



## How Common is Gain of Function

### Vidya Raghunathan\*

Central Instrument Facility, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi, India

\*Corresponding Author: Vidya Raghunathan, Central Instrument Facility, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi, India.

Email: vraghuna@nii.ac.in

Received: October 01, 2021

Published: October 18, 2021

© All rights are reserved by Vidya Raghunathan.

### Abstract

Evolution as we know from the time of Darwin C and Wallace AR moves in the direction of the survival of the fittest. With progressively greater understanding of biological systems and the microenvironment inside a cell, biological fitness at the molecular level has come to the fore. Evolution of metabolic regulation in parasites, as an adaptation to their respective environmental challenges is a case in point [1]. Kinetoplastida's intracellular organelle the glycosome that sequesters over a thousand proteins, is thought to have assisted in the metabolic adaptability and acquisition of parasitism [2].

**Keywords:** Glycosome; *Leishmania mexicana*

### Introduction

'Gain of function' is a little understood process but of vast importance in survival and evolution. Much of the work in this direction has been aided by high resolution structural studies through X-ray crystallography, electron microscopy, nuclear magnetic resonance or other spectroscopic methods. Evolutionary pressures that result in interconnectedness of species is mirrored within a cell as signaling pathways. For example there is a change in cellular status like pH, levels of secondary metabolites like calcium, state of phosphorylation of receptors etc. resulting from environmental stimuli. Many new cellular-stress signaling pathways are being discovered which get activated during oxidative stress, viral infection etc. [3]. Mutations and modifications in some cases assist the protein in going towards degradation but, mutations also assist in making proteins robust, resistant to degradation and adorn innovative modifications [4]. Interconnectedness of life molecules is exhibited in situations such as proteins that exist simply to assist

other proteins (for example chaperones), or domain swapping of proteins both of which help in increasing their functional longevity. Domains such as Zn-finger, EF-calcium binding, N-terminal or C-terminal signal sequences, protease activation sites, DNA-binding leucine zippers etc., are added to proteins to give the corresponding new functionality. Such domains or modules are important in evolution because they can be easily adapted in new situations [4]. Often the new function becomes the dominant one. These are convergent adaptive solutions wherein a pre-existing answer is adapted to a current problem; at the molecular level this is seen in the adaptation of common transcription factors between birds and mammals for example [5].

The pathway leading to the gain of new function in a protein from its evolutionarily pre-established role is intriguing as well as complex. This area of research is gaining prominence. Housekeeping enzymes were until recently thought to be too important for the cell to meddle with because they are highly conserved across

species and have not changed significantly in sequence or structure through evolution. Mutations do not alter the active site, basic functionality or stability of these enzymes. Generally, mutations are a natural consequence of biological interactions and information transfer between the sequence codes of genes and protein in a cell, a consequence of various correction pathways during cell division or a consequence of sudden or insistent exposure to a new micro-environment.

Domain switching adaptive changes, are both, fruitful and energy-inexpensive modes of altering protein functions. The important requirement is that a protein's stability expressed as free energy ( $\Delta G$ ) is maintained at a threshold level. Mutations that keep the stability above the threshold are well tolerated, as most mutations affect stability and not function [6]. Multiple and drastic environmental changes are dealt with in such a way as to maintain proteostasis, by degradation of aggregated, misfolded or unfolded proteins. A number of heat shock proteins of various classes has been identified in the proteome of *Leishmania infantum* [7]. Further demonstrating the importance of protein stability is the trade-off between protein stability and activity, a marking feature and general mechanism of all cold-adapted enzymes, where the flexible active site becomes tolerant of mutations. The thermolabile cold activated enzymes' behavior embodies the balance between protein stability and conformational flexibility. The general paradigm then becomes that enzymes evolve through a rational need to fit changing environmental niches [8] which may involve an exploratory component. One of the key type of inter-residue interactions in proteins that help adaptive evolution are called epistatic interactions [9].

Our studies with housekeeping enzymes from *Leishmania mexicana* are an example of how evolutionary pressures work, balancing protein stability and energy-expenditures within a cell.

