Universidade Estadual de Campinas
FACULDADE DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
DEPARTAMENTO DE TELEMÁTICA

Análise Algébrica dos Rotulamentos Associados ao Mapeamento do Código Genético

Tese apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Telecomunicações e Telemática.

por

Anderson José de Oliveira

Orientador: **Prof. Dr. Reginaldo Palazzo Júnior** FEEC/UNICAMP

Banca Examinadora:

Prof. Dr. Reginaldo Palazzo Júnior (FEEC/UNICAMP) (Presidente)
Dr. Nelson Afonso Lutaif (FCM/UNICAMP)
Prof. Dr. Carlos Eduardo Câmara (Unianchieta/Jundiaí)

Campinas - SP
2012

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

OL4a

Oliveira, Anderson José de
Análise algébrica dos rotulamentos associados ao mapeamento do código genético / Anderson José de Oliveira. --Campinas, SP: [s.n.], 2012.

Orientador: Reginaldo Palazzo Júnior.
Dissertação de Mestrado - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.

1. Código genético. 2. Polinômios . 3. Algebra booleana. I. Palazzo Júnior, Reginaldo. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.

Título em Inglês: Algebraic Analyses of the labels associated with the mapping of the genetic code
Palavras-chave em Inglês: Genetic code, Polynomials, Boolean algebra
Área de concentração: Telecomunicações e Telemática
Titulação: Mestre em Engenharia Elétrica
Banca examinadora: Carlos Eduardo Câmara, Nelson Afonso Lutaif
Data da defesa: 28/02/2012
Programa de Pós Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Anderson José de Oliveira

Data da Defesa: 28 de fevereiro de 2012

Título da Tese: "Análise Algébrica dos Rotulamentos Associados ao Mapeamento do Código Genético."

Prof. Dr. Reginaldo Palazzo Júnior (Presidente): _____

Prof. Dr. Carlos Eduardo Câmara: _____

Dr. Nelson Afonso Lutaif: _____

Agradecimentos

Ao Prof. Dr. Reginaldo Palazzo Júnior pela orientação durante todo esse período, além da paciência e tranquilidade que me passava em cada uma das reuniões que tivemos nesses dois anos, excelentes apontamentos e instruções, que fizeram com que esse trabalho fosse finalizado de maneira efetiva. Uma pessoa muito especial a qual respeito muito e deixo meus sinceros agradecimentos.

Aos professores membros da banca examinadora Dr. Nelson Afonso Lutaif e Prof. Dr. Carlos Eduardo Câmara, pelos apontamentos, sugestões e pela disponibilidade em analisar o trabalho.

Ao meu pai e minha mãe, que mesmo de longe sempre me apoiaram nessa escolha. Vocês foram e sempre serão muito importantes nas escolhas que faço na minha vida. Sem vocês não seria nada. Desde a época em que estudava numa escola rural, andando quase uma hora para chegar na mesma, vocês nunca me desanimaram, ao contário, sempre me incentivando e apoiando a continuar e estudar mais e mais.

Ao meu irmão, pelas inúmeras caronas... nossa, como era bom descer do ônibus e ver o "carrinho" parado ao lado da rodovia me esperando, minhas pernas agradeciam, rs.

Aos meus tios e tias, que sempre me incentivaram a correr atrás dos meus sonhos e nunca me abater diante de problemas, muito obrigado.

Aos meus grandes amigos de Campestre e Varginha, vocês que acompanharam toda minha batalha, desde a época que trabalhava durante o dia e estudava a noite, longas madrugadas estudando para prova e sempre me dizendo que eu "chegaria lá", agradeço de coração pela amizade. Alguns diziam que eu ficaria louco de tanto estudar, rs... mais louco?

E claro, não posso esquecer dos grandes amigos que fiz aqui em Campinas, em especial meus colegas do laboratório LTIA: Maicon e Cintya, que começaram comigo em 2010, com os quais fiz várias disciplinas, ou seja, sofremos um pouquinho juntos... mas o mais importante, se tornaram grandes amigos, nos quais tenho inteira confiança... As minhas companheiras da área biológica Andréa e Lu-

zinete, como é bom conversar com vocês, nossa sintonia bate muito, suas teses me ajudaram demais. A Lucila, com sua tranquilidade e excelentes conversas. A Clarice, Fernando e aos novos amigos de laboratório, Luiz e Diogo.

E por fim, o mais importante, a Deus, muito obrigado por tudo que tem acontecido em minha vida, pelos obstáculos, desafios, os quais sempre procurei enfrentar de cabeça erguida, e agora, por mais essa conquista. Com muita fé, coragem e estudo aqui estou realizando um grande sonho. E muitos ainda virão.

Enfim, agradeço a todos que direta ou indiretamente me ajudaram nesse período. É apenas o início, muita coisa boa vem por aí.

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

Tudo o que um sonho precisa para ser realizado é alguém que acredite que ele possa ser realizado.

Roberto Shinyashiki

Resumo

Uma área de pesquisa em franca expansão é a modelagem matemática do código genético, por meio da qual pode-se identificar as características e propriedades do mesmo. Neste trabalho apresentamos alguns modelos matemáticos aplicados à biologia, especificamente relacionado ao código genético. Os objetivos deste trabalho são: a) caracterização da hidropaticidade dos aminoácidos através da construção de reticulados booleanos e diagramas de Hasse associados a cada rotulamento do código genético, b) proposta de um algoritmo soma com transporte para efetuar a soma entre códons, ferramenta importante em análises mutacionais, c) representação polinomial dos códons do código genético, d) comparação dos resultados dos rotulamentos A, B e C em cada uma das modelagens construídas, e) análise do comportamento dos aminoácidos em cada um dos rotulamentos do código genético. Os resultados encontrados permitem a utilização de tais ferramentas em diversas áreas do conhecimento como bioinformática, biomatemática, engenharia genética, etc., devido a interdisciplinaridade do trabalho, onde elementos de biologia, matemática e engenharia foram utilizados.

Palavras-Chave: Reticulados booleanos, diagrama de Hasse, soma com transporte, polinômios, código genético, mutações.

Abstract

A research area in frank expansion is the mathematical modeling of the genetic code, through can identify the characteristics and properties of them. In this paper we present some mathematical models applied to biology, specifically related to the genetic code. The aims of this work are: a) a characterization of the hydropathy of the amino acids through the construction of boolean lattices and Hasse diagrams associated with each labeling of the genetic code, b) the proposal of a sum algorithm of transportation to make the sum of codons, important tool in mutational analysis, c) a polynomial representation of the codons of the genetic code, d) a comparing of the results of the A, B and C labels in each of the built modeling, e) an analysis of the behavior of the amino acids in each of the labels of the genetic code. The results allow the use of such tools in a lot of areas like bio informatics, biomathematics, genetic engineering, etc., due to the interdisciplinary of the paper, where elements of biology, mathematics and engineering were used.

Key-words: Boolean lattices, Hasse diagram, the sum of transportation, polynomials, genetic code, mutations.

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

A modelagem algébrica do código genético é algo que diversos autores tem pesquisado, no intuito de especificar as características e propriedades que o mesmo possui.

Em [1], Sanchez et al. utilizam da construção de um reticulado booleano e de um diagrama de Hasse, onde os 64 códons do código genético são dispostos de forma organizada com o objetivo de classificar os códons em hidrofóbicos e hidrofílicos. O critério utilizado para a construção das estruturas algébricas é a complementaridade biológica, que coincide com a complementaridade algébrica.

Além disso, em [2] é feito o uso de algumas operações entre códons a fim de analisar fenômenos mutacionais que ocorrem em sequências de DNA. Essa operação é realizada através de dois processos, um envolvendo a representação numérica de cada um dos códons no código genético e o outro através de um algoritmo que considera a importância biológica das bases nitrogenadas nos códons.

Todavia, em [3], observa-se a modelagem do código genético através da representação polinomial dos códons, utilizando a extensão de $GF(2)$ para $GF(64)$. Além dessa representação, foram apresentadas algumas operações e conceitos envolvendo os códons, como a distância de Hamming entre dois códons e o peso de Hamming de um códon.

Em [4], [5] e [6] vemos uma representação do código genético através de uma estrutura de um hipercubo, que biologicamente explica as substituições conservativas e não-conservativas de aminoácidos.

Em [7], Rocha propõe um modelo de comunicação biológico de importação de proteínas mitocondriais baseado em um sistema de comunicação padrão, com o objetivo de identificar estruturas matemáticas associadas às sequências de DNA.

Faria, em [8], vai além, apresentando a existência de códigos corretores de erros e protocolos de comunicação em sequências de DNA, usando para isso estruturas matemáticas, bem como conceitos relacionados a engenharia, destacando os códigos corretores de erros.

Além dos casos mencionados anteriormente, vários outros autores, através de outras técnicas pro-

curam modelar o código genético, de maneira a desvendar os "mistérios" da "máquina da vida".

1.1 Modelos Analisados

O modelo utilizado em [1] baseia-se na construção de dois reticulados booleanos, um primal e um dual e, a partir desses reticulados, a construção de um diagrama de Hasse, onde os códons são dispostos de maneira organizada, refletindo características biológicas relevantes no estudo do código genético, classificando os códons em hidrofóbicos ou hidrofílicos, além de apresentar a regularidade na separação desses códons de acordo com a hidropaticidade. Para alcançar esse objetivo se faz necessário considerar como referência a complementaridade biológica das bases nitrogenadas, a existência de dois elementos não-comparáveis, além de um elemento máximo e um elemento mínimo. Por meio desses elementos foram construídos os reticulados booleanos primal e dual, através de operações de álgebra booleana e o diagrama de Hasse correspondente, onde ocorreu a separação dos códons hidrofóbicos e hidrofílicos de maneira uniforme nas laterais do diagrama.

O modelo utilizado em [9] considera a tabela do código genético, onde os códons são localizados de acordo com a segunda base. É definida uma operação soma no conjunto das quatro bases do DNA, utilizando uma ordem específica das bases nitrogenadas rotuladas através de um alfabeto do anel $\mathbb{Z}_4 = \{0,1,2,3\}$.

Através dessa operação, dois códons são adicionados por meio de dois processos. Como foi feita uma bijeção dos 64 códons com o anel $\mathbb{Z}_{64}$, temos representados os códons de 0 até 63. O primeiro procedimento é efetuar a soma dos códons, tomando como base essa bijeção com $\mathbb{Z}_{64}$. O segundo procedimento é através de um algoritmo que leva em consideração a importância biológica das bases nitrogenadas nos códons, fundamentado pela ocorrência de erros em cada uma das bases. Erros na terceira base são mais frequentes que erros na primeira, que por sua vez, são mais frequentes que erros na segunda base. Além dessa operação, um outro ponto relevante a ser destacado é a separação dos códons pares e ímpares na tabela do código genético, com características biológicas muito interessantes.

Por meio dessas operações, observa-se a utilização de uma estrutura de espaço vetorial, relacionando estruturas matemáticas e biológicas. Além disso, a operação soma efetuada entre códons pode se tornar ferramenta eficaz em análises mutacionais.

Em [3] foi apresentada uma maneira de representarmos os códons através de polinômios, mapeando-se os nucleotídeos através de um alfabeto $\mathbb{Z}_2 \times \mathbb{Z}_2$. Essa representação leva em consideração a importância biológica da ordem das bases na representação do códon. Dessa representação polinomial,

uma distância de Hamming entre códons foi definida, bem como o peso de Hamming de um códon. Logo, percebe-se uma interessante modelagem matemática do código genético. Assim como o modelo utilizado em [9], as operações realizadas nesse espaço vetorial permite uma análise mutacional, tornando-se ferramenta relevante no estudo do código genético.

Rocha, em [7] propõe um modelo de um sistema de codificação e decodificação do mecanismo de importação de proteínas mitocondriais. Através do mapeamento das sequências de DNA, um código BCH sobre a estrutura de anel para a geração de sequências de DNA. O mapeamento foi feito por meio da bijeção de um alfabeto biológico $N = \{A, C, G, T/U\}$ com o alfabeto $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Esse mapeamento $N \to \mathbb{Z}_4$ consiste das 24 permutações, que podem ser divididas em três rotulamentos A, B e C de acordo com as formas geométricas, que produzem um diferente nível de não-linearidade para as sequências reproduzidas.

Na Figura 1.1 , vemos representados os três rotulamentos, embasados no mapeamento dos nucleotídeos. As permutações associadas ao rótulo A levam ao mapeamento $\mathbb{Z}_4 - linear$; as permutações associadas ao rótulo B levam ao mapeamento $\mathbb{Z}_2 \times \mathbb{Z}_2 - linear$, enquanto que as permutações associadas ao rótulo C levam ao mapeamento $Klein - linear$.

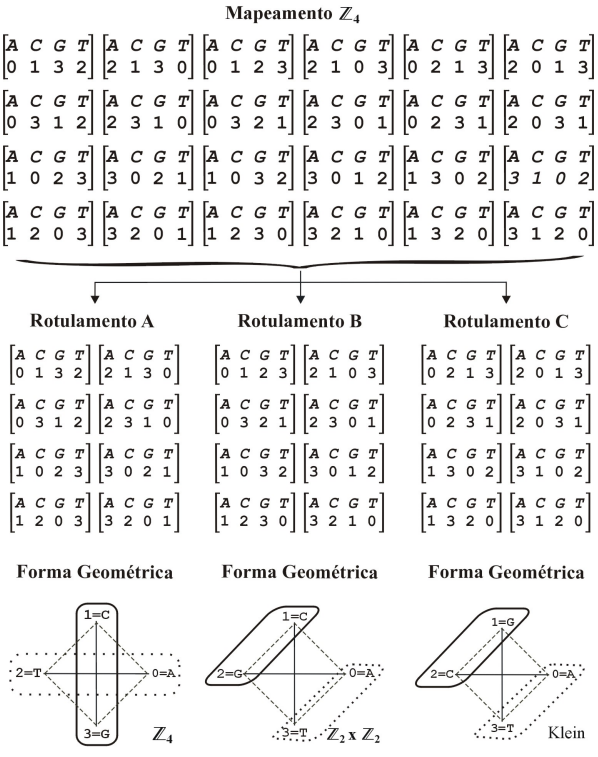

Figura 1.1: Rotulamentos A, B e C. Rocha, A.S.L., *Modelo de sistema de comunicações digital para o mecanismo de importação de proteínas mitocondriais através de códigos corretores de erros*, Tese de Doutorado, UNICAMP, 2010

Faria, em [8] apresenta a modelagem matemática do código genético através da comprovação da existência de códigos corretores de erros e protocolos de comunicação em sequências de DNA. Nesse trabalho, os rotulamentos A, B e C inicialmente apresentados em [7] são explicados de maneira detalhada. Além disso, é feita uma analogia do mapeamento dos nucleotídeos com uma modulação PSK.

A determinação dos rotulamentos proveniente da complementaridade biológica ou algébrica apresenta um interessante casamento entre o contexto biológico e o contexto matemático. No rotulamento A vemos a complementaridade biológica $(A-T)/(C-G)$ casada com a complementaridade algébrica $(00-11)/(10-01)$. No rotulamento B, não existe o casamento entre a complementaridade biológica e a complementaridade matemática. Nesse rotulamento vemos a união das bases $(A-G)/(C-T)$, por meio da complementaridade algébrica. Da mesma forma, no rotulamento C, não existe o casamento biológico e matemático. A união das bases $(A-C)/(G-T)$ se dá por meio da complementaridade algébrica.

Outro fator considerado para a classificação dos rotulamentos é o geométrico. A complementaridade biológica casada à complementaridade matemática gera um mapeamento não-linear, denominado $\mathbb{Z}_4$- linear. Como apresentado na Figura 1.1, observa-se que qualquer nucleotídeo necessita caminhar duas arestas para alcançar seu complementar biológico. No caso dos rotulamentos B e C o casamento entre a complementaridade biológica e matemática não ocorre, resultando em mapeamentos lineares. Cada nucleotídeo precisa caminhar apenas uma aresta para alcançar seu complementar biológico. O mapeamento associado ao rotulamento B é denominado $\mathbb{Z}_2\times\mathbb{Z}_2$ - linear, enquanto que o rotulamento C denomina-se Klein-linear, apresentados na Figura 1.1.

Ainda em [8] observa-se que para a geração de sequências de nucleotídeos de DNA existe um mapeamento casado entre as estruturas algébricas do codificador e do modulador, caracterizando os códigos geometricamente uniformes.

1.2 Apresentação do Problema

Muitos pesquisadores propuseram interessantes métodos para modelar matematicamente o código genético, com suas respectivas consequências biológicas.

Rocha, em [7] define três rotulamentos a partir do mapeamento das sequências de nucleotídeos do DNA de acordo com a característica geométrica que cada um deles gera, bem como de acordo com o casamento biológico/matemático. Nos trabalhos analisados em [1], [2], [9], [10], [11] e [3], observa-

mos que os métodos utilizados e os resultados obtidos estão de acordo com um dos rotulamentos: A, B ou C. Em [1], [2], [10], [11], [3], o rotulamento utilizado é o A e em [9] utilizou-se o rotulamento B. Será que os resultados encontrados para os rotulamentos em cada um dos casos são válidos para todos os rotulamentos?

Desta forma, no presente trabalho propomos a modelagem matemática dos rotulamentos associados ao mapeamento do código genético, através de algumas técnicas: primeiramente efetuamos a construção dos reticulados booleanos algébricos e os diagramas de Hasse para os rotulamentos B e C, como forma de compará-lo com o já existente do rotulamento A. Além disso, embasado no resultado encontrado em [9] relacionado à soma dos códons, propomos um novo método de soma de códons, o qual denominamos algoritmo soma com transporte, por meio do qual observamos a menor complexidade em relação ao algoritmo biológico utilizado em [9]. Para isso, usamos a bijeção do alfabeto $N = \{A, C, G, T/U\}$ com o anel $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Com isso, identificamos uma estrutura de espaço vetorial associada ao código genético e essa estrutura mostra a relevância do método proposto, uma vez que de maneira combinatorial, deslocamentos podem ser feitos no código genético, identificando sequências de DNA associadas a determinado rotulamento. Por fim, apresentamos uma representação polinomial de cada um dos 64 códons do código genético, identificamos a base canônica, determinamos algumas distâncias de Hamming entre códons, bem como encontramos o peso de Hamming de um códon, para os rotulamentos B e C, uma vez que resultados para o rotulamento A já são conhecidos. Para isso, utilizamos os elementos de $GF(64)$, obtidos através da extensão do corpo $GF(2)$, identificando mais uma vez uma estrutura de espaço vetorial, fundamentando os cálculos realizados para cada um dos rotulamentos.

Desta forma, utilizamos diversas ferramentas a fim de modelar o código genético matematicamente, embasadas nos rotulamentos A, B e C, [7] e [8]. Até onde é de nosso conhecimento, a comparação e análise de resultados de cada uma dessas técnicas usando os três rotulamentos não foi realizada anteriormente.

1.3 Organização do Trabalho

Este trabalho está organizado da seguinte maneira: no Capítulo 2 apresentamos os principais elementos teóricos utilizados no decorrer do trabalho. Devido a interdisciplinaridade do trabalho, elementos de álgebra, biologia e códigos corretores de erros serão apresentados.

No Capítulo 3 apresentamos o primeiro resultado das modelagens utilizadas. Efetuamos a construção dos reticulados booleanos para os rotulamentos B e C, bem como os diagramas de Hasse associados

aos mesmos e é feita uma análise algébrica e biológica do modelo apresentado.

Outro resultado de extrema relevância é apresentado no Capítulo 4, onde é proposto um algoritmo de soma entre códons, ferramenta eficaz em análises mutacionais. Além disso, é feito um estudo sobre o comportamento dos aminoácidos em cada um dos rotulamentos.

No Capítulo 5 é utilizada uma abordagem polinomial da representação dos códons no código genético, de maneira a determinar um espaço vetorial, onde vários cálculos foram efetuados.

Por fim, no Capítulo 6 são apresentadas as conclusões com as contribuições dos resultados encontrados, bem como sugestões e propostas para trabalhos futuros.

Capítulo 2

Elementos de Álgebra, Biologia e Códigos Corretores de Erros

O objetivo deste capítulo é apresentar os principais conceitos utilizados no decorrer do presente trabalho. Devido a interdisciplinaridade, serão apresentados elementos de álgebra, biologia, e códigos corretores de erros.

Na seção 2.1 serão apresentados alguns conceitos de álgebra abstrata, como as estruturas de grupos, anéis, estruturas isomorfas, corpos e extensões de corpos de Galois. Os cálculos envolvendo o código genético foram embasados nas estruturas citadas anteriormente. O mapeamento das bases nitrogenadas da estrutura do DNA (adenina - A, citosina - C, guanina - G e timina/uracila - T/U), considerou como alfabeto o conjunto $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ com as operações de soma e produto mod4. Para a representação polinomial dos códons foi utilizada a extensão de Galois de GF(2) para GF(64). Além disso, na seção 2.2 serão apresentados conceitos de álgebra booleana, através de suas principais operações. Na seção 2.3 será apresentada a aritmética binária, utilizando a operação soma com transporte. Essa operação foi usada na obtenção de alguns resultados da soma de códons nas tabelas do código genético para os rotulamentos A, B e C.

Os conceitos, teoremas, definições, exemplos e resultados apresentados dos elementos de álgebra podem ser encontrados em [12], [13], [14] , [15] e [16].

Na seção 2.4 serão apresentados os elementos de biologia, com a apresentação da célula, suas principais características, sua estrutura, bem como sua importância. Além disso, a estrutura do DNA e do RNA serão apresentadas, desde a duplicação do DNA até a síntese protéica. Os ácidos nucléicos, suas principais propriedades e funções também serão apresentados, bem como a tabela do código genético universal, além de um estudo dos principais aminoácidos codificados por cada uma das 64 trincas do código genético. Além disso, também serão apresentados conceitos relacionados à mutações

e a classificação dos códons em hidrofóbicos ou hidrofílicos.

Os conceitos relacionados aos elementos de biologia podem ser encontrados em [17], [18], [19], [20], [21], [22] e [23].

Na seção 2.5 apresentamos alguns conceitos relacionados aos códigos corretores de erros com um pequeno histórico, os códigos de bloco e os códigos geometricamente uniformes, que são utilizados na geração de sequências de DNA a partir da determinação de um espaço vetorial do código genético.

Os conceitos relacionados aos elementos dos códigos corretores de erros podem ser encontrados em [7], [8],[24], [25], [26], [27], [28], [16] e [29].

2.1 Álgebra Abstrata

Nesta seção serão apresentados os principais conceitos de álgebra abstrata usados no decorrer do trabalho. A mesma está dividida da seguinte forma: na subseção 2.1.1 apresentamos os principais conceitos, definições e exemplos de grupos. Na subseção 2.1.2 são apresentadas as estruturas binárias isomorfas. Na subseção 2.1.3 um estudo sobre a estrutura de anéis e por fim, na subseção 2.1.4 um estudo sobre os corpos, corpos de Galois e a construção de corpos de Galois em $GF(2^m)$.

