



# Caracterização das propriedades dos aminoácidos por meio do Diagrama de Hasse associado ao rotulamento $A$ do código genético

Characterization of the properties of amino acids using the Hasse diagram associated with the  $A$  labeling of the genetic code

Roberta Siqueira Fernandes

Universidade Federal de Alfenas

robertaf\_mat@hotmail.com

ORCID: 0000-0002-5286-704

Anderson José de Oliveira

Universidade Federal de Alfenas

anderson.oliveira@unifal-mg.edu.br

ORCID: 0000-0002-5467-9827

**Resumo.** A aplicação de estruturas matemáticas na modelagem do código genético é algo que diversos pesquisadores vêm realizando, buscando caracterizar matematicamente o que o mundo biológico realiza. Dessas estruturas, os reticulados booleanos, os diagramas de Hasse e a Distância de Hamming, têm sido utilizados no processo de análise de fenômenos mutacionais, possibilitando uma análise físico-química dos aminoácidos, além de permitir a caracterização algébrica do código genético. O matemático Claude E. Shannon foi o precursor da teoria matemática da comunicação e aplicou elementos de engenharia na biologia. As construções deste trabalho será por meio do mapeamento das bases nitrogenadas adenina, citosina, guanina, timina/uracila,  $\{A, C, G, T/U\}$  com o anel  $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ , onde é possível obter 24 permutações, as quais podem ser divididas em 3 rotulamentos ( $A$ ,  $B$  e  $C$ ), de acordo com as características geométricas dos mesmos. O objetivo deste trabalho é apresentar a construção dos reticulados booleanos e dos diagramas de Hasse primal e dual das permutações 0132 e 2310, respectivamente, do rotulamento  $A$ . Dessas construções será realizado uma análise das diferenças e semelhanças físico-químicas dos aminoácidos, através da caracterização biológica acerca das construções e dos cálculos das médias das distâncias de Hamming entre os códons, os quais poderão ser utilizados no processo de análise de fenômenos mutacionais.

**Palavras-chave.** Álgebra. Reticulados Booleanos. Distância de Hamming. Biologia.

**Abstract.** The application of mathematical structures to the modeling of the genetic code is something that several researchers have been carrying out, seeking to characterize mathematically what the biological world accomplishes. Of these structures, the Boolean lattices, the Hasse diagrams and the Hamming Distance, have been used in the process of analyzing mutational phenomena, enabling 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 the mathematical theory of communication and applied elements of engineering to biology. The constructions proposed in this paper make use of the mapping from the set consisting of the nitrogenous bases adenine, cytosine, guanine, and thymine/uracil,  $\{A, C, G, T/U\}$ , to the set  $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ , where it is possible to obtain 24 permutations, which can be divided into 3 labels ( $A$ ,  $B$  and  $C$ ), according to geometric characteristics. The objective of this work is to present the construction of the primal and dual Boolean lattices and Hasse diagrams of permutations 0132 and 2310, respectively, of the labeling  $A$ . An analysis of the differences and physicochemical similarities of the amino acids will be carried out on these constructions, through the biological characterization of the constructions and the calculation of the averages of the Hamming distances between the codons, which can be used to the process of analysis of mutational phenomena.

**Keywords.** Algebra. Boolean Lattices. Hamming Distance. Biology.

**Mathematics Subject Classification (MSC):** primary 06E05; secondary 92D20, 06A15.

## 1 Introdução

A modelagem algébrica do código genético tem sido alvo de diversas pesquisas, cujo objetivo é identificar as propriedades, as características e implicações do modelo a ser estudado. Além disso, tanto na Matemática quanto na Biologia, este representa um problema instigante e atraente para diversos pesquisadores.

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 de proteínas. Contudo, no século XX os biólogos conseguiram decifrar a linguagem da vida. Entretanto, ainda há muitos “mistérios da máquina da vida” para se desvendar. Com esse intuito, muitas pesquisas vêm sendo realizadas com estudos relacionados a problemas envolvendo o código genético, entre outros fenômenos biológicos, nas quais muitos pesquisadores utilizam da Matemática para contribuir na análise e descrição de conceitos teóricos.

