Shaping Space: The Possible and the Attainable in RNA Genotype-Phenotype Mapping

Walter Fontana and Peter Schuster

Institut für Theoretische Chemie, Universität Wien, A-1090 Wien, Austria,

Intenational Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria,

and Santa Fe Institute, Santa Fe, NM 87501, USA

* Mailing Address: Professor Peter Schuster
Institut für Theoretische Chemie der Universität Wien
Wähingerstraße 17, A-1090 Wien, Austria
Phone: ++43 1 40480 669
Fax: ++43 1 40480 660
E-Mail: pks@tbi.univie.ac.at

Abstract

Understanding which phenotypes are accessible from which genotypes is fundamental for understanding the evolutionary process. This notion of accessibility can be used to define a relation of nearness among phenotypes, independently of their similarity. Because of neutrality, phenotypes denote equivalence classes of genotypes. The definition of neighborhood relations among phenotypes relies, therefore, on the statistics of neighborhood relations among equivalence classes of genotypes in genotype space. The folding of RNA sequences (genotypes) into secondary structures (phenotypes) is an ideal case to implement these concepts. We study the extent to which the folding of RNA sequences induces a "statistical topology" on the set of minimum free energy secondary structures. The resulting nearness relation suggests a notion of "continuous" structure transformation. We can, then, rationalize major transitions in evolutionary trajectories at the level of RNA structures by identifying those transformations which are irreducibly discontinuous. This is shown by means of computer simulations. The statistical topology organizing the set of RNA shapes explains why neutral drift in sequence space plays a key role in evolutionary optimization.

Keywords: evolutionary trajectories, neutral evolution, neutral networks, optimization, RNA secondary structures, statistical topology

1. Introduction

Molecular genetics views the course of evolution as a lineage of genotypes, while paleontology sees a lineage of phenotypes as manifested in the fossil record. The problem is to understand how the two are related. There is widespread agreement that the temporal succession of phenotypes reflects the selective boundary conditions operating during the evolutionary process. In this context the notion of "fitness" proved useful to reason about the conditions under which a given mutant can invade a population. However, some would contend that an evolutionary history also reflects the variational constraints which are intrinsic to an evolving entity. The term "variational constraints" is used to collectively denote causes which channel evolution in fitness-independent ways. "Fitness" is a notion which emphasizes the fate of a genotype mediated by the reproductive success of its phenotype in a given demographic and environmental context, while variational constraints point at the fact that not all possible phenotypes are equally accessible (or accessible at all) through variation of a given genotype. When focussing on the variational process, the objective of understanding successions of evolutionary innovations becomes one of explaining how each innovation affects the potential for further evolution [1]. This requires reasoning about the "evolutionary potential" of an evolving entity. The primary theoretical difficulty derives from finding adequate representations of phenotypes, and a model of how they are generated from genotypes.

In the present paper we pursue these issues in the context of a very special phenotype and genotype-to-phenotype map. We show how a rather obvious concept of "phenotypic nearness" induces a meaningful topology on the set of possible phenotypes. This topology enables us to understand some key features of evolutionary trajectories as observed, for example, in computer experiments. Extensions to *in vitro* evolution appear feasible and straightforward.

Seen from a wider perspective, our contribution is limited in several regards. First, we are concerned with the simplest relevant (and the only currently available) genotype-phenotype mapping we know of: RNA folding. The situation is extreme in that we are dealing with a single molecule which can play both parts in the game, being simultaneously genotype and phenotype. An RNA molecule is a sequence that can be replicated. By folding back on itself it forms a shape which is the target of selection [19]. Furthermore, evolution is here reduced to the simplest case: independent (asexual) replication in a constant environment. Accordingly, we shall not be concerned with networks of interacting molecules, but rather focus on the problem of how individual RNA shapes evolve within a population under selection for a specific target shape. Understanding the "evolutionary potential" of an evolving entity may certainly be more interesting and daunting for complex functional organizations, such as cells or multicellular organisms. Nevertheless, we believe that the RNA case is fundamental and at least of heuristic interest in thinking about the bigger picture, because it combines conceptual simplicity with realism and experimental accessibility. Second, our present study is mostly a numerical investigation of the properties of RNA folding as captured by present day computational techniques and empirical parameters. The level of molecular shape we are concerned with is known as the secondary structure of RNA. It represents a biologically meaningful and widely used notion of structure which can be predicted from sequences by fast algorithms. Our work is, therefore, intermediate in abstraction: it approximates an empirical situation, while aiming at generally valid regularities which may serve as axioms for more abstract mathematical models (see, for example, [15]).

2. Generic properties of folding

The term "folding" is used here to denote a surjection $f : \mathcal{S} \mapsto \Sigma$ from the set $\mathcal{S}$ of all sequences of fixed length over the **AUGC**-alphabet onto the set Σ of all minimum free energy (mfe) secondary structures for that length. The map is established implicitly by a state-of-the-art folding algorithm based on thermodynamic data [8, 11, 21]. In past work we have found three generic properties of the RNA folding map [18]. They are surveyed in this paragraph, since they are of immediate interest in the context of the current study. By "generic properties" we mean statistically robust features obtained by folding large samples of sequences.

The folding map is many-to-one, reflecting sequence redundancy of structures. **Property 1** ("frequent structures") is a statement about the size of equivalence classes of sequences with respect to structure: Some structures occur significantly more frequently than others. The property of being "frequent" is made more precise by the observation that in the limit of long chains the fraction of such structures tends to zero (their number grows nevertheless exponentially), while the fraction of sequences folding into them tends to one [7][1]. The remaining properties hold only for frequent structures.

Property 2 ("neutral networks") is a statement about the connectivity within an equivalence class: Two sequences in a class are connected by paths of sequences within the same class which differ by one or two point mutations [18]. The structure-neutral sequences, therefore, form connected and extended networks in sequence space.

Property 3 ("shape space covering") is a statement about the mutual entanglement of networks belonging to different structures in the high dimensional sequence space: All frequent structures are realized within a small neighborhood of any arbitrarily chosen sequence (for example, 15 mutations are sufficient for a chain of length $l = 100$, see [16]).

It is important to realize that the contemplated features are largely independent of the predictive accuracy of the algorithm for individual sequences. That is, even if the algorithm fails to reproduce details of the actual secondary structure of a particular sequence (as determined, say, by phylogenetic comparisons), it still captures the logic of constrained base-pair optimization characteristic of RNA folding. There is indeed evidence that properties 2 and 3 hold in nature [2,

[1]A particular definition of "frequent" is given by the notion of "common" [17]: A structure is common, when it its formed by more sequences than the average, $\#(\text{all sequences}) \big/ \#(\text{all structures})$.

14]. Our claims are further substantiated by observing that the properties we call "generic" have been found to be insensitive to the choice of the criteria of structure formation (minimizing free energy or maximizing base pairing or kinetic folding), as long as one structure is assigned to every sequence. They are also numerically robust to variations in the set of empirical energy parameters or the thermodynamic level of description (one mfe structure versus the Boltzmann ensemble for a given sequence)[2] [20].

These statistical properties, in particular neutrality, have led to a mathematical model based on percolation in random graphs [15]. In recent work we began to link the properties of the folding map with features observed in evolutionary dynamics. Model RNA populations in a flow reactor were subject to selection for a prespecified target structure. In particular, we found diffusion of the population on a neutral network, and recovered the fixation rates expected from Kimura's theory [12], thus establishing a microscopic model for neutral evolution [10].

3. Secondary structure

In what follows it will be very useful to shift back and forth between two levels of resolution for secondary structures, fine grained and coarse grained structures. The fine grained level corresponds to the conventional definition of secondary structure, as the set P of paired positions which minimize free energy subject to the condition that if (i, j) and (k, l) are both in P, then $i < k < j$ implies $i < l < j$. This condition means that no pseudoknots are being considered. For example, a line-oriented representation of a "Y"-shaped secondary structure of length 29 would look like "`((((.(((...))).(((...))).))))`". Matching parentheses indicate positions which are paired with one another.

The coarse grained level is defined by discarding information about the size of stacks and loops. The only information retained about a structural element is

[2] Similar properties have been recently discovered in lattice models of protein folding [5, 6, 13].

its type: stack (symbolized by a pair of matching parentheses `()`), hairpin loop (`H`), internal loop (`I`), left (`B`) or right (`b`) bulge, and multiloop with more than two branches (`M`). Thus, the above "Y"-structure would read as "`((H)(H)M)`". Renderings which better convey the topological nature of secondary structure at both levels of graining can be seen in figure 1.

4. A relation of accessibility

A natural metric for sequences is given by the Hamming distance indicating the number of positions $d(a, b)$ at which two sequences a and b differ. This metric is natural for situations conserving chain length where point mutations are the exclusive source of variation. The Hamming metric then mirrors the physical processes interconverting sequences. In natural populations point mutations are indeed more frequent than insertions and deletions.

It is, however, far less clear what constitutes a natural metric for structures. Common practice defines distance measures for structures directly on some suitable representation of structure. For example, the root mean square deviation between two sets of three-dimensional coordinates, or the edit distance between tree representations of RNA secondary structures, or overlaps between contact maps of protein structures. However, from an evolutionary point of view any definition based on a syntactic notion of (dis)similarity is bound to be artificial, since there are no physical processes which directly and inheritably modify structures at this level of representation. To modify a structure evolutionarily, requires modifying its underlying sequence. It is at once clear, then, that a structure β which is highly dissimilar from a structure α on syntactic grounds might nonetheless be "near" to α on the count of being accessible from α by a small mutation in α's sequence. Alternatively, among two syntactically highly similar structures, one might nonetheless fail to be evolutionarily "accessible" from the other. Notice that such a relation of accessibility does not quantify distance, but expresses a weaker notion of neighborhood. Pursuing this line we are led to a topology rather than a metric on the set of phenotypes.

More specifically, we consider the set $\mathcal{S}_\alpha$ of sequences which fold into a given structure α and define its boundary $\mathcal{B}_\alpha \subset \mathcal{S}$ to consist of all sequences at Hamming distance 1 from any sequence in $\mathcal{S}_\alpha$. Similarly, we call the set of sequences at distance d from $\mathcal{S}_\alpha$ its d-boundary, and let "boundary" stand as a shorthand for 1-boundary. We next fold all sequences in $\mathcal{B}_\alpha$ to obtain their set $\Sigma_\alpha \subset \Sigma$ of mfe structures. We refer to Σ_α as the set of 1-accessible structures of α. The d-accessible structures are defined similarly. We will, however, focus only on 1-accessibility, or accessibility for short, since it turns out to be sufficient for the interpretation of computer experiments at low mutation rates.

In sum, we shall say that β is accessible from α, or $\beta \leftarrow \alpha$, if there exists a pair $a, b \in \mathcal{S}$ with $d(a, b) = 1$ and $f(a) = \alpha$ and $f(b) = \beta$. In this notation the set of structures accessible from α is written as $\Sigma_\alpha = \{\beta \mid \beta \leftarrow \alpha\}$. We defer a definition of "nearness" to a later section.

Recall that the generic property 2 above states that we can think of $\mathcal{S}_\alpha$ as an extended network of neighboring points in sequence space having equal structure α. This view is quite useful at times, but unless otherwise noted we continue thinking in terms of the **set** $\mathcal{S}_\alpha$. The only difference is that in the latter case the shape α is included in Σ_α ($\mathcal{S}_\alpha \cap \mathcal{B}_\alpha \neq \emptyset$, but $\{\text{neutral network}\} \cap \mathcal{B}_\alpha = \emptyset$), and the existence of neutral neighbors is expressed by the reflexivity of the accessibility relation.

5. Boundary statistics

No resources are available to completely identify the set of structure-neutral sequences, $\mathcal{S}_\alpha$, not even for moderate chain lengths, let alone to exhaustively fold its boundary. We must, therefore, resort to sampling $\mathcal{S}_\alpha$. We start by fixing a secondary structure α of length l, and generate by "inverse folding" [8] a sample of n sequences which have α as their mfe configuration. For each sequence in the sample we fold all its $3l$ neighbors, obtaining the structures of $3ln$ sequences in the boundary of $\mathcal{S}_\alpha$. These structures constitute a sample of Σ_α.

