



HAL
open science

Direct vs. global transition paths in Potts-like energy landscapes

Eugenio Mauri, Simona Cocco, Rémi Monasson

► **To cite this version:**

Eugenio Mauri, Simona Cocco, Rémi Monasson. Direct vs. global transition paths in Potts-like energy landscapes. 2023. hal-04307035v1

HAL Id: hal-04307035

<https://hal.science/hal-04307035v1>

Preprint submitted on 5 Apr 2023 (v1), last revised 13 Mar 2024 (v3)

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

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

Direct vs. global transition paths in Potts-like energy landscapes

Eugenio Mauri, Simona Cocco, Rémi Monasson^{1,*}

¹*Laboratory of Physics of the Ecole Normale Supérieure, CNRS UMR 8023 and PSL Research, Sorbonne Université, 24 rue Lhomond, 75231 Paris cedex 05, France*

(Dated: April 5, 2023)

Identifying transition paths between distant regions of an energy landscape is an important goal in statistical and computational physics, with relevant applications in evolutionary biology. We here consider the case of Potts-like landscapes, in which configurations are made of a large number of categorical variables, taking A distinct values. We show that, when $A \geq 3$, a phase transition arises, separating a regime in which transition paths connect directly the two edge configurations, from another regime, where paths can explore the energy landscape more globally to minimize the energy. This phase transition, controlled by the elasticity of the path, is first illustrated and studied in detail on a mathematically tractable Hopfield-Potts toy model. We then show that this direct-to-global phase transition is also found in energy landscapes inferred from protein-sequence data using Restricted Boltzmann machines.

I. INTRODUCTION

Characterizing transition paths in complex, rugged energy landscapes is a major goal 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, until a known homologous protein is reached, in such a way that all the intermediate sequences do not lose their biological activity. Solving this problem would be crucial to better understand the navigability of fitness landscapes [1]. From a statistical mechanics point of view, substantial efforts have been done to characterize how systems dynamically evolve in complex, *e.g.* glassy landscape to escape from meta-stable local minima to reach lower energy equilibrium configurations. In this context, recent works have focused on p -spin-like energy functions with quenched interactions, generally giving rises to very rugged landscapes [2, 3].

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 [4]. Consider a path joining two configurations, hereafter referred to as $\mathbf{v}_{\text{start}}$ and \mathbf{v}_{end} . A priori, intermediate configurations along the path can 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 sub-space with 2^N configurations, which we call direct space in the following. The question we address in the present work may be informally phrased as follows: under which conditions are good transition paths naturally living in 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 [5]. 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). 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. While this landscape is *a priori* unknown a vast use of data-driven models managed, over the past years, to capture the relation between protein sequences and their functionalities. Unsupervised machine-learning approaches such Boltzmann machines or Variational AutoEncoders could be trained from homologous sequence data. score the sequences, hence defining an empirical energy, and were in particular shown to be generative, *i.e.* they could be used to design novel proteins with functionalities comparable to natural proteins [6, 7].

In the present article, we investigate the issue of direct vs. global paths in the mean-field framework we recently introduced [8]. The energy of a path is defined as the sum of the energies of the intermediate configurations along the path, and of elastic contributions measuring the dissimilarities between successive configurations (Fig. 1). In the simplest setting, the energy corresponds to a Hopfield-Potts model with P patterns, defining a rank P pairwise coupling matrix between the v_i ’s. When N is sent to infinity while keeping P finite, this path model can be analytically solved using mean-field theory, with two sets of time-dependent order parameters, where the time t denotes the coordinate along the path: (1) the average magnetizations m_t^μ of configurations along the patterns $\mu = 1, \dots, P$; (2) the overlaps q_t between the configurations \mathbf{v}_t and \mathbf{v}_{t+1} on the path. We show that, depending on the stiffness coefficient of the elastic term acting on q_t two regimes can be encountered. For path under high tension transition paths are likely to remain confined within the direct space. For low tension, path 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-behaviour of the order parameters are analytically unveiled. Importantly, the mean-field formalism can be transferred to restricted Boltzmann machines (RBM) trained from natural protein sequence data [9, 10]. We show that the same direct-to-global phase transition is found in transi-

* eugenio.mauri@phys.ens.fr, remi.monasson@phys.ens.fr

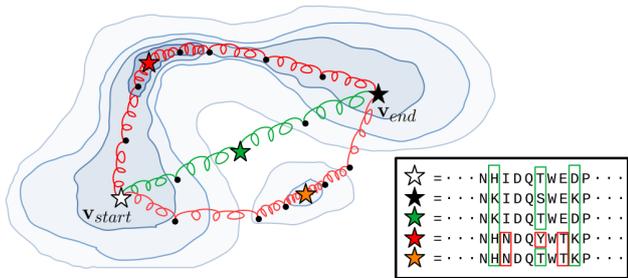


FIG. 1. **Mutational paths between two subfamilies in the sequence landscape associated to a protein family.** Darker blue levels correspond to increasing values of the protein fitness. Paths are either direct (green: each site carries the amino acid present at the same position in the initial or in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

tion paths built from such data-driven models.

This paper is organized as follows. In Section II we recall the transition-path framework introduced in [8], in particular the expression of the probability distribution of paths, and the mean-field free-energy as a function of the order parameters $\{m_t^\mu, q_t\}$ for a generic Hopfield-Potts energy. We also expose how to compute the entropy of transition paths interpolating between the anchoring (initial and final) configurations. In Section III we study in detail in the direct-to-global phase transition in a very simple Hopfield-Potts model with $P = 2$ non-orthogonal patterns. We apply our mean-field approach to inferred energy landscapes in Section IV, from *in silico* lattice-protein models [11] to benchmark our approach, and from natural protein sequence data associated to the WW domain, a short protein domain involved in signalling [10]. Conclusive remarks can be found in Section V.

II. TRANSITIONS PATHS: FRAMEWORK AND MEAN-FIELD THEORY

In this section we recall the general framework for transition paths introduced in [8]. We start by defining the probability distribution of mutational paths interpolating between two sequences. We then consider Hopfield-Potts models, and describe how to implement mean-field treatment of this class of models.

A. Mutational paths over sequence space

We aim to describe paths of sequences interpolating between two fixed targets, $\mathbf{v}_{\text{start}}$ and \mathbf{v}_{end} , in a landscape modeled as a probability distribution over N -dimensional configurations, $P_{\text{model}}(\mathbf{v})$. We define a path of T steps as a set of configurations $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{T-1}\}$, anchored at their extremities $\mathbf{v}_{\text{start}}$ and \mathbf{v}_{end} . The probability of

such a path reads

$$\mathcal{P}[\mathcal{V}|\mathbf{v}_{\text{start}}, \mathbf{v}_{\text{end}}] = \frac{1}{Z_{\text{path}}} \prod_{t=1}^{T-1} P_{\text{model}}(\mathbf{v}_t) \times \quad (1)$$

$$\pi(\mathbf{v}_{\text{start}}, \mathbf{v}_1) \times \prod_{t=1}^{T-2} \pi(\mathbf{v}_t, \mathbf{v}_{t+1}) \times \pi(\mathbf{v}_{T-1}, \mathbf{v}_{\text{end}}),$$

where Z_{path} ensures normalization and $\pi(\mathbf{v}, \mathbf{v}')$ is the 'transition' factor that increases with the similarity between the configurations \mathbf{v}, \mathbf{v}' . In this context, π plays the role of an 'elastic' term pushing the configurations along the path to be closer as shown in Fig. 1. In practice, the transition term $\pi(\mathbf{v}, \mathbf{v}')$ is a function of the overlap between sequences, $q(\mathbf{v}, \mathbf{v}') = \frac{1}{N} \sum_{i=1}^N \delta_{v_i, v'_i}$. We write

$$\pi(\mathbf{v}, \mathbf{v}') = e^{-N\Phi(q(\mathbf{v}, \mathbf{v}'))}, \quad (2)$$

where Φ is an energy potential, decreasing with the overlap between adjacent configurations along the path, with a minimal value for $q = 1$.

