

ISABELA JERÔNIMO BEZERRA DO Ó

**O PROTEOMA ESTRUTURAL ANTERIOR AO ÚLTIMO ANCESTRAL
UNIVERSAL COMUM E SUAS IMPLICAÇÕES PARA A EVOLUÇÃO
DAS PRIMEIRAS PROTEÍNAS**

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS

João Pessoa

2017

ISABELA JERONIMO BEZERRA DO Ó

**O PROTEOMA ESTRUTURAL ANTERIOR AO ÚLTIMO ANCESTRAL
UNIVERSAL COMUM E SUAS IMPLICAÇÕES PARA A EVOLUÇÃO
DAS PRIMEIRAS PROTEÍNAS**

Monografia apresentada ao Curso de Ciências Biológicas (Trabalho Acadêmico de conclusão de Curso), como requisito parcial à obtenção do grau de Bacharel em Ciências Biológicas da Universidade Federal da Paraíba.

Orientador: Sávio Torres de Farias

João Pessoa

2017

Catálogo na publicação
Biblioteca Setorial do CCEN/UFPB
Josélia Oliveira – CRB-15/113

B574p Bezerra do Ó, Isabela Jeronimo.
O proteoma estrutural anterior ao último ancestral universal comum e suas implicações para a evolução das primeiras proteínas / Isabela Jeronimo Bezerra do Ó. – João Pessoa, 2017.
51 p. : il. color.

Monografia (Bacharelado em Ciências Biológicas) – Universidade Federal da Paraíba.
Orientador: Prof.º Sávio Torres de Farias.

1. Bases da vida - tRNA. 2. Último ancestral universal comum.
3. Origem da vida. 4. Vias metabólicas. 5. Evolução de proteínas.
I. Título.

UFPB/BS-CCEN

CDU 577 (043.2)

ISABELA JERÔNIMO BEZERRA DO Ó

**O PROTEOMA ESTRUTURAL ANTERIOR AO ÚLTIMO ANCESTRAL
UNIVERSAL COMUM E SUAS IMPLICAÇÕES PARA A EVOLUÇÃO
DAS PRIMEIRAS PROTEÍNAS**

Monografia apresentada ao Curso de Ciências Biológicas (Trabalho Acadêmico de conclusão de Curso), como requisito parcial à obtenção do grau de Bacharel em Ciências Biológicas da Universidade Federal da Paraíba.

Data: _____

Resultado: _____

BANCA EXAMINADORA:

Sávio Torres de Farias, Doutor, UFPB

Gustavo Henrique Calazans Vieira, Doutor, UFPB

Pedro Cordeiro Estrela de Andrade Pinto, Doutor, UFPB

Patrício Adriano da Rocha, Doutor, UFPB

AGRADECIMENTOS

Primeiramente gostaria de agradecer ao apoio e dedicação do professor Sávio Torres de Farias, quem sugeriu o tema e me orientou durante o referente trabalho assim como nos últimos anos de minha graduação. Também gostaria de agradecer ao professor Pedro Cordeiro Estrela quem me orientou durante grande parte da minha graduação, notadamente seus auxílios, sugestões e instruções. À José Anderson Feijó e Livia Dias de Oliveira pela ajuda metodológica e sugestões também dadas pelos membros do Laboratório de Genética Evolutiva Paulo Leminski. Finalmente, agradeço aos demais professores e colegas que contribuíram direta ou indiretamente com o conteúdo e metodologia deste trabalho.

RESUMO

A origem e a evolução dos primeiros seres vivos é um grande desafio para a biologia moderna. A compreensão da constituição dos organismos no surgimento da vida é um passo importante para a compreensão da organização inicial do sistema biológico. Aqui, analisamos o proteoma estrutural anterior ao Último Ancestral Universal Comum (LUCA) baseado na reconstrução das sequências ancestrais e estrutura das proteínas envolvidas nas vias do carbono. Os resultados indicam que o sítio catalítico pode ter sido a primeira parte das proteínas iniciais, e possuíam funções básica de ligação em cofatores. Com o acréscimo das novas partes na estrutura, surgiu a função catalítica. Sugere-se também que os grupamentos estruturais iniciais pudessem participar da emergência das diferentes proteínas que atuam nos organismos modernos. A origem dos primeiros genes é discutida e é sugerido que a informação biológica inicial teve origem a partir de proto-tRNAs durante a formação do sistema de tradução primitiva.

Palavras-chave: tRNA; Último Ancestral Universal Comum; Origem da vida; vias metabólicas; evolução de proteínas.

ABSTRACT

The origin and evolution of the first living beings is a great challenge for modern biology. The comprehension of the constitution of organisms on the emergence of life is an important step to understand the initial organization of the biological system. Here, we analyzed the structural proteome before the Last Universal Common Ancestor (LUCA) based in the reconstruction of the ancestral sequences and structures of proteins involved in the carbon pathways. The results indicate that the catalytic site could have been the first part of these initial proteins, which carried basic function such as binding to cofactors. With the accretion of the new parts in the structure, the catalytic function emerged. Also it is suggested that the initial structural motifs could have participated in the emergence of the proteins that work in the modern organisms. The origin of the first genes is discussed and it is suggested that the initial biological information had origin from proto-tRNAs during the formation of the primitive translation system.

Keywords: tRNAs; Last Universal Common Ancestor; Origin of life; Metabolic pathways; Protein Evolution.

LISTA DE FIGURAS

Figura 1 - “Estruturas de sequências ancestrais obtidas pela abordagem Bottom Up. ”...página 21	
Figura 2 - “Estruturas para sequências ancestrais obtidas pela abordagem Top Down completa deleção (CD) e 99% de cobertura do local (PD99)”.....página 22	
Figura 3 - “Estruturas para sequências ancestrais obtidas por abordagem Top Down com 95% de cobertura do local”página 23	
Figura 4 - “Distância filogenética baseada no valor de RMSD”.....página 24	
Figura 5 - “Passos evolutivos propostos das estruturas. Alinhamento feito usando o servidor TM-Align.”página 25	

LISTA DE TABELAS E QUADROS

Tabela 1.....	página 26
Tabela 2.....	página 27
Tabela 3.....	página 28

LISTA DE ABREVIATURAS E SIGLAS

LUCA: Last Universal Common Ancestral (Último Ancestral Universal Comum)

DNA: Deoxyribonucleic Acid (Ácido Desoxirribonucleico)

RNA: Ribonucleic Acid (Ácido Ribonucleico)

tRNA: Transfer Ribonucleic Acid (Ácido Ribonucleico de Transferência)

TCA: Tricarboxylic Acids (Ácidos Tricarboxílicos)

RNY: “R” para purina, “N” para qualquer um dos quatro nucleotídeos, ”Y” para pirimidina.

GC: Guanina-Citosina

NCBI: *National Center for Biotechnology Information* (Centro Nacional de Informação Biotecnológica)

RMSD: Root-Mean-Square Deviation (Raiz do Desvio Quadrático Médio)

PDB: Protein Data-Base (Banco de Dados de Proteínas)

GOE: *Great Oxidation Event* (Grande evento de oxidação)

SUMÁRIO

INTRODUÇÃO.....	1
1 FUNDAMENTAÇÃO TEÓRICA.....	1
2 REFERÊNCIAS.....	4
3 ARTIGO CIENTÍFICO.....	7
Introdução.....	8
Material e Métodos.....	11
Resultados e Discussão.....	13
Conclusão.....	18
Referências.....	19
Figuras, Tabelas e Legenda.....	21
4 CONCLUSÃO.....	30
ANEXOS.....	32

INTRODUÇÃO

1. FUNDAMENTAÇÃO TEÓRICA

O éon Hadeano deu início à história da Terra como planeta. A desconfiança de sua relevância, o escasso registro geológico dessa época (Lunine, 2006) reflete a complexidade de responder perguntas como o surgimento da abundância hídrica no planeta ou, ao que é assunto nesta monografia; a origem da vida terrestre. Apesar do pouco conhecimento sobre esse tempo, há evidências (Halliday, 2000) sobre gigantescos impactos ocorridos através da colisão de corpos que chegavam ao tamanho de Marte contra a superfície terrestre. Esses fenômenos limitavam a existência de vida a uma probabilidade ínfima a ponto de podermos considerar que a vida como conhecemos muito provavelmente teve origem após o último grande impacto (Sleep, 1989). Posteriores impactos foram menores e, mesmo destruindo parte da biota local próxima à colisão, criavam uma possibilidade de sucessão ecológica (Fajardo-Cavazos, 2005) e alguns autores (Cockell, 2006) sugerem que poderiam ter criado ambientes pós-impactos particularmente favoráveis à origem da vida uma vez que formavam grandes áreas de superfície para reação e também introduziam novos reagentes como minerais relevantes para a manutenção da vida.

O impacto de Theia contra a Terra, o último grande impacto, culminou em uma elevação na densidade atmosférica formando um efeito estufa sobre todo o planeta. Esse fator, junto com outras características deste período, como a atividade vulcânica, resultou nas possíveis altas temperaturas (Kasting, 2006 and Pavlov, 1990) que podem ser evidenciadas através das taxas de $^{18}O/^{16}O$ em quartzos anteriores a 3,5 bilhões de anos. Essas altas temperaturas, a provável atmosfera e as prováveis moléculas existentes nesse período restringiam geoquimicamente a amplitude de moléculas orgânicas, o que eventualmente repercutiu na origem das primeiras vias metabólicas.