Our interest is not just in the accessible structures, but also in how often they occur. Each structure $\beta \leftarrow \alpha$ has two multiplicities associated with it. One multiplicity, $N(\beta, \alpha)$, counts the total number of sequence-neighborhoods of α in which structure β occurs at least once. We normalize it by the size N_α of $\mathcal{S}_\alpha$, and call it the neighborhood frequency: $\nu(\beta, \alpha) = N(\beta, \alpha)/N_\alpha$. It reflects the likelihood of finding structure β in the one-mutation neighborhood of a randomly chosen sequence of $\mathcal{S}_\alpha$. The other multiplicity refers to the total number of occurrences, $N_t(\beta, \alpha)$, of structure β in $\mathcal{B}_\alpha$. Each neighborhood of a sequence in $\mathcal{S}_\alpha$ is, therefore, weighted with the actual instances of β in that neighborhood. We normalize it by $3lN_\alpha$, and call it the occurrence frequency: $\vartheta(\beta, \alpha) = N_t(\beta, \alpha)/3lN_\alpha$. $\nu(\beta, \alpha)$ and $\vartheta(\beta, \alpha)$ are estimated by sampling Σ_α as mentioned above.

6. Statistical topology

In discussing the topological structure of a set, we may start with a notion of neighborhood for each member of that set. Often the set one starts with is already a metric space, and the neighborhoods of a point x are defined by means of ϵ-balls consisting of all points at a distance less than ϵ from x. As remarked previously, we lack an evolutionarily relevant metric on the set of shapes, and there need not exist one. Instead, we define sets playing a role similar to ϵ-balls as a neighborhood base. Let $\alpha \in \Sigma$ and consider the sets

$$\Psi_\epsilon(\alpha) = \{\beta \in \Sigma_\alpha \mid \rho(\beta, \alpha) \geq \epsilon\} \ ,$$

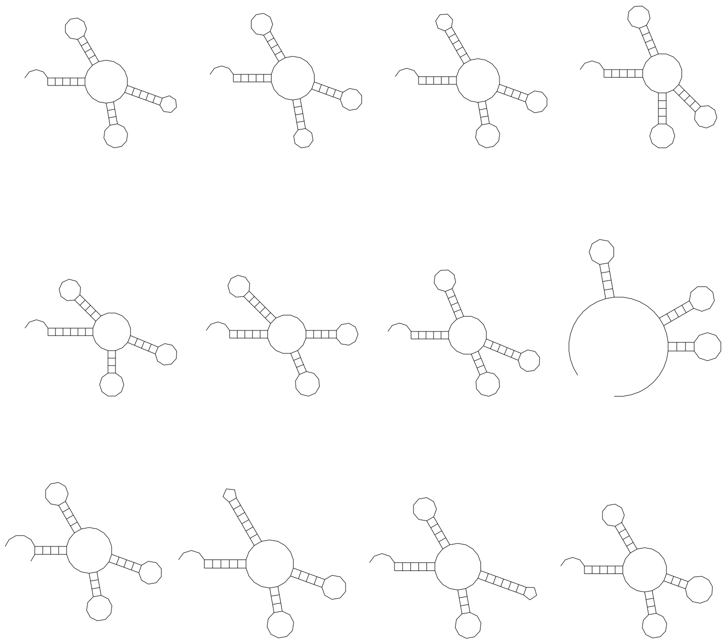
where $0 < \epsilon \leq 1$ and $0 \leq \rho(\beta, \alpha) \leq 1$ denotes a measure for the frequency of β in the boundary of $\mathcal{S}_\alpha$, such as $\nu(\beta, \alpha)$ or $\vartheta(\beta, \alpha)$ defined above. Clearly, $\epsilon_1 > \epsilon_2$ implies $\Psi_{\epsilon_1}(\alpha) \subset \Psi_{\epsilon_2}(\alpha)$. We note that $\rho(\beta, \alpha)$ is not a metric; in general $\rho(\beta, \alpha)$ is neither symmetric, nor does the triangle inequality hold. Technically speaking, a neighborhood of α is any set $\Psi_\epsilon(\cdot)$ containing α.

The actual topology of Σ depends on exactly what shapes those $\Psi_\epsilon(\alpha)$ contain for any $\alpha \in \Sigma$. The contents of these sets are not arbitrary, but rather a property of RNA folding (as captured by the folding algorithm), and must be obtained from a numerical investigation of $\rho(\beta, \alpha)$. A rigorous topology is invariably spoiled by the complexities of folding, and, hence, what matters here are statistical patterns. Moreover, given the combinatorial vastness of the space of possible sequences, we can only proceed by example. This means in particular, that our findings can be reasonably expected to hold only for the set of frequent shapes referred to previously.

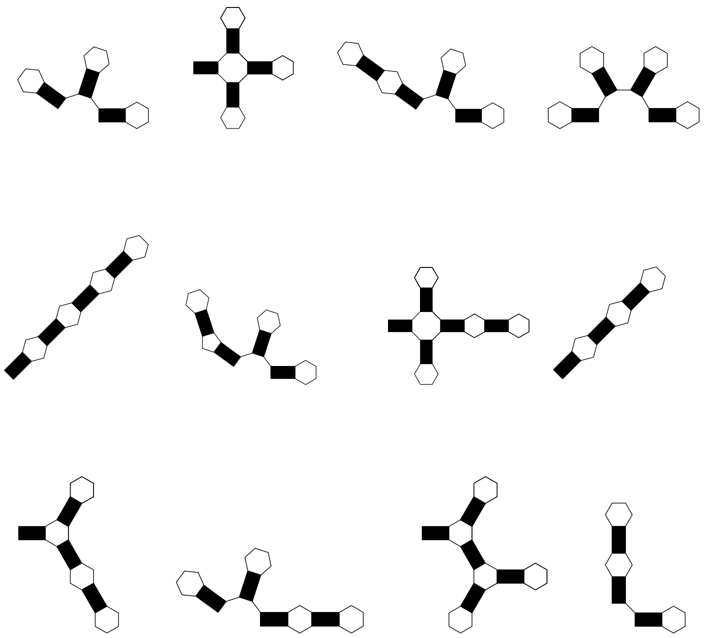
Neighborhoods in shape space can also be thought of as “correlation sets”, reflecting conditional structure correlations between pairs of nearest neighbors in sequence space, one of them folding into a fixed structure α. We call the topology based on such correlation sets a **statistical topology**. In the remainder of this paper we describe the contents of the $\Psi_\epsilon(\alpha)$, we then give a simple interpretation of the emerging topology, and make the connection with evolutionary trajectories obtained from computer simulations.

7. The set of all boundary shapes

As an example we start with the statistical profile of the shapes realized in the boundary of a tRNA clover-leaf structure, and consider first the contents of the biggest set in the neighborhood basis - the set of **all** shapes realized in the boundary of $\mathcal{S}_{\mathrm{tRNA}}$, $\Sigma_{\mathrm{tRNA}} = \{\beta \in \Sigma \mid \rho(\beta, \mathrm{tRNA}) > 0\}$ ($= \Psi_\epsilon(\mathrm{tRNA})$, where ϵ is the smallest frequency greater than zero, a lower bound being simply $1/4^l$). It is difficult to say something precise about the contents of this set, since our sample never catches all the shapes which occur just once in the boundary of $\mathcal{S}_{\mathrm{tRNA}}$. Some useful information, however, can be obtained by looking at the considerably smaller universe of coarse grained secondary structures defined previously. The question we ask is whether there is anything tRNA-specific to Σ_{tRNA} at the coarse grained level.



A



B

Figure 1: Shapes in the tRNA neighborhood. Part **A** shows the 12 highest ranked shapes with regard to neighborhood frequency (left to right, top to bottom). The series is the same with regard to the frequency of occurrence, except that #8 ranks first (and all others are shifted by one rank).
Part **B** shows the set of 12 topmost coarse grained structures ranked according to frequency of occurrence. All but the last shape are also found (in somewhat different ranking) within the top 15 with regard to neighborhood frequency. The last shape ranks 25th. All structures of **A** are represented by the first two coarse grained structures of this list. (The sample comprised 5,051 sequences with tRNA fine grained structure. Of the 1,151,628 one-error mutants a fraction of 0.39 were neutral with respect to the coarse grained tRNA structure, the remaining 703,476 sequences realized 5,881 different coarse grained structures.)

Our sample (described in figure 1) yielded 5,882 distinct coarse grained shapes. We denote the coarse grained Σ_{tRNA} with $\Sigma^{\mathrm{c}}_{\mathrm{tRNA}}$, and probe specificity by intersecting $\Sigma^{\mathrm{c}}_{\mathrm{tRNA}}$ with a pool of coarse grained random structures. The pool consisted of 1,578 unique shapes based on the coarse grained folds of 11,000 random sequences of length $l = 76$. 90.4 % of the shapes in the random pool were found in $\Sigma^{\mathrm{c}}_{\mathrm{tRNA}}$. The same procedure was applied to four other structures α of the same length $l = 76$. The sample size of their $\Sigma^{\mathrm{c}}_{\alpha}$ sets was about half the size of the tRNA sample. Their overlaps with the random pool were 82.4 %, 78.5 %, 73.3 %, and 81.7 %[3].

A further test set of 169 coarse grained structures was obtained from 10,000 random sequences of length $l = 45$. It's overlap with the $\Sigma^{\mathrm{c}}_{\alpha}$-sample of a randomly chosen structure α of that length was 87.6 %. Similarly, a pool of 32 shapes, generated by 20,000 random sequences of length $l = 29$, was intersected with the $\Sigma^{\mathrm{c}}_{\mathrm{Y}}$-sample (74 shapes) of the "Y" shape, "`((((.(((...))).(((...))).))))`". The overlap was 100 %.

We draw the conclusion that the set of coarse grained shapes realized in the boundary of a random structure contains the overwhelming majority of all coarse

[3] The corresponding structures were:
"`.........((((........)))).(((((.......))))).....(((((.......)))))...........`",
"`...((((.((((((....))))))..)).))....(((((...)))))......((.((((......))))))...`",
"`....((((................))))((((((((.(((.........))).)))..((((...)))).))))).`",
"`.......(((((((....))))))).(((((.((((((.....))))))....(((......)))....))))).`".

grained shapes realized by sequences of fixed length. While feasible sample sizes remain insufficient to collect true low frequency shapes, it is nonetheless tempting to conjecture that the boundary of a shape does contain **all frequent** coarse grained shapes realized by a given chain length.

Let us recall the generic property 3, shape space covering, which states that for any two frequent structures α and β, the distance between two sequences folding into α and β will not be larger than a certain value which is small compared to the diameter of sequence space. Our conjecture here corresponds to the obviously much stronger claim that this value is Hamming-distance one. Let us refer to it as the **adjacency conjecture**. At the fine grained level of secondary structure the adjacency conjecture is unlikely to hold. In fact, for the binary **GC**-only sequence space of length $l = 25$ the claim is false. Exhaustive folding of that space reveals, however, that there is a substantial fraction of shapes which are common to the shape boundaries Σ_α of the most frequent shapes α. This shared subset comprises between 30 % and 70 % of the individual Σ_α. It is likely to be even larger for **AUGC** sequences. In contrast, the intersection of the Σ_α of several rare structures turns out to be empty. It turns out that for coarse grained secondary structures the picture is different: In the case of **GC**-only sequences of length 25 the adjacency conjecture was found to hold.