Nessa temática, uma das propostas encontradas foi a de [14]. Utilizando dados experimentais, em seu estudo é apresentada a construção de um reticulado booleano do código

genético, além da construção de um diagrama de Hasse e dos cálculos das distâncias de Hamming entre os códons. Um outra proposta, é a de [11] que apresenta uma representação do código genético por meio de estruturas algébricas, com o objetivo de identificar as propriedades, características e implicações do modelo, como forma de explicar fenômenos biológicos. As estruturas algébricas utilizadas nesse estudo foram diagrama de Hasse, reticulados booleanos, grupos, corpos, anéis e extensões de Galois.

Os códons do código genético são formados por uma trinca de bases nitrogenadas com a possibilidade de 64 combinações distintas que são consideradas por [13] em seu estudo. Nesse trabalho, é apresentado um modelo por meio da construção da tabela do código genético. Para a construção, é feita uma bijeção do alfabeto biológico,  $\{A, C, G, T/U\}$ , com a seguinte representação:  $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{00, 01, 10, 11\}$ . 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 [13] leva em consideração a importância biológica das bases nitrogenadas, fundamentada pela ocorrência de erros estudadas na Biologia.

Em [4] e [12] é apresentada a modelagem do DNA através de uma analogia do alfabeto do código genético, que está relacionado com o conjunto de nucleotídeos  $N = \{A, C, G, T/U\}$  e o alfabeto 4-ário na estrutura de anel, denotado por  $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ . A princípio, sabe-se que é desconhecido o mapeamento entre  $N \leftrightarrow \mathbb{Z}_4$ , então a sequência de DNA é rotulada obtendo 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  $C$ , 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 a complementaridade dos nucleotídeos, na qual, tem-se que Adenina se emparelha com a Uracila e a Citosina se emparelha com a Guanina, ou vice-versa.

Na proposta de [5], 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  $GF(2^6)$ .

Com base nos estudos e por meio deste trabalho, temos como propósito a apresentação da construção dos reticulados booleanos primal e dual para as permutações 0132 e 2310 do rotulamento A. Além disso, serão apresentadas as construções dos Diagramas de Hasse primal e dual, a caracterização biológica acerca das construções realizadas e uma análise das diferenças e semelhanças físico-químicas dos aminoácidos acerca dos cálculos das

médias das distâncias de Hamming entre os códons, os quais poderão ser utilizados no processo de análise de fenômenos mutacionais.

## 2 Conceitos Fundamentais

Nesta Seção serão apresentados os principais conceitos teóricos utilizados neste trabalho. As referências utilizadas foram [1], [2], [3], [6], [7], [8], [9], [10], [11] e [14].

### 2.1 Elementos de Biologia

O Planeta Terra é habitado por organismos vivos extraordinariamente diversos e hoje, sabemos que todos os seres vivos, são formados por células e que todas essas unidades da matéria viva, segundo [1], 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 menor unidade morfofisiológica de um ser vivo, além de que, uma única célula é o veículo para toda a informação hereditária que define cada espécie. Elas podem ser apresentadas sob uma variedade de formas e tamanhos. Há aquelas que se movem rapidamente e outras que têm estruturas que se alteram. Existem organismos vivos que possuem uma única célula (unicelulares) e aqueles que são formados por mais de uma célula (pluricelulares). Além disso, elas podem ser diferenciadas em células eucariotas e células procariotas.

Os nucleotídeos são unidades de moléculas fundamentais dos ácidos nucleicos e que estão presentes no metabolismo como transportador de energia. Em todas as células vivas se encontram os ácidos nucleicos, que têm a função de conter, transmitir e traduzir informações genéticas dos seres vivos. Todos os organismos vivos apresentam ácidos nucleicos na forma de ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA).

O DNA é um polímero, ou seja, uma longa cadeia de nucleotídeos. Sua função é armazenar informação genética dos seres vivos. Conforme Alberts et al. (2010), o DNA é formado por uma dupla hélice com giro para a direita e parece uma escada de caracol molecular. 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. Já o RNA é um polímero de fita simples e 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) em sua composição e participa na tradução do DNA em proteína.

O DNA e o RNA são os responsáveis pela síntese proteica. Para que esse processo ocorra, é necessário que a célula transcreva e traduza a informação do código genético. 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.

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. Além disso, os aminoácidos podem ser classificados como hidrofóbicos (evitam o contato com a água) ou hidrofílicos (possuem afinidade com a água). Também são classificados pela carga (ácido, neutro, básico), estrutura (cíclico, acíclico) tamanho (pequeno, médio, grande), aromático e alifático.

