UNIVERSIDADE FEDERAL DE ALFENAS

LETÍCIA RIBEIRO PEREIRA

ANÁLISE MUTACIONAL VIA CÓDIGOS BCH NA IDENTIFICAÇÃO DE MALFORMAÇÕES OCULARES GRAVES

ALFENAS-MG

2022

LETÍCIA RIBEIRO PEREIRA

ANÁLISE MUTACIONAL VIA CÓDIGOS BCH NA IDENTIFICAÇÃO DE MALFORMAÇÕES OCULARES GRAVES

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do título de licenciada em Matemática pela Universidade Federal de Alfenas. Área de concentração: Matemática Aplicada.

Orientador: Prof. Dr. Anderson José de Oliveira.

ALFENAS-MG

2022

Agradecimentos

Primeiramente a Deus, por iluminar o meu caminho e permitir que chegasse até aqui.

Ao meu orientador, Anderson José, por toda paciência, compreensão e conhecimento transmitido ao longo deste trabalho. Te admiro muito, como pessoa e profissional. Agradeço também ao Diogo Guilherme, por colaborar com a realização da parte computacional da pesquisa.

As minhas colegas Bianca e Maria Flávia, agora amigas, pela oportunidade de estudar e aprender com elas durante a Iniciação Científica, a qual foi fundamental no processo de desenvolvimento desta monografia.

A minha família, por todo apoio e amor que tiveram comigo em toda a minha trajetória, em especial minha irmã, por sempre me ajudar diante das dificuldades enfrentadas em minha vida. Amo vocês.

Ao meu namorado por me incentivar, me acalmar nos momentos difíceis e não medir esforços para realizar os meus sonhos, como este.

A todos os meus amigos, em especial Valéria e Edmara, por me motivar, apoiar e por toda ajuda. Vocês tem grande parcela de contribuição na minha graduação e sempre serei muito grata por isso.

Por fim, gratidão a todos que de alguma forma fizeram parte da minha formação, contribuindo para que eu conseguisse concluir este sonho.

Resumo

Estudos relacionados à Biologia, Álgebra e Códigos Corretores de Erros estão em grande desenvolvimento nas últimas décadas, sendo um amplo assunto de pesquisa com várias possibilidades de aplicações. Devido a isso, é possível relacionar Álgebra e Biologia aos Códigos Corretores de Erros, uma vez que estruturas algébricas estão presentes em sua construção. Uma classe importante desses códigos, são os códigos BCH, que possuem características simples e um alto poder de detecção de erros, que podem ser utilizados para a reprodução de sequências de DNA. Desse modo, o objetivo deste trabalho é aplicar os códigos BCH no processo de geração de proteínas pelo algoritmo desenvolvido por Rocha, Faria e Palazzo Jr., a fim de reproduzir a sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA e analisar possíveis mutações que podem acarretar em doenças genéticas. A metodologia adotada consiste nas etapas de revisão de literatura, estudo do algoritmo de geração de proteínas, análise de uma sequência de DNA obtida no NCBI e sua aplicação no algoritmo estudado. Com este estudo, espera-se contribuir na análise das relações existentes entre Álgebra e Biologia, por meio dos Códigos Corretores de Erros, além de direcionar novos pesquisadores na área.

Palavras-chaves: códigos corretores de erros; sequência de DNA; estruturas algébricas.

Abstract

Studies related to Biology, Algebra and Error Correcting Codes have been under great development in the last decades, being a broad subject of research with several possibilities of applications. Because of this, it is possible to relate Algebra and Biology to the Error Correcting Codes, since algebraic structures are present in their construction. An important class of these codes are the BCH codes, which have simple characteristics and a high power of error detection, which can be used for the reproduction of DNA sequences. Thus, the objective of this work is to apply the BCH codes in the process of generating proteins by the algorithm developed by Rocha, Faria and Palazzo Jr., in order to reproduce a DNA sequence and analyze possible mutations that can lead to diseases genetics. The methodology consists of the steps of literature review, study of the protein generation algorithm, analysis of a DNA sequence obtained from the NCBI and its application in the studied algorithm. With this study, it is expected to contribute to the analysis of the existing relationships between Algebra and Biology, through Error Correcting Codes, in addition to directing new researchers in the area.

Keywords: error correcting codes; DNA sequence; algebraic structures.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

A tecnologia é uma importante ferramenta que surgiu em meio à Revolução Industrial e se tornou fundamental no século XXI, estando presente entre outras coisas, nas redes sociais, por meio do envio de mensagens. Este processo é realizado através de um sistema de comunicação digital, onde podem ser utilizados os códigos corretores de erros, que têm a capacidade de detectar e corrigir possíveis erros que possam ocorrer durante a transmissão ou armazenamento de mensagens.

Este sistema de comunicação pode ser relacionado ao sistema de transmissão de informação genética. A partir disso, é possível relacionar as sequências de DNA com as estruturas algébricas através dos códigos corretores de erros, sendo os erros as mutações genéticas. Uma classe importante desses códigos são os códigos BCH, que possuem características simples e um alto poder de detecção de erros, podendo ser aplicados em problemas biológicos.

De acordo com [1] e [2], pesquisadores da área de exatas, preocupados com a importância da redução de tempo para obter resultados e a redução de custos para a realização de experimentos laboratoriais, buscam métodos científicos capazes de reproduzir dados biológicos, por meio de modelos matemáticos de forma mais rápida e como consequência geram um menor custo.

Em [3], é realizado um trabalho com a proposta da modelagem matemática dos rotulamentos associados ao mapeamento do código genético para identificar suas características e propriedades. Tal modelagem, para o rotulamento A, é feita por meio do critério de complementaridade biológica, que coincide com a complementaridade algébrica, e nos casos dos rotulamentos B e C, utiliza-se a complementaridade algébrica. Para esta construção dos modelos, são utilizados os diagramas de Hasse e reticulados booleanos, que são associados a cada um dos rotulamentos do código genético.

Em [4] é apresentada a relevância dos códigos corretores de erros no processo de transmissão e/ou armazenamento de informações de forma confiável, que não altere a informação por meio dos erros. Um dos objetivos é mostrar a aplicação de polinômios nos códigos BCH sobre os corpos finitos, já que através desta estrutura é realizada a construção desses códigos, formando uma classe cíclica dos códigos corretores de erros e admitindo uma representação em termos de polinômios sobre extensões de $\mathbb{Z}_2$.

Em [5] tem-se o estudo da estrutura algébrica de um código BCH e a reprodução de uma sequência de DNA relacionada à proteína mitocondrial ATP6 utilizando esse código, a fim de localizar mutações e suas implicações biológicas causadas. Para isso, utilizou-se o algoritmo de geração de proteínas proposto em [6] e [2] e a construção de um código BCH sobre anéis, relacionando a presença de uma estrutura matemática presente no código genético por meio dos códigos corretores de erros.

Neste contexto, segundo [6], as estruturas algébricas, são um dos elementos fundamentais presentes no desdobramento do codificador genético e determinante dos códigos BCH expressos por anéis e por corpos, que são responsáveis pela identificação e reprodução das sequências de DNA.

Deste modo, a partir dos estudos e trabalhos desses pesquisadores e outros, ainda há diversas questões a serem analisadas, pesquisadas e respondidas. Assim, o objetivo desta monografia é apresentar a aplicação dos códigos BCH no processo de geração de proteínas pelo algoritmo desenvolvido por Rocha, Faria e Palazzo Jr., além de validar hipóteses levantadas, como a importância dos códigos BCH no algoritmo de geração de proteínas e uma aplicação no estudo de mutações que podem causar malformações oculares graves.

Este trabalho está estruturado da seguinte forma: no Capítulo 2 serão apresentados os principais conceitos teóricos fundamentais utilizados ao longo deste trabalho, referentes à Biologia, à Álgebra e aos Códigos Corretores de Erros. Na Seção 2.1 relacionada à Biologia, será apresentada desde a estrutura de uma célula até as mutações genéticas, sendo este último um dos alvos principais do estudo realizado. Na Seção 2.2 serão apresentadas as estruturas algébricas, além do processo de construção de Corpos de Galois. Posteriormente na Seção 2.3, serão apresentados os elementos de um sistema de comunicação, os Códigos Corretores de Erros, em especial os Códigos BCH. Na Seção 2.3 será apresentada a descrição do algoritmo de geração de proteínas, por meio do detalhadamento de seus passos. Já no Capítulo 3, será apresentada a geração da sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA via códigos corretores de erros, bem como a análise dos resultados obtidos. Por fim, no Capítulo 4 são apresentadas as considerações finais do trabalho.

2 Conceitos Teóricos Fundamentais

Neste capítulo serão apresentados os principais conceitos teóricos fundamentais utilizados no trabalho, referentes à Biologia, à Álgebra e aos Códigos Corretores de Erros.

2.1 Biologia

Nesta seção serão apresentados os principais elementos de Biologia utilizados neste estudo. As definições, teoremas, propriedades e aspectos teóricos apresentados nesta seção podem ser encontradas em [7], [8], [9], [10], [11], [12], [13], [14].

2.1.1 Estrutura de uma célula

As células surgiram há mais de três bilhões de anos, através da evolução e a formação de moléculas, que com o auxílio do meio aquoso foram envolvidas por uma membrana, formando as células, que são uma unidade que forma o corpo de todos os seres vivos, sendo capaz, individualmente, de obter energia, crescer e se reproduzir.

A célula possui três componentes básicos, a membrana plasmática, o citoplasma e o material genético. A membrana plasmática é uma membrana que envolve o citoplasma, um material fluido localizado no seu interior, e o material genético, responsável por carregar a informação genética dos seres vivos.

Diante destes componentes, é possível classificar as células em dois tipos principais: procariontes, onde o material genético fica disperso sobre o citoplasma e eucariontes, em que o material genético fica separado do citoplasma, por meio da carioteca, uma estrutura membranosa. Devido a esta delimitação do material genético, será estudada com mais detalhes a classe celular eucariótica que forma o corpo de todos os multicelulares, como os fungos, as plantas, os animais e os seres humanos.

A estrutura de uma célula eucariótica é formada além dos componentes básicos, por estruturas citoplasmáticas, tais como: as organelas (delimitadas por membranas), os ribossomos e os centríolos. Outra importante estrutura pertencente a este tipo de célula é o núcleo, formado através da carioteca, e estão presentes os cromossomos, composto basicamente por DNA (ácido desoxirribonucleico), que uma molécula que contém toda a informação genética de um ser vivo. A Figura 1 a seguir apresenta um exemplo de célula eucarionte e procarionte.

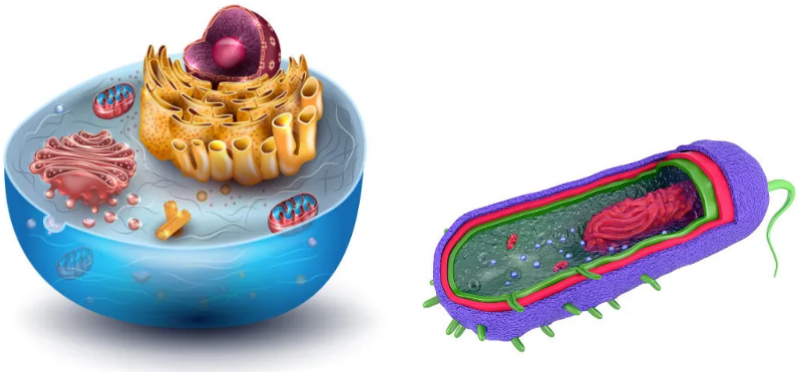

Figura 1 – Célula eucarionte, à esquerda, e procarionte, à direita.

Fonte: https://brasilescola.uol.com.br/biologia/celulas-procariontes.htm e https://brasilescola.uol.com.br/biologia/celulas-eucariontes.htm.

2.1.2 Nucleotídeos e ácidos nucleicos

As células possuem uma composição química, formada por substâncias inorgânicas, tendo origem mineral, como a água e os sais minerais. E por substâncias orgânicas, com origem vegetal ou animal, sendo os carboidratos, lipídios, proteínas e ácidos nucleicos. Esta última é uma importante substância relacionada à transmissão de características hereditárias presentes na informação genética do organismo, formadas por nucleotídeos, além de comandar e controlar todas as atividades das células.

Os ácidos nucleicos são uma repetição de moléculas menores, compostos por um grupo de fosfato (ácido fosfórico), uma molécula de açúcar (pentose) e uma base nitrogenada, como pode ser visto na Figura 2. Estas bases são divididas em bases púricas, que são adenina(A) e guanina(G) e bases pirimídicas, que são citosina(C), timina (T) e uracila(U).

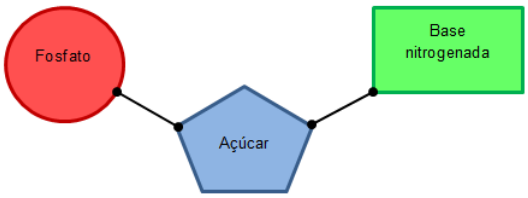

Figura 2 – Componentes químicos que formam um nucleotídeo.

Fonte: Próprio autor.

Desse modo, podem ser formados dois tipos de ácidos nucleicos: ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA). Tanto o DNA como o RNA são polímeros, isto é, moléculas formadas por várias unidades menores ligadas entre si de modo organizado.

O DNA funciona como um banco de informações genéticas, uma espécie de código, que transfere essas informações, com o objetivo de garantir a integridade da informação genômica e funcionando como molde para a síntese da molécula de RNA.

A determinação da estrutura do DNA por James Watson e Francis Crick em 1953 é, em geral, aceita como o marco do surgimento da biologia molecular moderna. A molécula de DNA é formada por uma fita dupla de vários nucleotídeos (dupla hélice), que são formados por um grupo de fosfato, uma pentose (desoxirribose, que é um açúcar composto por cinco átomos de carbono ligados, formando uma cadeia fechada) e uma base nitrogenada. As quatro bases nitrogenadas são: adenina e timina (purinas - maiores), citosina e guanina (pirimidinas - menores). As características de uma molécula de DNA são definidas por:

- Duas cadeias polinucleotídicas circundam um eixo comum formando a dupla hélice.
- Duas fitas de DNA são antiparalelas (possuem direções opostas), mas cada uma forma uma hélice para o lado direito.
- Bases ocupam o centro da hélice, e as cadeias de açúcar-fosfato estão dispostas na borda. A superfície da dupla hélice forma dois sulcos de largura desigual: a cavidade maior e a cavidade menor.
- Cada base está ligada a uma base da fita oposta por meio de pontes de hidrogênio. Adenina liga-se à timina e vice-versa, e a guanina liga-se com a citosina e vice-versa. Essas interações por pontes de hidrogênio (pareamento das bases complementares), resultam na associação específica das duas cadeias da fita dupla (Regra de Chargaff).

Uma molécula de DNA pode ser vizualizada na Figura 3 a seguir.

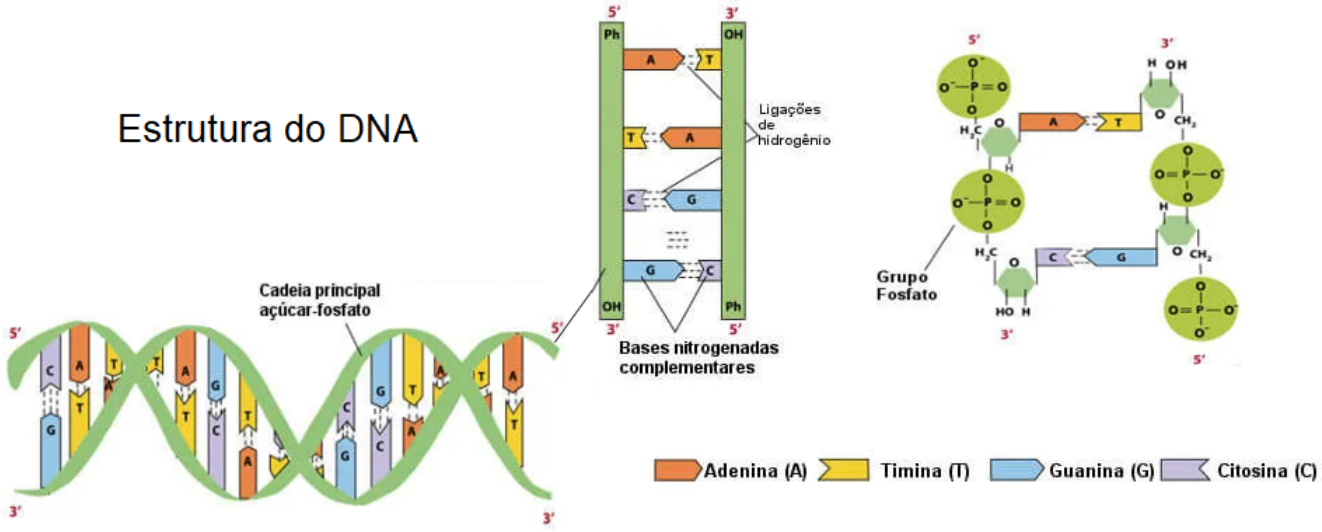

Figura 3 – Modelo da estrutura do DNA.

Fonte: https://brasilescola.uol.com.br/biologia/dna.htm.

Esse modelo sugere que 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 em qualquer fita. A dupla hélice de DNA forma espirais quando compactada dentro da célula.

A molécula de RNA (ácido ribonucleico) é formada a partir da molécula de DNA em um processo chamado de transcrição, onde a informação genética do DNA é transferida para uma molécula de RNAm, que veremos com mais detalhes no decorrer desta seção. Essa molécula apresenta informações com as quais é possível coordenar a produção de proteínas. Assim sendo, o RNA também participa do fluxo de informações genéticas pelos indivíduos.

