Inverse Ising inference by combining Ornstein-Zernike theory with deep learning

Alpha A. Lee*
School of Engineering and Applied Sciences and Kavli Institute of Bionano Science and Technology, Harvard University, Cambridge, MA 02138, USA

Inferring a generative model from data is a fundamental problem in machine learning. It is well-known that the Ising model is the maximum entropy model for binary variables which reproduces the sample mean and pairwise correlations. Learning the parameters of the Ising model from data is the challenge. We establish an analogy between the inverse Ising problem and the Ornstein-Zernike formalism in liquid state physics. Rather than analytically deriving the closure relation, we use a deep neural network to learn the closure from simulations of the Ising model. We show, using simulations as well as biochemical datasets, that the deep neural network model outperforms systematic field-theoretic expansions and can generalize well beyond the parameter regime of the training data. The neural network is able to learn from synthetic data, which can be generated with relative ease, to give accurate predictions on real world datasets.

Drawing meaningful interference from correlations amongst variables is a fundamental problem in science. The central challenge is to decipher the probability distribution that generates the correlations given a set of observations, and then predict properties of unknown samples with the inferred model. Many probabilistic models have been proposed in the literature [1, 2]. Focusing on capturing the sample mean and pairwise correlations, the simplest model, in the sense of maximum entropy, is a Boltzmann probability distribution with a Hamiltonian that contains terms linear and bilinear in the variables [3]. For continuous variables, this model amounts to a multivariate Gaussian distribution; if the variables are discrete, this model is precisely the inverse problem of the Ising model in statistical physics.

The inverse Ising model, also known in the machine learning literature as Boltzmann learning [4], has found applications in numerous scientific disciplines [5] such as understanding neural spike trains [6, 7], flocking behaviour of birds [8], and predicting the structure of proteins [9–11] and RNA [12] as well as their fitness landscapes [13–16] using evolutionary data. Applications outside natural sciences include understanding the US Supreme Court [17] and financial markets [18–21]. The inverse Ising model disentangles correlations in the dataset into pairwise interactions, and thus could distinguish the direct influence of a variable on another variable from indirect influence mediated through other variables.

However, exact maximum likelihood inference of the inverse Ising model is numerically challenging. Although the likelihood function is convex, maximizing it requires computing the partition function of an Ising model for each step of the optimisation algorithm [4]. To overcome this challenge, many approximate inference techniques have been developed [22]. Those approximate techniques are typically leading terms of asymptotic expansions that relate the sample mean and correlation to the Ising parameters in idealized limits where the coupling is tractable. Those limits include asymptotically small variable-variable correlations [23–25], small number of coupled clusters of variables [26–28], tree-like structure in the variable-variable interactions [29] etc. (For comprehensive reviews, see [5, 30].) A relatively recent class of algorithms maximize the pseudolikelihood rather than the likelihood [31, 32]. Although the pseudolikelihood method avoids partition function evaluations and converges to the maximum likelihood solution in the limit of infinite data, the algorithmic complexity scales with the number of data.

In this Letter, we eschew analytical asymptotic expansions and instead use a deep learning model to infer the relationship between the Ising parameters and sample mean and correlation. We will first motivate an analogy between the inverse Ising model and the Ornstein-Zernike formalism in liquid state physics. We show that a deep neural network, trained using simulations of Ising models, can approximate the closure to the Ornstein-Zernike equation for the inverse Ising model. The deep neural network achieves an accuracy beyond systematic field-theoretic expansions for biochemical datasets as well as simulations with disparately different parameters compared to the training data, showing that the inferred closure is generalizable. The neural network learns from synthetic data, which can be generated with relative ease, to give accurate predictions on real world datasets with a runtime that is independent of the dataset size.

