Review Article

CrossMark

Functionally tolerable mutations incorporated in proteins guides the evolution process

Abstract

Evolution, often considered with large sample size, positive natural selection drives the path of improvement in biological sequence information. Evolution in proteins is much more complex than in DNA. Factors assisting the evolutionary rate and models are focused in reference to proteins to emphasize their major functional role in evolution. Environmental stresses posing sudden fluctuations in the homeostasis level of functional proteins result in respective alterations in DNA as the functioning of DNA is dependent on proteins. Randomness generated in DNA due to pool of mutations leads to functional evolution of proteins and the ones possessing such evolved proteins are selected and produce individuals for further adaptive evolutionary cycles.

Keywords: evolution, selection, protein, DNA, mutation

Ashish Runthala
Birla Institute of Technology & Science, India

Correspondence: Ashish Runthala, Birla Institute of Technology & Science, Pilani, India, Tel: +91 – 07597971146, Email ashish.runthala@gmail.com

Introduction

Evolution is defined as adapted set of positively favorable changes across successive generations in heritable characteristics of different biological species.[1] It often imposes divergence in the organisms at every evolutionary level, be it genetic or epigenetic.[2] Canalization thus study the genotypic variants contributing to the similar phenotype.[3] Hence irrespective of genotypes, genetic alterations assisting the individual's phenotype solely guides the evolutionary adaptation. In this regard, the *molecular clock hypothesis* proposed by Zuckerkandl & Pauling[4] seems to be correct. Sarich,[5] Wilson[6] also demonstrated the divergence of humans and apes approximately 5million years ago through the same hypothesis.

Evolutionary factors

Evolution being highly dimensional phenomenon is almost impossible to explain simply. Even the mathematical analysis fails to correctly track the evolutionary rate in a population.[7,8] While diversions exist for every evolved factor among different species, there are some factors which constantly catalyze and assist the associated evolutionary track. Such factors including the Mutation (be it lethal or linked),[9–13] Genetic Recombination & count of such sites occurring during the crossing-overs[14–17] and Environmental Constraints form the primary tools to pave the path for evolution. Such factors normally act along with several other key constraints together to demarcate the evolutionary path. Such additional constraints include the Natural Selection and Adaptation. While Natural Selection is the major reason of successfully evolving species over time, the successful adaptation of the individuals against the aforementioned additional constraints implies their successful selection in the nature to grow and reproduce further. Successfully adapted mutation favoring the individual's survival is thus naturally selected and is thus taken care to pave the way for further evolution.

Based on these evolutionary rate and effect deciders, several evolutionary models have been developed. Prebiotic environmental mimicking experiment to simulate conditional synthesis of small life initiating molecules was conducted.[18] Such conditional constraints might have yielded lipid bilayers to distinguish the specific boundaries of life. Phospholipids aid several key activities in the cell including

transcription,[19] proper protein folding,[20,21] cellular and membrane transport functions[22,23] and signal transduction.[24–26] RNA World hypothesis was also postulated, as they can store information and can also work as self-replicating ribozymes.[27] Proteinaceous Era is also considered nowadays. Proteins dominate the cell's catalytic activity and their inability to code the life-supporting information and transfer it in evolving generations might have forced them to simply serve as the functionally assisting tools of a cell.[28]

Selective mutations guide functional divergence

Proteins mark the actual destiny of a mutation, be it liable for functional improvement of protein and cognitive inheritance, or should be considered trivial and dumped from the gene pool over a long time span. Acetyltransferase protein templates screened for a CASP (Critical Assessment of Structure Prediction) Round -VIII target T0400 through HHSearch[29] resulted in several hits, of which the closely relevant and functionally diverged proteins (related only due to a domain swapping) were considered for the study. The proportion of residue composition for all the standard 20 amino acids for all these sequences were then analyzed to yield the data summarized in Figure 1. By the rank of HHSearch score, the mutual culling of structures resulted in many hits, of which 15 hits (11 Close and 4 distantly related hits) were considered.[30]

