# Central Metabolic Pathways of Hyperthermophiles: Important Clues on how Metabolism Gives Rise to Life

R. S. Ronimus & H. W. Morgan

*Biological Sciences, University of Waikato, New Zealand*
*(Communicating author: ronimus@waikato.ac.nz)*

**Abstract.** Vital clues on life's origins within the galaxy exist here on present day Earth. Life is currently divided into the three domains Bacteria, Archaea and Eukarya based on the phylogeny of small ribosomal subunit RNA (16S/18S) gene sequences. The domains are presumed to share a "last universal common ancestor" (LUCA). Hyperthermophilic bacteria and archaea, which are able to thrive at 80°C or higher, dominate the bottom of the tree of life and are thus suggested to be the least evolved, or most "ancient". Geochemical data indicates that life first appeared on Earth approximately 3.8 billion years ago in a hot environment. Due to these considerations, hyperthermophiles represent the most appropriate microorganisms to investigate the origins of metabolism. The central biochemical pathway of gluconeogenesis/glycolysis (the Embden-Meyerhof pathway) which produces six carbon sugars from three carbon compounds is present in all organisms and can provide important hints concerning the early development of metabolism. Significantly, there are a number of striking deviations from the textbook canonical reaction sequence that are found, particularly in hyperthermophilic archaea. In this paper the phylogenetic istribution of enzymes of the pathway is detailed; overall, the distribution pattern provides strong evidence for the pathway to have developed from the bottom-up.

## 1. Introduction

There are three general theories that aim to delineate the environmental conditions in which life arose during the Archaean period. These are: (1), the hyperthermophilic ($\geq 80°C$), pyruvate-based chemo-autotrophic origin of metabolism; (2), the prebiotic soup; and (3), the bolide-mediated extraterrestrial seeding of the Earth with organic compounds (Orgel 1998; Maden 1995). The latter two models imply a heterotrophic origin for life and are usually associated with a mesophilic temperature, i.e. utilizing the accumulated organic material for energy, while in the hydrothermal autotrophic theory, $H_2$ gas is produced from $H_2S$ and $FeS$ during iron pyrite formation which then supplies the energy (Wächterhäuser 1988). Despite a large amount of empirical data and theoretical work investigating these prebiotic chemistry models it is still unclear, and controversial, which of the three led to the formation of life.

Estimates for when life first developed on Earth have placed this event between approximately 3.5 and 3.8 Ga years ago (relatively soon after Earth's formation). This time frame is based on three lines of evidence including the presence of stromatolites/microfossils, radioisotopic data (e.g. dating of hydroxyapatite) and phylogenetic analyses (Schidlowski 1988; Schopf et al. 2002; Feng, Cho, & Doolittle 1997). Geochemical modeling has indicated a temperature around 80°C for the ocean within this time frame (Baross 1998). Despite the disagreement about the temperature at which life first arose, most agree that the early atmosphere was reducing and had higher concentrations of $CO_2$ (80 to 600–old higher), the latter being an important determinant of reaction rates for carbon fixation.

An important aspect to a hydrothermal origin of life is that these vent sites were located at the bottom of the Archaean ocean, which formed circa 4.3 Ga ago, providing a significant level of protection from large bolide impacts (Baross 1998). The ability of some modern cultured hyperthermophiles to withstand autoclaving temperatures (121°C) for up to an hour is remarkable and evidence even exists for colonization of surfaces between 125 and 140°C (Madigan, Martinko, & Parker 2000). Deep sea hydrothermal activity (ocean depth, circa 3000 m) is thought to have been 3 to 5-fold higher when life formed than it is now. Additionally, vent ecosystems have recently been suggested to represent the environment that has the longest history of continuous microbial habitation (Reysenbach & Shock 2002). These hydrothermal habitats would have existed regardless of the ambient ocean or atmospheric temperatures (Baross 1998). Hyperthermophiles are widespread today, being found at deep-sea hydrothermal vents (black smokers), terrestrial hot springs and subterranean habitats. Geochemical modeling of the reduced hydrothermal vent fluids strongly suggests that many of life's normal metabolic reactions, including amino acid syntheses, are energetically favorable and that methanogenesis and sulfur/sulfate-reduction are favored between 80 and 125°C (Amend & Shock 2001).

One way to examine the origins of metabolism on Earth is by investigating the distribution of key enzymes that are part of highly conserved central metabolic pathways, within the context of the 16S/18S phylogenetic tree of life. Gluconeogenesis, the formation of glucose from pyruvate, forms the backbone of metabolism with connections being made to the tricarboxylic acid cycle (oxidative or reductive) and reductive pentose phosphate cycle, as well as amino acid, polysaccharide, lipid and nucleotide syntheses. Enzymes of the pathway are some of the most slowly evolving of all cellular proteins and can therefore be effectively used as molecular chronometers to assess the deep lineages of the universal tree of life (Fothergill-Gilmore 1986).

