

UNIVERSIDADE FEDERAL DE ALFENAS

ROBERTA SIQUEIRA FERNANDES

**APLICAÇÕES DO DIAGRAMA DE HASSE NA ANÁLISE DAS
PROPRIEDADES DOS AMINOÁCIDOS DO CÓDIGO GENÉTICO**

ALFENAS/MG

2021

ROBERTA SIQUEIRA FERNANDES

**APLICAÇÕES DO DIAGRAMA DE HASSE NA ANÁLISE DAS
PROPRIEDADES DOS AMINOÁCIDOS DO CÓDIGO GENÉTICO**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, Área de concentração em Estatística Aplicada e Biometria da Universidade Federal de Alfenas, MG, como parte dos requisitos para a obtenção do título de Mestre.

Linha de Pesquisa: Matemática Aplicada e Modelagem Matemática.

Orientador: Prof. Dr. Anderson José de Oliveira.

ALFENAS, MG

2021

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfnas

F363a Fernandes, Roberta Siqueira.
Aplicações do diagrama de Hasse na análise das propriedades dos aminoácidos do código genético. / Roberta Siqueira Fernandes. - Alfnas/MG, 2021.
92f.: il. --

Orientador: Anderson José de Oliveira.
Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfnas, 2021.
Bibliografia.

1. Álgebra abstrata. 2. Álgebra Booleana. 3. Matemática aplicada. 4. Biologia. I. Oliveira, Anderson José de. II. Título.

CDD-519

ROBERTA SIQUEIRA FERNANDES

APLICAÇÕES DO DIAGRAMA DE HASSE NA ANÁLISE DAS PROPRIEDADES DOS AMINOÁCIDOS DO CÓDIGO GENÉTICO

A Banca examinadora abaixo-assinada aprova a Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Aprovada em: 30 de abril de 2021.

Prof. Dr. Anderson José de Oliveira
Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Leandro Bezerra de Lima
Instituição: Universidade Federal de Mato Grosso do Sul - UFMS

Prof. Dr. José Paulo Carvalho dos Santos
Instituição: Universidade Federal de Alfenas - UNIFAL-MG



Documento assinado eletronicamente por **Anderson Jose de Oliveira, Professor do Magistério Superior**, em 30/04/2021, às 12:52, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **José Paulo Carvalho dos Santos, Professor do Magistério Superior**, em 30/04/2021, às 14:14, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Leandro Bezerra de Lima, Usuário Externo**, em 30/04/2021, às 15:06, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0503866** e o código CRC **3F5E3862**.

“The path isn’t a straight line; it’s a spiral. You continually come back to things you thought you understood and see deeper truths.” (Barry H. Gillespie)

*Aos meus pais, irmão,
amigos e familiares pelo
amor, carinho e por sempre
me apoiarem.*

AGRADECIMENTOS

Não foi fácil e ao me lembrar da trajetória me lembro de uma frase que li uma vez no Instagram, mas de que não me recordo o autor: “Nesse esforço que me quebro e remonto, lembro-me de que estudar envolve renúncias e renunciar não é para todos. É preciso trazer a descrição da estrada e o percurso trilhado para que jamais digam que foi sorte”. Devo lembrar, que em nenhum momento estive sozinha e por este motivo, tenho muito a agradecer a muitas pessoas. Primeiramente a Deus que permitiu que tudo isso acontecesse, ao longo de minha vida, e não somente nestes anos como universitária, mas em todos os momentos é o maior mestre que alguém pode conhecer. Agradeço pela paciência, sabedoria e por me lembrar que sou mais forte do que penso.

Aos meus pais, Roberto e Sueli, pelo amor, carinho, paciência e por sempre acreditarem no meu potencial. Vocês me ensinaram a ser forte, sem perder a doçura e sempre me impulsionaram a correr atrás dos meus sonhos. Vocês estavam por perto em todos os momentos. A vocês, minha eterna gratidão.

Ao meu irmão Jeferson e minha cunhada Michele obrigada pelos conselhos, incentivo, apoio e por tudo que tem feito por mim durante essa trajetória. Obrigada pelo presente que vocês me deram, minha sobrinha/afilhada Helena. Ela irá nos encher de alegria e amor. Aos meus familiares que sempre me incentivaram e aos meus afilhados Pedro Miguel e Pedro Miguel que sempre me alegraram. A minha prima Carol, que sempre foi meu orgulho, admiração e sempre esteve presente.

Aos meus amigos, quer estejam longe ou perto, quer sejam recentes ou antigos, agradeço ao apoio, as palavras de incentivo, a todos os momentos proporcionados, sejam eles alegres ou tristes, tudo seria tão mais difícil se vocês não estivessem comigo. Em especial, as minhas amigas de graduação Alice e Nayara, que sempre me ajudaram e compartilham comigo os medos, as alegrias e frustrações, as minhas amigas Michele e Ludmila, por serem pessoas tão especiais e companheiras. A minha amiga Talita pelos anos de amizade, que sempre me tirava de casa para os “rolêzinhos ou barzinhos” e por fazer esquecer, por alguns minutos a vida acadêmica.

Ao meu orientador, Anderson José de Oliveira, pela dedicação, incentivo, correções e paciência na execução deste trabalho. Além de professor, se tornou um grande amigo/parceiro nesta minha caminhada, aliás já estamos anos nisso, desde a graduação. Foram tantas conversas

e orientações. Agradeço também, pelo conhecimento e pela oportunidade de desenvolver tal pesquisa. Por estar presente durante toda a minha jornada e fazer parte de cada conquista. Eu posso dizer que a minha formação, inclusive pessoal, não teria sido a mesma sem a sua pessoa. Obrigada pelos almoços, viagens e “rolês aleatórios”.

À Universidade Federal de Alfenas, aos servidores do programa de Pós-Graduação em Estatística Aplicada e Biometria, pela oportunidade concedida para a realização do mestrado. Aos professores por me proporcionar o conhecimento não apenas racional, mas a manifestação do caráter e afetividade da educação no processo de formação profissional, por tanto que se dedicaram a mim, não somente por terem me ensinado, mas por terem me feito aprender. A palavra mestre, nunca fará justiça aos professores dedicados aos quais sem nominar terão os meus eternos agradecimentos.

A Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro e à CAPES, pois o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Enfim, a todos que contribuíram para a realização desse sonho, serei eternamente grata a vocês. Agradeço a todos de coração, que Deus vos abençoe.

RESUMO

A modelagem algébrica do código genético, cujo objetivo é caracterizar matematicamente o que o mundo biológico realiza, tornou-se um problema instigante e atraente para diversos pesquisadores. As estruturas matemáticas, como os reticulados booleanos, diagramas de Hasse e a distância de Hamming, têm sido utilizadas no processo de análise de fenômenos mutacionais, possibilitando uma análise físico-química dos aminoácidos, além de permitirem a caracterização algébrica do código genético. O matemático Claude E. Shannon foi o precursor da teoria de códigos e aplicou elementos de engenharia na biologia. Este trabalho propõe a construção dos reticulados booleanos e dos diagramas de Hasse primal e dual, utilizando uma permutação, escolhida aleatoriamente, de cada um dos rotulamentos A , B e C , que foram obtidos através de uma analogia do código genético, relacionada com o conjunto de nucleotídeos denotado por $N = \{A, C, G, T/U\}$ (A -adenina, C -citosina, G -guanina, T -timina/ U -uracila) e o alfabeto 4-ário na estrutura de anel, denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Da relação entre esses conjuntos é possível obter 24 permutações, as quais podem ser divididas em três rotulamentos (A , B e C), de acordo com as características geométricas e a complementaridade biológica dos mesmos. A partir dessas construções foi realizada uma análise das diferenças e semelhanças físico-químicas dos aminoácidos, através da caracterização biológica das construções e dos cálculos das médias das distâncias de Hamming entre os códons. Os resultados expostos neste trabalho contribuem para o desenvolvimento de ferramentas que podem ser aplicadas em diversas áreas do conhecimento, como engenharia genética, devido ao fato de terem sido utilizados elementos de Matemática, Biologia e Códigos Corretores de Erros, uma vez que podem servir como base para estudos relacionados a fenômenos mutacionais.

Palavras-Chave: Álgebra Abstrata. Álgebra Booleana. Matemática aplicada. Biologia.

ABSTRACT

The algebraic modeling of the genetic code, aiming to characterize mathematically what the biological world does, has become an exciting and attractive problem for several researchers. Mathematical structures, such as Boolean lattices, Hasse diagrams and Hamming distance, have been used in the process of analyzing mutational phenomena, allowing a physical-chemical analysis of amino acids, in addition to allowing the algebraic characterization of the genetic code. The mathematician Claude E. Shannon was the forerunner of code theory and applied elements of engineering in biology. This work proposes the construction of the Boolean lattices and the Hasse diagrams primal and dual, using a permutation, chosen randomly, from each of the labels A , B and C , which were obtained through an analogy of the genetic code, related to the set of nucleotides denoted by $N = \{A, C, G, T/U\}$ (A -adenine, C -cytosine, G -guanine, T -thymine/ U -uracil) and the 4-ary alphabet in the ring structure, denoted by $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. From the relationship between these sets, it is possible to obtain 24 permutations, which can be divided into three labels (A , B and C), according to the geometric characteristics and their biological complementarity. From these constructions, an analysis of the differences and physico-chemical similarities of the amino acids was carried out, through the biological characterization of the constructions and the calculations of the average Hamming distances among the codons, relevant for the study of mutational phenomena. The results exposed in this work contribute to the development of tools that can be applied in several areas of knowledge such as genetic engineering, due to the fact that elements of Mathematics, Biology and Error Correcting Codes were used, since it can serve as a basis for studies related to mutational phenomena.

Keywords: Abstract Algebra. Algebra Boolean. Application Mathematical. Biology.

LISTA DE FIGURAS

Figura 1 –	O alfabeto do código genético e o alfabeto 4-ário.	17
Figura 2 –	Rotulamentos A, B e C.	18
Figura 3 –	Informação hereditária. A informação hereditária na célula-ovo fertilizada determina a natureza de todo o organismo multicelular. Apesar das suas células iniciais parecerem superficialmente semelhantes, como indicado: uma célula-ovo de ouriço-do-mar dá origem a um ouriço-do-mar (A e B). Uma célula-ovo de camundongo dá origem a um camundongo (C e D). Uma célula-ovo da alga marinha <i>Fucus</i> origina uma alga marinha <i>Fucus</i> (E e F).	21
Figura 4 –	Célula procariota (a) e célula eucariota (b).	23
Figura 5 –	Estrutura dos nucleotídeos.	24
Figura 6 –	Unidades fundamentais do DNA.	24
Figura 7 –	Estrutura do DNA - dupla-hélice.	25
Figura 8 –	Nucleotídeos do DNA.	25
Figura 9 –	Nucleotídeos do RNA.	26
Figura 10 –	Processo de duplicação do DNA.	27
Figura 11 –	A informação codificada no DNA é convertida nas sequências de aminoácidos.	27
Figura 12 –	Do DNA à proteína.	28
Figura 13 –	Restos da divisão de $X^n + 1$ por $X^4 + X + 1$, com $n = 1; \dots; 15$	44
Figura 14 –	Relação identidade I_A	51
Figura 15 –	Diagrama de Hasse.	51
Figura 16 –	Elementos importantes em um sistema de comunicação.	52
Figura 17 –	Reticulado primal (0132).	59
Figura 18 –	Reticulado dual (2310).	59
Figura 19 –	Preenchimento da 4ª linha do diagrama de Hasse.	60
Figura 20 –	Diagrama de Hasse primal referente à permutação 0132 do Rotulamento A.	65
Figura 21 –	Diagrama de Hasse dual referente à permutação 2310 do Rotulamento A.	66
Figura 22 –	Reticulado primal (1032).	72

Figura 23 – Reticulado dual (3210).	72
Figura 24 – Diagrama de Hasse primal referente à permutação 1032 do Rotulamento <i>B</i>	74
Figura 25 – Diagrama de Hasse dual referente à permutação 3210 do Rotulamento <i>B</i>	75
Figura 26 – Reticulado primal (2031).	81
Figura 27 – Reticulado dual (0213).	81
Figura 28 – Diagrama de Hasse primal referente à permutação 2031 do Rotulamento <i>C</i>	83
Figura 29 – Diagrama de Hasse dual referente à permutação 0213 do Rotulamento <i>C</i>	84

LISTA DE TABELAS

Tabela 1	– Principais características das células procariotas e eucariotas.	22
Tabela 2	– Hidropacidade dos aminoácidos.	29
Tabela 3	– Combinações possíveis compostas pelas quatro bases nitrogenadas e seus respectivos aminoácidos.	30
Tabela 4	– Tábua de $G = \{e, a\}$	35
Tabela 5	– Tábua da adição de $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$	39
Tabela 6	– Tábua da multiplicação de $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$	39
Tabela 7	– Operação de adição módulo-2.	40
Tabela 8	– Operação de multiplicação módulo-2.	41
Tabela 9	– Lista de polinômios primitivos.	45
Tabela 10	– $GF(2^4)$ gerado por $p(X) = 1 + X + X^4$	48
Tabela 11	– Operação com o conectivo \wedge (conjunção).	50
Tabela 12	– Operação com o conectivo \vee (disjunção).	50
Tabela 13	– Associação estabelecida pela permutação 0132 primal do Rotulamento A	57
Tabela 14	– Associação estabelecida pela permutação 2310 dual do Rotulamento A	57
Tabela 15	– Operação primal \wedge (e) do Rotulamento A	57
Tabela 16	– Operação primal \vee (ou) do Rotulamento A	58
Tabela 17	– Operação dual \wedge (e) do Rotulamento A	58
Tabela 18	– Operação dual \vee (ou) do Rotulamento A	58
Tabela 19	– Médias das distâncias de Hamming entre pares de aminoácidos do Rotulamento A	68
Tabela 20	– Associação estabelecida pela permutação 1032 primal do Rotulamento B	71
Tabela 21	– Associação estabelecida pela permutação 3210 dual do Rotulamento B	71
Tabela 22	– Operação primal \wedge (e) do Rotulamento B	71
Tabela 23	– Operação primal \vee (ou) do Rotulamento B	71
Tabela 24	– Operação dual \wedge (e) do Rotulamento B	72
Tabela 25	– Operação dual \vee (ou) do Rotulamento B	72

Tabela 26 – Médias das distâncias de Hamming entre pares de aminoácidos do Rotulamento <i>B</i>	77
Tabela 27 – Associação estabelecida pela permutação 2031 primal do Rotulamento <i>C</i>	80
Tabela 28 – Associação estabelecida pela permutação 0213 dual do Rotulamento <i>C</i>	80
Tabela 29 – Operação primal \wedge (e) do Rotulamento <i>C</i>	80
Tabela 30 – Operação primal \vee (ou) do Rotulamento <i>C</i>	80
Tabela 31 – Operação dual \wedge (e) do Rotulamento <i>C</i>	81
Tabela 32 – Operação dual \vee (ou) do Rotulamento <i>C</i>	81
Tabela 33 – Médias das distâncias de Hamming entre pares de aminoácidos do Rotulamento <i>C</i>	85

SUMÁRIO

1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO	20
2.1	ELEMENTOS DE BIOLOGIA	20
2.1.1	A Célula	21
2.1.2	Tipos de Células	22
2.1.3	Nucleotídeos e Ácidos Nucleicos	23
2.1.4	A Duplicação do DNA	26
2.1.5	Síntese Proteica	27
2.1.6	Proteínas e Aminoácidos	28
2.1.7	O Código Genético	30
2.1.8	Mutações	32
2.2	ELEMENTOS DE ÁLGEBRA	33
2.2.1	Grupos	33
2.2.2	Anéis	35
2.2.3	Corpos	39
2.2.3.1	Aritmética de Corpos Binários	41
2.2.4	Construção de Corpos de Galois $GF(2^m)$	45
2.2.5	Relações de Ordem	49
2.2.5.1	Reticulados Booleanos	49
2.2.5.2	Diagrama de Hasse	50
2.3	CÓDIGOS CORRETORES DE ERROS	52
2.3.1	Códigos de Bloco	53
3	CONSTRUÇÃO E ANÁLISE DOS RETICULADOS BOOLEANOS E DIAGRAMAS DE HASSE ASSOCIADOS AOS ROTULAMENTOS DO CÓDIGO GENÉTICO	56
3.1	ROTULAMENTO A	57
3.1.1	Procedimento para a construção dos reticulados booleanos	57
3.1.2	Procedimento para a construção dos Diagramas de Hasse	59
3.1.3	Análise do diagrama primal	65
3.1.4	Análise do diagrama dual	66
3.1.5	Análise das diferenças e semelhanças físico-químicas dos aminoácidos por meio das distâncias de Hamming	68
3.2	ROTULAMENTO B	70
3.2.1	Procedimento para a construção dos reticulados booleanos	71
3.2.2	Procedimento para a construção dos Diagramas de Hasse	73
3.2.3	Análise do diagrama primal	74
3.2.4	Análise do diagrama dual	75
3.2.5	Análise das diferenças e semelhanças físico-químicas dos aminoácidos por meio das distâncias de Hamming	76
3.3	ROTULAMENTO C	79
3.3.1	Procedimento para a construção dos reticulados booleanos	80
3.3.2	Procedimento para a construção dos Diagramas de Hasse	82
3.3.3	Análise do diagrama primal	83
3.3.4	Análise do diagrama dual	84

3.3.5	Análise das diferenças e semelhanças físico-químicas dos aminoácidos por meio das distâncias de Hamming	85
4	CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS	88
4.1	Sugestões para Trabalhos Futuros	89
	REFERÊNCIAS	92

1 INTRODUÇÃO

Estudos realizados a partir da segunda metade do século XX, tornaram a natureza do código genético conhecida. Além disso, estudos envolvendo o código genético têm se tornado um problema instigante e atraente para diversos pesquisadores. Desse modo, a modelagem algébrica envolvendo o código genético tem sido alvo de diversas pesquisas, tendo como principais objetivos a identificação das propriedades, características e implicações do modelo a ser estudado.

O código genético é um sistema bioquímico que trata da tradução, ou decodificação, das informações contidas na estrutura das moléculas de DNA e RNA em sequências de proteínas, ou seja, conforme Jiménez–Montaño, Mora–Basáñez e Pöschel (1994, 1996), o código genético é tanto um sistema bioquímico quanto um sistema de comunicação. Até a década de 60, um dos maiores mistérios na Biologia, era como a sequência de nucleotídeos no DNA conseguia determinar a ordem dos aminoácidos na geração de proteínas. No século XX os biólogos conseguiram decifrar a “linguagem da vida”. Entretanto, ainda há muitas perguntas a serem respondidas com o intuito de desvendar os “mistérios da máquina da vida”. Nesse contexto, muitas pesquisas vêm sendo realizadas com estudos relacionados não só a problemas envolvendo o código genético, mas também a outros fenômenos biológicos, nas quais muitos pesquisadores utilizam da Matemática para contribuir na análise e descrição desses conceitos teóricos. Desta forma, vários autores procuraram estruturar e modelar o código genético e com isso, muitas hipóteses foram apresentadas.

Sánchez, Morgado e Grau (2004) por meio de dados experimentais, propõem a construção de reticulados booleanos, diagramas de Hasse e cálculos das médias das distâncias de Hamming entre os códons.

Em Jiménez-Montaño, Mora–Basáñez e Pöschel (1994, 1996) é apresentada uma interessante representação do código genético por meio de um hipercubo booleano hexadimensional. A estrutura proposta demonstra que no código genético existe um equilíbrio entre conservadorismo e inovação.

Uma outra proposta é a de Sánchez *et al.* (2005) na qual é apresentado um espaço vetorial n -dimensional das sequências de DNA relacionado ao corpo de Galois dos 64 códons do código genético ($GF(2^6)$). A modelagem do código genético é obtida por meio de um isomorfismo $\varphi = B(X) \rightarrow (\mathbb{Z}_2)^2$ e a importância biológica das posições das bases nos códons. A

importância da posição da base é fundamentada pela ocorrência de erros encontrada nos códons, analisada na Biologia. Os erros na terceira base são mais frequentes do que na primeira base e, por sua vez, são mais frequentes do que os erros na segunda base. Utilizando a extensão de $GF(2)$ para $GF(2^6)$, é apresentada uma forma de representar os códons através de polinômios. Também, por meio dos códons, foram apresentadas algumas operações, como o peso de Hamming entre um códon e a distância de Hamming entre dois códons.

Os códons do código genético são formados por uma trinca de bases nitrogenadas com a possibilidade de 64 combinações distintas que são consideradas por Sánchez, Morgado e Grau (2005) em seu estudo. Nesse trabalho, é apresentado um modelo por meio da construção da tabela do código genético. A construção é feita utilizando uma bijeção do alfabeto biológico com o alfabeto matemático $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{00, 01, 10, 11\}$. Desta forma, o seguinte mapeamento é proposto: $G - 00$, $U - 10$, $A - 01$ e $C - 11$, onde são definidas operações no conjunto das quatro bases do DNA, onde G representa guanina, U uracila, A adenina e C citosina. Vale ressaltar que o procedimento realizado por Sánchez, Morgado e Grau (2005) leva em consideração a importância biológica das bases nitrogenadas.

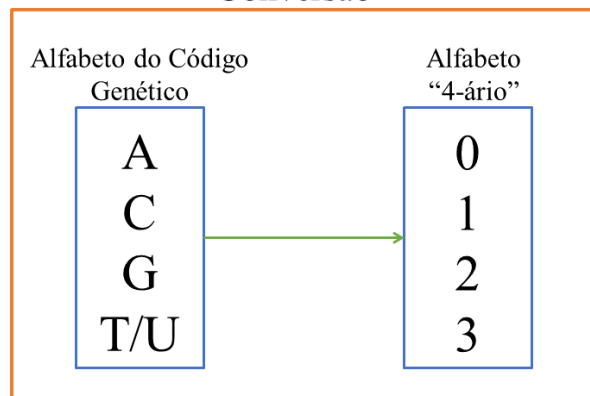
Segundo Sánchez e Grau (2006), além do código genético representar uma extensão do alfabeto de quatro letras das bases do DNA e RNA e estabelecer o emparelhamento entre as bases, devido as ligações de hidrogênio que ocorre entre elas, foi observada uma associação entre códons com U na posição da segunda base e a hidrofobia dos aminoácidos. Já os códons com A na segunda posição da base codificam aminoácidos hidrofílicos ou polares. Acredita-se que a ordem dos códons deva refletir nas propriedades físico-químicas dos aminoácidos. Partindo da ordem das quatro bases do DNA na rede booleana, Sánchez e Grau (2006) em seu estudo propõem a utilização de uma álgebra de Lie do código genético sobre o corpo de Galois das quatro bases do DNA, permitindo compreender melhor a lógica subjacente ao código genético. Concluiu-se então, que a atribuição dos códons e as propriedades físico-químicas dos aminoácidos não estão conectadas ao acaso, ou seja, a origem do código genético não foi aleatória. Além disso, o modelo proposto por Sánchez e Grau (2006) pode ajudar na compreensão de eventos mutacionais nos processos de evolução molecular.

Oliveira (2012) apresenta a modelagem algébrica do código genético, com o objetivo de identificar suas propriedades, características e implicações do modelo. É apresentada uma representação do código genético por meio de estruturas algébricas, que buscam formas de explicar fenômenos biológicos. São apresentadas as estruturas dos diagramas de Hasse, reti-

culados booleanos, estruturas de grupos, corpos, anéis e extensões de Galois. Os reticulados booleanos, bem como sua representação através do diagrama de Hasse, podem ser ferramentas eficazes na análise de algumas propriedades associadas ao código genético.

Em Faria (2011) e Rocha (2010) é apresentada a modelagem do DNA através de uma analogia do alfabeto do código genético, que está relacionado com o conjunto de nucleotídeos $N = \{A, C, G, T/U\}$ (A – adenina, C – citosina, G – guanina e T/U – timina/uracila) e o alfabeto 4-ário na estrutura de anel, denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, como pode ser observado na Figura 1.

Figura 1 – O alfabeto do código genético e o alfabeto 4-ário.
“Conversão”

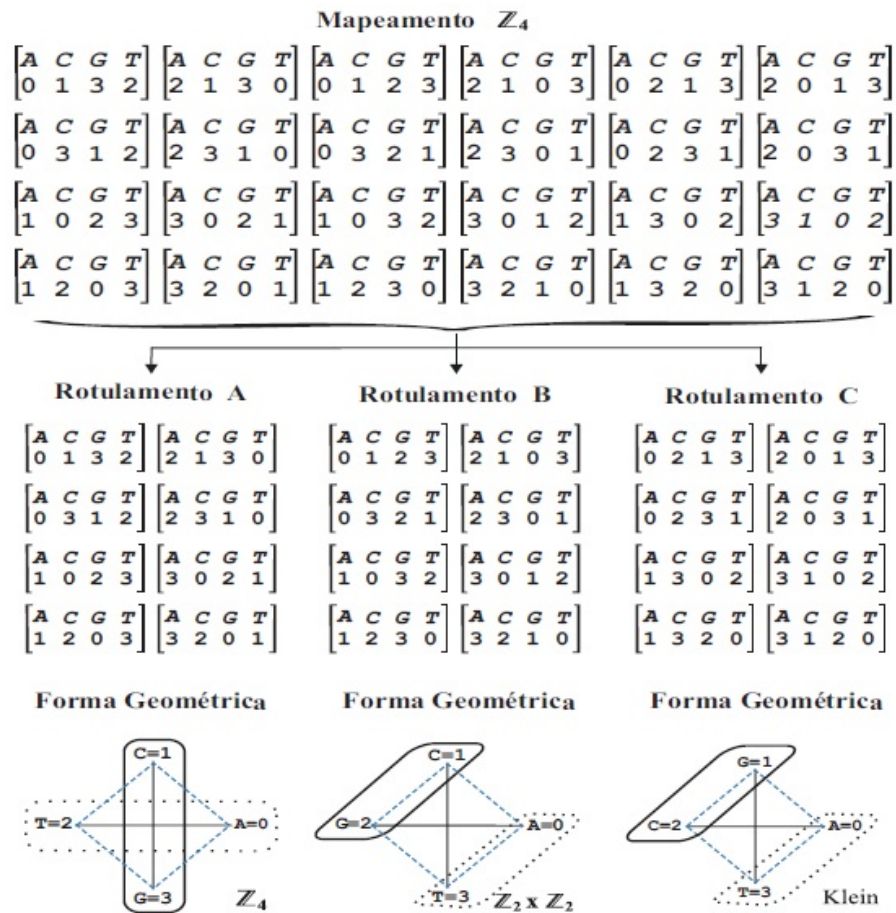


Fonte: Da autora.

Como a estrutura algébrica do alfabeto do código genético das sequências do DNA é desconhecida, temos que a conversão realizada na Figura 1 é necessária, desta forma, o alfabeto 4-ário, denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ está relacionado ao conjunto $N = \{A, C, G, T/U\}$.