O RNA é uma molécula composta por uma única fita de nucleotídeos ligados entre si, gerada pelo rompimento das pontes de hidrogênio da molécula de DNA, onde a pentose do RNA é sempre a ribose e as quatro bases nitrogenadas são: adenina, guanina, citosina e uracila, sendo a última exclusiva do RNA, como pode ser observado na Figura 4.

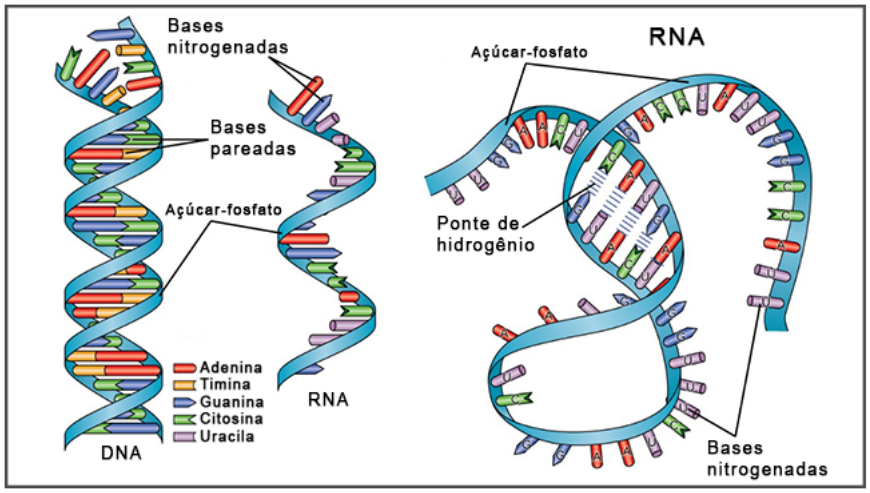

Figura 4 – Modelo da estrutura do RNA.

Fonte:https://www.significados.com.br/rna/

As moléculas de RNA são, geralmente, constituídas de uma única fita que se enrola entre si devido ao emparelhamento entre as bases complementares: a uracila liga-se com a adenina e a guanina liga-se com a citosina, seguindo a Regra de Chargaff. Elas são classificadas de acordo os papéis que desempenham no processo de transferência de informação. Existem três tipos de RNA: RNA mensageiro (RNAm), RNA transportador (RNAt) e RNA ribossômico (RNAr).

O RNA mensageiro é uma cópia das fitas de DNA, ficando responsável em levar as informações obtidas do DNA até o citoplasma, onde as proteínas serão produzidas. Como o RNA é uma cópia fiel de uma das fitas de DNA, é a partir dessa informação que o RNA mensageiro irá determinar quais são os aminoácidos necessários para a formação de determinada proteína, pois ele possui as trincas (códons) de bases nitrogenadas que definem cada aminoácido.

O RNA transportador é encarregado de transportar os aminoácidos que serão usados na formação das proteínas até o ribossomo, enquanto, que o RNA ribossômico faz parte da constituição dos ribossomos. É nos ribossomos que a sequência de bases do RNA mensageiro é interpretada e a proteína, de fato, sintetizada.

2.1.3 A síntese proteica e o código genético

As proteínas são polímeros formados de monômeros chamados de aminoácidos, que serão produzidas atráves das informações contidas na sequência de nucleotídeos que forma uma molécula de DNA, a qual será responsável por exercer diversas funções do nosso organismo. A síntese proteica pode ser dividida em duas fases, a transcrição e a tradução, como apresentado na Figura 5 a seguir.

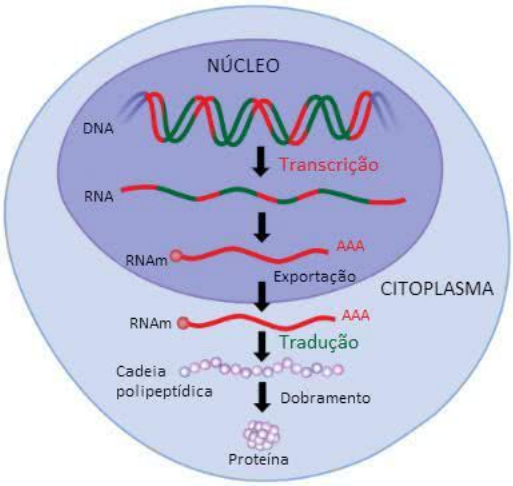

Figura 5 – Representação do processo de síntese proteica.

Fonte: https://www.todamateria.com.br/sintese-proteica/.

A transcrição consiste na conversão de DNA em RNA. Esse processo acontece por meio da enzima RNA polimerase, que se posiciona no início do gene separando as duas fitas de DNA, onde uma destas servirá como cadeia molde para que os seus nucleotídeos sejam emparelhados com os novos nucleotídeos de RNA (A,U,C,G) livres, desenvolvendo ao final o RNAm. Já a fase de tradução, consiste na conversão dos nucleotídeos do RNAt em aminoácidos, formando com a união destes aminoácidos uma cadeia polipeptídica.

Nessa fase o RNA migra para o citoplasma das células, onde guia a produção das proteínas no processo de tradução ou síntese proteica, que é realizado pelos ribossomos e envolve “decodificar"um RNA mensageiro (RNAm) e unir aminoácidos equivalentes a três sequências de bases nitrogenadas. Essas trincas de bases nitrogenadas são denominadas códons, cada um codificando um aminoácido específico da proteína. Existem 64 possíveis trincas, agrupadas três-a-três, que correspondem a 20 aminoácidos, sendo que mais de um códon pode corresponder ao mesmo aminoácido. Os aminoácidos são, portanto, as unidades básicas que formam as proteínas e a correspondência de um determinado aminoácido com um códon se dá por meio do código genético, como apresentado na Figura 6.

Os códons de parada ou terminação (stop), como os códons UAA, UAG e UGA, são utilizados para mostrar a interrupção da síntese de uma proteína e não determinam nenhum aminoácido. O códon de iniciação AUG utilizado para representar a iniciação da síntese do polipeptídeo (e também codifica a metionina). O DNA que abrange o códon de iniciação e vai até o códon de terminação é chamado de sequência codificadora. Além disso, os códons que determinam um aminoácido são semelhantes quimicamente, distinguindo-se somente em relação a uma base nitrogenada, o que faz

Trinca na fita de DNA codificador[1]	Códon do mRNA	Aminoácido[2]	Trinca na fita de DNA codificador[1]	Códon do mRNA	Aminoácido[2]
TTT	UUU	Fenilalanina	TAT	UAU	Tirosina
TTC	UUC	(Phe; F)	TAC	UAC	(Tyr; Y)
TTA	UUA		TAA	UAA	TERMINAÇÃO
TTG	UUG		TAG	UAG	
CTT	CUU	Leucina	CAT	CAU	Histidina
CTC	CUC	(Leu; L)	CAC	CAC	(His; H)
CTA	CUA		CAA	CAA	Glutamina
CTG	CUG		CAG	CAG	(Gln; Q)
ATT	AUU	Isoleucina (Ilê; I)	AAT	AAU	Asparagina
ATC	AUC		AAC	AAC	(Asn; N)
ATA	AUA		AAA	AAA	Lisina (Lys; K)
ATG		INICIAÇÃO / Metionina (Met; M)	AAG	AAG	
GTT	GUU		GAT	GAU	Ácido aspártico
GTC	GUC	Valina (Val; V)	GAC	GAC	(Asp; D)
GTA	GUA		GAA		Ácido
GTG	GUG		GAG	GAG	glutâmico (Glu; E)
TCT	UCU		TGT	UGU	Cisteína (Cys; C)
TCC	UCC	Serina (Ser; S)	TGC	UGC	
TCA	UCA		TGA	UGA	TERMINAÇÃO
TCG	UCG		TGG	UGG	Triptofano (Trp; W)
CCT	CCU		CGT	CGU	
CCC	CCC	Prolina (Pro; P)	CGC	CGC	Arginina (Arg; R)
CCA	CCA		CGA	CGA	
CCG	CCG		CGG	CGG	
ACT	ACU		AGT	AGU	Serina (Ser; S)
ACC	ACC	Treonina (Thr; T)	AGC	AGC	
ACA	ACA		AGA	AGA	Arginina (Arg; R)
ACG	ACG		AGG	AGG	
GCT	GCU		CGT	GGU	
GCC	GCC	Alanina (Ala; A)	GGC	GGC	Glicina (Gly; G)
GCA	GCA		GGA	GGA	
GCG	GCG		GGG	GGG	

[1] Conforme a convenção usual, as sequências de DNA são escritas em DNA equivalente ao RNA, que tem a sequência da fita que não é molde (é a fita codificadora)
[2] Os símbolos entre parênteses depois do nome de cada aminoácido são as abreviaturas padrão, em três letras e em uma letra.

Figura 6 – Código genético e as 64 possíveis combinações das bases nitrogenadas.

Fonte: [12].

com que as células se tornem mais resistentes a mutações, pois o aminoácido codificado é o mesmo, ainda que ocorra a troca da terceira base do códon.

2.1.4 Mutações

Mutações podem ser definidas como alterações que ocorrem no material genético, ou seja, na sequência de nucleotídeos que formam as cadeias do DNA. No processo de mutação esses nucleotídeos podem ser substituídos, perdidos ou acrescentados, modificando a sequência original. Essas mutações podem ocorrer espontaneamente, por meio de erros durante a cópia do material na divisão celular ou serem provocadas por agentes mutagênicos, como substâncias químicas, por exemplo drogas e bebidas alcoólicas. Além disso, a radiação solar também pode influenciar na geração de mutações.

Essas alterações no material genético podem trazer benefícios, prejuízos ou não se manifestarem no organismo e ocorrerem de forma aleatória. As mutações podem ser classificadas em: gênicas ou

cromossômicas.

As mutações gênicas se desenvolvem a partir da alteração da sequências de nucleotídeos do DNA. Como exemplo, temos a Heterocromia, uma mutação na coloração dos olhos, que causa um olho de cada cor. A Heterocromia pode não agregar prejuízo na vida da pessoa, mas talvez o olho mais claro receba mais iluminação durante determinado horário do dia, possibilitando uma dificuldade para enxergar.

Outro tipo de mutação existente são as mutações cromossômicas, que são alterações da estrutura ou do número de cromossomos. Como exemplo de mutações cromossômicas, temos a Síndrome de Turner, mutação que ocorre no número de cromossomos, surgindo somente em mulheres. Esta síndrome é uma monossomia do par 23, ou seja, do par sexual, onde o cariótipo de mulheres com essa síndrome será 45,X0, ou seja, tendo a ausência de outro X de caracterização do sexo feminino ou do Y que caracteriza o sexo masculino.

Em relação à estrutura, as mutações podem ser classificadas em mutações de pequena escala, as quais afetam um gene em um ou poucos nucleotídeos, como mutação de ponto, inserção e deleção, e mutações de grande escala como duplicação, deleção de regiões cromossômicas, translocação e inversão.

As mutações de pequena escala como a mutação de ponto, que ocorre quando somente uma base nitrogenada é alterada, podem ser classificadas como mutação "silenciosa", em que a base é modificada, no entanto, devido à característica de redundância do código genético, acaba por codificar o mesmo aminoácido; “missense” (sentido trocado), sendo feita a alteração de uma das bases do DNA, tendo como consequência a substituição de um aminoácido por outro diferente e por fim a mutação "sem sentido", onde produz um códon de STOP que impede que a proteína seja produzida integralmente.

Além disso, a mutação de pequena escala pode gerar a inserção que ocorre devido à adição de um ou mais nucleotídeos no DNA. Por fim, na deleção é retirado um ou mais nucleotídeos, alterando a composição da proteína e é aleatória. Um exemplo de mutação de pequena escala pode ser visto na Figura 7.

As mutações de grande escala, como a duplicação (amplificação), ocorrem pela criação de várias cópias de uma região cromossômica, aumentando a dosagem dos genes dentro dela. A deleção de regiões cromossômicas ocorre quando o cromossomo é desprovido de uma parte, ocasionando na perda dos genes nessas regiões. Translocação é a transferência do segmento de um cromossomo preso a outro cromossomo que não é homólogo ao seu. Na inversão ocorre a troca da orientação de um segmento do cromossomo. Um exemplo de mutação de grande escala pode ser visto na Figura 8.

As mutações podem afetar a função de uma proteína, gerando perda ou ganho de função ou até a morte do organismo que a possui. Por garantir variabilidade genética, as mutações são consideradas o mecanismo que permite a seleção natural. Desse modo, é por meio das mutações que características vantajosas serão multiplicadas nas gerações subsequentes ou características deletérias irão desaparecer.

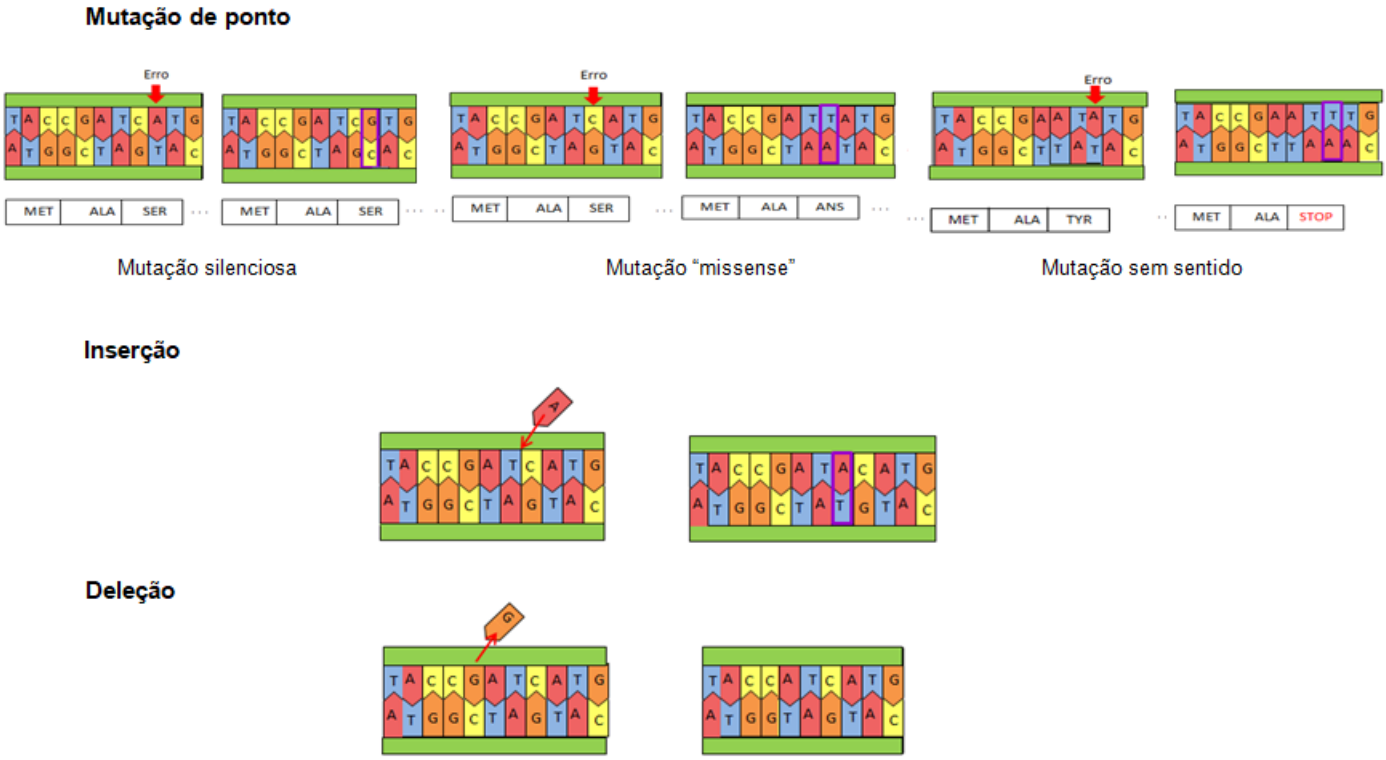

Figura 7 – Mutações de pequena escala.

Fonte: Próprio autor.

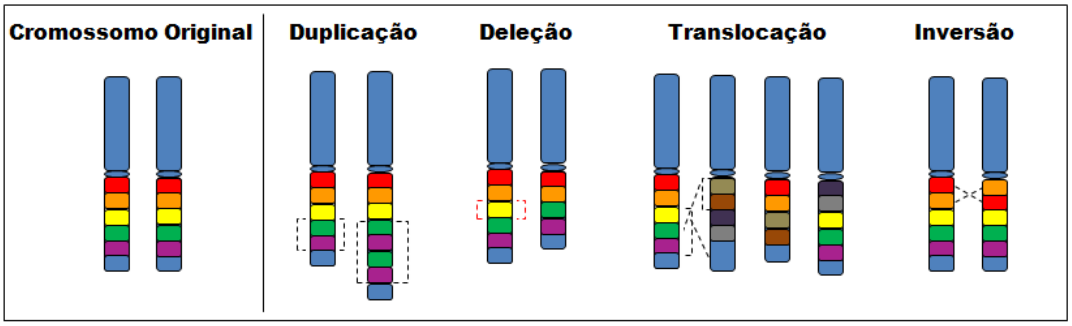

Figura 8 – Mutações de grande escala.

Fonte: Próprio autor.

2.2 Estruturas Algébricas

Nesta seção serão apresentadas as principais estruturas algébricas utilizadas neste trabalho, com algumas propriedades mais relevantes identificadas no estudo. As referências utilizadas são [3], [15] e [16].

2.2.1 Grupos

Definição 1. *Uma operação binaria $*$ sobre um conjunto S é uma regra que associa algum elemento de S a cada par ordenado (a, b) de elementos de S ($a*b$ denotará o elemento associado a (a, b) através de $*$).*

Definição 2. *Um conjunto G não vazio, ou seja, que possui elementos e uma operação $(a, b) \mapsto a * b$ sobre G, será um grupo se satisfazer os seguintes axiomas:*

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

Além dessas propriedades, se a operação satisfazer a propriedade de comutatividade: $a*b = b*a$, quaisquer que sejam $a, b \in G$, o grupo receberá o nome de grupo comutativo ou abeliano.

