

Review

Fitness Landscapes of Functional RNAs

Ádám Kun^{1,2,3,*} and Eörs Szathmáry^{1,3,4}

¹ Parmenides Center for the Conceptual Foundations of Science, Kirchplatz 1, 82049 Munich/Pullach, Germany; E-Mail: szathmary.eors@gmail.com

² MTA-ELTE-MTMT Ecology Research Group, Pázmány Péter sétány 1/C, 1117 Budapest, Hungary

³ Department of Plant Systematics, Ecology and Theoretical Biology, Institute of Biology, Eötvös University, Pázmány Péter sétány 1/C, 1117 Budapest, Hungary

⁴ MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, Pázmány Péter sétány 1/C, 1117 Budapest, Hungary

* Author to whom correspondence should be addressed; E-Mail: kunadam@elte.hu.

Academic Editor: Niles Lehman

Received: 25 May 2015 / Accepted: 3 August 2015 / Published: 21 August 2015

Abstract: The notion of fitness landscapes, a map between genotype and fitness, was proposed more than 80 years ago. For most of this time data was only available for a few alleles, and thus we had only a restricted view of the whole fitness landscape. Recently, advances in genetics and molecular biology allow a more detailed view of them. Here we review experimental and theoretical studies of fitness landscapes of functional RNAs, especially aptamers and ribozymes. We find that RNA structures can be divided into critical structures, connecting structures, neutral structures and forbidden structures. Such characterisation, coupled with theoretical sequence-to-structure predictions, allows us to construct the whole fitness landscape. Fitness landscapes then can be used to study evolution, and in our case the development of the RNA world.

Keywords: origin of life; ribozyme; aptamer; RNA; fitness landscape

1. Introduction

A fitness landscape is traditionally defined as a function that assigns to every genotype a numerical value linearly proportional to its fitness. Since only phenotypes have fitness, therefore a fitness landscape is a mapping from genotype to phenotype to fitness. Genotypes can be defined on the level of alleles.

Assume that there are only two alleles present for gene A: A1 and A2. A diploid individual thus could have three different genotypes: this is still a manageable genotype space. However, there are usually more than two alleles for any given gene, and there are many more genes even in the simplest bacteria. In the classical example of Sewall Wright [1] there are 10 allelomorphs in 1000 loci, which results in a space spanning 10^{1000} genotypes. Thus the main hurdle on the way to constructing fitness landscapes was already realized early in the development of evolutionary genetics: genotype spaces are vast. The size of the sequence space of an RNA or DNA of length L is 4^L . In case of peptides, this space is then mapped (via the genetic code) to peptide space with size of $20^{L/3}$. Sequence space is in turn mapped to phenotypes. Some would argue that mapping genotypes directly to fitness (or some functional properties, like enzyme activity) is enough, as ultimately we want to arrive at a fitness landscape. However, the paths available to us are not that direct. If we could attach a fitness value to all sequences in genotype space, then our work would be done: we would have a fitness landscape. Usually, we can only sample some part of the sequence space and we need to infer the whole fitness landscape from this limited data. Extending this data to the whole landscape requires some assumptions, which will be discussed in this review. The genotype-phenotype-fitness mapping corresponds to sequence-structure-function maps in biopolymers. An often coveted goal is to understand the structure-to-function map, and thus how function changes with changes in the structure.

Generally, a mutation can either change the structure or leave the structure intact. Changing the structure generally affects activity, except if the change happens far away from the active site or the substrate-binding parts. Mutations that do not affect the structure generally have little fitness effect [2,3], except when the exact chemical moiety present at the site is important for activity. We need to identify the **critical sites**, which are those where variation in the sequence destroys or severely decreases the activity of the molecule. *In vitro* generated mutants at these sites show low activity or no activity at all. Note that there are two kinds of critical site, and mutagenesis studies in themselves cannot distinguish between these two kinds. One kind we call a structural critical site, and the other kind a functional critical site. A **structural critical site** is important for the maintenance of structure; changes at this site would alter the structure extensively. A **functional critical site** contributes to function; for example, it binds the substrate, or lends a catalytic group to the active site. Change in a functional critical site might not alter the structure at all, but as the required chemical moiety is changed, it can no longer fulfill its function.

By identifying the critical sites and the structural features required for activity we can generalize the fitness landscape to areas of the genotype space not covered by mutants. In essence, a functional molecule has the right structural motifs (whatever their composition) and the right monomer at the functional critical sites. This generalization is usually missing, and unfortunately the data are not available to make this generalization. Without the generalization, we only have a list of genotypes and corresponding fitnesses for an insignificantly tiny part of the genotype space. This handful of genotype-fitness pairs might be located at and around a fitness peak (the wild-type sequence), but there could very well be many fitness peaks in the landscape and this particular one might not even be the highest [4]. If we accept that structure is an important determinant of phenotype and its knowledge is crucial for generalization, then we can state that RNAs lend themselves well to developing fitness landscapes.

The structure of an RNA can be approximated by theoretical calculations [5]. Such calculations are based on energetic parameters coming from experiments [6,7], and further refined by chemical probing and other biochemical analysis [6,8]. These approximation yields the so-called secondary structure of RNA (Figure 1). Secondary structure (or 2D structure) is the list of all intramolecular base-pairs the folded RNA harbours. The number of unique structures of this sequence space is estimated to be 2.35^L [9]. It is evident that there is by far fewer structures than sequences [9,10]. Current secondary structure predicting algorithms, for example the ViennaRNA Package [11], can predict more than 70% of the base pairs correctly based on an extensive set of known secondary structures [12]. The accuracy of these applications is constantly increasing [13], and they can predict the structure of longer sequences as well [14]. The function of the biological sequence ultimately depends on its 3-dimensional structure of which the secondary structure is only an approximation. Throughout the text we will mention instances when the secondary structure prediction needed to be amended because 3D interactions modified the structure. Such instances are examples where the usage of secondary structure as a proxy fails. The relative scarcity of these examples suggests that secondary structure is a good predictor of the 3D structure on which the function depends on.

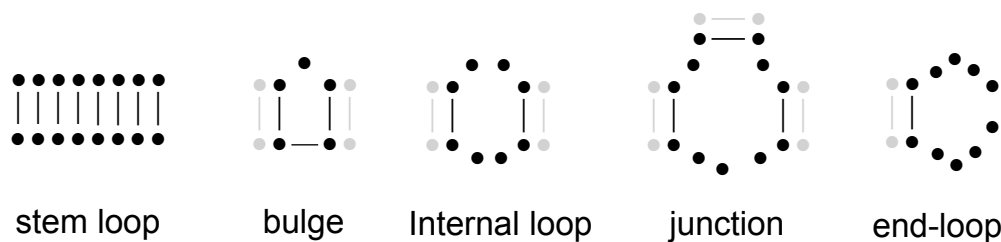


Figure 1. Secondary structure elements of RNAs. We can distinguish the following elements in a secondary structure: stem, bulge, internal loop, junction and end-loop. Nucleotides are depicted by a solid circle and hydrogen bonds by a line (except the horizontal line in the bulge, which depicts the covalent bond between the nucleotides). The greyed-out nucleotides do not belong to the given structure. Note that the exact number of nucleotides in a bulge, internal loop, junction and end-loop can vary. Examples of secondary structures can be seen in Figures 2–4.

The structure-function map can only be inferred by meticulous experimental analysis. Enzymatic activity or binding strength can be relatively easily measured, but knowing what certain parts of the structure do requires more investigation. In RNA viruses, like HIV, the underlying structure of the genome is important, as different sequences coding for the same protein (*i.e.*, synonymous mutation) can have different fitness [15,16]. Structural elements of the HIV genome, for example, were found to play role in slowing down translation so that the domains of the translated peptides can fold independently, and other structural elements allows frameshift during translation [17]. Understanding the fitness landscape underlying virus evolution [18,19] could help us predict their evolution and thus plan countermeasures in advance.

Fitness landscapes are not only theoretical constructs that might only be interesting to evolutionary biologist. They are also powerful conceptualization of the evolutionary potential of biological

entities. Furthermore, they can also be used to design and engineer novel functional variants of biomolecules. Here we have focused on the fitness landscapes of RNAs. Aptamers and ribozymes are now recognized as having therapeutic potential [20–25] in a wide array of diseases from virus infections [26,27] to cardio-vascular diseases [28]. They can also be used as biosensors [29] and in drug delivery [30–32]. Understanding of RNA structure–function landscapes not only gives us more insight into the development of the RNA world [33] but is of applied significance, too.

