The structure of genotype-phenotype maps makes fitness landscapes navigable

Sam F. Greenbury,[1, 2, 3] Ard A. Louis,[4] and Sebastian E. Ahnert[5, 6]

[1]*Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, UK*
[2]*NIHR Imperial Biomedical Research Centre, ITMAT Data Science Group, Imperial College London, UK*
[3]*Department of Metabolism, Digestion and Reproduction, Imperial College London, UK*
[4]*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, UK*
[5]*Department of Chemical Engineering and Biotechnology, University of Cambridge,*
Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom
[6]*The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB, United Kingdom*

Fitness landscapes are often described in terms of 'peaks' and 'valleys', implying an intuitive low-dimensional landscape of the kind encountered in everyday experience. The space of genotypes, however, is extremely high-dimensional, which results in counter-intuitive properties of genotype-phenotype maps, such as the close proximity of one phenotype to many others. Here we investigate how common structural properties of high-dimensional genotype-phenotype maps, such as the presence of neutral networks, affect the navigability of fitness landscapes. For three biologically realistic genotype-phenotype map models—RNA secondary structure, protein tertiary structure and protein complexes—we find that, even under random fitness assignment, fitness maxima can be reached from almost any other phenotype without passing through a fitness valley. This in turn implies that true fitness valleys are very rare. By considering evolutionary simulations between pairs of real examples of functional RNA sequences, we show that accessible paths are also likely to be utilised under evolutionary dynamics.

I. INTRODUCTION

Ever since they were first introduced in Sewall Wright's foundational paper [1], fitness landscapes have become an enduring and central concept in evolutionary biology [2–6]. In particular, a low-dimensional picture of fitness 'peaks' and fitness 'valleys' has played an important role in shaping intuition around evolutionary dynamics. A key prediction is that a population must typically traverse an unfavourable valley of lower fitness to move from one fitness peak to another. But, as already pointed out by Fisher [7] and many others since [4, 8–11], the space of genotypes is typically extremely high dimensional. As illustrated in Fig. 1, what appears to be a fitness valley in a lower-dimensional landscape could be easily bypassed when dimensions are added [9–11].

Two key open questions are: 1) Does the low-dimensional picture of fitness valleys hold for realistic high-dimensional genotype spaces? And 2), if we define *accessible paths* of point mutations between a low fitness phenotype and a high fitness phenotype as those with monotonically increasing fitness, are such paths sufficiently common that they can easily be found by an evolving population?

One way forward is to consider empirical fitness landscapes, where much recent progress has been made [5, 12], particularly for molecular phenotypes [5, 13–21]. This body of work has yielded important insights, such as the role of local epistatic interactions in sculpting evolutionary paths [22–24]. Nevertheless, ruling out high-dimensional bypasses is difficult in empirical studies because genotype spaces, which grow exponentially as K^L for alphabet size K and genotype length L, are almost always unimaginably vast [25]. They are also highly connected since distances are linear; two genotypes are at most L point mutations away, but are connected by up to $L!$ possible paths. For example, even for a very short $L = 20$ strand of RNA , there are up to $20! \approx 2 \times 10^{18}$ paths between any two genotypes. Empirical landscapes can typi-
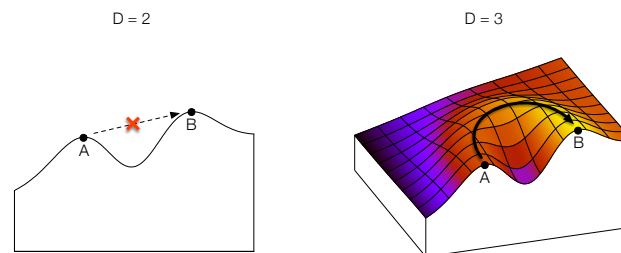

FIG. 1. Illustration of how increasing dimensionality can affect the navigability and presence of valleys in a fitness landscape.

cally only ever sample a small fraction of the full genotype space, so what may appear to be an isolated fitness peak, may in fact be accessible but the pathways are not feasible to experimentally identify.

A different strand of work, which can in principle address questions of global accessibility, has focused on model genotype-to-fitness landscapes [3, 6, 10, 11, 26, 27]. If fitness is assigned randomly to genotypes, as in Kingman's 'house of cards' model [28], then the probability of finding accessible paths is small. If instead there are correlations between fitness and the genotypes, then, depending on details of the model, accessible paths can be common [11, 29]. While again much progress has been made in this literature, it is not always clear how well these models capture true biological fitness landscapes.

Here we take a different approach, and build upon recent advances showing that many realistic genotype-phenotype (GP) maps share key structural features that enhance navigability. [30–32]. One important commonality is the existence of large neutral networks of genotypes that map to the same phenotype. Because of these networks, the mutational robustness ρ_p of a phenotype p (defined as the mean probability that

a point mutation leaves the phenotype unchanged) typically scales as the logarithm of phenotype frequency f_p (defined as fraction of genotype space occupied by phenotype p) for a wide range of GP maps [30–32]. If the genotypes of a phenotype were randomly distributed in genotype space, then the robustness would scale as $\rho_p \approx f_p$, which is much smaller than observed, highlighting the presence of neutral correlations in many realistic GP maps [30]. Large neutral networks play an important role in evolution because they allow all adjacent phenotypes of a neutral network to be reached by point mutations from any individual genotype in that network [31–34], and therefore may form part of accessible paths.

Our main contribution here is to show that commonly observed structural properties of GP maps greatly increase the number of accessible paths, or 'navigability', in associated fitness landscapes. In contrast to the genotype-to-fitness models studied by others (see above), we consider the genotype-phenotype (GP) map with the phenotype-to-fitness map as an additional layer on top. We first explore specific features of GP maps that affect the navigability: redundancy (large neutral sets), frequency of the unfolded or trivial phenotype, neutral correlations and high-dimensionality, and the effect of these quantities on the ruggedness of the landscape. We then focus on identifying whether accessible paths exist for fRNA phenotypes identified *in vivo* from the fRNA database [35], and simulate evolutionary dynamics to explore whether accessible paths might be utilised in biological evolution. Our findings show that certain structural properties of GP maps give rise to navigable fitness landscapes, and that the resulting accessible paths are indeed likely to be exploited in the course of biological evolution.

II. RESULTS

A. Several well-studied genotype-phenotype maps induce navigable fitness landscapes

A wide range of different GP maps share common structural properties, including a much larger number of genotypes than phenotypes (redundancy), a heavily skewed distribution in the number of genotypes per phenotype (phenotype bias), and close proximity of genotypes belonging to the same phenotype (which can also be described in terms of positive neutral correlations or large phenotypic robustness) [30, 31]. Here we consider the RNA secondary structure GP map for sequences of lengths $L = 12$ and $L = 15$ (RNA12, RNA15) [36–42], the Polyomino lattice self-assembly GP map ($S_{2,8}$, $S_{3,8}$) [30, 43], and several HP lattice protein folding GP maps (two compact GP maps HP5x5 and HP3x3x3, and two non-compact ones HP20 and HP25) [44–46].

We performed computational experiments in which fitness is assigned to phenotypes randomly, and two phenotypes are chosen randomly from the set of all phenotypes as the 'source' and 'target'.

TABLE I. RNA, Polyomino and compact HP GP maps all have navigable fitness landscapes ($\langle \psi \rangle > 0.6$) under random fitness assignment illustrating a lack of fitness valleys. By contrast, non-compact HP models have very low navigability ($\langle \psi \rangle \leq 0.013$).

GP map	K	L	N_P	f_{del}	$\log_{10} R$	$\langle \psi \rangle \pm SE(\langle \psi \rangle)$
RNA12	4	12	58	0.854	4.6	0.966 ± 0.002
RNA15	4	15	431	0.650	5.9	0.978 ± 0.002
$S_{2,8}$	8	8	14	0.537	5.8	0.913 ± 0.008
$S_{3,8}$	8	12	147	0.800	8.0	0.919 ± 0.003
HP5x5	2	25	549	0.816	4.1	0.995 ± 0.001
HP3x3x3	2	27	49,807	0.939	2.2	0.669 ± 0.005
HP20	2	20	5,311	0.976	0.7	0.004 ± 0.001
HP25	2	25	107,337	0.977	0.9	0.013 ± 0.001

The navigability $\langle \psi \rangle$ is defined as:

$$\langle \psi \rangle = \frac{1}{N} \sum_{k}^{N} \psi_{s_k t_k}$$

over a set of N source-target pairs (s_k, t_k), where ψ_{ij} is the probability that single-point mutation steps with monotonically increasing fitness (an accessible path) exist from a genotype of phenotype i to a genotype of phenotype j. In other words, the navigability is the average probability of an accessible path over the phenotypes of a GP map (see IV B 4).

In Table I, we report navigability for each GP map. The value of $\langle \psi \rangle$ is greater than 0.6 for all the GP maps we consider, apart from the non-compact HP models HP20 and HP25. The non-compact HP models have a navigability $\langle \psi \rangle \leq 0.013$ demonstrating these GP maps do not produce navigable fitness landscapes. These results suggest that the GP maps of RNA secondary structure, compact HP models, and the Polyomino model, have navigable fitness landscapes and contain very few fitness valleys under random fitness assignment. However, the lack of navigability in non-compact HP models highlights the need for further investigation of the effect of structural properties of the GP maps on navigability, which we pursue in the next section.