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 no Tabela 1, 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. Como os aminoácidos essenciais não podem ser sintetizados pelo corpo humano, temos de consegui-los por meio dos alimentos.

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

Tabela 1: Hidropacidade dos aminoácidos.

Hidrofilico (Polar)	Hidrofóbico (Apolar)
Asparigina (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: [11, p. 26].

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. 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. A Tabela 2 apresenta as 64 combinações possíveis compostas pelas quatro bases nitrogenadas e seus respectivos aminoácidos gerados. Nesta Tabela, os aminoácidos estão representados por abreviações, que são correspondentes à Tabela 1.

Tabela 2: Combinações possíveis compostas pelas quatro bases nitrogenadas e seus respectivos aminoácidos.

primeira posição	segunda posição				terceira posição
	A	C	G	U	
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: [11, p. 31].

A seguir serão apresentadas algumas características 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 um códon, sendo eles a metionina e o triptofano. Já os outros aminoácidos codificam mais de um códon, ou seja, três aminoácidos tem 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 possui seis códons possíveis, 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 as 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 apresentam 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.

## 2.2 Elementos de Álgebra

**Definição 1.** Um grupo é um par  $(G, *)$ , onde  $G$  é um conjunto não vazio e  $*$  é uma operação binária sobre  $G$ , satisfazendo: Associatividade:  $a * (b * c) = (a * b) * c$ ,

quaisquer que sejam  $a, b, c \in G$ ; *Existência de elemento neutro*: existe um elemento neutro  $e \in G$  tal que  $e * a = a * e = a$ , qualquer que seja  $a \in G$ ; *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 de  $G$ , então esse grupo é chamado de *grupo abeliano ou comutativo*.

**Definição 2.** Um grupo em que seu conjunto é finito, é denominado grupo finito, caso contrário, é denominado grupo infinito. Além disso, o número de elementos desse grupo é chamado ordem do grupo.

**Definição 3.** Um conjunto não vazio  $A$  e um par de operações sobre  $A$ , 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 as seguintes propriedades:

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

*Para a multiplicação*: Associatividade:  $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ , quaisquer que sejam  $a, b, c \in A$ ; A multiplicação é distributiva em relação a adição: se  $a, b, c \in A$ , então  $a \cdot (b + c) = a \cdot b + a \cdot c$  e  $(b + c) \cdot a = b \cdot a + c \cdot a$ .

**Definição 4.** Um anel  $(A, +, \cdot)$ , onde a operação  $\cdot$  é comutativa é denominado de anel comutativo.

**Definição 5.** Seja  $F$  um conjunto de elementos sobre o qual suas operações binárias, adição “+” e a multiplicação “ $\cdot$ ”, são definidas.  $F$ , junto com as duas operações binárias, é um corpo se as seguintes condições são satisfeitas:

1.  $F$  é um grupo comutativo sob  $+$ . O elemento identidade é o 0 (zero).
2. O conjunto dos elementos não nulos em  $F$  é um grupo comutativo sob  $\cdot$ . O elemento identidade é o 1 (um).
3. A multiplicação é distributiva sob adição, isto é, para quaisquer  $a, b, c$  em  $F$ ,

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

**Definição 6.** Um reticulado é um conjunto parcialmente ordenado no qual todo par de elementos do conjunto possui simultaneamente soma e produto.

$(A, \preceq)$  é um conjunto parcialmente ordenado (CPO) se  $\preceq$  for uma relação de ordem parcial definida num conjunto não vazio  $A$ . Uma relação  $\preceq$  em  $A$  diz-se de ordem parcial (ROP) se satisfaz as propriedades: Reflexiva:  $x \preceq x$  para todo  $x \in A$ ; Transitiva:  $x \preceq y$  e  $y \preceq z$  implica  $x \preceq z$ ; Anti-simétrica:  $x \preceq y$  e  $y \preceq x$  implica  $x = y$ .

### Reticulado como Álgebra: Conectivos $\wedge$ e $\vee$

Considere os valores-verdade  $F$  e  $V$  (Falso e Verdadeiro), bem como os conectivos lógicos  $\wedge$  e  $\vee$ . Logo, a estrutura  $\langle \{F, V\}, \wedge, \vee \rangle$  constitui um reticulado. 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 associativas, comutativas e as operações são fechadas) e, que os conectivos  $\wedge$  e  $\vee$  satisfazem à propriedade absorção.