2.1.1 Grupos

Definição 2.1 *Seja G um conjunto de elementos. Uma operação binária $*$ em G é uma regra que assinala para cada par de elementos a e b um único terceiro elemento $c = a * b$ em G. Quando uma operação binária $*$ é definida em G, dizemos que G é fechada sobre $*$. Uma operação binária $*$ em G é dita associativa se, para qualquer a, b e c em G:*

$$a * (b * c) = (a * b) * c$$

Definição 2.2 *Um conjunto G em que uma operação binária $*$ é definida é chamado **grupo** se as seguintes condições são satisfeitas:*

- *a operação binária $*$ é associativa;*
- *G contém um elemento e, tal que, para qualquer a em G, $a * e = e * a = a$. Este e é chamado elemento identidade de G;*
- *para qualquer elemento a em G, existe outro elemento a' em G, tal que $a * a' = a' * a = e$. O elemento a' é chamado inverso de a (a é também um inverso de a').*

Definição 2.3 *Um grupo G é dito comutativo se a operação binária $*$ também satisfaz a seguinte condição: para qualquer a e b em G, $a*b = b*a$.*

Teorema 2.1 *O elemento identidade em um grupo G é único.*

Teorema 2.2 *O inverso de um elemento no grupo é único.*

Exemplo 1 *Considere o conjunto de dois inteiros $G = \{0,1\}$. Seja definida uma operação binária, denotada por $+$, em G como segue:* $0+0=0, 0+1=1, 1+0=1, 1+1=0$

Esta operação binária é chamada de adição módulo 2. O conjunto $G = \{0,1\}$ é um grupo sobre adição módulo 2. A partir da definição segue que a adição módulo 2 $(+)$ em G é fechada sobre $+$, e $+$ é comutativa. Além disso, pode-se facilmente checar que $+$ é associativa. O elemento 0 é o elemento identidade. O inverso de 0 é ele mesmo e o inverso de 1 é também ele mesmo. Então, G com a operação $+$ é um grupo comutativo.

O número de elementos de um grupo é chamado de **ordem** desse grupo. Um grupo de ordem finita é chamado de **grupo finito**. Para qualquer inteiro positivo m, é possível construir um grupo de ordem m sobre uma operação binária que é muito parecido com a adição real. Observe o exemplo apresentado a seguir.

Exemplo 2 *Seja m um inteiro positivo. Conisidere o conjunto de inteiros $G = \{0,1,2,...,m-1\}$. Seja $+$ denotando a adição real. Defina uma operação binária $+$ em G como segue. Para quaisquer inteiros i e j em G, temos: $i+j = r$, onde r é o resto da divisão de $i+j$ por m. O resto r é um inteiro entre* 0 *e* $m-1$ *(algoritmo da divisão de Euclides) e está portanto em G. Logo, G é fechado para a operação binária $+$, e é chamada adição módulo m. O conjunto $G = \{0,1,2,...,m-1\}$ é um grupo sobre a adição módulo m, atendendo as condições indicadas na definição de grupo.*

Vamos considerar um caso particular para $m = 6$. Podemos construir um grupo finito para esse caso usando uma tabela com os elementos desse grupo para a operação $+$. O grupo aditivo da adição módulo 7 é apresentado na Tabela 2.1.

$+$	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

Tabela 2.1: Adição módulo 7

Além do grupo aditivo, podemos construir também grupos finitos com operação binária similar a multiplicação real. Considere o exemplo apresentado a seguir:

Exemplo 3 *Seja p um primo (i.e, $p = 2,3,5,7,11,...$). Considere o conjunto dos inteiros, $G = \{1,2,3,...,p-1\}$. Seja $\cdot$ denotando a multiplicação real. Defina uma operação binária $\cdot$ em G, como segue: para i e j em G, $i \cdot j = r$, onde r é o resto da divisão de $i \cdot j$ por p. Primeiramente, note que $i \cdot j$ não é divisível por p. Por isso, $0 < r < p$ e r é um elemento em G. Portanto, o conjunto G é fechado para a operação binária $\cdot$, que é referido como uma multiplicação módulo p. O conjunto $G = \{1,2,...,p-1\}$ é um grupo sobre a multiplicação módulo p, de acordo com as condições apresentadas na definição de grupo.*

Se p não for primo, o conjunto $G = \{1,2,...,p-1\}$ não é um grupo sobre a multiplicação módulo p. Veja a Tabela 2.2 que apresenta o grupo $G = \{1,2,3,4,5,6\}$ sobre a multiplicação módulo 7.

$\cdot$	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

Tabela 2.2: Multiplicação módulo 7

2.1.2 Estruturas binárias isomorfas

Considere a Tabela 2.3 para as operações binárias $*$ e $*'$ e nos conjuntos $S = \{a,b,c\}$ e $T = \{\clubsuit, \diamondsuit, \heartsuit\}$.

$*$	a	b	c
a	c	a	b
b	a	b	c
c	b	c	a

$*'$	♣	♢	♡
♣	♡	♣	♢
♢	♣	♢	♡
♡	♢	♡	♣

Tabela 2.3: Isomorfismo

Observe que de uma Tabela para a outra temos a correspondência um-para-um usando os símbolos:

$$a \leftrightarrow \clubsuit$$
$$b \leftrightarrow \diamondsuit$$
$$c \leftrightarrow \heartsuit$$

As duas tabelas se diferenciam apenas pelos símbolos (ou nomes) denotando os elementos e os símbolos $*$ e $*'$ para as operações. Pode-se criar uma nova tabela usando a seguinte correspondência um-para-um:

$$a \leftrightarrow y$$
$$b \leftrightarrow x$$
$$c \leftrightarrow z$$

com a operação $*$". A Tabela 2.4 com os novos elementos, está representada abaixo:

$*$"	y	x	z
y	z	y	x
x	y	x	z
z	x	z	y

Tabela 2.4: Isomorfismo

A partir das observações acima, pode-se compreender a seguinte definição:

Definição 2.4 *Sejam $< S, * >$ e $< S', *' >$ estruturas algébricas binárias. Um isomorfismo de S em S' é uma função mapeamento um-para-um ϕ tal que:*

$$\phi(x * y) = \phi(x) *' (y), \text{ para todo } x, y \in S.$$

Se tal mapeamento existe, então S e S' são estruturas binárias isomorfas e denotamos por $S \simeq S'$.

2.1.3 Anéis

Seja A um conjunto não vazio onde estejam definidas duas operações, as quais chamaremos de soma e produto em A e denotaremos (como em $\mathbb{Z}$) por $+$ e $*$.

Assim:

$$\begin{array}{rccc} +: & AXA & \to & A \\ & (a,b) & \mapsto & a+b \end{array} \quad e \quad \begin{array}{rccc} *: & AXA & \to & A \\ & (a,b) & \mapsto & a*b \end{array}$$

Chamaremos $< A,+,* >$ de anel se as seguintes 6 propriedades são verificadas quaisquer que sejam $a,b,c \in A$.

1. $(a+b)+c = a+(b+c) \to$ associatividade da soma.
2. $\exists 0 \in A$ tal que $a+0 = 0+a = a \to$ existência do elemento neutro para a soma;
3. $\forall x \in A$ existe um número $y \in A$, denotado por $y = -x$, tal que $x+y = y+x = 0 \to$ existência do inverso aditivo;
4. $a+b = b+a \to$ comutatividade da soma;
5. $(a*b)*c = a*(b*c) \to$ associatividade do produto;
6. $a*(b+c) = a*b+a*c; (a+b)*c = a*c+b*c \to$ distributividade à esquerda e à direita;

   Se um anel $< A,+,* >$ satisfaz a propriedade:

7. $\exists 1 \in A, 0 \neq 1$, tal que $x*1 = 1*x = x, \forall x \in A$, dizemos que $< A,+,* >$ é um **anel com unidade 1**.

   Se um anel $< A,+,* >$ satisfaz a propriedade:

8. $\forall x,y \in A \; x*y = y*x$, dizemos que $< A,+,* >$ é um **anel comutativo**.

   Se um anel $< A,+,* >$ satisfaz a propriedade:

9. $x,y \in A, x*y = 0 \Rightarrow x = 0$ ou $y = 0$, dizemos que $< A,+,* >$ é um **anel sem divisores de zero**.

   Se $< A,+,* >$ é um anel comutativo, com unidade e sem divisores de zero, dizemos que $< A,+,* >$ é um **domínio de integridade**.

   E, finalmente, se um domínio de integridade $< A,+,* >$ satisfaz a propriedade:

10. $\forall x \in A, x \neq 0, \exists y \in A$ tal que $x*y = y*x = 1$, dizemos que $< A,+,* >$ é um **corpo**.

Exemplo 4 *($\mathbb{Z}$,+,*), ($\mathbb{Q}$,+,*), ($\mathbb{R}$,+,*), ($\mathbb{C}$,+,*) são anéis.*

2.1.4 Corpos

Definição 2.5 *Seja F um conjunto de elementos sobre o qual duas operações binárias, a adição "+"e a multiplicação "*"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 em* $*$*.O elemento identidade é o* 1 *(um).*
3. *A multiplicação é distributiva sob adição, isto é, para quaisquer* a,b *e* c *em* F*,* $a\cdot(b+c) = a\cdot b + a\cdot c$

Corpos de Galois

Os corpos de Galois são corpos com número finito de elementos representado por $GF(p^m)$, onde p é um número primo, e m um inteiro positivo.

Exemplo 5 *Considere o conjunto* $\{0,1\}$ *cujas operações de adição e multiplicação módulo-2 são apresentadas na Tabela 2.5.*

Este conjunto é um corpo sob adição e multiplicação módulo-2, ou seja, é um corpo de Galois, $GF(2) = \{0,1\}$.

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Tabela 2.5: Adição e multiplicação módulo-2

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.

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

A construção de um corpo de Galois, a partir de um polinômio primitivo, resulta em uma representação em forma de potência, uma em forma de polinômio e uma em forma vetorial.

Teorema 2.3 *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.*

Teorema 2.4 *Um polinômio irredutível $p(x)$ de grau m é dito primitivo se o menor positivo n para o qual $p(x)$ divide x^n+1 é $n=2^m-1$.*

Exemplo 6 *Seja $m=4$ e considere o polinômio primitivo sobre $GF(2)$, $p(x)=1+x+x^4$. Admitindo que α seja uma raíz do polinômio, então $p(\alpha)=0$, ou seja:*

$$0=1+\alpha+\alpha^4 \Rightarrow \alpha^4=1+\alpha$$

A partir da relação acima pode-se construir um $GF(2^4)$ como se segue:

$$\alpha^5=\alpha\cdot\alpha^4=\alpha(1+\alpha)=\alpha+\alpha^2$$
$$\alpha^6=\alpha\cdot\alpha^5=\alpha(\alpha+\alpha^2)=\alpha^2+\alpha^3$$
$$\alpha^7=\alpha\cdot\alpha^6=\alpha(\alpha^2+\alpha^3)=\alpha^3+\alpha^4=\alpha^3+1+\alpha=1+\alpha+\alpha^3$$
$$\vdots$$
$$\alpha^{2^m-2}=\alpha^{14}=1+\alpha^3$$

Os elementos de $GF(2^4)$ estão apresentados na Tabela 2.6.

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

Tabela 2.6: $GF(2^4)$ gerado por $p(x)=1+x+x^4$

Por serem finitos, algumas operações sobre os corpos de Galois são realizadas de forma singular se comparadas com as operações equivalentes de álgebra comum.

Exemplo 7 *Seja o $GF(2^4)$ gerado por $p(x)=1+x+x^4$. A adição entre os termos α^3 e α^8 é:*

$$\alpha^3+\alpha^8=\alpha^3+1+\alpha^2=1+\alpha^2+\alpha^3=\alpha^{13}$$

Exemplo 8 *Seja o $GF(2^4)$ gerado por $p(x)=1+x+x^4$. O elemento de ordem mais alta do corpo é:*

$$\alpha^{2^m-2}=\alpha^{14}$$

Considere agora os elementos α^8 *e* α^{10}*. O produto entre esses dois elementos é:*

$$\alpha^8 \cdot \alpha^{10} = \alpha^{18}$$

Observe que $\alpha^{18} > \alpha^{14}$ *e assim, o expoente deve ser reduzido fazendo* $18 : (2^m - 1) = 18 : 15 = 1$ *e o resto é 3. Logo:*

$$\alpha^8 \cdot \alpha^{10} = \alpha^{18} = \alpha^3$$

2.2 Álgebra Booleana

Esta seção está dividida da seguinte maneira: na subseção 2.2.1 é apresentado o conceito de álgebra booleana. Na subseção 2.2.2 a dualidade na álgebra booleana. Em 2.2.3 vemos o conceito de ordem de uma álgebra booleana. Na subseção 2.2.4 apresentamos o projeto de cicuito de interruptores, onde são detalhados os conectivos lógicos de conjunção e disjunção, que serão utilizados na construção dos reticulados booleanos algébricos. Por fim, na subseção 2.2.5 vemos os conceitos dos diagramas de Hasse, que serão aplicados no mundo biológico.

2.2.1 Conceitos iniciais

A álgebra de Boole ou booleana é assim denominada em honra ao matemático George Boole (1813 - 1864), matemático britânico que em 1854, publicou "An investigation of the Laws of Thought", onde descreveu um sistema algébrico mais tarde designado por álgebra de Boole.

Definição 2.6 *Uma álgebra booleana é um conjunto B de elementos a, b, ... e duas operações binárias chamadas soma e produto, designadas respectivamente por* $+$ *e* $\cdot$*, tais que:*

1. *Lei do fecho: para qualquer* $a, b \in B$*, a soma* $a+b$ *e o produto* $a \cdot b$ *existem e são elementos únicos em B.*

2. *Lei comutativa:* $a+b = b+a$ *e* $a \cdot b = b \cdot a$

3. *Lei associativa:* $(a+b)+c = a+(b+c)$ *e* $(a \cdot b) \cdot c = a \cdot (b \cdot c)$

4. *Lei distributiva:* $a+(b \cdot c) = (a+b) \cdot (a+c)$ *e* $a \cdot (b+c) = (a \cdot b)+(a \cdot c)$

5. *Identidade: uma identidade aditiva* 0 *e uma identidade multiplicativa* U *existem tal que, para qualquer* $a \in B$*,* $a+0 = a$ *e* $a \cdot U = a$

6. *Complemento: para qualquer $a \in B$ existe um $a' \in B$ chamado de complemento de a, tal que:*

   $a + a' = U$

   $a \cdot a' = 0$

Exemplo 9 *Seja $B = \{1,0\}$ e sejam duas operações $+$ e $\cdot$ definidas em B da seguinte maneira na Tabela 2.7:*

$+$	1	0
1	1	1
0	1	0

$\cdot$	1	0
1	1	0
0	0	0

Tabela 2.7: Operações lógicas básicas

Portanto, B, ou mais precisamente, o terno $(B,+,\cdot)$ é uma álgebra booleana.

2.2.2 Dualidade na álgebra boolena

Por definição, o dual de qualquer proposição numa álgebra booleana $(B,+,\cdot)$ é a proposição derivada trocando $+$ e $\cdot$, e seus elementos identidade U e 0, na proposição original; por exemplo, o dual de:

$$(U + a) \cdot (b + 0) = b \text{ é } (0 \cdot a) + (b \cdot U) = b$$

Observe que o dual de cada axioma de uma álgebra booleana é também um axioma. Dessa forma, o princípio da dualidade se mantém, isto é:

Teorema 2.2.1 *(Princípio da dualidade) O dual de qualquer teorema numa álgebra booleana é também um teorema.*

2.2.3 Ordem na álgebra booleana

Considere o seguinte teorema:

Teorema 2.2.2 *Sejam $a, b \in B$ uma álgebra booleana. Assim, as seguintes condições são equivalentes:*

1. $a \cdot b' = 0$
2. $a + b = b$

3. $a' + b = U$

4. $a \cdot b = a$

Definição 2.7 *Sejam $a, b \in B$, uma álgebra booleana. Dizemos que a precede b, denotado por: $a \leq b$ se uma das propriedades do teorema anterior se mantiver.*

Teorema 2.2.3 *A relação numa álgebra booleana B, definida por $a \leq b$, é uma ordem parcial em B, isto é:*

1. *$a \leq a$ para cada $a \in b \rightarrow$ Lei reflexiva;*

2. *$a \leq b$ e $b \leq a$ implica em $a = b \rightarrow$ Lei anti-simétrica;*

3. *$a \leq b$ e $b \leq c$ implica em $a \leq c \rightarrow$ Lei transitiva.*

A menos que seja estabelecido em contrário, uma álgebra booleana é considerada, pelo teorema anterior, **parcialmente ordenada**.

2.2.4 Projeto de circuitos

Sejam A, B, ... chaves, e sejam A e $\overline{A}$ chaves que apresentam a propriedade de que se uma está fechada, a outra está aberta, e vice-versa. Duas chaves, digamos A e B, podem ser ligadas, por fios, em série ou em paralelo, conforme apresentado nas Figuras 2.1 e 2.2.

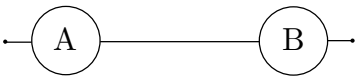

Figura 2.1: Combinação em série, $A \wedge B$

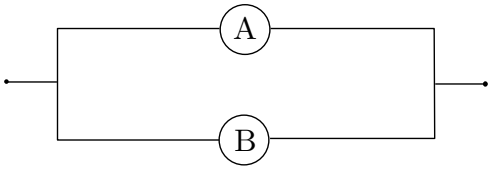

Figura 2.2: Combinação em paralelo, $A \vee B$

Note que $A \wedge B$ e $A \vee B$ denotam que as chaves A e B estão em série e em paralelo, respectivamente.

Um circuito elétrico booleano significa um arranjo de fios e chaves que pode ser montado com o uso repetido de combinações em série e em paralelo; pode assim ser descrito pelo uso dos sinais $\wedge$ e $\vee$.

Denotaremos que uma chave ou circuito está ligada por "1" e que uma chave ou circuito está desligada por "0". A Tabela 2.8 mostra a "tabela-verdade"dos circuitos $A \wedge B$ e $A \vee B$.

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

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

Tabela 2.8: Operações booleanas de conjunção e disjunção

A Tabela 2.9 mostra a relação entre uma chave A e uma chave $\overline{A}$.

A	$\overline{A}$
1	0
0	1

Tabela 2.9: Operação booleana de negação

As tabelas apresentadas serão usadas na construção dos resultados relacionados a contrução dos reticulados booleanos algébricos e dos diagramas de Hasse. Essas tabelas são idênticas às tabelas-verdade de conjunção, disjunção e negação, com a diferença de se usar nas mesmas 1 e 0 ao invés dos conectivos V e F.

2.2.5 Diagramas de Hasse

Escrevemos $x < y$ quando $x \leq y$ e $x \neq y$. Dado um conjunto parcialmente ordenado, também chamado *poset* $(A, \leq)$ e $x, y \in A$, dizemos que y cobre x se, e somente se, $x < y$ e não há outro elemento $z \in A$ tal que $x < z < y$. Um diagrama de Hasse do *poset* $(A, \leq)$ é uma representação gráfica onde vértices representam os elementos de A e dois elementos x e y são ligados por uma aresta se, e somente se y cobre x. Em um diagrama de Hasse, os elementos menores (com relação a ordem parcial) são em geral desenhados abaixo dos elementos maiores.

Exemplo 10 *O diagrama de Hasse do poset* $\{\{a,b,c\}, \subseteq\}$ *é mostrado na Figura 2.3:*

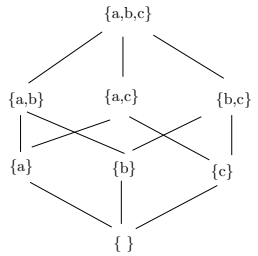

Figura 2.3: Diagrama de Hasse de$\{\{a,b,c\},\subseteq\}$

Seja $(A,\leq)$ um *poset.*

1. Chama-se menor elemento de A um elemento $0 \in A$ tal que $0 \leq x$ para todo $x \in A$;

2. Chama-se maior elemento de A um elemento $1 \in A$ tal que $x \leq 1$ para todo $x \in A$;

Teorema 2.2.4 *Em um poset $(A,\leq)$, se existe um menor elemento, ele é único. Similarmente, se existe um maior elemento, ele é único.*

1. Chama-se **elemento minimal** de A um elemento $a \in A$ tal que elemento algum de A precede estritamente a, isto é, para qualquer $x \in A$, se $x \leq A$, então $x = a$.

2. Chama-se **elemento maximal** de A um elemento $a \in A$ tal que elemento algum de A sucede estritamente a, isto é, para qualquer $x \in A$, se $x \geq A$, então $x = a$.

Teorema 2.2.5 *Se 0 é o menor elemento de $(A,\leq)$, então 0 é o único elemento minimal de $(A,\leq)$.*

Seja $(A,\leq)$, um *poset* e seja $X \subset A$.

1. Chama-se **limitante inferior** de X em A a todo elemento $a \in A$ tal que $a \leq x$ para todo $x \in X$.

2. Chama-se **limitante superior** de X em A a todo elemento $a \in A$ tal que $x \leq a$ para todo $x \in X$.

3. Chama-se **ínfimo** de X em A o maior elemento dos limitantes inferiores de X em A. O ínfimo de X é denotado $\bigwedge X$. O ínfimo de $\{x,y\}$ é denotado $x \wedge y$. O operador $\wedge$ é chamado **conjunção**.

4. Chama-se **supremo** de X em A o menor elemento dos limitantes superiores de X em A. O supremo de X é denotado $\bigvee X$. O supremo de $\{x,y\}$ é denotado $x \vee y$. O operador $\vee$ é chamado **união ou disjunção**.

Teorema 2.2.6 *Seja $a,b \in B$ uma álgebra booleana. Assim:*

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

$a \cdot b = infimo\{a,b\}$

Observação 2.1 *Qualquer conjunto A, parcialmente ordenado, tal que $infimo\{a,b\}$ e $supremo\{a,b\}$ existam para quaisquer elementos $a,b \in A$, é chamado um reticulado. Desse modo, uma álgebra booleana é um tipo especial de* ***reticulado****.*

Exemplo 11 *Seja $S = \{a,b,c\}$ e considere o conjunto de todos os subconjuntos próprios de S, conforme ilustrado no diagrama da Figura 2.4:*

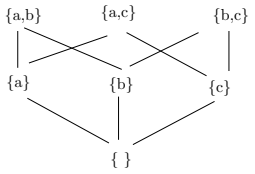

Figura 2.4: Diagrama de Hasse

Este conjunto tem três elementos maximais, $\{a,b\}$, $\{a,c\}$ e $\{b,c\}$. O menor elemento desse conjunto é $\{\}$ e, conforme teorema 2.2.5, é o único elemento minimal. O conjunto não possui maior elemento.