From the Figure 1 we can interpret that the first ten protein sequences show almost complete homology while the next 4 sequences, being mutually more sequence homologous are distantly related. This result shows that the functional domains of a protein remains conserved even during longer course of time although there are a lot of mutations in its DNA sequence during that period. Hence not all mutations necessarily result in functional change in a species. So, the average proportion of first 11 HHSearch ordered hits from 2Q7BA to 2BEIA is calculated to be 0.0475. The equivalent resultant value for the latter 4 very distant HHsearch hits with considerably lesser reliable scores compared to the former set of 11 hits is 0.0485. Proportion of Hydrophobic (P,A,G,L,I,F,W,M,V), Hydrophilic (D,E,H,R,K) and Neutral amino acids (C,S,T,Q,N,Y)in the former set is 0.4817, 0.2636 and 0.2054. Similar figures for the latter set are 0.5148, 0.2474 and 0.2070. Considering the evolutionary source of these proteins as shown in Table 1, it clearly let us make a firm belief

that over the evolutionary period for these set of proteins, propensity for alteration of hydrophobic amino acids is much more even though they form the protein core. The functionally selected and evolved alteration in these amino acids change the entire topology of the model structure and the fold topology is tweaked slightly.[31] Concomitant decrease in average percentage of hydrophobic amino acids leads to proportional increment in the number of hydrophilic amino acids, thus establishing the belief that the altered mutations would probably result in differential behavior of superficial protein surface towards the ligands and the solvent, for the altered efficiency of the protein function, which is common for all the considered proteins with minor differences in the domain, organism and phylogenic differences.

	Closely relevant top ranked HHSearch Hits												Low ranked Distant hits				
Sequence Length	164	171	148	171	157	140	152	170	175	163	156		181	168	181	163	
Amino Acid	2Q7BA	2J8MA	2AE6A	2I79A	2FIAA	2PDOA	1Y9KA	1GHEA	2OH1A	2VI7A	2BEIA		2QECA	1B6BA	1B87A	1CM0A	
ALANINE	0.0479	0.12571	0.09259	0.08092	0.04217	0.04294	0.05882	0.13873	0.07386	0.07738	0.08537		0.1326	0.09714	0.04762	0.02941	
ARGININE	0.07186	0.08	0.01852	0.0578	0.05422	0.07362	0.04706	0.0578	0.05114	0.125	0.06098		0.05525	0.07429	0.04762	0.02941	
ASPARAGINE	0.05389	0.03429	0.01235	0.04624	0.0241	0.04294	0.02353	0.01734	0.03409	0.02381	0.0122		0.0221	0.01714	0.04233	0.04706	
ASPARTIC	0.07186	0.06286	0.04938	0.04624	0.04217	0.06135	0.02353	0.05202	0.0625	0.07143	0.05488		0.0663	0.02286	0.07407	0.02941	
CYSTEINE	0.00599	0	0.00617	0	0.00602	0.01227	0.01176	0.01734	0.00568	0	0.03049		0	0.03429	0	0.01765	
GLUTAMINE	0.05389	0.05714	0.0679	0.05202	0.04819	0.01227	0.03529	0.05202	0.03977	0.04167	0.03049		0.02762	0.03429	0.03175	0.03529	
GLUTAMIC	0.07784	0.04	0.06173	0.09249	0.09036	0.09816	0.07647	0.05202	0.07386	0.05952	0.08537		0.0442	0.07429	0.08995	0.05294	
GLYCINE	0.03593	0.10857	0.05556	0.08092	0.04217	0.07975	0.07647	0.08671	0.06818	0.07738	0.09756		0.07182	0.07429	0.06878	0.07647	
HISTIDINE	0.01198	0.01714	0.04321	0.01734	0.03012	0.02454	0.01765	0.03468	0.01705	0.02976	0.0122		0.01657	0.06286	0.02116	0.04118	
ISOLEUCINE	0.07784	0.05143	0.08642	0.09249	0.05422	0.04908	0.08235	0.01156	0.08523	0.03571	0.07317		0.04972	0.02857	0.07407	0.08235	
LEUCINE	0.08982	0.09143	0.08025	0.13295	0.08434	0.09202	0.08824	0.13873	0.10227	0.1131	0.09146		0.0884	0.12571	0.08995	0.08235	
LYSINE	0.08982	0.01143	0.04321	0.04624	0.07831	0.03067	0.06471	0.03468	0.0625	0.01786	0.06098		0.03315	0.00571	0.04233	0.10588	
METHIONINE	0.01796	0.01714	0.02469	0.01734	0.05422	0.03681	0.03529	0.02312	0.03409	0.02976	0.0122		0.01657	0.01143	0.02116	0.02941	
PHENYLALANINE	0.07784	0.03429	0.04938	0.04046	0.06627	0.02454	0.02353	0.0289	0.05682	0.01786	0.04878		0.0442	0.06286	0.02116	0.05882	
PROLINE	0.02994	0.04	0.02469	0.00578	0.01807	0.03067	0.02353	0.0289	0.01705	0.02381	0.01829		0.09392	0.04571	0.05291	0.04118	
SERINE	0.02395	0.05143	0.0679	0.04624	0.04217	0.0184	0.05882	0.04624	0.06818	0.05952	0.03049		0.05525	0.05714	0.05291	0.02941	
THREONINE	0.02994	0.03429	0.03086	0.04046	0.05422	0.01227	0.03529	0.04046	0.04545	0.02976	0.02439		0.05525	0.02857	0.05291	0.04706	
TRYPTOPHAN	0.00599	0.02286	0.01235	0.01156	0.01807	0.01227	0	0.01156	0.02273	0.0119	0.01829		0.0221	0.01714	0.02116	0.00588	
TYROSINE	0.05988	0.02857	0.04938	0.0289	0.04819	0.03067	0.04118	0.03468	0.03977	0.04167	0.05488		0.03315	0.01143	0.04233	0.05294	
VALINE	0.0479	0.06857	0.03704	0.05202	0.04819	0.07362	0.07059	0.07514	0.03409	0.08333	0.04878		0.07182	0.07429	0.06349	0.06471	
Other amino acids	0.01798	0.02285	0.08642	0.01159	0.05421	0.14114	0.10589	0.01737	0.00569	0.02977	0.04875		0	0.03999	0.04234	0.04119	
Average	**0.0491**	**0.04886**	**0.04568**	**0.04942**	**0.04729**	**0.04294**	**0.04471**	**0.04913**	**0.04972**	**0.04851**	**0.04756**		**0.05**	**0.048**	**0.04788**	**0.04794**	
Average of Group	**0.04754**												**0.04846**				
											Average						**Average**
Proportion of Hydrophobic Amino Acid Changes	0.43112	0.56	0.46297	0.51444	0.42772	0.4417	0.45882	0.54335	0.49432	0.47023	0.4939	**0.48169**	0.59115	0.53714	0.4603	0.47058	**0.5148**
Proportion of Hydrophilic Amino Acid Changes	0.32336	0.21143	0.21605	0.26011	0.29518	0.28834	0.22942	0.2312	0.26705	0.30357	0.27441	**0.26365**	0.21547	0.24001	0.27513	0.25882	**0.2474**
Proportion of Neutral Amino Acid Changes	0.22754	0.20572	0.23456	0.21386	0.22289	0.12882	0.20587	0.20808	0.23294	0.19643	0.18294	**0.20542**	0.19337	0.18286	0.22223	0.22941	**0.2070**

