Evolutionary coupling range varies widely among enzymes depending on selection pressure

Julian Echave (jechave@unsam.edu.ar)

Instituto de Ciencias Físicas, Escuela de Ciencia y Tecnología,

Universidad Nacional de San Martín, Martín de Irigoyen 3100,

1650 San Martín, Buenos Aires, Argentina

(Dated: March 9, 2021)

Abstract

Recent studies proposed that active sites induce long-range evolutionary constraints in enzymes. However, the prevalence of such long-range evolutionary coupling is unknown and its relationship to the underlying physical coupling is unclear. Here I show that evolutionary coupling is not universally long range, but that range varies widely among enzymes, from $2\mathring{A}$ to $20\mathring{A}$, influencing from 10% to 87% of protein sites. Moreover, evolutionary coupling range is not determined by physical coupling, which is short-range for all enzymes, but by selection pressure. The reason for this is the selection-dependent non-linear function that maps physical coupling to evolutionary coupling.

As enzymes evolve, different sites evolve at different rates. The main reason for such variation of evolutionary rate among sites is selection for stability (Bastolla et al. 2017, Echave et al. 2016, Echave and Wilke 2017, Goldstein and Pollock 2017). Until recently, activity constraints were thought to affect just the few residues directly involved in catalysis and their immediate neighbours (Bartlett et al. 2002, Echave et al. 2016). However, recent studies have reported that active sites influence evolutionary rates at long distances (Dean et al. 2002, Jack et al. 2016, Sharir-Ivry and Xia 2018, 2019a,b). It seems reasonable to assume that such long-range evolutionary coupling could result from allosteric couplings, such as those involved in allosteric mutations (Guarnera and Berezovsky 2019, Perica et al. 2014), which would align with the notion that enzymes are particularly optimized for long-range allosteric coupling (Campitelli et al. 2020, Flechsig 2017, Naganathan 2019).

I have three main concerns with the previous view. First, much of the evidence for long-range evolutionary coupling to enzyme active sites is based on analyses that average over enzymes, which may lead to overestimates if range varies among them (Jack et al. 2016). Second, range may also have been overestimated because of imperfect disentangling of activity constraints from other influences. Finally, except for a few allosteric residues, I would expect physical coupling to be the typical short-range exponential decay expected for indirect effects (Maslov and Ispolatov 2007, Maslov et al. 2007, Naganathan 2019, Rajasekaran and Naganathan 2017, Rajasekaran et al. 2017a,b), which would leave long-range evolutionary coupling begging explanation.

To study whether evolutionary coupling to active sites is indeed long-range, to learn if it is universally so for all enzymes or only some, and to understand whether its range is due to enzymes being particularly well coupled or because of other reasons, I use a mechanistic biophysical model with selection on stability and activity (Echave 2019). This stability-activity model of enzyme evolution, M_{SA}, represents evolution as a sequential mutation-selection process in which, at each evolutionary time step, a random mutation is introduced and is either rejected or fixed, depending on a fixation probability that models selection, given by:

$$p^{\mathtt{fix}} = \min(1, e^{-a_S \Delta\Delta G}) \times \min(1, e^{-a_A \Delta\Delta G^*}) \tag{1}$$

where $\Delta\Delta G$ is the mutational change of folding free energy, $\Delta\Delta G^*$ is the mutational change of activation free energy, and a_S and a_A are positive parameters that measure selection pressure on stability and activity, respectively. In its current implementation, $\Delta\Delta G$ and

$\Delta\Delta G^*$ are calculated using the Linearly Forced Elastic Network Model (LFENM), which represents proteins as elastic networks of nodes connected by springs and mutations as perturbations of the springs (Echave 2008, 2019). In LFENM terms, $\Delta\Delta G$ is the mechanical stress introduced by the mutation minus the relaxation of part of this stress by modification of the mutant's conformation, and $\Delta\Delta G^*$ is the change in the deformation energy necessary for the active residues to assume an active conformation. The M_{SA} model has been validated on a dataset of monomeric enzymes of diverse sizes, structures, and functions (Echave 2019). Here, I use the same dataset to explore the extent of the evolutionary coupling between active sites and other residues.

