Networks in Molecular Evolution

PETER SCHUSTER & PETER F. STADLER

Institut für Theoretische Chemie und Molekulare Strukturbiologie Universität Wien, Währingerstrasse 17,A-1090 Wien, Austria
Tel: ++43 1 4277 52745, Fax: ++43 1 4277 52793, Email: studla@tbi.univie.ac.at
Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe NM 87501

Networks are a common theme at all levels of molecular evolution: Networks of metastable states and their connecting saddle points determine structure and folding kinetics of biopolymers. Neutral networks in sequence space explain the evolvability of both nucleic acids and polypeptides by linking Darwinian selection with neutral drift. Interacting replicators, be they simple molecules or highly complex mammals, form intricate ecological networks that are crucial for their survival. Chemical reactions are collected in extensive metabolic networks by means of specific enzymes; both the enzymes and the chemical reaction network that they govern undergoes evolutionary changes. Networks of gene regulation, protein-protein interaction, and cell signaling form the physico-chemical basis for morphogenesis and development. The nervous systems of higher animals form another distinct level of network architecture. We are beginning to understand the structure and function of each of the individual levels in some detail. Yet, their interplay largely remains still in the dark.

Keywords: Conformational diversity, metabolic networks, molecular evolution, neutral networks, sequence space.

1 INTRODUCTION

Networks of different kinds and sizes appear at different levels of molecular evolution. Networks have a natural representation as graphs or hypergraphs of various kinds. It is not surprising, therefore, that graph theoretical methods are of utmost importance for both, analyzing data and building models. Here we present a brief overview of networks with increasing complexity of the systems to be modeled.

The single biopolymer molecule is considered first. Conformational diversity of biopolymers as encapsulated by a free energy surface allows for straightforward application of tree-like networks representing the saddle hierarchies in conformation space. Replication and mutation are the basis of evolution based on asexual reproduction. The inherent limitations of such dynamical systems are overcome by another class of dynamical bimolecular processes, catalysis and inhibition representing the positive or negative expression of this interaction, respectively. The results are symbiosis and parasitism, where the former can give rise to dynamical networks of overall autocatalytic nature which may develop into functional units at a higher hierarchical level. Neutrality in evolution is a phenomenon that can be understood properly by the consideration of networks in genotype or sequence space. Networks involving the dynamics in entire cells or organisms will be mentioned at the end of our tour through structural and molecular biology.

Conformation spaces of biopolymers are highly complex multidimensional objects. Traditionally focusing on proteins, these objects were first approached in one dimension by considering the reaction coordinates of conformational changes or reversible denaturation. More recently, they were replaced by statistical concepts yielding folding funnels [7] which extend to two dimensions what

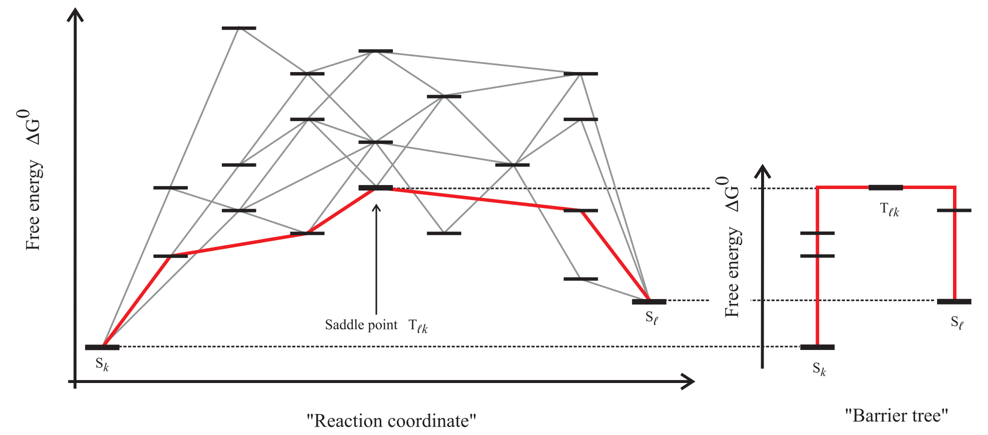

Figure 1: The saddle point that is crossed by the lowest energy path corresponds the the interior vertex in the barrier tree that connects two local minima.

in reality is a truly high-dimensional object. A simplified version of RNA structure allows, for the first time, to compute and to analyze the conformational space in its full complexity. The network of transitions can be simplified and illustrated by barrier trees and saddle point graphs [18].

The simplest examples of molecular evolution are populations of macromolecules that evolve *in vitro*, for example in test tubes. Experiments of this type have been performed since the late 1960s [51], demonstrating that 'naked' RNA molecules can exhibit Darwinian evolution, i.e., the interplay of mutation and selection that leads to adaptation and improvement of fitness, in the absence of cellular life. A theoretical frame for handling evolution of molecules has been developed almost simultaneously through combining the concepts of population genetics with the knowledge of molecular and structural biology [11]. The mechanistic way how autocatalysis shapes transitions from the evolution of molecules towards higher functional units is essentially comprised by the role of catalysis in autocatalytic kinetic networks [15]. Extensive theoretical and computational studies of RNA evolution *in vitro* by the Vienna group (for a recent review see [45]) have revealed the far-reaching consequences of the principles of RNA structure formation for evolutionary phenomena. Neutrality making genotypes indistinguishable for selection has long been postulated in population genetics [37,38] and inferred from empirical molecular sequence data [39]. More recently, using the concept of networks in sequence space, it was shown that neutrality is not only an unavoidable byproduct of molecular replication and mutation but contributes also in an essential way to the efficiency of optimization [20,33].