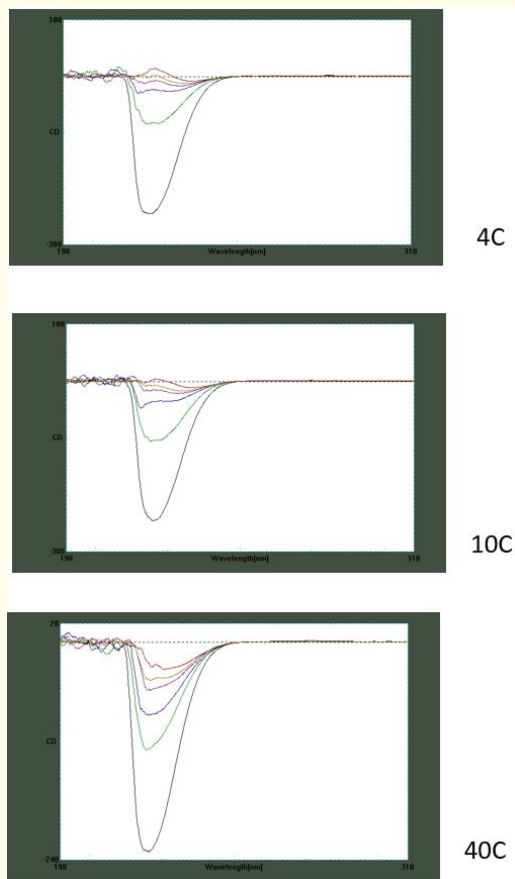
### Results of housekeeping enzyme, Phosphoglycerate kinase from *Leishmania mexicana*

Phosphoglycerate kinase (PGK) is a two domain protein of about 417 residues which catalyzes the reversible transfer of 1-phosphoryl group from 1,3-Diphosphoglycerate to ADP (adenosine diphosphate) producing ATP and 3-PGA in glycolysis. In parasitic protozoa like *Leishmania spp*, PGK molecular weight ranges from 417 residues to 479 residues although the three-dimensional structural

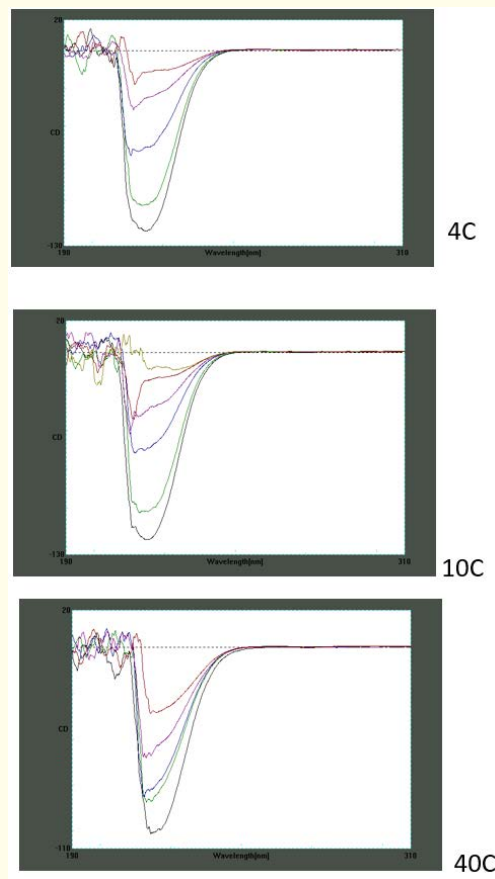
organization of the protein doesn't change and the active sites are completely conserved in order to retain its biological function. Typically, proteins that have a signal sequence in their N-terminus have it for the purpose of transport and then lose them through proteolytic cleavage (like chloroplast PGK). In *Leishmania spp*, PGK exists as 3 isoforms of very different molecular weights. The cytoplasmic PGK (PGKB) is 417 amino acids long and homologous to other PGK's. Another isoform called PGKC having a 62 residue addition at the C-terminal (making it 479 residues), gets sequestered in an intracellular organelle called glycosome. *Leishmania* PAS-PGK of 59 kDA is an isoform that is activated only at acidic pH [10]. PGK being a housekeeping enzyme is structurally stable and functionally preserved across most organisms, any mutation in the protein is tolerated provided these two factors are largely unaffected. *In vivo* folding studies have shown that rate of folding of a protein is affected by its stability within the cellular milieu and exhibits spatially heterogeneity even within a single cell [11]. Ebbinghaus, *et al.* showed that PGK folding kinetics were slower and thermal stability higher *in vitro* compared to *in vivo*. Thus PGKC\_Leishmania confined within an organelle, the glycosome, may have a different stability compared to its cytoplasmic counterpart. *In vitro*, *E.coli* expressed PGKC\_Lmexicana forms inclusion bodies in our studies whereas PGKB\_Lmexicana the cytoplasmic isoform remains soluble (data not shown). Within the cell PGKC stability may be enhanced through membrane association or interaction with membrane associated proteins although peroxin proteins that assist in glycosomal transport have not been experimentally shown to have an interaction with PGKC.

