

ARIOSVALDO PEREIRA DOS SANTOS JUNIOR

**ANÁLISE ESTRUTURAL E EVOLUTIVA DAS RNA POLIMERASE RNA
DEPENDENTE COM IMPLICAÇÕES NA TRANSIÇÃO DO MUNDO DE RNA
PARA UM MUNDO RNA/PROTEÍNA**

UNIVERSIDADE FEDERAL DA PARAIBA
DEPARTAMENTO DE BIOLOGIA MOLECULAR
CURSO DE BACHARELADO EM CIÊNCIAS BIOLÓGICAS

JOÃO PESSOA

2017

ARIOSVALDO PEREIRA DOS SANTOS JUNIOR

**ANÁLISE ESTRUTURAL E EVOLUTIVA DAS RNA POLIMERASE RNA
DEPENDENTE COM IMPLICAÇÕES NA TRANSIÇÃO DO MUNDO DE RNA
PARA UM MUNDO RNA/PROTEÍNA**

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

Orientador: Sávio Torres de Farias.

JOÃO PESSOA

2017

S231a Santos Junior, Ariosvaldo Pereira *Dos*.

Análise estrutural e evolutiva das RNA polimerase RNA dependente com implicações na transição do mundo de RNA para um mundo RNA/proteína / Ariosvaldo Pereira Santos Junior. - João Pessoa - 2017.
40f. : il.

Orientador: Sávio Torres de Farias.

Monografia (Graduação) – UFPB/CCEN.

1. Biologia molecular. 2. RNA dependente RNA polimerase. 3. tRNA. 4. Mundo de RNA. I.Título.

UFPB/BC

CDU: 577.2(043.2)

ARIOSVALDO PEREIRA DOS SANTOS JUNIOR

**ANÁLISE ESTRUTURAL E EVOLUTIVA DAS RNA POLIMERASE RNA
DEPENDENTE COM IMPLICAÇÕES NA TRANSIÇÃO DO MUNDO DE RNA
PARA UM MUNDO RNA/PROTEÍNA**

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

Data: _____

Resultado: _____

BANCA EXAMINADORA:

Sávio Torres de Farias, Doutor, UFPB

Gustavo Henrique Calazans Vieira, Doutor, UFPB

Pedro Cordeiro Estrela de Andrade Pinto, Doutor, UFPB

*Dedico este trabalho a minha amada Danielle Cristina.
Estas tatuada em minha alma no infinito do meu conhecer,
contigo aprendi a amar de forma inexplicada,
Saudade dos ventos polimórficos então fazes parte do meu ser,
assim me transformo na essência amada.*

AGRADECIMENTOS

Graças dou á minha progenitora Maria de Lourdes, por me fornecer toda energia vital, todo conhecimento que me passou, sendo sempre meu maior exemplo. A meu pai, por se exemplo de construção pessoal, moral e ética. Agradeço a toda minha família que sempre me apoiaram, em especial a pessoas tão influentes na mina vida, como Dayanna karla, Danielle Cristina, Kamilya Luiza. E a Vanessa por ser motivo de inspiração profissional.

A todas as pessoas que conheci durante minha vida acadêmica, as verdadeiras amizades que conquistei, aos momentos especiais que pode viver ao lado delas. Em especial a galera Bio sempre me inspirando a seguir em frente.

A todas as professoras e professores, orientadores, por fazerem parte do meu aprendizado não só profissional, mas como pessoa, que levarei por toda minha vida.

Ao professor Savio dou graças por permitir a conclusão deste trabalho, por todas as coisas que pude aprender e por me ser guia a um futuro.

Á Universidade Federal da Paraíba por me abrir portas, e me dá oportunidade de tornar aquilo que nunca sonhei.

Á natureza universal que sempre me proporcionou momentos especiais de expansão da consciência, assim como permitir transcender minha biologia, pois só assim me mantenho forte e seguro.

RESUMO

Entender a questões de como a vida se formou em nosso planeta, e como um sistema bioquímico simples emergiu não é fácil. A hipótese de que o RNA foi a primeira molécula da vida, se torna uma das mais plausíveis para a origem do maquinário bioquímico devido á dualidade que encontramos nessas moléculas, que tanto podem carregar informação biológica como atuar na catálise. A RNA dependente RNA polimerase (RDRP) é uma enzima de vírus de RNA. Em eucariotos funciona como amplificador de diversas moléculas de RNA com função de regulação da expressão gênica. Todas as RDRP possuem oito motivos de aminoácidos conservados, quatro desses motivos estão presentes em todas as classes de polimerases e residem na porção catalítica da enzima no domínio “palma”. Nós buscamos compreender a transição de um mundo de RNA para um mundo RNA/proteína, a partir de uma análise estrutural das RNA dependente RNA polimerase, evidenciando a hipótese “tRNA Core” mostrando a origem dos primeiros genes. Extraímos dos trabalhos de FARIAS et al, quatro sequências ancestrais de tRNA para a RDRP. Obtivemos 301 sequências para a enzima encontrada no NCBI, depois alinhamos e geramos uma filogenia, com o alinhamento e a árvore, no programa MEGA7, extraímos sequências de aminoácidos ancestrais, com as sequências primárias conseguimos gerar as estruturas tridimensionais dessas proteínas hipotéticas. Todas as estruturas ancestrais foram submetidas ao COACH e depois sobrepostas a todas as estruturas modernas da RDRP encontrada no PDB. Com os valores de RMSD das sobreposições construímos uma filogenia estrutural. Segundo os resultados do COACH todas as estruturas estavam se ligando a cofatores importantíssimos para a reação de catalise, assim como ácidos ribonucleicos. Nas sobreposições, as estruturas ancestrais se ligaram em todos os domínios proteicos das proteínas modernas, mas com prevalência ao sítio catalítico. Seguindo a filogenia, as ancestrais em sua grande maioria se encontraram na base da árvore. Os resultados sugerem que a enzima se originou a partir do domínio catalítico, sendo muito pequena e através da incorporação de pequenos aminoácidos da origem ao domínio catalítico mais parecido com o evidenciado atualmente. Seguindo a hipótese tRNA Core, é provável que um dos primeiros genes a se fixar ao código em formação fosse uma polimerase.

Palavras chave: RNA dependente RNA polimerase, tRNA, Mundo de RNA.

ABSTRACT

Understanding how life was formed on our planet, or even how a simple biochemical system emerged is not easy. The hypothesis that RNA was the first molecule of life, is one of the most plausible for the origin of the biochemical machinery. This is due to the duality we find in these molecules, which can both carry biological information and act in catalysis. One example of RNA having a catalytic action is RNA-dependent RNA polymerase (RDRP) is an enzyme found in retrovirus. In eukaryotes, it functions as an amplifier of several RNA molecules with the function of gene expression regulation. All RDRPs have eight conserved amino acid motifs, four of these motifs are known to be present in all classes of polymerases and reside in the catalytic portion of the enzyme in the palm domain. We sought to understand the transition from a RNA world to an RNA /protein world, from a structural analysis of RNA-dependent RNA polymerase, evidencing the "tRNA Core" hypothesis and showing the origin of the first genes. We extracted from the works of FARIAS et al, four ancestral sequences of tRNA for RDRP. We obtained 301 sequences for the enzyme found in NCBI, then we aligned these sequences and calculated a phylogeny, using the alignment and the tree, with the program MEGA 7 we extracted sequences of ancestral amino acids, with the primary sequences we were able to generate the three-dimensional structures of these hypothetical proteins. All ancestral structures were submitted to COACH and then superposed to all the modern RDRP structures found in the PDB. With the RMSD values of the overlaps we constructed a structural phylogeny. According to the results of COACH all the structures were linking to very important cofactors for the catalysis reaction, as well as ribonucleic acids. In overlaps, the ancestral structures have been bound in all protein domains in modern proteins, but with prevalence at the catalytic site. Following the phylogeny, the ancestors mostly found themselves at the base of the tree. The results suggest that the enzyme originated from the catalytic domain, being very small and through the incorporation of small amino acids from the origin to the catalytic domain more similar to the currently evidenced. Following the tRNA Core hypothesis, it is likely that one of the first genes to attach to the code in formation would be a polymerase.

Keywords: RNA-dependent RNA polymerase, tRNA Core, RNA world.

SUMÁRIO

INTRODUÇÃO	9
1 FUNDAMENTAÇÃO TEÓRICA	
2 OBJETIVOS.....	15
2 Objetivo geral	
2.1 Objetivos específicos	
3 METODOLOGIA.....	16
3.1 Reconstrução da estrutura ancestral para tRNA e análise de ligação	
3.2 Reconstrução da estrutura ancestral para RDRP e análise de ligação	
3.3 Análise Comparativa e Filogenia Estrutural	
4 RESULTADOS	19
4.1 Melhores ligantes COACH	
4.2 Alinhamento Estrutural	
4.3 Filogenia Estrutural	
5 DISCUSSÃO	30
6 CONCLUSÃO	34
REFERÊNCIAS	35

INTRODUÇÃO

1 FUNDAMENTAÇÃO TEÓRICA

Uma das questões mais intrigantes para a ciência é o surgimento da Vida, em que ponto um sistema simples capaz de se auto replicar, gerar variantes e ser cabível de evolução Darwiniana, pode emergir de um aglomerado de moléculas orgânicas, que passaram por um processo de evolução química[1]. Como esse maquinário bioquímico que conhecemos atualmente, emergiu de uma química prébiótica básica? Qual a primeira entidade auto replicante a se estabelecer? E o maior de todos os questionamentos, quem vem primeiro Ácidos nucleicos ou proteínas ?. Essas perguntas são bastantes difíceis de responder, pois ainda precisamos melhorar a gama de experimentos in vitro sobre química prébiótica, criar novas metodologias para poder analisar e interpretar dados moleculares, assim como limitações que sempre irão estar presentes em trabalhos sobre origem da vida, impedindo uma abordagem mais apurada dos fatos[2].

