Paulien Hogeweg
Theoretical Biology and Bioinformatics Group
Universiteit Utrecht
Padualaan 8, 3584 CH Utrecht
P.Hogeweg@bio.uu.nl

Overzichtsartikel

From bioinformatic pattern analysis to evolutionary dynamics

Ribonucleïnezuur (RNA) vervult twee verschillende rollen binnen een cel. Enerzijds heeft het molecuul de functie om informatie op te slaan (net zoals desoxyribonucleïnezuur (DNA)), anderzijds dient het als katalysator (op een manier zoals ook eiwit deze rol vervult). Bovendien kan RNA optreden als katalysator voor zijn eigen reproductie. Het molecuul kan daarom worden beschouwd als een minimaal model om de afbeelding tussen een genoom en een organisme te bestuderen. Paulien Hogeweg, hoogleraar bio-informatica aan de Universiteit Utrecht, geeft een overzicht van de theoretische inzichten die in de laatste tien jaar over dit onderwerp zijn verkregen.

The genomes of more than 50 organisms have now been fully sequenced and are available in public databases, providing a wealth of data and a great number of opportunities and challenges for fundamental bioinformatic research. Most people consider the release of the human genome (incomplete as it may be) as the milestone and since that moment the term 'post genomic era' has emerged, indicating that, once we have the genome sequence, there is still a long way to go to understand the functioning organism. The ultimate aim of bioinformatic research in the post genomic era is 'to compute the organism from its DNA sequence', or, in more technical terms, 'to compute the phenotype (i.e., the set of observable characteristics as determined by genotype and environment) from the genotype (i.e., the genetic constitution of the individual)'. Lofty, and as yet far away as this aim may be, in this paper we will try to go even one step beyond it. Beyond the 'prediction' of the phenotype from the genotype of existing entities, we aim at the understanding of which type of entities are likely to evolve, as well as which types are likely not to evolve, and the evolutionary dynamics which determines that.

Clearly, we can only make progress towards any of these goals by focusing on manageable subproblems. Thus, a much studied subproblem is the prediction of the way an RNA or a protein molecule folds into a functional shape, using the sequence information (see below).

Here we will focus on RNA. The reason that we focus on RNA and not on protein is that its possible secondary structures can be calculated relatively easily. RNA is best known for its intermediate role in producing proteins from DNA: DNA is transcribed into RNA, which is then translated into a protein. However, RNA has many more functions, as it acts as enzyme, i.e., as a catalyzer as well, both on its own and as part of RNA-protein complexes. In fact, the plethora of known functions is still increasing daily, and it includes also defense and gene regulation.

Because of these multiple roles in very fundamental processes, it is assumed that the so-called RNA world was a stage in pre-biotic evolution, in which RNA molecules served both for information storage and for catalysis, while the more specialized molecules, DNA and proteins, are later additions.

Here we will examine the properties of single RNA molecules. The enzymatic function depends primarily on the shape of the molecule. We will use the secondary structure as an approximation of the shape, and therewith of the 'phenotype' of the RNA sequence.

Thus we can, for this biologically interesting, simple special case, proceed to study the further questions outlined above. We will study the general properties of the genotype-phenotype mapping, the resulting evolutionary dynamics when, per definition, mutations occur at the level of the genotype (sequence) and the selection occurs at the level of the phenotype (secondary structure).

From genotype to phenotype

The figures 1a–d illustrate the relation between an RNA sequence and its higher order structures. The RNA sequence under consideration consists of 4 nucleotides: G,C,A, and U. The secondary structure is formed by the binding of the complementary base-pairs G-C, A-U and G-U. The figure shows three different but equivalent representations of this two dimensional intermediate. In figure 1a the RMA sequence is depicted as a string of parentheses and dots, where the opening and closing parentheses indicate a nucleotide bound to the downstream and upstream ones, respectively, and the dots show unbound bases (Konings and Hogeweg 1989). Figure 1b presents the conventional 'cloverleaf' structure, in which bound base-pairs are drawn adjacent, and in figure 1c the so-called 'mountain range representation' (Hogeweg and Hesper 1984) is shown, which retains the linear sequence at the bottom, and which represents the bonds by moving up when the nucleotide is bound to a nucleotide downstream (to the right), and down when it is bound to an upstream one. Unbound bases are given by plateaus. In figure 1d a possible tertiary structure is sketched. Although it is not

