J Mol Evol (1994) 39:71–79

Journal of Molecular Evolution

Pattern Generation in Molecular Evolution: Exploitation of the Variation in RNA Landscapes

Martijn A. Huynen, Paulien Hogeweg

Bioinformatics Group, Utrecht University, Padualaan 8, 3584 CH Utrecht, Netherlands

Received: 16 June 1993 / Revised and accepted: 10 December 1993

Abstract. Evolution of RNA secondary structure is studied using simulation techniques and statistical analysis of fitness landscapes. The transition from RNA sequence to RNA secondary structure leads to fitness landscapes that have local variations in their "ruggedness." Evolution exploits these variations. In stable environments it moves the quasispecies toward relatively "flat" peaks, where not only the master sequence but also its mutants have a high fitness. In a rapidly changing environment, the situation is reversed; evolution moves the quasispecies to a region where the correlation between secondary structures of "neighboring" RNA sequences is relatively low. In selection for simple secondary structures the movement toward flat peaks leads to pattern generation in the RNA sequences. Patterns are generated at the level of polynucleotide frequencies and the distribution of purines and pyrimidines. The patterns increase the modularity of the sequence. They thereby prevent the formation of alternative secondary structures after mutations. The movement of the quasispecies toward relatively rugged parts of the landscape results in pattern generation at the level of the RNA secondary structure. The base-pairing frequency of the sequences increases. The patterns that are generated in the RNA sequences and the RNA secondary structures are not directly selected for and can be regarded as a side effect of the evolutionary dynamics of the system.

Key words: RNA secondary structure — Fitness landscape — RNA landscape — Polynucleotide frequency — Genetic algorithm — Quasispecies

Correspondence to: M.A. Huynen

Introduction

Evolution is often visualized as an uphill walk on a "surface of selective value" (Wright 1932, 1967) or on a "fitness landscape" (Kaufmann 1989). Implicit to the notion of a fitness landscape is the idea of a genotype space, upon which the fitness landscape is defined. Such a space can consist of genotypes, where each point in the space represents a combination of genes, or of individual protein (Maynard-Smith 1970) or RNA/DNA sequences (Eigen 1985), where each point in the space represents a single gene. Given the paradigm of a fitness landscape, it is the structure of the fitness landscape that determines the evolutionary process and its products. In biotic genomes the various parts of the genome have to be coadapted to produce a desired behavior, which implies a high degree of epistatic interactions. Such interactions can be both between molecules, as in gene regulation, and within molecules, as in the formation of higher-order structures within RNAs or proteins. As was shown by Kauffman and co-workers (Kauffman 1989; Weinberger 1990), a high degree of epistatic interactions leads to very irregular or "rugged" fitness landscapes—i.e., small changes in the genotype lead to relatively large changes in fitness.

As a paradigm system for fitness landscapes that involve a high degree of epistatic interactions, we study landscapes in which fitness is a function of RNA sec-

ondary structure and the genotype is the RNA sequence. In the last decades algorithms have been developed that give reasonable estimates of the (minimal free energy) secondary structure of short RNA molecules (<300). (For a review see Konings 1989.) This makes it possible to study the transition from RNA sequence to RNA secondary structure and its evolutionary properties and to study the evolution of RNA secondary structure by simulation (Fontana and Schuster 1987; Huynen et al. 1993). The transition from RNA sequence to RNA secondary structure is very sensitive to changes in the RNA sequence. Changing 10–15% of the RNA sequence gives rise to a difference between the (minimal energy) RNA secondary structures that is hardly less than that between the secondary structures of two random RNA sequences (Fontana et al. 1993; Huynen et al. 1993; Bonhoeffer et al. 1993). Small changes in genotype space (RNA sequence) lead to large changes in the phenotype (RNA secondary structure). Thus the "RNA landscape"—i.e., the landscape that describes the change of RNA secondary structure for changes in the RNA sequence—is very rugged. In a recent article we showed that secondary structures that have been selected for a high frequency of base-pairing (>70%) are relatively stable to mutations and lie thus within a relatively smooth part of the RNA landscape (Huynen et al. 1993). In these results the evolution toward smooth parts of the landscape is mainly a side effect of the specific secondary structure that is selected for. In this article we "turn the idea around"; instead of analyzing how selection for a specific secondary structure affects the local structure of the RNA landscape we investigate whether we can also have selection for a specific local structure of the RNA landscape and how this affects the RNA sequence and secondary structure. Using genetic algorithms (Holland 1975) we simulate evolution of RNA secondary structure in two types of environments—a stable environment (a fixed fitness landscape) and a very unstable environment (a constantly changing fitness landscape). We want to emphasize that the object of this study is not to model the evolution of some specific biotic RNA secondary structure or to generate results that can quantitatively be compared to biotic data, but to study in general whether evolution "exploits" differences in the local ruggedness of the RNA landscape and what types of patterns can result from that. We illustrate the relation between the exploitation of local variations in the RNA landscape with a number of specific examples; the ideas that arise from these are, however, more generally applicable. There are a number of clues which point to the possibility of selection for the local structure of landscapes. Kaneko and Ikegami (1992) have shown in a model that when the mutation frequency itself is evolvable, it decreases once a population has reached an optimum. Reducing the mutation rate after reaching an optimum in a fixed landscape increases the inclusive fitness of fit individuals (the fitness of their progeny). An alternative way of increasing this fitness is to move toward relatively flat peaks in the fitness landscape. Here not only the fittest sequence, but also its mutants, have a high fitness. This idea is tightly linked to that of a quasispecies; a species does not represent a single point on the fitness landscape but is a collection of closely related genotypes that is centered around the fittest sequence (the master sequence) (Eigen and Schuster 1979; Eigen et al. 1989). Fitness then becomes a collective property that is determined by all the sequences in the quasispecies (Schuster 1989). Evolution toward "the flattest peak" increases the fitness of the quasispecies and has indeed been observed in a double-peaked landscape; the quasispecies ends up on the lower, relatively flat peak, provided that the mutation frequency is sufficiently high (Schuster 1989). Agur and co-workers (Agur 1987; Agur and Kersberg 1987) show that the smoothness of a landscape can be regulated by modulating the genotype–phenotype relation. In their model there is, in a changing fitness landscape, selection for individuals that can change their genotype–phenotype relation in such a way that the effect of mutations on the phenotype is increased. Another example of adaptation of landscapes is given by Kauffman and Johnsen (1991), who show in a model with multiple coevolving species adaptation of the ruggedness of the fitness landscapes per species. In our approach the relation between genotype (RNA sequence) and phenotype (RNA secondary structure) is fixed; thus there is only one landscape. Variations in the local ruggedness of the RNA landscape can only be exploited through evolutionary change of the RNA sequence itself.