## 2.  Material and Methods

Glycolytic/gluconeogenic enzyme amino acid sequences were used in BLASTP and PSI-BLAST searches of the non-redundant database (Altschul et al. 1997). The following databases were used: the Comprehensive Microbial Resource (www.tigr.org/tdb/mdb/mdbcomplete.html), COG (http://www.ncbi.nlm.-nih.gov/COG/), PEDANT (www.tigr.org/tdb/mdb/mdbcomplete.html), BRENDA

(www.brenda.bc.uni-koeln.de), conserved domain architecture retrieval tool (CD ART) and KEGG (http://www.genome. ad.jp/kegg/metabolism.html).

## 3. Results

The phylogenetic distribution of gluconeogenic/glycolytic enzymes from hyperthermophilic bacteria and archaea is summarised in Table 1. Included in the table are the reactions that are catalyzed, the common name of the enzymes, the enzyme classification numbers (EC numbers) and the number of gene sequence families that are required to explain the presence of the enzyme activities.

When glyceraldehyde-3-phosphate dehydrogenase (GAPDHs), triosephosphate isomerase (TPIs), phosphoglycerate kinase (PGKs) and enolase protein sequences were used in searches of the databases corresponding sequences for each enzyme were found to be widely distributed within the three domains, with all hyperthermophiles possessing related sequences. For enolases, there are two related genes present in some organisms (Pyrococcus species). For PGMs, the bacteria (Aquifex aeolicus and Thermotoga maritima) and some archaea (the acidophilic Thermoplasma acidophilum, Thermoplasma volcanium and Sulofolobus tokodaii; and the methanogens Methanosarcina acetivorans, Methanococcus jannaschii and Methanobacterium thermoautotrophicum) genes are present for both a cofactor-dependent phosphoglycerate mutase (dPGM) and a cofactor-independent PGM (iPGM). In others, only a single gene is present and, notably, this is always an iPGM-encoding gene sequence. Pyruvate kinase (PK), which catalyzes an irreversible glycolytic reaction, is not universally present. Phosphoenolpyruvate (PEP) synthase and pyruvate, phosphate dikinase (PPDK) gene sequences, which are reversible and related to each other, are present in all hyperthermophiles and mesophilic archaea.

In stark contrast to the near-universality of the three carbon transformation pathway enzymes, for all enzymes in the upper portion of the pathway, except possibly for aldolases, multiple sequence families are present (Table 1). For instance, a minimum of two sequence families is required to account for glucokinases/hexokinases (ADP- and ATP-dependent) and probably three in consideration of those present in eukaryotes (Ronimus & Morgan 2002). In addition, two sequence families are required to explain the presence of phosphoglucoisomerases (PGIs) and phosphofructokinases (Family A ATP- and pyrophosphate-dependent PFKs) and Family B/C PFKs (the Family B ATP- and Family C ADP-PFKs are suggested to be related; Ronimus & Morgan 2001). Some thermophilic archaea are obligate aerobes (T. acidophilum and T. volcanium) and utilise the Entner-Doudoroff (ED) pathway for processing glucose (as do Sulfolobus species) and therefore do not need a PFK (this reaction s bypassed in the ED pathway). Furthermore, three sequence families are present in organisms of Table 1 for fructose-1,6-bisphosphatases (FBPases), a gluconeogenic enzyme and normally a bypass reaction of PFK, and a fourth has been recently found in Bacillus subtilis (Ronimus & Morgan 2002). Fructose-1,6-bisphosphatase aldolases in organisms of Table 1, although still somewhat controversial (with many believing that there are essentially two classes of FBP aldolases with different reaction mechanisms), can be explained in a parsimonious manner by a

Table 1: Gene matrix table for hyperthermophilic prokaryotes and mesophilic archaea.

