UNIVERSIDADE FEDERAL DE ALFENAS

Débora Barbosa Souza

OS CÓDIGOS CORRETORES DE ERROS NO ESTUDO DE MUTAÇÕES GENÉTICAS

ALFENAS-MG

2024

DÉBORA BARBOSA SOUZA

OS CÓDIGOS CORRETORES DE ERROS NO ESTUDO DE MUTAÇÕES GENÉTICAS

Trabalho de Conclusão de Curso submetido à Universidade Federal de Alfenas, como requisito necessário para obtenção do grau de Licenciada em Matemática.
Orientadora: Profª. Dra. Cátia Regina de Oliveira Quilles Queiroz.

ALFENAS-MG
2024

UNIVERSIDADE FEDERAL DE ALFENAS

DÉBORA BARBOSA SOUZA

Esta Monografia foi julgada adequada para a obtenção do título de Licenciada em Matemática, sendo aprovada em sua forma final pela banca examinadora:

Documento assinado digitalmente
gov.br CATIA REGINA DE OLIVEIRA QUILLES QUEIROZ
Data: 10/07/2024 21:06:12-0300
Verifique em https://validar.iti.gov.br

Orientadora: Profª. Dra. Cátia Regina de Oliveira Quilles Queiroz
Universidade Federal de Alfenas - UNIFAL-MG

Documento assinado digitalmente
gov.br ANDERSON JOSE DE OLIVEIRA
Data: 11/07/2024 11:00:25-0300
Verifique em https://validar.iti.gov.br

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

Documento assinado digitalmente
gov.br EVANDRO MONTEIRO
Data: 11/07/2024 12:00:19-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Evandro Monteiro
Universidade Federal de Alfenas - UNIFAL-MG

Documento assinado digitalmente
gov.br MARCELO MOREIRA DA SILVA
Data: 12/07/2024 09:52:32-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Marcelo Moreira da Silva
Universidade Federal de Alfenas - UNIFAL-MG

ALFENAS-MG
2024

Agradecimentos

Agradeço ao meu marido, que foi meu pilar de força e superação, sua paciência, amor, apoio e compreensão foram fundamentais para perseverar nos momentos mais difíceis.

Aos meus pais, Isa e Paulo César, por todo suporte, amor e palavras de encorajamento, me incentivando e apoiando em todas as etapas da minha vida.

À minha orientadora Cátia Regina Oliveira Quilles Queiroz por toda dedicação, conhecimento, paciência, compreensão e amizade no desenvolvimento deste trabalho.

Aos meus amigos por todos os momentos vividos e aos novos amigos que o curso me presenteou, obrigada pelo companheirismo, dias de estudos, compartilhamento de ideias, risadas e ajuda mútua.

Aos professores do Departamento de Matemática que tanto contribuíram para minha formação.

"O futuro pertence àqueles que acreditam na beleza de seus sonhos."
(Eleanor Roosevelt)

Resumo

Os códigos corretores de erros visam identificar e, muitas vezes, corrigir interferências que podem ocorrer durante a transmissão de uma mensagem através de um canal de comunicação. Há códigos corretores de erros associados às sequências de DNA, que podem ser identificados como palavras código. Os códigos de bloco lineares BCH são o foco deste estudo, que auxiliarão na localização de mutações em uma molécula de DNA. Como a Álgebra fornece ferramentas fundamentais para o desenvolvimento dos algoritmos de codificação e decodificação, foi realizado um estudo teórico de conceitos algébricos, bem como biológicos, mostrando como ocorrem alterações na sequência dos nucleotídeos do material genético, durante a divisão celular. O Algoritmo de Codificação, já proposto na literatura, é detalhado passo a passo, com o auxílio do *Magma Computational Algebra System* na realização das operações matemáticas mais complexas, reproduzindo uma sequência de DNA com 63 nucleotídeos, escolhida aleatoriamente de um banco de dados conhecido como NCBI. Com este estudo espera-se contribuir para a associação entre Biologia, Matemática e Engenharia, além de auxiliar no desenvolvimento de uma metodologia para o diagnóstico de doenças e análises de mutações.

Palavras-chave: Álgebra. Códigos Cíclicos. Estrutura do DNA. Código BCH.

Abstract

Error correcting codes aim to identify and often correct interferences that may occur during the transmission of a message through a communication channel. There are error correcting codes associated with DNA sequences, which can be identified as code words. Linear block codes like BCH are the focus of this study, which will assist in locating mutations in a DNA molecule. Since Algebra provides fundamental tools for the development of encoding and decoding algorithms, a theoretical study of algebraic concepts, as well as biological ones, was conducted, showing how changes occur in the sequence of nucleotides of genetic material during cell division. The Encoding Algorithm, already proposed in the literature, is detailed step by step, with the support of the *Magma Computational Algebra System* in carrying out complex mathematical operations, reproducing a DNA sequence with 63 nucleotides, randomly chosen from a database known as NCBI. This study aims to contribute to the association between Biology, Mathematics, and Engineering, as well as to assist in the development of a methodology for disease diagnosis and mutation analysis.

Keywords: Algebra. Cyclic Codes. DNA Structure. BCH Code.

Sumário

1 Introdução

A utilização de sistemas de comunicação digital em várias esferas tem impulsionado o avanço de teorias matemáticas, que servem de suporte às novas tecnologias digitais, as quais integram a teoria da informação. A partir da dissertação de mestrado do matemático Claude E. Shannon, intitulada *A Symbolic Analysis of Relay and Switching Circuits* (1937), se deu início ao que mais tarde seria considerada a teoria da informação, possibilitando a criação de modelos de códigos capazes de detectar e/ou corrigir erros em um sistema de comunicação [1]. Nessa mesma década, a teoria de códigos corretores de erros teve seu início, impulsionada pelos estudos de Shannon, Golay e Hamming. Em 1948, Shannon demonstrou que, por meio de uma codificação adequada da informação, era possível controlar os erros introduzidos por canais ruidosos, mantendo a taxa de transmissão intacta, ficando conhecido como o "pai da teoria da informação" [2], [3].

Desta forma, pesquisadores passaram a buscar famílias de códigos eficientes e conjuntos de sinais associados a esses códigos, além de delinear decodificadores eficazes para os mesmos. Surge então a teoria da codificação, a partir da busca por códigos capazes de detectar e, se possível, corrigir erros que possam surgir durante os processos de transmissão e armazenamento de informações em um sistema de comunicação digital. Esses códigos são fundamentais em várias situações cotidianas, como na transmissão de informações digitalizadas, sinais de televisão e navegação na internet.

Nos sistemas de comunicação e armazenamento, existem dois tipos de códigos corretores de erros amplamente empregados, denominados códigos de bloco e códigos convolucionais. Os códigos de bloco são subdivididos em lineares e não lineares. Os códigos cíclicos constituem uma subclasse dos código de bloco lineares, destacando-se pela sua eficácia na detecção de erros, sendo os códigos BCH (Bose, Chaudhuri e Hocquenghem) exemplos desses códigos de vital relevância, os quais representam o foco deste trabalho.

É comum a utilização de modelos matemáticos para explicar fenômenos naturais, enquanto a Biologia tende a adotar abordagens mais experimentais para compreendê-los. A melhor compreensão da vida é uma busca constante realizada por pesquisadores de diversos países, que procuram vincular o genoma de seres vivos com estruturas matemáticas. Todavia, um grupo de pesquisadores da Universidade Estadual de Campinas (Unicamp) e da Universidade de São Paulo (USP) encontraram uma relação matemática entre um código numérico e a sequência do DNA (ácido desoxirribonucleico), ou seja, um código matemático que transcreve a sequência de DNA [4], [5], [6].

Assim, ao perceberem que as bases nitrogenadas se organizam segundo uma lógica matemática, mutações podem ser detectadas, através de um algoritmo para geração de

sequências de DNA, que identifica e reproduz tais sequências, com o uso da codificação.

A Álgebra desempenha um papel fundamental no estudo do processo de transmissão da informação, fornecendo ferramentas para o desenvolvimento de algoritmos de codificação e decodificação eficazes, garantindo uma transmissão confiável da informação e auxiliando na detecção e correção de erros. O algoritmo para a construção e geração de códigos BCH primitivos e não primitivos sobre o anel $\mathbb{Z}_4$, usando extensões de Galois, resulta na geração de palavras-código, que são identificadas como sequências de DNA com uma diferença em nucleotídeo (fita simples de DNA).

Com o avanço das pesquisas, espera-se que moléculas de DNA possam ser alteradas, restabelecendo a funcionalidade normal, como a volta da produção de insulina pelo pâncreas, além de auxiliar na produção de novos fármacos, melhoramento genético e biotecnologias de maneira geral.

O *Magma Computational Algebra System* é um pacote de *software* projetado para resolver problemas computacionalmente difíceis em álgebra, teoria dos números, geometria e combinatória. O *Magma* é distribuído pelo Grupo de Álgebra Computacional da Universidade de Sydney, sendo disponibilizado on-line gratuitamente no endereço: http://magma.maths.usyd.edu.au/calc/ [7]. O *software* foi utilizado nas operações que demandam maior tempo, como o cálculo dos polinômios minimais e no Passo 12 do algoritmo, para encontrar a matriz R, que será detalhada no Capítulo 3.

Deste modo, os objetivos deste trabalho se resumem em compreender conceitos básicos sobre genética, estudar estruturas algébricas de grupos, anéis, corpos e extensões de corpos, compreender os conceitos principais associados aos códigos corretores de erros, estudar as características principais dos códigos BCH, aplicar códigos BCH no estudo de mutações genéticas e entender os passos do algoritmo para identificar sequências de DNA.

Este trabalho está organizado da seguinte forma: no Capítulo 2, na primeira seção, são revisados conceitos de Biologia, desde as células, divisão celular, produção de proteínas e mutações, bem como conceitos teóricos fundamentais referentes à Álgebra na segunda seção. Na terceira seção são introduzidos os conceitos de códigos corretores de erros, especialmente os códigos de bloco lineares BCH, apresentando a construção do código $C(6, 3)$. No Capítulo 3 o Algoritmo de Codificação é detalhado passo a passo, visando localizar mutações na molécula de DNA, contando com o auxílio do *Magma Computational Algebra System* nas operações matemáticas mais complexas. Finaliza-se com as conclusões no Capítulo 4, seguidas das Referências.

2 *Revisão de Conceitos*

Neste capítulo serão apresentados os principais conceitos teóricos fundamentais utilizados neste trabalho, referentes à Biologia, à Álgebra e aos Códigos Corretores de Erros.

2.1 Conceitos Biológicos

Nesta seção serão apresentados os principais elementos de Biologia utilizados, sendo as principais referências [4], [6], [8], [9], [10], [11].

2.1.1 Células

A célula é uma unidade fundamental de diversos tamanhos e formas, presente em todos os organismos vivos, sendo classificadas em procarióticas, as quais são delimitadas por uma membrana plasmática, não possuindo um núcleo definido (material genético é disperso no citoplasma) e eucarióticas, que estão presentes nos membros animais e vegetais, cujo núcleo é definido e as organelas ficam dispersas no citoplasma, conforme Figura 1.

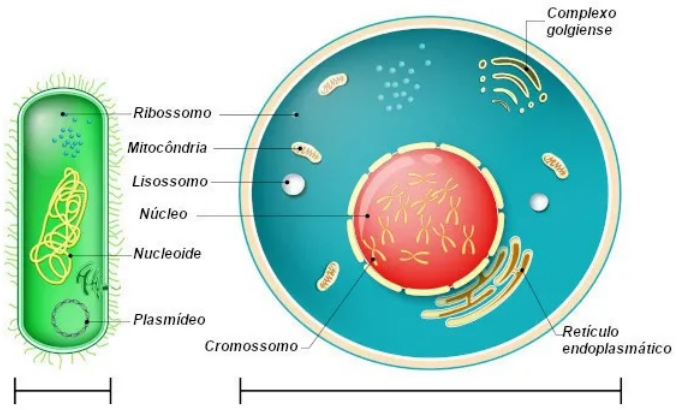

Figura 1 – Células eucarióticas e procarióticas.

Fonte: [22]

Moléculas pequenas transportam energia, transmitem sinais e se ligam para criar macromoléculas. Para obterem a energia necessária para a síntese de ATP (trifosfato de adenosina), as células quebram as moléculas dos alimentos e, ao quebrarem as ligações de ATP, liberam energia química, que pode ser empregada na maioria dos processos celulares, como na contração muscular.

As proteínas estruturam as células e promovem grande parte das tarefas celulares, sendo formadas a partir de, aproximadamente, 100 a 1000 aminoácidos, que são compostos por 20 tipos diferentes. As células mudam suas formas e se locomovem, permitindo a defesa contra infecções, transportando nutrientes, reparando e cicatrizando lesões. Comunicam-se através do envio de sinais, que são interpretados, gerando uma resposta, que vai desde a diferenciação em um tecido em particular até a sua morte que, por sua vez, pode ocorrer em resposta a agressões ou programação interna (apoptose), evitando toxicidade pela liberação de constituintes celulares.

2.1.2 Ciclo Celular

O ciclo celular, como pode ser visto na Figura 2, é composto por quatro estágios, e é caracterizado por uma série de eventos que preparam a célula para a divisão chamada de mitose, que é um processo assexual. As células possuem capacidade de reprodução, sendo que os cromossomos e o DNA (ácido desoxirribonucleico) que carregam são copiados durante a fase de síntese (S). As células eucarióticas, diferente das bactérias, necessitam de um tempo maior para crescer e se dividir, ao passo que esse controle permite um equilíbrio do crescimento de tecidos, possibilitando correções como substituição de células danificadas.

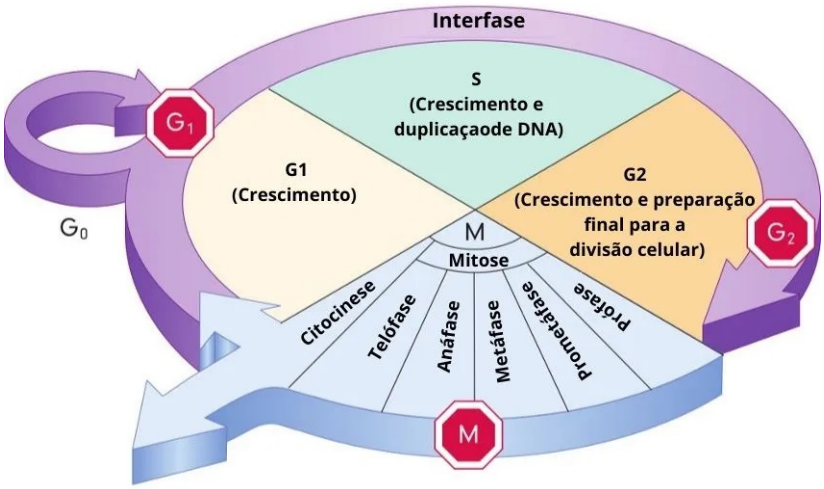

Figura 2 – Ciclo celular.

Fonte: [23]

Na reprodução sexuada (Figura 3) ocorre a produção da terceira célula, a partir da fusão de duas, contendo informações genéticas derivadas das parentais. Esse processo ocorre através da meiose, reduzindo o número de cromossomos na preparação, para então se fundirem posteriormente.

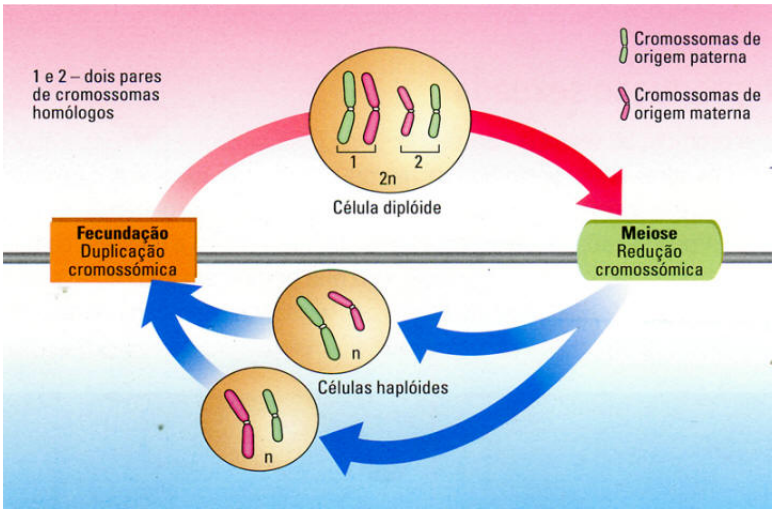

Figura 3 – Reprodução sexuada.

Fonte: [24]

2.1.3 Proteínas

As proteínas são repetições de aminoácidos, podendo ter quatro tipos de estruturas: primária, secundária, terciária e quaternária (vide Figura 4). A função é derivada da estrutura tridimensional, e a estrutura tridimensional é especificada pela sequência de aminoácidos.

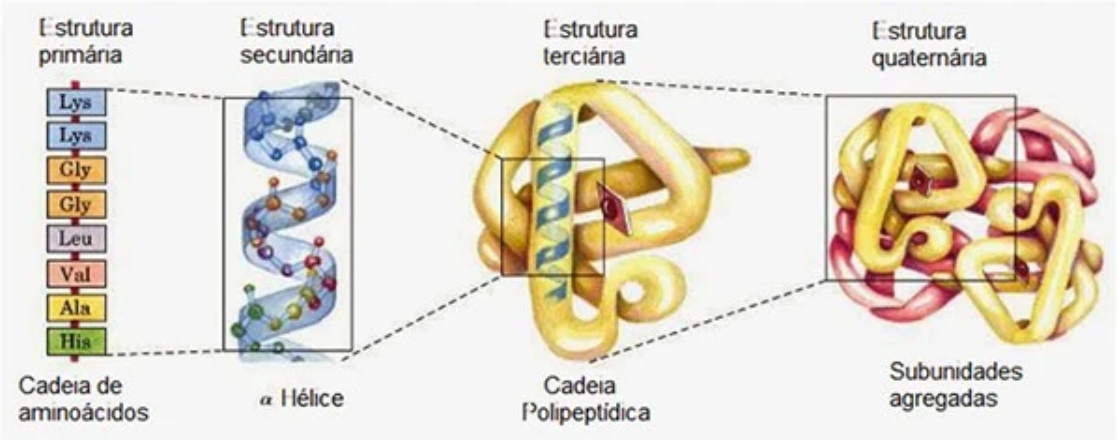

Figura 4 – Estrutura das proteínas.

Fonte: [25]

As proteínas possuem função estrutural, hormonal, energética, enzimática e de defesa, além de serem condutoras de gases, sendo formadas no processo chamado de tradução, onde ocorre a síntese de uma cadeia polipeptídica, através da decodificação do RNA (ácido ribonucleico) mensageiro e aquelas proteínas dobradas incorretamente são eliminadas por um sistema de verificação de erros das células.

Os ácidos nucleicos são macromoléculas constituídas por nucleotídeos, que formam o DNA e o RNA - componentes celulares que contêm as informações para a produção de proteínas, uma vez que determinam a sequência de aminoácidos e, consequentemente, a estrutura e função das proteínas de uma célula. No material genético estão os dados referentes a como, quando e onde deve ser produzido cada tipo de proteína, sendo que o DNA e o RNA (Figura 5) são as principais moléculas carregadoras de informações das células.

O DNA é um ácido desoxirribonucleico, cuja estrutura foi elucidada pela primeira vez em 1953 por James Watson e Francis Crick, sendo uma molécula helicoidal, formada por bases púricas (adenina e guanina) e pirimídicas (citosina e timina) que assumem distintas formas tautoméricas ("tautômeros são isômeros de conversão fácil, diferindo entre si apenas nas posições do hidrogênio"). Cada base é ligada a uma outra da base oposta, através de pontes de hidrogênio, sendo que as pirimídicas se ligam com as púricas, ou seja, ocorre um pareamento das bases complementares (adenina se liga com timina e guanina com citosina). Dessa forma, qualquer fita de DNA pode agir como molde para a síntese de sua fita complementar. Portanto, "a informação hereditária está codificada na sequência de bases em qualquer fita".

