The Physical Logic of Protein Machines

John M. McBride[1,*] and Tsvi Tlusty[1,2,†]

[1]*Center for Soft and Living Matter, Institute for Basic Science, Ulsan 44919, South Korea*
[2]*Departments of Physics and Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea*

Proteins are intricate molecular machines whose complexity arises from the heterogeneity of the amino acid building blocks and their dynamic network of many-body interactions. These nanomachines gain function when put in the context of a whole organism through interaction with other inhabitants of the biological realm. And this functionality shapes their evolutionary histories through intertwined paths of selection and adaptation. Recent advances in machine learning have solved the decades-old problem of how protein sequence determines their structure. However, the ultimate question regarding the basic logic of protein machines remains open: How does the collective physics of proteins lead to their functionality? and how does a sequence encode the full range of dynamics and chemical interactions that facilitate function? Here, we explore these questions within a physical approach that treats proteins as mechano-chemical machines, which are adapted to function via concerted evolution of structure, motion, and chemical interactions.

Contents

1. Introduction: protein machines 1
2. Protein biophysics and evolution 1
3. Why do protein machines move? 2
4. Genetic-mechano-chemical model of protein machines 3
5. Quantifying internal motion 6
6. Physics and evolution of protein motions 7
7. Protein evolution as effective internal motion 9
8. Discussion 10
References 11

1. Introduction: protein machines

Proteins are often described as molecular machines,[1–5] but what does this mean? While a rigorous consensus definition is lacking, broadly speaking, machines are devices or objects that transform a physical system to perform a function. The operation of machines naturally involves some dynamics. Many machines have moving parts and exert time-varying forces. Other machines change their internal state as they perform computation, and often computation and motion are coupled, as in mechanical computers. Proteins appear to fit into all of these criteria: they are complex objects which often have independently moving parts; motor proteins facilitate directional motion; binding to some molecules and not others is a form of computation; and they all evolve according to the task they perform – their *function*.

In this article, we will treat proteins as deformable mechano-chemical machines, whose physical behavior and response to forces are encoded in their DNA sequences, and whose evolution is directed towards improved function. We will address the question of why binding – which underlies most protein functions – requires *motion*. We discuss how to quantify internal motion involved in function, and in particular, examine quantities derived from finite strain theory. We then turn to the mechanistic question of how specific functional motion is encoded in protein sequences. Finally, we show how amino-acid substitutions lead to variations in protein structure that can be described as a kind of evolutionary motion captured by the same deformation metrics that describe the physical motion. Overall, the present study takes one step further toward understanding the physical logic of protein machines by proposing a general model of the genotype-to-phenotype map that takes into account the evolution of functional motion.

2. Protein biophysics and evolution

Understanding the logic of protein machines is an open challenge. A longstanding paradigm in biology has been that protein function is determined by structure, which is in turn determined by sequence. This has been a useful first approximation of the logic of proteins. But by now, we have a much more nuanced understanding, briefly summarized below, of the ongoing cycle of how genes encode proteins and their function, which in turn biases the evolution of genes (Fig. 1).

A protein, of course, is not simply a static object captured by a single structure, but rather a dynamic one, whereby an ensemble of structures, their dynamics, and surface chemistry intertwine to define a space of possible functions. A useful characterization of proteins must explain how these ensembles are shaped by the physico-chemical environment,[6–11] and how they respond to physical forces when interacting with other molecules.[12–16] Such responses can be characterized by a shift in the conformational landscape of a protein:[17–20] For example, conformations can change at binding sites due to induced fit,[12] or long-ranged, allosteric changes can occur throughout a protein;[14] in extreme cases, binding can cause disorder-to-order transitions, whereby a disordered region can form a stable secondary structure upon binding.[21]

The notion of function also needs to be re-examined, as some proteins can have many, varied functions (moonlighting),[22,23] and proteins can gain or lose function depending on the environment (gain of function, or pseudogenes).[24–26]

* jmmcbride@protonmail.com
† tsvitlusty@gmail.com

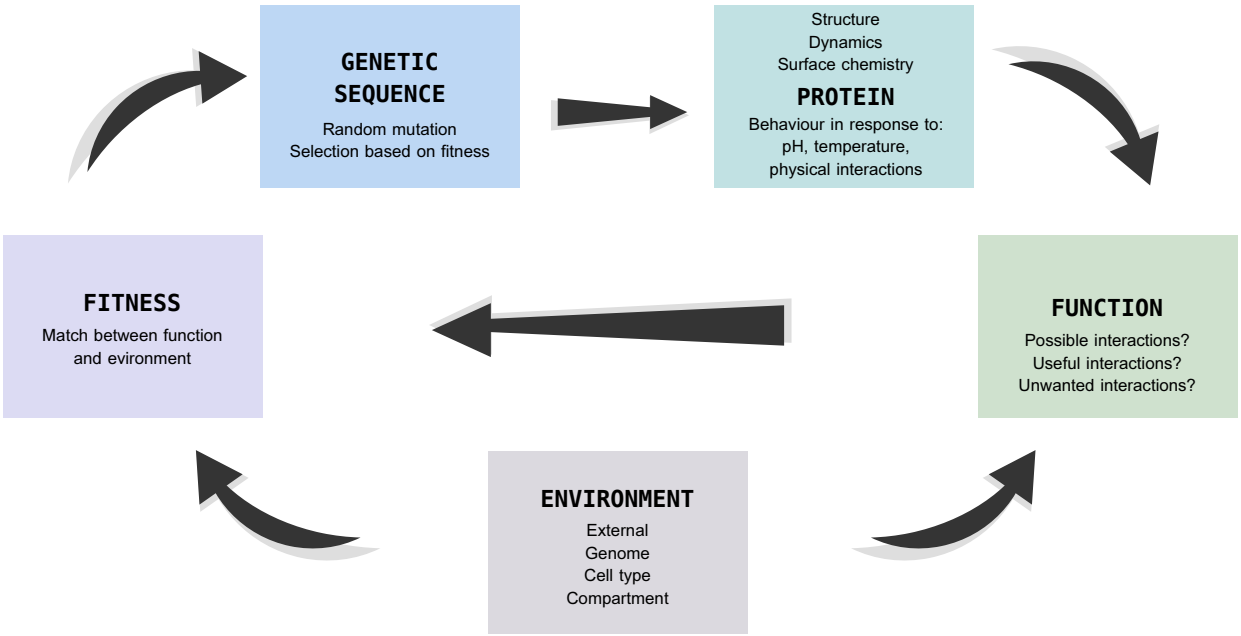

FIG. 1. **The cycle of evolution of protein machines**. *Genetic sequences* encode the protein's ensemble of molecular structures and its physical behavior. *Protein behavior* is defined by structure, dynamics, and surface chemistry. These aspects of protein behavior vary depending on the physicochemical environment and in response to physical forces encountered by the proteins when interacting with other molecules. *Functions* are constrained by the possible interactions a protein can have, but they are typically defined by the interactions that are *useful* for the host organism. Proteins exist and function within a *hierarchy of biological networks and environments*: proteins are often localized in specific cellular compartments and organelles, or only expressed in certain cell types; proteins often evolve to specifically interact with other molecules that are encoded by the host organism's genome, or encountered inside or outside the cell. The *fitness* contribution of a gene depends on how well the function matches the environment and the needs of the host organism, and this, in turn, acts to bias the selection of random mutations in the genetic sequence.

If proteins are expressed in new cell types, or located in new sub-cellular locations,[27] they gain access to new sets of molecules to interact with.[28,29] Varying external environments also provide a fluctuating landscape of interaction partners, which can switch proteins between functional and non-functional states (*e.g.*, antibodies, anti-toxins).[30,31] Function is also defined not only by interactions with target molecules but, not less importantly, by interactions or lack thereof with off-target molecules. Proteins need to interact *specifically* with certain molecules, while avoiding aggregation,[32] unwanted cross-talk in signalling,[33] and enzymatic side-reactions.[34] Thus, to wholly understand protein function, one needs to understand the range of behaviors that a protein is capable of, know what molecules they will encounter, and how it all links with the fitness of the host organism. For this, we need a physical model of the full genotype-to-phenotype map, which takes into account the place of the protein in the biological circuitry and all relevant counterparts.

3. Why do protein machines move?

All proteins move by their physical nature. The first theory of protein binding was the *lock-and-key* model, which hypothesized rigid shape matching between enzymes and substrates.[35] This model was proposed in 1894 by Emil Fischer, long before much was known about the size and complexity of proteins.[36] While geometry is indeed an important component of binding, by now, we are aware that proteins are large, flexible molecules that undergo internal motion (to which we refer hereafter simply as "motion" as we do not discuss whole-body translations and rotations).[37] Proteins are bound to move simply because their structures are held together by relatively weak bonds, which can easily break and reform. Moreover, the linear-chain topology makes it impossible to optimize all of the bonds between amino acids.[38,39] Thus the first answer to the question, "Why do proteins move?", is that they cannot help it. However, beyond the physical inevitability, there is much evidence that motion in proteins is also beneficial and that the extent and direction of movement are selected through evolution and can be harnessed to perform functions.

Mechanisms of binding involve movement. The lock-and-key model was eventually updated to include movement, first through the induced fit mechanism, and later through conformational selection:[40] In the induced fit model, proteins undergo conformational change only upon binding to substrate;[12]. In the conformational selection model, the unbound protein inherently samples many possible conformations, and binds when it achieves the optimally binding conformation.[41–43]

These two models can be described as extreme cases on a 2-dimensional continuum (Fig. 2) where the two coordinates are the degree of conformational flexibility in the absence of substrate (x axis) and the degree of conformational change upon binding (y axis). In reality, we expect that most binding events are described in the intermediate space in between the regions of induced fit, and conformational selection – sometimes called the population shift model.[44,45] Altogether these theories provide a comprehensive description of the mechanisms

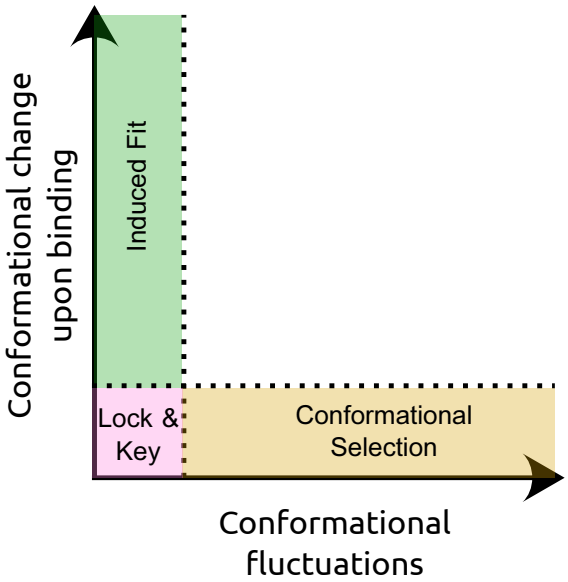

FIG. 2. Mechanisms of binding described in a 2D diagram whose coordinates are the degree of conformational fluctuations in the absence of the substrate, and the magnitude of conformational change upon binding to the substrate. The classical models – lock-and-key, induced fit, and conformational selection – are shown as limiting cases. In reality, absolute rigidity and deformation-free binding are impossible – proteins are inherently flexible and undergo thermal motion, while binding involves forces that will always induce some degree of conformational change.

of binding in terms of protein motion, but the mechanisms have little to say about *why* movement may be beneficial. In fact, if we purely considering molecular recognition, flexibility ought to be detrimental to binding, as more flexible proteins have greater reserves of conformational entropy that may be lost upon binding. To understand why proteins benefit from flexibility, we consider the fact that proteins exist in a milieu of many molecules, and they need to discriminate between target and off-target molecules.

