

UNIVERSIDADE FEDERAL DE ALFENAS

BRUNA PIRES ROCHA

**APLICAÇÕES DE ESTRUTURAS
ALGÉBRICAS NA CONEXÃO ENTRE
SISTEMAS DE COMUNICAÇÃO PADRÃO E
GENÉTICA**

Alfenas/MG

2018

BRUNA PIRES ROCHA

**APLICAÇÕES DE ESTRUTURAS ALGÉBRICAS NA
CONEXÃO ENTRE SISTEMAS DE COMUNICAÇÃO
PADRÃO E GENÉTICA**

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do título de Licenciado em Matemática pelo Instituto de Ciências Exatas da Universidade Federal de Alfenas. Área de concentração: Matemática Aplicada. Orientador: Anderson José de Oliveira.

Alfenas/MG

2018

BRUNA PIRES ROCHA

**APLICAÇÕES DE ESTRUTURAS ALGÉBRICAS NA
CONEXÃO ENTRE SISTEMAS DE COMUNICAÇÃO
PADRÃO E GENÉTICA**

A Banca examinadora abaixo-assinada, aprova a Monografia apresentada como parte dos requisitos para obtenção do título de Licenciado em Matemática pelo Instituto de Ciências Exatas da Universidade Federal de Alfenas. Área de concentração: Matemática Aplicada.

Aprovada em: / /

Banca Examinadora:

Prof. Dr. Anderson José de Oliveira
Instituto de Ciências Exatas
Presidente

Prof. Dr^a. Cátia Regina de Oliveira Quilles
Queiroz
Instituto de Ciências Exatas
Avaliador 1

Prof. Dr. Evandro Monteiro
Instituto de Ciências Exatas
Avaliador 2

Prof. Dr^a. Flávia da Ré Guerra
Departamento de Anatomia
Suplente

Dedico este trabalho primeiramente a Deus, aos meus pais, Nelma e Gilson (in memoriam), que são fontes de inspiração e coragem para prosseguir, aos meus amigos, namorado e familiares por sempre me apoiarem.

Agradecimentos

A Deus por me proporcionar a vida iluminando o meu caminho, me concedendo força, saúde e coragem durante esta caminhada acadêmica.

A minha mãe e meus avós, que sempre me incentivaram e me cobriram com orações, compreendendo toda dificuldade encontrada. Ao meu pai (in memoriam) que foi o meu maior exemplo de fé e perseverança, que desde muito cedo me direcionou ao melhor caminho e me deu a maior motivação para seguir sua profissão. Obrigada por tudo! Saudades eternas!

As minhas irmãs pelo incentivo moral, ao meu namorado e amigos que sempre me apoiaram nos momentos difíceis e me aturam nos muitíssimos momentos de drama na faculdade, rs.

Ao professor Anderson José de Oliveira pela excepcional orientação, paciência, disposição e competência durante o desenvolvimento do trabalho. Sou eternamente grata, pela tranquilidade que me passava em cada reunião sempre me atendendo prontamente, pela amizade que construímos, pelos excelentes apontamentos no trabalho que gerou a publicação de um artigo.

Aos professores membros da banca examinadora Dr^a. Cátia Regina de Oliveira Quilles Queiroz, Dr. Evandro Monteiro e Dr^a. Flávia da Ré Guerra pela disponibilidade em analisar o trabalho, com valiosas sugestões e apontamentos.

Enfim, agradeço a todos que colaboraram direta ou indiretamente, para que este trabalho fosse finalizado de maneira efetiva e contribuíram para minha formação.

Muito obrigada!

Resumo

Um dos grandes desafios para a comunidade científica é analisar a existência de uma estrutura matemática relacionada com o DNA. Este trabalho propõe uma análise dos modelos de sistemas de comunicação genética propostos por Gatlin (1972), Rolman-Roldan (1996), May et al. (2004), Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011). Estes últimos identificam, reproduzem e classificam matematicamente diferentes sequências de DNA. Para tanto, este trabalho tem como objetivo identificar a aplicação de estruturas algébricas na conexão entre os dogmas centrais da teoria de comunicações e da biologia molecular. Tendo em vista o objetivo traçado e a interdisciplinaridade que este estudo visa, foi feito um estudo bibliográfico acerca de conceitos biológicos, algébricos e de sistemas de comunicação, para a associação algébrica do código genético e um sistema de comunicação. Por meio dessa associação, percebe-se uma interessante conexão entre elementos biológicos (código genético), engenharia (constelação de sinais em um processo de transmissão de informação) e elementos de álgebra (extensão de Galois $GF(2^6)$). Os resultados apresentados neste trabalho contribuem em uma área de pesquisa em franca expansão, fazendo com que a biologia se torne uma ciência fundamentada teoricamente em conceitos matemáticos. Portanto, o desenvolvimento deste trabalho complementa a formação acadêmica e auxilia em melhoramentos genéticos, reduzindo tempo e custos laboratoriais.

Palavras-chave: Álgebra. Código genético. Informação. Analogia.

Abstract

One of the greatest challenges for the scientific community is to analyze the existence of a mathematical structure related to DNA. This work proposes an analysis of the models of genetic communication systems proposed by Gatlin (1972), Rolman-Roldan (1996), May et al. (2004), Rocha and Palazzo Júnior (2010) and Faria and Palazzo Júnior (2011). These latter identify, reproduce and classify mathematically different DNA sequences. Therefore, this work aims to identify the application of algebraic structures in the connection between the central dogmas of communications theory and molecular biology. In view of the objective outlined and the interdisciplinarity that this study aims for, it was made a bibliographic study about biological, algebraic and communication systems concepts, for the algebraic association of the genetic code and a communication system. Through this association, an interesting connection between biological elements (genetic code), engineering (signals constellation in an information transmission process) and elements of algebra (Galois extension $GF(2^6)$). The results presented in this work contribute in an area of research that is expanding, making biology a based science theoretically on mathematical concepts. Therefore, the development of this work complements the academic education and helps in genetic improvements, reducing time and labor costs.

Keywords: Algebra. Genetic code. Information. Analogy.

Lista de ilustrações

Figura 1 – Célula procarionte, à esquerda e Célula eucarionte, à direita.	16
Figura 2 – Nucleotídeos do DNA.	17
Figura 3 – Nucleotídeos do RNA.	18
Figura 4 – Esquema de duplicação do DNA.	19
Figura 5 – Esquema do processo de transcrição.	20
Figura 6 – Esquema da síntese proteica.	20
Figura 7 – Elementos básicos de um sistema de comunicação.	30
Figura 8 – Diagrama de blocos de um sistema de comunicação.	32
Figura 9 – O dogma central da teoria de comunicações.	33
Figura 10 – Esquema da síntese proteica.	34
Figura 11 – Sistema de comunicação de informação genética.	35
Figura 12 – Modelo proposto por Gatlin.	36
Figura 13 – Modelo proposto por Romam-Roldan.	37
Figura 14 – Modelo proposto por May.	37
Figura 15 – Modelo proposto por Rocha.	38
Figura 16 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genética.	39
Figura 17 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genômica.	40
Figura 18 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genética.	43
Figura 19 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genômica.	43
Figura 20 – Modelo de um sistema de comunicação de informação genética.	44
Figura 21 – O alfabeto do código genético e os alfabetos 4-ários dos CCEs.	45
Figura 22 – Associações entre $N \rightarrow GF(4)$ e $N \rightarrow \mathbb{Z}_4$	46
Figura 23 – Representação da estrutura de espaço vetorial sobre o corpo.	46
Figura 24 – Representação de Espaço Vetorial sobre o Anel.	47
Figura 25 – Codificador genético.	47
Figura 26 – Possibilidades de rotulamento entre os elementos do conjunto N e $GF(4)$	48
Figura 27 – Possibilidades de rotulamento entre os elementos do conjunto N e \mathbb{Z}_4	48
Figura 28 – Representação do mapeamento de $GF(4)$ e \mathbb{Z}_4	48
Figura 29 – Rotulamento e \mathbb{S}_4	49
Figura 30 – Rotulamentos A,B e C.	50
Figura 31 – Casamento entre contexto biológico e o contexto matemático.	50
Figura 32 – Mapeamentos do alfabeto \mathbb{Z}_4	51

Figura 33 – Constelação de 64 sinais.	51
Figura 34 – Associação algébrica do modelo de comunicação genética.	52

Lista de tabelas

Tabela 1	–	Hidropacidade dos aminoácidos.	21
Tabela 2	–	O código genético	22
Tabela 3	–	Tábua multiplicativa de $G = (+1, -1)$	24
Tabela 4	–	Tábua da Soma \mathbb{Z}_5	25
Tabela 5	–	Tábua da Multiplicação \mathbb{Z}_4	25
Tabela 6	–	Operação de adição em módulo - 5.	26
Tabela 7	–	Operação de multiplicação em módulo - 5.	27
Tabela 8	–	$GF(2^4)$ gerado por $p(X) = 1 + X + X^4$	29
Tabela 9	–	$GF(2^6)$ gerado por $p(X) = 1 + X + X^6$	54

Lista de abreviaturas e siglas

DNA	Ácido Desoxirribonucleico
RNA	Ácido Ribonucleico
RNA _m	RNA mensageiro
RNA _t	RNA transportador
RNA _r	RNA ribossômico
CCEs	Códigos Corretores de Erros

Sumário

1	INTRODUÇÃO	13
2	ELEMENTOS DE BIOLOGIA, ÁLGEBRA E SISTEMAS DE CO-MUNICAÇÃO	15
2.1	Elementos de Biologia	15
2.1.1	Células	16
2.1.2	Nucleotídeos e Ácidos Nucleicos	16
2.1.3	A Molécula de DNA	18
2.1.4	A Duplicação do DNA	18
2.1.5	O Processo de Transcrição	19
2.1.6	Síntese Proteica	20
2.1.7	Proteínas / Aminoácidos	21
2.1.8	O Código Genético	22
2.1.9	Mutações	23
2.2	Elementos de Álgebra	23
2.2.1	Grupos	24
2.2.2	Anéis	24
2.2.3	Corpos	25
2.2.4	Corpos de Galois	26
2.2.5	Construções de Corpos de Galois $GF(2^m)$	27
2.3	Sistema de Comunicação Digital	29
2.3.1	Sistema de Comunicação	29
3	CONEXÕES ENTRE OS DOGMAS DA TEORIA DE COMUNICA- ÇÕES E BIOLOGIA MOLECULAR	33
3.1	O Dogma Central da Teoria de Comunicações	33
3.2	O Dogma Central da Biologia Molecular	34
3.3	Analogias entre Sistemas de Comunicações e Sistemas de Informa- ção Genética	35
3.4	Comparação e Análise de Sistemas de Comunicação Biológico	35
3.4.1	Modelo de Comunicação de Gatlin	36
3.4.2	Modelo de Comunicação de Roman-Roldan	36
3.4.3	Modelo de Comunicação de May	37
3.4.4	Modelo de Comunicação de Rocha	38
3.4.5	Modelo de Comunicação de Faria	39
3.4.6	Comparações dos Modelos Analisados	40

4	CARACTERIZAÇÃO ALGÉBRICA DO MODELO DE CODIFICAÇÃO GENÉTICA PROPOSTO POR FARIA (2011)	42
4.1	Modelo de Codificação Genética e Genômica	42
4.1.1	Modelo de Codificação Genética	43
4.1.2	Elementos Algébricos Associados ao Codificador Genético	44
4.1.3	Conexão entre Código Genético, Constelação de Sinais e Extensões de Galois	51
5	CONCLUSÕES	57
	REFERÊNCIAS	59

1 Introdução

Embora aparentemente não apresentem relação, tanto a teoria de comunicações quanto a genética preocupam-se com a transferência de informação. Levando em consideração o fato de que a teoria de comunicações é realizada pelo homem, e a genética por um processo natural, ambas diferem entre si, principalmente por não operarem na mesma dimensão. Sendo assim, enquanto a teoria de comunicações está programada para enviar mensagens no espaço, de um lugar para outro, a genética envia mensagens hereditárias no tempo (BATTAIL, 2008).

Neste estudo, buscamos identificar a aplicabilidade de estruturas algébricas na conexão entre um sistema de comunicação padrão e um sistema de comunicação genético. O dogma central da teoria de comunicações, consiste no processo de transmitir uma informação desde um emissor até um receptor, podendo ocorrer erros, os chamados ruídos, que estão alocados no canal, comprometendo a informação original. Desta forma, um dos grandes desafios para a comunidade científica é analisar a existência de uma estrutura matemática relacionada com o DNA e com o sistema de comunicação. Existem diversos estudiosos que tratam sobre este tema, em especial, Claude Shannon, matemático americano, engenheiro eletrônico e criptógrafo que ficou conhecido como o “pai da teoria da informação”. Ele foi o precursor para a teoria de códigos e a teoria de comunicação.

Historicamente, na década de 1970, iniciou-se a aplicação da teoria da informação para análise de dados genéticos, porém sem muito sucesso. Após alguns anos, este estudo despertou novamente o interesse da aplicação da teoria de comunicações com o estudo do genoma. Com isso, muitos pesquisadores, assim como May et al. (2004), têm explorado o dogma central da genética a partir de um ponto de vista da transmissão de informação, cuja premissa central é a de que os genes sejam perpetuados sob a forma de sequências de ácidos nucleicos e gerados como proteínas.

Há um grande número de ferramentas matemáticas, principalmente algébricas, que são utilizadas para estudar diversos fenômenos biológicos. Os estudos de Rocha (2010), Faria (2011) e May (2004) são algumas das pesquisas recentes que têm focado nesta questão, com ênfase na proposta de modelos para o sistema de comunicação de informação genética e genômica, análogos ao sistema de comunicação digital. Assim, a aplicabilidade das estruturas algébricas tem sido alvo de diversos estudos, cujo intuito é representar matematicamente o que o mundo biológico realiza (OLIVEIRA, 2012).

Desta forma, podemos aplicar estruturas algébricas na conexão entre um sistema de comunicação padrão e um sistema de comunicação biológico. Esta afirmação pode ser mostrada a partir de respostas à seguinte questão: por meio das estruturas algébricas, como

é possível estabelecer conexões entre um sistema de comunicação padrão e um sistema de comunicação biológico?

Sendo assim, o propósito deste estudo é identificar a aplicação de estruturas algébricas na conexão entre os dogmas centrais da teoria de comunicações e da biologia molecular. Além disso, pretende-se compreender conceitos de biologia celular e molecular, compreender conceitos de álgebra abstrata como: grupos, anéis, corpos, corpos de Galois e suas propriedades, relacionar o sistema de comunicação com a estrutura do DNA e o processo de geração de proteínas e relacionar as analogias entre um sistema de comunicação padrão e um sistema de comunicação genética, a partir de alguns modelos propostos na literatura.

Para todo o embasamento interdisciplinar deste trabalho, serão apresentados os conceitos principais, apontando as principais aplicações, sendo organizado da seguinte forma: no Capítulo 2 serão apresentados, de forma sucinta, conteúdos de: Biologia (Seção 2.1), como biologia celular, biologia molecular, código genético e mutações; Álgebra (Seção 2.2) como: grupos, anéis, corpos e extensões de Galois; Sistemas de comunicação (Seção 2.3). No Capítulo 3 serão apresentadas as conexões entre a teoria de comunicações e o dogma da biologia molecular (Seções 3.1 e 3.2); analogias (Seção 3.3); comparação e análise de sistemas de comunicação biológica (Seção 3.4) e os modelos analisados no trabalho como: modelo de comunicação de Gatlin (Subseção 3.4.1); modelo de comunicação de Roman-Roldan (Subseção 3.4.2); modelo de comunicação de May (Subseção 3.4.3); modelo de comunicação de Rocha (Subseção 3.4.4); modelo de comunicação de Faria (Subseção 3.4.5). No Capítulo 4 será apresentada a caracterização algébrica do modelo de comunicação genética proposto por Faria e Palazzo Júnior (2011), cuja caracterização deste modelo requer uma associação com estruturas algébricas, em particular, a extensão de Galois $GF(2^6)$, que faz a relação do código genético e a constelação de sinais presente no modulador genético. Finalmente, no Capítulo 5 serão apresentadas as conclusões deste trabalho, bem como propostas de trabalhos futuros.

2 Elementos de Biologia, Álgebra e Sistemas de Comunicação

Em virtude da interdisciplinaridade deste trabalho, este capítulo introduz alguns conceitos de biologia, álgebra e de sistemas de comunicação, indispensáveis para o desenvolvimento do presente estudo.