A princípio, sabe-se que é desconhecido o mapeamento entre $N \leftrightarrow \mathbb{Z}_4$, então a sequência de DNA é rotulada obtendo-se 24 permutações entre $N \leftrightarrow \mathbb{Z}_4$. Como cada permutação possui suas características, pode-se então organizar esse mapeamento em três conjuntos, denominados **rotulamentos A, B e C** (Figura 2), contendo 8 permutações cada, de acordo com a caracterização geométrica de cada rotulamento, uma vez que, sabe-se que cada rotulamento produz um diferente nível de não-linearidade para as sequências. A diferença de cada rotulamento também está associada à complementaridade dos nucleotídeos, na qual, tem-se que adenina se emparelha com a timina/uracila e citosina se emparelha com guanina, ou vice-versa.

Figura 2 – Rotulamentos A, B e C.



Fonte: ROCHA, 2010, p. 74.

Na proposta de Fernandes (2018), foi possível obter as representações polinomial e vetorial associadas a cada um dos códons do código genético, levando em consideração a importância biológica das bases nitrogenadas, por meio da modelagem algébrica do código genético, utilizando a extensão de $GF(2)$ para $GF(2^6)$.

Cada molécula de DNA se diferencia pela sequência de bases nitrogenadas que elas apresentam, onde essa sequência formará o código genético. As regras para codificação das trinca de bases dos aminoácidos que formam o código genético ainda é um problema instigante e interessante, conforme Sánchez, Morgado e Grau (2004). O código genético é o sistema bioquímico, no qual são expostas as regras para descrever uma sequência de códons de um gene de uma sequência de nucleotídeos. Os códons são formados por uma trinca de bases nitrogenadas, com 64 combinações possíveis. As bases nitrogenadas são a adenina, citosina, guanina e timina/uracila, que são representadas pelas letras A, C, G e T/U, respectivamente, e representam o alfabeto do DNA.

Por meio deste trabalho, temos como propósito o detalhamento de algumas estruturas

matemáticas, que facilitam sua utilização em outras áreas de estudo, em particular a Biologia, através da construção de reticulados booleanos, além de sua representação pelo diagrama de Hasse, analisando algumas propriedades dos aminoácidos. Outra proposta é a de efetuar o cálculo das médias das distâncias de Hamming e o peso de Hamming dos códonos, a fim de analisar fenômenos mutacionais na estrutura do DNA.

Este trabalho está estruturado da seguinte maneira: no Capítulo 2 são apresentados os elementos de Biologia (2.1), Álgebra (2.2), Reticulados Booleanos e Diagramas de Hasse (Seções 2.2.5.1 e 2.2.5.2) e Códigos Corretores de Erros (Seção 2.3). No Capítulo 3 são apresentadas a construção e análise dos reticulados booleanos e diagramas de Hasse associados ao Rotulamento A (Seção 3.1), Rotulamento B (Seção 3.2) e Rotulamento C (Seção 3.3) do código genético. No Capítulo 4 são apresentadas as conclusões do trabalho e propostas para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Em virtude da interdisciplinaridade desta pesquisa, este capítulo introduz conceitos relacionados à Biologia, Álgebra e Códigos Corretores de Erros.

O presente capítulo está organizado da seguinte forma: Na seção 2.1 é realizada uma breve introdução dos principais conceitos referentes à Biologia, com o objetivo de compreender o que é célula, os tipos de célula, o que são nucleotídeos e ácidos nucleicos, como é realizada a síntese proteica, o processo de geração de proteínas, como funciona o código genético e a construção da tabela do código genético, a fim de compreender o sistema biológico e entender como ocorrem as mutações. Na seção 2.2 são apresentados os principais conceitos e definições das estruturas algébricas de grupos, anéis, corpos, extensões de Galois e relações de ordem. Esses são fundamentais para as construções que serão realizadas no decorrer deste estudo. Na seção 2.3 é apresentada uma breve introdução dos códigos corretores de erros, seus principais elementos em um sistema de comunicação e os código de bloco, permitindo assim calcular as distâncias de Hamming entre os códonos.

2.1 ELEMENTOS DE BIOLOGIA

Nesta seção, serão apresentados os principais conceitos biológicos, importantes para a compreensão deste estudo e podem ser encontrados de forma detalhada em Alberts *et al.* (2010), Alberts *et al.* (2017), Faria (2011), Griffiths (2006), Hib e Robertis (2006), Oliveira (2012), Petinari (2010), Rocha (2010) e Zaha, Ferreira e Passaglia (2014).

O Planeta Terra é habitado por organismos vivos extraordinariamente diversos. No entanto, nossos ancestrais não tinham nenhum conhecimento a respeito sobre o que eram células ou DNA, mas eles notaram que todos os organismos vivos apresentavam algo em comum. Segundo Alberts *et al.* (2017), esse algo eles chamaram de “vida”.

Na biologia celular estudamos as estruturas, funções e comportamentos das células, que nos dão vestígios sobre o entendimento das células e sua evolução, para que assim possamos lidar com os problemas históricos da vida no Planeta Terra. Edmund Beecher Wilson (1856-1939), foi o pioneiro em biologia celular e sempre enfatizava que “a chave para cada problema

biológico deve finalmente ser procurada na célula; para cada organismo vivo há, ou houve em algum momento, uma célula”, conforme Alberts *et al.* (2017).

Hoje, sabemos que todos os seres vivos, são formados por células e que todas essas unidades da matéria viva, segundo Alberts *et al.* (2010), compartilham de uma mesma maquinária para a maioria de suas funções básicas. As matérias vivas aparentam ser infinitamente variadas quanto ao seu exterior, porém, elas são semelhantes em seu interior.

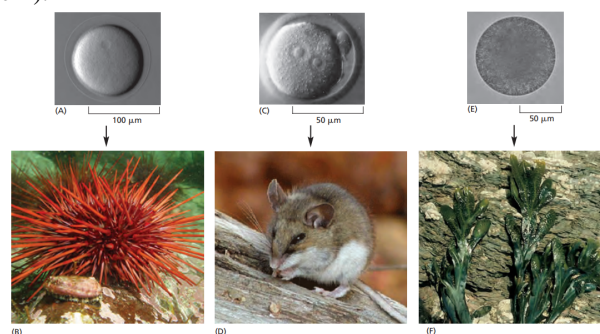
A célula é a unidade fundamental da vida em todas as formas de organismos vivos.

2.1.1 A Célula

A célula foi descoberta por Robert Hooke (1635-1703), no ano de 1665 ao examinar um pedaço de cortiça que era composto por diversas concavidades em forma de poliedro, em seu microscópio.

A célula é a menor unidade morfofisiológica de um ser vivo, além de que, uma única célula é o veículo para toda a informação hereditária (Figura 3) que define cada espécie, conforme Alberts *et al.* (2017) e podem ser apresentadas sob uma variedade de formas e tamanhos. Há aquelas que se movem rapidamente e outras que têm estruturas que se alteram. Além disso, Alberts *et al.* (2010) afirma que para algumas células o oxigênio é essencial, já para outras ele pode até matar. Existem organismos vivos que possuem uma única célula (unicelulares) e aqueles que são formados por mais de uma célula (pluricelulares).

Figura 3 – Informação hereditária. A informação hereditária na célula-ovo fertilizada determina a natureza de todo o organismo multicelular. Apesar das suas células iniciais parecerem superficialmente semelhantes, como indicado: uma célula-ovo de ouriço-do-mar dá origem a um ouriço-do-mar (A e B). Uma célula-ovo de camundongo dá origem a um camundongo (C e D). Uma célula-ovo da alga marinha Fucus origina uma alga marinha Fucus (E e F).



Fonte: ALBERTS *et al.* (2017, p. 2).

2.1.2 Tipos de Células

Devido ao grande número de organismos microscópicos que foram descobertos e por suas diferentes estruturas, nos anos 60, os microbiologistas entenderam que diferentes tipos de células poderiam ser distinguidas. Segundo Pettinari (2010), os microbiologistas Stanier e Van Niel (Stanier e Van Niel 1962) propuseram um novo critério de divisão, levando em consideração dois tipos de células: eucariotas ou eucariontes, aquelas em que o núcleo celular contém o material genético cercado por uma membrana e procariotas ou procariontes, os que não possuem essa estrutura. Porém, como os critérios estabelecidos só se adequavam aquela época, houve a necessidade de realizar estudos para diferenciar claramente as células eucariotas das procariotas. Contudo, mesmo com as discussões e os trabalhos de Woese (1979), até hoje ainda continuam sendo utilizadas as denominações eucariotas e procariotas.

As células procariotas são simples, possuem poucas organelas, não possuem um núcleo definido e organizado. O material genético se encontra em toda a célula.

As células eucariotas são mais complexas, seu núcleo é apresentado de forma mais organizada e apresentam muitas organelas. O material genético se encontra organizado no núcleo. Os protozoários, as algas, os fungos, os vegetais e os animais são seres que apresentam esse tipo de célula.

Essas células são constituídas por estruturas com diferentes funções. Na Tabela 1 são apresentadas as diferenças e principais características celulares de células procariotas e eucariotas.

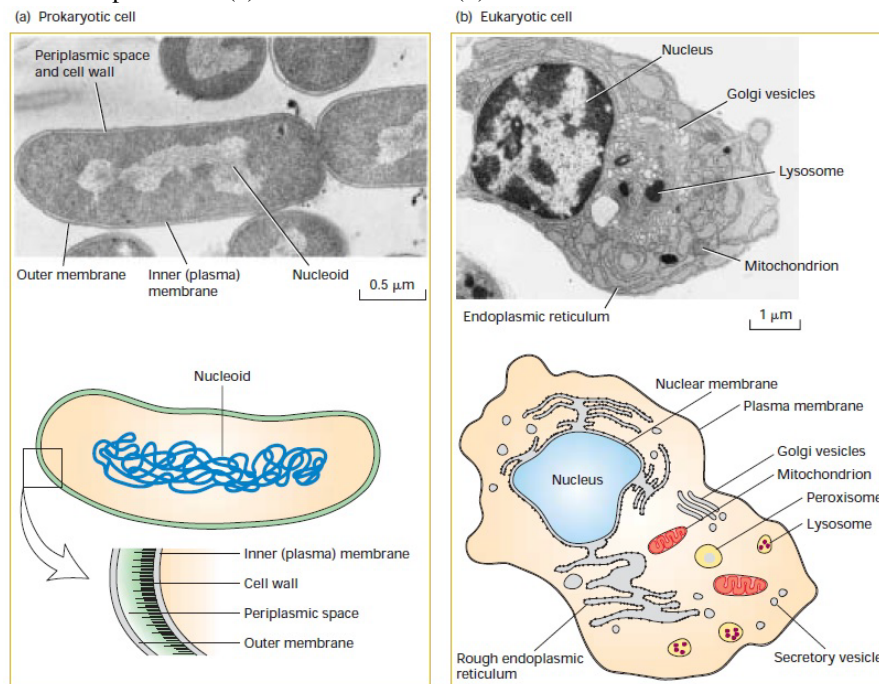
Tabela 1 – Principais características das células procariotas e eucariotas.

	Procariotas	Eucariotas
Organização	Principalmente unicelular	Principalmente pluricelular
Membrana citoplasmática	Bicamada fosfolipídica; rara presença de esteróis	Bicamada fosfolipídica; presença de esteróis e carboidratos
Núcleo	Ausente	Definido pela membrana nuclear
Citoplasma	Sem citoesqueleto	Citoesqueleto constituído
Motilidade	Flagelos simples	Flagelos complexos; pseudópodes; outros órgãos de locomoção mais complexos
Organelas	Poucas ou nenhuma	Presentes: lisossomos, complexo de Golgi, retículo endoplasmático (RE), mitocôndria e cloroplastos
Parede celular	Contém glicopeptídeos, lipídeos, proteína	Quando presente, contém quitina ou celulose

Fonte: ZAHA, FERREIRA E PASSAGLIA, 2014, p. 3.

A Figura 4 apresenta um exemplo de célula procariota (a) e célula eucariota (b).

Figura 4 – Célula procariota (a) e célula eucariota (b).



Fonte: FARIA, 2011, p. 23.

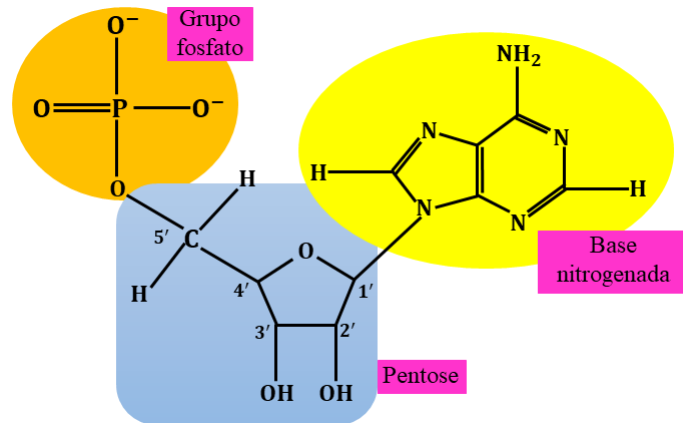
2.1.3 Nucleotídeos e Ácidos Nucleicos

Os nucleotídeos são unidades de moléculas fundamentais dos ácidos nucleicos e que estão presentes no metabolismo como transportador de energia. Ele participa da estrutura dos ácidos nucleicos através de ligações fosfodiéster*, formando cadeias longas com milhões de resíduos de comprimento. Além disso, agem como ativadores e inibidores importantes em várias vias do metabolismo intermediário da célula.

A Figura 5 apresenta como são formadas as moléculas dos nucleotídeos, sendo uma base nitrogenada, uma pentose e um ou mais radicais de fosfatos.

*Ligações fosfodiéster são ligações covalentes que ocorrem entre dois dos átomos de oxigênio de um grupo fosfato e grupos hidroxila de duas moléculas diferentes. Nesse tipo de ligação, o grupo fosfato atua como uma “ponte” de ligação estável entre as duas moléculas através de seus átomos de oxigênio.

Figura 5 – Estrutura dos nucleotídeos.

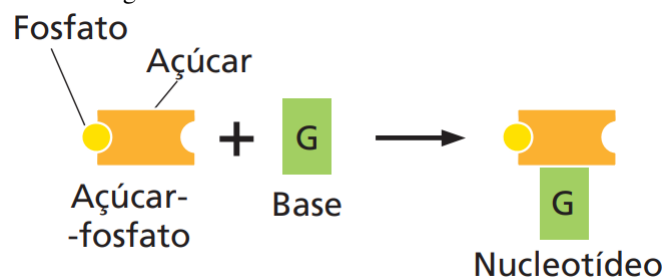


Fonte: Da autora.

Os ácidos nucleicos se encontram em todas as células vivas e têm como função armazenar, transmitir e traduzir informações genéticas. Todos os organismos vivos apresentam ácidos nucleicos na forma de Ácido Desoxirribonucleico (DNA) e Ácido Ribonucleico (RNA).

O Ácido Desoxirribonucleico, conhecido também como DNA é formado por uma longa cadeia de nucleotídeos e tem como função armazenar informação genética dos seres vivos. Cada nucleotídeo do DNA é constituído por um grupo fosfato, uma pentose (desoxirribose) e uma dentre quatro bases nitrogenadas: adenina (A), guanina (G), citosina (C) e timina (T), que estão interligadas por pontes de hidrogênio, conforme a Figura 6.

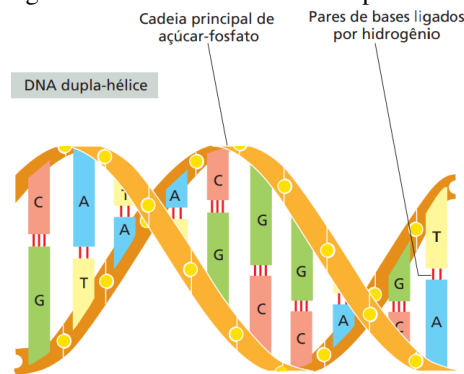
Figura 6 – Unidades fundamentais do DNA.



Fonte: ALBERTS *et al.*, 2017, p. 3.

Conforme Griffiths (2006), o DNA é formado por uma dupla hélice (Crick e Watson descobriram a hélice dupla de DNA) (Figura 7) com giro para a direita e parece uma escada de caracol molecular, que pode acomodar qualquer sequência de nucleotídeos sem alterar sua estrutura básica.

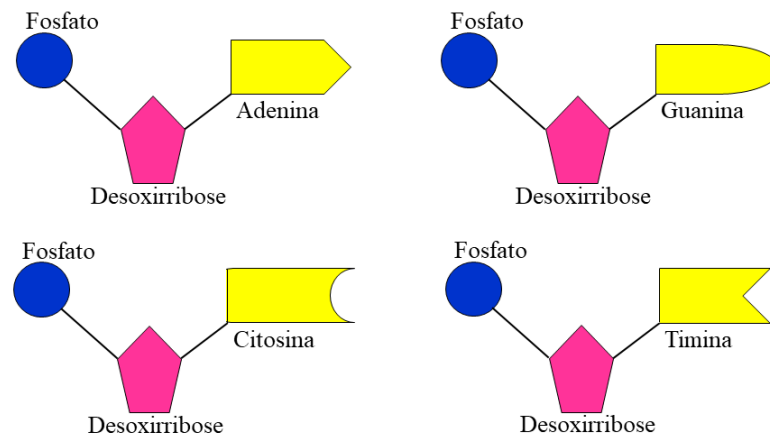
Figura 7 – Estrutura do DNA - dupla-hélice.



Fonte: ALBERTS *et al.*, 2017, p. 3.

No DNA existem quatro tipos diferentes de nucleotídeos, sendo estes compostos por três partes: um grupo fosfato, uma pentose (desoxirribose) e uma base nitrogenada (adenina e guanina, que são maiores, chamadas de purinas e timina e citosina, que são menores e chamadas de pirimidinas). Na Figura 8 são apresentados os nucleotídeos do DNA.

Figura 8 – Nucleotídeos do DNA.



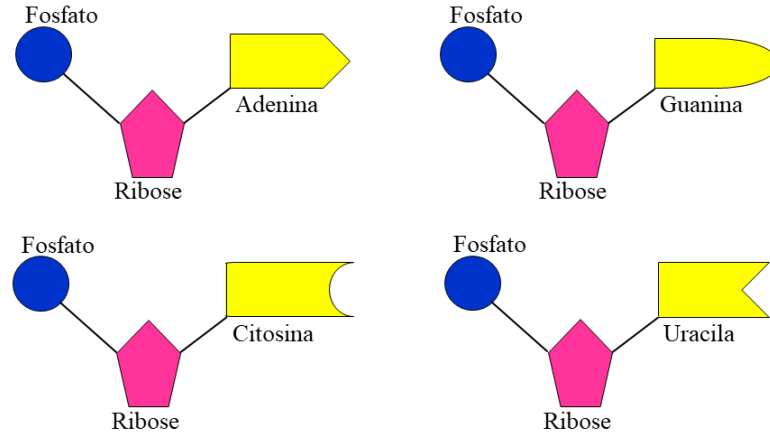
Fonte: Da autora.

Erwin Chargaff, determinou as proporções de adenina, citosina, guanina e timina presentes no DNA. A quantidade total de nucleotídeos de timina + citosina é sempre igual à quantidade total de nucleotídeos de adenina + guanina. A quantidade de timina é sempre igual à quantidade de adenina e a quantidade de citosina é sempre a mesma que de guanina. Porém, a quantidade de adenina + timina não é necessariamente igual à quantidade de guanina + citosina, ou seja, essa proporção pode variar de acordo com cada organismo.

O Ácido Ribonucleico, conhecido como RNA é um polímero de fita simples e não é formado por uma dupla hélice como o DNA. Possui uma pentose (ribose) em seus nucleotídeos em vez da desoxirribose encontrada no DNA. Logo, a diferença nesses açúcares é a presença

ou a ausência de apenas um átomo de oxigênio. O RNA possui a uracila (U) no lugar da timina (T) (Figura 9) em sua composição e participa na tradução do DNA em proteína.

Figura 9 – Nucleotídeos do RNA.



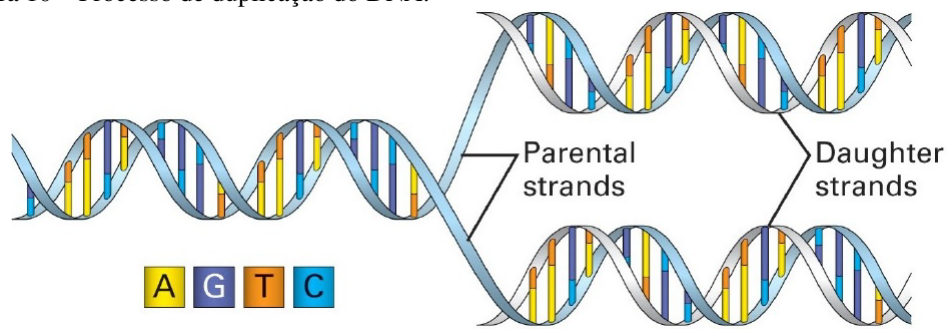
Fonte: Da autora.

Há três tipos de RNA: o RNA mensageiro, o RNA transportador e o RNA ribossômico. Esses tipos de RNA desempenham uma função específica no processo de síntese proteica. Contudo, nos organismos eucariontes e procariontes esses processos possuem algumas diferenças. Nas células eucariontes, a transcrição ocorre no núcleo e a tradução no citoplasma. Já nas células procariontes, tanto a transcrição quanto a tradução ocorrem próximas uma da outra.

2.1.4 A Duplicação do DNA

Para que ocorra o processo de duplicação do DNA, é necessário o rompimento das pontes de hidrogênio, para que assim aconteça o distanciamento das duas fitas. Com os nucleotídeos presentes nas células e livres, começam a ser atraídos pelas bases do DNA na parte em que as fitas estão separadas, formando assim outro filamento. A formação desse filamento é realizada obedecendo a regra de que adenina emparelha com timina e citosina emparelha com guanina, conhecida também como **Regra de Chargaff**. Desta forma, obtém-se duas moléculas de DNA idênticas entre si. Na Figura 10 vemos como esse processo ocorre.

Figura 10 – Processo de duplicação do DNA.



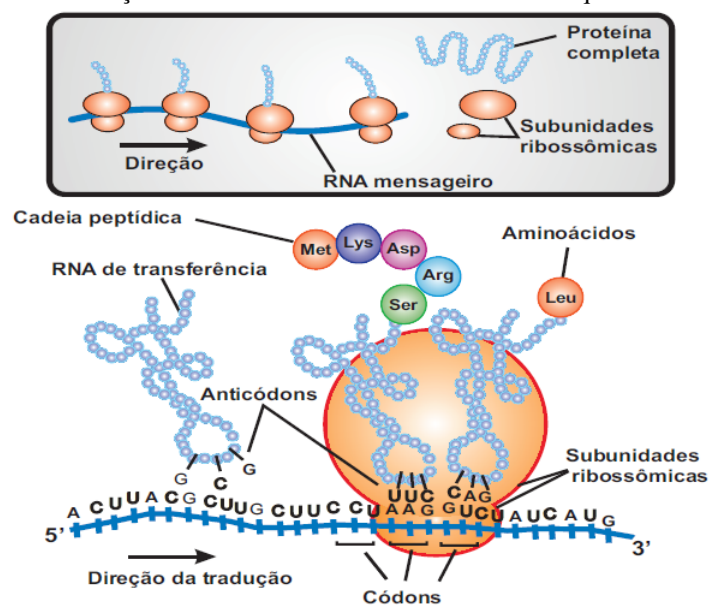
Fonte: OLIVEIRA, 2012, p. 26.

Como a molécula de RNA é uma fita simples, então apenas umas das fitas do DNA será usada como molde na produção do RNA, ou seja, através de uma molécula de DNA gera-se moléculas de RNA mensageiro que controlam a síntese proteica após deslocar-se para o citoplasma.

2.1.5 Síntese Proteica

Tanto o DNA quanto o RNA são responsáveis pela síntese proteica e para que isso ocorra, é necessário que a célula transcreva e traduza a informação do código genético, como pode ser observado na Figura 11.

Figura 11 – A informação codificada no DNA é convertida nas seqüências de aminoácidos.

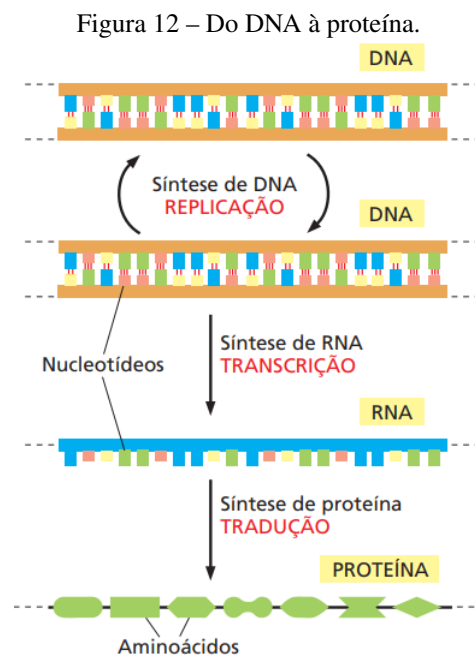


Fonte: FARIA, 2011, p. 41.

Como já observado, existem três tipos de RNA e cada um possui uma função específica.

O RNA mensageiro ou mRNA leva a mensagem do DNA, que está no núcleo, até o citoplasma. O RNA transportador ou tRNA é responsável por identificar e transportar os aminoácidos. Já o RNA ribossômico ou rRNA ajuda na decodificação da informação genética.

No processo de transcrição é necessário que ocorra a transferência da mensagem genética do DNA para o RNA. Já o processo de tradução da mensagem genética ocorre de modo que o código que é passado numa sequência de bases de RNA para uma sequência de aminoácidos da proteína. A Figura 12 apresenta como é realizado o processo do DNA à proteína.



Fonte: ALBERTS *et al.* 2017, p. 4.

2.1.6 Proteínas e Aminoácidos

As proteínas apresentam-se sob inúmeras formas e tamanhos e são as macromoléculas de maior abundância nos seres vivos. Elas estão presentes em todas as células e desempenham funções vitais dos seres vivos. Essas macromoléculas assumem funções biológicas distintas, tendo suas propriedades presentes nos músculos, cabelos, unhas, anticorpos, entre outros. A falta ou modificação de alguma proteína pode gerar algumas doenças e erros no metabolismo, como por exemplo, os portadores da doença fenilcetonúria, que é causada pelo acúmulo da proteína fenilalanina.

As proteínas são polímeros constituídos basicamente de aminoácidos. Os aminoácidos são unidades utilizadas na síntese proteica e por possuírem carbono são essenciais à vida

humana. Os aminoácidos são quimicamente estruturados da seguinte forma: um átomo de carbono, um átomo de hidrogênio, um grupo amina, um grupo carboxílico e uma cadeia lateral (grupo-R), que varia de acordo com cada aminoácido. Essa cadeia lateral é responsável por diferenciar os aminoácidos em tamanhos, funções e solubilidade em água. Os aminoácidos podem ser classificados como hidrofóbicos (tendem a evitar o contato com a água, ficando dessa forma “escondidos” no interior da proteína) ou hidrofílicos (ficam distribuídos na superfície da molécula ficando em contato com a água). Além do critério da hidropacidade, os aminoácidos também podem ser classificados pela carga (ácido, neutro, básico), estrutura (cíclico, acíclico) tamanho (pequeno, médio, grande), aromático e alifático. No presente trabalho uma das características que será analisada é a hidropacidade.

