 1 of entropy, volume and molecule number (and of order zero of Temperature, pressure and chemical potential). Euler's theorem then rules that:

$$1 = \frac{\partial \ln U}{\partial \ln S} + \frac{\partial \ln U}{\partial \ln V} + \sum_{j=1}^{n} \frac{\partial \ln U}{\partial \ln n_j}$$

The partial derivatives are given by the Gibbs-Duhem equation and inserting these, one obtains the Gibbs equation:

$$U = S \cdot T - P \cdot V + \sum_{j=1}^{n} n_j \cdot \mu_j$$

This inspired Gibbs to define the Gibbs free energy, as:

$$G \equiv U + P \cdot V - T \cdot S$$

Which then leads to:

$$G = \sum_{j=1}^{n} n_j \cdot \mu_j$$

Establishing the chemical potential also as the partial molar free energy and a the energy function of interest for isothermal, isobaric systems exchanging matter with their environment.

The functions of state entropy, volume and molecule number describe a system that is at equilibrium, and only partly systems that are away from equilibrium. For the latter systems the aspect of time or fluxes is missing. When searching for thermodynamic descriptions of systems away from equilibrium, it may be useful to consider the phenomena that keep the system away from equilibrium. These are the Gibbs energy dissipating processes, and more precisely the activities of these. In biochemical networks, virtually all these processes have material counterparts, *i.e.* the enzymes that catalyze them. These are in turn encoded by genes, constituting a further relationship with nucleic acids. The properties of biochemical networks at steady state can be considered functions of all process activities (here denoted by e_i) and many other properties. For such a property Z we write:

$$Z = z(e_1, e_2, ..., e_n, S, P, T, K_M, K_{eq}, k_{cat}, ...)$$

where one recognizes all the parameters of the enzyme kinetic rate equations. Z refers to a function that delivers the steady state value of Z. S refers to the concentrations of

external pathway substrates and products, which are parameters to the system of interest.

Above we transformed the system in terms of physical size, essentially by copying it lambda times, we here consider a transformation in terms of all activities: we consider the situation that all processes are simultaneously accelerated by the factor λ. Because we are at steady state, all processes should be balancing each other, such that there are no changes in time anymore of any property. As all processes are activated by the same factor this balance will be maintained, and nothing will change, except that all processes will proceed λ times faster. If Z refers to a flux, e.g. J, this implies that Z is a homogenous function of order 1 of all the processes rates. Using Euler's theorem one then obtains (10, 26):

$$1 = \frac{\partial \ln J}{\partial \ln e_1} + \frac{\partial \ln J}{\partial \ln e_2} + \frac{\partial \ln J}{\partial \ln e_3} + ... = C_1^J + C_2^J + C_3^J + ...$$

Where the coefficients denoted by capital C correspond to the flux control coefficients of MCA (10). One may here recognize the well-known summation law for flux control coefficients (24, 25). Similarly, realizing that the steady state concentrations are not changed, one sees that these are zero order homogeneous functions leading to the concentration control summation law:

$$0 = \frac{\partial \ln X}{\partial \ln e_1} + \frac{\partial \ln X}{\partial \ln e_2} + \frac{\partial \ln X}{\partial \ln e_3} + ... = C_1^X + C_2^X + C_3^X + ...$$

We here have two fundamental laws of non equilibrium biochemical networks that have been derived in much the same way as the Gibbs equation was derived in equilibrium thermodynamics. We therefore propose that the summation laws are aspects of the non equilibrium thermodynamic theory that was long sought after.

The concentration-control coefficients are the derivatives of the ensemble-averaged concentrations of the substances in the system with respect to the process activities. Because of the definition of the chemical potential, the concentration summation law can also be written as:

$$0 = \sum_{j=1}^{n} C_j^{[X]} = \sum_{j=1}^{n} \frac{\partial \mu_X}{\partial \ln e_j}$$

which now also shows as a thermodynamic law for non equilibrium steady state. The logarithm of the enzyme activity could also be written as the chemical potential of the enzyme:

$$0 = \sum_{j=1}^{n} C_j^{[X]} = \sum_{j=1}^{n} \frac{\partial \mu_X}{\partial \mu_{e_j}}$$

5. Discussion

We here discussed fundamental aspects surrounding the silicon cell approach. These are related to statistical thermodynamic properties of biological systems. The present silicon cell approach is suited for biochemical, signal-transduction, and gene-expression networks that fulfill a number of conditions. These entailing that the fluctuations in them are limited or follow the Poisson distribution. This assumption corresponds to the one required for the use of deterministic rate equations and is therefore quite acceptable for most biochemical networks.

