

**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE BIOLOGIA MOLECULAR**



DANIEL ANTONINO CORDEIRO

**AMINOACIL-TRNA SINTETASES, O CÓDIGO GENÉTICO E A ORIGEM DA
VIDA**

Prof. Dr. Sávio Torres de Farias

Orientador

João Pessoa
2015

DANIEL ANTONINO CORDEIRO

**AMINOACIL-TRNA SINTETASES, O CÓDIGO GENÉTICO E A ORIGEM DA
VIDA**

Monografia apresentada ao Curso de Ciências Biológicas da Universidade Federal da Paraíba, como requisito parcial à obtenção do título de Bacharel em Ciências Biológicas.

Prof. Dr. Sávio Torres de Farias

Orientador

João Pessoa
2015

DANIEL ANTONINO CORDEIRO

**AMINOACIL-TRNA SINTETASES, O CÓDIGO GENÉTICO E A ORIGEM DA
VIDA**

Monografia apresentada ao Curso de Ciências Biológicas, como requisito parcial à obtenção do grau de Bacharel em Ciências Biológicas.

Data:

Resultado:

BANCA EXAMINADORA:

Prof. Dr. Sávio Torres de Farias (orientador)
Departamento de Biologia Molecular - UFPB

Prof. Dr. Luís Fernando Marques-Santos
Departamento de Biologia Molecular - UFPB

Prof. Dr. Gustavo Henrique Calazans Vieira
Departamento de Sistemática e Ecologia - UFPB

RESUMO

A origem da vida é muitas vezes tida como a grande questão da biologia. Pergunta-se a respeito desde os primórdios da humanidade, e no meio científico ainda não temos uma resposta satisfatória. Os últimos anos foram um período de grande foco nessa área de estudo devido aos avanços nas técnicas moleculares e da bioinformática. As primeiras moléculas a iniciarem a maquinaria biológica, a origem do código genético e outras questões importantes acumularam evidências, mas permanecem em aberto. Neste trabalho teórico, fizemos um apanhado de todo o histórico dessa questão e exploramos o que temos até agora. Nossa abordagem foi feita através das aminoacil-tRNA sintetases (aaRS), enzimas responsáveis pela associação entre os aminoácidos e seus tRNAs correspondentes. As aaRS, portanto, definem as regras do código genético e podem fornecer dados cruciais no que diz respeito à origem da vida. Através de softwares como o BLAST, MEGA, RCSB PDB, TNT e o servidor I-TASSER, pudemos gerar dados sequenciais, estruturais e filogenéticos das aaRS que podem nos ajudar a formular hipóteses evolutivas que deem alguma luz à grande questão. Naturalmente, quaisquer resultados lançados aqui estarão longe de serem conclusivos, mas há fortes sugestões a serem consideradas: há uma clara ordem no código genético parcialmente refletida nas filogenias; foi verificada a presença de domínios de ligação ao anticódon em aaRS ancestrais geradas por alinhamento sequencial, tanto em sequência quanto em estrutura e dados propostos por outros trabalhos parecem corroborar a hipótese de um mundo ribonucleoprotéico, com um código genético rudimentar já estabelecido desde o início por relações esteroquímicas baseadas nas propriedades hidropáticas de tRNAs, sintetases e aminoácidos e que tenha passado por vários ganhos de especificidade e minimização de erros até se consolidar no que temos hoje.

Palavras-chave: origem da vida; aminoacil-tRNA sintetases; código genético

ABSTRACT

The origin of life is often taken as biology's ultimate question. It is a wonder since the dawn of humanity, and in the scientific community we still don't have a satisfactory answer. The last years have been a time of huge focus in this field of study due to the advances on molecular tools and bioinformatics. The first molecules to initiate the biological machinery, the origin of the genetic code and other main questions accumulated evidences, but remain open. In this theoretical work, we made a summary of all the history on this question and explored what we have so far. Our approach was made through the aminoacyl-tRNA synthetases (aaRS), enzymes responsible for the association between amino acids and their correspondent tRNAs. The aaRS, thus, define the rules of the genetic code and may provide crucial data regarding the origin of life. Through softwares like BLAST, MEGA, RCSB PDB, TNT and the I-TASSER server, we were able to generate sequential, structural and phylogenetic data from the aaRS that might help us formulate evolutionary hypothesis and shed some light on the subject. Naturally, any results shown here would be far from conclusive, but there are several suggestions to be considered: there's a clear order on the genetic code partially reflected on the phylogenies; it was verified the presence of anticodon-binding domains on ancestral aaRS generated by sequence alignment both in sequence and in structure and data proposed by other works seem to corroborate the hypothesis of a ribonucleoprotein world, with a rudimentary genetic code already established since the beginning by stereochemical relationships based on the hydrophobic properties of tRNAs, synthetases and amino acids and that had been through several increases on specificity and error minimization until becoming what we have today.

Keywords: origin of life; aminoacyl-tRNA synthetases; genetic code

SUMÁRIO

RESUMO

ABSTRACT

ÍNDICE

1. INTRODUÇÃO	1
1.1 A origem da vida	1
1.2 O código genético	4
1.3 As aminoacil-tRNA sintetases	6
1.4 A reconstrução de ancestrais	8
2. OBJETIVOS.....	10
2.1 Objetivos gerais	10
2.2 Objetivos específicos	10
3. MATERIAL E MÉTODOS	11
3.1 Domínios protéicos	11
3.2 Filogenias	11
4. RESULTADOS E DISCUSSÃO	13
4.1 Análise sequencial	13
4.2 Análise estrutural	14
4.3 Filogenia sequencial	16
4.4 Filogenia estrutural	19
5. CONSIDERAÇÕES FINAIS.....	21
6. REFERÊNCIAS.....	22

1. INTRODUÇÃO

1.1 A ORIGEM DA VIDA

De onde viemos é uma questão fundamental para toda a humanidade. Quando se fala em origem da vida, pode-se chegar de cara em várias correntes de pensamento das mais discrepantes naturezas. O que é vida? Os criacionistas e teólogos usariam argumentos espirituais. Segundo as religiões abraâmicas, a vida teria sido insuflada na matéria por Deus, não decorrendo de leis naturais. Psicólogos usariam argumentos psíquicos, e relacionariam a vida ao âmbito da mente. Para uma pessoa comum, ela estaria geralmente associada às experiências da existência. Todas estas constituem visões antropocêntricas da vida em si, que é um tema ainda relativamente recente no âmbito científico, se comparado com sua antiguidade no pensamento filosófico e religioso (DAMINELI; DAMINELI, 2007).

Seguindo a lógica das religiões abraâmicas, escritos de Aristóteles (384-322 a.C.) mencionam que haveria um princípio divino da vida animal (*pneuma*), e este “sopro vital” passaria para os descendentes por meio da reprodução. Porém, alguns organismos seriam capazes de surgir espontaneamente, sem serem frutos de outro ser vivo. Um exemplo seria o surgimento de ratos e insetos em carne podre. Acreditava-se que tais animais (matéria viva) teriam originado-se a partir da carne (matéria não-viva). Esta teoria ficou conhecida como *geração espontânea* ou *abiogênese*, e perdurou por todas as partes do mundo por dois milênios.

Até que Francesco Redi (1626-1697) demonstrou que os vermes brancos presentes na carne podre resultam da postura de ovos por insetos, e são apenas larvas que se desenvolvem em outro inseto. Seus experimentos bateram de frente com o que era tido como dogma na época. A comunidade científica entrou em polvorosa, e um confronto particularmente acirrado foi entre o inglês John Needham (1713-1781) e o italiano Lazzaro Spallanzani (1729-1799). Ambos realizaram experimentos similares e obtiveram diferentes resultados. Nos experimentos, líquidos que continham matéria orgânica foram fechados em recipientes, aquecidos, e posteriormente examinados para sinais de putrefação. Spallanzani havia selado hermeticamente seus recipientes, e ferveu seu conteúdo por longos períodos. Needham usou cortiças, e aqueceu o líquido sobre cinzas. No fim, os recipientes de Needham continham microorganismos, enquanto alguns dos recipientes de Spallanzani permaneceram completamente livres de

micróbios. Needham concluiu que a geração espontânea era uma consequência inevitável da existência de matéria orgânica, e Spallanzani concluiu que o aparecimento de seres vivos não seria observado se as devidas precauções fossem tomadas na esterilização. Spallanzani, entretanto, não obteve muito sucesso em convencer seus contemporâneos.

