

HIPERCUBOS BOOLEANOS APLICADOS NA CARACTERIZAÇÃO DE PROPRIEDADES DOS AMINOÁCIDOS ASSOCIADOS AO ROTULAMENTO A DO CÓDIGO GENÉTICO

Bianca Lapa Ribeiro

Universidade Federal de Alfenas

bianca.ribeiro@sou.unifal-mg.edu.br

Anderson José de Oliveira

Universidade Federal de Alfenas

anderson.oliveira@unifal-mg.edu.br

Comunicado por Z. Ramos

Resumo

A modelagem matemática do código genético possibilita, dentre outros aspectos, a análise, a interpretação e a caracterização de propriedades associadas aos aminoácidos e possíveis interferências em diversas situações, como o caso das mutações genéticas. Diagramas de Hasse, códigos de Gray e hipercubos booleanos representam algumas ferramentas matemáticas que podem ser empregadas nesse estudo. O código genético consiste na associação das trincas encontradas no RNA mensageiro, formadas pelas bases nitrogenadas e os aminoácidos que estão nas proteínas. A partir do mapeamento das bases nitrogenadas com a estrutura algébrica do anel $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, é possível obter 24 permutações, organizadas em três rotulamentos (A, B e C). Ademais, o código genético pode ser representado por um hipercubo booleano 6-dimensional, construído a partir da tabela do código de Gray. O objetivo deste trabalho é apresentar uma caracterização do código genético por meio do código de Gray e do hipercubo booleano, utilizando as permutações 1203 e 3021 do rotulamento A do código genético, a fim de analisar propriedades físico-químicas dos aminoácidos. A metodologia adotada neste trabalho baseia-se em uma natureza qualitativa e quantitativa, visando um estudo descritivo e abordagem aplicada. As construções realizadas possibilitam uma caracterização das propriedades associadas aos aminoácidos do código genético, permitindo uma classificação dos códons e possibilidades de futuras aplicações em fenômenos biológicos, como as mutações genéticas. Com isso, é possível compreender as conexões existentes entre Biologia, Álgebra e Geometria, além da análise do código genético.

Palavras-chave: Biologia. Códigos de Gray. Diagramas de Hasse.

Abstract

Mathematical modeling of the genetic code enables, among other aspects, the analysis, interpretation and characterization of properties associated with amino acids and possible interferences in various situations, such as genetic mutations. Hasse diagrams, Gray codes and Boolean hypercubes represent some mathematical tools that can be used in this study. The genetic code consists of the association of triplets found in messenger RNA, formed by nitrogenous bases and amino acids found in proteins. By mapping the nitrogenous bases with the algebraic structure of the ring $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, it is possible to obtain 24 permutations, organized into three labels (A, B and C). Furthermore, the genetic code can be represented by a 6-dimensional Boolean hypercube, constructed from the Gray code table. The aim of this paper is to present a characterization of the genetic code through the Gray code and the Boolean hypercube, using the permutations 1203 and 3021 of the A labeling of the genetic code, in order to analyze physicochemical properties of amino acids. The methodology adopted in this work is based on a qualitative and quantitative nature, aiming at a descriptive study and an applied approach. The constructions performed allow a characterization of the properties associated with the amino acids of the genetic code, allowing a classification of the codons and possibilities of future applications in biological phenomena, such as genetic mutations. With this, it is possible to understand the existing connections between Biology, Algebra and Geometry, in addition to the analysis of the genetic code.

Keywords: Biology. Gray Codes. Hasse Diagrams.

1 Introdução

No processo de modelagem matemática do código genético podem ser utilizadas diversas estruturas, com o objetivo de analisar, interpretar e caracterizar propriedades associadas aos aminoácidos, além de possíveis interferências em diversas situações. Algumas das abordagens utilizadas nesse processo incluem os chamados diagramas de Hasse, como apresentado em [3]. Além disso, por meio do código de Gray e do hiper-cubo booleano, diversas propriedades e características podem ser analisadas, conforme mostram Jimenez-Montano e La Mora-Basanez [8].

O código genético representa a associação de trinças encontradas no RNA mensageiro, formadas pelas bases nitrogenadas, e os aminoácidos que estão nas estruturas das proteínas. As trinças de bases nitrogenadas, mencionadas anteriormente, são denominadas códons. Existem 64 possíveis trinças, agrupadas três-a-três, que correspondem a 20 aminoácidos, os quais são as unidades básicas que formam as proteínas. A cor-

respondência de um determinado aminoácido com um códon se dá por meio do código genético, conforme apresentado em [1].

Em [2] é apresentada uma modelagem da estrutura do DNA, por meio de um mapeamento do alfabeto do código genético, que está relacionado com o conjunto de bases nitrogenadas $N = \{A, C, G, T/U\}$, (A - adenina, C - citosina, G - guanina e T/U - timina/uracila) e o alfabeto 4-ário da estrutura de anel, denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. As associações do alfabeto biológico com o alfabeto matemático possibilitam a identificação de 24 permutações, que são separadas em três rotulamentos, os quais tem como finalidade determinar qual a melhor associação de cada um dos símbolos do conjunto N com os correspondentes símbolos do conjunto \mathbb{Z}_4 e vice-versa.

As 24 permutações mencionadas anteriormente são apresentadas na Figura 1. Cada permutação possui suas próprias características, permitindo organizar esse mapeamento em três conjuntos, chamados de rotulamentos A, B e C (Figura 2), cada um contendo 8 permutações, conforme a caracterização geométrica associada a cada uma das permutações, conforme construção apresentada em Faria (2011).

Figura 1: Possibilidades de permutações entre os elementos dos conjuntos N e \mathbb{Z}_4 .

$N! = 4! = 24$ possibilidades de rotulamento para $\mathbb{Z}_4 = \{0, 1, 2, 3\}$

$\begin{bmatrix} A & C & G & T \\ 0 & 1 & 3 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 1 & 3 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 1 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 1 & 0 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 2 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 0 & 1 & 3 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 0 & 3 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 3 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 3 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 3 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 2 & 3 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 0 & 3 & 1 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 1 & 0 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 0 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 0 & 3 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 0 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 3 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 1 & 0 & 2 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 1 & 2 & 0 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 2 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 2 & 3 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 2 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 3 & 2 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 1 & 2 & 0 \end{bmatrix}$

Fonte: [2].

