The structural proteome for the primordial glycolysis/gluconeogenesis

I.J. Bezerra do Ó[1], T.G. Rego[2], M.V. José[3] and S.T. de Farias[1]

[1] Laboratório de Genética Evolutiva Paulo Leminsk, Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, Brasil
[2] Departamento de Informática, Universidade Federal da Paraíba, João Pessoa, Brasil
[3] Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad de México, Mexico

Corresponding author: S.T. de Farias
E-mail: stfarias@yahoo.com.br

ABSTRACT. Comprehending the constitution of early biological metabolism is indispensable for the understanding of the origin and evolution of life on Earth. Here, we analyzed the structural proteome before the Last Universal Common Ancestor based on the reconstruction of the ancestral sequences and structure of proteins involved in glycolysis/gluconeogenesis. The results are compatible with the notion that the first portions of the proteins were the areas homologous to the present-day catalytic sites. These "proto-proteins" had a simple function: binding to cofactors. Upon the accretion of new elements to the structure, the catalytic function could have emerged. Also, the first structural motifs might have have influenced the various types of proteins that are produced in modern organisms.

Key words: tRNA; LUCA; Origin of life; Metabolic pathways; Protein evolution

INTRODUCTION

In "On the Origin of Species by Means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life", Charles Darwin suggested a common ancestor for all living organisms. Since then, the concept of the Last Universal Common Ancestor (LUCA), has triggered discussions about the components of this biological entity (Penny

and Poole, 1999; Forterre et al., 2005; Glansdorff et al., 2008; Mushegian, 2008; Kim and Caetano-Anollés, 2011; Goldman et al., 2013).

One interpretation of the acronym LUCA refers to an already complex organism, with developed information processing, intricate metabolism and a DNA-based genome (Di Giulio, 2003). Thus, LUCA does not represent the earliest form of life, because the evolutionary processes had begun long before DNA appeared. Woese (1998) suggested the existence of communities of progenotes as precursors to LUCA. Progenotes were membrane-delimited organisms with few metabolic functions. In this scenario, these organisms or biological entities had RNA as an informational molecule, and an imprecise translation process, generating statistical proteins, i.e., sequences of peptides arranged probabilistically, which would be under the influence of natural selection. We can describe the progenotes as a community of subsystems that coevolved in such a way that at a particular time of development a higher degree of dependence became established among them, which culminated in the congregation of these subsystems (Woese, 1998).

It stands to reason to assume that not all steps of the modern metabolic pathways were present in a progenote; different steps of metabolism appeared in different progenotes. The ability to translate certain types of information was most likely one of the first properties, if not the first, to develop among progenotes (Eigen and Schuster, 1978; Woese, 1998).

According to Weiss et al. (2016), there were traces of carbon, nitrogen and energy metabolism in the genome of the LUCA. Carbon metabolism was most likely present in the earliest forms of life as we know it today, arguably not only by its vast presence in all life domains but also by its importance in generating energy and precursors for biological systems. Among the carbohydrate pathways, the two most universal and well preserved are the glycolysis/gluconeogenesis and citric acid cycles.

The antiquity of these pathways, specifically gycolysis/gluconeogenesis, is suggested by many features. First, glycolysis/gluconeogenesis is present, at least partially in all life forms, and all cellular life carry variations of carbohydrate metabolism. Second, they have a structural simplicity, meaning that the enzymes act independently and do not require membranes, which would be a necessary characteristic in an Archean ocean in a pre-membrane reality. Thirdly, it is independent of oxygen, another feature necessary at that time.

Additionally, the results of the two substrate-level phosphorylation reactions would provide enough energy for primitive metabolism. With pyruvate as the terminal electron acceptor, the redox balance of the pathway can be maintained. Another reason why this pathway should be prior to LUCA is that glycolysis evolved most likely as a reverse gluconeogenesis. The synthesis of hexose sugars would act as a storage of energy and biological precursors, which would be advantageous in a primitive scenario with scarce complex molecules.

At this point in evolutionary history, large-scale accumulation of sugar was not feasible. Glycolysis is not only responsible for energy output but its three-carbon compound products (e.g. dihydroxyacetone phosphate and glyceraldehyde-3-phosphate) act as a source of energy as well. These products are used in various other pathways, such as amino acid anabolism. Producing the three-carbon compounds was probably the first role of glycolysis, which is supported by it being constant in all life forms even if energy metabolism varies.

Storing carbon was probably advantageous as well, as it decreased organisms' dependence of the system surrounding it.