Methods

RNA Secondary Structure Determination. The enfold algorithm (Hogeweg and Hesper 1984) with the parameter set from Jaeger et al. (1989) is used to predict RNA secondary structure.

Comparison of RNA Secondary Structures. To compare RNA secondary structures we represent the structures as strings in which every position has a symbol depending on the direction of the base-pairing (upstream or downstream from the hairpin loop); if a base is not base-paired the symbol depends on whether or not it is in a hairpin loop (Konings and Hogeweg 1989). Dissimilarity between the strings is given by their nominal distance—that is, by the number of different symbols at corresponding positions (no alignment).

Types of Mutations (Genetic Operators). We will use the term *genetic operators* (Holland 1975) to denote operations (mutations) that change an RNA sequence. We use three kinds of genetic operators:

Point mutations are substitutions of nucleotides in the RNA string.
Insertions/deletions (indels) are made by inserting one nucleotide into the RNA string and removing one. The sites are chosen randomly and independently. One indel thus creates a shift of one nucleotide over the distance between the insertion and the deletion.
For cross-over one piece of the RNA string is substituted by another piece from another string in the population.

Evolution. Selection for a fixed secondary structure: A Genetic Algorithm (Holland 1975) was used to simulate the evolutionary search process. A population consists of 200 RNA strings (GNOMES) of 30 nucleotides. At each time-step 20 GNOMES are removed from the population. The chance of being removed is inversely proportional to the relative fitness of the GNOMES—i.e., "nonsurvival of the nonfittest." (For selection criteria see below.) Thereafter reproduction takes place. From the remaining population 20 GNOMES are randomly chosen and copied to create 20 new GNOMES. The genetic operators change the primary sequences of these new GNOMES. After point mutations or indels, (equal) cross-over between the new GNOMES takes place and the latter enter the population. The secondary structure and fitness of the newly formed GNOMES are then determined. In the initial population all GNOMES are identical, a setting that is biologically more relevant than the traditional setting for genetic algorithms in which all initial strings are chosen independently (Huynen and Hogeweg 1989). Each simulation is run for 30,000 time-steps; 25 simulations are performed for every set of genetic operators.

Evolution in a highly variable environment. The system we model is analogues to that which describes the interaction between a population of antibodies (A) and population of pathogens (B). It is, however, not meant to present any biotic system in particular; it is designed to study evolution in a highly variable environment. There are two populations: population A consists of 250 GNOMES, population B of 50. The sequence length is 50. Every time-step all GNOMES of A randomly encounter GNOMES of B. (The system is fully mixed.) The absolute fitness of a GENOME in A is the similarity of its secondary structure to that of the GNOME of B it encounters. GNOMES of A reproduce every time-step. GNOMES of B reproduce once in every five time-steps; their absolute fitness is 50 minus the maximum similarity of their secondary structure to that of the GNOMES of A they encountered in the last five time-steps. The "adaptive capacities" of population A (the chasing population) are clearly higher than those of population B (the population that is being chased). The idea behind this is that we wanted to increase the pressure in population B to change. Population sizes are constant; there is only competition within populations. Reproduction rates are constant, and fitness only affects the relative chance of death.

Selection Criteria. Selection for a single hairpin: The absolute fitness is the number of base pairs in the hairpin (up to a maximum of 12). Selection for a double hairpin: The absolute fitness is the product of the lengths of the two hairpins (up to a maximum of 5 per hairpin). The absolute fitness in the coevolving system is the absolute similarity (population A) or dissimilarity (population B) of secondary structures of GNOMES that "meet."

Frequencies of Occurrence of Genetic Operators. The frequency of occurrence of a genetic operator per newly formed string has a Poisson distribution. The mean number of point mutations per new GNOME is 1 in the selection for a fixed secondary structure and it is 0.1 in the coevolving system. It is kept relatively low in this system to keep the variation within both coevolving species low and to reduce dilution effects. (If the variation in the populations is high, GNOMES of A will encounter GNOMES of B they are not adapted to.) The mean number of indels per new GNOME is 0.25. The mean number of cross-over events per new GNOME is 1. There is no cross-over in the coevolving system.