Foi com os meticolosos e elegantes experimentos de Louis Pasteur (1822 – 1895) que a teoria da geração espontânea caiu por terra. Pasteur inicialmente esquentava o ar que era direcionado até um caldo nutritivo, destruindo os microorganismos e impedindo sua proliferação no caldo. Para responder às alegações de que a “força vital” teria sido destruída ao se aquecer o ar, ele refinou seu experimento utilizando frascos com pescoço de cisne. O formato dos pescoços dos frascos impedia que partículas viáveis caíssem no fundo. Ao quebrar-se o pescoço destes mesmos frascos, a proliferação de microorganismos tinha início. É de se notar que Pasteur não decretou a impossibilidade da geração espontânea, afinal, havia o problema do primeiro organismo. Se a matéria poderia ou não organizar-se a si mesma era uma questão que ainda prevalecia.

Hoje, temos evidências de que, de fato, ela pode. Charles Darwin (1809-1882) pôs em cheque as ideias da igreja acerca da origem da vida com a teoria da evolução e a ancestralidade nos seres vivos. Darwin imaginava que um caldo nutritivo contendo amônia, sais de fósforo, luz, calor e eletricidade pudesse ter dado origem a proteínas, e a partir das ideias de Aleksandr Ivanovich Oparin (1894-1980), que também havia elaborado modelos onde a vida seria produto da evolução bioquímica na Terra primitiva, a experimentação no ramo da origem da vida entrou em uma nova era. Oparin propôs que a organização de lipídios em pequenas estruturas esféricas, denominadas *coacervados*, poderiam ter dado origem às primeiras formas de vida. Tais estruturas, primeiro investigadas pelo químico H. G. Bungenberg de Jong em 1932, formam-se espontaneamente em soluções onde forças hidrofóbicas mantêm as moléculas orgânicas juntas. Em 1953, Stanley Miller (1930-2007) e seu famoso experimento demonstraram como aminoácidos poderiam surgir através de processos abióticos num ambiente redutor, que mimetizava as condições da Terra primitiva propostas por Harold Urey (1893-1981), onde a atmosfera seria composta primariamente por amônia, metano e hidrogênio. No experimento, os gases passavam por uma câmara, onde eram submetidos a descargas elétricas, depois condensados num recipiente de água e

evaporados novamente. Em poucos dias se observou um precipitado rico em aminoácidos. Tal resultado abriu novos horizontes para a área. Restava saber como poderiam esses aminoácidos originarem moléculas maiores, e provavelmente não seria na água, já que a formação das ligações peptídicas entre os aminoácidos libera uma molécula de água como produto. Sidney Fox (1912-1998), seguindo as idéias de Oparin, onde a água da Terra primitiva rica em aminoácidos era levada pelas chuvas para a superfície de rochas quentes, realizou um experimento onde ele aquecia uma solução líquida de aminoácidos numa superfície seca e aquecida. Formaram-se ligações peptídicas que resultavam em moléculas semelhantes às proteínas e delimitadas por membranas, denominadas *microesferas (proteinoídes)*. Assim como nos coacervados de Oparin, o experimento de Fox sugere que as microesferas poderiam ter se desenvolvido ao longo do tempo ao ponto de adquirirem as características de um ser vivo. O problema seria como esse organismo poderia se reproduzir para perpetuar-se no ambiente. Em outras palavras, na replicação da informação.

Outros estudos apontaram que a formação de açúcares e nucleotídeos também seria possível na Terra pré-biótica (FOX; DOSE, 1977). A hipótese do *mundo de RNA*, onde o RNA seria a molécula primordial data desde os papers de Francis Crick, em 1968. Na década de 80, Thomas R. Cech, Sidney Altman e colaboradores descobriram propriedades catalíticas em algumas moléculas de RNA, denominadas *ribozimas*, confirmando as suposições de Crick (KRUGER et al., 1982; GUERRIER-TAKADA et al., 1983). Sendo o RNA uma molécula muito mais simples do que o DNA e ter a capacidade de armazenar informação e catalisar reações, ela poderia ser considerada um modelo auto-replicativo que teria dado início à formação do primeiro organismo (DE DUVE, 1991; DWORKIN et al., 2003). Porém, a complexidade dos nucleotídeos, a forma insolúvel do fosfato na Terra e as condições altamente danosas da Terra primitiva impõem dificuldades nesta hipótese.

Há também a possibilidade de um mundo *ribonucleoprotéico*. Por essa linha nunca existiu um mundo de RNA; as proteínas já estariam associadas às ribozimas desde o início. Estudos envolvendo as intrínsecas relações entre RNAs e proteínas no ribossomo sugerem uma gradual coevolução entre as moléculas (HARISH; CAETANO-ANOLLÉS, 2012). Harish e Caetano-Anollés (2012) propõem que o sítio peptidil-transferase, responsável pela síntese protéica no ribossomo e localizado na subunidade maior, tenha aparecido posteriormente à subunidade menor, o que leva à sugestão de

que proteínas e RNAs tenham coevoluído para dar origem a um ribossomo moderno com múltiplas subunidades. Desta forma, a vida teria se iniciado da mesma forma como funciona hoje: através da interação entre proteínas e RNA.

1.2 O CÓDIGO GENÉTICO

Uma outra etapa crucial a se levar em conta quando se trata da origem da vida é a origem e evolução do código genético. Com uma olhada rápida na tabela do código pode-se perceber que a associação dos aminoácidos é claramente não-aleatória (KOONIN; NOVOZHILOV, 2009) (figura 1). Crick postulou que tal associação havia sido resultado de uma combinação aleatória de seus componentes que, pelo grande potencial evolutivo, prevaleceu na competição com outros sistemas e teria permanecido como o único sistema universal compartilhado entre todos os seres vivos (CRICK; 1968). Isso é conhecido como a *teoria do acidente congelado*, e hoje já é bastante contestada devido às amplas evidências de que o código genético não é literalmente universal e é passível a mudanças (KNIGHT et al., 2001; YOKOBORI et al., 2001; SANTOS et al., 2004; SENGUPTA et al., 2007).

	U	C	A	G	
U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Figura 1. O código genético.

A maleabilidade do código sugere que ele seja sujeito à evolução, e há basicamente três grandes teorias acerca de como essa evolução teria se dado, bem como a natureza e origem do código genético (KOONIN; NOVOZHILOV, 2009):

A *teoria estereoquímica* baseia-se nas relações estereoquímicas entre os aminoácidos e suas trincas cognatas. Essa teoria teve seus primeiros modelos postulados por Gamow, em 1954. Gamow propôs um mecanismo 'chave e fechadura' para os aminoácidos e os 'buracos' em forma de losango deixados por nucleotídeos no DNA (GAMOW, 1954). A partir daí, vários modelos foram elaborados utilizando a mesma base (WOESE et al., 1966; MELCHER, 1974; HENDRY et al., 1981; SHIMIZU, 1982; YARUS, 2005) e, apesar das numerosas sugestões mostrarem em geral afinidades fracas e não-específicas, pode-se argumentar que um código primordial baseado nessas conformações possa ter evoluído para um código mais preciso e elaborado (KOONIN; NOVOZHILOV, 2009).

A *teoria adaptativa* postula que a robustez do código foi modelada por pressões seletivas, num processo de minimização de erros. Há aqui uma distinção entre a pressão seletiva ser para reduzir o efeito de mutações deletérias (SONNEBORN, 1965) ou para reduzir os erros de tradução (WOESE, 1965). Também foram investigados os erros de transição e transversão, bem como os vieses de erro de tradução de cada base, separadamente (FREELAND; HURST, 1998a). Todos os estudos mostram a grande eficiência do código em sua minimização de erros, e este permanece sendo um dos mais eficientes em comparação a códigos alternativos quando se leva em conta os erros de tradução (FREELAND; HURST, 1998b).

A *teoria da coevolução* trata da origem do código genético através das vias biossintéticas dos aminoácidos, onde os códons de aminoácidos precursores teriam sido realocados para codificar os aminoácidos produtos destes precursores (WONG, 1975). A identificação dos pares precursor-produto de aminoácidos é o que torna esta teoria sensível, porém há vários trabalhos que a corroboram ao apresentar sugestões de um papel fundamental das vias metabólicas dos aminoácidos na estrutura do código (NIRENBERG et al., 1963; DI GIULIO, 2000; ARDELL; SELLA, 2002).