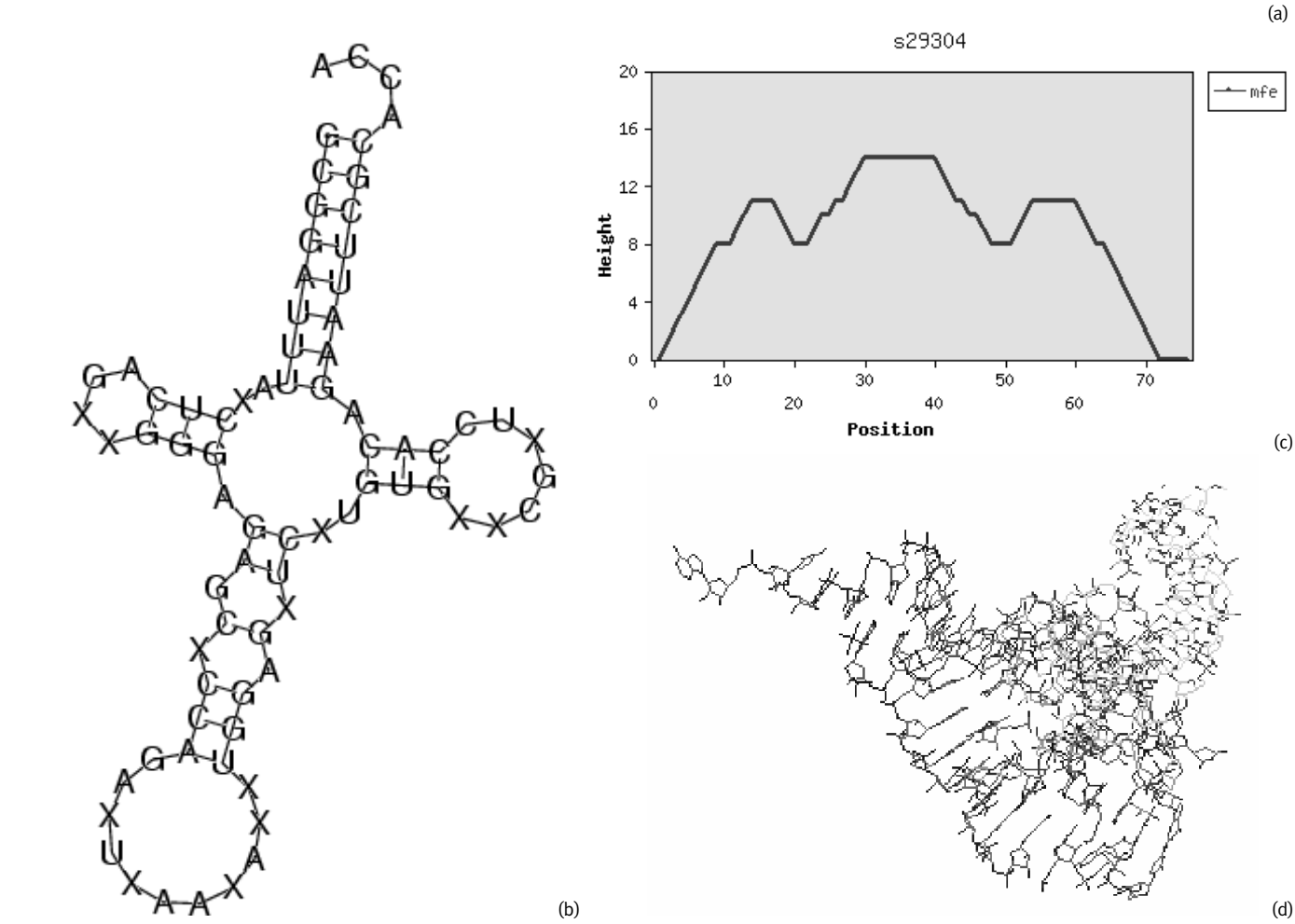


Figure 1 *The structure of $tRNA^{phe}$ of yeast.* The *primary structure* of a molecule is simply its sequence of constituting components. The building blocks of RNA are: guanine(G), adenine(A), cytosine(C), and uracil(U). In figure 1a an example of a primary structure is given in the top line. The string folds, creating new bondings between building blocks. The part of the configuration of the bondings that can be represented in the plane constitutes the so called *secondary structure* (figure 1b). Although there are strict conditions on pair forming, the number of possible secondary structures stemming from a given primary structure is huge, since the molecules are quite long. In figures 1a–c three ways to represent the secondary structure are given. The *tertiary structure* refers to the way the secondary structure folds in three dimensional space (figure 1d). It is clear that folding in 3-D adds again a large number of degrees of freedom to the system. Here, we consider only the step from primary to secondary structure and we assume that from the set of possible secondary structures the one with lowest energy is the most likely to occur. Thus, energy minimisation is used as optimization criterion.

clear at first sight, the secondary structure is mostly a true part of the tertiary structure and dominates it energetically.

Energy minimization

Our calculations are based on the observation that in nature structures with minimal energy are most likely to be formed. We calculate the free energy of a secondary structure from emperical stacking energies, i.e., the energies involved in the formation of specific base pairs.

Dynamic Programming methods

A dynamic programming (DP) algorithm for RNA secondary structure prediction was first developed by Zuker et al. (1981). The ideas behind DP are given in a separate frame. Hairpin loops and internal loops (i.e., loops with an unbound basis) — both are present in the cloverleaf of figure‘1b — are destabilizing, and they give rise to the occurrence of many relative minima in the free energy. However, an efficient algorithm that does find the minimal energy configuration, is available. We will use it as the basis of our exploration of the general properties of the RNA genotype phenotype mapping. Indeed, the availability of this algorithm spawned these investigations.