A união de dois aminoácidos gera um dipeptídeo, a junção de três aminoácidos forma um tripeptídeo, e assim segue sucessivamente, sendo que quando ocorre a união de vários aminoácidos temos uma cadeia polipeptídica. Essas estruturas são formadas por no máximo 20 tipos diferentes de aminoácidos, apresentados na Tabela 2, destes 12 são hidrofílicos (polares) e 8 hidrofóbicos (apolares). Os aminoácidos são classificados entre essenciais ou naturais de acordo com sua fonte primária. Os aminoácidos essenciais são sintetizados apenas pelos vegetais, são eles: leucina, isoleucina, lisina, metionina, treonina, triptofano, valina e fenilalanina. Os naturais são sintetizados tanto por animais quanto por vegetais, são eles: alanina, glicina, histidina, tirosina, arginina, ácido aspártico, asparagina, glutamina, serina, prolina, cisteína e ácido glutâmico.

Tabela 2 – Hidropacidade dos aminoácidos.

Hidrofílico (Polar)	Hidrofóbico (Apolar)
Asparagina (Asp)	Alanina (Ala)
Glutamina (Gln)	Leucina (Leu)
Arginina (Arg)	Valina (Val)
Histidina (His)	Isoleucina (Ile)
Lisina (Lys)	Prolina (Pro)
Cisteína (Cys)	Fenilalanina (Phe)
Glicina (Gly)	Metiotina (Met)
Serina (Ser)	Triptofano (Trp)
Treonina (Thr)	
Ácido aspártico (Asp)	
Ácido glutâmico (Gln)	
Tirosina (Tyr)	

Fonte: OLIVEIRA, 2012, p. 26.

2.1.7 O Código Genético

Até a década de 60, um dos maiores mistérios dentro da Biologia era de como a sequência de nucleotídeos no DNA conseguia determinar a ordem dos aminoácidos na geração da proteína. Contudo, no século XX os biólogos conseguiram decifrar a “linguagem da vida”.

Cada molécula de DNA se diferencia pela sequência de bases nitrogenadas que elas apresentam, e essa sequência é que vai formar o código genético, ou seja, essa sequência de bases do DNA se relaciona com a sequência correspondente de 20 aminoácidos para a formação das proteínas. Então, essa correspondência permite identificar o aminoácido específico de cada códon, como pode ser observado na Tabela 3. Esses códons são formados por uma trinca de bases nitrogenadas com 64 combinações possíveis. As bases são formadas pela adenina, citosina, guanina e timina/uracila, que são representadas pelas letras A, C, G e T/U respectivamente, e representam o alfabeto do DNA.

Tabela 3 – Combinações possíveis compostas pelas quatro bases nitrogenadas e seus respectivos aminoácidos.

primeira posição extremidade 5' ↓	segunda posição				terceira posição extremidade 3' ↓
	U	C	A	G	
A	Lys	Thr	Arg	Ile	A
	Asn	Thr	Ser	Ile	C
	Lys	Thr	Arg	Met	G
	Asn	Thr	Ser	Ile	U
C	Gln	Pro	Arg	Leu	A
	His	Pro	Arg	Leu	C
	Gln	Pro	Arg	Leu	G
	His	Pro	Arg	Leu	U
G	Glu	Ala	Gly	Val	A
	Asp	Ala	Gly	Val	C
	Glu	Ala	Gly	Val	G
	Asp	Ala	Gly	Val	U
U	STOP	Ser	STOP	Leu	A
	Tyr	Ser	Cys	Phe	C
	STOP	Ser	Trp	Leu	G
	Tyr	Ser	Cys	Phe	U

Fonte: OLIVEIRA, 2012, p. 31.

Os RNAs são produzidos a partir do DNA da seguinte forma: um grupo de três bases consecutivas do mRNA que são representados por trinca é o código correspondente a um aminoácido. O códon necessita então do auxílio do tRNA para identificar os aminoácidos. O tRNA é capaz de ligar-se aos aminoácidos que estão dissolvidos no citoplasma e transportar essas uni-

dades até o mRNA. O tRNA ainda possui em determinada região de sua molécula uma trinca de bases denominada anticódon e, de acordo com ele, o tRNA liga-se a um aminoácido específico. Este é o processo de transcrição.

O processo de tradução é aquele em que ocorre a leitura da mensagem contida no mRNA pelos ribossomos, traduzindo a linguagem de ácido nucleico para a linguagem de proteína.

A seguir serão apresentadas algumas características fundamentais sobre o código genético:

- a) cada códon corresponde a apenas um aminoácido, ou seja, o código genético é inequívoco;
- b) somente dois aminoácidos possuem apenas 1 códon, sendo eles a metionina e o triptofano. Já os outros aminoácidos codificam mais de um códon, ou seja, três aminoácidos têm seis códons possíveis, cinco aminoácidos têm quatro códons, um aminoácido tem três códons e nove aminoácidos têm dois códons possíveis;
- c) os códons UAA, UAG e UGA são códons que não codificam nenhum aminoácido. Esses códons são nomeados **STOP**, ou seja, são códons de pontuação dando a instrução pare;
- d) os códons com bases idênticas nas duas primeiras posições, possuem um aminoácido particular. Contudo, há exceção de quando o aminoácido possuir 6 códons possíveis, os quatro dos códons, possuem as primeiras bases em comum;
- e) quando os códons são semelhantes, eles costumam especificar aminoácidos semelhantes. Por exemplo, os códons da treonina (Thr) diferem dos códons da serina (Ser) num único nucleotídeo na posição 5'. Do mesmo modo, os códons que têm U na segunda posição codificam aminoácidos hidrofóbicos. Então, quando esses códons são alterados, seja na extremidade 3' ou 5', esse aminoácido hidrofóbico ainda é incorporado na proteína. Com base nessa propriedade do código genético e à sua degeneração, as células conseguem gerar uma certa tolerância devido às substituições ocasionadas a um único nucleotídeo, que nem sempre tem como consequência a produção de uma proteína não funcional. Deve-se ressaltar que, os códons que têm A na sua segunda posição codificam aminoácidos hidrofílicos;
- f) o código genético é “quase” **universal**, ou seja, todos os organismos utilizam o mesmo código para traduzir suas proteínas, porém, há algumas exceções que são conhecidas, como por exemplo, as mitocôndrias e certos protozoários ciliados apre-

sentam pequenas diferenças que estão relacionadas com os códons de parada. Essas diferenças aparecem da seguinte forma: em protozoários e mitocôndrias, os códons AGA e AGG são códons de parada, ao invés de codificar arginina e, na mitocôndria, AUA determina metionina em vez de isoleucina e UGA codifica o aminoácido triptofano em vez de STOP.

2.1.8 Mutações

As mutações são mudanças que ocorrem nas sequências de nucleotídeos do material genético de um organismo. De acordo com a alteração provocada, elas podem ser classificadas como desfavoráveis (quando ocasiona erros na sequência de proteínas, causando doenças hereditárias) ou favoráveis (quando são criadas novas proteínas que ajudam o organismo a se adaptar melhor a um ambiente). Nesta perspectiva, o efeito ocasionado na estrutura de um organismo também é classificado como mutações de pequena e de grande escala.

- a) **Mutações de pequena escala:** são aquelas que afetam um pequeno gene em um ou poucos nucleotídeos.
- **Mutação pontual:** É geralmente causada pelos erros na replicação do DNA, onde ocorre a troca de nucleotídeos, podendo ocorrer a transição ou a transversão dos mesmos, sendo a transição a troca de pirimidina por pirimidina ($C \longleftrightarrow U$) ou purina por purina ($A \longleftrightarrow G$), já na transversão ocorre a troca de uma pirimidina por uma purina ou vice-versa ($C/U \longleftrightarrow A/G$). Essas mutações pontuais também são classificadas em três tipos, são eles:
 1. *Mutação silenciosa:* é quando o códon codifica o mesmo aminoácido.
 2. *Mutação “missense”:* é quando um códon codifica um aminoácido distinto.
 3. *Mutação sem sentido:* é quando se codifica para um códon de parada (STOP), interrompendo a proteína antes de seu término.
 - **Inserção:** ocorre pela adição de um ou mais nucleotídeos na sequência do DNA. Essa mutação é ocasionada por erros no processo de replicação de elementos repetitivos.

- **Deleção:** além de serem mutações geralmente irreversíveis, ocorre a remoção de um ou mais nucleotídeos na sequência do DNA.
- b) **Mutações de grande escala:** dizem respeito aos rearranjos cromossômicos que acometem grandes segmentos de DNA.
- **Duplicação gênica:** é a criação de várias cópias de uma região cromossômica, que aumenta a dosagem dos genes dentro dela.
 - **Deleção de regiões cromossômicas:** é a mutação que ocasiona a perda dos genes presentes nessas regiões.
 - **Perda de heterozigosidade:** é onde ocorre a perda de um alelo por deleção num organismo que possuía dois alelos originalmente.

2.2 ELEMENTOS DE ÁLGEBRA

Nesta seção serão apresentados os principais conceitos algébricos, importantes para a compreensão deste estudo e podem ser encontrados em Domingues e Iezzi (2003), Gerônimo e Franco (2008), Hefez e Fernandez (2012), Costelo Jr. e Lin (1983), Oliveira (2012) e Sánchez, Morgado e Grau (2004).

2.2.1 Grupos

Definição 2.1 *Um grupo é um par $(G, *)$, onde G é um conjunto não vazio e $*$ é uma operação binária sobre G , satisfazendo:*

- a) *Associatividade: $a * (b * c) = (a * b) * c$, quaisquer que sejam $a, b, c \in G$;*
- b) *Existência de elemento neutro: existe um elemento neutro $e \in G$ tal que $e * a = a * e = a$, qualquer que seja $a \in G$;*
- c) *Existência de elemento inverso: para todo $a \in G$ existe um elemento $a' \in G$, tal que $a * a' = a' * a = e$.*

Se, além disso, a comutatividade for válida para o grupo G , isto é, $a * b = b * a$ para quaisquer dois elementos a, b de G , então esse grupo é chamado de *grupo abeliano ou comutativo*.

Exemplo 2.1 $(\mathbb{R}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{Z}, +)$, $(\mathbb{R} - \{0\}, \cdot)$, $(\mathbb{Q} - \{0\}, \cdot)$ são exemplos de grupos abelianos.

Os teoremas a seguir decorrem da definição de grupo.

Teorema 2.1 *O elemento neutro de um grupo é único.*

Dem.: Suponhamos que existam dois elementos neutros $e, e' \in G$, assim temos que,

$$e' = e' * e = e,$$

ou seja, $e' = e$.

Portanto, existe um e somente um elemento neutro em G . ■

Teorema 2.2 *O elemento inverso de um grupo é único.*

Dem.: Suponhamos que existam dois elementos inversos $a', a'' \in G$ para um elemento $a \in G$.

Então,

$$a' = a' * e = a' * (a * a'') = (a' * a) * a'' = e * a'' = a'',$$

ou seja, $a' = a''$.

Isto implica que a' e a'' são iguais e existe um único inverso para o elemento a . ■

Definição 2.2 *Um grupo em que seu conjunto é finito, é denominado grupo finito, caso contrário, é denominado de grupo infinito.*

Definição 2.3 *A ordem n de um elemento a pertencente a um grupo G é o menor inteiro positivo tal que $a^n = e$ onde e é a identidade do grupo.*

Definição 2.4 *As operações entre elementos de um grupo finito podem ser representadas em uma tabela de operações, conhecida como Tábua de Cayley.*

Exemplo 2.2 Considere $G = \{e, a\}$ um grupo sob $*$, de ordem 2. Para satisfazer as propriedades de grupo, necessariamente um elemento de G tem que ser um elemento neutro, além disso, o elemento a tem que ser o seu próprio elemento inverso, caso contrário, teríamos dois elementos neutros, o que contradiz a sua unicidade. Este grupo é representado pela tábua, apresentada na Tabela 4:

Tabela 4 – Tábua de $G = \{e, a\}$.

$*$	e	a
e	e	a
a	a	e

Fonte: Da autora.

Definição 2.5 Se um subconjunto H de um grupo G é fechado sob a operação binária sobre G e se H é um grupo sob esta operação binária, então H é chamado subgrupo de G . Escreve-se $H \leq G$.

2.2.2 Anéis

Definição 2.6 Um conjunto não vazio R e um par de operações sobre R , respectivamente uma adição $(x,y) \mapsto x + y$ e uma multiplicação $(x,y) \mapsto xy$ (ou $x \cdot y$), é chamado de anel se estiver sujeito às seguintes propriedades:

Para a adição:

- a) *Associatividade:* $a + (b + c) = (a + b) + c$, quaisquer que sejam $a, b, c \in R$;
- b) *Comutatividade:* $a + b = b + a$, quaisquer que sejam $a, b \in R$;
- c) *Existência de elemento neutro:* existe um elemento neutro $0_R \in R$ tal que, qualquer que seja $a \in R$, $a + 0_R = a = 0_R + a$;
- d) *Existência de elemento inverso:* qualquer que seja $a \in R$, existe um elemento em R , indicado por $-a$, tal que $a + (-a) = 0_R = (-a) + a$.

Para a multiplicação:

- a) *Associatividade:* $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, quaisquer que sejam $a, b, c \in R$;

b) A multiplicação é distributiva em relação a adição, à direita e à esquerda: se $a, b, c \in R$, então $a \cdot (b + c) = a \cdot b + a \cdot c$ e $(b + c) \cdot a = b \cdot a + c \cdot a$.

Definição 2.7 Um anel $(R, +, \cdot)$, onde a operação \cdot é comutativa é denominado de anel comutativo.

Definição 2.8 Se existe um elemento $1 \in R$ tal que $a \cdot 1 = 1 \cdot a = a$ para todo $a \in R$, então chamamos 1 a **unidade ou identidade** do anel R . Quando um anel R possui o elemento identidade dizemos que R é um anel com unidade, ou simplesmente um anel com 1 .

Definição 2.9 Seja $(R, +, \cdot)$ um anel. Um elemento $a \in R$, $a \neq 0$ é um divisor de zero à esquerda de R se existe $b \neq 0$ em R , tal que $a \cdot b = 0$. Analogamente, $a \neq 0$ é um divisor de zero à direita se existe $b \neq 0$ tal que $b \cdot a = 0$.

Exemplo 2.3 O conjunto \mathbb{Z} dos inteiros é um anel, onde $+$ e \cdot são as operações de adição e multiplicação usuais dos inteiros. E temos ainda, que a operação de multiplicação é comutativa e $1 \in \mathbb{Z}$ é o elemento neutro dessa operação. Sendo assim, o conjunto dos inteiros é um anel comutativo com unidade.

Teorema 2.3 Sejam $(R, +, \cdot)$ um anel. Então:

- a) O elemento neutro da soma denotado por 0 , ou 0_R , é único.
- b) Para todo $a \in R$, o oposto de a (o inverso com relação a soma), $-a$, é único.
- c) Valem as leis do cancelamento para a soma.
- d) Para todo $a \in R$, $a \cdot 0 = 0 = 0 \cdot a$.
- e) Para todo $a, b \in R$, $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$ e $(-a) \cdot (-b) = a \cdot b$.
- f) Se R é um anel com 1 , então 1_R é único.
- g) Se R tem mais do que um elemento e R tem 1_R , então $1_R \neq 0_R$.
- h) Se R é um anel no qual vale a lei do cancelamento à esquerda (respectivamente, à direita) para o produto, então R não tem divisores de zero.

Dem.:

a) Suponhamos que existam dois elementos neutros $0_R, 0'_R \in R$. Então,

$$0_R = 0_R + 0'_R = 0'_R,$$

ou seja, $0_R = 0'_R$.

Logo, existe um único elemento neutro em R .

b) Sejam a e a' os inversos de $a \in R$. Então,

$$a + (-a) = 0_R = (-a) + a \quad e \quad a' + (-a) = 0_R = (-a) + a'$$

Logo,

$$\begin{aligned} a &= a + 0_R = a + (a + a') = (a + a) + a' \\ &= 0_R + a' = a' \\ &a = a'. \end{aligned}$$

Portanto, o inverso de a é único.

c) Como R é um grupo abeliano sob a adição e $a, b, c \in R$, segue que para todo elemento $a \in R$ existe um inverso $-a$. Então,

$$\begin{aligned} a + b = a + c &\Rightarrow a + (-a) + b = a + (-a) + c \\ &\Rightarrow b = c. \end{aligned}$$

Portanto, $b = c$. O mesmo raciocínio é utilizado para mostrar que $b + a = c + a \Rightarrow b = c$.

d) Para todo $a \in R$, temos que

$$\begin{aligned} a \cdot 0_R + 0_R &= a \cdot 0_R \\ a \cdot 0_R + 0_R &= a \cdot (0_R + 0_R) \\ a \cdot 0_R + 0_R &= a \cdot 0_R + a \cdot 0_R. \end{aligned}$$

Segue pela lei do cancelamento que $0_R = a \cdot 0_R$.

e) Note que,

$$a \cdot (-b) + a \cdot b = a \cdot (-b + b) = a \cdot 0_R = 0_R. \quad (2.1)$$

Então, $a \cdot (-b) = -(a \cdot b)$. Do mesmo modo,

$$(-a) \cdot b + a \cdot b = (-a + a) \cdot b = 0_R \cdot b = 0_R. \quad (2.2)$$

Por isso, $(-a) \cdot b = -(a \cdot b)$. Portanto, de (2.1) e (2.2), segue que $a \cdot (-b) = a \cdot (-b) = -(a \cdot b)$.

f) Suponhamos que existam dois elementos neutros da multiplicação $1_R, 1'_R \in R$, então,

$$1_R = 1_R \cdot 1'_R = 1'_R.$$

Logo, se R é um anel com unidade, então 1_R é único.

g) Se $1_R = 0_R$, então dado $a \in R$, teríamos que:

$$a = a \cdot 1_R = a \cdot 0_R = 0_R.$$

Assim, $R \subset \{0_R\}$.

Então, $R = \{0_R\}$. Logo, R tem apenas um elemento.

h) Seja $a \in R$, com $a \neq 0_R$. Suponha que $a \cdot b = 0_R$. Então, vamos mostrar que $b = 0_R$.

Neste caso, R não possui divisores de zero.

Se $a \cdot b = 0_R = a \cdot 0_R$. Então,

$$a \cdot b = a \cdot 0_R.$$

Como $a \neq 0$ e pela lei do cancelamento para o produto, temos que $b = 0$.

■

Definição 2.10 Anel das classes de restos módulo m , $\forall m > 1$, é o conjunto $\mathbb{Z}_m = \{\overline{0}, \overline{1}, \overline{2}, \dots, \overline{m-1}\}$, em relação às operações assim definidas:

$$\overline{a} + \overline{b} = \overline{a + b},$$

$$\overline{a} \cdot \overline{b} = \overline{a \cdot b}.$$

O zero desse anel é classe $\bar{0}$ e oposto de um elemento $\bar{a} \in \mathbb{Z}_m$ é classe $\overline{m - a}$.

Exemplo 2.4 Os anéis da classe de resto módulo m , \mathbb{Z}_m ($m > 1$), são exemplos importantes de anéis finitos. Considere as tábuas do anel $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$, apresentadas nas Tabelas 5 e 6.

Tabela 5 – Tábua da adição de $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$.

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Fonte: Da autora.

Tabela 6 – Tábua da multiplicação de $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$.

·	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Fonte: Da autora.

2.2.3 Corpos

Definição 2.11 Seja F um conjunto de elementos sobre o qual duas operações binárias, adição “+” e a multiplicação “·”, são definidas. F , junto com as duas operações binárias, é um corpo se as seguintes condições são satisfeitas:

- F é um grupo comutativo sob “+”. O elemento identidade é o 0 (zero).
- O conjunto dos elementos não zero em F é um grupo comutativo sob “·”. O elemento identidade é o 1 (um).
- A multiplicação é distributiva sob adição à direita e à esquerda, isto é, para quaisquer

a, b, c em F ,

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

$$(b + c) \cdot a = b \cdot a + c \cdot a.$$

O número de elementos em um corpo é chamado de ordem do corpo. Um corpo com um número finito de elementos é chamado um *corpo finito*.

Observação 2.1 Em um corpo, o inverso aditivo de um elemento a é denotado por $-a$ e o inverso multiplicativo de a é denotado por a^{-1} , desde que $a \neq 0$.

Observação 2.2 Subtração de um corpo: a subtração de um elemento a por outro elemento b , ambos pertencentes a um corpo, é definido como:

$$a - b \triangleq a + (-b).$$

Observação 2.3 Divisão em um corpo: a divisão de um elemento a por um elemento não zero b , ambos pertencentes a um corpo, é definida como:

$$a \div b \triangleq a \cdot b^{-1}.$$

Definição 2.12 Um corpo de Galois é um corpo com um número finito de elementos representado por $GF(p)$, sendo p um número primo.

Exemplo 2.5 Considere o conjunto $\{0, 1\}$ cujas operações de adição e multiplicação módulo-2 são apresentadas nas Tabelas 7 e 8.

Tabela 7 – Operação de adição módulo-2.

+	0	1
0	0	1
1	1	0

Fonte: Da autora.

Este corpo é usualmente chamado de *corpo binário* e é denotado por $GF(2)$. Este corpo desempenha um papel importante na teoria de códigos e é amplamente usado em computadores digitais e sistemas de transmissão (ou armazenamento) de dados digitais.

Tabela 8 – Operação de multiplicação módulo-2.

·	0	1
0	0	0
1	0	1

Fonte: Da autora.

Exemplo 2.6 *Seja p um primo. O conjunto dos inteiros $\{0, 1, \dots, p-1\}$ é um grupo comutativo sob a adição módulo- p . Os elementos não nulos $\{1, 2, \dots, p-1\}$ formam um grupo comutativo sob a multiplicação módulo- p . Portanto, esse conjunto é um corpo de ordem p sob a adição e multiplicação módulo- p e é chamado de corpo primo, denotado por $GF(p)$.*

Para qualquer inteiro m é possível estender um corpo primo $GF(p)$ com p elementos para um corpo estendido $GF(p^m)$ com p^m elementos. A ordem de qualquer corpo finito estendido é potência de um primo.

É possível utilizar a maioria das técnicas de álgebra nos cálculos sobre corpos finitos.

2.2.3.1 Aritmética de Corpos Binários

Em geral, podemos construir códigos em símbolos de qualquer corpo de Galois $GF(q)$, em que q é um primo p ou uma potência de p , entretanto, códigos e símbolos de corpo binário $GF(2)$ ou suas extensões $GF(2^m)$ são mais amplamente utilizadas em transmissão de dados digitais e sistemas de armazenamento, porque a informação nestes sistemas é universalmente codificada na forma binária por razões práticas. Neste momento, nos concentraremos apenas em códigos binários e códigos com símbolos do corpo $GF(2^m)$. A maioria dos resultados apresentados podem ser generalizados para códigos com símbolos de qualquer corpo finito $GF(q)$, com $q \neq 2$ ou 2^m . Na aritmética binária a subtração é igual a adição. Nesta aritmética usamos adição e multiplicação módulo-2, definidas nas Tabelas 7 e 8.

Esta aritmética é equivalente à aritmética ordinária, exceto que consideramos 2 igual a 0, isto é, $1 + 1 = 2 = 0$. Observe que, como $1 + 1 = 0$, $1 = -1$. Daí, na aritmética binária, subtração é o mesmo que adição.

Para o estudo de códigos de bloco lineares, as operações polinomiais são de fundamental importância. Uma palavra binária pode ser representada por um polinômio na forma:

$$f(X) = f_0 + f_1X + f_2X^2 + \dots + f_nX^n.$$

Assim, a palavra binária (101011), onde o bit mais significativo é o bit mais à esquerda,

pode ser descrito pelo polinômio:

$$f(X) = 1 + 1X + 0X^2 + 1X^3 + 0X^4 + 1X^5,$$

$$f(X) = 1 + X + X^3 + X^5.$$

A adição (ou subtração) de dois polinômios $a(X) = 1 + X + X^3 + X^5$ e $b(X) = 1 + X^2 + X^3 + X^4 + X^7$ pode ser feita da forma

$$\begin{array}{r} 1 + 1X + 0X^2 + 1X^3 + 0X^4 + 1X^5 + 0X^6 + 0X^7 \\ + 1 + 0X + 1X^2 + 1X^3 + 1X^4 + 0X^5 + 0X^6 + 1X^7 \\ \hline 0 + 1X + 1X^2 + 0X^3 + 1X^4 + 1X^5 + 0X^6 + 1X^7 \\ a(X) + b(X) = X + X^2 + X^4 + X^5 + X^7. \end{array}$$

A multiplicação dos mesmos polinômios $a(X)$ e $b(X)$ é obtida fazendo

$$\begin{array}{r} 1 + 1X + 0X^2 + 1X^3 + 0X^4 + 1X^5 + 0X^6 + 0X^7 \\ \cdot 1 + 0X + 1X^2 + 1X^3 + 1X^4 + 0X^5 + 0X^6 + 1X^7 \\ \hline 1 + 1X + 0X^2 + 1X^3 + 0X^4 + 1X^5 + 0X^6 + 0X^7 \\ 1X^2 + 1X^3 + 0X^4 + 1X^5 + 0X^6 + 1X^7 + 0X^8 + 0X^9 \\ 1X^3 + 1X^4 + 0X^5 + 1X^6 + 0X^7 + 1X^8 + 0X^9 + 0X^{10} \\ 1X^4 + 1X^5 + 0X^6 + 1X^7 + 0X^8 + 1X^9 + 0X^{10} + 0X^{11} \\ 1X^7 + 1X^8 + 0X^9 + 1X^{10} + 0X^{11} + 1X^{12} + 0X^{13} + 0X^{14} \\ \hline 1 + 1X + 1X^2 + 1X^3 + 0X^4 + 1X^5 + 1X^6 + 1X^7 + 0X^8 + 1X^9 + 1X^{10} + 0X^{11} + 1X^{12} + 0X^{13} + 0X^{14} \end{array}$$

$$a(X) \cdot b(X) = 1 + X + X^2 + X^3 + X^5 + X^6 + X^7 + X^9 + X^{10} + X^{12}.$$

Evidentemente, se $a(X) = 0$, então $a(X) \cdot b(X) = 0$.