Results

Evolution in a Stable Environment

Selection for a Simple Secondary Structure Using Different Types of Mutations

The transition from RNA sequence to RNA secondary structure shows different sensitivity to different types of mutations (genetic operators); insertions/deletions (indels) cause more change in the secondary structure than do point mutations (Huynen et al. 1993). One can imagine that secondary structures that are relatively insensitive to one genetic operator may not be relatively insensitive to another genetic operator. This implies that if sequences have evolved with one type of genetic operator, and have moved during evolution to relatively smooth regions in the landscape (flat peaks) with respect to the genetic operator they evolved with, they might still be in a relatively rugged part with respect to another genetic operator. (Notice that the genotype–phenotype/fitness-landscape metaphor is defined on the base of nominal distance or number of point mutations between the genotypes; a "landscape with respect to a genetic operator" [other than point mutations] is not a very well-defined concept and is used here in a manner of speaking.) We let sequences evolve for a simple secondary structure (a single hairpin) under different regimes of genetic operators and tested the sensitivity of their secondary structures to the different genetic operators (Fig. 1).

The results show that secondary structures that evolved with indels are less affected both by indels and by point mutations in the RNA sequence than are secondary structures that evolved with point mutations. In the following we will use the phrase "sensitivity/resistance to mutations" to denote the effect of mutations in the RNA sequence on the (minimal energy) RNA secondary structure. The results also show that the use of cross-over during evolution increases the resistance of the secondary structures to both types of mutations.

The resistance of secondary structures to mutations is accompanied by a pattern at the sequence level (Fig. 2). There is an overrepresentation of polynucleotides (>4) consisting solely of purines or of pyrimidines. This pattern is more dominant in sequences that have evolved with indels than in sequences that have evolved with point mutations and is more dominant in sequences that have evolved with cross-over than in sequences that have evolved without cross-over.