Interestingly, the realization that these assumptions need to be made, suggests that the silicon-cell approach may be one of the types of approaches that non-equilibrium thermodynamics was looking for: it describes systems away from equilibrium. It does this by inserting the details of every molecular phenomenon at the ensemble level of enzyme-catalyzed reactions. The results of silicon-cell calculations are thereby also highly specific, *i.e.* they give the concentrations of all molecules as a function of time and all precise parameter values. Although the silicon cell therewith describes non-equilibrium processes, it may not qualify as thermodynamics, because it lacks any aspect of generality.

Parts of metabolic control analysis (MCA) on the other hand do describe non equilibrium systems in terms of generic properties. We have here shown that important principles of MCA are analogous to principles derived in equilibrium thermodynamics, and so are the derivations of these principles. The same may be true for the Hierarchical Control Analysis (27), which generalizes MCA to systems with signal transduction (28) and systems with variable gene expression. MCA also has other, famous laws/principles, i.e. the connectivity theorems. Also these have strong thermodynamic connotations including an origin in stability *vis-à-vis* fluctuations (10).

We here therefore postulate that HCA and MCA correspond to the non equilibrium thermodynamics that is most suited for most biological systems. We expect that taking this perspective and that of the silicon cell, and combining these more with thermodynamic considerations, even more new systems biology principles will emerge.

6. Acknowledgements

We thank the EU-FP6 program, NWO, Ecogenomics-EZ, IOP-genomics, for supporting our Systems Biology work, enabling us to obtain deeper insights also into the scientific position of systems biology.

References

1. Onsager L. (1931) Phys. Rev. 37, 405 - 426
2. Onsager L. (1931) Phys. Rev. 38, 2265 - 2279
3. Glansdorf P. and Prigogine I. (1971) *Thermodynamic Theory of Structure*, Stability and Fluctuations, John Wiley & Sons, New York,
4. Katchalsky A. & Curran P.F. (1967) Non-equilibrium thermodynamics in biophysics, Harvard University Press, Cambridge MA
5. Schroedinger, E. 1944. *What is Life? The Physical Aspect of the Living Cell*. Cambridge: Cambridge University Press.
6. Juretic D., and Westerhoff H.V. (1987) Biophys. Chem. 28, 21-34.
7. Kedem O. & Caplan S.R. (1965) Trans. Faraday Soc. 21, 1897 – 1911.
8. Mitchell P (1961) Nature 191, 144 – 148.
9. Stucki J.W. (1980) Eur. J. Biochem. 109, 257 – 283.
10. Westerhoff H.V. & Van Dam, K. (1987) Thermodynamics and Control of Biological Free-Energy Transduction, Elsevier, Amsterdam
11. Rottenberg H. (1973) Biophys. J. 13, 503 –511.
12. Van der Meer R., Westerhoff, H.V. & Van Dam, K. (1980) Biochim. Biophys. Acta 591, 488 – 493.
13. Cortassa S., Aon M.A., and Westerhoff H.V. (1991) Biophys. J. 60, 794 - 803
14. Turing A.M. (1952) Philos. Trans. R. Soc. Lond. B. 237, 37-72.

15, Nicolis G & Prigogine I (1977) Self-Organization in nonequilibrium systems, Wiley & Sons, New York.
16. Bakker R, Schouten JC, Giles CL, Takens F, & van den Bleek CM. (2000) Neural Comput. 12, 2355-2383.
17. Lawrence P.A. (1972) The making of a fly: (the genetics of animal design), Blackwell Science.
18. Popper K. (1963) Conjectures and Refutations, Routlegde and Kegan Paul, London
19. Westerhoff H.V. and Palsson B.O. (2004) Nature Biotechnol. 22, 1249-1252.
20. Alberghina L. & Westerhoff H.V. (2005) Systems Biology, Definitions & Perspectives, Springer
21. Snoep JL, Bruggeman F, Olivier BG & Westerhoff HV. (2006) Biosystems. 83, 207-216.
22. Westerhoff H.V., and Chen Y. (1985) Proc. Natl. Acad. Sci. USA 82, 3222 - 3226.
23. Elf J & Ehrenberg M. (2003) Genome Res. 13, 2475-2484.
24. Kacser H. & Burns, J. (1973) In Rate control of biological processes (Davies, D.D., ed.) pp. 65 – 104, Cambridge University Press.
25. Heinrich R & Rapoport T.A. (1974) Eur. J. Biochem. 42, 89 – 105.
26. Giersch C. (1988) Eur J Biochem. 174, 509-513.
27. Westerhoff H.V., Koster J.G., Van Workum M., & Rudd K.E. (1990) In Control of Metabolic Processes (Cornish-Bowden A., ed.), pp. 399 - 412, Plenum, New York.
28. Kahn D., & Westerhoff H.V. (1991) J. Theor. Biol. 153, 255 - 285.

.