

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

LIDIANE SILVA DO NASCIMENTO

SINGULARIDADES ORGÂNICAS:
ANÁLISE DAS ADAPTAÇÕES DAS PROTEÍNAS RIBOSSÔMICAS UNIVERSAIS DE
ORGANISMOS EXTREMÓFILOS HALÓFILOS DOS DOMÍNIOS *ARCHAEA* E
BACTERIA

JOÃO PESSOA - PB

2020

LIDIANE SILVA DO NASCIMENTO

SINGULARIDADES ORGÂNICAS:
ANÁLISE DAS ADAPTAÇÕES DAS PROTEÍNAS RIBOSSÔMICAS UNIVERSAIS DE
ORGANISMOS EXTREMÓFILOS HALÓFILOS DOS DOMÍNIOS *ARCHAEA* E
BACTERIA

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

Orientadora: Profª. Dra. Thaís Gaudencio do Rêgo

Coorientador: Prof. Dr. Sávio Torres de Farias

JOÃO PESSOA - PB

2020

Catálogo na publicação
Seção de Catalogação e Classificação

N244s Nascimento, Lidianne Silva do.

Singularidades orgânicas : análise das adaptações das proteínas ribossômicas universais de organismos extremófilos halófilos dos domínios archaea e bacteria / Lidianne Silva do Nascimento. - João Pessoa, 2020.
74 f. : il.

Orientação: Thaís Gaudencio do Rêgo.

Coorientação: Sávio Torres de Farias.

Monografia (Curso de Bacharelado em Ciências Biológicas) - UFPB/CCEN.

1. Microrganismos halófilos. 2. Proteínas ribossômicas. I. Rêgo, Thaís Gaudencio do. II. Farias, Sávio Torres de. III. Título.

UFPB/CCEN

CDU 57(043.2)

FOLHA DE APROVAÇÃO

LIDIANE SILVA DO NASCIMENTO

SINGULARIDADES ORGÂNICAS:

ANÁLISE DAS ADAPTAÇÕES DAS PROTEÍNAS RIBOSSÔMICAS UNIVERSAIS DE
ORGANISMOS EXTREMÓFILOS HALÓFILOS DOS DOMÍNIOS *ARCHAEA* E
BACTERIA

Monografia apresentada ao curso de Ciências Biológicas (Trabalho Acadêmico de Conclusão de Curso) como requisito parcial para a obtenção do grau de bacharelado em Ciências Biológicas da Universidade Federal da Paraíba.

Aprovado em: 17 / 04 / 2020

BANCA EXAMINADORA

Orientadora: Profª. Dra. Thaís Gaudencio do Rêgo
(Departamento de Informática - CI - UFPB, João Pessoa, PB)

Coorientador: Prof. Dr. Sávio Torres de Farias
(Departamento de Biologia Molecular - CCEN - UFPB, João Pessoa, PB)

Membro Interno: Prof. Dr. Luis Fernando Marques dos Santos
(Departamento de Biologia Molecular - CCEN - UFPB, João Pessoa, PB)

Membro Externo: Dr. Luiz Thibério Rangel
(Massachusetts Institute of Technology (MIT), Cambridge, MA, USA)

Dedico à minha amada família e ao Santos Futebol Clube.

AGRADECIMENTOS

Agradeço as leis fundamentais da Física por estruturarem um Universo extraordinariamente belo, com as condições ideais para o aparecimento da Química como a conhecemos, e a ambos, por estruturarem uma base físico-química para o fenômeno histórico mais impressionante e complexo de todo o Universo conhecido: os sistemas biológicos.

Agradeço a aleatoriedade, que permitiu com que a minha carga genética fosse selecionada para participar do fenômeno da existência neste pálido ponto azul terrestre.

Agradeço a oportunidade de sentir a sensibilidade de amar a realidade, de amar os sistemas recursivos da vida orgânica e social, a expectativa de observar a explosão de uma supernova, a visão de uma célula se dividindo, o aspecto delicado de um beija-flor voando pelo jardim da UFPB, e sobretudo, de amar o lugar onde estou nesse momento.

Agradeço pela estruturação cerebral e pelas experiências que me fizeram querer viver ao lado da Ciência, pois ela me fez enxergar o que é sentir o Universo.

Agradeço aos meus pais, Maria Inês, Antonio e Maria das Neves por todo amor e incentivo. Vocês são os responsáveis por toda a minha força. São as estrelas que brilham no meu céu interno quando penso em desistir. Sou fruto de uma soma multiplicada de vocês, muitas vezes desajeitada, mas completamente apaixonada por ter em cada centímetro do meu corpo bilhões de bases nitrogenadas que especificam que eu sou uma manifestação aleatória de vocês. Tudo de bom que fui, o que sou e o que serei, pertence a vocês.

Aos meus colegas de turma Adriennius Marques, Ana Clara Martins, Anna Beatriz Marques, Brenda Luany Almeida, Catarina Serrão, Dalette Suzzanne, Daniel de Souza, Hilquias da Silva, Jamilah Lucena, Mateus Santos Estrela, Natalia das Neves, Nathalia Correia, Rafael Vitor de Oliveira, Ruãn Pontes e Rubens de Almeida por cada atitude de solidariedade, empatia e por terem sido modelos intelectuais para mim. Sinto muito orgulho de ter dividido um curso e um pedaço do meu tempo de existência com vocês.

À Jamilah Lucena, por ser um centro de amizade, paciência e incentivo nessa caminhada para o término da nossa graduação. Sua força me fez enxergar a vida e os outros de uma forma mais sensível. Obrigada tendendo ao infinito.

À Camila Yasmin, por aceitar a difícil tarefa de me entender e me aceitar. Sua amizade é como um girassol na minha vida. Seu otimismo tende a se unir com o meu pessimismo e neutralizá-lo. Muito obrigada por sempre falar que tudo vai dar certo, mesmo que estatisticamente e humanamente isso seja quase impossível, sempre me acalmou. “Diante da

vastidão do tempo e da imensidão do universo, é um imenso prazer para mim dividir um planeta e uma época com você.” - Carl Sagan.

À Ruã Pontes, um ser genial que cruzou o meu caminho e me trouxe noções encantadoras, outrora desconhecidas. “Diante da vastidão do tempo e da imensidão do universo, é um imenso prazer para mim dividir um planeta e uma época com você.” - Carl Sagan.

Aos meus amigos, que transformaram meus desencontros sociais em piadas, e com isso, me fizeram enxergar a vida em grupo com outros olhos. Hoje, vocês são uma parte do que eu sou.

Aos meus colegas de laboratório, em especial a João e a Livia que foram como anjos nesses últimos dois meses. Um obrigada não é suficiente para expressar o meu agradecimento por tudo que fizeram e por tudo que compartilhamos. Guardarei no meu coração as expressões doces de vocês.

À minha orientadora Thaís Gaudencio do Rêgo, pela incrível experiência de viver um pouco ao seu lado. Por me acolher e me ensinar lições que guardarei eternamente. Obrigada por enxergar as pessoas de uma forma tão sensível. A senhora ilumina muitas mentes. Nunca conheci ninguém que tenha cruzado o seu caminho e não tivessem encontrado bondade e paciência. És um exemplo incrível de mulher, professora, orientadora e tantas outras faces que tens, que se mesclam e se complementam e permitem com que muitas pessoas floresçam. Thaís Gaudencio é um LABORATÓRIO INTEIRO, porque ela sempre conhece alguém que pode resolver qualquer coisa, e faz questão de fazer de tudo para que as ideias mais loucas dos seus alunos saiam das suas mentes e se materializem no papel. Nela, visualizei a verdadeira ciência colaborativa. Tenho certeza que todas as grandes cientistas do passado e do presente, sentiriam orgulho em conhecê-la.

Ao professor Luis Fernando Marques, por ser mais que um professor, foi um amigo em muitos momentos. Obrigada por toda a paciência durante grande parte da minha graduação. Se um dia eu chegar a ser professora, muito de mim, será baseado no que observei no senhor. Todos que possuem a oportunidade de cruzar o seu caminho intelectual, são privilegiados.

Ao professor Alexandre Vasconcellos, pelo abrigo no seu laboratório e por ser uma inspiração de dedicação diária como pesquisador e professor. Me sinto honrada por ter tido a oportunidade de ser sua aluna e observar um pouco do seu trabalho de perto.

Ao meu orientador Sávio Torres de Farias, por ter sido uma inspiração desde os meus primeiros dias no curso. Seus trabalhos foram um dos principais motivos pelo meu encantamento pela evolução molecular da vida. Muito obrigada por ter alimentado minha mente com suas ideias brilhantes e sua energia para inovar. Não tenho como agradecer e explicar o quão honrada sou por tê-lo como meu orientador. Tentar entender suas ideias, me

fizeram, no meu ver, amadurecer. O livro: A emergência dos sistemas biológicos, de sua autoria e do Francisco Prosdocimi, foi como um presente pessoal. Que esse seja o início de muitas outras aventuras no estudo da evolução da vida.

À todos os meus amores científicos: Albert Einstein, Isaac Newton, Marie Curie, Lynn Margulis, Charles Robert Darwin, Gregor Mendel, Rosalind Franklin, Nikola Tesla, Edwin Powell Hubble, Ernst Mayr, Oparin, Haldane, Miller e tantos outros que me proporcionaram dúvidas eternas sobre os seus trabalhos e o amor por imaginar. E em especial, ao Carl Sagan, por iluminar o mundo e o meu mundo pessoal, com a sua singular paixão pelo Universo humano e cósmico.

Aos professores, funcionários e alunos do Departamento de Biologia Molecular (DBM) e do Departamento de Sistemática e Ecologia (DSE), por me fornecerem ombros fortes para estabelecer uma agradável graduação.

Agradeço a todos aqueles que contribuíram para a estruturação da Bioinformática e da Astrobiologia. Enquanto a primeira me fornece acesso a níveis elevados de dados biológicos, a segunda me fornece perguntas que fazem com que eu olhe para a natureza com a mais pura sensação de deslumbramento. Juntas, me forneceram a possibilidade de estudar elementos antigos da história da vida na Terra com computadores modernos. É uma confusão temporal deliciosa. Obrigada!!!

Um agradecimento especial a minha banca, por terem aceitado gastar o que vocês possuem de mais valioso, o tempo, lendo e avaliando o meu trabalho.

As “Lidianes” do passado e do presente que não desistiram em meio a todos os dramas internos e externos, muitas vezes, quase cômicos.

*Où finit le télescope, le microscope commence.
Lequel des deux a la vue la plus grande?*

Onde o telescópio termina, o microscópio começa.
Qual dos dois oferece visão mais grandiosa?
~ Victor Hugo (Les Misérables, 1862)

“Somos feitos de poeira das estrelas”.
~ Carl Sagan, 1986

RESUMO

O grupo dos halófilos reúne organismos que, através de adaptações estruturais e fisiológicas, possuem a capacidade de sobreviver em ambientes com extremos de salinidade. Uma das grandes características apontadas em relação às adaptações moleculares dos halófilos é em relação ao caráter ácido da maioria das suas proteínas estudadas. É plenamente conhecido que os ribossomos em todas as formas de vida são formados por interações entre diversas proteínas e moléculas de RNA e, um dos papéis biológicos das proteínas ribossômicas, é a neutralização das cargas negativas do RNA ribossomal. Os estudos das sequências primárias dessas proteínas em organismos mesófilos apontam que uma das suas principais características é o alto teor de aminoácidos básicos, importantes para a estabilização do complexo ribonucleoproteico. Utilizando esses dados como ponto de partida, o presente trabalho teve como objetivo avaliar as adaptações das proteínas ribossômicas dos organismos halófilos dos domínios *Archaea* e *Bacteria*, através de análises estatísticas baseadas nos dados de sequências primárias disponíveis no NCBI. Com um total de 885 espécies de organismos extremófilos halofílicos pertencentes aos dois domínios da vida procariótica: *Archaea* (260) e *Bacteria* (625), e 190 espécies de organismos não extremófilos utilizados como controle, foram extraídas um total de 17.255 sequências correspondentes as 34 proteínas ribossomais universais. Após a quantificação dos aminoácidos dos arquivos multifasta correspondentes a cada grupo, testes estatísticos foram aplicados para avaliar e comparar a composição de aminoácidos dos mesmos. Como resultado, obtivemos dois perfis para as proteínas ribossômicas dos halófilos: as proteínas arqueanas, assim como o padrão apresentado em estudos anteriores, apresentou porcentagens maiores de aminoácidos ácidos e menores de aminoácidos hidrofóbicos aromáticos e alifáticos em comparação com as sequências mesofílicas controle; entretanto, as sequências dos halófilos bacterianos não mostraram diferenças estatísticas significativas em relação ao seu grupo controle. Concluímos que as proteínas ribossômicas dos halófilos, principalmente do domínio *Archaea*, possuem uma composição expressivamente ácida, esse resultado se torna interessante devido à interação íntima entre os elementos proteicos e os ácidos ribonucléicos que também possuem uma natureza ácida.

Palavras-chave: microrganismos halófilos; proteínas ribossômicas homólogas; composição de aminoácidos.

ABSTRACT

The group of halophiles brings together organisms that, through structural and physiological adaptations, can survive in environments with extreme salinity. One of the great characteristics pointed out about the molecular adaptations of halophiles relates to the acid character of most of their studied proteins. It is known that ribosomes in all forms of life are formed by interactions between various proteins and RNA molecules and one of the biological roles of ribosomal proteins is neutralization of negative ribosomal RNA charges. The studies of the primary sequences of these proteins in mesophilic organisms point out that one of their main characteristics is the high content of basic amino acids, important for the stabilization of the ribonucleoprotein complex. Using this data as a starting point, the present study aimed to evaluate the adaptations of ribosomal proteins of halophiles organisms in the *Archaea* and *Bacteria* domains, through basic statistical analysis based on primary sequence data available at NCBI. With a total of 885 species of extreme halophiles organisms belonging to two domains of prokaryotic life: *Archaea* (260) and *Bacteria* (625), and 190 species of non-extreme species used as control, were extracted a total of 17,255 associated sequences as 34 universal ribosomal proteins. After quantifying the amino acids in the multiphase files corresponding to each group, statistical tests were used to evaluate and compare their amino acid composition. As a result, we obtained two profiles for the ribosomal proteins of halophiles: the Archean proteins, as well as the pattern presented in previous studies, presented higher percentages of acidic amino acids and lower percentages of aromatic and aliphatic hydrophobic amino acids compared to the control mesophilic sequences; however, the sequences of the bacterial halophiles did not show statistically significant differences to their control group. We conclude that the ribosomal proteins of halophiles, mainly from the *Archaea* domain, have an expressively acidic composition, due to the close interaction between protein elements and ribonucleic acids.

Keywords: halophilic microorganisms; homologous ribosomal proteins; composition of amino acids.

LISTA DE FIGURAS

Figura 1- Distribuição dos organismos halófilos nos três domínios da vida celular.....	24
Figura 2- Mapa global dos níveis de salinidade dos oceanos terrestres.....	25
Figura 3- Ilhas de sal no Mar Morto, Israel.....	25
Figura 4- a) O Grande Lago Salgado (Great Salt Lake) de Utah. b) Colônias de arqueias halofílicas de Great Salt Lake que crescem em meios sólidos.....	26
Figura 5 - Os 20 aminoácidos separados com base nas propriedades das suas cadeias laterais.....	27
Figura 6 - Representação de ligações peptídicas.....	28
Figura 7- Os vinte aminoácidos que compõem as proteínas com suas cadeias laterais coloridas para o destacamento do seu perfil químico.....	29
Figura 8- Representação esquemática do ribossomo de <i>Thermus thermophilus</i> 70S a uma resolução de 5,5 Å, com seus domínios destacados (A, P e E).....	33
Figura 9- Esquema da estrutura ribonucleoproteína do ribossomo da espécie <i>Thermus thermophilus</i>	36
Figura 10- Esquema das intersecções do conjunto das proteínas ribossômicas entre os três domínios da vida celular.....	37
Figura 11- Esquema explicativo dos testes usados para avaliar a significância dos valores comparados.....	47
Figura 12 - Resultado do teste de correlação de Pearson.....	66

LISTA DE TABELAS

Tabela 1 - Os vinte aminoácidos e seus respectivos códons (código genético padrão).....	31
Tabela 2 - Quantidades de proteínas extraídas por grupo.....	43
Tabela 3 - Contagem das sequências das 34 proteínas referentes a cada grupo.....	44
Tabela 4 - Dados das médias, porcentagens, desvio padrão e resultados do teste t de Student dos 20 aminoácidos presentes nas 34 proteínas ribossômicas do domínio Bacteria (halófilos + controles).....	50
Tabela 5 - Dados das médias, porcentagens, desvio padrão e resultados do teste t de Student dos 20 aminoácidos presentes nas 34 proteínas ribossômicas do domínio Archaea (halófilos + controles).....	51
Tabela 6 - Porcentagem dos aminoácidos das 34 proteínas ribossômicas agrupados pelas características de suas cadeias laterais.....	52
Tabela 7 - Perfil dos aminoácidos hidrofóbicos e carregados das proteínas dos dois domínios.....	54
Tabela 8 - Porcentagem dos aminoácidos das 19 proteínas ribossômicas da subunidade maior (LSU) agrupados pelas características químicas das suas cadeias laterais.....	63
Tabela 9 - Porcentagem dos aminoácidos das 15 proteínas ribossômicas da subunidade menor (SSU) agrupados pelas características químicas das suas cadeias laterais.....	65
Tabela 10 - Tamanho médio das sequências das 34 proteínas ribossômicas dos quatro grupos comparados.....	67
Tabela 11 - Tamanho médio das proteínas através da comparação geral das 34 proteínas e por subunidade dos grupos de halófilos e seus controles.....	68

SUMÁRIO

1. INTRODUÇÃO	16
1.1. CONTEXTO E PROBLEMA	16
2. OBJETIVOS	20
2.1. OBJETIVO GERAL	20
2.2. OBJETIVOS ESPECÍFICOS	20
3. FUNDAMENTAÇÃO TEÓRICA	21
4. METODOLOGIA	40
4.1. DADOS DOS ORGANISMOS	40
4.2. PROTEÍNAS ANALISADAS NO ESTUDO	40
4.3. CONSTRUÇÃO DOS ARQUIVOS MULTIFASTA	41
4.4. SEQUÊNCIAS EXTRAÍDAS	42
4.5. ANÁLISE ESTATÍSTICA E DEMONSTRAÇÃO GRÁFICA	43
4.6. TESTES UTILIZADOS	45
5. RESULTADOS E DISCUSSÃO	47
6. CONSIDERAÇÕES FINAIS	69
7. PERSPECTIVAS FUTURAS	70
REFERÊNCIAS BIBLIOGRÁFICAS	71

1. INTRODUÇÃO

1.1. CONTEXTO E PROBLEMA

Diversos pontos sobre a emergência dos sistemas biológicos ainda são de difícil compreensão, devido a dificuldade de se explicar como algo tão complexo quanto uma célula viva poderia evoluir. Porém, já é bem estabelecido que os 3,8 bilhões de anos de evolução da vida na Terra configuraram um amplo espectro de adaptações, que proporcionaram aos seres vivos dominarem os mais diversos ambientes, desde os com condições consideradas amenas, até os ambientes mais inóspitos (DUARTE, 2016; REED, 2013).

Na classificação atualmente aceita, com base em análises genômicas, a vida celular é dividida em três domínios: *Archaea*, *Bacteria* e *Eukarya* (WOESE et al., 1990). Nos dois primeiros grupos, conhecidos como procariotos, a vida é restrita a formas unicelulares com uma complexidade celular considerada inferior à dos organismos eucariontes, sendo esse grupo caracterizado por possuir organismos unicelulares e multicelulares portadores de um sistema endomembranar que permitiu a evolução de uma organela específica chamada de núcleo, além de outros pontos que os diferenciam (ALBERTS, 2017). Mesmo com essa divisão básica, nos três domínios são encontrados exemplos das mais variadas adaptações.

A complexidade do fenômeno vida desafiou os estudiosos ao longo do histórico científico. Toda nova descoberta ao longo do tempo revela a singularidade dessa estruturação orgânica, organizada de uma forma insólita. No ano de 1964, o microbiologista norte-americano Thomas Brock fez uma descoberta que mudou noções paradigmáticas de até onde a vida poderia chegar. Ele descobriu que existiam organismos sobrevivendo ao redor de gêiseres do Parque Nacional do Yellowstone (EUA), que lançavam água a 82 °C, uma temperatura muito acima da tolerada pelos organismos conhecidos na época. Três anos após essa descoberta inicial, em 1967, ele publicou um trabalho onde descreve esse novo grupo de organismos, classificados dentro do domínio *Bacteria* e batizados de *Thermus aquaticus* (BROCK, 1967). Neste artigo ele descreve que os mesmos não apenas toleram as elevadas temperaturas, mas que essa condição deve ser respeitada para permitir que eles sobrevivam e se reproduzam (BROCK, 1967; DUARTE, 2016). A partir dessa descoberta, um novo grupo de organismos passou a ser conhecido: os extremófilos; gerando muitas perguntas e possibilidades para a Biologia.

Após esse descobrimento, estudos foram feitos em diversos ambientes outrora considerados inóspitos para a vida (MERINO et al., 2019). E a medida que o número de pesquisas aumentavam, um cenário biológico com diferentes estratégias metabólicas, que proporcionaram a capacidade de sobrevivência em ambientes com extremos de temperatura, salinidade, acidez, alcalinidade, pressão e radiação foi desabrochando (REED, 2013; DUARTE, 2016). Com esse fato, podemos perceber que o processo evolutivo não pontilha a Terra com vida em lugares específicos, ele faz com que a vida se ramifique pelo planeta, cobrindo-o em todas as suas extensões e condições.

