

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

Anna Posfai<sup>1</sup>, David M. McCandlish<sup>1</sup>, and Justin B. Kinney<sup>1,†</sup>

<sup>1</sup>Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724

This manuscript was compiled on May 13, 2024

**Quantitative models of sequence-function relationships, which 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. Here 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 the transformations of model parameters that compensate for these symmetry transformations are described by redundant irreducible matrix representations. Based on this finding, we describe an “embedding distillation” procedure that enables analytic calculation of the dimension of the space of gauge freedoms, as well as efficient computation of a sparse basis for this space. Finally, we show that the ability to interpret model parameters as quantifying allelic effects places strong constraints on the form that models can take, and in particular show that all nontrivial equivariant models of allelic effects must exhibit gauge freedoms. Our work thus advances the understanding of the relationship between symmetries and gauge freedoms in quantitative models of sequence-function relationships.**

sequence-function relationships | gauge freedoms | sequence space | permutation symmetry | representation theory

## Introduction

Understanding the quantitative nature of sequence-function relationships is a major goal of modern biology (1). To study a specific 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 often complicated, however, by gauge freedoms—directions in parameter space along which model parameters can be changed without altering model predictions. If any gauge freedoms are present in a model, the numerical values of individual model parameters cannot be meaningfully interpreted in the absence of additional constraints.

Researchers performing quantitative studies of sequence-function relationships routinely encounter gauge freedoms in their models. In practice, one of two methods is typically used to overcome the difficulties that such gauge freedoms can 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 (19–24). But despite being frequently encountered in the course of research, the gauge freedoms present in models of

sequence-function relationships have received little attention (though see e.g. 3, 5–7, 12, 25). In particular, the mathematical properties of these gauge freedoms have yet to be systematically studied.

In physics, by contrast, gauge freedoms are a topic of fundamental importance (26). Gauge freedoms are well-known to 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., 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 E&M in a form that is manifestly Lorentz invariant, one must instead formulate the equations 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\*. Nevertheless, working with the four-potential greatly simplifies the equations of E&M and often aids in both their solution and their physical interpretation.

Motivated by the connection between gauge freedoms and symmetries in physics, we investigated whether the gauge freedoms in mathematical models of sequence-function relationships have a connection to the symmetries of sequence space. Here 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. These models include many of the most commonly used models, including models with pairwise and/or higher-order interactions. Using techniques from the theory of matrix representations, we find that the gauge freedoms of these models arise when model parameters transform under redundant irreducible matrix representations of the symmetry group. Based on this finding, we describe an “embedding distillation” procedure that facilitates the analysis of the vector space formed by the gauge freedoms of a large class of commonly used models.

Finally, we investigate the connection between parameter interpretability and model transformation behavior. We show that the ability to interpret the individual parameters of an equivariant model as quantifying the effects of specific alleles requires that these parameters transform under a permutation representation of the symmetry group, rather than a more general matrix representation. A consequence is that

\*Results in quantum physics, such as the Aharonov-Bohm effect (28, 29), suggest a reality to the four-potential beyond what can be inferred solely from classical E&M, though there are arguments against this interpretation (30).

Please provide details of author contributions here.

Please declare any competing interests here.

† Correspondence: [jkinney@cshl.edu](mailto:jkinney@cshl.edu) (JBK)

all nontrivial models that satisfy this interpretation criterion have gauge freedoms. This shows in particular that models that have gauge freedoms can have important advantages over mathematically equivalent models that do not have gauge freedoms. A companion paper (31) reports specific gauge-fixing strategies that can be applied to the most commonly used models that can be interpreted as quantifying allelic effects.

## 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 (31) provides an expanded discussion of these results together with corresponding proofs.

**Sequence-function relationships.** Let  $\mathcal{A}$  denote an alphabet comprising  $\alpha$  distinct characters. Let  $\mathcal{S}$  denote the set of  $\alpha^L$  sequences of length  $L$  built from these characters. A model of a sequence-function relationship,  $f(s; \vec{\theta})$ , is defined to be a function that maps each sequence  $s \in \mathcal{S}$  to a complex number. The vector  $\vec{\theta}$  denotes the parameters of the model and is assumed to comprise  $M$  complex numbers.

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

$$f(s; \vec{\theta}) = \vec{\theta}^\dagger \vec{x}(s) = \sum_{i=1}^M \theta_i x_i(s), \quad [1]$$

where  $\vec{x}(\cdot)$  is a vector of  $M$  distinct sequence features, and each feature  $x_i(\cdot)$  is a function that maps sequences in  $\mathcal{S}$  to the complex numbers. We refer to the space  $\mathbb{C}^M$  in which these feature vectors live as feature space, and each specific feature vector  $\vec{x}(s)$  as the embedding of sequence  $s$ .

Note that we let both sequence embeddings  $\vec{x}$  and model parameters  $\vec{\theta}$  be complex. By contrast, ref. (31) limited embeddings and parameters to the reals. We choose here to work in complex spaces because, in addition to the added generality of the results, the algebraic completeness of the complex numbers simplifies some of our proofs. All of our results, however, hold if the parameters and embeddings are restricted to the reals. See SI Sec. 10 for details.

**Generalized one-hot (GO) models.** GO models are linear models in which the sequence features indicate only the presence or absence of specific characters at specific positions (1). An example of a GO is the pairwise-interaction model, which has the form

$$f_{\text{pair}}(s) = \theta_0 x_0(s) + \sum_l \sum_c \theta_l^c x_l^c(s) + \sum_{l < l'} \sum_{c, c'} \theta_{ll'}^{cc'} x_{ll'}^{cc'}(s), \quad [2]$$

where  $l, l' \in \{1, \dots, L\}$  index positions within sequences  $s$  and  $c, c' \in \mathcal{A}$  index characters at these positions. Pairwise-interaction models comprise three types of GO feature: the constant feature,  $x_0(s)$ , which equals one for every sequence  $s$ ; additive features,  $x_l^c(s)$ , which equal one if  $s_l = c$  and equal zero otherwise (where  $s_l$  denotes the character at position  $l$  in sequence  $s$ ); and pairwise features,  $x_{ll'}^{cc'}(s)$ , which equal one if both  $s_l = c$  and  $s_{l'} = c'$ , and which equal zero otherwise.

GO models are defined in a similar manner: as sums of terms that each have the form

$$\theta_{l_1 l_2 \dots l_K}^{c_1 c_2 \dots c_K} x_{l_1 l_2 \dots l_K}^{c_1 c_2 \dots c_K}(s). \quad [3]$$

Here,  $K \in \{0, \dots, L\}$  is a term-specific number,  $\{l_1, l_2, \dots, l_K\}$  is a term-specific set of positions, and  $\{c_1, c_2, \dots, c_K\}$  is a term-specific set of characters at the corresponding positions. Each feature  $x_{l_1 l_2 \dots l_K}^{c_1 c_2 \dots c_K}(s)$  is a  $K$ -order one-hot feature defined to be equal to one if  $s_k = c_k$  for all  $k \in \{1, \dots, K\}$  and equal to zero otherwise. For example, the pairwise-interaction model is a GO model that contains a  $K = 0$  term<sup>†</sup> as well as all possible terms of order  $K = 1$  and  $K = 2$ .

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

$$f(s; \vec{\theta}) = f(s; \vec{\theta} + \vec{g}) \quad \text{for all } s \in \mathcal{S}. \quad [4]$$

For linear sequence-function relationships, the set of gauge freedoms  $G$  is a vector space in  $\mathbb{C}^M$ .  $G$  is the orthogonal complement of the space spanned by sequence embeddings, which we denote by  $\text{span} \vec{x}$  (31). In what follows, we use  $\gamma$  to represent the dimension of  $G$ , and often refer to this quantity somewhat informally as the number of gauge freedoms.

Gauge freedoms arise from linear dependencies among sequence features. For example, one-hot pairwise-interaction models have  $M = 1 + \alpha L + \binom{L}{2} \alpha^2$  parameters, but  $\text{span} \vec{x}$  has only  $1 + (\alpha - 1)L + \binom{L}{2} (\alpha - 1)^2$  dimensions due the presence of  $L + \binom{L}{2} (2\alpha - 1)$  constraints on the embedding. Specifically,  $x_0(s) = \sum_{c'} x_l^{c'}(s)$  for all positions  $l$  (yielding 1 constraint per position), and both  $x_l^c(s) = \sum_{c'} x_{ll'}^{cc'}(s)$  and  $x_{ll'}^{cc'}(s) = \sum_{c''} x_{ll'}^{cc''}(s)$  for all characters  $c$  and for all pairs of positions  $l < l'$  [yielding  $2\alpha - 1$  independent constraints per pair of positions (31)]. The one-hot pairwise interaction model therefore has  $\gamma = L + \binom{L}{2} (2\alpha - 1)$  gauge freedoms; See also (3, 5, 7, 10).