| gene families | enzyme | reaction | EC number | genes identified |
|---|---|---|---|---|
| (2) | GlcK | glucose+ATP → G-6-P+ADP | EC 2.7.1.1/2 | AE,TM,APE,PAE, HB,TA,TV PAB,PF,PH,MJ,MA |
| | GlcK | glucose+ADP → G-6-P+AMP | EC 2.7.1.147 | |
| (2) | PGI | G-6-P ↔ F-6-P | EC 5.3.1.9 | AE,TM,APE,PAE,HB,MJ,MTH,ST, SSO,TA,TV |
| | PGI (*Pyrococcus*-like) | | EC 5.3.1.9 | AF,PAB,PF,PH,MK,MA |
| (2) | PFK (FamA) | F-6-P +ATP → F-1,6-P$_2$ | EC 2.7.1.11/90 | AE,TM(2) |
| | PFK (FamB) | | EC 2.7.1.11 | APE,PAE,AF(2),HB |
| | PFK (FamC) | F-6-P +ADP → F-1,6-P$_2$ | EC 2.7.1.146 | PA,PF,PH,MJ,MA |
| (3) | FBPase | F-1,6-P$_2$ → F-6-P | EC 3.1.3.11 | HB,AE(2),TM,APE,PAE,AF,PAB,PF, PH,MJ(2),MK,MA,MTH,ST,SSO |
| | FBPase (*glpX*) | | EC 3.1.3.11 | MA |
| | FBPase (T. KOD1) | | EC 3.1.3.11 | AE,APE,PAE,AF,PAB,PF,PH,MJ,MK, MTH,ST,SSO,TA,TV |
| (2) | aldolase | F-1,6-P$_2$ ↔ DHAP+ G-3-P | EC 4.1.2.13 | AE,TM |
| | aldolase (DhnA-like) | | EC 4.1.2.13 | AE,APE,AF(2),PAB,PF,PH,HB(2), MJ(2),MK,MA(4),MTH,ST,SSO |
| (1) | TPI | DHAP ↔ G-3-P | EC 5.3.1.1 | all species have one |
| (1) | GAPDH | G-3-P+P$_i$+NAD ↔1,3-BPG+NADH | EC 1.2.1.12 | all species have one |
| (1) | PGK | 1,3-BPG+ADP ↔ 3-PG+ATP | EC 2.7.2.3 | all species have one |
| (2) | iPGM | 3-PG ↔ 2-PG | EC 5.4.2.1 | AE,TM,APE,PAE,AF(2),PAB,PF,PH,HB, MJ(2),MK,MA,MTH(2),ST,SSO,TA,TV |
| | dPGM | | EC 5.4.2.1 | AE(2),TM,MA,ST,TA(2),TV(3) |
| (1) | enolase | 2-PG ↔ PEP+H$_2$O | EC 4.2.1.11 | all species have one (PAB,PF and PH have 2) |
| (1) | PK | PEP+ADP → pyr+ATP | EC 2.7.1.40 | TM,APE,PAE,PAB,PF,PH,HB,MJ,MA, ST,SSO,TA,TV |

Table 1: *continued*

| gene families | enzyme | reaction | EC number | genes identified |
|---|---|---|---|---|
| (1) | PEP synthase | ATP+pyr+H$_2$O $\leftrightarrow$ AMP+PEP+P$_i$ | EC 2.7.9.2 | all species have either a PEP synthase or PPDK |
| | PPDK | ATP+pyr+P$_i$ $\leftrightarrow$ AMP+PEP +PP$_i$ | EC 2.7.9.1 | |

Species name abbreviations: *Aquifex aeolicus* (AE), *Thermotoga maritima* (TM), *Aeropyrum pernix* (APE), *Pyrobaculum aerophilum* (PAE), *Archaeoglobus fulgidus* (AF), *Pyrococcus abysii* (PAB), *Pyrococcus furiosus* (PF), *Pyrococcus horikoshii* (PH), *Halobacterium* sp. NRC-1 (HB), *Methanococcus jannaschii* (MJ), *Methanopyrus kandleri* (MK), *Methanosarcina acetivorans* (MA), *Methanobacterium thermoautotrophicum* (MTH), *Sulfolobus tokodaii* (ST), *Sulfolobus solfataricus* (SSO), *Thermoplasma acidophilum* (TA), *Thermoplasma volcanium* (TV). Abbreviations: GlcK, glucokinase; PGI, phosphoglucoisomerase; PFK, phosphofructokinase; FBPase, fructose-1,6-bisphosphatase; T. KOD1, Thermococcus KOD1; aldolase, fructose-1,6-bisphosphate aldolase; TPI, triosephosphate isomerase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase (phosphorylating); PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; PPDK, pyruvate, phosphate dikinase; G-6-P, glucose-6-phosphate; F-6-P, fructose-6-phosphate; F-1,6-P$_2$, fructose-1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; G-3-P, D-glyceraldehyde-3-phosphate; NAD, nicotinamide adenine dinucleotide; 2-PG, 2-phosphoglycerate; 3-PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; 1,3-BPG, 1,3-bisphosphoglycerate; pyr, pyruvate. Numbers in parentheses present in the genes identified column are the number of related genes (i.e. paralogous) present in that particular species.

single ancestral gene giving rise to all aldolases (Galperin, Aravind, & Koonin 2000; Ronimus & Morgan 2002).

