Out-of-equilibrium selection pressure enhances inference from protein sequence data

Nicola Dietler, Cyril Malbranke, and Anne-Florence Bitbol*
Institute of Bioengineering, School of Life Sciences,
École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland and
SIB Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland
(Dated: December 20, 2025)

Homologous proteins have similar three-dimensional structures and biological functions that shape their sequences. The resulting coevolution-driven correlations underlie methods from Potts models to AlphaFold, which infer protein structure and function from sequences. Using a minimal model, we show that fluctuating selection strength and the onset of new selection pressures improve coevolution-based inference of structural contacts. Our conclusions extend to realistic synthetic data and to the inference of interaction partners. Out-of-equilibrium noise arising from ubiquitous variations in natural selection thus enhances, rather than hinders, the success of inference from protein sequences.

Introduction. During evolution, proteins evolve through random mutations of their sequences and natural selection for function. This coupling between sequence and function leads to coevolution between amino acid sites in multiple sequence alignments (MSAs) of homologous proteins, which share ancestry, three-dimensional structure, and function. Amino-acid identities at contacting positions in a protein's three-dimensional structure are correlated, due to the need to preserve physico-chemical complementarity. This has allowed to predict structural contacts from sequences, using mutual information [1] and Potts models [2–5], which are pairwise maximum entropy models inferred on MSAs. Attention coefficients in protein language models trained on MSAs capture coevolution [6, 7], a key factor underlying AlphaFold's success in protein structure prediction [7]. Beyond structure, coevolution allows to infer interaction partners [8–11] and functional sectors of collectively correlated amino acids [12–15], as well as to investigate protein evolution [16–21].

Natural environments often change over time [22–27], causing variable selection pressures on proteins [22]. How do such fluctuations impact inference from sequences? To address this, we consider a minimal model of protein sequences evolving under fluctuating selection to preserve structure. We show that the resulting out-of-equilibrium noise can enhance the inference of structural contacts via Potts models. This conclusion extends to realistic synthetic data. It also holds in a simple model of the onset of a new selection pressure, where a random ancestral sequence evolves under structural selection with phylogeny. Finally, we show that out-of-equilibrium selection also enhances the inference of interaction partners. Our results suggest that time-dependent external drives can facilitate inference by pairwise maximum entropy models.

Model and methods. We consider a minimal model, where protein sequences are represented as length-L sequences of Ising spins $(\sigma_1, \ldots, \sigma_L)$, defined on the nodes of a fixed Erdős-Rényi random graph. Structural contacts are modeled by pairwise ferromagnetic couplings on the graph edges, all set to 1 for simplicity, leading to the Hamiltonian

$$H = -\sum_{(i,j)\in\mathcal{E}} \sigma_i \sigma_j \,, \tag{1}$$

where $\mathcal{E}$ denotes the set of graph edges, see also [28]. We sample independent equilibrium sequences using a Metropolis–Hastings algorithm by proposing spin flips at random sites and accepting them with probability

$$p = \min\left[1, \exp(-\Delta H/T)\right] \,, \tag{2}$$

where ΔH represents the energy variation associated to the flip, and T the Monte Carlo sampling temperature, see Fig. S1(a) for a schematic and Fig. S2 for information about equilibration. Selection strength is controlled by T, with higher T corresponding to weaker selection [28]. We model variable selection strength using a telegraph process that switches between T_1 and $T_2 > T_1$ with equal transition rates $1/\tau$. We start from independent equilibrium sequences at temperature T_1, see Fig. S1(b).

To go beyond this minimal model, we generate realistic synthetic data using a Potts model inferred from natural sequences via bmDCA [29, 30], an approach whose generative power was experimentally validated [30].

Finally, to model the onset of a new selection pressure, we start from a random ancestral sequence and apply selection for structure using the minimal model described above. We evolve different sequences from the same ancestor via independent Markov chains, producing a simple star phylogeny, see Fig. S1(c).

Fluctuating selection enhances contact inference. In our model, selection strength is controlled by the sampling temperature T. At very low T, only spin flips that decrease the energy (Eq. 1) are accepted, leading to low sequence variability and poor inference. Conversely, at very high T, almost all spin flips are accepted, producing noisy sequences that also hinder inference. As a result, inference is optimal at intermediate temperatures.

* anne-florence.bitbol@epfl.ch

In the present minimal model featuring a ferromagnetic-paramagnetic phase transition (Fig. S3 and [28, 31]), inference is most successful around and moderately above this transition [28, 32]. With our parameter values, it occurs at $T_C = 4$ (Fig. S4). Inference performance also depends on the number of available sequences (Fig. S5 and [31, 32]). In natural data, this number is limited by experimental sampling and by biological evolution, leading to finite-size effects that hinder inference [3, 4, 33]. Thus motivated, we work with finite MSAs throughout.

How does the time variability in selection pressure impact inference from sequence data? To answer this within our minimal model, we start from independent sequences equilibrated at $T_1 = 1$ and evolve them under a time-varying sampling temperature that switches between $T_1 = 1$ and $T_2 = 15$.

Fig. 1 shows the True Positive (TP) fraction – defined as the fraction of true contacts among the top N inferred Ising couplings, where N is the actual number of contacts – versus the number of accepted spin flips, representing mutations. Panel (a) shows that the average TP fraction increases with time before plateauing, indicating that fluctuating selection strength enhances inference performance. Faster temperature switches further improve inference by increasing sequence diversity. The effective number of sequences, M_{eff} [2], defined by grouping sequences with Hamming distance smaller than 0.2, reaches 1980. This is close to the value for an MSA generated at $T = 15$ (2007), and to the total number of sequences per MSA (2048), and far above the nearly frozen $M_{\text{eff}} = 2$ at $T = 1$. Inference under variable temperature outperforms both fixed-temperature cases at T_1 and T_2, and matches the best fixed-temperature case, see Fig. S4, demonstrating that variable selection strength can substantially improve inference.

Fig. 1(b) shows that, in single realizations of the process, switching to strong selection (T_1) after a weak selection phase (T_2) yields a transient improvement in inference performance (see also Fig. S6). Increasing selection strength suppresses noise that partly obscures structural correlations, before significantly reducing sequence diversity. Thus, switching between weak and strong selection temporarily enhances structural correlations in sequences. If the system is held at a fixed temperature for times comparable to the equilibration times (see Fig. S2), inference performance relaxes to its equilibrium value at that temperature. The observed improvement is therefore a transient, out-of-equilibrium effect. Faster temperature alternations, on timescales significantly below equilibration, amplify these effects and further enhance inference performance, as seen in Fig. 1(a).

What is the impact of the temperatures T_1 and T_2 between which the system switches? In our minimal model, inference is optimal around and slightly above the ferromagnetic-paramagnetic transition temperature T_C (see above). Accordingly, alternating between $T_1 < T_C$ and $T_2 > T_C$ improves inference, whereas switching between two temperatures below or above T_C does not (Fig. S7). Furthermore, for fast switches, the data resembles equilibrium data generated at an intermediate temperature between T_1 and T_2 (Fig. S8), consistent with