Teorema 1. *O elemento neutro é único.*

Demonstração. Suponha que existam dois elementos neutros, $e, e' \in G$. Assim temos:

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

ou seja, $e' = e$.
Logo, existe um único elemento neutro em G.

■

Teorema 2. *O elemento inverso é único.*

Demonstração. Suponha que existam dois elementos inversos, $a', a'' \in G$ para algum elemento $a \in G$. Assim temos:

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

Ou seja, $a' = a''$.
Desse modo, existe um único inverso para o elemento a em G.

■

Definição 3. *Seja G um grupo. O número de elementos de G é chamado de ordem de G, denotada por $|G|$. Um grupo de ordem finita é chamado grupo finito. Caso contrário, é denominado de grupo infinito.*

Definição 4. *As operações entre os elementos de um grupo finito podem ser representadas em uma tabela de operações conhecida como Tábua de Cayley, conforme apresentado na Tabela 1.*

Tabela 1 – Representação do grupo $(G, *)$, em que $G = \{e, ..., b, a\}$.

$*$	e	$\cdots$	a
e	e	$\cdots$	a
$\vdots$	$\vdots$		
a	a		$a * a$

Definição 5. *Grupos aditivos de classes de restos são para qualquer inteiro $m > 1$ o conjunto das classes de resto módulo m, ou seja, $\mathbb{Z}_m = \{\overline{0}, \overline{1}, \overline{2}, ..., \overline{m-1}\}$ é o conjunto quociente de $\mathbb{Z}$ pela relação de congruência, módulo m. Este grupo é representado por $(\mathbb{Z}_m, +)$, chamado grupo aditivo das classes de resto módulo m, onde sua ordem é m.*

Teorema 3. *$(\mathbb{Z}_m, +)$ é um grupo abeliano.*

Demonstração. A operação + é associativa:
Dados $\overline{a}$, $\overline{b}$, $\overline{c} \in \mathbb{Z}_m$, temos :

$$\overline{a} + (\overline{b} + \overline{c}) = \overline{a} + (\overline{b+c}) = \overline{a + (b+c)} = \overline{(a+b)+c} = \overline{(a+b)} + \overline{c} = (\overline{a} + \overline{b}) + \overline{c}.$$

O elemento neutro $e = \overline{0}$:
De fato,

$$\text{para todo } \overline{a} \in \mathbb{Z}_m \text{ , temos que } \overline{a} + \overline{0} = \overline{a+0} = \overline{a}.$$

Do mesmo modo,

$$\overline{0} + \overline{a} = \overline{0+a} = \overline{a}.$$

Por fim,

$$\text{dado } \overline{x} \in \mathbb{Z}_m, \text{ então } (\overline{x})^{-1} = \overline{n-x}.$$

De fato,

$$\overline{x} + (\overline{n-x}) = \overline{n} = \overline{0}.$$

Assim como,

$$(\overline{n-x}) + \overline{x} = \overline{n} = \overline{0}.$$

Além disso,

+ é comutativa em $\mathbb{Z}_m$, pois $\overline{a} + \overline{b} = \overline{a+b} = \overline{b+a} = \overline{b} + \overline{a}$, para todo $\overline{a}$, $\overline{b} \in \mathbb{Z}_m$.

■

Exemplo 1. $(\mathbb{Z}_4, +)$ *é um grupo.*
De fato, sabemos que $\mathbb{Z}_4$=$\{\overline{0}, \overline{1}, \overline{2}, \overline{3}\}$. *Sendo assim, dados* $\overline{1}$, $\overline{2}$, $\overline{3} \in \mathbb{Z}_4$, *temos pelo Teorema* 3 *que:*
$+$ *é associativa.*
Além disso, temos $e = \overline{0}$, *o elemento neutro.*
Por fim, temos a existência de simétrico.

Portanto, $(\mathbb{Z}_4, +)$ *é um grupo.*

Definição 6. *Um grupo multiplicativo de classes de restos, representado por* $(\mathbb{Z}_m, \cdot)$ *goza das propriedades associativa e comutativa. Além disso, a classe* $\overline{1}$ *é o seu elemento neutro. No entanto, devido* $\overline{0}$ *de* $\mathbb{Z}_m$ *não possuir inverso e o restante do conjunto nem sempre ser um grupo multiplicativo, há uma restrição da multiplicação módulo* m, *que será apresentada no Teorema 4 a seguir.*

Teorema 4. *Se* p *é um inteiro primo positivo, então* $(\mathbb{Z}_p^*, \cdot)$ *é um grupo abeliano.*

Demonstração. $\cdot$ é uma operação associativa, pois:

$$\overline{a} \cdot (\overline{b} \cdot \overline{c}) = \overline{a} \cdot (\overline{b \cdot c}) = \overline{a \cdot (b \cdot c)} = \overline{(a \cdot b) \cdot c} = \overline{(a \cdot b)} \cdot \overline{c} = (\overline{a} \cdot \overline{b}) \cdot \overline{c}$$

$\forall\ \overline{a}, \overline{b}, \overline{c} \in \mathbb{Z}_p^*$. Além disso, o elemento neutro de $\cdot$ é $e = \overline{1}$, pois:

$$\overline{a} \cdot \overline{1} = \overline{a \cdot 1} = \overline{a} \text{ e } \overline{1} \cdot \overline{a} = \overline{1 \cdot a} = \overline{a}.$$

Finalmente, dado $\overline{a} \in \mathbb{Z}_p^*$, com p primo, segue que $mdc(a, p) = 1$. Então, pelo Teorema de Bezout,[1] existem $r, s \in \mathbb{Z}$, tais que $ar + ps = 1$. Assim, $\overline{ar + ps} = \overline{1}$

$$(\overline{a} \cdot \overline{r}) + (\overline{p} \cdot \overline{s}) = \overline{1}$$

$\therefore\ \overline{a} \cdot \overline{r} = \overline{1}$, ou seja, todo elemento de $\mathbb{Z}_p^*$ é invertível. Portanto, $(\mathbb{Z}_p^*, \cdot)$ é um grupo. Como

$$\overline{a} \cdot \overline{b} = \overline{a \cdot b} = \overline{b \cdot a} = \overline{b} \cdot \overline{a}, \forall \overline{a}, \overline{b} \in \mathbb{Z}_m,$$

concluímos que $(\mathbb{Z}_p^*, \cdot)$ é um grupo abeliano.

■

2.2.2 Anéis

Definição 7. *Um conjunto não vazio* R *juntamente com duas operações binárias* $+$ *e* $\cdot$, *recebe o nome de anel, denotado por* $(R, +, \cdot)$, *quando satisfazer as seguintes condições:*

i) $(R, +)$ *é um grupo abeliano, ou seja:*

- *Associatividade:* $a + (b + c) = (a + b) + c$, $\forall a, b, c \in R$;

[1] Teorema de Bezout: Sejam a, p inteiros, não ambos nulos e seja $d = mdc(a, p)$. Então existem inteiros r, s tais que $d = ar + ps$.

- *Elemento neutro:* $\exists 0 \in R$*; tal que, qualquer que seja* $a \in R, a + 0 = 0 + a = a$*;*
- *Elemento oposto:* $\forall a \in R$*,* $\exists -a \in R$*;* $a + (-a) = 0 = (-a) + a$*.*
- *Comutatividade:* $\forall a, b \in R$*, então* $a + b = b + a$*;*

ii) *A multiplicação satisfaz a propriedade associativa, ou seja,* $\forall a, b, c \in R$*,* $a \cdot (b \cdot c) = (a \cdot b) \cdot c$*.*

iii) *A multiplicação é distributiva em relação à adição, à direita e à esquerda, ou seja,* $\forall a, b, c \in R$*,* $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ *e* $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$*.*

Definição 8. *Um anel* $(R, +, \cdot)$ *em que o conjunto* R *é finito chama-se anel finito. Se* R *é um anel finito, as tábuas da adição e da multiplicação podem ser instrumentos úteis para visualizar algumas de suas propriedades.*

Exemplo 2. *Considere as tábuas do anel* $\mathbb{Z}_4 = \{0, 1, 2, 3\}$*, apresentadas na Tabela 2.*

Tabela 2 – Tábuas da adição e da multiplicação do anel $\mathbb{Z}_4$.

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Exemplo 3. $\forall m \in \mathbb{Z}$*, com* $m > 1$*, temos que* $(\mathbb{Z}_m, +, \cdot)$ *é um anel, que recebe o nome de anel dos inteiros módulo* m*, sendo* $\cdot$ *uma operação comutativa e* $\overline{1}$ *é o elemento neutro desta operação.*

Exemplo 4. $(\mathbb{Z}_3, +, \cdot)$ *é um anel.*
De fato, sabemos que $\mathbb{Z}_3 = \{\overline{0}, \overline{1}, \overline{2}\}$*. Sendo assim,* $\overline{0}, \overline{1}, \overline{2} \in \mathbb{Z}_3$*. Com isso, da Definição 7-i):* $(\mathbb{Z}_3, +)$ *é um grupo abeliano.* $+$ *é comutativa:*

$$\overline{0} + \overline{1} = \overline{0 + 1} = \overline{1 + 0} = \overline{1} + \overline{0} = \overline{1}$$

$+$ *é associativa:*

$$\overline{0} + (\overline{1} + \overline{2}) = \overline{0} + (\overline{1 + 2}) = \overline{0 + (1 + 2)} = \overline{(0 + 1) + 2} = \overline{(0 + 1)} + \overline{2} = (\overline{0} + \overline{1}) + \overline{2}.$$

Existência de elemento neutro. Dado $\overline{2} \in \mathbb{Z}_3$*, temos que* $2 \in \mathbb{Z}$*. Também sabemos que* $0 \in \mathbb{Z}$ *e* $0 + 2 = 2 + 0$*. Então:*

$$\overline{2} = \overline{2 + 0} = \overline{2} + \overline{0} \text{ e } \overline{2} = \overline{0 + 2} = \overline{0} + \overline{2}.$$

Isto é, $\overline{0}$ *é o elemento neutro de* $\mathbb{Z}_3$*.*
Existência de elemento simétrico. Dado $\overline{2} \in \mathbb{Z}_3$*, temos que* $3 \in \mathbb{Z}$*. Também* $-2 \in \mathbb{Z}$ *e* $a - a = -a + a = 0$*. Então:*

$$\overline{0} = \overline{(-2)+2} = \overline{-2} + \overline{2} \text{ e } \overline{0} = \overline{2-2} = \overline{2} + \overline{(-2)}.$$

Isto é, $\overline{(-2)}$ *é o simétrico de* $\overline{2}$*. Da Definição 7-ii)a multiplicação é associativa:*

$$\overline{0} \cdot (\overline{1} \cdot \overline{2}) = \overline{0} \cdot (\overline{1 \cdot 2}) = \overline{0 \cdot (1 \cdot 2)} = \overline{(0 \cdot 1) \cdot 2} = (\overline{0 \cdot 1}) \cdot \overline{2} = (\overline{0} \cdot \overline{1}) \cdot \overline{2}.$$

Da Definição 7-iii) a multiplicação é distributiva em relação à adição:

$$\overline{0} \cdot (\overline{1} + \overline{2}) = \overline{0} \cdot (\overline{1+2}) = \overline{0 \cdot (1+2)} = \overline{0 \cdot 1 + 0 \cdot 2} = \overline{0 \cdot 1} + \overline{0 \cdot 2} = \overline{0} \cdot \overline{1} + \overline{0} \cdot \overline{2}.$$

Portanto, $(\mathbb{Z}_3, +, \cdot)$ *é um anel.*

2.2.3 Corpos

Definição 9. *Seja* $(R, +, \cdot)$ *um anel. Dizemos que* $(R, +, \cdot)$ *é corpo se* $(R - \{0\}, \cdot)$ *é um grupo abeliano.*

Teorema 5. $(\mathbb{Z}_p, +, \cdot)$ *com p primo, é um corpo.*

Demonstração. Vejamos se $(\mathbb{Z}_p - \{\overline{0}\}, \cdot)$ é um grupo abeliano.
$\cdot$ é associativa e comutativa, pois $(\mathbb{Z}_p, +, \cdot)$ é um anel comutativo.
$\overline{1} \in \mathbb{Z}_p - \{\overline{0}\}$ e dado $\overline{a} \in \mathbb{Z}_p - \{\overline{0}\}$, temos que,

$$\overline{1} \cdot \overline{a} = \overline{a} = \overline{a} \cdot \overline{1}.$$

$\therefore \overline{1}$ é elemento neutro para $\cdot$.
Dado $\overline{a} \in \mathbb{Z}_p - \{\overline{0}\}$, então $\overline{a} \neq \overline{0}$.
Logo, $a \in \mathbb{Z}$ e $p \nmid a$. Assim, $mdc(a, p) = 1$ o que implica pelo Teorema de Bezout que existem $r, s \in \mathbb{Z}$, tais que $ar + ps = 1$. Então,

$$\overline{ar + ps} = \overline{1},\ (\overline{a} \cdot \overline{r}) + (\overline{p} \cdot \overline{s}) = \overline{1}.$$

$\therefore \overline{a} \cdot \overline{r} = \overline{1}$, ou seja $\overline{r} = (\overline{a})^{-1}$.
Portanto,$(\mathbb{Z}_p, +, \cdot)$ é um corpo. ■

Definição 10. *Um anel* $(R, +, \cdot)$ *tal que:*

i) *a operação* $\cdot$ *é comutativa, recebe o nome de anel comutativo.*

ii) $\cdot$ *tem elemento neutro, recebe o nome de anel com unidade, ou simplesmente, anel com* 1*. Tal elemento será indicado por* 1 *ou* 1_R*.*

Definição 11. *Um domínio ou anel de integridade, é um anel comutativo com* 1 *e sem divisores de zero, ou seja, um* $(R, +, \cdot)$ *comutativo com* 1 *é domínio* $\Leftrightarrow \forall a, b \in R$*, tais que* $a \cdot b = 0$*, então* $a = 0$ *ou* $b = 0$*.*

Definição 12. *Um anel $(R, +, \cdot)$ é um anel com divisão, ou um quase corpo, se $(R - \{0\}, \cdot)$ é um grupo, ou seja, $\exists 1 \in R$ e $\forall a \in R, a \neq 0, \exists b \in R - \{0\}$, tal que $a \cdot b = b \cdot a = 1$. O elemento b recebe o nome de inverso de a e é denotado por a^{-1}.*

Teorema 6. *Todo corpo é um domínio, mais ainda, todo anel com divisão não tem divisores de zero.*

Demonstração. Se F é um corpo, então $(F - \{0\}, \cdot)$ é um grupo abeliano. (F é um anel comutativo com 1, no qual todo elemento não nulo tem inverso multiplicativo.) Se $a, b \in F$, são tais que:

$$a \cdot b = 0 \text{ e } a \neq 0,$$

então $\exists a^{-1} \in F - 0$, tal que $a \cdot a^{-1} = 1 = a^{-1} \cdot a$. Assim,

$$b = 1 \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot 0 = 0.$$

$\therefore F$ não tem divisores de zero $\Rightarrow F$ é um domínio. ■

Definição 13. *Um corpo K que apresenta um número finito de elementos, sendo esse número sua ordem, é chamado de corpo finito.*

Exemplo 5. *O conjunto $\mathbb{Z}_p = \{1, ..., p-1\}$, para todo p primo é um corpo finito.*

Definição 14. *Um corpo de Galois é um corpo com número finito de elementos e é representado por $GF(p)$, onde p é número primo.*

Para qualquer inteiro positivo m é possível estender um corpo primo $GF(p)$ com p elementos para um corpo estendido $GF(p^m)$ com p^m elementos. A ordem de qualquer corpo finito estendido é potência de um primo. Podemos construir códigos a partir de $GF(2)$ ou $GF(2^m)$ e, consequentemente, a aritmética usada é binária.

Definição 15. *Um polinômio $p(x)$ de grau m sobre $GF(2)$ é dito irredutível se ele não é divisível por nenhum outro polinômio sobre $GF(2)$ de grau menor que m mas maior que zero.*

Para qualquer $m \geq 1$ existe um polinômio irredutível de grau m.

Definição 16. *Um polinômio irredutível $p(x)$ de grau m é dito primitivo se o menor positivo inteiro n para $p(x)$ divide $X^n + 1$ for $n = 2^m - 1$.*

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

Em [5] é apresentado um método para construção de um corpo de Galois sobre $GF(2)$ com 2^m elementos, sendo $m \geq 1$. Sejam 0 e 1 os elementos de $GF(2)$, o símbolo α e a operação de multiplicação $\cdot$ definida da seguinte forma:

$$0 \cdot \alpha^j = \alpha^j \cdot 0 = 0 \ ;$$

$$1 \cdot \alpha^j = \alpha^j \cdot 1 = \alpha^j \ ;$$

.

.

.

$$\alpha^i \cdot \alpha^j = \alpha^j \cdot \alpha^i = \alpha^{i+j}.$$

Assim, temos um conjunto de elementos sobre o qual a operação "$\cdot$" é definida:

$$F = \{0, 1, \alpha, \alpha^2, \alpha^3, ..., \alpha^j, ...\}.$$

Vamos estabelecer uma condição sobre o elemento α para considerar que o conjunto F tenha 2^m elementos, sendo fechado para a operação de multiplicação definida anteriormente. Seja $p(x)$ um polinômio primitivo de grau m sobre $GF(2)$ e seja α uma raiz de $p(x)$, ou seja, p$(\alpha) = 0$.