The probability of a path can therefore be rewritten as a Boltzmann distribution $\mathcal{P}[\mathcal{V}] \propto e^{-N\mathcal{E}(\mathcal{V})}$, where the path energy \mathcal{E} is given by

$$\mathcal{E}(\mathcal{V}|\mathbf{v}_{\text{start}}, \mathbf{v}_{\text{end}}) = -\frac{1}{N} \sum_{t=1}^{T-1} \log P_{\text{model}}(\mathbf{v}_t) +$$

$$+ \Phi(q(\mathbf{v}_{\text{start}}, \mathbf{v}_1)) + \sum_{t=1}^{T-2} \Phi(q(\mathbf{v}_t, \mathbf{v}_{t+1})) +$$

$$+ \Phi(q(\mathbf{v}_{T-1}, \mathbf{v}_{\text{end}})) \quad (3)$$

This energy landscape promotes paths of configurations with high scores according to the model under consideration, as well as paths where intermediate configurations are not far away from each other in order to guarantee smoothness in the interpolation. A key role is played by the potential Φ , which controls the elastic properties of the path. In [8], we considered two choices for Φ , defining to different scenarios of mutational dynamics along the paths. The first one, denoted by Cont, make sure that any two contiguous configurations along the path, \mathbf{v}_t and \mathbf{v}_{t+1} , differ on a bounded (and small) number of sites. The second choice for Φ is inspired by Kimura's theory of neutral evolution [12] and hereafter called Evo. It enforces a constant dynamical rate for each site on the configurations, 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 Φ in order to avoid small overlaps between adjacent sequences, *i.e.* $q \ll 1$ associated to large jumps along the path. More precisely, we want to enforce that the total number of changes along the path does not scale linearly with T . To do so we impose a hard wall condition to the potential: $\Phi = +\infty$ if $q < q_c = 1 - \gamma/T$. This choice allows the path to explore at most $T \times N(1 - q_c) = \gamma N$ mutations in T steps. Defining the Hamming distance between the target configurations $\mathbf{v}_{\text{start}}$ and \mathbf{v}_{end} as $D = N(1 - q(\mathbf{v}_{\text{start}}, \mathbf{v}_{\text{end}}))$, we see it is sufficient to choose $\gamma > D/N$ to interpolate

between the two target configurations. Larger values of γ will authorize more flexible paths. In practice, we set

$$\Phi_{\text{Cont}}(q) = \frac{1}{T^2 |q - q_c|} = \frac{1}{T^2 |q - 1 + \frac{\gamma}{T}|}, \quad (4)$$

where the $1/T^2$ scaling in the potential guarantees the existence of continuous solution in the large- T limit as shown in Section III; Other choices of potentials with hard-wall constraints give similar results.

In the Evo scenario, the potential Φ is chosen to emulate neutral evolution with a certain mutation rate μ . According to [12], the probability of a site i to be mutated after a time interval Δt under rate μ is (with A denoting the number of possible states on each sites)

$$p_{\neq} = \frac{1}{A} \left(1 - \exp \left(-\frac{A}{A-1} \mu \Delta t \right) \right), \quad (5)$$

while the probability of not mutating is $p_{=} = 1 - (A-1)p_{\neq}$. Setting $\Delta t = 1$, we can now write the transition probability between two configurations \mathbf{v} and \mathbf{v}' having overlap q as

$$\pi(\mathbf{v}, \mathbf{v}') = (p_{=})^{Nq} (p_{\neq})^{N(1-q)}. \quad (6)$$

Taking the logarithm of the previous equality, see Eq. (2), we obtain

$$\Phi_{\text{Evo}}(q) = (1-q) \ln \left(1 + \frac{A}{e^{\mu A/(A-1)} - 1} \right), \quad (7)$$

up to an irrelevant additive constant. In the Evo scenario, the partition function Z_{path} in Eq. (2) is proportional to the transition probability between $\mathbf{v}_{\text{start}}$ and \mathbf{v}_{end} under a dynamics driven by random mutations and selection parameterized by the mutation rate μ and the fitness function $\log P_{\text{model}}$.

B. Mean-field theory of Hopfield-Potts paths

While P_{model} can be in principle any distribution over the configuration space, we hereafter focus on a simple class of Hopfield-Potts models, with energy

$$E(\mathbf{v}) = -\frac{1}{2N} \sum_{\mu=1}^M \sum_{i,j=1}^N w_{i\mu}(v_i) w_{j\mu}(v_j), \quad (8)$$

which includes M patterns $w_{i\mu}(v_i)$. The probability distribution, $P_{\text{model}}(\mathbf{v}) \propto \exp[-E(\mathbf{v})]$, can equivalently be defined as the marginal distribution of a model, whose energy is a function of the visible configuration \mathbf{v} and the hidden real-valued vector \mathbf{h} :

$$E_{\text{full}}(\mathbf{v}, \mathbf{h}) = -\sum_{i,\mu} w_{i\mu}(v_i) h_{\mu} + \frac{N}{2} \sum_{\mu} h_{\mu}^2. \quad (9)$$

The partition function Z_{path} defined in Eq. (2) with the energy function in Eq. 8 can be computed in the $N \rightarrow \infty$ limit using mean-field theory. To this aim we define the average projections of \mathbf{v} along the patterns

w_{μ} at each step of the path as $m_t^{\mu} = \frac{1}{N} \sum_i \langle w_{i\mu}(v_{it}) \rangle$, while the average overlap between adjacent sequences is given by $q_t = \langle q(\mathbf{v}_t, \mathbf{v}_{t+1}) \rangle$. Here, the average is computed over the path distribution in Eq. (2), $\langle \cdot \rangle = \sum_{\mathcal{V}} (\cdot) \exp(-\beta \mathcal{E}(\mathcal{V})) / Z_{\text{path}}(\beta)$, where β is an inverse temperature. Using the projections $\mathbf{m} = \{m_t^{\mu}\}$ and overlaps $\mathbf{q} = \{q_t\}$ as the order parameters of our model, we can write the partition function as

$$Z_{\text{path}}(\beta) = \int d\mathbf{m} d\mathbf{q} \exp \left[\frac{N\beta}{2} \sum_{\mu,t} (m_t^{\mu})^2 - N\beta \sum_t \Phi(q_t) + N \mathcal{S}(\mathbf{m}, \mathbf{q}) \right], \quad (10)$$

where we have defined the entropy as

$$\mathcal{S}(\mathbf{m}, \mathbf{q}) = \frac{1}{N} \log \sum_{\mathcal{V}} \prod_{\mu,t} \delta \left(\frac{1}{N} \sum_i w_{i\mu}(v_{i,t}) - m_t^{\mu} \right) \times \prod_t \delta \left(\frac{1}{N} \sum_{i=1}^N \delta_{v_{i,t}, v_{i,t+1}} - q_t \right). \quad (11)$$

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

$$\mathcal{S}(\mathbf{m}, \mathbf{q}) = \frac{1}{N} \log \int \frac{d\hat{\mathbf{m}}}{2\pi/N} \frac{d\hat{\mathbf{q}}}{2\pi/N} \exp [-N \mathbf{m} \cdot \hat{\mathbf{m}} - N \mathbf{q} \cdot \hat{\mathbf{q}}] \times \sum_{\mathcal{V}} \prod_i \exp \left[\sum_{\mu,t} \hat{m}_t^{\mu} w_{i\mu}(v_{i,t}) + \sum_t \hat{q}_t \delta_{v_{i,t}, v_{i,t+1}} \right]. \quad (12)$$

In the large- N limit we obtain

$$\mathcal{S}(\mathbf{m}, \mathbf{q}) = \min_{\hat{\mathbf{m}}, \hat{\mathbf{q}}} \left[-\mathbf{m} \cdot \hat{\mathbf{m}} - \mathbf{q} \cdot \hat{\mathbf{q}} + \frac{1}{N} \sum_i \log Z_i^{\text{1D}}(\hat{\mathbf{m}}, \hat{\mathbf{q}}) \right], \quad (13)$$

where

$$Z_i^{\text{1D}} = \sum_{\{v_i\}} \exp \left[\sum_{\mu,t} \hat{m}_t^{\mu} w_{i\mu}(v_t) + \sum_t \hat{q}_t \delta_{v_t, v_{t+1}} \right] \quad (14)$$

is the partition function of a 1D-Potts models with nearest-neighbour interactions. The auxiliary variables $\hat{\mathbf{m}}^*$ and $\hat{\mathbf{q}}^*$ therefore fulfill the following set of coupled implicit equations:

$$m_t^{\mu} = \frac{1}{N} \sum_i \frac{\partial \log Z_i^{\text{1D}}}{\partial \hat{m}_t^{\mu}}(\hat{\mathbf{m}}^*, \hat{\mathbf{q}}^*), \quad q_t = \frac{1}{N} \sum_i \frac{\partial \log Z_i^{\text{1D}}}{\partial \hat{q}_t}(\hat{\mathbf{m}}^*, \hat{\mathbf{q}}^*). \quad (15)$$

We conclude, according to Eq. (10), that the path free-

energy is given by

$$\begin{aligned} f_{\text{path}}(\beta) &= \lim_{N \rightarrow \infty} -\frac{1}{N\beta} \log Z_{\text{path}}(\beta) \\ &= \min_{\mathbf{m}, \mathbf{q}} f_{\text{path}}(\beta, \mathbf{m}, \mathbf{q}), \end{aligned} \quad (16)$$

where we have defined the free-energy functional

$$f_{\text{path}}(\beta, \mathbf{m}, \mathbf{q}) = -\frac{1}{2} \sum_{\mu, t} (m_t^\mu)^2 + \sum_t \Phi(q_t) - \frac{1}{\beta} \mathcal{S}(\mathbf{m}, \mathbf{q}). \quad (17)$$

The minimum of f_{path} is reached in $\mathbf{m}^*, \mathbf{q}^*$ solutions of

$$\hat{\mathbf{m}}^* = \beta \mathbf{m}^*, \quad \hat{\mathbf{q}}^* = -\beta \Phi'(\mathbf{q}^*), \quad (18)$$

which, together with Eq.(15), form a closed set of self-consistent equations for the order parameters.