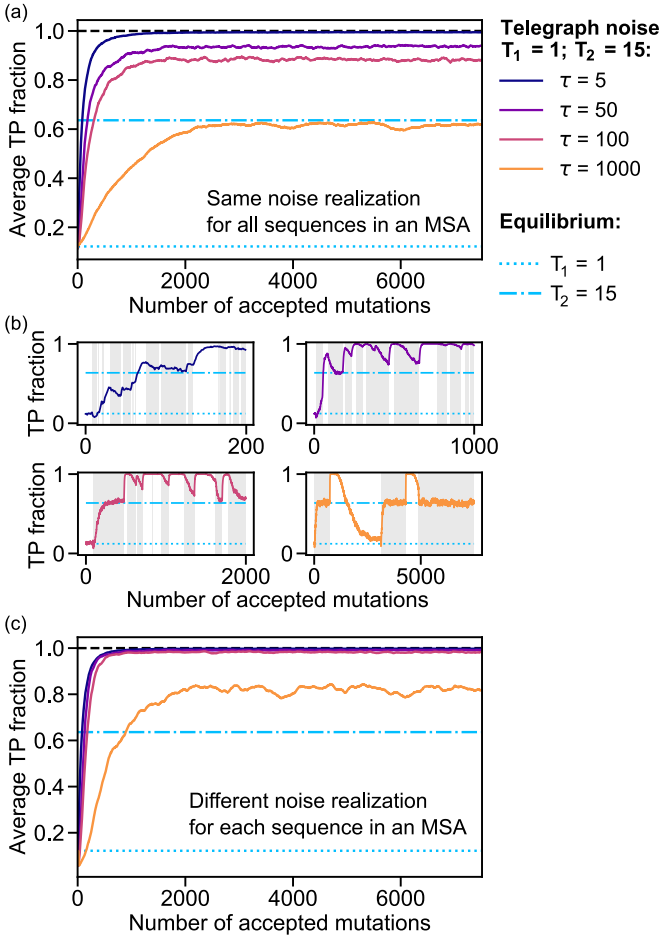

FIG. 1. **Impact of fluctuating selection strength on contact prediction.** TP fraction versus number of accepted mutations per sequence, while switching selection strength (i.e. the sampling temperature) via a telegraph process with timescale τ. (a) For each value of τ, the TP fraction is averaged over 1000 replicates. In each replicate, the same telegraph process is used for all sequences in the MSA. Different replicates use different realizations of the data generation and of the telegraph process. (b) The TP fraction is shown for single realizations of the telegraph process and data generation. Gray background: T_2; white background: T_1. (c) Same as in (a), except that a different telegraph process is used for each sequence in an MSA. In all panels, TP fractions for equilibrium sequences generated at T_1 and T_2 are shown for reference. Sequences are generated using our minimal model, see Fig. S1(a-b), with $T_1 = 1$ and $T_2 = 15$. For each realization, we generate an MSA of 2048 sequences of length 200 using an Erdős-Rényi random graph with probability 0.02 to represent contacts, and infer contacts via mean-field DCA (mfDCA) [3, 4], which is computationally efficient, and performs well on minimal data [28], with pseudocount 0.01.

the observed inference performance.

So far, all sequences in an MSA were subjected to the same fluctuating selection strength. However, in natural data, different sequences (e.g. from different species) may experience different selection histories. Fig. 1(c) shows inference performance when each sequence in an MSA is subjected to an independent realization of the temperature fluctuations. Inference is enhanced even more strongly than in Fig. 1(a), indicating that heterogeneous fluctuating selection further amplifies the effect.

Extension to realistic synthetic data. How do the observations made with our minimal model generalize to more realistic data? To address this question, we infer a Potts model from a natural MSA of the protein family PF0004 (AAA ATPase) from Interpro/Pfam [34] via bmDCA [29, 30]. We then generate sequences from this Potts model, either at equilibrium or under fluctuating selection strength (see Model and Methods).

Fig. 2 shows the TP fraction versus the number of accepted mutations for this realistic synthetic data. As in the minimal model, alternating between strong and weak selection improves inference performance, with faster alternations between low and high temperature yielding larger gains. Inference performance then becomes comparable to the best equilibrium performance, obtained at temperature $T = 1$ (Fig. S9), even though both alternating temperatures yield poorer inference at equilibrium. Thus, our conclusions from the minimal model extend to realistic synthetic data.

Impact of selection onset with phylogeny. Beyond fluctuating environments, out-of-equilibrium effects also arise when a new selection pressure emerges. We model this by generating sequences starting from a random ancestral sequence, under selection for structural contacts at fixed T in the minimal model, using a simple star phylogeny (see Model and Methods). This setup captures the sudden onset of a specific selection pressure, starting from an ancestor that was not subject to it.

Fig. 3 shows the TP fraction versus the number μ of accepted mutations per branch in the star phylogeny. For $T < 4$, i.e. below the ferromagnetic-paramagnetic transition of the minimal model, the TP fraction exhibits a maximum at intermediate μ. Inference then performs significantly better than the equilibrium TP fraction approached at large μ. This enhancement coincides with a transient increase of sequence diversity (Fig. S10), and persists for a balanced binary branching tree (Fig. S11). Hence, under strong selection, out-of-equilibrium noise originating from the random ancestor enhances contact inference. The beneficial effect of out-of-equilibrium noise evidenced for fluctuating selection thus also applies to the onset of a new selection pressure.

Extension to interaction partner inference. So far, we considered structural contact inference. Coevolution-based methods are also used to infer protein interaction partners [8–10, 38]. To test whether our results extend

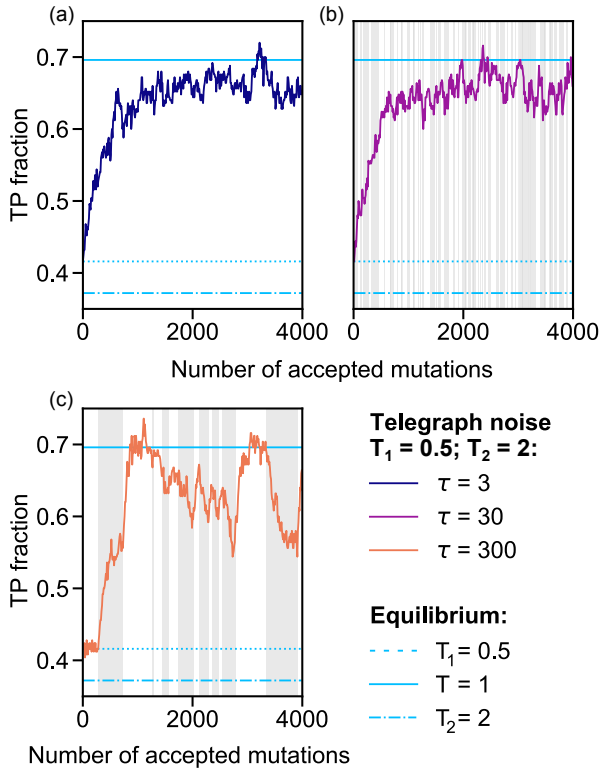

FIG. 2. **Impact of fluctuating selection strength on contact prediction: PF0004 family.** As in Fig. 1(b), the TP fraction is shown versus the number of accepted mutations under switching selection strength via a telegraph process, but for realistic data composed of MSAs of 70,000 sequences generated from a Potts model inferred on a natural MSA of 39,277 sequences from the PF0004 family (with length 132), following Refs. [29, 35], in particular using regularization strengths 0.01. Gray background: T_2; white background: T_1 (alternations not shown in the first panel for readability). Inference performances for equilibrium sequences generated at T_1, T_2, and $T = 1$, are shown for reference. Contact inference is performed using plmDCA [36, 37], with regularization strengths set to 0.01 and no phylogenetic reweighting.

to this problem, we perform partner inference within our minimal model using the method of [38]. As for contact inference, partner inference performance peaks at an intermediate temperature (Fig. S4). Fig. 4 shows that fluctuating selection strength enhances partner inference, indicating that the beneficial effect of out-of-equilibrium selection extends beyond contact inference.

Discussion. Using a minimal model, we showed that alternating between weak and strong selection enhances structural contact inference from sequences. This effect extends to realistic synthetic sequences and to the onset of a new selection pressure. Moreover, out-of-equilibrium selection also enhances interaction partner inference.

Under strong selection, limited sequence variability hinders coevolution-based inference. Alternating with weak selection, or starting from an unselected ancestor transiently increases diversity, enhancing inference.