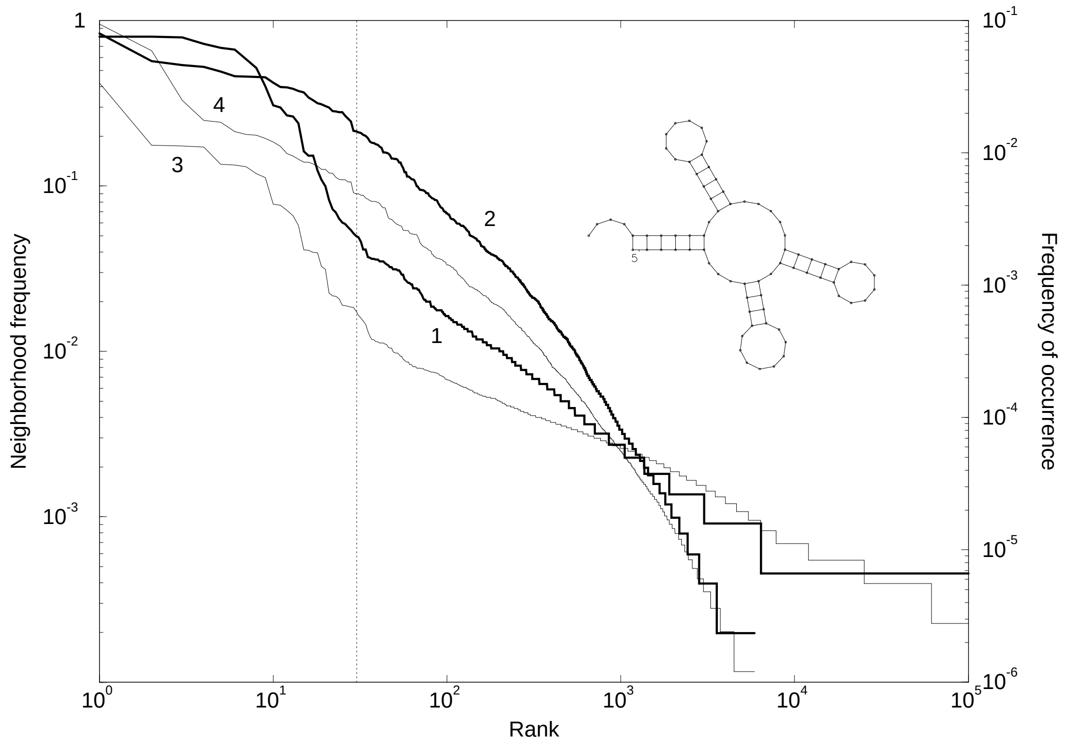
As a consequence of the shape space covering property, an evolutionary process has to explore only a restriced (yet still substantial) l-dimensional volume of sequence space, in order to find an arbitrary frequent shape. However, when advantageous mutants are not immediatley accessible to a given population of sequences, mutation dynamics and selection pressure confine the population to drift on a (much) lower dimensional slice of sequence space given by the neutral network of the temporarily fittest shape [10]. Huynen [9] pointed out that while drifting, a population keeps exploring the network's one-error boundary. The point added by the adjacency conjecture is that selective confinement to a neutral network does, in principle, **not** preempt the process from eventually reaching **any** of the frequent coarse grained shapes.

8. Not all shapes are equally important

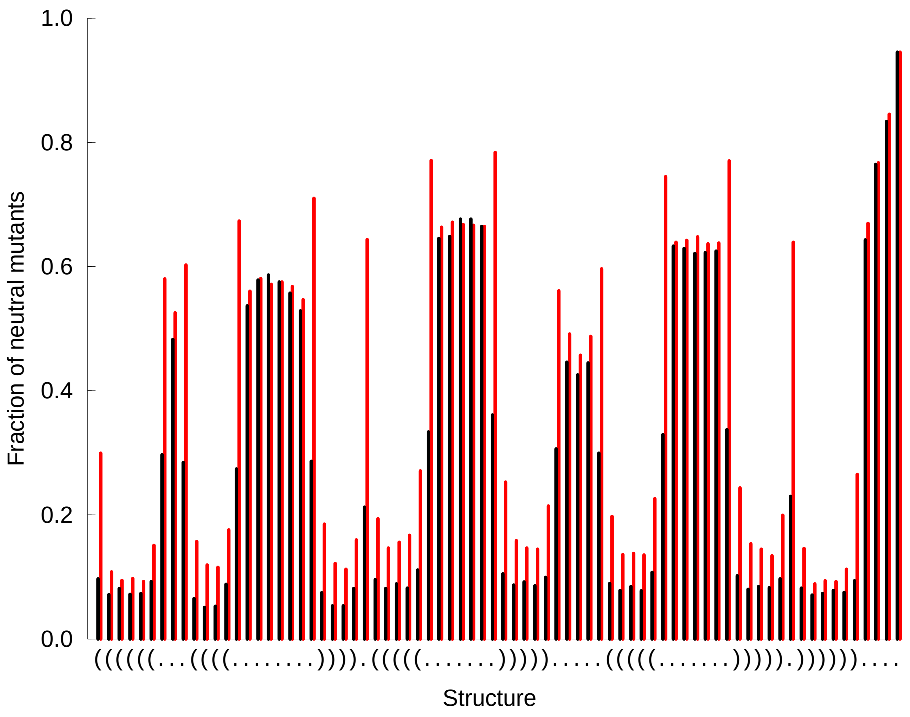
The contents of the other Ψ_ϵ(tRNA) depend on the frequency of accessible shapes. Figure 2 shows a log-log plot of the rank ordered distributions of the neighborhood frequency, $\nu(\beta, \alpha)$ (curve 1), and the occurrence frequency, $\vartheta(\beta, \alpha)$ (curve 3), for fine grained structures. The main feature is the existence of two regimes with distinct exponents. The first regime includes a small set of about 30 shapes which are considerably more frequent (and thin out substantially faster with increasing rank) than the dominant tail regime which covers several orders of magnitude. This signals the existence of structures which are characteristic neighbors of the tRNA, in the sense of occurring with high probability among the one-error mutants of only those random sequences which fold into the tRNA structure.

The most frequent structure is the reference structure α, and percolating neutrality is expressed by the fact that $\nu(\alpha, \alpha) = 1$. This being the case for any frequent shape α (generic property 2), we will omit the reference structure α from its fine grained distribution data and the rankings. Figure 1A shows the 12 most frequent fine grained structures ranked according to $\nu(\beta, \alpha)$. Figure 1B summarizes the top neighboring structures at the coarse grained level. The coarse grained "neutral" is included here, because it represents a variety of distinct fine grained variants. However, in the corresponding frequency distributions, figure 2A curves 3 and 4, the contribution due to fine grained, or "true", neutrality has been subtracted. The two topmost coarse grained structures in figure 1B represent most of the high frequency fine grained neighborhood. Other frequent coarse grained structures include variants lacking the multiloop, or having stems interrupted by (small) internal loops.

With one exception, all high ranking fine grained structures are very similar to the reference. They arise from shortening or lengthening a stacking region by one base pair with the concomitant lengthening or shortening of the affected loop region. These structures all share the same coarse grained shape. In reference structures with long stacks, the blocking of a base pair often produces small "bub-


Neighborhood frequency
Frequency of occurrence
Rank
1
2
3
4
5

A


Fraction of neutral mutants
Structure
((((((...((((........)))).(((((.......))))).....(((((.......))))).))))))....

B

Figure 2: Shapes in the tRNA boundary. 2199 sequences folding into the tRNA clover-leaf reference structure (length l=76, inset of plot **A**) were sampled. All their one-error mutants ($2199 \cdot 76 \cdot 3 = 501{,}372$ sequences) were folded. A fraction of 0.28 (142,847 sequences) had the same structure as the reference. The remaining 358,525 sequences folded into 141,907 distinct shapes. Curve 1 is a log-log plot of their rank ordered neighborhood frequency (thick line, left ordinate) and curve 3 shows their rank ordered frequency of occurrence (thin line, right ordinate). The dotted vertical line is meant to separate regions with different scaling. In a second sample (of 5,051 sequences with reference structure) we collected the statistics pertaining to coarse grained shapes. The corresponding data sets 2 and 4 are analogous to curves 1 and 3, respectively. The fraction of fine grained neutrals was subtracted from the coarse grained neutral.
Plot **B** shows the neutral mutation frequency as a function of the position mutated (sample of 6,597 sequences per position) for the fine grained case (black impulses) and for the coarse grained case (sample of 15,153 sequences per position; grey impulses). The abscissa shows the reference structure with pairs of matching parentheses representing base pairs at the corresponding positions along the sequence. Note the lower, but non-zero, level of neutrality in the paired regions due to **GU** pairing.

bles" (internal loops or bulges), while conserving the overall hairpin architecture. In figure 2B the high frequency neighbors of the tRNA shape can be immediately identified by comparing the fraction of neutral mutations per position of the fine grained (black) and the coarse grained (grey) case. Large differences between both levels flag exactly those positions whose modification yields fine grained variants indistinguishable from the coarse grained reference structure.

The notable exception to the series of slight variants is the three-hairpin structure ranked #8 in terms of neighborhood frequency (0.52). It even ranks first in terms of frequency of occurrence (0.03; all others shown in figure 1A shift down in rank by one). This shape deserves further attention and we shall give it the name: tRNA_8.

Consider now the Ψ_ϵ of tRNA_8 (see figure 3 for shapes and figure 4, curve 1, for the frequency distribution). Observations analogous to those made for the tRNA case apply here as well. The main point, however, concerns the relation between the two structures: the tRNA structure was never sampled in the boundary of tRNA_8. Every second sequence in $\mathcal{S}_{\text{tRNA}}$ is susceptible to the destruction of the multiloop closing stem upon a single point mutation in that region. In strong contrast, extremely few sequences in $\mathcal{S}_{\mathrm{tRNA}_8}$ meet the constraints for the creation of any closing stem from an open region in one mutation. Evidently, the relation between phenotypes induced by a frequency weigthed genotypic accessibility is not symmetric.

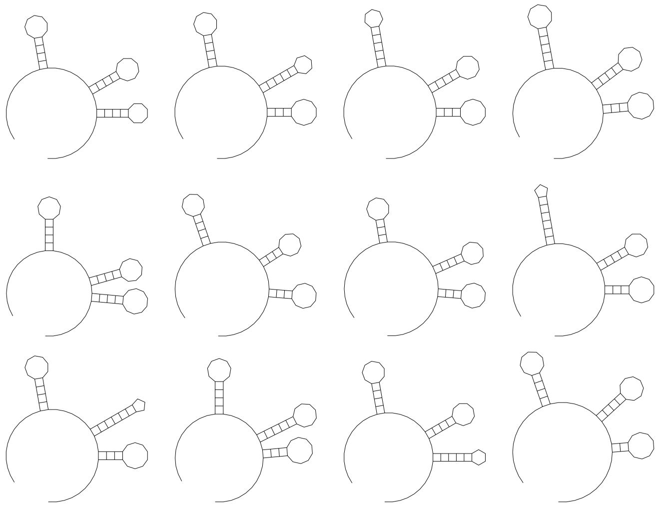
This statistical asymmetry has a counterpart in sequence space at the level of the corresponding neutral networks. The neutral network of the tRNA_8 structure is substantially larger than the one of the tRNA. The fact that the tRNA_8 network is persistently found one step away from the tRNA network, suggests a kind of "embedding": seen from the smaller network, the larger one appears almost everywhere in its boundary, while from the viewpoint of the larger one, the smaller appears almost nowhere.

Similar asymmetric relations hold between the tRNA_8 structure and some of its topmost boundary shapes. For example, with high frequency any of the remaining stacks of tRNA_8 can disappear in a single point mutation (see the shape ranked #4 in the coarse grained neighborhood, figure 3). Two-stack structures of this kind are even found among the high frequency ranks of the tRNA boundary. This is due to the existence of sequences whose tRNA structure contains a hairpin stem which is stabilized only in the context of the multiloop. If a mutation destroys the multiloop, the multiloop sensitive stack opens as well.