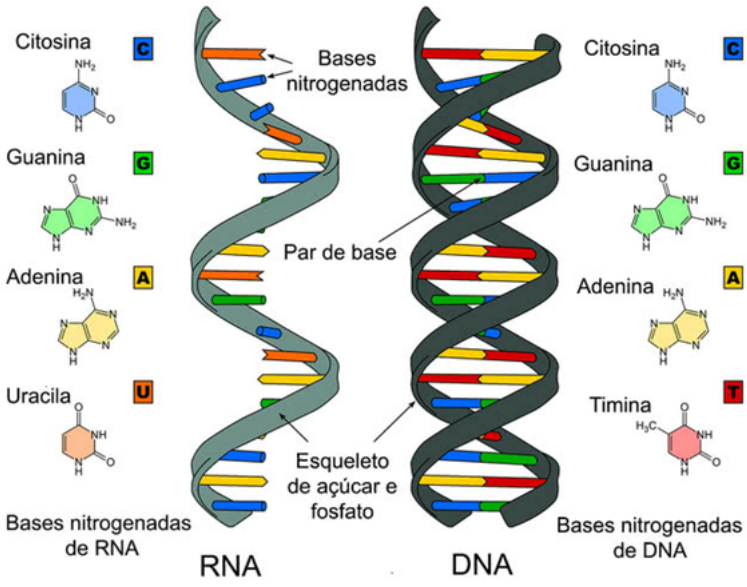

Figura 5 – DNA e RNA.

Fonte: [26]

Transcrição

As células fazem uso de dois processos em série para converter a informação codificada no DNA em proteínas. No primeiro, denominado de transcrição (Figura 6), a região codificante de um gene é copiada sob a forma de uma versão em fita simples de RNA

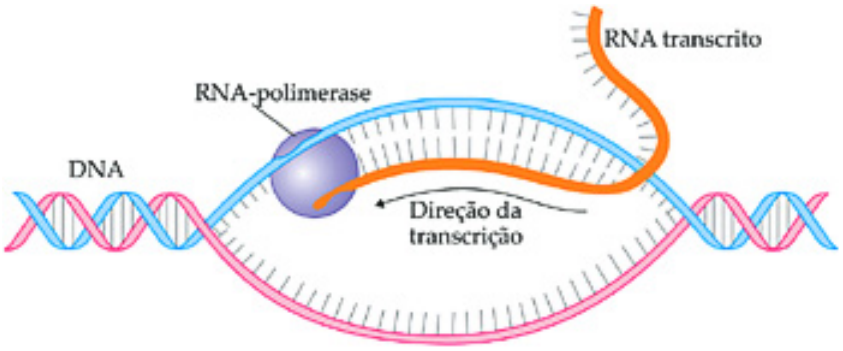

Figura 6 – Transcrição.

Fonte: [27]

a partir da dupla fita de DNA. Uma grande enzima, a RNA polimerase, catalisa a ligação dos nucleotídeos na cadeia de RNA, usando o DNA como molde. Em células eucarióticas, o produto inicial de RNA é processado em uma molécula de RNA mensageiro (mRNA) menor, a qual é transportada para o citoplasma. Neste compartimento, o ribossomo, uma enorme máquina molecular complexa, composta de RNAs de proteínas, se encarrega de efetuar o segundo processo, denominado tradução.

Tradução

Durante a tradução, vista na Figura 7, o ribossomo organiza e liga os aminoácidos seguindo uma ordem estabelecida, a qual é ditada pela sequência do mRNA, de acordo com um código genético praticamente universal. A partir da sequência do mRNA, a cada três bases da fita única, um aminoácido é codificado. Cada trinca é denominada códon e existem 64 códons que correspondem a 20 aminoácidos, apresentados na Tabela 1.

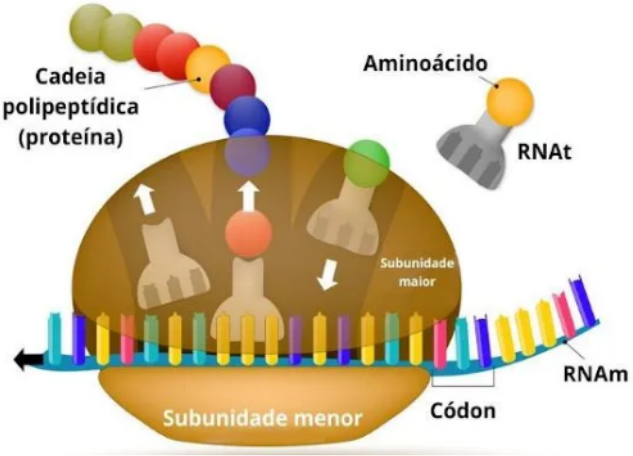

Figura 7 – Tradução.

Fonte: [28]

Tabela 1 – Lista de Aminoácidos.

Nome	**Símbolo**	**Códons**
Glicina	Gly, Gli	GGU, GGC, GGA e GGG
Alanina	Ala	GCU, GCC, GCA e GCG
Leucina	Leu	UUA, UUG, CUU, CUC, CUA e CUG
Valina	Val	GUU, GUC, GUA e GUG
Isoleucina	Ile	AUU, AUC e AUA
Prolina	Pro	CCU, CCC, CCA e CCG
Fenilalanina	Phe, Fen	UUU e UUC
Serina	Ser	UCU, UCC, UCA, UCG, AGU e AGC
Treonina	The,Thr	ACU, ACC, ACA e ACG
Cisteina	Cis, Cys	UGU e UGC
Tirosina	Tyr, Tir	UAU e UAC
Asparagina	Asn	AAU e AAC
Glutamina	Gln	CAA e CAG
Ácido aspártico	Asp	AAG e GAU
Ácido glutâmico	Glu	GAA e GAG
Arginina	Arg	CGU, CGC, CGA, CGG, AGA e AGG
Lisina	Lys, Lis	AAA e AAG
Histidina	His	CAU e CAC
Triptofano	Trp, Tri	UGG
Metionina	Met	AUG
Parada (stop)		UAA, UAG e UGA

Portanto, o código dito degenerado é aquele com mais de um códon correspondendo ao mesmo aminoácido. Os códons de iniciação (AUG - Metionina) e de finalização (UAA, UAG e UGA – stop) indicam que a sequência de aminoácidos da proteína deve ser inicializada e finalizada, respectivamente.

Cerca de 95% do DNA cromossomal humano é não codificante, ou seja, não codificam mRNAs ou qualquer outro RNA necessário ao organismo. Esse DNA contém regiões semelhantes, porém não idênticas, uma vez que são capazes de identificar uma pessoa pelas variações dessas sequências. As mutações podem ser provocadas quando elementos de DNA "móvel" se deslocam para outros locais no genoma.

A arquitetura celular é mantida com o empacotamento do DNA. O material genético apresenta-se na forma de uma nucleoproteína complexa (cromatina) dispersa no núcleo, quando as células não estão em divisão (intérfase). O dobramento e a compactação adicional da cromatina, durante a mitose, produz os cromossomos em metáfase. A espécie humana possui 46 cromossomos, sendo 22 pares de autossomos e 1 par de sexuais. Na mitose, após a replicação do DNA, os cromossomos apresentam-se com duas cromátides irmãs unidas pelo centrômero; sendo que o número, o tamanho e a forma dos cromossomos estabelecem o cariótipo (distinto em cada espécie). Nas extremidades dos cromossomos, há sequências especiais de DNA, os telômeros (Figura 8), cuja função principal é manter a estabilidade

estrutural dos cromossomos, e a cada divisão, os telômeros são encurtados, até que de tão curtos não permitem mais a replicação dos cromossomos, fazendo com que as células percam a capacidade de divisão (estudos sobre os telômeros podem auxiliar no tratamento de várias doenças).

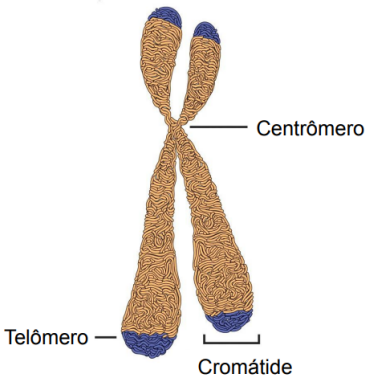

Figura 8 – Telômero.

Fonte: [4]

Dessa forma, o genoma de um organismo, que é seu conteúdo específico de DNA, pode estar dividido em múltiplos cromossomos, cada um com uma molécula de DNA, contendo as informações genéticas presentes em qualquer fita. Diante disso, as moléculas são extremamente grandes, sendo descritas em termos do número de pares de bases (pb) por milhares de pares de bases (quilobases em pares ou kb) e quanto mais complexo o organismo, mais DNA contém (raras exceções).

2.1.4 Mutações

As mutações ocorrem devido a alterações na sequência dos nucleotídeos do material genético, durante a divisão celular, podendo ser causadas por exposições a radiação ultravioleta ou ionizante, mutagênicos químicos ou vírus. Elas provocam variações no conjunto de genes populacional, sendo benéficas ou não. No caso das favoráveis, elas podem se acumular levando a mudanças evolutivas adaptativas, já as desfavoráveis são reduzidas por seleção natural e mecanismos de reparo de DNA. Nesse caso, ao causar erros na sequência de proteínas, elas tornam-se parciais ou não funcionais e, se possuírem papel importante, podem resultar em uma doença (doença genética). As neutras não influenciam na aptidão dos indivíduos. As mutações podem ainda ser germinativas, quando são transmitidas aos descendentes ou somáticas, caso contrário, e a "mutação de novo" é aquela que não foi herdada de nenhum dos pais.

A maioria dos organismos dispõe de mecanismos reparadores de DNA, capazes de corrigir grande parte das mutações, antes que consigam provocar mudanças permanentes. Há muitas formas de alterar a sequência de um gene, sendo que as mutações genéticas

podem afetar a saúde de várias maneiras, dependendo do local de ocorrência ou função proteica alterada.

As mutações podem ser classificadas de diversas maneiras, sendo alguns exemplos as de pequena escala, como a mutação de ponto (Figura 9), que geralmente é causada por mutagênicos ou erros na replicação do DNA, em que há a troca de um único nucleotídeo por outro; a transição, que é a mais comum, onde purina é trocada por purina ou pirimidina por pirimidina; e o caso contrário, que é chamado de transversão.

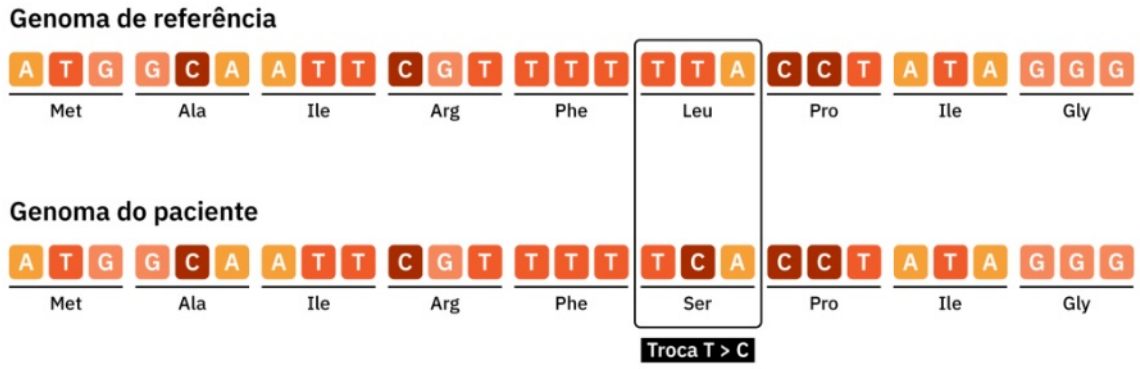

Figura 9 – Mutação de ponto.

Fonte: [29]

A mutação pontual pode ser revertida por outra mutação pontual e pode ser classificada em três tipos, dependendo do tipo de expressão apresentada pelo códon mutado:

1. Mutação silenciosa - o códon codifica para o mesmo aminoácido, não gerando mudanças na função proteica;
2. Mutação de sentido trocado - codifica para um aminoácido diferente, mudando a função proteica e, consequentemente, resultando em mudanças fenotípicas observáveis;
3. Mutação sem sentido - é a inserção de um códon de parada, interrompendo a proteína antes de seu término.

A mutação de inserção, também de pequena escala, decorre da adição de um ou mais nucleotídeos na sequência de DNA, comumente causada por transposons (componentes genéticos que se movimentam em uma sequência de DNA por transposição) ou erros durante a replicação de elementos repetitivos, modificando o quadro de leitura do gene.

Já na deleção, ocorre a remoção de um ou mais nucleotídeos na sequência de DNA, de maneira aleatória, modificando o quadro de leitura do gene. Normalmente,

são irreversíveis, uma vez que a reversão por elementos de transposição é improvável de acontecer.

As mutações de grande escala são classificadas em quatro tipos:

1. Amplificação - ocorre a criação de várias cópias de uma região cromossômica, aumentando a dosagem dos genes dentro dela;
2. Deleção - ocorre a perda de genes;
3. Inserção - tem-se a união de partes do DNA, incluindo translocação cromossômica (troca de porções de cadeias de DNA entre cromossomos não homólogos), deleção de interstício (deleção de um segmento de DNA de um cromossomo, agrupando, assim, genes anteriormente distantes) e inversão cromossômica (inversão da orientação de um segmento do cromossomo);
4. Perda de heterozigozidade - ocorre a perda de um alelo por deleção ou recombinação num organismo que originalmente possuía dois alelos.

Proteínas parciais ou não funcionais podem ser geradas a partir das mutações maléficas, levando a quadros de doenças genéticas ou até mesmo câncer, no caso de células malignas. Se uma célula germinal possuir uma mutação, descendentes poderão ser portadores da mesma mutação (doenças hereditárias). Na maioria dos casos, as mutações gênicas são reparadas pelo próprio organismo, através de um sistema celular de reparação de DNA, protegendo de doenças.

As mutações benéficas provocam mudanças, que permitem aos organismos se adaptarem melhor ao ambiente ao seu redor, levando a mudanças evolutivas e adaptativas.

2.2 Álgebra

Nesta seção, faremos uma revisão de estruturas algébricas, como grupos, anéis e extensão de corpos. As referências utilizadas foram [12], [13], [14] e [15].

2.2.1 Grupos

Definição 1. Um conjunto não vazio G sobre o qual uma operação binária $*$ é definida, é chamado de *grupo* se as seguintes condições são satisfeitas:

1. $(a * b) * c = a * (b * c), \forall\ a, b, c \in G$; (associatividade)
2. Existe $e \in G$ tal que $a * e = e * a = a$, $\forall\ a \in G$; (elemento neutro)

3. Para cada $a \in G$, existe $a' \in G$ tal que $a * a' = a' * a = e$. (todo elemento é simetrizável)

Se, além disso, sua operação binária é comutativa para todo $a, b \in G$, um grupo G é chamado de *grupo comutativo* ou *abeliano*, ou seja, $a * b = b * a, \forall\ a, b \in G$.

Definição 2. Um grupo $(G,*)$ é um *grupo finito* se o conjunto G é finito. Assim, a *ordem do grupo* é o número de elementos de G, sendo denotado por $o(G)$ ou $|G|$.

Propriedades de grupos

1. O elemento neutro de $(G,*)$ é único;

   Demonstração: Suponha que existam dois elementos neutros e e e', então

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

   □

2. O elemento simétrico de um grupo é único;

   Demonstração: Suponha que existam dois elementos inversos a' e a''. Então

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

   □

3. $(a * b)' = b' * a', \forall\ a, b \in G$;

   Demonstração: Sejam $a, b \in G$.

   $$\begin{aligned}
   (a * b)' * (a * b) &= e \\
   (a * b)' * (a * b) * b' &= e * b' \\
   (a * b)' * a * (b * b') &= b' \\
   (a * b)' * a * e &= b' \\
   (a * b)' * a &= b' \\
   (a * b)' * a * a' &= b' * a' \\
   (a * b)' * e &= b' * a' \\
   (a * b)' &= b' * a'
   \end{aligned}$$

   □

4. $(a')' = a, \forall\ a \in G$;

   Demonstração: Sejam $a \in G$, a' o elemento simétrico de a e $e \in G$ o elemento neutro de G.

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

   □

   Dizemos que $a \in G$ é regular, se $a * x = a * y \Rightarrow x = y$, para todo $x, y \in G$.

5. Todo $a \in G$ é regular;

   Demonstração: Suponhamos $a * x = a * y, \forall\ a, x, y \in G$
   Como todo elemento de G é simetrizável, então existe $a' \in G$ tal que

   $$\begin{aligned} (a' * a) * x &= (a' * a) * y \\ e * x &= e * y \\ x &= y. \end{aligned}$$

   De modo análogo, $x * a = y * a \Rightarrow x = y$ □

Alguns grupos importantes

- $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{Q}^*, \cdot)$, $(\mathbb{R}^*, \cdot)$, $(\mathbb{C}^*, \cdot)$ são grupos abelianos.

- Grupos aditivos de classes de restos.

  $(\mathbb{Z}_m, +)$ é um grupo aditivo, chamado de grupo aditivo de classe de restos módulo m. Seja $m > 1, m \in \mathbb{N}$ e $\mathbb{Z}_m = \{\overline{0}, \overline{1}, \overline{2}, ..., \overline{m-1}\}$, a soma de dois elementos de $\mathbb{Z}_m$ é definida por

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

  tal que r é o resto da divisão de $a + b$ por m, sendo chamado de adição módulo m.

- Grupos multiplicativos de classes de restos.

  $(\mathbb{Z}_m^*, .)$ é um grupo multiplicativo, chamado de grupo multiplicativo de classe de restos módulo m se, e somente se, m é um número primo. Seja $m > 1, m \in \mathbb{N}$ e $\mathbb{Z}_m^* = \{\overline{0}, \overline{1}, \overline{2}, ..., \overline{m-1}\}$, a multiplicação de dois elementos de $\mathbb{Z}_m^*$ é definida por

  $$\overline{a}.\overline{b} = \overline{a.b} = r,$$

  tal que r é o resto da divisão de $a.b$ por m, sendo chamado de multiplicação módulo m.

Exemplo 1. $(\mathbb{Z}, .)$ não é grupo, pois nem todos os elementos têm simétrico. Note que $4 \in \mathbb{Z}$, mas $\frac{1}{4} \notin \mathbb{Z}$ tal que $4 \cdot \frac{1}{4} = 1$.

Subgrupos

Definição 3. Seja H um subconjunto não vazio de G. O subconjunto H é um *subgrupo* de G se H é fechado sob a operação do grupo G e satisfaz todas as condições de um grupo. Se $(G, *)$ é um grupo, então $H = \{e\}$ e $H = G$ são subgrupos de G, chamados de subgrupos triviais. Os demais, caso existam, são não triviais.

Exemplo 2. $(\mathbb{Z}, +)$ é um subgrupo de $(\mathbb{Q}, +)$

Teorema 2.2.1. Sejam G um grupo sob a operação binária $*$ e H um subconjunto não vazio de G. Então H é um subgrupo de G se as seguintes condições são válidas:

(i) H é fechado sob a operação binária $*$;

(ii) Para qualquer elemento $a \in H$, o simétrico de a também está em H ou $\forall a, b \in H, a * b' \in H$.

Demonstração:

Da condição (i), H é fechado sob a operação $*$;

A condição (ii) diz que todo elemento de H tem um simétrico em H;

As condições (i) e (ii) garantem que o elemento identidade de G está em H.

Como os elementos de H estão em G, a condição associativa sob $*$ vale automaticamente.

Daí, H satisfaz todas as condições de um grupo, logo é um subgrupo de G.