Podemos considerar que a noção de extremo se deriva, além do padrão de sobrevivência antropocêntrico e da maioria das espécies macroscópicas conhecidas, também da noção de esterilização empregada para promover a inviabilidade celular de microrganismos indesejados. Quando se busca meios para a preservação de alimentos ou limpeza, utilizam-se medidas como: aumento ou diminuição drástica de temperatura, adição de elevados níveis de sal, ou até mesmo a utilização de alguns tipos de ácidos. Por isso, imaginar e descobrir organismos vivendo em ambientes vulcânicos, lagos hipersalinos, desertos, geleiras, sedimentos oceânicos a elevadas profundidades, ambientes com extremos de pH, fendas e fumarolas hidrotermais e ambientes com elevados níveis de radiação foi uma descoberta que provocou revoluções importantes em princípios fundamentais que se baseia a vida na Terra (CAVICCHIOLI, 2006; MERINO et al., 2019; RAMPELOTTO, 2013).

Uma das grandes questões humanas que se mescla com essas descobertas é a possibilidade de vida fora da Terra. Olhar para o céu e observar os corpos celestes, fez com que diversos pensadores se perguntassem se a vida também poderia ser encontrada em outros pontos extraterrestres (MERINO et al., 2019). Essa indagação trouxe, inicialmente, tanto combustível para a especulação científica e filosófica, como diversos exemplos de repressão do livre pensamento. Contudo, o desejo de entender o nosso lugar no espaço e no tempo nos trouxe a um momento singular da história do pensamento científico e tecnológico.

A astrobiologia se configura em uma empreitada científica necessária para sistematizar e lidar com os elevados níveis de conhecimentos acumulados em todas as áreas envolvidas com o estudo da evolução da vida e dos fenômenos relacionados a ela, e com futuros dados que surgirão ao longo do tempo. A astrobiologia é uma área emergente da ciência, que é notavelmente multidisciplinar, onde astrônomos, biólogos, físicos, geólogos e químicos podem utilizar seus conhecimentos específicos para contribuir para a construção de

uma grande cadeia de conhecimentos, que visa desvendar os intrincados passos da matéria orgânica viva, desde o início até os seus padrões atuais (RODRIGUES, 2016). A mesma herdou o estereótipo da exobiologia, mas foge do contexto de restringir suas pesquisas e buscas a vida extraterrestre inteligente. Suas principais questões se baseiam na amostra terrestre, com isso, suas perguntas giram em torno das questões sobre a emergência dos sistemas biológicos terrestres, sua evolução, distribuição e seu futuro espacialmente e temporalmente (RODRIGUES, 2016).

Com a descoberta dos microrganismos extremófilos, a astrobiologia ganhou um grupo de organismos que possibilitou aproximar, ainda mais, os estudos sobre as condições de outros corpos celestes com os estudos sobre a diversidade bioquímica da vida. Os estudos sobre as adaptações desses organismos passaram a ser de suma importância para o entendimento sobre a evolução da vida em si e seus limites.

Atualmente, já são bem conhecidas as principais biomoléculas que compõem a vida: os ácidos nucléicos, carboidratos, lipídios e as proteínas. Entre essas 4 famílias de macromoléculas, as proteínas são os blocos construtores celulares mais abundantes. Estão relacionadas com processos fisiológicos e estruturais, sendo responsáveis por executar praticamente todas as funções celulares (LESK, 2001). As proteínas que são enzimas, fornecem complexas superfícies moleculares no interior das células para a catálise de diversas reações químicas. Proteínas transmembranares formam canais e bombas que controlam o tráfego de pequenas moléculas para dentro e para fora da célula. Outras estão relacionadas com os processos de comunicação celular, onde sinais químicos são recebidos e processados para gerar respostas em locais específicos, integrando os complexos multicelulares e os subsistemas celulares. Proteínas do citoesqueleto moldam as formas das células, geram resistência e proporcionam o tráfego de organelas e de outras proteínas em várias direções. Proteínas que interagem com o DNA estruturam processos fundamentais para a vida, como os processos de replicação e transcrição; organizam as longas fitas em estruturas compactas; condensam e descondensam partes específicas que gera especificidade para a célula (Alberts, 2017; LESK, 2001). Por conta desses inúmeros papéis nos diversos contextos celulares, é de suma importância compreender profundamente a composição e as múltiplas atividades proteicas, para entendermos de forma abrangente a história evolutiva e o funcionamento dos sistemas biológicos.

Para manter a viabilidade celular, os organismos extremófilos são dotados de várias adaptações moleculares, que variam entre estruturais e fisiológicas. Devido às diferentes funções que exercem, as proteínas recebem intensa atenção quando o assunto é entender as manobras naturais dos organismos vivos. No entanto, as mesmas são apontadas como de difícil estudo em extremófilos, graças à grande variabilidade de adaptações para uma determinada condição, porque não se procura apenas uma única tendência ou característica nas suas moléculas. Essa falta de padrão é derivada dos recursos independentes desenvolvidos por esses organismos, que são personalizados para cada ambiente (REED, 2013).

O processo biológico que transforma a informação genética em informação proteica é chamado de tradução, sendo um dos mais fundamentais e importantes para a viabilidade celular. Trabalhos com os de Farias e Prosdocimi (2019) apontam que a organização inicial de partes ancestrais desse sistema foi o insuflar da vida na matéria orgânica. Nas células modernas, o sistema de tradução é composto por uma orquestrada interação entre alguns tipos de RNAs e várias proteínas. Essa relação promove a tradução de uma sequência nucleotídica em uma sequência de aminoácidos, por meio da aplicação de regras que são conhecidas como código genético (ALBERTS, 2017; LECOMPTE et al., 2002).

Devido ao seu papel fundamental no processo de tradução do código genético, que leva à produção proteica, os ribossomos são estruturas importantes para a viabilidade celular de todos os organismos dos três domínios da vida, sejam eles extremófilos ou não. Devido a isso, mecanismos moleculares que promovem a estabilidade dos seus componentes estruturais (proteínas e RNAs ribossomais) são indispensáveis. Com base nisso, quais são as adaptações das proteínas ribossômicas que promovem a estabilidade necessária para o funcionamento das mesmas nos ambientes com elevado nível de salinidade? Entender a evolução dos componentes ribossômicos de diferentes grupos de organismos extremófilos, pode auxiliar no entendimento de questões ainda mais profundas para as Ciências Biológicas, como a origem do ribossomo e da vida.

2. OBJETIVOS

2.1. OBJETIVO GERAL

- O objetivo geral desta pesquisa se configura em analisar as adaptações presentes nas sequências primárias de proteínas ribossômicas de organismos extremófilos halofílicos dos domínios *Archaea* e *Bacteria* e comparar com padrões de organismos não extremófilos.

2.2. OBJETIVOS ESPECÍFICOS

- Comparar através de testes estatísticos a constituição de aminoácidos das 34 proteínas ribossômicas homólogas de organismos halófilos dos domínios *Archaea* e *Bacteria*, e comparar com grupos controles não extremófilos dos respectivos domínios;
- Calcular o tamanho médio das sequências primárias das 34 proteínas representantes de cada grupo.

3. FUNDAMENTAÇÃO TEÓRICA

3.1. CARACTERIZAÇÃO GERAL DOS ORGANISMOS EXTREMÓFILOS

Após as revoluções provocadas pelo desenvolvimento da microscopia ótica e da teoria celular, a vida foi dividida em seres procariontes e eucariontes. Inicialmente, todos os organismos celulares não portadores de um sistema de endomembranas foram enquadrados no domínio *Bacteria*. Só após as análises moleculares comparativas com base em rRNA 16S, feitas no final do Século XX, é que a vida celular foi dividida em três domínios: *Archaea*, *Bacteria* e *Eukarya* (WOESE et al., 1990). A origem e ancestralidade de cada domínio ainda é discutível (STALEY, 2017), porém, os padrões de complexidade fisiológicos e estruturais dos mesmos são bem conhecidos.

A descoberta de organismos extremófilos trouxe considerações importantes para o entendimento sobre os limites físico-químicos que a organização biológica é capaz de suportar, e ricas opções para o setor industrial. Esse grupo de organismos possui adaptações que lhe proporcionaram a capacidade de viver em ambientes hostis para a maioria da grande massa biológica (FARIAS & BONATO, 2003). São distribuídos em diversos nichos com extremos de temperatura, salinidade, pressão, radiação, com extremos de pH, alguns podem crescer em resíduos tóxicos, solventes orgânicos, metais pesados ou em vários outros habitats que antes eram considerados inóspitos para toda a vida. Extremófilos foram encontrados a profundidades de 6,7 km dentro da crosta terrestre; mais de 10 km no fundo do oceano - a pressões de até 110 MPa (mega pascal); de ácido extremo (pH 0) a condições básicas extremas (pH 12,8); e de saídas hidrotermais a 122 ° C para água do mar congelada, a -20 ° C (DUARTE, 2016; RAMPELOTTO, 2013; REED, 2013).

A classificação dos extremófilos é decorrente das condições em que crescem. Organismos que sobrevivem em ambientes com valores de pH ácidos são conhecidos como acidófilos. No outro extremo da escala de pH, estão os organismos conhecidos como alcalófilos, que possuem adaptações para sobreviverem em ambientes com níveis básicos de pH. Em relação a temperatura, a vida se distribui em um espectro amplo, e nos seus extremos estão os organismos termofílicos/hipertermofílicos que estão relacionados com ambientes com altas temperaturas, e no ponto oposto do termômetro, em ambientes com baixas temperaturas, estão os organismos psicofílicos, que possuem a capacidade de metabolizar nesses ambientes. Ambientes com níveis elevados de pressão, radiação e salinidade são lares

de organismos barofílicos, radioresistentes e halofílicos, respectivamente (MADIGAN et al., 2016; DUARTE, 2016; MERINO et al., 2019; RAMPELOTTO, 2013; REED, 2013).

Essa classificação geral não enquadra uma gama ampla de organismos, conhecidos como poli-extremófilos, que são adaptados a ambientes com pluralidade de condições extremas (MERINO et al., 2019). Alguns exemplos são muitas fontes termais, que em decorrência dos fenômenos geológicos, são ácidas ou alcalinas ao mesmo tempo, e geralmente ricas em elementos metálicos dissolvidos; o oceano profundo, por exemplo, é geralmente frio, oligotrófico (muito baixo teor de nutrientes) e exposto a alta pressão (RAMPELOTTO, 2013).

Esses fascinantes organismos, além de estarem distribuídos por todo o planeta, possuem membros nos três domínios da vida celular: *Archaea*, *Bacteria* e *Eukarya*. Sendo o domínio *Archaea* um grupo destacado, devido ao grande número de organismos extremófilos pertencentes a ele; mas esse grupo também inclui eucariotos como protistas (por exemplo, algas, fungos e protozoários) e organismos multicelulares (RAMPELOTTO, 2013).

3.2. ORGANISMOS HALÓFILOS

É amplamente conhecido o processo milenar de salgar alimentos como carnes, a fim de preservá-los. A base da técnica é simples e se baseia no processo osmótico da água, ou seja, a mesma tem a tendência de sair de um ambiente menos concentrado para um meio mais concentrado. Durante o processo, a retirada da umidade do tecido ocorre paralelamente com a entrada de sal, gerando um ambiente com condições impróprias para o metabolismo da maioria dos microorganismos conhecidos. Entretanto, um grupo específico de organismos consegue sobreviver em diferentes concentrações salinas que em muitos grupos levaria ao colapso funcional metabólico (SIGLIOCCOLO et al., 2008).

Os organismos halofílicos são divididos em halotolerantes, halófilos moderados e extremos. Kushner 1978, definiu halófilos moderados como organismos que crescem de maneira ideal entre 0,5 e 2,5 M de sal, enquanto espécies que crescem na ausência de sal, bem como na presença de concentrações relativamente altas de sal são designadas halotolerantes (ou extremamente halotolerantes, se o crescimento ultrapassar 2,5 M). As espécies halofílicas moderadas apresentam adaptações que permitem que as mesmas cresçam em amplas faixas de concentrações de sal, que variam desde as concentrações apresentadas

pela água doce, até o ponto de saturação do sal, sendo essa capacidade complexa vista como surpreendente, pois demanda um arcabouço de estratégias alto (VENTOSA, 1998). Em relação ao terceiro grupo, para ser considerado halófilo extremo as espécies devem requerer, no mínimo, 1,5 M (cerca de 9%) de NaCl para seu crescimento, sendo a maioria dependente de 2 a 4 M de NaCl (12-23%) para o crescimento ótimo (SIGLIOCCOLO et al. 2011). E de uma forma quase geral, os halófilos extremos são capazes de crescer em 5,5 M de NaCl (32%, que é o limite de saturação para o NaCl), porém, é mostrado que algumas espécies apenas cresçam lentamente nesta salinidade (MADIGAN et al., 2016). Vale destacar, que os padrões de tolerância e exigência de sal de muitas espécies está relacionado com outras variáveis importantes para o crescimento, como temperatura e composição do meio. A temperatura de crescimento deve ser especificada, especialmente para a definição da faixa de sal mais baixa que permite o crescimento (VENTOSA, 1998).

Existe uma famosa frase entre os microbiólogos especialistas em halófilos: "Tudo está em toda parte: mas o ambiente seleciona". A mesma é uma citação derivada do livro de Lourens Baas Becking, de 1934. Ela é atribuída como base para a compreensão da distribuição de microrganismos halofílicos em todo o mundo (MA et al., 2010). Os estudos científicos sobre microrganismos sobrevivendo em altas concentrações salinas é antigo no contexto da microbiologia de extremófilos. Trabalhos como o de Baxter (2014) cita trabalhos a partir de 1838, e Zheng (1997) referencia exemplos datados desde 1913, sendo a década de 1970 a passagem histórica onde esse grupo de microrganismos é colocado como significativo em teorias biológicas fundamentais e perspectivas importantes na aplicação prática (ZHENG, 1997). Assim como sua ampla distribuição no mundo, os organismos halofílicos possuem exemplos nos três domínios da vida celular, como mostrado na Figura 1, porém, os domínios procariontes apresentam uma vasta diversidade de espécies, quando comparado com o domínio eucarioto.

O domínio *Bacteria* possui inúmeros representantes halofílicos classificados nos seus subgrupos, principalmente halófilos moderados, a maioria pertencendo a Halomonadaceae, uma família de Proteobacteria (classe Gammaproteobacteria). Essas espécies bacterianas são aeróbicas, anaeróbicas, quimio-heterotróficas, foto-heterotróficas e / ou fotoautótrofas. Eles são onipresentemente encontrados em lagos e áreas costeiras em todo o mundo (EDBEIB et al., 2016). Em relação ao domínio *Archaea*, o Filo de destaque quando se refere a halófilos é o Euryarchaeota, que contempla um grupo grande e filogeneticamente diverso de arqueias.

Incluindo arqueias metanogênicas, muitos gêneros de arqueias extremamente halofílicas, e organismos hipertermófilos; sendo os principais gêneros de arqueias extremamente halofílicas são: *Halobacterium*, *Haloferax*, *Natronobacterium* (MADIGAN et al., 2016; EDBEIB et al., 2016).

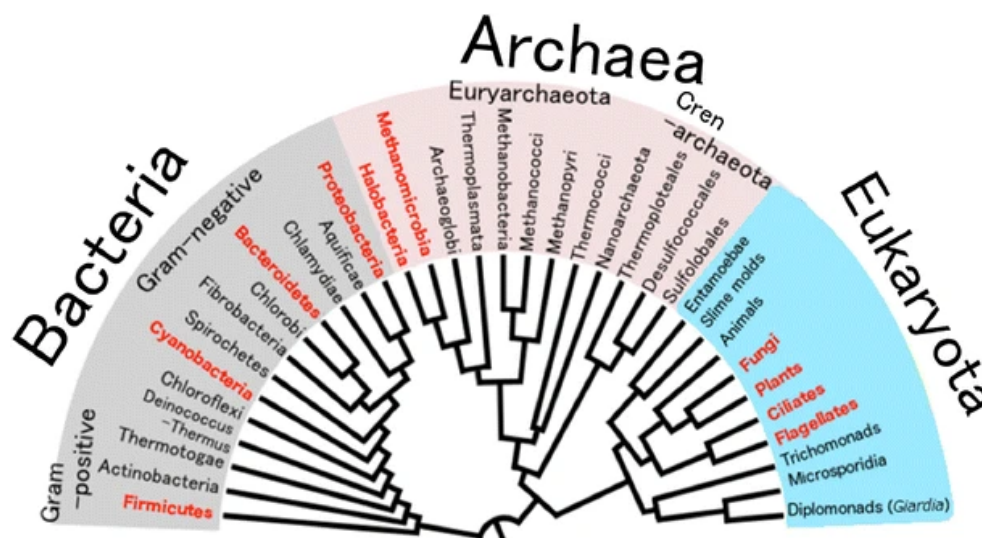


Figura 1: Distribuição dos organismos halófilos nos três domínios da vida celular. Em vermelho estão destacados os grupos que possui no mínimo uma espécie halofílica: (*Archaea*: Classes Methanomicrobia e Halobacteria); (*Bacteria*: Filos Proteobacteria, Bacteroidetes, Cyanobacteria e Firmicutes); (*Eukaryota*: Reinos Fungi, Plants, Ciliates e Flagellates). Fonte: EDBEIB et al., 2016

3.3. AMBIENTES SALINOS

As primeiras imagens do nosso planeta tiradas do espaço revelaram, materialmente, uma perspectiva que muitos navegantes marinhos já sabiam: "o planeta Terra é azul". Isso é derivado das vastas extensões oceânicas que dominam a superfície terrestre. Esse tipo de condição planetária, estruturou inúmeros ambientes com diferentes condições de salinidade, sendo os ambientes hipersalinos comuns em todo planeta e os extremamente hipersalinos mais raros. Atualmente, devido a atividades humanas, os ambientes salinos são separados em dois grupos, os naturalmente salinos, como salinas solares e lagos salinos, e ambientes artificialmente salinos, como a superfície de alimentos altamente salgados (MADIGAN et al., 2016). Na imagem abaixo, é mostrado um mapa global com os índices mais atualizados sobre as taxas de salinidade dos oceanos terrestres (Figura 2). Nele é possível observar regiões destacadas em vermelho, onde a cor representa os mais elevados índices de salinidade. Localizada em um desses pontos, está uma região marinha conhecida como Mar

Morto em Israel, com uma concentração salina cerca de 10 vezes maior que a de qualquer oceano. Mesmo possuindo um nome relacionado a falta de exemplares da estruturação orgânica viva, o Mar Morto transborda de organismos de vários grupos, como fungos, protozoários, arqueas, bactérias e vírus (OREN, 1999; WEI, 2015). Na Figura 3 é representada uma das características desse ambiente, a formação de ilhas salinas derivadas do processo de evaporação.

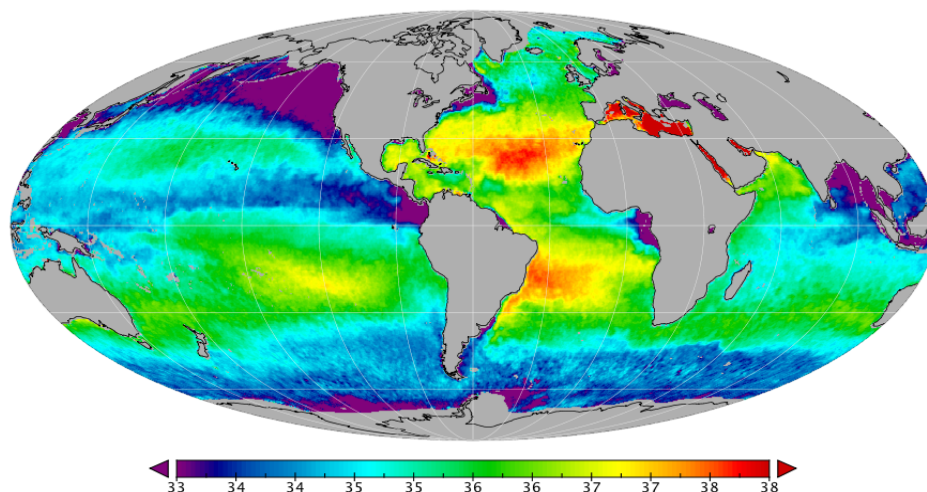


Figura 2: Mapa global dos níveis de salinidade dos oceanos terrestres. Fonte: *NASA Salinity*, 2020. Disponível em: <https://salinity.oceansciences.org/smap-salinity.htm>.



Figura 3: Ilhas de sal no Mar Morto, Israel. Fonte: <https://twistedifter.com/2012/06/10-things-you-didnt-know-about-the-dead-sea/>

Um outro exemplo de ambiente altamente salino é o Grande Lago Salgado (*Great Salt Lake*) de Utah, que apresenta uma beleza visual derivada de uma paleta de cores que variam espacialmente e sazonalmente de rosa a roxo e laranja no braço norte e de azul a verde no

braço sul. Essa amostra natural é derivada das características geológicas do lago e dos microorganismos presentes no mesmo, que possuem carotenóides ou clorofila. O braço norte do Grande Lago Salgado possui uma cor rosa devido, em grande parte, às arqueias halofílicas que habitam a água hipersalina, enquanto o braço sul é de cor verde devido à rica diversidade de espécies de algas (Figura 4) (BAXTER, 2018).