Figure 1 Residue Composition Data for the 11 close HH Search Hits for CASP8 target T0400 and 4 distant low scored hits, showing the proportional differences in the percentages of Hydrophobic, Hydrophilic and Neutral amino acids.

Table 1 Evolutionary origin of the proteins

	Sequence length	Amino acid	Source/Functional deviations
Closely Relevant Top Ranked HHSearch Hits	164	**2Q7BA**	*Streptococcus agalactiae/-*
	171	**2J8MA**	*Pseudomonas aeruginosa /-*
	148	**2AE6A**	*Enterococcus faecalis* / GNAT family
	171	**2I79A**	*Streptococcus pneumoniae* / GNAT family
	157	**2FIAA**	*Enterococcus faecalis*
	140	**2PDOA**	*Shigella flexneri* / GNAT Family
	152	**1Y9KA**	*Bacillus cereus* / IAA family
	170	**1GHEA**	Tabtoxin Resistance Protein
	175	**2OH1A**	*Listeria monocytogenes* / GNAT family
	163	**2VI7A**	*Pseudomonas aeruginosa*
	156	**2BEIA**	*Homo sapiens* / Thialysine n-acetyltransferase
Low Ranked Distant Hits	181	**2QECA**	*Corynebacterium glutamicum* / Histone acetyltransferase
	168	**1B6BA**	*Ovis aries* / SEROTONIN N-ACETYLTRANSFERASE
	181	**1B87A**	*Enterococcus faecium*/ Aminoglycoside 6'-N-Acetyltransferase
	163	**1CM0A**	*Homo sapiens* / PCAF/Coenzyme-A Complex

