

Optima and Simplicity in Nature

Kamaludin Dingle^{1,2,3,*}

¹*Department of Chemical Engineering and Biotechnology, Cambridge University, UK*

²*Department of Computing and Mathematical Sciences, California Institute of Technology, USA*

³*CAMB, Department of Mathematics and Natural Sciences,
Gulf University for Science and Technology, Kuwait*

Why are simple, regular, and symmetric shapes common in nature? Many natural shapes arise as solutions to energy minimisation or other optimisation problems, but is there a general relation between optima and simple, regular shapes and geometries? Here we argue from algorithmic information theory that for objective functions common in nature — based on physics and engineering laws — optimal geometries will be simple, regular, and symmetric. Further, we derive a null model prediction that if a given geometry is an optimal solution for one natural objective function, then it is *a priori* more likely to be optimal or close to optimal for another objective function.

Key words: Optimisation; simplicity; Kolmogorov complexity; physics

I. INTRODUCTION

Certain simple shapes, such as spheres, spirals, and fractal branching patterns, abound in nature. Why are they so common? One perspective on geometries in nature is that they arise via solutions to some form of optimisation problem, where the objective function stems from basic physics and engineering principles [1]. This suggests there might be some general relation between optima and simplicity [2]; but what precisely is the connection, if any? There are also several examples of simple, regular, or symmetric geometries that are simultaneously optimal, or near optimal, for different objective functions. Consider these motivating examples: (a) There exists a universally optimal arrangements of points on a sphere, which is optimal for a range of different energy functions [3]; (b) In studies of lattice proteins — a simplified biophysics model of protein folding — it has been observed that simple and regular lattice protein folds are the most designable (ie have the most sequences assigned) [4], are the fastest folders [5], the fastest unfolders [6], the most thermally stable [4], and are highly robust to genetic mutations [7] (see also [8]); (c) Scale free network architectures, with their regular patterns have been found to be highly robust [9], and also efficient [10] due to having small network diameters; (d) Spheres represent the optimal convex 3D shape for reduced thermal loss due to minimising surface area for a given volume, but at the same time this shape is believed to be optimal for low density packing of convex shapes [11].

Here we adopt an algorithmic information theory [12–14] approach, and argue that if the objective function to be optimised is ‘simple’ (in a technical sense, discusses below) and the set of geometries over which the the objective function is evaluated is also ‘simple’, then:

(1) *Optimal geometries will be simple, symmetric, or regular*
Further, perhaps more surprisingly we derive a null-model expectation that:

(2) *Given that a geometry is optimal for one function, then it is more likely to also be an optimal geometry for another function, compared to a null model based on random functions*

We also argue that these types of optimisation problem occur commonly in nature, hence explaining at least some of the ubiquitous appearances of simple natural shapes and forms. Before deriving these statements, in the next section we give some brief theoretical background on algorithmic information theory and formally describe the the optimisation problems that we will study.

II. BACKGROUND AND PROBLEM SET UP

Theoretical framework: Developed within theoretical computer science, *algorithmic information theory* [12–14] (AIT) sits at the intersection of computation, computability theory, and information theory. AIT is fundamentally based on estimating the information content, or complexity, of discrete objects or patterns, such as discrete sequences, or geometries. These information estimates can yield mathematical relations and bounds, which can be used for various scientific and mathematical predictions. In AIT, information content is quantified by *Kolmogorov complexity*, $K(x)$, which measures the complexity of an individual object x in terms of the amount of information required to describe or generate x . Another way to think about $K(x)$ is via compression, where $K(x)$ measures the size of the compressed version of x . If x is a simple string eg containing repeating patterns like $x = 1010101010101010$ then it is easy to compress, and hence $K(x)$ will be small. Contrast this with a randomly generated bit string of length n , which is highly unlikely to contain any patterns, and hence cannot be compressed. In general, a sequence of length n is called “complex” or “random” if $K(x) \approx n$ bits, and “simple” if $K(x) \ll n$.