B. Common properties of GP maps are associated with navigability

1. GP maps with fewer phenotypes and fewer deleterious genotypes are more navigable

Having showed that three distinct GP maps give rise to navigable fitness landscapes under random fitness assignment, we explore the relationship between structural properties GP maps and navigability. Specifically, we consider the redundancy R of a GP map, measured as the average number of genotypes per non-deleterious phenotype (see Eq. (1)), and the deleterious frequency f_{del}. The deleterious frequency describes the fraction of genotype space that does not map to a well-defined phenotype. In the case of RNA secondary structure the deleterious phenotype would correspond to the un-
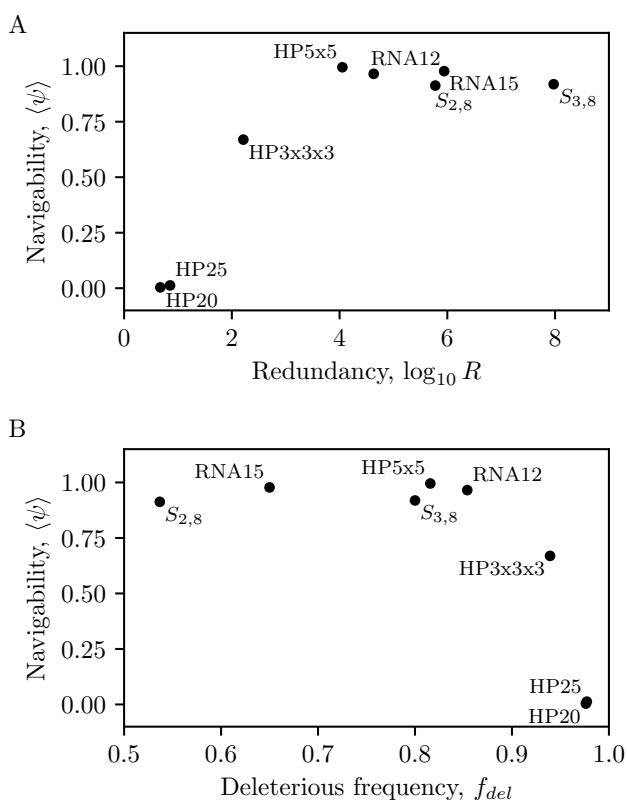
FIG. 2. Navigability of each GP map is plotted in relation to (A) redundancy $\log_{10} R$ and (B) deleterious frequency f_{del}. We find that there is a positive association with navigability and redundancy while a negative association with respect to deleterious frequency for large f_{del}.

folded RNA strand (i.e. the absence of any secondary structure). In the HP model it corresponds to the absence of a unique folded ground state. In the Polyomino model it corresponds to unbounded or non-deterministic assembly. In Fig. 2A we plot navigability against redundancy, while in Fig. 2B navigability is shown against the deleterious frequency with the numerical values provided in Table I. We observe a general increase in navigability for greater redundancy and smaller f_{del}. HP3x3x3 presents an example of particular interest by maintaining navigability ($\langle\psi\rangle = 0.669$) with less redundancy ($\log_{10} R = 2.2$) and large deleterious frequency ($f_{del} = 0.939$).

The results across different GP maps provide some intuition for factors that determine navigability. With decreasing redundancy, it becomes more difficult to access all phenotypes as they begin to occupy smaller fractions of the overall space. As f_{del} increases, more neighbours of a given genotype will have a fitness of 0, therefore localising phenotypes to smaller components in the GP map, increasing the likelihood of each genotype having no neighbouring genotypes with greater fitness.

2. Positive neutral correlations increase navigability

We next consider how neutral correlations, which fundamentally arise from a very general picture of constrained and unconstrained portions of genotype sequences [47–49] and lead to greatly enhanced mutational robustness [30], affect navigability. The level of correlations in a given GP map can be adjusted by taking two genotypes g_1 and g_2 at random and assigning the phenotype of g_1 to g_2 and vice versa. Such random swaps remove the local correlations that are intrinsic to the GP map. The total number of swaps applied is parameterised as s. With increasing s, we decorrelate the GP map towards a random phenotype assignment to the set of genotypes. While s parameterises the decorrelation process, it is not on a scale that captures the level of correlations present relative to either the original GP map or fully randomised GP map where there are no correlations. Therefore, a measure of correlations $c(s)$ (Eq. (9)) after s swaps is captured by relating phenotype robustness ρ_p and frequency f_p averaged across the phenotypes of the GP map for a given number of swaps s. When $c(s) = 1$, the correlations are equal to the original GP map, when $c(s) = 0$, the correlations are that of the randomised null model. Positive neutral correlations are present for $c(s) > 0$. By measuring the navigability $\langle\psi\rangle$ after a given number of swaps s, we measure the extent to which neutral correlations $c(s)$ affect navigability.

In Fig. 3A, we plot how navigability varies with $c(s)$ in $S_{2,8}$, RNA12, HP5x5 and HP3x3x3 GP maps, a subset of the GP maps in the previous section that are both small enough to be tractable here, and have sufficiently large navigability such that the effect of reducing correlations and dimensionality may be sizeable. All four GP maps, on average, show greater navigability for greater $c(s)$ with an approximately linear decay in navigability with decreasing $c(s)$, saturating at a lower value specific to each GP map: 0.378 ± 0.005 for RNA12, 0.100 ± 0.003 for HP5x5, 0.000 ± 0.000 for HP3x3x3, and 0.949 ± 0.002 for $S_{2,8}$, substantial reductions apart from for $S_{2,8}$. In $S_{2,8}$, the navigability $\langle\psi\rangle$ takes a greater value for the decorrelated GP map ($c < 1$) than for the original one ($c = 1$). This is because not all phenotypes are directly accessible from each other in the original GP map. However, a slight randomisation increases phenotype inter-connectivity due to the fact that the number of phenotypes for $S_{2,8}$ is smaller than the number of local mutations ($N_P < (K-1)L$). We expect that in GP maps of longer sequence length L, the role of positive neutral correlations will become even more pronounced. We explore this in Section II C with respect to fRNA phenotypes.

