

## Transition paths in Potts-like energy landscapes: General properties and application to protein sequence models

Eugenio Mauri <sup>\*</sup>, Simona Cocco, and Rémi Monasson<sup>†</sup>Laboratory of Physics, *École Normale Supérieure*, CNRS UMR 8023, and PSL Research, Sorbonne Université, 24 Rue Lhomond, 75231 Paris Cedex 05, France

(Received 5 April 2023; accepted 7 August 2023; published 28 August 2023)

We study transition paths in energy landscapes over multicategorical Potts configurations using the mean-field approach introduced by Mauri *et al.* [*Phys. Rev. Lett.* **130**, 158402 (2023)]. Paths interpolate between two fixed configurations or are anchored at one extremity only. We characterize the properties of “good” transition paths realizing a trade-off between exploring low-energy regions in the landscape and being not too long, such as their entropy or the probability of escape from a region of the landscape. We unveil the existence of a phase transition separating a regime in which paths are stretched in between their anchors from another regime where paths can explore the energy landscape more globally to minimize the energy. This phase transition is first illustrated and studied in detail on a mathematically tractable Hopfield-Potts toy model, then studied in energy landscapes inferred from protein sequence data.

DOI: [10.1103/PhysRevE.108.024141](https://doi.org/10.1103/PhysRevE.108.024141)

### I. INTRODUCTION

Characterizing transition paths in complex, rugged energy landscapes is a relevant issue in statistical physics and in other fields. In evolutionary biology, for instance, a fundamental problem is to sample mutational paths that, starting from a given protein (described as a given sequence of amino acids), introduce single mutations at each step, in such a way that all the intermediate sequences do not lose their biological activity. Characterizing these paths would be crucial to better understand the navigability of fitness landscapes [1,2]. From a statistical mechanics point of view, substantial efforts have been done to characterize how systems dynamically evolve in complex, e.g., glassy, landscapes to escape from metastable local minima and reach lower-energy equilibrium configurations. In this context, recent works have focused on  $p$ -spin-like energy functions with quenched interactions, generally giving rise to very rugged landscapes [3,4].

Of particular interest is the case of energy functions  $E(\mathbf{v})$  defined over Potts-like configurations  $\mathbf{v} = (v_1, v_2, \dots, v_N)$ , where the variables  $v_i$  can take one out of  $A$  categorical values [5]. Consider a path starting from a configuration  $\mathbf{v}_{\text{start}}$ , and exploring  $T$  subsequent configurations. The last configuration (extremity) of the path,  $\mathbf{v}_{\text{end}}$ , can be free or fixed, depending on the problem of interest. Intermediate configurations along the path can *a priori* take any of the  $A^N$  possible values, which we refer to as global configuration space below. However, the initial and final configurations define for each variable  $i$  (at most) two categorical values, defining a subspace with  $2^N$  configurations, which we call direct space in the following. The question we address in the present work may be infor-

mally phrased as follows: Under which conditions are good transition paths naturally living in or close to the direct space rather than in the global one?

This question is of conceptual interest, but has also practical consequences. Consider again the case of mutational paths joining two protein sequences (see Fig. 1). Due to the huge number of possible paths in the global space ( $A = 20$ ), mutagenesis experiments generally restrict to direct paths [6]. However, constraining paths to be direct may preclude the discovery of much better global paths, involving mutations and their reversions and reaching more favorable regions in the sequence space (Fig. 1). Mutational models have demonstrated that exploring the fitness landscape beyond the direct space can enhance adaptation [7]. Additionally, the existence of such beneficial “global” mutations could provide valuable insights into the properties of the fitness landscape, e.g., the presence of high-fitness regions responsible for the deviation of the paths from the direct subspace.

Whether paths remain direct or explore the global space will depend on their length, on their “elastic” properties (defined by the mutation process), as well as on the nature of the energy (minus fitness) landscape. In particular, fitness landscapes can be complex and with many good regions (surrounding local maxima) that attract the path outside the direct space. In this article, we introduce a minimal landscape model, corresponding to a Hopfield-Potts model with  $P$  patterns, defining a rank- $P$  pairwise coupling matrix between the  $v_i$ 's. The energy of a path is then defined as the sum of the Hopfield-Potts energies of the intermediate configurations, and of elastic contributions measuring the dissimilarities between successive configurations (Fig. 1). When  $N$  is sent to infinity while keeping  $P$  finite, transition paths in this landscape can be analytically studied using the mean-field framework introduced in [8]. Two sets of time-dependent order parameters, where the time  $t$  denotes the coordinate

<sup>\*</sup>eugenio.mauri@phys.ens.fr<sup>†</sup>remi.monasson@phys.ens.fr

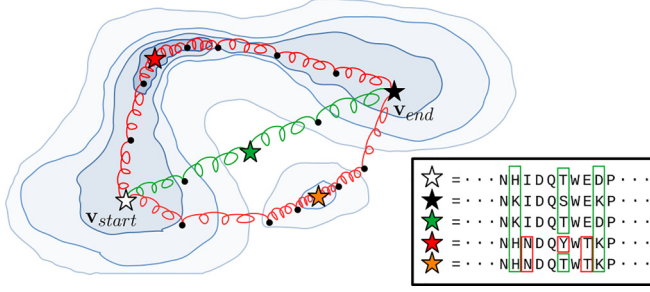


FIG. 1. Mutational paths between two subfamilies in the sequence landscape associated to a protein family. Darker blue levels correspond to increasing values of the protein fitness. Paths are either direct (green, central line: each site carries the amino acid present at the same position in the initial or in the final sequence) or global (red, top and bottom lines: no restriction on amino acids), making possible the exploration of high-fitness regions.

along the path, are needed in the mean-field theory: (1) the average projections  $m_t^\mu$  of configurations along the patterns  $\mu = 1, \dots, P$ , and (2) the overlaps  $q_t$  between successive configurations  $\mathbf{v}_t$  and  $\mathbf{v}_{t+1}$  along the path. We explain how the mean-field theory allows for a detailed study of the statistical properties of transition paths, such as their entropy, the escape probabilities from a local minima, and their direct vs global nature. In particular, we show that, depending on the stiffness coefficient of the elastic term acting on  $q_t$ , two regimes can be encountered. Paths under high tension are likely to remain confined within the direct space. For low tension, paths are likely to explore the global path to minimize their energy. The nature of the phase transition, such as the critical tension and the time behavior of the order parameters, is analytically unveiled. We also compute the entropy of transition paths interpolating between the anchoring (initial and final) configurations.

Importantly, the mean-field formalism can be extended to restricted Boltzmann machines (RBMs) trained on natural protein sequence data [9,10]. While protein sequence landscapes are *a priori* unknown, the vast usage of data-driven models has allowed capturing the relation between protein sequences and their functionalities over the past years. Unsupervised machine-learning approaches, such as Boltzmann machines or variational autoencoders, can be trained on homologous sequence data to score the sequences, thus defining an empirical energy. These models have been shown to be generative, meaning they can be used to design novel proteins with functionalities comparable to natural proteins [11,12]. In this context, RBMs can be seen as a natural extension of Hopfield-Potts models, with the weights connecting the visible (sequence) and hidden (representation) layers playing the role of patterns, and the energy not necessarily being quadratic in the projections  $m^\mu$ . Hereafter, we apply a RBM to sequence data from both *in silico* and real proteins, demonstrating that direct-to-global phase transitions are found in transition paths built from these data-driven models.

This paper is organized as follows. Section II provides the main definitions and an overview of the basic properties of transition paths in Hopfield-Potts landscapes. In Sec. III, we revisit the transition-path framework introduced in [8],

focusing on the expression of the mean-field free energy as a function of the order parameters  $m_t^\mu$  and  $q_t$  for a generic Hopfield-Potts energy. We also conduct a detailed study of the direct-to-global phase transition in the minimal case of  $P = 2$  nonorthogonal patterns. In Sec. IV, we apply our mean-field approach to energy landscapes inferred from *in silico* lattice-protein models [13] to benchmark our approach, and to natural protein sequence data associated with the WW domain, a short protein domain involved in signaling [10]. Conclusive remarks can be found in Sec. V.

## II. DEFINITIONS AND OVERVIEW OF THE RESULTS

### A. Mutational paths over configuration space

We consider an energy landscape  $E_{\text{model}}(\mathbf{v})$  over  $N$ -dimensional Potts configurations  $\mathbf{v}$  (see Fig. 1).  $E_{\text{model}}$  can be either derived from first principles or inferred from some available data using machine-learning methods. Following [8], we associate to each path  $\mathcal{V} = \{\mathbf{v}_{\text{start}}, \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{T-1}, \mathbf{v}_{\text{end}}\}$  an energy  $\mathcal{E}(\mathcal{V})$ . This energy is the sum of the energies of the intermediate configurations along the path, and of elastic contributions decreasing with the similarities between pairs of successive configurations. We denote by  $\Phi$  the elastic potential. The energy of a path (divided by  $N$ ) is then

$$\mathcal{E}(\mathcal{V}) = \frac{1}{N} \sum_{t=1}^{T-1} E_{\text{model}}(\mathbf{v}_t) + \Phi(q(\mathbf{v}_{\text{start}}, \mathbf{v}_1)) + \sum_{t=1}^{T-2} \Phi(q(\mathbf{v}_t, \mathbf{v}_{t+1})) + \Phi(q(\mathbf{v}_{T-1}, \mathbf{v}_{\text{end}})), \quad (1)$$

where the overlap  $q(\mathbf{v}_t, \mathbf{v}_{t+1}) = \frac{1}{N} \sum_i \delta_{v_{i,t}, v_{i,t+1}}$  measures the similarity between adjacent sequences.

The probability of the path is then defined as the Boltzmann distribution

$$\mathcal{P}[\mathcal{V}] = \frac{1}{Z_{\text{path}}} e^{-\beta N \mathcal{E}(\mathcal{V})}, \quad (2)$$

where  $\beta$  is an inverse temperature and  $Z_{\text{path}}$  ensures normalization. This distribution promotes paths where intermediate configurations have low energies  $E_{\text{model}}$  and are not far away from each other in order to guarantee smoothness in the interpolation. A key role is played by the potential  $\Phi$ , which controls the elastic properties of the path. In [8], we considered two choices for  $\Phi$ , corresponding to distinct scenarios for the mutational dynamics. The first one, denoted by Cont, makes sure that any two contiguous configurations along the path,  $\mathbf{v}_t$  and  $\mathbf{v}_{t+1}$ , differ by a bounded (and small compared to  $N$ ) number of sites. The second choice for  $\Phi$  is inspired by Kimura's theory of neutral evolution (when the fitness landscape is not present) [14] and hereafter called Evo. It enforces a constant mutation rate for each variable (see below).

In the Cont scenario, we aim to build paths that continuously interpolate between the two target configurations as  $T$  grows. Hence, we choose  $\Phi$  in order to avoid small overlaps  $q$  between adjacent sequences, which would signal

large jumps along the path. In practice, we set

$$\Phi_{\text{Cont}}(q) = \frac{1}{T^2|q - q_c|} = \frac{1}{T^2|q - 1 + \frac{\gamma}{T}|}, \quad (3)$$

where the  $1/T^2$  scaling in the potential guarantees the existence of continuous solution in the large- $T$  limit as shown in Sec. III B; other choices of potentials with hard-wall constraints give similar results. The parameter  $\gamma$  controls the elasticity of the path. Its minimal value is  $D/N$ , where  $D$  is the Hamming distance between the extremities  $\mathbf{v}_{\text{start}}$  and  $\mathbf{v}_{\text{end}}$ . Larger values of  $\gamma$  will authorize more flexible paths.

In the Evo scenario, the potential  $\Phi$  is chosen to emulate neutral evolution with a certain mutation rate  $\mu$ , and is given by

$$\Phi_{\text{Evo}}(q) = (1 - q) \ln \left( 1 + \frac{A}{e^{\mu A/(A-1)} - 1} \right). \quad (4)$$

In this setting, paths can be seen as alternating steps of random mutations (starting from  $\mathbf{v}_{\text{start}}$ ) and of selection, parametrized by, respectively, the mutation rate  $\mu$  and the effective “logarithmic fitness”  $-E_{\text{model}}$ . The transition path is conditioned to end in  $\mathbf{v}_{\text{end}}$ . In standard evolutionary dynamics, paths are not constrained by their final configuration, but only by their initial one. Such paths are anchored at one extremity only. However, if the configuration (genome) of an organism is observed after some evolutionary time, it is legitimate to ask about the distribution of putative paths followed by the organism that interpolate between this “final” and the known initial configurations. As a result of this conditioning, the transitions paths are now anchored at both extremities.

The internal parameters of the Evo and Cont scenarios, respectively,  $\mu$  and  $\gamma$ , control the “elasticity” of the paths. When paths are flexible enough, our model allows for back-mutations, e.g., the same sites can undergo a mutation and its reverse along the path, reaching back to the original state. The presence of back-mutations may be observed in evolving systems with high mutational variability, such as viruses or directed evolution experiments. However, in other cases, such as organisms with large genomes, mutations are rarer, and the dynamical processes are better described by strong-selection-weak-mutation (SSWM) evolutionary models [15,16]. In such contexts, back-mutations become extremely unlikely, imposing constraints on  $\mu$  and  $\gamma$  for the plausibility of our model.