Fitness landscapes of microbes [34], peptides [35], viruses [36], and RNAs [37] are available to some extent. A recent review of fitness landscapes concluded that while we have begun to appreciate the need for such landscapes, especially as they turn out to be rugged and not smooth, we are only beginning to collect the required data [38]. Here we would like to review as well as extend what we know about the fitness landscapes of aptamers and ribozymes (functional RNA). Furthermore, we would like to highlight areas badly in need of more data.

2. Whole Genotype–Fitness Map

A full fitness landscape requires a map from every possible genotype to fitness. The realization of every possible genotype is usually technically not feasible due to the combinatorial explosion of the genotype space. Library sizes employed in SELEX [39–42] are about 10^{13} – 10^{15} , which can cover the whole sequence space of up to 25mers. Thus, for very short sequences, the genotype space can be realized and by subsequent selection probed. We need to appreciate this possibility! A fresh 1 TB HDD can only hold a tiny fraction of this information (2.5×10^{11} positions if we store each nucleotide in 2 bits, so each byte stores 4 nucleotides). This is a perfect example of when “wet” computers can store more information per volume or mass, and do more computations per unit time (*i.e.*, evaluation of function) than digital computers.

The possibility of scanning a whole genotype space was realized in Irene Chen’s lab [43]. They have started from a pool of nearly all 24mers (there are 281,474,976,710,656 different sequences) in thousand-fold excess. Then they selected for GTP binding. The aptamers were sequenced and thus the resulting fitness landscape maps sequences to GTP binding strength. As expected, there are peaks in the landscape around which the functional sequences congregate [43]. Based on the results we can infer some general characteristics:

- (1) *There are multiple solutions to being a GTP aptamer.* The fact that there is neither a common sequence motif nor a common structural motif that would tie together the fitness peaks suggests that there are multiple, independent solutions to a GTP aptamer. Thus the sequence space can be populated with different aptamers that have nothing in common except that they have the very same function.
- (2) *There are uncommon structural solutions as well.* In order to assess how frequent or infrequent the structures reported in S14 of the paper by Jiménez *et al.* [43] were, we folded 1 million random sequences of length 24 and estimated the frequency of the common structures from this sample. We know that this is a tiny fraction of the 2.814×10^{14} unique 24mers, but the common structures should be well represented even in this small sample. The structural repertoire of 24mers is rather limited. The structural space is dominated by the unstructured structure with roughly 18% of the

sequences folding into this structure. Four of the peaks were found to be unstructured according to the Vienna Package. We cannot exclude the possibility of such a structure to bind the ligand, but it is more likely that an induced fit mechanism determines the final structure, or that the structure is stabilized by interactions that cannot be predicted by the employed 2D folding algorithm. Most of the other peaks were also folding into quite common structures (Table 1). A common structure is defined as those structures that are formed by more sequences than the average structure [44], *i.e.*, $N_c > 4^L / S_L$, where N_c is the number of sequences folding into a common structure, S_L is the number of distinct structures of length L . Most of the sequences fold into one of the common structures [44]. For example, more than 90% of the 25 nt long sequences comprising only G and C nucleotides fold into a common structure. Consequently, while there could be many rare structures, very few sequences fold into them. However, there was one peak (m20j22) with a structure that we have not found in our sample of structures for 10 million sequences. We can safely assume that this structure is uncommon. While we expect that evolution will mostly (if not always) find the common structures, it is reassuring to know that uncommon structures can also have function, just they are unlikely to be found by evolution. Aptamers and ribozymes we know are all folding into common structures not because only common structures can have function, but because they are the ones that can be reached by evolution.

Table 1. Structures and their frequencies among common structures of fitness peaks of a GTP aptamer.

Structure ¹	Peaks folding into this structure ²	Frequency in our sample ³	Frequency in our larger sample ⁴
.....	m18j13, m14j12, m04j04, m02j01	18.7%	18.8%
(((.....)))	m19j09	3.4%	3.4%
(((.....)))	m15j18, m10j11, m07j07	2.4%	2.5%
(((.....)))	m16j20	1.8%	1.9%
(((.....)))	m17j08	0.8%	0.8%
(((.....)))	m01j03	0.6%	0.6%
(((.....)))	m03j02	0.4%	0.4%
((.....))	m05j10	0.3%	0.3%
(((.....)))	m06j06	0.05%	0.006%
((.....))...((.....))	m20j22	not found in our sample	not found in our sample

¹ We assume that leading and trailing unpaired bases are not important, and they were left out; ² Naming of the peaks is as in the original paper; ³ Based on 1 million random sequences; ⁴ Based on 10 million random sequences.

3. Lessons from Aptamers

The landscape described previously is an important milestone in the study of fitness landscapes. Strictly speaking it represents a genotype-to-fitness map, without too much reference to the phenotype. Knowing the structural and sequence (phenotype) requirement of a GTP aptamer would allow us to answer questions like “Is there a sequence that need to be in the loop in order to bind GTP?”, “Is there a minimum stem-loop length for the aptamer?” or “Is there any structural significance to the internal loop in m05j10 or m06j06?” Selection of longer aptamers followed by deep-sequencing and structure probing could provide us with insight, albeit at the expense of looking at the whole sequence space. Aptamer research, due to relatively shorter sequences and easier selection, has offered some further insight into the structural requirement of certain functions. Most studies on aptamers or ribozymes only report the selected sequences [45–47] though some report the folded structures [48,49] as well. Studies that actively try to deduce a consensus structure [50,51] lead us one step further in understanding structure–function relationships. Recent studies go even further and implement the consensus in order to obtain the minimum motif(s) required for activity.

For example, aptamers selected to bind to the 5'-untranslated region of the HIV have a common structural element and a common sequence motif: the “*consensus octamer GGCAAGGA is always placed in an apical loop flanked by complementary sequences that form a double-stranded region of at least 4 bp in length*” [52] (Figure 2a). In the *in vitro* selected aptamers (64 nt) this motif could be found in various contexts (see Figure 4 in [52]), but an engineered minimal loop with the right structure and sequence can also inhibit HIV. A binding aptamer against the trans-activation responsive (TAR) RNA part of HIV-1 also demonstrate that sequence complementarity in an end-loop is not sufficient to ensure functionality [53]. The requirement for a binding aptamer was found to be “(1) a loop complementary to the entire loop of the target RNA, (2) a double-stranded stem region comprising generally more than 4 bp adjacent to the loop, and (3) a closing GA pair” [53] (Figure 2b). These aptamers contain perfect examples of critical sites: certain sequences need to be located in the right structural context. Changing either the sequence or the structural context can ruin the aptamer.

A similar result was obtained for the aptamer that binds to the reverse transcriptase of HIV: the strongly binding (6/5)AL aptamer needs to have an internal loop of 6 nucleotides on one side, and 5 on the other, flanked by at least 6 base pairs on one side and 10 on the other (Figure 2c) [54]. The flanking stems can be longer, and there could be slight variation in the sequence of the internal loop. Which stem ends in a loop, and which bears the start/end of the sequence, does not seem to matter. Another aptamer selected against the same target show a consensus structure which is a “*simplified stem-loop structure in which the UCAA bulge is flanked on one side by two highly conserved base pairs (AC/GU) and on the other by a largely generic helix that is interrupted by a single unpaired U residue*” [55] (Figure 2d). While in-depth analysis of the structure showed that there might be other requirements for an active inhibitor (functional aptamer), the UCAA bulge and the flanking base pairs are a constant feature, and quite a number of other structural elements can be changed without loss of function [55].

The consensus structures/sequences of Figure 2 are sufficient to ensure aptamer binding, but they are certainly not the only structures that are functional. The studies [52,54,55] demonstrate that a small motif embedded in a larger structure can be functional. Thus (especially for aptamers) we need to look

for small motifs that are common to the selected structures, and there is no need to strictly adhere (pun intended) to maintaining much larger structures.