De acordo com Alves et al (2002), no início da evolução do metabolismo, as enzimas provavelmente eram escassas e essas possuíam baixa efetividade e especificidade, portanto possuíam capacidades catalíticas reduzidas. No livro “Functional Metabolism: Regulation and Adaptation” (2004), é apontado que uma forma de catálise é dada através da associação física que pode ser dada diretamente entre as moléculas em questão ou pelo meio de um polímero estabilizante. Aqui neste trabalho, estamos considerando a possibilidade das primeiras moléculas catalíticas terem agido dessa forma. Um hiperciclo incipiente do “universo dos polímeros” (chamado por Storey de “polymer world”) teria sido vantajoso sobre outros hiperciclos dado a capacidade de produção de andaimes poliméricos longos por servirem para

armazenamento de energia e matéria prima. Esse sistema alcançaria altos níveis de eficiência catalítica ao passo que as ligações ao “andaime” se tornassem mais prováveis quimicamente. Podemos observar ainda hoje possíveis rastros da funcionalidade de “andaimes”; o citoesqueleto de eucaliptos é um exemplo de um sistema existente de andaime; organiza componentes celulares no espaço e permite a existência de sistemas cooperativos maiores e mais complexos (Storey, 2004).

As primeiras vias metabólicas provavelmente possuíam características restringidas pelas limitações dos períodos geológicos as quais surgiram. Deste modo, é esperado que as vias modernas que possuem histórias evolutivas antigas transpareçam essa antiguidade através de determinados atributos. Registros apontam (Dutkiewicz et al, 2006) que a vida surgiu muito antes do Grande Evento de Oxidação (GOE). Da mesma forma que é lógico supor que elementos oxidados fossem de inexistentes a no máximo raros antes do GOE, esperamos que as primeiras funções biológicas independam de oxigênio. Um caso paralelo é a relação entre as vias metabólicas e membranas biológicas; consideramos que estas surgiram também posteriormente às primeiras reações químicas que originaram as primeiras vias metabólicas homólogas às vias atuais. Além dessas restrições esperamos que as primeiras funções biológicas estejam espalhadas filogeneticamente pelos domínios da vida.

Existe uma lista de vias que atendem às particularidades explanadas anteriormente, limitando as vias que podem ter surgido anterior à primeira célula e próximo à origem da vida. Resultados encontrados por Farias et al (2016) seguem o esperado de acordo com o exposto no parágrafo anterior. Os autores encontraram as seguintes vias em seus resultados como sendo vias anteriores ao Último Ancestral Universal Comum: Vias de aminoácidos, Vias do Carbono, Vias de lipídeos, Vias de nucleotídeos, Transcrição e Tradução ou Replicação de RNA. Nesta monografia, analisamos o encontrado referente as Vias do Carbono. Nessa notamos que as enzimas encontradas estão presentes principalmente nas atuais Glicólise e Gliconeogênese.

A Glicólise é a sequência de reações que metaboliza uma molécula de glicose em duas moléculas de piruvato com a concomitante produção, através da energia livre, de duas moléculas de adenosina trifosfato. Esse é um dos processos metabólicos mais universais conhecidos, ocorrendo em diversos tipos de células e, pelo menos parcialmente, em todos os organismos. As sequências de aminoácidos das enzimas que fazem a Glicólise são muito bem conservadas ao longo de todos os domínios da vida, sugerindo que provavelmente houve uma origem comum. (Freeman, 2002)

Glicose é um dos monossacarídeos formados a partir de formaldeído sob condições pré-bióticas, o que torna possível sua disponibilidade como uma fonte de combustível para os

sistemas bioquímicos primitivos, ou seja, a glicose pode ter sido uma das primeiras fontes de energia para os sistemas vivos (Freeman, 2002). Adicionalmente, a Glicólise possui uma estrutura simples, dado que as enzimas agem independentemente e os passos da via não dependem de membrana. Não obstante à sua simplicidade, produziria energia suficiente a uma célula primitiva. Além disso, o fato das reações glicolíticas dependerem de oxigênio as tornam possíveis na terra primitiva quando as porcentagens de oxigênio eram baixas. Esses são alguns dos principais pontos que sugerem que essa via tenha surgido bastante cedo na história evolutiva da terra, provavelmente anterior ao surgimento da Última Célula Ancestral Comum (Storey, 2005).

A Glicólise provavelmente teve como função primordial a síntese de hexoses, esses monossacarídeos que eram escassos nesse ponto da história da terra são e provavelmente eram importantes e comuns fontes de energia (Freeman, 2002). Provavelmente, a Gliconeogênese, a rota de produção da glicose a partir de compostos aglicanos, teria surgido previamente à Glicólise (Silva, 2009), agindo posteriormente como provedor da glicose utilizado nessa via mais tardia. Isso pode ser embasado com o fato da Gliconeogênese ser ainda mais difundida que a Glicólise nos domínios da vida. Além disso, sete dos dez passos da Glicólise são reversíveis, as enzimas podem funcionar na via anabólica e catabólica (Lubert, S 1995), com a exceção de três passos entre os intermediários formados através da Glicólise e da Gliconeogênese, todas reações são palíndromos.

Apesar do amplo conhecimento sobre as vias glicolítica e gliconeogênica, ainda há muito para ser explorado sobre a origem dessas vias. Nesta monografia entendemos que a seleção positiva em organismos vivos para existência dessas estruturas bioquímicas teria sido dada pela pressão positiva do aumento da velocidade de reação que as enzimas possibilitavam. O estado mais primitivo dessas estruturas teria sido o de polímeros de ancoramento (andaimes), onde os intermediários se ligariam a essas enzimas ancestrais e graças à proximidade física a probabilidade de interação entre as moléculas aumentaria.

Algumas das reações dos intermediários das vias do carbono (presentes nas vias da Glicólise e da pentose fosfato) reagem independentemente da existência dessas enzimas. Keller et al (2014) forneceram evidências que a estrutura das vias da Glicólise e da pentose fosfato poderiam ser fundamentalmente formadas a partir do que se aceita ser o meio químico estabelecido no oceano primitivo. Então, em condições semelhantes aquelas aceitas como pré-bióticas os intermediários dessas vias sofrem conversões espontâneas com resultados próximos aqueles encontrados nas vias modernas. Sabendo disto, supomos que a pressão seletiva para as enzimas teria sido apenas para acelerar essas reações que já aconteciam.

2. REFERÊNCIAS

1. "Chapter 16 Glycolysis and Gluconeogenesis" Freeman, W. H. 2002. *NCBI*
2. "Functional Metabolism: Regulation and Adaptation" Storey, K.B. 2005. *John Wiley & Sons*
3. "The Chemical Logic Behind Gluconeogenesis" Silva, P. 2009. *Retrieved*
4. "Biochemistry" (Quarta edição). Lubert, S 1995. *W.H. Freeman and Company*
5. "Functional Metabolism: Regulation and Adaptation" Storey, K.B. 2004. *John Wiley and Sons*;
6. "O Atualismo entre uniformitaristas e catastrofistas". Farias F. 2014 *Revista da Sociedade Brasileira de História da Ciência*;
7. "Origin and evolution of metabolic pathways" Fani, R; Fondi, M. 2009. *ScienceDirect*;
8. "Curso Online Introdução a bioinformática" F, Prodocini. 2007
9. "Evolution of enzymes in metabolism: a network perspective" - Alves, R; Chaleil, R.A.; Sternberg M.J. 2002. *PubMed*;
10. "Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean "Keller, M.A.; Turchyn, V.T.A.; Ralser, M. 2014. *Molecular Systems Biology*;
11. "Enzyme recruitment in evolution of new function" Jensen, R. A. 1976 *Microbiology*;
12. "EWIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction" Overbeek, R., Larsen, N., Pusch, G. D., D'Souza, M., Selkov, E., Kyrpides, N. 2000. *Nucleic Acids Research*;
13. "My Name is LUCA—The Last Universal Common Ancestor" Poole, A. 2002. *Actionbioscience*;
14. "Changing perspectives on the origin of eukaryotes." Katz, L.A. 1998. *Trends Ecology & Evolution*;
15. "Were bacteria the first forms of life on Earth?" Jeffares, D.C.; Poole, A.M.2000. *Actionbioscience*;
16. "A Study of Enzymes – Volume 2" Kuby, S.A. 1990. *CRC Press*;
17. "Origins of life: Common ancestry put to the test" Steel M.; Penny D. 2010. *Nature*;
18. "Physical conditions on the early Earth." Lunine, J. I.2006. *Philosophical Transactions of the Royal Society B: Biological Sciences*;
19. "Annihilation of ecosystems by large asteroid impacts on the early Earth." Sleep N.H, Zahnle K.J, Kasting J.F, Morowitz H.J. 1989 *Nature*;

20. "The origin and emergence of life under impact bombardment." Cockell C. 2006 *Philosophical Transactions of the Royal Society B: Biological Sciences*;
21. "Bacillus subtilis spores on artificial meteorites survive hypervelocity atmospheric entry: implications for lithopanspermia." Fajardo-Cavazos P, Link L, Melosh H.J, Nicholson W.L. 2005 *Astrobiology*;
22. "Atmospheric composition and climate on the early Earth." Kasting J.F. 2006 *Philosophical Transactions of the Royal Society B: Biological Sciences*;
23. "Greenhouse warming by CH₄ in the atmosphere of early." Pavlov A.A, Kasting J, Brown L.L, Rages K.A, Freedman R. 1990. *Journal of Geophysical Research*;
24. "Origin and diagenesis of cherts: an isotopic perspective." Knauth L.P. 1992. *Isotopic signatures and sedimentary records*;
25. "Terrestrial accretion rates and the origin of the Moon". Halliday, A. N. 2000. *Earth and Planetary Science Letters*;
26. "Biomarkers from Huronian oil-bearing fluid inclusions: an uncontaminated record of life before the Great Oxidation Event". Dutkiewicz, A.; Volk, H.; George, S. C.; Ridley, J.; Buick, R. 2006. *Geology*;
27. "The Natural History of Oxygen". Dole, M. 1965." *The Journal of General Physiology*.