Since I am interested only on coupling related to activity, I calculate activity constraints by turning off selection on stability. First, I obtain the model parameters a_{A} and a_{S} by fitting site-specific amino-acid substitution rates predicted by M_{SA} with empirical rates inferred from multiple sequence alignments (for details, see (Echave 2019)). Then, I set $a_{\mathrm{S}} = 0$ in equation (1) to obtain:

$$p_A^{\mathtt{fix}} = \min(1, e^{-a_{\mathrm{A}}\Delta\Delta G^*}) \tag{2}$$

This is the fixation probability in the hypothetical scenario in which stability constraints do not operate. Averaging over mutations at a given site, we get the site's acceptance rate ω_A:

$$\omega_A =< p_A^{\mathtt{fix}} > . \tag{3}$$

As measure of the degree of activity constraint, it is convenient to use $1 - \omega_A$, the average rejection probability, that varies between 0 for the no-constraint case (no mutations are rejected) and 1 for the infinite-constraint case (all mutations are rejected). Using equations (2) and (3), it follows that

$$1 - \omega_{\mathrm{A}} = 1 - < \min(1, e^{-a_{\mathrm{A}}\Delta\Delta G^*}) > \tag{4}$$

In the right hand side of this equation, $\Delta\Delta G^*$ is the impact of mutating a given residue on the enzyme's activation energy through changes of the active site distribution of conformations. Therefore, it represents the *physical coupling* between the active site and the mutated residue. The left hand term of equation (3), $1 - \omega_{\mathrm{A}}$, is the average number of mutations rejected at the focus site because of their effect on activity ($\Delta\Delta G^*$). Therefore $1 - \omega_{\mathrm{A}}$ represents the *evolutionary coupling* between the active site and the mutated site. It

is important to note that

$$1 - \omega_{\mathrm{A}} \neq 0 \iff \Delta\Delta G^{*} \neq 0. \tag{5}$$

Thus a given residue is evolutionarily coupled to the active site only if it is physically coupled. However, because of the non-linear nature of equation (4), these couplings may behave very differently.

The dependence of physical and evolutionary coupling on distance to the active site for three illustrative examples is shown in Figure 2. In this figure, the coupling range is measured by $d_{1/2}$, the distance at which coupling is half the maximum. Physical coupling ($< \Delta\Delta G^{*} >$) is similar for the three cases, it decreases exponentially with increasing distance, and it is very short range ($d_{1/2}$ is 2.1$\mathring{A}$, 2.0$\mathring{A}$, and 2.0$\mathring{A}$ for 1OYG, 1QK2, and 1PMI, respectively, Figure 1A). In contrast, evolutionary coupling ($1 - \omega_{\mathrm{A}}$) varies among the examples, it is not an exponential but a sigmoid, and its range varies widely ($d_{1/2}$ is 5.4$\mathring{A}$, 8.7$\mathring{A}$, and 15.6$\mathring{A}$ for 1OYG, 1QK2, and 1PMI, respectively). According to equation 4, evolutionary coupling depends on physical coupling ($\Delta\Delta G^{*}$) and selection pressure (a_{A}). Since physical coupling behaves similarly for the three cases of Figure 1, the variation of evolutionary coupling among these examples must be determined by the selection parameter a_A, that is 29.2, 80.2, and 800 for 1OYG, 1QK2, and 1PMI, respectively (Figure 1B).

The distance-dependence of physical and evolutionary coupling of other enzymes is similar to that described in the previous paragraph. Physical coupling ($\Delta\Delta G^{*}$) is a short-range exponential decay very similar for all the enzymes studied (Figure 2A). In contrast, the distance-dependence of evolutionary coupling ($1 - \omega_{\mathrm{A}}$) varies from a short range exponential decay to a long-range sigmoidal decay, depending on the selection pressure parameter a_{A} (Figure 2B). More quantitatively, the range of evolutionary coupling varies widely, from 1.9$\mathring{A}$ to 19.7$\mathring{A}$ (Figure 2C). Accordingly, the fraction of sites subject to functional constraints varies between 10% and 87% (Figure 2D).