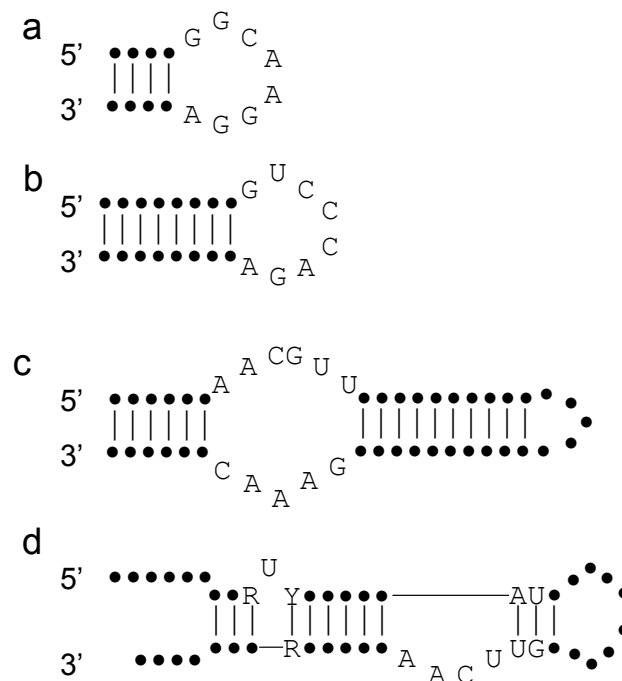


Figure 2. Consensus structures of aptamers selected against HIV-1. Dots represent nucleotides without sequence specificity. Vertical lines represent hydrogen bonds, whereas horizontal lines represent the link between nucleotides (these links are omitted between nucleotides close to each other of the figure). (a) Aptamer against the 5'-UTH region of HIV, (b–d) aptamers against the reverse transcriptase of HIV.

4. Lessons from Ribozymes

Aptamers usually bind via a stretch of well-defined nucleotides, and because of their relatively short length they cannot have too rich a structure. Quite a few aptamers are made of an end-loop with a small stem. This is only as far as we can go with the study of aptamers in our quest for a fitness landscape. Ultimately we would like to understand the fitness landscapes of enzymes (genes). As ribozymes are generally around 50–300 nucleotides long (very small ribozymes are also known [56,57]), their whole genotype space cannot be produced. We can only hope to study the mutational neighbourhood of the wild-type sequence. A sequence of length L has $3L$ single point mutants, *i.e.*, mutants that only differ at one position. For large sequences, this number of mutations can also be quite high, not to mention that the fitness of the sequences have to be characterized as well. This is not unfeasible either, as demonstrated recently for the *E. coli* TEM-1 β -lactamase. Nearly all single-point mutants in genome space (2536/2583) and in protein space (15,167/18,081) were generated and their fitness was estimated [58].

Still, there are very few ribozymes studied in enough detail for any meaningful inference of a functionality landscape. For the most part, we can only consider the natural ribozymes [59]. Unfortunately even among these few ribozymes some need to be excluded for practical reasons: the

Hepatitis Delta Virus and like ribozymes [60] and the RNase P [61] have a pseudo-knot [62], a structural element that cannot be computed by the most widely used secondary structure-predicting algorithms. This is also true for the well-studied artificial Class I ligase [63] and derived ribozymes (e.g., the RNA dependent RNA polymerase ribozyme [64–67]). Enough data might not be available for the more recently discovered natural ribozymes like the *glmS* ribozyme [68], and the twister ribozymes [69]. This leaves the Group I [70] and II introns [71], *Neurospora* VS [72], the hairpin [73], the hammerhead ribozyme [74], the Diels-Alderase ribozyme [75] and the class II ligase [76] for consideration.

The Class II ligase would be the perfect subject for a fitness landscape: it is short (54 nt) and does not have a pseudo-knot. Furthermore, and more importantly, each of its nucleotides was replaced by the other 3 nucleotides not present at the given site, and activity was measured [77]. There is a catch, however: even the fittest mutant has an order of magnitude less activity than the master sequence. That is very unusual, as some part of the ribozyme is probably not that important, and some sequence variations are generally tolerated. The activity of the wild type is described in another paper [78]. The methodology employed to assay enzymatic activity was different in the two studies, and thus it is not surprising to find different activities for the same mutations (Table 2). This raises the question of whether the activity of the wild-type ribozyme would be the same in the second study. The activities highly correlate (linear fit with $R^2 = 0.976$), but the slope is mainly determined by the activities of mutations G2U, G2A, G2C and G49 (and the data points are not independent). According to this admittedly flawed statistic, the activity of the wild-type sequence should be 3.7 times lower in the second study. However, the wild-type ribozyme still exhibit orders of magnitude higher activity than any of the mutants. Such a high fitness peak compared to its mutational neighborhood (*i.e.*, mutants differing at one site) is interesting, and merits further study.

Table 2. Activities of mutants of the Class II Ligase.

Mutation of the Class II Ligase	Activity from [77]	Activity from [78]
G1A	0.00173	3.80×10^{-4}
G2A	0.03043	1.10×10^{-1}
G2C	0.00599	3.50×10^{-2}
G2U	0.03584	1.30×10^{-1}
A3G	0.00147	6.30×10^{-4}
A3C	0.00108	8.00×10^{-5}
A3U	0.00134	1.60×10^{-4}
G49A	0.01048	2.10×10^{-2}
G47A	0.00111	2.00×10^{-5}
G47C	0.00089	1.90×10^{-5}
G47U	0.00108	9.50×10^{-6}

The core part of the class II ribozyme (the 3-way junction and arms P2 and P3) was predicted well by the Vienna RNA package [11,79]. Unfortunately, the P1 stem containing the template to which the substrate is hybridized cannot be predicted well. Such a discrepancy between the predicted and the chemically determined secondary structure hints at non-canonical intramolecular interactions. For

example, the 49mer Diels-Alderase ribozyme's [80] predicted secondary structure has a base pairing between G26-C44 (Figure 3a, this pairing is not shown). However, the GGAG sequence, which actually has the substrate tethered to it, binds with bases in the internal loop (Figure 3a, hydrogen bonds with dashed lines) [81]. Thus the G26-C44 pair cannot form. Similarly for the VS ribozyme, the secondary-structure predicted by RNAfold [11] and the one predicted by chemical probing [82] are different. NMR studies [83] corroborated the theoretical secondary-structure with the addition that the active site harbours a number of non-canonical base-pairs (A⁺-C, G-A and U-G). This structure is probably the ground-state structure [83], which then transforms into the active structure determined by chemical probing [82] (Figure 3b).

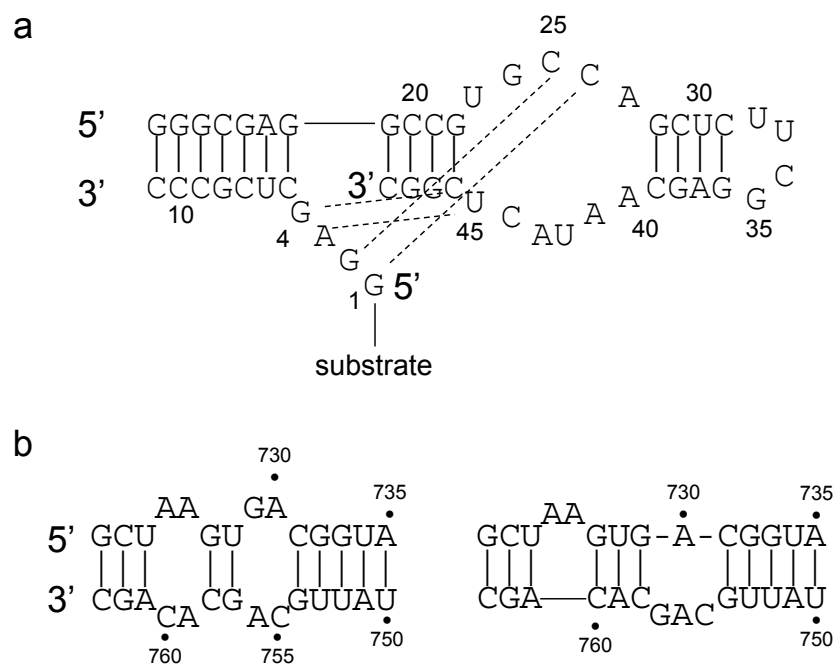


Figure 3. (a) The Diels-Alderase ribozyme's secondary structure. Numbering is as in [82]. Dashed lines represent hydrogen bonds that are present but not predicted by conventional secondary structure predicting algorithms. (b) The predicted secondary structure of the VS ribozyme according to folding algorithms and NMR (**left**) and the active conformation determined by chemical probing (**right**).

