



UNIVERSIDADE
ESTADUAL de LONDRINA

LARISSA PASSOS

**MODELAGEM MATEMÁTICA E SIMULAÇÕES
COMPUTACIONAIS DA PROBABILIDADE DE
FORMAÇÃO DE PEPTÍDEOS SOB CONDIÇÕES
PREBIÓTICAS**

LARISSA PASSOS

**MODELAGEM MATEMÁTICA E SIMULAÇÕES
COMPUTACIONAIS DA PROBABILIDADE DE
FORMAÇÃO DE PEPTÍDEOS SOB CONDIÇÕES
PREBIÓTICAS**

Dissertação de mestrado apresentada ao PGMAC da Universidade Estadual de Londrina, como requisito parcial para a obtenção do Título de MESTRE em Matemática Aplicada e Computacional.

Orientador: Prof. Dr. Paulo Laerte Natti

Coorientador: Prof. Dr. Dimas A. Morozin Zaia

Londrina
2026

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina**

Dados Internacionais de Catalogação -na-Publicação (CIP)

S232c	Passos, Larissa. Modelagem matemática e simulações computacionais da probabilidade de formação de peptídeos sob condições prebióticas / Larissa Passos. – Londrina, 2026. 75 f. : il. Orientador: Paulo Laerte Natti. Coorientador: Dimas Augusto Morozin Zaia. Dissertação (Mestrado em Matemática Aplicada e Computacional) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Matemática Aplicada e Computacional, 2026. Inclui Bibliografia. 1. Modelagem matemática - Teses. 2. Química prebiótica - Teses. 3. Formação de peptídeos - Teses. 4. Simulações computacionais - Teses. 5. Origem da vida - Teses. I. Natti, Paulo Laerte. II Morozin Zaia, Dimas Augusto. III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Matemática Aplicada e Computacional. IV. Título. 519.681-7
-------	--

LARISSA PASSOS

**MODELAGEM MATEMÁTICA E SIMULAÇÕES
COMPUTACIONAIS DA PROBABILIDADE DE
FORMAÇÃO DE PEPTÍDEOS SOB CONDIÇÕES
PREBIÓTICAS**

Dissertação de mestrado apresentada ao PGMAC da Universidade Estadual de Londrina, como requisito parcial para a obtenção do Título de MESTRE em Matemática Aplicada e Computacional.

BANCA EXAMINADORA

Prof. Dra. Nancy Lopes Garcia
Universidade Estadual de Campinas

Prof. Dr. Jorge Alberto Quillfeldt
Universidade Federal do Rio Grande do Sul

Prof. Dr. Tiago Vinicius Alves
Universidade Federal da Bahia

Prof. Dr. Dimas Augusto Morozin Zaia
Universidade Estadual de Londrina

Prof. Dr. Paulo Laerte Natti
Universidade Estadual de Londrina

Londrina, 30 de janeiro de 2026.

Dedico este trabalho aos meus avós.

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Paulo Laerte Natti, pela orientação ao longo de todo o desenvolvimento deste trabalho.

Ao meu coorientador, Prof. Dr. Dimas Augusto Morozin Zaia, pela sugestão do tema, bem como pelas correções, sugestões e contribuições relevantes prestadas durante todo o processo de elaboração desta pesquisa.

Ao Prof. Dr. Eliandro Rodrigues Cirilo, pela colaboração na elaboração do código numérico utilizado nas simulações.

Ao Programa de Pós-Graduação em Matemática Aplicada e Computacional (PGMAC) da UEL.

À minha família, pelo apoio e incentivo ao longo de toda a trajetória acadêmica, especialmente nos momentos mais desafiadores.

Ao meu namorado, Mario, pelo companheirismo, apoio e incentivo durante todo o período de desenvolvimento deste trabalho.

Às minhas amigas Maria Graziela e Nathalia, pela amizade, apoio e pelas palavras de incentivo que tornaram essa caminhada mais leve.

À CAPES, pelo apoio financeiro essencial à realização deste trabalho.

PASSOS, Larissa. **Modelagem matemática e simulações computacionais da probabilidade de formação de peptídeos sob condições prebióticas**. 2026. 76 páginas. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Universidade Estadual de Londrina, Londrina, 2026.

RESUMO

Como um conjunto de aminoácidos foram selecionados para formar as proteínas essenciais para a vida? Esse questionamento incentivou pesquisadores ao longo dos anos a realizarem experimentos em busca de respostas. Esse trabalho apresenta a formulação e a aplicação de um modelo matemático para calcular a probabilidade de formação de peptídeos e proteínas. A elaboração do modelo foi baseada nos dados experimentais da abundância relativa dos aminoácidos em dois cenários: um representando a Terra primitiva e outro representando um ambiente extra-terrestre. Também foram utilizados dados experimentais do rendimento das ligações peptídicas em ambiente simulado terrestre. As simulações numéricas, considerando os 4 aminoácidos GADV (Glicina — G, Alanina — A, Ácido Aspártico — D e Valina — V). Posteriormente também foi incluída a Leucina, implementadas no software GNU Octave, mostraram que apenas algumas sequências de aminoácidos apresentam alta probabilidade de formação, enquanto a grande maioria das sequência de aminoácidos possíveis tem probabilidade extremamente baixa. Discutimos se esse padrão estatístico pode estar relacionado à seleção natural que ocorreu antecedendo a vida. Em nossas simulações numéricas esse resultado é verificado tanto para o cenário em que os aminoácidos foram formados na Terra primitiva como no cenário em que foram trazidos por meteoritos.

Palavras-chave: Formação de peptídeos. Hipóteses GADV. Modelagem matemática. Química prebiótica. Simulações computacionais.

PASSOS, Larissa. **Mathematical Modeling and Computational Simulations of the Probability of Peptide Formation under Prebiotic Conditions**. 2026. 76 páginas. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Universidade Estadual de Londrina, Londrina, 2026.

ABSTRACT

How was a set of amino acids selected to form the proteins essential for life? This question has encouraged researchers over the years to conduct experiments in search of an answer to this question. This work presents the formulation and application of a mathematical model to calculate the probability of peptide and protein formation. The development of the model was based on experimental data of the relative abundance of amino acids in two scenarios: one representing the primitive Earth and the other representing an extraterrestrial environment. Experimental data on the yield of peptide bond formation in a simulated terrestrial environment were also used. The numerical simulations, considering the four amino acids GADV (Glycine — G, Alanine — A, Aspartic Acid — D, and Valine — V). Later, leucine was also included and implemented in the GNU Octave software, showed that only a few amino acid sequences have a high probability of formation, while the vast majority of possible amino acid sequences have an extremely low probability. We discuss whether this statistical pattern may be related to natural selection that occurred before life. In our numerical simulations, this result is observed for both the scenario in which amino acids were formed on primitive Earth and for the scenario in which they were brought by meteorites.

Keywords: Computational simulations. GADV-protein hypothesis. Mathematical modeling. Peptide formation. Prebiotic chemistry.

SUMÁRIO

1	INTRODUÇÃO	17
2	Fundamentação Teórica	19
2.1	QUÍMICA PREBIÓTICA E O AMBIENTE DA TERRA PRIMITIVA	19
2.1.1	Experimentos históricos e hipóteses sobre a origem da vida	19
2.1.2	Ambientes favoráveis para as reações químicas	23
2.2	FORMAÇÃO DE PEPTÍDEOS	24
3	Metodologia	26
3.1	REVISÃO BIBLIOGRÁFICA DE DADOS EXPERIMENTAIS	26
3.2	NORMALIZAÇÃO DOS DADOS	28
3.2.1	Normalização dos rendimentos de formação de dipeptídeos	28
3.2.2	Normalização da abundância dos aminoácidos	29
3.3	LINHAS DE TENDÊNCIA	30
4	O modelo matemático	32
4.1	O MODELO DE PROBABILIDADE DE FORMAÇÃO DE PEPTÍDEOS	32
4.2	CONDIÇÃO DE CONTORNO PARA CADEIA DE <i>n</i> -AMINOÁCIDOS	32
4.3	MODELO COMPUTACIONAL	33
5	Simulações Numéricas e Discussão	36
5.1	SOBRE O RENDIMENTO DA LIGAÇÃO VAL-ASP	36
5.2	RESULTADOS COM 4 AMINOÁCIDOS: ABUNDÂNCIA TERRESTRE	37
5.3	RESULTADOS COM 4 AMINOÁCIDOS: ABUNDÂNCIA EXTRATERRESTRE	46
5.4	RESULTADOS COM 5 AMINOÁCIDOS: INCLUSÃO DA LEUCINA	56
5.5	ANÁLISE COMPARATIVA DAS RAZÕES DE RELEVÂNCIA E CORTES DE PROBABILIDADE	65
6	Conclusão	68
A	Apêndice A: Aplicação do Modelo Probabilístico para a Formação de Dipeptídeos e Tripeptídeos	70
	REFERÊNCIAS	72

LISTA DE FIGURAS

2.1	Descrição do experimento de Francesco Redi.	20
2.2	Aparelho utilizado por Miller para o experimento. Fonte: Adaptado de [24] . .	23
2.3	Coleta da mistura resultante no experimento de Miller em cromatografia em papel. Fonte: Adaptado de [13]	23
2.4	Esquema comumente utilizado no estudo de ciclos de hidratação e desidratação. Fonte: Adaptado de [29]	25
3.1	Estruturas resultantes da formação de dipeptídeos Gly–Ala e Ala–Gly. Fonte: Autora	27
4.1	Fluxograma do algoritmo de simulação computacional da probabilidade de formação de peptídeos. Fonte: Autora.	35
5.1	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 95%. Fonte: Autora	39
5.2	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 90 %. Fonte: Autora	40
5.3	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 80 %. Fonte: Autora	41
5.4	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 70 %. Fonte: Autora	42
5.5	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 60 %. Fonte: Autora	43
5.6	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 50 %. Fonte: Autora	45
5.7	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 95%. Fonte: Autora	48

5.8	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 90%. Fonte: Autora	49
5.9	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 80%. Fonte: Autora	51
5.10	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 70%. Fonte: Autora	52
5.11	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 60%. Fonte: Autora	53
5.12	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 50%. Fonte: Autora	55
5.13	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 95%. Fonte: Autora	58
5.14	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 90%. Fonte: Autora	59
5.15	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 80%. Fonte: Autora	60
5.16	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 70%. Fonte: Autora	61
5.17	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 60%. Fonte: Autora	63
5.18	Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 50%. Fonte: Autora	64
5.19	Evolução do número de peptídeos relevantes, considerando ambiente terrestre e os 4 aminoácidos GADV. Fonte: Autora	67

LISTA DE TABELAS

2.1	Etapas da formação de vida a partir de moléculas simples segundo a hipótese Oparin-Haldane.	21
3.1	Rendimento em % de dipeptídeos <i>AB</i> formados a partir da massa original de aminoácidos após cinco ciclos.	27
3.2	Abundância relativa dos aminoácidos GADV em fontes terrestres e extraterrestres (glicina = 1).	28
3.3	Valores normalizados de rendimento de dipeptídeos <i>AB</i> formados a partir da massa original de aminoácidos, após o quinto ciclo. Valores normalizados em relação ao rendimento do dipeptídeo Gly-Gly.	29
3.4	Abundâncias relativas normalizadas dos aminoácidos GADV em fontes terrestres e extraterrestres	29
5.1	Valores de rendimento em % de dipeptídeos <i>AB</i> formados a partir da massa original de aminoácidos, após o quinto ciclo. Valores normalizados em relação ao rendimento do dipeptídeo Gly-Gly.	37
5.2	Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%. . . .	38
5.3	Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 90%. . . .	40
5.4	Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 80%. . . .	41
5.5	Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 70%. . . .	42
5.6	Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 60%. . . .	43
5.7	Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 50%. . . .	44
5.8	Razão de relevância para a formação de cadeias de 100-aminoácidos (proteínas) em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%.	46
5.9	Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%. . . .	47
5.10	Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 90%. . . .	49

5.11 Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 80%.	50
5.12 Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 70%.	52
5.13 Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 60%.	53
5.14 Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 50%.	54
5.15 Razão de relevância para a formação de cadeias de 100-aminoácidos (proteínas) em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%.	56
5.16 Valores normalizados do rendimento em % de dipeptídeos AB formados a partir da massa original de 5 aminoácidos após o quinto ciclo	57
5.17 Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 95%.	57
5.18 Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 90%.	59
5.19 Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 80%.	60
5.20 Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 70%.	61
5.21 Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 60%.	62
5.22 Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 50%.	63
5.23 Razão de relevância para a formação de cadeias de 100-aminoácidos (proteínas) em ambiente terrestre, com aminoácidos, considerando corte de probabilidade em 95%.	65
5.24 Comparação das razões de relevância (ρ) nos três cenários com corte de probabilidade de 95%.	66
5.25 Evolução do número de cadeias peptídicas relevantes em função do corte percentual, considerando ambiente terrestre e os 4 aminoácidos GADV	66
A.1 Cálculo das probabilidades P_{AB} para combinações de 2 aminoácidos	70
A.2 Cálculo das probabilidades P_{ABC} para combinações de 3 aminoácidos	71

LISTA DE DEFINIÇÕES

Abundância relativa de aminoácidos: Medida da quantidade relativa de cada aminoácido disponível em determinado ambiente prebiótico.

Aminoácido: Molécula orgânica que constitui a unidade básica dos peptídeos.

AB_D : Abundância normalizada do aminoácido localizado à direita da ligação peptídica.

AB_E : Abundância normalizada do aminoácido localizado à esquerda da ligação peptídica.

Corte de probabilidade: Percentual adotado para identificar quantas das cadeias mais prováveis são necessárias para representar uma fração pré-definida da probabilidade total de formação.

Dipeptídeo: Peptídeo formado por dois aminoácidos ligados por uma ligação peptídica.

E_0 : Condição inicial do modelo probabilístico, associada à probabilidade acumulada antes da formação da primeira ligação peptídica.

Ligação peptídica: Ligação química que une aminoácidos entre si, permitindo a formação de peptídeos e proteínas.

Peptídeo: Cadeia formada pela ligação de dois ou mais aminoácidos.

Proteína: Macromoléculas formadas por longas cadeias de aminoácidos unidos por ligações peptídicas.

P_{AB} : Probabilidade de formação da ligação peptídica entre os aminoácidos A e B , calculada a partir da abundância dos aminoácidos e do rendimento da ligação correspondente.

Razão de relevância (ρ): Medida definida como a razão entre o número de cadeias peptídicas que permanecem após a seleção das cadeias com maior probabilidade de formação e o número total de cadeias possíveis para um mesmo número de aminoácidos. Essa razão indica qual fração do conjunto total de cadeias é considerada na análise após a aplicação do critério probabilístico.

Rendimento da ligação peptídica: Medida da eficiência da formação de ligações peptídicas, indicando a quantidade de peptídeos formados em relação ao total de aminoácidos disponíveis.

Rendimento normalizado da ligação peptídica (R): Rendimento normalizado da ligação peptídica entre dois aminoácidos, obtido a partir de dados experimentais de simulação prebiótica.

SIPF: Experimento de simulação prebiótica baseado em ciclos sucessivos de hidratação e desidratação, nos quais misturas de aminoácidos são submetidas a condições que favorecem reações de condensação. Durante esses ciclos, ocorre a formação de ligações peptídicas entre aminoácidos, resultando na produção de dipeptídeos. A eficiência dessas ligações é quantificada por meio do rendimento relativo de cada ligação peptídica, valores utilizados como parâmetros neste trabalho.

Tripeptídeo: Peptídeo formado por três aminoácidos ligados por uma ligação peptídica.

***n*:** Número de aminoácidos que compõem uma cadeia peptídica, utilizado como medida do tamanho da cadeia.

1 INTRODUÇÃO

Esse trabalho desenvolve e analisa um modelo matemático para a probabilidade de síntese de peptídeos em condições prebióticas. O estudo descreve o desenvolvimento e a aplicação de um modelo matemático probabilístico elaborado a partir de dados experimentais, tais como os dados de rendimentos de ligações peptídicas em ambientes simulados [1, 2] e também os dados de abundância relativa (terrestre e extraterrestres) dos aminoácidos [3].

Os peptídeos são polímeros formados por aminoácidos, ou seja, cadeias de aminoácidos unidos por ligações químicas. Eles têm funções essenciais para a vida, por exemplo: muitos deles desempenham papéis como hormônios ou reguladores da liberação de hormônios, alguns são neuropeptídeos (endorfina, vasopressina, oxitocina); neurotransmissores; antibióticos naturais, adoçantes, entre outras funções. Os peptídeos com 10 a 20 aminoácidos combinados são chamados de oligopeptídeos, já ligação peptídica com vários aminoácidos é chamada de polipeptídeos. Proteínas são polipeptídeos formados a partir de ligações de 50 aminoácidos até milhares de aminoácidos, mas essa nomenclatura pode variar de autor para autor, alguns autores consideram proteínas os polipeptídeos a partir de 20 aminoácidos [4]. As proteínas também desempenham um papel muito importante nos organismos vivos, exercendo diferentes funções, um exemplo é que elas atuam em todos os tecidos do nosso corpo, algumas atuam realizando o transporte de substâncias, além de serem fundamentais para a formação de enzimas [4–6]. Além disso, peptídeos e proteínas predominam nas células procarióticas, podendo representar até 75% de sua composição [7, 8].

O estudo desse tema é importante para compreender a base molecular da vida e como as estruturas biológicas se originaram, pois ao compreender como peptídeos e proteínas foram formados, podemos tentar construir um caminho que levou à origem da vida. É um fato que a síntese de peptídeos e proteínas não ocorreu de maneira aleatória, e fatores como a seleção de aminoácidos desempenham um papel importante nas funcionalidades das cadeias formadas. Assim, ao propormos um modelo matemático probabilístico, baseado em dados experimentais, conseguimos investigar como a abundância relativa dos aminoácidos e os rendimentos das ligações peptídicas influenciaram na formação de estruturas moleculares complexas na Terra primitiva.

Nosso estudo visa descrever a formação de peptídeos e proteínas a partir de dados de rendimentos de ligações peptídicas e dados de quantidades de aminoácidos sintetizados simulando ambientes terrestre e extraterrestre assim como análise de meteoritos. Vários trabalhos neste contexto mediram os rendimentos dessas reações, citamos os experimentos de Rode [1, 2], Leman [9] e Huber [10]. Neste contexto, elaboramos um modelo matemático para calcular a probabilidade da formação de peptídeos, utilizando os dados obtidos no experimento de Rode [1, 2] e dados de abundância relativa (terrestre e extraterrestres) dos aminoácidos [3, 11]. Este trabalho visa contribuir para o estudo dos processos químicos envolvidos na origem da

vida, por meio de uma abordagem baseada em modelagem matemática, análise estatística e simulações numéricas.

O objetivo geral deste trabalho é desenvolver e aplicar um modelo matemático probabilístico, baseado em dados experimentais de abundância de aminoácidos e de rendimento de ligações peptídicas, com o propósito de investigar a formação de peptídeos e proteínas sob condições prebióticas. Por meio de simulações computacionais, busca-se compreender como fatores químicos e probabilísticos influenciam a seleção de cadeias peptídicas mais prováveis em diferentes cenários ambientais.