Figura 2: Rotulamentos A, B e C.

Rotulamento A		Rotulamento B		Rotulamento C	
$\begin{bmatrix} A & C & G & T \\ 0 & 1 & 3 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 1 & 3 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 1 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 1 & 0 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 2 & 1 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 0 & 1 & 3 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 0 & 3 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 3 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 3 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 3 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 2 & 3 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 2 & 0 & 3 & 1 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 1 & 0 & 2 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 0 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 0 & 3 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 0 & 1 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 3 & 0 & 2 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 1 & 0 & 2 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 1 & 2 & 0 & 3 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 2 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 2 & 3 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 2 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 3 & 2 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 3 & 1 & 2 & 0 \end{bmatrix}$

Fonte: [2].

Outra importante contribuição ao processo de modelagem e análise do código genético é apresentada em [6], por meio de um hipercubo booleano e sua relação com um código binário, mostrando construções para as dimensões de 1 a 4 do hipercubo. Em [8] é apresentada uma representação do código genético através de um hipercubo booleano

6-dimensional, construído a partir da tabela do código de Gray, em que os vértices de cada cubo são representados pelos códons do código genético.

O hipercubo representa simultaneamente todo o conjunto de códons e mantém o controle de quais códons são adjacentes. A construção do hipercubo booleano é realizada a partir de todas as ligações, feitas por meio da identificação da diferença de um *bit* existente de um vértice para o outro, desde a dimensão 1 até a dimensão 6.

A partir das informações apresentadas anteriormente e importantes questões ainda em aberto acerca da modelagem matemática do código genético, o objetivo deste trabalho é apresentar a caracterização do código genético por meio da construção e análise dos hipercubos booleanos, utilizando as permutações 1203 e 3021 do rotulamento A.

2 Referencial teórico

Nesta seção serão apresentados os principais elementos teóricos que darão embasamento para a obtenção dos resultados deste trabalho.

2.1 Biologia molecular

De acordo com Alberts *et al.* [1], a célula é uma unidade de matéria viva que forma o corpo de todos os seres vivos, sendo capaz, individualmente, de obter energia, crescer e se reproduzir. A célula possui três componentes básicos: a membrana plasmática, o citoplasma e o material genético. A membrana plasmática envolve a célula, delimitando o citoplasma - um material fluido que contém organelas e outras estruturas celulares - e protege o material genético, que carrega as informações genéticas dos seres vivos.

Ainda de acordo com Alberts *et al.* [1], os nucleotídeos são uma repetição de moléculas menores, compostos por um grupo de fosfato (ácido fosfórico), uma molécula de açúcar e uma base nitrogenada. Estas bases são divididas em bases púricas, que são adenina(A) e guanina(G); bases pirimídicas, que são citosina(C), timina (T) e uracila(U). Desse modo, podem ser formados dois tipos de ácidos nucleicos: ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA).

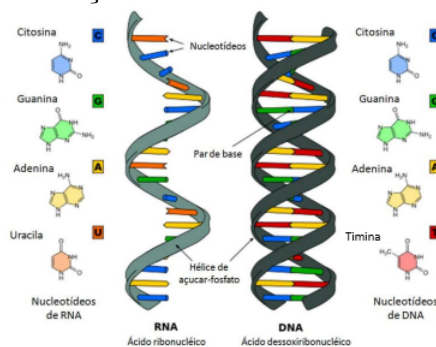
De acordo com Hepp e Nonohay [7], o DNA funciona como um banco de informações genéticas, uma espécie de código, que transfere essas informações, com o objetivo de garantir a integridade da informação genômica e funcionando como molde para a síntese da molécula de RNA. A molécula de DNA é formada por uma fita dupla de vários nucleotídeos (dupla hélice), que são formados por um grupo de fosfato, uma pentose (desoxirribose, que é um açúcar composto por cinco átomos de carbono, formando uma cadeia fechada) e uma base nitrogenada. As quatro bases nitrogenadas são: adenina

e timina (purinas - maiores), citosina e guanina (pirimidinas - menores). A adenina liga-se apenas à timina e a citosina apenas à guanina.

A molécula de RNA (ácido ribonucleico) é formada a partir da molécula de DNA em um processo chamado de transcrição. Essa molécula apresenta informações com as quais é possível coordenar a produção de proteínas. Assim sendo, o RNA também participa do fluxo de informações genéticas pelos indivíduos. O RNA é uma molécula composta por uma única fita de nucleotídeos ligados entre si. A pentose do RNA é sempre a ribose e as quatro bases nitrogenadas são: adenina, guanina, citosina e uracila, sendo a última exclusiva do RNA, (Alberts *et al.*, [1]).

As moléculas de RNA são, geralmente, constituídas de uma única fita que se enrola entre si devido ao emparelhamento entre as bases complementares: a uracila liga-se com a adenina e a guanina liga-se com a citosina. Existem três tipos de RNA: RNA mensageiro (RNAm ou mRNA), RNA transportador (RNAt) e RNA ribossômico (RNAr), conforme apresentado em [7]. A Figura 3 apresenta as diferenças estruturais entre o RNA e o DNA.

Figura 3: Diferenças estruturais entre o DNA e o RNA.



Fonte: [7].

As proteínas são importantes para diversas funções do nosso organismo e a sequência de nucleotídeos que forma uma molécula de DNA representa a informação necessária à produção de todas elas. A síntese proteica pode ser dividida em dois passos principais, a transcrição, que consiste na conversão de DNA em RNA, e a tradução, que consiste na conversão dos nucleotídeos em aminoácidos.

O código genético é a associação das trincas encontradas no RNA mensageiro, formadas pelas bases nitrogenadas e os aminoácidos que estão nas proteínas. As trincas de bases nitrogenadas são denominadas códon, cada um codificando um aminoácido específico da proteína. A Figura 4 apresenta a estrutura do código genético.

Figura 4: O código genético.