3. Large dimensionality increases navigability and decreases ruggedness

We now examine the effect of *dimensionality* of the GP map. The dimensionality of the entire GP map is defined as L, the length of the sequence. During the search for an accessible path from the source to target phenotype, all bases can be mutated, making use of the full dimensionality of the GP map. We can, however, reduce the dimensionality of the search by
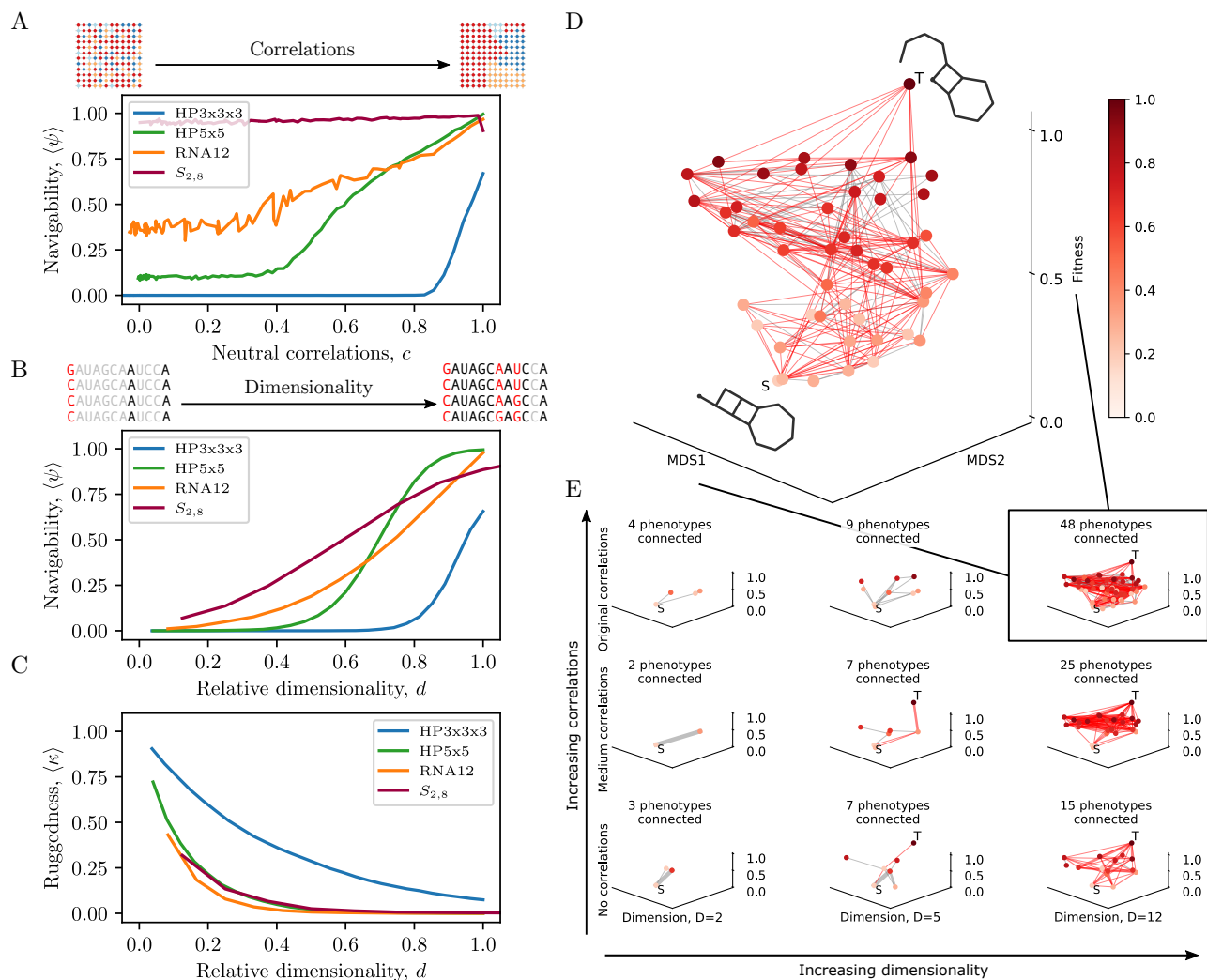
FIG. 3. **(A) Navigability emerges as positive neutral correlations are added to HP3x3x3, HP5x5, RNA12 and $S_{2,8}$ GP maps.** The level of neutral correlations is adjusted through genotype swaps, and the extent of positive neutral correlations after s swaps is measured on a scale c between the original GP map ($c = 1$) and the random null model's correlations ($c = 0$). A caricature of the genotype space, coloured according to phenotypes, is shown for low neutral correlations (top left) and high neutral correlations (top right). **(B) Greater dimensionality of the GP map increases navigability for $S_{2,8}$, HP3x3x3, HP5x5 and RNA12 GP maps.** During the search from a randomly chosen source phenotype to a target phenotype, we only allow D ($d = D/L$) of the total L bases to be mutated to explore genotype space. A caricature of a sequence with grey bases ($L - D$) not mutable, black bases mutable (D) and red bases varying across sequences, is depicted for low dimensionality (top left, $d = 3/12$) and high dimensionality (top right, $d = 11/12$). The GP maps show differing tolerance with respect to navigability under a change in dimensionality, $S_{2,8}$ permitting navigability for low dimensionality significantly more than HP3x3x3, for example. **(C) With increasing dimensionality, landscape ruggedness decreases.** We measure landscape ruggedness $\langle \kappa \rangle$ as the average proportion of all genotypes encountered that are local fitness maxima (no neutral neighbours or neighbours with increased fitness). Ruggedness decreases in all GP maps as dimensionality increases, but the level of ruggedness is GP map dependent. **(D) A schematic of the joint effect of dimensionality and correlations on navigability through visualisation of the phenotype connectivity network.** An example is illustrated of the search for an accessible path in a specific random instance of a fitness landscape with the phenotype network of RNA12. Phenotypes are nodes and the edges are possible transitions between genotypes of those phenotypes given the random fitness assignments. Edges that are red are transitions that may lead to the target phenotype from the source phenotype. Inaccessible transitions are shown in grey. The vertical axis is fitness. The horizontal plane is a two-dimensional embedding of the phenotype space of RNA12 derived through a multidimensional scaling (MDS) that uses the pairwise Hamming distances between the dot-bracket representations of the phenotypes. It follows that proximity in the horizontal plane corresponds to similar dot-bracket phenotypes. **(E) The phenotype network is shown for three levels of correlations (original, medium, and no correlations) and three levels of dimensionality ($D = 2, 6, 12$).** Navigability and connectivity in the phenotypic network visibly increases with both increasing correlations and dimensionality.

allowing only a random set of D sites (where $D < L$) to be mutated during a given search for an accessible path from

source to target. We then consider $\langle \psi \rangle$ as a function of the *relative dimensionality* $d = D/L \; \forall D \in \{1, ..., L\}$.

In Fig. 3B, we plot navigability $\langle \psi \rangle$ as a function of d. Reduced dimensionality severely reduces the navigability of fitness landscapes, with a sigmoidal relationship between $\langle \psi \rangle$ and d. All the curves show an increase from low navigability to high navigability as $d \to 1$ of the full GP map. The critical value of d, and general scale and shape, is different across the four GP maps indicating a complex dependence on other GP map properties.

In addition to identifying an accessible path during the search from source to target, we also count the number of genotypes that do not have a neutral neighbour or neighbour with greater fitness. In other words, the proportion of genotypes that are local fitness peaks, therefore providing a measure of landscape *ruggedness*. The average proportion of genotypes that are local fitness peaks across source-target phenotype pairs and fitness assignments in a given GP map, is represented as $\langle \kappa \rangle$. In Fig. 3C, the ruggedness for each relative dimensionality $d = D/L$ is plotted in the same four GP maps. We observe increasing dimensionality reduces ruggedness and, as relative dimensionality drops below a certain level, ruggedness sharply increases. Of note is HP3x3x3, where ruggedness is greater at a given relative dimensionality than for the other GP maps. Where all bases may mutate at $d = 1$, around 7 in 100 genotypes are local peaks ($\langle \kappa \rangle = 0.07$) but navigability remains high ($\langle \psi \rangle = 0.66$), demonstrating that partially rugged landscapes can still be navigable.

We illustrate an example of a source-target search in a schematic of the RNA12 GP map in Fig. 3D. We choose a random source and target pair and, during the search for an accessible path, keep track of all phenotypes encountered, their fitness and any transition between phenotypes. Each phenotype is represented as a node, edges as transitions between phenotypes, and the value on the vertical axis as the fitness. The $N_P = 58$ phenotypes of this GP map are assigned coordinates in the horizontal plane using multidimensional scaling (MDS) based on the pairwise Hamming distance between phenotypes [50]. This allows phenotypes that are similar to each other to be located in similar parts of the MDS1-MDS2 plane. The source and target phenotypes are labelled S and T respectively, edges that may form accessible paths are coloured red, and the remaining edges grey. This depiction of the fitness landscape immediately shows that it is highly connected with many accessible paths.

In Fig. 3E, with the same schematic source-target pair and fitness assignments as Fig. 3D, we illustrate the joint effect of neutral correlations and dimensionality on connectivity and navigability. We show the navigability of the phenotype network for three different degrees of correlation (no correlations, some correlations, original correlations) and three different dimensionalities ($D = 2, 6, 12$). The top right of the 9 plots is the original GP map that is also shown enlarged in Fig. 3D. We observe that decreasing both correlations and dimensionality of the search visibly reduces the navigability of the landscape through increasingly restricted networks. In the case of $D = 2$, the dimensionality in which fitness valleys are often visualised in the literature, phenotypic connectivity is sparse, making the landscape unnavigable. The increase in navigability with increases in both dimensionality and correlations highlight that both the structure of the underlying GP map and the high-dimensional nature of the evolutionary search are essential for fitness landscapes to be navigable.

C. Navigability of functional RNA fitness landscapes

Next we focus on the RNA secondary structure GP map by specifically choosing source and target phenotypes that have been observed in nature. This is important as only a small subset of all possible phenotypes are typically seen in real biological systems [51, 52] and it is navigability among this subset that has most relevance for evolutionary processes.

1. Fitness valleys are not observed between short fRNAs

We sample RNA secondary structures from the functional RNA database (fRNAdb) [35]. We consider pairs of fRNA phenotypes from the database with a given sequence length L, assigning a random fitness $F_{source} \in [0, 1)$ and $F_{target} = 1$, with random uniform assignment of fitness for all non-trivial phenotypes found during the search process. We consider larger L than earlier, specifically in the range $L \in [20, 40]$. We perform two distinct types of search by either permitting or preventing neutral mutations in exploring a given genotype's mutational neighbourhood. This provides a means to directly measure the role of neutral correlations in facilitating navigability for larger L. As the sequence length increases the number of phenotypes grows as $N_P \approx 1.76^L$ [53] producing a large computational overhead to track all phenotypes encountered during a search. In Section IV F, we describe in detail the more complex approach taken to measure navigability for larger L, which is necessary due to the increased computational expense.

In Table II, the navigability $\langle \psi \rangle$ for fitness landscapes with fRNA of sequence length $L = 20 - 40$ is reported along with the proportion of searches that were aborted and whether or not neutral mutations were permitted. With neutral mutations allowed, navigability is almost always 1.0, suggesting that fitness landscapes with fRNAdb source and targets are highly navigable. For $L > 30$ the proportion of aborted searches increases, leading to the greater potential for this estimate to be biased. However, there is a strong indication that with a greater computational threshold, similarly large navigability would be achieved at even larger L fRNA landscapes due to the observed scaling of $\langle \psi \rangle$ with the computational threshold (see Section A).

