Long-term evolution is surprisingly predictable in lattice proteins

Michael E. Palmer, Amav Moudgil and Marcus W. Feldman

Department of Biology, Stanford University, Gilbert Hall, Stanford, CA 94305-5020, USA

It has long been debated whether natural selection acts primarily upon individual organisms, or whether it also commonly acts upon higher-level entities such as lineages. Two arguments against the effectiveness of long-term selection on lineages have been (i) that long-term evolutionary outcomes will not be sufficiently predictable to support a meaningful long-term fitness and (ii) that short-term selection on organisms will almost always overpower long-term selection. Here, we use a computational model of protein folding and binding called ‘lattice proteins’. We quantify the long-term evolutionary success of lineages with two metrics called the $k$-fitness and $k$-survivability. We show that long-term outcomes are surprisingly predictable in this model: only a small fraction of the possible outcomes are ever realized in multiple replicates. Furthermore, the long-term fitness of a lineage depends only partly on its short-term fitness; other factors are also important, including the ‘evolvability’ of a lineage—its capacity to produce adaptive variation. In a system with a distinct short-term and long-term fitness, evolution need not be ‘short-sighted’: lineages may be selected for their long-term properties, sometimes in opposition to short-term selection. Similar evolutionary basins of attraction have been observed in vivo, suggesting that natural biological lineages will also have a predictive long-term fitness.

1. Introduction

The fitness of a lineage is defined as the expected factor of increase in the number of its member organisms over a single generation. It provides a quantitative measure of how successful a lineage is expected to be, in a particular environmental context, one generation in the future. This measure of fitness is useful because it is predictive: in a new experimental replicate in the same biotic and abiotic environmental context, the lineage can be expected to increase by the same factor.

The standard fitness depends only upon the current standing variation in a lineage. There has been significant discussion of systems’ ability to produce variation, sometimes referred to as evolvability, although definitions have varied widely. Kirschner & Gerhart [1, p. 8420] call evolvability, ‘an organism’s capacity to generate heritable phenotypic variation’, without reference to the fitness of that variation. The view of Houle [2], which associates evolvability with current standing variation and the immediate response of the population to selection, has fallen out of favour. A consensus now supports Wagner & Altenberg [3, p. 970], who proposed, ‘the genome’s ability to produce adaptive variants’. In this sense, evolvability is about the adaptive variation that may be produced, rather than the current standing variation, i.e. the quantitative genetics ‘M-matrix’ as opposed to the ‘G-matrix’ [4]. Altenberg [5, p. 60] provides a rare quantitative definition, ‘the probability that a population generates individuals fitter than any existing’. Other quantitative definitions are ad hoc for particular systems [6–9].

However, we note that the ‘ability to produce adaptive variants’ alone cannot predict evolutionary outcomes, in general. Other factors are important to the long-term success of a lineage: for example, avoidance of production of deleterious variation. Furthermore, this definition of evolvability explicitly...
excludes standing variation, but clearly a lineage must survive in the immediate term (i.e. one generation) in order to succeed in the long term.

Palmer & Feldman [10] introduced two quantitative metrics of the expected evolutionary success of lineages. We suggest that these metrics properly frame evolutionary dynamics in both the short and long term, properly incorporate the affects of both standing and generated variation, and can be predictive like the standard fitness. The first of these metrics is \( W_i(t) \), the ‘k-fitness’ at generation \( t \), defined as the ratio of the expected number of members of a lineage at generation \( t + k \), to the expected number at generation \( t \). (Note that \( W_i(t) \) is the standard, one-generation fitness at generation \( t \).) The second is \( S_i(t) \), the ‘k-survivability’ at generation \( t \), defined as the probability that the lineage will survive to generation \( t + k \), given that it has survived to generation \( t \). We say that a lineage survives (does not go extinct) if it has any living descendants.

Palmer & Feldman [10] previously showed that the \( k \)-fitness and \( k \)-survivability metrics can be predictive in simple in silico models. The question driving this paper is whether these metrics are predictive in more biologically realistic models. In this study, we use an in silico model of evolved ‘lattice proteins’, a model of protein folding and binding, to demonstrate that long-term outcomes are indeed predictable in a biologically realistic model, where the ‘outcome’ of an experiment with several competing lineages is the set of lineages that survive to the end of the experiment. We quantify this predictability with an entropy measure. Furthermore, we find that outcomes are surprisingly predictable; that is, even when the \textit{a priori} uncertainty is very large, such as when many competing lineages are initially present, the evolutionary process removes much of this uncertainty in the long term. We quantify this removal of uncertainty with a negentropy measure. Finally, we show that while outcomes are predictable, they are not fully determined by the initial standing variation, nor predicted by the standard one-generation fitness; thus lattice proteins (evolved as described here) possess a distinct short- and long-term fitness. It is in such a system that long-term selection can sometimes act in opposition to short-term selection.