A divisão entre um polinômio $f(X)$ e um polinômio $g(X)$, no caso em que $f(X)$ possui grau maior ou igual do que $g(X)$ e $g(X)$ não é zero, resulta em um quociente $q(X)$ e um resto $r(X)$ e a operação pode ser escrita na forma do *algoritmo de divisão de Euclides* como:

$$f(X) = q(X)g(X) + r(X).$$

Exemplo 2.7 Seja $f(X) = X^6 + X^5 + X^4 + X + 1$ e $g(X) = X^3 + X + 1$. Então a divisão

de $f(X)$ por $g(X)$ pode ser feita conforme será mostrado a seguir:

$$\begin{array}{r}
 X^6 + X^5 + X^4 \quad \quad \quad + X + 1 \\
 \hline
 X^6 \quad \quad + X^4 + X^3 \\
 \hline
 X^5 \quad \quad + X^3 \quad \quad + X + 1 \\
 \hline
 X^5 \quad \quad + X^3 + X^2 \\
 \hline
 X^2 + X + 1 \text{ (resto)}
 \end{array}
 \quad
 \begin{array}{l}
 |X^3 + X + 1 \\
 \hline
 X^3 + X^2 \text{ (quociente)}
 \end{array}$$

Um polinômio $f(X)$ sobre $GF(2)$, com número par de termos, é divisível por $X + 1$.

Segundo o Teorema do Resto (Domingues e Iezzi, 2003, p. 297), temos que o resto da divisão de f por $X - a$ é $f(a)$. Neste caso, como $-1 = 1$ em $GF(2)$, então ser divisível por $X + 1$ é o mesmo que ser divisível por $X - 1$, o que corresponde ao fato de $f(1) = 0$, segundo o Teorema do Resto. Agora, tendo f um número par de termos, cujos coeficientes são compostos apenas por 1's, ao avaliarmos $f(1)$ obtemos necessariamente um número par, que em $GF(2)$ é igual a zero. Portanto, $f(1) = 0$ e o resultado segue.

Definição 2.13 Um polinômio $p(X)$ sobre $GF(2)$ de grau m é dito irredutível sobre $GF(2)$ se ele não for divisível por nenhum outro polinômio sobre $GF(2)$ de grau menor que m , mas maior que zero.

Exemplo 2.8 Os polinômios: $X^2 + X + 1$; $X^3 + X + 1$; $X^4 + X + 1$ são irredutíveis (Costelo Jr. e Lin (1983)).

Existem apenas dois polinômios de grau 1 sobre $GF(2)$, que são: X e $X + 1$.

Seja $f_1(X) = X^2 + X + 1$. O Teorema do Resto nos diz que o fato de $f_1(0) = 1 \neq 0$ e $f_1(1) = 3 = 1 \neq 0$, resulta que f_1 não pode ser divisível por X e nem por $X + 1$. Portanto, $f_1(X)$ é um polinômio irredutível de grau 2.

Seja $f_2(X) = X^3 + X + 1$. Com um raciocínio análogo, concluímos que $f_2(X)$ não é divisível por X e nem por $X + 1$, pois $f_2(0) \neq 0$ e $f_2(1) \neq 0$. Como $f_2(X)$ não é divisível por nenhum polinômio de grau 1, ele não pode ser divisível por um polinômio de grau 2. Lembrando que a única maneira de fatorar $f_2(X)$, sem que um dos fatores seja um polinômio constante, e no caso em que um polinômio tem grau 1 e o outro tem grau 2, para que a soma dos graus resulte em $3 = \text{grau } f_2$. Portanto, $f_2(X)$ é irredutível sobre $GF(2)$.

Seja $f_3(X) = X^4 + X + 1$. Como $f_3(0) \neq 0$ e $f_3(1) \neq 0$, concluímos que $f_3(X)$ não se fatora como um produto de um polinômio de grau 1 por um polinômio de grau 3. A

possibilidade que resta é através do produto de dois polinômios de grau 2. Existem quatro polinômios de grau 2 sobre $GF(2)$, que são: X^2 ; $1 + X^2$; $X + X^2$ e $1 + X + X^2$. Uma vez que o produto entre eles nunca resulta em $f_3(X)$, concluímos que $f_3(X)$ é um polinômio irredutível de grau 4.

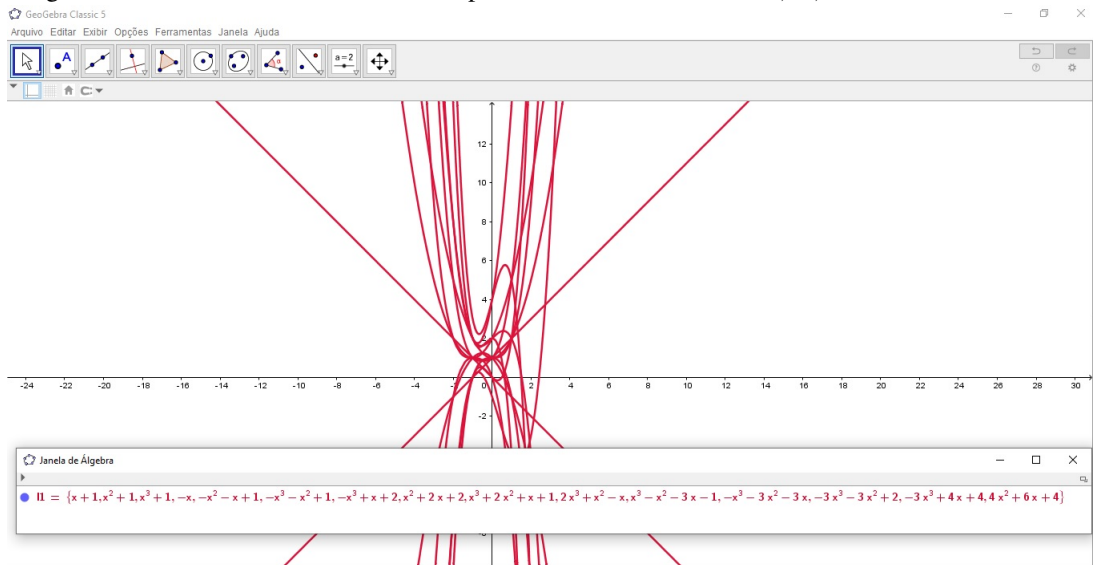
Definição 2.14 Um polinômio irredutível $p(X)$ de grau m é dito primitivo se o menor inteiro positivo n para o qual $p(X)$ divide $X^n + 1$ é $n = 2^m - 1$.

Exemplo 2.9 O polinômio $p(X) = X^4 + X + 1$ é irredutível e primitivo, pois $p(X)$ divide $X^{15} + 1$ e nenhum $X^n + 1$ para $1 \leq n < 15$ (Costelo Jr. e Lin (1983)).

Usando o GeoGebra, podemos listar os restos da divisão de $X^n + 1$ por $X^4 + X + 1$, com $n = 1; \dots; 15$, utilizando o comando:

Sequência (Resto (X^{m+1}, X^4+X+1), $m, 1, 15$) .

Figura 13 – Restos da divisão de $X^n + 1$ por $X^4 + X + 1$, com $n = 1; \dots; 15$.



Fonte: Da autora.

Observe na Figura 13, que os restos obtidos são:

$x + 1, x^2 + 1, x^3 + 1, -x, -x^2 - x + 1, -x^3 - x^2 + 1, -x^3 + x + 2, x^2 + 2x + 2, x^3 + 2x^2 + x + 1, 2x^3 + x^2 - x, x^3 - x^2 - 3x - 1, -x^3 - 3x^2 - 3x, -3x^3 - 3x^2 + 2, -3x^3 + 4x + 4, 4x^2 + 6x + 4.$

Note que o único resto cujos coeficientes são todos pares, ou seja, resto nulo em $GF(2)$, é o proveniente de $n = 15$.

Por outro lado,

$$p(X) = X^4 + X^3 + X^2 + X + 1$$

é irredutível, porém não é primitivo, pois além de $p(X)$ dividir $X^{15} + 1$ ele também divide $X^5 + 1$, uma vez que sobre $GF(2)$, temos:

$$x^{15} + 1 = (x + 1).(x^2 + x + 1).(x^4 + x^3 + x^2 + x + 1).(x^8 + x^7 + x^5 + x^4 + x^3 + x + 1)$$

e

$$x^5 + 1 = (x + 1).(x^4 + x^3 + x^2 + x + 1).$$

Não é fácil reconhecer um polinômio primitivo. Para um dado m , pode haver mais do que um polinômio primitivo de grau m . A Tabela 9 apresenta apenas um polinômio primitivo para cada valor de m . Os polinômios apresentados são os que possuem o menor número de termos.

Tabela 9 – Lista de polinômios primitivos.

m	$p(X)$
3	$1 + X + X^3$
4	$1 + X + X^4$
5	$1 + X^2 + X^5$
6	$1 + X + X^6$
7	$1 + X^3 + X^7$
8	$1 + X^2 + X^3 + X^4 + X^8$
9	$1 + X^4 + X^9$
10	$1 + X^3 + X^{10}$
11	$1 + X^2 + X^{11}$
12	$1 + X + X^4 + X^6 + X^{12}$
13	$1 + X + X^3 + X^4 + X^{13}$
14	$1 + X + X^6 + X^{10} + X^{14}$
15	$1 + X + X^{15}$
16	$1 + X + X^3 + X^{12} + X^{16}$
17	$1 + X^3 + X^{17}$
18	$1 + X^7 + X^{18}$
19	$1 + X + X^2 + X^5 + X^{19}$
20	$1 + X^3 + X^{20}$
21	$1 + X^2 + X^{21}$
22	$1 + X + X^{22}$
23	$1 + X^5 + X^{23}$
24	$1 + X + X^2 + X^7 + X^{24}$

Fonte: COSTELO JR. E LIN, 1983, p.29.

2.2.4 Construção de Corpos de Galois $GF(2^m)$

Nesta seção será apresentado um método para construir corpos de Galois de 2^m elementos ($m > 1$) do corpo binário $GF(2)$. Começamos com os dois elementos 0 e 1 de $GF(2)$ e um novo símbolo α . Então definimos uma multiplicação “ \cdot ” para introduzir uma sequência de

potência de α como segue:

$$\begin{aligned}
 0 \cdot 0 &= 0, \\
 0 \cdot 1 &= 0 = 1 \cdot 0, \\
 1 \cdot 1 &= 1, \\
 0 \cdot \alpha &= 0 = \alpha \cdot 0, \\
 1 \cdot \alpha &= \alpha = \alpha \cdot 1, \\
 \alpha^2 &= \alpha \cdot \alpha, \\
 \alpha^3 &= \alpha \cdot \alpha \cdot \alpha, \\
 &\vdots \\
 \alpha^j &= \alpha \cdot \alpha \cdot \dots \cdot \alpha \quad (j - \text{vezes}) \\
 &\vdots
 \end{aligned} \tag{2.3}$$

Segue da definição de multiplicação precedente que:

$$\begin{aligned}
 0 \cdot \alpha^j &= \alpha^j \cdot 0 = 0, \\
 1 \cdot \alpha^j &= \alpha^j \cdot 1 = \alpha^j, \\
 \alpha^i \cdot \alpha^j &= \alpha^j \cdot \alpha^i = \alpha^{i+j}.
 \end{aligned}$$

Agora, temos que o seguinte conjunto de elementos sobre os quais a operação multiplicação “ \cdot ” está definida:

$$F = \{0, 1, \alpha, \alpha^2, \dots, \alpha^j, \dots\}.$$

O elemento 1 é algumas vezes denotado por α^0 .

Suponha que o conjunto F contenha apenas 2^m elementos e seja fechado sobre a multiplicação “ \cdot ” definida em (2.3).

Seja $p(X)$ um polinômio primitivo de grau m sobre $GF(2)$. Assumimos que $p(\alpha) = 0$ (isto é, α é uma raiz de $p(X)$). Como $p(X)$ divide $X^{2^m-1} + 1$, temos:

$$X^{2^m-1} + 1 = q(X)p(X).$$

Se trocarmos X por α , obtemos:

$$\alpha^{2^m-1} + 1 = q(\alpha)p(\alpha).$$

Como $p(\alpha) = 0$, temos:

$$\alpha^{2^m-1} + 1 = q(\alpha) \cdot 0.$$

Considere $q(\alpha)$ como um polinômio de α sobre $GF(2)$, segue que $q(\alpha) \cdot 0 = 0$. Como resultado, obtemos:

$$\alpha^{2^m-1} + 1 = 0.$$

Adicionando 1 a ambos os lados de $\alpha^{2^m-1} + 1 = 0$ (usando adição módulo-2), obtemos:

$$\alpha^{2^m-1} = 1 = \alpha^0. \quad (2.4)$$

Assim, sob a condição que $p(\alpha) = 0$, o conjunto F se torna finito e contendo os seguintes elementos:

$$F^* = \{0, 1, \alpha, \alpha^2, \dots, \alpha^{2^m-2}\}. \quad (2.5)$$

F^* possui ordem 2^m e os seus elementos não nulos são fechados sob a operação multiplicação “.”.

Para construir $GF(2^m)$ de $GF(2)$, serão desenvolvidas duas representações para os elementos não nulos de $GF(2^m)$: a representação por potência e a representação polinomial. A representação por potência é conveniente para a multiplicação e a representação polinomial é conveniente para a adição.

Exemplo 2.10 *Seja $m = 4$ e considere o polinômio primitivo $p(X) = 1 + X + X^4$ sobre $GF(2)$. Admitindo que α seja uma raiz do polinômio, então $p(\alpha) = 0$, ou seja,*

$$0 = 1 + \alpha + \alpha^4 \quad \Rightarrow \quad \alpha^4 = 1 + \alpha \quad (2.6)$$

Usando a relação (2.6), pode-se construir um $GF(2^4)$ como se segue:

$$\alpha^5 = \alpha \cdot \alpha^4 = \alpha(1 + \alpha) = \alpha + \alpha^2$$

$$\alpha^6 = \alpha \cdot \alpha^5 = \alpha(\alpha + \alpha^2) = \alpha^2 + \alpha^3$$

$$\alpha^7 = \alpha \cdot \alpha^6 = \alpha(\alpha^2 + \alpha^3) = \alpha^3 + \alpha^4 = \alpha^3 + 1 + \alpha = 1 + \alpha + \alpha^3$$

$$\alpha^8 = \alpha \cdot \alpha^7 = \alpha(1 + \alpha + \alpha^3) = \alpha + \alpha^2 + \alpha^4 = \alpha + \alpha^2 + 1 + \alpha = 1 + \alpha^2$$

$$\alpha^9 = \alpha \cdot \alpha^8 = \alpha(1 + \alpha^2) = \alpha + \alpha^3$$

$$\alpha^{10} = \alpha \cdot \alpha^9 = \alpha(\alpha + \alpha^3) = \alpha^2 + \alpha^4 = \alpha^2 + 1 + \alpha = 1 + \alpha + \alpha^2$$

$$\alpha^{11} = \alpha \cdot \alpha^{10} = \alpha(1 + \alpha + \alpha^2) = \alpha + \alpha^2 + \alpha^3$$

$$\alpha^{12} = \alpha \cdot \alpha^{11} = \alpha(\alpha + \alpha^2 + \alpha^3) = \alpha^2 + \alpha^3 + \alpha^4 = \alpha^2 + \alpha^3 + 1 + \alpha = 1 + \alpha + \alpha^2 + \alpha^3$$

$$\alpha^{13} = \alpha \cdot \alpha^{12} = \alpha(1 + \alpha + \alpha^2 + \alpha^3) = \alpha + \alpha^2 + \alpha^3 + \alpha^4 = \alpha + \alpha^2 + \alpha^3 + 1 + \alpha = 1 + \alpha^2 + \alpha^3$$

$$\alpha^{14} = \alpha \cdot \alpha^{13} = \alpha(1 + \alpha^2 + \alpha^3) = \alpha + \alpha^3 + \alpha^4 = \alpha + \alpha^3 + 1 + \alpha = 1 + \alpha^3$$

$$\alpha^{15} = \alpha \cdot \alpha^{14} = \alpha(1 + \alpha^3) = \alpha + \alpha^4 = \alpha + 1 + \alpha = \underbrace{1}_{\text{voltou!}}$$

As representações por potência, polinomial e vetorial para $GF(2^4)$ são apresentadas na Tabela 10.

Tabela 10 – $GF(2^4)$ gerado por $p(X) = 1 + X + X^4$.

POTÊNCIA	POLINOMIAL	VETORIAL
0	0	(0000)
$\alpha^0 = 1$	1	(1000)
α^1	α	(0100)
α^2	α^2	(0010)
α^3	α^3	(0001)
α^4	$1 + \alpha$	(1100)
α^5	$\alpha + \alpha^2$	(0110)
α^6	$\alpha^2 + \alpha^3$	(0011)
α^7	$1 + \alpha + \alpha^3$	(1101)
α^8	$1 + \alpha^2$	(1010)
α^9	$\alpha + \alpha^3$	(0101)
α^{10}	$1 + \alpha + \alpha^2$	(1110)
α^{11}	$\alpha + \alpha^2 + \alpha^3$	(0111)
α^{12}	$1 + \alpha + \alpha^2 + \alpha^3$	(1111)
α^{13}	$1 + \alpha^2 + \alpha^3$	(1011)
α^{14}	$1 + \alpha^3$	(1001)

Fonte: Da autora.

2.2.5 Relações de Ordem

Nesta seção serão apresentados os principais elementos sobre Relações de Ordem, detalhados da seguinte maneira: na subseção 2.2.5.1 a estrutura de Reticulados Booleanos e na subseção 2.2.5.2 os diagramas de Hasse.

2.2.5.1 Reticulados Booleanos

Definição 2.15 *Uma relação \preceq em A diz-se de ordem parcial (ROP) se satisfaz as propriedades:*

- a) *Reflexiva: $x \preceq x$ para todo $x \in A$.*
- b) *Transitiva: $x \preceq y$ e $y \preceq z$ implica $x \preceq z$.*
- c) *Anti-simétrica: $x \preceq y$ e $y \preceq x$ implica $x = y$.*

Definição 2.16 *(A, \preceq) é um conjunto parcialmente ordenado (CPO) se \preceq for uma relação de ordem parcial definida num conjunto não vazio A .*

Definição 2.17 *Um Reticulado é um conjunto parcialmente ordenado no qual todo par de elementos do conjunto possui simultaneamente soma e produto.*

Reticulado como Álgebra: Conectivos \wedge e \vee

Considere os valores verdade F e V (Falso e Verdadeiro), bem como os conectivos lógicos \wedge e \vee . Se considerarmos 0 e 1 ao invés de F e V para os conectivos lógicos temos que a estrutura $\langle \{F, V\}, \wedge, \vee \rangle$ constitui um reticulado, ou seja, a conjunção e disjunção (Tabelas 11 e 12) levam a um reticulado booleano atendendo as suas propriedades. Então, utilizando o estudo da lógica, pode-se verificar que as estruturas $\langle \{F, V\}, \wedge \rangle$ e $\langle \{F, V\}, \vee \rangle$ são semigrupos abelianos (pois satisfazem as propriedades associativa, comutativa e as operações são fechadas) e, que os conectivos \wedge e \vee satisfazem à propriedade de absorção, ou seja, $a \vee (a \wedge b) = a$ e $a \wedge (a \vee b) = a$.

Tabela 11 – Operação com o conectivo \wedge (conjunção).

A	B	$A \wedge B$
1	1	1
1	0	0
0	1	0
0	0	0

Fonte: Da autora.

Tabela 12 – Operação com o conectivo \vee (disjunção).

A	B	$A \vee B$
1	1	1
1	0	1
0	1	1
0	0	0

Fonte: Da autora.

Definição 2.18 Um reticulado booleano $B(X)$ é um conjunto parcialmente ordenado de elementos com as seguintes propriedades:

a) $B(X)$ contém dois elementos chamados de elementos mínimo e máximo, denotados por 0 e 1 , respectivamente, que são limites universais, ou seja: $0 \leq \alpha \leq 1$ para todos os $\alpha \in X$ e satisfaz as propriedades especiais:

- Interseção: $0 \wedge \alpha = 0$ e $1 \wedge \alpha = \alpha$.
- União: $0 \vee \alpha = \alpha$ e $1 \vee \alpha = 1$.

b) Para todos os elementos $\alpha \in X$ existe o elemento $\neg\alpha \in X$, chamado complemento do elemento α , de tal forma que:

$$\alpha \wedge \neg\alpha = 0 \quad e \quad \alpha \vee \neg\alpha = 1.$$

c) Em $B(X)$, as operações \wedge e \vee satisfazem a lei distributiva, isto é:

$$(\alpha \wedge \beta) \vee (\alpha \wedge \delta) = \alpha \wedge (\beta \vee \delta) \quad e \quad (\alpha \vee \beta) \wedge (\alpha \vee \delta) = \alpha \vee (\beta \wedge \delta).$$

2.2.5.2 Diagrama de Hasse

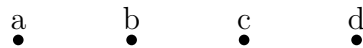
Definição 2.19 Quando um conjunto possui um número finito de elementos, a relação de ordem possui uma representação gráfica adequada para as suas propriedades. Essa representação é denominada “**diagrama de Hasse**” ou “**diagrama de linha**”, que é uma forma mais simplificada do diagrama sagital. Esse tipo de representação torna mais evidente o comportamento dos elementos do conjunto dado pela relação.

Considere uma relação de ordem R sobre um conjunto A , a construção se faz da seguinte maneira:

- a) Cada elemento do conjunto A é representado por um ponto;
- b) Se um elemento x estiver relacionado com um elemento y , representaremos isso por um segmento de reta unindo ambos;
- c) A propriedade reflexiva é omitida na representação, ou seja, não colocaremos os laços em torno de cada elemento de A , como se faz no diagrama sagital;
- d) A propriedade transitiva fica subentendida na representação, ou seja, xRy e yRz , temos xRz , mas o segmento indicativo de xRz não é colocado;
- e) A representação será orientada de baixo para cima, ou seja, se xRy , então o elemento x será colocado em uma posição abaixo do elemento y .

Exemplo 2.11 Considere o conjunto $A = \{a, b, c, d\}$. A relação identidade I_A é representada pelo diagrama da Figura 14. Observe que não há segmentos ligando os pontos.

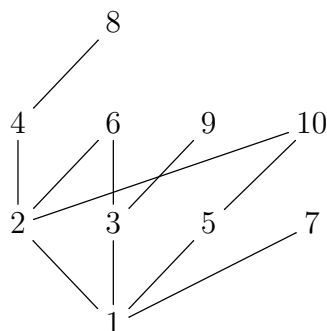
Figura 14 – Relação identidade I_A .



Fonte: Da autora.

Exemplo 2.12 A relação $R = \{(x, y) \in \mathbb{N} \times \mathbb{N} \mid x \text{ divide } y\}$, tem como diagrama parcial a representação apresentada na Figura 15. Escolhemos para representar apenas um subconjunto de \mathbb{N} . Nesse caso, $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

Figura 15 – Diagrama de Hasse.



Fonte: Da autora.

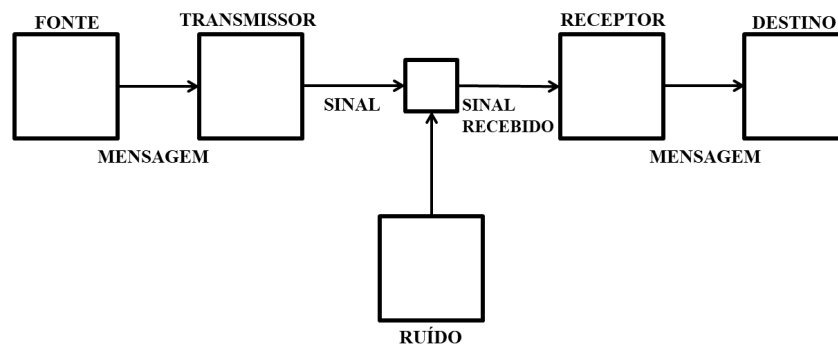
2.3 CÓDIGOS CORRETORES DE ERROS

Nesta seção, serão apresentados os principais conceitos de Códigos Corretores de Erros, importantes para a compreensão deste estudo e podem ser encontrados em MacWilliams e Sloane (1981) e Oliveira (2012).

Um canal de comunicação pode apresentar uma série de perturbações que podem dificultar o entendimento dos dados enviados, essas perturbações podem ser distorções, ruídos, etc (OLIVEIRA, 2012). Um sistema de comunicação pode ser considerado um conjunto de mecanismos que possibilita a transmissão de informação de um transmissor para um determinado receptor, através de um canal de comunicação.

Um sistema de comunicação possui alguns elementos importantes, apresentados na Figura 16.

Figura 16 – Elementos importantes em um sistema de comunicação.



Fonte: Da autora.

1. **Fonte de informação:** produz a mensagem que será enviada.
2. **Transmissor:** possibilita enviar a mensagem com um sinal adequado.
3. **Canal:** meio utilizado para enviar a mensagem do transmissor para o receptor.
4. **Receptor:** realiza as ações inversas ao que foi feito pelo transmissor, reconstruindo a mensagem.
5. **Destino:** para quem / onde a mensagem foi destinada.

A seguir serão apresentadas algumas definições importantes relacionadas ao estudo da teoria de códigos.

Definição 2.20 *Código é um conjunto de símbolos usados na transmissão e recepção de mensagens. Os elementos básicos para se construir um código são:*

- **Alfabeto:** *um conjunto finito de elementos. Cada um desses elementos chama-se dígito. Quando o número de elementos do alfabeto é q , diz-se que o código é q -ário.*
- **Palavras-código:** *é uma sequência finita de dígitos. O número de dígitos de uma palavra-código é o seu comprimento.*

Definição 2.21 *Codificação é o processo de mapeamento, ou seja, é uma conversão de uma dada sequência de dígitos (alfabeto fonte) em uma outra sequência de dígitos (alfabeto do código).*

Matematicamente, uma codificação é uma função injetiva que a cada símbolo do alfabeto fonte faz corresponder uma palavra-código. Dada uma codificação f , com alfabeto fonte $F = \{s_1, s_2, \dots, s_n\}$, as palavras $f(s_1), f(s_2), \dots, f(s_n)$ são palavras-código e o código, C , é o conjunto dessas palavras-código.

$$C = \{f(s_1), f(s_2), \dots, f(s_n)\}.$$

2.3.1 Códigos de Bloco

Os códigos de bloco se caracterizam pelo fato do processo de codificação ser feito sobre blocos de bits ou blocos de símbolos. Um feixe de bits ou símbolos é segmentado em blocos de k bits ou símbolos, a partir dos quais são geradas palavras-código com n bits ou símbolos. Assim, a notação que caracteriza um código de bloco é $C(n, k)$. Se k bits estão contidos em um bloco de n bits, então a quantidade de bits de redundância introduzidos no processo de codificação é $n - k$. A quantidade de palavras-código diferentes que podem ser geradas é q^k . Para os códigos binários, teremos 2^k palavras-código.

O processo de codificação por bloco consiste em transformar um segmento da mensagem, m , com k bits em palavra-código ou **vetor código**, c , com n bits. As alterações sofridas pelo vetor código podem ser representadas por meio de um **vetor erro**, e . Um vetor código, c , somado com um vetor erro, e , resulta em um **vetor recebido**, r .

$$r = c \oplus e,$$

onde, \oplus representa a soma direta, módulo-2, entre os vetores c e e .

O vetor recebido é entregue ao decodificador, cuja finalidade é transformar o vetor recebido no **vetor decodificado**, que consiste na melhor estimativa do vetor transmitido. A partir do **vetor código estimado**, c' , a melhor estimativa da mensagem, m' , é reproduzida na saída do decodificador.