Where neutral mutations are disallowed, we find that navigability is markedly reduced below 1.0, although still substantially greater than zero ($\langle \psi \rangle \in [0.38, 0.64]$). The proportion of aborted searches is negligible. This finding is intriguing as it highlights that positive neutral correlations are important, but not essential, for the existence of accessible paths. A possible explanation lies in the vast number of phenotypes $N_P \approx 1.76^L$ available in the GP map, coupled with its high dimensionality. As fitness is randomly assigned and novel variation is only a few mutations away, there is a pool

TABLE II. The navigability $\langle\psi\rangle$ for length $L = 20 - 40$ fRNAs, the number of unique targets tested, the number of phenotypes in the fRNA database, the proportion of runs that are aborted and the estimated navigability. Results for simulations with and without neutral mutations are shown in the left-hand and right-hand sets of columns respectively. For non-aborted runs with neutral mutations permitted, random observed fRNA landscapes are almost completely navigable. When neutral mutations are prohibited, navigability is severely reduced, but still substantial.

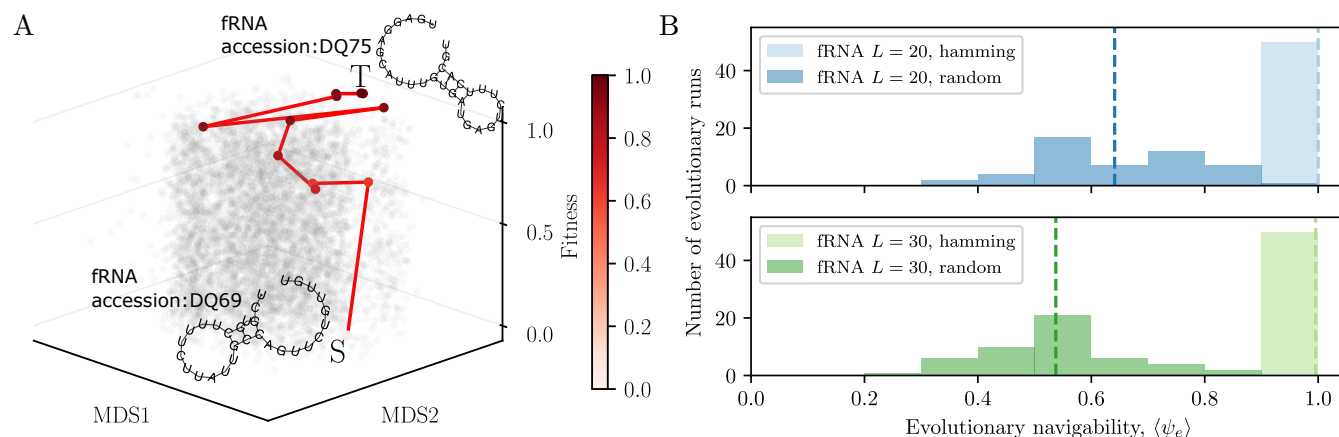|	Neutral mutations				No neutral mutations			
L	Targets	N_P^{fRNAdb}	$\langle\psi\rangle$	Aborted	Targets	N_P^{fRNAdb}	$\langle\psi\rangle$	Aborted
20	50	14350	0.987 ± 0.004	0.002	50	14350	0.530 ± 0.016	0.001
25	49	12958	0.999 ± 0.001	0.015	49	12958	0.661 ± 0.015	0.001
30	50	42195	1.000 ± 0.000	0.065	50	42195	0.638 ± 0.015	0.001
35	46	752	1.000 ± 0.000	0.190	46	752	0.599 ± 0.016	0.001
40	47	662	1.000 ± 0.000	0.517	47	662	0.380 ± 0.015	0.001

FIG. 4. **(A) Example of an accessible path for a specific $L = 30$ fRNA source-target pair.** As introduced in Fig. 3D, phenotypes are nodes whose coordinates are derived from a multidimensional scaling (MDS) embedding of the phenotype similarities based on Hamming distance, while the vertical axis is fitness. We show the vast extent of phenotypes discovered during the search as 'grey' nodes, a shortest accessible path connecting the source and target phenotypes with red edges, and the phenotypes along this path shaded in proportion to fitness. The example illustrates the interconnected nature of the fitness landscapes for a concrete fRNA example, where the properties of the GP map are key in facilitating navigability. **(B) Evolutionary dynamics for fRNA with the distribution of ψ for randomly chosen phenotypes that belonged to the fRNA database were considered for the fixed $L = 20$ and $L = 30$ GP maps.** The navigability for $L = 20$ and $L = 30$ fRNA for 50 different target fRNA phenotypes are illustrated using histograms of $\langle\psi_e\rangle$ for each target phenotype. The dark shaded bars show the proportion of successful searches for random fitness assignment, and the light bars for Hamming fitness assignment. Mean navigability of $\langle\psi_e\rangle > 0.5$ is observed for random fitness assignment and $\langle\psi_e\rangle > 0.9$ for Hamming fitness assignment.

of non-neutral phenotypes with possibly larger fitness, potentially within a small mutational radius.

In Fig. 4, we use the representation introduced in Fig. 3D to illustrate an accessible path in fRNA. For the successful traversal between a specific source and target fRNA, we see a vast array of background, 'greyed out' phenotypes discovered during the search for an accessible path, as well as a shortest accessible path connecting 10 different phenotypes with the node colour and their vertical axis coordinate showing their fitness. This illustration further highlights the hyper-connectedness and high-dimensional bypasses present in fRNA GP maps that are afforded through exponentially increasing redundancy, positive neutral correlations, and high dimensionality. The phenotype network also serves again as an alternative depiction of the fitness landscape in which the effect of GP map structure on the course of potential evolutionary explorations may be grasped more intuitively.

Summarising our results, we have demonstrated that fRNA GP maps have navigable fitness landscapes up to $L = 30$ fRNA. They are highly likely to be navigable for even larger *in vivo* fRNAs due to the observed scaling of both the GP map properties and navigability with respect to the computational threshold. Neutral mutations drastically increase navigability but do not solely determine the presence of accessible paths.

D. Evolutionary dynamics make use of accessible paths between fRNAs

Having considered whether accessible paths *exist* in a variety of GP maps, we consider whether these accessible paths are *utilised* under evolutionary dynamics.

It is conceivable that, while accessible paths exist in a fitness landscape, they may not be frequently used due to the

entropic effects associated with the evolutionary search process. For example, if there are many mutational paths that lead to a local fitness maximum compared to a single path leading to the globally fittest phenotype, the increased number of ways to reach the local peak may result in populations taking one of these more prevalent paths and becoming trapped, necessitating passage across a fitness valley to reach the fittest phenotype.

We simulated evolutionary dynamics with a Wright-Fisher process, implemented via a genetic algorithm, and considered two different fitness assignment schemes: (a) random and (b) using a given phenotype's dot-bracket Hamming distance to the target phenotype.

We chose $N_s = 50$ source phenotypes for each of $N_t = 20$ target phenotypes. During an evolutionary search, the change in fitness of the majority phenotype in the population was measured. The population was initialised from a population of genotypes that map to the source phenotype, and the fitness of the target was set to 1. We consider only the set of evolutionary simulations where the population was able to reach the target phenotype. We define *evolutionary navigability* $\langle \psi_e \rangle$ as the average probability that the population's majority phenotype reaches a target phenotype from a source phenotype (both randomly chosen) via an accessible path.

We consider only the polymorphic dynamical regime ($N\mu L \gg 1$, where N is population size, μ is point mutation rate and L is genotype length). This case provides dynamics that are most likely to be associated with entropic regimes due to rapid discovery and exploration of mutational pathways that lead to more prevalent local fitness peaks as opposed to global ones. A greater mutation rate also increases the ability to cross fitness valleys making it a valuable test under which to consider whether evolutionary accessible paths continue to be used. Further details of the evolutionary simulation are provided in the methods (see Section IV G).

In Fig. 4B the navigability for $L = 20$ and $L = 30$ fRNA is illustrated with histograms binning the value of $\langle \psi_e \rangle$ for each of $N_t = 50$ target phenotypes. The darker-shaded bars show the proportion of successful searches for random fitness assignment, and the lighter-shaded bars for the fitness assignment based on dot-bracket Hamming distance from the target phenotype. The mean navigability $\langle \psi_e \rangle$ is shown as a vertical dashed line. For random fitness assignment we find decreased navigability values of $\langle \psi_e \rangle \approx 0.64$ for $L = 20$ and $\langle \psi_e \rangle \approx 0.54$ for $L = 30$ compared to the non-evolutionary scenario, for which $\langle \psi \rangle \approx 1$. While this is a reduction relative to the potential navigability present in the landscape, this still suggests that accessible paths are utilised for the majority of targets.

Under Hamming distance fitness assignment we found that accessible paths are taken much more frequently with all target phenotypes having $\langle \psi_e \rangle = 1.0$ for $L = 20$ and $\langle \psi_e \rangle > 0.94$ for $L = 30$. This provides additional evidence of evolutionary navigability with a plausible alternative fitness assignment and identifies the potential importance of *phenotypic correlations* within the GP map (in addition to the genotypic correlations discussed above) for evolutionary navigability.

III.   SUMMARY AND DISCUSSION

In this paper, we considered the navigability of fitness landscapes and, specifically, whether fitness valleys are prevalent in high-dimensional fitness landscapes based on biologically realistic GP maps. We examined three such GP maps with common structural properties and found that they were highly navigable, suggesting that fitness valleys are largely absent. We generalised this by demonstrating navigability in GP maps with longer RNA sequences using phenotypes contained in the fRNAdb database of fRNA observed in nature. Finally, we considered the question of whether accessible paths not only exist, but are also utilised by populations subject to evolutionary dynamics. We found that accessible paths are followed frequently by populations in evolutionary simulations, and are therefore likely to play an important role in real evolutionary settings.