2. Methods

2.1. Lineages

We define an asexual lineage as a single founding individual, and zero or more generations of its descendants. A sexual lineage consists of a founding population, and zero or more generations of its descendants. In either case, a lineage is monophyletic. For sexuals, we assume that a lineage begins at an irreversible speciation event. The founders of a sexual lineage are all the members of one of the nascent species. This definition implies that two distinct lineages cannot fuse together. A lineage can continue to speciate, however, and all members of sub-lineages thus formed are also members of the original lineage. The non-fusing requirement is automatically satisfied for asexuals. Note that new lineages are founded frequently, especially in asexuals: any asexual individual is itself the founder of a new lineage at every generation. And, any asexual individual is a member of many lineages: the lineages founded by each of its ancestors.

All lineages simulated here are asexual, but the metrics defined in §2.2 apply equally well to sexual lineages, given the non-fusing constraint.

2.2. Metrics of expected evolutionary success

We define the \( k \)-generation fitness and \( k \)-generation survivability of lineages as follows [10]: \( \pi(N(t) = n) \) is defined as the probability that a lineage has \( n \) members at generation \( t \). This can be estimated with a sufficient number of longitudinal experimental replicates in which lineage membership is periodically counted. \( N(t) \) is the expected number of members of the lineage at generation \( t \), namely

\[
N(t) = \sum_n n \pi(N(t) = n).
\]

\( W_i(t) \) is the ‘k-fitness’ at generation \( t \), namely the ratio of the expected number of members of the lineage at generation \( t + k \), to the expected number at generation \( t \):

\[
W_i(t) = \frac{N(t+k)}{N(t)} \quad (t \geq 0, k \geq 1).
\]

\( S_i(t) \) is the ‘k-survivability’ at generation \( t \), namely the probability that the lineage will survive to generation \( t + k \), given that it has survived to generation \( t \):

\[
S_i(t) = \pi(N(t+k) \neq 0 | N(t) \neq 0).
\]

(To refer explicitly to the \( k \)-fitness or \( k \)-survivability of a lineage \( i \) among several lineages, we will add a superscript, as in \( W_i(t) \), \( S_i(t) \), but we will usually omit it for brevity.)

Because selection on lineages is ultimately determined by extinction, the \( k \)-survivability, \( S_i(t) \), is the more fundamental measure of evolutionary success. Nonetheless, \( W_i(t) \) is also useful because, for short time spans, extinctions of lineages may be rare. In this case, the \( k \)-fitness is a practical proxy for the \( k \)-survivability.

We refer to \( k = 1 \) as the immediate term. For example, \( W_i(t) \) (which is equivalent to the standard one-generation fitness used in population genetics) is the \textit{immediate fitness} at generation \( t \), and \( S_i(t) \) is the \textit{immediate survivability} at generation \( t \). We will refer to \( k \) slightly greater than one (\( k > 1 \)) as the \textit{short term}, and \( k \gg 1 \) as the \textit{long term}. For the immediate term (\( k = 1 \)), the evolutionary outcome depends on the standing variation in the lineage at generation \( t \), but cannot depend on the spectrum of variation it may produce. As \( k \) increases, the variation produced by each lineage becomes increasingly important to the outcome.

We will estimate \( W_i(t) \) and \( S_i(t) \) by conducting multiple experimental replicates of a given experimental scenario. The specification of a scenario will typically include: the set of founding protein sequences; the ligand (or ligands) used as binding targets; all details of the lattice protein model; and all details of the population genetic model. Multiple replicates of one scenario will be identical in the scenario specification, and differ only stochastically. Typically, the replicates will differ stochastically when we draw random numbers to determine which mutations occur, or when we sample to produce a finite number of offspring. Thus, the metrics \( W_i(t) \) and \( S_i(t) \) characterize the evolutionary process in the context of that scenario, with stochasticity entering in a well-prescribed way.

It is possible to define ‘broader’ scenarios in which stochasticity enters at other points. For example, in this paper, we typically fix the ligand across replicates. However, one could instead define a procedure for generating a class of ligands, and use a different, random ligand in each replicate. In that case, the metrics \( W_i(t) \) and \( S_i(t) \) would characterize the behaviour of the lineages when binding to a class of ligands, rather than to a particular ligand. In another example, we might allow environments (here, the ligand comprises the only ‘environment’) to vary stochastically over time, differently across replicates; then, our metrics will apply to a particular regime of environmental change. In summary, for any scenario, we must define what is constant across replicates, and what varies across replicates; our metrics will be specific to this context (as is the standard fitness).
Our metrics may also be applied to living systems. Two important requirements are (i) that we can periodically count the membership of a lineage and (ii) that a sufficient number of replicate experiments can be conducted to estimate $\pi(N(t) = n)$ for a particular experimental scenario. For long-lived organisms, there are obvious experimental difficulties, but the lifespan of microorganisms is short enough such that experiments of many generations can be conducted. The most closely related work of which we are aware is that of Woods et al. [11], who measured the equivalent of $W_t(t)$ (the seven-generation fitness) of multiple bacterial lineages at two time points $t_1$ and $t_2$. They did not measure $W_t(t)$ over the long term. However, the technology for such measurement does currently exist: random ‘DNA barcode’ sequences may be inserted into individual cells [12–14]; descendants of these initial lineage founders can then be reliably counted for many generations. Replicate longitudinal experiments of this type could generate estimates of $\pi(N(t) = n)$.