3. ARTIGO CIENTÍFICO

O proteoma estrutural anterior ao último ancestral universal comum e suas implicações para a evolução das primeiras proteínas

Isabela Jerônimo Bezerra do Ó¹, Thais Gaudêncio Rego², Marco V. José³ and Sávio Torres de Farias¹

¹Laboratório de Genética Evolutiva Paulo Leminsk, Departamento de Biologia Molecular, Universidade Federal da Paraíba, João Pessoa, Brazil.

²Departamento de Informática, Universidade Federal da Paraíba, João Pessoa, Brazil

³Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad de México CDMX, C.P. 04510

Resumo

A origem e a evolução dos primeiros seres vivos é um grande desafio para a biologia moderna. A compreensão da constituição dos organismos no surgimento da vida é um passo importante para a compreensão da organização inicial do sistema biológico. Aqui, analisamos o proteoma estrutural anterior ao Último Ancestral Universal Comum (LUCA) baseado na reconstrução das sequências ancestrais e estrutura das proteínas envolvidas nas vias do carbono. Os resultados indicam que o sítio catalítico pode ter sido a primeira parte das proteínas iniciais, e possuíam funções básica de ligação em cofatores. Com o acréscimo das novas partes na estrutura, surgiu a função catalítica. Sugere-se também que os grupamentos estruturais iniciais pudessem participar da emergência das diferentes proteínas que atuam nos organismos modernos. A origem dos primeiros genes é discutida e é sugerido que a informação biológica inicial teve origem a partir de proto-tRNAs durante a formação do sistema de tradução primitiva.

Palavras-chave: tRNA; Último Ancestral Universal Comum; Origem da vida; vias metabólicas; evolução de proteínas.

Introdução

No seu livro *Sobre a Origem das Espécies por Meio da Seleção Natural ou a Preservação de Raças Favorecidas na Luta pela Vida*, (1859) Charles Darwin sugeriu um ancestral comum para todos os organismos vivos. O conceito de Último Ancestral Universal Comum (LUCA) desencadeia até hoje discussões sobre os componentes desta entidade biológica (Penny and Poole, 1999; Forterre et al., 2005; Mushegian, 2008; Glansdorff et al., 2008; Kim and Caetano-Anollés, 2011; Goldman et al., 2013).

LUCA refere-se a um organismo já complexo, com processamento de informação desenvolvido, metabolismo intrincado e um genoma baseado em DNA (Di Giulio, 2003). Assim, LUCA não representa a primeira forma de vida, porque os processos evolutivos tinham começado muito antes do DNA. Carl Woese (1998) sugeriu a existência de comunidades de “progenotes” como precursores do LUCA. Os progenotes eram organismos delimitados por membranas com poucas funções metabólicas. Nesse cenário, esses organismos ou entidades biológicas tinham como molécula informacional o RNA e um processo de tradução impreciso, gerando proteínas estatísticas mas que já estariam sob a influência da seleção natural. Podemos descrever os progenotes como uma comunidade de subsistemas que se desenvolveram separadamente, mas evoluíram de forma que, em determinado momento de desenvolvimento, foi estabelecido um estágio mais elevado de dependência entre eles, que culminou na congregação desses subsistemas (Woese, 1998).

É lógico supor que nem todos os passos das modernas vias metabólicas estavam presentes em um progenote, e que diferentes etapas do metabolismo apareceram em diferentes progenotes. A capacidade de traduzir certos tipos de informação foi provavelmente uma das primeiras propriedades, se não a primeira, a se desenvolver entre os progenotes que antecederam a vida conhecida (Eigen and Schuster, 1978; Woese, 1998). De acordo com Weiss et al (2016),

havia vestígios de metabolismo de carbono, nitrogênio e energia no genoma do LUCA. O metabolismo de carbono estava mais provavelmente presente nas formas mais antigas de vida como conhecemos hoje, e isso não só pela sua vasta presença em todos os domínios da vida, mas também pela sua importância na geração de energia e precursores de sistemas biológicos. Das vias dos carboidratos os dois mais universais e bem preservados são a Glicólise / Gliconeogênese e o Ciclo do Ácido Cítrico.

Glicólise / Gliconeogênese está presente, pelo menos parcialmente em todas as formas de vida, e toda a vida celular traz variações do metabolismo de carboidratos. Ao comparar formas alternativas de Glicólise e ciclo do TCA em uma grande variedade de táxons, o núcleo dessas vias e suas porções mais prováveis que podem estar presentes nos organismos mais antigos podem ser elucidadas. A antiguidade dessas vias, especificamente Glicólise / Gliconeogênese, é sugerida por uma série de características, como sua simplicidade estrutural, o que significa que as enzimas atuam independentemente e não necessitam de membrana, o que seria uma característica necessária em um oceano Arqueano em uma realidade pré-membrana. Além disso, é independente do oxigênio, outra característica necessária nesse período.

Adicionalmente, os resultados das duas reações de fosforilação ao nível do substrato proporcionariam energia suficiente para o metabolismo primitivo. O piruvato como receptor terminal de elétron mantém o equilíbrio redox da via. Outra razão para sustentar a ocupação desse papel anterior ao LUCA, é que provavelmente a Glicólise evoluiu como uma Gliconeogênese reversa. A síntese de açúcares de hexose atuaria como armazenamento de energia e precursores biológicos, e seria vantajoso num cenário primitivo com raras moléculas complexas. Neste momento do tempo evolutivo, não era possível uma acumulação intensa de açúcar. No entanto, a Glicólise não é apenas responsável pela produção de energia, mas seus compostos de três carbonos também funcionam como fonte de energia. Estes produtos são utilizados numa série de outras vias, tais como o anabolismo de aminoácidos. Este foi

provavelmente o primeiro papel da Glicólise, que é apoiado por esta ser constante em todas as formas de vida, enquanto o metabolismo energético varia.

Foi provavelmente vantajoso para um organismo ser capaz de armazenamento de carbono, uma vez que aumentaria sua independência em relação ao sistema circundante. Isto expandiria a probabilidade de sobreviver em um ambiente incerto, dado eventos estocásticos que diminuiriam a disponibilidade de nutrientes.

Nos primeiros estágios evolutivos do metabolismo, como conhecemos hoje, as grandes moléculas informacionais, possivelmente não tinham estabilidade suficiente, e os primeiros dispositivos de codificação devem ter refletido as restrições químicas de sua origem. Eigen e Winkler-Oswatitsch (1981) propuseram que os eventos moleculares iniciais foram realizados por tRNAs, principalmente porque: (1) Têm predominância de emparelhamento GC, necessário para a estabilidade relativa; (2) sua estrutura secundária é estável, aumentando sua resistência à decomposição hidrolítica; (3) Os códons podem atuar como portadores de mensagem; (4) possuem cerca de 50 a 100 resíduos, ideal para condições de polimerização pre-biótica; (5) a árvore filogenética tem origem antiga; (6) possui uma estrutura simétrica.

Eigen e Schuster (1978) sugeriram que os primeiros anticódons eram do tipo RNY. Estes são códons onde a primeira base é uma purina, a terceira uma pirimidina, e a segunda é qualquer um deles. Assim, os primeiros módulos eram ricos em Alanina, Serina, Treonina, Asparagina, Ácido Aspártico, Valina, Isoleucina e Glicina. Farias et al (2016) reconstruíram a sequência ancestral de tRNAs que podem ter atuado como os primeiros genes desses progenotes. Eles propuseram um proteoma para os progenotes baseados na tradução de tRNAs ancestrais com um padrão RNY. Entre as vias metabólicas encontradas em organismos modernos, algumas já estavam possivelmente presentes no LUCA e nos progenotes.

Farias et al (2016), analisando as similaridades de tRNAs traduzidos com estruturas modernas, encontraram similaridade com as seguintes vias: Vias de aminoácidos, Vias do

carbono, Vias de lipídeos, Vias de nucleotídeos, Transcrição e Tradução ou Replicação de RNA. Aqui utilizamos os dados obtidos por Farias et al (2016) e suas sugestões para o metabolismo precoce para reconstruir dois estágios diferentes de evolução. Um (Bottom Up) anterior ao Último Ancestral Universal Comum no qual é baseado no proteoma estabelecido pelos ancestrais de tRNA e um outro (Top Down) baseado na reconstrução de sequências de ancestrais de proteínas modernas envolvidas no metabolismo de Glicólise / Gliconeogênese e compostos de três carbonos. Comparamos ambas as estratégias para compreender a história evolutiva destas vias metabólicas.

Materiais e Métodos

Abordagens de reconstrução de ancestrais

Duas abordagens diferentes foram usadas para reconstruir ancestrais.

Abordagem Bottom Up

O primeiro, “Bottom Up” refere-se às reconstruções baseadas nas sequências encontradas por Farias et al (2016). O proteoma sugerido por eles foi feito usando sequências ancestrais de tRNAs seguindo o padrão RNY, que quando traduzidos apresentam correspondências com proteínas modernas envolvidas no metabolismo de três carbonos.

Abordagem Top Down

A outra abordagem foi uma reconstrução usando sequências das proteínas modernas escolhidas com base no proteoma ancestral sugerido por Farias et al (2016) que foram as seguintes: Gliceratoquinase, Triose fosfato isomerase, Glicose-6-fosfato-1-dehidrogenase, Glicose-6-fosfato-isomerase, Fosfogliceratoquinase, Glicerol-3-fosfato-dehidrogenase e Transquetolase.

Reconstrução de Sequência Ancestral

Obtivemos 400 sequências para cada tipo de enzimas que fazem parte da Glicólise / Gliconeogênese e da via dos três carbonos no NCBI (National Center for Biotechnology Information). As sequências abrangiam os três domínios da vida.

As sequências foram alinhadas com o suporte do software MAFFT v.7 (Kato and Standley, 2016). Para alinhar as sequências, utilizamos BLOSUM45 para a matriz de pontuação e todos os outros parâmetros permaneceram como padrão. Para reconstruir a sequência ancestral mais provável, utilizamos máxima verossimilhança com o suporte do software MEGA7 (Kumar et al., 2015). Para essa abordagem, utilizamos diferentes níveis de rigor para inferir a sequência de ancestrais por máxima verossimilhança, foram utilizados deleção completa, 95% e 99% de cobertura de cobertura.

