Journal of Theoretical Biology 332 (2013) 218–227

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Exact results for amplitude spectra of fitness landscapes

Johannes Neidhart, Ivan G. Szendro, Joachim Krug*

Institute of Theoretical Physics, University of Cologne, Zülpicher Straße 77, 50937 Cologne, Germany

AUTHOR-HIGHLIGHTS

- Amplitude spectra quantify the epistatic interactions between genetic loci.
- We compute spectra analytically from the corresponding fitness correlations.
- For several commonly used models, the spectra have a surprisingly simple form.
- We fit the analytic formulae to different data sets and extract epistasis parameters.
- We show that Fourier decomposition is useful for analyzing empirical fitness data.

ARTICLE INFO

Article history:
Received 9 January 2013
Received in revised form
18 April 2013
Accepted 6 May 2013
Available online 14 May 2013

Keywords:
Fitness landscapes
Sequence space
Epistasis
Fourier decomposition
Experimental evolution

ABSTRACT

Starting from fitness correlation functions, we calculate exact expressions for the amplitude spectra of fitness landscapes as defined by Stadler [1996. Landscapes and their correlation functions. J. Math. Chem. 20, 1] for common landscape models, including Kauffman's *NK*-model, rough Mount Fuji landscapes and general linear superpositions of such landscapes. We further show that correlations decaying exponentially with the Hamming distance yield exponentially decaying spectra similar to those reported recently for a model of molecular signal transduction. Finally, we compare our results for the model systems to the spectra of various experimentally measured fitness landscapes. We claim that our analytical results should be helpful when trying to interpret empirical data and guide the search for improved fitness landscape models.

1. Introduction

In evolutionary processes, populations acquire changes to their gene content by mutational or recombinational events during reproduction. If those changes improve the adaptation of the organism to its environment, individuals carrying the modified genome have a better chance to survive and leave more offspring in the next generation. Through the interplay of repeated mutation and selection, the genetic structure of the population evolves and beneficial alleles increase in frequency. In a constant environment the population may thus end up in a well adapted state, where beneficial mutations are rare or entirely absent and only combinations of several mutations can further increase fitness.

To describe this kind of process, Wright (1932) introduced the notion of a fitness landscape. Here, the genotype is encoded by the coordinates of some suitable space and the degree of adaptation or reproductive success is modeled as a real number, called fitness, which is identified with the height of the landscape above the corresponding genotype. The evolutionary process of repeated mutation and selection is thus depicted as a hill climbing process. Mutations lead to the exploration of new genotypes and selection forces populations to move preferentially to genotypes with larger fitness. If more than one mutation is necessary to increase fitness, the population has reached a local fitness peak. Note that some caution is necessary when applying this picture, as the way in which genotypes are connected to one another does not correspond to the topology of a low-dimensional Euclidean space but is more appropriately described by a graph or network (see below). The underlying structure is well known from other areas of science, such as spin glasses in statistical physics (Binder and Young, 1986; Mézard et al., 1987) and optimization problems in computer science (Garey and Johnson, 1979).

The concept of fitness landscapes has been very fruitful for the understanding of evolutionary processes. While earlier work in this field has been largely theoretical and computational, in recent years an increasing amount of experimental fitness data for mutational landscapes has become available (Lunzer et al., 2005; Weinreich et al., 2006; Poelwijk et al., 2007; O'Maille et al., 2008;

---

* Corresponding author. Tel.: +49 2214702818.
E-mail address: krug@thp.uni-koeln.de (J. Krug).

0022-5193/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2013.05.002

Lozovsky et al., 2009; Brown et al., 2010; Hall et al., 2010; da Silva et al., 2010; Costanzo et al., 2011; Chou et al., 2011; Khan et al., 2011; Tan et al., 2011), see Szendro et al. (2013) for a review. Analysis of such data sets provides us with the possibility of a better understanding of the biological mechanisms that shape fitness landscapes and helps us to build better models. Thus, identifying properties of fitness landscapes that yield relevant information on evolution is an important task.

