

# On the Probability of Reaching High Peaks in Fitness Landscapes by Adaptive Walks

Yang Li <sup>1</sup> Jianzhi Zhang <sup>1,\*</sup>

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

\*Corresponding author: E-mail: [jianzhi@umich.edu](mailto:jianzhi@umich.edu)

Associate editor: Haipeng Li

## Abstract

Adaptive evolution can be described by an uphill walk in a fitness landscape. However, climbing the global peak in a multipeak landscape is improbable because of the high chance of being trapped at a local peak. Nonetheless, over three-quarters of simulated adaptive walks in the fitness landscape of the *Escherichia coli* dihydrofolate reductase (DHFR) gene were reported to end at the highest 14% of peaks, suggesting that biological systems may be substantially more evolvable than commonly thought. To investigate the cause and generality of this observation, we estimate in empirical and theoretical fitness landscapes the probability of reaching high peaks by adaptive walks ( $P_{HP}$ ), where high peaks refer to the highest 1, 5, 14, or 25% of all peaks. We find that (i)  $P_{HP}$  varies substantially among landscapes, (ii)  $P_{HP}$  in empirical landscapes is generally comparable to or smaller than that in same-size Rough Mount Fuji landscapes of similar ruggedness, and (iii) lowering landscape ruggedness boosts  $P_{HP}$ . As observed in DHFR, we find in every examined landscape a positive correlation between the fitness of a peak and its basin size, which is the number of genotypes that can reach the peak through adaptive walks. Yet, this correlation does not guarantee a large  $P_{HP}$  because of the influences of other factors. We conclude that evolvability depends on the specific fitness landscape and that the large  $P_{HP}$  in the DHFR landscape is not a general property of empirical or theoretical fitness landscapes.

**Keywords:** fitness landscape, adaptive walk, Rough Mount Fuji model, landscape ruggedness, peak accessibility

## Introduction

Evolution is often described as a walk of a population in a fitness landscape depicting the relationship between genotypes and their fitness (Wright 1932; McCandlish 2011; De Visser and Krug 2014). As in geographic landscapes, fitness landscapes contain (fitness) peaks where any single-nucleotide change in the genotype results in a fitness reduction. Past empirical studies revealed that fitness landscapes typically contain multiple peaks (i.e. rugged) (Weinreich et al. 2005; De Visser and Krug 2014; Nahum et al. 2015; Lyons et al. 2020; Bank 2022); the highest peak is known as the global peak, while all other peaks are referred to as local peaks. Unless the population of concern is small, evolution is generally an uphill walk in the fitness landscape, with every step associated with a fitness gain or at least not a fitness drop. Much has been learned through theoretical analyses of adaptation in fitness landscapes (Franke et al. 2011; Franke and Krug 2012; Manhart and Morozov 2014; Neidhart et al. 2014; Ferretti et al. 2018; Krug and Oros 2024). It is clear that, in a fitness landscape with many peaks, the probability of arriving at the global peak through an adaptive walk starting from a random genotype is slim, because the population is easily trapped in one of the local peaks (Orr 2005, 2009). In other words, the evolvability of biological systems may be severely limited due to the ruggedness of fitness landscapes.

Recently, Papkou et al. (2023) measured the fitness landscape of a nine-nucleotide segment of the *Escherichia coli* gene *folA*, corresponding to three amino acids of its encoded

dihydrofolate reductase (DHFR). The mapped landscape includes the fitness estimates of 99.7% of all possible genotypes of the nine-nucleotide segment. In a sublandscape consisting of 135,178 genotypes (see below), 514 fitness peaks were identified. Based on the fitness distribution of these peaks, Papkou et al. classified the 74 highest peaks—all having amino acid Glu or Asp at position 27—as high peaks. They reported that an adaptive walk starting from a randomly picked genotype in the sublandscape has a probability ( $P_{HP}$ ) of 76.5% to arrive at one of the high peaks, despite that the high peaks constitute only  $74/514 \approx 14\%$  of all peaks. This unexpected observation suggests that, notwithstanding the difficulty in reaching the global peak by an adaptive walk in a rugged landscape, reaching a high peak is highly probable. Papkou et al. explained the unexpectedly large  $P_{HP}$  by a positive correlation between the fitness of a peak and its basin size, which is the number of genotypes that can reach the peak by adaptive walks; that is, higher peaks are more accessible than lower peaks.

Papkou et al.'s finding implies that the evolvability of biological systems may be substantially higher than currently thought, provided that evolvability includes reaching any of the high peaks. However, because Papkou et al. studied only one gene segment, it is unclear whether their finding extends to other fitness landscapes and whether the explanation offered for the large  $P_{HP}$  is general. In the present work, we address these questions by investigating multiple empirical and theoretical fitness landscapes. We show that  $P_{HP}$  varies greatly among empirical and theoretical fitness landscapes and that

Received: October 31, 2024. Revised: February 25, 2025. Accepted: March 10, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [reprints@oup.com](mailto:reprints@oup.com) for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

the large  $P_{HP}$  of the DHFR sublandscape is not general. We further show that while the positive correlation between peak fitness and basin size appears universal, this correlation does not guarantee a large  $P_{HP}$  because of the influences of other factors.

## Results

### Variation in the Probability of Reaching High Peaks Among Empirical Landscapes

We started by replicating the published estimation of the probability of reaching various peaks in the fitness landscape of DHFR (Papkou et al. 2023). Of the 261,382 genotypes with fitness estimates (Table 1), Papkou et al. considered 18,029 genotypes to be functional and the rest to be nonfunctional due to their low fitness. They then removed nonfunctional genotypes with no functional neighbors, where a neighbor refers to a genotype that differs from the focal genotype by one nucleotide change. Of the remaining genotypes, they identified a giant component (i.e. the largest connected subgraph) of 135,178 genotypes consisting of 18,019 functional genotypes and 117,159 nonfunctional genotypes (Table 1). They then treated this giant component as the DHFR landscape in subsequent analysis. To avoid confusion, we will refer to this giant component as the DHFR sublandscape; the landscape consisting of all 261,382 genotypes with estimated fitness will be referred to as the DHFR landscape. We first followed Papkou et al. to simulate adaptive walks in the DHFR sublandscape. Specifically, starting from a randomly picked genotype in the sublandscape, we simulated an adaptive walk (see Materials and Methods) till the population reached a fitness peak. Such an adaptive walk was repeated for  $10^6$  randomly picked starting genotypes. We then drew the cumulative probability distribution of reaching peaks of various fitness (Fig. 1a), which corresponds to Fig. 3E in Papkou et al. (2023). We found that the probability of reaching the highest 14% of all peaks is  $P_{14\%} = 76.4\%$ , virtually identical to the reported value (76.5%). We then conducted the same analysis in the DHFR landscape and considered the highest 14% of all 4,055 peaks in the DHFR landscape as high peaks, which yielded  $P_{14\%} = 69.1\%$  (Fig. 1a). Note that if 14% of the total number of peaks is not an integer, we considered the largest integer smaller than 14% of the total number of peaks to be the number of high peaks.

