Modelling “evo-devo” with RNA

Walter Fontana

SFI WORKING PAPER: 2002-09-053

SANTA FE INSTITUTE

Modelling 'evo-devo' with RNA

Walter Fontana

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, NM 87501, USA
walter@santafe.edu

Keywords: genotype-phenotype map, neutral networks
shape space, topology, RNA folding

to appear in *BioEssays*

Summary

The folding of RNA sequences into secondary structures is a simple yet biophysically grounded model of a genotype-phenotype map. Its computational and mathematical analysis has uncovered a surprisingly rich statistical structure characterized by shape space covering, neutral networks and plastogenetic congruence. I review these concepts and discuss their evolutionary implications.

1 Introduction

Phenotype refers to the physical, organizational and behavioral expression of an organism during its lifetime. Genotype refers to a heritable repository of information that instructs the production of molecules whose interactions, in conjunction with the environment, generate and maintain the phenotype. The processes linking genotype to phenotype are known as development. They intervene in the genesis of phenotypic novelty from genetic mutation. Evolutionary trajectories therefore depend on development. In turn, evolutionary processes shape development, creating a feed-back known as "evo-devo" [1,2].

The main thrust of this review is to show that some key aspects of this feed-back are present even in the microcosm of RNA folding. In a narrow sense, the relation between RNA sequences and their shapes is treated as a problem in biophysics. Yet, in a wider sense, RNA folding can be regarded as a minimal model of a genotype-phenotype relation[3].

The RNA model is not a representation of organismal development. The regulatory networks of gene expression and signal transduction that coordinate the spatiotemporal unfolding of complex molecular processes in organismal development (for recent overviews see[4,5]) have no concrete analogue in the RNA sequence-to-structure map. Developmental processes themselves evolve and this too is outside the scope of the rather simple notion of RNA folding considered here. Yet, the RNA folding map transparently implements concepts like epistasis and phenotypic plasticity, thus enabling the study of constraints to variation, canalization, modularity, phenotypic robustness and evolvability. As detailed in this review, the statistical architecture of the sequence-to-structure map in RNA offers explanations for patterns of

phenotypic evolution, such as directionality and the partially punctuated nature of evolutionary change. This statistical architecture is critically shaped by "developmental neutrality", that is, the extent to which many genotypes map into the same phenotype. The RNA model is an *abstract* analogue of development in that it grounds a discussion of these issues within a simple biophysical framework. The fact that neutrality is the key structuring factor does not hinge on the developmental mechanisms generating neutrality, but, of course, the extent of neutrality does. Whether the features present in the RNA map carry over to more complex genotype-phenotype maps (and to which ones at that) will depend on the genetic robustness of developmental mechanisms.

An important purpose of simple and abstract models in biology is to sharpen the questions – perhaps even to understand what the questions are. In this vein, the RNA model contributes in making more precise what we mean when we speak of "phenotype *space*". Like any other phenotype, an RNA shape cannot be heritably modified in a direct way, but requires a change in the underlying sequence. This indirection in transforming one phenotype into another makes the structure of phenotype space dependent on the structure of genotype space *and* the mapping from genotype to phenotype. The absence of a formal theory addressing the latter dependence in the neo-Darwinian school of evolutionary thought has led to unwarranted assumptions about the structure of phenotype space and to much confusion about continuity and discontinuity in evolution (for a discussion see[6]).

2 RNA phenotypes

Secondary structure. RNA structure can be defined at several levels of resolution. The level known as secondary structure is presently the best compromise between theoretical tractability and empirical accessibility on a large scale. The term secondary structure denotes any planar pattern of base pair contacts (Fig. 1). It is a topological concept and should not be confused with some kind of two-dimensional structure. In fact, all representations in Fig. 1 are equivalent. Planarity means that pairings between positions in different loop regions are not considered. In the circle representation of Fig. 1B, chords never cross. The secondary structure repertoire of a sequence consists of all base pairings that are compatible with the rules A·U, G·C and

G·U.

A secondary structure graph can be uniquely decomposed into loops and stacks[7]. A stack is a run of adjacent pairs, corresponding to a double-helical arrangement in the three-dimensional structure. A loop consists of varying numbers of unpaired bases and stacks that originate from it (Fig. 1A). The major stabilizing free energy contribution comes from stacking interactions between adjacent base pairs. The G·C/G·C stacking interaction is roughly 3 (2) times the A·U/A·U (A·U/G·C) interaction[8]. Loops, in contrast, are destabilizing. The formation of a stack necessarily causes the formation of a loop. This "frustrated" energetics can generate a very rugged energy landscape over the secondary structure space of a sequence. Because the energetically relevant units are loops and stacks rather than individual base pairs, a structure that minimizes free energy is oftentimes unique and quite different from one that maximizes the base pair count. For example, the 77 nucleotide histidine $tRNA^{his}$ (EMBL accession RH1660) has one minimum free energy secondary structure (22 base pairs) and 149, 126 secondary structures realizing the maximum of 26 base pairs[9].

The free energy contributions of stack and loop elements have been empirically determined[8,10–13]. Combinatorial algorithms[7,14–16], based on a powerful optimization technique known as dynamic programming[17], reference these parameters in computing the minimum free energy structure of a sequence. This procedure, however, does *not* consider the dynamical folding *process* by which a sequence acquires its structure.

Base pairings that break planarity are called pseudoknots and are considered to be tertiary structure elements. Thermodynamic[18] and kinetic[19] algorithms that account for pseudoknots have been developed recently. Although widespread in naturally occurring RNAs, pseudoknots will not be considered here.

The secondary structure participates as a geometric, kinetic and thermodynamic scaffold[20] in the formation of the three-dimensional structure, which involves bringing secondary structure elements into proximity by means of pseudoknots, non-standard base pairings and bivalent counterions. Its correlation with functional properties of the tertiary structure is evidenced by phylogenetic conservation[21]. Further details on the biophysical and computational aspects of RNA secondary structure can be found in several reviews [7,22–24].

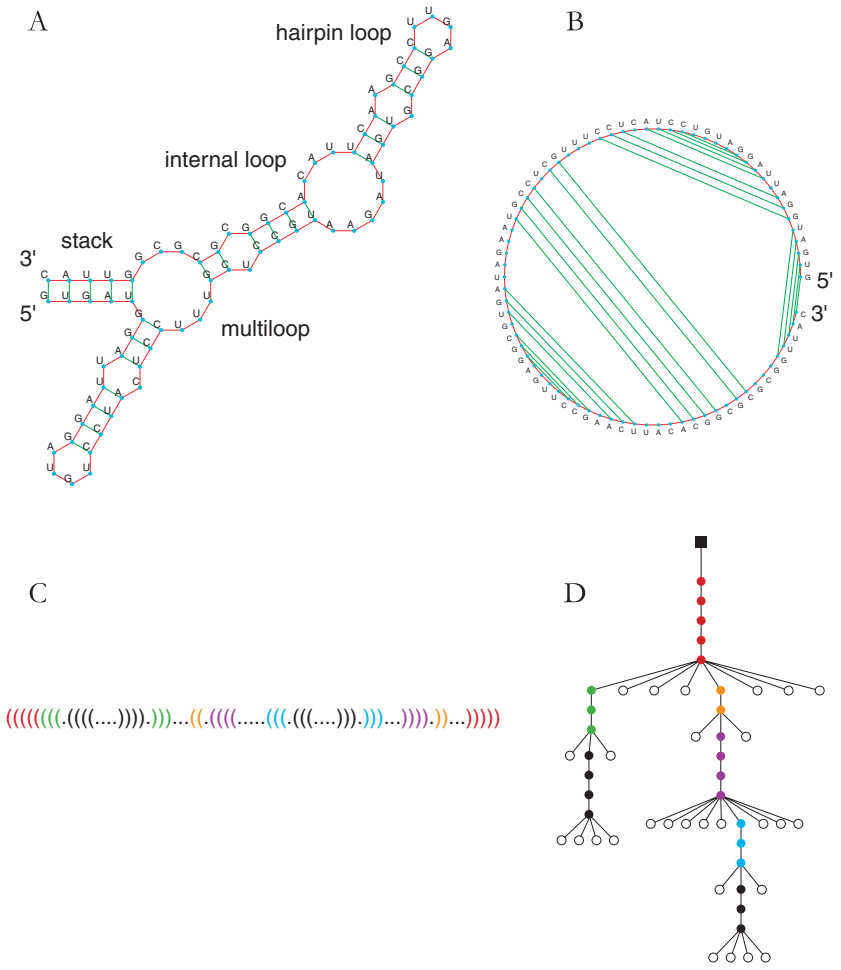

Figure 1: **RNA secondary structure representations.** A secondary structure is a graph of contacts between nucleotides at positions $i = 1, \ldots, n$ along the sequence. Position 1 is the 5'-end. The graph has two types of edges: the backbone connecting nucleotide i with nucleotide $i+1$ (red) and hydrogen-bonded base pairings between non-adjacent positions (green). The set of base-pairings, P, must satisfy two conditions: (i) each nucleotide can pair with at most one other nucleotide (green edges in **A** or **B**), and (ii) pairings cannot cross (this is best expressed by representation **B**, where chords, standing for base pairs, are not allowed to intersect). Condition (ii) expresses (outer)planarity. **(A)** Typical visualization of a structure. **(B)** Circle representation. **(C)** Line-oriented representation. A dot stands for an unpaired position and a pair of matching parentheses indicates paired positions. **D** Tree representation. Base pairs are internal nodes and unpaired positions correspond to leaves. The top node (square) keeps the tree rooted for structures with dangling 5'- or 3'-ends and joints. **(A-D)** contain exactly the same structure information.