**Definição 7.** Um reticulado booleano  $B(X)$  é um conjunto parcialmente ordenado de elementos com as seguintes propriedades:

1.  $B(X)$  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:

(a) Interseção:  $0 \wedge \alpha = 0$  e  $1 \wedge \alpha = \alpha$ .

(b) União:  $0 \vee \alpha = \alpha$  e  $1 \vee \alpha = 1$ .

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

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

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

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

**Definição 8.** 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.

## 2.3 Códigos Corretores de Erros

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, segundo [11]. 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.

Os principais elementos de um sistema de comunicação são apresentados na Figura 1.

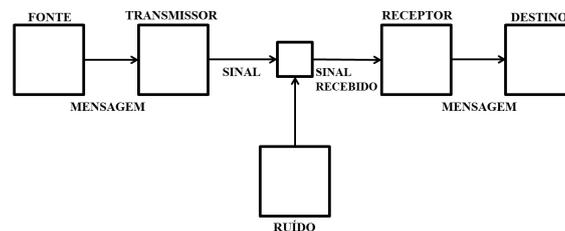


Figura 1: Elementos importantes em um sistema de comunicação.

Fonte: Autor.

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 9.** *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 10.** *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).*

### 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. Uma sequência de bits ou símbolos é segmentada 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**,  $r$ .

$$r = c \oplus e.$$

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 11.** *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 12.** *A distância de Hamming entre dois vetores códigos,  $v$  e  $x$ , cuja notação é  $d(v, x)$ , é definido como sendo o número de posições em que os dígitos dos dois vetores 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  $v \oplus x$ .*

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

## 3 Desenvolvimento

Para o estudo do Rotulamento A, foi utilizada a permutação 0132, para o reticulado primal. Como consequência, utilizando a complementaridade biológica, sendo que adenina

se emparelha com a uracila e citosina se emparelha com a guanina, 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 a permutação escolhida primal, foi estabelecida a seguinte associação, levando em consideração a ordem das bases apresentadas anteriormente e a associação com elementos do conjunto  $\mathbb{Z}_2 \times \mathbb{Z}_2$ :  $A - 00, C - 10, G - 01$  e  $U - 11$ . Estabelecida a associação da permutação primal, obtemos a associação da permutação dual, sendo  $\mathbb{Z}_2 \times \mathbb{Z}_2$ :  $A - 11, C - 01, G - 10$  e  $U - 00$ .

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 obtidas das permutações escolhidas, já mencionadas. Para a realização dessas operações, utiliza-se os dados apresentados nas Tabelas 3 e 4.

 Tabela 3: Operação com o conectivo  $\wedge$ .

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

Fonte: Autor.

 Tabela 4: Operação com o conectivo  $\vee$ .

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

Fonte: Autor.

Utilizando as operações da álgebra booleana e a permutação 0132 do rotulamento  $A$ , obtemos os dados das Tabelas 5 e 6.

 Tabela 5: Operação Primal  $\wedge$  (e).

$\wedge$	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: Autor.

$\wedge$	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: Autor.

 Tabela 6: Operação Primal  $\vee$  (ou).

$\vee$	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: Autor.

$\vee$	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: Autor.

Utilizando as operações da álgebra booleana e a permutação 2310 do rotulamento  $A$ , obtemos os dados das Tabelas 7 e 8.

O reticulado booleano primal será obtido a partir da Tabela 6. Observe que:

Tabela 7: Operação Dual  $\wedge$  (e).

$\wedge$	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: Autor.

$\wedge$	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: Autor.

 Tabela 8: Operação Dual  $\vee$  (ou).

$\vee$	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: Autor.

$\vee$	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: Autor.

- A base A (adenina) se liga as bases G (guanina), C (citosina) e U (uracila).
- A base G (guanina) se liga a ela mesma e a base U (uracila).
- A base C (citosina) se liga a ela mesma e a 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 8. Assim, obtemos as representações nas Figuras 2a e 2b.