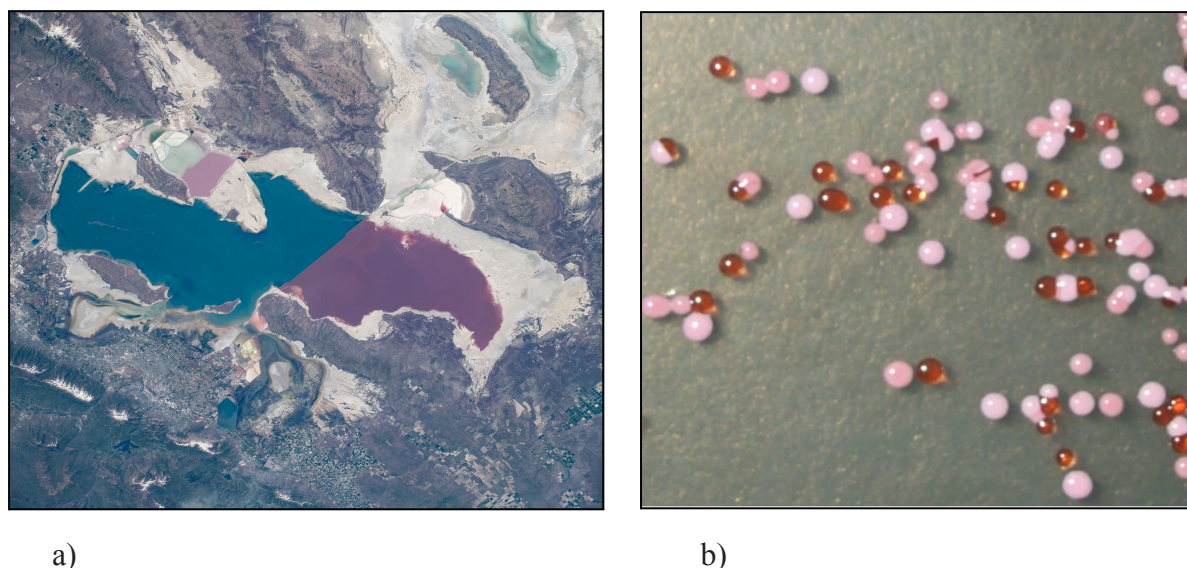


Figura 4: a) O Grande Lago Salgado (*Great Salt Lake*) de Utah. Foto tirada a partir da Estação Espacial Internacional orbitando 255 milhas acima do sudoeste dos Estados Unidos. Fonte: NASA, 2019. b) Colônias de arqueias halofílicas de *Great Salt Lake* que crescem em meios sólidos. Fonte: BAXTER, 2018.

3.4. PROTEÍNAS: AS MOLÉCULAS MULTIFUNCIONAIS DA VIDA

"No drama da vida em escala molecular, as proteínas estão onde a ação está" (LESK, 2001). Seus intrincados e variados padrões proporcionam a essas macromoléculas a capacidade de exercer grande parte das funções celulares. Grande parte da configuração molecular que permite com que diferentes espécies sobrevivam em ambientes com condições como as citadas acima, estão presentes nas suas proteínas. Por isso, é necessário entendê-las com profundidade. Uma consideração interessante é que todas as tarefas que elas podem executar são baseadas em um princípio comum, os vinte aminoácidos que as formam. Por essa razão o estudo de proteínas, sua composição, estrutura, dinâmica e função é tão importante. Precisamos entender como essas moléculas se dobram, como se agrupam em complexos e como funcionam se quisermos responder a perguntas cruciais para o entendimento da vida em ambientes extremos e não extremos; por que a vida como a

conhecemos evoluiu desta maneira e neste planeta e não em nenhum outro lugar, pelo menos por enquanto (BREDA et al., 2008). E para isso, precisamos entender primariamente os padrões específicos de composição dos aminoácidos.

Atualmente, o estudo das proteínas é centralizado em uma área multidisciplinar, a bioinformática, a partir dela, os conhecimentos específicos das ciências biológicas, física, química, matemática e da informática contribuem nesta empreitada, que permite um novo nível de conhecimento sobre a organização da vida (BREDA et al., 2008).

Essas moléculas são formadas por 20 blocos construtores, que de forma geral, esses monômeros possuem uma estrutura padrão comum: todos possuem um grupo ácido carboxílico (COOH), um grupo amino (NH₃), um hidrogênio (H) e uma cadeia lateral (R) que difere de um aminoácido para outro e varia em termos de tamanho, polaridade, hidrofobicidade, carga, forma, volume etc., todos ligados a um único átomo de carbono denominado carbono-alfa (C) (BREDA et al., 2008; ALBERTS, 2017). Com base nas características de suas cadeias laterais, os aminoácidos podem ser divididos em quatro grupos: apolares ou hidrofóbicos, polar não-carregado, carregado positivamente e carregado negativamente (ALBERTS, 2017; LESK, 2008). Na Tabela 5 abaixo, destacados por cores estão os aminoácidos de cada grupo.

Ácido aspártico	Asp	D	Negativa
Ácido glutâmico	Glu	E	Negativa
Arginina	Arg	R	Positiva
Lisina	Lys	K	Positiva
Histidina	His	H	Positiva
Asparagina	Asn	N	Polar não carregada
Glutamina	Gln	Q	Polar não carregada
Serina	Ser	S	Polar não carregada
Treonina	Thr	T	Polar não carregada
Tirosina	Tyr	Y	Polar não carregada

Alanina	Ala	A	Apolar
Glicina	Gly	G	Apolar
Valina	Val	V	Apolar
Leucina	Leu	L	Apolar
Isoleucina	Ile	I	Apolar
Prolina	Pro	P	Apolar
Fenilalanina	Phe	F	Apolar
Metionina	Met	M	Apolar
Triptofano	Trp	W	Apolar
Cisteína	Cys	C	Apolar

Figura 5: Os 20 aminoácidos separados com base nas propriedades das suas cadeias laterais. Onde cada uma das cores separa um grupo específico. Fonte: ALBERTS, 2017.

As proteínas, de uma forma simplificada, apresentam quatro níveis estruturais: estrutura primária, secundária, terciária e quaternária. A primária é a mais simples de todas e não é a forma funcional da proteína, ela é derivada de ligações conhecidas como ligações peptídicas entre diferentes combinações de aminoácidos (Figuras 6 e 7), que são ligados

entre si através da combinação da região C-terminal de um aminoácido com a região N-terminal de outro (ALBERTS, 2017; BREDA et al., 2008; LESK, 2008).

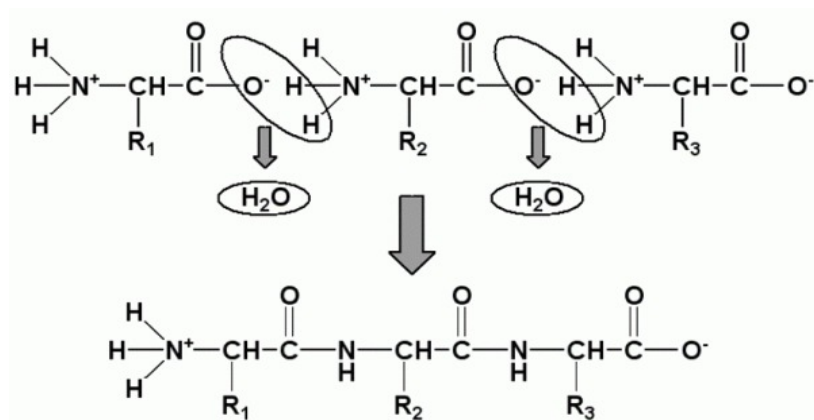


Figura 6: Representação de ligações peptídicas que geram uma cadeia de proteína através de uma série de reações de condensação, normalmente ocorrendo *in vivo* no ribossomo durante a síntese de proteínas. Fonte: BREDA et al., 2008.

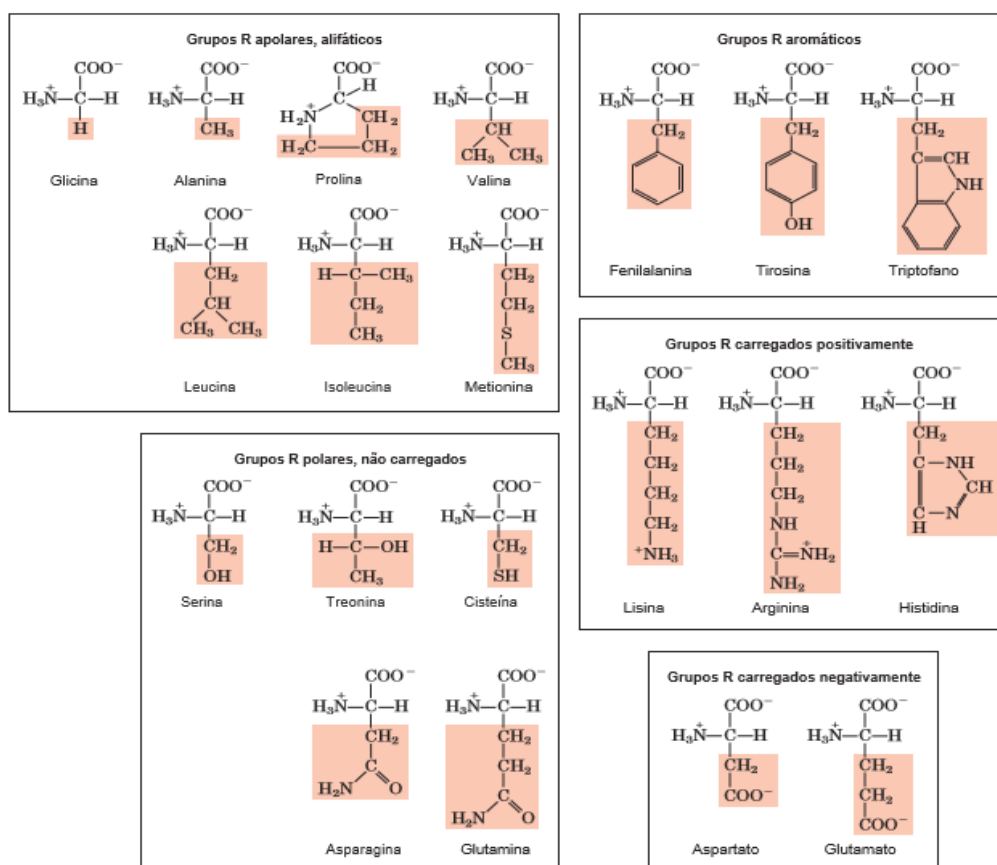


Figura 7: Os vinte aminoácidos que compõem as proteínas com suas cadeias laterais coloridas para o destaque do seu perfil químico. Fonte: NELSON et al., 2014.

Mesmo representando a forma mais básica de uma proteína, a sequência primária, na falta de estruturas tridimensionais estabelecidas, podem fornecer informações valiosas, pois carrega a informação mais básica sobre uma proteína, sua constituição de aminoácidos, que fornecem a base para o entendimento mais abrangente sobre a estrutura funcional proteica e como ela pode estar relacionada no contexto dos subsistemas celulares (LESK, 2008).

3.5. AS LIGAÇÕES QUE FAZEM A VIDA FUNCIONAR: O PROCESSO DE TRADUÇÃO DA INFORMAÇÃO GENÉTICA

O processo de tradução da informação genética é um dos mais fundamentais para todas as formas de vida; sendo caracterizado como a conversão da informação contida em uma molécula de RNA mensageiro (mRNA) em uma proteína, com base em regras conhecidas como código genético. Esse processo complexo e conservado em todos os organismos celulares é executado e regulado através de ações conjuntas de diversas moléculas de proteínas e de ácidos ribonucleicos (RNAs) (FREDRICK et al., 2009). De uma forma geral, as peças que compõem esse processo são: RNAs mensageiros (mRNAs); moléculas de RNAs transportadores (tRNAs); proteínas conhecidas como aminoacil-tRNA-sintetases; e uma máquina macromolecular, conhecida como ribossomo (ALBERTS et al., 2010).

Os mRNAs são derivados do processo de transcrição do DNA (ácido desoxirribonucleico) em RNA, e contém a mensagem codificada na forma de códons, que são trincas de nucleotídeos que correspondem aos aminoácidos (Tabela 1). Porém, a interação entre os códons com seus aminoácidos correspondentes é feito através de moléculas adaptadoras, chamadas de RNAs transportadores (tRNAs), também derivadas do processo de transcrição. Um outro elemento de suma importância para o processo de tradução são as proteínas conhecidas aminoacil-tRNA-sintetases. O papel biológico conhecido desse grupo de proteínas é a capacidade de realizar o reconhecimento e a ligação de aminoácidos específicos às suas moléculas de tRNAs apropriadas (ALBERTS et al., 2010; FREDRICK et al., 2009).

AMINOÁCIDOS	CÓDONS
Alanina (A)	GCA GCC GCG GCU
Cisteína (C)	UGC UGU
Ácido Aspártico (D)	GAC GAU
Ácido Glutâmico (E)	GAA GAG
Fenilalanina (F)	UUC UUU
Glicina (G)	GGA GGC GGG GGU
Histidina (H)	CAC CAU
Isoleucina (I)	AUA AUC AUU
Lisina (K)	AAA AAG
Leucina (L)	UUA UUG CUA CUC CUG CUU
Metionina (M)	AUG
Asparagina (N)	AAC AAU
Prolina (P)	CCA CCC CCG CCU
Glutamina (Q)	CAA CAG
Arginina (R)	AGA AGG CGA CGC CGG CGU
Serina (S)	AGC AGU UCA UCC UCG UCU
Treonina (T)	ACA ACC ACG ACU
Valina (V)	GUA GUC GUG GUU
Triptofano (W)	UGG
Tirosina (Y)	UAC UAU

Tabela 1: Os vinte aminoácidos e seus respectivos códons (código genético padrão). Fonte: ALBERTS et al., 2010.

A centralidade do processo de tradução está relacionada com uma máquina macromolecular sofisticada, conhecida como ribossomo, responsável pela integração de todos os papéis descritos acima. É nos ribossomos onde a gênese das proteínas acontece. Sua estrutura intrincada permite o reconhecimento dos mRNAs, a interação dos seus códons com os anticódons dos tRNAs e a união dos aminoácidos em uma cadeia contínua. Durante o processo de tradução, quando o ribossomo está na sua forma ativa, o mesmo possui duas subunidades: a subunidade ribossômica pequena (SSU, **do inglês** *small subunit*) e a subunidade ribossômica maior (LSU, **do inglês** *large subunit*) (ALBERTS, 2017; MALLIK, 2013; SMITH et al., 2008). A subunidade pequena é estruturada de tal forma, que fornece de forma eficiente uma região sobre a qual os tRNAs podem ser pareados sobre os códons do mRNA, enquanto a subunidade maior realiza as catálises das ligações peptídicas que unem os aminoácidos, formando uma cadeia proteica crescente (FREDRICK et al., 2009). Mesmo com várias etapas ocorrendo simultaneamente, a adição dos monômeros a cadeia polipeptídica ocorre de maneira rápida, sendo a velocidade diferencial quando se analisa o funcionamento dos ribossomos eucarióticos, que adicionam aproximadamente dois aminoácidos a cada segundo; enquanto os ribossomos bacterianos executam ainda mais rapidamente, a taxas cerca de 20 aminoácidos por segundo (ALBERTS, 2017).

A ação do ribossomo para proporcionar o advento da tradução é baseado na ação de 4 sítios específicos: o primeiro está localizado na subunidade menor (SSU), conhecido como sítio de ligação ao mRNA; os outros 3 (denominados sítio A, sítio P e sítio E) são para tRNAs, localizados na subunidade maior (LSU), (Figura 8) (ALBERTS, 2017; FREDRICK et al., 2009).

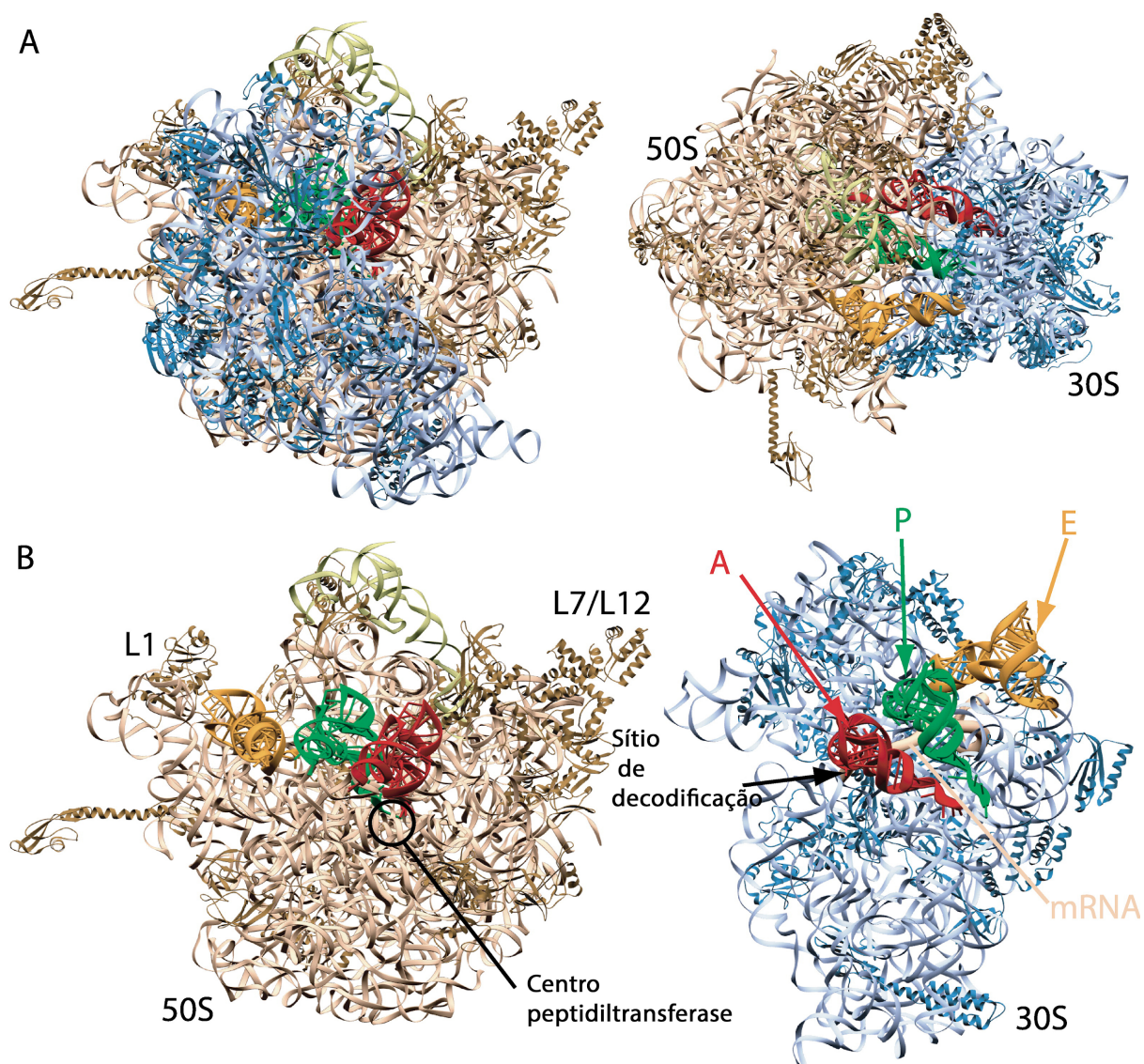


Figura 8: Representação esquemática do ribossomo de *Thermus thermophilus* 70S a uma resolução de 5,5 Å, com seus domínios destacados (A, P e E). A) Visualização do ribossomo com ambas as subunidades. B) Imagens das subunidades ribossomais separadas, na esquerda está a subunidade maior (LSU) e na direita a subunidade menor (SSU). Fonte: RAMAKRISHNAN, 2002.

No citoplasma, a subunidade menor interage inicialmente com o mRNA recém sintetizado. Com essa interação, a subunidade maior completa o complexo, permitindo com que os tRNAs interajam com os seus sítios. Os mesmos se ligam com alta afinidade aos sítios A e P apenas se seus anticódons formarem pares de bases com o códon complementar na molécula de mRNA, que está ligada ao ribossomo. Os sítios A e P estão suficientemente

próximos para que suas duas moléculas de tRNA sejam forçadas a formarem pares de bases com códons adjacentes da molécula de mRNA (ALBERTS, 2017; FREDRICK et al., 2009). Essa característica do ribossomo mantém a fase de leitura correta no mRNA (ALBERTS, 2017).

Quatro etapas definem o processo: na etapa 1, um tRNA ligado a um aminoácido forma pares de bases com o códon do mRNA posicionado na subunidade menor. Na etapa 2, a extremidade carboxila da cadeia polipeptídica é liberada do tRNA no sítio P (pelo rompimento da ligação de alta energia entre o tRNA e seu aminoácido) e é ligada ao grupo amino livre do aminoácido ligado ao tRNA no sítio A, formando uma nova ligação peptídica. Essa reação central da síntese de proteínas é catalisada por uma peptidiltransferase contida na subunidade ribossomal maior. Na etapa 3, a subunidade maior se move em relação ao mRNA que está ligado à subunidade menor, o que interfere nas hastesceptoras dos dois tRNAs que se encontram nos sítios E e P da subunidade maior. Na etapa 4, uma nova série de alterações conformacionais move a subunidade menor e o mRNA a ela ligado exatamente três nucleotídeos, ejetando o tRNA ligado ao sítio E e reiniciando o ribossomo para que ele esteja pronto para receber o próximo RNA transportador. A etapa 1 é, então, repetida com a chegada de um novo aminoacil-tRNA, e assim por diante. Esse ciclo de quatro etapas é repetido cada vez que um aminoácido é adicionado à cadeia polipeptídica, e a cadeia cresce a partir de sua extremidade amino em direção à extremidade carboxila (ALBERTS, 2017; FREDRICK et al., 2009; RAMAKRISHNAN, 2002).