Large informational molecules most probably did not have sufficient stability on primitive Earth, and the first coding devices must have reflected the chemical constraints of their origin. Eigen and Winkler-Oswatitsch (1981) proposed that the initial molecular events were carried out by tRNAs, mostly because: (1) they have predominant GC pairing, necessary for relative stability within lengths of tRNAs; (2) they have stable folded secondary structures, enhancing their resistance to hydrolytic decomposition; (3) the codons act as messenger carriers; (4) they have around 50 to 100 residues, which is ideal for prebiotic polymerization conditions; (5) ancient family tree; and (6) symmetric structure.

Eigen and Schuster (1978) suggested that the first anticodons were of the type RNY (purine-any nucleotide-pyrimidine). These are codons where the first base is a purine, the third a pyrimidine, and the second is any of the two. Hence, the first modules were rich in alanine, serine, threonine, asparagine, aspartic acid, valine, isoleucine, and glycine.

Farias et al .(2016a) reconstructed the ancestral sequence of tRNAs that may have acted as the first genes of these progenotes. They proposed a proteome for the progenotes based on the translation of ancestral tRNAs with a RNY pattern. Among the various metabolic pathways found in modern organisms, some were already present in LUCA and in the progenotes. Farias et al (2016a), by analyzing the similarities of translated tRNAs with modern proteins, found possible homologies with the following pathways: amino acids pathways, glycolysis/gluconeogenesis, lipid pathways, nucleotide pathways, translation, and transcription or RNA replication.

Here we use the data obtained by Farias et al. (2016a) and their suggestions for the early metabolism to reconstruct two different stages of evolution. One (bottom up) prior to LUCA, which will be based on the proteome established by tRNA ancestor molecules and another (top down) based on the reconstruction of ancestor sequence from modern proteins involved in the metabolism of glycolysis/ gluconeogenesis and three carbon compounds and compare the two strategies to understand the evolutionary history of this metabolic pathway.

MATERIAL AND METHODS

Ancestral Approaches

Two different approaches were used to reconstruct ancestors.

Bottom Up approach

The first one, "bottom up", refers to the reconstructions based on the sequences found by Farias et al. (2016a) – (supplementary information). The proteome suggested by them was made using sequences of ancestral tRNAs following the RNY pattern, which when translated presented matches with modern proteins involved in three-carbon metabolism.

Top Down approach

The other approach was a reconstruction using sequences of the modern proteins chosen based on the ancestral proteome suggested by Farias et al. (2016a), which were the

following: glycerate kinase, triosephosphate isomerase, glucose 6-phosphate 1-dehydrogenase 2, glucose 6-phosphate isomerase, phosphoglycerate kinase, glycerol 3-phosphate dehydrogenase and transketolase.

Ancestral sequence reconstruction

We obtained 400 sequences for each type of enzyme that is part of the glycolysis/gluconeogenesis and three carbon metabolic pathways from the National Center for Biotechnology Information (NCBI). The sequences covered the three domains of life, namely Archaea, Bacteria and Eukarya. The sequences were aligned with the support of the software MAFFT v.7 (Katoh and Standley, 2016). To align the sequences, we used BLOSUM45 for the scoring matrix, and all the other parameters remained as defaults. To reconstruct the most probable ancestral sequence, we used maximum likelihood with the support of software MEGA7 (Kumar et al., 2016). For this approach, we used different levels of strictness for inferring the ancestor sequence by maximum likelihood; 95% for complete deletion and 99% site coverage cutoff. The ancestral sequences obtained are available in the supplementary information.

Structural analysis

The protein structures were reconstructed by homology using the web server I-Tasser (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Zhang, 2008). The templates of modern proteins and the ligands were obtained in Protein Data Base (www.rcsb.org) (Berman et al., 2000).

The structural alignments were performed using the software TM-Align (http://zhanglab.ccmb.med.umich.edu/TM-align/)(Zhang and Skolnick, 2005). For analysis of the ligands to the ancestral structure in both approaches (top-down and bottom-up), we used the software Coach (http://zhanglab.ccmb.med.umich.edu/COACH/) (Yang et al., 2013).

RESULTS

The glycolysis/gluconeogenesis pathways central to modern metabolism, as they are conserved in all domains of life. These pathways are the core of carbon metabolism in modern cells. It has been suggested that glycolysis may have originated as a mechanism that extracted energy and biological precursors from abiotically formed amino acids (Kenneth et al., 2004). Sutherland (2017) suggested the importance of compounds with three carbons in the emergence of the first metabolic pathways. Farias et al. (2016b) proposed that these compounds worked as a distribution center of carbon to other metabolic pathways in development (Jheeta, 2015).