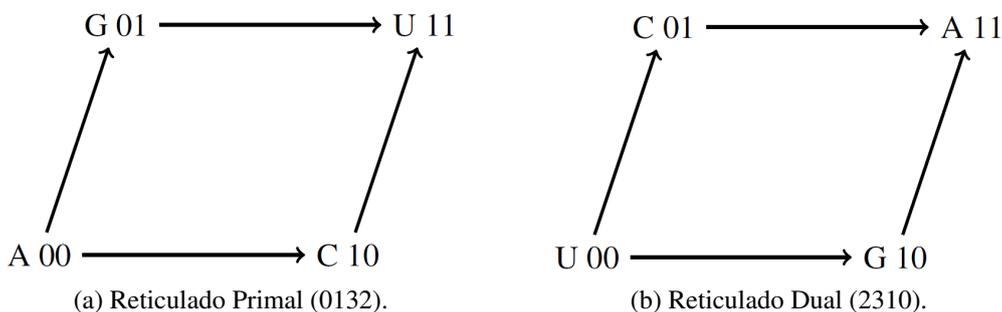


Figura 2: Reticulados booleanos primal e dual associados ao rotulamento A.

Fonte: Autor.

Vamos agora apresentar a construção dos diagramas de Hasse primal e dual, utilizando as associações já estabelecidas. 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 hidropaticidade e outras características. Este, é composto por 64 códons, dispostos em 7 linhas. Cada linha é apresentada da seguinte forma:

Na **1ª linha** tem-se o códon que apresenta seu elemento máximo, atribuído por 11.

Na **7ª linha** tem-se o códon que apresenta elemento mínimo, atribuído por 00.

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 complementariedade biológica:  $A \leftrightarrow U$  e  $C \leftrightarrow G$ .

A **3ª linha** e a **5ª linha** são constituídas por quinze códons cada uma, e segue o mesmo procedimento da 2ª linha e 6ª linha. 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.

Além disso, 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 em sua representação vetorial, como pode ser observado nas Figuras 3 e 4.

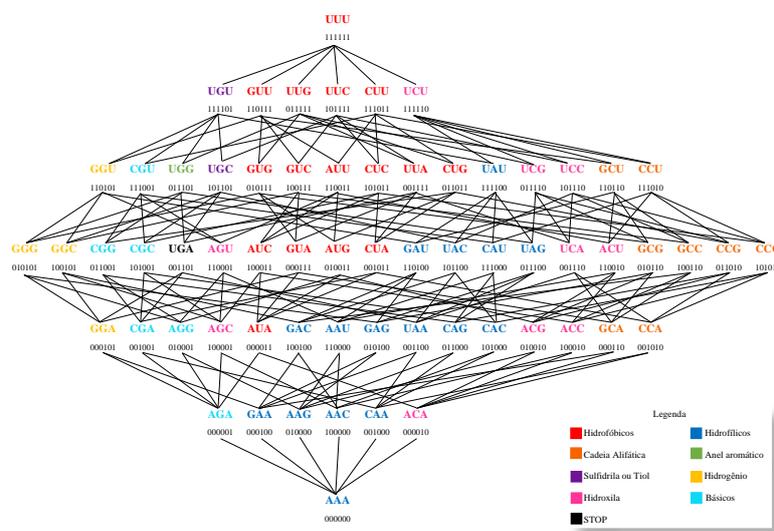


Figura 3: Diagrama de Hasse primal do rotulamento A para a permutação 0132.

Fonte: Autor.

Observe que os diagramas das Figuras 3 e 4, são construídos por um determinado número de arestas, que foi proposto por [14]. Sua construção não é feita de maneira arbitrária e cada aresta determina a distância de Hamming entre os códons.

A correspondência entre a ordem do códon e as propriedades dos aminoácidos é refletida no diagrama de Hasse do código genético.

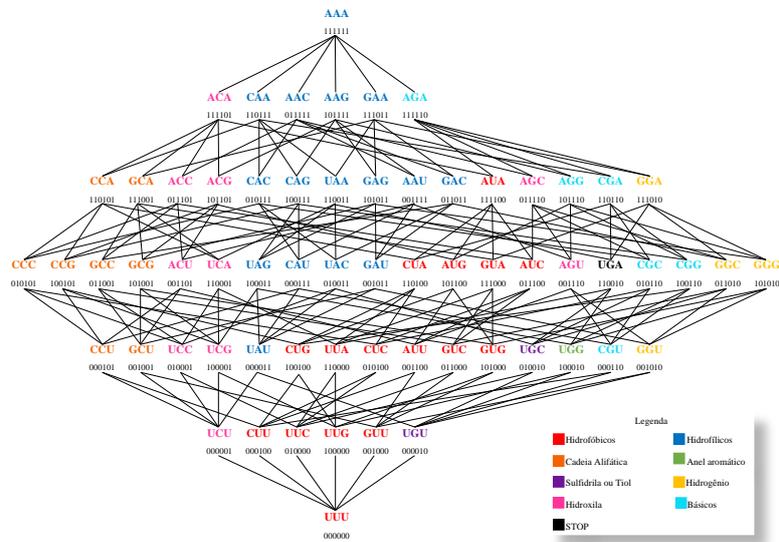


Figura 4: Diagrama de Hasse dual do rotulamento A para a permutação 2310.  
Fonte: Autor.