Although the secondary structures derived by the dynamic programming method are pretty good, especially for relatively short sequences, the minimal energy configuration the DP algorithm finds does not always coincide with empirically determined ones. For one thing, it does not deal with the more complex features of secondary structures as pseudo-knots (see for example figure 4). Dynamic programming extensions to include these features have been developed (Rivas and Eddy,1998) but become very cumbersome in time and space requirements, respectively ($O(N^6)$ and $O(N^4)$), and are therefore not very useful. However, also in ‘normal’ secondary structures prediction errors are rampant in long RNA’s. This can be due to mistakes in the stacking energy data (and these have indeed been modified and improved

Dynamic programming

Dynamic programming is a general method for sequential decision making. It can be used for models that are dynamic in nature or problems in which the sequential aspect is introduced in an artificial way. An example of the latter is the shortest path in a (road) network, with distances $d(x, y)$ between crossings x and y. Let $V(x, n)$ be the shortest path from any x to the destination z using n streets. Then $V(x, n+1)$ can be expressed in $V(y, n)$ in the following way:

$$V(x, n+1) = \min_y \{d(x, y) + V(y, n)\}.$$

If we take $V(x, 0) = \infty$ and $V(z, n) = 0$, then computing $V(x, n)$ for increasing n leads us to the shortest path.

A dynamic programming algorithm can also be used for finding the minimal-energy secondary structure. The energy of a structure depends on the way pairs are formed in the RNA molecule. The idea is that the minimal energy folding of a sequence can be determined from the way its subsequences can be folded. For example, the minimal energy for the sequence from position i to j can be expressed in the minimal energies of subsequences such as $(i+1, j-1)$, depending on whether i and j are paired or not. By starting the calculation with sequences of length 2 we can recursively compute the minimal energy folding for sequences of arbitrary length.

over the years), unrecognized base modifications, longer range energy contributions, or due to the fact that secondary structures occurring in vivo are not neccessarily minimal energy structures.

To try to amend such shortcomings in secondary structure predictions a number of alternative methods are developed and used.

Comparative evolutionary methods

RNA molecules in various species which share a common evolutionary ancestry (*homologous* RNA's), especially those with conserved catalytic activity, presumably fold into the same structure. Thus, when several sequences are available, the variation among those can be exploited for finding the functional structure. To do this one simply determines all energetically feasible, possibly overlapping, secondary structure elements (i.e., hairpins) of all available sequences and one selects a subset which occurs in all sequences. Moreover, so called compensatory base substitutions, where two bases are mutated such that the binding is preserved, support the functionality of hairpins, and therewith their eligibility in the consensus structure (see for example Gutell 1993). This approach is very useful when the homologs are available, and indeed the use of homology information is the predominant method in finding protein secondary structures, for which no simple dynamic programming algorithms are available. For our purposes a single sequence to secondary structure calculation is, however, needed.

Simulation methods

While, from a mathematical point of view, the dynamic programming energy minimization algorithms are far superior to stepwise stochastic energy minimization, that use local structure transitions, the latter sometimes better recover biologically functional structures. For example, Flamm et al. (2000) developed such a simulation method, with additions and deletions of single bonds, and single step 'sliding' of bonds (involving a break and a formation of an adjacent bond) as elementary steps. Using this algorithm, they have shown, for example, that in tRNA functional structure is not the energetic minimum, but that it is the structure with the largest probability of forming. Similarly they showed that the RNA structure, which can be replicated in an in-vitro evolution experiment, is not the minimal energy fold, but is the one which is formed far more frequently than the minimal energy fold.

Conclusion

Standard methods from dynamic programming (DP) for the calculation of RNA secondary structure are not perfect, and one should use in addition other approaches to study particular RNA sequences. Indeed, new methods are developed to combine, for example, comparative and DP methods. (For an overview see Zuker 2000.) Nevertheless, the DP algorithm captures the general properties of the relation between the primary and secondary structure of RNA. We will use it to uncover these general properties in the next sections.

Characterizing RNA primary to secondary structure mapping

The first thing to note about the RNA primary to secondary structure mapping is that there are many more primary structures than there are secondary structures. This is immediately clear from the "bracket representation" (figure 1a) of RNA secondary structure, which consists of only three symbols, whereas there are four nucleotides. Moreover, while there are a-priori no constraints on the ordering and frequencies of nucleotides, there are constraints on the secondary structure: all opening and closing brackets should match, stability of helices requires runs of at least two brackets, and hairpins should have a length of at least 3 long. These constraints yield $1.4848\, n^{-3/2}(1.8488)^n$ shapes for strings of length n (Schuster et al., 1994). Exhaustive calculation of all minimal energy secondary structures of sequences of length 30, consisting entirely of G's and C's, yields the numbers from the table below.

$1.07\ 10^9$	sequences
$2.18830\ 10^5$	secondary structures
$2.2718\ 10^4$	'typical structures'
93.4%	seq's in typical structure

Thus, not only do the sequences outnumber the secondary structures by far, more than 90% of the sequences fold in about 10% of the secondary structures. Moreover, it has been shown (using a 'reverse' folding algorithm) that these 10% typical shapes percolate through the whole sequence space. Thus maximally different sequences can fold into the same secondary structure, and there is a path between them of sequences which do also fold into this secondary structure, each being only 1 or 2 mutations from the next one. These properties generalize to longer sequences and to GCAU sequences: the frequency distribution appears to follow a Zipf law ($f(x) = 1.24\ (71.2 + x)^{1.23}$) , and the frequent ones together form a giant component (Schuster et al. 1994).

