LA-UR- 97-228

CONF-970937--1

Title: EVOLUTIONARY DYNAMICS ON RANDOM STRUCTURES

Author(s): SIMON M. FRASER
CHRISTIAN M. REIDYS

Submitted to: OPTIMIZATION AND SIMULATIONS CONFERENCE
SINGAPORE
SEPTEMBER 1-4, 1997

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

Los Alamos
NATIONAL LABORATORY

Form No. 836 R5
ST 2629 10/91

Evolutionary Dynamics on Random Structures

By

Simon M. Fraser[b] and Christian M. Reidys[a,b]

[b]Santa Fe Institute
1399 Hyde Park Rd., Santa Fe, NM 87501, USA

[a] Los Alamos National Laboratory
TSA/DO-SA, 87548 New Mexico

*Mailing Address:
[b]Santa Fe Institute
1399 Hyde Park Rd., Santa Fe, NM 87501, USA
Phone: **(505) 984-8800 Fax: **(505) 982-0565
E-Mail: {smfr, duck}@santafe.edu

Abstract:

In this paper we consider the evolutionary dynamics of populations of sequences, under a process of selection at the phenotypic level of structures. We use a simple graph-theoretic representation of structures which captures well the properties of the mapping between RNA sequences and their molecular structure. Each sequence is assigned to a structure by means of a sequence-to-structure mapping. We will make the basic assumption that every fitness landscape can be factorized through the structures. The set of all sequences that map into a particular random structure can then be modeled as a random graph in sequence space, the so-called neutral network. We analyze in detail how an evolving population searches for new structures, in particular how they switch from one neutral network to another. We verify that that transitions occur directly between neutral networks, and study the effects of different population sizes and the influence of the relatedness of the structures on these transitions. In fitness landscapes where several structures exhibit high fitness, we then study evolutionary paths on the structural level taken by the population during its search. We present a new way of expressing structural similarities which are shown to have relevant implications for the time evolution of the population.

1 Introduction

An understanding of the mapping between genotypes and phenotypes is of central importance in evolutionary theory, as well as bearing on biologically-inspired computational optimization techniques. For real organisms, the properties of this mapping are almost completely unknown, although, for the simple paradigmatic example of the mapping of RNA sequences into secondary structures, the surprising properties of this mapping are being elucidated [7]. In this more restricted case, the properties of this mapping are relevant for the understanding of evolutionary optimization of biopolymers, and the theory of molecular evolution [8], and indicate why the folding of molecular sequences into their spatial structures is of central interest in biophysics [9].

In this paper we use an even more coarse-grained representation of biomolecules as "random structures", and study the dynamics of populations of sequences which replicate according to their fitness. The dynamics of such populations of sequences replicating with mutation is of course closely related to the underlying fitness landscape. In this paper we will make the basic assumption that each fitness landscape can be factorized through a set of such "structures", i.e. that there exists a unique mapping that assigns fitness values to structures. Therefore all sequences mapping into a particular structure have equal fitness, and the preimage of a structure (i.e. the set of all sequences that are mapped into that structure) forms a so-called neutral network in sequence space [7], upon which the sequences move by point mutations while still mapping into the same structure.

Let us first describe our basic framework: sequence space is a graph $\mathcal{Q}_\alpha^n$ (where α is the size of the alphabet of which sequences are composed, i.e. 4 for the **A**, **U**, **G**, **C** alphabet of RNA sequences, and n is the sequence length), its vertices are n-tuples $V = (x_1, \ldots, x_n)$ and any two sequences are adjacent if they differ in exactly one base. The term "structure", can reflect various levels of coarse graining. Here we will consider "structure" to consist of a list of all pairs of coordinates of the sequence that are joined by means of chemical bonds (the underlying *contact graph*), and a multi-set of relations on the edges of the graph which specify the base-pairing rules (for example, the allowable Watson-Crick base pairs) [5, 6]. For each structure there exists a set of compatible sequences, that is the set of all sequences whose coordinates fulfill the relations imposed by the edges of the contact graph of the structure. The *preimage* of the structure is then the subset of this set of compatible sequences, containing sequences which are mapped into the structure by the mapping algorithm. The preimage thus forms a random graph in sequences space, the neutral network. By choosing an ordering among the set of structures we then obtain a mapping from sequences into structures simply by iterating this random process.