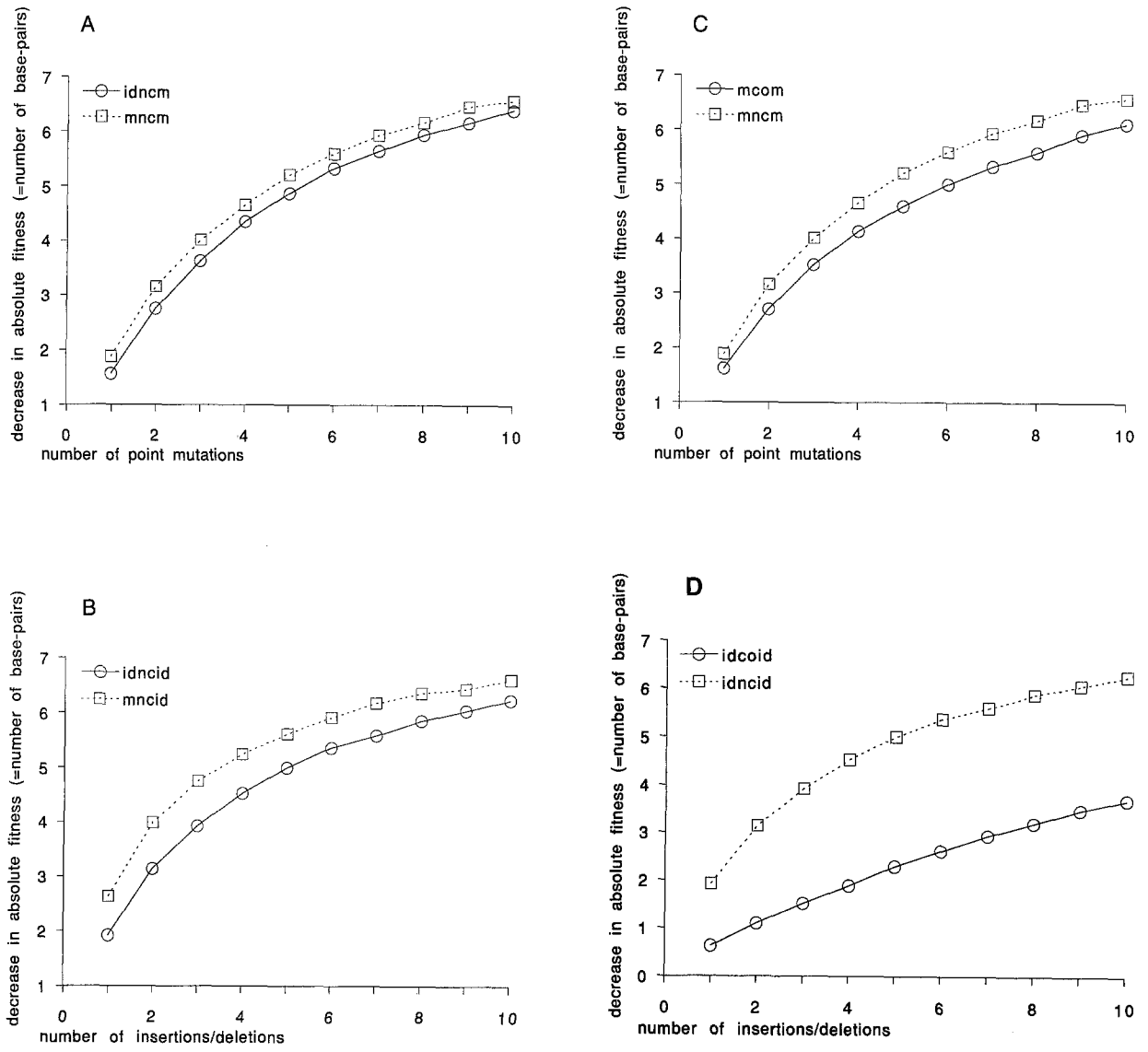

Fig. 1. Decrease in fitness (= decrease in number of base-pairs in a single hairpin) after mutation for sequences that have evolved with different combinations of genetic operators for a single hairpin structure (sequence length 30 nucleotides, maximum fitness is reached with 12 base-pairs). The sequences that have evolved with insertions/deletions (indels) are specifically resistant to indels. The use of crossover increases the resistance to mutations that occur during evolution. **A** Decrease in fitness after point mutations, for sequences that have evolved with point mutations (*mncm*) or with indels (*idncm*). **B** Decrease in fitness after indels, for sequences that have evolved with point mutations (*mncid*) or indels (*idncid*). **C** Decrease in fitness after point mutations for sequences that have evolved with point mutations and with (*mcom*) or without crossover (*mncm*). **D** Decrease in fitness after indels for sequences that have evolved with indels and with (*idcoid*) or without crossover (*idncid*).

The correlation between the pattern generation and the resistance to mutations points to a causal relation. Polynucleotides increase the modularity of the sequence and can thereby prevent large "swaps" in secondary structure after mutations. This, however, does not explain the relatively high frequency of polynucleotides in sequences that have evolved with indels. Nor does it explain the high increase in resistance to indels in these sequences (relative to the increase in resistance to point mutations). Apparently polypurines and polypyrimidines create secondary structures that are specifically robust to indels. The reason for this can be understood from the sequence level; shifting a polynucleotide of purines or pyrimidines causes little harm at the sequence level itself and will thus cause relatively little change at the secondary structure. Another reason for the extremely high frequency of polynucleotides in sequences that evolved with indels and cross-over is the combination of the genetic operators themselves; the combined action of indels and cross-over causes polynucleotides de novo and leads even in neutral evolution to an overrepresentation of polynucleotides of purines or pyrimidines. The frequencies are, however, much lower than the ones found in our simulations of selection for a single hairpin structure (data not shown).

The use of cross-over during evolution increases pat-

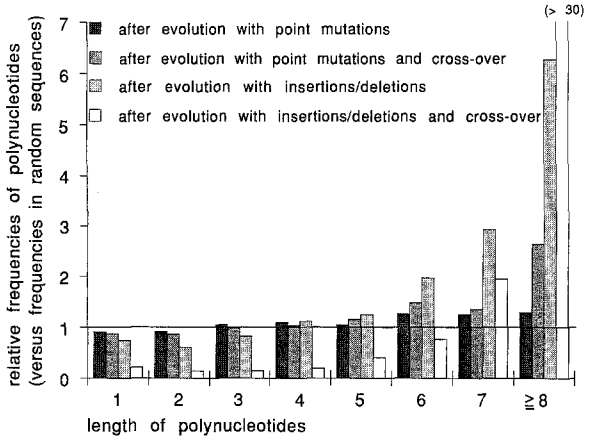

Fig. 2. Relative frequencies of polynucleotides consisting solely of purines or pyrimidines in sequences that have evolved with different combinations of genetic operators for a single hairpin structure. The expected frequency is set to 1 (= the frequency in a random sequence with the same frequency of purines and pyrimidines). Shown is the mean of 25 runs. The overrepresentation of long stretches of purines or pyrimidines is more pronounced in sequences that have evolved with crossover than in sequences that have evolved without crossover, and more pronounced in sequences that have evolved with indels that in sequences that have evolved with point mutations.

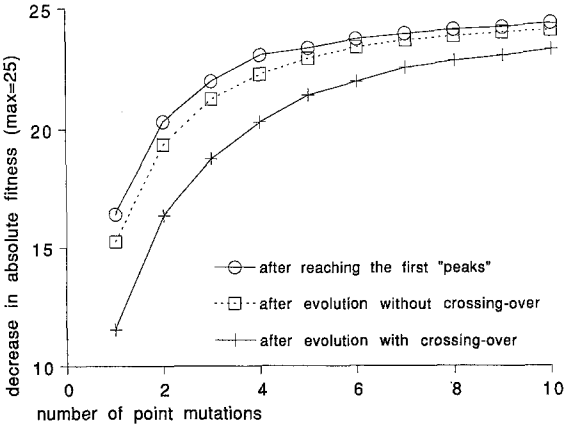

Fig. 3. Decrease in fitness after point mutations in sequences that have evolved for a double-hairpin structure (fitness is the product of the length of the two hairpins, max = 25) with point mutations and with or without crossover. The upper curve shows the effect of point mutations on the secondary structure of sequences that during the simulation have reached the required secondary structure for "the first time" (generally within 500 generations). The lower curves are of populations that have evolved for 30,000 generations. Shown is the mean of 25 runs. As is shown, the resistance to mutations increases during evolution, especially if crossover occurs.

