Manuscript submitted to **Biophysical** *Journal*
Letters

Evolutionary coupling range varies widely among enzymes depending on selection pressure

Julian Echave[1,*]

[1]Instituto de Ciencias Físicas, Escuela de Ciencia y Tecnología, Universidad Nacional de San Martín, Martín de Irigoyen 3100, 1650 San Martín, Buenos Aires, Argentina
[*]Correspondence: jechave@unsam.edu.ar

ABSTRACT Recent studies proposed that enzyme active sites induce evolutionary constraints at long distances. The physical origin of such long-range evolutionary coupling is unknown. Here, I use a recent biophysical model of evolution to study the relationship between physical and evolutionary couplings on a diverse data set of monomeric enzymes. I show that evolutionary coupling is not universally long-range. Rather, range varies widely among enzymes, from 2Å to 20Å. Furthermore, the evolutionary coupling range of an enzyme does not inform on the underlying physical coupling, which is short-range for all enzymes. Rather, evolutionary coupling range is determined by functional selection pressure.

SIGNIFICANCE Until recently, only residues near enzyme active sites were thought to be evolutionarily constrained. However, recent studies proposed that active sites induce long-range evolutionary constraints. This seems to conflict with the common finding that physical couplings in proteins are short-range. This raises the question of how short-range physical couplings may cause long-range evolutionary couplings. Here, I show that the function that maps physical coupling into evolutionary coupling depends on functional selection pressure. Under weak selection, both couplings are similarly short-range; under strong selection, short-range physical coupling is non-linearly turned into long-range evolutionary coupling. Thus, due to a huge variation of selection pressure, evolutionary coupling range varies widely among enzymes, from very short (2 Å) to very long (20 Å).

As enzymes evolve, different sites evolve at different rates. The main reason for such variation of evolutionary rate among sites within proteins is selection for stability (1–4). Until recently, activity constraints were thought to affect just the few residues directly involved in catalysis and their immediate neighbours (1, 5). However, recent studies have reported that active sites influence evolutionary rates at long distances, slowing down the evolution of residues as distant as 30 Å (6–10). It seems reasonable to assume that such long-range evolutionary coupling could result from long-range physical couplings, such as those involved in allosteric mutations (11, 12). This would align with the notion that enzymes are evolutionarily designed to optimize long-range coupling (13–15). However, except perhaps for a few allosteric residues, I would expect physical coupling to be a be the typical short-range exponentially decreasing function of distance characteristic of indirect through-the-contact-network couplings (14, 16–20). If this is the case, it would leave long-range evolutionary coupling begging explanation.

The aim of this work is to verify whether physical coupling is short-range and, in that case, to study how such a short-range physical coupling may give rise to a long-range evolutionary coupling. To this end, I use the stability-activity model of enzyme evolution, MSA (21). I previously showed that this model reproduces quantitatively the observed slow increase of evolutionary rate with distance from the active site that led to the proposal of long-range evolutionary couplings (7). This makes MSA suitable for exploring the physical underpinnings of long-range evolutionary coupling.

Before describing the model, I start with some definitions. The evolutionary rate, K, is the number of amino acid substitutions per unit time along an evolutionary trajectory; K_0 is the rate for the case in which all mutations are neutral; $\omega = \frac{K}{K_0}$ is the rate relative to the neutral rate; and $1 - \omega = \frac{K_0 - K}{K_0}$ is the relative slowdown with respect to the neutral evolution case. As selection pressure increases, evolution slows down: K and ω decrease, and $1 - \omega$, the relative slowdown, increases. K, ω, and $1 - \omega$ contain exactly the same information. Therefore, in what follows, to measure the effect of selection on evolution, I will mostly use $1 - \omega$, the evolutionary slowdown.