To investigate the generality of the above finding, we collected from the literature five representative empirical fitness landscapes (Table 1): (i) an *E. coli* Shine–Dalgarno sequence translational activity landscape of nine variable nucleotide sites each with four states (Kuo et al. 2020); (ii) a yeast tRNA fitness landscape of 10 variable nucleotide sites including six two-state sites and four three-state sites (Domingo et al. 2018); (iii) an ACE2 affinity landscape of SARS-CoV-2 Omicron BA.1 spike receptor-binding domain (RBD), with 15 variable nucleotide sites each with two states (Moulana et al. 2022); (iv) a sea anemone (*Entacmaea quadricolor*) fluorescent protein fluorescence intensity landscape of 13 variable amino acid sites each with two states (Poelwijk et al. 2019); and (v) a Streptococcal protein G domain B1 (GB1) immunoglobulin-binding affinity landscape of four variable amino acid sites each with 20 states (Wu et al. 2016). For simplicity, we refer to them as fitness landscapes with the understanding that fitness is approximated by different phenotypic values in these landscapes and the assumption that

**Table 1** Summary statistics of the empirical fitness landscapes studied

Landscapes	No. of variable sites	No. of genotypes included (completeness)	No. of peaks	Mean no. of peaks in corresponding maximally rugged landscapes $\pm$ SD	Ratio <sup>a</sup>	$\sigma$ of the corresponding RMF landscape
<i>E. coli</i> DHFR fitness	9 four-state sites	261,382 (99.7%)	4,055	9,342 $\pm$ 67	0.434	0.91
<i>E. coli</i> DHFR sublandscape	9 four-state sites	135,178 (N.A.)	514	5,913 $\pm$ 55 (1,487 $\pm$ 20) <sup>b</sup>	0.087 (0.346)	0.50
<i>E. coli</i> Shine–Dalgarno sequence translational activity	9 four-state sites	197,890 (75.5%)	2,388	8,932 $\pm$ 93	0.267	0.83
Yeast tRNA fitness	6 two-state sites and 4 three-state sites	4,176 (80.6%)	85	313 $\pm$ 11	0.272	0.71
SARS-CoV-2 Omicron BA.1 spike RBD affinity for ACE2	15 two-state sites	32,739 (99.9%)	135	2,042 $\pm$ 29	0.066	0.38
<i>E. quadricolor</i> fluorescent protein fluorescence intensity	13 two-state sites	8,192 (100%)	75	589 $\pm$ 13	0.127	0.62
Streptococcal GB1 immunoglobulin-binding affinity	4 twenty-state sites	149,362 (93.3%)	182	2,012 $\pm$ 34	0.090	0.32

<sup>a</sup>Ratio of the number of peaks in the empirical landscape to the mean number of peaks in 30 randomly generated corresponding maximally rugged landscapes.

<sup>b</sup>In the parentheses are values for corresponding maximally rugged landscapes generated by shuffling the fitness of “functional” genotypes only.

higher phenotypic values are selectively favored. We analyzed these five landscapes and drew the cumulative probability distribution of reaching peaks of various fitness as in the analysis of the DHFR landscape.

While the cumulative probability distributions of all six landscapes are more or less S-shaped, the specifics of the distributions differ greatly (Fig. 1). We find that although  $P_{14\%}$  exceeds 14% in every landscape examined, its variation among the six landscapes is substantial—from 32.4% to 96.9%. Because it was somewhat arbitrary to classify the highest 14% of all peaks as high peaks in DHFR, we respectively treated the global peak or highest 1%, 5%, or 25% of peaks as high peaks in additional analyses. The results (supplementary fig. S1, Supplementary Material online) are qualitatively similar to those of  $P_{14\%}$  (Fig. 1).

### Variation in the Probability of Reaching High Peaks Among Theoretical Fitness Landscapes

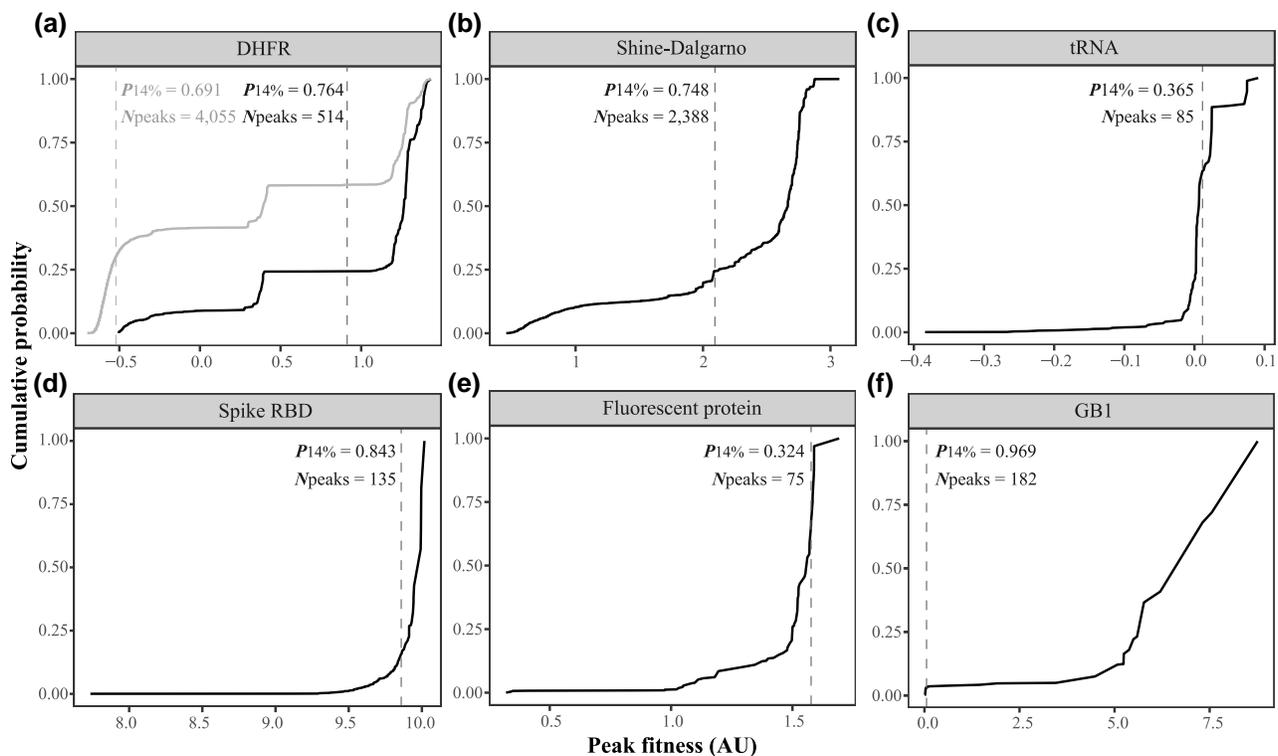
Because the above six empirical landscapes vary in many aspects (species, gene, number of variable sites, number of states per site, fitness measure, etc.), it is not easy to find why  $P_{HP}$  varies among them. We thus resorted to theoretical fitness landscapes, where we could study how  $P_{HP}$  changes with key landscape parameters.

We first generated a series of Rough Mount Fuji (RMF) landscapes (Aita et al. 2000; Aita and Husimi 2000; Neidhart et al. 2014) of nine variable sites each with four states, because two of the empirical landscapes (DHFR and Shine–Dalgarno) have

nine variable sites with four states per site. The ruggedness of an RMF landscape can be adjusted by the parameter  $\sigma \geq 0$ , which describes the extent to which the fitness effects of genetic changes are nonadditive (see Materials and Methods). An RMF landscape with  $\sigma = 0$  is smooth and contains only one peak. The larger the  $\sigma$ , the higher the nonadditivity and ruggedness of the RMF landscape. The cumulative probability distributions of reaching peaks of various fitness show that as  $\sigma$  increases from 0.5 to 10,  $P_{14\%}$  reduces from 96.3% to 26.4% (Fig. 2 and supplementary fig. S2, Supplementary Material online). Similarly,  $P_1$  (probability of reaching the global peak),  $P_{1\%}$ ,  $P_{5\%}$ ,  $P_{14\%}$ , and  $P_{25\%}$  decline as  $\sigma$  rises (supplementary fig. S3, Supplementary Material online). Hence, for the series of RMF landscapes examined, landscape ruggedness is an important determinant of  $P_{HP}$ .