Na Seção 2.1 serão apresentados os elementos de biologia, com a definição, estrutura e características das células, bem como a importância dos nucleotídeos e dos ácidos nucleicos. Além disso, serão apresentadas as principais características e estruturas da molécula de DNA e RNA, desde a duplicação do DNA até a síntese proteica. Também serão apresentados os principais aminoácidos que são codificados pelas 64 trincas (códon) do código genético. Por fim, as características e os tipos de mutações.

Na Seção 2.2 serão apresentados alguns conceitos de álgebra, tais como grupos, anéis, corpos e corpos de Galois. Essas estruturas são primordiais para o embasamento dos cálculos e das conexões com o código genético. O mapeamento das bases nitrogenadas que estruturam o DNA: Adenina (A), Citosina (C), Guanina (G) e Timina (T), bem como o RNA: Adenina (A), Citosina (C), Guanina (G) e Uracila (U), tem como alfabeto o conjunto $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, obedecendo as operações de soma e de produto módulo 4 e também, o alfabeto 4-ário que corresponde ao $GF(4) = \{0, 1, \alpha, \alpha^2\}$.

Na Seção 2.3 serão apresentados os principais elementos de um sistema de comunicação, bem como suas características.

2.1 Elementos de Biologia

Nesta Seção serão apresentados alguns conceitos de biologia que serão utilizados neste trabalho. Na Subseção 2.1.1 são apresentados as principais características e funcionalidades das células. Na Subseção 2.1.2, o estudo se refere aos nucleotídeos e ácidos nucleicos. Na Subseção 2.1.3 é detalhada a estrutura do ácido desoxirribonucleico, o DNA. O processo de duplicação, transcrição e tradução do DNA é explanado nas Subseções 2.1.4, 2.1.5, 2.1.6, respectivamente. Na Subseção 2.1.7 são apresentados os principais aminoácidos, sua estrutura e funções. O código genético é apresentado na Subseção 2.1.8. Por fim, na Subseção 2.1.9, é apresentado um estudo sobre as mutações e suas consequências, benéficas ou maléficas, nos organismos.

Os conceitos relacionados com os elementos de biologia podem ser encontrados em detalhes em Alberts et al. (2010), Marzzoco e Torres (2013), Pamphile e Vicentini (2011)

e Oliveira e Palazzo Júnior (2012).

2.1.1 Células

A célula é a menor porção de matéria com vida do organismo, são microscópicas e só conseguem ser visualizadas com aparelhos especiais. Foi descoberta por Robert Hooke, em 1665. Ele as descobriu quando examinava no microscópio um pedaço de cortiça e percebeu que ele era composto por compartimentos, os quais designou células. As principais funções das células são: captar nutrientes, utilizando-os na produção de energia, eliminar resíduos do metabolismo e capacidade de movimentação (em alguns casos).

As células se dividem em dois grupos, procariontes ¹ e eucariontes ². A Figura 1 apresenta um exemplo de célula procarionte e célula eucarionte.

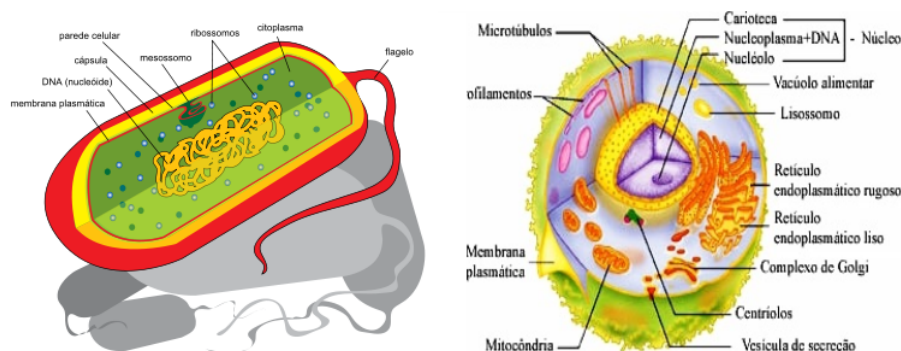


Figura 1 – Célula procarionte, à esquerda e Célula eucarionte, à direita.

Fonte: Almeida (2012).

De acordo com (ALBERTS et al., 2010) a origem das células está diretamente relacionada com o surgimento da vida em nosso planeta. Estudos com fósseis determinaram que as células procarióticas surgiram a aproximadamente $3,5 - 3,0 \times 10^9$ anos e as eucarióticas a aproximadamente $1,5 \times 10^9$ anos.

2.1.2 Nucleotídeos e Ácidos Nucleicos

Os ácidos nucleicos são macromoléculas de suma importância biológica. São encontradas em todas as células vivas, atribuindo-se aos mesmos as importantes funções

¹ Células procariontes: Sua principal característica é a ausência de núcleo (carioteca) e da maioria das organelas, tais como: mitocôndria, cloroplasto e complexo de Golgi. Aqui estão incluídas as bactérias e as cianobactérias (algas azuis).

² Células eucariontes: Apresentam um núcleo, um compartimento bem definido onde o material genético fica isolado do citoplasma. São células mais complexas que as procarióticas e, existem estudos que abordam que a célula eucarionte é um aperfeiçoamento de células procariontes. Possuem membrana nuclear individualizada e vários tipos de organelas. Aqui estão incluídos os animais, vegetais, fungos e leveduras, que são seres pluricelulares.

de conter, transmitir e traduzir as informações genéticas de um determinado organismo. Vale ressaltar que, a vida, como a conhecemos está intimamente relacionada à química dos nucleotídeos e dos ácidos nucleicos. Portanto, a maioria dos organismos vivos possuem ácidos nucleicos na forma de ácido desoxirribonucleico (DNA) e ácido ribonucleico (RNA).

O DNA é o grande responsável pelo armazenamento da informação genética, através de uma sequência de bases (adenina, timina, citosina e guanina) que estão interligadas por pontes de hidrogênio e também é composto por um grupo fosfato e uma pentose (desoxirribose), sendo que pode ser transcrita ou traduzida. Ela é transcrita em moléculas de RNA que depois são traduzidas em proteínas, por meio dos aminoácidos. Os nucleotídeos do DNA são apresentados na Figura 2.

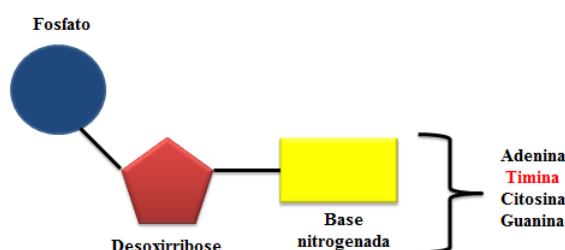


Figura 2 – Nucleotídeos do DNA.

As bases nitrogenadas são unidas por meio de pontes de hidrogênio, sendo estas complementares, onde sempre a adenina se liga à timina, por meio de duas pontes de hidrogênio e a citosina se liga à guanina, por meio de três pontes de hidrogênio. Esta complementariedade foi descoberta por Erwin Chargaff, em meados da década de 40.

Cada ácido nucleico tem um açúcar, chamado pentose. No RNA a pentose é chamada de ribose e no DNA ela é chamada de desoxirribose. A única diferença do ponto de vista de composição da molécula entre estes dois açúcares é que a desoxirribose possui um átomo de oxigênio a menos.

O DNA é constituído por uma fita dupla de nucleotídeos. Em contrapartida, o RNA é composto por uma longa fita de nucleotídeos simples que estão ligadas entre si. Possui em sua estrutura um grupo fosfato, uma pentose (ribose) e uma base nitrogenada (adenina, guanina, citosina e uracila), conforme apresentado na Figura 3.

Existem três tipos de RNA: O RNA mensageiro (RNAm), o RNA tranpostador (RNAt) e o RNA ribossômico (RNAr), cada um com uma função específica no processo da síntese protéica e se diferem, quanto aos organismos procarióticos e eucarióticos. Nos organismos procarióticos, o processo de transcrição e tradução ocorrem bem próximos um dos outros. Já nos organismos eucarióticos, a transcrição ocorre no núcleo e a tradução no citoplasma da célula.

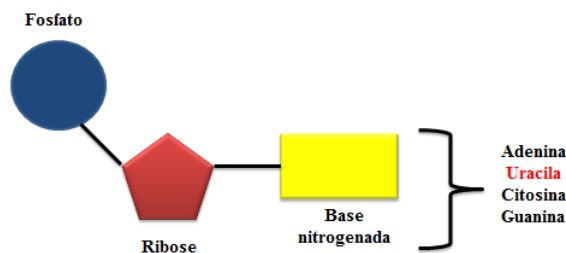


Figura 3 – Nucleotídeos do RNA.

É de suma importância o entendimento da estrutura e características da molécula de DNA, antes de analisar os processos de transcrição e tradução da célula.

2.1.3 A Molécula de DNA

O DNA é um ácido desoxirribonucleico, cuja estrutura mais difundida é a dupla hélice de nucleotídeos. Segundo (PAMPHILE; VICENTINI, 2011) o DNA descreve as estruturas primárias de todas as proteínas, RNAs e, por meio das enzimas, afeta diretamente todos os processos constituintes celulares, como forma, tamanho e função dos seres vivos.

O modelo da molécula de DNA elaborado por Watson e Crick em 1953, foi muito significativo para o desenvolvimento de pesquisas na área da Biologia Molecular e suas principais características são:

1. Duas cadeias polinucleotídicas (chamadas fitas), formando a dupla hélice.
2. As duas fitas de DNA são antiparalelas.
3. As bases nitrogenadas ocupam o centro da hélice e as cadeias de açúcar-fosfato estão na periferia.
4. Cada base está ligada a outra por meio de pontes de hidrogênio, formando um par de base planar, onde adenina se liga com timina e vice-versa e guanina se liga com citosina e vice-versa.

2.1.4 A Duplicação do DNA

O processo de duplicação do DNA ocorre por meio da presença da enzima DNA *polimerase*. Com a presença dessa enzima, ocorrem as seguintes etapas:

1. Primeiramente, ocorre o rompimento das pontes de hidrogênio, que ligam as bases nitrogenadas e as duas fitas se afastam.

2. Os nucleotídeos livres na célula se encaixam nas duas fitas que se afastaram, respeitando a complementariedade da Regra de Chargaff.
3. Quando as duas fitas originais tiverem sido completadas por novos nucleotídeos, obtemos duas moléculas de DNA idênticas entre si.

Na Figura 4 é apresentado o processo da duplicação do DNA.

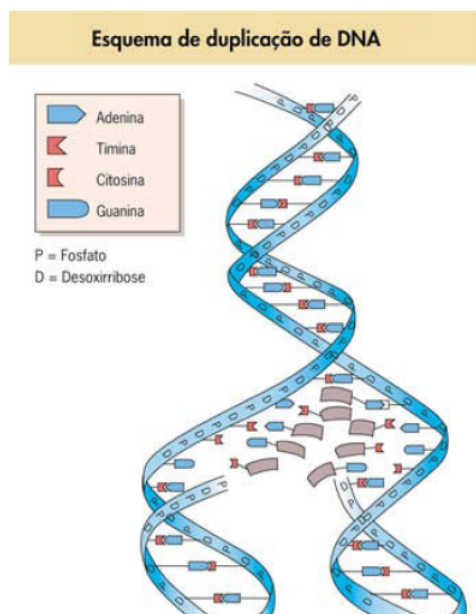


Figura 4 – Esquema de duplicação do DNA.

Fonte: Marques (2014).

2.1.5 O Processo de Transcrição

A transcrição consiste no processo em que o RNA é sintetizado a partir de um molde de DNA, ou seja, o DNA produz moléculas de RNAm e, tais moléculas migram para o citoplasma e controlam a produção das proteínas. Uma importante diferença em relação ao processo de duplicação do DNA é que uma das fitas do DNA serve de molde para produzir a molécula de RNAm. Sendo assim, o RNA produzido será, portanto, uma fita simples e não dupla. Tal processo ocorre através das seguintes etapas:

1. É necessária a presença da enzima RNA polimerase.
2. As duas fitas de DNA se afastam, quando as pontes de hidrogênio se desfazem.
3. Denominamos por fita ativa, o encaixe dos nucleotídeos livres do RNA numa das fitas.

4. A molécula de RNA destaca-se de seu molde de DNA e migra para o citoplasma.
5. Através do pareamento das duas fitas de DNA, reconstitui-se a molécula original.

A Figura 5 apresenta o processo de transcrição.

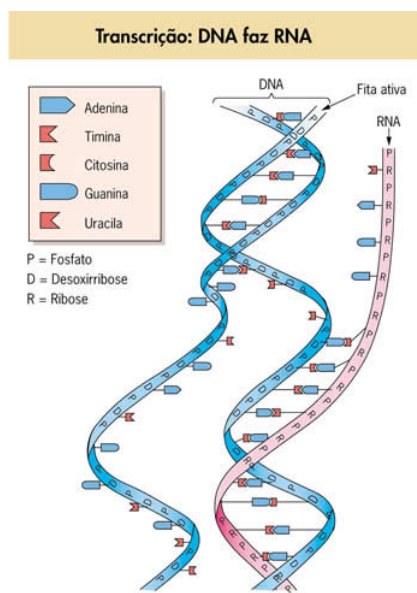


Figura 5 – Esquema do processo de transcrição.

Fonte: Marques (2014).

2.1.6 Síntese Proteica

De acordo com Alberts et al. (2010), a síntese proteica é considerada o fator central da biologia molecular, sendo o DNA e o RNA fundamentais neste processo. Em suma, este processo consiste na transcrição do DNA pelo RNAm, em seguida a informação é traduzida pelo RNAr, moléculas de proteínas e pelo RNAt, que transporta os aminoácidos, cuja sequência formará a nova molécula de proteína.

Na Figura 6, é apresentado um esquema básico da síntese proteica. É importante destacar que o processo visto no esquema ocorre de maneira diferente entre as células eucarióticas e procarióticas.

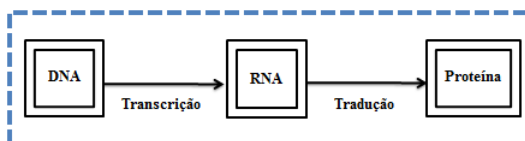


Figura 6 – Esquema da síntese proteica.

2.1.7 Proteínas / Aminoácidos

As proteínas apresentam-se sob inúmeras formas e tamanhos, sendo as mais importantes das macromoléculas biológicas, e exercem funções metabólicas cruciais ao bom funcionamento da célula. Ressalta-se ainda que as proteínas representam uma classe especial de moléculas. Segundo (MARZZOCO; TORRES, 2013), uma dada proteína, por si, representa um polímero ³, sendo constituída por subunidades denominadas aminoácidos.

Os aminoácidos são unidades utilizadas na síntese proteica e por possuírem carbono são essenciais à vida humana. Também podem ser classificados como hidrofóbicos ⁴ ou hidrofílicos ⁵. Um dos critérios para a classificação dos aminoácidos é a hidropacidade, sendo estes também classificados pela carga (ácido, neutro, básico), estrutura (cíclico, acíclico) tamanho (pequeno, médio, grande), aromático e alifático.

Dos 20 aminoácidos existentes, 12 são hidrofílicos (polares) e 8 hidrofóbicos (apolares), apresentados na Tabela 1. Os aminoácidos são classificados entre essenciais ou naturais de acordo com sua fonte primária. Os aminoácidos essenciais são sintetizados apenas pelos vegetais são eles: leucina, isoleucina, lisina, metionina, treonina, triptofano, valina e fenilalanina. Os naturais são sintetizados tanto por animais tanto por vegetais, são eles: alanina, glicina, histidina, tirosina, arginina, ácido aspártico, asparagina, glutamina, serina, prolina, cisteína e ácido glutâmico.

Tabela 1 – Hidropacidade dos aminoácidos.

Hidrofílico (Polar)	Hidrofóbico (Apolar)
Asparagina (Asp)	Alanina (Ala)
Glutamina (Gln)	Leucina (Leu)
Arginina (Arg)	Valina (Val)
Histidina (His)	Isoleucina (Ile)
Lisina (Lys)	Prolina (Pro)
Cisteína (Cys)	Fenilalanina (Phe)
Glicina (Gly)	Metionina (Met)
Serina (Ser)	Triptofano (Trp)
Treonina (Thr)	
Ácido aspártico (Asp)	
Ácido glutâmico (Gln)	
Tirosina (Tyr)	

Fonte: Oliveira e Palazzo Júnior (2012).

³ Polímero: são compostos por macromoléculas de peso molecular elevado formados por moléculas menores (monômeros).