## B. Minimal Hopfield-Potts model for transition paths

We now introduce a minimal setting, where the properties of transition paths can be analytically characterized.

### 1. The Hopfield-Model landscape

We first define the energy landscape for Potts configurations. We consider a Hopfield model for categorical data, hereafter referred to as Hopfield-Potts. There are  $A \geq 3$  states per site (called  $a, b, c$ , and so on). The energy of our minimal Hopfield-Potts (MHP) model reads

$$E_{\text{MHP}}(\mathbf{v}) = -\frac{J}{2N} \sum_{i,j} (w_{1i}(v_i)w_{1j}(v_j) + w_{2i}(v_i)w_{2j}(v_j)), \quad (5)$$

where the two patterns  $\mathbf{w}$  are constructed as follows:

$$\begin{aligned} w_{1i}(v_i) &= \delta_{v_i,a} + \omega \delta_{v_i,c}, \\ w_{2i}(v_i) &= \delta_{v_i,b} + \omega \delta_{v_i,c}, \end{aligned} \quad (6)$$

and  $\omega$  is a positive parameter that controls how much the two patterns overlap. The coupling strength  $J$  is supposed to be large, but its precise value does not affect the qualitative description below. The energy of a configuration  $\mathbf{v}$  is a quadratic function of its two projections along the patterns, denoted as  $m^\mu(\mathbf{v}) = \frac{1}{N} \sum_i w_{i\mu}(v_i)$  ( $\mu = 1, 2$ ):

$$E_{\text{MHP}}(\mathbf{v}) = -\frac{J}{2} N [(m^1(\mathbf{v}))^2 + (m^2(\mathbf{v}))^2]. \quad (7)$$

The MHP model is therefore intrinsically of mean-field nature, and can be easily solved in the large- $N$  limit.

A sketch of the free energy of the MHP model in the  $(m^1, m^2)$  plane is shown in Fig. 2. Depending on the value of  $\omega$  two cases must be distinguished:

(i) For  $\omega < \frac{1}{2}$ , the only minima of the free energy are  $(m^1, m^2) = (m^*, 0)$  and  $(0, m^*)$ , with  $m^* \simeq 1$ . As a consequence, the only configurations with non-negligible probabilities are the two patterns themselves.

(ii) For  $\omega > \frac{1}{2}$ , a new local minimum will appear at  $(m^1, m^2) = (\omega, \omega)$ , which we refer to as a symmetric minimum later. This local minimum becomes global when  $\omega > \frac{1}{\sqrt{2}}$ . Therefore, the energy landscape includes a region far away from the pattern-associated configuration that is energetically favorable.

### 2. Transition paths anchored at both extremities

In this landscape, we will consider paths of configurations anchored at both extremities, i.e., such that  $\mathbf{v}_{\text{start}} = \{a, a, a, \dots, a\}$  and  $\mathbf{v}_{\text{end}} = \{b, b, b, \dots, b\}$ . For the sake of simplicity, we will restrict ourselves to the Cont potential [see Eq. (3)]. As the Hamming distance between the two edges of the path is equal to  $D = N$ , the flexibility parameter  $\gamma$  must be larger than 1.

The properties of the mutational paths associated to this energy landscape can be analytically characterized. The mean-field theory associated to paths is more sophisticated than for single configurations. Explicit expressions can nevertheless be derived for the average projections  $m_t^1, m_t^2$  of intermediate configurations  $\mathbf{v}_t$  and for the average overlap  $q_t$  between successive sequences  $\mathbf{v}_t, \mathbf{v}_{t+1}$ . Detailed calculations and results are reported in Sec. III. Briefly speaking, we find the following (see Fig. 2):

(i) For  $\omega < \frac{1}{2}$ , the optimal path connecting the starting and ending configurations is direct. Due to the absence of favorable regions in the landscape outside the neighborhoods of the anchors, paths have no incentive to explore the landscape: they directly interpolate between  $\mathbf{v}_{\text{start}}$  and  $\mathbf{v}_{\text{end}}$  to minimize their elastic energy.

(ii) For  $\omega > \frac{1}{2}$ , the symmetric minimum attracts mutational paths and makes them leave the direct space. If paths are sufficiently long and flexible they are deviated by this minimum, and explore the global configuration space.

While the precise locus of the paths is specific to the MHP model, the coincidence of the onset of the transition with the existence of favorable regions in the configuration space is a

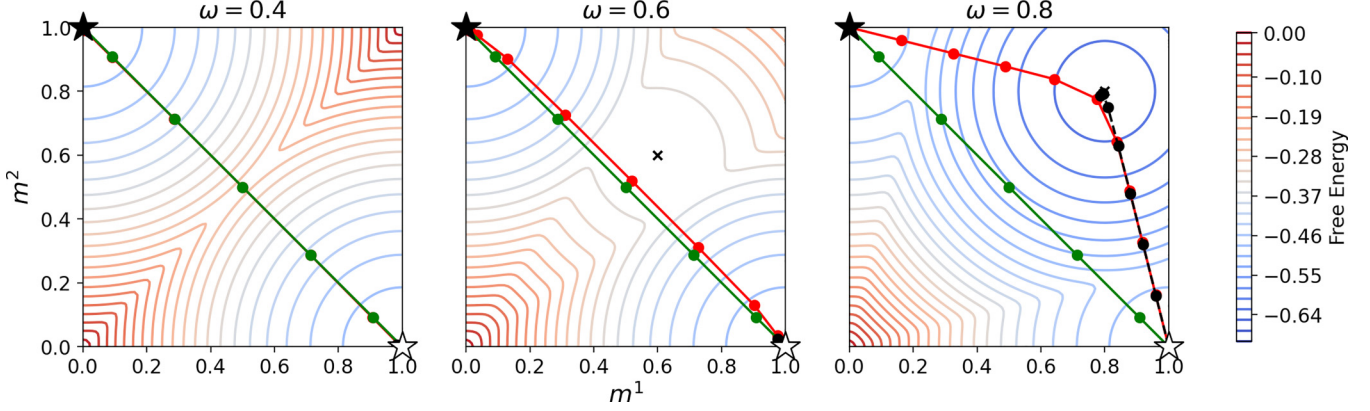


FIG. 2. Sketch of the direct-to-global phase transition in the MHP model. Green (light gray) paths correspond to paths constrained in the direct space, while red (gray) are global (free to explore any configuration). Stars represent the initial and final configurations (see corners of the free-energy landscape). Each plot represents the Hopfield-Potts energy landscape for a fixed value of  $\omega$ , showing the crossover between direct and global transition paths as the symmetric minimum of the free energy becomes more and more attractive, i.e., as the overlap between the patterns  $\omega$  increases and as the length of the path increases. Black dashed lines correspond to paths computed without fixing the end, showing a transition between paths staying close to the initial minimum (i.e., the white star) and paths that jump into the intermediate minimum when this becomes stable for higher values of  $\omega$ . Here, the parameters  $\beta$  and  $\gamma$  appearing in Eqs. (2) and (3) are equal to, respectively, 6 and 3. The length of all paths defined in Eq. (1) is set to  $T = 10$ , but only points that are different from the end points (i.e., white and black stars) are shown.

general phenomenon. The nature of optimal transition paths is therefore intimately related to the structure of the energy landscape.

### 3. Transition paths anchored at one extremity

We consider the case in which  $\mathbf{v}_{\text{end}}$  is not fixed. In this scenario, this final configuration is very likely, for large  $N$ , to lie in one of the global minima of the free energy. If the starting configuration is attached to another minimum, then paths can explore the space in its diversity, see Fig. 2 (right) for an illustration.

Our mean-field theory can be adapted to the case of paths anchored at one extremity, and allows us to estimate the time, i.e., the minimal length necessary for a path to escape some region  $\mathcal{R}$  of the configuration space. In practice, a region is defined as the local minimum of the mean-field free energy containing  $\mathbf{v}_{\text{start}}$ . We define the probability of paths of length  $T$  to stay in  $\mathcal{R}$  through the ratio of the statistical weight, defined in Eq. (2), of all the paths constrained to end in  $\mathcal{R}$  and the weight associated to unconstrained paths, i.e., free to wander in the configuration space. This probability reads

$$P_{\text{stay}}(\mathcal{R}|T) = \frac{\sum_{\mathcal{V}: \mathbf{v}_{\text{end}} \in \mathcal{R}} e^{-N\mathcal{E}(\mathcal{V})}}{\sum_{\mathcal{V}} e^{-N\mathcal{E}(\mathcal{V})}} \underset{N \gg 1}{\sim} \frac{e^{-Nf_{\text{path}}^{\text{const.}}(\mathbf{v}_{\text{start}}, \mathcal{R}|T)}}{e^{-Nf_{\text{path}}^{\text{unconst.}}(\mathbf{v}_{\text{start}}|T)}}, \quad (8)$$

where  $f_{\text{path}}^{\text{const.}}$  and  $f_{\text{path}}^{\text{unconst.}}$  are the free energies associated to, respectively, constrained and unconstrained paths. The calculation of these free energies is reported in Secs. III A and III E. For the MHP model, we observe that, above a certain crossover value of  $\omega$  (depending on  $T$ ), the path is very likely

to escape from the local minimum at  $(m^1, m^2) = (m^*, 0)$  and to end up in  $(\omega, \omega)$  (see Fig. 2).

### C. Transition paths with restricted Boltzmann machines inferred from protein sequence data

Our mean-field approach for transition paths can be applied to more complex Hopfield-Potts energies than Eq. (7), i.e., with more than two patterns, and/or with nonquadratic dependence on the projections  $m^\mu$ . This is the case for the so-called restricted Boltzmann machines, a class of unsupervised architectures that can be trained from data.

#### 1. Restricted Boltzmann machines and landscape inference

Generally speaking, unsupervised machine learning aims to infer an energy landscape through the inference of a probabilistic model  $P_{\text{model}}(\mathbf{v})$  from data configurations,  $\mathbf{v}_b$ ,  $b = 1, \dots, B$ . We consider RBM, a bipartite neural network, in which data configurations  $\mathbf{v}$  are carried by an  $N$ -dimensional layer of visible neurons, and representations  $\mathbf{h}$  of these data are extracted by an  $M$ -dimensional layer of real-valued hidden (latent) units. The two layers interact through the weights  $w$ . The joint probability distribution of visible and hidden configurations is given by, up to a normalization constant,

$$P_{\text{RBM}}(\mathbf{v}, \mathbf{h}) \propto \exp \left( \sum_i g_i(v_i) + \sum_\mu h_\mu I_\mu(\mathbf{v}) - \sum_\mu \mathcal{U}_\mu(h_\mu) \right), \quad (9)$$

where  $I_\mu(\mathbf{v}) = \sum_i w_{i,\mu}(v_i)$  is the input to hidden unit  $\mu$ . The  $g_i$ 's and  $\mathcal{U}_\mu$ 's are local potentials acting on, respectively, visible and hidden units. Note that the weight  $w_{i,\mu}(v_i)$  between visible unit  $i$  and hidden unit  $\mu$  depends on the category  $v_i$  of the visible unit. The hidden potentials  $\mathcal{U}_\mu$  are chosen among

the class of double rectified linear units (DReLU):

$$\mathcal{U}_\mu(h) = \frac{1}{2}\gamma_{\mu,+}h_+^2 + \frac{1}{2}\gamma_{\mu,-}h_-^2 + \theta_{\mu,+}h_+ + \theta_{\mu,-}h_-, \quad (10)$$

where  $h_+ = \max(h, 0)$  and  $h_- = \min(h, 0)$  [10]. All the parameters of the model (weights and local potentials) are learned by maximizing  $\prod_{b=1}^B P_{\text{RBM}}(\mathbf{v}_b)$  using persistent contrastive divergence [17]; regularization over model parameters can also be enforced. Here,  $P_{\text{RBM}}(\mathbf{v}) = \int d\mathbf{h} P_{\text{RBM}}(\mathbf{v}, \mathbf{h})$  is the marginal distribution for configurations. As a result, the RBM energy is

$$E_{\text{RBM}}(\mathbf{v}) = -\ln P_{\text{RBM}}(\mathbf{v}) = -\sum_{i=1}^N g_i(v_i) - \sum_{\mu=1}^M \Gamma_\mu(I_\mu(\mathbf{v})), \quad (11)$$

where  $\Gamma_\mu(I) = \ln \int dh e^{-\mathcal{U}_\mu(h)+hI}$  and irrelevant additive constants have been omitted.