We cannot estimate $\pi(N(t) = n)$ for extinct natural lineages, because we cannot conduct multiple experimental replicates: we have only the single ‘replicate’ comprising natural history. However, the standard one-generation fitness has the same limitation, and it has been an extremely useful tool in biology.

### 2.3. An entropy measure of the predictability of evolutionary outcomes

$W_t(t)$ and $S_t(t)$ describe the evolutionary process in a given scenario: they summarize what has happened in the replicates we have conducted, and they yield a prediction about what the outcome will be in a new replicate. We next define a measure of predictability in a given evolutionary scenario.

Call $s(i, j, k)$ the ‘survival state’ of lineage $i$ in replicate experiment $j$ at generation $k$. $s(i, j, k) = 1$ if lineage $i$ has any living members at generation $k$ of replicate $j$, and 0 otherwise. The ‘joint survival state’ $M(i, j, k)$ is a binary vector composed of elements $s(i, j, k)$, ordered by $i$; it indicates the realized survival state of all lineages at generation $k$ of replicate $j$. If there are $L$ lineages, there are $2^L$ possible joint survival states. By averaging over multiple replicates, we can estimate $p_m(k)$, the probability that each joint survival state will occur at time step $k$. We then compute the entropy of the joint survival state at generation $k$,

$$H_{\text{joint}}(k) = -\sum_k p_m(k) \log_2(p_m(k)),$$

as a measure of the unpredictability of evolutionary outcomes in a particular evolutionary system. Note that

$$0 \leq H_{\text{joint}}(k) \leq L \quad \text{(in units of bits),}$$

$H_{\text{joint}}(k)$ estimates our average uncertainty in predicting the outcome of a new experimental replicate at generation $k$. If $H_{\text{joint}}(k) = 0$, then the same state always occurs at generation $k$: the outcome is perfectly predictable. If $H_{\text{joint}}(k) = L$, then the outcomes are equally likely, and thus maximally unpredictable.

We also compute $H(k)$, the entropy of the survival state of each distinct lineage $i$. Below, we will show that these lineage entropies, $H_i(k)$, are useful for interpreting the variation in the joint entropy, $H_{\text{joint}}(k)$, over time. The $k$-survivability of lineage $i$, $S_i(0)$, is the probability of its survival to generation $k$ (given that it was present at generation 0), so $H(k)$ turns out to be a simple function of $S_i(0)$:

$$H(k) = -S_i(0) \log_2(S_i(0)) - (1 - S_i(0)) \log_2(1 - S_i(0)).$$

Because there are only two possible survival states of one lineage,

$$0 \leq H(k) \leq 1 \quad \text{(in units of bits).}$$

The joint entropy and the lineage entropies obey the following inequalities:

$$H_{\text{joint}}(k) \geq \max[H_i(k)]$$

and

$$H_{\text{joint}}(k) \leq \sum_i H_i(k).$$

### 2.4. Lattice protein model

The lattice protein model [15,16] is a well-studied, simple model of protein folding, in which the amino acid residues of a protein sequence are constrained to points on a grid (either two-dimensional, as here, or three-dimensional). Although the lattice protein model is much less physically realistic than modern, atomic-level protein folding models, it is also vastly cheaper, computationally. This permits the evolution of populations of lattice proteins, over many generations, in multiple replicates, on a modern computer cluster. Despite their relative simplicity, lattice proteins retain important properties of real proteins.

The heritable genotype in the model consists of a sequence of amino acid residues, of length 12 for the simulations presented here, chosen from the 20 biological amino acids. Occasional mutations (rates below) may change the residue at a particular locus.

The fitness of one protein sequence is computed by first ‘folding’ the sequence into its minimum energy conformation, and then computing its binding affinity to a target ligand, as follows.

#### 2.4.1. Protein folding

A protein sequence of a finite length may ‘fold’ into a finite number of possible conformations on the two-dimensional lattice. Examples of folded lattice proteins are shown in figure 1; the folded proteins are indicated by upper-case letters. All pairs of abutting amino acid residues contribute to the total interaction energy $E(C_i)$ of a conformation $C_i$, according to a set of pairwise interaction energies taken from Miyazawa & Jernigan [17, table V]. The stability of a particular folded conformation $C_i$ is given by

$$\Delta_G(C_i) = E(C_i) + T \ln(Q(T)) - e^{-E(C_i)/T},$$

where $Q(T) = \sum e^{-E(C_i)/T}$ is the partition function. In vivo, real proteins will spend time in a number of possible conformations; more time is spent in lower-energy conformations. Here, we make the simplification of considering only the most stable conformation for a particular sequence. If the lowest $\Delta_G$ for a sequence is greater than zero, then the sequence has no energetic ‘preference’ for folding, and receives a fitness of zero.