Farias et al. (2016b) reconstructed the ancestral sequence from tRNAs and compared the ancestor sequences translated with the modern proteins. The results suggested that the enzymes of glycolysis/gluconeogenesis metabolism could have emerged from proto-tRNAs as well as the first genes, the proteins involved in the metabolism of three carbon compounds being the core of the process. Here, we call this approach bottom up, where the structures are derived from translated sequences of the tRNA ancestor.

To understand the structural evolutionary history of the glycolysis/gluconeogenesis enzymes, we used a second approach based on the reconstruction of the ancestral sequences and structures from the modern enzymes involved in glycolysis/gluconeogenesis and three carbon compounds (top down approach). We used the same enzymes that had matches with translated sequences of the tRNA ancestor (see materials and methods), to make comparisons between the structures obtained from different approaches. From the ancestral sequences obtained from modern enzymes, we used 95%complete deletion and 99% site coverage cutoff, which may capture similarities with different steps in the evolutionary history of these enzymes.

Ancestral structures reconstruction

Figure 1 shows the reconstructed structures by homology of the tRNA ancestor sequences translated in the bottom up approach. In general, the structures are composed of simple motifs, basically, loops and small alpha helices. Table 1 shows the most similar modern structures with RMSD values (root-mean-square deviation) between ancestor by bottom up approach and modern proteins, and the organism to which the protein belongs.
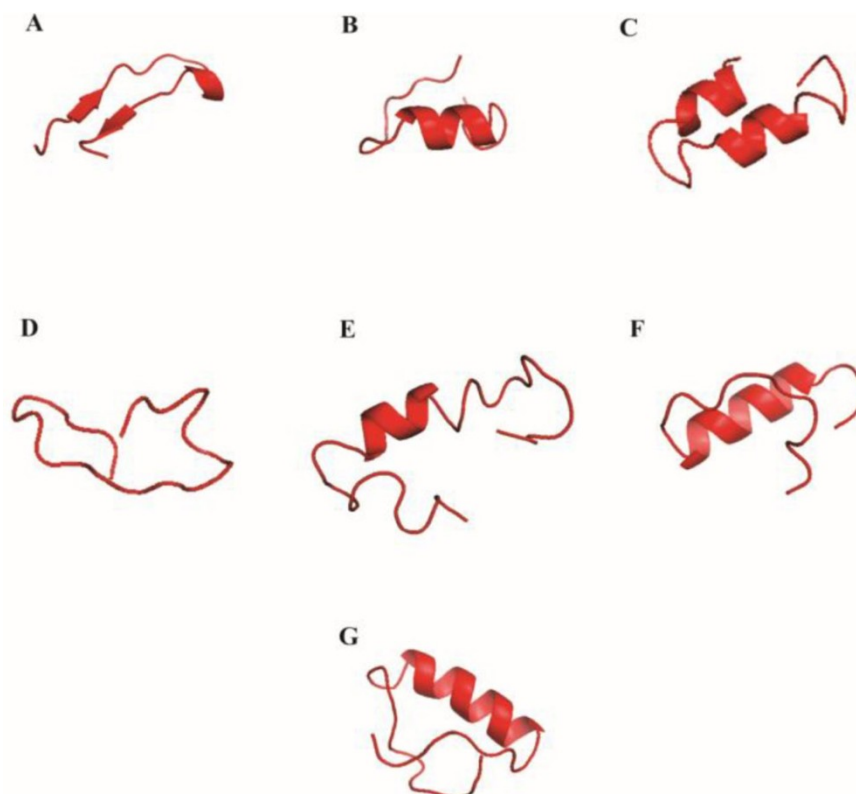
Figure 1. Structures of ancestral sequences obtained by the bottom up approach. A) glyceraldehyde-3-phosphate-dehydrogenase, B) glucose 6-phosphate1-dehydrogenase, C) glucose 6 phosphate isomerase, D) glycerate kinase, E) phosphoglycerate kinase, F) triose phosphate isomerase and G) transketolase.

Table 1. Ancestral proteins obtained with the bottom up approach; the best similar structure in modern organisms and the RMSD distance between the ancestral protein and modern protein.