I will henceforth refer to the (minimum free energy) secondary structure as (mfe) shape. A wealth of different phenotypes can be defined and computed at this level of structure (Fig. 2).

Energy (kinetic) landscape. Thermodynamic folding algorithms only map a sequence into its global mfe shape. In contrast, kinetic algorithms are concerned with how a sequence folds and hence with the rates and paths through configuration space that constrain and promote folding[19,25–28]. A sequence may fold reliably into a shape other than the mfe shape or it may switch between metastable shapes with a long lifetime relative to the molecule's interaction time scale[29,30].

In modelling the folding process, the key concept is the *energy (or kinetic) landscape* (Fig. 2A). The configuration space of the many shapes compatible with a given sequence is defined in terms of elementary "moves" that interconvert shapes. The free energy associated with each shape gives rise to an energy landscape over the configuration space and the energy differences between adjacent shapes determine (roughly) the transition probabilities[19,28,31]

The energy landscape of a sequence is the RNA analogue of Waddington's developmental or epigenetic landscape[32]. Sequences folding into the same mfe shape can differ profoundly in their energy landscapes. In this limited sense, the RNA model is capable of mimicking an "evolution of development". The analogy breaks down when the mechanisms of development themselves evolve[33]. After all, an RNA sequence doesn't code for the base pairing rules.

Phenotypic plasticity. A sequence can wiggle between alternative low-energy shapes as a consequence of energy fluctuations comprising a few kT (k is the Boltzmann constant and T the absolute temperature). The phenotypic plasticity of a sequence is quantified by the probability p_{ij} that positions i and j are paired with one another. The p_{ij} are obtained from a calculation of the partition function[34], $Z = \sum_i \exp(-E(S_i)/kT)$, where $E(S_i)$ is the free energy of shape S_i. Fig. 2B depicts a matrix of such probabilities (upper triangle), compared to a rendition in the same format of the mfe shape (lower triangle). To further express plasticity as a single number, the matrix of base pairing probabilities can be collapsed into an entropy-like quantity[35].

A useful description of plasticity, to which I shall return later, is the set of all shapes within a free energy interval of, say, $5kT$ (3 kcal/mol at 37°C) from the groundstate[9], Fig. 2C. This set is termed the *plastic repertoire*[36].

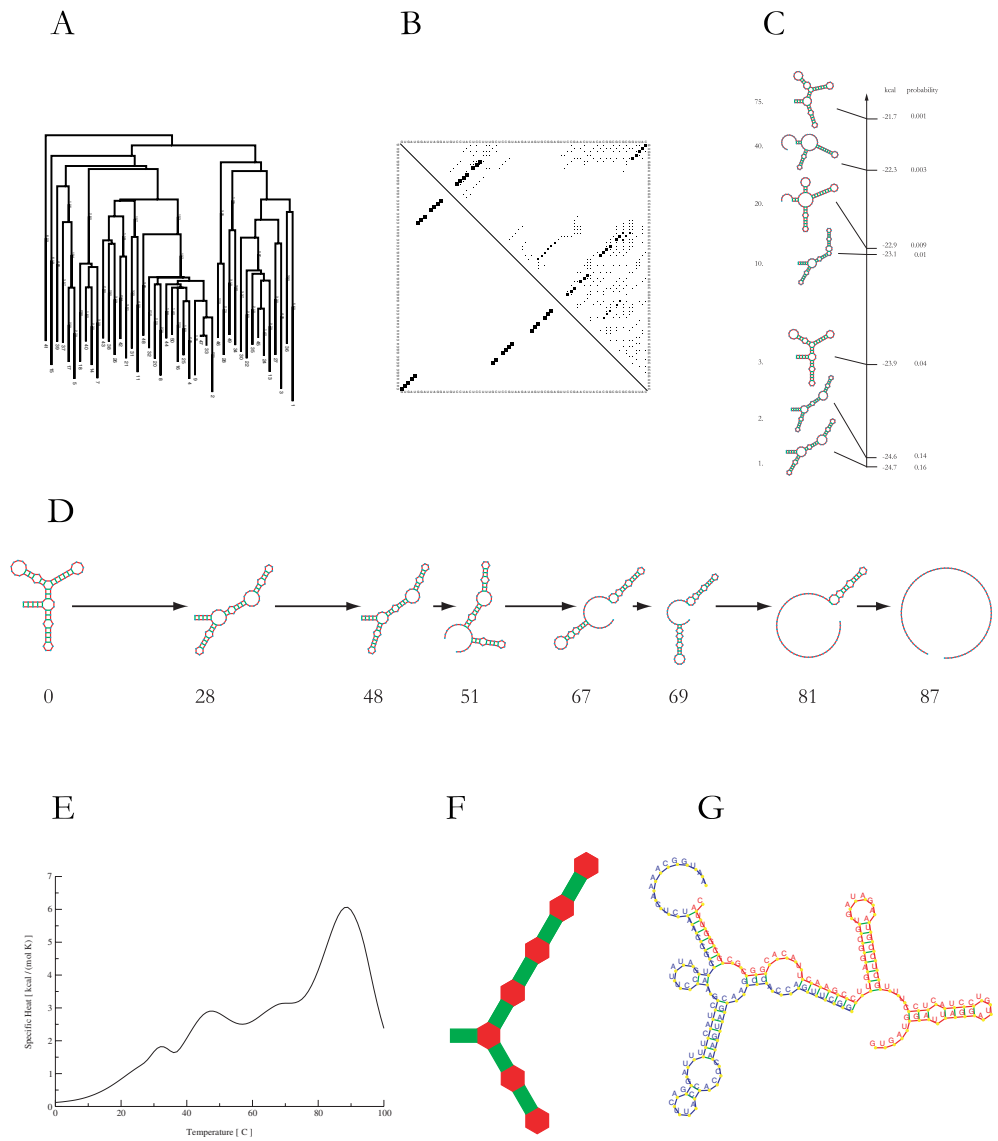

Figure 2: **RNA phenotypes.** All examples refer to the sequence of Fig. 1. **(A)** Developmental landscape. The graph shows the barrier-tree[28] of the low-energy portion of the free energy landscape. The barrier-tree expresses the likelihood of a conversion from one shape configuration into another. This is how it should be read. The vertical dimension means free energy. The leaves of the tree are shapes (not shown) that correspond to local energy minima and the highest point on the

direct route from one minimum to any other indicates the energy barrier that controls the likelihood of that route. The likelihood (at constant temperature) of switching between any two configurations is exponential in the (negative) height of that barrier. High barriers mean very long crossing times (or low crossing probabilities). The structure of this barrier landscape determines the folding paths and rates. **(B)** The matrix of base-pairing probabilities. The size of a dot at location (i, j) in the upper triangle depicts the probability that position i is paired with position j ($i < j$). For comparison, the lower triangle displays the pairing pattern of the mfe shape. While the mfe pattern is predominant in the upper triangle, dots of significant size exist at alternative positions, indicating different folds. **(C)** Plastic repertoire. The set of shapes within $5kT$ (3 kcal/mol at 37° C) from the mfe shape. Only a few shapes are shown and their relative energy is indicated on the vertical bar whose total length corresponds to 3 kcal/mol. **(D)** Norm of reaction I (melting profile). The series of mfe shapes as the temperature is raised from 0° C to the temperature at which this sequence loses all secondary structure. **(E)** Norm of reaction II (specific heat). The specific heat is given by $\mathcal{H} = -T\partial^2 G/\partial T^2$ with $G = -RT \log Z$ (G is the Gibbs free energy, R the gas constant, T the absolute temperature and Z the partition function). $\mathcal{H}$ profiles the changes in the statistical weights of the shape configurations available to an RNA sequence as the temperature is raised. The humps in the specific heat indicate the major melting transitions at 28, 48, 67 and 87 degrees Celsius shown in (D). This function can be measured by differential scanning calorimetry. **(F)** Loop-structure: the relative arrangement of loops disregarding the size of stacks and loops. **(G)** Joint shape of two hybridized RNA sequences.

Knowing the partition function Z and assuming thermodynamic equilibration, we compute the fraction of time a molecule spends in any shape S_i of its plastic repertoire as $p(S_i) = \exp(-E(S_i)/kT)/Z$. The extent to which the mfe shape is the most occupied configuration varies significantly among sequences.

Plasticity so-defined emphasizes an *intrinsic* phenotypic variance, induced by molecular energy fluctuations at nonzero temperature. A more traditional use of the term, known as *norm of reaction*[37], refers to persistent phenotypic change as a function of environmental parameters. The biophysical analogue in RNA is the melting profile, that is, the suite of mfe shapes as a function of temperature (Fig. 2D), or its statistical equivalent, the specific heat (Fig. 2E).