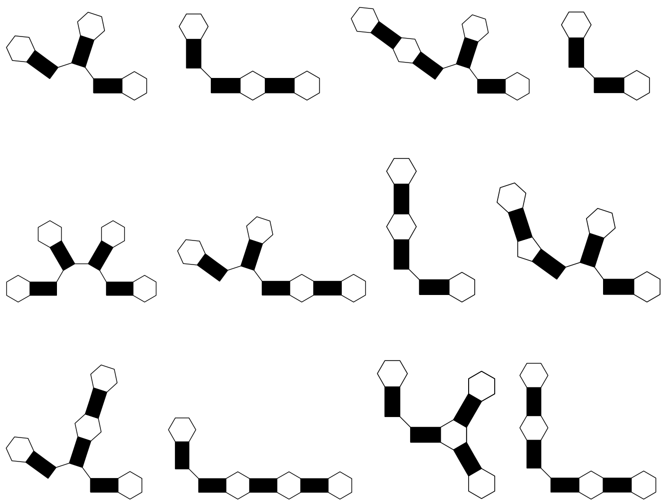
At the coarse grained level, the intersection of $\Sigma^{c}_{\mathrm{tRNA}}$ with $\Sigma^{c}_{\mathrm{tRNA}_8}$ (3,344 coarse grained structures in the sample) shows an overlap of 81.1 %, as expected from the previous intersection results with random pools. All top 30 coarse grained structures realized in the boundary of $\mathcal{S}_{\mathrm{tRNA}_8}$ occur among the top 91 realized in the boundary of $\mathcal{S}_{\mathrm{tRNA}}$.

9. The choice of frequency

Consider a structure which occurs once among the one-error mutants of each of 10,000 sequences in the sample, and one that occurs 100 times around 100 sequences. Both are indistinguishable in terms of the occurrence frequency, $\vartheta(\beta, \alpha)$,



A



B

Figure 3: Shapes in the $tRNA_8$ neighborhood. The arrangement of shapes is as in figure 1. The highest ranked shape not shown (#13) in terms of neighborhood frequency lacks the $5'$ hairpin stem. The sample consisted of 2,200 sequences folding into the $tRNA_8$ structure. Of the 501,600 one-error mutants a fraction of 0.36 were neutral with respect to the fine grained structure. The remaining sequences in the boundary sample realized 130,668 distinct fine grained shapes and 3,344 distinct coarse grained shapes.

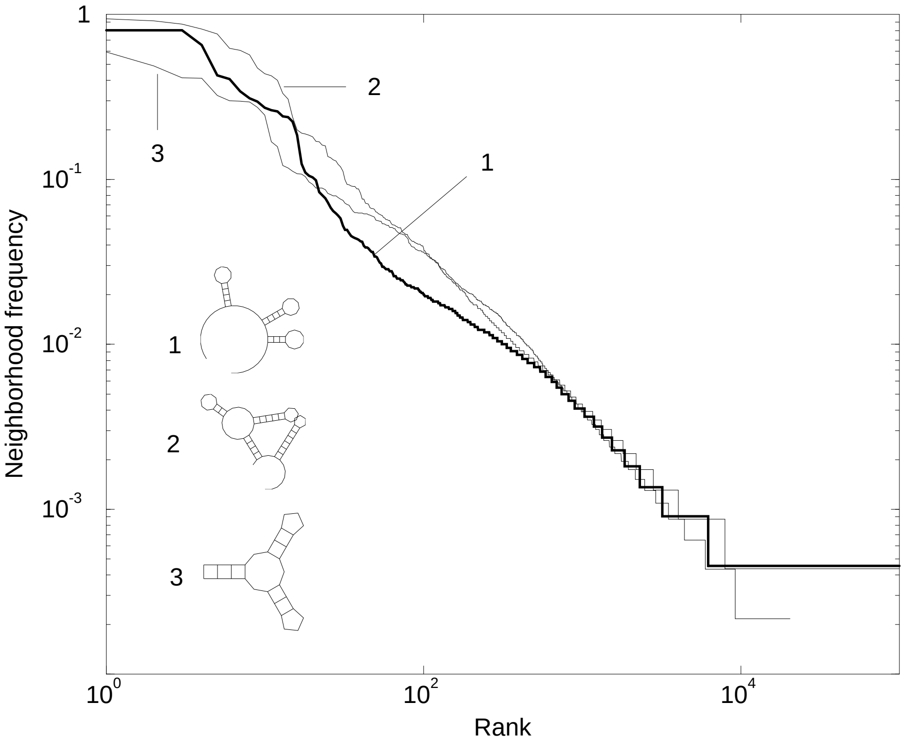


Figure 4: RNA boundary distributions for non-tRNA structures. Log-log plots of rank ordered neighborhood frequency distributions of other RNA shapes. Curve 1 (thick line) belongs to tRNA_8, curve 2 belongs to a randomly chosen structure of the same length (l=76), and curve 3 to a Y-shaped structure of length l=29.

but the former ranks much higher with respect to the neigborhood frequency, $\nu(\beta, \alpha)$. Whether the two frequencies emphasize different notions of neighborhood, depends on the actual distribution of occurrences of β per sequence neighborhood of α.

Figure 5 shows three typical scenarios for the tRNA case. The neutral structure has a wide distribution ranging from 25 to as much as 109 occurrences per one-error neighborhood. This is in contrast to most accessible variants, an example of which is shown in the inset of figure 5. Their occurrences range typically between 1 and less than 10 per sequence neighborhood, which doesn't make a big difference between $\nu(\beta, \alpha)$ and $\vartheta(\beta, \alpha)$. The exception is again tRNA_8, whose distribution is bimodal. This bimodality is found with all accessible structures lacking the multiloop closing stem, and indicates that there are many sequences in which an extended shape feature – here the multiloop closing stem – is marginally stable. For example, the tRNA multiloop closing stem comprises 12 nucleotides, yielding 36 possible one-error mutants each of them blocking the affected base pair with probability 4/6 or 5/6, depending on whether or not it is a **GU** pair. Assuming the limiting case that the stack is so marginally stable that removal of any one pair destroys the entire stem, one expects 24 to 30 instances per sequence neighborhood. The maximum for tRNA_8 in figure 5 is 31. In general, one expects the frequency with which an extended shape feature is marginally realized to match roughly the frequency with which it is stably realized (giving rise to only few occurrences per neighborhood), since strong and weak stacking interactions between Watson-Crick pairs are balanced. Non-Watson-Crick **GU** pairing affects this balance only slightly, because excessive **GU** pairing destabilizes a stack to such an extent that it doesn't form at all.

The majority of frequently realized modifications are limited to local shape features, such as individual base pairs. These can necessarily be realized only a few times per neighborhood, as there are only few positions for a mutation to affect the feature in the first place. The corresponding occurrence distributions are similar to each other, and resemble the one shown in the inset of figure 5.

While the shape space neighborhood structures induced by $\nu(\beta,\alpha)$ and $\vartheta(\beta,\alpha)$ appear quite similar, we prefer $\nu(\beta,\alpha)$ as it treats large and small shape features on an equal footing.

10. Nearness

The form of the rank-ordered boundary shape distributions (figures 2 and 4) indicates an α-(shape)-dependent value δ separating the characteristic set of high frequency structures from a low frequency background shared to a large extent with any other reference shape. This suggests to describe the topological structure of

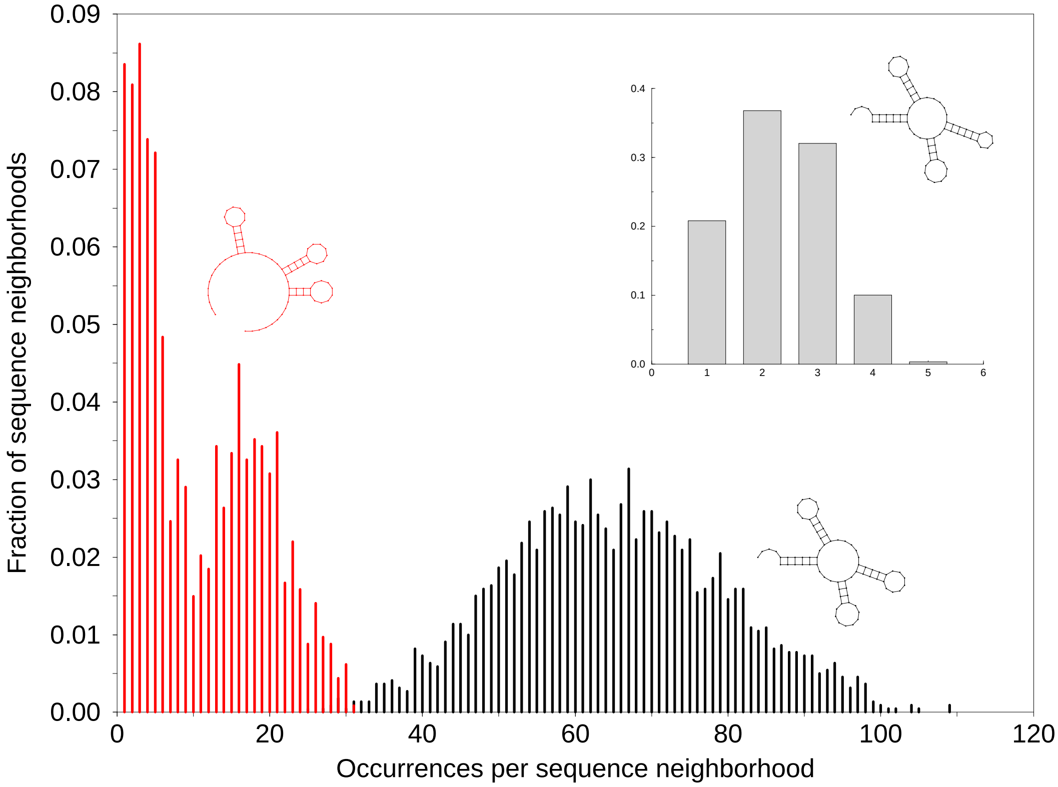


Figure 5: Occurrence distribution in sequence neighborhhoods. The plot shows for three tRNA boundary shapes how often each of them occurs in how many sequence neighborhoods (Each neighborhood allows for a maximum of 76·3=228 occurences). The right side (dark) and the left side (light) of the plot refer to the neutral shape (tRNA) and the tRNA_8, respectively. (The two plots barely overlap). For tRNA_8, the area under the low incidence region roughly equals the area under the high incidence region (0.54 up to the minimum at 10, separating both regions), indicating that the multiloop closing stem is realized stably and marginally with approximately the same frequency. The inset shows the distribution for the most frequent boundary shape after the neutral one. See text for details.

RNA shape space by considering for each shape α only the shapes accessible from α with a frequency of at least δ. For moderate chain lengths, like those considered here, δ is not sharply defined. The purpose of the procedure, however, is to emphasize the major qualitative aspects of the shape space organization. In this vein, the neighborhood base becomes $\Psi^\delta_\epsilon(\alpha) = \{\beta \in \Sigma_\alpha \mid \rho(\beta, \alpha) \geq \epsilon \geq \delta\}$. We refer to the largest such set, $\Psi^\delta_\delta(\alpha)$, as the **characteristic set** of α. The set of all boundary shapes, Σ_α, is obtained by relaxing the threshold δ, $\Sigma_\alpha = \lim_{\delta\to 0}\Psi^\delta_\epsilon = \Psi^0_0$.

At both levels of resolution, coarse and fine graining, the highest ranking structures occur with almost the same frequency, as evidenced by the initial flat region of the distribution. To caricature the qualitative aspects, we may also discretize the variation of ϵ by a suitable step size ξ (ξ^{-1} integer), $\epsilon = 1 - n\xi, n = 1, 2, \ldots, \xi^{-1}$, considering $\Psi^\delta_{[n]}(\alpha) = \{\beta \in \Sigma_\alpha \mid \rho(\beta,\alpha) \geq 1 - n\xi \geq \delta\}$. The smallest set, $\Psi^\delta_{[1]}(\alpha)$, contains all shapes accessible from α with frequencies in the top range $[1-\xi, 1]$.

We finally proceed to define nearness. A shape β is defined to be **near** a set of shapes Γ, if every neighborhood $\Psi^\delta_{[n]}(\beta)$ contains a shape of Γ. By abuse of language we call a shape β near a shape α, if β is near the characteristic set of α. In this sense the tRNA$_8$ shape is near the tRNA, but not vice versa. However, for many shapes β in the top set $\Psi^\delta_{[1]}(\alpha)$ the nearness relation to α is symmetric. Consider, for example, a hairpin structure α with a single stack of length s. Almost every sequence folding into α will have among its 1-error mutants some in which the loop closing terminal base pair of the stack has been destroyed, yielding a shape β with a single stack of length $s-1$. Conversely, given a sequence which folds into β, it is easy to access α by rebuilding that base pair through a single mutation.