In summary, the range of influence of active sites varies widely among enzymes. Evolutionary coupling is not universally short range (Bartlett et al. 2002) or long-range (Jack et al. 2016), but range varies from very short to very long. Fundamentally, the most interesting result is that this wide range variation is not due to variations of physical coupling range, but to variations of activity selection pressure. For all the enzymes studied, physical coupling displays a very similar short-range exponential decrease with distance from the active residues. These similar physical couplings lead to different evolutionary couplings,

depending on a parameter, a_{A}, that measures functional selection pressure. This is easily explained using equation (4) (see Figure 3). For very small a_{A} linearising the exponential leads to: $1-\omega_{\mathrm{A}} \approx a_{\mathrm{A}} < \Delta\Delta G^{*} >$, so that at very low selection pressures both couplings are short-range exponential decays. At larger a_{A}, the short-range exponential physical coupling is non-linearly mapped into a long-range sigmoidal evolutionary coupling. From a complementary perspective, increasing a_{A} amplifies $\Delta\Delta G^{*}$, turning small physical couplings into large evolutionary couplings, which puts weakly coupled distant sites under the influence of evolutionary constraints.

To finish, I would like to mention two research directions implied by the finding that evolutionary coupling range depends on selection pressure. First, since selection pressure varies among enzymes according to their metabolic role, it would be interesting to explore coupling range in the context of metabolic networks (Aguilar-Rodrıguez and Wagner 2018, Alvarez-Ponce et al. 2016, Boucher et al. 2016, Poyatos 2020). Second, these results suggest the intriguing possibility of tuning evolutionary coupling range by manipulating selection pressure in enzyme evolution experiments.

ACKNOWLEDGEMENTS

This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (grant number PIP 112 201501 00385 CO) and by Agencia Nacional de Promoción Científica y Tecnológica (grant number PICT-2016-4209).

DATA AVAILABILITY

The data and code underlying this article are available in Zenodo, at `http://doi.org/10.5281/zenodo.4309233`

Aguilar-Rodrıguez J, Wagner A, 2018. Metabolic determinants of enzyme evolution in a genome-scale bacterial metabolic network. *Genome Biol Evol* 10:3076–3088.

Alvarez-Ponce D, Sabater-Munoz B, Toft C, Ruiz-Gonzalez MX, Fares MA, 2016. Essentiality is a strong determinant of protein rates of evolution during mutation accumulation experiments in Escherichia coli. *Genome Biol Evol* 8:2914–2927.

Bartlett GJ, Porter CT, Borkakoti N, Thornton JM, 2002. Analysis of catalytic residues in enzyme active sites. *J Mol Biol* 324:105–121.

Bastolla U, Dehouck Y, Echave J, 2017. What evolution tells us about protein physics, and protein physics tells us about evolution. *Curr Opin Struct Biol* 42:59–66.

Boucher JI, Bolon DNA, Tawfik DS, 2016. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. *Protein Sci* 25:1219–1226.

Campitelli P, Modi T, Kumar S, Banu Ozkan S, 2020. The role of conformational dynamics and allostery in modulating protein evolution. *Annu Rev Biophys* 49:267–288.

Dean AM, Neuhauser C, Grenier E, Golding GB, 2002. The pattern of amino acid replacements in alpha/beta-barrels. *Mol Biol Evol* 19:1846–1864.

Echave J, 2008. Evolutionary divergence of protein structure: The linearly forced elastic network model. *Chem Phys Lett* 457:413–416.

Echave J, 2019. Beyond Stability Constraints: A Biophysical Model of Enzyme Evolution with Selection on Stability and Activity. *Mol Biol Evol* 36:613–620.

Echave J, Spielman SJ, Wilke CO, 2016. Causes of evolutionary rate variation among protein sites. *Nat Rev Genet* 17:109–121.

Echave J, Wilke CO, 2017. Biophysical Models of Protein Evolution: Understanding the Patterns of Evolutionary Sequence Divergence. *Annu Rev Biophys* 46:85–103.

Flechsig H, 2017. Design of elastic networks with evolutionary optimized long-range communication as mechanical models of allosteric proteins. *Biophys J* 113:558–571.

Goldstein RA, Pollock DD, 2017. Sequence entropy of folding and the absolute rate of amino acid substitutions. *Nat Ecol Evol* 1:1923–1930.