Binding specificity necessitates internal motion. The induced fit theory was motivated by observations on amino acid binding by tRNA synthetases which were incompatible with the lock-and-key theory. Koshland[12] noted that differences van der Waals interaction energy are insufficient to explain how isoleucyl-tRNA synthetase selects its target, the amino acid isoleucine, over valine which differs only by a single methyl group (-CH_3). The lock-and-key model can explain the selection of valine over isoleucine (by the corresponding valyl-tRNA synthetase) – since isoleucine is larger, it can be sterically excluded. However, lock-and-key cannot explain the inverse selection since valine should fit into the binding pocket that also fits the larger isoleucine. The proposed solution was the induced fit model, whereby the binding of isoleucine must lead to a conformational change that unlocks the binding pose, while valine may not induce such a change. This theory, thus, offers one good reason why proteins might have evolved to move.

The idea that movement is needed for binding specificity was revisited in the conformational proofreading theory.[15,46,47] This model gives a quantitative explanation of induced fit and conformational selection, which takes into account the inherent flexibility of proteins and the energetics of recognition. According to this theory, proteins may interact with off-target as well as target molecules due to conformational flexibility, if the two molecules are similar (by either induced fit or conformational selection). This leads to the prediction that, for molecular discrimination, the optimal protein may have some mismatch with the target: often, a small amount of mismatch between protein and target will slightly decrease the affinity to target, while disproportionally affecting binding to off-target ligands.

4. Genetic-mechano-chemical model of protein machines

In our recent work, we produced a detailed quantitative model of protein binding that builds on these previous models, with the aim of elucidating the mechanistic determinants of binding specificity and how they are encoded in the evolution of the gene.[16] The model is presented in two variants – a basic mechano-chemical machine (denoted **MeCh**), and a genetic-mechano-chemical machine (**G-MeCh**), which adds the genetic component. Here, we briefly describe the main concepts, while a full discussion of the model can be found in McBride *et al.*[16].

We start with the simple variant of the model, **MeCh**, which takes into account flexibility, protein-ligand mismatch, and chemical binding energy (Fig. 3A). For simplicity, we keep the initial structure constant (although see McBride *et al.*[16] for a relaxation of this constraint), and model the protein as a 2-dimensional spring network of "amino acids", a, (each representing a tightly-connected bunch of amino acids) on a hexagonal lattice.

Protein flexibility is described by connecting the nearest neighbors with harmonic springs, such that deformation incurs an energy penalty

$$\mathcal{E} = \frac{1}{2} \sum_{\langle i,j \rangle} K_{ij} \left(r_{ij} - \ell_{ij} \right)^2 , \tag{1}$$

where the summation $\langle i, j \rangle$ is over all pairs of amino acids a_i and a_j connected by a bond, K_{ij} is the spring constant of the bond, r_{ij} is its length, and ℓ_{ij} its equilibrium length. The initial configuration is not deformed, $\mathcal{E} = 0$, since all distances are set equal to the equilibrium bond lengths, $r_{ij} = \ell_{ij}$. In the **MeCh** model, all spring constants are equal, $K_{ij} = K$. For simplicity, we treat ligands as rigid molecules. In reality, however, ligands can also be flexible, especially if the ligand is a protein or peptide, and the model can be extended to account for such effects. Flexibility also determines the conformational entropy of the protein, and conformational entropy loss upon binding is calculated under the assumption that binding results in a more rigid binding site; note that we also include a separate parameter to account for other sources of binding entropy change.

Next, we consider protein-ligand mismatch and binding energy. In reality, mismatch and binding energy are inseparable, since amino acid substitutions at the binding site lead to concurrent changes in shape and chemistry. Nevertheless, we treat mismatch and binding energy as separable components, both for simplicity, and so that we may examine their separate contributions to binding specificity. Mismatch is entirely encoded in the shape – the protein binding pocket has an angle of 60°, and ligands are wedge-shaped

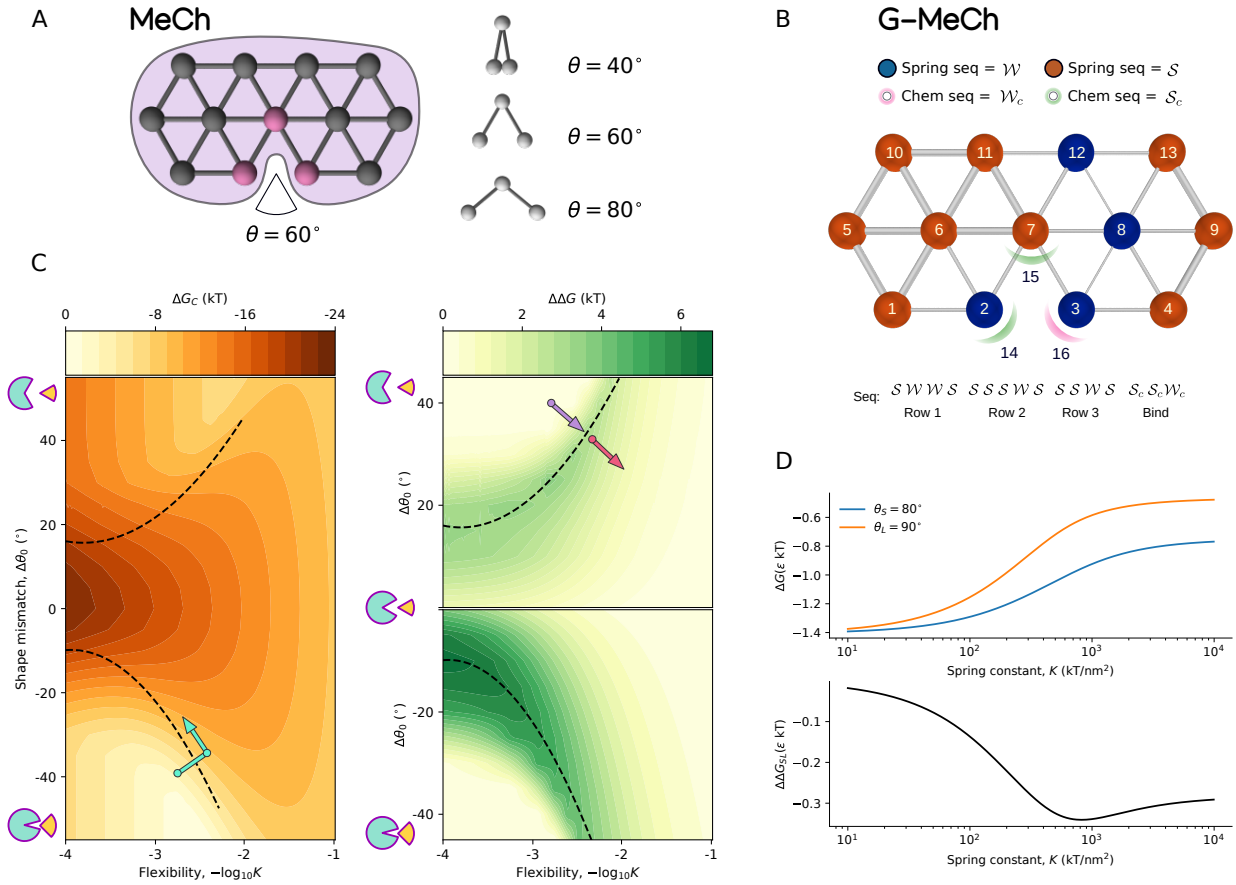

FIG. 3. A: Mechano-chemical discriminator model (**MeCh**): Coarse-grained model of a protein as spring network with a chemical binding site. In this example, the model includes $N = 13$ nodes referred to as 'amino acids' represent small sets of tightly bound coordinated amino acids. These amino acids interact with their neighbors via harmonic potentials. The binding site consists of three loci which bind to corresponding loci on wedge-shaped ligands. Spring constants are all equal to $K_{ij} = K$, and each binding loci can contribute a maximum of ε kT to the binding energy.
B: Genetic, mechano-chemical **G-MeCh** discriminator model: The model is the same as **MeCh**, but now the energy and spring constants are a function of a genetic sequence $\mathbf{g} = \{a_i\}$ with two amino acids species $\mathcal{S}$ and $\mathcal{W}$; springs can be either weak, intermediate, or strong depending on the amino acid identity of the neighbors, $K_{ij}(a_i, a_j)$; chemical binding energy can be either weak or strong depending on the sequence.
C: Affinity and specificity phase diagrams in the **MeCh** model, as a function of flexibility and mismatch between the protein pocket ($\theta_0 = 60°$) and the ligand shape, $\Delta\theta_0$. Positive shape mismatch indicates that the binding pocket is larger than the ligand, and vice versa.
D: Free energy of binding, ΔG, for large ($\theta_L = 90°$) and small ($\theta_S = 80°$) ligands, and the difference in free energy $\Delta\Delta G_{SL} = \Delta G_S - \Delta G_L$, as a function of spring constant K for the **MeCh** model.

with angles ranging from $5 - 175°$. Chemical binding energy is modeled as a short-ranged interaction between three amino acids at the binding pocket with three corresponding loci on the ligand,

$$\mathcal{C} = \sum_i \varepsilon_i e^{-|\mathbf{r}_i^b - \mathbf{r}_i^p|^2/\sigma^2} . \tag{2}$$

Here, ε_i is the energy scale of binding locus i of the ligand, $\mathbf{r}_i^p$ and $\mathbf{r}_i^b$ are the positions, respectively, of the amino acids a_i at the binding pocket and the ligand binding locus i, and we set the length scale of the interaction at $\sigma = 0.3\,\text{nm}$. In the **MeCh** model, the energy scale is the same for all binding sites, $\varepsilon_i = \varepsilon$. The energy scale, ε, is equivalent to the maximum chemical binding energy afforded by a ligand's accessible chemical groups.

While **MeCh** models a general, mechano-chemical discrimination machine, the **G-MeCh** model aims to describe more concretely proteins (Fig. 3B). Therefore, instead of allowing the spring constants and energy scale to vary continuously, the mechano-chemical properties of the protein are defined by genetic sequence $\mathbf{g} = \{a_i\}$ that determines the spring constants $K_{ij}(a_i, a_j)$ in Eq.(1) and the binding interactions $\varepsilon_i(a_i)$ in Eq.(2). This results in a protein that undergoes discrete changes via mutations to its sequence, and is a heterogeneous rather than homogeneous spring network.

The simplicity of these models allows us to examine a large space of possible proteins, where we can calculate binding energy for a large range of ligands, and all possible genotypes. The complete survey of the genotype-phenotype map sheds light on the evolution of specific binding in proteins, and how this depends on the interplay between internal protein motion, shape mismatch, and binding energy.