We next investigated a series of NK landscapes (Kauffman and Weinberger 1989) of nine sites each with four states. We varied  $K$  from 0 to 8 to increase the landscape ruggedness from zero to maximum. In general,  $P_{14\%}$  declines from 100% to 24.8% as  $K$  increases from 0 to 8 (supplementary fig. S4, Supplementary Material online). Note, however, that  $P_{14\%}$  cannot be fairly compared when  $K=0$  (only 1 peak) or 1 (27 peaks) due to the small number of peaks. Similar results were seen for  $P_1$ ,  $P_{1\%}$ ,  $P_{5\%}$ , and  $P_{25\%}$  (supplementary fig. S5, Supplementary Material online).

We also simulated a series of fitness landscapes of 5 to 10 variable sites each with four states under the House of Cards (HoC) model (Kingman 1978), in which the fitness of a genotype is a random variable between 0 and 1. These



**Fig. 1.** Cumulative probabilities of reaching peaks of various fitness values by adaptive walks in six empirical fitness landscapes: *E. coli* DHFR fitness landscape/sublandscape (a), *E. coli* Shine–Dalgarno sequence translational activity landscape (b), yeast tRNA fitness landscape (c), landscape of SARS-CoV-2 Omicron BA.1 spike RBD affinity for ACE2 (d), *E. quadricolor* fluorescent protein fluorescence intensity landscape (e), and Streptococcal GB1 immunoglobulin-binding affinity landscape (f). In each panel, peaks are ordered from low to high fitness on the x axis, and the y axis shows the probability of reaching peaks with fitness no greater than a given value on the x axis. The black dashed line separates the highest 14% of peaks (to the right of the line) and the rest of the peaks. The probability of reaching the highest 14% of peaks ( $P_{14\%}$ ) and the number of peaks ( $N_{\text{peaks}}$ ) in each landscape are indicated. In a), the gray and black curves show the results in the DHFR landscape and subsandscape, respectively, and the left (gray) and right (black) dashed lines are for the DHFR landscape and subsandscape, respectively.

landscapes are all maximally rugged because the fitness difference between two genotypes is independent of the extent of their genotypic difference. We found that varying the number of variable sites has minimal impacts on  $P_{14\%}$ , which is always around 25% (supplementary fig. S6, Supplementary Material online). Varying the number of variable sites also has minimal impacts on  $P_1$ ,  $P_{1\%}$ ,  $P_{5\%}$ , and  $P_{2.5\%}$  (supplementary fig. S7, Supplementary Material online).

### Comparing $P_{HP}$ Between Empirical and RMF Landscapes

The analysis of theoretical landscapes revealed an important determinant of  $P_{HP}$ —landscape ruggedness. To assess whether the  $P_{HP}$  of an empirical landscape reflects the expected  $P_{HP}$  given its landscape ruggedness, we compared  $P_{HP}$  between an empirical landscape and a RMF landscape of the same size (i.e. same number of variable sites and number of states per site) and similar ruggedness quantified by the fraction of genotypes that are peaks (see Materials and Methods), because the RMF model was previously shown to reproduce key features of various empirical landscapes (Aita et al. 2000; Szendro et al. 2013; Neidhart et al. 2014). For instance, in an analysis of 10 empirical landscapes, Szendro et al. (2013) used the RMF model to interpolate the behavior of various ruggedness measures between the limits of a HoC model and an additive landscape and found good agreement with the trends in the empirical data.

We first created a series of RMF landscapes of nine four-state sites with different  $\sigma$ . By varying the cutoff that distinguished “functional” from “nonfunctional” genotypes, we followed Papkou et al. (2023) to identify the giant component in each of these RMF landscapes with  $\sim 135,178$  genotypes (see Materials and Methods). These giant components, referred to as RMF sublandscapes, were then compared with the DHFR sublandscape in terms of  $P_{HP}$ . We found that the DHFR sublandscape is only minimally rugged, corresponding to a same-size RMF sublandscape with  $\sigma = 0.50$  (Table 1). Interestingly,  $P_{HP}$  is generally smaller in the DHFR sublandscape than in the corresponding RMF sublandscape (see  $P_{5\%}$ ,  $P_{14\%}$ , and  $P_{2.5\%}$  in Fig. 3a and  $P_1$  and  $P_{1\%}$  in supplementary fig. S8a, Supplementary Material online).

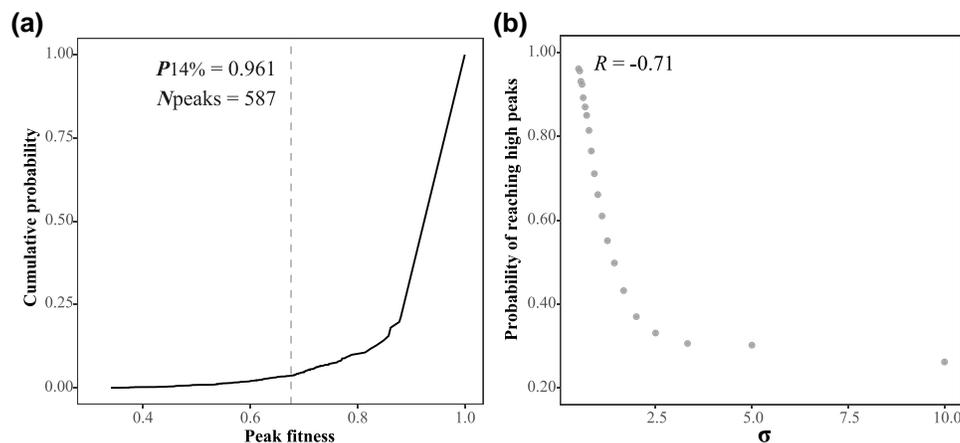
Another measure of the ruggedness of the DHFR sublandscape is its number of peaks as a fraction of the number of peaks in the corresponding HoC landscape, which can be generated by randomly shuffling the fitness of all genotypes in the DHFR sublandscape. This ruggedness measure is 8.7% based on 30 rounds of shuffling (Table 1), meaning that the number of peaks in the DHFR sublandscape is 8.7% of that in the corresponding maximally rugged landscape. Because the DHFR sublandscape was constructed with functional genotypes and their nonfunctional neighbors, it may be more appropriately compared with HoC landscapes created by shuffling the fitness of only functional genotypes. Based on this shuffling method, the number of peaks in the DHFR sublandscape is 34.6% of that in the corresponding maximally rugged landscape (Table 1).

We similarly compared RMF landscapes of nine four-state sites with the DHFR landscape. The DHFR landscape corresponds to a same-size RMF landscape with  $\sigma = 0.91$  in terms of the fraction of genotypes that are peaks (Table 1).  $P_{HP}$  is similar between the DHFR landscape and the corresponding RMF landscape (see  $P_{5\%}$ ,  $P_{14\%}$ , and  $P_{2.5\%}$  in Fig. 3b and  $P_1$  and  $P_{1\%}$  in supplementary fig. S8b, Supplementary Material online).