2.3 Aritmética Binária

Para realizar os cálculos envolvendo a soma de códons no código genético para os rotulamentos A, B e C, propusemos um método utilizando uma soma com transporte de 1. Esse cálculo fundamenta-se na adição binária realizada como a adição decimal. Se tomarmos dois números decimais 56719 e 31863 e adicioná-los, teremos a soma 88582. Podemos analisar os detalhes desta operação da seguinte maneira:

Transporte	0	0	1	0	1	
		5	6	7	1	9
Parcelas	+	3	1	8	6	3
Soma		8	8	5	8	2

Somando a primeira coluna, números decimais 9 e 3, resulta o dígito 2 com um transporte de 1. O transporte é então somado à próxima coluna. Adicionado à segunda coluna, $(1+1+6)$, resulta o

número 8, sem transporte. Este processo continua até que todas as colunas (incluindo os transportes) tenham sido somadas. A soma representa o valor numérico das parcelas.

Quando você soma dois números binários, você realiza a mesma operação.

Veja abaixo um resumo das regras de adição com números binários:

$0+0=0$

$0+1=1$

$1+1=0 \longrightarrow$ com transporte de 1

$1+1+1=1 \longrightarrow$ com transporte de 1

Veja um exemplo ilustrando o processo de adição binária, através da soma de 1101 com 1101.

$$\begin{array}{rccccc} \textit{Transporte} & 1 & 1 & 0 & 1 & \\ \textit{Parcela} & & 1 & 1 & 0 & 1 \\ + & 1 & 1 & 0 & 1 & \\ \hline & 1 & 1 & 0 & 1 & 0 \end{array}$$

Na primeira coluna, 1 mais 1 resulta 0 com transporte de 1 para a segunda coluna. Na segunda coluna, 0 mais 0 resulta 0 sem transporte. A este resultado, o transporte da primeira coluna é somado. Assim, 0 mais 1 resulta 1 sem transporte.

Estas duas adições na segunda coluna dão uma soma total de 1 com um transporte de 0.

Na terceira coluna, 1 mais 1 resulta 0 com um transporte de 1. Nesta soma, o transporte da segunda coluna é somado. Isto resulta uma soma da terceira coluna de 0 com um transporte de 1 para a coluna 4.

Na coluna quatro, 1 mais 1 resulta 0 com um transporte de 1. Para esta soma, o transporte da terceira coluna é somado. Isto resulta uma soma da quarta coluna de 1 com um transporte para a quinta coluna.

Na quinta coluna não há parcelas. Assim $1101_2 + 1101_2$ é igual a 11010_2.

Em nosso trabalho estendemos o conceito de soma com transporte de 1 para o alfabeto Z_4, usando procedimento análogo aos exemplos apresentados anteriormente.

2.4 Elementos de Biologia

Nesta seção apresentamos os conceitos de biologia utilizados neste trabalho. Na subseção 2.4.1 apresentamos as células, suas principais características e funcionalidades. Na subseção 2.4.2 um estudo sobre os nucleotídeos e os ácidos nucléicos. Na subseção 2.4.3 detalhamos a estrutura do ácido desoxiribonucléico, o DNA, cujo processo de duplicação e síntese protéica são detalhados na subseção 2.4.4. Os principais aminoácidos, sua estrutura e funções serão apresentados na subseção 2.4.5. Na

subseção 2.4.6 apresentamos o código genético, uma analogia de um sistema de comunicações padrão e um sistema de comunicações biológico, além da tabela universal do código genético. Por fim, na subseção 2.4.7 apresentamos um estudo sobre as mutações e suas consequências, benéficas ou maléficas, nos organismos.

2.4.1 A célula

Há mais de três bilhões de anos, sob condições não inteiramente claras e num instante de tempo difícil de compreender, surgiram moléculas orgânicas, através de combinações e recombinações de elementos como carbono, hidrogênio, oxigênio, nitrogênio, enxofre e fósforo.

Com o passar do tempo, diversas células se desenvolveram e tanto a química como a estrutura das células tornaram-se complexas.

A célula é a unidade básica da vida em todas as formas de organismos vivos, da mais simples bactéria ao mais complexo animal.

As células podem ser classificadas em duas categorias: procariotas e eucariotas. Essa classificação é feita com base nas diferenças macroscópicas e bioquímicas.

Nas células procariotas, o material genético está presente em toda a célula. Aqui estão inclusas as bactérias e rickettsia. Os genes são encontrados geralmente agrupados em operons. Um operon é um conjunto de genes que estão relacionados e que estão sob o controle de um único promotor (região reguladora). A Figura 2.5 apresenta um modelo de célula procarionte.

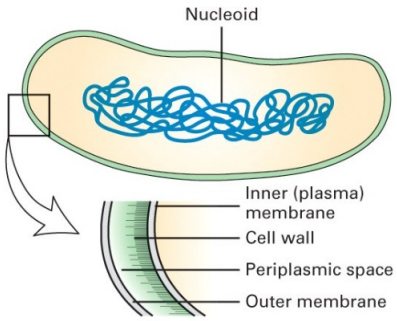

Figura 2.5: Célula procarionte. Lodish *et al. Molecular Cell Biology*, 5th Edition.

Nas células eucariotas, o material genético está organizado no núcleo, um compartimento bem definido. Incluem as células de leveduras, fungos, vegetais e animais. Também chamadas de eucélulas, são mais complexas que as procariotas, possuindo membrana nuclear individualizada e vários tipos de organelas. A Figura 2.6 apresenta um modelo de célula eucarionte.

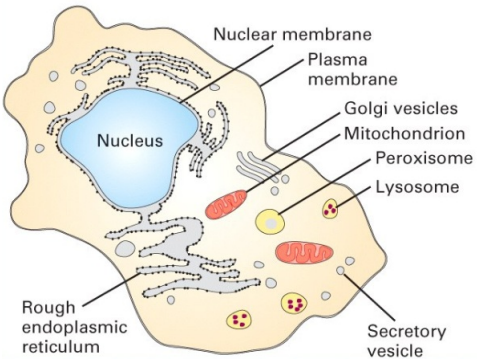

Figura 2.6: Célula eucarionte. Lodish *et al. Molecular Cell Biology*, 5th Edition.

Nessas células, os processos de transcrição e tradução ocorrem de forma separada, a transcrição ocorre no núcleo e a tradução ocorre no citoplasma. Uma surpreendente descoberta sobre genes eucarióticos, em 1977 foi de que os genes eucarióticos continham pedaços extras, que não aparecem no RNA mensageiro do gene codificado, os chamados íntrons, que devem ser removidos antes do RNA mensageiro ser traduzido.

2.4.2 Nucleotídeos e ácidos nucléicos

Os nucleotídeos são unidades moleculares que quando ligadas entre si formam os ácidos nucléicos. Esses nucleotídeos participam na transferência de energia e junto com os ácidos nucléicos desempenham funções estruturais e catalíticas nas células. As formas poliméricas dos nucleotídeos, os ácidos nucléicos (DNA e RNA), são os participantes básicos no armazenamento e codificação da informação genética.

A vida, como a conhecemos está intimamente ligada à química dos nucleotídeos e dos ácidos nucléicos.

O DNA e o RNA são substâncias químicas envolvidas na transmissão de caracteres hereditários e na produção de proteínas, compostos que são o principal constituinte dos seres vivos. Os ácidos nucléicos são encontrados em todas as células e são conhecidos em português pelas siglas ADN (ácido desoxiribonucléico) e ARN (ácido ribonucléico). De acordo com a moderna Biologia, o DNA faz RNA, que faz proteína (embora existam exceções, os retrovírus, como o vírus da Aids).

Existem quatro diferentes tipos de nucleotídeos no DNA, sendo que esses nucleotídeos são compostos por três partes: um grupo fosfato, uma pentose (desoxiribose) 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 2.7 vemos os nucleotídeos do DNA.

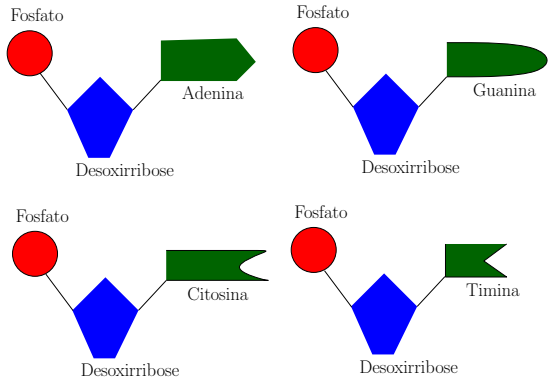

Figura 2.7: Nucleotídeos do DNA

Essas bases nitrogenadas são unidas através de pontes de hidrogênio, obedecendo uma complementaridade, conhecida como **regra de Chargaff**, descoberta no final dos anos 40 por Erwin Chargaff, onde se vê adenina ligando-se com timina e citosina ligando-se com guanina, sendo dois o número de pontes de hidrogênio entre adenina e timina e três entre citosina e guanina.

O RNA é também uma longa fita de nucleotídeos ligados entre si, mas ao contrário do DNA que é constituído por uma fita dupla, o RNA é constituído por uma fita simples.

Assim como o DNA, o RNA possui em sua estrutura um grupo fosfato, uma pentose e uma base nitrogenada, mas com algumas diferenças em relação ao DNA. No RNA, a pentose é a ribose e as bases nitrogenadas são: adenina, guanina, citosina e uracila, ou seja, ao invés de timina, exclusiva do DNA, o RNA possui a uracila, que por sua vez, é exclusiva do RNA. Na Figura 2.8 vemos os nucleotídeos do RNA.

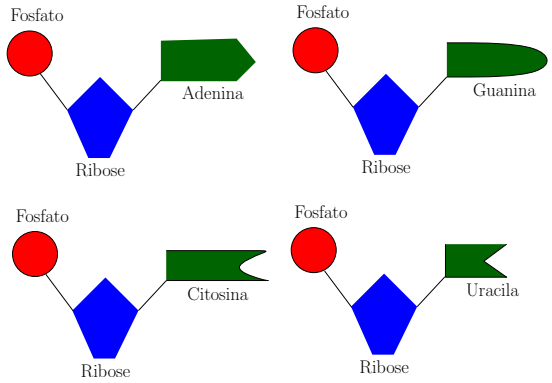

Figura 2.8: Nucleotídeos do RNA

Existem três tipos de RNA: o RNA mensageiro, o RNA transportador e o RNA ribossômico, cada um desempenhando uma função específica no processo de síntese protéica. Esse processo possui algumas diferenças nos organismos procariontes e eucariontes. Nos procariontes, a transcrição e a tradução ocorrem próximas uma da outra. Já nos eucariontes, a transcrição ocorre no núcleo e a tradução no citoplasma da célula.

Antes de apresentar os processos de transcrição e tradução é de vital importância entendermos a molécula de DNA, sua estrutura e importância.

2.4.3 A molécula de DNA

O DNA é um ácido desoxiribonucléico, cuja estrutura mais difundida é a dupla hélice(ou duplex).

A determinação da estrutura do DNA por James Watson e Francis Crick em 1953 é, em geral, aceita como o marco da biologia molecular moderna. O modelo de Watson e Crick possui as seguintes características principais:

1. duas cadeias polinucleotídicas (chamadas fitas) , formando a dupla hélice.

2. as duas fitas de DNA são antiparalelas.

3. as bases nitrogenadas ocupam o centro da hélice e as cadeias de açucar-fosfato estão na periferia.

4. cada base está ligada a outra por meio de pontes de hidrogênio formando um par de base planar, onde adenina se liga com timina e vice-versa e guanina se liga com citosina e vice-versa.

Cada fita de DNA pode atuar como um molde para a síntese de sua fita complementar e, consequentemente, a informação hereditária está codificada na sequência de bases de qualquer fita.

A dupla hélice de DNA forma espirais quando compactada dentro da célula.

2.4.4 A duplicação do DNA e a síntese protéica

O processo de duplicação do DNA ocorre através da presença da enzima DNA polimerase, onde primeiramente ocorre o rompimento das pontes de hidrogênio, com o afastamento das duas fitas. Em seguida, nucleotídeos livres na célula se encaixam nas fitas que se afastaram, respeitando a complementaridade adenina-timina e citosina-guanina. Depois de completadas as fitas originais, teremos duas moléculas de DNA idênticas entre si.

Percebe-se que cada molécula-filha conserva metade dos caracteres da molécula-mãe, em virtude da fita original. Por isso, o processo de duplicação do DNA é chamado de duplicação semiconservativa. Na Figura 2.9 vemos como ocorre o processo de duplicação do DNA.

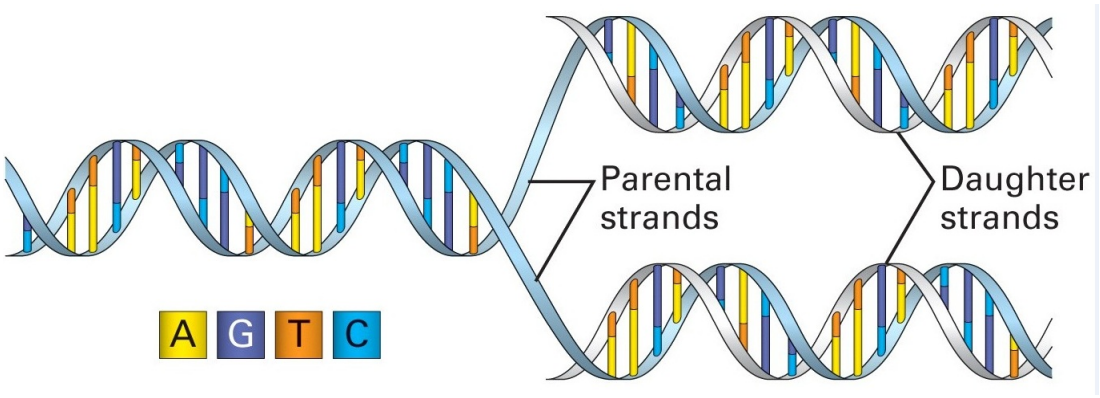

Figura 2.9: Duplicação do DNA. Lodish *et al. Molecular Cell Biology*, 5th Edition.

A partir de uma molécula de DNA gera-se moléculas de RNA mensageiro que controlam a síntese protéica após migrarem para o citoplasma. Conforme já observado, a molécula de RNA é uma fita simples, ou seja, apenas uma das fitas da dupla hélice do DNA será usada como molde para a produção do RNA.

A produção do RNA mensageiro se dá a partir da presença da enzima RNA polimerase. Com a presença dessa enzima, ocorre o rompimento das pontes de hidrogênio e o afastamento das fitas do DNA. Em seguida, nucleotídeos livres de RNA no núcleo se encaixam em uma das fitas simples do DNA, que passa a ser chamada de fita ativa. Após o encaixe, a molécula de RNA se destaca da fita molde de DNA e migra para o citoplasma, onde passará por mais um estágio antes da realização da síntese protéica. Por fim, as duas fitas de DNA tornam a se parear, voltando à molécula original.

Migrando para o citoplasma, o RNA mensageiro se liga aos ribossomos. Neles, existem três sítios: o sítio A (relacionado à entrada de aminoácidos), o sítio P (relacionado à formação do polipeptídeo) e o sítio onde se liga ao RNA mensageiro.

O RNA transportador carrega os aminoácidos, levando-os até o ribossomo, onde penetram pelo sítio A. O RNA mensageiro deve reconhecer o aminoácido que chega ao ribossomo para ocorrer a síntese protéica.

A sequência de bases que codificarão um aminoácido é chamada códon e é formado por três bases nitrogenadas. Cada códon codifica apenas um aminoácido, mas um aminoácido pode ser codificado por mais de um códon.

Os códons do RNA mensageiro são reconhecidos pelo RNA transportador, que vai se ligar a um determinado aminoácido e também vai ser reconhecido por um grupo de 3 nucleotídeos na molécula de RNA mensageiro. No RNA transportador temos o anticódon, que reconhece a posição do aminoácido no RNA mensageiro, unindo seu anticódon ao códon do RNA mensageiro.

Deslocando-se para o citoplasma, o RNA transportador se liga a aminoácidos, deslocando-os até os pontos onde ocorrerá a síntese protéica. A síntese é realizada pelas organelas chamadas ribossomos, no RNA ribossômico. Na Figura 2.10 vemos o processo de síntese protéica, desde o processo de ativação do DNA, passando pelo processo de transcrição, a retirada dos íntrons e o processo de tradução, sintetizando a proteína.

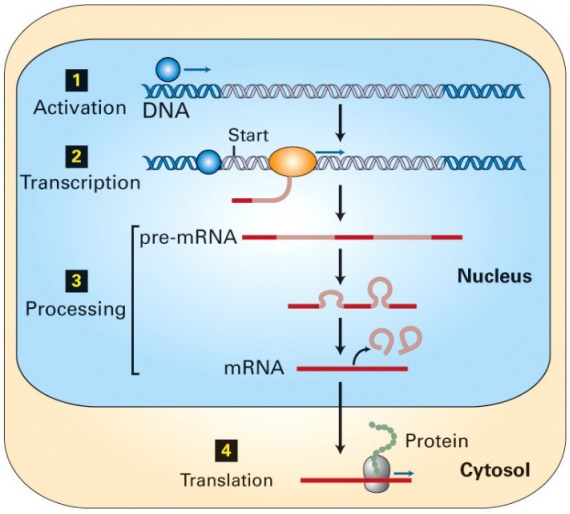

Figura 2.10: A síntese protéica. Lodish *et al. Molecular Cell Biology*, 5th Edition.

2.4.5 Aminoácidos

Os estudos modernos sobre proteínas e aminoácidos devem muito aos experimentos do século XIX e início do século XX. Os aminoácidos, por conterem carbono, são essenciais à vida e são as unidades estruturais que compõem as proteínas. Vários aminoácidos estão entre os compostos orgânicos que, acredita-se, surgiram nos primórdios da história na Terra.

Estrutura

A análise de um grande número de proteínas de quase todas as fontes conhecidas, mostrou que todas elas são compostas por 20 aminoácidos-padrão.

Os aminoácidos comuns são conhecidos como α-aminoácidos porque possuem um grupo amino primário $(-NH_2)$ ligado ao carbono α, que é o carbono próximo ao grupo carboxílico $(-COOH)$. A prolina é uma exceção, possuindo um grupo amino secundário $(-NH-)$. A Figura 2.11 a seguir apresenta a estrutura geral de um α-aminoácido.

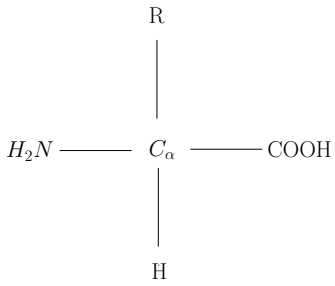

Figura 2.11: Estrutura geral de um α-aminoácido

Esses 20 aminoácidos diferentes possuem propriedades físico-químicas específicas. Se considerarmos a interação com a água como parâmetro de classificação, os aminoácidos podem ser classificados como hidrofílicos, quando "tem afinidade" com água, ou hidrofóbicos, se não "tem afinidade" de água. Como o ambiente celular é essencialmente aquoso, esta característica de hidropaticidade desempenha um papel de suma importância na determinação da estrutura tridimensional das proteínas. Os aminoácidos hidrofóbicos tendem a evitar o contato com a água, ficando dessa forma "escondidos" no interior da proteína, ao passo que os hidrofílicos distribuem-se na superfície da molécula ficando em contato com a água.

A hidropaticidade é um dos critérios de classificação dos aminoácidos, que podem ser classificados também pela carga (ácido, neutro, básico), estrutura (cíclico, acíclico), tamanho (pequeno, médio, grande), aromático, alifático. No presente trabalho a característica amplamente analisada é a hidropaticidade. As demais características serão abordadas em outra ocasião.

Dos 20 aminoácidos, 12 deles são hidrofílicos (polares) e 8 são hidrofóbicos (apolares). Na Tabela 2.10 vemos a classificação dos aminoácidos de acordo com a hidropaticidade. Além disso, vamos a abreviatura de cada um dos aminoácidos que serão utilizadas na construção da tabela do código genético.

De acordo com a fonte primária, ou seja, onde os aminoácidos são produzidos, eles podem ser classificados em essenciais ou naturais.

Os aminoácidos essenciais são sintetizados apenas pelos vegetais. São eles: isoleucina, leucina, lisina, metionina, treonina, triptofano, valina e fenilalanina.

Os naturais são produzidos tanto por animais quanto por vegetais. São eles: alanina, glicina, histidina, tirosina, arginina, ácido aspártico, asparagina, glutamina, serina, prolina, cisteína e ácido glutâmico.

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)	Metionina (Met)
Serina (Ser)	Triptofano (Trp)
Treonina (Thr))	
Ácido aspártico (Asp))	
Ácido glutâmico (Gln)	
Tirosina (Tyr)	

Tabela 2.10: Hidropaticidade dos aminoácidos

Principais aminoácidos e suas funções

No decorrer do presente trabalho, alguns dos 20 aminoácidos terão uma atenção especial, em virtude de sua aplicabilidade. Estão listados abaixo os mais utilizados e suas principais características:

1. **Glicina** - Codificado pelas trincas GGU, GGC, GGA, GGG. É o mais simples dos aminoácidos, estando presente na maioria das proteínas. Serve como precursor em diversas espécies químicas.

2. **Prolina** - Possui uma estrutura quimicamente coesa e rígida, sendo o mais rígido dos vinte que são codificados geneticamente. É codificado pelas trincas CCA, CCC, CCG, CCU.

3. **Fenilalanina** - É um composto natural presente em todas as proteínas (vegetais ou animais). É um componente essencial da dieta diária, de forma que o corpo não consegue funcionar sem sua presença. É codificado pelas trincas UUC e UUU.

4. **Lisina** - Possui cadeia lateral muito polar, ou seja, altamente hidrofílico. Auxilia no crescimento ósseo, ajudando na formação do colágeno. É codificado pelas trincas AAA e AAG.

5. **Triptofano** - É um aminoácido essencial para a nutrição humana, codificado apenas pelo códon UGG. Responsável pela produção de serotonina, ou seja, atua como antidepressivo, uma vez que eleva os níveis de serotonina, responsável pela sensação de bem estar.

6. **Tirosina** - É um aminoácido que não pode ser completamente sintetizado pelos animais. É utilizado na síntese de adrenalina. É codificado pelas trincas UAC e UAU.

2.4.6 O código genético

O DNA de cada célula de todos os organismos vivos contém, pelo menos, uma cópia (ou, raramente, várias) dos genes que carregam a informação para produzir cada proteína que o organismo necessita.

Para a transmissão dessa informação genética existem os processos de transcrição e tradução. No processo de transcrição, todos os RNAs celulares são produzidos a partir do DNA, primeiramente o RNA mensageiro, em seguida, o RNA transportador e, por fim, o RNA ribossômico.

A síntese protéica ocorre a partir do DNA, mas não diretamente do mesmo. Para tal, existe o RNA mensageiro, que serve como intermediário. A sequência de nucleotídeos do RNA mensageiro é "lida" para produzir milhares de proteínas. Esse é o processo de tradução, que ocorre nos ribossomos.