One important question, then, is how neutral networks are embedded in sequence space. It has been shown [5, 6] that the graph structure of the union of the contact graphs of two structures is of particular relevance. For example it encodes the coupling of the corresponding two neutral networks in terms of the time in which a population can switch

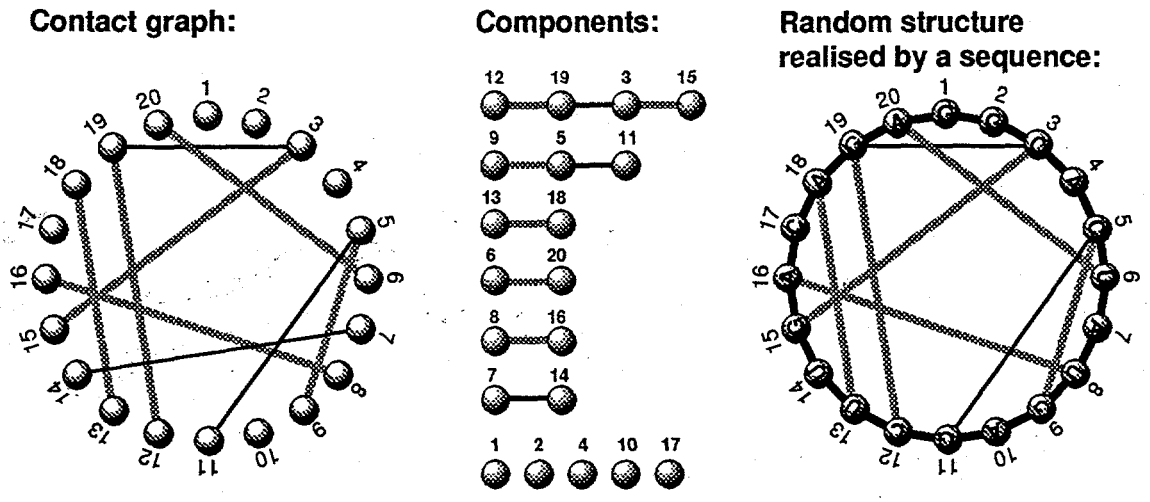

Figure 1: *A contact graph, consisting of an ordered set of vertices (numbered), between which there can be either secondary (gray) or tertiary (thin black) edges, together with its set of components. On the right hand side, the bases of one compatible sequence are shown on the vertices, indicating that, in the random structure, certain relations associated with the edges have to be fulfilled (in this case Watson–Crick base-pairing rules).*

from one net to the other. This graph contains information how close the neutral networks come in sequence space and how likely it is that bi-compatible sequences, i.e. sequences that are simulataneously compatible to both structures, will exist. For random structures, we can determine precisely the structure of this union graph, the properties of which show dramatic changes with increasing proportions of tertiary interactions in the constituent structures [1]. More specifically, it exhibits a phase transition reflected in the sudden emergence of a giant component in the graph above a certain threshold for this proportion. Below the critical value, the neutral networks of any two random structures come very close in sequence space, and it is of central interest how populations perform the transition from one network to the other.

This transition phenomenon clearly depends on the population size, the random structures and the mutation rate per replication event. It is also of interest whether these transition happen directly or via the 1 or 2-mutant hull around the neutral nets. Clearly, if a fraction of the population has realized a reasonably fit structure, sequences on this net will produce relatively many offspring. Because of mutations, some proportion of these offspring will

be mapped into other structures, and in a sense "fall off" the neutral network of the fitter structure. This scenario allows populations to search for better structures with their variant offspring while keeping the mean fitness high. Once a neutral net is found that corresponds to a fitter structure, the population performs a transition to this network.

The paper is structured as follows: first we review the concepts of random structures, compatible sequences and neutral networks. Second, we analyze transitions between random structures of equal fitness. Third, we study the ease with which transitions occur when there is some structural similarity between the underlying contact graphs.