A general way to represent the sequence to structure mapping is to compare the difference in primary structure to the difference in secondary structure. This can be expressed for example as correlation length. A more complete representation is the representation as 'RNA landscape' (figure 2) (Fontana and Schuster 1987, Huynen et al., 1993). Here the frequency distribution of structure differences for a given sequence difference is plotted in 3D.

This landscape is obtained as follows. First, the number of point mutations is fixed. An example of a point mutation is a replacement of an A by a G base. Next, the specified number of point mutations is applied in a random fashion and the corresponding change in secondary structure is calculated. Possible measures for such a change are discussed in the next section. The number of times a specific change in secondary structure is met is plotted as the local height of the landscape.

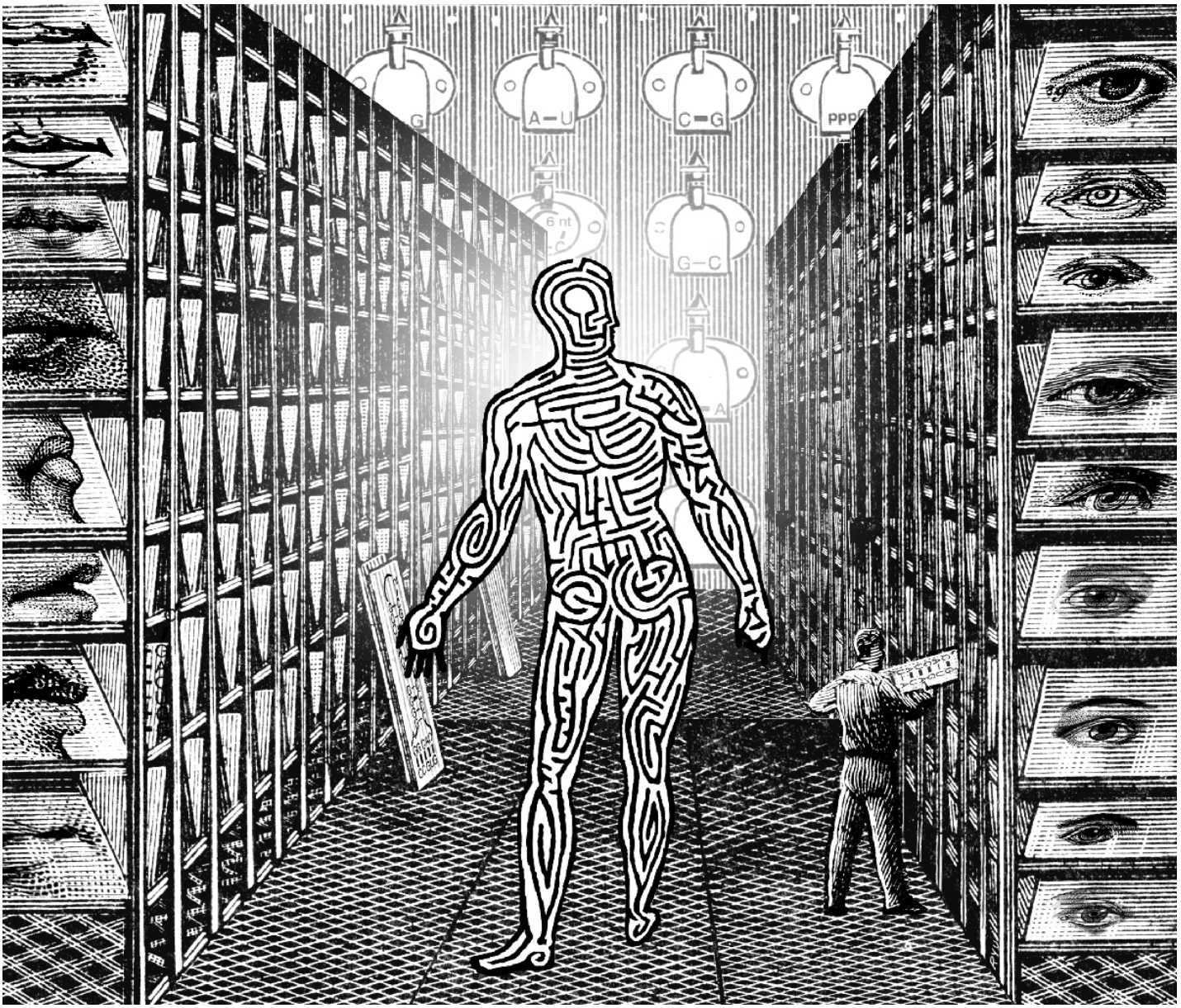
From the figure we see that (a) the correlation length is rather small: the distribution does not change after more than ca. 10 mutations and (b) although a single base change generally changes the secondary structure little or not at all, it may cause the minimal energy structure to become entirely different, as seen from the long tail of the front-most curve, which corresponds to changing only one base. Taken together, the RNA landscape and the percolating 'typical' structures indicate the following characteristics of the primary to secondary structure mapping:

- A small neighborhood in sequence space maps to all regions in structure space. Thus, given an RNA sequence, any ('typical') secondary structure is relatively 'close by' and can be reached with relatively few mutations.
- 10% of the points in structure space map to all regions in sequence space. Thus, any ('typical') RNA secondary structure, can be formed by very different sequences, and can therefore be compatible with other constraints (e.g., for coding a particular protein).