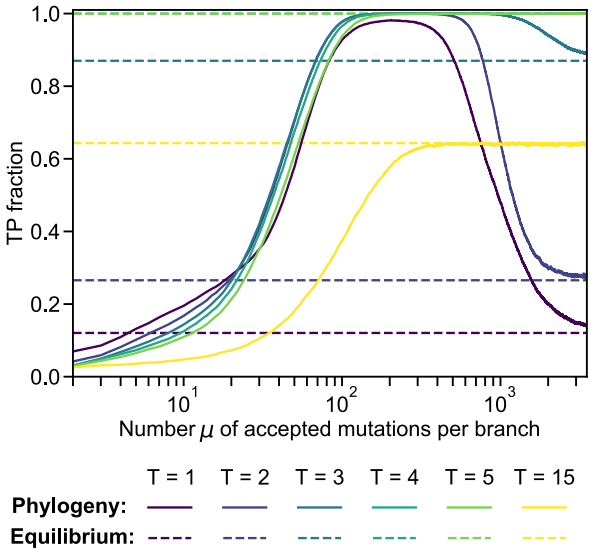

FIG. 3. **Impact of selection onset on contact inference with phylogeny.** The TP fraction is shown versus the number μ of accepted mutations per branch of the star phylogenetic tree, starting from a random ancestral sequence. Sequences with length 200 are generated as explained in Fig. S1(c), using the same Erdős-Rényi graph as in Fig. 1, along a star phylogenetic tree with 2048 branches. Along each branch, Monte Carlo sampling is performed until μ mutations are accepted, giving 2048 final sequences, which constitute our MSA. We infer contacts via mfDCA [3, 4, 28] with pseudocount 0.01. Results are averaged over 100 replicates, each starting from a different random ancestor.

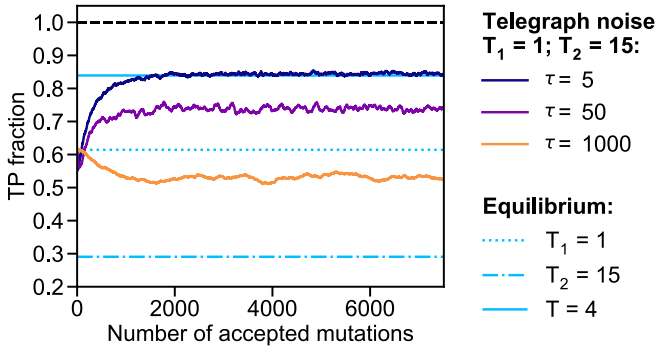

FIG. 4. **Impact of fluctuating selection strength on interaction partner prediction.** TP fraction versus number of accepted mutations, while switching selection strength (i.e. the sampling temperature) via a telegraph process with timescale τ. For each value of τ, the TP fraction is averaged over 1000 replicates. TP fractions for equilibrium sequences generated at T_1 and T_2 are shown for reference. MSAs comprising 1024 sequences are generated as in Fig. 1(a), and then randomly split into a training set of 400 sequences and a testing set of 624 sequences. The latter is randomly divided in sets of 4 sequences each, representing species, and each sequence is split in two halves of equal length representing two interaction partners. Pairings between partners are blinded in each species, and predicted using mfDCA scores with pseudocount 0.01 [38].

Overall, our results reveal that, far from hindering it, out-of-equilibrium selection can facilitate the extraction of structural and functional information from protein sequences. While we relied on synthetic data, which allowed us to vary selection pressures and control ancestral sequences, the extension to realistic data suggests that these principles are generic and relevant for natural proteins, where variable selection pressures are ubiquitous [22, 23]. Hence, our findings suggest that the success of coevolution-based inference in proteins may be partly caused by fluctuating selection intensity.

We considered a simple form of selection variability, switching between two strengths. Other forms of fluctuations of selection strengths, including general random walks, are expected to yield similar results, provided that weak and strong selection are included and the fluctuations are faster than equilibration times. Besides, while we focused on structural selection pressures, which are a major constraint on protein evolution [39], our work could be extended to other evolutionary constraints, such as those shaping collectively correlated amino acid groups [12–15], and to RNA structure inference [40]. Additional perspectives include selection pressure variations that cannot be captured by an overall scaling of selection strength, e.g. affecting specific fields or couplings. Finally, although we focused on inference based on Potts models, coevolution also underlies recent deep learning approaches, including AlphaFold [7], indicating that out-of-equilibrium fluctuations may similarly benefit these approaches.

Beyond natural evolution, our results have implications for directed evolution experiments [41]. In recent studies [42, 43], structural prediction was performed using sequences evolved *in vitro* under constant selection strength. It was later shown [44] that tuning this strength can improve contact inference. Our results suggest that dynamically varying selection strength during experiments, e.g. by varying antibiotic concentration, could also enhance inference.

More broadly, pairwise maximum entropy models are widely used for inference beyond protein sequence data. A prominent example is neuronal population activity [45–47]. Neurons are often driven by common input stimuli that vary in time, and could play a role analogous to fluctuating selection strength in proteins. Such inputs have been found to impact inference [48], increase inferred couplings in some cases [49], and generate signatures reminiscent of criticality [50, 51], which are frequently observed in biological systems [52]. Our results thus suggest a potentially general principle: time-dependent external drives can enhance inference in systems described by pairwise maximum entropy models.

Code accessibility. Our code is available at `https://github.com/Bitbol-Lab/OutOfEquilibrium-ProtSeq`.

This research was partly funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 851173, to A.-F. B.).

[1] S. D. Dunn, L. M. Wahl, and G. B. Gloor, Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction, Bioinformatics **24**, 333 (2008).

[2] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing, Proc. Natl. Acad. Sci. U.S.A. **106**, 67 (2009).

[3] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc. Natl. Acad. Sci. U.S.A. **108**, E1293 (2011).

[4] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, Protein 3D structure computed from evolutionary sequence variation, PLoS ONE **6**, e28766 (2011).

[5] J. P. Barton, E. De Leonardis, A. Coucke, and S. Cocco, ACE: adaptive cluster expansion for maximum entropy graphical model inference, Bioinformatics **32**, 3089 (2016).

[6] R. M. Rao, J. Liu, R. Verkuil, J. Meier, J. Canny, P. Abbeel, T. Sercu, and A. Rives, MSA Transformer, Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research, **139**, 8844 (2021).

[7] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature **596**, 583 (2021).

[8] L. Burger and E. van Nimwegen, Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method, Mol. Syst. Biol. **4**, 165 (2008).

[9] A.-F. Bitbol, R. S. Dwyer, L. J. Colwell, and N. S. Wingreen, Inferring interaction partners from protein sequences, Proc. Natl. Acad. Sci. U.S.A. **113**, 12180 (2016).

[10] A. F. Bitbol, Inferring interaction partners from protein sequences using mutual information, PLoS Comput. Biol. **14**, e1006401 (2018).

[11] U. Lupo, D. Sgarbossa, and A.-F. Bitbol, Pairing interacting protein sequences using masked language modeling, Proc. Natl. Acad. Sci. USA **121**, e2311887121 (2024).

[12] N. Halabi, O. Rivoire, S. Leibler, and R. Ranganathan, Protein sectors: evolutionary units of three-dimensional structure, Cell **138**, 774 (2009).

[13] O. Rivoire, K. A. Reynolds, and R. Ranganathan, Evolution-based functional decomposition of proteins, PLoS Comput. Biol. **12**, e1004817 (2016).

[14] S.-W. Wang, A.-F. Bitbol, and N. Wingreen, Revealing evolutionary constraints on proteins through sequence analysis, PLOS Comput. Biol. **15**, e1007010 (2019).