From a practical point of view, despite the care brought in numerically solving these equations a small disagreement between the left and right hand sides of Eq.(15) may subsist. As the number of order parameters scales proportionally to T and M , these inaccuracies must be taken into account when estimating the entropy \mathcal{S} . To compute the latter we therefore estimate f_{path} at different inverse temperatures β and use the identity

$$\mathcal{S} = -\frac{df_{\text{path}}}{d(1/\beta)}. \quad (19)$$

This procedure gives a more precise estimate of the entropy than directly plugging the values of the order parameters in Eq.(13).

III. DIRECT-TO-GLOBAL PHASE TRANSITION IN HOPFIELD-POTTS MODELS

As stated in the introduction, our main purpose is to understand under which conditions transition paths between two configurations are direct, *i.e.* remain confined to the Potts symbols present in the initial and in the final configurations anchoring the path, or global, *i.e.* possibly exploring favorable regions in the configuration space. To measure how far a configuration \mathbf{v} is from the direct subspace, we introduce the distance

$$d_{\text{DS}}(\mathbf{v}) = \frac{1}{N} \sum_i (1 - \delta_{v_i^{\text{start}}, v_i})(1 - \delta_{v_i^{\text{end}}, v_i}). \quad (20)$$

By definition, d_{DS} vanishes if the configuration is within the direct subspace, and is strictly positive otherwise. Its maximal value is 1.

A. The model

We consider a Hopfield-Potts (HP) model with $M = 2$ patterns and $A \geq 3$ states per site (called a, b and c and so on). We will consider the thermodynamic limit

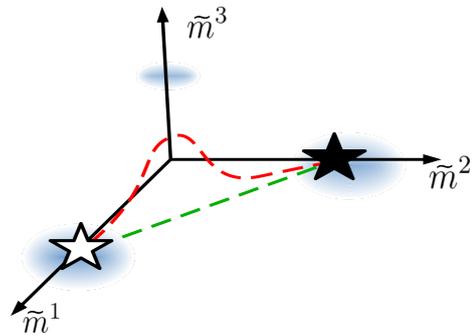


FIG. 2. Sketch of transition paths in HP models presented in Sec. III. Direct solutions (in green) linearly interpolate in the input space the two minima of the energy landscape. The global solutions are pushed away from the direct ones by the presence of a third minima emerging from the overlap ω between the two pattern of the model defined in Eq. 21.

$N \rightarrow \infty$. The two patterns \mathbf{w} are constructed as follows:

$$\begin{aligned} w_{1i}(v_i) &= \delta_{v_i, a} + \omega \delta_{v_i, c}, \\ w_{2i}(v_i) &= \delta_{v_i, b} + \omega \delta_{v_i, c}, \end{aligned} \quad (21)$$

where ω is a positive parameter. The HP model probability reads

$$P_{\text{HP}}(\mathbf{v}) = \frac{1}{Z_{\text{HP}}} e^{\frac{\beta}{2N} \sum_{i,j} w_{1i}(v_i) w_{1j}(v_j) + w_{2i}(v_i) w_{2j}(v_j)}, \quad (22)$$

where Z_{HP} is the normalisation constant. In this landscape, we will focus on paths of configurations starting from $\mathbf{v}^{\text{start}} = \{a, a, a, \dots, a\}$ and ending in $\mathbf{v}^{\text{end}} = \{b, b, b, \dots, b\}$. The parameter ω controls the overlap between the two patterns, along the direction c , which is outside the direct space. For the sake of simplicity, we will restrict ourselves to the Cont potential, see Eq.(4). As the Hamming distance between the two edges of the path is equal to $D = N$, the flexibility parameter γ must be larger than 1.

We study the properties of the transition paths with mean-field theory, see Section IIB. To make the theory easier to interpret, we will not consider the projections m_t^μ along the patterns \mathbf{w}_μ ($\mu = 1, 2$) defined above at different steps t along the paths, but rather the projections (denoted by \tilde{m}_t^μ , with $\mu = 1, 2, 3$, along the vectors $\delta_{v_i, a}, \delta_{v_i, b}, \omega \delta_{v_i, c}$. 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

$$\begin{aligned} f_{\text{path}}(\beta, \tilde{\mathbf{m}}, \mathbf{q}) &= \sum_t \left(\frac{1}{2} (\tilde{m}_t^1)^2 + \frac{1}{2} (\tilde{m}_t^2)^2 + (\tilde{m}_t^3)^2 + \right. \\ &\left. + \tilde{m}_t^3 (\tilde{m}_t^1 + \tilde{m}_t^2) \right) + \sum_t (\Phi(q_t) - q_t \Phi'(q_t)) - \frac{1}{\beta} \log Z_{1\text{D}}, \end{aligned} \quad (23)$$

where

$$Z_{1D} = \sum_{\{v_t\}} \exp \left[\beta \left(\sum_t \delta_{v_t,a} (\tilde{m}_t^1 + \tilde{m}_t^3) + \delta_{v_t,b} (\tilde{m}_t^2 + \tilde{m}_t^3) + \omega \delta_{v_t,c} (\tilde{m}_t^1 + \tilde{m}_t^2 + 2\tilde{m}_t^3) - \Phi'(q_t) \delta_{v_t,v_{t+1}} \right) \right] \quad (24)$$

As we shall see, this model undergoes a first order phase transition in the regime where $\beta \times T$ is large controlled by the overlap between patterns, ω , the length of the path, T , and the stiffness of the Cont potential, γ . We will show the existence of a stretched regime when either T and ω are small or γ is large. In this regime the minimum of the free energy corresponds to the direct solution from $\mathbf{v}^{\text{start}}$ to \mathbf{v}^{end} that one obtains by restricting the sum in Z_{1D} over the first two colors only. We will refer to this solution as $\#_{\text{dir}}$. If either T and ω are large or γ 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 non zero value of \tilde{m}_t^3 .

B. 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 $\#_2$, 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.(15):

$$\tilde{m}_t^{1,\text{dir}} = \frac{1}{Z_{1D}^{\text{dir}}} \sum_{\{v_t=a,b\}} \delta_{v_t,a} e^{-\beta E_{1D}(\{v_t\})}, \quad (25)$$

$$\tilde{m}_t^{2,\text{dir}} = \frac{1}{Z_{1D}^{\text{dir}}} \sum_{\{v_t=a,b\}} \delta_{v_t,b} e^{-\beta E_{1D}(\{v_t\})}, \quad (26)$$

$$q_t^{\text{dir}} = \frac{1}{Z_{1D}^{\text{dir}}} \sum_{\{v_t=a,b\}} \delta_{v_t,v_{t+1}} e^{-\beta E_{1D}(\{v_t\})}. \quad (27)$$

where

$$E_{1D} = - \sum_t \left(\tilde{m}_t^{1,\text{dir}} \delta_{v_t,a} + \tilde{m}_t^{2,\text{dir}} \delta_{v_t,b} - \Phi'(q_t^{\text{dir}}) \delta_{v_t,v_{t+1}} \right). \quad (28)$$

The partition function Z_{1D}^{dir} is the same as in Eq.(24) 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 $T \gg 1$ (remember 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

$$\tilde{m}_t^{1,\text{dir}} = \tilde{m} \left(\tau = \frac{t}{T} \right), \quad (29)$$

where

$$\tilde{m}(\tau) = \begin{cases} 1 & \text{for } \tau < \hat{x} \\ 1 - \frac{\tau - \hat{x}}{1 - 2\hat{x}} + \eta(\tau) & \text{for } \tau \in (\hat{x}, 1 - \hat{x}) \\ 0 & \text{for } \tau > 1 - \hat{x} \end{cases} \quad (30)$$