Naturalmente, é possível se considerar relações entre as teorias, principalmente entre as abordagens estereoquímicas e adaptativas, mas, assim como a emergência do metabolismo e da replicação, as questões fundamentais acerca da origem e evolução do

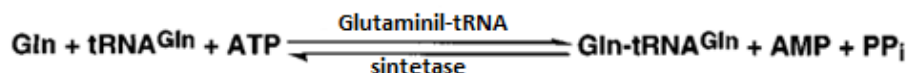
código genético permanecem em aberto até hoje. Mas há um grupo de moléculas antigas que estão relacionadas diretamente com tudo isso, e seu estudo pode revelar imprescindíveis dados a respeito do surgimento das primeiras formas de vida. Estas são as aminoacil-tRNA sintetases (aaRS) e elas são o grande foco deste trabalho.

1.3 AS AMINOACIL-tRNA SINTETASES

As aaRS são o grande elo entre proteínas e ácidos nucléicos. São enzimas de múltiplos domínios que fazem a ligação entre os tRNAs e seus respectivos aminoácidos, definindo as regras do código genético e mediando a síntese protéica. Se proteínas já existiam previamente à própria maquinaria de tradução é, como exposto anteriormente, uma das grandes discussões a respeito da emergência das primeiras formas de vida. Mas certamente, entre as primeiras proteínas pós-tradução, estariam aquelas relacionadas ao carregamento dos tRNAs. O papel central das sintetases na tradução sugere que sua história e a do código genético estejam interligadas (WOESE et al., 2000).

Existem 20 delas, cada uma associada a um dos aminoácidos e seus tRNA cognatos. As sintetases atuam normalmente por um mecanismo de duas etapas (figura 2): inicialmente, o aminoácido e uma molécula de ATP se ligam aos seus respectivos sítios e reagem, formando aminoacil-adenilato e liberando pirofosfato. O aminoácido, agora ativado, é transferido ao RNA, que também possui um sítio de ligação na sintetase, liberando AMP. Há também uma via indireta de acilação, tRNA-dependente, conhecida em alguns organismos como sendo o método essencial para se carregar certos tRNAs (CURNOW et al., 1997; CURNOW et al., 1998). Neste caso, uma sintetase não-discriminante carrega um tRNA com um aminoácido “precursor”. Este aminoácido é então convertido para o aminoácido correto por uma outra enzima (WOESE et al., 2000). Estes processos são altamente específicos, não residindo necessariamente em alguma propriedade da estrutura primária ou secundária isoladamente, e envolvem mecanismos de revisão (LING et al., 2009). A fidelidade da síntese protéica depende das aaRS, já que erros na ligação dos aminoácidos a seus RNAs transportadores correspondentes não serão corrigidos no ribossomo. As sintetases também participam de muitos outros processos além da tradução, como o processamento de RNA, síntese de rRNA e apoptose (LEE et al., 2004), observados principalmente em eucariotos. Tal versatilidade é produto da gradual incorporação de domínios ao longo do tempo evolutivo (CAETANO-ANOLLÉS et al., 2013).

Acilação direta



Modificação de aminoácido tRNA-dependente

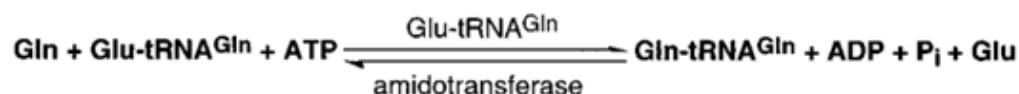
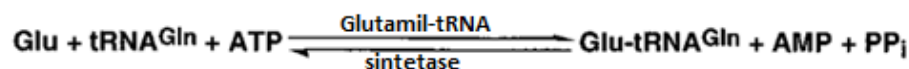


Figura 2. Esquema dos mecanismos de acilação, no caso, para a formação do glutaminil-tRNA. *B. subtilis* emprega a Glu-tRNA^{Gln} amidotransferase num processo indireto de acilação.

As sintetases são um grupo diverso de proteínas, podendo ser monoméricas, diméricas ou tetraméricas. Elas são divididas em dois grupos, cada um contendo dez delas, divididos por critérios estruturais: as sintetases de classe I possuem um domínio catalítico N-terminal, com motivos de folhas- β paralelas e α -hélices alternadas (*dobras de Rossmann*), enquanto que nas sintetases de classe II, há uma grande folha- β antiparalela circundada por α -hélices. Pertencem à classe I Arg-RS, Cys-RS, Gln-RS, Glu-RS, Ile-RS, Leu-RS, Met-RS, Trp-RS, Tyr-RS e Val-RS. Pertencem à classe II Ala-RS, Asn-RS, Asp-RS, Gly-RS, His-RS, Lys-RS, Phe-RS, Pro-RS, Ser-RS e Thr-RS. Como uma notável exceção, Lys-RS foi documentada como sendo de classe I em algumas arqueobactérias (IBBA et al. 1997). Os dois grupos são monofiléticos, praticamente não têm motivos em comum, encontram o RNA por diferentes ângulos e acilam os aminoácidos a hidroxilas diferentes (IBBA; SÖLL, 2000). O fato dos grupos não terem quase nada em comum é um dos grandes mistérios da biologia, e sugere diferentes origens evolutivas para cada classe de aaRS. É também interessante observar a simetria entre os dois grupos não-homólogos, possivelmente associada a um acoplamento de duas sintetases numa mesma haste aceptora (RIBAS DE POUPLANA; SCHIMMEL, 2001). Sun e Caetano-Anollés (2008) indicam que estudos filogenômicos da sequência e estrutura de milhares de tRNAs mostram uma origem mais tardia da

porção inferior do tRNA, que carrega o domínio de ligação ao anticódon. Desta forma, as primeiras sintetases teriam suas especificidades voltadas exclusivamente para a interação com a haste aceptora, e o sistema codificante teria surgido a partir das graduais agregações de domínios das aaRS num mundo de proteínas. Farias (2013) e Farias et al. (2014) sugerem um cenário onde o sistema codificante tenha surgido a partir da coevolução entre sintetases e tRNAs, onde mudanças na segunda base dos anticódons que resultam em mudanças em sua hidropatia teriam sido a força que induziu isso. O sistema codificante teria estado presente desde o início neste caso.

Além de todo o mistério acerca da origem de suas classes, as filogenias das aaRS também muitas vezes fogem da classificação padrão dos grandes domínios de seres vivos (DOOLITTLE; HANDY, 1998), na qual *Archaea* é o grupo mais próximo de *Eukaria* e teria dividido um ancestral comum com *Bacteria* em algum ponto. Por outro lado, algumas das enzimas são inteiramente consistentes com o modelo (DOOLITTLE; HANDY, 1998), e é bastante provável que as aaRS tenham passado por várias transferências horizontais de genes (FOURNIER et al., 2011), o que ao mesmo tempo explica tais anomalias filogenéticas e torna incrível o fato das enzimas reterem tantos traços evolutivos antigos. Provavelmente as linhagens envolvidas neste processo de transferência horizontal datam de antes do Último Ancestral Comum Universal (LUCA).

1.4 A RECONSTRUÇÃO DE ANCESTRAIS

A era genômica tornou possível o sequenciamento de proteínas e ácidos nucleicos, o que possibilitou toda a base para estudos evolutivos moleculares. Através do alinhamento de múltiplas sequências já existentes e de uma árvore filogenética correspondente, podemos produzir métodos para a reconstrução de ancestrais e gerar inferências estatísticas para testar hipóteses evolutivas, verificar a plausibilidade de modelos e inclusive sintetizar em laboratório proteínas ancestrais inferidas (CHANG; DONOGHUE, 2000). Os procedimentos de reconstrução de sequências ancestrais foram aplicados com sucesso em estudos como a análise da evolução da especificidade de substratos em proteases (CHANDRASEKHARAN et al., 1996), a recriação de uma ancestral rodopsina dos arcossauros (CHANG et al, 2002) e a paleobioquímica inferida de bactérias antigas (GAUCHER et al., 2003). Para se inferir filogenias, há três métodos principais: parsimônia, distância e máxima verossimilhança, todos revisados em

Felsenstein (1988). O avanço da bioinformática trouxe uma vasta seleção de softwares dedicados à reconstrução de sequências ancestrais que aplicam tais métodos.

As história evolutiva das aaRS pode nos fornecer informações cruciais para o entendimento das origens da relação entre proteínas e RNAs, do código genético e da vida em si. Partindo das condições dos prováveis componentes de uma Terra primordial, e através da reconstrução das formas ancestrais hipotéticas das moléculas, podemos elaborar modelos de como um primeiro organismo poderia ter se consolidado. Através de softwares e bancos de dados, pudemos gerar tanto sequências e estruturas tridimensionais de aaRS ancestrais, quanto árvores filogenéticas para cada uma de suas classes. Os dados das filogenias dão margem a várias ideias de como as moléculas mais antigas poderiam ter interagido entre si e que tipo de força poderia estar agindo para isso, e a análise dos domínios encontrados nas sequências ancestrais é um grande indicativo da presença ou ausência de uma associação inicial entre proteína e RNA. As análises das aaRS podem nos dar sugestões de uma emergência tardia do código genético, ou se isso já fosse algo consolidado desde o início.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Gerar sugestões da formação das primeiras formas de vida através das aaRS.