3.6. CONSTITUIÇÃO DO RIBOSSOMO

O complexo ribonucleoproteico conhecido como ribossomo é uma estrutura molecular altamente complexa, estruturada com base em dezenas de pequenas proteínas (50-80), normalmente com 50 a 300 aminoácidos e (3 ou 4) grandes moléculas de RNAs ribossômicos (rRNA) com 120 a 4.500 nucleotídeos (Figura 9) (RAMAKRISHNAN, 2002; WILSON et al., 2008). Esses dois elementos são essenciais para as funções ribossomais descritas acima. Por conta da característica multifuncional, na década de 1970, as proteínas ribossômicas eram vistas como as gerenciadoras principais na execução das atividades do ribossomo, enquanto os RNAs ribossômicos (rRNA) eram vistos como moléculas puramente estruturais, que promoviam as posições corretas das proteínas para o bom funcionamento do

ribossomo. Todo esse pensamento se transformou a partir da década seguinte, com base nas descobertas das atividades catalíticas dos RNAs, juntamente com os resultados mostrando um envolvimento direto de rRNAs em funções essenciais para a tradução, por exemplo, as interações *Shine–Dalgarno* entre o RNA mensageiro (mRNA) e rRNA durante o início da síntese proteica bacteriana, bem como a predominância de rRNA nos centros de decodificação e peptidil-transferase (PTF) (NIKOLAY et al., 2015; WILSON et al., 2008).

Embora o ácido ribonucleico ribossômico (rRNA) tenha a principal reivindicação de certos aspectos da função do ribossomo, como decodificação e atividade peptidiltransferase, também existem processos importantes dominados por proteínas no ribossomo, como por exemplo o poro de entrada do RNA mensageiro (mRNA), o local de ligação ao fator de tradução e saída do túnel ribossômico. Atualmente, o caminho do meio, é colocado como o mais correto, pois, após estudos mais aprofundados sobre as estruturas tridimensionais dos ribossomos, constatou-se que para que as funções ribossômicas fossem exercidas, havia tanto a atuação efetiva de proteínas, como dos rRNAs (NIKOLAY et al., 2015; WILSON et al., 2008).

Devido a divergência evolutiva e os processos evolutivos posteriores, os ribossomos dos três domínios da vida apresentam significativas diferenças, o tamanho dos rRNAs e número de proteínas se diferenciam. O ribossomo bacteriano possui 58 proteínas (24/34) em comparação com o do domínio *Archaea*, que possui 68 proteínas (28/40) e o do domínio *Eukarya*, que possui 78 proteínas (32/46). Existem 34 proteínas ribossômicas (15/19) identificadas como homólogas aos três domínios celulares: uL1, uL2, uL3, uL4, uL5, uL6, uL10e, uL11, uL12, uL13, uL14, uL15, uL16, uL18, uL22, uL23p, uL24, uL29 e uL30 que compõem a subunidade maior (LSU) e uS2, uS3, uS4, uS5, S7, uS8, uS9, uS10, uS11, uS12, uS13, uS14, uS15, uS17, uS19 (Figura 10) (LECOMPTE et al., 2002; SMITH et al., 2008).

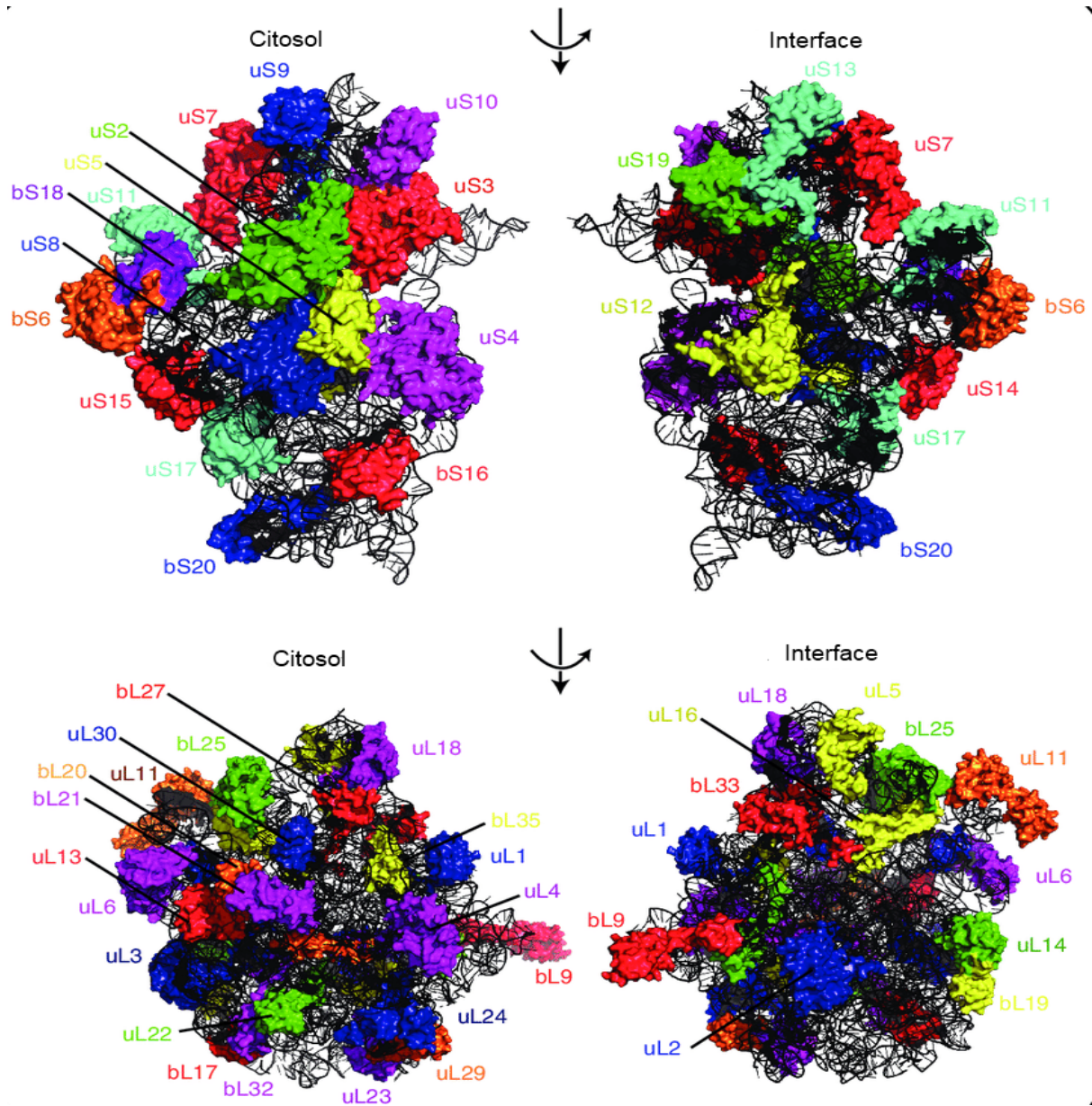


Figura 9: Esquema da estrutura ribonucleoproteína do ribossomo da espécie *Termus. thermophilus*. A subunidade menor (SSU) está sendo mostrada acima e a subunidade maior (LSU) está esquematizada abaixo. As proteínas estão coloridas, com seus respectivos nomes, enquanto o rRNA é representado em preto. Fonte: NIKOLAY et al., 2015.

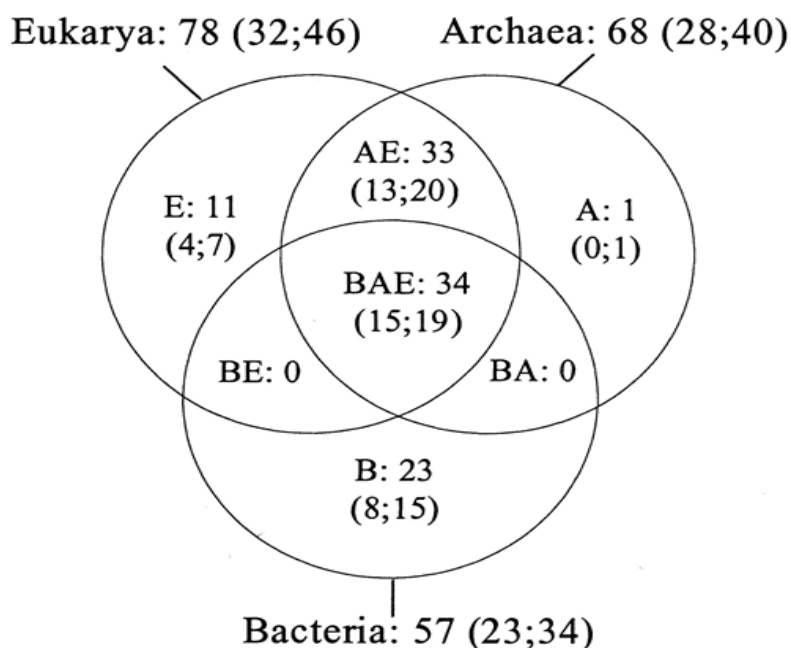


Figura 10: Esquema das intersecções do conjunto das proteínas ribossômicas entre os três domínios da vida celular. No centro está representado o conjunto das 34 proteínas homólogas aos três domínios da vida (BAE). Fonte: LECOMPTE et al., 2002.

Devido à enorme cooperatividade dos elementos do ribossomo, a atribuição funcional específica às proteínas do mesmo é uma tarefa delicada (NIKOLAY et al., 2015). De uma forma geral, as proteínas ribossômicas são muito básicas (ponto Isoelétrico-pI ~10,1 médio em comparação com pI = 4 a 5 para a maioria dos fatores de tradução), sugerindo que uma função geral das proteínas ribossômicas seja a neutralização das cargas negativas dos resíduos de fosfato no rRNA (LOTT et al., 2013; WILSON et al., 2008). Algumas exceções são S1 e S6 na subunidade pequena e as proteínas L7/L12 na subunidade grande. A proteína L7/L12 interage diretamente com a proteína ribossomal L10, e não com o rRNA; e as proteínas S6 faz contato extenso com a S18, e a S1 interage com a S21, S11 e S18 (WILSON et al., 2008). Essa colocação está relacionada com o princípio básico da eletrostática, o princípio da atração e repulsão, que fundamenta que cargas com o mesmo sinal elétrico se repelem, e com sinais contrários se atraem (SILVA, 2011). Assim como grande parte da matéria conhecida, a massa viva é formada por átomos, constituídos de maneira geral por partículas neutras, positivas e negativas. Com isso, muitas moléculas biológicas são dotadas de cargas que são fundamentais na estruturação da vida como a conhecemos. A estrutura dos ácidos nucleicos DNA e RNA, é baseada em cinco bases nitrogenadas (timina, citosina,

uracila, adenina e guanina) e em um esqueleto de açúcar-fosfato. Esse esqueleto rico em átomos eletricamente carregados dão aos ácidos nucleicos uma carga final negativa (NELSON et al., 2014)).

Um dos exemplos biológicos mais expressivos dessa relação de carga e interação é a formação da cromatina e dos cromossomos. Nas células, o DNA se encontra organizado na forma de cromatina. Essa estruturação é de extrema importância, em vista do longo tamanho das moléculas de DNA comparadas com as dimensões do núcleo das células eucarióticas e as das células procarióticas. O nível mais baixo dessa condensação é chamado de nucleossomo, onde o princípio desse enovelamento está relacionado com as cargas das moléculas proteicas associadas e a do DNA. As proteínas histonas, protagonistas do processo, possuem uma estrutura complexa com propriedades que permitem uma interação íntima com o DNA. Na formação de cada nucleossomo nas células eucarióticas, 142 ligações de hidrogênios são formadas entre o DNA e o cerne de histonas. Quase metade dessas ligações formam-se entre os aminoácidos da estrutura das histonas e o esqueleto de fosfodiéster do DNA. Além disso, interações hidrofóbicas e pontes salinas também mantêm o DNA ligado às proteínas no nucleossomo. E um ponto importante diz respeito a constituição de aminoácidos dos cernes das histonas, onde mais de um quinto são lisina ou arginina, aminoácidos com cadeias laterais básicas, onde suas cargas positivas neutralizam a carga negativa do esqueleto fosfodiéster do DNA (ALBERTS, 2017). Com base nesse exemplo, pode-se observar como as cargas das moléculas atuam na promoção de interações dos complexos proteína-ácido nucleico.

3.7. TRABALHOS RELACIONADOS COM O ESTUDO DAS PROTEÍNAS DE ORGANISMOS HALÓFILOS

Devido a diversidade de grupos existentes, foi observado que diferentes halófilos enfrentavam a questão da alta salinidade de formas diferenciadas, de modo que o estudo da vida microbiana em altas concentrações de sal pode responder a muitas perguntas básicas sobre a adaptação de microrganismos aos seus ambientes (MA et al., 2010).

Todos os microrganismos halofílicos compartilham uma propriedade básica: seu citoplasma deve ser pelo menos isosmótico com o meio circundante (MA et al., 2010; SAUM

et al., 2008). De uma forma geral, é relatado basicamente duas estratégias que permitem que os microrganismos extremófilos halotolerantes vivam em altas concentrações de sal: a estratégia de “alto teor de sal” no ambiente citosólico, principalmente potássio e cloreto e a estratégia com “baixo teor de sal”. No primeiro caso, é relatado que as proteínas são dotadas de características marcantes nas suas estruturas para a configuração de moléculas estáveis e ativas na presença de concentrações molares de sal. A segunda estratégia tem como característica um padrão citosólico salino muito próximo ao encontrado em organismos que não estão relacionadas com ambientes com níveis elevados de salinidade. O ponto central dessa estratégia é estruturado a partir de vias específicas de biossíntese e/ou acumulação de solutos orgânicos que não interferem significativamente na atividade das enzimas normais. Essas duas estratégias promovem ambientes citosólicos diferenciados, e com isso, pressões diferentes para as proteínas intracelulares, porém as proteínas de membrana plasmática que possuem domínios extracelulares e proteínas estruturais localizadas na parede celular, mesmo nos organismos com a adaptação com baixo teor de sal no citoplasma, possuem o perfil da primeira estratégia (FEDYUKINA et al., 2014; MA et al., 2010; KRAEGELOH, 2002; SAUM et al., 2008).

A presença de um soluto como o NaCl em ambientes aquáticos, leva a modificações em características básicas da água, como pontos de congelamento e ebulição, assim como nas pressões de vapor e osmótica, sendo essas alterações desencadeadas a partir da diminuição do potencial químico da água (CORREIA, 2014; KRAEGELOH, 2002). Nos movimentos a curta distância, como nos sistemas osmóticos, a água desloca-se de regiões de maior potencial hídrico (maior energia livre) para regiões de potencial hídrico mais baixo (menor energia livre) (CORREIA, 2014). A diminuição do potencial químico da água depende em grande parte da mudança na entropia da mesma. Quando o sal está presente, ele tende a provocar interferências na estrutura ordenada da água, aumentando assim a entropia, ou seja, deixando-a positiva (KRAEGELOH, 2002).

Em organismos que não são adaptados a esse tipo de ambiente, pode-se entender o que a diferença entre o potencial da água citoplasmática e de um ambiente rico em sal pode causar no sistema celular. Comparada com a água com sal dissolvido, a água citoplasmática apresenta um potencial químico maior, com isso, a água citoplasmática tenderá a ir em direção ao ambiente extracelular circundante, provocando dano drásticos nas células, como redução do volume celular, que leva a inviabilidade de crescimento devido ao aglomeramento

molecular interno, portanto, levando a taxas de difusão reduzidas de proteínas e metabólitos (KRAEGELOH, 2002).

O meio fisiológico dos ribossomos dos grupos não halofílicos possui concentrações de sal moderado, com cerca de 150 mM, enquanto o ambiente dos ribossomos de organismos halofílicos é caracterizado por concentrações muito maiores de sal, de 2-5 M (FEDYUKINA *et al.*, 2013). Isso vale principalmente para o domínio *Archaea*, onde são referenciados maiores números de espécies halofílicas extremas (VENTOSA *et al.*, 1998).

Devido ao alto teor de sal no ambiente fisiológico, sabe-se que o ponto isoelétrico (pI) das proteínas dos halófilos é geralmente mais baixo do que o pI das proteínas dos organismos não-halofílico. Portanto, o proteoma dos halófilos é dominado por biomoléculas com uma carga negativa líquida a pH fisiológico. Essa propriedade se manifesta na alta abundância de ácido aspártico (Asp) e/ou ácido glutâmico (Glu) nas proteínas de organismos halofílicos. Esses resíduos são mais eficazes na captação de água de hidratação em solução. Dessa forma, altas quantidades de Asp e Glu permitem que proteínas de halófilos concorram com sucesso com o alto teor de sal em massa para a hidratação da proteína (FEDYUKINA *et al.*, 2013).

Outras adaptações incluem baixo teor de lisina e aumento no número de aminoácidos hidrofóbicos com cadeias laterais mais curtas. Diferentes funções foram sugeridas para o papel dos resíduos ácidos. Eles são, provavelmente, importantes para a ligação de moléculas de água essenciais, mas também podem desempenhar um papel crítico na ligação de sal ou na prevenção da agregação. As outras alterações na composição de aminoácidos nas proteínas são provavelmente importantes para a redução da hidrofobicidade, a fim de impedir a agregação e manter a flexibilidade em altas concentrações de sal. Também são encontradas características estruturais que podem ser importantes para a adaptação halofílica, como um aumento no número de pontes de salinas. Isso, no entanto, não é usado por todas as proteínas, indicando que diferentes estratégias evoluíram para lidar com condições muito salinas (BOLHUIS, 2008).

4. METODOLOGIA

A pesquisa realizada encontra-se dentro do contexto das pesquisas explicativas, que visam identificar os elementos que estão relacionados com a determinação ou que contribuem para que um fenômeno ocorra com características específicas (OLIVEIRA, 2011), sendo os dados coletados, classificados dentro do misto qualitativos/quantitativos.

4.1. DADOS DOS ORGANISMOS

Foram selecionadas 885 espécies de organismos extremófilos halofílicos pertencentes aos dois domínios da vida procariótica: *Archaea* (260) e *Bacteria* (625), do bancos de dados *HaloDom* que é restrito a informações sobre halófilos, disponível em: <http://halodom.bio.auth.gr/?view=home>), em 10 de fevereiro de 2020; sendo as espécies classificadas como halófilos moderados ou extremos. Em relação ao grupo controle, foram selecionados 190 espécies de organismos não-extremófilos, sendo 160 do domínio *Bacteria* e 30 do domínio *Archaea*. O menor número de espécies arqueais do grupo controle foi devido a dificuldade em encontrar exemplares não-extremófilos na literatura e em bancos de dados.

4.2. PROTEÍNAS ANALISADAS NO ESTUDO

As proteínas analisadas são as 34 proteínas ribossomais consideradas homólogas entre os três domínios da vida, sendo 19 proteínas (**uL1, uL2, uL3, uL4, L5, uL6, uL10e, uL11, uL12, uL13, uL14, uL15, uL16, uL18, uL22, uL23p, uL24, uL29 e uL30**) que compõem a subunidade maior (LSU) e 15 (**uS2, uS3, uS4, uS5, S7, uS8, uS9, uS10, uS11, uS12, uS13, uS14, uS15, uS17, uS19**) envolvidas na estrutura da subunidade menor (SSU) (SMITH et al., 2008; NIKOLAY et al., 2015; WILSON et al., 2008).

4.3. CONSTRUÇÃO DOS ARQUIVOS MULTIFASTA

As sequências de aminoácidos correspondentes a cada proteína foram extraídas do NCBI (*National Center for Biotechnology Information*), disponível em: <https://www.ncbi.nlm.nih.gov/>, entre os dias 10 de fevereiro e 10 de Março de 2020, utilizando um programa de busca específica de múltiplas sequências escrito na linguagem de programação *Python*.

A seleção das sequências pelo programa implementado foi feita com base nos elementos descritivos presentes na primeira linha do formato FASTA (é um formato baseado em letras, onde as sequências de nucleotídeos e aminoácidos são representadas por letras individuais que representam cada monômero, para identificação das sequências, o formato fasta conta com uma linha de comentário que é inicializada a partir de um sinal de maior (>)). Para ser selecionada, a sequência FASTA de cada proteína deveria conter o nome da proteína, o nome em inglês da subunidade ribossomal ao qual pertence ou a representação numérica 50 S ou 30 S, e o nome de uma das espécies da lista de organismos.

Após a extração, todas as sequências organizadas em arquivos multifasta foram visualizadas individualmente para a certificação. Com isso, observou-se que alguns grupos de proteínas continham sequências erradas (fatores de tradução que atuam em proteínas ribossomais, por isso, continham a descrição das suas proteínas alvos na linha de comentário das sequências fasta), ou parciais. Todas as sequências erradas foram substituídas manualmente pelas sequências corretas, utilizado a busca direta no NCBI, com base no nome da proteína e da espécie, e no caso de não disponibilidade da proteína ribossômica correspondente a procura, a sequência foi descartada. No caso das parciais, em alguns casos, a proteína completa não foi encontrada pelo mecanismo de busca manual, porém, nestes casos, as sequências não foram apagadas dos arquivos, por apresentarem tamanhos acima de 50% do tamanho médio das proteínas do grupo ao qual pertenciam.

Foram organizados 136 arquivos multifasta no formato txt (34 x 4), sendo 34 por grupo (*Archaea* - Halófilos; *Bacteria* - Halófilos; *Archaea* - Controle; e *Bacteria* - Controle), ou seja, 1 para cada grupo de proteínas.

4.4. SEQUÊNCIAS EXTRAÍDAS

Com base nos parâmetros de seleção utilizados, descritos acima, um total de 17.255 sequências das 34 proteínas ribossômicas homólogas aos três domínios da vida, foram extraídas para a categoria dos halófilos dos domínios *Archaea* e *Bacteria* e dos grupos controles. Foram 4.427 sequências fasta do grupo *Archaea*-halófilos (2.467 da subunidade maior (LSU) e 1.960 correspondente à subunidade menor (SSU)); 8.889 sequências do grupo *Bacteria* halófilos, sendo 4.941 das proteínas da subunidade maior (LSU) e 3.948 sequências da subunidade menor (SSU); 569 sequências do grupo *Archaea* (controle) com 305 da subunidade maior (LSU) e 264 correspondentes à subunidade menor (SSU); 3.370 arquivos fasta correspondentes ao grupo *Bacteria* (Controle), com 1.920 das proteínas da subunidade maior (LSU), e 1.450 sequências da subunidade menor (SSU) (Tabela 2).