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

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

$$q_t^{\text{dir}} = 1 - \frac{\text{nb. mutations}}{N} = 1 + \tilde{m}_{t+1}^{1,\text{dir}} - \tilde{m}_t^{1,\text{dir}} = 1 + \frac{1}{T} \partial_\tau \tilde{m}(\tau) \quad (31)$$

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

$$q(\tau) = \begin{cases} 1 & \text{for } \tau < \hat{x} \\ 1 + \frac{1}{T} \left(\frac{-1}{1-2\hat{x}} + \eta'(\tau) \right) & \text{for } \tau \in (\hat{x}, 1 - \hat{x}) \\ 1 & \text{for } \tau > 1 - \hat{x} \end{cases} \quad (32)$$

Our goal is to inject the above Ansätze into Eq. (27) and determine the function η and the value of \hat{x} that solve the equation at the 0-th order in T . First, we expect the effective coupling $-\Phi'(q(\tau))$ between neighbouring 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}^{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

$$\gamma - \frac{1}{1 - 2\hat{x}} + \eta'(\tau) \equiv \frac{\xi(\tau)}{\sqrt{T}}. \quad (33)$$

The partition function can then be rewritten as

$$Z_{1D}^{\text{dir}} = T \int_0^1 d\tau \exp \left[\beta T \left(\int_0^\tau dy \tilde{m}(y) + \int_\tau^1 dy (1 - \tilde{m}(y)) - \frac{1}{\xi(\tau)^2} \right) \right] \quad (34)$$

where we explicitly integrate over the reduced ‘time’ τ at which the $a \rightarrow b$ mutation occurs. When $\beta T \gg 1$, the exponential integral in the partition function should not depend on τ 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 factor of βT with respect to τ we obtain the following differential equation for $\tau \in (\hat{x}, 1 - \hat{x})$:

$$\tilde{m}(\tau) - (1 - \tilde{m}(\tau)) - \frac{d}{d\tau} \left(\frac{1}{\xi(\tau)^2} \right) = 0, \quad (35)$$

or, equivalently in the large T limit,

$$1 - 2 \frac{\tau - \hat{x}}{1 - 2\hat{x}} + 2 \frac{\xi'(\tau)}{\xi(\tau)^3} = 0. \quad (36)$$

Solving this differential equation leads to

$$\xi(\tau) = \left[\frac{1}{\xi(\hat{x})^2} - \frac{\tau^2 - \hat{x}^2 - (\tau - \hat{x})}{(1 - 2\hat{x})} \right]^{-\frac{1}{2}}. \quad (37)$$

In order to ensure the continuity of $\Phi'(q(\tau))$ in $\tau = \hat{x}$, we choose $\xi(\hat{x}) = \gamma\sqrt{T}$. Integrating Eq.(33) over τ we obtain

$$\eta(\tau) - \eta(\hat{x}) = \frac{1}{T^{1/2}} \int_{\hat{x}}^{\tau} \xi(y) dy + \left(\frac{1}{1 - 2\hat{x}} - \gamma \right) (\tau - \hat{x}). \quad (38)$$

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 γ and of T . In particular, we can expand \hat{x} for large T as

$$\hat{x} = \frac{1}{2} - \frac{1}{2\gamma} - \frac{\pi}{2\sqrt{\gamma^3 T}} + o(T^{-\frac{1}{2}}). \quad (39)$$

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

$$\tilde{m}^\infty(\tau) = 1 - \gamma(\tau - \hat{x}^\infty) \quad (40)$$

if $\hat{x}^\infty \leq \tau \leq 1 - \hat{x}^\infty$. It is easy to check that Eqs.(25),(27) are fulfilled at zeroth 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.(38) and integrating function ξ , we find that γ has to satisfy the equation

$$\gamma = 1 + \frac{2}{\sqrt{T}} \arctan \left(\frac{\gamma\sqrt{T}}{2} \right). \quad (41)$$

The root of this equation, which we denote by $\gamma^*(T)$ is plotted in Figure 3. We may now conclude:

- 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 Figure 3. For such small γ 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. (38) can be satisfied by fixing the initial value of the function ξ ,

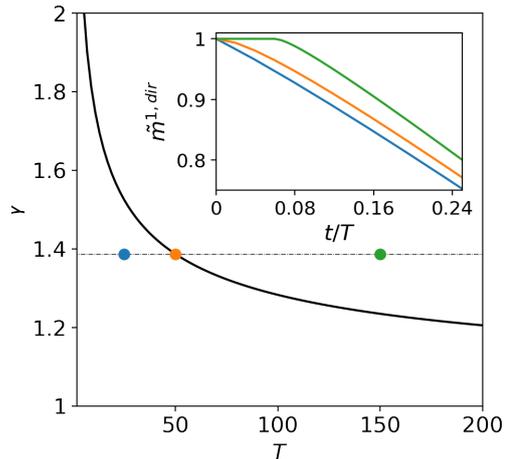


FIG. 3. **The ‘understretched’ and ‘overstretched’ sub-regimes for direct paths.** The solid black line represents the root $\gamma^*(T)$ of Eq. (41). The three colored dots on the black dashed line $\gamma = 1.387$ correspond to $T = 25$ (blue), 50 (orange), 150 (green). The blue and green dots respectively 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 $\tilde{m}_t^{1,dir}$ with those combinations of parameters are shown in the inset plot for $t/T \leq 0.24$. In the simulations $\beta = 6$.

i.e. $\xi(0)$. In particular, we find

$$\xi(\hat{x} = 0) = 2 \tan \left(\frac{\sqrt{T}(\gamma - 1)}{2} \right). \quad (42)$$

- 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 Figure 3. We hereafter call this regime *understretched*. All the analytical results reported in Eqs.39,40 are in excellent agreement with the numerical resolution of the self-consistent equations for the order parameters, see Figure 4.

C. The direct-to-global phase transition

The solution $\#_{dir}$ 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_{path} is located in $\tilde{m}_3 = 0$. We compute below the first derivative of the free energy along the third magnetization \tilde{m}_t^3 :

$$\left. \frac{\partial f_{path}}{\partial \tilde{m}_t^3} \right|_{\#_{dir}} = 1 - \langle \delta_{v_t,a} + \delta_{v_t,b} + 2\omega \delta_{v_t,c} \rangle_{1D} |_{\#_{dir}}. \quad (43)$$

By studying the sign of this derivative we will show the existence of a critical value of ω appearing in the pat-

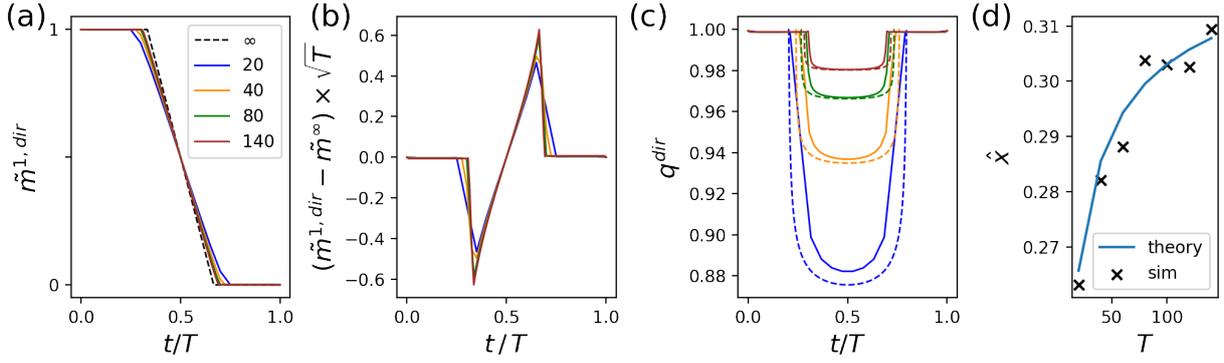


FIG. 4. **Mean-field solution of the HP model in the understretched regime for direct paths.** (a) Numerical solutions for $\tilde{m}_t^{1,dir}$ for different values of T (values showed in legend) compared with the limit solution \tilde{m}_t^∞ in Eq.(40) for $T \rightarrow \infty$ (black dashed line). (b) Scaling of the difference between $\tilde{m}_t^{1,dir}$ computed numerically and \tilde{m}_t^∞ for large T . (c) Numerical solutions for q_t^{dir} (solid lines) compared with the respective theoretical estimation (dashed lines) evaluated using \hat{x} according to Eq.(39). (d) Numerical estimation of \hat{x} (black crosses: the value corresponds to the moment $\tilde{m}_t^{1,dir}$ becomes < 1) vs. theoretical scaling from Eq.(39) (blue line). The parameters of the simulations are $\beta = 6$, $\gamma = 3$.