Definição 2.22 O peso de Hamming de um vetor v , cuja notação é $\omega(v)$, é definido como sendo o número de elementos não nulos em v . Para um vetor binário, o peso de Hamming é igual ao número de dígitos “1” contidos em v .

Definição 2.23 A distância de Hamming entre dois vetores códigos, v e x , cuja notação é $d(v, x)$, é definida como sendo o número de posições em que os dígitos dos dois vetores que são diferentes entre si. Para o caso binário, a distância de Hamming pode ser determinada facilmente pela propriedade de adição módulo-2, pois ela é igual ao número de dígitos “1” contidos no vetor resultante da operação $v \oplus x$.

$$d(v, x) = \omega(v \oplus x).$$

Exemplo 2.13 Determine a distância de Hamming entre os vetores $v = 10110$ e $x = 10101$.

$$d(v, x) = \omega(v \oplus x) = \omega(10110 \oplus 10101) = \omega(00011) = d(v, x) = 2.$$

Definição 2.24 Uma métrica em um conjunto M é uma função,

$$\begin{aligned} d : M \times M &\rightarrow \mathbb{R} \\ (x, y) &\mapsto d(x, y) \end{aligned}$$

satisfazendo:

- (a) $d(x, y) \geq 0$ e $d(x, y) = 0 \Leftrightarrow x = y$;
- (b) $d(x, y) = d(y, x)$;
- (c) $d(x, z) \leq d(x, y) + d(y, z), \forall x, y, z \in M$.

Proposição 2.1 A distância de Hamming é uma métrica.

Dem.: Seja um código corretor de erros um subconjunto próprio de A^n , para algum número natural n . Dados dois elementos u e v em A^n , definimos a distância de Hamming entre eles como

$$d(u, v) = \#\{i; u_i \neq v_i, i \in [n]\},$$

onde $\#$ representa a cardinalidade, ou seja, o número de elementos.

(a)

$$d(u, v) \in \mathbb{N} \setminus \{0\}, \forall u, v \in A^n \Rightarrow d(u, v) \geq 0, \forall u, v \in A^n.$$

(b)

$$\begin{aligned} d(u, v) &= \#\{i; u_i \neq v_i, i \in [n]\} \\ &= \#\{i; v_i \neq u_i, i \in [n]\} \\ &= d(v, u). \end{aligned}$$

(c) Defina S :

$$\begin{aligned} S : A^n \times A^n &: \rightarrow \mathcal{P}([n]) \\ (u, v) &\mapsto \{i; u_i \neq v_i, i \in [n]\} \end{aligned}$$

de modo que $d(u, v) = \#S(u, v)$.

Daí, sejam $u, v, w \in A^n$.

$$\begin{aligned} i \in S(u, v) &\Rightarrow u_i \neq v_i \Rightarrow u_i \neq w_i \text{ ou } v_i \neq w_i \\ &\Rightarrow i \in S(u, w) \text{ ou } i \in S(v, w) \\ &\Rightarrow S(u, v) \subset S(u, w) \cup S(v, w) \\ &\Rightarrow \#S(u, v) \leq \#(S(u, w) \cup S(v, w)) \leq \#S(u, w) + \#S(v, w) \\ &\therefore d(u, v) \leq d(u, w) + d(v, w), \forall u, v, w \in A^n. \end{aligned}$$

■

3 CONSTRUÇÃO E ANÁLISE DOS RETICULADOS BOOLEANOS E DIAGRAMAS DE HASSE ASSOCIADOS AOS ROTULAMENTOS DO CÓDIGO GENÉTICO

Neste capítulo serão apresentadas as contribuições deste trabalho, por meio das construções dos reticulados booleanos, diagramas de Hasse e as médias das distâncias de Hamming entre os pares de aminoácidos, como as distâncias médias entre os seus respectivos códons, por meio de uma permutação escolhida para cada um dos rotulamentos: *A*, *B* e *C*.

Para a construção dos reticulados booleanos devemos construir as “tabelas verdade” utilizando os conectivos lógicos “ \wedge ” e “ \vee ” das operações da álgebra booleana, por meio das associações que serão obtidas da permutação escolhida em cada rotulamento. Para a realização dessas operações, utiliza-se os dados apresentados nas Tabelas 11 e 12.

O diagrama de Hasse é uma forma organizada de apresentar os códons do código genético, com o objetivo de analisar as propriedades dos aminoácidos e classificar os códons de acordo com a sua hidropacidade e outras características. O diagrama é composto por 64 códons, dispostos em 7 linhas. Cada linha apresenta um determinado número de códons e sua construção detalhada é observada no Rotulamento *A*.

Conforme a proposta de Sánchez, Morgado e Grau (2004), apresentamos em cada rotulamento as médias das distâncias de Hamming entre os pares de aminoácidos, como as distâncias médias entre os seus respectivos códons. Pode-se notar que os valores das médias das distâncias de Hamming maiores ou iguais a 4 aparecem em negrito, pois são considerados valores altos para essas distâncias, de acordo com Sánchez, Morgado e Grau (2004). Um fato a ser destacado, é que os resultados expressos na Tabela 19 do Rotulamento *A*, são iguais ao caso proposto por Sánchez, Morgado e Grau (2004), que utilizou outra permutação no mapeamento do código genético.

Segundo Sánchez, Morgado e Grau (2004), quando se obtém valores maiores nas médias das distâncias de Hamming entre pares de códons, é sabido que elas são as mais perigosas, pois pode-se alterar as propriedades dos aminoácidos e as funções biológicas das proteínas, acarretando até riscos de grandes mutações genéticas. Pode-se notar que aminoácidos com grandes diferenças em suas propriedades apresentaram altos valores nas médias das distâncias de Hamming.

3.1 ROTULAMENTO A

Para o estudo do Rotulamento A, será utilizada a permutação 0132, para o reticulado primal. Como consequência, a permutação 2310 do Rotulamento A será utilizada para a construção do reticulado dual. Uma observação é que a atribuição $\{0, 1, 2, 3\}$ de \mathbb{Z}_4 é feita em relação a ordem $\{A, C, G, U\}$ em N , ou seja, as bases são estabelecidas na seguinte ordem: adenina (A), citosina (C), guanina (G) e por fim, uracila (U). Utilizando as permutações escolhidas, foram estabelecidas as seguintes associações, apresentadas nas Tabelas 13 e 14, levando em consideração a ordem das bases apresentadas anteriormente e a associação com elementos do conjunto $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Tabela 13 – Associação estabelecida pela permutação 0132 primal do Rotulamento A.

A	C	G	U
0	1	3	2
00	10	01	11

Fonte: Da autora.

Tabela 14 – Associação estabelecida pela permutação 2310 dual do Rotulamento A.

A	C	G	U
2	3	1	0
11	01	10	00

Fonte: Da autora.

A seguir será apresentado o procedimento para a construção dos reticulados booleanos primal e dual.

3.1.1 Procedimento para a construção dos reticulados booleanos

Utilizando as operações da álgebra booleana (Tabelas 11 e 12) e a permutação 0132 do Rotulamento A, obtemos os dados das Tabelas 15 e 16.

Tabela 15 – Operação primal \wedge (e) do Rotulamento A.

\wedge	00	01	10	11
00	00	00	00	00
01	00	01	00	01
10	00	00	10	10
11	00	01	10	11

Fonte: Da autora.

\wedge	A	G	C	U
A	A	A	A	A
G	A	G	A	G
C	A	A	C	C
U	A	G	C	U

Fonte: Da autora.

Tabela 16 – Operação primal \vee (ou) do Rotulamento A.

\vee	00	01	10	11
00	00	01	10	11
01	01	01	11	11
10	10	11	10	11
11	11	11	11	11

Fonte: Da autora.

\vee	A	G	C	U
A	A	G	C	U
G	G	G	U	U
C	C	U	C	U
U	U	U	U	U

Fonte: Da autora.

Utilizando as operações da álgebra booleana (Tabelas 11 e 12) e a permutação 2310 do Rotulamento A, obtemos os dados das Tabelas 17 e 18.

Tabela 17 – Operação dual \wedge (e) do Rotulamento A.

\wedge	00	01	10	11
00	00	00	00	00
01	00	01	00	01
10	00	00	10	10
11	00	01	10	11

Fonte: Da autora.

\wedge	U	C	G	A
U	U	U	U	U
C	U	C	U	C
G	U	U	G	G
A	U	C	G	A

Fonte: Da autora.

Tabela 18 – Operação dual \vee (ou) do Rotulamento A.

\vee	00	01	10	11
00	00	01	10	11
01	01	01	11	11
10	10	11	10	11
11	11	11	11	11

Fonte: Da autora.

\vee	U	C	G	A
U	U	C	G	A
C	C	C	A	A
G	G	A	G	A
A	A	A	A	A

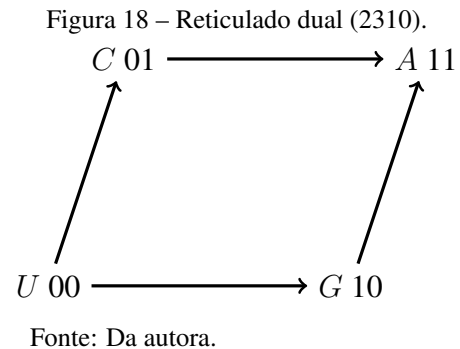
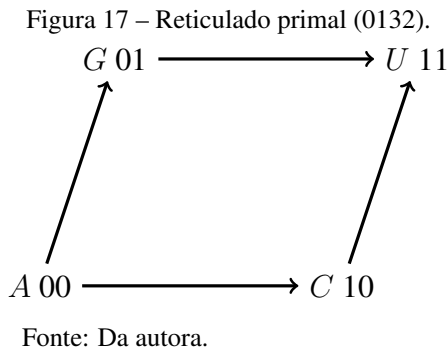
Fonte: Da autora.

O reticulado booleano primal será obtido a partir da Tabela 16. Observe que:

- A base A (adenina) se liga às bases G (guanina), C (citosina) e U (uracila).
- A base G (guanina) se liga a ela mesma e à base U (uracila).
- A base C (citosina) se liga a ela mesma e à base U (uracila).
- A base U (uracila) se liga apenas a ela mesma.

Conforme essas observações, obtém-se a construção seguindo o caminho que as bases percorrem. O reticulado booleano dual será obtido de maneira análoga à construção do reticulado primal, sendo ele a partir da Tabela 18. Assim, obtemos as representações nas Figuras 17 e 18.

Reticulados primal e dual associados ao Rotulamento A.



3.1.2 Procedimento para a construção dos Diagramas de Hasse

O diagrama de Hasse é composto por 64 códons, dispostos em 7 linhas, seguindo o procedimento:

- Na **1ª linha** tem-se o códon que apresenta seu elemento máximo, atribuído por 11. Por exemplo: $UUU \longleftrightarrow 111111$.
- Na **7ª linha** tem-se o códon que apresenta elemento mínimo, atribuído por 00. Por exemplo: $AAA \longleftrightarrow 000000$.
- A **2ª linha** é constituída por seis códons organizados da esquerda para a direita. Biologicamente, como se a leitura fosse no sentido $5' - 3'$.
- A **6ª linha** é constituída no sentido oposto à **2ª linha** ($3' - 5'$), ou seja, da direita para a esquerda, respeitando sempre a complementaridade biológica: $A \leftrightarrow U$ e $C \leftrightarrow G$.

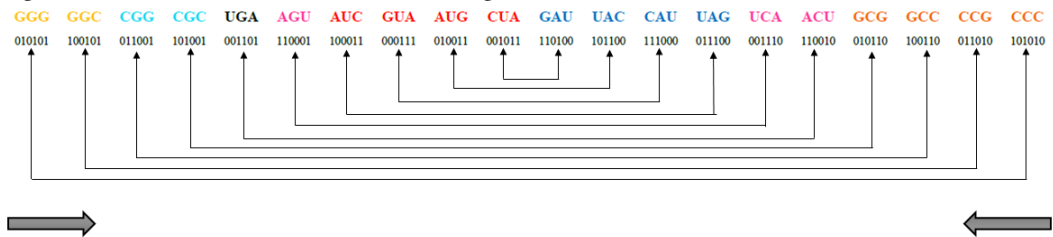
Como exemplo, considere na Figura 20, na segunda linha seguindo da esquerda para a direita, o primeiro códon é o **UGU**. Logo, seu complementar biológico na 6ª linha, da direita para a esquerda, é o códon **ACA**. Ou seja, $UGU \longleftrightarrow ACA$.

- A **3ª linha** e **5ª linha** são constituídas por quinze códons cada uma, e segue o mesmo procedimento dos passos c) e d).

- f) A **4ª linha** é constituída por vinte códons e a escrita dos códons é realizada respeitando a complementaridade biológica seguindo das laterais para o centro.

Considerando a proposta de Sánchez, Morgado e Grau (2004), a Figura 19 representa como é realizado o preenchimento da 4ª linha do diagrama de Hasse da Figura 20.

Figura 19 – Preenchimento da 4ª linha do diagrama de Hasse.



Fonte: Da autora.

Observe que o número de códons existente em cada linha, se dá devido a sua associação algébrica, ou seja, está relacionado com o número de zeros (0) ou uns (1) que cada códon possui. Por exemplo, para Figura 20, temos que:

- a) Na 1ª linha, o códon possui seis uns (1) e ele é o único códon que possui essa quantidade de uns (1).

$$UUU \leftrightarrow 111111$$

- b) Na 2ª linha, os códons são representados por um elemento zero (0) e cinco elementos uns (1). Com isso, tem-se seis códons, apresentados a seguir:

$$\begin{array}{lll} UGU \leftrightarrow 111101 & GUU \leftrightarrow 110111 & UUG \leftrightarrow 011111 \\ UUC \leftrightarrow 101111 & CUU \leftrightarrow 111011 & UCU \leftrightarrow 111110 \end{array}$$

- c) Na 3ª linha, os códons são representados por dois elementos zeros (0) e quatro elementos uns (1). Temos então, que esta linha possui quinze códons, sendo eles:

$$\begin{array}{lll} GGU \leftrightarrow 110101 & CGU \leftrightarrow 111001 & UGG \leftrightarrow 011101 \\ UGC \leftrightarrow 101101 & GUG \leftrightarrow 010111 & GUC \leftrightarrow 100111 \\ AUU \leftrightarrow 110011 & CUC \leftrightarrow 101011 & UUA \leftrightarrow 001111 \\ CUG \leftrightarrow 011011 & UAU \leftrightarrow 111100 & UCG \leftrightarrow 011110 \\ UCC \leftrightarrow 101110 & GCU \leftrightarrow 110110 & CCU \leftrightarrow 111010 \end{array}$$

- d) Na 4ª linha, os códons são representados por três elementos zeros (0) e três elementos uns (1). Com esses elementos, obtemos vinte códons, sendo eles:

GGG ↔ 010101	GGC ↔ 100101	CGG ↔ 011001	CGC ↔ 101001
UGA ↔ 001101	AGU ↔ 110001	AUC ↔ 100011	GUA ↔ 000111
AUG ↔ 010011	CUA ↔ 001011	GAU ↔ 110100	UAC ↔ 101100
CAU ↔ 111000	UAG ↔ 011100	UCA ↔ 001110	ACU ↔ 110010
GCG ↔ 010110	GCC ↔ 100110	CCG ↔ 011010	CCC ↔ 101010

- e) Na 5ª linha, os códons são representados por quatro elementos zeros (0) e dois elementos uns (1). Nesta linha, obtemos quinze códons, sendo eles:

GGA ↔ 000101	CGA ↔ 001001	AGG ↔ 010001
AGC ↔ 100001	AUA ↔ 000011	GAC ↔ 100100
AAU ↔ 110000	GAG ↔ 010100	UAA ↔ 001100
CAG ↔ 011000	CAC ↔ 101000	ACG ↔ 010010
ACC ↔ 100010	GCA ↔ 000110	CCA ↔ 001010

- f) Na 6ª linha os códons são representados por cinco elementos zeros (0) e um elemento um (1). Logo, é possível obter seis códons nesta linha, sendo eles:

AGA ↔ 000001	GAA ↔ 000100	AAG ↔ 010000
AAC ↔ 100000	CAA ↔ 001000	ACA ↔ 000010

- g) Na 7ª linha, o códon possui seis elementos zeros (0) e ele é o único códon que possui essa quantidade de zero.

AAA ↔ 000000

A construção do diagrama de Hasse, proposto por Sánchez et al. (2004), é feita por um determinado número de arestas. Cada aresta determina a distância de Hamming entre códons. Mas sua construção não é feita de forma arbitrária, ela segue o seguinte procedimento:

- a) Analisando o diagrama do elemento mínimo ao máximo, temos que, o códon mínimo percorre 6 arestas até chegar ao códon máximo. Observe os cálculos, obtidos da 7ª linha até a 1ª linha, com base na proposta de Sánchez et al. (2004).

- Cálculo da distância de Hamming dos elementos da 7ª linha em relação aos da 6ª linha.

$$\begin{aligned}
 d_H(AAA, AGA) &= d_H(000000, 000001) = 1 & d_H(AAA, GAA) &= d_H(000000, 000100) = 1 \\
 d_H(AAA, AAG) &= d_H(000000, 010000) = 1 & d_H(AAA, AAC) &= d_H(000000, 100000) = 1 \\
 d_H(AAA, CAA) &= d_H(000000, 001000) = 1 & d_H(AAA, ACA) &= d_H(000000, 000010) = 1
 \end{aligned}$$

- Cálculo da distância de Hamming dos elementos da 7ª linha em relação aos da 5ª linha.

$$\begin{aligned}
d_H(AAA, GGA) &= d_H(000000, 000101) = 2 & d_H(AAA, AGG) &= d_H(000000, 010001) = 2 \\
d_H(AAA, AGC) &= d_H(000000, 100001) = 2 & d_H(AAA, CGA) &= d_H(000000, 001001) = 2 \\
d_H(AAA, GAG) &= d_H(000000, 010100) = 2 & d_H(AAA, GAC) &= d_H(000000, 100100) = 2 \\
d_H(AAA, AAU) &= d_H(000000, 110000) = 2 & d_H(AAA, AUA) &= d_H(000000, 000011) = 2 \\
d_H(AAA, UAA) &= d_H(000000, 001100) = 2 & d_H(AAA, CAG) &= d_H(000000, 011000) = 2 \\
d_H(AAA, CAC) &= d_H(000000, 101000) = 2 & d_H(AAA, GCA) &= d_H(000000, 000110) = 2 \\
d_H(AAA, ACG) &= d_H(000000, 010010) = 2 & d_H(AAA, ACC) &= d_H(000000, 100010) = 2 \\
d_H(AAA, CCA) &= d_H(000000, 001010) = 2 & d_H(AAA, AGG) &= d_H(000000, 010001) = 2
\end{aligned}$$

- Cálculo da distância de Hamming dos elementos da 7ª linha em relação aos da 4ª linha.

$$\begin{aligned}
d_H(AAA, GGG) &= d_H(000000, 010101) = 3 & d_H(AAA, GGC) &= d_H(000000, 100101) = 3 \\
d_H(AAA, CGG) &= d_H(000000, 011001) = 3 & d_H(AAA, AGU) &= d_H(000000, 110001) = 3 \\
d_H(AAA, CGC) &= d_H(000000, 101001) = 3 & d_H(AAA, UGA) &= d_H(000000, 001101) = 3 \\
d_H(AAA, UAG) &= d_H(000000, 011100) = 3 & d_H(AAA, GUA) &= d_H(000000, 000111) = 3 \\
d_H(AAA, AUG) &= d_H(000000, 010011) = 3 & d_H(AAA, GAU) &= d_H(000000, 110100) = 3 \\
d_H(AAA, CUA) &= d_H(000000, 001011) = 3 & d_H(AAA, UAC) &= d_H(000000, 101100) = 3 \\
d_H(AAA, CAU) &= d_H(000000, 111000) = 3 & d_H(AAA, AUC) &= d_H(000000, 100011) = 3 \\
d_H(AAA, ACU) &= d_H(000000, 110010) = 3 & d_H(AAA, GCG) &= d_H(000000, 010110) = 3 \\
d_H(AAA, UCA) &= d_H(000000, 001110) = 3 & d_H(AAA, GCC) &= d_H(000000, 100110) = 3 \\
d_H(AAA, CCG) &= d_H(000000, 011010) = 3 & d_H(AAA, CCC) &= d_H(000000, 101010) = 3
\end{aligned}$$

- Cálculo da distância de Hamming dos elementos da 7ª linha em relação aos da 3ª linha.

$$\begin{aligned}
d_H(AAA, GGU) &= d_H(000000, 110101) = 4 & d_H(AAA, UGG) &= d_H(000000, 011101) = 4 \\
d_H(AAA, UGC) &= d_H(000000, 101101) = 4 & d_H(AAA, CGU) &= d_H(000000, 111001) = 4 \\
d_H(AAA, GUG) &= d_H(000000, 010111) = 4 & d_H(AAA, GUC) &= d_H(000000, 100111) = 4 \\
d_H(AAA, AUU) &= d_H(000000, 110011) = 4 & d_H(AAA, UAU) &= d_H(000000, 111100) = 4 \\
d_H(AAA, UUA) &= d_H(000000, 001111) = 4 & d_H(AAA, CUG) &= d_H(000000, 011011) = 4 \\
d_H(AAA, CUC) &= d_H(000000, 101011) = 4 & d_H(AAA, GCU) &= d_H(000000, 110110) = 4 \\
d_H(AAA, UCG) &= d_H(000000, 011110) = 4 & d_H(AAA, UCC) &= d_H(000000, 101110) = 4 \\
d_H(AAA, CCU) &= d_H(000000, 111010) = 4
\end{aligned}$$

- Cálculo da distância de Hamming dos elementos da 7ª linha em relação aos da 2ª linha.

$$\begin{aligned}
d_H(AAA, UGU) &= d_H(000000, 111101) = 5 & d_H(AAA, GUU) &= d_H(000000, 110111) = 5 \\
d_H(AAA, UUG) &= d_H(000000, 011111) = 5 & d_H(AAA, UUC) &= d_H(000000, 101111) = 5 \\
d_H(AAA, CUU) &= d_H(000000, 111011) = 5 & d_H(AAA, UCU) &= d_H(000000, 111110) = 5
\end{aligned}$$

- Cálculo da distância de Hamming dos elementos da 7ª linha em relação aos da 1ª linha.

$$d_H(AAA, UUU) = d_H(000000, 111111) = 6$$

Observação 3.1 *Se os cálculos forem efetuados da 1ª linha à 7ª linha, iremos obter os mesmos resultados.*

b) Da 7ª linha para 6ª linha, cada códon é ligado por uma única aresta. Conforme os cálculos anteriores, a distância do códon AAA até o seu correspondente na 6ª linha, é igual a 1.

c) Da 6ª linha para a 5ª linha, cada códon é ligado por cinco arestas. Essas arestas só são ligadas se a distância de um códon para outro for igual a 1. Os cálculos da 6ª linha à 5ª linha, podem ser vistos a seguir:

- Do códon AGA partem 5 arestas

$$d_H(AGA, GGA) = d_H(000001, 000101) = 1$$

$$d_H(AGA, AGG) = d_H(000001, 010001) = 1$$

$$d_H(AGA, AGC) = d_H(000001, 100001) = 1$$

$$d_H(AGA, CGA) = d_H(000001, 001001) = 1$$

$$d_H(AGA, AUA) = d_H(000001, 000011) = 1$$

- Do códon AAC partem 5 arestas.

$$d_H(AAC, AGC) = d_H(100000, 100001) = 1$$

$$d_H(AAC, GAC) = d_H(100000, 100100) = 1$$

$$d_H(AAC, AAU) = d_H(100000, 110000) = 1$$

$$d_H(AAC, CAC) = d_H(100000, 101000) = 1$$

$$d_H(AAC, ACC) = d_H(100000, 100010) = 1$$

- Do códon GAA partem 5 arestas

$$d_H(GAA, GGA) = d_H(000100, 000101) = 1$$

$$d_H(GAA, GAG) = d_H(000100, 010100) = 1$$

$$d_H(GAA, GAC) = d_H(000100, 100100) = 1$$

$$d_H(GAA, UAA) = d_H(000100, 001100) = 1$$

$$d_H(GAA, GCA) = d_H(000100, 000110) = 1$$

- Do códon CAA partem 5 arestas.

$$d_H(CAA, CGA) = d_H(001000, 001001) = 1$$

$$d_H(CAA, UAA) = d_H(001000, 001100) = 1$$

$$d_H(CAA, CAG) = d_H(001000, 011000) = 1$$

$$d_H(CAA, CAC) = d_H(001000, 101000) = 1$$

$$d_H(CAA, CCA) = d_H(001000, 001010) = 1$$

- Do códon AAG partem 5 arestas.

$$d_H(AAG, AGG) = d_H(010000, 010001) = 1$$

$$d_H(AAG, GAG) = d_H(010000, 010100) = 1$$

$$d_H(AAG, AAU) = d_H(010000, 110000) = 1$$

$$d_H(AAG, CAG) = d_H(010000, 011000) = 1$$

$$d_H(AAG, ACG) = d_H(010000, 010010) = 1$$

- Do códon ACA partem 5 arestas.

$$d_H(ACA, AUA) = d_H(000010, 000011) = 1$$

$$d_H(ACA, GCA) = d_H(000010, 000110) = 1$$

$$d_H(ACA, ACG) = d_H(000010, 010010) = 1$$

$$d_H(ACA, ACC) = d_H(000010, 100010) = 1$$

$$d_H(ACA, CCA) = d_H(000010, 001010) = 1$$

d) Da 5ª linha para a 4ª linha, cada códon é ligado por quatro arestas. Essas arestas só são ligadas se a distância de um códon para outro for igual a 1. Os cálculos são obtidos de maneira análoga ao passo anterior.

e) Da 4ª linha para 3ª linha, cada códon é ligado por três arestas. Essas arestas só são ligadas se a distância de um códon para outro for igual a 1, cálculos análogos aos passos anteriores.

f) Da 3ª linha para a 2ª linha, cada códon é ligado por duas arestas. Como no passo anterior,

essas arestas só são ligadas se a distância de um códon para outro for igual a 1.