With the help of the MSA model, I define physical and evolutionary coupling measures, and derive the formula that relates them. (The MSA model is described in detail in the Supporting Material document (SM), in SM Section 1 and SM Section 3,

and in (21).) MSA predicts that the evolutionary slowdown of an enzyme residue r due to selection on stability and activity is given by (Eq. S19):

$$1 - \omega(r) = 1 - \langle \min(1, e^{-a_S \Delta\Delta G(r)}) \min(1, e^{-a_A \Delta\Delta G^*(r)}) \rangle \quad (1)$$

where $\Delta\Delta G(r)$ and $\Delta\Delta G^*(r)$ are, respectively, mutational changes of folding free energy and activation free energy; a_S and a_A are positive parameters that represent selection pressure on stability and activity, respectively; and $\langle \cdots \rangle$ stands for averaging over mutations. Eq. 1 relates the evolutionary slowdown of residue r to the effects of mutating this residue on stability and activity. Since here I am only interested on selection on activity, I consider a hypothetical scenario in which selection on stability is turned off. Replacing $a_S = 0$ into Eq. 1 it follows that the evolutionary slowdown of site r due to selection on activity is given by (Eq. S20):

$$1 - \omega_A(r) =< \min(1, e^{-a_A \Delta\Delta G^*(r)}) > \quad (2)$$

The mutational activation free-energy change $\Delta\Delta G^*(r)$ is due to the distortion of the active-site caused by mutating residue r (21) (Eq. S45, SM Section 1.1, and SM Section 3.3). Therefore, $\Delta\Delta G^*(r)$ represents the *physical coupling* between the enzyme's active site and residue r. This causes the slowdown $1 - \omega_A(r)$, which, therefore, can be considered a measure of *evolutionary coupling*. Thus, Eq. 2 governs how evolutionary coupling ($1 - \omega_A$) depends on physical coupling ($\Delta\Delta G^*$) and functional selection pressure (a_A). (For notational simplicity, I will drop the explicit reference to residue r whenever possible.)

I studied the relationship between physical and evolutionary couplings on a data set of monomeric enzymes used previously (21) (SM Section 1.2). Briefly, for each protein, I calculated $\Delta\Delta G$ and $\Delta\Delta G^*$ for all mutations at all residues using the Linearly Forced Elastic Network Model (LFENM). Then, I obtained the model parameters a_S and a_A by fitting MSA predictions to empirical rates. Finally, I calculated the residue-dependent evolutionary couplings, $1 - \omega_A$, and physical couplings, $\langle \Delta\Delta G^* \rangle$. ($\langle \Delta\Delta G^*(r) \rangle$ is the average of $\Delta\Delta G^*(r)$ over mutations at r.) For details of the calculation, see SM Section 1.1. I consider MSA to be validated in a previous study (21). However, for completeness, in SM Section 2 I show the excellent agreement between MSA predictions and empirical rates (SM Section 2.1) and I discuss the adequacy of using LFENM to calculate $\Delta\Delta G$ and $\Delta\Delta G^*$ (SM Section 2.2). In what follows, I focus on the study of physical and evolutionary couplings.

For clarity, I start by considering three illustrative examples (Figure 1). I measure coupling range using $d_{1/2}$, the distance at which coupling is half the maximum. Physical coupling ($\langle \Delta\Delta G^* \rangle$) is similar for the three cases, it decreases exponentially with increasing distance, and it is very short range ($d_{1/2}^{\text{phys}}$ is 2.1 Å, 2.0 Å, and 2.0 Å for 1OYG, 1QK2, and 1PMI, respectively, Figure 1A). In contrast, evolutionary coupling ($1 - \omega_A$) varies among the examples, it is not an exponential but a sigmoid, and its range varies widely ($d_{1/2}^{\text{evo}}$ is 5.4 Å, 8.7 Å, and 15.6 Å for 1OYG, 1QK2, and 1PMI, respectively; Figure 1B). It will shown below that this variation of evolutionary coupling range is due to the variation of functional selection pressure acting on these three enzymes (a_A, is 29.2, 80.2, and 800 for 1OYG, 1QK2, and 1PMI, respectively).