⁴ Hidrofóbicos: evitam o contato com a água, eles se "escondem" no interior da proteína.

⁵ Hidrofílicos: distribuem-se na superfície da molécula ficando em contato com a água.

2.1.8 O Código Genético

O DNA de cada célula possui as informações genéticas do indivíduo, essas informações são passadas através da transcrição e tradução. No processo de transcrição todos os RNA's são feitos a partir de um DNA.

A síntese proteica ocorre a partir do DNA, mas não é ele que sintetiza, pois como intermediário temos o RNAm. A sequência do RNAm é lida e assim são sintetizadas muitas proteínas é esse o processo de tradução, que ocorre nos ribossomos.

As proteínas são polímeros formados através da combinação aleatória de 20 tipos de aminoácidos. Existem 64 possíveis combinações, utilizando as quatro bases nitrogenadas do DNA/RNA agrupadas três-a-três, caracterizando a especificação dos 20 distintos aminoácidos.

A síntese proteica é feita em blocos de três nucleotídeos, que são os códons, cada códon é responsável por um aminoácido específico. A correspondência do códon para um aminoácido é o que denominamos de código genético, apresentado na Tabela 2, com seus respectivos aminoácidos, abreviação e códons correspondentes.

Tabela 2 – O código genético

Primeira Base	Segunda Base				Terceira Base
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Fonte: Oliveira e Palazzo Júnior (2012).

Todos os organismos utilizam do mesmo código para produzir suas proteínas, para parar a tradução temos os chamados STOP códons que indicam que é o momento de parar a síntese proteica. Vale ressaltar que, o código é dito universal, pois todos os organismos usufruem de um mesmo código na tradução de proteínas com algumas exceções tais como, em certos protozoários e mitocôndrias, que apresentam pequenas diferenças.

2.1.9 Mutações

As mutações são mudanças que ocorrem nas sequências de nucleotídeos do material genético de um organismo. De acordo com a alteração provocada, elas podem ser classificadas como desfavoráveis (quando ocasiona erros na sequência de proteínas, causando doenças hereditárias) ou favoráveis (quando são criadas novas proteínas que ajudam o organismo a se adaptar melhor a um ambiente). Nesta perspectiva, o efeito ocasionado na estrutura de um organismo também é classificado como mutações de pequena e de grande escala.

Mutações de pequena escala: são aquelas que afetam um pequeno gene em um ou poucos nucleotídeos.

1. **Mutação pontual:** é geralmente causada pelos erros na replicação do DNA, onde ocorre a troca de nucleotídeos, podendo ocorrer a transição ou a transversão dos mesmos, sendo a transição a troca de pirimidina por pirimidina ($C \longleftrightarrow U$) ou purina por purina ($A \longleftrightarrow G$), já na transversão ocorre a troca de uma pirimidina por uma purina e vice-versa ($C/U \longleftrightarrow A/G$). Essas mutações pontuais também são classificadas em três tipos, são eles:

- (i) **Mutação silenciosa:** é quando o códon codifica o mesmo aminoácido.
- (ii) **Mutação “missence”:** é quando um códon codifica um aminoácido distinto.
- (iii) **Mutação sem sentido:** é quando se codifica para um códon de parada (STOP), interrompendo a proteína antes de seu término.

2. **Inserção:** ocorre pela adição de um ou mais nucleotídeos na sequência do DNA. Essa mutação é ocasionada por erros no processo de replicação de elementos repetitivos.

3. **Deleção:** além de serem mutações geralmente irreversíveis, ocorre a remoção de um ou mais nucleotídeos na sequência do DNA.

Mutações de grande escala:

1. **Duplicação gênica:** é a criação de várias cópias de uma região cromossômica, que aumenta a dosagem dos genes dentro dela.

2. **Deleção de regiões cromossômicas:** é a mutação que ocasiona a perda dos genes presentes nessas regiões.

3. **Perda de heterozigosidade:** é onde ocorre a perda de um alelo por deleção num organismo que possuía dois alelos originalmente.

2.2 Elementos de Álgebra

Nesta Seção serão apresentados os conceitos principais de estruturas algébricas, tais como grupos, anéis, corpos e corpos de Galois, usados no decorrer do trabalho. A mesma está dividida na seguinte forma: na Subseção 2.2.1 é apresentada a definição, os

principais conceitos e exemplo de grupos. Na Subseção 2.2.2 é apresentada a estrutura de anéis. Na Subseção 2.2.3 é apresentada um estudo sobre corpos e, finalmente na Subseção 2.2.4 é apresentada um estudo sobre os corpos de Galois.

Os conceitos, propriedades e definições sobre estruturas algébricas são encontrados em: Domingues e Iezzi (2003), Garcia e Lequain (2010) e Lin e Costelo (1983).

2.2.1 Grupos

Definição 1: Um conjunto $G \neq \emptyset$ com uma operação $*$ onde:

$$\begin{aligned}\gamma: G \times G &\longrightarrow G \\ (a, b) &\longmapsto a * b,\end{aligned}$$

munido das seguintes propriedades:

1. **associativa:** $a * (b * c) = (a * b) * c, \forall a, b, c \in G,$
 2. \exists **elemento neutro** e para a operação $*$: $e * a = a = a * e, \forall a \in G,$
 3. $\forall a \in G, \exists b$ (**elemento inverso**): $a * b = b * a = e, \forall a, b \in G,$
- é um grupo.

Definição 2: Um grupo $(G, *)$ é chamado de **grupo abeliano** (ou grupo comutativo) se a operação binária $*$ satisfaz a seguinte condição: para quaisquer a e b em G , $a * b = b * a$.

Definição 3: Um grupo $(G, *)$ de ordem finita é chamado de **grupo finito**. Sendo assim, o número de elementos de G é chamado de **ordem do grupo**, (notação $(o(G))$) e a tabela da operação $*$ se denomina **tábua do grupo**.

Exemplo 1: É fácil verificar que $G = (+1, -1)$ é um grupo multiplicativo. Sua ordem é 2 e sua tabela é apresentada na Tabela 3.

Tabela 3 – Tábua multiplicativa de $G = (+1, -1)$.

$*$	$+1$	-1
$+1$	$+1$	-1
-1	-1	$+1$

2.2.2 Anéis

Definição 4: Um conjunto não vazio A e um par de operações binárias: uma adição $(a, b) \longmapsto a + b$ e uma multiplicação $(a, b) \longmapsto a \cdot b$, é denominado de anel se as seguintes propriedades forem verificadas:

Para a *soma*:

- (i) **associatividade**: $\forall a, b, c \in A, (a + b) + c = a + (b + c)$;
- (ii) **comutatividade**: $\forall a, b \in A, a + b = b + a$;
- (iii) existe um **elemento neutro** 0_A tal que: $\forall a \in A, a + 0_A = a$;
- (iv) existe **elemento oposto**: $\forall a \in A, \exists (-a) \in A$ tal que $a + (-a) = 0_A$;

Para a *multiplicação*:

- (v) **associatividade**: $\forall a, b, c \in A, a \cdot (b \cdot c) = (a \cdot b) \cdot c$;
- (vi) **distributividade** (em relação a adição): $\forall a, b, c \in A, a \cdot (b + c) = a \cdot b + a \cdot c$ e $(a + b) \cdot c = a \cdot c + b \cdot c$.

cujas notações são dadas por: $(A, +, \cdot)$, denotará um anel A com as operações $+$ e \cdot .

Definição 5: Um anel $(A, +, \cdot)$, onde a operação \cdot é comutativa é denominado de **anel comutativo**.

Definição 6: Um anel $(A, +, \cdot)$ em que o conjunto A é finito é denominado de **anel finito** e é representado da forma $\mathbb{Z}_m (m > 1)$. Sendo um anel finito, suas propriedades (i)-(vi) podem ser verificadas, através das tábuas de adição e multiplicação.

Exemplo 2: Observe as tábuas do anel $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, apresentadas nas Tabelas 4 e 5.

Tabela 4 – Tábua da Soma \mathbb{Z}_5 .

$+$	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 5 – Tábua da Multiplicação \mathbb{Z}_4 .

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

2.2.3 Corpos

Definição 7: Seja K um conjunto de elementos onde estejam definidas duas operações binárias, a adição $(+)$ e a multiplicação (\cdot) . Então, K é um corpo se as condições forem

satisfeitas:

- (i) K é um **grupo comutativo** sob a adição. O elemento identidade é o 0 (zero).
- (ii) O conjunto dos elementos não nulos em K é um grupo comutativo sob a multiplicação. O elemento identidade é o 1 (um).
- (iii) A multiplicação é distributiva em relação a adição, isto é, para quaisquer a, b e c em K , $a \cdot (b + c) = a \cdot b + a \cdot c$.

cujas notação é dada por $(K, +, \cdot)$, denotará um corpo K com as operações $+$ e \cdot .

Exemplo 3: Os anéis numéricos $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ e $(\mathbb{C}, +, \cdot)$ são corpos, pois todos os seus elementos não nulos admitem inversos multiplicativos. No entanto, o anel $(\mathbb{Z}, +, \cdot)$ não é um corpo pois, apenas os seus elementos 1 e -1 admitem inversos multiplicativos.

2.2.4 Corpos de Galois

Definição 8: É um corpo com um número finito de elementos representado por $GF(p^m)$, onde p é um número primo, e m um inteiro positivo.

A partir da definição, temos as seguintes propriedades:

1. Para qualquer a , $a \cdot 0 = 0 \cdot a$.
2. Para quaisquer dois elementos não nulos a e b , temos que $a \cdot b \neq 0$.
3. Se $a \cdot b = 0$ e, $a \neq 0$, então $b = 0$.
4. Para quaisquer dois elementos a e b , temos que $-(a \cdot b) = (-a) \cdot b = a \cdot (-b)$.
5. Para $a \neq 0$, temos que, se $a \cdot b = a \cdot c$, então $b = c$.

Exemplo 4: Considere o conjunto $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$. Este conjunto é um Corpo de Galois, cujas operações de adição e multiplicação módulo - 5 são apresentadas nas Tabelas 6 e 7.

Tabela 6 – Operação de adição em módulo - 5.

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Tabela 7 – Operação de multiplicação em módulo - 5.

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Para qualquer inteiro m é possível estender um corpo primo $GF(p)$ com p elementos para um corpo estendido $GF(p^m)$ com p^m elementos.

Definição 9: Um polinômio sobre $GF(2)$ de grau m é dito ser irredutível se $p(X)$ é não divisível por nenhum outro polinômio sobre $GF(2)$ de grau menor que m e maior que zero.

Exemplo 5: Os polinômios $X^2 + X + 1$, $X^3 + X + 1$ e $X^4 + X + 1$ são irredutíveis. Para qualquer $m \geq 1$ existe um polinômio irredutível de grau m . Com isso, temos como um importante resultado que qualquer polinômio irredutível sobre $GF(2)$ de grau m divide $X^{2^m-1} + 1$.

Definição 10: Um polinômio irredutível $p(X)$ 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 6: O polinômio $X^4 + X + 1$ é irredutível e primitivo, pois $p(X)$ divide $X^{15} + 1$ mas não divide nenhum $X^n + 1$ para $1 \leq n \leq 15$.

Entretanto, o polinômio $p(X) = X^4 + X^3 + X^2 + X + 1$ é irredutível mas não é primitivo, pois além de $p(X)$ dividir $X^{15} + 1$ ele também divide $X^5 + 1$.

2.2.5 Construções de Corpos de Galois $GF(2^m)$

Nesta Seção apresentamos um método para construir Corpos de Galois de 2^m elementos ($m > 1$) do corpo binário $GF(2)$. Consideramos os dois elementos de $GF(2)$, 0 e 1 e um novo símbolo α e a operação de multiplicação “.” da seguinte forma:

$$\begin{aligned} 0 \cdot \alpha^j &= \alpha^j \cdot 0 = 0 \\ 1 \cdot \alpha^j &= \alpha^j \cdot 1 = \alpha^j \\ \alpha^i \cdot \alpha^j &= \alpha^j \cdot \alpha^i = \alpha^{i+j} \end{aligned}$$

onde $i, j = 0, 1, 2, \dots$

Logo, formamos o conjunto $F = \{0, 1, \alpha, \alpha^2, \dots, \alpha^j, \dots\}$, sobre o qual a operação de “.” está definida. O elemento 1 é algumas vezes denotado por α^0 .

Assumindo a condição sobre o elemento α tal que o conjunto F contenha apenas 2^m elementos e seja fechado sobre a multiplicação. Seja $p(X)$ um polinômio primitivo de

grau m sobre $GF(2)$. Assim, assumimos que $p(\alpha) = 0$, isto é, α é raiz de $p(X)$. Como $p(X)$ divide $X^{2^m-1} + 1$, temos que $X^{2^m-1} + 1 = q(X) \cdot p(X)$. Trocando X por α , obtemos:

$$\begin{aligned} X^{2^m-1} + 1 &= q(\alpha) \cdot p(\alpha) \\ X^{2^m-1} + 1 &= q(\alpha) \cdot 0 \\ X^{2^m-1} + 1 &= 0 \\ X^{2^m-1} &= 1 \end{aligned}$$

Logo, sob a condição que $p(\alpha) = 0$, o conjunto F se torna finito e obtemos um conjunto F^* que contém os elementos: $F^* = \{0, 1, \alpha, \dots, \alpha^{2^m-2}\}$.

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

Exemplo 7: Seja $m = 4$. O polinômio $p(X) = 1 + X + X^4$ é um polinômio primitivo sobre $GF(2)$.

Estabelecendo $p(\alpha) = 0$, temos que $1 + \alpha + \alpha^4 = 0$, ou seja, $\alpha^4 = 1 + \alpha$. Usando esta relação, podemos construir $GF(2^4)$ como se segue:

$$\begin{aligned} \alpha^5 &= \alpha \cdot \alpha^4 = \alpha \cdot (1 + \alpha) = \alpha + \alpha^2 \\ \alpha^6 &= \alpha \cdot \alpha^5 = \alpha \cdot (\alpha + \alpha^2) = \alpha^2 + \alpha^3 \\ \alpha^7 &= \alpha \cdot \alpha^6 = \alpha \cdot (\alpha^2 + \alpha^3) = \alpha^3 + \alpha^4 = \alpha^3 + 1 + \alpha = 1 + \alpha + \alpha^3 \\ \alpha^8 &= \alpha \cdot \alpha^7 = \alpha \cdot (1 + \alpha + \alpha^3) = \alpha + \alpha^2 + \alpha^4 = \alpha + \alpha^2 + 1 + \alpha = 1 + \alpha^2 \\ \alpha^9 &= \alpha \cdot \alpha^8 = \alpha \cdot (1 + \alpha^2) = \alpha + \alpha^3 \\ \alpha^{10} &= \alpha \cdot \alpha^9 = \alpha \cdot (\alpha + \alpha^3) = \alpha^2 + \alpha^4 = \alpha^2 + 1 + \alpha = 1 + \alpha + \alpha^2 \\ \alpha^{11} &= \alpha \cdot \alpha^{10} = \alpha \cdot (1 + \alpha + \alpha^2) = \alpha + \alpha^2 + \alpha^3 \\ \alpha^{12} &= \alpha \cdot \alpha^{11} = \alpha \cdot (\alpha + \alpha^2 + \alpha^3) = \alpha^2 + \alpha^3 + \alpha^4 = 1 + \alpha + \alpha^2 + \alpha^3 \\ \alpha^{13} &= \alpha \cdot \alpha^{12} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^3) = \alpha + \alpha^2 + \alpha^3 + 1 + \alpha = 1 + \alpha^2 + \alpha^3 \\ \alpha^{14} &= \alpha \cdot \alpha^{13} = \alpha \cdot (1 + \alpha^2 + \alpha^3) = \alpha + \alpha^3 + \alpha^4 = \alpha + \alpha^3 + 1 + \alpha = 1 + \alpha^3 \end{aligned}$$

A Tabela 8 apresenta a representação por potência, a representação polinomial e a representação vetorial da extensão de $GF(2^4)$.

Tabela 8 – $GF(2^4)$ gerado por $p(X) = 1 + X + X^4$.

Representação por Potência	Representação Polinomial	Representação 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}	$1 + \alpha + \alpha^2$	(1100)
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)

2.3 Sistema de Comunicação Digital

Esta Seção tem como objetivo apresentar uma descrição geral sobre sistemas de comunicação, presente na Engenharia de Telecomunicações ⁶, cuja abrangência se estende em diversas áreas no cotidiano, sendo as mais comuns: sistemas de comunicações via satélite, sistema de telefone móvel, sistema de propagação via rádio e redes de telecomunicações. O estudo acerca destes conceitos podem ser encontrados em Faria e Palazzo Júnior (2011), Lin e Costello (1983) e Shannon (1948).