The expression of  $E_{\text{RBM}}$  above shows that RBMs are a generalized class of Hopfield-Potts models. In addition to local potentials acting on the visible units ( $g_i$ ), the energy depends on the configuration  $\mathbf{v}$  through the inputs  $I_\mu(\mathbf{v})$  only. These inputs play the same role as the projections  $m^\mu(\mathbf{v})$  in the Hopfield-Potts framework; both quantities are simply related through  $m^\mu(\mathbf{v}) = \frac{1}{N}I_\mu(\mathbf{v})$ . We stress that the dependence of the energy upon the inputs is generally nonquadratic. Standard Hopfield-Potts models are recovered for  $\mathcal{U}(h) \propto h^2$ , implying  $\Gamma(I) \propto I^2$ . The number of patterns is, in the context of RBM, equal to the number  $M$  of hidden units. In practice,  $M$  is a hyperparameter which is fixed during learning through cross-validation procedures. The Hopfield-Potts nature of RBMs allows us to straightforwardly extend our mean-field approach to these data-driven models (see Sec. IV A).

## 2. Applications to proteins

We apply in Sec. IV our analytical mean-field Hopfield-Potts framework to the RBM energy landscapes inferred from sequence data of real and synthetic protein families. All necessary information about training and sequence data can be found in Secs. IV B and IV C and in [8].

We observe the same kind of direct-to-global transition as the one discussed for the MHP model above. Moreover, we compute, for the WW domain, the entropy of paths as a function of their length for both Cont and Evo potentials, with or without fixed end extremity, as well as the probability of staying in the initial region of the energy landscape. An outcome of this work, of practical relevance to mutagenesis experiments, is the prediction of the sites  $i$  and amino acids  $v_i$ , where mutations outside the direct space are expected to be highly beneficial. These predictions could be used to propose and test new mutations along transition paths, and offer a controlled way to explore the sequence space beyond the amino acids present in the initial and final proteins. In our toy model of lattice proteins such reversed mutations are essential to stabilize the protein when paths join two functionally distinct regions, and show switching from one specificity to another.

## III. MEAN-FIELD THEORY AND DIRECT-TO-GLOBAL TRANSITION FOR THE MINIMAL HOPFIELD-POTTS MODEL

In this section, we describe the mean-field theory treatment of paths in the MHP landscape following [8], solve the corresponding self-consistent equations for the order parameters  $\{m_t^1, m_t^2, q_t\}$  along the path, and then characterize the nature of the transition. For the sake of generality, expressions are written for a generic number  $M$  of patterns  $w_{i\mu}(v)$ , and then applied to the case of the  $M = 2$  patterns in Eq. (6).

### A. Mean-field theory of transition paths

The partition function  $Z_{\text{path}}$  defined in Eq. (2) with the energy function in Eq. (5) can be expressed as an integral over the projections  $\mathbf{m} = \{m_t^\mu\}$  of intermediate configurations on the patterns and over the overlaps  $\mathbf{q} = \{q_t\}$  between successive configurations:

$$Z_{\text{path}}(\beta) = \int d\mathbf{m} d\mathbf{q} \exp \left[ \frac{N\beta}{2} \sum_{\mu,t} (m_t^\mu)^2 - N\beta \sum_t \Phi(q_t) + N \mathcal{S}(\mathbf{m}, \mathbf{q}) \right], \quad (12)$$

where we have defined the entropy as

$$\mathcal{S}(\mathbf{m}, \mathbf{q}) = \frac{1}{N} \ln \sum_{\mathcal{V}} \prod_{\mu,t} \delta \left( \frac{1}{N} \sum_i w_{i\mu}(v_{i,t}) - m_t^\mu \right) \times \prod_t \delta \left( \frac{1}{N} \sum_{i=1}^N \delta_{v_{i,t}, v_{i,t+1}} - q_t \right). \quad (13)$$

Using integral representations of the Dirac  $\delta$ 's, we may express the entropy as an integral over the auxiliary variables  $\hat{\mathbf{m}} = \{\hat{m}_t^\mu\}$  and  $\hat{\mathbf{q}} = \{\hat{q}_t\}$ :

$$\mathcal{S}(\mathbf{m}, \mathbf{q}) = \frac{1}{N} \ln \int \frac{d\hat{\mathbf{m}} d\hat{\mathbf{q}}}{(2\pi/N)^2} \exp[-N\mathbf{m} \cdot \hat{\mathbf{m}} - N\mathbf{q} \cdot \hat{\mathbf{q}}] \times \sum_{\mathcal{V}} \prod_i \exp \left[ \sum_{\mu,t} \hat{m}_t^\mu w_{i\mu}(v_{i,t}) + \sum_t \hat{q}_t \delta_{v_{i,t}, v_{i,t+1}} \right]. \quad (14)$$

In the large- $N$  limit we obtain

$$\mathcal{S}(\mathbf{m}, \mathbf{q}) = \min_{\hat{\mathbf{m}}, \hat{\mathbf{q}}} \left[ -\mathbf{m} \cdot \hat{\mathbf{m}} - \mathbf{q} \cdot \hat{\mathbf{q}} + \frac{1}{N} \sum_i \ln Z_i^{\text{1D}}(\hat{\mathbf{m}}, \hat{\mathbf{q}}) \right], \quad (15)$$

where

$$Z_i^{\text{1D}} = \sum_{\{v_i\}} \exp \left[ \sum_{\mu,t} \hat{m}_t^\mu w_{i\mu}(v_t) + \sum_t \hat{q}_t \delta_{v_t, v_{t+1}} \right] \quad (16)$$

is the partition function of a one-dimensional (1D) Potts model with nearest-neighbor interactions. We note that in the case of paths with both ends fixed the starting and final elements of this sum are fixed, while for paths with free ends

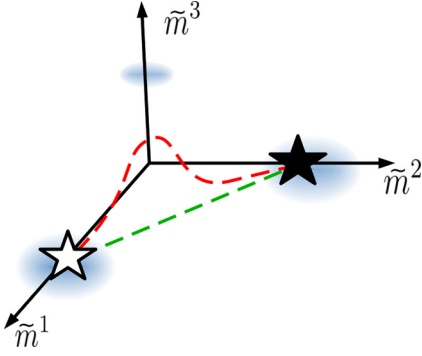


FIG. 3. Transition paths in the three-dimensional (3D) space of projections  $\tilde{\mathbf{m}}$  for the MHP model presented in Sec. II B. Direct solutions [in green (light gray)] linearly interpolate in the input space the two minima of the energy landscape. The global solutions are pushed away from the direct ones by the presence of a third minima emerging from the overlap  $\omega$  between the two patterns of the model defined in Eq. (6).

we also sum over the last element  $v_T$ . At the saddle point, the auxiliary variables fulfill the following set of coupled implicit equations:

$$\begin{aligned} m_t^\mu &= \frac{1}{N} \sum_i \frac{\partial \ln Z_i^{\text{1D}}}{\partial \hat{m}_t^\mu}(\hat{\mathbf{m}}, \hat{\mathbf{q}}), \\ q_t &= \frac{1}{N} \sum_i \frac{\partial \ln Z_i^{\text{1D}}}{\partial \hat{q}_t}(\hat{\mathbf{m}}, \hat{\mathbf{q}}). \end{aligned} \quad (17)$$

We conclude, according to Eq. (12), that the path free energy is given by

$$\begin{aligned} f_{\text{path}}(\beta) &= \lim_{N \rightarrow \infty} -\frac{1}{N\beta} \ln Z_{\text{path}}(\beta) \\ &= \min_{\mathbf{m}, \mathbf{q}} f_{\text{path}}(\beta, \mathbf{m}, \mathbf{q}), \end{aligned} \quad (18)$$

where we have defined the free-energy functional

$$f_{\text{path}}(\beta, \mathbf{m}, \mathbf{q}) = -\frac{1}{2} \sum_{\mu, t} (m_t^\mu)^2 + \sum_t \Phi(q_t) - \frac{1}{\beta} \mathcal{S}(\mathbf{m}, \mathbf{q}). \quad (19)$$

The minimum of  $f_{\text{path}}$  is reached for the roots of

$$\hat{\mathbf{m}} = \beta \mathbf{m}, \quad \hat{\mathbf{q}} = -\beta \Phi'(\mathbf{q}), \quad (20)$$

which, together with Eqs. (17), form a closed set of self-consistent equations for the order parameters.

### B. Free energy for paths

To make the theory easier to interpret in the case of the MHP model, we introduce the three projections (denoted by  $\tilde{m}_t^\mu$ , with  $\mu = 1, 2, 3$ ) along the vectors  $\delta_{v_i, a}$ ,  $\delta_{v_i, b}$ , and  $\omega \delta_{v_i, c}$  (see Fig. 3). While introducing an additional order parameter compared to the number of patterns makes the computation slightly more lengthy, it offers the major advantage to allow for immediate distinction between direct ( $\tilde{m}^3 = 0$ ) and global ( $\tilde{m}^3 > 0$ ) paths. With this choice, we rewrite the free energy

of the path as

$$\begin{aligned} f_{\text{path}}(\beta, \tilde{\mathbf{m}}, \mathbf{q}) &= \sum_t \left( \frac{1}{2} (\tilde{m}_t^1)^2 + \frac{1}{2} (\tilde{m}_t^2)^2 + (\tilde{m}_t^3)^2 \right. \\ &\quad \left. + \tilde{m}_t^3 (\tilde{m}_t^1 + \tilde{m}_t^2) \right) + \sum_t (\Phi(q_t) \\ &\quad - q_t \Phi'(q_t)) - \frac{1}{\beta} \ln Z_{\text{1D}}, \end{aligned} \quad (21)$$

where

$$\begin{aligned} Z_{\text{1D}} &= \sum_{\{v_i\}} \exp \left[ \beta \left( \sum_t \delta_{v_i, a} (\tilde{m}_t^1 + \tilde{m}_t^3) + \delta_{v_i, b} (\tilde{m}_t^2 + \tilde{m}_t^3) \right. \right. \\ &\quad \left. \left. + \omega \delta_{v_i, c} (\tilde{m}_t^1 + \tilde{m}_t^2 + 2\tilde{m}_t^3) - \Phi'(q_t) \delta_{v_i, v_{i+1}} \right) \right]. \end{aligned} \quad (22)$$

As we will see, this model undergoes a first-order phase transition in the regime where  $\beta T$  is large controlled by the overlap between patterns,  $\omega$ , the length of the path,  $T$ , and the stiffness of the Cont potential,  $\gamma$ . We will show the existence of a stretched regime when either  $T$  and  $\omega$  are small or  $\gamma$  is large. In this regime, the minimum of the free energy corresponds to the direct solution from  $\mathbf{v}_{\text{start}}$  to  $\mathbf{v}_{\text{end}}$  that one obtains by restricting the sum in  $Z_{\text{1D}}$  over the first two colors only. We will refer to this solution as  $\#_{\text{dir}}$ . If either  $T$  and  $\omega$  are large or  $\gamma$  is small, a floppy regime arises and  $\#_{\text{dir}}$  is no longer a minimum of the free energy, and the latter is minimized by global paths introducing novel mutations at intermediate steps with nonzero value of  $\tilde{m}_t^3$ .

### C. Minimization of the path free energy in the direct subspace

To understand this phase transition, we first have to find a solution of the direct problem  $\#_{\text{dir}}$ , that is, the set of parameters  $\{\tilde{m}_t^{1, \text{dir}}, \tilde{m}_t^{2, \text{dir}}, q_t^{\text{dir}}\}$ . The direct solution is found by solving the following coupled equations similar to Eq. (17):

$$\tilde{m}_t^{1, \text{dir}} = \frac{1}{Z_{\text{1D}}^{\text{dir}}} \sum_{\{v_i = a, b\}} \delta_{v_i, a} e^{-\beta E_{\text{1D}}(\{v_i\})}, \quad (23)$$

$$\tilde{m}_t^{2, \text{dir}} = \frac{1}{Z_{\text{1D}}^{\text{dir}}} \sum_{\{v_i = a, b\}} \delta_{v_i, b} e^{-\beta E_{\text{1D}}(\{v_i\})}, \quad (24)$$

$$q_t^{\text{dir}} = \frac{1}{Z_{\text{1D}}^{\text{dir}}} \sum_{\{v_i = a, b\}} \delta_{v_i, v_{i+1}} e^{-\beta E_{\text{1D}}(\{v_i\})}, \quad (25)$$

where

$$E_{\text{1D}} = - \sum_t (\tilde{m}_t^{1, \text{dir}} \delta_{v_i, a} + \tilde{m}_t^{2, \text{dir}} \delta_{v_i, b} - \Phi'(q_t^{\text{dir}}) \delta_{v_i, v_{i+1}}). \quad (26)$$

The partition function  $Z_{\text{1D}}^{\text{dir}}$  is the same as in Eq. (22) with the sum running over the states  $a, b$  only, and  $\tilde{m}^3 = 0$ .