Nas Figuras 3 e 4 observa-se uma simetria na distribuição dos códons. A construção dos diagramas de Hasse para as permutações 0132 e 2310 apoiou-se na complementariedade 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 dos diagramas, tanto na Figura 3 quanto na Figura 4. Além disso, o códon que codifica um aminoácido hidrofóbico é sempre complementar a um códon que codifica um aminoácido hidrofílico, ou seja, observe na Figura 3, 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'.

No diagrama de Hasse primal (Figura 3), nas laterais no diagrama, podemos encontrar os aminoácidos hidrogênio, básicos, hidroxila e cadeia alifática. Além disso, na lateral esquerda, podemos encontrar os aminoácidos sulfidríla 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 (Figura 3), 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 codificam 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.

No diagrama de Hasse dual (Figura 4), o códon UGG que é um anel aromático que codifica o aminoácido triptofano, 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 codificam o aminoácido prolina. No diagrama eles estão próximos e estão presentes na lateral esquerda.

Conforme a proposta de [14], serão apresentadas a seguir as médias das distâncias de Hamming entre os pares de aminoácidos, como as distâncias médias entre os seus respectivos códons (Tabela 9). Foram calculadas as distâncias de Hamming entre cada um dos códons que codificam determinado aminoácido, em seguida, foi realizada a média entre essas distâncias.

A distância de Hamming entre os códons é o número de posições (base) em que os códons se diferem, por meio da associação estabelecida do conjunto  $\mathbb{Z}_2 \times \mathbb{Z}_2$ . 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 [14]. Um fato a ser destacado, é que os resultados expressos na Tabela 9, para a permutação 0132 do Rotulamento A, são iguais ao caso proposto por [14], que utilizou outra permutação no mapeamento do código genético.

Tabela 9: Médias das distâncias de Hamming entre pares de aminoácidos.