O código genético						Os aminoácidos e seus símbolos			Códons
1ª posição (extremidade 5')	2ª posição				3ª posição (extremidade 3')				
↓	U	C	A	G	↓	A	Ala	Alanina	GCA GCC GCG GCU
U	Phe	Ser	Tyr	Cys	U	C	Cys	Cisteína	UGC UGU
	Phe	Ser	Tyr	Cys	C	D	Asp	Ácido aspártico	GAC GAU
	Leu	Ser	PARADA	PARADA	A	E	Glu	Ácido glutâmico	GAA GAG
	Leu	Ser	PARADA	Trp	G	F	Phe	Fenilalanina	UUC UUU
C	Leu	Pro	His	Arg	U	G	Gly	Glicina	GGA GGC GGG GGU
	Leu	Pro	His	Arg	C	H	His	Histidina	CAC CAU
	Leu	Pro	Gln	Arg	A	I	Ile	Isoleucina	AUA AUC AUU
	Leu	Pro	Gln	Arg	G	K	Lys	Lisina	AAA AAG
A	Ile	Thr	Asn	Ser	U	L	Leu	Leucina	UUA UUG CUA CUC CUG CUU
	Ile	Thr	Asn	Ser	C	M	Met	Metionina	AUG
	Ile	Thr	Lys	Arg	A	N	Asn	Aspargina	AAC AAU
	Met	Thr	Lys	Arg	G	P	Pro	Prolina	CCA CCC CCG CCU
G	Val	Ala	Asp	Gly	U	Q	Gln	Glutamina	CAA CAG
	Val	Ala	Asp	Gly	C	R	Arg	Arginina	AGA AGG CGA CGC CGG CGU
	Val	Ala	Glu	Gly	A	S	Ser	Serina	AGC AGU UCA UCC UCG UCU
	Val	Ala	Glu	Gly	G	T	Thr	Treonina	ACA ACC ACG ACU
						V	Val	Valina	GUA GUC GUG GUU
						W	Trp	Triptofano	UGG
						Y	Tyr	Tirosina	UAC UAU

Fonte: [1].

2.2 Álgebra

Os elementos apresentados nesta seção tomam como referência [5], [6], [4] e [9], onde informações adicionais podem ser consultadas.

2.2.1 Reticulados Booleanos

Definição 2.1. Uma relação \preceq em A diz-se de ordem parcial (ROP) se satisfaz as propriedades:

- Reflexiva: $x \preceq x, \forall x \in A$.
- Transitiva: $x \preceq y$ e $y \preceq z \Rightarrow x \preceq z$.
- Anti-simétrica: $x \preceq y$ e $y \preceq x \Rightarrow x = y$.

Além disso, se, para quaisquer $x, y \in A$, tivermos xRy ou yRx , a relação será chamada relação de ordem total sobre A . O conjunto A , nesse caso, será chamado conjunto totalmente ordenado.

Definição 2.2. (A, \preceq) é um conjunto parcialmente ordenado (CPO) se \preceq for uma relação de ordem parcial definida num conjunto não-vazio A .

Definição 2.3. Um reticulado é um conjunto parcialmente ordenado no qual todo par de elementos do conjunto possui simultaneamente soma e produto, onde a soma está associada à disjunção e o produto está associado à conjunção.

2.2.2 Reticulado como Álgebra: Conectivos \wedge e \vee

Sejam 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 $(\{F, V\}, \wedge, \vee)$ constitui um reticulado, ou seja, a conjunção e disjunção (Tabela 1) levam a um reticulado booleano atendendo as suas propriedades.

Tabela 1: Operação com os conectivos \wedge (conjunção) e \vee (disjunção).

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

Fonte: Fernandes e Oliveira (2021).

2.2.3 Diagrama de Hasse

Definição 2.4. Seja (A, \preceq) um conjunto parcialmente ordenado. O “diagrama de Hasse” de A é o grafo cujos vértices são os elementos de A e em que há uma aresta entre a e b (com a abaixo de b) se, e somente se, $a \prec b$ e não existe $c \in A$ tal que $a \prec c \prec b$.

2.2.4 Distância de Hamming

Definição 2.5. A distância de Hamming, denotada por $d_H(x, y)$, é definida como o número de elementos em que dois vetores $x = \{x_1, \dots, x_n\}$ e $y = \{y_1, \dots, y_n\}$ se diferem. 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 $x \oplus y$.

2.2.5 Código de Gray

Conforme apresentado em Jimenez-Montano, La Mora-Basanez e Poeschrl (2002), o código de Gray é uma forma de codificação binária em que dois números consecutivos

diferem em apenas um bit, reduzindo erros em transições de estado, frequentemente usado em circuitos digitais e sistemas de comunicação.

A Figura 5 apresenta a representação do código de Gray. Na primeira e na quarta colunas são mostrados os vetores 6-binários (palavras-código). Na segunda e na quinta colunas aparecem os códons correspondentes. Finalmente, na terceira e na sexta colunas os aminoácidos em notação de letra única. Os dois primeiros dígitos correspondem à primeira base, os dois seguintes à segunda base e os dois últimos à última base, de acordo com a codificação binária das bases, apresentada em Jimenez-Montano, La Mora-Basanez e Poeschel (2002).

Figura 5: Código de Gray.