Predictions from the genetic-mechano-chemical model:

(A) *Suboptimality of Lock-and-key*: Lock-and-key binding is not the optimal strategy for evolving binding specificity; it only works when off-target molecules are larger (leading

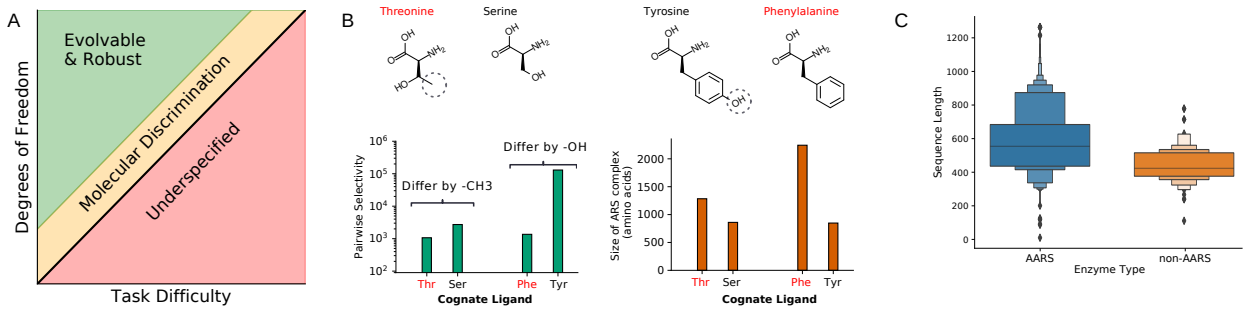

FIG. 4. **Size-Specificity-Scaling Hypothesis**. A: The difficulty of a discrimination task imposes a constraint on the minimum sequence length that can lead to successful completion of the task. Sequences longer than the minimum length are more evolvable and robust. B: Amino-acyl tRNA-synthetase (AARS) pairwise selectivity (ratio of binding constant of cognate ligand to a similar non-cognate ligand) and number of residues in the AARS complex, for threonine, serine, phenylaline and tyrosine. Ligands that are more difficult to selectively bind to are highlighted in red. C: Sequence length distribution of AARS vs non-AARS proteins that also bind to amino acids.

to steric exclusion), in the unrealistic case of extreme rigidity and perfect match between protein and ligand. Rigid proteins do not lose much conformational entropy upon binding, which can lead to very high binding affinity. Although affinity is higher for the matching ligand, similar off-target ligands will still bind even if the binding pose is not perfect (Fig. 3C). This is the aforementioned insight that led to the development of induced-fit model.[12]

(B) *Mismatch is essential*: We show, in agreement with the *conformational proofreading* model,[15] that some mismatch between protein and target is required for discrimination, in order to increase the energy penalty for binding to off-target ligands (Fig. 3C). Through the **MeCh** model, we explain mechanistically that this is due to the non-linear nature of binding as a function of protein motion. Deformation energy (stretching of intramolecular bonds) increases at a greater rate for the off-target molecules than the target molecules. As a result, optimal specificity is achieved when the protein incurs a heavy deformation penalty which can be paid off by binding to the target molecule, but not to the off-target molecule (Fig. 3D).

(C) *Discrimination requires finetuning of multiple coupled factors:* In principle, better discrimination is achieved by the most rigid proteins, but the benefit of high rigidity is marginal. Instead of maximizing rigidity, it is much more important for proteins to carefully tune the interplay between (i) ligand mismatch, (ii) the energy scale of binding (*i.e.*, number, and strength of potential bonds), and (iii) flexibility (Fig. 3C). As a result, there are no monotonic relations between specificity and any of these three components.

This novel prediction is at odds with conventional wisdom in the fields of enzyme and antibody evolution, where it is often assumed that specificity monotonically decreases with increasing flexibility.[48–52] This prediction can explain reportedly 'counterintuitive' behavior in a family of esterases, where specificity increases with increasing flexibility.[53] Since the esterases in question have large, open binding pockets, and specificity is found to increase with decreasing binding pocket size,[54] we can infer the approximate location of these esterases on the phase diagram (Fig. 3C, purple arrow), and predict that further increases in flexibility will eventually lead to a decrease in specificity (Fig. 3C, red arrow).

Likewise, the predictions of the affinity-specificity phase diagram are compatible with the results of a study on antibody evolution, where an evolutionary trajectory selecting for affinity first led to increased flexibility, followed by decreased flexibility (Fig. 3C, cyan arrows).[55] Our model predicts that this evolutionary trajectory is expected when the initial antibody is a poor match for the substrate. In general, we speculate that the reason for the inferred monotonic relation between flexibility and specificity could be a bias from observing wild-type proteins; if proteins are optimized for specificity, then they will fall along the optimal line (Fig. 3C, dashed line), along which flexibility and specificity are proportional to each other.

(D) *Distant residues are essential*: A key finding from the **G-MeCh** model about the role of flexibility in molecular discrimination is that residues *far* from the binding site are important for discrimination. While mutations at distal residues have relatively small effects on binding modes, compared to residues at the binding site, these small effects are exactly what is needed for achieving precision in binding interactions.

(E) *Why are proteins so large?* The **G-MeCh** model shows that *larger* proteins are able to discriminate between a greater range of ligands, and are more evolvable and robust.[16] This can be explained by the findings that good specificity is achieved by precise tuning of flexibility, mismatch, and chemical binding energy, and that greater precision is achieved through fine-tuning via mutations at residues distal to the binding site. Extra amino acids in the protein sequence afford greater degrees of freedom to tune for accurate discrimination.

This leads to the hypothesis that the difficulty of achieving a certain degree of specificity between ligands sets a minimum size constraint on proteins. In other words, the minimum size of a functional protein scales with the specificity requirements of the discrimination task. We call this the *size-specificity-scaling hypothesis*. Proteins need to be at least as big as some minimum size, and being larger increases evolvability and robustness, thus facilitating evolution (Fig. 4A). The requirements of specificity thus can explain why proteins are so large – they are not simply structural supports for active sites, but rather expansive molecular machines that are precisely calibrated for function.

The size-specificity-scaling hypothesis might be hard to test because the task difficulty lacks a rigorous mathematical definition. But we can circumvent this issue by breaking down task difficulty into two components: the difficulty of specifically binding to one molecule over another, X_{diff}, and the required specificity, $\Delta\Delta G$, between the two molecules. While X_{diff} is hard to define, we can test the hypothesis even without a definition by controlling for X_{diff}.

Thankfully, there exists a natural test case for this hypothesis – amino-acyl tRNA synthetases (AARS). For certain amino acids, we can perform pairwise comparisons where we know that X_{diff} is higher in one case than the opposite case. For example, distinguishing tyrosine from phenylalanine is easier than the reverse task due to the increased chemical binding energy thanks to the extra hydroxyl group on tyrosine. Likewise, distinguishing threonine from serine is easier than the reverse case because the extra methyl group on threonine does not afford much greater binding energy, but it allows serine to be favoured through steric repulsion of threonine.

As predicted by theory, the AARSs with easier tasks not only have higher selectivity, but are also smaller (Fig. 4B). Furthermore, non-AARS amino-acid binding proteins are much smaller than AARSs (Fig. 4C), which is what we expect given that AARSs require much greater selectivity, due to prohibitively high cost of incorrect protein translation. We will return to this in the discussion section with a conceptually simple, and feasible experiment to test this hypothesis.

Recap: why do proteins move? Molecular discrimination – which is more difficult than mere recognition – is predicted to benefit from internal motion in proteins by the theories of induced fit, conformational proofreading, and our genetic-mechano-chemical model of proteins[12,15,16], which provides a physical basis to this hypothesis. Many essential protein functions require carefully calibrated functional motion: AARSs need to discriminate between amino acids with exceptional precision, and transcription factors need to discriminate between a huge range of similar targets.[56] Antibodies need to bind to pathogens while avoiding products of the host organism, otherwise leading to auto-immune disorders,[57] and enzymes need to simultaneously optimize for both binding to substrates and also populating the transition state, while simultaneously facilitating product release.[58–60]

While we so far presented evidence for how specificity requirements necessitate functional motion in proteins, there may be other important reasons for the evolution of conformational dynamics, such as promiscuity,[61] signalling,[62] and evolvability.[49,63,64] Specificity is essential to many protein functions, but there are also many promiscuous "generalist" proteins with multiple interaction partners.[65,66] For example, proteins involved in host defense, such as broadly-neutralizing antibodies,[67,68] or promiscuity in antibiotic resistance,[69] benefit from the ability to ward off diverse attacks. Signalling is aided by long-ranged, allosteric dynamics, which enable transduction of force across membranes,[70] and switch-like inhibition/activation.[71] Furthermore, it has been hypothesized that promiscuity is an evolutionary driver for functional diversification.[72,73] To summarize, proteins move first because they have to, but the precise way in which they move is dictated by evolutionary history and biological function.

5. Quantifying internal motion

Protein motion can be measured using various techniques that access specific timescales, spatial resolution, and different degrees of molecular and atomic resolution. No one method is superlative, as they each have strengths and deficiencies. Here, we are interested in the methods that provide structural or dynamical information in terms of the full atomic coordinates of proteins, both experimentally (X-ray crystallography, NMR spectroscopy, cryo-EM) and computationally (molecular dynamics, elastic network models, structure prediction algorithms). In particular, we will discuss how one should mathematically approach measuring differences between structures.

What is a good metric of internal motion? When measuring motion between protein conformations, we need to specify both the set of atoms to compare, and the frame of reference in which to compare them. We can consider *local* motion between neighboring amino acids, or the *global* motion that compares all amino acids. Most methods involve some form of *structural alignment*, which defines a frame of reference. The alignments are needed to eliminate differences in atomic coordinates due to whole-body translations and rotations. However, there is no single, universally correct method of alignment, since proteins consist of many moving parts that can individually undergo translations and rotations (Fig. 5). Ideally, one desires methods for measuring protein motion on both local and global scales, that are insensitive to the choice of alignment.

The most common metric of protein structural change is root mean square deviation (RMSD), which measures the average displacement between atoms in a target and reference structure, after aligning the target to the reference structure; this is a useful measure of global motion.[74] Similarly, one can examine the displacement per amino acid to get a local measure of motion. The main issue with this measurement is that displacement can be large in areas that do not undergo any local rearrangements (Fig. 5), so certain residues exhibit large shifts from the reference structure, while locally retaining exactly the same configuration.

Many methods have been developed to overcome this problem, yet they all share a common weakness. Methods such as template-modelling score, global distance test, and local distance difference test, are designed to measure similarity instead of difference.[75–77] As a consequence, they underweight outliers (flexible regions tend to move a lot, and this is to be ignored) and are normalized to produce a score instead of a true *metric* in the mathematical sense. In the following, we will consider how to measure local motion in proteins in a way that avoids the inherent problems associated with internal motion, while also maintaining consistency across different magnitudes of motion.