Guarnera E, Berezovsky IN, 2019. On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr Opin Struct Biol* 56:18–27.

Jack BR, Meyer AG, Echave J, Wilke CO, 2016. Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes. *PLOS Biol* 14:e1002452.

Maslov S, Ispolatov I, 2007. Propagation of large concentration changes in reversible protein-binding networks. *Proc Natl Acad Sci U S A* 104:13655–13660.

Maslov S, Sneppen K, Ispolatov I, 2007. Spreading out of perturbations in reversible reaction networks. *New J Phys* 9.

Naganathan AN, 2019. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Curr Opin Struct Biol* 54:1–9.

Perica T, Kondo Y, Tiwari SP, McLaughlin SH, Kemplen KR, Zhang XW, Steward A, Reuter N, Clarke J, Teichmann SA, 2014. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science* 346:12.

Poyatos JF, 2020. Genetic buffering and potentiation in metabolism. *PLoS Comput Biol* 16:1–15.

Rajasekaran N, Naganathan AN, 2017. A self-consistent structural perturbation approach for determining the magnitude and extent of allosteric coupling in proteins. *Biochem J* 474:2379–2388.

Rajasekaran N, Sekhar A, Naganathan AN, 2017a. A Universal Pattern in the Percolation and Dissipation of Protein Structural Perturbations. *J Phys Chem Lett* page acs.jpclett.7b02021.

Rajasekaran N, Suresh S, Gopi S, Raman K, Naganathan AN, 2017b. A General Mechanism for the Propagation of Mutational Effects in Proteins. *Biochemistry* 56:294–305.

Sharir-Ivry A, Xia Y, 2018. Nature of long-range evolutionary constraint in enzymes: Insights from comparison to pseudoenzymes with similar structures. *Mol Biol Evol* 35:2597–2606.

Sharir-Ivry A, Xia Y, 2019a. Non-catalytic Binding Sites Induce Weaker Long-Range Evolutionary Rate Gradients than Catalytic Sites in Enzymes. *J Mol Biol* 431:3860–3870.

Sharir-Ivry A, Xia Y, 2019b. Using Pseudoenzymes to Probe Evolutionary Design Principles of Enzymes. *Evol Bioinforma* 15:117693431985593.