Figure 2 shows the distance-dependence of physical and evolutionary couplings for all the enzymes studied. Physical coupling is a short-range exponential decline with distance, very similar for all cases (Figure 2A). In contrast, the distance-dependence of evolutionary coupling varies among proteins, from a short range exponential decline to a long-range sigmoidal decline (Figure 2B). Because physical coupling is very similar for all enzymes, from Eq. 2 it follows that the variation among enzymes of evolutionary coupling must be determined by the selection parameter a_A. This is confirmed in Figure 2C, that shows that evolutionary coupling range is independent of physical coupling range, and Figure 2D, that shows that the variation of evolutionary coupling range is almost completely explained by the selection pressure parameter a_A.

Thus, according to Figure 2, evolutionary coupling range varies widely among enzymes, from 1.9 Å to 19.7 Å, as a result of the variation of parameter a_A over more than 4 orders of magnitude, from 6×10^{-2} to 2×10^3. This huge variation of a_A represents a variation of the functional selection pressure under which enzymes evolve (SM Section 2.3). This variation can be explained by Eq. 2, that non-linearly maps a short-range exponentially decreasing physical coupling into a sigmoidally decreasing evolutionary coupling whose range increases with selection pressure (SM Section 2.4). In summary, the fundamental finding of this work is that long-range evolutionary coupling is not due to enzymes being particularly well-designed for long-range physical coupling, but it is a consequence of a non-linear amplification of physical coupling under strong functional selection pressure.

The previous findings provide a mechanism that explains how functional constraints slow down enzyme evolution. *A priori*, the decrease of the rate of evolution with increasing selection pressure could be uniformly distributed among all residues. However, the present findings imply a different picture: increasing functional selection pressure increases the range of influence of the enzyme's active site on other residues (Figure 2 D); as this range increases, more sites become functionally constrained (Figure 3A) and the active site becomes more tightly coupled to the rest of the protein (Figure 3 B); as a result, the enzyme's evolution slows down (Figure 3 C). In this work, I have derived a_A from patterns of rate variation among sites within proteins. However, the model predicts that a_A should also influence rate variation among proteins, which would connect rate variation within proteins with rate variation among proteins. Further work is needed to test this important prediction.

To finish, I mention another two research directions suggested by this work. First, for enzymes, functional selection pressure depends on metabolic role (22–25). Specifically, the main functional constraints are enzyme-specific metabolic flow and enzyme essentiality (23). Therefore, these properties should correlate with parameter a_A of the present work, and, as a consequence, metabolic role should affect evolutionary coupling range. This prediction should be verified. Second, the present findings indicate the intriguing possibility of manipulating evolutionary coupling range by adjusting selection pressure, which could be explored using enzyme evolution experiments.

SUPPORTING CITATIONS

References (26–37)

SUPPORTING MATERIAL

See supporting_material.pdf.

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

REFERENCES

1. Echave, J., S. J. Spielman, and C. O. Wilke, 2016. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* 17:109–121.

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

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

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

5. Bartlett, G. J., C. T. Porter, N. Borkakoti, and J. M. Thornton, 2002. Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* 324:105–121.

6. Dean, A. M., C. Neuhauser, E. Grenier, and G. B. Golding, 2002. The pattern of amino acid replacements in alpha/beta-barrels. *Mol. Biol. Evol.* 19:1846–1864.

7. Jack, B. R., A. G. Meyer, J. Echave, and C. O. Wilke, 2016. Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes. *PLOS Biol.* 14:e1002452.

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

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

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

11. Perica, T., Y. Kondo, S. P. Tiwari, S. H. McLaughlin, K. R. Kemplen, X. W. Zhang, A. Steward, N. Reuter, J. Clarke, and S. A. Teichmann, 2014. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science* 346:12.

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

13. Flechsig, H., 2017. Design of Elastic Networks with Evolutionary Optimized Long-Range Communication as Mechanical Models of Allosteric Proteins. *Biophys. J.* 113:558–571.