*	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	<b>4,33</b>	2,5	3,5	2,5	3,5	<b>4,5</b>	2,5	3,5	<b>4</b>	<b>5</b>	<b>4</b>	1,33
C	2	1,5	0,5	2,5	2,83	3	2,83	1,5	<b>4,5</b>	3,83	3,5	2,5	1,5	<b>4,5</b>	3,5	3,5	2,5	<b>4</b>	<b>5</b>	<b>4</b>	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	<b>4,67</b>	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	<b>4</b>	<b>4</b>	<b>4</b>	<b>5</b>	<b>5</b>	2	3	<b>4</b>	3,67
L	3,67	2,5	2,83	2,61	2,94	2,67	1,39	1,83	2,17	2,39	<b>4,5</b>	<b>4,83</b>	3,83	<b>4,17</b>	<b>4,5</b>	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	<b>4,5</b>	3,5	2,5	<b>5,5</b>	<b>4,5</b>	<b>4,5</b>	3,5	3	<b>4</b>	3	3,17
M	3	3	<b>4,5</b>	2,5	3,5	2	2,17	3,5	0	1,33	3,5	<b>4,5</b>	<b>5,5</b>	2,5	3,5	3,5	<b>4,5</b>	3	2	3	<b>4,33</b>
I	3	<b>4,33</b>	3,83	2,72	3,28	2	2,39	2,83	1,33	0,89	<b>4,17</b>	3,83	<b>4,83</b>	3,17	2,83	<b>4,17</b>	3,83	3	2	3	<b>4,78</b>
E	2	2,5	3,5	3,5	3,17	3	<b>4,5</b>	<b>4,5</b>	3,5	<b>4,17</b>	0,5	1,5	2,5	1,5	2,5	2,5	3,5	2	3	<b>4</b>	1,83
D	2	3,5	2,5	3,83	2,83	3	<b>4,83</b>	3,5	<b>4,5</b>	3,83	1,5	0,5	1,5	2,5	1,5	3,5	2,5	2	3	<b>4</b>	2,83
Y	3	2,5	1,5	3,5	2,5	<b>4</b>	3,83	2,5	<b>5,5</b>	<b>4,83</b>	2,5	1,5	0,5	3,5	2,5	2,5	1,5	3	<b>4</b>	3	1,83
K	3	3,5	<b>4,5</b>	2,5	3,5	<b>4</b>	<b>4,17</b>	<b>5,5</b>	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	<b>4,5</b>	3,5	2,83	3,17	<b>4</b>	<b>4,5</b>	<b>4,5</b>	3,5	2,83	2,5	1,5	2,5	1,5	0,5	2,5	1,5	3	2	3	3,83
Q	<b>4</b>	2,5	3,5	2,17	3,17	<b>5</b>	3,17	<b>4,5</b>	3,5	<b>4,17</b>	2,5	3,5	2,5	1,5	2,5	0,5	1,5	<b>4</b>	3	2	1,83
H	<b>4</b>	3,5	2,5	2,5	2,83	<b>5</b>	3,5	3,5	<b>4,5</b>	3,83	3,5	2,5	1,5	2,5	1,5	1,5	0,5	<b>4</b>	3	2	2,83
A	3	<b>4</b>	<b>4</b>	<b>4,67</b>	2,67	2	3,67	3	3	3	2	2	3	3	3	<b>4</b>	<b>4</b>	1	2	3	3,33
T	<b>4</b>	<b>5</b>	<b>5</b>	3,67	3	3	3,33	<b>4</b>	2	2	3	3	<b>4</b>	2	2	3	3	2	1	2	<b>4,33</b>
P	<b>5</b>	<b>4</b>	<b>4</b>	3,33	2,67	<b>4</b>	2,33	3	3	3	<b>4</b>	<b>4</b>	3	3	3	2	2	3	2	1	3,33
STOP	2,67	1,33	2,17	2,83	2,94	3,67	3,17	3,17	<b>4,33</b>	<b>4,78</b>	1,83	2,83	1,83	2,83	3,83	1,83	2,83	3,33	<b>4,33</b>	3,33	0,89

Fonte: Autor.

\*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 - Argina, S - Serina, T - Treonina, V - Valina, W - Triptofano, Y - Tirosina.

Segundo [14], quando se obtém valores maiores nas médias das distâncias de Ham-

ming 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 de distâncias de Hamming. Por exemplo, as distâncias médias entre o códon que codifica o aminoácido glicina (G) e os códons que codificam os aminoácidos glutamina (Q), histidina (H), triptofano (W) e prolina (P) foram maiores ou iguais a 4. O códon associado ao triptofano e os códons 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 códons 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 9.

Com base nas características de cada aminoácido codificado pelos seus respectivos códons, vamos analisar algumas diferenças e semelhanças biológicas acerca dos cálculos obtidos na Tabela 9.

O códon UGG codifica o aminoácido triptofano (W) e os códons AUU, AUC e AUA codificam o aminoácido isoleucina (I). Na Tabela 9, 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. Como suas características são bem diferentes, sua distância pode ser observada nos Diagramas de Hasse (Figuras 3 e 4), nos quais os códons aparecem distantes uns dos outros. Os códons AAU e AAC codificam o aminoácido asparagina (N), que é hidrofílico. Na Tabela 9, a distância média entre esses códons e o aminoácido triptofano (W) é igual a 4,5. Como são bem diferentes, nota-se nos Diagramas de Hasse (Figuras 3 e 4), que os códons estão distantes, além de estarem separados. 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 3 e 4, sua distância pode ser observada no Diagrama de Hasse, onde, 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. Como são aminoácidos diferentes entre si, nas Figuras 3 e 4, sua distância pode ser observada nos Diagramas de Hasse. 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 9 é igual a 4. Como suas características são bem diferentes, sua distância pode ser observada nos Diagramas de Hasse (Figuras 3 e 4), onde os códons aparecem distantes uns dos outros.

