

# Using Pseudoenzymes to Probe Evolutionary Design Principles of Enzymes

Avital Sharir-Ivry  and Yu Xia

Department of Bioengineering, McGill University, Montreal, QC, Canada.

Evolutionary Bioinformatics  
Volume 15: 1–4  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1176934319855937



**ABSTRACT:** Enzymes are governed by unique evolutionary design principles as their catalytic sites were shown to induce long-range evolutionary conservation gradients. We have recently used a comparative bioinformatics approach to disentangle structural determinants from other possible determinants of the evolutionary conservation gradients. The approach is based on comparing the evolutionary patterns of enzymes to those of pseudoenzymes with the same tertiary structure where the catalytic functionality is turned off. This approach provides a way to evaluate several hypotheses regarding the origin of the observed evolutionary conservation gradient in enzymes. The conclusions from such comparative analyses are important for a better understanding of the unique evolutionary design principles of enzymes, which can in turn potentially guide the design of new and improved enzymes.

**KEYWORDS:** enzyme evolution, pseudoenzymes, evolutionary conservation gradient, ligand binding, allosteric sites

**RECEIVED:** May 6, 2019. **ACCEPTED:** May 15, 2019.

**TYPE:** Commentary

**FUNDING:** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Natural Sciences and Engineering Research Council of Canada grants RGPIN-2019-05952 and RGPAS-2019-00012, Canada Foundation for Innovation grants JELF-33732 and IF-33122, and Canada Research Chairs program.

**DECLARATION OF CONFLICTING INTERESTS:** The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**CORRESPONDING AUTHOR:** Yu Xia, Department of Bioengineering, McGill University, Montreal, Quebec H3A 0E9, Canada. Email: brandon.xia@mcgill.ca

**COMMENT ON:** Sharir-Ivry A, Xia Y. Nature of long-range evolutionary constraint in enzymes: insights from comparison to pseudoenzymes with similar structures. *Mol Biol Evol.* 2018;35:2597-2606. doi:10.1093/molbev/msy177. PubMed PMID: 30202983. Website: <https://www.ncbi.nlm.nih.gov/pubmed/30202983>.

Enzymes enable life by allowing important chemical reactions in the cell to occur in reasonable timescales. Enzymatic function is very complex as it involves binding substrates as well as catalyzing their transformation into products by lowering the energy barrier of the chemical reaction. Currently, much effort is dedicated to the design of new enzymes catalyzing unnatural reactions or the improvement of functionality in existing enzymes. A notable example of such enzyme design methods is the directed evolution method for which the recent Nobel Prize in Chemistry was given, where evolutionary principles are used in the laboratory to drive the realization of a desired functionality in enzymes.<sup>1</sup> In order to efficiently design new and improved enzymes, we need to better understand evolutionary design principles that govern the formation of conservation and variation patterns observed in existing enzymes.

Recently, a nearly linear long-range evolutionary conservation gradient was shown to extend from the main functional site of enzymes, the catalytic site.<sup>2</sup> The gradient is such that the closer a residue is to the catalytic site, the stronger the evolutionary selective pressure it experiences. The conservation gradient extends up to ~28Å in distance and affects ~80% of the protein residues. This gradient was shown to be significantly stronger than that from protein-protein interaction sites. This observation raises the following fundamental question: what are the key physical, structural, and functional factors that drive the formation of this unique long-range evolutionary conservation gradient in enzymes?

There are three main hypotheses regarding the origin of the evolutionary conservation gradient from catalytic sites in enzymes. This phenomenon could be driven by factors related

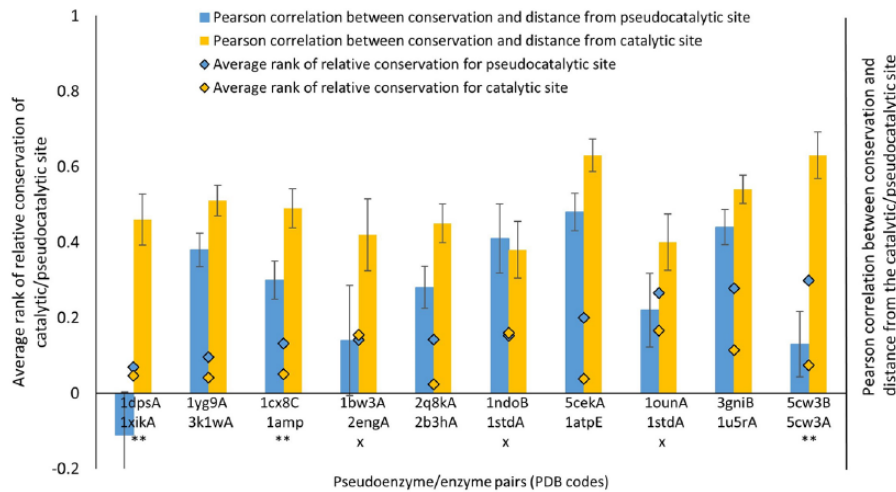
to protein tertiary structure, by the percolation of evolutionary selective pressure from the catalytic site, or by other factors related to enzyme function. These three hypotheses are not necessarily mutually exclusive, and their combination could potentially drive the formation of the long-range evolutionary conservation gradient. It is therefore difficult to test each hypothesis separately and pinpoint the origin of this phenomenon.