14. Naganathan, A. N., 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.

15. Campitelli, P., T. Modi, S. Kumar, and S. Banu Ozkan, 2020. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* 49:267–288.

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

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

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

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

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

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

22. Boucher, J. I., D. N. A. Bolon, and D. S. Tawfik, 2016. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. *Protein Sci.* 25:1219–1226.

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

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

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

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

27. Echave, J., and F. M. Fernández, 2010. A perturbative view of protein structural variation. *Proteins Struct. Funct. Bioinforma.* 78:173–180.

28. Ming, D., and M. E. Wall, 2005. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.* 95:198103doi10.

29. Furnham, N., G. L. Holliday, T. A. P. de Beer, J. O. B. Jacobsen, W. R. Pearson, and J. M. Thornton, 2014. The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* D485–D489.

30. Webb, E. C., 1992. Enzyme Nomenclature. Wiley, San Diego.

31. McCandlish, D. M., and A. Stoltzfus, 2014. Modeling Evolution Using the Probability of Fixation: History and Implications. *Q. Rev. Biol.* 89:225–252.

32. Echave, J., E. L. E. Jackson, and C. O. C. Wilke, 2015. Relationship between protein thermodynamic constraints and variation of evolutionary rates among sites. *Phys. Biol.* 12:025002.

33. Huang, T.-T., M. L. Del Valle Marcos, J.-K. Hwang, and J. Echave, 2014. A mechanistic stress model of protein evolution accounts for site-specific evolutionary rates and their relationship with packing density and flexibility. *BMC Evol. Biol.* 14:78.

34. Marcos, M. L., and J. Echave, 2015. Too Packed to Change: Side-Chain Packing and Site-Specific Substitution Rates in Protein Evolution. *PeerJ* 3:e911.

35. Stein, R. L., 2011. Kinetics of Enzyme Action: Essential Principles for Drug Hunters. Wiley, Hoboken, New Jersey.

36. Schowen, R. L., 1978. Catalytic power and Transition-State stabilization. *In* R. D. Gandour, and R. L. Schowen, editors, Transition States of Biochemical Processes, Springer, Boston, chapter 20, 77–144.

37. Micheletti, C., P. Carloni, and A. Maritan, 2004. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins* 55:635–45.

38. Tomasello, G., I. Armenia, and G. Molla, 2020. The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics* 36:2909–2911.