More formally, the Kolmogorov complexity $K_U(x)$ of a string x with respect to a (prefix optimal) universal Turing machine [15] (U), is defined [12–14] as

$$K_U(x) = \min_p \{|p| : U(p) = x\} \quad (1)$$

where p is a binary program for U , and $|p|$ denotes the length of the (halting) program p in bits. Due to the invariance theorem [16] for any two optimal UTMs U and V , $|K_U(x) - K_V(x)| \leq c$ so that the complexity of x is independent of the choice of the machine, to within an additive constant c . Hence we conventionally drop the subscript U in $K_U(x)$, and speak of ‘the’ Kolmogorov

* Correspondence: dinglek@caltech.edu

complexity $K(x)$. $K(x)$ is formally uncomputable, meaning that there cannot exist a general algorithm that for any arbitrary string returns the value of $K(x)$. In practice, it is commonly approximated by standard data compression algorithms [16]. Further, a large number of studies have shown that AIT and Kolmogorov complexity can be successfully applied in physics, including thermodynamics [17–19], entropy estimation [20, 21], in addition to applications in engineering and statistics [16, 22–24]. More details and technicalities can be found in standard AIT references [16, 25–27].

Problem description: We consider a general optimization problem to take the form: find the optimal geometry x^* where

$$x^* = \arg \left(\max_x (f(x)) \right) \quad (2)$$

with $x \in \mathcal{X}$, and $f(x)$ is the objective function to maximise. The set \mathcal{X} is assumed to be some finite set of discrete geometries, configurations, or sequences, such as binary strings of length n bits, graphs on n nodes, the set of self avoiding walks of length n on a lattice, etc. The value of n is a ‘natural’ measure of the size of the problem. We assume that we can enumerate all possible elements within \mathcal{X} . Computational complexity is irrelevant for our purposes, so it does not matter if enumerating all geometries would require exponentially large amounts of time. We assume that $f(x)$ is some computable function that can be evaluated eg on a computer. Combinatorial optimization problems often come in this form [28].

III. RESULTS

Bounding the complexity of optima: We can describe any $x \in \mathcal{X}$, given the value of n , by first generating the set \mathcal{X} using n , then producing an ordering of the elements of \mathcal{X} using the function f (ie ordering elements from most optimal to least), and then describing x exactly with its rank r in the ordering. Thus we can bound the Kolmogorov complexity of optima by implementing the following algorithm:

- i) Enumerate in order all elements of \mathcal{X}
- ii) For each $x_i \in \mathcal{X}$, evaluate $f(x_i)$
- iii) List x_i in descending order of optimality: $x_1^*, x_2^*, \dots, x_{|\mathcal{X}|}^*$
- iv) Set $x^* = x_1^*$, print x^* and halt

Any element in the list $x_1^*, x_2^*, \dots, x_{|\mathcal{X}|}^*$ can be identified by its rank r in the list, where eg $r = 1$ corresponds the optimum x^* , $r = 2$ indicates the second highest value for f , and $r = |\mathcal{X}|$ is assigned to the geometry with the lowest value of the objective function. If evaluating the objective function for different elements leads to the same function value, then this could lead to multiple elements with the same rank r . To avoid this and keep all ranks unique, we will assume that elements with the same value are also ranked according to the order they appear in the enumeration. For example, if $f(x_{45}) = f(x_{52})$ and x_{45} has rank r , then x_{52} will have rank $r + 1$ because it appeared later in the original enumeration.

Putting all the preceding together we can bound the complexity of any geometry x_r^* by

$$K(x_r^*|n) \leq K(\mathcal{X}|n) + K(f) + K(r|n) + O(1) \quad (3)$$

From this bound we see that the complexity $K(\mathcal{X}|n)$ of the set \mathcal{X} as well as the complexity $K(f)$ of the function f are important quantities. We will now examine these quantities in more detail.

