last changes: August 5, 1992

Error Thresholds on Correlated Fitness Landscapes

SEBASTIAN BONHOEFFER[1,2,*] AND PETER F. STADLER[1]

[1]Institut für theoretische Chemie
Universität Wien
Währingerstr. 17, A-1090 Wien, Austria

[2]Department of Zoology
University of Oxford
South Parks Road, OX1 3PS, Oxford, U. K.

Correspondence should be sent to S. Bonhoeffer at [2]

Abstract

The evolution of molecular quasispecies on two different complex fitness landscapes, the Sherrington Kirkpatrick spin glass and the Graph Bipartitioning landscape, is investigated in dependence on replication fidelity and population size. Three different regimes of replication fidelity are detected. At high copying fidelity one obtains a population with high mean fitness localized around a time independent consensus sequence. For large mutation frequencies the population is spread over the entire sequence space and the average fitness vanishes. In the intermediate regime the population is centered around a consensus sequence which moves in sequence space. Although the genetic information is lost over time, the population maintains an average fitness significantly above zero. A theoretical approximation is derived for the error threshold as a function of the sequence length.

1 Introduction: The Quasispecies Model

About 20 years ago Eigen (1971) introduced a dynamic theory of molecular evolution, the *quasispecies model*, in terms of the chemical kinetics of erroneously and independently replicating biopolymers. Within the framework of this theory initial distributions of polynucleotide sequences $\alpha, \beta, \ldots$ converge asymptotically towards a stationary mutant distribution uniquely characterized by the replication rate constants A_α and the mutation frequencies $Q_{\beta\alpha}$. This stationary distribution is commonly called the quasispecies as a generalization of the biological notion of species. The reaction network of Eigen's model is composed of error prone copy reactions of the form

$$\alpha \xrightarrow{Q_{\beta\alpha}A_\alpha} \alpha + \beta$$

and an unspecific dilution flux

$$\alpha \xrightarrow{\Phi} \emptyset$$

The dilution flux Φ is adjusted such as to keep the total population size of macromolecules constant. In the limit of infinitely large populations the dynamics are governed by the following system of ordinary differential equations (Eigen, 1971)

$$\dot{x}_\alpha = \sum_\beta Q_{\alpha\beta} A_\beta x_\beta - x_\alpha \sum_\beta A_\beta x_\beta \tag{1}$$

Explicit solutions in terms of eigenvalues of the matrix $W_{\alpha\beta} = Q_{\alpha\beta}A_\beta$ are known (Thompson & McBride, 1974, Jones *et al.*, 1976).

For the sake of mathematical tractability almost all investigations of the above model consider only point mutations as mutation events. As a consequence the macromolecules are restricted to constant chain length n. This allows to define a distance d (Hamming distance) between macromolecules as the minimal number of point mutations which is needed to transform one molecule into the other (Hamming, 1986). The set of all sequences together with the Hamming distance forms the *sequence space* (Maynard-Smith, 1970, Eigen, 1971) and the *fitness landscape* (Wright, 1932) is then obtained simply by assigning a replication rate constant A_α to each point α in the sequence space $\mathcal{S}$. This concept allows to look at evolution as an optimization procedure on a fitness landscape.

A further simplification s the *uniform error model* based on the assumption that point mutations occur independently with a probability p at each position of the sequence (*uniform error model*)(Eigen, 1971, Eigen *et al.*, 1988, Eigen *et al.*, 1989). Therefore the matrix element $Q_{\beta\alpha}$, which gives the probability to mutate sequence α into sequence β

in a single copying event depends only on the Hamming distance d and the single digit accuracy $q = 1 - p$.

$$Q_{\beta\alpha} = p^{d_{\alpha\beta}} q^{n-d_{\alpha\beta}} \tag{2}$$

The uniform error model was investigated in detail for various kinds of simple model fitness landscapes including single peak landscapes (Eigen & Schuster, 1977), double peak landscapes (Schuster & Swetina, 1988), essentially uncorrelated landscapes (McCaskill, 1984), flat fitness landscapes (Derrida & Peliti, 1991), and highly symmetric landscapes with few local optima (Tarazona, 1992). The effect of population size has been studied by Nowak and Schuster (1989) by simplifying the dynamics to a birth and death process. For correlated landscapes with many local optima extensive computer simulations have been published based on both RNA folding landscapes (Fontana & Schuster, 1987, Fontana *et al.*, 1989) and spin glass models (Amitrano *et al.*, 1989) focusing on the dynamics of evolution.

The most prominent feature of the quasispecies model is the existence of an *error threshold* (Eigen, 1971, Eigen & Schuster, 1977), i.e., a minimal replication accuracy necessary to maintain the genetic information in the population. Above the error threshold the distribution of sequences is centered around a consensus sequence, which depends only on the replication accuracy. The consensus sequence is defined as the sequence of the symbols which are predominant in each position. For $q \to 1$ the consensus sequence coincides with the global optimum of the fitness landscape. At lower replication rates the consensus sequence does not necessarily coincide with the global optimum, since the average fitness of the quasispecies is optimized. This is easily understood if one thinks of a global optimum lying one a very narrow peak and a broad local optimum with slightly smaller fitness (Schuster & Swetina, 1988).

