Symmetry, gauge freedoms, and the interpretability of sequence-function relationships

Anna Posfai,[1] David M. McCandlish,[1] and Justin B. Kinney[1,*]

[1] *Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory*

(Dated: March 17, 2025)

Quantitative models that describe how biological sequences encode functional activities are ubiquitous in modern biology. One important aspect of these models is that they commonly exhibit gauge freedoms, i.e., directions in parameter space that do not affect model predictions. In physics, gauge freedoms arise when physical theories are formulated in ways that respect fundamental symmetries. However, the connections that gauge freedoms in models of sequence-function relationships have to the symmetries of sequence space have yet to be systematically studied. In this work we study the gauge freedoms of models that respect a specific symmetry of sequence space: the group of position-specific character permutations. We find that gauge freedoms arise when model parameters transform under redundant irreducible matrix representations of this group. Based on this finding, we describe an "embedding distillation" procedure that enables both analytic calculation of the number of independent gauge freedoms and efficient computation of a sparse basis for the space of gauge freedoms. We also study how parameter transformation behavior affects parameter interpretability. We find that in many (and possibly all) nontrivial models, the ability to interpret individual model parameters as quantifying intrinsic allelic effects requires that gauge freedoms be present. This finding establishes an incompatibility between two distinct notions of parameter interpretability. Our work thus advances the understanding of symmetries, gauge freedoms, and parameter interpretability in models of sequence-function relationships.

INTRODUCTION

Understanding the quantitative nature of sequence-function relationships is a major goal of modern biology [1]. To study a sequence-function relationship of interest, researchers often propose a mathematical model, fit the parameters of the model to data, then biologically interpret the resulting parameter values. This interpretation step is complicated, however, by gauge freedoms—directions in parameter space along which model parameters can be changed without altering model predictions. When gauge freedoms are present in a model, the values of individual model parameters cannot be meaningfully interpreted without additional constraints. In standard Potts models of proteins, for example, the values of the parameters representing interactions between amino acids cannot be directly interpreted as quantifying interaction strength. This is because gauge freedoms make it possible to change any specific coupling parameter of interest without affecting model predictions by also making appropriate compensatory changes to other model parameters [2–6].

Researchers who study sequence-function relationships using quantitative models routinely encounter gauge freedoms. In practice, one of two methods is used to overcome the difficulties that gauge freedoms present. One method, called "gauge fixing", removes gauge freedoms by introducing additional constraints on model parameters [2–18]. Another method limits the mathematical models that one uses to models that do not have any gauge freedoms in the first place [19–24]. Despite being frequently encountered in the course of research, the gauge freedoms present in models of sequence-function relationships have received only limited attention [e.g., 2, 4–6, 12, 25]. In particular, the mathematical properties of these gauge freedoms have yet to be systematically studied.

In physics, by contrast, gauge freedoms are well recognized as a topic of fundamental importance [26]. Gauge freedoms arise when a physical theory is expressed in a form that manifestly respects fundamental symmetries. For example, the classical theory of electricity and magnetism (E&M) is invariant to Lorentz transformations, i.e., relativistic changes in an observer's velocity [27]. Lorentz invariance is obscured, however, when the equations of E&M are expressed directly in terms of electric and magnetic fields. To express these equations in a form that is manifestly Lorentz invariant, one must instead formulate them in terms of an electromagnetic four-potential. Doing this introduces gauge freedoms because the four-potential, unlike electric and magnetic fields, is neither directly measurable nor uniquely determined by the configuration of a physical system.[28] Nevertheless, working with the four-potential simplifies the equations of E&M and can aid in both their solution and their physical interpretation.

Motivated by the connection between gauge freedoms and symmetries in physics, we asked whether gauge freedoms in models of sequence-function relationships have a connection to the symmetries of sequence space, i.e., the possible ways of transforming the space of sequences without altering the Hamming distances between sequences. In this work we study the gauge freedoms of linear models that are equivariant under a specific symmetry group of sequence space—the group of position-specific character permutations (PSCP). Here, "linear models" are models that can be expressed as a sum of

* jkinney@cshl.edu

sequence features, each multiplied by a corresponding parameter; "PSCP" encompasses transformations that permute the identities of the individual characters (e.g., DNA bases or protein amino acids) at one sequence position, as well as transformations built from combinations of such permutations; and "equivariant" describes models for which linear transformations of the model parameters are able to compensate for the effects of PSCP transformations of sequences. Equivariant linear models include many of the most commonly used models in the literature, such as models with pairwise and/or higher-order interactions.

Using techniques from the theory of matrix representations of the symmetric group [29], we find that the gauge freedoms of these linear equivariant models arise when model parameters transform under redundant irreducible representations of PSCP. Based on this finding, we introduce an "embedding distillation" procedure that, for any linear equivariant model, facilitates both analytical and computational analyses of the vector space of gauge freedoms. We also study the connection between parameter interpretability and model transformation behavior. We find that in many (and possibly all) nontrivial models, the ability to interpret model parameters as quantifying the intrinsic effects of alleles requires that the model have gauge freedoms. This finding shows that models having gauge freedoms can have important advantages over models that have no gauge freedoms.

A companion paper [30] reports specific gauge-fixing strategies that can be applied to an important subset of the linear equivariant models, one that includes the most commonly used models of sequence-function relationships. It also describes specific ways of using these gauge-fixing strategies to assist in the development and biological interpretation of such models.

BACKGROUND

We now establish definitions and notation used in Results. We also review basic results regarding gauge freedoms in mathematical models of sequence-function relationships. Our companion paper [30] provides an expanded discussion of these results together with corresponding proofs.

Sequence-function relationships

Let $\mathcal{A}$ denote an alphabet comprising α distinct characters, let $\mathcal{S}$ denote the set of α^L sequences of length L built from these characters, and let $s_l \in \mathcal{A}$ denote the character at position l in any sequence $s \in \mathcal{S}$. A real-valued model of a sequence-function relationship, $f(s;\vec{\theta})$, is defined to be a function that maps each sequence s to a real number. The vector $\vec{\theta}$ denotes the parameters of the model and is assumed to comprise M real numbers. For technical reasons it is sometimes useful to consider complex-valued models of sequence-function relationships, which are defined analogously.

Linear models

Linear models of sequence-function relationships are linear in $\vec{\theta}$ and thus have the form

$$f(s;\vec{\theta}) = \vec{\theta}^\top \vec{x}(s) = \sum_{i=1}^{M} \theta_i x_i(s), \tag{1}$$

for all $s \in \mathcal{S}$. Here, $\vec{x}(\cdot)$ is an M-dimensional vector of sequence features, and each feature $x_i(\cdot)$ is a function that maps $\mathcal{S}$ to $\mathbb{R}$. We refer to the space $\mathbb{R}^M$ in which $\vec{x}$ and $\vec{\theta}$ live as feature space.[31]

An example of a linear model is the pairwise one-hot model, which has the form

$$f^{\,\text{ohe}}_{\text{pair}} = \theta_0 x_0 + \sum_{l} \sum_{c} \theta^c_l x^c_l + \sum_{l<l'} \sum_{c,c'} \theta^{cc'}_{ll'} x^{cc'}_{ll'}, \tag{2}$$

where the arguments of both the model and features have been kept implicit. In Eq. 2, $l, l' \in \{1, \ldots, L\}$ index the positions within each sequence, and $c, c' \in \mathcal{A}$ index the possible characters at these positions. We use the superscript "ohe" here and in what follows to indicate mathematical objects (such as embeddings, models, and representations) that are based on one-hot embeddings. Pairwise one-hot models, in particular, make use of the pairwise one-hot embedding $\vec{x}^{\,\text{ohe}}_{\text{pair}}(s)$, the elements of which represent three types of features: the constant feature, $x_0(s)$, which equals one for every sequence s; additive one-hot features, $x^c_l(s)$, which equal one if $s_l = c$ and equal zero otherwise; and pairwise one-hot features, $x^{cc'}_{ll'}(s)$, which equal one if both $s_l = c$ and $s_{l'} = c'$, and equal zero otherwise.

Gauge freedoms

Gauge freedoms are transformations of model parameters that do not affect model predictions. Formally, a gauge freedom is any vector $\vec{g} \in \mathbb{R}^M$ that satisfies

$$f(s;\vec{\theta}) = f(s;\vec{\theta} + \vec{g}) \quad \text{for all} \quad s \in \mathcal{S}. \tag{3}$$

For linear sequence-function relationships the set of gauge freedoms, denoted by G, forms a vector space in $\mathbb{R}^M$. It is readily shown that G is the orthogonal complement of the space spanned by sequence embeddings [30]. In what follows, we use γ to represent the dimension of G, i.e., the number of (independent) gauge freedoms.

Gauge freedoms arise from linear dependencies among sequence features. By inspection we see that $f^{\,\text{ohe}}_{\text{pair}}$ has

$$M^{\,\text{ohe}}_{\text{pair}} = 1 + \alpha L + \binom{L}{2}\alpha^2 \tag{4}$$

parameters. However, it turns out that the space spanned by the corresponding embedding $\vec{x}^{\,\text{ohe}}_{\text{pair}}$ has only $1 + (\alpha - 1)L + \binom{L}{2}(\alpha - 1)^2$ dimensions. This difference reflects the presence of $L + \binom{L}{2}(2\alpha - 1)$ constraints on the features, namely: $x_0 = \sum_{c'} x_l^{c'}$ for all positions l (yielding 1 constraint per position); and $x_l^c = \sum_{c'} x_{ll'}^{cc'}$, $x_{l'}^{c'} = \sum_c x_{ll'}^{cc'}$ for all pairs of positions $l < l'$ and all choices of character c or c' (yielding $2\alpha - 1$ independent constraints per pair of positions). The model $f^{\,\text{ohe}}_{\text{pair}}$ therefore has

$$\gamma^{\,\text{ohe}}_{\text{pair}} = L + \binom{L}{2}(2\alpha - 1) \tag{5}$$

gauge freedoms. See our companion paper [30] for more details, as well as [2, 4, 6, 10] for earlier treatments of gauge freedoms in the pairwise one-hot model.