Metabolic and genetic networks, finally, bring us to the currently hottest topics in molecular genetics. Sequencing of entire genomes, DNA chip technologies, and other methods handling the information of whole cells and organisms produce enormous quantities of data and this wealth of still unexploited information calls for new global and comprehensive concepts of regulation and control. No doubt the solution to the problem lies in a still to develop novel network theory that makes use of modules and hierarchical control elements in order to mediate between partly autonomous local agents and global properties.

2 NETWORKS OF METASTABLE STATES

Let us consider the energy landscape of a particular biopolymer sequence. The "move set" of conformational changes arranges the set of conformations as a graph with edges connecting structures that can be directly inter-converted. This implies a notion of neighborhood that allows to speak of *local minima* of the energy landscape, i.e., of *metastable structures*. A vertex x is a local minimum of all its neighbors, $y \in N(x)$, satisfy $f(x) \leq f(y)$.

The basins of attraction of two metastable states x and y are connected by saddle points, which are defined as the maxima along the paths $\mathbf{p} \in \mathbb{P}_{xy}$ that connect x and y with smallest maximal energy. The height of the saddle point is therefore

$$\hat{f}(x,y) = \min_{\mathbf{p}\in\mathbb{P}_{xy}} \max_{z\in\mathbf{p}} f(z) \tag{1}$$

The saddle points can be arranged in a tree such that the leaves of the tree are the metastable states while the internal nodes corresponds to the saddle points separating the local optima from each other [4, 18, 24], Fig. 1. The height of the energy barrier can be directly read off the barrier tree, see Fig. 2. From the energy barriers aggregate parameters describing the energy landscape can be obtained, such as *depth* and *difficulty*, that play an important role e.g. in the theory of simulated annealing [8], or the fractal dimension of the barrier tree itself [21].

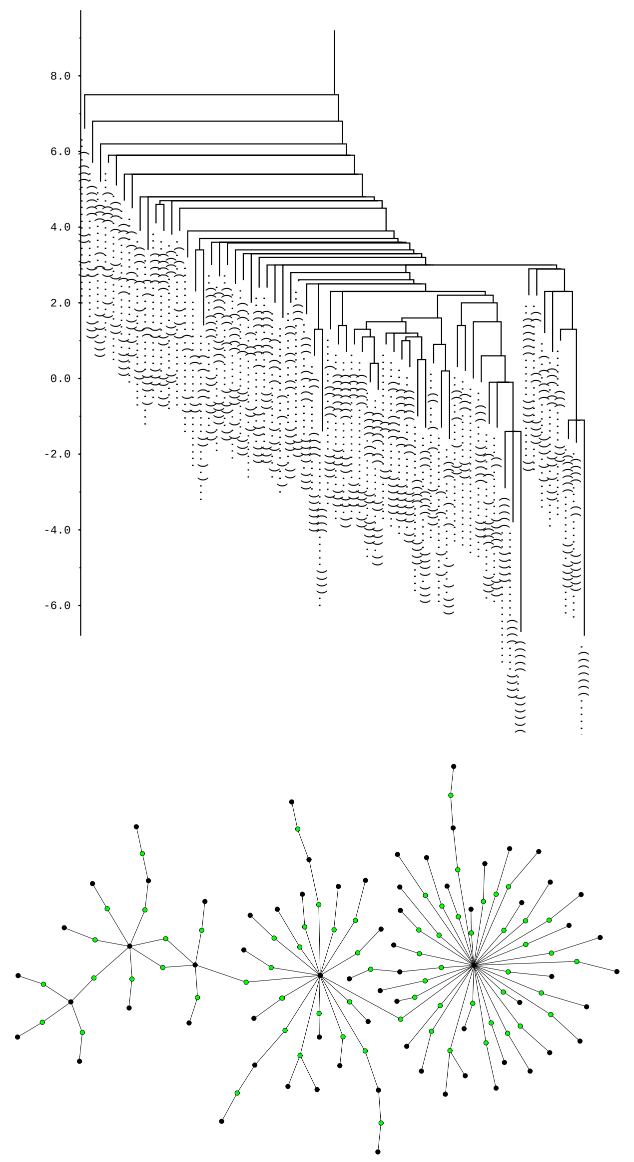

Figure 2: Top: Barrier Tree of an engineered the RNA switch. Bottow: Saddle point network. The sequence of this molecule is ACCGGUUUCUGGUCGAGCUGGCUCG.

We say that a saddle point s *directly* connects x and y if there are two paths of steepest descent that start at neighbor $x', y' \in N(s)$ and end in the local minima x and y, respectively. The metastable states and their directly connecting saddle points therefore form a network that encapsulates the most salient features of the topology of the energy landscape [19]. As an example we show the barrier tree and the saddle point graph of a small RNA switch, i.e., an RNA molecule that may fold into two alternative conformation with (almost) the same folding energy. In case of the molecule in Fig. 2 the saddle-point network is almost a tree. This need not be the case in general, however. Examples of mesh-like saddle point networks arise e.g. in the highly degenerate landscape of certain combinatorial optimization problem such as the *low autocorrelated binary string problem* [19].