We identified that universal structural properties of GP maps can facilitate navigability, namely: genotypic redundancy, the frequency of the deleterious phenotype, positive neutral correlations, and high dimensionality as a proportion of sequence length. These are important factors that are not characterised in a direct genotype-to-fitness mapping and are necessary to provide navigability. Additionally, we demonstrated that the phenotype network is arguably a more useful way to conceptualise evolutionary exploration. Visualising the fitness landscape in this way avoids the misleading intuitions of fitness valleys that can arise from the low-dimensional fitness landscape metaphor.

While we found fitness landscapes to be generally navigable under evolutionary dynamics, this navigability was lower than one might expect given the potential availability of accessible paths in these landscapes. We suggest two possible reasons for the reduction in the evolutionary setting: 1) the population truly arrives at a local optima and has no choice but to cross a fitness valley, and 2) due to the stochastic nature of the evolutionary dynamics the fitness of the population's majority phenotype may drop, but not all members will necessarily have a reduced fitness and therefore, for the majority to return to a greater fitness, a fitness valley may not need to be crossed. This may lead to an underestimate of the true evolutionary navigability. The sensitivity of evolutionary navigability to alternative definitions is an important area for future exploration. The replacement of the random fitness assignment with one based on Hamming distance improved navigability drastically and highlights the role that *phenotypic correlations* play in GP maps in addition to the genotypic correlations discussed in [30].

A central assumption was that function and fitness are directly related to shape of the physical structure alone. This is an assumption made ubiquitously in the study of self-assembly GP maps where the structure is the sole component of the phenotype [31, 32]. Importantly, this will not always hold for all biological systems. For example, where a specific sequence is necessary to facilitate binding of a protein, an additional sequence constraint is imposed on top of that required to specify the structure. This additional specificity potentially reduces both the redundancy of the phenotype and the di-

mensionality available for accessing alternate genotypes. Our findings regarding landscape navigability should therefore be considered in the context of the GP map properties that facilitate accessible paths. If these properties are not present in the system in question, the fitness landscape is unlikely to be navigable. This is supported, for example, in ref. [54] where fragmented fitness peaks are identified in a rare exhaustive empirical fitness landscape study, but where fitness was specifically determined by the ability for GTP to bind rather than by specific secondary structure itself.

The metaphor of the fitness landscape has endured for almost a century of research in evolutionary biology. It is often discussed in intuitive terms, as a low-dimensional landscape. This can be problematic, as it obscures counter-intuitive properties of high-dimensional spaces, which real fitness landscapes are. Moreover, much of the literature on fitness landscapes does not consider genotype-phenotype maps and their properties, such as the ubiquity of neutral networks and their correlations in genotype space. Our contribution demonstrates that specific GP map properties, in combination with high-dimensionality, make fitness landscapes navigable. We show that accessible paths are not only available in three different biologically realistic GP maps, but also that they are followed in simulated evolutionary dynamics of functional RNA structures. These findings demonstrate that fitness valleys are largely absent in three biological GP maps. Given that the relevant GP map properties have been found in numerous other GP maps, it is highly likely that fitness valleys are indeed uncommon across a wide range of biological systems. Our findings support work on the role of high-dimensionality in promoting accessibility [11], as well as attempts to create an up-to-date metaphor for evolutionary adaptation [55]. A fuller understanding of the role of the GP map in structuring the high dimensional fitness landscape could provide vital insights into areas such as the arrival of drug resistance [56, 57] or the mutational progressions of cancer [58].

IV. METHODS

A. Self-assembly GP maps

We consider three GP maps for different systems of biological self-assembly: the RNA secondary structure GP map [37] for secondary structure of RNA sequences, the HP lattice model for protein tertiary structure [44, 59] and the Polyomino model for protein quaternary structure [43]. The phenotype in each is solely related to the assembled structure. We briefly summarise the GP maps below with detailed comparisons between the three GP maps found in ref. [30].

- *RNA secondary structure*: we use the Vienna package [37] (version 1.8.5) with default parameters to convert RNA sequences to dot-bracket secondary structures. GP maps are represented as RNAL with sequences of length L.

- *HP lattice model*: we follow refs. [45, 46] and consider energetic interactions between non-adjacent pairs

to have values $E_{\text{HH}} = -1$, with $E_{\text{HP}} = E_{\text{PP}} = 0$, where H are hydrophobic and P are polar amino acids. If a sequence has a unique lowest energy structure, its phenotype is that structure, otherwise it is considered degenerate. We consider both the non-compact GP map for all folds of a given length referred to as HPL and also only the set of compact structures referred to as HPlxwxh.

- *Polyomino model*: we follow refs. [30, 43] and consider the GP maps S_{N_t,N_c} where N_t is the number of assembly kit tiles and N_c with the default self-assembly process used.

The GP maps may be further characterised by their genotype sequence length L, base K, number of genotypes $N_G = K^L$ and number of phenotypes N_P. The redundancy n_p of a given phenotype p is the number of genotypes that map to p and this is normalised by the size of the genotype space to give the frequency $f_p = n_p/K^L$. The overall redundancy R of a GP map is defined as the average number of genotypes per non-deleterious phenotype:

$$R = K^L(1 - f_{del})/(N_P - 1) \tag{1}$$

We provide Table III to summarise the characteristic properties used to differentiate the GP maps.

A particular feature of all three GP maps is a single phenotype that is of a different nature to the others: for RNA secondary structure this is the unfolded 'trivial' structure, the HP lattice model it is sequences that have a degenerate ground state and for the Polyomino model it is when there is either unbounded or non-deterministic growth (UND). We refer to this phenotype here as the *deleterious* or *del* phenotype as, in each GP map, we consider it low fitness due to the non-specificity of the structural phenotype. We assign a fitness of zero for *del* throughout this work. While this is a strong assumption, given the large-scale dominance of the *del* phenotype in Polyomino and HP GP maps, we expect this assumption to exacerbate the presence of valleys rather than introducing a bias towards navigability.

B. Measuring landscape navigability

1. Definitions and formulation

In order to establish the presence of fitness valleys in a fitness landscape, we consider whether it is possible to reach the fittest phenotype from any given point in the genotype space via a path where the fitness increases monotonically defined as an *accessible path* [11, 60]. *Landscape navigability* has previously been defined as the proportion of accessible paths to a given genotype from all other genotypes [17]. To briefly summarise, here we specifically define the navigability as the average probability that a randomly chosen phenotype pair have at least one accessible path between them, given a fitness assignment process to phenotypes. We denote accessibility with ψ, where $\psi = 1$ indicates the presence of at least one accessible path between two phenotypes for a specific set of fitness

Term and description	Representation
Alphabet	$\mathcal{A}$
Genotype space	$\mathcal{G}$
Phenotype space	$\mathcal{P}$
Alphabet size	$|\mathcal{A}| = K$
Genotype length	L
Dimensionality: the number of bases that may be mutated in the genotype	D
Relative dimensionality	$d = D/L$
1-mutant genotypes of genotype g	$\sigma(g)$
Number of 1-mutation neighbours of genotype g	$|\sigma(g)| = (K-1)L$
Number of genotypes	$N_G = K^L$
Number of phenotypes	N_P
GP map	$M : \mathcal{G} \to \mathcal{P}$
Phenotype redundancy: the number of genotypes that map to phenotype p	n_p
Phenotype frequency: the fraction of genotypes that map to phenotype p	f_p
GP map redundancy: the average number of genotypes per non-deleterious phenotype	$R = K^L(1 - f_{del})/(N_P - 1)$
Fitness of phenotype p	F_p
Presence of at least one accessible path with a binary event indicator	ψ
Probability of an accessible path existing between phenotype p and q	ψ_{pq}
Navigability: average probability of an accessible path for phenotype pairs of a GP map	$\langle\psi\rangle$
Evolutionary navigability	$\langle\psi_e\rangle$
Ruggedness	$\langle\kappa\rangle$
Neutral correlations	c
RNA GP map for genotype sequence length L	RNAL
Polyomino GP map with N_t kit tiles and N_c interface types	S_{N_t, N_c}
HP lattice GP map for genotype sequence length L	HPL
HP lattice GP map for compact phenotypes in the grid box $l \times w \times h$	HP$lxwxh$

TABLE III. **Terminology.** A summary of terms and their representations used in the paper.

assignments, and $\psi = 0$ indicating no accessible paths. When $\psi = 0$, a fitness valley must be traversed between the phenotypes. With this notation, we use $\langle\psi\rangle$ to represent navigability of fitness landscapes for a given GP map.