□

Grupos Cíclicos

Definição 4. Sejam $a \in G$ e $m \in \mathbb{Z}$, tal que G é um grupo multiplicativo, a *potência m-ésima de a* é denotada por a^m e definida da seguinte maneira:

- Se $m \geq 0$,

$$\begin{aligned} a^0 &= e \qquad \text{(elemento neutro de } G) \\ a^m &= a^{m-1} \cdot a, \qquad \text{se } m \geq 1; \end{aligned}$$

- Se $m < 0$,

$$a^m = (a^{-m})^{-1}.$$

Propriedades
Sejam G um grupo multiplicativo e $a \in G, m, n \in \mathbb{Z}$:

1. $a^m \cdot a^n = a^{m+n}$;
2. $(a^m)^n = a^{m \cdot n}$;
3. $a^{-m} = (a^m)^{-1} = (a^{-1})^m$.

Definição 5. Seja $a \in G$, tal que G é um grupo aditivo, define-se *múltiplo* de a, para todo inteiro m, da seguinte maneira:

- Se $m \geq 0$,

$$\begin{aligned} 0 \cdot a &= e \qquad \text{(elemento neutro de } G) \\ m \cdot a &= (m-1) \cdot a + a, \qquad \text{se } m \geq 1; \end{aligned}$$

- Se $m < 0$,

$$m \cdot a = (-m) \cdot (-a).$$

Propriedades
Sejam G um grupo aditivo e $a \in G, m, n \in \mathbb{Z}$:

1. $m \cdot a + n \cdot a = (m+n) \cdot a$;
2. $m \cdot (n \cdot a) = (m \cdot n) \cdot a$;
3. $(-m) \cdot a = m \cdot (-a) = -(m \cdot a)$.

Se G é um grupo multiplicativo cíclico gerado por a, então $G = \{a^m | \, m \in \mathbb{Z}\} = [a]$. Se G é um grupo aditivo cíclico gerado por a, então $G = \{m \cdot a | \, m \in \mathbb{Z}\} = [a]_+$.

Definição 6. Um grupo G é chamado *grupo cíclico* se $G = [a]$ para algum $a \in G$. O elemento a é chamado gerador do grupo G.

Exemplo 3. O grupo multiplicativo $G = \{1, -1\}$ é cíclico, pois $G = \{(-1)^m | \, m \in \mathbb{Z}\} = [-1]$.

Teorema 2.2.2. Dados $(G, \cdot)$, se $a \in G$, então o subconjunto $[a]$ é um subgrupo de G, chamado *subgrupo cíclico* gerado por a.

Demonstração: Como $e = a^0$, então $e \in [a]$ e $[a] \neq \emptyset$. Agora, dados elementos distintos $x, y \in [a]$, então existem $m, n \in \mathbb{Z}$ tais que $x = a^m$ e $y = a^n$. Logo,

$$x.y' = x.y^{(-1)} = a^m.a^{(-n)} = a^{m-n},$$

assim $xy' \in [a]$.

Portanto, $[a]$ é subgrupo de G, pelo Teorema 2.2.1. □

Dessa forma, o teorema 2.2.2 diz que todo elemento a de um grupo G é gerador de um subgrupo cíclico. Tal subgrupo será indicado por: $[a] = \{a^m | \, m \in \mathbb{Z}\}$.

Definição 7. Dado um elemento a de um grupo G, a *ordem* de a é o menor inteiro $h > 0$, tal que $a^h = e$, denotada por $o(a)$ ou $|a|$.

Definição 8. Dado um elemento a de um grupo G, se $a^m = e \Leftrightarrow m = 0$, o elemento a tem ordem zero e o grupo gerado por a é um *grupo cíclico infinito.*

Definição 9. Seja $G = [a]$ um grupo cíclico. G é um *grupo cíclico finito* se a ordem do elemento a for um número natural $h > 0$. Neste caso, $G = \{e, a, a^2, ..., a^{h-1}\}$.

Exemplo 4. O elemento $i \in \mathbb{C}^*$ tem ordem 4, pois $i^1 = i, i^2 = -1, i^3 = -i$ e $i^4 = 1$.

Exemplo 5. No grupo multiplicativo do conjunto dos números reais, o elemento 2 tem ordem zero.

Teorema 2.2.3. Seja a um elemento de ordem $h > 0$ de um grupo G. Então $a^m = e$ se, e somente se, $h \mid m$.

Demonstração: ($\Rightarrow$) Suponhamos que $a^m = e$, então pelo algoritmo da divisão, existem inteiros q e r tais que

$$m = hq + r \quad (0 \leq r < h).$$

Assim,

$$\begin{aligned} e &= a^m \\ &= a^{hq+r} \\ &= (a^h)^q a^r \\ &= e^q a^r \\ &= e a^r \\ &= a^r. \end{aligned}$$

Ou seja, $a^r = e$. Como $r < h$ e a ordem de a é h, segue que $r = 0$. Portanto, $m = hq$, ou seja, $h \mid m$.

($\Leftarrow$) Se $h \mid m$, então $m = hq$, para algum $q \in \mathbb{Z}$. Então,

$$a^m = a^{hq} = (a^h)^q = e^q = e.$$

Logo, $a^m = e$. □

Classes Laterais

Definição 10. Seja H um subgrupo do grupo $(G, *)$. Dado $a \in G$, $a*H = \{a*x/x \in H\}$ é chamada *classe lateral à esquerda* de H, definida por a. Analogamente, $H*a = \{x*a/x \in H\}$ é chamada *classe lateral à direita* de H, definida por a. Se G é um grupo comutativo, então $a*H = H*a, \forall a \in G$.

Exemplo 6. Dado o grupo $G = \{1, i, -1, -i\}$ e seu subgrupo $H = \{1, -1\}$. Temos que:

$$1 \cdot H = \{1, -1\} = H \cdot 1$$

$$(-1) \cdot H = \{-1, 1\} = H \cdot (-1)$$

$$i \cdot H = \{i, -i\} = H \cdot i$$

$$(-i) \cdot H = \{-i, i\} = H \cdot (-i).$$

Totalizando duas classes laterais.

Teorema 2.2.4. A união de todas as classes laterais à esquerda (ou à direita) *módulo H* é igual a G.

Demonstração: Seja e o elemento neutro de G, então $e \in H$. Logo, todo elemento $a \in G$ pertence à classe aH, pois $a = a \cdot e$. Agora, se cada elemento de G está em uma classe lateral à esquerda, *módulo H*, então a união de todas elas é igual a G. □

Teorema 2.2.5. $aH = bH \Leftrightarrow b^{-1}a \in H, \forall a, b \in G$.

Demonstração: ($\Rightarrow$) Suponhamos que $aH = bH, \forall a, b \in G$

$$a \;=\; a \cdot e \in aH = bH \Rightarrow a \in bH \Rightarrow a = bh_1, h_1 \in H \Rightarrow b^{-1}a = b^{-1}bh_1 = h_1 \in H.$$

Portanto, $b^{-1}a \in H$.
($\Leftarrow$) Suponhamos que $b^{-1}a \in H$, então existe $h_1 \in H$ tal que $b^{-1}a = h_1$, logo $a = bh_1$.

Seja $y \in aH$, então $y = ah_2$, para algum $h_2 \in H$. Logo, para $h_1, h_2 \in H$, $y = (bh_1)h_2 \Rightarrow y = b(h_1h_2) \Rightarrow y \in bH$.

Portanto, $aH \subset bH$.

Seja $y \in bH$, logo $y = bh_2$, para algum $h_2 \in H$. Daí, para $h_1, h_2 \in H$, $y = ah_1^{-1}h_2 \Rightarrow y = a(h_1^{-1}h_2) \Rightarrow y \in aH$.

Assim, $bH \subset aH$.

Portanto, $aH = bH$. □

Teorema 2.2.6. Sejam aH e bH duas classes laterais *módulo* H. Então, $aH \cap bH = \emptyset$ ou $aH = bH$.

Demonstração: Suponhamos que exista $x \in aH \cap bH$. Então existem $h_1, h_2 \in H$ tal que $x = ah_1 = bh_2$. Disso segue que $b^{-1}a = h_2h_1^{-1} \in H$.
Pelo Teorema 2.2.5, temos que $aH = bH$. □

Definição 11. Seja G um grupo finito. O número de classes laterais *módulo* H em G é chamado de *índice de* H em G e denotado por $(G : H)$.

Teorema 2.2.7. *(Teorema de Lagrange)* Seja H um subgrupo de um grupo finito G. Então $o(H)|o(G)$ e $o(G) = o(H) \cdot (G : H)$.

Demonstração: Suponhamos que $(G : H) = r$ e seja $G/H = \{a_1H, a_2H, ..., a_rH\}$ o conjunto de todas as classes laterais à esquerda. Sabemos que $G = a_1H \cup a_2H \cup ... \cup a_rH$. Como cada elemento de G pertence a apenas uma dessas classes laterais (Teoremas 2.2.5 e 2.2.6) e o número de elementos de cada uma dessas classes é o mesmo de H, ou seja, é dado por $o(H)$, temos que

$$a_1H \cup a_2H \cup ... \cup a_rH = G \Rightarrow o(H) + o(H) + ... + o(H) = o(G)$$

$$\Rightarrow r \cdot o(H) = o(G) \Rightarrow o(H)|o(G)$$

e

$$o(G) = r \cdot o(H) = (G : H) \cdot o(H).$$

□

2.2.2 Anéis

Definição 12. Um conjunto não vazio A, munido de duas operações (soma e produto) é um *anel*, se satisfaz as seguintes propriedades:

1. $(A, +)$ é grupo abeliano;
2. $(A, \cdot)$ é associativa;
3. A multiplicação é distributiva em relação à adição, ou seja, $\forall a, b \in A, a(b+c) = ab+ac$ e $(a + b)c = ac + bc$.

Nestas condições, o anel é denotado por $(A, +, \cdot)$. Se além disso,

4. $ab = ba, \forall a, b \in A$, dizemos que A é um *anel comutativo* (ou *abeliano*);

5. $\exists\ 1 \in A$ tal que $1 \cdot a = a \cdot 1 = a, \forall a \in A$, dizemos que A é um *anel com identidade* (ou *unidade*).

Exemplo 7. $(\mathbb{Z}, +, \cdot), (\mathbb{Q}, +, \cdot), (\mathbb{R}, +, \cdot), (\mathbb{C}, +, \cdot)$ são anéis abelianos.

Definição 13. Seja $(A, +, \cdot)$ um anel. Dizemos que um subconjunto não vazio $L \in A$, é um *subanel* de A se:

(i) L é fechado para ambas as operações de A, isto é, $\forall a, b \in L, a + b \in L$ e $ab \in L$;

(ii) $(L, +, \cdot)$ é anel.

Teorema 2.2.8. Sejam $(A, +, \cdot)$ um anel e $L \subset A$ um subconjunto não vazio. Então L é um subanel de A se, e somente se, para todo $a, b \in L$, $a - b \in L$ e $a \cdot b \in L$, ou seja, L é fechado para a subtração e multiplicação de A.

Demonstração. ($\Rightarrow$) Suponhamos L é subanel não vazio de A. Sejam $a, b \in L$, por definição de subanel, $a \cdot b \in L$, então resta mostrar que $a - b \in L$.
Como L é anel, então $(L, +)$ é grupo, logo $\forall\ b \in L$, temos que $-b \in L$. Daí, $a + (-b) \in L$. Logo $a - b \in L$. Portanto, $\forall a, b \in L, a \cdot b \in L$ e $a - b \in L$.

($\Leftarrow$) Suponhamos que $\forall a, b \in L$, $a - b \in L$ e $a \cdot b \in L$. Vamos mostrar que L é subanel.

(i) $(L, +)$ é grupo abeliano
De fato:

- Tomando $a = b$, obtemos $a - a = 0 \in L$ (elemento neutro);
- Tomando $a = 0$, obtemos $a - b = -b \in L$ (simétrico) ;
- Dados $a, -b \in L$, então $a - (-b) = a + b \in L$ (fechado para adição);
- $\forall\ a, b, c \in L$. Se $a, b, c \in A$, então $a + (b + c) = (a + b) + c$ (associativa);
- $\forall\ a, b \in L$. Se $a, b \in A$, então $a + b = b + a$ (comutativa).

(ii) $(L, \cdot)$ é associativa.
$\forall\ a, b, c \in L$. Se $a, b, c \in A$, então $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

(iii) $\forall\ a, b, c \in L$, $a, b, c \in A$, logo $a(b + c) = ab + ac$ e $(a + b)c = ac + bc$. Logo, a multiplicação é distributiva em relação à adição em L.
Portanto, L é subanel de A.

□

Definição 14. Sejam $(A, +, \cdot)$ um anel e $L \subset A$ um subanel. Se o elemento neutro de A é igual ao elemento neutro de L para a multiplicação, então o subanel L é chamado de *subanel unitário*.

Exemplo 8. Os anéis $\mathbb{Z}_m$ das classes de resto, módulo m, também são anéis comutativos com unidade, pois o resto da divisão de ab por m é igual ao resto da divisão de ba por m, e a unidade é a classe $\overline{1}$, pois $\overline{a} \cdot \overline{1} = \overline{a} = \overline{1} \cdot \overline{a}$, com $\overline{a}, \overline{b} \in \mathbb{Z}_m$.

Anéis de integridade

Definição 15. Um anel A comutativo com unidade é denominado *anel de integridade*, se satisfaz: $\forall\ a, b \in A$, $a \cdot b = 0 \Rightarrow a = 0$ ou $b = 0$ (lei do anulamento do produto).

Se a e b são elementos não nulos de um anel A tais que $a \cdot b = 0$ ou $b \cdot a = 0$, a e b são chamados de *divisores próprios do zero em A*.

Exemplo 9. Os anéis $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ e $\mathbb{C}$ são anéis de integridade.

Exemplo 10. No anel $\mathbb{Z}_6$ os elementos $\overline{2}$ e $\overline{3}$ são divisores próprios do zero, pois são não nulos e, contudo, $\overline{2} \cdot \overline{3} = 0$. Logo, $\mathbb{Z}_6$ não é *anel de integridade*.

Teorema 2.2.9. Um anel de classes de restos $\mathbb{Z}_m$ é um anel de integridade se, e somente se, m é um número primo.

Demonstração. ($\Rightarrow$) Suponhamos que $\mathbb{Z}_m$ é anel de integridade e vamos mostrar que m é primo. Vamos supor por absurdo que m não é primo, então existem $a, b \in \mathbb{Z}$ tal que $1 < a, b < m$ e $m = a \cdot b$. Assim, $\overline{a}, \overline{b} \in \mathbb{Z}_m$, $\overline{a}, \overline{b} \neq \overline{0}$, mas $\overline{a} \cdot \overline{b} = \overline{m} = \overline{0}$ (Absurdo).
Portanto, m é primo.

($\Leftarrow$) Suponhamos que m é primo e vamos mostrar que $\mathbb{Z}_m$ é um anel de integridade
Sejam $\overline{a}, \overline{b} \in \mathbb{Z}_m$, tal que $\overline{a} \cdot \overline{b} = \overline{ab} = \overline{0}$. Então, $\overline{a} \cdot \overline{b} = \overline{mq}$ (com $q \in \mathbb{Z}$), ou seja, $m|ab$. Como m é primo, $m|a$ ou $m|b$. Assim, temos que $\overline{a} = \overline{0}$ ou $\overline{b} = \overline{0}$. Ou seja, se m é primo, então $\mathbb{Z}_m$ não possui divisores de zero. Portanto, por definição, $\mathbb{Z}_m$ é um anel de integridade. □

2.2.3 Corpos

Definição 16. Se $\mathbb{K}$ é um anel comutativo com unidade que satisfaz: $\forall\ a \in \mathbb{K}, a \neq 0, \exists\ b \in \mathbb{K} / ab = 1$, ou seja, todo elemento não nulo é invertível, $\mathbb{K}$ é chamado *corpo*.

Exemplo 11. Os anéis $\mathbb{Q}, \mathbb{R}$ e $\mathbb{C}$, são corpos, mas o anel $\mathbb{Z}$ não é um corpo, pois os elementos 1 e -1 são os únicos invertíveis.

Teorema 2.2.10. Todo corpo $\mathbb{K}$ é um anel de integridade.

Demonstração. Sejam $a, b \in \mathbb{K}$ tais que $ab = 0$. Se $a = b = 0$ está provado. Suponhamos que $a \neq 0$, então existe $a^{-1} \in \mathbb{K}$ invertível. Multiplicando a igualdade $ab = 0$ por $a^{-1} \in \mathbb{K}$, tal que

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

Logo, a multiplicação de dois elementos de $\mathbb{K}$ é nula apenas se um dos elementos é nulo. Portanto $\mathbb{K}$ é um anel de integridade. □

Quando o anel de integridade é finito a recíproca do teorema 2.2.10 é verdadeira.

Definição 17. Seja $\mathbb{K}$ um corpo. Um subconjunto não vazio $\mathbb{M} \subset \mathbb{K}$ é *subcorpo* de $\mathbb{K}$ se:

(i) $1 \in \mathbb{M}$;

(ii) $\forall\ a, b \in \mathbb{M}, a - b \in \mathbb{M}$ e $a \cdot b^{-1} \in \mathbb{M}$.

Exemplo 12. $\mathbb{Q}$ é subcorpo de $\mathbb{R}$.

2.2.4 Extensões de Corpos

Para qualquer primo p existe um *corpo finito* de p elementos. De fato, para qualquer inteiro positivo m, é possível estender o corpo primo $GF(p)$ para um *corpo* de p^m elementos, o qual é chamado de uma *extensão do corpo GF(p)* e é denotado por $GF(p^m)$. Além disso, a *ordem* de qualquer corpo finito é uma potência de um primo.

Definição 18. Um corpo de Galois (nome em homenagem ao seu descobridor) é um corpo com número finito de elementos e é representado por $GF(p)$, onde p é número primo.

Usando a aritmética binária, podemos construir códigos a partir de $GF(2)$ ou $GF(2^m)$.

Definição 19. Um polinômio $p(x)$ sobre $GF(2)$ de grau m é dito ser *irredutível*, se ele não é divisível por qualquer polinômio sobre $GF(2)$ de grau menor que m, mas maior que zero.

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

Exemplo 13. O polinômio $x^4 + x + 1$ é um polinômio irredutível de *grau* 4, pois não é divisível por qualquer polinômio sobre $GF(2)$ de grau menor que 4, mas maior que zero.

Definição 20. Um polinômio $p(x)$ de grau m é dito ser *primitivo*, se o menor inteiro positivo n para o qual $p(x)$ divide $x^n + 1$ é $n = 2^m - 1$.

Exemplo 14. O polinômio $x^4 + x + 1$ é irredutível e primitivo, pois $p(x)$ divide $x^{15} + 1$, mas não divide nenhum $x^n + 1$, $1 \leqslant n < 15$.

Exemplo 15. O polinômio $x^4 + x^3 + x^2 + x + 1$ é irredutível, mas não é primitivo, pois $p(x)$ divide $x^{15} + 1$, mas também $x^5 + 1 (5 \neq 2^m - 1, \forall m \in \mathbb{N})$.

Analisando um método para construir *Corpos de Galois* de 2^m elementos ($m > 1$) do corpo binário $GF(2)$, serão utilizado os elementos 0 e 1 de $GF(2)$ e um novo símbolo α, definindo a multiplicação "$\cdot$" da seguinte forma:

$$
\begin{gathered}
0 \cdot 0 = 0 \\
0 \cdot 1 = 1 \cdot 0 = 0 \\
1 \cdot 1 = 1 \\
0 \cdot \alpha = \alpha \cdot 0 = 0 \\
1 \cdot \alpha = \alpha \cdot 1 = \alpha \\
\alpha^2 = \alpha \cdot \alpha \\
\alpha^3 = \alpha \cdot \alpha \cdot \alpha \\
. \\
. \\
. \\
\alpha^j = \alpha \cdot \alpha \cdots \alpha \ (j \text{ vezes}).
\end{gathered}
$$

A partir da definição anterior de multiplicação, segue que:

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

Portanto, dispomos de um conjunto de elementos no qual a operação de multiplicação está estabelecida: $F = \{0, 1, \alpha, \alpha^2, ..., \alpha^j, ...\}$.

