HAL
open science

Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape.

Alexandre Dawid, Daniel J Kiviet, Manjunatha Kogenaru, Marjon de Vos, Sander J Tans

To cite this version:

Alexandre Dawid, Daniel J Kiviet, Manjunatha Kogenaru, Marjon de Vos, Sander J Tans. Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape.. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2010, 20 (2), pp.026105. 10.1063/1.3453602 . hal-00867970

HAL Id: hal-00867970

https://hal.science/hal-00867970v1

Submitted on 30 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AIP | Chaos
An Interdisciplinary Journal of Nonlinear Science

Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape

Alexandre Dawid, Daniel J. Kiviet, Manjunatha Kogenaru, Marjon de Vos, and Sander J. Tans

Citation: Chaos **20**, 026105 (2010); doi: 10.1063/1.3453602
View online: http://dx.doi.org/10.1063/1.3453602
View Table of Contents: http://chaos.aip.org/resource/1/CHAOEH/v20/i2
Published by the AIP Publishing LLC.

Additional information on Chaos

Journal Homepage: http://chaos.aip.org/
Journal Information: http://chaos.aip.org/about/about_the_journal
Top downloads: http://chaos.aip.org/features/most_downloaded
Information for Authors: http://chaos.aip.org/authors

ADVERTISEMENT

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

Multiple peaks and reciprocal sign epistasis in an empirically determined genotype-phenotype landscape

Alexandre Dawid, Daniel J. Kiviet, Manjunatha Kogenaru, Marjon de Vos, and Sander J. Tans[a)]
AMOLF Institute, Science Park 104, 1098 XG Amsterdam, The Netherlands

(Received 5 February 2010; accepted 20 May 2010; published online 30 June 2010)

Insight into the ruggedness of adaptive landscapes is central to understanding the mechanisms and constraints that shape the course of evolution. While empirical data on adaptive landscapes remain scarce, a handful of recent investigations have revealed genotype-phenotype and genotype-fitness landscapes that appeared smooth and single peaked. Here, we used existing *in vivo* measurements on *lac* repressor and operator mutants in *Escherichia coli* to reconstruct the genotype-phenotype map that details the repression value of this regulatory system as a function of two key repressor residues and four key operator base pairs. We found that this landscape is multipeaked, harboring in total 19 distinct optima. Analysis showed that all direct evolutionary pathways between peaks involve significant dips in the repression value. Consistent with earlier predictions, we found reciprocal sign epistatic interactions at the repression minimum of the most favorable paths between two peaks. These results suggest that the occurrence of multiple peaks and reciprocal epistatic interactions may be a general feature in coevolving systems like the repressor-operator pair studied here. © *2010 American Institute of Physics*. [doi:10.1063/1.3453602]

It has long been recognized that the evolution of new functions is not only determined by the external forces of natural selection, but also by diverse internal limitations of the evolving biological system itself. Apart from the hard constraints imposed by physical and chemical laws, the Darwinian process of repeated selection of heritable changes can also give rise to adaptive limitations when some of the genetic changes that are required to reach a more adapted genotype are not unconditionally favorable. One of the most striking situations arises when no single genetic change is favorable while combinations of multiple genetic changes are, as it can lead to evolutionary stasis. This scenario can be seen as an entrapment in a local optimum in a multipeaked adaptive landscape. While in recent years methodologies have been developed to determine such adaptive landscapes empirically, the evidence for the existence of multiple peaks have been rather scarce and indirect. Here we analyze published experimental data on the expression regulation of mutants of the *lac* repressor and operator. We report the presence of multiple peaks in repression, as the key residues and base pairs for the binding specificity are varied in the transcription factor and its target DNA binding site. Together with our finding, the existence of multiple homologous repressor-operator pairs in *Escherichia coli* indicates that evolution has been able to avoid the frustration associated with local suboptima, and exploits the wide range of solutions available in the genetic space despite the presence of genetic barriers.

[a)]Electronic mail: tans@amolf.nl.

I. INTRODUCTION

Determining the architecture of adaptive landscapes is central to understanding the course of evolution. The stepwise adaptation of living systems to new environments by natural selection results from the intricate relationships between genotype and phenotype and between phenotype and fitness.[1] Ever since Wright[2] introduced the metaphor of an adaptive landscape, its shape has been hotly debated, but nonetheless essentially remained unknown due to insufficient empirical data.[1,3–12]

The architecture of adaptive landscapes is tightly related to the notion of epistasis (Fig. 1).[2,13] Epistasis provides a way to classify how elementary genetic changes correlate in terms of their effect on phenotype and fitness. For magnitude epistasis or in absence of epistasis, mutations give rise to either a positive or a negative fitness or phenotypic effect, regardless of the genetic background [Fig. 1(a), top]. This results in adaptive landscapes that are smooth and single peaked [Fig. 1(b), left]. In the case of sign epistasis, the sign of the fitness or phenotypic effect of a mutation does depend on the genetic background [Fig. 1(a), bottom left], such that only a fraction of the total paths to the optimum are selectively accessible, i.e., contain only steps that confer a performance increase. A third class of epistatic interactions is that of reciprocal sign epistasis, in which two genetic changes are individually deleterious but jointly advantageous [Fig. 1(a), bottom right]. It has been suggested that reciprocal sign epistatic interactions play a central role to generate adaptive landscapes with multiple distinct peaks [Fig. 1(b)].[10] The occurrence of multiple peaks can give rise to entrapment on local suboptima, which frustrates adaptation to the global optimum.