The network structure of a biopolymer's energy landscape determines its folding properties. Heuristic notions such as the "folding funnel" [10, 24, 57] postulated for most natural proteins can be described at a microscopic level in terms of network properties. For example, one might speak of a funnel if there are no local optima with significant basins of attraction and/or substantial energy barriers.

3 AUTOCATALYTIC NETWORKS

Differential selection as a consequence differences in replication rates is one of the pillars of the Darwinian paradigm. One of the simplest models deals with polymeric molecules that replicate according to the reaction scheme

$$\mathrm{I}_k + \mathrm{I}_l + (M) \longrightarrow 2\mathrm{I}_k + \mathrm{I}_l \ , \qquad (2)$$

i.e., one polymer species I_k is copied by another polymer species I_l of the same type, thereby consuming building material (M) which is assumed to be present in excess.

Although postulated and analyzed already in the nineteen-seventies [11,14], no direct experimental implementation of this system was possible since the known biopolymers with the required properties were either (obligatory) templates, like RNA molecules, or protein based replicases. But recently an RNA replicating ribozyme was obtained by means of artificial evolution [35], and an RNA assay for the reaction (2) is now within reach. Under the simplest assumptions on the environmental boundary conditions, the replication mechanism (2) leads to the dynamical system [29,30,47,49]

$$\dot{x}_k = x_k\Big(\sum_{i=1}^{n} A_{ki}x_i - \Phi(t)\Big)\ ;\ \Phi(t) = \sum_{i,j} A_{ij}x_ix_j \ . \qquad (3)$$

where x_k is the relative concentration of the species I_k. The flux term $\Phi(t)$ is adjusted such that the replicator equation fulfils the conservation law $\sum_k \dot{x}_k = 0$. Equ.(3) describes also the behavior in a *continuously stirred tank reactor* with sufficiently small influx rates. With properly chosen parameters A_{ij}, replicator equations develop the full scenario of complex dynamics for the minimum number of independent variables, e.g. oscillations for $n = 3$ and deterministic chaos for $n = 4$.

The *second order* replicator equation (3) is invariant under the transformation $A_{kj} \to A_{kj} - A_{kk}$, i.e., the dynamics depends only on the strength of the interactions relative to the self-catalysis. An important class are the so-called *autocatalytic networks* [48] for which $A_{kj} - A_{kk} > 0$. The structure of the network can therefore be depicted by a graph in which the catalytic function is represented by an arrow (Fig.3).

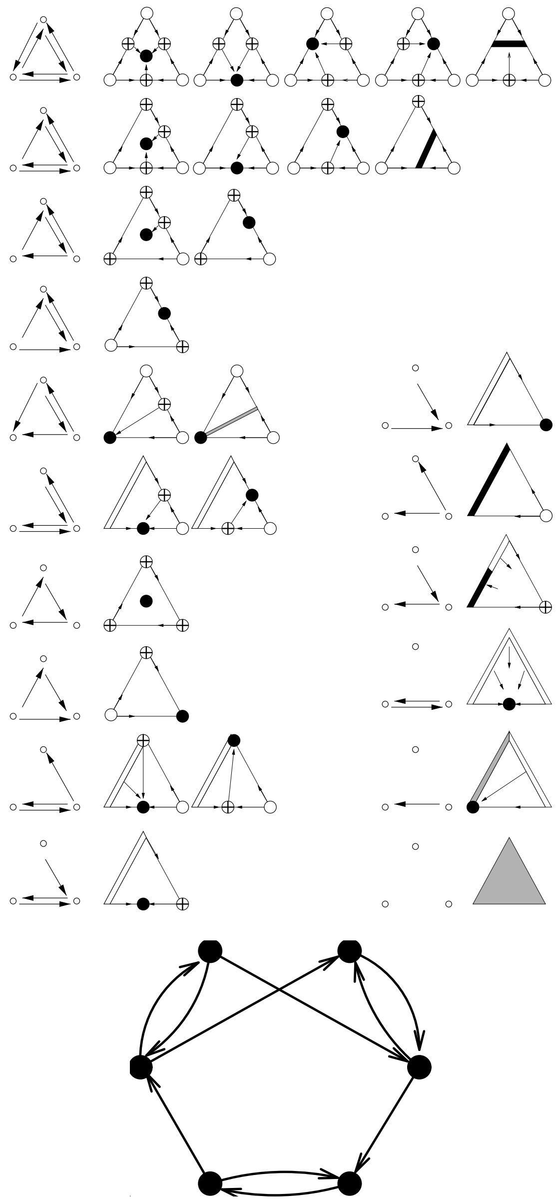

Figure 3: Autocatalytic Networks. *Above:* Phase portraits of the autocatalyic networks with $n = 3$ species. This is only part of the classification of three-species replicator equations [53]. Symbols: sink •, source ∘, saddle ⊕. Lines of fixed points are shaded according to their stability.
Below: Non-Hamiltonian autocatalytic network compatible with permanence [1].

There is a connection between the graph representation of the reaction networks and their dynamical behavior. For example, if the network is *permanent* [31,48] (i.e., no member dies out over time) then the graph is strongly connected [50]. The hypercycle model [14] of cooperating replicators is probably the best-known example of a permanent catalytic network. For $n \leq 5$ the network must be "hypercycle like" by containing a Hamiltonian cycle [1].