Bottom Up	Most Similar Structure	Organism	RMSD
Glucose-6-phosphate Isomerase	Iron ABC transporter solute-binding protein (4HN9)	*Eubacterium eligens*	1.69
Phosphoglycerate-kinase	RNA binding domain (4GH9)	*Marburg virus*	2.67
Triosephosphate-isomerase	Hydrolase (Crystal structure of had family enzyme bt-2542 target e fi-501088) magnesium complex (4DFD)	*Bacteroides thetaiotaomicron*	1.46
Glycerate-kinase	Kinase domain of protein tyrosine kinase 2 beta (3CC6)	*Homo sapiens*	1.95
Glyceraldehyde-3-phosphate-dehydrogenase	Enolase superfamily member complexed with Mg and glycerol (3R25)	*Vibrionales bacterium*	2.10
Glucose-6-phosphate-1-dehydrogenase	CRISPR-associated protein Cas6e (4DZD)	*Escherichia coli*	1.78
Transketolase	Acyl-CoA dehydrogenase fold (5AHS)	*Advenella mimigardefordensis*	1.99

These results are in agreement with basic functions for very early proteins in the beginning of the biological systems, when sophisticated catalytic functions were evolving. The initial proteins were very simple and were able to bind to a variety of substrates as they were rudimentary and had low affinity, based on the work of Frances Arnold, the 2018 Nobel Prize Laureate in chemistry (Newton, 2018). The similarities with proteins that are not involved in carbon metabolism can indicate that these basic motifs worked as mounting blocks in the origin of the first proteins and genes.

The most similar modern structures showed basic biological functions such as hydrolase, RNA binding domain, and enolase, with a range of RMSD values of 1.46 to 2.67. These results can indicate the basic structural blocks that participated in the formation of many different pathways, and the combination of these structural parts was relevant to primordial diversification of essential pathways and for the establishment and maintenance of the emerging biological systems.

Figure 2 shows the structures constructed by homology by the top down approach with complete deletion and 99% site coverage cutoff and in Figure 3 with 95% site coverage cutoff. The comparison between the structures obtained from the top down and the bottom up approaches indicates that the structures from the top down approach are more complex.
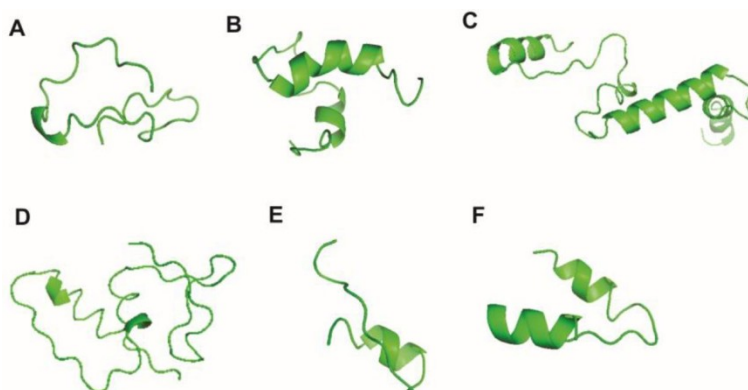

Figure 2. Protein structures are constructed based on the ancestral sequences obtained by top down approach with complete deletion (CD) and 99% site coverage cutoff (PD99). A) glyceraldehyde-3-phosphate-dehydrogenase (PD99), B) glucose 6-phosphate1-dehydrogenase (CD), C) glucose 6 phosphate Isomerase (PD99), D) glycerate kinase (CD), (E) triose phosphate isomerase (CD) and (F) transketolase (PD99).
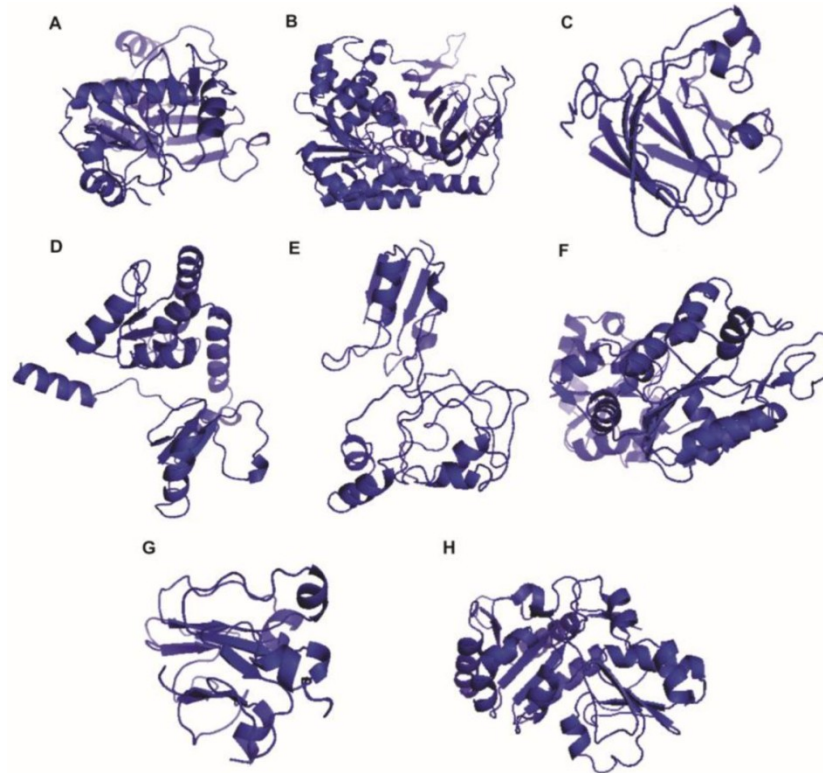
Figure 3. Structures of ancestral sequences obtained by top down approach with 95% site coverage cutoff. A) glyceraldehyde-3-phosphate-dehydrogenase, B) glucose 6-phosphate1-dehydrogenase, C) glucose 6 phosphate Isomerase (I), D) glucose 6 phosphate isomerase (II), E) glycerate kinase, F) phosphoglycerate kinase, G) triose phosphate isomerase and H) transketolase.