Thus, while the secondary structure can be a good proxy for the 3-dimensional structure, some interactions out of the plane or the interaction with the substrate can modify the structure.

We have constructed a functionality landscape (used as a proxy for fitness) for the *Neurospora* VS ribozyme and the hairpin ribozyme [84,85]. Here we discuss the fitness landscape of the VS ribozyme in more detail. The VS ribozyme is a moderately sized ribozyme with a rich structure (Figure 4). The full 3D structure of the ribozyme is not yet available, but NMR has been extensively employed to study important localities of the ribozyme [83,86–94].

Basically, the end-loop of region V binds the substrate, so that the internal loop in region VI can align with the cleavage site. Structurally this means that the length of region V [95–97], the linker region III [95] and the part of region VI between the J236 junction and the active site internal-loop are important.

In the case of region V, it needs to be the right length or at most one base-pair shorter or longer [97]. As it needs to align with region I (the substrate), a change in the substrate (increase in the length of region Ib) necessitates a decrease in the length of region V [97]. The length of the rest of the stem-loops can be varied in the *trans*-acting ribozyme [95,98]. We note here that the length of region II was not varied; this part is generally considered to be unimportant for the *trans*-acting ribozyme. It would still be nice to know for sure. However, the bulge at position 652 in stem-loop II is required, though its chemical nature is not important [82,98]. Activity decrease to about fifth of the wild-type value for A652C and A652G, but the activity of A652G is one tenth that of the wild-type ribozyme. The former mutations do not change the predicted secondary structure, while the latter does (albeit causing only local rearrangement). We know that pairing of this bulge does not affect activity much [82,95], but removal of the bulge or putting it on the other side of the stem-loop does. Similarly, changing the nucleotides of the bulge at positions 725/725 in region VI does not change activity (actually, it increases activity, except for the A725G mutation which also changes the structure), but removal of the bulge decreases activity [98]. The above is generally true for the bulge at position 718 in region III as well. No data is available for change of nucleotide, but removal and inversion of the bulge has the same deleterious effect [95] as previously. The predicted secondary-structure would only change for A718U, and then only slightly.

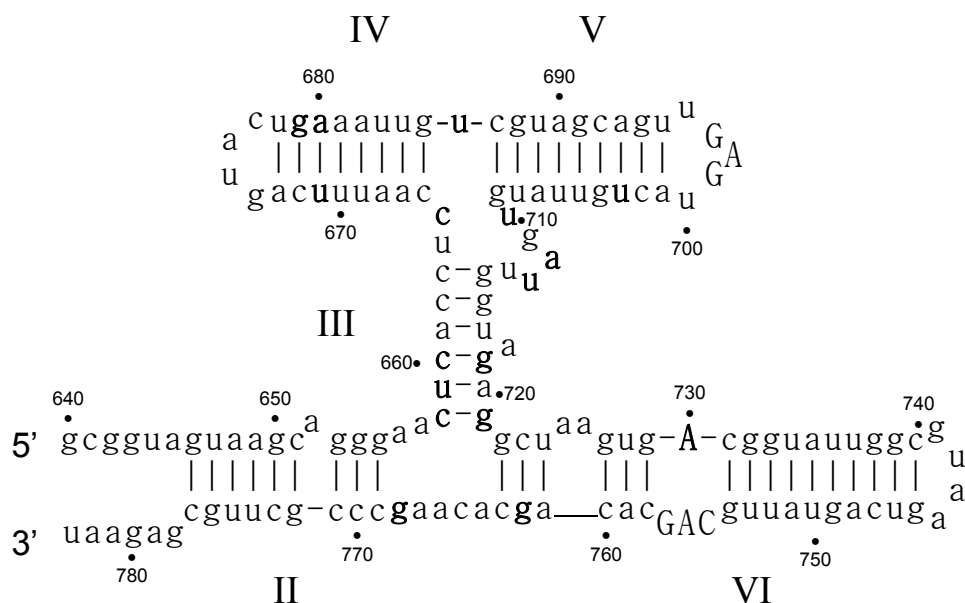


Figure 4. Secondary structure of the *Neurospora* VS ribozyme. Numbering is according to convention [82]. Positions in bold represent structural critical sites. Capitalized positions represent functional critical sites. (A730 is both).

The junction 3-4-5 contains a number of critical sites. The nucleotides U710/G711/A712 form a canonical U turn (5'-UNR-3' R = A or G) [99]. Thus at position 711 there could be any ribozyme, and indeed changes at these site are neutral. The A712C and A712U mutations would interfere with the U turn, and they have low activity (though they do not change the predicted secondary-structure). Interestingly, the A712G mutation has a low activity, though the U turn could theoretically form. But this mutation also rearranges the minimum free energy structure, and thus it has a low activity. Changing

the position 686 again changes the predicted secondary-structure, and accordingly the U686A mutation has low activity (activity for U686C and U686G is not available). Similarly, C665G and C665A decrease activity and change the structure. Unfortunately the activity of C665U is not available, as it might not decrease activity, since the predicted secondary-structure does not change. Thus the positions 710, 712, 686 and 665 are structural critical sites.

The catalytic site of the ribozyme is the internal loop in region VI of the ribozyme [98], especially the adenine at position 730. The nucleotides C755, A756 and G757 are also important, and most of their mutations change the structure as well. The A730 nucleotide is both a critical functional site and a critical structural site.

According to the above we can distinguish four kinds of structural elements:

- **Critical structure** contains functional critical sites. Figure 2 shows a number of such structural elements, and the end loop of region V and the internal loop of region VI of the VS ribozyme are further examples.
- **Connecting structure** contains no functional critical sites, but it has structural critical sites, and is important for the positioning of the functional critical sites. Region III is such a structure: it connects the substrate binding region V and the active site region VI.
- **Neutral structure** is the part of the ribozyme that can be freely changed, even removed. Region IV and II, and the distal part of region VI are examples of such structures.
- **Forbidden structure** is one that would decrease or abolish activity if present. Examples are the inverted bulges in region II, III and IV of the VS ribozyme. The information content of absence is often overlooked [100]. For a general fitness landscape we also need to know what can be there and what cannot.

The substrate region (region I) of the VS ribozyme has received some focus in recent years. For example, the G638 site in the substrate loop is also a functional critical site. The structure of the G638A mutant is the same as the WT substrate, yet the catalytic rate is nearly 4 orders of magnitudes lower. “The results suggest that the main requirement for the nucleobase at position 638 for cleavage under standard conditions is the C6 carbonyl, or the imino proton on N1, or both” [101].

We have folded all possible point mutants of the VS ribozyme (Supplementary Table S1) in order to determine structural critical sites. Sites at which the predicted secondary structure changed considerably if a point-mutation was introduced are potential structural critical sites. The mutations A648C, C773A, U703A, U703G, U705A, G694A, A680C, and G679C change the predicted structure at more than 34 places, which translates to considerable structural rearrangements. Unfortunately no activity data is available for any of these sites.

5. Double Mutations and the Question of Epistasis

The local view of a fitness landscape usually falls short when one needs to assess the effect of multiple mutations. We mostly know what happens to the ribozyme (or gene, or organism) if we introduce one mutation; unfortunately there is hardly any data on the effect of multiple mutations. People might say that there is a huge amount of literature on epistasis [102]. This is true, but most studies on epistasis (or any interaction of mutations) only record the genotype and the resultant fitness effect.

Suppose we introduce two mutations into a sequence. What will happen? Based on the mutational neighbourhood of ribozymes and other sequences, we can safely say that if one of the mutated sites is a critical site then activity is abolished. Can we say anything more? Can we say that if none of the mutation affects activity then activity will not be affected? No, we cannot. The minimal free energy structure can very well change, and the whole sequence can fold into a very different structure, even if the individual changes do not affect activity. Can we say that if both mutations only slightly affect fitness, then they both together will again only slightly affect fitness (maybe to a slightly greater extent than the individual ones)? Again, we cannot. For example, restoring a base pair usually restores activity. Mutations of base pairs are thus examples of sign epistasis [103], in which the individual mutations decrease activity but both together restore (increase) activity. Sign epistasis is usually defined the other way around, *i.e.*, substitutions that are beneficial individually, but lethal or very deleterious together [104–108].