0	0	0	0	1	1	A	A	C	N	0	0	1	1	1	1	A	C	C	T
0	0	0	0	1	0	A	A	A	N	0	0	1	1	1	0	A	C	C	T
0	0	0	0	0	0	A	A	A	K	0	0	1	1	0	0	A	C	C	T
0	0	0	0	0	1	A	A	G	K	0	0	1	1	0	1	A	C	C	T
1	0	0	0	0	0	U	A	A	G	1	0	1	1	0	0	U	C	C	S
1	0	0	0	0	0	U	A	A	A	1	0	1	1	0	0	U	C	C	S
1	0	0	0	1	0	U	A	A	U	1	0	1	1	1	0	U	C	C	S
1	0	0	0	1	1	U	A	A	C	1	0	1	1	1	1	U	C	C	S
1	1	0	0	1	1	C	A	A	C	1	1	1	1	1	1	U	C	C	P
1	1	0	0	1	0	C	A	A	U	1	1	1	1	1	0	C	C	C	P
1	1	0	0	0	0	C	A	A	A	1	1	1	1	0	0	C	C	C	P
1	1	0	0	0	1	C	A	A	G	1	1	1	1	0	1	G	C	C	P
0	1	0	0	0	0	G	A	A	E	0	1	1	1	0	0	G	C	C	A
0	1	0	0	0	1	G	A	A	U	0	1	1	1	1	0	G	C	C	A
0	1	0	0	1	1	G	A	A	C	0	1	1	1	1	1	G	C	C	A
0	1	1	0	1	1	G	U	U	C	0	1	0	1	1	1	G	G	C	A
0	1	1	0	1	0	G	U	U	A	0	1	0	1	1	0	G	G	C	A
0	1	1	0	0	0	G	U	U	A	0	1	0	1	0	0	G	G	C	A
0	1	1	0	0	1	G	U	U	G	0	1	0	1	0	1	G	G	C	A
1	1	1	0	0	1	C	U	U	G	1	1	0	1	0	1	G	G	C	A
1	1	1	0	0	0	C	U	U	L	1	1	0	1	0	0	C	C	C	A
1	1	1	0	1	0	C	U	U	L	1	1	0	1	1	0	C	C	C	A
1	1	1	0	1	1	C	U	U	C	1	1	0	1	1	1	C	C	C	A
1	0	1	0	1	1	U	U	U	C	1	0	0	1	1	1	U	G	C	A
1	0	1	0	1	0	U	U	U	A	1	0	0	1	1	0	U	G	C	A
1	0	1	0	0	0	U	U	U	A	1	0	0	1	0	0	U	G	C	A
1	0	1	0	0	1	U	U	G	A	1	0	0	1	0	1	U	G	C	A
0	0	1	0	0	1	A	U	G	A	0	0	0	1	0	1	A	G	G	A
0	0	1	0	0	0	A	U	U	A	0	0	0	1	0	0	A	G	G	A
0	0	1	0	1	0	A	U	U	C	0	0	0	1	1	0	A	G	G	A
0	0	1	0	1	1	A	U	C	I	0	0	0	1	1	1	A	G	G	A

Fonte: Jimenez-Montano, La Mora-Basanez e Poeschel (2002).

2.2.6 Hipercubo Booleano

De acordo com Jimenez-Montano, La Mora-Basanez e Poeschel (2002), um hipercubo n -dimensional, denotado por Q_n , consiste em 2^n nós (vértices), cada um endereçado por um número de identificação único de n bits. Existe uma ligação entre dois nós de Q_n se, e somente se, seus endereços de nó diferem em exatamente uma posição de bit. Dois nós em um hipercubo são considerados adjacentes se houver uma ligação presente entre eles.

Nesse sentido, a distância de Hamming entre quaisquer dois nós de cubo é o número de bits que diferem em seus endereços, os quais são representações binárias dos nós dentro de uma estrutura de cubo, em que cada nó é identificado por uma sequência

de bits. O número de saltos necessários para alcançar um nó de outro nó é igual à distância entre os dois nós.

3 Resultados e discussão

Para a construção dos hipercubos booleanos associadas às permutações do rotulamento A do código genético foi escolhida a permutação 1203, para o reticulado primal e, como consequência, a permutação 3021 do rotulamento A foi utilizada para a construção do reticulado dual.

A atribuição $\{0, 1, 2, 3\}$, referente ao anel \mathbb{Z}_4 , é feita em relação à 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). As Tabelas 2 e 3 apresentam as associações estabelecidas pelas permutações 1203 e 3021, respectivamente primal e dual, referentes ao rotulamento A.

Tabela 2: Associação estabelecida pela permutação 1203 primal do rotulamento A.

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

Fonte: dos autores.

Tabela 3: Associação estabelecida pela permutação 3021 dual do rotulamento A.

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

Fonte: dos autores.

Inicialmente serão apresentadas as tabelas das operações primal e dual para as permutações escolhidas do rotulamento A. Em seguida serão construídos os reticulados booleanos, primal e dual. O procedimento a ser utilizado para o caso do rotulamento A pode ser encontrado de forma detalhada em Fernandes e Oliveira (2021). As operações mencionadas anteriormente são apresentadas nas Tabelas 4, 5, 6 e 7.

A Figura 6 apresenta os reticulados booleanos primal e dual relacionados às permutações utilizadas para o rotulamento A.

Tabela 4: Operação primal \wedge (e) do rotulamento A.

\wedge	00	01	10	11	\wedge	G	U	A	C
00	00	00	00	00	G	G	G	G	G
01	00	01	00	01	U	G	U	G	U
10	00	00	10	10	A	G	G	A	A
11	00	01	10	11	C	G	U	A	C

Fonte: dos autores.

Tabela 5: Operação primal \vee (ou) do rotulamento A.

\vee	00	01	10	11	\vee	G	U	A	C
00	00	01	10	11	G	G	U	A	C
01	01	01	11	11	U	U	U	C	C
10	10	11	10	11	A	A	C	A	C
11	11	11	11	11	C	C	C	C	C

Fonte: dos autores.

Tabela 6: Operação dual \wedge (e) do rotulamento A.

\wedge	00	01	10	11	\wedge	C	A	U	G
00	00	00	00	00	C	C	C	C	C
01	00	01	00	01	A	C	A	C	A
10	00	00	10	10	U	C	C	U	U
11	00	01	10	11	G	C	A	U	G

Fonte: dos autores.

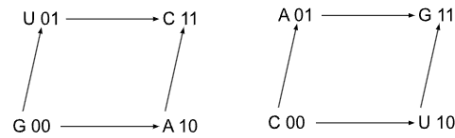
Tabela 7: Operação dual \vee (ou) do rotulamento A.

\vee	00	01	10	11	\vee	C	A	U	G
00	00	01	10	11	C	C	A	U	G
01	01	01	11	11	A	A	A	G	G
10	10	11	10	11	U	U	G	U	G
11	11	11	11	11	G	G	G	G	G

Fonte: dos autores.

A seguir serão apresentados os diagramas de Hasse, tanto para a permutação do caso

Figura 6: Reticulados booleanos primal (1203) e dual (3021) do rotulamento A.