Error thresholds are closely related to cooperative phase transitions (Demetrius, 1987, Leuthäuser, 1987, Tarazona, 1992); in the limit of infinite chain length they were shown to be equivalent to an order-disorder transition. So far error thresholds have not been studied for highly correlated complex fitness landscapes. These are of special interest since recent work on landscapes derived from RNA secondary structures (Fontana *et al.*, 1991, Fontana *et al.*, 1992a, Bonhoeffer *et al.*, 1992, Tacker *et al.*, 1992) revealed that biophysical landscapes are in fact highly correlated. Landscapes derived from some well known combinatorial optimization problems, e.g., the traveling salesman problem (Lawler *et al.*, 1985), the Graph Bipartitioning problem (Fu & Anderson, 1986), as well as long range spin glasses like the Sherrington Kirkpatrick model (Sherrington & Kirkpatrick, 1975) exhibit similar correlation structure (Stadler & Happel, 1992, Stadler & Schnabl, 1992, Weinberger & Stadler, 1992). Since these landscapes are computationally less costly by orders of magnitude compared to the more realistic RNA fitness landscapes, they allow for a detailed study of error thresholds on correlated landscapes.

The error threshold q_e also gives an upper bound for the genome size of organisms by

$$n_{\max} = \frac{\ln \sigma}{1 - q_e} \tag{3}$$

σ is the superiority parameter weighting the fitness of the wildtype against the fitness of the mutant distribution (Eigen, 1971). Interestingly some small RNA viruses like Qβ and HIV operate with single digit accuracies close to the error threshold (Domingo *et al.*, 1988, Preston *et al.*, 1988, Roberts *et al.*, 1988)

2 Fitness landscapes

In general a (fitness) landscape is a mapping $f : \mathcal{C} \longrightarrow \mathbb{R}$ assigning a real number to each configuration $\alpha \in \mathcal{C}$. The configuration space $\mathcal{C}$ is defined as the set of all possible configurations together with a metric distance $d_{\alpha\beta}$, which is not necessarily the Hamming distance. Landscapes can be classified by means of their autocorrelation functions

$$\rho(d) = \frac{\langle f(\alpha)f(\beta)\rangle_{d_{\alpha\beta}=d} - \langle f\rangle^2}{\langle f^2\rangle - \langle f\rangle^2} \tag{4}$$

where $\langle\,.\,\rangle$ denotes the average over the entire configuration space and $\langle\,.\,\rangle_{d_{\alpha\beta}=d}$ is the average over all pairs of configuration with given Hamming distance d. RNA free energy landscapes (Fontana *et al.*, 1991, Fontana *et al.*, 1992a, Bonhoeffer *et al.*), RNA landscapes of melting activation energy (Tacker *et al.*, 1992), landscapes of combinatorial optimization problems (Stadler & Happel, 1992, Stadler & Schnabl 1992), spin glass models (Weinberger & Stadler, 1992) and Kauffman's Nk-model (Kauffman & Levin, 1987 Kauffman *et al.*, 1988, Weinberger, 1990) exhibit an autocorrelation function of the form

$$\rho_E(d) = 1 - c\frac{d}{n} + \mathcal{O}(\frac{d}{n})^2 \tag{5}$$

where c is a constant independent of the chain length n.

In the context of the quasispecies model the fitness of a sequence α is simply the reaction rate constant A_α of its copy reaction. Reaction rate constants can commonly be related to activation energies by the Arrhenius law (Laidler, 1969)

$$A_\alpha = Be^{-E(\alpha)/RT} \tag{6}$$

B is the frequency factor which is assumed to be independent of the particular sequence α. For the sake of computationally tractability we choose the Sherrington Kirkpatrick spin glass (SK) and the Graph Bipartitioning Problem (GBP) instead of RNA landscapes in order to model activation energy landscapes.

The SK model is well known from statistical physics and was applied to prebiotic evolution by Anderson (1983). A particular spin configuration can be represented as a string $\alpha = \alpha_1, \ldots, \alpha_n$ where the $\alpha_i = \pm 1$ according to spin up or spin down. In the same way the genetic sequence can be coded, where $+1$ now stands for a purine and -1 stands for a pyrimidine. The energy of a spin glass configuration is

$$H(\alpha) = \frac{1}{n}\sum_{i<j} J_{ij}\alpha_i\alpha_j \tag{7}$$

where the J_{ij} are Gaussian distributed random variables with mean zero and variance 1. J_{ij} accounts for the strength of the spin interactions and n is the sequence length. The natural distance measure between two spin configuration is here also the Hamming distance. A single spin flip corresponds to a point mutation.

The GBP is known as a hard ($\mathcal{NP}$ - complete) optimization problem (Garey & Johnson, 1979). Given a set of an even number n of points with a connection pattern between them, a configuration is a partition of the set into two subsets A and B containing each half of the total number of points. A configuration is as a string $\alpha = \alpha_1, \ldots, \alpha_n$ where $\alpha_i = +1$ if the point i is in A and $\alpha_i = -1$ if $i \in B$. The task is to minimize the number of connections between A and B. Note that $\sum_{i=1}^{n} \alpha_i = 0$ since both A and B contain an equal number of points. The connection pattern can be described by a matrix J_{ij}, where $J_{ij} = 1$ if point i is connected with j , else $J_{ij} = 0$. Now the cost function of a configuration is

$$C(\alpha) = \sum_{i<j} J_{ij}(1 - \alpha_i\alpha_j) = \sum_{i<j} J_{ij} - \sum_{i<j} J_{ij}\alpha_i\alpha_j \tag{8}$$

The first term is only a constant which gives the total number of connections. Since we are only interested in optimization, we may omit it and use the energy function

$$H(\alpha) = -\sum_{i<j} J_{ij}\alpha_i\alpha_j \tag{9}$$

In contrast to the SK model and molecular genetics, exchange mutations are here the natural way of transforming one configuration into another. The Hamming distance is twice the minimal number of exchange mutations.