[15] N. Dietler, A. Abbara, S. Choudhury, and A.-F. Bitbol, Impact of phylogeny on the inference of functional sectors from protein sequence data, PLOS Comput. Biol. **20**, 1 (2024).

[16] J. A. de la Paz, C. M. Nartey, M. Yuvaraj, and F. Morcos, Epistatic contributions promote the unification of incompatible models of neutral molecular evolution, Proc. Natl. Acad. Sci. USA **117**, 5873 (2020).

[17] L. Di Bari, M. Bisardi, S. Cotogno, M. Weigt, and F. Zamponi, Emergent time scales of epistasis in protein evolution, Proc. Natl. Acad. Sci. USA **121**, e2406807121 (2024).

[18] M. De Leonardis, A. Pagnani, and P. Barrat-Charlaix, Reconstruction of ancestral protein sequences using autoregressive generative models, Mol. Biol. Evol. **42**, msaf070 (2025).

[19] A. Biswas, I. Choudhuri, E. Arnold, D. Lyumkis, A. Haldane, and R. M. Levy, Kinetic coevolutionary models predict the temporal emergence of HIV-1 resistance mutations under drug selection pressure, Proc. Natl. Acad. Sci. U.S.A. **121**, e2316662121 (2024).

[20] J. Gizzio, A. Thakur, A. Haldane, C. B. Post, and R. M. Levy, Evolutionary sequence and structural basis for the distinct conformational landscapes of Tyr and Ser/Thr kinases, Nat. Commun. **15**, 6545 (2024).

[21] S. Alvarez, C. M. Nartey, N. Mercado, J. A. de la Paz, T. Huseinbegovic, and F. Morcos, In vivo functional phenotypes from a computational epistatic model of evolution, Proc. Natl. Acad. Sci. U.S.A. **121**, e2308895121 (2024).

[22] V. Mustonen and M. Lässig, Molecular evolution under fitness fluctuations, Phys. Rev. Lett. **100**, 108101 (2008).

[23] G. Bell, Fluctuating selection: the perpetual renewal of adaptation in variable environments, Phil. Trans. R. Soc. B **365**, 87 (2010).

[24] A. Melbinger and M. Vergassola, The impact of environmental fluctuations on evolutionary fitness functions, Sci. Rep. **5**, 15211 (2015).

[25] K. Wienand, E. Frey, and M. Mobilia, Evolution of a fluctuating population in a randomly switching environment, Phys. Rev. Lett. **119**, 158301 (2017).

[26] A. Taitelbaum, R. West, M. Assaf, and M. Mobilia, Population dynamics in a changing environment: Random versus periodic switching, Phys. Rev. Lett. **125**, 048105 (2020).

[27] I. Meyer and N. M. Shnerb, Evolutionary dynamics in fluctuating environment, Phys. Rev. Res. **2**, 023308 (2020).

[28] N. Dietler, U. Lupo, and A.-F. Bitbol, Impact of phylogeny on structural contact inference from protein sequence data, J. R. Soc. Interface **20**, 20220707 (2023).

[29] M. Figliuzzi, P. Barrat-Charlaix, and M. Weigt, How Pairwise Coevolutionary Models Capture the Collective Residue Variability in Proteins?, Mol. Biol. Evol. **35**, 1018 (2018).

[30] 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, An evolution-based model for designing chorismate mutase enzymes, Science **369**, 440 (2020).

[31] C. A. Gandarilla-Pérez, P. Mergny, M. Weigt, and A.-F. Bitbol, Statistical physics of interacting proteins: Impact of dataset size and quality assessed in synthetic sequences, Phys. Rev. E **101**, 032413 (2020).

[32] V. Ngampruetikorn, V. Sachdeva, J. Torrence, J. Humplik, D. J. Schwab, and S. E. Palmer, Inferring couplings in networks across order-disorder phase transitions, Phys. Rev. Research **4**, 023240 (2022).

[33] S. Ovchinnikov, H. Kamisetty, and D. Baker, Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information, Elife **3**, e02030 (2014).

[34] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, and A. Bateman, Pfam: The protein families database in 2021, Nucleic Acids Res. **49**, D412 (2020).

[35] U. Lupo, D. Sgarbossa, and A.-F. Bitbol, Protein language models trained on multiple sequence alignments learn phylogenetic relationships, Nat. Commun. **13**, 6298 (2022).

[36] M. Ekeberg, C. Lovkvist, Y. Lan, M. Weigt, and E. Aurell, Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models, Phys. Rev. E **87**, 012707 (2013).

[37] M. Ekeberg, T. Hartonen, and E. Aurell, Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences, J. Comput. Phys. **276**, 341 (2014).

[38] A. Gerardos, N. Dietler, and A. F. Bitbol, Correlations from structure and phylogeny combine constructively in the inference of protein partners from sequences, PLOS Comput. Biol. **18**, e1010147 (2022).

[39] F. Morcos, N. P. Schafer, R. R. Cheng, J. N. Onuchic, and P. G. Wolynes, Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection, Proc. Natl. Acad. Sci. U.S.A. **111**, 12408 (2014).

[40] M. B. Zerihun, F. Pucci, E. K. Peter, and A. Schug, pydca v1.0: a comprehensive software for direct coupling analysis of RNA and protein sequences, Bioinformatics **36**, 2264 (2019).

[41] F. H. Arnold, Directed evolution: Bringing new chemistry to life, Angew. Chem. Int. Ed. **57**, 4106 (2018).

[42] M. Fantini, S. Lisi, P. De Los Rios, A. Cattaneo, and A. Pastore, Protein Structural Information and Evolutionary Landscape by In Vitro Evolution, Mol. Biol. Evol. **37**, 1179 (2019).

[43] M. A. Stiffler, F. J. Poelwijk, K. P. Brock, R. R. Stein, A. Riesselman, J. Teyra, S. S. Sidhu, D. S. Marks, N. P. Gauthier, and C. Sander, Protein structure from experimental evolution, Cell Syst. **10**, 15 (2020).

[44] M. Bisardi, J. Rodriguez-Rivas, F. Zamponi, and M. Weigt, Modeling Sequence-Space Exploration and Emergence of Epistatic Signals in Protein Evolution, Mol. Biol. Evol. **39**, msab321 (2021).

[45] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature **440**, 1007 (2006).

[46] L. Meshulam, J. L. Gauthier, C. D. Brody, D. W. Tank, and W. Bialek, Collective Behavior of Place and Nonplace Neurons in the Hippocampal Network, Neuron **96**, 1178 (2017).

[47] L. Meshulam and W. Bialek, Statistical mechanics for networks of real neurons, Rev. Mod. Phys. **97**, 045002 (2025).

[48] J. Tyrcha, Y. Roudi, M. Marsili, and J. Hertz, The effect of nonstationarity on models inferred from neural data, J. Stat. Mech. **2013**, P03005 (2013).

[49] Y. Roudi, J. Tyrcha, and J. Hertz, Ising model for neural data: Model quality and approximate methods for extracting functional connectivity, Phys. Rev. E **79**, 051915 (2009).

[50] D. J. Schwab, I. Nemenman, and P. Mehta, Zipf's Law and Criticality in Multivariate Data without Fine-Tuning, Phys. Rev. Lett. **113**, 068102 (2014).

[51] V. Ngampruetikorn, I. Nemenman, and D. J. Schwab, Extrinsic vs intrinsic criticality in systems with many components, Phys. Rev. Res. **7**, 013188 (2025).

[52] T. Mora and W. Bialek, Are biological systems poised at criticality?, J. Stat. Phys. **144**, 268 (2011).