Circular dichroism at different temperatures illustrate the difference in stability *in vitro* between PGK's from Yeast and *Leishmania mexicana mexicana* (Figure 1A and 1B). Cytoplasmic PGK, PGKB\_Lmexicana, shows less cooperative unfolding as compared to Yeast PGK with perhaps more intermediates involved.

Our modelling studies and structural analysis have revealed a novel function of the C-domain extension of PGKC in membrane binding. PGKC\_Lmexicana binding to the membrane may be important for substrate transport through the membrane [12]. The multi-functionality of *Leishmania spp* PGKC is unique, whereby the basic function is retained and added upon. That the 62 residue new domain at the C-terminus of PGKC (as compared to PGKB) is not destabilizing to the protein may be due to a specific micro-



**Figure 1A:** Far UV Circular dichroism spectra of Yeast PGK at varying temperatures. Yeast PGK was purchased from Sigma Chem. Co. India.



**Figure 1B:** Far UV Circular dichroism spectra of PGKB\_ *Lmexicana* at varying temperatures. Plasmid encoding PGKB\_ *Lmexicana* gene was transformed into *E. coli* and expressed and purified as described in our previous work [12].

environment within the glycosome and/or to the presence of epistatic interactions in the folded PGKC [13]. Mutational selection which normally purges mutations that don't maintain the existing scaffold and function, clearly is not in play here. The gain of function is apparently significant enough to overcome energy barriers and evolutionary pressures. This supports the belief that, within the cellular milieu and due to constant environmental pressures, proteins needn't be too stable but rather just stable enough to stay functional.

Figure 1B gives a first glimpse that the conformational landscape of PGK\_ *Lmexicana* may have many conformations in the

ensemble. Sabelko., *et al.* suggest that PGK follows 'strange folding kinetics' which means that the downhill folding is a multi-time scale process with an array of temporary conformations quite unlike a 2-state or intermediate-involved kinetics [14,15]. More detailed studies of PGK\_ *Lmexicana* folding will lead to interesting insights with regards to the structural domain organization and inter-domain interactions.

### Discussion and Conclusion

We take the example of PGK\_ *Lmexicana* to unravel gain of function in this protein. The high adaptability of the metabolic pathway

is key to survival for trypanosomatida in a hostile environment just as infection is key to viral propagation. *Leishmania spp* parasite has to survive in different hosts and also harsh milieu for which, glycosome plays the role of metabolic hub, energy reserve and ATP homeostasis [16]. Functional promiscuity is amply demonstrated in the case of PGK *Leishmania spp*. PAS-PGK has the PAS domain that primarily regulates activity with respect to pH change [10]. The encapsulation of PGKC into the glycosome would change the protein environment and thereby its thermostability. PGKC has an additional domain at the C-terminus containing a membrane binding helix that ends with a double arginine RR and a membrane dimerization motif GXXXG. The GXXXG motif lies on the surface of PGKC as part of a small hydrophobic patch formed by L462, L463 and I464 (region <sup>462</sup>LLIGIFIG<sup>469</sup> of PGKC *Lmexicana*) [13]. GXXXG motif with the flanking  $\beta$ -branched residues give a stabilizing effect to the folded structure of the 62-mer domain which sits in the groove between the two main domains [13,17]. The 62-mer that has the potential to interact with the membrane forms a stable fold in a membrane-free environment through epistatic interactions. PGKC *Lmexicana*, having acquired a new domain, retains the closed fold of the old protein, but stabilizes into a folded structure through a new hydrophobic patch strengthened by epistatic interactions [13]. The interactions that stabilize the new domain may weaken through membrane interaction, implying that membrane association of PGKC is transient. Inter-residue interaction seem to drive/control evolutionary process here. The evolvability of this core metabolic protein PGKC in response to environmental pressures goes against the current thinking that secondary metabolism responds to environmental changes whereas core metabolism remains unchanged [18]. Rather, challenges and resistance to change may be more severe but not impossible, in the case of core metabolism.