Como estamos tratando de algo muito distante da nossa realidade, fica claro que é um tempo bastante longo do passado. Logo são variáveis que nem conhecemos com tanta clareza, por exemplo quais as condições de temperatura, PH, salinidade dos mares, pressão atmosférica, entre várias outras do nosso planeta a 4 bilhões de anos atrás, que foram essenciais para a emergência da vida[3]. Temos dados atuais que há cerca de 4 milhões de anos atrás já tínhamos células bacterianas capazes de obter sua própria energia, e isso nos mostra que a vida celular é muito antiga em nosso planeta[4]. É provável, que a formação do código genético foi essencial para essa vida celular se desenvolver e evoluir, pois quando o genoma se formou estabeleceu o que chamamos de “Dogma Central” na Biologia da vida[5].

Desde os antigos gregos, passando pelos experimentos de Miller até os dias atuais, os pesquisadores têm tentado mostrar como a vida começou até a formação da primeira célula[6]. Hoje vários estudos têm tentado demonstrar um genoma para o último ancestral celular universal (LUCA), mostrando que o ancestral de todas as células já tinha um sistema bioquímico bem estabelecido, e que evoluiu muito provavelmente de um pré-LUCA[7]. Mas outros estudos pelo contrário, tem proposto que antes da formação desse pré-LUCA, haveria entidades replicativas que estavam em constante simbiose umas com as outras, havendo troca constante de informação, ou seja, havia um genoma antes do LUCA[8], com genes essenciais para uma entidade auto-replicativa evoluir, e com essa troca de informação, foi possível algumas classes dessas entidades darem origem a população de LUCA[9].

Existem vários estudos que tentam mostrar a formação de moléculas orgânicas no laboratório, simulando o ambiente prébiótico da terra sobre condições controladas[10-13]. Desde 1953 sabemos que a formação de aminoácidos é plausível de ocorrer na sopa prébiótica. Vários estudos com outras moléculas foram realizados, por exemplo, estudos com formamidina e cianeto de hidrogênio, tem demonstrado que em algum ponto de aglomeração dessas moléculas temos a formação de purinas, pirimidinas, ácidos graxos, diversos cofatores como NAD, FAD, ATP[12,13,14]. Vários desses estudos tem focado em meteoritos ricos em água, minerais radioativos e compostos orgânicos, que sugerem que pode ser plausível a formação de todas essas moléculas essenciais para os organismos vivos, na terra há 4 bilhões de anos[15].

Em 1986 foi lançada a hipótese do mundo RNA, em que foi proposto que o RNA foi a primeira molécula da vida a se formar [16]. Devido a dualidade que encontramos hoje em moléculas de RNA, que tanto podem armazenar informação genética, como possuir atividade de catálise[17-19], por exemplo a parte do ribossomo que exerce a ligação peptídica é uma ribozima, parte do rRNA[20,21]. A existência desse pequeno momento na emergência da vida, em que moléculas de RNA possuíam funções tanto de fenótipo quanto de genótipo, torna o mundo de RNA uma hipótese promissora para a formação das primeiras entidades auto-replicativas [22].

O mundo RNA é uma das mais plausíveis hipóteses para explicar a origem da vida, devido à variedade de observações que encontramos nos organismos atuais, onde o RNA com suas diversas funções é parte chave na decodificação da informação. Temos várias moléculas de RNA que passam por processos de auto-splicing, a vasta gama de ribozimas demonstrada em laboratório, assim como aquelas encontradas naturalmente[23,24,25]. Outra é o controle da expressão gênica sendo guiada por moléculas de RNA de interferência. E a descoberta atual de que quase todo o genoma é transcrito, mas apenas uma pequena porcentagem é codificada em proteínas, enquanto a maior parte produz RNAs não codificantes que vão atuar como reguladores de expressão[26,27]. Todos esses pontos fortalecem ainda mais a ideia de que, esse mundo de RNA tenha existido e todas essas relíquias/fósseis moleculares nos deixam pistas acerca desse mundo. Por exemplo, a replicação do DNA, necessita de uma vasta gama de enzimas proteicas, assim como no reparo e recombinação, o mundo de RNA nos deixa claro que situações como essas, poderiam se formar muito mais simples quando não existia DNA ou Proteína, sendo o RNA a primeira molécula a exercer papéis tanto de genoma como catalise, que posterior essas funções foram melhor exercidas por DNA e proteína respectivamente[28].

Outros estudos também focando em saber como ocorreu a formação de moléculas de RNA e posteriormente o surgimento do DNA, tem mostrado que era plausível a existência de quimeras de RNA/DNA, e que essas quimeras poderiam ser muito mais estáveis do que moléculas de RNA por si só, sendo percussores da formação de cadeias apenas de DNA, pela melhor estabilidade exercida pela molécula[29].

As RNA-dependente RNA-polimerase (RdRp) são enzimas essenciais para todos os vírus com genoma de RNA[30]. Diferente de outras polimerases, as RdRp possuem alta taxa de erro na hora da incorporação do nucleotídeo adequado (em ordem de 10^{-4}). Por isso, são cruciais para gerar a variabilidade viral e permiti sua evolução. Segundo os trabalhos de Nasir et al [31], onde afirmam que os vírus de RNA são a classe de viroses mais antiga, essa enzima tem papéis fundamentais para sua evolução, pois é provável que essas entidades sejam fósseis moleculares de um passado remoto [31,32]. Estas enzimas também participam de vários processos biológicos em eucariontes, como amplificação de microRNAs, que exercem papéis no controle da expressão genica, atuando no desenvolvimento de nematóides, e protegendo plantas contra ácidos nucleicos patógenos [33,34].

As RdRp apresentam em sua conformação estrutural 3D, arquitetura de uma mão direita com três subdomínios funcionais, denominados, dedos, palma e polegar. Em todas as polimerases o subdomínio palma é bastante conservado, sendo os outros subdomínios diferentes em cada classe. Nas RdRp o subdomínio Palma possui dois resíduos conservados de Aspartato, o qual interagem com dois íons metálicos bivalentes para a ocorrência do ataque nucleofílico, onde essa interação permite a incorporação do ribonucleotídeo de entrada da cadeia em formação. O subdomínio dos dedos estar envolvido na interação com a molécula molde e incorporação de nucleotídeos. O subdomínio polegar estar relacionado a formação do grampo de ligação no momento em que a reação ocorre. Estar localizado oposto aos dedos, na porção N-terminal da enzima, sendo o domínio mais variável [35-37].

Seis motivos estruturais (A – F) são bastante conservados em todas as RdRp. Todos esses motivos estão localizados no subdomínio palma, com exceção do motivo F, que se encontra nos dedos. Todos esses motivos têm mostrado estarem relacionados a processos importantíssimos na correta incorporação e reorganização de nucleotídeos. Vários desses motivos são compartilhados com outras polimerases, indicando sua importância fundamental na função enzimática. Nas RdRp estes motivos encontram-se na cavidade central que está responsável pela ligação de substratos e cofatores [38,39].

As RdRp têm um papel central no início da evolução da vida, quando o RNA primariamente supostamente se comportava como material genético. Acredita-se que essa enzima teria importantes funções. Periodicamente as primeiras formas de vida poderiam ser baseadas em RNA, algumas moléculas desses RNAs poderiam possuir ação catalítica a partir de mudanças estruturais na cadeia, com ribozimas polimerases exercendo função de replicação desse genoma de RNA [40-41]. Posteriormente quando a função catalítica passou a ser exercida por proteínas, as enzimas que replicam o genoma teriam sido as primeiras a se formar devido a sua importância, sendo a polimerase de RNA parte dos primeiros genes. Ainda essas polimerases são capazes de inserir mutações, processo base para ocorrência de variabilidade genética, o que pode ter levado à origem de outras enzimas polimerases[42]. LEON et al[40] sugere que uma transcriptase reversa teve papel importantíssimo para o surgimento da molécula de DNA.

Farias et al [8,9] sugeriram a constituição do proteoma antes do último ancestral comum universal baseado na tradução das moléculas ancestrais de tRNA. Entre as proteínas modernas que tinham semelhanças com as sequências ancestrais de tRNA traduzidas, a RdRp era a única polimerase que tinha correspondência.

Estes resultados reforçam a natureza do material genético primitivo baseado em RNA e o surgimento precoce da RdRp no sistema biológico.