Activation energies are positive by definition. Therefore we define for both, the SK and the GBP landscape,

$$E(\alpha) = H(\alpha) + h \tag{10}$$

where h is a sufficiently large positive constant. Note that an additive or multiplicative constant does not change the autocorrelation function $\rho(d)$.

Our simulations of the quasispecies model are based on a simple algorithm. At each time step an individual from the population is selected for a replicating process with a probability proportional to its fitness. This individual is then replicated allowing for point mutations at every position with probability p. Finally the replication product replaces a randomly chosen individual in the population. Hence, the population size is kept constant.

The probability of an individual to be chosen for replication is given by

$$S(\alpha) = \frac{A_\alpha}{\sum\limits_{\beta\in\mathcal{P}} A_\beta} = \frac{e^{E(\alpha)/RT}}{\sum\limits_{\beta\in\mathcal{P}} e^{E(\beta)/RT}} = \frac{e^{-H(\alpha)/w}}{\sum\limits_{\beta\in\mathcal{P}} e^{-H(\beta)/w}} \tag{11}$$

where w is a shorthand for RT. The "temperature" is here simply a scaling parameter tuning the selective advantage, since no explicit assumptions on the absolute magnitude of the activation energies have been specified. α and β refer to single individuals in the population rather than to a species. Therefore concentrations do not enter in the above equation.

3 The error threshold

In order to locate the error threshold we simulated the evolution of a population at constant error rate for 200000 replication steps, which proved long enough to equilibrate each of the following parameters of the population.

- $\hat{f}$: the maximal fitness in the population
- $\bar{f}$: the average fitness of the population
- $\bar{d}$: the average Hamming distance in the population
- $\bar{d}_c$: the average Hamming distance to the consensus sequence in the population
- d_{co}: the distance of the consensus sequence to the global optimum o.

The convergence of these quantities can be drastically improved if, instead of a random population, a population equilibrated at a nearby error rate is used. It was carefully checked that this does not affect the results.

Figure 1 impressively illustrates the existence of a stable consensus sequence for error rates below the error threshold approaching it from below or above. The evolution of the consensus sequence by starting at the global optimum and varying the error rate of

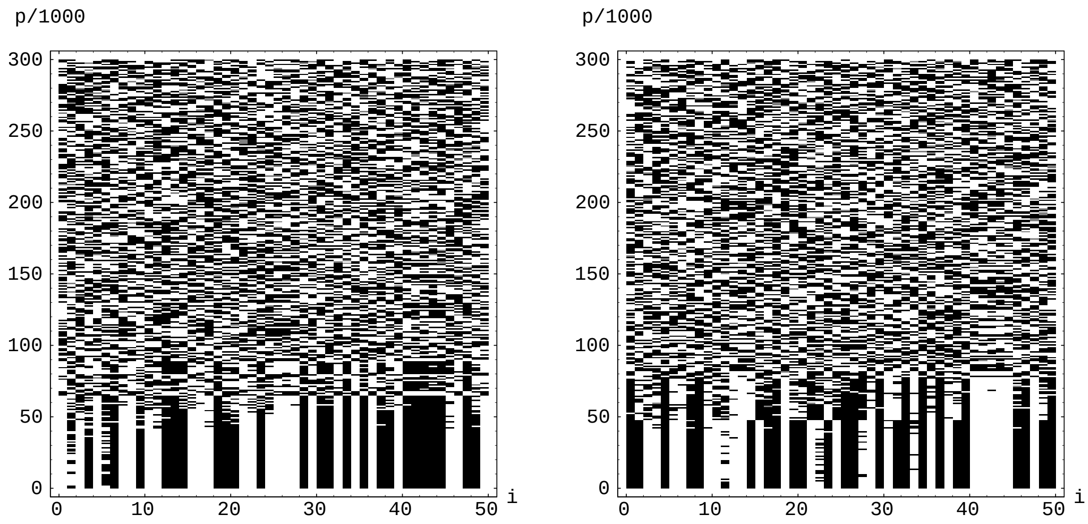

Figure 1: The consensus sequence in a population of 500 members on a GBP landscape with $n = 50$. If the majority of sequences have a $+1, -1$ in position i the field is plotted white or black, respectively. The field is plotted gray, if the position is undecided. *Left*: The simulation was started with a homogeneous population at the global optimum. Then the population was equilibrated (200000 replication steps) at a constant error rate p. Afterwards the error rate was increased by $\Delta p = 0.001$ and the computation was continued with the current population. Note the neutral exchange mutation (position $2 \leftrightarrow 6$) in the left plot. *Right*: Here we started with a random population on the same landscape at $p = 0.3$ and then step by step decreased the error rate (200000 replication steps, stepsize $\Delta p = 0.001$). Hence the time evolution is opposite to the p-axis. The population does not evolve to the global optimum of the left plot but to a local optimum with nearly the same fitness differing around 1% (see also the bottom right plot of fig. 2). The error threshold is characterized by the loss of the consensus sequence, i.e., the genetic information of the population. Beyond the error threshold the consensus sequence is no longer constant in time, the plots therefore show only a "snapshot". Slightly above the error threshold the consensus sequence flips back and forth erratically between the two complementary neighborhoods of the global optima.

replication is shown in figure 1. As a model fitness landscape we used here the Graph Bipartitioning landscape with $n = 50$. We see the loss of the information contained in the consensus sequence for error rates above the error threshold. In the left plot on the bottom there are neutral exchange mutations between position 2 and 6. When the error threshold region is reached we see a flip of the consensus sequence to its complement. Inspection of equ. 9 immediately tells us that complementary sequences have identical fitness.