Recently, we introduced a new approach that uses pseudoenzymes to probe evolutionary design principles of enzymes. Pseudoenzymes and their enzyme counterparts share highly similar tertiary structures; however, unlike enzymes, pseudoenzymes do not exhibit catalytic functions. Hence, by directly comparing pseudoenzymes and their enzyme counterparts, we can assess how turning catalytic function on and off specifically affects the evolutionary properties of these proteins. This approach enables us to disentangle different structural and functional contributions to the long-range evolutionary conservation gradient in enzymes. Below we describe how our comparative approach can shed light on the validity of each of the three hypotheses mentioned above.

## Factors Related to Tertiary Structure

Several residue-level structural determinants have been found to correlate significantly with residue evolutionary rate. The most dominant structural determinant of residue evolutionary rate is the residue's degree of solvent exposure or degree of packing.<sup>3-5</sup> The more exposed or less packed a site is, the less it is subject to selective pressure, and the faster it evolves. Similar trends were observed independent of the hydrophobicity of the residue<sup>3</sup> or





**Figure 1.** Pseudoenzymes with high average rank of relative conservation of pseudocatalytic site within the protein.

Shown are average rank of relative conservation of catalytic/pseudocatalytic site within the protein (yellow and blue diamonds respectively, where 0 is the highest ranking and 1 is the lowest ranking) as well as the Pearson correlation between conservation and distance from that site (yellow and blue bars respectively). None of the differences between average rank of pseudocatalytic and catalytic site pairs are statistically significant ( $P > .05$ ). All differences between Pearson correlations are statistically significant at the 0.05 level unless otherwise mentioned (\*\*represents statistical significance at the 0.01 level, x represents no statistical significance).

the solvent environment,<sup>6</sup> suggesting that the instantaneous rate of residue evolution is primarily constrained by the tight packing in the protein interior. Other determinants were shown to have an effect as well; however, they are usually not mutually exclusive from solvent exposure and packing. These structural determinants represent the necessity to maintain the stability of the native structure and its proper folding. Hence, sites which are more important for maintaining the stability of the native structure will usually experience stronger selective pressure.<sup>7</sup>

Enzymes are known to utilize protein folds that are generally different from those of non-enzymes.<sup>8,9</sup> Since structural determinants play significant roles in shaping sequence conservation patterns in proteins, it is reasonable to hypothesize that structural determinants are also the main driving force behind the long-range evolutionary conservation gradients from catalytic sites in enzymes. According to this hypothesis, catalytic sites are preferentially located in an optimal position within the protein such that structural determinants change gradually with distance from the catalytic site. While it was shown that solvent exposure cannot account for the observed conservation gradient in enzymes,<sup>2</sup> it is possible that other known or yet to be discovered structural determinants are responsible for this phenomenon.

We used a unique approach to test the validity of this hypothesis, by directly comparing the evolutionary conservation patterns in enzymes and pseudoenzymes sharing nearly identical tertiary structures. Unlike their enzyme counterparts, the catalytic function is turned off in pseudoenzymes, due to either missing catalytic residues or blocking of the entrance to the catalytic site.<sup>10,11</sup> We have shown that despite sharing nearly identical tertiary structures, the conservation gradient from the pseudocatalytic site in pseudoenzymes is significantly weaker than the conservation gradient from the catalytic site in their enzyme counterparts. Hence, backbone-based tertiary structural

determinants (such as solvent exposure, residue packing, among many others) cannot be the main driving force for the strong evolutionary conservation gradient from catalytic sites in enzymes.