FARIAS et al [9,43], sugerem um novo modelo para a origem do sistema biológico, onde as moléculas de tRNA orquestraram a organização inicial da vida. Na emergência do sistema biológico primitivo, o tRNA poderia trabalhar como o material genético primitivo e dar origem ao mRNA e rRNA, bem como, gerar por conformações estruturais alternativas às primeiras ribozimas. Através do processo de dimerização, moléculas de proto-tRNA poderiam dar origem ao Centro Peptidil Transferase(PTC)[44]. Com o surgimento do rRNA e mRNA, este sistema poderia originar o sistema de tradução primitiva, que possibilitou a síntese das primeiras proteínas.

Esse sistema de tradução primitiva, permitiu a transição de um mundo de RNA para um mundo de RNA/proteína e posteriormente, para o mundo DNA/RNA/ proteína[43,44].

2 OBJETIVO GERAL

Compreender a origem e a evolução estrutural das RNA polimerase RNA dependentes

2.1 OBJETIVOS ESPECIFICOS

- Avaliar as estruturas ancestrais derivadas de tRNAs como prováveis moléculas originárias das RdRp.
- Reconstruir sequências ancestrais das RdRp a partir de sequências modernas, determinar a estrutura tridimensional e as etapas evolutivas até a formação da enzima funcional.
- Compreender as relações filogenéticas estruturais, buscando observar mecanismos transicionais durante o processo evolutivo.

3 METODOLOGIA

3.1 Reconstrução da estrutura ancestral para tRNA e análise de ligação

Farias et al [9] sugeriram que a RNA polimerase RNA dependente tinha origem a partir da junção de tRNA ancestrais. Quatro sequências ancestrais para RNA polimerase RNA dependente foram utilizadas neste trabalho como sugerido por Farias et al. Cada sequência ancestral foi submetida ao servidor web I-Tasser [45] (<http://zhanglab.ccmb.med.umich.edu/I-TASSER>) para prever por homologia o melhor modelo estrutural. Para uma melhor resolução estrutural das proteínas ancestrais, cada modelo ancestral estrutural foi submetido ao programa ModRefiner[46] (<http://zhanglab.ccmb.med.umich.edu/ModRefiner>), que refinou todas as hastes dando uma melhor resolução para cada modelo ancestral (Figura 1). Após o refinamento, cada modelo estrutural foi submetido ao programa COACH[47] (<http://zhanglab.ccmb.med.umich.edu/COACH/>) para inferência sobre qual molécula se liga melhor ao local de ligação das Proteínas ancestrais.

3.2 Reconstrução da estrutura ancestral para RDRP e análise de ligação

Obtivemos 301 sequências da polimerase encontrada no NCBI (<http://ncbi.nlm.nih.gov/>), para todos os vírus de RNA assim como organismos celulares, todas com mais de 500 aminoácidos. Cada sequência foi submetida ao Uniprot [48](<http://www.uniprot.org/blast/>) para verificar a confiabilidade das mesmas, sendo excluídas sequências redundantes. Posteriormente alinhamos as sequências utilizando o MAFFT[49](<http://mafft.cbrc.jp/alignment/server/>), para posterior inferência sobre qual o melhor modelo evolutivo para o alinhamento usando o ProtTest 3.2[50], que analisou e o melhor modelo foi o WAG+G.

Com o alinhamento foi gerada uma árvore filogenética utilizando T-Rex[51] (<http://www.trex.uqam.ca/index.php?action=inference&project=trex>),

construída pelo método de Máxima Verossimilhança. Com o alinhamento e a árvore filogenética foi possível no programa MEGA7[52], construir uma filogenia ancestral, obtendo também as sequências ancestrais. Não foi possível o cálculo com 100% de deleção de gaps, assim extraímos duas sequências ancestrais a 90%.

Cada sequência ancestral foi submetida ao I-Tasser[45] para predição da melhor estrutura. Para melhor resolução estrutural dessas proteínas hipotéticas, cada uma delas foi submetida ao ModRefiner [46], que refinou todas as estruturas dando uma melhor resolução para cada modelo (Figura 2). Após refinado cada modelo estrutural foi submetido ao COACH [47] para inferência sobre qual molécula melhor se liga ao seu sítio.

COACH é uma abordagem de meta-servidor para predição do local de ligação e os melhores ligantes da proteína alvo. O COACH irá gerar previsões do sítio de ligação do ligante complementar utilizando vários métodos comparativos, que reconhecem modelos de ligação de ligantes a partir da base de dados da função de proteína BioLiP por comparações de subestrutura e correspondência de sequência. Estas previsões serão combinadas com resultados de outros métodos para gerar previsões finais do sítio de ligação do ligante[47].

3.3 Análise Comparativa e Filogenia Estrutural

Foram extraídas todas as sequências das 24 RdRp cristalizadas encontrada no PDB[53]. Depois, foi realizado todo o processo descrito acima com essas sequências. No programa MEGA7 poder-se construir a árvore ancestral assim como sequências ancestrais com 100% de deleção de gaps. Extraímos uma sequência ancestral, que depois de submetida ao I-Tasser, ModRefiner (Figura 3), foi lançada ao COACH[45-52].

Foram obtidas as estruturas cristalizadas modernas para RdRp, de vírus de RNA, que incluem : RNA Polimerase DSRNA Bacteriophage(PDB: 1HHS); RNA Polimerase Rabbit Hemorrhagic Disease Vírus(PDB: 1KHV); RNA Polimerase Sapporo Virus(PDB: 2CKW); RNA polimerase Hepatitis C (PDB: 2D41); RNA polimerase Neurospora Crassa (PDB: 2J7N); RNA Polimerase Birnavirus (PDB: 2PGG); RNA Polimerase Infectious bursal disease vírus (PDB: 2PUS); RNA polimerase Rotavirus (PDB: 2R7T); RNA polimerase Infectious Pancreatic Necrosis Vírus(PDB: 2YI8); RNA polimerase cypoviruses (PDB: 3JA4); RNA polimerase Enterovirus A (PDB: 3N6L); RNA polimerase Norwalk vírus (PDB: 3UQS); RNA polimerase Rotavirus A (PDB: 4AU6); RNA polimerase Thosea asigna vírus (PDB:4XHA); Rhinovirus A (PDB: 1XR7); Enterovirus C(PDB: 3OL6); Foot-and-mouth disease virus (PDB: 1U09); Cardiovirus A (PDB: 4NZ0); Japanese encephalitis virus (PDB: 4HDH); Dengue vírus (PDB: 2J7U); Bovine viral diarrhea vírus 1 (PDB: 1S48); vírus Qbeta (PDB: 3MMP); Reovirus (PDB: 1MUK); La Crosse Bunyavirus (PDB: 5AMQ);

Cada estrutura extraída do PDB[53], foi submetida ao ModRefiner[46]. As estruturas modernas foram sobrepostas às duas proteínas ancestrais hipotéticas e a todas as 4 estruturas de tRNA. Usando o TM-align[54](<http://zhanglab.ccmb.med.umich.edu/TM-align/>), foi possível verificar em qual região elas estão se ligando, junto com seus valores de RMSD.

Com os valores de RMSD das sobreposições criamos uma matriz de sobreposição estrutural, que através do método Neighbor-Joining no T-Rex[51], deu origem à uma árvore filogenética estrutural para todas as proteínas. Em todos os programas foram utilizados parâmetros padrão.

4 RESULTADOS

4.1 Melhores ligantes COACH

Os resultados do COACH para nossa estrutura controle é apresentada na Tabela 1, mostrando como esperado que os melhores ligantes são moléculas essenciais para atividade de polimerização[12,13,32], como observado para Ancescont.

Tabela 1: Resultados do COACH, com melhores ligantes para nossa estrutura controle.

Grupo Controle	COACH	TM-SITE	S-SITE
Ancescont	Ácidos Ribonucleicos	Magnésio, Cálcio, Zinco	Magnésio

Na figura 1, são mostradas as sequências e os modelos estruturais gerados por homologia com sequências ancestrais, como sugerido por Farias et al.[9]. As estruturas obtidas mostraram uma conformação simples com loops e hélices e as sequências propostas em tRNA1, tRNA 2 e tRNA 3 mostraram semelhanças entre elas, com pequenas diferenças na constituição de aminoácidos. A sequência tRNA 4 é mais divergente e pode representar um segundo módulo na formação da proteína primordial.

Na Tabela 2, é mostrada a análise ao substrato e à ligação do cofator às sequências ancestrais derivadas do tRNA ancestral. Os dados indicam que as estruturas ancestrais têm a capacidade de se ligar em moléculas simples como magnésio, cálcio, ribonucleótidos entre outros. O cofator mais frequentemente encontrado nesta análise foi o magnésio que é o cofator das RdRp moderna. Estas análises indicaram que na origem da RdRp a parte mais antiga da proteína é o local de ligação do cofator. Nas proteínas modernas, o local de ligação do cofator está localizado no domínio palma, o que é considerado pela análise evolutiva o domínio mais antigo na RdRp moderna [38].