 © 2010 American Institute of Physics

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

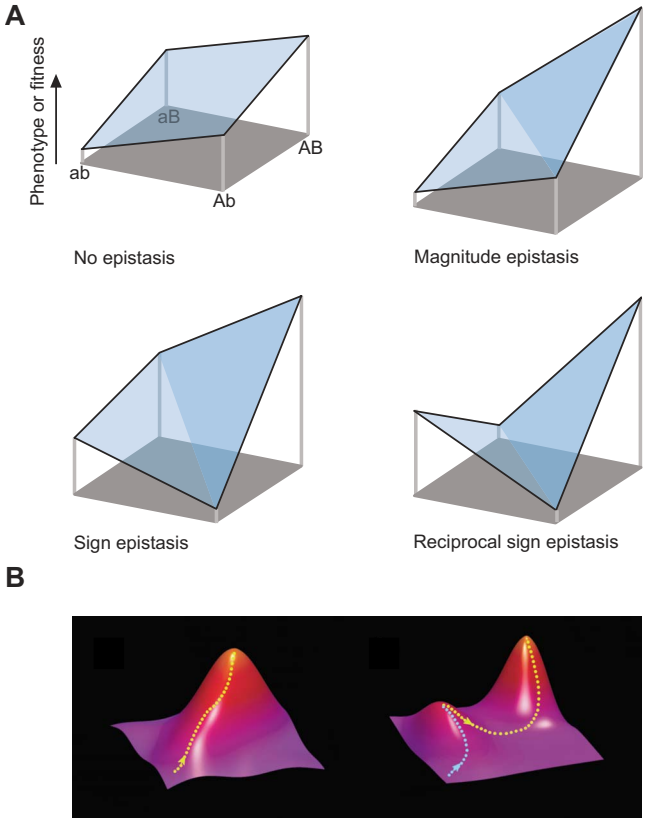

FIG. 1. (Color online) Relationship between epistasis and landscape ruggedness. (a) Schematic representation of different classes of epistatic interactions between mutations at two different genetic loci: a→A and b→B. In the absence of epistasis, mutation a→A yields the same phenotypic or fitness effect in genetic backgrounds b or B and vice versa. With magnitude epistasis, the phenotypic or fitness effect differs in magnitude depending on the genetic background. For sign epistasis, the sign of the fitness effect depends on the genetic background; as a result some paths are selectively inaccessible. In the case of reciprocal sign epistasis mutations are individually deleterious but collectively advantageous. (b) Continuous surfaces that serve to illustrate ruggedness in fitness landscapes. As a disclaimer, note that several features of these surfaces do not correspond to fitness landscapes. The left panel shows a single peaked surface where all the paths toward optimum are monotonously increasing in fitness. The left panel shows a multipeaked surface. All paths from the suboptimal peak to the optimal peak encounter a decrease in fitness.

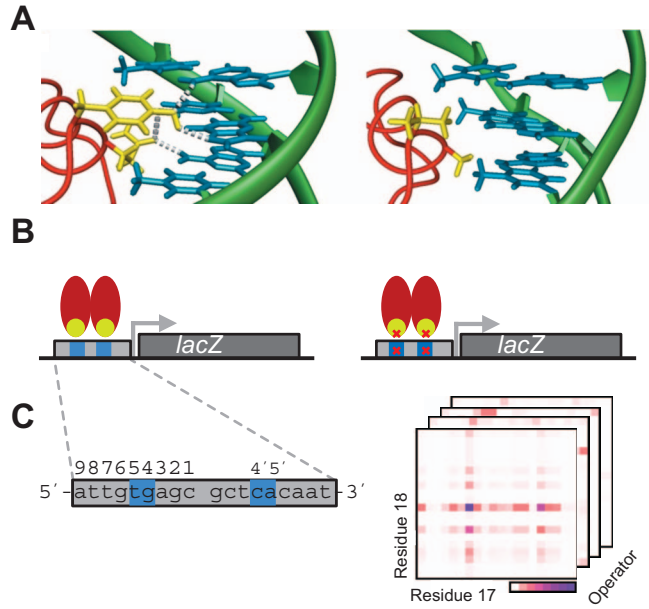

FIG. 2. (Color) Description of the studied system. (a) Structures illustrating the molecular interactions between the key residues in the *lac* repressor (yellow) and the key base pairs in the operator sequence (blue). The left panel shows a wild-type *E. coli* repressor-operator system, where the side chain of the key residues 17 and 18 from the repressor forms hydrogen bonds (dotted gray lines) with bases 4 and 5 of the symmetrical half operator. The left panel shows another repressor-operator pair. (b) Cartoon representation of the above three-dimensional structure with downstream reporter gene *lacZ* whose expression level is controlled by binding of the *lac* repressor (red) to the *lac* operator. The left panel represents the same but for another pair with mutations in the repressor and the operator (red crosses). (c) Representation of the data set. Genotype-phenotype map showing repression values as a function of residues 17 and 18 and all 16 operator variants based on the *in vivo* measurements. Low and high repression values are indicated by light and dark colors, respectively.