2.2 OBJETIVOS ESPECÍFICOS

Verificar a viabilidade das interações proteína-RNA num estado primitivo pela análise dos domínios protéicos das formas ancestrais das aaRS;

Analisar possíveis associações entre as primeiras moléculas através da filogenia das aaRS.

3. MATERIAL E MÉTODOS

3.1 DOMÍNIOS PROTÉICOS

Submetemos as sequências ancestrais de cada sintetase a uma análise de domínios pelo banco de dados BLAST. Usamos também o servidor I-TASSER (ZHANG, 2008) para fornecer a estrutura tridimensional das ancestrais. Em seguida, através da ferramenta de comparação RCSB PDB (PRLIC et al., 2010), obteve-se o alinhamento estrutural entre os domínios de ligação ao anticódon de sintetases modernas com as ancestrais, a fim de se verificar tanto a funcionalidade dos domínios ancestrais (caso presentes) quanto possíveis sites que posteriormente teriam essa função.

3.2 FILOGENIAS

O programa MEGA 5 (TAMURA et al., 2011) foi usado em todos os testes para se obter as filogenias de sequência. Após o alinhamento pelo algoritmo ClustalW, buscamos o modelo evolutivo de uma série de proteomas modernos de cada uma das sintetases pelo método da máxima verossimilhança para gerar as ancestrais. Foram 136 sequências usadas para Alanil-RS, 135 para Arginil-RS, 9 para Asparaginil-RS, 208 para Aspartil-RS, 140 para Cisteinil-RS, 139 para Glutamil-RS, 48 para Glutaminil-RS, 197 para Glicil-RS, 162 para Histidil-RS, 99 para Isoleucil-RS, 51 para Leucil-RS, 148 para Lisil-RS, 145 para Metionil-RS, 114 para Fenilalanil-RS, 125 para Prolil-RS, 142 para Seril-RS, 144 para Treonil-RS, 151 para Triptofanil-RS, 148 para Tirosil-RS e 136 para Valil-RS. Os modelos apontados foram: WAG para Ala (G: 1.03); rtREV para Arg (G: 1.19); WAG para Asn (G: 5.10); WAG para Asp (G: 0.91); WAG para Cys (G: 0.93); WAG para Gln (G: 0.96); WAG para Glu (G: 1.54, I: 0.03); WAG para Gly (G: 2.33); WAG para His (G: 1.34); WAG para Ile (G: 1.51, I: 0.05); WAG para Leu (G: 1.90); WAG para Lys (G: 1,78); WAG para Met (G: 1.39); WAG para Phe (G: 1.48); WAG para Pro (G: 1.76, I: 0.02); WAG para Ser (G: 1.07); WAG para Thr (G: 1.26, I: 0.02); WAG para Trp (G: 2.26); WAG para Tyr (G: 2.22, I: 0.03) e WAG para Val (G: 1.03). Seguindo os respectivos modelos e fazendo um teste de *bootstrap* com 1000 replicações, geramos uma árvore filogenética por máxima verossimilhança para cada uma das sintetases, a partir das quais as sequências ancestrais foram traçadas e extraídas. Quando não se foi possível gerar um único ancestral, admitiram-se dois. Geradas as ancestrais, elas foram separadas por classe para novos testes de filogenia. Os

modelos apontados foram JTT para a classe I e WAG para a classe II, e os métodos foram novamente por máxima verossimilhança e *bootstrapping* com 1000 replicações.

Após obtermos a estrutura tridimensional para cada uma das sequências ancestrais, comparamos a distância entre todas as estruturas uma a uma através da ferramenta de comparação RCSB PDB. O valor de distância é dado através do RMSD (*Root-Mean-Square Deviation*), que é a distância média do esqueleto de carbonos alfa em duas estruturas sobrepostas. Os valores obtidos foram plotados em uma matriz e a árvore de distância foi calculada com o auxílio do programa TNT (GOLOBOFF; NIXON, 2008), por parsimônia e com 1000 replicações no *bootstrap*.

4. RESULTADOS E DISCUSSÃO

4.1 ANÁLISE SEQUENCIAL

A análise dos domínios revelou domínios de ligação ao anticódon em alguns dos ancestrais. A presença de tais domínios são um indicativo de que a relação entre as proteínas e o RNA fosse algo solidificado desde o início, principalmente quando registrada em sintetases de aminoácidos considerados primitivos, como na Gly-RS (figura 3).