Análise Estrutural

As estruturas das proteínas foram reconstruídas por homologia utilizando o servidor web I-Tasser (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (Zhang, 2008). As estruturas foram refinadas através do programa ModRefiner (<http://zhanglab.ccmb.med.umich.edu/ModRefiner/>) (Xu, 2010). Os modelos de proteínas modernas e os ligantes foram obtidos na Protein Data Base (www.rcsb.org) (Berman et al., 2000). Os alinhamentos estruturais foram realizados utilizando o software TM-Align (<http://zhanglab.ccmb.med.umich.edu/TM-align/>) (Zhang and Skolnick, 2005). Para a análise dos ligantes às estruturas ancestrais em ambas as abordagens (Top Down e Bottom Up), nós utilizamos o software Coach (<http://zhanglab.ccmb.med.umich.edu/COACH/>) (Yang et al., 2013).

Análise Filogenética

Construímos uma matriz com o valor de RMSD (a raiz do desvio quadrático médio das posições atômicas) calculado entre cada uma das estruturas calculadas assim como a estrutura moderna mais semelhante com o ancestral Top Down 95%.

Utilizamos SeaView (Gouy et al., 2010) para calcular a árvore de Neighbour joining baseada nas distâncias de Mahalanobis (Huberty, 2014) com a matriz supracitada.

Resultados e Discussão

A origem e a evolução das primeiras vias metabólicas ainda são um desafio relevante para a biologia moderna uma vez que poderíamos compreender o surgimento da vida e sua diversificação a partir desses estudos. Dessa forma, entender o seu núcleo é fundamental. No centro do metabolismo moderno estão as vias de Glicólise / Gliconeogênese, que são conservadas em todos os domínios da vida e seu desenvolvimento desempenhou um papel fundamental no metabolismo de carbono em células modernas. Foi sugerido que a Glicólise pode ter-se originado como um mecanismo que extraiu energia e precursores biológicos de aminoácidos formados abioticamente (Kenneth et al., 2004). Sutherland (2017) indicou a importância dos compostos com três carbonos na emergência das primeiras vias metabólicas. Farias et al (2016) propuseram que esses compostos funcionassem como um centro de distribuição de carbono para outras vias metabólicas em desenvolvimento. Farias et al. (2016) reconstruíram as sequências ancestrais de tRNAs e compararam as sequências ancestrais traduzidas com proteínas modernas. Os resultados sugeriram que as enzimas do metabolismo Glicólise / Gliconeogênese poderiam ter surgido a partir dos proto-tRNAs, assim como, os primeiros genes podem ter surgido destes, e no núcleo do processo estariam as proteínas

envolvidas no metabolismo de compostos com três carbonos. Aqui, chamamos essa abordagem de Bottom Up, onde as estruturas são derivadas de sequências traduzidas do ancestral tRNA.

Para entender a história evolutiva estrutural das enzimas da Glicólise / Gliconeogênese, utilizamos uma segunda abordagem baseada na reconstrução das sequências ancestrais e estruturas das enzimas modernas envolvidas na Glicólise / Gliconeogênese e três compostos de carbono (abordagem Top Down). Usamos apenas as mesmas enzimas que tinham correspondências com as sequências traduzidas do ancestral tRNA (ver Materiais e Métodos), a fim de fazer comparações entre as estruturas obtidas a partir de diferentes abordagens. A partir das sequências ancestrais obtidas através de enzimas modernas, utilizamos deleção completa, 95% e 99% de cobertura de cobertura local, o que pode captar diferentes etapas na história evolutiva dessas enzimas.

Reconstrução de estruturas ancestrais

Na fig. 1, é mostrado as estruturas reconstruídas por homologia das sequências de ancestrais de tRNA traduzidas na abordagem Bottom Up. Em geral, as estruturas são compostas por grupamentos simples, basicamente, loops e pequenas alfa-hélices. Na Tabela 1, são mostradas as estruturas modernas mais semelhantes em termos de valor de RMSD entre ancestrais por abordagem Bottom Up e estruturas modernas, e os organismos dos quais essas estruturas pertencem. Esses resultados estão de acordo com as funções básicas para proteínas muito precoces no início dos sistemas biológicos, quando funções catalíticas mais sofisticadas ainda estavam em desenvolvimento. As proteínas iniciais eram muito simples e tinham como principal atividade a ligação de alguns substratos com baixa afinidade. As semelhanças com proteínas que não estão envolvidas no metabolismo de carbono podem indicar que esses

grupamentos estruturais básicos funcionaram como blocos de montagem na origem das primeiras proteínas e genes. As estruturas modernas mais semelhantes possuem funções biológicas básicas tais como hidrólise, domínio de ligação a RNA, enolase, com valores de RMSD variando de 1,46 a 2,67. Esses resultados podem indicar que os blocos estruturais básicos participaram na formação de diversas vias, e a combinação dessas partes estruturais foram relevantes para a diversificação primordial de vias essenciais e para o estabelecimento e manutenção dos sistemas biológicos emergentes.

Na fig. 2, são mostradas as estruturas construídas por homologia pela abordagem de Top Down com deleção completa e corte de cobertura de 99% no local e na Fig. 3 com 95% de cobertura de cobertura do local. A comparação entre as estruturas obtidas das abordagens Top Down e de Bottom Up indica que as estruturas da abordagem de Down exibem uma estrutura mais complexa. Na Tabela 2, são mostradas as estruturas modernas mais semelhantes com valores de RMSD entre ancestrais por abordagem Top Down e proteínas modernas. Estes resultados refletem as limitações do método de reconstrução de sequências ancestrais de proteínas modernas, onde é possível identificar os primeiros domínios funcionais, mas não a evolução profunda desses domínios, antes do estabelecimento da especificidade moderna. No entanto, estes resultados também mostram a eficiência do método para compreender a história evolutiva das proteínas até o primeiro domínio funcional. Nossos resultados mostram que as estruturas ancestrais sugeridas são semelhantes às estruturas de diferentes extremófilos que vivem em ou habitats ácidos ou em altas temperaturas, ou ambos. Isto pode indicar que estas primeiras enzimas antes e depois do LUCA foram adaptadas a cenários extremos. Sugere-se que o impacto de Theia contra Terra criou energia suficiente para derreter o manto da Terra. A água condensada formou uma atmosfera densa obstruindo o calor para sair o que formou um efeito estufa (Nisbet et al., 2007). Mesmo que ainda haja um debate, diferentes evidências parecem

mostrar que a Terra pode ter atingido temperaturas extremamente altas no éon Hadeano e isso pode ter influenciado a evolução biológica (Sleep et al., 2007; Zahnle et al., 2007).

Análise de ligação de substratos

Na Tabela 3, é mostrada a análise da ligação do substrato para as estruturas ancestrais obtidas para as abordagens Bottom Up e para Top Down. Em relação a abordagem Bottom Up, os resultados da ligação de substrato indicam que estas estruturas podem interagir com compostos simples que podem ter funcionado como cofatores para funções básicas. Algumas das estruturas ligam com aminoácidos, a maioria deles ligam a coenzimas que têm o potencial de transferência de energia, e todos eles ligam com hidrocarbonetos pequenos. A interação entre as estruturas ancestrais e os nucleotídeos também sugere que a relação coevolutiva entre esses compostos foi estabelecida anteriormente no ambiente primordial. Caetano-Anolles et al (2012) sugeriu que na origem das vias metabólicas iniciais, o aparecimento progressivo de complexidade era a regra, e que as primeiras proteínas ou peptídeos trabalharam ligando vários cofatores, e o acréscimo de novas partes tornou possível essas proteínas primitivas operarem em catálise e funções moleculares. Tal como nos sistemas biológicos modernos, o modelo sugere que os ácidos nucleicos também funcionaram como cofatores no início da vida.

Para as estruturas obtidas através da metodologia Top Down com completa deleção ou com corte de cobertura de 99% podemos observar um padrão similar ao encontrado pela abordagem Bottom Up em relação aos ligantes mais prováveis. Para as estruturas obtidas pela abordagem com corte de cobertura de 95% os substratos mais prováveis foram aqueles similares as estruturas homólogas modernas. Esses resultados sugerem que ao longo da história evolutiva, o acréscimo de novas partes estruturais na proteína tornou as estruturas mais complexas e com mais especificidade o que tornou possível uma estabilização das primeiras vias metabólicas.

Distância evolutiva das estruturas ancestrais e passos evolutivos

Reconstruímos uma filogenia baseada em distância entre todas as estruturas utilizando o valor de RMSD. Foram comparadas as estruturas ancestrais de ambas as abordagens, Top Down e Bottom Up, bem como, as proteínas modernas. O resultado pode ser visto na Fig. 4. O padrão obtido na árvore de distância reforça o sugerido de que na origem das vias metabólicas os primeiros grupamentos estruturais funcionaram como blocos de montagem e com o acréscimo de novos grupamentos as proteínas ganharam novas funções e especificidade. Embora quando analisadas, as sequências para cada linhagem evolutiva exibam mais semelhanças entre seus homólogos, a estrutura mostrou um padrão difuso de similaridades, sugerindo que na origem a combinação de grupamentos estruturais simples foi utilizada para organizar diferentes proteínas, sendo possível que todas as proteínas do metabolismo de carbono tenham tido um ancestral único. Na Fig. 5, é mostrada a reconstrução dos passos evolutivos. Os dados sugerem que a porção inicial das proteínas eram grupamentos estruturais simples próximos ao sítio catalítico moderno e durante o acréscimo evolutivo de novos grupamentos a pressão seletiva agiu de sorte a proteger e isolar o sítio catalítico, o que foi tornando as proteínas mais específicas. Alves et al. (2002) mostraram que blocos funcionais de características químicas semelhante evoluíram dentro das redes metabólicas. Os ancestrais Bottom Up são todos estruturalmente semelhantes, mais do que em relação aos seus homólogos. Estes ancestrais também tinham substratos semelhantes, mas ligados a uma variedade de ligantes. Isso poderia indicar que eles tinham originado de um grupamento comum que tinha uma ação não específica. Estas primeiras moléculas "*enzyme-like*" poderiam ter agido em diferentes pontos do metabolismo e, ao longo da evolução, ganharam especificidade. Além de ganhar especificidade, a estrutura das enzimas tornou-se maior e mais complexa. A evolução provavelmente ocorreu em torno do sítio catalítico seguindo o modelo de acréscimo. Sustentamos que a primeira ação destas enzimas

relacionadas com a Glicólise foi efetivamente na direção gluconeogênica. A síntese de hidrocarbonetos teria sido positivamente selecionada uma vez que proporciona tanto armazenamento de energia como precursores biossintéticos.

