

# Networks in Molecular Evolution

*A Common Theme at All Levels*

**PETER SCHUSTER AND PETER F. STADLER**

## 1. INTRODUCTION

**N**etworks 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 [1] which extend to two dimensions what in reality is a

---

*Peter Schuster is professor of theoretical chemistry at the University of Vienna since 1973, an external faculty member of the Santa Fe Institute since 1991, and most recently, he was appointed Vice President of the Austrian Academy of Sciences. He holds a Ph.D. in Chemistry and Physics from the University of Vienna. From 1991–1995 he was founding director of the Institute for Molecular Biotechnology in Jena, Germany, a research institute devoted to transfer of knowledge from basic research in the life sciences to technological applications. Since beginning his Ph.D. work in 1963, Professor Schuster has worked in the areas of organic chemistry, theoretical chemistry, molecular spectroscopy, chemical kinetics, and molecular evolution. He has developed the concepts of molecular quasispecies and hypercycles together with the nobel laureate Manfred Eigen. His current research is dealing with the role of neutrality in evolutionary dynamics. He developed the concept of neutral networks and designed the RNA model as a tool to study evolutionary phenomena, a tool that is simple enough to allow for mathematical analysis and computer simulation yet sufficiently detailed to be subjected to experimental testing; E-mail: pks@tbi.univie.ac.at.*

*Peter Stadler holds a Ph.D. in Chemistry from the University of Vienna. Born December 24, 1965 in Vienna, Peter Stadler is married with two children; he holds a PostDoc at the Max Planck Institute for Biophysical Chemistry in Goettingen, Germany. He has been Assistant Professor (since 1991) and Associate Professor (since 1994) of Theoretical Chemistry at the University of Vienna. Currently he is Professor for Bioinformatics in Leipzig since September 2002 and Santa Fe Institute external faculty since 1994. Main topics of his work are the theory of fitness landscapes, sequence-structure maps of biopolymers, the mathematical structure of phenotypic evolution, and the dynamics of molecular replicators. E-mail: peter.stadler@bioinf.uni-leipzig.de.*

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 [2].

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 [3], 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 [4]. The mechanistic way that autocatalysis shapes transitions from the evolution of molecules towards higher functional units is essentially comprised by the role of catalysis in autocatalytic kinetic networks [5]. Extensive theoretical and computational studies of RNA evolution *in vitro* by the Vienna group (for a recent review see [6]) 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 [7,8] and inferred from empirical molecular sequence data [9]. 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 [10,11].

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,  $y \in N(x)$ , if it satisfies  $f(x) \leq f(y)$ .

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  $\mathbf{p} \in \mathbb{P}_{x,y}$  that connect  $x$  and  $y$  with smallest maximal energy. The height of the saddle point is therefore

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

The saddle points can be arranged in a tree such that the leaves of the tree are the metastable states, whereas the internal nodes correspond to the saddle points separating the local optima from each other [2,12,13] (Figure 1). The height of the energy barrier can be directly read off the barrier tree, see Figure 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 [14], or the fractal dimension of the barrier tree itself [15].

We say that a saddle point  $s$  *directly* connects  $x$  and  $y$  if there are two paths of steepest descent that start at neighbors  $x', y' \in N(s)$  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 [16]. 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 Figure 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 problems such as the *low autocorrelated binary string problem* [16].