Supplementary material for "Out-of-equilibrium selection pressure enhances inference from protein sequence data"

Nicola Dietler, Cyril Malbranke, and Anne-Florence Bitbol*
Institute of Bioengineering, School of Life Sciences,
École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland and
SIB Swiss Institute of Bioinformatics, CH-1015 Lausanne, Switzerland
(Dated: December 20, 2025)

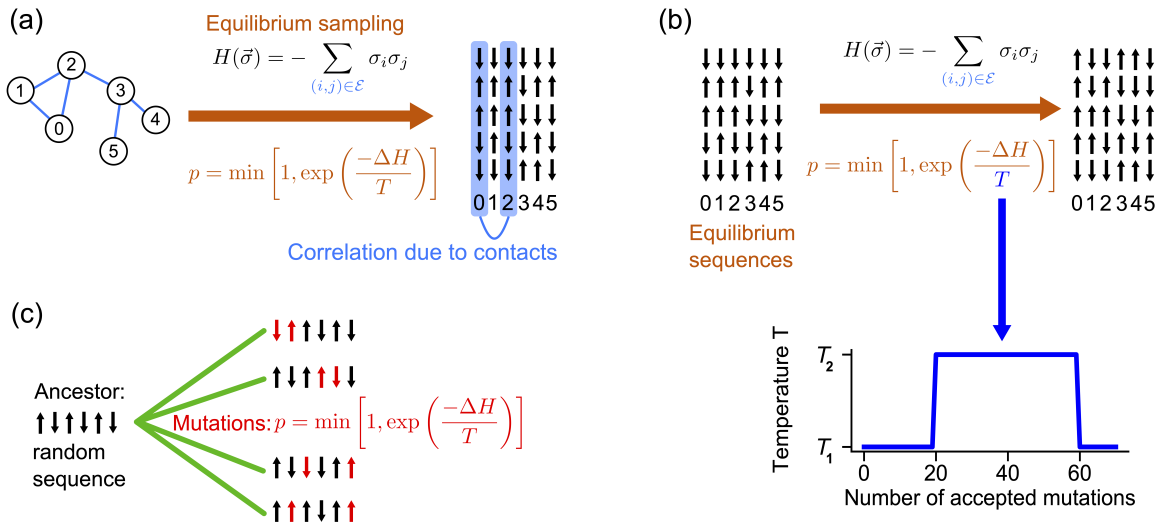

FIG. S1. **Minimal model**. (a) We model structural contacts as couplings set to 1 on the edges of an Erdős-Rényi random graph. Couplings between other nodes are set to 0. Equilibrium sampling of independent sequences is performed using a Metropolis–Hastings algorithm under the Hamiltonian in Eq. (1), with spin flip acceptance probability in Eq. (2), starting from random initial sequences. The MSA formed from the resulting sequences features correlations arising from couplings, shown between columns 0 and 2 (blue). (b) To generate sequences with fluctuating selection strength, we start from sequences generated at equilibrium with temperature T_1. We then evolve each of these sequences using the same Metropolis–Hastings algorithm as before, but with a sampling temperature T that switches between T_1 and T_2. (c) To minimally model the impact of emerging selection during evolution along a phylogeny, we take a random sequence as ancestor, and we evolve it along a star phylogenetic tree (green) where mutations (red) are accepted with the probability in Eq. (2). This represents selection for structural contacts being switched on. On each branch of the tree, μ mutations are accepted, with $\mu = 2$ in our schematic.

* anne-florence.bitbol@epfl.ch

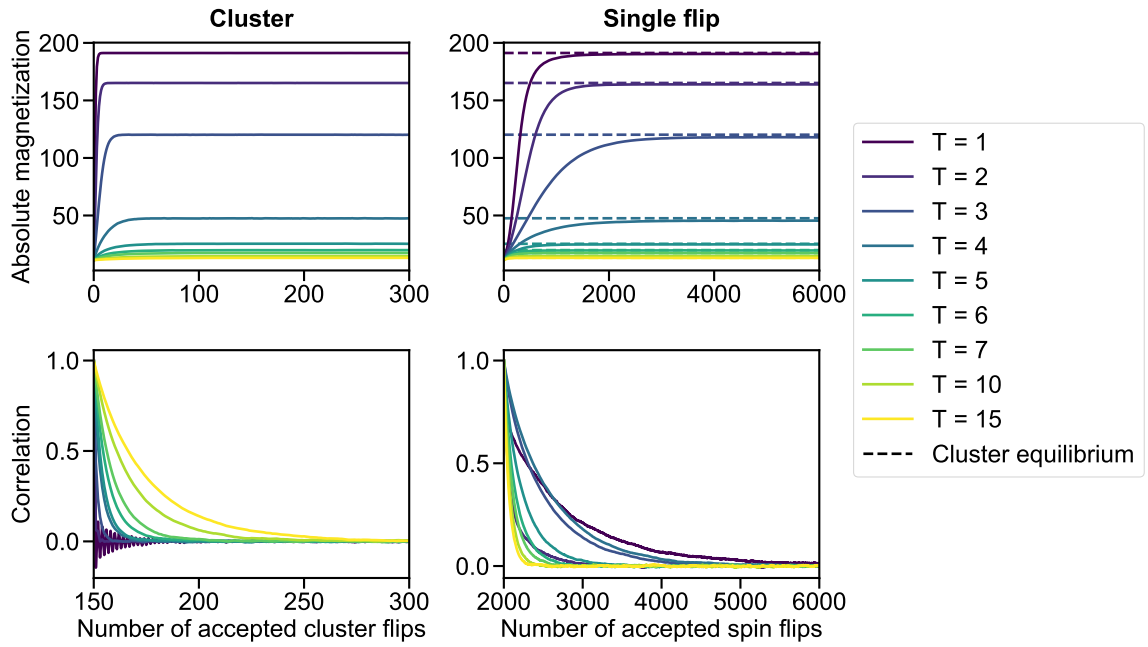

FIG. S2. **Equilibration of sequences.** We report results regarding the generation of equilibrium sequences at fixed sampling temperature by two Metropolis Monte Carlo algorithms, namely the Wolff cluster algorithm [1] (left panels) and the single flip algorithm (right panels). Note that, elsewhere in this work, the single flip algorithm is used if not stated otherwise, given its biological relevance (single flips representing mutations). We start from sequences where each spin is chosen uniformly at random. Absolute magnetization (top) and correlation of the absolute magnetization with that of the initial state (bottom) are shown versus the number of accepted cluster flips (left) or single spin flips (right). As equilibrium is reached, magnetisation plateaus and correlation tends to zero. This process is faster with the cluster algorithm. For comparison, the final value or magnetization reached with this algorithm is shown on the top right panel. At each sampling temperature, 2048 sequences are generated, each starting from a different random sequence. Results are averaged over 100 replicates.

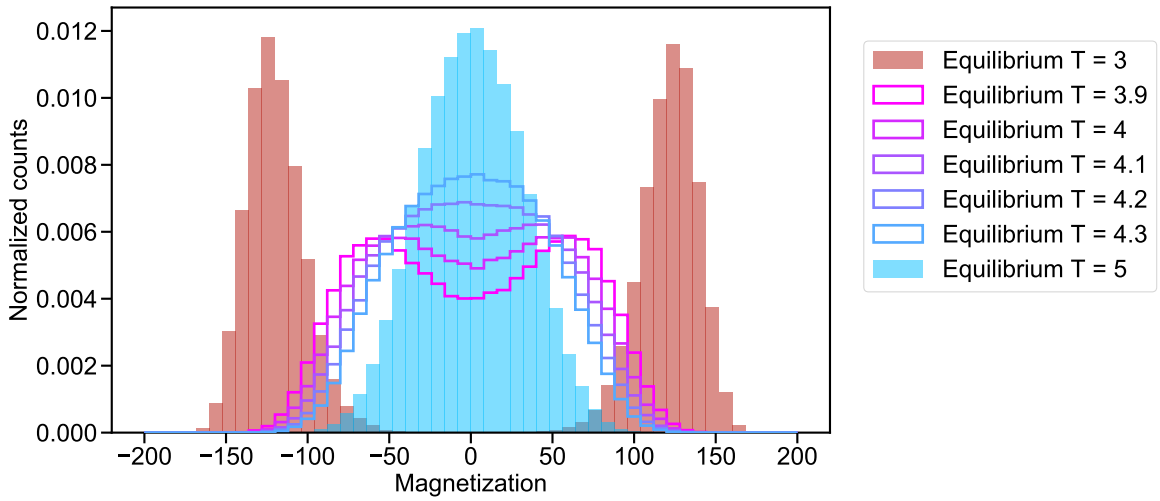

FIG. S3. **Magnetization histograms evidencing a ferromagnetic-paramagnetic transition.** We report histograms of the magnetization at different sampling temperatures. Sequences are generated at equilibrium using the cluster algorithm, starting from random sequences and using 300 accepted cluster flips for each sequence. Each histogram is computed over 2048 sequences. The same Erdős-Rényi graph as in Fig. 1 and elsewhere is used to represent contacts.