The difference between the two plasticities (intrinsic variance and norm of reaction) is best understood in terms of the free energy landscape underlying the folding behavior of a sequence. The topography of a free energy landscape depends non-monotonically on the temperature (the environment). Plasticity understood as a norm of reaction refers to transitions between mfe shapes as the free energy landscape is deformed by temperature, while plasticity understood as intrinsic phenotypic variance refers to transitions between different shapes on a constant free energy landscape.

Loop-structure. A shape can be coarse-grained by disregarding the size of loops and the length of stacks, retaining only the relative arrangements of loops (Fig. 2F). This skeletal morphology will be referred to as *loop-structure.*

Interaction phenotypes. Base pairs may be formed within or between molecules. Straightforward extensions of standard thermodynamic and kinetic folding algorithms yield the joint shape acquired by two (or more) sequences[22]. This defines a natural notion of interaction (Fig. 2G) which could form the basis for RNA models of coevolution.

3 Neutrality, epistasis, canalization

I begin with some terminology. Sequences that differ from a reference sequence by n point mutations are called *n-error neighbors.* A *neutral mutation* is a nucleotide substitution that preserves the mfe shape (but it may affect everything else, such as free energy, plastic repertoire and kinetic landscape). A one-error neighbor that preserves the same mfe shape is termed a *neutral neighbor.* A sequence position that allows for at least one neutral mutation is termed a *neutral position* (Fig. 3). The *neutrality* of a sequence is the fraction of neutral (one-error) neighbors.

Neutrality is here defined with respect to mfe shape, not fitness. Fitness is a function from phenotypes to numbers and if phenotype is defined as mfe shape, then neutrality extends to fitness as well. If phenotype and fitness are defined in terms of the plastic repertoire of a sequence, I shall still refer to sequences that share the same mfe shape as neutral, even when their plastic repertoires (and fitness) differ.

Epistasis means that the phenotypic consequences of a mutation at gene i depend on the genetic background provided by the remaining genotype[38]. This

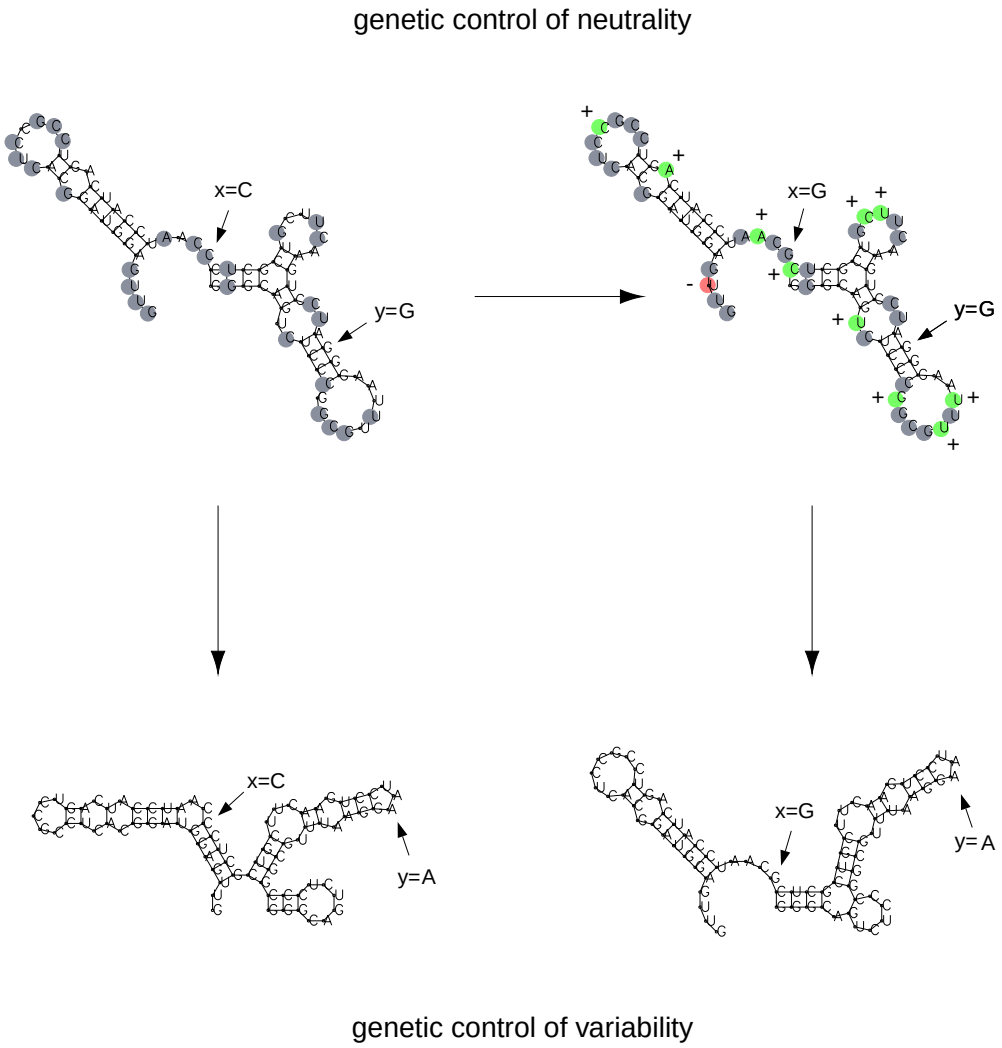

Figure 3: **Epistasis.** Bullets indicate a neutral position. In the top left sequence, position x is neutral because the substitution of G for C preserves the shape, as shown at the top on the right. Yet, neutral positions *do* change as a consequence of a neutral mutation. The green and red bullets indicate positions that have become or stopped being neutral, respectively. The lower part illustrates that the neutral mutation from C to G at x influences the consequences of swapping A for G at the (non-neutral) position y.

dependency is mediated by networks of interactions among gene products. The same concept applies to RNA, when substituting "gene" with "sequence position". The transparency (but also the limitation) of the RNA genotype-phenotype model derives from the identity of epigenetic and epistatic interactions, since phenotype is defined directly in terms of interactions among sequence positions.

A mutation changes the base pairing possibilities of a sequence and hence the network of epistatic interactions. The mfe shape shown at the top left of Fig. 3 remains the same if C is substituted by G at the position labelled x. Yet, whether x is C or G determines which mfe shape is obtained as a result of mutating position y from G to A. More subtly, the neutral substitution from C to G at x alters the number and identity of neutral positions.

The tendency of a sequence to adopt a different shape upon mutation (*variability*) is a prerequisite for its capacity to evolve in response to selective pressures (*evolvability*). In this sense, variability underlies evolvability[2,39]. Fig. 3 illustrates that variability (quantified as the number of non-neutral neighbors) is sequence dependent. Variability can therefore evolve[39,40].

Canalization[41–43] is a biological concept related to robustness in physics and engineering[44] aimed at quantifying a system's resilience to perturbation. Biologists distinguish between environmental and genetic canalization, depending on the nature of the perturbation. In our highly simplified RNA context, genetic canalization is phenotypic robustness to mutation and environmental canalization is phenotypic robustness to environmental change or noise. Neutrality, as defined here, is basically a measure of genetic canalization, while plasticity is the converse of environmental canalization.

4 The statistical deep-structure of the RNA folding map and its consequences

When RNA folding was employed as a toy-model to study evolution in populations of individuals equipped with a biophysically grounded genotype-phenotype relation[45], it soon became clear that the evolutionary dynamics must be understood in terms of the statistical characteristics of folding. Frequency distributions of structural elements and shape correlation functions in sequence space were estimated by means of random walks and by folding

large ensembles of random RNA sequences of varying length, nucleotide alphabet and composition[46–49]. This simple shift towards a statistical view of the folding map brought into focus features that were insensitive to variations in the free energy parameters, the algorithmic details and the accuracy of prediction. Here I focus on this "deep structure" of the folding map and its consequences for evolutionary innovation and dynamics.

4.1 One phenotype, many genotypes

There are significantly more sequences than secondary structures[50]. An asymptotic upper bound on the number of possible shapes, $\mathcal{S}_n$, for sequences of length n is $\mathcal{S}_n = 1.48\, n^{-3/2} 1.85^n$, compared to 4^n possible sequences[51]. Exhaustive folding indicates that the number of actually realized shapes is considerably smaller than this upper bound[52]. The situation is not altered significantly by accounting for pseudoknots[53]. To appreciate the degree of degeneracy of the folding map, consider the set of all possible binary GC sequences of length 30 (GC-30). $1.07 \cdot 10^9$ $(= 2^{30})$ sequences fold into only $218,820$ shapes[52,54].

The frequency of shapes is strongly biased. The rank-ordered frequency distribution follows qualitatively a generalized Zipf-law[50], $f(r) = A(B + r)^{-\gamma}$, where r denotes rank (the most frequent shape has rank 1) and $f(r)$ is the fraction of sequences folding into the shape of rank r. For AUGC-sequences of meaningful size, such distributions can be presently computed for loop-structures only. Abundancy distributions for fully resolved secondary structures were obtained by exhaustively folding all GC-30 sequences[52]. The constants A, B and γ depend on sequence length and nucleotide alphabet. Examples for γ-values are 1.7 (AUGC-100, loop-structures) and 2.9 (GC-30, full secondary structures). The Zipf-distribution assigns a high abundancy to a tiny number of shapes compared to those in the power-tail. A *frequent shape* may be defined as one realized by more sequences than the average [55], $4^n/S_n$, which amounts to 4907 in the case of GC-30 sequences. With this definition, only 10.4% of the GC-30 shapes are frequent, yet 93% of all GC-30 sequences fold into them. As sequence length increases, a decreasing percentage of shapes is frequent, while being realized by an increasing percentage of sequences[52].