We now derive the analytical expression for the mean-field solution when  $T \gg 1$  (remembering that  $N$  was sent to infinity first). Due to exchange symmetry  $a \leftrightarrow b$  we have  $\tilde{m}_t^{2, \text{dir}} = 1 - \tilde{m}_t^{1, \text{dir}}$ . We then look for a direct solution of the form

$$\tilde{m}_t^{1, \text{dir}} = \tilde{m} \left( \tau = \frac{t}{T} \right), \quad (27)$$

where

$$\tilde{m}(\tau) = \begin{cases} 1 & \text{for } \tau < \hat{x} \\ 1 - \frac{\tau - \hat{x}}{1 - 2\hat{x}} + \eta(\tau) & \text{for } \tau \in (\hat{x}, 1 - \hat{x}) \\ 0 & \text{for } \tau > 1 - \hat{x}, \end{cases} \quad (28)$$

where  $\hat{x}$  depends on  $T$  and the function  $\eta(\tau)$  vanishes at large  $T$ . We will show below that  $\eta$  is of the order of  $1/\sqrt{T}$ .

As the number of mutations at each step  $t$  is equivalent to the difference in the projection  $\tilde{m}^{1,\text{dir}}$  between steps  $t$  and  $t + 1$ , we write

$$\begin{aligned} q_t^{\text{dir}} &= 1 - \frac{\text{no. mutations}}{N} = 1 + \tilde{m}_{t+1}^{1,\text{dir}} - \tilde{m}_t^{1,\text{dir}} \\ &= 1 + \frac{1}{T} \partial_\tau \tilde{m}(\tau) \end{aligned} \quad (29)$$

to dominant order in  $T$ . Hence, the overlap order parameters are fully determined once the projection is, with the explicit expression  $q_t^{\text{dir}} = q(\tau = t/T)$  and

$$q(\tau) = \begin{cases} 1 & \text{for } \tau < \hat{x} \\ 1 + \frac{1}{T} \left( \frac{-1}{1-2\hat{x}} + \eta'(\tau) \right) & \text{for } \tau \in (\hat{x}, 1 - \hat{x}) \\ 1 & \text{for } \tau > 1 - \hat{x}. \end{cases} \quad (30)$$

Our goal is to inject the above *Ansätze* into Eq. (25) and determine the function  $\eta$  and the value of  $\hat{x}$  that solve the equation at zero order in  $T$ . First, we expect the effective coupling  $-\Phi'(q(\tau))$  between neighboring  $v_t, v_{t+1}$  in the energy  $E_{1D}$  to scale linearly with the size of the system  $T$ . The reason is that, given a configuration  $\{v_t\}$  appearing in the sum of  $Z_{1D}^{\text{dir}}$ , every couple of adjacent sites  $v_t$  and  $v_{t+1}$  occupying different states, i.e., for every mutation along the path, would produce an energetic penalty  $-\Phi'(q_t)\delta_{v_t, v_{t+1}}$  of the order of  $T$ . The partition function will thus be dominated by the configurations  $v_t = a$  for  $\tau < \hat{x}$  and  $v_t = b$  for  $\tau > 1 - \hat{x}$ , that is, by configurations with a single mutation along the path.

Computing the derivative of the Cont potential, we obtain  $-\Phi'(q(\tau)) = |\gamma - 1/(1 - 2\hat{x}) + \eta'(\tau)|^{-2}$ . Therefore, we expect

$$\gamma - \frac{1}{1 - 2\hat{x}} + \eta'(\tau) \equiv \frac{\xi(\tau)}{\sqrt{T}}. \quad (31)$$

The partition function can then be rewritten as

$$\begin{aligned} Z_{1D}^{\text{dir}} &= T \int_0^1 d\tau \exp \left[ \beta T \left( \int_0^\tau dy \tilde{m}(y) \right. \right. \\ &\quad \left. \left. + \int_\tau^1 dy (1 - \tilde{m}(y)) - \frac{1}{\xi(\tau)^2} \right) \right], \end{aligned} \quad (32)$$

where we explicitly integrate over the reduced “time”  $\tau$  at which the  $a \rightarrow b$  mutation occurs. When  $\beta T \gg 1$ , the exponential integral in the partition function should not depend on  $\tau$  as the mutation may take place with uniform probability in the interval  $(\hat{x}, 1 - \hat{x})$ ; hence, the mutations will happen at different times depending on the site  $i$ . Differentiating the term in the factor of  $\beta T$  with respect to  $\tau$ , we obtain the following differential equation for  $\tau \in (\hat{x}, 1 - \hat{x})$ :

$$\tilde{m}(\tau) - (1 - \tilde{m}(\tau)) - \frac{d}{d\tau} \left( \frac{1}{\xi(\tau)^2} \right) = 0, \quad (33)$$

or, equivalently in the large- $T$  limit,

$$1 - 2 \frac{\tau - \hat{x}}{1 - 2\hat{x}} + 2 \frac{\xi'(\tau)}{\xi(\tau)^3} = 0. \quad (34)$$

Solving this differential equation leads to

$$\xi(\tau) = \left[ \frac{1}{\xi(\hat{x})^2} - \frac{\tau^2 - \hat{x}^2 - (\tau - \hat{x})}{(1 - 2\hat{x})} \right]^{-\frac{1}{2}}. \quad (35)$$

In order to ensure the continuity of  $\Phi'(q(\tau))$  in  $\tau = \hat{x}$ , we choose  $\xi(\hat{x}) = \gamma\sqrt{T}$ . Integrating Eq. (31) over  $\tau$ , we obtain

$$\eta(\tau) - \eta(\hat{x}) = \frac{1}{T^{1/2}} \int_{\hat{x}}^\tau \xi(y) dy + \left( \frac{1}{1 - 2\hat{x}} - \gamma \right) (\tau - \hat{x}). \quad (36)$$

Last of all, upon imposing the boundary condition  $\eta(\hat{x}) = \eta(1 - \hat{x}) = 0$ , we also determine  $\hat{x}$  as a function of  $\gamma$  and of  $T$ . In particular, we can expand  $\hat{x}$  for large  $T$  as

$$\hat{x} = \frac{1}{2} - \frac{1}{2\gamma} - \frac{\pi}{2\sqrt{\gamma^3 T}} + o(T^{-\frac{1}{2}}). \quad (37)$$

Consequently,  $\tilde{m}(\tau) = \tilde{m}^\infty(\tau) + \mathcal{O}(T^{-\frac{1}{2}})$  with  $\tilde{m}^\infty(\tau) = 1$  if  $\tau < \hat{x}^\infty = \frac{1}{2}(1 - \frac{1}{\gamma})$ ,  $\tilde{m}^\infty(\tau) = 0$  if  $\tau > 1 - \hat{x}^\infty$ , and

$$\tilde{m}^\infty(\tau) = 1 - \gamma(\tau - \hat{x}^\infty) \quad (38)$$

if  $\hat{x}^\infty \leq \tau \leq 1 - \hat{x}^\infty$ . It is easy to check that Eqs. (23) and (25) are fulfilled at zero order by this solution.

The solution above holds as long as  $\hat{x}$  does not hit the boundary, i.e., provided  $\hat{x} > 0$ . When  $\hat{x} = 0$ , using Eq. (36) and integrating the function  $\xi$ , we find that  $\gamma$  has to satisfy the equation

$$\gamma = 1 + \frac{2}{\sqrt{T}} \arctan \left( \frac{\gamma\sqrt{T}}{2} \right). \quad (39)$$

The root of this equation, which we denote by  $\gamma^*(T)$ , is plotted in Fig. 4. We may now conclude the following:

(i) If  $\gamma < \gamma^*(T)$  we have  $\hat{x} = 0$ : the projection  $\tilde{m}(\tau)$  is smaller than 1 as soon as  $\tau > 0$  (see inset in Fig. 4). For such small  $\gamma$  the paths are not flexible enough and the full “time”  $T$  at their disposal is needed to join the anchoring edges. We call this regime *overstretched*. Notice that the boundary conditions  $\eta(\hat{x} = 0) = 0$  in Eq. (36) can be satisfied by fixing the initial value of the function  $\xi$ , i.e.,  $\xi(0)$ . In particular, we find

$$\xi(\hat{x} = 0) = 2 \tan \left( \frac{\sqrt{T}(\gamma - 1)}{2} \right). \quad (40)$$

(ii) If  $\gamma > \gamma^*(T)$ , we have  $\hat{x} > 0$ . The available number of intermediate sequences along the path,  $T$ , is larger than what is actually needed to join the two edges. A fraction ( $= 2\hat{x}$ ) of these intermediate sequences are mere copies of the initial and final configurations (see inset of Fig. 4). We hereafter call this regime *understretched*. All the analytical results reported in Eqs. (37) and (38) are in excellent agreement with the numerical resolution of the self-consistent equations for the order parameters (see Fig. 5).

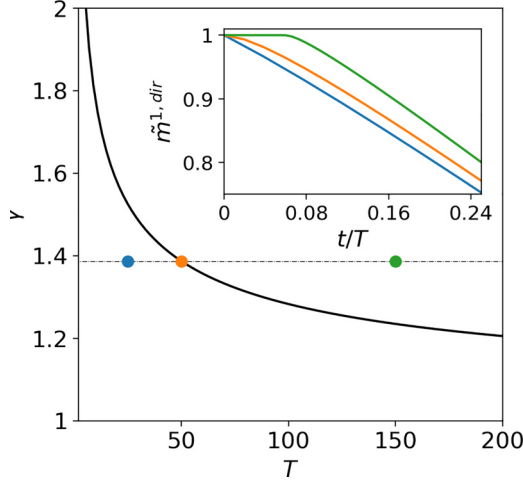


FIG. 4. The “understretched” and “overstretched” subregimes for direct paths. The solid black line represents the root  $\gamma^*(T)$  of Eq. (39). The three colored dots on the black dashed line  $\gamma = 1.387$  correspond to  $T = 25$  (blue), 50 (orange), and 150 (green). The blue and green dots respectively correspond to the overstretched [ $\hat{x} = 0$ ,  $\gamma < \gamma^*(T)$ ] and understretched [ $\hat{x} > 0$ ,  $\gamma > \gamma^*(T)$ ] direct regimes. The orange dot locates the crossover point [ $\gamma = \gamma^*(T)$ ]. Inset: Numerical solutions for  $\tilde{m}_t^{1,\text{dir}}$  with those combinations of parameters are shown in the inset plot for  $t/T \leq 0.24$ . In the simulations,  $\beta = 6$ .

#### D. The direct-to-global phase transition

The solution  $\#_{\text{dir}}$  we have derived above assumes that  $\tilde{m}_3$  vanishes at all time. This assumption is correct as long as the minimum of the free energy  $f_{\text{path}}$  is located in  $\tilde{m}_3 = 0$ . We compute below the first derivative of the free energy along the third projection  $\tilde{m}_3^3$ :

$$\left. \frac{\partial f_{\text{path}}}{\partial \tilde{m}_t^3} \right|_{\#_{\text{dir}}} = 1 - \langle \delta_{v_i,a} + \delta_{v_i,b} + 2\omega \delta_{v_i,c} \rangle_{1D} \Big|_{\#_{\text{dir}}}. \quad (41)$$

By studying the sign of this derivative, we will show the existence of a critical value of  $\omega$  appearing in the patterns of the MHP model [see Eq. (6)]. This critical value, hereafter

denoted by  $\omega_c$ , separates a regime where the direct solution is stable ( $\omega < \omega_c$ ) and a regime where it is not and the true mean-field solution is global ( $\omega > \omega_c$ ).

Two classes of competing configurations must be considered: the direct (dir) ones, which start in  $v^{\text{start}} = a$  and turn into  $v^{\text{end}} = b$  at some time  $\tau \in (\hat{x}, 1 - \hat{x})$ ; the global (glob) ones, which start in  $a$  then change to  $c$  at some time  $\tau \equiv x \in (0, 1/2)$ , then turn into  $b$  when  $\tau = 1 - x$ . We estimate below the energies  $E_{\text{dir}}$  and  $E_{\text{glob}}$  corresponding to the two scenarios. In particular, when  $E_{\text{dir}} < E_{\text{glob}}$ , the direct configurations dominate the average on the right-hand side of Eq. (41), leading to

$$\left. \frac{\partial f_{\text{path}}}{\partial \tilde{m}_t^3} \right|_{\#_{\text{dir}}} = 0 \quad \forall t. \quad (42)$$

Conversely, when  $E_{\text{dir}} > E_{\text{glob}}$ , we will have

$$\left. \frac{\partial f_{\text{path}}}{\partial \tilde{m}_t^3} \right|_{\#_{\text{dir}}} = 1 - 2\omega \text{ for } t \in (x, 1 - x), \text{ 0 otherwise.} \quad (43)$$

Hence, the direct solution will be unstable if, in addition,  $\omega > \frac{1}{2}$ . As we will check explicitly below, this condition is always met when  $E_{\text{dir}} > E_{\text{glob}}$ .