#### 2.4.2. Protein binding to a target ligand

For all sequences that do fold stably (i.e. $\Delta_G < 0$), we compute the maximum binding affinity to a target ligand, indicated by lowercase letters in figure 1. The ligand is a fixed-conformation protein of length six residues. For a given sequence, call $C_{\text{min}}$ the conformation with the lowest $\Delta_G$. All possible position and rotation combinations of the ligand in relation to the protein are tested to find the minimum binding energy (sum of the interaction energies of abutting residues of the sequence and ligand), relative to conformation $C_{\text{min}}$. Call $\text{BE}_{\text{min}}(C_{\text{min}})$ the lowest binding energy over all these combinations. The fitness of each protein sequence is $f = e^{-\text{BE}_{\text{min}}(C_{\text{min}})}$. See [15,16,18] for additional details.

### 2.5. Population genetic model

Reproduction is asexual, and generations do not overlap. Each deme (local population) has associated with it a single target ligand. In some of the experiments below, the population is subdivided into multiple demes. At each generation, the fitness of
each protein sequence in a deme is computed as above. Each of the \( N \) sequences is cloned and placed into the offspring generation with probability proportional to its fitness until there are \( N \) offspring. Random mutations are applied, and if there are multiple demes, random migration occurs among them.

2.6. Experimental design

We screened many random lattice protein sequences of length 12 to generate 1600 sequences that fold stably \((\Delta G_f < 0)\), in an approximately uniform distribution of \( \Delta G_f \) between \(-3.5 \) and \( 0 \) kcal mol\(^{-1}\). In an experimental scenario using \( L \) lineages (e.g. \( L = 4 \)), we select \( L \) ‘founding’ sequences out of the 1600 pre-screened stable sequences. We generate \( N/L \) clones of each founding sequence to construct \( L \) initial lineages with a total of individuals. In a given replicate of a scenario, the membership of each lineage will vary over time, and lineages may go extinct. In each replicate, we record the number of living members of each of the \( L \) lineages at each generation, in order to estimate \( \pi[N(t)] = \pi \), which allows us to compute \( W_k(t) \) and \( S_k(t) \), for each lineage, for a given scenario.

In the first set of experiments below, we generated 32 scenarios, each of which comprised a fixed set of \( L = 4 \) founders (chosen from the 1600 stable sequences), and a single fixed ligand. A single deme \((D = 1)\) of fixed population size, \( N = 16384 \), was used. We ran \( R = 512 \) replicates of each scenario (which will ensure a 95% probability of the total number of outcomes will be estimated within 5% [19]), for a duration of \( r_{\text{max}} = 2000 \) generations each. Stochasticity across replicates of a given scenario enters only during mutation, and sampling to the population size \( N \). Mutations were applied at the rate of \( \mu = 0.0005 \) per residue per generation (producing about 10 mutations per generation).

In the second set of experiments below, the number of founders was increased to \( L = 128 \). Again, we generated 32 such scenarios, and \( N = 16384 \), \( D = 1 \), \( R = 512 \) and \( r_{\text{max}} = 2000 \).

In the third set of experiments below, we subdivided the population into \( D = 32 \) demes of \( N = 512 \) individuals each (total population, \( N_D = 16384 \)). For each of 16 scenarios, we selected \( L = 128 \) founders, assigning them to fixed demes within a scenario. One random ligand was assigned to each deme, and this was fixed within a scenario. Again, we performed \( R = 512 \) replicates; however, \( r_{\text{max}} \) was increased to 10 000 generations, because population subdivision slows the attainment of equilibrium. Here, additional stochasticity enters in the selection of random migrants between the demes.

3. Results

3.1. Evolution is predictable in lattice proteins

In figure 2, each of the three rows presents the results of a distinct experimental ‘scenario’ of our first set of experiments, which involve a single deme and \( L = 4 \) founding lineages. (We conducted 32 such scenarios, selecting three of them for inclusion in figure 2.) Each of the pre-screened stable founding sequences is assigned a unique number from 1 to 1600; each lineage is uniquely identified by the sequence number of its founder. The legends in each row identify the four lineages included in the scenario: for example, in the
first row of figure 2, sequences 0736, 1051, 0742 and 0024 were cloned to found the four lineages. In the legends, the lineage numbers are ordered by the total ‘weight’ [20] of their lineage, i.e. the sum of the counts of their members over all generations. (This ‘weight’ is a simple way to rank the lineages in a given scenario.) The three columns show the \( k \)-fitness, \( k \)-survivability, and the entropy, \( H(k) \), of the survival state, including the entropy of the joint survival state of all lineages (black). For \( R = 512 \) replicates, the maximum measurable entropy is \( \log_2(512) = 9 \) bits, as indicated by the dotted line in each of the right panels. However, for these scenarios, the actual entropy remains well below this maximum measurable level. (The folded proteins shown in figure 1 belong to the green lineage (0665) in the third row of figure 2. See also figure 8.)

