

PEDRO HENRIQUE LOPES FERREIRA DANTAS

**HISTÓRIA NATURAL DAS AMINOACIL-TRNA SINTETASE DE CLASSE I E SEU
PAPEL NO ESTABELECIMENTO DO CÓDIGO GENÉTICO**

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

João Pessoa

2020

PEDRO HENRIQUE LOPES FERREIRA DANTAS

**HISTÓRIA NATURAL DAS AMINOACIL-TRNA SINTETASE DE CLASSE I E SEU
PAPEL NO ESTABELECIMENTO DO CÓDIGO GENÉTICO**

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 pela
Universidade Federal da Paraíba.

Orientador: Profº Drº Sávio Torres de Farias

João Pessoa

2020

Catálogo na publicação
Seção de Catalogação e Classificação

D192h Dantas, Pedro Henrique Lopes Ferreira.

História natural das aminoacil-tRNA sintetase de classe I e seu papel no estabelecimento do código genético / Pedro Henrique Lopes Ferreira Dantas. - João Pessoa, 2020.

35 f. : il.

Orientação: Sávio Torres de Farias.

Trabalho de Conclusão de Curso na modalidade Artigo (Graduação/Bacharelado em Ciências Biológicas) - UFPB/CCEN.

1. Aminoacil-tRNA sintetase. 2. Código genético. 3. tRNA. I. Farias, Sávio Torres de. II. Título.

UFPB/CCEN

CDU 575.113.1(043.2)

PEDRO HENRIQUE LOPES FERREIRA DANTAS

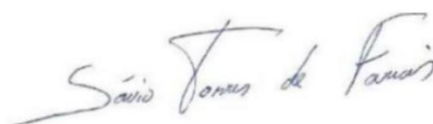
**HISTÓRIA NATURAL DAS AMINOACIL-TRNA SINTETASE DE CLASSE I E SEU
PAPEL NO ESTABELECIMENTO DO CÓDIGO GENÉTICO**

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

Data: 10/12/2020

Resultado: Aprovado

Banca examinadora:



Sávio Torres de Farias, Doutor, Universidade Federal da Paraíba



Francisco Prosdócimi, Doutor, Universidade Federal do Rio de Janeiro



Fabiano Bezerra Menegidio, Doutor, Universidade de Mogi das Cruzes

AGRADECIMENTOS

Agradeço primeiramente a minha família, especialmente meus pais, Elias e Saledilma, que sempre se esforçaram para que eu tivesse uma educação de qualidade, que me trouxe até aqui. Agradeço aos meus irmãos, João e Ana, por sempre estarem comigo, vocês são muito especiais para mim, amo vocês.

Agradeço às amizades que trago desde a escola, e que acompanharam meu trajeto, sempre carregarei vocês comigo: Bruna, Eric, Heloisa, Jéssica, Victória e Marina. Agradeço também a todas as amizades que cativei dentro da UFPB, especialmente aos amigos de turma que iniciaram esse percurso comigo, levarei vocês comigo para a vida: Eduardo, Sara, Ivo. Agradeço aos meus veteranos, que me acolheram e me proporcionaram tantas experiências maravilhosas, vocês são parte de mim: Mateus, Rafael, Jamilah, Natália, Adriennius, Catarina. Agradeço especialmente à Ana Clara, você que sempre me apoiou e esteve comigo neste projeto e em outros da minha vida, amo você de todo o coração. Agradeço também às amizades que fiz ao longo de todas as turmas do curso de Ciências Biológicas; aos meus ictiólogos preferidos, Jade e Marcus, amo vocês e agradeço por tornarem essa trajetória menos árdua; à minha querida bandinha, Igor, André e Katarina, obrigado pelo ombro amigo, vocês são muito especiais para mim; ao meu amigo gaúcho, Anderson, obrigado pela vivência e pelo companheirismo.

Sou grato a todos os professores que de alguma forma contribuíram para a minha formação. Agradeço ao professor Sávio pela oportunidade e experiência, foi enriquecedora e irei levar seus ensinamentos para minha vida acadêmica. Agradeço à Universidade Federal da Paraíba por esses 4 anos, foi uma experiência incrível em um curso que eu amo, onde aprendi tanto, não apenas acerca de ensino, pesquisa e extensão, mas também sobre as lutas que devemos sempre travar. Por fim, agradeço a mim por ter lutado até o fim nesta jornada e por ter acreditado que sou capaz.

RESUMO

O código genético é a linguagem pela qual a informação contida no DNA possa ser traduzida em proteínas. A relação da informação contida no DNA e o código genético é mantida pela maquinaria de tradução, que é responsável pela identificação do códon e associação com seu respectivo aminoácido. Dentre as proteínas envolvidas na tradução, destacam-se as aminoacil-tRNA sintetase (aaRS), enzimas responsáveis pela associação de cada aminoácido com seu respectivo tRNA. As aaRS compreendem 20 proteínas altamente conservadas, divididas simetricamente entre duas classes: classe I (ArgRS, CysRS, GlnRS, GluRS, IleRS, LeuRS, MetRS, TrpRS, TyrRS, ValRS) e classe II (AlaRS, AsnRS, AspRS, GlyRS, HisRS, LysRS, PheRS, ProRS, SerRS, ThrRS). Para uma melhor compreensão do estabelecimento do código genético, se faz necessário diversas análises a respeito da evolução dos componentes da maquinaria de tradução, uma vez que ambos estão relacionados. Dessa forma, esse trabalho se propõe a analisar a evolução molecular das aaRS de classe I, o papel desempenhado pelos tRNA e a implicação desses eventos no estabelecimento do código genético. Para tal análise, foram geradas sequências ancestrais para 9 aaRS da classe I e posteriormente essas sequências foram modeladas. Estudos de docking molecular entre as estruturas ancestrais modeladas com o loop do anticódon e o braço acceptor do tRNA cognato de cada aaRS foram conduzidos. Nossos resultados apontam para uma evolução gradual das estruturas das aaRS, com a aquisição de alguns domínios como o domínio responsável pela revisão dos aminoácidos, e o domínio de ligação aos peptídeos (CP). É hipotetizado que a aquisição dessas estruturas partiu de eventos de duplicação gênica, transferência gênica horizontal e mecanismos de troca e fusão de domínios. Além disso, as estruturas dos tRNA, quando ligadas às estruturas ancestrais, mostraram uma maior preferência pela região catalítica das aaRS, levando a crer que essa região foi a primeira a surgir nas aaRS. Em um primeiro momento, essa ligação entre o tRNA e as aaRS era forte, atuando na estabilidade das proteínas, e foi diminuída ao longo do tempo. Esse diminuição na força de ligação pode ser reflexo da plasticidade adquirida pelas duas estruturas, uma vez que ambas foram adquirindo complexidade, sendo posteriormente cooptadas para carregar informações. Dessa forma, é possível delinear um modelo onde o tRNA e as aaRS se estabilizavam mutuamente e medida que adquiriam complexidade, sendo subsequentemente cooptadas de uma função estrutural, para uma função informacional.

Palavras-chave: código genético, aminoacil-trna sintetase, trna.

ABSTRACT

The genetic code is the language whose data stored in the DNA can be translated into proteins. The relation between this stored data in the DNA and the genetic code is held by the machinery of translation, which is responsible for the identification of the codons and the association with its respective amino acid. Among the proteins involved in the translation, it is possible to highlight the aminoacyl-tRNA synthetase (aaRS), the enzymes responsible for the association of each amino acid with its respective tRNA. AaRS comprise 20 highly conserved proteins, symmetrically divided between two classes: class I (ArgRS, CysRS, GlnRS, GluRS, IleRS, LeuRS, MetRS, TrpRS, TyrRS, ValRS) and class II (AlaRS, AsnRS, AspRS, GlyRS, HisRS, LysRS, PheRS, ProRS, SerRS, ThrRS). For a better comprehension of the genetic code establishment, several analyzes are necessary regarding the translation machinery component evolution, since all of them are related. Thus, this paper proposes to analyze the class I aaRS molecular evolution, the role played by the tRNA and the implication of these events over the genetic code establishment. For the mentioned analysis, ancestral sequences for 9 aaRS of class I were generated, and later these sequences were modeled. Molecular docking studies between ancestral structures modeled with the anticodon loop and the cognate tRNA acceptor arm of each aaRS were conducted. Our results indicate the progressive evolution of the aaRS structures, with the acquisition of some domains, such as the domain responsible for the amino acid revision, and the peptide-binding domain (CP). It is hypothesized that the acquisition of these structures started from gene duplication events, horizontal gene transfer and domain exchange and merger mechanism. Besides that, tRNA structures show preference for the aaRS catalytic region when linked to ancestral structures, generating the belief that this region was the first to arise in the aaRS. At a first moment, this link between the tRNA and aaRS was firm, acting on the stability of proteins, and has been decreased over time. This decrease in bond strength can be a reflection of the plasticity acquired by the two structures, since both were acquiring complexity, being subsequently co-opted to load information. This way, it is possible to outline a model in which the tRNA and the aaRS can mutually stabilize as they acquire complexity, being subsequently co-opted from a structure function to an informational function.