**Fixing the gauge.** Fixing the gauge is the process of removing gauge freedoms by restricting  $\vec{\theta}$  to a subset of parameter space,  $\Theta$ , called the gauge. Linear gauges are choices of  $\Theta$  that are themselves vector spaces. One useful property of linear gauges is that gauge-fixing can be accomplished by projection. Specifically, for any linear gauge  $\Theta$ , there exists a projection matrix  $P$  that projects each parameter vector  $\vec{\theta} \in \mathbb{C}^M$  onto an equivalent parameter vector  $\vec{\theta}_{\text{fixed}}$  that lies in  $\Theta$ , i.e.,

$$\vec{\theta}_{\text{fixed}} = P \vec{\theta}. \quad [5]$$

Given  $\Theta$ , the projection matrix  $P$  is uniquely defined by the requirement that  $P$  is idempotent, the image  $P$  is  $\Theta$ , and the kernel of  $P$  is  $G$ . Our companion paper (31) describes a parametric family of linear gauges (including an explicit formula for the projection matrix) that includes many of the most commonly used gauges as special cases.

## Results

In what follows, we define the group of position-specific character permutations, as well as the linear models that are equivariant under this group. Next, we use methods from the

<sup>†</sup> Here and in what follows,  $K = 0$  corresponds to feature  $x_0$  and parameter  $\theta_0$ .

170 theory of group representations (32) to identify all possible  
 171 equivariant linear models. In the process, we also describe a  
 172 procedure we call “embedding distillation” that allows one to  
 173 compute the gauge freedoms of any equivariant linear model.  
 174 After demonstrating embedding distillation on the one-hot  
 175 pairwise-interaction model, we apply embedding distillation to  
 176 GEO models and derive expressions for the number of gauge  
 177 freedoms (i.e., the dimension of the space of gage freedoms)  
 178 of a variety of commonly used models. Finally, we explore  
 179 the relationship between model transformation behavior and  
 180 parameter identifiability.

181 **Position-specific character permutations.** Let  $H_{\text{CP}}^l$  denote the  
 182 group of permutations among the  $\alpha$  possible characters at  
 183 position  $l$  in a sequence. Note that  $H_{\text{CP}}^l$  is isomorphic to  
 184 the symmetric group on  $\alpha$  elements,  $S_\alpha$  (32). The group of  
 185 position-specific character permutations is given by the direct  
 186 product of all  $H_{\text{CP}}^l$ , i.e.,

$$187 \quad H_{\text{PSCP}} = H_{\text{CP}}^1 \times \cdots \times H_{\text{CP}}^L. \quad [6]$$

188 Given any  $h \in H_{\text{PSCP}}$ , the transformation of a sequence  $s$  by  
 189  $h$  is written as  $hs$ , and the transformation of sequence space  
 190  $\mathcal{S}$  by  $h$  is written as  $h\mathcal{S}$ .

191 **Equivariant embeddings and equivariant models.** A representa-  
 192 tion  $R$  of a group  $H$  is a function that maps each  $h \in H$  to  
 193 a complex matrix  $R(h)$  such that  $R(h_1 h_2) = R(h_1)R(h_2)$  for  
 194 all  $h_1, h_2 \in H$ . In what follows, we say that an embedding  $\vec{x}$   
 195 is “equivariant” if and only if there is a representation  $R$  of  
 196  $H_{\text{PSCP}}$  such that

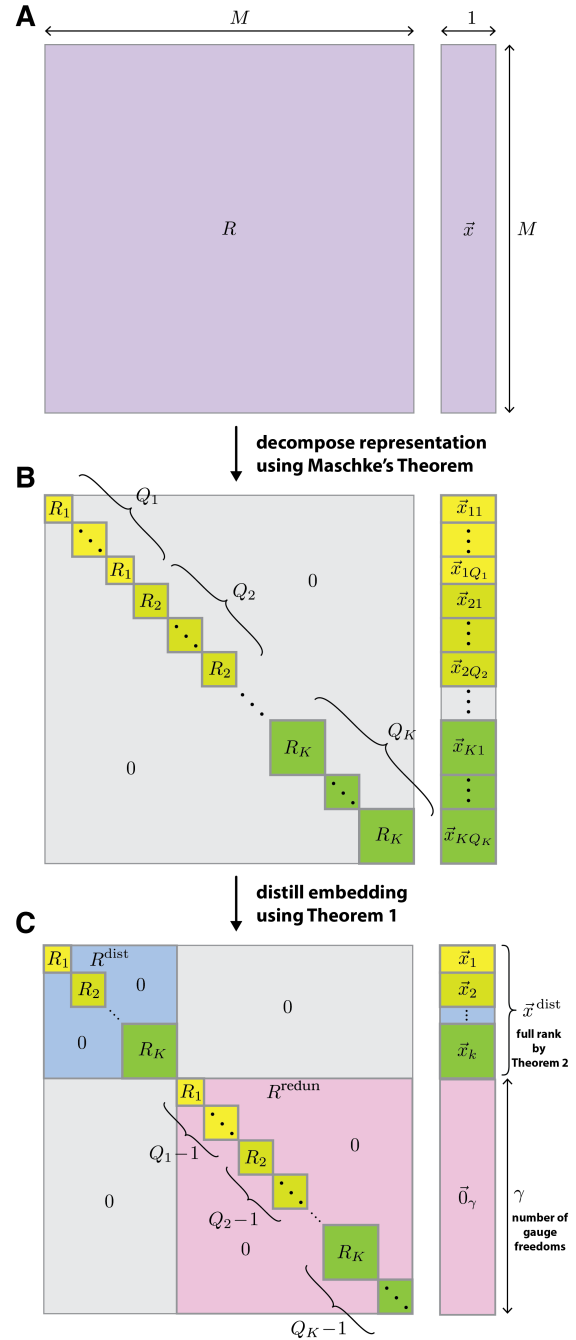
$$197 \quad \vec{x}(hs) = R(h)\vec{x}(s) \quad [7]$$

198 for all  $h \in H_{\text{PSCP}}$  and all  $s \in \mathcal{S}$  (Fig. 1A). We also say that a  
 199 linear model is equivariant if and only if it is defined with an  
 200 equivariant embedding  $\vec{x}$  in Eq. 1. For any equivariant model,  
 201 the transformation of sequence space by any  $h \in H_{\text{PSCP}}$  can be  
 202 compensated for by a corresponding transformation of model  
 203 parameters. Specifically, the sequence-space transformation  
 204  $\mathcal{S} \rightarrow h\mathcal{S}$ ,  $h \in H_{\text{PSCP}}$ , is compensated for by the parameter  
 205 transformation  $\vec{\theta} \rightarrow R(h)^{-1}\vec{\theta}$ , in the sense that  $f(s; \vec{\theta}) =$   
 206  $f(hs; R(h)^{-1}\vec{\theta})$  for every  $s \in \mathcal{S}$  and  $\vec{\theta} \in \mathbb{C}^M$  (see SI Sec.  
 207 3.2). Using terminology from representation theory, every  
 208  $R(h)$  is an  $M \times M$  matrix where  $M$  is called the degree of  $R$   
 209 (denoted  $\text{deg } R$ ). Similarly  $\vec{x}(s)$  is an  $M$ -dimensional vector,  
 210 where  $m$  is called the degree of  $\vec{x}$  (denoted  $\text{deg } \vec{x}$ ).

211 **Maschke decomposition of equivariant embeddings.** Every  
 212 group representation is either reducible or irreducible. A  
 213 representation is irreducible if and only if it has no proper  
 214 invariant subspace. Maschke’s theorem, a basic result in repre-  
 215 sentation theory, says that all representations of finite groups  
 216 are equivalent (i.e., equal up to a similarity transformation) to  
 217 a direct sum of irreducible representations. Any representation  
 218  $R$  of  $H_{\text{PSCP}}$  can therefore be expressed as

$$219 \quad R \simeq \bigoplus_{k=1}^K Q_k R_k, \quad [8]$$

220 where  $\simeq$  denotes equivalence, each  $R_k$  is an irreducible repre-  
 221 sentation of  $H_{\text{PSCP}}$ , all  $R_k$  are pairwise inequivalent, and  $Q_k$   
 222 denotes the multiplicity of  $R_k$  in the direct sum.