The evolution of the consensus sequence by decreasing the error rate is shown in the right plot of figure 1. We used exactly the same landscape as in the left plot. Obviously the error threshold does not depend on whether it is approached from below or from above. A more accurate estimate of the error threshold can be obtained from the maximal fitness $\hat{f}$ and the distance of the consensus sequence to the optimum d_{co}. In figure 2 we show $\hat{f}$, $\bar{f}$, $\bar{d}$, d_c and d_{co} for the same simulation runs. In the three plots on the left side we increased the error rate of replication from 0 to 0.3 and started from the global optimum.

In the right hand side plots we decreased the error rate from 0.3 to 0. In the left plot the consensus sequence is at the global optimum, in the right plot the evolution leads the consensus sequence to local optimum with nearly the same fitness value (see bottom right plot fig. 2).

The same quantities are plotted in figure 3 for SK landscapes of length $n = 50$ (left) and $n = 100$ (right). Fig. 2 and fig. 3 look qualitatively the same, although the GBP and the SK landscape are quite different in their detailed structure. The constant c in equ. 5 differs of a factor 2 and the GBP landscape is much more neutral than the SK landscape, since H has only integer values for the GBP. These plots divide into three regimes of replication fidelity. increasing chain length. The first region of high replication accuracy is characterized by the persistence of the best sequence in the population and the fixation of the consensus sequence to a local optimum. In this region evolutionary optimization leads to a "steady state", i. e., the population progresses towards a local optimum and remains there. This means that the information of the consensus sequences is inherited to future generations. The speed of the optimization can be tuned by the replication precision. Optimization will be the fastest close to the error threshold, since too exact copying reduces the chance of producing new advantageous mutants. In nature a number of viruses have shown to operate very close to their error threshold. This enhances their flexibility to adapt on a continuously changing fitness landscape.

At the error treshold the localization of the quasispecies breaks down. Hence, the second region is characterized by a moving consensus sequence and non-vanishing mean fitness of the population. The movement of the consensus sequence sequence space is equivalent to the loss of genetic information. The replication procedure is too erroneous to maintain the information contained in the consensus sequence for future generations, however, the genetic relations in the populations are not yet destroyed. The population is as a diffuse but rather small cloud erratically staggering through sequence space. The consensus sequence is still well defined. The population behaves like a quasispecies on a flat fitness landscape (Derrida & Peliti, 1991).

In the third region the mean fitness is very small and the consensus sequence is no longer well defined. Here the replication is so randomized that the mean fitness vanishes. In the deterministic treatment (Eigen *et al.*, 1989) the second and the third domain are indistuinguishable, since time averages over the concentrations tend to $1/2^n$ in both regions.

Figure 3 shows $\hat{f}$, $\bar{f}$, $\bar{d}$, d_c and d_{co} for two different SK landscapes of length $n = 50$ and $n = 100$. The GBP and the SK landscape are quite different in their detailed structure. c in equ. 5 differs of a factor 2 and the GBP landscape is much more neutral than the SK landscape, since H has only integer values for the GBP. Nevertheless fig. 2 and 3 look qualitatively the same. This indicates that presumably the existence of the above described three different regions is a generic feature of complex correlated fitness landscapes. Comparing the plots on the left and the right side of fig. 3 shows, that the error threshold shifts towards 0 for increasing chain length.

The genetic information of a sequence in the population can only be maintained if its probability to be selected and replicated correctly is larger than to be removed, i.e.,

$$QS(\alpha) = q^n S(\alpha) > \frac{1}{N} \tag{12}$$

The error threshold is characterized by the loss of the fittest sequence o in the population, hence

$$Q_e S(o) = q_e^n S(o) \approx \frac{1}{N} \tag{13}$$

where N is the population size.