Spurred by systematic laboratory reconstructions of the evolutionary intermediates for a handful of well-characterized phenotypes, recent years have seen a renewed interest in the structure of adaptive landscapes.[3,10,11,14] These efforts have revealed the existence of sign epistatic interactions and single peaked landscapes. Here we investigate the structure of the genotype-phenotype landscape for the repression of the *lac* operon by the *lac* repressor and its operator. Using *in vivo* measured data from Müller-Hill and co-workers,[15] we have previously reconstructed this landscape to investigate the divergence between two repressor-operator pairs.[16] Here, we aim to determine whether the repression value for a single repressor and operator exhibits more than one distinct peak as a function of its genotype.

In this effort, we developed a recursive algorithm to search for peaks within the genotype space of the repressor-operator system. This analysis showed that the genotype-phenotype landscape is multipeaked, encompassing in total 19 well-defined optima. Our result contrasts with previous studies that showed single peaked adaptive landscapes,[3,14] which we suggest may be understood from the lock-key architecture of the here studied system. This finding, together with the observation that several repressor-operator pairs homologous to the *lac* system exist in *Escherichia coli*, suggests that evolution is able to overcome the frustration of a multipeaked landscape to exploit a wide diversity of interactions that are available in the genetic space.

II. DESCRIPTION OF THE SYSTEM

Recognition of DNA by proteins plays a central role in the regulation of transcription in all organisms. The lactose operon of *Escherichia coli* serves as a key biological system to study gene transcription regulation ever since Monod and Jacob[17] discovered it. Transcription regulation of this operon by binding of the *lac* repressor (LacI) to its operator regions in the *lac* promoter (Fig. 2) is understood in great detail and continues to be of great value in the study of gene regulation.[15,18–27] LacI is a prototypic member of the large GalR-LacI family of prokaryotic transcription factors, a group that has more than 1000 members.[28,29] Members of this family possess a conserved N-terminal DNA binding domain, and a central highly versatile domain that, under the same scaffold, functions as a binding pocket for different

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

types of small signaling molecules and promotes oligomerization of the complex by protein-protein interaction between the monomers. Binding of a signaling molecule to the receiving pocket allosterically regulates binding of the transcription factor to the target DNA sequence and thereby modulates mRNA production from the promoter of the operon.[29]

The *lac* system of *E. coli* is well-suited to start addressing the structure of adaptive landscapes for molecular interactions. Residues determining the specific binding between the *lac* repressor and its operator have been identified and circumscribed to only ten base pairs,[25] reducing to a large extent the genotypic search space: essentially, two key residues, Tyr-17 and Gln-18, from the recognition helix of the *lac* repressor are responsible for specific recognition of key base pairs 4 and 5 (and symmetrically related base pairs) in the palindromic *lac* operator sequence,[30] altogether reducing the determinant factors to six base pairs for the codons of residues 17 and 18 of the repressor, and four base pairs in the palindromic operator [Fig. 2(a)]. We note that other residues (e.g., Ser-21, Arg-22 of the recognition helix and base pair 6 of the operator) do have an effect on affinity, although less on specificity. Müller-Hill and co-workers have measured *in vivo* the repression values of repressor-operator pair variants obtained by extensive base pair substitutions at the aforementioned ten key positions.[15] Repression values for repressor-operator pair variants were determined as the ratio of repressed and unrepressed expression of a downstream β-galactosidase (*lacZ*) reporter gene, as measured via a standard Miller[31] assay. The measured data set covers 1286 out of a total of 6400 possible homodimeric repressor-palindromic operator variants (two amino acids and two independent base pairs).[15,31,32] From the measured data it has been observed that mutations in the key residues of the repressor (residues 17 and 18) contribute additively to the repression value, but the mutations in the key base pairs in the operator (base pairs 4 and 5) do not.

Based on this observed additivity between the repressor residues, the repression values for those mutants for which there were no measured data were determined by interpolation.[15,32] Additionally, to obtain the complete mapping between genotype and phenotype (i.e., repression value), we have extended the data set to also describe nonpalindromic operators, which constitute necessary intermediates for a step-by-step mutation process. To this aim, we have used the observation that each monomer of the dimeric repressor contributes additively to the binding energy with DNA.[16,33] Briefly, extrapolated repression values are calculated according to the general Eq. (2) of Ref. 16,

$$F_{o1o2} = 1 + \mathrm{sqrt}((F_{o1} - 1)^{*}(F_{o2} - 1)).$$

For a palindromic operator, $F_{o1o2}=F_{o1}=F_{o2}$, where $F_{o(1/2)}$ is the product of two factors (one for each of the key residues) taken from Table II of Ref. 15. F_{o1} and F_{o2} may also be unequal, thus yielding the repression value for a nonpalindromic operator. In total, our genotype data set consists of around 10^6 sequences (all combinations of ten independent base pairs) of repressor-operator variants, constituted of homodimeric repressors and palindromic or nonpalindromic operators (Fig. 2).