One such property that has attracted considerable interest is epistasis (de Visser et al., 2011). Epistasis implies that the change in fitness that is caused by a specific mutation depends on the configurations at other loci, or groups of loci, in the genome. In other words, epistasis is the interaction between different loci in their effect on fitness. Interactions that only affect the strength of the mutational effect are referred to as magnitude epistasis, while interactions that change a mutation from beneficial to deleterious or vice versa are referred to as sign epistasis (Weinreich et al., 2005). In the absence of sign epistasis, the fitness landscape contains only a single peak and fitness values fall off monotonically with distance to that peak. If sign epistasis is present, the landscape can present several peaks and valleys, which has important implications for the mutational accessibility of the different genotypes (Poelwijk et al., 2007; Franke et al., 2011; Franke and Krug, 2012) and shortens the path to the next fitness optimum (Orr, 2002; Joyce et al., 2008; Poelwijk et al., 2010; Neidhart and Krug, 2011; Crona et al., 2013). Thus, the absence of sign epistasis implies a smooth landscape, while landscapes with sign epistasis are rugged.

Beyond the question of the presence of epistasis, one would like to be able to make more detailed statements about *how much* of it is present or *in which way* epistasis is realized in the landscape. A very helpful tool to answer these kinds of questions is the Fourier decomposition of fitness landscapes introduced in Weinberger (1991a). This decomposition makes use of graph theory to expand the landscape into components that correspond to interactions between loci. The coefficients of the decomposition corresponding to interactions between a given number of loci can be combined to yield the *amplitude spectrum*. Calculating amplitude spectra numerically for data obtained from models or experiments is straightforward in principle, but so far only a small part of the information contained in the spectra is actually used. To improve this situation, it is important to understand how biologically meaningful features of a fitness landscape are reflected in its amplitude spectrum.

In this paper, we take a first step in this direction by analytically calculating spectra for some of the most popular landscape models: the *NK*-model introduced by Kauffman and Weinberger (1989) and Kauffman (1993), two versions of the rough Mount Fuji (RMF) model (Franke et al., 2011; Aita et al., 2000), and a generic model with correlations that decay exponentially with distance on the landscape. Thanks to the linearity of the amplitude decomposition, linear superpositions of these landscapes can also be treated. We calculate the spectra by exploiting their connection to fitness correlation functions originally established in Stadler and Happel (1999). Moreover, we compare some experimentally obtained spectra to the predictions of the models to see what features can be explained by these models and which cannot. In the next section we begin by introducing the definitions of fitness landscapes and their amplitude spectra on more rigorous mathematical grounds.

2. Fitness landscapes and their amplitude spectra

2.1. Sequence space and epistasis

On the molecular level the genotype of an organism is encoded in a sequence of letters taken from the alphabet $\mathfrak{A}=\{T,C,G,A\}$ of nucleotide base pairs with cardinality $|\mathfrak{A}|=4$. Point mutations replace single letters by others, altering the sequence and therefore the properties of the organism. A similar description applies to the space of proteins, where the cardinality of the encoding alphabet equals the number of amino acids (Maynard Smith, 1970). By contrast, in the context of classical genetics the units making up the genotype are genes occurring in different variants (alleles), which again can be described as letters in some alphabet (Wright, 1932). This provides a coarse-grained view of the genome in which also complex mutational events are represented by replacing one allele by another.

For simplicity, fitness landscapes are often defined on sequences composed of elements of a *binary* alphabet $\mathfrak{A}^B$, where a common choice is $\mathfrak{A}^B=\{0,1\}$. In this paper we prefer the symmetric alphabet $\mathfrak{A}^B=\{-1,1\}$ for mathematical convenience (Neher and Shraiman, 2011). Referring to the discussion in the preceding paragraph, we emphasize that the elements of the binary alphabet do not generally stand for bases or encoded proteins but rather indicate whether a particular mutation is present in a gene or not (Szendro et al., 2013). Therefore, the restriction to single changes in the sequence does not imply that the treatment is limited to point mutations.