We begin by stating the inverse Ising model: We are given N sequences of p variables, $\{\mathbf{s}^\alpha\}_{\alpha=1}^N$, each variable s_i can take values ± 1. We seek parameters $\{h_i, J_{ij}\}$ such that the probability distribution defined for $\boldsymbol{\sigma} \in \{-1, 1\}^p$

$$p(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\left(\sum_i h_i \sigma_i + \sum_{i>j} J_{ij} \sigma_i \sigma_j \right) \qquad (1)$$

has the property that the first two moments agree with the first two moments of the variables estimated from the data, i.e.

$$\langle \sigma_i \rangle_p = \langle s_i \rangle_{\text{data}}, \quad \langle \sigma_i \sigma_j \rangle_p = \langle s_i s_j \rangle_{\text{data}}. \qquad (2)$$

Inferring $\{h_i, J_{ij}\}$ is challenging because the correlation between variables i and j, $C_{ij} = \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle$, can be large even if $J_{ij} = 0$ when there is an intervening variable k such that J_{ik} and J_{kj} are large. In other words, the pairwise correlations C_{ij} one detect in a dataset is a measure of the interaction between i and j that is mediated by all other variables.

To make further progress, we follow the Ornstein-Zernike formalism [33, 34] in liquid state physics to deconvolve direct interactions and indirect correlations in the inverse Ising model. We will first review the formalism by discussing the simple case of a one-component, homogenous and isotropic liquid. Molecules interact with a pairwise additive potential $v(\mathbf{r}_{12})$, where $\mathbf{r}_{12} = |\mathbf{r}_1 - \mathbf{r}_2|$ is the distance between particles 1 and 2. The liquid structure is characterised by the radial distribution function, $g(\mathbf{r}_{12})$, which the probability of observing a molecule located at distance $\mathbf{r}_{12}$ away from the molecule at the origin. Ornstein and Zernike noticed that $g(\mathbf{r}_{12})$ can be long ranged even when $v(\mathbf{r}_{12})$ is short ranged because $g(\mathbf{r}_{12})$ accounts for both the direct interaction between two molecules as well as the indirect interactions due to the molecule interacting with surrounding molecules. Their crucial insight is to introduce a quantity known as the direct correlation function, $c(\mathbf{r}_{12})$, and write the total correlation function $h(\mathbf{r}_{12}) \equiv g(\mathbf{r}_{12}) - 1$ as

$$h(\mathbf{r}_{12}) = c(\mathbf{r}_{12}) + \int \mathrm{d}\mathbf{r}_3 \; c(\mathbf{r}_{13}) c(\mathbf{r}_{32}) + \int \mathrm{d}\mathbf{r}_3 \mathrm{d}\mathbf{r}_4 \; c(\mathbf{r}_{13}) c(\mathbf{r}_{34}) c(\mathbf{r}_{42}) + \cdots \tag{3}$$

where the first term captures the direct influence of molecule 1 on molecule 2, the second term captures the influence of molecule 1 on molecule 2 that is mediated by molecule 3, and higher order terms capture correlations that can be attributed to more intervening molecules. Equation (3) can be rewritten in a compact form

$$h(\mathbf{r}_{12}) = c(\mathbf{r}_{12}) + \int \mathrm{d}\mathbf{r}_3 \; c(\mathbf{r}_{13}) h(\mathbf{r}_{32}), \tag{4}$$

which is known as the Ornstein-Zernike equation (we have scaled the classic direct and total correlation functions by the local density to make the link with the Ising model more apparent later). To close the problem, we need a relation between $c(\mathbf{r}_{12})$, $h(\mathbf{r}_{12})$, and $v(\mathbf{r}_{12})$. Many closure relationships have been developed in the literature using diagrammatic expansions. The crucial feature of many closures is that they are *local* and take the form

$$f(c(\mathbf{r}_{12}), h(\mathbf{r}_{12}), v(\mathbf{r}_{12}); \rho) = 0 \tag{5}$$

where f is a real-valued function (*not* a functional) and ρ is the density of the liquid [34]. For example, the widely used Hypernetted-chain approximation [35] reads $c(\mathbf{r}_{12})/\rho = h(\mathbf{r}_{12})/\rho - \log(h(\mathbf{r}_{12})/\rho + 1) - v(\mathbf{r}_{12})$. Although the locality of the closure relationship is not exact [36], decades of comparing liquid state theory predictions with simulations have shown that it is a very reasonable approximation for a wide variety of inter-particle interactions. We note that the homogeneity of the fluid is not an essential assumption behind Equation (4). The salient idea is modelling the total correlation as a sum of convolutions of direct correlations.