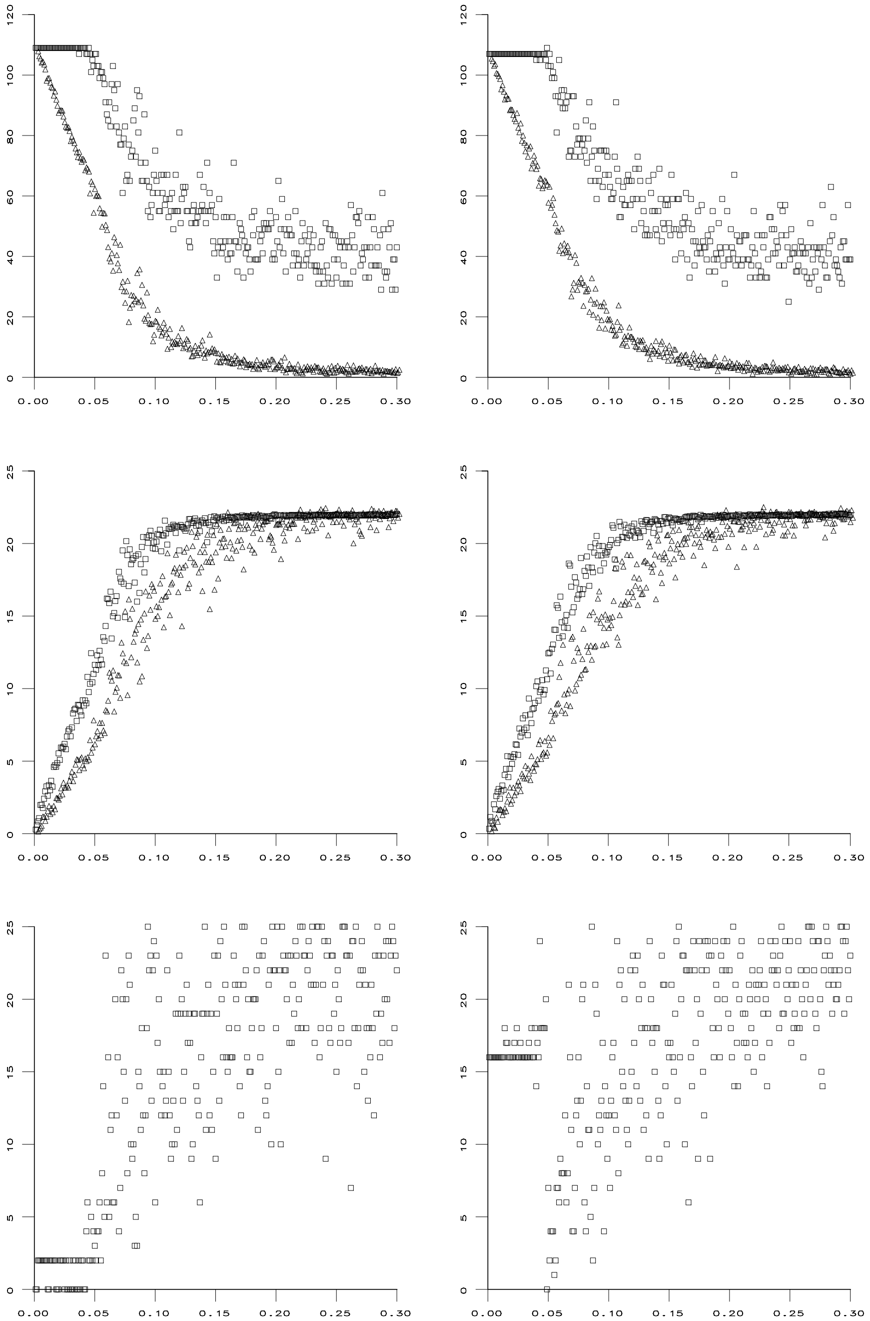

Figure 2: The evolution of a population on a Graph Bipartitioning landscape with $n = 50$. In the three plots on the left side we increased the error rate of replication from 0 to 0.3 and started at the global optimum. In the right hand side plots we decreased the error rate from 0.3 to 0. Hence, the time evolution of the system is here opposite to the X-axis. In the upper two plots we show the highest fitness value ($\square$) and the mean fitness ($\triangle$) in the population. The loss of the optimum is a characteristic of the error threshold. In the middle two plots we show the mean Hamming distance in the population ($\square$) and the mean Hamming distance to the consensus sequence ($\triangle$). The bottom plots show the distance of the consensus sequence to the local optimum. Since the GBP is symmetric the natural distance measure is $\min(d, n-d)$, where d is the Hamming distance. Note, that the population ended in a different local optimum (bottom right)