Nesse sentido, este trabalho tem como objetivos específicos:

- reunir e organizar dados experimentais da literatura referentes à abundância relativa de aminoácidos e aos rendimentos de ligações peptídicas obtidos em experimentos de simulação prebiótica;
- normalizar esses dados experimentais, de modo a torná-los compatíveis com a formulação matemática utilizada no modelo;
- formular um modelo matemático probabilístico para o cálculo da probabilidade de formação de cadeias peptídicas de diferentes comprimentos, considerando tanto a disponibilidade dos aminoácidos quanto a eficiência das ligações entre eles;
- estabelecer condições iniciais que permitam a aplicação do modelo a cadeias longas de aminoácidos;
- implementar computacionalmente o modelo matemático proposto, possibilitando a realização de simulações numéricas envolvendo um grande número de combinações possíveis de sequências peptídicas;
- analisar os resultados obtidos a partir de diferentes cortes de probabilidade, comparando cenários prebióticos distintos e investigando a influência da abundância dos aminoácidos e dos rendimentos das ligações na seleção de sequências peptídicas mais relevantes.

Nesta dissertação, no Capítulo 2, apresentamos a fundamentação teórica da química prebiótica, apresentando os mecanismos experimentais de síntese de peptídeos e experimentos importantes para o estudo da origem da vida. Já no Capítulo 3, apresentamos a metodologia adotada, incluindo a base de dados utilizada, a normalização dos rendimentos e abundâncias dos aminoácidos. No Capítulo 4, modelo matemático probabilístico desenvolvido e a implementação computacional. No Capítulo 5 apresentamos os resultados e discussões, contendo análises das simulações realizadas para diferentes ambientes (terrestre e extraterrestre) e para diferentes cortes de probabilidade. Por fim, no Capítulo 6, são apresentadas as conclusões do estudo. No Apêndice A explicamos detalhadamente o cálculo da probabilidade de formação de dipeptídeos e tripeptídeos, a partir do modelo probabilístico desenvolvido.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 QUÍMICA PREBIÓTICA E O AMBIENTE DA TERRA PRIMITIVA

A química prebiótica é um campo de estudo que investiga como reações químicas na Terra primitiva ajudaram no surgimento da vida que conhecemos hoje. Ela investiga como ocorreram esses processos que levaram moléculas simples a se combinarem para formar moléculas complexas tais como biopolímeros, estruturas coacervadas ou qualquer molécula que seja essencial para os seres vivos [12]. Assim, muitos pesquisadores simularam ambientes da Terra primitiva para tentar compreender como compostos orgânicos poderiam ter se formado nas condições da Terra primitiva. Um clássico exemplo de simulação é o famoso experimento feito por Stanley L. Miller (1930-2007) e Harold C. Urey (1893-1981) [13], que inclusive foi o experimento responsável por iniciar o que hoje chamamos de Química Prebiótica.

Os estudos e simulações em Química Prebiótica são realizados para investigar reações químicas que poderiam ocorrer em ambientes prebióticos, ou seja, sem vida conhecida. Nesse contexto, são importantes os dados sobre a composição geoquímica e geológica na Terra primitiva, a temperatura do planeta e dos oceanos, a intensidade das radiações que atingiam o planeta naquela época e as moléculas que já existiam no planeta. Embora esses dados sejam cruciais, pode ser uma tarefa complexa consegui-los, pois existem muitas lacunas sobre as condições existentes na Terra antes da origem da vida. Isto pode parecer um beco sem saída, mas na realidade abre a perspectiva que devemos apenas mostrar, que dentro de determinadas condições, a vida pode emergir a partir de moléculas simples, como os aminoácidos, por meio de processos químicos. Por fim, destacamos que sequer sabemos se a vida teve início na Terra ou se foi trazida à Terra [12, 14, 15].

2.1.1 Experimentos históricos e hipóteses sobre a origem da vida

Quando se fala sobre o surgimento da vida, uma das primeiras explicações propostas entre filósofos e cientistas foi a teoria da geração espontânea, ou seja, a ideia de que um ser vivo poderia surgir de matéria em estado de putrefação, como, por exemplo, um alimento em decomposição. Mesmo que, nos dias de hoje, essa ideia pareça ser algo improvável, no passado ela era debatida e até mesmo incentivava alguns filósofos a fazerem experimentos, um exemplo disso é o experimento feito pelo médico Johann Baptista van Helmont (1580-1644), onde ele seguiu uma espécie de “receita” para a formação de camundongos. O experimento foi feito colocando algumas roupas sujas dentro de um jarro e cobrindo-as de trigo, e após 21 dias seriam formados camundongos adultos. Atualmente é fácil supormos o que aconteceu para as aparições dos camundongos, ou seja, Helmont provavelmente deixou o jarro aberto, o que atraiu os camundongos para dentro do jarro para comer o trigo. Na época esse experimento foi aceito, visto que representava a crença predominante entre os estudiosos do assunto [12].



Figura 2.1: Descrição do experimento de Francesco Redi.

Fonte: [17]

A teoria da geração espontânea começou a ser questionada quando Francesco Redi (1626-1697), médico de Florença, fez um experimento obtendo resultado que refutava tudo o que era acreditado até o momento. Este experimento já tinha sido realizado anteriormente por outros estudiosos do assunto. No experimento em questão, representado na Figura 2.1, Redi colocou carne crua (uma cobra, alguns peixes, um pedaço de vitela) dentro de frascos, alguns frascos eram vedados com papel e barbante e outros eram deixados abertos. De acordo com a teoria da geração espontânea, após alguns dias deveria nascer insetos em todos os frascos, mas o que Redi observou foi que apenas nos frascos abertos tinham moscas, já nos frascos vedados nenhuma mosca foi observada. Redi conseguiu demonstrar que os insetos eram atraídos pela carne e não nascidos na matéria inanimada (a carne em decomposição), indo totalmente contra a teoria da geração espontânea [12, 16].

Mesmo após o experimento de Redi a teoria da geração espontânea ainda era acreditada por uma parcela de pesquisadores, porém com o passar dos anos, e após outros experimentos, cada vez mais a teoria da geração espontânea foi perdendo credibilidade, sendo que no século XIX, Louis Pasteur (1822-1895), químico francês, realizou um experimento que comprovou que a vida não poderia surgir de matéria em decomposição, e esse experimento é descrito em sua obra. O experimento de Pasteur foi feito utilizando um líquido rico em nutrientes, como um caldo de carne, que ele ferveu em um recipiente aberto com gargalo em formato de S, conhecido como “pescoço de cisne”, que permite a entrada de ar no recipiente, mas impede a entrada de partículas de poeira, ou seja, apenas a entrada de ar era permitida chegar até o líquido. Pasteur esquentou o líquido no recipiente até que ele fosse fervido, com isso acabou matando qualquer microorganismo presente no recipiente, tornando o líquido totalmente estéril. Após esse processo de esterilização, Pasteur deixou o líquido esfriando, mantendo o gargalo do recipiente intacto, o que permitia que apenas o ar entrasse no frasco. Após meses observando o líquido no frasco, foi notório que o líquido não apresentava nenhum sinal de vida, com isso, Pasteur conseguiu constatar que a vida não iria surgir simplesmente com o contato com o ar. Ainda para confirmar sua hipótese, Pasteur quebrou o gargalo do recipiente e em cerca de 24,

36 ou 48 horas depois, o caldo começou a apresentar sinais de vida, como bolores. Com isso, o experimento derrubou, então, a teoria da geração espontânea, e dando base para Teoria da Biogênese, que foi aceita pelos cientistas daquela época [18, 19].

Em 1924, o bioquímico Aleksander I. Oparin (1894-1980) criou um esquema para estudar a origem da vida. Já em 1928, o geneticista John B. S. Haldane (1892-1964) propôs um esquema semelhante, dando origem à hipótese de Oparin-Haldane [12], que é representada na Tabela 2.1.

Etapa	Descrição e Resultados
Reações químicas entre moléculas simples	Moléculas como CH ₄ , CO, CO ₂ , H ₂ , H ₂ S, HCN, NH ₃ , H ₂ O, etc. interagem entre si, dando origem a moléculas mais complexas, como aminoácidos, açúcares, etc.
Acúmulo e combinação de moléculas complexas	Após milhões de anos, moléculas complexas se unem, formando biopolímeros, como peptídeos, polissacarídeos e nucleotídeos.
Formação de estruturas coacervadas	Os biopolímeros reagem entre si, dando origem a estruturas que se assemelham a células.
Origem da vida	Após milhões de anos, dentro das estruturas coacervadas, reações químicas tornam-se tão complexas, que as estruturas são consideradas vivas.

Tabela 2.1: Etapas da formação de vida a partir de moléculas simples segundo a hipótese Oparin-Haldane.

Uma outra hipótese é a teoria da síntese de RNA, conhecida como mundo de RNA, popularizada pelo físico e bioquímico Walter Gilbert (1932-) em 1986. Essa hipótese tenta explicar o surgimento de vida na Terra, defendendo a possibilidade de síntese espontânea de moléculas de RNA em ambiente prebiótico, uma vez que o RNA possui a capacidade de se autorreplicar e catalisar suas próprias reações [20].

Uma outra hipótese é a da Panspermia, que começou a ser discutida mais seriamente no século XX. Svante A. Arrhenius (1859-1927), um físico-químico sueco, sugeriu que a vida que conhecemos hoje poderia não ter sido sintetizada aqui na Terra, mas sim sintetizada em outro planeta e trazida para a Terra por meio de meteoros, cometas, etc..., e aqui encontrou um ambiente com condições favoráveis. A panspermia é uma hipótese que atraiu muitas críticas. Muitos cientistas defendem que os organismos não iriam conseguir sobreviver no espaço e que essa hipótese não resolve o problema do início da vida, apenas transfere o problema para fora do planeta Terra [21].

O experimento de S. L. Miller (1930-2007), em 1953 [13], foi um grande marco, pois esse experimento deu início ao que hoje é conhecido como química prebiótica e também foi o primeiro experimento que trouxe evidências de que a hipótese de Oparin-Haldane

poderia estar correta [12].

Miller utilizou um equipamento semelhante àquele representado na Figura 2.2, que imitava o ambiente da Terra primitiva. Foi utilizado um frasco de 5 litros onde gases (metano, amônia, hidrogênio e vapor de água) circulavam, e o tubo em formato de U impedia que o fluxo acontecesse de forma contrária. Esses gases eram submetidos a uma descarga elétrica, que simulava situações que poderiam ter ocorrido na atmosfera antiga, como raios. Essa descarga elétrica gerava radicais livres que são moléculas/átomos reativos que contribuíam para a síntese de moléculas. Resumidamente, o equipamento utilizado por Miller aquecia a água e o vapor se misturava com os outros gases citados na Tabela 2.1. Essa mistura de gases era transportada para outra parte do aparelho e sofria descargas elétricas, e após isso, o vapor esfriava e voltava para seu estado líquido no recipiente inicial, para ser aquecido e repetido todo o processo, novamente. Esse processo foi repetido durante uma semana e nesse período os compostos formados se juntavam à água líquida. No final do primeiro dia a cor da água já estava visivelmente rosa, e a do fim da semana já estava vermelha. A mistura resultante foi coletada para análise por cromatografia em papel, que consiste em um processo para separar produtos químicos ou substâncias coloridas, como pode ser visto na Figura 2.3. Conforme evidenciado na Figura 2.3, foram encontrados aminocidos glicina, α -alanina, β -alanina e outros compostos orgânicos, sendo que a presença dos aminoácidos ácido aspártico e α -amino-n-butírico não era uma certeza devido à baixa intensidade das manchas, e por fim as manchas *A* e *B* não foram identificadas. Assim, de fato Miller conseguiu sintetizar aminoácidos a partir de moléculas simples [13].

Embora esse experimento tenha muita importância no estudo sobre a origem da vida, representando um marco na pesquisa sobre o assunto, hoje em dia ele é considerado desatualizado, sendo que do ponto de vista atual da química prebiótica o experimento não apresenta com precisão as condições da Terra primitiva, visto que estudos mostram que provavelmente a Terra primitiva não era composta por metano e amônia em grandes quantidades, mas sim por dióxido de carbono, nitrogênio e vapor de água, e essa composição torna a atmosfera da Terra primitiva parcialmente oxidante e não redutora como pensava Miller [22].

Uma hipótese mais recente, conhecida como Hipótese GADV [23], foi proposta por Kenji Ikehara (1944-) na década de 2000. De acordo com a Hipótese GADV de Ikehara, as primeiras proteínas formadas na Terra primitiva eram compostas apenas por quatro aminoácidos: Glicina (G), Alanina (A), Ácido Aspártico (D) e Valina (V). Esses aminoácidos teriam sido os mais abundantes e mais facilmente sintetizados em condições prebióticas, apresentando propriedades químicas complementares que facilitariam a formação de estruturas proteicas estáveis e funcionais.

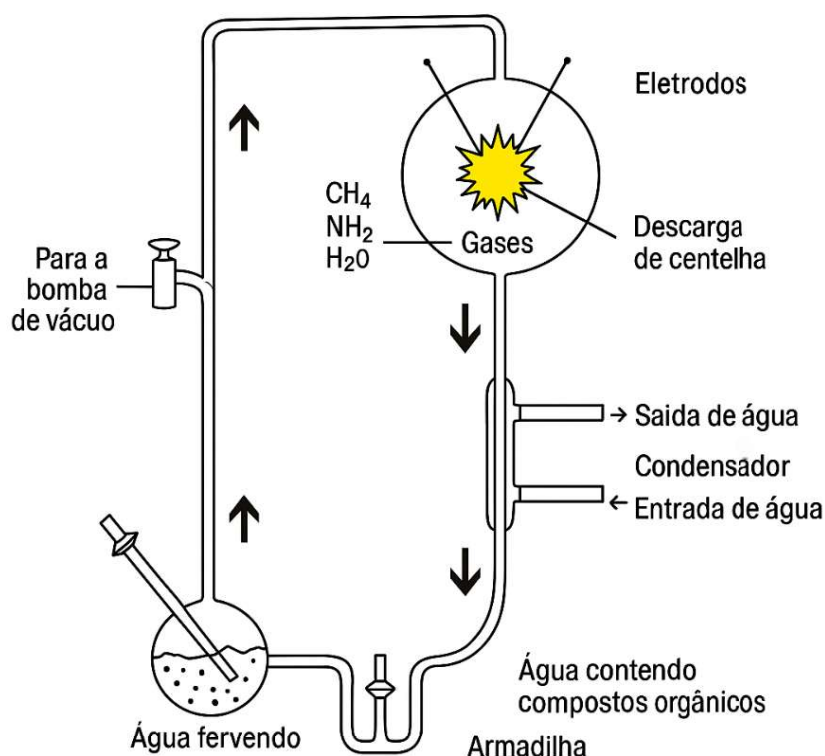


Figura 2.2: Aparelho utilizado por Miller para o experimento.
Fonte: Adaptado de [24]

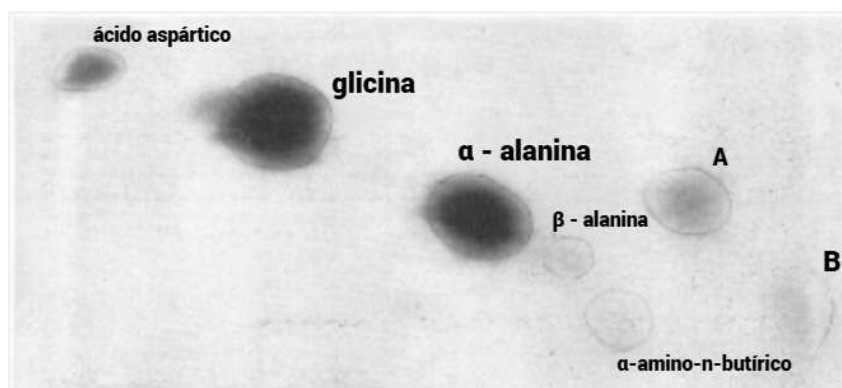


Figura 2.3: Coleta da mistura resultante no experimento de Miller em cromatografia em papel.
Fonte: Adaptado de [13]

2.1.2 Ambientes favoráveis para as reações químicas

Alguns tipos de ambientes podem ter contribuído para as reações químicas entre aminoácidos na Terra primitiva. Os ambientes que têm sido estudados são: a) **fontes hidrotermais**, onde as altas temperaturas e pressões alteram as propriedades da água, permitindo que ela favoreça diversas reações químicas que poderiam ter sido importantes para a origem da vida; b) **poças de água do mar ou pequenas lagoas**, que sofrem um processo de hidratação e desidratação devido ao aumento das marés e em seguida do calor do sol, processo esse

que favorece a formação principalmente de polímeros; c) **vulcões terrestres ativos e locais de choque de cometas ou meteoritos**, pois misturas de aminoácidos em estado sólido, como por exemplo, aminoácidos misturados em areia, facilitam a formação de ligações peptídicas. Importante ressaltar que apesar de reações em estado sólido serem importantes em algumas etapas, elas não podem estar presentes no processo completo da origem da vida, pois nos organismos vivos atuais as reações químicas ocorrem em ambiente aquoso [12].

Por outro lado, um outro ambiente em que reações químicas poderiam ter ocorrido para contribuir para a origem da vida é o **ambiente interestelar**, onde moléculas importantes para a vida já foram detectadas [12].

2.2 FORMAÇÃO DE PEPTÍDEOS

Peptídeos são moléculas formadas por ligações químicas entre aminoácidos. Alguns métodos de sintetizar os peptídeos foram propostos, sendo alguns deles: a) agentes de condensação [25]; b) fusão de aminoácidos [26]; c) catálise por argilas e minerais [27]; d) polimerização térmica em estado sólido [28]. Muitos desafios foram encontrados nos experimentos feitos para reproduzir esses métodos, como por exemplo, a dificuldade para reproduzir com precisão as condições da Terra primitiva, ou que poucos peptídeos eram formados mesmo após longo tempo de reação do experimento, e que alguns métodos requeriam uma quantidade excessiva de aminoácidos ácidos ou básicos.

Um outro mecanismo, que para o nosso trabalho é muito importante, é o mecanismo constituído de ciclos de hidratação e desidratação representado na Figura 2.4, que basicamente são fases que alternam entre a fase seca, quando a água evapora, e a fase úmida, quando a água é adicionada. Quanto mais esses ciclos de hidratação e desidratação são realizados, maiores ficam as cadeias de peptídeos. O mecanismo para a formação de peptídeos, induzido por sal (SIPF), é uma reação que ocorre em um ambiente com apenas aminoácidos, NaCl (cloreto de sódio) e íon Cu(II). Os sais, por exemplo NaCl (cloreto de sódio), absorvem a água, fazendo com que a formação de ligação peptídica seja mais favorável termodinamicamente. O íon Cu(II) atua como catalisador na formação da ligação peptídica, formando um complexo monoclorocuprato, que é um tipo de estrutura onde se ligam os aminoácidos entre si [1, 2].