KEY WORDS: genetic code, aminoacyl-trna synthetase, trna.

SUMÁRIO

1. FUNDAMENTAÇÃO TEÓRICA.....	8
2. REFERÊNCIAS.....	12
3. HISTÓRIA NATURAL DAS AMINOACIL-TRNA SINTETASE DE CLASSE I E SEU PAPEL NO ESTABELECIMENTO DO CÓDIGO GENÉTICO (Manuscrito).....	15
3.1 Introdução.....	15
3.2 Materiais e Métodos.....	17
3.3 Resultados e discussão.....	18
3.4 Referências.....	22
3.5 Figuras e legenda.....	26
3.6 Tabelas.....	31
4. CONSIDERAÇÕES FINAIS.....	34

1. Fundamentação teórica

O código genético é a linguagem utilizada pelas células para que a informação genética codificada nos ácidos nucleicos seja traduzida em proteínas. Sabe-se hoje que o código genético é compartilhado universalmente entre todas as formas de vida conhecidas e que sua composição é de 64 trincas de nucleotídeos que codificam, com pequenas variações, para os 20 aminoácidos básicos além de códons de início e de parada (NELSON; LEHNINGER; COX, 2008). Teorias a respeito da origem e evolução do código genético são extensivamente introduzidas e discutidas no campo da origem da vida.

Após o código ser determinado em 1965, Francis Crick propôs em 1968 que o código genético é universal porque qualquer alteração, após a expansão e incorporação dos 20 aminoácidos básicos, seria altamente desvantajosa e no máximo deletéria. Dessa forma, ele conclui que a atribuição dos códons à seus aminoácidos seria apenas um atributo do acaso (CRICK, 1968). Essa teoria ficou conhecida como a “teoria do acidente congelado” e é tida como um código genético padrão (CGP), tendo como contraponto o surgimento de outras variações e de novas hipóteses subsequentemente. Além da teoria do acidente congelado de Crick, é possível delinear a existência de três linhas de pensamento a respeito da origem e evolução do código genético: a teoria estereoquímica, a teoria de minimização de erros e a teoria da coevolução.

A teoria estereoquímica hipotetiza que a relação dos aminoácidos e seus respectivos códons é estabelecida por semelhanças nas propriedades físico-químicas. O primeiro modelo baseado nessa hipótese foi proposto em 1954 por Gamow, logo após a resolução da estrutura do DNA por Crick, Watson e Franklin. Em seu trabalho, Gamow correlaciona os aminoácidos com os “bolsões em forma de losango” formados entre os nucleotídeos do DNA (GAMOW, 1954). Outras versões da teoria estereoquímica surgiram subsequentemente. Pelc e Welton estabeleceram em 1966 uma correlação entre um aminoácido e um determinado códon, correlação essa que foi baseada na observação de quantas pontes de hidrogênio ou interações apolares podem ser feitas entre um dado aminoácido e um nucleotídeo na terceira posição de um códon. Baseado no número de ligações, eles estabeleceram o quão bem cada aminoácido adequa-se a seu respectivo códon (PELC; WELTON, 1966). O trabalho de Pelc e Welton foi posteriormente posto em cheque por Woese e colaboradores, que apontaram para a falta de consideração de Pelc e Welton para com as demais ligações formadas entre as pentoses e os grupamentos fosfatos do DNA (WOESE et al., 1966a). Além deste, outros trabalhos apontaram para dados não convincentes a respeito da relação físico-química entre os aminoácidos e seus

respectivos códons (WOESE, 1968; ZUBAY; DOTY, 1958). Todos esses trabalhos fizeram com que a teoria estereoquímica caminhasse a passos lentos nos anos subsequentes.

Yarus e colaboradores, empregando novas abordagens, lançaram luz sob os aspectos físico-químicos entre aminoácidos e nucleotídeos. Utilizando 337 sítios de interação de 8 aminoácidos (Arg, Gln, His, Ile, Leu, Phe, Trp, Tyr) com aproximadamente 18.500 ribonucleotídeos, os pesquisadores observaram a frequência das trincas de ribonucleotídeos na interação do RNA com seu aminoácido cognato. Neste âmbito, 5 dos 8 aminoácidos (Arg, His, Ile, Phe, Trp) mostraram uma significativa afinidade pelas suas respectivas trincas dentro dos sítios de ligação ao RNA (YARUS; WIDMANN; KNIGHT, 2009). Entretanto, Koonin & Novozhilov (2017) alertam para o fato de que dos cinco aminoácidos que mostraram significativa afinidade por suas respectivas trincas de ribonucleotídeos, quatro (Arg, His, Phe, Trp) são tidos como novos aminoácidos, incorporados recentemente no código genético, possuindo longos grupamentos laterais, o que poderia justificar afinidades observadas, deixando novamente em aberto a hipótese do estabelecimento do código genético por interações estereoquímicas.

A teoria da minimização de erros postula que o código genético padrão é robusto, pouco sujeito a mutações e erros de tradução, e que ele foi estabelecido de uma forma que viesse a reduzir os efeitos deletérios dessas alterações (MASSEY, 2008). Para estabelecer esses postulados, métricas físico-químicas (escala PR, hidropatia, volume molecular) (FARIAS; MOREIRA; GUIMARÃES, 2007; WOESE et al., 1966b) e não físico-químicas (BUHRMAN et al., 2011) foram utilizadas a fim de determinar a robustez do código como conhecemos hoje. Em 1998, Hurst e Freeland propuseram que o código genético é um em um milhão, sendo a probabilidade de que haja outro código, tão ou mais eficiente quanto o código genético padrão, de 10^{-6} (FREELAND; HURST, 1998). Estudos subsequentes seguiram tentando determinar o nível de otimização do código genético padrão, utilizando diferentes abordagens e uma combinação de métodos, o que levou ao entendimento de que o código genético padrão é robusto e bem adaptado para conter os efeitos deletérios de mutações, porém, existem outros vários códigos que são muito mais robustos (MASSEY, 2008, 2016). Esses estudos trazem esclarecimentos sobre a robustez do código genético, mas também lançam mão a respeito dos processos evolutivos que conferiram ao código genético tal característica. Massey propôs em 2015 que o código genético emergiu a partir de processos evolutivos neutros não-adaptativos (MASSEY, 2015). Tais observações foram baseadas em um modelo que postula que a informação genética era contida primordialmente apenas na segunda base dos códons, sendo posteriormente expandida para a primeira base e então para a terceira (MASSEY, 2006). Esse

modelo ficou conhecido como “2-1-3” e é suportado por estudos posteriores que chegaram à mesma conclusão, de que a segunda base dos códons foi a primeira a adquirir propriedade informativa (DE FARIAS, 2013; SUN; CAETANO-ANOLLÉS, 2008). Por fim, ao passo que Koonin e Novozhilov propõem que a robustez do código genético surgiu como um subproduto da expansão do mesmo (KOONIN; NOVOZHILOV, 2017), Firnberg e Ostermeier levantam a questão de que uma extrema taxa de minimização de erros poderia restringir o potencial evolutivo das proteínas (FIRNBERG; OSTERMEIER, 2013).

Em 1963 Nirenberg e colaboradores apontaram para uma relação entre aminoácidos de uma determinada rota biossintética e seus respectivos códons (NIRENBERG et al., 1963). Após alguns outros trabalhos explorarem essas observações (DILLON, 1973; JUKES, 1966; PELC, 1965), Wong em 1975 chegou a conclusão da importância dessa relação para a origem do código genético, estabelecendo então o que hoje é conhecido como a teoria da coevolução (WONG, 1975). Como mencionado anteriormente, essa teoria propõe que o código genético reflete o mapa das rotas biossintéticas dos aminoácidos, onde tais aminoácidos só poderiam ser incorporados no código genético à medida que se tornassem disponíveis (WONG, 1975). Em seu famoso experimento em 1953, Miller mostrou ser possível a síntese de aminoácidos a partir de condições pré bióticas (MILLER, 1953). Esse aminoácidos precursores, seriam então incorporados no código genético a medida que também serviriam para a síntese de novos produtos, no caso, novos aminoácidos, que seriam subsequentemente incorporados no código genético a partir das mudanças nos códons que codificam para os aminoácidos precursores (DI GIULIO, 2008; WONG, 1975). Em um exemplo dessa teoria, Taylor e Coates indicaram uma relação entre as vias bioquímicas glicolíticas e do ácido cítrico com a incorporação de alguns aminoácidos no código genético, onde os aminoácidos codificados pelos códons GNN (Gli, Ala, Val, Asp, Glu) aparecem como os primeiros produtos das suas respectivas rotas metabólicas (TAYLOR; COATES, 1989). Com exceção da prolina, esses aminoácidos também constam como os primeiros a serem incorporados no código genético, de acordo com análises cronológicas posteriores (BERNHARDT; PATRICK, 2014; TAMURA, 2015; TRIFONOV, 2004). Utilizando filogenias de sequências ancestrais de tRNA, Di Giulio também mostrou que a teoria da coevolução é melhor suportada em comparação com a teoria estereoquímica (DI GIULIO, 1995).