These properties seem to be 'ideal' for evolution (Schuster et al. 1994). We will study this in the next section. Here we can already answer one of the questions posed in the introduction: We expect to encounter biologically functional RNA secondary structures only in the subclass of 10% 'typical' secondary structures: others are hard to find, and hard to maintain in evolving systems.

Evolutionary dynamics on RNA landscapes

It is common wisdom that optimization on so-called rugged landscapes is difficult due to local peaks. The RNA landscapes are rugged, the



illustratie: Ryu Tajiri

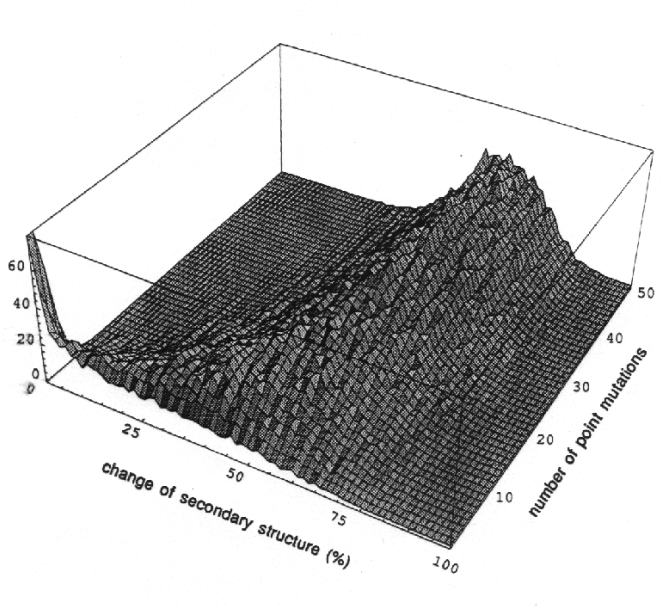


Figure 2 RNA landscape

correlation length is small as shown above. In this section we will study evolutionary optimization on this particular rugged landscape. We will show that the ruggedness is not necessarily an obstacle to optimization.

We study evolution on RNA landscapes by assuming that the 'fitness' depends on the secondary structure. We choose a certain secondary structure as the one with the highest fitness. Note that when we choose this secondary structure as the secondary structure in which an arbitrary RNA sequence folds, or when we choose it from a biological sequence database, chances are that the secondary structure is a 'typical' one as defined above. Fitness of other sequences is defined in terms of distance to the optimal structure in structure space. Distance between secondary structures can be measured in various ways, e.g. as tree distance, possibly using course graining by omitting length of stacks and loops, but also simply as the number of non-equal signs in the string representation of secondary structure.

The evolutionary dynamics is defined similar to that used for optimization in evolutionary computation. A population of RNA molecules is subjected to a replication, mutation, and selection regime. Each molecule has the same probability of replication. During replication mutations occur in the strings with a certain probability. The probability of 'death' of a molecule depends on its 'fitness'. This leads to a variable population, which on average will increase its fitness upto a point where selection and mutation balance. We assume chemostat conditions (i.e., the total RNA population is kept constant). We use only point mutations (i.e., changes of nucleotides), but similar behavior is observed when other changes like insertions, deletions and crossing over are used.

Epochal evolution and molecular clock

A typical evolutionary time course is shown in figure 3, which shows the approach towards the fittest structure from an arbitrary initial population (Huynen et al., 1996). The time course shows a pattern of 'punctuated equilibria' or 'epochal evolution'.

This implies that the system walks for some time along a so-called 'neutral path' on which the fitness is constant, although the sequence continues to change at a constant rate (i.e., is 'clocklike').

On the neutral path the walk through the space of all possible RNA molecules behaves as a diffusionlike pocess, with a diffusion coefficient depending on the number of neutral neighbors (λ), population size (N), replication rate (a), and the mutation rate per position (μ), and the sequence length (l) and can be approximated by: $D = \lambda(5al\mu/(3+4N\mu))$ (Huynen et al., 1996). When the population on a neutral path approaches a point where a higher fitness neutral path is 'near' (i.e., 1 or 2 mutations away) it can jump to the higher neutral path, and subsequently diffuses along the new path. This 'scenario' provides a unification between the concept of neutral evolution and of adaptive evolution where the former may facilitate the latter (Huynen 1996, Zuckerkandl 1997, Fontana and Schuster 1998a,b). Along the neutral path the population will continue to encounter in its neighborhood new structures, although there is a set of structures which remains in its 'shadow' all along the path.

Evolution of mutational robustness

The picture sketched above: diffusion on a network of neutral paths (also called a neutral network) and occasional shifts to higher neutral paths, would suggest that once the highest neutral paths is reached, no further evolution occurs, and that the population properties, at least on average, remain constant. Early simulation studies on RNA evolution have shown that this is not the case. In particular the 'shape' of the RNA landscape (compare figure 2) around the population changes during a long term evolutionary trajectory which remains on a certain fitness level. The average fitness of close mutants increases and so does the correlation length. (Huynen and Hogeweg 1994). The effect is small, but significant, for evolution with just point mutations, but becomes much stronger when crossovers and insertion/deletions are included.