III. ALGORITHM

To find local repression optima in the genetic space of the repressor-operator pairs, the repression value of each point in the space is compared with the repression values of its nearest (single point mutation distant) neighbors. If at least one neighbor has a better repression value, then the point is not an optimum. If all the neighbors have lower repression values, then the point is an optimum. However, it might be that while none of the neighbors have a higher repression value, not all of them have lower repression values—that is, there might exist neutral neighbors. This does not necessarily mean that both the assessed point and the neutral neighbors are not optima; on the contrary, they might all together constitute an optimum plateau.

Therefore, during the assessment process of each point of the genetic space, each time a neutral neighbor is found a recursive procedure is started to determine (i) the extent of the associated neutral region and (ii) to test each point of the neutral region for optimality before concluding about the optimality of the entire region—i.e., if one of the points of the neutral region has a neighbor not in the neutral region and with a higher repression value, then the region is not an optimum.

At the end of the procedure, each point of the genotype map is defined either as "nonoptimum" or as "optimum-j," where j is an integer number that differentiates each distinct and independent local optimum, and that is the same for all neutral points of an optimum plateau.

IV. RESULTS

We have reconstructed the genotype-phenotype landscape detailing the repression values (defined as the ratio of unrepressed and repressed expression levels of the downstream *lacZ* gene) for variants of the *lac* repressor-operator system, and analyzed the ruggedness of the landscape. The genotype space contains about 10^6 variants, covering all possible combinations of mutations in the repressor and the operator, at the positions known to determine their binding specificity [i.e., base pairs 4, 5, and symmetrically positioned base pairs 4′ and 5′ in the operator, as well as base pairs coding for residues 17 and 18 in the recognition helix of the *lac* repressor[25] (Fig. 2)]. A particular operator-repressor variant of the explored genotype space is represented by the sequence at the four key positions in the operator (respectively base pairs 5, 4, 4′, and 5′—see Fig. 2), followed by symbols of the amino acids present respectively at residues 17 and 18 of the LacI protein. Thereby, the wild-type genotype would for instance be designated tgcaYQ.

The algorithm described above was used to search for peaks in repression values throughout the entire delineated genotype space. This analysis revealed 19 distinct peaks, i.e., areas of high repression values within the genotype space, isolated from each other by genotypes of strictly lower repression levels. Table I lists the genotypes and associated repression values of the 19 peaks of the landscape.

In order to quantify the distinctness of the peaks, we analyzed their relative distance and the decreases in repression values along the paths between them. On average two

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

TABLE I. Genotypes and repression values for the 19 independent peaks in the phenotype landscape of *lac* operator-repressor pair variants. Nonpalindromic operators are indicated by their two equivalent reverse-complement sequences.

Peak rank	Genotype		Repression level
	Operator	Repressor	
1	tg···ca	SM	520
2	tt···aa	HM	500
3	tg···ca	SQ	325
4	gt···ac	KS; KT; KM[a]	300
5	aa···tt	IS	300
6	aa···tt	SS	225
7	ta···ta	SG; IG; PG[b]	220
8	tg···ac/gt···ca	KM	219
9	ac···gt	PK	200
10	ac···gt	MK	200
11	aa···ac/gt···tt	SS	184
12	at···at	VM	160
13	ag···ac/gt···ct	PS	150
14	ag···ct	PS	150
15	tg···ac/gt···ca	KQ	100
16	gt···ac	KQ	100
17	tg···ca	SN	91
18	ga···tc	QT	90
19	gt···ac	VQ	50

[a]S, T, and M are connected, so 1 optimum.
[b]S, I, and P are connected, so 1 optimum.

peaks are separated by a Hamming distance of six mutations, with Hamming distances ranging between two and nine mutations. Note that a specific situation occurs in the case of serine which is encoded by two independent groups of codons separated by two mutations. This results in the existence of distinct peaks when this amino acid is present in the repressor. For simplicity, we have not distinguished these peaks in Table I.

Next, we looked more closely at the paths between two peaks separated by the average Hamming distance (six mutations). In particular, we considered the peaks atgcPK and tgcaSQ (respectively, peaks 9 and 3 in Table I). These two peaks have repression values of 200 and 325, respectively. The peak tgcaSQ is the optimum that is closest to the wild-type sequence tgcaYQ. Note that for simplicity we have excluded the cases of reverse mutations and restricted our analysis to direct paths between the peaks.

For a Hamming distance of h between two peaks, one can follow $h!$ different direct paths. Figure 3 presents the histogram of the smallest repression values encountered along each of the $6!=720$ paths going from peaks acgtPK to tgcaSQ. The vast majority of paths (>600) decreases down to repression values of 2 or less, which represents more than a 100-fold reduction in repression. The 12 most optimal paths, i.e., the paths that involve the least drastic dips, still decrease down to repression values of around 20. Thus, to evolve from one peak to the other, the system has to overcome a loss of at least tenfold in repression.