Conclusão

A compreensão da origem e evolução das vias metabólicas é a chave para a compreensão do surgimento da vida. Aqui, analisamos a origem e evolução das proteínas envolvidas na Glicólise / Gliconeogênese. Os resultados sugerem que a informação biológica inicial pode ter tido origem a partir de proto-tRNAs e os primeiros peptídeos agiam de forma simples ligando cofatores e com a complexificação da estrutura proteica por acréscimo dos novos grupamentos, foi possível a sofisticação do domínio catalisador. Estes resultados indicam que no surgimento dos primeiros peptídeos, os grupamentos simples estruturais funcionaram como blocos de montagem, e pela combinação dos diferentes grupamentos surgiram diferentes proteínas que agem nos organismos modernos. A evolução das proteínas envolvidas na Glicólise / Gliconeogênese ocorreu por pressão seletiva para proteger o domínio catalítico, o que possibilitou a melhora da catálise. Sugerimos que a Gliconeogênese precede a Glicólise e com a acumulação do carbono no sistema biológico inicial foi obtido um equilíbrio entre as vias.

Informação Adicional

- Interesses financeiros concorrentes

Os autores declaram não haver interesses financeiros concorrentes.

-Auxilio financeiro

MVJ foi financiado por PAPIIT-IN224015; UNAM; México.

- Conflito de interesse

Nenhum

Referências

- ALVES, R., CHALEIL, R.A & STERNBERG, M.J. (2002). Evolution of enzymes in metabolism: a network perspective. *Journal of Molecular Biology*, 19;320(4):751-70.
- BERMAN, H.M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T.N., WEISSIG, H., SHINDYALOV, I.N & BOURNE, P.E. (2000). The Protein Data Bank. *Nucleic Acids Research*, 1;28(1):235-42.
- CAETANO-ANOLLÉS, G., KIM, K.M & CAETANO-ANOLLÉS, D. (2012). The phylogenomic roots of modern biochemistry: origins of proteins, cofactors and protein biosynthesis. *Journal of Molecular Evolution*, 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. *Journal of Molecular Evolution*, 57(6):721-30.
- EIGEN, M & SCHUSTER, P. (1978) The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, 64(11):541-65.
- EIGEN, M & WINKLER-OSWATITSCH, R. (1981) Transfer-RNA, an early gene? *Naturwissenschaften*, 68(6):282-92.
- FARIAS, S.T, REGO, T.G & JOSE, M.V. (2016). A proposal of the proteome before the last universal common ancestor (LUCA). *International Journal of Astrobiology*, 15 (1): 27-31. doi:10.1017/S1473550415000464.
- FORTERRE, P., GRIBALDO, S & BROCHIER, C. (2005). Luca: the last universal common ancestor. *Médecine Sciences (Paris)*, 21(10):860-5.
- GLANSDORFF, N., XU, Y & LABEDAN, B. (2008). The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biology Direct*, 9;3:29. doi: 10.1186/1745-6150-3-29.
- GOLDMAN, A.D., BERNHARD, T.M., DOLZHENKO, E & LANDWEBER, L.F. (2013). LUCApedia: a database for the study of ancient life. *Nucleic Acids Research*, (Database issue):D1079-82. doi: 10.1093/nar/gks1217.
- GOUY, M., GUINDON S. & GASCUEL O. (2010). SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27(2):221-224.
- HUBERTY, C. J. (2014). Mahalanobis Distance. *Wiley StatsRef: Statistics Reference Online*.
- KATOH, K & STANDLEY, D.M. (2016). A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*, 1;32(13):1933-42. doi: 10.1093/bioinformatics/btw108.

KENNETH, S.B. (2004). *Functional Metabolism: Regulation and Adaptation*. John Wiley & Sons, INC., Publication.

KIM, K.M & CAETANO-ANOLLÉS, G. (2011). The proteomic complexity and rise of the primordial ancestor of diversified life. *BMC Evolutionary Biology*, 25;11:140.doi:10.1186/1471-2148-11-140.

KUMAR, S., STECHER, G & TAMURA, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33 (7): 1870-1874. doi:10.1093/molbev/msw054.

MUSHEGIAN, A. (2008). Gene content of LUCA, the last universal common ancestor. *Frontiers in Biosciences*, 1;13:4657-66.

NISBET, E., ZAHNLE, K., GERASIMOV, M.V, HELBERT, J., JAUMANN, R., HOFMANN, B.A., BENZERARA, K & WESTALL, F. (2007). Creating habitable zones, at all scales, from planets to mud micro-habitats, on Earth and on Mars. *Space Science Reviews*, 129:79–121.

PENNY, D & POOLE, A. (1999). The nature of the last universal common ancestor. *Current Opinion in Genetics and Development*, 9(6):672-7.

SLEEP, N.H. (2007). Plate tectonics through time, *Treatise on Geophysics Volume 9*, (ed. Schubert G.), pp. 101–117 Oxford, Elsevier. SUTHERLAND, J.D. (2017). Studies on the origin of life – the end of the beginning. *Nature reviews chemistry*, 01:12. DOI/10.1038/s41570-016-0012.

WEISS, M.C., SOUSA, F.L., MRNJAVAC, N., NEUKIRCHEN, S., ROETTGER, M., NELSON-SATHI, S & MARTIN W.F. (2016). The physiology and habitat of the last universal common ancestor. *Nature Microbiology*, 25;1(9):16116. doi: 10.1038/nmicrobiol.2016.116.

WOESE, C. (1998) The universal ancestor. *Proceedings of the National Academy of Sciences of the United States of America*, 9;95(12):6854-9.

YANG, J., ROY, A & 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., NISBET, E., SELSIS, F & SLEEP, N.H. (2007). Emergence of a Habitable Planet. *Space Science Reviews*, 129:35–78.

ZHANG, Y & SKOLNICK J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 22;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.

XU, D. & Zhang Y. (2011). Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. *Biophysical Journal*, vol 101, 2525-2534.

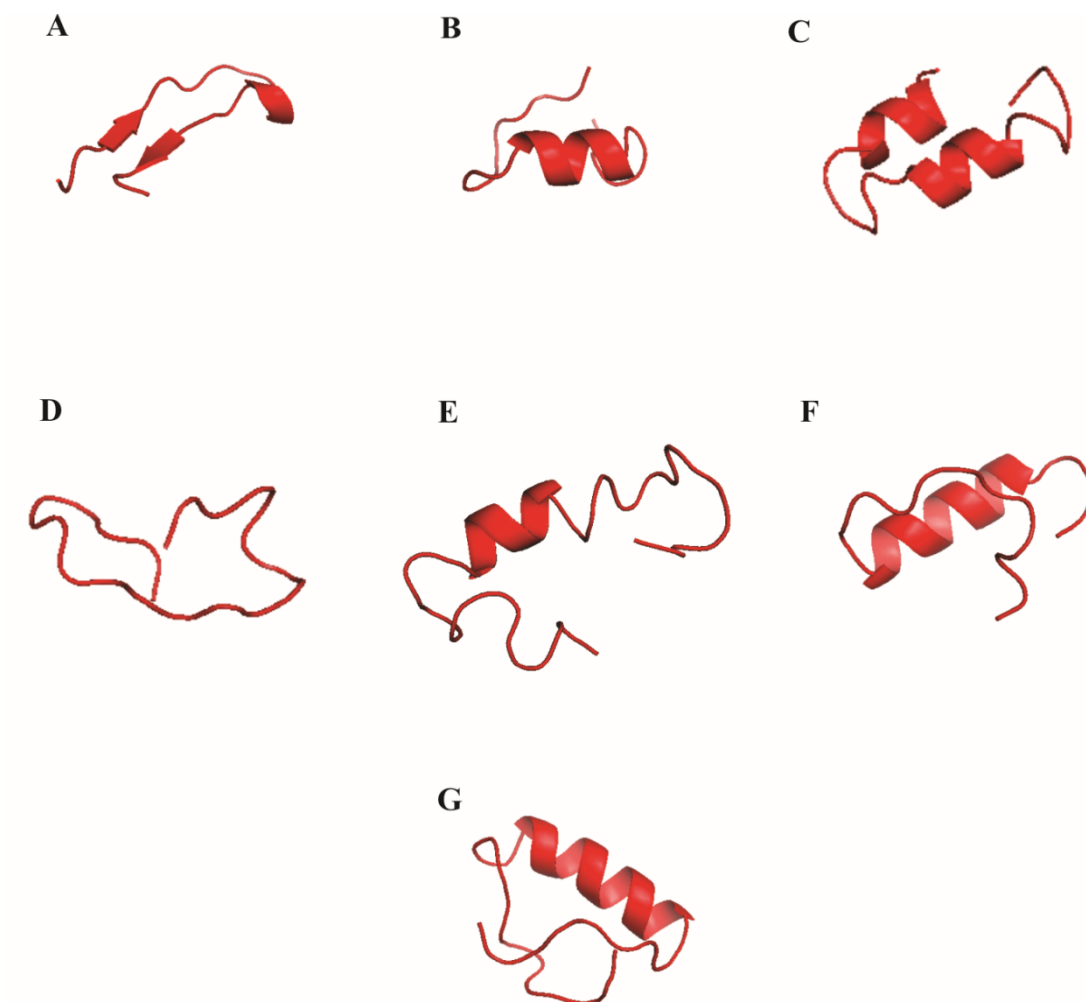
Figuras e Legendas*Figuras***Figura 1.**

Figura 2.