**Fig. 1. Embedding distillation.** (A) Given an  $M$ -dimensional embedding  $\vec{x}$  that is equivariant under  $H_{\text{PSCP}}$ , let  $R$  be the representation of  $H_{\text{PSCP}}$  that acts on  $\vec{x}$ . (B) By Maschke’s theorem,  $R$  can be decomposed into a direct sum of irreducible representations,  $R_k$  ( $k \in \{1, \dots, K\}$ ), each of which occurs with multiplicity  $Q_k$  (Eq. 8). Similarly,  $\vec{x}$  can be decomposed into a direct sum of irreducible embeddings  $\vec{x}_{k,q}$  ( $q \in \{1, \dots, Q_k\}$ ), where each  $\vec{x}_{k,q}$  transforms under  $R_k$  (Eq. 9). (C) By Theorem 1, an additional similarity transformation can be performed that, for each value of  $k$ , zeroes out all but one  $\vec{x}_{k,q}$ ; the remaining  $\vec{x}_{k,q}$  is denoted by  $\vec{x}_k$  (Eq. 11 and Eq. 12). Consequently,  $\vec{x}$  decomposes into a direct sum of a distilled embedding,  $\vec{x}^{\text{dist}}$ , and a zero vector,  $\vec{0}_\gamma$ , having dimension  $\gamma$  (Eq. 11).  $\vec{x}^{\text{dist}}$  is given by the direct sum of all  $\vec{x}_k$  (Eq. 12) and is full rank by Theorem 2. The distilled representation,  $R^{\text{dist}}$ , describes how  $\vec{x}^{\text{dist}}$  transforms and contains one copy of each  $R_k$ . The redundant representation,  $R^{\text{redun}}$ , operates on  $\vec{0}_\gamma$ , and encapsulates the  $Q_k - 1$  redundant copies of each  $R_k$ .  $\gamma$ , the degree of  $R^{\text{redun}}$ , is equal to the number of gauge freedoms (Eq. 15).

223 In what follows, we say that a sequence embedding is  
 224 irreducible if and only if it transforms under an irreducible  
 225 representation of  $H_{\text{PSCP}}$ . One consequence of Eq. 8 is that  
 226 any embedding  $\vec{x}$  that transforms under  $R$  can be decomposed  
 227 as

$$228 \quad \vec{x} \simeq \bigoplus_{k=1}^K \bigoplus_{q=1}^{Q_k} \vec{x}_{kq}, \quad [9]$$

229 where each  $\vec{x}_{kq}$  is an irreducible embedding that transforms  
 230 under  $R_k$ . This decomposition is illustrated in Fig. 1B. We  
 231 assume in what follows that all  $\vec{x}_{kq}$  are nonzero, but this  
 232 assumption can be removed without fundamentally changing  
 233 our results.<sup>‡</sup>

234 **Distillation of equivariant embeddings.** We now describe how  
 235 equivariant models are analyzed via the distillation of their  
 236 embeddings. In SI Sec. 5.1, we prove the following:

237 **Theorem 1** *Any two nonzero sequence embeddings that trans-*  
 238 *form under the same irreducible representation of  $H_{\text{PSCP}}$  are*  
 239 *equal up to a constant of proportionality.*

240 Using Theorem 1, then performing additional similarity trans-  
 241 formations to remove the constants of proportionality, we  
 242 obtain,

$$243 \quad \vec{x} \simeq \bigoplus_{k=1}^K Q_k \vec{x}_k, \quad [10]$$

244 where  $\vec{x}_k$  is any one of the  $\vec{x}_{kq}$ , and  $Q_k$  is the multiplicity of  $\vec{x}_k$   
 245 in the direct sum. Additional similarity transformations can  
 246 then be performed to zero out all except one copy of  $\vec{x}_k$ . We  
 247 therefore find that there is an invertible “distillation matrix”  
 248  $T$  such that

$$249 \quad T\vec{x} = \vec{x}^{\text{dist}} \oplus \vec{0}_\gamma, \quad [11]$$

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

$$251 \quad \vec{x}^{\text{dist}} = \bigoplus_{k=1}^K \vec{x}_k, \quad [12]$$

252 is the distilled embedding. Similarly, the matrix representation  
 253  $R$  decomposes as

$$254 \quad TRT^{-1} = \vec{R}^{\text{dist}} \oplus \vec{R}^{\text{redund}} \quad [13]$$

255 where the distilled representation,  $R^{\text{dist}} = \bigoplus_{k=1}^K R_k$ , con-  
 256 tains one copy of each  $R_k$ , and the redundant representation,  
 257  $R^{\text{redund}} = \bigoplus_{k=1}^K (Q_k - 1)R_k$ , contains all of the other copies  
 258 of each  $R_k$  that are present in  $R$ . These decompositions are  
 259 illustrated in Fig. 1C.

260 **Identification of gauge freedoms in equivariant models.** To  
 261 identify the gauge freedoms of any equivariant model, we use  
 262 the fact that  $\vec{x}^{\text{dist}}$  is full rank. This is a consequence of the  
 263 following Theorem, which is proven in SI Sec. 3.4:

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

<sup>‡</sup>See SI Sec. 5.2 for a statement of our main results when this assumption is removed.

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

$$268 \quad \vec{g} = T^\dagger [\vec{0}_{M-\gamma} \oplus \vec{g}_\gamma], \quad [14] \quad 270$$

271 for some  $\gamma$ -dimensional vector  $\vec{g}_\gamma$ . The space of gauge transfor-  
 272 mations,  $G$ , is therefore given by the set of vectors having the  
 273 form in Eq. 14. In particular, the number of gauge freedoms  
 274 is,

$$275 \quad \gamma = \deg \vec{x} - \deg \vec{x}^{\text{dist}} = \deg R^{\text{redund}}. \quad [15]$$

276 We thus see that the number of gauge freedoms is equal to the  
 277 sum of the degrees of all redundant irreducible representations  
 278 in  $R$ .

279 From Eq. 14, we also see that  $G$  is spanned by the last  $\gamma$   
 280 column vectors of  $T^\dagger$ . One can therefore compute a basis for  $G$   
 281 simply by computing  $T$ , and computing  $T$  only requires keeping  
 282 track of the similarity transformations needed to express  $\vec{x}$  in  
 283 the distilled form shown in Eq. 11.

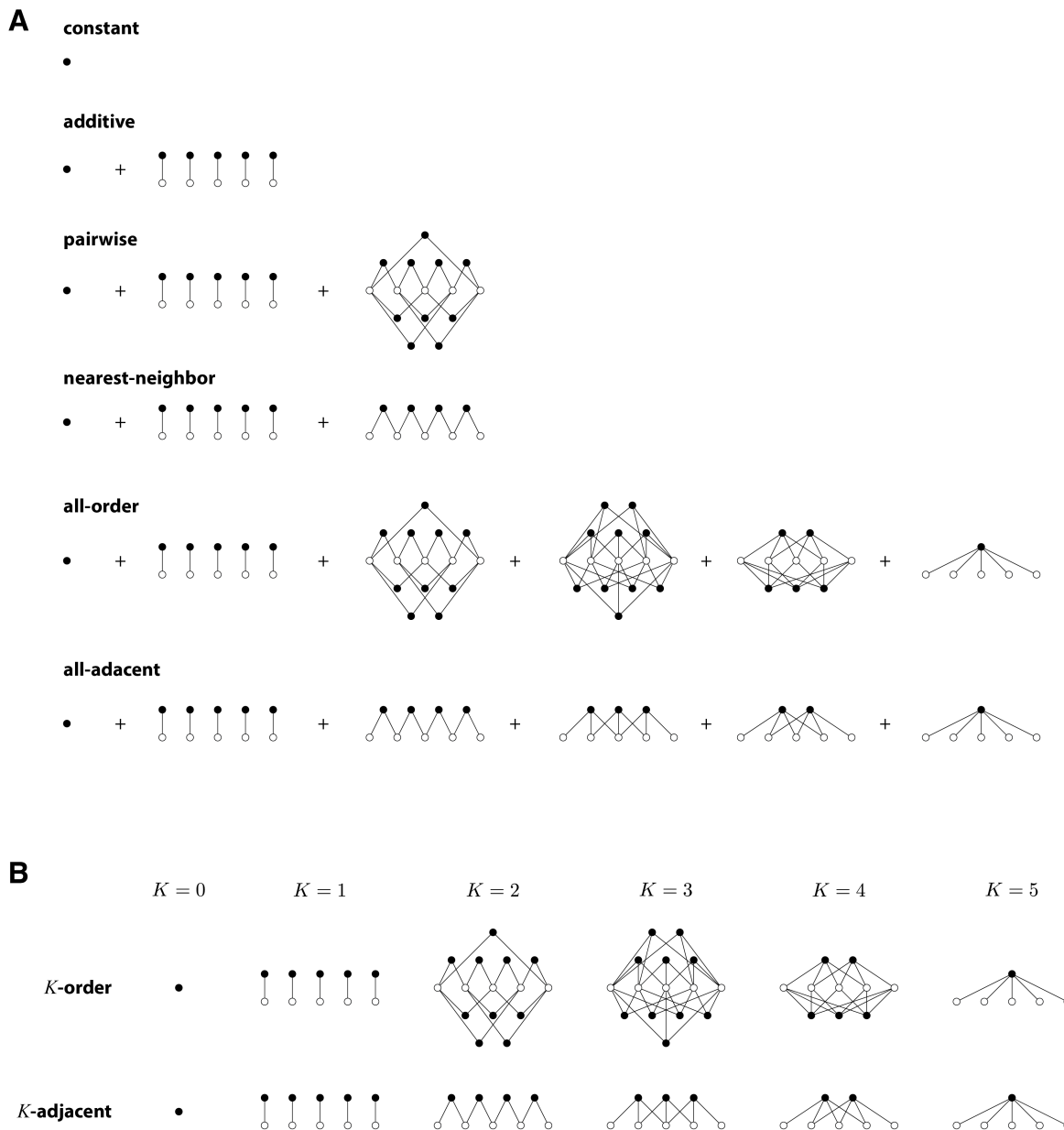
284 **Identification of all equivariant embeddings.** The specific struc-  
 285 ture of  $H_{\text{PSCP}}$  allows us to identify all possible inequivalent  
 286 irreducible equivariant embeddings,  $\vec{x}_k$ . Because  $\vec{x}_k$  is irre-  
 287 reducible and  $H_{\text{PSCP}}$  is a product group,  $\vec{x}_k$  can be expressed  
 288 as