O esquema de produção de proteínas é similar a um sistema de comunicação padrão, desde o transmissor até o receptor, passando por um canal ruidoso. Nas Figuras 2.12 e 2.13 vemos uma analogia entre um sistema de comunicação e um sistema biológico, através do dogma central da teoria de comunicações e dogma central da genética.

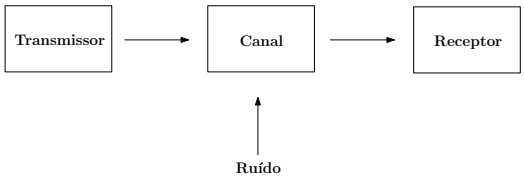

Figura 2.12: Dogma central da teoria de comunicações

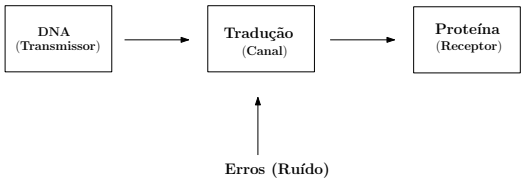

Figura 2.13: Sistema de comunicação da informação genética

Enquanto a teoria de comunicações lida com a transmissão de dados ou de informação de um ponto a outro, onde a informação a ser transmitida por meio do sistema de comunicações está sempre sujeita a um conjunto de interferências, chamadas de ruídos, na biologia a informação é perpetuada através da replicação do DNA e traduzida através da transcrição e da tradução, até a síntese protéica, porém nesses processos podem ocorrer erros, análogo aos ruídos do sistema de comunicações padrão.

As proteínas são polímeros construídos por 20 diferentes tipos de aminoácidos. A especificação dos 20 diferentes aminoácidos se dá através de 64 possíveis combinações das quatro bases nitrogenadas agrupadas três-a-três.

A síntese protéica portanto, é feita em blocos de três nucleotídeos (trincas), que são denominados códons, cada um correspondendo sempre a um mesmo aminoácido. A correspondência do códon a um determinado aminoácido se dá através do chamado código genético, apresentado na Tabela 2.11.

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

Tabela 2.11: O código genético

Três dos 64 códons determinam o fim da síntese protéica, os chamados códons STOP. Os 61 códons restantes são utilizados para codificar os 20 aminoácidos, sendo que um deles, além de codificar o aminoácido metionina, determina o início da síntese protéica, o códon AUG. Na tabela vemos os 20 aminoácidos, seus códigos e os códons correspontes.

Uma observação de extrema relevância é que o código genético é dito **universal**, ou seja, todos os organismos usam o mesmo código para traduzir suas proteínas. As únicas exceções conhecidas ocorrem em certos protozoários ciliados e nas mitocôndrias, que apresentam pequenas diferenças. As diferenças são as seguintes: em protozoários e mitocôndrias, AGA e AGG são códons de parada, ao invés de codificar arginina e, na mitocôndria, ainda existem mais duas diferenças: AUA determina

metionina em vez de isoleucina e UGA codifica triptofano em vez de STOP.

Outro fato de suma importância que pode ser observado na tabela do código genético é a de que os dois primeiros nucleotídeos dos códons de um mesmo aminoácido são geralmente os mesmos. Isso minimiza o efeito da mutação, fazendo assim com que a célula seja resistente a produção de proteínas defeituosas. Se, por exemplo, houver uma mutação no DNA resultando na troca da terceira base do códon, não mudará o aminoácido codificado.

2.4.7 Mutações

As mutações são mudanças que ocorrem na sequência de nucleotídeos do material genético de um organismo, que podem ser causadas por erros durante a divisão celular, exposição a radiação ultravioleta ou ionizante, mutagênicos químicos ou vírus.

De acordo com a alteração provocada, as mutações podem ser classificadas como desfavoráveis (ou deletérias) ou favoráveis (benéficas ou vantajosas).

Mudanças no DNA causadas por mutações podem causar erros na sequência das proteínas, criando proteínas não-funcionais. A alteração de uma proteína que desempenha papel importante no organismo pode levar ao surgimento de uma doença, chamada de doença genética. No entanto, a maioria das mutações não tem impacto na saúde.

Se uma mutação estiver presente numa célula germinal, pode originar descendentes portadores dessa mutação em todas as suas células, ocasionando as chamadas doenças hereditárias.

Em relação às mutações benéficas pode-se observar que representam uma pequena porcentagem em relação a ocorrência das mutações. Estas mutações levam a novas versões de proteínas que ajudam o organismo e futuras gerações a se adaptarem melhor a mudanças em seu ambiente.

De acordo com o efeito ocasionado na estrutura de um organismo, as mutações podem ser classificadas em **mutações de pequena escala** e **mutações de grande escala**.

- **Mutações de pequena escala** - afetam um pequeno gene ou poucos nucleotídeos.

  ⋆ **Mutação pontual**: Há a troca de um nucleotídeo por outro, podendo ser a **transição**, onde ocorre a troca de uma purina por outra purina ($A \leftrightarrow G$) ou uma pirimidina por outra pirimidina ($C \leftrightarrow U$) ou a **transversão**, onde ocorre a troca de uma purina por uma pirimidina, ou vice-versa ($C/U \leftrightarrow A/G$). Essas mutações pontuais podem ser chamadas de **silenciosas**, quando o códon modificado codifica o mesmo aminoácido, **"missense"**, quando o códon modificado codifica um aminoácido diferente e, **sem sentido**, quando codifica um códon STOP e interrompe a proteína antes de seu término.

- ⋆ **Inserção**: Ocorre quando há a adição de um ou mais nucleotídeos na sequência de DNA.
- ⋆ **Deleção**: Ocorre a remoção de um ou mais nucleotídeos da sequência de DNA.

- **Mutações de grande escala**

- ⋆ **Amplificação**: Criação de várias cópias de uma região cromossômica, aumentando a dosagem dos genes dentro dela.
- ⋆ **Inserção**:Une partes do DNA anteriormente separados, potencialmente unindo genes, de tal forma que surjam genes fundidos funcionalmente distintos.
- ⋆ **Deleção de regiões cromossômicas**: Perda dos genes presentes nas regiões.
- ⋆ **Perda de heterozigosidade**:Ocorre a perda de um alelo por deleção ou recombinação num organismo que originalmente possuia dois alelos.

As mutações podem afetar a função de uma proteína, fazendo com que ocorra **perda de função**, **ganho de função** ou até a **morte do organismo** que a possui.

As mutações são consideradas o mecanismo que permite a ação da seleção natural, já que insere a variação genética sobre a qual ela irá agir, fornecendo as novas características vantajosas que sobrevivem e se multiplicam nas gerações subsequentes ou as características deletérias que aparecem em organismos mais fracos.

2.5 Códigos Corretores de Erros

Um canal de comunicação pode apresentar uma série de interferências que dificultam a correta interpretação dos sinais transmitidos. Essas interferências se manifestam através de ruídos, imperfeições, distorções, etc. O objetivo de um sistema de comunicação é a transmissão de informação através de uma fonte para um destinatário por meio de um canal de comunicação com a maior confiabilidade possível.

A busca por bons códigos e bons conjuntos de sinais associados aos mesmos se inicia a partir do trabalho de Shannon [25] em 1948, onde provou que para taxas de transmissão de informação menor que a capacidade do canal, existe um código que permite a transmissão com probabilidade de erro arbitrariamente pequena. Com isso, deu-se o início de pesquisas em teoria de codificação.

Várias linhas de pesquisa surgiram englobando códigos lineares, não-lineares, conjunto de sinais, códigos de Slepian [28], constelações de sinais provenientes de reticulados, etc.

Em 1982, Ungerboeck [26] mostrou que, por meio do particionamento do conjunto de sinais, ganhos significativos de codificação eram obtidos. Era o surgimento da modulação codificada.

Forney [27], estendeu o conceito de Ungerboeck, englobando os códigos de Slepian e os códigos reticulados, apresentando uma nova classe de códigos chamados **códigos geometricamente uniformes**.

Os códigos lineares constituem uma classe importante de códigos por possuírem uma estrutura algébrica, a qual facilita o processo de decodificação. No caso dos não-lineares, a falta de uma estrutura algébrica aumenta a complexidade do processo de decodificação.

Neste trabalho um conceito amplamente discutido é o dos códigos geometricamente uniformes, os quais possuem diversas propriedades simétricas de vital relevância: todas as regiões de Voronoi são congruentes; o perfil de distâncias é o mesmo para qualquer palavra-código; as palavras-código possuem a mesma probabilidade de erro; e o grupo gerador é isomorfo a um grupo de permutações que atua transitivamente nas palavras-código.

Nesta seção apresentamos os principais conceitos dos códigos de blocos na subseção 2.5.1, os códigos geometricamente uniformes, com as principais definições, exemplos e teoremas, na subseção 2.5.2 e por fim, na subseção 2.5.3 o casamento de conjunto de sinais a um grupo.

2.5.1 Códigos de bloco

Os códigos de bloco são caracterizados pelo fato do processo de codificação ser feito sobre blocos de bits ou bloco de símbolos. Isso quer dizer que 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.

A taxa de codificação de um código de bloco é definida como a relação entre o número de bits de informação e o número de bits da palavra-código, ou seja, $R = k/n$.

Um código de bloco linear binário é um subespaço vetorial com 2^k vetores do espaço vetorial constituído de todos os 2^n vetores com n elementos de $\{0,1\}$, apresentado na Figura 2.14. Considerando as duas condições necessárias de caracterização de um subespaço vetorial aplicadas aos códigos de blocos lineares, temos:

- a soma de duas palavras-códigos quaisquer resulta em outra palavra-código.
- o vetor nulo ou vetor todo zero é também uma palavra-código.

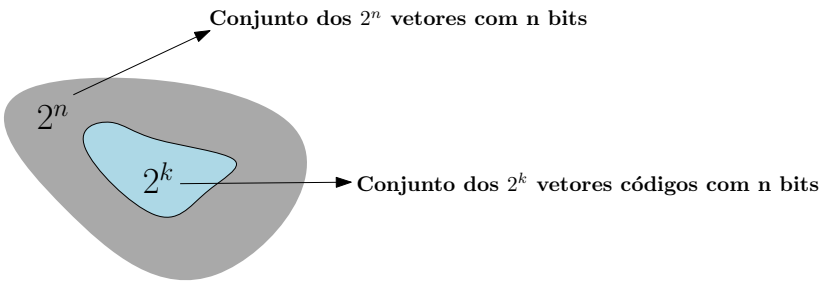

Figura 2.14: Representação dos códigos de blocos lineares como um subespaço vetorial de um espaço vetorial V_n

Nota-se que o subespaço vetorial constitui um conjunto dos vetores códigos ou vetores válidos. Portanto, qualquer vetor de comprimento n que não pertença ao subespaço vetorial está no espaço vetorial, porém, é um vetor não válido.

Definição 2.8 *[29]A distância de Hamming entre dois vetores $x = x_1...x_n$ e $y_1...y_n$ é o número de posições em que eles se diferenciam e denotado por $dist(x,y)$.*

Definição 2.9 *[29]O peso de Hamming de um vetor $x = x_1...x_n$ é o número de coordenadas não-nulas de x_i e denotado por $w_t(x)$.*

Exemplo 12 $dist(10111,00101) = 2$, $w_t(101110) = 4$

2.5.2 Códigos geometricamente uniformes

As definições e resultados aqui apresentados estão em [27].

Definição 2.10 *[27] Seja $\mathbb{S}$ um conjunto de sinais em um espaço métrico $(\mathbb{M},d)$. Dizemos que $\mathbb{S}$ é um* ***código geometricamente uniforme*** *se para quaisquer s_1 e $s_2 \in \mathbb{S}$, existe uma isometria μ_{s_1,s_2}, tal que:*

$$\mu_{s_1,s_2}(s_1 = s_2),\ e\ \mu_{s_1,s_2}(\mathbb{S}) = \mathbb{S}$$

Em outras palavras, a ação do grupo de simetrias, $\Gamma(\mathbb{S})$, de $\mathbb{S}$ é transitiva. Se $\mathbb{S}$ for finito, dizemos que $\mathbb{S}$ é uma **constelação uniforme** e se $\mathbb{S}$ for infinito dizemos que é um **arranjo regular**.

Em geral, o grupo de simetrias de um conjunto de sinais geometricamente uniforme possui mais elementos do que o necessário para gerá-lo. Para isso, consideremos a seguinte definição.

Definição 2.11 *[27] Seja $\mathbb{S}$ um código geometricamente uniforme. Um grupo gerador mínimo $U(\mathbb{S})$ de $\mathbb{S}$, é um subgrupo do grupo de simetrias de $\mathbb{S}$ que satisfaz:*

$\forall s_0 \in \mathbb{S}$, $\mathbb{S} = \{\mu(s_0), \mu \in U(\mathbb{S})\}$, *e a função* $m : U(\mathbb{S}) \longrightarrow \mathbb{S}$, *dada por* $m(\mu) = \mu(s_0)$ *é injetora.*

Exemplo 13 *[24]A constelação de sinais* $M-PSK$ *é um código geometricamente uniforme. De fato, seu grupo de simetrias é* $\mathbb{S}_M$*, isto é, o grupo de permutação de* M *elementos, e um grupo gerador natural é o subgrupo das rotações* R_M *que é isomorfo a* $\mathbb{Z}_M$*.*

Teorema 2.5.1 *[27] O produto cartesiano de conjuntos de sinais geometricamente uniformes é um conjunto de sinais geometricamente uniforme.*

Definição 2.12 *[27] Seja* $\mathbb{S}$ *um conjunto de sinais geometricamente uniforme com grupo gerador mínimo* $U(\mathbb{S})$*. Uma partição geometricamente uniforme* $\mathbb{S}/\mathbb{S}'$*, é uma partição de* $\mathbb{S}$ *induzida por um subgrupo normal* $\mathbb{U}'$ *de* $\mathbb{U}(\mathbb{S})$*. Os elementos de* $\mathbb{S}/\mathbb{S}'$ *são os subconjuntos de* $\mathbb{S}$ *que correspondem às classes laterais de* $\mathbb{U}'$ *em* $U(\mathbb{S})$*.*

Definição 2.13 *[27] Sejam* $\mathbb{S}/\mathbb{S}'$ *uma partição geometricamente uniforme e* $\mathbb{G}$ *um grupo isomorfo a* $\mathbb{S}/\mathbb{S}'$*. Um* ***rotulamento isométrico*** *é uma função injetora* $m : \mathbb{G} \longrightarrow \mathbb{S}/\mathbb{S}'$ *dada pela composição do isomorfismo entre* $\mathbb{G}$ *e* $U(\mathbb{S})/U'(\mathbb{S}')$ *e a função injetora induzida por m de* $U(\mathbb{S})/U'(\mathbb{S}')$ *em* $\mathbb{S}/\mathbb{S}'$*.*

Exemplo 14 *[24]Sejam* $U(\mathbb{S})$ *e* $U'(\mathbb{S})$ *dados por* $U(\mathbb{S}) = \{e, r, r^2, r^3\}$ *e* $U'(\mathbb{S}) = \{e\}$*. Então* $\mathbb{S}/\mathbb{S}' = \{00, 10, 11, 01\}$

Como $U(\mathbb{S})/U'(\mathbb{S})$ é isomorfo a $\mathbb{Z}_4$, este isomorfismo nos leva ao rotulamento isométrico dos elementos de ${\mathbb{Z}_2}^2$, sob a distância de Hamming e dos elementos de $\mathbb{Z}_4$ sob a métrica de Lee, apresentados na Tabela 2.12.

00	$\rightarrow$	0
10	$\rightarrow$	1
11	$\rightarrow$	2
01	$\rightarrow$	3

Tabela 2.12: Rotulamento isométrico

2.5.3 Conjunto de sinais casados a grupos

Conjunto de sinais casado a grupos, [28], é a forma mais adequada de considerar o modulador e o codificador como um único bloco, associando a cada elemento da palavra-código um sinal a ser transmitido. As definições apresentadas a seguir estão em [28].

Definição 2.14 *[28] Seja $(\mathbb{M}, d)$ um espaço métrico. Dizemos que um* ***conjunto de sinais*** *finito S em $\mathbb{M}$ está* ***casado a um grupo*** *G se existe uma função sobrejetora $\mu : G \longrightarrow S$, tal que:*

$$d(\mu(g), \mu(g')) = d(\mu(g^{-1} * g'), \mu(e)),\ \forall g, g' \in G,$$

onde e é o elemento neutro de G. A função μ é denominada ***mapeamento casado****. Se μ é uma injetora, então μ^{-1} é chamada* ***rotulamento casado****.*

Definição 2.15 *[28] Um mapeamento casado $\mu : G \longrightarrow S$, tal que H é um subgrupo de G (ou seja, $H = \mu^{-1}\mu(e)$, onde e é o elemento neutro em G), e não contém subgrupos normais não triviais de G, é chamado* ***mapeamento efetivamente casado****. Nesse caso, dizemos que S está* ***efetivamente casado*** *a G.*

Capítulo 3

Reticulados Booleanos Algébricos e Diagramas de Hasse

Diversos modelos matemáticos são utilizados no estudo das características e propriedades do código genético. Os reticulados booleanos algébricos, bem como os diagramas de Hasse associados aos mesmos representam um desses modelos e são ferramentas eficazes na análise de algumas propriedades associadas ao código genético.

Através da construção dos reticulados booleanos algébricos podemos observar as ligações entre as bases nitrogenadas da estrutura do DNA por meio de operações booleanas. Efetuada a construção dos reticulados booleanos, os mesmos fundamentam a construção dos diagramas de Hasse, estruturas matemáticas que facilitam a ordenação dos códons do código genético, uma vez que os mesmos são dispostos de forma organizada, através de alguns critérios, propiciando uma melhor visualização desses códons e consequências biológicas.

Os diagramas de Hasse representam uma rica estrutura matemática, representando um caso de POSET(conjunto parcialmente ordenado), que podem ser utilizados para diversos estudos, como a representação de estruturas secundárias e terciárias de proteínas.

A relevância da construção dos reticulados booleanos algébricos e dos diagramas de Hasse é pelo fato dessas estruturas refletirem o papel biológico e físico-químico na distribuição dos códons associados a cada aminoácido. Características como a hidropaticidade dos aminoácidos podem ser analisadas através da construção dessas estruturas.

Através do rotulamento dos nucleotídeos adenina, citosina, guanina, timina/uracila das sequências de DNA, denotados por $N = \{A, C, G, T/U\}$ em um alfabeto 4-ário $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, para a estrutura de anel, obtém-se um conjunto de 24 permutações $N \rightarrow \mathbb{Z}_4$, as quais podem ser classificadas em três rotulamentos A, B e C, relacionados às formas geométricas que geram, apresentado em [7]. Em [1],

Sanchez et al. utilizam um mapeamento que reflete o rotulamento A, obtendo para o mesmo resultados algébricos e biológicos. Propomos a construção dessas estruturas para os rotulamentos B e C, com o objetivo de analisar as características biológicas e algébricas associadas aos mesmos, bem como uma comparação com o modelo construído para o rotulamento A.

Este capítulo está organizado da seguinte maneira: na seção 3.1 apresentamos, de forma detalhada, o modelo proposto em [1] e os resultados desse estudo. Na seção 3.2 apresentamos a proposta de construção dos reticulados booleanos algébricos e o diagrama de Hasse para o rotulamento B, através de um rótulo selecionado de maneira randômica das 8 permutações associadas a esse rotulamento. Na seção 3.3, apresentamos a construção dos reticulados booleanos e do diagrama de Hasse para o rotulamento C, utilizando também um rótulo aleatório das 8 permutações possíveis. Na seção 3.4 são apresentados alguns comentários e resultados acerca da montagem dessas estruturas.

3.1 Reticulados Boolenos Algébricos e Diagramas de Hasse Associados ao Rotulamento A

Considerando a complementaridade biológica das bases nitrogenadas adenina(A), timina/uracila(T/U), guanina(G) e citosina(C), Sanchez et al. em [1] apresentam a construção de um reticulado Booleano e um diagrama de Hasse, que refletem as propriedades físico-químicas dos aminoácidos, os quais cada trinca codifica.

Para a construção de tais reticulados, Sanchez et al. afirmam que é necessário a existência de dois elementos não comparáveis, de um elemento máximo e um elemento mínimo. Esses reticulados booleanos, chamados primal e dual darão origem ao diagrama de Hasse, considerando a seguinte associação:

$$U \rightarrow UUU \qquad C \rightarrow CCC \qquad G \rightarrow GGG \qquad A \rightarrow AAA$$

O critério utilizado em [1] para a construção dos reticulados booleanos e respectivo diagrama de Hasse foi o seguinte:

- os códons GGG e CCC serão considerados como elementos máximo / mínimo, respectivamente, pois codificam aminoácidos com pequenas diferenças, a glicina e a prolina. Isto permite a comparação entre os mesmos.

- os códons UUU e AAA serão considerados como elementos não-comparáveis, uma vez que mesmo tendo o mesmo número mínimo de pontes de hidrogênio, eles codificam aminoácidos com polaridades extremas, como a fenilalanina e a lisina.

Através das informações apresentadas anteriormente, é feita a construção de dois reticulados booleanos, chamados primal e dual. Primeiramente, o elemento máximo no reticulado primal será a citosina(C) e o elemento mínimo a guanina(G). No reticulado dual, teremos a guanina(G) como elemento máximo e a citosina(C) como elemento mínimo. Desta forma, teremos dois reticulados booleanos - $(B(X),\vee,\wedge)$ - reticulado primal e $(B'(X),\wedge,\vee)$ - reticulado dual, onde $X=\{U,C,G,A\}$.

A Figura 3.1 apresenta os reticulados primal e dual. O isomorfismo entre esses reticulados com o reticulado booleano $((\mathbb{Z}_2)^2,\vee,\wedge)$ e $((\mathbb{Z}_2)^2,\wedge,\vee)$, onde $\mathbb{Z}_2=\{0,1\}$,$\vee$ - conectivo lógico ou(disjunção) e $\wedge$ - conectivo lógico e(conjunção), nos permite fazer a seguinte representação: $G\leftrightarrow 00, A\leftrightarrow 01, U\leftrightarrow 10, C\leftrightarrow 11$. Para o reticulado dual, temos: $C\leftrightarrow 00, U\leftrightarrow 01, A\leftrightarrow 10, G\leftrightarrow 11$.

Em [7] e [8] observamos um rotulamento dessas representações binárias através de um alfabeto 4-ário, $\mathbb{Z}_4=\{0,1,2,3\}$. A representação binária possui as seguintes associações:

$$00\rightarrow 0 \qquad 10\rightarrow 1 \qquad 11\rightarrow 2 \qquad 01\rightarrow 3$$

Essa associação de um alfabeto 4-ário para um alfabeto binário é detalhada em [8], através dos conceitos da G-linearidade.