Van Nimwegen et al. (1999) have demonstrated this effect analytically. Assuming a large fitness difference between 'on' and 'off' the neutral path, they show that the diffusion over a neutral net according to a mutation selection regime will lead to areas with relative high connectivity in the neutral network.

Note that the connectivity is a global property of the landscape, while the landscape is 'sampled' only locally by the population. It is interesting to compare this result with two well-known random walks: a blind ant, which chooses a random neighbor and only makes a step when on the neutral net, spends equal amounts of time at

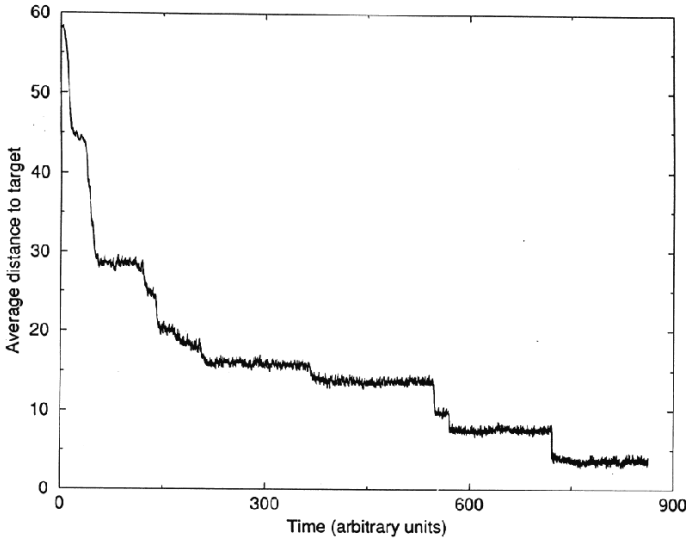


Figure 3 Punctuated evolutionary dynamics

each node on the neutral net; on the other hand, a myopic ant, which chooses a random neighbor on the neutral net, sees an average connectivity of $D = \hat{d} + Var(d)/\hat{d}$. The result depends on mutation rate and population size: it holds when their product is large enough; for very small values the connectivity 'seen' by the population remains average. They checked their analytical results with simulations on an RNA landscape, constructing the connectivity matrix by exhaustive search and showed an excellent fit.

It is a well-known experimental observation that one can obtain an increase of fitness of a population in a new environment relatively easily, but that the mutational robustness of such newly evolved population is low relative to the wild type. Bioinformatic analysis of viral RNA's reveals a larger mutational robustness than expected from random sequences (Wagner and Stadler, 1999). The above results demonstrate that this is an automatic result of evolutionary dynamics and that no explicit selection on robustness is needed (as is often assumed).

Long neutral paths and shortcuts

In a rugged fitness landscape neutral paths will 'meander'. The question arises which route will be chosen by the evolutionary dynamics, the long one along the neutral path, or a much shorter one for which a ditch should be crossed? The results mentioned above suggest that the neutral path is followed. Van Nimwegen derived the following (simplified) expression comparing the length of the neutral path which can be transferred in the time that it takes to cross a ditch of a particular width (i.e., w, number of mutations) and height (σ, fitness drop), given population size n and mutation rate μ:

$$V = \frac{n}{w!}\left(\frac{log(\sigma)}{\mu}\right)^{w-1}.$$

This result was obtained by tracing the ancestry trees of mutants in the ditch, to find the crossing probability. When the mutation rate is not too large, the result shows the following.

- A very long neutral path can be transferred in the time it takes to cross even a shallow and narrow ditch. Thus, we should indeed expect that the evolutionary dynamics typically follows the neutral path, without shortcuts.
- The width, rather than the depth of the ditch, determines the crossing time. This can be understood qualitatively by the fact that the width relates to the number of steps that the non-fit mutant should survive, while being negatively selected with a strength proportional to the depth of the ditch.

For large mutation rates the population 'perceives' the ditch not as a ditch and crosses it by diffusion as on the neutral path. From this analysis we can conclude:

- It is the structure of the neutral networks rather than the ruggedness of the landscape which determines the evolutionary dynamics. This result is shown directly by Barnett (1998) by including neutral path in a family of landscapes categorized by ruggedness (Kauffman 1993),
- Evolutionary dynamics is qualitatively different from energy minimization dynamics as used in many other optimization techniques with respect to the crossing of 'energy barriers': in case of energy minimization the probability depends on the height of the barrier whereas in evolutionary dynamics it depends primarily on the width (due to the competition with the rest of the population).

Discussion and perspectives

In this paper we used a simple, but biologically relevant example of the 'coding' of secondary RNA structure in the RNA sequence.