### Percolation of Selective Pressure From Catalytic Site

This hypothesis is based on the assumption that if a site is under strong selective pressure, then its neighboring residues in the tertiary structure also tend to be more conserved than expected. As a result, the strong selective pressure exerted on catalytic sites percolates into their surrounding residues via residue-residue contacts to create the observed strong evolutionary conservation gradient in enzymes. According to this hypothesis, the weaker conservation gradients from pseudocatalytic sites in pseudoenzymes is primarily due to these pseudocatalytic sites being under weaker selective pressure.

We can obtain clues as to the validity of the percolation hypothesis by comparing the conservation gradients in pseudoenzymes and their enzyme counterparts. We examined the top ten pseudoenzymes in our dataset<sup>12</sup> for which the average rank of the sequence conservation of the pseudocatalytic site within the protein is highest (Figure 1, where 0 represents the highest rank of conservation within the protein and 1 represents the lowest rank of conservation within the protein). For eight of the ten cases in Figure 1, the average rank of relative conservation of the catalytic site within the enzyme protein is higher than that of the counterpart pseudocatalytic site within the pseudoenzyme protein. Moreover, in the two cases in which the average rank of the catalytic site is lower than that of the counterpart pseudocatalytic site (1bw3A-2engA and 1ndoB-1stdA), the conservation gradient from the pseudocatalytic site is actually not significantly different from the conservation gradient from the catalytic site. These observations support the possible involvement of percolation in creating the conservation

**Table 1.** Pseudoenzymes with larger conformational changes than the counterpart enzyme induce weaker conservation gradients. Maximal RMSD for identical chains in the PDB (according to PDBFlex) for enzyme and pseudoenzyme pairs where the maximal RMSD is higher in the pseudoenzyme, as well as the respective Pearson correlations between conservation and distance from the catalytic/pseudocatalytic site.

ENZYME (PDB CODE)	PSEUDOENZYME (PDB CODE)	MAXIMAL RMSD (ENZYME)	MAXIMAL RMSD (PSEUDOENZYME)	PEARSON CORRELATION (ENZYME)	PEARSON CORRELATION (PSEUDOENZYME)	P-VALUE FOR THE DIFFERENCE BETWEEN CORRELATION
1std	1oun	1.155	2.432	0.4	0.2	0.0445
1amp	1cx8	0.857	1.753	0.49	0.3	0.0045
5cw3A	5cw3B	1.252	1.306	0.65	0.1	<10 <sup>-4</sup>
4k8v	4nxt	2.521	2.735	0.29	0.02	0.0011
2vk5	5tih	0.517	1.673	0.35	0.01	0.0021
1pmaB	1pmaA	0.624	1.051	0.55	0.34	0.0068
1b70A	1b70B	1.108	1.129	0.45	0.23	0.0096

RMSD: simply root-mean-square deviation.

gradient, such that strongly conserved sites induce stronger conservation gradients. On the other hand, we can see the case of 1dpsA-1xikA, where the relative conservation of the catalytic and pseudocatalytic sites are similarly high; however, the conservation gradient from the catalytic site is significantly stronger than that from the pseudocatalytic site. Hence, more work is required to fully elucidate the role that percolation plays in shaping the observed conservation gradient from catalytic sites.

### Functional Determinants

Other potential driving forces for the evolutionary conservation gradient from catalytic sites in enzymes involve different aspects of the complex enzymatic function including substrate binding, allosteric function, and catalysis.

To test the hypothesis that substrate binding functionality of the catalytic site is the main determinant of the induced conservation gradient, the approach of comparing enzymes and pseudoenzymes that share nearly identical tertiary structure was utilized again.<sup>12</sup> This time we focused on the subset of pseudoenzymes in which the ligand binding function is retained in the pseudocatalytic site. These pseudoenzymes preserve the tertiary structure as well as ligand binding functionality of their enzyme counterparts, where only the catalytic activity is turned off. Ligand-binding pseudocatalytic sites were shown to induce significantly weaker conservation gradients compared with the counterpart catalytic sites. We thus conclude that the presence or absence of binding function alone cannot be the main determinant of the strong conservation gradient from catalytic sites in enzymes. At the same time, it remains to be determined whether or not possible difference in binding affinity plays a role.