2.3.1 Sistema de Comunicação

A transmissão da informação é a motivação do desenvolvimento dos códigos corretores de erros (CCE's) porém, sua aplicação e realização estão totalmente vinculadas à teoria das comunicações. Shannon (1948) publicou o artigo intitulado A Mathematical Theory of Communication, tal publicação enfatizou os aspectos fundamentais do sistema de comunicação a princípio simples mas extraordinariamente eficiente na detecção e na resolução dos problemas técnicos.

Podemos considerar que um sistema de comunicação é um conjunto de meios físicos, equipamentos, ou até mesmo de um organismo, cujo principal objetivo é transferir dados de uma fonte de informação para um determinado destino (destinatário), de maneira confiável,

⁶ Engenharia de Telecomunicações: é uma área de especialização da Engenharia Elétrica e da Engenharia Eletrônica.

permitindo que a mensagem seja recebida de forma fidedigna à informação original, por meio de um canal de comunicação.

É importante saber como são os modelos de redes de sistema de comunicação. A **fonte** gera símbolos a serem transmitidos, pertencentes a um alfabeto finito que na maioria dos casos é binário, o **emissor** possui o papel de converter esses dados em sinais adequados, por meio de processos de modulação e codificação, para o **sistema de transmissão**, presente entre a fonte e o destinatário, que por sua vez transforma os dados em sinais, transportando-os para o **receptor**, que transforma os sinais novamente em dados, que chegam ao **destino** que consome os dados.

Durante essa transmissão temos o canal atuando como um filtro, fazendo com que o sinal seja transmitido de forma correta, distorcendo sua forma de onda. Todavia, pode-se ocorrer no processo de transmissão de sinal efeitos indesejáveis que afetam o sinal, por meio de interferências, ruídos e distorções que impedem a passagem dos sinais para o receptor fazendo com que eles cheguem de modo indesejado.

A **distorção** é a resposta imperfeita do sistema ao sinal utilizado e desaparece quando o mesmo é desligado. A **interferência** é uma contaminação do sinal que ocorre devido a fontes humanas, que podem ser linhas cruzadas que atrapalham a transmissão dos sinais, os **ruídos** são sinais aleatórios que podem afetar de tal forma a transmissão que pode ocorrer perda total dos sinais.

Os elementos de um sistema de comunicação são apresentados na Figura 7.

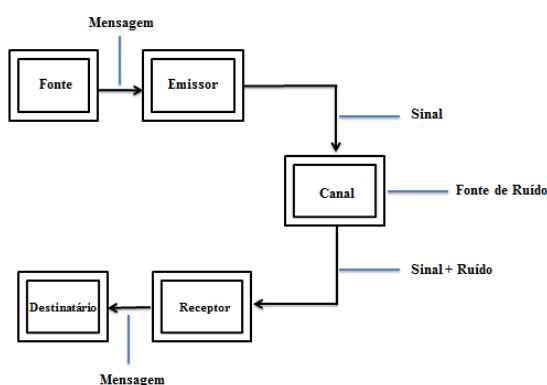


Figura 7 – Elementos básicos de um sistema de comunicação.

O modelo do sistema de comunicação digital é constituído pela inserção do codificador da fonte, codificador de canal, modulador, demodulador, decodificador de canal e decodificador de fonte, sendo representado através de um diagrama de blocos, como mostra a Figura 8. O objetivo deste sistema é conectar uma fonte a um destinatário, onde, cada um dos blocos componentes de um sistema de comunicação é definido da seguinte maneira:

1. **Fonte:** gera a informação (dados) a transmitir (máquina, ser humano), podendo gerar uma sequência de símbolos discretos ou um sinal.
2. **Codificador de fonte:** converte o sinal da saída da fonte em uma sequência de dígitos binários, extraíndo o excesso de redundância de informação e como consequência produz uma sequência de informação com o objetivo de deixar a transmissão mais confiável.
3. **Codificador de canal:** transforma a sequência da saída do codificador de fonte em uma palavra-código pela adição de redundâncias, cujo objetivo é combater os efeitos do ruído introduzido através do canal. Cada símbolo da sequência codificada (palavra-código), é representado por dígitos binários na sinalização binária. Em contrapartida, no uso de mais do que dois sinais (por exemplo: q sinais), não temos bits, e sim dígitos de um alfabeto q -ário.
4. **Modulador:** tem a função de converter os símbolos digitais da saída do codificador de canal em forma de ondas analógicas para a transmissão.
5. **Canal:** é o meio físico que conecta o transmissor ao receptor. O sinal modulado a ser transmitido através do canal pode ou não sofrer a ação do ruído.
6. **Demodulador:** com a recepção do sinal vindo do canal, o demodulador converte a sequência de forma de ondas corrompidas pelo ruído em uma sequência de símbolos para o decodificador de canal.
7. **Decodificador de canal:** decodifica a sequência produzida pelo demodulador de maneira mais confiável (com a menor probabilidade de erro, se houver), produzindo uma estimativa dos dígitos na saída do decodificador de fonte.
8. **Decodificador de fonte:** transforma a sequência estimada na saída do decodificador de canal em uma estimativa na saída da fonte, reproduzindo os dados emitidos pela fonte de informação.
9. **Destinatário:** é o receptor da informação transmitida (ser humano, máquina ou até mesmo um organismo).

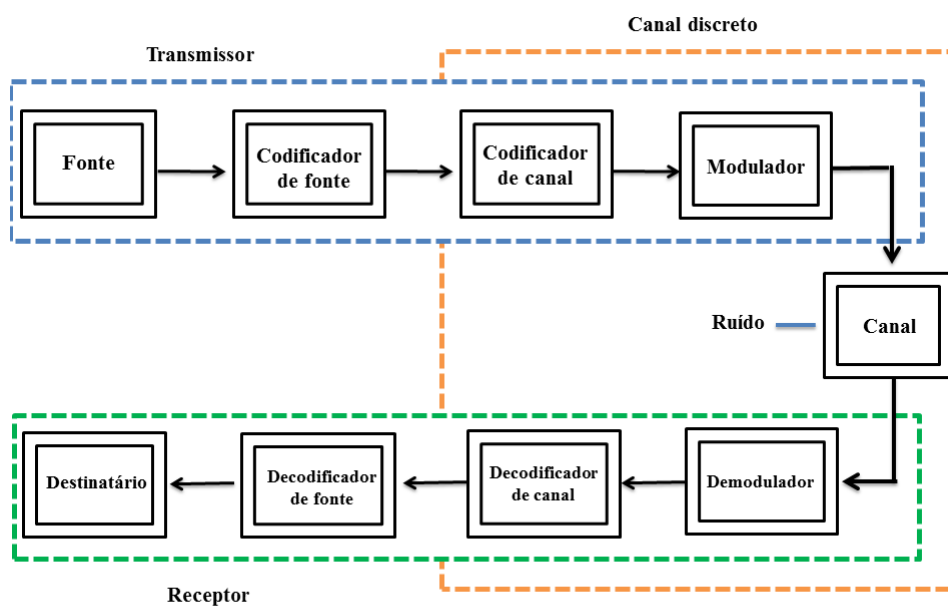


Figura 8 – Diagrama de blocos de um sistema de comunicação.

Fonte: Adaptado de Faria e Palazzo Júnior (2011).

3 Conexões entre os Dogmas da Teoria de Comunicações e Biologia Molecular

De acordo com Faria e Palazzo Júnior (2011), um dos maiores desafios em teorias da informação genética, comunicação genética e codificação genética, dentro da comunidade científica, é analisar a existência de uma estrutura matemática relacionada com o DNA. Sendo assim, é necessário utilizar um modelo de sistema de comunicação padrão com o sistema de comunicação genética, cujo objetivo é identificar, reproduzir e classificar matematicamente sequências distintas de DNA.

O objetivo deste capítulo é apresentar as conexões existentes entre a teoria de comunicação padrão e a teoria de comunicação genética, analisando e comparando métodos teóricos de codificação, mostrando um elo entre sistemas de comunicação, estruturas algébricas e elementos biológicos.

Os conceitos, definições e propriedades apresentados podem ser encontrados de forma detalhada em: Gatlin (1972), Rolman-Roldan (1996) May et al. (2004), Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011).

3.1 O Dogma Central da Teoria de Comunicações

No Capítulo 2, foi apresentado de forma detalhada, o diagrama de blocos de um sistema de comunicação, onde toda transmissão da informação está sujeita a erros, devido às imperfeições e ruídos, alocados no canal.

O diagrama de bloco principal no dogma central da teoria de comunicações, está ilustrado na Figura 9.

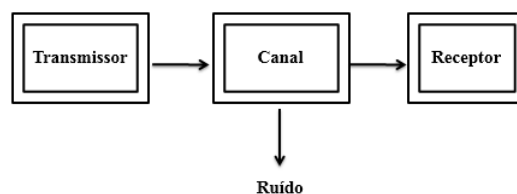


Figura 9 – O dogma central da teoria de comunicações.

Podemos conceituar os blocos deste sistema de comunicação como:

1. **Transmissor:** gera a informação, transmitida através de uma mensagem. Esta

mensagem pode ser uma voz, um sinal elétrico, ou quaisquer outras formas que a informação possa assumir.

2. **Canal:** é o meio físico que conecta o transmissor ao receptor, através do qual a informação será transmitida, podendo assumir a forma de uma linha telefônica e até mesmo um organismo biológico, dentre outros. O sinal modulado a ser transmitido através do canal sofre a ação do ruído, tornando-se umas das grandes preocupações da codificação, a fim de proteger a informação contra erros na transmissão.
3. **Receptor:** é o usuário que recebe a mensagem transmitida pelo canal, podendo ser uma pessoa, uma máquina ou até mesmo um organismo vivo qualquer.

Através do sistema de comunicação padrão, muitos pesquisadores exploram as conexões com o fluxo da informação genética, com o objetivo de modelar diversos sistemas biológicos.

3.2 O Dogma Central da Biologia Molecular

Francis Crick, em 1958, descreveu o dogma central da biologia molecular, visando relacionar o DNA, RNA e as proteínas. Na biologia molecular, o processamento de informações se dá através da replicação do DNA e é traduzida através de dois processos:

1. **Transcrição:** converte a informação do DNA em moléculas de RNA.
2. **Tradução:** converte a informação do RNA em proteínas.

No Capítulo 2, analisamos separadamente os processos de duplicação do DNA, o processo de transcrição e o processo de tradução, caracterizando o esquema da síntese proteica, no dogma central da biologia molecular, ilustrado na Figura 10.

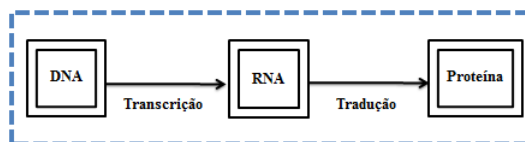


Figura 10 – Esquema da síntese proteica.

3.3 Analogias entre Sistemas de Comunicações e Sistemas de Informação Genética

Através das informações fornecidas anteriormente, apresentamos a seguir, as semelhanças existentes entre o dogma central da teoria de comunicações e o dogma central da biologia molecular, fazendo as seguintes associações:

1. No sistema de comunicação, quem gera a informação a ser transmitida é o transmissor. Biologicamente, o responsável por esta função é o DNA.
2. Do ponto de vista da comunicação, os processos de transcrição e tradução, representam os processos de codificação de canal e de modulação, e os eventuais erros, devido ao ruído introduzido no canal. Biologicamente, os processos de transcrição e tradução têm como objetivo a transmissão da informação. Neste processo podem ocorrer erros, ocasionando interferências na informação, como por exemplo, a não-leitura de um códon, que acarreta na perda do pareamento do ribossomo. Pode-se relacionar tais erros a processos mutacionais.
3. O receptor pode ser modelado como o local onde a informação está sendo enviada, no sistema de comunicação. Biologicamente, especificamente neste caso, a informação é a proteína.

Sendo assim, podemos adotar o sistema de comunicação da informação genética, através de um diagrama de blocos, apresentado na Figura 11.

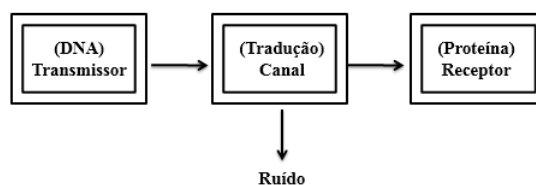


Figura 11 – Sistema de comunicação de informação genética.

3.4 Comparação e Análise de Sistemas de Comunicação Biológico

A premissa central da genética é a de que os genes são perpetuados sob a forma de sequências de ácidos nucleicos e gerados como proteínas (MAY et al.(2004)). Diversos investigadores (GATLIN (1972); ROLMAN-ROLDAN(1996); MAY et al (2004); MAY (2004); ROCHA e PALAZZO JÚNIOR(2010); FARIA e PALAZZO JÚNIOR (2011)), desenvolveram modelos que tentam capturar várias informações sobre aspectos teóricos do sistema genético.

Nesta Seção, apresentamos alguns dos importantes modelos para os sistemas de comunicação biológico, evidenciando as semelhanças e diferenças entre os modelos analisados.

Foram analisados os modelos propostos por Gatlin, Roman-Roldan, May, Rocha e Faria, encontrados em Gatlin (1972), Roman-Roldan (1996), May (2004), May et al. (2004), Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011).

3.4.1 Modelo de Comunicação de Gatlin

Gatlin (1972) apresenta um dos trabalhos pioneiros sobre as propriedades teóricas da informação em sistemas genéticos. A Figura 12 apresenta a visão da teoria da comunicação proposta por Gatlin do sistema genético. Em seu modelo, as sequências de bases do DNA representam a mensagem codificada gerada por uma fonte, posteriormente, o DNA codificado passa por um canal (definido pela transcrição e tradução), que Gatlin refere-se a todos os mecanismos para produzir proteínas. A sequência de aminoácidos é a mensagem recebida, que refere-se ao receptor/destinatário.

Em seu trabalho, não é explícito onde se encaixa a replicação ou se ela considera o processo de replicação como parte do codificador. Apesar da existência de erros de transcrição e de tradução, a replicação também introduz erros, sendo uma fonte significativa de ruído que deve ser abordada separadamente.

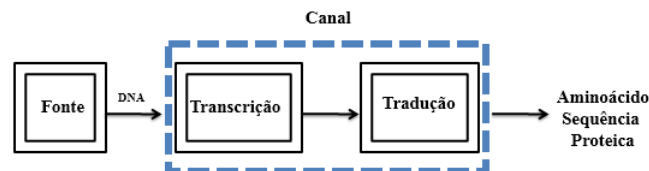


Figura 12 – Modelo proposto por Gatlin.

Fonte: Adaptado de May, et al. (2004).

3.4.2 Modelo de Comunicação de Roman-Roldan

Roman-Roldan et al. (1996), apresenta o modelo de informação biológica como um canal de comunicação, que contempla a sequência de DNA como a entrada e as sequências de aminoácidos (forma a proteína) como a saída do canal, ou seja, seu modelo consiste em modelar o processo biológico relacionado à síntese de proteínas, conforme apresentado na Figura 13.

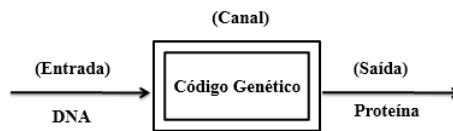


Figura 13 – Modelo proposto por Romam-Roldan.

Fonte: Adaptado de May, et al. (2004).

Roman-Roldan também designa o código genético, como o processo de mapeamento dos códons para aminoácidos, como o canal de transmissão no qual o DNA é transmitido e a proteína é recebida. Também não aborda o papel da replicação do DNA na informação genética na transmissão.

3.4.3 Modelo de Comunicação de May

O objetivo do modelo proposto por May é interpretar o mecanismo de iniciação da tradução em organismos procarióticos. Sendo assim, May et al. (2004), define em seu modelo o RNAm como a saída do canal de comunicação e incorpora um decodificador que traduz o RNAm em cadeias de aminoácidos formadora de proteína. Originalmente, o canal consiste na replicação do DNA e no processo de transcrição durante os quais erros são introduzidos na sequência nucleotídica. Este modelo é apresentado na Figura 14.