The set \mathcal{X} may in general be simple or complex. Many optimization problems in physics take the form looking for an optimal x over *all possible* x within some set, such as all possible graphs up to some size, all possible protein configurations, all possible branching patterns, all possible topologies with some condition. Perhaps surprisingly, the ‘set of all’ objects of some kind typically has very low Kolmogorov complexity. For example, the set of all binary strings of length n has complexity only $K(n) \lesssim \log_2(n)$, because it can be described by the program ‘Enumerate and print all binary strings of length n ’. (Contrast this with the complexity of a typical element of the set of all binary strings of length n , which is likely to have complexity $\approx n$ bits, far higher than $\log_2(n)$ bits.) Indeed, the countably infinite set of *all* binary strings $\{0,1,00,01,\dots\}$ has very low information content, because it can be generated by a fixed length program, and hence has complexity $O(1)$ bits. Similarly, the set of all graphs on n nodes has very low complexity, at most $\approx \log_2(n)$ bits. A set can be complex also, if many bits are required to specify it precisely. For example if \mathcal{X} is made from randomly selecting $M \ll 2^n$ strings out of all binary strings of length n , then in this case $K(\mathcal{X}) \approx Mn$ bits because the only way to precisely generate this set is by individually specifying each element, each of which has complexity $\approx n$ bits (with high probability).

An objective function $f(x)$ may in general be simple or complex. Examples of simple functions are (low order) polynomials, trigonometric functions, and many of the equations describing the laws of physics (eg $F = ma$). These all are classed as simple $O(1)$ complexity functions, because they can be computed via fixed sized algorithms. An example of a complex function is $f(x) = x + 0110\dots11$ where the latter binary string is a randomly chosen string of length n . In this case, $K(f) \approx n$ bits, which may be arbitrarily large. A noteworthy and less trivial example of a complex function arises in the well-know *assignment problem* from combinatorial optimisation [28]: If there are n people and n jobs, and a matrix C_{ij} which gives the affinity of person i for job j , the problem is to find the assignment of the jobs to the people which maximises the total affinity. The space of configurations/sequences is the space of the $n!$ permutations of the n indices. Here, the objective function $f(x)$ is in general not simple, because the Kolmogorov complexity of the matrix C_{ij} will typically be $O(n^2)$.

Simple optimal geometries: We have seen that for many optimisation problems relevant to physics, simple sets with simple objective functions can be expected. In these cases, both $K(\mathcal{X}|n)$ and $K(f)$ will be small. Further, highly optimal structures x will have rank values $r \approx 1$ (and $r = 1$ for the optimum) and hence $K(r|n) \approx 0$. Therefore the right

hand side of Eq. (3) will be very small, and

$$K(x_r^*|n) \leq K(\mathcal{X}|n) + K(f) + K(r|n) \approx 0 \Rightarrow K(x_r^*|n) \approx 0 \quad (4)$$

so that any x_r^* with $r \approx 1$ must be simple, and x^* must have very low, $O(1)$ complexity.

We have proven that highly optimal solutions must have low Kolmogorov complexity, but this is not necessarily the same as simple or symmetric in the usual senses of the words. It is possible, in principle, that these solutions appear complex or pseudo random. However, we can expect in many natural systems that near optimal solutions will in fact be simple, or regular, or symmetric, because common physics-based objective functions are unlikely to contain UTMs which can compute arbitrary algorithms and generate pseudo random patterns. Notable exceptions include chaotic systems, which are common in nature, and hence may possible lead to pseudo random optima. See Vitanyi [22, 29] for a similar argument for why common patterns in natural science are unlikely to be pseudo random.

Note that $K(r|n) \approx 0$ also holds for $r \approx |\mathcal{X}|$. For example if there are 2^n possible geometries, then the last rank will be $r = 2^n$, and hence $K(r|n) = K(2^n|n) = O(1)$. This also implies the *least* optimal geometry must also have low Kolmogorov complexity and hence be simple (assuming a simple set and function). In general, any geometries with extreme values with respect to the objective function will be simple.

Example of a simple optimum: Define \mathcal{X} to be all binary strings of length n , and let the objective function be $f(x) = \sum^n x_i$, with x_i the i th bit of the string x (i.e. $f(x)$ counts the number of 1's in the string). The set \mathcal{X} has complexity $K(\mathcal{X}|n) = O(1)$, and $K(f) = O(1)$. It follows then from Eq. (4) that whatever is the optimal string x^* with $r = 1$, it must have $K(x^*|n) = O(1)$ and hence be very simple. In this somewhat trivial example, we can confirm this prediction by noting that the optimal solution is $x^* = 111\dots 11$ with n 1's, and we know that $K(111\dots 11|n) = O(1)$.

Null expectation for simultaneously optimising multiple objective functions: In the Introduction we saw several examples of optimal geometries that simultaneously have high or optimal values for different objective functions. But are these rare exceptional cases, or instances of a more common trend? We now derive a surprising null expectation connecting optima for different objective functions.