11. The substructure relation

It is instructive to compare this topology with a different relation. Secondary structures are partially ordered by the subset relation on the set of their base pairs P_α: $\alpha < \beta$ (read: α is a substructure of β), if $P_\alpha \subset P_\beta$. This relation can be visualized as a directed graph on Σ. In the present context we are interested only in the undirected version of that graph, that is, two structures are connected by an edge if they are comparable. Obviously, two structures are comparable, if they arise from one another by either removing or adding base pairs (subject to the no-pseudoknot condition). Conversely, two structures are incomparable, if their interconversion involves both removing **and** adding base pairs. In the latter case it is useful to distinguish whether the interconversion is a **generalized shift**. We define a change of base pairing to be a generalized shift, if for each base pair changed at least one base remains paired. This includes the standard shift, where paired strands slide past each other, typically by a few positions (figure 6). A generalized shift, however, also covers transformations such as the "roll-over", the "flip", and the "double flip" sketched in figure 6.

Structures that differ by generalized shifts are called **shift-incomparable**. For example, $\kappa \equiv$ "`(((.....)))..((((...))))`" and $\lambda \equiv$ "`((((...))))..(((.....)))`" are incomparable but not shift-incomparable, while κ and $\mu \equiv$ "`(((......))).((((...))))`" are shift-incomparable, as are κ and $\nu \equiv$ "`.(((.....))).((((...))))`".

Our previous observations indicate that if a structure α is near a structure β, then α and β are comparable. The converse is not quite true, however. Consider, for example, the pair $\alpha = \text{tRNA}_8$ and $\beta = \text{tRNA}$. Rather, if two structures are comparable and their symmetric difference $\Delta(\alpha, \beta) = (P_\alpha - P_\beta) \cup (P_\beta - P_\alpha)$ is small, such as one or two base pairs, then both α and β are near one another.

Notice that, by definition, if β is near α, we can pass from α to β directly, that is, in one step, without leaving the characteristic set of α. For the sake of simplicity, let us refer to the characteristic set of α as "the" neighborhood of α. If β is not near α, the one-step transition from α to β requires leaving the neighborhood of α (figure 7A). A transition from α to β need not be in one step, but may occur in several stages. If this prevents neighborhood boundaries from being crossed, we call the transition **continuous** (figure 7B). An example is given by the previously mentioned structures κ and λ. They are incomparable, and a direct transition from, say, κ to λ leaves the neighborhood of κ. However, the transition is reducible, since there is an intermediate structure $\kappa' =$ "`((((...)))).. ((((...))))`" which is comparable with κ and $\Delta(\kappa, \kappa')$ is small. From the numerical neighborhood profile of κ we can infer that the transition from κ to κ' does not leave the neighborhood

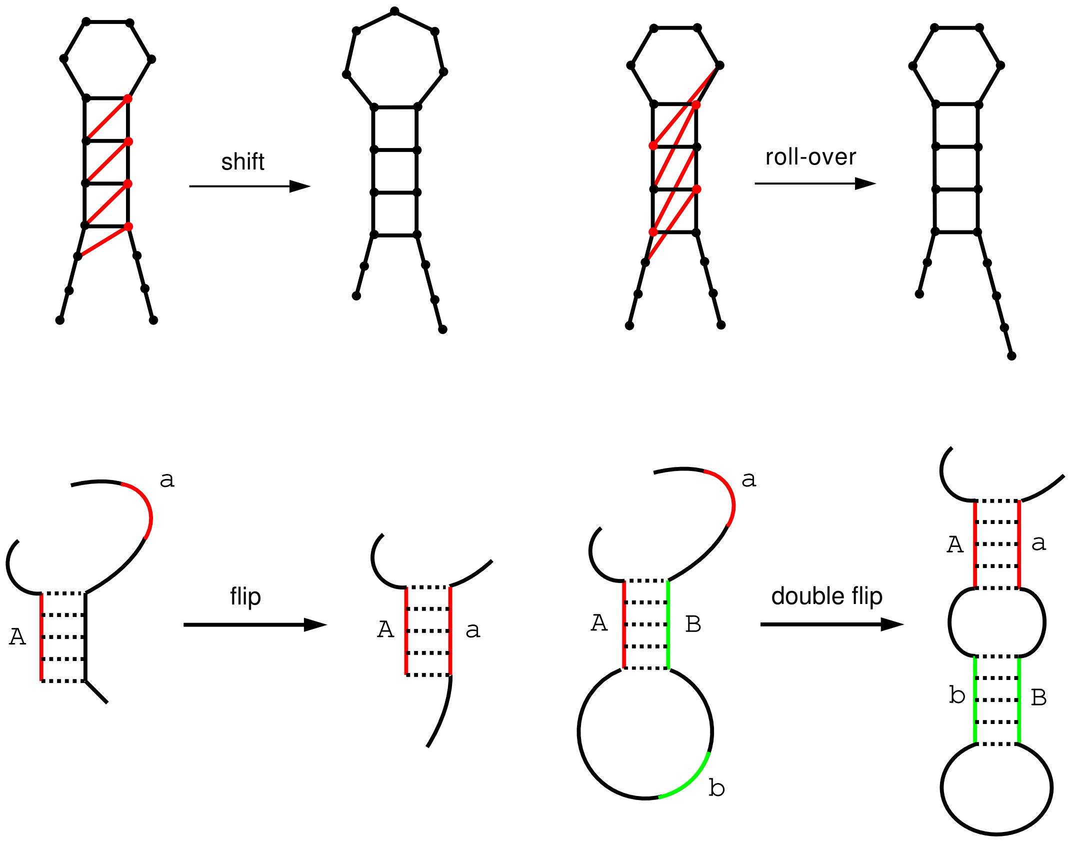


Figure 6: Generalized shifts. The figure shows the shift types bundled under the term generalized shift. In a standard shift (upper left) one strand of a stacked region slides past the other. (Light lines indicate the new pairing pattern.) The result is the growth of a loop. In a "roll-over" (upper right) both strands of a stack shift by the same amount; as a result the loop maintains its size and, in the example, "rolls" towards the $5'$ end. A "flip" denotes a big slide where the new position of the shifted strand does not overlap with the old position (lower left), while a "double flip" refers to the analogous situation in which both strands flip. In all cases, for each base pair involved, at least one position remains paired before and after the change.

of κ. Moreover, since κ' is similarly related to λ, the transition from κ' to λ doesn't leave the neighborhood of κ' either. Hence, there is a continuous multi-step transition from κ to λ. In contrast, there is no continuous transition from tRNA$_8$ to the tRNA. The table 1 summarizes the observed regularities.

In general, if β is not near α and there exists a series of structures

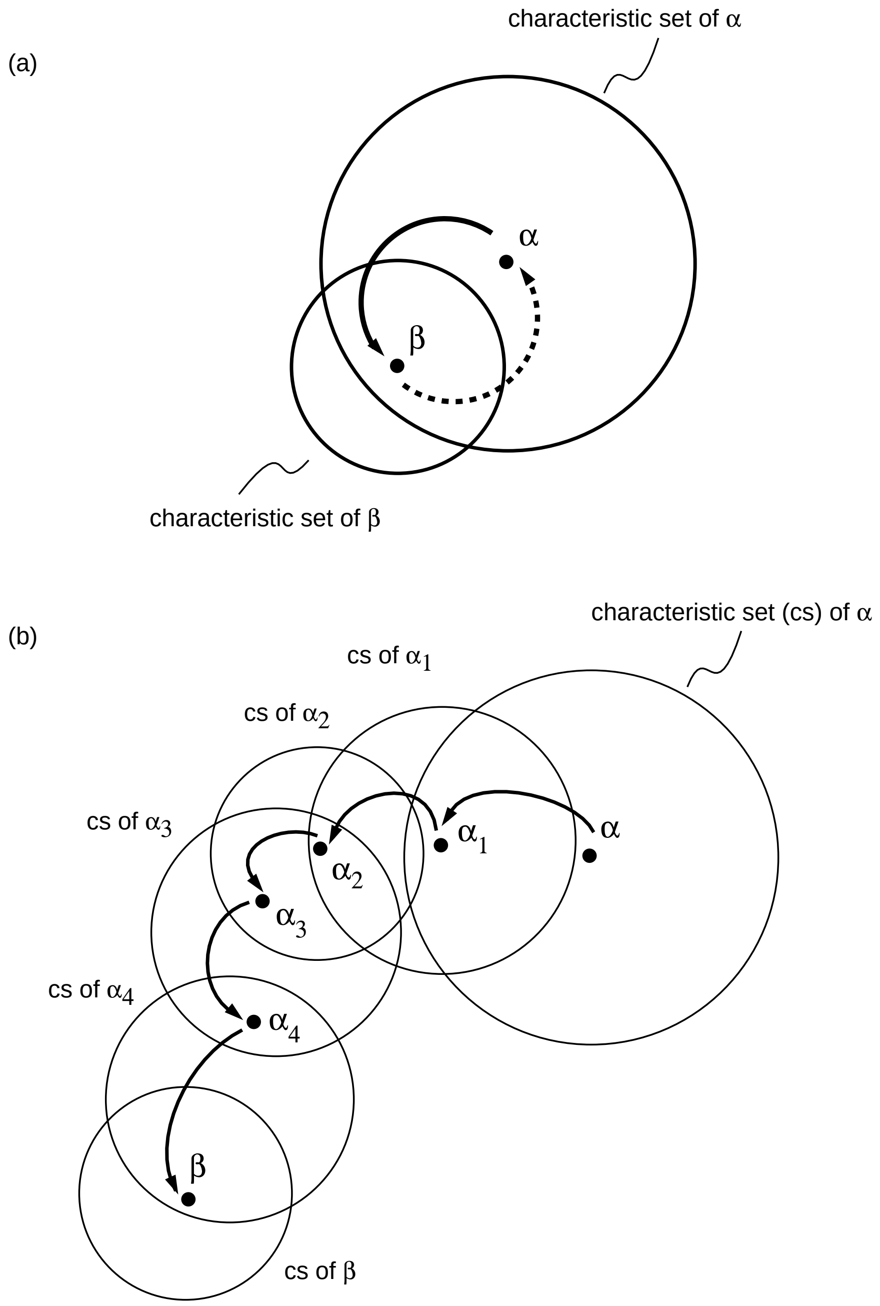


Figure 7: Continuity of transitions. The upper drawing shows a situation in which β is in the characteristic set of α, but not vice versa. By definition this simply means that accessing β from α is easy on average, while accessing α from β is not. The former transition is termed continuous, the latter discontinuous (dotted arrow). The lower drawing illustrates how a discontinuous one-step transition from α to β becomes continuous by means of a suitable series of intermediates accessible through continuous one-step transitions.

Table 1: Substructure relations and the nature of one-step transitions. Transitions from α to β are classified by "yes" and "no" depending on whether or not they remain in the neighborhood of α. Asterisks mark irreducibly discontinuous transitions (see text).

Substructure relation α, β			Transition
Case	Class	Δ	continuous
1	comparable	small (1bp)	yes
2	comparable	stack length, $\beta < \alpha$	yes
3	comparable	large, $\alpha < \beta$	no *
4	shift-incomparable	—	no *
5	incomparable	—	no

$\alpha \equiv \alpha_0 \ \alpha_1 \ \cdots \ \alpha_{i-1} \ \alpha_i \ \cdots \ \alpha_n \equiv \beta$ such that α_i is near α_{i-1}, the overall (multistep) transition from α to β is continuous. A transition is **irreducibly discontinuous**, if no such series exists. The irreducibly discontinuous transitions are realized by two scenarios: (1) α and β are comparable, but in passing from α to β a long stacking region, such as a multiloop closing stem, must be created from scratch (the symmetric difference is large and $\alpha < \beta$), or (2) α and β are shift-incomparable. Both cases have one feature in common: the transition cannot be done incrementally on thermodynamic grounds. Case (1) reflects the fact that a minimum stack size is needed to compensate for the destabilizing free energy contribution resulting from the loop created by the new stack. This nucleation size depends on the nature of the stacking pairs, the terminal mismatches, and the nature and the size of the loop. Case (2) reflects the fact that shifting a stretch of contiguous base pairs requires their synchronous displacement. The pairs cannot shift in random sequential order without violating obvious steric constraints (and, formally, the no-pseudoknot constraint). Moreover, if a shift were to happen in stages, unpaired bubbles arise which need stabilization by a minimum stack size on either side. Shifts of long stretches are indeed likely to happen in stages, but typically not shifts involving stacks of size 4. The irreducibly discontinuous transitions are, hence, determined by the thermodynamics of folding.