Desta forma, podemos perceber que em [1] foi utilizado o rotulamento: $G\leftrightarrow 0, A\leftrightarrow 3, U\leftrightarrow 1, C\leftrightarrow 2$ para o reticulado primal e $C\leftrightarrow 0, U\leftrightarrow 3, A\leftrightarrow 1, G\leftrightarrow 2$ para o reticulado dual.

Conforme já foi apresentado, as 24 permutações entre os elementos do alfabeto genético $N=\{A,C,G,T/U\}$ com os do alfabeto 4-ário $\mathbb{Z}_4=\{0,1,2,3\}$ geram 3 rotulamentos, chamados em [7] de rotulamentos A, B e C, cada um com uma característica geométrica.

Observa-se que as representações dos reticulados primal e dual fazem parte do rotulamento A. São elas: 3201 para o caso primal e 1023 para o caso 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, sempre a primeira base adenina (A), em seguida citosina (C), a seguir guanina (G) e por fim, uracila (U).

Os reticulados primal e dual do rotulamento A, apresentados em [1] são mostrados na Figura 3.1.

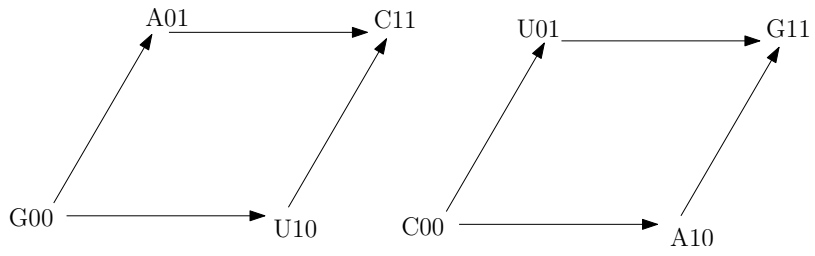

Figura 3.1: Reticulados booleanos primal e dual

Os reticulados booleanos representados na Figura 3.1 foram construídos de acordo com as operações da álgebra booleana, capítulo 2, e apresentados nas Tabelas 3.1 e 3.2.

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

Tabela 3.1: Primal (ou)

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

Tabela 3.2: Primal (e)

O caso dual é apresentado através das Tabelas 3.3 e 3.4

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

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

Tabela 3.3: Dual (ou)

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

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

Tabela 3.4: Dual (e)

Através dos reticulados booleanos primal e dual foi construído um diagrama de Hasse, tomando conforme já dito, um elemento máximo, um elemento mínimo e dois elementos não comparáveis. A Tabela 3.5 ilustra o diagrama de Hasse referente ao rotulamento A. As propriedades de simetria deste

diagrama são determinadas pela função $NOT : XYZ \rightarrow\sim (XYZ)$, de modo que se as bases complementares de $X_1X_2X_3$ são as bases $X_1'X_2'X_3'$ ($X_i, X_i' \in \{A,C,G,U\}, i = 1,2,3$), então a imagem do códon ${}^{5'}X_1X_2X_3{}^{3'}$ é ${}^{3'}X_1X_2X_3{}^{5'}$.

						GGG										
					GUG	UGG	GGU	GGA	AGG	GAG						
		UUG	GUU	GUA	AUG	UGU	UGA	CGG	GCG	GGC	AGU	AGA	UAG	GAU	GAA	AAG		
UUU	UUA	AUU	CUG	AUA	GUC	GCU	UGC	CGU	UCG	AGC	GCA	ACG	CGA	CAG	UAU	GAC	UAA	AAU	AAA
		UUC	CUU	CUA	AUC	UCU	UCA	CCG	CGC	GCC	ACU	ACA	UAC	CAU	CAA	AAC		
						CUC	UCC	CCU	CCA	ACC	CAC							
									CCC							

Tabela 3.5: Diagrama de Hasse segundo o rotulamento A

Observamos na Tabela 3.5 que a imagem simétrica de um códon que codifica um aminoácido hidrofílico (códons com A na segunda posição) é sempre um códon que codifica aminoácido hidrofóbico (códons com U na segunda posição). Considere a cadeia $(GUG, UGG, GGU, GGA, AGG, GAG)$. A imagem da mesma é a anti-cadeia $(CAC, ACC, CCA, CCU, UCC, CUC)$, tendo os elementos um-a-um uma imagem, ou seja, a imagem de GUG é CAC e assim por diante.

Podemos observar no diagrama de Hasse do rotulamento A, representado na Tabela 3.5 que os códons hidrofóbicos e hidrofílicos estão separados nas cadeias laterais do diagrama, representados em vermelho e azul, respectivamente.

Esse resultado se refere a um caso do rotulamento A. A seguir, veremos um exemplo para o rotulamento B e um exemplo para o rotulamento C e a consequente implicação biológica na construção dos reticulados booleanos e dos diagramas de Hasse.

3.2 Modelo Proposto para o Rotulamento B

Selecionamos um rótulo dentre as oito permutações referentes ao rotulamento B. O rótulo escolhido foi $ACGU = 1230$. Utilizando a notação binária, temos $A = 10, C = 11, G = 01, U = 00$. Este rótulo dará origem ao reticulado booleano primal. O reticulado booleano dual será construído a partir do rótulo $ACGU = 3012$, ou seja, $A = 01, C = 00, G = 10, U = 11$. Como temos ACGU para o primal, o dual será UGCA, portanto, o rótulo 2103. As Tabelas 3.6 e 3.7 ilustram as operações ou e e. Nas Tabelas 3.8 e 3.9 apresentadas a seguir vemos o caso dual.

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

Tabela 3.6: Primal (ou)

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

Tabela 3.7: Primal (e)

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

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

Tabela 3.8: Dual (ou)

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

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

Tabela 3.9: Dual (e)

Os reticulados booleanos referentes a essas tabelas são ilustrados na Figura 3.2

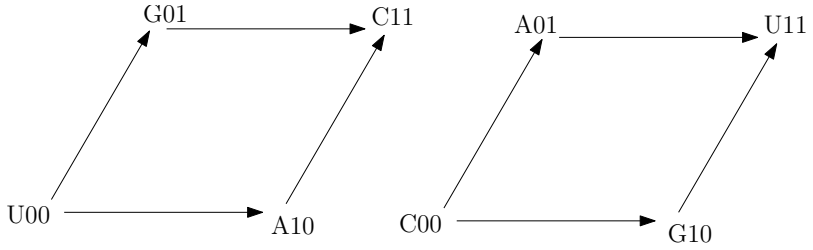

Figura 3.2: Reticulados booleanos primal e dual

No modelo em consideração, observa-se que os elementos mínimo/máximo dos reticulados são a uracila(U) e a citosina(C), respectivamente, ambas bases nitrogenadas pirimidinas, que biologicamente não são ligadas entre si por pontes de hidrogênio, a menos de uma mutação. Ao considerarmos as trincas UUU e CCC, as mesmas codificam aminoácidos com poucas diferenças, a prolina e a fenilalanina, respectivamente.

Portanto, o diagrama de Hasse desse rotulamento segue uma relação de complementaridade algébrica, ou seja, $00-11$ e $01-10$. Desta forma, as bases nitrogenadas referentes a essa complementaridade algébrica são: $U-C$ e $G-A$.

Até onde é de nosso conhecimento, o diagrama de Hasse referente ao rotulamento B, bem como os reticulados booleanos primal e dual ainda não foram apresentados na literatura.

A disposição de códons do código genético no diagrama de Hasse com relação ao rotulamento B, é mostrada na Tabela 3.10.

								CCC										
						CGC	GCC	CCG	CCA	ACC	CAC							
		GGC	CGG	CGA	AGC	GCG	GCA	UCC	CUC	CCU	ACG	ACA	GAC	CAG	CAA	AAC		
GGG	GGA	AGG	UGC	AGA	CGU	CUG	GCU	UCG	GUC	ACU	CUA	AUC	UCA	UAC	GAG	CAU	GAA	AAG	AAA
		GGU	UGG	UGA	AGU	GUG	GUA	UUC	UCU	CUU	AUG	AUA	GAU	UAG	UAA	AAU		
						UGU	GUU	UUG	UUA	AUU	UAU							
								UUU										

Tabela 3.10: Diagrama de Hasse segundo o rotulamento B

Observando a Tabela 3.10, podemos perceber que a união das bases se dá por meio da complementaridade algébrica. Se tomarmos como exemplo a cadeia $CGC, GCC, CCG, CCA, ACC, CAC$, sua imagem será a anti-cadeia $UGU, GUU, UUG, UUA, AUU, UAU$. Note que a primeira cadeia é lida no sentido $5'-3'$ e a imagem da mesma é lida no sentido $3'-5'$ e a complementaridade das bases se dá da seguinte forma: guanina(G) com adenina(A) e citosina(C) com uracila(U), ou seja, purina com purina e pirimidina com pirimidina.

Ao contrário do que ocorre no rotulamento A, onde são separados códons hidrofóbicos e hidrofílicos nas cadeias laterais, no rotulamento B observamos, com exceção dos códons STOP e do códon UGG, que codifica o aminoácido triptofano, os demais aminoácidos das cadeias laterais são todos códons hidrofílicos (em vermelho), deixando os hidrofóbicos na parte central do diagrama.

3.3 Modelo Proposto para o Rotulamento C

Assim como fizemos para o rotulamento B, selecionamos um rótulo dentre as oito permutações do rotulamento C e construímos os reticulados booleanos primal e dual e o respectivo diagrama de Hasse.

O rótulo selecionado foi $ACGU = 3102$, cuja representação binária é $01,10,00,11$. O dual será o rótulo $UGCA = 0231$, que em binário é $00,11,01,10$.

Nas Tabelas 3.11 e 3.12 temos as operações booleanas que culminarão com a montagem dos reticulados booleanos primal e dual e o diagrama de Hasse correspondente. Nas Tabelas 3.13 e 3.14 vemos o caso dual.

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

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

Tabela 3.11: Primal (ou)

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

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

Tabela 3.12: Primal (e)

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

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

Tabela 3.13: Dual (ou)

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

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

Tabela 3.14: Dual (e)

No rotulamento C, os elementos mínimo/máximo são a uracila(U) e a guanina(G), uma purina e uma pirimidina, respectivamente. As trincas UUU e GGG codificam aminoácidos com características bem distintas, a fenilalanina e a glicina. Efetuaremos agora a montagem dos reticulados booleanos referentes às Tabelas 3.11 a 3.14. Os mesmos se encontram na Figura 3.3.

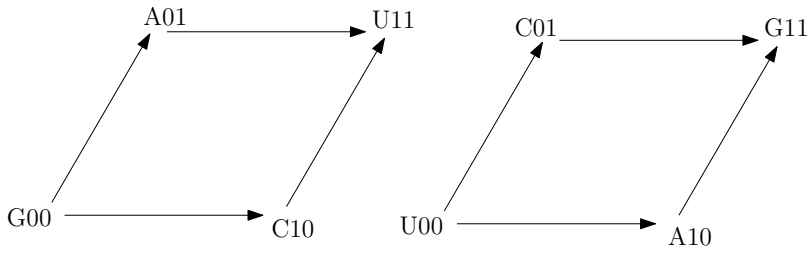

Figura 3.3: Reticulados booleanos primal e dual

Finalmente, construímos o diagrama de Hasse do rotulamento C, como mostra a Tabela 3.15.

								UUU										
						UAU	AUU	UUA	UUC	CUU	UCU							
		AAU	UAA	UAC	CAU	AUA	AUC	GUU	UGU	UUG	CUA	CUC	ACU	UCA	UCC	CCU		
AAA	AAC	CAA	GAU	CAC	UAG	UGA	AUG	GUA	AGU	CUG	UGC	CGU	GUC	GCU	ACA	UCG	ACC	CCA	CCC
		AAG	GAA	GAC	CAG	AGA	AGC	GGU	GUG	UGG	CGA	CGC	ACG	GCA	GCC	CCG		
						GAG	AGG	GGA	GGC	CGG	GCG							
								GGG										

Tabela 3.15: Diagrama de Hasse segundo o rotulamento C

Analisando a complementaridade das bases nitrogenadas do rotulamento C, notamos a união de uma base purina com uma pirimidina, ou seja, adenina(A) com citosina(C) e uracila(U) com guanina(G). Vamos tomar como exemplo a cadeia $UAU, AUU, UUA, UUC, CUU, UCU$. Sua imagem no diagrama de Hasse é encontrada através da complementaridade das bases, portanto, será a anti-cadeia $GAG, AGG, GGA, GGC, CGG, GCG$. No diagrama de Hasse desse rotulamento, vemos na cadeia lateral esquerda códons hidrofílicos, representados em vermelho e na cadeia lateral direita metade dos códons são hidrofóbicos e a outra metade hidrofílicos, ou seja, não mantém uma regularidade como nos rotulamentos A e B.

3.4 Análise dos Resultados

Através da construção dos reticulados booleanos e dos diagramas de Hasse para os rotulamentos B e C, algumas interpretações biológicas poderão ser apresentadas.

Iniciaremos com algumas análises relacionadas no rotulamento B. No mesmo existe uma regularidade na separação dos códons do código genético, uma vez que como pode-se observar no diagrama, os códons em vermelho possuem características hidrofílicas, com exceção dos códons STOP e do códon UGG, que codifica o aminoácido triptofano. Os códons em preto possuem características hidrofóbicas, localizando-se na parte central do diagrama.

No rotulamento C, ao contrário do rotulamento B, onde existe uma regularidade na separação dos códons pela hidropaticidade, os códons não mantém essa regularidade, o que ocorre é a separação dos códons nas classes laterais do diagrama de acordo com a segunda base, em uma das laterais vemos códons com a adenina (A) sendo a segunda base, na outra códons com a citosina (C) sendo a segunda base. Este fato também ocorre no rotulamento B, porém as bases que ocupam a segunda posição do códon são adenina (A) e guanina (G).

O critério biológico analisado é a hidropaticidade dos aminoácidos, apresentado no capítulo 2. Para o rotulamento A, vemos a separação dos códons hidrofóbicos e hidrofílicos nas cadeias laterais do diagrama. Essa regularidade na separação dos códons também ocorre para rotulamento B, porém para o rotulamento C a mesma não se mantém.

Algebricamente, podemos observar que a complementaridade das bases nitrogenadas para os rotulamentos B e C não obedece a regra de pareamento biológica adenina-timina, citosina-guanina, conforme ocorre no rotulamento A, onde ocorre o casamento da complementaridade biológica e matemática. Nota-se para os rotulamentos B e C o pareamento matemático (00-11) e (01-10), que não coincide com o pareamento biológico, observa-se no rotulamento B o pareamento uracila (U) - citosina e guanina (G) - adenina (A) e para o rotulamento C o pareamento adenina (A) - citosina (C) e guanina (G) - uracila (U).

No caso do rotulamento A observa-se o casamento algébrico, biológico e regularidade na separação dos códons em hidrofóbicos e hidrofílicos, fato que não ocorre nos rotulamentos B e C, onde essas três características não estão casadas.

Para o rotulamento A, a construção baseou-se inicialmente no contexto biológico, considerando a complementaridade das bases nitrogenadas e atrelado a essa complementaridade associou-se o mapeamento algébrico. Para os rotulamentos B e C, considerou-se inicialmente a complementaridade algébrica e a ela associou-se uma complementaridade biológica. Isso influenciou na construção das

estruturas dos reticulados booleanos algébricos e diagramas de Hasse.

O casamento entre o contexto biológico e o contexto matemático, classificando os mapeamentos em não-linear e linear é detalhado em [8], fundamentando a construção dos reticulados booleanos e dos diagramas de Hasse.

Capítulo 4

Operações Algébricas no Código Genético

A representação do código genético é feita geralmente através de uma tabela com quatro colunas, onde os códons são localizados de acordo com a segunda base.

Modelar algebricamente o código genético é algo que diversos autores têm pesquisado no intuito de identificar suas propriedades, características e implicações biológicas. Uma modelagem matemática muito interessante se baseia no cálculo associado aos códons do código genético, como por exemplo, a soma entre códons e a separação dos códons pela paridade.

Sanchez et al. em [9] realizam a construção do código genético através de um mapeamento que reflete o rotulamento B, usando a associação dos elementos do conjunto $N = \{A, C, G, T/U\}$ com os elementos do anel $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Essa associação identifica uma estrutura de espaço vetorial associada ao código genético. Desta forma, podemos realizar operações usando esta estrutura de espaço vetorial, de maneira combinatorial, de forma que essa estrutura matemática seja capaz de identificar algumas sequências de DNA, ou seja, no bloco codificador estamos identificando o rotulamento, que casado ao código corretor de erros adequeado gera a sequência de DNA desejada. Observe a Figura 4.1.

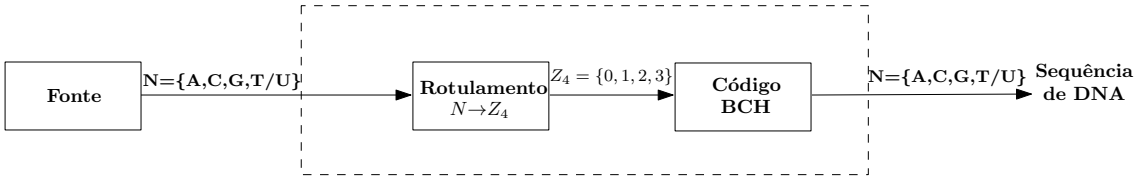

Figura 4.1: Codificador genético

A associação binária de cada um dos elementos de $\mathbb{Z}_4$ permite-nos relacioná-lo a uma modulação $4-PSK$ conforme Figura 4.2.

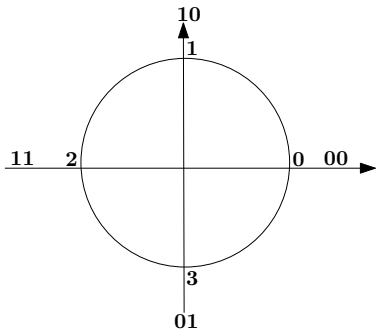

Figura 4.2: Conjunto de sinais $4-PSK$ e sua representação binária

Como as estruturas algébricas do codificador e do modulador (constelação de sinais) são isomorfas, pode-se dizer que esse isomorfismo é uma mapeamento casado (MC), conforme apresentado em [8], satisfazendo a propriedade dos códigos geometricamente uniformes.

Utilizando as operações sobre o anel $\mathbb{Z}_4$, em [9] é proposta uma operação soma entre códons usando dois procedimentos distintos: um levando em consideração o número correspondente a cada códon, por meio do anel $\mathbb{Z}_{64}$ e outro através de um algoritmo que tem como fundamentação um critério biológico. Além disso, realizam a separação dos códons por meio da paridade dos mesmos. Propomos uma operação soma com transporte, fazendo uso dessa estrutura de espaço vetorial, que diminui a complexidade do algoritmo proposto em [9], uma vez que resumimos o mesmo a uma simples operação de soma com transporte de 1 em $\mathbb{Z}_4$. Além disso, estendemos esses cálculos para os rotulamentos A e C, efetuando a soma entre códons através dos três métodos (soma em $\mathbb{Z}_{64}$, algoritmo biológico e algoritmo proposto). Outro aspecto pesquisado foi o comportamento dos aminoácidos no código genético em cada um dos rotulamentos, usando assim como fizemos no capítulo 3, o critério biológico de hidropaticidade.

Através dessas operações vemos a utilização da estrutura de espaço vetorial, relacionando as estruturas matemáticas e biológicas. O que é realizado no mundo matemático se reflete no mundo biológico.

Além disso, identificamos uma estrutura de grupo associada a esse contexto biológico e a soma realizada entre códons por meio dessa estrutura, torna-se ferramenta eficaz em análises mutacionais. Por exemplo, um organismo atacado por uma bactéria pode sofrer alterações em sua estrutura genética, causada pela alteração de um ou mais códons, podendo ser uma mutação silenciosa, como também uma mutação que acarrete a troca do aminoácido codificado.

Este capítulo está organizado da seguinte maneira: na seção 4.1 apresentamos o modelo utilizado

em [9], com a montagem do código genético associado ao rotulamento B, bem como a operação soma entre códons através dos métodos propostos em [9]. Em seguida, apresentamos o método soma com transporte em consideração. Na seção 4.2 mostramos a aplicação do método soma com transporte, bem como adaptamos o algoritmo com características biológicas para os rotulamentos A e C. Além disso, comentamos sobre a paridade dos códons nos rotulamentos A e C. Na seção 4.3 apresentamos um estudo sobre o comportamento dos aminoácidos em cada um dos rotulamentos de acordo com o critério de hidropaticidade.

4.1 Soma Algébrica no Código Genético usando o Rotulamento B

Em [9] é realizada a construção do código genético considerando a seguinte ordem das bases nitrogenadas $\{A,C,G,U\}$ e $\{U,G,C,A\}$ para os casos primal e dual, respectivamente. A relação entre a ordenação do conjunto das bases nitrogenadas com os elementos do anel $\mathbb{Z}_4 = \{0,1,2,3\}$, implica no rotulamento B, uma vez que teremos para o caso primal o rótulo 0123 e para o caso dual o rótulo 3210. Realizada a construção do código genético para os casos primal e dual, define-se uma operação soma no conjunto das quatro bases do DNA, conforme mostrada na Tabela 4.1, considerando a ordem do anel $\mathbb{Z}_4 = \{0,1,2,3\}$.

$+$	A	C	G	U
A	A	C	G	U
C	C	G	U	A
G	G	U	A	C
U	U	A	C	G

$+$	U	G	C	A
U	U	G	C	A
G	G	C	A	U
C	C	A	U	G
A	A	U	G	C

Tabela 4.1: Tabelas soma - primal e dual

As Tabelas 4.3 e 4.4 ilustram a construção do código genético para o rotulamento B. Essas tabelas contém a representação dos códons, o aminoácido codificado por cada um desses códons, bem como a bijeção do grupo abeliano do código genético com o anel $\mathbb{Z}_{64}$.

Antes de efetuar a construção do código genético, vemos listados os 20 aminoácidos e suas respectivas abreviaturas na Tabela 4.2, os quais serão utilizados na construção das tabelas do código genético para cada um dos rotulamentos.

As hipóteses consideradas são: a ordem dos códons deve refletir as propriedades físico-químicas dos mesmos, bem como a posição da base no códon e a interação códon-anticódon, onde tipos químicos (purina e pirimidina), bem como o número de pontes de hidrogênio tem uma regra impor-

tante. A importância da posição da base é sugerida pela frequência de erros encontrada nos códons. Erros na terceira base são mais frequentes que na primeira, e, por sua vez, estes são mais frequentes que erros na segunda base. Portanto, em ordem de importância das bases temos: a segunda, a primeira e a terceira base.