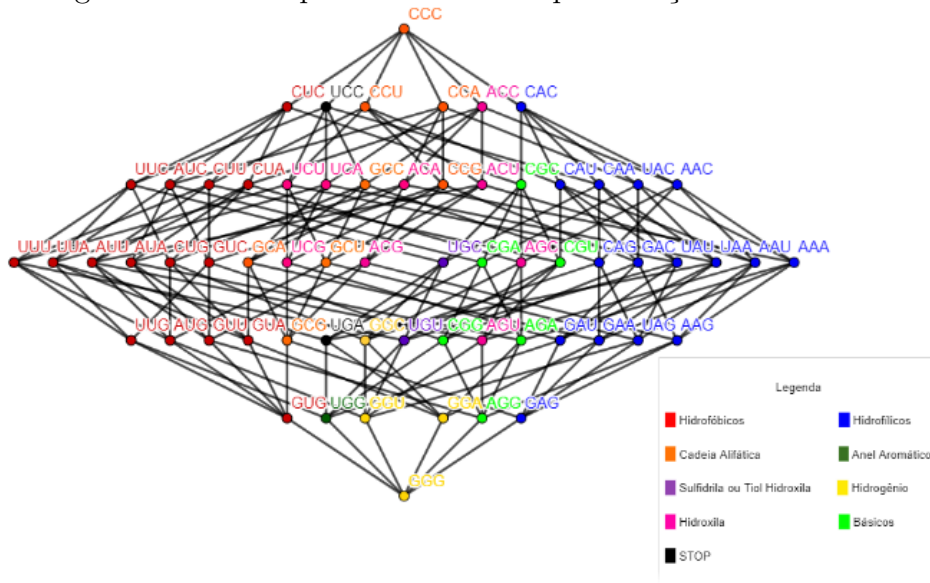
Fonte: Adaptado de Fernandes e Oliveira (2021).

primal, quanto para a do caso dual. O detalhamento para essas construções podem ser encontrados em Fernandes e Oliveira (2021).

3.1 Diagramas de Hasse associados ao rotulamento A

Os diagramas de Hasse, primal e dual, referentes às permutações 1203 e 3021, respectivamente, são apresentados nas Figuras 7 e 8. Nas legendas são apresentadas as características de cada um dos códons, por meio das cores escolhidas.

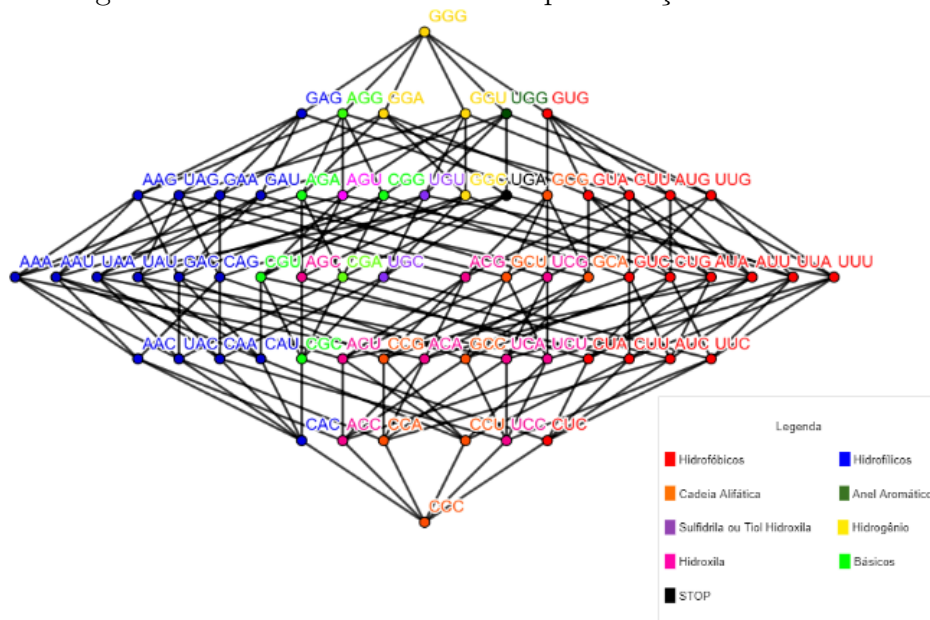
Figura 7: Diagrama de Hasse primal referente à permutação 1203 do rotulamento A.



Fonte: Adaptado de Fernandes e Oliveira (2021).

A seguir serão apresentadas as tabelas do código de Gray referentes às permutações selecionadas do rotulamento A. Esta construção tomou como referência o trabalho de

Figura 8: Diagrama de Hasse dual referente à permutação 3021 do rotulamento A.



Fonte: Adaptado de Fernandes e Oliveira (2021).

Jimenez-Montano, La Mora-Basanez e Poeschel (2002), mas já representa uma contribuição inédita deste trabalho, dadas as características apresentadas na construção e análises advindas, uma vez que nestas construções, além da apresentação da representação matemática, por meio da associação binária a cada um dos códons e seus aminoácidos, foi apresentada uma classificação e separação dos códons, por meio das cores associadas a cada uma das propriedades físico-químicas dos aminoácidos, em consonância com as construções dos diagramas de Hasse apresentados anteriormente.

3.2 Código de Gray - Rotulamento A

A Tabela 8 apresenta o código de Gray associado à permutação 1203 (primal) do rotulamento A do código genético, na qual os dígitos 00 correspondem à base *G*, os dígitos 11 correspondem à base *C*, os dígitos 10 correspondem à base *A* e os dígitos 01 correspondem à base *U*.

A Tabela 9 apresenta o código de Gray associado à permutação 3021 (dual) do rotulamento A do código genético, na qual os dígitos 00 correspondem à base *C*, os dígitos 11 correspondem à base *G*, os dígitos 10 correspondem à base *U* e os dígitos 01 correspondem à base *A*.

Tabela 8: Código de Gray associado à permutação 1203 do código genético.