These observations are the key to understanding the evolutionary dynamics of simulated RNA populations to which we turn next.

12. Boundary conditions

The influence of the genotype-to-phenotype map on evolutionary dynamics is potentially blurred by the composition with a phenotype-to-fitness function. The latter introduces a further type of neutrality resulting from different shapes having the same fitness. As a consequence, phenotypically distinct neutral networks in genotype space are merged into one larger fitness-neutral network. While the phenotype-to-fitness map can be largely arbitrary and directly influenced in laboratory evolution experiments, we believe that the two cases studied here delimit a wide class of meaningful functions. We have in mind a fairly generic situation in which RNA sequences evolve towards a shape capable of binding to some chosen molecule, as is indeed the case in SELEX-type experiments producing small RNA binders called "aptamers" [3]. Since we are interested in evolutionary trajectories rather than searching for a shape with some predefined property, we specify in advance what the final shape ought to look like. Our mapping from shapes to replication rate constants (fitness), then, becomes a simple function of the syntactic distance of a given shape to a prespecified target shape (see below). Given a distance measure d, a shape replicates faster, the more it resembles the target. In all simulations reported here, the replication rate r_i of a sequence i of length l with shape α at distance $d(\alpha, \tau)$ from a target shape τ is given by $r_i = (0.01 + d(\alpha, \tau)/l)^{-1}$. Using an exponential or a linear function did not make any difference with regard to the issues we are interested in. We used a rate of 8 errors per 1000 nucleotides copied (For a tRNA length of $l = 76$ this corresponds to a per nucleotide error rate of $p = 0.001$). At this rate, the difference between parent and a modified offspring sequence is mostly one point mutation. These conditions are thus appropriate for a statistical topology based on 1-accessibility.

Two quite distinct distance measures on shapes were considered. One is given by the Hamming distance between two shapes in their parenthesized representation, and the other is given by the "base pair distance". Both treat corresponding sequence positions which differ in their pairing state as errors contributing to distance (unpaired versus paired and upstream-paired versus downstream-paired). The difference, however, derives from base pair distance treating a base **pair** as a unit, while Hamming distance treats a paired **position** as the unit. As a consequence, base pair distance will count as errors situations which do not contribute towards Hamming distance. For example, consider two shift-incomparable shapes, "`((((....))))`" and "`.((((...))))`". Their Hamming distance is 2 (only positions 1 and 5 differ), while their base pair distance is 9 (all of the paired positions differ).

Our simulation of an RNA population subject to selection in a constrained flow reactor according to stochastic chemical kinetics is a continuous time model of Spiegelman's classic serial transfer experiments [19]. It's implementation is described elsewhere [4, 10].

13. Evolution in phenotype space: the relay series

In order to study the influence of the statistical topology on evolutionary dynamics, we focus on the temporal succession of shapes rather than individual sequences. Stated in terms of sequences this means that we focus on the succession of equivalence classes of sequences represented by a neutral network corresponding to a shape. If we were to track over time which shape gives rise to which shape, we would end up with a vast and highly interconnected network of phylogenies (Circular paths at the level of individual sequences are a common result of mutational backflow in the population and this backflow must be substantially larger, if we consider entire equivalence classes of sequences with respect to shapes). Besides the computational complexities associated with handling hundreds of megabytes of history data pouring from a simulation of this kind, we would still need to know which pathways are the relevant ones.

Our solution to this involves two steps. First, we consider the phylogenetic network of only those shapes that literally made history: we only record information about events which generate a "relative innovation", that is, a shape which is new in the population at the time t of its appearance. This does not neccesarily imply that the shape hasn't been in the population in the past; it could have gone extinct some time and it might be "rediscovered" at time t. This provision removes the large backflow among extant shapes within a population, which doesn't proceed into new phenotypic territory. For each relative innovation α we record entry times, l_i^α, and exit times h_i^α.

In a second step, after the evolutionary process has found the target or has been stopped, we trace back through the history data in the following way. Consider the set of "live intervals" $\mathcal{L}_\alpha = \{[l_i^\alpha, h_i^\alpha],\ l_i^\alpha < h_i^\alpha < l_{i+1}^\alpha\}$, marking the presence of shape α in the flow reactor during the simulation. Each live interval $[l_i^\alpha, h_i^\alpha]$ of α has a unique ancestor with shape β which spawned that interval at time l_i^α, meaning that a sequence folding into β produced at time l_i^α a mutant which folded into α, and α was not in the population at that time. Let ω be the target shape, and $[l_i^\omega, h_i^\omega]$ one of its live intervals whose ancestor is ω_{-1}. Among the $\mathcal{L}_{\omega_{-1}}$ there is a unique live interval $[l_j^{\omega_{-1}}, h_j^{\omega_{-1}}]$ containing the time instant l_i^ω, and we proceed searching for the unique ancestor of $[l_j^{\omega_{-1}}, h_j^{\omega_{-1}}]$. Upon repeating this procedure we eventually end up at one of the initial shapes. At this point we have reconstructed a chain of shapes $\alpha \equiv \omega_{-n}\ \omega_{-n+1}\ \cdots\ \omega_{-i}\ \cdots\ \omega_{-1}\ \omega_0 \equiv \beta$ connecting an initially present shape α with the target (or final) shape β. This chain is uninterrupted in time, in the sense that for every $n \geq i \geq 1$, ω_{-i} is ancestor of ω_{-i+1} and there exists a pair $[l_r^{\omega_{-i}}, h_r^{\omega_{-i}}]\ [l_s^{\omega_{-i+1}}, h_s^{\omega_{-i+1}}]$ with $l_r^{\omega_{-i}} < l_s^{\omega_{-i+1}} < h_r^{\omega_{-i}}$. The chain depends on the live interval of the final shape β from where the trace starts, but it is unique for that interval. On rare occasions there may be more than one such interval for the target shape, since stochastic fluctuations may wipe out the target which must be generated again to get established in the population. By default we consider the last live interval of the target shape at the time the simulation has been stopped.

Because of the chain's connectedness in time and its uniqueness relative to a live interval of the final shape, we think of it as the causal chain of phenotypic innovations leading from α to β. We term it the **relay series**, in analogy to a relay-race in which a team races to goal in stages, each runner covering a segment of the trail and handing over the baton to the successor. Of course, our relay series can only be known in retrospect.

In computer experiments it is easy to record a unique identifier for every mutation event. When reconstructing the relay series, we obtain the identifier of each event that caused the passage from one relay shape to the next. Equipped with this hindsight, we rerun the same history (by using the same random seeds), but this time selectively recording the actual succession of sequences underlying the relay series (Recording such information *a priori* would flood most computer systems, because we have no foresight into the sequence of stochastic events and thus the relay series is accessible only by backtracking of trajectories).

Notice that the relay series is not defined by appealing to concentration or fitness and, hence, the relay shapes need neither coincide with the succession of dominant nor fittest shapes in the population. Moreover, the definition does not prevent the relay series from containing cycles. Finally, the uniqueness of the relay series (per target live interval) refers only to a given computer experiment. Different simulations may proceed through different phenotypic paths leading from α to β. In the next section we study the relationship between the relay series and the shape space topology.

14. Patterns of history

We monitor a macroscopic and a microscopic aspect of the evolutionary process. The macroscopic one is given by the time evolution of the average distance (average fitness) of the population to the target shape. The microscopic one is a description of the temporal succession of phenotypes, as given by the relay series.

We discuss evolution towards a tRNA shape as an exemplar representing a variety of simulations carried out with different target shapes.

14.1. Hamming distance

In the following simulations the shape distance function is Hamming distance. Figure 8 juxtaposes the micro and macro aspects mentioned. The relay series shown in figure 9 consists of 42 shapes, indexed from 41 (start shape) through 0 (target shape). Figure 8 also shows the live intervals of all relay shapes separated vertically by index. The step trace indicates the time segment of the relay history occupied by each relay shape. The left boundary of each segment coincides (by definition) with the beginning of a live interval generated by the previous relay shape. The right boundary marks the time instant at which the corresponding relay shape has generated its successor in the relay series.

The shape space topology induced by the folding map has little influence on the early phases of evolution. This results from the fact that whatever change a random shape undergoes, it is likely to narrow the gap to most targets. Moreover, major changes are likely to occur when shapes contain long unpaired random regions. Chains with long unpaired stretches, or even the open chain, are not frequent structures for the lengths considered here, and our topology fails. Once the opportunities deriving from this initial "latent" structural variability have been exploited for a quick gain in fitness, the stage is set for the shape space topology to shape the long course of the remaining history.

In order to show that the shape space neighborhoods underlie the relay series, we consider the sets of live intervals of two consecutive relay shapes, ω_{-i} and ω_{-i+1}, during a period in which the fitness average of the population remains constant. In such a period relay transitions occur typically among fitness neutral shapes, and most relay shapes are significantly populated during some of their live intervals. If ω_{-i+1} is near ω_{-i}, we should observe a series of live intervals belonging to ω_{-i+1}, indicating its intermittent presence starting with the arrival of ω_{-i}. This signals the fact that ω_{-i+1} is repeatedly generated from ω_{-i}, and,

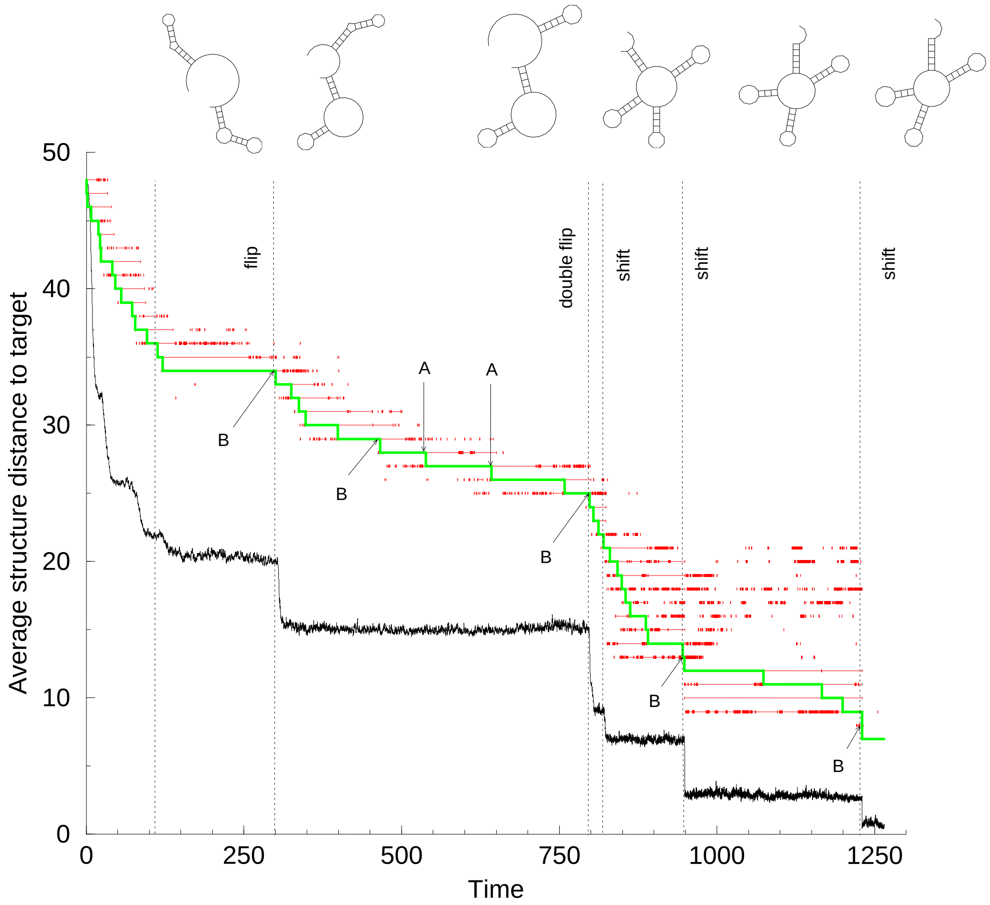


Figure 8: Major transitions towards a tRNA shape. The figure shows how much optimization has progressed at the macro level by plotting the average Hamming distance to the target structure. The fitness curve is superimposed by the relay trace showing the flow of causality from start shape to target (see text for definitions). The major transitions are marked by vertical lines. The corresponding generalized shifts are named, and the shapes before and after the transition are shown (Except for the first standard shift to avoid congestion of the figure). The flow reactor was stochastically constrained to maintain an average of 1,000 sequences and the error rate was 0.001 per nucleotide.