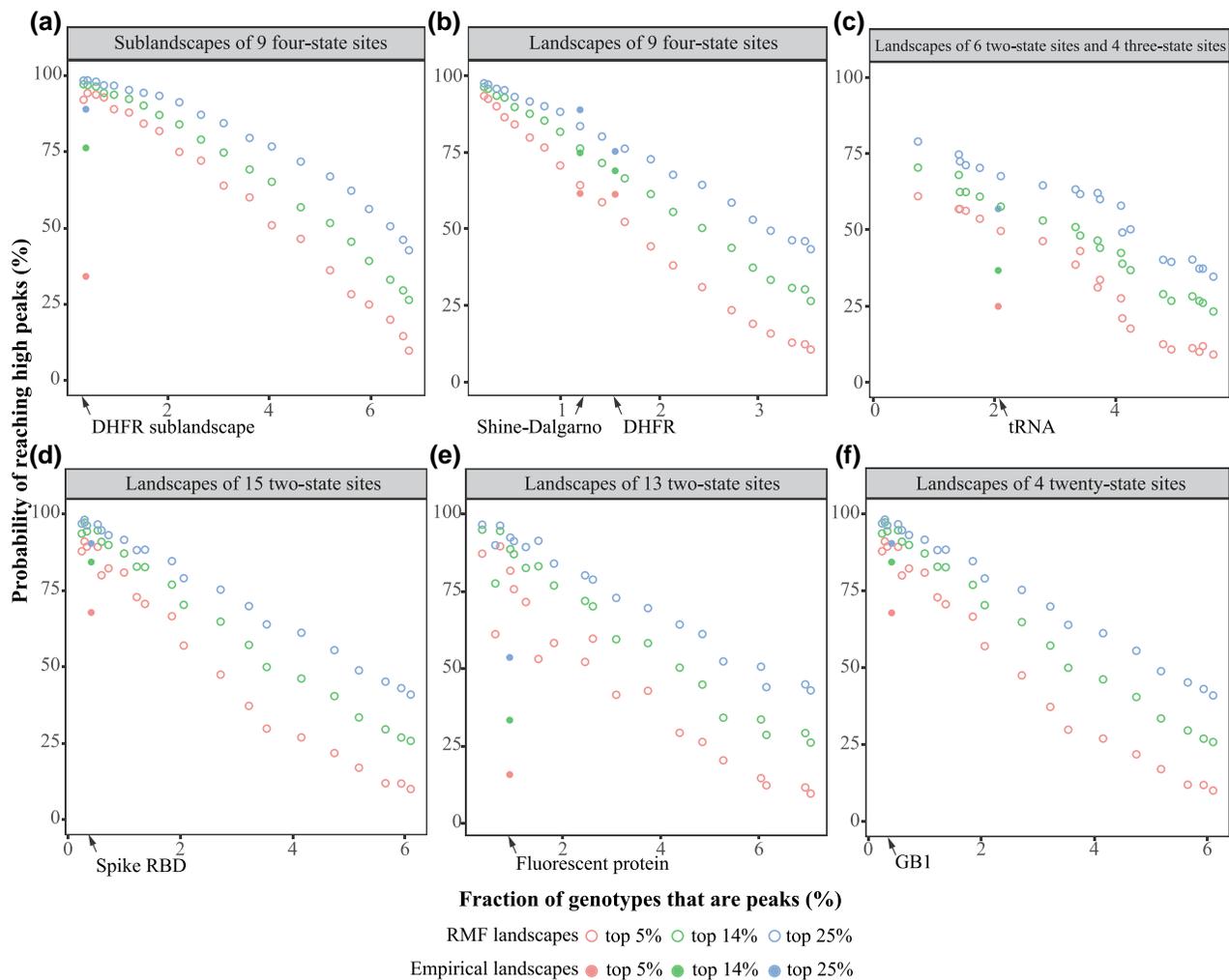
The Shine–Dalgarno sequence translational activity landscape corresponds to a same-size RMF landscape with  $\sigma = 0.83$  in terms of the fraction of genotypes that are peaks (Table 1), and its  $P_{HP}$  is similar to that in the corresponding RMF landscape (Fig. 3b and supplementary fig. S8b, Supplementary Material online).

$P_{HP}$  is relatively small for the yeast tRNA fitness landscape (Fig. 1c). We found that this landscape corresponds to a same-size RMF of  $\sigma = 0.71$  (Table 1).  $P_{HP}$  of the corresponding RMF landscape is also relatively small, although larger than that of the tRNA fitness landscape (Fig. 3c and supplementary fig. S8c, Supplementary Material online).

$P_{HP}$  is relatively large for the SARS-CoV-2 spike RBD affinity landscape (Fig. 1d). We found that this landscape corresponds to a same-size RMF of  $\sigma = 0.38$  (Table 1) and that  $P_{HP}$  of the corresponding RMF landscape is also relatively large (Fig. 3d and supplementary fig. S8d, Supplementary Material online).



**Fig. 2.** Negative correlation between the parameter  $\sigma$  and the probability of reaching high peaks by adaptive walks in RMF landscapes. a) Cumulative probability of reaching peaks of varying fitness in a RMF landscape with  $\sigma = 0.5$ . The dashed line separates the highest 14% of peaks (to the right of the line) and the rest of the peaks. The probability of reaching the highest 14% of peaks ( $P_{14\%}$ ) and the number of peaks ( $N_{\text{peaks}}$ ) in the landscape are indicated. b)  $P_{14\%}$  declines with  $\sigma$  in RMF landscapes. Pearson’s correlation ( $R$ ) between  $P_{14\%}$  and  $\sigma$  is indicated ( $P < 4.9 \times 10^{-4}$ ).



**Fig. 3.** Probability of reaching high peaks in an empirical landscape and that in its corresponding same-size RMF landscape of comparable ruggedness. High peaks refer to the highest 5%, 14%, or 25% of all peaks. In each panel, a series of RMF landscapes of different  $\sigma$  are simulated. For a given color, each open circle represents an RMF landscape, of which the fraction of genotypes that are peaks is indicated on the x axis and the probability of reaching high peaks is indicated on the y axis. Similarly, for a given color, a solid dot represents an empirical landscape of which the fraction of genotypes that are peaks is indicated on the x axis and the probability of reaching high peaks is indicated on the y axis. Comparisons for the DHFR sublandscape (a), DHFR landscape (b), Shine–Dalgarno landscape (b), tRNA landscape (c), spike RBD landscape (d), fluorescent protein landscape (e), and GB1 landscape (f).

For the fluorescent protein landscape, the corresponding same-size RMF landscape has a  $\sigma$  of 0.62 (Table 1). The observed  $P_{HP}$  is generally below the expected value from the corresponding RMF landscape (Fig. 3e and supplementary fig. S8e, Supplementary Material online).