Discussion

70% Mutation is shown to be lethal with damaging effects.[9] This is mostly the case in all living systems, as most of the mutations normally alter the protein expression pathway and thus alter the relevant biochemical network of proteins. So only the logistics of biased evolution exists.[32,33] For example, if a mutating nucleotide at a locus is favored more in genera than a different nucleotide, it essentially implies the conserved nature of the structurally evolved structural decoy due to the biased mutation. Such mutational biases also assist the track of natural selection and adaptation. And also, such a succession of biased mutations along with differential sets of natural selection, adaptation and evolution results due to the altered environmental constraints at different geographic loci. Differential geographic coordinate loci favor a phenotype more often than others to acclimatize with existing constraints. If n genes are considered for an evolutionary check experiment, and m is the available set of constraints imposing the evolutionary pressure in form of discrete sets of boosters of evolutionary rate, there is the possibility of m x n variant sets of genes in different persons among the total population of t individuals. The reason to take the variant sets as m x n is simply that nature works on multitude of variations. Nothing seems to be conserved for nature. It is just that any variations maximally affecting the expression of housekeeping genes pose an intolerable pressure to survive for the likelihood and hence mostly vanishes. Ones which still survive are normally better than the initial set of individuals and hence surpass the count of initial population over a long term span. So if count of c individuals survives the exposed set of evolutionary pressurizing constraints in the milieu of the organism, it simply means that a proportion of c/t individuals are adapted over the m x n set of constraints. Hence, these systems make up the path clear for further dynamic evolutionary alterations. Such a set imposed constraints is dynamic due to the activities, side-products, physical or chemical products of all the individuals physically available or unavailable.

This iteratively assists and guides the evolutionary track making sure that un-evolved ancestors can't survive in the altered set of conditions. Best example for such a fact is the efficient survival of current genera of *homo sapiens*, compared to the ancestral extinct systems who faced different survival problems that are unavailable now. Simply it is just the available set of ever-dynamic constraints paving the evolutionary path through different expression of genes in relation to altered adaptation and survival rate of individuals. Hence, it is just the altered trend in the stability dynamics of proteins which assist the organism survive better in the ever-changing conditions, both being inter-linked together. Protein stability governed by core hydrophobic residues is often conserved across evolving genera and the exposed loops and turns often altered as a resultant activity of the factors boosting the evolutionary rate. Evolution is thus highly dimensional and fixed patterns never exist in almost every case. Nevertheless, still a few rules do exist. Any mutation altering the conserved chunk of an important protein is not tolerated and hence discarded from the genera set. Any better efficient genotype sustaining the evolutionary pressure better is favored and slowly leads the count of others in the population to make the next set of individuals for further evolution.

Conclusion

Mutations favoring the functional efficiencies of proteins are selected in the nature. Therefore over the time span, when organisms evolve for increased functional efficiency of proteins, we see a probable, reliable and obvious increment in the proportion of hydrophilic amino acids. As per Figure 1, percentage of hydrophilic amino acids has increased in Set I of 11 proteins as compared to Set II of 4 proteins, which are much divergent to Set I. It is interesting to note that percentage of neutral amino acids has not majorly altered over the time span. Such selected mutants benchmarking the positive or tolerable substitutions/mutations in the DNA/genetic makeup of the cell, are retained for further adaptive functional mutations. In contrast to such aforesaid adaptive substitutions, mutations altering the functional efficacy of the protein product for a gene don't show the positive selection feature in further mutation events. Briefly, a cell is hard biased for all the mutations assisting its efficiency of functional proteins in different important biochemical pathways and all the mutations infringing such important biased and necessarily required segments of DNA are ill sustained in population over a longer time span.

Acknowledgements

Conflict of interest

Author declares that there is no conflict of interest.

References

1. Hall BK, Hallgrímsson B. *Strickberger's Evolution*. 4th ed. Jones & Bartlett; 2008. 762p.

2. Jablonka E, Raz G. Transgenerational epigenetic inheritance: Prevalence, mechanisms and implications for the study of heredity and evolution. *Q Rev Biol*. 2009;84(2):131–176.