Como $p(x)$ divide $x^{2^m-1} + 1$ temos que:

$$x^{2^m-1} + 1 = q(x) \cdot p(x),$$

em que $q(x)$ é um polinômio qualquer sobre $GF(2)$.

Substituindo x por α:

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

Adicionando 1 em ambos lados da igualdade, obtemos:

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

Portanto, existe um elemento $\alpha^{2^m-1} \neq 0$ a partir do qual os elementos do conjunto F tornam-se finitos. Assim, $F = \{0, 1, \alpha, \alpha^2, \alpha^3, ..., \alpha^{2^m-2}\}$ é um corpo de Galois com 2^m elementos.

No processo de construção de $GF(2^m)$ a partir $GF(2)$, os elementos não nulos de $GF(2^m)$ podem ser representados por potências, polinômios ou vetores.

Exemplo 6. *Seja $m = 5$ e considere o polinômio primitivo sobre $GF(2)$, $p(x) = 1 + x^2 + x^5$. O conjunto F será dado da seguinte forma:*

$$F = \{0, 1, \alpha, \alpha^2, \alpha^3, ..., \alpha^{2^5-2}\} \Rightarrow F = \{0, 1, \alpha, \alpha^2, \alpha^3, ..., \alpha^{30}\}.$$

Admitindo que α seja uma raiz desse polinômio, ou seja, $p(\alpha) = 0$, temos:

$$0 = 1 + \alpha^2 + \alpha^5 \Rightarrow \alpha^5 = 1 + \alpha^2.$$

A partir dessa relação podemos construir $GF(2^5)$:

$$\alpha^6 = \alpha^5 \cdot \alpha = (1+\alpha^2)\cdot\alpha = \alpha + \alpha^3$$
$$\alpha^7 = \alpha^6 \cdot \alpha = (\alpha+\alpha^3)\cdot\alpha = \alpha^2 + \alpha^4$$
$$\alpha^8 = \alpha^7 \cdot \alpha = (\alpha^2+\alpha^4)\cdot\alpha = \alpha^3 + \alpha^5 = \alpha^3 + 1 + \alpha^2 = 1 + \alpha^2 + \alpha^3$$
$$\alpha^9 = \alpha^8 \cdot \alpha = (1+\alpha^2+\alpha^3)\cdot\alpha = \alpha + \alpha^3 + \alpha^4$$
$$\alpha^{10} = \alpha^9 \cdot \alpha = (\alpha+\alpha^3+\alpha^4)\cdot\alpha = \alpha^2 + \alpha^4 + \alpha^5 = \alpha^2 + \alpha^4 + 1 + \alpha^2 = 1 + \alpha^4$$
$$\alpha^{11} = \alpha^{10} \cdot \alpha = (1+\alpha^4)\cdot\alpha = \alpha + \alpha^5 = \alpha + 1 + \alpha^2 = 1 + \alpha + \alpha^2$$
$$\alpha^{12} = \alpha^{11} \cdot \alpha = (1+\alpha+\alpha^2)\cdot\alpha = \alpha + \alpha^2 + \alpha^3$$
$$\alpha^{13} = \alpha^{12} \cdot \alpha = (\alpha+\alpha^2+\alpha^3)\cdot\alpha = \alpha^2 + \alpha^3 + \alpha^4$$
$$\alpha^{14} = \alpha^{13} \cdot \alpha = (\alpha^2+\alpha^3+\alpha^4)\cdot\alpha = \alpha^3 + \alpha^4 + \alpha^5 = \alpha^3 + \alpha^4 + 1 + \alpha^2 = 1 + \alpha^2 + \alpha^3 + \alpha^4$$
$$\alpha^{15} = \alpha^{14} \cdot \alpha = (1+\alpha^2+\alpha^3+\alpha^4)\cdot\alpha = \alpha + \alpha^3 + \alpha^4 + \alpha^5 = \alpha + \alpha^3 + \alpha^4 + 1 + \alpha^2 = 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4$$
$$\alpha^{16} = \alpha^{15} \cdot \alpha = (1+\alpha+\alpha^2+\alpha^3+\alpha^4)\cdot\alpha = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + 1 + \alpha^2 = 1 + \alpha + \alpha^3 + \alpha^4$$
$$\alpha^{17} = \alpha^{16} \cdot \alpha = (1+\alpha+\alpha^3+\alpha^4)\cdot\alpha = \alpha + \alpha^2 + \alpha^4 + \alpha^5 = \alpha + \alpha^2 + \alpha^4 + 1 + \alpha^2 = 1 + \alpha + \alpha^4$$
$$\alpha^{18} = \alpha^{17} \cdot \alpha = (1+\alpha+\alpha^4)\cdot\alpha = \alpha + \alpha^2 + \alpha^5 = \alpha + \alpha^2 + 1 + \alpha^2 = 1 + \alpha$$
$$\alpha^{19} = \alpha^{18} \cdot \alpha = (1+\alpha)\cdot\alpha = \alpha + \alpha^2$$
$$\alpha^{20} = \alpha^{19} \cdot \alpha = (\alpha+\alpha^2)\cdot\alpha = \alpha^2 + \alpha^3$$
$$\alpha^{21} = \alpha^{20} \cdot \alpha = (\alpha^2+\alpha^3)\cdot\alpha = \alpha^3 + \alpha^4$$
$$\alpha^{22} = \alpha^{21} \cdot \alpha = (\alpha^3+\alpha^4)\cdot\alpha = \alpha^4 + \alpha^5 = \alpha^4 + 1 + \alpha^2 = 1 + \alpha^2 + \alpha^4$$
$$\alpha^{23} = \alpha^{22} \cdot \alpha = (1+\alpha^2+\alpha^4)\cdot\alpha = \alpha + \alpha^3 + \alpha^5 = \alpha + \alpha^3 + 1 + \alpha^2 = 1 + \alpha + \alpha^2 + \alpha^3$$
$$\alpha^{24} = \alpha^{23} \cdot \alpha = (1+\alpha+\alpha^2+\alpha^3)\cdot\alpha = \alpha + \alpha^2 + \alpha^3 + \alpha^4$$
$$\alpha^{25} = \alpha^{24} \cdot \alpha = (\alpha+\alpha^2+\alpha^3+\alpha^4)\cdot\alpha = \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 = \alpha^2 + \alpha^3 + \alpha^4 + 1 + \alpha^2 = 1 + \alpha^3 + \alpha^4$$
$$\alpha^{26} = \alpha^{25} \cdot \alpha = (1+\alpha^3+\alpha^4)\cdot\alpha = \alpha + \alpha^4 + \alpha^5 = \alpha + \alpha^4 + 1 + \alpha^2 = 1 + \alpha + \alpha^2 + \alpha^4$$
$$\alpha^{27} = \alpha^{26} \cdot \alpha = (1+\alpha+\alpha^2+\alpha^4)\cdot\alpha = \alpha + \alpha^2 + \alpha^3 + \alpha^5 = \alpha + \alpha^2 + \alpha^3 + 1 + \alpha^2 = 1 + \alpha + \alpha^3$$
$$\alpha^{28} = \alpha^{27} \cdot \alpha = (1+\alpha+\alpha^3)\cdot\alpha = \alpha + \alpha^2 + \alpha^4$$
$$\alpha^{29} = \alpha^{28} \cdot \alpha = (\alpha+\alpha^2+\alpha^4)\cdot\alpha = \alpha^2 + \alpha^3 + \alpha^5 = \alpha^2 + \alpha^3 + 1 + \alpha^2 = 1 + \alpha^3$$
$$\alpha^{30} = \alpha^{29} \cdot \alpha = (1+\alpha^3)\cdot\alpha = \alpha + \alpha^4$$

.

As representações vetoriais, polinomiais e por potência de $GF(2^5)$são apresentadas na Tabela 3.

2.3 Códigos Corretores de Erros

Nesta seção serão apresentados os conceitos fundamentais referentes aos códigos corretores de erros, utilizados neste trabalho. As referências utilizadas foram [1], [2], [3], [4], [5], [6], [17], [18], [19], [20], [21], [22], [23].

2.3.1 Elementos de um sistema de comunicação

Definição 17. *Um sistema de comunicação é um conjunto de mecanismos que tem como objetivo transmitir informações de uma fonte a um destinatário via um canal de comunicação. E pode ser dividido em dois sistemas:*

Tabela 3 – $GF(2^5)$ gerado por $p(x) = 1 + x^2 + x^5$.

Representações					
Por Potência	**Polinomial**	**Vetorial**	**Por Potência**	**Polinomial**	**Vetorial**
0	0	(00000)	α^{15}	$1+\alpha+\alpha^2+\alpha^3+\alpha^4$	(11111)
$\alpha^0 = 1$	1	(10000)	α^{16}	$1+\alpha+\alpha^3+\alpha^4$	(11011)
α	α	(01000)	α^{17}	$1+\alpha+\alpha^4$	(11001)
α^2	α^2	(00100)	α^{18}	$1+\alpha$	(11000)
α^3	α^3	(00010)	α^{19}	$\alpha+\alpha^2$	(01100)
α^4	α^4	(00001)	α^{20}	$\alpha^2+\alpha^3$	(00110)
α^5	$1+\alpha^2$	(10100)	α^{21}	$\alpha^3+\alpha^4$	(00011)
α^6	$\alpha+\alpha^3$	(01010)	α^{22}	$1+\alpha^2+\alpha^4$	(10101)
α^7	$\alpha^2+\alpha^4$	(00101)	α^{23}	$1+\alpha+\alpha^2+\alpha^3$	(11110)
α^8	$1+\alpha^2+\alpha^3$	(10110)	α^{24}	$\alpha+\alpha^2+\alpha^3+\alpha^4$	(01111)
α^9	$\alpha+\alpha^3+\alpha^4$	(01011)	α^{25}	$1+\alpha^3+\alpha^4$	(10011)
α^{10}	$1+\alpha^4$	(10001)	α^{26}	$1+\alpha+\alpha^2+\alpha^4$	(11101)
α^{11}	$1+\alpha+\alpha^2$	(11100)	α^{27}	$1+\alpha+\alpha^3$	(11010)
α^{12}	$\alpha+\alpha^2+\alpha^3$	(01110)	α^{28}	$\alpha+\alpha^2+\alpha^4$	(01101)
α^{13}	$\alpha^2+\alpha^3+\alpha^4$	(00111)	α^{29}	$1+\alpha^3$	(10010)
α^{14}	$1+\alpha^2+\alpha^3+\alpha^4$	(10111)	α^{30}	$\alpha+\alpha^4$	(01001)

- *sistemas analógicos: contém dispositivos que manipulam quantidades físicas, podendo variar ao longo de uma faixa contínua de valores, onde são representadas na forma analógica.*
- *sistemas digitais: são aqueles que sofrem ruídos durante o processo de transmissão da informação, podendo causar erros e por consequência impede que a informação seja reproduzida fielmente ao destinatário.*

Um sistema de comunicação possui os seguintes elementos:

1. Transmissor: responsável por gerar a informação, no qual localizam-se :
   - Fonte: local em que gera o sinal.
   - Codificador de fonte: realiza a conversão do sinal da saída da fonte em uma sequência de dígitos binários, que são os códigos.
2. Canal: região em que é transmitida a informação e onde podem ser introduzidos os ruídos.
   - Codificador de canal: transforma a sequência da saída do codificador de fonte em uma palavra-código (dígitos binários), através da redundância para eliminar os efeitos ruidosos adquiridos no canal.
   - Modulador: converte a saída do codificador de canal para uma forma adequada para ser transmitida.

- Demodulador: com o sinal recebido do canal, se estima sua versão digital e é enviada para o codificador de canal.
- Decodificador de canal: realiza uma tentativa de corrigir alguns erros que possam aparecer nos dígitos fornecidos pelo demodulador, estimando os dígitos na saída do codificador da fonte.

3. Receptor: representa o usuário que vai receber a informação.

- Decodificador de fonte: é o local que transforma a sequência estimada na saída do decodificador de canal em uma estimativa na saída da fonte.
- Destinatário: é quem recebe a informação transmitida.

A Figura 9 apresenta um diagrama de blocos de um sistema de comunicação.

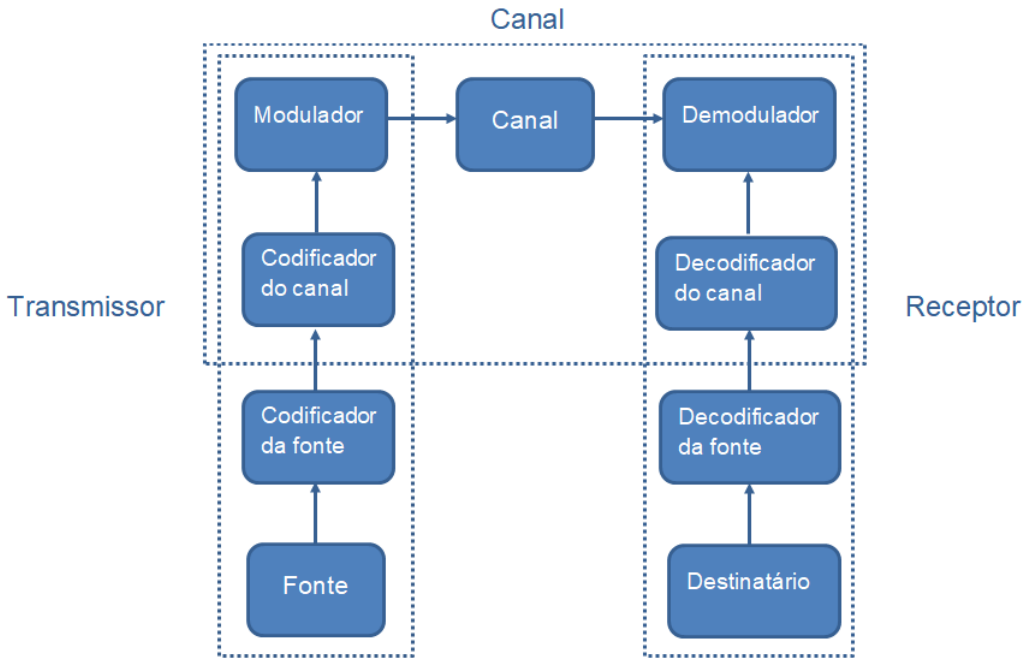

Figura 9 – Diagrama de blocos de um sistema de comunicação.

Fonte: Próprio autor.

2.3.2 Códigos corretores de erros

A utilização dos códigos para correção de erros se desenvolveu por volta de 1947, através de Richard W. Hamming no laboratório Bell de Tecnologia, onde ele trabalhava na análise dos erros que surgiam na transmissão de informações. Neste local, Hamming foi capaz de desenvolver uma teoria, que possibilita não só detectar os erros, mas corrigi-los.

Posteriormente com o auxílio de Shannon (1948) e Golay(1949), Hamming desenvolveu os primeiro trabalhos com os códigos corretores de erros, que possibilitou por meio de pesquisas e estudos,

um grande avanço na área de comunicação móvel, aparelhos de armazenamento de dados, internet, comunicação via satélite, entre outras.

Com isso, os códigos corretores de erros são utilizados no processo de transmissão e armazenamento de dados, sendo capazes de detectar e corrigir possíveis erros que possam surgir, garantindo a confiabilidade durante esses processos.

Atualmente, os códigos corretores de erros podem ser divididos em dois grupos: códigos de blocos e códigos convolucionais, sendo que cada um dos grupos possui diversos códigos, para as mais diversas aplicações.

Os códigos convolucionais possuem caráter probabilístico, pois a princípio, alguns matemáticos procuraram estimar a probabilidade de erros das "melhores" famílias de códigos de blocos e tinham o objetivo de compreender a codificação e a decodificação de um ponto de vista probabilístico, considerando a noção de decodificação sequencial. Nesse contexto, a decodificação sequencial exigiu a criação de uma nova classe de códigos sem blocos, com comprimento indefinido, representáveis por uma árvore em que a decodificação é feita percorrendo toda a extensão dessa árvore.

Os códigos de bloco são caracterizados pelo fato do processo de codificação ser feito sobre blocos de bits ou bloco de símbolos, isto é, 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 caracterizada como a relação entre o número de bits de informação e o número de bits da palavra-código, ou seja, $R = \frac{k}{n}$.

Definição 18. *Um código de bloco C de comprimento n sobre um alfabeto A é qualquer subconjunto do conjunto A^n das sequências $c = \{c_i | 1 \leq i \leq n\}$. Um código de bloco é caracterizado por três parâmetros: comprimento, dimensão e distância mínima, ou seja, (n, k, d_{min}).*

Definição 19. *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 $\omega_t(x)$.*

Exemplo 7. *Seja a palavra-código $x = (1010011)$.*
O peso de Hamming dessa palavras-código é $\omega(x) = 4$.

Definição 20. *A distância de Hamming, denotada por $dist(x, y)$, é definida como o número de elementos em que dois vetores $x = \{\$x_1, ..., x_n\}$ e $y = \{\$y_1, ..., y_n\}$ se diferem. Para o caso binário, a distância de Hamming pode ser determinada facilmente pela propriedade de adição módulo-2, pois ela é igual ao número de dígitos "1" contidos no vetor resultante da operação $v + x$.*

$$dist(x, y) = \omega(x + y).$$

Exemplo 8. *Determine a distância de Hamming entre os vetores $x = 10101$ e $y = 10110$.*

$$dist(x, y) = \omega(x + y) = \omega(10110 + 10101) = \omega(00011) = dist(x, y) = 2.$$

Definição 21. *A distância mínima d_{min} de um código de blocos C, é a menor distância de Hamming entre dois vetores distintos quaisquer desse código. Ou seja, $d_{min} = \{dist(x, y) : x, y \in C, x \neq y\}$.*

Exemplo 9. *Determine a distância mínima do código* $C = \{(01011), (11110)\}$ *e* $\{(01101)\}$.

$$d_{min} = dist((01011), (11110), (01101)) = \omega(01011 + 11110 + 10101) = \omega(11000) = 2.$$

Definição 22. *A capacidade de correção de erros é o número máximo de erros que podem ser corrigidos por palavra-código, e é dada por:* $t = (d_{min} - 1)/2$*, onde* t *é o maior inteiro não superior a* $(d_{min} - 1)/2$.