If we want to make some generalizations about the effect of double and multiple mutations, we should treat base pairs differently. We suggest that base pairs should be considered *one* unit of an element in a structure, and not as two independent bases. A given base-pair can either mutate into a mispair (usually lowering or abolishing activity) or mutate into a base-pair, in which case activity does not necessarily change. There are 47 base-pairs in the *cis*-acting VS ribozyme (Supplementary Table S1). Out of these base pairs, only 6 were such that changing the base pair to a canonical base pair changed the predicted secondary structure. In a further 11 cases, the structure changes if the base-pair is changed to a GU or an UG base-pair. Otherwise (30 cases), the secondary structure does not change if the base-pair is changed to any of the other possible base-pairs. However, not all such predicted neutral changes are actually neutral (Supplementary Table S1). A mispair (an open base-pair compared to wild-type structure) usually has a mild deleterious effect on activity [85]. More than two mispairs in the same region surely abolish activity, though, which again demonstrates that the structure can slightly vary, but too much structural change leads to an inactive ribozyme. The base-pairs flanking the junctions are, however, critical. In some cases, changing them to other base-pairs also changes the structure (658:721 CG→AU; 663:715 CG→GU; 655:769 GC→UG). Thus structural critical sites can be generalized to **structural critical units**, encompassing critical base-pairs.

Outside the realms of base-pairs and critical sites, there could still be double and multiple mutations whose effect we need to know in order to characterize a fitness landscape. We do not expect complex epistatic interaction between the sites; if there were some, then at least one of the sites would be a critical site. The most common assumption is a multiplicative (additive) fitness effect, in which case, if mutants individually have, for example, half the activity of the wild type, then the double mutant would have 25% activity ($0.5 \cdot 0.5$). An early and pioneering study by Lehman and Joyce [109] corroborated this expectation for the *Tetrahymena* ribozyme. Since then others also found either multiplicative fitness effects or some synergistic effects, in which cases the double mutant has higher activity than expected by the multiplicative model. Such studies include the study of the RNA virus *Tobacco etch potyvirus* [110]; the study of kinase ribozymes [111,112]; and our own collection of mutants of small, natural ribozymes [85]. In conclusion, the easiest way to deal with multiple mutations is to assume mutational independence (multiplicative effects), although it slightly overestimates the decrease in fitness due to multiple mutations.

We still hasten to add that more data on triple and higher order mutants as well as systematic investigation of multiple mutations of known ribozymes would greatly help us to assess the combined effects of multiple mutations.

6. Far from the Wild-Type Region of the Landscape

We know more and more about the immediate mutational neighbourhood of RNA viruses, aptamers and ribozymes. This alone allows us to build better representations of fitness landscapes. The ultimate goal is to also know the fitness landscape far from the wild-type sequence. To our knowledge no such investigation has been conducted. We know that there are multiple fitness peaks [43] and that the same secondary structure can be realized with very different sequences [10,113]. Hayashi *et al.* [114] have randomized a part of the fd phage, thereby abolishing infectivity. They then introduced random mutations and measured infectivity. Fitness increased smoothly to about 40% the original level, but then the ruggedness of the landscape became dominant and evolution slowed down. The peptide evolved to a sequence different from the wild-type, thus arriving at another fitness peak. This means that there are fitness peaks of equal height away from the original one. We do not know if these sequences were structurally similar to the original structure (and this was on a peptide anyway). The studies discussed in the section on aptamers (see above) are informative here as well: they demonstrate that as long as core structural/functional elements are present, other parts of the aptamer can change without loss of activity. We can therefore assume that the fitness landscape has a similar structure surrounding functional but sequence-wise different (as much as sequence can differ, c.f. functional critical sites) sequences.

7. Summary

- There are multiple solutions to the same functionality.
- Functionality often depends on small motifs, and the rest of the aptamer need only not interfere with the motif or to allow the required stacking of motif parts.
- Even very rare structures can have a function, though evolution will rarely, if ever, find them.
- There are critical sites where substitutions are not tolerated. These sites can be either functional critical or structural critical sites/units. Change in a functional critical site does not change the structure of the RNA, but abolishes activity. Change in a structural unit (a single nucleotide or a base-pair) abolishes activity through change in the structure.
- Structural elements or bases not appearing at certain places are as informative as permitted structures. While these structural elements, by definition, are not part of the structure, they show the limit of structural changes. We call these features forbidden structures.
- Apart from critical and forbidden structures, a structure can be characterized by uncovering connecting structures and neutral structures too. Connecting structures link the structural critical sites, but their exact forms are not prescribed. Neutral structures can be changed (to some extent) freely without affecting activity.
- RNA structure is even important for peptides, as the structure of the mRNA or RNA viruses can influence the replication, expression and activity of the translated peptide even when the resultant peptides would have the very same sequence (*i.e.*, there are only synonymous changes).

Acknowledgments

Financial support has been provided by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no (294332). Ádám Kun acknowledges support by the European Union and co-financed by the European Social Fund (grant agreement no. TAMOP 4.2.1/B-09/1/KMR-2010-0003) and from the Hungarian Research Grants (OTKA K100299). Ádám Kun gratefully acknowledges a János Bolyai Research Fellowship of the Hungarian Academy of Sciences. This work was carried out as part of EU COST action CM1304 “Emergence and Evolution of Complex Chemical Systems”.

Conflicts of Interest

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Wright, S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In Proceedings of the Sixth International Congress on Genetics, New York, NY, USA, 1932; pp. 355–366.
2. Lafontaine, D.A.; Norman, D.G.; Lilley, D.M.J. The structure and active site of the varkud satellite ribozyme. *Biochem. Soc. Trans.* **2002**, *30*, 1170–1175. [[CrossRef](#)] [[PubMed](#)]
3. Fedor, M. Structure and function of the hairpin ribozyme. *J. Mol. Biol.* **2000**, *297*, 269–291. [[CrossRef](#)] [[PubMed](#)]
4. Lehman, N. Assessing the likelihood of recurrence during RNA evolution *in vitro*. *Artif. Life* **2004**, *10*, 1–22. [[CrossRef](#)] [[PubMed](#)]
5. Higgs, P.G. RNA secondary structure: Physical and computational aspects. *Q. Rev. Biophys.* **2000**, *33*, 199–253. [[CrossRef](#)] [[PubMed](#)]
6. Mathews, D.H.; Disney, M.D.; Childs, J.L.; Schroeder, S.J.; Zuker, M.; Turner, D.H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7287–7292. [[CrossRef](#)] [[PubMed](#)]
7. Mathews, D.H.; Sabina, J.; Zuker, M.; Turner, H. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.* **1999**, *288*, 911–940. [[CrossRef](#)] [[PubMed](#)]
8. Deigan, K.E.; Li, T.W.; Mathews, D.H.; Weeks, K.M. Accurate shape-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 97–102. [[CrossRef](#)] [[PubMed](#)]
9. Haslinger, C.; Stadler, P.F. RNA structure with pseudo-knots: Graph-theoretical and combinatorial properties. *Bull. Math. Biol.* **1999**, *61*, 437–467. [[CrossRef](#)] [[PubMed](#)]
10. Schuster, P.; Fontana, W.; Stadler, P.F.; Hofacker, I.L. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **1994**, *255*, 279–284. [[CrossRef](#)] [[PubMed](#)]