Having introduced the Ornstein-Zernike formalism and closure, we return to the inverse Ising problem. The sample correlation, C_{ij}, is a discrete analogy of the total correlation function $h(\mathbf{r}_{ij})$ in the Ornstein-Zernike formalism. Therefore, we introduce the direct correlation D_{ij} – the discrete analogue of the direct correlation function $c(\mathbf{r}_{ij})$ – and replace the integrals in Equation (4) as a sum over lattice sites, i.e.

$$C_{ij} = \delta_{ij} + D_{ij} + \sum_k D_{ik} D_{kj} + \sum_{k,l} D_{ik} D_{kl} D_{lj} + \cdots . \tag{6}$$

In matrix form, $C = (\mathbb{I} - D)^{-1}$ or

$$D = \mathbb{I} - C^{-1}. \tag{7}$$

Taking $J = -C^{-1}$ is known as the mean-field approximation or Direct Coupling Analysis [10, 37].

Now we introduce the key assumption of this Letter: we posit that the locality heuristic of closure relations for liquids applies to the inverse Ising problem. This crucial assumption will be verified later by comparing with simulations. We assume that there exists a function of four real variables $F(\cdot, \cdot, \cdot, \cdot)$ such that

$$J_{ij} \approx F\left(C_{ij}, [C^{-1}]_{ij}, \langle s_i \rangle, \langle s_j \rangle\right). \tag{8}$$

Unlike the case of homogeneous fluids, we need an extra equation to determine the fields h_i. Motivated by Percus' expression of the free energy of inhomogeneous fluids in terms of the direct correlation function [38], we posit the existence of a function $G(\cdot, \cdot, \cdot, \cdot)$ such that

$$h_i \approx G\left(\tanh^{-1}\langle s_i \rangle, [C^{-1}]_{ii}, \sum_{j \neq i} J_{ij} \langle s_j \rangle, \sum_{j \neq i} C_{ij} \langle s_j \rangle\right), \tag{9}$$

where J_{ij} is given by F. As an extra check, surveying a recent review [5], the first few terms of most approximate analytical theories of the inverse Ising model do take the form of Equations (8)-(9) with different F and G depending on the approximations.

F and G are complicated functions. Rather than attempting to analytically determine what they are, we will use a deep learning approach and approximate F and G by a multilayer neural network, a richly parameterised interpolating function. Our approach is inspired by phenomenological closures in liquid state theory based on interpolating between rigorously derived closures [39, 40], although a neural network offers far greater flexibility

[41] and therefore accuracy if it is trained with sufficient amount of data.

The training data is prepared by running 20 simulations with $p = 50$ Ising sites. In each simulation, the Ising parameters are chosen from a normal distribution, $J_{ij} = J_{ji} \sim \mathcal{N}(0, \beta_J/\sqrt{p})$ and $h_i \sim \mathcal{N}(0, \beta_h)$, i.e. the Sherrington-Kirkpatrick model [42] with a random field. The variances of the distributions are fixed in each simulation but vary across the 20 simulations, with $\beta_J \sim$ uniform$(0.5, 2.5)$ and $\beta_h \sim$ uniform$(0.5, 2.5)$. The one point and two point correlations are computed from 5×10^8 steps of Markov Chain Monte Carlo (MCMC), sampled at every 500 steps. We approximate F using a three layer neural network, with 9, 7 and 5 rectified linear units (`ReLU`) in the first, second and third layer. G is approximated using a two layer neural network with 2 `tanh` units in both layers. The neural network architecture is determined by randomly holding out 10% of training data and testing the accuracy of neural networks with less than three layers and 10 `ReLU` or `tanh` units per layer. We implement the neural network in `SciKit.Learn` [43] using the function `MLPRegressor` and the `LBFGS` algorithm to optimize hyperparameters. Henceforth we will use the same neural network for the rest of this Letter.