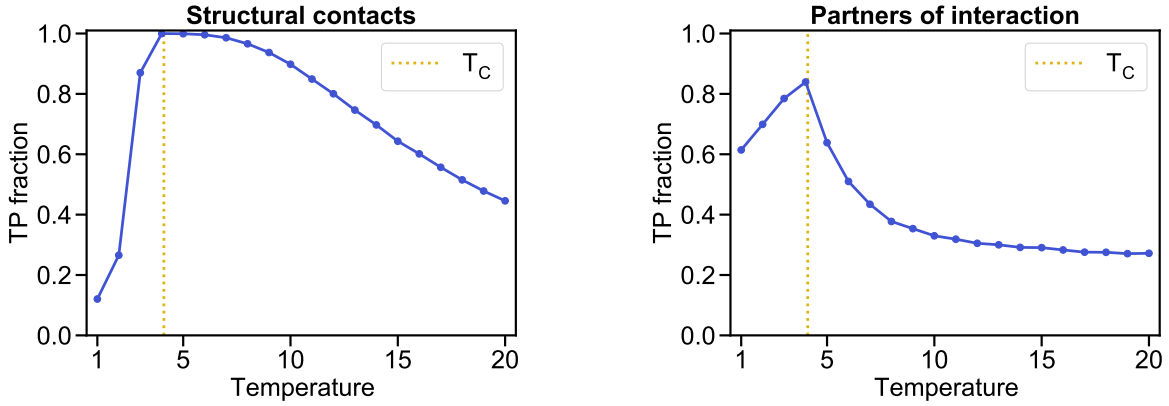

FIG. S4. **Impact of sampling temperature on contact and partner predictions at equilibrium.** We show the TP fraction as a function of sampling temperature for contact prediction (left panel) and interaction partner prediction (right panel). Sequences are generated at equilibrium for different temperatures using the minimal model, see Fig. S1(a-b), and the cluster algorithm (starting from random sequences and using 300 accepted cluster flips), see Fig. S2. For structural contact inference, each TP fraction is computed over 2048 sequences, using the same method as in Fig. S1. For partner inference, each TP fraction is computed over 624 sequences, using the exact same method (and data split approach) as in Fig. 4. The approximate temperature T_C of the ferromagnetic-paramagnetic transition, see Fig. S3, is indicated by a vertical dotted line. All results are averaged over 100 replicates.

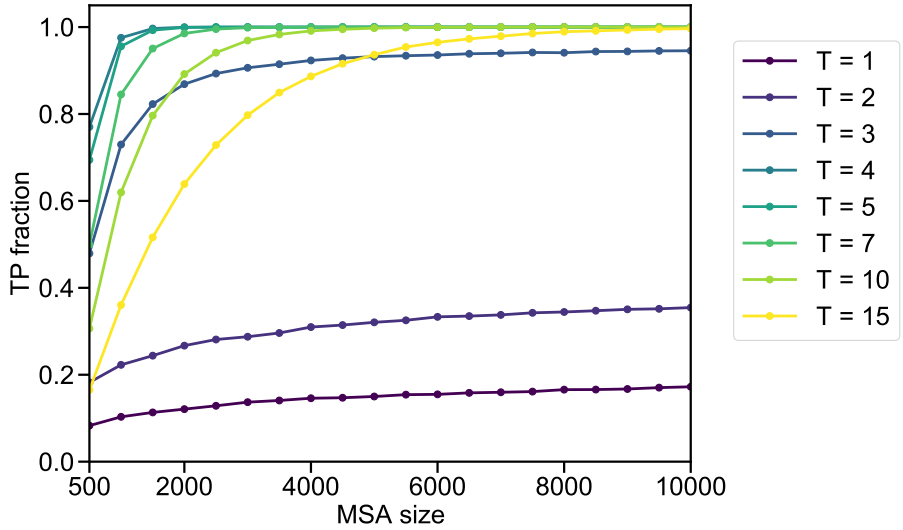

FIG. S5. **Impact of MSA size on structural contact prediction.** We report the TP fraction for structural contact prediction versus the MSA size used to infer contacts. MSAs are generated at equilibrium for each sampling temperature considered (using the cluster algorithm, starting from random sequences and using 300 accepted cluster flips for each sequence). As elsewhere, we infer contacts via mfDCA [2–4] with pseudocount 0.01.

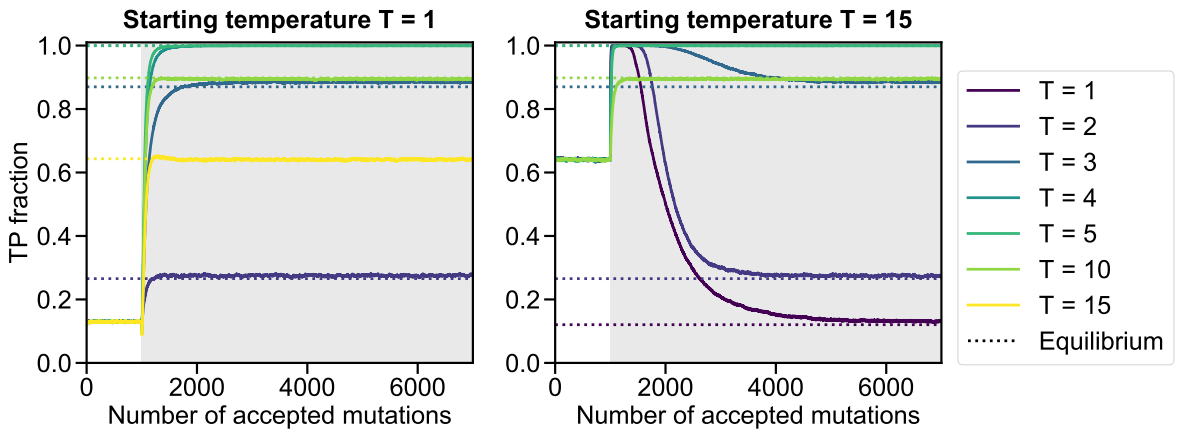

FIG. S6. **Impact of switching the sampling temperature on the performance of contact inference.** We report the TP fraction for contact inference when starting from an equilibrated MSA at temperature $T = 1$ (left panel) or $T = 15$ (right panel) and switching to a different temperature. White background: initial temperature; grey background: new temperature. The initial MSA is generated with the cluster algorithm (starting from random sequences and using 300 accepted cluster flips) and then single flips are applied, 1000 being accepted at the initial temperature, before the switch occurs. In each case, the TP fraction is calculated using MSAs of 2048 sequences, and an averaging over 100 replicates is performed.