11. Lorenz, R.; Bernhart, S.; Honer zu Siederdisen, C.; Tafer, H.; Flamm, C.; Stadler, P.; Hofacker, I. Viennarna package 2.0. *Algorithms Mol. Biol.* **2011**, *6*. [[CrossRef](#)] [[PubMed](#)]
12. Andronescu, M.; Bereg, V.; Hoos, H.H.; Condon, A. RNA strand: The RNA secondary structure and statistical analysis database. *BMC Bioinform.* **2008**, *9*, 340–340. [[CrossRef](#)] [[PubMed](#)]
13. Doshi, K.J.; Cannone, J.J.; Cobaugh, C.W.; Gutell, R.R. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinform.* **2004**, *5*. [[CrossRef](#)] [[PubMed](#)]
14. Dowell, R.D.; Eddy, S.R. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. *BMC Bioinform.* **2004**, *5*. [[CrossRef](#)] [[PubMed](#)]
15. Han, J.-Q.; Townsend, H.L.; Jha, B.K.; Paranjape, J.M.; Silverman, R.H.; Barton, D.J. A phylogenetically conserved RNA structure in the poliovirus open reading frame inhibits the antiviral endoribonuclease RNase L. *J. Virol.* **2007**, *81*, 5561–5572. [[CrossRef](#)] [[PubMed](#)]
16. Acevedo, A.; Brodsky, L.; Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **2014**, *505*, 686–690. [[CrossRef](#)] [[PubMed](#)]
17. Watts, J.M.; Dang, K.K.; Gorelick, R.J.; Leonard, C.W.; Bess, J.W., Jr.; Swanstrom, R.; Burch, C.L.; Weeks, K.M. Architecture and secondary structure of an entire hiv-1 RNA genome. *Nature* **2009**, *460*, 711–716. [[CrossRef](#)] [[PubMed](#)]
18. Tsetsarkin, K.A.; Chen, R.; Yun, R.; Rossi, S.L.; Plante, K.S.; Guerbois, M.; Forrester, N.; Perng, G.C.; Sreekumar, E.; Leal, G.; *et al.* Multi-peaked adaptive landscape for chikungunya virus evolution predicts continued fitness optimization in aedes albopictus mosquitoes. *Nat. Commun.* **2014**, *5*. [[CrossRef](#)] [[PubMed](#)]
19. Luksza, M.; Lassig, M. A predictive fitness model for influenza. *Nature* **2014**, *507*, 57–61. [[CrossRef](#)] [[PubMed](#)]
20. Marton, S.; Reyes-Darias, J.A.; Sánchez-Luque, F.J.; Romero-López, C.; Berzal-Herranz, A. *In vitro* and *ex vivo* selection procedures for identifying potentially therapeutic DNA and RNA molecules. *Molecules* **2010**, *15*, 4610–4638. [[CrossRef](#)] [[PubMed](#)]
21. Bagheri, S.; Kashani-Sabet, M. Ribozymes in the age of molecular therapeutics. *Curr. Mol. Med.* **2004**, *4*, 489–506. [[CrossRef](#)] [[PubMed](#)]
22. Thiel, K.W.; Giangrande, P.H. Therapeutic applications of DNA and RNA aptamers. *Oligonucleotides* **2009**, *19*, 209–222. [[CrossRef](#)] [[PubMed](#)]
23. Bunka, D.H.J.; Platonova, O.; Stockley, P.G. Development of aptamer therapeutics. *Curr. Opin. Pharmacol.* **2010**, *10*, 557–562. [[CrossRef](#)] [[PubMed](#)]
24. Burnett, J.C.; Rossi, J.J. RNA-based therapeutics: Current progress and future prospects. *Chem. Biol.* **2012**, *19*, 60–71. [[CrossRef](#)] [[PubMed](#)]
25. Santosh, B.; Yadava, P.K. Nucleic acid aptamers: Research tools in disease diagnostics and therapeutics. *BioMed Res. Int.* **2014**, *2014*, 540451. [[CrossRef](#)] [[PubMed](#)]
26. Bunka, D.H.J.; Stockley, P.G. Aptamers come of age—At last. *Nat. Rev. Microbiol.* **2006**, *4*, 588–596. [[CrossRef](#)] [[PubMed](#)]
27. Shum, K.-T.; Zhou, J.; Rossi, J. Aptamer-based therapeutics: New approaches to combat human viral diseases. *Pharmaceuticals* **2013**, *6*, 1507–1542. [[CrossRef](#)] [[PubMed](#)]

28. Wang, P.; Yang, Y.; Hong, H.; Zhang, Y.; Cai, W.; Fang, D. Aptamers as therapeutics in cardiovascular diseases. *Curr. Med. Chem.* **2011**, *18*, 4169–4174. [[CrossRef](#)] [[PubMed](#)]
29. Strehlitz, B.; Reinemann, C.; Linkorn, S.; Stoltenburg, R. Aptamers for pharmaceuticals and their application in environmental analytics. *Bioanal. Rev.* **2012**, *4*, 1–30. [[CrossRef](#)] [[PubMed](#)]
30. Reinemann, C.; Strehlitz, B. Aptamer-modified nanoparticles and their use in cancer diagnostics and treatment. *Swiss Med. Wkly.* **2014**, *144*, w13908. [[CrossRef](#)] [[PubMed](#)]
31. Aravind, A.; Yoshida, Y.; Maekawa, T.; Kumar, D.S. Aptamer-conjugated polymeric nanoparticles for targeted cancer therapy. *Drug Deliv. Transl. Res.* **2012**, *2*, 418–436. [[CrossRef](#)] [[PubMed](#)]
32. Barbas, A.S.; Mi, J.; Clary, B.M.; White, R.R. Aptamer applications for targeted cancer therapy. *Future Oncol.* **2010**, *6*, 1117–1126. [[CrossRef](#)] [[PubMed](#)]
33. Kun, Á.; Szilágyi, A.; Könyű, B.; Boza, G.; Zachár, I.; Szathmáry, E. The dynamics of the RNA world: Insights and challenges. *Ann. N. Y. Acad. Sci.* **2015**, *1341*, 75–95. [[CrossRef](#)] [[PubMed](#)]
34. Colegrave, N.; Buckling, A. Microbial experiments on adaptive landscapes. *BioEssays* **2005**, *27*, 1167–1173. [[CrossRef](#)] [[PubMed](#)]
35. Carneiro, M.; Hartl, D.L. Adaptive landscapes and protein evolution. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1747–1751. [[CrossRef](#)] [[PubMed](#)]
36. Elena, S.F.; Sanjuán, R. RNA viruses as complex adaptive systems. *Biosystems* **2005**, *81*, 31–41. [[CrossRef](#)] [[PubMed](#)]
37. Athavale, S.S.; Spicer, B.; Chen, I.A. Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Curr. Opin. Chem. Biol.* **2014**, *22*, 35–39. [[CrossRef](#)] [[PubMed](#)]
38. De Visser, J.A.G.M.; Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **2014**, *15*, 480–490. [[CrossRef](#)] [[PubMed](#)]
39. Stoltenburg, R.; Reinemann, C.; Strehlitz, B. Selex—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol. Eng.* **2007**, *24*, 381–403. [[CrossRef](#)] [[PubMed](#)]
40. Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage t4 DNA polymerase. *Science* **1990**, *249*, 505–510. [[CrossRef](#)] [[PubMed](#)]
41. Ellington, A.D.; Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346*, 818–822. [[CrossRef](#)] [[PubMed](#)]
42. Robertson, D.L.; Joyce, G.F. Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **1990**, *344*, 467–468. [[CrossRef](#)] [[PubMed](#)]
43. Jiménez, J.I.; Xulvi-Brunet, R.; Campbell, G.W.; Turk-MacLeod, R.; Chen, I.A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14984–14989. [[CrossRef](#)] [[PubMed](#)]
44. Schuster, P. Genotypes with phenotypes: Adventures in an RNA toy world. *Biophys. Chem.* **1997**, *66*, 75–110. [[CrossRef](#)]
45. Bayrac, A.T.; Sefah, K.; Parekh, P.; Bayrac, C.; Gulbakan, B.; Oktem, H.A.; Tan, W. *In vitro* selection of DNA aptamers to glioblastoma multiforme. *ACS Chem. Neurosci.* **2011**, *2*, 175–181. [[CrossRef](#)] [[PubMed](#)]
46. Schlosser, K.; Li, Y. Diverse evolutionary trajectories characterize a community of RNA-cleaving deoxyribozymes: A case study into the population dynamics of *in vitro* selection. *J. Mol. Evol.* **2005**, *61*, 192–206. [[CrossRef](#)] [[PubMed](#)]