Vamos utilizar nesse trabalho os dados experimentais obtidos pelo químico Bernd M. Rode (1946-2022) [1, 2], que realizou experimentos usando o mecanismo (SIPF) para obter rendimentos das ligações peptídicas entre os aminoácidos estudados.

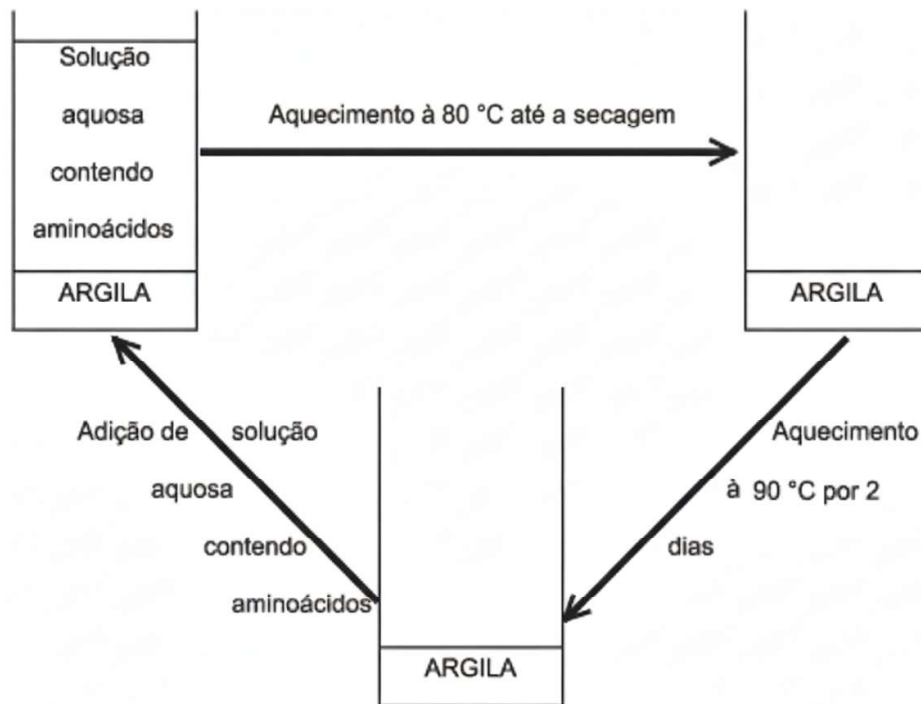


Figura 2.4: Esquema comumente utilizado no estudo de ciclos de hidratação e desidratação.
Fonte: Adaptado de [29]

3 METODOLOGIA

3.1 REVISÃO BIBLIOGRÁFICA DE DADOS EXPERIMENTAIS

Os dados experimentais nesse trabalho contêm dois conjuntos de dados principais: os rendimentos de ligações peptídicas (Tabela 3.1) [1, 2], e os dados de abundância relativa de cada aminoácido (Tabela 3.2) [3].

Sobre os rendimentos de ligações peptídicas, os trabalhos de Rode [1, 2] utilizaram um mecanismo de formação conhecido como SIPF (Salt-Induced Peptide Formation). Essa reação é realizada na presença de NaCl (cloreto de sódio), que age retirando água do meio de reação, facilitando a formação de peptídeos; e de íons Cu(II) que catalisam a reação [1, 30]. Os experimentos conduzidos por Rode consideraram 9 aminoácidos, enquanto utilizamos no nosso trabalho apenas as combinações entre 4 aminoácidos. O motivo para que esses 4 aminoácidos fossem escolhidos é baseado na Hipótese do Mundo de Proteínas [GADV], proposta por Ikehara [23], segundo a qual as primeiras proteínas formadas na Terra primitiva eram compostas por apenas quatro aminoácidos: Glicina, Alanina, Ácido Aspártico e Valina.

A Tabela 3.1 apresenta os valores dos rendimentos das ligações peptídicas dos 4 aminoácidos após 5 ciclos SIPF. Mesmo após o quinto ciclo, o rendimento para a obtenção de alguns dipeptídeos foi muito baixo para ser detectado. Assim, o valor do rendimento da ligação do dipeptídeo Val - Asp, por meio de processo SIPF, não foi encontrado em nossa revisão bibliográfica. Para lidar com essa ausência de dados, adotou-se o valor do menor rendimento detectado, que foi utilizado como referência, como descrito na seção 4.1. O dado correspondente ao rendimento da ligação do dipeptídeo Val - Val foi obtido em [30]. Já os demais valores dos rendimentos das ligações de dipeptídeos A - B foram obtidos em [1].

Os valores da diagonal da Tabela 3.1 representam porcentagens de rendimentos de homodímeros, dipeptídeos compostos por dois aminoácidos iguais e os valores que não se encontram na diagonal representam a porcentagem de rendimento de heterodímeros, dipeptídeos formados por dois aminoácidos diferentes.

Pode-se notar que os valores de rendimento dos dipeptídeos AB não são iguais aos valores do rendimento dos dipeptídeos BA. Isso acontece porque a ligação peptídica apresenta direcionalidade, sempre é formada do grupo amino livre (N-terminal) de um aminoácido para o grupo carboxila livre (C-terminal) de outro aminoácido. Ou seja, a ordem dos aminoácidos na sequência é muito importante, pois influencia a reatividade e a estabilidade do dipeptídeo formado. Logo, o rendimento obtido por AB pode ser diferente do rendimento obtido por BA [31]. Essa diferença pode ser visualizada na Figura 3.1, que ilustra as diferentes estruturas resultantes da formação de dipeptídeos.

Tabela 3.1: Rendimento em % de dipeptídeos AB formados a partir da massa original de aminoácidos após cinco ciclos.

$A \downarrow / B \rightarrow$	GLY	ALA	ASP	VAL
GLY	6,57	0,91	0,24	2,05
ALA	1,20	1,86	0,37	1,17
ASP	0,73	0,19	0,38	0,15
VAL	0,89	0,26		0,96

Gly = Glicina, Ala = Alanina, Asp = Ácido Aspártico, Val = Valina. Os ciclos são sucessivos processos de hidratação e desidratação, que simulam as condições ambientais da Terra primitiva e promovem reações de condensação entre aminoácidos. Os dados desta tabela foram obtidos de Rode [1, 2].

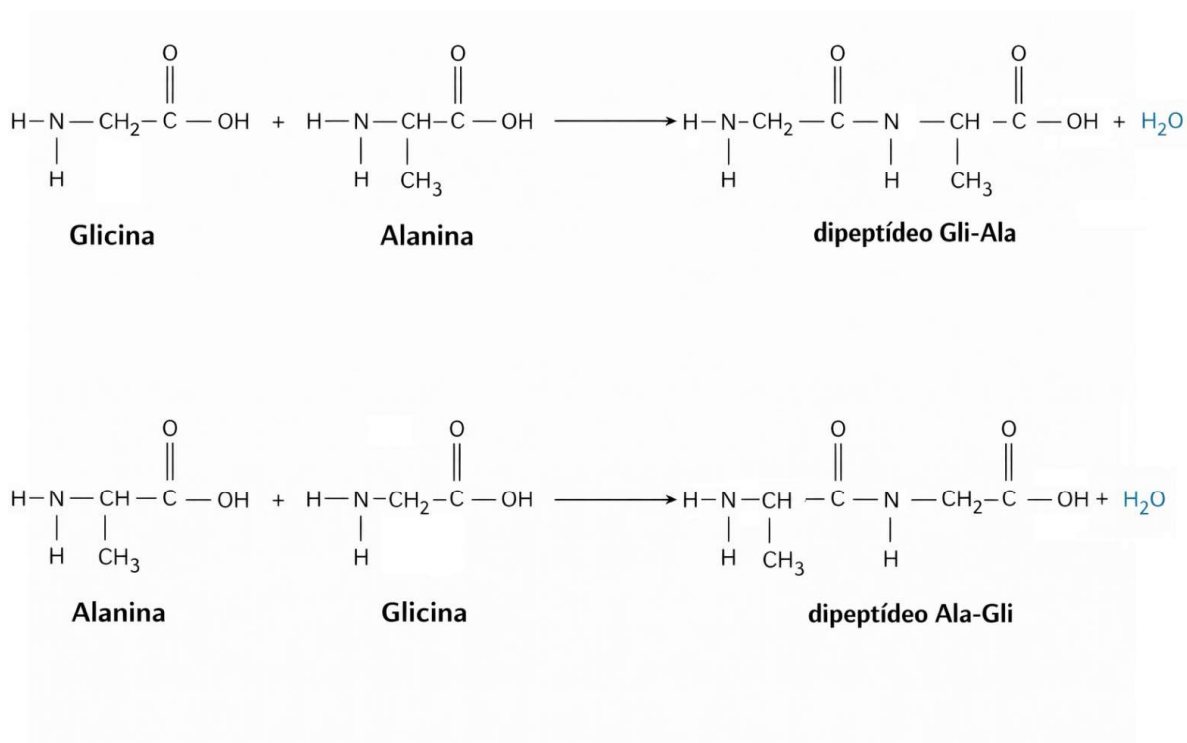


Figura 3.1: Estruturas resultantes da formação de dipeptídeos Gly–Ala e Ala–Gly. Fonte: Autora

Além dos rendimentos das ligações dos peptídeos, o modelo também depende dos dados de abundância de aminoácidos em ambiente prebiótico, que foram estimados a partir de diversos experimentos simulando a Terra prebiótica, experimentos com ISM (meio interestelar) e outras análises complementares descritas em [3]. Foram considerados dois ambientes: terrestres e extraterrestres. A Tabela 3.2 apresenta os valores das abundâncias, relativas ao

aminoácido Glicina, dos aminoácidos Alanina, Ácido Aspártico e Valina, tanto em ambiente terrestre como extraterrestre.

Tabela 3.2: Abundância relativa dos aminoácidos GADV em fontes terrestres e extraterrestres (glicina = 1).

Aminoácidos	Terrestre		Extraterrestre	
	Maior	Menor	Maior	Menor
Gly	1,00	1,00	1,00	1,00
Ala	2,00	0,30	1,50	0,50
Asp	0,15	0,02	0,80	0,20
Val	0,30	0,05	0,30	0,10

GADV = (Gly, Ala, Asp, Val), Gly = Glicina, Ala = Alanina, Asp = Ácido Aspártico, Val = Valina. Fontes terrestres: descargas elétricas em misturas gasosas, fontes hidrotermais, reações em meio aquoso com sais e minerais e ciclos de hidratação/desidratação. Fontes extraterrestres: análises de meteoritos e experimentos simulando ambiente interestelar. Os dados desta tabela foram obtidos de Zaia et al. (2008). Fonte: Adaptado de [3].

3.2 NORMALIZAÇÃO DOS DADOS

3.2.1 Normalização dos rendimentos de formação de dipeptídeos

Antes de aplicar o modelo matemático de probabilidade de formação de peptídeos e proteínas, os valores da Tabela 3.1 foram normalizados, atribuindo o valor 1,00 para o rendimento da reação de formação de dipeptídeos de valor mais abundante (Gly - Gly) e ajustando os demais valores proporcionalmente. Segue a Tabela 5.1, que apresenta os valores normalizados à unidade dos rendimentos das reações de formação de dipeptídeos.

Tabela 3.3: Valores normalizados de rendimento de dipeptídeos AB formados a partir da massa original de aminoácidos, após o quinto ciclo. Valores normalizados em relação ao rendimento do dipeptídeo Gly-Gly.

$A \downarrow / B \rightarrow$	GLY	ALA	ASP	VAL
GLY	1,00	0,139	0,037	0,312
ALA	0,182	0,283	0,056	0,178
ASP	0,111	0,029	0,058	0,023
VAL	0,135	0,040		0,146

Gly = Glicina, Ala = Alanina, Asp = Ácido Aspártico, Val = Valina. Os ciclos são sucessivos processos de hidratação e desidratação, que simulam as condições ambientais da Terra primitiva e promovem reações de condensação entre aminoácidos. Os dados desta tabela foram obtidos de Rode [1,2] e por ajustes numéricos.

3.2.2 Normalização da abundância dos aminoácidos

De modo análogo, para utilização no modelo matemático de probabilidade de formação de peptídeos e proteínas, os dados da abundância relativa dos aminoácidos GADV encontrados na Tabela 3.2 também foram normalizados, tomando como 1,00 o valor de maior abundância em cada cenário (terrestre e extraterrestre). Em ambos os cenários, o aminoácido com maior abundância é a Alanina. A Tabela 3.4 apresenta os valores das abundâncias dos aminoácidos, normalizados à unidade.

Tabela 3.4: Abundâncias relativas normalizadas dos aminoácidos GADV em fontes terrestres e extraterrestres

Aminoácidos	Terrestre		Extraterrestre	
	Maior	Menor	Maior	Menor
Gly	0,50	0,50	0,67	0,67
Ala	1,00	0,15	1,00	0,33
Asp	0,075	0,01	0,53	0,13
Val	0,15	0,025	0,20	0,07

GADV=(Gly, Ala, Asp, Val), Gly = Glicina, Ala = Alanina, Asp = Ácido Aspártico, Val = Valina. Fontes terrestres: descargas elétricas em misturas gasosas, fontes hidrotermais, reações em meio aquoso com sais e minerais e ciclos de hidratação/desidratação. Fontes extraterrestres: análises de meteoritos e experimentos simulando ambientes interestelares.

Fonte: Adaptado de [3].

3.3 LINHAS DE TENDÊNCIA

Em situações apresentadas no capítulo 5 a utilização da ferramenta estatística de linha de tendência para realizar estimativas e previsões é bastante útil. De modo geral, a linha de tendência é uma forma de visualizar padrões.

Consideremos três modelos de regressão com ajustes linear, quadrático e exponencial:

- ajuste linear ($y = ax + b$),
- ajuste exponencial ($y = ae^{bx}$),
- ajuste quadrático ($y = ax^2 + bx + c$).

Para os três ajustes acima, o coeficiente de determinação R^2 , definido por:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.1)$$

onde:

- $\sum_{i=1}^n (y_i - \hat{y}_i)^2$: soma dos quadrados dos resíduos (erros).
- $\sum_{i=1}^n (y_i - \bar{y})^2$: soma dos quadrados totais, em relação à média dos dados.
- y_i : valores observados.
- $\hat{y}_i$: valores ajustados pelo modelo.
- $\bar{y}$: média dos n valores observados.

O coeficiente de determinação R^2 é uma medida estatística que indica a qualidade do modelo ajustado. Seus valores variam entre 0 e 1, onde $R^2 = 1$ representa um ajuste perfeito dos dados observados ao modelo ajustado [32].

Porém, é importante destacar que o R^2 tradicional, dado em 3.1, quando aplicado a ajustes não lineares, pode levar a interpretações equivocadas, por esse motivo utilizamos o Pseudo- R^2 de Nagelkerke para identificar a qualidade dos ajustes não lineares. Essa métrica funciona de forma semelhante ao R^2 da regressão linear, variando entre 0 e 1 [33]. Ele é calculado a partir da razão de verossimilhança do modelo ajustado em relação ao modelo nulo. A verossimilhança mede a probabilidade de observar um determinado conjunto de dados, dado uma família de modelos com diferentes valores de parâmetros, ou seja, é uma medida de quão bem o modelo explica os dados. A verossimilhança é definida como:

$$R_{Nagelkerke}^2 = \frac{1 - \left(\frac{L_0}{L_m} \right)^{2/n}}{1 - L_0^{2/n}},$$

onde:

- L_0 : verossimilhança do modelo nulo, isto é, indica o quão prováveis são os dados observados se o modelo estiver correto.
- L_m : verossimilhança do modelo ajustado, ou seja, o quanto os dados são prováveis dado o modelo com os parâmetros estimados.
- n : número total de observações.

Para tornar essas definições mais claras, apresentamos a seguir as expressões matemáticas de L_0 e L_m , que correspondem, respectivamente, à verossimilhança do modelo nulo e do modelo ajustado.

$$L_0 = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(y_i - \bar{y})^2}{2\sigma_0^2}\right),$$

onde σ_0^2 é a variância total dos dados.

Já no modelo ajustado a verossimilhança é calculada utilizando os valores previstos pelo modelo:

$$L_m = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma_m^2}\right),$$

onde σ_m^2 é a variância dos resíduos do modelo ajustado.

4 O MODELO MATEMÁTICO

4.1 O MODELO DE PROBABILIDADE DE FORMAÇÃO DE PEPTÍDEOS

Para calcularmos a probabilidade de formação de peptídeos (cadeias de n -aminoácidos) em um ambiente simulado, desenvolvemos um modelo matemático. Este modelo está baseado em parâmetros específicos, como abundância de aminoácidos, rendimento de ligações peptídicas e condições iniciais e de contorno. Nosso modelo matemático de probabilidade de formação de peptídeos é dado por:

$$P_{AB} = AB_E \cdot E_0 \cdot R \cdot AB_D, \quad (4.1)$$

onde

- AB_E (abundância do aminoácido esquerdo): Representa a abundância do aminoácido A à esquerda do par dos aminoácidos próximos vizinhos interagentes formados pelo aminoácido acrescentado e o aminoácido do n -peptídeo com o qual houve a ligação. Esses valores são retirados da Tabela 3.4.
- AB_D (abundância do aminoácido direito): Representa a abundância do aminoácido B à direita do par dos aminoácidos próximos vizinhos interagentes formados pelo aminoácido acrescentado e o aminoácido do n -peptídeo com o qual houve a ligação. Esses valores são retirados da Tabela 3.4.
- E_0 (condição inicial): Representa a probabilidade acumulada nos passos anteriores, àquela da formação do peptídeo atual. Na primeira etapa, temos $E_0 = 1$, o que indica que as condições iniciais para esse caso são normalizadas à unidade, sem influência de iterações anteriores.
- R (rendimento da ligação): Representa o rendimento da ligação peptídica que se deseja formar entre os aminoácidos AB. Esses valores são retirados da Tabela 5.1.

4.2 CONDIÇÃO DE CONTORNO PARA CADEIA DE n -AMINOÁCIDOS

Os exemplos apresentados no Apêndice A mostram o cálculo do modelo probabilístico para a construção de cadeias peptídicas pequenas, considerando cadeias de 2 e 3 tipos de aminoácidos. No entanto, o objetivo é que o modelo possa ser usado para descrever cadeias de aminoácidos mais complexas. Assim, propomos uma hipótese (condição de contorno) para o modelo do cálculo da probabilidade, para que possa ser calculado a formação de um peptídeo de tamanho n .

Consideramos então um modelo iterativo. Note que os dados experimentais apresentados nas Tabelas 3.3 e 5.1 referem-se aos rendimentos de ligações entre dois aminoácidos, não entre um aminoácido e um peptídeo. Então, para calcular um peptídeo formado de 3 aminoácidos, consideramos as probabilidades de formação dos peptídeos de 2 aminoácidos como condições iniciais E_0 e também uma condição (hipótese) de contorno:

- **Condição de contorno:** Quando um novo aminoácido é adicionado a uma cadeia já construída (pela direita ou esquerda), consideramos como rendimento da ligação o valor correspondente entre esse novo aminoácido e o aminoácido próximo vizinho contido na cadeia existente, naquele ponto de ligação. Partimos do pressuposto de que esse rendimento depende apenas dessa interação local de próximos vizinhos, e que os demais aminoácidos da cadeia não afetam o valor do rendimento da ligação.