To test the generality and accuracy of the model, we simulate Ising models with $p = 70$ sites (note that the model is only trained on $p = 50$ sites), $J_{ij} = J_{ji} \sim \mathcal{N}(0, \beta/\sqrt{p})$ and $h_i \sim \mathcal{N}(0, 0.3\beta)$. A large value of β corresponds to stronger coupling thus further away from the perturbative regimes that underlie all analytical theories. Figure 1A-B shows the root mean square error of recovering J_{ij} and h_i as a function of β. The neural network model is more accurate than the Thouless-Anderson-Palmer approximation (TAP), a third order high temperature expansion [44]. Importantly, the neural network is accurate even for $\beta \in (2.5, 3.5)$, which is outside the coupling strengths in the training data, demonstrating generalizability. The neural network approximation is also robust to sampling noise. Figure 1C-D shows that our neural network still outperforms the TAP approximation when C_{ij} and $\langle s_i \rangle$ are estimated from less MCMC data.

The neural network approximation is generalizable and accurate even when the couplings are non-Gaussian distributed. Figure 2A shows that the neural network approximation can accurately recover the coupling parameter for a one-dimensional ferromagnetic Ising model with constant nearest-neighbour coupling $J_{ij} = J_{ji} = J(\delta_{i,j+1} + \delta_{i+1,j})$. The coupling matrix recovered from the neural network (Figure 2B) is strongly localised on the off-diagonal elements (although slightly underestimating J) despite all the training data has a coupling matrix that is delocalised. Inspired by the use of inverse Ising models in protein structure and fitness prediction [10, 15, 16], Figure 2C shows that the neural network approximation can recover a coupling matrix that has

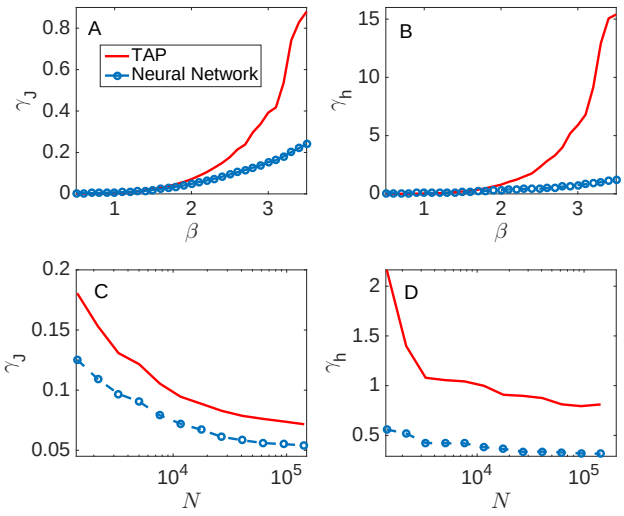

FIG. 1. The neural network model is accurate, generalisable and robust to sampling noise. The RMS error, γ, of predicting (A) J_{ij} and (B) h_i as a function of "inverse temperature" β. (C)-(D) The RMS error for predicting J_{ij} and h_i as a function of N, the number of MCMC samples used to estimate C_{ij} and $\langle s_i \rangle$, for $\beta = 2$.

the structure of a protein contact map. The neural network crucially recovers the pattern of large and small entries of J (Figure 2D), i.e. the contact map. In Figure 2C-D, we consider the Bovine pancreatic trypsin inhibitor protein (PDB ID: 5PTI), which is a benchmark example in ref [45] and comprises $p = 58$ amino acids. For this example, we assume that we know the "ground truth" couplings (to what extent are amino acid interactions pairwise additive is a separate question [46]), and we take $J_{ij} = e^{-d_{ij}/7\mathring{A}}/\sqrt{p}$, where the d_{ij} is the distance between the C_β atoms of residues i and j and the decay length $7\mathring{A}$ is motivated by the typical distance criterion of a residue-residue contact. The correlation matrix is computed using MCMC sampling with $p \times 10^7$ steps [47].