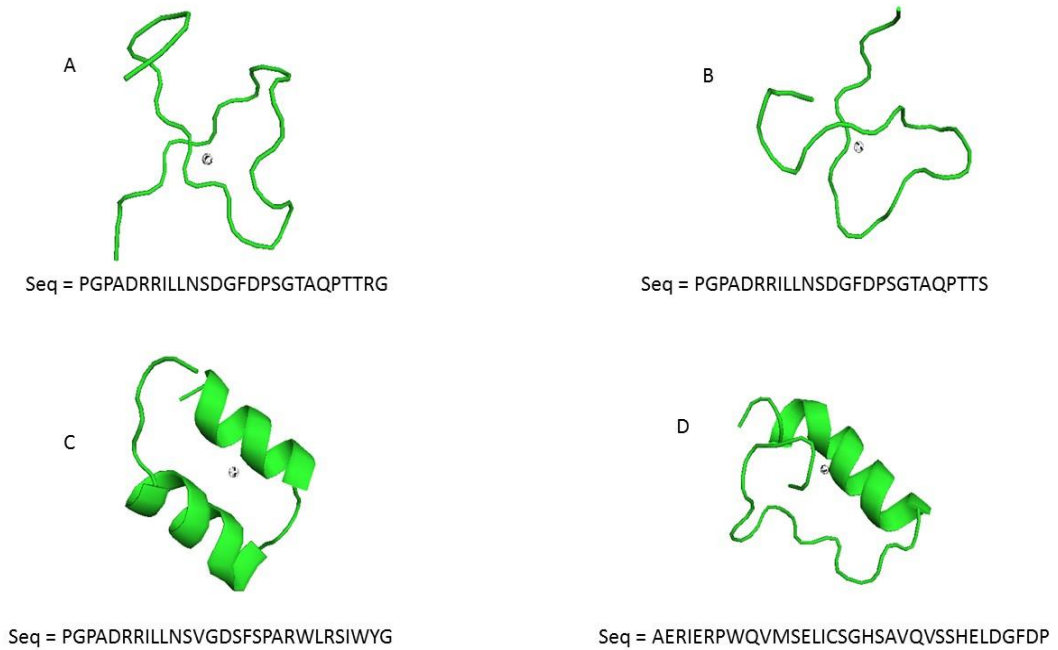


Figura 1: Modelo estrutural com suas respectivas sequências geradas usando os concatêmeros de proto-tRNA traduzidos como sugerido por Farias et al, com o metal representado no centro de cada molécula. A- tRNA 1. B – tRNA 2. C – tRNA 3. D- tRNA 4.

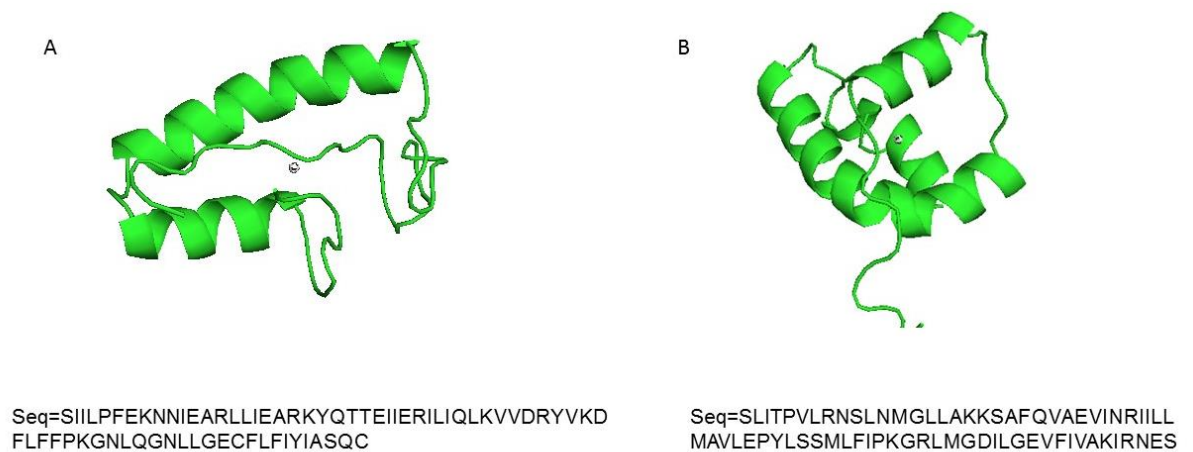
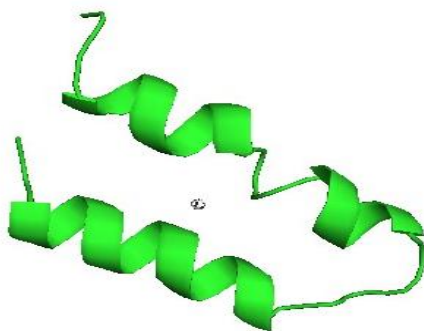


Figura 2: Estruturas ancestrais para RDRP refinadas. A – Equivalente a Ancestral 90, com metal no centro da proteína. B – Equivalente a Ancestral 90.1 com seu respectivo metal no centro da molécula.



Seq = GMSATSIMGDLQALIEVTFLKRTFRLDPVLPMMKENMQEHHG

Figura 3: Estrutura controle, com o metal representado no centro da molécula, e sua respectiva sequência de aminoácidos.

Estes dados estão de acordo com a sugestão de Farias et al, onde os tRNAs por junção, deram origem aos primeiros genes e que RdRp é uma enzima muito antiga no sistema biológico. Ancestralmente a RdRp replicaria o RNA, essa molécula genética primitiva pode então armazenar as informações biológicas iniciais e com o surgimento desta enzima no início da organização do sistema biológico temos a criação e perpetuação do primeiro módulo informativo que funcionou como o impulso inicial para o Metabolismo necessário à emergência da vida.

Tabela 2: Resultado da melhor ligação do substrato para os modelos gerados pela tradução dos proto-tRNA traduzidos.

ANCESTRAIS tRNA	COACH	TM-SITE	S-SITE
tRNA 1	NAD, Potássio	Magnésio, Potássio, Cálcio	Magnésio, AMP, Cálcio
tRNA 2	Cálcio	Zinco, cálcio	Magnésio, AMP, Cálcio
tRNA 3	Magnésio	Magnésio, Ácidos Ribonucleicos, Cálcio	Clorofila, D- manose
tRNA 4	Iodo	Magnésio, Zinco, Cálcio	FAD, Zinco, ATP

Para as estruturas ancestrais da RdRp, os melhores ligantes foram ácidos ribonucleicos assim como cofatores importantes para a ação catalítica como o magnésio (Tabela 3). Mostrando que estruturas ancestrais em sua conformação simples tem preferência nos sítios de ligação por uma diversidade de moléculas, que de alguma forma interagem para a melhora da atividade catalítica da enzima[30,35].

Tabela 3: Melhores ligantes para as ancestrais geradas a partir das sequências modernas para a RNA polimerase RNA dependente.

Ancestrais	COACH	TM-SITE	S-SITE
RDRP			
Ances 90	Peptídeos, clorofila	Bacterioclorofila	Magnésio
Ances 90.1	Ácidos Ribonucleicos	Magnésio	Ácidos Ribonucleicos, fosfatidilcolina

4.2 Alinhamento Estrutural

A estrutura controle (Figura 3) em sua grande maioria se ligou no domínio catalítico das proteínas Virais, assim como em regiões próximas de incorporação de nucleotídeos. Como esperado, a estrutura ancestral tem preferência de ligação por regiões com função principalmente envolvidas na reação de polimerização. Das 24 sobreposições realizadas, nove delas a estrutura controle estava se ligando no domínio catalítico, oito na região do polegar, três na região dos dedos e quatro sobreposições a estrutura estavam ligando em sítios próximos de onde há a incorporação do metal. Esses resultados servem como base para todas as sobreposições geradas com as ancestrais hipotéticas.

Na Tabela 4, é mostrado o valor de RMSD para o alinhamento estrutural entre os tRNA e as proteínas modernas. É possível observar que os ancestrais de tRNA têm correspondência com todas as proteínas modernas e com domínios diferentes nestas proteínas. O alinhamento estrutural com o menor valor de RMSD foi entre o tRNA 2 no domínio dos dedos do Cardiovírus A com um valor de RMSD de 1,97 (4NZ0), o tRNA 3 no domínio Palma da RNA Polimerase DSRNA Bacteriófago com RMSD de 1,93 (1HHS) e no domínio polegar temos a RNA polimerase Rotavirus com RMSD de 1,44 (2R7T). Na Figura 4, é mostrado o alinhamento estrutural que tinha o menor valor de RMSD.

Estes resultados sugerem que a RdRp tem origem a partir das junções dos proto-tRNAs que foram traduzidos no surgimento do sistema de tradução primitiva, onde o RNA foi a molécula informativa. Também está sugerindo que o pequeno domínio que tinha origem a partir de proto-tRNAs funcionou como blocos de construção e que o surgimento dos dedos e polegar ocorreu pela duplicação do primeiro domínio que teve origem no domínio palma.

Tabela 4: Resultados do TM-aling. Cada estrutura moderna sobreposta as proteínas hipotéticas de tRNA, com seus respectivos valores de RMSD.