terms of the HP model, see Eq.(21). This critical value, hereafter denoted by ω_c , separating 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_{dir} and E_{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. (43), leading to

$$\left. \frac{\partial f_{\text{path}}}{\partial \tilde{m}_t^3} \right|_{\#_{\text{dir}}} = 0 \quad \forall t. \quad (44)$$

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

$$\left. \frac{\partial f_{\text{path}}}{\partial \tilde{m}_t^3} \right|_{\#_{\text{dir}}} = 1 - 2\omega \text{ for } t \in (x, 1 - x), \quad 0 \text{ otherwise.} \quad (45)$$

Hence, the direct solution will be unstable if, in addition, $\omega > \frac{1}{2}$. As we shall 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:

$$E_{\text{dir}} = -T \left(\hat{x} + \frac{1}{2} \right) + \frac{1}{\gamma^2}, \quad (46)$$

while the global ones have energy

$$E_{\text{glob}}(x) = \begin{cases} -T(2x + \omega(1 - 2x)) + \frac{2}{\gamma^2} & \text{for } x \leq \hat{x} \\ -T \left(2\hat{x} + 2 \int_{\hat{x}}^x dy \left(1 - \frac{y - \hat{x}}{1 - 2\hat{x}} \right) + \right. \\ \left. + \omega(1 - 2x) - \frac{2}{|\xi(x)|^2} \right) & \text{for } x \in (\hat{x}, 1/2) \end{cases} \quad (47)$$

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 ω for the phase transition:

$$\omega_c^{\text{under}}(\gamma, T) = \frac{1}{2} + \frac{1}{T\gamma^2(1 - 2\hat{x})} \simeq \frac{1}{2} + \frac{1}{T\gamma} \quad (48)$$

for large T .

2. Overstretched regime

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

$$E_{\text{dir}} = -\frac{T}{2} + \frac{T}{\xi(0)^2}, \quad (49)$$

while the global configurations correspond to energy

$$E_{\text{glob}} = -T\omega + \frac{2T}{\xi(0)^2}. \quad (50)$$

Here, $\xi(0)$ is given by Eq.(42). The condition $E_{\text{glob}} < E_{\text{dir}}$ leads to a new critical value for ω :

$$\omega_c^{\text{over}}(\gamma, T) = \frac{1}{2} + \frac{1}{4 \tan^2(\sqrt{T}(\gamma - 1)/2)}. \quad (51)$$

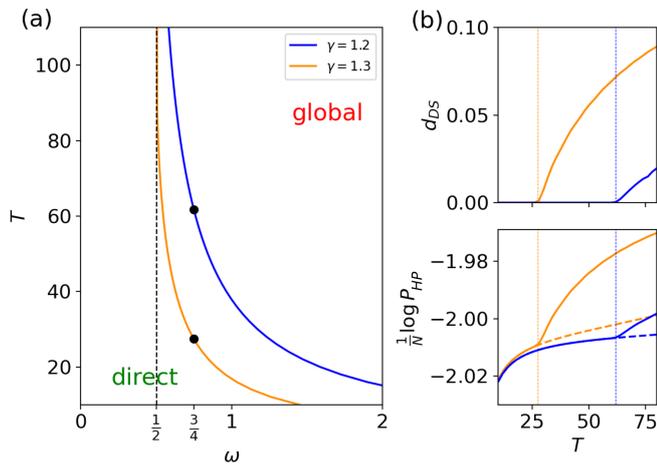


FIG. 5. **Crossover between direct and global transition paths in the HP model.** (a). Critical line $\omega_c(\gamma, T)$ vs. T for two values of γ , see Eq.(52). The black dots show the crossovers for $\omega = \frac{3}{4}$. (b) Distance d_{DS} to the direct space (Top) and $(\log P_{HP})/N$ averaged over intermediate sequences (Bottom; solid line: global, dashed: direct) vs. path length T ; same parameters as in (a).

3. Comparison with numerics

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

$$\omega_c(\gamma, T) = \begin{cases} \omega_c^{\text{over}}(\gamma, T) & \text{for } \gamma < \gamma^*(T) \\ \omega_c^{\text{under}}(\gamma, T) & \text{for } \gamma > \gamma^*(T) \end{cases}. \quad (52)$$

The phase diagram in the (ω, T) plane is shown in Figure 5 for different values of the flexibility parameter γ .

While the transition formally takes place in the limit $\beta \times T \rightarrow \infty$, a cross-over is observed for finite T and β . We show in Figure 6 the coincidence of the average log-likelihoods of intermediate sequences along direct and global paths at large T for small ω , and the higher quality of global paths for large ω . Notice that these results are valid when T is sent to large values while keeping β fixed. If β 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 6 shows that, for small $\beta \times T$, global paths are indeed of lesser quality (probability) than their direct counterparts, even at high ω .

IV. MUTATIONAL PATHS WITH INFERRED MODELS OF PROTEIN SEQUENCES

In this section, we aim to expand our mean-field analysis of transition paths to models of protein sequences, where we use unsupervised machine learning to infer an energy landscape from homologous sequence data.

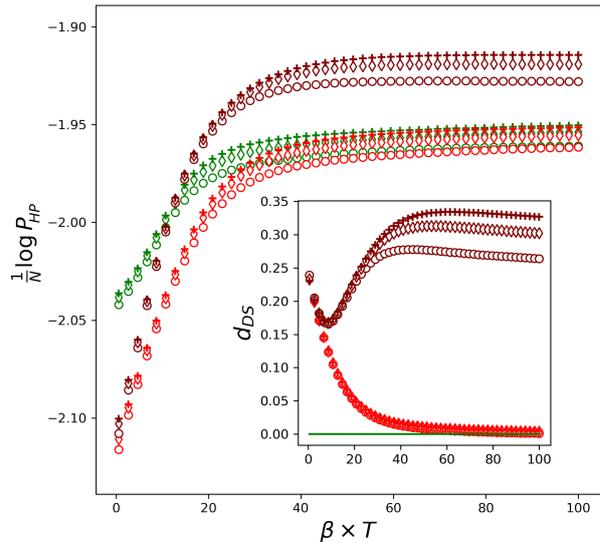


FIG. 6. **Average log-likelihood along the paths for the HP model** as a function of $\beta \times T$. Inset plot shows the average distance to direct space. Symbols stands for different T (circles for $T = 20$, diamonds for $T = 30$ and pluses for $T = 40$). Green symbols represents direct solutions (which are of course independent of ω), Red symbols represents global solutions with $\omega = 0.4$ and maroon symbols represent global solutions for $\omega = 0.7$. Here $\gamma = 2$. For high values of $\beta \times T$ 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 Figure 5.

A. Restricted Boltzmann Machine for landscape inference

Generally speaking, unsupervised machine learning applied to protein sequence data aims to reconstruct the fitness landscape of a protein family as a probabilistic model $P_{\text{model}}(\mathbf{v})$ over the sequences \mathbf{v} . Sequences with high probabilities according to the model are predicted to have high fitnesses. Given a multi-sequence alignment (MSA) of homologous proteins of length N and an alphabet of size $A = 21$ (20 amino acids plus the gap symbol), an unsupervised model adjusts its internal parameters to infer the probability distribution that better represents the data through maximum likelihood approach. We use Restricted Boltzmann Machines (RBM) as unsupervised models of protein families [9, 10]. RBMs are bipartite neural network where the N visible neurons (representing the amino acid at each site) interact with M hidden units through the weights matrix w . The joint probability distribution of visible and hidden configuration (respectively \mathbf{v} and \mathbf{h}) is given by

$$P_{\text{RBM}}(\mathbf{v}, \mathbf{h}) \propto \exp \left(\sum_i g_i(v_i) + \sum_{\mu} h_{\mu} I_{\mu}(\mathbf{v}) - \sum_{\mu} \mathcal{U}_{\mu}(h_{\mu}) \right), \quad (53)$$

where $I_{\mu}(\mathbf{v}) = \sum_i w_{i,\mu}(v_i)$ is the input to hidden unit μ . The g_i 's and \mathcal{U}_{μ} 's are local potentials acting on, respectively, visible and hidden units, and the $w_{i\mu}$'s are the interactions between the two layers. The hidden potentials \mathcal{U}_{μ} are chosen among the class of double Rectified Linear Units (dReLU):

$$\mathcal{U}_{\mu}(h) = \frac{1}{2}\gamma_{\mu,+}h_+^2 + \frac{1}{2}\gamma_{\mu,-}h_-^2 + \theta_{\mu,+}h_+ + \theta_{\mu,-}h_-, \quad (54)$$

where $h_+ = \max(h, 0)$ and $h_- = \min(h, 0)$. All the parameters of the model are learned by maximizing the posterior distribution $\prod_{p=1}^P P_{\text{model}}(\mathbf{v}_p|\mathbf{w})p_{\text{prior}}(\mathbf{w})$ over the sequences \mathbf{v}_p of the MSA of the family using Persistent Contrastive Divergence [13]. Here, the model probability distribution of the sequences is given by the marginal probability $P_{\text{model}}(\mathbf{v}) = \int d\mathbf{h} P_{\text{RBM}}(\mathbf{v}, \mathbf{h})$ while the prior distribution over the hyper-parameters accounts for regularization terms. We use below the same models as in [8], where all necessary information about training and data can be found.