In Table 2, the most similar modern structures with RMSD values between ancestor by top down approach and modern proteins are shown. These results reflect the limitations of the method to reconstruct ancestral sequences from modern proteins, where it is possible to identify the first functional domains, but not the complete evolution of these domains, before the establishment of the modern specificity. However, these results also show the efficiency of the method to understand the evolutionary history of proteins until the first functional domain.

Our results show that the suggested ancestral structures are similar to structures of different extremophiles that lived in either acidic or high temperature habitats, or both. This may indicate that these first enzymes before and after LUCA were adapted to extreme scenarios to reflect the environmental conditions on the Earth some 4.0 billion years ago (Bell et al, 2015; Jheeta, 2017). It is suggested that Theia's impact with the Earth created enough energy to melt Earth's mantle. The condensed water formed a dense atmosphere obstructing the dissipation of heat, which resulted in a greenhouse effect (Nisbet et al., 2007). Even though there is still a debate, various pieces of evidence seem to show that the Earth might have reached extremely high temperatures in the hadean eon and this might have influenced biological evolution (Sleep et al., 2007; Zahnle et al., 2007).

Table 2. Ancestral proteins obtained by Top down approach; the best similar structure in modern organisms and the RMSD distance between the ancestral protein and modern protein. In parentheses, the protein database code.

	MOST SIMILAR STRUCTURE			ORGANISM			RMSD		
	Top Down Complete Deletion	Top Down 99% Threshold	Top Down 95% Threshold	Top Down Complete Deletion	Top Down 99% Threshold	Top Down 95% Threshold	Top Down Complete Deletion	Top Down 99% Threshold	Top Down 95% Threshold
Glucose-6-phosphate isomerase	not applicable	glucose-6-phosphate isomerase (3IFS)	Glucose-6-phospho-isomerase in complex with manganese, and 5-phospho-D-arabinotate. (4LUK)	not applicable	*Bacillus anthracis*	*(1)Pyrococcus furiosos. (2)Thermococcus litoralis*		1.62	1.07
Phosphoglycerate-kinase	not applicable	not applicable	Phosphoglycerate-kinase (1PHP)	not applicable	not applicable	*Geobcacillus stearothermophilus*	Not Applicable	Not Applicable	0.85
Triosephosphate-isomerase	Chondroitin polymerase (2Z86)	not applicable	Triosephosphate-isomerase complex with 2-phosphoglycolic acid (1BTM)	*Escherichia coli*	not applicable	*Geobcacillus stearothermophilus*	1.74	Not Applicable	2.37
Glycerate-kinase:	Glycerate kinase (3CWC)	Not Applicable	Glycerate kinase (1TO6)	*Salmonella typhimurium*	not applicable	*Nisseria meningitidis*	1.44	Not Applicable	2.79
Glyceraldehyde-3-phosphate-dehydrogenase	Not Applicable	Solution structure of Bacteriophage Lambda GPW (1HYW)	Glycosomal glyceraldehyde-3-phosphate dehydrogenase (4LSM)	not applicable	*Escherichia virus Lambda*	*Trypanosoma cruzi*	Not Applicable	1.87	0.92
Glucose-6-phosphate-1-dehydrogenase	Enoyl-CoA hydratase (3PES)	Not Applicable	Glucose-6-phosphate-1-dehydrogenase (1H9B)	*Mycobacterium smegmatis*	not applicable	*Leuconostoc mesenteroides*	2.47	Not Applicable	1.23
Transketolase	Not Applicable	Selenomethionine incorporated apo D-serine deaminase	Transketolase with thiamin diphosphate and CA+ (3M34)	not applicable	*Salmonella typhymurium*	*Campylobacter jejuni*	Not Applicable	1.32	0.99