##### 1. Understretched regime

The energy of the direct configurations (for  $T \gg 1$ ) is given by

$$E_{\text{dir}} = -T \left( \hat{x} + \frac{1}{2} \right) + \frac{1}{\gamma^2}, \quad (44)$$

while the global ones have energy

$$E_{\text{glob}}(x) = \begin{cases} -T(2x + \omega(1 - 2x)) + \frac{2}{\gamma^2} & \text{for } x \leq \hat{x} \\ -T \left( 2\hat{x} + 2 \int_{\hat{x}}^x dy \left( 1 - \frac{y - \hat{x}}{1 - 2\hat{x}} \right) \right. \\ \left. + \omega(1 - 2x) - \frac{2}{|\xi(x)|^2} \right) & \text{for } x \in (\hat{x}, 1/2), \end{cases} \quad (45)$$

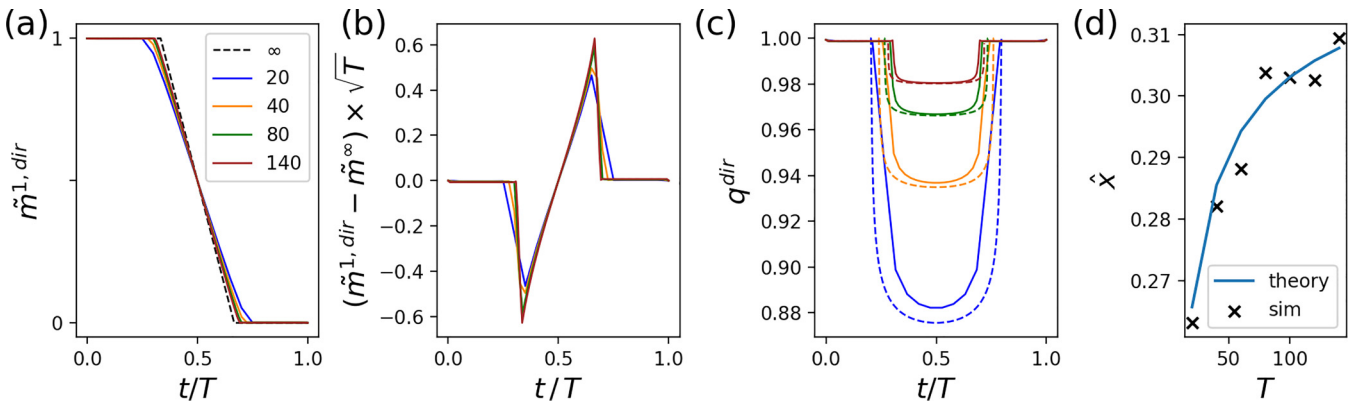


FIG. 5. Mean-field solution of the MHP model in the understretched regime for direct paths. (a) Numerical solutions for  $\tilde{m}_t^{1,\text{dir}}$  for different values of  $T$  (values shown in legend) compared with the limit solution  $\tilde{m}^\infty$  in Eq. (38) for  $T \rightarrow \infty$  (black dashed line). (b) Scaling of the difference between  $\tilde{m}_t^{1,\text{dir}}$  computed numerically and  $\tilde{m}_t^\infty$  for large  $T$ . (c) Numerical solutions for  $q_t^{\text{dir}}$  (solid lines) compared with the respective theoretical estimation (dashed lines) evaluated using  $\hat{x}$  according to Eq. (37). (d) Numerical estimation of  $\hat{x}$  (black crosses: the value corresponds to the moment  $\tilde{m}_t^{1,\text{dir}}$  becomes  $< 1$ ) vs theoretical scaling from Eq. (37) (blue solid line). The parameters of the simulations are  $\beta = 6$ ,  $\gamma = 3$ .



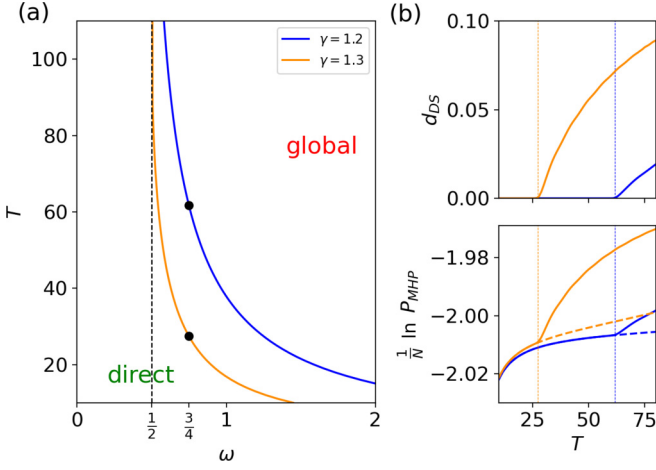


FIG. 6. Crossover between direct and global transition paths in the MHP model. (a) Critical line  $\omega_c(\gamma, T)$  vs  $T$  for two values of  $\gamma$  [see Eq. (50)]. The black dots show the crossovers for  $\omega = \frac{3}{4}$ . (b) Distance  $d_{DS}$  to the direct space (top) and  $(\ln P_{MHP})/N$  averaged over intermediate sequences (bottom; solid line, global; dashed line, direct) vs path length  $T$ ; same parameters as in (a).

which is minimal for  $x = \hat{x}$  when  $\omega \in (1/4, 1)$  and for  $x = 0$  when  $\omega > 1$ . Here the condition  $E_{\text{glob}} < E_{\text{dir}}$  provides the critical value of  $\omega$  for the phase transition,

$$\omega_c^{\text{under}}(\gamma, T) = \frac{1}{2} + \frac{1}{T\gamma^2(1 - 2\hat{x})} \simeq \frac{1}{2} + \frac{1}{T\gamma}, \quad (46)$$

for large  $T$ .

### 2. Overstretched regime

In the overstretched case, the energy of the direct configurations is given by

$$E_{\text{dir}} = -\frac{T}{2} + \frac{T}{\xi(0)^2}, \quad (47)$$

while the global configurations correspond to energy

$$E_{\text{glob}} = -T\omega + \frac{2T}{\xi(0)^2}. \quad (48)$$

Here,  $\xi(0)$  is given by Eq. (40). The condition  $E_{\text{glob}} < E_{\text{dir}}$  leads to a new critical value for  $\omega$ :

$$\omega_c^{\text{over}}(\gamma, T) = \frac{1}{2} + \frac{1}{4 \tan^2(\sqrt{T}(\gamma - 1)/2)}. \quad (49)$$

### 3. Comparison with numerics

Putting together the two regimes studied above, we find that the transition takes place at

$$\omega_c(\gamma, T) = \begin{cases} \omega_c^{\text{over}}(\gamma, T) & \text{for } \gamma < \gamma^*(T) \\ \omega_c^{\text{under}}(\gamma, T) & \text{for } \gamma > \gamma^*(T). \end{cases} \quad (50)$$

The phase diagram in the  $(\omega, T)$  plane is shown in Fig. 6 for different values of the flexibility parameter  $\gamma$ .

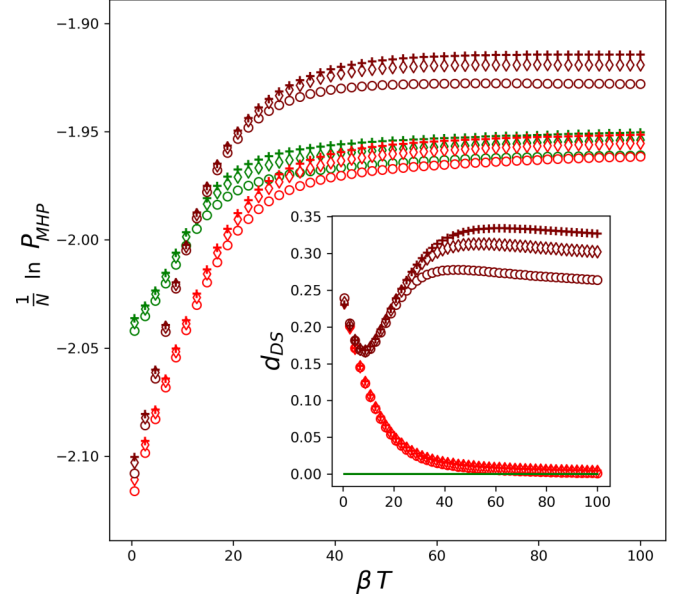


FIG. 7. Average log-likelihood along the paths for the MHP model as a function of  $\beta T$ . Inset plot shows the average distance to direct space. Symbols stand for different  $T$  (circles for  $T = 20$ , diamonds for  $T = 30$ , and pluses for  $T = 40$ ). Green (light gray) symbols represent direct solutions (which are of course independent of  $\omega$ ), red (gray or lower) symbols represent global solutions with  $\omega = 0.4$ , and maroon (dark gray or upper) symbols represent global solutions for  $\omega = 0.7$ . Here  $\gamma = 2$  (Cont potential). For high values of  $\beta T$ , we see that the global paths for  $\omega < 0.5$  converge towards the direct ones, while, for  $\omega > 0.5$ , the two classes of paths remain separated, in agreement with the phase transition shown in Fig. 6.

While the transition formally takes place in the limit  $\beta T \rightarrow \infty$ , a crossover is observed for finite  $T$  and  $\beta$ . We show in Fig. 7 the coincidence of the average log-likelihoods of intermediate sequences along direct and global paths at large  $T$  for small  $\omega$ , and the higher quality of global paths for large  $\omega$ . Notice that these results are valid when  $T$  is sent to large values while keeping  $\beta$  fixed. If  $\beta$  is small, e.g., of the order of  $\frac{1}{T}$ , the domination of global paths on direct paths is due to the larger entropy of the former. Figure 7 shows that, for small  $\beta T$ , global paths are indeed of lesser quality (probability) than their direct counterparts, even at high  $\omega$ .

To better distinguish global from direct paths, we introduce the distance

$$d_{DS}(\mathbf{v}) = \frac{1}{N} \sum_i (1 - \delta_{(v_{\text{start}})_i, v_i})(1 - \delta_{(v_{\text{end}})_i, v_i}). \quad (51)$$

By definition,  $d_{DS}$  vanishes if the configuration is within the direct subspace, and is strictly positive otherwise. Its maximal value is 1. We show in Fig. 7 (inset) the behavior of  $d_{DS}$  for two values of the flexibility parameter controlling the Cont potential, below and above the transition point.

### E. Escaping from local minimum: Paths anchored at origin

We have so far considered paths anchored at both extremities. Our mean-field formalism can be extended to the case of paths in which the final configuration is not fixed. In this context the goal is to characterize the most likely behavior of

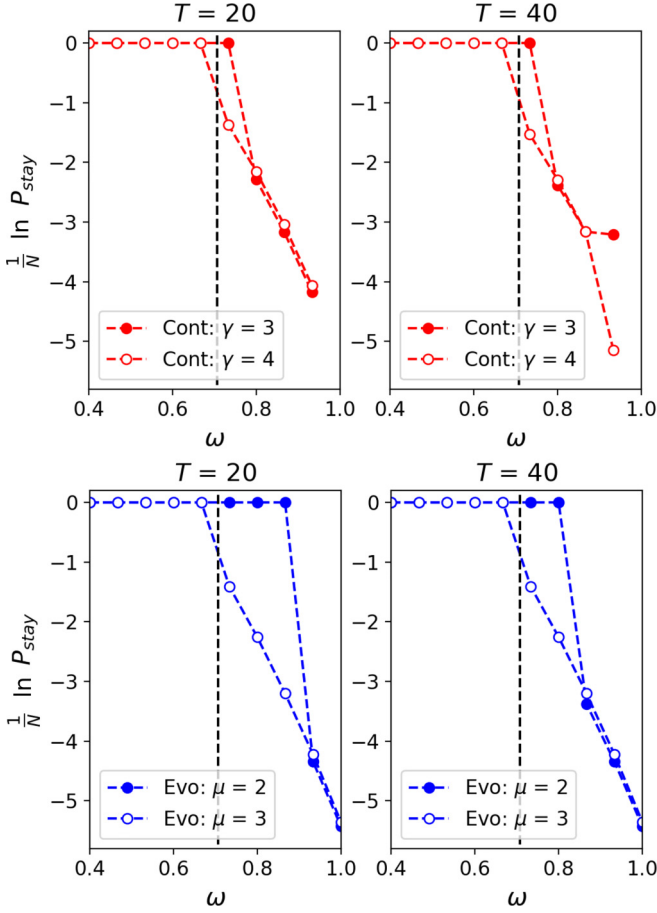


FIG. 8. Probability of stay in the initial local minimum close to the configuration  $\mathbf{m}^{\text{start}} = (1, 0)$  after  $T = 20$  (left) and  $T = 40$  (right) steps. The probabilities are plotted against different values of the overlap  $\omega$ . Cont (top) and Evo (bottom) scenarios are respectively plotted in red and blue. The dark dashed line corresponds to  $\omega = 1/\sqrt{2}$ , when the minimum at  $\mathbf{m} = (\omega, \omega)$  becomes global. Here,  $\beta = 6$ .

a path in the energy landscape under a mutational dynamics encoded in the interaction potential  $\Phi$ .

A natural question in this scenario is to estimate when and in which conditions a configuration escapes from a local minimum to reach a more stable configuration. In the MHP model, we consider paths starting in  $\mathbf{v}_{\text{start}} = \{a, a, a, \dots, a\}$ , and unconstrained at the other extremity. The properties of these paths can be computed through Eq. (18), upon relaxing the condition at the extremity when computing  $Z_i^{\text{1D}}$  using the transfer matrix in Eq. (16).