Consider a finite and simple set \mathcal{X} as above, and two different objective functions f_1 and f_2 , with respective optima $x_{f_1}^*, x_{f_2}^* \in \mathcal{X}$. If these two functions are random functions, such that their values are random variables, then the null model prediction for the probability that the two optimal geometries coincide is

$$P(x_{f_1}^* = x_{f_2}^*) = \frac{1}{|\mathcal{X}|} \quad (5)$$

assuming that all function values are distinct. This probability is typically exponentially small, because $|\mathcal{X}|$ is typically exponential in n , eg $|\mathcal{X}| = 2^n$ if \mathcal{X} consists of all binary strings. Hence, under the random function null model it is

highly unlikely that the optimum geometry for one function coincides with that of another function.

The situation is very different for optima from simple functions: if f_1 and f_2 are simple having $O(1)$ complexity, then both $x_{f_1}^*$ and $x_{f_2}^*$ must be simple by Eq. (4). Because there are exponentially few simple geometries, $x_{f_1}^*$ and $x_{f_2}^*$ must both come from this exponentially small set. Hence given that we already know $x_{f_1}^*$ is simple and therefore in this small set, it is *a priori* far more likely that $x_{f_1}^*$ and $x_{f_2}^*$ coincide. More formally, let $\mathcal{X}_{O(1)} \subset \mathcal{X}$ be such that all elements are simple with $O(1)$ complexity. In this scenario, the null probability expectation that the two optima coincide is

$$P(x_{f_1}^* = x_{f_2}^*) = \frac{1}{|\mathcal{X}_{O(1)}|} \gg \frac{1}{|\mathcal{X}|} \quad (6)$$

because $|\mathcal{X}_{O(1)}| = O(1) \ll |\mathcal{X}|$ due to the fact that there are only $O(1)$ simple geometries. While $O(1)$ is not a precise quantitative value, nonetheless we understand that this probability is much higher than the exponentially small value of $1/|\mathcal{X}|$. $1/|\mathcal{X}_{O(1)}|$ may not be a high value, but it is much higher than expected by random functions.

The preceding is an *a priori* argument, or expected null model for simple optima. It does not mean that all simple optima will be optima for all simple functions. Indeed, if we chose $f_2(x) = -f_1(x)$ then the optimum for f_1 will be the *worst* performing geometry for f_2 . This illustrates that extrema must be simple, but not necessarily optimal.

Another reason why optima may not coincide is due the fact that while optima (under the described conditions) must be simple, it does not mean that all simple geometries must be optima. Some simple geometries may have low or medium values with respect to the objective function. This is related to the occurrence of low complexity, low probability patterns in studies of simplicity bias [24, 30].

The above argument can be extended beyond exactly coinciding optima, to geometries which are close to optima ie with rank $r \approx 1$. We can ask, given we have an optimal solution $x_{f_1}^*$ for f_1 , and we have a set $\mathcal{X}_{f_2}^r \subset \mathcal{X}$ consisting of the r most optimal geometries in \mathcal{X} ranked according to f_2 , what is the probability that $x_{f_1}^* \in \mathcal{X}_{f_2}^r$? This is an interesting question, because if $x_{f_1}^* \in \mathcal{X}_{f_2}^r$ then it means that the optimal geometry for one objective function f_1 is also close to optimal for another apparently unrelated objective function f_2 . To proceed, if f_1 and f_2 are random functions, then we expect

$$P(x_{f_1}^* \in \mathcal{X}_{f_2}^r) = \frac{r}{|\mathcal{X}|} \quad (7)$$

which is a very small number for highly optimal geometries with $r \approx 1$ because $r \ll |\mathcal{X}|$. Contrast this with the case where functions are simple, then the null model prediction is

$$P(x_{f_1}^* \in \mathcal{X}_{f_2}^r) = \frac{r}{|\mathcal{X}_{O(1)}|} \gg \frac{r}{|\mathcal{X}|} \quad (8)$$

because both $x_{f_1}^* \in \mathcal{X}_{O(1)}$ and $\mathcal{X}_{f_2}^r \subset \mathcal{X}_{O(1)}$ due to being simple.