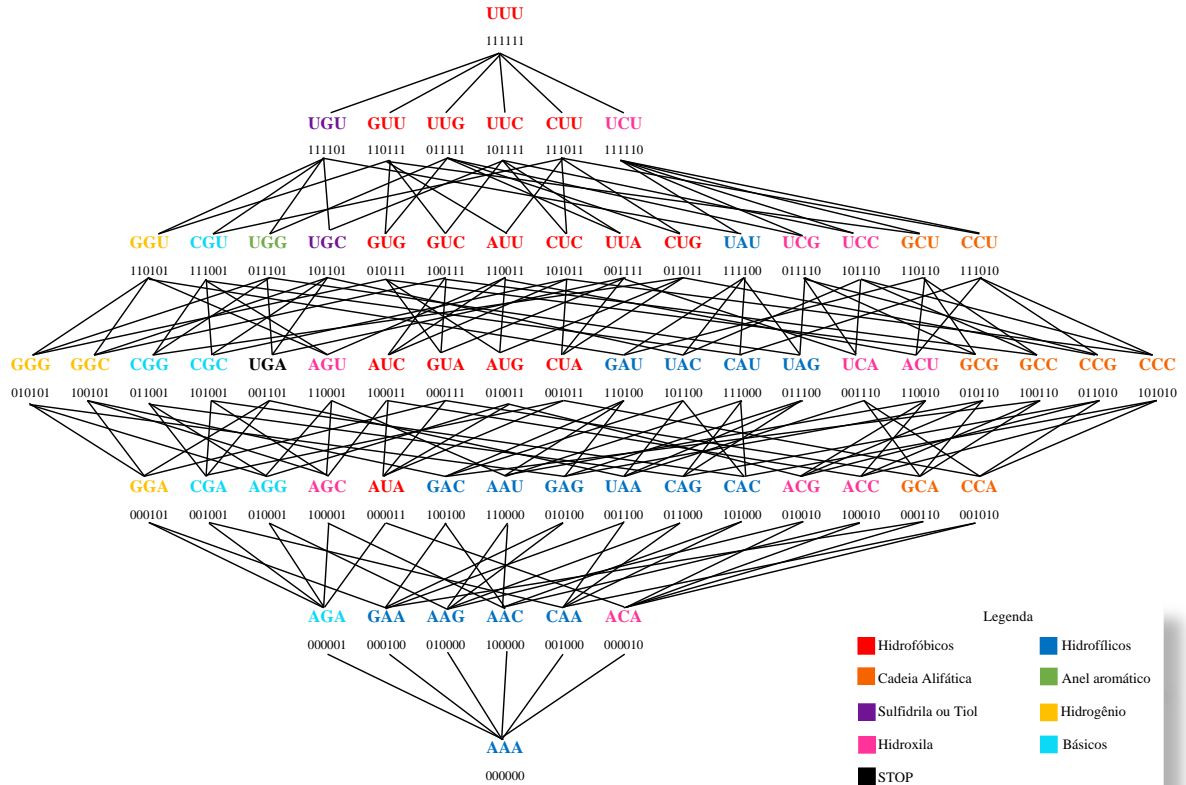
- g) Da 2ª linha para a 1ª linha, cada códon é ligado por uma única aresta. Logo, cada códon da 2ª linha possui a distância até o códon GGG igual a 1.

Observação 3.2 *Vale destacar que, o processo descrito, foi realizado de baixo para cima (do elemento mínimo para o máximo). Logo, se efetuarmos os cálculos de cima para baixo, o processo é análogo. Então, temos que da 1ª linha para a 2ª linha cada códon possui seis arestas. Da 2ª linha para a 3ª linha, cada códon possui cinco arestas. Da 3ª linha para a 4ª linha, cada códon possui quatro arestas. Da 4ª linha para a 5ª linha, cada códon possui três arestas. Da 5ª linha para a 6ª linha, cada códon possui duas arestas e por fim, da 6ª linha para a 7ª linha, cada códon possui uma única aresta.*

A Figura 20 apresenta o diagrama de Hasse primal para a permutação 0132 do rotulamento A , escolhida de forma aleatória. O padrão mencionado anteriormente pode ser visto através da análise da legenda, a qual apresenta as características dos aminoácidos, dispostos de forma organizada no diagrama de Hasse. A Figura 21 apresenta o diagrama de Hasse dual para a permutação 2310 do rotulamento A . Sua construção é análoga à construção do diagrama de Hasse primal.

3.1.3 Análise do diagrama primal

Figura 20 – Diagrama de Hasse primal referente à permutação 0132 do Rotulamento A.



Fonte: Da autora.

A correspondência entre a ordem do códon e as propriedades dos aminoácidos é refletida no diagrama de Hasse do código genético. Na Figura 20 observa-se uma simetria na distribuição dos códons. A construção do diagrama de Hasse para a permutação 0132 apoiou-se na complementaridade das bases nitrogenadas que assim associou-se ao mapeamento algébrico.

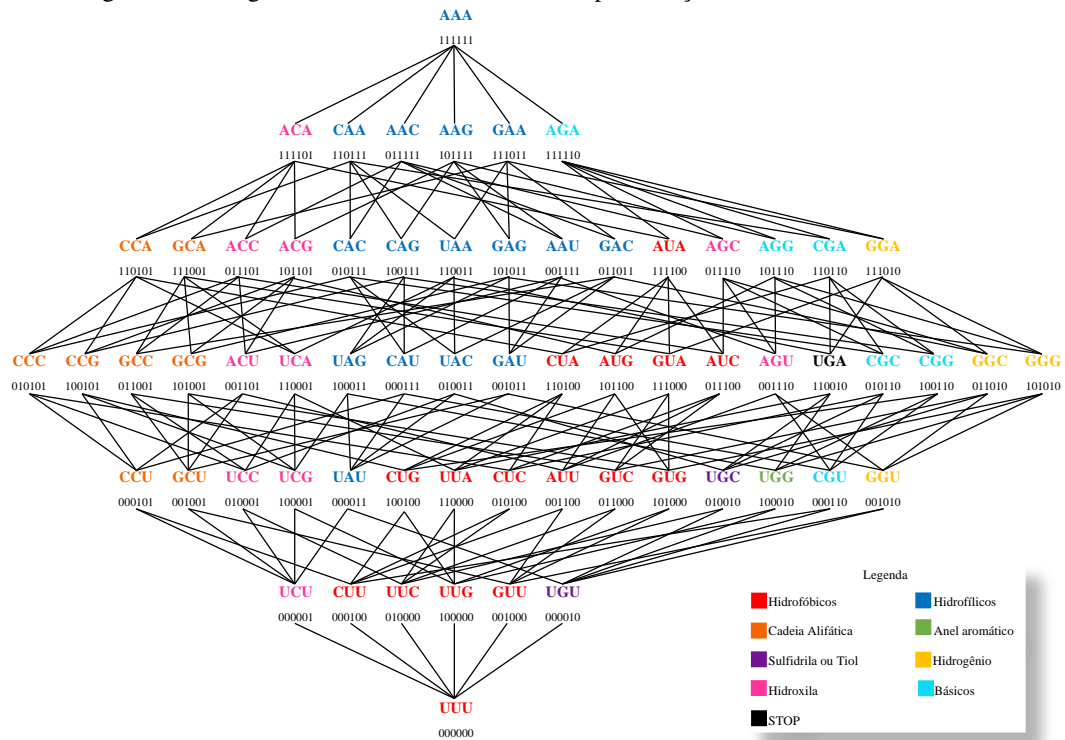
De acordo com a legenda, pode-se observar que os códons hidrofóbicos, aqueles que não possuem “afinidade” com a água (códons com *U* na segunda posição estão representados em vermelho) estão separados dos códons hidrofílicos, aqueles que possuem “afinidade” com a água (códons com *A* na segunda posição estão representados em azul). Estes mesmos códons estão distribuídos no centro da Figura 20. O códon que codifica um aminoácido hidrofóbico é sempre complementar a um códon que codifica um aminoácido hidrofílico, ou seja, observe que a cadeia (UGU, GUU, UUG, UUC, CUU, UCU), lida no sentido 5' – 3' é a imagem da anti-cadeia (AGA, GAA, AAG, AAC, CAA, ACA), lida no sentido 3' – 5'.

Nas laterais no diagrama, podemos encontrar os aminoácidos hidrogênio, básicos, hidroxila e cadeia alifática. Além disso, na lateral esquerda, podemos encontrar os aminoácidos sulfidrina ou tiol, anel aromático e o códon STOP, sendo este utilizado para interromper a proteína antes de seu término.

O códon UGG é um anel aromático que codifica o aminoácido triptofano, responsável pela sensação de bem estar, ou seja, ele atua como um antidepressivo (cor verde na lateral esquerda). As trincas GGU, GGC, GGA e GGG, classificados como hidrogênio, são os códons que codificam o aminoácido glicina, considerado o mais simples e está presente na maioria das proteínas. No diagrama, eles podem ser encontrados na lateral esquerda e estão próximos um do outro. Os códons CCA, CCC, CCG e CCU são classificados como cadeia alifática e codifica o aminoácido prolina, sendo este o mais rígido dos vinte aminoácidos, ou seja, possui uma estrutura quimicamente coesa e rígida. No diagrama eles estão próximos e estão presentes na lateral direita.

3.1.4 Análise do diagrama dual

Figura 21 – Diagrama de Hasse dual referente à permutação 2310 do Rotulamento A.



Fonte: Da autora.

Na Figura 21 observa-se uma simetria na distribuição dos códons. A construção do diagrama de Hasse para a permutação 2310 apoiou-se na complementaridade das bases nitrogenadas que assim associou-se ao mapeamento algébrico.

De acordo com a legenda, pode-se observar que os códons hidrofílicos no diagrama dual (códons com *A* na segunda posição estão representados em azul e possuem “afinidade” com a água), além de separados, ficaram na parte superior no centro do diagrama, com relação dos códons hidrofóbicos (códons com *U* na segunda posição estão representados em vermelho e não possuem “afinidade” com a água) que estão situados na parte inferior do diagrama da Figura 21. O códon que codifica um aminoácido hidrofílico é sempre complementar a um códon que codifica um aminoácido hidrofóbico, ou seja, observe que a cadeia (ACA, CAA, AAC, AAG, GAA, AGA), lida no sentido 5' – 3' é a imagem da anti-cadeia (UCU, CUU, UUC, UUG, GUU, UGU), lida no sentido 3' – 5'.

Na laterais no diagrama dual, assim como o diagrama primal, podemos encontrar os aminoácidos hidrogênio, básicos, hidroxila e cadeia alifática. Na lateral direita, podemos encontrar os aminoácidos sulfidril ou tiol, anel aromático e o códon STOP, sendo este utilizado para interromper a proteína antes de seu término.

O códon UGG é um anel aromático que codifica o aminoácido triptofano, nesse diagrama (Figura 21) está na cor verde e encontra-se na lateral direita. As trincas GGU, GGC, GGA e GGG, classificados como hidrogênio, são os códons que codificam o aminoácido glicina. Note que eles estão próximos um do outro e estão na lateral direita. Os códons CCA, CCC, CCG e CCU são classificados como cadeia alifática e codifica o aminoácido prolina. No diagrama eles estão próximos e estão presentes na lateral esquerda.

Desta forma, percebe-se uma interessante caracterização algébrica associada ao mapeamento do código genético, permitindo uma análise biológica acerca da construção do diagrama de Hasse para as permutações 0132 e 2310 pertencentes ao Rotulamento *A*, de acordo com as características dos aminoácidos.

3.1.5 Análise das diferenças e semelhanças físico-químicas dos aminoácidos por meio das distâncias de Hamming

Conforme a proposta de Sánchez, Morgado e Grau (2004), serão apresentadas a seguir as médias das distâncias de Hamming entre os pares de aminoácidos, como as distâncias médias entre os seus respectivos códon (Tabela 19). Pode-se notar que os valores das médias das distâncias de Hamming maiores ou iguais a 4 aparecem em negrito.

Tabela 19 – Médias das distâncias de Hamming entre pares de aminoácidos do Rotulamento A.

*	G	W	C	R	S	V	L	F	M	I	E	D	Y	K	N	Q	H	A	T	P	STOP
G	1	2	2	2,67	3,33	2	3,67	3	3	3	2	2	3	3	3	4	4	3	4	5	2,67
W	2	0	1,5	2,17	3,17	3	2,5	2,5	3	4,33	2,5	3,5	2,5	3,5	4,5	2,5	3,5	4	5	4	1,33
C	2	1,5	0,5	2,5	2,83	3	2,83	1,5	4,5	3,83	3,5	2,5	1,5	4,5	3,5	3,5	2,5	4	5	4	2,17
R	2,67	2,17	2,5	1,39	3,5	3,67	2,61	3,5	2,5	2,72	3,5	3,83	3,5	2,5	2,83	2,17	2,5	4,67	3,67	3,33	2,83
S	3,33	3,17	2,83	3,5	2,72	3	2,94	2,5	3,5	3,28	3,17	2,83	2,5	3,5	3,17	3,17	2,83	2,67	3	2,67	2,94
V	2	3	3	3,67	3	1	2,67	2	2	2	3	3	4	4	4	5	5	2	3	4	3,67
L	3,67	2,5	2,83	2,61	2,94	2,67	1,39	1,83	2,17	2,39	4,5	4,83	3,83	4,17	4,5	3,17	3,5	3,67	3,33	2,33	3,17
F	3	2,5	1,5	3,5	2,5	2	1,83	0,5	3,5	2,83	4,5	3,5	2,5	5,5	4,5	4,5	3,5	3	4	3	3,17
M	3	3	4,5	2,5	3,5	2	2,17	3,5	0	1,33	3,5	4,5	5,5	2,5	3,5	3,5	4,5	3	2	3	4,33
I	3	4,33	3,83	2,72	3,28	2	2,39	2,83	1,33	0,89	4,17	3,83	4,83	3,17	2,83	4,17	3,83	3	2	3	4,78
E	2	2,5	3,5	3,5	3,17	3	4,5	4,5	3,5	4,17	0,5	1,5	2,5	1,5	2,5	2,5	3,5	2	3	4	1,83
D	2	3,5	2,5	3,83	2,83	3	4,83	3,5	4,5	3,83	1,5	0,5	1,5	2,5	1,5	3,5	2,5	2	3	4	2,83
Y	3	2,5	1,5	3,5	2,5	4	3,83	2,5	5,5	4,83	2,5	1,5	0,5	3,5	2,5	2,5	1,5	3	4	3	1,83
K	3	3,5	4,5	2,5	3,5	4	4,17	5,5	2,5	3,17	1,5	2,5	3,5	0,5	1,5	1,5	2,5	3	2	3	2,83
N	3	4,5	3,5	2,83	3,17	4	4,5	4,5	3,5	2,83	2,5	1,5	2,5	1,5	0,5	2,5	1,5	3	2	3	3,83
Q	4	2,5	3,5	2,17	3,17	5	3,17	4,5	3,5	4,17	2,5	3,5	2,5	1,5	2,5	0,5	1,5	4	3	2	1,83
H	4	3,5	2,5	2,5	2,83	5	3,5	3,5	4,5	3,83	3,5	2,5	1,5	2,5	1,5	1,5	0,5	4	3	2	2,83
A	3	4	4	4,67	2,67	2	3,67	3	3	3	2	2	3	3	3	4	4	1	2	3	3,33
T	4	5	5	3,67	3	3	3,33	4	2	2	3	3	4	2	2	3	3	2	1	2	4,33
P	5	4	4	3,33	2,67	4	2,33	3	3	3	4	4	3	3	2	2	2	3	2	1	3,33
STOP	2,67	1,33	2,17	2,83	2,94	3,67	3,17	3,17	4,33	4,78	1,83	2,83	1,83	2,83	3,83	1,83	2,83	3,33	4,33	3,33	0,89

Fonte: Da autora.

Nota: *Representação dos aminoácidos por um único símbolo. A - Alanina, C - Cisteína, D - Ác. Aspártico, E - Ác. Glutâmico, F - Fenilalanina, G - Glicina, H - Histidina, I - Isoleucina, K - Lisina, L - Leucina, M - Metionina, N - Asparagina, P - Prolina, Q - Glutamina, R - Arginina, S - Serina, T - Treonina, V - Valina, W - Triptofano, Y - Tirosina.

As distâncias médias entre o códon que codifica o aminoácido glicina (G) e os códon que codificam os aminoácidos glutamina (Q), histidina (H), treonina (T) e prolina (P) foram maiores ou iguais a 4. O códon associado ao triptofano e os códon associados a isoleucina (I), asparagina (N), alanina (A), treonina (T) e prolina (P) foram maiores ou iguais a 4. O códon associado a cisteína (C) e os códon associados a metionina (M), lisina (K), alanina (A), treonina (T) e prolina (P) também foram maiores ou iguais a 4. Os outros casos podem ser observados na Tabela 19. Esses aminoácidos apresentam grandes diferenças físico-químicas entre si.

Com base nas características de cada aminoácido codificado pelos seus respectivos códon, a seguir serão apresentados alguns casos de análise de algumas diferenças e semelhanças biológicas acerca dos cálculos obtidos na Tabela 19 e associá-los ao diagrama de Hasse primal (Figura 20) e ao diagrama de Hasse dual (Figura 21).

O códon UGG codifica o aminoácido triptofano (W) e os códons AUU, AUC e AUA codificam o aminoácido isoleucina (I). Na Tabela 19, a distância média entre o códon UGG e os códons AUU, AUC e AUA é 4,33. O triptofano (W) é um anel aromático e a isoleucina (I) é um aminoácido hidrofóbico. Os códons AAU e AAC codificam o aminoácido asparagina (N), que é hidrofílico. Na Tabela 19 a distância média entre esses códons e o códon que codifica o aminoácido triptofano (W) é igual a 4,5. Os códons GCU, GCC, GCA, GCG codificam o aminoácido alanina (A), e a sua distância média com o códon que codifica o aminoácido triptofano (W) é igual a 4. A alanina (A) possui cadeia alifática, sendo este diferente do aminoácido triptofano (W). Nas Figuras 20 e 21, sua distância pode ser observada nos diagramas de Hasse, nos quais, além dos códons estarem separados, há uma distância entre os códons. Os códons ACU, ACC, ACA e ACG codificam o aminoácido treonina (T), sendo uma hidroxila. A sua distância média com o códon do aminoácido triptofano (W) é igual a 5. Os códons CCU, CCC, CCA e CCG codificam o aminoácido prolina (P), que possui cadeia alifática e a distância média com o códon do aminoácido triptofano (W), baseado na Tabela 19 é igual a 4. Os códons UGU e UGC codificam o aminoácido cisteína (C), que tem como característica sulfidril ou tiol. A distância média com o códon que codifica o aminoácido triptofano (W) é igual a 1,5 (Tabela 19). Note que, nos diagramas de Hasse, eles estão próximos um do outro, podendo assim, ter características comuns entre eles.

Os códons GGU, GGC, GGA e GGG codificam o aminoácido glicina (G), que possui como característica de hidrogênio e os códons CAA e CAG codificam o aminoácido glutamina (Q) que é hidrofílico. Pode-se notar que a média da distância entre os códons desses aminoácidos na Tabela 19 é igual a 4. Os códons CAU e CAC codificam o aminoácido histidina (H), que possui a mesma característica do aminoácido glutamina (Q), sendo eles hidrofílicos. A distância média dos códons do aminoácido histidina (H) entre os códons do aminoácido glicina (G) é igual a 4 (Tabela 19). Os códons ACU, ACC, ACA e ACG codificam o aminoácido treonina (T) que é uma hidroxila. Devido as suas diferenças de características com o aminoácido glicina (G), pode-se notar que a distância média entre os códons desses aminoácidos na Tabela 19 é igual a 4. Os códons CCU, CCC, CCA e CCG codificam o aminoácido prolina (P), que possui cadeia alifática e a distância média com os códons do aminoácido glicina (G), baseado na Tabela 19 é igual a 5. Os códons GUU, GUC, GUA e GUG codificam o aminoácido valina (V), que são classificados como hidrofóbicos. A distância média com os códons que codificam o aminoácido glicina (G), na Tabela 19 é igual a 2. Note que nos diagramas de Hasse, os códons estão

próximos, o que justifica o valor da sua média.

Os códons CGU, CGC, CGA, CGG, AGA e AGG codificam o aminoácido arginina (R), que são classificados como básicos e os códons GCU, GCC, GCA e GCG que codificam o aminoácido alanina (A), tem cadeia alifática. Pode-se notar que a distância média entre os códons desses aminoácidos na Tabela 19 é igual a 4,67.

O códon UGG codifica o aminoácido triptofano (W) e os códons UGA, UAA e UAG codificam o aminoácido STOP, sendo este utilizado para interromper a proteína antes de seu término. Na Tabela 19, a distância média entre esses aminoácidos é 1,33. Podemos notar no diagrama que os códons UAA e UAG possuem características hidrofílicas, sendo diferente do aminoácido triptofano, que é um anel aromático. Porém o códon UGA, está localizado perto do códon UGG, o que justifica o fato da média entre os códons ter sido 1,33. Um outro caso é o códon AUG que codifica o aminoácido metionina (M) e os códons AUU, AUC e AUA que codificam o aminoácido isoleucina (I). Ambos os aminoácidos são semelhantes, pois possuem características hidrofílicas. Nos diagramas (Figuras 20 e 21) podemos observar que os códons estão próximos um do outro, justificando o fato de que na Tabela 19, a distância média entre os aminoácidos metionina (M) e isoleucina (I) é 1,33.

Pode-se observar os outros casos na Tabela 19 com os diagramas (Figuras 20 e 21) construídos. Em todos os casos, as distâncias que foram maiores ou iguais a 4 os aminoácidos apresentam diferenças significativas entre si, e isso se reflete na distância entre eles nos diagramas de Hasse primal e dual, da mesma forma que as distâncias com valores menores refletem as semelhanças dos aminoácidos e refletem na distância nos diagramas construídos.

3.2 ROTULAMENTO B

Para o estudo do Rotulamento B , será utilizada a permutação 1032, para o reticulado primal, nesse caso obtemos a associação apresentada na Tabela 20. Como consequência, a permutação 3210 do Rotulamento B será utilizada para a construção do reticulado dual. Na Tabela 21 será apresentada a associação estabelecida. Neste caso, não foi utilizada a complementaridade biológica das bases, mas sim, a complementaridade algébrica, na qual 00 - 11 e 01 - 10. Desta forma, as bases nitrogenadas referentes a essa complementaridade algébrica são: $A - G$ e $C - U$, ou vice-versa.

Tabela 20 – Associação estabelecida pela permutação 1032 primal do Rotulamento *B*.

A	C	G	U
1	0	3	2
10	00	01	11

Fonte: Da autora.

Tabela 21 – Associação estabelecida pela permutação 3210 dual do Rotulamento *B*.

A	C	G	U
3	2	1	0
01	11	10	00

Fonte: Da autora.

A seguir será apresentado o procedimento para a construção dos reticulados booleanos primal e dual associados às permutações selecionadas para o Rotulamento *B*.

3.2.1 Procedimento para a construção dos reticulados booleanos

Utilizando as operações da álgebra booleana (Tabelas 11 e 12) e a permutação 1032 do Rotulamento *B* (primal), obtemos os dados das Tabelas 22 e 23.

Tabela 22 – Operação primal \wedge (e) do Rotulamento *B*.

\wedge	00	01	10	11
00	00	00	00	00
01	00	01	00	01
10	00	00	10	10
11	00	01	10	11

Fonte: Da autora.

\wedge	C	G	A	U
C	C	C	C	C
G	C	G	C	G
A	C	C	A	A
U	C	G	A	U

Fonte: Da autora.

Tabela 23 – Operação primal \vee (ou) do Rotulamento *B*.

\vee	00	01	10	11
00	00	01	10	11
01	01	01	11	11
10	10	11	10	11
11	11	11	11	11

Fonte: Da autora.

\vee	C	G	A	U
C	C	G	A	U
G	G	G	U	U
A	A	U	A	U
U	U	U	U	U

Fonte: Da autora.

Utilizando as operações da álgebra booleana (Tabelas 11 e 12) e a permutação 3210 do Rotulamento *B* (dual), obtemos os dados das Tabelas 24 e 25.

Tabela 24 – Operação dual \wedge (e) do Rotulamento B .

\wedge	00	01	10	11
00	00	00	00	00
01	00	01	00	01
10	00	00	10	10
11	00	01	10	11

Fonte: Da autora.

\wedge	U	A	G	C
U	U	U	U	U
A	U	A	U	A
G	U	U	G	G
C	U	A	G	C

Fonte: Da autora.

Tabela 25 – Operação dual \vee (ou) do Rotulamento B .

\vee	00	01	10	11
00	00	01	10	11
01	01	01	11	11
10	10	11	10	11
11	11	11	11	11

Fonte: Da autora.

\vee	U	A	G	C
U	U	A	G	C
A	A	A	C	C
G	G	C	G	C
C	C	C	C	C

Fonte: Da autora.

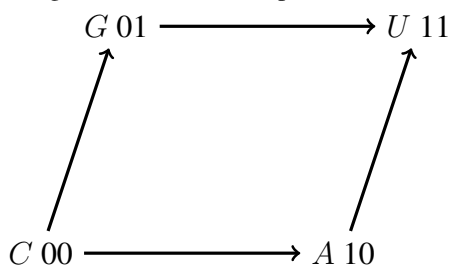
O reticulado booleano primal será obtido a partir da Tabela 23. Observe que:

- A base C (citosina) se liga às bases G (guanina), A (adenina) e U (uracila).
- A base G (guanina) se liga a ela mesma e à base U (uracila).
- A base A (adenina) se liga a ela mesma e à base U (uracila).
- A base U (uracila) se liga apenas a ela mesma.

A construção é realizada seguindo o caminho que as bases percorrem. O reticulado booleano dual será obtido de maneira análoga à construção do reticulado primal, sendo ele a partir da Tabela 25. Assim, obtemos as representações das Figuras 22 e 23.

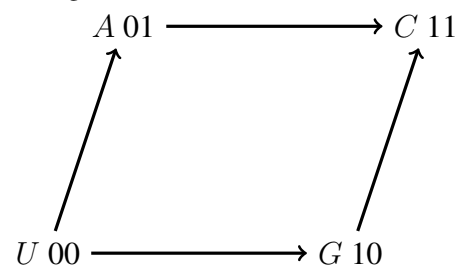
Reticulados primal e dual associados ao Rotulamento B .

Figura 22 – Reticulado primal (1032).



Fonte: Da autora.

Figura 23 – Reticulado dual (3210).



Fonte: Da autora.

3.2.2 Procedimento para a construção dos Diagramas de Hasse

Para a construção dos diagramas de Hasse associados ao Rotulamento B , seguiremos o mesmo procedimento do Rotulamento A . Ao contrário do Rotulamento A , onde foi utilizada a complementaridade biológica das bases, no Rotulamento B será considerada a complementaridade algébrica, sendo assim, $A - G$ e $C - U$. Novamente, iremos construir os diagramas de Hasse, a fim de analisar as propriedades dos aminoácidos e classificar os códon de acordo com a sua hidropacidade e outras características.

No diagrama de Hasse primal (Figura 24) na 1ª linha, está o códon que apresenta seu elemento máximo, sendo ele $UUU \leftrightarrow 111111$. No diagrama de Hasse dual (Figura 25), o elemento máximo é o $CCC \leftrightarrow 111111$.