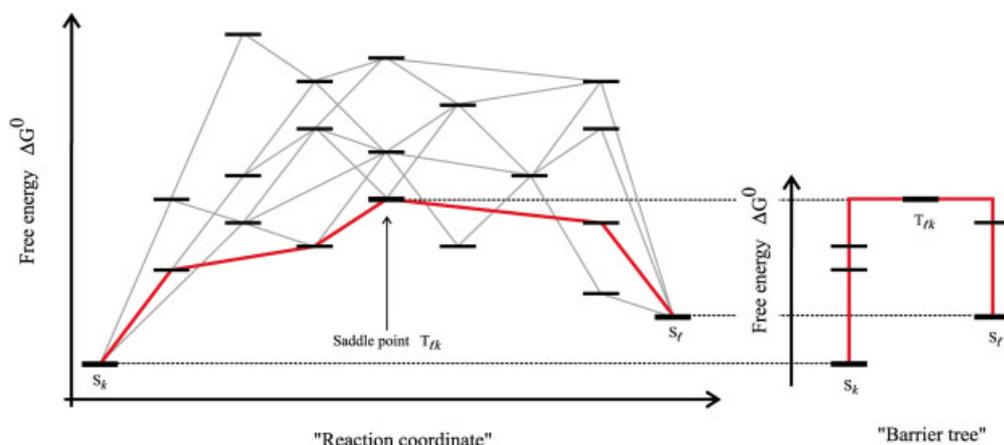
The network structure of a biopolymer’s energy landscape determines its folding properties. Heuristic notions such as the “folding funnel” [13,17,18] postulated for most natural proteins can be described at a microscopic level in terms of network properties. For example, one might speak of a funnel if there are no local optima with significant basins of attraction and/or substantial energy barriers.

## 3. AUTOCATALYTIC NETWORKS

Differential selection as a consequence of differences in replication rates is one of the pillars of the Darwinian par-

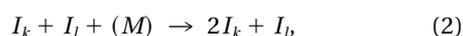
**A theoretical frame for handling evolution of molecules has been developed almost simultaneously through combining the concepts of population genetics with the knowledge of molecular and structural biology.**

FIGURE 1



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

adigm. One of the simplest models deals with polymeric molecules that replicate according to the reaction scheme



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 1970s [4,19], 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 [20], 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 [21–24]

$$\dot{x}_k = x_k \left( \sum_{i=1}^n A_{ki} x_i - \Phi(t) \right); \quad \Phi(t) = \sum_{i,j} A_{ij} x_i x_j, \quad (3)$$

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  $\sum_k \dot{x}_k = 0$ . Equation (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  $A_{kj} \rightarrow A_{kj} - A_{kk}$  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* [25] 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 (Figure 3).

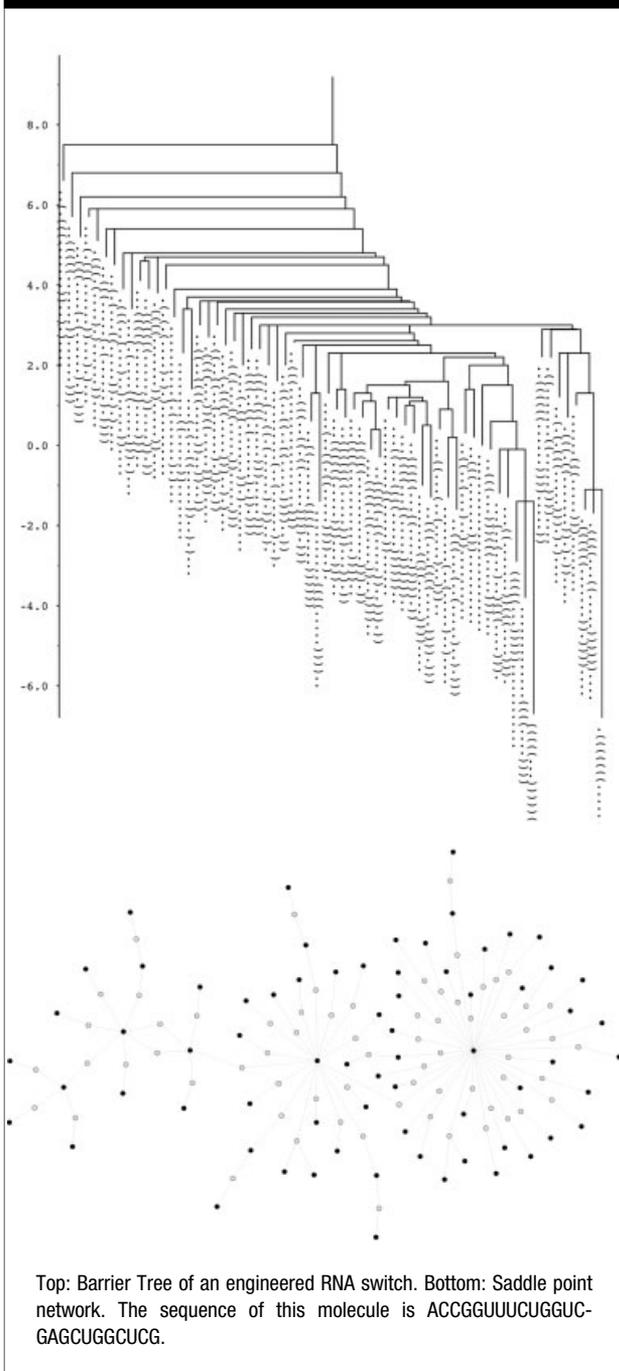
There is a connection between the graph representation of the reaction networks and their dynamical behavior. For example, if the network is *permanent* [25,28] (i.e., no member dies out over time), then the graph is strongly connected [29]. The hypercycle model [19] of cooperating replicators is probably the best-known example of a permanent catalytic network. For  $n \leq 5$  the network must be “hypercycle-like” by containing a Hamiltonian cycle [27].