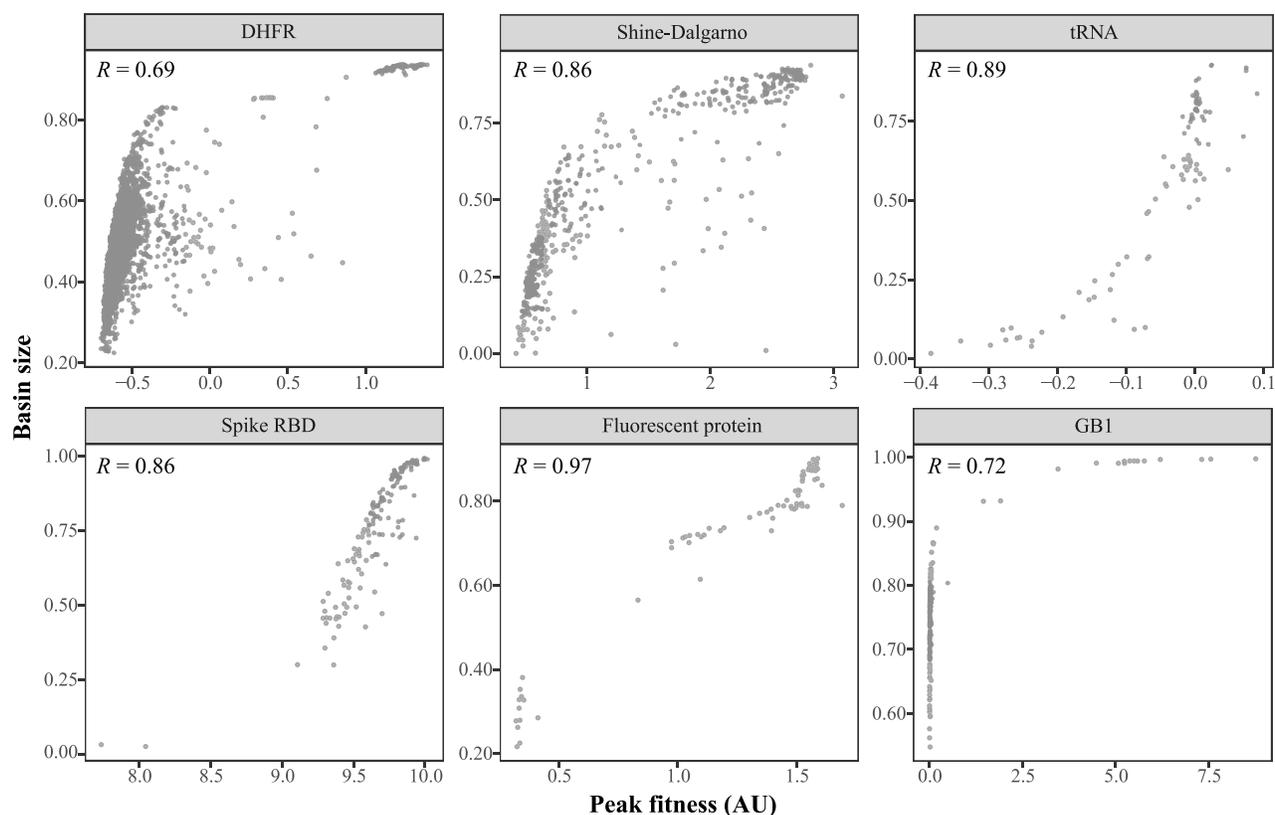
The GB1 immunoglobulin affinity landscape corresponds to a same-size RMF landscape of  $\sigma=0.32$  (Table 1). Hence, its large  $P_{HP}$  (Fig. 1f) is expected; in fact, its  $P_{HP}$  is below the expectation from the corresponding RMF landscape (Fig. 3f and supplementary fig. S8f, Supplementary Material online).

Empirical landscapes are usually incomplete, meaning that the fitness information is lacking for some genotypes. The presence of up to 24.5% of missing data in some of the empirical landscapes studied (Table 1) raises the question of whether our comparison between an empirical landscape and a same-size complete RMF landscape is appropriate. To address this question, we generated incomplete RMF landscapes by removing a fraction of genotypes in two different ways. In the first way, we removed 1%, 10%, or 25% of genotypes randomly. In the second way, we removed 1%, 10%, or 25% of the least fit genotypes, because fitness quantification tends to be difficult

or inaccurate for low-fitness genotypes, especially in high-throughput experiments (Li et al. 2016). We found that  $P_{HP}$  is hardly affected by the above fractions of missing data except when 25% of randomly chosen genotypes are removed; even then,  $P_{HP}$  is usually reduced only slightly (supplementary fig. S9, Supplementary Material online). For the DHFR sublandscape, we already compared it with RMF sublandscapes of similar numbers of genotypes (Fig. 3a). Taken together, the comparisons in Fig. 3 and supplementary fig. S8, Supplementary Material online are expected to be reliable.

### A Positive Correlation Between Peak Height and Basin Size is General but Does Not Guarantee a Large $P_{HP}$

Papkou et al. discovered in their analysis of the DHFR sublandscape that the fitness of a peak positively correlates with its basin size, which is the number of genotypes that can reach the peak through adaptive walks. They suggested that this positive correlation may explain their observed large  $P_{HP}$ . The positive correlation between peak height and basin size



**Fig. 4.** Correlation between peak fitness and basin size in empirical landscapes. Each dot is a peak, with the x axis showing its fitness and y axis showing its basin size measured by the fraction of genotypes in the landscape that can reach the peak by an adaptive walk. Pearson's correlation coefficient ( $R$ ) is indicated.  $R$  is significantly different from 0 in all panels ( $P < 2.2 \times 10^{-16}$  in all panels). Note that the result for the DHFR sublandscape is presented.

( $R_{\text{ph-bs}}$ ) makes intuitive sense, because everything else being equal, a fitter genotype should on average be more likely than a less fit genotype to be the end point of an adaptive walk owing to the simple fact that natural selection pushes a population to higher fitness. We thus examined whether a positive  $R_{\text{ph-bs}}$  exists in the empirical landscapes studied here. Because Papkou et al. already examined the DHFR sublandscape, we here analyzed the DHFR landscape instead. Indeed,  $R_{\text{ph-bs}}$  ranges from 0.69 to 0.97 in the six empirical landscapes. Interestingly, the fluorescent protein landscape has the largest  $R_{\text{ph-bs}}$  (Fig. 4) but smallest  $P_{14\%}$  (Fig. 1) among the six landscapes, although  $R_{\text{ph-bs}}$  and  $P_{14\%}$  are not significantly correlated ( $R = -0.75$ ,  $P = 0.086$ ).

To investigate the relationships among landscape ruggedness,  $R_{\text{ph-bs}}$ , and  $P_{\text{HP}}$ , we examined a series of RMF landscapes of nine sites each with four states. We found that as  $\sigma$  and ruggedness rise,  $R_{\text{ph-bs}}$  (Fig. 5 and supplementary fig. S10, Supplementary Material online) and  $P_{\text{HP}}$  (supplementary figs. S2 and S3, Supplementary Material online) decrease. That  $R_{\text{ph-bs}}$  tends to decrease with landscape ruggedness is generally true for NK landscapes of nine sites each with four states (supplementary fig. S11, Supplementary Material online). The HoC landscapes examined have different numbers of variable sites but similar ruggedness and  $P_{\text{HP}}$  (supplementary fig. S7, Supplementary Material online), and their  $R_{\text{ph-bs}}$  values are similar too (supplementary fig. S12, Supplementary Material online).