Finite strain theory for proteins. For a more rigorous metric of local motion, we look to finite strain theory, which gives us a natural framework for analyzing deformation (Fig. 5).[78] Strain can be broken down into affine and non-affine strains. The affine component includes compression/expansion, shear, and twist motions, which can be de-

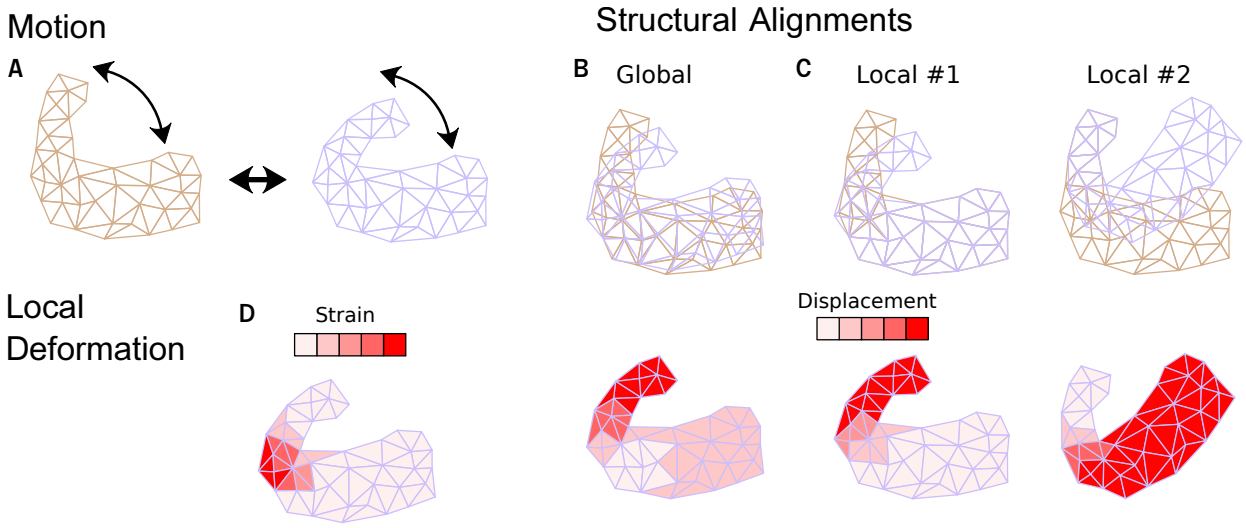

FIG. 5. **Quantifying protein motion and deformation.** A: The internal motion of protein includes separate movements of its sub-units. B,C show possible alignments of the original and deformed structure, where the top figures are the alignment and the bottom ones show the corresponding displacement. B: *Global* measures of motion capture large-scale motion relative to some reference configuration, but they misrepresent local motion due to the use of global alignments. C: *Local* alignments of only a subset of amino acids illustrate this. B, and C demonstrate that measures derived from displacement, such as RMSD, strongly depend on the alignment. D: An alternative approach is to measure *strain*, which is a local measure of deformation. Strain is low if parts of a protein move together as a rigid body, and high if their relative positions change, which implies stretching, compressing and breaking of bonds, and the formation of new bonds. Thus, strain naturally captures local deformation relevant to function.

scribed as linear transformations of the coordinates, and the non-affine component includes all the non-linear residuals. The affine part of strain is captured by the deformation gradient tensor, $\mathbf{F}$, which describes the transformation between the two configurations, $\mathbf{r}$ and $\mathbf{r}'$, as

$$\mathbf{r}' = \mathbf{F}\,\mathbf{r}\ , \tag{3}$$

The diagonal elements of $\mathbf{F}$ correspond to expansion or compression, while the off-diagonal elements relate to shear and twist motions.

In cases where deformation is a smooth function in space (*e.g.*, continuum mechanics), deformation is entirely affine, and the linear relation (3) holds. In discrete systems, as is found at atomic nanoscale, this is no longer the case, and the residual of (3) (*i.e.*the deviation from the linear relation) reveals a non-affine component of the overall strain.[79–81] Here, we explain how to calculate the affine, non-affine quantities for proteins.[82–84]

We consider the case of measuring strain between two protein conformations, which are specified by the positions of their C_α atoms, $\mathbf{r}$ and $\mathbf{r}'$ (each amino acid has one such atom). Strain is a local quantity, which we measure at each amino acid i, by comparing the relative positions of the neighbors of i, $j \in N_i$, between conformations $\mathbf{r}$ and $\mathbf{r}'$. Neighbors of residue i are those residues that are within a distance γ of residue i; *i.e.*, if $r_{ij} = |\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j| \leq \gamma$. We then define a neighborhood tensor $\mathbf{D}_i$ as the tensor of the $n_i = |N_i|$ distance vectors $\mathbf{r}_{ij}$. Thus, we can compute the deformation gradient tensor at each amino acid i of a protein, $\mathbf{F}_i$ as the solution of linear relation,

$$\mathbf{D}'_i = \mathbf{F}_i\,\mathbf{D}_i\ . \tag{4}$$

We find $\mathbf{F}_i$ using linear regression, which minimizes the deviation from Eq.(4) due to non-affine deformation. Then, we can extract the bulk (diagonal), shear (off-diagonal), and non-affine (the residual deviation) components. For proteins, it appears that the magnitude of the bulk component is generally lower than the magnitude of the shear component[82], which is in turn lower than (yet, the same order of magnitude and correlated with) the non-affine component.[84]

To formulate a single metric for measuring deformation that incorporates both affine and non-affine components of strain, we use what we term *effective strain* (ES). This is simply the average relative change in distance vectors between neighbors,

$$S_i = \left\langle \frac{|\Delta\mathbf{r}_{ij}|}{r_{ij}} \right\rangle = \frac{1}{n_i}\sum_{j\in N_i} \frac{\left|\mathbf{r}_{ij} - \mathbf{r}'_{ij}\right|}{r_{ij}}\ . \tag{5}$$

At each amino acid, calculation of ES first requires a local structural alignment of the neighborhood tensors of the deformed and reference states; the corresponding rotation is found using the Kabsch algorithm.[85] See Eckmann *et al.*[83] and McBride *et al.*[84] for more details on calculating strain. All of the above methods have been implemented in the python package *PDAnalysis* (github.com/mirabdi/PDAnalysis).

6. Physics and evolution of protein motions

Proteins are easily deformable due to the soft nature of the non-covalent bonds holding them together, which are easily stretched or broken. Breaking of multiple bonds can occur cooperatively which leads to bistability,[87] or even multiple conformational states separated by energy barriers.[88] Bonds stretch and break due to physical forces from binding and colliding with other molecules, or from thermal motion in the solvent. Here, we will consider protein deformation as a response to local forces (*e.g.*,

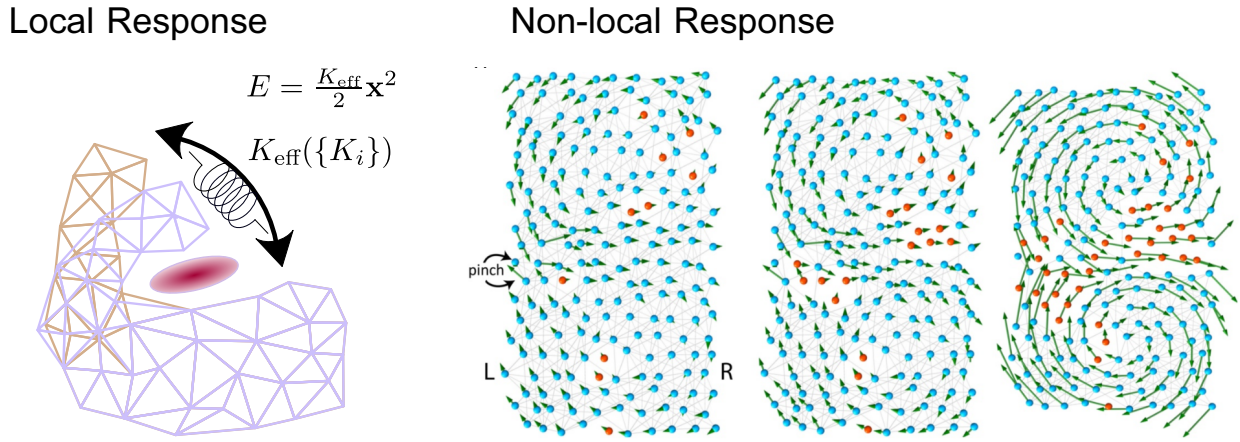

FIG. 6. **How is movement encoded in protein structure?**
Local Response: We consider a protein as an elastic network where neighbors interact via harmonic potentials. A local response occurs when ligand binding causes displacement of amino acids at the binding site. This local deformation is apportioned non-locally throughout the spring network, thereby conveying long-range elastic forces. For any specific displacement vector $\mathbf{x}$, the spring network can be described as a single spring with an effective spring constant $K_{\text{eff}}(\{K_i\})$, where K_i are the individual spring constants of each spring i. This means that mutations far from the binding site can affect the response of the protein to binding by altering K_{eff} for $\mathbf{x}$. *Non-local Response:* Protein responses to ligand binding can be long-ranged (*e.g.*, allosteric regulation, signal transduction), such that a molecule binds at one site and causes motion at a distal site. In a 2-dimensional physical model of proteins,[86] with two types of amino acids (HP model: 'H', blue, forms strong bonds; 'P', red, forms weak bonds), the protein is evolved so that a pinch force $\mathbf{f}$ at one end produces a response $\mathbf{v}$ at the other end. The typical solution obtained through evolution is a channel of weakly bonded amino acids that undergo high strain in response to a pinch, which allows force to be transduced through the protein.

induced fit), and as a response to non-local forces (*e.g.*, allosteric inhibition/activation). We will examine two models,[16,86,89], where, for simplicity, only elastic forces are considered without explicit bond breaking.

Elastic network models. Motion at binding sites in response to force, whether large-scale or minute, is determined by the many-body interaction network of the amino acids, which is encoded in the gene. The amino acids closest to the binding site are most critical for protein function as they largely determine the geometry and chemistry of the binding site. Nevertheless, mutations at distal residues have also been shown to affect binding and catalysis, independently from confounding effects on folding stability.[90] In some cases, these residues have been found to impact dynamics by controlling the probability of conformational sub-states.[91,92]

One possible mechanism for such non-local effects of distant amino acids can be understood by treating a protein as an elastic network, in which mechanical forces are transmitted over long ranges.[93,94] To intuit the basic idea, consider the very simple case of a 1D protein made of a linear chain of N springs with constants $\{K_i\}$. Assume that successful binding requires that amino acids at the binding site are displaced by x relative to its neighbors. If this deformation remains localized in a single bond with spring constant K_1, then the elastic energy is $\mathcal{E}_1 = K_1 x^2/2$. But if we let this displacement distribute over the N springs, the overall energy relaxes to $\mathcal{E}_N = K_{\text{eff}} x^2/2$, where the effective spring constant is the harmonic mean, $K_{\text{eff}}(\{K_i\}) = [\sum_{i=1}^{N} K_i^{-1}]^{-1}$, and the motion of each spring is inversely related to its strength, $x_i \sim 1/K_i$. Proteins are much more complex, of course, as they consist of networks of interconnected bonds, also in parallel, but the basic principle remains the same: the response to any localized force can be described using a single effective spring constant, which is a simple function of all spring constants and network connectivity, $K_{\text{eff}}(\{K_i\})$(*Fig. 5*).[16,95] This explains how the internal motion $\mathbf{x}$ is encoded in the topology and strength of the bonds. Next, we explain how this can be encoded in a protein's gene sequence.