	<i>ARCHAEA</i> (H)	<i>BACTERIA</i> (H)	CONTROLE (A)	CONTROLE (B)
LSU	2.467	4.941	305	1.920
SSU	1.960	3.948	264	1.450
TOTAL	4.427	8.889	569	3.370

Tabela 2: Quantidades de proteínas extraídas por grupo. A primeira linha em azul estão os grupos (entre parênteses estão as letras H: que representa o nome halófilos; enquanto as letras A e B significam *Archaea* e *Bacteria*, respectivamente). Na primeira coluna em azul estão as abreviações das subunidades (LSU) maior, e (SSU) menor; as demais colunas em branco contém os valores das sequências extraídas. Fonte: Autoria Própria.

Devido ao fato de nem todos os organismos apresentarem todo o genoma sequenciado e anotado, e nem possuírem todas as informações sobre suas sequências proteicas disponíveis, a quantidade de sequências extraídas não é equivalente ao número total bruto esperado, ou seja $865 \times 34 = 29.410$ sequências do grupo dos halófilos, e $190 \times 34 = 6.460$ do grupo controle. Na tabela 3 abaixo está a quantificação das sequências fasta obtidas para cada proteína de cada grupo. As linhas em azul contém os nomes de cada uma das 34 proteínas, enquanto a primeira coluna está o nome abreviado dos grupos; e nas demais colunas está a quantificação das sequências extraídas.

*	uL1	uL2	uL3	uL4	uL5	uL6	uL10	uL11	uL12	uL13	uL14	uL15
A(H)	143 (0)	137 (0)	140 (1)	157 (0)	152 (0)	152 (0)	64 (0)	142 (0)	22 (0)	144 (0)	96 (0)	156 (0)
B(H)	294 (0)	298 (1)	303 (0)	310 (1)	289 (1)	306 (2)	7 (0)	303 (0)	246 (2)	275 (1)	248 (0)	309 (0)
A(C)	19 (0)	17 (0)	18 (0)	18 (0)	18 (0)	18 (0)	10 (0)	19 (0)	3 (0)	18 (0)	17 (0)	19 (0)
B(C)	95 (0)	111 (0)	93 (0)	112 (0)	109 (0)	114 (0)	17 (0)	99 (0)	81 (0)	94 (0)	112 (0)	124 (0)
	uL16	uL18	uL22	uL23p	uL24	uL29	uL30	uS2	uS3	uS4	u S5	uS7
A(H)	137 (0)	142 (1)	150 (0)	88 (0)	155 (0)	135 (0)	155 (0)	161 (0)	160 (0)	138 (0)	136 (0)	152 (0)
B(H)	270 (0)	305 (1)	275 (0)	105 (0)	285 (0)	256 (2)	257 (0)	312 (1)	285 (0)	273 (0)	260 (1)	269 (1)
A(C)	17 (0)	18 (0)	18 (0)	7 (0)	18 (0)	16 (0)	17 (0)	18 (0)	18 (0)	18 (0)	18 (0)	18 (0)
B(C)	104 (0)	124 (0)	124 (0)	59 (0)	115 (0)	124 (0)	109 (0)	112 (0)	116 (1)	84 (0)	91 (0)	104 (0)
		uS8	uS9	uS10	uS11	uS12	uS13	uS14	uS15	uS17	uS19	
A(H)		54 (0)	123 (0)	146 (0)	101 (0)	96 (0)	148 (1)	116 (1)	149 (0)	138 (1)	142 (0)	
B(H)		214 (4)	260 (4)	169 (0)	247 (1)	259 (0)	295 (2)	283 (0)	280 (1)	284 (1)	258 (0)	
A(C)		18 (0)	18 (0)	16 (0)	16 (0)	17 (0)	19 (0)	17 (0)	19 (0)	16 (0)	18 (0)	
B(C)		51 (0)	74 (0)	94 (0)	95 (0)	73 (0)	123 (0)	105 (0)	113 (0)	109 (0)	106 (0)	

*A(H) : *Archaea* (Halófilos); B(H): *Bacteria* (Halófilos); A(C): *Archaea* (Controle); e B(C): *Bacteria* (Controle)

Tabela 3: Contagem das sequências das 34 proteínas referentes a cada grupo. Os números entre parênteses representam a quantidade de sequências parciais nos arquivos multifasta das 34 proteínas de cada grupo. Fonte: Autoria Própria.

4.5. ANÁLISE ESTATÍSTICA E DEMONSTRAÇÃO GRÁFICA

Para a obtenção das informações iniciais das características de cada proteína e das mesmas agrupadas por subunidade, foram feitas as contagens dos aminoácidos de cada sequência através de um programa escrito na linguagem *Python*. A partir da contagem, foram contabilizados os valores das frequências absoluta e relativa dos aminoácidos, médias, variâncias e do desvio padrão, sendo os dados organizados em planilhas do Google Drive e do software Excel, para promover a comparação entre os grupos. A frequência absoluta (FA),

de uma forma simplificada, é a contagem de vezes que um determinado elemento aparece em uma amostra. Já a frequência relativa (FR) é a representação em porcentagem do valor da frequência absoluta em relação a amostra inteira. A média é uma medida de tendência central, e o desvio-padrão é um estimador da variabilidade dos dados, muito usado quando a variável analisada é do tipo contínua ou paramétrica.

Um desvio padrão grande significa que os valores amostrais estão bem distribuídos em torno da média, enquanto um desvio padrão pequeno indica que eles estão condensados próximos a média. De uma forma simplificada, quanto menor o desvio, mais homogênea é a amostra. Testes de normalidade, homocedasticidade, e testes de significância foram feitos utilizando bibliotecas da linguagem *Python*, para proporcionar a comparação dos grupos (Figura 11). Por fim, a representação gráfica foi feita utilizando o Software estatístico Minitab 19, disponível em: <https://www.minitab.com/pt-br/>.

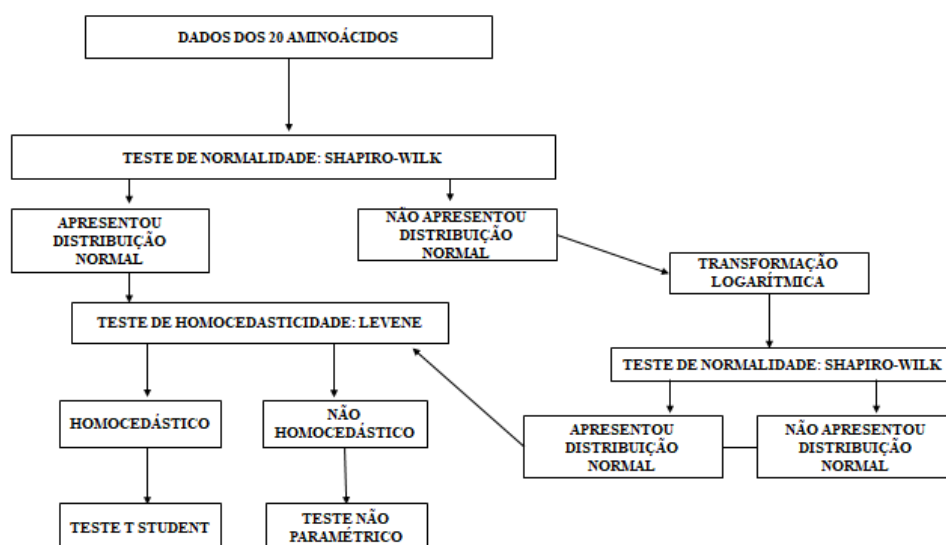


Figura 11: Esquema explicativo dos testes usados para avaliar a significância dos valores comparados. Fonte: Autoria Própria.

4.6. TESTES UTILIZADOS

4.6.1. Teste de normalidade

Para a escolha adequada dos testes de significância para a análise dos dados, é necessário verificar qual é o tipo de distribuição dos mesmos, ou seja, se os dados são paramétricos ou não paramétricos. Esse tipo de verificação é imprescindível, tanto nos dados do grupo de forma global, quanto na análise de subgrupos ou em comparações múltiplas. Existe uma grande variedade de testes estatísticos que verificam o ajuste dos dados à distribuição normal a partir de diferentes pressupostos e algoritmos. Porém, todos os testes pressupõem a hipótese de normalidade dos dados (H_0), retornando um $p\text{-valor} > 0,05$, se resultarem na aderência aos parâmetros de normalidade (MIOT, 2017). O teste de normalidade utilizado neste trabalho foi o Shapiro-Wilk, que em diversas simulações apresentou-se com um excelente desempenho (MIOT, 2017).

4.6.2. Normalização dos dados

Para a realização da estatística paramétrica existem alguns requisitos, como “a normalidade da distribuição dos erros, homogeneidade das variâncias, e a aditividade dos efeitos dos fatores de variação”. A transformação dos dados amostrais é uma opção quando esses requisitos não são preenchidos pelos dados da amostra experimental, como uma última tentativa para a não realização de testes não-paramétricos (CAMPOS, 2000). Existe uma quantidade significativa de tipos de transformações diferentes, sendo a utilizada neste trabalho a transformação logarítmica, como tentativa de se obter uma distribuição normal dos dados.

4.6.3. Teste de homocedasticidade: Levene

Antes de executar um teste comparativo de médias, é necessário verificar se os dados são homocedásticos, ou seja, se há homogeneidade de variâncias ou não. O teste de Levene se enquadra neste contexto, sendo considerado um teste robusto em suas medições até para amostras que fogem da normalidade (ALMEIDA et al., 2008).

4.6.4. Teste *t* de *student*

O conhecido *t*-Student é um dos testes mais utilizados para avaliar a diferença entre duas médias. Ele consiste basicamente em testar a hipótese H_0 *versus* H_1 , e com base no valor de p , é atribuído se há ou não diferenças significativas. Para a sua realização é necessário confirmar o padrão de distribuição dos dados, se os mesmos são normais ou não, caso sejam paramétricos, o teste *t* pode ser utilizado no conjunto de dados (BARROS et al., 2005).

4.6.5. Teste de correlação de Pearson

O coeficiente de correlação de Pearson tem origem no trabalho conjunto de Karl Pearson e Francis Galton. De uma forma simplificada a correlação é uma medida de associação entre duas variáveis, ou seja, a medida de associação bivariada do grau de relacionamento entre duas variáveis. Para compreendê-la é necessário considerar dois conceitos importantes: "associação" e "linearidade". Em termos estatísticos, duas variáveis se associam quando elas guardam semelhanças na distribuição dos seus escores. Mais precisamente, elas podem se associar a partir da distribuição das frequências ou pelo compartilhamento de variância. No caso da correlação de Pearson (r) vale esse último parâmetro, ou seja, ele é uma medida da variância compartilhada entre duas variáveis (FILHO et al., 2009).

5. RESULTADOS E DISCUSSÃO

Os resultados mostrados e discutidos a seguir são derivativos de uma pesquisa geradora de dados quantitativos. De forma resumida, o objetivo que levou a realização desse levantamento científico, girou em torno da dúvida sobre as adaptações básicas das proteínas ribossômicas de organismos halófilos dos domínios *Archaea* e *Bacteria*. Para tanto, utilizou-se de um dos recursos mais abundantes atualmente nas ciências biológicas, dados de sequências proteicas disponíveis em bancos de dados públicos.

Utilizando a abordagem comparativa com base em avaliações estatísticas básicas, buscou-se avaliar, inicialmente, os padrões de composição de aminoácidos de 34 proteínas ribossômicas universais em organismos halófilos e controles não extremófilos. Devido a sua estruturação, os arquivos de sequências são materiais relativamente fáceis de se manipular, por causa da codificação simplificada com base em uma organização na forma de listas de letras, onde cada uma equivale a um monômero. O arcabouço dessa seção é com base em gráficos, tabelas, descrições dos resultados e uma abordagem discursiva após cada bloco de resultados. Priorizou-se esse formato com a perspectiva de promover uma leitura mais dinâmica e com um relacionamento mais direto entre os dados representativos e escritos.

5.1. ANÁLISE ESTATÍSTICA DOS DADOS E DEMONSTRAÇÃO GRÁFICA

5.1.1. Comparação das médias e porcentagens das 34 proteínas ribossômicas dos domínios procariotos

A primeira avaliação consistiu na quantificação das médias e das porcentagens dos 20 aminoácidos das 34 proteínas dos 4 grupos (*Archaea*-halófilos + controle e *Bacteria*-halófilos + controle). Com isso, objetivou-se comparar os halófilos com os seus respectivos grupos controles, além, de verificar as diferenças entre os dois grupos de extremófilos.

No Gráfico 1 são apresentadas as médias. Os dados gráficos mostram 3 pontos importantes: as linhas que representam os grupos do domínio *Bacteria* (vermelha e roxa) apresentam sobreposições ou muita proximidade na maioria dos pontos, onde pode-se inferir uma diferença pouco significativa entre a composição de aminoácidos das proteínas de

organismos halófilos bacterianos e seus controles não-halófilos. Isso é corroborado com os dados do teste t presente na (Tabela 4), onde nenhum valor foi concebido com um resultado significativo. O segundo ponto é o comportamento dos grupos do domínio *Archaea*, assim como no caso bacteriano, existe uma proximidade entre halófilos e controle, porém, em alguns pontos existem diferenças significativas em relação às médias: os aminoácidos Alanina (A) ($p = 0,0345$), Ácido aspártico (D) ($p = 0,0125$), Ácido glutâmico (E) ($p = 0,0234$), Isoleucina (I) ($0,0491$) e o aminoácido Lisina (K) ($p = 0,0298$) tiveram valores significativamente diferentes entre os grupos (Tabela 5). O terceiro ponto que pode ser visualizado no gráfico é a maior proximidade entre a linha do grupo controle arqueal (verde) com as dos grupos bacterianos (roxa e vermelho).

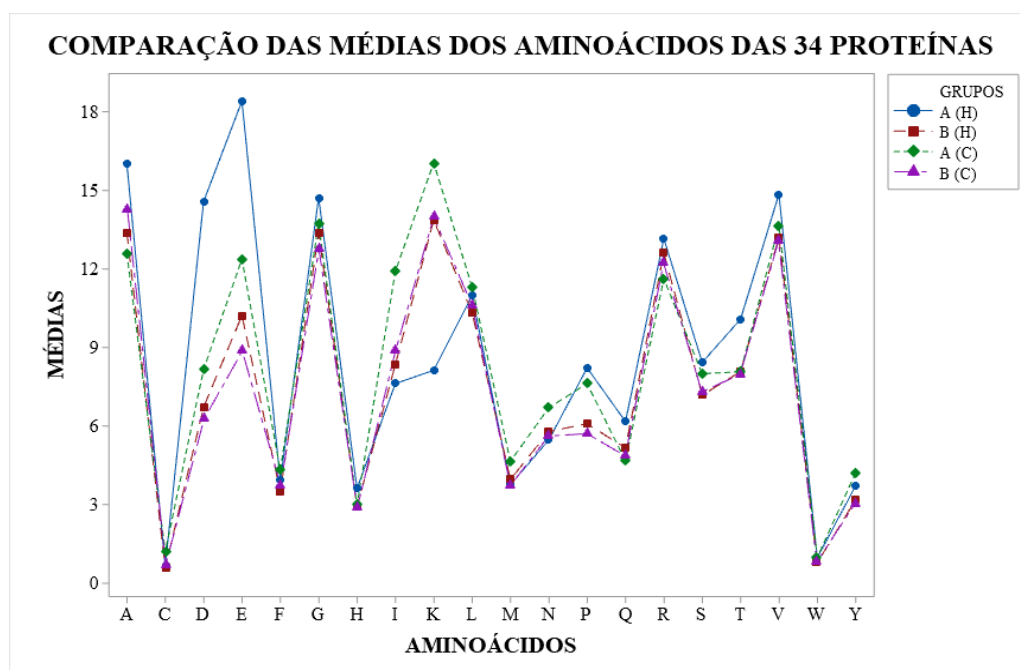


Gráfico 1: Representação das médias dos 20 aminoácidos das 34 proteínas ribossômicas dos quatro grupos analisados. Na pequena caixa à direita está a relação entre as linhas e os grupos: A (H) em azul é a linha que representa os valores do grupo *Archaea*-halófilos; B (H) equivale a *Bacteria*-halófilos codificado em vermelho; A(C) em verde é o grupo controle do domínio *Archaea*; e simbolizado pela linha roxa, o grupo B(C), controle não halófilo bacteriano. As letras no eixo x representam os aminoácidos com base no código de uma letra dos mesmos. Fonte: Autoria Própria.

Aminoácidos	Médias (B-H)	DP*	(%)	Médias (B-C)	DP*	(%)	<i>p-value</i>
Alanina (A) - H. Alifático	13,38	7,27	8,98%	14,26	7,67	9,71%	0,4538
Cisteína(C) - Polar N.C	0,56	0,86	0,37%	0,66	0,97	0,45%	0,4359
Ácido aspártico (D) - Negativo	6,7	4,1	4,49%	6,24	3,82	4,25%	0,8763
Ácido glutâmico (E) - Negativo	10,2	5,34	6,85%	8,84	4,86	6,02%	0,0993
Fenilalanina (F) - H. Aromático	3,48	2,44	2,34%	3,7	2,57	2,52%	0,7654
Glicina (G) - H. Alifático	13,35	7,66	8,96%	12,75	7,63	8,68%	0,8211
Histidina (H) - Positivo	2,89	2,2	1,94%	2,83	2,2	1,93%	0,8972
Isoleucina (I) - H. Alifático	8,33	4,53	5,59%	8,85	4,57	6,03%	0,9311
Lisina (K) - Positivo	13,85	5,22	9,29%	13,97	5,39	9,52%	0,8236
Leucina (L) - H. Alifático	10,31	4,8	6,92%	10,59	4,93	7,21%	0,5687
Metionina (M) - H. Alifático	3,94	2,18	2,64%	3,7	2,16	2,52%	0,8876
Asparagina (N) - Polar N.C	5,77	3,35	3,87%	5,59	3,53	3,81%	0,5541
Prolina (P) - H. Alifático	6,07	3,63	4,07%	5,68	3,65	3,87%	0,7913
Glutamina (Q) - Polar N.C	5,14	3,06	3,45%	4,83	3,02	3,29%	0,4349
Arginina (R) - Positivo	12,59	6,27	8,45%	12,21	6,27	8,31%	0,8622
Serina (S) - Polar N.C	7,16	3,62	4,80%	7,28	3,84	4,96%	0,9160
Treonina (T) - Polar N.C	8,06	3,75	5,41%	7,95	3,73	5,41%	0,9874
Valina (V) - H. Alifático	13,18	6,07	8,85%	13,05	6,25	8,89%	0,7321
Triptofano (W) - H. Aromático	0,75	0,96	0,50%	0,77	1,15	0,52%	0,6989
Tirosina (Y) - H. Aromático	3,17	2,31	2,13%	3	2,28	2,01%	0,8896

Tabela 4: Dados das médias, porcentagens, desvio padrão e resultados do teste t de *Student* dos 20 aminoácidos presentes nas 34 proteínas ribossômicas do domínio *Bacteria* (halófilos + controles). Fonte: Autoria Própria.

Aminoácidos	Médias (A-H)	DP	(%)	Médias (A-C)	DP	(%)	<i>p-value</i>
Alanina (A) - H. Alifático	16,5	7,19	9,40%	12,57	6,69	7,64%	0,0345
Cisteína(C) - Polar N.C	0,64	0,99	0,37%	1,18	1,23	0,71%	0,0567
Ácido aspártico (D) - Negativo	14,55	8,24	8,39%	8,14	3,95	4,94%	0,0125
Ácido glutâmico (E) - Negativo	18,39	13,84	10,60%	12,35	6,92	7,50%	0,0234
Fenilalanina (F) - H. Aromático	3,89	2,32	2,24%	4,3	2,44	2,61%	0,5988
Glicina (G) - H. Alifático	14,7	7,69	8,47%	13,72	6,7	8,34%	0,4569
Histidina (H) - Positivo	3,59	2,22	2,07%	2,97	2,13	1,81%	0,7538
Isoleucina (I) - H. Alifático	7,6	3,87	4,38%	11,89	4,91	7,23%	0,0491
Lisina (K) - Positivo	8,11	4,04	4,67%	16,03	6,19	9,74%	0,0298
Leucina (L) - H. Alifático	10,96	5,08	6,32%	11,3	5,35	6,87%	0,8675
Metionina (M) - H. Alifático	3,74	1,88	2,16%	4,63	2,21	2,81%	0,7897
Asparagina (N) - Polar N.C	5,45	2,73	3,14%	6,68	3,32	4,06%	0,5632
Prolina (P) - H. Alifático	8,2	5,32	4,73%	7,6	4,33	4,62%	0,2985
Glutamina (Q) - Polar N.C	6,15	2,98	3,55%	4,65	2,7	2,83%	0,0985
Arginina (R) - Positivo	13,14	5,87	7,60%	11,61	5,05	7,06%	0,8755
Serina (S) - Polar N.C	8,41	3,75	4,85%	7,98	4,06	4,85%	0,7984
Treonina (T) - Polar N.C	10,06	4,48	5,80%	8,05	4,2	4,89%	0,0992
Valina (V) - H. Alifático	14,83	6,83	8,55%	13,63	6,78	8,28%	0,6549
Triptofano (W) - H. Aromático	0,95	1,01	0,55%	0,97	1,03	0,59%	0,9159
Tirosina (Y) - H. Aromático	3,7	2,24	2,13%	4,19	2,39	2,54%	0,6732

Tabela 5: Dados das médias, porcentagens, desvio padrão e resultados do teste t de *Student* dos 20 aminoácidos presentes nas 34 proteínas ribossômicas do domínio *Archaea* (halófilos + controles). Os números marcados em vermelho são valores significativos em relação à diferença entre as médias. Fonte: Autoria Própria.