A number of typical paths are illustrated in Fig. 4, where the respective mutations and repression values at each step

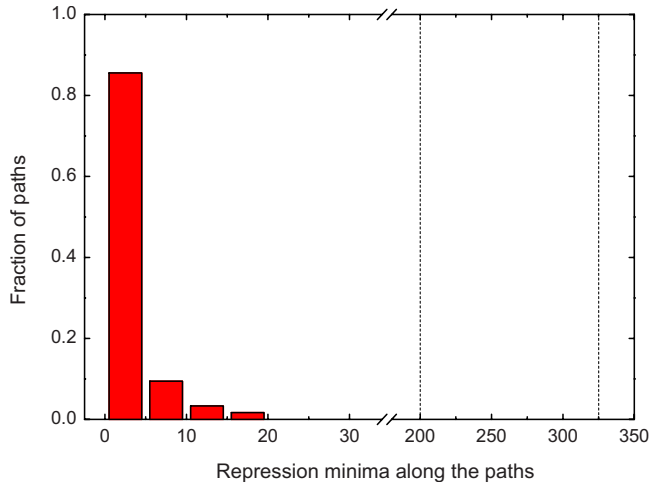

FIG. 3. (Color online) Histogram of the minimal repression values along direct paths between peaks acgtPK and tgcaSQ. Dotted lines indicate the repression levels at the beginning and at the end of the mutational path (repression values of 200 and 350 for genotypes acgtPK and tgcaSQ, respectively).

are indicated. In this graph, path 1 is one of the least likely paths, exhibiting a 200-fold drop in repression value at step 3 where the repression decreases to a value of 1. In this path, the operator is mutated first, resulting in disruption of its palindromic symmetry, and decreasing the repression value to about 80. At the second and third steps, the operator experiences additional mutations that bring it closer to the final sequence, although still maintaining the sequence asymmetry initially introduced. Ultimately, the repression shrinks to 1 at the third step. Subsequently, in steps 4 and 5 two mutations occur in the repressor. The first of these mutations, lysine (L) to glutamine (K), compensates for the mutations in the operator and restores the repression level to about 100, while the second mutation in the repressor, a proline (P) to serine

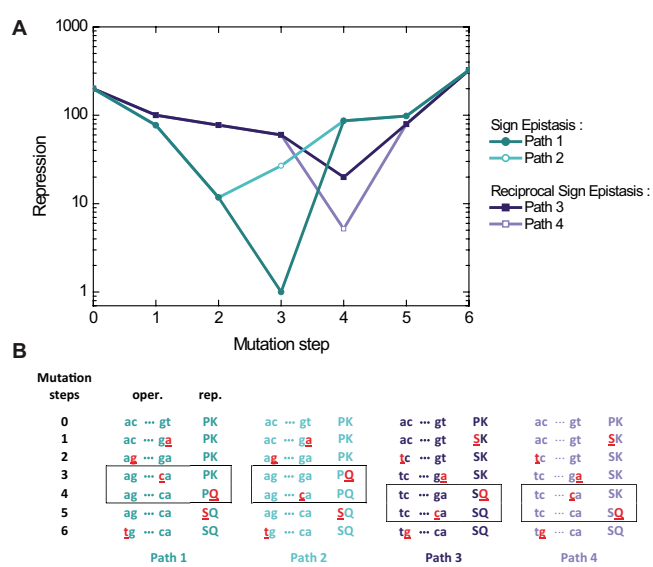

FIG. 4. (Color online) Examples of direct evolutionary paths between peaks acgtPK and tgcaSQ. (a) Repression values of the intermediate mutants along each path. (b) Intermediate sequences along the mutational paths. Mutations are underlined. Steps exhibiting epistatic interactions are outlined.

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

(S) transition, is almost neutral. Finally, the last *a* to *t* mutation from step 6 restores the symmetry of the operator, bringing the repression value to 350 at the tgcaSQ peak.

A close alternative to path 1 would be path 2, where all mutation steps are the same as in path 1 except for a permutation of the mutations occurring at steps 3 and 4, affecting respectively the operator and the repressor [see Fig. 4(b), paths 1 and 2, outlined steps]. With this new mutation order, instead of a decrease at step 3 followed by a restoration of the repression level at step 4, now both mutations (K to Q at steps 3 and *g* to *c* at step 4) increase the repression level, thus making this path more favorable. These two alternative paths show that the *g* to *c* mutation in the second half of the operator with the K or Q amino acid in the second key residue of the repressor exhibit a sign epistatic interaction.