$$289 \quad \vec{x}_k \simeq \bigotimes_{l=1}^L \vec{x}_l^k, \quad [16]$$

290 where each  $\vec{x}_l^k$  is an embedding that depends only on the  
 291 character at position  $l$  and that transforms under an irreducible  
 292 representation of  $H_{\text{CP}}^l$ . Moreover, because  $H_{\text{CP}}^l$  is isomorphic  
 293 to  $S_\alpha$  and  $S_\alpha$  supports only two inequivalent embeddings (see  
 294 SI Sec. 4.3 for proof), there are only two inequivalent choices  
 295 for each  $\vec{x}_l^k$ : the trivial embedding and the simplex embedding.  
 296 The trivial embedding, denoted  $\vec{x}^{\text{triv}}$ , maps every sequence to  
 297 a one-dimensional vector and transforms under what is called  
 298 the “trivial representation” of  $S_\alpha$ . The simplex embedding,  
 299 denoted  $\vec{x}_l^{\text{sim}}$ , maps sequences to the  $\alpha$  vertices of an  $\alpha - 1$   
 300 dimensional simplex and transforms under what is called the  
 301 “standard representation” of  $S_\alpha$ . One example of the simplex  
 302 embedding is the tetrahedral embedding of DNA and RNA  
 303 (20, 22). Note: to lessen the notational burden in what follows,  
 304 we avoid writing  $\vec{x}^{\text{triv}}$  within tensor products over positions  
 305  $l$ , and only show factors that contribute nontrivially to these  
 306 products.

307 We now identify all equivariant embeddings  $\vec{x}$ . Because  
 308 there are 2 inequivalent choices for each  $\vec{x}_l^k$  ( $\vec{x}^{\text{triv}}$  or  $\vec{x}_l^{\text{sim}}$ ),  
 309 there are  $2^L$  inequivalent choices for  $\vec{x}_k$ , and thus  $\binom{2^L}{K}$  in-  
 310 equivalent choices for the set  $\{\vec{x}_k\}_{k=1}^K$ . Letting  $K$  in Eq. 12  
 311 range from 0 to  $2^L$ , we find that there are  $\sum_{K=0}^{2^L} \binom{2^L}{K} = 2^{2^L}$   
 312 inequivalent choices for  $\vec{x}^{\text{dist}}$ . Every equivariant embedding  $\vec{x}$   
 313 can therefore be expressed, using one of these  $2^{2^L}$  inequivalent  
 314 distilled embeddings  $\vec{x}^{\text{dist}}$  together with a zero vector  $\vec{0}_\gamma$  and  
 315 an invertible matrix  $T$ . Conversely, choosing any of the  $2^{2^L}$   
 316 inequivalent distilled embeddings  $\vec{x}^{\text{dist}}$ , any non-negative inte-  
 317 ger  $\gamma$ , and any invertible matrix  $T$  of the appropriate size will  
 318 yield an equivariant embedding  $\vec{x}$  via Eq. 11. We therefore  
 319 find that, modulo the choice of the similarity matrix  $T$  and  
 320 number of gauge freedoms  $\gamma$ , there are  $2^{2^L}$  distinct choices for  
 321  $\vec{x}$ .





**Fig. 2.** Structure of GEO models. (A,B) Models analyzed in Table 1, illustrated for  $L = 5$ . Open circles represent sequence positions. Closed circles represent sets of parameters that are closed under the action of  $H_{PSCP}$ , as in Eq. 23. Edges indicate position indices shared by all the parameters within each closed set. (A) Structure of specific models of interest. (B) Structure of  $K$ -order models and  $K$ -adjacent models for various interaction orders  $K$ .

322 **Analytical analysis of pairwise-interaction models.** We now  
 323 demonstrate the embedding distillation procedure on the  
 324 pairwise-interaction model. First we specify the pairwise-  
 325 interaction embedding,  $\vec{x}_{\text{pair}}$ , as a direct sum of direct products  
 326 of simpler embeddings:

$$327 \quad \vec{x}_{\text{pair}} = \vec{x}^{\text{triv}} \oplus \left\{ \bigoplus_l \vec{x}_l^{\text{ohc}} \right\} \oplus \left\{ \bigoplus_{l < l'} \vec{x}_l^{\text{ohc}} \otimes \vec{x}_{l'}^{\text{ohc}} \right\}, \quad [17]$$

where  $\vec{x}_l^{\text{ohc}}$  is a position-specific one-hot embedding of dimension  $\alpha$  given by

$$328 \quad \vec{x}_l^{\text{ohc}}(s) = \begin{bmatrix} x_l^{c_1}(s) \\ \vdots \\ x_l^{c_\alpha}(s) \end{bmatrix} \quad [18] \quad 330$$

331 for all sequences  $s$ , where  $c_1, \dots, c_\alpha$  denote the elements of  $\mathcal{A}$ .  
 332 The number of model parameters is equal to the dimension  
 333 of  $\vec{x}_{\text{pair}}$ , which is seen from Eq. 17 to be  $\text{deg } \vec{x}_{\text{pair}} = 1 + L\alpha +$   
 334  $\binom{L}{2}\alpha^2$ .

335 The gauge freedoms of pairwise-interaction models arise  
 336 because  $\vec{x}_{\text{pair}}$  is not full rank. The reduced rank of  $\vec{x}_{\text{pair}}$  is  
 337 a consequence of the fact that  $\vec{x}_l^{\text{ohc}}$  is reducible. To derive a

model type	interaction orders	no. parameters ( $M$ )	no. gauge freedoms ( $\gamma$ )
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	$1 + L\alpha + (L - 1)\alpha^2$	$L + (L - 1)(2\alpha - 1)$
all-order	0, 1, $\dots$ , $L$	$(\alpha + 1)^L$	$(\alpha + 1)^L - \alpha^L$
all-adjacent	0, 1, $\dots$ , $L$	$1 + \frac{\alpha}{(\alpha - 1)^2} [\alpha^{L+1} - (L + 1)\alpha + L]$	$1 + \frac{\alpha}{(\alpha - 1)^2} [2\alpha^L - \alpha^{L-1} - (L + 1)\alpha + L]$
$K$ -order	$K$	$\binom{L}{K}\alpha^K$	$\binom{L}{K}\alpha^K - \sum_{k=0}^{K-1} \binom{L}{k}(\alpha - 1)^k$
hierarchical $K$ -order	0, 1, $\dots$ , $K$	$\sum_{k=0}^K \binom{L}{k}\alpha^k$	$\sum_{n=0}^K \binom{L}{n} [\alpha^n - (\alpha - 1)^n]$
$K$ -adjacent <sup>†</sup>	$K$	$(L - K + 1)\alpha^K$	$(L - K)\alpha^{K-1}$
hierarchical $K$ -adjacent <sup>†</sup>	0, 1, $\dots$ , $K$	$1 + \sum_{k=1}^K (L - k + 1)\alpha^k$	$(L - K)\alpha^{K-1} + 1 + \sum_{k=1}^{K-1} (L - k + 1)\alpha^k$