O estabelecimento do código genético segue sendo um terreno fértil para debates e discussões, porém Koonin e Novozhilov levantam uma crítica, pontuando que a emergência do código genético e a evolução do sistema de tradução não podem ser discutidos separadamente, sendo que a evolução do código genético só pode ser entendida à luz da evolução do sistema

de tradução (KOONIN; NOVOZHILOV, 2017). Ainda que alguns pesquisadores questionem o papel de componentes da maquinaria de tradução (e.g. aminoacil-tRNA sintetase) no estabelecimento do código genético (DI GIULIO, 2017), mais análises se fazem necessárias para a compreensão de como a maquinaria de tradução atuou no estabelecimento do código genético. Sabendo disso, neste trabalho propomos uma ampla análise da evolução das aminoacil-tRNA sintetase e sua relação com seus tRNAs cognatos e com o código genético.

2. Referências

- BERNHARDT, H. S.; PATRICK, W. M. Genetic Code Evolution Started with the Incorporation of Glycine, Followed by Other Small Hydrophilic Amino Acids. **Journal of Molecular Evolution**, v. 78, n. 6, p. 307–309, jun. 2014.
- BUHRMAN, H. et al. Some Mathematical Refinements Concerning Error Minimization in the Genetic Code. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, v. 8, n. 5, p. 1358–1372, set. 2011.
- CRICK, F. H. C. The origin of the genetic code. **Journal of Molecular Biology**, v. 38, n. 3, p. 367–379, dez. 1968.
- DE FARIAS, S. T. Suggested phylogeny of tRNAs based on the construction of ancestral sequences. **Journal of Theoretical Biology**, v. 335, p. 245–248, out. 2013.
- DI GIULIO, M. The phylogeny of trnas seems to confirm the predictions of the coevolution theory of the origin of the genetic code. **Origins of life and evolution of the biosphere**, v. 25, n. 6, p. 549–564, dez. 1995.
- DI GIULIO, M. An extension of the coevolution theory of the origin of the genetic code. **Biology Direct**, v. 3, n. 1, p. 37, 2008.
- DI GIULIO, M. The aminoacyl-tRNA synthetases had only a marginal role in the origin of the organization of the genetic code: Evidence in favor of the coevolution theory. **Journal of Theoretical Biology**, v. 432, p. 14–24, nov. 2017.
- DILLON, L. S. The origins of the genetic code. **The Botanical Review**, v. 39, n. 4, p. 301–345, out. 1973.
- FARIAS, S. T. DE; MOREIRA, C. H. C.; GUIMARÃES, R. C. Structure of the genetic code suggested by the hydropathy correlation between anticodons and amino acid residues. **Origins of Life and Evolution of Biospheres**, v. 37, n. 1, p. 83–103, 1 fev. 2007.
- FIRNBERG, E.; OSTERMEIER, M. The genetic code constrains yet facilitates Darwinian evolution. **Nucleic Acids Research**, v. 41, n. 15, p. 7420–7428, ago. 2013.
- 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, set. 1998.

GAMOW, G. Possible Relation between Deoxyribonucleic Acid and Protein Structures. **Nature**, v. 173, n. 4398, p. 318–318, fev. 1954.

JUKES, T. H. Molecules and evolution. 1966.

KOONIN, E. V.; NOVOZHILOV, A. S. Origin and Evolution of the Universal Genetic Code. **Annual Review of Genetics**, v. 51, p. 45–62, 27 2017.

MASSEY, S. Genetic Code Evolution Reveals the Neutral Emergence of Mutational Robustness, and Information as an Evolutionary Constraint. **Life**, v. 5, n. 2, p. 1301–1332, 24 abr. 2015.

MASSEY, S. E. A Sequential “2-1-3” Model of Genetic Code Evolution That Explains Codon Constraints. **Journal of Molecular Evolution**, v. 62, n. 6, p. 809–810, jun. 2006.

MASSEY, S. E. A Neutral Origin for Error Minimization in the Genetic Code. **Journal of Molecular Evolution**, v. 67, n. 5, p. 510–516, nov. 2008.

MASSEY, S. E. The neutral emergence of error minimized genetic codes superior to the standard genetic code. **Journal of Theoretical Biology**, v. 408, p. 237–242, nov. 2016.

MILLER, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. **Science**, v. 117, n. 3046, p. 528–529, 15 maio 1953.

NELSON, D. L.; LEHNINGER, A. L.; COX, M. M. **Lehninger principles of biochemistry**. [s.l.] Macmillan, 2008.

NIRENBERG, M. W. et al. On the Coding of Genetic Information. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 28, n. 0, p. 549–557, 1 jan. 1963.

PELC, S. R. Correlation between Coding-Triplets and Amino-Acids. **Nature**, v. 207, n. 4997, p. 597–599, 7 ago. 1965.

PELC, S. R.; WELTON, M. G. E. Stereochemical Relationship Between Coding Triplets and Amino-Acids. **Nature**, v. 209, n. 5026, p. 868–870, fev. 1966.

SUN, F.-J.; 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, 30 jul. 2008.

TAMURA, K. Beyond the Frozen Accident: Glycine Assignment in the Genetic Code. **Journal of Molecular Evolution**, v. 81, n. 3–4, p. 69–71, out. 2015.

TAYLOR, F. J. R.; COATES, D. The code within the codons. **Biosystems**, v. 22, n. 3, p. 177–187, jan. 1989.

TRIFONOV, E. N. The Triplet Code From First Principles. **Journal of Biomolecular Structure and Dynamics**, v. 22, n. 1, p. 1–11, ago. 2004.

WOESE, C. R. et al. On the Fundamental Nature and Evolution of the Genetic Code. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 31, n. 0, p. 723–736, 1 jan. 1966a.

WOESE, C. R. et al. The molecular basis for the genetic code. **Proceedings of the National Academy of Sciences**, v. 55, n. 4, p. 966–974, 1 abr. 1966b.

WOESE, C. R. The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. **Proceedings of the National Academy of Sciences**, v. 59, n. 1, p. 110–117, 1 jan. 1968.

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, 1 maio 1975.

YARUS, M.; WIDMANN, J. J.; KNIGHT, R. RNA–Amino Acid Binding: A Stereochemical Era for the Genetic Code. **Journal of Molecular Evolution**, v. 69, n. 5, p. 406–429, nov. 2009.

ZUBAY, G.; DOTY, P. Nucleic acid interactions with metal ions and amino acids. **Biochimica et Biophysica Acta**, v. 29, n. 1, p. 47–58, jul. 1958.

3. ARTIGO CIENTÍFICO

HISTÓRIA NATURAL DAS AMINOACIL-TRNA SINTETASE DE CLASSE I E SEU PAPEL NO ESTABELECIMENTO DO CÓDIGO GENÉTICO

Pedro Henrique Lopes Ferreira Dantas¹ e Sávio Torres de Farias¹

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

3.1 Introdução

Nos sistemas biológicos os ácidos nucleicos são responsáveis por armazenar a informação genética que pode ser posteriormente traduzida em proteínas. Esses dois componentes, proteínas e ácidos nucleicos, são cruciais para o perfeito funcionamento dos organismos vivos, e a correspondência entre ambos é garantida pelo código genético, que por sua vez é amparado por processos de transcrição e de tradução.

A tradução é um dos processos mais bem conservados ao longo da evolução, e não obstante, seus componentes também o são. Uma das peças centrais no processo de tradução, presente em todos os domínios da árvore da vida, são as aminoacil-tRNA sintetases (aminoacil-tRNA ligase, aaRS), enzimas responsáveis pela ativação e ligação de um aminoácido ao seu respectivo tRNA. Esse processo, a aminoacilação, é catalisado em duas etapas, onde primeiro o aminoácido é ativado por uma molécula de ATP e logo em seguida é transferido para a extremidade 3' do tRNA cognato (CRNKOVIĆ; VARGAS-RODRIGUEZ; SÖLL, 2019).

Por várias décadas, as aaRS têm sido alvo de diversas hipóteses a respeito de sua evolução, organização estrutural e relação com o código genético. As aaRS foram divididas em duas classes estruturalmente e evolutivamente não relacionadas, cada uma contendo 10 proteínas. Essa divisão foi baseada em correlações estruturais e funcionais observadas no modo como as aaRS de cada classe aminoacilam o braço acceptor do tRNA (ERIANI et al., 1990). Estudos filogenéticos conduzidos posteriormente suportam essa divisão e lançam uma luz a respeito dos fatores que influenciaram na evolução e diversificação dessas duas classes de proteínas (NAGEL; DOOLITTLE, 1991).