Aminoácido	Abreviatura	Aminoácido	Abreviatura
Alanina	A	Leucina	L
Arginina	R	Lisina	K
Asparagina	N	Metionina	M
Ácido Aspártico	D	Fenilalanina	F
Cisteína	C	Prolina	P
Glutamina	Q	Serina	S
Ácido Glutâmico	E	Treonina	T
Glicina	G	Triptofano	W
Histidina	H	Tirosina	Y
Isoleucina	I	Valina	V
STOP	UAG, UGA, UAA		

Tabela 4.2: Aminoácidos

Depois de efetuar a construção da tabela da operação soma do código genético e as tabelas do código genético primal e dual, será apresentada uma operação soma entre dois códons XYZ e $X'Y'Z'$.

Dois métodos foram utilizados em [9] para efetuar a soma entre códons: o algoritmo soma em $\mathbb{Z}_{64}$ e o algoritmo SMG (Sanchez, Morgado e Grau).

Observando o algoritmo SMG (Sanchez, Morgado e Grau), vemos que o mesmo se baseia em uma operação soma com transporte de 1. Desta forma, com o intuito de otimizar os cálculos da soma entre códons propomos o algoritmo soma com transporte.

	A			C			G			U			
	nº	I	II	nº	I	II	nº	I	II	nº	I	II	
A	0	AAA	K	16	ACA	T	32	AGA	R	48	AUA	I	A
	1	AAC	N	17	ACC	T	33	AGC	S	49	AUC	I	C
	2	AAG	K	18	ACG	T	34	AGG	R	50	AUG	M	G
	3	AAU	N	19	ACU	T	35	AGU	S	51	AUU	I	U
C	4	CAA	Q	20	CCA	P	36	CGA	R	52	CUA	L	A
	5	CAC	H	21	CCC	P	37	CGC	R	53	CUC	L	C
	6	CAG	Q	22	CCG	P	38	CGG	R	54	CUG	L	G
	7	CAU	H	23	CCU	P	39	CGU	R	55	CUU	L	U
G	8	GAA	E	24	GCA	A	40	GGA	G	56	GUA	V	A
	9	GAC	D	25	GCC	A	41	GGC	G	57	GUC	V	C
	10	GAG	E	26	GCG	A	42	GGG	G	58	GUG	V	G
	11	GAU	D	27	GCU	A	43	GGU	G	59	GUU	V	U
U	12	UAA	STOP	28	UCA	S	44	UGA	STOP	60	UUA	L	A
	13	UAC	Y	29	UCC	S	45	UGC	C	61	UUC	F	C
	14	UAG	STOP	30	UCG	S	46	UGG	W	62	UUG	L	G
	15	UAU	Y	31	UCU	S	47	UGU	C	63	UUU	F	U

Tabela 4.3: Código genético segundo o rotulamento B - primal

U				G			C			A			
	nº	I	II	nº	I	II	nº	I	II	nº	I	II	
U	0	UUU	F	16	UGU	C	32	UCU	S	48	UAU	Y	U
	1	UUG	L	17	UGG	W	33	UCG	S	49	UAG	STOP	G
	2	UUC	F	18	UGC	C	34	UCC	S	50	UAC	Y	C
	3	UUA	L	19	UGA	STOP	35	UCA	S	51	UAA	STOP	A
G	4	GUU	V	20	GGU	G	36	GCU	A	52	GAU	D	U
	5	GUG	V	21	GGG	G	37	GCG	A	53	GAG	E	G
	6	GUC	V	22	GGC	G	38	GCC	A	54	GAC	D	C
	7	GUA	V	23	GGA	G	39	GCA	A	55	GAA	E	A
C	8	CUU	L	24	CGU	R	40	CCU	P	56	CAU	H	U
	9	CUG	L	25	CGG	R	41	CCG	P	57	CAG	Q	G
	10	CUC	L	26	CGC	R	42	CCC	P	58	CAC	H	C
	11	CUA	L	27	CGA	R	43	CCA	P	59	CAA	Q	A
A	12	AUU	I	28	AGU	S	44	ACU	T	60	AAU	N	U
	13	AUG	M	29	AGG	R	45	ACG	T	61	AAG	K	G
	14	AUC	I	30	AGC	S	46	ACC	T	62	AAC	N	C
	15	AUA	I	31	AGA	R	47	ACA	T	63	AAA	K	A

Tabela 4.4: Código genético segundo o rotulamento B - dual

4.1.1 Algoritmo Soma com Transporte

O algoritmo soma com transporte é descrito da seguinte maneira:

Passo 1 - Especificar o rótulo utilizado.

Passo 2 - Somar as terceiras bases em $\mathbb{Z}_4$. Se o valor encontrado for superior a 3, transporte 1 para a próxima soma.

Passo 3 - Somar as primeiras bases em $\mathbb{Z}_4$, adicionando 1, caso a soma do passo 2 seja superior a 3. Se o valor encontrado for superior a 3, transporte 1 para a próxima soma.

Passo 4 - Somar as segundas bases em $\mathbb{Z}_4$, adicionando 1, caso a soma anterior seja superior a 3.

Para somas isoladas entre códons, caso a soma do passo 4 seja superior a 3, não existirá mais transporte. Porém, se a soma estiver sendo feita entre códons em uma proteína, o transporte é feito para o próximo códon, considerando a ordem de soma das bases nitrogenadas, ou seja, somando-se as terceiras bases, em seguidas as primeiras bases e por fim, as segundas bases.

O método proposto se aplica aos rotulamentos A, B e C, uma vez que estamos trabalhando com o anel $\mathbb{Z}_4$ e qualquer permutação não irá alterar a soma.

Observe os exemplos utilizando os três métodos para soma de códons no código genético para o rotulamento B (0123), caso primal.

Exemplo 15 *Efetuar a seguinte soma: GAA + CUC, usando o algoritmo soma com transporte.*

Passo 1 - $\{A,C,G,U\} \to \{0,1,2,3\}$. *Então, temos:* $200 + 131$

Passo 2 - $0+1=1$ *mod*4 - *soma das terceiras bases.*

Passo 3 - $2+1=3$ *mod*4 - *soma das primeiras bases.*

Passo 4 - $0+3=3$ *mod*4 - *soma das segundas bases.*

Neste caso não foi necessário utilizar o transporte de 1, uma vez que em todas as somas os valores foram menores ou iguais a 3. Em outros exemplos, o transporte será necessário.

Portanto, $200+131=331$*, que corresponde ao códon UUC.*

4.1.2 Algoritmo Soma em $\mathbb{Z}_{64}$

O algoritmo soma em $\mathbb{Z}_{64}$ é descrito da seguinte maneira:

Passo 1 - Especificar o rótulo utilizado.

Passo 2 - Identificar o elemento de $\mathbb{Z}_{64}$ correspondente aos códons, multiplicando a primeira posição por 4^1, a segunda posição por 4^2 e a terceira posição por 4^0 e, em seguida, somando esse produto, ou seja, obedecendo a ordem de importância biológica das bases nitrogenadas no códon.

Passo 3 - Somar os elementos correspondentes a cada códon reduzidos módulo 64.

Exemplo 16 *Efetuar a seguinte soma: GAA + CUC, usando o algoritmo soma em* $\mathbb{Z}_{64}$*.*

Passo 1 - $\{A,C,G,U\} \to \{0,1,2,3\}$*. Então, temos:* $200+131$

Passo 2 - $GAA = 2\cdot 4^1+0\cdot 4^2+0\cdot 4^0=8$ *e* $CUC = 1\cdot 4^1+3\cdot 4^2+1\cdot 4^0=53$

Passo 3 - $8+53=61$ *mod*64*, que corresponde ao códon UUC.*

Portanto, $GAA+CUC=UUC$*.*

4.1.3 Algoritmo SMG (Sanchez, Morgado e Grau)

O algoritmo SMG, proposto em [9] é descrito da seguinte maneira:

Passo 1 - as bases correspondendo a terceira posição são adicionadas de acordo com a tabela soma.

Passo 2 - se a base resultante da operação soma é anterior à base adicionada (a ordem no conjunto de bases), então o novo valor é escrito e a base C(ou G, se o dual for usado), é adicionado à próxima posição.

Passo 3 - as outras bases são adicionadas de acordo com a tabela soma, passo 2, indo da primeira para a segunda base.

Note que a operação realizada de soma entre dois códons leva em consideração a importância das bases. A operação soma entre dois códons é obtida a partir da base menos importante biologicamente (terceira posição) para a base mais importante (segunda posição).

Exemplo 17 *Considere a soma entre os códons UCU e CAC. Usando o algoritmo SMG, temos:*

Passo 1 *- $A+C=C$ (soma das bases da terceira posição). Como estamos utilizando o rótulo $\{A,C,G,U\}$ e a base C não antecede a base A, efetuamos a soma das primeiras bases.*

Passo 2 *- $G+C=U$ (soma das bases da primeira posição). Assim como no caso anterior, a base U não antecede as bases G e C. Desta forma, somaremos as segundas bases.*

Passo 3 *- $A+U=U$ (soma das bases da segunda posição).*

Desta forma, a soma entre UCU e CAC resulta no códon UUC.

4.2 Aplicação das Operações no Código Genético para os Rotulamentos A e C

Utilizamos procedimento análogo ao rotulamento B na construção do código genético para os rotulamentos A e C, bem como efetuamos operações soma envolvendo códons desses dois rotulamentos, utilizando os três métodos (algoritmo soma em $\mathbb{Z}_{64}$, algoritmo SMG adaptado para os rotulamentos A e C e algoritmo soma com transporte).

Escolhemos de maneira aleatória o rótulo 3201 para o caso primal do rotulamento A. Logo, o rótulo dual será 0132.

Como temos a bijeção entre as bases $\{G,U,C,A\}$ e o anel $\mathbb{Z}_4=\{0,1,2,3\}$, define-se uma operação soma no conjunto das quatro bases do DNA (primal e dual). As tabelas soma para o rótulo A selecionado (caso primal e dual) estão dispostas a na Tabela 4.5:

$+$	G	U	C	A
G	G	U	C	A
U	U	C	A	G
C	C	A	G	U
A	A	G	U	C

$+$	A	C	U	G
A	A	C	U	G
C	C	U	G	A
U	U	G	A	C
G	G	A	C	U

Tabela 4.5: Tabelas soma - primal e dual

Observe a construção do código genético para os rótulos selecionados do rotulamento A para o caso primal e dual. No mesmo estão a representação numérica, os códons e os aminoácidos que codificam cada um dos códons. O código genético para os casos citados estão apresentados nas Tabelas 4.6 e 4.7.

De posse do código genético, bem como as tabelas soma, vamos efetuar a soma de códons, usando para isso os três métodos empregados para o rotulamento B.

Exemplo 18 *Efetuar a seguinte soma: UCU + CAC, usando a tabela primal.*

1º - Algoritmo soma com transporte

Passo 1 - $\{G,U,C,A\} \to \{0,1,2,3\}$. *Então temos:* $121+232$

Passo 2 - $1+2=3$ *mod*4 - *soma das terceiras bases.*

Passo 3 - $1+2=3$ *mod*4 - *soma das primeiras bases.*

Passo 4 - $2+3=1$ *mod*4 - *soma das segundas bases.*

Portanto, $121+232=313$*, que corresponde ao códon AUA.*

2º - Algoritmo soma em $\mathbb{Z}_{64}$

Passo 1 - $\{G,U,C,A\} \to \{0,1,2,3\}$. *Então temos:* $121+232$.

Passo 2 - $UCU = 1\cdot 4^1+2\cdot 4^2+1\cdot 4^0=37$ *e* $CAC = 2\cdot 4^1+3\cdot 4^2+2\cdot 4^0=58$.

Passo 3 - $37+58=31$ *mod*64*, que corresponde ao códon AUA.*

Portanto, $UCU+CAC=AUA$.

3º - Algoritmo SMG (Sanchez, Morgado e Grau) *- vamos fazer uma adaptação do algoritmo desenvolvido em [9].*

Passo 1 *-* $U+C=A$ *(soma das bases da terceira posição). Como estamos trabalhando com o rótulo* $\{G,U,C,A\}$ *e a base A não antecede a base C, passemos a soma das primeiras bases.*

Passo 2 *-* $U+C=A$ *(soma das bases da primeira posição). Idem caso anterior. Vamos efetuar a soma das segundas bases.*

Passo 3 *-* $C+A=U$ *(soma das bases da segunda posição).*

Dessa forma, a soma entre UCU e CAC resulta no códon AUA.

Exemplo 19 *Efetuar a seguinte soma:* $ACU+UUG$*, usando a tabela dual.*

1º - Algoritmo soma com transporte

Passo 1 - $\{A,C,U,G\} \to \{0,1,2,3\}$. *Então temos:* $012+223$.

Passo 2 - $2+3=1$ *mod*4 - *soma das terceiras bases.*

Passo 3 - $0+2+1=3$ *mod*4 - *soma das primeiras bases, com transporte de 1.*

Passo 4 - $1+2=3$ *mod*4 - *soma das segundas bases.*

Portanto, $012+223=331$*, que corresponde ao códon GGC.*

2º - Algoritmo soma em $\mathbb{Z}_{64}$

Passo 1 - $\{A,C,U,G\} \to \{0,1,2,3\}$. *Então temos:* $012+223$.

Passo 2 - $ACU = 0\cdot 4^1+1\cdot 4^2+2\cdot 4^0=18$ *e* $UUG = 2\cdot 4^1+2\cdot 4^2+3\cdot 4^0=43$.

Passo 3 - $18+43=61$ *mod*64*, que corresponde ao códon GGC.*

Portanto, $ACU + UUG = GGC$.

3º - Algoritmo SMG (Sanchez, Morgado e Grau) *- vamos fazer uma adaptação do algoritmo desenvolvido em [9].*

Passo 1 *- $U + G = C$ (soma das bases da terceira posição). Como estamos utilizando o rótulo $\{A,C,U,G\}$ e a base C antecede a base G, acrescentamos a base C (elemento 1)na próxima soma.*

Passo 2 *- $A + U = U + C = G$ (soma das bases da primeira posição com o elemento 1, representado pela base C). A base G não antecede C.*

Passo 3 *- $C + U = G$ (soma das bases da segunda posição).*

Logo, temos $ACU + UUG = GGC$

Exemplo 20 *Efetuar a seguinte soma: $GUU + GAA$, usando a tabela primal.*

1º - Algoritmo soma com transporte

Passo 1 *- $\{G,U,C,A\} \to \{0,1,2,3\}$. Então temos:* $011 + 033$.

Passo 2 *- $1 + 3 = 0\ mod 4$ - soma das terceiras bases.*

Passo 3 *- $0 + 0 + 1 = 1\ mod 4$ - soma das primeiras bases, com transporte de 1.*

Passo 4 *- $1 + 3 = 0\ mod 4$ - soma das segundas bases.*

Portanto, $011 + 033 = 100$*, que corresponde ao códon UGG.*

2º - Algoritmo soma em $\mathbb{Z}_{64}$

Passo 1 *- $\{G,U,C,A\} \to \{0,1,2,3\}$. Então temos:* $011 + 033$.

Passo 2 *- $GUU = 0 \cdot 4^1 + 1 \cdot 4^2 + 1 \cdot 4^0 = 17$ e $GAA = 0 \cdot 4^1 + 3 \cdot 4^2 + 3 \cdot 4^0 = 51$.*

Passo 3 *- $17 + 51 = 4\ mod 64$, que corresponde ao códon UGG.*

Portanto, $GUU + GAA = UGG$.

3º - Algoritmo SMG (Sanchez, Morgado e Grau) *- vamos fazer uma adaptação do algoritmo desenvolvido em [9].*

Passo 1 *- $U + A = G$ (soma das bases da terceira posição). Como estamos utilizando o rótulo $\{G,U,C,A\}$ e a base G antecede as base U e A, acrescentamos a base U (elemento 1)na próxima soma.*

Passo 2 *- $G + G = G + U = U$ (soma das bases da primeira posição com o elemento 1, representado pela base U). A base U não antecede G.*

Passo 3 *- $U + A = G$ (soma das bases da segunda posição).*

Logo, temos $GUU + GAA = UGG$

G				U			C			A			
	nº	I	II	nº	I	II	nº	I	II	nº	I	II	
G	0	GGG	G	16	GUG	V	32	GCG	A	48	GAG	E	G
	1	GGU	G	17	GUU	V	33	GCU	A	49	GAU	D	U
	2	GGC	G	18	GUC	V	34	GCC	A	50	GAC	D	C
	3	GGA	G	19	GUA	V	35	GCA	A	51	GAA	E	A
U	4	UGG	W	20	UUG	L	36	UCG	S	52	UAG	STOP	G
	5	UGU	C	21	UUU	F	37	UCU	S	53	UAU	Y	U
	6	UGC	C	22	UUC	F	38	UCC	S	54	UAC	Y	C
	7	UGA	STOP	23	UUA	L	39	UCA	S	55	UAA	STOP	A
C	8	CGG	R	24	CUG	L	40	CCG	P	56	CAG	Q	G
	9	CGU	R	25	CUU	L	41	CCU	P	57	CAU	H	U
	10	CGC	R	26	CUC	L	42	CCC	P	58	CAC	H	C
	11	CGA	R	27	CUA	L	43	CCA	P	59	CAA	Q	A
A	12	AGG	R	28	AUG	M	44	ACG	T	60	AAG	K	G
	13	AGU	S	29	AUU	I	45	ACU	T	61	AAU	N	U
	14	AGC	S	30	AUC	I	46	ACC	T	62	AAC	N	C
	15	AGA	R	31	AUA	I	47	ACA	T	63	AAA	K	A

Tabela 4.6: Código genético segundo o rotulamento A - primal

A				C			U			G			
	nº	I	II	nº	I	II	nº	I	II	nº	I	II	
A	0	AAA	K	16	ACA	T	32	AUA	I	48	AGA	R	A
	1	AAC	N	17	ACC	T	33	AUC	I	49	AGC	S	C
	2	AAU	N	18	ACU	T	34	AUU	I	50	AGU	S	U
	3	AAG	K	19	ACG	T	35	AUG	M	51	AGG	R	G
C	4	CAA	Q	20	CCA	P	36	CUA	L	52	CGA	R	A
	5	CAC	H	21	CCC	P	37	CUC	L	53	CGC	R	C
	6	CAU	H	22	CCU	P	38	CUU	L	54	CGU	R	U
	7	CAG	Q	23	CCG	P	39	CUG	L	55	CGG	R	G
U	8	UAA	STOP	24	UCA	S	40	UUA	L	56	UGA	STOP	A
	9	UAC	Y	25	UCC	S	41	UUC	F	57	UGC	C	C
	10	UAU	Y	26	UCU	S	42	UUU	F	58	UGU	C	U
	11	UAG	STOP	27	UCG	S	43	UUG	L	59	UGG	W	G
G	12	GAA	E	28	GCA	A	44	GUA	V	60	GGA	G	A
	13	GAC	D	29	GCC	A	45	GUC	V	61	GGC	G	C
	14	GAU	D	30	GCU	A	46	GUU	V	62	GGU	G	U
	15	GAG	E	31	GCG	A	47	GUG	V	63	GGG	G	G

Tabela 4.7: Código genético segundo o rotulamento A - dual

Para o rotulamento C, escolhemos, também de forma aleatória o rótulo 1320 para o caso primal. Seu dual será o rótulo 2013.

No rotulamento C temos a bijeção $\{U,A,G,C\}$ com o anel $\mathbb{Z}_4 = \{0,1,2,3\}$ para o caso primal e a bijeção $\{C,G,A,U\}$ com o anel $\mathbb{Z}_4 = \{0,1,2,3\}$ para o caso dual. Define-se uma operação soma no conjunto das quatro bases do DNA (primal e dual). A Tabela 4.8 apresenta a soma para o rótulo C selecionado (caso primal e dual).

$+$	U	A	G	C
U	U	A	G	C
A	A	G	C	U
G	G	C	U	A
C	C	U	A	G

$+$	C	G	A	U
C	C	G	A	U
G	G	A	U	C
A	A	U	C	G
U	U	C	G	A

Tabela 4.8: Tabelas soma - primal e dual

Nas Tabelas 4.9 e 4.10 temos a construção do código genético para os rótulos selecionados do rotulamento C (primal e dual). Assim como no caso dos rotulamentos A e B, observa-se a representação numérica, os códons e os aminoácidos codificados por cada um dos códons.

Vamos efetuar a soma entre dois códons aplicando os três algoritmos utilizados nos rotulamentos A e B.

Exemplo 21 *Efetuar a seguinte soma: $AGA + GCG$ para o caso primal.*

1º - Algoritmo soma com transporte

Passo 1 - $\{U,A,G,C\} \to \{0,1,2,3\}$. *Então temos:* $121 + 232$.

Passo 2 - $1+2 = 3\ mod 4$ - *soma das terceiras bases.*

Passo 3 - $1+2 = 3\ mod 4$ - *soma das primeiras bases.*

Passo 4 - $2+3 = 1\ mod 4$ - *soma das segundas bases.*

Portanto, $121 + 232 = 313$*, que corresponde ao códon CAC.*

2º - Algoritmo soma em $\mathbb{Z}_{64}$

Passo 1 - $\{U,A,G,C\} \to \{0,1,2,3\}$. *Então temos:* $121 + 232$.

Passo 2 - $AGA = 1 \cdot 4^1 + 2 \cdot 4^2 + 1 \cdot 4^0 = 37$ *e* $GCG = 2 \cdot 4^1 + 3 \cdot 4^2 + 2 \cdot 4^0 = 58$.

Passo 3 - $37 + 58 = 31\ mod 64$*, que corresponde ao códon CAC.*

Portanto, $AGA + GCG = CAC$.

3º - Algoritmo SMG (Sanchez, Morgado e Grau) *- vamos fazer uma adaptação do algoritmo desenvolvido em [9] para o rotulamento C.*

Passo 1 *-* $A + G = C$ *(soma das bases da terceira posição). Como estamos utilizando o rótulo* $\{U,A,G,C\}$ *e a base C não antecede a base G, vamos somar as primeiras bases.*

Passo 2 *-* $A + G = C$ *(soma das bases da primeira posição). Idem caso anterior. Vamos agora somar as segundas bases.*

Passo 3 *-* $G + C = A$ *(soma das bases da segunda posição).*

Portanto, $AGA + GCG = CAC$.