Para garantir que o conjunto F tenha 2^m elementos e seja fechado em relação à operação de multiplicação definida anteriormente, introduziremos uma condição relativa ao elemento α. Seja $p(x)$ um *polinômio primitivo* de grau m em $GF(2)$ e considere α uma raiz de $p(x)$, ou seja, $p(\alpha) = 0$. Como $p(x)$ divide $x^{2^m-1} + 1$, temos que:

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

Substituindo x por α, obtemos:

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

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

Assim, sob a condição que $p(\alpha) = 0$, o conjunto F se torna finito e contém os elementos $F = \{0, 1, \alpha, \alpha^2, ..., \alpha^{2^m-2}\}$, que é um *corpo de Galois* de 2^m elementos, denotado por $GF(2^m)$.

No processo de construção de $GF(2^m)$ a partir de $GF(2)$, os elementos não nulos de $GF(2^m)$ podem ser representados por potências, polinômios ou vetores, sendo as representações por potência e polinomial convenientes para multiplicação e adição, respectivamente.

Exemplo 16. Seja $m = 4$ e considere o polinômio $p(x) = 1 + x + x^4$ primitivo sobre $GF(2)$. O conjunto F será dado da seguinte maneira:

$$F = \{0, 1, \alpha, \alpha^2, ..., \alpha^{2^4-2}\} \Rightarrow F = \{0, 1, \alpha, \alpha^2, ..., \alpha^{14}\}.$$

Seja $p(\alpha) = 0$, temos que $1 + \alpha + \alpha^4 = 0 \Rightarrow \alpha^4 = 1 + \alpha$. Assim, $\alpha^5 = \alpha \cdot \alpha^4 = \alpha \cdot (1 + \alpha) = \alpha + \alpha^2$. Com base nessa relação, podemos construir $GF(2^4)$, conforme a Tabela 2.

Tabela 2 – Representação dos elementos de $GF(2^4)$

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

Sejam os corpos $\mathbb{L}$ e $\mathbb{K}$ onde $\mathbb{K} \subseteq \mathbb{L}$. Um elemento $\alpha \in \mathbb{L}$ é denominado *elemento algébrico* sobre $\mathbb{K}$ se for solução de um polinômio com coeficientes em $\mathbb{K}$. Dizemos que $\mathbb{L}$ é *algébrico* sobre $\mathbb{K}$ se todo elemento de $\mathbb{L}$ é algébrico sobre $\mathbb{K}$.

Definição 21. Sejam $\mathbb{K}$ e $\mathbb{L}$ corpos. $\mathbb{L}$ é uma *extensão algébrica* de $\mathbb{K}$ se, e somente se, $\mathbb{K}$ é um subcorpo de $\mathbb{L}$.

Neste caso, $\mathbb{K} \subset \mathbb{L}$, $\mathbb{K}$ é um corpo mantendo as mesmas operações matemáticas definidas em $\mathbb{L}$, e com a identidade multiplicativa de $\mathbb{K}$ sendo igual à identidade multiplicativa de $\mathbb{L}$.

Todo corpo $\mathbb{K}$ é uma extensão de si mesmo, e as seguintes extensões são chamadas de extensões naturais: $\mathbb{Q} \subset \mathbb{R}$, $\mathbb{Q} \subset \mathbb{C}$ e $\mathbb{R} \subset \mathbb{C}$.

Definição 22. Dado um corpo $\mathbb{K}$ e um subconjunto $\mathbb{S} \subseteq \mathbb{K}$, o *subanel de* $\mathbb{K}$ *gerado por* $\mathbb{S}$ é a interseção de todos os subanéis de $\mathbb{K}$ que contém $\mathbb{S}$. O *subcorpo de* $\mathbb{K}$ *gerado por* $\mathbb{S}$ é a interseção de todos os subcorpos de $\mathbb{K}$ que contém $\mathbb{S}$.

Exemplo 17. O subanel de $\mathbb{R}$ gerado por $\{1\}$ é $\mathbb{Z}$ e o subcorpo de $\mathbb{R}$ gerado por $\{1\}$ é $\mathbb{Q}$.

Lema 2.2.11. Sejam $\mathbb{K}$ um corpo e $\mathbb{S}$ um subconjunto de $\mathbb{K}$, onde $1_{\mathbb{K}}$ é a identidade multiplicativa. Se R é o subanel de $\mathbb{K}$ gerado por $\mathbb{S}$, então R é um domínio e $\mathbb{F}$, o subcorpo de $\mathbb{K}$ gerado por $\mathbb{S}$, é o corpo de frações de R.

Demonstração. Por hipótese, R é um subanel, então $R \subset \mathbb{F}$ pois todo subcorpo é um subanel. Agora, como $1_{\mathbb{K}} \in \mathbb{S} \subseteq R$, temos que $1_R = 1_{\mathbb{K}} = 1$. Mais ainda, como $R \subseteq \mathbb{K}$, temos que R é um domínio.

Seja $\mathbb{F}' = \{a \cdot b^{-1}; a, b \in R, b \neq 0\}$ o corpo de frações de R. Desde que $R \subseteq \mathbb{F}$ e $\mathbb{F}'$ é o menor corpo que contém R, temos que $\mathbb{F}' \subseteq \mathbb{F}$. Mas temos $\mathbb{S} \subseteq R \subseteq \mathbb{F}' \subseteq \mathbb{K}$, ou seja, $\mathbb{F}'$ é um subcorpo de $\mathbb{K}$ que contém $\mathbb{S}$. Então, por definição, $\mathbb{F} \subseteq \mathbb{F}'$. Assim, $\mathbb{F}' = \mathbb{F}$. □

Definição 23. O *corpo primo* de $\mathbb{K}$, subcorpo de $\mathbb{K}$ gerado por $\{1_{\mathbb{K}}\}$, é a interseção de todos os subcorpos de $\mathbb{K}$.

Definição 24. Dada $\mathbb{L}$ uma extensão do corpo $\mathbb{K}$, o *grau da extensão*, ou grau de $\mathbb{L}$ sobre $\mathbb{K}$, denotada por $[\mathbb{L} : \mathbb{K}]$ é a dimensão de $\mathbb{L}$ sobre $\mathbb{K}$. Se o grau de $\mathbb{L}$ sobre $\mathbb{Q}$ é finito, então $\mathbb{L}$ é um *corpo de números*.

Exemplo 18. $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$ é uma extensão finita de grau 2. De fato, $\{1, \sqrt{2}\}$ gera $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2}; a, b \in \mathbb{Q}\}$ como $\mathbb{Q}$-espaço vetorial. Suponhamos que $a + b\sqrt{2} = 0$, com $a, b \in \mathbb{Q}$. Se $b \neq 0$, então $\sqrt{2} = -a \cdot b^{-1} \in \mathbb{Q}$, o que é um absurdo. Logo, $b = 0$ e, consequentemente, $a = 0$. Portanto, $\{1, \sqrt{2}\}$ é linearmente independente sobre $\mathbb{Q}$. Então, $[\mathbb{Q}(\sqrt{2}) : \mathbb{Q}] = 2$.

A extensão de grau 2 sobre o corpo $\mathbb{Q}$ dos números racionais do Exemplo 18 é chamada *corpo quadrático*. Todo corpo quadrático é da forma $\mathbb{Q}(\sqrt{d})$, onde d é um inteiro livre de quadrados, ou seja, $d \not\equiv 0 \pmod 4$.

Seja $A \subseteq R$ um anel. O conjunto dos elementos de R que são inteiros sobre A é chamado *anel dos inteiros* de A em R ou *fêcho inteiro* de A em R, denotado por $\mathbb{I}_R$.

Teorema 2.2.12. Seja $\mathbb{K} = \mathbb{Q}(\sqrt{d})$ um corpo quadrático, onde d é um inteiro livre de quadrados. Além disso, se

(i) $d \not\equiv 1 \pmod 4$, então o anel dos inteiros $\mathbb{I}_{\mathbb{K}}$ é $\mathbb{Z}[\sqrt{d}]$ e $\{1, \sqrt{d}\}$ é base de $\mathbb{Z}[\sqrt{d}]$ como um $\mathbb{Z}$-*módulo*.

(ii) $d \equiv 1 \pmod 4$, então o anel dos inteiros $\mathbb{I}_{\mathbb{K}}$ é $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ e $\{1, \frac{1+\sqrt{d}}{2}\}$ é base de $\mathbb{Z}[\frac{1+\sqrt{d}}{2}]$ como um $\mathbb{Z}$-*módulo*.

Uma demonstração do Teorema 2.2.12 pode ser encontrada em [21].

2.3 Códigos Corretores de Erros

Nesta seção serão apresentados os conceitos fundamentais referentes aos códigos corretores de erros, utilizados neste trabalho, sendo as referências utilizadas [3], [16], [17], [18], [19] e [20].

Um código corretor de erros (CCE) consiste, em sua essência, em uma abordagem organizada para incorporar informações adicionais em qualquer dado que precisa ser transmitido ou armazenado, possibilitando a detecção e/ou correção de erros ao recuperar a informação. Os CCEs desempenham um papel fundamental em nosso dia a dia, manifestando-se em diversas situações, como quando lidamos com informações digitalizadas, ao assistir à televisão ou navegar na internet, por exemplo.

Em um sistema de comunicação a *fonte de informação* produz a mensagem que será enviada, sendo o *transmissor* responsável por enviar a mensagem com sinal adequado. O *codificador de fonte* realiza a conversão do sinal da saída da fonte em uma sequência de dígitos, que são os códigos. Através do *canal*, que é o meio utilizado para enviar a mensagem do *transmissor* para o *receptor*, onde podem ser introduzidos os ruídos, o *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. O *decodificador de canal* realiza uma tentativa de corrigir alguns erros que possam aparecer, estimando os dígitos na saída do *codificador da fonte*. O *receptor* representa o usuário que vai receber a informação, onde o *decodificador de fonte* transforma a sequência estimada na saída do *decodificador de canal* em uma estimativa na saída da fonte, chegando a mensagem ao destinatário. Esse processo pode ser esquematizado através do diagrama da Figura 10.

Os CCEs são classificados como códigos de árvores, sendo divididos nas classes dos códigos de blocos e dos códigos de treliças, que podem ser lineares ou não lineares, sendo a primeira a mais aplicada na prática, baseada nas estruturas algébricas de grupo, anel, corpos finitos e suas extensões, sistematizando os processos de codificação, decodificação e análise de desempenho dos códigos.

Os elementos básicos para se construir um código são o *alfabeto*, que é um conjunto finito de elementos, sendo que cada um deles chama-se *dígito* e as *palavras-código* que

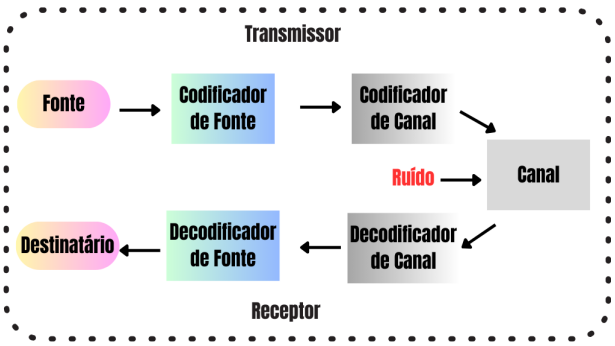

Figura 10 – Diagrama do funcionamento de um sistema de comunicação.

Fonte: Próprio autor.

são uma sequência finita de dígitos, tal que o número de dígitos de uma palavra-código denomina-se *comprimento*.

Definição 25. Codificação é o processo de mapeamento, ou seja, é uma conversão de uma dada sequência de dígitos (alfabeto fonte) em outra sequência de dígitos (alfabeto do código).

2.3.1 Códigos de Bloco Lineares

Os códigos de bloco caracterizam-se pelo fato do processo de codificação ocorrer através da transformação de conjuntos de bits ou símbolos, chamados blocos. Em outras palavras, uma sequência de bits ou símbolos é dividida em blocos compostos por k bits ou símbolos, a partir dos quais são geradas palavras-código com n bits ou símbolos.

Um código de bloco linear é construído com base em uma estrutura algébrica, como corpo, anel ou grupo, e sua identificação envolve principalmente três parâmetros: seu comprimento finito, sua dimensão e sua distância mínima, ou seja, (n, k, d_{min}).

Definição 26. Um *código de bloco* C de comprimento n sobre um alfabeto A é qualquer subconjunto do conjunto A^n das sequências $c = \{c_i | 1 \leq i \leq n\}$.

Definição 27. A *dimensão de um código* C é dada por $k = log|A||C|$, onde $|\cdot|$ denota a cardinalidade do conjunto.

2.3.2 Códigos de Hamming

Os *códigos de Hamming*, desenvolvidos por Richard Hamming, são códigos de blocos lineares que se fundamentam na inclusão de bits de paridade, sendo possível detectar erros ao adicionar bits de paridade a um conjunto específico de bits de dados. O seu uso permite a transferência e armazenamento de dados de forma segura.

Definição 28. O *peso de Hamming* de um vetor v, cuja notação é $\omega(v)$, é o número de elementos não-nulos em v. Para um vetor binário, o *peso de Hamming* é igual ao número de dígitos "1" contidos em v.

Exemplo 19. Seja a palavra-código $x = (0011010)$. O *peso de Hamming* é $\omega(x) = 3$

Definição 29. A *distância de Hamming* entre dois vetores códigos v e x, cuja notação é $d(v, x)$, é o número de posições em que os dígitos dos dois vetores são diferentes entre si. Para o caso binário, a *distância de Hamming* pode ser determinada por $d(v, x) = \omega(v \oplus x)$, tal que $\oplus$ denota a soma direta para o caso binário módulo 2.

Exemplo 20. A *distância de Hamming* entre os vetores $v = 11001$ e $x = 10111$ será dada por $d(v, x) = \omega(v \oplus x) = \omega(11001 \oplus 10111) = \omega(01110) = 3$.

Definição 30. A *distância mínima* de um código de bloco C, denotada por d_{min}, é a *menor distância de Hamming* entre dois vetores distintos quaisquer desse código. Ou seja, $d_{min} = \{d(x, y) : x, y \in C, x \neq y\}$.

Exemplo 21. A *distância mínima* do código $C = \{(01011), (11110)\}$ e $\{(01101)\}$ é dada por $d_{min} = \{d((01011), (11110), (01101))\} = \omega(01011 + 11110 + 01101) = \omega(11000) = 2$.

Definição 31. A *taxa de codificação* R_c de um código de bloco é a razão entre o número de bits de informação k e o número de bits da palavra-código n, ou seja, $R_c = \frac{k}{n}$. Este é um importante parâmetro de caracterização de um código, indicando seu desempenho.

Um código de bloco divide a sequência contínua de dígitos de informação na entrada do codificador em blocos de k símbolos e processa esses blocos de forma independente, conforme o código utilizado. Cada possível bloco de informação é associado a uma *n-upla* de símbolos de canal, onde $n > k$. O resultado é transmitido, sujeito a ser corrompido pelo ruído, e decodificado independentemente dos outros blocos. A principal classe dos códigos de bloco é a dos códigos lineares.

As capacidades de detecção de erros e de correção de erros de um código de bloco (n, k, d_{min}), estão relacionados com a distância mínima, denotadas, respectivamente, por ϱ e t, sendo dadas por $\varrho = d_{min} - 1$ e $t = \lfloor \frac{d_{min}-1}{2} \rfloor$. Isso significa que quanto maior for a distância mínima do código, maior será a sua capacidade de corrigir e de detectar erros. Os códigos eficazes costumam ser longos, o que torna impraticável descrevê-los através de listas extensas de palavras-código. Logo, os códigos de blocos lineares podem facilitar as operações de codificação e decodificação, através da associação de estruturas matemáticas.

Definição 32. Um *código de bloco* de comprimento n com 2^k *palavras-código* é um código linear se, e somente se, as suas 2^k *palavras-código* formam um subespaço de dimensão k do espaço vetorial formado pelas 2^k $n - uplas$ possíveis em $GF(2)$.

Seja $B = \{v_1, v_2, ..., v_k\}$ uma base do subespaço vetorial C de dimensão k. Assim, qualquer *palavra-código* $v = (a_1, a_2, ..., a_k) \in C$ pode ser escrita como combinação linear dos vetores da base: $v = a_1v_1 + a_2v_2 + ... + a_kv_k$, em que $a_i \in \{0, 1, ..., q-1\}$ e a soma é realizada em *módulo q*.

A matriz geradora do código (n, k, d) é formada por esses vetores que compõem uma base para o subespaço C, sendo organizados em linhas. Essas linhas na matriz G constituem o código linear C, no qual cada palavra-código é uma combinação linear das linhas de G. Portanto, se a dimensão do espaço vetorial C é k, o número de linhas em G é igual a k, uma vez que as linhas em G são linearmente independentes. Assim, a matriz G é dada por:

$$G = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{pmatrix}.$$

Qualquer correspondência que relacione uma $k-upla$ a uma palavra-código pode servir como um método de codificação, expresso pela fórmula

$$v = u \cdot G,$$

em que u representa a $k-upla$ de símbolos de informação a serem codificadas e a $n-upla$ v é a palavra-código correspondente.

Além disso, podemos escrever a palavra-código v como:

$$v \cdot H^T = 0, \tag{2.1}$$

onde H representa a *matriz verificação de paridade* de C, sendo de ordem $n-k$. Qualquer vetor que seja ortogonal às linhas da matriz geradora G está contido no espaço vetorial gerado por essas linhas. Com a equação 2.1, podemos verificar se uma palavra-código v pertence a um código C.

Definição 33. Seja um código C com *matriz de verificação de paridade* H, a *síndrome* de um vetor $v \in F_q$ é o vetor $v \cdot H^T = s$.

O argumento *síndrome* é utilizado para corrigir erros em códigos lineares, sendo fundamental em um processo de decodificação, para detectar erros e possivelmente corrigi-los. A expressão *padrão de erro* designa a diferença entre a palavra-código recebida e a palavra-código enviada. Um código de bloco linear $C(n, k)$ com capacidade de correção de t erros é capaz de corrigir um total de 2^{n-k} padrões de erros.

Na subseção 2.3.3 serão ilustrados os conceitos apresentados anteriormente por meio da construção do código $C(6,3)$.

2.3.3 Construção do Código $C(6,3)$

Seja o subespaço vetorial formado pelos oito vetores:

$$\{(000000),(110101),(101100),(011110),(011001),(101011),(110010),(000111)\}.$$

Ao escolher três vetores LI (Linearmente Independentes), como $(101100),(011110)$ e (000111), estes geram todo o subespaço vetorial e formam uma base do subespaço vetorial, conforme apresentado a seguir.

Os vetores são LI, pois:

$\alpha(101100)+\beta(011110)+\delta(000111)=(000000)$,

$$\Longrightarrow \begin{cases} \alpha = 0 \\ \beta = 0 \\ \alpha+\beta+\delta = 0 \Rightarrow \delta = 0. \end{cases}$$

Além disso, os três vetores geram todo o subespaço vetorial:

$$\begin{aligned}
000 &= 0(101100)+0(011110)+0(000111)=(000000)\\
001 &= 0(101100)+0(011110)+1(000111)=(000111)\\
010 &= 0(101100)+1(011110)+0(000111)=(011110)\\
011 &= 0(101100)+1(011110)+1(000111)=(011001)\\
100 &= 1(101100)+0(011110)+0(000111)=(101100)\\
101 &= 1(101100)+0(011110)+1(000111)=(101011)\\
110 &= 1(101100)+1(011110)+0(000111)=(110010)\\
111 &= 1(101100)+1(011110)+1(000111)=(110101).
\end{aligned}$$

A partir da base do subespaço vetorial, obtém-se a matriz geradora G, em que cada vetor da base equivale a uma linha da matriz geradora. Dessa forma, a matriz G será:

$$G=\begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix}=\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

Esta matriz encontra-se na forma *não sistemática*. Para se obter uma G', chamada de *sistemática*, efetua-se operações com as linhas da G de modo que a matriz fique da forma: $G'=[P_{k\times(n-k)}|I_{k\times k}]$.