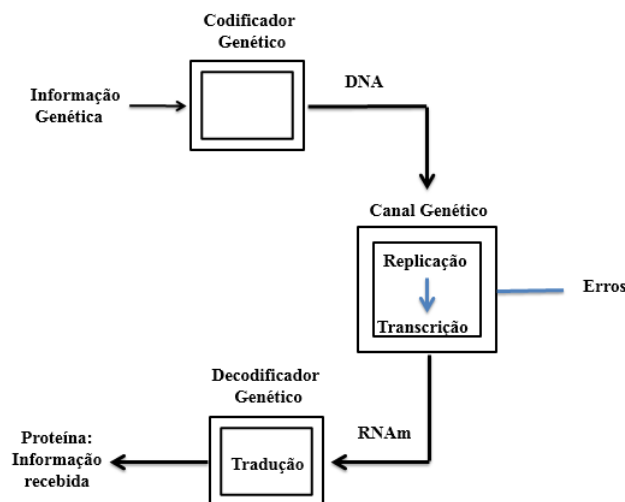


Figura 14 – Modelo proposto por May.

Fonte: Adaptado de May, et al. (2004).

Do ponto de vista da comunicação, as mutações (transferidas do pai para a prole), podem ser vistas como os erros ocorridos na transmissão, quando corrompem a mensagem

original. Durante a replicação, os erros que podem ocorrer são: eliminação de bases de DNA e a inserção de bases de DNA incorretas. Tais erros ou mutações no código genético pode corromper, e distorcer o sinal genético, que é vital para a sobrevivência do organismo (MAY, 2004).

3.4.4 Modelo de Comunicação de Rocha

Rocha e Palazzo Júnior (2010), propõem um modelo de comunicação biológico para a importação de proteínas mitocondriais, conforme a Figura 15, com o objetivo de identificar estruturas matemáticas associadas às sequências de DNA. O processo de codificação é realizado pelo bloco transmissor, e o processo de decodificação é realizado pelo bloco receptor.

O codificador é o responsável pelas transformações das bases nitrogenadas $\{A, C, G, T\}$ em um alfabeto matemático $\{0, 1, 2, 3\}$, usado pelo código através da estrutura algébrica de anel, estabelecendo o mapeamento. O código BCH refere-se as palavras-códigos na saída do codificador que está relacionada à sequência de direcionamento em termos de nucleotídeos, construídas através de extensões de Galois, enquanto que o modulador consiste do código genético, do RNAt e do ribossomo e, a palavra-código na saída se refere à sequência de direcionamento em termos de aminoácidos.

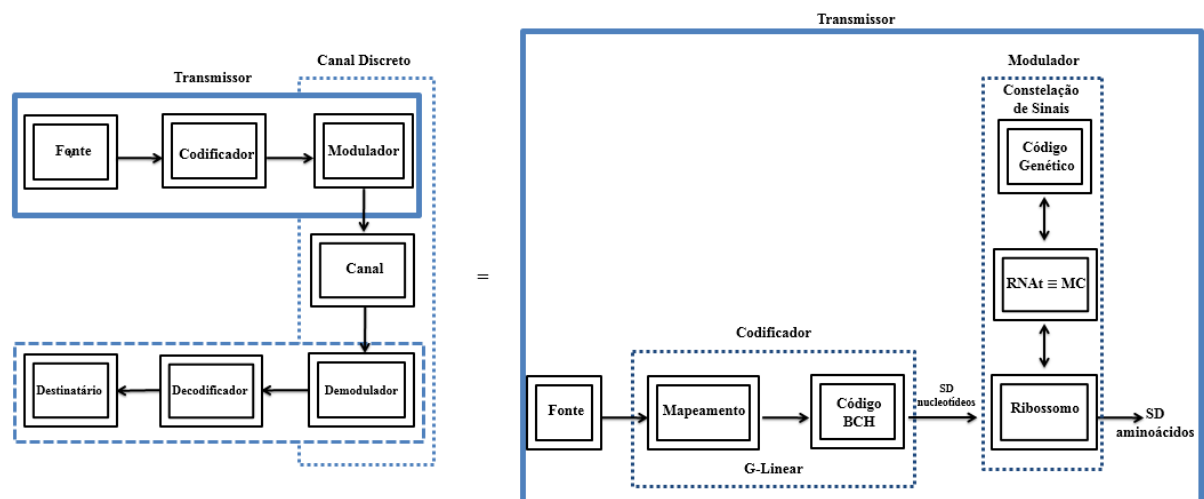


Figura 15 – Modelo proposto por Rocha.

Fonte: Adaptado de Rocha e Palazzo Júnior (2010).

3.4.5 Modelo de Comunicação de Faria

Faria e Palazzo Júnior (2011) propõem um modelo de sistema de comunicação de informação genética e um sistema de comunicação de informação genômica análogos ao modelo de um sistema de comunicação digital, Figuras 16 e 17, respectivamente.

O modelo de informação genética é constituído de um **codificador genético** e de um **modulador genético**. A palavra código na saída do codificador relaciona-se com a sequência de nucleotídeos (RNAm) e na saída do modulador com a sequência dos aminoácidos (proteína). A função do RNAt é realizar o mapeamento do código genético, porém é necessária uma caracterização matemática no contexto de um sistema de comunicação digital, onde os 64 códons representam os sinais da constelação de sinais.

O modelo de informação genômica associa-se com o bloco do codificador de canal. Portanto, o processo de codificação está relacionado com os possíveis erros nesse canal, sendo esse modelo responsável pela identificação e reprodução na informação genômica contida na dupla hélice do DNA.

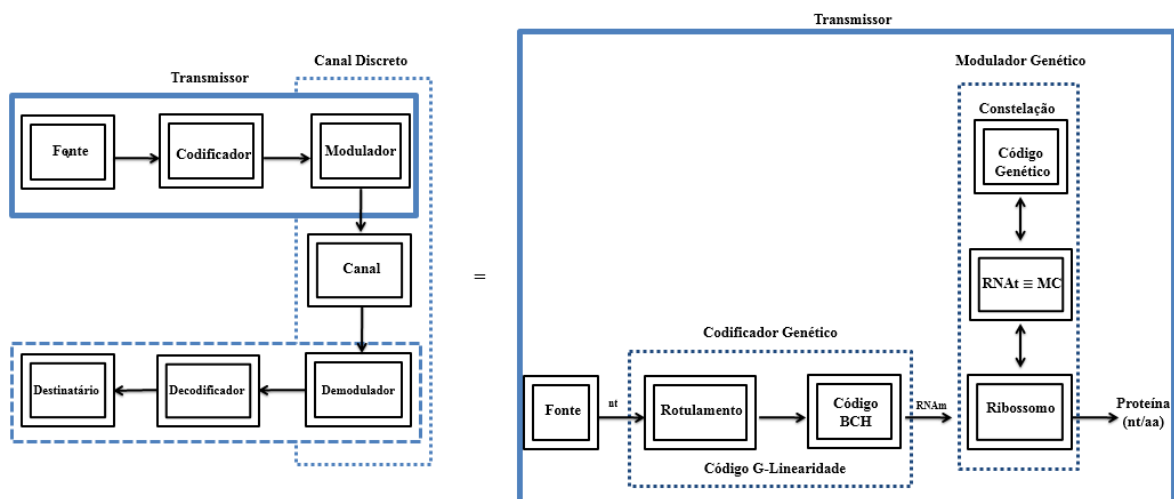


Figura 16 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genética.

Fonte: Adaptado de Faria e Palazzo Júnior (2011).

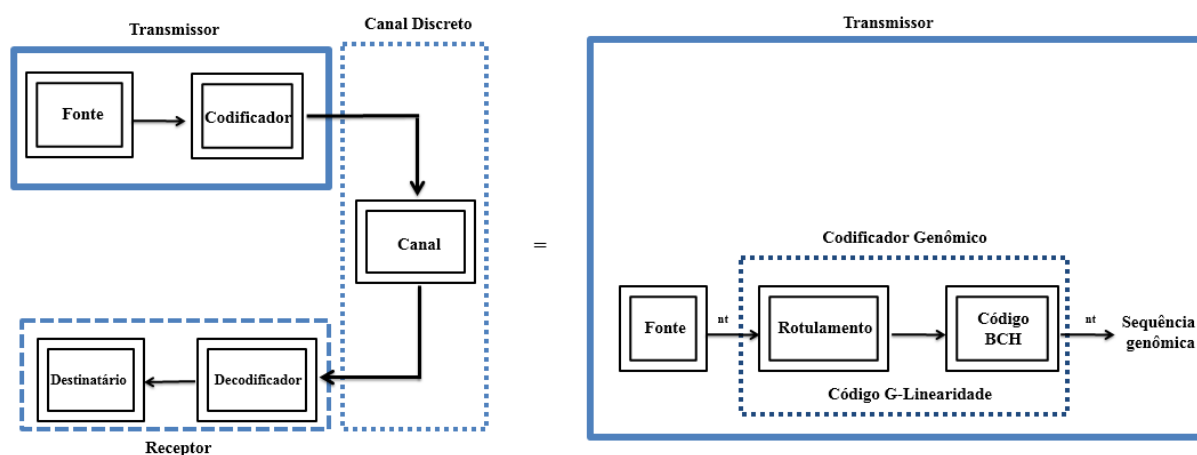


Figura 17 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genômica.

Fonte: Adaptado de Faria e Palazzo Júnior (2011).

3.4.6 Comparações dos Modelos Analisados

No modelo proposto por Gatlin, percebe-se que não deixa claro onde ocorre o processo de replicação, apenas destaca que este processo é uma fonte significativa de ruído, que deve ser abordada separadamente. Contudo, além da sua visão teórica da informação, gerada em processos genéticos, ele também modela para um programa de computador, que não é a ênfase deste trabalho.

Semelhante a Gatlin, Rolman-Roldan aborda a sequência de DNA como a entrada do canal, e a proteína como a saída. Assume-se que o modelo de Rolman-Roldan é totalmente paralelo ao de Gatlin pois, não aborda o papel da replicação durante a transmissão. Ambos abordam o ruído como mecanismos genéticos responsáveis pela síntese proteica (transcrição e tradução). Ressalta-se ainda que, não é explícito a presença ou a ausência de redundâncias na entrada do canal (DNA), no modelo de Rolman-Roldan.

As propostas de Rolman-Roldan e de Gatlin, se diferem da proposta de May pois, ela caracteriza o RNAm como a saída do canal de comunicação e, seu modelo aborda o processo de replicação do DNA, como sendo o canal que introduz os erros, o que não é explícito nos outros modelos. Contudo, May não aborda os erros de transcrição e tradução e, embora eles não se propaguem com os erros da replicação, é necessária uma análise mais detalhada de seus impactos no sistema de comunicação genética.

Rocha se difere de todas as propostas, uma vez que ela apresenta um modelo de sistema de comunicação para a importação de proteínas, que consiste em caracterizar os processos de codificação e de decodificação no direcionamento dessas proteínas. Conforme mencionado, em seu modelo, a palavra-código na saída codificador se refere aos nucleotídeos

e na saída do modulador se refere aos aminoácidos. O mapeamento realizado em seu modelo consiste da associação das bases nitrogenadas do DNA e do RNA com a estrutura algébrica de anel. Os erros ocorridos no processo de codificação estão localizados em um canal discreto.

A proposta de Faria é bem semelhante à proposta apresentada em Rocha. Entretanto, Faria expande o mapeamento para um rotulamento presente no codificador genético, que consiste em 24 possibilidades (conjunto das bases nitrogenadas do DNA/RNA) e, além de fazer a associação dessas bases nitrogenadas com a estrutura algébrica de anel, também é feita a associação com a extensão de Galois de grau 2 de $GF(2)$. A caracterização dos modelos apresentados por Faria está vinculada tanto ao processo de identificação da estrutura matemática, quanto ao alfabeto do código corretor de erro (CCE), cujo objetivo é reproduzir, identificar e classificar as sequências de DNA.

4 Caracterização Algébrica do Modelo de Codificação Genética proposto por Faria (2011)

Neste Capítulo apresentamos os principais elementos para a determinação e a caracterização dos modelo de codificação genética que é semelhante a um sistema de comunicação. Este modelo é responsável pela identificação e reprodução da informação genética que está contida na sequência de nucleotídeos de uma fita simples de DNA. A caracterização deste modelo está vinculada ao processo de identificação da estrutura matemática quanto ao alfabeto do código corretor de erro cujo objetivo é reproduzir, identificar e classificar as sequências de DNA.

Estes conceitos podem ser encontrados de forma detalhada em Faria e Palazzo Júnior (2011), Matoso e Queiroz (2013) e Rocha e Palazzo Júnior (2010).

4.1 Modelo de Codificação Genética e Genômica

A codificação genética e codificação genômica, leva ao questionamento se existe alguma forma de CCE na estrutura do DNA. Alguns autores afirmam que isso seria um problema sem solução. Battail (2008), considera a identificação de algum CCE que seja capaz de reproduzir uma determinada sequência de DNA como um problema insolúvel. Nesse sentido, a proposta deste trabalho é analisar um encaminhamento positivo a esse problema, de acordo com os pressupostos de Faria (2011). Um grande avanço é que as sequências de nucleotídeos e dos correspondentes aminoácidos de uma fita simples do DNA com características biológicas distintas e comprimentos variados são identificadas como palavras-código de CCEs.

Diante disso, apresenta-se um modelo de sistema de comunicação de informação genética, como mostra a Figura 18, e um modelo de um sistema de comunicação de informação genômica, como mostra a Figura 19, análogos ao modelo de um sistema de comunicação digital. O modelo de um sistema de comunicação de informação genética está diretamente relacionado com o desmembramento do canal discreto. A caracterização do modulador e do demodulador, fundamenta o contexto da informação biológica trafegar de um ponto para outro intracelular (dentro da célula) ou intercelular (entre a célula). O codificador e o modulador, juntos, são responsáveis pela identificação e reprodução da informação genética contida no RNA mensageiro (RNAm).

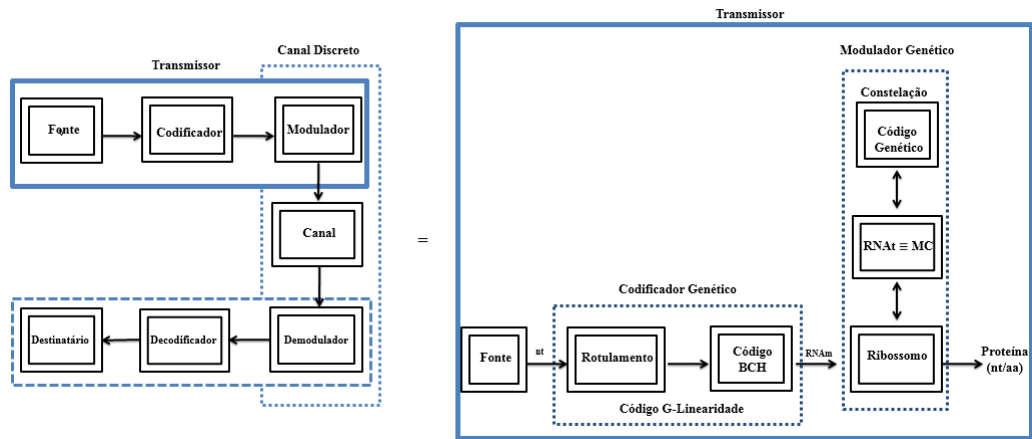


Figura 18 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genética.

Um sistema de comunicação de informação genômica associa-se com o bloco do codificador de canal. O modulador, canal e demodulador juntam-se em um *canal discreto* podendo ter ou não memória. Portanto, o processo de codificação está relacionado com os possíveis erros nesse canal, sendo esse modelo responsável pela identificação e reprodução da informação genômica contida na dupla hélice do DNA.

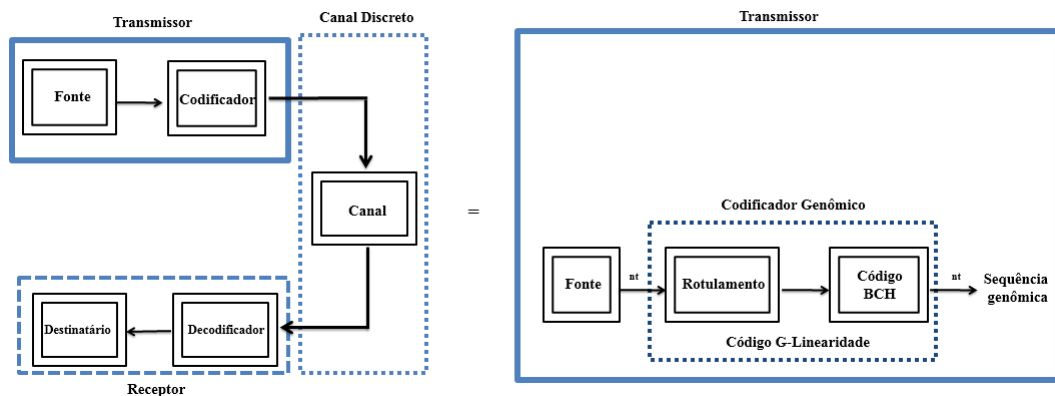


Figura 19 – Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genômica.