Analysis of substrates binding

Table 3 shows the analysis of substrate binding for the ancestral structures obtained for the bottom up and the top down approaches. For the bottom up approach, the results of substrate binding indicate that these structures could interact with simple compounds that may have worked as cofactors for basic functions. Some of the structures bind to amino acids, most of them bind to coenzymes that have the potential of energy transfer, and all of them bind to small hydrocarbons. The interaction between the ancestral structures and nucleotides also suggest that the coevolutionary relationships between these compounds were established earlier in the primordial environment.

Caetano-Anolles et al. (2012) suggested that in the origin of the initial metabolic pathways, the progressive appearance of complexity was the rule, and that the first proteins or peptides worked by binding various cofactors, and the accretion of new parts made it possible for these primitive proteins to be able to developp catalysis and molecular functions. As in modern biological systems, the model suggests that nucleic acids worked as cofactors in the beginnings of life.

For the structures obtained by the top down approach with complete deletion and 99% site coverage cutoff, we observed a similar pattern of substrate binding with the structures obtained from bottom up approach. For the structures obtained by the top down

approach with 95% site coverage cutoff, the best substrates were similar to the modern proteins. These results suggest that during the evolutionary history, the accretion of new structural parts in the protein made the structure more complex and specific. Increased specificity enabled the fixation of the first metabolic pathways.

Table 3. The ancestral structures obtained by bottom up and top down approaches and the best ligands for each ancestral protein.

	Bottom Up	Top Down Complete Deletion/ 99% Threshold	Top Down 95%
Glucose-6-phopate-isomerase	Ethylene glycol, Nucleic Acid, S-Adenosylmethionine and methionine	Bicarbonate, Chlorophyl, Bacterioruberin, Cholesterol, Zinc	Glucose-6-phosphate and 6-phophogluconate
Phosphoglycerate-kinase	Guanosine Triphosphate	(Not Calculated)	Adenosine Diphosphate and 3-phophoglycerate
Triosephosphate-isomerase	Nucleic Acid, Arginine and Calcium	Magnesium, Nucleic acid, Glucoside and Urea	3-phosphoglycerate
Glycerate-kinase	Adenosine Diphosphate, Mannose, Inositol and Nucleic Acid	Iron, glucoside, Polyphenylene sulfide , Beta-glucose and Nucleic Acid	Phosphate
Glyceraldehyde-3-phosphate-dehydrogenase	Calcium, Benzenemethanamine, Nitride and Tryptamine	Magnesium, Flavin mononucleotide and Lauric Acid	Glyceraldehyde-3-phosphate, Nicotinamide adenine dinucleotide and Glycerate-3-phosphate
Glucose-6-phosphate-1-dehydrogenase	Peptides, Nucleic Acid, Catechol and Adenosine Triphosphate	Carbon, Diethyl-4-methylbenzylphosphonate, [FE4S4] and Glutamic acid	5-phospho-darabinohydroNot Applicableamic acid, Iron, 1-Malate, Zinc ion, Adehydo-n-acetyl-d-glucosamine
Transketolase	Iodide, Nucleic Acid and Calcium ions	Calcium ions, Nucleic acid and Alanine	Thiamine thiazolone diphosphate, 5-O-phosphonobeta-d-ribofuranose, D-fructose-6-phosphate-thiamin-diphosphate-adduct

Evolutionary steps in the natural history of glycolysis/gluconeogenesis

The evolutionary steps are reconstructed in Figure 4. The data suggest that the initial portions of the proteins were simple structural motifs nearly homologous to the modern catalytic sites. During the evolutionary accretion of new motifs, selective pressure worked to protect and isolate the catalytic site, making the protein more specific. Alves et al. (2002) showed that functional blocks of similar chemistry evolved within the metabolic networks. The bottom up ancestors was all structurally close, more than they were to their homologs. Doolittle (2000) discusses, based on the propositions of Woese (1998), the possibility that all genes originate in small RNA molecules that may have developed in a period prior to the appearance of cellular structures. This model is in line with our data and reopens the discussion on how we should understand the origin and evolution of biological information and the evolutionary history relationship between the diverse gene families. These ancestors also had similar substrates but bound to a variety of ligands. This could indicate that they had a common moiety that had a non-specific action. These first "enzyme-like" molecules could have acted at different points of the metabolism, and throughout evolution they gained specificity. Other than gaining specificity, the structure of the

enzymes became larger and more complex. Evolution probably acted around the catalytic site following the accretion model.