tern generation and decreases the effect of mutations on the evolved secondary structures (Fig. 1). One can look upon cross-over as a kind of noise in the system, the source of which lies in the system itself. Thus there is not only selection for secondary structures that are insensitive to the noise, but also for harmless noise. This idea will be illustrated in the next part of the results, which is on selection for a more complicated secondary structure. The relatively high resistance to mutations in secondary structures that have evolved with cross-over implies that sequences that evolve without cross-over do not reach the flattest peaks of the landscape. Even if we use the sequences that were generated by evolution with cross-over as a starting point for evolution without cross-over, the resistance to mutations and the frequencies of the polynucleotides drop toward the level that was reached when evolution started with random sequences (data not shown). There is a paradoxical situation here; on the one hand mutations lead to the selection pressure that moves the quasispecies toward relatively smooth regions; on the other hand mutations are responsible for the genetic drift that moves the quasispecies away from the smoothest regions.

Selection for a Double-Hairpin Structure

We let sequences evolve for a double-hairpin structure. Figure 3 shows the resistance to mutations at the time when the population has adapted to the required secondary structure for the first time, and at the end of evolution (in this case 30,000 time-steps). The resistance to mutations has increased during the simulation. It appears that initially the quasispecies climbs in the fitness landscape a peak that is relatively near, and that the movement toward relatively flat peaks is more a long-term process.

During selection for a double-hairpin structure there is in the sequences a pattern generation that is associated with the development of resistance to mutations. Figure 4 shows the development of frequencies of polynucleotides in time. The pattern that has developed is different from that which has developed during selection for a single hairpin; besides an overrepresentation of long polypurines and polypyrimidines (>4), there is also an overrepresentation of monopurines and monopyrimidines. The specific pattern that is generated is shown in Fig. 5. One hairpin is made of one stretch of purines on one side and one stretch of pyrimidines on the other side, whereas the other hairpin is made of alternating purines and pyrimidines. This pattern is very efficient in preventing alternative foldings. Not only does it prevent the two hairpins from merging to one large hairpin, where the 5′ side of the 5′ hairpin interacts with the 3′ side of the 3′ hairpin, and the 3′ side of the 5′ hairpin interacts with the 5′ side of the 3′ hairpin, but it also prevents the two hairpins from merging to one small hairpin where the 3′ side of the 5′ hairpin interacts with the 3′ side of the 3′ hairpin or the 5′ side of the 5′ hairpin interacts with the 5′ side of the 3′ hairpin. The same type of pattern formation, although less pronounced, was observed with selection for two hairpins of length 8 in sequences of length 44 (data not shown).

The movement toward flat peaks reduces the difference in selection strength between mutants and should therefore give rise to more variation in the population. Indeed, if one monitors the amount of variation in the populations, it is seen to increase long after the popu-

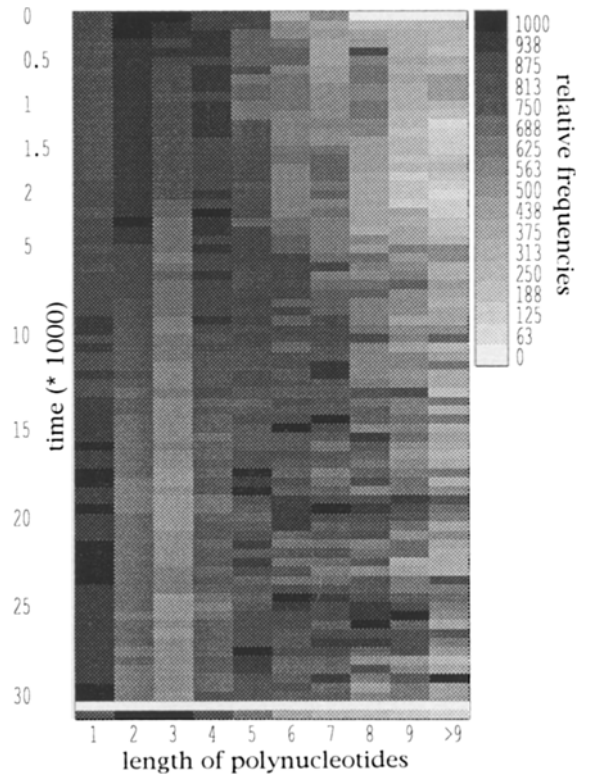

Fig. 4. Development, during evolution for a double-hairpin structure, of patterns in the frequencies of polynucleotides consisting solely of purines or pyrimidines. Shown are the relative frequencies of polynucleotides of length 1, 2, 3 . . . >9 during time for 25 runs (scaling only within columns, the maximum frequency that is reached is set to 1000). The bottom row shows expected frequencies in random sequences. The data show a clear overrepresentation of the polynucleotides of length 1 ("mono-purines" and "-pyrimidines") and of polynucleotides longer than 4; there is, however, a considerable amount of variation in the relative frequencies of the very long polynucleotides.

lations have reached the desired secondary structure (Fig. 6), and is more-or-less parallel to the pattern generation in the populations. It is interesting to note that during the development of an increase in resistance to mutations the fitness of the whole population increases only very little (Fig. 6), particularly in comparison to the decrease of the effect of mutations on the fitness of sequences with the maximum fitness (Fig. 3). The effect of the increase in resistance to mutations seems to be partly nullified by the increase in the variation in the population.