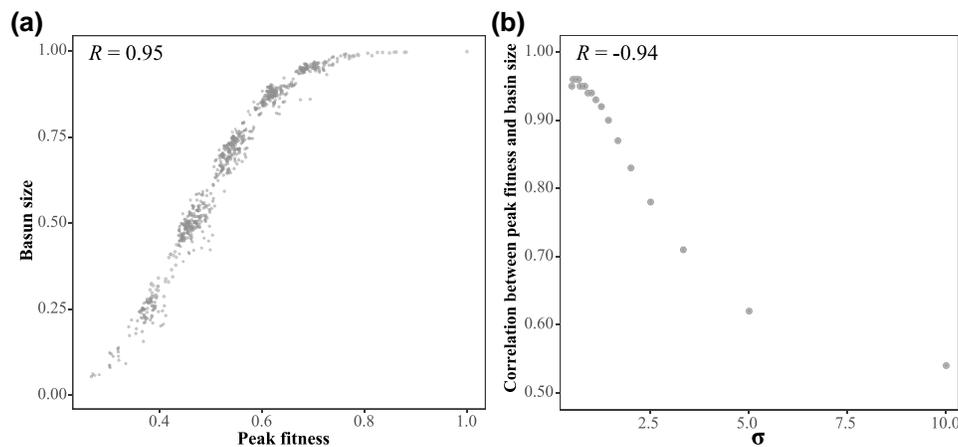
In all empirical and theoretical landscapes examined except for a smooth NK landscape,  $R_{\text{ph-bs}}$  is significantly positive and is at least moderate in size. Yet,  $P_{14\%}$  varies substantially among these landscapes. Therefore, although a positive  $R_{\text{ph-bs}}$  may

explain why the probability of reaching high peaks exceeds the fraction of peaks considered high, it does not guarantee a  $P_{14\%}$  that is much higher than 14% as in the DHFR sublandscape. Using the simulated RMF landscapes (Fig. 5 and supplementary fig. S2, Supplementary Material online), we performed a multiple linear regression in which  $P_{14\%}$  is the dependent variable and landscape ruggedness (measured by the fraction of genotypes that are peaks) and  $R_{\text{ph-bs}}$  are two independent variables. In the regression, the coefficient for ruggedness is  $\sim 63$  times the coefficient for  $R_{\text{ph-bs}}$ , suggesting that ruggedness plays a much greater role than  $R_{\text{ph-bs}}$  in determining  $P_{14\%}$ . Note that in the above regression, we normalized ruggedness and  $R_{\text{ph-bs}}$  so that they both have a mean of 0 and a SD of 1 to allow a fair comparison. Qualitatively similar results were obtained when the dependent variable was  $P_1$ ,  $P_{1\%}$ ,  $P_{5\%}$ , or  $P_{25\%}$ .

## Discussion

Our investigation of empirical and theoretical fitness landscapes revealed that the probability of reaching high peaks ( $P_{\text{HP}}$ ) by an adaptive walk is generally greater than the proportion of the highest peaks that are considered high peaks. Nonetheless,  $P_{\text{HP}}$  varies substantially among empirical as well as theoretical landscapes, and landscape ruggedness is a major determinant of  $P_{\text{HP}}$ , with higher ruggedness generally corresponding to smaller  $P_{\text{HP}}$ .

As mentioned, the proportion of the highest peaks considered to be high peaks by Papkou et al. (2023) was somewhat arbitrary. That is why we considered multiple proportions (top 1, top 1%, 5%, 14%, and 25%) in our analysis; the results obtained,



**Fig. 5.** Correlation between peak fitness and basin size in RMF landscapes of nine four-state sites. a) Correlation between peak fitness and basin size when  $\sigma = 0.5$ . Each dot is a peak, with the x axis showing its fitness and y axis showing its basin size measured by the fraction of genotypes in the landscape that can reach the peak by an adaptive walk. Pearson's correlation ( $R$ ) is indicated ( $P < 2.2 \times 10^{-16}$ ). b) Relationship between parameter  $\sigma$  and the correlation between peak fitness and basin size across RMF landscapes. Pearson's correlation ( $R$ ) between  $\sigma$  and the correlation between peak fitness and basin size is indicated ( $P < 7.6 \times 10^{-10}$ ).

however, were qualitatively similar. One may argue that the proportion of the highest peaks considered high peaks should be landscape-specific, depending on the fitness distribution of all peaks in the landscape. We do not disagree. However, such definitions of high peaks would make it difficult to compare among empirical landscapes as well as between empirical and theoretical landscapes. We thus leave the investigation of a more realistic definition of high peaks and its impact on  $P_{\text{HP}}$  to future research.

In our study, the theoretical fitness landscapes were simulated without error, but the fitness values in the empirical landscapes were experimentally measured so must contain error. A previous systematic analysis found that random fitness estimation errors cause an overestimation of landscape ruggedness (Song and Zhang 2021), which is predicted to cause an underestimation of  $P_{\text{HP}}$ . In other words, the true  $P_{\text{HP}}$  may be larger than estimated here and in Papkou et al. (2023). Indeed, Papkou et al. (2023) commented that their finding of a large  $P_{\text{HP}}$  is qualitatively unchanged when measurement error is considered. However, most of the empirical landscapes analyzed here do not contain measurement error information, so the impact cannot be directly assessed using standard methods (Franke et al. 2011). This said, we note that our key finding through comparing  $P_{\text{HP}}$  between an estimated empirical landscape and its corresponding same-size RMF landscape of similar ruggedness is most likely unaffected, provided that the relationship between landscape ruggedness and  $P_{\text{HP}}$  is not drastically different between the empirical and RMF landscapes near the ruggedness of the empirical landscape.

We compared empirical landscapes with corresponding same-size RMF landscapes of similar ruggedness because the RMF model was shown to reproduce key features of various empirical landscapes (Aita et al. 2000; Szendro et al. 2013; Neidhart et al. 2014). There is no doubt that the RMF model is simple and cannot reproduce all aspects of empirical landscapes. However, without evidence of the superiority of other theoretical landscapes over the RMF model, our comparison between empirical and RMF landscapes seems appropriate. This comparison demonstrated that  $P_{\text{HP}}$  of an empirical landscape is generally similar to or smaller than that of its corresponding RMF landscape, providing no evidence for higher-than-expected evolvability in empirical landscapes. This means that the observation of a large  $P_{14\%}$  in the DHFR sublandscape (Papkou et al. 2023) is not unexpected given the sublandscape's

size and ruggedness. The initial surprise at the large  $P_{14\%}$  of DHFR was likely due to our ignorance on its expected value.

In both empirical and theoretical landscapes, we found landscape ruggedness to be a major determinant of the probability of reaching high peaks, consistent with prior theoretical results, which showed that landscape ruggedness plays a critical role in shaping evolutionary dynamics and that increased landscape ruggedness constrains evolvability (Franke et al. 2011; Franke and Krug 2012; Manhart and Morozov 2014; Neidhart et al. 2014; Ferretti et al. 2018; Krug and Oros 2024).

In every empirical or theoretical landscape examined, we observed a positive correlation between the fitness of a peak and its basin size that was initially reported in DHFR by Papkou et al. (2023). While this positive correlation may explain why  $P_{\text{HP}}$  exceeds the fraction of the highest peaks considered high peaks, it does not guarantee a large  $P_{\text{HP}}$  due to the influences of other factors. Specifically, in the analysis of RMF landscapes, we found that  $P_{\text{HP}}$  is influenced substantially more by landscape ruggedness than by the above correlation, underscoring the primary impact of landscape ruggedness on  $P_{\text{HP}}$ .

We focused on low-dimension fitness landscapes (i.e. with relatively few variable sites) because all high-dimension, empirical fitness landscapes contain a substantial fraction of missing data. However, real-world fitness landscapes have orders of magnitude more variable sites than those considered here and the applicability of our findings to real-world fitness landscapes is unclear. For instance, fitness peaks inaccessible in low-dimension landscapes often become accessible in landscapes of extremely high dimensions (Greenbury et al. 2022), and crossing fitness valleys becomes less of a problem than finding a fitter genotype in the sea of many equally fit genotypes (Van Nimwegen and Crutchfield 2000; Greenbury et al. 2022). Major improvements are needed to make empirical fitness landscape studies more relevant to real-world fitness landscapes.