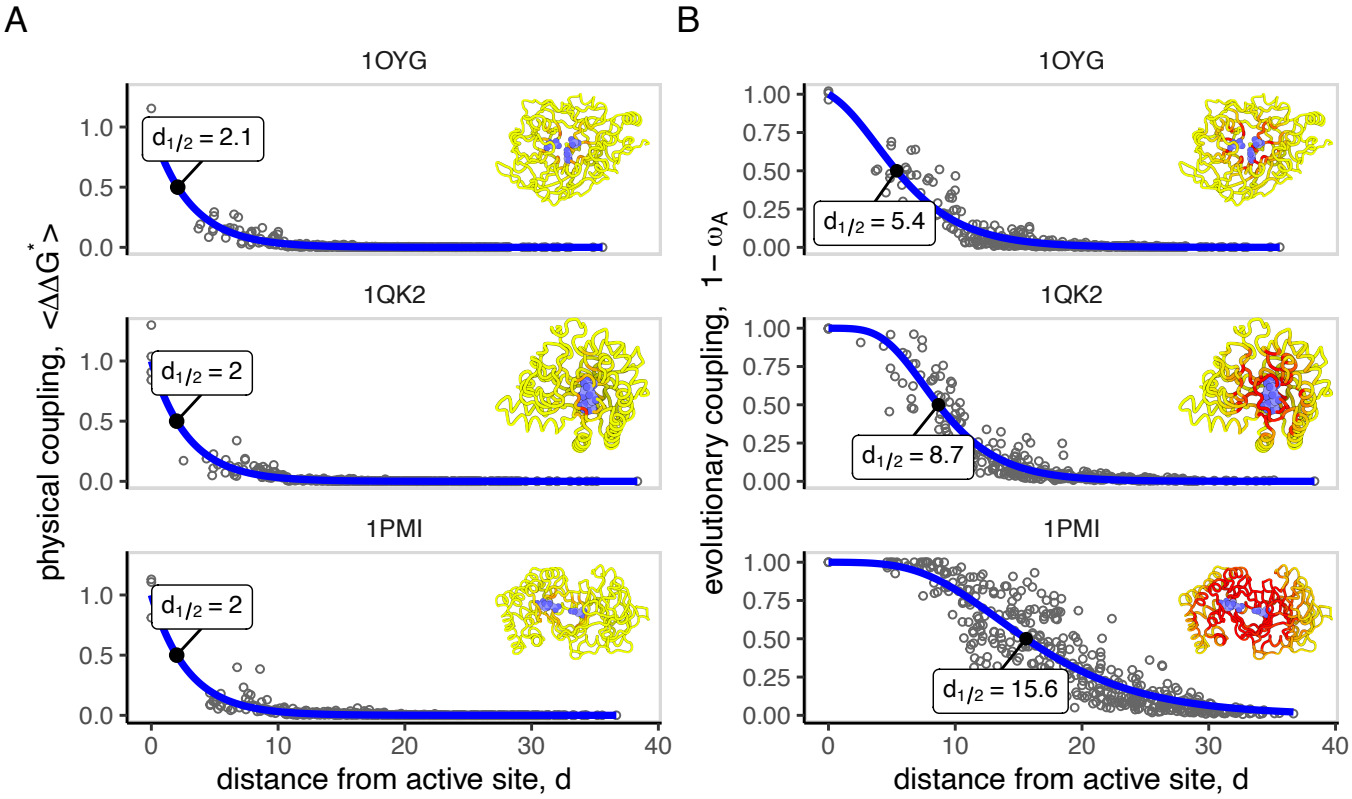

Figure 1: **Coupling between the active site and other residues for three illustrative examples.** The three cases shown are the enzymes with PDB IDs 1OYG, 1QK2, and 1PMI. **A**: physical coupling, as measured by the change in activation free energy that results from mutations; each point corresponds to a protein site; a site's $\langle \Delta\Delta G^* \rangle$ is the average of $\Delta\Delta G^*$ over mutations; the smooth line is an exponential fit $\langle \Delta\Delta G^* \rangle = b_0 e^{-b_1 d}$, where d is the distance from the closest active-site residue, measured in Å; 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 $1 - \omega_A$, the relative slowdown of evolution due to selection on activity; each point corresponds to a protein site; the smooth line is a function $1 - \omega_A = 1 - e^{-c_0 e^{-c_1 d}}$, fit to the points; 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 $d = 0$. 3D images were made with https://3dproteinimaging.com/protein-imager (38).