Fixing the gauge

Fixing the gauge is the process of removing gauge freedoms by restricting $\vec{\theta}$ to a subset Θ of parameter space called "the gauge". For example, the commonly used "zero-sum gauge" [4, 6] for the pairwise one-hot model is the subspace of parameter space in which the additive parameters at every position sum to zero when marginalized over characters ($\sum_c \theta_l^c = 0$ for every l) and the pairwise parameters at all pairs of positions sum to zero when marginalized over the characters at either position ($\sum_c \theta_{ll'}^{cc'} = 0$ and $\sum_{c'} \theta_{ll'}^{cc'} = 0$ for every l, l', c, c').

Linear gauges are choices of Θ that are vector spaces. The zero-sum gauge is one such linear gauge. A useful property of linear gauges is that gauge-fixing can be accomplished through linear projection. Specifically, for any linear gauge Θ, there exists a projection matrix P that projects each parameter vector $\vec{\theta} \in \mathbb{R}^M$ onto an equivalent parameter vector $\vec{\theta}_{\text{fixed}} \in \Theta$ via $\vec{\theta}_{\text{fixed}} = P\,\vec{\theta}$. Our companion paper describes a parametric family of linear gauges (including an explicit formula for the corresponding projection matrices) that includes as special cases many of the most commonly used gauges in the literature [30].

RESULTS

We begin this section by formally defining the group of PSCP transformations, as well as the notion of model equivariance under this group. We then illustrate, for two example pairwise-interaction models, how transformation behavior under PSCP impacts both gauge freedoms and a specific type of parameter interpretability, namely the ability to assign intrinsic effects to individual alleles. Next we formally investigate this relationship more generally using methods from the theory of group representations [29]. In doing so, we establish an "embedding distillation" procedure that, for any equivariant model, enables analytic calculation of the number of gauge freedoms. We also establish an algorithm that enables the efficient computation of a sparse basis for the space of gauge freedoms. We conclude by revisiting the issue of parameter interpretability in light of these results.

Position-specific character permutations (PSCP)

Different transformations of sequence space impact models of sequence-function relationships in different ways. Here we focus on PSCP transformations. These transformations of sequence space form a mathematical group, which we denote by H_{PSCP}. The action of a transformation $h \in H_{\text{PSCP}}$ on a sequence $s \in \mathcal{S}$ is written hs. H_{PSCP} is a symmetry group of sequence space in that its transformations preserve the Hamming distances between sequences. There are other symmetry groups of sequence space as well, but we ultimately find that these symmetry groups do not have the same connections to gauge freedoms that H_{PSCP} does (discussed below and in SI Sec. 7).

Equivariance

We also focus on equivariant linear models of sequence-function relationships. These are models for which both embeddings and parameters transform linearly under H_{PSCP}. The specific sets of matrices that encode these linear transformations are called "representations" [29]. In general, a representation R of a group H is a function that maps each $h \in H$ to a matrix $R(h)$ in a way that preserves the multiplicative structure of H, i.e., $R(h_1 h_2) = R(h_1)R(h_2)$ for any two group elements $h_1, h_2 \in H$. The degree of the representation R (denoted $\deg R$) is the dimension of the vector space on which R acts. Two different examples of representations for the same group are described and illustrated below (see Eqs. 8 and 11, as well as Fig. 1)

Formally, we say that an embedding $\vec{x}$ is equivariant in H if and only if there is a representation R of H such that

$$\vec{x}(hs) = R(h)\,\vec{x}(s) \tag{6}$$

for all $h \in H$ and all $s \in \mathcal{S}$. We also say that a model is equivariant if and only if it has an equivariant embedding. For an equivariant model whose embedding transforms as in Eq. 6, the transformation of $\mathcal{S}$ by any $h \in H$ can be compensated for by the transformation of $\vec{\theta}$ by $R(h)^{-1\top}$, in the sense that

$$f\left(s; \vec{\theta}\,\right) = f\left(hs; R(h)^{-1\top}\,\vec{\theta}\,\right) \tag{7}$$

for every $s \in \mathcal{S}$ and every $\vec{\theta} \in \mathbb{R}^M$ (see SI Sec. 3.2). Although linear models of sequence-function relationships

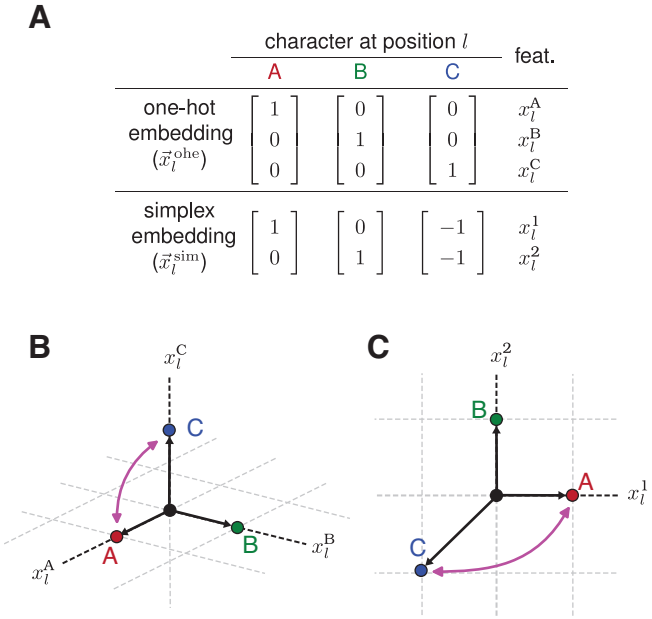

Figure 1. Transformation behavior of two single-position embeddings. (A) Two possible embeddings of characters at position l of a sequence built from the three-character alphabet $\mathcal{A} = \{\text{A}, \text{B}, \text{C}\}$: the three-dimensional one-hot embedding $\vec{x}_l^{\text{ohe}}$ and the two-dimensional simplex embedding $\vec{x}_l^{\text{sim}}$. The elements of $\vec{x}_l^{\text{ohe}}$ are the three one-hot sequence features x_l^{A}, x_l^{B}, and x_l^{C}. The two elements of $\vec{x}_l^{\text{sim}}$ are denoted x_l^1 and x_l^2. (B) The three-dimensional embedding $\vec{x}^{\text{ohe}}$ for each possible character at position l. (C) The two-dimensional embedding $\vec{x}_l^{\text{sim}}$ for each possible character at position l. Pink arrows in panels B and C indicate the transformation of each embedding vector induced by permuting the characters A and C.

can be equivariant in a variety of symmetry groups H, we use the term "equivariant" to specifically refer to equivariance under H_{PSCP} unless otherwise noted.

One-hot models

The most commonly used equivariant models are based on single-position one-hot embeddings. Such models are arguably the most intuitive, as their features are built from the indicator functions for single-position alleles (e.g., the nucleotides in a DNA sequence or the amino acids in a protein sequence). We denote the single-position one-hot embedding for position l as $\vec{x}_l^{\text{ohe}}$ and define it to be a binary vector of dimension α with features $x_l^{c_1}, \ldots, x_l^{c_\alpha}$, where $c_1, \ldots, c_\alpha$ is an ordering of the characters in $\mathcal{A}$. For example, Fig. 1A shows $\vec{x}_l^{\text{ohe}}$ for the three-character alphabet $\mathcal{A} = \{\text{A}, \text{B}, \text{C}\}$.

The embedding $\vec{x}_l^{\text{ohe}}$ transforms under what is known as a "permutation representation" [29]. We denote this representation as R_l^{ohe}. For example, consider the transformation $h_{\text{A}\leftrightarrow\text{C}}$ that exchanges characters A and C at every position in a sequence. The effect of this transformation on $\vec{x}_l^{\text{ohe}}$ (Fig. 1B) is equivalent to multiplying $\vec{x}_l^{\text{ohe}}$ by the matrix

$$R_l^{\text{ohe}}(h_{\text{A}\leftrightarrow\text{C}}) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{8}$$

This and all other matrices in the representation R_l^{ohe} are permutation matrices, in that all matrix elements are 0 or 1, and each row and column contains a single 1. Consequently, multiplying a vector by one of these matrices changes the order of the elements in that vector, but does not change the overall set values that those elements take. We refer to $\vec{x}_l^{\text{ohe}}$ and other embeddings that transform under permutation representations as permutation embeddings; their corresponding models are called permutation models.

The embeddings of many different models can be built by taking direct sums of Kronecker products of $\vec{x}_l^{\text{ohe}}$. For example, the pairwise one-hot model of Eq. 2 is based on the embedding

$$\vec{x}_{\text{pair}}^{\text{ohe}} = \vec{x}^{\text{triv}} \oplus \left\{ \bigoplus_l \vec{x}_l^{\text{ohe}} \right\} \oplus \left\{ \bigoplus_{l<l'} \vec{x}_{ll'}^{\text{ohe}} \right\}, \tag{9}$$

where $\vec{x}^{\text{triv}}$ denotes the trivial embedding (defined to be the 1-dimensional vector [1] for all sequences) and the Kronecker product

$$\vec{x}_{ll'}^{\text{ohe}} = \vec{x}_l^{\text{ohe}} \otimes \vec{x}_{l'}^{\text{ohe}} \tag{10}$$

yields an α^2-dimensional embedding having elements $x_{ll'}^{cc'}(s) = x_l^c(s)\, x_{l'}^{c'}(s)$ for all $s \in \mathcal{S}$ and all characters $c, c' \in \mathcal{A}$. The direct sums in Eq. 9 yield $\vec{x}_{\text{pair}}^{\text{ohe}}$ by stacking the component embeddings on top of one another in the resulting column vector. Note that, because $\vec{x}_l^{\text{ohe}}$ is a permutation embedding, so is $\vec{x}_{\text{pair}}^{\text{ohe}}$. In fact, any embedding constructed from a direct sum of Kronecker products of $\vec{x}_l^{\text{ohe}}$ is a permutation embedding. We call this class of models the "generalized one-hot models".