Permanent autocatalytic networks have been studied extensively as a possibility to overcome limitations imposed on replication-mutation ensembles by the *error-threshold* of molecular quasispecies model [4,30,31], 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 [32].

The networks of ecology behave similarly to the autocatalytic networks because the replicator Equation (3) and the Lotka-Volterra equations are related by a transfor-

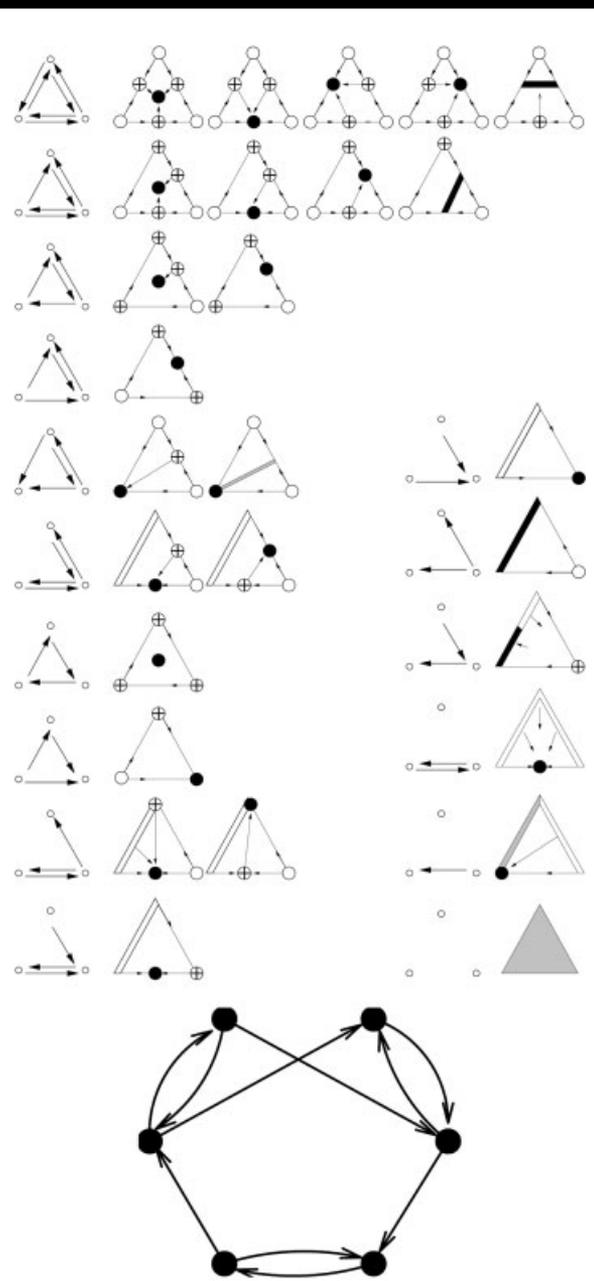
**The hyper cycle model of cooperations replicators is the best-known example of a permanent catalytic network. Permanent coexistence is an exceedingly rare and fragile phenomenon that requires very special interaction structures.**