IV. DISCUSSION

We have shown that optimal discrete geometries (or configurations) selected by low Kolmogorov complexity (simple

objective functions from low Kolmogorov complexity sets, must themselves have low Kolmogorov complexity: optima for simple objective functions must be simple. However, geometries with very *low* scores with respect to some objective function are also expected to be simple, hence not always optima. We argued that these types of optimisations are common in nature, due to the fact that many objective functions derive from generic physics and engineering-based constraints, and these constraints can be described by simple $O(1)$ equations and laws. Our result is perhaps a generalisation of the finding of Cohn and Kumar [31] that symmetrical configurations can often be built using very simple potential functions, and perhaps also the statement of Bormashenko [32] that the symmetry of biological structures follow the symmetry of media in which the structure is functioning. In ref. [33] we argued that much of the symmetry in biology can be explained by the fact that simpler shapes are easier to generate via efficient genomic programs in DNA, which is a kind of ‘internal’ argument for simplicity. Here we have shown that symmetry and simplicity can arise from a kind of ‘external’ argument, that of the simplicity of the external constraints.

Here we also argued that given a geometry is optimal for one simple objective function, it is therefore more likely to be an optimal geometry for another simple objective function, as compared to a null expectation based on random functions. This is a rather surprising result, given that we typically conceive of exponentially large search spaces. Our argument is somewhat similar to that of Valle-Perez et al [34] who proposed that a reason for the success of deep neural

networks in machine learning is that they are biased towards simpler functions, and natural functions are also tend to be simple, such that a coincidence of functions is more likely than expected by a random function null model.

The work here has focussed on discrete geometries, configurations, or sequences, such as binary strings, discrete graphs, branching patterns, and lattice shapes. However, continuous curves, functions and shapes can be discretised often quite easily [23, 35], and because of this the arguments presented here apply to a wide array of shapes and geometries.

Looking at individual systems, we may be able to use system specific theory to directly infer that some optimal shape for one objective function is linked to optimal shapes for some other objective function, eg employing thermodynamics to understand why high mutational robustness and thermal stability are linked in natural proteins [36]. Importantly, such examples do not undermine our work here, in which we have attempted to show *general theory* linking optima, simplicity, and coinciding optima. Whether or not particular cases can or cannot be understood with existing system specific theory, does not affect our arguments. Here we have taken an information content approach from algorithmic information theory as a theoretical framework for investigating these questions in general.

Acknowledgements: We acknowledge financial support from the Gulf University for Science and Technology for a Seed Grant (grant number 263301) and a Summer Faculty Fellowship. We thank Paris Flood and Nora Martin for discussions related to this work.

-
- [1] S. Hilderbrandt and A. Tromba. *The parsimonious universe: shape and form in the natural world*. Springer-Verlag, New York, 1996.
- [2] Henry Cohn. Order and disorder in energy minimization. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. 1: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 2416–2443. World Scientific, 2010.
- [3] Henry Cohn and Abhinav Kumar. Universally optimal distribution of points on spheres. *Journal of the American Mathematical Society*, 20(1):99–148, 2007.
- [4] H. Li, R. Helling, C. Tang, and N. Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273(5275):666, 1996.
- [5] R. Mélin, H. Li, N.S. Wingreen, and C. Tang. Designability, thermodynamic stability, and dynamics in protein folding: a lattice model study. *The Journal of chemical physics*, 110:1252, 1999.
- [6] C.L. Dias and M. Grant. Designable structures are easy to unfold. *Physical Review E*, 74(4):042902, 2006.
- [7] Sam F Greenbury, Steffen Schaper, Sebastian E Ahnert, and Ard A Louis. Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. *PLoS computational biology*, 12(3):e1004773, 2016.
- [8] E.D. Nelson, L.F. Teneyck, and J.N. Onuchic. Symmetry and kinetic optimization of proteinlike heteropolymers. *Physical review letters*, 79(18):3534–3537, 1997.
- [9] S. Valverde, R.F. Cancho, and R.V. Sole. Scale-free networks from optimal design. *EPL (Europhysics Letters)*, 60:512, 2002.
- [10] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90(5):58701, 2003.
- [11] A. Haji-Akbari, M. Engel, A.S. Keys, X. Zheng, R.G. Petschek, P. Palfy-Muhoray, and S.C. Glotzer. Disordered, quasicrystalline and crystalline phases of densely packed tetrahedra. *Nature*, 462(7274):773–777, 2009.
- [12] R.J. Solomonoff. A preliminary report on a general theory of inductive inference. 1960.
- [13] A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problems of information transmission*, 1(1):1–7, 1965.
- [14] Gregory J Chaitin. A theory of program size formally identical to information theory. *Journal of the ACM (JACM)*, 22(3):329–340, 1975.
- [15] Alan Mathison Turing. On computable numbers, with an application to the entscheidungsproblem. *J. of Math*, 58(345–363):5, 1936.
- [16] M. Li and P.M.B. Vitanyi. *An introduction to Kolmogorov complexity and its applications*. Springer-Verlag New York Inc, 2008.
- [17] C.H. Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.
- [18] Artemy Kolchinsky and David H Wolpert. Thermodynamic costs of turing machines. *Physical Review Research*, 2(3):033312, 2020.
- [19] W.H. Zurek. Algorithmic randomness and physical entropy. *Physical Review A*, 40(8):4731, 1989.