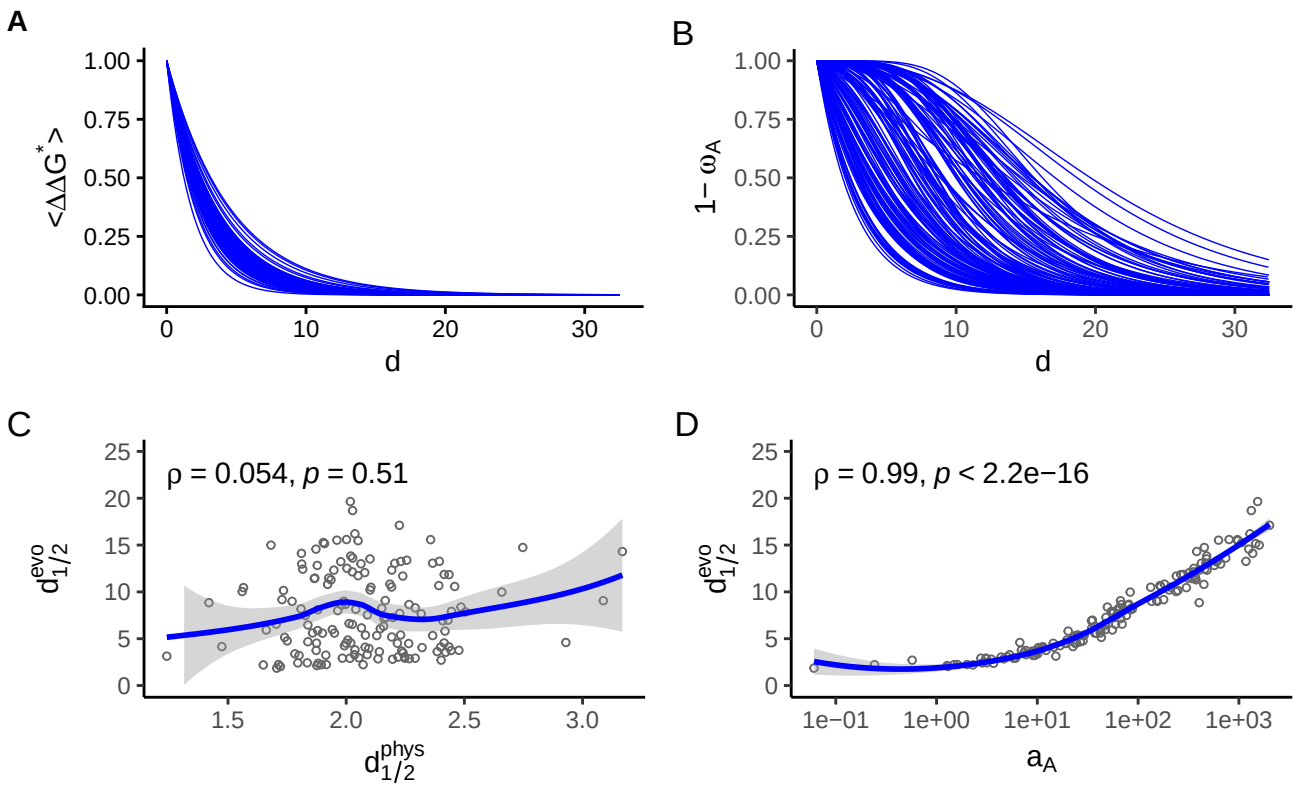

Figure 2: **Evolutionary coupling range increases with selection pressure on activity.** Coupling between the active site and other residues for each protein of a data set of 157 monomeric enzymes of diverse sizes, structures, and functions. **A**: physical coupling, as measured by $\langle \Delta\Delta G^* \rangle$, the activation free energy change averaged over mutations at the mutated site; each line is the smooth fit $\langle \Delta\Delta G^* \rangle = b_0 e^{-b_1 d}$ for one protein of the data set, where d is the distance from the closest active-site residue in Å. **B**: evolutionary coupling, as measured by $1 - \omega_A$, the relative slowdown due to selection on activity; each line is the smooth fit $1 - \omega_A = 1 - e^{-c_0 e^{-c_1 d}}$ for one protein of the data set; Couplings are scaled so that all smooth fits are 1 at $d = 0$. **C**: range of evolutionary coupling, measured by $d_{1/2}^{\text{evo}}$, the distance at which $1 - \omega_A$ becomes half of its maximum, vs. range of physical coupling, measured by $d_{1/2}^{\text{phys}}$, the distance at which $\langle \Delta\Delta G^* \rangle$ becomes half of its maximum. **D**: range of evolutionary coupling, measured by $d_{1/2}^{\text{evo}}$, vs. functional selection pressure, measured by model parameter a_A. In **C** and **D**, each point represents one protein, ρ is Spearman's correlation coefficient, p its p-value, and the lines are local regression fits.