**FIGURE 2**



mation [33]. Permanence in Lotka-Volterra systems is obviously related to the loss of species in ecosystems. A survey of replicator systems [34] thus implies that permanence is also an exceedingly rare and fragile phenomenon in ecological networks that requires very special interaction structures. Those natural ecosystems that are known to be stable over long periods thus owe their stability to an efficient evolutionary selection process.

**FIGURE 3**



Autocatalytic networks. Top: Phase portraits of the autocatalytic networks with  $n = 3$  species. This is only part of the classification of three-species replicator equations [26]. Symbols: sink ●, source ○, saddle ⊕. Lines of fixed points are shaded according to their stability. Bottom: Non-Hamiltonian autocatalytic network compatible with permanence [27].

#### 4. NEUTRAL NETWORKS IN SEQUENCE SPACE

The evolutionarily relevant relation between genotypes  $l$  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,  $\mathcal{P}$ . 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 (Figure 4). Distance between two genotypes  $I$  and  $I'$  of length  $n$  is readily expressed by means of the Hamming metric  $d^h(I, I')$ , 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

$$\psi : (\mathcal{Q}_n; d^h) \rightarrow (\mathcal{P}; d^s) \quad S = \psi(I) \quad (4)$$

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

As indicated in Figure 4 there is ample evidence 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 [35] 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 pre-image  $\psi^{-1}(S)$  by means of the Hamming distance one criterium for edges. The quantity

**The network structure of a biopolymer's energy landscape determines its folding properties.**

that allows for global 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  $\bar{\lambda}$ .

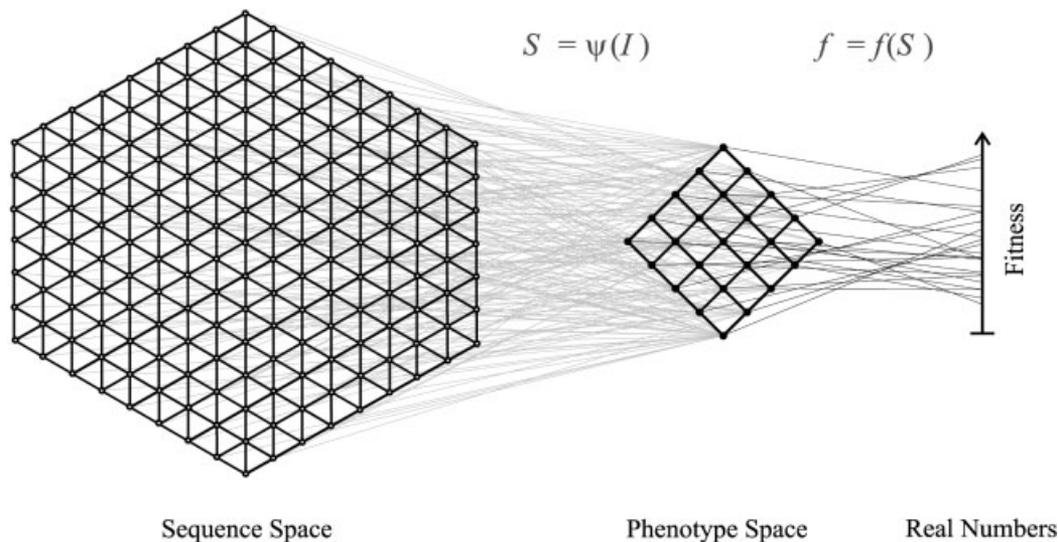
Neglecting the influence of the distribution of neutral sequences over  $\mathcal{Q}_n$ , the degree of neutrality will be the higher, the larger the pre-image.