É proposto que a aminoacilação era um papel desempenhado por moléculas de RNA catalisadores com pouca especificidade e poder catalítico, mas que realizavam esse papel antes do surgimento das aaRS. De fato, alguns estudos suportam a ideia de que as estruturas determinantes para a aminoacilação, o braço acceptor do tRNA e a loop do anticódon, já estavam

presentes antes do surgimento das aaRS (RIBAS DE POUPLANA; SCHIMMEL, 2001). Esses RNAs precursores teriam dado lugar à enzimas ancestrais que catalisariam a ativação de um grande número de moléculas de aminoácidos que ficariam disponíveis para os RNAs utilizarem. Outro progenitor também poderia ter surgido para dar suporte na ativação de aminoácidos que não pudessem ser ativados por essa primeira enzima. A evolução e diversificação desses dois progenitores de aaRS então seria guiada por eventos subsequentes de duplicação gênica que aperfeiçoaram a especificidade e a capacidade catalítica (NAGEL; DOOLITTLE, 1995).

Entretanto, estudos recentes têm contestado a relevância da duplicação gênica na evolução das aaRS. Treangen e Rocha propuseram que, ao menos em procariotos, a transferência gênica horizontal (HGT) tem um papel mais relevante na expansão das famílias proteicas, ao contrário da duplicação gênica (TREANGEN; ROCHA, 2011). Essa transferência gênica horizontal teria sido seguida de diversos outros eventos como a recombinação, troca e fusão de domínios (SAHA et al., 2009).

As aaRS são proteínas muito susceptíveis à pressões evolutivas devido a sua natureza independente do restante da maquinaria celular, tendo essas proteínas contato direto apenas com seus principais substratos, o ATP, os aminoácidos e o tRNA. De certa forma então é de se esperar que esses substratos exerçam alguma pressão seletiva na diversificação das aaRS. É hipotetizado que um dos determinantes para a diversificação das aaRS pode ter sido a sua relação com os aminoácidos e como a característica química destes dirigiram a evolução ao menos da região N-terminal das aaRS, responsável pelo reconhecimento e ativação dos aminoácidos em estruturas modernas (DE FARIAS; GUIMARÃES, 2008; NAGEL; DOOLITTLE, 1995). Além disso, estudos recentes sugerem que as características físico-químicas dos aminoácidos tiveram um papel mais relevante no estabelecimento do código genético do que as aaRS, tendo essas realizado apenas um papel intermediário devido a sua relação com os aminoácidos (DI GIULIO, 2017). É argumentado que a estrutura do tRNA, outra molécula que interage com as aaRS, evoluiu a partir de eventos de ligação e deleção entre proto-tRNAs (ROOT-BERNSTEIN et al., 2016). Outra hipótese é de que o tRNA é fruto de modificações como a adição e duplicação de nucleotídeos, eventos esses que partiram da extremidade -CCA do braço aceptor, região responsável pela ligação ao aminoácido (SUN; CAETANO-ANOLLÉS, 2008), colocando assim em xeque a hipótese de que o tRNA teria surgido a partir do braço do anticódon (SZATHMÁRY, 1999). Definido como código operacional, diversos outros estudos, baseados na observação de que a aminoacilação pode ocorrer sem a necessidade do reconhecimento do anticódon, suportam esse modelo onde a região de ligação ao aminoácido no braço aceptor teria surgido antes da loop do anticódon

(HAMANN; HOU, 1995; HOU; SCHIMMEL, 1988; PARK; SCHIMMEL, 1988). Nesse contexto, (CAETANO-ANOLLÉS; WANG; CAETANO-ANOLLÉS, 2013) propuseram que nas aaRS, a região catalítica teria surgido antes da região de ligação ao anticódon. Nesse cenário, as aaRS primitivas eram limitadas à conservação da região catalítica e de reconhecimento, que seriam posteriormente cooptadas para atuar em processos mais complexos na maquinaria de tradução.

Farias e colaboradores propuseram um modelo onde o tRNA e as aaRS co-evoluíram a partir de mudanças na segunda base do anticódon, mudança essa que alteraria as propriedades estereoquímicas do anticódon, criando uma pressão seletiva para a diversificação das aaRS (DE FARIAS, 2013; DE FARIAS et al., 2019; DE FARIAS; DO RÊGO; JOSÉ, 2014).

Assim sendo, neste trabalho nós analisamos a história evolutiva das aaRS de classe I a partir de observações estruturais de sequências ancestrais, sua relação com as regiões acceptoras e do anticódon do tRNA, como essa relação se estabeleceu e quais as implicações na determinação do código genético e do sistema de tradução.

3.2 Materiais e Métodos

3.2.1 Banco de dados

Foram obtidas 4533 sequências do GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) referentes às 9 aaRSs. A quantidade de sequências compreende proporcionalmente, para cada uma das 9 aaRS, os 3 domínios da vida (Archaea, Bacteria e Eukarya).

3.2.2 Reconstrução de sequências ancestrais

Para a reconstrução de sequências ancestrais, inicialmente foi realizado o alinhamento das sequências obtidas para cada aaRS, através do software MAFFT (Versão 7.0) utilizando a matriz de substituição BLOSUM45 com uma penalidade de 1.53 para aberturas de gaps (<https://mafft.cbrc.jp/alignment/server/index.html>) (KATO; ROZEWICKI; YAMADA, 2019). Um estudo comparando ferramentas de alinhamento múltiplos apontou o MAFFT como o melhor software para esse tipo de abordagem (VIALLE; TAMURI; GOLDMAN, 2018). Para determinar o melhor modelo evolutivo, o alinhamento foi submetido ao software ModelTest-NG (Versão 0.1.5) (DARRIBA et al., 2020) hospedado no servidor Cipres (<https://www.phylo.org/>). Para reconstrução de sequências ancestrais, Hanson e colaboradores concluíram o melhor método para tal abordagem é a máxima verossimilhança (ML) em comparação com análises bayesianas (BA) (HANSON-SMITH; KOLACZKOWSKI;

THORNTON, 2010). Dessa forma, as árvores filogenéticas foram inferidas utilizando o software IQ-Tree (<http://www.iqtree.org/>) (Versão 1.6.12) (NGUYEN et al., 2015). As sequências ancestrais foram obtidas por meio do software MEGAX (Versão 10.0) a partir do método de máxima verossimilhança (KUMAR et al., 2018).

3.2.3 Modelagem molecular

Os modelos tridimensionais das proteínas ancestrais foram gerados por meio do servidor I-Tasser (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) (ROY; KUCUKURAL; ZHANG, 2010). O refinamento das estruturas foi feito através do servidor GalaxyWeb (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>) (KO et al., 2012). O alinhamento estrutural dos modelos tridimensionais foi performado com o auxílio do servidor TM-Align (<https://zhanglab.ccmb.med.umich.edu/TM-align/>) (ZHANG, 2005). Para os alinhamentos estruturais, foram escolhidas as estruturas de aaRS modernas, obtidas no banco de dados PDB ([rcsb.org](https://www.rcsb.org/)) (BERMAN, 2000), com a melhor resolução e associadas ao seu tRNA cognato. A validação da estrutura tridimensional das proteínas modeladas foi feito por meio do servidor PROCHECK (<https://servicesn.mbi.ucla.edu/PROCHECK/>) (LASKOWSKI et al., 1993).

3.2.4 Docking molecular

Para as análises de docking molecular, foram utilizadas as estruturas tridimensionais dos tRNAs cognatos, especificamente do loop do anticódon e da região do braço aceptor. O docking molecular foi então realizado entre o braço aceptor e do anticódon com a sua respectiva proteína ancestral. Para o docking, foi utilizado o programa Hex 8.0.0 (RITCHIE; VENKATRAMAN, 2010) de acordo com os seguintes parâmetros: tipo de correlação - shape+electro+DARS, FFT mode - 5 dimensões, método de amostragem - range angles, pós-processamento - DARS minimization. Os demais parâmetros foram utilizados na configuração padrão.

3.3 Resultados e Discussão

3.3.1 Evolução estrutural das aaRS de Classe I

As sequências ancestrais foram geradas a partir de sequências modernas obtidas para cada uma das 9 aaRSs (Tabela 1). Foram utilizados dois pontos de cortes para a reconstrução das sequências ancestrais. O primeiro ponto de corte com deleção completa retorna uma sequência ancestral com as posições alinhadas em todas as sequências modernas (100%),

enquanto o ponto de corte de deleção parcial corresponde às posições alinhadas em 99% das sequências modernas. Para cada ponto de corte, foram geradas duas sequências ancestrais, resultando em 4 sequências ancestrais para cada aaRS.