3. Jablonka E, Lamb MJ. The changing concept of epigenetics. *Ann N Y Acad Sci*. 2002;981:82–96.

4. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel HJ, editors. *Evolving Genes and Proteins*. New York: Academic Press; 1965. p. 97–166.

5. Sarich VM, Wilson AC. Immunological time scale for hominid evolution. *Science*. 1967;158(3805):1200–1203.

6. Wilson AC, Sarich VM. A molecular time scale for human evolution. *Proc Natl Acad Sci*. 1969;63(4):1088–1093.

7. Kimura M. Evolutionary rate at the molecular level. *Nature*. 1968;217:624–626.

8. Lewontin RC, Hubby JL. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura. *Genetics*. 1966;54(2):595–609.

9. Sawyer SA, Parsch J, Zhang Z, et al. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. *Proc Natl Acad Sci*. 2007;104(16):6504–6510.

10. Comeron JM, Kreitman M. Population, evolutionary and genomic consequences of interference selection. *Genetics*. 2002;161(1):389–410.

11. Gillespie JH. Genetic drift in an infinite population. The pseudohitch-hiking model. *Genetics*. 2000;155(2):909–919.

12. Gillespie JH. The neutral theory in an infinite population. *Gene*. 2000;261(1):11–18.

13. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res*. 1966;8(3):269–294.

14. Waldman AS, Liskay RM. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proc Natl Acad Sci*. 1987;84(15):5340–5344.

15. Waldman AS, Liskay RM. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol Cell Biol.* 1988;8(12):5350–5357.

16. Yang D, Waldman AS. Fine-resolution analysis of products of intrachromosomal homeologous recombination in mammalian cells. *Mol Cell Biol.* 1997;17(7):3614–3628.

17. Lukacsovich T, Waldman AS. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics.* 1998;151(4):1559–1568.

18. Miller U, Stanley L. Production of Amino Acids Under Possible Primitive Earth Conditions. *Science.* 1953;117(3046):528–529.

19. Boeneman K, Crooke E. Chromosomal replication and the cell membrane. *Curr Opin Microbiol.* 2005;8(2):143–148.

20. Bogdanov M, Umeda M, Dowhan W. Phospholipid-assisted refolding of an integral membrane protein. Minimum structural features for phosphatidylethanolamine to act as a molecular chaperone. *J Biol Chem.* 1999;274(18):12339–12345.

21. Bogdanov M, Dowhan W. Lipid-assisted protein folding. *J Biol Chem.* 1999;274(52):36827–36830.

22. Huijbregts RP, Topalof L, Bankaitis VA. Lipid metabolism and regulation of membrane trafficking. *Traffic.* 2000;1(3):195–202.

23. Bankaitis VA, Morris AJ. Lipids and the exocytotic machinery of eukaryotic cells. *Curr Opin Cell Biol.* 2003;15(4):389–395.

24. Exton JH. Regulation of phospholipase D. *FEBS Lett.* 2002;531(1):58–61.

25. Sigal YJ, McDermott MI, Morris AJ. Integral membrane lipid phosphatases/ phosphotransferases: common structure and diverse functions. *Biochem J.* 2005;387(pt 2):281–293.

26. Wang X, Devaiah SP, Zhang W, et al. Signaling functions of phosphatidic acid. *Prog Lipid Res.* 2006;45(3):250–278.

27. Gilbert W. Origin of life: The RNA world. *Nature.* 1986;319:618–620.

28. Runthala A. Evolution: Boosters, Drivers and Plausible Models. *Research and Reviews: A Journal of Life Sciences.* 2012;2(2):1–6.

29. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics.* 2005;21(7):951–960.

30. Runthala A. Protein Structure Prediction: Challenging Targets for CASP10. *J Biomol Struct Dyn.* 2012;30(5):607–615.

31. Sabharwal NA, Runthala A. Functional Protein Domains Evolve Very Specifically Over Mutations. *J Proteomics Genomics.* 2014;1(1):102.

32. Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci.* 2007;104(Suppl 1):8597–8604.

33. Petrov DA, Sangster TA, Johnston JS, et al. Evidence for DNA loss as a determinant of genome size. *Science.* 2000;287(5455):1060–1062.