A capacidade de correção de erros, t*, está relacionada a distância mínima do código da seguinte maneira:* $d_{min} \leq 2t + 1$*, ou seja, quanto maior a capacidade de correção de erros de um código, maior a sua distância mínima.*

Definição 23. *Um código de bloco de comprimento* n *e* 2^k *palavras-código, é um código linear se, e só se, as suas* 2^k *palavras-código formam um subespaço de dimensão* k *em relação ao espaço formado pelas* 2^n *n-uplas possíveis em* $GF(2)$.

Definição 24. *Sejam* $C(n, k)$ *um código linear sobre* F_q *(ou* $GF(q)$*) um corpo finito com* q *elementos e* $\beta = \{g_0, g_1, ..., g_{k-1}\}$ *uma base de* C*. A matriz* G *de ordem* $k \times n$ *cujas linhas são* $g_0, g_1, ..., g_{k-1}$*, é denominada matriz geradora do código linear* C*. Uma matriz geradora,* G*, é aquela que permite obter os vetores códigos,* c_j*, correspondentes às mensagens* m_i*, a partir do produto interno determinado por:*

$$c_j = m_i.G$$

, onde

$$G = \begin{bmatrix} g_0 \\ g_1 \\ . \\ . \\ . \\ g_{k-1} \end{bmatrix} = \begin{bmatrix} g_{00} & g_{01} & g_{02} & \cdots & g_{0,n-1} \\ g_{10} & g_{11} & g_{12} & \cdots & g_{1,n-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ g_{k-1,0} & g_{k-1,1} & g_{k-1,2} & \cdots & g_{k-1,n-1} \end{bmatrix}$$

e $g_0, g_1, ..., g_{k-1}$, são os vetores geradores.

Definição 25. *Associada à matriz geradora,* G*, existe uma outra matriz de dimensão* $(n - k) \times n$*, chamada matriz de verificação de paridade, que é denotada por* H*, com* $n - k$ *linhas linearmente independentes, de forma que as linhas da matriz* G *sejam ortogonais às linhas da matriz* H.

Ou seja, $G \cdot H^T = 0$, em que H^T é a matriz transposta de H. Com isso, podemos descrever um código linear (n, k) gerado por G de uma forma diferente, pois uma palavra-código v do código gerado por G é palavra código se somente se $v \cdot H^T = 0$, sendo 0 o vetor nulo.

$$H = \begin{bmatrix} h_0 \\ h_1 \\ . \\ . \\ . \\ h_{n-k-1} \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} & ... & h_{0,n-1} \\ h_{10} & h_{11} & h_{12} & ... & h_{1,n-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ h_{n-k-1,0} & h_{n-k-1,1} & h_{n-k-1,2} & ... & h_{nk-1,n-1} \end{bmatrix}.$$

2.3.3 Códigos BCH

Os códigos cíclicos formam uma importante subclasse dos códigos de bloco lineares. Um código de bloco C é denominado cíclico se qualquer deslocamento de uma palavra código resulta em uma outra palavra código. Estes códigos possuem grandes possibilidades de aplicações práticas, como no compact disc (CD) e no Nasa Deep Space Standard para a comunicações via satélite.

Seja um código linear C e $v = (v_0, v_1, v_2, ..., v_{n-1})$ um vetor de C. Se as componentes do vetor v forem deslocadas uma posição para a direita e a última componente v_{n-1} for deslocada para primeira posição à esquerda, obtemos o seguinte vetor: $v(1) = (v_{n-1}, v_0, v_1, v_2, ..., v_{n-2})$. Esse processo de deslocamento das componentes de um vetor é chamado de descolamento cíclico de v.

Definição 26. *Um código linear $C(n, k)$ é um código cíclico se qualquer deslocamento cíclico de uma palavra-código de C resulta em uma outra palavra-código do código C.*

Podemos representar a palavra código $v = (v_0, v_1, v_2, ..., v_{n-1})$ pelo polinômio-código $v(x) = v_0 + v_1x + v_2x^2 + ... + v_{n-1}x^{n-1}$.

Definição 27. *Os códigos de Hamming são códigos de blocos lineares, desenvolvidos por Richard Hamming, baseados na adição de bits de paridade. Assim, é possível detectar erros por meio da adição de bits de paridade a um determinado número de bits de dados.*

Os códigos BCH (Bose, Chaudhuri e Hocquenghen) são uma importante classe cíclica de códigos corretores de erros, em que são utilizados corpos finitos em sua construção, admitindo representação em termos de polinômios sobre $GF(p)$. Os códigos BCH binários são uma generalização dos códigos de Hamming e podem corrigir múltiplos erros.

Definição 28. *Para qualquer inteiro $m \geq 3$ e $t < 2^{m-1}$ existe um código BCH binário com capacidade de correção de t erros, com os seguintes parâmetros:*

- *Comprimento do bloco: $n = 2^m - 1$;*
- *Número de dígitos de verificação de paridade: $n - k \leq mt$;*
- *Distância mínima: $d_{min} \geq 2^t + 1$.*

O código BCH definido anteriormente é gerado por um polinômio que é especificado em termos de suas raízes no corpo finito $GF(2^m)$.

Definição 29. *Seja* $g(x) = 1 + g_1x + g_2x^2 + ... + g_{n-k-1}x^{n-k-1} + g_{n-k}x^{n-k}$ *um polinômio não nulo de grau mínimo* $n-k$ *de um código cíclico binário* $C(n,k)$*. O polinômio* $g(x)$ *é chamado de polinômio gerador de* $C(n,k)$*.*

Definição 30. *Seja* $f(x)$ *um polinômio com coeficientes em* $GF(2)$*. Se um elemento* β *de* $GF(2^m)$ *é uma raiz de* $f(x)$*, então o polinômio* $f(x)$ *também tem como raízes* β^{2^l} *para qualquer* $l \geq 0$*. Este elemento é chamado de conjugado de* β*.*

O polinômio gerador de um código BCH é o polinômio de menor grau sobre o corpo de Galois $GF(2^m)$ que tem $\alpha, \alpha^2, \alpha^3, ..., \alpha^{2t}$ e seus conjugados como todas suas raízes. Se α é um elemento primitivo de $GF(2^m)$, então o código BCH resultante é um código BCH primitivo.

Definição 31. *O polinômio mínimo* $\phi(x)$ *de um elemento* β *em* $GF(2^m)$ *é dado por:*

$$\phi(x) = \prod_{i=1}^{t-1}(x + \beta^{2^i}) \ .$$

Seja $\phi(x)$ o polinômio mínimo de α^i. Então o polinômio gerador $g(x)$ é dado pelo mínimo múltiplo comum dos polinômios mínimos $\{\phi_1(x), \phi_2(x), ..., \phi_{2t}(x)\}$:

$$g(x) = MMC\{\phi_1(x), \phi_2(x), ..., \phi_{2t}(x)\} \ .$$

Definição 32. *O polinômio gerador* $g(x)$ *de um código BCH binário com comprimento* $2^m - 1$ *e capacidade de correção de* t *erros é dado por:*

$$g(x) = MMC\{\phi_1(x), \phi_3(x), ..., \phi_{2t-1}(x)\}.$$

Um código BCH de comprimento $2^m - 1$, com capacidade de correção $t = 1$ (um único erro) é gerado por $g(x) = \phi_1(x)$, no qual $\phi_1(x)$ é um polinômio primitivo de grau m, e é um código de Hamming.

Exemplo 10. *Considere o Exemplo 7 e realize a construção de um polinômio gerador do código BCH com capacidade de correção de cinco erros a partir de* $GF(2^5)$*, gerado por* $p(x) = 1 + x^2 + x^5$*.*

Primeiramente obtemos todos elementos de $GF(2^5)$ *gerado por* $p(x) = 1 + x^2 + x^5$*, apresentados na Tabela 3 no Exemplo 7.*

Em seguida, admitindo um dado elemento β *de* $GF(2^5)$*, encontramos as suas raízes conjugadas,* β^{2^l} *, com* $l \geq 0$*. Por exemplo, admitindo que* $\beta = \alpha$*, temos as seguintes raízes conjugadas referentes:*

$$(\alpha)^{2^1} = \alpha^2; (\alpha)^{2^2} = \alpha^4; (\alpha)^{2^3} = \alpha^8; (\alpha)^{2^4} = \alpha^{16},$$

do mesmo modo, encontramos as raízes conjugadas para $\beta = \alpha^3$*;* $\beta = \alpha^5$*;* $\beta = \alpha^7$*;* $\beta = \alpha^{11}$ *e* $\beta = \alpha^{15}$*, encontradas na Tabela 4. Com isso, construímos os polinômios mínimos referentes às raízes conjugadas encontradas anteriormente, por meio do seguinte produtório:*

Tabela 4 – Raízes conjugadas de $GF(2^5)$ geradas por $p(x) = 1 + x^2 + x^5$.

β	β^{2^l}	**Raízes conjugadas**
0	0	0
1	1	1
α	$\alpha^2, \alpha^4, \alpha^8, \alpha^{16}$	$\alpha, \alpha^2, \alpha^4, \alpha^8, \alpha^{16}$
α^3	$\alpha^6, \alpha^{12}, \alpha^{24}, \alpha^{48} = \alpha^{17}$	$\alpha^3, \alpha^6, \alpha^{12}, \alpha^{17}, \alpha^{24}$
α^5	$\alpha^{10}, \alpha^{20}, \alpha^{40} = \alpha^9, \alpha^{80} = \alpha^{18}$	$\alpha^5, \alpha^9, \alpha^{10}, \alpha^{18}, \alpha^{20}$
α^7	$\alpha^{14}, \alpha^{28}, \alpha^{56} = \alpha^{25}, \alpha^{112} = \alpha^{19}$	$\alpha^7, \alpha^{14}, \alpha^{19}, \alpha^{25}, \alpha^{28}$
α^{11}	$\alpha^{22}, \alpha^{44} = \alpha^{13}, \alpha^{88} = \alpha^{26}, \alpha^{176} = \alpha^{21}$	$\alpha^{11}, \alpha^{13}, \alpha^{21}, \alpha^{22}, \alpha^{26}$
α^{15}	$\alpha^{30}, \alpha^{60} = \alpha^{29}, \alpha^{120} = \alpha^{27}, \alpha^{240} = \alpha^{23}$	$\alpha^{15}, \alpha^{23}, \alpha^{27}, \alpha^{29}, \alpha^{30}$

$$\phi(x) = \prod_{i=0}^{t-1}(x + \beta^{2^i}).$$

Devemos encontrar $2t - 1$ *polinômios mínimos, pois o polinômio gerador é dado pelo produto dos polinômios mínimos* $\phi_1(x)$, $\phi_3(x)$, ..., $\phi_{2t-1}(x)$. *Deste modo, como o código em questão tem capacidade de correção de 5 cinco erros* $(t = 5)$,

$$2t - 1 = 2 \cdot 5 - 1 = 9.$$

Como $\phi_9(x)$ *não corresponde a nenhuma raíz conjugada, vamos calcular os seguintes polinômios mínimos referentes às raízes que encontramos:* $\phi_1(x), \phi_3(x)$, $\phi_5(x)$ *e* $\phi_7(x)$.

Para $\phi_1(x)$, *temos a seguinte relação:*

$$\phi_1(x) = (x + \alpha)(x + \alpha^2)(x + \alpha^4)(x + \alpha^8)(x + \alpha^{16}).$$

Efetuando os cálculos, obtemos:

$$\phi_1(x) = x^5 + x^4(\alpha^{16} + \alpha^8 + \alpha^4 + \alpha^2 + \alpha) + x^3(\alpha^{24} + \alpha^{20} + \alpha^{18} + \alpha^{17} + \alpha^{10} + \alpha^9 + \alpha^8 + \alpha^6 + \alpha^5 + \alpha^3) + x^2(\alpha^{28} + \alpha^{26} + \alpha^{25} + \alpha^{22} + \alpha^{21} + \alpha^{19} + \alpha^{14} + \alpha^{13} + \alpha^{11} + \alpha^7) + x(\alpha^{30} + \alpha^{29} + \alpha^{27} + \alpha^{23} + \alpha^{15}) + \alpha^{31}$$

$$\phi_1(x) = x^5 + x^4(0) + x^3(0) + x^2(1) + x(0) + 1$$

$$\phi_1(x) = x^5 + x^2 + 1.$$

Para $\phi_3(x)$, $\phi_5(x)$ *e* $\phi_7(x)$ *o processo é análogo, atentando-se sempre às raízes conjugadas referentes a cada polinômio mínimo. Os polinômios mínimos são apresentados na Tabela 5.*

Para obter o polinômio gerador, dado por $g(x) = MMC\{\phi_1(x), \phi_3(x), ..., \phi_{2t-1}(x)\}$, *efetuamos o produto dos polinômios mínimos* $\phi_1(x)$, $\phi_3(x)$, ..., $\phi_{2t-1}(x)$, *encontrados anteriormente.*

Neste caso, o polinômio gerador pode ser descrito da seguinte forma:

$$g(x) = MMC\{\phi_1(x), \phi_3(x), \phi_5(x), \phi_7(x)\}$$

Tabela 5 – Polinômios minimais de $GF(2^5)$ gerados por $p(x) = 1 + x^2 + x^5$.

Raízes conjugadas	**Polinômios mínimos**
0	$\phi(x) = x$
1	$\phi_0(x) = x + 1$
$\alpha, \alpha^2, \alpha^4, \alpha^8, \alpha^{16}$	$\phi_1(x) = (x+\alpha)(x+\alpha^2)(x+\alpha^4)(x+\alpha^8)(x+\alpha^{16})$ $\phi_1(x) = x^5 + x^2 + 1$
$\alpha^3, \alpha^6, \alpha^{12}, \alpha^{17}, \alpha^{24}$	$\phi_3(x) = (x+\alpha^3)(x+\alpha^6)(x+\alpha^{12})(x+\alpha^{17})(x+\alpha^{24})$ $\phi_3(x) = x^5 + x^4 + x^3 + x^2 + 1$
$\alpha^5, \alpha^9, \alpha^{10}, \alpha^{18}, \alpha^{20}$	$\phi_5(x) = (x+\alpha^5)(x+\alpha^9)(x+\alpha^{10})(x+\alpha^{18})(x+\alpha^{20})$ $\phi_5(x) = x^5 + x^4 + x^2 + x + 1$
$\alpha^7, \alpha^{14}, \alpha^{19}, \alpha^{25}, \alpha^{28}$	$\phi_7(x) = (x+\alpha^7)(x+\alpha^{14})(x+\alpha^{19})(x+\alpha^{25})(x+\alpha^{28})$ $\phi_7(x) = x^5 + x^3 + x^2 + x + 1$

$$g(x) = (x^5 + x^2 + 1) \cdot (x^5 + x^4 + x^3 + x^2 + 1) \cdot (x^5 + x^4 + x^2 + x + 1) \cdot (x^5 + x^3 + x^2 + x + 1),$$

e obtemos:

$$g(x) = x^{20} + x^{18} + x^{17} + x^{13} + x^{10} + x^9 + x^7 + x^6 + x^4 + x^2 + 1,$$

gerando um código BCH cíclico com $d_{min} \geq 11$.

É possível estabelecer uma relação entre o grau do polinômio gerador e os parâmetros do código BCH. Dado o polinômio gerador $g(x)$ *e sendo seu grau* $gr(g(x)) = 20$*, podemos encontrar os parâmetros do código BCH com capacidade de correção de cinco erros através da seguinte igualdade:*

$$gr(g(x)) = n - k. \ (I)$$

O parâmetro n *de um código BCH* (n, k) *é dado por* $n = 2^m - 1$*. Como* $m = 5$*, pois o código é gerado por* $p(x) = 1 + x^2 + x^5$ *a partir de* $GF(2^5)$*, temos que:*

$$n = 2^m - 1 \Rightarrow n = 2^5 - 1 \Rightarrow n = 31. \ (II)$$

De (I) *e* (II)*, segue que:*

$$20 = n - k \Rightarrow 20 = 31 - k \Rightarrow k = 11.$$

Portanto, o código em questão é o código $BCH(31, 11)$.

2.4 Algoritmo de Geração de Proteínas

Nesta seção serão apresentados os passos referentes ao algoritmo de geração de proteínas utilizado neste trabalho. As referências utilizadas foram [1], [2], [3], [4], [5], [6].

2.4.1 Descrição do algoritmo de geração de proteínas

O algoritmo de geração de proteínas, proposto por [6] e [2], identifica e reproduz diferentes sequências de DNA por meio dos códigos BCH, possibilitando o reconhecimento de uma estrutura de códigos corretores de erros em sequências de DNA, além de permitir uma nova classificação dessas sequências sob um ponto de vista matemático. Esse algoritmo pode ser descrito em 17 passos, conforme foi detalhado em [5].

Passo 1: Especifica-se a estrutura matemática e o alfabeto do código. Tem-se um alfabeto 4-ário do código genético, que é composto pelas bases nitrogenadas e denotado pelo conjunto $N = \{A, C, G, T/U\}$. Assim, ele é relacionado com o alfabeto 4-ário dos CCE's sobre uma estrutura de anel indicado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, seguindo as operações de adição e multiplicação módulo 4.

Passo 2: Determina-se a extensão de Galois. O comprimento da sequência de DNA deve obedecer a restrição de $n = 2^r - 1$, sendo r o grau da extensão.