The neural network remains accurate for non-Gaussian distributed couplings thanks to the approximate locality of the Ornstein-Zernike closure. The functions F and G is independent of the J_{ij} distribution and can be accurately approximated as long as the training data spans the four-dimensional input space. Therefore, we can solve the inverse Ising models for a general coupling matrix by training the neural network only with data from Gaussian coupling matrices. We will go beyond synthetic data where the ground truth is known and apply our method to two problems in computational biology and chemistry.

Fitness landscape of HIV-1 Gag: Developing a predictive model for the fitness of HIV virus as a function of the amino acid sequence of the Group-specific antigen (Gag) protein is a challenge in HIV vaccine development. Recent works have shown that the fitness landscape can

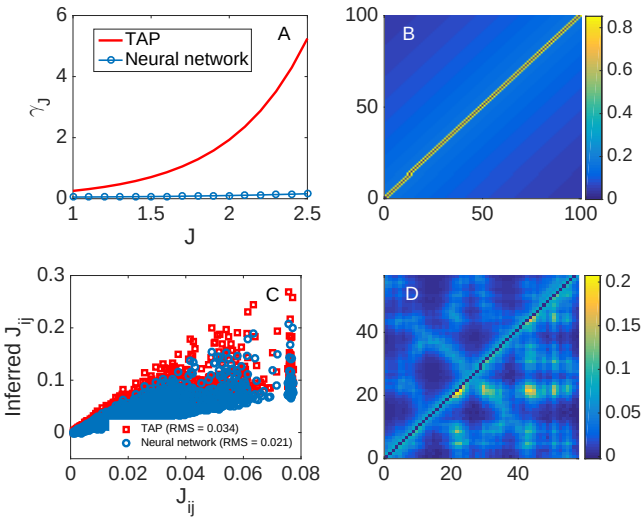

FIG. 2. The neural network model is accurate even for very non-Gaussian coupling matrices. (A) The root mean square error, γ, of inferring J for the nearest neighbour Ising model. (B) The inferred coupling matrix for $J = 2.5$. (C) The actual entries of the coupling matrix, J_{ij}, against the inferred J_{ij} for a coupling matrix that has the structure of a protein contact map. (D) The top left corner shows the entries of the actual coupling matrix and the bottom right shows the entries of the inferred coupling matrix.

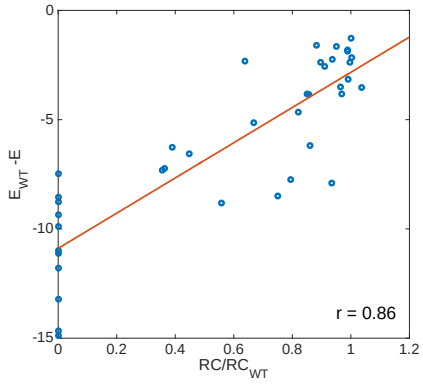

FIG. 3. The log probability of a mutant, E, is strongly correlated with the replication capacity RC. The subscript WT denotes wild type. Details of sequences and experiments can be found in ref [49]. A pseudocount of 0.2 is used to correct for undersampling [50], and the largest eigenvector of the correlation matrix, an artefact of amino acid conservation, is projected out before being processed by the neural network.

be inferred from the statistics of sequences found in patients [14]. The hypothesis, loosely speaking, is that the frequency of observing conserved sites and sets of correlated mutations reflect the contribution of those residues to fitness [48]. Therefore, the fitness of an unknown sequence can be predicted by the log probability of that sequence computed using a generative model for the sequences observed in patients. We replicate the analysis of ref [49] for the HIV-1 Gag protein using the Ising representation for sequences, except the Ising parameters are inferred using the neural network. Figure 3 shows that the log probability predicted by our model is highly correlated with experimental measurements of replication capacity, with a correlation coefficient $r = 0.86$, slightly higher than the state-of-the-art model [49] ($r = 0.83$).