All possible sequences of a given length N constructed from the alphabet $\mathfrak{A}$ with cardinality $|\mathfrak{A}|=\kappa$ form a metric space called the *Hamming space* $\mathbb{H}_\kappa^N$. It can be expressed as $\mathbb{H}_\kappa^N=(\mathcal{K}_\kappa)^N=\mathcal{K}_\kappa\otimes\cdots\otimes\mathcal{K}_\kappa$, where $\otimes$ denotes the Cartesian product and $\mathcal{K}_\kappa$ is the complete graph with κ nodes. For a binary alphabet $\mathbb{H}_2^N$ are hypercubes. Their metric is called the *Hamming distance*

$$\begin{aligned} d:\ &\mathbb{H}_2^N\times\mathbb{H}_2^N\to\mathbb{N}\cup\{0\}\\ &(\sigma,\sigma')\mapsto\sum_{i=1}^{N}(1-\delta_{\sigma_i,\sigma'_i}) \end{aligned} \tag{1}$$

which equals the number of single mutational steps required to transform one sequence into the other. To quantify the degree of adaptation or reproductive success of an organism carrying the genotype σ, a real number F called fitness is assigned to the corresponding sequence according to

$$\begin{aligned} F:\ &\mathbb{H}_2^N\to\mathbb{R}\\ &\sigma\mapsto F(\sigma). \end{aligned} \tag{2}$$

To precisely define the different notions of epistasis introduced above, we consider two sequences $\sigma,\sigma'\in\mathbb{H}_2^N$ with $d(\sigma,\sigma')<N$. Let $\sigma=(\sigma_1,\dots,\sigma_i,\dots,\sigma_N)$ and $\sigma'=(\sigma'_1,\dots,\sigma_i,\dots,\sigma'_N)$, and denote the sequences with a mutation at the i-th locus by $\sigma^{(i)}$ and $\sigma'^{(i)}$, respectively, with $\sigma_i^{(i)}=\sigma'^{(i)}_i=-\sigma_i$. If $F(\sigma)-F(\sigma^{(i)})\neq F(\sigma')-F(\sigma'^{(i)})$ for some i, the fitness landscape is called *epistatic*. If $\mathrm{sgn}(F(\sigma)-F(\sigma^{(i)}))=\mathrm{sgn}(F(\sigma')-F(\sigma'^{(i)}))$ the effect is called *magnitude epistasis*, while for $\mathrm{sgn}(F(\sigma)-F(\sigma^{(i)}))=-\mathrm{sgn}(F(\sigma')-F(\sigma'^{(i)}))$ it is called *sign epistasis*. Furthermore, the landscape is said to contain *reciprocal sign epistasis* if there are pairs of mutations such that $-\mathrm{sgn}(F(\sigma)-F(\sigma^{(i,j)}))=\mathrm{sgn}(F(\sigma)-F(\sigma^{(i)}))=\mathrm{sgn}(F(\sigma)-F(\sigma^{(j)}))$,with $\sigma^{(i,j)}$ denoting the sequence mutated at loci i and j (Poelwijk et al., 2007). A landscape with sign epistasis is said to be *rugged*, while landscapes containing no epistasis or only magnitude epistasis are called *smooth*. Non-epistatic landscapes are also called *additive*, as here the individual effects of mutations add up independently.

The presence of sign epistasis severely limits which paths on the landscape are accessible to evolution (Poelwijk et al., 2007; Weinreich et al., 2005; Franke et al., 2011). Landscapes that display reciprocal sign epistasis may contain several local fitness maxima (Poelwijk et al., 2010), while those that do not have a single maximum. The existence of reciprocal sign epistasis is a necessary but not sufficient condition for the existence of multiple maxima. For an example of a sufficient condition for multiple maxima based on local properties of the landscape see (Crona et al., 2013).