In our **G-MeCh** model, we use a binary sequence of amino acids whose identities encode the strengths of springs connecting the $n \leq 6$ adjacent amino acids.[16] This means that each mutation can alter the effective strength K_{eff} by a discrete amount. Since each amino acid affects K_{eff} to a different degree, larger proteins with more amino acids offer more ways to fine-tune K_{eff}, which enables more precise interactions between proteins and ligands. In summary, the simple spring network model explains (i) why and how distant residues are involved in deformation due to binding, (ii) how the collective properties of many amino acids can be reduced to a low-dimensional representation in terms of functional motion, and (iii) how this is encoded in sequence.[83,86,96]

Protein sequences encode non-local response. Since forces can propagate throughout proteins, they could evolve functions, such as signal transduction and regulation, involveing long-ranged communication between amino acids, known as allostery.[13,14] Much of our understanding of allostery comes from elastic network models and molecular dynamics (MD) simulations, where it has been shown that the way force propagates along a network depends on its topology,[62,97–100] and that allosteric transduction pathways in proteins are facilitated by low-energy, slow modes in the intrinsic dynamics of proteins.[101] Here, we will revisit this question with a focus on how proteins *evolve* such allosteric function, through a minimal model of protein function and evolution.[86,89]

Models are always some compromise between "realism" (how close they are to real proteins) and simplicity, which allows fast computations and the ability to see something in the high-dimensional spaces of protein evolution. The main difficulty in studying protein evolution is the "curse of dimensionality": the need to obtain large enough statistics of long evolutionary paths, such that we can filter out the noise of random fluctuations. Nevertheless, if we wish to understand the physical response of the system to localized forces (i.e. binding induced allostery), there exists an old physical concept that achieves exactly this: the Green function.[102]

The Green function describes the motion of the whole system as a response to a local perturbation: If a ligand binds at a certain site, many amino acids will move, depending on the network of intramolecular forces, and the interactions with the salty water at the protein surface (*e.g.*, the formation of hydrogen bonds with water molecules). As described below, we use Green's functions to build a simplified model of protein as an evolving elastic network whose interactions are encoded in a gene (Fig. 6).[83,86] To make the model tractable for evolutionary simulations, we make the following simplifications: (i) The protein is described as a 2-dimensional network. (ii) There are only two types of amino acids: H - hydrophobic and P - polar, following the HP model.[103] The genetic space contains genes $\mathbf{g}$, which are strings of zeros and ones. (iii) Intra-protein interactions are harmonic, and bonds are at minimum energy in the ground state. All three assumptions can be relaxed if one wishes to be more realistic about a certain aspect.

To formally define the Green function, we consider a toy protein made of a harmonic spring network. The elastic response of the protein is captured by an elasticity tensor $\mathbf{H}$. Basically, this tensor is a generalized spring constant $\mathbf{H}$ that tells us what force field $\mathbf{f}$ occurs in response to a motion field $\mathbf{u}$, through the linear relation, $\mathbf{f} = \mathbf{H}\,\mathbf{u}$ (the fields $\mathbf{u}$ and $\mathbf{f}$ assign force and displacement vectors to each amino acid of the protein). The elasticity tensor $\mathbf{H}$ depends on the strength and $\{K_i\}$ connectivity of the spring between the amino acids, which are encoded by the gene. Therefore, we can write $\mathbf{H}$ as a function of the gene $\mathbf{H}(\mathbf{g})$. From $\mathbf{H}$, we can derive the effective spring constant for any displacement $K_{\text{eff}}(\{K_i\})$. As usual, the Green function is then the inverse tensor $\mathbf{G}(\mathbf{g}) = \mathbf{H}(\mathbf{g})^{-1}$, which is evidently also a function of the gene. The function $\mathbf{G}(\mathbf{g})$ solves the inverse problem: what is the motion induced by a force field $\mathbf{u}(\mathbf{g})$? *i.e.*, $\mathbf{u}(\mathbf{g}) = \mathbf{G}(\mathbf{g})\,\mathbf{f}$.[83,86]

In this theoretical picture, the performance of the protein is measured by the similarity of the induced response $\mathbf{u}(\mathbf{g}) = \mathbf{G}(\mathbf{g})\,\mathbf{f}$ and a reference functional collective mode $\mathbf{v}$, which can represent for example the motion during induced-fit binding. We can therefore define a performance measure F (kind of "fitness") of our toy protein as the overlap, $F(\mathbf{g}) = \mathbf{v}^{\mathsf{T}} \cdot \mathbf{u} = \mathbf{v}^{\mathsf{T}}\mathbf{G}(\mathbf{g})\,\mathbf{f}$. Thus, in this simple model, we have an explicit *genotype to phenotype map*, $\mathbf{g} \to \mathbf{u}(\mathbf{g}) = \mathbf{G}(\mathbf{g})\,\mathbf{f}$, and a corresponding *fitness landscape*, $\mathbf{g} \to F(\mathbf{g}) = \mathbf{v}^{\mathsf{T}}\mathbf{G}(\mathbf{g})\,\mathbf{f}$.

With this simple mathematical description, we can perform many simulations of evolutionary trajectories, starting from a random gene $\mathbf{g}_0$ until we find a functional gene $\mathbf{g}^*$, where functionality means having a fitness above a certain threshold, $F(\mathbf{g}^*) \geq F^*$. A typical simulation of how evolution reaches a functional protein is shown in (Fig. 6). In general, the model explains how functional proteins with regions of localized high strain (a.k.a. "shear bands") emerge during evolution. In this physical picture, mutations induce local perturbations in the amino acid network, which shape the genotype-to-phenotype map. In particular, mutations induce significant structural deformation at high-strain regions (Fig. 6), and the evolutionary interaction between these mutations (epistasis) corresponds to physical interaction among these "defects". This epistasis shapes the genotype-to-phenotype map.

7. Protein evolution as effective internal motion

Like *physical* perturbations (binding, physicochemical environment changes), *evolutionary* perturbations – that is, structural changes resulting from mutations – can induce effective motion in the conformational free energy landscape of a protein, albeit on vastly different timescales. Evolutionary perturbations may cause allosteric changes that affect structure rather than (or in addition to) dynamics, and these perturbations may have a physical basis similar to that of perturbations induced by physical forces.[89,104,105] Despite mounting evidence, the exact mapping between the response to evolutionary perturbations and physical perturbations is still unclear. The mapping is easy to deduce in simple models but is much more elaborate in real proteins. In the following, we consider the effect of evolutionary perturbations on structure, using effective strain (ES) as a natural metric.

To this end, we use DeepMind's protein structure prediction algorithm, AlphaFold (AF), to predict the effect of mutation on structure as a strain field.[84] As a case study, we measure strain between the WT CypA and a double mutant (S99T, C115S), using three types of structural data (Fig. 7A): structures from the protein data bank (PDB); AF-predicted structures; an "average" structure ($\langle$AF$\rangle$) created by averaging over local neighborhoods of multiple AF-predicted structures, which has been shown to give a more precise estimate of deformation due to mutation (as opposed to stochastic noise).[84,106] We see that the strain in PDB structures depends on distance from the nearest mutated residue, δ_{m}, within the first 1 nm; beyond this high strain is an artefact of high variance between repeat measurements in flexible regions of PDB structures.[84] In contrast, AF has much lower measurement variance,[84], and we see a clear exponential dependence of strain on δ_{m}; this dependence is exceptionally clear for strain calculated using $\langle$AF$\rangle$ structures.

To understand the range of mutation effects, we report analyses of a set of 3,901 proteins taken from McBride *et al.*[84], in which pairs of proteins differ by up to three amino acid substitutions. The geometric mean of strain, $\langle S_i \rangle$, as a function of δ_{m} shows that PDB and AF-predicted structures permit reliable estimation of mutation effects up to about 15 Å from mutated residues (Fig. 6B); at high δ_{m}, strain increases since extremal residues tend to be in flexible loops and tails. Averaging ($\langle$AF$\rangle$) across many AF-predicted structures enables measurement of fine-grained mutation effects up to 25 Å from mutated residues. Altogether, our results demonstrate shows how evolutionary perturbations (*i.e.*, mutations) induce structural changes that can be seen as effective internal motion of proteins. The natural measure of this effective motion is strain, which

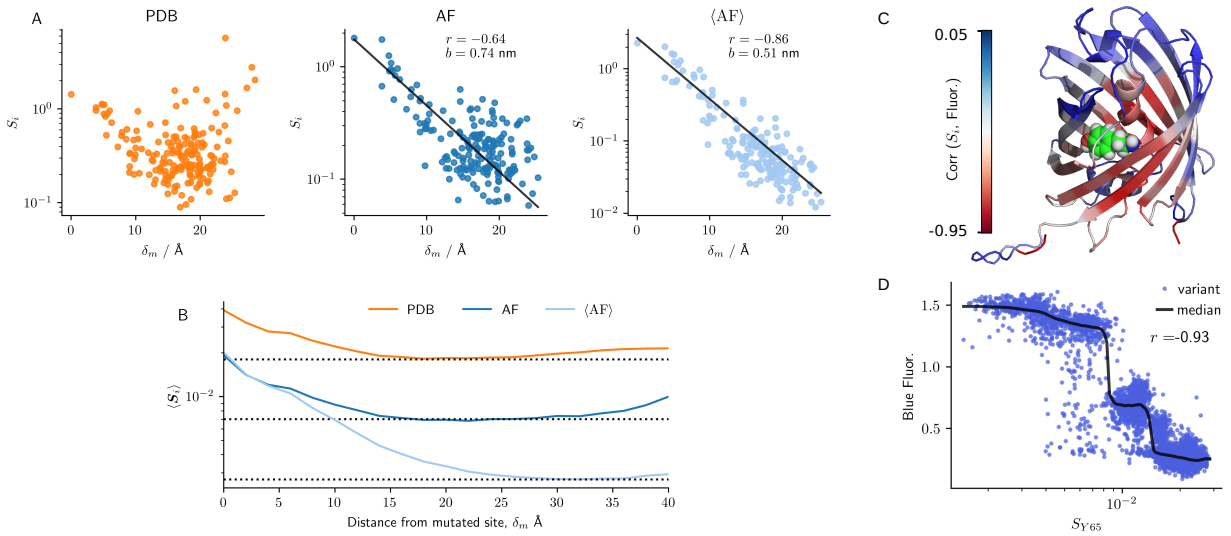

FIG. 7. **Protein structure evolution as effective motion**. A: Effective strain S_i vs distance from a mutated site δ_m, calculated using wild-type (WT) CypA (`6U5C_A`) and a double mutant (S99T, C115S, `6BTA_A`); results are shown for PDB structures, AF-predicted structures, and effective strain (ES) calculated using "averaged" structures obtained from multiple AF predictions (⟨AF⟩);[84] black lines indicate fits to the function $S_i = ae^{-\delta_m/b}$. B: Average strain $\langle S_i \rangle$ (over 12,627 pairs of structures[84]) vs δ_m for PDB, AF, and ⟨AF⟩ structures. C: Blue fluorescent protein (BFP) colored by how well deformation at each residue S_i correlates with fluorescence; chromophore-binding site (Y65) is shown as spheres. D: Fluorescence vs ES (compared to WT) at residue Y65 for 8,191 BFP variants; moving median line and Pearson's r are shown.

decays exponentially with distance from the perturbation.