4.1.1 Modelo de Codificação Genética

Este modelo é constituído de um **codificador genético** e de um **modulador genético**. A palavra código na saída do codificador relaciona-se com a sequência de nucleotídeos (RNAm) e na saída do modulador com a sequência dos aminoácidos (proteína), Figura 20. Embora seja bastante conhecido no contexto biológico, o RNAt é quem realiza o mapeamento do código genético, porém é necessária uma caracterização matemática no

contexto de um sistema de comunicação digital, onde os 64 possíveis códons (as trincas dos aminoácidos) representam os sinais da constelação de sinais. Estes 64 códons podem ser relacionados como extensões de Galois $GF(2^6)$.

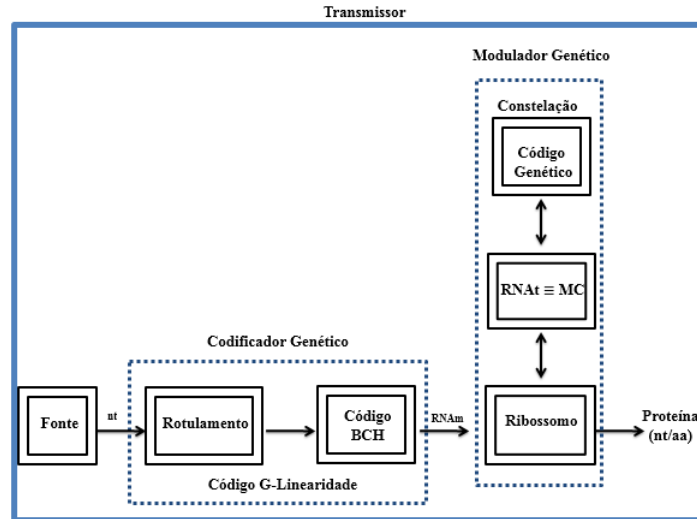


Figura 20 – Modelo de um sistema de comunicação de informação genética.

Quando temos uma comunicação digital isomorfa, ou seja, quando as estruturas algébricas do codificador e do modulador é um isomorfismo, denota-se como um **mapeamento casado** (MC). Deste modo, o codificador genético consiste de um bloco mapeamento associado a um código corretor de erros (bloco rotulamento e bloco código BCH).

O modulador genético consiste do código genético, do RNA transportador e do ribossomo. É importante ressaltar que o código genético pode ser visto como uma constelação de sinais, onde cada códon representa um sinal nessa constelação.

Com o objetivo de facilitar o entendimento do processo de caracterização do modelo de um sistema de comunicação de informação genética, conforme ilustrado na Figura 20, serão apresentados à descrição do desdobramento do bloco do codificador genético.

4.1.2 Elementos Algébricos Associados ao Codificador Genético

A utilização dos códigos BCH iniciou-se com o objetivo de identificar e reproduzir diferentes sequências de DNA no codificador genético. Primeiramente constrói-se o código BCH na estrutura algébrica de anel, em seguida em estrutura de corpo e extensões de Galois.

1. Determinação da Estrutura Algébrica do Alfabeto do Código Genético

Nesse processo de identificação da estrutura algébrica são pertinentes:

- a) O alfabeto do código genético e o alfabeto do código BCH:

No processo de codificação genética, o alfabeto 4-ário do código genético está relacionado ao conjunto formado pelos nucleotídeos denotado por $N = \{A, C, G, T/U\}$ correspondendo à adenina (A), citosina (C), guanina (G) e timina (T) ou uracila (U). Por esta razão utilizamos o alfabeto 4-ário que corresponde ao $GF(4) = \{0, 1, \alpha, \alpha^2\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, como apresenta a Figura 21.

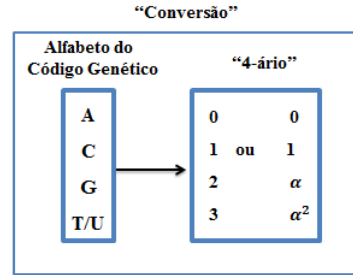


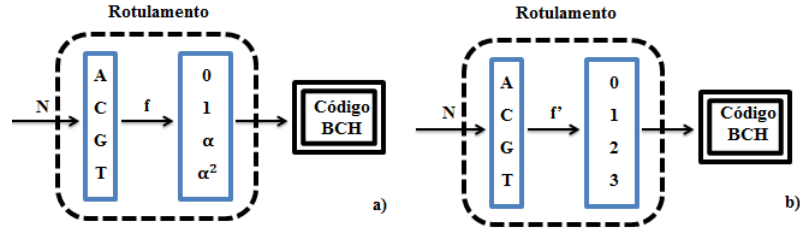
Figura 21 – O alfabeto do código genético e os alfabetos 4-ários dos CCEs.

Como a estrutura algébrica do alfabeto do código genético das sequências de DNA apresentado é desconhecida, é necessário utilizar um processo de “conversão” do alfabeto do código genético para o alfabeto utilizado nos CCEs, onde os alfabetos 4-ários dos CCEs estão relacionados ao conjunto $N = \{A, C, G, T/U\}$.

Inicialmente, associa-se os elementos do conjunto $N = \{A, C, G, T/U\}$ com os elementos dos conjuntos $GF(4) = \{0, 1, \alpha, \alpha^2\}$ e posteriormente com o conjunto $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, como mostrado na Figura 22 a) e b), respectivamente.

Conforme mencionado, considera-se o alfabeto do código $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ como uma estrutura algébrica de anel, obedecendo as operações de adição e multiplicação módulo 4. Na estrutura algébrica de corpo considera-se o alfabeto do código $GF(4) = \{0, 1, \alpha, \alpha^2\}$ obedecendo as operações de adição e multiplicação módulo $(x^2 + x + 1)$.

Sendo $m = 2$, podemos considerar $(x^2 + x + 1)$ como um polinômio sobre $GF(2)$, $p(x) = 1 + x + x^2$. Admitindo que α seja uma raiz do polinômio, então $p(\alpha) = 0$, ou seja, $0 = 1 + \alpha + \alpha^2 \implies \alpha^2 = 1 + \alpha$. A partir desta relação pode-se construir um $GF(2^2)$ como se segue: $\alpha^3 = \alpha \cdot \alpha^2 = \alpha \cdot (\alpha + 1) = \alpha^2 + \alpha = \alpha + 1 + \alpha = 1$


 Figura 22 – Associações entre $N \longrightarrow GF(4)$ e $N \longrightarrow \mathbb{Z}_4$.

b) A Estrutura de Espaço Vetorial Sobre Corpo $GF(2)$:

Considere $GF(2^2) = \{0, 1, \alpha, \alpha^2\}$ como a extensão de Galois de grau 2 de $GF(2)$, isto é,

$$GF(2^2)[x] \cong \frac{GF(2)[x]}{\langle p(x) \rangle} = \frac{GF(2)[x]}{\langle x^2+x+1 \rangle} = \{a_0 + a_1x; a_0, a_1 \in GF(2)\},$$

onde $p(x)$ é o polinômio primitivo considerado. Note que o par a_0a_1 assume os valores 00, 10, 01 e 11. Esses elementos binários estão associados aos elementos de $GF(4)$, sendo 00–0, 10–1, 01– α e 11– α^2 , conforme ilustrado na Figura 23.

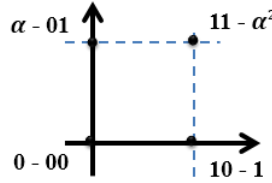


Figura 23 – Representação da estrutura de espaço vetorial sobre o corpo.

c) Módulo (Espaço Vetorial) sobre Anel:

A associação de espaço vetorial ao conjunto $N = \{A, C, G, T/U\}$, proveniente dos números complexos, é mostrada através da raiz quarta da unidade, ou seja, $x^4 - 1$, sendo a solução:

$$S = \{e^{j\frac{0\pi}{2}}, e^{j\frac{1\pi}{2}}, e^{j\frac{2\pi}{2}}, e^{j\frac{3\pi}{2}}\}.$$

Observe que S é um grupo multiplicativo, cujo elemento gerador é $e^{j\frac{\pi}{2}}$. Por esta razão, o que se nota é que existe um isomorfismo entre o grupo multiplicativo S e o grupo aditivo (módulo 4) S' , denotado por:

$$S = \{e^{j\frac{0\pi}{2}}, e^{j\frac{1\pi}{2}}, e^{j\frac{2\pi}{2}}, e^{j\frac{3\pi}{2}}\} \longrightarrow S' = \{0, 1, 2, 3\}.$$

Assim como S e S' são isomorfos segue que, o espaço vetorial associado ao \mathbb{Z}_4 é como ilustrado na Figura 24.

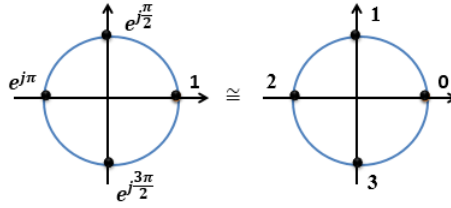


Figura 24 – Representação de Espaço Vetorial sobre o Anel.

Com relação ao espaço vetorial, estruturas algébricas como corpo e anel, são introduzidas no bloco codificador. Porém, estas não são as únicas estruturas algébricas existentes, pois estamos apresentando quais as possibilidades dessas estruturas matemáticas que estão associados ao código genético que sejam capazes de identificar sequências de DNA, ou seja, a inserção de um alfabeto associado a um código.

Para caracterizar o código corretor de erro presente no codificador genético, é necessário o desdobramento do rotulamento, bem como, a identificação do mapeamento do corpo e do anel trabalhado, conforme ilustrado na Figura 25.

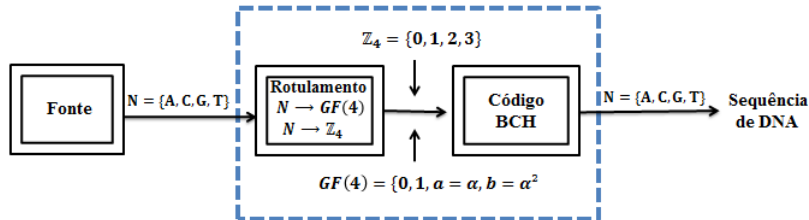


Figura 25 – Codificador genético.

2. Determinação do rotulamento/mapeamento no codificador genético

Todas as 24 possibilidades de associações dos elementos do conjunto $N = \{A, C, G, T/U\}$ com os elementos dos conjuntos $GF(4) = \{0, 1, a = \alpha, b = \alpha^2\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ denominamos rotulamentos, são apresentados nas Figuras 26 e 27 respectivamente. O objetivo destes rotulamentos é a determinação da melhor associação de cada um dos símbolos do conjunto N (bases nitrogenadas) com os correspondentes símbolos nos conjuntos do corpo e do anel e vice-versa.

$N! = 4! = 24$ possibilidades de rotulamento para $GF(4) = \{0, 1, a, b\}$

$\begin{bmatrix} A & C & G & T \\ 0 & 1 & b & a \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ b & 0 & a & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ a & 1 & 0 & b \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & b & 0 & a \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 0 & b & 1 & a \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ b & a & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ a & b & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & b & a & 0 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 1 & 0 & a & b \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & 1 & a & b \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ b & 0 & 1 & a \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ a & 0 & 1 & b \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ 1 & a & 0 & b \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & b & a & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ b & a & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ a & 0 & b & 1 \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ a & 1 & b & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & 0 & b & a \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & a & 1 & b \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ b & 1 & 0 & a \end{bmatrix}$
$\begin{bmatrix} A & C & G & T \\ a & b & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 1 & a & b & 0 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ 0 & a & b & 1 \end{bmatrix}$	$\begin{bmatrix} A & C & G & T \\ b & 1 & a & 0 \end{bmatrix}$

Figura 26 – Possibilidades de rotulamento entre os elementos do conjunto N e $GF(4)$.

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

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

Figura 27 – Possibilidades de rotulamento entre os elementos do conjunto N e \mathbb{Z}_4 .

A estrutura algébrica das sequências de DNA e do CCE relaciona-se à estrutura do alfabeto do código genético, sendo identificada através da estrutura geométrica das sequências de DNA aos alfabetos binários dos CCEs, realizando o mapeamento entre os alfabetos 4 - ário e binário. Denominamos de mapeamento, a decomposição binária dos elementos de $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ e $GF(4) = \{0, 1, \alpha, \alpha^2\}$, como ilustrado na Figura 28.

Mapeamento $GF(4)$		Mapeamento \mathbb{Z}_4
$00 \leftrightarrow 0$		$00 \leftrightarrow 0$
$10 \leftrightarrow 1$		$10 \leftrightarrow 1$
$01 \leftrightarrow \alpha$	e	$11 \leftrightarrow 2$
$11 \leftrightarrow \alpha^2$		$01 \leftrightarrow 3$

Figura 28 – Representação do mapeamento de $GF(4)$ e \mathbb{Z}_4 .

É necessária a realização de um rotulamento entre os conjuntos $N \longleftrightarrow GF(4)$ e $N \longleftrightarrow \mathbb{Z}_4$. É de suma importância levarmos em consideração a complementariedade das bases nitrogenadas, obedecendo a regra de Chargaff, na análise destes rotulamentos da

seguinte maneira:

a) **Rotulamento/Mapeamento no $GF(4) = \{0, 1, \alpha, \alpha^2\}$:**

A Figura 23 representa um mapeamento linear proveniente da extensão do $GF(2)$ para o $GF(4)$. As sequências de DNA são rotuladas através das 24 permutações entre $N \rightarrow GF(4)$, sendo o rotulamento entre $N \rightarrow GF(4)$ desconhecido. Cada sequência possui 24 palavras-código que correspondem às 24 permutações. Já o rotulamento recíproco do conjunto $GF(4) \rightarrow N$, possui em cada uma dessas 24 palavras-código, 24 palavras-código iguais com relação aos aminoácidos e nucleotídeos. Na Figura 29 é representado o rotulamento S_4 que é análogo ao grupo das permutações no $GF(4)$.

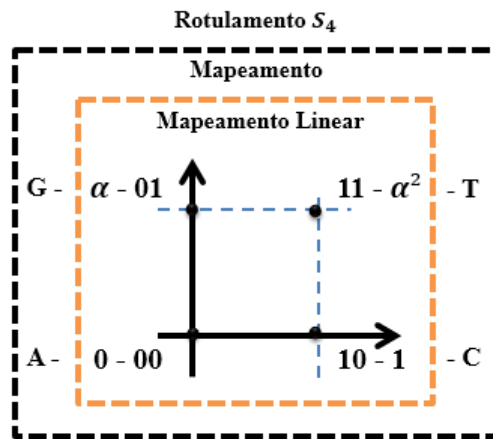


Figura 29 – Rotulamento e S_4 .

b) **Rotulamento/Mapeamento no $\mathbb{Z}_4 = \{0, 1, 2, 3\}$:**

Neste caso, como o mapeamento entre $N \rightarrow \mathbb{Z}_4$ é desconhecido, procedemos da seguinte maneira: toda sequência de DNA é rotulada pelas 24 permutações entre $N \rightarrow \mathbb{Z}_4$. Sendo assim, para cada sequência reproduzida pelo código existem oito palavras-código que correspondem às 8 permutações. Já o rotulamento recíproco do conjunto $\mathbb{Z}_4 \rightarrow N$, possui em cada uma dessas 8 palavras-código, 8 palavras-código iguais com relação aos aminoácidos e nucleotídeos, resultando em três conjuntos com oito permutações em cada um, conforme ilustrado na Figura 30.

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

Figura 30 – Rotulamentos A,B e C.

De acordo com Faria e Palazzo Jr. (2011, p.95):

A determinação dos rotulamentos (composição mapeamento/mapeamento) é proveniente da complementaridade biológica (A-T) e (C-G), que por sua vez deve prevalecer sobre complementaridade matemática (00-11) e (01-10).