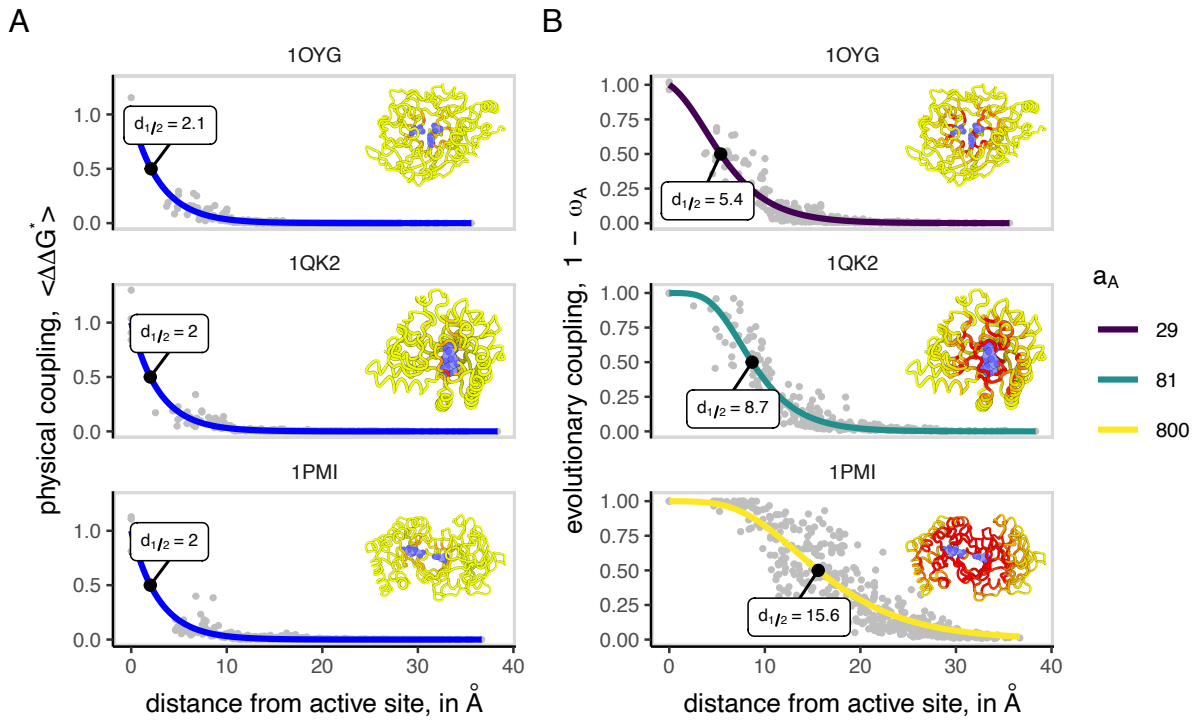

FIG. 1. **Coupling between the active site and other residues for three illustrative examples.** The three cases shown are the enzymes with pdb ids 1OYG, 1QK2, and 1PMI. **A**: physical coupling, as measured by the change in activation free energy that results from mutations; each point corresponds to a protein site; a site's $< \Delta\Delta G^* >$ is the average over mutations; the smooth line is an exponential fit $< \Delta\Delta G^* >= b_0 e^{-b_1 d}$, where d is the distance from the closest active-site residue; the point for which coupling is 1/2 of its maximum is displayed in black; the insets show the 3D protein structures coloured from yellow to red according to increasing physical coupling. **B**: evolutionary coupling, as measured by the average proportion of mutations rejected due to their effect on activation energy, $1-\omega_{\mathrm{A}}$; each point corresponds to a protein site; the smooth line is a function $1-\omega_{\mathrm{A}} = 1 - e^{-b_0 e^{-b_1 d}}$, fit to the points; these lines are coloured according to the functional selection pressure parameter a_{A}; the point for which coupling is 1/2 of the maximum is displayed in black; the insets show the 3D protein structures coloured from yellow to red according to increasing evolutionary coupling. For the sake of comparison, couplings are scaled so that the smooth fits are 1 at $d = 0$.