Uma comparação mais geral foi feita através das análises das porcentagens dos aminoácidos agrupados pelas características das suas cadeias laterais (hidrofóbicos alifáticos, hidrofóbicos aromáticos, negativo, polar não-carregado e positivo). Com base na Tabela 6 abaixo é possível observar que ambos os grupos de extremófilos apresentam porcentagens de aminoácidos hidrofóbicos (alifáticos e aromáticos) mais baixas do que os seus respectivos controles, porém, estatisticamente não foi calculada uma significância aceitável para inferir um menor nível de hidrofobicidade através dessa comparação. Assim, como no caso dos aminoácidos hidrofóbicos, a porcentagem de aminoácidos carregados positivamente, principalmente no grupo halófilo arqueal (*Archaea* H), foi expressivamente menor do que no seu controle, porém, a diferença não foi significativa com base no teste t de Student.

Os únicos resultados que apresentaram valores estatísticos significativos foram os referentes aos aminoácidos negativos. Na comparação entre os dois grupos do domínio *Archaea*, a porcentagem dos negativos dos halófilos se mostrou significativamente mais alta do que a do seu grupo controle. Ao compararmos os dois grupos de halófilos, o único valor significativo também está relacionado com os aminoácidos negativos (Tabela 6).

	<i>Archaea</i> (H)	Controle (A)	<i>p-value</i>	<i>Bacteria</i> (H)	Controle (B)	<i>p-value</i>	<i>p-value***</i> (A+B)
H. Alifático*	44,01%	45,79%	0,645561	46,01%	46,91%	0,395479	0,830133
H. Aromático**	4,92%	5,74%	0,144998	4,97%	5,05%	0,787402	0,598237
Negativo	18,99%	12,44%	0,033986	11,34%	10,27%	0,320805	0,012481
Polar N.C	17,71%	17,34%	0,837287	17,90%	17,92%	0,945791	0,847915
Positivo	14,34%	18,61%	0,517113	19,68%	19,76%	0,82858	0,34394

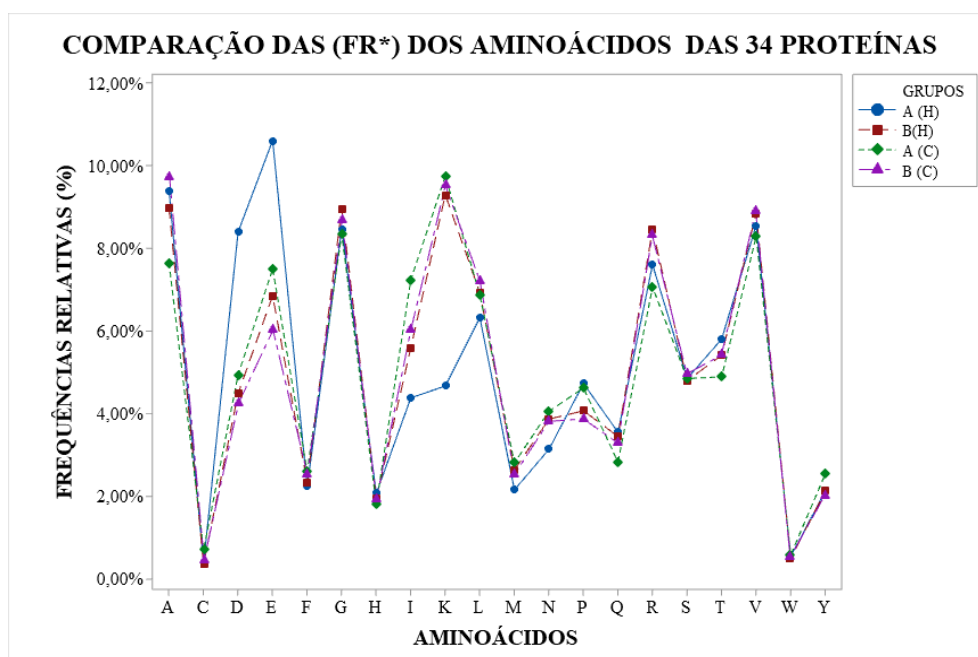
(*Hidrofóbico Alifático; **Hidrofóbico Aromático; ***valor de p da comparação entre os halófilos dos domínios *Archaea* (A) e *Bacteria* (B)).

Tabela 6: Porcentagem dos aminoácidos das 34 proteínas ribossômicas agrupados pelas características de suas cadeias laterais. Na primeira coluna (azul) estão os tipos de aminoácidos com base em suas cadeias laterais; na primeira linha (azul) estão os nomes dos grupos e a representação do valor de p; nas demais caixas (brancas) estão os valores das porcentagens de cada grupo de aminoácido de cada perfil de organismos e os valores de significância do teste t comparativo, sendo os significativos marcados de vermelho. Fonte: Autoria Própria.

5.1.2. Perfil dos aminoácidos

No Gráfico 2 abaixo são mostradas as frequências relativas de cada um dos aminoácidos nos quatro grupos analisados. Esse gráfico foi construído com base nas informações das Tabelas 4 e 5. O padrão encontrado neste gráfico é similar ao das médias, mostrando um perfil de porcentagens bem parecido entre os grupos na maioria dos pontos. A linha azul representa o domínio *Archaea*, a única que apresenta distância mais expressiva do seu controle em cinco pontos: o relacionado ao aminoácido Alanina (A); os relacionados aos dois aminoácidos negativos, ácido aspártico (D) e ácido glutâmico (E); o ponto relativo a isoleucina (I); e ao aminoácido Lisina (K).

No perfil mostrado na Tabela 7, vemos que o padrão dos três aminoácidos apolares mais representativos não muda (Alanina, Glicina, e Valina) em nenhum dos grupos de ambos os domínios, porém, destacado em verde, no domínio *Archaea* é possível perceber que a Alanina é a mais representativa dos três nos organismos halófilos arqueais, enquanto a glicina é mais abundante no controle. Outro ponto, talvez o mais significativo sobre as diferenças dos hidrofóbicos dos grupos do domínio *Archaea*, é que a porcentagem de Isoleucina no controle é quase duas vezes maior do que nos halófilos. Outro ponto importante diz respeito aos aminoácidos positivos, onde o mais representativo nos halófilos arqueais é a Arginina (R), enquanto no seu controle há um maior nível de Lisina (K).



* (FR) Significa Frequência Relativa

Gráfico 2: Frequência relativa (%) dos vinte aminoácidos das 34 proteínas. Fonte: Autoria Própria.

<i>Archaea-halófilos</i>	<i>Archaea-controle</i>	<i>Bacteria-halófilos</i>	<i>Bacteria-controle</i>
Alanina - 9,40%	Glicina - 8,34%	Alanina - 8,98%	Alanina - 9,71%
Valina - 8,55%	Valina - 8,28%	Glicina - 8,96%	Valina - 8,89%
Glicina - 8,47%	Alanina - 7,64%	Valina - 8,85%	Glicina - 8,68%
Leucina - 6,32%	Isoleucina - 7,23%	Leucina - 6,92%	Leucina - 7,21%
Prolina - 4,73%	Leucina - 6,87%	Isoleucina - 5,59%	Isoleucina - 6,03%
Isoleucina - 4,38%	Prolina - 4,62%		
Fenilalanina - 2,24%	Metionina - 2,81%		
Ácido glutâmico - 10,60%	Ácido glutâmico - 7,50%	Ácido glutâmico - 6,85%	Ácido glutâmico - 6,02%
Ácido aspártico - 8,39%	Ácido aspártico - 4,94%	Ácido aspártico - 4,49%	Ácido aspártico - 4,25%
Arginina - 7,60%	Lisina - 9,74%	Lisina - 9,29%	Lisina - 9,52%
Lisina - 4,67%	Arginina - 7,06%	Arginina - 8,45%	Arginina - 8,31%
Histidina - 2,07%	Histidina - 1,81%	Histidina - 1,94%	Histidina - 1,93%

Tabela 7: Perfil dos aminoácidos hidrofóbicos e carregados das proteínas dos dois domínios. Os aminoácidos coloridos em verde estão destacados para mostrar o perfil dos três aminoácidos hidrofóbicos mais abundantes, porém, com padrão de posicionamento diferente em cada grupo. O bloco da isoleucina está destacado em rosa para destacar a diferença das porcentagens em cada grupo. Assim, como os aminoácidos positivos enquadrados em laranja, para destacar a proporção de arginina e lisina em cada grupo. Fonte: Autoria Própria.

Os resultados expostos até o momento mostraram dois perfis para as proteínas ribossomais de organismos halófilos. As proteínas halofílicas do domínio *Archaea* se mostraram significativamente diferentes das proteínas do seu controle não halófilo, e diferente das proteínas bacterianas halofílicas. Enquanto isso, as bactérias halofílicas tiveram um padrão próximo ao mesofílico, porém, considerações sobre algumas diferenças apareceram e foram feitas ao longo da discussão.

O padrão diferenciado observado nas proteínas halofílicas arqueais é, principalmente, derivado do seu maior nível de aminoácidos ácidos (ácido aspártico (D) e ácido glutâmico

(E)), que se mostraram significativos (Tabela 5), diminuição de carga positiva (menores porcentagens de lisina (K)), que também se mostrou significativa em relação ao controle, e uma diminuição na porcentagem dos aminoácidos hidrofóbicos. Esse resultado se torna interessante à luz dos conhecimentos sobre as propriedades dos ácidos nucleicos, que inclui o RNA ribossomal (rRNA), e também quando se avalia juntamente as características das proteínas ribossômicas de organismos classificados como não extremófilos (mesófilos).

Os ácidos nucleicos possuem uma carga negativa derivada do seu esqueleto de açúcar-fosfato (FEDYUKINA et al. 2014); e as proteínas ribossômicas dos organismos mesófilos são avaliadas como ricas em resíduos de aminoácidos positivos, principalmente lisina, sendo essa dualidade de carga, com base em princípios fundamentais da eletrostática, importante para a estabilidade do complexo ribonucleoproteico (OREN, 2011).

Com base em métodos computacionais, Burton e colaboradores (2012) analisaram a dinâmica das proteínas ribossômicas durante a montagem do ribossomo. Neste estudo é apontado que as proteínas ribossômicas são carregadas positivamente e a porcentagem de resíduos positivos nessas proteínas é, aproximadamente, o dobro da média: Lys (K) + Arg (R) é 18,7% para *Escherichia coli* e 21,2%, para *Thermus thermophilus*. Além disso, os resíduos positivos constituem uma grande proporção de resíduos de contato com o RNA: 39% para *E. coli* e 46%, para *T. thermophilus*. Com base nos seus resultados, afirma a importância das interações carga-carga na montagem do ribossomo (BURTON et al., 2012).

Com base nessas considerações, resultados que atribuem um elevado nível de carga negativa para as proteínas ribossômicas promovem reformulações interessantes quando se pensa em estabilidade ribossômica. Dados similares aos deste presente trabalho, que fornecem subsídios baseados em uma abordagem relativamente diferente, foram obtidos por Fedyukina e colaboradores (2014) com uma abordagem baseada em características de estruturas tridimensionais. Suas análises se deram em torno das proteínas da subunidade 50S de 10 espécies, onde foi encontrada uma distinção entre a carga líquida de proteínas ribossômicas de organismos halofílicos e não halofílicos (FEDYUKINA et al., 2014).

Os autores apontam que 67% das proteínas ribossômicas dos halófilos são carregadas negativamente, enquanto apenas 15% das proteínas ribossômicas dos não halófilos compartilham essa propriedade. Além disso, outro ponto interessante colocado nos seus resultados é que a hidrofobicidade tende a ser menor para proteínas ribossômicas de halófilos, do que para as proteínas correspondentes de não halófilos, sendo esse padrão verificado na

amostragem dos três domínios da vida, porém, o domínio *Archaea* possui valores mais significativos que os demais domínios (FEDYUKINA et al. 2014). Essa mesma colocação também foi verificada nos resultados deste presente trabalho, mesmo que esse comportamento não tenha se mostrado significativo a partir de teste comparativo, em escala numérica, as proteínas ribossômicas analisadas tiveram porcentagens menores dos dois tipos de aminoácidos hidrofóbicos (alifáticos e aromáticos) em ambos os grupos: *Archaea*-halófilos e *Bacteria*-halófilos.

Bolhuis e colaboradores (2010), Hutcheon e colaboradores (2005), Kennedy e colaboradores (2001), Joo e colaboradores (2005) e Mevarech e colaboradores (2000) também apontaram que as principais adaptações encontradas nas proteínas halofílicas são o excesso de resíduos ácidos, baixa quantidade de lisinas e o aumento do número de aminoácidos hidrofóbicos com cadeias laterais pequenas, levando à diminuição de resíduos hidrofóbicos com cadeias laterais longas ou aromáticas, e isso promove consequentemente a diminuição da hidrofobicidade da molécula (HUTCHEON et al., 2005; KENNEDY et al., 2001; JOO et al., 2005; MEVARECH et al., 2000).

Com base nesse último ponto, no presente estudo, tivemos como resultado significativo nos halófilos arqueais uma maior porcentagem do aminoácido alanina, classificado como um dos hidrofóbicos mais simples, perdendo apenas para a glicina (sua cadeia lateral é formada apenas por um grupo metil), e uma diminuição em cerca de 40% na porcentagem de isoleucina (juntamente com seu isômero leucina, é um dos maiores aminoácidos alifáticos; sua cadeia lateral contém quatro carbonos totalmente protonados dispostos em forma de "L" que é volumosa e muito hidrofóbica), em relação ao seu controle (Tabela 7) (NELSON et al., 2014)). Os resultados de Paul e colaboradores (2008) indicam claramente que as proteínas halófilas preferem usar ácido aspártico, ácido glutâmico, valina e treonina à custa de lisina, metionina, leucina, isoleucina e cisteína, que são relacionados com níveis maiores de hidrofobicidade e rigidez em estruturas proteicas. Isso leva a considerações de como esses atributos observados estruturam o arranjo macromolecular ribossômico de forma funcional nas células dos organismos que vivem imersos em concentrações salinas altas.

Em diferentes trabalhos, funções foram sugeridas para o papel dos resíduos ácidos nas proteínas halofílicas. Eles são colocados como importantes para a ligação de moléculas de água essenciais para a hidratação, mas também podem desempenhar um papel crítico na

interação com íons que formam os diferentes tipos de sais, proporcionando que mais moléculas de água fiquem livres, resultando assim, na prevenção da agregação, sendo esses elementos essenciais para a viabilidade metabólica de qualquer organismo celular vivo. Quando uma quantidade muito grande de sal está presente em meio aquoso, seus íons dissolvidos tendem a interagir com as moléculas de água presentes no meio, quando isso ocorre no citoplasma, o excesso de sal compete pelo solvente com as moléculas orgânicas, com isso, as moléculas sem adaptações halofílicas tendem a se agruparem. A redução da quantidade de aminoácidos hidrofóbicos, ou a substituição de resíduos com cadeias laterais altamente hidrofóbicas por outros com cadeias laterais menos hidrofóbicas (redução da hidrofobicidade) também é colocado como importante fator para o impedimento da agregação, a fim de manter a flexibilidade em altas concentrações de sal, pois o sal age, em pequena quantidade, para promover uma maior solubilidade em moléculas carregadas, como proteínas, mas, em excesso, pode causar agregação molecular, que impede a difusão de moléculas no citoplasma (HUTCHEON et al., 2005; KENNEDY et al., 2001).

Um outro ponto relacionado a esses, é o aumento das pontes salinas, considerada uma adaptação estrutural importante para a halofilia e derivadas dos dois elementos citados acima: aumento de aminoácidos polares com capacidade de realizar interações iônicas e ligações de hidrogênio e diminuição da hidrofobicidade (BOLHUIS et al., 2010). Paul e colaboradores (2008) enfatizam esse ponto: “o maior uso de resíduos de aminoácidos carregados negativamente leva à organização de uma rede de íons de sal hidratada na superfície de proteínas e à formação de pontes salinas com resíduos básicos estrategicamente posicionados” (PAUL et al., 2008).

Outro trabalho que proporciona uma visão composicional/estrutural é o de Sigliocco e colaboradores (2011), que obteve como resultados características que demonstram como esses padrões de sequência primária estruturam proteínas halofílicas estáveis. Partindo das características superficiais das proteínas halofílicas, demonstraram uma redução significativa da superfície apolar na amostra arqueana e um aumento paralelo da área polar formada pelos átomos de oxigênio da cadeia lateral, enquanto os átomos da cadeia lateral de nitrogênio diminuem sua contribuição na superfície. Esses resultados refletem o aumento da frequência superficial de aspartato e de glutamato, combinada à diminuição de resíduos de lisina (SIGLIOCCO et al., 2011).

Um outro acréscimo interessante dos resultados de Sigliocco e colaboradores (2011) como comparativo com os resultados deste presente estudo, é que ele atribui significância para os valores observados nas proteínas halófilas bacterianas. Nos resultados deste presente trabalho, como revelado no início da seção de resultados e discussão, o grupo halofílico bacteriano apareceu com resultados não significativos em relação ao seu controle, além disso, o perfil dos aminoácidos não demonstrava distinção, porém, quando analisa-se as comparações utilizando apenas o atributo numérico de menor, maior ou igual ($<$, $>$ ou $=$), as porcentagens dos aminoácidos ácidos e as porcentagens dos aminoácidos hidrofóbicos se mostraram diferentes do seu controle, pois nas proteínas halofílicas, as porcentagens negativas eram maiores e as dos aminoácidos hidrofóbicos eram menores.

Nos seus resultados, Sigliocco e colaboradores (2011) observaram uma diminuição significativa da área da superfície apolar na proteína sintetizada pelos organismos halofílicos na amostra que utilizava como estratégia a produção de solutos orgânicos, estratégia tipicamente bacteriana, salvo exceções, em paralelo com uma tendência aparente para aumentar a presença de átomos de oxigênio de cadeia lateral. A maioria dos osmólitos compatíveis produzidos por procariontes halofílicos são espécies neutras, que contêm grupos carregados (por exemplo, derivados de aminoácidos); portanto, a manutenção da solubilidade e estabilidade das proteínas na concentração decimolar de osmólitos exige remodelação superficial parcialmente semelhante à observada na amostra dos organismos halófilos extremos, implicando diminuição da superfície apolar; no entanto, deve-se destacar que mesmo havendo esse remodelamento com a mesma tendência, a extensão geral da modificação da superfície é muito menor do que a observada nas proteínas halófilas extremas, que possuem níveis altos de sais no citosol (SIGLIOCCOLO et al., 2011).

Outro trabalho que atribui diferença entre as composições de proteínas pela simples comparação de contagens é o Kennedy e colaboradores (2001). Com base nos números, é observada uma super-representação de resíduos carregados negativamente e sub-representação de resíduos carregados positivamente nos homólogos halobacterianos. (KENNEDY et al., 2001). Com base nesses exemplos, podemos considerar que os dados obtidos para as proteínas ribossômicas dos halófilos bacterianos estão relacionados com o perfil padrão das proteínas halofílicas, porém, menos expressivo que o das proteínas ribossômicas arqueanas.

Com base nos resultados desse trabalho e nos relacionados, juntamente com a apresentação dos registros científicos sobre as características das proteínas ribossômicas não halofílicas e do RNA, apontados até o momento, é possível entender as implicações que os mesmos trazem para a interpretação sobre os mecanismos interacionais ribonucleoproteico nos organismos halofílicos. A maior representação de aminoácidos ácidos nas proteínas ribossômicas halófilas, apresentadas aqui e nos proteomas analisados por alguns autores, está relacionada com o ambiente onde essas proteínas estão inseridas, e como isso afeta a solubilidade das mesmas, sendo importante salientar, que uma diminuição na solubilidade proteica afeta de maneira desfavorável a sua funcionalidade.

Com base na literatura, existem dois elementos relacionados com a solubilidade proteica: os extrínsecos: força iônica, composição iônica, pH e temperatura; e os intrínsecos: composição dos aminoácidos (PELEGRINE, 2003). Trevino e colaboradores (2007) investigaram como os determinantes intrínsecos, ou seja, o conjunto de alguns aminoácidos afetam a solubilidade das proteínas. Eles apontaram que o ácido aspártico, o ácido glutâmico e a serina contribuem de maneira mais favorável à solubilidade das proteínas. Além disso, os resultados sugerem que a contribuição da lisina e da arginina para a solubilidade da proteína é complexa e depende drasticamente da carga líquida da proteína. Os resultados também sugerem que o ácido aspártico, o ácido glutâmico e a serina contribuem significativamente mais favoravelmente do que outros aminoácidos hidrofílicos, como asparagina, glutamina e treonina (TREVINO et al., 2007).

Essas remodelações evolutivas nas proteínas halofílicas, incluindo as proteínas ribossômicas, permitem que elas sejam estáveis e solúveis na alta concentração citoplasmática de NaCl / KCl. Nos organismos procariotos, a síntese de todas as proteínas acontece no citoplasma, e como durante a síntese as proteínas não estão na sua forma enovelada, os resíduos hidrofóbicos são expostos a altas concentrações de sal, levando a interações não adequadas de suas cadeias laterais, que podem competir com o arranjo intramolecular adequado, dentro da conformação correta (PAUL et al., 2008). Provavelmente, para minimizar essa possibilidade, todas as proteínas halofílicas solúveis têm um número menor de resíduos hidrofóbicos. O aumento da carga negativa na superfície das proteínas halófilas neutraliza a constante dielétrica mais baixa com alta salinidade e, portanto, proporciona maior solubilidade da proteína (KENNEDY et al., 2001; HUTCHEON et al., 2005; PAUL et al., 2008).