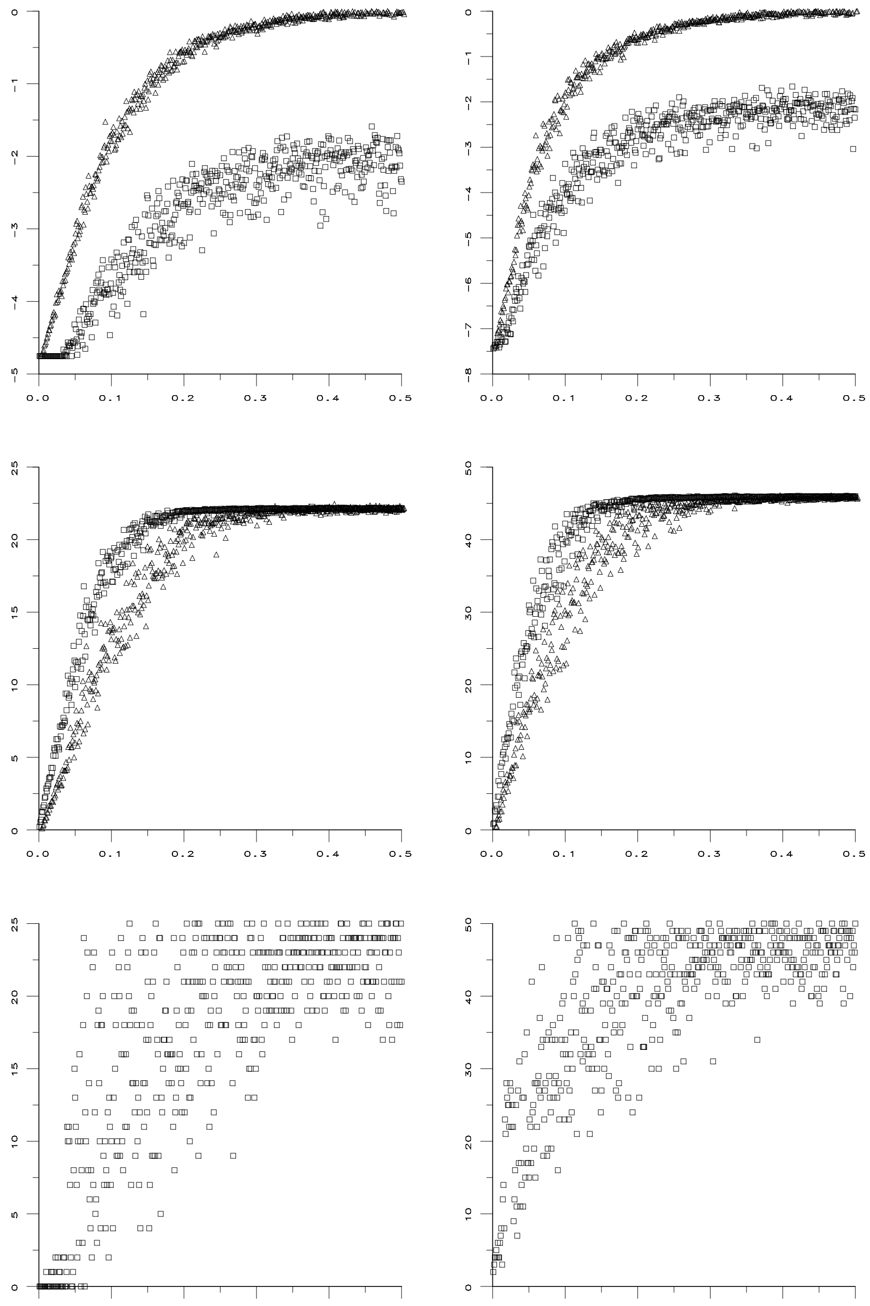

Figure 3: The evolution of a population on a Sherrington Kirkpatrick landscape with $n = 50$ (left) and $n = 100$ (right). In all plots we increased the error rate of replication from 0 to 0.5 and started at the global optimum. In the upper two plots we show the highest fitness value ($\Box$) and the mean fitness ($\triangle$) in the population. In the middle two plots we show the mean Hamming distance in the population ($\Box$) and the mean Hamming distance to the consensus sequence ($\triangle$). The distance is $\min(d, n-d)$, where d is the Hamming distance. In the bottom plots the minimal distance of the consensus sequence to the local optimum or its compliment is shown.

At the error threshold most sequences will have a fitness comparable to near neighbors of the optimum. The average value of H can therefore be approximated using the autocorrelation function as $\rho(d^\star)H(o)$, where $d^\star$ is close to one, since only sequences with a fitness comparable to the one or two error neighbors of o can survive due to the exponetial convergence of the fitness values on the activation energies. Now we can give an approximation for the denominator of equ. 11

$$\sum_{\beta\in\mathcal{P}} e^{-H(\beta)/w} \approx N e^{-\rho(d^\star)H(o)/w} \tag{14}$$

Using this and equs. 11 and 5 we get

$$S(o) \approx \frac{1}{N} e^{-(1-\rho(d^\star))H(o)/w} \approx \frac{1}{N} e^{-cd^\star H(o)/wn} \approx \frac{1}{q_e^n N} \tag{15}$$

Hence,

$$q_e \approx e^{cd^\star H(o)/wn^2} \approx 1 + \frac{cd^\star H(o)}{wn^2} \tag{16}$$

and the mutation rate at the error threshold is given by

$$p_e \approx -\frac{cd^\star}{w}\frac{H(o)}{n^2} \tag{17}$$

Theoretical estimates for the global optimum of the SK and the GBP are available in the literature (Mezard *et al.*, 1987). For the SK model in our normalization one finds

$$H(o) = -\sqrt{\frac{2}{\pi}}\, n^{1/2} \tag{18}$$

and for the GBP we get

$$H(o) = -\frac{1}{\sqrt{2\pi}}\, n^{3/2} \tag{19}$$

Figure 4 shows the error threshold for the SK model as a function of the sequence length. With equ. 18 the approximation 17 becomes

$$p_e \approx \sqrt{\frac{2}{\pi}}\,\frac{4d^\star}{w}\, n^{-3/2} \tag{20}$$

with $c = 4$ for the SK model (Stadler & Schnabl, 1992). In fig. 4 this function is drawn as a solid line with $w = 0.5$. A least square fit for $N = 500$ yields $d^\star = 1.78$. In the limit infinite population size $d^\star \approx 2$. The data obtained by numerical simulation is in very good agreement with the theoretical approximation.

In figure 5 we see the error treshold as a function of the scaling parameter w for a GBP landscape of length $n = 50$. The solid line corresponds to equation 17 with $H(o) = -111$, $c = 8$ (Stadler & Happel, 1992) and $n = 50$. Also here the numerical data agrees very well with the approximation for the error threshold.

For the population size dependence of the error threshold on single peak landscapes Nowak and Schuster (1989) derived a $1/\sqrt{N}$ law. In figure 6 the error threshold is plotted against the square root of the population size for SK landscapes of lengths $n = 25, 50$. The solid lines are mean square fits of the numerical data. Our data is consistant with this prediction.