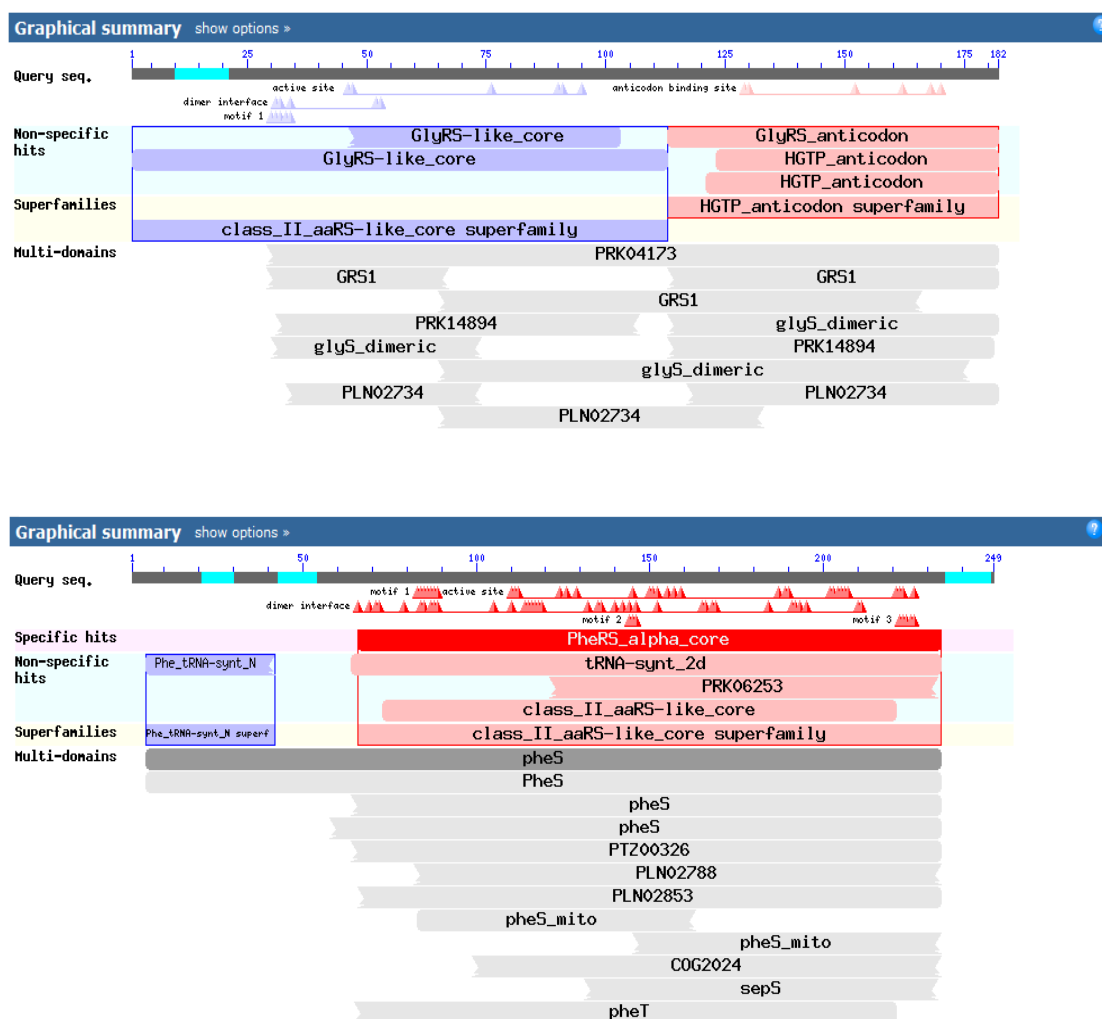


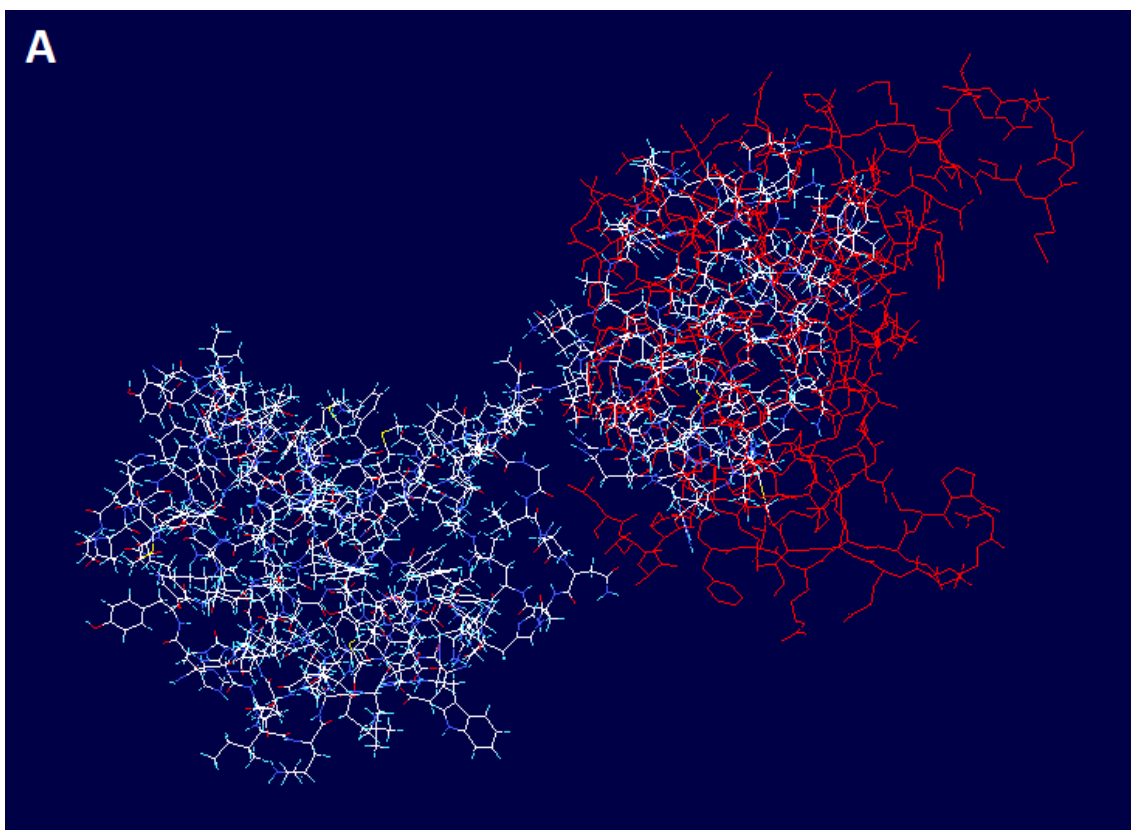
Figura 3. Presença do domínio de ligação ao anticódon em Gly-RS, mais especificamente Gly₂ (acima), e ausência do mesmo em Phe-RS (abaixo).

O domínio de ligação ao anticódon foi encontrado em Arg-RS, Ile-RS (Ile₁ e Ile₂), Met-RS e Val-RS para a classe I; e em Asn-RS, Gly-RS (Gly₂), Lys-RS, Pro-RS (Pro₂) e Thr-RS para a classe II. O domínio catalítico estava presente em todas as ancestrais, com exceção de His-RS. É bastante provável que aminoácidos fossem

abundantes na Terra primordial (MILLER, 1953; PARKER et al., 2011; BADA, 2013), e é natural pensarmos que as aaRS tenham se associado inicialmente a eles. Mas a associação com os anticódons aparentemente não é algo tão posterior quanto se pensava.

4.2 ANÁLISE ESTRUTURAL

A ocorrência de um domínio numa sequência linear não nos dá uma correspondência funcional, já que o arranjo da proteína pode pôr certas áreas em regiões inacessíveis. Para se checar o arranjo dos domínios nas ancestrais, fizemos um alinhamento entre elas e domínios de sintetases modernas. Verificada a presença do domínio na sequência primária, e com arranjo similar à conformação moderna nas ancestrais, há uma forte sugestão da emergência inicial do sistema codificante. O alinhamento na ancestral Gly_2 é evidenciado na figura 4A. A figura 4B mostra o não-alinhamento em Phe, para fins comparativos.



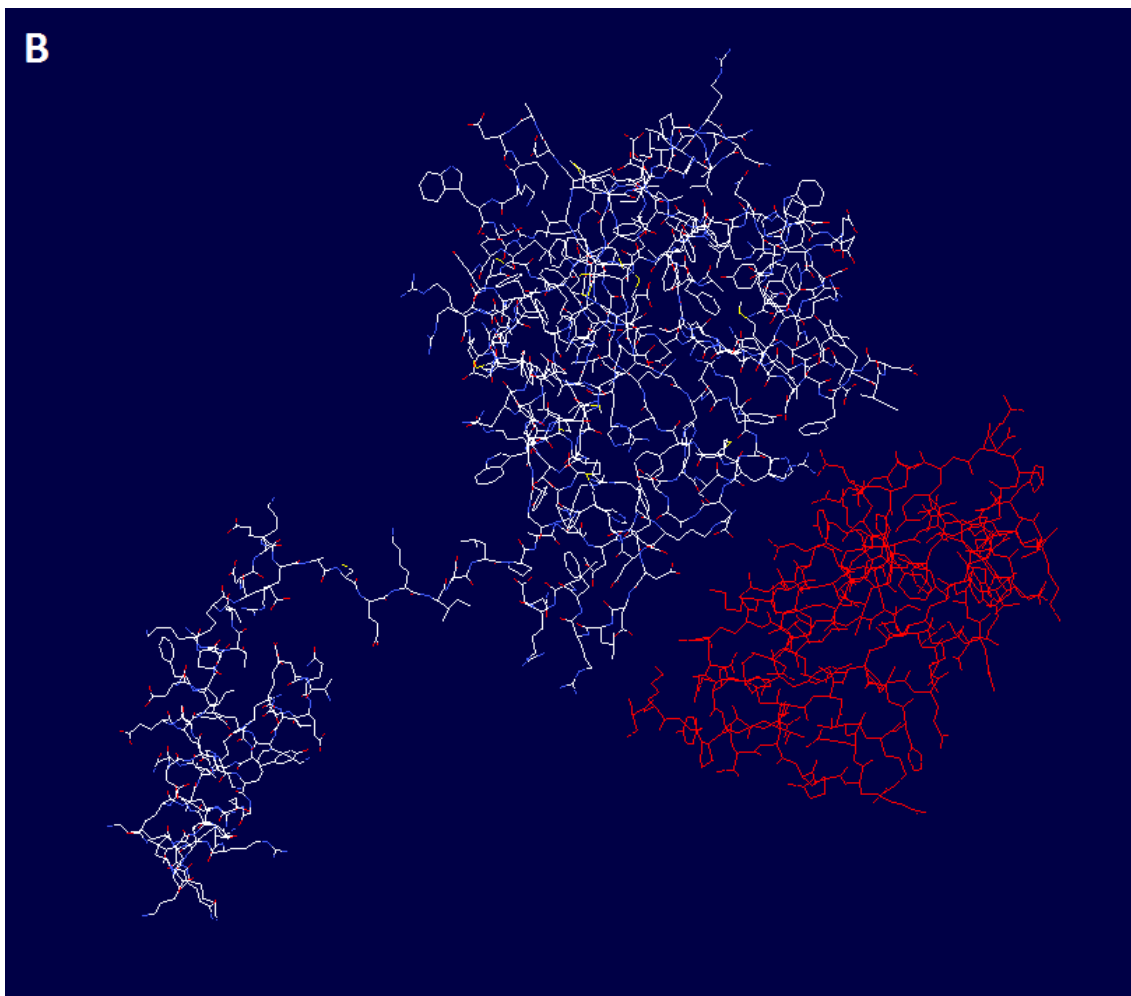


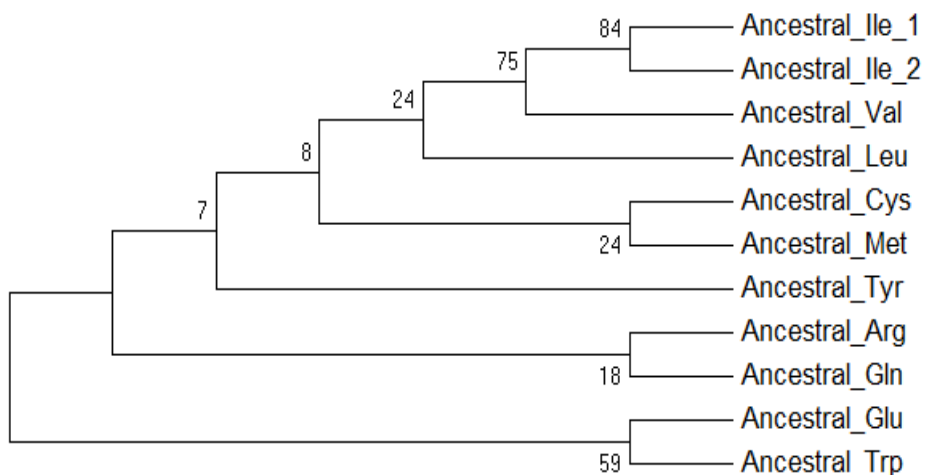
Figura 4. Alinhamento entre as ancestrais (em branco) e o domínio de ligação ao anticódon modernos (em vermelho). (A) Ancestral Gly_2 e o domínio de ligação ao anticódon de *Thermus thermophilus*, razoavelmente alinhados. (B) Ancestral Phe e o domínio de ligação ao anticódon humano mitocondrial, não alinhados.