Neste caso, tem-se $n=6$ e $k=3$, logo $G'=[P_{3\times3}|I_{3\times3}]$. Assim, G' pode ser obtida realizando-se as seguintes operações:

$$g'_0 = g_0 = (101100)$$

$$g'_1 = g_0 + g_1 = (101100) + (011110) = (110010)$$

$$g'_2 = g_1 + g_2 = (011001)$$

$$G' = \begin{bmatrix} g'_0 \\ g'_1 \\ g'_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Os vetores mensagens e seus respectivos vetores códigos são apresentados na Tabela 3:

Tabela 3 – Palavras-código do código C(6, 3) na forma sistemática

m	$m \cdot G'$	$c = m \cdot G'$
000	0(101100) + 0(110010) + 0(011001)	(000000)
001	0(101100) + 0(110010) + 1(011001)	(011001)
010	0(101100) + 1(110010) + 0(011001)	(110010)
011	0(101100) + 1(110010) + 1(011001)	(101011)
100	1(101100) + 0(110010) + 0(011001)	(101100)
101	1(101100) + 0(110010) + 1(011001)	(110101)
110	1(101100) + 1(110010) + 0(011001)	(011110)
111	1(101100) + 1(110010) + 1(011001)	(000111)

A partir de G', obtém-se a *Matriz Verificadora de Paridade* que será da forma $H = [I_{(n-k)\times(n-k)}|P^T]$, ou seja, $H = [I_{3\times3}|P^T]$.

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}_{3\times6}.$$

Ao calcular a *distância mínima*, obtém-se os resultados apresentados na Tabela 4, onde verifica-se que, para este código, a *distância mínima* é igual a 3, logo o código será da forma $C(6, 3, 3)$.

As capacidades de correção (t) e detecção (ϱ) de erros do código serão dadas por: $\varrho = d_{min} - 1$ e $t = \lfloor \frac{d_{min}-1}{2} \rfloor$, ou seja, $\varrho = 3 - 1 = 2$ e $t = \lfloor \frac{3-1}{2} \rfloor = 1$.

Portanto, o código $C(6, 3, 3)$ tem capacidade de detectar até dois erros e corrigir apenas um.

A *síndrome de erros* do código $C(6, 3, 3)$ dada por $s = v \cdot H^T$ será apresentada na Tabela 5, lembrando que:

Tabela 4 – Distância mínima do código C(6, 3).

m	c	ω
000	(000000)	0
001	(011001)	3
010	(110010)	3
011	(101011)	4
100	(101100)	3
101	(110101)	4
110	(011110)	4
111	(000111)	3

$$H = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}_{3\times 6}.$$

Logo,

$$H^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}_{6\times 3}.$$

Tabela 5 – Síndrome de erros do código $C(6, 3, 3)$.

$e(t = 1)$	s
100000	100
010000	010
001000	001
000100	101
000010	110
000001	011

2.3.4 Códigos BCH

Os códigos cíclicos representam uma categoria fundamental dentro dos códigos de blocos lineares. Um código de bloco C é considerado cíclico quando qualquer deslocamento de uma palavra código resulta em outra palavra código. Esse processo é conhecido como *deslocamento cíclico de v*, quando considerado um código linear C e um vetor v associado a C.

O código de Hamming é um código de bloco linear utilizado no processamento de sinal e nas telecomunicações. Uma extensão dos códigos de Hamming são os códigos

BCH (Bose, Chaudhuri e Hocquenghen) binários, que destacam-se pela capacidade de corrigir múltiplos erros. Todavia, a eficácia desses códigos é notável em situações em que o comprimento das palavras-código é relativamente curto, caso contrário, quando as palavras-código são extensas, a performance desses códigos é comprometida devido às taxas de transmissão mais baixas. Os códigos *BCH* constituem uma significativa categoria cíclica de códigos corretores de erros. Em sua formulação, empregam corpos finitos e são representados por meio de polinômios sobre $GF(p)$, onde p é um número primo. O código *BCH* é originado a partir de um polinômio, cujas raízes são expressas no corpo finito $GF(2^m)$.

Definição 34. Para todo inteiro $m \geqslant 3$ e $t < 2^m - 1$, é possível encontrar um código *BCH* binário com a capacidade de corrigir até t erros, caracterizado pelos seguintes parâmetros:

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

Definição 35. Seja $g(x) = 1 + g_1x + g_2x^2 + \cdots + g_{n-k-1}x^{n-k-1} + g_{n-k}x^{n-k}$ um polinômio não nulo de grau mínimo $n - k$ de um código cíclico binário $C(n, k)$. O polinômio $g(x)$ é denominado de polinômio gerador de $C(n, k)$.

Definição 36. Seja $f(x)$ um polinômio com coeficientes em $GF(2)$. Se um elemento β pertencente a $GF(2^m)$ é uma raiz de $f(x)$, então o polinômio $f(x)$ também tem como raízes β^{2^l} para qualquer $l \geqslant 0$. Esse elemento é designado como o 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 possui $\alpha, \alpha^2, \alpha^3, \cdots, \alpha^{2t}$ e seus conjugados como raízes. Quando α é um elemento primitivo de $GF(2^m)$, o código *BCH* resultante é classificado como um código *BCH* primitivo.

Definição 37. 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 $\{\phi_1(x), \phi_2(x), \cdots, \phi_{2t}(x)\}$:

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

Definição 38. 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:

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

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*.

Essa construção mencionada anteriormente será importante para o processo da codificação que será estudado. No próximo capítulo será utilizada a construção de códigos BCH primitivos e não primitivos, ambos sobre o anel local Z_q de ordem $n = (p^r - 1)$, onde $q = p^k$, $p = k = 2$ e r é o grau da *extensão de Galois* e as raízes do polinômio gerador estão na extensão do anel Z_q. Se a ordem do corpo base, p, e o comprimento das palavras-código, n, são primos entre si, isto é, $mdc(p, n) = 1$, então $x^n - 1$ não possui multiplicidade de raízes.

3 Algoritmo de Codificação

Neste capítulo serão apresentados os principais resultados deste trabalho, sendo feita inicialmente uma analogia entre os sistemas de comunicação e de informação genômica como esquematizado na Figura 11, em que o DNA representa o transmissor da mensagem, o canal são os processos de transcrição e tradução, onde podem ocorrer os erros, que seriam as mutações, e as proteínas representam o receptor.

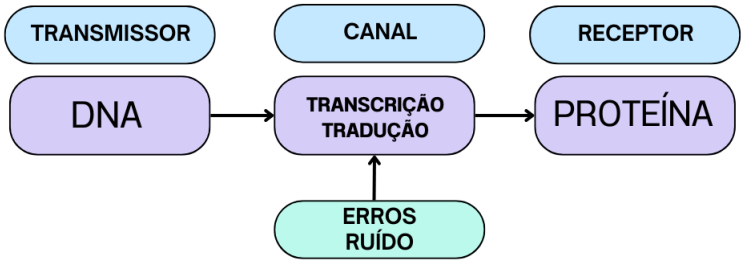

Figura 11 – Analogia entre o sistema de informação genômica e o sistema de comunicação simplificado.

Fonte: Próprio autor.

A *palavra-código* será denotada pela sequência de dígitos $a = (a_1, a_2, \cdots, a_n)$, a sequência de DNA do *NCBI (National Center for Biotechnology Information) - Seq*, será denotada por $b = (b_1, b_2, \cdots, b_n)$, onde $a_i's$, $b_i's \in \mathbb{Z}_4$, e os *padrões de erros* dados pela diferença entre a palavra-código e a sequência de DNA do *NCBI*, serão denotados por $D(a, b)$.

O passo a passo do algoritmo de codificação para geração de sequências de DNA, que descreve a construção de códigos BCH sobre o *anel de Galois* $\mathbb{Z}_4$, desenvolvido por [4] e [6], é descrito a seguir:

Passo 1 - Especificar a estrutura matemática e o alfabeto do código;
Passo 2 - Determinar a extensão de Galois;
Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois;
Passo 4 - Determinar a extensão do corpo $GF(2)$;
Passo 5 - Determinar a extensão do anel $\mathbb{Z}_4$;
Passo 6 - Determinar o grupo das unidades para o código BCH primitivo, quando o comprimento da sequência de DNA for igual a $n = (2^r - 1)$, ou, determinar o subgrupo das unidades para o código BCH não primitivo, quando o comprimento da sequência de DNA for um submúltiplo de $n = (2^r - 1)$;

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

1^o) Cálculos das raízes dos polinômios minimais;

2^o) Cálculos dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, ..., n-1$;

3^o) Cálculos dos polinômios geradores para todos os valores de t relacionados à distância de Hamming $d_H \geq 2t + 1$;

Passo 8 - Determinar o polinômio gerador da matriz $H, h(x)$;
Passo 9 - Determinar a matriz G e a sua transposta G^T;
Passo 10 - Determinar a matriz H e a sua transposta H^T;
Passo 11 - Rotular a sequência de DNA usando o Passo 1;
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 $D(a, b) = 2$;
Passo 13 - Comparar todas as palavras-código armazenadas no Passo 12 com a sequência de DNA do $NCBI$ e mostrar onde os erros ocorreram;
Passo 14 - Voltar para o Passo 7 e determinar outro $g(x)$;
Passo 15 - Repetir os Passos 8 ao 12 para o $g(x)$ obtido no Passo 14, até que se esgotem todas as possibilidades de $g(x)$;
Passo 16 - Voltar para o Passo 3 e escolher outro $p(x)$, e, então, repetir os Passos 4 ao 14 até esgotar todos os $p(x)$ do Passo 3;
Passo 17 - Fim.

A seguir será apresentado um exemplo, analisando uma sequência de DNA contendo 63 nucleotídeos, por ser este o menor comprimento de uma sequência de DNA que satisfaz as condições dos códigos BCH.

Exemplo 22. Vamos analisar a sequência de DNA do sinal interno, SI, de uma proteína mitocondrial, GI número 832917, com um comprimento de n = 63 nucleotídeos (Seq.36), dada por

$$\{5' - GCCGTTCATGTTTACTCTGGGTTGCCTTGGTGGGGAACTATCGCGGCCACCACCATCCTCATT - 3'\}$$

Essa sequência de DNA pode ser obtida através da busca no banco de dados do *NCBI*, um *website* oficial do governo dos Estados Unidos, acessando o endereço eletrônico <https://www.ncbi.nlm.nih.gov/>, cuja estrutura da *home page* pode ser vista na Figura 12. Como o exemplo analisado será da sequência específica com número de identificação 832917, basta realizar a busca através dessa numeração. Porém, apesar de ter sido utilizada neste trabalho a mesma sequência de [4], a sequência foi atualizada no banco de dados no *NCBI* e já não possui o comprimento de 63 nucleotídeos. Uma sequência com esse comprimento pode ser obtida realizando-se a busca conforme mostra a Figura 13. Para isso, seleciona-se a opção "Gene" e insere-se "63[Gene Lenght]", o que resulta em 13.779 possibilidades.

Figura 12 – Página inicial do *NCBI*.

Fonte: [30]

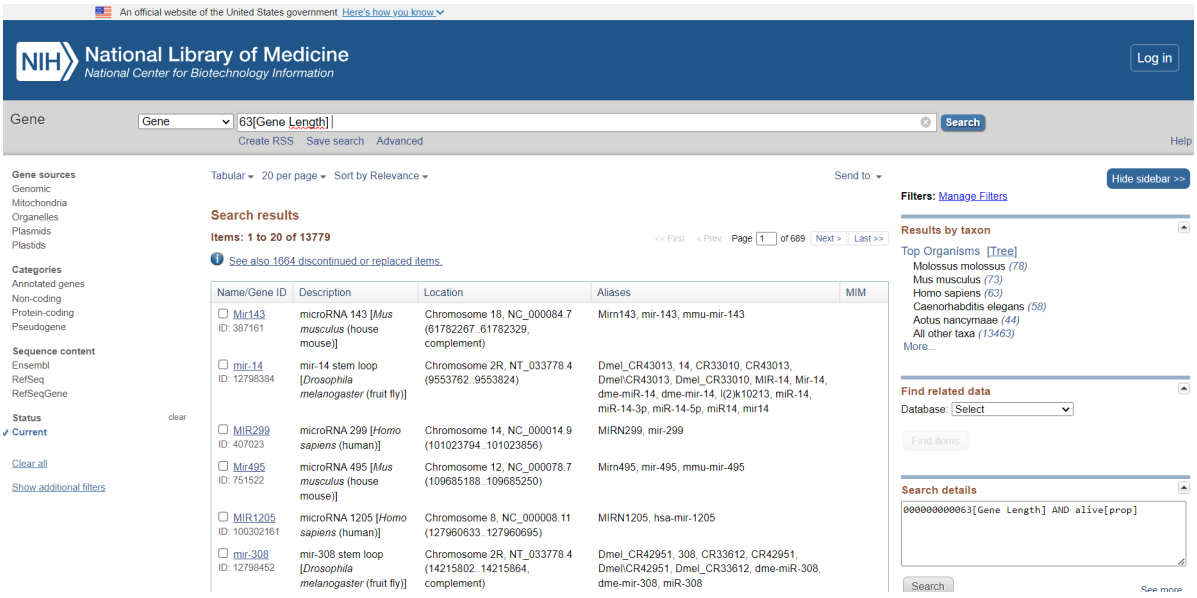

Figura 13 – Busca realizada no *NCBI* de uma sequência de DNA contendo 63 nucleotídeos.

Fonte: [31]

O código BCH primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (63, k, d_H)$ tem a capacidade de produzir e replicar sequências de DNA de comprimento $n = (2^r - 1) = (2^6 - 1) = 63$, com $D(a, b) = 1$ e $D(a, b) = 2$, permitindo até dois nucleotídeos diferindo da sequência de DNA do *NCBI*, onde r é o grau da extensão de Galois.

Passo 1 - Especificar a estrutura matemática e o alfabeto do código: O conjunto de nucleotídeos denotado por $N = \{A, C, G, T/U\}$, representando Adenina, Citosina, Guanina e Timina ou Uracila, está relacionado ao alfabeto quaternário do código genético. Por essa razão, será adotado o alfabeto quaternário representado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, seguindo as operações de adição e multiplicação módulo 4 (conforme as Tabelas 6 e 7), conferindo-lhe uma estrutura algébrica de anel.

Tabela 6 – Adição módulo 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

Tabela 7 – Multiplicação módulo 4.

·	**0**	**1**	**2**	**3**
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Passo 2 - Determinar a extensão de Galois: Para garantir a unicidade da fatoração de $x^n - 1$ no grupo das unidades, $GR^*(4, r)$, é necessário que o comprimento da sequência de DNA seja ímpar, expresso como $n = (2^r - 1)$, onde n representa o número de elementos não nulos no corpo de Galois, $GF(2^r)$, ou, de forma equivalente, os elementos que possuem inverso. 2^r indica a cardinalidade do conjunto $GF(2^r)$, enquanto r denota o grau de extensão do corpo de Galois. Os valores de r serão empregados na extensão do corpo $GF(2)$ no **Passo 4**.

Será analisada a sequência de DNA do sinal interno, SI, de uma proteína mitocondrial, GI número 832917, com um comprimento de $n = 63$ nucleotídeos ($Seq.36$). Portanto, o grau r do polinômio primitivo utilizado na extensão de Galois do corpo $GF(2)$ é $r = 6$, visto que $n = 2^r - 1 = 2^6 - 1 = 63$. Assim, esse valor de $r = 6$ será aplicado na extensão do corpo $GF(2)$ no **Passo 4**.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois: os $p(x)$ relacionados ao grau de extensão de Galois $r = 6$ podem ser encontrados em [20]:

- $p_1(x) = x^6 + x + 1$
- $p_2(x) = x^6 + x^4 + x^3 + x + 1$
- $p_3(x) = x^6 + x^5 + 1$

- $p_4(x) = x^6 + x^5 + x^2 + x + 1$
- $p_5(x) = x^6 + x^5 + x^3 + x^2 + 1$
- $p_6(x) = x^6 + x^5 + x^4 + x + 1$

Passo 4 - Determinar a extensão do corpo $GF(2)$: o corpo $GF(2^r)$ é construído através da divisão do anel de todos os polinômios com coeficientes em $GF(2)$, denotado por $GF(2)[x]$, por um ideal gerado por qualquer polinômio primitivo de grau $r = 6$. Esse processo constitui a extensão do corpo $GF(2)$ da seguinte forma:

Seja o corpo de Galois $GF(2^r) = GF(2^6) = GF(64) = F_{64}$ definido por

$$\frac{F_2[x]}{p(x)} \cong \frac{F_2[x]}{x^6 + x + 1} = \left\{ a_0 + a_1 x + a_2 x^2 + ... + a_5 x^5 : a_i' s \in F_2 \right\},$$

onde $p(x)$ é o polinômio primitivo $p_1(x)$ do **Passo 3**.

Seja α uma raiz de $x^6+x+1 = 0$, logo $\alpha^6+\alpha+1 = 0$, implicando em $\alpha^6 = \alpha+1$, pois os coeficientes dos polinômios que compõem o conjunto dos elementos de F_{64} pertencem a F_2. Os elementos de F_{64} estão representados na Tabela 8.

Tabela 8 – Elementos de F_{64} em notação de r-uplas com $p_1(x)$.