**Table 1. Number of parameters and gauge freedoms of various GEO models. Columns show model type, the orders of interaction included in each model, the number of parameters of each model, and the number of gauge freedoms of each model. See SI Sec. 6 for derivations of these results. GEO, generalized equivariant one-hot. <sup>†</sup>Assumes  $K \geq 1$ .**

338 distilled version of  $\vec{x}_{\text{pair}}$  that is full rank, we reexpress  $\vec{x}_l^{\text{ohc}}$   
 339 as a direct sum of irreducible embeddings using

$$340 \quad \vec{x}_l^{\text{ohc}} \simeq \vec{x}^{\text{triv}} \oplus \vec{x}_l^{\text{sim}}; \quad [19]$$

341 see SI Sec. 2.4 for details. Plugging Eq. 19 into Eq. 17, ex-  
 342 panding the direct product, and grouping like terms, we get

$$343 \quad \vec{x}_{\text{pair}} \simeq \left[ 1 + L + \binom{L}{2} \right] \vec{x}^{\text{triv}} \oplus \left\{ \bigoplus_l L \vec{x}_l^{\text{sim}} \right\} \oplus \left\{ \bigoplus_{l < l'} \vec{x}_l^{\text{sim}} \otimes \vec{x}_{l'}^{\text{sim}} \right\}, \quad [20]$$

344 where the scalar coefficients denote the multiplicity of each  
 345 term in the direct sum. Because  $\vec{x}^{\text{triv}}$ ,  $\vec{x}_l^{\text{sim}}$ , and  $\vec{x}_l^{\text{sim}} \otimes \vec{x}_{l'}^{\text{sim}}$   
 346 are irreducible and pairwise inequivalent, the distillation of  
 347  $\vec{x}_{\text{pair}}$  is seen from Eq. 20 to be

$$348 \quad \vec{x}_{\text{pair}}^{\text{dist}} = \vec{x}^{\text{triv}} \oplus \left\{ \bigoplus_l \vec{x}_l^{\text{sim}} \right\} \oplus \left\{ \bigoplus_{l < l'} \vec{x}_l^{\text{sim}} \otimes \vec{x}_{l'}^{\text{sim}} \right\}. \quad [21]$$

349 From this we observe that  $\deg \vec{x}_{\text{pair}}^{\text{dist}} = 1 + L(\alpha - 1) + \binom{L}{2}(\alpha - 1)^2$ .  
 350 The number of gauge freedoms then follows from Eq. 15:

$$351 \quad \gamma = L + \binom{L}{2}(2\alpha - 1), \quad [22]$$

352 which matches the well-known result (3).

353 **Generalized equivariant one-hot (GEO) models.** For a GO  
 354 model to be equivariant, it is sufficient for the model to be  
 355 expressible as a sum of equivariant terms, each term of the  
 356 form

$$357 \quad \sum_{c_1 \in \mathcal{A}} \dots \sum_{c_K \in \mathcal{A}} \theta_{l_1 l_2 \dots l_K}^{c_1 c_2 \dots c_K} x_{l_1 l_2 \dots l_K}^{c_1 c_2 \dots c_K}(s), \quad [23]$$

358 for some term-specific choice of  $K$  and term-specific set of  
 359 positions  $\{l_1, \dots, l_K\}$ . Observe that GEO models differ from  
 360 GO models in that, for every set of positions used to define a  
 361 term, a GEO model sums over all possible characters at all  
 362 positions in the set, whereas a GO model need not include  
 363 terms for every possible choice of characters. An example of  
 364 GO models that are not GEO models are those based on  
 365 wild-type embeddings, i.e., embeddings that exclude features  
 366 that involve character-positions pairs that occur in a chosen  
 367 “wild-type” sequence.

The embeddings of GEO models all have the following form. 368  
 Let  $A_j$  denote a set of sequence positions, and let  $\{A_j\}_{j=1}^J$  369  
 denote the sets of positions used to construct an GEO model 370  
 with sequence embedding  $\vec{x}$ . By analogy to Eq. 17,  $\vec{x}$  can then 371  
 be written as 372

$$373 \quad \vec{x} = \bigoplus_{j=1}^J \bigotimes_{l \in A_j} \vec{x}_l^{\text{ohc}}. \quad [24]$$

Because each direct product in Eq. 24 yields an embedding of 374  
 dimension  $\alpha^{|A_j|}$ , the full dimension of  $\vec{x}$  (and thus the number 375  
 of model parameters) is 376

$$377 \quad \deg \vec{x} = \sum_{j=1}^J \alpha^{|A_j|}. \quad [25]$$

**Analytical analysis of GEO models.** Now we derive the cor- 378  
 responding distilled embedding. Using Eq. 19 to decompose 379  
 each  $\vec{x}_l^{\text{ohc}}$  in terms of  $\vec{x}^{\text{triv}}$  and  $\vec{x}_l^{\text{sim}}$ , then expanding each 380  
 tensor product and grouping the resulting terms, we find that 381  
 $\vec{x}$  is given by Eq. 10 where 382

$$383 \quad \vec{x}_k = \bigotimes_{l \in B_k} \vec{x}_l^{\text{sim}}, \quad [26]$$

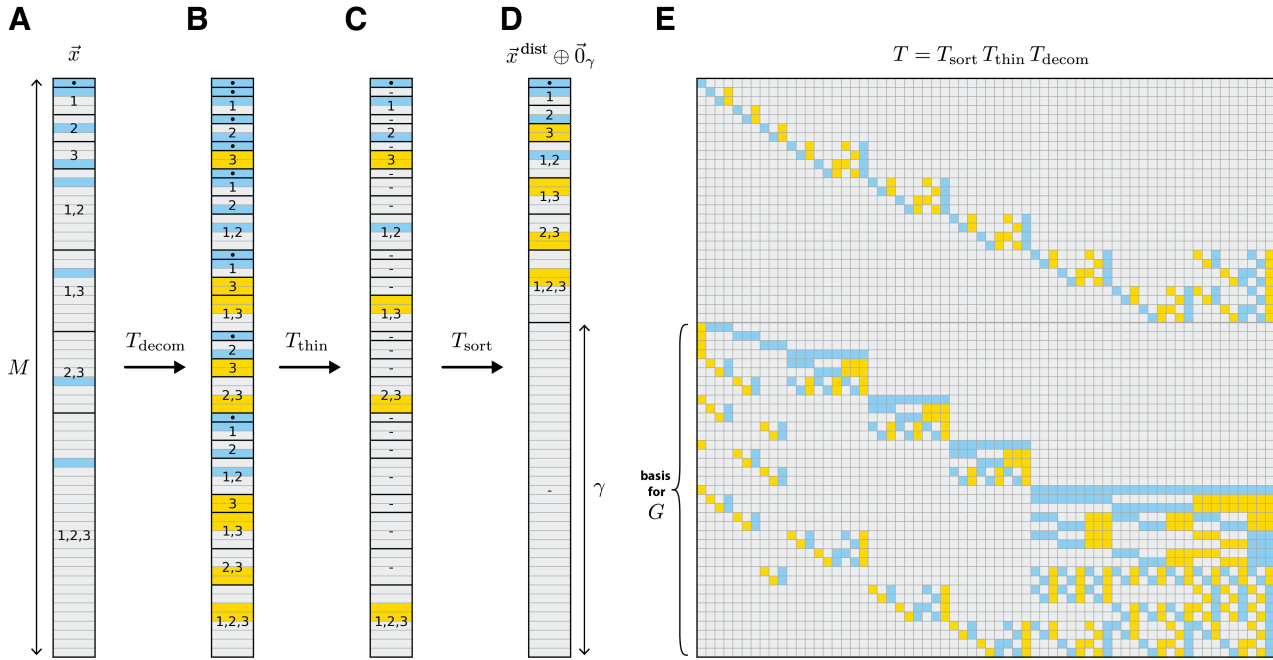
where each  $B_k$  ( $k \in \{1, \dots, K\}$ ) denotes a subset of positions 384  
 that occurs among at least one of the  $A_j$ , and  $Q_k$  denotes the 385  
 number of  $A_j$  in which  $B_k$  occurs.<sup>§</sup> By inspection we see that 386  
 each  $\vec{x}_k$  in Eq. 26 has dimension  $(\alpha - 1)^{|B_k|}$ . Therefore, the 387  
 dimension of  $\vec{x}$  can alternatively be written as 388

$$389 \quad \deg \vec{x} = \sum_{k=1}^K Q_k (\alpha - 1)^{|B_k|}. \quad [27]$$

Every  $\vec{x}_k$  is irreducible because every  $\vec{x}_l^{\text{sim}}$  is irreducible. Con- 390  
 sequently, the distilled embedding  $\vec{x}^{\text{dist}}$  is given by Eq. 12 and 391  
 has dimension 392

$$393 \quad \deg \vec{x}^{\text{dist}} = \sum_{k=1}^K (\alpha - 1)^{|B_k|}. \quad [28]$$

<sup>§</sup>Formally,  $\{B_k\}_{k=1}^K = \bigcup_{j=1}^J P(A_j)$  where  $P(\cdot)$  denotes the powerset, and  $Q_k = \sum_{j=1}^J 1_{P(A_j)}(B_k)$  where  $1_{P(A)}(\cdot)$  is the indicator function for membership in  $P(A)$ .