Two points are demonstrated here, 1) functional gain does not happen with one or two mutations within the same fold and 2) promiscuous mutations/additions in part of the protein sequence outside of exposed loops or bends does occur. Now I want to come to the controversial issue (being presently debated) of gain of function in the context of the RNA virus SARS-CoV-2. RNA viruses show ten-thousand fold higher mutation rates compared to any other organisms. It was shown in 1986 that the high nucleotide substitution rates in hemagglutinin (HA1) and neuraminidase of Influenza virus was not due to positive selection pressure owing to host immune system but due to high viral mutation rates [19].

RNA viral proteins tend to be loosely packed and demonstrate less destabilizing free energy effects of mutations. The robust threshold stability of normal proteins is not seen in RNA viral proteins and their average stability loss per mutation is less compared to non-viral proteins. The furin binding sequence presumably acquired through natural mutation on the spike protein of SARS-CoV-2 has 3/4 contiguous amino acids added to the sequence of the Spike, which are more solvent exposed than in other coronaviruses. This furin cleavage site is not found in the closely related SARS-CoV/SARS-like CoV's, but is present in MERS-CoV, Zika virus, Ebola Zaire virus, HIV, Avian H5N1 where its presence increases pathogenesis through furin activation. An endoprotease discovered in 1990, Furin cleaves at the consensus site RXK/RR↓ [20]. Anthrax toxin protective agent and avian influenza virus hemagglutinin are furin substrates; furin acts from the cell surface to activate these diverse pathogens. Spike protein (S) of SARS-CoV-2 being 76% similar to that in SARS-CoV has a furin like protease recognition site present in the vicinity of the S-protein maturation site [21]. While furin enzyme cleaves at S2 site at KR↓SF, in SARS-CoV-2 a solvent exposed PRRAR↓SV is present on S1 which can also be cleaved during virus egress. This is the gain of function which is intriguing, in its differences compared to other betacoronaviruses. Papa G., *et al.* have shown that the furin site is not essential for infection or cell-cell fusion in the case of SARS-CoV-2 [22], whereas, virulence of many other viruses is directly correlated with their ability to incorporate a consensus furin cleavage site within their envelop proteins. Papa G., *et al.* also say that antivirals that target furin, which have been found effective for example in the treatment of HIV1, are unlikely to be effective in COVID patients. The exact location of the furin binding site is close to the proteolytic processing site but seems not to be required for proteolytic processing [22]. Thus the furin site could have been a consequence of natural mutation but in structure and stability terms, this seems very unlikely. Just like the trypanosomatids use the glycosome for metabolic flexibility, to enter into a quiescent state to minimize host immune response, SARS-CoV-2 may need versatility during infection, but as discussed above positive selection of furin site for increased infectivity is not evident.

Most natural mutations do not affect the active site or protein stability, with the exception being psychrophilic enzymes where destabilizing the structure goes hand in hand with acquiring broader substrate specificity and lowering substrate affinity [23]. Acquiring a new domain or additional functions must occur balancing a

huge number of interaction energies within the protein, maintaining crucial hydrophobic domains and tertiary interactions, cellular and environmental factors. If the spike protein is not rigid enough to restrict mutations it is valid to ask the question how does it retain it? Epistatic interactions are one of the key interactions that help stabilize evolving protein structures; very important during evolution, a compensatory mutation occurs to stabilize the adverse effect of a prior mutation in a nearby location as seen in the case of PGKC\_Leishmania. The occurrence of epistatic interactions and negative epistatic interaction (which lead to steep decline in protein fitness) demonstrates how intricate and complicated the process of acquiring new function could be for a protein. A few examples are cited here now. Small changes in sequence cause conformational switching as in prions. Inter-protein interactions are more susceptible to environmental conditions than inherent stability of each protein component of the complex. In protein-protein complexes, pairs of correlated mutations establish themselves in a population during evolution [for example the cytochrome C/cytochrome oxidase heterodimer [24]. The flexibility at a given site on the protein or a physiological role of that particular site, both, may direct evolutionary rate and type for overall gain. Polymorphism of proteins has a higher entropy as compared to deleterious mutations [25].