How a single-position embedding transforms has important consequences for how the parameters of models constructed from that embedding are interpreted. For the pairwise one-hot model, the fact that $\vec{x}_l^{\text{ohe}}$ transforms under a permutation representation implies that both $\vec{x}_{\text{pair}}^{\text{ohe}}$ and $\vec{\theta}_{\text{pair}}^{\text{ohe}}$ transform under permutation representations as well. A consequence of this is that the individual parameters in $\vec{\theta}_{\text{pair}}^{\text{ohe}}$ can be interpreted as quantifying the intrinsic effects of individual alleles. For example, the transformation $h_{\text{A}\leftrightarrow\text{C}}$ induces a permutation of parameters that exchanges $\theta_l^{\text{A}} \leftrightarrow \theta_l^{\text{C}}$ at all positions l, exchanges $\theta_{ll'}^{\text{AA}} \leftrightarrow \theta_{ll'}^{CC}$ at all pairs of positions $l < l'$, and so on. Model parameters therefore track their corresponding alleles: θ_l^{A} tracks sequences that have A at position l, $\theta_{ll'}^{\text{AA}}$ tracks sequences that have AA at positions l and l', etc..

The fact that $\vec{x}_l^{\text{ohe}}$ transforms under a permutation representation also means that the features therein are not linearly independent. For example, the three embedding vectors in Fig. 1B lie within a two-dimensional affine

subspace defined by the constraint $x_l^A + x_l^B + x_l^C = 1$. As we will see, a consequence of such constraints is that embeddings (like $\vec{x}_{\text{pair}}^{\text{ohe}}$) that are built from direct sums of Kronecker products of single-position one-hot embeddings will yield models that have gauge freedoms. So although the parameters of generalized one-hot models can be interpreted as quantifying intrinsic allelic effects, the numerical values of individual parameters cannot (at least in some cases) be interpreted in the absence of gauge-fixing constraints.

Simplex models

Simplex embeddings mathematically represent alleles in a more compact but less intuitive way than the one-hot embeddings discussed above do. Single-position simplex embeddings encode the α characters of $\mathcal{A}$ using zero-centered vectors of dimension $\alpha - 1$, and thus have fewer dimensions than corresponding alleles. Simpilex embeddings can be defined in multiple ways that differ from one another by similarity transformations, i.e., change-of-basis transformations. Here we adopt a particularly convenient definition: $\vec{x}_l^{\text{sim}}(s)$ is defined to be an $\alpha - 1$ dimensional vector, the i'th element of which is $x_l^{c_i}(s)$ if $s_l \neq c_\alpha$ and -1 if $s_l = c_\alpha$. We use the superscript "sim" here and in what follows to indicate mathematical objects that are based on simplex embeddings. Figs. 1A,C illustrate $\vec{x}_l^{\text{sim}}$ for the three-character alphabet. Unlike $\vec{x}_l^{\text{ohe}}$, $\vec{x}_l^{\text{sim}}$ transforms under a non-permutation representation, which we denote as R_l^{sim}. For example, the effect of $h_{\text{A}\leftrightarrow\text{C}}$ on $\vec{x}_l^{\text{sim}}$ is equivalent to multiplication by the matrix

$$R_l^{\text{sim}}(h_{\text{A}\leftrightarrow\text{C}}) = \begin{bmatrix} -1 & 0 \\ -1 & 1 \end{bmatrix}. \tag{11}$$

As with one-hot embeddings, the embeddings of many different models can be built from direct sums of direct products of $\vec{x}_l^{\text{sim}}$. For example, a simplex embedding analogous to $\vec{x}_{\text{pair}}^{\text{ohe}}$ can be constructed as

$$\vec{x}_{\text{pair}}^{\text{sim}} = \vec{x}^{\text{triv}} \oplus \left\{ \bigoplus_l \vec{x}_l^{\text{sim}} \right\} \oplus \left\{ \bigoplus_{l<l'} \vec{x}_l^{\text{sim}} \otimes \vec{x}_{l'}^{\text{sim}} \right\}. \tag{12}$$

The corresponding pairwise simplex model has the form

$$f_{\text{pair}}^{\text{sim}} = \theta_0 x_0 + \sum_l \sum_{i=1}^{\alpha-1} \theta_l^i x_l^i + \sum_{l<l'} \sum_{i,j=1}^{\alpha-1} \theta_{ll'}^{ij} x_{ll'}^{ij}, \tag{13}$$

where x_l^i denotes the i'th element of $\vec{x}_l^{\text{ohe}}$, and where $x_{ll'}^{ij}(s) = x_l^i(s) x_{l'}^j(s)$ for all $s \in \mathcal{S}$. We use $\vec{\theta}_{\text{pair}}^{\text{sim}}$ to denote the parameters of this model. Note that these parameters are indexed using numerical superscripts ranging from 1 to $\alpha - 1$, rather than by characters in $\mathcal{A}$.

Pairwise simplex models describe the same sequence-function relationships that pairwise one-hot models do, i.e., given a set of parameters for one of these models, there exists a corresponding set of parameters for the other model that yields the same predictions over all sequences. However, because $\vec{x}_l^{\text{sim}}$ has lower dimension than $\vec{x}_l^{\text{ohe}}$, $\vec{\theta}_{\text{pair}}^{\text{sim}}$ contains fewer parameters than $\vec{\theta}_{\text{pair}}^{\text{ohe}}$. Inspection of Eq. 12 shows that the number of parameters in $\vec{\theta}_{\text{pair}}^{\text{ohe}}$ is, in fact,

$$\dim \vec{x}_{\text{pair}}^{\text{sim}} = 1 + (\alpha - 1)L + \binom{L}{2}(\alpha - 1)^2. \tag{14}$$

This reduction in the number of parameters entirely eliminates gauge freedoms, as can be seen from the fact that

$$\gamma_{\text{pair}}^{\text{ohe}} = \dim \vec{x}_{\text{pair}}^{\text{ohe}} - \dim \vec{x}_{\text{pair}}^{\text{sim}}. \tag{15}$$

The lack of gauge freedoms in $f_{\text{pair}}^{\text{sim}}$ is one example of the fact that, as we will see, models defined using (non-redundant) simplex embeddings do not have gauge freedoms. In fact, multiple groups [20, 22, 23] have argued for the use of simplex models, rather than one-hot models, based on the simplex models not having gauge freedoms.

We argue, however, that the parameters of simplex models are fundamentally more difficult to interpret as allelic effects than are the parameters of one-hot models. Because $\vec{x}_l^{\text{sim}}$ does not transform under a permutation representation, neither does $\vec{x}_{\text{pair}}^{\text{sim}}$ and neither does $\vec{\theta}_{\text{pair}}^{\text{sim}}$. In the case of the three-character alphabet, one sees from Eq. 12 that $h_{\text{A}\leftrightarrow\text{C}}$ induces a transformation of model parameters that maps $\theta_l^1 \to -\theta_l^1$, $\theta_l^2 \to -\theta_l^1 + \theta_l^2$, $\theta_{ll'}^{22} \to \theta_{ll'}^{11} - \theta_{ll'}^{12} - \theta_{ll'}^{21} + \theta_{ll'}^{22}$, and so on. The fact that these parameters change in ways described by nontrivial linear combinations means that individual parameters cannot be interpreted as quantifying individual allelic effects.

Maschke decomposition

We now use methods from the theory of group representations to formally investigate the general connection between model transformation behavior and gauge freedoms. Maschke's theorem, a foundational result in representation theory, states that every matrix representation of a finite group is equivalent to a direct sum of irreducible matrix representations. Here the term "equivalent" means that there is a similarity transformation (i.e., a change of basis) that maps one representation to another; we use the symbol $\simeq$ to denote equivalence in what follows. The term "irreducible" means that the representation has no proper invariant subspace.

Consider for example R_l^{ohe}, the representation that describes how $\vec{x}_l^{\text{ohe}}$ transforms under H_{PSCP}. The group H_{PSCP} is isomorphic to the symmetric group (i.e., the group of permutations), the representations of which are well understood [29]. In this context, R_l^{ohe} is called the "defining representation" and is well known to be reducible. Specifically, R_l^{ohe} has two proper invariant subspaces. One subspace has dimension 1 and is spanned by the vector $[1\ 1\ \cdots\ 1]^\top$. The other subspace has dimension $\alpha - 1$ and consists of the set of α-dimensional

vectors whose elements sum to zero. The first of these subspaces transforms under the "trivial representation", which is simply the 1x1 matrix [1] and which we denote by R^{triv}. The other subspace transforms (after an appropriate change of coordinates) under the representation R_l^{sim}. R_l^{sim} is called the "standard representation" and is well known to be irreducible. The Maschke decomposition of R_l^{ohe} is therefore given by

$$R_l^{\text{ohe}} \simeq R^{\text{triv}} \oplus R_l^{\text{sim}}, \tag{16}$$

where the direct sum on the right-hand side yields a block diagonal matrix created from R^{triv} and R_l^{sim}.