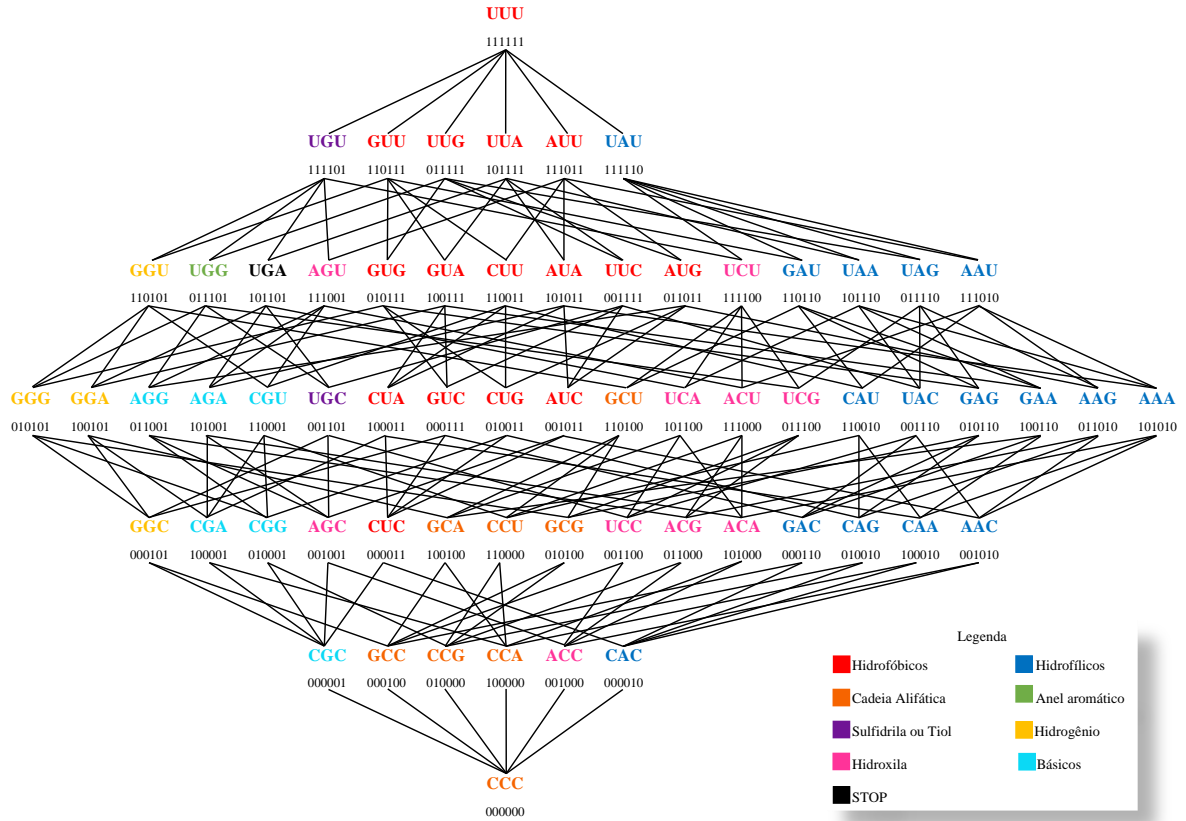
Na 7ª linha, nos diagramas de Hasse (Figuras 24 e 25), estão os códon que possuem elemento mínimo, sendo eles, na Figura 24, o códon $CCC \leftrightarrow 000000$ e na Figura 25, o códon $UUU \leftrightarrow 000000$.

As outras linhas dos diagramas de Hasse, seguem o mesmo procedimento utilizado no Rotulamento A . Porém, levando em consideração o complementar algébrico de cada códon. Nos diagramas, o número de códon existentes em cada linha, se dá devido a representação algébrica de cada códon, estando relacionada ao número de zeros (0) ou uns (1) existente em cada códon. Além disso, assim como no Rotulamento A , no Rotulamento B , determinamos as médias das distâncias de Hamming entre os códon através das arestas, que podem ser observadas nas construções (Figuras 24 e 25).

A Figura 24 apresenta o diagrama de Hasse primal para a permutação 1032 do Rotulamento B , escolhida de forma aleatória. O padrão mencionado anteriormente pode ser visto através da análise da legenda, a qual apresenta as características dos aminoácidos, dispostos de forma organizada no diagrama de Hasse. A Figura 25 apresenta o diagrama de Hasse dual para a permutação 3210 do Rotulamento B . Sua construção é análoga à construção do diagrama de Hasse primal.

3.2.3 Análise do diagrama primal

Figura 24 – Diagrama de Hasse primal referente à permutação 1032 do Rotulamento *B*.



Fonte: Da autora.

A construção do diagrama de Hasse primal para a permutação 1032 apoiou-se na complementaridade algébrica das bases nitrogenadas associadas ao mapeamento algébrico no Rotulamento *B*.

Na Figura 24, podemos perceber que a união das bases se dá por meio da complementaridade algébrica. Se tomarmos como exemplo a cadeia UGU, GUU, UUG, UUA, AUU, UAU, sua imagem será a anti-cadeia CGC, GCC, CCG, CCA, ACC, CAC. Temos que a primeira cadeia é lida no sentido $5' - 3'$ e a imagem é lida no sentido $3' - 5'$ e a complementaridade das bases é realizada da seguinte forma: adenina (*A*) com guanina (*G*) e citosina (*C*) com uracila (*U*).

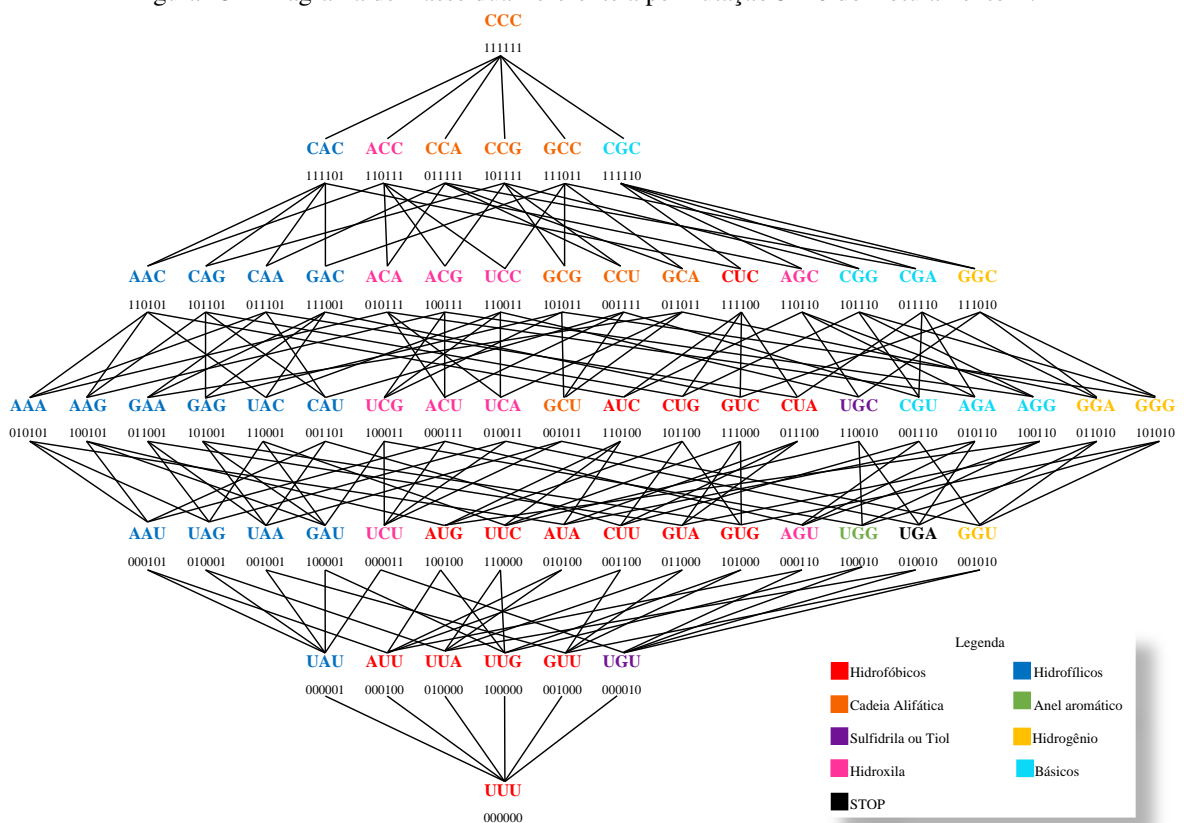
De acordo com a legenda, diferente do diagrama de Hasse do Rotulamento *A*, o códon que codifica um aminoácido hidrofóbico (códons com *U* na segunda posição, representados em vermelho) não é complementar de um códon que codifica um aminoácido hidrofílico (códons com *A* na segunda posição, representados em azul). Os códons hidrofílicos e hidrofóbicos

ficaram separados no diagrama. Os códons hidrofílicos estão presentes na cadeia lateral direita do diagrama e os códons hidrofóbicos estão presentes no centro do diagrama.

Na lateral esquerda do diagrama da Figura 24, estão dispostos os códons que codificam os aminoácidos hidrogênio, básicos, anel aromático e STOP. Além disso, o códon UGG (verde) que codifica o aminoácido triptofano ficou ao lado do códon UGA (preto) sendo este utilizado para interromper a proteína antes de seu término. No centro do diagrama, estão os códons que codificam os aminoácidos hidroxila e cadeia alifática.

3.2.4 Análise do diagrama dual

Figura 25 – Diagrama de Hasse dual referente à permutação 3210 do Rotulamento *B*.



Fonte: Da autora.

A construção do diagrama de Hasse dual para a permutação 3210 apoiou-se na complementaridade algébrica das bases nitrogenadas associadas ao mapeamento algébrico no Rotulamento *B*. Assim como no Rotulamento *A* e no diagrama de Hasse primal (Figura 24), na Figura 25, pode-se observar um padrão na distribuição dos códons.

Na Figura 25, pode-se perceber que a união das bases se dá por meio da complementaridade algébrica. Se tomarmos como exemplo a cadeia CAC, ACC, CCA, CCG, GCC, CGC, sua imagem será a anti-cadeia UAU, AUU, UUA, UUG, GUU, UGU. A primeira cadeia é lida no sentido $5' - 3'$ e a imagem é lida no sentido $3' - 5'$ e a complementaridade das bases é da seguinte forma: adenina (*A*) com guanina (*G*) e citosina (*C*) com uracila (*U*).

De acordo com a legenda, diferente do diagrama de Hasse do Rotulamento *A*, o códon que codifica um aminoácido hidrofóbico (códon com *U* na segunda posição, representados em vermelho) não é complementar a um códon que codifica um aminoácido hidrofílico (códon com *A* na segunda posição, representados em azul). Os códon hidrofílicos e hidrofóbicos ficaram separados no diagrama. Na Figura 24 os códon hidrofílicos estão presentes na cadeia lateral direita do diagrama e os códon hidrofóbicos estão presentes no centro superior do diagrama. Porém, na Figura 25, os códon hidrofílicos estão presentes na lateral esquerda e os códon hidrofóbicos estão presentes no centro inferior do diagrama.

Na lateral direita do diagrama da Figura 25, estão os códon que codificam os aminoácidos hidrogênio, básicos, anel aromático e STOP. Além disso, o códon UGG (verde) que codifica o aminoácido triptofano ficou ao lado do códon UGA (preto) sendo este utilizado para interromper a proteína antes de seu término. Pode-se notar, que nesse diagrama, ele está presente na lateral direita inferior. No centro do diagrama, estão os códon que codificam os aminoácidos hidroxila e cadeia alifática.

Comparando os diagramas de Hasse (Figuras 24 e 25) pode-se notar que o digrama de Hasse primal (Figura 24) foi rotacionado em relação ao diagrama de Hasse dual (Figura 25).

3.2.5 Análise das diferenças e semelhanças físico-químicas dos aminoácidos por meio das distâncias de Hamming

Conforme a proposta de Sánchez, Morgado e Grau (2004), serão apresentadas a seguir as médias das distâncias de Hamming entre os pares de aminoácidos, como as distâncias médias entre os seus respectivos códon (Tabela 26). Pode-se notar que os valores das médias das distâncias de Hamming maiores ou iguais a 4 aparecem em negrito, pois são considerados valores altos para essas distâncias, de acordo com Sánchez, Morgado e Grau (2004), conforme mencionado na análise realizada no Rotulamento *A*.

Tabela 26 – Médias das distâncias de Hamming entre pares de aminoácidos do Rotulamento B.

*	G	W	C	R	S	V	L	F	M	I	E	D	Y	K	N	Q	H	A	T	P	STOP
G	1	2	2	2,33	2,83	2	3	3	4	4	3	3	4	5	5	4	4	2	4	3	3,33
W	2	0	1	1	2	3	3,33	2	2	3,33	4	4	3	4	4	5	5	3	3	4	2,67
C	2	1	1	2,67	2	3	3,33	2	3	3	4	4	3	4	4	5	5	3	3	4	2,33
R	2,33	2,67	2,67	1,44	3	3,33	2,78	3,67	2,67	2,67	4,33	4,33	4,67	3,67	3,67	3,33	3,33	3,33	2,67	2,33	4
S	2,83	2	2	3	1,94	3,83	3,76	2,83	3,18	3,17	3,83	3,83	2,83	3,17	3,17	4,17	4,17	2,83	2,17	3,17	2,5
V	2	3	3	3,33	1,94	1	2	2	3	3	2	2	3	4	4	3	3	3	5	4	3
L	3	3,33	3,33	2,78	3,72	2	1,89	2,58	2	2	3	3	3,33	3	3	2,5	2,67	4	4	3,67	3,33
F	3	2	2	3,67	2,83	2	2,58	1	2	2	3	3	2	3	3	4	4	4	4	5	2
M	4	2	3	2,67	3,18	3	2	2	0	1,33	4	4	3	2	2	3	3	5	3	4	3,33
I	4	3,33	3	2,67	3,18	3	2	2	1,33	1,11	4	4	3	2	2	3	3	5	3	4	2,89
E	3	4	4	4,33	3,83	2	3	3	4	4	1	1	2	3	3	2	2	2	4	3	2,67
D	3	4	4	4,33	3,83	2	3	3	4	4	1	1	2	3	3	2	2	2	4	3	2,67
Y	4	3	3	4,67	2,83	3	3,33	2	3	3	2	2	1	2	2	3	3	3	3	4	1,67
K	5	4	4	3,67	3,17	4	3	3	2	3	3	2	2	1	1	2	2	4	2	3	2,67
N	5	4	4	3,67	3,17	4	3	3	2	2	3	3	2	1	1	2	2	4	2	3	2,67
Q	4	5	5	3,33	4,17	3	2,5	4	3	3	1	2	3	2	2	1	1	3	3	2	3,67
H	4	5	5	3,33	4,17	3	2,67	4	3	3	2	2	3	2	2	1	1	3	3	2	3,67
A	2	3	3	3,33	2,83	3	4	4	5	5	2	2	3	4	4	3	3	1	3	2	3
T	4	3	3	2,67	2,17	5	4	4	3	3	4	4	3	2	2	3	3	3	1	2	3
P	3	4	4	2,33	3,17	4	3,67	5	4	4	3	3	4	3	3	2	2	2	2	1	4
STOP	3,33	2,67	2,33	4	2,5	3	3,33	2	3,33	2,89	2,67	2,67	1,67	2,67	2,67	3,67	3,67	3	3	4	1,33

Fonte: Da autora.

Nota: *Representação dos aminoácidos por um único símbolo. A - Alanina, C - Cisteína, D - Ác. Aspártico, E - Ác. Glutâmico, F - Fenilalanina, G - Glicina, H - Histidina, I - Isoleucina, K - Lisina, L - Leucina, M - Metionina, N - Asparagina, P - Prolina, Q - Glutamina, R - Arginina, S - Serina, T - Treonina, V - Valina, W - Triptofano, Y - Tirosina.

As distâncias médias entre o códon que codificam o aminoácido glicina (G) e os códon que codificam os aminoácidos metionina (M), isoleucina (I), tirosina (Y), lisina (K), asparagina (N), glutamina (Q), histidina (H) e treonina (T) foram maiores ou iguais a 4. Entre os códon associados ao triptofano (W) e os códon associados ácido glutâmico (E), ácido aspártico (D), lisina (K), asparagina (N), glutamina (Q), histidina (H) e prolina (P) foram maiores ou iguais a 4. Entre os códon associados a cisteína (C) e os códon associados a ácido glutâmico (E), ácido aspártico (D), lisina (K), asparagina (N), glutamina (Q), histidina (H) e prolina (P) também foram maiores ou iguais a 4. Entre os códon associados a arginina (R) e os códon associados ao ácido glutâmico (E), ácido aspártico (D), tirosina (Y) e STOP foram maiores ou iguais a 4. Entre os códon associados ao aminoácido serina (S) e os códon que codificam os aminoácidos glutamina (Q) e histidina (H) foram maiores ou iguais a 4. Os outros casos podem ser observados na Tabela 26.

Com base nas características de cada aminoácido codificado pelos seus respectivos códon, serão apresentados alguns exemplos da análise de algumas diferenças e semelhanças biológicas acerca dos cálculos obtidos na Tabela 26 e associá-los assim com o diagrama de Hasse primal (Figura 24) e o diagrama de Hasse dual (Figura 25).

O códon UGG codifica o aminoácido triptofano (W) e os códon GAA e GAG codificam o aminoácido ácido glutâmico (E). A distância média entre o códon UGG e os códon GAA e GAG é 4. O triptofano (W) é um anel aromático e o ácido glutâmico (E) é um aminoácido

hidrofílico. Os códons GAU e GAC codificam o aminoácido ácido aspártico (D), que é hidrofílico. Na Tabela 26 a distância média entre esses códons e o aminoácido triptofano (W) é igual a 4. Os códons AAA e AAG codificam o aminoácido lisina (K), e a sua distância média com o códon que codifica o aminoácido triptofano (W) é igual a 4. A lisina (K) é hidrofílico, sendo este diferente do aminoácido triptofano (W). Os códons AAU e AAC codificam o aminoácido asparagina (N), sendo um aminoácido hidrofílico. A sua distância média com o códon do aminoácido triptofano (W) é igual a 4. Os códons CAA e CAG codificam o aminoácido glutamina (Q), que possui característica hidrofílica e a distância média com o códon do aminoácido triptofano (W), baseado na Tabela 26 é igual a 5. Os códons CAU e CAC codificam o aminoácido histidina (H), que é hidrofílico e a distância média com o códon do aminoácido triptofano (W) é igual a 5. Os códons CCU, CCC, CCA e CCG codificam o aminoácido prolina (P), que possui cadeia alifática e a distância média com o códon do aminoácido triptofano, baseado na Tabela 26 é igual a 4. Já os códons CGU, CGC, CGA, CGG, AGA e AGG codificam o aminoácido arginina (R), que são códons com características básicas. A distância média com o códon de aminoácido triptofano (W) é igual a 1. Nos diagramas, esses códons aparecem próximos um do outro.

Os códons CGU, CGC, CGA, CGG, AGA e AGG codificam o aminoácido arginina (R), que possui como característica básico e os códons GAA e GAG codificam o aminoácido ácido glutâmico (E) que é hidrofílico. Pode-se notar que a média da distância entre os códons desses aminoácidos é igual a 4,33. Os códons GAU e GAC codificam o aminoácido ácido aspártico (D), que é hidrofílico. A distância média dos códons do aminoácido arginina (R) entre os códons do aminoácido ácido aspártico (D) é igual a 4,33. Os códons UAU e UAC codificam o aminoácido tirosina (Y), que também possui características hidrofílicas. A distância média com os códons do aminoácido arginina (R), é igual a 4,67. Os códons UAA, UGA e UAG codificam o aminoácido STOP, um aminoácido importante, que é utilizado para interromper a proteína antes de seu término. Além disso, os códons UAA e UAG possuem características hidrofílicas. A distância média dos códons do aminoácido arginina (R) entre os códons do aminoácido STOP é igual a 4. Nas Figuras 24 e 25, podemos observar que o códon UGA está localizado perto dos códons que codificam o aminoácido arginina, porém, os códons UAA e UAG aparecem distantes dos códons CGU, CGC, CGA, CGG, AGA e AGG devido as suas características, o que explica o fato da média entre eles ser igual a 4.

Os códons UCU, UCC, UCA, UCG, AGU e AGC codificam o aminoácido serina (S),

classificado como hidroxila e os códons CAA e CAG que codificam o aminoácido glutamina (Q), hidrofílico. A distância média entre os códons desses aminoácidos é igual a 4,17. Os códons CAU e CAC codificam o aminoácido histidina (H), sendo ele classificado como hidrofílico. Também, pode-se notar que a distância média entre os códons que codificam o aminoácido serina (S) e os códons que codificam o aminoácido histidina é 4,17.

O códon UGG codifica o aminoácido triptofano (W) e os códons UGA, UAA e UAG codificam o aminoácido STOP, sendo este utilizado para interromper a proteína antes de seu término. Na Tabela 26, a distância média entre esses aminoácidos é 1,33. Podemos notar no diagrama que os códons UAA e UAG possuem características hidrofílicas, sendo diferente do aminoácido triptofano, que é um anel aromático. Porém o códon UGA, está localizado próximo do códon UGG, o que justifica o fato da média entre os códons ter sido 1,33. Um outro caso é o códon AUG que codifica o aminoácido metionina (M) e os códons AUU, AUC e AUA que codificam o aminoácido isoleucina (I). Ambos os aminoácidos são semelhantes, pois possuem características hidrofílicas. Nos diagramas (Figuras 24 e 25) nota-se que os códons estão próximos um do outro, justificando o fato de que na Tabela 26, a distância média entre os aminoácidos metionina (M) e isoleucina (I) ser 1,33.

Pode-se observar a conexão entre os outros casos na Tabela 26 com os diagramas (Figuras 24 e 25) construídos. Se realizarmos a análise de todos os casos, nos quais, as distâncias médias entre as distâncias de Hamming foram maiores ou iguais a 4 ou valores menores, as suas distâncias estão refletidas nos diagramas de Hasse, com base nas diferenças biológicas entre os códons.

3.3 ROTULAMENTO *C*

Para o estudo do Rotulamento *C*, será utilizada a permutação 2031, para o reticulado primal, obtendo a associação apresentada na Tabela 27. Como consequência, a permutação 0213 do Rotulamento *C* será utilizada para a construção do reticulado dual. Na Tabela 28 é apresentada a associação estabelecida. Neste caso, assim como no Rotulamento *B*, não foi utilizada a complementaridade biológica das bases, mas sim, a complementaridade algébrica, na qual 00 - 11 e 01 - 10. Desta forma, as bases nitrogenadas referentes a essa complementaridade algébrica são: $A - C$ e $G - U$, ou vice-versa.

Tabela 27 – Associação estabelecida pela permutação 2031 primal do Rotulamento C.

A	C	G	U
2	0	3	1
11	00	01	10

Fonte: Da autora.

Tabela 28 – Associação estabelecida pela permutação 0213 dual do Rotulamento C.

A	C	G	U
0	2	1	3
00	11	10	01

Fonte: Da autora.

A seguir será apresentado o procedimento para a construção dos reticulados booleanos primal e dual associados ao Rotulamento C, analogamente ao que foi realizado para os rotulamentos A e B.

3.3.1 Procedimento para a construção dos reticulados booleanos

Utilizando as operações da álgebra booleana (Tabelas 11 e 12) e a permutação 2031 do Rotulamento C (primal), obtemos os dados das Tabelas 29 e 30.

Tabela 29 – Operação primal \wedge (e) do Rotulamento C.

\wedge	00	01	10	11
00	00	00	00	00
01	00	01	00	01
10	00	00	10	10
11	00	01	10	11

Fonte: Da autora.

\wedge	C	G	U	A
C	C	C	C	C
G	C	G	C	G
U	C	C	U	U
A	C	G	U	A

Fonte: Da autora.

Tabela 30 – Operação primal \vee (ou) do Rotulamento C.

\vee	00	01	10	11
00	00	01	10	11
01	01	01	11	11
10	10	11	10	11
11	11	11	11	11

Fonte: Da autora.

\vee	C	G	U	A
C	C	G	U	A
G	G	G	A	A
U	U	A	U	A
A	A	A	A	A

Fonte: Da autora.

Utilizando as operações da álgebra booleana (Tabelas 11 e 12) e a permutação 0213 do Rotulamento C (dual), obtemos os dados das Tabelas 31 e 32.

Tabela 31 – Operação dual \wedge (e) do Rotulamento C .

\wedge	00	01	10	11
00	00	00	00	00
01	00	01	00	01
10	00	00	10	10
11	00	01	10	11

Fonte: Da autora.

\wedge	A	U	G	C
A	A	A	A	A
U	A	U	A	U
G	A	A	G	G
C	A	U	G	C

Fonte: Da autora.

Tabela 32 – Operação dual \vee (ou) do Rotulamento C .

\vee	00	01	10	11
00	00	01	10	11
01	01	01	11	11
10	10	11	10	11
11	11	11	11	11

Fonte: Da autora.

\vee	A	U	G	C
A	A	U	G	C
U	U	U	C	C
G	G	C	G	C
C	C	C	C	C

Fonte: Da autora.

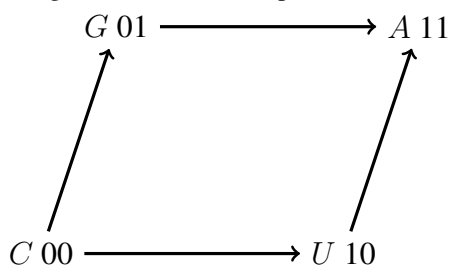
O reticulado booleano primal será obtido a partir da Tabela 30. Observe que:

- A base C (citosina) se liga às bases G (guanina), U (uracila) e A (adenina).
- A base G (guanina) se liga a ela mesma e à base A (adenina).
- A base U (uracila) se liga a ela mesma e à base U (uracila).
- A base A (adenina) se liga apenas a ela mesma.

Os reticulados primal e dual são apresentados nas Figuras 26 e 27.

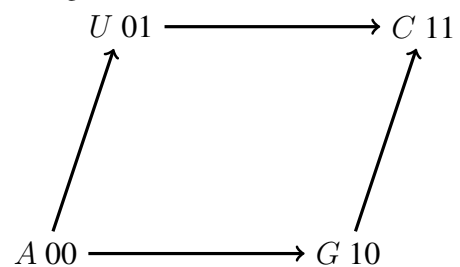
Reticulados primal e dual associados ao Rotulamento C .

Figura 26 – Reticulado primal (2031).



Fonte: Da autora.

Figura 27 – Reticulado dual (0213).



Fonte: Da autora.

3.3.2 Procedimento para a construção dos Diagramas de Hasse

Para a construção dos diagramas de Hasse do Rotulamento C , seguiremos o mesmo procedimento dos rotulamentos A e B . Lembrando que, no Rotulamento A foi considerada a complementaridade biológica das bases e no Rotulamento B foi considerada a complementaridade algébrica, sendo $A - G$ e $C - U$. Já no Rotulamento C também iremos utilizar a complementaridade algébrica, sendo assim, $A - C$ e $G - U$.

Na 1ª linha, está o códon que apresenta seu elemento máximo, sendo ele $AAA \leftrightarrow 111111$. No diagrama de Hasse dual (Figura 29), o elemento máximo é o $CCC \leftrightarrow 111111$.