A frequent shape compromises between two opposing trends. First, it must

be realizable with sufficient thermodynamic stability. Otherwise, mutations would be too likely to alter the shape, reducing the number of sequences folding into it. Second, it must be frequent on combinatorial grounds. While long stacks enhance the thermodynamic stability of a shape, they lower its combinatorial realizability by constraining the choice of nucleotides at paired positions. The open chain is combinatorially best, but thermodynamically among the worst (for longer sequences). The opposite is the case for a long hairpin. Frequent shapes occupy a middle ground by allocating base pairs to separate stacks, since each stack creates a loop that enhances combinatorial realizability. One is tempted to speculate that among these shapes are also the most "interesting" ones, since the diversity of structural elements can be exploited at the tertiary level to create elements with potential functionality, such as "pockets", "arms", "tweezers", "spacers" and the like.

4.2 Neutral networks

A sequence folding into a frequent shape has typically a significant fraction of neutral one- or two-error neighbors. The same holds for these neighbors. This results in an extensive, mutationally connected network of sequences, for which we coined the term *neutral network*[50] (see, for a schematic example, the "green" network in Fig. 4). Models based on random graphs formalize neutral networks as a percolation phenomenon[56]. The possibility of changing the genotype while preserving its phenotype is both a manifestation of phenotypic robustness to genetic mutations and a key factor underlying evolvability. This only seems contradictory. Imagine a population with phenotype A in a situation where phenotype B would be advantageous. Phenotype B, however, may not be accessible in the vicinity of the population's current location in genotype space. In the mythical image of a rugged fitness (or adaptive) landscape[57], the population would be stuck at a local peak, forever waiting for an exceedingly unlikely event to deliver the right combination of several mutations. Yet, if phenotype A has an associated neutral network in genotype space, the population can drift on that network into far away regions, vastly improving its chances of encountering the neutral network associated with phenotype B[3,58–60], see Fig. 4. Neutral networks enable phenotypic innovation by permitting the gradual accumulation of silent mutations. These alter the web of epistatic interactions, enabling a subsequent mutation to become phenotypically consequential. Recall that neutral mutations also

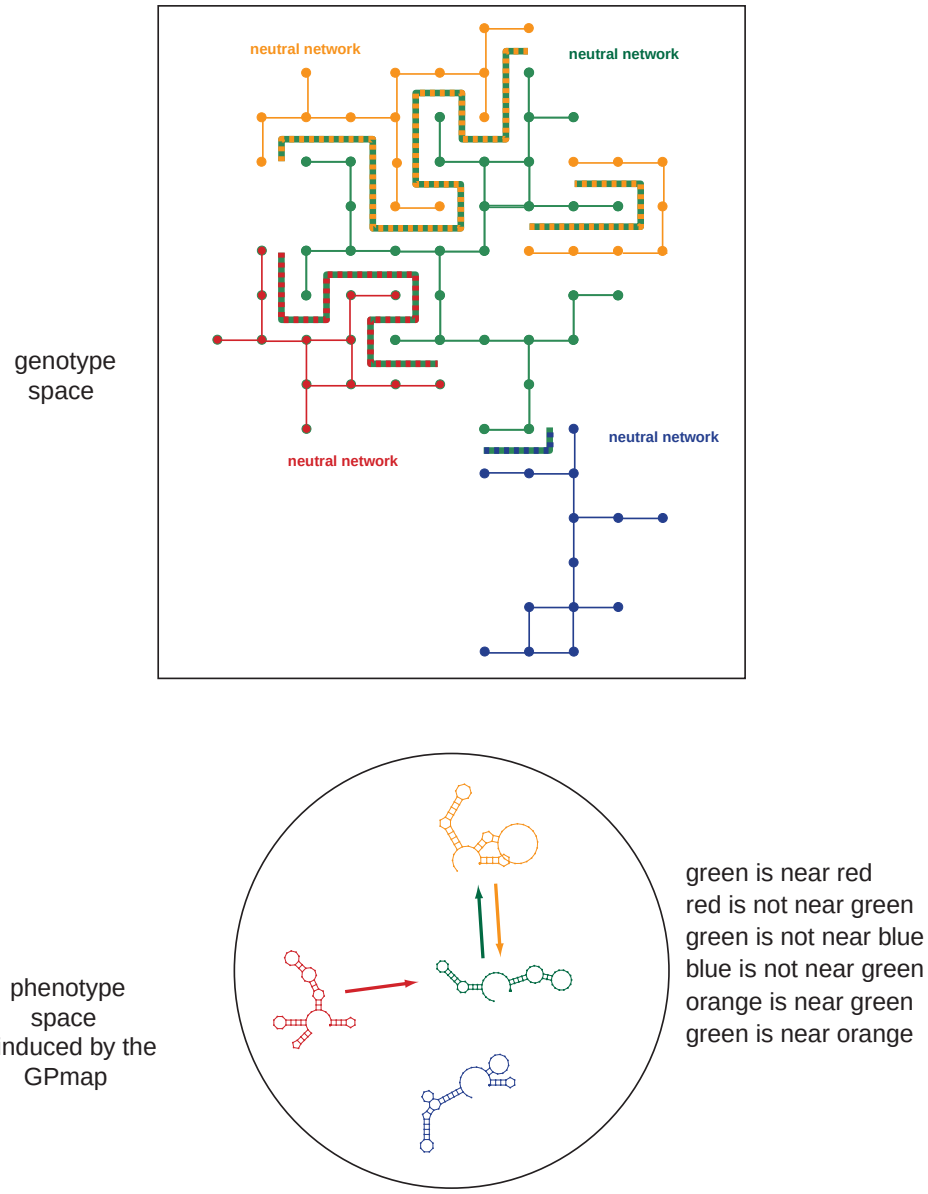

Figure 4: **Neutral networks induce the topology of shape space.** A schematic depiction of neutral networks in sequence space. (For a more accurate representation, the reader ought to imagine at least a 100-dimensional space.) A population located in the upper right corner of the network of sequences that possess the "green" phenotype cannot access the "blue" phenotype in its genetic

vicinity. Yet, the population can diffuse on the green network until it encounters the blue network. Neighborhood between shapes (phenotypes) is defined in terms of the fraction of shared boundary between the corresponding networks in sequence space (indicated by thick lines with alternating colors), see text for details. In this schematic of four networks, the red network is near the green one, because a random step out of red has a high probability of yielding green. The green network, however, is not near the red one, because a random step out of green has a low probability of yielding red. This effect results from very differently sized networks, like those associated with the loss and formation of a stack, Fig. 6A. The alternative case is one in which the networks have similar size, but border one another only rarely; green and blue are not in each other's neighborhood. This correpsonds to the shift transformation of Fig. 6A.

---

influence the degree of neutrality (Fig. 3), causing the connectivity within a neutral network to be highly variable.

The existence of neutral paths in RNA sequence space was impressively demonstrated in a recent experiment. Schultes and Bartel[61] constructed an *intersection sequence*[28,62] between two evolutionarily unrelated catalytic RNAs with no fortuitous shape similarities – a class III self-ligating ribozyme evolved in vitro and a naturally occurring hepatitis delta virus self-cleaving ribozyme. The intersection sequence was located about 40 mutations from each original ribozyme and performed both catalytic tasks at highly reduced rates (Fig. 5). Within a few mutations from the intersection sequence, two sequences were found, each specialized to one task with a catalytic activity comparable to the original ribozyme. Starting from these sequences, two paths of about 40 steps were identified that led all the way to each original sequence, while maintaining the level of catalytic activity. This demonstrates the existence of neutral paths for different ribozyme folds and their close apposition, compatible with the scenario described above.

The importance of neutrality for the diffusive motion of allele frequencies and their rate of fixation was established by Kimura[63,64] and has since been extended in numerous ways[65–67]. The concept of neutral networks, however, has brought into focus issues that go beyond the dynamics of allele frequencies. First, a population does not move entirely randomly over a network, but tends to concentrate at highly interconnected regions[68]. Thus, a

selection/mutation balance on a neutral network automatically yields phenotypes that are relatively robust to mutations[68]. Second, evolution requires the hereditary transmission of a phenotype rather than any *particular* genotype. In a neutral network, the loss of a genotype does not imply the loss of its phenotype, because many neighboring mutants still map into that same phenotype. This significantly increases the mutation rate threshold at which heredity breaks down[56,58]. Third, the partitioning of sequence space into neutral networks suggests a notion of nearness between phenotypes that imposes a new topology on phenotype space (Fig. 4) with a corresponding formalization of evolutionary continuity[3,6,69,70] detailed below.