Equivalently, we can think of Maschke decomposition in terms of embeddings. Thinking in terms of embeddings can be helpful for deriving the specific invertible matrix that performs the similarity transformation needed to express a Maschke decomposition as an equality instead of an equivalence. When multiplied by an appropriate similarity transformation matrix T, $\vec{x}_l^{\text{ohe}}$ can be expressed as a direct sum of the trivial embedding $\vec{x}^{\text{triv}}$ (which is simply the 1-dimensional vector [1]) and the simplex embedding $\vec{x}_l^{\text{sim}}$, i.e.,

$$T\vec{x}_l^{\text{ohe}}(s) = \vec{x}^{\text{triv}}(s) \oplus \vec{x}_l^{\text{sim}}(s), \tag{17}$$

for all sequences s. This allows us to express the equivalence relation in Eq. 16 as an equality, as it implies that

$$TR_l^{\text{ohe}}(h)T^{-1} = R^{\text{triv}}(h) \oplus R_l^{\text{sim}}(h) \tag{18}$$

for all group elements h. Based on the definition of the embeddings $\vec{x}_l^{\text{ohe}}$ and $\vec{x}_l^{\text{sim}}$ above, one can readily show that the similarity transformation matrix T is given by

$$T = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}. \tag{19}$$

This matrix T will be used later when defining an algorithm for distilling general equivariant embeddings.

Decomposition of equivariant embeddings

Maschke's Theorem implies that any representation R of H_{PSCP} can be expressed as

$$R \simeq \bigoplus_{k=1}^{K} Q_k R_k, \tag{20}$$

where the R_k are distinct irreducible representations of H_{PSCP} and each Q_k is a natural number that denotes the multiplicity of R_k in the direct sum. R is thus equivalent to a block-diagonal representation formed by placing Q_k copies of each R_k along the diagonal and setting all other matrix elements to zero (see Fig. 2B). One consequence of Eq. 20 is that any embedding $\vec{x}$ that transforms under R can be decomposed as

$$\vec{x} \simeq \bigoplus_{k=1}^{K} \bigoplus_{q=1}^{Q_k} \vec{x}_{kq}, \tag{21}$$

where each $\vec{x}_{kq}$ is an embedding that transforms under R_k. In what follows, we say that embeddings like $\vec{x}_{kq}$ are irreducible because they transform under irreducible representations. We also assume that all $\vec{x}_{kq}$ are nonzero, but this assumption can be removed without fundamentally changing our results; see SI Sec. 5.2 for details. The Maschke decompositions of R and $\vec{x}$ are illustrated in Fig. 2A,B.

Distillation of equivariant embeddings

We now describe an "embedding distillation" procedure that connects the Maschke decomposition of $\vec{x}$ to the gauge freedoms of the corresponding model. In SI Sec. 5.1 we prove the following:

Theorem 1 *Any two nonzero sequence embeddings that transform under the same irreducible representation of H_{PSCP} are equal up to a constant of proportionality.*

Using Theorem 1 we find that there is a similarity transformation matrix T_{decom} such that

$$T_{\text{decom}}\,\vec{x} = \bigoplus_{k=1}^{K} Q_k \vec{x}_k, \tag{22}$$

where, for each k, $\vec{x}_k$ denotes any one of the irreducible embeddings $\vec{x}_{kq}$ in Eq. 21 and Q_k denotes the multiplicity of each term in the direct sum. Next we perform a similarity transformation (described by a matrix T_{thin}) that "thins out" the embedding by setting all except one copy of each $\vec{x}_k$ to zero. Finally, we perform a similarity transformation (described a matrix T_{sort}) that "sorts" the remaining nonzero embeddings, arranging them in series at the top of the resulting embedding vector. We thus find that applying the cumulative similarity transformation given by

$$T_{\text{dist}} = T_{\text{sort}}\,T_{\text{thin}}\,T_{\text{decom}} \tag{23}$$

to the embedding $\vec{x}$ yields

$$T_{\text{dist}}\,\vec{x} = \vec{x}^{\,\text{dist}} \oplus \vec{0}_\gamma, \tag{24}$$

where $\vec{0}_\gamma$ is a γ-dimensional vector of zeros and

$$\vec{x}^{\,\text{dist}} = \bigoplus_{k=1}^{K} \vec{x}_k \tag{25}$$

is a "distilled embedding". When applied to the representation R, this distillation procedure yields

$$T_{\text{dist}}\,R\,T_{\text{dist}}^{-1} = \vec{R}^{\,\text{dist}} \oplus \vec{R}^{\,\text{redun}}, \tag{26}$$

where the "distilled representation", $R^{\text{dist}} = \bigoplus_{k=1}^{K} R_k$, comprises one copy of each R_k present in Eq. 20, and where the redundant representation, $R^{\text{redun}} = \bigoplus_{k=1}^{K} (Q_k - 1) R_k$, sweeps up the remaining copies of each R_k. The final distilled versions of R and $\vec{x}$ are illustrated in Fig. 2C. Explicit formulae for constructing T_{decom}, T_{thin}, and T_{dist} are given in SI Sec. 8.

Identification of gauge freedoms in equivariant models

To identify the gauge freedoms of an equivariant model, we use the fact that $\vec{x}^{\,\text{dist}}$ (defined in Eq. 25) is full rank. This is a consequence of the following Theorem, which is proven in SI Sec. 3.4:

Theorem 2 *For each $k \in \{1, \ldots, K\}$, let $\vec{x}_k$ be a nonzero embedding that transforms under an irreducible representation R_k of the group H_{PSCP}. Then the direct sum of all $\vec{x}_k$ is full rank if and only if all R_k are pairwise inequivalent.*

Because $\vec{x}^{\,\text{dist}}$ is full rank, $\vec{g}^{\,\top} \vec{x}(s) = 0$ for all $s \in \mathcal{S}$ if and only if

$$\vec{g} = T_{\text{dist}}^{\top} \left[\vec{0}_{M-\gamma} \oplus \vec{g}_{\gamma} \right], \tag{27}$$

where T_{dist} is the distillation matrix in Eq. 23 and $\vec{g}_{\gamma}$ is any vector in $\mathbb{R}^{\gamma}$. The space of gauge transformations G is therefore given by the set of vectors $\vec{g}$ that have the form in Eq. 27. In particular, the number of gauge freedoms is seen to be

$$\gamma = \dim \vec{x} - \dim \vec{x}^{\,\text{dist}} = \deg R^{\text{redun}}. \tag{28}$$

We thus see that the number of gauge freedoms of an equivariant linear model is equal to the sum of the degrees of all the redundant irreducible representations under which that model's embedding (or equivalently, that model's parameter vector) transforms.

Identification of all equivariant models

The mathematical structure of a group defines the models that transform equivariantly under that group. In the case of H_{PSCP}, the relatively simple group structure allows the straight-forward identification of all inequivalent distilled embeddings and thus the identification of all equivariant linear models.

H_{PSCP} can be written as a direct product of simpler groups:

$$H_{\text{PSCP}} = H_1 \times \cdots \times H_L, \tag{29}$$

where each H_l denotes the group of character permutations at sequence position l. Each irreducible representation R_k of H_{PSCP} can therefore be expressed as the

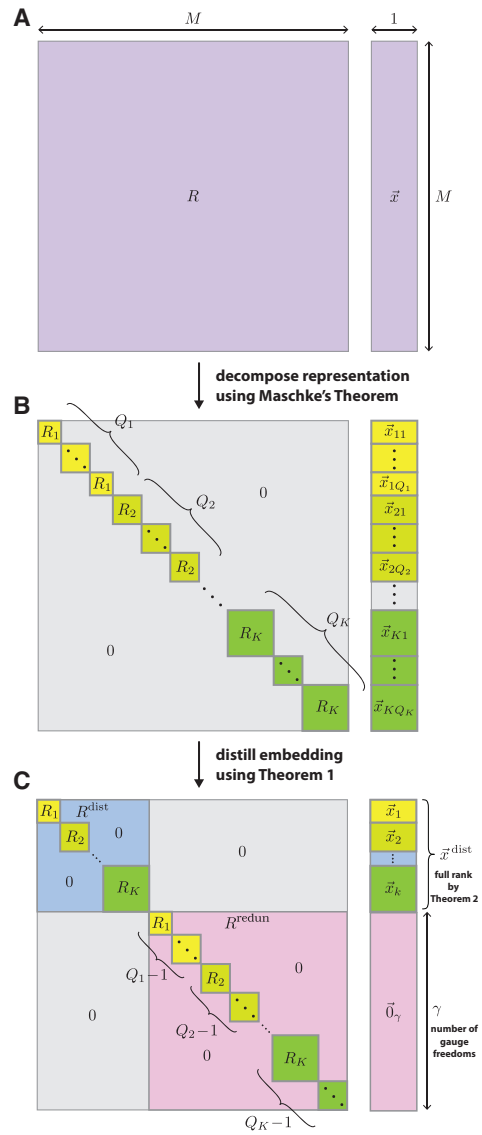

Figure 2. **Embedding distillation**. (A) Given an M-dimensional embedding $\vec{x}$ that is equivariant under H_{PSCP}, let R be the representation that describes how $\vec{x}$ transforms. (B) By Maschke's theorem, R can be decomposed into a direct sum of irreducible representations, R_k ($k \in \{1, \ldots, K\}$), each of which occurs with multiplicity Q_k (Eq. 20). Similarly, $\vec{x}$ can be decomposed into a direct sum of irreducible embeddings $\vec{x}_{kq}$ ($q \in \{1, \ldots, Q_k\}$), where each $\vec{x}_{kq}$ transforms under R_k (Eq. 21). (C) By Theorem 1, an additional similarity transformation can be performed that, for each value of k, zeroes out all but one $\vec{x}_{kq}$ and sorts the remaining embeddings; each remaining $\vec{x}_{kq}$ is denoted by $\vec{x}_k$. Consequently, $\vec{x}$ decomposes into a direct sum of a distilled embedding $\vec{x}^{\,\text{dist}}$ and a zero vector $\vec{0}_{\gamma}$ having some dimension γ (Eq. 24). $\vec{x}^{\,\text{dist}}$ is given by the direct sum of all $\vec{x}_k$ (Eq. 25) and is full rank by Theorem 2. The distilled representation R^{dist} describes how $\vec{x}^{\,\text{dist}}$ transforms and is given by a direct sum of one copy of each R_k. The redundant representation R^{redun} operates on $\vec{0}_{\gamma}$ and comprises the $Q_k - 1$ remaining copies of each R_k. The resulting number of gauge freedoms is γ (see Eq. 28).