thus, "unavoidable". Conversely, if ω_{-i} is near ω_{-i+1}, we should observe a similar pattern of live intervals for ω_{-i}, once ω_{-i+1} has become the relay shape. Consider, for example, the patterns around the relay transitions marked A in figure 8. The relay segment of the incoming shape ω_{-i} (the ancestor) is followed by a series

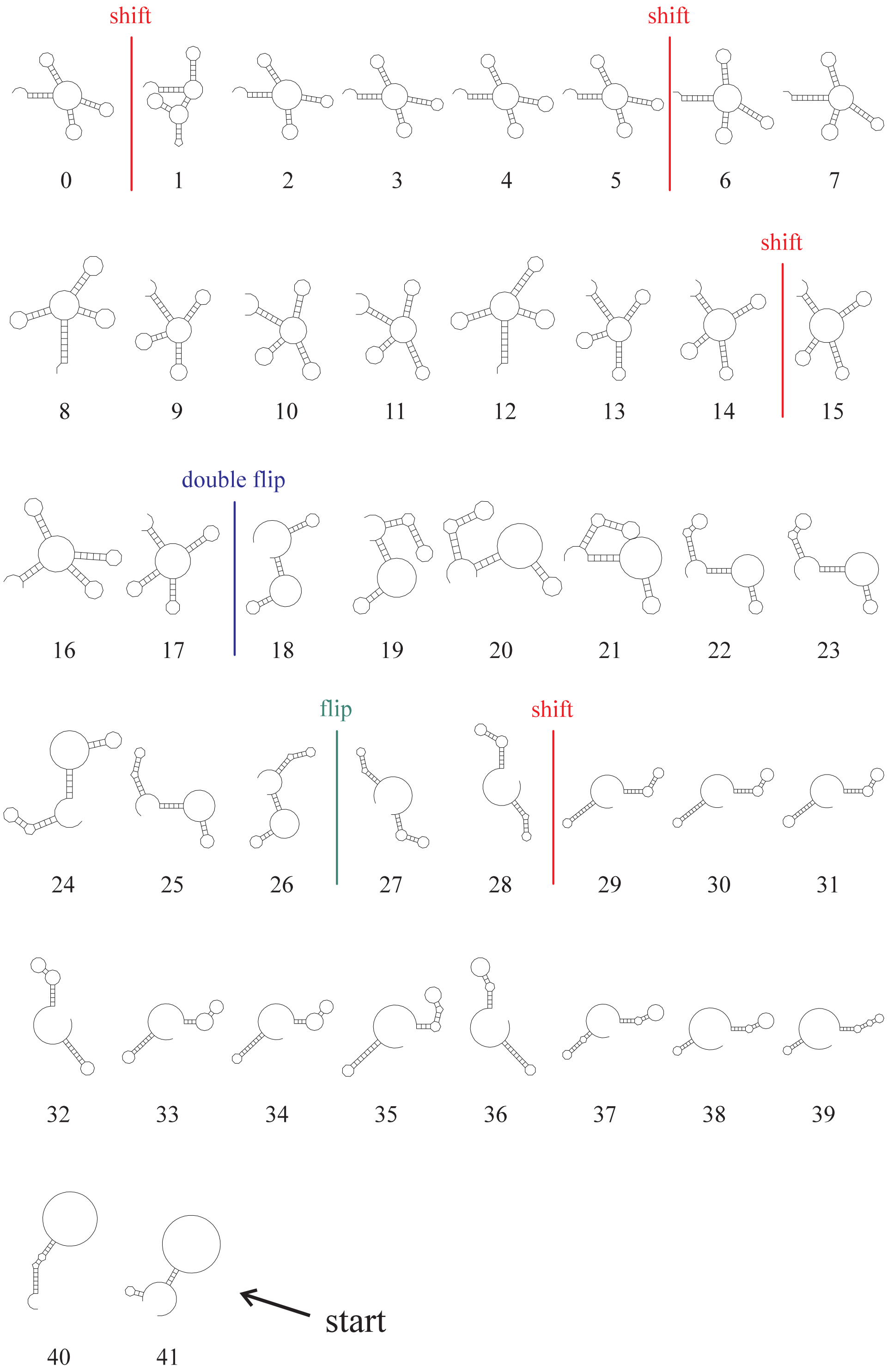


Figure 9: Relay series. The full series of relay shapes for the simulation in figure 8 is shown. See text for details.

of (non-relay) live intervals, and the relay segment of the outgoing shape ω_{-i+1} (the offspring) is preceded by a series of (non-relay) live intervals. This pattern indicates that ω_{-i} and ω_{-i+1} are both near each other. A pattern in which either one of the series of non-relay live intervals is absent, indicates that the shape associated with that series is near the one lacking it, but not vice versa. Thus, the patterns at the instants marked B in figure 8 signal that the transition from ω_{-i} to ω_{-i+1} leaves the characteristic set of ω_{-i}. Yet in all but the last cases ω_{-i} is near ω_{-i+1}, since the presence of ω_{-i+1} entails the presence of ω_{-i}.

During the extended periods of constant average fitness most relay shape passages are continuous, in the sense of proceeding within neighborhoods. On rare occasions a discontinuous relay transition – an escape from a neighborhood – occurs without having an impact on fitness (such as instant B on the plateau around time 450 in figure 8). The main observation, however, runs in the opposite direction: each change in fitness (vertical dotted lines in figure 8) is associated with the escape from a shape space neighborhood. Either the fitness change is caused directly by the escape, or it is shortly preceded by it. This observation holds for all target choices we made. In fact, that choice can only affect which neighborhood escapes have an impact on fitness.

Then, the basic pattern of a phenotypic path is the following. In the early phase of the process some random restructuring may take place, and the overlap between stacks present and the target is maximized by growing or shrinking stacks one or two base pairs at a time. According to the shape space topology outlined before these latter adjustments are continuous. This phase of the process necessarily stops when no continuous improvements are possible. At this point some stacks are positioned more or less correctly, while others consist of one strand overlapping with its target analogue, yet pairing to an entirely misplaced region. The fine tuning of stack positions without loss of fitness must proceed through shifts (for example, the first transition indicated in figure 8). Similarly, if one strand of a stack is correct, but pairs wrongly, a generalized shift of the “flip”-type (figure 6) must occur (second transition indicated in figure 8). The latter case is put to an extreme when the position and the pairing orientation (upstream or downstream) of both strands of a stack agree with the target, but the strands do not pair with each other in the target. To correct such a situation under strong selection, a double flip (figure 6) must occur (see figure 8 and 9). By means of flips and double flips the correct overall architecture of the shape is eventually achieved, which is then fine tuned by further standard shifts. At high replication accuracy, the vast majority of generalized shifts is triggered by a single point mutation.

Recall that the statistical topology structures the set of shapes only in the high accessibility regime up to the threshold value δ suggested by the neighborhood frequency distribution. Yet, the set of all boundary shapes Σ_α extends far beyond it. The escape from the neighborhood system of α, that is, from its characteristic set, is possible, because other shapes (presumably all frequent coarse grained shapes) are accessible from α with low probability. This includes shapes β which differ by a generalized shift from α. When selection confines a population to a dominant master shape α, escaping α's neighborhood can only occur through random drift on the neutral network of α. This random drift enables a specific sequence context to be set without loss of viability, such that a single point mutation can trigger the structural rearrangement to β. This causes long waiting times on average. A statistics of such waiting times and a study of how they depend (if at all) on the proximity to the target, remains to be done.

Notice that standard shift differences to the target are also a major cause of neutrality in this context. Consider, for example, a target structure (or substructure) $\beta \equiv$ "`(((....))).`" and a shift-incomparable shape $\alpha \equiv$ "`.(((...))).`" at Hamming distance 2. Elongating α's stack by a base pair, "`((((...))))`", constitutes a transition to a shape near α and leaves the Hamming distance to β unchanged. This is precisely what happens during neutral drift periods. It is easily seen that a situation with multiple stacks can lead to cycles in the relay series. There are two instances of this sort in figure 9: relay shapes #12 and #7 are identical, as well as shapes #5 and #3.

A few details at the sequence level are worth mentioning. The mutation which caused the double flip from relay shape #18 to relay shape #17 (figure 9) occurred at a site remote from the sites involved in the flipping. It extended the hairpin stack near the 3' end by one base pair, which provided sufficient stabilization for the double flip to occur. The necessary sequence context for this event arose during the preceding long phase of random drift on the fitness-neutral network of shapes #26 through #18.

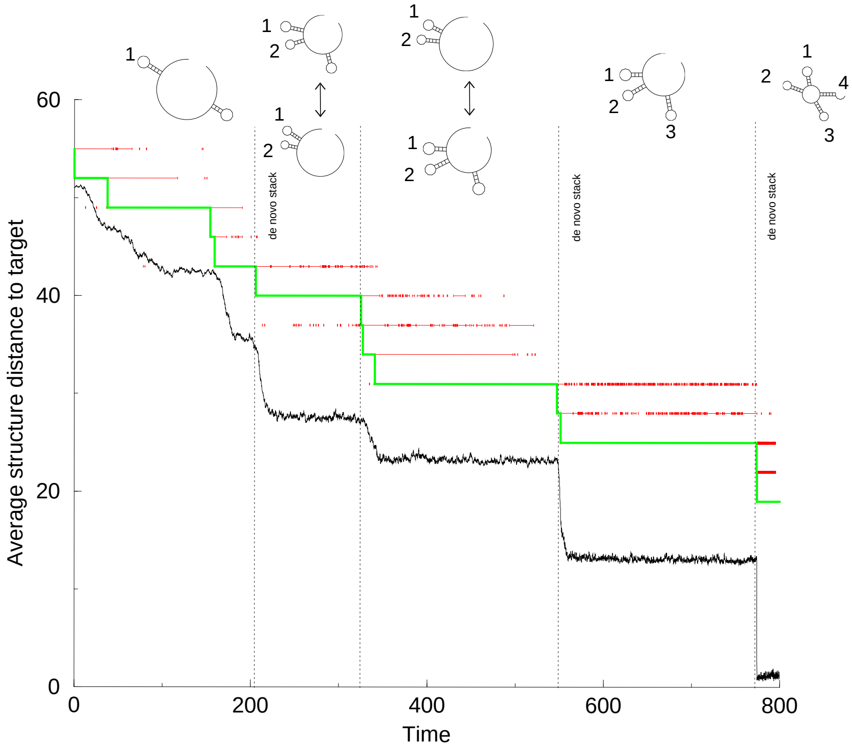
Furthermore, shape #1 in figure 9 seems a strange intermediate in the shift event leading from shape #2 to the target (#0). The shift actually happened in two stages (first #2 to #1, then #1 to target). If the two mutation events had occurred in the reverse order, a different intermediate shape would have made the shift process more obvious. With the specific sequences involved, the actual order of events forced a more dramatic constriction of the multiloop. A computer experiment whose initial population consisted of the population at time 1000 in figure 8, confirmed that shape #1 is not a necessary relay shape for the corresponding shift. There are alternative histories. That simulation (not shown) produced the target once after 73 time units. A fluctuation then wiped it from the population, and the target shape was produced again some 20 time units later, when it finally took over. Two relay series, differing slightly in their final stages, are associated with these two live intervals. In the first relay series the target was produced by a shift directly from what is shape #4 in figure 9. In the second relay series the target arose via #3 ($\equiv$ #5) by a silent roll-over and a standard shift.