We contend that the first action of these enzymes related to glycolysis was actually in the gluconeogenic direction. The synthesis of hydrocarbons would be positively selected since it provided both stored energy and biosynthetic precursors.
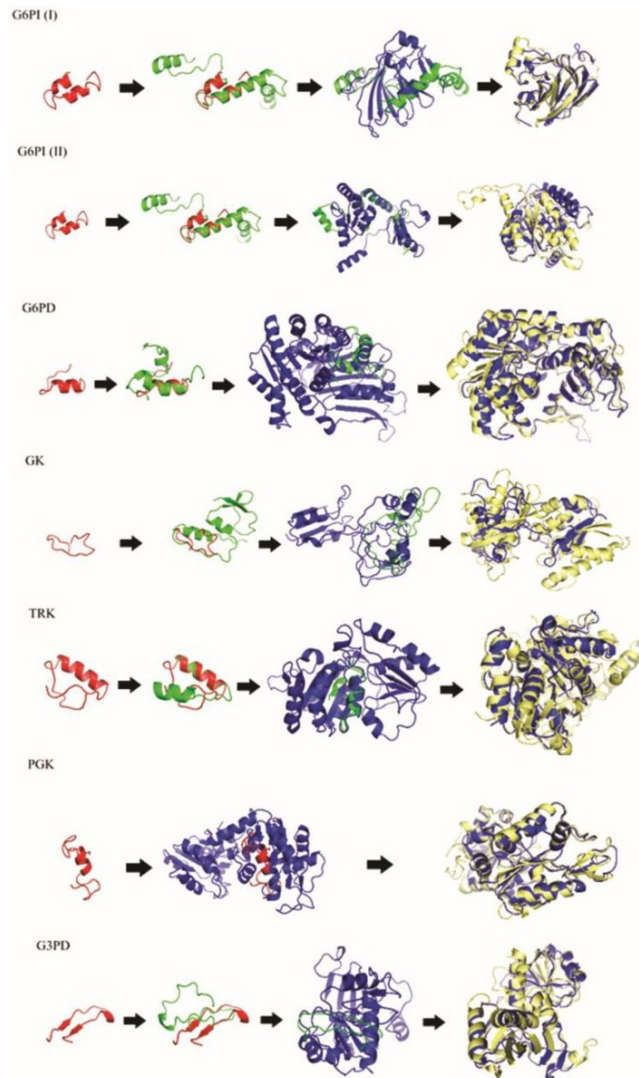

Figure 4. Proposed evolutionary steps of structures. Alignment made using TM-Align server. In red the bottom up ancestors, in green the top down ancestors calculated with complete deletion or the top down ancestors calculated with partial deletion with 99% threshold, in blue the top down ancestors calculated with 95% threshold and in pale yellow the modern enzymes.   gyceraldehyde-3-phosphate-dehydrogenase (G3PD), glucose 6-phosphate1-dehydrogenase G6PD, glucose 6 phosphate isomerase (G6PI (I) and (II)), glycerate kinase (GK), phosphoglycerate kinase (PGK), triose phosphate isomerase (TPI) and transketolase (TRK).

CONCLUSIONS

The comprehension of the origin and evolution of the metabolic pathways is a keystone to the understanding of the emergence of life itself. Here, we analyzed the origin and evolution of the proteins involved in glycolysis/gluconeogenesis. The results endorse the idea that the initial genetic information may have originated from proto-tRNAs. Also, the first peptides worked simply by binding cofactors. Accretion of new motifs resulted in complexification of the protein structure, which allowed for the sophistication of the catalytic domain.

Our work indicates that in the emergence of the first peptides, simple structural motifs worked as mounting blocks, and by the combination of the different motifs, different proteins emerged that work in modern organisms. The evolution of the proteins involved in glycolysis/ gluconeogenesis occurred by selective pressure to protect the catalytic domain, which culminated in the improvement of catalysis. We suggest that gluconeogenesis preceded glycolysis and, with the accumulation of the carbon in the initial biological system, a balance between the pathways was obtained.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Alves R, Chaleil RA and Sternberg MJ (2002). Evolution of enzymes in metabolism: a network perspective. *J. Mol. Biol*. 320(4): 751-70.