Generic properties of neutral networks [36] 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:

$$\lambda_{cr} = 1 - \kappa^{-1/(\kappa-1)}, \quad (6)$$

where  $\kappa$  is the size of the genetic alphabet. For example,  $\kappa = 4$  for the canonical genetic alphabet {A, U, G, C},  $\kappa = 3$  for {A, U, G}, and  $\kappa = 2$  for {G, C}. If  $\bar{\lambda} < \lambda_{cr}$  is fulfilled, the network consists of many isolated parts with one dominating giant component. On the other hand, the network is generically connected if  $\bar{\lambda} > \lambda_{cr}$ . The critical value  $\lambda_{cr}$  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.

**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.

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 [37] the phenotype is replaced by the minimum free energy structure of RNA. This simplifying assumption is met indeed by RNA evolution experiments *in vitro* [38] as well as by the design of RNA molecules through artificial selection [39]. 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 (Figure 5). RNA secondary structures are planar graphs and they can be represented, for example, by trees. They allow for straight-forward 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  $n$  [40]. 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:

$$|\mathcal{Q}|/|\mathcal{P}| = 0.67 n^{3/2} \times (2.16)^n.$$

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  $\{\mathbf{A}, \mathbf{U}, \mathbf{G}, \mathbf{C}\}$  sequence space,  $\mathcal{Q}^{(\mathbf{AUGC})}$  (For structures formed from sequences over a  $\{\mathbf{G}, \mathbf{C}\}$  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 cast 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

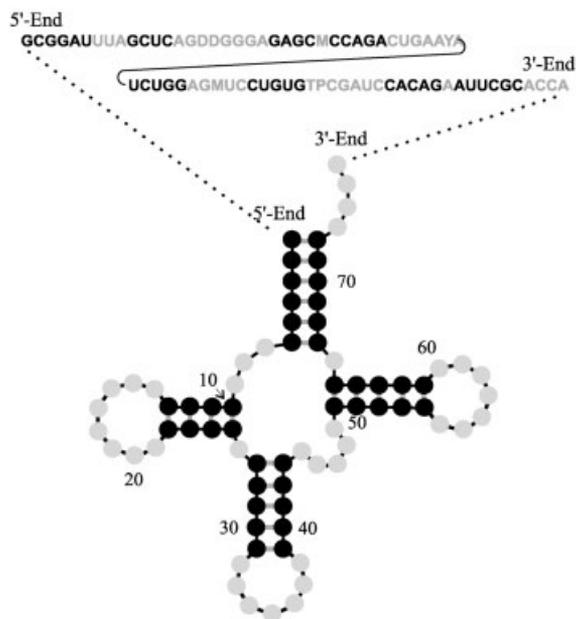
parts of sequence space and the protein landscape is therefore "holey." The concept of holey landscapes has been transferred also to the much more sophisticated problem of evolution of higher organisms and speciation [41].

**Neutral networks in sequence space explain the evolvability of both nucleic acids and polypeptides by linking Darwinian selection with neutral drift.**

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 [42]. 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 [36] 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 one sequence that can form both. Schultes and Bartel [42] 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 [43,44]. So far, these predictions are in agreement with experiments (see, for example, [45]).

**FIGURE 5**

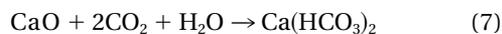


Nucleotide sequence and secondary structure of phenylalanyl-transfer-RNA.