Binário	Códon	Aminoácido	Binário	Códon	Aminoácido
000011	GGC	glicina	000010	GGA	glicina
000000	GGG	glicina	000001	GGU	glicina
100001	AGU	serina	100000	AGG	arginina
100010	AGA	arginina	100011	AGC	serina
110011	CGC	arginina	110010	CGA	arginina
110000	CGG	arginina	110001	CGU	arginina
010001	UGU	cisteína	010000	UGG	triptofano
010010	UGA	STOP	010011	UGC	cisteína
011011	UAC	tirosina	011010	UAA	STOP
011000	UAG	STOP	011001	UAU	tirosina
111001	CAU	histidina	111000	CAG	glutamina
111010	CAA	glutamina	111011	CAC	histidina
101011	AAC	asparagina	101010	AAA	lisina
101000	AAG	lisina	101001	AAU	asparagina
001001	AUG	aspartato	001000	GAU	glutamato
001010	GAG	glutamato	001011	GAA	aspartato
001111	GCC	alanina	001110	GCA	alanina
001100	GCG	alanina	001101	GCU	alanina
101101	ACU	treoina	101100	ACG	treoina
101110	ACA	treoina	101111	ACC	treoina
111111	CCC	prolina	111110	CCA	prolina
111100	CCG	prolina	111101	CCU	prolina
011101	UCU	serina	011100	UCG	serina
011110	UCA	serina	011111	UCC	serina
010111	UUC	leucina	010110	UUA	fenilalanina
010100	UGG	leucina	010101	UUU	fenilalanina
110101	CUU	isoleucina	110100	CUG	metionina
110110	CUC	isoleucina	110111	CGC	isoleucina
100111	AUC	leucina	100110	AUA	leucina
100100	AUG	leucina	100101	AUU	leucina
000101	GUU	valina	000100	GUG	valina
000110	GUA	valina	000111	GUC	valina

Fonte: do autor.

Tabela 9: Código de Gray associado à permutação 3021 do código genético.

Binário	Códon	Aminoácido	Binário	Códon	Aminoácido
000011	CCG	prolina	000010	CCU	prolina
000000	CCC	prolina	000001	CCA	prolina
100001	UCA	serina	100000	UCC	serina
100010	UCU	serina	100011	UCG	serina
110011	GCG	alanina	110010	GCU	alanina
110000	GCC	alanina	110001	GCA	alanina
010001	ACA	treonina	010000	ACC	treonina
010010	ACU	treonina	010011	ACG	treonina
011011	UAC	leucina	011010	UAA	leucina
011000	UAG	leucina	011001	UAU	leucina
111001	AUG	valina	111000	AUU	valina
111010	AUC	valina	111011	AUA	valina
101011	UUG	leucina	101010	UUU	fenilalanina
101000	UUC	leucina	101001	UUA	fenilalanina
001001	CUA	isoleucina	001000	CUC	isoleucina
001010	CUU	isoleucina	001011	CUG	isoleucina
001111	CGG	arginina	001110	CGU	arginina
001100	CGC	arginina	001101	CGA	arginina
101101	UGA	STOP	101100	UGC	cisteína
101110	UGU	cisteína	101111	UGG	triptofano
111111	GGG	glicina	111110	GGU	glicina
111100	GGC	glicina	111101	GGA	glicina
011101	AGA	arginina	011100	AGC	serina
011110	AGU	serina	011111	AGG	arginina
010111	AAG	lisina	010110	AAU	arparagina
010100	AAC	arparagina	010101	AAA	lisina
110101	GAA	glutamato	110100	GAC	aspartato
110110	GAU	aspartato	110111	GAG	glutamato
100111	UAG	STOP	100110	UAU	tirosina
100100	UAC	tirosina	100101	UAA	STOP
000101	CAA	glutamina	000100	CAC	histidina
000110	CAU	histidina	000111	CAG	glutamina

Fonte: do autor.

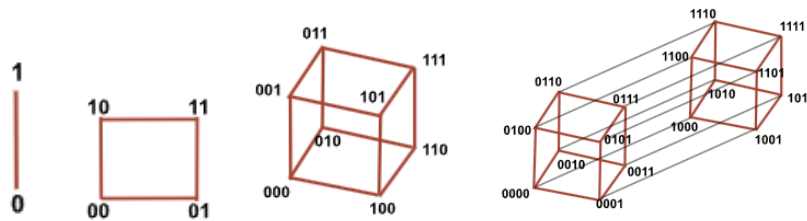
Na Subseção 3.3 serão apresentadas as construções dos hipercubos booleanos relacionados às permutações escolhidas do rotulamento A, além das conexões dessas estruturas com as tabelas do código de Gray e os diagramas de Hasse construídos anteriormente.

3.3 Construção e análise dos hipercubos booleanos do código genético

Nas construções dos hipercubos apresentadas a seguir, todas as ligações entre os vértices são feitas a partir da diferença de um bit de um vértice para o outro, conforme proposto por Hage e Harary (2002).

- Para as dimensões 1, 2, 3 e 4, estão representadas as combinações do alfabeto $\mathbb{Z}_2 = \{0, 1\}$, \mathbb{Z}_2^2 , \mathbb{Z}_2^3 e \mathbb{Z}_2^4 , respectivamente, conforme Figura 9.

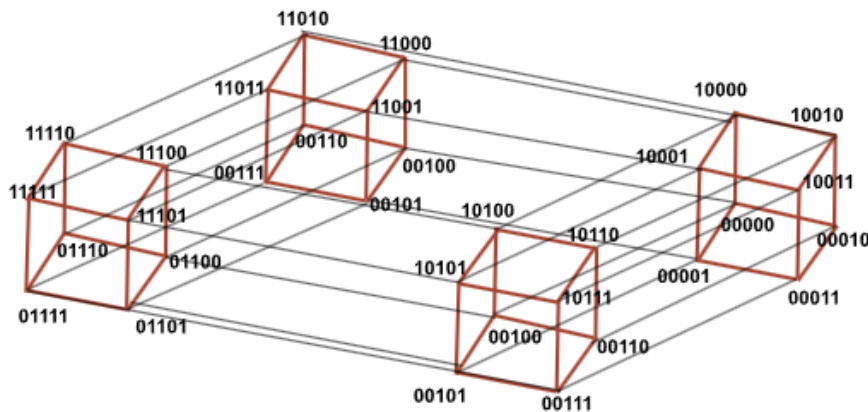
Figura 9: Dimensões 1, 2, 3 e 4.