ESTRUTURA PDB	DEDOS	PALMA	POLEGAR
1HHS	tRNA 2 RMSD= 2.67	tRNA 1 RMSD= 3.27 tRNA 3 RMSD= 1.93 tRNA 4 RMSD= 2.21	
1KHV		tRNA 1 RMSD= 2.46 tRNA 2 RMSD= 3.40 tRNA 3 RMSD= 1.96	tRNA 4 RMSD= 2.99
1MUK	tRNA 1 RMSD= 3.39	tRNA 2 RMSD= 3.54	tRNA 3 RMSD= 2.21 tRNA 4 RMSD= 2.27
1S48	tRNA 1 RMSD= 2.80	tRNA 2 RMSD= 3.48 tRNA 4 RMSD= 3.01	tRNA 3 RMSD= 1.95
1U09	tRNA 1 RMSD= 3 tRNA 2 RMSD= 3.67	tRNA 3 RMSD= 2.15 tRNA 4 RMSD= 3.08	
1XR7		tRNA 1 RMSD= 3.40 tRNA 2 RMSD= 3.26 tRNA 4 RMSD= 2.72	tRNA 3 RMSD=1.97

ESTRUTURA PDB	DEDOS	PALMA	POLEGAR
2CKW	tRNA 2 RMSD= 3.73	tRNA 3 RMSD = 2.49	tRNA 1 RMSD= 2.87 tRNA 4 RMSD= 2.72
2D41	tRNA 1 RMSD= 3.49 tRNA 2 RMSD= 3.37 tRNA 3 RMSD= 3.15	tRNA 4 RMSD= 2.39	
2J7N	tRNA 2 RMSD= 2.82	tRNA 1 RMSD= 3.66	tRNA 3 RMSD=2.95 tRNA 4 RMSD= 3.29
2J7U	tRNA 1 RMSD= 3.30 tRNA 4 RMSD= 2.66	tRNA 3 RMSD= 2.46	tRNA 2 RMSD= 3
2PGG	tRNA 1 RMSD= 3.26	tRNA 4 RMSD= 2.51	tRNA 2 RMSD=3.30 tRNA 3 RMSD=2.59
2PUS		tRNA 1 RMSD= 2.76 tRNA 2 RMSD= 2.30	tRNA 3 RMSD= 2.84 tRNA 4 RMSD= 2.84
2R7T	tRNA 2 RMSD= 2.91 tRNA 4 RMSD= 2.49	tRNA 1 RMSD = 3.45	tRNA 3 RMSD= 1.44
2YI8	tRNA 1 RMSD= 3.33 tRNA 4 RMSD= 3.08	tRNA 3 RMSD= 3.21	tRNA 2 RMSD= 2.79
3JA4	tRNA 1 RMSD= 3.18 tRNA 3 RMSD= 2.12	tRNA 2 RMSD= 3.26	tRNA 4 RMSD= 2.81
3MMP	tRNA 1 RMSD= 2.70 tRNA 3 RMSD= 2.03	tRNA 2 RMSD= 3.36 tRNA 4 RMSD= 3.10	
3N6L	tRNA 2 RMSD= 4.09	tRNA 1 RMSD= 3.23 tRNA 3 RMSD= 2.43 tRNA 4 RMSD= 2.78	
3OL6	tRNA 2 RMSD= 4.1	tRNA 1 RMSD= 3.16 tRNA 4 RMSD= 2.74	tRNA 3 RMSD= 2.45
3UQS		tRNA 2 RMSD= 3.24 tRNA 3 RMSD= 2.01	tRNA 1 RMSD= 3.50 tRNA 4 RMSD= 3.14
4AU6	tRNA 2 RMSD= 2.35 tRNA 4 RMSD= 2.51	tRNA 1 RMSD= 3.20	tRNA 3 RMSD= 1.77
4HDH	tRNA 1 RMSD= 3.46 tRNA 3 RMSD= 2.28 tRNA 4 RMSD= 2.70	tRNA 2 RMSD= 3.44	
4NZ0	tRNA 2 RMSD= 1.97	tRNA 1 RMSD= 2.29 tRNA 3 RMSD= 2.35	tRNA 4 RMSD= 3.25
4XHA	tRNA 3 RMSD= 2.05	tRNA 1 RMSD= 3.12 tRNA 2 RMSD= 2.86 tRNA 4 RMSD= 3.52	
5AMQ	tRNA 4 RMSD= 3.11	tRNA 1 RMSD= 3.23 tRNA 2 RMSD= 3.43 tRNA 3 RMSD= 2.27	

O domínio catalítico pode ter sua formação com uma sequência pequena de aminoácidos, com o Loop de ligação ao metal em sua estrutura tridimensional como parecido com as estruturas de tRNA (Figura 1), que posteriormente foi complexificado com a incorporação de mais aminoácidos levando ao desenvolvimento do domínio palma primitivo, que poderia ser parecido com umas das Ances 90 ou 90.1(Figura 2). Com o advento do maquinário bioquímico moderno, as estruturas apenas incorporaram novos resíduos que formou novos domínios, gerando a diversidade estrutural que vemos nessas proteínas polimerases modernas[34,35,37,38].

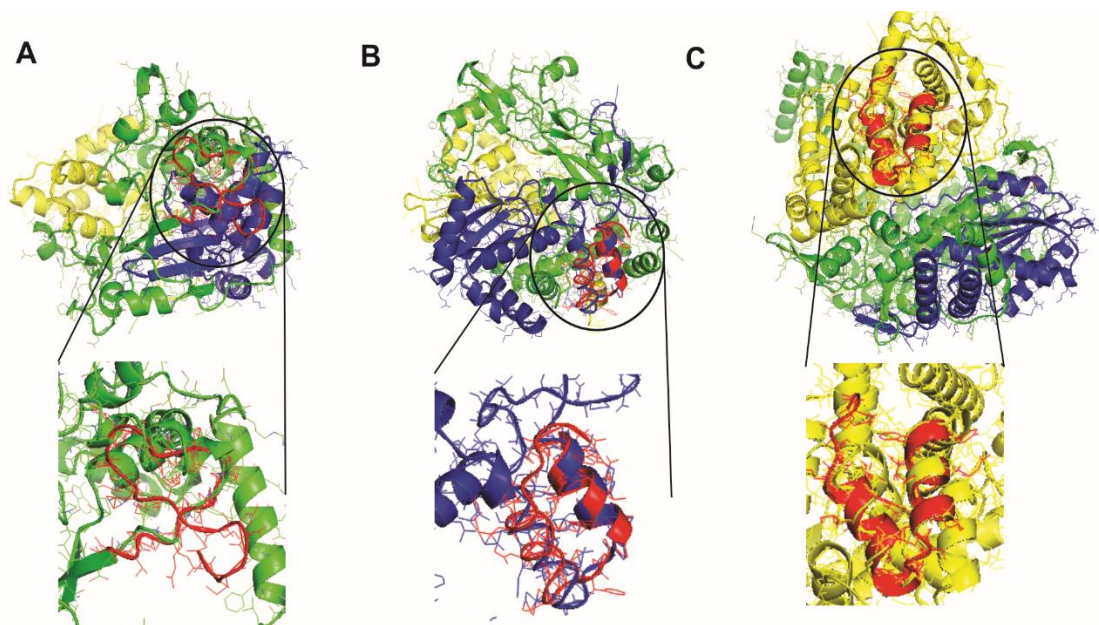


Figura 4: Alinhamento estrutural entre os modelos gerados pela tradução de proto-tRNA com RdRp moderna com o menor valor de RMSD. Em verde o domínio Dedos, em azul o domínio Palma, em amarelo o domínio polegar e em vermelho o modelo estrutural gerado pela tradução de concatâmeros de proto-tRNA. Em (A) o alinhamento estrutural entre o tRNA 2 e RdRp do Cardiovirus A, em (B) o alinhamento estrutural entre o tRNA 3 e RNA Polimerase DSRNA Bacteriophage e em (C) o alinhamento estrutural entre o tRNA 3 e RNA polimerase Rotavirus. Em destaque a porção estrutural alinhada entre cada modelo de tRNA e a proteína moderna.

As estruturas ancestrais para a RdRp 90 e 90.1 se ligaram em todos os domínios, tendo, porém, o domínio catalítico prevalente entre as sobreposições como pode ser observado na tabela 5. Esses resultados nos mostram que as estruturas ancestrais para a RdRp eram inicialmente pequenos domínios com afinidade por cofatores iônicos, que melhoram a funcionalidade dessas estruturas simples. Todas as estruturas modernas tiveram principalmente a estrutura Ances 90 se sobrepondo ao domínio catalítico, sendo muito provável a conservação do domínio palma que se formou a partir de estruturas mais simples[38,39].

Tabela 5: Resultado das sobreposições das ancestrais para a RDRP com as estruturas modernas virais, com seus respectivos valores de RMSD.