Permanent autocatalytic networks have been studies extensively as a possibility to overcome limitations imposed on replication-mutation ensembles by the *error-threshold* of molecular quasispecies model [11–13] which places a restrictive limit on the length of self-replicating polymers in the absence of a sophisticated mechanism that reduces copying errors. These cooperative networks are, however, susceptible to parasites, a threat that can at least in part be alleviated by spatial pattern formation [6].

The networks of ecology behave similar to the autocatalytic networks because the replicator equation (3) and the Lotka-Volterra equations are related by a transformation [28]. Permanence in Lotka-Volterra systems is obviously related to the loss of species in ecosystems. A survey of replicator systems [52] thus implies that permanence is also an exceedingly rare and fragile phenomenon in ecological networks that requires very special interaction structures. Those natural ecosystems which are known to be stable over long periods thus owe their stability to an efficient evolutionary selection process.

4 NEUTRAL NETWORKS IN SEQUENCE SPACE

The evolutionarily relevant relation between genotypes I and phenotypes S can be visualized as a mapping from sequence space, $\mathcal{Q}_n$ with n being the sequence length, into a space of phenotypes, $\mathcal{S}$. Phenotypes are evaluated through selection by a differential fitness criterion f. Accordingly, a mapping from phenotype or shape space into the real numbers completes the weighting of genetic variants (Fig.4). Distance between two genotypes I$'$ and I$''$ of length n is readily expressed by means of the Hamming metric $d^h(\mathrm{I}', \mathrm{I}'')$, whereas the derivation of a distance in shape space $(d^s(S', S''))$ that is useful for understanding optimization and adaptation is much more subtle. Formally, the two maps can be written as

$$\psi : (\mathcal{Q}_n; d^h) \rightarrow (\mathcal{S}; d^s) \qquad S = \psi(\mathrm{I}) \qquad \text{and} \qquad (4)$$

$$f : (\mathcal{S}; d^s) \rightarrow \mathbb{R} \qquad f = f(S) \qquad (5)$$

As indicated in Fig. 4 there is ample evidence that for redundancy in genotype-phenotype maps in the sense that many genotypes cannot be distinguished by a evolutionarily relevant coarse grained notion of phenotypes which, in turn, give rise to fitness values that cannot be faithfully separated through selection. Such indistinguishable genotypes form graphs in sequence space called **neutral networks** [46] when the genotypes are interpreted as nodes and all Hamming distance one neighbors are connected by an edge. In other words, the neutral network of phenotype S in $\mathcal{Q}_n$ is the graph $\mathcal{G}(S)$ formed on the preimage $\psi^{-1}(S)$ by means of the Hamming distance one criterium for edges. The quantity that allows for global

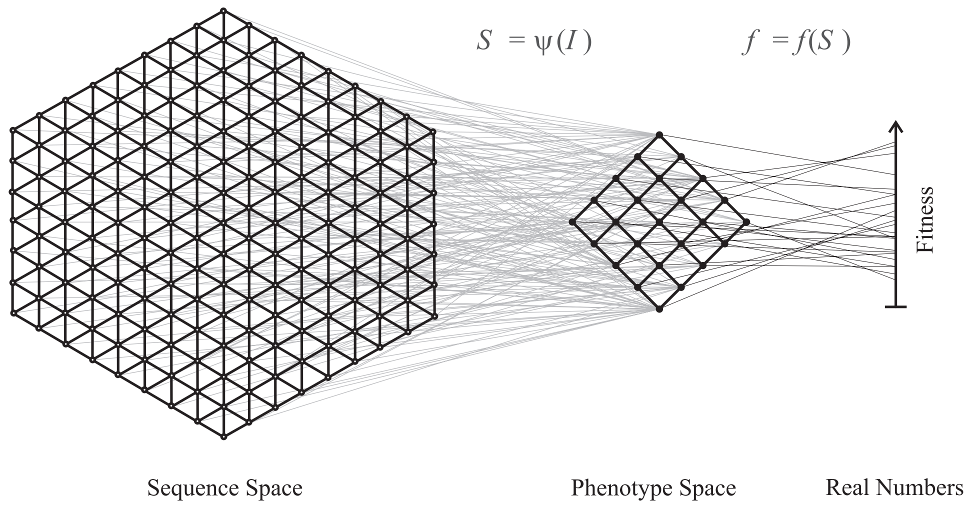

Figure 4: Mapping from sequence space into phenotype space and into fitness values. Evaluation of genotypes through variation and selection is performed in two steps: (i) Development of phenotypes through unfolding of genotypes and (ii) assignment of fitness values to individual phenotypes. The first step is modelled by means of a many-to-one mapping.

characterization of neutral networks is the fraction of neutral Hamming distance one neighbors averaged over all members of the net called the degree of neutrality $\bar{\lambda}$. Neglecting the influence of the distribution of neutral sequences over $\mathcal{Q}_n$, the degree of neutrality will be the higher, the larger the pre-image is.