Note que só existem dados experimentais de rendimentos de ligação entre dois aminoácidos. Por exemplo, não existem dados experimentais de R para ligação entre um dipeptídeo e um aminoácido. Nesta situação, para o valor de R em nosso modelo, consideramos apenas a interação de próximos vizinhos, ou seja, o valor de R em nosso modelo matemático é aquele da interação entre o aminoácido adicionado, à direita ou esquerda do peptídeo já formado, com o aminoácido próximo vizinho do peptídeo. Esse processo (hipótese de contorno) se repete para o cálculo da probabilidade da formação de peptídeos maiores, até serem constituídos de n aminoácidos.

4.3 MODELO COMPUTACIONAL

A simulação proposta foi implementada no software GNU Octave, versão 6.4.0. O código foi desenvolvido para realizar o cálculo da probabilidade de formação de cadeias de peptídeos sob condições prebióticas, com base em dados experimentais de abundância relativa e rendimento reacional para 4 aminoácidos: Glicina (G), Alanina (A), Ácido Aspártico (D), Valina (V), podendo ser adaptado para mais aminoácidos [1, 3]. O código permite a possibilidade de calcular a probabilidade para cadeias com grande número de aminoácidos, algo que manualmente seria uma tarefa inviável.

Os dados de entrada do modelo são:

- O **tamanho das cadeias** de aminoácidos a serem formadas;
- A **proporção acumulada de probabilidade** desejada para seleção das cadeias mais prováveis.

A partir desses parâmetros fornecidos pelo usuário, o algoritmo realiza as seguintes etapas:

1. Inicializa os vetores de abundância relativa e de rendimento das ligações entre pares de aminoácidos;
2. Gera todas as combinações possíveis de aminoácidos (sequência peptídica) até o tamanho que o usuário definiu nos dados de entrada;
3. Calcula a probabilidade de formação de cada cadeia específica, utilizando a equação 4.1;
4. Ordena as cadeias por valor de probabilidade em ordem decrescente;
5. Seleciona as cadeias mais prováveis, ou seja, aquelas com maiores probabilidades de formação, até atingir a proporção acumulada de probabilidade definida pelo usuário;
6. Exporta os resultados para um arquivo `.csv`, com todas as cadeias de combinações de aminoácidos (cadeias peptídicas) e seus respectivos valores de probabilidade de formação de cada cadeia de aminoácidos.

O tempo total de execução do código é exibido no final, permitindo avaliar o desempenho computacional para cada tamanho de cadeia de aminoácidos. A Figura 4.1 apresenta o fluxograma do algoritmo desenvolvido.

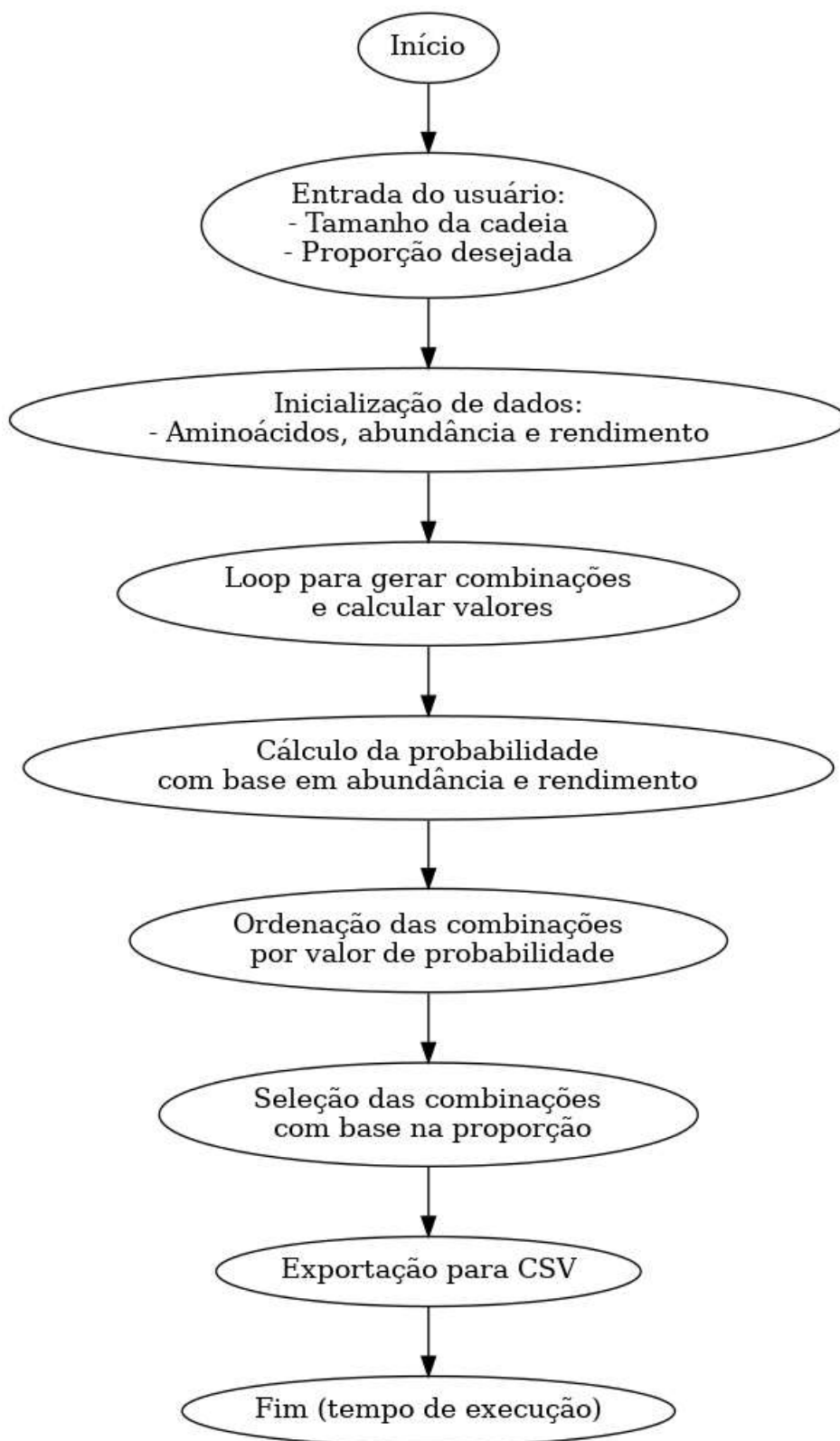


Figura 4.1: Fluxograma do algoritmo de simulação computacional da probabilidade de formação de peptídeos. Fonte: Autora.

5 SIMULAÇÕES NUMÉRICAS E DISCUSSÃO

A simulação de formação de peptídeos é fundamental para entendermos sobre a origem da vida. Utilizando o modelo matemático e realizando sua implementação, realizamos simulações para calcular as probabilidades de formação de n -peptídeos com combinações de aminoácidos específicos, sendo eles glicina, alanina, ácido aspártico e valina. Tendo como objetivo entender como as diferentes combinações de aminoácidos podem influenciar na formação de peptídeos, buscamos investigar se determinadas sequências são mais favoráveis e se há alguma tendência para a seleção natural de peptídeos.

Ressalta-se que as simulações foram executadas em um computador com 4,00 GB de memória RAM, processador Intel(R) Celeron(R) N4020 CPU (1.10 GHz) e placa gráfica integrada Intel(R) UHD Graphics 600 com 512 MB de memória.

5.1 SOBRE O RENDIMENTO DA LIGAÇÃO VAL-ASP

Observamos que o rendimento de Val–Asp não foi detectado nos experimentos de Rode [1, 2] e para o mecanismo SIPF, sugerindo que o valor real estaria próximo ou abaixo do limite de detecção experimental. O menor rendimento de combinação entre os 9 aminoácidos relatados por Rode foi para Lys–Val, com valor de 0,003 [1, 30]. Como o menor valor de detecção foi 0,003, utilizaremos esse valor para o rendimento da ligação Val–Asp. Esse valor encontra-se no limiar de detecção nos experimentos de Rode, o que é consistente com a ausência de detecção nos experimentos originais. Dessa forma, o valor utilizado está em conformidade com os dados observados nos experimentos.

Tabela 5.1: Valores de rendimento em % de dipeptídeos AB formados a partir da massa original de aminoácidos, após o quinto ciclo. Valores normalizados em relação ao rendimento do dipeptídeo Gly-Gly.

$A \downarrow / B \rightarrow$	GLY	ALA	ASP	VAL
GLY	1,00	0,139	0,037	0,312
ALA	0,182	0,283	0,056	0,178
ASP	0,111	0,029	0,058	0,023
VAL	0,135	0,040	0,003*	0,146

Gly = Glicina, Ala = Alanina, Asp = Ácido Aspártico, Val = Valina. Os ciclos são sucessivos processos de hidratação e desidratação, que simulam as condições ambientais da Terra primitiva e promovem reações de condensação entre aminoácidos. Os dados desta tabela foram obtidos de Rode [1, 2] e por ajustes numéricos.

5.2 RESULTADOS COM 4 AMINOÁCIDOS: ABUNDÂNCIA TERRESTRE

No primeiro momento fizemos a simulação utilizando os dados da abundância de aminoácidos terrestres [1, 2], apresentamos na Tabela 3.4. Também fizemos um corte estatístico com o objetivo de investigar para cadeias de n -aminoácidos, quantos m , $m < n$ desses peptídeos são necessários para atingir uma fração significativa (95%, 90%, 80%, 70%, 60% e 50%) do total das probabilidades de possíveis formações desses peptídeos.

Nesse cálculo executamos o código para diferentes tamanhos de cadeia de aminoácidos ($n = 2, 3, 4, 5, 6, 7, 8, 9$ e 10) e obtivemos a probabilidade total de formação de cada uma das 4^n cadeias de aminoácidos. Organizamos as combinações por probabilidade decrescente e somamos as probabilidades até atingirmos a probabilidade de corte desejada (50% a 95%).

- Corte em 95% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$.

A Tabela 5.2 apresenta os resultados da análise que foi realizada, onde a primeira coluna apresenta a quantidade de aminoácidos nas cadeias peptídicas, a segunda coluna indica os valores da razão de relevância ρ entre o número de peptídeos devido ao corte em 95% da probabilidade total e o número total 4^n de peptídeos e a terceira coluna indica o tempo em segundos para a execução do código.

Tabela 5.2: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)	Tempo para execução do código (em segundos)
2	$0,375 = 6/16$	0,029747
3	$0,171875 = 11/64$	0,047045
4	$0,0859375 = 22/256$	0,064073
5	$0,0390625 = 40/1024$	0,154939
6	$0,01953125 = 80/4096$	0,471832
7	$0,00933838 = 153/16384$	1,819561
8	$0,00450134 = 295/65536$	7,045964
9	$0,00220489 = 578/262144$	28,364451
10	$0,00104045 = 1091/1048576$	115,201531

Fonte: Autora

Observa-se que o tempo de execução do código cresce à medida que o número de aminoácidos na cadeia aumenta. O tempo computacional apresenta crescimento exponencial ajustado pela equação $y = 0,0014e^{1,0615x}$, com $R^2 = 0,9888$, tornando inviável a realização dos cálculos para n crescentes no computador utilizado neste trabalho. Por exemplo, de acordo com o ajuste exponencial do tempo computacional, para $n = 12$, o tempo computacional seria de $t = 477s$ e para $n = 15$, seria $t = 11528s$. Os demais resultados apresentados nesta seção (cortes de 90%, 80%, 70%, 60% e 50%) demandam tempo computacional semelhante, uma vez que o algorítmico empregado é o mesmo.

A Figura 5.1 apresenta a distribuição dos valores da Tabela 5.2, mostrando como a razão de relevância ρ diminuiu conforme o número de aminoácidos nas cadeias peptídicas aumenta, sugerindo visualmente uma taxa de decaimento. A linha de tendência ajustada (em vermelho) reforça esse comportamento, com um coeficiente de determinação $R^2 = 0,9995$, indicando um ótimo ajuste ao modelo exponencial ($y = 1,5776 \cdot e^{-0,732x}$).

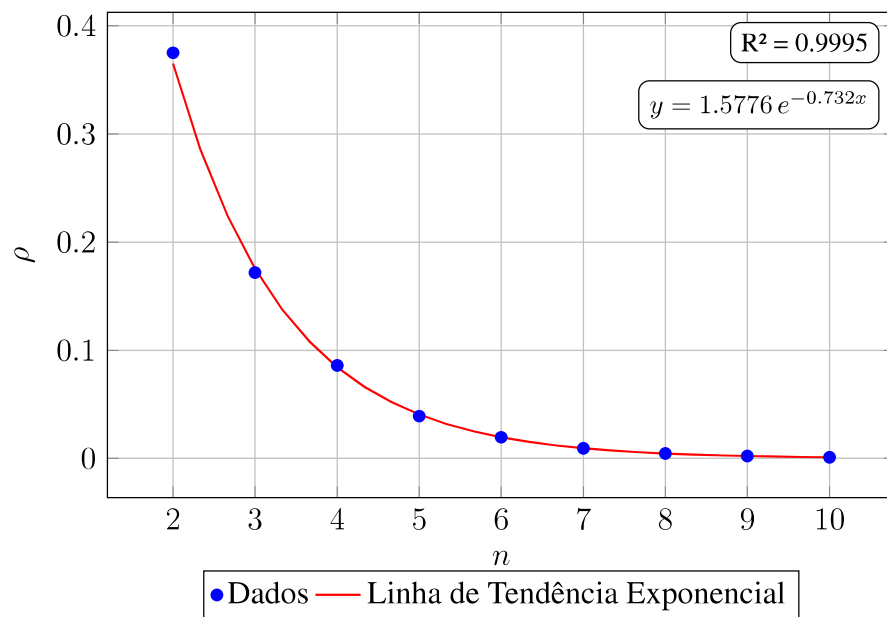


Figura 5.1: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 95%.
Fonte: Autora

- Corte em 90% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$.

Repetimos a análise precedente para probabilidade de corte em 90%. Podemos observar que o padrão acima se repete, onde a Tabela 5.3 e o gráfico apresentado na Figura 5.2 mostram o decaimento exponencial das probabilidades conforme o número n de aminoácidos nas cadeias peptídicas aumenta. Para um corte de 90%, a linha de tendência exponencial ($y = 1.4195e^{-0.782x}$, $R^2 = 0.9995$) confirma que o estudo segue um padrão estatístico já observado.

Tabela 5.3: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 90%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,3125 = 5/16$
3	$0,140625 = 9/64$
4	$0,05859375 = 15/256$
5	$0,02734375 = 28/1024$
6	$0,01293945 = 53/4096$
7	$0,00579833 = 95/16384$
8	$0,00274658 = 180/65536$
9	$0,00128555 = 337/262144$
10	$0,00057888 = 607/1048576$

Fonte: Autora

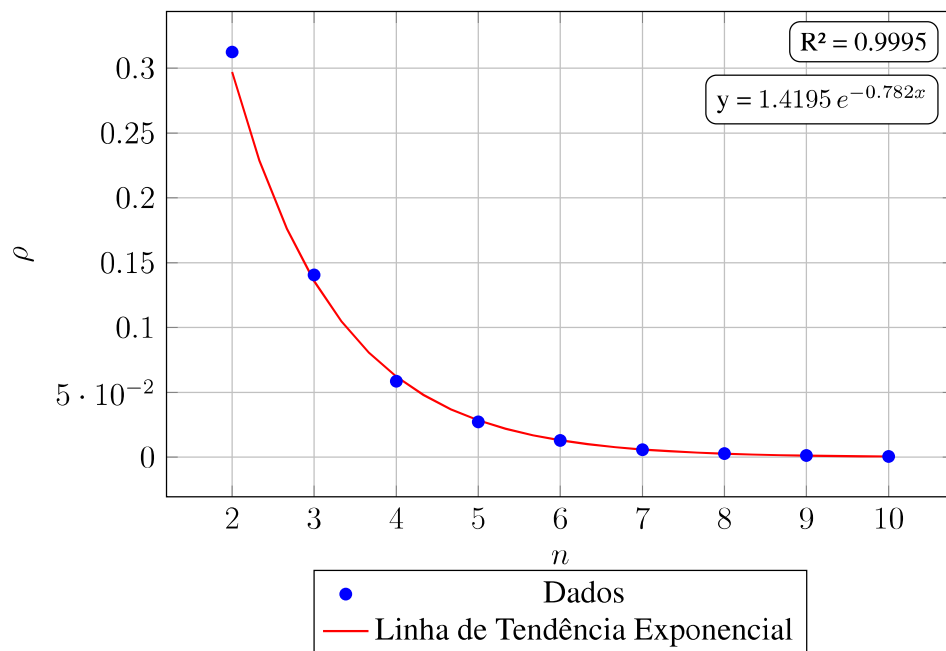


Figura 5.2: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 90 %.
Fonte: Autora

- Corte em 80% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$.

Nesse corte, também observamos um padrão de decaimento exponencial nos valores da razão ρ . A Tabela 5.4 mostra os valores calculados e a Figura 5.3 ilustra esses valores. A linha de tendência ajustada ($y = 1,0255e^{-0,819x}$) reforça a regularidade estatística observada, com um coeficiente de determinação $R^2 = 0,997$, o que confirma que o ajuste exponencial

segue sendo adequado.

Tabela 5.4: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 80%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,1875 = 3/16
3	0,09375 = 6/64
4	0,0390625 = 10/256
5	0,01757812 = 18/1024
6	0,00732421 = 30/4096
7	0,00323486 = 53/16384
8	0,00149536 = 98/65536
9	0,00064849 = 170/262144
10	0,00028228 = 296/1048576

Fonte: Autora

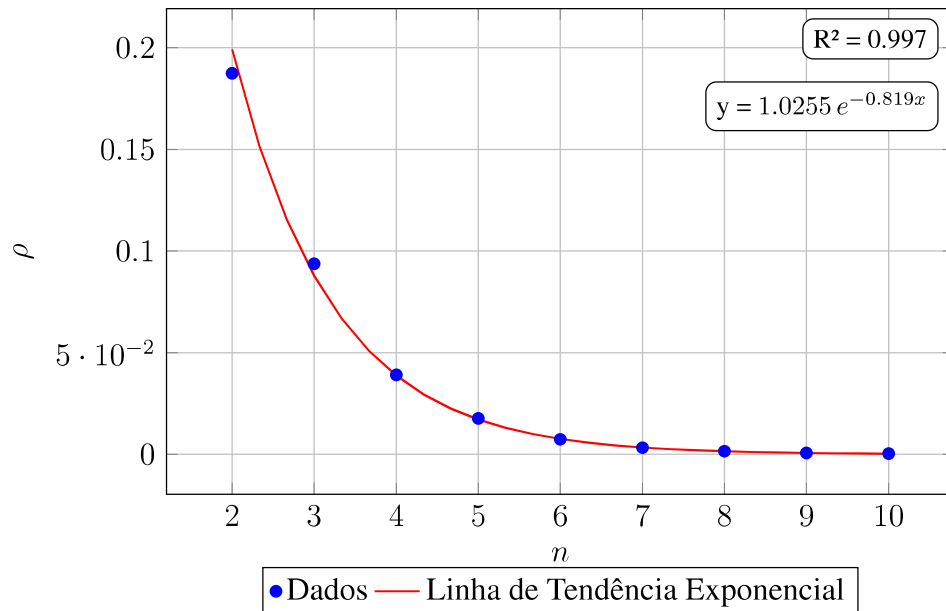


Figura 5.3: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 80 %.
Fonte: Autora

- Corte em 70% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$.