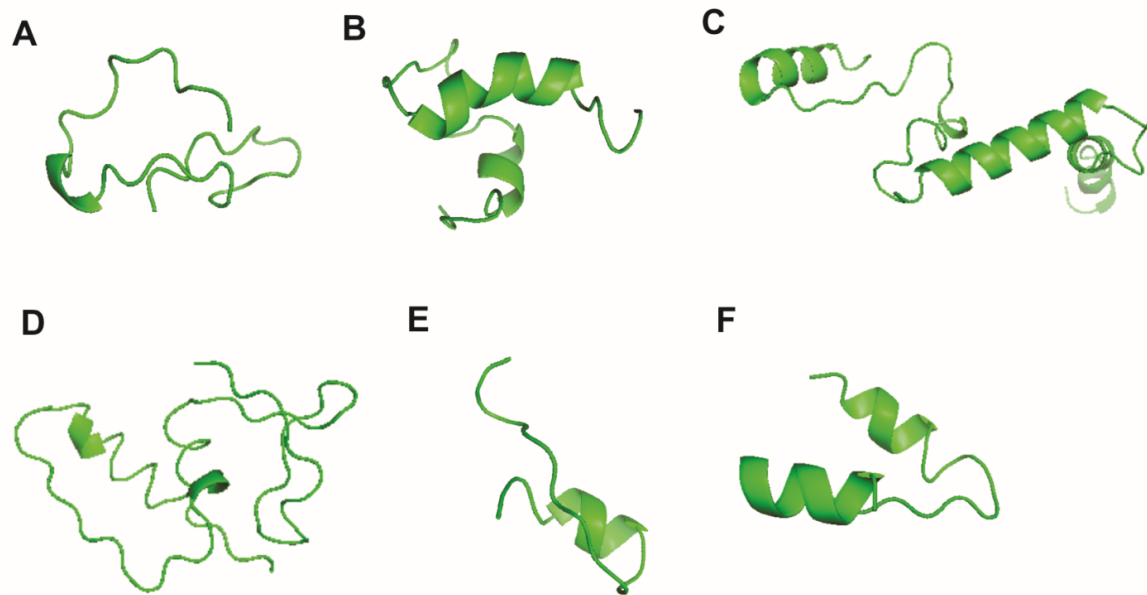


Figure 3.

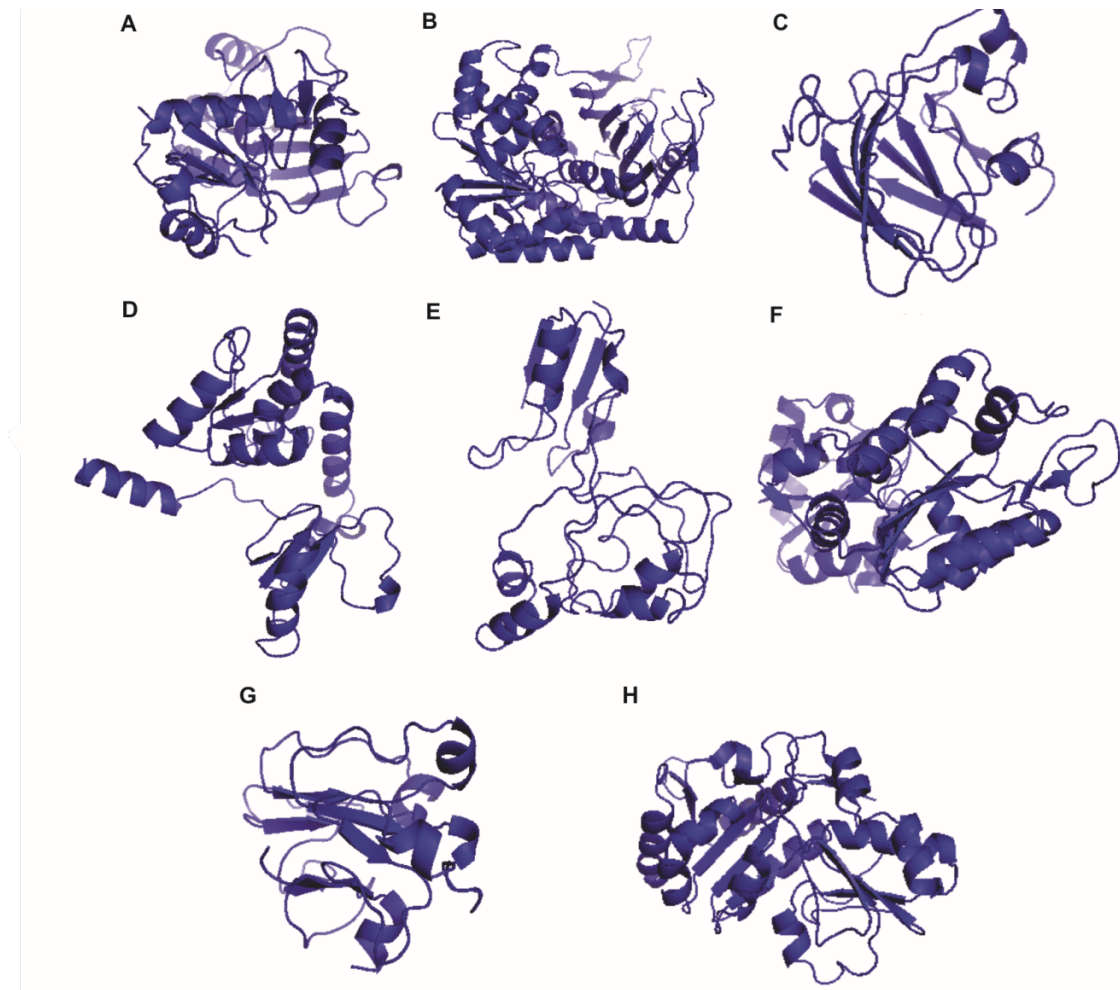


Figura 4.

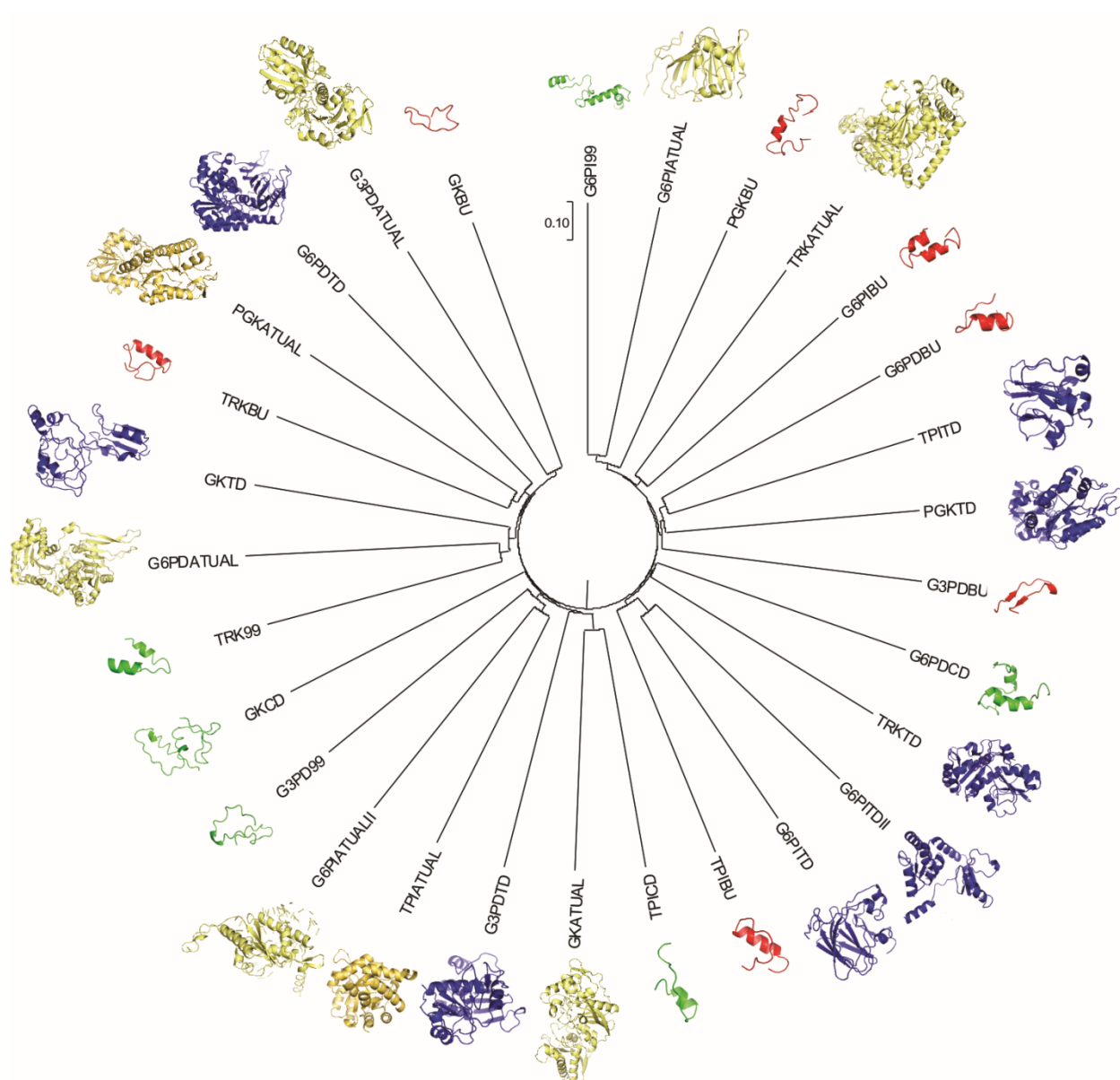


Figura 5.