In the evolution for a relatively complicated secondary structure we see again that cross-over plays an important role in the development of patterns and in the increase of resistance to mutations. Cross-over leads to a selection at the subgenomic level. (See also Hogeweg and Hesper 1992.) In the results presented here, cross-over leads to the formation of two hairpin structures that are relatively independent of each other. The actual number of potential secondary structure interactions within the molecule is thereby decreased.

Evolution in a Rapidly Changing Environment

In order to study whether it is also possible for the quasispecies to evolve toward relative rugged parts of the landscape, we developed a system with a rapidly changing environment. The system consists of two populations of RNA strings. The fitness of strings in population A is positively dependent on the similarity between their secondary structure and the secondary structure of strings in population B, whereas the fitness of strings in population B is negatively dependent on this similarity. Population A thus "chases" population B in a Red Queen type of evolution. The situation is very like that of antibodies in an immune system (*A*) which try to match an epitope of a fast-evolving pathogen (*B*).

Because there is no specific secondary structure connected with the fitness criterion, we can compare the sensitivity of the secondary structures of sequences to mutations with that of random sequences. Figure 7 shows the relation of dissimilarity between sequences and between secondary structures for random sequences and for the sequences of population *A* and population *B*. The sensitivity to mutations in the chased population (*B*) is clearly higher than that in random sequences. Population *A* also shows a trend, although less pronounced, to move toward relatively rugged regions.

One can understand this higher sensitivity if one realizes that it is in the interest of sequences of *B* to change their secondary structure rapidly and thus stay ahead of the chasing population. The fitness landscape for *B* changes rapidly, but its changes are in a way predictable, since a secondary structure that gives a high fitness at time X will give a low fitness at a time $X + t$, so anything is better than staying at the same spot; thus there is selection for (phenotypically) rapidly evolving quasispecies. For the chasing population the situation is more complex; population *A* has no way of knowing where population *B* will move to. The sensitivity to mutations of the secondary structures in population *A* is affected mainly by their evolution toward the specific secondary structures of population *B*. With respect to models of the interaction between an immune system and a rapidly changing virus (e.g., the AIDS model of Nowak et al. 1990) these results imply that the quasispecies of the virus will move toward regions where the effect of mutations on the higher-order structure of the antigen is maximized, and thus the cross-reactivity of antibodies with antigens is minimized. Indeed, the *env* gene of lentiviruses, which codes for the protein that is located on the surface of the virus, has a relative low redundancy in its codon usage, which means that a rela-

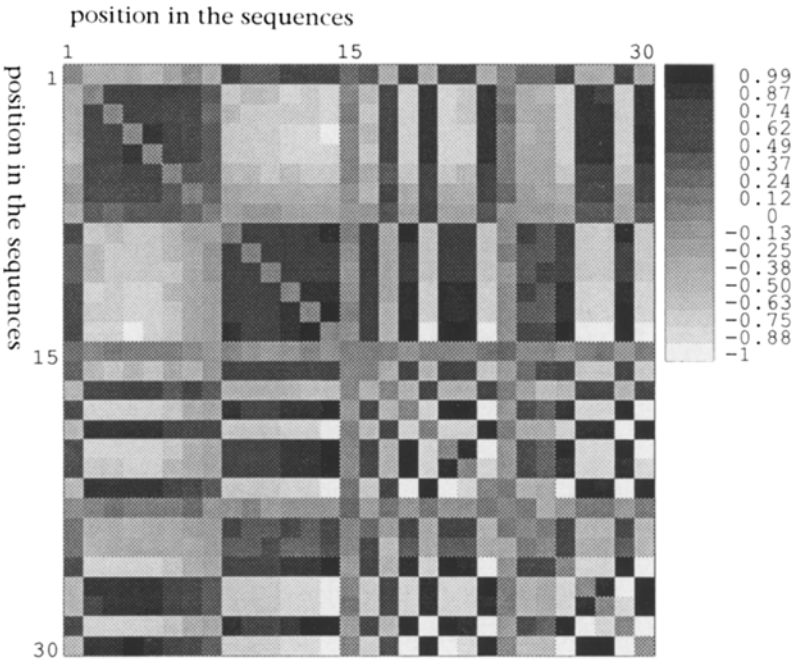

Fig. 5. The distribution of purines and pyrimidines in the sequences of one population after evolution for a double-hairpin secondary structure. Shown is the "co-occurrence" of the purines or pyrimidines between the different positions in all the sequences of one population (N = 200, the original distribution which lies between 0 and 200 is drawn between −1 and +1). The "co-occurrence" of purine contents with themselves (the diagonal) is set to zero, its gray-level serves as reference. The presentation visualizes that interaction is only possible between purines and pyrimidines, and not between purines and purines or between pyrimidines and pyrimidines. White regions can only interact with black regions, and regions with alternating black and white can only interact with other regions with alternating black and white. Thus, the only way the presented structure can possibly fold, even after a few mutations, is by forming two unconnected hairpins.