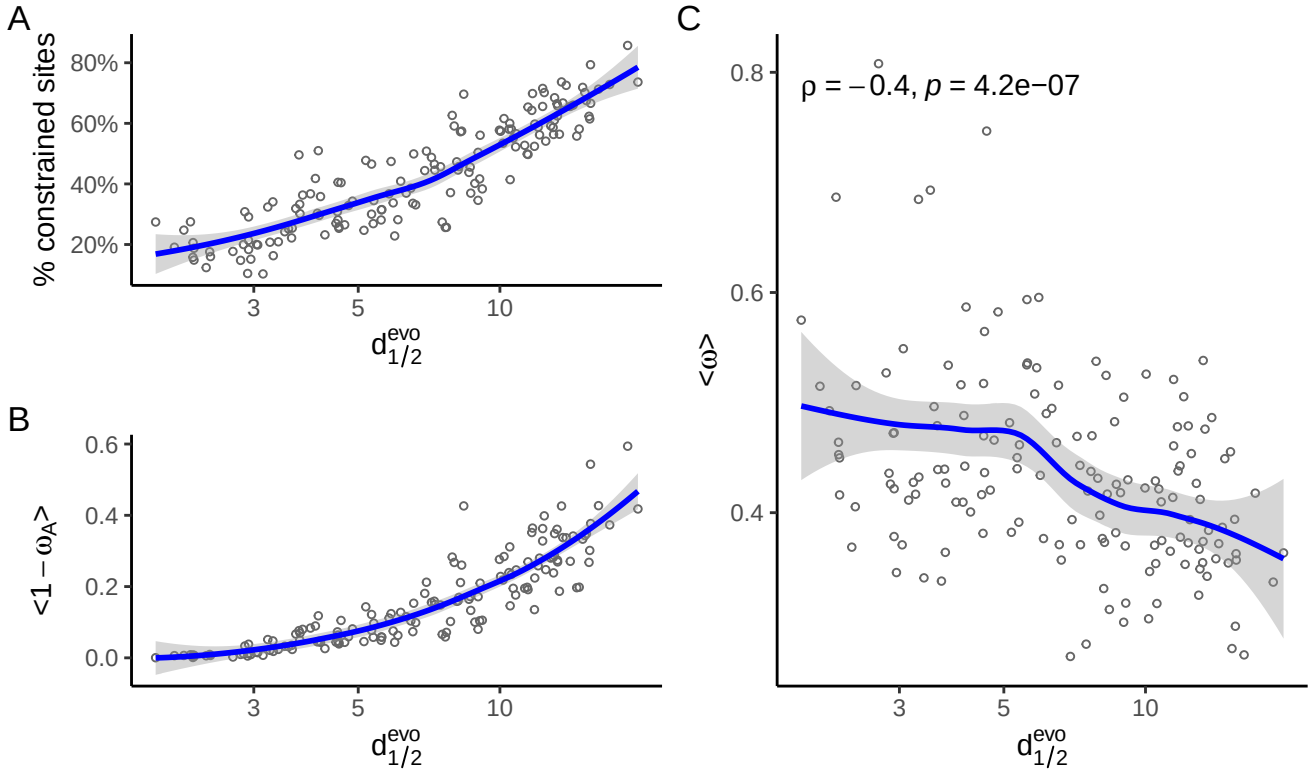

Figure 3: **Increasing evolutionary coupling range slows down enzyme evolution.** Evolutionary coupling range is measured by $d_{1/2}^{\text{evo}}$, the distance at which evolutionary coupling is half its maximum. **A**: Increase of the fraction of functionally constrained sites; the number of activity-constrained residues was calculated using $n_A = e^{-\sum_r p_r \ln p_r}$, where r denotes residue and $p_r = (1-\omega_A^r)/\sum_r (1-\omega_A^r)$. So defined, n_A, which varies between 1 and the total number of sites, measures how distributed over sites $1-\omega_A$ is. **B**: Increase of overall enzyme coupling, measured by $1-\omega_A$ averaged over residues. **C**: Predicted decrease of the protein rate of evolution, measured by the relative rate ω averaged over residues; ρ is Spearman's correlation coefficient, p its p-value. In all panels, each point represents one protein and lines are local regression fits.