**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 [46]).

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

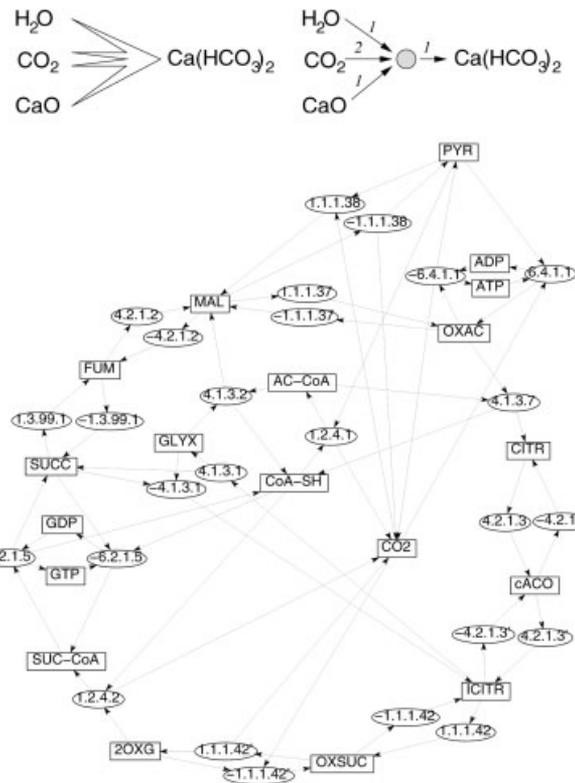


Another way of writing this is equation would be

$$\rho = -1 \cdot \text{CaO} - 2 \cdot \text{CO}_2 - 1 \cdot \text{H}_2\text{O} + 1 \cdot \text{Ca}(\text{HCO}_3)_2 \quad (8)$$

This representation suggests ways to view a reaction as (formal) linear combination of substrates  $\rho = \sum_x s_{\rho,x} x$  with *stoichiometric coefficients*  $s_{\rho,x}$ . Equivalently, a reaction can be thought of as directed hyperedges in a directed weighted hypergraph that connects the substrates on the 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 (Figure 6). The bipartite graphs are particularly appropriate in the context of metabolic reaction networks since there the reactions are specif-

**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 ( $\text{H}_2\text{O}$  and  $\text{H}^+$  omitted); reactions (in ovals) are labelled by enzyme that catalyzes the reaction.

ically catalyzed by the 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 [47–50] 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  $\ker \mathbf{S}^+ = \{\vec{J} | \mathbf{S} \vec{J} = \vec{0}\}$  of (the transpose of) the stoichiometric matrix  $\mathbf{S} = (s_{\rho,x})$ . The extremal flux vectors  $\vec{J}$  spanning this cone are closely related to the (directed) cycles of the network graph [51].

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 [52] (Figure 6). Functional differences can be explained in terms of differences of metabolic pathways [53,54], allowing a first glimpse on a genotype-pheno-

type map that is much more complex than biopolymer folding.

On a global scale, recent surveys [55–57] have revealed that metabolic reaction networks belong to the class of small world networks [58] 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 [51] 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 [59]).