Kronecker product

$$R_k \simeq \bigotimes_{l=1}^{L} R_l^k, \tag{30}$$

where each R_l^k is an irreducible representation of H_l (see Theorem 1.11.3 of [29]). An embedding $\vec{x}_k$ that transforms under R_k will therefore have the form

$$\vec{x}_k \simeq \bigotimes_{l=1}^{L} \vec{x}_l^k, \tag{31}$$

where $\vec{x}_l^k$ is an irreducible embedding that transforms under H_l. In SI Sec. 4.3 we show that H_l supports only two inequivalent irreducible embeddings (regardless of alphabet size): $\vec{x}^{\,\mathrm{triv}}$ and $\vec{x}_l^{\,\mathrm{sim}}$. Each $\vec{x}_l^k$ must therefore be equivalent to one of these two embeddings. Ignoring the factors of $\vec{x}^{\,\mathrm{triv}}$ because they do not impact Kronecker products, Eq. 31 becomes

$$\vec{x}_k \simeq \bigotimes_{l \in B_k} \vec{x}_l^{\,\mathrm{sim}}, \tag{32}$$

where B_k is a subset of the positions $\{1, \ldots, L\}$. There are 2^L possible choices for each subset B_k, and thus 2^L inequivalent irreducible embeddings $\vec{x}_k$. Since each $\vec{x}_k$ can appear at most once on the left-hand side of Eq. 25, there are 2^{2^L} inequivalent distilled embeddings $\vec{x}^{\,\mathrm{dist}}$.

For each choice of $\vec{x}^{\,\mathrm{dist}}$ there are an infinite number of possible choices for T_{dist} and γ that can be used to define $\vec{x}$ (via Eq. 24). The number of possible equivariant embeddings $\vec{x}$, and thus the number of equivariant models f, is therefore infinite. However, all models corresponding to a specific $\vec{x}^{\,\mathrm{dist}}$ have the same expressivity, i.e., the set of sequence-function relationships that each model describes (considered over all possible values of model parameters) is the same. We therefore consider these models to be equivalent, and conclude that there are a total of 2^{2^L} inequivalent equivariant linear models on sequences of length L.

Analytical analysis of generalized one-hot models

We now use the embedding distillation procedure to compute the number of gauge freedoms of all generalized one-hot models. This derivation is based on the Maschke decomposition $\vec{x}_l^{\,\mathrm{ohe}} \simeq \vec{x}^{\,\mathrm{triv}} \oplus \vec{x}_l^{\,\mathrm{sim}}$ from Eq. 17.

We first demonstrate this calculation on the pairwise one-hot model. Plugging the decomposition of $\vec{x}_l^{\,\mathrm{ohe}}$ into the definition for $\vec{x}_{\mathrm{pair}}^{\,\mathrm{ohe}}$ in Eq. 9, then expanding the Kronecker products and grouping like terms, we find that

$$\vec{x}_{\mathrm{pair}}^{\,\mathrm{ohe}} \simeq \left[1 + L + \binom{L}{2}\right] \vec{x}^{\,\mathrm{triv}} \oplus \left\{\bigoplus_{l} L\, \vec{x}_l^{\,\mathrm{sim}}\right\} \oplus \left\{\bigoplus_{l<l'} \vec{x}_l^{\,\mathrm{sim}} \otimes \vec{x}_{l'}^{\,\mathrm{sim}}\right\}, \tag{33}$$

where the scalar coefficients correspond to the Q_k in Eq. 22. We derive the corresponding distilled embedding by replacing each of these coefficients with 1. Doing so reveals the distillation of $\vec{x}_{\mathrm{pair}}^{\,\mathrm{ohe}}$ to be $\vec{x}_{\mathrm{pair}}^{\,\mathrm{sim}}$. The result for $\gamma_{\mathrm{pair}}^{\,\mathrm{ohe}}$ in Eq. 5 is therefore just a manifestation of Eq. 28.

We now extend this approach to all generalized one-hot models. The embedding $\vec{x}_{\mathrm{goh}}$ of any generalized one-hot model can be written as

$$\vec{x}_{\mathrm{goh}} = \bigoplus_{j=1}^{J} \bigotimes_{l \in A_j} \vec{x}_l^{\,\mathrm{ohe}}. \tag{34}$$

where $A_1, \ldots, A_J$ denote J (not necessarily distinct) sets of positions. Because the dimension of $\vec{x}_l^{\,\mathrm{ohe}}$ is α, the number of corresponding model parameters is

$$M_{\mathrm{goh}} = \dim \vec{x}_{\mathrm{goh}} = \sum_{j=1}^{J} \alpha^{|A_j|}. \tag{35}$$

Decomposing $\vec{x}_l^{\,\mathrm{ohe}}$ in terms of $\vec{x}^{\,\mathrm{triv}}$ and $\vec{x}_l^{\,\mathrm{sim}}$, expanding each Kronecker product, then grouping the resulting terms, we find that

$$\vec{x}_{\mathrm{goh}}^{\,\mathrm{dist}} = \bigoplus_{k=1}^{K} \bigotimes_{l \in B_k} \vec{x}_l^{\,\mathrm{sim}}, \tag{36}$$

where $B_1, \ldots, B_K$ denote the distinct subsets of positions that occur among all the A_j. Because the dimension of $\vec{x}_l^{\,\mathrm{sim}}$ is $\alpha - 1$,

$$\dim \vec{x}_{\mathrm{goh}}^{\,\mathrm{dist}} = \sum_{k=1}^{K} (\alpha - 1)^{|B_k|}. \tag{37}$$

The number of gauge freedoms of the generalized one-hot model having embedding $\vec{x}_{\mathrm{goh}}$ is therefore given by

$$\gamma_{\mathrm{goh}} = \sum_{j=1}^{J} \alpha^{|A_j|} - \sum_{k=1}^{K} (\alpha - 1)^{|B_k|}. \tag{38}$$

Table I reports the number of gauge freedoms calculated in this manner for a variety of generalized one-hot models (illustrated in Fig. 3). SI Sec. 6 provides expanded descriptions for each generalized one-hot model, as well as detailed derivations of the results in Table I.

A result of this analysis is that all generalized one-hot models have gauge freedoms, save models for which the direct sum in Eq. 34 includes only one term. To see this, observe that Eq. 22 gives

$$\dim \vec{x}_{\mathrm{goh}} = \sum_{k=1}^{K} Q_k (\alpha - 1)^{|B_k|}, \tag{39}$$

where each multiplicity value Q_k is equal to the number of sets A_j that contain B_k. Using this together with Eq. 36 and Eq. 28 gives

$$\gamma_{\mathrm{goh}} = \sum_{k=1}^{K} (Q_k - 1)(\alpha - 1)^{|B_k|}. \tag{40}$$