2 Random structures, compatible sequences and neutral networks

Let us first review some terminology of graph theory. A *graph* X consists of a tuple (vX, eX) and a map $o \otimes t : eX \longrightarrow vX \times vX$. vX is called the *vertex set* and eX the *edge set.* An element $P \in X$ is called a *vertex* of X; an element $y \in X$ is called an *edge.* The vertex $o(y)$ is called the *origin* of y and the vertex $t(y)$ is called the *terminus* of y; $o(y), t(y)$ are called the *extremities* of some edge y. There is an obvious notion of Y being a *subgraph* of X. We call a subgraph Y *induced*, if for any $P, P' \in Y$ being extremities of an edge $y \in X$, it follows $y \in Y$. A *path* in X is a sequence $(Q_1, y_1, Q_2, y_2, \ldots, y_n, Q_{n+1})$, where $Q_i \in X$, $y_i \in X$, $o(y_i) = Q_i$ and $t(y_i) = Q_{i+1}$. A path such that $Q_1 = Q_{n+1}$ is called a *cycle.* X is called *connected* if any two vertices are vertices of a path of X. A connected graph without cycles is called a *tree.* Being connected is an equivalence relation in X, and the maximal connected subsets of vertices are called *components* of X.

Let m be the number of secondary bonds, and c_2 the fraction of nucleotides involved in tertiary bonds. Then the contact graph is a graph on the n indices of the coordinates of a sequence, whose edge set is the union of the edge sets of two random graphs. The first one is a 1-regular graph on $2m$ vertices, obtained by picking m pairs of indices without replacement. The second one is a random graph on n vertices, obtained by selecting the

remaining $\binom{n}{2} - m$ edges with independent probability $p = \frac{c_2}{n}$. Here, c_2 is the probability of a specific nucleotide being involved in a tertiary interaction. It has been shown in [5] that in the limit of long sequences almost all vertices of the random contact graphs are contained in tree components of logarithmic size (relative to sequence length).

A *random structure*, s_n, on n nucleotides of a finite alphabet $\mathcal{A}$ consists of:

- a contact graph $X_1 \otimes X_2$
- a family of symmetric relations $(\mathcal{R}_*, \mathcal{R}_y)_{y \in X_2}$, where $\mathcal{R}_*, \mathcal{R}_y \subset \mathcal{A} \times \mathcal{A}$.

Each $\mathcal{R}_y$ is supposed to have the property: for all $a \in \mathcal{A}$ there exists at least one $b \in \mathcal{A}$ with the property: $a\mathcal{R}_y b$. The relation $\mathcal{R}_*$ is motivated by Watson-Crick *base-pairing rules* observed in RNA secondary-structures. For $y \in X_2$ the relation $\mathcal{R}_y$ corresponds to a specific (tertiary) interaction rule that might be context dependent.

A number of asymptotic results on contact graphs and their union has been proven in [5]. As is typical for random graph results, these hold in the limit of long sequences. First, an upper bound for the expected number of paths of length ℓ in contact graphs is given by

$$n[c_1 + c_2]^\ell . \tag{1}$$

This implies that the contact graphs decompose with probability 1 into small components. In contact graphs the largest component is at most of the order $C \ln(n)$ with constant $C > 0$ for $c_2 < 0.25$. Next we turn to the structure of the union graph. Following [5] there is a phase transition in $(X_1 \cup X_1') \otimes (X_2 \cup X_2')$ concerning the emergence of a largest component of order $C'n$ with constant $C' > 0$. The latter transition is expressed in terms of c_2, the average number of tertiary interactions per nucleotide and c_1, the fraction of secondary interactions. The exact result reads that for small values of c_2 the components of the union graph have sizes bounded by $C \ln(n)$ and, in the limit of large sequence length, for c_1, c_2 such that

$$8c_1[2 - c_1]c_2 > 1 , \tag{2}$$

there exists a unique large component of size Kn, $K > 0$, with probability tending to 1.

A vertex (sequence) $V \in \mathcal{Q}_\alpha^n$ is called *compatible* to s_n if and only if

- for all bonds y of the partial 1-factor graph X_1 its nucleotides indexed by the extremities $\{o(y), t(y)\}$ have the property $P_{o(y)}\mathcal{R}_* P_{t(y)}$ (note that since $\mathcal{R}_*$ is symmetric we also have $P_{t(y)}\mathcal{R}_* P_{o(y)}$)
- its nucleotides fulfill for all tertiary bonds $y \in X_2$: $P_{o(y)}\mathcal{R}_y P_{t(y)}$.