Outro estudo sobre o efeito da carga líquida na solubilidade, atividade e estabilidade proteica foi feito por Shaw e colaboradores (2001), utilizando como modelo a proteína Ribonuclease Sa (RNase Sa), uma proteína ácida com um ponto isoelétrico de 3,5, que não contém resíduos de lisina (K). Neste estudo, eles substituíram os resíduos de ácido aspártico e ácido glutâmico pelos resíduos de lisina isoladamente e em combinação com outros aminoácidos, para produzir variantes com diferentes pontos isoelétricos. Uma colocação importante, é que as proteínas tendem a ser menos solúveis perto do seu ponto isoelétrico. A substituição de três resíduos ácidos por lisinas elevaram o ponto isoelétrico quase à neutralidade e a substituição de cinco resíduos ácidos por lisinas elevou o ponto isoelétrico acima de 10 (SHAW *et al.*, 2001). Ao relacionarmos esse ponto com o pH citosólico das células vivas, próximo a neutralidade ($\sim 7,0$), é possível verificar que a substituição de aminoácidos ácidos por básicos fizeram com que a Ribonuclease Sa (RNase Sa) ficasse menos solúveis. Com isso, podemos inferir que o perfil ácido das proteínas ribossômicas halofílicas (principalmente do domínio *Archaea*) em pH citosólico, será mais solúvel (SHAW *et al.*, 2001).

Existe uma relação entre solubilidade proteica e pH em química de proteínas. A solubilidade de uma proteína é proporcional ao quadrado da carga líquida na proteína. Consequentemente, espera-se que as proteínas sejam menos solúveis perto de seus pontos isoelétricos, e a solubilidade deve aumentar à medida que o pH aumenta ou diminui à medida que a magnitude da carga líquida aumenta (SHAW *et al.*, 2001). Como a tendência é uma diminuição da solubilidade perto dos valores isoelétricos, as proteínas ribossômicas com o perfil mais ácido, terá a tendência de ser mais solúvel em um ambiente rico em sal.

Assim como a composição proteômica, as adaptações halofílicas também estão associadas a uma assinatura genômica específica. Os halófilos obrigatórios geralmente contêm genomas ricos em GC (bem acima de 60%). Atribui-se que um alto conteúdo de GC nos genomas halofílicos ajude a evitar a formação de dímeros de timidina induzida por radiação ultravioleta (UV) e o possível acúmulo de mutações em seu habitat especializado (lagoas costeiras rasas), caracterizado por altos níveis desse tipo de radiação.

A análise feita por Paul e colaboradores (2008) revelou que os genomas de todos os halófilos obrigatórios mostram abundância de dinucleótido, principalmente de CG, GA / CT e CA / GT) definidas em comparação com os não-halófilos. A maior frequência de dinucleotídeos GA, AC e GT, na primeira e na segunda posições do códon, reflete a

exigência de resíduos de ácido aspártico, ácido glutâmico, treonina e valina em sequências de proteínas halófilas.

Os autores apontam ainda os códons preferidos pelos halófilos: Arginina (CGA e CGG), Valina (GUC), Treonina (ACG), Leucina (CUC) e Cisteína (UGU). Os códons GAC e GAG, correspondentes a ácido aspártico e ácido glutâmico, respectivamente, estão super-representados, correspondendo a uma abundância de resíduos ácidos (KENNEDY et al., 2001). Todas essas descobertas apoiam fortemente a noção de evolução convergente, não apenas no nível da composição do proteoma, mas também no nível da organização do genoma dos microrganismos adaptados aos ambientes com alto teor de sal (PAUL et al., 2008).

5.1.3. Comparação das médias e porcentagens das proteínas halofílicas separadas por subunidades (LSU e SSU)

Nos quatro gráficos abaixo (3 e 4) correspondentes às médias e as porcentagens dos aminoácidos das 19 proteínas ribossomais da subunidade maior (LSU); e (5 e 6) correspondentes aos valores médios e das porcentagens das 15 proteínas da subunidade menor (SSU), mostraram similaridade com os gráficos que representavam os dados das 34 proteínas juntas. Com isso, pode-se entender um padrão composicional em todo o ribossomo.

Com base nas porcentagem dos aminoácidos agrupados pelas características das suas cadeias laterais (hidrofóbicos alifáticos, hidrofóbicos aromáticos, negativo, polar não carregado e positivo), foi possível observar nas proteínas separadas por subunidade, o mesmo perfil encontrado nas 34 proteínas analisadas em grupo. Nas Tabelas 8 (das 19 proteínas-LSU) e 9 (15 proteínas-SSU), as porcentagens dos aminoácidos hidrofóbicos (alifáticos e aromáticos) são menores nos organismos halófilos em comparação com os seus controles, porém, não apresentaram significância estatística. O mesmo caso ocorreu com os aminoácidos positivos, a porcentagem dos mesmos é mais baixa nos organismos extremófilos analisados, principalmente no grupo dos halófilos do domínio *Archaea*, porém, não há significância estatística. Em relação aos negativos, apresentaram diferença significativa entre o grupo *Archaea* halófilo e controle, na comparação das 15 proteínas (Tabela 9), mas em relação aos outros grupos a significância não existiu.

COMPARAÇÃO DAS MÉDIAS DOS AMINOÁCIDOS DAS 19 PROTEÍNAS (LSU)

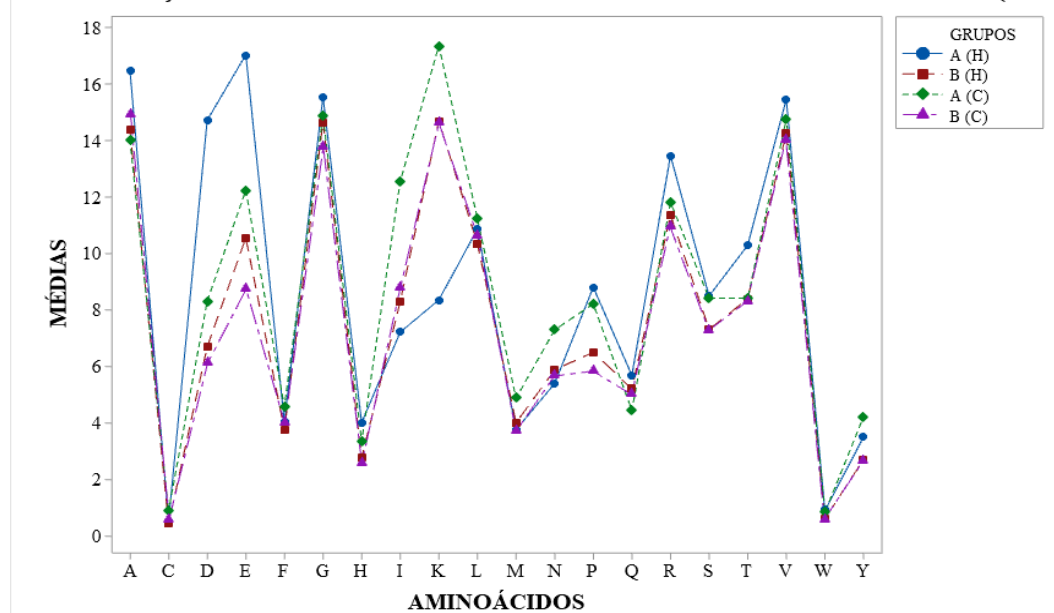
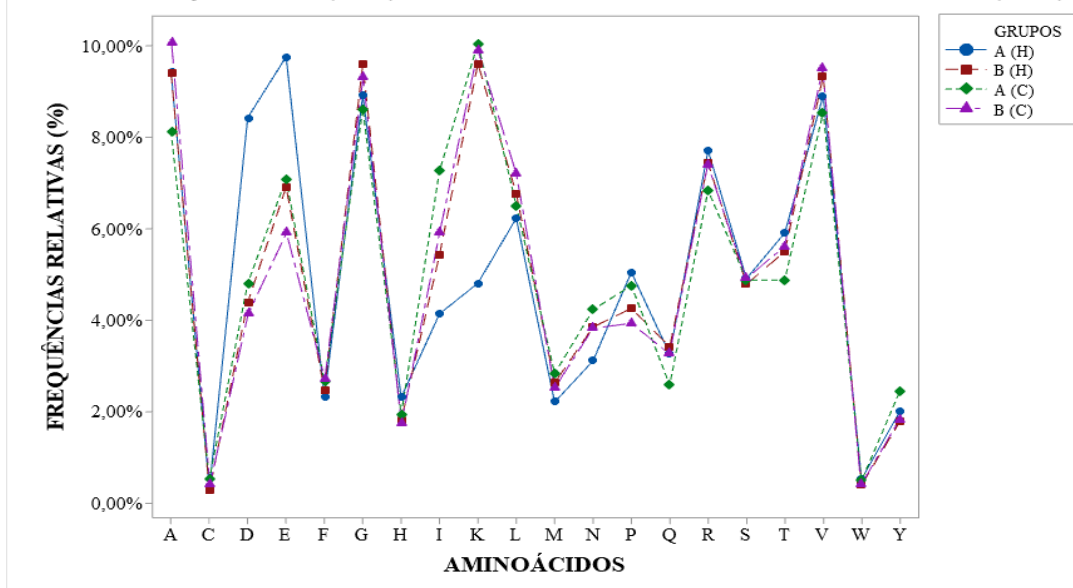


Gráfico 3: Médias dos 20 aminoácidos nas 19 proteínas ribossômicas da subunidade maior (LSU).

COMPARAÇÃO DAS (FR*) DOS AMINOÁCIDOS DAS 19 PROTEÍNAS (LSU)



*Significa Frequência Relativa

Gráfico 4: Frequência relativa (%) dos vinte aminoácidos das 19 proteínas ribossômicas da subunidade maior (LSU).

	<i>Archaea</i> (H)	Controle (A)	p-value	<i>Bacteria</i> (H)	Controle (B)	p-value	p-value** (A+B)
H. Alifático	44,84%	46,60%	0,652874	47,38%	48,39%	0,373661	0,18619
H. Aromático	4,82%	5,56%	0,224354	4,62%	4,88%	0,379897	0,614671
Negativo	18,17%	11,88%	0,09543	11,28%	10,03%	0,344042	0,108879
Polar N.C	17,46%	17,04%	0,832882	17,79%	17,94%	0,614714	0,748057
Positivo	14,80%	18,80%	0,566277	18,83%	19,02%	0,648329	0,518628

Tabela 8: Porcentagem dos aminoácidos das 19 proteínas ribossômicas da subunidade maior (LSU) agrupados pelas características químicas das suas cadeias laterais.

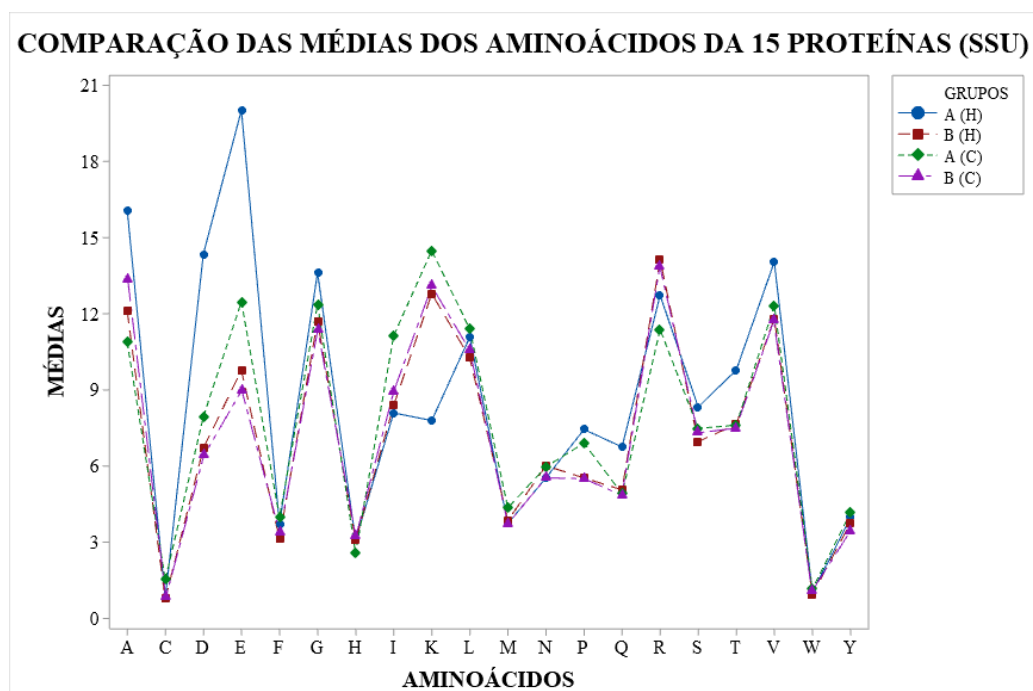
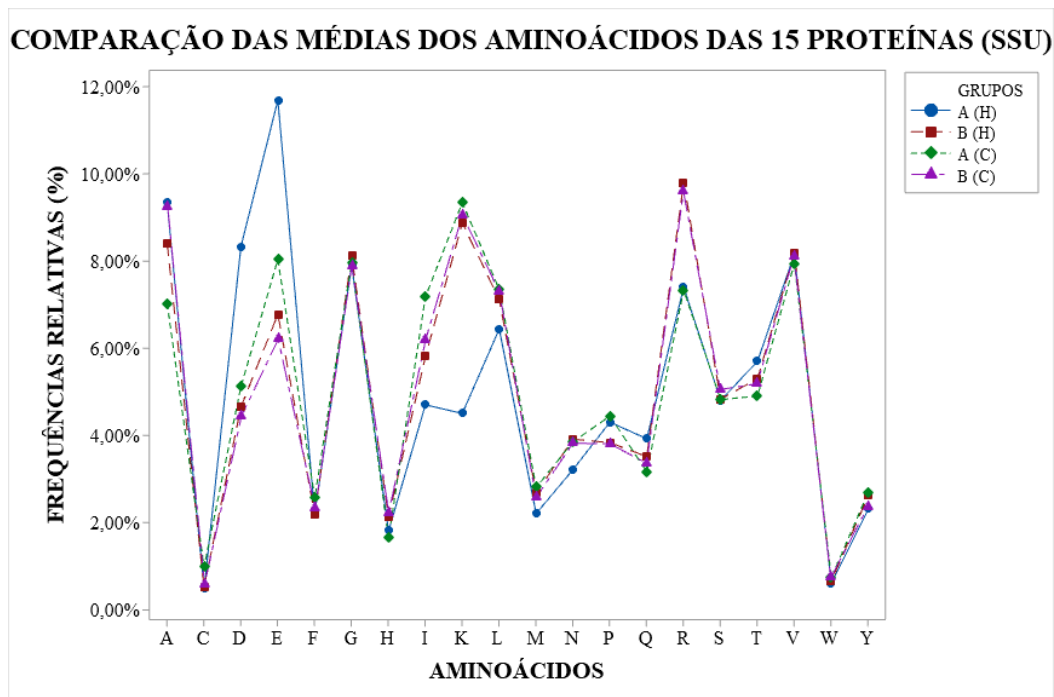


Gráfico 5: Médias dos 20 aminoácidos nas 15 proteínas ribossômicas da subunidade menor (SSU).



*Significa Frequência Relativa

Gráfico 6: Frequência relativa (%) dos vinte aminoácidos das 15 proteínas ribossômicas da subunidade menor (SSU).

	<i>Archaea</i> (H)	Controle (A)	<i>p-value</i>	<i>Bacteria</i> (H)	Controle (B)	<i>p-value</i>	<i>p-value**</i> (A+B)
H. Alifático	43,06%	44,73%	0,675874	44,17%	45,12%	0,370484	0,567900283
H. Aromático	5,09%	5,97%	0,062343	5,43%	5,36%	0,86867	0,376120533
Negativo	20,04%	13,16%	0,040659	11,43%	10,64%	0,265502	0,090334471
Polar N.C	18,09%	17,70%	0,810606	18,06%	17,94%	0,747885	0,977798005
Positivo	13,70%	18,33%	0,449181	20,81%	20,89%	0,840826	0,180561738

Tabela 9: Porcentagem dos aminoácidos das 15 proteínas ribossômicas da subunidade menor (SSU) agrupados pelas características químicas das suas cadeias laterais.

5.1.4. Análise de correlação de Pearson

O resultado da correlação de Pearson foi de uma correlação positiva, e baseado no valor de r , a relação positiva entre os dois grupos de extremófilos é forte, pois o valor se

aproxima de 1; isso implica que as quantidades dos aminoácidos de *Archaea*-halófilos segue a mesma tendência (comportamento) que em *Bacteria*-halófilos (Gráfico 7 e Figura 12). Devido a natureza das proteínas analisadas, consideradas altamente conservadas (YUTIN et al., 2012), uma correlação positiva era esperada. Com base no resultado da correlação, podemos inferir que os dois grupos de proteínas possuem o mesmo perfil adaptativo em relação a composição de aminoácidos.

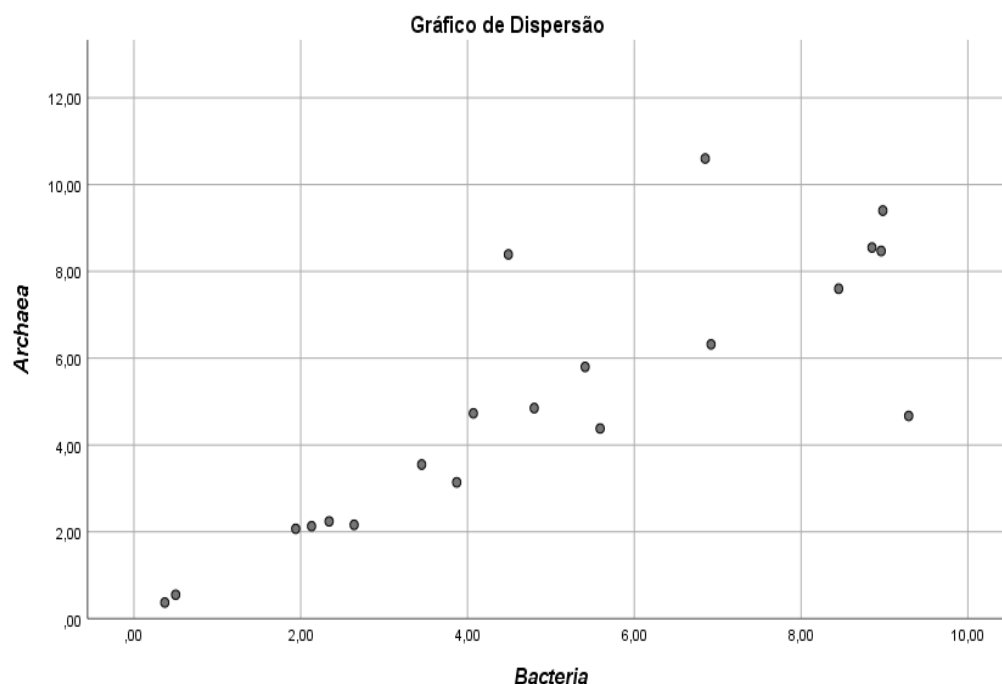


Gráfico 7: Avaliação da dispersão das porcentagens dos 20 aminoácidos das 34 proteínas dos dois grupos de halófilos.

		<i>Archaea</i>	<i>Bacteria</i>
Archaea	Correlação de Pearson	1	,837**
	Sig. (2 extremidades)		,000
	N	20	20
Bacteria	Correlação de Pearson	,837**	1
	Sig. (2 extremidades)	,000	
	N	20	20

**. A correlação é significativa no nível 0,01 (2 extremidades).

Valor do teste de Shapiro-Wilk:

Archaea (Halófilos): sig = 0,393

Bacteria (Halófilos): sig = 0,199

Figura 12: Resultado do teste de correlação de Pearson ($r = 0,837$; $p = 0,0004$).

5.1.5. Tamanho médio das proteínas

Devido aos resultados obtidos em relação a composição dos aminoácidos das proteínas ribossômicas dos organismos extremófilos, em relação aos seus controles, e entre os dois grupos de halófilos, foi calculado o tamanho médio das proteínas dos quatro grupos, com o objetivo de entender se as quantidades diferenciais de alguns aminoácidos são frutos, de uma forma mais acentuada, da substituição de uma classe de aminoácidos por outra (ou seja, troca de positivos por negativos e vice-versa), ou se há um acréscimo de um determinado aminoácido, sem detrimento direto de outro.

Na Tabela 10 abaixo são apresentados os tamanhos médios das 34 proteínas ribossômicas dos grupos analisados: *Archaea*-Halófilos A(H); *Archaea*-Controle A(C); *Bacteria*-Halófilos B(H); e *Bacteria*-Controle B(C).

	uL1	uL2	uL3	uL4	uL5	uL6	uL10	uL11	uL12	uL13	uL14	uL15
A(H)	211,62	239,61	336,79	248,56	175,25	178,2	175,67	158,94	115,04	146,69	132	164,05
B(H)	231,67	275,58	215,32	206,76	181,34	178,01	140,42	143,26	123,2	144,8	121,82	147,49
A(C)	212,26	239,35	335,77	254,44	170,38	178,94	166,8	160,15	103	139,94	132	142,26
B(C)	231,52	276,13	215,56	206,07	180,81	178,15	137,76	142,41	124,82	144,43	122,11	149,45
	uL16	uL18	uL22	uL23	uL24	uL29	uL30	uS2	uS3	uS4	uS5	uS7
A(H)	175,82	183,57	156	83,72	116,23	71,28	154,2	265,07	317,79	174,63	215,42	205,38
B(H)	140,23	119,34	116,73	96,8	105,17	66,79	60,76	257,41	230,43	203,55	173,08	155,85
A(C)	165,41	186,88	152,88	85,85	113,16	67,87	152,7	210,11	264,66	197	215,83	188,16
B(C)	138,6	118,7	119,8	100,55	107,61	68,46	60,69	266,06	240,32	204,53	178,08	156,58
		uS8	uS9	uS10	uS11	uS12	uS13	uS14	uS15	uS17	uS19	
A(H)		130	131,98	107,19	129,28	142	171,22	61,47	156,06	126,52	140,35	
B(H)		130,08	134,31	104,68	129,95	128,72	120,77	86,38	89,35	87,84	91,85	
A(C)		129,88	133,33	102,06	129,06	142,05	155,52	48,41	139,21	107,62	139,83	
B(C)		131,41	135,05	103,12	129,91	128,42	121,39	91,95	89	86,22	91,54	

Tabela 10: Tamanho médio das sequências das 34 proteínas ribossômicas dos quatro grupos comparados.