## Materials and Methods

### Empirical Landscapes

We considered six empirical landscapes. The published *E. coli* DHFR gene fitness landscape has nine variable nucleotide sites each with four states, corresponding to three amino acids in DHFR (Papkou et al. 2023). The DHFR landscape contains

261,382 genotypes, whereas the DHFR sublandscape contains 135,178 genotypes. Kuo et al. (2020) measured the translational activity of 197,890 variants of the Shine–Dalgarno sequence in *E. coli*, producing a translational activity landscape of nine nucleotide sites each with four states. Domingo et al. (2018) measured the fitness of 4,176 variants of a yeast tRNA gene, producing a fitness landscape of 10 variable nucleotide sites including six two-state sites and four three-state sites. Moulana et al. (2022) measured the ACE2-binding affinity of 32,739 variants of the SARS-CoV-2 Omicron BA.1 spike RBD, producing a binding affinity landscape of 15 variable nucleotide sites each with two states. Poelwijk et al. (2019) measured the fluorescence intensity of 8,192 variants of a fluorescent protein from *E. quadricolor*, producing a fluorescence intensity landscape of 13 variable amino acid sites each with two states. Lastly, Wu et al. (2016) measured the immunoglobulin-binding affinity of 149,361 variants of a Streptococcal protein GB1, producing a binding affinity landscape of four variable amino acid sites each with 20 states. For simplicity, we refer to all these landscapes as fitness landscapes, where fitness is approximated by different phenotypes in different landscapes.

Note that, in empirical landscapes, some genotypes do not have corresponding fitness values due to missing data in experiments. When counting peaks and performing adaptive walks, we do not consider these genotypes as if they do not exist. We assessed the impact of missing data on our analysis by artificially removing various fractions of genotypes from theoretical landscapes (supplementary fig. S9, Supplementary Material online).

### RMF Landscapes

In RMF landscapes (Aita et al. 2000; Aita and Husimi 2000; Neidhart et al. 2014), the fitness of genotype  $x$  equals  $F(x) = -D(x, x^*) + \eta_x$ , where  $x^*$  is the genotype with the highest fitness,  $D(x, x^*)$  is the Hamming distance (i.e. number of sites that differ) between  $x$  and  $x^*$ , and  $\eta_x$  is a random fitness effect sampled from a normal distribution with mean = 0.5 and SD =  $\sigma$ . All fitness values in each landscape were then linearly normalized to the range between 0 and 1.

We simulated RMF sublandscapes for comparison with the DHFR sublandscape as follows. First, we simulated a RMF landscape of nine four-state sites with a given  $\sigma$ . Second, from this landscape, we designated the  $m$  genotypes with the lowest fitness as nonfunctional and the rest as functional. Third, we removed from the landscape all nonfunctional genotypes with no functional neighbors. Fourth, we identified the giant component from the network composed of the remaining genotypes and counted the number ( $n$ ) of genotypes in the giant component. We repeated steps 2 to 4 with varying  $m$  until  $n$  approached 135,178, the size of the DHFR sublandscape. For the RMF sublandscapes considered in Fig. 3a, the final  $n$  deviated from 135,178 by a mean of 3.8% and a maximum of 6.5%.

When comparing an empirical landscape with its corresponding same-size RMF landscape, we required that they have similar fractions of peaks. Specifically, we computed the absolute value of the difference between the fraction of peaks in the empirical landscape and that in the RMF landscape, which is then divided by the fraction of peaks in the empirical landscape. The above value was on average 0.026 in all analyses, with a maximum of 0.063.

### NK Landscapes

In the NK model, parameter  $N$  denotes the genotype length or the number of variable sites whereas parameter  $K$ , which ranges from 0 to  $N-1$ , denotes the number of other sites with which each focal site interacts (Kauffman and Weinberger 1989; Kauffman 1993). As  $K$

increases, the ruggedness of the landscape increases. We followed the code available at [https://github.com/Mac13 kW/NK\\_model](https://github.com/Mac13 kW/NK_model) to generate our NK landscapes (Hwang et al. 2018; Song and Zhang 2021). The fitness of genotype  $x$  equals  $F(x) = \frac{1}{N} \sum_{i=1}^N F_i(x)$ ,

where  $F_i(x)$  is the fitness associated with the  $i$ th site. When  $K=0$ , each site in the sequence acts independently, resulting in a single-peak landscape. As  $K$  increases, the landscape ruggedness rises because each site's fitness contribution depends on its state and the states of  $K$  other (randomly picked) sites. All fitness values in each landscape were linearly normalized to the range between 0 and 1.

Papkou et al. (2023) stated that the DHFR sublandscape corresponds to an NK landscape of the maximum ruggedness (i.e. a HoC landscape) in terms of the number of peaks. They arrived at this conclusion by multiplying the expected fraction of peaks (1/28) in a maximally rugged NK landscape of nine four-state sites (Kauffman and Levin 1987) by the number of functional genotypes (18,029) in the sublandscape. However, their calculation was incorrect because they considered only functional genotypes in the sublandscape but the chance probability for a functional genotype to be a peak is higher than that (1/28) for a random genotype.

### HoC Landscapes

In HoC landscapes, the fitness of a genotype is a random variable sampled from a normal distribution with mean = 0.5 and SD = 0.1; random variables outside the range from 0 to 1 are discarded. Because there is no relationship between the similarity of two genotypes and that of their fitness values, HoC landscapes have the maximal ruggedness (Kingman 1978).

### Landscape Shuffling

We compared each empirical landscape with its corresponding maximally rugged landscape created by shuffling the fitness values of all genotypes in the empirical landscape. The mean number of peaks and the SD from 30 shuffled landscapes are reported in Table 1 for each empirical landscape. For the DHFR sublandscape, we additionally created its corresponding maximally rugged landscapes by shuffling the fitness values of all “functional” genotypes.

### Simulating Adaptive Walks Under the Strong Selection, Weak Mutation Regime

Adaptive walks were simulated under the strong selection, weak mutation (SSWM) regime. That is, at each step, a neighboring genotype fitter than the current genotype was chosen randomly, and this process was repeated until a fitness peak was reached. We simulated one adaptive walk from each of  $10^6$  randomly picked starting genotypes. We also simulated adaptive walks using Kimura's fixation probability (Kimura 1983; de Oliveira and Campos 2004) and found virtually identical results, as noted previously (Papkou et al. 2023). Using the SSWM regime is likely more appropriate than using Kimura's formula given that the phenotype considered is not fitness in most of the landscapes studied here.

### Basin Size

The basin of a fitness peak comprises all variants that can reach the peak by an adaptive walk. Specifically, we first identify neighbors of the peak that can reach the peak by an adaptive walk. For each of these “viable” neighbors, we then find its neighbors that can reach it by an adaptive walk. This

iterative analysis allows finding all genotypes that can reach the peak by an adaptive walk. Note that we require each adaptive walk to have a monotonic increase in fitness.

## Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

## Acknowledgments

The authors thank Siliang Song for technical assistance, Maciej Workiewicz and Andrei Papkou for answering questions about their computer code or publication, and Siliang Song, Daohan Jiang, and two reviewers for valuable comments on earlier versions of the manuscript.

## Funding

This work was supported by the research grant R35GM139484 from the U.S. National Institutes of Health to J.Z.