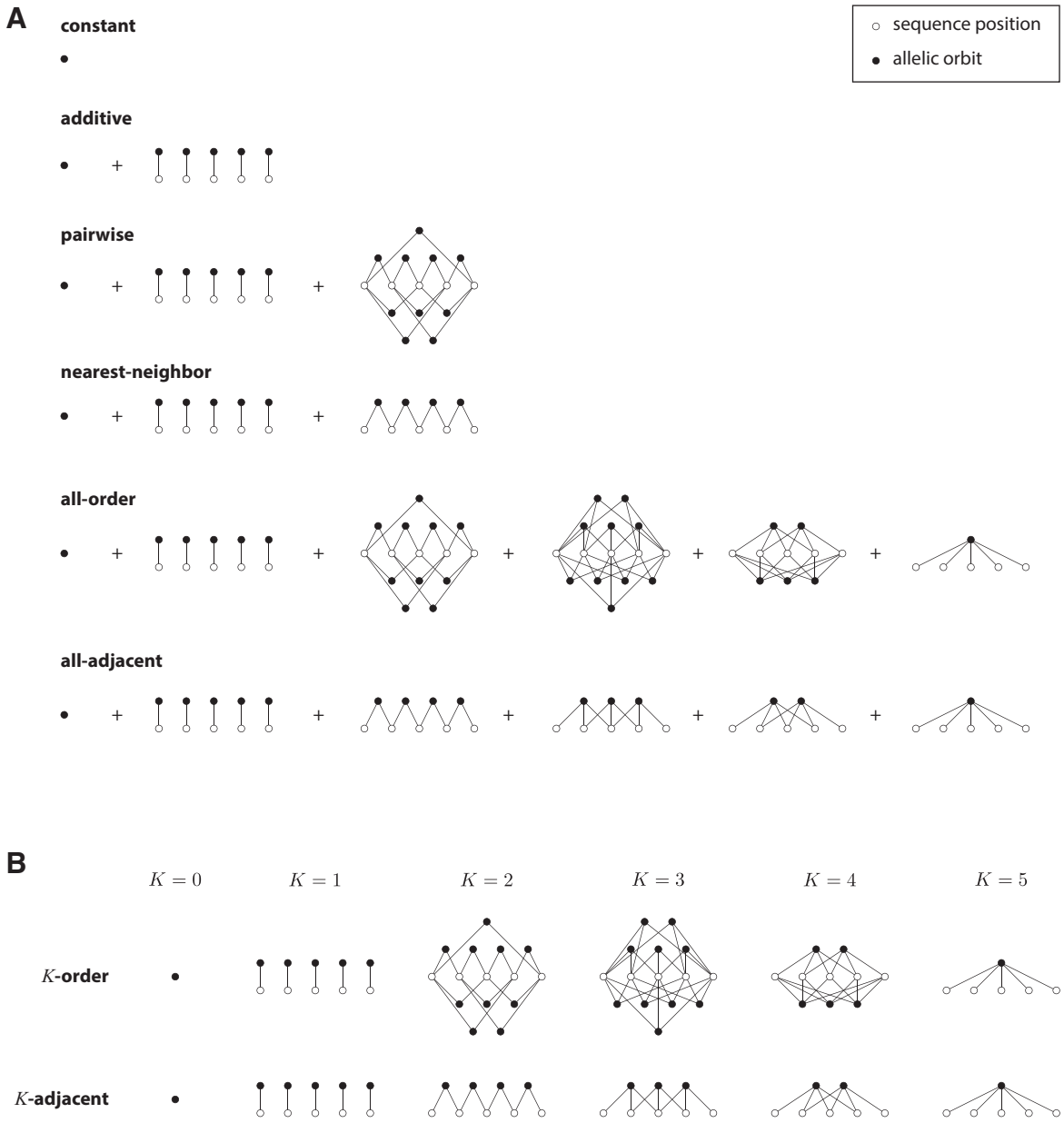

Figure 3. Structure of generalized one-hot models analyzed in Table I for sequences of length $L = 5$. Open circles represent sequence positions. Closed circles represent allelic orbits, i.e., sets of sequence features that are closed under the action of H_{PSCP}. Edges indicate position indices shared by the features in each allelic orbit. (A) Structure of constant, additive, pairwise, nearest-neighbor, all-order, and all-adjacent models. (B) Structure of K-order models and K-adjacent models for various interaction orders K.

We thus find that $\gamma_{\mathrm{goh}} = 0$ if and only if none of the Q_k are greater than 1. But since the empty set is a subset of every A_j, it will always be among the B_k, and the corresponding multiplicity value will be $Q_k = J$. Gauge freedoms are therefore present in all generalized one-hot models for which $J \geq 2$. Conversely, $\gamma_{\mathrm{goh}} = 0$ when $J = 1$ because all B_k occur with multiplicity $Q_k = 1$. Gauge freedoms are therefore absent in all generalized one-hot models for which $J = 1$.

Computational analysis of generalized one-hot models

Embedding distillation also allows one to efficiently compute a sparse basis for the space of gauge freedoms G_{goh} of any generalized one-hot model. Eq. 27 reveals that G_{goh} is spanned by the last γ_{goh} row vectors of T_{dist}. One can therefore compute a basis for G_{goh} by computing T_{dist}. This is notable because computing T_{dist} only requires keeping track of the similarity transformations

model type	interaction orders	number of parameters (M_{goh})	number of gauge freedoms (γ_{goh})
constant	0	1	0
additive	$0, 1$	$1 + L\alpha$	L
pairwise	$0, 1, 2$	$1 + L\alpha + \binom{L}{2}\alpha^2$	$L + \binom{L}{2}(2\alpha - 1)$
nearest-neighbor	$0, 1, 2^{\ddagger}$	$1 + L\alpha + (L-1)\alpha^2$	$L + (L-1)(2\alpha - 1)$
all-order	$0, 1, \ldots, L$	$(\alpha+1)^L$	$(\alpha+1)^L - \alpha^L$
all-adjacent	$0, 1, 2^{\ddagger}, \ldots, L^{\ddagger}$	$1 + \frac{\alpha}{(\alpha-1)^2}\left[\alpha^{L+1} - (L+1)\alpha + L\right]$	$1 + \frac{\alpha}{(\alpha-1)^2}\left[2\alpha^L - \alpha^{L-1} - (L+1)\alpha + L\right]$
K-order	K	$\binom{L}{K}\alpha^K$	$\binom{L}{K}\alpha^K - \sum_{k=0}^{K}\binom{L}{k}(\alpha-1)^k$
hierarchical K-order	$0, 1, \ldots, K$	$\sum_{k=0}^{K}\binom{L}{k}\alpha^k$	$\sum_{k=0}^{K}\binom{L}{k}\left[\alpha^k - (\alpha-1)^k\right]$
K-adjacent‡	$K^{\ddagger}$	$(L-K+1)\alpha^K$	$(L-K)\alpha^{K-1}$
hierarchical K-adjacent‡	$0, 1, 2^{\ddagger}, \ldots, K^{\ddagger}$	$1 + \sum_{k=1}^{K}(L-k+1)\alpha^k$	$(L-K)\alpha^{K-1} + 1 + \sum_{k=1}^{K-1}(L-k+1)\alpha^k$

Table I. Analytical results for various generalized one-hot models, computed using Eqs. 35 and 38. See SI Sec. 6 for derivations of these results. K-adjacent models assume $K \geq 1$. ‡Only includes interactions among adjacent positions.

needed to express $\vec{x}_{\text{goh}}$ in the distilled form shown in Eq. 24. This computation is far less demanding than computing a basis for G_{goh} using Gaussian elimination or singular value decomposition when (as is often the case) the number of possible sequences is far greater than the number of model parameters.

In Eq. 23 we described how to construct T_{dist} from a product of three matrices: T_{decom}, T_{thin}, and T_{sort}. Explicit formulae for computing these matrices, as well as their inverses, are provided in SI Sec. 8. For these formulae we observe that each matrix, as well as its inverse, is sparse in the large L limit when the maximal order of interaction described by the model is fixed. The resulting distillation matrix T_{dist} is therefore also sparse, as is its inverse. It also turns out that every nonzero element of T_{dist} is +1 or -1. Taking the last γ_{goh} rows of T_{dist} thus provides a basis for G_{goh} consisting of sparse vectors whose only nonzero elements are $+1$ and -1. Having sparse matrices for T_{dist} and T_{dist}^{-1} also allows us to compute a sparse gauge-fixing projection matrix P; see SI Sec. 8 for details. Fig. 4 illustrates the actions of T_{decom}, T_{thin}, and T_{sort} on an example embedding vector for the all-order interaction model corresponding to $L = 3$ and $\alpha = 3$. Fig. 4 also illustrates the corresponding distillation matrix T_{dist}.

Other symmetry groups

The proof of Theorem 1 in SI Sec. 5.1, and thus our embedding distillation procedure, applies only to the symmetry group H_{PSCP}. There are other symmetry groups of sequence space besides H_{PSCP}, however, and it is worth asking whether Theorem 1, and thus Eqs. 24-28, hold for those groups as well.

One symmetry group is the group of global character permutations, H_{GCP}. This group comprises transformations that apply the same permutation to characters at every position in a sequence. Another is the group of position permutations, H_{PP}. This group comprises transformations that permute the positions in a sequence without otherwise changing the characters therein. SI Sec. 7.1 shows that Theorem 1 does not hold for either H_{GCP} or H_{PP}. Consequently, one cannot compute distilled embeddings using the irreducible representations of either group.

A third symmetry group is H_{Ham}, which describes combinations of position permutations and position-specific character permutations. H_{Ham} is the largest symmetry group that preserves Hamming distances [32], and includes H_{PSCP}, H_{PP}, and H_{GCP} as subgroups. Theorem 1 does hold for H_{Ham}, due the fact that H_{PSCP} is a subgroup (see SI Sec. 7.2). However, the set of models that are equivariant under H_{Ham} is a subset of the models that are equivariant under H_{PSCP}, and the irreducible representations of H_{Ham} are more complex than those of H_{PSCP}. H_{PSCP} is therefore more useful than H_{Ham} for analyzing gauge freedoms.

Transformation behavior and parameter interpretability

We now return to the connection between parameter transformation behavior and parameter interpretability. Our above discussion of pairwise models suggested that the ability to interpret individual parameters as quantifying intrinsic allelic effects required the presence of gauge freedoms. We now formalize this observation and conjecture an extension to all linear equivariant models.

Mathematically, we define a generalized allele a to be any subset of $\mathcal{S}$, and say that any sequence $s \in \mathcal{S}$ has allele a if $s \in a$. The corresponding "allelic feature" x_a is defined to be the indicator function on $\mathcal{S}$ for whether a sequence has allele a. An "allelic model" is defined to be a linear sequence-function relationship in which every