**Fig. 3.** Illustrated distillation computation. (A) Embedding  $\vec{x}$  of sequence  $s = ABC$  for an all-order interaction model based on the alphabet  $\mathcal{A} = \{A, B, C\}$ . Embedding has degree  $M = 64$ . (B) Result of multiplication by the decomposition matrix,  $T_{\text{decom}}$ . (C) Result of subsequent multiplication by the thinning matrix  $T_{\text{thin}}$ . (D) Result of subsequent multiplication by the sorting matrix  $T_{\text{sort}}$ , which yields  $\vec{x}^{\text{dist}} \oplus \vec{0}_\gamma$  with  $\gamma = 37$  gauge freedoms. In B-D, dots indicate  $\vec{x}^{\text{triv}}$ , dashes indicate zero vectors, and numbers indicate  $\vec{x}_l^{\text{sim}}$  or Kronecker products thereof for specified positions  $l$ . (E) Distillation matrix  $T$  that implements the full distillation procedure in A-D. Last  $\gamma$  rows of  $T$  provide a sparse basis for the gauge space,  $G$ . In A-E, vector and matrix elements are colored using: blue, +1; yellow, -1; gray, 0.

394 Using Eq. 15, the number of gauge freedoms of the embedding  
395  $\vec{x}$  is thus seen to be

$$396 \quad \gamma = \sum_{k=1}^K (Q_k - 1)(\alpha - 1)^{|B_k|}. \quad [29]$$

397 This result provides a way to analytically compute the number  
398 of gauge freedoms of any GEO model. Table 1 reports the  
399 number of gauge freedoms thus computed for a variety of such  
400 models. SI Sec. 6 provides expanded descriptions for each  
401 model, as well as detailed computations of the results in Table  
402 1.

403 We note that the only GEO models that have no gauge  
404 freedoms are those that have embeddings built from only one  
405 tensor product in Eq. 24. To see this, observe from Eq. 29 that  
406  $\gamma = 0$  if and only if none of the  $Q_k$  are greater than 1. This  
407 requires that none of the  $B_k$  are subsets of two or more  $A_j$ .  
408 But the empty set,  $\emptyset$ , is a subset of every  $A_j$ , which means  
409 that  $Q_k = J$  whenever  $B_k = \emptyset$ . Gauge freedoms will therefore  
410 be present unless  $J = 1$ , i.e. the direct sum in Eq. 24 includes  
411 only one term.

412 **Computational analysis of GEO models.** To derive a basis  
413 for the space of gauge freedoms, we must choose a specific  
414 realization of the irreducible embeddings  $\vec{x}^{\text{triv}}$  and  $\vec{x}^{\text{sim}}$ . In

what follows we choose  $\vec{x}^{\text{triv}}(s) = [1]$  and

$$415 \quad \vec{x}_l^{\text{sim}}(s) = \begin{cases} \begin{bmatrix} x_l^{c_1}(s) \\ \vdots \\ x_l^{c_{\alpha-1}}(s) \end{bmatrix} & \text{if } s_l \neq c_\alpha, \\ \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix} & \text{if } s_l = c_\alpha, \end{cases} \quad [30] \quad 416$$

417 for all sequences  $s$ , where  $c_1, \dots, c_\alpha$  represent an ordering of  
418 the characters in  $\mathcal{A}$ . With these choices in hand, Eq. 19 can  
419 be written as an equality:

$$420 \quad T^{(1)} \vec{x}^{\text{ohc}} = \vec{x}^{\text{triv}} \oplus \vec{x}^{\text{sim}}, \quad [31]$$

421 where  $T^{(1)}$  is an  $\alpha \times \alpha$  matrix given by

$$422 \quad T^{(1)} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}. \quad [32]$$

423 Using  $T^{(1)}$  one can compute the distillation matrix for any  
424 GEO model as the product of three matrices:

$$425 \quad T = T_{\text{sort}} T_{\text{thin}} T_{\text{decom}}. \quad [33]$$

426 The effects of these three matrices are illustrated in Fig. 3.  
427 The ‘‘decomposition matrix’’,  $T_{\text{decom}}$ , decomposes the one-hot

embedding  $\vec{x}$  (Fig. 3A) into a direct sum of irreducible embeddings (Fig. 3B). The “thinning matrix”,  $T_{\text{thin}}$ , then zeros out all except the first copy of each irreducible embedding (Fig. 3C). The “sorting matrix”,  $T_{\text{sort}}$ , then rearranges the direct sum so that the remaining nonzero embeddings come first (Fig. 3D). **SI Sec. 8** provides explicit algorithms for constructing  $T_{\text{decom}}$ ,  $T_{\text{thin}}$ , and  $T_{\text{sort}}$ , as well as the inverse of each of these three matrices, for a large class of GEO models. Each of these six matrices has only  $O(L)$  nonzero elements, and the algorithm for constructing each matrix has  $O(L)$  computational complexity. The resulting distillation matrix  $T$ , as well as its inverse, are also sparse. Moreover, every nonzero element of  $T$  is +1 or -1 (Fig. 3E). Because the last  $\gamma$  columns of  $T^\dagger$  provide a basis for  $G$ , we thus obtain a basis for the gauge space consisting of sparse vectors whose only nonzero elements are +1 and -1.

This result can also be used to efficiently fix the gauge of any GEO. Define the projection matrix

$$P = T^\dagger \Big|_{M-\gamma} T^{-1\dagger}, \quad [34]$$

where  $\Big|_{M-\gamma}$  denotes that the last  $\gamma$  columns of a matrix have been set to zero.  $P$  projects parameter vectors  $\vec{\theta}$  onto the space spanned by the first  $M - \gamma$  columns of  $T^\dagger$ . Moreover, by expanding  $P$  as

$$P = T_{\text{decom}}^\dagger T_{\text{thin}}^\dagger T_{\text{sort}}^\dagger I \Big|_{M-\gamma} T_{\text{sort}}^{-1\dagger} T_{\text{thin}}^{-1\dagger} T_{\text{decom}}^{-1\dagger}, \quad [35]$$

and applying each matrix factor to  $\vec{\theta}$  individually, this projection can be performed in  $O(L)$  time. Projection by  $P$  therefore provides an efficient way to remove gauge freedoms by projecting model parameters into a linear gauge. We note, however, that the resulting linear gauge is not one of the parametric gauges discussed in our companion paper (31).

**Interpretability of pairwise-interaction models.** The ability to interpret the parameters of equivariant models as quantifying allelic effects is closely related to how those parameters transform under  $H_{\text{PSCP}}$ . To illustrate this point, we consider two equivariant models: a pairwise-interaction GEO model with embedding  $\vec{x}_{\text{pair}}$ , and the corresponding distilled model with embedding  $\vec{x}_{\text{pair}}^{\text{dist}}$ , both operating on sequences built from a three-character alphabet,  $\mathcal{A} = \{A, B, C\}$ . The two embeddings encode the same set of interactions but do so in different ways:  $\vec{x}_{\text{pair}}$  is built from the single-position one-hot encodings  $\vec{x}_l^{\text{oh}}$ , whereas  $\vec{x}_{\text{pair}}^{\text{dist}}$  is built from the single-position simplex encodings  $\vec{x}_l^{\text{sim}}$  (Fig. 4A). And as we show above, the GEO model has gauge freedoms whereas the distilled model does not.