Predicting binding to estrogen receptor α: A key challenge in drug discovery and computational toxicity is predicting whether an unknown molecule will bind to a particular receptor. We consider a high throughput screening study of agonists of estrogen receptor α (PubChem AID: 743079), a key receptor in the endocrine system which is a target of the Tox21 computational toxicology initiative [51, 52]. Each molecule in the dataset is converted to a string of ± 1 by the 2048-bit Morgan6 Fingerprint [53] using the package `rdKit` [54]. The Morgan fingerprint records the presence (1)/absence (-1) of chemical groups. A separate Ising model is inferred for the set of molecules that binds ($\{J_{ij}^{\rm b}, h_i^{\rm b}\}$) and does not bind ($\{J_{ij}^{\rm n-b}, h_i^{\rm n-b}\}$). The high throughput screening data poses a challenge as the number of variables is significantly more than the number of samples. To remove sampling noise, we clean the correlation matrix using techniques from random matrix theory originally developed in finance [55] but recently applied to chemoinformatics [56, 57] – only eigenvectors with statistically significant eigenvalues are kept, and the rest are "clipped" , i.e.

$$\tilde{C} = \sum_{i=1}^{p} \lambda_i \theta(\lambda_{\rm MP} - \lambda_i) \mathbf{e}_i^T \mathbf{e}_i + \frac{\mathbb{I}}{p} \sum_{i=1}^{p} \lambda_i \theta(\lambda_i - \lambda_{\rm MP}) \quad (10)$$

where $\theta(x)$ is the Heaviside function, $\mathbf{e}_i$ (λ_i) the eigenvector (eigenvalue) of the original correlation matrix and $\lambda_{\rm MP} = (1 + \sqrt{p/N})^2$ the cutoff determined by random matrix theory; $\tilde{C}$ is then converted back to the covariance matrix by scaling with the standard deviations of the variables. We score a molecule $\mathbf{f}$ by the log probability ratio

$$E(\mathbf{f}) = \sum_{i<j} f_i f_j (J_{ij}^{\rm b} - J_{ij}^{\rm n-b}) + \sum_i f_i (h_i^{\rm b} - h_i^{\rm n-b}), \quad (11)$$

and the molecule is predicted to bind if $E < \epsilon$, were ϵ is a parameter that controls the tradeoff between false positive and true positive. Figure 4 shows that the inverse Ising model accurately predicts binding (out-of-sample AUC = 0.75); for reference the TAP method achieves AUC of 0.68 and a support vector machine model achieves AUC of 0.70 [58].

In conclusion, we demonstrate a method that combines Ornstein-Zernike theory with a highly accurate closure parameterised using deep learning to solve the inverse Ising problem. We illustrate how our method can be used in real world datasets by considering examples in

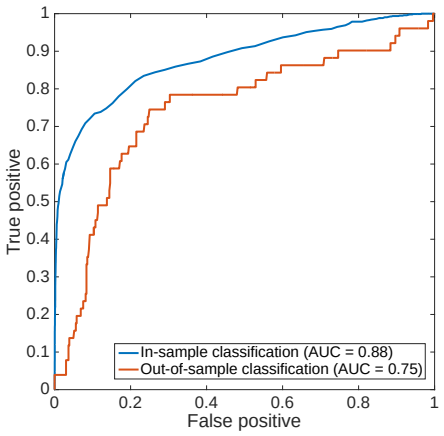

FIG. 4. The inverse Ising model is an accurate classifier for whether a molecule binds to the estrogen receptor α. The training and verification sets are taken from the Tox21 challenge [51, 52].

computational biology and chemoinformatics. We anticipate the strategy of parametrising Ornstein-Zernike closure with data to be useful not only for inverse Ising models but also in liquid state physics, predication of solvation thermodynamics, and the inverse design of liquid structure [59, 60].