Foi possível observar que 20 das 34 proteínas analisadas, ou seja 58,8%, apresentaram um tamanho médio maior no domínio *Archaea* (uL1; uL3; uL4; uL10; uL11; uL14; uL15; uL16; uL18; uL22; uL24; uL30; uS3; uS5; uS7; uS12; uS13; uS15; uS17; e uS19). Dentre essas proteínas, 4 apresentaram comportamento especial (as marcadas em vermelho):

- uL15 apresenta um tamanho médio maior significativo apenas no grupo das archaeas halófilas, nos outros três grupos, o tamanho médio é de 140 aminoácidos;
- uL30 possui um tamanho médio duas vezes maior no domínio arqueal (A-H e A-C) do que nas espécies bacterianas;
- uS2 apresenta-se maior nas arqueas halófilas, em comparação ao seu controle
- uS3 possui o tamanho médio de 317,79 aminoácidos no grupo A(H), que é 16,71% maior do que a média do grupo A(C), 27,48% maior que a média do grupo B(H), e 24,37% maior do que a médias do grupo B(C).

Em relação as quatorze restantes, seis proteínas (17,6%) apresentaram um tamanho médio maior nos grupos das bactérias (B(H) e B(C)): uL2; uL5; uL12; uL23; uS4; e uS14. E oito (23,5%) apresentaram tamanhos muito próximos entre os 4 grupos: uL6; uL13; uL29; uS2; uS8; uS9; uS10; e uS11.

Na Tabela 11 abaixo estão os valores referentes a avaliação das 34 proteínas juntas e separadas por subunidades. As avaliações gerais dos tamanhos médios das proteínas ribossômicas apontam para o mesmo perfil descrito acima, ou seja, os organismos halófilos, de ambos os domínios, possuem proteínas, em média, maiores do que os seus grupos controles, sendo a única exceção a comparação entre os resultados das proteínas das subunidade menor, entre os halófilos bacterianos e seu controle, onde o controle se apresenta com uma média de 144,3 aminoácidos, que é 0,3% maior que o grupo halófilico.

	<i>Archaea</i> (halófilos)	<i>Bacteria</i> (halófilos)	<i>Archaea</i> (controle)	<i>Bacteria</i> (controle)
34 Proteínas	173,5	149	164,5	146,8
19 PROTEÍNAS (LSU)	174,8	153	172,9	148,7
15 PROTEÍNAS (SSU)	171,9	143,9	154,8	144,3

Tabela 11: Tamanho médio das proteínas através da comparação geral das 34 proteínas e por subunidade dos grupos de halófilos e seus controles.

Em relação a comparação entre os dois grupos de extremófilos, foi possível observar que os halófilos do domínio *Archaea* possuem proteínas ribossômicas maiores que o grupo do domínio *Bacteria* nas três avaliações, com as seguintes diferenças em porcentagem: 14,12% quando avaliadas as 34 proteínas juntas; 12,5% quando comparadas as médias das 19 proteínas da subunidade maior; e 16,3 % com relação as 15 proteínas da subunidade menor.

Tiessen e colaboradores (2012) apontam que os tamanhos das sequências das proteínas são relevantes para a estrutura bioquímica e para a sua função biológica, e que a distribuição estatística dos comprimentos de proteínas pode fornecer informações valiosas sobre a evolução dos proteomas estudados. A partir de análises comparativas, foi verificado que há diferenças significativas entre o tamanho médio das proteínas dos três domínios da vida, onde o comprimento médio das proteínas arqueais (283 aminoácidos) foi o menor, seguido pelo comprimento médio das proteínas bacterianas (320) e eucarióticas (472) (TIESEN et al. 2012).

O resultado apresentado no presente trabalho, referente ao tamanho médio das sequências é contrário ao apresentado por Tiessen e colaboradores (2012), porém, tratando-se das proteínas ribossômicas, uma sequência média maior apresentada pelos grupos arqueais pode ser baseada na complexidade de todo o seu sistema de tradução, que é consideravelmente próximo ao eucarionte (ALBERTS, 2017; MADIGAN et al., 2016). O próprio número de proteínas que formam o ribossomo no domínio *Archaea* (68 proteínas) é maior que o apresentado pelo domínio *Bacteria* (57 proteínas) (NIKOLAY et al., 2015).

Entretanto, quando avaliamos a diferença do tamanho médio das proteínas halofílicas em relação ao seu controle, ou seja, dentro do mesmo domínio, percebe-se que as proteínas dos halófilos são maiores. Com base nisso, pode-se inferir um ganho de aminoácidos, principalmente com carga negativa nessas proteínas, que ao invés de substituição direta, pode

proporcionar tanto a conservação dos domínios específicos dessas proteínas, como a capacidade de adaptação aos ambientes inóspitos onde estão inseridas.

Apoiando essa perspectiva, três estudos anteriores relataram uma relação positiva e negativa entre o tamanho médio das proteínas com o conteúdo de alguns aminoácidos. White (1992) relatou que existe uma dependência das frequências de aminoácidos em relação ao comprimento da sequência. Usando um conjunto não redundante de 2.275 proteínas, descobriu que a frequência de cisteína (C) aumenta dramaticamente para sequências menores que 100 aminoácidos. Também relatou, sendo esse ponto o mais importante para os resultados dessa pesquisa, que as frequências de arginina (R) e lisina (K) aumentam para sequências curtas, enquanto as frequências de ácido aspártico (D) e ácido glutâmico (E) diminuem.

Tiessen e colaboradores (2012) baseados em mais de 1,2 milhões de proteínas, também observaram que o comprimento da proteína correlaciona-se negativamente com as porcentagens de cisteína, metionina, lisina, fenilalanina, arginina, triptofano, tirosina, e positivamente com as porcentagens de ácido aspártico, ácido glutâmico, glutamina, serina e treonina. Kiraga e colaboradores (2007) restringe os resultados aos aminoácidos carregados e também encontrou uma correlação negativa dos aminoácidos básicos (lisina e arginina) e positiva com os aminoácidos ácidos (ácido aspártico e ácido glutâmico) em relação a tamanho da cadeia polipeptídica.

6. CONSIDERAÇÕES FINAIS

Os resultados obtidos com essa pesquisa proporcionaram responder o questionamento inicial do trabalho, que girou em torno das adaptações das proteínas ribossômicas dos organismos halofílicos dos domínios *Archaea* e *Bacteria*. Configuraram-se subsídios para inferir que as proteínas ribossômicas halofílicas possuem características que as diferenciam do padrão mesofílico.

A composição das sequências primárias das proteínas ribossômicas halofílicas se mostraram mais ácidas, menos hidrofóbicas e menos positivas que os controles, padrão que é descrito para outros elementos proteômicos não ribossomais dos organismos halofílicos.

Porém, destaca-se a avaliação conjunta com o aspecto funcional das proteínas ribossômicas, onde um dos papéis biológicos descritos é a neutralização das cargas negativas do RNA, que age em prol da estabilidade do complexo ribonucleoprotéico.

Nossos resultados corroboram com os dados que apontam que o domínio *Archaea* possui, principalmente, organismos ligados à adaptação com altos níveis de sal citosólico, repercutindo no seu padrão proteico. Os halófilos bacterianos, com base nos nossos resultados, estão mais relacionados à adaptação que não apresenta assinaturas tão marcantes nas suas proteínas (produção de osmólitos), o que explica a proximidade desse grupo com o seu controle.

De uma forma geral, podemos concluir que uma abordagem com base em análises estatísticas de sequências primárias de proteínas podem fornecer resultados relevantes no entendimento de padrões adaptativos de organismos extremófilos que, posteriormente, podem ser melhor desenvolvidos através de outras abordagens.

7. PERSPECTIVAS FUTURAS

Para o desenvolvimento, como continuação da pesquisa aqui apresentada, em trabalhos futuros, pretende-se:

- Verificar a significância dos dados individuais das 34 proteínas (em anexo) para promover estudos sobre considerações individuais das mesmas;
- Analisar estruturalmente as proteínas ribossômicas através de software de interação de resíduos de aminoácidos e modelagem por homologia;
- Verificar a significância das médias do tamanho médio das proteínas;
- Fazer análises de correlação entre o tamanho médio das proteínas e o uso dos aminoácidos;
- Verificar as frequências do uso de códons das proteínas ribossômicas dos grupos halófilos dos dois domínios e comparar com o padrão genômico.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBERTS, B. et al. **Biologia Molecular da Célula**. 5. ed. Porto Alegre: Editora Artmed, 2010.
- ALBERTS, B. et al. **Biologia Molecular da Célula**. 6. ed. Porto Alegre: Editora Artmed, 2017.
- ALMEIDA, A.; ELIAN, S.; NOBRE, J. **Modificações e alternativas aos testes de Levene e de Brown e Forsythe para igualdade de variâncias e médias**. Revista Colombiana de Estadística, volume 31, no. 2, pp. 241 a 260, 2008.
- BARROS, E. & MAZUCHELI, J. **Um estudo sobre o tamanho e poder dos testes t-Student e Wilcoxon** = A study about the size and power of t-Student and Wilcoxon tests. Acta Scientiarum : Language and Culture, 27, 2005.
- BAXTER, B. K. **Great Salt Lake microbiology: a historical perspective**. *International microbiology : the official journal of the Spanish Society for Microbiology* vol. 21,3: 79-95, 2018.
- BOLHUIS, A., KWAN, D., & THOMAS, J. R. **Halophilic adaptations of proteins**. In K. S. Siddiqui, T. Thomas, & V. Uversky (Eds.), *Protein Adaptation in Extremophiles: Design, Selection and Applications* (pp. 71-104). (Molecular Anatomy and Physiology of Proteins). New York, U. S. A.: Nova Science Publishers, 2008.
- BOLHUIS, A., KWAN, D., THOMAS, J.R. **Halophilic adaptations of proteins** (Capítulo de Livro). Ribosomal Proteins and Protein Engineering: Design, Selection and Applications, Páginas: 143-174, 2010.
- BREDA, A.; VALADARES, N. F.; GARRATT, O. N. **Protein Structure, Modelling and Applications**. Cap. 6. In: GRUBER, A. et al. *Bioinformatics in Tropical Disease Research A Practical and Case-Study Approach* Editores. Bethesda (MD): National Biotechnology Information Center (EUA) ; 2008.
- BROCK, T. D. **Life at High Temperatures: Evolutionary, ecological, and biochemical significance of organisms living in hot springs is discussed**. *Science*, 158(3804), 1012–1019, 1967.
- BURTON, B.; ZIMMERMANN, M. T.; JERNIGAN, R. L.; WANG, Y. **A Computational Investigation on the Connection between Dynamics Properties of Ribosomal Proteins and Ribosome Assembly**. PLOS Computational Biology 8(5): e1002530, 2012.
- CAVICCHIOLI, R. **Cold-adapted archaea**. *Nat Rev Microbiol* 4, 331–343, 2006.
- CORREIA, S. **Potencial hídrico**. Rev. Ciência Elem., V2(1):003, 2014.

DASSARMA, S., & DASSARMA, P. **Halophiles and their enzymes: negativity put to good use.** *Current opinion in microbiology*, 25, 120–126, 2015.

DUARTE, R.T.D.; RIBEIRO, C. G. & PELLIZARI, V. H. **Vida ao Extremo: A magnífica versatilidade da vida microbiana em ambientes extremos da Terra.** Capítulo 8. Astrobiologia [livro eletrônico] : uma ciência emergente / Núcleo de Pesquisa em Astrobiologia. -- São Paulo : Tikinet Edição : IAG/USP, 2016. 10 Mb ; ePUB e PDF.

DUMORNÉ, K.; CÓRDOVA, D.C.; ASTORGA-ELÓ, M.; & RENGANATHAN, P. **Extremozymes: A Potential Source for Industrial Applications.** *Journal of microbiology and biotechnology*, 27(4), 649–659, 2017.

EDBEIB, M. F. et al. **Halophiles: biology, adaptation, and their role in decontamination of hypersaline environments.** *World journal of microbiology & biotechnology* vol. 32,8: 135, 2016.

FARIAS, S. T.; & BONATO, M. C. **Preferred amino acids and thermostability.** *Genetics and Molecular Research : GMR*. 2003.

FARIAS, S.; & PROSDOCIMI, F. **A emergência dos sistemas biológicos: uma visão molecular sobre a origem da vida.** 2019.

FEDYUKINA, D. V. et al. **Charge segregation and low hydrophobicity are key features of ribosomal proteins from different organisms.** *The Journal of biological chemistry* vol. 289,10: 6740-6750, 2014.

FREDRICK, K.; & MICHAEL, I. **Protein synthesis: Errors rectified in retrospect.** *Nature* vol. 457,7226: 157-8, 2009.

HUTCHEON, G. W.; VASISHT, N.; BOLHUIS, A. **Characterisation of a highly stable alpha-amylase from the halophilic archaeon *Haloarcula hispanica*.** *Extremophiles*; 9(6):487–495, 2005.

JOO, W. A.; & KIM, C. W. **Proteomics of Halophilic archaea.** *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 815(1-2), 237–250, 2005.

KENNEDY, S. P.; NG, W. V.; SALZBERG, S. L.; HOOD, L., & DASSARMA, S. **Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence.** *Genome research*, 11(10), 1641–1650, 2001.

KIRAGA, J. et al. **The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms.** *BMC Genomics*. 8: 8-10.1186/1471-2164-8-8, 2007.

KLEIN, D. J.; MOORE, P. B.; STEITZ, T. A. **The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit.** *J Mol Biol*. 2004;340(1):141–177. doi:10.1016/j.jmb. 2004. 03.076

KRAEGELOH, A.; KUNTE, H. J. **Novel insights into the role of potassium for osmoregulation in *Halomonas elongata*.** *Extremophiles*, v. 6, p. 453–462, 2002.

KUSHNER, D. J. **Life in high salt and solute concentrations: halophilic bacteria.** In: KUSHNER, D. J., editor. *Microbial life in extreme environments*. London, United Kingdom: Academic Press, Ltd.; pp. 317–368, 1978.

LECOMPTE, O. et al. **Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale.** *Nucleic acids research* vol. 30,24, 2002.

LESK, A. **Introdução à Bioinformática.** Tradução da 1a edição. Editora Artmed, Porto Alegre, Brasil, 2001.

LESK, A. **Introdução à Bioinformática.** Tradução da 2a edição. Editora Artmed, Porto Alegre, Brasil, 2008.

LOTT, B. B.; WANG, Y.; & NAKAZATO, T. **A comparative study of ribosomal proteins: linkage between amino acid distribution and ribosomal assembly.** *BMC Biophys* 6, 13, 2013.

MA, Y.; GALINSKI, E. A.; GRANT, W. D.; OREN, A.; & VENTOSA, A. **Halophiles: life in saline environments.** *Applied and environmental microbiology*, 76(21), 6971–6981, 2010.

MADIGAN, M. T. et al. **Microbiologia de Brock.** 14. ed. Porto Alegre: ArtMed, 2016.

MADIGAN, M. T.; CLARK, D. P.; STAHL, D.; & MARTINKO, J. M. **Em Brock Biology of Microorganisms.** Ch. 2, 86–111. Benjamin Cummings Press, 2010.

MALLIK, S.; KUNDU, S. **A comparison of structural and evolutionary attributes of *Escherichia coli* and *Thermus thermophilus* small ribosomal subunits: signatures of thermal adaptation.** *PLoS One*. 2013;8(8):e69898. Published: 2013.

MELNIKOV, S.; MANAKONGTREETCHEEP, K.; SÖLL, D. **Revising the Structural Diversity of Ribosomal Proteins Across the Three Domains of Life.** *Molecular Biology and Evolution*, Volume 35, Issue 7, Pages 1588–1598, July, 2018.

MERINO, N. et al. **Living at the Extremes: Extremophiles and the Limits of Life in a Planetary Context.** *Frontiers in microbiology* vol. 10 780. 15 Apr. 2019.

MEVARECH, M.; FROLOW, F.; & GLOSS, L. M. **Halophilic enzymes: proteins with a grain of salt.** *Biophysical chemistry*, 86(2-3), 155–164, 2000.

MIOT, H. A. **Avaliação da normalidade dos dados em estudos clínicos e experimentais.** *Jornal vascular brasileiro* vol. 16,2: 88-91, 2017.

NELSON, D. L.; COX, M. M. **Princípios de bioquímica de Lehninger.** Porto Alegre: Artmed, 2011. 6. ed. Porto Alegre: Artmed, 2014.

NIKOLAY, R.; VAN DEN BRUCK, D.; ACHENBACH, J.; & NIERHAUS, K.H. **Ribosomal Proteins: Role in Ribosomal Functions**. In eLS, John Wiley & Sons, Ltd (Ed.), 2015.

OLIVEIRA, M. F. **Metodologia científica: um manual para a realização de pesquisas em Administração / Maxwell Ferreira de Oliveira**. -- Catalão: UFG (2011). 72 p.: il.

OREN, A. **Microbiological studies in the Dead Sea: future challenges toward the understanding of life at the limit of salt concentrations**. *Hydrobiologia* 405, 1–9, 1999.

PAUL, S.; BAG, S. K.; DAS, S.; HARVILL, E. T.; & DUTTA, C. **Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes**. *Genome biology*, 9(4), R70, 2008.

PELEGRINE, D. H. G.; & GASPARETTO, C. A. **Estudo da solubilidade das proteínas da clara do ovo em função da temperatura e do pH**. *Revista Ciências Exatas e Naturais*, Vol. 5, no 2, Jul/Dez, 2003.

RAMAKRISHNAN, V. **Ribosome structure and the mechanism of translation**. *Cell* vol. 108,4: 557-72, 2002.

RAMPELOTTO, P. H. **Extremophiles and Extreme Environments**. *Life* 3, no. 3: 482-485, 2013.

REED, C. J.; LEWIS, H.; TREJO, E.; WINSTON, V.; & EVILIA, C. **Protein adaptations in archaeal extremophiles**. *Archaea (Vancouver, B.C.)*, 373275, 2013.

RODRIGUES, F.; GALANTE, D.; & AVELLAR, M. G. B. **Astrobiologia: Estudando a vida no Universo**. Capítulo 1. Astrobiologia [livro eletrônico] : uma ciência emergente / Núcleo de Pesquisa em Astrobiologia. -- São Paulo : Tikinet Edição : IAG/USP, 2016. 10 Mb ; ePUB e PDF

SAUM, S.H.; MÜLLER, V. **Regulation of osmoadaptation in the moderate halophile *Halobacillus halophilus*: chloride, glutamate and switching osmolyte strategies**. *Aquat. Biosyst.* 4, 4, 2008.

SHAW, K. L. et al. **The effect of net charge on the solubility, activity, and stability of ribonuclease Sa**. *Protein science : a publication of the Protein Society* vol. 10,6: 1206-15, 2001.

SIGLIOCCOLO, A.; PAIARDINI, A.; PISCITELLI, M. et al. **Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface**. *BMC Struct Biol* 11, 50, 2011.

SILVA, J. N. **Uma abordagem histórica e experimental da Eletrostática**. *Estação Científica (UNIFAP)*, [S.l.], v. 1, n. 1, p. 99-113, mar. 2011. ISSN 2179-1902.

SMITH, T. F. et al. **The origin and evolution of the ribosome**. *Biology Direct* 3:16. 2008.

STALEY, J. T. **Domain Cell Theory supports the independent evolution of the Eukarya, Bacteria and Archaea and the Nuclear Compartment Commonality hypothesis.** *Open Biol.* 7: 170041. 2017.

TIESSEN, A.; PÉREZ-RODRÍGUEZ, P.; & DELAYE-ARREDONDO, L.J. **Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes.** *BMC Res Notes* 5, 85, 2012.

TREVINO, S. R.; SCHOLTZ, J. M.; & PACE, C. N. **Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa.** *Journal of molecular biology*, 366(2), 449–460, 2007.

VAN DER LINDEN, M. G.; DE FARIAS, S. T. **Correlation between use of codes and thermostability.** *Extremophiles* 10, 479–481, 2006.

VENTOSA, A. et al. **Biology of moderately halophilic aerobic bacteria.** *Microbiology and molecular biology reviews* : MMBR vol. 62,2, 504-44, 1998.

WEI, X.; JIANG, Y.; CHEN, X. et al. ***Amycolatopsis flava* sp. nov., a halophilic actinomycete isolated from Dead Sea.** *Antonie van Leeuwenhoek* 108, 879–885, 2015.

WHITE, S. H. **The amino acid preferences of small proteins: implications for protein stability and evolution.** *J Mol Biol.* 227 (4): 991-995, 1992.

WILSON, D. N. et al. **The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning.** *Proceedings of the National Academy of Sciences of the United States of America* vol. 105,36: 13339-44, 2008.

WOESE, C. R; KANDLER, O.; WHEELIS, M. L. **Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eukarya.** *Proceedings of the National Academy of Sciences* 87:4576-4579, 1990.

YUTIN, N.; PUIGBÒ, P.; KOONIN, E. V.; & WOLF, Y. I. **Phylogenomics of prokaryotic ribosomal proteins.** *PloS one*, 7(5), e36972, 2012.

ZHENG, M. **An introduction to saline lakes on the Qinghai-Tibet plateau.** 1st ed. Dordrecht: Kluwer Academic Publisher; 1997.