The set of compatible vertices with respect to the random structure s_n is called $C(s_n)$. Suppose a random structure s_n, is fixed. Then its preimage is necessarily contained in $C(s_n)$. The contact-graph induces a partition of the indices $\{1, \ldots, n\}$ into its components. Accordingly, we can regroup the indices of the nucleotides of a compatible sequence into the components of the contact graph. Formally we can now consider each multi-set $(P_{i_1}, \ldots, P_{i_k})$, consisting of nucleotides whose indices belong to a component of the contact graph, to be an element of a new alphabet, $\mathcal{A}_k$. Accordingly, we can rewrite a compatible sequence as $(A_{i_1}, \ldots, A_{i_\ell})$ (ℓ being the number of components of the contact graph). In general the set of compatible sequences is the vertex set of $\prod_{i=1}^{h} \mathcal{Q}_{\alpha_i}^{n_i}$ where $\sum_i i \cdot n_i = n$. $\alpha_i = |\mathcal{A}_i|$, h is the number of components, and n_i the length of the i-th component of the contact graph. Next we construct the preimage of the random structure s_n. It will be a random induced graph by selecting the vertices in each factor $\mathcal{Q}_{\alpha_i}^{n_i}$ with independent probability λ_i. Note that "vertex" here corresponds to a multi-set $(P_{i_1}, \ldots, P_{i_k})$ consisting of nucleotides whose indices belong to a component of the contact graph of s_n. In this sense "vertex" can be viewed as a certain segment of the sequence. λ_i (i being the index of a component) can be interpreted as the stability of the random structure with respect to a mutation that has (i) occurred in the i-th component and that has (ii) led to a compatible sequence. To summarize, the preimage of a random structure is obtained by selecting certain segments of sequences in $\mathcal{Q}_\alpha^n$ at random. For this process the mathematical structure of randomly induced subgraphs of generalized n-cubes is of particular relevance. It has been shown [7] that $\lambda^* = 1 - \sqrt[\alpha-1]{\alpha^{-1}}$ is a threshold value for density

and connectivity. These results suggest that the preimage of a random structure consists above the threshold of large connected subgraphs in sequence space. In order to define a complete mapping into random structures by the above procedure we only have to iterate the above process. We obtain mappings $f : \mathcal{Q}_\alpha^n \longrightarrow \{s_n\}$ by constructing the corresponding preimages as random graphs as follows: we fix a mapping $r : \{s_n\} \to \mathbb{N}$ having the property $j \leq i \implies r(s_j) \geq r(s_i)$ and set

$$f_r^{-1}(s_0) = \Gamma_n[s_0] \quad f_r^{-1}(s_i) = \Gamma_n[s_i] \setminus \bigcup_{j<i} [\Gamma_n[s_i] \cap \Gamma_n[s_j]] \, .$$

3 Dynamics on random structures

3.1 The basic replication scheme

In this section we will study the time evolution of finite *populations* $\mathbf{V}$ of sequences which are replicated with a probability p of mutation at each nucleotide. The replication event itself is a point process; more precisely a birth-death process in which sequences are chosen for replication with respect to their fitness, and randomly for deletion. A population $\mathbf{V}$, of size N, is a (finite) multi-set of sequences $(V_i \,|\, i \in N)$ where $\{V_i \,|\, i \in N\} \subset \mathcal{Q}_\alpha^n$ and $N > 1$. The theory of point processes provides a powerful tool by identifying $(V_i \,|\, i \in N)$ with an integer valued measure $\phi : \mathcal{Q}_\alpha^n \to \mathbb{R}$,

$$\mathbf{V} = (V_i \,|\, i \in N) \quad \longleftrightarrow \quad \phi = \sum_{i=1}^{N} g_{V_i}, \text{ where } \quad g_{V_i}(v) = \begin{cases} 1 & \text{for } V \neq V_i \\ 0 & \text{otherwise} \, . \end{cases}$$

We call the set of sequences where ϕ is nonzero the *support* of ϕ. Clearly, the *restriction* of ϕ to a subgraph $Y < \mathcal{Q}_\alpha^n$ corresponds to considering subpopulations on the vertices of Y. Note that $\phi(f^{-1}(s))$ is the number of elements of $\mathbf{V}$ contained in $f^{-1}(s)$. The time evolution of ϕ is then obtained by a mapping from $(V_i \,|\, i \in N)$ to the family $(V_i' \,|\, i \in N)$ as follows: we select an ordered pair (V_l, V_k) where $V_l, V_k \in \{V_i \,|\, i \in N\}$. For this purpose let $\mathrm{res}_s\phi$ be the restriction of ϕ to all sequences that are mapped into s. Clearly the subpopulation that corresponds to $\mathrm{res}_s\phi$ consists of sequences all having fitness $f(s)$.