As aaRS de classe I são proteínas altamente conservadas, com pequenas variações ao longo dos domínios da vida, possuindo uma região catalítica que se liga ao braço aceptor do tRNA e uma região de ligação ao anticódon, tendo também como principal característica dois motivos estruturalmente conservados, HIGH e KMSKS, além de um domínio Rossmann-fold presente no sítio catalítico das enzimas (CUSACK; HÄRTLEIN; LEBERMAN, 1991). Dentre as estruturas das sequências ancestrais da deleção completa referente às 9 aaRS analisadas neste trabalho (ArgRS, CisRS, GluRS, IleRS, LeuRS, MetRS, TirRS, TrpRS, ValRS), 8 delas, com exceção da TrpRS, tiveram a região catalítica conservada. Em contrapartida, 6 das 9 aaRS analisadas tiveram a região de ligação ao loop do anticódon conservadas (Figura 1). Esses achados corroboram estudos prévios que atestaram o sítio catalítico como o primeiro domínio das aaRS a surgir, tendo os demais domínios surgido posteriormente a partir dessa região catalítica (CAETANO-ANOLLÉS; WANG; CAETANO-ANOLLÉS, 2013). De fato, já foi comprovado que aaRS com ausência de uma sítio de ligação ao anticódon se mostraram capazes de catalisar o processo de aminoacilação, ainda que isso afetasse a capacidade da aaRS de discriminar tRNAs não cognatos (SCHWOB; SÖLL, 1993). As estruturas geradas a partir de sequências da deleção parcial (99%) mostram um aumento na complexidade das aaRS, a partir do surgimento de novas estruturas, como a região aceptor do loop do anticódon e a região de conexão ao peptídeo (CP) (Figura 2). A aquisição dessas novas estruturas pode ter se dado após diversos eventos de transferência gênica horizontal. Em estudos comparando genomas de procariotos, foi visto que genes transferido -xenólogos- estão relacionados com a aquisição de novas funcionalidades e tendem a ter uma evolução mais rápida, enquanto genes duplicados -parálogos- evoluem lentamente e são relacionados com a dosagem gênica (TREANGEN; ROCHA, 2011). Diversos estudos já apontaram para a relevância da transferência gênica horizontal na história evolutiva das aaRS (WOESE et al., 2000). Após esses eventos de transferência gênica horizontal, as aaRS estariam mais susceptíveis à diferentes pressões seletivas, que levariam à ocorrência de outras modificações por meio de recombinações entre genes que codificam as aaRS. Isso explicaria, por exemplo, a presença de um domínio DALR, comumente encontrado em aaRS de classe I, em GliRS, uma enzima pertencente à classe II (WOLF et al., 1999). Além da recombinação, as aaRS estariam sujeitas a processos de troca e fusão de domínios. Esses processos levariam à criação de variáveis quiméricas das aaRS, que poderiam apresentar alguma vantagem evolutiva (SAHA et al., 2009). Isso foi visto em estudos,

onde uma proteína quimérica produzida a partir de outras duas proteínas com propriedades e atividades semelhantes entre si tiveram uma maior estabilidade e atividade catalítica (NIXON; WARREN; BENKOVIC, 1997). Todos esses eventos levaram a um aumento da complexidade das aaRS, complexidade essa que consequentemente contribuiu para o ganho de especificidade de cada aaRS por seu tRNA e aminoácido cognato.

3.3.2 Coevolução entre as aaRS de classe I e o tRNA

O docking molecular foi performedo entre as estruturas tridimensionais da deleção completa e da deleção parcial com o loop do anticódon e o braço acceptor do tRNA cognato de cada aaRS. Para o docking, toda a superfície das proteínas foi considerada, a fim de que o braço acceptor e o loop do anticódon encontrassem a região de ligação com menor energia. As energias de ligação foram compiladas na tabela 2.

De acordo com nossos resultados, o loop do anticódon e o braço acceptor do tRNA, com apenas duas exceções (LeuRS e ValRS com o braço acceptor do tRNA), se ligaram com uma menor energia livre nas estruturas da deleção completa, enquanto que nas estruturas da deleção parcial, houve um aumento da energia livre (tabela 2). Uma menor energia livre entre as aaRS com o braço acceptor do tRNA e o loop do anticódon nos leva a hipotetizar que inicialmente, o tRNA e as aaRS se ligavam mais fortemente. Em um trabalho avaliando a evolução estrutural da GliRS, Farias e colaboradores teorizam que o loop do anticódon agia primitivamente como um estabilizador conformacional para o sítio catalítico da GliRS (DE FARIAS et al., 2019), o que poderia justificar a forte ligação entre as estruturas da deleção completa das aaRS com o tRNA em nossas análises. Outra observação feita diz respeito à característica estrutural das regiões onde o tRNA se ligou nas aaRS. Grande parte das regiões das proteínas que interagiram com o loop do anticódon do tRNA não assumiam nenhuma conformação secundária, seja folha-beta ou alfa-hélice. Essas estruturas características desempenham um importante papel na estabilidade conformacional da proteína a partir das ligações de hidrogênio entre os aminoácidos adjacentes (NELSON; LEHNINGER; COX, 2008). Portanto é possível hipotetizar que essas interações tRNA-aaRS em certo ponto podem ter auxiliado a proteína a se estabilizar, na ausência de uma estrutura secundária nativa. Esse papel desempenhado pelo loop do anticódon do tRNA faria sentido em um cenário onde ainda não havia a presença de um código genético, fato corroborado por estudos que afirmam que primitivamente o ribossomo produzia apenas peptídeos aleatórios (AGMON, 2009; BELOUSOFF et al., 2010).

Neste trabalho, avaliamos também a ligação às aaRS do braço acceptor e do loop do anticódon do tRNA. Diversos trabalhos tratam a respeito da origem de ambas as regiões, e

discutem qual região emergiu primeiro no tRNA. A hipótese mais aceita atualmente atesta para o código operacional, onde o braço acceptor é tido como a primeira região a surgir nos tRNA. Essa hipótese é defendida por inúmeros estudos (HOU; SCHIMMEL, 1988; PARK; SCHIMMEL, 1988; SCHWOB; SÖLL, 1993), enquanto que outros atestam para o surgimento inicial do loop do anticódon (SZATHMÁRY, 1999). O resultado do docking molecular nos mostrou que em 7 das 9 aaRS, o braço acceptor e o loop do anticódon se ligaram praticamente na mesma região. Baseado nesses resultados, nós propomos que o código operacional e o código do anticódon coexistiram, competindo entre si pela ligação à região catalítica das aaRS, enquanto consequentemente atuavam na estabilização dessas proteínas. A região catalítica das aaRS então serviu como condicionante para que eventos de fusão entre o braço acceptor e o loop do anticódon ocorressem, eventos esses semelhantes aos hipotetizados por Root-Bernstein e colaboradores (ROOT-BERNSTEIN et al., 2016). Esse condicionamento do braço acceptor e do loop do anticódon em uma mesma região também pode refletir uma estrutura primitiva dos tRNAs, como atestam Möller e Janssen, que em 1992, baseados na análise estatística de 1400 sequências de tRNA, propuseram que o braço acceptor e o loop do anticódon surgiam a partir de uma única região nas posições 3-5 de tRNAs primitivos, dessa forma, tendo uma natureza química semelhante, um “imprinting” do loop do anticódon no braço acceptor dos tRNAs. Esses tRNAs primitivos então se expandiram a partir de eventos de duplicação e clivagem, auxiliando também no estabelecimento do código genético nos mRNAs (MÖLLER; JANSSEN, 1992). Esse aumento na complexidade dos tRNAs até as estruturas atuais foi acompanhado por um aumento na energia livre entre as aaRS e os tRNAs (tabela 2), garantindo certa plasticidade na relação estrutural de ambos, e que poderia também ser reflexo da cooptação do tRNA para que este passasse a atuar no fluxo da informação. Essa expansão estrutural dos tRNAs também teria atuado como força seletiva para a diversificação estrutural das aaRS, que após se estabelecerem conformacionalmente, sofreriam nova pressão seletiva a partir de mudanças na segunda base do anticódon, como proposto por Farias e colaboradores (DE FARIAS, 2013; DE FARIAS; DO RÊGO; JOSÉ, 2014). Dessa forma, nós propomos que a complexidade do tRNA e das aaRS emergiu em um processo de coevolução, tendo o tRNA atuado inicialmente como um estabilizador conformacional das aaRS, sendo posteriormente cooptado para atuar no fluxo da informação e no estabelecimento do processo de tradução e do código genético.

Informação adicional

- Interesses financeiros concorrentes

Os autores declaram não haver interesses financeiros concorrentes.

- Conflito de interesse

Nenhum

3.4 Referências

AGMON, I. The Dimeric Proto-Ribosome: Structural Details and Possible Implications on the Origin of Life. **International Journal of Molecular Sciences**, v. 10, n. 7, p. 2921–2934, 30 jun. 2009.

BELOUSOFF, M. J. et al. On the development towards the modern world: A plausible role of uncoded peptides in the rna-world. **Origins of Life and Evolution of the Biosphere**, v. 40, n. 4, p. 415–420, 2010.

BERMAN, H. M. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 1 jan. 2000.

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, 21 ago. 2013.

COX, Michael M.; NELSON, David L. **Lehninger principles of biochemistry**. Wh Freeman, 2008.

CRNKOVIĆ, A.; VARGAS-RODRIGUEZ, O.; SÖLL, D. Plasticity and Constraints of tRNA Aminoacylation Define Directed Evolution of Aminoacyl-tRNA Synthetases. **International Journal of Molecular Sciences**, v. 20, n. 9, p. 2294, 9 maio 2019.