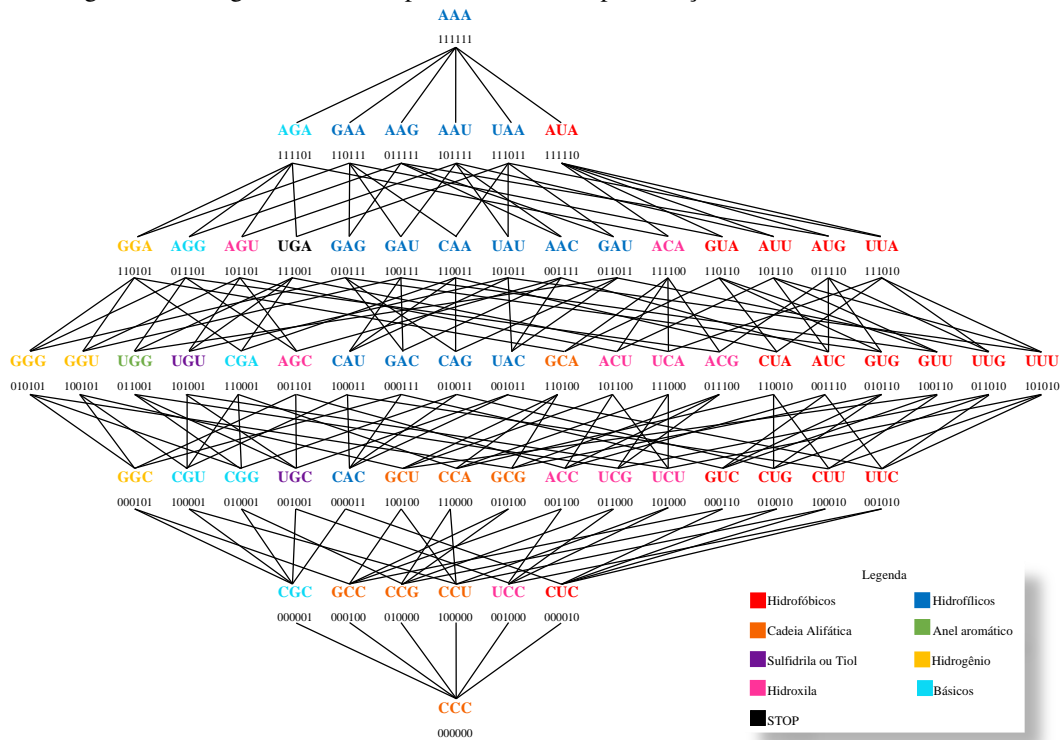
Na 7ª linha, estão os códons que possuem elemento mínimo, sendo eles, na Figura 28, o códon $CCC \leftrightarrow 000000$ e na Figura 29, o códon $AAA \leftrightarrow 000000$.

As outras linhas do diagrama de Hasse, seguem o mesmo procedimento dos rotulamentos A e B , levando em consideração o complementar algébrico de cada códon. Nos diagramas, o número de códons existentes em cada linha, se dá devido a representação algébrica de cada códon, estando relacionada ao número de zeros (0) ou uns (1) existente em cada códon. Além disso, assim como nos rotulamentos A e B , no rotulamento C , determinamos as distâncias de Hamming entre os códons através das arestas, que podem ser observadas nas construções (Figuras 28 e 29).

A Figura 28 apresenta o diagrama de Hasse primal para a permutação 2031 do Rotulamento C , escolhida de forma aleatória. A Figura 29 apresenta o diagrama de Hasse dual para a permutação 0213 do Rotulamento C . Sua construção é análoga à construção do diagrama de Hasse primal.

3.3.3 Análise do diagrama primal

Figura 28 – Diagrama de Hasse primal referente à permutação 2031 do Rotulamento C.



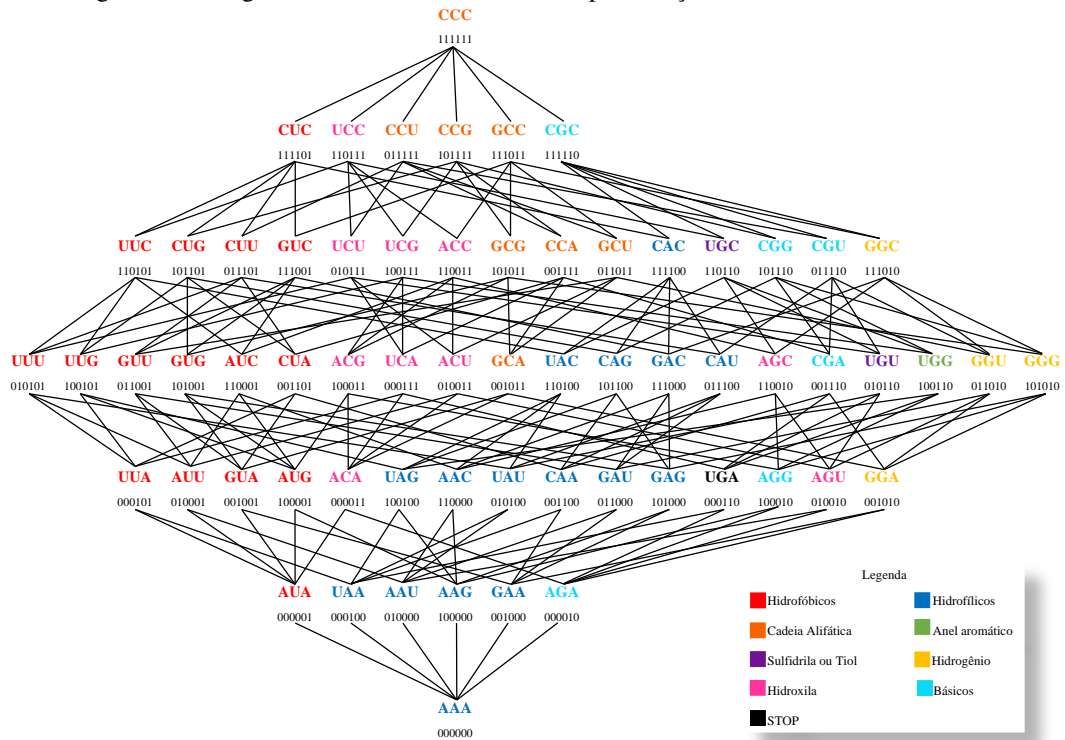
Fonte: Da autora.

Analisando a complementaridade das bases nitrogenadas do Rotulamento C, notamos a união das bases adenina (A) com citosina (C) e guanina (G) com uracila (U). Vamos tomar como exemplo a cadeia CUC, UCC, CCU, CCG, GCC, CGC tem como imagem a anti-cadeia AUA, UAA, AAU, AAG, GAA, AGA, seguindo a complementaridade algébrica dos códons. Na lateral direita podemos encontrar os códons hidrofóbicos (vermelho) e no centro superior no diagrama, podemos encontrar os códons hidrofílicos (azul). Diferente dos códons hidrofóbicos não vemos uma regularidade na distribuição dos códons hidrofílicos.

O códon UGG (anel aromático), neste diagrama, ficou distante do códon UGA (STOP), com relação aos diagramas dos rotulamentos A e B, e podemos encontrá-los na cadeia lateral direita. Os códons com característica cadeia alifática estão no centro inferior do diagrama e os códons com característica hidroxila, no centro do diagrama. Na lateral direita, encontramos os códons com característica hidrogênio, básicos, sulfidrina ou tiol e alguns códons com característica hidroxila.

3.3.4 Análise do diagrama dual

Figura 29 – Diagrama de Hasse dual referente à permutação 0213 do Rotulamento C.



Fonte: Da autora.

Na cadeia lateral direita estão os códon hidrofóbicos, representados em vermelho e no centro inferior do diagrama os códon hidrofílicos, representados em azul. Da mesma forma que o diagrama primal, não vemos uma regularidade na distribuição dos códon hidrofílicos.

O códon UGG (anel aromático), neste diagrama, ficou distante do códon UGA (STOP), com relação aos diagramas dos rotulamentos A e B. Os códon com característica cadeia alifática estão no centro superior do diagrama. Também, temos os códon com característica hidroxila no centro do diagrama. Na lateral esquerda, encontramos os códon com característica hidrogênio, básicos, sulfidrila ou tiol e alguns códon com característica hidroxila.

3.3.5 Análise das diferenças e semelhanças físico-químicas dos aminoácidos por meio das distâncias de Hamming

A Tabela 33 apresenta a média das distâncias de Hamming entre os pares de aminoácidos. Da mesma forma que os casos analisados para os rotulamentos *A* e *B*, pode-se notar que os valores das médias das distâncias de Hamming maiores ou iguais a 4 aparecem em negrito, pois são considerados valores altos para essas distâncias, de acordo com Sánchez, Morgado e Grau (2004).

Tabela 33 – Médias das distâncias de Hamming entre pares de aminoácidos do Rotulamento *C*.

*	G	W	C	R	S	V	L	F	M	I	E	D	Y	K	N	Q	H	A	T	P	STOP
G	1	3	3	2	3,33	3	4,33	5	4	4	2	2	4	3	3	3	3	2	3	3	3,67
W	3	0	1,5	1,83	2,17	5	3,5	3,5	3	4,33	3,5	4,5	2,5	2,5	3,5	2,5	3,5	4	3	3	1,33
C	3	1,5	0,5	2,17	1,83	5	3,83	2,5	4,5	3,83	4,5	3,5	1,5	3,5	2,5	3,5	2,5	4	3	3	2,17
R	2	1,83	2,17	1,83	2,83	4	3,67	4,17	4,17	4,39	2,83	3,17	3,17	3,17	3,5	2,5	2,83	3	3,33	2,67	2,5
S	3,33	2,17	1,83	2,83	1,83	2,67	3,39	2,5	3,17	2,94	4,5	4,17	2,83	3,5	3,17	4,17	3,83	3	2	2,67	2,83
V	3	5	5	4	2,67	1	2,33	3	2	2	2	2	4	3	3	3	3	2	3	3	4,33
L	4,33	3,5	3,83	3,67	3,39	2,33	1,39	1,83	2,5	2,67	3,17	3,5	2,83	3,5	3,83	2,17	2,5	3,33	3,67	2,38	2,83
F	5	3,5	2,5	4,17	2,5	3	1,83	0,5	2,5	1,83	4,5	3,5	1,5	3,5	2,5	3,5	2,5	4	3	3	2,83
M	4	3	4,5	4,17	3,17	2	2,5	2,5	0	1,33	2,5	3,5	3,5	1,5	2,5	3,5	3,5	3	2,25	4	3
I	4	4,33	3,83	4,39	2,94	2	2,67	1,83	1,33	0,89	3,17	2,83	2,83	2,17	1,83	4,17	3,83	3	2	3,92	3,44
E	2	3,5	4,5	2,83	4,5	2	3,17	4,5	2,5	3,17	0,5	1,5	3,5	1,5	2,5	1,5	2,5	3	4	4	2,67
D	2	4,5	3,5	3,17	4,17	2	3,5	3,5	3,5	2,83	1,5	0,5	2,5	2,5	1,5	2,5	1,5	3	4	4	3,83
Y	4	2,5	1,5	3,17	2,83	4	2,83	1,5	3,5	2,83	3,5	2,5	0,5	2,5	1,5	2,5	1,5	5	4	4	1,83
K	3	2,5	3,5	3,17	3,5	3	3,5	3,5	1,5	2,17	1,5	2,5	2,5	0,5	1,5	2,5	3,5	4	3	5	1,83
N	3	3,5	2,5	3,5	3,17	3	3,83	2,5	2,5	1,83	2,5	1,5	1,5	1,5	0,5	3,5	2,5	4	3	5	2,83
Q	3	2,5	3,5	2,5	4,17	3	2,17	3,5	3,5	4,17	1,5	2,5	2,5	2,5	3,5	0,5	1,5	4	5	3	1,83
H	3	3,5	2,5	2,83	3,83	3	2,5	2,5	3,5	3,83	2,5	1,5	1,5	3,5	2,5	2,5	0,5	4	5	3	2,83
A	2	4	4	3	3	2	3,33	4	3	3	3	5	4	4	4	4	4	1	2	2	4,67
T	3	3	3	3,33	2	3	3,67	3	2,25	2	4	4	4	3	3	3	5	2	1	3	3,67
P	3	3	3	2,67	2,67	3	2,38	3	4	3,92	4	4	4	5	5	5	3	2	3	1	3,67
STOP	3,67	1,33	2,17	2,5	2,83	4,33	2,83	2,83	3	3,44	2,67	3,83	1,83	1,83	2,83	2,83	2,83	4,67	3,67	3,67	0,89

Fonte: Da autora.

Nota: *Representação dos aminoácidos por um único símbolo. A - Alanina, C - Cisteína, D - Ác. Aspártico, E - Ác. Glutâmico, F - Fenilalanina, G - Glicina, H - Histidina, I - Isoleucina, K - Lisina, L - Leucina, M - Metionina, N - Asparagina, P - Prolina, Q - Glutamina, R - Arginina, S - Serina, T - Treonina, V - Valina, W - Triptofano, Y - Tirosina.

Pode-se notar que aminoácidos com grandes diferenças em suas propriedades apresentaram altos valores de distâncias de Hamming, da mesma forma dos casos analisados para os rotulamentos *A* e *B*. Por exemplo, as distâncias médias entre os códons que codificam o aminoácido glicina (G) e os códons que codificam os aminoácidos leucina (L), fenilalanina (F), metionina (M), isoleucina (I) e tirosina (Y) foram maiores ou iguais a 4. Os códons associados ao aminoácido triptofano (W) e os códons associados ao aminoácido valina (V), isoleucina (I), ácido aspártico (D) e alanina (A) foram maiores ou iguais a 4. Os códons associados a cisteína (C) e os códons associados à valina (V), metionina (M), ácido glutâmico (E) e alanina (A) também foram maiores ou iguais a 4. Os códons associados a arginina (R) e os códons associados à valina (V), fenilalanina (F), metionina (M) e isoleucina (I) foram maiores ou iguais a 4. Os códons associados ao aminoácido serina (S) e os códons que codificam os aminoácidos ácido

glutâmico (E), ácido aspártico (D) e glutamina (Q) foram maiores ou iguais a 4. Os outros casos podem ser observados na Tabela 33.

A seguir serão apresentados alguns exemplos com algumas diferenças e semelhanças biológicas acerca dos cálculos obtidos na Tabela 33 e associá-los assim com o diagrama de Hasse primal (Figura 28) e o diagrama de Hasse dual (Figura 29).

O códon UGG codifica o aminoácido triptofano (W) e os códons GUU, GUC, GUA e GUG codificam o aminoácido valina (V). A distância média entre o códon UGG e os códons GUU, GUC, GUA e GUG é igual a 5. O triptofano (W) é um anel aromático e a valina (V) é um aminoácido hidrofóbico. Os códons AUU, AUC e AUA codificam o aminoácido isoleucina (I), que é hidrofóbico. A distância média entre esses códons e o códon do aminoácido triptofano é igual a 4,33. Os códons GAU e GAC codificam o aminoácido ácido aspártico (D), e sua distância média com o códon que codifica o aminoácido triptofano (W) é igual a 4,5. O ácido aspártico (D) é hidrofílico, sendo este diferente do triptofano. Os códons GCU, GCC, GCA e GCG codificam o aminoácido alanina (A), sendo uma cadeia alifática. A sua distância média com o códon do aminoácido triptofano é igual a 4. Já os códons UGU e UGC codificam o aminoácido cisteína (C), que são códons com características sulfidríla ou tiol. A distância média com o códon de aminoácido triptofano é igual a 1,5. Podemos observar que nos diagramas das Figuras 28 e 29 esses códons estão próximos do códon do aminoácido triptofano (W), concluindo assim, que eles possuem características em comum.

Os códons UGU e UGC codificam o aminoácido cisteína (C), que são códons com características sulfidríla ou tiol e os códons GUU, GUC, GUA e GUG codificam o aminoácido valina (V), que possui característica hidrofóbica. Pode-se notar que a média da distância entre os códons desses aminoácidos é igual a 5. O códon AUG codifica o aminoácido metionina (M), sendo este hidrofóbico. A distância média dos códons do aminoácido cisteína (C) e os códons do aminoácido metionina (M) é igual a 4,5. Os códons GAA e GAG codificam o aminoácido ácido glutâmico (E), que possuem características hidrofílicas. A distância média com os códons do aminoácido cisteína (C), é igual a 4,5. Os códons GCU, GCC, GCA e GCG codificam o aminoácido alanina (A), sendo uma cadeia alifática. Pode-se notar que a média das distâncias entre os códons do aminoácido cisteína (C) na é igual a 4. Já os códons CGU, CGC, CGA, CGG, AGA e AGG codificam o aminoácido arginina (R), que são códons com características básicas. A distância média com o códon do aminoácido cisteína (C) é igual a 2,17. Pode-se observar que nos diagramas das Figuras 28 e 29 esses códons estão próximos do códon do aminoácido

triptofano (W), concluindo assim, que eles possuem características em comum.

Os códons CGU, CGC, CGA, CGG, AGA e AGG codificam o aminoácido arginina (R), que são códons classificados como básicos e os códons GUU, GUC, GUA e GUG codificam o aminoácido valina (V), que possui característica hidrofóbica. Pode-se notar que a distância média entre os códons desses aminoácidos é igual a 4. Os códons UUU e UUC codificam o aminoácido fenilalanina (F). Como é um aminoácido com característica hidrofóbica, pode-se notar que a sua distância média entre o aminoácido arginina (R) é igual a 4,17. O códon AUG codifica o aminoácido metionina (M), sendo este hidrofóbico. A distância média dos códons do aminoácido arginina (R) entre os códons do aminoácido metionina (M) é igual a 4,17. Já os códons GGU, GGC, GGA e GGG codificam o aminoácido glicina (G), que são códons que tem hidrogênio como característica. A distância média com o códon do aminoácido arginina (R) é igual a 2. Pode-se notar que tanto os códons que codificam o aminoácido arginina (R) quanto os códons que codificam o aminoácido glicina (G) estão próximos um do outro, nos diagramas (Figuras 28 e 29), concluindo assim que eles possuem características em comum.

Os códons ACU, ACC, ACA e ACG codificam o aminoácido treonina (T), que é classificado como hidroxila. A distância média com o códon do aminoácido serina (S) é igual a 2. Como os aminoácidos serina (S) e treonina (T) são classificados como hidroxila, possuindo as mesmas características, faz sentido o fato delas estarem tão próximas uma da outra nos diagramas (Figuras 28 e 29). Assim, justifica o valor da média entre eles ser menor.

Pode-se observar os outros casos na Tabela 33 com os diagramas (Figuras 28 e 29) construídos. Em todos os casos, as distâncias que foram maiores ou iguais a 4 os aminoácidos apresentam diferenças significativas entre si, e isso se reflete na distância entre eles nos diagramas de Hasse primal e dual.

4 CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

As ferramentas matemáticas abordadas neste trabalho podem ser utilizadas em estudos relacionados ao código genético, em particular a fenômenos mutacionais, auxiliando na modelagem dos mesmos. Há tempos que diversos pesquisadores buscam representar o que mundo biológico realiza, a fim de identificar suas características, propriedades, funções e implicações do modelo estudado. Além disso, a modelagem matemática associada ao código genético é uma área de pesquisa que está em constante expansão e aprimoramento de técnicas, pois vários autores procuram estruturar e modelar o código genético, com o intuito de validar suas hipóteses.

Este trabalho mostra algumas aplicações de estruturas matemáticas aplicadas ao código genético, em particular o diagrama de Hasse, que permite uma melhor visualização dos códons do código genético, possibilitando analisar as propriedades dos aminoácidos e classificar os códons de acordo com a sua hidropacidade e outras características, tornando-se uma ferramenta eficaz na análise de diversos fenômenos biológicos, em particular, estudos envolvendo análises mutacionais.

Neste estudo, percebe-se uma interessante caracterização algébrica acerca das construções realizadas. Com os diagramas de Hasse foi possível uma análise biológica de cada rotulamento, de acordo com as características dos aminoácidos. Apenas no Rotulamento *A* podemos destacar que o códon que codifica um aminoácido hidrofóbico é sempre complementar a um códon que codifica um aminoácido hidrofílico. No Rotulamento *A* foi utilizada a complementaridade biológica e nos rotulamentos *B* e *C* foram utilizadas as complementaridades algébricas. Nessas construções foi possível observar que tanto no Rotulamento *A* quanto no Rotulamento *B* houve um equilíbrio na distribuição dos códons de acordo com suas características físico-químicas. Além disso, os códons com as mesmas características permaneceram juntos. Já no Rotulamento *C* não houve uma regularidade na distribuição dos códons hidrofílicos e os códons UGG e UGA, com relação aos diagramas dos rotulamentos *A* e *B*, estão distantes um do outro, o que nos motiva a estudos futuros sobre este Rotulamento para analisar essas diferenças com os rotulamentos *A* e *B*.

Conforme a proposta de Sánchez, Morgado e Grau (2004) foram calculadas as médias das distâncias de Hamming entre os pares de aminoácidos para cada um dos rotulamentos. Com base nas características de cada aminoácido, foi realizada um análise das diferenças e semelhan-

ças biológicas acerca dos cálculos obtidos, relacionando com as construções realizadas. Como mencionado no Capítulo 3, para as diferenças foram consideradas como parâmetro as médias maiores ou iguais a 4, pois segundo Sánchez, Morgado e Grau (2004) quando se obtém esses valores nas médias das distâncias de Hamming entre pares de códons, pode-se alterar as propriedades dos aminoácidos e as funções biológicas das proteínas, acarretando riscos de grandes mutações genéticas. Um fato a ser destacado, é que no rotulamento *A*, os valores obtidos das médias são iguais ao caso proposto por Sánchez, Morgado e Grau (2004), porém utilizou-se outra permutação. Como em cada rotulamento utilizamos um tipo de permutação, temos que as associações geradas em cada caso foram diferentes, ou seja, os valores das tabelas obtidos foram diferentes, entretanto há alguns casos em que os resultados foram iguais. Ao associar cada rotulamento com suas respectivas construções, notamos que em todos os casos, as distâncias estão refletidas nos diagramas de Hasse, com base nas diferenças e semelhanças biológicas entre os códons. Portanto, os diagramas de Hasse refletem de forma significativa as características físico-químicas dos aminoácidos baseadas nos cálculos das distâncias de Hamming, de forma a poder utilizar esse resultados em estudos de fenômenos mutacionais.

Pode-se observar a interdisciplinaridade presente neste estudo, uma vez que foi possível relacionar elementos de Matemática, Biologia e Códigos Corretores de Erros, que torna-se relevante em diversas áreas de pesquisa, como engenharia genética, uma vez que pode servir como base para estudos relacionados a fenômenos mutacionais.

4.1 Sugestões para Trabalhos Futuros

A seguir serão apresentadas algumas sugestões para trabalhos futuros, como forma de dar continuidade a este trabalho.

- a) Relacionar as construções dos diagramas de Hasse para os rotulamentos *A*, *B* e *C* com códigos corretores de erros, em particular, os códigos de subespaço;
- b) Apresentar a conexão entre as estruturas dos diagramas de Hasse e os elementos de uma extensão de Galois $GF(2^6)$;
- c) Estudar as conexões entre as construções realizadas e a estrutura de um hipercubo relacionado ao código genético;

- d) Efetuar análises mutacionais acerca das distâncias médias de Hamming obtidas para os rotulamentos A , B e C .

REFERÊNCIAS

- ALBERTS, B. *et al.* *Biologia molecular da célula*. 5. ed. Porto Alegre: ArtMed, 2010.
- ALBERTS, B. *et al.* *Biologia molecular da célula*. 6. ed. Porto Alegre: ArtMed, 2017.
- COSTELO JR., D. J.; LIN, S. *Error Control coding: fundamentals and applications*. Englewood Cliffs: Prentice Hall, 1983.
- DOMINGUES, H. H.; IEZZI, G. *Álgebra moderna*. 4. ed. São Paulo: Atual, 2003.
- FARIA, L. C. B. D. *Existências de códigos corretores de erros e protocolos de comunicação em sequências de DNA*. 2011. 322f. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2011.
- FERNANDES, R. S. *Aplicações de estruturas algébricas na modelagem do código genético*. 2018. 48 f. Trabalho de Conclusão de Curso (Graduação em Matemática Licenciatura) - Instituto de Ciências Exatas, Universidade Federal de Alfenas, 2018.
- GERÔNIMO, J. R.; FRANCO, V. S. *Fundamentos da matemática: uma introdução à lógica matemática, teoria dos conjuntos, relações e funções*. 2. ed. Maringá: Eduem, 2008.
- GRIFFITHS, A. J. F. *et al.* *Introdução à genética*. Rio de Janeiro: Guanabara Koogan, 2006.
- HEFEZ, A.; FERNANDEZ, C. S. *Introdução à álgebra linear*. Rio de Janeiro: SBM, 2012.
- HIB, J.; ROBERTIS, E. M. F. *Bases da biologia celular e molecular*. 4. ed. Rio de Janeiro: Guanabara Koogan, 2006.
- JIMÉNEZ-MONTAÑO, M. A.; MORA-BASÁÑEZ, C. R.; PÖSCHEL, T. On the hypercube structure of the genetic code. *Bioinformatics and Genome Research*, p. 445, 1994.
- JIMÉNEZ-MONTAÑO, M. A.; MORA-BASÁÑEZ, C. R.; PÖSCHEL, T. The hypercube structure of the genetic code explains conservative and non-conservative aminoacid substitutions in vivo an in vitro, *BioSystems*, 39 p. 117–125, 1996.
- MACWILLIAMS, F. J.; SLOANE, N. J. A. *The theory of Error-Correcting Codes*. Elsevier Science: North-Holland, B.V., 1981.
- OLIVEIRA, A. J. *Análise algébrica dos rotulamentos associados ao mapeamento do código genético*. 2012. 105f. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2012.
- PETTINARI, M. J. Las bacterias y nosotros, tan diferentes... y tan parecidos. Mitos y verdades de las diferencias entre eucariotas y procariotas. *Química Viva [en línea]*, v. 9, n. 1, p. 3-11, 2010. Disponible en: <https://www.redalyc.org/articulo.oa?id=86312852002>. Acesso em 5 May 2020.

ROCHA, A. S. L. *Modelo de sistema de comunicação digital para o mecanismo de importação de proteínas mitocondriais através de códigos corretores de erros*. 2010. 165f. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2010.

SÁNCHEZ, R. *et al.* A new DNA sequences vector space on a genetic code Galois field, *MATCH Commun. Math. Comput. Chem.*, v. 54, 2005.

SÁNCHEZ, R.; GRAU, R.. A novel Lie algebra of the genetic code over the Galois field of four DNA bases. *Mathematical Biosciences*, v. 202, p. 156–174, 2006.

SÁNCHEZ, R.; MORGADO, E.; GRAU, R. Gene algebra from a genetic code algebraic structure. *Journal of Mathematical Biology*, v. 51, n. 4, p. 431-457, 2005.

SÁNCHEZ, R.; MORGADO, E.; GRAU, R. The genetic Code Boolean Lattice. *MATCH Commun. Math. Comput. Chem*, v.52, n. 52, p. 29-46, 2004.

ZAHA, A.; FERREIRA, H. B.; PASSAGLIA, L. M. P. *Biologia molecular básica*. 5. ed. Porto Alegre: ArtMed, 2014.