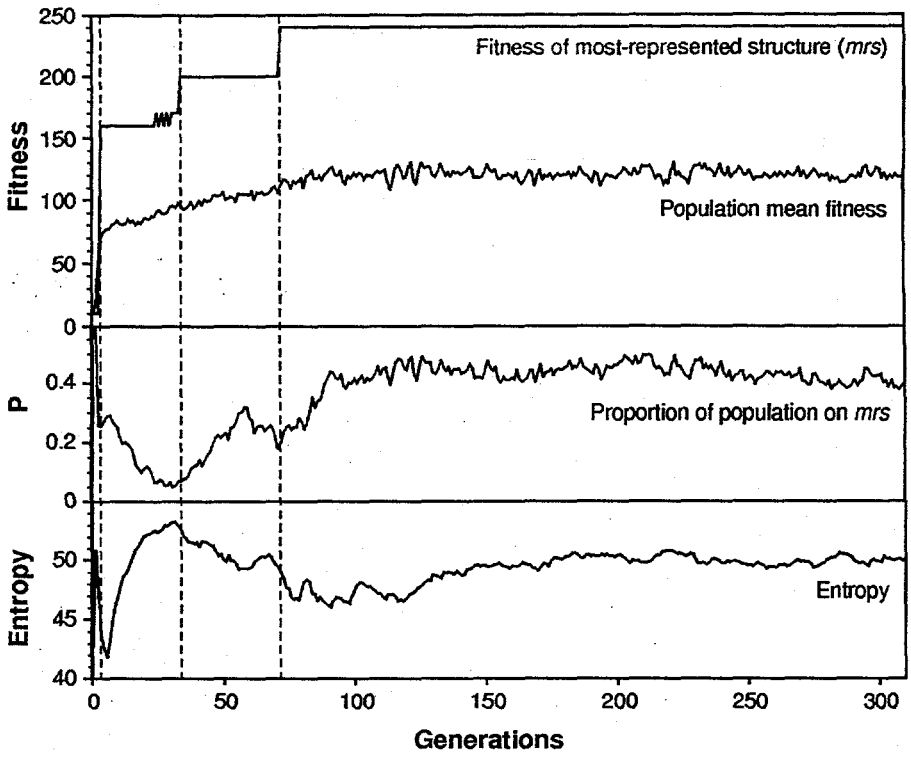

Figure 2: *A typical time evolution of a population of* 2000 *sequences of length* 40 *in a landscape of* 10^4 *random structures* $f_{\eta,\xi}(s_i) \in \mathbb{R}$ *where* $\mu\{f_{\eta,\xi} = k\xi\} = \frac{\eta^k}{k!}e^{-\eta}$, *and* $\lambda = 0.8$. *Displayed are mean fitness of the population, the fitness of the most represented structure (*mrs*), the fraction of elements of the population realizing the* mrs *and the Renyi-entropy as functions of time. The highest possible fitness in this experiment was* 250.

Accordingly the average fitness of ϕ reads

$$f_\phi = \sum_s \phi(f^{-1}(s))f(s)\,.$$

Now, the first coordinate V_l of the above ordered pair is chosen, according to fitness, with probability $f(s_{V_l})/f_\phi$ among the elements of $\mathbf{V}$. The second coordinate of the above pair is selected with uniform probability on $(V_i \neq V_l \,|\, i \in N)$, i.e. $1/(N-1)$. We select those pairs of sequences at equidistant time steps, and for a population of size N we refer to a *generation* as N such time steps.

Next, in the error-prone replication step, we map the first sequence, $V_l = (x_1, ..., x_n)$, into the sequence $V^* = (x'_1, ..., x'_n)$. This is performed by assigning to each coordinate x_i a

$x_i' \neq x_i$ with probability p where all $x_i' \neq x_i$ are equally distributed and leave the coordinate fixed otherwise. This random mapping $\iota \mapsto \upsilon^*$ is called *replication*. Finally, we delete the second coordinate of the pair (V_l, V_k), that is V_k and have a mapping $(V_l, V_k) \mapsto (V_l, V^*)$. Thereby we obtain a new family by substituting the V_k by the V^*. This process is referred to as the *replication-deletion process*.