Elementos de F_{64}	$(\alpha^0, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5)$	Elementos de F_{64}	$(\alpha^0, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5)$
0	(000000)	$\alpha^{32} = 1 + \alpha^3$	(100100)
1	(100000)	$\alpha^{33} = \alpha + \alpha^4$	(100100)
α	(010000)	$\alpha^{34} = \alpha^2 + \alpha^5$	(001001)
α^2	(001000)	$\alpha^{35} = 1 + \alpha + \alpha^3$	(110100)
α^3	(000100)	$\alpha^{36} = \alpha + \alpha^2 + \alpha^4$	(011010)
α^4	(000010)	$\alpha^{37} = \alpha^2 + \alpha^3 + \alpha^5$	(001101)
α^5	(000001)	$\alpha^{38} = 1 + \alpha + \alpha^3 + \alpha^4$	(110110)
$\alpha^6 = 1 + \alpha$	(110000)	$\alpha^{39} = \alpha + \alpha^2 + \alpha^4 + \alpha^5$	(011011)
$\alpha^7 = \alpha + \alpha^2$	(011000)	$\alpha^{40} = 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^5$	(111101)
$\alpha^8 = \alpha^2 + \alpha^3$	(001100)	$\alpha^{41} = 1 + \alpha^2 + \alpha^3 + \alpha^4$	(101110)
$\alpha^9 = \alpha^3 + \alpha^4$	(000110)	$\alpha^{42} = \alpha + \alpha^3 + \alpha^4 + \alpha^5$	(010111)
$\alpha^{10} = \alpha^4 + \alpha^5$	(000011)	$\alpha^{43} = 1 + \alpha + \alpha^2 + \alpha^4 + \alpha^5$	(111011)
$\alpha^{11} = 1 + \alpha + \alpha^5$	(110001)	$\alpha^{44} = 1 + \alpha^2 + \alpha^3 + \alpha^5$	(101101)
$\alpha^{12} = 1 + \alpha^2$	(101000)	$\alpha^{45} = 1 + \alpha^3 + \alpha^4$	(100110)
$\alpha^{13} = \alpha + \alpha^3$	(010100)	$\alpha^{46} = \alpha + \alpha^4 + \alpha^5$	(010011)
$\alpha^{14} = \alpha^2 + \alpha^4$	(001010)	$\alpha^{47} = 1 + \alpha + \alpha^2 + \alpha^5$	(111001)
$\alpha^{15} = \alpha^3 + \alpha^5$	(000101)	$\alpha^{48} = 1 + \alpha^2 + \alpha^3$	(101100)
$\alpha^{16} = 1 + \alpha + \alpha^4$	(110010)	$\alpha^{49} = \alpha + \alpha^3 + \alpha^4$	(010110)
$\alpha^{17} = \alpha + \alpha^2 + \alpha^5$	(011001)	$\alpha^{50} = \alpha^2 + \alpha^4 + \alpha^5$	(001011)
$\alpha^{18} = 1 + \alpha + \alpha^2 + \alpha^3$	(111100)	$\alpha^{51} = 1 + \alpha + \alpha^3 + \alpha^5$	(110101)
$\alpha^{19} = \alpha + \alpha^2 + \alpha^3 + \alpha^4$	(011110)	$\alpha^{52} = 1 + \alpha^2 + \alpha^4$	(101010)
$\alpha^{20} = \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(001111)	$\alpha^{53} = \alpha + \alpha^3 + \alpha^5$	(010101)
$\alpha^{21} = 1 + \alpha + \alpha^3 + \alpha^4 + \alpha^5$	(110111)	$\alpha^{54} = 1 + \alpha + \alpha^2 + \alpha^4$	(111010)
$\alpha^{22} = 1 + \alpha^2 + \alpha^4 + \alpha^5$	(101011)	$\alpha^{55} = \alpha + \alpha^2 + \alpha^3 + \alpha^5$	(011101)
$\alpha^{23} = 1 + \alpha^3 + \alpha^5$	(100101)	$\alpha^{56} = 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4$	(111110)
$\alpha^{24} = 1 + \alpha^4$	(100010)	$\alpha^{57} = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(011111)
$\alpha^{25} = \alpha + \alpha^5$	(010001)	$\alpha^{58} = 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(111111)
$\alpha^{26} = 1 + \alpha + \alpha^2$	(111000)	$\alpha^{59} = 1 + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(101111)
$\alpha^{27} = \alpha + \alpha^2 + \alpha^3$	(011100)	$\alpha^{60} = 1 + \alpha^3 + \alpha^4 + \alpha^5$	(100111)
$\alpha^{28} = \alpha^2 + \alpha^3 + \alpha^4$	(001110)	$\alpha^{61} = 1 + \alpha^4 + \alpha^5$	(100011)
$\alpha^{29} = \alpha^3 + \alpha^4 + \alpha^5$	(000111)	$\alpha^{62} = 1 + \alpha^5$	(100001)
$\alpha^{30} = 1 + \alpha + \alpha^4 + \alpha^5$	(110011)	$\alpha^{63} = 1$	(100000)
$\alpha^{31} = 1 + \alpha^2 + \alpha^5$	(111001)		

Passo 5 - Determinar a extensão do anel $\mathbb{Z}_4$: o anel $GR(p^k, r) = GR(4, 6)$ é construído através da divisão do anel $\mathbb{Z}_4$ (conjunto de todos os polinômios com coeficientes em $\mathbb{Z}_4$) pelo ideal gerado pelo mesmo $p(x)$ utilizado para realizar a extensão do corpo no **Passo 4**, ou seja,

$\frac{\mathbb{Z}_4[x]}{p(x)} \cong \frac{\mathbb{Z}_4[x]}{x^6+x+1} = \{b_0 + b_1x + b_2x^2 + ... + b_5x^5 : b_i's \in \mathbb{Z}_4\}$, onde $p(x)$ é o polinômio primitivo $p_1(x)$ do **Passo 3**.

Tanto na extensão do corpo quanto na do anel, α é uma raiz do polinômio primitivo, logo $\alpha^6 = -\alpha - 1$, mas como os coeficientes dos polinômios estão em $\mathbb{Z}_4$, $\alpha^6 = 3\alpha + 3$. Seja $f = \alpha = (010000)$, todos os elementos não nulos e inversíveis do grupo cíclico de $GR^*(4, 6)$ são obtidos por meio da potenciação de f, conforme ilustrado na Tabela 9.

Tabela 9 – Elementos do grupo cíclico do grupo $GR^*(4, 6)$ em notação de r-uplas com $p_1(x)$.

$GR^*(4,6)$	$(\alpha^0, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5)$	$GR^*(4,6)$	$(\alpha^0, \alpha^1, \alpha^2, \alpha^3, \alpha^4, \alpha^5)$
1	(100000)	$f^{10} = x^{10} = \alpha^{10} = 3\alpha^4 + 3\alpha^5$	(000033)
$f = x = \alpha$	(010000)	$f^{11} = x^{11} = \alpha^{11} = 1 + \alpha + 3\alpha^5$	(110003)
$f^2 = x^2 = \alpha^2$	(001000)	$f^{12} = x^{12} = \alpha^{12} = 1 + \alpha^2$	(101000)
$f^3 = x^3 = \alpha^3$	(000100)	. . .	. . .
$f^4 = x^4 = \alpha^4$	(000010)	$f^{121} = x^{121} = \alpha^{121}$	(331313)
$f^5 = x^5 = \alpha^5$	(000001)	$f^{122} = x^{122} = \alpha^{122}$	(103131)
$f^6 = x^6 = \alpha^6 = 3 + 3\alpha$	(330000)	$f^{123} = x^{123} = \alpha^{123}$	(300313)
$f^7 = x^7 = \alpha^7 = 3\alpha + 3\alpha^2$	(033000)	$f^{124} = x^{124} = \alpha^{124}$	(100031)
$f^8 = x^8 = \alpha^8 = 3\alpha^2 + 3\alpha^3$	(003300)	$f^{125} = x^{125} = \alpha^{125}$	(300003)
$f^9 = x^9 = \alpha^9 = 3\alpha^3 + 3\alpha^4$	(000330)	$f^{126} = x^{126} = \alpha^{126}$	(100000)

Passo 6 - Determinar o grupo das unidades: A partir do **Passo 5**, conclui-se que f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 6)$, onde $d \geqslant 1 \in Z$, e f^d gera o subgrupo cíclico cuja ordem é 63 em $GR^*(4, 6)$. Assim, temos que $n \cdot d = 63 \cdot d = 126$, o que implica que $d = 2$. Como resultado, $f^2 = (001000) = \alpha^2$ gera um subgrupo cíclico de ordem 63 em $GR^*(4, 6)$. Portanto, $\beta = \alpha^2$ é o elemento primitivo que gera o subgrupo cíclico $G_n = G_{63}$. Assim, temos que:

$$
\begin{aligned}
\beta &= \alpha^2 \rightarrow (001000) \\
\beta^2 &= (\alpha^2)^2 = \alpha^4 \rightarrow (000010) \\
\beta^3 &= (\alpha^2)^3 = \alpha^6 = 3 + 3\alpha \rightarrow (330000) \\
\beta^4 &= (\alpha^2)^4 = \alpha^8 = \alpha^2 \cdot \alpha^6 = \alpha^2 \cdot (3 + 3\alpha) = 3\alpha^2 + 3\alpha^3 \rightarrow (003300) \\
\beta^5 &= (\alpha^2)^5 = \alpha^{10} = \alpha^2 \cdot \alpha^8 = \alpha^2 \cdot (3\alpha^2 + 3\alpha^3) = 3\alpha^4 + 3\alpha^5 \rightarrow (000033) \\
\beta^6 &= (\alpha^2)^6 = \alpha^{12} = \alpha^2 \cdot \alpha^{10} = \alpha^2 \cdot (3\alpha^4 + 3\alpha^5) = 3\alpha^6 + 3\alpha^7 = 3 \cdot (3 + 3\alpha) + 3\alpha \cdot (3 + 3\alpha) \\
&= 9 + 9\alpha + 9\alpha + 9\alpha^2 = 1 + 2\alpha + \alpha^2 \rightarrow (121000) \\
\beta^7 &= (\alpha^2)^7 = \alpha^{14} = \alpha^2 \cdot \alpha^{12} = \alpha^2 \cdot (1 + 2\alpha + \alpha^2) = \alpha^2 + 2\alpha^3 + \alpha^4 \rightarrow (001210)
\end{aligned}
$$

$$
\begin{aligned}
\beta^8 &= (\alpha^2)^8 = \alpha^{16} = \alpha^2 \cdot \alpha^{14} = \alpha^2 \cdot (\alpha^2 + 2\alpha^3 + \alpha^4) = \alpha^4 + 2\alpha^5 + \alpha^6 = \\
&= 3 + 3\alpha + \alpha^4 + 2\alpha^5 \rightarrow (330012) \\
\beta^9 &= (\alpha^2)^9 = \alpha^{18} = \alpha^2 \cdot \alpha^{16} = \alpha^2 \cdot (3 + 3\alpha + \alpha^4 + 2\alpha^5) = 3\alpha^2 + 3\alpha^3 + \alpha^6 + 2\alpha^7 = \\
&= 3\alpha^2 + 3\alpha^3 + 3 + 3\alpha + 2\alpha + 2\alpha^2 = 3 + \alpha + \alpha^2 + 3\alpha^3 \rightarrow (311300) \\
\beta^{10} &= (\alpha^2)^{10} = \alpha^{20} = \alpha^2 \cdot \alpha^{18} = \alpha^2 \cdot (3 + \alpha + \alpha^2 + 3\alpha^3) = 3\alpha^2 + \alpha^3 + \alpha^4 + 3\alpha^5 \rightarrow (003113) \\
&\vdots \\
\beta^{62} &= (\alpha^2)^{62} = \alpha^{124} = \alpha^2 \cdot \alpha^{122} = \alpha^2 \cdot (1 + 3\alpha^2 + \alpha^3 + 3\alpha^4 + \alpha^5) = \alpha^2 + 3\alpha^4 + \alpha^5 + 3\alpha^6 + \alpha^7 \\
&= \alpha^2 + 3\alpha^4 + \alpha^5 + 1 + \alpha + 3\alpha + 3\alpha^2 = 1 + 3\alpha^4 + \alpha^5 \rightarrow (100031) \\
\beta^{63} &= (\alpha^2)^{63} = \alpha^{126} = \alpha^2 \cdot \alpha^{124} = \alpha^2 \cdot (1 + 3\alpha^4 + \alpha^5) = \alpha^2 + 3\alpha^6 + \alpha^7 = \\
&= \alpha^2 + 1 + \alpha + 3\alpha + 3\alpha^2 = 1 \rightarrow (100000)
\end{aligned}
$$

como mostrado na Tabela 10. Este elemento primitivo será utilizado na construção de um código BCH de comprimento $n = 63$ sobre $\mathbb{Z}_4$. Quando o comprimento n da palavra-código desejada for igual à cardinalidade de G_n, procederemos à construção de um código BCH primitivo, onde f gera um grupo cíclico de ordem $n \cdot 2$ em $GR^*(4, r)$.

Tabela 10 – Elementos de G_{63}

$G_{63} \rightarrow \alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5$	$G_{63} \rightarrow \alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5$	$G_{63} \rightarrow \alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5$
$\beta \rightarrow (001000)$	$\beta^{22} \rightarrow (303101)$	$\beta^{43} \rightarrow (100303)$
$\beta^2 \rightarrow (000010)$	$\beta^{23} \rightarrow (032031)$	$\beta^{44} \rightarrow (012003)$
$\beta^3 \rightarrow (330000)$	$\beta^{24} \rightarrow (103320)$	$\beta^{45} \rightarrow (011120)$
$\beta^4 \rightarrow (003300)$	$\beta^{25} \rightarrow (221033)$	$\beta^{46} \rightarrow (220111)$
$\beta^5 \rightarrow (000033)$	$\beta^{26} \rightarrow (123210)$	$\beta^{47} \rightarrow (321201)$
$\beta^6 \rightarrow (121000)$	$\beta^{27} \rightarrow (331232)$	$\beta^{48} \rightarrow (032212)$
$\beta^7 \rightarrow (001210)$	$\beta^{28} \rightarrow (131312)$	$\beta^{49} \rightarrow (312322)$
$\beta^8 \rightarrow (330012)$	$\beta^{29} \rightarrow (313313)$	$\beta^{50} \rightarrow (201123)$
$\beta^9 \rightarrow (311300)$	$\beta^{30} \rightarrow (300133)$	$\beta^{51} \rightarrow (233011)$
$\beta^{10} \rightarrow (311300)$	$\beta^{31} \rightarrow (120001)$	$\beta^{52} \rightarrow (321330)$
$\beta^{11} \rightarrow (301031)$	$\beta^{32} \rightarrow (030200)$	$\beta^{53} \rightarrow (113213)$
$\beta^{12} \rightarrow (102010)$	$\beta^{33} \rightarrow (000302)$	$\beta^{54} \rightarrow (302132)$
$\beta^{13} \rightarrow (331020)$	$\beta^{34} \rightarrow (022003)$	$\beta^{55} \rightarrow (131021)$
$\beta^{14} \rightarrow (223310)$	$\beta^{35} \rightarrow (011220)$	$\beta^{56} \rightarrow (210310)$
$\beta^{15} \rightarrow (332233)$	$\beta^{36} \rightarrow (220112)$	$\beta^{57} \rightarrow (332103)$
$\beta^{16} \rightarrow (120322)$	$\beta^{37} \rightarrow (310201)$	$\beta^{58} \rightarrow (010321)$
$\beta^{17} \rightarrow (203203)$	$\beta^{38} \rightarrow (032102)$	$\beta^{59} \rightarrow (213103)$
$\beta^{18} \rightarrow (013032)$	$\beta^{39} \rightarrow (022321)$	$\beta^{60} \rightarrow (013131)$
$\beta^{19} \rightarrow (132130)$	$\beta^{40} \rightarrow (213223)$	$\beta^{61} \rightarrow (103131)$
$\beta^{20} \rightarrow (111321)$	$\beta^{41} \rightarrow (233132)$	$\beta^{62} \rightarrow (100031)$
$\beta^{21} \rightarrow (210113)$	$\beta^{42} \rightarrow (130331)$	$\beta^{63} \rightarrow (100000)$

Passo 7 - Determinar o polinômio gerador da matriz $G, g(x)$: serão calculados os polinômios geradores $g(x)$ das matrizes geradoras G dos códigos. Os polinômios geradores

dos códigos de comprimento n, tem como raízes os elementos na sequência,

$$\{(\beta^i), (\beta^i)^p, (\beta^i)^{p^2}, (\beta^i)^{p^3}, \cdots, (\beta^i)^{p^{r-1(mod\ n)}}\}.$$

Estes polinômios são dados por $g(x) = mmc(M_1(x), M_2(x), \cdots, M_{n-1}(x))$, onde $M_i(x)$ é o polinômio minimal associado ao elemento β_i, $\{i = 1, 2, \cdots, n-1\}$, sendo β um elemento primitivo em G_n.

Na palavra-código, cujo comprimento é $n = 63$, serão analisados os valores de $1 \leqslant t \leqslant 31$, sendo que para cada valor, teremos uma distância equivalente, bem como seus respectivos polinômios minimais envolvidos nos cálculos dos $g(x)$.

1º Cálculo das raízes dos polinômios minimais: para cada polinômio minimal $M_i(x) = M_i$, com $i = 1, 2, \cdots, 62$, temos:

$$\begin{aligned}
M_1(x) &= \{(\beta^1), (\beta^1)^2, \cdots, (\beta^1)^{2^{6-1(mod\ 63)}}\} \rightarrow M_1 = \{\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}\},\\
M_2(x) &= \{(\beta^2), (\beta^2)^2, \cdots, (\beta^2)^{2^{6-1(mod\ 63)}}\} \rightarrow M_2 = \{\beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta\},\\
M_3(x) &= \{(\beta^3), (\beta^3)^2, \cdots, (\beta^3)^{2^{6-1(mod\ 63)}}\} \rightarrow M_3 = \{\beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}\},\\
&\vdots\\
M_{62}(x) &= \{(\beta^{62}), (\beta^{62})^2, \cdots, (\beta^{62})^{2^{6-1(mod\ 63)}}\} \rightarrow M_{62} = \{\beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}\}.
\end{aligned}$$

As raízes dos polinômios minimais sobre G_{63} podem ser vistas na Tabela 11, observando-se que alguns possuem as mesmas raízes. Portanto, estes polinômios minimais são iguais, como mostrados na Tabela 12.

2º Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, \cdots, 62$: como alguns polinômios são iguais, já que possuem as mesmas raízes, são calculados com o auxílio do *Magma*:

$$M_1(x) = M_2(x) = M_4(x) = M_8(x) = M_{16}(x) = M_{32}(x) =$$
$$(x-\beta)(x-\beta^2)(x-\beta^4)(x-\beta^8)(x-\beta^{16})(x-\beta^{32}) = x^6 + 2x^3 + 3x + 1;$$

$$M_3(x) = M_6(x) = M_{12}(x) = M_{24}(x) = M_{48}(x) = M_{33}(x) =$$
$$(x-\beta^3)(x-\beta^6)(x-\beta^{12})(x-\beta^{24})(x-\beta^{33})(x-\beta^{48}) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1;$$

$$M_5(x) = M_{10}(x) = M_{17}(x) = M_{20}(x) = M_{34}(x) = M_{40}(x) =$$
$$(x-\beta^5)(x-\beta^{10})(x-\beta^{17})(x-\beta^{20})(x-\beta^{34})(x-\beta^{40}) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1;$$

$$M_7(x) = M_{14}(x) = M_{28}(x) = M_{35}(x) = M_{56}(x) = M_{49}(x) =$$
$$(x-\beta^7)(x-\beta^{14})(x-\beta^{28})(x-\beta^{35})(x-\beta^{49})(x-\beta^{56}) = x^6 + x^3 + 1;$$

$$M_9(x) = M_{18}(x) = M_{36}(x) = (x-\beta^9)(x-\beta^{18})(x-\beta^{36}) = x^3+3x^2+2x+3;$$

$$M_{11}(x) = M_{22}(x) = M_{25}(x) = M_{37}(x) = M_{44}(x) = M_{50}(x) =$$
$$(x-\beta^{11})(x-\beta^{22})(x-\beta^{25})(x-\beta^{37})(x-\beta^{44})(x-\beta^{50}) = x^6+3x^5+x^3+x^2+2x+1;$$

$$M_{13}(x) = M_{19}(x) = M_{26}(x) = M_{38}(x) = M_{41}(x) = M_{52}(x) =$$
$$(x-\beta^{13})(x-\beta^{19})(x-\beta^{26})(x-\beta^{38})(x-\beta^{41})(x-\beta^{52}) = x^6+2x^5+x^4+x^3+3x+1;$$

$$M_{15}(x) = M_{30}(x) = M_{39}(x) = M_{51}(x) = M_{57}(x) = M_{60}(x) =$$
$$(x-\beta^{15})(x-\beta^{30})(x-\beta^{39})(x-\beta^{51})(x-\beta^{57})(x-\beta^{60}) = x^6+x^5+3x^4+3x^2+2x+1;$$

$$M_{21}(x) = M_{42}(x) = (x-\beta^{21})(x-\beta^{42}) = x^2+x+1;$$

$$M_{23}(x) = M_{29}(x) = M_{43}(x) = M_{46}(x) = M_{53}(x) = M_{58}(x) =$$
$$(x-\beta^{23})(x-\beta^{29})(x-\beta^{43})(x-\beta^{46})(x-\beta^{53})(x-\beta^{58}) = x^6+x^5+x^4+2x^2+3x+1;$$

$$M_{27}(x) = M_{45}(x) = M_{54}(x) = (x-\beta^{27})(x-\beta^{45})(x-\beta^{54}) = x^3+2x^2+x+3;$$

$$M_{31}(x) = M_{47}(x) = M_{55}(x) = M_{59}(x) = M_{61}(x) = M_{62}(x) =$$
$$(x-\beta^{31})(x-\beta^{47})(x-\beta^{55})(x-\beta^{59})(x-\beta^{61})(x-\beta^{62}) = x^6+3x^5+2x^3+1;$$

Tabela 11 – Raízes dos polinômios minimais sobre G_{63}.