B. Mean-field theory

Due to the bipartite structure of their interaction graph, the mean-field theory developed in Section II B for the Hopfield-Potts model can be easily extended to the case of RBM. Two differences are: (1) the effective energy is not a quadratic function of the magnetizations $m_t^{\mu} = \frac{1}{N} \sum_{i=1}^N \langle w_{i\mu}(v_{it}) \rangle$ when the hidden potentials $\mathcal{U}(h)$ are not quadratic in h ; (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

$$f_{\text{path}}(\mathbf{m}, \mathbf{q}) = - \sum_{\mu,t} \Gamma_{\mu}(m_t^{\mu}) + \sum_t \Phi(q_t) - \frac{1}{\beta} \mathcal{S}(\mathbf{m}, \mathbf{q}), \quad (55)$$

where $\Gamma_{\mu}(m) = \frac{1}{N} \ln \int dh e^{N m h - \mathcal{U}_{\mu}(h)}$ and the entropy \mathcal{S} is given by Eq.(13) with

$$Z_i^{\text{1D}} = \sum_{\{v_t\}} \exp \left(\beta \sum_t g_i(v_t) + \sum_{t,\mu} \hat{m}_t^{\mu} w_{i\mu}(v_t) + \sum_t \hat{q}_t \delta_{v_t, v_{t+1}} \right). \quad (56)$$

Z_i can be efficiently estimated through products of $A \times A$ -dimensional transfer matrices, where A is the number of Potts states. For global paths, $A = 21$ (20 amino acids plus the gap symbol), while $A = 2$ for direct paths.

This mean-field theory is exact when $N \rightarrow \infty$ [14] 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_{path} we can compute any observable, such as the average frequencies of amino acids on site i at intermediate step t on the path:

$$\langle \delta_{v_{i,t}, a} \rangle = \frac{\partial f_{\text{path}}}{\partial (\beta g_{i,t}(a))} = \sum_{\{v_{t'}\}} \frac{\delta_{v_{i,t}, a}}{Z_i} \exp \left(\beta \sum_{t'} g_{i,t'}(v_{t'}) + \sum_{t',\mu} \hat{m}_{t'}^{\mu} w_{i\mu}(v_{t'}) + \sum_{t'} \hat{q}_{t'} \delta_{v_{t'}, v_{t'+1}} \right), \quad (57)$$

where $\hat{m}_t^{\mu} = \beta \Gamma'_{\mu}(m_t^{\mu})$ and $\hat{q}_t = -\beta \Phi'(q_t)$, see Eq.(18).

C. Lattice Proteins

We now apply our mean-field theory to characterise global and direct paths. We start by considering the toy-model of Lattice Proteins (LP) [15]. The model considers sequences of $N = 27$ amino acids that may fold in one out of $\sim 10^5$ possible 3-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 neighbouring nodes on the lattice). In particular, the total energy of sequence \mathbf{v} with respect to structure \mathbb{S} is given by

$$\mathcal{E}_{\text{LP}}(\mathbf{v}|\mathbb{S}) = \sum_{i<j} c_{ij}^{\mathbb{S}} E_{MJ}(v_i, v_j), \quad (58)$$

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 $E_{MJ}(v_i, v_j)$ represents the amino-acid physico-chemical interactions given by the the Miyazawa-Jernigan knowledge-based potential [16]. The probability to fold into a specific structure is written as

$$p_{\text{nat}}(\mathbb{S}|\mathbf{v}) = \frac{e^{-\mathcal{E}_{\text{LP}}(\mathbf{v}|\mathbb{S})}}{\sum_{\mathbb{S}'} e^{-\mathcal{E}_{\text{LP}}(\mathbf{v}|\mathbb{S}')}}, \quad (59)$$

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

To test our mean-field 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 [11]. Then we numerically compute the MF solutions for paths connecting two far away target sequence with high p_{nat} for both the global and direct cases. The trajectories of the inputs m_t^{μ} and of the overlaps q_t reveal which and when latent factors of RBM enter into play throughout the interpolation between the initial and final sequences.

Figure 7(a) shows the trajectories of inputs associated

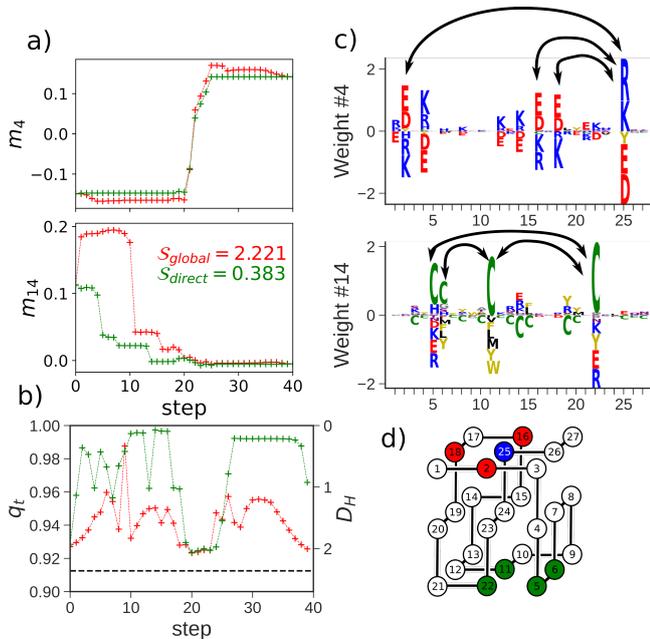


FIG. 7. **Mean-field description of mutational paths** in lattice proteins. (a) Values of two relevant inputs vs. number t of mutations along paths of length $T = 40$. Red and green lines correspond to, respectively, global and direct paths. Parameters: $\beta = 3$, $\gamma = 3.5$. In the inset we show the entropy for the global and direct solutions. (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 dark line shows q_c . (c) Logos of the attached weights $w_{i,\mu}(v)$. (d) Reference structure for the Lattice Protein model.

to the weights in Fig. 7(c) (associated with 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 (Figure 7(d)) and are consequentially relevant for its stability. While the logo of w_4 show that the interaction between sites 2,16,18 and 25 can be realized through electrostatic forces between charged amino acids [17], w_4 tells that contacts between sites 5,6,11 and 22 can be realized through disulfide bonds between Cysteines (C). The dynamics explains how optimal paths exploit Cysteine-Cysteine interactions (not present in the initial and final sequences) in order to maintain the structural stability of the protein when the signs of the charge along the electrostatic chain are flipped (Fig. 7(a)). The exploration of favourable regions in the landscape is made possible by the slightly higher number of mutations between successive sequences in the former case than in the latter, see Fig. 7(b). Along global paths, most of the intermediate mutational steps do not abruptly affect the inputs nor the probability, with the exception of the bump in the overlap q at step ~ 10 , possibly related to the presence of preparatory mutations for the Cys-related transition in Fig. 7(a).