- [20] Ram Avinery, Micha Kornreich, and Roy Beck. Universal and accessible entropy estimation using a compression algorithm. *Physical review letters*, 123(17):178102, 2019.
- [21] Stefano Martiniani, Paul M Chaikin, and Dov Levine. Quantifying hidden order out of equilibrium. *Physical Review X*, 9(1):011031, 2019.
- [22] Paul MB Vitányi. Similarity and denoising. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984), 2013.
- [23] Kamaludin Dingle, Chico Q Camargo, and Ard A Louis. Input–output maps are strongly biased towards simple outputs. *Nature communications*, 9(1):761, 2018.
- [24] Kamaludin Dingle, Guillermo Valle Pérez, and Ard A Louis. Generic predictions of output probability based on complexities of inputs and outputs. *Scientific reports*, 10(1):1–9, 2020.
- [25] C.S. Calude. *Information and randomness: An algorithmic perspective*. Springer, 2002.
- [26] P. Gács. *Lecture notes on descriptive complexity and randomness*. Boston University, Graduate School of Arts and Sciences, Computer Science Department, 1988.
- [27] Alexander Shen, Vladimir A Uspensky, and Nikolay Vereshchagin. *Kolmogorov complexity and algorithmic randomness*, volume 220. American Mathematical Society, 2022.
- [28] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, USA, 2009.
- [29] Paul Vitányi. How incomputable is kolmogorov complexity? *Entropy*, 22(4):408, 2020.
- [30] Mohamed Alaskandarani and Kamaludin Dingle. Low complexity, low probability patterns and consequences for algorithmic probability applications. *Preprints*, 2022.
- [31] Henry Cohn and Abhinav Kumar. Algorithmic design of self-assembling structures. *Proceedings of the National Academy of Sciences*, 106(24):9570–9575, 2009.
- [32] Edward Bormashenko. Fibonacci sequences, symmetry and ordering in biological patterns, their sources, information origin and the landauer principle. *Preprints*, 2022080107, 2022.
- [33] Iain G Johnston, Kamaludin Dingle, Sam F Greenbury, Chico Q Camargo, Jonathan PK Doye, Sebastian E Ahnert, and Ard A Louis. Symmetry and simplicity spontaneously emerge from the algorithmic nature of evolution. *Proceedings of the National Academy of Sciences*, 119(11):e2113883119, 2022.
- [34] Guillermo Valle-Perez, Chico Q Camargo, and Ard A Louis. Deep learning generalizes because the parameter-function map is biased towards simple functions. *arXiv preprint arXiv:1805.08522*, 2018.
- [35] Kamaludin Dingle, Rafiq Kamal, and Boumediene Hamzi. A note on a priori forecasting and simplicity bias in time series. *arXiv preprint arXiv:2203.05391*, 2022.
- [36] Jesse D Bloom, Jonathan J Silberg, Claus O Wilke, D Allan Drummond, Christoph Adami, and Frances H Arnold. Thermodynamic prediction of protein neutrality. *Proceedings of the National Academy of Sciences*, 102(3):606–611, 2005.