Bell EA, Boehnke P, Harrison TM and Mao WL (2015). Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc. Natl. Acad. Sci. U S A*. 112(47): 14518-14521

Berman HM, Westbrook J, Feng Z, Gilliland G, et al. (2000). The Protein Data Bank. *Nucleic Acids Res*. 28(1): 235-42.

Caetano-Anollés G, Kim KM and Caetano-Anollés D (2012). The phylogenomic roots of modern biochemistry: origins of proteins, cofactors and protein biosynthesis. *J. Mol. Evol*. 74(1-2): 1-34. doi: 10.1007/s00239-011-9480-1.

Di Giulio M (2003). The universal ancestor and the ancestor of bacteria were hyperthermophiles. *J. Mol. Evol*. 57(6): 721-30.

Doolittle WF (2000). The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol*. 10(3): 355-358. https://doi.org/10.1016/s0959-440x(00)00096-8.

Eigen M and Schuster P. (1978). The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*. 64(11): 541-65.

Eigen M and Winkler-Oswatitsch R (1981). Transfer-RNA, an early gene? *Naturwissenschaften*. 68(6): 282-92.

Farias ST, Rego TG and Jose MV (2016a). A proposal of the proteome before the last universal common ancestor (LUCA). *Int. J. Astrobiol*. 15 (1): 27-31. doi:10.1017/S1473550415000464.

Farias ST, Rêgo TG and José MV (2016b). tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. *Life (Basel)*. 6(2): E15. doi: 10.3390/life6020015.

Forterre P, Gribaldo S and Brochier C (2005). Luca: the last universal common ancestor. *Med. Sci. (Paris)*. 21(10): 860-5.

Glansdorff N, Xu Y and Labedan B (2008). The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol. Direct*. 3: 29. doi: 10.1186/1745-6150-3-29.

Goldman AD, Bernhard TM, Dolzhenko E and Landweber LF (2013). LUCApedia: a database for the study of ancient life. *Nucleic Acids Res*. 41(Database issue): D1079-82. doi: 10.1093/nar/gks1217.

Katoh K and Standley DM (2016). A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*. 32(13): 1933-42. doi: 10.1093/bioinformatics/btw108.

Kenneth SB (2004). Functional Metabolism: Regulation and Adaptation. *John Wiley & Sons, INC., Publication*.

Kim KM and Caetano-Anollés G (2011). The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evol. Biol*. 11: 140. doi:10.1186/1471-2148-11-140.

Kumar S, Stecher G and Tamura K (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. 33(7): 1870-1874. doi:10.1093/molbev/msw054.

Jheeta S (2015). The Routes of emergence of life from LUCA during the RNA and viral world: a conspectus. *Life*. 5: 1445-1453. doi:10.3390/life5021445.

Jheeta S (2017). The Landscape of the Emergence of Life. *Life*. 7: 27. doi:10.3390/life7020027.

Mushegian A (2008). Gene content of LUCA, the last universal common ancestor. *Front. Biosci*. 13: 4657-66.

Newton J (2018). The potential of promiscuity. *Chem World*. 15(11): 28-29.

Nisbet E, Zahnle K, Gerasimov MV, Helbert J, et al. (2007). Creating habitable zones, at all scales, from planets to mud micro-habitats, on Earth and on Mars. *Space Sci. Rev*. 129: 79-121.

Penny D and Poole A (1999). The nature of the last universal common ancestor. *Curr. Opin. Genetics Dev*. 9(6): 672-7.

Sleep NH (2007). Plate tectonics through time, Treatise on Geophysics Volume 9, (ed. Schubert G.), pp. 101-117 Oxford, Elsevier.

Sutherland JD (2017). Studies on the origin of life – the end of the beginning. *Nat. Rev. Chem*. 1: 12. DOI/10.1038/s41570-016-0012.

Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, et al. (2016). The physiology and habitat of the last universal common ancestor. *Nat. Microbiol*. 1(9): 16116. doi: 10.1038/nmicrobiol.2016.116.

Woese C (1998). The universal ancestor. *Proc. Natl. Acad. Sci. U S A*. 95(12): 6854-9.

Yang J, Roy A and Zhang Y (2013). Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*. 29: 2588-2595. doi:10.1093/bioinformatics/btt447.

Zahnle K, Arndt N, Cockell C, Halliday A, et al. (2007). Emergence of a Habitable Planet. *Space Sci. Rev*. 129:35–78.

Zhang Y, Skolnick J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 33(7): 2302-9.

Zhang Y (2008). I-TASSER server for 3D structure prediction. *BMC Bioinformatics*. 9: 40. doi:10.1186/1471-2105-9-40.