## Data Availability

Computer code and data are available at GitHub (<https://github.com/yangli-evo/theoretical-fitness-landscapes>).

## References

- Aita T, Husimi Y. Adaptive walks by the fittest among finite random mutants on a Mt. Fuji-type fitness landscape II. Effect of small non-additivity. *J Math Biol*. 2000;41:207–231. <https://link.springer.com/article/10.1007/s002850000046>.
- Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: application to prolyl endopeptidase and thermolysin. *Biopolymers*. 2000;54:64–79. [https://doi.org/10.1002/\(SICI\)1097-0282\(200007\)54:1<64::AID-BIP70>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-0282(200007)54:1<64::AID-BIP70>3.0.CO;2-R).
- Bank C. Epistasis and adaptation on fitness landscapes. *Annu Rev Ecol Syst*. 2022;53:457–479. <https://doi.org/10.1146/annurev-ecolsys-102320-112153>.
- de Oliveira VM, Campos PR. Dynamics of fixation of advantageous mutations. *Phys A: Stat Mech Appl*. 2004;337:546–554. <https://doi.org/10.1016/j.physa.2004.02.007>.
- De Visser JAG, Krug J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet*. 2014;15:480–490. <http://doi.org/10.1038/nrg3744>.
- Domingo J, Diss G, Lehner B. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*. 2018;558:117–121. <http://doi.org/10.1038/s41586-018-0170-7>.
- Ferretti L, Weinreich D, Tajima F, Achaz G. Evolutionary constraints in fitness landscapes. *Heredity (Edinb)*. 2018;121:466–481. <http://doi.org/10.1038/s41437-018-0110-1>.
- Franke J, Klözer A, de Visser JAG, Krug J. Evolutionary accessibility of mutational pathways. *PLOS Comp Biol*. 2011;7(8):e1002134. <https://doi.org/10.1371/journal.pcbi.1002134>.
- Franke J, Krug J. Evolutionary accessibility in tunably rugged fitness landscapes. *J Stat Phys*. 2012;148:706–723. <http://doi.org/10.1007/s10955-012-0507-9>.
- Greenbury SF, Louis AA, Ahnert SE. The structure of genotype-phenotype maps makes fitness landscapes navigable. *Nat Ecol Evol*. 2022;6(11):1742–1752. <https://doi.org/10.1038/s41559-022-01867-z>.
- Hwang S, Schmiegel B, Ferretti L, Krug J. Universality classes of interaction structures for NK fitness landscapes. *J Stat Phys*. 2018;172(1):226–278. <https://doi.org/10.1007/s10955-018-1979-z>.
- Kauffman S, Levin S. Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol*. 1987;128(1):11–45. [https://doi.org/10.1016/S0022-5193\(87\)80029-2](https://doi.org/10.1016/S0022-5193(87)80029-2).
- Kauffman SA. *The origins of order: self-organization and selection in evolution*. USA: Oxford University Press; 1993.
- Kauffman SA, Weinberger ED. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J Theor Biol*. 1989;141(2):211–245. [https://doi.org/10.1016/S0022-5193\(89\)80019-0](https://doi.org/10.1016/S0022-5193(89)80019-0).
- Kimura M. *The neutral theory of molecular evolution*. Cambridge University Press; 1983.
- Kingman JF. A simple model for the balance between selection and mutation. *J Appl Probab*. 1978;15(1):1–12. <https://doi.org/10.2307/3213231>.
- Krug J, Oros D. Evolutionary accessibility of random and structured fitness landscapes. *J Stat Mech: Theory Exp*. 2024;2024(3):034003. <https://doi.org/10.1088/1742-5468/ad3197>.
- Kuo S-T, Jahn R-L, Cheng Y-J, Chen Y-L, Lee Y-J, Hollfelder F, Wen J-D, Chou H-HD. Global fitness landscapes of the Shine-Dalgarno sequence. *Genome Res*. 2020;30(5):711–723. <https://doi.org/10.1101/gr.260182.119>.
- Li C, Qian W, Maclean CJ, Zhang J. The fitness landscape of a tRNA gene. *Science*. 2016;352(6287):837–840. <https://doi.org/10.1126/science.aae0568>.
- Lyons DM, Zou Z, Xu H, Zhang J. Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. *Nat Ecol Evol*. 2020;4(12):1685–1693. <https://doi.org/10.1038/s41559-020-01286-y>.
- Manhart M, Morozov AV. Statistical physics of evolutionary trajectories on fitness landscapes. In: *First-passage phenomena and their applications*. Singapore: World Scientific; 2014. p. 416–446.
- McCandlish DM. Visualizing fitness landscapes. *Evolution*. 2011;65(6):1544–1558. <https://doi.org/10.1111/j.1558-5646.2011.01236.x>.
- Moulana A, Dupic T, Phillips AM, Chang J, Nieves S, Roffler AA, Greaney AJ, Starr TN, Bloom JD, Desai MM. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat Commun*. 2022;13(1):7011. <https://doi.org/10.1038/s41467-022-34506-z>.
- Nahum JR, Godfrey-Smith P, Harding BN, Marcus JH, Carlson-Stevermer J, Kerr B. A tortoise-hare pattern seen in adapting structured and unstructured populations suggests a rugged fitness landscape in bacteria. *Proc Natl Acad Sci U S A*. 2015;112(24):7530–7535. <https://doi.org/10.1073/pnas.1410631112>.
- Neidhart J, Szendro IG, Krug J. Adaptation in tunably rugged fitness landscapes: the rough Mount Fuji model. *Genetics*. 2014;198(2):699–721. <https://doi.org/10.1534/genetics.114.167668>.
- Orr HA. The genetic theory of adaptation: a brief history. *Nat Rev Genet*. 2005;6(2):119–127. <https://doi.org/10.1038/nrg1523>.
- Orr HA. Fitness and its role in evolutionary genetics. *Nat Rev Genet*. 2009;10(8):531–539. <https://doi.org/10.1038/nrg2603>.
- Papkou A, Garcia-Pastor L, Escudero JA, Wagner A. A rugged yet easily navigable fitness landscape. *Science*. 2023;382(6673):eadh3860. <https://doi.org/10.1126/science.adh3860>.
- Poelwijk FJ, Socolich M, Ranganathan R. Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat Commun*. 2019;10(1):4213. <https://doi.org/10.1038/s41467-019-12130-8>.
- Song S, Zhang J. Unbiased inference of the fitness landscape ruggedness from imprecise fitness estimates. *Evolution*. 2021;75(11):2658–2671. <https://doi.org/10.1111/evo.14363>.
- Szendro IG, Schenk MF, Franke J, Krug J, De Visser JAG. Quantitative analyses of empirical fitness landscapes. *J Stat Mech: Theory Exp*. 2013;2013((0111)):P01005. <https://doi.org/10.1088/1742-5468/2013/01/P01005>.
- Van Nimwegen E, Crutchfield JP. Metastable evolutionary dynamics: crossing fitness barriers or escaping via neutral paths? *Bull Math Biol*. 2000;62(5):799–848. <https://doi.org/10.1006/bulm.2000.0180>.
- Weinreich DM, Watson RA, Chao L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*. 2005;59:1165–1174. <https://doi.org/10.1111/j.0014-3820.2005.tb01768.x>.
- Wright S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*. 1932;1:355–366.
- Wu NC, Dai L, Olson CA, Lloyd-Smith JO, Sun R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife*. 2016;5:e16965. <https://doi.org/10.7554/eLife.16965>.