O alinhamento ocorreu em Arg-RS, Cys-RS, Gln-RS, Ile-RS (Ile_1 e Ile_2), Met-RS e Val-RS para a classe I; e em Asn-RS, Gly-RS (Gly_2), Lys-RS e Thr-RS para a classe II. Temos então que as ancestrais Arg, Ile_1, Ile_2, Met, Val, Asn, Gly_2, Lys e Thr possuem o domínio tanto na sequência, quanto numa aproximação em estrutura. Cys e Gln só alinham em estrutura, mas não têm o domínio na sequência, enquanto Pro_2 tem o domínio na sequência, mas não alinha em estrutura. Nesses casos, interpretamos como uma consolidação inicial que se definiria melhor com o tempo evolutivo, seja por mutações, seja por rearranjos. Isto desafia a noção de que as proteínas precederam os ácidos nucleicos já que, como dito anteriormente, é bastante provável que as aaRS estejam entre as primeiras proteínas formadas. Sua vantagem seletiva é muito grande. Ribas de Pouplana et al. (1998) sugerem, através de uma

análise evolutiva da Lys-RS, que os tRNAs precederam suas sintetases, o que é mais provável. Esta segunda linha de pensamento vai de acordo com o mundo de RNA, mas dadas as complicações relacionadas a esta hipótese e o fato dos dados aqui apresentados sugerirem uma associação já inicial entre proteínas e RNA, a hipótese do mundo ribonucleoprotéico parece ainda mais plausível. Vimos que há sugestões de uma coevolução inicial e, desta forma, um sistema estaria cobrindo as pressões negativas sobre o outro, com um código genético rudimentar já estabelecido.

4.3 FILOGENIA SEQUENCIAL

As árvores filogenéticas (figura 5) mostraram vários grupos peculiares. Em busca de prováveis explicações para o que os dados revelaram, foram analisadas características estruturais, genéticas e bioquímicas tanto das sintetases quanto de seus aminoácidos correspondentes. Já é proposto que haja um padrão de diversificação entre sintetases e seus aminoácidos (NAGEL; DOOLITTLE, 1995). Houveram agrupamentos em que não se foi possível encontrar explicações satisfatórias, mas com os dados apresentados nesse e em outros trabalhos citados, temos o suficiente para propor um modelo evolutivo para a coevolução entre sintetases e tRNAs e consequente formação do código genético.



Classe I

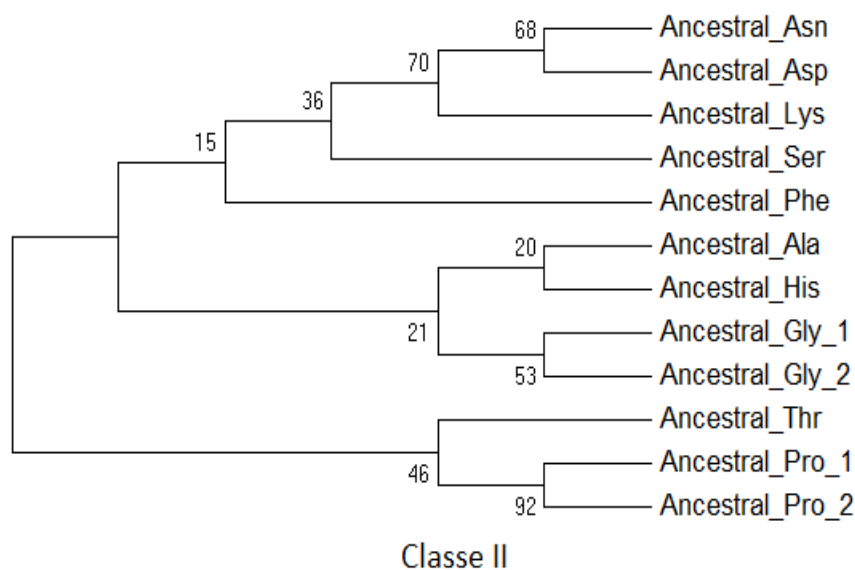


Figura 5. Árvores filogenéticas por máxima verossimilhança.

Dentro da classe I, Val-RS e Ile-RS são sintetases com sequências bastante similares (BROWN; DOOLITTLE, 1995), diferindo um pouco de Leu-RS, e as três são específicas para aminoácidos hidrofóbicos. Met-RS também tem um certo nível de similaridade com as três, embora menor. Ao olhar a tabela do código genético, podemos ver que os anticódons para Val, Ile, Leu e Met têm conformações similares. Cys-RS também aparece compondo este primeiro grupo de enzimas, sendo estruturalmente similar às outras mesmo sendo muito menor. Aparentemente os aminoácidos da Cys-RS formam uma dobra de ligação ao nucleotídeo bem mais compacta do que a de seus relativos (HOU et al., 1991). Aqui, Cys-RS encontra-se particularmente próxima a Met-RS, a outra sintetase específica pra aminoácidos que contém enxofre em sua composição e igualmente hidrofílicos. Quanto ao pareamento entre Gln-RS e Arg-RS, os anticódons de Gln para Arg diferenciam apenas em sua segunda base e ambos os aminoácidos são hidrofílicos e derivados do α -Cetogluturato. Por outro lado, não há traços como os mencionados em comum entre Glu-RS e Trp-RS. Prováveis motivos de estarem agrupadas são desconhecidos. Além disso, Tyr-RS aparece sozinha, apesar de suas notáveis semelhanças com Trp-RS (DOUBLIÉ et al., 1995).

Analisando a classe II, a árvore agrupa Asn-RS, Asp-RS e Lys-RS, todas estruturalmente similares (BERTHET-COLOMINAS et al, 1998) e de aminoácidos hidrofílicos derivados do Oxaloacetato. Em outro clado, temos um grupo com His-RS, Gly-RS, Thr-RS e Pro-RS (MARTINIS et al, 1999), também estruturalmente similares.

No primeiro grupo, a maior proximidade também é refletida em seus anticódons, e o mesmo pode se dizer do agrupamento entre Thr-RS e Pro-RS. Também vemos Ala-RS no segundo grupo próxima a Gly-RS. Além de terem anticódons similares, ambos aminoácidos são os mais simples em sua estrutura e hidrofóbicos. Já Phe-RS e Ser-RS aparecem sem pareamentos próximos. Sabemos que Phe-RS é a única codificante para um aminoácido aromático na classe II, mas Ser-RS é semelhante a Thr-RS em sua estrutura, em seus anticódons e nas características de seus aminoácidos. É estranho a enzima estar posicionada desta forma na árvore.