ESTRUTURA PDB	DEDOS	PALMA	POLEGAR
1HHS		90 RMSD = 3.14	90.1 RMSD = 4.02
1KHV		90 RMSD = 3.76	90.1 RMSD = 3.75
1MUK	90 RMSD = 3.12 90.1 RMSD = 3.12		
1S48		90 RMSD = 4.19	90.1 RMSD = 4.01
1U09		90 RMSD = 4.01	90.1 RMSD = 3.54
1XR7		90 RMSD = 3.17	90.1 RMSD = 4.02
2CKW		90 RMSD = 3.62	90.1 RMSD = 3.78
2D41		90 RMSD = 4.43	90.1 RMSD = 4.11
2J7N		90 RMSD = 3.24	90.1 RMSD = 3.63
2J7U		90 RMSD = 3.94	90.1 RMSD = 4.13
2PGG		90.1 RMSD = 3.30	90 RMSD = 3.26

ESTRUTURA PDB	DEDOS	PALMA	POLEGAR
2PUS		90 RMSD = 3.38 90.1 RMSD = 4.22	
2R7T	90 RMSD = 3.50 90.1 RMSD = 3.54		
2Y18		90.1 RMSD = 3.26	90 RMSD = 3.12
3JA4	90 RMSD = 3.67		90.1 RMSD = 4.36
3MMP		90 RMSD = 3.38 90.1 RMSD = 4.36	
3N6L		90 RMSD = 3.90	90.1 RMSD = 3.93
3OL6		90 RMSD = 4	90.1 RMSD = 3.98
3UQS		90 RMSD = 3.32	90.1 RMSD = 3.84
4AU6	90 RMSD = 3.98	90.1 RMSD = 4.40	
4HDH		90 RMSD = 4.40	90.1 RMSD = 3.71
4NZ0		90 RMSD = 3.94	90.1 RMSD = 3.88
4XHA		90 RMSD = 3.22	90.1 RMSD = 4.26
5AMQ	90.1 RMSD = 3.77	90 RMSD = 3.76	

4.3 Filogenia Estrutural

A árvore filogenética construída com os valores de RMSD das sobreposições (Figura 5) nos mostra um pouco sobre a evolução estrutural das proteínas Polimerases de RNA das viroses com as ancestrais hipotéticas. Como podemos observar, as RdRp apesar de seus domínios conservados, são bastante diferentes entre si, pois cada grupo de viroses possui suas próprias peculiaridades. Na base da árvore temos as ancestrais de tRNA 1 e 2, como nossas ancestrais hipotéticas. Esses resultados levam a hipótese de que inicialmente moléculas de tRNA poderiam originar os primeiros genes[8,43,44].

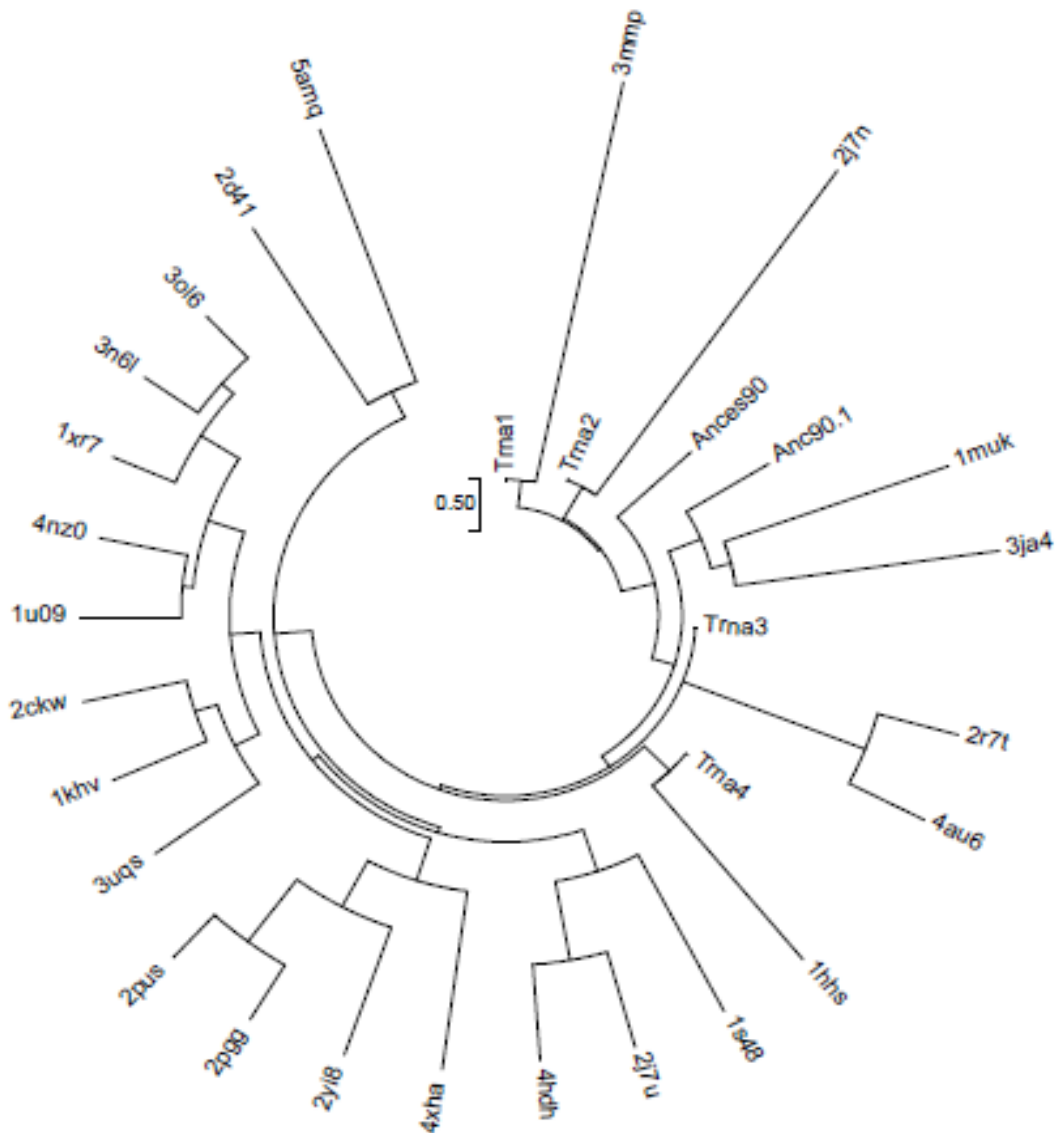


Figura 5: Filogenia Estrutural das Proteínas para a polimerase de RNA. Árvore construída pelo método Neighbor-joining, a partir de uma matriz com valores de RMSD, com código do PDB. 3mmp vírus Qbeta; 2j7n Neurospora Crassa; 1muk Reovirus; 3ja4 cypoviruses; 2r7t Rotavirus; 4au6 Rotavirus A; 1hhs DSRNA Bacteriophage; 1s48 Bovine viral diarrhea; 2j7u Dengue vírus; 4hdh Japanese encephalitis vírus; 4xha Thosea asigna vírus; 2yi8 Infectious Pancreatic Necrosis Vírus; 2pgg RNA Polimerase Birnavirus; 2pus Infectious bursal disease vírus; 3uqs Norwalk vírus; 1khv Rabbit Hemorrhagic Disease Vírus; 2ckw Sapporo Virus; 1u09 Foot-and-mouth disease vírus; 4nz0 Cardiovirus A; 1xr7 Rhinovirus A; 3n6l Enterovirus A; 3ol6 Enterovirus C; 2d41 Hepatitis C; 5amq La Crosse Bunyavirus;

5 DISCUSSÃO

Todas as moléculas ancestrais hipotéticas para a RdRp e tRNA aqui geradas, estão se ligando a cofatores, principalmente íons divalentes que são essenciais para atividade de polimerização enzimática[13,42,43]. Nós podemos inferir que o sítio catalítico foi o primeiro a se formar, como esperado devido a sua função. Os outros domínios muito prováveis foram incorporados mais tarde na evolução da polimerase[38,39]. Todas as classes de polimerase compartilham o sítio catalítico, principalmente resíduos de aminoácidos responsável por incorporação dos metais magnésio, tendo vários estudos mostrando que dentre essas polimerases a RdRp são o grupo mais antigo[36].

Seguindo a hipótese de que os tRNA deram origem aos primeiros genes[8,43,44], seria bastante provável na terra prébiótica a presença de vários cofatores assim como minerais em determinadas concentrações, com a presença de moléculas de ácido ribonucleico entre várias outras moléculas orgânicas disponíveis naquele tempo[6,7,10-12]. Haveria então a matéria-prima necessária para evolução de genomas RNA de entidades primitivas como vírus de RNA e os ancestrais progenotes do LUCA[7-9,28,31], onde as funções catalíticas inicialmente estavam se formando, assim poderia haver também uma troca de informação gênica por transferência horizontal, sendo a relação viroses com progenotes um evento importantíssimo para a evolução dos organismos celulares [6,12,13].

Inicialmente, uma ribozima com atividade de polimerase poderia ter aumentado a sua atividade por ligação em cofatores simples, como magnésio, assim, exercendo as funções de replicação de alguma informação armazenada em moléculas de RNA [24,25]. Com o surgimento do sistema de tradução primitiva, as primeiras proteínas a serem formadas pela tradução da junção de proto-tRNAs foi a enzima RdRp[8].