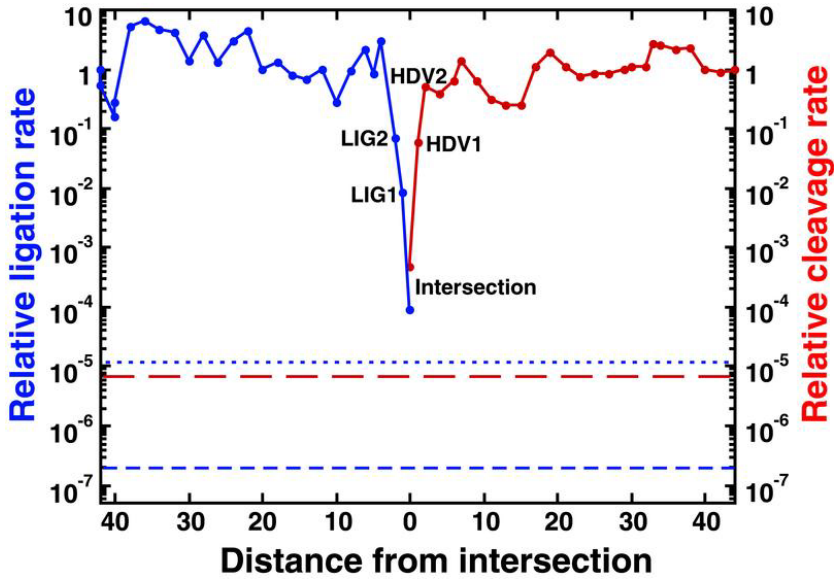

Figure 5: **The Schultes-Bartel neutral path experiment.** Schultes and Bartel[61] have established the existence and apposition of neutral paths by stepwise changing sequences over more than 40 positions, while retaining structure and functional activity. See text for details. (Reprinted with permission from E. A. Schultes and D. P. Bartel, "One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds", *Science*, **289**, 448–452 (2000). Copyright 2000 American Association for the Advancement of Science.)

4.3 Shape space covering

Neutral networks of frequent shapes are sponge-like objects that live in a high-dimensional sequence space. These networks are strongly interwoven. Imagine a (hyper)sphere centered at some random sequence. How big must its radius be, for it to contain at least one point of the neutral network of each frequent shape? Numerical procedures[50] and analytical estimates[54] yield answers that are significantly smaller than the radius $(n/2)$ of sequence space. For AUGC sequences of length $n = 100$ the radius of that sphere is about 15. In other words, given a random sequence of length 100, 15 point mutations are, on average, sufficient to realize any frequent shape. A relatively small volume of sequence space around each random sequence realizes the entire statistically relevant portion of shape space. In this case, the haystack in which evolution has to look for the needle has been reduced by a factor of 10^{37}. We refer to this phenomenon as *shape space covering*[50], borrowing a term coined by Perelson and Oster in an immunological context [71].

Changing perspective on the same issue, consider all shapes found in the one-error neighborhood off a neutral network. These shapes are, in principle, accessible to a population that is drifting on the network at small mutation rates[59]. It turns out that the vast majority of loop-structures (Fig. 2F) formed by randomly generated sequences (of fixed length) occur in the one-error neighborhood of the neutral network of any frequent shape[3]. Thus, for a frequent shape α and a frequent loop-structure Ω, there likely exists at least one sequence folding into α such that a single point mutation tips it into some shape whose loop-structure is Ω.

4.4 The topology of RNA shape space

A space is formally a set with a structure that derives from relationships among its elements. A relation of distance gives rise to a metric space. For example, the distance between two RNA sequences is the number of positions in which they differ (Hamming distance). For a metric space to have biological relevance, the distance must be defined in terms of naturally occurring operations that interconvert elements, such as point mutations in sequence space. This raises the question about the structure of shape space.

A distance (or similarity measure) between shapes could be defined in terms of formal "edit"-operations on the shape representations of Fig. 1. This is perfectly useful for defining selection or sorting criteria. Yet, a shape space so-defined is of no help in understanding evolutionary histories, because there are no physical operations that interconvert shapes *heritably.* To heritably convert one shape into another, requires mutating the underlying sequence, and this forces the folding (or genotype-phenotype) relation into the picture.

Recall that a neutral network is the set of all sequences adopting a particular shape. Fontana and Schuster construct shape space by defining a relation of accessibility between two shapes, α and β, in terms of the adjacency of their corresponding neutral networks in sequence space[3,69], Fig. 4. The *boundary* of a neutral network consists of all sequences that are one mutation off the network. The intersection of the neutral network of β with the boundary of the neutral network of α, relative to the total boundary of α's network, is a measure of the probability that one step off a random point on the neutral network of α there is a sequence folding into β. Accessibility, however, is not symmetric and therefore not a distance (Fig. 4). To wit: the loss and formation of a stack. A stack in shape α will be only marginally stable in most sequences realizing α. As a consequence, point mutations are more likely to cause its loss. In contrast, creating a stack in a single mutation requires specially poised sequences. RNA folding is a simple mechanism giving rise to strongly asymmetric transition probabilities. A shape β may be significantly easier to access from shape α than the other way around.

To convert accessibility distributions into a binary attribute of nearness, we define the *neighborhood* of shape α as the set containing α and all shapes accessible from α above a certain likelihood[3,69] – a "frequent neighbor", akin to the notion of a frequent shape. The major implications of this construction do not depend on the exact value of the cutoff point[6]. Because accessibility is asymmetric, shape β may be near (read: in the neighborhood of) α, but α may not be near β. This construction of shape-neighborhood is technically consistent with the formalization of the neighborhood concept in topology [6,72]. Phenotype space has thus been organized into neighborhoods without assuming a distance. To non-topologists, this may seem counterintuitive, since common sense conceives neighborhood in terms of "small distance".

The notion of neighborhood is sufficient to define *continuity* mathematically. In the present context, a genetic path is continuous, if the offspring of a

sequence is near that sequence in the topology of genotype space. Specifically, offspring and parent differ by one point mutation. Similarly, a path in phenotype space is continuous if the phenotype of the offspring is near the phenotype of the parent in the accessibility topology just defined. Think of an *evolutionary trajectory* as a time series of genotypes (G) with their associated phenotypes (P) – a GP path. A GP path is continuous, if it is continuous at the genetic and the associated phenotypic level. The key question now becomes: given any two shapes, is there a continuous GP path connecting them? The answer is no. There is a well-defined class of shape transformations (Fig. 6A) that is *dis*continuous along *any* continuous genetic path [3]. A transition from one neutral network to another involving such a shape transformation, while possible by a single point mutation, depends on very special, hence relatively rare, sequences. The two classes of discontinuous transformations are stack formations and shifts (Fig. 6A). They both require the *simultaneous* change of several base pairs, since partial rearrangements result in thermodynamically unstable intermediates. The notion of discontinuity reflects precisely those transformations that are difficult to achieve by virtue of the mechanisms underlying the map from genotype to phenotype.

A few observations deserve emphasis. First, (dis)continuity cross-cuts morphological (dis)similarity. Some transitions between similar shapes are discontinuous (e.g. the shift in Fig. 6A) and some transitions between dissimilar shapes are continuous (e.g. the loss of a stack). Second, the notion of discontinuity defined here is *not* related to sudden jumps in fitness or the discreteness in the variation of a trait. The classes of discontinuity are caused by the genotype-phenotype map and thus remain the same *regardless* of the further mapping from phenotypes to fitness. (Of course, the particular shapes observed at discontinuous transitions will depend on the fitness map.) Third, the dynamical signature of this phenotype topology is punctuation (Fig. 6B). A population of replicating and mutating sequences under selection drifts on the neutral network of the currently best shape until it encounters a gateway to a network that conveys some advantage or is fitness-neutral. That encounter, however, is evidently not under the control of selection. While similar to the phenomenon of punctuated equilibrium recognized by Eldredge and Gould[74] in the fossil record of species evolution, punctuation in evolving RNA populations occurs in the *absence* of externalities (such as meteorite impact or abrupt climate change in the species case). We refer to it as *intrinsic punctuation*, since it reflects the variational properties of the underlying