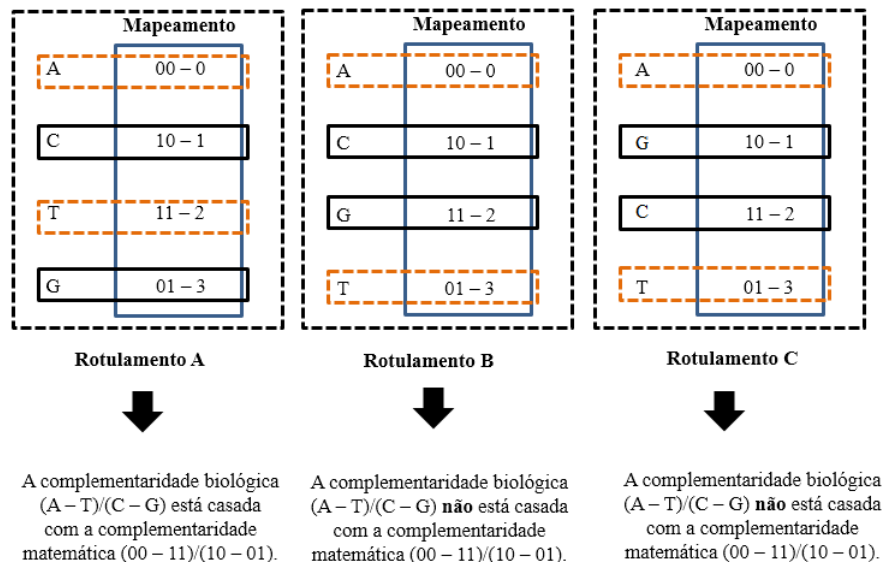
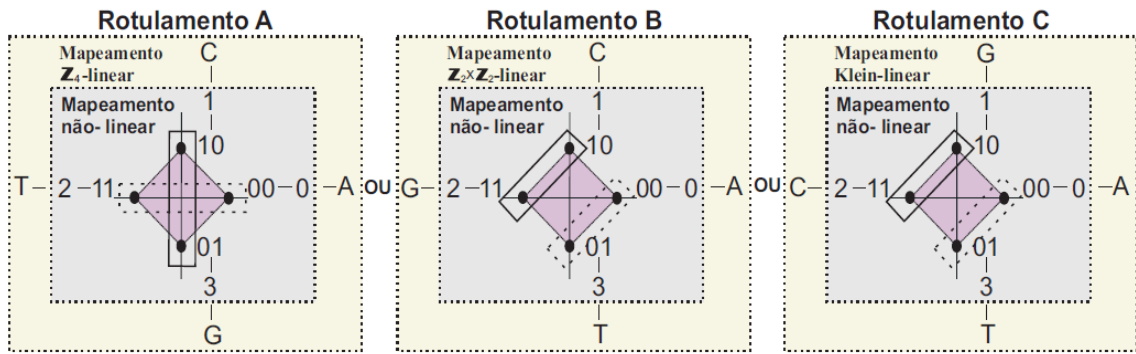


Figura 31 – Casamento entre contexto biológico e o contexto matemático.

Fonte: Adaptado de Faria e Palazzo Júnior (2011).

Todavia, essa complementariedade biológica pode ou não estar casada com a complementariedade matemática.


 Figura 32 – Mapeamentos do alfabeto \mathbb{Z}_4 .

Fonte: Faria e Palazzo Júnior (2011).

Observe na Figura 32 o mapeamento dos rotulamentos com relação ao conjunto \mathbb{Z}_4 . No caso do rotulamento A, vemos que qualquer um dos nucleotídeos necessita caminhar duas arestas para alcançar seu complementar, enquanto que no rotulamento B e C é necessário caminhar uma aresta somente.

4.1.3 Conexão entre Código Genético, Constelação de Sinais e Extensões de Galois

Nesta Seção, apresentamos o detalhamento da caracterização algébrica do modelo de sistema de comunicação de informação genética, proposto por Faria e Palazzo Júnior (2011), em relação ao código genético presente no modulador, que é caracterizado por uma constelação de sinais (Figura 33) que representa os 64 possíveis códons (as trincas dos aminoácidos), conforme apresentado na tabela do código genético do Capítulo 2.

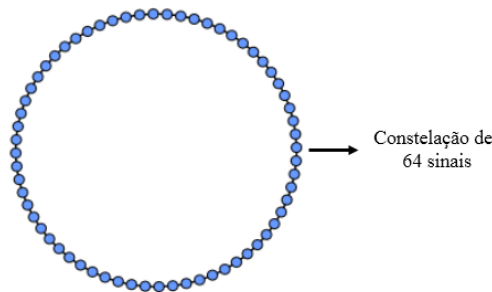


Figura 33 – Constelação de 64 sinais.

A associação algébrica aqui definida para fazer a relação do código genético e a constelação de sinais é através de extensões de Galois, conforme apresenta a Figura 34.

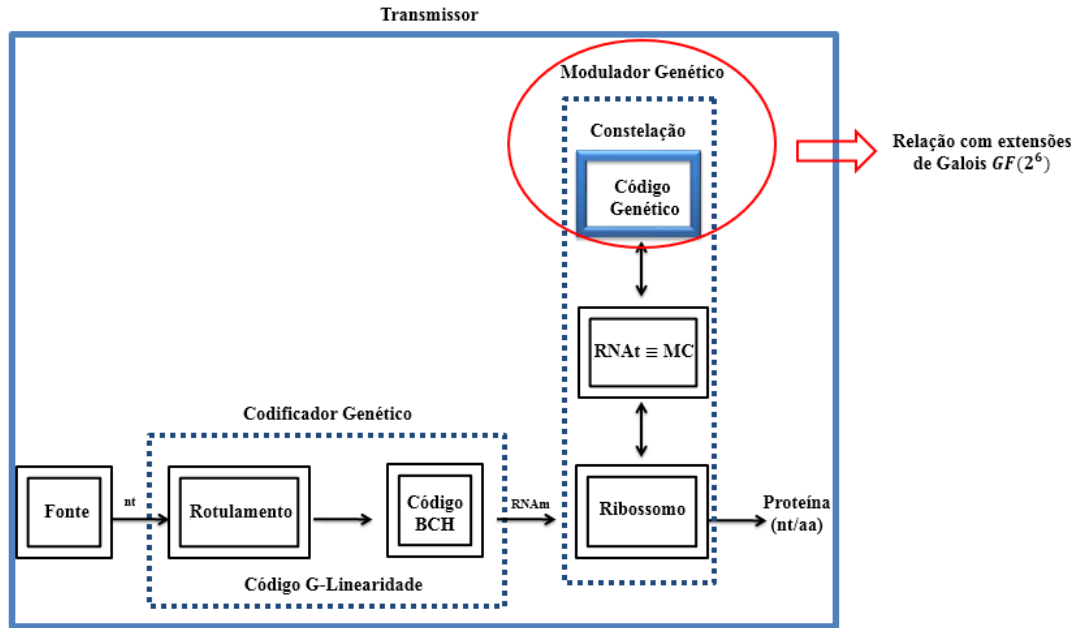


Figura 34 – Associação algébrica do modelo de comunicação genética.

Seja $m = 6$. O polinômio $p(X) = 1 + X + X^6$ é um polinômio primitivo sobre $GF(2)$.

Estabelecendo $p(\alpha) = 0$, temos que $1 + \alpha + \alpha^6 = 0$, ou seja, $\alpha^6 = 1 + \alpha$. Usando esta relação, podemos construir $GF(2^6)$ como se segue:

$$\alpha^7 = \alpha \cdot \alpha^6 = \alpha \cdot (1 + \alpha) = \alpha + \alpha^2$$

$$\alpha^8 = \alpha \cdot \alpha^7 = \alpha \cdot (\alpha + \alpha^2) = \alpha^2 + \alpha^3$$

$$\alpha^9 = \alpha \cdot \alpha^8 = \alpha \cdot (\alpha^2 + \alpha^3) = \alpha^3 + \alpha^4$$

$$\alpha^{10} = \alpha \cdot \alpha^9 = \alpha \cdot (\alpha^3 + \alpha^4) = \alpha^4 + \alpha^5$$

$$\alpha^{11} = \alpha \cdot \alpha^{10} = \alpha \cdot (\alpha^4 + \alpha^5) = \alpha^5 + 1 + \alpha = 1 + \alpha + \alpha^5$$

$$\alpha^{12} = \alpha \cdot \alpha^{11} = \alpha \cdot (1 + \alpha + \alpha^5) = \alpha + \alpha^2 + \alpha^6 = \alpha + \alpha^2 + 1 + \alpha = 1 + \alpha^2$$

$$\alpha^{13} = \alpha \cdot \alpha^{12} = \alpha \cdot (1 + \alpha^2) = \alpha + \alpha^3$$

$$\alpha^{14} = \alpha \cdot \alpha^{13} = \alpha \cdot (\alpha + \alpha^3) = \alpha^2 + \alpha^4$$

$$\alpha^{15} = \alpha \cdot \alpha^{14} = \alpha \cdot (\alpha^2 + \alpha^4) = \alpha^3 + \alpha^5$$

$$\alpha^{16} = \alpha \cdot \alpha^{15} = \alpha \cdot (\alpha^3 + \alpha^5) = \alpha^4 + \alpha^6 = \alpha^4 + 1 + \alpha = 1 + \alpha + \alpha^4$$

$$\alpha^{17} = \alpha \cdot \alpha^{16} = \alpha \cdot (1 + \alpha + \alpha^4) = \alpha + \alpha^2 + \alpha^5$$

$$\alpha^{18} = \alpha \cdot \alpha^{17} = \alpha \cdot (\alpha + \alpha^2 + \alpha^5) = \alpha^2 + \alpha^3 + \alpha^6 = \alpha^2 + \alpha^3 + 1 + \alpha = 1 + \alpha + \alpha^2 + \alpha^3$$

$$\alpha^{19} = \alpha \cdot \alpha^{18} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^3) = \alpha + \alpha^2 + \alpha^3 + \alpha^4$$

$$\alpha^{20} = \alpha \cdot \alpha^{19} = \alpha \cdot (\alpha + \alpha^2 + \alpha^3 + \alpha^4) = \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$$

$$\alpha^{21} = \alpha \cdot \alpha^{20} = \alpha \cdot (\alpha^2 + \alpha^3 + \alpha^4 + \alpha^5) = \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6 = \alpha^3 + \alpha^4 + \alpha^5 + 1 + \alpha = 1 + \alpha + \alpha^3 + \alpha^4 + \alpha^5$$

$$\begin{aligned}
 \alpha^{22} &= \alpha \cdot \alpha^{21} = \alpha \cdot (1 + \alpha + \alpha^3 + \alpha^4 + \alpha^5) = \alpha + \alpha^2 + \alpha^4 + \alpha^5 + \alpha^6 = \alpha + \alpha^2 + \alpha^4 + \alpha^5 + 1 + \alpha = \\
 &= 1 + \alpha^2 + \alpha^4 + \alpha^5 \\
 \alpha^{23} &= \alpha \cdot \alpha^{22} = \alpha \cdot (1 + \alpha^2 + \alpha^4 + \alpha^5) = \alpha + \alpha^3 + \alpha^5 + \alpha^6 = \alpha + \alpha^3 + \alpha^5 + 1 + \alpha = 1 + \alpha^3 + \alpha^5 \\
 \alpha^{24} &= \alpha \cdot \alpha^{23} = \alpha \cdot (1 + \alpha^3 + \alpha^5) = \alpha + \alpha^4 + \alpha^6 = \alpha + \alpha^4 + 1 + \alpha = 1 + \alpha^4 \\
 \alpha^{25} &= \alpha \cdot \alpha^{24} = \alpha \cdot (1 + \alpha^4) = \alpha + \alpha^5 \\
 \alpha^{26} &= \alpha \cdot \alpha^{25} = \alpha \cdot (\alpha + \alpha^5) = \alpha^2 + \alpha^6 = \alpha^2 + 1 + \alpha = 1 + \alpha + \alpha^2 \\
 \alpha^{27} &= \alpha \cdot \alpha^{26} = \alpha \cdot (1 + \alpha + \alpha^2) = \alpha + \alpha^2 + \alpha^3 \\
 \alpha^{28} &= \alpha \cdot \alpha^{27} = \alpha \cdot (\alpha + \alpha^2 + \alpha^3) = \alpha^2 + \alpha^3 + \alpha^4 \\
 \alpha^{29} &= \alpha \cdot \alpha^{28} = \alpha \cdot (\alpha^2 + \alpha^3 + \alpha^4) = \alpha^3 + \alpha^4 + \alpha^5 \\
 \alpha^{30} &= \alpha \cdot \alpha^{29} = \alpha \cdot (\alpha^3 + \alpha^4 + \alpha^5) = \alpha^4 + \alpha^5 + \alpha^6 = \alpha^4 + \alpha^5 + 1 + \alpha = 1 + \alpha + \alpha^4 + \alpha^5 \\
 \alpha^{31} &= \alpha \cdot \alpha^{30} = \alpha \cdot (1 + \alpha + \alpha^4 + \alpha^5) = \alpha + \alpha^2 + \alpha^5 + \alpha^6 = \alpha + \alpha^2 + \alpha^5 + 1 + \alpha = 1 + \alpha^2 + \alpha^5 \\
 \alpha^{32} &= \alpha \cdot \alpha^{31} = \alpha \cdot (1 + \alpha^2 + \alpha^5) = \alpha + \alpha^3 + \alpha^6 = \alpha + \alpha^3 + 1 + \alpha = 1 + \alpha^3 \\
 \alpha^{33} &= \alpha \cdot \alpha^{32} = \alpha \cdot (1 + \alpha^3) = \alpha + \alpha^4 \\
 \alpha^{34} &= \alpha \cdot \alpha^{33} = \alpha \cdot (\alpha + \alpha^4) = \alpha^2 + \alpha^5 \\
 \alpha^{35} &= \alpha \cdot \alpha^{34} = \alpha \cdot (\alpha^2 + \alpha^5) = \alpha^3 + \alpha^6 = \alpha^3 + 1 + \alpha = 1 + \alpha + \alpha^3 \\
 \alpha^{36} &= \alpha \cdot \alpha^{35} = \alpha \cdot (1 + \alpha + \alpha^3) = \alpha + \alpha^2 + \alpha^4 \\
 \alpha^{37} &= \alpha \cdot \alpha^{36} = \alpha \cdot (\alpha + \alpha^2 + \alpha^4) = \alpha^2 + \alpha^3 + \alpha^5 \\
 \alpha^{38} &= \alpha \cdot \alpha^{37} = \alpha \cdot (\alpha^2 + \alpha^3 + \alpha^5) = \alpha^3 + \alpha^4 + \alpha^6 = \alpha^3 + \alpha^4 + 1 + \alpha = 1 + \alpha + \alpha^3 + \alpha^4 \\
 \alpha^{39} &= \alpha \cdot \alpha^{38} = \alpha \cdot (1 + \alpha + \alpha^3 + \alpha^4) = \alpha + \alpha^2 + \alpha^4 + \alpha^5 \\
 \alpha^{40} &= \alpha \cdot \alpha^{39} = \alpha \cdot (\alpha + \alpha^2 + \alpha^4 + \alpha^5) = \alpha^2 + \alpha^3 + \alpha^5 + \alpha^6 = \alpha^2 + \alpha^3 + \alpha^5 + 1 + \alpha = \\
 &= 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^5 \\
 \alpha^{41} &= \alpha \cdot \alpha^{40} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^3 + \alpha^5) = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^6 = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + 1 + \alpha = \\
 &= 1 + \alpha^2 + \alpha^3 + \alpha^4 \\
 \alpha^{42} &= \alpha \cdot \alpha^{41} = \alpha \cdot (1 + \alpha^2 + \alpha^3 + \alpha^4) = \alpha + \alpha^3 + \alpha^4 + \alpha^5 \\
 \alpha^{43} &= \alpha \cdot \alpha^{42} = \alpha \cdot (\alpha + \alpha^3 + \alpha^4 + \alpha^5) = \alpha^2 + \alpha^4 + \alpha^5 + \alpha^6 = \alpha^2 + \alpha^4 + \alpha^5 + 1 + \alpha = \\
 &= 1 + \alpha + \alpha^2 + \alpha^4 + \alpha^5 \\
 \alpha^{44} &= \alpha \cdot \alpha^{43} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^4 + \alpha^5) = \alpha + \alpha^2 + \alpha^3 + \alpha^5 + \alpha^6 = \alpha + \alpha^2 + \alpha^3 + \alpha^5 + 1 + \alpha = \\
 &= 1 + \alpha^2 + \alpha^3 + \alpha^5 \\
 \alpha^{45} &= \alpha \cdot \alpha^{44} = \alpha \cdot (1 + \alpha^2 + \alpha^3 + \alpha^5) = \alpha + \alpha^3 + \alpha^4 + \alpha^6 = \alpha + \alpha^3 + \alpha^4 + 1 + \alpha = 1 + \alpha^3 + \alpha^4 \\
 \alpha^{46} &= \alpha \cdot \alpha^{45} = \alpha \cdot (1 + \alpha^3 + \alpha^4) = \alpha + \alpha^4 + \alpha^5 \\
 \alpha^{47} &= \alpha \cdot \alpha^{46} = \alpha \cdot (\alpha + \alpha^4 + \alpha^5) = \alpha^2 + \alpha^5 + \alpha^6 = \alpha^2 + \alpha^5 + 1 + \alpha = 1 + \alpha + \alpha^2 + \alpha^5 \\
 \alpha^{48} &= \alpha \cdot \alpha^{47} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^5) = \alpha + \alpha^2 + \alpha^3 + \alpha^6 = \alpha + \alpha^2 + \alpha^3 + 1 + \alpha = 1 + \alpha^2 + \alpha^3 \\
 \alpha^{49} &= \alpha \cdot \alpha^{48} = \alpha \cdot (1 + \alpha^2 + \alpha^3) = \alpha + \alpha^3 + \alpha^4 \\
 \alpha^{50} &= \alpha \cdot \alpha^{49} = \alpha \cdot (\alpha + \alpha^3 + \alpha^4) = \alpha^2 + \alpha^4 + \alpha^5 \\
 \alpha^{51} &= \alpha \cdot \alpha^{50} = \alpha \cdot (\alpha^2 + \alpha^4 + \alpha^5) = \alpha^3 + \alpha^5 + \alpha^6 = \alpha^3 + \alpha^5 + 1 + \alpha = 1 + \alpha + \alpha^3 + \alpha^5 \\
 \alpha^{52} &= \alpha \cdot \alpha^{51} = \alpha \cdot (1 + \alpha + \alpha^3 + \alpha^5) = \alpha + \alpha^2 + \alpha^4 + \alpha^6 = \alpha + \alpha^2 + \alpha^4 + 1 + \alpha = 1 + \alpha^2 + \alpha^4 \\
 \alpha^{53} &= \alpha \cdot \alpha^{52} = \alpha \cdot (1 + \alpha^2 + \alpha^4) = \alpha + \alpha^3 + \alpha^5 \\
 \alpha^{54} &= \alpha \cdot \alpha^{53} = \alpha \cdot (\alpha + \alpha^3 + \alpha^5) = \alpha^2 + \alpha^4 + \alpha^6 = \alpha^2 + \alpha^4 + 1 + \alpha = 1 + \alpha + \alpha^2 + \alpha^4 \\
 \alpha^{55} &= \alpha \cdot \alpha^{54} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^4) = \alpha + \alpha^2 + \alpha^3 + \alpha^5
 \end{aligned}$$