Evolutionary deformation is linked to function. We have shown that evolutionary perturbations give rise to changes in structure, akin to motion due to physical perturbations. Ideally, we would discuss the effect of evolutionary perturbations on dynamics as well, but this is less clear – recent augmentations to the AlphaFold algorithm may enable such investigations, but it is yet to be seen whether AlphaFold can inform the effect of mutations on dynamics.[107,108] Nevertheless, we ask whether one can deduce how a mutation affects function merely from knowing the structural change such a perturbation induces.

To address this question, we studied three protein families and examined the relationship between evolutionary strain and phenotype for proteins whose genotype-to-phenotype space has been partially mapped via deep mutational scanning experiments.[84] Here we highlight one such protein, the blue fluorescent protein (BFP, Fig. 7C).[109] We find that by measuring the strain induced at the chromophore binding site by mutations at a range of sites, we can predict with great accuracy the effect of these mutations on fluorescence (Fig. 7C-D). Fluorescence dims if strain at the chromophore binding site exceeds 0.008, and if it passes exceeds 0.013 then fluorescence is lost entirely.

Our findings shine light on the elaborate mapping between sequence and function – that allosteric deformations at a functionally important residue can be sufficient to explain much of the variance in function. Furthermore, we have shown that AF-predicted strain correlates with other important phenotypes, such as catalytic activity and stability.[84,106] These results demonstrate how deep mutational scanning, along with AlphaFold, have opened up a scintillating new avenue for understanding how proteins work and evolve.

8. Discussion

We have presented a theory of why motion is essential for protein function, and how it is encoded in structure and sequence. Although there are multiple hypotheses about why protein motion is important for function, we have proposed that it is absolutely essential for achieving *superlative discrimination* between molecules. This motion is encoded throughout proteins, due to the long-ranged nature of force transduction in elastic (and plastic) networks. In the following, we will highlight several predictions that emerge from our physical and mechano-chemical models of proteins, and suggest experimental tests for these predictions.

Binding-induced strain highlights functional residues. Specific binding and allosteric force transduction is achieved by a precise calibration of the dynamical response to physical interactions between proteins and ligands (Fig. 6). The residues that contribute most to this dynamical response are those that undergo high strain upon binding.[110] These are exactly residues that have an outsize effect on the effective spring constant K_{eff} that dictates the internal motion of the protein in response to binding. In particular, the existence of high-strain regions is found to be crucial for long-range transduction of force (Fig. 6). We show that the appropriate metric for measuring deformation is strain, and that computing the effective strain is a simple method for reliably estimating deformation (Fig. 5).

What follows is a potential answer to the longstanding question of how residues *far* from binding sites affect the binding. We predict that mutations at residues that experience high strain upon binding should lead to large changes in function, even if they are far from the binding site. At first, one might think that the structures in the PDB offer a quick way to test this hypothesis. However, PDB data should be used with great caution since strain measured using PDB structures is often a sign of thermal fluctuations

in flexible regions[84], although the strain may still be informative in cases of significant motions. Instead, based on our experience of using AF, we recommend measuring strain between conformational ensembles, by averaging over multiple structures to arrive at a more precise measure of functional deformation.[84,106] Thus, strain measurements can be used to target mutations far from the binding site that are predicted to have a significant effect on catalysis and/or binding, which can be measured experimentally.

Possible links of physical strain and evolutionary strain. Theoretical results from elastic network models of force propagation,[16,83,86,89] along with empirical and computational estimation of effects of mutations with distance from mutated sites,[84,106] show that force can propagate throughout networks of connected residues. It stands to reason that the same underlying physics can explain both phenomena. In order to show an equivalence between evolutionary and physical perturbations, we can study strain in response to both force and mutation. We can also study free energy landscapes before and after mutation, and before and after binding. Is there a rigorous equivalence, even partially, between the effects of binding and mutation? Answering this would yield a unified theory of proteins where physics and evolution are intertwined.

Towards a phase diagram of specificity for real proteins. We presented a theoretical affinity-specificity phase diagram for a minimal model of protein-ligand binding.[16] This study can help conceptualize the factors leading to successful molecular discrimination and their interplay. However, by studying a thermodynamic model using an elastic network, we ignore two important factors: binding kinetics and plastic deformation. To truly understand how binding specificity works, we need precise measurements of real proteins. One benefit of the minimal model that we would like to retain is the simplicity of the specificity phase diagram, which has only three dimensions: mismatch, flexibility, and energy level. Furthermore, we demonstrated that the dimension can be even further reduced, where the optimal mismatch is a 1D function of ε/K. To achieve similar simplicity in real proteins is difficult as we need to know how many dimensions are required to characterize specificity and what one should measure.

Creating an affinity-specificity phase diagram for real proteins is challenging because of the sheer number of required measurements of multiple proteins: protein-ligand mismatch, bond energy, flexibility, affinity, and specificity. Shape mismatch is conceptually simple but a clear metric for estimating shape mismatch between arbitrary 3-dimensional shapes is lacking (although advances in machine learning may make this possible).[111] As an alternative, we suggest estimating shape mismatch indirectly, by measuring instead how much a protein needs to deform upon binding. This can be estimated using strain and studied using experimental structural data[112], machine learning predictors,[113,114] flexible docking,[115] or MD simulations.[116,117] Bond energy can be calculated using a similar approach, and is 1D by definition. In contrast, flexibility is hard to reduce to a single dimension since the binding mode must be identified, and when ligands are large molecules (*e.g.*, proteins, nucleotides), or when there are multiple simultaneously-binding subtrates, there may be several binding modes. But at least in the case of small molecules, it may be possible to represent a single binding mode in terms of one effective spring constant K_{eff}. This can be measured using MD simulations, or elastic network models. Finally, unlike the above quantities, affinity and specificity can be measured by existing fast and accurate experimental assays. Additionally, such measurements are readily available in datasets,[118–120] or published alongside high-throughput experiments.[121] Computational tools such as MD[122] and machine learning[123] can also be used to predict affinity. Thus we expect that, although it requires much work, the determination of an affinity-specificity phase diagram is well within reach using existing data and methods. This direction may open a new era of theory in protein science, where theories of binding are quantitative and predictive,[16,124] rather than qualitative and descriptive.[12–14]

Does molecular discrimination limit protein size? The *size-specificity-scaling* hypothesis predicts that specificity requirements impose a minimum size constraint on proteins (Fig. 4A). The most obvious way to test this would be to take a protein and to try to make it smaller without decreasing its functional specificity. However, this is extremely impractical due to the severely deleterious impact of deletions.[125] Inverting this approach seems considerably more tractable: take proteins of different size that are known to bind to the same molecule, and try to engineer mutants using directed evolution that maximize specificity. Amino-acid binding proteins are a promising test case for this approach since there are many of these, ranging in size from hundreds to a thousand amino acids (Fig. 4C). Our hypothesis has two clear predictions: larger proteins will be able to achieve higher selectivity, and they will also be easier to evolve.

Acknowledgements This work was supported by the Institute for Basic Science, Project Code IBS-R020-D1.

[1] M. Kurzyński, Protein machine model of enzymatic reactions gated by enzyme internal dynamics, Biophysical Chemistry **65**, 1 (1997).
[2] B. Alberts, The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists, Cell **92**, 291 (1998).
[3] K. Kinbara and T. Aida, Toward Intelligent Molecular Machines: Directed Motions of Biological and Artificial Molecules and Assemblies, Chem. Rev. **105**, 1377 (2005).
[4] D. S. Goodsell, *The Machinery of Life*, 2nd ed. (Copernicus Books, New York, 2009).
[5] H. Flechsig and A. S. Mikhailov, Simple mechanics of protein machines, J. R. Soc. Interface. **16**, 20190244 (2019).
[6] S. N. Timasheff, The Control of Protein Stability and Association by Weak Interactions with Water: How Do Solvents Affect These Processes?, Annu. Rev. Biophys. Biomol. Struct. **22**, 67 (1993).
[7] M. Schaefer, M. Sommer, and M. Karplus, pH-Dependence of Protein Stability: Absolute Electrostatic Free Energy Differences between Conformations, J. Phys. Chem. B **101**, 1663 (1997).

[8] F. Meersman, L. Smeller, and K. Heremans, Protein stability and dynamics in the pressure–temperature plane, Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics **1764**, 346 (2006).

[9] E. P. O'Brien, B. R. Brooks, and D. Thirumalai, Effects of pH on Proteins: Predictions for Ensemble and Single-Molecule Pulling Experiments, J. Am. Chem. Soc. **134**, 979 (2012).

[10] D. W. Anderson, F. Baier, G. Yang, and N. Tokuriki, The adaptive landscape of a metallo-enzyme is shaped by environment-dependent epistasis, Nat Commun **12**, 3867 (2021).

[11] L. M. Pandey, Physicochemical factors of bioprocessing impact the stability of therapeutic proteins, Biotechnology Advances **55**, 107909 (2022).

[12] D. E. Koshland, Application of a Theory of Enzyme Specificity to Protein Synthesis, Proc. Natl. Acad. Sci. U.S.A. **44**, 98 (1958).

[13] D. E. Koshland, G. Némethy, and D. Filmer, Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits *, Biochemistry **5**, 365 (1966).

[14] J. Monod, J. Wyman, and J.-P. Changeux, On the nature of allosteric transitions: A plausible model, Journal of Molecular Biology **12**, 88 (1965).

[15] Y. Savir and T. Tlusty, Conformational Proofreading: The Impact of Conformational Changes on the Specificity of Molecular Recognition, PLoS ONE **2**, e468 (2007).

[16] J. M. McBride, J.-P. Eckmann, and T. Tlusty, General Theory of Specific Binding: Insights from a Genetic-Mechano-Chemical Protein Model, Molecular Biology and Evolution **39**, msac217 (2022).

[17] K.-i. Okazaki and S. Takada, Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms, Proc. Natl. Acad. Sci. U.S.A. **105**, 11182 (2008).

[18] A. Cembran, J. Kim, J. Gao, and G. Veglia, NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding, Phys. Chem. Chem. Phys. **16**, 6508 (2014).

[19] S. A. Wankowicz, S. H. de Oliveira, D. W. Hogan, H. van den Bedem, and J. S. Fraser, Ligand binding remodels protein side-chain conformational heterogeneity, eLife **11**, e74114 (2022).

[20] T. R. Stachowski and M. Fischer, Large-Scale Ligand Perturbations of the Protein Conformational Landscape Reveal State-Specific Interaction Hotspots, J. Med. Chem. **65**, 13692 (2022).