The most likely path between optima acgtPK and tgcaSQ is path 3 depicted in Fig. 4, which exhibits the smallest dip among all possible paths. Here, the first mutation occurs in the repressor with the transition from P to S, which brings the repression level to about 100. The repression level then stays almost constant during the next two mutation steps that occur in the operator. Interestingly, in this pathway the palindromic symmetry, initially broken by the *a* to *t* mutation in the operator sequence at step 2, is immediately restored at step 3 with a *t* to *a* mutation in the second half of the operator. The following mutation is the K to Q transition in the repressor at step 4, which reduces the repression level to 20. This is the lowest repression level along this path, constituting a tenfold drop relative to the initial repression value at the peak. The repression level is then progressively restored as the two remaining symmetric *g* and *c* key bases of the operator mutate, respectively, to *c* and *g* to give the final palindromic operator.

In fact, path 3 belongs to a group of 12 best paths that are essentially equivalent. Indeed, due to the sequence symmetry of the operator, mutations of base pairs 5 and 5′ at steps 2 and 3 can occur indistinctively in the reverse order, as well as mutations at base pairs 4′ and 4 at step 5. Additionally, the P to S transition at step 1 can occur at step 2 or 3 with only a negligible decrease in repression values at some steps along the respective paths. Combining all these possible permutations produces a group of 12 paths, all having the same shape as path 3 with their minima in repression level at step 4.

The best alternative path to path 3, apart from the 12 aforementioned paths, is path 4, which differs only in the order of the mutations leading to, and following, the deepest drop in repression value at step 4. Permuting the order of these mutations—that is, the K to Q transition in the repressor protein and the first *c* to *g* mutation in the operator [Fig. 4(b), outlined steps]—results only in a deeper global dip in repression value at step 4 compared to path 3. Similar to paths 1 and 2, paths 3 and 4 also show that the effect of the *g* to *c* mutation in position 4′ can change sign, depending on residue 18 of the operator (K or Q). Additionally, however, the effect of this K to Q mutation now also changes sign depending on the position 4′ of the operator (*g* or *c*). Thus, the two mutations exhibit a reciprocal sign epistatic interaction [Fig. 1(a)].

Reciprocal sign epistasis occurs when two mutations are individually deleterious but jointly result in a positive effect [Fig. 1(a)]. Such a situation captures, at the level of individual mutation steps, the constraints created by a multipeaked landscape. Our analysis of paths 3 and 4 shows that the choice between alternative best paths between two peaks reduces to a choice between two routes in a reciprocal sign epistasis pattern that is located where these paths encounter their deepest drop in repression values. In other words, the lowest point in the optimal path between two peaks results from a reciprocal sign epistasis interaction. This observation illustrates that reciprocal sign epistatic interaction stands at key locations of a multipeaked landscape, and is in line with a theoretical investigation of ours, which indicates that reciprocal sign epistasis is an essential ingredient for the existence of multiple peaks.[34]

From Table I, we can also identify several peaks that are in close proximity to each other, being separated by a Hamming distance of only two. Three different situations can be discerned among the 13 cases. First, two different peaks can have the same repressor, while their operators differ by two mutations. This is for instance the case for peaks 6 and 11 or 13 and 14. The opposite situation also exists, where several peaks share the same operator sequence but the associated repressors differ by two mutations. This holds, for instance, for peaks 1, 3, and 17. The intermediate situation, where each of the operator and repressor variants differs by only one mutation between two peaks, also exists. This special case is encountered between peaks 11 and 13, both carrying a nonpalindromic operator.

Examination of the direct paths between these proximal peaks reveals an interesting pattern. When two peaks differ only by their operators (or by one mutation in the repressor and one in the operator—which is the case only between peaks 11 and 13), there are only weakly separated, with the minimal dip among the different paths being less than a factor of 2. Notably, when two proximal peaks differ by their operators, at least one of them is nonpalindromic. Thus, we do not observe two proximal peaks that differ only by their operators with both of them being palindromic. This observation might explain why those peaks are only weakly separated.

In contrast, when two proximal peaks differ only by their repressor sequence, which occurs in half of the cases of proximal peaks, the minimum drop in repression is larger than a factor of 5, or even a factor of 10 or 100 in three of the cases (between peaks 1 and 3, 3 and 17, and 8 and 15). The sole exception concerns the paths between peaks 16 and 19, for which the minimal drop in repression is small (i.e., less than a factor of 2). Therefore, while some peaks are close to each other in sequence space, they can still be separated by genotypes having substantial reduced repression values.

Note that due to the degeneracy of the genetic code, there exist silent mutations for most of the codons of the amino acids. Therefore, all peaks in the landscape form in fact a small plateau of neutral variants (see, however, Kimchi-Sarfaty *et al.*[35] for an example of phenotypic effect due to "silent" mutations). The presence of neutral variants at the peaks results in the existence of parallel identical groups

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

of paths between peaks. For instance, for the acgtPK to tgcaSQ transition depicted in Fig. 4, because P and S can be encoded, respectively, by the triplet *ccn* and *ucn* (where *n* can be any base), and K and Q, respectively, by the triplets *aar* and *car* (where *r* can be either *a* or *g*), there are in fact six identical "channels" of direct paths to go from one peak to the other (one channel for each combination of sequences at *n* and *r*). Each of these channels is therefore constituted of the same group of 720 paths described previously, differing only by the base sequence at positions *n* and *r* in the codons. For example, in the paths of Fig. 4, *n* and *r* have been arbitrarily chosen to be *c* and *a*—although this choice is not apparent and could have been different without altering the result. Since these channels are independent from each other, the validity of our previous discussion on epistasis and constraints is unaltered.