Passo 3: Determina-se todos os polinômios primitivos $p(x)$ relacionados à extensão de Galois, determinada no passo anterior.

Passo 4: Determina-se a extensão do corpo $GF(2)$. Para isso, considera-se uma extensão do corpo $GF(2)$ e um polinômio primitivo $p(x)$. Então toma-se um elemento α como raiz desse polinômio e a partir disso, determina-se todos os elementos desse conjunto, que são representados por potências de α. O último elemento que compõe esse conjunto é dado pela seguinte relação: α^{2^m-2}.

Passo 5: Determina-se a extensão do anel $\mathbb{Z}_4$, que é dada pelo quociente do conjunto de todos os polinômios com coeficientes em $\mathbb{Z}_4$ pelo ideal gerado pelo polinômio primitivo $p(x)$ encontrado no Passo 3. Como os coeficientes dos polinômios estão em $\mathbb{Z}_4$, realiza-se uma operação módulo 4 e obtém-se todos os elementos não nulos e invertíveis do grupo cíclico.

Passo 6: Determina-se o grupo das unidades, que é a construção do subgrupo cíclico, a partir do grupo cíclico gerado no Passo 5.

Passo 7: Determina-se o polinômio gerador da matriz G, $g(x)$. Para cada valor de t, que é a quantidade de erros, tem-se um polinômio gerador $g(x)$ diferente, e consequentemente um novo código. Então, são consideradas todas as possibilidades relacionadas à distância mínima $d_H \leq 2t + 1$, para $1 \leq t \leq \frac{(n-1)}{2}$. Para obter o polinômio gerador $g(x)$, primeiro calcula-se as raízes dos polinômios minimais, em seguida encontra-se os polinômios minimais $M_i(x)$ para $i = 1, 2, ..., 2t$ e, por último, calcula-se os polinômios geradores para $1 \leq t \leq \frac{(n-1)}{2}$. O polinômio gerador do código BCH de comprimento n tem como raízes os elementos $(\beta_i), (\beta_i)p, ..., (\beta_i)p^{r-1(\mod n)}$ e é dado por:

$$g(x) = mmc(M_1(x), M_2(x), ..., M_{2t}(x)).$$

O polinômio gerador $g(x)$ do código é relacionado à matriz geradora G do código BCH sobre o anel $\mathbb{Z}_4$, com parâmetros (n, k, d_H).

Passo 8: Determina-se o polinômio gerador da matriz verificação de paridade H, $h(x)$, obtido por meio do quociente de x^{n-1} pelo polinômio gerador obtido no Passo 7.

Passo 9: A matriz geradora G, relacionada ao polinômio gerador encontrado, será determinada. Sendo o polinômio $g(x) = g_0 + g_1x + g_2x^2 + ... + g_{r-1}x^{r-1} + g_rx^r$, tem-se a matriz geradora do código dada pelo deslocamento dos coeficientes da esquerda para direita, com dimensões $k \times n$.

Passo 10: Determina-se a matriz H e a sua transposta H^T. Dado o polinômio verificação de paridade $h(x) = h_0 + h_1x + ... + h_kx^k$, tem-se a matriz H dada pelo deslocamento dos coeficientes do polinômio gerador $h(x)$ da direita para a esquerda, com dimensões $(n-k) \times n$. A matriz H^T de dimensão $n \times (n-k)$ é obtida realizando a troca das linhas pelas colunas.

Passo 11: A sequência de DNA é rotulada utilizando o Passo 1. É analisado se o código BCH pode reproduzir a sequência, fazendo uma analogia entre o alfabeto 4-ário do código genético e o alfabeto 4-ário do código BCH para estrutura de anel, para realizar o mapeamento entre o conjunto $N = \{A, C, G, T/U\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Toda sequência de DNA será considerada como uma das 24 permutações entre $N \to \mathbb{Z}_4$ que podem ser divididas entre três grupos de oito permutações, chamados rotulamento A, rotulamento B e rotulamento C. As permutações do rotulamento A caracterizam um mapeamento em $\mathbb{Z}_4$ – linear e toda sequência de DNA reproduzida será não linear, pois o mapeamento $\mathbb{Z}_4$ – linear é não linear enquanto o código BCH sobre estrutura de anel é linear. Os mapeamentos $\mathbb{Z}_2 \times \mathbb{Z}_2$ e Klein – linear, relacionados aos rotulamentos B e C, respectivamente, são lineares e como o código BCH também é linear as sequências de DNA reproduzidas serão lineares. As permutações da sequência são dispostas nas 24 linhas de uma matriz P e podem ser descritas conforme a Tabela 6.

Tabela 6 – Relação entre as linhas da matriz P e as 24 permutações.

Fonte: [5]

Linha = Caso	$N \to \mathbb{Z}_4$	Linha = Caso	$N \to \mathbb{Z}_4$
L 1 = Caso 01	(A, C, G, T) = (0, 1, 2, 3)	L 13 = Caso 13	(A, C, G, T) = (2, 0, 1, 3)
L 2 = Caso 02	(A, C, G, T) = (0, 1, 3, 2)	L 14 = Caso 14	(A, C, G, T) = (2, 0, 3, 1)
L 3 = Caso 03	(A, C, G, T) = (0, 2, 1, 3)	L 15 = Caso 15	(A, C, G, T) = (2, 1, 0, 3)
L 4 = Caso 04	(A, C, G, T) = (0, 2, 3, 1)	L 16 = Caso 16	(A, C, G, T) = (2, 1, 3, 0)
L 5 = Caso 05	(A, C, G, T) = (0, 3, 2, 1)	L 17 = Caso 17	(A, C, G, T) = (2, 3, 0, 1)
L 6 = Caso 06	(A, C, G, T) = (0, 3, 1, 2)	L 18 = Caso 18	(A, C, G, T) = (2, 3, 1, 0)
L 7 = Caso 07	(A, C, G, T) = (1, 0, 2, 3)	L 19 = Caso 19	(A, C, G, T) = (3, 0, 1, 2)
L 8 = Caso 08	(A, C, G, T) = (1, 0, 3, 2)	L 20 = Caso 20	(A, C, G, T) = (3, 0, 2, 1)
L 9 = Caso 09	(A, C, G, T) = (1, 2, 0, 3)	L 21 = Caso 21	(A, C, G, T) = (3, 1, 0, 2)
L 10 = Caso 10	(A, C, G, T) = (1, 2, 3, 0)	L 22 = Caso 22	(A, C, G, T) = (3, 1, 2, 0)
L 11 = Caso 11	(A, C, G, T) = (1, 3, 0, 2)	L 23 = Caso 23	(A, C, G, T) = (3, 2, 0, 1)
L 12 = Caso 12	(A, C, G, T) = (1, 3, 2, 0)	L 24 = Caso 24	(A, C, G, T) = (3, 2, 1, 0)

Passo 12: Verifica-se se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(a, b) = 0$, $D(a, b) = 1$ e $D(a, b) = 2$.

Passo 13: São comparadas todas as palavras-código armazenadas no passo anterior com a sequência

de DNA original, mostrando onde os erros ocorreram. As palavras-código armazenadas no passo 12 em forma do alfabeto do código $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ serão convertidas em nucleotídeos usando o alfabeto do código genético $N = \{A, C, G, T/U\}$. Em seguida, as palavras-código são comparadas, uma a uma, com a sequência de DNA original, mostrando onde os nucleotídeos diferem.

Passo 14: Volta-se para o Passo 7 e determina-se outro $g(x)$, sendo determinada outra distância mínima e em seguida calcula-se o polinômio gerador relacionado a esta nova distância.

Passo 15: Repete-se os Passos 8 ao 12 para o polinômio gerador $g(x)$ obtido no passo 14, até que se esgote todas as possibilidades de $g(x)$. O algoritmo determina todas as palavras-código encontradas com 0, 1 e 2 nucleotídeos de diferença, com todos os polinômios geradores.

Passo 16: Volta-se para o Passo 3 para escolher outro $p(x)$, e então, repete-se os Passos 4 ao 14 até que se esgotem todos os $p(x)$ do Passo 3.

Passo 17: Chega-se ao fim após esgotarem todos os polinômios primitivos.

A Figura 10 apresenta um fluxograma com o detalhamento do algoritmo descrito anteriormente.

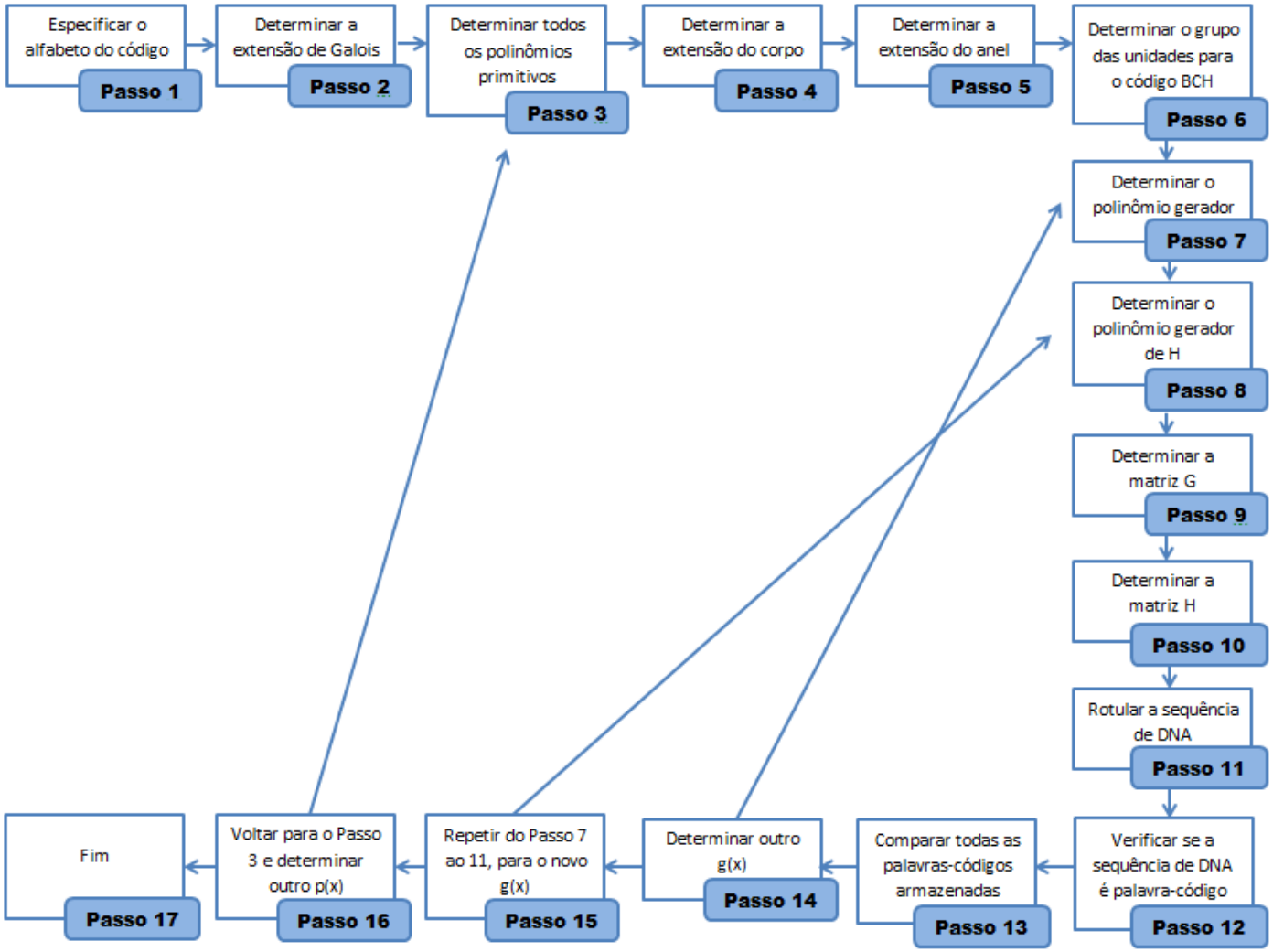

Figura 10 – Fluxograma do algoritmo de geração de proteínas.

Fonte: Próprio autor.

3 Resultados

Neste capítulo serão apresentados os resultados obtidos durante o estudo realizado neste trabalho de conclusão de curso. As referências utilizadas foram [1], [5],[24], [25], [26], [27], [28], [29], [30], [31], [32] e [33].

3.1 Análise da sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA

Nesta seção são apresentadas a reprodução e a análise da sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA, por meio do algoritmo de geração de proteínas, descrito na Seção 2.4 do Capítulo 2.

3.1.1 Geração da sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA

A sequência Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA foi escolhida para ser reproduzida pelo algoritmo de geração de proteínas por possuir comprimento $n = 63$ nucleotídeos, levando em consideração as restrições relacionadas ao comprimento para análise, que é expressa por $n = 2^r - 1$. Além disso, devido à mutação acarretar doenças genéticas, como veremos adiante.

Para identificar possiveis mutações, foi executado o algoritmo de geração de proteinas proposto por [2] e [6], com a finalidade de verificar se a sequência poderia ser identificada e reproduzida por meio dos códigos corretores de erros, em especial os códigos BCH.

Para essa execução do algoritmo foi necessário percorrer os 17 passos apresentados na Subseção 2.3.4. Considere então a construção do código BCH primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (63, k, d_H)$ como é apresentado a seguir.

Passo 1 - Especificar a estrutura matemática e o alfabeto do código

O alfabeto 4-ário do código genético ao conjunto formado por nucleotídeos, representado por $N = \{A, C, G, T/U\}$ que corresponde respectivmente à adenina, citosina, guanina e timina/uracila. Já o alfabeto 4-ário dos códigos corretores de erros, denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Todas as operações algébricas necessárias irão obedecer as operações de adição e multiplicação módulo 4, apresentadas na Tabela 2.

Passo 2 - Determinar a extensão de Galois

Para o algoritmo ser executado, o comprimento (n) da sequência de DNA deve obedecer a seguinte restrição:

$$n = 2^r - 1,$$

onde r é o grau da extensão do corpo de Galois. Como $n = 63$, temos que:

$$63 = 2^r - 1 \Rightarrow 2^r = 64 \Rightarrow r = 6.$$

Logo, o grau dos polinômios primitivos a ser usado na extensão de Galois do corpo $GF(2^6)$ é 6.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois

Nesta etapa temos o armazenamento de todos os polinômios primitivos $p\,(x)$ de grau igual a $r = 6$, os quais estão apresentados na Tabela 7.

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

Tabela 7 – Polinômios primitivos $p\,(x)$ de grau 6.

Passo 4 - Determinar a extensão do corpo $GF(2)$

Considere o corpo $GF(2^r) = GF(2^6) = GF(64)$ dado por:

$$\frac{F_2}{\langle p\,(x)\rangle} \cong \frac{F_2\,[x]}{\langle 1x^6 + 1x^4 + 1x^3 + 1x^1 + 1\rangle} = \left\{a_0 + a_1x + a_2x^2 + \cdots + a_5x^5 :\ a_i's \in F_2\right\},$$

Vamos construir um corpo dotado de 64 elementos, sendo formado a partir das classes residuais de polinômios (sobre GF (2)), módulo $1x^6 + 1x^4 + 1x^3 + 1x^1 + 1$, como realizado no Exemplo 11. A Tabela 8 apresenta as representações por potência e vetorial.

Passo 5 - Determinar a extensão do anel $\mathbb{Z}_4$

Consideremos agora o anel $\mathrm{GR}\left(p^k, r\right) = \mathrm{GR}\,(4, 6)$, dado por:

$$\frac{\mathbb{Z}_4\,[x]}{\langle p\,(x)\rangle} \cong \frac{\mathbb{Z}_4\,[x]}{\langle 1x^6 + 1x^4 + 1x^3 + 1x^1 + 1\rangle} = \left\{b_0 + b_1x + b_2x^2 + \cdots + b_5x^5 :\ b_i's \in \mathbb{Z}_4\right\},$$

Agora estamos aptos a construir a extensão do anel em $\mathrm{GR}\,(4, 6)$, sendo formado a partir das classes residuais de polinômios sobre $\mathrm{GR}\,(4, 6)$ módulo $p\,(x)$ (ou $1x^6 + 1x^4 + 1x^3 + 1x^1 + 1$ por ser o nosso exemplo em questão).

Seja β uma raiz de $p\,(x)$. Então, $\beta^6 + \beta^4 + \beta^3 + \beta^1 + 1 = 0$ levando a $\beta^6 = -\beta^4 - \beta^3 - \beta^1 - 1$. Porém, note que agora estamos trabalhando em $\mathbb{Z}_4$. Sendo assim, devemos adequar os coeficientes dos polinômios. Logo: $\beta^6 = 3\beta^4 + 3\beta^3 + 3\beta^1 + 3$.

Para calcular o valor de β_i, com $i > 6$, é necessário fazer uma composição dos índices anteriores, por exemplo: para se obter β_7 é necessário fazer a composição:

Tabela 8 – Elementos de $GF(64)$.