Para o corte de 70% verifica-se o mesmo padrão dos casos anteriores. Na Tabela 5.5 e na Figura 5.4 a função de ajuste exponencial utilizada neste caso ($y = 0.8946e^{-0.864x}$)

apresenta um coeficiente de determinação $R^2 = 0.9925$, mostrando que os dados continuam seguindo esse padrão.

Tabela 5.5: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 70%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,1875 = 3/16
3	0,0625 = 4/64
4	0,02734375 = 7/256
5	0,01074218 = 11/1024
6	0,00488281 = 20/4096
7	0,00207519 = 34/16384
8	0,00086975 = 57/65536
9	0,00038528 = 101/262144
10	0,00016880 = 177/1048576

Fonte: Autora

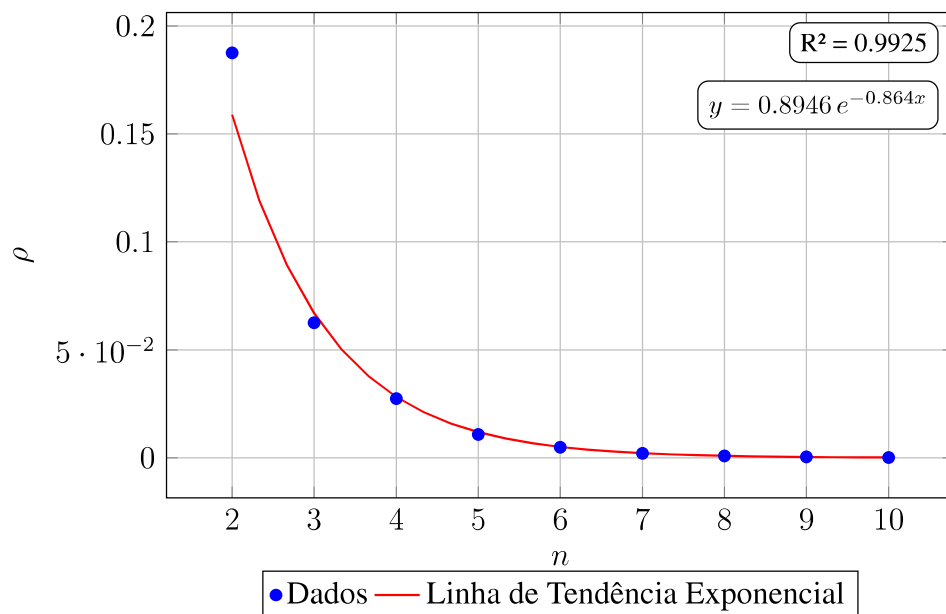


Figura 5.4: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 70 %.
Fonte: Autora

- Corte em 60% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$.

A tendência continua a mesma no corte para 60% das probabilidades, onde poucas cadeias de aminoácidos são responsáveis por uma parte significativa da probabilidade total, neste caso, 60%. A Figura 5.5 confirma como os valores da razão ρ decrescem rapidamente à medida que aumentamos o número de cadeias de aminoácidos. O ajuste exponencial ($y = 0.6875e^{-0.891x}$) com $R^2 = 0.9988$ enfatiza a consistência do comportamento observado também neste nível de corte.

Tabela 5.6: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 60%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,125 = 2/16
3	0,046875 = 3/64
4	0,01953125 = 5/256
5	0,0078125 = 8/1024
6	0,00292968 = 12/4096
7	0,00134277 = 22/16384
8	0,00056457 = 37/65536
9	0,00023269 = 61/262144
10	0,00009632 = 101/1048576

Fonte: Autora

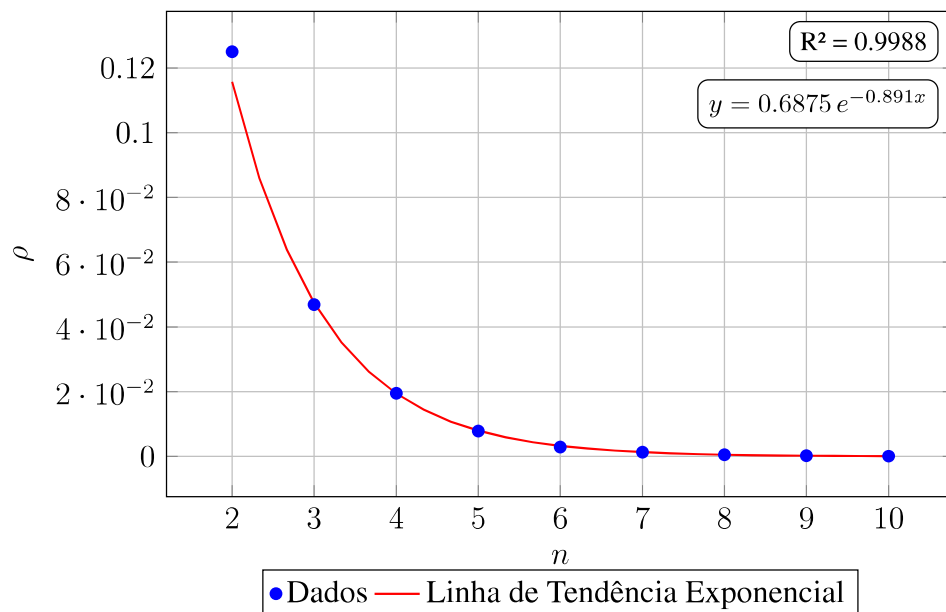


Figura 5.5: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 60 %.
Fonte: Autora

- Corte em 50% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$.

No corte de 50%, é ainda mais evidente que apenas um pequeno número de combinações de aminoácidos representam metade da probabilidade total de formação de peptídeos. Essa informação é relevante, pois indica que há um subconjunto privilegiado de cadeias peptídicas muito mais prováveis de ocorrer do que outras. O ajuste exponencial dado por ($y = 0.643e^{-0.938x}$) confirma o padrão de decaimento com um excelente coeficiente de determinação $R^2 = 0.9809$

Tabela 5.7: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 50%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,125 = 2/16
3	0,03125 = 2/64
4	0,015625 = 4/256
5	0,00585937 = 6/1024
6	0,00219726 = 9/4096
7	0,00079345 = 13/16384
8	0,00035095 = 23/65536
9	0,00014877 = 39/262144
10	0,00005912 = 62/1048576

Fonte: Autora

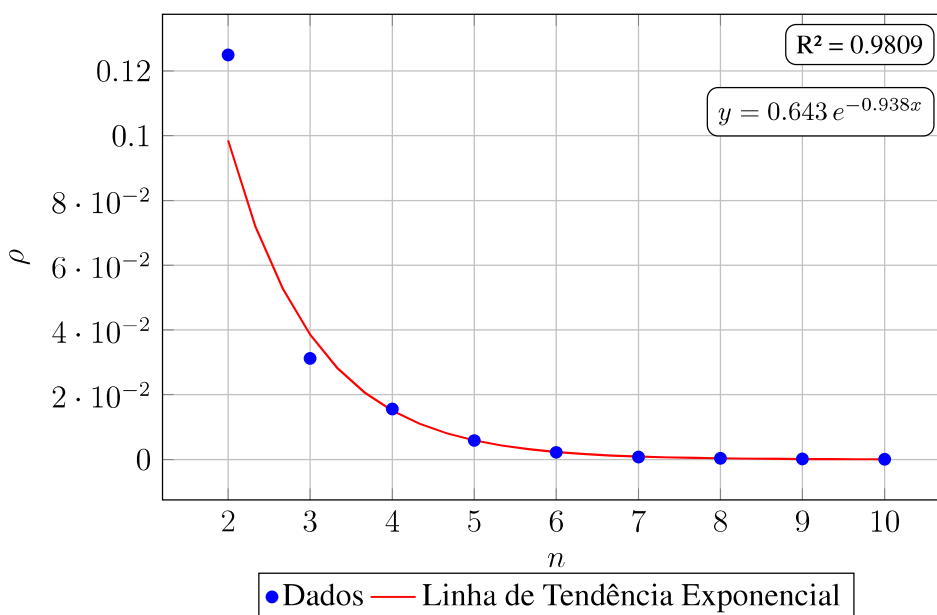


Figura 5.6: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 50 %. Fonte: Autora

- Corte em 95% para cadeias de aminoácidos GADV e estimativa para $n = 100$.

Vamos utilizar um ajuste com linha de tendência exponencial para estimar como seria a probabilidade de formação de peptídeo com $n = 100$ aminoácidos, extrapolando os resultados obtidos para o corte de probabilidade em 95%, para cadeias de 100 aminoácidos. Note que para o corte de probabilidade em 95% o ajuste da linha de tendência exponencial teve um coeficiente de determinação $R^2 = 0,9995$. Essa qualidade do ajuste torna aceitável utilizarmos os resultados da Tabela 4.1 e o gráfico de ajuste exponencial na Figura 4.1 para extrapolar resultados para cadeias com 100 aminoácidos, que se caracterizam como proteínas. A Tabela 5.8 apresenta esse resultado.

Tabela 5.8: Razão de relevância para a formação de cadeias de 100-aminoácidos (proteínas) em ambiente terrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,375 = 6/16$
3	$0,171875 = 11/64$
4	$0,0859375 = 22/256$
5	$0,0390625 = 40/1024$
6	$0,01953125 = 80/4096$
7	$0,00939941 = 154/16384$
8	$0,00450134 = 295/65536$
9	$0,00216901 = 577/262144$
10	$0,00104312 = 1094/1048576$
.	...
.	...
.	...
100	$2,32 \cdot 10^{-32} = 3,728 \cdot 10^{28} / 1,607 \cdot 10^{60}$

Fonte: Autora

Podemos observar que a razão de relevância ρ de formação de proteínas com 100 aminoácidos é $\rho = 2,32 \cdot 10^{-32}$.

5.3 RESULTADOS COM 4 AMINOÁCIDOS: ABUNDÂNCIA EXTRATERRESTRE

Também realizamos os cálculos utilizando a abundância relativa de aminoácidos de origem extraterrestre, isto é, a concentração de aminoácidos que chegaram à Terra por meio de meteoros, meteoritos e cometas. O cálculo utilizado é o mesmo descrito na seção 4.1, com a diferença de que agora vamos utilizar os dados da abundância dos aminoácidos extraterrestre, dados na Tabela 3.4. Já os valores dos rendimentos de concentrações de dipeptídeos continuam os mesmo, visto que, apesar que os aminoácidos sejam extraterrestres, a síntese de peptídeos acontece na Terra. Abaixo apresentamos os cortes de probabilidades para 95%, 90%, 80%, 70%, 60% e 50% da probabilidade total.

- Corte em 95% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$

Neste corte de 95% da probabilidade total, observamos um valor da razão de relevância ρ levemente elevada em relação à seção 4.2, mas novamente houve um decaimento exponencial da probabilidade de formação de peptídeos à medida que o tamanho da cadeia de

n -aminoácidos aumenta. O ajuste com $R^2 = 0,9965$ indica que o modelo ($y = 2,4525 \cdot e^{-0,721 \cdot x}$) representa bem os dados. Esse resultado mostra que, com base na abundância extraterrestre, a razão de relevância ρ diminui conforme o número de aminoácidos nas cadeias peptídicas aumentam.

Tabela 5.9: Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)	Tempo para execução do código (em segundos)
2	$0,625 = 10/16$	0,026983
3	$0,265625 = 17/64$	0,046275
4	$0,12890625 = 33/256$	0,069192
5	$0,06445312 = 66/1024$	0,188121
6	$0,03100585 = 127/4096$	0,494543
7	$0,01501464 = 246/16384$	1,888938
8	$0,00732421 = 480/65536$	7,081056
9	$0,00356292 = 934/262144$	28,016297
10	$0,00170803 = 1791/1048576$	115,673763

Fonte: Autora

Observa-se comportamento semelhante no que se refere ao tempo de execução do algoritmo. O tempo computacional apresenta crescimento exponencial ajustado pela equação $y = 0,0014e^{1,0706x}$, com $R^2 = 0,9889$, tornando inviável a realização dos cálculos para n crescentes no computador utilizado neste trabalho. Por exemplo, de acordo com o ajuste exponencial do tempo computacional, para $n = 12$, o tempo computacional seria de $t = 531s$ e para $n = 15$, seria $t = 13224s$.

Assim como anteriormente, os demais cortes de probabilidade desta seção (90%, 80%, 70%, 06% e 50%) apresentam demanda computacional equivalente, uma vez que o procedimento algorítmico permanece inalterado.

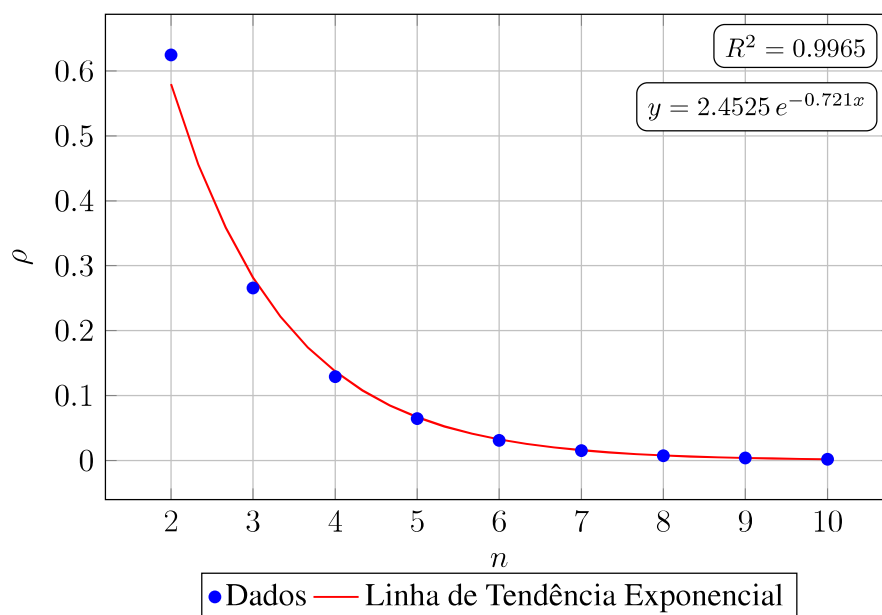


Figura 5.7: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 95%. Fonte: Autora

- Corte em 90% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$

Com 90% do total das probabilidade de formação das cadeias de aminoácidos, a tendência de decaimento se mantém como mostrado na Tabela 5.10 e na Figura 5.8. O ajuste exponencial ($y = 2,0235 \cdot e^{-0,785 \cdot x}$) tem $R^2 = 0,999$. Isso mostra que, embora o corte seja menor (90%), a relação de relevância segue a mesma lógica e reforça a consistência da ideia que apenas a pequena parte do total das possibilidades representam uma parte relevante da probabilidade total de formação de peptídeos.

Tabela 5.10: Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 90%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,4375 = 7/16
3	0,1875 = 12/64
4	0,08203125 = 21/256
5	0,03808593 = 39/1024
6	0,01782226 = 73/4096
7	0,00823974 = 135/16384
8	0,00369262 = 242/65536
9	0,00165939 = 435/262144
10	0,00075531 = 792/1048576

Fonte: Autora

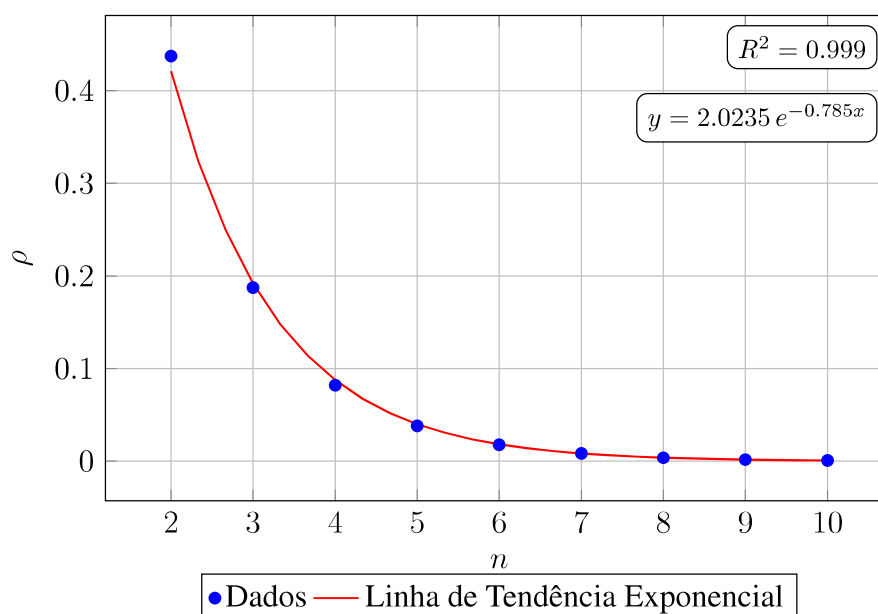


Figura 5.8: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 90%. Fonte: Autora

- Corte em 80% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$

No corte de 80%, poucas cadeias já concentram quase toda a probabilidade desejada, neste caso, 80%. A Tabela 5.11 e a Figura 5.9 mostram os valores da razão de relevância ρ . Vê-se que mesmo com um corte ainda relativamente alto, a grande maioria das combinações de aminoácidos possíveis tem uma contribuição muito pequena no total das probabilidades de formação de peptídeos. O modelo exponencial ajustado foi $y = 1,4143 \cdot e^{-0,848 \cdot x}$ com $R^2 = 0.9996$.