We now focus on how the features and parameters of these two models are affected by the transformation  $h \in H_{\text{PSCP}}$  that exchanges the characters A and C at all positions  $l$ . For the GEO model,  $h$  induces a permutation of embedding coordinates (Fig. 4B) and thus of model parameters. Consequently,  $h$  preserves the set of values taken by the GEO parameters; it simply permutes which parameters have which values. This mirrors the action of  $h$  on the alleles that drive model predictions:  $h$  permutes sequences and thus the one-position and two-position alleles they contain, but does not alter the full set of alleles present among the full set of sequences. And in fact we see that individual parameter values track their corresponding alleles:  $\theta_l^A$  and  $\theta_l^C$  switch values,  $\theta_{ll}^{\text{AA}}$  and  $\theta_{ll}^{\text{CC}}$  switch

values, etc.. The transformation behavior of the GEO model is therefore consistent with individual parameters quantifying the effects of individual alleles.

For the distilled model, however,  $h$  induces a non-permutation transformation of embedding coordinates (Fig. 4C) and thus of model parameters. Using the embedding shown in Fig. 4A, one finds that the value of  $\theta_l^1$  transforms to  $-\theta_l^1 + \theta_l^2$ , the value of  $\theta_{ll}^{11}$  transforms to  $\theta_{ll}^{11} - \theta_{ll}^{12} - \theta_{ll}^{21} + \theta_{ll}^{22}$ , etc.. The transformation  $h$  therefore changes the full set of values taken by the distilled model parameters. Consequently, the individual parameters of this model cannot be interpreted as quantifying the effects of individual alleles.

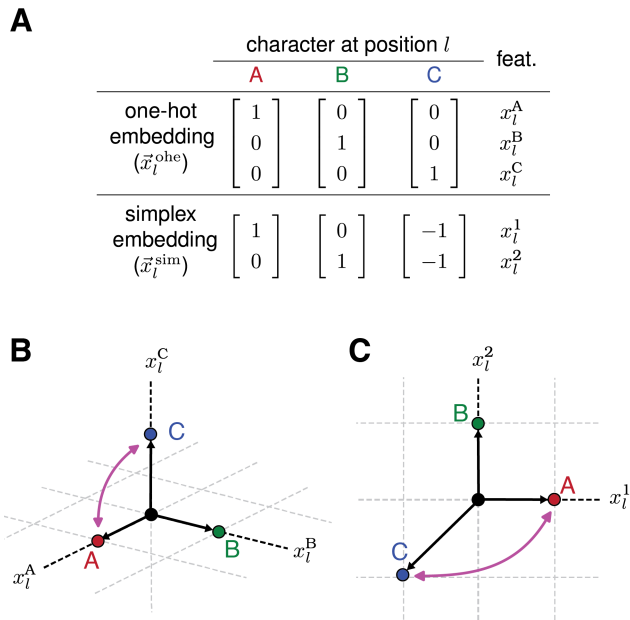
**Nontrivial equivariant allelic models have gauge freedoms.** To clarify the connection between the interpretation and transformation behavior of model parameters, we now formalize the notion of an allele, and allelic effect, and related concepts. We define an allele  $a$  to be a pattern of characters that is either present or absent in every sequence. The corresponding allelic set  $\mathcal{S}_a$  is defined to be the set of sequences that have allele  $a$ , and the corresponding allelic feature  $x_a$  is defined to be the indicator function for membership in  $\mathcal{S}_a$ . An allelic model is defined to be a linear sequence-function relationship in which every feature is an allelic feature. The effect of allele  $a$  is defined, in the context of a specific allelic model, to be the parameter  $\theta_a$  that multiplies the allelic feature  $x_a$ .

Requiring an allelic model to be equivariant puts strong constraints on which alleles it can describe, and on how the corresponding allelic features and allelic effects transform. Given a specific allele  $a$ , the action of  $H_{\text{PSCP}}$  on  $a$  generates a set of alleles  $\mathcal{O}$ , which we call an allelic orbit. If the allelic model is equivariant, the allelic sets  $\mathcal{S}_{a'}$  corresponding to all  $a' \in \mathcal{O}$  will tile sequence space without overlaps. This requirement greatly constrains the set of possible alleles such a model can describe. Moreover, the model must include one feature  $x_{a'}$  for every allele  $a' \in \mathcal{O}$ . These features will then transform among themselves according to a permutation representation. See **SI Sec. 9** for details.

An equivariant allelic model must therefore contain features that can be partitioned into a set of complete allelic orbits. The features of the model will then transform under a direct sum of permutation representations, one for each allelic orbit. Because every permutation representation contains the trivial representation in its Maschke decomposition, the allelic model will have at least as many gauge freedoms as the number of allelic orbits minus one. Perhaps more intuitively, the sum of all allelic features corresponding to each orbit is equal to one for all sequences. Therefore, each orbit’s features are sufficient to represent a constant function on sequence space. Including features from multiple orbits therefore overparameterizes the model and introduces gauge freedoms. We emphasize, however, that additional gauge freedoms can be present as well, so this result only provides a lower bound on  $\gamma$ .

It is readily seen that all GEO models are allelic models. In a GEO model, each allelic orbit corresponds to a position set  $A_j$  in Eq. 24, and the number of allelic orbits is given by  $J$ . Our lower-bound on the number of gauge freedoms recapitulates the finding above that only GEO models with  $J = 1$  have no gauge freedoms. We also show in **SI Sec. 9** that, given a model defined by a direct sum of direct products of single-position embeddings, the corresponding GEO model has the smallest number of gauge freedoms possible.





**Fig. 4.** Transformation behavior of two single-position embeddings. (A) Two single-position embeddings,  $\bar{x}_l^{\text{ohc}}$  and  $\bar{x}_l^{\text{sim}}$ , for the three-character alphabet  $\mathcal{A} = \{A, B, C\}$ . The specific features corresponding to each element of  $\bar{x}_l^{\text{ohc}}$  and  $\bar{x}_l^{\text{sim}}$  are also shown. (B) The three-dimensional one-hot embedding,  $\bar{x}_l^{\text{ohc}}(c)$ , for each  $c \in \mathcal{A}$ . (C) The two-dimensional simplex embedding,  $\bar{x}_l^{\text{sim}}(c)$ , for each  $c \in \mathcal{A}$ . Pink arrows indicate the transformation of each embedding vector induced by permuting characters A and C at position  $l$ .

the dimension of the space of gauge freedoms in a large class of commonly used models, as well as the efficient computation of a sparse basis for this space. From a conceptual standpoint, the results link the gauge freedoms of models of sequence-function relationships to the transformation behavior of these models under a specific symmetry group of sequence space.

We also investigated the link between parameter transformation behavior and parameter interpretability. In doing so, we identified a tension between two different notions of parameter interpretability: in all nontrivial equivariant models, the ability to interpret the values of model parameters in the absence of gauge-fixing constraints is incompatible with the ability to interpret parameters as quantifying allelic effects. Consequently, models that do have gauge freedoms (including nontrivial additive models, pairwise-interaction models, etc.) have important advantages over equally expressive models that do not have gauge freedoms.

We now return to the analogy with theoretical physics. 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 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 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 remain dominant in the literature. We suggest that a major reason for this may be that researchers prefer to use models that both (a) reflect symmetries of sequence space and (b) have parameters that can be interpreted as allelic effects. As we showed, these criteria require the use of over-parameterized models. And 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 modes can be biological interpreted. This paper and its companion (31) have only addressed gauge freedoms and symmetries in linear models of sequence-function relationships. Some work has explored the gauge freedoms and symmetries of nonlinear models of sequence-function relationships (33, 34), but important questions remain. The sloppy modes (35, 36) present in sequence-function relationships are also important to understand but, to our knowledge, these have yet to be systematically studied. Addressing these problems is becoming increasingly urgent,

We therefore see that there is an incompatibility between two distinct notions of parameter interpretability. In all except a limited class of models, the ability to interpret parameters as quantifying allelic effects is incompatible with the ability to interpret parameter values in the absence of gauge-fixing constraints. The only exceptions to this rule are single-orbit allelic models, but these models are trivial in the following sense:<sup>†</sup> each sequence has only one allele, the effect of which is the model's prediction for the sequence. In a single-orbit allelic model, each sequence has only one allele—and thus one feature and one parameter—that contributes to its activity. The parameters are therefore essentially just a catalog of allelic effects. By contrast, the reason researchers quantitatively model sequence-function relationships in the first place is to deconvolve the influence of multiple co-occurring alleles. We conclude that, among nontrivial equivariant models (i.e., models that support co-occurring alleles), the ability to interpret model parameters as quantifying allelic effects requires that the model 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 models that are equivariant under the group of position-specific character permutations (denoted  $H_{\text{PSCP}}$ ), gauge freedoms arise due to model parameters transforming according to redundant irreducible matrix representations of  $H_{\text{PSCP}}$ . From a practical standpoint, this result facilitates the analytic calculation of

<sup>†</sup>This is the same sense in which the "trivial gauge" described in (31) is trivial.