Os códons GGU, GGC, GGA e GGG codificam o aminoácido glicina (G), que possui como característica de hidrogênio e o códons CAA e CAG codificam o aminoácido gluta-

mina (Q) que é hidrofílico. Pode-se notar que a média da distância entre os códons desses aminoácidos na Tabela 9 é igual a 4. Logo, nas Figuras 3 e 4, essa distância pode ser visualizada nos Diagramas de Hasse, nos quais, os códons estão separados de acordo com suas características e estão distantes uns dos outros. 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 9). Devido as diferenças em suas características, nas Figuras 3 e 4, pode-se observar que os códons estão distantes um do outro, nos Diagramas de Hasse. 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 9 é igual a 4. Então, nas Figuras 3 e 4, essa distância pode ser visualizada no Diagrama de Hasse, onde os códons estão separadas de acordo com suas características e estão distantes um do outro. 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 9 é igual a 5. Como suas características são diferentes, sua distância pode ser observada nos Diagramas de Hasse (Figuras 3 e 4), onde os códons aparecem distantes um do outro.

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 9 é igual a 4,67. Logo, nas Figuras 3 e 4, essa distância pode ser observada no Diagrama de Hasse, onde os códons estão separados de acordo com suas características e estão distantes uns dos outros.

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 9, 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 3 e 4) podemos observar que os códons estão próximos um do outro, justificando o fato de que na Tabela 9, a distância média entre os aminoácidos metionina (M) e isoleucina (I) é 1,33.

Pode-se observar os outros casos na Tabela 9 com os Diagramas (Figuras 3 e 4) cons-

truídos. Além disso, se fizermos a análise de todos os casos, nos quais, as distâncias médias entre as distâncias de Hamming foram maiores ou iguais a 4, as suas distâncias estão refletidas no Diagrama de Hasse, com base nas diferenças biológicas entre os códons.

## 4 Considerações Finais

Com base nas construções realizadas neste estudo, percebe-se uma interessante caracterização algébrica associada ao mapeamento do código genético, no qual foi possível uma análise biológica para as permutações 0132 e 2310 do rotulamento  $A$ , baseados nas características dos aminoácidos. Além disso, pode-se observar com a Tabela 9, que o diagrama de Hasse reflete de forma significativa as características físico-químicas dos aminoácidos baseados nos cálculos das distâncias de Hamming, de forma a poder utilizar esses resultados em estudos de fenômenos mutacionais.

## Agradecimentos

À Universidade Federal de Alfenas - UNIFAL, ao Programa de Pós-Graduação em Estatística Aplicada e Biometria e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Capes.

## Referências

- [1] ALBERTS, B. et al. **Biologia Molecular da Célula**. 5ª edição. Porto Alegre: Art-Med, 2010.
- [2] ALBERTS, B. et al. **Biologia Molecular da Célula**. 6ª edição. Porto Alegre: Art-Med, 2017.
- [3] DOMINGUES, H. H.; IEZZI, G. **Álgebra Moderna**. 4ª edição. São Paulo: Atual, 2003.
- [4] FARIA, L. C. B. D.; PALAZZO JÚNIOR, R. **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.
- [5] FERNANDES, R. S; OLIVEIRA, A. J. **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, Alfenas, 2018.

- [6] GERÔNIMO, J. R.; FRANCO, V. S. **Fundamentos da Matemática: uma introdução à lógica matemática, teoria dos conjuntos, relações e funções.** 2ª edição. Maringá: Eduem, 2008.
- [7] HEFEZ, A.; FERNANDEZ, C. S. **Introdução à Álgebra Linear.** 1ª edição. Rio de Janeiro: SBM, 2012.
- [8] HIB, J.; ROBERTIS, E. M. F. **Bases da biologia celular e molecular.** Tradução de Antônio Francisco Dieb Paulo. 4ª edição. Rio de Janeiro: Guanabara Koogan, 2006.
- [9] LIN, S.; COSTELO JR, D. J. **Error Control Coding: Fundamentals and Applications.** Englewood Cliffs: Prentice Hall, 1983.
- [10] MACWILLIAMS, F. J.; SLOANE, N. J. A. **The theory of Error-Correcting Codes.** Elsevier Science B.V., 1981.
- [11] OLIVEIRA, A. J.; PALAZZO JÚNIOR, R. **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.
- [12] ROCHA, A. S. L.; PALAZZO JÚNIOR, R. **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.
- [13] 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.
- [14] SÁNCHEZ, R.; MORGADO, E.; GRAU, R. The genetic Code Boolean Lattice. **MATCH Commun. Math.Comp.Chem**, v.52, n. 52, p. 29-46, 2004.