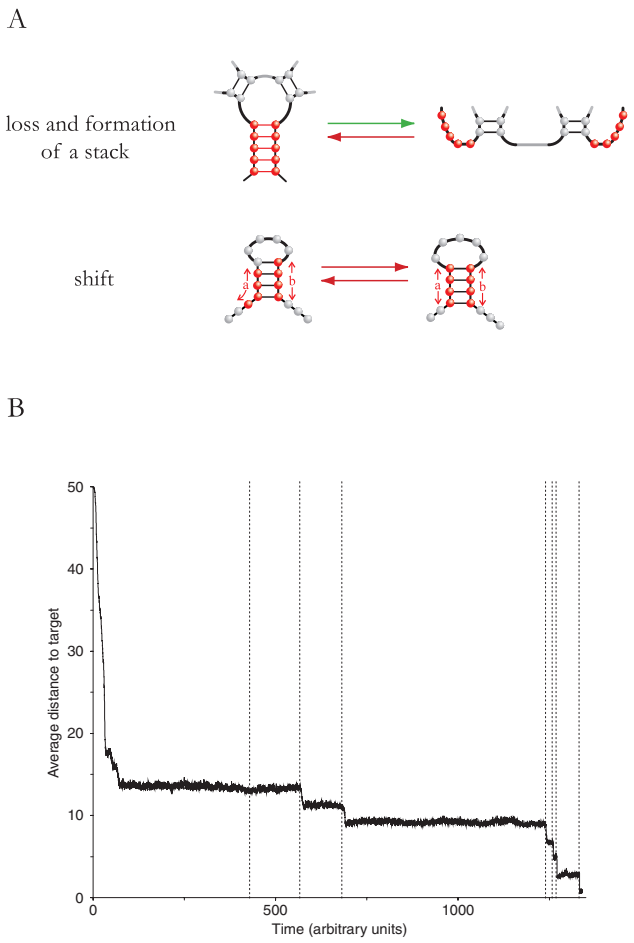

Figure 6: **Intrinsic punctuation.** **(A)** Discontinuous shape transformations in RNA. A green (red) arrow indicates a (dis)continuous transition. There are several other transformations in the shift category (not shown), see[3,69]. Continuous transitions, other than the wholesale loss of a stack, involve the elongation or shortening of a stack by one base pair (not shown). **(B)** Punctuation in evolving RNA populations. A population of RNA sequences evolves under selection for a specific target shape. The average fitness shows periods of stasis punctuated by sudden improvements. (Fitness is maximal when the distance to the target shape has become zero.) Yet, the phenotypic discontinuities (marker lines), as revealed by an ex post reconstruction of the evolutionary trajectory, are not always congruent with the fitness picture. The first two jumps in fitness are continuous in the genotype-phenotype picture and a crucial discontinuous transition is fitness neutral (first marker line)[69,73].

developmental architecture.

4.5 Plastogenetic congruence, canalization and modularity

The RNA folding map exhibits a consequential correlation between plasticity and variability. The mfe shapes realized in the one-error neighborhood of a sequence are, with high frequency, a subset of the shapes in the plastic repertoire of that sequence[36] (Fig. 7A). A point mutation oftentimes stabilizes a shape that is suboptimal in the parent, promoting it to mfe shape in the mutant. We termed this correlation *plastogenetic congruence*[36] to emphasize that plasticity and (genetic) variability are correlated, because both depend on the *same* structure-forming mechanisms. The significance of such congruence consists in directly coupling a change in plasticity to a change in genetic variability. While genetic variability is a property affecting the future evolution of a lineage, plasticity affects an individual in the present and can be an easy target of selection. Since plasticity correlates with the shape of things to come, selection on plasticity has a direct impact on a future capacity. In RNA, environmental canalization (the reduction of plasticity by tightening the genetic determination of the mfe shape) entails genetic canalization (the curtailing of phenotypic novelty accessible by mutation). This was hypothesized by Wagner et al.[75] on the basis of population genetic models. Selection pressures favoring the reduction of plasticity may arise from the fitness costs associated with plasticity[37].

Recall that a neutral network is, by definition, neutral only with respect to the mfe shape. If fitness depends on plasticity, the fitness neutrality of a neutral network is broken. Under conditions that favor low plasticity, selection will drive a population into regions of the network consisting of sequences that fold into the currently best shape as stably as possible. Fig. 7B illustrates the remarkable degree of canalization attainable in RNA. As many as 1200 shapes are within $5kT$ from the groundstate of a random sequence, while a sequence thermodynamically optimized to fold into the same mfe shape has as few as 5 alternatives in its plastic repertoire. Sequences with a highly stabilized mfe shape still form a large (and now almost fitness-neutral) subnetwork of the network for that shape. A population is therefore still able to drift in sequence space. Increased thermodynamic stability (low plasticity) implies, however, an increased buffering of the mfe shape with

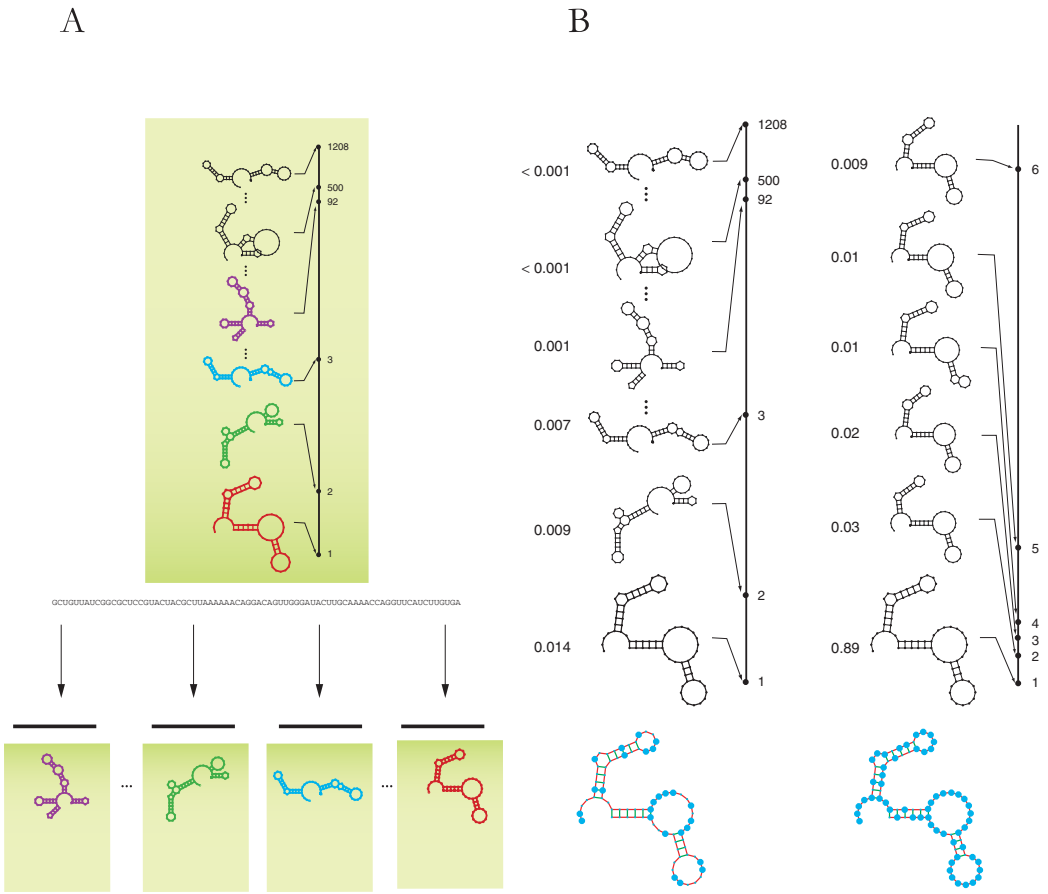

Figure 7: **Plastogenetic congruence.** **(A)** Plasticity mirrors variability. **(B)** Canalization. Both sequences fold into the same mfe shape. The sequence on the left is randomly chosen and the one on the right has been evolved for reduced plasticity. Genetic canalization has increased dramatically in the latter, as indicated by the blue dots marking neutral positions.

respect to genetic mutations (neutrality). In addition, the shapes in the plastic repertoire of a low-plasticity sequence tend to be morphologically similar to the mfe shape (Fig. 7B). By plastogenetic congruence, mutations that alter the mfe shape will do so only slightly. Stated in the language of neutral networks, only few networks of highly correlated shapes reach into the highly neutral regions of another network. Populations confined to those regions (on fitness grounds) have lost the potential for phenotypic innovation. The loss occurs in a manner imposed by the genotype-phenotype map, for if the highly neutral regions were random subnetworks of neutral networks, that loss would be, at worst, partial. We termed this situation *neutral confinement* [36].

The most striking feature distinguishing low from high plasticity sequences is the *modularity* of their mfe shape[36]. Modularity means autonomy of shape pieces on the basis of genetic (contextual), kinetic (developmental) and thermophysical (environmental) criteria. Consider the melting of a modular shape with rising temperature (Fig. 8A). The modules are identified by sharp and distinct melting temperatures at which they disappear as a whole, without disturbing the integrity of the remaining module(s). In the same vein, the shape of a module is largely insensitive to the sequence of flanking segments and can be therefore "cut" and "pasted". The organization of the energy landscapes is perhaps most illuminating (Fig. 8B). The energy landscape of a high-plasticity sequence allows for deep and frequent misfolds that are difficult to reverse. The low plasticity sequence, in contrast, has a perfect folding funnel (a notion coined in the context of protein folding[76,77]), which guides the folding process reliably and quickly into the native shape. This is Waddington's picture of developmental canalization at the level of an individual molecule.

The previously described loss of evolvability caused by the loss of plasticity occurred with respect to point mutations as the source of genetic change. Intriguingly, that process ends up with sequences whose shapes are highly modular. Once modularity has originated, however, the production of phenotypic novelty (and hence evolvability) could be regained by shifting the mechanisms of genetic change to recombination[36].