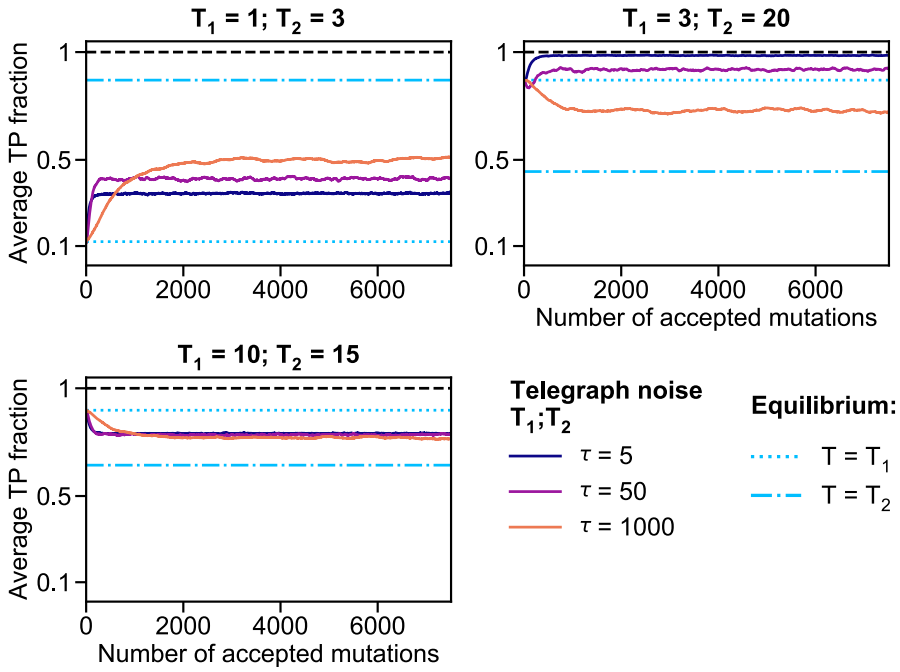

FIG. S7. **Impact of fluctuating selection strength on contact prediction, using different sampling temperatures.** As in Fig 1(a), we report the TP fraction versus the number of accepted mutations, while switching selection strength (i.e. the sampling temperature) via a telegraph process with timescale τ, but different pairs of temperatures $(T_1; T_2)$ are employed (one pair in each panel). For each value of τ, the TP fraction is averaged over 1000 replicates. In each replicate, the same telegraph process is used for all sequences in the MSA. Different replicates use different realizations of the data generation and of the telegraph process. In all panels, TP fractions for equilibrium sequences generated at T_1 and T_2 are shown for reference. Sequences are generated using our minimal model, see Fig. S1(a-b). For each realization, we generate an MSA of 2048 sequences of length 200 using the same Erdős-Rényi random graph with probability 0.02 to represent contacts as in Fig. S1, and we infer contacts via mean-field DCA (mfDCA) [2–4], with pseudocount 0.01.

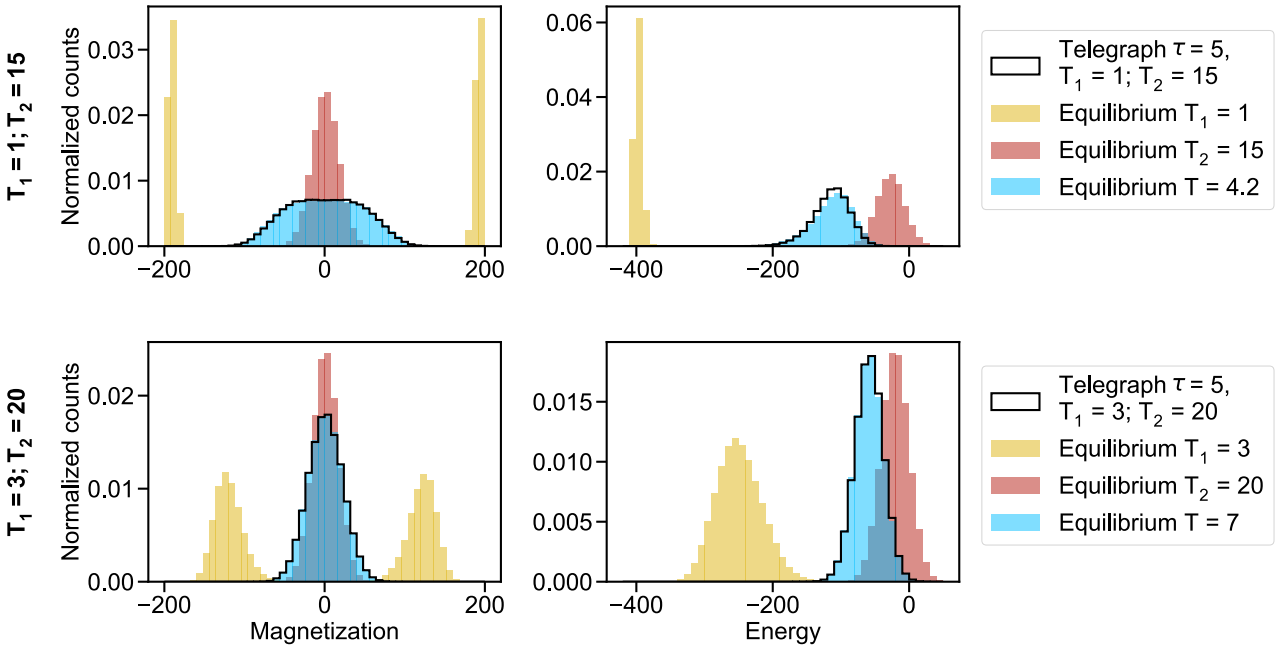

FIG. S8. **Characterization of sequences subjected to fluctuating selection strength.** Histograms of the magnetization (left panels) and statistical energy (right panels) of each sequence are shown for sequences generated at equilibrium and for sequences subjected to fluctuating selection strength (i.e. sampling temperature), modeled by a telegraph process with timescale τ. All sequences are generated using the minimal model explained in Fig. S1. In each case, 2048 sequences are generated. Those subjected to fluctuating selection strength are each generated starting from a different ancestor, but always using the same realization of the telegraph noise, with $\tau = 5$, for a total duration such that 5000 mutations are accepted, using the same sequence length of 200 and the same Erdős-Rényi random graph with probability 0.02 representing contacts as in Fig. 1. Two different pairs of temperatures (T_1, T_2) are considered for the telegraph process (top and bottom panels).

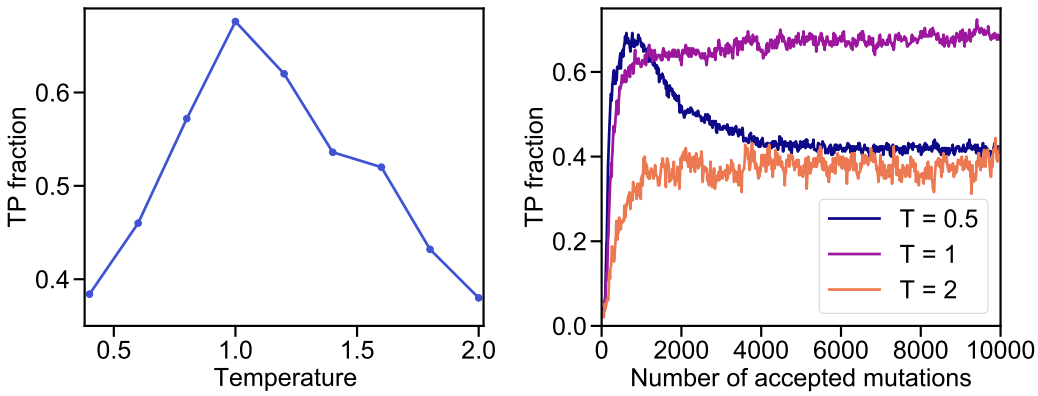

FIG. S9. **Impact of sampling temperature on structural contact prediction: PF0004 family.** Left panel: As in Fig. S4(a), the TP fraction for contact prediction is shown versus the sampling temperature, but for realistic data generated from a Potts model inferred on the PF0004 family. Right panel: The TP fraction for contact prediction is shown versus the number of accepted mutations, at different sampling temperatures, starting from random sequences. In both panels, in each case, we consider MSAs of 70,000 sequences generated from a Potts model inferred on a natural MSA of 39,277 sequences from the PF0004 family (with length 132), following Refs. [5, 6]. In the left panel, data obtained after 10000 accepted mutations is used at each temperature to infer contacts. This realistic data is thus similar to that considered in Fig. 2, except that here, we focus on data equilibrated at constant temperature, while in Fig. 2, we considered data generated under switching selection strength. As in Fig. 2, contact inference is performed using plmDCA [7, 8], with regularization strengths set to 0.01 and no phylogenetic reweighting.