Generic properties of neutral networks [41] are readily derived by means of a random graph model. The neutral network, in essence, is based on random introduction of nodes into $\mathcal{Q}_n$ whose number is determined by the predetermined value of $\bar{\lambda}$. Theory predicts a phase transition like change in the appearance of neutral networks with increasing degree of neutrality at a critical value:

$$\lambda_{\mathrm{cr}} \;=\; 1 - \kappa^{-\frac{1}{\kappa-1}} \;, \tag{6}$$

where κ is the size of the genetic alphabet. For example, $\kappa = 4$ for the canonical genetic alphabet $\{\mathbf{A}, \mathbf{U}, \mathbf{G}, \mathbf{C}\}$, $\kappa = 3$ for $\{\mathbf{A}, \mathbf{U}, \mathbf{G}\}$, and $\kappa = 2$ for $\{\mathbf{G}, \mathbf{C}\}$. If $\bar{\lambda} < \lambda_{\mathrm{cr}}$ is fulfilled, the network consists of many isolated parts with one dominating giant component. On the other hand, the network is generically connected if $\bar{\lambda} > \lambda_{\mathrm{cr}}$. The critical value λ_{cr} is the connectivity threshold. This property of neutral networks reminds of percolation phenomena known from different areas of physics, although the high symmetry of sequence space, with all points being equivalent, introduces a difference in the two concepts.

Although qualitatively there is ample evidence for neutrality in natural evolution as well as in experiments under controlled conditions in the lab, very little is known about regularities in general genotype phenotype relations. In the RNA model [44] the phenotype is replaced by the minimum free energy structure of RNA. This simplifying assumption is met indeed by RNA evolution experiments *in vitro* [5] as well as by the design of

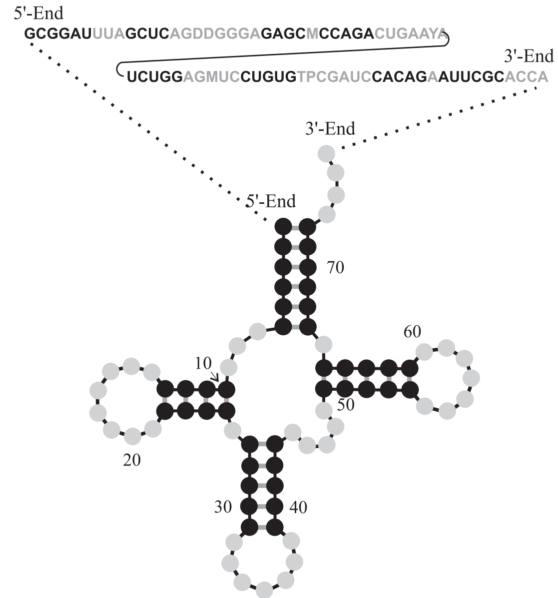

Figure 5: Nucleotide sequence and secondary structure of phenylalanyl-transfer-RNA.

RNA molecules through artificial selection [56]. Three-dimensional RNA structures are very complex objects still and they are replaced by so-called secondary structures that allow for efficient handling of large numbers of individuals in computer simulation experiments (Fig.5). RNA secondary structures are planar graphs and they can be represented, for example, by trees. They allow for straightforward application of combinatorial methods for counting structural features. For example, asymptotical formulas are available for the number of thermodynamically acceptable structures as functions of chain length

n [27]. Genotype-phenotype mappings are dependent on the ratio of sequences to structures and in case of the RNA model this ratio can be calculated analytically:

$$|\mathcal{Q}|/|\mathcal{S}| \;=\; 0.67\, n^{\frac{3}{2}} \times (2.16)^n \;.$$

It grows exponentially with chain length n and reaches several orders of magnitude already at $n \geq 10$. Efficient computer algorithms and implementations for RNA secondary structure prediction are available and RNA genotype-phenotype mappings are readily computed by folding all sequences of entire hypercubes and exhaustive enumeration. Although these studies are obviously limited to short RNAs, since the number of objects like sequences and structures that can be handled on conventional present day computers is limited to some 10^9, four properties of the RNA map were derived:

(i) **More sequences than structures**. For sequence spaces of chain lengths $n \geq 10$ there are orders of magnitude more sequences than structures and hence, the map is many-to-one (as predicted analytically above).

(ii) **Few common and many rare structures**. Relatively few common structures are opposed by a relatively large number of rare structures, some of which are formed by a single sequence only ("relatively" points at the fact that the numbers of both common and rare structures increase exponentially with n, but the exponent for the common structures is smaller than that for the rare ones).

(iii) **Shape space covering**. The distribution of neutral genotypes, these are sequences that fold into the same structure, is approximately random in sequence space. As a result it is possible to define a spherical ball, with a diameter d_{cov} being much smaller than the diameter of sequence space (n), which contains on the average for every common structure at least on sequence that folds into it.

(iv) **Existence and connectivity of neutral networks**. Neutral networks, being pre-images of phenotypes or structures in sequence space, of common structures are connected unless specific and readily recognizable special features of RNA structures require specific non-random distribution in the $\{\mathbf{A,U,G,C}\}$ sequence space, $\mathcal{Q}^{(\mathsf{AUGC})}$ (For structures formed from sequences over a $\{\mathbf{G,C}\}$ alphabet the connectivity threshold is higher, whereas, at the same time, the mean number of neutral neighbors is smaller).