636 not just because of the rapidly expanding use of quantita-  
637 tive models of sequence-function relationships, but also be-  
638 cause of the emerging use of surrogate models for interpreting  
639 sequence-function relationships described by genomic deep  
640 neural networks (37).

## 641 Materials and Methods

642 See Supplemental Information for full derivations of the mathe-  
643 matical results. Python code implementing the embedding distil-  
644 lation algorithm described the section “Computational analysis of  
645 GEO models”, as well as used for generating Fig. 3, is available at  
646 [https://github.com/jbkinney/23\\_posfai](https://github.com/jbkinney/23_posfai).

647 **ACKNOWLEDGMENTS.** We thank Peter Koo and Vijay Bala-  
648 subramanian for helpful discussions. This work was supported by  
649 NIH grant R35 GM133613 (AP, DMM), NIH grant R35 GM133777  
650 (AP, JBK), NIH grant R01 HG011787 (JBK), the Alfred P. Sloan  
651 foundation (DMM), and additional funding from the Simons Center  
652 for Quantitative Biology at CSHL (DMM, JBK).

653 1. JB Kinney, DM McCandlish, Massively parallel assays and quantitative sequence-function  
654 relationships. *Annu. Rev. Genomics Hum. Genet.* **20**, 99–127 (2019) Wrote.  
655 2. JB Kinney, G Tkacik, CG Callan, Precise physical models of protein-DNA interaction from  
656 high-throughput data. *Proc. Natl. Acad. Sci.* **104**, 501–506 (2007) Wrote.  
657 3. M Weigt, RA White, H Szurmant, JA Hoch, T Hwa, Identification of direct residue contacts in  
658 protein-protein interaction by message passing. *Proc. Natl. Acad. Sci.* **106**, 67–72 (2009).  
659 4. DS Marks, et al., Protein 3D Structure Computed from Evolutionary Sequence Variation. *PLoS*  
660 *ONE* **6**, e28766 (2011).  
661 5. M Ekeberg, C Lovkvist, Y Lan, M Weigt, E Aurell, Improved contact prediction in proteins:  
662 Using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **87**, 012707 (2013).  
663 6. M Ekeberg, T Hartonen, E Aurell, Fast pseudolikelihood maximization for direct-coupling  
664 analysis of protein structure from many homologous amino-acid sequences. *J. Comput. Phys.*  
665 **276**, 341–356 (2014).  
666 7. RR Stein, DS Marks, C Sander, Inferring Pairwise Interactions from Biological Data Using  
667 Maximum-Entropy Probability Models. *PLoS Comput. Biol.* **11**, e1004182 (2015).  
668 8. JP Barton, ED Leonardi, A Coucke, S Cocco, ACE: adaptive cluster expansion for maximum  
669 entropy graphical model inference. *Bioinformatics* **32**, 3089–3097 (2016).  
670 9. A Haldane, WF Flynn, P He, RM Levy, Coevolutionary Landscape of Kinase Family Proteins:  
671 Sequence Probabilities and Functional Motifs. *Biophys. J.* **114**, 21–31 (2018).  
672 10. S Cocco, C Feinauer, M Figliuzzi, R Monasson, M Weigt, Inverse statistical physics of protein  
673 sequences: a key issues review. *Reports on Prog. Phys.* **81**, 032601 (2018).  
674 11. A Haldane, RM Levy, Influence of multiple-sequence-alignment depth on Potts statistical  
675 models of protein covariation. *Phys. Rev. E* **99**, 032405 (2019).  
676 12. HT Rube, et al., Probing molecular specificity with deep sequencing and biophysically inter-  
677 pretable machine learning. *bioRxiv* p. 2021.06.30.450414 (2021).  
678 13. S Zamuner, PDL Rios, Interpretable Neural Networks based classifiers for categorical inputs.  
679 *arXiv* (2021).  
680 14. C Feinauer, B Meynard-Piganeau, C Lucibello, Interpretable pairwise distillations for generative  
681 protein sequence models. *PLoS Comput. Biol.* **18**, e1010219 (2022).  
682 15. A Gerardos, N Dietler, AF Bitbol, Correlations from structure and phylogeny combine construc-  
683 tively in the inference of protein partners from sequences. *PLoS Comput. Biol.* **18**, e1010147  
684 (2022).  
685 16. C Hsu, H Nisonoff, C Fannjiang, J Listgarten, Learning protein fitness models from evolutionary  
686 and assay-labeled data. *Nat. Biotechnol.* **40**, 1114–1122 (2022).  
687 17. C Feinauer, E Borgonovo, Mean Dimension of Generative Models for Protein Sequences.  
688 *bioRxiv* p. 2022.12.12.520028 (2022).  
689 18. HT Rube, et al., Prediction of protein-ligand binding affinity from sequencing data with inter-  
690 pretable machine learning. *Nat. Biotechnol.* **40**, 1520–1527 (2022).  
691 19. ED Weinberger, Fourier and Taylor series on fitness landscapes. *Biol. Cybernetics* **65**, 321–330  
692 (1991).  
693 20. CT Zhang, R Zhang, Analysis of distribution of bases in the coding sequences by a diagram-  
694 matic technique. *Nucleic acids research* **19**, 6313–7 (1991).  
695 21. PF Stadler, Spectral landscape theory in *Evolutionary Dynamics: Exploring the Interplay of*  
696 *Selection, Accident, Neutrality and Function*, eds. J Crutchfield, P Schuster. (Oxford Univ.  
697 Press, Oxford), pp. 231 – 271 (2003).  
698 22. GD Stormo, Maximally efficient modeling of DNA sequence motifs at all levels of complexity.  
699 *Genetics* **187**, 1219 – 1224 (2011-04).  
700 23. FJ Poelwijk, V Krishna, R Ranganathan, The Context-Dependence of Mutations: A Linkage of  
701 Formalisms. *PLoS Comput. Biol.* **12**, e1004771 (2016).  
702 24. DH Brookes, A Aghazadeh, J Listgarten, On the sparsity of fitness functions and implications  
703 for learning. *Proc. Natl. Acad. Sci.* **119**, e2109649118 (2022).  
704 25. A Tareen, et al., MAVE-NN: learning genotype-phenotype maps from multiplex assays of  
705 variant effect. *Genome Biol.* **23**, 98 (2022).  
706 26. JD Jackson, LB Okun, Historical roots of gauge invariance. *Rev. Mod. Phys.* **73**, 663–680  
707 (2001).  
708 27. JD Jackson, *Classical electrodynamics*. (John Wiley & Sons), (1998).  
709 28. Y Aharonov, D Bohm, Significance of electromagnetic potentials in the quantum theory. *Phys.*  
710 *review* **115**, 485 (1959).

29. M Peshkin, A Tonomura, *The Aharonov-Bohm Effect*. (Springer Verlag), (2005). 711  
30. L Vaidman, Role of potentials in the Aharonov-Bohm effect. *Phys. Rev. A* **86**, 040101 (2012). 712  
31. A Posfai, J Zhou, DM McCandlish, JB Kinney, Gauge fixing for sequence-function relationships. 713  
*In prep.* (2024). 714  
32. BE Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric* 715  
*Functions*, Graduate Texts in Mathematics. (Springer), 2 edition, (2001) Read in early 2022. 716  
33. JB Kinney, GS Atwal, Parametric Inference in the Large Data Limit Using Maximally Informative 717  
Models. *Neural computation* **26**, 637–653 (2014-04) Wrote. 718  
34. GS Atwal, JB Kinney, Learning Quantitative Sequence-Function Relationships from Massively 719  
Parallel Experiments. *J. Stat. Phys.* **162**, 1203–1243 (2016) Wrote. 720  
35. BB Machta, R Chachra, MK Transtrum, JP Sethna, Parameter space compression underlies 721  
emergent theories and predictive models. *Science* **342**, 604 – 607 (2013). 722  
36. MK Transtrum, et al., Perspective: Sloppiness and emergent theories in physics, biology, and 723  
beyond. *The J. Chem. Phys.* **143**, 010901 – 14 (2015). 724  
37. E Seitz, DM McCandlish, JB Kinney, PK Koo, Interpreting cis-regulatory mechanisms from 725  
genomic deep neural networks using surrogate models. *bioRxiv* (2023). 726