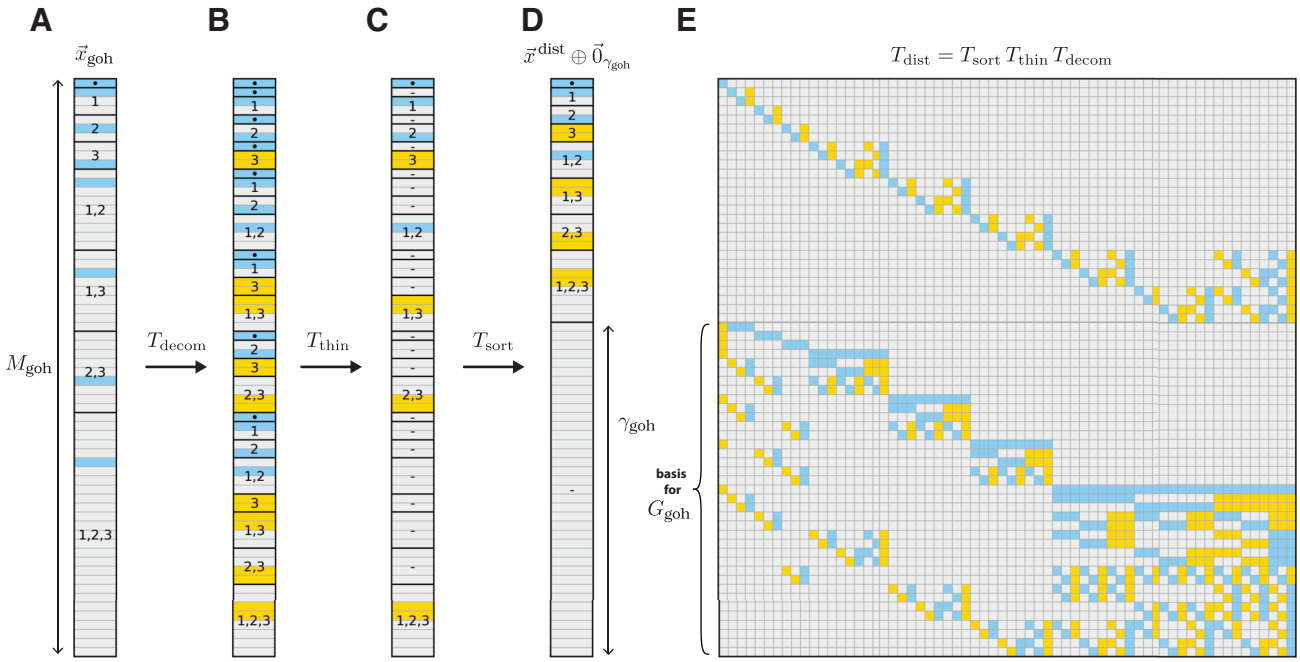

Figure 4. Embedding distillation for an example generalized one-hot model. (A) Embedding $\vec{x}_{\rm goh}$ of the $L = 3$ sequence $s = \text{ABC}$ for an all-order one-hot model based on the alphabet $\mathcal{A} = \{\text{A, B, C}\}$. This embedding has degree $M_{\rm goh} = 64$. (B) Result of multiplication by the decomposition matrix $T_{\rm decom}$. (C) Result of subsequent multiplication by the thinning matrix $T_{\rm thin}$. (D) Result of subsequent multiplication by the sorting matrix $T_{\rm sort}$, which yields $\vec{x}^{\,\rm dist} \oplus \vec{0}_{\gamma_{\rm goh}}$ with $\gamma_{\rm goh} = 37$ being the number of gauge freedoms. In B-D, dots indicate $\vec{x}^{\,\rm triv}$, dashes indicate zero vectors, and numbers indicate the positions l contributing to each component. (E) Distillation matrix $T_{\rm dist}$ that implements the full distillation procedure in A-D. The last $\gamma_{\rm goh}$ rows of $T_{\rm dist}$ provide a sparse basis for the gauge space, $G_{\rm goh}$. In A-E, vector and matrix elements are colored according to their numerical values: blue represents +1, yellow represents -1, and gray represents 0.

feature is an allelic feature. In the context of an allelic model, the parameter θ_a that multiplies x_a is said to be an "allelic effect." The parameters of a linear model can therefore be interpreted as allelic effects if and only if every one of the corresponding features is an indicator function for membership in some subset of $\mathcal{S}$.

For an allelic model to have parameters that describe intrinsic allelic effects, the model must be a "permutation model", i.e., the features and parameters of the model must transform under a permutation representation of $H_{\rm PSCP}$. Requiring an allelic model to be a permutation model puts strong constraints on which sets of alleles it can describe. Because $H_{\rm PSCP}$ permutes sequences, it also permutes alleles. Given a specific allele a, we call the set of alleles created by the action of $H_{\rm PSCP}$ on a an "allelic orbit". It is readily seen that, for an allelic model to be a permutation model, the set of alleles it describes must consist of some number J of complete allelic orbits.

All allelic models that comprise $J \geq 2$ allelic orbits have gauge freedoms. To see this, observe that the features in each orbit transform among themselves according to a permutation representation. The features of the full model will therefore transform under a direct sum of J permutation representations. Because every permutation representation contains the trivial representation in its Maschke decomposition, the decomposition of the full model's representation will contain at least J copies of the trivial representation. The model will therefore have at least $J - 1$ gauge freedoms, though additional gauge freedoms can be present as well.

This result is reflected in our above analytic analysis of generalized one-hot models. All generalized one-hot models are allelic permutation models (though the converse is not true; see SI Sec. 9.1), and each allelic orbit of a generalized one-hot model corresponds to a position set A_j in Eq. 34. The lower-bound on the number of gauge freedoms identified here recapitulates the finding above that generalized one-hot models have no gauge freedoms if and only if $J = 1$.

An allelic permutation model that does not have gauge freedoms must therefore comprise only one allelic orbit. An example of a model with only one allelic orbit is a one-hot model of length $L = 1$, e.g., a model describing the effects of only one nucleotide position in a DNA sequence or one amino acid position in a protein sequence. Are single-orbit allelic models useful in practice? We argue that the answer is essentially no. In SI Sec. 9.1 we show that single-orbit generalized one-hot models cannot describe co-occurring alleles. We regard such models as trivial because the entire reason for quantitatively modeling sequence-function relationships is to deconvolve the influence of co-occurring alleles. There are single-orbit al-

lelic permutation modelsthat describe co-occurring alleles, but all the examples of these we have analyzed either have gauge freedoms or are mathematically equivalent to generalized one-hot models (see SI Sec. 9.1). Moreover, among models whose embeddings are built from direct sums of Kronecker products of single-position embeddings, the generalized one-hot models have the fewest gauge freedoms (see SI Sec. 9.2). Based on these findings, we conjecture that all nontrivial allelic permutation models (i.e., all models whose parameters describe intrinsic allelic effects) have gauge freedoms.

DISCUSSION

Motivated by the connection between gauge freedoms and symmetries in physics, we investigated the relationship between gauge freedoms and symmetries in quantitative models of sequence-function relationships. We found that, for linear models that are equivariant under H_{PSCP} (i.e., the group of PSCP transformations), gauge freedoms arise due to model parameters transforming under redundant irreducible matrix representations. From a conceptual standpoint, this result links the gauge freedoms of models of sequence-function relationships to the transformation behavior of these models under a specific symmetry group of sequence space. From a practical standpoint, this result facilitates the analytic calculation of the number of independent gauge freedoms in a large class of commonly used models. It also enables an embedding distillation algorithm that can efficiently compute a sparse basis for the space of gauge freedoms. This latter capability may prove to be useful particularly when studying models with very large numbers of parameters. Such models are increasingly common, as massively parallel reporter assays, deep mutational scanning experiments, and other multiplex assays of variant effect can now readily measure the activities of hundreds of thousands of sequences in a single experiment [e.g. 33, 34].

We also investigated the link between parameter transformation behavior and parameter interpretability. In doing so, we identified an incompatibility between two different notions of parameter interpretability: in linear models that are equivariant under H_{PSCP}, the ability to interpret individual parameters as quantifying intrinsic allelic effects requires that these parameters transform under a permutation representation of H_{PSCP}. But in many (and possibly in all) nontrivial models, this requirement is incompatible with the ability to interpret the values of individual parameters in the absence of gauge-fixing constraints. Consequently, models that have gauge freedoms can have advantages over equally expressive models that do not have gauge freedoms.

It should be noted that there are indeed useful models that do not have gauge freedoms. One such class of models are the "wild-type" one-hot models, the features of which are limited to those describing mutations away from a specific sequence of interest [e.g., 34, 35]. Note that wild-type models differ in an important way from one-hot models expressed in the wild-type gauge (described in [30]): the latter models have specific parameters set to zero, whereas the former models lack these parameters entirely.

The parameters of wild-type models have a close connection to the quantities that one can actually experimentally measure, i.e., activity differences between alleles. However, these parameters do not transform under a permutation representation of H_{PSCP} and so do not quantify intrinsic allelic effects. Indeed, wild-type models are quite close in spirit to the representation of E&M explicitly in terms of electric and magnetic fields: while these fields are the directly measurable manifestation of E&M, they transform in complicated ways under changes in velocity and so do not provide the theoretical clarity–the *intrinsic* description of E&M–that the electromagnetic four-potential does.

Another class of useful models that do not have gauge freedoms are models whose features represent sequence-dependent physical properties, such as the chemical properties of amino acids [36, 37] or the physical shape of the DNA double helix [38, 39]. These models are not equivariant, however, and their parameters describe the effects of physical properties of alleles, not the effects of alleles themselves. Notably, both classes of model reflect inductive biases that break H_{PSCP} symmetry.

In classical field theories like E&M, there are specific symmetries that are well-established by experiment and that any mathematical formulation of the theory must be consistent with. This does not, however, mean that the equations of the theory must transform in a simple way under those symmetries. Mathematically formulating physical theories so that the equations themselves manifestly respect the symmetries of the theory generally requires over-parameterizing the equations, thereby introducing gauge freedoms. Physicists often find it worthwhile to do this, as having fundamental symmetries be manifestly reflected in one's equations can greatly facilitate the interpretation and application of those equations. Solving such equations, however, requires fixing the gauge—introducing additional constraints that make the solution of the equations unique.

Unlike in physics, there is no experimentally established requirement that models of sequence-function relationships be equivariant under any symmetries of sequence space. The specific mathematical form one uses for such models is subjective, and different models are commonly used in different contexts. Citing the ambiguities caused by gauge freedoms, some have argued for restricting one's choice of model to those that have no gauge freedoms. Nevertheless, models that have gauge freedoms are still common in the literature. We suggest that a major reason for this may be that researchers often prefer to use models that manifestly reflect symmetries of sequence space, and therefore have parameters that are interpretable as intrinsic allelic effects. As we showed, these criteria often (and possibly in all nontriv-

ial cases) require the use of over-parameterized models. In this way, the origin of gauge freedoms in models of sequence-function relationships does mirror the origin of gauge freedoms in physical theories.