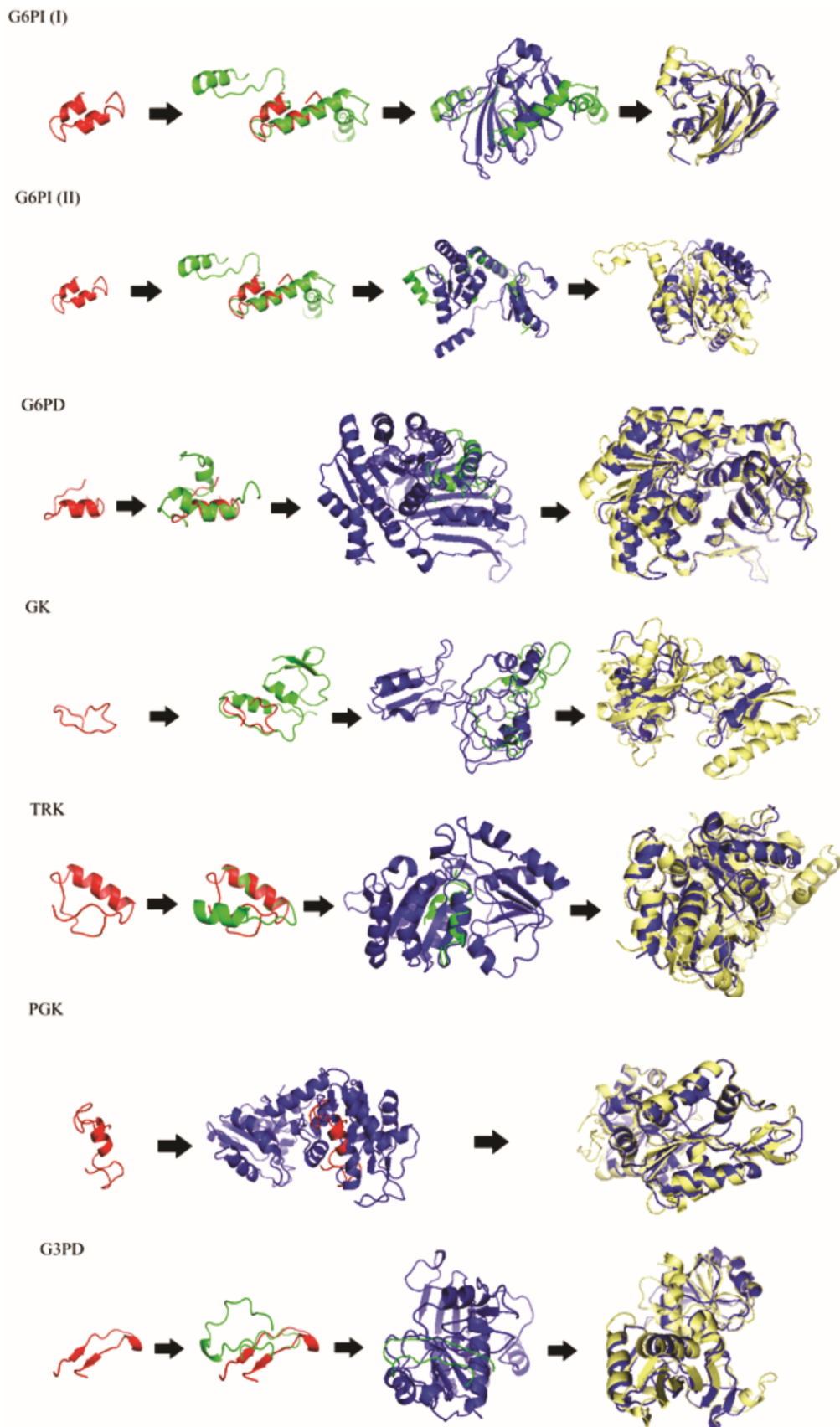


Tabela 1.

Bottom Up	Most Similar Structure	Organism	RMSD
Glucose-6-phosphate Isomerase	Iron ABC transporter solute-binding protein (4HN9)	<i>Eubacterium eligens</i>	1.69
Phosphoglycerate-kinase	RNA binding domain (4GH9)	<i>Marburg virus</i>	2.67
Triosephosphate-isomerase	Hydrolase (Crystal structure of had family enzyme bt-2542 target e fi-501088) magnesium complex (4DFD)	<i>Bacteroides thetaiotaomicron</i>	1.46
Glycerate-kinase	Kinase domain of protein tyrosine kinase 2 beta (3CC6)	<i>Homo sapiens</i>	1,95
Glyceraldehyde-3-phosphate-dehydrogenase	Enolase superfamily member complexed with Mg and Glycerol (3R25)	<i>Vibrionales bacterium</i>	2.10
Glucose-6-phosphate-1-dehydrogenase	CRISPR-associated protein Cas6e (4DZD)	<i>Escherichia coli</i>	1.78
Transketolase	Acyl-CoA dehydrogenase fold (5AHS)	<i>Advenella mimigardefordensis</i>	1.99

Tabela 2.

	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 Aplicable	<i>Bacillus anthracis</i>	(1) <i>Pyrococcus furiosos</i> . (2) <i>Thermococcus litoralis</i>		1.62	1.07
Phosphoglycerate-kinase	Not Applicable	Not Applicable	Phosphoglycerate kinase (1PHP)	Not Aplicable	Not Aplicable	<i>Geobcacidillus stearothermophilus</i>	Not Applicable	Not Applicable	0.85
Triosephosphate-isomerase	Chondroitin polymerase (2Z86)	Not Applicable	Triosephosphate-isomerase complex with 2-phosphoglycolic acid (1BTM)	<i>Escherichia coli</i>	Not Aplicable	<i>Geobcacidillus stearothermophilus</i>	1.74	Not Applicable	2.37
Glycerate-kinase:	Glycerate kinase (3CWC)	Not Applicable	Glycerate kinase (1TO6)	<i>Salmonella typhimurium</i>	Not Aplicable	<i>Nisseria meningitidis</i>	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 Aplicable	<i>erichia virus Lar</i>	<i>Trypanosoma cruzi</i>	Not Applicable	1.87	0.92
Glucose-6-phosphate-1-dehydrogenase	Enoyl-CoA hydratase (3PES)	Not Applicable	Glucose-6-phosphate-1-dehydrogenase (1H9B)	<i>Mycobacterium smegmatis</i>	Not Aplicable	<i>Leuconostoc mesenteroides</i>	2.47	Not Applicable	1.23
Transketolase	Not Applicable	Selenomethionine incorporated apo D-serine deaminase	Transketolase with thiamin diphosphate and CA+ (3M34)	Not Aplicable	<i>Salmonella typhimurium</i>	<i>Campylobacter jejuni</i>	Not Applicable	1.32	0.99

Tabela 3.

	Bottom Up	Top Down Complete Deletion/ 99% Threshold	Top Down 95%
Glucose-6-phosphate-isomerase	Ethylene glycol, Nucleic Acid, S-Adenosylmethionine and methionine	Bicarbonate, Chlorophyll, Bacterioruberin, Cholesterol, Zinc	Glucose-6-phosphate and 6-phosphogluconate
Phosphoglycerate-kinase	Guanosine Triphosphate	(Not Calculated)	Adenosine Diphosphate and 3-phosphoglycerate
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 trioxide, Diethyl-4-methylbenzylphosphonate, [FE4S4] and Glutamic acid	5-phospho-darabinohydroxamic acid, Iron, 1-Malate, Zinc ion, Adenyl-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

Legendas

Figura 1. Estruturas das sequências ancestrais obtidas pela abordagem Bottom Up.

A) Gliceraldeído-3-fosfato-desidrogenase, B) Glucose 6-fosfato 1-desidrogenase, C) Glucose 6 Fosfato Isomerase, D) Glicerato Quinase, E) Fosfoglicerato quinase, F) Triose fosfato isomerase e G) Transquetolase

Figura 2. Estruturas das sequências ancestrais obtidas pela abordagem Top Down completa deleção (CD) e 99% de cobertura do local (PD99). A) Gliceraldeído-3-fosfato-desidrogenase (PD99), B) Glucose 6-fosfato 1-desidrogenase (CD), C) Glucose 6 Isomerase de fosfato (PD99), D) Glicerato quinase (CD), (E) Triose fosfato isomerase (CD) e (F) Transquetolase (PD99).

Figura 3. Estruturas das sequências ancestrais obtidas pela abordagem Top Down com 95% de cobertura do local. A) Gliceraldeído-3-fosfato-desidrogenase, B) Glucose 6-fosfato 1-

desidrogenase, C) Glucose 6 Fosfato Isomerase (I), D) Glucose 6 Fosfato Isomerase (II), E) Glicerato quinase, F) Fosfoglicerato quinase, G) Triose fosfato isomerase e H) transcetolase.