CUSACK, S.; HÄRTLEIN, M.; LEBERMAN, R. Sequence, structural and evolutionary relationships between class 2 aminoacyl-tRNA synthetases. **Nucleic Acids Research**, v. 19, n. 13, p. 3489–3498, 1991.

DARRIBA, D. et al. ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. **Molecular Biology and Evolution**, v. 37, n. 1, p. 291–294, 1 jan. 2020.

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

DE FARIAS, S. T. et al. Structural evolution of Glycyl-tRNA synthetases alpha subunit and its implication in the initial organization of the decoding system. **Progress in Biophysics and Molecular Biology**, v. 142, p. 43–50, mar. 2019.

DE 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, 28 ago. 2014.

DE FARIAS, S. T.; GUIMARÃES, R. C. Aminoacyl-tRNA synthetase classes and groups in prokaryotes. **Journal of Theoretical Biology**, v. 250, n. 2, p. 221–229, jan. 2008.

DI GIULIO, M. The aminoacyl-tRNA synthetases had only a marginal role in the origin of the organization of the genetic code: Evidence in favor of the coevolution theory. **Journal of Theoretical Biology**, v. 432, p. 14–24, nov. 2017.

ERIANI, G. et al. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. **Nature**, v. 347, n. 6289, p. 203–206, set. 1990.

HAMANN, C. S.; HOU, Y.-M. Enzymic Aminoacylation of tRNA Acceptor Stem Helices with Cysteine Is Dependent on a Single Nucleotide. **Biochemistry**, v. 34, n. 19, p. 6527–6532, 16 maio 1995.

HANSON-SMITH, V.; KOLACZKOWSKI, B.; THORNTON, J. W. Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty. **Molecular Biology and Evolution**, v. 27, n. 9, p. 1988–1999, set. 2010.

HOU, Y.-M.; SCHIMMEL, P. A simple structural feature is a major determinant of the identity of a transfer RNA. **Nature**, v. 333, n. 6169, p. 140–145, maio 1988.

KATOH, K.; ROZEWICKI, J.; YAMADA, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1160–1166, 19 jul. 2019.

KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. **Nucleic Acids Research**, v. 40, n. W1, p. W294–W297, 1 jul. 2012.

KUMAR, S. et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 1 jun. 2018.

LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of

protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, 1 abr. 1993.

MÖLLER, W.; JANSSEN, G. M. C. Statistical evidence for remnants of the primordial code in the acceptor stem of prokaryotic transfer RNA. **Journal of Molecular Evolution**, v. 34, n. 6, p. 471–477, jun. 1992.

NAGEL, G. M.; DOOLITTLE, R. F. Evolution and relatedness in two aminoacyl-tRNA synthetase families. **Proceedings of the National Academy of Sciences**, v. 88, n. 18, p. 8121–8125, 15 set. 1991.

NAGEL, GLENN M.; DOOLITTLE, RUSSELL F. Phylogenetic Analysis of the Aminoacyl-tRNA synthetases. **Journal of Molecular Evolution**, v. 40, n. 5, maio 1995.

NGUYEN, L.-T. et al. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. **Molecular Biology and Evolution**, v. 32, n. 1, p. 268–274, jan. 2015.

NIXON, A. E.; WARREN, M. S.; BENKOVIC, S. J. Assembly of an active enzyme by the linkage of two protein modules. **Proceedings of the National Academy of Sciences**, v. 94, n. 4, p. 1069–1073, 18 fev. 1997.

PARK, S. J.; SCHIMMEL, P. Evidence for interaction of an aminoacyl transfer RNA synthetase with a region important for the identity of its cognate transfer RNA. **The Journal of Biological Chemistry**, v. 263, n. 32, p. 16527–16530, 15 nov. 1988.

RIBAS DE POUPLANA, L.; SCHIMMEL, P. Operational RNA Code for Amino Acids in Relation to Genetic Code in Evolution. **Journal of Biological Chemistry**, v. 276, n. 10, p. 6881–6884, 9 mar. 2001.

RITCHIE, D. W.; VENKATRAMAN, V. Ultra-fast FFT protein docking on graphics processors. **Bioinformatics**, v. 26, n. 19, p. 2398–2405, 1 out. 2010.

ROOT-BERNSTEIN, R. et al. tRNA evolution from the proto-tRNA minihelix world. **Transcription**, v. 7, n. 5, p. 153–163, 19 out. 2016.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p. 725–738, abr. 2010.

SAHA, R. et al. A chimaeric glutamyl:glutaminyl-tRNA synthetase: implications for evolution.

Biochemical Journal, v. 417, n. 2, p. 449–455, 15 jan. 2009.

SCHWOB, E.; SÖLL, D. Selection of a “minimal” glutaminyl-tRNA synthetase and the evolution of class I synthetases. **The EMBO journal**, v. 12, n. 13, p. 5201–5208, 15 dez. 1993.

SUN, F.-J.; CAETANO-ANOLLÉS, G. The Origin and Evolution of tRNA Inferred from Phylogenetic Analysis of Structure. **Journal of Molecular Evolution**, v. 66, n. 1, p. 21–35, jan. 2008.

SZATHMÁRY, E. The origin of the genetic code: amino acids as cofactors in an RNA world. **Trends in Genetics**, v. 15, n. 6, p. 223–229, jun. 1999.

TREANGEN, T. J.; ROCHA, E. P. C. Horizontal Transfer, Not Duplication, Drives the Expansion of Protein Families in Prokaryotes. **PLoS Genetics**, v. 7, n. 1, p. e1001284, 27 jan. 2011.

VIALLE, R. A.; TAMURI, A. U.; GOLDMAN, N. Alignment Modulates Ancestral Sequence Reconstruction Accuracy. **Molecular Biology and Evolution**, v. 35, n. 7, p. 1783–1797, 1 jul. 2018.

WOESE, C. R. et al. Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process. **Microbiology and Molecular Biology Reviews**, v. 64, n. 1, p. 202–236, 1 mar. 2000.

WOLF, Y. I. et al. Evolution of aminoacyl-tRNA synthetases--analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. **Genome Research**, v. 9, n. 8, p. 689–710, ago. 1999.

ZHANG, Y. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic Acids Research**, v. 33, n. 7, p. 2302–2309, 11 abr. 2005.

Figuras e Legendas

Figuras

Figura 1.