The above replication-deletion scheme will be used to generate the time evolution of a population. The generic parameters for the following experiments are mapping probability $\lambda = 0.8$, fraction of secondary bonds $c_1 = 0.6$, fraction of tertiary bonds $c_2 = 0.05$, and an error rate p such that $pn = 1$. A typical run is shown in figure 2, in which a certain fraction of the population realizes the structure with the highest fitnesses discovered so far. This structure will be referred to as the most frequent structure realized by the population (*mrs*). According to the replication-deletion process described above, its error mutants search for better structures. Given that many error mutants find better structures the population spreads on the corresponding neutral networks and is rather delocalized. If the *mrs* has relatively high fitness the population is more localized on its neutral network. These findings are not really apparent in the trends in mean fitness. Close inspection of the entropy curve shows that when a fitter structure is discovered, the entropy peaks low. In this situation the population becomes more localized and then begins to diffuse on the newly found neutral network, whence the entropy increases again.

fitness	index	number of random sequences folding into s
250	7894	2
240	6661	4
220	816	3
220	7624	2
220	5634	3
220	4574	2
220	3096	9
220	2927	7
220	2388	2
220	1204	1
220	1143	4

The likelihood of a randomly-generated sequence being mapped into the structure found by the evolving population is of interest here. In the table below the numbers of 10^6 random sequences that were mapped into the eleven most-fit structures in the run are shown, showing that the chances of a random sequence realising one of the more fit structures is extremely low.

3.2 A stochastic phenomenon in flat landscapes: Transitions

There are various ways of representing how a population of sequences is organized in sequence space. The standard methods, like cluster analysis, are in fact too fine-grained. Generalizing a representation in [7], we introduce pairs of distances with respect to two neutral networks. The distances are obtained by counting simply the numbers of incompatible base pairs with respect to each structure s_1 and s_2 respectively. The population then has a certain number of elements in say, (i, k), that is these elements have i incompatible sequences w.r.t. s_1 and k w.r.t. s_2 respectively.

In our first experiment, we use this representation for a transition as shown in figure 3, where both structures have equal fitness. Therefore the transition phenomenon does not necessarily depend on the presence of a fitness gradient; it is a stochastic phenomenon.

Our second experiment consists of the following. Let s_1, s_2 be two random structures. We set their fitness to 10 and all other structures have fitness 1. We construct the neutral networks with respect to s_1 and s_2 with $\lambda = 0.8$ and initialize a random process by selecting N random sequences on the neutral networks w.r.t. s_1. We choose the error rate p such that $np \approx 1$ and start the replication-deletion scheme described above. Then we repeat the experiment increasing the population size, the results of which are shown in Figure 4. These show that for small population sizes, the vast majority of the population realises the same structure. Increasing the population size allows the population to split between both the neutral networks.

Third we analyze to what extent the transition phenomenon depends on structural sim-