$M_1 = (\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32})$	$M_{22} = (\beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}, \beta^{11})$	$M_{43} = (\beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}, \beta^{53})$
$M_2 = (\beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta)$	$M_{23} = (\beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}, \beta^{43})$	$M_{44} = (\beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}, \beta^{22})$
$M_3 = (\beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33})$	$M_{24} = (\beta^{24}, \beta^{48}, \beta^{33}, \beta^3, \beta^6, \beta^{12})$	$M_{45} = (\beta^{45}, \beta^{27}, \beta^{54})$
$M_4 = (\beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta, \beta^2)$	$M_{25} = (\beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}, \beta^{44})$	$M_{46} = (\beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}, \beta^{23})$
$M_5 = (\beta^5, \beta^{10}, \beta^{20}, \beta^{40}, \beta^{17}, \beta^{34})$	$M_{26} = (\beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}, \beta^{13})$	$M_{47} = (\beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}, \beta^{55})$
$M_6 = (\beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}, \beta^3)$	$M_{27} = (\beta^{27}, \beta^{54}, \beta^{45})$	$M_{48} = (\beta^{48}, \beta^{33}, \beta^3, \beta^6, \beta^{12}, \beta^{24})$
$M_7 = (\beta^7, \beta^{14}, \beta^{28}, \beta^{56}, \beta^{49}, \beta^{35})$	$M_{28} = (\beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}, \beta^7, \beta^{14})$	$M_{49} = (\beta^{49}, \beta^{35}, \beta^7, \beta^{14}, \beta^{28}, \beta^{56})$
$M_8 = (\beta^8, \beta^{16}, \beta^{32}, \beta\cdot\beta^2, \beta^4)$	$M_{29} = (\beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}, \beta^{46})$	$M_{50} = (\beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}, \beta^{25})$
$M_9 = (\beta^9, \beta^{18}, \beta^{36})$	$M_{30} = (\beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}, \beta^{15})$	$M_{51} = (\beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}, \beta^{57})$
$M_{10} = (\beta^{10}, \beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}, \beta^5)$	$M_{31} = (\beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47})$	$M_{52} = (\beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}, \beta^{26})$
$M_{11} = (\beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}, \beta^{37})$	$M_{32} = (\beta^{32}, \beta, \beta^2, \beta^4, \beta^8, \beta^{16})$	$M_{53} = (\beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}, \beta^{58})$
$M_{12} = (\beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}, \beta^3, \beta^6)$	$M_{33} = (\beta^{33}, \beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48})$	$M_{54} = (\beta^{54}, \beta^{45}, \beta^{27})$
$M_{13} = (\beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}, \beta^{38})$	$M_{34} = (\beta^{34}, \beta^5, \beta^{10}, \beta^{20}, \beta^{40}, \beta^{17})$	$M_{55} = (\beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}, \beta^{59})$
$M_{14} = (\beta^{14}, \beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}, \beta^7)$	$M_{35} = (\beta^{35}, \beta^7, \beta^{14}, \beta^{28}, \beta^{56}, \beta^{49})$	$M_{56} = (\beta^{56}, \beta^{49}, \beta^{35}, \beta^7, \beta^{14}, \beta^{28})$
$M_{15} = (\beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}, \beta^{39})$	$M_{36} = (\beta^{36}, \beta^9, \beta^{18})$	$M_{57} = (\beta^{57}, \beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}, \beta^{60})$
$M_{16} = (\beta^{16}, \beta^{32}, \beta, \beta^2, \beta^4, \beta^8)$	$M_{37} = (\beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}, \beta^{50})$	$M_{58} = (\beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}, \beta^{29})$
$M_{17} = (\beta^{17}, \beta^{34}, \beta^5, \beta^{10}, \beta^{20}, \beta^{40})$	$M_{38} = (\beta^{38}, \beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}, \beta^{19})$	$M_{59} = (\beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}, \beta^{61})$
$M_{18} = (\beta^{18}, \beta^{36}, \beta^9)$	$M_{39} = (\beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}, \beta^{51})$	$M_{60} = (\beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}, \beta^{15}, \beta^{30})$
$M_{19} = (\beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}, \beta^{16}, \beta^{41})$	$M_{40} = (\beta^{40}, \beta^{17}, \beta^{34}, \beta^5, \beta^{10}, \beta^{20})$	$M_{61} = (\beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}, \beta^{62})$
$M_{20} = (\beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}, \beta^{16}, \beta^{10})$	$M_{41} = (\beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}, \beta^{52})$	$M_{62} = (\beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31})$
$M_{21} = (\beta^{21}, \beta^{42})$	$M_{42} = (\beta^{42}, \beta^{21})$	

Tabela 12 – Polinômios minimais que são iguais sobre G_{63}.

$M_1 = M_2 = M_4 = M_8 = M_{16} = M_{32}$	$M_{13} = M_{19} = M_{26} = M_{38} = M_{41} = M_{52}$
$M_3 = M_6 = M_{12} = M_{24} = M_{33} = M_{48}$	$M_{15} = M_{30} = M_{39} = M_{51} = M_{57} = M_{60}$
$M_5 = M_{10} = M_{17} = M_{20} = M_{34} = M_{40}$	$M_{21} = M_{42}$
$M_7 = M_{14} = M_{28} = M_{35} = M_{56} = M_{49}$	$M_{23} = M_{29} = M_{43} = M_{46} = M_{53} = M_{58}$
$M_9 = M_{18} = M_{36}$	$M_{27} = M_{45} = M_{54}$
$M_{11} = M_{22} = M_{25} = M_{37} = M_{44} = M_{50}$	$M_{31} = M_{47} = M_{55} = M_{59} = M_{61} = M_{62}$

3º Cálculo dos polinômios geradores para $1 \leqslant t \leqslant 31$: o polinômio gerador para cada valor de t é dado por

$$g(x) = mmc(M_1(x), M_2(x), \cdots, M_{n-1}(x)),$$

formado pelos polinômios minimais que são distintos entre si e possuem raízes $\beta \cdots \beta^{2t}$, como pode ser visto na Tabela 13.

Tabela 13 – Polinômios geradores.

t	$d_H \geqslant 2t+1$	$(\beta^1, ..., \beta^{2t})$	$g(x) = MMC(M_1(x), ..., M_{n-1}(x))$	$C(n, k, d_H)$
1	$d_H \geqslant 3$	(β^1, β^2)	$g(x) = MMC(M_1)$	$C(63, 57, 03)$
2	$d_H \geqslant 5$	$(\beta^1, \beta^2, \beta^3, \beta^4)$	$g(x) = MMC(M_1, M_3)$	$C(63, 51, 05)$
3	$d_H \geqslant 7$	$(\beta^1, ..., \beta^6)$	$g(x) = MMC(M_1, M_3, M_5)$	$C(63, 45, 07)$
4	$d_H \geqslant 9$	$(\beta^1, ..., \beta^8)$	$g(x) = MMC(M_1, M_3, M_5, M_7)$	$C(63, 39, 09)$
5	$d_H \geqslant 11$	$(\beta^1, ..., \beta^{10})$	$g(x) = MMC(M_1, M_3, M_5, M_7, M_9)$	$C(63, 36, 11)$
6	$d_H \geqslant 13$	$(\beta^1, ..., \beta^{12})$	$g(x) = MMC(M_1, M_3, M_5, M_7, M_9, M_{11})$	$C(63, 30, 13)$
7	$d_H \geqslant 15$	$(\beta^1, ..., \beta^{14})$	$g(x) = MMC(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13})$	$C(63, 24, 15)$
8	$d_H \geqslant 17$	$(\beta^1, ..., \beta^{16})$	$g(x) = MMC(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15})$	$C(63, 18, 17)$
9	$d_H \geqslant 19$	$(\beta^1, ..., \beta^{18})$	$g(x) = MMC(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15})$	$C(63, 18, 19)$
10	$d_H \geqslant 21$	$(\beta^1, ..., \beta^{20})$	$g(x) = MMC(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15})$	$C(63, 18, 21)$
11	$d_H \geqslant 23$	$(\beta^1, ..., \beta^{22})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{13}, M_{15}, M_{21})$	$C(63, 16, 23)$
12	$d_H \geqslant 25$	$(\beta^1, ..., \beta^{24})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{15}, M_{21}, M_{23})$	$C(63, 10, 25)$
13	$d_H \geqslant 27$	$(\beta^1, ..., \beta^{26})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{15}, M_{21}, M_{23})$	$C(63, 10, 27)$
14	$d_H \geqslant 29$	$(\beta^1, ..., \beta^{28})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{21}, M_{23}, M_{27})$	$C(63, 07, 29)$
15	$d_H \geqslant 31$	$(\beta^1, ..., \beta^{30})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{21}, M_{23}, M_{27})$	$C(63, 07, 31)$
16	$d_H \geqslant 33$	$(\beta^1, ..., \beta^{32})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 33)$
17	$d_H \geqslant 35$	$(\beta^1, ..., \beta^{34})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 35)$
18	$d_H \geqslant 37$	$(\beta^1, ..., \beta^{36})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 37)$
19	$d_H \geqslant 39$	$(\beta^1, ..., \beta^{38})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 39)$
20	$d_H \geqslant 41$	$(\beta^1, ..., \beta^{40})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 41)$
21	$d_H \geqslant 43$	$(\beta^1, ..., \beta^{42})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 43)$
22	$d_H \geqslant 45$	$(\beta^1, ..., \beta^{44})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 45)$
23	$d_H \geqslant 47$	$(\beta^1, ..., \beta^{46})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 47)$
24	$d_H \geqslant 49$	$(\beta^1, ..., \beta^{48})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 49)$
25	$d_H \geqslant 51$	$(\beta^1, ..., \beta^{50})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 51)$
26	$d_H \geqslant 53$	$(\beta^1, ..., \beta^{52})$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 53)$
27	$d_H \geqslant 55$	$(\beta^1, ..., \beta 54)$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 55)$
28	$d_H \geqslant 57$	$(\beta^1, ..., \beta 56)$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 57)$
29	$d_H \geqslant 59$	$(\beta^1, ..., \beta 58)$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 59)$
30	$d_H \geqslant 61$	$(\beta^1, ..., \beta 60)$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 61)$
31	$d_H \geqslant 63$	$(\beta^1, ..., \beta 62)$	$g(x) = MMC(M_1, M_3, M_5, \cdots, M_{23}, M_{27}, M_{31})$	$C(63, 01, 63)$

Considerando que a distância mínima do código seja $d_H = 3$, então o polinômio gerador do código é dado por $g_1(x) = x^6 + 2x^3 + 3x + 1$, que gera o código desejado e está associado à matriz geradora G do código BCH sobre $\mathbb{Z}_4$ com parâmetros $(n, k, d_H) = (63, 57, 3)$.

Passo 8 - Determinar o polinômio gerador da matriz $H, h(x)$: o polinômio gerador $h(x)$, cujos coeficientes pertencem a $\mathbb{Z}_4$, da matriz verificação de paridade H é obtido através da relação:

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

$$h(x) = x^{57} + 2x^{54} + x^{52} + 3x^{51} + x^{47} + 2x^{46} + x^{45} + 2x^{44} + 3x^{42} + x^{41} + 3x^{40} + 3x^{39} + x^{37} + 2x^{35} + 2x^{34} + x^{33} + x^{32} + 3x^{31} + 2x^{29} + x^{28} + 2x^{26} + 3x^{25} + 2x^{24} + 3x^{23} + x^{22} + 2x^{21} + 3x^{20} + x^{19} + x^{18} + 3x^{16} + 3x^{15} + 2x^{14} + 2x^{13} + x^{12} + 3x^{11} + x^9 + 2x^8 + 3x^7 + x^5 + 3x^4 + x^3 + 3x^2 + 3x + 3$$

Passo 9 - Determinar a matriz G e a sua transposta G^T: o polinômio gerador $g_1(x) = x^6 + 2x^3 + 3x + 1$ está relacionado à matriz geradora G. Ao realizar os deslocamentos dos coeficientes do polinômio $g(x)$ da esquerda para a direita, obtemos a matriz G com dimensão 57×63:

$$G = \begin{bmatrix} 130200100000000000000000000000000000000000000000000000000000000 \\ 013020010000000000000000000000000000000000000000000000000000000 \\ 001302001000000000000000000000000000000000000000000000000000000 \\ 000130200100000000000000000000000000000000000000000000000000000 \\ \vdots \\ 000000000000000000000000000000000000000000000000000001302001000 \\ 000000000000000000000000000000000000000000000000000000130200100 \\ 000000000000000000000000000000000000000000000000000000013020010 \\ 000000000000000000000000000000000000000000000000000000001302001 \end{bmatrix}$$

Sua transposta G^T de dimensão 63×57 é obtida trocando-se as linhas por colunas.

Passo 10 - Determinar a matriz H e a sua transposta H^T: a matriz H de dimensão 6×63 é obtida através do deslocamento dos coeficientes da direita para a esquerda do polinômio gerador $h(x)$ determinado no **Passo 8**:

$$H = \begin{bmatrix} 000001002013000121203133010221130210232312311033221301230131333 \\ 000010020130001212031330102211302102323123110332213012301313330 \\ 000100201300012120313301022113021023231231103322130123013133300 \\ 001002013000121203133010221130210232312311033221301230131333000 \\ 010020130001212031330102211302102323123110332213012301313330000 \\ 100201300012120313301022113021023231231103322130123013133300000 \end{bmatrix}$$

Sua transposta H^T de dimensão 63×3 é obtida trocando-se as linhas por colunas.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1: o código BCH primitivo sobre $GR(4, 6)$ será analisado se é capaz de reproduzir a sequência de sinal interno ($SI = Seq.36$) de uma proteína mitocondrial - GI número 832917, com um comprimento de $n = 63$ nucleotídeos. Serão determinadas as 24 permutações entre o alfabeto do código genético $N = \{A, C, G, T/U\}$ e o alfabeto do código BCH $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ da sequência de DNA a ser analisada. Dado que o mapeamento entre $N \to \mathbb{Z}_4$ é desconhecido, consideramos todas as permutações entre esses dois conjuntos. As 24 linhas da matriz P correspondem às 24 permutações da SI. Seja a SI do $NCBI$ igual a

$$\{5' - GCCGTTCATGTTTACTCTGGGTTGCCTTGGTGGGGAACTATCGCGGCCACCACCATCCTCATT - 3'\},$$

então

$$P = \begin{bmatrix}
211233103233301313222332113322322220013031212211011011031131033 \\
311322102322201212333223112233233330012021313311011011021121022 \\
122133203133302323111331223311311110023032121122022022032232033 \\
322311201311102121333113221133133330021012323322022022012212011 \\
233211301211103131222112331122122220031013232233033033013313011 \\
133122302122203232111221332211211110032023131133033033023323022 \\
200233013233310303222332003322322221103130202200100100130030133 \\
300322012322210202333223002233233331102120303300100100120020122 \\
022033213033312323000330223300300001123132020022122122132232133 \\
322300210300012020333003220033033331120102323322122122102202100 \\
033022312022213232000220332200200001132123030033133133123323122 \\
233200310200013030222002330022022221130103232233133133103303100 \\
100133023133320303111331003311311112203230101100200200230030233 \\
300311021311120101333113001133133332201210303300200200210010211 \\
011033123033321313000330113300300002213231010011211211231131233 \\
311300120300021010333003110033033332210201313311211211201101200 \\
033011321011123131000110331100100002231213030033233233213313211 \\
133100320100023030111001330011011112230203131133233233203303200 \\
100122032122230202111221002211211113302320101100300300320020322 \\
200211031211130101222112001122122223301310202200300300310010311 \\
011022132022231212000220112200200003312321010011311311321121322 \\
211200130200031010222002110022022223310301212211311311301101300 \\
022011231011132121000110221100100003321312020022322322312212311 \\
122100230100032020111001220011011113320302121122322322302202300
\end{bmatrix}$$

As linhas da matriz P estão associadas às 24 permutações entre $N \to \mathbb{Z}_4$, onde cada uma dessas permutações foi estabelecida como um caso, conforme ilustrado na Tabela 14.

Tabela 14 – Relação entre as linhas da matriz P e as 24 permutações

$Linha = Caso$	$N \to Z_4$	$Linha = Caso$	$N \to Z_4$
$L1 = Caso\ 01$	$(A,C,G,T) = (0,1,2,3)$	$L13 = Caso\ 13$	$(A,C,G,T) = (2,0,1,3)$
$L2 = Caso\ 02$	$(A,C,G,T) = (0,1,3,2)$	$L14 = Caso\ 14$	$(A,C,G,T) = (2,0,3,1)$
$L3 = Caso\ 03$	$(A,C,G,T) = (0,2,1,3)$	$L15 = Caso\ 15$	$(A,C,G,T) = (2,1,0,3)$
$L4 = Caso\ 04$	$(A,C,G,T) = (0,2,3,1)$	$L16 = Caso\ 16$	$(A,C,G,T) = (2,1,3,0$
$L5 = Caso\ 05$	$(A,C,G,T) = (0,3,2,1)$	$L17\ = Caso\ 17$	$(A,C,G,T) = (2,3,0,1)$
$L6 = Caso\ 06$	$(A,C,G,T) = (0,3,1,2)$	$L18\ = Caso\ 18$	$(A,C,G,T) = (2,3,1,0)$
$L7 = Caso\ 07$	$(A,C,G,T) = (1,0,2,3)$	$L19\ = Caso\ 19$	$(A,C,G,T) = (3,0,1,2)$
$L8 = Caso\ 08$	$(A,C,G,T) = (1,0,3,2)$	$L20\ = Caso\ 20$	$(A,C,G,T) = (3,0,2,1)$
$L9 = Caso\ 09$	$(A,C,G,T) = (1,2,0,3)$	$L21\ = Caso\ 21$	$(A,C,G,T) = (3,1,0,2)$
$L10 = Caso\ 10$	$(A,C,G,T) = (1,2,3,0)$	$L22\ = Caso\ 22$	$(A,C,G,T) = (3,1,2,0)$
$L11 = Caso\ 11$	$(A,C,G,T) = (1,3,0,2)$	$L23\ = Caso\ 23$	$(A,C,G,T) = (3,2,0,1)$
$L12 = Caso\ 12$	$(A,C,G,T) = (1,3,2,0)$	$L24\ = Caso\ 24$	$(A,C,G,T) = (3,2,1,0)$

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

utilizado para determinar quais das sequências (com 0 e 1 nucleotídeo de diferença da sequência original) são palavras-código dos códigos $(63, k, d_H)$, segue a seguinte abordagem:

As linhas da matriz geradora do código (n, k, d_H) do **Passo 9** formam uma base do espaço vetorial conhecido como o código linear C. Portanto, as combinações lineares das linhas de G resultam em palavras-código de C. Assim, o processo de codificação pode ser expresso como $v = u \cdot G$, onde u representa a informação e v é a palavra-código correspondente, neste caso, as sequências de DNA a serem analisadas. Para todas as palavras-código v, a relação a seguir é válida: $v \cdot H^T = 0$.

A capacidade de correção de erros de um código está relacionada ao número de palavras-código: no caso em questão, temos 4^k palavras-código, onde $k = n - r$. É importante observar que quanto maior o valor de k, maior será o número de palavras-código, o que implica em uma maior complexidade computacional para gerar todas as 4^k palavras-código. Para lidar com esse problema, ao invés de gerar todas as palavras-código para comparar com a sequência de DNA, consideramos que a sequência sob a aplicação de cada uma das 24 permutações do **Passo 11** é uma palavra-código, correspondendo ao padrão de erro denotado por $D(a, b) = 0$, ou seja, a diferença entre a palavra-código e a sequência de DNA do *NCBI* é nula. Assim, para determinar se cada uma dessas 24 possibilidades é de fato uma palavra-código, usamos a relação $v \cdot H^T = 0$, onde v é a possível palavra-código e H^T é a transposta da matriz de verificação de paridade determinada no **Passo 10**.