Another functional attribute of enzyme function that could potentially be a main determinant of the conservation gradient in enzymes is its allosteric function. Here, a ligand binding

event in a binding site distant from the catalytic site (called allosteric site) shifts the catalytic site conformation into its active conformation. None of the enzymes in our dataset are known to have allosteric function and yet they are all able to induce strong conservation gradient from their catalytic site, which implies that these gradients are not dependent on allosteric function. It is possible, however, that these enzymes do have allosteric function that is not known or is relatively weak. To further obtain clues as to the validity of the hypothesis using the enzyme/pseudoenzyme pairs in our dataset, we compared their conformational flexibility using PDBFlex.<sup>13</sup> For each protein, we looked at the maximal RMSD between its identical chains (sequence identity > 95%) in the PDB. We look at pairs for which the pseudoenzyme is more flexible than the enzyme (Table 1), which could potentially point to possible allosteric functionality in the pseudoenzyme. In all these cases, the conservation gradient is significantly stronger in the more rigid enzyme, implying that allosteric function is unlikely to be the main determinant of the conservation gradient from catalytic sites. That being said, a systematic analysis is required to determine for certain whether or not allosteric function is a determinant of the strong evolutionary conservation gradients from catalytic sites.

Finally, catalysis is the function in which the enzyme preferentially stabilizes the transition state of the chemical reaction over the reactant state. The hypothesis that this catalytic function is the driving force behind the formation of the conservation gradient from catalytic sites aligns with the “chemistry driven” view of enzyme evolution in which homologous enzymes that evolved from promiscuous ancestors tend to share similar reaction mechanistic strategies.<sup>14</sup> Further work is required to test this hypothesis fully. Recently, a mechanistic model that incorporates enzyme catalytic function in addition to protein folding energetics was shown to better predict

conservation gradients from catalytic sites,<sup>15</sup> pointing toward the validity of this hypothesis.

In summary, we have introduced a unique methodology to investigate how different structural and functional properties affect protein evolution at the residue level. This method is based on comparing two groups of proteins where certain biophysical properties of interest are “turned-on” for one group of proteins and “turned-off” for the other group of proteins, while keeping other biophysical properties nearly identical. The comparison of evolutionary patterns between these two groups of proteins is shown to be a powerful tool to pinpoint the causal structural and functional determinants of protein evolution at the residue level.

### Author Contributions

All authors contributed to the discussion and manuscript writing.

### ORCID iD

Avital Sharir-Ivry  <https://orcid.org/0000-0002-3566-2472>

### REFERENCES

1. Arnold FH. Directed evolution: bringing new chemistry to life. *Angew Chem Int Ed Engl.* 2018;57:4143-4148.
2. Jack BR, Meyer AG, Echave J, Wilke CO. Functional sites induce long-range evolutionary constraints in enzymes. *PLoS Biol.* 2016;14:e1002452.
3. Franzosa EA, Xia Y. Structural determinants of protein evolution are context-sensitive at the residue level. *Mol Biol Evol.* 2009;26:2387-2395.
4. Marcos ML, Echave J. Too packed to change: side-chain packing and site-specific substitution rates in protein evolution. *PeerJ.* 2015;3:e911.
5. Sharir-Ivry A, Xia Y. The impact of native state switching on protein sequence evolution. *Mol Biol Evol.* 2017;34:1378-1390.
6. Franzosa EA, Xue R, Xia Y. Quantitative residue-level structure-evolution relationships in the yeast membrane proteome. *Genome Biol Evol.* 2013;5:734-744.
7. Echave J, Spielman SJ, Wilke CO. Causes of evolutionary rate variation among protein sites. *Nat Rev Genet.* 2016;17:109-121.
8. Hegyi H, Gerstein M. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J Mol Biol.* 1999;288:147-164.
9. Martin AC, Orengo CA, Hutchinson EG, et al. Protein folds and functions. *Structure.* 1998;6:875-884.
10. Todd AE, Orengo CA, Thornton JM. Sequence and structural differences between enzyme and nonenzyme homologs. *Structure.* 2002;10:1435-1451.
11. Eyers PA, Murphy JM. The evolving world of pseudoenzymes: proteins, prejudice and zombies. *BMC Biol.* 2016;14:98.
12. Sharir-Ivry A, Xia Y. Nature of long-range evolutionary constraint in enzymes: insights from comparison to pseudoenzymes with similar structures. *Mol Biol Evol.* 2018;35:2597-2606.
13. Hrabe T, Li Z, Sedova M, Rotkiewicz P, Jaroszewski L, Godzik A. PDBFlex: exploring flexibility in protein structures. *Nucleic Acids Res.* 2016;44:D423-D428.
14. Gerlt JA, Babbitt PC. Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu Rev Biochem.* 2001;70:209-246.
15. Echave J. Beyond stability constraints: a biophysical model of enzyme evolution with selection on stability and activity. *Mol Biol Evol.* 2019;36:613-620.