Neste ponto, é interessante comparar a filogenia das aaRS com a filogenia dos tRNAs. Farias (2013) propôs uma filogenia dos tRNAs por uma metodologia bastante semelhante. O trabalho dele sugere que não haja uma correlação entre a diversificação dos tRNAs e um aminoácido em particular e, portanto, a correlação entre as duas moléculas poderiam ter sido estabelecidas indiretamente por proto-aaRS. Relações de hidropatia podem ter sido o centro dessa interação inicial (FARIAS et al., 2007). Tais relações foram utilizadas por Woese et al. (1966) para demonstrar a distribuição ordenada de códons e aminoácidos, e é proposto que estejam primariamente associadas ao crescimento gradual da cadeia de aminoácidos das aaRS (FARIAS et al., 2007). Woese (1965) propõe que a associação dos aminoácidos a seus códons tenha sido o início da formação do código genético, num gradual processo de redução de erros de tradução. Isso segue a linha da teoria adaptativa do código, e consegue explicar as similaridades entre aminoácidos de códons também similares. Essencialmente, podemos ter certeza de que as primeiras células operavam com uma forma rudimentar de tradução, onde não havia tanta diferenciação entre aminoácidos. Provavelmente, em tais células, haviam várias ambiguidades na associação de códons e poucas chances de se produzir enzimas funcionais. Neste cenário, as escalas de hidropatia entre tRNAs, sintetases e aminoácidos seriam um provável ponto de partida para os pareamentos tendo em vista a notável semelhança entre este viés de associação e a forma como o código se apresenta (Woese et al., 1966), seguindo desta vez uma abordagem estereoquímica das origens do código. Também vemos na filogenia proposta aqui que, além de similaridades nas características hidropáticas dos aminoácidos e na conformação dos anticódons, sintetases cujos aminoácidos são derivados do mesmo composto também é algo que tende a aparecer nos grupos. Isso dá margem para uma interpretação coevolutiva na qual os códons tenham sido realocados à medida que um

aminoácido fosse gerado a partir de outro ou de um mesmo substrato, adquirindo mais especificidade com o tempo.

4.4 FILOGENIA ESTRUTURAL

Tendo em mente que a polaridade das moléculas é influenciada pelo arranjo de seus átomos, a filogenia gerada a partir das matrizes de distância refletem melhor o caráter estereoquímico mencionado anteriormente (figura 6).

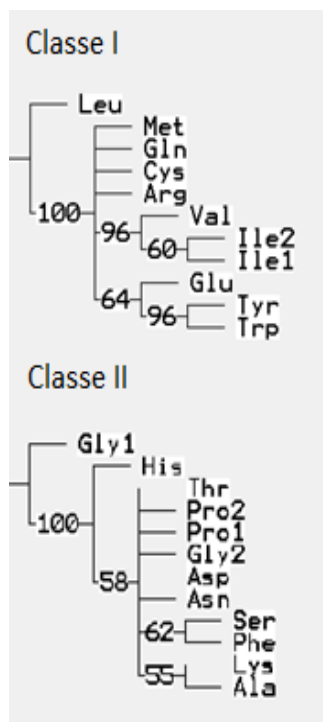


Figura 6. Filogenias estruturais por parsimônia. As ancestrais usadas como *outgroup* (Leu-RS e Gly-RS) foram escolhidas com base em características basais delas mesmas e dos aminoácidos correspondentes.

Na classe I vemos claramente que as aaRS de aminoácidos hidrofóbicos se agrupam, com Tyr-Trp em um pareamento, e Val-Ile continuando juntas em outro. Como ressaltado, Tyr-RS não aparece junto a Trp-RS na filogenia sequencial apesar de suas similaridades. Provavelmente essa homologia vem mais de características estruturais, o que é corroborado por aparecerem juntas aqui e pelo fato de que Tyr e Trp são aminoácidos aromáticos de estrutura complexa, provavelmente tendo aparecido só depois de certo tempo evolutivo. Todas as outras sintetases codificam pra aminoácidos hidrofílicos (com exceção da *outgroup* Leu).

A filogenia estrutural de classe II não mostrou padrões discerníveis; os pareamentos Ser-Phe e Lys-Ala não são congruentes com os que vimos até então. Tais convergências não ocorrerem com a classe I no que diz respeito às estruturas nos leva a crer que pressões seletivas pudessem ser diferentes de uma para outra (visto que os dois grupos são monofiléticos e têm origens evolutivas diferentes), de forma que aspectos relacionados à estrutura, como a hidropatia, ainda não repercutissem em vantagens para a classe II.

5. CONSIDERAÇÕES FINAIS

Os resultados mostrados aqui estão longe de serem conclusivos, mas procuramos cobrir tanto a origem do código quanto os primeiros passos para a emergência de uma maquinaria biológica funcional através das aaRS. A presença dos domínios de ligação ao anticódon aparentemente funcionais nas ancestrais dá força à teoria do mundo ribonucleoprotéico, com um código rudimentar já estabelecido no início das interações proteína-RNA. No modelo proposto, o código teria começado a se organizar pela associação de aminoácidos a seus códons por relações estereoquímicas de hidropatia, com baixa especificidade e alta percentagem de erros de tradução. Com o tempo, o código foi adquirindo especificidade, possivelmente pela realocação de códons à medida em que novos aminoácidos eram gerados no sistema.

6. REFERÊNCIAS

ARDELL, D. H.; SELLA G. No accident: genetic code freeze in error-correcting patterns of the standard genetic code. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 357, n. 1427, p. 1625-1642, 2002.

BADA, J. L. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. **Chemical Society Reviews**, v. 42, n. 5, p. 2186-2196, 2013.

BERTHET-COLOMINAS, C.; SEIGNOVERT, L.; HÉRTLEIN, M.; GROTLI, M.; CUSACK, S.; LEBERMAN, R. The crystal structure of asparaginyl-tRNA synthetase from *Thermus thermophilus* and its complexes with ATP and asparaginyl-adenylate: the mechanism of discrimination between asparagine and aspartic acid. **EMBO Journal**, v. 17, n. 10, p. 2947-2960, 1998.

BROWN, J. R.; DOOLITTLE, W. F. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. **Proceedings of the National Academy of Sciences of the United States of America**, v. 92, n. 7, p. 2441-2445, 1995.

CAETANO-ANOLLÉS, G.; WANG, M.; CAETANO-ANOLLÉS, D. Structural phylogenomics retrodicts the origin of the genetic code and uncovers the evolutionary impact of protein flexibility. **PLoS One**, v. 8, n. 8, p. e72225, 2013.

CECH, T. R.; BASS B. L. Biological catalysis by RNA. **Annual Review of Biochemistry**, v. 55, p. 599-629, 1986.

CHANDRASEKHARAN, U. M.; SANKER, S.; GLYNIAS, M. J.; KARNIK, S. S.; HUSAIN, A. Angiotensin II-forming activity in a reconstructed ancestral chymase. **Science**, v. 271, n. 5248, p. 502-505, 1996.

CHANG, B. S.; DONOGHUE, M. J. Recreating ancestral proteins. **Trends in Ecology and Evolution**, v. 15, n. 3, p. 109-114, 2000.

CHANG, B. S.; JÖNSSON, K.; KAZMI, M. A.; DONOGHUE, M. J.; SAKMAR, T. P. Recreating a functional ancestral archosaur visual pigment. **Molecular Biology and Evolution**, v. 19, n. 9, p. 1493-1489, 2002.

CURNOW, A. W.; HONG, K. W.; YUAN, R.; KIM, S. I.; MARTINS, O.; WINKLER, W.; HENKIN, T. M.; SÖLL, D. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 22, p. 11819-11826, 1997.

CURNOW, A. W.; TUMBULA, D. L.; PELASCHIER, J. T.; MIN, B.; SÖLL D. Glutamyl-tRNA^{Gln} amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 22, p. 12838-12843, 1998.

CRICK, F. H. The origin of the genetic code. **Journal of Molecular Biology**, v. 38, n. 3, p. 367-379, 1968.

DAMINELI, A.; DAMINELI, D. S. C. Origens da vida. **Estudos Avançados**, v. 21, n. 59, p. 263-284, 2007.

DE DUVE, C. **Blueprint for a Cell: The Nature and Origin of Life**, Burlington, Neil Patterson, 1991.

DI GIULIO, M. The origin of the genetic code cannot be studied using measurements based on the PAM matrix because this matrix reflects the code itself, making any such analyses tautologous. **Journal of Theoretical Biology**, v. 208, n. 2, p. 141-144, 2000.

DOOLITTLE, R. F.; HANDY, J. Evolutionary anomalies among the aminoacyl-tRNA synthetases. **Current Opinion in Genetics & Development**, v. 8, n. 6, p. 630-636, 1998.

DOUBLIÉ, S.; BRICOGNE, G.; GILMORE, C.; CARTER, C. W. Jr. Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. **Structure**, v. 3, n. 1, p. 17-31, 1995.

DWORKIN, J. P.; LAZCANO, A.; MILLER, S. L. The roads to and from the RNA world. **Journal of Theoretical Biology**, v. 222, n. 1, p. 127-134, 2003

FARIAS, S. T. Suggested phylogeny of tRNAs based on the construction of ancestral sequences. **Journal of Theoretical Biology**, v. 335, p. 245-248, 2013.