Interestingly, several of the peaks in the genotype-phenotype map of the *lac* repressor-operator occur for nonpalindromic operator sequences (see Table I). Considering the symmetry constraint imposed by the homodimeric repressor protein, this finding might seem counterintuitive as for the given repressor sequence, one would expect to find a better optimum with a symmetrized palindromic operator. Closer inspection reveals that all the nonpalindromic optima we have found need more than just one mutation for their operator sequence to become palindromic, making them just sufficiently isolated to constitute local optima (see Table I). Furthermore, these nonpalindromic optima are not global optima and their symmetrized versions always have a higher or at least equal repression value, thereby forming higher or equivalent peaks unless they are themselves buried into another higher peak.

V. DISCUSSION

In recent years a number of adaptive landscapes have been determined empirically through the genetic reconstruction of neighboring genotypes. These efforts have identified sign epistatic interactions, either at the genotype-phenotype level[3] or at the genotype-fitness level,[14] thereby showing that paths can be selectively inaccessible [Fig. 1(b)]. Nevertheless, some paths to the global optimum remained selectively accessible, indicating that the landscapes were single peaked.[3,14] Here we report the presence of multiple peaks in the landscape detailing the repression value of the *lac* regulation system as a function of key operator base pairs and repressor residues. The peaks are distinct: they consist of repressor-operator pairs capable of high repression values, which are surrounded by genotypic variants of lower repression levels. Our assumption of complete additivity between mutations in the key residues of the repressor might lead to an underestimate of the ruggedness of the landscape. Relaxing this assumption would only lead to a more rugged landscape. However, despite this assumption, distinct peaks are identified in the genotype-phenotype space.

A rationale for the existence of multiple peaks in the case of the *lac* regulatory system can be found by considering the analogy between the operator-repressor interaction and a key fitting a lock. Forming a new lock and matching key by stepwise mutations presents a dilemma: mutating the key first is not viable because it does not fit the old lock, and vice versa. This dilemma can arise for a recognition function between two components that can change both, in contrast, for instance, with an enzymatic reaction, where only one component changes by evolution. However, it is not a necessary consequence. The dilemma can in principle be resolved by the molecular equivalent of a master key: an intermediate transcription factor that is able to bind intermediate operator sequences, thus bridging two peaks.[10] Our study shows that such a master key does not exist for the *lac* repressor-operator system.

A multipeaked landscape reflects the widespread presence of epistatic interactions across the genotypic space. Indeed, among the mutations that bring the system to an optimum, there must necessarily be some that have a decreasing effect if introduced from another optimum. Otherwise the system would be single peaked. In other words, some of the mutations in one binding partner will only be beneficial when the other partner has already been modified, and vice versa. The requirement of such a reciprocal sign-epistatic interaction for multipeaked landscapes, which can also be theoretically addressed in a more rigorous manner,[34] is supported by our analysis: as predicted, such interactions appeared present along paths exhibiting the highest minimum.

It has frequently been recognized that a multipeaked landscape can constrain a stepwise Darwinian evolution process by trapping the evolving population in local suboptima—i.e., peaks lower than the global optima. Given the existing diversity of recognition within the GalR-LacI family of transcription factors,[36] the results suggest that evolution has been able to overcome entrapment on suboptimal peaks. Different scenarios may be considered for escaping suboptima. First, certain environments may free the system from a selective pressure temporarily, allowing new recognitions to be achieved through neutral drift. Alternatively, the participation of the system within a larger network of interacting components may alleviate the constraints. For instance, a duplication event may allow one of the duplicate repressors to compensate repression-decreasing mutations in the diverging copy.[16] One might also hypothesize the existence of hidden paths, involving substitutions beyond the key residues. However, this implies longer paths in an expanded genotypic space, which also occurs at the expense of reduced probability.[37]

Finally, we also would like to discuss the limitations of our approach. First, our analysis is based on phenotypic rather than fitness data. In order to address the evolutionary dynamics in a quantitative manner, the relation between repression characteristics and fitness should be determined, which also involves the nature of the environmental changes. Second, not all evolutionary intermediates have been directly characterized, but rather have been interpolated using the assumption that the two residues contribute additively to the repression value. While this assumption does not change our main conclusion that the *lac* repressor-operator system exhibits a multipeaked landscape, it will be of interest to reconstruct all intermediates between two peaks.

Specific molecular interactions are ubiquitous in biological systems and essential to their complexity and their ability

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions

to survive. One may therefore expect that multiple peaks in phenotype and fitness, as well as the underlying reciprocal sign epistatic interactions, be equally pervasive. It will be intriguing to explore how these elementary interactions shape the course of evolution of more elaborate biological functions.