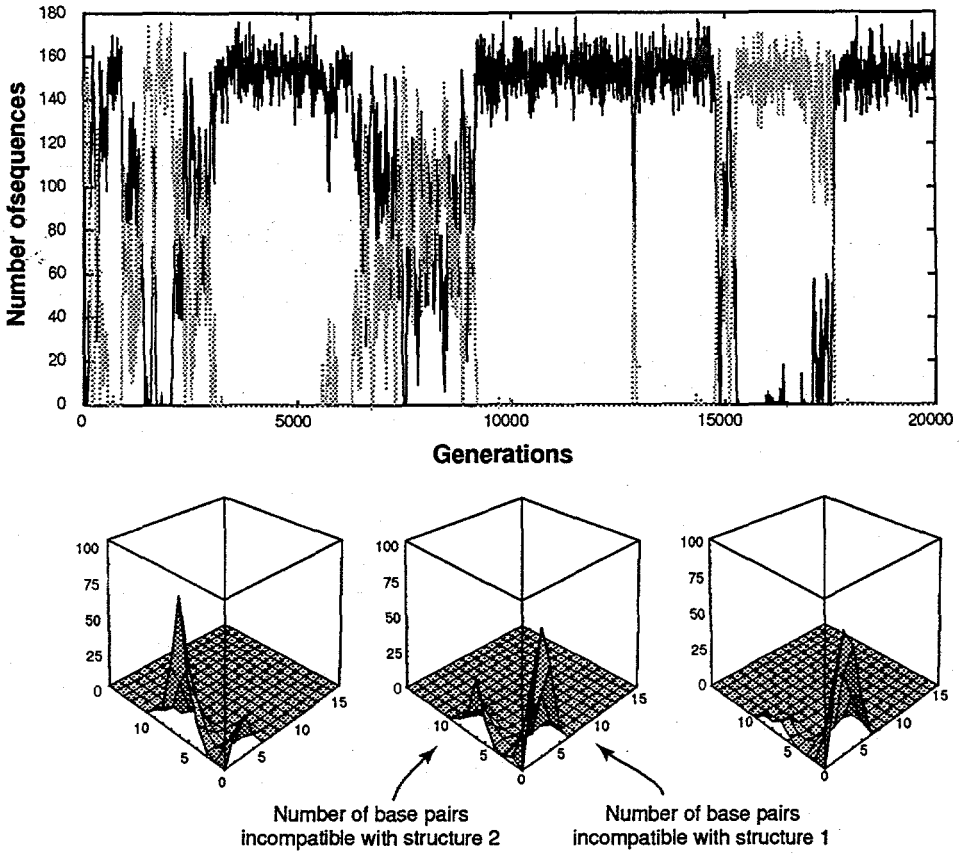

Figure 3: *Transitions between neutral networks of equally fit structures, in a population of* 300 *sequences of length* 30. *Displayed are: Upper: the numbers of sequences on the neutral network of* s_1 *(black) and* s_2 *(gray) as a function of time. Lower: the numbers of sequences with respect to the distance pairs* (i, k), *as described in the text, for a sample transition, in steps of 2 generations.*

ilarities. For this purpose we choose a random structure s_0 and consider two further structures s_1, s_2 with $n = 35$. While s_1 is a random structure and chosen independent of s_0, s_2 shares a fraction σ of its secondary bonds with s_0. We initialize a population of 500 random sequences on the neutral network of s_0 and let the population evolve until either a transition to one of the neutral networks of s_1 or s_2 has occurred or we terminate the run after 10^4 generations. The table below shows summary data from 10^3 runs of this experiment, the columns showing the proportions of runs terminating on the neutral networks of each of the structures s_1 and s_2 respectively. It is clear that increasing structural similarity makes transitions more likely.

fraction of common secondary bounds σ	fraction of runs that terminate with s_1	fraction of runs that terminate with s_2
0.1	0.3604	0.6436
0.15	0.3604	0.6436
0.2	0.2560	0.7440
0.25	0.2240	0.7760
0.3	0.1280	0.8720
0.35	0.1210	0.8790

4 Conclusions

Random structures, as well as secondary structures [7], induce neutral networks in sequence space, i.e. extended, practically connected subgraphs consisting of all sequences that are all mapped into a particular random structure. Neutral networks are either connected or consist of a few components, whose size depends on the fraction of neutral point mutants (with respect to the structure). They are stable under point mutations and allow hence for *neutral evolution* [4]. These properties of mappings of sequences into structures are generic properties and of central importance for the understanding of molecular dynamics.

Under the basic assumption that every sequence realizes a particular structure that exclusively determines its fitness, the time evolution of those structures exhibits unique features. Small populations diffuse on neutral networks and search for fitter structures by their variant offspring that fall off the net. Once a fitter structure is found small populations perform a transition to the neutral network that corresponds to the fitter structure. Interestingly this phenomenon is observed even if the new structure has equal fitness i.e. in a flat landscape. In a neutral evolution complete populations switch between two neutral networks associated to two structures having the same fitness.

The transition phenomenon is caused by the interplay of two effects. The first is the dieout time of a population located on one of the neutral networks, s_1 say, induced by the replication-deletion process. The second effect is the flow of elements which perform