I thank M. P. Brenner and R. Monasson for insightful discussions and comments, and J. P. Barton for making available the data of Ref [49]. I am supported by the George F. Carrier Fellowship.

* alphalee@g.harvard.edu

[1] Y. Bengio, A. Courville, and P. Vincent, IEEE transactions on pattern analysis and machine intelligence **35**, 1798 (2013).
[2] R. Salakhutdinov, Annual Review of Statistics and Its Application **2**, 361 (2015).
[3] E. T. Jaynes, Physical review **106**, 620 (1957).
[4] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognitive Science **9**, 147 (1985).
[5] H. C. Nguyen, R. Zecchina, and J. Berg, arXiv preprint arXiv:1702.01522 (2017).
[6] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature **440**, 1007 (2006).
[7] S. Cocco, S. Leibler, and R. Monasson, Proceedings of the National Academy of Sciences **106**, 14058 (2009).
[8] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Proceedings of the National Academy of Sciences **109**, 4786 (2012).
[9] A. Lapedes, B. Giraud, and C. Jarzynski, arXiv preprint arXiv:1207.2484 (2012).
[10] D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina, and C. Sander, PloS one **6**, e28766 (2011).
[11] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Proceedings of the National Academy of Sciences **108**, E1293 (2011).
[12] E. De Leonardis, B. Lutz, S. Ratz, S. Cocco, R. Monasson, A. Schug, and M. Weigt, Nucleic Acids Research **43**, 10444 (2015).
[13] K. Shekhar, C. F. Ruberman, A. L. Ferguson, J. P. Barton, M. Kardar, and A. K. Chakraborty, Physical Review E **88**, 062705 (2013).
[14] A. L. Ferguson, J. K. Mann, S. Omarjee, T. Ndungu, B. D. Walker, and A. K. Chakraborty, Immunity **38**, 606 (2013).
[15] M. Figliuzzi, H. Jacquier, A. Schug, O. Tenaillon, and M. Weigt, Molecular Biology and Evolution **33**, 268 (2015).
[16] T. A. Hopf, J. B. Ingraham, F. J. Poelwijk, M. Springer, C. Sander, and D. S. Marks, Nature Biotechnology **35**, 128 (2017).
[17] E. D. Lee, C. P. Broedersz, and W. Bialek, Journal of Statistical Physics **160**, 275 (2015).
[18] J.-i. Maskawa, Physica A: Statistical Mechanics and its Applications **311**, 563 (2002).
[19] T. Bury, Journal of Statistical Mechanics: Theory and Experiment **2013**, P11004 (2013).
[20] H.-L. Zeng, R. Lemoy, and M. Alava, Journal of Statistical Mechanics: Theory and Experiment , P07008 (2014).
[21] S. S. Borysov, Y. Roudi, and A. V. Balatsky, The European Physical Journal B **88**, 1 (2015).
[22] M. Opper and D. Saad, *Advanced mean field methods: Theory and practice* (MIT press, 2001).
[23] H. J. Kappen and F. d. B. Rodríguez, Neural Computation **10**, 1137 (1998).
[24] T. Tanaka, Physical Review E **58**, 2302 (1998).
[25] V. Sessak and R. Monasson, Journal of Physics A: Mathematical and Theoretical **42**, 055001 (2009).
[26] S. Cocco and R. Monasson, Physical Review Letters **106**, 090601 (2011).
[27] S. Cocco and R. Monasson, Journal of Statistical Physics **147**, 252 (2012).
[28] J. P. Barton, E. De Leonardis, A. Coucke, and S. Cocco, Bioinformatics **32**, 3089 (2016).
[29] H. C. Nguyen and J. Berg, Journal of Statistical Mechanics: Theory and Experiment **2012**, P03004 (2012).
[30] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, and M. Weigt, arXiv preprint arXiv:1703.01222 (2017).
[31] E. Aurell and M. Ekeberg, Physical Review Letters **108**, 090201 (2012).
[32] A. Decelle and F. Ricci-Tersenghi, Physical Review Letters **112**, 070603 (2014).
[33] L. S. Ornstein and F. Zernike, in *Proc. Acad. Sci. Amsterdam*, Vol. 17 (1914) pp. 793–806.
[34] J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, 2013).
[35] J. Van Leeuwen, J. Groeneveld, and J. De Boer, Physica **25**, 792 (1959).
[36] R. Fantoni and G. Pastore, The Journal of chemical physics **120**, 10681 (2004).
[37] A. R. Kinjo, Biophysics and Physicobiology **12**, 117 (2015).
[38] J. Percus, Journal of statistical physics **52**, 1157 (1988).
[39] F. J. Rogers and D. A. Young, Physical Review A **30**, 999 (1984).
[40] G. Zerah and J.-P. Hansen, Journal of Chemical Physics **84**, 2336 (1986).