All four properties of genotype-phenotype mappings can be casted into quantitative expressions for the RNA model and, at least, the features (i), (ii), and (iv) seem to hold for the more complicated protein spaces as well. Proteins, in contrast to RNA molecules, do not form structures but aggregate in aqueous solution when their constituents are too hydrophobic. This means that no useful structures will be available in certain parts of sequence space and the protein landscape is "holey" therefore. The concept of holey landscapes has been transferred also to the much more sophisticated problem of evolution of higher organisms and speciation [25].

Sequences at the intersection of the compatible sets of two neutral networks upon the same sequence space were found to be of actual interest since they can simultaneously carry the properties of the different RNA folds. For example, they can exhibit catalytic activities of two ribozymes [43]. The set of nodes of the neutral network $\mathcal{G}(S)$ is embedded in a compatible set $C(S)$ which includes all sequences that can form the structure S as suboptimal **or** minimum free energy conformation: $G(S) \subseteq C(S)$. The intersection theorem [41] states that for all pairs of structures S' and S'' the intersection $C(S') \cap C(S'')$ is non-empty. In other words, for each arbitrarily chosen pair of structures there will be at least on sequence that can form both. Schultes & Bartel [43] presented one particularly interesting experimental case. RNA switches, mentioned in section 2 above, are other applications of the intersection theorem.

Analogous results were reported for proteins. Computational studies predict the existence of neutral networks and shape space covering also for polypeptides [2, 3]. So far, these predictions are in agreement with experiments(see, for example, [36]).

5 METABOLIC REACTION NETWORKS

Once we leave the realm of *in vitro* evolution and proceed to living organism we have to consider at least two additional classes of networks: Metabolic networks describe the chemical pathways that process nutrients to build up essential substances and the networks regulating gene expression (see Andreas Wagner's contribution in this issue for a discussion of the latter).

A reaction network, for simplicity referred to as *network* is a pair $\mathfrak{N} = (S, \mathcal{R})$ of a set S of substrates and a set $\mathcal{R}$ of reactions "using" these substrates. For instance, consider

$$\mathsf{CaO} + 2\mathsf{CO_2} + \mathsf{H_2O} \rightarrow \mathsf{Ca(HCO_3)_2} \tag{7}$$

Another way of writing this is equation would be

$$\rho = -1 \cdot \mathsf{CaO} - 2 \cdot \mathsf{CO_2} - 1 \cdot \mathsf{H_2O} + 1 \cdot \mathsf{Ca(HCO_3)_2} \tag{8}$$

This representation suggests to view a reaction as (formal) linear combination of substrates $\rho = \sum_{\mathsf{x}} s_{\rho,\mathsf{x}} \mathsf{x}$ with *stoichiometric coefficients* $s_{\rho,\mathsf{x}}$. Equivalently, a reaction can be thought of as a directed hyperedge in a directed weighted hypergraph that connects the substrates on the

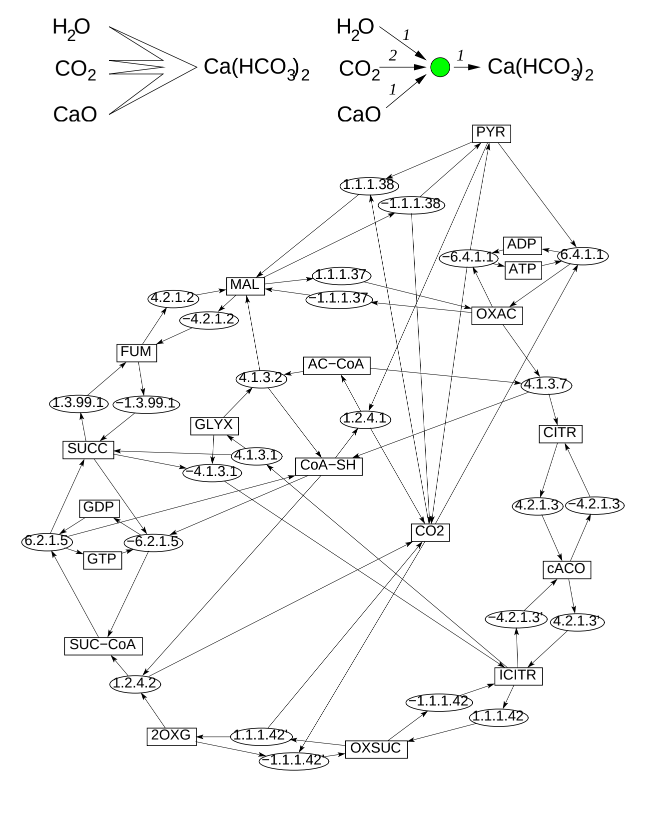

Figure 6: Description of a reaction as hyperedge connecting its substrate vertices and as (part of) the bipartite network graph in which both substrates and reactions are represented as vertices.
Below the bipartite graph of the Krebs citric acid cycle and its alternative pathways is shown (H_2O and H^+ omitted); reactions (in ovals) are labelled by enzyme that catalyzes the reaction.

left hand side of the reaction with its right hand side. Equivalently, we may draw a bipartite graph in which substrates and reactions are different classes of vertices (Fig. 6). The bipartite graphs are particularly appropriate in the context of metabolic reaction networks since there the reactions are specifically catalyzed by enzymes or enzyme complexes. The catalysts appear as one class of vertices while the small substrates such as sugars or amino acids form the other class.