[21] S. Gianni, J. Dogan, and P. Jemth, Coupled binding and folding of intrinsically disordered proteins: What can we learn from kinetics?, Current Opinion in Structural Biology **36**, 18 (2016).

[22] C. J. Jeffery, Protein moonlighting: What is it, and why is it important?, Phil. Trans. R. Soc. B **373**, 20160523 (2018).

[23] E. Bertolini, G. Babbi, C. Savojardo, P. L. Martelli, and R. Casadio, MultifacetedProtDB: A database of human proteins with multiple functions, Nucleic Acids Research , gkad783 (2023).

[24] X. Zhuang, C. Yang, K. R. Murphy, and C.-H. C. Cheng, Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids, Proc. Natl. Acad. Sci. U.S.A. **116**, 4400 (2019).

[25] L. Gerasimavicius, B. J. Livesey, and J. A. Marsh, Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure, Nat Commun **13**, 3895 (2022).

[26] P. M. Harrison and M. Gerstein, Studying Genomes Through the Aeons: Protein Families, Pseudogenes and Proteome Evolution, Journal of Molecular Biology **318**, 1155 (2002).

[27] M.-C. Hung and W. Link, Protein localization in disease and therapy, Journal of Cell Science **124**, 3381 (2011).

[28] I. M. Nooren, NEW EMBO MEMBER'S REVIEW: Diversity of protein-protein interactions, The EMBO Journal **22**, 3486 (2003).

[29] M. N. Gupta and V. N. Uversky, Moonlighting enzymes: When cellular context defines specificity, Cell. Mol. Life Sci. **80**, 130 (2023).

[30] J. J. Marchalonis, M. K. Adelman, S. F. Schluter, and P. A. Ramsland, The antibody repertoire in evolution: Chance, selection, and continuity, Developmental & Comparative Immunology **30**, 223 (2006).

[31] F. Hayes and L. Van Melderen, Toxins-antitoxins: Diversity, evolution and function, Critical Reviews in Biochemistry and Molecular Biology **46**, 386 (2011).

[32] G. Invernizzi, E. Papaleo, R. Sabate, and S. Ventura, Protein aggregation: Mechanisms and functional consequences, The International Journal of Biochemistry & Cell Biology **44**, 1541 (2012).

[33] G. Vert and J. Chory, Crosstalk in Cellular Signaling: Background Noise or the Real Thing?, Developmental Cell **21**, 985 (2011).

[34] B. Vögeli and T. J. Erb, 'Negative' and 'positive catalysis': Complementary principles that shape the catalytic landscape of enzymes, Current Opinion in Chemical Biology **47**, 94 (2018).

[35] E. Fischer, Einfluss der configuration auf die wirkung der enzyme, Berichte der deutschen chemischen Gesellschaft **27**, 2985 (1894).

[36] W. H. Rastetter, Enzyme engineering, Appl Biochem Biotechnol **8**, 423 (1983).

[37] Z. Bu and D. J. Callaway, Proteins MOVE! Protein dynamics and long-range allostery in cell signaling, in *Advances in Protein Chemistry and Structural Biology*, Vol. 83 (Elsevier, 2011) pp. 163–221.

[38] D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes, Localizing frustration in native proteins and protein assemblies, Proc. Natl. Acad. Sci. U.S.A. **104**, 19819 (2007).

[39] D. U. Ferreiro, E. A. Komives, and P. G. Wolynes, Frustration in biomolecules, Quarterly Reviews of Biophysics **47**, 285 (2014).

[40] T. Holyoak, Molecular Recognition: Lock-and-Key, Induced Fit, and Conformational Selection, in *Encyclopedia of Biophysics*, edited by G. C. K. Roberts (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013) pp. 1584–1588.

[41] F. B. Straub and G. Szabolcsi, Remarks on the dynamic aspects of enzyme structure, in *Molecular Biology: Problems and Perspectives*, edited by A. E. Braunstein (Nauka, Moscow, 1964) pp. 182–7.

[42] B. G. Vértessy and F. Orosz, From "fluctuation fit" to "conformational selection": Evolution, rediscovery, and integration of a concept, Bioessays **33**, 30 (2011).

[43] C.-J. Tsai, B. Ma, and R. Nussinov, Folding and binding cascades: Shifts in energy landscapes, Proc. Natl. Acad. Sci. U.S.A. **96**, 9970 (1999).

[44] P. Csermely, R. Palotai, and R. Nussinov, Induced fit, conformational selection and independent dynamic segments: An extended view of binding events, Trends in Biochemical Sciences **35**, 539 (2010).

[45] D. Bucher, B. J. Grant, and J. A. McCammon, Induced Fit or Conformational Selection? The Role of the Semi-closed State in the Maltose Binding Protein, Biochemistry **50**, 10530 (2011).

[46] Y. Savir and T. Tlusty, Reca-mediated homology search as a nearly optimal signal detection system, Molecular Cell **40**, 388 (2010).

[47] Y. Savir and T. Tlusty, The Ribosome as an Optimal Decoder: A Lesson in Molecular Recognition, Cell **153**, 471 (2013).

[48] J. Zimmermann, E. L. Oakman, I. F. Thorpe, X. Shi, P. Abbyad, C. L. Brooks, S. G. Boxer, and F. E. Romesberg, Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics, Proc. Natl. Acad. Sci. U.S.A. **103**, 13722 (2006).

[49] A. Pabis, V. A. Risso, J. M. Sanchez-Ruiz, and S. C. Kamerlin, Cooperativity and flexibility in enzyme evolution, Current Opinion in Structural Biology **48**, 83 (2018).

[50] D. Petrović, V. A. Risso, S. C. L. Kamerlin, and J. M. Sanchez-Ruiz, Conformational dynamics and enzyme evolution, Journal of The Royal Society Interface **15**, 20180330 (2018).

[51] M. H. V. Van Regenmortel, Specificity, polyspecificity, and heterospecificity of antibody-antigen recognition, Journal of Molecular Recognition **27**, 627 (2014).

[52] A. K. Mishra and R. A. Mariuzza, Insights into the structural basis of antibody affinity maturation from next-generation sequencing, Frontiers in Immunology **9**, 117 (2018).

[53] C. Nutschel, C. Coscolín, B. David, D. Mulnaes, M. Ferrer, K.-E. Jaeger, and H. Gohlke, Promiscuous esterases counterintuitively are less flexible than specific ones, Journal of Chemical Information and Modeling **61**, 2383 (2021).

[54] M. Martínez-Martínez, C. Coscolín, G. Santiago, J. Chow, P. J. Stogios, R. Bargiela, C. Gertler, J. Navarro-Fernández, A. Bollinger, S. Thies, C. Méndez-García, A. Popovic, G. Brown, T. N. Chernikova, A. García-Moyano, G. E. K. Bjerga, P. Pérez-García, T. Hai, M. V. Del Pozo, R. Stokke, I. H. Steen, H. Cui, X. Xu, B. P. Nocek, M. Alcaide, M. Distaso, V. Mesa, A. I. Peláez, J. Sánchez, P. C. F. Buchholz, J. Pleiss, A. Fernández-Guerra, F. O. Glöckner, O. V. Golyshina, M. M. Yakimov, A. Savchenko, K.-E. Jaeger, A. F. Yakunin, W. R. Streit, P. N. Golyshin, V. Guallar, M. Ferrer, and T. I. Consortium, Determinants and prediction of esterase substrate promiscuity patterns, ACS Chemical Biology **13**, 225 (2018).

[55] V. Ovchinnikov, J. E. Louveau, J. P. Barton, M. Karplus, and A. K. Chakraborty, Role of framework mutations and antibody flexibility in the evolution of broadly neutralizing antibodies, eLife **7**, e33038 (2018).

[56] A. Jolma, J. Yan, T. Whitington, J. Toivonen, K. R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, K. Palin, J. M. Vaquerizas, R. Vincentelli, N. M. Luscombe, T. R. Hughes, P. Lemaire, E. Ukkonen, T. Kivioja, and J. Taipale, DNA-Binding Specificities of Human Transcription Factors, Cell **152**, 327 (2013).

[57] M. Mahler and M. J. Fritzler, Epitope specificity and significance in systemic autoimmune diseases, Annals of the New York Academy of Sciences **1183**, 267 (2010).

[58] A. Peracchi, The Limits of Enzyme Specificity and the Evolution of Metabolism, Trends in Biochemical Sciences **43**, 984 (2018).

[59] O. Rivoire, Geometry and Flexibility of Optimal Catalysts in a Minimal Elastic Model, J. Phys. Chem. B **124**, 807 (2020).

[60] O. Rivoire, How Flexibility Can Enhance Catalysis, Phys. Rev. Lett. **131**, 088401 (2023).

[61] S. D. Copley, Shining a light on enzyme promiscuity, Current Opinion in Structural Biology **47**, 167 (2017).

[62] S. J. Wodak, E. Paci, N. V. Dokholyan, I. N. Berezovsky, A. Horovitz, J. Li, V. J. Hilser, I. Bahar, J. Karanicolas, G. Stock, P. Hamm, R. H. Stote, J. Eberhardt, Y. Chebaro, A. Dejaegere, M. Cecchini, J.-P. Changeux, P. G. Bolhuis, J. Vreede, P. Faccioli, S. Orioli, R. Ravasio, L. Yan, C. Brito, M. Wyart, P. Gkeka, I. Rivalta, G. Palermo, J. A. McCammon, J. Panecka-Hofman, R. C. Wade, A. Di Pizio, M. Y. Niv, R. Nussinov, C.-J. Tsai, H. Jang, D. Padhorny, D. Kozakov, and T. McLeish, Allostery in Its Many Disguises: From Theory to Applications, Structure **27**, 566 (2019).

[63] K. E. Johansson and K. Lindorff-Larsen, Structural heterogeneity and dynamics in protein evolution and design, Current Opinion in Structural Biology **48**, 157 (2018).

[64] P. Campitelli, T. Modi, S. Kumar, and S. B. Ozkan, The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution, Annu. Rev. Biophys. **49**, 267 (2020).

[65] I. Nobeli, A. D. Favia, and J. M. Thornton, Protein promiscuity and its implications for biotechnology, Nat Biotechnol **27**, 157 (2009).

[66] A. Cumberworth, G. Lamour, M. M. Babu, and J. Gsponer, Promiscuity as a functional trait: Intrinsically disordered regions as central players of interactomes, Biochemical Journal **454**, 361 (2013).

[67] D. Corti and A. Lanzavecchia, Broadly Neutralizing Antiviral Antibodies, Annu. Rev. Immunol. **31**, 705 (2013).

[68] Y. Chen, X. Zhao, H. Zhou, H. Zhu, S. Jiang, and P. Wang, Broadly neutralizing antibodies to SARS-CoV-2 and other human coronaviruses, Nat Rev Immunol **23**, 189 (2023).

[69] C. Fröhlich, J. Z. Chen, S. Gholipour, A. N. Erdogan, and N. Tokuriki, Evolution of β-lactamases and enzyme promiscuity, Protein Engineering, Design and Selection **34**, gzab013 (2021).

[70] J. T. Groves and J. Kuriyan, Molecular mechanisms in signal transduction at the membrane, Nat Struct Mol Biol **17**, 659 (2010).