Using Eq.(57) we can compute the amino acids frequencies at each site along the path and use this information to estimate the average log-likelihood and p_{nat} at each step. To estimate the p_{nat} we use this frequencies to build an independent site model that approxi-

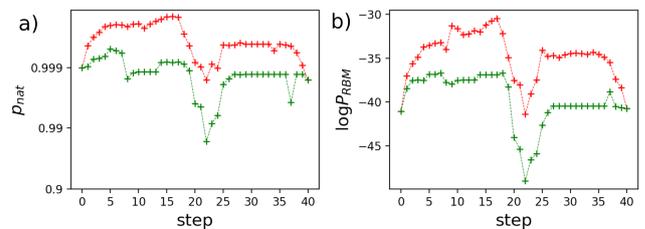


FIG. 8. **Average value of p_{nat} and log-likelihood along the paths** estimated from the mean-field global (red) and direct (green) solutions shown in Figure 7.

mate 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_{nat} from this samples. The results shown in Figure 8 confirm very good values for the probabilities of intermediate sequences along the path, both for p_{nat} (Figure 8(a)) and for the model P_{RBM} (Figure 8(b)). We also observe that sequences along the global paths have substantially higher probabilities than along direct paths for the values of T and γ considered.

D. Application to WW domain

We apply the above approach to RBM models learnt from sequence data of the WW family extracted from public database (PFAM id: PF00397) [18, 19]. WW is a small protein 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:

$$\mathbf{v}_{\text{start}} = \text{LPAGWEMAKTSS-GQRYFLNHIDQTTTQQDP} ,$$

$$\mathbf{v}_{\text{end}} = \text{LPKPWIVKISRNRNPYFFNTETHESLWEP} .$$

$\mathbf{v}_{\text{start}}$ was shown to have strong binding affinity to PPxY (x = any amino acid) motifs [20] (called class I - WW domains), while $\mathbf{v}_{\text{start}}$ binds to pTP or pTS motifs (p=phosphorylated site) [21] (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 Figure 9. This figure shows in particular the presence of a cross-over, 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 γ , 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 γ , we would like to know what sites will be the first to mutate outside the direct space immediately after we release the

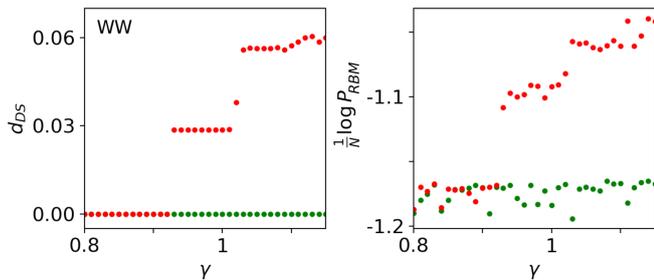


FIG. 9. **Direct-to-global phase transition in WW domain.** Mean-field estimates of d_{DS} (Left) and of $(\log P_{RBM})/N$ (Right; red: global paths, green: direct) vs. γ for mutational paths of the WW domain of length $T = 10$. In all panels $\beta = 3$.

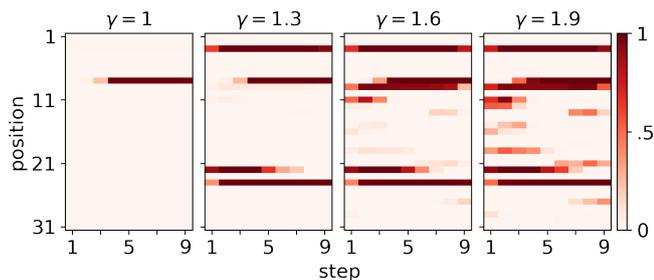


FIG. 10. **Probability of non-direct amino acids along direct paths** as a function of the step t (x -axis) and of the sequence site i (y -axis) for the WW domain. Results are shown for four values of γ , see panels. Parameter: $\beta = 3$.

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. (57) to compute the frequencies of each amino acid $\langle \delta_{v_{it}, a} \rangle$ in the global space (where the transfer matrix that defines Z_i^{1D} is of size 21×21) around the direct solution. Then we compute the probability assigned to non-direct amino acids at some point by the direct mean-field solution, $p_{DS}^{\text{out}}(i, t)$, as:

$$p_{DS}^{\text{out}}(i, t) = 1 - \langle \delta_{v_{it}, v_{\text{start}, i}} \rangle_{\# \text{dir}} - \langle \delta_{v_{it}, v_{\text{end}, i}} \rangle_{\# \text{dir}}. \quad (60)$$

Results for different values of γ are shown in Figure 10. As expected for higher values of γ the interaction potential Φ_{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 non-direct mutations. The difference between these global mutations computed on the direct solution and the global solution is shown in Figure 11, where we used Eq. (57) 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.

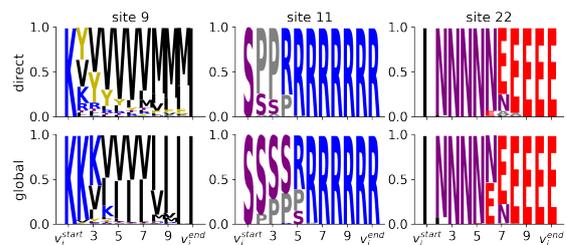


FIG. 11. **Logos of the amino-acid frequencies** at three arbitrarily chosen sites along a path of length $T = 10$ joining two WW domains. (Top) logos computed using the MF direct solution; the two amino acids allowed on each site in the direct subspace are the ones corresponding to v^{start} and v^{end} . The other amino acids are candidate for mutations outside the direct space. (Bottom) logos computed using the MF global solution. Here $\gamma = 1.6$ and $\beta = 3$.

2. Entropy of 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. To compute this entropy we follow Eq.(19). In the case of RBM the latter equation can be written as

$$\mathcal{S}_{\text{path}} = -\frac{\beta}{N} \sum_{i,t} \langle g_{i,t}(a) \rangle + \frac{1}{N} \sum_i \log Z_i - \frac{\beta}{N} \left(\Gamma'(\mathbf{m}) \frac{\partial}{\partial \hat{\mathbf{m}}} - \Phi'(\mathbf{q}) \frac{\partial}{\partial \hat{\mathbf{q}}} \right) \sum_i \log Z_i. \quad (61)$$

Estimates of $\mathcal{S}_{\text{path}}$ in the Cont and Evo scenarios are given in Figure 12(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 T in the Cont scenario. This behaviour can be understood in the following toy model. We consider a uniform (flat) landscape P_{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 μ and on A only. Hence, the entropy is approximately added 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 hard wall in Φ_{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 ρN 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 ρ is $-\rho \log \rho - (1-\rho) \log(1-\rho) \simeq \frac{\gamma}{T} \log T$ for large T . Hence, the total entropy (per site) of the paths of length T is expected to scale as $\sim \log T$.

3. Transition and escape probabilities

The partition function in Eq.(16) is computed on the ensemble of paths fixed at both ends to be equal to sequence \mathbf{v}_{start} and \mathbf{v}_{end} . One can easily redo the computation by relaxing one of these constraints. In particular one can compute the sum over all the paths anchored on the starting sequence \mathbf{v}_{start} but free to take any configuration at the other extremity. We show in Figure 12(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 (and the opposite for the Cont scenario). This arises from the fact that the unconstrained Evo solution jumps directly from the initial configuration to the closest minimum of the energy landscape, while the constrained Evo solution has to interpolate distant target sequences. The presence of an hard wall potentials in the Cont scenario forbids both solutions to remain in the same configurations for long times and 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 Cont constrained solution has to smoothly interpolate between distant region in such a way that the energy along the path is optimized, this makes the number of accessible paths lower with respect to the unconstrained solution.

Knowledge of the corresponding partially unconstrained free energy allows us to compute the probability to go from \mathbf{v}_{start} to \mathbf{v}_{end} in T dynamical steps:

$$P(\mathbf{v}_{start} \rightarrow \mathbf{v}_{end}|T) = \frac{\sum_{\mathcal{V}}^{\text{const.}} e^{-N\mathcal{E}(\mathcal{V};\mathbf{v}_{start},\mathbf{v}_{end})}}{\sum_{\mathcal{V}}^{\text{unconst.}} e^{-N\mathcal{E}(\mathcal{V};\mathbf{v}_{start})}} \underset{N \gg 1}{\sim} \frac{e^{-Nf_{\text{path}}^{\text{const.}}(\mathbf{v}_{start},\mathbf{v}_{end}|T)}}{e^{-Nf_{\text{path}}^{\text{unconst.}}(\mathbf{v}_{start}|T)}}. \quad (62)$$

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 μ (per step) combined with selection with probability P_{model} . We show in Figure 12(b) the transition probabilities for the Cont and Evo scenarios. The Evo scenario shows an optimal length T^* for which the probability is maximised, 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 Cont scenario by the stiffness of Φ_{Cont} , which increases with T .