Metabolic flux analysis [9, 16, 40, 42] consists in finding a basis of elementary "flux modes" that describe the dominant reaction pathways within the network. From the mathematical point of view the problem is to find the extremal rays of the cone defined by the intersection of the non-negative orthant $\{\vec{J}|J_i \geq 0\}$ and the kernel $\ker \mathbf{S}^+ = \{\vec{J}\mathbf{S} = \vec{o}\}$ of (the transpose of) the stoichiometric matrix $\mathbf{S} = (s_{x,\rho})$. The extremal flux vectors $\vec{J}$ spanning this cone are closely related to the (directed) cycles of the network graph [26].

Comparative studies of metabolic networks reveal a surprising variety of pathways even in the core of the metabolism. For instance, most organisms use only small parts of the citric acid cycle [32] (Fig.6). Functional differences can be explained in terms of differences of metabolic pathways [22, 23], allowing a first glimpse on a genotype-phenotype map that is much more complex than biopolymer folding.

On a global scale, recent surveys [17, 34, 54] have revealed that metabolic reaction networks belong to the class of small world networks [55] in the wider sense: they have a diameter that is much smaller than what one would expect for an uncorrelated random graph with the same number of vertices and edges and a power-law distribution of vertex degrees. The evolutionarily oldest metabolites play the role of "hubs" exhibiting the largest vertex degree. A comparison with reaction networks from planetary atmospheres [26] suggests that small world features are a generic property of chemical reaction networks rather than the product of evolution. On the other hand, there is an over-abundance of small cycles that could be the result of an evolutionary optimization for resilience against perturbations.

We have discussed here only four classes of the networks which play a central role in molecular biology. Networks of gene regulation, both at level of the single cell and the signaling networks of multicellular organisms are just as important. In higher organisms neural networks form an additional layer of complexity (see Olaf Sporn's contribution in this issue).

While we begin to understand structure and function of networks within each class, for which ample data are available at least in some cases, little is known on the interaction of the various levels. The mechanisms governing crosstalk between hierarchical levels are, nevertheless, equally important if not more relevant for proper functioning of organisms. They are in the center of top-down control and regulation, without which every form of organization is doomed to break down. What makes this type of control tricky in but not only in biology is the partial autonomy of agents at the lower level. More and more detailed knowledge on communication between cells will help us to explore successfully the subtle interplay between local independence and global control.

REFERENCES

(1) E. Amann. *Permanence in Catalytic Networks.* Phd thesis, Universität Wien, Institut für Mathematik, 1989.

(2) A. Babajide, R. Farber, I. L. Hofacker, J. Inman, A. S. Lapedes, and P. F. Stadler. Exploring protein sequence space using knowledge based potentials. *J. Theor. Biol.*, 212:35–46, 2001.

(3) A. Babajide, I. L. Hofacker, M. J. Sippl, and P. F. Stadler. Neutral networks in protein space: A

computational study based on knowledge-based potentials of mean force. *Folding & Design*, 2:261–269, 1997.

(4) O. M. Becker and M. Karplus. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.*, 106:1495–1517, 1997.

(5) C. K. Biebricher and W. C. Gardiner. Molecular evolution of RNA *in vitro*. *Biophys. Chem.*, 66:179–192, 1997.

(6) M. C. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D*, 48:17 – 28, 1991.

(7) J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function, and Genetics*, 21:167–195, 1995.

(8) O. Catoni. Simulated annealing algorithms and Markov chains with rate transitions. In J. Azema, M. Emery, M. Ledoux, and M. Yor, editors, *Seminaire de Probabilites XXXIII*, volume 709 of *Lecture Notes in Mathematics*, pages 69–119. Springer, Berlin/Heidelberg, 1999.

(9) B. L. Clarke. Stoichiometric network analysis. *Cell Biophys.*, 12:237–253, 1988.

(10) K. A. Dill and H. S. Chan. From Levinthal to payhways to funnels. *Nature Struct. Biol.*, 4:10–19, 1997.

(11) M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465–523, 1971.

(12) M. Eigen, J. McCaskill, and P. Schuster. The molecular quasispecies. *Adv. Chem. Phys.*, 75:149–263, 1989.

(13) M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, 64:541–565, 1977.

(14) M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. *Naturwissenschaften*, 65:7–41, 1978.

(15) M. Eigen and P. Schuster. *The Hypercycle*. Springer-Verlag, Berlin, 1979.

(16) D. A. Fell. *Understanding the Control of Metabolism*. Portland Press, London, 1997.

(17) D. A. Fell and A. Wagner. Structural properties of metabolic networks: implications for evolution and modelling of metabolism. In J.-H. S. Hofmeyr, J. M. Rohwer, and J. L. Snoep, editors, *Animating the cellular map*, pages 79–85, Stellenbosch, ZA, 2000. Stellenbosch University Press.

(18) C. Flamm, W. Fontana, I. Hofacker, and P. Schuster. RNA folding kinetics at elementary step resolution. *RNA*, 6:325–338, 2000.

(19) C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger. Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, 216:155–173, 2002.

(20) W. Fontana and P. Schuster. Continuity in evolution: On the nature of transitions. *Science*, 280:1451–1455, 1998.

(21) J. F. Fontanari and P. F. Stadler. Fractal geometry of spin-glass models. *J. Phys. A: Math. Gen.*, 35:1509–1516, 2002.