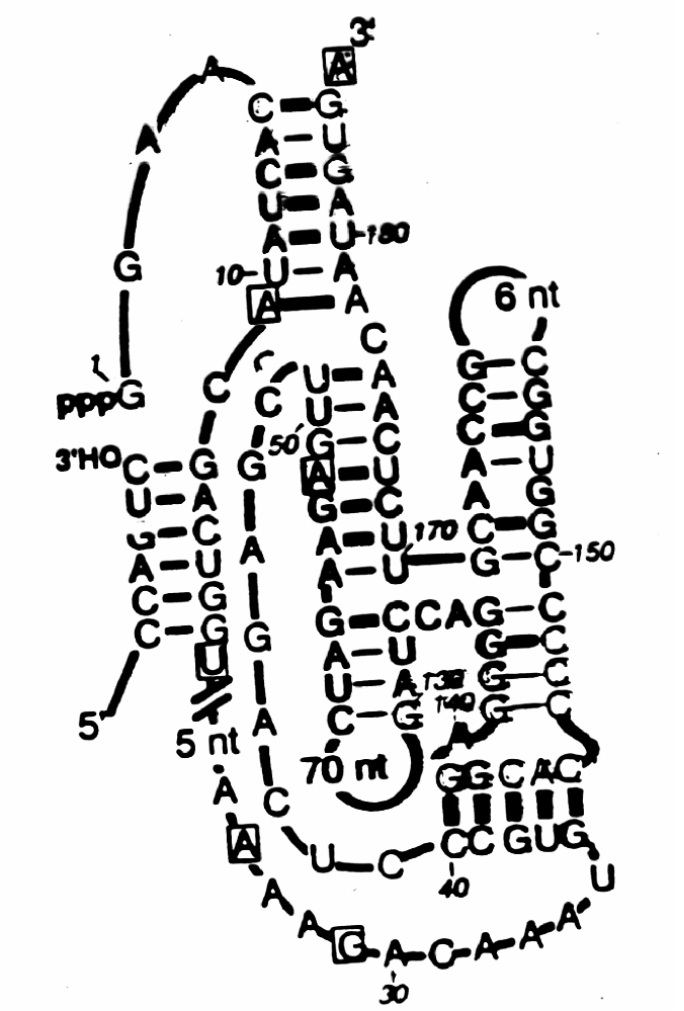


Figure 4 Evolved ligase (Ekland et al., 1995)

In studying the evolutionary dynamics on RNA landscapes, we study the relationship between a particular 'coding' structure and evolutionary optimization. The relation between coding structure and evolutionary optimization can be studied from three different perspectives, i.e.,

- given a coding structure, how is the optimization dynamics
- given a problem, how should it be coded to obtain efficient optimization
- given evolutionary dynamics what kind of coding structure will be selected?

In this paper we took the first perspective and studied the static structure of the landscape and the evolutionary dynamics on the landscape. The obtained insights in RNA evolution have been discussed above. In addition, the theory of RNA evolution gives some unexpected insights in the other perspectives.

- We have seen that a very redundant and rugged coding structure can be beneficial for evolutionary optimization, whereas reasoning from the second perspective it would seem that one should minimize the size of the search space and minimize the ruggedness.
- We have seen that, although landscapes are usually characterized by their average properties, the evolutionary dynamics exploits the variation in the landscape and so can 'choose' to a certain extent the

coding structure it encounters. We have seen that, for the evolutionary scenario considered here, i.e., evolution towards a fixed target structure, it biases the coding structure towards larger mutational robustness.

Interestingly, in-vitro evolution experiments, in which certain binding and catalytic properties of RNA are maximized, are very successful: almost any desired catalytic function (including many but not yet all which are indispensable for the putative RNA world) can be obtained easily from randomized initial sequences. In such experiments one uses large populations (circa 10^{15} molecules, but this is still vanishing small relative to all possible sequences), no mutation and only few selection steps. The success of these experiments bears out forcefully the redundancy of RNAs at the functional level. The complex nested pseudo-knot secondary structure of figure 4 is an example of a selected catalytic domain. It has 92 conserved positions, and is found in a pool of 10^{15} molecules with a random stretch of 220 nucleotides. The same experiment yielded several equally efficient catalysts, some much simpler. Interestingly, the frequency of complex structures found is much larger than expected and as compared to smaller catalytic domains: apparently catalysis is more abundant in complex structures. (Ekland et al., 1995, Sabeti et al., 1997).

We used the RNA sequence to secondary structure mapping in these investigations, as the only biological mapping which can be computed realistically. Similar results have been obtained by using rough lattice models for protein folding. Moreover, the main features discussed here, i.e., redundant genotype-phenotype mapping, neutral networks, epochal evolution, and increase of mutational robustness during long term evolution appear also to occur *mutatis mutandis* in more complex situations such as in experiments on evolution of morphogenesis through gene regulation and differential adhesion (Hogeweg 2000).

The image of evolution on a static fixed dimensional fitness landscape with simple neighborhood relations as considered here is an over-simplification. Mutations other than point mutations, e.g. duplications, insertions, and deletions change the sequence length and thus the dimensionality of the landscape, cross-overs change the neighborhood relations and through co-evolution the landscape is altered in the same time-scale as the movement of the population over the landscape. Nevertheless, the concepts developed here are useful for the analysis of such more complicated situations. For example, in the case of co-evolution we have shown that instead to evolving to 'flatter' parts of the landscape, evolution is to more rugged parts of the landscape and that insertion, deletions, and cross-overs make the bias in long term evolution more pronounced (Huynen and Hogeweg, 1994).