Acompanhado pelo processo de compartimentalização, com o controle de solutos no ambiente interno, a função de polimerização por proteína poderia ocorrer muito mais eficientemente do que uma ribozima, o que melhoraria o funcionamento do sistema bioquímico primitivo. Inicialmente, a enzima foi formada apenas pelo loop-catalítico com capacidade de ligação no cofator e moléculas simples como ribonucleótidos, como podemos ver nas Figuras 1 e 2, que posteriormente com a duplicação e diversificação do domínio catalítico inicial emergiram as outras partes da proteína. Com o estabelecimento do primeiro domínio na RdRp, as variantes poderiam ser geradas e mutações poderiam ocorrer, assim, era possível gerar novas proteínas, com a conservação do local catalítico[33].

Na tabela 4, observamos que todas as proteínas modernas possuem correspondências com moléculas de tRNA. Como observado na figura 4, as RdRp tiveram sua origem a partir do domínio catalítico, que emergiu das junções de proto-tRNAs. Com a formação desse domínio, através de duplicação, poderia ocorrer o surgimento dos outros domínios primitivos que vão originar os domínios dedos e polegar. O primeiro domínio formado de proto-tRNAs provavelmente funcionou como bloco de construção da organização inicial da enzima.

Com o surgimento de uma variante com propriedades da transcriptase reversa, e subsequentemente uma polimerase de DNA, ocorreu um passo fundamental para originar os primeiros genomas baseados em DNA, sendo a ponte para se mover de um mundo de RNA/Proteína para um mundo

DNA/RNA/Proteína [16,17,40]. Quando as moléculas de DNA emergiram, por variação, apareceram outras classes de polimerases. Com a complexificação do sistema biológico, emergiram funções como replicação, reparo e recombinação, as taxas de erro de incorporação foram selecionadas negativamente e as novas variantes surgiram para trabalhar em processos específicos durante a complexificação do sistema biológico inicial.

É sabido que esta polimerase seja a mais antiga dentre todas as polimerases DNA/RNA dependente[34-36]. Evidências atuais de vários trabalhos, nos levam a sugerir que os vírus, principalmente os de RNA, remetem a um passado muito antigo, sendo fósseis moleculares de um tempo prebiótico[28,31,42]. No tempo em que uma comunidade de protocelulas/progenote estavam em formação, seria muito provável que através de transferência horizontal de genes entre esses progenotes e as entidades virais, pudesse haver a emergência de um LUCA[6-9]. Essas entidades virais poderiam ajudar a inserir novos genes que se favorável a adaptação desses progenotes, levariam ao posterior desenvolvimento e evolução do ancestral de todas as células, e conseqüentemente a formação da diversidade celular que vemos atualmente.

Nas sobreposições entre as estruturas modernas para a RdRp e as estruturas ancestrais hipotéticas geradas a partir das sequências de aminoácidos, chamadas de Ances 90 e 90.1(Tabela 5). Tivemos em sua grande maioria a Ances 90 se encontrando no domínio palma ou próximo de onde há a ligação do metal. A maioria das estruturas ancestrais hipotéticas de tRNA estavam se ligando no domínio catalítico, sendo algumas próximas dos resíduos de incorporação de íons divalentes, e várias se ligando nos subdomínios dedos e polegar. Esses resultados podem servir como base para corroborar a hipótese de moléculas de tRNA originando os primeiros genes no maquinário bioquímico primitivo[43,44].

Além dessas tRNAs se ligarem a diversos cofatores, estão em sua grande maioria se ligando no subdomínio palma, sugerindo que foi o primeiro a se formar na evolução das polimerases, devido a sua função de polimerização, que começou com moléculas de RNA e posterior molécula de DNA, por isso se manteve conservada entre todas as classes de polimerases[33-39].

Seguindo a filogenia com dados estruturais (Figura 5), podemos observar que a estrutura mais basal é tRNA1 seguidas das outras tRNAs. Na base também encontramos a Ances 90 sendo a mais ancestral para as outras classes. Apesar da alta taxa de resíduos conservados as proteínas modernas virais possuem diferenças significativas de estrutura, que são característica para cada tipo de vírus[30]. Podemos inferir que a evolução estrutural das RdRp se iniciaram como um loop que se ligava a íons magnésio para melhorar a ocorrência da reação. Depois com a incorporação de outros resíduos esses íons tenderam a se inserir para dentro da enzima, enquanto os outros domínios estavam se formando, seguindo uma melhora na reação de polimerização e diminuição nas taxas de incorporação de nucleotídeos errados.

6 CONCLUSÃO

Responder a questões tão profundas não é fácil, principalmente em relação de como a vida surgiu na terra e foi capaz de se desenvolver. Ainda precisamos de uma vasta gama de experimentos que demonstre mais precisamente como as primeiras moléculas se formaram em um ambiente caótico para depois se organizar de forma a originar os primeiros sistemas biológicos. Nosso trabalho leva a sugerir, que a enzima RdRp se originou a partir do domínio catalítico “Palma”, sendo muito pequena e através da incorporação de pequenos resíduos de aminoácidos emerge o domínio catalítico mais parecido com o evidenciado hoje. Posteriormente, os outros domínios seriam incorporados. As viroses de RNA foram essenciais no processo de formação dos primeiros genomas, seria muito provável que no momento da fixação de genes ao código genético haveria uma troca de informação muito grande entre entidades antigas (Vírus de RNA, progenotes/LUCA). Principalmente quando as funções de catálise passaram a ser exercidas por proteínas, genes de polimerização foram um dos primeiros a se formar e a se fixarem ao genoma. Além disso, com o surgimento de variantes e o aparecimento de outras classes de polimerases poderia ocorrer o aparecimento dos primeiros genomas de DNA. Como os resíduos envolvidos na catálise são altamente conservados, mostrando que tiveram uma origem em comum, é plausível sugerir duplicação seguida de mutação para o desenvolvimento de novas proteínas, dando origem as várias classes de polimerase que vemos hoje nos organismos modernos. Entretanto, vários outros estudos precisam ser realizados para podermos entender com mais clareza como ocorreu a formação dos primeiros genomas.

REFERÊNCIAS

- 1- KAWAMURA, K. **A Hypothesis: Life Initiated from Two Genes, as Deduced from the RNA World Hypothesis and the Characteristics of Life-Like Systems.** *Life*, Vol. 6, 29, 2016.
- 2- LAZCANO, A.; MILLER, S. L. **The Origin and Early Evolution Review of Life: Prebiotic Chemistry, the Pre-RNA World.** *Cell*, Vol. 85, 793–798, 1996.
- 3- HIGGS, P. G.; LEHMAN, N. **The RNA World: molecular cooperation at the origins of life.** *Nature Reviews Genetics*, Vol.16, 7-17, 2014.
- 4- DODD, M. S.; PAPINEAU, D.; GRENNE, T.; SLACK, J. F.; RTTNER, M.; PIRAJNO, F.; O'NEIL, J.; LITTLE, C. T. S. **Evidence for early life in Earth's oldest hydrothermal vent precipitates.** *Nature*, Vol. 543, 60-64, 2017.
- 5- PRESSMAN, A.; BLANCO, C.; CHEN, I. A. **The RNA World as a Model System to Study the Origin of Life.** *Current Biology*, Vol. 25, 953–963, 2015.
- 6- JHEETA, S. **The Routes of Emergence of Life from LUCA during the RNA and Viral World: A Conspectus.** *Life*, Vol. 5, 1445-1453, 2015.
- 7- WEISS, M. C.; SOUSA, F. L.; MRNJAVAC, N.; NEUKIRCHEN, S.; ROETTGER, M.; SATHI, S. N.; MARTIN, W. F. **The physiology and habitat of the last universal common ancestor.** *Nature microbiology*, Vol. 1, 116, 2016.
- 8- FARIAS, S. T.; RÊGO, T.G.; JOSE, M. V. **A proposal of the proteome before the Last Universal Common Ancestor.** *International Journal of Astrobiology*, Vol. 15, 27-31, 2015.
- 9- FARIAS S. T.; PROSDOCIMI, F. **Buds of the tree: the highway to the last universal common ancestor.** *International Journal of Astrobiology*, Vol.16, 105-113, 2016.
- 10- PEARCE, B. K. D.; PUDRITZ, R. E. **Meteorites and the RNA World: A Thermodynamic Model of Nucleobase Synthesis within Planetesimals.** *Astrobiology*, Vol. 16, 853-872, 2016.