U				A			G			C			
	nº	I	II	nº	I	II	nº	I	II	nº	I	II	
U	0	UUU	F	16	UAU	Y	32	UGU	C	48	UCU	S	U
	1	UUA	L	17	UAA	STOP	33	UGA	STOP	49	UCA	S	A
	2	UUG	L	18	UAG	STOP	34	UGG	W	50	UCG	S	G
	3	UUC	F	19	UAC	Y	35	UGC	C	51	UCC	S	C
A	4	AUU	I	20	AAU	N	36	AGU	S	52	ACU	T	U
	5	AUA	I	21	AAA	K	37	AGA	R	53	ACA	T	A
	6	AUG	M	22	AAG	K	38	AGG	R	54	ACG	T	G
	7	AUC	I	23	AAC	N	39	AGC	S	55	ACC	T	C
G	8	GUU	V	24	GAU	D	40	GGU	G	56	GCU	A	U
	9	GUA	V	25	GAA	E	41	GGA	G	57	GCA	A	A
	10	GUG	V	26	GAG	E	42	GGG	G	58	GCG	A	G
	11	GUC	V	27	GAC	D	43	GGC	G	59	GCC	A	C
C	12	CUU	L	28	CAU	H	44	CGU	R	60	CCU	P	U
	13	CUA	L	29	CAA	Q	45	CGA	R	61	CCA	P	A
	14	CUG	L	30	CAG	Q	46	CGG	R	62	CCG	P	G
	15	CUC	L	31	CAC	H	47	CGC	R	63	CCC	P	C

Tabela 4.9: Código genético segundo o rotulamento C - primal

C				G			A			U			
	nº	I	II	nº	I	II	nº	I	II	nº	I	II	
C	0	CCC	P	16	CGC	R	32	CAC	H	48	CUC	L	C
	1	CCG	P	17	CGG	R	33	CAG	Q	49	CUG	L	G
	2	CCA	P	18	CGA	R	34	CAA	Q	50	CUA	L	A
	3	CCU	P	19	CGU	R	35	CAU	H	51	CUU	L	U
G	4	GCC	A	20	GGC	G	36	GAC	D	52	GUC	V	C
	5	GCG	A	21	GGG	G	37	GAG	E	53	GUG	V	G
	6	GCA	A	22	GGA	G	38	GAA	E	54	GUA	V	A
	7	GCU	A	23	GGU	G	39	GAU	D	55	GUU	V	U
A	8	ACC	T	24	AGC	S	40	AAC	N	56	AUC	I	C
	9	ACG	T	25	AGG	R	41	AAG	K	57	AUG	M	G
	10	ACA	T	26	AGA	R	42	AAA	K	58	AUA	I	A
	11	ACU	T	27	AGU	S	43	AAU	N	59	AUU	I	U
U	12	UCC	S	28	UGC	C	44	UAC	Y	60	UUC	F	C
	13	UCG	S	29	UGG	W	45	UAG	STOP	61	UUG	L	G
	14	UCA	S	30	UGA	STOP	46	UAA	STOP	62	UUA	L	A
	15	UCU	S	31	UGU	C	47	UAU	Y	63	UUU	F	U

Tabela 4.10: Código genético segundo o rotulamento C - dual

4.3 Comportamento dos Aminoácidos no Código Genético

Um outro ponto de extrema relevância a ser discutido é o comportamento que os aminoácidos possuem em cada um dos rotulamentos. Nas tabelas do código genético apresentadas anteriormente, vemos em uma das colunas as representações dos aminoácidos codificados por trincas.

O critério utilizado para análise é a hidropaticidade, ou seja, separar os aminoácidos com comportamento hidrofóbico dos aminoácidos com comportamento hidrofílico.

Para realizar esta análise foi construída uma tabela contendo os 20 aminoácidos, suas abreviaturas,

bem como onde os mesmos aparecem nos rotulamentos A, B e C, para os casos primal e dual, como mostra a Tabela 4.11.

Aminoácido	Abreviatura	Rotulamento A	Rotulamento B	Rotulamento C
Alanina	A	32,33,34,35 / 28,29,30,31	24,25,26,27 / 36,37,38,39	56,57,58,59 / 4,5,6,7
Arginina	R	8,9,10,11,12,15 / 48,51,52,53,54,55	32,34,36,37,38,39 / 24,25,26,27,29,31	37,38,44,45,46,47 / 16,17,18,19,25,26
Asparagina	N	61,62 / 1,2	1,3 / 60,62	20,23 / 40,43
Ácido aspártico	D	49,50 / 13,14	9,11 / 52,54	24,27 / 36,39
Cisteína	C	5,6 / 57,58	45,47 / 16,18	32,35 / 28,31
Glutamina	Q	56,59 / 4,7	4,6 / 57,59	29,30 / 33,34
Ácido glutâmico	E	48,51 / 12,15	8,10 / 53,55	25,26 / 37,38
Glicina	G	0,1,2,3 / 60,61,62,63	40,41,42,43 / 20,21,22,23	40,41,42,43 / 20,21,22,23
Histidina	H	57,58 / 5,6	5,7 / 56,58	28,31 / 32,35
Isoleucina	I	29,30,31 / 32,33,34	48,49,51 / 12,14,15	4,5,7 / 56,58,59
Leucina	L	20,23,24,25,26,27 / 36,37,38,39,40,43	52,53,54,55,60,62 / 1,3,8,9,10,11	1,2,12,13,14,15 / 48,49,50,51,61,62
Lisina	K	60,63 / 0,3	0,2 / 61,63	21,22 / 41,42
Metionina	M	28 / 35	50 / 13	6 / 57
Fenilalanina	F	21,22 / 41,42	61,63 / 0,2	0,3 / 60,63
Prolina	P	40,41,42,43 / 20,21,22,23	20,21,22,23 / 40,41,42,43	60,61,62,63 / 0,1,2,3
Serina	S	13,14,36,37,38,39 / 24,25,26,27,49,50	28,29,30,31,33,35 / 28,30,32,33,34,35	36,39,48,49,50,51 / 12,13,14,15,24,27
Treonina	T	44,45,46,47 / 16,17,18,19	16,17,18,19 / 44,45,46,47	52,53,54,55 / 8,9,10,11
Triptofano	W	4 / 59	46 / 17	34 / 29
Tirosina	Y	53,54 / 9,10	13,15 / 48,50	16,19 / 44,47
Valina	V	16,17,18,19 / 44,45,46,47	56,57,58,59 / 4,5,6,7	8,9,10,11 / 52,53,54,55
STOP	STOP	7,52,55 / 8,11,56	12,14,44 / 19,49,51	17,18,33 / 30,45,46

Tabela 4.11: Aminoácidos nos rotulamentos A, B e C

Além desta tabela, uma outra foi construída no intuito de classificar os aminoácidos codificados em cada um dos 64 códons do código genético. Na Tabela 4.12, temos uma coluna com a representação numérica das trincas, além de três colunas, relativas aos rotulamentos A, B e C, e a classificação do aminoácido correspondente a cada número em cada um dos rotulamentos como hidrofóbico ou hidrofílico. Usaremos a letra L para representar os aminoácidos hidrofílicos e a letra B para os aminoácidos hidrofóbicos. Além disso, usaremos a palavra STOP para os códons de finalização.

Rep. Num.	Rot. A	Rot. B	Rot. C	Rep. Num.	Rot. A	Rot. B	Rot. C	Rep. Num.	Rot. A	Rot. B	Rot. C
0	L	L	B	22	B	B	L	43	B	L	L
1	L	L	B	23	B	B	L	44	L	STOP	L
2	L	L	B	24	B	B	L	45	L	L	L
3	L	L	B	25	B	B	L	46	L	B	L
4	B	L	B	26	B	B	L	47	L	L	L
5	L	L	B	27	B	B	L	48	L	B	L
6	L	L	B	28	B	L	L	49	L	B	L
7	STOP	L	B	29	B	L	L	50	L	B	L
8	L	L	B	30	B	L	L	51	L	B	L
9	L	L	B	31	B	L	L	52	STOP	B	L
10	L	L	B	32	B	L	L	53	L	B	L
11	L	L	B	33	B	L	STOP	54	L	B	L
12	L	STOP	B	34	B	L	B	55	STOP	B	L
13	L	L	B	35	B	L	L	56	L	B	B
14	L	STOP	B	36	L	L	L	57	L	B	B
15	L	L	B	37	L	L	L	58	L	B	B
16	B	L	L	38	L	L	L	59	L	B	B
17	B	L	STOP	39	L	L	L	60	L	B	B
18	B	L	STOP	40	B	L	L	61	L	B	B
19	B	L	L	41	B	L	L	62	L	B	B
20	B	B	L	42	B	L	L	63	L	B	B
21	B	B	L								

Tabela 4.12: Hidropaticidade dos aminoácidos

Observamos que nos três rotulamentos existe uma certa regularidade na separação dos aminoácidos hidrofóbicos e hidrofílicos. Uma observação extremamente interessante é sobre o aminoácido triptofano(W), que é hidrofóbico, e nos três rotulamentos aparece isolado, entre dois aminoácidos hidrofílicos. Esse aminoácido é codificado apenas pelo códon UGG, como podemos observar na Tabela

4.12, o mesmo aparece no número 4 do rotulamento A, no número 46 do rotulamento B e no número 34 do rotulamento C.

Outro ponto de destaque é o fato de que em todos os rotulamentos o códon que codifica o triptofano está próximo do códon STOP, ou seja, uma mutação na última base desse aminoácido finalizaria a síntese de uma proteína .

Outro ponto relevante a ser destacado é em relação ao aminoácido tirosina. O mesmo é codificado pelas trincas UAU e UAC. Nos três rotulamentos este aminoácido aparece muito próximo de códons de finalização (STOP). Biologicamente, a mesma análise mutacional feita para o triptofano pode ser feita para a tirosina.

Biologicamente, tanto o triptofano quanto a tirosina, aminoácidos em destaque na análise, ajudam na sensação de bem-estar, evitando o stress. Ambos são passíveis de uma mutação na última base e a finalização da síntese de uma proteína, podendo acarretar mudanças na mesma, desde sua estrutura até sua funcionalidade.

4.4 Análise dos Resultados

Fizemos a construção do código genético para os rotulamentos A e C, fazendo a bijeção do alfabeto biológico $N = \{A, C, G, T/U\}$ com o alfabeto matemático $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Por meio dessa associação, identificamos uma estrutura de espaço vetorial associada ao código genético, relacionada a uma modulação $4 - PSK$. Esse conjunto de sinais casado a estrutura do codificador identifica um código geometricamente uniforme. Notamos que podemos efetuar as operações de soma entre códons para os rotulamentos A, B e C utilizando de três processos: o algoritmo soma com transporte, o qual propusemos, usando a operação soma com transporte de 1 no anel $\mathbb{Z}_4$, o algoritmo soma em $\mathbb{Z}_{64}$ e o algoritmo SMG (Sanchez, Morgado e Grau), o qual adaptamos para os rotulamentos A e C. Além disso, por meio da construção das tabelas do código genético e das somas efetuadas entre os códons, identificamos uma estrutura de grupo ligada a esse contexto biológico.

Além da operação soma entre códons, uma outra observação importante a ser feita no estudo do código genético é a separação dos códons pela paridade, ou seja, separá-los em pares e ímpares. Para o caso primal do rotulamento B, os códons que terminam em A e G são pares e os que terminam em U e C são ímpares. Para o caso dual ocorre uma inversão, códons terminados em U e C são pares e terminados em A e G são ímpares.

Para o rotulamento A, códons terminados em G e C são pares e terminados em A e U são ímpares (caso primal) e códons terminados em A e U são pares e terminados em G e C são ímpares (caso dual).

A mesma análise foi feita para o rotulamento C. Para o mesmo, códons terminados em U e G são pares e terminados em C e A são ímpares (caso primal) e códons terminados em C e A são pares e terminados em U e G são ímpares (caso dual).

Analisando cada um dos rotulamentos e as especificações dos códons serem pares ou ímpares, podemos observar uma ligação com os reticulados booleanos e diagramas de Hasse apresentados no capítulo anterior. Nos mesmos, observamos o pareamento algébrico entre as bases, ou seja, a complementaridade algébrica.

Observando o rotulamento A, caso primal temos nos reticulados booleanos e diagramas de Hasse a complementaridade biológica em consonância com a complementaridade algébrica, uma vez que temos o pareamento A - U e G - C. Pela análise feita anteriormente, os códons terminados em A e U são ímpares e terminados em G e C são pares.

A mesma análise pode ser feita para os rotulamentos B e C, onde os pareamentos são algébricos, não coincidindo com o pareamento biológico. Em ambos os casos, vemos uma forte ligação entre a análise algébrica/biológica feita pelos reticulados booleanos e diagramas de Hasse e a análise feita através da montagem do código genético.

No rotulamento B, temos a complementaridade algébrica, onde uracila se une com citosina e guanina se une com adenina. Observando a análise de separação de códons pares e ímpares, vemos que códons terminados em A e G são pares e terminados em U e C são ímpares, para o caso primal.

Analisando o rotulamento C, vemos a complementaridade guanina - uracila e adenina - citosina. Analisando a paridade para o rotulamento C, vemos que códons terminados em G e U são pares e terminados em C e A são ímpares, também para o caso primal.

Desta forma, vemos que o pareamento das bases nitrogenadas nos rotulamentos A, B e C levam em consideração a separação dos códons em pares e ímpares.

O comportamento dos aminoácidos nos rotulamentos A, B e C foi outro ponto de extrema relevância, onde tomamos como objeto de estudo a hidropaticidade em relação a cada um dos rotulamentos, bem como o destaque para dois aminoácidos, o triptofano e a tirosina, devido a suas posições nas tabelas do código genético para cada um dos rotulamentos.

Notamos uma regularidade na separação dos códons hidrofóbicos e hidrofílicos nas tabelas do código genético, esta análise pôde ser observada na montagem dos diagramas de Hasse no capítulo 3.

Os aminoácidos triptofano e tirosina merecem destaque pela proximidade de ambos em todos os rotulamentos, ao códon STOP, mostrando que uma mutação na última base desses aminoácidos ocasiona a troca dos mesmos por um códon STOP, finalizando o processo de síntese protéica.

Capítulo 5

Representação Polinomial dos Códons

Além das abordagens apresentadas acerca da análise algébrica do código genético através dos diagramas de Hasse, dos reticulados booleanos algébricos e da utilização de um espaço vetorial determinada pelo anel $\mathbb{Z}_4$ para a realização de operações envolvendo códons para os rotulamentos A, B e C, uma outra abordagem de extrema relevância é a utilização de uma representação polinomial dos códons.

O objetivo deste capítulo é apresentar uma abordagem polinomial da representação dos códons, de maneira a determinar um espaço vetorial, onde pode ser realizada uma série de operações e cálculos de forma estruturada e organizada.

Para casos envolvendo sequências de DNA pequenas, o uso do anel $\mathbb{Z}_4$ visto no capítulo 4 é eficaz, mas ao utilizar sequências maiores, o uso de uma representação polinomial se torna relevante e os cálculos mais eficazes.

Sanchez et al. em [3], utilizam a representação $G-00$, $U-10$, $A-01$ e $C-11$ para efetuar a construção da tabela do código genético, ou seja, vemos que a representação é feita através do alfabeto $\mathbb{Z}_2\times\mathbb{Z}_2$, correspondendo ao rótulo 3201 e que de acordo com Rocha, em [7], representa uma das oito permutações do rotulamento A. Essa representação foi feita obedecendo-se uma ordem de importância das bases, onde o grau dos termos do polinômio diminui a partir da base biológica mais importante para a base biológica menos importante. Em seguida, alguns cálculos relacionados à distância de Hamming entre dois códons foram efetuados, bem como o peso de Hamming de um códon.

A representação polinomial desses códons utiliza os elementos de $GF(64)$, que são obtidos através da extensão do corpo $GF(2)$.

Rocha em [7] utiliza essa extensão em um dos passos da construção de um código BCH para sequências de DNA de comprimento 63, de forma a obter as sequências de DNA desejadas.

O corpo $GF(2^r)$ é obtido por meio da extensão do corpo $GF(2)$ através de um ideal gerado por

um dos polinômios primitivos de grau 6, onde a raíz é α. Os polinômios primitivos de grau 6 são:

1. $x^6+x^5+x^3+x^2+1$
2. x^6+x+1
3. $x^6+x^5+x^2+x+1$
4. $x^6+x^4+x^3+x+1$
5. $x^6+x^5+x^4+x+1$
6. x^6+x^5+1

A extensão é realizada da seguinte forma: considere o corpo de Galois $GF(2^r) = GF(2^6) = GF(64)$ dado por:

$\frac{GF(2)[x]}{<p(x)>} = \{a_0+a_1x+a_2x^2+...+a_5x^5, a_i's \in F_2\}$, onde $p(x)$ é o polinômio primitivo.

Propomos a construção da tabela do código genético para os rotulamentos B e C, bem como os cálculos da distância de Hamming entre dois códons e o peso de Hamming de um códon, a fim de comparar os resultados com os efetuados para o rotulamento A, além de mostrar que os procedimentos relacionados ao rotulamento A podem ser estendidos para os rotulamentos B e C.

Este capítulo está organizado da seguinte maneira: na seção 5.1 apresentamos o modelo utilizado por Sanchez et al. em [3]. Na seção 5.2 apresentamos a proposta de representação dos códons para os rotulamentos B e C e uma série de exemplos relacionados a essas representações. Por fim, na seção 5.3 apresentamos alguns comentários, comparações entre os rotulamentos e apontamentos relevantes relacionados à representação polinomial dos códons.

5.1 Modelo de Representação Polinomial dos Códons para o Rotulamento A

Sanchez et al. em [3], utilizam a representação $G-00$, $U-10$, $A-01$ e $C-11$ para efetuar a construção da tabela do código genético, que corresponde ao rotulamento A.

A representação binária das bases não ocorreu de forma arbitrária, as bases complementares na molécula de DNA estão em correspondência com a representação binária e a forma complementar. Note que esta não é uma codificação de base arbitrária, este é o resultado de um isomorfismo entre

dois reticulados booleanos, $\varphi : B(X) \rightarrow ((\mathbb{Z}_2)^2, \wedge, \vee)$,onde $B(X) = \{A, C, G, T/U\}$, $\mathbb{Z}_2 = \{0,1\}$, $\wedge$ - conectivo lógico e(conjunção) e $\vee$ - conectivo lógico ou(disjunção).

Os coeficientes na representação polinomial dos códons do código genético obedecem uma ordem de importância da posição das bases nos códons e o isomorfismo $\varphi : B(X) \rightarrow (\mathbb{Z}_2)^2$ permitem-nos apresentar uma função $\psi : GF(64) \rightarrow C_g$, de forma que:

$$a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5 \rightarrow (f_1(a_2a_3), f_2(a_4a_5), f_3(a_0a_1))$$

Para todo $X_1X_2X_3 \in C_g$ existe um polinômio $p(x) \in GF(64)$ e vice-versa, tal que:

$$\psi(p(x)) = X_1X_2X_3.$$

Observa-se que os coeficientes a_4 e a_5, dos monômios de maiores graus correspondem à segunda base do códon. Os coeficientes que correspondem a primeira base são a_2 e a_3 e, por fim, os coeficientes a_0 e a_1, com menores graus correspondem a terceira posição do códon. Ou seja, o grau dos termos do polinômio diminui da base biológica mais importante para a base biológica menos importante.

A Tabela 5.1 apresenta a construção do código genético com a representação polinomial para cada um dos 64 códons. Uma observação interessante é que para todo códon a sequência de dígitos binários da representação polinomial é o inverso da sequência de dígitos binários do número inteiro correspondente.

G					U				A				C				
	nº	GF(64)	I	II	nº	GF(64)	I	II	nº	GF(64)	I	II	nº	GF(64)	I	II	
G	0	000000	GGG	G	16	000010	GUG	V	32	000001	GAG	E	48	000011	GCG	A	G
	1	100000	GGU	G	17	100010	GUU	V	33	100001	GAU	D	49	100011	GCU	A	U
	2	010000	GGA	G	18	010010	GUA	V	34	010001	GAA	E	50	010011	GCA	A	A
	3	110000	GGC	G	19	110010	GUC	V	35	110001	GAC	D	51	110011	GCC	A	C
U	4	001000	UGG	W	20	001010	UUG	L	36	001001	UAG	STOP	52	001011	UCG	S	G
	5	101000	UGU	C	21	101010	UUU	F	37	101001	UAU	Y	53	101011	UCU	S	U
	6	011000	UGA	STOP	22	011010	UUA	L	38	011001	UAA	STOP	54	011011	UCA	S	A
	7	111000	UGC	C	23	111010	UUC	F	39	111001	UAC	Y	55	111011	UCC	S	C
A	8	000100	AGG	R	24	000110	AUG	M	40	000101	AAG	K	56	000111	ACG	T	G
	9	100100	AGU	S	25	100110	AUU	I	41	100101	AAU	N	57	100111	ACU	T	U
	10	010100	AGA	R	26	010110	AUA	I	42	010101	AAA	K	58	010111	ACA	T	A
	11	110100	AGC	S	27	110110	AUC	I	43	110101	AAC	N	59	110111	ACC	T	C
C	12	001100	CGG	R	28	001110	CUG	L	44	001101	CAG	Q	60	001111	CCG	P	G
	13	101100	CGU	R	29	101110	CUU	L	45	101101	CAU	H	61	101111	CCU	P	U
	14	011100	CGA	R	30	011110	CUA	L	46	011101	CAA	Q	62	011111	CCA	P	A
	15	111100	CGC	R	31	111110	CUC	L	47	111101	CAC	H	63	111111	CCC	P	C

Tabela 5.1: Código genético segundo o rotulamento A

Observe os exemplos a seguir onde são detalhadas as características mencionadas anteriormente.

Exemplo 22 *Tomemos como referência o número 19 da Tabela 5.1. O mesmo é representado na forma binária como* 110010 *(leitura da esquerda para a direita) , referente ao códon GUC e a representação polinomial é* $1 + x + x^4$. *Como surgiram essas representações?*

$$\textit{Temos } 19 \rightarrow 010011 \rightarrow 110010 \rightarrow GUC \rightarrow 1+x+x^4$$

A primeira representação é a numérica, contida no anel $\mathbb{Z}_{64}$, dos códons do código genético. Em seguida, vemos a representação binária do número 19. A seguir, conforme mencionado anteriormente, vemos a sequência de dígitos binários da representação polinomial, que como pode-se notar é o inverso da sequência de dígitos binários do número inteiro correspondente. O códon GUC, representação que aparece logo após a representação binária do polinômio foi estabelecido pela ordem de importância das bases. Perceba que a representação binária 11 corresponde a base C(terceira base), 00 corresponde a base G(primeira base) e 10 corresponde a base U(segunda base). O grau aumenta da base biológica menos importante (terceira) para a base biológica mais importante (segunda). Por fim, o polinômio que representa o códon é $1+x+x^4$.

Através dessa representação algébrica do código genético, é possível estabelecer várias operações entre os códons no espaço vetorial gerado pela extensão de Galois para código genético. Para o rotulamento A, o códon *GGU* representa a base canônica desse espaço e o mesmo codifica o aminoácido de estrutura mais simples, a glicina.

Além disso, a distância de Hamming, d_H, entre códons é estabelecida, como o número de dígitos diferentes entre dois códons.

Exemplo 23 *Consideremos os códons GAU e CCA. A distância de Hamming entre os mesmos será:*

$$d_H(000110, 101111) = 3$$

O peso de Hamming, W_H, de um códon é definido como o número de dígitos não nulos da representação polinomial de um códon.