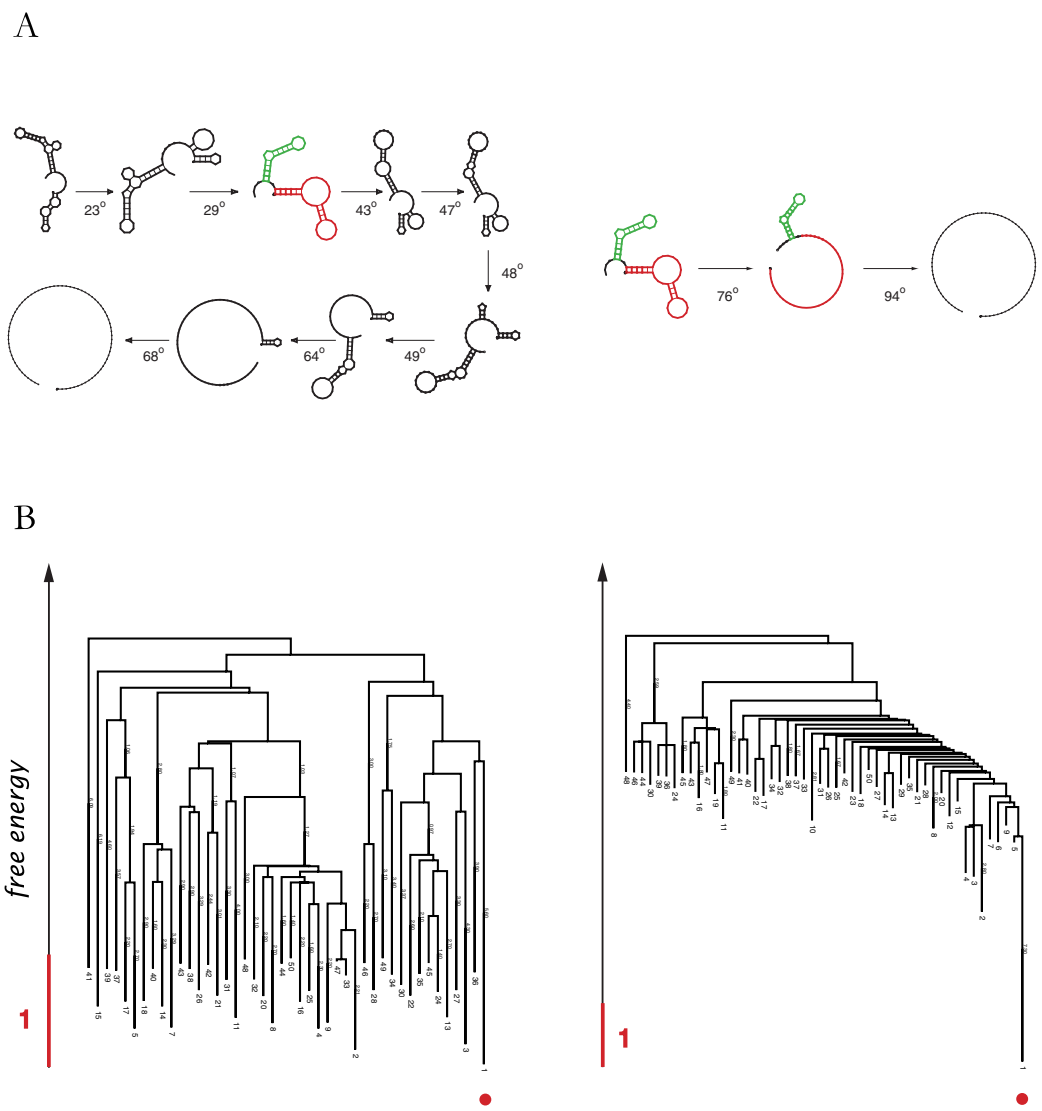

Figure 8: **Modularity.** **(A)** Melting profile and **(B)** free energy landscape of a random sequence (left) and a canalized sequence (right) sharing the same mfe shape at 37° C.

5 Conclusions: the topology of the possible

The arrival of a new phenotype must precede its survival in a population. Understanding the dynamics of evolution requires, therefore, charting the space of possible phenotypes and their transformations by genetic coin tosses. Part of the difficulty is to define an experimentally and theoretically tractable genotype-phenotype model whose mechanisms are reasonably well understood. The folding of RNA sequences into shapes is such a model. While RNA folding is of clear relevance to the molecular biology of the cell and its evolutionary history[78], it seems a leap of faith to claim that developmental biologists can learn anything of interest from it. As emphasized in the Introduction, RNA folding is clearly not a model of organismal development. Yet, it has an abstract connection with development in offering *a* realization of concepts – epistasis, plasticity, pathways (and networks of pathways) through state space, canalization, modularity, to mention only a few – that play an important role in thinking about development. While the detailed realization of these concepts is specific to RNA, their consequences and evolutionary interrelations may be more general.

Among the consequences brought to light by the computational and mathematical analysis of the RNA model are shape space covering, neutral networks and plastogenetic congruence. The conceptually deepest consequence is the emerging topology (accessibility structure) of RNA shape space, which leads one to question the widespread but unwarranted assumption of a highly symmetric Euclidean vector space as an adequate model of phenotype spaces in general. The departure from a metric structure reconciles known patterns of phenotypic evolution with a developmental perspective. This relieves the notion of fitness from having to do too much explanatory work.

Some of the consequences reviewed here are already generalizable to more complex ingredients of genotype-phenotype mappings, most notably protein folding[79–82] and the behavior of cellular control networks[83]. Recent models of signaling networks exhibit large connected volumes in the space of kinetic parameters that generate the same biological behaviors[83]. Are these behavioral equivalence classes in parameter space analogous to neutral networks in RNA sequence space?

The nature of discontinuous phenotype transformations described here has prompted Günter Wagner to speculate[84] that it may be experimentally im-

possible to reconstruct and demonstrate which genetic changes caused a major evolutionary innovation, precisely because the required genetic backgrounds are so rare and unlikely to be conserved in any extant species by virtue of neutral drift. But rather than ending with an unsettling note on the limits of what is knowable in evolution, I emphasize the decade-long convergence between experiment[85–88] and theory towards characterizing the distribution of structural and functional properties in RNA sequence space. The Schultes-Bartel experiment[61] has made contact and, perhaps, this constitutes a transition.

Acknowledgements

My research is supported by Jim Rutt's Proteus Foundation and by core grants to the Santa Fe Institute from the John D. and Catherine T. MacArthur Foundation, the National Science Foundation, and the U.S. Department of Energy.

References

1. Maynard-Smith J, Burian R, Kauffman S A, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. Developmental constraints and evolution. Quart. Rev. Biol. 1985; 60:265–287.

2. Gould S J. The Structure of Evolutionary Theory. Belknap/Harvard University Press Cambridge, MA 2002;.

3. Fontana W, Schuster P. Continuity in evolution: On the nature of transitions. Science 1998; 280:1451–1455.

4. Carroll S B, Grenier J K, Weatherbee S D. From DNA to diversity. Blackwell Science Malden, Massachusetts 2001;.

5. Wilkins A S. The evolution of developmental pathways. Sinauer Associates, Inc. Sunderland, Massachusetts 2002;.

6. Stadler B M R, Stadler P F, Wagner G, Fontana W. The topology of the possible: Formal spaces underlying patterns of evolutionary change. J. Theor. Biol. 2001; 213:241–274.

7. Zuker M, Sankoff D. RNA secondary structures and their prediction. Bull.Math.Biol. 1984; 46;(4):591–621.

8. Turner D H, Sugimoto N, Freier S. RNA structure prediction. Annual Review of Biophysics and Biophysical Chemistry 1988; 17:167–192.

9. Wuchty S, Fontana W, Hofacker I L, Schuster P. Complete suboptimal folding of RNA and the stability of secondary structures. Biopolymers 1999; 49:145–165.

10. Jaeger J A, Turner D H, Zuker M. Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA 1989; 86:7706–7710.

11. He L, Kierzek R, SantaLucia J, Walter A E, Turner D H. Nearest-neighbour parameters for G-U mismatches. Biochemistry 1991; 30:11124.

12. Walter A E, Turner D H, Kim J, Lyttle M H, Muller P, Mathews D H, Zuker M. Coaxial stacking of helices enhances binding of oligoribonucleotides and improves prediction of RNA folding. Proc. Natl. Acad. Sci. 1994; 91:9218 – 9222.

13. Mathews D H, Sabina J, Zuker M, Turner D H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 1999; 288:911–940.

14. Waterman M S, Smith T F. RNA secondary structure: A complete mathematical analysis. Mathematical Biosciences 1978; 42:257–266.

15. Nussinov R, Jacobson A B. Fast algorithm for predicting the secondary structure of single-stranded RNA. Proc. Natl. Acad. Sci. USA 1980; 77; (11):6309–6313.

16. Zuker M, Stiegler P. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research 1981; 9:133–148.

17. Bellman R E. Dynamic Programming. Princeton University Press Princeton, NJ 1957;.

18. Rivas E, Eddy S R. A dynamic programming algorithm for RNA structure prediction including pseudoknots. J. Mol. Biol. 1999; 285:2053–2068.

19. Isambert H, Siggia E D. Modeling RNA folding paths with pseudoknots: Application to hepatitis delta virus ribozyme. Proc. Natl. Acad. Sci. USA 2000; 97:6515–6520.

20. Banerjee A R, Jaeger J A, Turner D H. Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. Biochemistry 1993; 32:153–163.

21. Gutell R R. Evolutionary chracteristics of RNA: Inferring higher order structure from patterns of sequence variation. Curr. Opin. Struct. Biol. 1993; 3:313–322.