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. Bryngelson, J.D.; Onuchic, J.N.; Socci, N.D.; Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Genet* 1995, 21, 167–195.
2. Flamm, C.; Fontana, W.; Hofacker, I.; and Schuster, P. RNA folding kinetics at elementary step resolution. *RNA* 2000, 6, 325–338.
3. Spiegelman, S. An approach to experimental analysis of precellular evolution. *Quart Rev Biophys* 1971, 4, 213–253.
4. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 1971, 58, 465–523.
5. Eigen, M.; Schuster, P. *The Hypercycle*; Springer-Verlag: Berlin, 1979.
6. Schuster, P. Molecular insight into the evolution of phenotypes. In *Evolutionary Dynamics—Exploring the Interplay of Accident, Selection, Neutrality, and Function*; Crutchfield, J.P., Schuster, P., Eds.; Oxford University Press, New York, 2002, pp 163–218.
7. Kimura, M. Evolutionary rate at the molecular level. *Nature* 1968, 217, 624–626.
8. Kimura, M. *The Neutral Theory of Molecular Evolution*. Cambridge University Press: Cambridge, UK, 1983.
9. King, J.L.; Jukes, T.H. Non-Darwinian evolution: Random fixation of selectively neutral variants. *Science* 1969, 164, 788–798.
10. Fontana, W.; Schuster, P. Continuity in evolution: On the nature of transitions. *Science* 1998, 280, 1451–1455.
11. Huynen, M.A.; Stadler, P.F.; Fontana, W. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc Natl Acad Sci USA* 1996, 93, 397–401.
12. Becker, O.M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J Chem Phys* 1997, 106, 1495–1517.
13. Garstecki, P.; Hoang, T.X.; Cieplak, M. Energy landscapes, supergraphs, and “folding funnels” in spin systems. *Phys Rev E* 1999, 60, 3219–3226.
14. Catoni, O. Simulated annealing algorithms and Markov chains with rate transitions. In: Azema, J., Emery, M., Ledoux, M., Yor, M., Eds.; *Seminaire de Probabilites XXXIII*, volume 709 of *Lecture Notes in Mathematics*. Springer: Berlin/Heidelberg, 1999, pp 69–119.
15. Fontanari, J.F.; Stadler, P.F. Fractal geometry of spin-glass models. *J Phys A: Math Gen* 2002, 35, 1509–1516.
16. Flamm, C.; Hofacker, I.L.; Stadler, P.F.; Wolfinger, M.T. Barrier trees of degenerate landscapes. *Z Phys Chem* 2002, 216, 155–173.
17. Dill, K.A.; Chan, H.S. From Levinthal to payhways to funnels. *Nat Struct Biol* 1997, 4, 10–19.
18. Wolynes, P.G.; Onuchic, J.N.; Thirumalai, D. Navigating the folding routes. *Science* 1995, 267, 1619–1620.
19. Eigen, M.; Schuster, P. The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. *Naturwissenschaften* 1978, 65, 7–41.
20. Johnston, W.K.; Unrau, P.J.; Lawrence, M.J.; Glasner, M.E.; Bartel, D.P. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* 2001, 292, 1319–1325.
21. Hofbauer, J.; Sigmund, K. *Dynamical Systems and the Theory of Evolution*; Cambridge University Press: Cambridge UK, 1988.
22. Hofbauer, J.; Sigmund, K. *Evolutionary Games and Population Dynamics*; Cambridge University Press: Cambridge, UK, 1998.
23. Schuster, P.; Sigmund, K. Replicator dynamics. *J Theor Biol* 1983, 100, 533–538.
24. Schuster, P.; Sigmund, K.; Wolff, R. Mass action kinetics of selfreplication in flow reactors. *J Math Anal Appl* 1980, 78, 88–112.
25. Schuster, P.; Sigmund, K.; Wolff, R. Dynamical systems under constant organization III: Cooperative and competitive behaviour of hypercycles. *J Diff Eqns* 1979, 32, 357–368.
26. Stadler, P.F.; Schuster, P. Dynamics of small autocatalytic reaction networks I: Bifurcations, permanence and exclusion. *Bull Math Biol* 1990, 52, 485–508.
27. Amann, E. *Permanence in Catalytic Networks*. Phd thesis, Universität Wien, Institut für Mathematik, 1989.
28. Hutson, V.; Schmitt, K. Permanence and the dynamics of biological systems. *Math Biosci* 1992, 111, 1–71.
29. Sigmund, K.; Schuster, P. Permanence and uninvasibility for deterministic population models. In *Stochastic Phenomena and Chaotic Behaviour in Complex Systems*, volume 21 of *Synergetics*; Schuster, P., Ed.; Springer-Verlag: New York, 1984, pp 186–205.
30. Eigen, M.; McCaskill, J.; Schuster, P. The molecular quasispecies. *Adv Chem Phys* 1989, 75, 149–263.