Fonte: Adaptado de Jimenez-Montano, La Mora-Basanez e Poeschel (2002).

- Na dimensão 5, a representação geométrica é dada por quatro cubos, em que cada um dos vértices desses cubos estão representados pelas combinações de \mathbb{Z}_2^5 , conforme Figura 10.

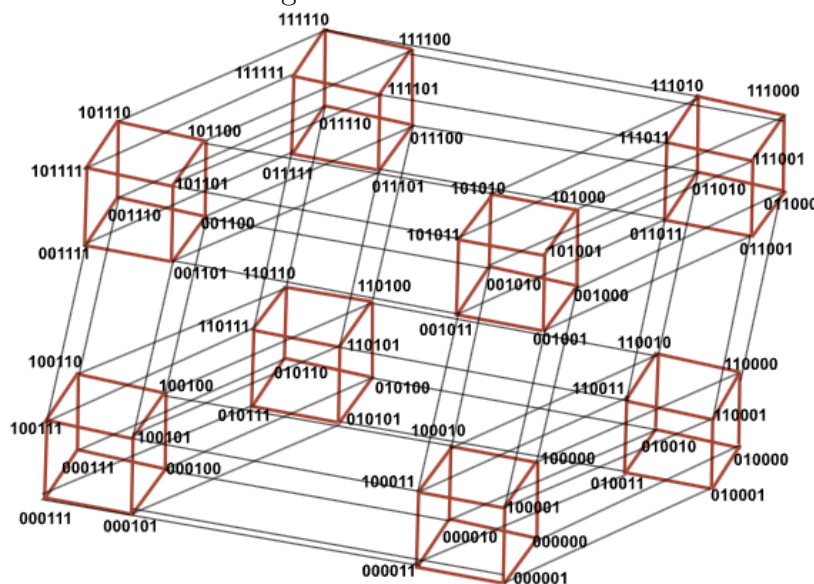
Figura 10: Dimensão 5.



Fonte: Adaptado de Jimenez-Montano, La Mora-Basanez e Poeschel (2002).

A partir das construções anteriores, será apresentada na Figura 11, a representação disponível em Jimenez-Montano, La Mora-Basanez e Poeschel (2002), onde é feita uma associação do hipercubo booleano 6-dimensional com uma representação do código genético.

Figura 11: Dimensão 6.



Fonte: Adaptado de Jimenez-Montano, La Mora-Basanez e Poeschel (2002).

Esses elementos não estão representados de maneira aleatória, a organização se dá de modo que a ligação entre os vértices tenha a diferença de um bit.

Em Jimenez-Montano, La Mora-Basanez e Poeschel (2002), é feita a proposta da construção do hipercubo booleano 6-dimensional, associado ao código genético, mas não se explicita um procedimento de construção. A seguir será apresentado um detalhamento dessa construção, representando uma contribuição deste trabalho.

1. Nas duas primeiras linhas da tabela do código de Gray, cada códon representa o vértice de uma face do cubo. Assim, ligando-se o primeiro códon com o segundo, o segundo com o terceiro, o terceiro com o quarto e o quarto com o primeiro, tem-se a primeira face do cubo;
2. Na terceira e quarta linhas da tabela do código de Gray, cada códon representa o vértice de uma face do cubo. Assim, ligando-se o primeiro códon com o segundo, o segundo com o terceiro e o terceiro com o quarto e o quarto com o primeiro, tem-se a segunda face do cubo;

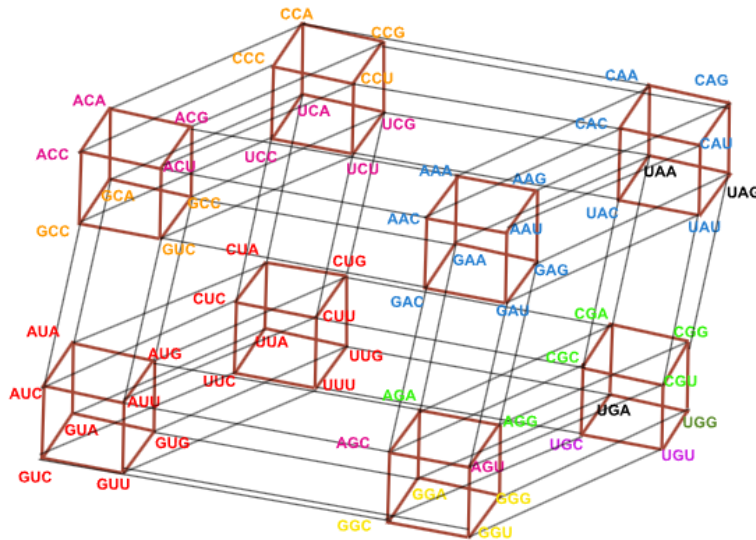
3. Construindo uma aresta que liga os vértices formados pela primeira face com os formados pela segunda face, obtém-se um cubo. Esses vértices devem ser ligados de forma que a diferença entre os códons que os representa seja de uma base;
4. Efetuando o mesmo processo de forma sucessiva para as próximas linhas, quatro a quatro, tem-se os 8 cubos que serão os “vértices” do hipercubo 6-dimensional;
5. Ligando-se esses “vértices”, é formado o hipercubo 6-dimensional.

A partir do procedimento para a construção do hipercubo booleano 6-dimensional, a associação com o código genético será feita utilizando a permutação 1203 do rotulamento A para o caso primal e a permutação 3021 do rotulamento A para o caso dual e, após a construção, serão realizadas as análises das construções realizadas.

3.4 Hipercubos booleanos associados ao rotulamento A

Os hipercubos booleanos associados às permutações 1203 (primal) e 3021 (dual) do rotulamento A estão apresentados nas Figuras 12 e 13.

Figura 12: Hipercubo booleano associado à permutação 1203 (primal) do rotulamento A.



Fonte: do autor.

As cores são representadas de acordo com as propriedades dos aminoácidos. Desta forma, no hipercubo booleano primal pode-se observar que os códons hidrofóbicos (representados pela cor vermelha) são aqueles que não possuem “afinidade” com a água

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.