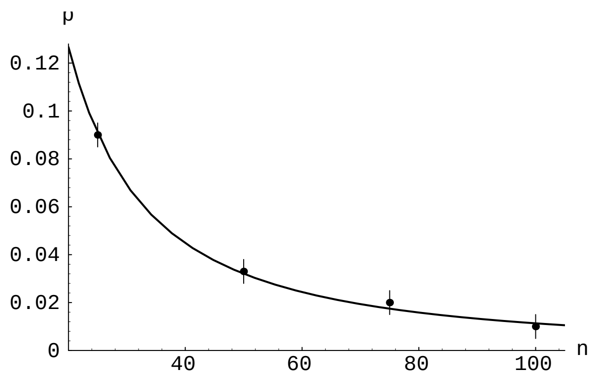

Figure 4: The error threshold versus the sequence length for the SK model. The solid line corresponds to the theoretical approximation equ. 17 with $d^\star = 1.78$. The points are derived from a numerical simulation

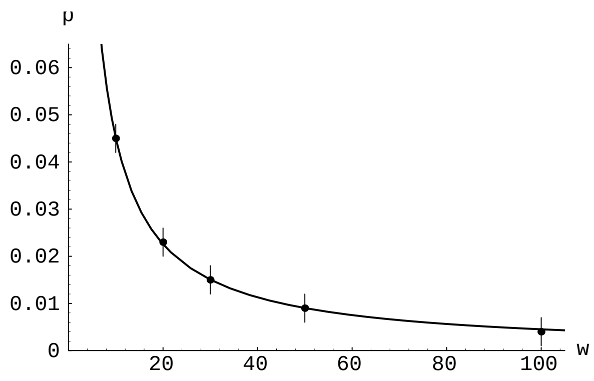

Figure 5: The error threshold versus the scaling parameter w in equation for the GBP landscape. The solid line corresponds to the theoretical approximation equ. 17 with $d^\star = 1.27$. The points are derived from a numerical simulation

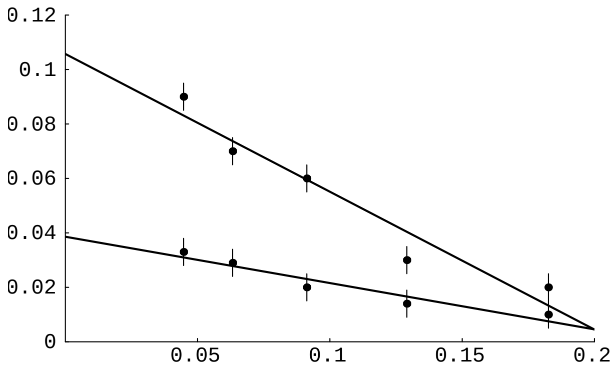

Figure 6: The error threshold p (Y-Axis) versus $N^{-1/2}$ (X-axis) for the GBP landscape ($N = 30, 60, 120, 250, 500$). The upper line corresponds to a GBP landscape of length $n = 25$, the bottom line corresponds to $n = 50$. The population size dependence of the error threshold seems to obey $N^{-1/2}$ law.

4 Conclusion and future directions

Our computer simulations of a quasispecies on correlated fitness landscapes show clearly three distinct regimes of copying fidelity. Populations with high replication accuracy evolve to a local optimum of fitness. The majority of the members in the population have a sequence close to the consensus sequence and the mean fitness is close to the global optimum. High replication accuracy guarantees the conservation of the genetic information for future generations. Populations with very unprecise replication spread uniformly over the sequence space. Their consensus sequence is ill defined and their mean fitness vanishes. In between there is region which is charaterized on one hand by the loss of the optimum in the population and on the other hand by non-vanishing mean fitness genetic relatedness. The population diffuses through sequence space below the peaks of the fitness landscape.

Probably this feature of complex correlated landscapes could be used for constructing improved optimization algorithms resembling simulated annealing. Starting with a population below the error threshold one increases the copying fidelity after an equilibration time stepwize beyond the error threshold. In the beginning the population will spread over the entire sequence space and probably such an optimization algorithm leads to a more exhaustive search for optimal solutions for very large optimization problems.

A simple theoretical approximation for the error threshold on complex landscapes was derived, which agrees very well with the numerical simulations. As models for fitness landscapes we used landscapes of combinatorial optimization problems, which have a similar correlation structure. Such landscapes are intrinsically isotropic, i.e., they are statistically equivalent in all directions. How close realistic biophysical landscapes are to isotropy has to be checked out. Strictly speaking, our approximation of the error threshold is valid only for isotropic landscapes, but we believe to describe the generic behavior correctly.

*Acknowledgements: Stimulating discussions with P. Schuster and I. L. Hofacker are greatfully acknowledged. This work was supported in part by *****

References

[1] C. Amitrano, L. Peliti, and M. Saber. Population dynamics in a spin-glass model of chemical evolution. *J. Mol. Evol.* , 29:513–525, 1989.

[2] P. W. Anderson. Suggested model for prebiotic evolution: The use of chaos. *Proc. Natl. Acad. Sci. USA*, 80:3386–3390, 1983.

[3] S. Bonhoeffer, J. McCaskill, P. F. Stadler, and P. Schuster. RNA multi-structure landscapes, a study based on temperature dependant partition functions. *submitted to Europ. Biophys. J.* , 1992.

[4] L. Demetrius. Random spin models and chemical kinetics. *J. Chem. Phys.* , 87:6939–6964, 1987.

[5] B. Derrida and L. Peliti. Evolution in a flat fitness landscape. *Bull. Math. Biol.* , 53:355–382, 1991.