31. Eigen, M.; Schuster, P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften* 1977, 64, 541–565.
32. Boerlijst, M.C.; Hogeweg, P. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D* 1991, 48, 17–28.
33. Hofbauer, J. On the occurrence of limit cycles in the Volterra-Lotka equation. *Nonlinear Analysis* 1981, 5, 1003–1007.
34. Stadler, P.F.; Happel, R. The probability for permanence. *Math Biosci* 1993, 113, 25–50.
35. Schuster, P.; Fontana, W.; Stadler, P.F.; Hofacker, I.L. From sequences to shapes and back: A case study in RNA secondary structures. *Proc Roy Soc Lond B* 1994, 255, 279–284.
36. Reidys, C.; Stadler, P.F.; Schuster, P. Generic properties of combinatory maps. Neutral networks of RNA secondary structure. *Bull Math Biol* 1997, 59, 339–397.
37. Schuster, P. Evolution *in silico* and *in vitro*: The RNA model. *Biol Chem* 2001, 382, 1301–1314.
38. Biebricher, C.K.; Gardiner, W.C. Molecular evolution of RNA *in vitro*. *Biophys Chem* 1997, 66, 179–192.
39. Wilson, D.S.; Szostak, J.W. *In Vitro* selection of functional nucleic acids. *Annu Rev Biochem* 1999, 68, 611–647.
40. Hofacker, I.L.; Schuster, P.; Stadler, P.F. Combinatorics of RNA secondary structures. *Discr Appl Math* 1998, 89, 177–207.
41. Gavrillets, S. Evolution and speciation on holey adaptive landscapes. *Trends Ecol Evol* 1997, 12, 307–312.
42. Schultes, E.A.; Bartel, D.P. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science* 2000, 289, 448–452.
43. Babajide, A.; Farber, R.; Hofacker, I.L.; Inman, J.; Lapedes, A.S.; Stadler, P.F. Exploring protein sequence space using knowledge based potentials. *J Theor Biol* 2001, 212, 35–46.
44. Babajide, A.; Hofacker, I.L.; Sippl, M.J.; Stadler, P.F. Neutral networks in protein space: A computational study based on knowledge-based potentials of mean force. *Folding & Design* 1997, 2, 261–269.
45. Keefe, A.D.; Szostak, J.W. Functional proteins from a random-sequence library. *Nature* 2001, 410, 715–718.
46. Wagner, A. The large-scale structure of metabolic networks: a glimpse at life's origins? *Complexity* 2002, 8(1), 15.
47. Clarke, B.L. Stoichiometric network analysis. *Cell Biophys* 1988, 12, 237–253.
48. Fell, D.A. *Understanding the Control of Metabolism*; Portland Press: London, 1997.
49. Olivera, J.S.; Bailey, C.G.; Jones-Olivera, J.B.; Dixon, D.A. An algebraic-combinatorial model for the identification and mapping of biochemical pathways. *Bull Math Biol* 2001, 63, 1163–1196.
50. Schilling, C.H.; Letscher, D.; Palsson, B.O. Theory for systemic definition of metabolic pathways and their use in interpreting metabolic function from pathway-oriented perspective. *J Theor Biol* 2000, 203, 229–248.
51. Gleiss, P.M.; Stadler, P.F.; Wagner, A.; Fell, D.A. Relevant cycles in chemical reaction network. *Adv Complex Syst* 2001, 4, 207–226.
52. Huynen, M.A.; Dandekar, T.; Bork, P. Variation and evolution of the citric-acid cycle: A genomic perspective. *Trends Microbiol* 1999, 7, 281–291.
53. Forst, C.V.; Schulten, K. Evolution of metabolisms: A new method for the comparison of metabolic pathways using genomic information. *J Comput Biol* 1999, 6, 343–360.
54. Forst, C.V.; Schulten, K. Phylogenetic analysis of metabolic pathways. *J Mol Evol* 2001, 52, 471–489.
55. Fell, D.A.; Wagner, A. Structural properties of metabolic networks: implications for evolution and modelling of metabolism. In: *Animating the Cellular Map*, Hofmeyr, J.-H.S., Rohwer, J.M., Snoep, J.L., Eds.; Stellenbosch University Press: Stellenbosch, ZA, 2000, pp. 79–85.
56. Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabasi, A. The large-scale organization of metabolic networks. *Nature* 2000, 407, 651–654.
57. Wagner, A.; Fell, D.A. The small world inside large metabolic networks. Technical Report 00-07-041, Santa Fe Institute, 2000.
58. Watts, D.J. *Small Worlds*; Princeton University Press: Princeton, NJ, 1999.
59. Sporns, O. Network analysis complexity, and brain function. *Complexity* 2002, 8(1).