- 11-ORGEL, L. E. **Prebiotic Chemistry and the Origin of the RNA World.** *Critical Reviews in Biochemistry and Molecular Biology*, Vol. 39, 99–123, 2004.
- 12-ROTELLI, L.; RODRIGUEZ, J.M.T.; CAMBERO, C.E.M.; CAROTA, E.; BOTTA, L.; MAURO, E.D.; SALADINO, R. **The key role of meteorites in the formation of relevant prebiotic molecules in a formamide/water environment.** *Scientific Reports*, Vol. 6, 2016.
- 13-RAGGI, L.; BABA, J.L.; LAZCANO, A. **On the lack of evolutionary continuity between prebiotic peptides and extant enzymes.** *Physical Chemistry Chemical Physics*, Vol. 18, 2016.
- 14-KUMAR, V. A. **Evolution of specific 3'-5'-linkages in RNA in prebiotic soup: a new hypothesis.** *Organic and Biomolecular Chemistry*, Vol. 14, 10123-10133, 2016.
- 15-LEHMAN, N. **The RNA World: 4,000,000,050 years old.** *Life*, Vol. 5, 1583–1586, 2015.
- 16- Gilbert, W. **The RNA world.** *Nature*, Vol. 319, 618, 1986.
- 17-MULLER, U.F. **Re-creating an RNA World.** *Cellular and Molecular Life Sciences*, Vol. 63, 1278–1293, 2006.
- 18-RANA, A. K.; ANKRI, S. **Reviving the RNA World: An Insight Into the Appearance of RNA Methyltransferases.** *Frontiers in Genetics*, Vol. 7, 99, 2016.
- 19-VLASSOV, A. V.; KAZAKOV, S. A.; JOHNSTON, B. H.; LANDWEBER, L. F. **The RNA World on Ice: A New Scenario for the Emergence of RNA Information.** *Journal of Molecular Evolution*, Vol. 61, 264–273, 2005.
- 20-BOWMAN, C. J.; HUD, V. N.; WILLIAMS, L. D. **The Ribosome Challenge to the RNA World.** *Journal of Molecular Evolution*, Vol. 80, 143-161, 2015.
- 21-POOLE, A. M.; JEFFARES, D. C.; HOEPPNER, P. M.; PENNY, D. **Does the Ribosome Challenge our Understanding of the RNA World?.** *Journal of Molecular Evolution*, Vol. 82, 1-4, 2015.

- 22-LEHMAN, N. **RNA in evolution**. WIREs RNA, Vol. 1, 202–213, 2010.
- 23-ELLINGTON, A. D.; CHEN, X.; ROBERTSON, M.; SYRETT, A. **Evolutionary origins and directed evolution of RNA**. The International Journal of Biochemistry & Cell Biology, Vol. 41, 254–265, 2009.
- 24-TAGAMI, S.; ATTWATER, J.; HOLLIGER, P. **Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function**. Nature Chemistry, Vol. 9, 325–332, 2017.
- 25-HIGGS, P. G.; KIM, Y. E. **Co-operation between Polymerases and Nucleotide Synthetases in the RNA World**. PLOS Computational Biology, Vol. 12, 2016.
- 26-SARKIES, P.; MISKA, E. A. **Is there social RNA?**. Science, Vol. 341, 467–468, 2013.
- 27-ADELMAN, K.; EGAN, E. **Non-coding RNA: More uses for genomic junk**. Nature, Vol 543, 183–185, 2017.
- 28-FORTERRE, P. **The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells**. Biochimie, Vol. 87, 793–803, 2005.
- 29-GAVETTE, J. V.; STOOP, M.; HUD, N. V.; KRISHNAMURTHY, R. **RNA–DNA Chimeras in the Context of an RNA World Transition to an RNA/DNA World**. Angewandter Chemie (International ed. In English), Vol. 55, 13204–13209, 2016.
- 30-CERNY´, J.; BOLFIKOVA´, B.C.; VALDES, J.J.; GRUBHOFFER, L.; RUZEK, D. **Evolution of Tertiary Structure of Viral RNA Dependent Polymerases**. PLOS ONE, Vol. 9, 2014.
- 31-NASIR, A.; ANOLLES, G. C. **A phylogenomic data-driven exploration of viral origins and evolution**. Science Advances Virology, Vol. 1, 2015.
- 32-AHLQUIST, P. **RNA-Dependent RNA Polymerases, Viruses, and RNA Silencing**. Science, Vol. 296, 1270–1273, 2002.

- 33-ZONG, J.; YAO, X.; YIN, J.; ZHANG, D.; MA, H. **Evolution of the RNA-dependent RNA polymerase (RdRP) genes: Duplications and possible losses before and after the divergence of major eukaryotic groups.** *Gene*, Vol. 447, 29-39, 2009.
- 34- IYER, L.M.; KOONIN, E.V.; ARAVIND, L. **Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases.** *BMC Structural Biology*, Vol. 3, 2003.
- 35-JACOME, R.; BECERRA, A.; LEON, S. P.; LAZCANO, A. **Structural Analysis of Monomeric RNA Dependent Polymerases: Evolutionary and Therapeutic Implications.** *PLOS ONE*, Vol. 10, 2015.
- 36-MONTTINEN, H.A.M.; RAVNTTI, J.J.; STUART, D.I.; PORANEN, M.M. **Automated Structural Comparisons Clarify the Phylogeny of the Right-Hand-Shaped Polymerases.** *Molecular Biology and Evolution*. Vol. 31, 2741–2752, 2014.
- 37-OREILLY, K.E.; KAO, C.C. **Analysis of RNA-Dependent RNA Polymerase Structure and Function as Guided by Known Polymerase Structures and Computer Predictions of Secondary Structure.** *Virology*, Vol. 252, 287–303, 1998.
- 38-POCH, O.; SAUVAGET, I.; DELARUE, M.; TORDO, N. **Identification of four conserved motifs among the RNA dependent polymerase encoding elements.** *The EMBO Journal*, Vol. 8, 3867-3874, 1989.
- 39-NG, K.K.S.; ARNOLD, J.J.; CAMERON, CE. **Structure-function relationships among RNA-dependent RNA polymerases.** *Current Topics Microbiology and Immunology*, Vol. 320, 137-156, 2008.
- 40-LEON, P.E. **Inhibition of Ribozymes by Deoxyribonucleotides and the Origin of DNA.** *Journal of Molecular Evolution*, Vol. 47, 122–126, 1998.
- 41-ČERNÝ, J.; BOLFÍKOVÁ, C. B.; ZANOTTO, P.M.A.; GRUBHOFFER, L.; RŮŽEK, D. **A deep phylogeny of viral and cellular right-hand polymerases.** *Infection Genetics and Evolution*, Vol. 36, 275-286, 2015.

- 42-PRIETO, F.R.; MORALES, H.R.; JACOME, R.; BECERRA, A.; LAZCANO, A. **Coenzymes, viruses and the RNA world**. *Biochimie*, Vol. 94, 1467-1473, 2012.
- 43-FARIAS, S.T.; RÊGO, T.G.; JOSE, M.V. **tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World**. *Life*, Vol. 6, 15, 2015.
- 44-FARIAS, S.T.; RÊGO, T.G.; JOSE, M.V. **Origin and evolution of the Peptidyl Transferase Center from proto-tRNAs**. *FEBS Open Bio*, Vol. 4, 175–178, 2014.
- 45-ZHANG, Y. **I-TASSER server for protein 3D structure prediction**. *BMC Bioinformatics*, Vol. 9, 2008.
- 46-XU, D.; ZHANG, Y. **Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization**. *Biophysical Journal*, Vol. 101, 2525-2534, 2011.
- 47-YANG, J.; ROY, A.; ZHANG, Y. **Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment**. *Bioinformatics*, Vol. 29, 2588-2595, 2013.
- 48-The UniProt Consortium. **UniProt: the universal protein knowledgebase**. *Nucleic Acids Research*, Vol. 45, 158-169, 2017.
- 49-KAZUTAKA, K.; KAZUHARU, M.; KEI-ICHI, K.; TAKASHI, M. **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform**. *Nucleic Acids Research*, Vol. 30, 3059–3066, 2002.
- 50-DARRIBA, D.; TABOADA GL.; DOALLO, R.; POSADA, D. **ProtTest 3: fast selection of best-fit models of protein evolution**. *Bioinformatics*, Vol. 27, 1164-1165, 2011.
- 51-DIALLO, B. A.; ALPHA, B.; MAKARENKOV, V. **T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks**. *Nucleic Acids Research*, Vol. 40, 573-579, 2012.
- 52-KUMAR, S.; STECHER, G.; TAMURA, K. **MEGA7: Molecular Evolutionary Genetics Analysis version 7.0**. *Molecular Biology and Evolution*, Vol. 33, 1870-1874, 2015.
- 53-ROSE, P. W.; PRLIĆ, A.; ALTUNKAYA, A.; BI, C.; BRADLEY, A. R.; CHRISTIE, C. H.; COSTANZO, L. D.; DUARTE, J. M.; DUTTA, S.; FENG, Z.; GREEN, R. K.; GOODSSELL, D. S.; HUDSON, B.; KALRO, T.;

LOWE, R.; PEISACH, E.; RANDLE, C.; ROSE, A. S.; SHAO, C.; TAO, Y.P.; VALASATAVA, Y.; VOIGT, M.; WESTBROOK, J.D.; WOO, J.; YANG, H.; YOUNG, J. Y.; ZARDECKI, C.; BERMAN, H.M.; BURLEY, S. K. **The RCSB protein data bank: integrative view of protein, gene and 3D structural information.** Nucleic Acids Research, Vol. 45, 271-281, 2017.

54-ZHANG, Y.; SKOLNICK, J. **TM-align: A protein structure alignment algorithm based on TM-score.** Nucleic Acids Research, Vol. 33, 2302-2309, 2005.