(22) C. V. Forst and K. Schulten. Evolution of metabolisms: A new method for the comparison of metabolic pathways using genomic information. *J. Comput. Biol.*, 6:343–360, 1999.

(23) C. V. Forst and K. Schulten. Phylogenetic analysis of metabolic pathways. *J. Mol. Evol.*, 52:471–489, 2001.

(24) P. Garstecki, T. X. Hoang, and M. Cieplak. Energy landscapes, supergraphs, and "folding funnels" in spin systems. *Phys. Rev. E*, 60:3219–3226, 1999.

(25) S. Gavrilets. Evolution and speciation on holey adaptive landscapes. *Trends in Ecology and Evolution*, 12:307–312, 1997.

(26) P. M. Gleiss, P. F. Stadler, A. Wagner, and D. A. Fell. Relevant cycles in chemical reaction network. *Adv. Complex Syst.*, 4:207–226, 2001.

(27) I. L. Hofacker, P. Schuster, and P. F. Stadler. Combinatorics of RNA secondary structures. *Discr. Appl. Math.*, 89:177–207, 1998.

(28) J. Hofbauer. On the occurrence of limit cycles in the Volterra-Lotka equation. *Nonlinear Analysis*, 5:1003–1007, 1981.

(29) J. Hofbauer and K. Sigmund. *Dynamical Systems and the Theory of Evolution*. Cambridge University Press, Cambridge U.K., 1988.

(30) J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamicals*. Cambridge University Press, Cambridge U.K., 1998.

(31) V. Hutson and K. Schmitt. Permanence and the dynamics of biological systems. *Math. Biosc.*, 111:1–71, 1992.

(32) M. A. Huynen, T. Dandekar, and P. Bork. Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.*, 7:281–291, 1999.

(33) M. A. Huynen, P. F.Stadler, and W. Fontana. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401, 1996.

(34) H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.

(35) W. K. Johnston, P. J. Unrau, M. J. Lawrence,

M. E. Glasner, and D. P. Bartel. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science*, 292:1319–1325, 2001.

(36) A. D. Keefe and J. W. Szostak. Functional proteins from a random-sequence library. *Nature*, 410:715–718, 2001.

(37) M. Kimura. Evolutionary rate at the molecular level. *Nature*, 217:624–626, 1968.

(38) M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK, 1983.

(39) J. L. King and T. H. Jukes. Non-Darwinian evolution: Random fixation of selectively neutral variants. *Science*, 164:788–798, 1969.

(40) J. S. Olivera, C. G. Bailey, J. B. Jones-Olivera, and D. A. Dixon. An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bull. Math. Biol.*, 63:1163–1196, 2001.

(41) C. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatory maps. Neutral networks of RNA secondary structure. *Bull. Math. Biol.*, 59:339–397, 1997.

(42) C. H. Schilling, D. Letscher, and B. O. Palsson. Theory for systemic definition of metabolic pathways and their use in interpreting metabolic function from pathway-oriented perspective. *J. Theor. Biol.*, 203:229–248, 2000.

(43) E. A. Schultes and D. P. Bartel. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289:448–452, 2000.

(44) P. Schuster. Evolution *in silico* and *in vitro*: The RNA model. *Biol. Chem.*, 382:1301–1314, 2001.

(45) P. Schuster. Molecular insight into the evolution of phenotypes. In J. P. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics – Exploring the Interplay of Accident, Selection, Neutrality, and Function*. Oxford University Press, New York, 2002. In press.

(46) P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284, 1994.

(47) P. Schuster and K. Sigmund. Replicator dynamics. *J.Theor.Biol.*, 100:533–538, 1983.

(48) P. Schuster, K. Sigmund, and R. Wolff. Dynamical systems under constant organization III: Cooperative and competitive behaviour of hypercycles. *J. Diff. Eqns.*, 32:357–368, 1979.

(49) P. Schuster, K. Sigmund, and R. Wolff. Mass action kinetics of selfreplication in flow reactors. *J. Math. Anal. Appl.*, 78:88–112, 1980.

(50) K. Sigmund and P. Schuster. Permanence and uninvadability for deterministic population models. In P. Schuster, editor, *Stochastic Phenomena and Chaotic Behaviour in Complex Systems*, volume 21 of *Synergetics*, pages 186–205, New York, 1984. Springer-Verlag.

(51) S. Spiegelman. An approach to experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 4:213–253, 1971.

(52) P. F. Stadler and R. Happel. The probability for permanence. *Math. Biosc.*, 113:25–50, 1993.

(53) P. F. Stadler and P. Schuster. Dynamics of small autocatalytic reaction networks I: Bifurcations, permanence and exclusion. *Bull. Math. Biol.*, 52:485–508, 1990.

(54) A. Wagner and D. A. Fell. The small world inside large metabolic networks. Technical Report 00-07-041, Santa Fe Institute, 2000.

(55) D. J. Watts. *Small Worlds*. Princeton University Press, Princeton NJ, 1999.

(56) D. S. Wilson and J. W. Szostak. *In Vitro* selection of fuctional nucleic acids. *Annu. Rev. Biochem.*, 68:611–647, 1999.

(57) P. G. Wolynes, J. N. Onuchic, and D. Thirumalai. Navigating the folding routes. *Science*, 267:1619–1620, 1995.