47. Ameta, S.; Winz, M.-L.; Previti, C.; Jäschke, A. Next-generation sequencing reveals how RNA catalysts evolve from random space. *Nucleic Acids Res.* **2014**, *42*, 1303–1310. [[CrossRef](#)] [[PubMed](#)]
48. Cho, M.; Xiao, Y.; Nie, J.; Stewart, R.; Csordas, A.T.; Oh, S.S.; Thomson, J.A.; Soh, H.T. Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15373–15378. [[CrossRef](#)] [[PubMed](#)]
49. Thiel, W.H.; Bair, T.; Peek, A.S.; Liu, X.; Dassie, J.; Stockdale, K.R.; Behlke, M.A.; Miller, F.J., Jr.; Giangrande, P.H. Rapid identification of cell-specific, internalizing RNA aptamers with bioinformatics analyses of a cell-based aptamer selection. *PLoS ONE* **2012**, *7*, e43836. [[CrossRef](#)] [[PubMed](#)]
50. Musafia, B.; Oren-Banaroya, R.; Noiman, S. Designing anti-influenza aptamers: Novel quantitative structure activity relationship approach gives insights into aptamer—Virus interaction. *PLoS ONE* **2014**, *9*, e97696. [[CrossRef](#)] [[PubMed](#)]
51. Fischer, N.O.; Tok, J.B.H.; Tarasow, T.M. Massively parallel interrogation of aptamer sequence, structure and function. *PLoS ONE* **2008**, *3*, e2720. [[CrossRef](#)] [[PubMed](#)]
52. Sanchez-Luque, F.J.; Stich, M.; Manrubia, S.; Briones, C.; Berzal-Herranz, A. Efficient hiv-1 inhibition by a 16 nt-long RNA aptamer designed by combining *in vitro* selection and *in silico* optimisation strategies. *Sci. Rep.* **2014**, *4*. [[CrossRef](#)] [[PubMed](#)]
53. Ducongé, F.; Toulmé, J.J. *In vitro* selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the tar RNA element of hiv-1. *RNA* **1999**, *5*, 1605–1614. [[CrossRef](#)] [[PubMed](#)]
54. Ditzler, M.A.; Lange, M.J.; Bose, D.; Bottoms, C.A.; Virkler, K.F.; Sawyer, A.W.; Whatley, A.S.; Spollen, W.; Givan, S.A.; Burke, D.H. High-throughput sequence analysis reveals structural diversity and improved potency among RNA inhibitors of hiv reverse transcriptase. *Nucleic Acids Res.* **2013**, *41*, 1873–1884. [[CrossRef](#)] [[PubMed](#)]
55. Whatley, A.S.; Ditzler, M.A.; Lange, M.J.; Biondi, E.; Sawyer, A.W.; Chang, J.L.; Franken, J.D.; Burke, D.H. Potent inhibition of HIV-1 reverse transcriptase and replication by nonpseudoknot, “ucaa-motif” RNA aptamers. *Mol. Ther. Nucleic Acids* **2013**, *2*, e71. [[CrossRef](#)] [[PubMed](#)]
56. Chumachenko, N.V.; Novikov, Y.; Yarus, M. Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *J. Am. Chem. Soc.* **2009**, *131*, 5257–5263. [[CrossRef](#)] [[PubMed](#)]
57. Illangasekare, M.; Yarus, M. A tiny RNA that catalyzes both aminoacyl-tRNA and peptidyl-RNA synthesis. *RNA* **1999**, *5*, 1482–1489. [[CrossRef](#)] [[PubMed](#)]
58. Firnberg, E.; Labonte, J.W.; Gray, J.J.; Ostermeier, M. A comprehensive, high-resolution map of a gene’s fitness landscape. *Mol. Biol. Evol.* **2014**, *31*, 1581–1592. [[CrossRef](#)] [[PubMed](#)]
59. Doudna, J.A.; Cech, T.R. The chemical repertoire of natural ribozymes. *Nature* **2002**, *418*, 222–228. [[CrossRef](#)] [[PubMed](#)]
60. Sharmeen, L.; Kuo, M.Y.P.; Dinner-Gottlieb, G.; Taylor, J. Antigenomic RNA of human hepatitis delta viruses can undergo self-cleavage. *J. Virol.* **1988**, *62*, 2674–2679. [[PubMed](#)]
61. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA moiety of ribonuclease p is the catalytic subunit of the enzyme. *Cell* **1983**, *35*, 849–857. [[CrossRef](#)]

62. Perrotta, A.T.; Been, M.D. A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature* **1991**, *350*, 434–436. [[CrossRef](#)] [[PubMed](#)]
63. Bartel, D.P.; Szostak, J.W. Isolation of a new ribozyme from a large pool of random sequences. *Science* **1993**, *261*, 1411–1418. [[CrossRef](#)] [[PubMed](#)]
64. Johnston, W.K.; Unrau, P.J.; Lawrence, M.S.; Glasen, M.E.; Bartel, D.P. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **2001**, *292*, 1319–1325. [[CrossRef](#)] [[PubMed](#)]
65. Zaher, H.S.; Unrau, P.J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **2007**, *13*, 1017–1026. [[CrossRef](#)] [[PubMed](#)]
66. Attwater, J.; Wochner, A.; Holliger, P. In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* **2013**, *5*, 1011–1018. [[CrossRef](#)] [[PubMed](#)]
67. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **2011**, *332*, 209–212. [[CrossRef](#)] [[PubMed](#)]
68. Winkler, W.C.; Nahvi, A.; Roth, A.; Collins, J.A.; Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **2004**, *428*, 281–286. [[CrossRef](#)] [[PubMed](#)]
69. Roth, A.; Weinberg, Z.; Chen, A.G.Y.; Kim, P.B.; Ames, T.D.; Breaker, R.R. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **2014**, *10*, 56–60. [[CrossRef](#)] [[PubMed](#)]
70. Kruger, K.; Grabowski, P.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of *tetrahymena*. *Cell* **1982**, *31*, 147–157. [[CrossRef](#)]
71. Peebles, C.L.; Perlman, P.S.; Mecklenburg, K.L.; Pertillo, M.L.; Tabor, J.H.; Jarrell, K.A.; Cheng, H.-L. A self-splicing RNA excises an intron lariat. *Cell* **1986**, *44*, 213–223. [[CrossRef](#)]
72. Saville, B.J.; Collins, R.A. A site-specific self-cleavage reaction performed by a novel RNA in *neurospora* mitochondria. *Cell* **1990**, *61*, 685–696. [[CrossRef](#)]
73. Hampel, A.; Tritz, R.R. RNA catalytic properties of the minimum (-)strsv sequences. *Biochemistry* **1989**, *28*, 4929–4933. [[CrossRef](#)] [[PubMed](#)]
74. Forster, A.C.; Symons, R.H. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active site. *Cell* **1987**, *49*, 211–220. [[CrossRef](#)]
75. Seelig, B.; Jäschke, A. A small catalytic RNA motif with diels-alderase activity. *Chem. Biol.* **1999**, *6*, 167–176. [[CrossRef](#)]
76. Eklund, E.H.; Szostak, J.W.; Bartel, D.P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **1995**, *269*, 364–370. [[CrossRef](#)] [[PubMed](#)]
77. Pitt, J.N.; Ferré-D'Amaré, A.R. Rapid construction of empirical RNA fitness landscapes. *Science* **2010**, *330*, 376–379. [[CrossRef](#)] [[PubMed](#)]
78. Pitt, J.N.; Ferré-D'Amaré, A.R. Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J. Am. Chem. Soc.* **2009**, *131*, 3532–3540. [[CrossRef](#)] [[PubMed](#)]
79. Hofacker, I.L.; Fontana, W.; Stadler, P.F.; Bonhoeffer, S.; Tacker, M.; Schuster, P. Fast folding and comparison of RNA secondary structures. *Monatshefte Chem.* **1994**, *125*, 167–188. [[CrossRef](#)]