Tabela 5.11: Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 80%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,25 = 4/16
3	0,109375 = 7/64
4	0,046875 = 12/256
5	0,020507 = 21/1024
6	0,008544 = 35/4096
7	0,00366210 = 60/16384
8	0,00155639 = 102/65536
9	0,00066757 = 175/262144
10	0,00027942 = 293/1048576

Fonte: Autora

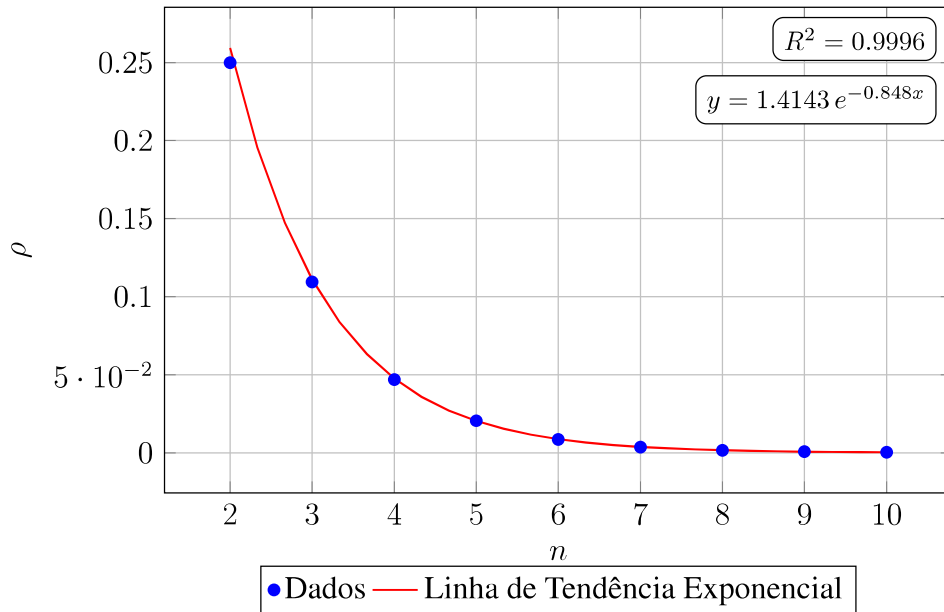


Figura 5.9: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 80%. Fonte: Autora

- Corte em 70% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$

Considerando agora 70% da probabilidade total, podemos observar na Figura 5.10 que o comportamento se repete, ou seja, conforme o número n de aminoácidos na cadeia peptídica aumenta, o valor da razão de relevância ρ diminui, indicando que a maioria das combinações de aminoácidos apresentam valores de probabilidade irrelevantes para a probabilidade total.

Mesmo as cadeias com apenas 3 ou 4 aminoácidos já têm os valores da razão de relevância ρ muito reduzidas, mas conforme o número de cadeias de aminoácidos aumentam, o valor da razão ρ é cada vez menor. Isso reforça a ideia de que, em um ambiente prebiótico, pouquíssimas combinações de aminoácidos seriam favorecidas. O modelo exponencial ajustado foi $y = 1,1993 \cdot e^{-0,915 \cdot x}$ com $R^2 = 0.9997$.

Tabela 5.12: Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 70%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,1875 = 3/16
3	0,078125 = 5/64
4	0,03125 = 8/256
5	0,01269531 = 13/1024
6	0,00488281 = 20/4096
7	0,00189208 = 31/16384
8	0,00076293 = 50/65536
9	0,00030899 = 81/262144
10	0,00012588 = 132/1048576

Fonte: Autora

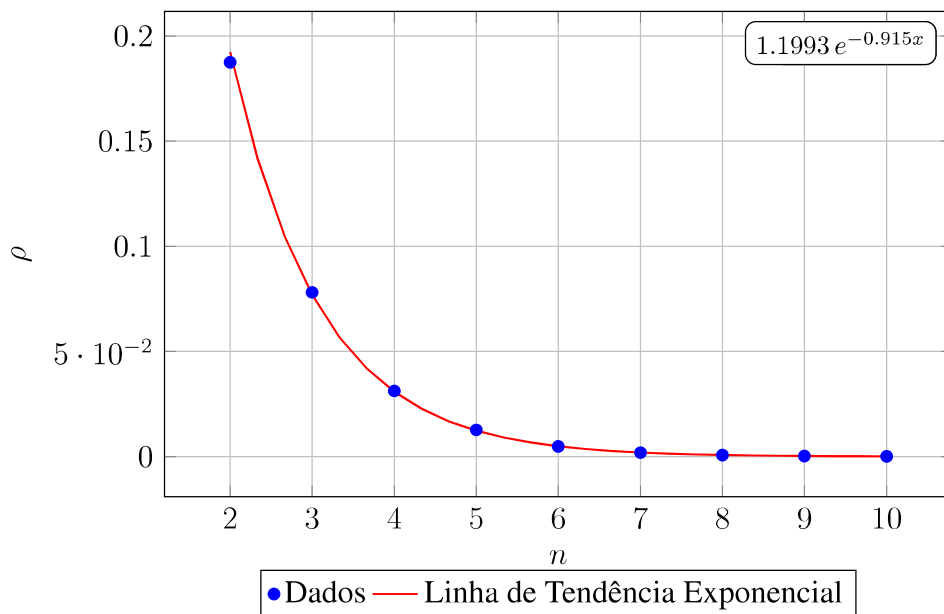


Figura 5.10: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 70%. Fonte: Autora

- Corte em 60% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$

No corte de 60% das probabilidades totais, fica cada vez mais evidente que poucas cadeias de aminoácidos concentram quase toda a probabilidade desejada. Pode-se observar na Tabela 5.13 e na Figura 5.11, conforme n cresce, que o modelo expressa um comportamento altamente seletivo, favorecendo fortemente um número muito limitado e pequeno

de cadeias de aminoácidos. O modelo exponencial ajustado foi $y = 0,9765 \cdot e^{-0,969 \cdot x}$ com $R^2 = 0.9873$.

Tabela 5.13: Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 60%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,125 = 2/16$
3	$0,0625 = 4/64$
4	$0,01953125 = 5/256$
5	$0,0078125 = 8/1024$
6	$0,00292968 = 12/4096$
7	$0,00109863 = 18/16384$
8	$0,00041198 = 27/65536$
9	$0,00015258 = 40/262144$
10	$0,00005912 = 62/1048576$

Fonte: Autora

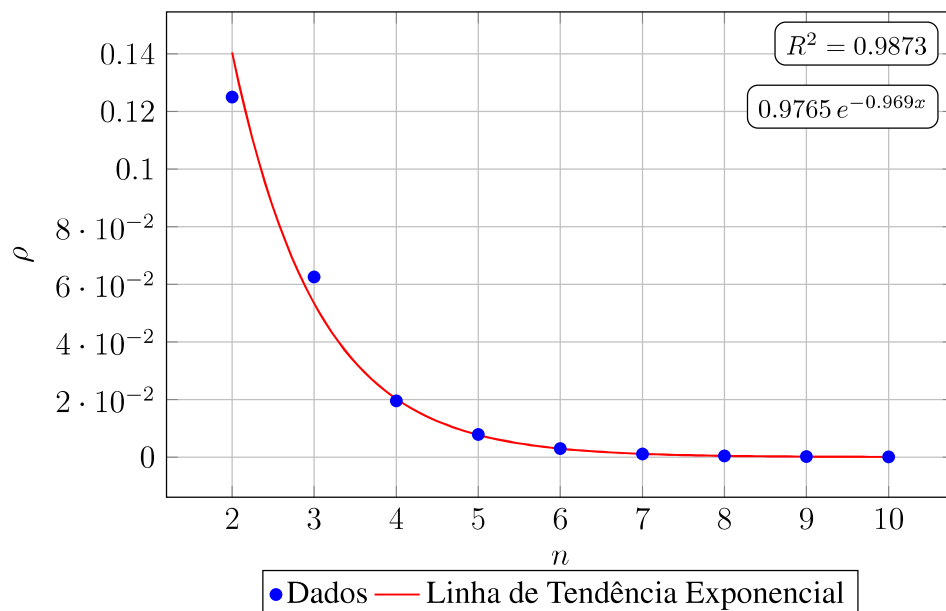


Figura 5.11: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 60%. Fonte: Autora

- Corte em 50% para cadeias de aminoácidos GADV com $2 \leq n \leq 10$

No corte de 50%, o padrão visto nos estudos anteriores se reforça. A metade da probabilidade total de formação dos peptídeos está concentrada em um número extremamente reduzido de cadeias de aminoácidos, enquanto a maior parte das outras combinações de aminoácidos possuem valores extremamente baixos de probabilidade de ocorrer, praticamente desprezíveis. O modelo exponencial ajustado foi $y = 1,0032 \cdot e^{-1,047 \cdot x}$ com $R^2 = 0.9994$.

Tabela 5.14: Razão de relevância para a formação de peptídeos em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 50%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,125 = 2/16
3	0,046875 = 3/64
4	0,015625 = 4/256
5	0,00488281 = 5/1024
6	0,00170898 = 7/4096
7	0,00061035 = 10/16384
8	0,00022888 = 15/65536
9	0,00008010 = 21/262144
10	0,00002956 = 31/1048576

Fonte: Autora

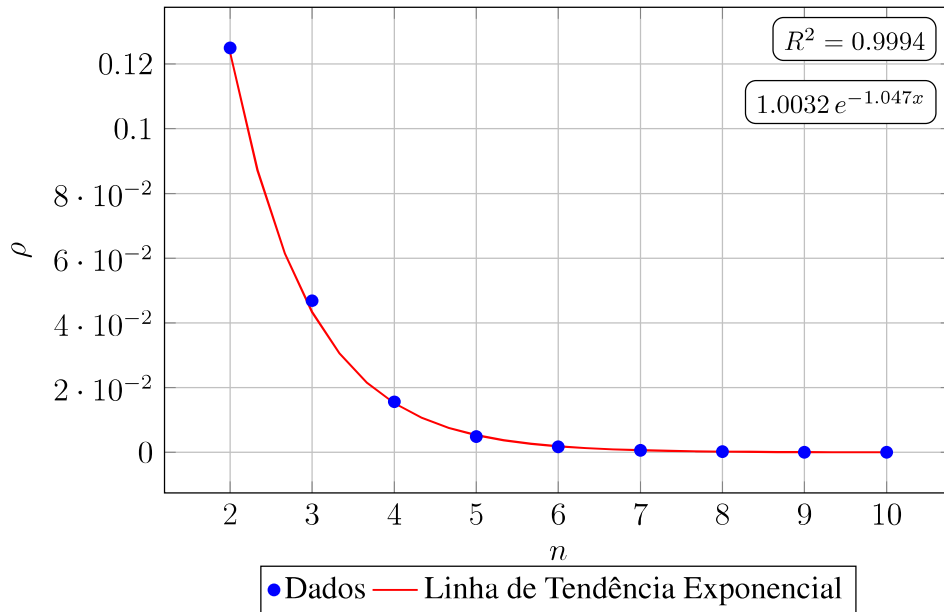


Figura 5.12: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente extraterrestre, considerando os 4 aminoácidos GADV e corte de probabilidade em 50%. Fonte: Autora

- Corte em 95% para cadeias de aminoácidos GADV e estimativa para $n = 100$.

Como fizemos na seção 4.2, também usaremos a ferramenta de linha de tendência exponencial para extrapolar os valores da Tabela 5.9, para cadeias de 100 aminoácidos (proteínas). A linha de tendência exponencial $y = 2,4525 \cdot e^{-0,721 \cdot x}$ ajustada na Figura 5.7 tem um coeficiente de determinação $R^2 = 0.9965$, o que torna aceitável extrapolar os resultados da Tabela 5.9. Assim obtemos os dados da Tabela 5.15.

Comparando as Tabelas 5.8 e 5.15, observamos que os valores obtidos da razão de relevância ρ são minimamente maiores para os casos de cadeias de aminoácidos extraterrestres em comparação com os casos de abundância terrestre, isso se dá devido às diferenças nas abundâncias relativas desses aminoácidos. Contudo é notório que, por mais que em cadeias peptídicas com aminoácidos extraterrestres os valores obtidos da razão de relevância ρ são minimamente maiores, a formação de cadeias longas, como com 100 aminoácidos, por exemplo, continuam tendo uma valor da razão de relevância ρ extremamente baixos em ambos os cenários.

Tabela 5.15: Razão de relevância para a formação de cadeias de 100-aminoácidos (proteínas) em ambiente extraterrestre, a partir dos 4 aminoácidos GADV, considerando corte de probabilidade em 95%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,5533 = 9/16$
3	$0,2688 = 17/64$
4	$0,1306 = 33/256$
5	$0,0634 = 65/1024$
6	$0,0308 = 126/4096$
7	$0,0150 = 245/16384$
8	$0,0073 = 478/65536$
9	$0,0035 = 917/262144$
10	$0,0017 = 1782/1048576$
.	...
.	...
.	...
100	$1,03 \cdot 10^{-31} = 1,65 \cdot 10^{29} / 1,607 \cdot 10^{60}$

Fonte: Autora

5.4 RESULTADOS COM 5 AMINOÁCIDOS: INCLUSÃO DA LEUCINA

Para finalizar as simulações numéricas, ampliamos os cálculos acrescentando mais um aminoácido, a Leucina, que tem uma abundância de 0,05 [3], elevando a quantidade de possíveis combinações de cadeias peptídicas para 5^n . Queremos observar qual o padrão da razão de relevância ρ devido a inclusão de um novo aminoácido. Quanto à escolha da Leucina, isso se deve ao fato de existirem dados experimentais necessários para nossos cálculos.

Repetimos os mesmos procedimentos das simulações realizadas acima, mas agora com um aminoácido a mais. A tabela de rendimentos agora é dada pela Tabela 5.16.

- Corte em 95% para cadeias de aminoácidos GAVD+Leucina com $2 \leq n \leq 10$

Acrescentando a Leucina ao conjunto de aminoácidos, a distribuição de probabilidades se torna mais complexa, pelo fato de o número total combinações aumentar. No corte de 95%, a partir de cadeias com 5 ou 6 aminoácidos, os valores da razão de relevância ρ se reduzem de forma significativa. Podemos ver na Tabela 5.17 e na Figura 5.16 que o ajuste exponencial ($y = 1,7815 \cdot e^{-0,953 \cdot x}$) com $R^2 = 0,9997$ permanece excelente com a inclusão da Leucina nas cadeias de n -aminoácidos.

Tabela 5.16: Valores normalizados do rendimento em % de dipeptídeos AB formados a partir da massa original de 5 aminoácidos após o quinto ciclo

$A \downarrow / B \rightarrow$	GLY	ALA	ASP	VAL	LEU
GLY	1,00	0,139	0,037	0,312	0,447
ALA	0,182	0,283	0,056	0,178	0,030
ASP	0,111	0,029	0,058	0,023	0,015
VAL	0,135	0,040	0,003	0,1461	0,024
LEU	0,088	0,041	0,016	0,057	0,045

Fonte: Adaptado Rode [1], p. 311.

Tabela 5.17: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 95%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)	Tempo para execução do código (em segundos)
2	$0,28 = 7/25$	0,029965
3	$0,104 = 13/125$	0,041633
4	$0,0384 = 24/625$	0,079729
5	$0,0144 = 45/3125$	0,233287
6	$0,00576 = 90/15625$	0,887886
7	$0,002176 = 170/78125$	4,058331
8	$0,00086016 = 336/390625$	19,889247
9	$0,00033228 = 649/1953125$	150,028431
10	$0,00013752 = 1343/9765625$	1512,315892

Fonte: Autora

Observa-se que, com a inclusão de um quinto aminoácido, os tempos de execução permanecem da mesma ordem de grandeza dos casos anteriores para valores menores de n . As diferenças tornam-se mais perceptíveis apenas para $n = 9$, quando o aumento do número de combinações possíveis passa a impactar de forma mais evidente o tempo de processamento.

Esse comportamento está associado ao crescimento combinatório do número de cadeias possíveis à medida que se aumenta tanto o tamanho da cadeia quanto o conjunto de aminoácidos considerados. Vejamos, o ajuste de crescimento do tempo computacional é do tipo $y = 0,0005e^{1,363x}$, com $R^2 = 0,9752$, tornando inviável a realização dos cálculos para n crescentes no computador utilizado neste trabalho. Por exemplo, de acordo com o ajuste

exponencial do tempo computacional, para $n = 12$, o tempo computacional seria de $t = 6340s$ e para $n = 15$, seria $t = 378000s$.

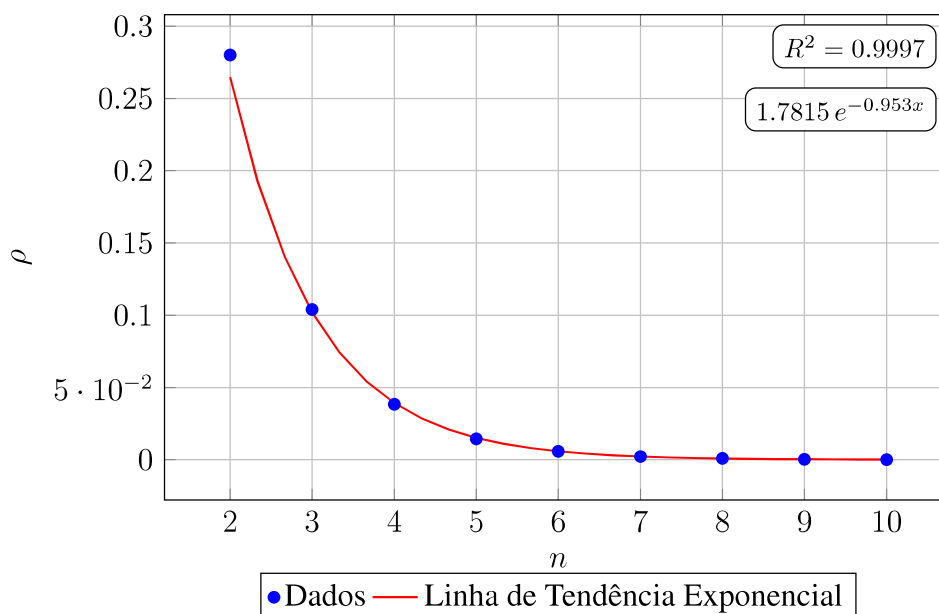


Figura 5.13: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 95%. Fonte: Autora

- Corte em 90% para cadeias de aminoácidos GAVD+Leucina com $2 \leq n \leq 10$

Para o corte de 90% das probabilidades totais de formação de cadeias peptídicas, como visto na Tabela 5.18 e na Figura 5.14, as razões de relevância associadas as cadeias com 6 ou mais aminoácidos já são bastante reduzidas, indicando que com o acréscimo de um novo aminoácido, o modelo continua favorecendo apenas uma pequena fração das possíveis combinações de aminoácidos. O modelo exponencial ajustado foi $y = 1,3822 \cdot e^{-0,99 \cdot x}$ com $R = 0.9998$.