## 4. Discussion

One of the most important implications from the data in Table 1 is that the enzymes of the lower portion of the pathway, or trunk pathway, are universally distributed in all species of the three domains. In other words, the activities of GAPDHs, TPIs, PGKs and enolases can each be accounted for by a separate single sequence family, i.e. all GAPDHs, for example, are related to each other and are derived from a single common ancestral gene. The same analogy holds for the other three lower universally distributed enzymes (TPIs, PGKs and enolases). Phosphoglycerate mutases (PGMs) are a special case but the cofactor-independent iPGMs are found in all hyperthermophiles suggesting that these were present in the last universal common ancestor (LUCA). In contrast, all of the enzymes of the upper portion of the pathway, involving transformations of six carbon sugars (Ronimus & Morgan 2001; Ronimus & Morgan 2002), are encoded by multiple gene families. The example of highest multiplicity is represented by fructose-1,6-bisphosphatses (FBPases) where three families are required to explain their ctivity amongst organisms of Table 1, and four amongst the three domains (Ronimus & Morgan 2002). Overall, the data strongly supports the view that the gluconeogenic pathway developed from the bottom-up. If the pathway had developed in the glycolytic direction one would expect single gene sequence families at the top of the pathway and increasing levels of diversity (multiplicity) in the lower portion of the pathway, but the opposite is the case.

An interesting aspect of the chemistry of reactants/substrates of the gluconeogenic pathway is the prevalence of carboxylic acid and phosphate groups, both negatively charged, which would have aided in their retention in the cytoplasm of cellular life. According to Wächterhäuser's theory (1988), the early development of metabolism is envisaged to have taken place on the positively charged surface of pyrite which would have acted to concentrate metabolites ensuring rapid rates of synthesis and an associated reduction in the rate of their degradation through hydrolysis. Additionally, the thermal stability of the substrates themselves is relevant to the discussion. For instance, several of the lower pathway substrates are significantly less stable than six carbon sugars of the upper portion of the pathway (the half-life of glyceraldehyde-3-phosphate is 3.4 min at 80°C whereas the half-life for fructose-1,6-phosphate is over 3 hours at 90°C; Ronimus & Morgan 2002). Pyruvate possesses more connections within the scheme of metabolism than any other metabolite and, interestingly, is not phosphorylated. Overall, the multiple inter-connecting reactions between phosphoenolpyruvate, pyruvate and oxaloacetate, and between pyruvate and acetyl-CoA, malate, acetyl-phosphate, methylglyoxal, acetate and the amino acids valine, leucine and isoleucine lend support to a pivotal role for these enzymatic reactions in the origins of metabolism. Their prevalence also supports a gluconeogenic origin for the glycolytic pathway. The combination of the high level of conservation of enzymes of the lower glycolytic pathway and their universal presence in all life on Earth supports the notion that metabolism would

develop in a similar manner in life on other planets. Evolutionary constraints (thermodynamics, chemistry of individual metabolites, thermal stability, etc.) throughout the development of life on Earth has ensured that the gluconeogenic pathway is central to metabolism and developed in the upwards direction, i.e. from 3-carbon to 6-carbon compounds.

## References

Altschul, S. F., Madden, T. L. , Schäffer, A. A. , Zhang, Z., Miller, W., & Lipman, D. J. 1997, Nucl. Acids Res., 25, 3389

Amend, J. P., & Shock, E. L. 2001, FEMS Microbiol. Rev., 25, 175

Baross, J. A. 1998,in Thermophiles: the keys to molecular evolution and the origin of life?, ed. J. Wiegel & M. W. W. Adams, (USA: Taylor and Francis Philadelphia), 3

Feng, D-F., Cho, G., & Doolittle, R. F. 1997, Proc. Nat. Acad. Sci., 94, 13028

Fothergill-Gilmore, L. A. 1986, Trends Biochem. Sci., 11, 47

Galperin, M. Y., Aravind, L., & Koonin, E. V. 2000, FEMS Microbiol. Lett., 183, 259

Maden, B. E. H. 1995, Trends Biochem. Sci., 20, 337

Madigan, M. T., Martinko, J. M., & Parker, J. 2000, Brock: biology of microorganisms, (New Jersey: Prentice Hall International)

Orgel, L. E. 1998, Trends Biochem. Sci., 23, 491

Reysenbach, A. L., & Shock, E. 2002, Science, 296, 1077

Ronimus, R. S., & Morgan, H. W. 2001, Extremophiles, 5, 357

Ronimus, R. S., & Morgan, H. W. 2002 Archaea, 1, 1

Schidlowski M. 1988, Nature, 333, 313

Schopf, J. W., Kudryavtsev, A. B., Agresti, D. G., Wdowiak, T. J., & Czaja, A. D. 2002, Nature, 416, 73

Wächterhäuser, G. 1988, Microbiol. Rev., 52, 452