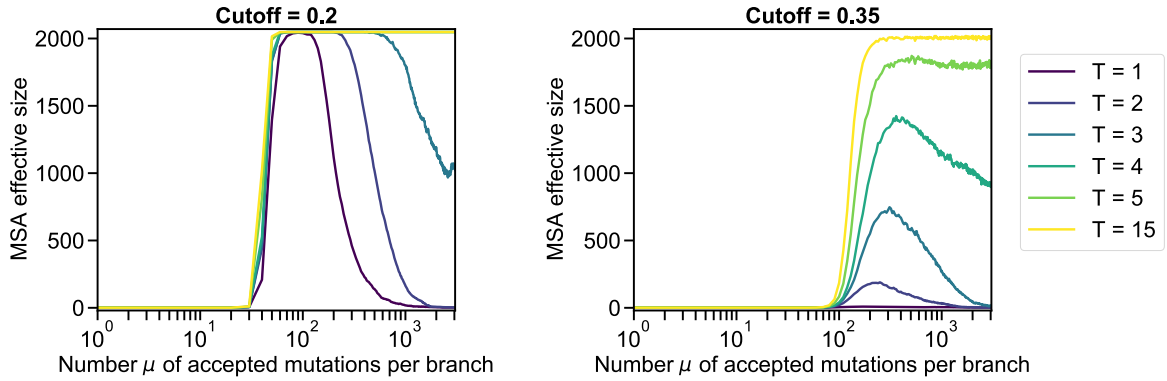

FIG. S10. **Effective size of MSAs in the star phylogeny where selection is switched on.** We report the effective size of the MSA versus the number μ of mutations per branch for the datasets corresponding to Fig. 3, where we considered a star phylogeny starting from a random ancestral sequence. We compute the effective depth [9] of each MSA as $M_{\text{eff}} = \sum_{i=1}^{M} w_i$, where for each sequence i, w_i is the inverse of the number of sequences such that their Hamming distance to sequence i is smaller than or equal to a cutoff value, which we set to 0.2 or to 0.35 in each panel. Recall that the actual number of sequences in each MSA is 2048, implying that $M_{\text{eff}} \leq 2048$.

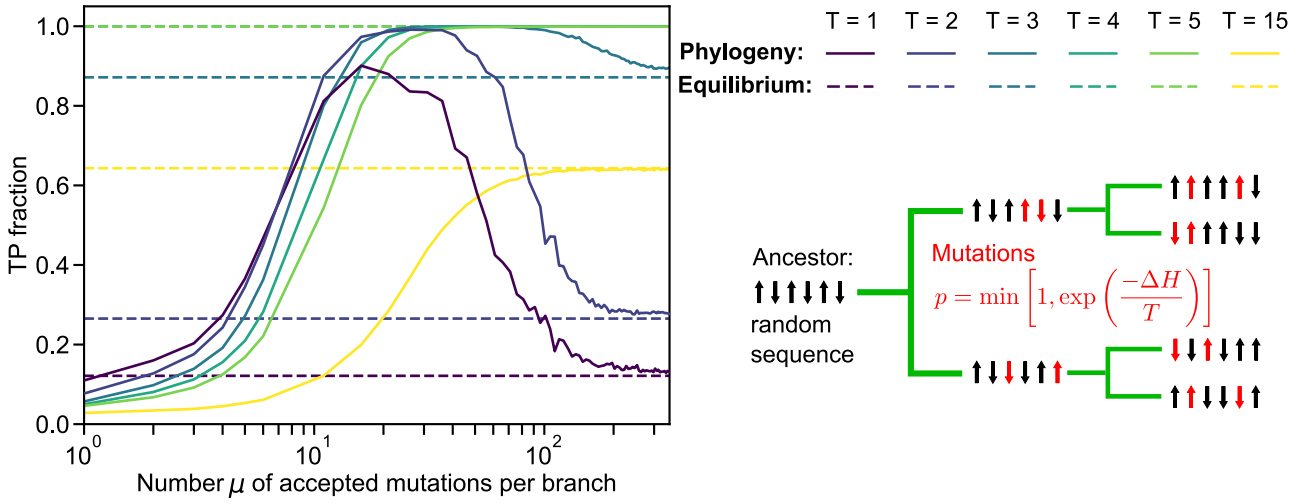

FIG. S11. **Impact of switching selection on on contact inference using a binary branching tree.** As in Fig. 3, the TP fraction is shown versus the number μ of accepted mutations per branch of a phylogenetic tree, starting from a random ancestral sequence. However, here, we use a binary branching tree (see schematic on the right) instead of a star phylogenetic tree. Sequences are generated as explained in Fig. S1(a) and in the schematic on the right, using the same Erdős-Rényi graph as in Fig. 1. The phylogenetic tree comprises 11 generations of binary branching events, resulting in 2048 sequences at the leaves, which constitute our MSA, and μ mutations are accepted on each branch. We infer contacts via mfDCA [2–4] with pseudocount 0.01. Results are averaged over 100 replicates, each starting from a different random ancestor. Schematic (right): To minimally model the impact of selection being switched on, we take a random sequence as the ancestor, and we evolve it along a perfectly balanced binary branching tree (green) where mutations (red) are accepted with the probability in Eq. (2), yielding selection for structural contacts. On each branch of the tree, Monte Carlo sampling is performed until μ mutations are accepted, with $\mu = 2$ in our schematic. Hence, all branches in this simple tree have the same length.

[1] W. Krauth, *Statistical Mechanics: Algorithms and Computations* (Oxford University Press, Oxford Master Series in Physics, 2006).

[2] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, Protein 3D structure computed from evolutionary sequence variation, PLoS ONE **6**, e28766 (2011).

[3] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, Proc. Natl. Acad. Sci. U.S.A. **108**, E1293 (2011).

[4] N. Dietler, U. Lupo, and A.-F. Bitbol, Impact of phylogeny on structural contact inference from protein sequence data, J. R. Soc. Interface **20**, 20220707 (2023).

[5] U. Lupo, D. Sgarbossa, and A.-F. Bitbol, Protein language models trained on multiple sequence alignments learn phylogenetic relationships, Nat. Commun. **13**, 6298 (2022).

[6] M. Figliuzzi, P. Barrat-Charlaix, and M. Weigt, How Pairwise Coevolutionary Models Capture the Collective Residue Variability in Proteins?, Mol. Biol. Evol. **35**, 1018 (2018).

[7] M. Ekeberg, C. Lovkvist, Y. Lan, M. Weigt, and E. Aurell, Improved contact prediction in proteins: using pseudolikelihoods to infer Potts models, Phys. Rev. E **87**, 012707 (2013).

[8] M. Ekeberg, T. Hartonen, and E. Aurell, Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences, J. Comput. Phys. **276**, 341 (2014).

[9] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing, Proc. Natl. Acad. Sci. U.S.A. **106**, 67 (2009).