Tabela 5.18: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 90%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,2 = 5/25$
3	$0,072 = 9/125$
4	$0,0256 = 16/625$
5	$0,0096 = 30/3125$
6	$0,003584 = 56/15625$
7	$0,0012928 = 101/78125$
8	$0,00050432 = 197/390625$
9	$0,00018636 = 364/1953125$
10	$0,00007301 = 713/9765625$

Fonte: Autora

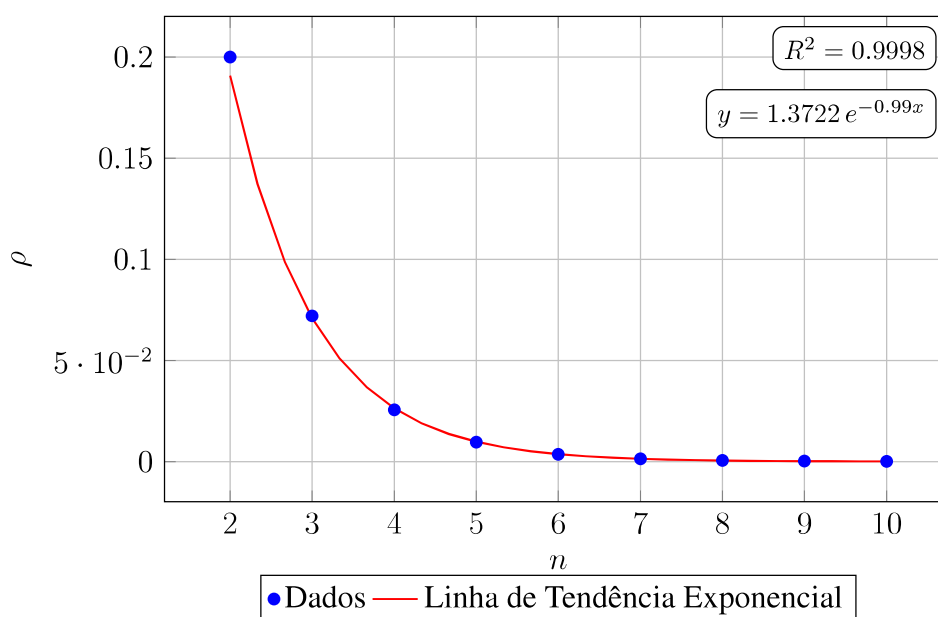


Figura 5.14: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 90%. Fonte: Autora

- Corte em 80% para cadeias de aminoácidos GAVD+Leucina com $2 \leq n \leq 10$

Para o corte de 80% das probabilidades totais, o comportamento seletivo que o modelo expressa permanece: os valores da razão de relevância ρ decaem, como pode ser observado na Tabela 5.19 e na Figura 5.15. A razão de relevância ρ com 9 aminoácidos é muito baixa, o que demonstra que a grande maioria das cadeias possíveis possui probabilidades

praticamente nulas de ocorrer. Mesmo com mais combinações disponíveis devido ao quinto aminoácido incluído nos cálculos, poucas são efetivamente prováveis de ocorrer. O modelo exponencial ajustado foi

$$y = 1,1501 \cdot e^{-1,049 \cdot x}$$

com $R^2 = 0.9998$.

Tabela 5.19: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 80%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,16 = 4/25
3	0,048 = 6/125
4	0,016 = 10/625
5	0,00608 = 19/3125
6	0,001984 = 31/15625
7	0,0007168 = 56/78125
8	0,00026368 = 103/390625
9	0,00009113 = 178/1953125
10	0,00003420 = 334/9765625

Fonte: Autora

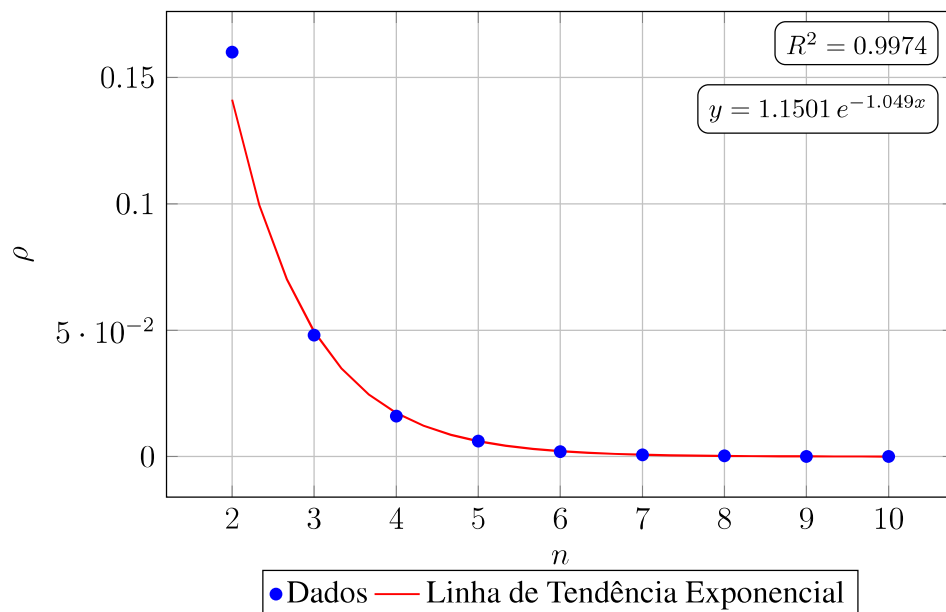


Figura 5.15: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 80%. Fonte: Autora

- Corte em 70% para cadeias de aminoácidos GAVD+Leucina com $2 \leq n \leq 10$

Com o corte de 70% da probabilidades totais, pode ser observado da Tabela 5.20 e da Figura 5.16 que o decaimento da razão de relevância ρ mantém um padrão exponencial ($y = 0,8875 \cdot e^{-1,08 \cdot x}$ com $R = 0,9948$).

Tabela 5.20: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 70%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,12 = 3/25
3	0,032 = 4/125
4	0,0112 = 7/625
5	0,00384 = 12/3125
6	0,001344 = 21/15625
7	0,0004608 = 36/78125
8	0,00014848 = 58/390625
9	0,00005376 = 105/1953125
10	0,00001966 = 192/9765625

Fonte: Autora

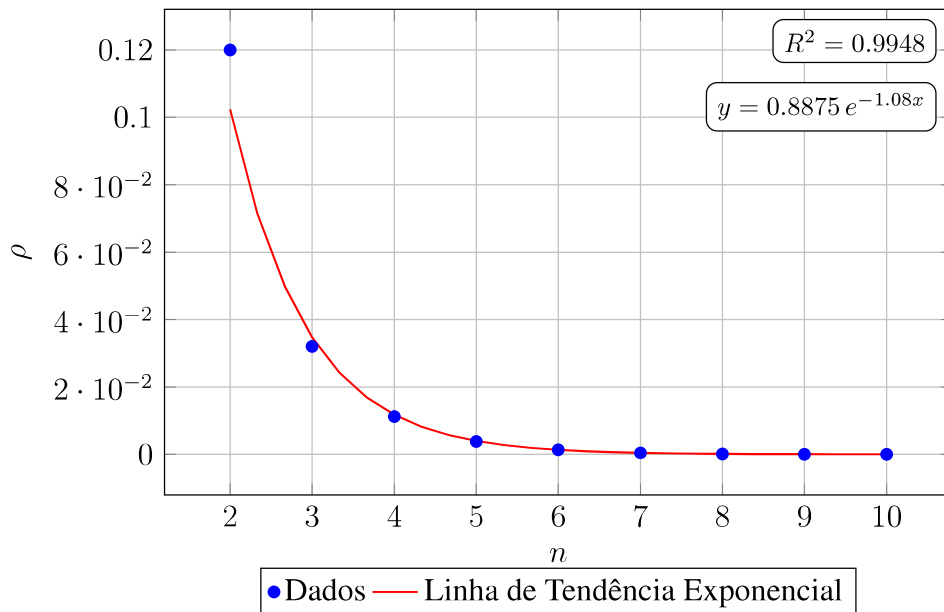


Figura 5.16: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 70%. Fonte: Autora

- Corte em 60% para cadeias de aminoácidos GAVD+Leucina com $2 \leq n \leq 10$

Neste caso, para 60% das probabilidade totais, observa-se o mesmo padrão anterior, ou seja, conforme o número n de aminoácidos na cadeia peptídica aumenta, os valores da razão de relevância ρ decaem, como pode ser observado na Tabela 5.21 e na Figura 5.17. O modelo exponencial ajustado foi $y = 0,6699 \cdot e^{-1,104 \cdot x}$ com $R = 0.999$.

Tabela 5.21: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 60%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	0,08 = 2/25
3	0,024 = 3/125
4	0,008 = 5/625
5	0,00256 = 8/3125
6	0,000832 = 13/15625
7	0,0002944 = 23/78125
8	0,00009984 = 39/390625
9	0,000032256 = 63/1953125
10	0,00001126 = 110/9765625

Fonte: Autora

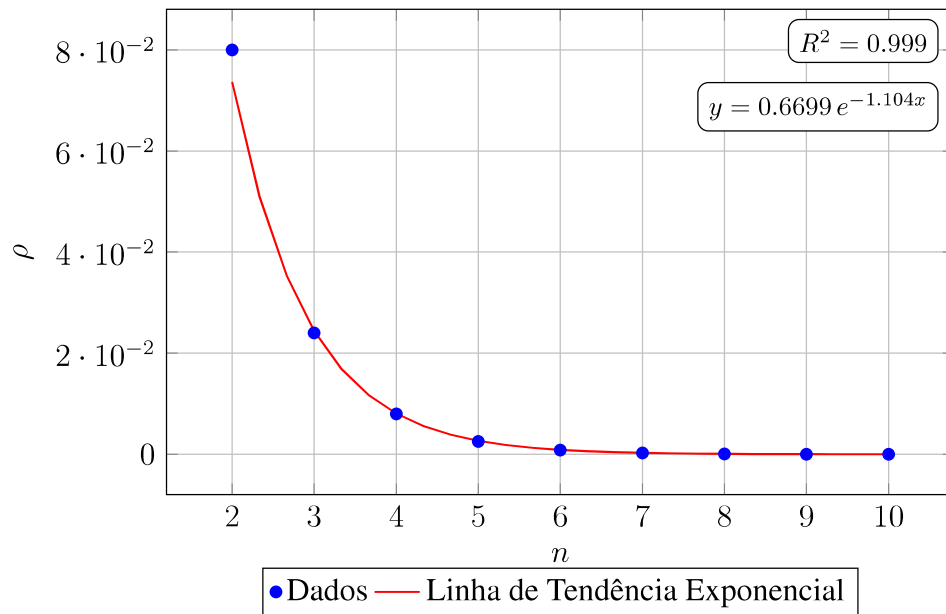


Figura 5.17: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 60%. Fonte: Autora

- Corte em 50% para cadeias de aminoácidos GAVD+Leucina com $2 \leq n \leq 10$

No corte de 50% das probabilidades totais, podemos observar na Tabela 5.22 e na Figura 5.18 que o valor da razão de relevância ρ diminui novamente, mantendo um comportamento seletivo que persiste mesmo com maior diversidade de combinações, agora com 5 aminoácidos.

Tabela 5.22: Razão de relevância para a formação de peptídeos em ambiente terrestre, a partir de 5 aminoácidos, considerando corte de probabilidade em 50%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,08 = 2/16$
3	$0,016 = 2/125$
4	$0,0064 = 4/625$
5	$0,00192 = 6/3125$
6	$0,000576 = 9/15625$
7	$0,0001792 = 14/78125$
8	$0,00006144 = 24/390625$
9	$0,00002048 = 40/1953125$
10	$0,00000665 = 65/9765625$

Fonte: Autora

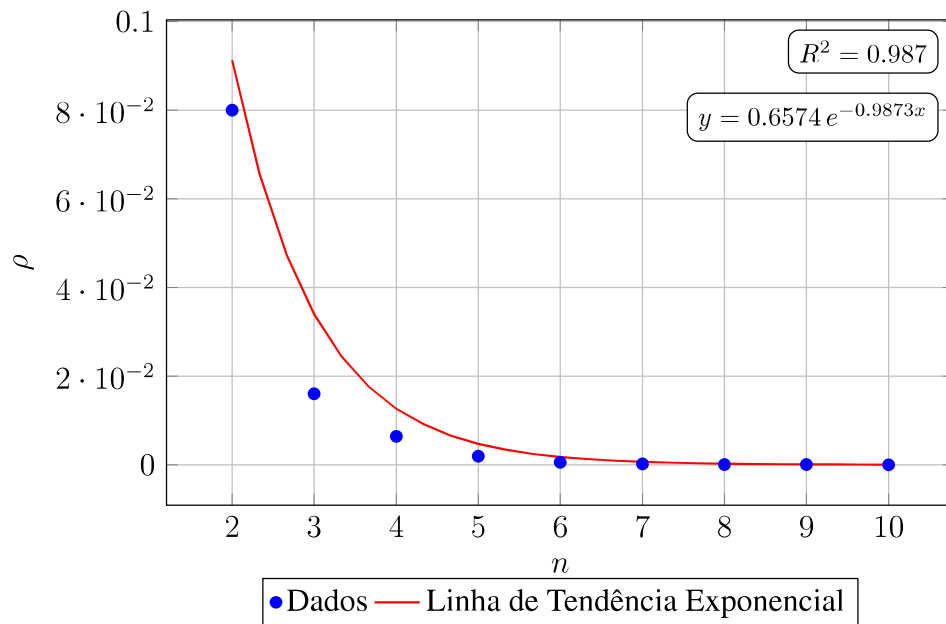


Figura 5.18: Gráfico da razão de relevância ρ , em função do tamanho n da cadeia peptídica, em ambiente terrestre, considerando 5 aminoácidos e corte de probabilidade em 50%. Fonte: Autora

- Corte em 95% para cadeias de aminoácidos GAVD+Leucina e estimativa para $n = 100$

Como fizemos anteriormente para 4 aminoácidos, realizamos novamente uma estimativa de corte de 95%, mas para cadeias com $n = 100$ aminoácidos (proteínas). De maneira análoga, extrapolamos os valores da Tabela 5.17, utilizando a linha de tendência encontrada no gráfico, visto que foi obtido um coeficiente de determinação $R^2 = 0.9997$.

Tabela 5.23: Razão de relevância para a formação de cadeias de 100-aminoácidos (proteínas) em ambiente terrestre, com aminoácidos, considerando corte de probabilidade em 95%.

n (número de aminoácidos na cadeia)	ρ (razão de relevância)
2	$0,264 = 6/25$
3	$0,102 = 12/125$
4	$0,040 = 25/625$
5	$0,015 = 46/3125$
6	$0,0058 = 90/15625$
7	$0,0023 = 179/78125$
8	$0,00088 = 343/390625$
9	$0,00034 = 664/1953125$
10	$0,00013 = 1269/9765625$
.	...
.	...
.	...
100	$5,65 \cdot 10^{-42} = 4,46 \cdot 10^{28} / 7,9 \cdot 10^{69}$

Fonte: Autora

Podemos notar com a inclusão da Leucina que há um padrão análogo ao caso de 4 aminoácidos: um decaimento exponencial nos valores da razão de relevância ρ conforme o número n de aminoácidos aumentam nas cadeias peptídicas. A extrapolação para 100 aminoácidos reforça uma razão ρ extremamente baixa, o que é uma tendência que já vimos anteriormente.

5.5 ANÁLISE COMPARATIVA DAS RAZÕES DE RELEVÂNCIA E CORTES DE PROBABILIDADE

Um resultado obtido ao considerarmos 5 aminoácidos é que o fator de relevância diminui, se comparado com os valores obtidos para cadeias com 4 aminoácidos, como pode ser observado na Tabela 5.24. Note que ao considerarmos cadeias peptídicas com 5 aminoácidos, o número de possíveis cadeias aumenta consideravelmente, enquanto o número de cadeias probabilisticamente relevantes se mantém da mesma ordem dos valores obtidos com 4 aminoácidos.

Tabela 5.24: Comparação das razões de relevância (ρ) nos três cenários com corte de probabilidade de 95%.

n (número de aminoácidos na cadeia)	Terrestre	Extraterrestre	Com Leucina
2	0,37500	0,62500	0,28000
3	0,17187	0,26562	0,10400
4	0,08593	0,12890	0,03840
5	0,03906	0,06445	0,01440
6	0,01953	0,03100	0,00576
7	0,00933	0,01501	0,00217
8	0,00451	0,00732	0,00086
9	0,00220	0,00352	0,00033
10	0,00104	0,00170	0,00013

Fonte: Autora

Analizamos agora especificamente o número de cadeias peptídicas formadas por 10 aminoácidos terrestres GADV necessárias para atingir diferentes cortes percentuais da probabilidade total acumulada. Como estamos considerando uma simulação com cadeias com 10 aminoácidos GADV, o número de combinações possíveis a serem consideradas é de 1.048.576. Queremos avaliar como a variação do corte percentual influencia a quantidade de cadeias peptídicas necessárias para representar tal parcela da probabilidade total. A Tabela 5.25 apresenta o número de combinações de aminoácidos probabilisticamente relevantes para cada nível de corte, enquanto a Figura 5.19 representa graficamente esses valores e o ajuste exponencial correspondente.

Tabela 5.25: Evolução do número de cadeias peptídicas relevantes em função do corte percentual, considerando ambiente terrestre e os 4 aminoácidos GADV

corte	número de cadeias peptídicas relevantes
50%	62
60%	101
70%	177
80%	296
90%	607
95%	1091

Fonte: Autora

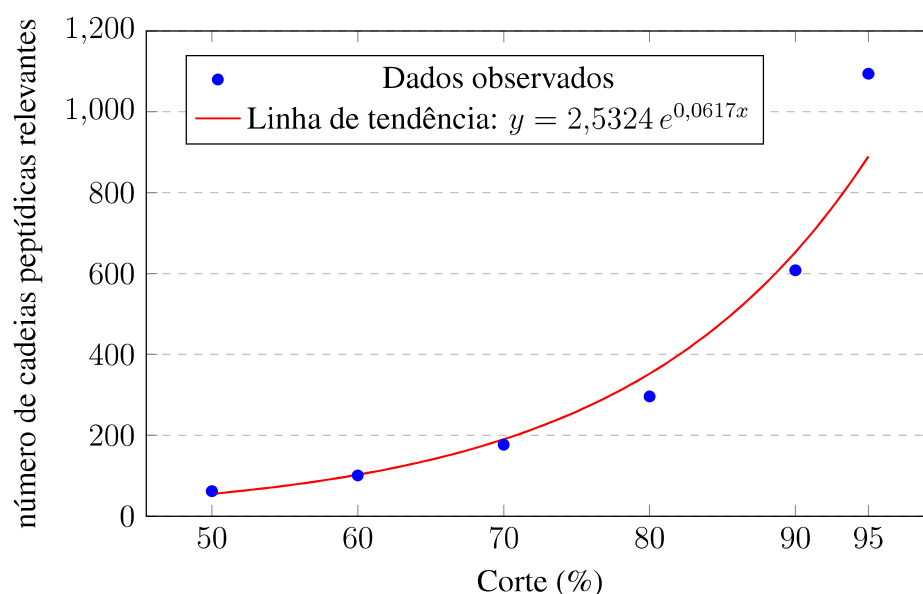


Figura 5.19: Evolução do número de peptídeos relevantes, considerando ambiente terrestre e os 4 aminoácidos GADV. Fonte: Autora

Observamos um crescimento não linear, de fato realizamos um ajuste com linha de tendência exponencial no número de combinações de aminoácidos relevantes à medida que o corte aumenta. Por exemplo, para um corte de 50% das cadeias peptídicas mais prováveis, obtivemos 62 cadeias peptídicas relevantes em 1.048.576 possíveis cadeias, enquanto com um corte em 95% o número de cadeias peptídicas relevantes subiu para 1094 cadeias peptídicas relevantes. Esse comportamento sugere uma relação exponencial entre o percentual de corte e o número de cadeias peptídicas relevantes, o que pode ser observado pela linha de tendência ajustada ($y = 2,5324 \cdot e^{0,0617x}$ com $R^2 = 0,9624$) na Figura 5.19.