Pode-se observar também os aminoácidos sulfidril ou tiol (representados pela cor roxa), os básicos (representados pela cor verde claro). Por fim, o códon STOP (representado pela cor preta), que é utilizado para interromper a proteína antes de seu término. Com isso, essas construções apresentam uma caracterização de importantes propriedades envolvendo a estrutura do código genético e poderão ser utilizadas em situações envolvendo estruturas de proteínas, além de possibilitar análises com possíveis alterações realizadas na construção dessas estruturas.

Vale ressaltar que as construções apresentadas nesta seção representam uma contribuição inédita, pois, até onde é de nosso conhecimento, ainda não foram apresentadas construções de hipercubos booleanos associados ao contexto dos rotulamentos propostos por [2] e também ainda não é conhecido o processo de classificação dos códons por meio de suas características físico-químicas, possibilitando assim uma classificação dos códons, organização dos mesmos em uma estrutura geométrica e as conexões dessas construções com outras realizadas por [3], [4] e [8].

Portanto, os resultados apresentados neste trabalho possibilitam novas perspectivas e possibilidades de estudos associados ao âmbito da Matemática, a fim de analisar outros aspectos teóricos, como a relação com os elementos da extensão de Galois $GF(2^6)$ e a análise de suas propriedades, bem como aspectos aplicados, como na análise de possíveis ajustes e interpretações dos códons dispostos nos hipercubos, como em mutações genéticas, na análise de sequências de DNA, os quais poderão contribuir significativamente em áreas da Biotecnologia, Bioinformática, Biologia e Medicina.

4 Considerações finais

Pode-se notar algumas aplicações de estruturas matemáticas no contexto do código genético, em particular, o código de Gray e o hipercubo booleano, a partir de construções realizadas de diagramas de Hasse, por meio das permutações escolhidas do rotulamento A, associado ao código genético, visto que essas estruturas matemáticas facilitam a visualização dos códons, permitindo a análise das propriedades dos aminoácidos e a classificação dos códons com base em suas características, tornando-se uma ferramenta valiosa na análise de diversos fenômenos biológicos.

A partir das construções realizadas, foi possível notar que o rotulamento A segue uma abordagem biológica, com as cores representando propriedades e características físico-químicas dos aminoácidos aos códons. Sua organização é bem estruturada, com grupos hidrofóbicos nas bordas e hidroxila e básicos no centro, refletindo proteínas com uma estrutura definida e funções específicas.

Desta maneira, por meio deste trabalho pôde-se notar a conexão existente entre elementos de Álgebra, Geometria e Biologia. Assim, essa conexão é significativa em várias áreas de pesquisa, uma vez que pode-se fundamentar estudos sobre fenômenos biológicos.

Como perspectivas futuras, visando a aplicação de tais construções em um contexto prático, espera-se que possam ser analisadas possíveis alterações na disposição dos códons nos hipercubos, analisando por exemplo, casos em que ocorra alguma(s) troca(s) de base(s), como nas mutações genéticas e verificar os impactos dessa(s) troca(s) em possíveis sequências reais, em que alguma(s) mutação(ões) possa(m) ser identificada(s), buscando assim uma forma de classificar a(s) mutação(ões) dessas sequências, por meio dos hipercubos construídos.

Outra possibilidade de aprofundamento seria a aplicação dos hipercubos booleanos na análise de mutações associadas a doenças genéticas. No caso da Síndrome de Leigh, causada por alterações em genes mitocondriais que afetam a produção de energia celular, a representação dos códons no hipercubo permitiria identificar como trocas de bases deslocam aminoácidos entre diferentes classes físico-químicas. Esse mapeamento poderia evidenciar, por exemplo, a transição de um códon hidrofílico para outro hidrofóbico, contribuindo para a instabilidade da proteína resultante e para o comprometimento de funções bioquímicas essenciais. Assim, os hipercubos booleanos auxiliam na caracterização e interpretação de mutações, ampliando as possibilidades de investigação em doenças como a Síndrome de Leigh.

Agradecimentos

À Universidade Federal de Alfenas e à 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.

Referências

- [1] Alberts, B.; et al. *Biologia molecular da célula*. Artmed Editora, 2017.
- [2] Faria, L. C. B. *Existências de códigos corretores de erros e protocolos de comunicação em sequências de DNA*. Tese (Doutorado), Universidade Estadual de Campinas, Campinas, SP, 2011.

- [3] Fernandes, R. S.; Oliveira, A. J. Análise das propriedades físico-químicas dos aminoácidos por meio das distâncias de Hamming associadas ao rotulamento C do código genético. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics* **8** (2021), no. 1.
- [4] Fernandes, R. S.; Oliveira, A. J. Caracterização das propriedades dos aminoácidos por meio do diagrama de Hasse associado ao rotulamento A do código genético. *Brazilian Electronic Journal of Mathematics* **2** (2021), no. 4, 81–100.
- [5] Gerônimo, J. R.; Franco, V. S. *Fundamentos de matemática: uma introdução à lógica matemática, teoria dos conjuntos, relações e funções*. Eduem, Maringá, 2008.
- [6] Hage, P.; Harary, F. Què es un hipercubo? Un código binario para relaciones de parentesco. *Análisis de redes: aplicaciones en ciencias sociales* (J. G. Mendieta and S. Schmidt, eds.), 1a Ed., UNAM, 2002, pp. 15–22.
- [7] Hepp, D.; Nonohay, J. S. A importância das técnicas e análises de DNA. *ScientiaTec* **3** (2016), no. 2, 114–124.
- [8] Jimenez-Montano, M. A.; La Mora-Basanez, C. R.; Poeschel, T. On the hypercube structure of the genetic code. *arXiv preprint cond-mat/0204044* (2002), 1–14.
- [9] Lin, S.; Costello, D. J., Jr. *Error control coding: fundamentals and applications*. 2nd ed., Prentice Hall, Englewood Cliffs, 1983.

Recebido em 10 de Fevereiro de 2025

Revisado em 28 de Outubro de 2025

Aceito em 18 de Novembro de 2025