2. Fitness landscapes

In conjunction with the GP map M, a fitness landscape instance is defined by the set of phenotype fitnesses $\mathcal{F} := \{F_{p_i}\}_{i=1}^{N_P}$, with i denoting the ith indexed phenotype p_i. We refer to the *source* phenotype p and *target* phenotype q in the search for an accessible path from $p \to q$. We consider two fitness assignments in this paper:

- *Random fitness*: random samples $F_{p_i} \sim \text{Uniform}(0,1)$ with target phenotype q having $F_q = 1$

- *Hamming distance*: where the similarity of phenotype p compared to a phenotype q is measured by the number of matching positions in the aligned phenotype string representation given by $F(p,q) = 1 - \sum_j^L \delta(p^{(j)}, q^{(j)})/L$, where $p^{(j)}$ is the string character representing phenotype p at the jth base position and $F(p,q)$ is the fitness of phenotype p compared to a target phenotype q

$F_{del} = 0$ for all fitness assignments.

3. Navigability estimation

The probability of an accessible path ($\psi = 1$) between a source phenotype p and target phenotype q, given a random fitness landscape instance $\mathcal{F}$, is deterministic with a binary outcome. We can define the probability of ψ more explicitly as a function of p, q and $\mathcal{F}$ as follows:

$$\psi(p, q, \mathcal{F}) := P(\psi = 1 | p \to q, \mathcal{F}) \qquad (2)$$

where

$$\psi(p, q, \mathcal{F}) = \begin{cases} 1 & \text{if at least one accessible path exists} \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

We can take the expectation over $\mathcal{F}$ yielding the mean probability of an accessible path from p to q as:

$$\psi_{pq} = E_{\mathcal{F}}[\psi(p, q, \mathcal{F})] \qquad (4)$$

With this notation, we can define the *navigability* for the GP map as the expectation over Eq. (4) for phenotypes p and q sampled uniformly at random:

$$\langle\psi\rangle = E_{p,q}[\psi_{pq}] \qquad (5)$$

We can estimate this probability of reaching a given target phenotype q from a uniform randomly chosen source phenotype p by computationally measuring $\psi(p, q, \mathcal{F})$ for N_s randomly chosen sources for each of N_t randomly chosen targets, with a new random fitness landscape instance $\mathcal{F}$ for each

pair. We use $I_T(s,t)$ to indicate whether the computational estimate for source index s with target index t was inside the computational threshold T and completed the search without aborting. The estimate can be written as:

$$\langle\psi\rangle = \frac{1}{N_c}\sum_{t=1}^{N_t}\sum_{s=1}^{N_s} I_T(s,t)\psi(p_{st}, q_t, \mathcal{F}_{st}) \qquad (6)$$

where p_{st} and q_t are the source and target phenotypes of sth source for the tth target, the number of completed runs is $N_c = \sum_{t,s} I_T(s,t)$ and the aborted proportion α:

$$\alpha = 1 - \frac{N_c}{N_t N_s} \qquad (7)$$

The estimate of the navigability of a fitness landscape with GP map has an associated Bernoulli standard error (derived from an estimate of the corrected sample standard deviation):

$$SE(\langle\psi\rangle) = \sqrt{\frac{\langle\psi\rangle\,(1 - \langle\psi\rangle)}{N_c - 1}} \qquad (8)$$

We next describe in more detail the computational algorithm for estimating $\langle\psi\rangle$.

4. Navigability estimation algorithm

For a given source and target phenotype, in each random landscape instance, we perform the following computational algorithm to measure ψ. We first provide some definitions:

- GP map M: is a function $M : \mathcal{G} \to \mathcal{P}$ where $\mathcal{G}$ is the space of genotypes and $\mathcal{P}$ is the space of phenotypes, such that we can write the phenotype p of genotype g as $p = M(g)$

- Dimensionality: We define the set of sequence positions that may be mutated as $\mathcal{D}$, with the size of $|\mathcal{D}|$ being the dimensionality D. When $|\mathcal{D}| = L$ all base positions are mutable. Relative dimensionality is defined as the dimensionality relative to sequence length $d = D/L$

- Alphabet: sequences have a set of $\mathcal{A}$ possible letters at a given site. The size of $|\mathcal{A}| = K$ is the base.

- u_0 contains genotypes whose 1-mutant neighbours are yet to be considered in a given search for an accessible path

- u_1 contains genotypes that have already had their 1-mutant neighbours considered in a given search for an accessible path

The algorithm proceeds with a Breadth First Search (BFS):

1. A random genotype g that maps to the source phenotype is chosen and added to u_0

2. Set the first element of u_0 as g

3. For base $a \in \mathcal{A}$ at position j and for each position $j \in \mathcal{D}$, measure genotype neighbour g' and phenotype $p' = M(g')$

4. If $F_{p'} \geq F_p$ and $g' \notin u_1$, add g' to u_0

5. Move g from u_0 to u_1

6. If $|u_0| = 0$ or $|u_0| + |u_1| > T$ (computational threshold) or the target phenotype is found, return 'aborted' or ψ respectively. Otherwise return to step 2

The algorithm finishes with either u becoming empty, or the combined size of u_0 and u_1 becoming larger than a pre-defined threshold T (introduced in Section IV B 1), beyond which computational progress may become unfeasible. We discard these aborted runs from the measurement of navigability $\langle\psi\rangle$ using the indicator function I_T of the previous section (Section IV B 3).

As described in Eq. (6) we pick N_s source phenotypes uniformly at random for each of the N_t target phenotypes also chosen at random. We set $N_t = 20$ and $N_s = 50$. The uncertainty in the estimate of the navigability $\langle\psi\rangle$ is reported as the standard error $SE(\langle\psi\rangle)$ across the ensemble of measurements.

C. Removing correlations

In order to measure the effect of positive neutral correlations [30], we perform genotype swaps and then repeat the measurement of $\langle\psi\rangle$. This process involves constructing a new GP map M_s from the original GP map $M_{s=0} := M$ where s is the number of pairs of genotypes whose phenotype's have been swapped. More precisely, a swap involves selecting two genotypes g_1 and g_2 with uniform random probability and setting $M_s(g_1) = M_{s-1}(g_2)$ and $M_s(g_2) = M_{s-1}(g_1)$. It follows that $M_{s\to\infty}$ is the uncorrelated random null model GP map with no positive neutral correlations as used in ref. [30]. As shown in ref. [30], the random null model has $\rho_p \approx f_p$ when there are no positive neutral correlations. Therefore, we additionally define the correlations c present in a given GP map M_s by comparing the logarithm of the average robustness-to-frequency ratio in a given GP map against the original GP map, generating a scale for measuring correlations in M_s:

$$c(s) = \frac{\log_{10}\left\langle\frac{\rho_p(s)}{f_p(s)}\right\rangle_p}{\log_{10}\left\langle\frac{\rho_p(0)}{f_p(0)}\right\rangle_p} \qquad (9)$$

where for $s = 0$ we have $c(0) = 1$, and for $\lim_{s\to\infty} c(s) \approx 0$ the expectation for the random model. Therefore, the scale yields positive values for c where there is, on average, greater robustness than frequency. The process of removing correlations gradually from the original GP map ($s = 0$) to the random null model ($s \to \infty$) provides a range over which the relationship between positive neutral correlations and navigability may be considered in GP maps. We measure the navigability of $S_{2,8}$, RNA12, HP3x3x3 and HP5x5 by taking 100

evenly spaced values for s on the range $s = [0, K^L]$ and measuring $\langle \psi \rangle$ and $c(s)$ for each.

D. Restricting dimensionality

To measure the role of dimensionality we restrict the dimensionality of a search for an accessible path from source to target by only allowing a set of $\mathcal{D}$ randomly chosen positions along the sequence to be mutated in the 1-mutant neighbour measurement in Step 3 of the navigability algorithm above (Section IV B 4). The dimensionality D is the number of positions that may be mutated $|\mathcal{D}|$, and the relative dimensionality $d := D/L$. When $D = L$ we have the original dimensionality, while for $D = 1$ only a single sequence position may be mutated. The GP map M itself is not changed under this dimensional restriction but rather the connectivity of genotypes and therefore the connectivity of the fitness landscape.

We measure the navigability of $S_{2,8}$, RNA12, HP3x3x3 and HP5x5 by taking evenly spaced values for D on the range $D \in [1, L]$.

E. Measuring ruggedness

For fitness landscapes, related to navigability is the concept of landscape *ruggedness*. We measure $\kappa(g)$, whether a genotype is a local fitness maximum, during the search from source to target. The average proportion of genotypes that are local fitness maxima provides a measure of ruggedness [26]. Whether a genotype g is a local fitness peak is determined by the fitness of all accessible 1-mutant neighbours g', such that:

$$\kappa(g) = \begin{cases} 1 & \text{if } F_{M(g')} < F_{M(g)} \forall g' \in \sigma(g) \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where we have the function $\sigma(g)$ which returns the set of 1-mutants of genotype g. We calculate the ruggedness for a landscape by taking the average of $\kappa(g)$ over all genotypes and all source-target pairs once the search has completed. We denote the ruggedness as $\langle \kappa \rangle$.

F. Navigability in the functional RNA database

In Section II C, we examine navigability in a specific subset of RNA phenotypes, namely those that are found in the functional RNA database (fRNAdb) [35]. For a given length we use all phenotypes in proportion to their occurrence in the fRNAdb apart from the trial structure which we exclude as it is assigned zero fitness here. We randomly choose $N_t = 50$ targets with $N_s = 20$ randomly chosen sources from this set.