[6] E. Domingo, J. J. Holland, and P. Ahlquist. *RNA Genetics*, volume 1-3. CRC press, Boca Raton, Fl, 1988.

[7] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 10:465–523, 1971.

[8] M. Eigen, J. McCaskill, and P. Schuster. Molecular quasi-species. *J. Phys. Chem.* , 92:6881–6891, 1988.

[9] M. Eigen, J. McCaskill, and P. Schuster. The molecular quasispecies. *Adv. Chem. Phys.* , 75:149–263, 1989.

[10] M. Eigen and P. Schuster. The hypercycle. a principle of natural selforganization. Part A: The emergence of the hypercycle. *Naturwissenschaften*, 64:541–565, 1977.

[11] W. Fontana, T. Griessmacher, W. Schnabl, P. F. Stadler, and P. Schuster. Statistics of landscapes based on free energies. *Mh. Chem.* , 122:795–819, 1991.

[12] W. Fontana, D. A. Konings, P. F. Stadler, and P. Schuster. Statistics of RNA secondary structures. *submitted to Biopolymers*, 1992.

[13] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaptation. *Phys. Rev. A*, 40:3301–3321, 1989.

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

[15] Y. T. Fu and P. W. Anderson. Application of statistical mechanics to $\mathcal{NP}$ -complete problems in combinatorial optimization. *J. Phys. A: Math. Gen.* , 19:1605–1620, 1986.

[16] M. Garey and D. Johnson. *Computers and Intractability. A Guide to the Theory of $\mathcal{NP}$ Completeness.* Freeman, San Francisco, 1979.

[17] R. W. Hamming. *Coding and Information Theory.* Prentice Hall, Englewood Cliffs, N. J. , 2nd edition, 1986.

[18] B. L. Jones, R. H. Enns, and S. S. Ragnekar. On the theory of selection of coupled macromolecular systems. *Bull. Math. Biol.* , 38:12–28, 1976.

[19] S. A. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* , 128:11–45, 1987.

[20] S. A. Kauffman, E. D. Weinberger, and A. S. Perelson. Maturation of the immune response via adaptive walks on affinity landscapes. In A. S. Perelson, editor, *Theoretical Immunology Part I*, Santa Fe Institute Studies in the Sciences of Complexity. Addison Wesley, Reading, Ma, 1988.

[21] K. J. Laidler. *Theories of Chemical Reaction Rates.* McGraw-Hill Book Company, 1969.

[22] E. Lawler, J. Lenstra, A. R. Kan, and D. Shmoys. *The Traveling Salesman Problem. A Guided Tour of Combinatorial Optimization.* John Wiley & Sons, 1985.

[23] I. Leuthäusser. Statistical mechanics of Eigen's evolution model. *J. Stat. Phys.* , 48:343–360, 1987.

[24] J. Maynard-Smith. Natural selection and the concept of a protein space. *Nature*, 225:563–564, 1970.

[25] J. McCaskill. A localization threshold for macromolecular quasispecies from continously distributed replication rates. *J. Chem. Phys.* , 80:5194–5202, 1984.

[26] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin Glass Theory and Beyond.* World Scientific Lecture Notes in Physics Vol. 9. World Scientific, Singapore, 1987.

[27] M. Nowak and P. Schuster. Error tresholds of replication in finite populations, mutation frequencies and the onset of muller's ratchet. *J. Theor. Biol.* , 137:375–395, 1989.

[28] B. D. Preston, B. J. Poiesz, and L. A. Loeb. Fidelity of HIV-1 reverse transcriptase. *Science*, 242:1168–1171, 1988.

[29] J. D. Roberts, K. Bebenek, and T. A. Kunkel. The accuracy of reverse transcriptase from HIV-1. *Science*, 242:1171–1173, 1988.

[30] P. Schuster and J. Swetina. Stationary mutant distributions and evolutionary optimization. *Bull. Math. Biol.* , 50:635–660, 1988.

[31] D. Sherrington and S. Kirkpatrick. Solvable model of a spin glass. *Phys. Rev. Lett.* , 35:1792–1796, 1975.

[32] P. F. Stadler and R. Happel. Correlation structure of the landscape of the graph-bipartitioning problem. *J. Phys. A: Math. Gen.* , 25:3103–3110, 1992.

[33] P. F. Stadler and W. Schnabl. The landscape of the traveling salesman problem. *Phys. Lett. A*, 161:337–344, 1992.

[34] J. Swetina and P. Schuster. Self-Replication with errors, a model for polynucleotide replication. *Biophys. Chem.* , 16:329–345, 1982.

[35] M. Tacker, P. F. Stadler, W. Fontana, and P. Schuster. Statistics of RNA melting kinetics. *in preparation*, 1992.

[36] P. Tarazona. Error thresholds for molecular quasispecies as phase transitions: From simple landscapes to spin glass models. *Phys. Rev. A [15] (Stat. Phys.* , 45:6038–6050, 1992.

[37] C. J. Thompson and J. L. McBride. On Eigen's theory of selforganization of molecules and the evolution of biological macromolecules. *Math. Biosci.* , 21:127–142, 1974.

[38] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybern.* , 63:325–336, 1990.

[39] E. D. Weinberger and P. F. Stadler. Why some fitness landscapes are fractal. *in preparation*, 1992, 1992.

[40] S. Wright. The role of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proccedings of the 6th International Congress on Genetics*, volume 1, pages 356–366, 1932.