In the top left panel of figure 2, where the \( k \)-fitness is plotted for the first scenario, lineage 0736 (red) is shown to be the best adapted to the target ligand in the short term. Its \( W_k(0) \) is approximately 2.0, which can be read along the \( y \)-axis (note that generations are plotted on a log scale, and the \( y \)-axis corresponds to generation 1). That is, in the first generation, the membership of the red lineage increases by a factor of approximately 2.0 on average (over \( R = 512 \) replicates). It quickly comes to dominate the population, with a medium- and long-term \( W_k(0) \) of approximately 4; this is the maximum possible \( k \)-fitness, because each lineage starts with \( N/L \) members (with \( L = 4 \)) and can increase to at most \( N \). In the centre panel of the first row, where \( S_k(0) \) is plotted, it can be seen that lineages 0742 (blue) and 0024 (purple) have a 0 per cent estimated probability of surviving past generation 16, or 20, respectively. Lineage 1051 (green) has an almost 100 per cent chance of extinction shortly thereafter, leaving lineage 0736 with an almost 100 per cent chance of surviving (again, as estimated over 512 replicates) in the medium and long term.

The entropy of the survival state of each lineage \( H_k(k) \) is shown in the right panel of each row (coloured lines), along with the entropy of the joint survival state \( H_\text{joint}(k) \) (black line). For \( R = 512 \) replicates, the maximum observable joint entropy is \( \log_2(512) = 9 \) bits; this maximum reading would be achieved if a different outcome (joint survival state) occurred in every replicate. The horizontal dotted line across the top of each panel in the right column of figure 2 indicates this maximum possible measurement.

In the top right panel of figure 2, after about generation 30, the joint entropy (black line) for the first scenario is almost zero: in almost all replicates, the same survival state occurs, corresponding to the survival of lineage 0736 and the extinction of all other lineages. There is almost no uncertainty in the joint survival state after this point; thus, evolution is very predictable after \( k = 30 \). If we performed a new replicate of this scenario, we could be very confident that lineage 0736 (red) would again dominate in the long term. (There is a small chance that lineage 1051 (green) would survive.) Notice that before generation 4 there is also
zero uncertainty: all lineages reliably survive until at least generation 3. The $k$ of highest joint uncertainty (nearly 1.5 bits) is $k = 7$. In an additional replicate, it would be uncertain whether the blue and purple lineages would yet be extinct, as can be seen by the plots of $H(k)$ for lineages 0742 and 0024 (blue and purple lines in top right panel). A secondary peak in joint uncertainty occurs because the green lineage tends to go extinct at an uncertain moment between generations 10 and 30.

In the second row of figure 2, lineage 1184 (green) is initially better adapted to the target ligand, and increases rapidly in membership; however, lineage 1116 (red) adapts more rapidly, and supersedes lineage 1184. Thus, although the green lineage has a superior $W_t(0)$ of about 2.0 (as can be read along the y-axis of the left panel in the second row of figure 2), the red lineage has a superior $W_{0000}(0)$ of 3.75. Ultimately, the red lineage has approximately 95 per cent chance of surviving in the long term, and the green lineage has approximately 5 per cent chance of surviving in the long term (see middle panel of second row). (The blue lineage also has a small chance of surviving in the long term.) Thus, while the green lineage is initially better adapted to the ligand, the red lineage has higher long-term survivability, to the extent that red drives green extinct approximately 95 per cent of the time.

Note that, at generation 4, the $k$-fitness (left panel, second row) of lineage 1116 reaches a minimum of approximately 0.3. Because each lineage was initialized with 4096 members, this means that the red lineage had approximately 1229 members at generation 4, on average this means that the red lineage had approximately 1229

0.3. Because each lineage was initialized with 4096 members, and increases

rapidly in membership; however, lineage 1116 (red) adapts more rapidly, and supersedes lineage 1184. Thus, although the green lineage has a superior $W_t(0)$ of about 2.0 (as can be read along the y-axis of the left panel in the second row of figure 2), the red lineage has a superior $W_{0000}(0)$ of 3.75. Ultimately, the red lineage has approximately 95 per cent chance of surviving in the long term, and the green lineage has approximately 5 per cent chance of surviving in the long term (see middle panel of second row). (The blue lineage also has a small chance of surviving in the long term.) Thus, while the green lineage is initially better adapted to the ligand, the red lineage has higher long-term survivability, to the extent that red drives green extinct approximately 95 per cent of the time.