Representações					
Por Potência	**Vetorial**	**Por Potência**	**Vetorial**	**Por Potência**	**Vetorial**
0	(000000)	α^{21}	(111010)	α^{43}	(001101)
1	(100000)	α^{22}	(011101)	α^{44}	(101011)
α^1	(010000)	α^{23}	(100011)	α^{45}	(111000)
α^2	(001000)	α^{24}	(111100)	α^{46}	(011100)
α^3	(000100)	α^{25}	(011110)	α^{47}	(001110)
α^4	(000010)	α^{26}	(001111)	α^{48}	(000111)
α^5	(000001)	α^{27}	(101010)	α^{49}	(101110)
α^6	(101101)	α^{28}	(010101)	α^{50}	(010111)
α^7	(111011)	α^{29}	(100111)	α^{51}	(100110)
α^8	(110000)	α^{30}	(111110)	α^{52}	(010011)
α^9	(011000)	α^{31}	(011111)	α^{53}	(100100)
α^{10}	(001100)	α^{32}	(100010)	α^{54}	(010010)
α^{11}	(000110)	α^{33}	(010001)	α^{55}	(001001)
α^{12}	(000011)	α^{34}	(100101)	α^{56}	(101001)
α^{13}	(101100)	α^{35}	(111111)	α^{57}	(111001)
α^{14}	(010110)	α^{36}	(110010)	α^{58}	(110001)
α^{15}	(001011)	α^{37}	(011001)	α^{59}	(110101)
α^{16}	(101000)	α^{38}	(100001)	α^{60}	(110111)
α^{17}	(010100)	α^{39}	(111101)	α^{61}	(110110)
α^{18}	(001010)	α^{40}	(110011)	α^{62}	(011011)
α^{19}	(000101)	α^{41}	(110100)	α^{63}	(100000)
α^{20}	(101111)	α^{42}	(011010)		

$$\beta_7 = \beta \times \beta_6 = \beta \times \left(3\beta^4 + 3\beta^3 + 3\beta^1 + 3\right) = 3\beta^5 + 3\beta^4 + 3\beta^2 + 3\beta.$$

O que resultaria em $\beta_7 \rightarrow (033033)$ (considerando a leitura, da direita para a esquerda, dos coeficientes resultantes). Para os demais casos o raciocínio é análogo ao apresentado para os cálculos de α's.

$$\begin{aligned} \beta_8 &= \beta \times \beta_7 = \cdots \\ \beta_9 &= \beta \times \beta_8 = \cdots \\ \vdots &= \cdots \\ \beta_{126} &= \beta \times \beta_{125} = \cdots \end{aligned}$$

De maneira análoga ao cálculo da extensão do corpo, o processo de composição dos novos β's é apresentado nos trabalhos [2, 6, 1].

A Tabela 9 apresenta todos os elementos não nulos e inversíveis do grupo cíclico do grupo $\mathrm{GR}^*(4, 6)$:

Tabela 9 – Elementos do grupo cíclico do grupo $\mathrm{GR}^*(4,6)$ em notação de r-uplas.

Representações					
Por Potência	**Vetorial**	**Por Potência**	**Vetorial**	**Por Potência**	**Vetorial**
1	(100000)	β^1	(010000)	β^2	(001000)
β^3	(000100)	β^4	(000010)	β^5	(000001)
β^6	(303303)	β^7	(030330)	β^8	(302132)
β^9	(133312)	β^{10}	(011333)	β^{11}	(100032)
β^{12}	(010003)	β^{13}	(300303)	β^{14}	(030032)
β^{15}	(201003)	β^{16}	(321203)	β^{17}	(032322)
β^{18}	(201032)	β^{19}	(220301)	β^{20}	(321333)
β^{21}	(032131)	β^{22}	(003211)	β^{23}	(301020)
β^{24}	(333001)	β^{25}	(233100)	β^{26}	(023110)
β^{27}	(202111)	β^{28}	(020011)	β^{29}	(301102)
β^{30}	(333013)	β^{31}	(033103)	β^{32}	(100011)
β^{33}	(212201)	β^{34}	(320123)	β^{35}	(030012)
β^{36}	(102100)	β^{37}	(111311)	β^{38}	(312030)
β^{39}	(132100)	β^{40}	(112113)	β^{41}	(213011)
β^{42}	(023103)	β^{43}	(103211)	β^{44}	(111022)
β^{45}	(312201)	β^{46}	(231220)	β^{47}	(223320)
β^{48}	(222132)	β^{49}	(020013)	β^{50}	(202203)
β^{51}	(323321)	β^{52}	(032132)	β^{53}	(001213)
β^{54}	(000323)	β^{55}	(103133)	β^{56}	(113210)
β^{57}	(310220)	β^{58}	(332123)	β^{59}	(233212)
β^{60}	(122222)	β^{61}	(311123)	β^{62}	(233110)
β^{63}	(023113)	β^{64}	(000311)	β^{65}	(011132)
β^{66}	(103311)	β^{67}	(011030)	β^{68}	(233301)
β^{69}	(122031)	β^{70}	(212201)	β^{71}	(021220)
β^{72}	(013021)	β^{73}	(130201)	β^{74}	(332321)
β^{75}	(000331)	β^{76}	(133130)	β^{77}	(312012)
β^{78}	(231203)	β^{79}	(030023)	β^{80}	(121002)
β^{81}	(220300)	β^{82}	(310232)	β^{83}	(111223)
β^{84}	(201120)	β^{85}	(123213)	β^{86}	(223222)
β^{87}	(133023)	β^{88}	(121300)	β^{89}	(310330)
β^{90}	(021233)	β^{91}	(130123)	β^{92}	(100313)
β^{93}	(320033)	β^{94}	(332203)	β^{95}	(012323)
β^{96}	(031030)	β^{97}	(111103)	β^{98}	(333310)
β^{99}	(232232)	β^{100}	(222320)	β^{101}	(210232)
β^{102}	(122120)	β^{103}	(201113)	β^{104}	(322313)
β^{105}	(002031)	β^{106}	(301300)	β^{107}	(210330)
β^{108}	(232130)	β^{109}	(112310)	β^{110}	(303031)
β^{111}	(101002)	β^{112}	(102102)	β^{113}	(121111)
β^{114}	(131210)	β^{115}	(002220)	β^{116}	(320020)
β^{117}	(211101)	β^{118}	(002211)	β^{119}	(003322)
β^{120}	(223231)	β^{121}	(010121)	β^{122}	(113210)
β^{123}	(302020)	β^{124}	(311301)	β^{125}	(011330)
β^{126}	(203133)				

Passo 6 - Determinar o grupo das unidades

Nesta etapa do algoritmo será construído o subgrupo cíclico, sendo formado por elementos da extensão do anel e será baseado no parâmetro d. Como $d = 2$ teremos um subgrupo formado por 63 elementos com $f^2 \rightarrow (001000)$ sendo este considerado como o elemento primitivo que gera o subgrupo cíclico G_{63}. A Tabela 10 apresenta os elementos constituintes do subgrupo:

$GR^*(4,6)$	$(\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$	$GR^*(4,6)$	$(\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$	$GR^*(4,6)$	$(\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$
$(f^2)^1 = \beta^2$	(001000)	$(f^2)^2 = \beta^4$	(000010)	$(f^2)^3 = \beta^6$	(330330)
$(f^2)^4 = \beta^8$	(113013)	$(f^2)^5 = \beta^{10}$	(302031)	$(f^2)^6 = \beta^{12}$	(102123)
$(f^2)^7 = \beta^{14}$	(232212)	$(f^2)^8 = \beta^{16}$	(310230)	$(f^2)^9 = \beta^{18}$	(113212)
$(f^2)^{10} = \beta^{20}$	(313000)	$(f^2)^{11} = \beta^{22}$	(003130)	$(f^2)^{12} = \beta^{24}$	(110101)
$(f^2)^{13} = \beta^{26}$	(030130)	$(f^2)^{14} = \beta^{28}$	(110011)	$(f^2)^{15} = \beta^{30}$	(320023)
$(f^2)^{16} = \beta^{32}$	(230031)	$(f^2)^{17} = \beta^{34}$	(101003)	$(f^2)^{18} = \beta^{36}$	(012021)
$(f^2)^{19} = \beta^{38}$	(213333)	$(f^2)^{20} = \beta^{40}$	(123210)	$(f^2)^{21} = \beta^{42}$	(331122)
$(f^2)^{22} = \beta^{44}$	(201113)	$(f^2)^{23} = \beta^{46}$	(303312)	$(f^2)^{24} = \beta^{48}$	(311301)
$(f^2)^{25} = \beta^{50}$	(032102)	$(f^2)^{26} = \beta^{52}$	(022303)	$(f^2)^{27} = \beta^{54}$	(011230)
$(f^2)^{28} = \beta^{56}$	(110222)	$(f^2)^{29} = \beta^{58}$	(203300)	$(f^2)^{30} = \beta^{60}$	(002033)
$(f^2)^{31} = \beta^{62}$	(121101)	$(f^2)^{32} = \beta^{64}$	(030200)	$(f^2)^{33} = \beta^{66}$	(000302)
$(f^2)^{34} = \beta^{68}$	(022021)	$(f^2)^{35} = \beta^{70}$	(213033)	$(f^2)^{36} = \beta^{72}$	(123211)
$(f^2)^{37} = \beta^{74}$	(320111)	$(f^2)^{38} = \beta^{76}$	(322120)	$(f^2)^{39} = \beta^{78}$	(223001)
$(f^2)^{40} = \beta^{80}$	(031223)	$(f^2)^{41} = \beta^{82}$	(231103)	$(f^2)^{42} = \beta^{84}$	(013322)
$(f^2)^{43} = \beta^{86}$	(202331)	$(f^2)^{44} = \beta^{88}$	(101122)	$(f^2)^{45} = \beta^{90}$	(203213)
$(f^2)^{46} = \beta^{92}$	(303333)	$(f^2)^{47} = \beta^{94}$	(120110)	$(f^2)^{48} = \beta^{96}$	(331131)
$(f^2)^{49} = \beta^{98}$	(102010)	$(f^2)^{50} = \beta^{100}$	(331310)	$(f^2)^{51} = \beta^{102}$	(333203)
$(f^2)^{52} = \beta^{104}$	(010303)	$(f^2)^{53} = \beta^{106}$	(011110)	$(f^2)^{54} = \beta^{108}$	(330001)
$(f^2)^{55} = \beta^{110}$	(032333)	$(f^2)^{56} = \beta^{112}$	(121000)	$(f^2)^{57} = \beta^{114}$	(001210)
$(f^2)^{58} = \beta^{116}$	(330302)	$(f^2)^{59} = \beta^{118}$	(021321)	$(f^2)^{60} = \beta^{120}$	(213022)
$(f^2)^{61} = \beta^{122}$	(200332)	$(f^2)^{62} = \beta^{124}$	(130131)	$(f^2)^{63} = \beta^{126}$	(100000)

Tabela 10 – Elementos de G_{63}.

Passo 7 - Determinar o polinômio gerador da matriz G, $g(x)$

Neste passo será obtido o polinômio gerador $g\,(x)$. Para isso devemos determinar o polinômio

$(\phi(x))$, comum a todos os polinômios primitivos, utilizando o subgrupo cíclico G_{63}, específico de cada polinômio primitivo.

Para podermos resolver este polinômio devemos primeiro isolar todos os índices de β para um dado grau do polinômio.

$\phi(x) = x^6 + (3\beta^{32} + 3\beta^{16} + 3\beta^8 + 3\beta^4 + 3\beta^2 + 3\beta^1) \cdot x^5 + (1\beta^{48} + 1\beta^{40} + 1\beta^{24} + 1\beta^{36} + 1\beta^{20}$
$+1\beta^{12} + 1\beta^{34} + 1\beta^{18} + 1\beta^{10} + 1\beta^6 + 1\beta^{33} + 1\beta^{17} + 1\beta^9 + 1\beta^5 + 1\beta^3) \cdot x^4 + (3\beta^{56} + 3\beta^{52}$
$+3\beta^{44} + 3\beta^{28} + 3\beta^{50} + 3\beta^{42} + 3\beta^{26} + 3\beta^{38} + 3\beta^{22} + 3\beta^{14} + 3\beta^{49} + 3\beta^{41} + 3\beta^{25} + 3\beta^{37}$
$+3\beta^{21} + 3\beta^{13} + 3\beta^{35} + 3\beta^{19} + 3\beta^{11} + 3\beta^7) \cdot x^3 + (1\beta^{60} + 1\beta^{58} + 1\beta^{54} + 1\beta^{46} + 1\beta^{30} + 1\beta^{57}$
$+1\beta^{53} + 1\beta^{45} + 1\beta^{29} + 1\beta^{51} + 1\beta^{43} + 1\beta^{27} + 1\beta^{39} + 1\beta^{23} + 1\beta^{15}) \cdot x^2 + (3\beta^{62} + 3\beta^{61} + 3\beta^{59}$
$+3\beta^{55} + 3\beta^{47} + 3\beta^{31}) \cdot x^1 + 1\beta^{63}.$

Desta forma, o polinômio gerador é dado por:

$$g(x) = 1x^6 + 3x^5 + 1x^3 + 1x^2 + 2x^1 + 1.$$

Passo 8 - Determinar o polinômio gerador da matriz H, $h(x)$

Para a obtenção do polinômio $h(x)$, o cálculo é realizado por meio da seguinte relação:

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{63} - 1}{1x^6 + 3x^5 + 1x^3 + 1x^2 + 2x^1 + 1}$$

$h(x) = 1x^{57} + 1x^{56} + 1x^{55} + 2x^{53} + 2x^{52} + 2x^{51} + 1x^{50} + 3x^{47} + 1x^{43} + 3x^{42} + 3x^{40} + 3x^{39} + 2x^{38} +$
$3x^{36} + 1x^{34} + 3x^{33} + 2x^{32} + 3x^{31} + 1x^{29} + 1x^{28} + 3x^{27} + 2x^{26} + 1x^{25} + 3x^{24} + 3x^{23} + 1x^{22} + 2x^{21} + 1x^{19} +$
$1x^{18} + 2x^{17} + 3x^{14} + 2x^{13} + 1x^{12} + 3x^{10} + 2x^9 + 2x^8 + 3x^7 + 1x^6 + 3x^5 + 3x^4 + 1x^3 + 1x^2 + 2x + 3.$

Passo 9 - Determinar a matriz G

De acordo com o Passo 7, foi obtido o polinômio gerador $g(x) = 1x^6 + 3x^5 + 1x^3 + 1x^2 + 2x^1 + 1$. Assim pode-se encontrar a matriz G com dimensão $k \times n$, através dos deslocamentos dos coeficientes do polinômio $g(x)$ da esquerda para a direita. Como $n = 63$, k será dado pelo seguinte método: $k = n - r \Rightarrow k = 63 - 6 \Rightarrow k = 57$. Desta forma, a matriz G possui dimensão 57×63:

$$G = \begin{bmatrix}
1211031000000000000000000000000000000000000000000000000000000000 \\
0121103100000000000000000000000000000000000000000000000000000000 \\
0012110310000000000000000000000000000000000000000000000000000000 \\
0001211031000000000000000000000000000000000000000000000000000000 \\
0000121103100000000000000000000000000000000000000000000000000000 \\
0000012110310000000000000000000000000000000000000000000000000000 \\
\vdots \\
0000000000000000000000000000000000000000000000000121103100000 \\
0000000000000000000000000000000000000000000000000012110310000 \\
0000000000000000000000000000000000000000000000000001211031000 \\
0000000000000000000000000000000000000000000000000000121103100 \\
0000000000000000000000000000000000000000000000000000012110310 \\
0000000000000000000000000000000000000000000000000000001211031
\end{bmatrix}$$

Passo 10 - Determinar a matriz H e a sua transposta H^T

Neste passo será construída a matriz verificação de pariedade H por meio do polinômio $h(x)$.

Assim pode-se encontrar a matriz H com dimensão $r \times n$, através dos deslocamentos dos coeficientes do polinômio $h(x)$ da direita para a esquerda . Como $n = 63$ e $r = 6$,obtemos desse modo, a matriz H com dimensão 6×63:

$$H = \begin{bmatrix}
000003211331322301230021102133123110323103023303100030012220111 \\
000032113313223012300211021331231103231030233031000300122201110 \\
000321133132230123002110213312311032310302330310003001222011100 \\
003211331322301230021102133123110323103023303100030012220111000 \\
032113313223012300211021331231103231030233031000300122201110000 \\
321133132230123002110213312311032310302330310003001222011100000
\end{bmatrix}$$

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1

Nesse passo será verificado se o código BCH primitivo sobre anel é capaz de reproduzir a sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA. Essa sequência deve ser rotulada com as 24 permutações, de acordo com a Tabela 6, resultando na matriz P.

Seja a sequência do $NCBI$ igual a:

TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG

Desta forma, obtém-se a matriz P, dada por:

$$P = \begin{bmatrix}
3232110210213300130212220222110222030121023033322111022101112 2 \\
2323110310312200120313330333110333020131032022223311103310111 33 \\
3131220120123300230121110111220111030212013033331122201120222 11 \\
1313220320321100210323330333220333010232031011113322203320222 33 \\
1212330230231100310232220222330222010323021011112233302230333 22 \\
2121330130132200320131110111330111020313012022221133301130333 11 \\
3232001201203311031202221222001222131020123133332200012201000 22 \\
2323001301302211021303331333001333121030132122223300013301000 33 \\
3030221021023311231020001000221000131202103133330022210021222 00 \\
0303221321320011201323331333221333101232130100003322213321222 33 \\
2020331031032211321030001000331000121303102122220033310031333 00 \\
0202331231230011301232221222331222101323120100002233312231333 22 \\
3131002102103322032102111211100211123201021323333110002110200 11 \\
1313002302301122012303332333002333212030231211113300023302000 33 \\
3030112012013322132010002000112000232101203233330011120012111 00 \\
0303112312310022102313332333112333202131230200003311123312111 33 \\
1010332032031122312030002000332000212303201211110033320032333 00 \\
0101332132130022302131112111332111202313210200001133321132333 11 \\
2121003103102233023101113111003111323010312322221100031103000 11 \\
1212003203201133013202223222003222313020321311112200032203000 22 \\
2020113013012233123010003000113000323101302322220011130013111 00 \\
0202113213210033103212223222113222303121320300002211132213111 22 \\
1010223023021133213020003000223000313202301311110022230023222 00 \\
0101223123120033203121113111223111303212310300001122231123222 11
\end{bmatrix}$$

Passo 12 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(a,b) = 0$**,** $D(a,b) = 1$ **e** $D(a,b) = 2$

Neste passo é analisado se as sequências são palavras-código dos códigos (n, k, d_H) usando a equação $v \cdot H^T = 0$, da seguinte maneira:

- Para a análise de sequências de DNA com até 1 nucleotídeo de diferença da sequência de DNA original, consideramos as 4356 possiveis palavras-código para cada sequência de DNA analisada. Todas as palavras-código identificadas serão armazenadas.
- Já para a análise de sequências de DNA com até 2 nucleotídeos de diferença da sequência de DNA original, serão considerados todas as combinações 2 a 2 dos n nuclelotídeos de comprimento da sequência para as 24 permutações. Todas as palavras-código identificadas serão armazenadas.