There is still much to understand about the relationship between models of sequence-function relationships, the symmetries of these models, and how these models can be biologically interpreted. This paper and its companion [30] have only addressed gauge freedoms and symmetries in linear models. Some work has explored the gauge freedoms and symmetries of nonlinear models of sequence-function relationships [40, 41], but important questions remain. The sloppy modes [42, 43] present in these models are also important to understand, and to our knowledge these have yet to be systematically investigated. Addressing these problems is becoming increasingly urgent due to the expanding interest in interpretable quantitative models of sequence-function relationships [e.g., 44].

See Supplemental Information for full derivations of the mathematical results presented above. Python code implementing the embedding distillation algorithm described the section "Computational analysis of generalized one-hot models" and used for generating Fig. 4 is available at `https://github.com/jbkinney/24_posfai2`.

ACKNOWLEDGMENTS

We thank Peter Koo and Vijay Balasubramanian for helpful discussions. This work was supported by NIH grant R35 GM133613 (AP, DMM), NIH grant R35 GM133777 (AP, JBK), NIH grant R01 HG011787 (JBK), the Alfred P. Sloan foundation (DMM), and additional funding from the Simons Center for Quantitative Biology at CSHL (DMM, JBK).

[1] J. B. Kinney and D. M. McCandlish, Massively parallel assays and quantitative sequence-function relationships, Annual Review of Genomics and Human Genetics **20**, 99 (2019), wrote.

[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, Proceedings of the National Academy of Sciences **106**, 67 (2009), 0901.1248.

[3] 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).

[4] M. Ekeberg, C. L ovkvist, Y. Lan, M. Weigt, and E. Aurell, Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models, Physical Review E **87**, 012707 (2013), 1211.1281.

[5] M. Ekeberg, T. Hartonen, and E. Aurell, Fast pseudolikelihood maximization for direct-coupling analysis of protein structure from many homologous amino-acid sequences, Journal of Computational Physics **276**, 341 (2014), 1401.4832.

[6] R. R. Stein, D. S. Marks, and C. Sander, Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models, PLoS Computational Biology **11**, e1004182 (2015).

[7] J. B. Kinney, G. Tkacik, and C. G. Callan, Precise physical models of protein-DNA interaction from high-throughput data, Proceedings of the National Academy of Sciences **104**, 501 (2007), wrote.

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

[9] A. Haldane, W. F. Flynn, P. He, and R. M. Levy, Coevolutionary Landscape of Kinase Family Proteins: Sequence Probabilities and Functional Motifs, Biophysical Journal **114**, 21 (2018).

[10] S. Cocco, C. Feinauer, M. Figliuzzi, R. Monasson, and M. Weigt, Inverse statistical physics of protein sequences: a key issues review, Reports on Progress in Physics **81**, 032601 (2018), 1703.01222.

[11] A. Haldane and R. M. Levy, Influence of multiple-sequence-alignment depth on Potts statistical models of protein covariation, Physical Review E **99**, 032405 (2019), 1812.04162.

[12] H. T. Rube, C. Rastogi, S. Feng, J. F. Kribelbauer, A. Li, B. Becerra, L. A. N. Melo, B. V. Do, X. Li, H. H. Adam, N. H. Shah, R. S. Mann, and H. J. Bussemaker, Probing molecular specificity with deep sequencing and biophysically interpretable machine learning, bioRxiv , 2021.06.30.450414 (2021).

[13] S. Zamuner and P. D. L. Rios, Interpretable Neural Networks based classifiers for categorical inputs, arXiv 10.48550/arxiv.2102.03202 (2021), 2102.03202.

[14] C. Feinauer, B. Meynard-Piganeau, and C. Lucibello, Interpretable pairwise distillations for generative protein sequence models, PLoS Computational Biology **18**, e1010219 (2022).

[15] A. Gerardos, N. Dietler, and A.-F. Bitbol, Correlations from structure and phylogeny combine constructively in the inference of protein partners from sequences, PLoS Computational Biology **18**, e1010147 (2022), 2111.11373.

[16] C. Hsu, H. Nisonoff, C. Fannjiang, and J. Listgarten, Learning protein fitness models from evolutionary and assay-labeled data, Nature Biotechnology **40**, 1114 (2022).

[17] C. Feinauer and E. Borgonovo, Mean Dimension of Generative Models for Protein Sequences, bioRxiv , 2022.12.12.520028 (2022).

[18] H. T. Rube, C. Rastogi, S. Feng, J. F. Kribelbauer, A. Li, B. Becerra, L. A. N. Melo, B. V. Do, X. Li, H. H. Adam, N. H. Shah, R. S. Mann, and H. J. Bussemaker, Prediction of protein-ligand binding affinity from sequencing data with interpretable machine learning, Nature Biotechnology **40**, 1520 (2022).

[19] E. D. Weinberger, Fourier and taylor series on fitness landscapes, Biological cybernetics **65**, 321 (1991).

[20] C. T. Zhang and R. Zhang, Analysis of distribution of bases in the coding sequences by a diagrammatic technique., Nucleic acids research **19**, 6313 (1991).

[21] P. F. Stadler, Spectral landscape theory, in *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality and Function*, edited by J. Crutchfield and P. Schuster (Oxford Univ. Press, Oxford, 2003) pp. 231 – 271.

[22] G. D. Stormo, Maximally efficient modeling of DNA sequence motifs at all levels of complexity., Genetics **187**, 1219 (2011-04).

[23] F. J. Poelwijk, V. Krishna, and R. Ranganathan, The Context-Dependence of Mutations: A Linkage of Formalisms, PLOS Computational Biology **12**, e1004771 (2016), 1502.00726.

[24] D. H. Brookes, A. Aghazadeh, and J. Listgarten, On the sparsity of fitness functions and implications for learning, Proceedings of the National Academy of Sciences **119**, e2109649118 (2022).

[25] A. Tareen, M. Kooshkbaghi, A. Posfai, W. T. Ireland, D. M. McCandlish, and J. B. Kinney, MAVE-NN: learning genotype-phenotype maps from multiplex assays of variant effect, Genome Biology **23**, 98 (2022).

[26] J. D. Jackson and L. B. Okun, Historical roots of gauge invariance, Reviews of Modern Physics **73**, 663 (2001), hep-ph/0012061.

[27] J. D. Jackson, *Classical electrodynamics* (John Wiley & Sons, 1998).

[28] Results in quantum physics, such as the Aharanov-Bohm effect [45, 46], suggest a reality to the four-potential beyond what can be inferred solely from classical E&M, though there are arguments against this interpretation [47].

[29] B. E. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, 2nd ed., Graduate Texts in Mathematics (Springer, 2001) read in early 2022.

[30] A. Posfai, J. Zhou, D. M. McCandlish, and J. B. Kinney, Gauge fixing for sequence-function relationships, bioRxiv 10.1101/2024.05.12.593772 (2024).

[31] Our analysis is readily extended to models in which $\vec{x}$ and $\vec{\theta}$ are complex vectors [e.g. 24]. All the results in Supplemental Information are, in fact, derived for this more general class of models. Here we restrict our discussion to the reals only to simplify the presentation.

[32] R. Happel and P. F. Stadler, Canonical approximation of fitness landscapes, Complexity **2**, 53 (1996).

[33] J. B. Kinney, A. Murugan, C. G. Callan, and E. C. Cox, Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence, Proceedings of the National Academy of Sciences **107**, 9158 (2010), wrote.

[34] C. A. Olson, N. C. Wu, and R. Sun, A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain., Current biology : CB **24**, 2643 (2014).

[35] P. Baeza-Centurion, B. Miñana, J. M. Schmiedel, J. Valcárcel, and B. Lehner, Combinatorial genetics reveals a scaling law for the effects of mutations on splicing, Cell **176**, 549 (2019).

[36] S. Kawashima, P. Pokarowski, M. Pokarowska, A. Kolinski, T. Katayama, and M. Kanehisa, Aaindex: amino acid index database, progress report 2008, Nucleic acids research **36**, D202 (2007).

[37] K. K. Yang, Z. Wu, C. N. Bedbrook, and F. H. Arnold, Learned protein embeddings for machine learning, Bioinformatics **34**, 2642 (2018).

[38] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig, The role of DNA shape in protein-DNA recognition., Nature **461**, 1248 (2009).

[39] L. Yang, T. Zhou, I. Dror, A. Mathelier, W. W. Wasserman, R. Gordân, and R. Rohs, TFBSshape: a motif database for DNA shape features of transcription factor binding sites, Nucl Acids Res **42**, D148 (2013).

[40] J. B. Kinney and G. S. Atwal, Parametric Inference in the Large Data Limit Using Maximally Informative Models, Neural computation **26**, 637 (2014-04), wrote.

[41] G. S. Atwal and J. B. Kinney, Learning Quantitative Sequence–Function Relationships from Massively Parallel Experiments, Journal of Statistical Physics **162**, 1203 (2016), wrote, 1506.00054.

[42] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Parameter space compression underlies emergent theories and predictive models., Science **342**, 604 (2013), 1303.6738.

[43] M. K. Transtrum, B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna, Perspective: Sloppiness and emergent theories in physics, biology, and beyond, The Journal of Chemical Physics **143**, 010901 (2015).

[44] E. Seitz, D. M. McCandlish, J. B. Kinney, and P. K. Koo, Interpreting cis-regulatory mechanisms from genomic deep neural networks using surrogate models, bioRxiv 10.1101/2023.11.14.567120 (2023).

[45] Y. Aharonov and D. Bohm, Significance of electromagnetic potentials in the quantum theory, Physical review **115**, 485 (1959).

[46] M. Peshkin and A. Tonomura, *The Aharonov-Bohm Effect* (Springer Verlag, 2005).

[47] L. Vaidman, Role of potentials in the aharonov-bohm effect, Physical Review A **86**, 040101 (2012).