Biological systems are multilevel systems and the 'distance' between genotype and phenotype is much larger than the 'distance' between RNA primary and secondary structure as examined here. Both for bioinformatic pattern analysis and for bioinformatic modeling of the function and evolution of such systems the great challenge is to face the complexity, and to uncover the function/consequences of the complexity and the evolutionary dynamics which leads to it.

Acknowledgements

The theory on RNA evolution which I have reviewed in this paper has been developed over the last decade, primarily by the group of Peter Schuster with Walter Fontana and Peter Stadler (Vienna) and by my former students Martijn Huynen and Erik van Nimwegen. I thank them for sharing their interesting work with me.

References

1 Barnett, L. (1998) *Ruggedness and neutrality: The NKp family of fitness landscapes*, in C. Adami et al., eds. Artificial Life VI (MIT press) p. 18–27.

2 Ekland E.H., Szostak J.W., Bartel DP (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. Science. 1995 Jul 21;269(5222):364–70.

3 Flamm C., Fontana W., Hofacker IL, Schuster P. (2000) RNA folding at elementary step resolution. RNA. 2000 Mar;6(3):325–38.

4 Fontana W., Schuster P. (1987) A computer model of evolutionary optimization. Biophys. Chem. 1987 May 9;26(2-3):123–47.

5 Fontana W., Schuster P. (1998a) Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. J. Theor. Biol. 1998 Oct 21;194(4):491–515.

6 Fontana W., Schuster P. (1998b) Continuity in evolution: on the nature of transitions. Science. 1998 May 29;280(5368):1451–5.

7 Gutell R.R. (1993) Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. Curr. Opin. Struct. Biol. 1993, 3: 313–322.

8 Hogeweg P., Hesper B. (1984) Energy directed folding of RNA sequences. Nucleic. Acids. Res. 1984 Jan 11;12(1 Pt 1):67–74.

9 Hogeweg P. (2000) Shapes in the shadow: evolutionary dynamics of morphogenesis. Artif. Life. 2000 Winter;6(1):85–101.

10 Huynen M.A. (1996) Exploring phenotype space through neutral evolution. J. Mol. Evol. 1996 Sep;43(3):165–9.

11 Huynen M.A., Bork P. (1998) Measuring genome evolution. Proc. Natl. Acad. Sci. USA 1998 May 26;95(11):5849–56.

12 Huynen M.A., Hogeweg P. (1994) Pattern generation in molecular evolution: exploitation of the variation in RNA landscapes. J. Mol. Evol. 1994 Jul;39(1):71–9.

13 Huynen M.A., Stadler P.F., Fontana W. (1996) Smoothness within ruggedness: the role of neutrality in adaptation. Proc. Natl. Acad. Sci. USA 1996 Jan 9;93(1):397–401.

14 Huynen M.A., Konings D.A., Hogeweg P. (1993) Multiple coding and the evolutionary properties of RNA secondary structure. J. Theor. Biol. 1993 Nov 21;165(2):251–67.

15 S.A. Kauffman (1993) The origin of order. Oxford Univ. Press 709pp.

16 Kimura M. (1993) The neutral theory of molecular evolution. Cambridge Univ Press.

17 Konings D.A., Hogeweg P. (1989) Pattern analysis of RNA secondary structure similarity and consensus of minimal-energy folding. J. Mol. Biol. 1989 Jun 5;207(3):597–614.

18 van Nimwegen E., Crutchfield J.P., Huynen M.(1999) Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. USA 1999 Aug 17;96(17):9716–20.

19 van Nimwegen E. (1999) The statistical dynamics of Epochal Evolution, thesis Utrecht University.

20 van Nimwegen E., Crutchfield J.P., Melanie Mitchell M. (1999) Statistical Dynamics of the Royal Road Genetic Algorithm, Theoretical Computer Science, 1999 Nov 6;229(1–2): 41–102.

21 van Nimwegen E., Crutchfield J.P. (2000) Metastable evolutionary dynamics: crossing fitness barriers or escaping via neutral paths? Bull. Math. Biol. 2000 Sep;62(5):799–848.

22 Rivas E., Eddy S.R.(1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. J. Mol. Biol. 1999 Feb 5;285(5):2053–68.

23 Sabeti P.C., Unrau P.J., Bartel D.P. (1997) Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. Chem. Biol. 1997 Oct;4(10):767–74.

24 Schuster P., Fontana W., Stadler P.F., Hofacker I.L. (1994) From sequences to shapes and back: a case study in RNA secondary structures. Proc. R. Soc. Lond. B. Biol. Sci. 1994 Mar 22;255(1344):279-84.

25 Wagner A., Stadler P.F. (1999) Viral RNA and evolved mutational robustness. J. Exp. Zool. 1999 Aug 15;285(2):119–27.

26 Zuckerkandl E. (1997) Neutral and Nonneutral Mutations: The Creative Mix-Evolution of Complexity in Gene Interaction Systems J. Mol. Evol. 1997 Apr;44(4):470.

27 Zuker M., Stiegler P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic. Acids. Res. 1981, 9: 133–148.

28 Zuker M. (2000) Calculating nucleic acid secondary structure. Curr. Opin. Struct. Biol. 2000 Jun;10(3):303–10.