14.2. Base pair distance

The dominance of generalized shifts in structural transitions is linked to the evaluation of shapes by means of a Hamming metric which considers each strand of a stack separately. The Hamming metric implies that selection pressure can hold in place one strand of a stacking region, while the other is free to shift or to flip. Indeed, starting from different random initial conditions, the route to the major structural tRNA feature – the multiloop – always involved a double flip. Generalized shift differences to the target are small in the Hamming metric. Yet, shift transitions are difficult to achieve, and the evolutionary process guided by Hamming distance "hangs" whenever such a transition is required. Direct formation of a multiloop closing stem (the other kind of irreducibly discontinuous transition) can be forced to occur, when starting with a homogeneous population consisting of the tRNA_8 shape. As expected, a long period of drift precedes the closing of the multiloop (not shown).

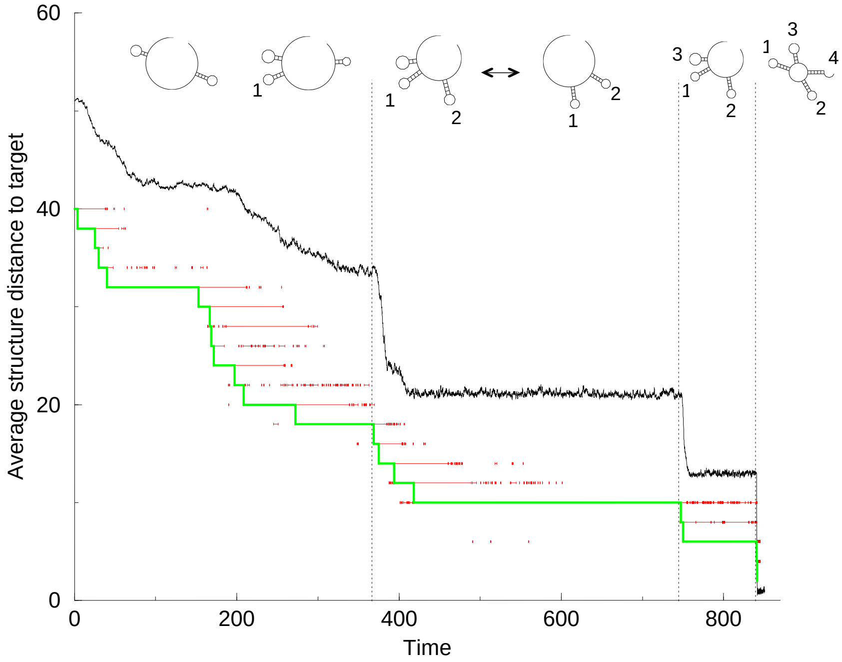
It is, thus, instructive to consider the impact of base pair distance on evolutionary trajectories. In terms of base pair distance shifts appear as large differences, and selection pressure is exerted on individual base pairs rather than individual positions, that is, the two strands of a helix cannot evolve independently. Figure 10 shows two runs of the optimization process for tRNA, both starting from identical initial conditions. The macroscopic picture exhibits the same phenomena as discussed in detail for the Hamming case. The main difference, however, is at the microlevel. Transitions are now predominantly *de novo* constructions of stacking regions rather than shifts. Whether a sequence segment folds into a stack shifted by one position relative to the target or whether that segment doesn't fold at all makes only a slight difference from the point of view of base pair distance. To effectively shift a stack, the stack is undone and subsequently generated from scratch in the shifted position. As indicated in figure 10, during some neutral drift periods both neutral shape versions, with and without a wrongly positioned stack, were present in high concentration.

The relationship between neighborhood transitions and the relay series is seen by means of the live interval trace in much the same way as in the Hamming case. There is one seeming exception in figure 10A at the second transition. That transition is not about the creation of a new stack, but rather about the shortening and lengthening by one base pair of the stacks labelled "1" and "2", respectively. However, according to the shape space topology described previously, these transitions are continuous and should not be preceded by a long period of drift. In fact, consider the live interval trace of the relay shape β succeeding the relay



Average structure distance to target
Time
de novo stack
de novo stack
de novo stack
0
20
40
60
0
200
400
600
800

A



Average structure distance to target
Time
0
20
40
60
0
200
400
600
800

B

Figure 10: Evolutionary trajectory under base pair distance. As in figure 8 the evolutionary approach towards a tRNA target shape is shown in terms of average structure distance and the relay series. The difference to figure 8 is that the similarity between structures is measured as a base pair distance. This distance function is virtually "blind" to shifts. As a result, the *de novo* creation of a stacking region – the second type of discontinuous transition in the shape space topology – becomes prominent. Selected relay shapes documenting the arrival of relevant structure motifs are shown. Correctly positioned stacks are labelled in the order of their appearance. Neutrality between coexisting dominant shapes is indicated by double arrows.

shape α present during the drift period. The trace shows that β was generated rapidly and repeatedly after α entered the population – as it should be, since β is obviously near α. The problem is that the base pair distance to the target is large at this point in the process, and given our specific fitness function, a distance improvement of 2 units is too weak to trigger a sharp selective response. Once β accumulated in the population, and generated its own mutants, it gave immediately rise to a (near) shape with a lengthened stack labelled "2". This increased the fitness gain to 4 units, triggering a fast selection response. Such a spurious drift period is absent from figure 10B.

The correctly positioned stacks of the relay shapes shown in figure 10 are numbered in the order of their appearance. The modularity of RNA secondary structures permits many permutations in the sequence of transition events leading to a target shape. However, in all runs perfomed, the multiloop closing stem was the last to form.

Hamming distance caricatures better than base pair distance the *in vivo* or *in vitro* evaluation of RNA shapes under selection constraints similar to those considered here. More sophisticated distance criteria, such as tree edit distance [18], produce a picture which is very similar to the Hamming case.

The main point of the comparison is to show that even dramatic changes in the phenotypic evaluation function only affect which of the two discontinuous transitions dominates, not the fact that they dominate. The high degree to which the RNA folding map constrains evolutionary trajectories in a recognizable way seems largely independent of the fitness map imposed on the phenotypes.

The shape space organization appears not nearly as obvious for binary sequences. **GC**-only sequences of a given shape cannot exhibit neutrality within paired regions. This is enough to make the escape from shape space neighborhoods through neutral drift extremely difficult. While **GC**-only sequences with a tRNA shape were easily found by inverse folding, extensive simulations failed entirely to evolve anything close to a tRNA shape in a **GC**-only setting. A similar situation might hold for **AU**-only sequences. However, our inability to find **AU**-only tRNA shapes by inverse folding suggests that they don't even exist.

15. Conclusions

In order to understand evolutionary histories, we need to organize the set of possible phenotypes in a way that reflects their attainability through genotypic changes. This induces a different organization than one based on moves performed on some direct representation of the phenotypes. The latter emphasizes a purely syntactic similarity of phenotypes, while the former connects with the genotypic level in a fashion mediated by development (here folding).

"Development" and "genetic changes" are treated here in their probably simplest non-trivial realization given by RNA folding and point mutations. Since our folding algorithm computes the shapes (phenotypes) from sequences (genotypes) by a mathematical procedure which does not necessarily reflect the actual process of folding, we deal with a static (yet statistically accurate) genotype-to-phenotype map.

The set of RNA minimum free energy secondary structures is organized as a topological space by means of a frequency weighted relation of accessibility. One-error mutants of sequences sharing a common minimum free energy structure preferentially fold into certain shapes. Accessibility, thus, captures nearest neighbor correlations at the level of neutral networks, which are equivalence classes of sequences with respect to shape.

The numerical analysis of the accessibility relation implied by a widely used folding algorithm yielded three basic findings. First, for any frequent shape α, there exists a set of accessible shapes which is characteristic for α from a frequency point of view. Second, a notion of nearness, based on this characteristic set, leads to the notion of a "continuous transformation" of structure. Informally, the transformation of a structure α into β is continuous, if it proceeds through a sequence of neighboring genotypes such that the structure of each offspring is near to the structure of its parent. The discontinuous transitions are precisely those structural changes that cannot be sequentialized in an incremental fashion, but rather require the synchronized change of several base pairs. Such transitions involve a generalized shift or the formation of a longer stacking region, such as a multiloop closing stem. Third, independently of fitness criteria, the RNA shape space topology strongly influences evolutionary trajectories approaching a target in all but the very early stages of the process.

Effecting a discontinuous transition in shape space by a small genetic change (here one point mutation) poses stringent conditions on candidate sequences. Optimization generally requires discontinuous transitions in shape space, and this entails long average waiting times during which selection pressure confines the population to drift along neutral networks. Transitions coincide with an escape from a shape space neighborhood made possible by the fact that the boundary of a neutral network contains beyond its characteristic set a very large number of very low frequency shapes. Our study suggests that this set of shapes virtually includes all frequent coarse grainedstructures. It is, however, precisely the neutral drift resulting from the hung process which enables the transition to actually occur. Neutral drift eventually leads to a suitable sequence context which gives rise to the major structural rearrangement upon a single point mutation.

The temporal sequence of major structural transitions are not made predictable by the shape space topology considered here. The indeterminacy derives from the additive modularity of RNA secondary structure. A complex base pairing pattern, such as the tRNA clover leaf, can be assembled by constructing the components (hairpins) in virtually any order. What becomes predictable, however, is the nature of the major structural transitions, and the fact that they must be preceded by a period of neutral drift leading to generalized shifts or *de novo* stack formation. Once the early phase of evolution has elapsed, a fairly educated guess about the number of such transitions needed to reach a target is possible.

Point mutations alone are probably impractical in the evolutionary design of large structures in the laboratory. The use of chain elongations and concatenations is likely to be more effective. In view of what we have shown here, it will be important to understand how the shape space topology responds to these changes in sequence space.

Acknowledgements

Discussions with Leo Buss, Ivo Hofacker, and Christian Reidys are gratefully acknowledged. We thank Martijn Huynen for critical comments on previous drafts of this paper. Stephan Kopp has kindly provided the boundary analysis of an exhaustively folded sequence space. Financial support was provided by the Austrian *Fonds zur Förderung der wissenschaftlichen Forschung* (Projects P-10578 and P-11065), by IIASA Laxenburg, Austria, by the Commission of the European Union (Contract Study PSS*0884), and by the integrative core research at the Santa Fe Institute.

References

[1] L. W. Buss. *The Evolution of Individuality.* Princeton University Press, Princeton, 1987.

[2] E. H. Ekland, J. W. Szostak, and D. P. Bartel. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, 269:364–370, 1995.

[3] A. D. Ellington. RNA selection. Aptamers achieve the desired recognition. *Current Biology*, 4:427–429, 1994.

[4] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophys. Chem.*, 26:123–147, 1987.

[5] S. Govindarajan and R. A. Goldstein. Why are some protein structures so common. *Proc. Natl. Acad. Sci. USA*, 93:3341–3345, 1996.

[6] S. Govindarajan and R. A. Goldstein. The foldability landscape of model proteins. in press, 1997.

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

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

[9] M. A. Huynen. Exploring phenotype space through neutral evolution. *J. Mol. Evol.*, 43:165–169, 1996.

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

[11] J. Jaeger, D. Turner, and M. Zuker. Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA*, 86:7706–7710, 1989.

[12] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK, 1983.

[13] H. Li, R. Helling, C. Tang, and N. Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273:666–669, 1996.

[14] J. Pütz, J. D. Puglisi, C. Florentz, and R. Giegé. Identity elements for specific aminoacylation of yeast $\mathrm{tRNA}^{\mathrm{Asp}}$. *Science*, 252:1696–1699, 1991.

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

[16] P. Schuster. How to search for RNA structures. Theoretical concepts in evolutionary biotechnology. *Journal of Biotechnology*, 41:239–257, 1995.

[17] P. Schuster. Genotypes with phenotypes: Adventures in an RNA toy world. *Biophys. Chem.*, 66:75–110, 1997.

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

[19] S. Spiegelman. An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 4:213–253, 1971.

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

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