To estimate the escape probability, we define the region  $\mathcal{R}$  associated to the minimum of the free-energy landscape close to the initial configuration  $\mathbf{m}^{\text{start}} = (1, 0)$ . Then, we evaluate the probability  $P_{\text{stay}}$  of remaining in that region after a certain number of steps,  $T$ , using Eq. (8). The escape probability is computed as  $P_{\text{escape}} = 1 - P_{\text{stay}}$ . In Fig. 8, we show the estimated  $\ln P_{\text{stay}}$  in the Cont and Evo scenarios for different values of  $\omega$  and  $T$ . When  $\omega > 1/\sqrt{2}$ , the local minimum in  $(\omega, \omega)$  depicted in Fig. 2 becomes global, and the path is attracted towards this minimum. For finite  $T$ , higher values

of  $\omega$  are required to overcome the elastic constraint due to  $\Phi$  to remain close to  $\mathbf{v}_{\text{start}}$ .

#### IV. PATHS IN DATA-DRIVEN PROTEIN MODELS

In this section, we aim to expand our mean-field analysis to model the landscapes inferred by the restricted Boltzmann machine from data. Data configurations  $\mathbf{v}$  are sequences of the same protein family given in a multisequence alignment (MSA) of length  $N$  and an alphabet of size  $A = 21$  (20 amino acids plus the gap symbol). The sequences with high probabilities according to the inferred model are predicted to have high fitnesses.

##### A. Mean-field theory

Due to the bipartite structure of their interaction graph, the mean-field theory of the Hopfield-Potts model presented in Sec. III A can be easily extended to the case of RBMs. Two differences are that (1) the effective energy is not a quadratic function of the projections  $m_t^\mu$  when the hidden potentials  $\mathcal{U}(h)$  are not quadratic in  $h$  and (2) the 1D partition function now depends on the potentials  $g_i(v_i)$  acting on the visible units. The expression for the path free energy is now

$$f_{\text{path}}(\mathbf{m}, \mathbf{q}) = -\frac{1}{N} \sum_{\mu, t} \Gamma_\mu(Nm_t^\mu) + \sum_t \Phi(q_t) - \frac{1}{\beta} \mathcal{S}(\mathbf{m}, \mathbf{q}), \quad (52)$$

where  $\Gamma_\mu$  is defined after Eq. (11) and the entropy  $\mathcal{S}$  is given by Eq. (15) with

$$Z_i^{\text{1D}} = \sum_{\{v_t\}} \exp \left( \beta \sum_t g_i(v_t) + \sum_{t, \mu} \hat{m}_t^\mu w_{i\mu}(v_t) + \sum_t \hat{q}_t \delta_{v_t, v_{t+1}} \right). \quad (53)$$

$Z_i^{\text{1D}}$  can be efficiently estimated through products of  $(A \times A)$ -dimensional transfer matrices, where  $A$  is the number of Potts states. For global paths,  $A = 21$ , while  $A = 2$  for direct paths. This mean-field theory is exact when  $N \rightarrow \infty$  [18] and the numbers of hidden units,  $M$ , and of steps,  $T$ , remain finite, but it is already an accurate approximation for some finite- $N$  cases, as will be shown below.

Once the mean-field solution has been determined through minimization of  $f_{\text{path}}$  we can compute any observable, such as the average frequencies of amino acids on site  $i$  at intermediate step  $t$  on the path:

$$\langle \delta_{v_t, a} \rangle = \frac{\partial f_{\text{path}}}{\partial (\beta g_{i,t}(a))} = \sum_{\{v_{t'}\}} \frac{\delta_{v_t, a}}{Z_i} \exp \left( \beta \sum_{t'} g_{i,t'}(v_{t'}) + \sum_{t', \mu} \hat{m}_{t'}^\mu w_{i\mu}(v_{t'}) + \sum_{t'} \hat{q}_{t'} \delta_{v_{t'}, v_{t'+1}} \right), \quad (54)$$

where  $\hat{m}_t^\mu = \beta \Gamma'_\mu(Nm_t^\mu)$  and  $\hat{q}_t = -\beta \Phi'(q_t)$  [see Eq. (20)].

### B. Application to sequence data from lattice-protein models

We start by considering the toy model of lattice proteins (LPs) [13,19]. The model considers sequences of  $N = 27$  amino acids that may fold in one out of  $\sim 10^5$  possible three-dimensional conformations, defined by all possible self-avoiding walks going through the nodes of the  $3 \times 3 \times 3$  cubic lattice.

Given a structural conformation  $\mathbb{S}$ , the probability of a sequence  $\mathbf{v}$  to fold into that structure is given by the interaction energies between amino acids in contact in the structure (occupying neighboring nodes on the lattice). In particular, the total energy of sequence  $\mathbf{v}$  with respect to structure  $\mathbb{S}$  is given by

$$\mathcal{E}_{\text{LP}}(\mathbf{v}|\mathbb{S}) = \sum_{i < j} c_{ij}^{\mathbb{S}} E_{\text{MJ}}(v_i, v_j), \quad (55)$$

where  $c^{\mathbb{S}}$  is the contact map ( $c_{ij}^{\mathbb{S}} = 1$  if sites are in contact and 0 otherwise), while the pairwise energy  $E_{\text{MJ}}(v_i, v_j)$  represents the amino-acid physicochemical interactions given by the Miyazawa-Jernigan knowledge-based potential [20]. The probability to fold into a specific structure is written as

$$p_{\text{nat}}(\mathbb{S}|\mathbf{v}) = \frac{e^{-\mathcal{E}_{\text{LP}}(\mathbf{v}|\mathbb{S})}}{\sum_{\mathbb{S}'} e^{-\mathcal{E}_{\text{LP}}(\mathbf{v}|\mathbb{S}')}}, \quad (56)$$

where the sum runs over the entire set of folds on the cubic lattice. The function  $p_{\text{nat}}$  represents a suitable landscape that maps each sequence to a score measuring the quality of its folding.

To test our mean-field (MF) theory, we first train a RBM over sequences sampled from the probability distribution  $\propto p_{\text{nat}}^{\beta_s}(\cdot|\mathbb{S})$  for a specific structure  $\mathbb{S}$  (with  $\beta_s = 10^3$ ) using Monte Carlo simulation [13]. Then, we numerically compute the MF solutions for paths connecting two faraway target sequences with high  $p_{\text{nat}}$  for both the global and direct cases:

- (1)  $\mathbf{v}_{\text{start}} = \text{DRGIQCLAQMFEKEMRKKRRKCYLECD}$ ,
- (2)  $\mathbf{v}_{\text{end}} = \text{RECCA VCHQRFKDKIDEDYEDAWLKCN}$ .

These two configurations are characterized by a flip of the charge (from negative to positive) of the amino acids in the site 25 (from E to K) and of the neighboring sites [see Fig. 9(d)] to keep an attractive interaction between such sites in order to guarantee the stability of the fold. The trajectories of the inputs  $m_i^\mu$  and of the overlaps  $q_t$  reveal which and when latent factors of RBM enter into play in the transition.

Figure 9(a) shows the trajectories of inputs associated to the weights in Fig. 9(c) (corresponding to hidden variable  $\mu = 4$  and 14). These two hidden variables are strongly activated at sites that are in contact in the tertiary structure of the protein [Fig. 9(d)] and are consequentially relevant for its stability. While the logo of  $w_4$  shows that the interaction between site 25 and its neighbors can be realized through electrostatic forces between charged amino acids,  $w_4$  tells that contacts between sites 5, 6, 11, and 22 can be realized through disulfide bonds between cysteines (C). The dynamics of the projection  $m^{14}$  [Fig. 9(a)] explains how global optimal paths exploit cysteine-cysteine interactions (not present in the initial and final sequences) in order to maintain the structural stability through transient mutations to C-C in sites 5, 6, and

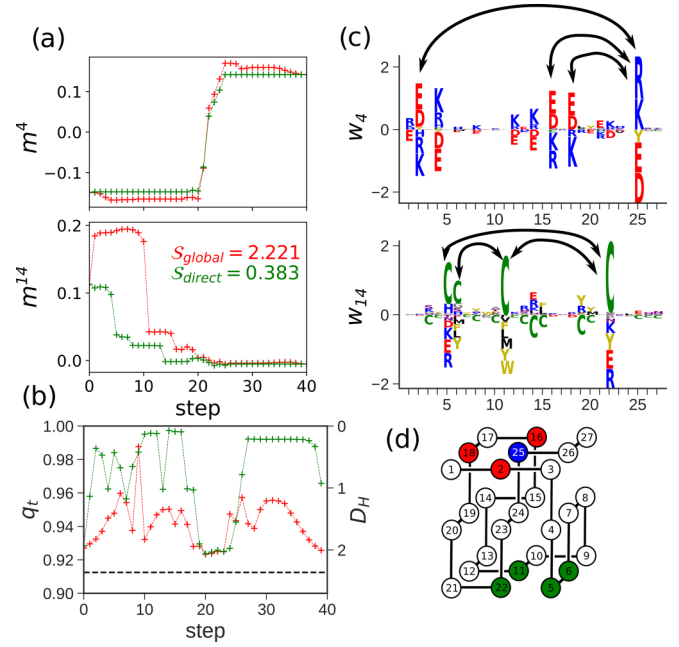


FIG. 9. Mean-field description of mutational paths in lattice proteins with the Cont potential. (a) Values of two relevant inputs vs number  $t$  of mutations along paths of length  $T = 40$ . Red (gray) and green (light gray) lines correspond to, respectively, global and direct paths. Parameters:  $\beta = 3$ ,  $\gamma = 3.5$ . Inset: The entropy for the global and direct solutions. See Sec. IV C 2 for details. (b) Overlap  $q_t$  (left scale) and average number of mutations  $D_H = N(1 - q_t)$  (right scale) between sequences at steps  $t$  and  $t + 1$  vs  $t$ . The black dotted line shows  $q_c$ . (c) Logos of the attached weights  $w_{i,\mu}(v)$ . Positively charged amino acids are in blue (dark gray), negatively charged ones in red (gray). (d) Reference structure for the lattice-protein model.

11 of the protein. These C-C bonds are then lost in the final configuration, as clearly seen by the decrease of the projection  $m^{14}$ . Along global paths, most of the intermediate mutational steps do not abruptly change the order parameters, with the exception of the bump in the overlap  $q$  at step  $\sim 10$ , possibly related to the presence of preparatory mutations for the cysteine-related transition in Fig. 9(a).

Using Eq. (54) we can compute the amino acid frequencies at each site along the path and use this information to estimate the average log-likelihood and  $p_{\text{nat}}$  at each step. To estimate the  $p_{\text{nat}}$  we use these frequencies to build an independent site model that approximates the true marginal distribution of sequences, then we use this model to sample many sequences at a given step  $t$  and compute the average  $p_{\text{nat}}$  from these samples. The results shown in Fig. 10 confirm very good values for the probabilities of intermediate sequences along the path, both for  $p_{\text{nat}}$  [Fig. 10(a)] and for the model  $P_{\text{RBM}}$  [Fig. 10(b)]. We also observe that sequences along the global paths have substantially higher probabilities than along direct paths for the values of  $T$  and  $\gamma$  considered.

### C. Application to the WW domain

We apply the above approach to RBM models learned from sequence data of the WW family extracted from a public database (PFAM id: PF00397) [21,22]. WW is a small protein

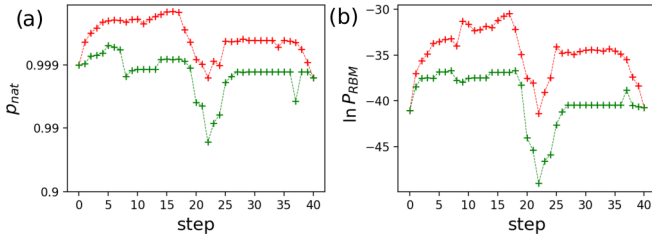


FIG. 10. Average value of  $p_{\text{nat}}$  and log-likelihood along the paths for lattice proteins estimated from the mean-field global (red) and direct (green) solutions shown in Fig. 9.

module with  $\sim 30$ – $40$  amino acids, able to specifically bind to peptidic ligands. In particular, we will study paths interpolating between two proteins known to have different binding activity:

- (1)  $\mathbf{v}_{\text{start}} = \text{LPAGWEMAKTSS-GQRYFLNHIDQTTTQWQDP}$ ,
- (2)  $\mathbf{v}_{\text{end}} = \text{LPKPWIVKISRNRNPYFFNTETHESLWEPP}$ .

$\mathbf{v}_{\text{start}}$  was shown to have strong binding affinity to PPxY ( $x$  = any amino acid) motifs [23] (called class I WW domains), while  $\mathbf{v}_{\text{end}}$  binds to  $p$ TP or  $p$ TS motifs ( $p$  = phosphorylated site) [24] (called class IV WW domains).