Note that, at generation 4, the $k$-fitness (left panel, second row) of lineage 1116 reaches a minimum of approximately 0.3. Because each lineage was initialized with 4096 members, this means that the red lineage had approximately 1229 members at generation 4, on average this means that the red lineage had approximately 1229

members at generation 4, although the number may fluctuate across replicates. If the total population size ($N = 16384$) were smaller (not shown), the red lineage might have gone extinct early in some replicates, rather than discovering the adaptation(s) that led it to a high $k$-fitness in the long term. (See Palmer & Feldman [10] for a discussion of the relationship between population size and the conflict between long-term and short-term fitness. This conflict is also discussed by Clune et al. [21].) In the rightmost panel of the second row, we see a joint survival entropy (black line) of only 0.4 bits in the long term. The plots of the lineage entropies (coloured lines in rightmost panel) indicate that the purple and blue lineages contribute to higher joint entropy mostly in the initial generations ($k$ less than about 35). Later, the joint entropy is due to uncertainty in the survival state of the red and green lineages.

In the third row of figure 2, again the lineage that does best initially (purple, 0788) is not the one that dominates in the long term (red, 0140); although the purple lineage has a $W_t(0)$ of approximately 1.6 (as seen in the bottom left panel), it has a 100 per cent chance of extinction by approximately generation 30 (as seen in the bottom centre panel). A maximum uncertainty in the joint survival state (approx. 1.8 bits) arises around generation 13, but the long-term joint entropy is less than one bit.

For this first set of experiments ($L = 4$, in a single deme), we performed 32 different scenarios, with $R = 512$ replicates each. (Just three of them were included in figure 2.) Figure 3 plots the entropy of the joint survival state, averaged over all 32 scenarios. This indicates how predictable the $k$-generation evolutionary outcome will be in general for such scenarios (i.e. for $L = 4$, $N = 16384$, with sequences and ligands selected in the same way). In the long term ($k > 500$), there is low uncertainty in the outcome: only about 0.9 bits, more certain than the

3.2. Is evolution still predictable for larger numbers of competing lineages?

For our second set of experiments, we increase $L$ to 128. In figure 4, we plot several scenarios (again, one scenario per row) with a larger number of founding lineages: $L = 128$. As before, $N = 16384$, $D = 1$, and $R = 512$ replicates. For clarity, only the five lineages with the heaviest ‘weight’ (sum of member count across generations) are explicitly named in the panel legends for each of three scenarios (rows); however, all 128 lineages are plotted until extinction.

The top row of figure 4 shows a scenario for which a single lineage, 0957 (red), is initially fittest, and also dominates in the long term. It attains a long-term fitness of about 60; the maximum possible fitness is 128 in this case (because each lineage is initialized with $N/L$ members and can increase to at most $N$, where $L = 128$). As seen in the top centre panel, the red lineage has an approximately 45 per cent chance of long-term survival, with several other lineages having non-zero chances of surviving in the long term.

In the top right panel of figure 4, the maximum observable joint entropy is indicated with a horizontal dotted line: because there are $R = 512$ replicates, the maximum observable joint entropy is $\log_2(512) = 9$ bits. Note that because there are $L = 128$ lineages, there are $2^{128}$ possible joint survival states, or a maximum actual joint entropy of $\log_2(2^{128}) = 128$ bits. Thus, between generations 3 and 6, the measured entropy is ‘clipped’ by the dotted horizontal line: the actual entropy is higher than we can observe with only 512 replicates. This is because many of the 128 lineages go extinct in the first six generations, but their precise order of extinction is uncertain. However, in all three of the scenarios shown, the actual entropy decreases to well below the maximum observable nine bits in the long term.

The second row shows a scenario in which the initially best-adapted lineage does not dominate in the long term: the green lineage (0898) is initially more fit in the short term, but later fails to adapt as well as the red lineage (0001), which dominates in the long term. The long-term outcome is quite certain: only around 0.1 bits of joint entropy.
In the third row, a variety of lineages may survive to the long term in different replicates; this produces (as seen in the bottom right panel) a higher long-term joint entropy of about 3.7 bits: the long-term outcome in this scenario is much more uncertain.

In figure 5, we plot the average joint entropy over 32 different $L = 128$ scenarios. Surprisingly, the average joint entropy in the long term is still quite low: only about two bits! This is comparable to only four (or so) of the 128 lineages being in contention to dominate in the long term.

While it is interesting that the $L = 4$ scenarios yielded an average joint uncertainty of 0.9 bits in the long term (figure 3), it seems even more remarkable that the $L = 128$ scenarios should yield an average joint uncertainty of only two bits (figure 5), given that there are so many more possible outcomes when $L = 128$. How should we compare these two situations quantitatively?