In order to examine navigability between functional RNAs, we must consider sequences longer than $L = 15$. In doing so, we introduce additional computational overhead given the increasing neutral set size resulting in the condition $|u_0| + |u_1| > T$ being more likely to be met. Therefore to maximise the

number of non-aborted runs, we perform a modified Depth-First Search (DFS) where we attempt to greedily follow paths of increasing gradient until we reach the max fit phenotype. If the path fails, instead of moving back one step as in a standard DFS, we go all the way back to the start of the walk and pick an unexplored neighbour with the lowest fitness to begin a new uphill walk. In this way, we maximise the exploration of new phenotypes by always starting our deep walks from the lowest point while still maintaining the ability to perform long walks during the search.

We write the modified DFS algorithm explicitly as:

1. A random genotype g that maps to the source phenotype is chosen and added to u_0.

2. Set the first element of u_0 as g, and $p = M(g)$

3. For each alternative base $a \in \mathcal{A}$ at position j and for each position j in $\mathcal{D}$, measure genotype neighbour g' and phenotype $p' = M(g')$

4. If any g' has $F_{p'} > F_p$ and $g' \notin u_1$ and $g' \notin u_0$, add g' to front of u_0 and return to step 2

5. If any g' have $p = p'$ and $|u_0| = 1$, add one such neutral case to the back of u_0 if $g' \notin u_0$ and $g' \notin u_1$

6. Move g from u_0 to u_1

7. If $|u_0| = 0$ or $|u_0| + |u_1| > T$ (computational threshold) or the target phenotype is found, return 'aborted' or ψ respectively. Otherwise return to step 2.

We note that for searches where neutral mutations are not permitted as part of the search, step 5 of the above is omitted.

G. Navigability estimation under evolutionary dynamics

We measured fitness landscape navigability as the average probability that a given source-target pair could be connected by way of an accessible path. We extend this definition to the more strict requirement of *evolutionary navigability* where the evolutionary dynamics of a population is considered instead of just the existence of an accessible path in crossing the fitness landscape.

We measure $\langle \psi_e \rangle$ as the proportion of source-target pairs for which the target is reached without the majority population phenotype undergoing a decrease in fitness before finding the target. The majority was taken as being a phenotype that occupied more than 50% of the population's phenotypes. If no phenotype met this condition in a given generation, then the majority phenotype fitness is not updated.

Evolutionary dynamics were performed using Wright-Fisher dynamics [61, 62], and the additional parameters used for each evolutionary dynamical runs were the following:

- GA parameters: $N_{\text{gen}} = 10,000, N_{\text{pop}} = 100, \mu = 0.05$

- fRNA parameters: $L = 20$ and 30

Evolutionary runs that are terminated after $N_{\text{gen}} = 10,000$ generations are treated in the same manner as those that are aborted when estimating $\langle \psi \rangle$. Therefore, evolutionary navigability $\langle \psi_e \rangle$ is the fraction of evolutionary runs that successfully evolved to the target phenotype through the majority population phenotype taking an accessible path across all runs excluding those that were terminated at N_{gen} generations.

V. ACKNOWLEDGEMENTS

VI. AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: SFG, AAL, SEA. Performed the experiments: SFG. Analysed the data: SFG, AAL, SEA. Supervised the work: SEA. Wrote the paper: SFG, AAL, SEA.

[1] Sewall Wright. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proceedings 6th International Congress on Genetics*, volume 1, pages 356–366, 1932.

[2] Stuart A. Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford university press, 1993.

[3] Erik Svensson and Ryan Calsbeek. *The adaptive landscape in evolutionary biology*. Oxford University Press, 2012.

[4] Massimo Pigliucci. Landscapes, surfaces, and morphospaces: what are they good for. *The adaptive landscape in evolutionary biology*, pages 26–38, 2012.

[5] J Arjan GM de Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15(7):480–490, 2014.

[6] Inês Fragata, Alexandre Blanckaert, Marco António Dias Louro, David A Liberles, and Claudia Bank. Evolution in the light of fitness landscape theory. *Trends in ecology & evolution*, 34(1):69–82, 2019.

[7] Ronald Aylmer Fisher. *The genetical theory of natural selection*. Clarendon Press, 1958.

[8] Robert M May. *Stability and complexity in model ecosystems*. Princeton university press, 1973.

[9] Michael Conrad and Werner Ebeling. M.V. Volkenstein, evolutionary thinking and the structure of fitness landscapes. *BioSystems*, 27(3):125–128, 1992.

[10] Sergey Gavrilets. *Fitness landscapes and the origin of species (MPB-41)*. Princeton University Press, 2004.

[11] Jasper Franke, Alexander Klözer, J. Arjan G. M. de Visser, and Joachim Krug. Evolutionary Accessibility of Mutational Pathways. *PLoS Computational Biology*, 7(8):e1002134, 2011.

[12] Suman G. Das, Susana O.L. Direito, Bartlomiej Waclaw, Rosalind J. Allen, and Joachim Krug. Predictable properties of fitness landscapes induced by adaptational tradeoffs. *eLife*, 9:1–24, 2020.

[13] Daniel M. Weinreich, Richard A. Watson, and Lin Chao. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*, 59(6):1165–1174, 2005.

[14] Maurício Carneiro and Daniel L. Hartl. Adaptive landscapes and protein evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(SUPPL. 1):1747–1751, 2010.

[15] Nicholas C. Wu, Lei Dai, C. Anders Olson, James O. Lloyd-Smith, and Ren Sun. Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife*, 5:1–21, 2016.

[16] Claudia Bank, Sebastian Matuszewski, Ryan T. Hietpas, and Jeffrey D. Jensen. On the (un)predictability of a large intragenic fitness landscape. *Proceedings of the National Academy of Sci-*

ences of the United States of America, 113(49):14085–14090, 2016.

[17] José Aguilar-Rodríguez, Joshua L. Payne, and Andreas Wagner. A thousand empirical adaptive landscapes and their navigability. *Nature Ecology & Evolution*, 1(2):0045, 2017.

[18] Júlia Domingo, Guillaume Diss, and Ben Lehner. Pairwise and higher-order genetic interactions during the evolution of a tRNA. *Nature*, 558(7708):117–121, 2018.

[19] Jia Zheng, Joshua L Payne, and Andreas Wagner. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science*, 365(6451):347–353, 2019.

[20] Victoria O. Pokusaeva, Dinara R. Usmanova, Ekaterina V. Putintseva, Lorena Espinar, Karen S. Sarkisyan, Alexander S. Mishin, Natalya S. Bogatyreva, Dmitry N. Ivankov, Arseniy V. Akopyan, Sergey Ya Avvakumov, Inna S. Povolotskaya, Guillaume J. Filion, Lucas B. Carey, and Fyodor A. Kondrashov. An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. *PLOS Genetics*, 15(4):e1008079, 2019.

[21] Andreas Wagner. *Life Finds a Way: What Evolution Teaches Us about Creativity*. Oneworld Publications, 2019.

[22] Frank J Poelwijk, Daniel J Kiviet, Daniel M Weinreich, and Sander J Tans. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, 445(7126):383–386, 2007.

[23] Alexander E Lobkovsky and Eugene V Koonin. Replaying the tape of life: quantification of the predictability of evolution. *Frontiers in genetics*, 3:246, 2012.

[24] Daniel L. Hartl. What can we learn from fitness landscapes? *Current Opinion in Microbiology*, 21:51–57, 2014.

[25] Ard A Louis. Contingency, convergence and hyper-astronomical numbers in biological evolution. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 58:107–116, 2016.

[26] Stuart Kauffman and Simon Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11 – 45, 1987.

[27] Marcin Zagorski, Zdzislaw Burda, and Bartlomiej Waclaw. Beyond the hypercube: evolutionary accessibility of fitness landscapes with realistic mutational networks. *PLoS computational biology*, 12(12):e1005218, 2016.

[28] John FC Kingman. A simple model for the balance between selection and mutation. *Journal of Applied Probability*, 15(1):1–12, 1978.

[29] Bjørn Østman and Christoph Adami. Predicting Evolution and Visualizing High-Dimensional Fitness Landscapes. In *Recent*

Advances in the Theory and Application of Fitness Landscapes, pages 509–526. Springer, 2014.

[30] Sam F. Greenbury, Steffen Schaper, Sebastian E. Ahnert, and Ard A. Louis. Genetic Correlations Greatly Increase Mutational Robustness and Can Both Reduce and Enhance Evolvability. *PLoS Computational Biology*, 12(3):1–27, 2016.

[31] S. E. Ahnert. Structural properties of genotype–phenotype maps. *Journal of The Royal Society Interface*, 14(132):20170275, 2017.

[32] Susanna Manrubia, José A. Cuesta, Jacobo Aguirre, Sebastian E. Ahnert, Lee Altenberg, Alejandro V. Cano, Pablo Catalán, Ramon Diaz-Uriarte, Santiago F. Elena, Juan Antonio García-Martín, Paulien Hogeweg, Bhavin S. Khatri, Joachim Krug, Ard A. Louis, Nora S. Martin, Joshua L. Payne, Matthew J. Tarnowski, and Marcel Weiß. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of Life Reviews*, 38:55–106, 2021.