$$\begin{aligned}
 \alpha^{56} &= \alpha \cdot \alpha^{55} = \alpha \cdot (\alpha + \alpha^2 + \alpha^3 + \alpha^5) = \alpha^2 + \alpha^3 + \alpha^4 + \alpha^6 = \alpha^2 + \alpha^3 + \alpha^4 + 1 + \alpha = \\
 &= 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 \\
 \alpha^{57} &= \alpha \cdot \alpha^{56} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4) = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 \\
 \alpha^{58} &= \alpha \cdot \alpha^{57} = \alpha \cdot (\alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5) = \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6 = 1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 \\
 \alpha^{59} &= \alpha \cdot \alpha^{58} = \alpha \cdot (1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5) = \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6 = \\
 &= \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 + 1 + \alpha = 1 + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5 \\
 \alpha^{60} &= \alpha \cdot \alpha^{59} = \alpha \cdot (1 + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5) = \alpha + \alpha^3 + \alpha^4 + \alpha^5 + \alpha^6 = 1 + \alpha^3 + \alpha^4 + \alpha^5 \\
 \alpha^{61} &= \alpha \cdot \alpha^{60} = \alpha \cdot (1 + \alpha^3 + \alpha^4 + \alpha^5) = \alpha + \alpha^4 + \alpha^5 + \alpha^6 = \alpha + \alpha^4 + \alpha^5 + 1 + \alpha = 1 + \alpha^4 + \alpha^5 \\
 \alpha^{62} &= \alpha \cdot \alpha^{61} = \alpha \cdot (1 + \alpha^4 + \alpha^5) = \alpha + \alpha^5 + \alpha^6 = \alpha + \alpha^5 + 1 + \alpha = 1 + \alpha^5
 \end{aligned}$$

A Tabela 9 apresenta as representações por potência, polinomial e vetorial da extensão $GF(2^6)$.

Tabela 9 – $GF(2^6)$ gerado por $p(X) = 1 + X + X^6$.

Representação por Potência	Representação Polinomial	Representação vetorial
0	0	(000000)
1	1	(100000)
α	α	(010000)
α^2	α^2	(001000)
α^3	α^3	(000100)
α^4	α^4	(000010)
α^5	α^5	(000001)
α^6	$1 + \alpha$	(110000)
α^7	$\alpha + \alpha^2$	(011000)
α^8	$\alpha^2 + \alpha^3$	(001100)
α^9	$\alpha^3 + \alpha^4$	(000110)
α^{10}	$\alpha^4 + \alpha^5$	(000011)
α^{11}	$1 + \alpha + \alpha^5$	(110001)
α^{12}	$1 + \alpha^2$	(101000)
α^{13}	$\alpha + \alpha^3$	(010100)
α^{14}	$\alpha^2 + \alpha^4$	(001010)
α^{15}	$\alpha^3 + \alpha^5$	(000101)
α^{16}	$1 + \alpha + \alpha^4$	(110010)
α^{17}	$\alpha + \alpha^2 + \alpha^5$	(011001)
α^{18}	$1 + \alpha + \alpha^2 + \alpha^3$	(111100)
α^{19}	$\alpha + \alpha^2 + \alpha^3 + \alpha^4$	(011110)
α^{20}	$\alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(001111)
α^{21}	$1 + \alpha + \alpha^3 + \alpha^4 + \alpha^5$	(110111)
α^{22}	$1 + \alpha^2 + \alpha^4 + \alpha^5$	(101011)
α^{23}	$1 + \alpha^3 + \alpha^5$	(100101)
α^{24}	$1 + \alpha^4$	(100010)
α^{25}	$\alpha + \alpha^5$	(010001)

Representação por Potência	Representação Polinomial	Representação vetorial
α^{26}	$1 + \alpha + \alpha^2$	(111000)
α^{27}	$\alpha + \alpha^2 + \alpha^3$	(101100)
α^{28}	$\alpha^2 + \alpha^3 + \alpha^4$	(001110)
α^{29}	$\alpha^3 + \alpha^4 + \alpha^5$	(000111)
α^{30}	$1 + \alpha + \alpha^4 + \alpha^5$	(110011)
α^{31}	$1 + \alpha^2 + \alpha^5$	(101001)
α^{32}	$1 + \alpha^3$	(100100)
α^{33}	$\alpha + \alpha^4$	(010010)
α^{34}	$\alpha^2 + \alpha^5$	(001001)
α^{35}	$1 + \alpha + \alpha^3$	(110100)
α^{36}	$\alpha + \alpha^2 + \alpha^4$	(011010)
α^{37}	$\alpha^2 + \alpha^3 + \alpha^5$	(001101)
α^{38}	$1 + \alpha + \alpha^3 + \alpha^4$	(110110)
α^{39}	$\alpha + \alpha^2 + \alpha^4 + \alpha^5$	(011011)
α^{40}	$1 + \alpha + \alpha^2 + \alpha^3 + \alpha^5$	(111101)
α^{41}	$1 + \alpha^2 + \alpha^3 + \alpha^4$	(1011110)
α^{42}	$\alpha + \alpha^3 + \alpha^4 + \alpha^5$	(010111)
α^{43}	$1 + \alpha + \alpha^2 + \alpha^4 + \alpha^5$	(111011)
α^{44}	$1 + \alpha^2 + \alpha^3 + \alpha^5$	(101101)
α^{45}	$1 + \alpha^3 + \alpha^4$	(100110)
α^{46}	$\alpha + \alpha^4 + \alpha^5$	(010011)
α^{47}	$1 + \alpha + \alpha^2 + \alpha^5$	(111001)
α^{48}	$1 + \alpha^2 + \alpha^3$	(101100)
α^{49}	$\alpha + \alpha^3 + \alpha^4$	(010110)
α^{50}	$\alpha^2 + \alpha^4 + \alpha^5$	(001011)
α^{51}	$1 + \alpha + \alpha^3 + \alpha^5$	(110101)
α^{52}	$1 + \alpha^2 + \alpha^4$	(101010)
α^{53}	$\alpha + \alpha^3 + \alpha^5$	(010101)
α^{54}	$1 + \alpha + \alpha^2 + \alpha^4$	(111010)
α^{55}	$\alpha + \alpha^2 + \alpha^3 + \alpha^5$	(011101)
α^{56}	$1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4$	(111110)
α^{57}	$\alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(011111)
α^{58}	$1 + \alpha + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(111111)
α^{59}	$1 + \alpha^2 + \alpha^3 + \alpha^4 + \alpha^5$	(101111)
α^{60}	$1 + \alpha^3 + \alpha^4 + \alpha^5$	(100111)
α^{61}	$1 + \alpha^4 + \alpha^5$	(100011)
α^{62}	$1 + \alpha^5$	(100001)

Através da Tabela 9, as representações por potência, polinomial e vetorial da extensão de Galois $GF(2^6)$ gerado pelo polinômio $p(X) = 1 + X + X^6$, cada elemento pode ser associado a uma trinca de códons do código genético (apresentada no Capítulo 2) e a constelação de sinais.

Por meio dessa associação, percebe-se uma interessante conexão entre elementos bio-

lógicos (código genético), engenharia (constelação de sinais em um processo de transmissão de informação) e elementos de álgebra (extensões de Galois).

O detalhamento dessa representação para os rotulamentos A, B e C, presente no codificador genético, será apresentado em trabalhos futuros.

5 Conclusões

Tanto a teoria de comunicações quanto a genética preocupam-se com a transferência de informação e, cientistas tem estudado e explorado os conceitos dessas teorias a décadas, buscando representar matematicamente o que o mundo biológico realiza. No processo de transmitir uma informação desde um emissor até um receptor, podem ocorrer erros, os chamados ruídos, que estão alocados no canal, podendo comprometer a informação original. Biologicamente, muitos investigadores buscam fazer a associação do sistema de comunicação com a geração de proteínas e as ferramentas matemáticas apresentadas neste trabalho apresentam uma interessante conexão entre elementos biológicos, matemáticos e de engenharia, mostrando a interdisciplinaridade que este estudo visa.

Ao apresentarmos as semelhanças existentes entre a teoria de comunicações com o dogma da biologia molecular, percebe-se que os pesquisadores desenvolveram em seus modelos diversas informações sobre aspectos teóricos do sistema genético e, alguns buscaram suprir limitações não abrangidas em outros modelos, limitações estas que corroboram com os resultados apresentados nos trabalhos de Rocha e Palazzo Júnior (2010) e Faria e Palazzo Júnior (2011), cujo sistema de codificação biológico consiste de um codificador e de um modulador, construídos através de estruturas algébricas (anel e extensão de Galois), relacionando com o sistema de comunicação padrão.

Neste trabalho foram apresentadas interessantes relações e conexões, mostrando um elo entre sistema de comunicação, estruturas algébricas e elementos biológicos, buscando uma contribuição em uma área de pesquisa em franca expansão, fazendo com que a biologia se torne uma ciência fundamentada teoricamente em conceitos matemáticos e de engenharia.

Dessa forma, os objetivos propostos neste estudo foram alcançados e, a partir de estruturas algébricas, em particular, as extensões de Galois, estabelecemos as conexões da constelação de 64 sinais presente no modulador genético de um sistema de codificação genética semelhante a um sistema de comunicação. Sendo assim, este estudo interdisciplinar auxilia em melhoramentos genéticos, reduzindo tempo e custos laboratoriais e, consequentemente, auxilia na formação acadêmica de um matemático.

Através deste trabalho, foi submetido e aceito um artigo a ser publicado no Volume Temático da IV Semana da Matemática da UNIFAL-MG na Revista SIGMAE, também foi apresentado um resumo na forma de painel no XXXVII Congresso Nacional de Matemática Aplicada e Computacional (CNMAC) em Setembro 2017, na cidade de São José dos Campos-SP e, este trabalho também foi selecionado e apresentado em comunicação oral no Evento da IV Semana da Matemática (SEMAT) em Maio de 2017, na cidade de

Alfenas-MG, e está publicado nos Anais da IV SEMAT.

Este trabalho leva a várias questões a serem respondidas acerca das conexões apresentadas e possibilidades de aplicações em diversas áreas. A fim de dar prosseguimento a este estudo, apresentamos algumas sugestões, tais como: o detalhamento algébrico, através de extensões de Galois, para a representação dos rotulamentos A, B e C, presente no codificador genético da proposta de Faria e Palazzo Júnior; um estudo acerca do código BCH e sua importância na correção de erros; um estudo acerca do processo de decodificação em um sistema de comunicação. Enfim, é um estudo de extrema relevância para a continuação da pesquisa em trabalhos futuros.

Referências

ALBERTS, B. et al. **Biologia Molecular da Célula**. 5. ed. Porto Alegre: Artmed, 2010.

ALMEIDA, R. **Procarionte X Eucarionte**. 2012. Disponível em: <<http://cadernodebiologia.blogspot.com.br/2012/04/perguntas-e-respostas-procarionte-x.html>>. Acesso em: 30 mar. 2017.

BATTAIL, G. **An Outline of Informational Genetics**. [S.l.]: Morgan & Claypool Publishers, 2008.

DOMINGUES, H. H.; IEZZI, G. **Álgebra Moderna**. 4. ed. São Paulo: Atual, 2003.

FARIA, L. C. B.; PALAZZO JÚNIOR, R. **Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA**. 2011. 147 f. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2011.

GARCIA, A.; LEQUAIN, Y. **Elementos de Álgebra**. 5. ed. Rio de Janeiro: IMPA, 2010.

GATLIN, L. L. I. **Information Theory and the Living System**. New York: Columbia University Press, 1972.

LIN, S.; COSTELO JR, D. J. **Error Control Coding: Fundamentals and Applications**. New Jersey: Prentice Hall, Englewood Cliffs, 1983.

MARQUES, N. **Duplicação do DNA**. 2014. Disponível em: <<http://bionel.blogspot.com.br/p/duplicacao-do-dna.html>>. Acesso em: 30 mar. 2017.

MARZZOCO, A.; TORRES, B.B. **Bioquímica Básica**. 3. ed. [S.l.]: Guanabara-Koogan, 2013.

MATOSO, D. E. C.; QUEIROZ, C. R. O. Q. Modelos de Codificação Genômica e Genética gerados através de Códigos Corretores de Erros. **Sigmae**, Alfenas, v. 2, n. 2, p. 43-59, 2013.

MAY, E. E. Towards a Biological Coding Theory Discipline. **New Thesis**, v. 1, n. 1, p. 19-38, 2004.

MAY, et al. An error-correcting code framework for genetic sequence analysis. **Journal**

of **The Franklin Institute**, v. 341, n. 23, p. 89-109, 2004.

OLIVEIRA, A. J.; PALLAZO JÚNIOR, R. **Análise Algébrica dos Rotulamentos Associados ao Mapeamento do Código Genético**. 2012. 105 f. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2012.

PAMPHILE, J. A. (Org.); VICENTINI, V. E. P. (Org.). **Genética**. Maringá: EDUEM, 2011.

ROCHA, A. S. L.; PALAZZO JÚNIOR, R.; SILVA-FILHO, M.C. **Modelo de sistema de comunicações digital para o mecanismo de importação de proteínas mitocondriais através de códigos corretores de erros**. 2010. 155 f. Tese (Doutorado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2010.

ROMAN-ROLDAN, R.; BERNAOLA-GALVAN, P.; OLIVER, J. L. Application of information theory to DNA sequence analysis. **Pattern Recognition**, v. 29, n. 7, p. 1187-1194, 1996.

SHANNON, C. E. A mathematical theory of communication. **The Bell System Technical Journal**, v. 27, n. 3, p. 379-423, 1948.