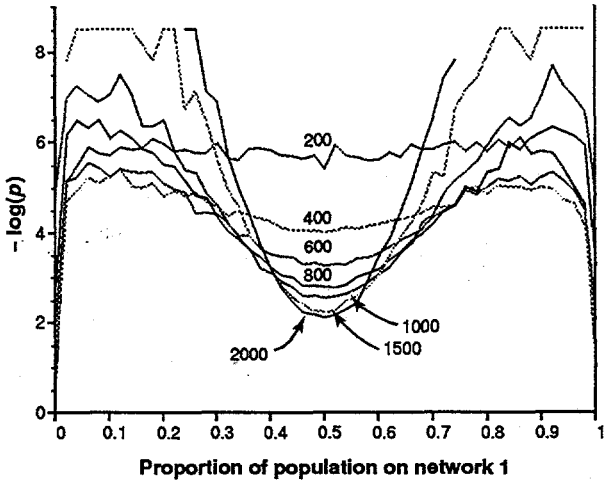

Figure 4: *Here two neutral networks w.r.t. the structures s_1, s_2 are given. Both structures have fitness* 10 *and all other structures are assumed to have fitness* 1. *We display the negative logarithm of a multi-set of probability distributions indexed by population sizes. The distributions represent the numbers of elements in the population that are on the neutral network w.r.t. s_1. The time evolution of the population bases on the replication-deletion scheme described in the text.*

a transition from s_2 to s_1, i.e. the mutants of s_2 sequences which are mapped into s_1. It is well known that the dieout time of a population of size N in a Markov process scales with e^{-N} [3], therefore below a certain fraction of the population on s_1, there cannot be sufficient flow from the network of s_2 in order to maintain it. Consequently, a small population spread across two neutral networks is in an unstable state, and will rapidly move onto either one of the networks. This is clearly shown in Figure 3. The transition phenomenon is restricted to fairly small population sizes. Above a certain population size we observe that the population splits onto both networks and searches in parallel. Hence for given error rate p and given structures s_1, s_2 there is a critical population size above which populations split among neutral networks corresponding to structures with equal fitness and and below which the population does not split (Figure 4).

Such transitions have been recently reported for RNA secondary structures [2]. In this case, a detailed analysis is rather difficult, because the union graph corresponding to the two structures cannot be described probabilisticly, although a group theoretical argument can be used [7]. For random structures, we can do this—transitions are closely related

to the properties of the union graph of the two underlying structures. These properties are amenble to analysis, having a complete probability space at hand. This allows us to determine what graph structures are typical for the union graph, and therefore to predict typical transition rates between two random structures. The properties of the union of two contact graphs of random structures depend heavily on the fraction of tertiary interactions [5, 1]. For a $c_1 = 0.6$ the critical fraction of tertiary interactions for the emergence of a giant component in the union graph of two random structures has $c_2 = 0.15$ as an upper bound (eq. (2)). For $c_1 = 1$ a lower bound on the critical fraction would be 0.125. The existence of this dramatic change, which is a phase transition in the limit of long sequences, does not depend on c_1 as long as $c_1 > 0$. Known 3D-structures (for example t-RNA) have values of c_2 which are well below this critical threshold, with about 4-6% nucleotides involved in tertiary interactions.

Of course, evolutionary adaptations on the structural level are rarely between completely unrelated structures, with continuing function dependent on at least some structural similarity. Here, we measure similarity in terms of shared secondary edges between structures, which will have significant impact on the structure of the union graph. As we have shown, transitions between such structures are much more likely. This suggests that evolutionary paths may exist, consisting of pairwise similar structures along which the population evolves in time.

Acknowledgments We want to thank Christopher L. Barrett for helpful discussion. SF is funded by DARPA under grant ONR N0014-95-1-1000.

References

[1] Reidys C.M. and Fraser S.M. Evolution on random structures. *Bull. Math. Bio.*, 1997. submitted.

[2] Weber J. *Evolution on RNA secondary structures.* PhD thesis, University of Jena, 1997. PhD Thesis.

[3] S. Karlin and H.M. Taylor. *A first Course in Stochastic Processes.* ACADEMIC PRESS, INC., 1975.

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

[5] C.M. Reidys. Mapping in random-structures. *SIAM Journal of Discrete Mathematics and Optimization*, 1996. submitted, May 1996.

[6] C.M. Reidys. Random induced subgraphs of generalized n-cubes. *Advances in Applied Mathematics*, 1997. accepted.

[7] C.M. Reidys, F.P. Stadler, and P.K. Schuster. Generic properties of combinatory maps and neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 1995. in press.

[8] Peter Schuster. How to search for RNA structures. *Journal of Biotechnology*, 41:239–257, 1995.

[9] Peter K. Schuster. Molecular evolution. *Physica D*, ??:279–284, 1996.