FARIAS, S. T.; DO RÊGO, T. G.; JOSÉ, M. V. Evolution of transfer RNA and the origin of the translation system. **Frontiers in Genetics**, v. 5, p. 303, 2014.

FARIAS, S. T.; MOREIRA, C. H.; GUIMARÃES, R. C. Structure of the genetic code suggested by the hydrophathy correlations between anticodons and amino acid residues. **Origins of Life and Evolution of the Biosphere**, v. 37, n. 1, p. 83-103, 2007.

FELSENSTEIN, J. Phylogenies from molecular sequences: inference and reliability. **Annual Review of Genetics**, v. 22, p. 521-565, 1988.

FOURNIER, G. P.; ANDAM, C. P.; ALM, E. G.; GOGARTEN, J. P. Molecular evolution of aminoacyl-tRNA synthetase proteins in the early history of life. **Origins of Life and Evolution of the Biosphere**, v. 41, n. 6, p. 621-632, 2011.

FOX, S. W.; DOSE, K. **Molecular Evolution and the Origin of Life**, New York, Marcel Dekker, 1977.

FREELAND, S. J.; HURST, L. D. Load minimization of the genetic code: history does not explain the pattern. **Proceedings of the Royal Society B: Biological Sciences**, v. 265, n. 1410, p. 2111-2119, 1998b.

FREELAND, S. J.; HURST, L. D. The genetic code is one in a million. **Journal of Molecular Evolution**, v. 47, n. 3, p. 238-248, 1998a.

GAMOW, G. Possible relation between deoxyribonucleic acid and protein structures. **Nature**, v. 173, p. 318, 1954.

GAUCHER, E. A.; THOMSON, J. M.; BURGAN, M. F.; BENNER, S. A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. **Nature**, v. 425, n. 6955, p. 285-288, 2003.

GOLOBOFF, P. A.; NIXON, F. S. K. TNT: a program for phylogenetic analysis. **Cladistics**, v. 24, p. 774-786, 2008.

GUERRIER-TAKADA, C.; GARDINER, K.; MARSH, T.; PACE, N.; ALTMAN, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. **Cell**, v. 35, n. 3 Pt 2, p. 849-857, 1983.

HARISH, A.; CAETANO-ANOLLÉS, G. Ribosomal history reveals origins of modern protein synthesis. **PLoS One**, v. 7, n. 3, p. e32776, 2012.

HOU, Y. M.; SHIBA, K.; MOTTES, C.; SCHIMMEL, P. Sequence determination and modeling of structural motifs for the smallest monomeric aminoacyl-tRNA synthetase. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 3, p. 976-980, 1991.

IBBA, M.; SÖLL, D. Aminoacyl-tRNA synthesis. **Annual Review of Biochemistry**, v. 69, p. 617-650, 2000.

IBBA, M.; MORGAN, S.; CURNOW, A. W.; PRIDMORE, D. R.; VOTHKNECHT, U. C.; GARDNER, W.; LIN, W.; WOESE, C. R.; SÖLL, D. A euryarchaeal lysyl-tRNA synthetase: resemblance to class I synthetases. **Science**, v. 278, n. 5340, p. 1119-1122, 1997.

KNIGHT, R. D.; FREELAND, S. J.; LANDWEBER, L. F. Rewiring the keyboard: evolvability of the genetic code. **Nature Reviews Genetics**, v. 2, p. 49-58, 2001.

KOONIN, E. V.; NOVOZHILOV, A. S. Origin and evolution of the genetic code: the universal enigma. **IUBMB Life**, v. 61, n. 2, p. 99-111, 2009.

KRUGER, K.; GRABOWSKI, P. J.; ZAUG, A. J.; SANDS, J.; GOTTSCHLING, D. E.; CECH, T. R. Self-splicing RNA: autoexcision and auto-cyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. **Cell**, v. 31, n. 1, p. 147-157, 1982

LEE, S. W.; CHO, B. H.; PARK, S. G.; KIM, S. Aminoacyl-tRNA synthetase complexes: beyond translation. **Journal of Cell Science**, v. 117, Pt. 17, p. 3725-3734, 2004.

LING, J.; REYNOLDS, N.; IBBA, M. Aminoacyl-tRNA synthesis and translational quality control. **Annual Review of Microbiology**, v. 63, p. 61-78, 2009.

MARTINIS, S. A.; PLATEAU, P.; CAVARELLI, J.; FLORENTZ, C. Aminoacyl-tRNA synthetases: a new image for a classical family. **Biochimie**, v. 81, n. 7, p. 683-700, 1999.

MELCHER, G. Stereospecificity of the genetic code. **Journal of Molecular Evolution**, v. 3, n. 2, p. 121-140, 1974.

- MILLER, S. L. A production of amino acids under possible primitive earth conditions. **Science**, v. 117, n. 3046, p. 528-589
- NAGEL, G. M.; DOOLITTLE, R. F. Phylogenetic analysis of the aminoacyl-tRNA synthetases. **Journal of Molecular Evolution**, v. 40, n. 5, p. 487-498, 1995.
- NIREMBERG, M. W.; JONES, W.; LEDER, P.; CLARK, B. F. C.; SLY, W. S.; PESTKA, S. On the coding of genetic information. **Cold Spring Harbor Symposium on Quantitative Biology**, v. 28, p. 549-557, 1963.
- PARKER, E. T.; CLEAVES, H. J.; DWORKIN, J. P.; GLAVIN, D. P.; CALLAHAN, M.; AUBREY, A.; LAZCANO, A.; BADA J. L. Primordial synthesis of amines and amino acids in a 1958 Miller H₂S-rich spark discharge experiment. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 14, p. 5526-5531, 2011.
- PRLIC, A.; BLIVEN, S.; ROSE, P. W.; BLUHM, W. F.; BIZON, C.; GODZIK, A.; BOURNE, P. E. Pre-calculated protein structure alignments at the RCSB PDB website. **Bioinformatics**, v. 26, n. 23, p. 2983-2985, 2010.
- RIBAS DE POUPLANA, L.; SCHIMMEL, P. Aminoacyl-tRNA synthetases: potential markers of genetic code development. **Trends in Biochemical Sciences**, v. 26, n. 10, p. 591-596, 2001.
- RIBAS DE POUPLANA, L.; TURNER, R. J.; STEER, B. A.; SCHIMMEL, P. Genetic code origins: tRNAs older than their synthetases? **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 19, p. 11295-11300, 1998.
- SANTOS, M. A. S.; MOURA, G.; MASSEY, S. E.; TUIITE, M. F. Driving change: the evolution of alternative genetic codes. **Trends in Genetics**, v. 20, n. 2, p. 95-102, 2004.
- SENGUPTA, S.; YANG, X.; HIGGS, P. G. The mechanisms of codon reassignments in mitochondrial genetic codes. **Journal of Molecular Evolution**, v. 54, n. 6, p. 662-688, 2007.
- SHIMIZU, M. Molecular basis for the genetic code. **Journal of Molecular Evolution**, v. 18, n. 5, p. 297-303, 1982.
- SONNEBORN, T. M. Degeneracy of the genetic code: extent, nature, and genetic implications. **Evolving Genes and Proteins**, p. 377-397, 1965.
- SUN, F.; CAETANO-ANOLLÉS, G. Evolutionary patterns in the sequence and structure of transfer RNA: a window into early translation and the genetic code. **PLoS One**, v. 3, n. 7, p. e2799, 2008.
- TAMURA, K.; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. **Molecular Biology and Evolution**, v. 28, n. 10, p. 2731-2739, 2011.

WOESE, C. R. On the evolution of the genetic code. **Proceedings of the National Academy of Sciences of the United States of America**, v. 54, n. 6, p. 1546-1552, 1965.

WOESE, C. R.; DUGRE, D. H.; SAXINGER, W. C.; DUGRE, S. A. Molecular basis for the genetic code. **Proceedings of the National Academy of Sciences of the United States of America**, v. 55, n. 4, p. 966-974, 1966.

WOESE, C. R.; OLSEN, G. J.; IBBA, M.; SÖLL, D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. **Microbiology and Molecular Biology Review**, v. 64, n. 1, p. 202-236, 2000.

WONG, J. T. F. A co-evolution theory of the genetic code. **Proceedings of the National Academy of Sciences**, v. 72, n. 5, p. 1909-1912, 1975.

YARUS, M. An RNA–amino acid complex and the origin of the genetic code. **New Biologist**, v. 3, n. 2, p. 183-189, 1991.

YOKOBORI, S.; SUZUKI, T.; WATANABE, K. Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. **Journal of Molecular Evolution**, v. 53, n. 4, p. 314-326, 2001.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**, v. 9, p. 40, 2008.