### 3.3. A negentropy measure of uncertainty removed by an evolutionary process

In thermodynamics and information theory, a quantity called the negentropy [22,23] is used to describe the difference between the maximum possible uncertainty in a system and the current uncertainty. If $H_{\text{max}}$ is the maximum entropy of a system and $H_{\text{curr}}$ is the current entropy, then the ‘negentropy’ is defined as

$$N = H_{\text{max}} - H_{\text{curr}}.$$ 

This quantity is also useful for describing evolutionary processes. For an evolving system of $L$ lineages, there are $2^L$ possible values of the ‘joint survival state’ vector. The most uncertain distribution is when all states have an equal probability of $1/2^L$, producing a maximum entropy of $H_{\text{joint max}} = L$ bits. Thus, if $H_{\text{joint}}(k)$ is the joint entropy of a set of $L$ lineages at generation $k$, then the joint negentropy at generation $k$ is

$$N_{\text{joint}}(k) = L - H_{\text{joint}}(k).$$

This tells us how much uncertainty is removed by observing many replicates of a particular evolutionary process (as specified by the scenario).
\( N_{\text{joint}}(k) \) is equivalent to the mutual information [24,25] \( I(X;Y) \) between \( X \), the outcome of a single new replicate, and \( Y \), the set of outcomes of the \( R \) replicates:

\[
N_{\text{joint}} = I(X;Y) = H(X) - H(X|Y).
\]

This can be read as ‘the uncertainty in \( X \) (without knowing \( Y \)) minus the uncertainty in \( X \) given \( Y \).’

We will use the following qualitative terminology. A scenario is predictable, but not surprisingly so when \( H_{\text{joint}}(k) \) is low, but \( N_{\text{joint}}(k) \) is also low. A scenario is surprisingly predictable when \( H_{\text{joint}}(k) \) is low, and \( N_{\text{joint}}(k) \) is high. A scenario is simply not predictable when \( H_{\text{joint}}(k) \) is high. In our first set of experiments (\( L = 4 \)), in the long term, \( H_{\text{joint}}(k) \) is 0.9 (according to figure 3), yielding \( N_{\text{joint}}(k) \) is 3.1. In our second set of experiments (\( L = 128 \)), \( H_{\text{joint}}(k) \) is 2.0 (according to figure 5), yielding \( N_{\text{joint}}(k) \) is 126. In both cases, evolution is quite predictable; but in the latter case, the predictability is much more surprising.

### 3.4. A structured metapopulation still yields surprising predictability

Note that, in the Wright–Fisher model with fixed population size and fixed fitnesses, one lineage will eventually fix in a single deme. We excluded this knowledge of the evolutionary process when computing the maximum possible entropy. However, assuming this knowledge a priori would eliminate much uncertainty: the number of possible long-term outcomes would go from \( 2^L \) to \( L \). By contrast, in a structured population of multiple demes, it may take much longer for a single lineage to fix [26]: multiple lineages may persist for a long time in a subset of the demes, especially if selection is different in each deme. This raises the question of whether evolution will still be so surprisingly predictable in multiple demes.

Therefore, we conducted a third set of experiments, in which the population was divided into \( D = 32 \) demes, with migration among them at a rate of \( \text{mig} = 0.01 \) per individual per generation. \( L = 128 \), and the assignment of founders to demes was fixed per scenario. The population size in each deme was \( N = 512 \), for a total population of \( ND = 16384 \), and \( R = 512 \). Because evolution can take longer to reach a stationary distribution in a structured population, we increased \( \text{gen}_{\text{max}} \) to 10 000 generations. A different target ligand was used in each of the 32 demes (fixed per scenario), so that for a lineage to dominate the entire metapopulation, it must adapt to bind a different ligand in every deme. Nevertheless, it was still possible to find scenarios in which a single lineage typically dominates in the long term.

The first row of figure 6 shows such a scenario: the red lineage (0097) dominates from the beginning, and has an estimated 100 per cent survival probability in the long term. The long-term entropy in the joint survival state of about 3.5 bits is due to the possibility that a variety of lineages may coexist (i.e. in different demes) with the ever-present red lineage.
still very predictable.

In summary, for interesting, non-degenerate experimental setups, we have shown that evolution can be surprisingly predictable in lattice proteins.

3.5. The likelihood of long-term success is a property of the lineage

In figure 8, we plot the evolutionary trajectories of individual sequences from the third scenario (row) of figure 2. Each of the four panels plots the trajectories of the members of four lineages: 0140 (red), 0665 (green), 1018 (blue) and 0788 (purple). In figure 2, \(N(t)\), the expected number of members at generation \(t\), estimated across all replicates (\(R = 512\)), is shown on a log scale. The black line in each panel shows the total \(N(t)\) for the specified lineage as a whole. The coloured lines in each panel show \(N(t)\) for each unique sequence belonging to the lineage; all sequences that attain an \(N(t)\) of 10 or more at some generation \(t\) are plotted, and the 10 sequences attaining the highest peak values of \(N(t)\) are labelled with their sequence names. In all panels, the founding sequence of each lineage (i.e. FCTF KIIN CEWV for lineage 0140 (red), MVNL TLPV VTLM for lineage 0665 (green), FLEL TCLN NPCF for lineage 1018 (blue) and IWPK AHML SHNY for lineage 0788 (purple)) goes extinct within 10 generations. We include these plots to illustrate how natural it is to consider the lineage as an entity that possesses a long-term fitness and survivability: no single sequence completely determines the long-term expected success of the lineage; rather, success is determined partly by the immediate fitness of the founding sequence, and partly by the fitness of descendant sequences.