Neste mesmo passo, determinaremos quais das sequências de DNA apresentando o padrão de erro $D(a, b) = 1$ nucleotídeo de diferença da sequência de DNA do *NCBI* são palavras-código dos códigos (n, k, d_H).

a) $D(a, b) = 1$ **nucleotídeo de diferença**

Para analisar as sequências de DNA com até 1 nucleotídeo de diferença da sequência de DNA do $NCBI$, representada por $D(a, b) = 1$, consideramos as outras três possibilidades de nucleotídeos em cada posição na sequência de DNA para cada permutação. Isso resulta em um total de três possibilidades de nucleotídeo em cada posição, multiplicado pelo comprimento da sequência n e pelas 24 possibilidades de permutações. Portanto, para cada sequência de DNA, temos $3 \cdot 63 \cdot 24 = 4536$ possibilidades. Utilizando a relação $v \cdot H^T = 0$, as palavras-código identificadas são então armazenadas.

Como resultado da geração com 1 nucleotídeo de diferença $(D(a, b) = 1)$, obtemos a matriz R, onde cada linha representa uma palavra-código encontrada.

$$R = \left[\begin{array}{l}
12213320313330232311133122311131111002303212112202202203223203 3 \\
32231120131110212133311322133313333002101232332202202201221201 1 \\
03302231202221323200022033200020000113212303003313313312332312 2 \\
23320031020001303022200233022202222113010323223313313310330310 0 \\
10013302313332030311133100311131111220323010110020020023003023 3 \\
30031102131112010133311300133313333220121030330020020021001021 1 \\
01102213202223121200022011200020000331232101001131131132112132 2 \\
21120013020003101022200211022202222331030121221131131130110130 0
\end{array}\right]$$

Para encontrar a matriz R, foi utilizado o *software Magma*, seguindo a seguinte rotina:

1. Determinação da matriz H e sua transposta:

>H:=Matrix(IntegerRing(),6,63),
[0, 0, 0, 0, 0, 1, 0, 0, 2, 0, 1, 3, 0, 0, 0, 1, 2, 1, 2, 0, 3, 1, 3, 3, 0, 1, 0, 2, 2, 1, 1, 3, 0, 2, 1, 0, 2, 3, 2, 3, 1, 2, 3, 1, 1, 0, 3, 3, 2, 2,
1, 3, 0, 1, 2, 3, 0, 1, 3, 1, 3, 3, 3,
0, 0, 0, 0, 1, 0, 0, 2, 0, 1, 3, 0, 0, 0, 1, 2, 1, 2, 0, 3, 1, 3, 3, 0, 1, 0, 2, 2, 1, 1, 3, 0, 2, 1, 0, 2, 3, 2, 3, 1, 2, 3, 1, 1, 0, 3, 3, 2, 2, 1, 3,
0, 1, 2, 3, 0, 1, 3, 1, 3, 3, 3, 0,
0, 0, 0, 1, 0, 0, 2, 0, 1, 3, 0, 0, 0, 1, 2, 1, 2, 0, 3, 1, 3, 3, 0, 1, 0, 2, 2, 1, 1, 3, 0, 2, 1, 0, 2, 3, 2, 3, 1, 2, 3, 1, 1, 0, 3, 3, 2, 2, 1, 3, 0,
1, 2, 3, 0, 1, 3, 1, 3, 3, 3, 0, 0,
0, 0, 1, 0, 0, 2, 0, 1, 3, 0, 0, 0, 1, 2, 1, 2, 0, 3, 1, 3, 3, 0, 1, 0, 2, 2, 1, 1, 3, 0, 2, 1, 0, 2, 3, 2, 3, 1, 2, 3, 1, 1, 0, 3, 3, 2, 2, 1, 3, 0, 1,
2, 3, 0, 1, 3, 1, 3, 3, 3, 0, 0, 0,
0, 1, 0, 0, 2, 0, 1, 3, 0, 0, 0, 1, 2, 1, 2, 0, 3, 1, 3, 3, 0, 1, 0, 2, 2, 1, 1, 3, 0, 2, 1, 0, 2, 3, 2, 3, 1, 2, 3, 1, 1, 0, 3, 3, 2, 2, 1, 3, 0, 1, 2,
3, 0, 1, 3, 1, 3, 3, 3, 0, 0, 0, 0,
1, 0, 0, 2, 0, 1, 3, 0, 0, 0, 1, 2, 1, 2, 0, 3, 1, 3, 3, 0, 1, 0, 2, 2, 1, 1, 3, 0, 2, 1, 0, 2, 3, 2, 3, 1, 2, 3, 1, 1, 0, 3, 3, 2, 2, 1, 3, 0, 1, 2, 3,
0, 1, 3, 1, 3, 3, 3, 0, 0, 0, 0, 0]);

> $Transpose(H)$;

2. Determinação da palavra-código v:

   No exemplo apresentado a seguir é testada a terceira linha da matriz P trocando-se o nucleotídeo 3 da 28^a posição por 1, ou seja, Guanina no lugar de Timina:

> $v := Matrix(IntegerRing(), 1, 63)$,
[1, 2, 2, 1, 3, 3, 2, 0, 3, 1, 3, 3, 3, 0, 2, 3, 2, 3, 1, 1, 1, 3, 3, 1, 2, 2, 3, 1, 1, 1, 3, 1, 1, 1, 1, 0, 0, 2, 3, 0, 3, 2, 1, 2, 1, 1, 2, 2, 0, 2, 2,
0, 2, 2, 0, 3, 2, 2, 3, 2, 0, 3, 3]);

> $v * Transpose(H)$;

   Ao realizar a multiplicação de v por H^T, obtemos o seguinte resultado:

   [152 156 136 144 136 152]

Como a operação é módulo 4, este é o vetor nulo [000000] = 0, portanto é palavra-código e será armazenada na primeira linha da matriz R.

O processo é repetido para todas as linhas da matriz P, trocando-se as bases nitrogenadas e realizando-se a relação $v \cdot H^T$, ou seja, 4536 operações, a fim de se encontrar as oito palavras-código armazenadas em R.

Passo 13 - Comparar todas as palavras-código armazenadas no Passo 12 com a sequência de DNA do *NCBI* e mostrar onde os erros ocorreram: todas as palavras-código armazenadas na etapa anterior estão rotuladas no alfabeto do código $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, e serão convertidas em nucleotídeos usando o alfabeto do código genético $N = \{A, C, G, T\}$. Após a conversão, as palavras-código são comparadas, uma a uma, com a sequência de DNA original, mostrando onde os nucleotídeos diferem e armazenando os resultados.

Passo 14 - Voltar ao Passo 7 e determinar outro $g(x)$: a seguir, mostram-se outros polinômios geradores e seus respectivos polinômios primitivos, que podem ser utilizados para construir códigos com parâmetros $(63, 57, 3)$:

- $p_{01}(x) = x^6 + x + 1$
  - $g_{01}(x) = x^6 + 2x^3 + 3x + 1$
- $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$
  - $g_{02}(x) = x^6 + 2x^5 + x^4 + x^3 + 3x + 1$
- $p_{03}(x) = x^6 + x^5 + 1$
  - $g_{03}(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$
- $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$
  - $g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$
- $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$
  - $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$
- $p_{06}(x) = x^6 + x^5 + x^4 + x + 1$
  - $g_{06}(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$

Passo 15 - Repetir os Passos 8 ao 12 para o $g(x)$ **obtido no Passo 14, até que se esgote todas as possibilidades de** $g(x)$.

Passo 16 - Voltar ao Passo 3 e escolher outro $p(x)$ **para então repetir os Passos 4 ao 14 até que se esgote todos os** $p(x)$ **do Passo 3**.

Passo 17 - Fim.

A partir do caso analisado, com a utilização dos passos do algoritmo, pode-se observar na Tabela 15 que as oito palavras-código obtidas e armazenadas na matriz R são distintas em termos do alfabeto do código $Z_4 = \{0, 1, 2, 3\}$, porém são as mesmas quando rotuladas através do alfabeto do código genético $N = \{A, C, G, T\}$, ou seja, resultam em uma única sequência de DNA, destacando-se ainda que, nas oito palavras-código, a diferença do dígito correspondente ao nucleotídeo da sequência de DNA ocorre na mesma posição.

Tabela 15 – As 24 permutações para $D(a,b) = 1$.

24 casos de permutações	**Palavras-código** $n = 63$	**Quantidade de palavras-código**	**Sequência de DNA reproduzida**
01: $(A,C,G,T) = (0,1,2,3)$	$00000 \cdots 00000$	0	-
02: $(A,C,G,T) = (0,1,3,2)$	$00000 \cdots 00000$	0	-
03: $(A,C,G,T) = (0,2,1,3)$	$122133 \cdots 232033$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
04: $(A,C,G,T) = (0,2,3,1)$	$322311 \cdots 212011$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
05: $(A,C,G,T) = (0,3,2,1)$	$00000 \cdots 00000$	0	-
06: $(A,C,G,T) = (0,3,1,2)$	$00000 \cdots 00000$	0	-
07: $(A,C,G,T) = (1,0,2,3)$	$00000 \cdots 00000$	0	-
08: $(A,C,G,T) = (1,0,3,2)$	$00000 \cdots 00000$	0	-
09: $(A,C,G,T) = (1,2,0,3)$	$00000 \cdots 00000$	0	-
10: $(A,C,G,T) = (1,2,3,0)$	$00000 \cdots 00000$	0	-
11: $(A,C,G,T) = (1,3,0,2)$	$033022 \cdots 323122$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
12: $(A,C,G,T) = (1,3,2,0)$	$233200 \cdots 303100$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
13: $(A,C,G,T) = (2,0,1,3)$	$100133 \cdots 030233$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
14: $(A,C,G,T) = (2,0,3,1)$	$300311 \cdots 010211$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
15: $(A,C,G,T) = (2,1,0,3)$	$00000 \cdots 00000$	0	-
16: $(A,C,G,T) = (2,1,3,0$	$00000 \cdots 00000$	0	-
17: $(A,C,G,T) = (2,3,0,1)$	$00000 \cdots 00000$	0	-
18: $(A,C,G,T) = (2,3,1,0)$	$00000 \cdots 00000$	0	-
19: $(A,C,G,T) = (3,0,1,2)$	$00000 \cdots 00000$	0	-
20: $(A,C,G,T) = (3,0,2,1)$	$00000 \cdots 00000$	0	-
21: $(A,C,G,T) = (3,1,0,2)$	$011022 \cdots 121322$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
22: $(A,C,G,T) = (3,1,2,0)$	$211200 \cdots 101300$	1	GCCGTTCATGTTTACTCTGGGTTGCCTG $\cdots$
23: $(A,C,G,T) = (3,2,0,1)$	$00000 \cdots 00000$	0	-
24: $(A,C,G,T) = (3,2,1,0)$	$00000 \cdots 00000$	0	-

Assim, basta analisar uma sequência, como o caso 03 mostrado na Tabela 16, em que *aaO* são os aminoácidos originais, *NtO* os nucleotídeos originais, *RtO* o rotulamento original, *RtG* o rotulamento gerado, *NtG* os nucleotídeos gerados e *aaG* os aminoácidos gerados. Percebe-se que houve uma substituição de nucleotídeo na décima trinca, resultando na mudança do aminoácido nessa posição, onde o Triptofano (Tri) foi trocado por Glicina (Gli), ou seja, houve uma mutação de ponto de sentido trocado, porém não sabemos se, biologicamente, as sequências com alterações nos aminoácidos continuarão desempenhando suas respectivas funções.

Tabela 16 – Sequência de DNA de sinal interno com 63 nucleotídeos e $D(a,b) = 1$.

aaO	Ala	Val	His	Val	Tir	Ser	Gli	Leu	Pro	Tri	Tri	Gli	The	Ile	Ala	Ala	The	The	Ile	Leu	Ile
NtO	GCC	GTT	CAT	GTT	TAC	TCT	GGG	TTG	CCT	TGG	TGG	GGA	ACT	ATC	GCG	GCC	ACC	ACC	ATC	CTC	ATT
RtO	122	133	203	133	302	323	111	331	223	311	311	110	023	032	121	122	022	022	032	232	033
RtG	122	133	203	133	302	323	111	331	223	111	311	110	023	032	121	122	022	022	032	232	033
NtG	GCC	GTT	CAT	GTT	TAC	TCT	GGG	TTG	CCT	GGG	TGG	GGA	ACT	ATC	GCG	GCC	ACC	ACC	ATC	CTC	ATT
aaG	Ala	Val	His	Val	Tir	Ser	Gli	Leu	Pro	Gli	Tri	Gli	The	Ile	Ala	Ala	The	The	Ile	Leu	Ile

Dessa forma, percebemos que as sequências de DNA têm uma estrutura matemática, o que permite que sejam identificadas, reproduzidas e classificadas por meio de códigos corretores de erros. O *software Magma*, que possui uma linguagem de programação própria, com uma interface que permite a definição e manipulação de estruturas matemáticas, possibilitou trabalhar com os anéis dos números inteiros, matrizes e multiplicação de vetores por matrizes, poupando tempo nas operações.

4 Conclusão

Os códigos corretores de erros desempenham um papel fundamental na detecção de erros que, em muitos casos, podem ser corrigidos durante o processo de transmissão da informação em um sistema de comunicação. Estes erros podem comprometer o recebimento correto da informação enviada. Os códigos BCH mostraram-se significativos, pela simplicidade nos processos de codificação e decodificação, apesar da limitação relacionada aos comprimentos das sequências.

Com a implementação do algoritmo foi possível identificar e reproduzir uma sequência de DNA com 63 nucleotídeos, resultando em palavras-código com um nucleotídeo de diferença da sequência original. O *software Magma* foi fundamental para a realização das operações matemáticas mais complexas, que apesar de demandar esforço, reduziu significativamente o tempo empregado em comparação ao método manual.

Observou-se que os códigos corretores de erros são eficientes na localização de mutações na molécula de DNA, porém estudos biológicos são necessários para analisar consequências resultantes da troca de bases nitrogenadas nas sequências de DNA. Em pesquisas futuras, sequências mais longas poderão ser exploradas, permitindo a análise das características biológicas e das possíveis consequências resultantes das mutações.

Com o estudo de conceitos fundamentais de álgebra e dos códigos corretores de erros, bem como elementos de biologia, foi possível complementar a formação matemática, contemplando áreas fora da grade curricular, possibilitando uma abordagem interdisciplinar e ampliando conexões entre diferentes campos do conhecimento. Além disso, possibilitou a participação em eventos acadêmicos de destaque, como o Congresso Nacional de Matemática Aplicada e Computacional (CNMAC), realizado em setembro de 2023, no Centro de Convenções de Bonito/MS, onde um resumo do trabalho intitulado “Análise de mutações genéticas e genômicas através de códigos corretores de erros” foi apresentado e publicado nos Anais do evento, ampliando assim o alcance e o impacto do estudo realizado.

Referências

1 SHANNON, C.E.; **A Mathematical Theory of Communications**. BSTJ 27, 1948.

2 GOLAY, M.J.E. Notes on digital coding. **Proceedings of the I. R. E.** (IEEE), v. 27, p. 112-156, 1948.

3 MILIES, C.P. **Introdução à Teoria dos Códigos Corretores de Erros**. In: COLÓQUIO DE MATEMÁTICA DA REGIÃO CENTRO-OESTE. Campo Grande, Departamento de Matemática, 2009.

4 FARIA, L. C. B. **Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA**. Tese (Doutorado) — Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, Campinas, São Paulo - Brasil, 2011.

5 OLIVEIRA, M. Equações da Vida – A mesma estrutura de códigos une sequências de DNA e comunição digital. **Revista Pesquisa Fapesp**, ed.178, dez. 2010.

6 ROCHA, A.S.L. **Modelo de Sistema de Comunicações Digital para o Mecanismo de Importação de Proteínas Mitocondriais Através de Códigos Corretores de Erros**, 2010. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2010.

7 MAGMA. **Magma:** Computer - Algebra. Disponível em: <http://magma.maths.usyd.edu.au/calc/> Acesso em: 13 de janeiro de 2024.

8 ALBERTS, B.; JOHNSON, A.; LEWIS, J., et al. **Biologia molecular da célula**. Artmed, 5ª Ed. pág. 1485-1524, 2008.

9 BATTAIL, G. An Outline of Informational Genetics. **Morgan & Claypool Publishers**, 2008.

10 CARVALHO, H. F.; RECCO-PIMENTEL, S.M. **A célula**. Manole Ltda, 2001.

11 WATSON; BAKER; BELL, et al. **Biologia Molecular do gene**. Artmed, 5ª Ed., 2006.

12 BOLDRINI, J. L. **Álgebra Linear**. São Paulo: Harber ltda,1986.

13 DIAS, I. **Teoria de Anéis - Notas de Aulas**, ICMC - USP, 2001, Disponível em: <https://sites.icmc.usp.br/iresdias/material/sma306.pdf>

14 SAMUEL, P. **Algebraic Theory of Numbers**, Herman, Paris, 1967.

15 DOMINGUES, H. H. **Álgebra moderna**. 4. ed. ref. São Paulo: Atual, 2003.

16 ABRANTES, S. A. Códigos Corretores de Erros em Comunicações digitais. **FEUP edições**. P. 135-171, Março. 2010.

17 LIN, S.; COSTELLO JR., D. J. **Error Control Coding**. 2 ed. Prentice Hall, 2004.

18 INTERLANDO, J.C.; PALAZZO Jr., R.; GERÔNIMO, J.R.; et al. **Códigos Corretores de Erros sobre Estruturas de Corpos, Anéis e Grupos**, DT-FEEC-UNICAMP. 1998.

19 MC WILLIANS, F.J.; SLOANE, N.J.A. **The Theory of Error Correcting Codes**. North- Holland Publishing Company, 1977.

20 PETERSON ,W.W.; WELDOM Jr., E.J. **Error-Correcting Codes**, 2nd ed. MIT Press, 1972.

21 STEWART, I.; TALL, D. **Algebraic Number Theory**. New York: Chapman-Hall, 1987.

22 MUNDO EDUCAÇÃO. **Diferença entre células procarióticas e eucarióticas**. Disponível em: <https://mundoeducacao.uol.com.br/biologia/diferenca-entre-celulas-procariotas-eucariotas.htm> Acesso em: 10 de setembro de 2023.

23 BIOLOGIA NET. **Ciclo celular**. Disponível em: <https://www.biologianet.com/biologia-celular/ciclo-celular.htm> Acesso em: 10 de setembro de 2023.

24 PASSEI DIRETO. **Reprodução sexuada**. Disponível em: <https://www.passeidireto.com/arquivo/70551150/reproducao-sexuada> Acesso em: 10 de setembro de 2023.

25 CURSO ENEM GRATUITO. **Proteínas:** o que são, funções e a estrutura dos aminoácidos. Disponível em: <https://cursoenemgratuito.com.br/proteinas-bioquimica-celular/> Acesso em: 10 de setembro de 2023.

26 DIFERENÇA. **DNA e RNA.** Disponível em: <https://www.diferenca.com/dna-e-rna/> Acesso em: 10 de setembro de 2023.

27 PROENEM. **Transcrição e tradução.** Disponível em: <https://proenem.com.br/enem/biologia/transcricao-e-traducao/> Acesso em: 10 de setembro de 2023.

28 BIOLOGIA NET. **Síntese proteica.** Disponível em: <https://www.biologianet.com/biologia-celular/sintese-proteica.htm> Acesso em: 10 de setembro de 2023.

29 MENDELICS. **Alterações genéticas:** Glossário de genética III. Disponível em: <https://blog.mendelics.com.br/glossario-de-genetica-parte-3/> Acesso em: 10 de setembro de 2023.

30 NIH. **National Library of Medicine**. Disponível em: <https://www.ncbi.nlm.nih.gov/> Acesso em: 3 de janeiro de 2024.

31 NIH. **National Library of Medicine**. Disponível em: <https://www.ncbi.nlm.nih.gov/gene/?term=63%5BGene+Lenght%5D> Acesso em: 3 de janeiro de 2024.