Exemplo 24 *O peso de Hamming $W_H(X_1X_2X_3)$ do códon CUA é $W_H(101100) = 3$.*

Como consequência,

$$d_H(X,Y) = W_H(X+Y)$$

Essas operações são importantes para a definição de um pseudo-produto interno entre dois códons.

5.2 Representação Polinomial dos Códons para os Rotulamentos B e C

De forma similar ao caso considerado, iremos utilizar a representação polinomial para os rotulamentos B e C. Desta forma, selecionamos um rótulo no conjunto associado ao rotulamento B e C. Em seguida, efetuamos as operações correspondentes aos rotulamentos B e C.

Para o rotulamento B, selecionamos o rótulo 1230, que para a construção do código genético usaremos a representação $\mathbb{Z}_2 \times \mathbb{Z}_2$, ou seja, $U-00$, $A-10$, $G-01$, $C-11$. A Tabela 5.2 ilustra o código genético referente ao rotulamento B com a representação polinomial de cada um dos códons.

U					A				G				C				
	nº	GF(64)	I	II	nº	GF(64)	I	II	nº	GF(64)	I	II	nº	GF(64)	I	II	
U	0	000000	UUU	F	16	000010	UAU	Y	32	000001	UGU	C	48	000011	UCU	S	U
	1	100000	UUA	L	17	100010	UAA	STOP	33	100001	UGA	STOP	49	100011	UCA	S	A
	2	010000	UUG	L	18	010010	UAG	STOP	34	010001	UGG	W	50	010011	UCG	S	G
	3	110000	UUC	F	19	110010	UAC	Y	35	110001	UGC	C	51	110011	UCC	S	C
A	4	001000	AUU	I	20	001010	AAU	N	36	001001	AGU	S	52	001011	ACU	T	U
	5	101000	AUA	I	21	101010	AAA	K	37	101001	AGA	R	53	101011	ACA	T	A
	6	011000	AUG	M	22	011010	AAG	K	38	011001	AGG	R	54	011011	ACG	T	G
	7	111000	AUC	I	23	111010	AAC	N	39	111001	AGC	S	55	111011	ACC	T	C
G	8	000100	GUU	V	24	000110	GAU	D	40	000101	GGU	G	56	000111	GCU	A	U
	9	100100	GUA	V	25	100110	GAA	E	41	100101	GGA	G	57	100111	GCA	A	A
	10	010100	GUG	V	26	010110	GAG	E	42	010101	GGG	G	58	010111	GCG	A	G
	11	110100	GUC	V	27	110110	GAC	D	43	110101	GGC	G	59	110111	GCC	A	C
C	12	001100	CUU	L	28	001110	CAU	H	44	001101	CGU	R	60	001111	CCU	P	U
	13	101100	CUA	L	29	101110	CAA	Q	45	101101	CGA	R	61	101111	CCA	P	A
	14	011100	CUG	L	30	011110	CAG	Q	46	011101	CGG	R	62	011111	CCG	P	G
	15	111100	CUC	L	31	111110	CAC	H	47	111101	CGC	R	63	111111	CCC	P	C

Tabela 5.2: Código genético segundo o rotulamento B

Exemplo 25 *Vamos considerar o número 54 da Tabela 5.2. O mesmo é representado por* 011011*, referente ao códon ACG e a representação polinomial é* $x+x^2+x^4+x^5$.

O número 54 é a representação em $\mathbb{Z}_{64}$ *do códon ACG, bem como* 110110 *é sua representação binária, que lida da direita para a esquerda identifica os coeficientes do polinômio relacionado a esse códon, ou seja,* $x+x^2+x^4+x^5$. *A ordem de importância das bases levou ao códon ACG. A representação binária* 01 *corresponde à base G (terceira base), 10 corresponde à base A (primeira base) e 11 corresponde à base C (segunda base).*

O espaço vetorial do código genético associado ao rotulamento B tem como base canônica o códon UUA, que codifica o aminoácido leucina.

Exemplo 26 *A distância de Hamming entre os códons CGU e AUC é:*

$$d_H(CGU, AUC) = d_H(001101, 111000) = 4$$

Exemplo 27 *O peso de Hamming do códon AUG é:*

$$W_H(AUG) = W_H(011000) = 2$$

O rotulamento C também foi utilizado na construção do código genético. Para isso, usamos o rótulo 1320, ou seja, em $\mathbb{Z}_2 \times \mathbb{Z}_2$ temos $U-00$, $A-10$, $C-01$, $G-11$. O código genético com este rotulamento é apresentado na Tabela 5.3.

U					A				C				G				
	nº	GF(64)	I	II	nº	GF(64)	I	II	nº	GF(64)	I	II	nº	GF(64)	I	II	
U	0	000000	UUU	F	16	000010	UAU	Y	32	000001	UCU	C	48	000011	UGU	S	U
	1	100000	UUA	L	17	100010	UAA	STOP	33	100001	UCA	STOP	49	100011	UGA	S	A
	2	010000	UUC	F	18	010010	UAC	Y	34	010001	UCC	C	50	010011	UGC	S	C
	3	110000	UUG	L	19	110010	UAG	STOP	35	110001	UCG	W	51	110011	UGG	S	G
A	4	001000	AUU	I	20	001010	AAU	N	36	001001	ACU	S	52	001011	AGU	T	U
	5	101000	AUA	I	21	101010	AAA	K	37	101001	ACA	R	53	101011	AGA	T	A
	6	011000	AUC	I	22	011010	AAC	N	38	011001	ACC	S	54	011011	AGC	T	C
	7	111000	AUG	M	23	111010	AAG	K	39	111001	ACG	R	55	111011	AGG	T	G
C	8	000100	CUU	L	24	000110	CAU	H	40	000101	CCU	R	56	000111	CGU	P	U
	9	100100	CUA	L	25	100110	CAA	Q	41	100101	CCA	R	57	100111	CGA	P	A
	10	010100	CUC	L	26	010110	CAC	H	42	010101	CCC	R	58	010111	CGC	P	C
	11	110100	CUG	L	27	110110	CAG	Q	43	110101	CCG	R	59	110111	CGG	P	G
G	12	001100	GUU	V	28	001110	GAU	D	44	001101	GCU	G	60	001111	GGU	A	U
	13	101100	GUA	V	29	101110	GAA	E	45	101101	GCA	G	61	101111	GGA	A	A
	14	011100	GUC	V	30	011110	GAC	D	46	011101	GCC	A	62	011111	GGC	G	C
	15	111100	GUG	V	31	111110	GAG	E	47	111101	GCG	A	63	111111	GGG	A	G

Tabela 5.3: Código genético segundo o rotulamento C

Exemplo 28 *Considere o códon AGU, representado pelo número 52 da Tabela 5.3. Temos:*

$$52 \to 110100 \to 001011 \to AGU \to x^2 + x^4 + x^5$$

O número 52 é a representação numérica em $\mathbb{Z}_{64}$*, bem como* 110100 *é sua representação binária. Em seguida, vemos os coeficientes do polinômio associado ao códon AGU, que nada mais é que a representação binária lida da direita para a esquerda. Mais uma vez vemos a importância da posição das bases na representação do códon,* 00 *é a associação binária atribuída a base U(terceira base),* 10 *à base A(primeira base) e por fim* 11*, associado à base G(segunda base).*

Da mesma forma que a base canônica associada ao espaço vetorial do código genético do rotulamento B é a trinca UUA, para o rotulamento C, a base canônica também é a trinca UUA, que codifica o aminoácido leucina.

Exemplo 29 *A distância de Hamming entre os códons CUG e UAU é:*

$$d_H(CUG, UAU) = d_H(110100, 000010) = 4$$

Exemplo 30 *O peso de Hamming do códon AAG é:*

$$W_H(AAG) = W_H(111010) = 4$$

5.3 Análise dos Resultados

Neste capítulo foram construídas tabelas do código genético para os rotulamentos B e C, expressando sua representação polinomial e algumas operações associadas a essa representação, usando para isso uma estrutura de espaço vetorial através de uma extensão do corpo $GF(2)$ para o corpo $GF(64)$, englobando os 64 códons que fazem parte do código genético. Podemos relacionar essa estrutura polinomial ao contexto de comunicação digital, onde os 64 códons, ou trincas, representam os sinais da constelação de sinais.

Da mesma forma que apresentamos o casamento do codificador com o modulador para o caso combinatorial, onde trabalhamos com a constelação $4-PSK$, culminando com um código geometricamente uniforme, para a representação polinomial também observamos esse casamento, uma vez que agora estamos trabalhando com o corpo $GF(64)$ a partir de uma representação em $GF(2)$. O mapeamento usado no codificador está casado com a constelação de 64 sinais (trincas) do código genético, presentes no modulador, que participam da construção de sequências de DNA dos mais variados comprimentos, como apresentos em [7] e [8].

Desta forma, temos o modelo algébrico dos rotulamentos A, B e C. Em todos eles são válidas as representações polinomiais, bem como as operações definidas no decorrer do capítulo. Os cálculos realizados se referem ao caso primal, podendo ser estendidos para o caso dual.

É importante ressaltar que nossa preocupação neste capítulo foi com a representação polinomial dos códons como forma de modelarmos algebricamente esta estrutura biológica refletindo biologicamente o que é feito matematicamente. As implicações biológicas acerca do cálculo das distâncias de Hamming e do peso de Hamming de um códon estão relacionadas a mutações e suas consequências são propostas para trabalhos futuros.

Capítulo 6

Conclusões e Perspectivas Futuras

O estudo do código genético, suas características, propriedades e funções é algo que diversos pesquisadores tem estudado e a modelagem matemática associada ao mesmo é uma área de pesquisa em franca expansão, buscando representar matematicamente o que o mundo biológico realiza. Este trabalho mostra algumas aplicações de estruturas matemáticas no código genético, tornando-se ferramenta eficaz na análise de diversos fenômenos biológicos, como a caracterização dos códons de acordo com o critério de hidropaticidade, bem como ferramenta eficaz em estudos de análises mutacionais.

Essas construções, cujos resultados representam as contribuições deste trabalho estão nos capítulos 3, 4 e 5.

Na seção 6.1 apresentamos o desenvolvimento do trabalho, através do detalhamento de cada um dos capítulos da tese, bem como os resultados encontrados em cada uma das modelagens realizadas. Na seção 6.2 apresentamos as contribuições do trabalho, através da análise dos resultados e sua relevância algébrica e biológica. Na seção 6.3 apresentamos algumas sugestões e propostas para trabalhos futuros. Por fim, na seção 6.4 apresentamos algumas considerações finais e o fechamento do trabalho.

6.1 Desenvolvimento do Trabalho

No capítulo 2 são apresentados os elementos de álgebra, biologia e códigos corretores de erro utilizados no decorrer do trabalho.

No capítulo 3 é apresentado o primeiro resultado, através da construção do reticulado booleano e do diagrama de Hasse, estruturas matemáticas que organizam os códons por meio de operações de álgebra booleana, refletindo características biológicas associadas aos aminoácidos codificados. A hidropaticidade, ou seja, a classificação dos aminoácidos em hidrofóbicos ou hidrofílicos foi a propri-

edade analisada.

Através do rotulamento dos nucleotídeos das sequências de DNA, adenina, citosina, guanina, timina/uracila, representadas pelo alfabeto $N = \{A, C, G, T/U\}$ foi feito um mapeamento para o alfabeto do anel $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Através desse rotulamento, 24 permutações são obtidas, classificadas de acordo com [7] em três rotulamentos A, B e C, de acordo com a característica geométrica gerada por cada um dos rotulamentos.

Os reticulados booleanos e o diagrama de Hasse relacionados ao rotulamento A foram apresentados em [1]. Neste trabalho, foram construídos os reticulados booleanos e os diagramas de Hasse associados aos rotulamentos B e C, os quais apresentam características biológicas e algébricas distintas do modelo construído para o rotulamento A.

As construções realizadas identificaram uma regularidade na separação dos códons para o rotulamento B, separando os códons hidrofóbicos e hidrofílicos, regularidade não encontrada no rotulamento C, onde os mesmos não se separam.

Outra observação interessante é em relação ao casamento da complementaridade biológica e matemática. No rotulamento A, ocorre o casamento da complementaridade biológica $(A-U)/(C-G)$ com a complementaridade algébrica $(00-11)/(01-10)$, ao passo que esse casamento não ocorre nos rotulamentos B e C, onde prevalece a complementaridade algébrica. Em relação ao pareamento biológico, observa-se no rotulamento B $(U-C)/(G-A)$ e no rotulamento C $(A-C)/(G-U)$.

No capítulo 4 apresentamos a segunda contribuição, através das operações algébricas entre códons, como a soma entre códons e a separação dos códons pela paridade. A operação soma entre códons pode ser uma ferramenta de vital importância em análises mutacionais. Nesta segunda abordagem foi utilizado o mapeamento $N \to \mathbb{Z}_4$, identificando uma estrutura de espaço vetorial associada ao código genético, permitindo assim a realização de operações, de forma combinatorial. Nesse contexto, propusemos um algoritmo de soma entre códons que diminui a complexidade em relação a algoritmos usados em outras abordagens.

O algoritmo proposto utiliza a operação soma com transporte de 1 no anel $\mathbb{Z}_4$. Na construção das tabelas do código genético identifica-se uma estrutura de grupo associada, ou seja, qualquer soma realizada entre códons resultará em um códon da tabela do código genético.

Em relação a paridade, observa-se uma forte ligação com os resultados apresentados na construção dos reticulados booleanos e diagramas de Hasse, uma vez que prevalece o pareamento algébrico na identificação dos códons pares e ímpares em cada um dos rotulamentos.

A hidropaticidade dos aminoácidos, analisada no capítulo 3, através da construção dos reticulados booleanos e diagramas de Hasse é outro aspecto interessante analisado nas tabelas do código genético

em cada um dos rotulamentos A, B e C. A regularidade ou ausência dela na separação dos códons observada no diagrama de Hasse também é observada nas tabelas do código genético para cada um dos rotulamentos.

O último resultado desse trabalho é apresentado no capítulo 5, onde é utilizada uma representação polinomial dos códons na tabela do código genético para os rotulamentos B e C, visto que para o rotulamento A a mesma já foi determinada. O objetivo desta abordagem é pelo fato da representação polinomial ser ferramenta eficaz na realização de operações entre códons, que mais uma vez pode ajudar na análise de mutações.

Utilizado a extensão de $GF(2)$ para $GF(64)$, encontramos a representação polinomial de cada um dos códons para os rotulamentos A, B e C, levando em consideração a importância biológica de cada uma das bases nos códons. Essa representação permite a realização de uma série de operações, como soma e produto entre códons, cálculo da distância de Hamming e peso de Hamming dos códons, elementos que podem ajudar em pesquisas relacionadas a análises mutacionais.

A relevância deste trabalho está no fato de efetuarmos a construção de vários modelos matemáticos para os três rotulamentos das permutações associadas à bijeção $N \to \mathbb{Z}_4$ e a análise algébrica e biológica associadas a essas construções.

6.2 Contribuições do Trabalho

Os resultados apresentados no decorrer do trabalho contribuem para os campos de engenharia genética, bioinformática, biomatemática, bem como matemática aplicada, uma vez que devido a interdisciplinaridade do mesmo, áreas como biologia, álgebra e códigos foram analisadas. As seguintes contribuições podem ser destacadas:

- A utilização de ferramentas matemáticas no contexto biológico.
- Caracterização e análise dos rotulamentos A, B e C através da construção dos reticulados booleanos, diagramas de Hasse.
- Comparação das construções dos diagramas de Hasse para os rotulamentos A, B e C.
- Identificação de uma estrutura de grupo associada ao mapeamento do código genético.
- Desenvolvimento de um algoritmo de soma com transporte, facilitando a realização da operação soma entre códons.

- Análise da paridade dos códons e sua relação com as construções dos diagramas de Hasse.
- Comparação do algoritmo soma com transporte com os algoritmos soma em $\mathbb{Z}_{64}$ e algoritmo SMG.
- Relacão das operações entre códons e análise mutacionais.
- Utilização de uma representação polinomial dos códons associada a cada um dos rotulamentos do código genético.
- Comparação das representações polinomiais dos rotulamentos A, B e C.
- Análise do comportamento dos principais aminoácidos nas modelagens construídas.
- Determinação de um espaço vetorial para o código genético e sua relação com a geração de sequências de DNA.

Um dos resultados desse trabalho, referente à construção dos reticulados booleanos e diagramas de Hasse foi apresentado no congresso [30].

6.3 Sugestões para Trabalhos Futuros

Algumas sugestões para trabalhos futuros são apresentadas a seguir, com o objetivo de dar continuidade na pesquisa.

- O uso dos reticulados booleanos e diagramas de Hasse, bem como outras estruturas de POSETS(conjuntos parcialmente ordenados) para representar estruturas secundárias e terciárias de proteínas.
- Implementar o algoritmo soma com transporte, de maneira a realizar testes com sequências de DNA e suas possíveis mutações.
- Realizar um estudo mais detalhado dos aminoácidos triptofano e tirosina, devido a sua proximidade de códons STOP no código genético.
- Estudar de maneira detalhada o pseudo-produto interno entre dois códons e sua relação com as mutações.
- Analisar as consequências biológicas acerca do cálculo das distâncias de Hamming e do peso de Hamming entre códons.

6.4 Comentário Final

Nesse trabalho foram utilizadas diversas ferramentas matemáticas com o objetivo de modelar o código genético, apresentando suas características e propriedades. A interdisciplinaridade do mesmo permitiu relacionar elementos biológicos, algébricos e de codificação, tornando-se uma abordagem relevante em diversas áreas de pesquisa, como engenharia genética, biomatemática, bioinformática, etc.

Diversos estudiosos tem pesquisado o assunto e a tendência é o aprimoramento cada vez maior de técnicas que relacionam estas diversas áreas do conhecimento, que poderão desvendar os mistérios da "máquina da vida".

Referências Bibliográficas

[1] Sánchez R.; Morgado E.; Grau R. "The genetic Code Boolean Lattice, " *MATCH Commun.Math.Comp.Chem*, **52** pp. 29–46, (2004).

[2] Sánchez R.; Grau R. "A genetic code Boolean structure. II. The genetic information system as a Boolean information system, " *Bulletin of Mathematical Biology*, **67** pp. 1017–1029, (2005).

[3] Sánchez R.; Perfetti L.A.; Grau R.; Morgado E. "A new DNA sequences vector space on a genetic code Galois field, " *MATCH Commun. Math. Comput. Chem.*, **54** ,(2005) .

[4] Sánchez R.; Grau R.; Morgado E. "A novel Lie algebra of the genetic code over the Galois field of four DNA bases, " *Mathematical Biosciences*, **202** pp. 156–174, (2006).

[5] 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** pp. 117–125, (1996).

[6] Jiménez-Montaño M.A.; Mora-Basáñez C.R.; Pöschel T. "On the hypercube structure of the genetic code, " *Bioinformatics and genome research*, pp. 445, (1994).

[7] Rocha A. S. L., R., Palazzo Jr., M.C. Silva-Filho "Modelo de sistema de comunicações digital para o mecanismo de importação de proteínas mitocondriais através de códigos corretores de erros". 2010. 155 f. Tese (Doutorado em Engenharia Elétrica). DT-FEEC, UNICAMP.

[8] Faria L. C. B., R. Palazzo Jr. "Existências de códigos corretores de erros e protocolos de comunicação em sequências de DNA". 2011. 298 f. Tese (Doutorado em Engenharia Elétrica). DT-FEEC, UNICAMP.

[9] Sánchez R.; Morgado E.; Grau R. "Gene algebra from a genetic code algebraic structure, " *J.Math.Biol.*, **51** pp. 431–457, (2005).

[10] Sánchez R.; Grau R.; Morgado E. "Genetic code Boolean algebras, " *WSEAS Transactions on Biology and Biomedicine*, **1** pp.190–197 , (2004).

[11] Sánchez R.; Morgado E.; Grau R. "A genetic code Boolean structure. I. The meaning of boolean deductions, " *Bulletin of Mathematical Biology*, **67** pp. 1–14, (2005).

[12] Hoernes G.E; Heilweil M.F *Introduction to Boolean Algebra and Logic Design*. McGraw-Hill Book Company, (1964). 306 p.

[13] Lipschutz S. tradução de Fernando Vilain Heusi da Silva *Teoria dos Conjuntos*. McGraw-Hill Book Company, (1972). 333 p.

[14] Fraleigh J.B. *A First course in abstract algebra*. Addison Wesley, (2003). 520 p.

[15] Gonçalves A. *Introdução à álgebra*. Adílson Gonçalves, (1979). 194 p.

[16] Lin S. Costello, Jr. D.J. *Error Control Coding*. Pearson Prentice Hall, (2004). 1262 p.

[17] Rocha A. S. L. "Modelo Matemático para a Previsão de Recombinação Sítio-Específica do DNA ". 2004. Dissertação (Mestrado em Engenharia Elétrica). DT-FEEC, UNICAMP.

[18] Nicholl D.S.T. *An introduction to genetic engeneering*. Cambridge, (2008). 340 p.

[19] Carvalho H.F; Recco-Pimentel S.M. *A célula 2001*. Manole Ltda, (2001). 292 p.

[20] Lodish, Berk, Matsudaira, Kaiser, Scott, Zipursky, Darnell. *Molecular Cell Biology*. Artmed, (2005). 1056 p.

[21] Nelson D.L; Cox M.M. *Lehninger Princípios de Bioquímica*. Sarvier, (2002). 982 p.

[22] De Robertis E.M.F; HIB J. *Bases da Biologia Celular e Molecular*. Guanabara Koogan, (2006).

[23] Turner P.C, et al. *Biologia Molecular*. Guanabara Koogan, (2004).

[24] Gerônimo J.R., R., Palazzo Jr.,Interlando J.C. "Extensão da Z_4 linearidade via grupo de simetrias". 1997. 167 f. Tese (Doutorado em Engenharia Elétrica). DT-FEEC, UNICAMP.

[25] Shannon, C.E.. "A mathematical theory of communication, " *Bell Syst. Tech. J.*, **27** pp. 397–423, (julho 1948) and pp. 623 – 656, (outubro 1948).

[26] Ungerboeck. G. "Channel coding with multilevel / phase signals, " *IEEE Trans. Inform. Theory*, **IT-28** pp. 56–67, (1982).

[27] Forney, Jr. G.D. “Geometrically uniform codes, ” *IEEE Trans. Inform. Theory*, **IT-37** pp. 1241–1260, (1991).

[28] Loeliger H.A. “Signal sets matched to groups, ” *IEEE Trans. Inform. Theory*, **IT-37** pp. 1675–1682, (1991).

[29] Macwilliams F.J., Sloane N.J.A.. *The theory of Error-Correcting Codes*. Elsevier Science B.V. , (1981). 764 p.

[30] Oliveira A.; R., Palazzo Jr. “Análise Algébrica do Código Genético através do Diagrama de Hasse e do Reticulado Booleano, ” *Anais do I Congresso de Matemática Aplicada e Computacional da Região Sudeste - I CMAC Sudeste*, pp. 228–230, (2011).