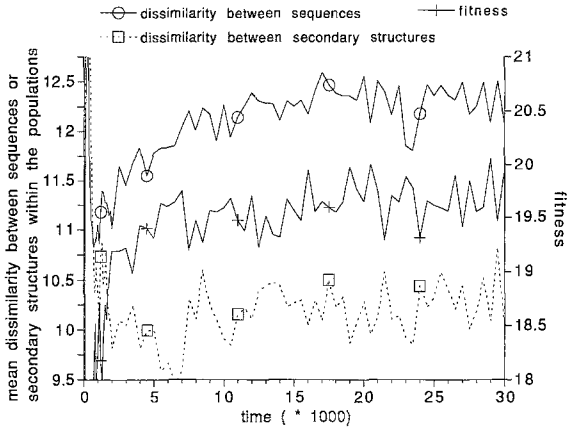

Fig. 6. The amount of variation within the populations, both genotypic and phenotypic, and the mean fitness during evolution for a double-hairpin structure. The variation shown is the mean difference between two sequences per population. As the data show, the genotypic variation increases long after the desired secondary structures are reached. (All 25 populations reach the required secondary structure within 2000 generations, more than 85% within 500 generations.)

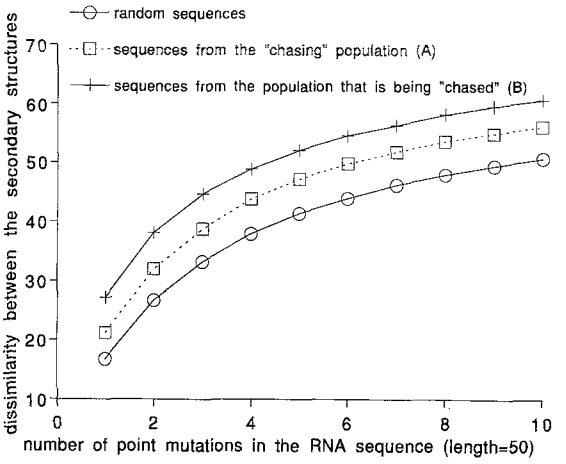

Fig. 7. The effect of point mutations on the secondary structure of random sequences (equal probabilities for all nucleotides in all positions) and of sequences in a co-evolving system after 5000 time-steps. The fitness criterion for sequences in population A is the similarity of their secondary structure to the secondary structure of sequences in population B, whereas for population B the fitness depends on the dissimilarity of its secondary structures to those of A (population A thus "chases" population B in a Red Queen-like fashion). Both populations show an increased effect of point mutations on their secondary structures; the effect, however, is strongest in population B.

tive high frequency of mutations on its RNA sequence will affect the amino-acid sequence (Huynen and Hogeweg 1989).

The evolution toward relatively rugged parts of the landscape leads to pattern generation at the RNA secondary structure level. That is, in order to increase the effect of mutations, the amount of secondary structure (= frequency of base-pairing) is increased. In random sequences of length 50 the mean frequency of base-pairing is 34%. In the population that is chased (*B*) it goes up to an average of 53%, whereas in the chasing population (*A*) it becomes 42%. One might expect such an increase in base-pairing frequency to lead to more stable secondary structures instead of the less stable ones observed here. This effect, however, only appears to play a role with very high frequencies of base-pairing (>70%) in long RNA sequences (150 nucleotides) (Huynen et al. 1993).

The qualitative results reported here (an increase in the effect of mutations on the secondary structure and pattern generation in the secondary structure) do not depend on the specific parametrization of the model and were observed for a wide array of models, including models with population dynamics or a spatial dimension (Huynen 1993).

Discussion

Patterns in sequence data are generally attributed to direct selection pressures, historical accidents (pseudogenes), and mutation pressures/random drift. According to the results of our simulations, patterns in sequence data can also be a side effect of the evolutionary dynamics, which in this case exploit the variations in the ruggedness of a fitness landscape. Another example of how the evolutionary dynamics can generate patterns in sequence data is the evolution of multiple coding as a side effect of cross-over and high maturation rates (Hogeweg and Hesper 1992).

By evolving toward relatively smooth or rugged areas of the RNA landscape the quasispecies actually increases or decreases the occurrence of phenotype affecting mutations. Kaneko and Ikegami (1992; Ikegami and Kaneko, 1992) have studied models where the mutation frequency itself was part of the genotype, and subject to mutation and selection. They observed a decrease of the mutation frequency in fixed fitness landscapes after the population had reached an optimum, whereas the mutation frequency was maintained at a high level in a constantly changing fitness landscape. This is analogous to our results on varying the local ruggedness of the landscape if one only considers mutations that affect the phenotype. From a biological point of view an important difference between varying mutations rates and varying the (local) ruggedness of the landscape is that mechanisms for varying mutation rates generally affect large parts of the genome, although they do not necessarily affect the whole genome (e.g., local somatic hypermutation regions in B cells—French et al. 1989), whereas the ruggedness of the genotype–phenotype transition in different parts of the genome can be varied locally. However, an advantage of varying mutation rates is that mechanisms for this can be turned on or off on an evolutionary relatively short time scale. Both mechanisms can of course coexist, which allows for fine tuning of the phenotype-affecting mutation rates in the whole genome.