The framework above can be extended to compute the probabilities of escaping from some minima of the free energy corresponding to the starting sequence towards some region \mathcal{R} of the sequence space in T steps. We define the escape probability through

$$P_{\text{escape}}(\mathcal{R}|T) = 1 - P_{\text{stay}}(\mathcal{R}|T), \quad (63)$$

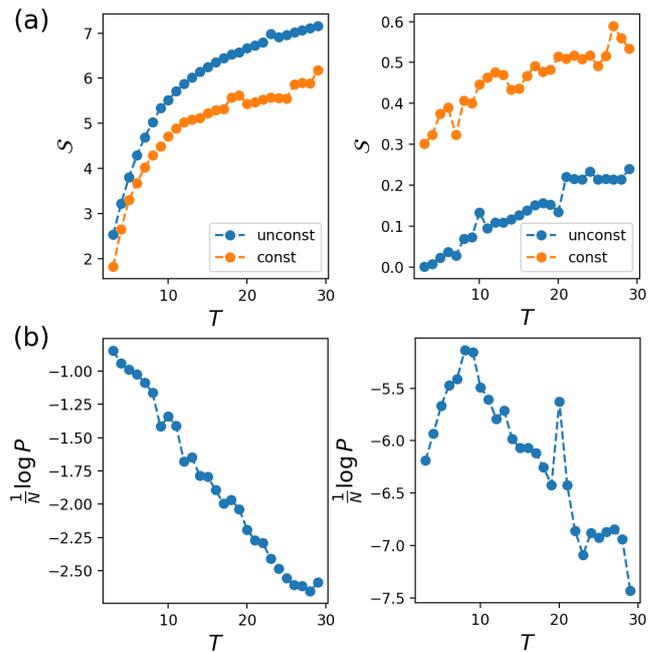


FIG. 12. **Entropies and probabilities of transition for the Cont (left) and Evo (right) potentials.** (a) Entropy S_{path} of paths as a function of T . (b) Transition probability as a function of T . Results are shown for paths joining the two WW domain wild-type sequences (constrained) and paths anchored by the starting sequence and free at the other extremity (unconstrained). Parameters for Evo: $\mu = 10^{-4}$, $\beta = 1$; for Cont: $\gamma = 3$, $\beta = 1$.

with

$$P_{\text{stay}}(\mathcal{R}|T) = \frac{\sum_{\mathcal{V}}^{\text{const.}} e^{-N\mathcal{E}(\mathcal{V};\mathbf{v}_{start},\mathcal{R})}}{\sum_{\mathcal{V}}^{\text{unconst.}} e^{-N\mathcal{E}(\mathcal{V};\mathbf{v}_{start})}} \underset{N \gg 1}{\sim} \frac{e^{-Nf_{\text{path}}^{\text{const.}}(\mathbf{v}_{start},\mathcal{R}|T)}}{e^{-Nf_{\text{path}}^{\text{unconst.}}(\mathbf{v}_{start}|T)}}, \quad (64)$$

where the sum at the numerator is restricted to the paths ending in \mathcal{R} , while the second sum is constrained only at the beginning. This leads to two free-energy optimisation problems: one for the constrained paths ensemble and one for the unconstrained one. Minimizing the free energy (Eq.(55)) for both scenarios and comparing them as in the last line of Eq. (64) leads to an estimation of the escaping probability. In Figure 13 we plot the probability of remaining in a certain local minima of the WW domain energy landscape in the Evo scenario. For different values of μ we are able to estimate at which time an evolving configuration is supposed to escape from the minima. 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 μ .

V. CONCLUSION

In the present work, we have focused on the nature of transition paths in Potts-like energy landscapes in high dimension N . These paths join two configura-

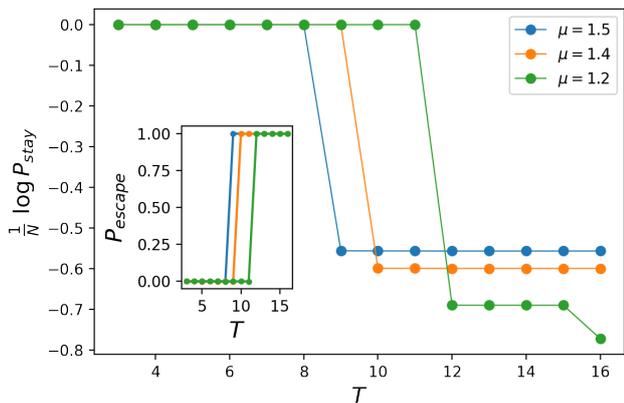


FIG. 13. **Probability of remaining (main panel, log scale along y -axis) and of escaping from (inset) the neighborhood of \mathbf{v}_{start}** for the WW domain, computed using Eq.(64) in the Evo scenario ($\beta = 1$). Three different values of the mutation rate μ are considered.

tions through a set of intermediates under two conditions. 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.

We have considered two kinds of elastic potentials. The Evo potential corresponds to evolutionary paths, *i.e.* it mimics random mutations at a constant rate μ [22], while the Potts-like energy plays the role of the selective pressure driving the evolution. To avoid local maxima in the landscape successive intermediate along Evo paths may occasionally differ by more than the average number of mutations, μN . To avoid such ‘jumps’ between configurations, we have introduced another potential, referred to as Cont, which uses a hard wall to impose a maximal number of mutations and ensure a smooth interpolation between sequences along the path.

Using mean-field theory, we have computed the typical properties of Evo and Cont paths in two contexts. The first one, called direct (*dir*) 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 (*glob*), 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 state A per site.

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 [23], where each data configuration activate a finite number of hidden units.

Last of all, we have tested our method on to data-driven models of natural proteins, 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 are definitely needed to improve our approach, *e.g.* by considering finite- N fluctuations around the mean-field theory solution. From an 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 is here not only theoretical but also practical, *e.g.* to gain intuition on how many random sequences can evolve a given functionality under selective pressure. In addition, 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 ICFP Labex fellowship of the Physics Department at ENS.

[1] S. F. Greenbury, A. A. Louis, and S. E. Ahnert, Nature Ecology & Evolution **6**, 1742 (2022).

[2] D. A. Stariolo and L. F. Cugliandolo, EPL (Europhysics

- Letters) **127**, 16002 (2019).
- [3] V. Ros, G. Biroli, and C. Cammarota, *SciPost Physics* **10**, 002 (2021).
- [4] F.-Y. Wu, *Reviews of modern physics* **54**, 235 (1982).
- [5] F. J. Poelwijk, M. Socolich, and R. Ranganathan, *Nat. Commun.* **10**, 1 (2019).
- [6] 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).
- [7] A. Hawkins-Hooker, F. Depardieu, S. Baur, G. Couairon, A. Chen, and D. Bikard, *PLOS Computational Biology* **17**, 1 (2021).
- [8] E. Mauri, S. Cocco, and R. Monasson, arXiv preprint arXiv:2204.10553 (2022).
- [9] A. Fischer and C. Igel, in *Iberoamerican congress on pattern recognition* (Springer, 2012) pp. 14–36.
- [10] J. Tubiana, S. Cocco, and R. Monasson, *eLife* **8**, e39397 (2019).
- [11] H. Jacquin, A. Gilson, E. Shakhnovich, S. Cocco, and R. Monasson, *PLOS Computational Biology* **12**, 1 (2016).
- [12] M. Kimura, *The neutral theory of molecular evolution* (Cambridge University Press, 1983).
- [13] T. Tieleman, in *ICML '08: Proceedings of the 25th international conference on Machine learning* (Association for Computing Machinery, New York, NY, USA, 2008) pp. 1064–1071.
- [14] Note that in order for Γ_μ 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.* $w_{i\mu}$, $g_i = \mathcal{O}(1)$.
- [15] K. F. Lau and K. A. Dill, *Macromolecules* **22**, 3986 (1989).
- [16] S. Miyazawa and R. L. Jernigan, *J. Mol. Biol.* **256**, 623 (1996).
- [17] Positive charged amino acids: Arginine (R), Histidine (H), Lysine (K). Negative charged amino acids: Aspartic acid (D), Glutamic acid (E).
- [18] M. Sudol, *Progress in biophysics and molecular biology* **65**, 113 (1996).
- [19] M. Sudol and T. Hunter, *Cell* **103**, 1001 (2000).
- [20] X. Espanel and M. Sudol, *Journal of Biological Chemistry* **274**, 17284 (1999).
- [21] W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffe, and R. Ranganathan, *Nature* **437**, 579 (2005).
- [22] I. Leuthäusser, *Journal of statistical physics* **48**, 343 (1987).
- [23] J. Tubiana and R. Monasson, *Physical review letters* **118**, 138301 (2017).