[41] B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, in *Advances in Neural Information Processing Systems* (2016) pp. 3360–3368.

[42] D. Sherrington and S. Kirkpatrick, Physical Review Letters **35**, 1792 (1975).

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, Journal of Machine Learning Research **12**, 2825 (2011).

[44] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philosophical Magazine **35**, 593 (1977).

[45] S. Cocco, R. Monasson, and M. Weigt, PLoS Comput Biol **9**, e1003176 (2013).

[46] H. Jacquin, A. Gilson, E. Shakhnovich, S. Cocco, and R. Monasson, PLoS Comput Biol **12**, e1004889 (2016).

[47] The largest eigenvector of the correlation matrix, which reflects different thermodynamic states [61], is projected out.

[48] A. K. Chakraborty and J. P. Barton, Reports on Progress in Physics **80**, 032601 (2017).

[49] J. K. Mann, J. P. Barton, A. L. Ferguson, S. Omarjee, B. D. Walker, A. Chakraborty, and T. Ndung'u, PLoS Comput Biol **10**, e1003776 (2014).

[50] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis: probabilistic models of proteins and nucleic acids* (Cambridge University Press, 1998).

[51] R. Huang, M. Xia, S. Sakamuru, J. Zhao, S. A. Shahane, M. Attene-Ramos, T. Zhao, C. P. Austin, and A. Simeonov, Nature Communications **7** (2016).

[52] R. Huang and M. Xia, Frontiers in Environmental Science **5**, 3 (2017).

[53] D. Rogers and M. Hahn, Journal of Chemical Information and Modeling **50**, 742 (2010).

[54] "RDKit: Open-source cheminformatics," `http://www.rdkit.org`.

[55] J.-P. Bouchaud and M. Potters, in *The Oxford Handbook of Random Matrix Theory* (Oxford University Press, 2011).

[56] A. A. Lee, M. P. Brenner, and L. J. Colwell, Proceedings of the National Academy of Sciences **113**, 13564 (2016).

[57] A. A. Lee, M. P. Brenner, and L. J. Colwell, arXiv preprint arXiv:1702.06001 (2017).

[58] A. Mayr, G. Klambauer, T. Unterthiner, and S. Hochreiter, Frontiers in Environmental Science **3**, 80 (2016).

[59] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, Soft Matter **12**, 2663 (2016).

[60] B. A. Lindquist, S. Dutta, R. B. Jadrich, D. J. Milliron, and T. M. Truskett, Soft Matter **13**, 1335 (2017).

[61] H. C. Nguyen and J. Berg, Physical Review Letters **109**, 050602 (2012).