Figura 4. Distância filogenética baseada no valor de RMSD. Em vermelho, as estruturas das sequência ancestrais da abordagem Bottom up, Gliceraldeído-3-fosfato-desidrogenase (G3PDBU), Glucose 6-fosfato1-desidrogenase (G6PDBU), Glucose 6 fosfato isomerase (G6PIBU), Glicerato quinase (GKBU), fosfoglicerato quinase (PGKBU) , Triose fosfato isomerase (TPIBU) e Transquetolase (TRKBU). Em verde, as estruturas das sequência ancestrais da abordagem Top Down com supressão completa e 99% de cobertura de cobertura de local, Gliceraldeído-3-fosfato-desidrogenase (G3PD99), Glucose 6-fosfato1-desidrogenase (G6PDCD), Glucose 6 Fosfato Isomerase (G6PI99), Glicerato Quinase (GKCD), Triose fosfato isomerase (TPICD) e Transquetolase (TRK99). Em azul, as estruturas das sequência ancestrais da abordagem Top Down com 95% de cobertura de cobertura de local, Gliceraldeído-3-fosfato-desidrogenase (G3PDTD), Glucose 6-fosfato1-desidrogenase (G6PDTD), Glucose 6 Phosphate Isomerase (G6PITDI), Glucose-6-Isomerase Fosfato (G6PITDII), Glicerato Quinase (GKTD), Fosfoglicerato quinase (PGKTD), Triose fosfato isomerase (TPITD) e Transquetolase (TRKTD). Em amarelo pálido, as enzimas modernas coletadas na Base de Dados de Proteínas (PDB), Gliceraldeído-3-fosfato-desidrogenase (G3PDATUAL), Glucose 6-fosfato1-desidrogenase (G6PDATUAL), Glucose 6 Fosfato Isomerase (G6PIATAULI), Glucose 6 Fosfato Isomerase G6PIATUALII), Glicerato Quinase (GKATUAL), Fosfoglicerato quinase (PGKATUAL), Triose fosfato isomerase (TPIATUAL) e Transquetolase (TRKATUAL).

Figura 5. Passos evolutivos propostos das estruturas. Alinhamento feito usando o servidor TM-Align. Em vermelho, os ancestrais de Bottom Up, em verde os ancestrais Top Down calculados com supressão completa ou os Top Down calculados com supressão parcial com limiar de 99%, em azul os ancestrais Top Down calculados com limiar de 95% e em amarelo pálido as enzimas modernas. Gliceraldeído-3-fosfato-desidrogenase (G3PD), Glucose 6-fosfato1-desidrogenase G6PD, Glucose 6 Fosfato Isomerase (G6PI (I) e (II)), Glicerato Quinase (GK), Fosfoglicerato quinase (PGK), Triose fosfato isomerase TPI) e Transquetolase (TRK).

Tabela 1. Proteínas ancestrais obtidas pela abordagem Bottom Up, a estrutura presente em organismos modernos mais semelhante em termos de RMSD e o valor de RMSD entre a proteína ancestral e a proteína moderna.

Tabela 2. Proteínas ancestrais obtidas pela abordagem Top Down, a estrutura presente em organismos modernos mais semelhante em termos de RMSD e o valor de RMSD entre a proteína ancestral e a proteína moderna.

Tabela 3. As estruturas ancestrais obtidas pela abordagem Bottom up e Top Down, e os mais prováveis ligantes para cada proteína ancestral.

4. CONCLUSÃO

Os nossos resultados possibilitaram inferências sobre diferentes ângulos da evolução das proteínas analisadas. Dada a provável antiguidade e relevância das vias aqui estudadas, compreender a origem e surgimento de especificidade das enzimas relacionadas nos permite propor modelos sobre a evolução de outras vias metabólicas, e assim nos aproximar do que pode ter ocorrido há aproximadamente 4 bilhões de anos com o desenvolvimento das primeiras formas de vida como se conhece hoje.

De acordo com a nossa interpretação dos dados obtidos e trabalhos anteriores, na Terra do éon Hadeano teriam surgido as primeiras biomoléculas com valor informacional. Essas estruturas teriam sido homólogas ao que hoje chamamos de RNAs transportadores (tRNAs). Essas moléculas possuem uma série de características que possibilitam a ação de carreadoras de informação, como uma estabilidade conferida graças a predominância de pareamento entre guanina e citosina, e a sua estrutura secundária que aumenta sua resistência à hidrólise. Os tRNAs fariam naquela época uma ação semelhante àquela que desempenham hoje: levariam moléculas menores, como o caso de aminoácidos, de um lugar a outro; dessa forma podemos entender os códons como aquilo que continha a mensagem biológica.

Consideramos que os tRNAs presentes nesse momento da história evolutiva eram do tipo RNY. Esses são códons que a primeira base é uma purina, a segunda pode ser qualquer base e a terceira uma pirimidina. Assim, os primeiros módulos seriam ricos em Alanina, Serina, Treonina, Asparagina, Ácido Aspártico, Valina, Isoleucina e Glicina. Partindo disso, Farias et al (2016) propuseram um proteoma para os progenotes com base na tradução de tRNAs ancestrais com um padrão RNY. A tradução dos tRNAs ancestrais obtidos mostraram enzimas de vias provavelmente muito antigas, como o caso das vias dos carboidratos, mais especificamente, proteínas envolvidas na glicolise/gliconeogênese.

Continuando a reconstrução da história evolutiva, esses tRNAs formariam proteínas estatísticas por afinidade química que já seriam alvo da seleção natural, assim, com o tempo certas estruturas se tornariam mais prováveis, ou seja, com maiores amostras na “população” em comparação a outras estruturas. Dentre essas estruturas, aquelas que possuíam a capacidade de geração de energia e de obtenção de precursores de componentes biológicos seriam favorecidas, uma vez que diminuiriam a dependência dessas vias no meio.

Compreendemos que a falta de especificidade das estruturas obtidas as quais julgamos serem mais antigas indica que essas “pré-enzimas” agissem de modo amplo no metabolismo. Em outras palavras, os primeiros polipeptídeos ligavam-se em diferentes substratos, tendo assim

diversas funções no metabolismo. Com a evolução, essas estruturas foram se complexificando e aumentando de tamanho de sorte que o centro catalítico se tornou menos periférico e mais protegido, com *dockings* mais eficientes e específicos.

Além disso, na análise de distância das estruturas, observamos que as estruturas mais antigas são mais próximas entre si do que são com suas homólogas. Juntando com a similaridade em relação aos substratos que se ligavam, nós sugerimos que as enzimas das vias analisadas poderiam ter tido uma origem comum. A primeira molécula que deu origem ao que hoje é a via da Glicólise/ Gliconeogênese provavelmente foi inespecífica em relação aos substratos, com o sítio catalítico exposto e agia de forma ampla no metabolismo. Ao longo do processo evolutivo as enzimas foram se tornando cada vez mais específicas, maiores e mais complexas.

Em relação ao ambiente que essas primeiras enzimas foram formadas, de acordo com a análise de similaridade com estruturas atuais, notamos que a maioria das moléculas eram estruturalmente semelhante a proteínas ou partes de proteínas presentes em organismos extremófilos de altas temperaturas e/ou alta acidez. Isso pode indicar que os primeiros passos evolutivos dessas vias foram dados em ambientes com pH baixo e temperaturas altas, o que concorda com estudos anteriores.

ANEXOS

ANEXO A – Resumo da formatação da revista científica "Quaternary Reviews of Biophysics" Segundo o site (<https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/information/instructions-contributors>):

Os editores sugerem o uso de Word para PC ou Macintosh.

O texto deve ser escrito em inglês, espaçamento duplo com uma margem larga ao redor de todo o texto. É pedido numeração das linhas e das páginas. Imagens, Tabelas e legendas devem estar ao final do manuscrito e devem ser digitadas separadamente.

Organização do manuscrito:

Página do Título

Incluindo o título, que deve ser curto mas informativo e refletir o conteúdo do trabalho, os nomes dos autores, o nome e o endereço dos departamentos e instituições atribuídas para cada autor. O endereço de e-mail, telefone e fax dos autor responsável.

Resumo

Cada artigo deve incluir um resumo indicando os principais pontos em menos de 300 palavras.

Introdução

Deve ser o mais curto possível, simples, sugerem por volta de três parágrafos.

Material e métodos

Deve haver informação suficiente para que o leitor possa repetir o trabalho. Técnicas descritas detalhadamente em outros trabalhos não precisam ser repetidas e sim citadas.

Resultados

Deve-se restringir aos resultados obtidos, se necessário usar testes estatísticos.

Discussão

Os resultados (incluindo a referência a números e tabelas) não devem ser repetidos em pormenor nem devem ser introduzidas novas informações. A especulação é encorajada, mas não deve ir além de hipóteses razoáveis e testáveis. A discussão não deve tentar ser uma mini-revisão.

Agradecimentos

Nessa seção você pode reconhecer pessoas ou organizações que forneceram aconselhamento, apoio (não financeiro). O apoio financeiro formal e o financiamento devem ser listados na seguinte seção

Apoio financeiro

Forneça detalhes das fontes de apoio financeiro para todos os autores, incluindo números de subsídios.

Conflito de Interesse

Por favor, forneça detalhes de todas as relações financeiras, profissionais e pessoais conhecidas com o potencial de viés do trabalho. Onde não existem conflitos de interesse conhecidos, inclua a seguinte declaração: "Nenhum".

Material Suplementar

Material adicional (por exemplo, conjuntos de dados, tabelas grandes) relevante para o papel pode ser submetido para publicação somente on-line, onde eles são disponibilizados através de um link do papel. O papel deve ficar sozinho sem esses dados. O material suplementar deve ser citado em um lugar relevante no texto do artigo.

Referencias

As referências devem ser de autor e ano. Quando um artigo citado tem três ou mais autores, o estilo 'Smith et al. 2013' deve ser usado em todas as ocasiões.

No final do artigo, as referências devem ser listadas em ordem alfabética, com um título completo de cada artigo. Os títulos da revista devem ser fornecidos na íntegra. Os usuários do EndNote podem usar o arquivo de estilo distribuído para "Biological Reviews".

A exatidão das referências é da responsabilidade do (s) autor (es). As referências devem ser verificadas em relação ao texto para assegurar a) que a ortografia dos nomes dos autores e as datas dadas são coerentes e b) que todos os autores citados no texto (na ordem das datas, se houver mais de um) são indicados na referência Lista e vice-versa.