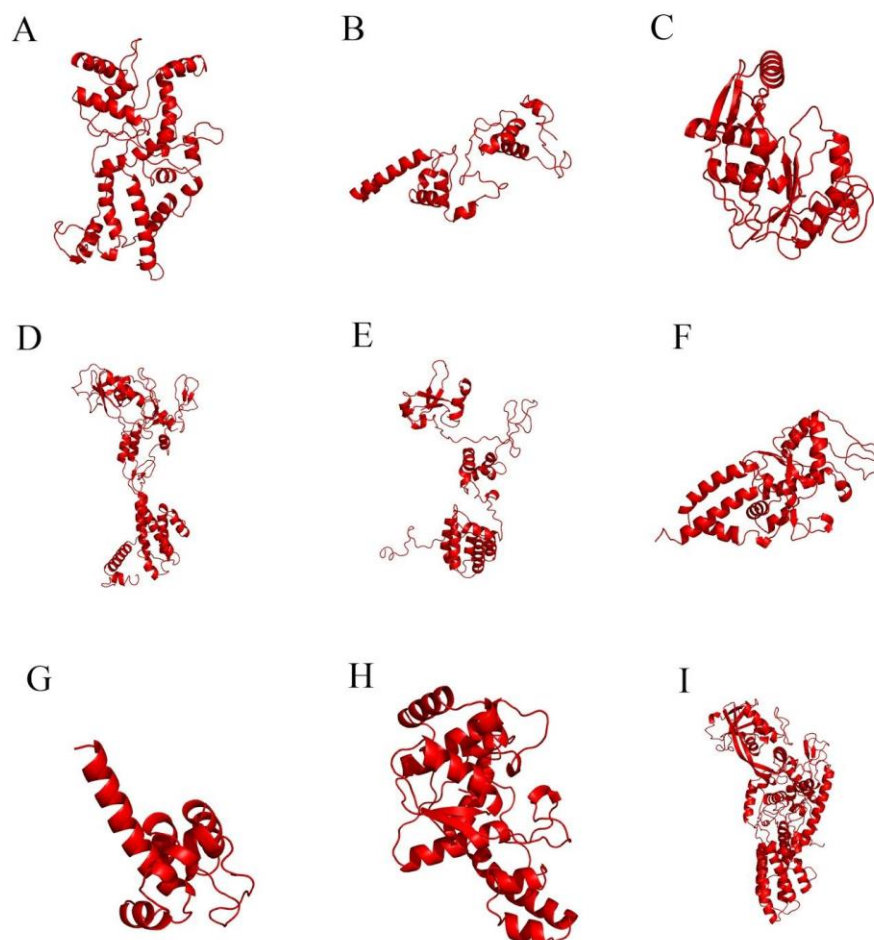


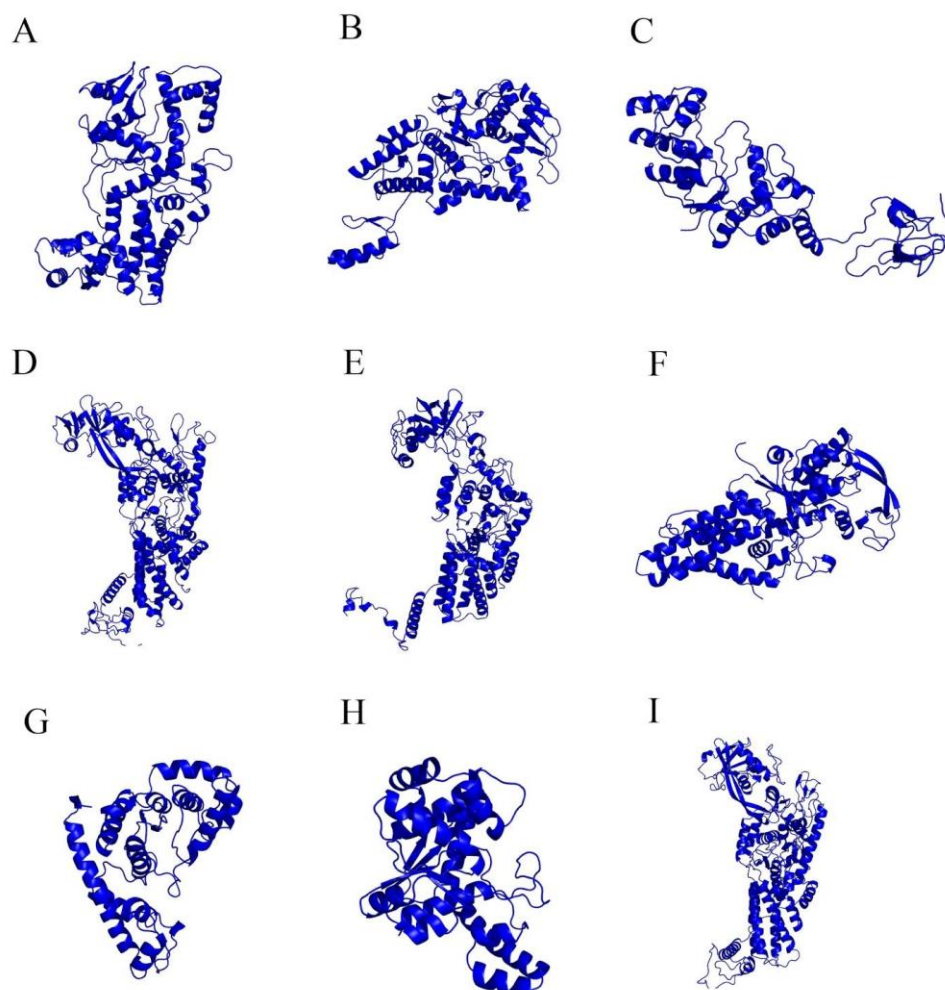
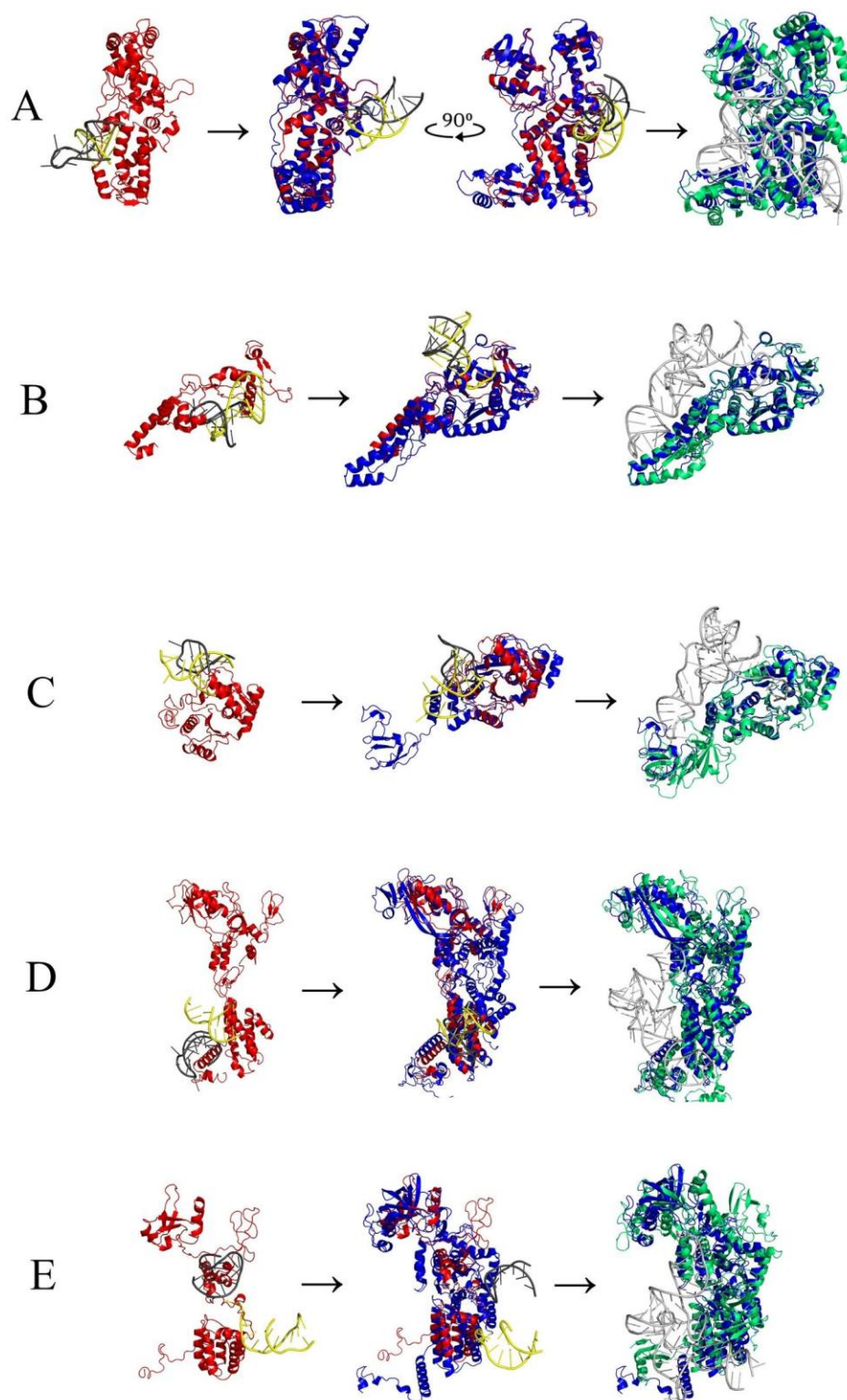
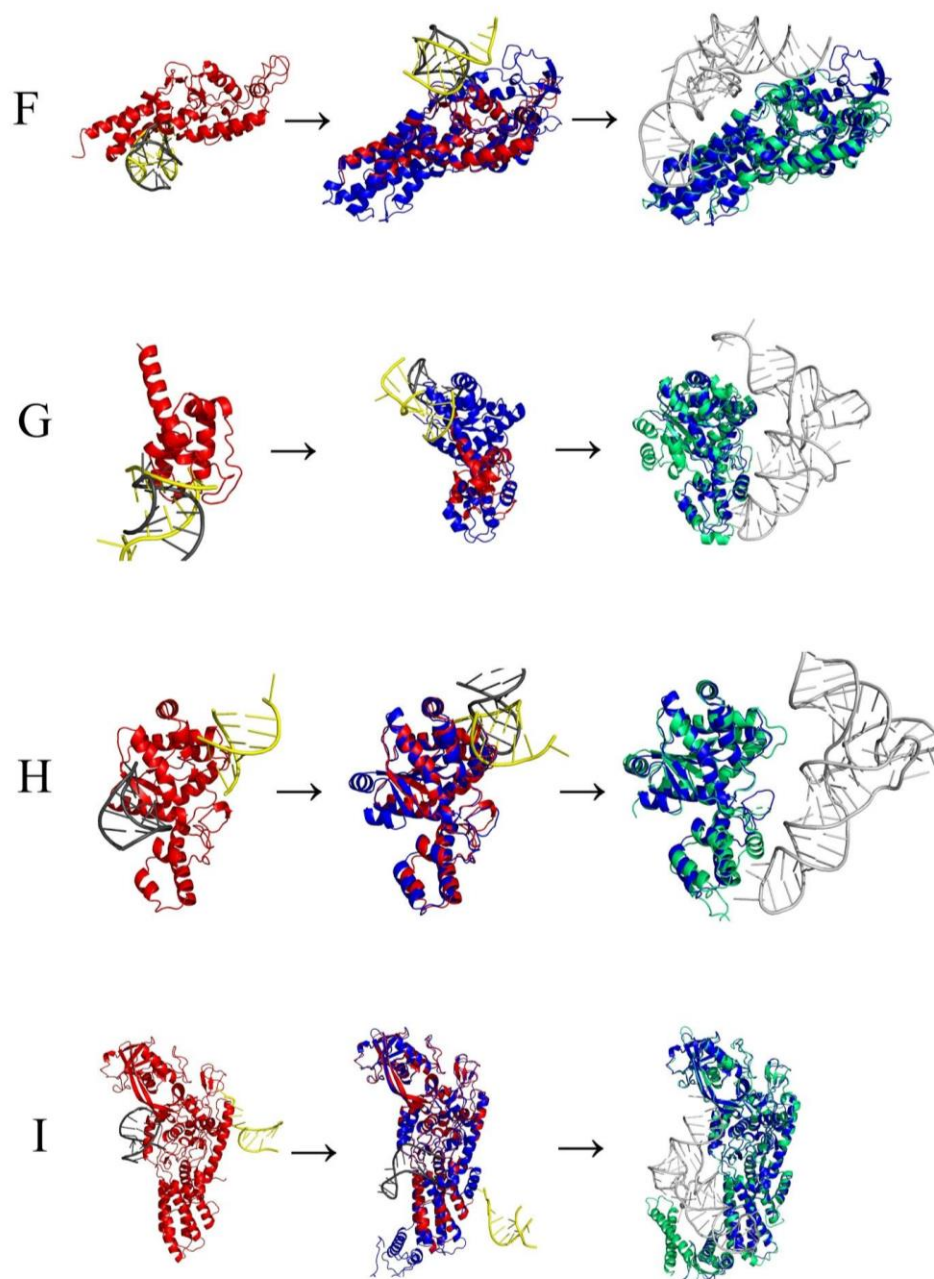
Figura 2.