[71] D. Kern and E. R. Zuiderweg, The role of dynamics in allosteric regulation, Current Opinion in Structural Biology **13**, 748 (2003).

[72] O. Khersonsky, C. Roodveldt, and D. Tawfik, Enzyme promiscuity: Evolutionary and mechanistic aspects, Current Opinion in Chemical Biology **10**, 498 (2006).

[73] E. Campbell, M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, N. Tokuriki, and C. J. Jackson, The role of protein dynamics in the evolution of new enzyme function, Nat Chem Biol **12**, 944 (2016).

[74] I. Kufareva and R. Abagyan, Methods of protein structure comparison, in *Homology Modeling: Methods and Protocols*, edited by A. J. W. Orry and R. Abagyan (Humana Press, Totowa, NJ, 2012) pp. 231–257.

[75] A. Zemla, Lga: A method for finding 3d similarities in protein structures, Nucleic Acids Res. **31**, 3370 (2003).

[76] Y. Zhang and J. Skolnick, Scoring function for automated assessment of protein structure template quality, Proteins. **57**, 702 (2004).

[77] V. Mariani, M. Biasini, A. Barbato, and T. Schwede, Lddt: A local superposition-free score for comparing protein structures and models using distance difference tests, Method. Biochem. Anal. **29**, 2722 (2013).

[78] J. Lubliner, *Plasticity Theory*, dover ed ed., Dover Books on Engineering (Dover Publications, Mineola, N.Y, 2008).

[79] C. P. Broedersz and F. C. MacKintosh, Molecular motors stiffen non-affine semiflexible polymer networks, Soft Matter **7**, 3186 (2011).

[80] Y. Wang, J. Gong, D. Wirtz, and B. W. Schafer, Affine and non-affine deformations quantified in cytoskeletal networks through three-dimensional form-finding model, Journal of the Mechanical Behavior of Biomedical Materials **72**, 52 (2017).

[81] D. D. Prakashchand, N. Ahalawat, S. Bandyopadhyay, S. Sengupta, and J. Mondal, Nonaffine Displacements Encode Collective Conformational Fluctuations in Proteins, J. Chem. Theory Comput. **16**, 2508 (2020).

[82] M. R. Mitchell, T. Tlusty, and S. Leibler, Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings, Proc. Natl. Acad. Sci. U.S.A. **113**, 10.1073/pnas.1609462113 (2016).

[83] J.-P. Eckmann, J. Rougemont, and T. Tlusty, *Colloquium* : Proteins: The physics of amorphous evolving matter, Rev.

Mod. Phys. **91**, 031001 (2019).
[84] J. M. McBride, K. Polev, A. Abdirasulov, V. Reinharz, B. A. Grzybowski, and T. Tlusty, *AlphaFold2 Can Predict Single-Mutation Effects*, Preprint (Biochemistry, 2022).
[85] W. Kabsch, A solution for the best rotation to relate two sets of vectors, Acta Crystallogr. A **32**, 922 (1976).
[86] S. Dutta, J.-P. Eckmann, A. Libchaber, and T. Tlusty, Green function of correlated genes in a minimal mechanical model of protein evolution, Proc. Natl. Acad. Sci. U.S.A. **115**, 10.1073/pnas.1716215115 (2018).
[87] J. E. Ferrell, Self-perpetuating states in signal transduction: Positive feedback, double-negative feedback and bistability, Current Opinion in Cell Biology **14**, 140 (2002).
[88] Y. Wang, C. Tang, E. Wang, and J. Wang, Exploration of Multi-State Conformational Dynamics and Underlying Global Functional Landscape of Maltose Binding Protein, PLoS Comput Biol **8**, e1002471 (2012).
[89] T. Tlusty, A. Libchaber, and J.-P. Eckmann, Physical Model of the Genotype-to-Phenotype Map of Proteins, Phys. Rev. X **7**, 021037 (2017).
[90] C. J. Markin, D. A. Mokhtari, F. Sunden, M. J. Appel, E. Akiva, S. A. Longwell, C. Sabatti, D. Herschlag, and P. M. Fordyce, Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics, Science **373**, eabf8761 (2021).
[91] A. Tousignant and J. N. Pelletier, Protein Motions Promote Catalysis, Chemistry & Biology **11**, 1037 (2004).
[92] A. N. Naganathan, Modulation of allosteric coupling by mutations: From protein dynamics and packing to altered native ensembles and function, Current Opinion in Structural Biology **54**, 1 (2019).
[93] M. M. Tirion, Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis, Phys. Rev. Lett. **77**, 1905 (1996).
[94] I. Bahar, A. R. Atilgan, and B. Erman, Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential, Folding and Design **2**, 173 (1997).
[95] E. Eyal and I. Bahar, Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models, Biophysical Journal **94**, 3424 (2008).
[96] J.-P. Eckmann and T. Tlusty, Dimensional reduction in complex living systems: Where, why, and how, BioEssays **43**, 2100062 (2021).
[97] L. Yan, R. Ravasio, C. Brito, and M. Wyart, Architecture and coevolution of allosteric materials, Proc. Natl. Acad. Sci. U.S.A. **114**, 2526 (2017).
[98] J. W. Rocks, N. Pashine, I. Bischofberger, C. P. Goodrich, A. J. Liu, and S. R. Nagel, Designing allostery-inspired response in mechanical networks, Proc. Natl. Acad. Sci. U.S.A. **114**, 2520 (2017).
[99] H. Flechsig, Design of Elastic Networks with Evolutionary Optimized Long-Range Communication as Mechanical Models of Allosteric Proteins, Biophysical Journal **113**, 558 (2017).
[100] D. Thirumalai, C. Hyeon, P. I. Zhuravlev, and G. H. Lorimer, Symmetry, Rigidity, and Allosteric Signaling: From Monomeric Proteins to Molecular Machines, Chem. Rev. **119**, 6788 (2019).
[101] Y. Zhang, P. Doruker, B. Kaynak, S. Zhang, J. Krieger, H. Li, and I. Bahar, Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior, Current Opinion in Structural Biology **62**, 14 (2020).
[102] G. Green, *An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism* (T. Wheelhouse, Nottingham, England, 1828).
[103] K. A. Dill, Theory for the folding and stability of globular proteins, Biochemistry **24**, 1501 (1985).
[104] J. Echave, Evolutionary divergence of protein structure: The linearly forced elastic network model, Chemical Physics Letters **457**, 413 (2008).
[105] Q.-Y. Tang and K. Kaneko, Dynamics-evolution correspondence in protein structures, Phys. Rev. Lett. **127**, 098103 (2021).
[106] J. M. McBride and T. Tlusty, *AlphaFold-predicted Deformation Probes Changes in Protein Stability*, Preprint (Biophysics, 2023).
[107] D. Del Alamo, D. Sala, H. S. Mchaourab, and J. Meiler, Sampling alternative conformational states of transporters and receptors with AlphaFold2, eLife **11**, e75751 (2022).
[108] H. K. Wayment-Steele, A. Ojoawo, R. Otten, J. M. Apitz, W. Pitsawong, M. Hömberger, S. Ovchinnikov, L. Colwell, and D. Kern, Predicting multiple conformations via sequence clustering and AlphaFold2, Nature 10.1038/s41586-023-06832-9 (2023).
[109] F. J. Poelwijk, M. Socolich, and R. Ranganathan, Learning the pattern of epistasis linking genotype and phenotype in a protein, Nat. Commun. **10**, 4213 (2019).
[110] R. Ravasio, S. M. Flatt, L. Yan, S. Zamuner, C. Brito, and M. Wyart, Mechanics of Allostery: Contrasting the Induced Fit and Population Shift Scenarios, Biophysical Journal **117**, 1954 (2019).
[111] M. Simonovsky and J. Meyers, DeeplyTough: Learning Structural Comparison of Protein Binding Sites, J. Chem. Inf. Model. **60**, 2356 (2020).
[112] R. Wang, X. Fang, Y. Lu, C.-Y. Yang, and S. Wang, The PDBbind Database: Methodologies and Updates, J. Med. Chem. **48**, 4111 (2005).
[113] J. Tubiana, D. Schneidman-Duhovny, and H. J. Wolfson, ScanNet: An interpretable geometric deep learning model for structure-based protein binding site prediction, Nat Methods **19**, 730 (2022).
[114] M. L. Hekkelman, I. De Vries, R. P. Joosten, and A. Perrakis, AlphaFill: Enriching AlphaFold models with ligands and cofactors, Nat Methods **20**, 205 (2023).
[115] J. Fan, A. Fu, and L. Zhang, Progress in molecular docking, Quant Biol **7**, 83 (2019).
[116] T. Siebenmorgen and M. Zacharias, Computational prediction of protein–protein binding affinities, WIREs Comput Mol Sci **10**, e1448 (2020).
[117] S. Wan, A. P. Bhati, S. J. Zasada, and P. V. Coveney, Rapid, accurate, precise and reproducible ligand–protein binding free energy prediction, Interface Focus. **10**, 20200007 (2020).
[118] M. K. Gilson, T. Liu, M. Baitaluk, G. Nicola, L. Hwang, and J. Chong, BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology, Nucleic Acids Res **44**, D1045 (2016).
[119] J. Jankauskaitė, B. Jiménez-García, J. Dapkūnas, J. Fernández-Recio, and I. H. Moal, SKEMPI 2.0: An updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation, Bioinformatics **35**, 462 (2019).
[120] D. Lingė, M. Gedgaudas, A. Merkys, V. Petrauskas, A. Vaitkus, A. Grybauskas, V. Paketurytė, A. Zubrienė, A. Zakšauskas, A. Mickevičiūtė, J. Smirnovienė, L. Baranauskienė, E. Čapkauskaitė, V. Dudutienė, E. Urniežius, A. Konovalovas, E. Kazlauskas, K. Shubin, H. B. Schiöth, W.-Y. Chen, J. E. Ladbury, S. Gražulis, and D. Matulis, PLBD: Protein–ligand binding database of thermodynamic and kinetic intrinsic parameters, Database **2023**, baad040 (2023).
[121] A. K. Aditham, C. J. Markin, D. A. Mokhtari, N. DelRosso, and P. M. Fordyce, High-throughput affinity measurements of transcription factor and DNA mutations reveal affinity and specificity determinants, Cell Systems **12**, 112 (2021).
[122] A. De Ruiter and C. Oostenbrink, Advances in the calculation of binding free energies, Current Opinion in Structural Biology **61**, 207 (2020).

[123] D. Jones, H. Kim, X. Zhang, A. Zemla, G. Stevenson, W. F. D. Bennett, D. Kirshner, S. E. Wong, F. C. Lightstone, and J. E. Allen, Improved Protein–Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference, J. Chem. Inf. Model. **61**, 1583 (2021).

[124] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Speeding molecular recognition by using the folding funnel: The fly-casting mechanism, Proc. Natl. Acad. Sci. U.S.A. **97**, 8868 (2000).

[125] M. Topolska, A. Beltran, and B. Lehner, *Deep Indel Mutagenesis Reveals the Impact of Insertions and Deletions on Protein Stability and Function*, Preprint (Genomics, 2023).