ACKNOWLEDGMENTS

We thank Bianca Mladek for her comments on the manuscript. This work is part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijke Onderzoek (NWO). A.D. is supported by a long term fellowship from EMBO.

[1]R. Korona, C. H. Nakatsu, L. J. Forney, and R. E. Lenski, Proc. Natl. Acad. Sci. U.S.A. **91**, 9037 (1994).
[2]S. Wright, Proceedings of the Sixth International Congress of Genetics, 1932, Vol. 1, p. 356.
[3]M. Lunzer, S. P. Miller, R. Felsheim, and A. M. Dean, Science **310**, 499 (2005).
[4]R. A. Fisher, Heredity **4**, 117 (1950).
[5]S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[6]S. A. Kauffman and S. Johnsen, J. Theor. Biol. **149**, 467 (1991).
[7]S. A. Kauffman and E. D. Weinberger, J. Theor. Biol. **141**, 211 (1989).
[8]Y. Kim, Evolution (Lawrence, Kansas) **61**, 1847 (2007).
[9]D. M. Weinreich, R. A. Watson, and L. Chao, Evolution (Lawrence, Kansas) **59**, 1165 (2005).
[10]F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich, and S. J. Tans, Nature (London) **445**, 383 (2007).
[11]A. Buckling, M. A. Wills, and N. Colegrave, Science **302**, 2107 (2003).
[12]S. Wright, Genetics **16**, 97 (1931).
[13]M. C. Whitlock, P. C. Phillips, F. B. G. Moore, and S. J. Tonsor, Annu. Rev. Ecol. Syst. **26**, 601 (1995).
[14]D. M. Weinreich, N. F. Delaney, M. A. Depristo, and D. L. Hartl, Science **312**, 111 (2006).
[15]N. Lehming, J. Sartorius, B. Kisters-Woike, B. von Wilcken-Bergmann, and B. Müller-Hill, EMBO J. **9**, 615 (1990).
[16]F. J. Poelwijk, D. J. Kiviet, and S. J. Tans, PLOS Comput. Biol. **2**, e58 (2006).
[17]F. Jacob and J. Monod, J. Mol. Biol. **3**, 318 (1961).
[18]K. Adler, K. Beyreuther, E. Fanning, N. Geisler, B. Gronenborn, A. Klemm, B. Müller-Hill, M. Pfahl, and A. Schmitz, Nature (London) **237**, 322 (1972).
[19]A. Barker, R. Fickert, S. Oehler, and B. Müller-Hill, J. Mol. Biol. **278**, 549 (1998).
[20]C. E. Bell and M. Lewis, Curr. Opin. Struct. Biol. **11**, 19 (2001).
[21]M. P. Calos, D. Galas, and J. H. Miller, J. Mol. Biol. **126**, 865 (1978).
[22]C. Coulondre and J. H. Miller, J. Mol. Biol. **117**, 525 (1977).
[23]E. R. Eismann and B. Müller-Hill, J. Mol. Biol. **213**, 763 (1990).
[24]R. Fickert and B. Müller-Hill, J. Mol. Biol. **226**, 59 (1992).
[25]N. Lehming, J. Sartorius, M. Niemöller, G. Genenger, B. von Wilcken-Bergmann, and B. Müller-Hill, EMBO J. **6**, 3145 (1987).
[26]J. H. Miller, D. Ganem, P. Lu, and A. Schmitz, J. Mol. Biol. **109**, 275 (1977).
[27]J. Suckow, P. Markiewicz, L. G. Kleina, J. Miller, B. Kisters-Woike, and B. Müller-Hill, J. Mol. Biol. **261**, 509 (1996).
[28]M. Weickert and S. Adhya, J. Biol. Chem. **267**, 15869 (1992).
[29]L. Swint-Kruse and K. S. Matthews, Curr. Opin. Microbiol. **12**, 129 (2009).
[30]J. R. Sadler, H. Sasmor, and J. L. Betz, Proc. Natl. Acad. Sci. U.S.A. **80**, 6785 (1983).
[31]J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, New York, 1972).
[32]N. Lehming, Ph.D. thesis, Universität zu Köln, 1990.
[33]M. Hollis, D. Valenzuela, D. Pioli, R. Wharton, and M. Ptashne, Proc. Natl. Acad. Sci. U.S.A. **85**, 5834 (1988).
[34]F. J. Poelwijk, S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans "Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes," J. Theor. Biol. (submitted).
[35]C. Kimchi-Sarfaty, J. M. Oh, I.-W. Kim, Z. E. Sauna, A. M. Calcagno, S. V. Ambudkar, and M. M. Gottesman, Science **315**, 525 (2006).
[36]C. C. Nguyen and J. M. H. Saier, FEBS Lett. **377**, 98 (1995).
[37]J. T. Bridgham, E. A. Ortlund, and J. W. Thornton, Nature (London) **461**, 515 (2009).

Downloaded 30 Sep 2013 to 193.48.255.141. This article is copyrighted as indicated in the abstract. Reuse of AIP content is subject to the terms at: http://chaos.aip.org/about/rights_and_permissions