Each mutation in accumulating mutations will reduce protein fitness. This then opens the question of how easily can multiple residues such as furin domain get added. A naturally acquired furin domain is inconsistent with our understanding of the process of mutability of protein structure versus maintenance of protein function. Functional promiscuity which is key in evolution is not evident here unlike the case of *Leishmania spp* PGKC. Coevolutionary intelligence and convergence works at a molecular level as seen with psychrophilic enzymes whose molecular adaptations arise from both genetic coding and long-term selections. Variation of substitution rates over different sites in the protein and variation in rates over time together need to be factored in [5], especially in case of a convergent evolution such as acquisition of common set of amino acids of a desired specific function. Information theory based analysis of viral proteins, shows high protein polymorphism rather than mutability [25] however, how this applies to the furin domain of SARS-CoV-2 remains to be established.

### Acknowledgement

This study and review was supported by the National Institute of Immunology, NII during the trying times of the two COVID waves of 2020/21. I wish to thank Director, NII for scientific support.

### Bibliography

1. Fernandes PM., *et al.* "Kinetic and structural studies of Trypanosoma and Leishmania phosphofructokinases show evolutionary divergence and identify AMP as a switch regulating glycolysis versus gluconeogenesis". *FEBS Journal* 287.13 (2020): 2847-2861.
2. Gualdron-Lopez M., *et al.* *International Journal for Parasitology* 42 (2012): 1-20.
3. Eldeeb MA., *et al.* *Trends in Biochemical Sciences* 44.3 (2019): 181-183.
4. Rorick MM and Wagner GP. *Genome Biology and Evolution* 3 (2011): 456-475
5. Goldstein RA., *et al.* *Molecular Biology and Evolution* 32.6 (2015): 1373-1381.
6. Tokuriki N and Tawfik DS. *Current Opinion in Structural Biology* 19 (2009): 596-604.
7. Sanchiz A., *et al.* "The Experimental Proteome of *Leishmania infantum* Promastigote and Its Usefulness for Improving Gene Annotations". *Genes* 11 (2020): 1036.
8. Santiago M., *et al.* "Discovery, Molecular Mechanisms, and Industrial Applications of Cold-Active Enzymes". *Frontiers in Microbiology* 7 (2016): 1408.
9. Gupta A and Adami C. *PLOS Genetics* 12 (2016): e1005960.
10. Biswas S., *et al.* "Regulation of *Leishmania* major PAS domain-containing phosphoglycerate kinase by cofactor Mg<sup>2+</sup> ion at neutral pH". *FEBS Journal* (2020).
11. Ebbinghaus S., *et al.* *Nature Mathematics* 7 (2010): 319-323.
12. Kaushik S., *et al.* *Molecular and Biochemical Parasitology* 185 (2012): 27-35.
13. Srinivasan S., *et al.* *Allergy, Drugs and Clinical Immunology ADCI* 2.1 (2018): 44-55.
14. Sabelko J., *et al.* *Proceedings of the National Academy of Sciences of the United States of America* 96.11 (1999): 6031-6036.



15. Osvath S., et al. *Journal of Molecular Biology* 333 (2003): 187-199.
16. Jamdhade MD., et al. *OMICS: A Journal of Integrative Biology* 19.3 (2015): 157-170.
17. Senes A., et al. *Journal of Molecular Biology* 296 (2000): 921-936.
18. Tokuriki N., et al. *Science* 324 (2009): 203-207.
19. Saitou N and Nei M. *Molecular Biology and Evolution* 3.1 (1986): 57-74.
20. G Thomas. *Nature Reviews Molecular Cell Biology* 3.10 (2002): 753-766.
21. Coutard B., et al. *Antiviral Research* 176 (2020): 104742.
22. Papa G., et al. *Plos Pathogens* 17.1 (2021): e1009246.
23. Feller G. "Psychrophilic Enzymes: From Folding to Function and Biotechnology". *Scientifica (Cairo)* (2013).
24. Rawson PD and Burton RS. *Proceedings of the National Academy of Sciences of the United States of America* 99 (2002): 12955-12958.
25. Adami C. *Physics of Life Reviews* 1 (2004): 3-22.

**Volume 4 Issue 11 November 2021**

**© All rights are reserved by Vidya Raghunathan.**