### 1. Direct-to-global phase transition

Direct-to-global transitions are observed in mutational paths joining natural WW sequences (see Fig. 11). This figure shows in particular the presence of a crossover, when the path length  $T$  is kept fixed, between direct and global solution at a value of  $\gamma \sim 0.92$  and another jump at  $\gamma \sim 1.3$ , corresponding to the insertion of a novel mutation outside the direct space.

To further study the difference between direct and global solutions at different values of  $\gamma$ , we can compute what and where the first relevant mutations that push the solutions outside the direct space should be considered. Differently stated, given a direct path computed for certain value of length  $T$  and potential stiffness  $\gamma$ , we would like to know what sites will be the first to mutate outside the direct space immediately after we release the constraint on the path to be direct (i.e., we compute the mean-field solution only considering as accessible sites the ones present at the target sequences). To do so, we use Eq. (54) to compute the frequencies of each amino acid  $\langle \delta_{v_{it}, a} \rangle$  in the global space (where the transfer matrix

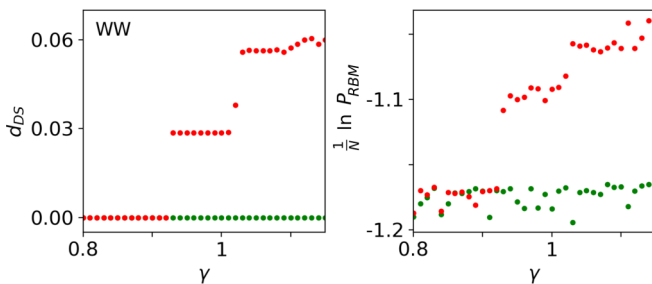


FIG. 11. Direct-to-global phase transition in the WW domain. Mean-field estimates of  $d_{\text{DS}}$  (left) and of  $(\ln P_{\text{RBM}})/N$  (right; red (top), global paths; green (bottom), direct) vs  $\gamma$  for mutational paths of the WW domain of length  $T = 10$ . In all panels,  $\beta = 3$ .

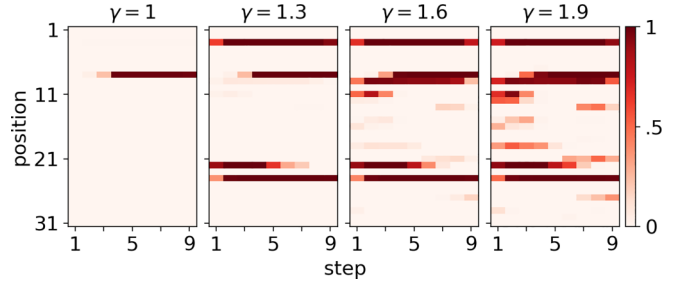


FIG. 12. Probability of nondirect amino acids along direct paths as a function of the step  $t$  ( $x$  axis) and of the sequence site  $i$  ( $y$  axis) for the WW domain. Results are shown for four values of  $\gamma$  (see panels). Parameter:  $\beta = 3$ .

that defines  $Z_i^{\text{1D}}$  is of size  $21 \times 21$ ) around the direct solution. Then, we compute the probability assigned to nondirect amino acids at some point by the direct mean-field solution,  $p_{\text{DS}}^{\text{out}}(i, t)$ , as

$$p_{\text{DS}}^{\text{out}}(i, t) = 1 - \langle \delta_{v_{it}, v_{\text{start}, i}} \rangle_{\#_{\text{dir}}} - \langle \delta_{v_{it}, v_{\text{end}, i}} \rangle_{\#_{\text{dir}}}. \quad (57)$$

Results for different values of  $\gamma$  are shown in Fig. 12. As expected for higher values of  $\gamma$  the interaction potential  $\Phi_{\text{Cont}}$  becomes less stiff and allows the emergence of more mutations escaping the direct space. In the case of  $\gamma = 1$  the Cont potential is stiff enough to allow only one mutation outside the direct space. In particular, this mutation appears in the middle of the path and stays until the very end (before returning to the final state at step 10), showing that the path has to reach a proper region of the sequence space before engaging nondirect mutations. The difference between these global mutations computed on the direct solution and the global solution is shown in Fig. 13, where we used Eq. (54) to compute the frequencies of each amino acid. This approach can be useful to improve mutagenesis experiments by suggesting a minimal number of mutations outside the direct space that can already improve the quality of the intermediate sequences.

Differences between direct and global solutions in the case of the WW domain can be observed in Fig. 14. Here, we plot the values of two relevant inputs along both types of paths. The two weights have been chosen between those that maximize the difference between direct and global solutions. In particular, we see that the projection along the weight  $w_{32}$  for the direct solution remains almost constant compared to the global case. On the other hand, projection along the weight  $w_{41}$  shows a switch in both cases, with global solutions showing a stronger activity.

### 2. Entropy of doubly anchored paths

Our mean-field theory allows us to compute other quantities of interest, such as the number of relevant transition paths. Knowing the entropy of the distribution of paths would be useful for example to estimate how rare the transition between two regions of the sequence space is.

From a practical point of view, despite the care brought in numerically solving Eqs. (17) a small disagreement between the left- and right-hand sides may subsist. As the number of order parameters scales proportionally to  $T$  and  $M$ , these inaccuracies must be taken into account when estimating the

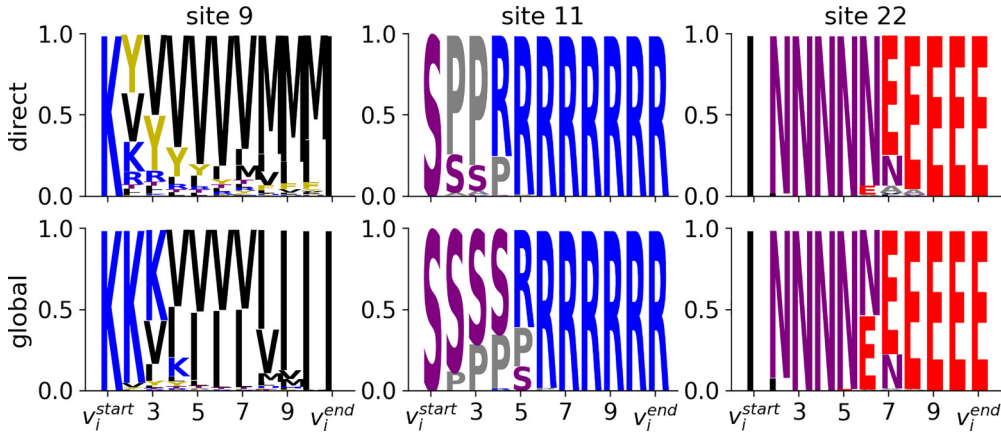


FIG. 13. Logos of the amino-acid frequencies at three arbitrarily chosen sites along a path of length  $T = 10$  joining two WW domains. Top: Logos computed using the MF direct solution; the two amino acids allowed on each site in the direct subspace are the ones corresponding to  $v^{\text{start}}$  and  $v^{\text{end}}$ . The other amino acids are candidates for mutations outside the direct space. Bottom: Logos computed using the MF global solution. Here,  $\gamma = 1.6$  and  $\beta = 3$ .

entropy  $\mathcal{S}_{\text{path}}$ . To compute the latter we therefore estimate  $f_{\text{path}}$  at different inverse temperatures  $\beta$  and use the identity

$$\mathcal{S}_{\text{path}} = -\frac{df_{\text{path}}}{d(1/\beta)}. \quad (58)$$

This procedure gives a more precise estimate of the entropy than directly plugging the values of the order parameters into Eq. (15). In the case of the RBM we obtain

$$\begin{aligned} \mathcal{S}_{\text{path}} = & -\frac{\beta}{N} \sum_{i,t} \langle g_{i,t}(a) \rangle + \frac{1}{N} \sum_i \ln Z_i \\ & - \frac{\beta}{N} \left( \Gamma'(N\mathbf{m}) \frac{\partial}{\partial \mathbf{m}} - \Phi'(\mathbf{q}) \frac{\partial}{\partial \hat{\mathbf{q}}} \right) \sum_i \ln Z_i. \end{aligned} \quad (59)$$

Estimates of  $\mathcal{S}_{\text{path}}$  in the Cont and Evo scenarios are shown in Fig. 15(a). The first important aspect to be noted regards the scaling of  $\mathcal{S}_{\text{path}}$  with the path length  $T$ : while in the Evo scenario the entropy seems to grow linearly with  $T$ , we notice a slower growth with  $T$  in the Cont scenario. This behavior

can be understood in the following toy model. We consider a uniform (flat) landscape  $P_{\text{model}}$ , without constraint on the final sequence. In the Evo scenario, it is easy to show that each time step corresponds, on average, to a constant number of mutations whose value depends on  $\mu$  and on  $A$  only. Hence, the entropy is approximately added to the logarithm of this number at each step, and the total entropy will scale linearly with  $T$ . In the Cont scenario, the number of possible configurations at each step is bounded from above by the

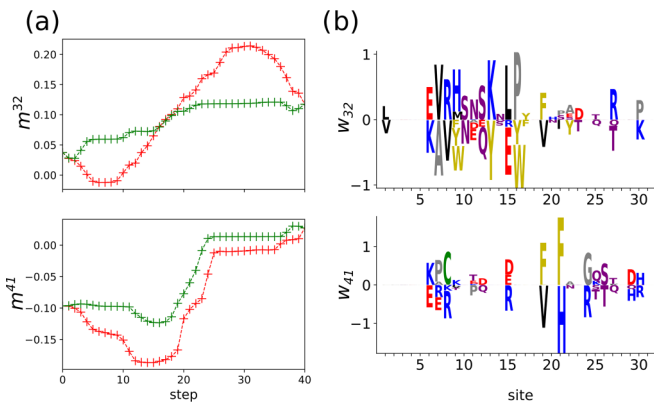


FIG. 14. Direct and global transition path in the WW domain. (a) Values of two relevant inputs. Green (light gray) and red (gray) paths correspond to direct and global solutions, respectively. (b) Logo of the weights associated to the inputs. Simulation parameters:  $\gamma = 1.6$ ,  $T = 40$ ,  $\beta = 3$ .

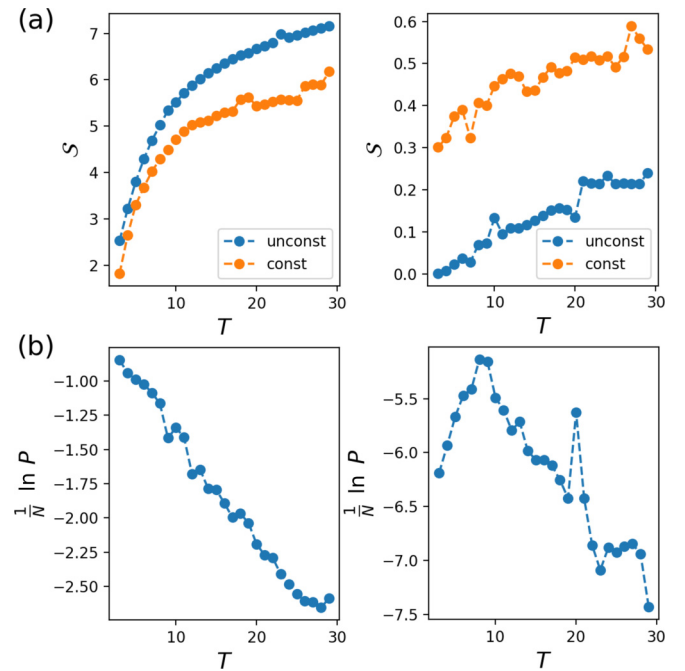


FIG. 15. Entropies and probabilities of transition for the Cont (left) and Evo (right) potentials. (a) Entropy  $\mathcal{S}_{\text{path}}$  of paths as a function of  $T$ . Results are shown for paths joining the two WW domain wild-type sequences (constrained) and paths anchored by the starting sequence and free at the other extremity (unconstrained). (b) Probability of a transition path as a function of  $T$ . Parameters for Evo,  $\mu = 10^{-4}$ ,  $\beta = 1$ ; for Cont,  $\gamma = 3$ ,  $\beta = 1$ .

hard wall in  $\Phi_{\text{Cont}}$ , defined by the overlap  $q_c = 1 - \gamma/T$ . Considering that  $\Phi_{\text{Cont}}(q > q_c) \ll 1$ , each sequence along a path will have on average  $\rho N$  mutations with respect to the previous sequence, where  $\rho = \gamma/T$  is the mutation probability per site. We then estimate the entropy of a binary variable (mutation or no mutation on each site) with probability  $\rho$  as  $-\rho \ln \rho - (1 - \rho) \ln(1 - \rho) \simeq \frac{\gamma}{T} \ln T$  for large  $T$ . Hence, the total entropy (per site) of the paths of length  $T$  as expected to scale as  $\sim \ln T$ .

### 3. Case of paths anchored at the origin

The partition function for paths in Eq. (18) is computed on the ensemble of paths fixed at both ends to be equal to sequence  $\mathbf{v}_{\text{start}}$  and  $\mathbf{v}_{\text{end}}$ . One can easily redo the computation when the last extremity is left free. We show in Fig. 15(a) the entropies of these partially unconstrained paths for the Cont and Evo potentials.