Significant neutral divergence is apparent at high \(k\) in the red and green panels. (The 10 highest weight sequences in the upper right panel (green, 0665) are the same sequences shown in folded form in figure 1.)

4. Discussion

In a particular model, there are two ways that \(W_k(t)\) and \(S_k(t)\) could fail to have the predictive utility of the standard one-generation fitness: (i) evolutionary outcomes could simply not be very predictable over the long term or (ii) outcomes could be so strongly predicted by the short-term fitness that the long-term metrics are uninteresting. In the lattice protein model, we have shown that long-term evolutionary outcomes in a single deme and in multiple demes are predictable, and surprisingly so: the entropy of the joint survival state is low in the long term, whereas its negentropy is high. We have shown that the immediate-term adaptedness of a lineage contributes to, but does not entirely determine, its chance of success in the long term. The long-term success of a lineage depends, additionally, on its tendency to generate adaptive variation, its tendency to avoid deleterious variation, its ability to diversify in phenotype, and its ability to physically disperse [10].

That long-term outcomes are predictable, but not fully determined by the short-term fitness, implies that lineages may be selected for their long-term properties, sometimes in opposition to short-term selective forces; evolution need not be short-sighted [21]. Selection at some time scales and in some situations may best be summarized as selection on organisms. However, in some cases, often in the longer term, lineage-level selection may provide the best model by describing evolution both simply and with sufficient accuracy. We might slightly increase our descriptive accuracy by including the voluminous details of organismal-level selection; but this would hide higher-level consistencies that could be concisely captured as lineage-level selection.

The ‘viewpoint’ of the lineage-as-individual is quite alien to that of the organism-as-individual. To the lineage, organisms are expendable, until their numbers become small. The successful lineage is one that continues to have some ability to diversify in phenotype, and its ability to physically disperse [10].
Figure 8. Plots of expected counts, $N(t)$, for each unique sequence belonging to the four lineages shown in the third scenario of figure 2 ($L = 4$). All sequences attaining $N(t) \geq 10$ at some $t$ are plotted; the 10 sequences with the highest peak $N(t)$ are labelled. Long-term success is best considered as a property of a lineage. (The lineages shown are taken from the scenario in the third row of figure 2. The 10 highest ‘weight’ sequences of lineage 0665 (green) are shown in folded form in figure 1.)

Bloom et al. [27] claimed that the stability of a folded protein promoted ‘evolvability’ (defined differently). By contrast, we found a weak negative correlation between the stability of the founding sequence and its long-term fitness (not shown). We have come to suspect that there may be no simple property of a sequence that strongly predicts its long-term fitness. Instead, the mutational landscape, the fitness landscape (binding process) and the genotype–phenotype map (folding process), together with the other scenario details, define a multi-dimensional set of evolutionary ‘pathways’ that the descendants of a given founder are likely to follow. These pathways cannot be simply predicted from the high-level properties of a single sequence, e.g. from its initial stability. Although these high-dimensional pathways are difficult for us to visualize, they are persistent across replicates of a given scenario. They describe basins of attraction that pull the lineages through the evolutionary process, to repeatable outcomes. The long-term fitness and survivability of a lineage characterize these basins of attraction in a given scenario. Moreover, analogous to how the survival state of all lineages $i$ can be combined into a joint survival state, the entire evolutionary system can be considered to be moving along a much higher-dimensional set of pathways, from its initial state (all lineages with equal numbers of members) to a restricted set of possible outcomes (one or a few particular lineages surviving).

We have previously discussed the metrics $W_i(t)$ and $S_i(t)$ in the context of simpler models [10]. In this study, our intent has been to demonstrate the metrics in the more biologically realistic model of lattice proteins. The lattice protein model resembles real protein folding, but is simpler in several ways: (i) here, it is in two dimensions, rather than three; (ii) we consider only the lowest-energy conformation of the protein, and the lowest-energy binding site; and (iii) the proteins are short. Nonetheless, these model proteins do retain important features of real proteins, producing a complex evolutionary landscape that has some compatibility to that of real proteins. If long-term predictability had been absent in lattice protein evolution, or if the short-term fitness had completely predicted the long-term fitness, then we would have held out little hope for the utility of $W_i(t)$ and $S_i(t)$ in biological models.

Do real proteins, and real lineages of organisms, also have a predictive long-term fitness that is distinct from their short-term fitness? There is evidence in real proteins [28–30] and in real bacteria [11,31] that the number of selectively permissible evolutionary pathways from a maladapted to an adapted state may be quite low, which would enhance predictability of outcomes. We think it likely that such a long-term fitness, and surprising predictability, will soon be measured in experimental evolution with micro-organisms.

This work was supported in part by NIH grant no. GM28016. Many thanks to Jesse Bloom for graciously sharing his lattice protein folding code [15,18,27,32]. Thanks to our anonymous reviewers for important insights and suggestions.
References