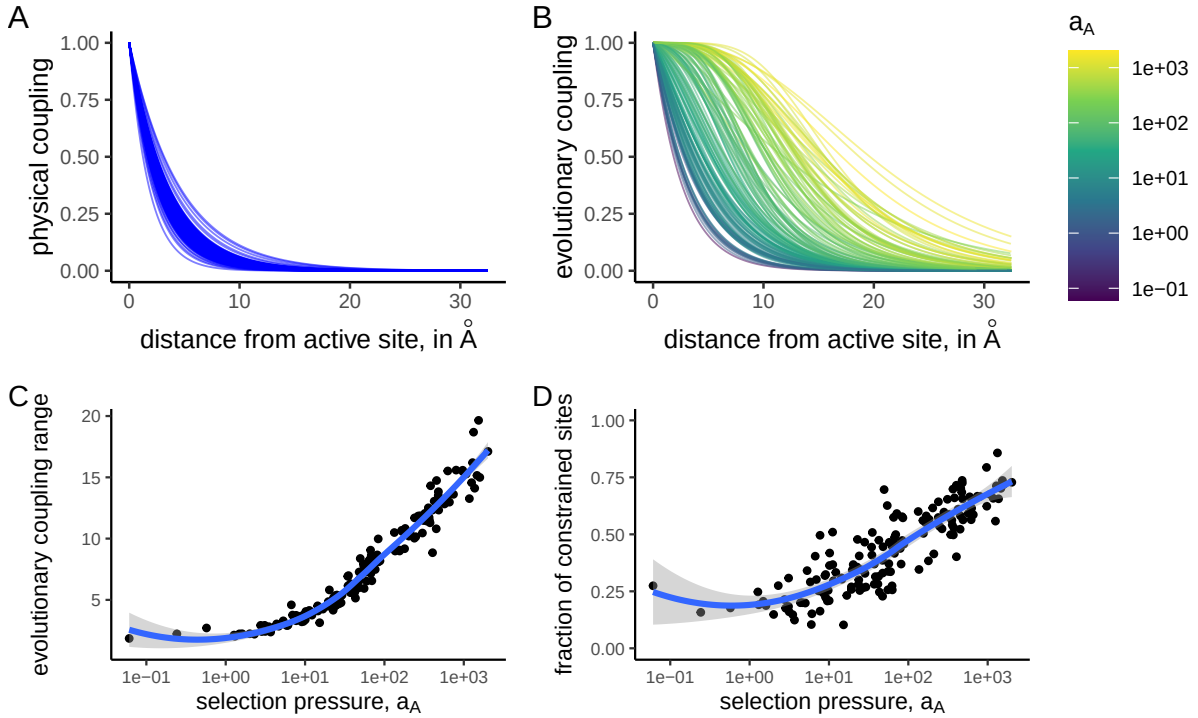

FIG. 2. **Evolutionary coupling range increases with selection pressure on activity.** The dataset consists of 157 monomeric enzymes of diverse sizes, structures, and functions. **A**: physical coupling, as measured by the change in activation free energy that results from mutations; each line is the smooth fit $< \Delta\Delta G^* >= b_0 e^{-b_1 d}$ for one protein of the dataset, where d is the distance from the closest active-site residue. **B**: evolutionary coupling, as measured by $1 - \omega_{\mathrm{A}}$, the average proportion of mutations rejected due to their effect on activation energy; each line is the smooth fit $1 - \omega_{\mathrm{A}} = 1 - e^{-b_0 e^{-b_1 d}}$ for one protein of the dataset. Lines are coloured according to the value of the parameter a_A, that measures functional selection pressure. Couplings are scaled so that all smooth fits are 1 at $d = 0$. **C**: range of evolutionary coupling, measured by the distance at which $1 - \omega_{\mathrm{A}}$ becomes half of its maximum, vs. functional selection pressure, measured by model parameter a_A. **D**: fraction of functionally constrained sites vs. selection pressure. The number of activity-constrained residues was calculated using $n_A = e^{-\sum_r p_r \ln p_r}$, where r denotes residue and $p_r = (1 - \omega_{\mathrm{A}}^r)/\sum_r (1 - \omega_{\mathrm{A}}^r)$. So defined, n_A, which varies between 1 and the total number of sites, measures how distributed over sites $1 - \omega_{\mathrm{A}}$ is.

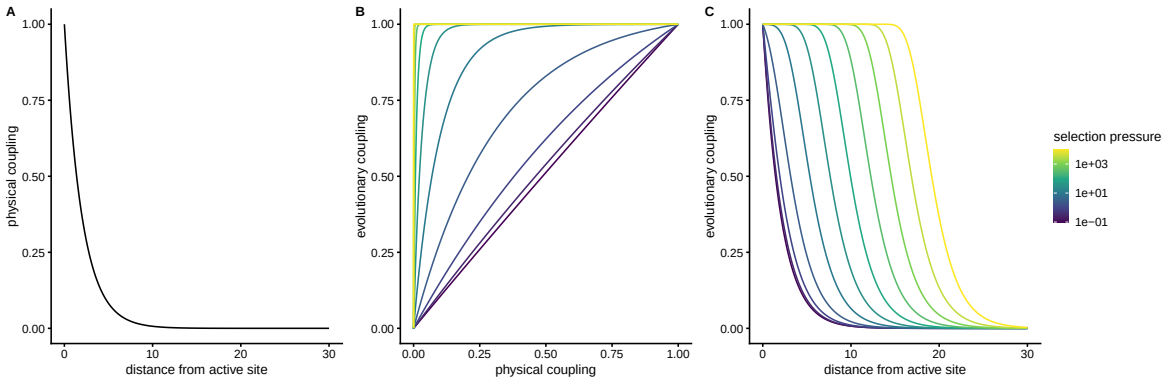

FIG. 3. **Selection pressure controls how physical coupling is mapped into evolutionary coupling.** **A**: typical exponentially decreasing physical coupling. **B**: evolutionary coupling $(1-\omega_{\mathrm{A}})$ vs. physical coupling for increasing functional selection pressure: equation (4). **C**: distance-dependence of evolutionary coupling, which is the composition of panels **A** and **B**. Note that the same short-range physical coupling (panel **A**) is mapped into evolutionary couplings with increasingly longer range as functional selection pressure increases (panel **C**).