80. Keiper, S.; Bebenroth, D.; Seelig, B.; Westhof, E.; Jaschke, A. Architecture of a diels-alderase ribozyme with a preformed catalytic pocket. *Chem. Biol.* **2004**, *11*, 1217–1227. [[CrossRef](#)] [[PubMed](#)]
81. Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Hobartner, C.; Polonskaia, A.; Phan, A.T.; Wombacher, R.; Micura, R.; *et al.* Structural basis for diels-alder ribozyme-catalyzed carbon-carbon bond formation. *Nat. Struct. Mol. Biol.* **2005**, *12*, 218–224. [[CrossRef](#)] [[PubMed](#)]
82. Beattie, T.L.; Olive, J.E.; Collins, R.A. A secondary-structure model for the self-cleaving region of the *Neurospora VS* RNA. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 4686–4690. [[CrossRef](#)] [[PubMed](#)]
83. Flinders, J.; Dieckmann, T. The solution structure of the VS ribozyme active site loop reveals a dynamic “hot-spot”. *J. Mol. Biol.* **2004**, *341*, 935–949. [[CrossRef](#)] [[PubMed](#)]
84. Kun, Á.; Mauro, S.; Szathmáry, E. Real ribozymes suggest a relaxed error threshold. *Nat. Genet.* **2005**, *37*, 1008–1011. [[CrossRef](#)] [[PubMed](#)]
85. Kun, Á.; Maurel, M.-C.; Santos, M.; Szathmáry, E. Fitness landscapes, error thresholds, and cofactors in aptamer evolution. In *The Aptamer Handbook*; Klussmann, S., Ed.; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005; pp. 54–92.
86. Campbell, D.O.; Bouchard, P.; Desjardins, G.; Legault, P. NMR structure of varkud satellite ribozyme stem-loop v in the presence of magnesium ions and localization of metal-binding sites. *Biochemistry* **2006**, *45*, 10591–10605. [[CrossRef](#)] [[PubMed](#)]
87. Campbell, D.O.; Legault, P. Nuclear magnetic resonance structure of the varkud satellite ribozyme stem-loop v RNA and magnesium-ion binding from chemical-shift mapping. *Biochemistry* **2005**, *44*, 4157–4170. [[CrossRef](#)] [[PubMed](#)]
88. Michiels, P.J.; Schouten, C.H.; Hilbers, C.W.; Heus, H.A. Structure of the ribozyme substrate hairpin of *Neurospora VS* RNA: A close look at the cleavage site. *RNA* **2000**, *6*, 1821–1832. [[CrossRef](#)] [[PubMed](#)]
89. Hoffmann, B.; Mitchell, G.T.; Gendron, P.; Major, F.; Andersen, A.A.; Collins, R.A.; Legault, P. Nmr structure of the active conformation of the varkud satellite ribozyme cleavage site. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 7003–7008. [[CrossRef](#)] [[PubMed](#)]
90. Flinders, J.; Dieckmann, T. A ph controlled conformational switch in the cleavage site of the VS ribozyme substrate RNA. *J. Mol. Biol.* **2001**, *308*, 665–679. [[CrossRef](#)] [[PubMed](#)]
91. Bouchard, P.; Lacroix-Labonté, J.; Desjardins, G.; Lampron, P.; Lisi, V.; Lemieux, S.; Major, F.; Legault, P. Role of slv in sli substrate recognition by the *Neurospora VS* ribozyme. *RNA* **2008**, *14*, 736–748. [[CrossRef](#)] [[PubMed](#)]
92. Bouchard, P.; Legault, P. Structural insights into substrate recognition by the *neurospora varkud* satellite ribozyme: Importance of u-turns at the kissing-loop junction. *Biochemistry* **2013**, *53*, 258–269. [[CrossRef](#)] [[PubMed](#)]
93. Bouchard, P.; Legault, P. A remarkably stable kissing-loop interaction defines substrate recognition by the *Neurospora Varkud* Satellite ribozyme. *RNA* **2014**, *20*, 1451–1464. [[CrossRef](#)] [[PubMed](#)]

94. Desjardins, G.; Bonneau, E.; Girard, N.; Boisbouvier, J.; Legault, P. Nmr structure of the a730 loop of the *Neurospora* VS ribozyme: Insights into the formation of the active site. *Nucleic Acids Res.* **2011**, *39*, 4427–4437. [[CrossRef](#)] [[PubMed](#)]
95. Lafontaine, D.A.; Norman, D.G.; Lilley, D.M.J. The global structure of the VS ribozyme. *EMBO J.* **2002**, *21*, 2461–2471. [[CrossRef](#)] [[PubMed](#)]
96. Rastogi, T.; Collins, R.A. Smaller, faster ribozymes reveal the catalytic core of *Neurospora* VS RNA. *J. Mol. Biol.* **1998**, *277*, 215–224. [[CrossRef](#)] [[PubMed](#)]
97. Lacroix-Labonté, J.; Girard, N.; Lemieux, S.; Legault, P. Helix-length compensation studies reveal the adaptability of the VS ribozyme architecture. *Nucleic Acids Res.* **2012**, *40*, 2284–2293. [[CrossRef](#)] [[PubMed](#)]
98. Lafontaine, D.A.; Wilson, T.J.; Norman, D.G.; Lilley, D.M.J. The a730 loop is an important component of the active site of the VS ribozyme. *J. Mol. Biol.* **2001**, *312*, 663–674. [[CrossRef](#)] [[PubMed](#)]
99. Bonneau, E.; Legault, P. Nuclear Magnetic Resonance Structure of the III–IV–V Three-Way Junction from the Varkud Satellite Ribozyme and Identification of Magnesium-Binding Sites Using Paramagnetic Relaxation Enhancement. *Biochemistry* **2014**, *53*, 6264–6275. [[CrossRef](#)] [[PubMed](#)]
100. Jakó, É.; Itzész, P.; Szenes, Á.; Kun, Á.; Szathmáry, E.; Pál, G. *In silico* detection of tRNA sequence features characteristic to aminoacyl-tRNA synthetase class membership. *Nucleic Acids Res.* **2007**, *35*, 5593–5609. [[CrossRef](#)] [[PubMed](#)]
101. Wilson, T.J.; McLeod, A.C.; Lilley, D.M.J. A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO J.* **2007**, *26*, 2489–2500. [[CrossRef](#)] [[PubMed](#)]
102. Elena, S.F.; Solé, R.V.; Sardanyés, J. Simple genomes, complex interactions: Epistasis in RNA virus. *Chaos: Interdiscip. J. Nonlinear Sci.* **2010**, *20*. [[CrossRef](#)] [[PubMed](#)]
103. Weinreich, D.M.; Watson, R.A.; Chao, L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **2005**, *59*, 1165–1174. [[CrossRef](#)] [[PubMed](#)]
104. Cabanillas, L.; Arribas, M.; Lazaro, E. Evolution at increased error rate leads to the coexistence of multiple adaptive pathways in an RNA virus. *BMC Evolut. Biol.* **2013**, *13*. [[CrossRef](#)] [[PubMed](#)]
105. Weinreich, D.M.; Delaney, N.F.; DePristo, M.A.; Hartl, D.L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **2006**, *312*, 111–114. [[CrossRef](#)] [[PubMed](#)]
106. Salverda, M.L.M.; de Visser, J.A.G.M.; Barlow, M. Natural evolution of tem-1 β -lactamase: Experimental reconstruction and clinical relevance. *FEMS Microbiol. Rev.* **2010**. [[CrossRef](#)] [[PubMed](#)]
107. De Visser, J.; Arjan, G.M.; Park, S.-C.; Krug, J. Exploring the effect of sex on empirical fitness landscapes. *Am. Nat.* **2009**, *174*, S15–S30. [[CrossRef](#)] [[PubMed](#)]
108. Hayden, E.J.; Wagner, A. Environmental change exposes beneficial epistatic interactions in a catalytic RNA. *Proc. R. Soc. Lond. B* **2012**, *279*, 3418–3425. [[CrossRef](#)] [[PubMed](#)]
109. Lehman, N.; Joyce, G.F. Evolution *in vitro*: Analysis of a lineage of ribozymes. *Curr. Biol.* **1993**, *3*, 723–734. [[CrossRef](#)]
110. Lalic, J.; Elena, S.F. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. *Heredity* **2012**, *109*, 71–77. [[CrossRef](#)] [[PubMed](#)]

111. Curtis, E.A.; Bartel, D.P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **2005**, *12*, 994–1000. [[CrossRef](#)] [[PubMed](#)]
112. Curtis, E.A.; Bartel, D.P. Synthetic shuffling and *in vitro* selection reveal the rugged adaptive fitness landscape of a kinase ribozyme. *RNA* **2013**, *19*, 1116–1128. [[CrossRef](#)] [[PubMed](#)]
113. Huynen, M.A. Exploring phenotype space through neutral evolution. *J. Mol. Evol.* **1996**, *43*, 165–169. [[CrossRef](#)] [[PubMed](#)]
114. Hayashi, Y.; Aita, T.; Toyota, H.; Husimi, Y.; Urabe, I.; Yomo, T. Experimental rugged fitness landscape in protein sequence space. *PLoS ONE* **2006**, *1*, e96. [[CrossRef](#)] [[PubMed](#)]

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).