[33] Andreas Wagner. Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society B: Biological Sciences*, 275(1630):91–100, 2008.

[34] Steffen Schaper, Iain G. Johnston, and Ard A. Louis. Epistasis can lead to fragmented neutral spaces and contingency in evolution. *Proceedings of the Royal Society B: Biological Sciences*, 279(1734):1777–1783, 2012.

[35] Taishin Kin, Kouichirou Yamada, Goro Terai, Hiroaki Okida, Yasuhiko Yoshinari, Yukiteru Ono, Aya Kojima, Yuki Kimura, Takashi Komori, and Kiyoshi Asai. fRNAdb: A platform for mining/annotating functional RNA candidates from non-coding RNA sequences. *Nucleic Acids Research*, 35(SUPPL. 1):145–148, 2007.

[36] Peter Schuster, Walter Fontana, Peter F. Stadler, and Ivo L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1344):279–284, 1994.

[37] I.L. Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of rna secondary structures. *Monatshefte für Chemie/Chemical Monthly*, 125(2):167–188, 1994.

[38] Walter Fontana. Modelling 'evo-devo' with RNA. *BioEssays*, 24(12):1164–1177, 2002.

[39] Matthew C. Cowperthwaite, Evan P. Economo, William R. Harcombe, Eric L. Miller, and Lauren Ancel Meyers. The ascent of the abundant: How mutational networks constrain evolution. *PLoS Comput Biol*, 4(7):e1000110, 07 2008.

[40] Jacobo Aguirre, Javier M. Buldú, Michael Stich, and Susanna C. Manrubia. Topological structure of the space of phenotypes: The case of RNA neutral networks. *PLoS ONE*, 6(10):e26324, 10 2011.

[41] Steffen Schaper and Ard A. Louis. The arrival of the frequent: How bias in genotype-phenotype maps can steer populations to local optima. *PLoS ONE*, 9(2):e86635, 02 2014.

[42] Andreas Wagner. *The origins of evolutionary innovations: a theory of transformative change in living systems*. Oxford University Press, 2011.

[43] Sam F. Greenbury, Iain G. Johnston, Ard A. Louis, and Sebastian E. Ahnert. A tractable genotype–phenotype map modelling the self-assembly of protein quaternary structure. *Journal of The Royal Society Interface*, 11(95), 2014.

[44] Ken A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24(6):1501–1509, 1985.

[45] Anders Irbäck and Carl Troein. Enumerating designing sequences in the HP model. *Journal of Biological Physics*, 28(1):1–15, 2002.

[46] Evandro Ferrada and Andreas Wagner. A comparison of genotype-phenotype maps for RNA and proteins. *Biophysical Journal*, 102(8):1916 – 1925, 2012.

[47] S F Greenbury and S E Ahnert. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype–phenotype maps. *Journal of The Royal Society Interface*, 12(113):20150724, 2015.

[48] Susanna Manrubia and José A. Cuesta. Distribution of genotype network sizes in sequence-to-structure genotype-phenotype maps. *Journal of the Royal Society Interface*, 14(129), 2017.

[49] Marcel Weiß and Sebastian E Ahnert. Phenotypes can be robust and evolvable if mutations have non-local effects on sequence constraints. *Journal of The Royal Society Interface*, 15(138):20170618, 2018.

[50] Ingwer Borg and Patrick JF Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

[51] Kamaludin Dingle, Fatme Ghaddar, Petr Šulc, and Ard A Louis. Phenotype bias determines how natural RNA structures occupy the morphospace of all possible shapes. *Molecular Biology and Evolution*, 09 2021. msab280.

[52] Iain G Johnston, Kamaludin Dingle, Sam F Greenbury, Chico Q Camargo, Jonathan PK Doye, Sebastian E Ahnert, and Ard A Louis. Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution. *Molecular Biology and Evolution*, 2021.

[53] Kamaludin Dingle, Steffen Schaper, and Ard A. Louis. The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA. *Interface Focus*, 5(6):20150053, 2015.

[54] José I Jiménez, Ramon Xulvi-Brunet, Gregory W Campbell, Rebecca Turk-MacLeod, and Irene A Chen. Comprehensive experimental fitness landscape and evolutionary network for small rna. *Proceedings of the National Academy of Sciences*, 110(37):14984–14989, 2013.

[55] Pablo Catalán, Clemente F. Arias, Jose A. Cuesta, and Susanna Manrubia. Adaptive multiscapes: An up-to-date metaphor to visualize molecular adaptation. *Biology Direct*, 12(1):1–15, 2017.

[56] C. Brandon Ogbunugafor, C. Scott Wylie, Ibrahim Diakite, Daniel M. Weinreich, and Daniel L. Hartl. Adaptive Landscape by Environment Interactions Dictate Evolutionary Dynamics in Models of Drug Resistance. *PLoS Computational Biology*, 12(1):1–20, 2016.

[57] Daniel Nichol, Mark Robertson-Tessi, Alexander R.A. Anderson, and Peter Jeavons. Model genotype–phenotype mappings and the algorithmic structure of evolution. *Journal of the Royal Society Interface*, 16(160), 2019.

[58] Ramon Diaz-Uriarte. Cancer progression models and fitness landscapes: A many-to-many relationship. *Bioinformatics*, 34(5):836–844, 2018.

[59] Kit Fun Lau and Ken A. Dill. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules*, 22(10):3986–3997, 1989.

[60] Daniel M Weinreich. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, 312(5770):111–114, 2006.

[61] Warren John Ewens. *Mathematical Population Genetics: Theoretical Introduction*, volume 1. Springer, New York, NY, 2004.

[62] Lorens A. Imhof and Martin A. Nowak. Evolutionary game dynamics in a wright-fisher process. *Journal of Mathematical Biology*, 52(5):667–681, 2006.

Appendix A: Impact of computational thresholds on discovery of estimation of navigability
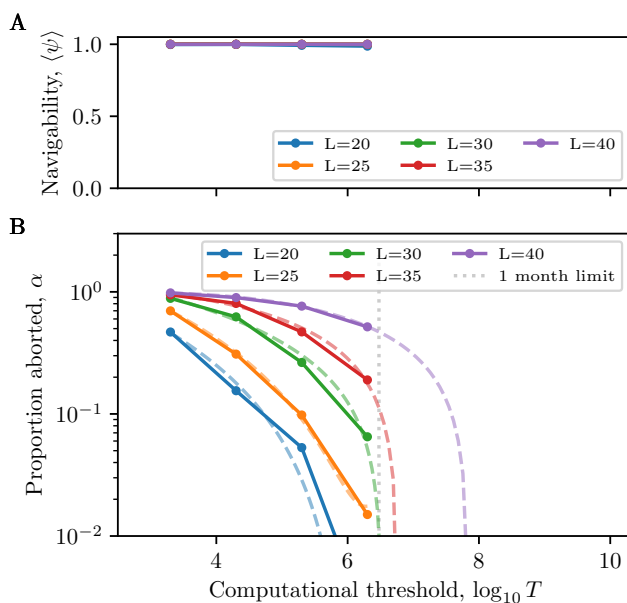

FIG. 5. **(A) Navigability $\langle \psi \rangle$ for different length L for increasing computational threshold T.** Navigability is approximately 1.0 for all computational thresholds suggesting that navigability may be persist for larger computational thresholds. **(B) Proportion of estimations aborted for four different thresholds for different fRNA length L.** Dashed lines provide quadratic interpolations to illustrate potential computational thresholds for which a given abortion threshold may be reached if the fit holds for extrapolation. As a guide, we highlight the computational limit corresponding to one month of chronological time given available computational resources.

To allow us to consider the plausibility of navigable landscapes for longer fRNA ($L > 20$), we explore the effect of changing the computational threshold T (Section IV B 1) at which the search for an accessible path is aborted. We test four orders of magnitude for the threshold $|u_0| + |u_1| < T$ condition: $N_{\text{thresh}} = \{2 \times 10^3, 2 \times 10^4, 2 \times 10^5 \text{ and } 2 \times 10^6\}$. In each case we attempt $N_t = 50$ target phenotypes and for each target $N_s = 20$ source phenotypes and attempt to identify an accessible path, where we record whether a search was successful, unsuccessful or aborted.

In Fig. 5A and Fig. 5B we plot navigability and the proportion of runs that are aborted respectively for the different thresholds against the length of the fRNA sequences. The change in the proportion of aborted runs is pertinent for understanding both how navigability changes when increasing the threshold and also what level of T is required to be able to reasonably estimate navigability for a given length L. With respect to the first point, in Fig. 5A navigability $\langle \psi \rangle \approx 1$ for all lengths L and thresholds T, showing that almost all non-aborted runs have accessible paths. Extrapolating this observation we should expect high navigability for longer length $L > 30$ if greater computation resource were available. With respect to the required computational thresholds for a given length L, we observe, very roughly, that around 50% aborted proportion is reached for $L = 20$ at $T = 2 \times 10^3$, for $L = 35$ at $T = 2 \times 10^5$ and $L = 40$ at $T = 2 \times 10^6$. Extrapolating with quadratic fits we could hypothesise that the aborted threshold could be reduced to 10% for $L = 40$ at between $[2 \times 10^7, 2 \times 10^8]$.