An important question is whether these results also hold for genotype–phenotype landscapes that are based on a relation other than the RNA sequence–RNA secondary structure relation. Basically the results depend on whether there are variations in the local ruggedness of the phenotype/fitness landscape. An obvious example of another genotype–phenotype relation is the relation between protein sequence and protein higher-order structure. Because in proteins there is a greater choice of building blocks which can be used to form higher-order structures (20 amino acids vs four nucleotides) we expect to find at least as much variation in the correlation of protein landscapes as we find in RNA landscapes.

It has often been argued that selection does not operate at a level higher than that of the individual. This precludes selection for properties that affect the evolutionary process itself. Here we observe selection for the effect of mutations on the phenotype, which is possible because the selection acts at the level of the quasispecies (Eigen and Schuster 1979; Schuster 1989). This type of selection can be referred to as metaselection—that is, selection for properties that affect the "evolvability." In this sense it resembles one of the explanations for the evolution of sex (for a review see Hamilton et al. 1990)—namely, that sex evolved because it facilitates evolutionary adaptation of hosts to resist parasites.

Acknowledgments. We thank S.M. McNab for linguistic advice and Danielle Konings for comments on the original manuscript. The investigations were supported by the Foundation for Biological Research (BION), which is subsidized by the Netherlands Organization for Scientific Research.

References

Agur Z (1987) Resilience and variability in pathogens and hosts. IMA J Math Appl Med Biol 4:295–307

Agur Z, Kersberg M (1987) The emergence of phenotypic novelties through progressive change. Am Natur 129:862–875

Bonhoeffer S, McCaskill JS, Stadler PF, Schuster P (1993) RNA multi-structure landscapes. Eur Biophys J 22:13–24

Eigen M (1985) Macromolecular evolution: dynamical ordering in sequence space. In: Pines D (ed) Emerging synthesis in science. Santa Fe Institute Studies in the Sciences of Complexity. Addison Wesley, Reading, MA, vol I, pp 25–69

Eigen M, McCaskill JS, Schuster P (1989) The molecular quasispecies. Adv Chem Phys 75:149–263

Eigen M, Schuster P (1979) The hypercycle: a principle of natural self-organization. Springer, Berlin

Fontana W, Konings DAM, Stadler PF, Schuster P (1993) Statistics of RNA secondary structure. Biopolymers 33:1389–1404

Fontana W, Schuster P (1987) A computer model for evolutionary optimization. Biophys Chem 26:123–147

French D, Laskov R, Scharff M (1989) The role of somatic hypermutation in generation of antibody diversity. Science 244:1152–1157

Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as ad adaptation to resist parasites (a review). Proc Natl Acad Sci USA 87:3566–3573

Hogeweg P, Hesper B (1984) Energy directed folding of RNA sequences. Nucleic Acids Res 12:67–74

Hogeweg P, Hesper B (1992) Evolutionary dynamics and the coding structure of sequences; multiple coding as a consequence of crossover and high mutation rates. Comput Chem 16:171–182

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor

Huynen MA (1993) Evolutionary dynamics and pattern generation in

the sequence and secondary structure of RNA: a bioinformatic approach. PhD thesis, University of Utrecht

Huynen MA, Hogeweg P (1989) Genetic algorithms and information accumulation during the evolution of gene regulation. In: Schaffer JD (ed) Proceedings of the Third International Conference on Genetic Algorithms. Morgan Kaufmann, San Mateo, pp 225–230

Huynen MA, Konings DAM, Hogeweg P (1993) Multiple coding and the evolutionary properties of RNA secondary structure. J Theor Biol 165:251–267

Ikegami T, Kaneko K (1992) Evolution of host-parasitoid network through homeochaotic dynamics. CHAOS 2:397–407

Jaeger JA, Turner DH, Zuker M (1989) Improved predictions of secondary structures for RNA. Proc Natl Acad Sci USA 86:7706–7710

Kaneko K, Ikegami T (1992) Homeochaos: dynamic stability of a symbiotic network with population dynamics and evolving mutation rates. Physica D 56:406–429

Kauffman SA (1989) Adaptation on rugged fitness Landscapes. In: Stein DL (ed) Lectures in the sciences of complexity, vol 1. Addison Wesley, New York, pp 619–712

Kauffman SA, Johnsen S (1991) Co-evolution to the edge of chaos: coupled fitness landscape, poised states, and co-evolutionary avalanches. In: Langton CG, Taylor C, Farmer JD, Rasmussen S (eds) Artificial life II. Addison Wesley, New York, pp 325–369

Konings DAM (1989) Pattern analysis of RNA secondary structure: a bioinformatic approach. PhD thesis, Utrecht University

Konings DAM, Hogeweg P (1989) Pattern analysis of RNA secondary structure. J Mol Biol 207:597–614

Maynard-Smith J (1970) Natural selection and the concept of a protein space. Nature 255:563–566

Nowak MA, May RM, Anderson RM (1990) The evolutionary dynamics of HIV-1 quasispecies and the development of immunodeficiency disease. AIDS 4:1095–1103

Schuster P (1989) Optimization of RNA structure and properties. In: Perelson AS, Kauffman SA (eds) Molecular evolution on rugged landscapes: proteins, RNA and the immune system. Addison Wesley, Redwood City, pp 47–71

Weinberger E (1990) Correlated and uncorrelated fitness landscapes and how to tell the difference. Biol Cybern 63:325–336

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the Sixth International Congress on Genetics 1, pp 356–366

Wright S (1967) Surfaces of selective value. Proc Natl Acad Sci U S A 58:165–172