Figura 3.





Legendas

Figura 1. Estrutura das sequências ancestrais da deleção total (CD). A) Arginina-tRNA sintetase, B) Cisteína-tRNA sintetase, C) Glutamato-tRNA sintetase, D) Isoleucina-tRNA sintetase, E) Leucina-tRNA sintetase, F) Metionina-tRNA sintetase, G) Triptofano-tRNA sintetase, H) Tirosina-tRNA sintetase, I) Valina-tRNA sintetase.

Figura 2. Estrutura das sequências ancestrais da deleção parcial (PD). A) Arginina-tRNA sintetase, B) Cisteína-tRNA sintetase, C) Glutamato-tRNA sintetase, D) Isoleucina-tRNA sintetase, E) Leucina-tRNA sintetase, F) Metionina-tRNA sintetase, G) Triptofano-tRNA sintetase, H) Tirosina-tRNA sintetase, I) Valina-tRNA sintetase.

Figura 3. Passos evolutivos propostos das estruturas sobrepostas da deleção completa-CD (vermelho), deleção parcial-PD (azul) e das estruturas modernas (verde-limão), associadas às estruturas do tRNA (loop do anticódon-preto e braço acceptor-amarelo) e o tRNA moderno (branco). Alinhamento estrutural foi feito através do software TM-align. A produção das imagens foi feita através do Pymol. A) Arginina-tRNA sintetase, B) Cisteína-tRNA sintetase, C) Glutamato-tRNA sintetase, D) Isoleucina-tRNA sintetase, E) Leucina-tRNA sintetase, F) Metionina-tRNA sintetase, G) Triptofano-tRNA sintetase, H) Tirosina-tRNA sintetase, I) Valina-tRNA sintetase.

Tabelas e legendas

Tabela 1. Quantidade de sequências para cada aaRS. O número de sequências para cada aaRS compreende proporcionalmente os três domínios da vida (Archaea, Bacteria e Eukarya).

Proteína	Número de sequências
Arginina-tRNA sintetase	600
Cisteína-tRNA sintetase	600
Glutamato-tRNA sintetase	216
Isoleucina-tRNA sintetase	600
Leucina-tRNA sintetase	600
Metionina-tRNA sintetase	417
Tirosina-tRNA sintetase	600
Triptofano-tRNA sintetase	600
Valina-tRNA sintetase	300
Total de Sequências	4533

Tabela 2. Relação da energia livre após o docking entre as aaRS e as estruturas do tRNA (e-total - kj/mol).

Aminoacil-tRNA sintetase	Energias – Loop do Anticódon		Energias – Braço acceptor	
	CD	PD	CD	PD
ArgRS	-2288.7	-497.5	-2391.8	-610.5
CisRS	-4202.1	-880.6	-3129.5	-884.5
GluRS	-3173.0	-2210.8	-3856.4	-2843.9
IleRS	-1671.4	-951.2	-1756.8	-705.6
LeuRS	-2568.6	-2132.2	-1462.6	-1613.4
MetRS	-1217.7	1579.9	-1604.9	2725.3

TirRS	-2472.7	-2023.4	-2260.8	-1077.3
TrpRS	-2368.9	-844.0	-2372.6	-1414.2
ValRS	205.3	903.2	497.7	-219.6

Tabela 3. A tabela abaixo exibe cada aaRS, da deleção completa (CD) e da deleção parcial (PD) e seu respectivo template (Query) utilizado para a modelagem molecular. Também exibe o RMSD (Root-mean-square deviation) e o percentual de identidade entre cada template e as estruturas ancestrais. O template é escolhido durante o processo de modelagem molecular, onde o I-Tasser elenca as 10 estruturas mais similares à nossa proteína a partir do alinhamento estrutural com proteínas resolvidas experimentalmente no PDB (Protein Data Bank).

Proteína		CD	PD
ArgRS	Query	4q2tA	6ao8A
	RMSD	2.38	1.11
	Identidade	0.677	0.272
CysRS	Query	1u0bB	1u0bB
	RMSD	2.14	0.94
	Identidade	0.539	0.564
GluRS	Query	4griA	5zdkA
	RMSD	2.61	1.54
	Identidade	0.312	0.351
IleRS	Query	1ivsB	1jzsA
	RMSD	2.05	1.14
	Identidade	0.260	0.441
LeuRS	Query	5ah5A	6lpfA
	RMSD	2.58	2.40
	Identidade	0.414	0.746
MetRS	Query	1rqgA	1rqgA

	RMSD	3.09	1.55
	Identidade	0.330	0.379
TyrRS	Query	2cybA	2cybA
	RMSD	1.33	0.85
	Identidade	0.382	0.563
TrpRS	Query	6gs1A	5ekdA
	RMSD	3.95	1.57
	Identidade	0.052	0.252
ValRS	Query	1ivsA	1ivsA
	RMSD	0.81	0.80
	Identidade	0.318	0.315

Tabela 4. Cada aaRS e sua respectiva estrutura moderna utilizada no alinhamento estrutural para construção do passo-a-passo evolutivo. Foram escolhidas estruturas resolvidas experimentalmente do banco de dados PDB, com a melhor resolução e associadas à seu tRNA cognato.

Proteína	Estrutura Moderna
ArgRS	2zue
CysRS	1u0b
GluRS	1qtq
IleRS	1ffv
LeuRS	4as1
MetRS	2csx
TyrRS	1j1u
TrpRS	2dr2
ValRS	1ivs

4. Considerações Finais

Nossos resultados levantam esclarecimentos a respeito dos aspectos envolvidos na relação das aaRS com o tRNA, mas também levantam muitas outras dúvidas e questionamentos. Devido a isso, análises subsequentes devem ser propostas, a fim de que haja um maior aprofundamento na evolução dessas proteínas, análises essas que também devam contemplar a classe II, outra classe de aaRS muito mais heterogênea e antiga que a classe I.

Em nossa interpretação, as aaRS foram uma das primeiras proteínas a surgir, fruto de uma síntese de peptídeos aleatória, onde não havia a presença de um código, e em um mundo onde as reações eram catalisadas por RNA. As aaRS então firmaram uma estável relação com algumas estruturas do tRNA, estruturas essas capazes de se ligar aos aminoácidos, mas ainda assim em um contexto onde não havia uma síntese coordenada de peptídeos. Essas estruturas dos tRNAs então passaram a atuar como estabilizadoras conformacionais das aaRS, a medida que as aaRS iam adquirindo complexidade a partir de eventos de transferência e fusão de domínios proteicos. As estruturas iniciais dos tRNAs, condicionadas à estrutura das aaRS, também evoluíram estruturalmente a partir de eventos de duplicação e deleção, servindo também como uma pressão seletiva para a diversificação das aaRS. Esses eventos teriam ocorrido paralelamente, em um processo de coevolução, levando ambas as estruturas a atuarem no fluxo da informação, no processo que hoje conhecemos como tradução. Tal processo é imprescindível para a manutenção e fidelidade do código genético.

Essas análises refletem apenas uma parte de um todo. Os passos subsequentes dizem respeito a integração desse modelo com outros modelos propostos para a evolução do ribossomo, outra molécula envolvida na tradução e também imprescindível para a manutenção do código genético.