22. Hofacker I L, Fontana W, Stadler P F, Bonhoeffer S, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. Mh. Chem. 1994; 125:167–188.

23. Waterman M S. Introduction to Computational Biology: Sequences, Maps and Genomes. Chapman & Hall London 1995;.

24. Higgs P. RNA secondary structure: physical and computational aspects. Quart. Rev. Biophys. 2000; 33:199–253.

25. Mironov A, Lebedev V F. A kinetic model of RNA folding. BioSystems 1993; 30:49–56.

26. Gultyaev A P, van Batenburg F H D, Pleij C W A. The computer simulation of RNA folding pathways using a genetic algorithm. J. Mol. Biol. 1995; 250:37–51.

27. Morgan S R, Higgs P G. Evidence for kinetic effects in the folding of large RNA molecules. J. Chem. Phys. 1996; 105:7152–7157.

28. Flamm C, Fontana W, Hofacker I L, Schuster P. RNA folding at elementary step resolution. RNA 2000; 6:325–338.

29. Biebricher C K, Luce R. In vitro recombination and terminal elongation of RNA by Qb replicase. EMBO J. 1992; 11:5129–5135.

30. Flamm C, Hofacker I L, Maurer-Stroh S, Stadler P F, Zehl M. Design of multistable RNA molecules. RNA 2001; 7:254–265.

31. Jacob C, Breton N, Daegelen P. Stochastic theories of the activated complex and the activated collision: The RNA example. J. Chem. Phys. 1997; 107:2903–2912.

32. Waddington C H. The Strategy of the Genes. MacMillan Co. New York, NY 1957;.

33. Griesemer J. The units of evolutionary transition. Selection 2000; 1: 67–80.

34. McCaskill J S. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers 1990; 29:1105–1119.

35. Huynen M, Gutell R, Konings D. Assessing the reliability of RNA folding using statistical mechanics. J.Mol.Biol 1997; 267:1104–1112.

36. Ancel L, Fontana W. Plasticity, evolvability and modularity in RNA. J. of Exp. Zoology (Molecular and Developmental Evolution) 2000; 288: 242–283.

37. Scheiner S M. Genetics and evolution of phenotypic plasticity. Annu. Rev. Ecol. Syst. 1993; 24:35–68.

38. Wagner G P, Laubichler M D, Bagheri-Chaichian H. Genetic measurement theory of epistatic effects. Genetica 1998; 102/103:569–580.

39. Wagner G P, Altenberg L. Complex adaptations and the evolution of evolvability. Evolution 1996; 50:967–976.

40. Wagner A. Does evolutionary plasticity evolve? Evolution 1996; 50(3: 1008–1023.

41. Waddington C H. Canalization of development and the inheritance of acquired characters. Nature 1942; 3811:563–565.

42. Wilkins A S. Genetic analysis of animal development. John Wiley and Sons New York, NY 1986;.

43. Scharloo W. Canalization: Genetic and developmental aspects. Annu. Rev. Ecol. Syst. 1991; 22:65–93.

44. Csete M E, Doyle J C. Reverse engineering of biological complexity. Nature 2002; 295:1664–1669.

45. Fontana W, Schuster P. A Computer Model of Evolutionary Optimization. Biophys. Chem. 1987; 26:123–147.

46. Fontana W, Stadler P F, Bornberg-Bauer E G, Griesmacher T, Hofacker I L, Tacker M, Tarazona P, Weinberger E D, Schuster P. RNA folding and combinatory landscapes. Phys. Rev. E 1993; 47:2083–2099.

47. Fontana W, Konings D A M, Stadler P F, Schuster P. Statistics of RNA secondary structures. Biopolymers 1993; 33:1389–1404.

48. Tacker M, Stadler P F, Bornberg-Bauer E G, Hofacker I L, Schuster P. Algorithm independent properties of RNA secondary structure predictions. Eur.Biophys.J. 1996; 25:115–130.

49. Schultes E, Hraber P T, LaBean T. Estimating the contributions of selection and self- organisation in RNA secondary structure. J. Mol. Evol. 1999; 49:76–83.

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

51. Hofacker I L, Schuster P, Stadler P F. Combinatorics of RNA secondary structures. Discr. Appl. Math. 1999; 89:177–207.

52. Grüner W, Giegerich R, Strothmann D, Reidys C, Weber J, Hofacker I L, Stadler P F, Schuster P. Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. Mh.Chem. 1996; 127:355–374.

53. Haslinger C, Stadler P F. RNA structures with pseudo-knots. Bull. Math. Biol. 1999; 61:437–467.

54. Grüner W, Giegerich R, Strothmann D, Reidys C, Weber J, Hofacker I L, Stadler P F, Schuster P. Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structure of neutral networks and shape space covering. Mh.Chem. 1996; 127:375–389.

55. Schuster P. Landscapes and molecular evolution. Physica D 1997; 107: 351–365.

56. Reidys C, Stadler P F, Schuster P. Generic properties of combinatory maps - Neutral networks of RNA secondary structures. Bull. Math. Biol. 1997; 59:339–397.

57. Wright S. Evolution in mendelian populations. Genetics 1931; 16:97–159.

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

59. Huynen M A. Exploring phenotype space through neutral evolution. J. Mol. Evol. 1996; 43:165–169.

60. van Nimwegen E, Crutchfield J P. Metastable evolutionary dynamics: Crossing fitness barriers or escaping via neutral paths? Bull. Math. Biol. 2000; 62:799–848.

61. Schultes E A, Bartel D P. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. Science 2000; 289:448–452.

62. Reidys C, Stadler P F. Bio-molecular shapes and algebraic structures. Computers Chem. 1996; 20:85–94.

63. Kimura M. Evolutionary rate at the molecular level. Nature 1968; 217: 624–626.

64. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge University Press Cambridge, UK 1983;.

65. Derrida B, Peliti L. Evolution in a flat fitness landscape. Bull. Math. Biol. 1991; 53:355–382.

66. Gavrilets S. Evolution and speciation on holey adaptive landscapes. Trends in Ecology and Evolution 1997; 12:307–312.

67. Newman M E J, Engelhardt R. Effects of selective neutrality on the evolution of molecular species. Proc. Roy. Soc. (London) B 1998; 265: 1333–1338.

68. van Nimwegen E, Crutchfield J P, Huynen M. Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. USA 1999; 96:9716–9720.

69. Fontana W, Schuster P. Shaping space: The possible and the attainable in RNA genotype-phenotype mapping. J. Theor. Biol. 1998; 194:491–515.

70. Cupal J, Kopp S, Stadler P F. RNA shape space topology. Artificial Life 2000; 6:3–23.

71. Perelson A S, Oster G F. Theoretical studies on clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. J. Theor. Biol. 1979; 81:645–670.

72. Gaal S A. Point Set Topology. Academic Press New York 1964;.

73. Schuster P, Fontana W. Chance and necessity in evolution: Lessons from RNA. Physica D 1999; 133:427–452.

74. Eldredge N, Gould S J. Punctuated equilibria: an alternative to phyletic gradualism. In Schopf T J M, editor, *Models in Paleobiology* pages 82–115. Freeman, Cooper & Co. San Francisco, CA 1972;.

75. Wagner G P, Booth G, Bagheri-Chaichian H. A population genetic theory of canalization. Evolution 1997; 51:329–347.

76. Leopold P E, Montal M, Onuchic J N. Protein folding funnels - a kinetic approach to the sequence structure relationship. Proc. Natl. Acad. Sci. USA 1992; 89:8721–8725.

77. Dill K A, Chan H S. From Levinthal to pathways to funnels. Nature Structural Biology 1997; 4:10–19.

78. Gilbert W. The RNA world. Nature 1986; 319:618.

79. Vendruscolo M, Maritan A, Banavar J R. Stability threshold as a selection principle for protein design. Phys. Rev. Lett 1997; 78:3967–3970.

80. Bussemaker H J, Thirumalai D, Bhattacharjee J K. Thermodynamic stability of folded proteins against mutations. Phys. Rev. Lett. 1997; 79: 3530–3533.

81. Bornberg-Bauer E, Chan H S. Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. Proc. Natl. Acad. Sci. USA 1999; 96:10689–10694.

82. Buchler N E G, Goldstein R A. Surveying determinants of protein structure designability across different energy models and amino-acid alphabets: A consensus. J. Chem. Phys. 2000; 112:2533–2547.

83. von Dassow G, Meir E, Munro E, Odell G M. The segment polarity network is a robust developmental module. Nature 2000; 406:188–192.

84. Wagner G P. What is the promise of developmental evolution? Part II: A causal explanation of evolutionary innovations may be impossible. Journal of Experimental Zoology (Mol Devol Evol) 2001; 291:305–309.

85. Spiegelman S. An approach to experimental analysis of precellular evolution. Quart. Rev. Biophys. 1971; 4:213–253.

86. Ellington A D, Szostak J W. In vitro selection of RNA molecules that bind specific ligands. Nature 1990; 346:818–822.

87. Bartel D P, Szostak J W. Isolation of new ribozymes from a large pool of random sequences. Science 1993; 261:1411–1418.

88. Wright M C, Joyce G F. Continuous in vitro evolution of catalytic function. Science 1997; 276:614–617.