6 CONCLUSÃO

Desenvolvemos neste trabalho um modelo matemático probabilístico para estimar a probabilidade de formação de peptídeos sob condições prebióticas e realizamos simulações computacionais deste modelo. Para isso, consideramos dois possíveis cenários: aminoácidos de fontes terrestres e aminoácidos de fonte extraterrestre. A escolha de quais aminoácidos investigar foi baseada na hipótese GADV, que diz que as primeiras proteínas formadas eram compostas por apenas quatro aminoácidos: Glicina, Alanina, Ácido Aspártico e Valina [23].

Os resultados das simulações nos mostraram que a probabilidade de formação dos peptídeos não ocorre de forma uniforme entre as diferentes sequências possíveis de aminoácidos. Na verdade, observamos um comportamento para a probabilidade de formação de peptídeos e proteínas caracterizado por um decaimento exponencial, em que apenas algumas sequências de aminoácidos apresentam alta probabilidade de formação, enquanto a maioria das combinações possíveis ocorre com probabilidade muito baixa. Esse padrão na probabilidade de formação de cadeias de aminoácidos foi observado para os cenários com aminoácidos terrestres, extraterrestres e também para quando adicionamos a Leucina, um quinto aminoácido nas simulações.

A reduzida quantidade de cadeias de aminoácidos prováveis de ocorrer, sugere que, mesmo em ambientes com diferentes aminoácidos e abundâncias, a taxa de formação de certos peptídeos é consideravelmente mais favoráveis. Além disso, ao extrapolarmos os resultados para cadeias com 100 aminoácidos, observamos que embora o número de sequências possíveis de aminoácidos aumente exponencialmente, uma quantidade extremamente reduzida dessas combinações representam uma porcentagem significativa da probabilidade total.

Observou-se que em cenários de aminoácidos terrestres, a razão de relevância ρ é levemente menor em relação as simulações com aminoácidos extraterrestres, isso se deve à diferença entre as abundâncias relativas entre os aminoácidos terrestres e extraterrestres. Já quando acrescentamos a Leucina nos cálculos e passamos a ter combinações entre 5 aminoácidos, os valores de relevância ρ diminuem se comparado com os casos com apenas 4 aminoácidos. Um resultado importante em todas análises realizadas neste texto, para ambientes da Terra primitiva, é que o modelo expressa um comportamento altamente seletivo das cadeias peptídicas possíveis, indicando que na Química Prebiótica ocorre um forte favorecimento de um número muito limitado e pequeno de cadeias de aminoácidos.

Enfim, algumas outras questões importantes talvez possam ser respondidas por nosso estudo probabilístico de formação de peptídeos, certamente após um maior número de simulações numéricas. Por exemplo, por que as proteínas dos seres vivos atuais são constituídas majoritariamente de 20 aminoácidos? Devido às fontes terrestres e extraterrestres de moléculas orgânicas, é possível que uma grande variedade de moléculas orgânicas já estavam presentes na Terra primitiva [34–36]. No caso específico de aminoácidos, experimentos simu-

lando ambientes terrestres prebióticos mostraram que é possível se obter uma grande variedade de aminoácidos [3]. Destacamos também que a análise de meteoritos tem mostrado uma grande variedade de aminoácidos, sendo que no meteorito Murchison foram identificados mais de 80 diferentes aminoácidos [37]. Portanto aqui temos uma pergunta, ou seja, se na Terra antes da origem da vida existia uma grande variedade de aminoácidos, por que as proteínas dos seres vivos atuais são constituídas majoritariamente de 20 aminoácidos?

Uma tentativa de resposta a pergunta acima formulada, pode ser obtida pela utilização de três diferentes informações: a) dados experimentais do rendimento de formação de dipeptídeos, b) dados de proporção de aminoácidos produzidos por experimentos simulando ambientes terrestres e extraterrestres, assim como análise de meteoritos [1,3,23] e c) verificação por meio de simulações numéricas da hipótese formulada no final da seção 4.3, ou seja, que ao considerarmos mais aminoácidos, o fator de relevância sempre diminui.

Para diversos aminoácidos, Rode obteve o rendimento experimental na formação dos dipeptídeos AB e BA [1, 2], observando valores decrescentes para esses rendimentos, o que evidencia um favorecimento para alguns aminoácidos e ligações. Quanto à proporção relativa de aminoácidos produzidos em experimentos simulando ambientes terrestres, eles fornecem um dado quantitativo da proporção de aminoácidos existentes na Terra primitiva. Também, os dados de análise de meteoritos e de experimentos simulando ambientes extraterrestres fornecem dados quantitativos da proporção de aminoácidos de fontes extraterrestres [3]. Notamos que os valores de abundância também favorecem mais alguns aminoácidos, praticamente tornando cadeias com alguns aminoácidos muito improváveis. Quanto à hipótese formulada no final da seção 4.3, ou seja, que ao considerarmos mais aminoácidos o fator de relevância diminui, mais simulações devem ser realizadas.

Outra possibilidade de estudos futuros é sobre o modelo matemático utilizado. Podemos aprimorar o modelo matemático para a probabilidade de formação de cadeias de aminoácidos? Outros parâmetros químicos além do rendimento e abundância de aminoácidos devem ser incluídos no modelo matemático? Essas questões permanecem em aberto.

A APLICAÇÃO DO MODELO PROBABILÍSTICO PARA A FORMAÇÃO DE DI-PEPTÍDEOS E TRIPEPTÍDEOS

•Exemplo A: formação de dipeptídios considerando Glicina e Alanina

Para esse caso temos as seguintes possíveis ligações peptídicas (em vermelho é o aminoácido acrescentado a ligação): **GG**; **GG**; **AG**; **GA**; **GA**; **AG**; **AA**; **AA**. Utilizando o modelo para a probabilidade de formação de peptídeo temos:

Tabela A.1: Cálculo das probabilidades P_{AB} para combinações de 2 aminoácidos

Combinação	Cálculo	P_{AB}
$G \text{ } \color{red}{G}$ e $\color{red}{G} G$	$0,5 \times (1 \times 1) \times 0,5 = 0,5 \times (1 \times 1) \times 0,5$	0,25
$G \text{ } \color{red}{A}$ e $\color{red}{G} A$	$0,5 \times (1 \times 0,139) \times 1 = 0,5 \times (1 \times 0,139) \times 1$	0,0695
$A \text{ } \color{red}{G}$ e $\color{red}{A} G$	$1 \times (1 \times 0,182) \times 0,5 = 1 \times (1 \times 0,182) \times 0,5$	0,091
$A \text{ } \color{red}{A}$ e $\color{red}{A} A$	$1 \times (1 \times 0,283) \times 1 = 1 \times (1 \times 0,283) \times 1$	0,283

G = Glicina, A = Alanina. Fonte: autora

Neste caso, temos 4 probabilidades diferentes de formação de dipeptídeos ($n = 2$).

Exemplo B: formação de tripeptídios considerando Glicina e Alanina

Neste caso faremos a suposição de que o dipeptídeo tem um comportamento semelhante ao aminoácido quando da entrada do terceiro aminoácido, em vermelho. Para esse caso temos as seguintes possíveis ligações: **GGG**; **GGG**; **AGG**; **GGA**; **GGA**; **GAG**; **AGA**; **GAA**; **GAG**; **AGG**; **AAG**; **AGA**; **GAA**; **AAG**; **AAA**; **AAA**

Utilizando o modelo para a probabilidade de formação de peptídeo temos 8 probabilidades diferentes de formação de tripeptídeos ($n = 3$). Podemos notar que o cálculo probabilístico que estamos fazendo é realizado de forma iterativa. Veja que primeiro determinamos todas as possibilidades de formação de dipeptídeos e após isso utilizamos esse resultado como parâmetro E_0 , que representa o acúmulo de probabilidade de formação de dipeptídeos, para calcularmos as probabilidades de formação de tripeptídeos ($n = 3$).

Para fins de explicação, veja que nos cálculos para 3 aminoácidos, o parâmetro E_0 é o acúmulo de formação de peptídeo da etapa anterior. Já o parâmetro R é o rendimento da ligação peptídica entre dois aminoácidos, o aminoácido acrescentado na ligação e o seu próximo vizinho.

Tabela A.2: Cálculo das probabilidades P_{ABC} para combinações de 3 aminoácidos

Combinação	Cálculo	P_{ABC}
$G \begin{bmatrix} \text{GG} \end{bmatrix}$ e $\begin{bmatrix} \text{GG} \end{bmatrix} G$	$0,5 \times (0,25 \times 1) \times 0,5 = 0,5 \times (0,25 \times 1) \times 0,5$	0,0625
$G \begin{bmatrix} \text{GA} \end{bmatrix}$ e $\begin{bmatrix} \text{GG} \end{bmatrix} A$	$0,5 \times (0,25 \times 0,139) \times 1 = 0,5 \times (0,0695 \times 1) \times 0,5$	0,017375
$G \begin{bmatrix} \text{AG} \end{bmatrix}$ e $\begin{bmatrix} \text{GA} \end{bmatrix} G$	$1 \times (0,0695 \times 0,182) \times 0,5 = 0,5 \times (0,091 \times 0,139) \times 1$	0,0063245
$A \begin{bmatrix} \text{GG} \end{bmatrix}$ e $\begin{bmatrix} \text{AG} \end{bmatrix} G$	$0,5 \times (0,091 \times 1) \times 0,5 = 1 \times (0,25 \times 0,182) \times 0,5$	0,02275
$A \begin{bmatrix} \text{AG} \end{bmatrix}$ e $\begin{bmatrix} \text{AA} \end{bmatrix} G$	$1 \times (0,283 \times 0,182) \times 0,5 = 0,5 \times 1 \times (0,091 \times 0,283) \times 1$	0,025753
$A \begin{bmatrix} \text{GA} \end{bmatrix}$ e $\begin{bmatrix} \text{AG} \end{bmatrix} A$	$0,5 \times (0,091 \times 0,139) \times 1 = 1 \times (0,0695 \times 0,182) \times 0,5$	0,0063245
$G \begin{bmatrix} \text{AA} \end{bmatrix}$ e $\begin{bmatrix} \text{GA} \end{bmatrix} A$	$1 \times (0,0695 \times 0,283) \times 1 = 0,5 \times (0,283 \times 0,139) \times 1$	0,0196685
$A \begin{bmatrix} \text{AA} \end{bmatrix}$ e $\begin{bmatrix} \text{AA} \end{bmatrix} A$	$1 \times (0,283 \times 0,283) \times 1 = 1 \times (0,283 \times 0,283) \times 1$	0,080089

G = Glicina, A = Alanina. Fonte: autora

REFERÊNCIAS

- [1] Bernd M Rode, Artur H Eder, and Yongyos Yongyai. Amino acid sequence preferences of the salt-induced peptide formation reaction in comparison to archaic cell protein composition. *Inorganica chimica acta*, 254(2):309–314, 1997. DOI: [https://doi.org/10.1016/S0020-1693\(96\)05178-X](https://doi.org/10.1016/S0020-1693(96)05178-X).
- [2] Bernd M Rode, Hoang L Son, Yuttana Suwannachot, and Juraj Bujdak. The combination of salt induced peptide formation reaction and clay catalysis: a way to higher peptides under primitive earth conditions. *Origins of Life and Evolution of the Biosphere*, 29(3):273–286, May 1999. DOI: 10.1023/a:1006540101290.
- [3] Dimas A M Zaia, Cássia Thaïs B V Zaia, and Henrique De Santana. Which amino acids should be used in prebiotic chemistry studies? *Review Orig Life Evol Biosph*, 38(6):469–488, December 2008. DOI: 10.1007/s11084-008-9150-5.
- [4] International Union of Pure, Applied Chemistry, and International Union of Biochemistry. Nomenclature of amino acids and peptides. *Pure and Applied Chemistry*, 56(5):595–624, 1984. DOI:10.1111/j.1432-1033.1984.tb07877.x.
- [5] Alessandra Machado, Cleber W. Liria, Patrícia B. Proti, César Remuzgo, and M. Têrêsa M. Miranda. Sínteses química e enzimática de peptídeos: princípios básicos e aplicações. *Química Nova*, 27(5):781–789, September 2004. DOI: 10.1590/S0100-40422004000500018.
- [6] António José Meireles Ribeiro. Proteína. *Revista Ciência Elementar*, 2(3):229, September 2014. DOI:<http://doi.org/10.24927/rce2014.229>.
- [7] Sandip Kumar, Michaël L Cartron, Nic Mullin, Pu Qian, Graham J Leggett, C Neil Hunter, and Jamie K Hobbs. Direct imaging of protein organization in an intact bacterial organelle using high-resolution atomic force microscopy. *ACS nano*, 11(1):126–133, November 2017. DOI: <https://doi.org/10.1021/acsnano.6b05647>.
- [8] Wolfgang Nitschke, Orion Farr, Nil Gaudu, Chloé Truong, François Guyot, Michael J Russell, and Simon Duval. The winding road from origin to emergence (of life). *Life*, 14(5):607, May 2024. DOI: 10.3390/life14050607.
- [9] Luke Leman, Leslie Orgel, and M Reza Ghadiri. Carbonyl sulfide-mediated prebiotic formation of peptides. *Science*, 306(5694):283–286, 2004. DOI: 10.1023/a:1006540101290.

- [10] Claudia Huber and Gunter Wächtershäuser. Peptides by activation of amino acids with co on (ni, fe) s surfaces: implications for the origin of life. *Science*, 281(5377):670–672, July 1998. DOI: 10.1126/science.281.5377.670.
- [11] Dimas AM Zaia and Cássia Thaís BV Zaia. A few experimental suggestions using minerals to obtain peptides with a high concentration of l-amino acids and protein amino acids. *Symmetry*, 12(12):2046, October 2020. DOI: <https://doi.org/10.3390/sym12122046>.
- [12] Dimas A. M. Zaia. Da geração espontânea à química prebiótica. *Química Nova*, 26(2):260–264, 2003. DOI: <https://doi.org/10.1590/S0100-40422003000200020>.
- [13] Stanley L. Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529, February 1953. DOI: 10.1126/science.117.3046.528.
- [14] Albert Eschenmoser. The search for the chemistry of life’s origin. *Tetrahedron*, 63(52):12821–12844, December 2007. DOI: <https://doi.org/10.1016/j.tet.2007.10.012>.
- [15] Joshua Jortner. Conditions for the emergence of life on the early earth: summary and reflections. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1474):1877–1891, October 2006. DOI: 10.1098/rstb.2006.1909.
- [16] Francesco Redi. *Esperienze intorno alla generazione degl’insetti*. All’insegna della Stella, Firenze, Itália, 1996.
- [17] BiologiaNet. Experimento de redi. <https://www.biologianet.com/origem-universo-vida/experimento-redi.htm>, s.d. Acesso em: 10 jun. 2025.
- [18] Louis Pasteur. *Mémoire sur les corpuscules organisés qui existent dans l’atmosphère. Examen de la doctrine des générations spontanées*. Mallet-Bachelier, Paris, França, 1862.
- [19] Lilian Al-Chueyr Pereira Martins. Pasteur e a geração espontânea: uma história equivocada. *Filosofia e História da Biologia*, 4(1):65–100.
- [20] Walter Gilbert. Origin of life: The rna world. *nature*, 319(20):618, February 1986. DOI: 10.1038/319618a0.
- [21] N. C. Wickramasinghe. Panspermia according to hoyle. *Astrophysics and Space Science*, 285(2):535–538, 2003. DOI: 10.1023/A:1025437920710.
- [22] Dimas A. M. Zaia. Algumas controvÉrsias sobre a origem da vida. *Química Nova*, 31(6):1599–1602, August 2008. DOI:<https://doi.org/10.1590/S0100-40422008000600054>.

- [23] Kenji Ikehara. Possible steps to the emergence of life: the [gadv]-protein world hypothesis. *Chem Rec*, 5(2):107–118, 2005. DOI: 10.1002/tcr.20037.
- [24] Fabiana Santos Gonçalves. Experimento de miller. <https://www.infoescola.com/evolucao/experimento-de-miller/>, 2006–2025. Consultado em 10 de junho de 2025 às 15:01.
- [25] Pierre Canavelli, Saidul Islam, and Matthew W Powner. Peptide ligation by chemo-selective aminonitrile coupling in water. *Nature*, 571(7766):546–549, 2019. DOI: <https://doi.org/10.1038/s41586-019-1371-4>.
- [26] Sidney W Fox and Kaoru Harada. Thermal copolymerization of amino acids to a product resembling protein. *Science*, 128(3333):1214–1214, November 1958. DOI: 10.1126/science.128.3333.1214.
- [27] James P Ferris, Aubrey R Hill Jr, Rihe Liu, and Leslie E Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381(6577):59–61, May 1996. DOI:10.1038/381059a0.
- [28] Yuling Yang, Zhibiao Wang, Jin Bai, and Hai Qiao. Prebiotic peptide synthesis: How did longest peptide appear? *Journal of Molecular Evolution*, 93(5):1–19, 2025. DOI: 10.1007/s00239-025-10237-9.
- [29] Dimas AM Zaia and Cássia Thaïs BV Zaia. Adsorção de aminoácidos sobre minerais e a origem da vida. *Química Nova*, 29(5):786–789, March 2006. DOI: 10.1590/S0100-40422006000400027.
- [30] B. M. Rode, W. Flader, C. Sottriffer, and A. Righi. Are prions a relic of an early stage of peptide evolution? *Peptides*, 20(12):1513–1516, December 1999. DOI: [https://doi.org/10.1016/S0196-9781\(99\)00163-1](https://doi.org/10.1016/S0196-9781(99)00163-1).
- [31] Jeremy M. Berg, John L. Tymoczko, and Lubert Stryer. *Biochemistry*. W. H. Freeman, New York, 2002.
- [32] Andy Field. *Discovering statistics using IBM SPSS statistics*. Sage publications limited, 1 Oliver’s Yard, 55 City Road, London, EC1Y 1SP, United Kingdom, 2024.
- [33] Nico JD Nagelkerke et al. A note on a general definition of the coefficient of determination. *biometrika*, 78(3):691–692, 1991. DOI: <https://doi.org/10.1093/biomet/78.3.691>.
- [34] Norio Kitadai and Shigenori Maruyama. Origins of building blocks of life: A review. *Geoscience Frontiers*, 9(4):1117–1153, 2018. DOI: <https://doi.org/10.1016/j.gsf.2017.07.007>.

- [35] Kepa Ruiz-Mirazo, Carlos Briones, and Andres de la Escosura. Prebiotic systems chemistry: new perspectives for the origins of life. *Chemical reviews*, 114(1):285–366, January 2014. DOI: 10.1021/cr2004844.
- [36] Dimas Augusto Morozin Zaia. A origem da vida e a química prebiótica. *Semina: Ciências Exatas e Tecnológicas*, 25(1):3–8, 2004. DOI: 10.5433/1679-0375.2004v25n1p3.
- [37] Sandra Pizzarello. The chemistry that preceded life’s origin: a study guide from meteorites. *Chemistry & biodiversity*, 4(4):680–693, 2007. DOI: <https://doi.org/10.1002/cbdv.200790058>.