In the Evo scenario the unconstrained solution shows lower entropy than the constrained one, while it has a higher entropy in the Cont scenario as intuitively expected. This apparently surprising finding can be explained as follows. For the constrained, doubly anchored paths  $\mathbf{v}_{\text{end}}$  has relatively high energy [see Fig. 10(b)], and many paths connect this last sequence to  $\mathbf{v}_{\text{start}}$ . Conversely, in the unconstrained case, paths are attracted to a lower free-energy minimum, and there are fewer paths connecting the initial configuration to this final region. The presence of a hard wall in the Cont scenario forbids both solutions to remain in the same configurations for long times and to then jump directly to another distant point in sequence space. Hence, Cont solutions will explore many more different configurations, making their entropy higher with respect to their Evo counterparts. Moreover, since the constrained solution in the Cont case has to smoothly interpolate between distant regions in such a way that the energy along the path is optimized, this makes the number of accessible paths lower than in the unconstrained solution.

Our mean-field formalism allows us to compute the probability to go from  $\mathbf{v}_{\text{start}}$  to  $\mathbf{v}_{\text{end}}$  in  $T$  dynamical steps (see [8]). This probability acquires an evolutionary interpretation in the case of the Evo potential. It estimates the probability to join the two sequences in  $T$  steps consisting of mutations at rate  $\mu$  (per step) combined with selection with probability  $P_{\text{model}}$ . We show in Fig. 15(b) the transition probabilities for the Cont and Evo scenarios. The Evo scenario shows an optimal length  $T^*$  for which the probability is maximized, while, in the Cont scenario, the transition probability decreases linearly with  $T$ . This may be explained from the fact that the Evo potential emulates a mutational dynamics in which  $T^*$  plays the role of an evolutionary distance between the two edge sequences. On the contrary the emergence of this optimal  $T^*$  is forbidden in the Cont scenario by the stiffness of  $\Phi_{\text{Cont}}$ , which increases with  $T$ .

Furthermore, the framework above allows us to compute the probability of remaining in the minimum of the free-energy landscape corresponding to the starting sequence towards some region  $\mathcal{R}$  of the sequence space in  $T$  steps. We define  $P_{\text{stay}}(\mathcal{R}|T)$  as in Eq. (8). In Fig. 16, we plot the probability of remaining in the region associated to  $\mathbf{v}_{\text{start}}$  for the WW domain energy landscape in the Evo scenario. For

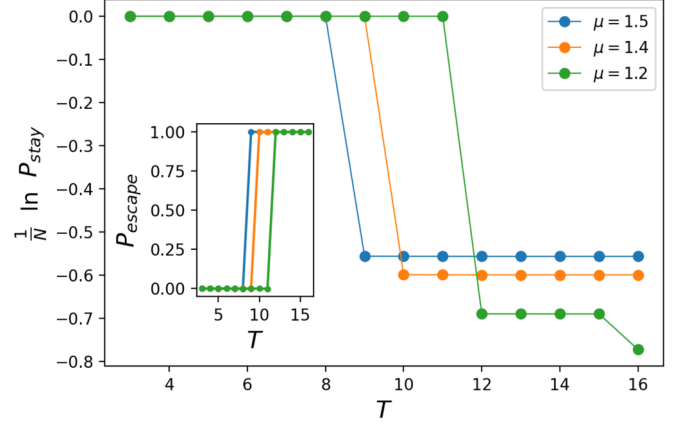


FIG. 16. Probability of remaining in (main panel, logarithmic scale along y axis) and of escaping from (inset) the neighborhood of  $\mathbf{v}_{\text{start}}$  for the WW domain, computed using Eq. (8) in the Evo scenario ( $\beta = 1$ ). Three different values of the mutation rate  $\mu$  are considered.

different values of  $\mu$ , we are able to estimate at which time an evolving configuration is supposed to escape from the minimum. We observe the existence of a trade-off between the time and the probability of sojourn in the starting region depending on the value of  $\mu$ .

## V. CONCLUSION

In the present work, we have focused on the study of transition paths in Potts-like energy landscapes in high dimension  $N$ . These paths can be anchored at the initial and final extremity, or at the origin only. Paths explore the energy landscape under conflicting constraints. First, contiguous configurations along the path should differ little from each other, in a way controlled by an elastic potential. Second, intermediate configurations should have low energies, or, equivalently, high probabilities in the landscape.

We have considered two kinds of elastic potentials. The first one, referred to as Cont, ensures a smooth interpolation between sequences along the path and avoid “jumps” between configurations. The second potential, called Evo, is inspired by evolutionary biology, i.e., it mimics random mutations at a constant rate  $\mu$  [25], while the energy landscape plays the role of the selective pressure driving the evolution. To avoid local maxima in the landscape, successive intermediates along Evo paths may occasionally differ by more than the average number of mutations,  $\mu N$ .

Using mean-field theory, we have computed the typical properties of Evo and Cont paths in two contexts. The first one, called direct, interpolates between two edge sequences, assigning on each site along the path one of the two active states present at the fixed edges. If the Hamming distance between the two extremities of the path is  $D$ , there are  $2^D$  distinct direct intermediate sequences. The second one, called global, may introduce novel mutations along the path compared to the target sequences, allowing for a deeper exploration of the energy landscape. While global paths can find better (i.e., with lower energy) intermediate sequences, they are associated to higher elastic potential energy due to the fact that global paths are in general longer (in terms of total number of mutations)

than direct paths. Whether the subspace of direct paths is statistically dominant in the set of all possible global paths depends on their length and on their flexibility, controlled by the elastic potential.

In the Cont case, we have unveiled the existence of a direct-to-global phase diagram controlled by the stiffness of the interacting potential and the total number of steps of the path, together with the inner structure of the energy landscape. We have analytically described this phase diagram for the so-called Hopfield-Potts model, with only two interaction patterns with projections outside the direct subspace of controllable amplitude. We have analytically located the direct-to-global phase transition in a low-temperature, high-length regime as a trade-off between long, flexible paths with low-energy intermediate configurations and short, stiff paths minimizing the number of mutations to go from one sequence to another. In this low-temperature regime, the direct-to-global transition is essentially not affected by the number  $A$  of Potts states (colors). Conversely, in the high-temperature regime, that is, if the fluctuations of the energy are smaller than or comparable to the inverse of the path length, paths tend to be global due to thermal fluctuations and the entropy of the system will depend on the total number of accessible states  $A$  per site.

This direct-to-global phase transition takes place due to the conditioning on the final extremity of the paths. While evolutionary paths are generally not constrained in this way, there exist relevant situations in which conditioning is important. For instance, consider a directed evolution experiment starting from a wild-type sequence (of DNA, RNA, protein). Samples of the pool of sequences are retained at each round of selections or mutations. After several rounds, a sequence is obtained, and one asks for the possible transition paths that led to this outcome from the wild type. This well-posed question can be addressed with the methods proposed in this work, and compared to sequences sampled at intermediate rounds. In addition, irrespective of conditioning at the end of the path, we have shown that the direct-to-global transition is intimately related to the presence of an attractive region in the energy-fitness landscape (Fig. 3).

From a statistical mechanics point of view, the mean-field approach followed here computes transition paths for a given

realization of the quenched disorder. This is made possible by the fact that, formally, the number  $M$  of patterns in the Hopfield-Potts model (or of hidden units in the RBM) is finite as  $N \rightarrow \infty$ . We plan in future to extend our approach with  $M$  scaling linearly with  $N$ . A possible application, in the case of RBM, would be the so-called compositional phase of [26], where each data configuration activates a finite number of hidden units. In particular, in this scenario we aim to describe the free energy of the system as only a finite number of patterns are active, while the other acts as a white noise.

Last of all, we have tested our method for computing the transition path onto data-driven models of natural proteins, extending the previous work [8] by showing how we could compute different quantities of interest, such as the entropy, i.e., the number of relevant transition paths, the transition probability between two sequences, and the escape probability from confined regions of the sequence space. Future work is definitely needed to improve our approach, e.g., by considering finite- $N$  fluctuations around the mean-field theory solution. From a biological point of view, understanding the shape and the connectivity of the protein fitness landscape and its entropy is of fundamental importance in the field of natural evolutionary processes and also for directed evolution experiments. The motivation here is not only theoretical but also practical, e.g., to gain intuition on how many random sequences can evolve a given functionality under selective pressure. As stressed in [8], inferring the optimal path and its optimal length ( $T$ ) with the Evo potential is an extension of the reconstruction of phylogenetic trees and of the optimal evolutionary distance between two ancestral sequences for epistatic fitness landscapes, inferred from data. Finally, better characterizing transition paths could help predict escaping mutations, e.g., allowing a virus to escape from the control of the immune system, and is therefore of primary importance in the development of effective drugs or vaccines.

#### ACKNOWLEDGMENTS

This work was supported by the ANR-19 Decrypted CE30-0021-01 project. E.M. is funded by a International Centre for Fundamental Physics (ICFP) Labex fellowship of the Physics Department at École Normale Supérieure (ENS).

- 
- [1] S. F. Greenbury, A. A. Louis, and S. E. Ahnert, *Nat. Ecol. Evol.* **6**, 1742 (2022).
  - [2] A. Papkou, L. Garcia-Pastor, J. A. Escudero, and A. Wagner, *bioRxiv* 2023.02.27.530293 (2023), doi:<https://doi.org/10.1101/2023.02.27.530293>.
  - [3] D. A. Stariolo and L. F. Cugliandolo, *Europhys. Lett.* **127**, 16002 (2019).
  - [4] V. Ros, G. Biroli, and C. Cammarota, *SciPost Phys.* **10**, 002 (2021).
  - [5] F.-Y. Wu, *Rev. Mod. Phys.* **54**, 235 (1982).
  - [6] F. J. Poelwijk, M. Socolich, and R. Ranganathan, *Nat. Commun.* **10**, 4213 (2019).
  - [7] N. C. Wu, L. Dai, C. A. Olson, J. O. Lloyd-Smith, and R. Sun, *eLife* **5**, e16965 (2016).
  - [8] E. Mauri, S. Cocco, and R. Monasson, *Phys. Rev. Lett.* **130**, 158402 (2023).
  - [9] A. Fischer and C. Igel, in *Iberoamerican Congress on Pattern Recognition* (Springer, Berlin, 2012), pp. 14–36.
  - [10] J. Tubiana, S. Cocco, and R. Monasson, *eLife* **8**, e39397 (2019).
  - [11] W. P. Russ, M. Figliuzzi, C. Stocker, P. Barrat-Charlaix, M. Socolich, P. Kast, D. Hilvert, R. Monasson, S. Cocco, M. Weigt, and R. Ranganathan, *Science* **369**, 440 (2020).
  - [12] A. Hawkins-Hooker, F. Depardieu, S. Baur, G. Couairon, A. Chen, and D. Bikard, *PLoS Comput. Biol.* **17**, e1008736 (2021).

- [13] H. Jacquin, A. Gilson, E. Shakhnovich, S. Cocco, and R. Monasson, [PLoS Comput. Biol.](#) **12**, e1004889 (2016).
- [14] M. Kimura, *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, UK, 1983).
- [15] M. Manhart and A. V. Morozov, Statistical physics of evolutionary trajectories on fitness landscapes, in *First-Passage Phenomena and Their Applications* (World Scientific, Singapore, 2014), Chap. 1, pp. 416–446.
- [16] J. H. Gillespie, [Evolution](#) **38**, 1116 (1984).
- [17] T. Tieleman, Training restricted Boltzmann machines using approximations to the likelihood gradient, in *ICML '08: Proceedings of the 25th International Conference on Machine Learning* (Association for Computing Machinery, New York, 2008), pp. 1064–1071.
- [18] Note that in order for  $\Gamma_\mu$  to be well defined in this limit, we are formally supposing that the local fields on the hidden space are scaling as  $N$ , i.e.,  $\gamma_{\mu,\pm}, \theta_{\mu,\pm} = \mathcal{O}(N)$ . The weights do not scale with  $N$ , i.e.,  $w_{i\mu}, g_i = \mathcal{O}(1)$ .
- [19] K. F. Lau and K. A. Dill, [Macromolecules](#) **22**, 3986 (1989).
- [20] S. Miyazawa and R. L. Jernigan, [J. Mol. Biol.](#) **256**, 623 (1996).
- [21] M. Sudol, [Prog. Biophys. Mol. Biol.](#) **65**, 113 (1996).
- [22] M. Sudol and T. Hunter, [Cell](#) **103**, 1001 (2000).
- [23] X. Espanel and M. Sudol, [J. Biol. Chem.](#) **274**, 17284 (1999).
- [24] W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffe, and R. Ranganathan, [Nature \(London\)](#) **437**, 579 (2005).
- [25] I. Leuthäusser, [J. Stat. Phys.](#) **48**, 343 (1987).
- [26] J. Tubiana and R. Monasson, [Phys. Rev. Lett.](#) **118**, 138301 (2017).