Os demais passos seguirão a estrutura apresentada na Subseção 2.4.1.

Como resultado das simulações da sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA, para $D(a,b) = 1$, diferindo em um nucleotídeo da sequência do NCBI, foram obtidas 8 palavras código, listadas a seguir.

Palavra código: 3 de rotulamento B :: (A = 0, C = 1, G = 2, T = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 3232110210213300130212220222110222030121023033322111022101112 2
GLb: 3232110210213300130212220222110222030121023003332211102210111 22
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (A = 0, T = 1, G = 2, C = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 1212330230231100310232220222330222010323021011112233302230333 22
GLb: 1212330230231100310232220222330222010323021001112233302230333 22
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (C = 0, A = 1, T = 2, G = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 2323001301302211021303331333001333121030132122223300013301000 33
GLb: 2323001301302211021303331333001333121030132112223300013301000 33
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (T = 0, A = 1, C = 2, G = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 0303221321320011201323331333221333101232130100003322213321222 33
GLb: 0303221321320011201323331333221333101232130110003322213321222 33
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (G = 0, C = 1, A = 2, T = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 3030112012013322132010002000112000232101203233330011120012111 00
```

```
GLb: 3030112012013322132010002000112000232101203233300111200121110 0
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (G = 0, T = 1, A = 2, C = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 1010332032031122312030002000332000212303201211110033320032333 00
GLb: 1010332032031122312030002000332000212303201221110033320032333 00
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (C = 0, G = 1, T = 2, A = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 2121003103102233023101113111003111323010312322221100031103000 11
GLb: 2121003103102233023101113111003111323010312332221100031103000 11
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

Palavra código: 3 de rotulamento B :: (T = 0, G = 1, C = 2, A = 3)

```
Oaa: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..Y..F..G..P..G..T..R.
Ont: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTATTTTGGCCCAGGCACCCGG
OLb: 0101223123120033203121113111223111303212310300001122231123222 11
GLb: 0101223123120033203121113111223111303212310330001122231123222 11
Gnt: TGTGCCAGCAGCTTAACTAGCGGGAGGGCCAGGGATACGCAGTAATTTGGCCCAGGCACCCGG
Goo: .C..A..S..S..L..T..S..G..R..A..R..D..T..Q..*..F..G..P..G..T..R.
```

rotulamento A	0
rotulamento B	8
rotulamento C	0

Total de palavras-código: 8.

Com a análise de apenas uma diferença, observe que, na posição da trinca 15 houve uma troca do nucleotídeo timina (TAT) por um nucleotídeo adenina (TAA), ocasionando a troca de aminoácido nesta posição, sendo a tirosina (Tyr) substituída pelo códon de parada (stop).

3.1.2 Ánalise da mutação identificada na sequência de DNA

Biologicamente, a troca de único nucleotídeo, T→A , realiza uma transversão , que ocasiona a codificação de um aminoácido diferente, no caso a tirosina (TAT) é substituída por um códon de parada (TAA). Temos então uma mutação sem sentido na sequência de DNA, no qual a substituição do nucleotídeo resulta em um códon de parada (stop), que altera uma trinca de bases nitrogenadas e a substituição de um aminoácido por outro.

Tal mudança de acordo com [24], gera uma mutação no gene OTX2, que é necessário no anterior do cérebro e principalmente do olho, para a formação de fotorreceptores, que é responsável por converter imagens em impulsos elétricos [25]. Além disso, OTX2 está presente na expressão da rodopsina, composto químico encontrado na retina dos olhos, tendo a função de conversão das luzes nos impulsos elétricos que o cérebro é capaz de interpretar como visão [26].

Em [24], foi realizada uma pesquisa em uma família que possui essa mutação hereditária, gerando uma mutação na proteína Y179X. Esta mudança foi analisada através de exames clínicos, como ressonância magnética e testes neuropsicométricos de OTX2 em dois irmãos, representados por pacientes $4A$, $4B$ e da mãe dos pacientes, a paciente $4C$.

Esta alteração causa malformação ocular grave, pode afetar não só fenótipos oculares, acarretando principalmente a microftalmia, mas também os fenótipos neurológicos, variando de grave atraso de desenvolvimento cognitivo. Como pode ser observado na Tabela 11 a seguir, que apresenta as características clínicas identificadas nos paciente $4A$ e $4B$, sendo o foco do estudo.

Tabela 11 – Características clínicas.

Fonte: [26].

	$4A$	$4B$
Sexo	Feminino	Masculino
Idade	33 anos	25 anos
Fenótipo da mãe	Pigmentar retinopatia[1]	Pigmentar retinopatia
Visão	Nenhuma	6/60
Microftalmia[2]	Bilateral	Bilateral leve
Diâmetro da córnea (mm)	D=4 ; E=4	D=11 ; E=11
Anomalias oculares	Corectopia[3] , Coloboma[4]	Sinéquias[5]
Catarata[6]	Olho E, início tardio.	Não
Pigmentar retinopatia	Não	Aglomerados no centro do olho
Dificuldade de aprendizagem	Grave	Suave
Convulsões	A partir dos 14 anos de forma severa	Não
Outras características	Falas atrasadas	Nistagmo[7]

Nota: D= direita e E= esquerda; 1 doença localizada na retina dos olhos que causa perda de visão progressiva [27]. 2 anomalia ocular, que causa volume ocular reduzido[28]. 3 deslocamento da pupila [29]. 4 alteração na estrutura do olho, com ausência de parte da palpebra e íris [30]. 5 aderências entre a íris e a cápsula anterior do cristalino (elemento do ocular) [31]. 6 doença ocular que afeta o cristalino do olho, reduzindo a visão [32]. 7 movimentos involuntários dos olhos [33].

A Figura 11 apresenta os pacientes $4A$ e $4B$ mencionados anteriormente.

Figura 11 – Paciente $4A$ à esquerda e paciente $4B$ à direita.

Fonte:[26].

Como foi destacado anteriormente, esta mutação na sequência identificada nos irmãos foi herdada da mãe (paciente $4C$), que possui distrofia retiniana progressiva. No entanto, existem menos defeitos oculares estruturais do que seus filhos. Essa diferença ocular entre o filho $4B$ e sua mãe, pode ser identicada na Figura 12 a seguir.

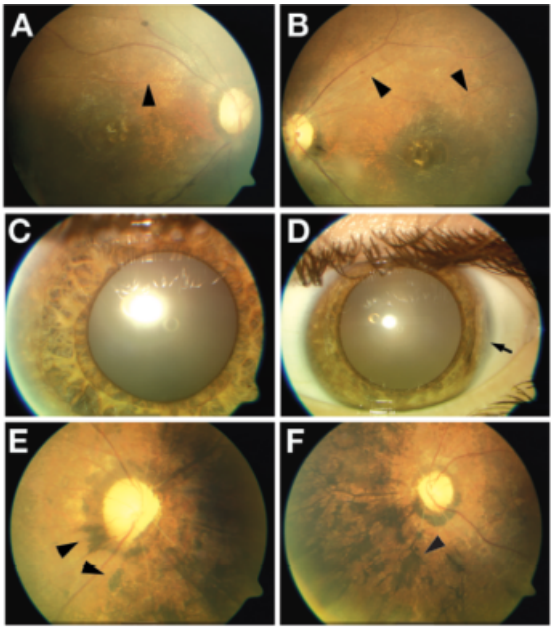

Figura 12 – Imagem ocular mais recente dos pacientes $4B$ e $4C$.

Fonte: [26].

Nota: Paciente $4B$ aos 28 anos: fundo direito (A) e fundo esquerdo (B), ambos com discos ópticos pálidos, máculas atróficas e vasos retinianos finos (pontas de seta); íris direita (C) e íris esquerda (D). A íris esquerda possui sinéquias anteriores periféricas (seta). Paciente $4C$ (mosaico mãe dos pacientes $4A$ e $4B$) aos 51 anos: fundo direito (E) e fundo esquerdo (F), mostrando retinopatia com aglomerados de pigmento acumulados (seta) [24].

4 Considerações Finais

O estudo realizado nesta monografia se baseia na utilização dos códigos BCH na reprodução de uma sequência de DNA via um algoritmo de geração de proteínas, para identificar possíveis mutações genéticas. Para isso, foi estudada a construção deste código, que ocorre por meio da estrutura algébrica de corpos finitos e que admitem representações em termos de polinômios sobre extensões de corpos de Galois. Além disso, foram analisados todos os passos da descrição do algoritmo e uma pesquisa mais detalhada sobre mutações, para melhor realização do trabalho.

Como resultado, foi possível reproduzir a sequência de DNA Homo sapiens T cell receptor beta chain (BV6S4-BJ2S3) mRNA através do algoritmo de geração de proteínas, onde foi identificada uma mutação do tipo sem sentido, que ocasiona a codificação de um aminoácido diferente, no caso o códon de parada (TAA) é codificada no lugar da tirosina (TAT). Tal mutação apresenta uma implicação biológica, neste caso relacionado com malformações oculares graves, que causa nos pacientes problemas oculares e neurológicos.

Essas aplicações foram importantes no processo de conectar a Matemática, Biologia e a Engenharia por meio do estudo da utilização de estruturas algébricas na construção de códigos BCH e suas aplicações nos estudos mutacionais, através do algoritmo. Este Trabalho de Conclusão de Curso surgiu como extensão de uma Iniciação Científica, desenvolvida de forma remota, devido à pandemia gerada pelo vírus SARS-CoV-2, de setembro de 2020 a agosto de 2021, por meio de reuniões e apresentações de seminários, realizados na plataforma Google Meet.

Durante a Iniciação Científica foram apresentados cinco trabalhos nos eventos: X Encontro Regional de Matemática Aplicada e Computacional do Rio Grande do Sul - X ERMAC RS, XL Congresso Nacional de Matemática Aplicada e Computacional (CNMAC 2021), II Workshop em Corpos Finitos e Aplicações, VII Simpósio Integrado UNIFAL-MG e Encontro Regional de Matemática Aplicada e Computacional - ERMAC RJ 2021. Além disso, foi submetido um resumo sobre a pesquisa realizada neste Trabalho de Conclusão de Curso para o evento 4º Encontro de Biomatemática, a ser realizado de 26 a 29 de abril 2022, para uma possível apresentação no formato de pôster.

Esta pesquisa contribuiu de maneira significativa na minha formação, proporcionando um crescimento pessoal e profissional, além da oportunidade de participar de eventos, que auxiliam na divulgação da pesquisa e no direcionamento de novos pesquisadores na área.

Referências

[1] PEREIRA, G. D. **Uma Abordagem Computacional para a Análise de Sequências de DNA por meio dos Códigos Corretores de Erros**. 2014. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP.

[2] 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 (Doutorado em Engenharia Eletrica). DT-FEEC, UNICAMP, 2010.

[3] OLIVEIRA, A. J. **Análise Algébrica dos Rotulamentos Associados ao Mapeamento do Código Genético**. Dissertação (Mestrado em Engenharia Elétrica) Mestrado, FEEC-UNICAMP, 2012.

[4] MARCET, A. P. **Códigos Corretores de Erros BCH: Uma Aplicação de Polinômios em Corpos Finitos**. 2019. 77 f. Dissertação (Mestrado Profissional em Matemática em Rede Nacional - PROFMAT) - Faculdade de Formação de Professores, Universidade do Estado do Rio de Janeiro, RJ.

[5] OLIVEIRA, A. N. **Códigos BCH Aplicados no Processo de Análise de Fenômenos Mutacionais**. 2020. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Programa de Pós-Graduação em Estatística Aplicada e Biometria, Alfenas, MG.

[6] FARIA. L. C. B. **Existências de códigos corretores de erros e protocolos de comunicação em sequências de DNA**. Tese (Doutorado em Engenharia Elêtrica). DT-FEEC, UNICAMP, 2011.

[7] CARVALHO, H. F.; RECCO-PIMENTEL, S.M. A célula. Manole Ltda, 2001.

[8] NELSON, D. L.; COX, M. M.; LEHNINGER, A. Princípios de Bioquímica. Sarvier, 2002.

[9] MEC; SEPS; FUNBEC. Série "Ciências para o 1º grau"- CORPO HUMANO: Funções e Nutrição. São Paulo: Hamburg: Centro de Estudos de Ciências de São Paulo,1994.

[10] LAURENCE, J. Biologia: ensino médio, volume único. 1.ed. São Paulo: Nova Geração, 2005.

[11] ZAHA, A.; FERREIRA, H. B.; PASSAGLIA, L. M. P.. Biologia molecular básica. Artmed Editora, 2014.

[12] NICHOLAS,F.W. Introdução à genética veterinária. 3.ed. Porto Alegre: Artmed, 2012.

[13] BRASIL ESCOLA. Mutações gênicas.
Disponível em: <https://www.youtube.com/watch?$v = 0d_2J1b7G9g$ > Acesso em: 19 de outubro de 2021.

[14] BRASIL ESCOLA. Mutações cromossômicas.
Disponível em: <https://www.youtube.com/watch?$v = Kb_uiUCAW4o$ > Acesso em: 19 de outubro de 2021.

[15] DOMINGUES, H. H. Algebra moderna. 4: ed. ref. São Paulo: Atual, 2003.

[16] FRALEIGH, J.B. A first course in abstract algebra. Addison Wesley, 2003.

[17] LIN, S.; COSTELLO, D. J. Jr. Error Control Coding. 2ª edição. Prentice Hall, 2004.

[18] ROCHA, B. P.; OLIVEIRA, A. J. Aplicações de Estruturas Algébricas na Conexão entre Sistemas de Comunicação Padrão e Genética. Sigmae, Alfenas, v.6, n.2, p. 1-14. 2014.

[19] ANDRADE, A. A.; SHAH, T; QAMAR, A. Uma construção de códigos BCH. Revista Eletrônica Paulista de Matemática, v. 2, n. 1, p. 1, 2013.

[20] ANDRADE, A. A. **Uma contribuição à construção e decodificação de códigos de bloco lineares sobre anéis finitos**. 1996. 116f. Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP.

[21] DUARTE GONZALEZ, M. E. **Modelagem da síntese de proteínas e sua estrutura organizacional através de códigos corretores de erros**. 2017. 1 recurso online (174 p.). Tese (Doutorado em Engenharia Elétrica) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP.

[22] SOUZA, T. A. **Algebra de Corpos Finitos Aplicada à Teoria da Codificação: Estudo do Codificador BCH**. 2012. Trabalho de Conclusão de Curso (Bacharelado em Matemática) - Universidade Federal da Paraíba, João Pessoa, PB.

[23] TRANCOSO, M. C. R. **Sistematização da Codificação e Descodificação de Códigos BCH**. 1995. Dissertação (Mestrado em Engenharia Electrotécnica) - Faculdade de Engenharia da Universidade do Porto, Porto, Portugal.

[24] RAGEE, N. K. et al. Heterozygous Mutations of OTX2 Cause Severe Ocular Malformations. **American Society of Human Genetics**, p. 1008-1022, 2005.

[25] MANUAL MSD. Estrutura e função dos olhos.
Disponível em: <https://www.msdmanuals.com/pt-br/casa/dist Acesso em: 21 de novembro de 2021.

[26] FOCUS MEDICINA DOS OLHOS. Como funciona a visão. Disponível em: <http://focusmedicinadosolhos.com.br/index.php/artigos/como-funciona-a-visao/> Acesso em: 21 de novembro de 2021.

[27] SAÚDE E BEM ESTAR. Retinite Pigmentosa, Retinose Pigmentar. Disponível em: https://pt.overleaf.com/project/612fc91a3e4aa52f230d538f <https://www.saudebemestar.pt/pt/clinica/oftalmologia/retinite-pigmentosa/> Acesso em: 21 de novembro de 2021.

[28] IDECO OFTALMOLOGISTA EM SÃO CARLOS. Microftalmia: uma anomalia ocular congênita Disponível em: <http://www.ideco.med.br/blog/2020/09/22/microftalmia-uma-anomalia-ocular-congenita/> Acesso em: 21 de novembro de 2021.

[29] A REVISTA DA OFTALMOLOGIA UNIVERSO VISUAL. O que é síndrome de Axenfeld-Rieger e quais os critérios diagnósticos?. Disponível em: <https://universovisual.com.br/secaodesktop/noticias/274/o-que-e-sindrome-de-axenfeld-rieger-e-quais-os-criterios-diagnosticos> Acesso em: 21 de novembro de 2021.

[30] + TUA SAÚDE. Coloboma: o que é, tipos, sintomas e tratamento. Disponível em: <https://www.tuasaude.com/coloboma-ocular/> Acesso em: 21 de novembro de 2021.

[31] CLÍNICA DE OLHOS. Uveítes.
Disponível em: <https://danielparente.com.br/doenca/19/uveites.html> Acesso em: 21 de novembro de 2021.

[32] + TUA SAÚDE. Catarata: o que é, sintomas, tipos causas e tratamento. Disponível em: <https://www.tuasaude.com/catarata/> Acesso em: 21 de novembro de 2021.

[33] REDE D'OR SÃO LUIZ. Nistagmo.
Disponível em: <https://www.rededorsaoluiz.com.br/doencas/nistagmo> Acesso em: 21 de novembro de 2021.