

Julio Cesar Holanda de Albuquerque

Graduado em Matemática – IMMEC-UNICAMP

PROPOSTA DE CONSTELAÇÕES DE SINAIS PARA O CÓDIGO GENÉTICO

Dissertação de mestrado apresentada à Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Telecomunicações e Telemática. Aprovada pela banca examinadora no dia 25 de Novembro de 2008.

Banca Examinadora:

Prof. Dr. Reginaldo Palazzo Junior
FEEC/UNICAMP

Prof. Dr. Yuzo Iano
FEEC/UNICAMP

Prof. Dr. Carlos Eduardo Câmara
FCC/USF

Campinas
2008

COMISSÃO JULGADORA - TESE DE MESTRADO

Candidato: Julio Cesar Holanda de Albuquerque

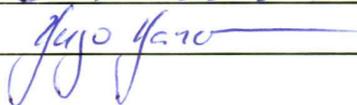
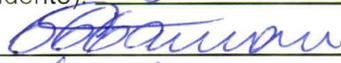
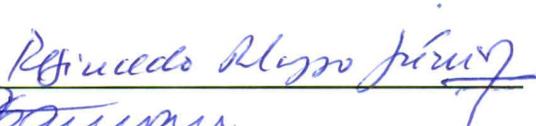
Data da Defesa: 25 de novembro de 2008

Título da Tese: "Proposta de Constelações de Sinais para o Código Genético"

Prof. Dr. Reginaldo Palazzo Júnior (Presidente):

Prof. Dr. Carlos Eduardo Câmara:

Prof. Dr. Yuzo Iano:



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

AL15p Albuquerque, Julio Cesar Holanda de
Proposta de constelações de sinais para o código
genético / Julio Cesar Holanda de Albuquerque.
--Campinas, SP: [s.n.], 2008.

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

1. Modulação digital. 2. Codigo genetico. 3. Grupos
de permutação. 4. Modelos matematicos. I. Júnior
Palazzo, Reginaldo. II. Universidade Estadual de
Campinas. Faculdade de Engenharia Elétrica e de
Computação. III. Título.

Título em Inglês: Proposal of signal constellations for the genetic code

Palavras-chave em Inglês: Signal Constellation, Genetic Code, Partition of the Genetic
Code, Permutation Groups, Mathematical models

Área de concentração: Telecomunicações e Telemática

Titulação: Mestre em Engenharia Elétrica

Banca examinadora: Reginaldo Palazzo Júnior, Carlos Eduardo Câmara, Yuzo Iano

Data da defesa: 25/11/2008

Programa de Pós Graduação: Engenharia Elétrica

DEDICADA AO MEU PAI; JOSÉ MIGUEL DE ALBUQUERQUE (IN MEMORIAM) E A TODA MINHA FAMÍLIA

Agradecimentos

Agradeço,

ao Prof. Dr. Reginaldo Palazzo Jr., por suas valiosas sugestões, comentários e pela sua paciência;

em especial ao amigo e grande irmão Giuliano (Machão!!!), por todas críticas e sugestões, uma pessoa notável e estimado por todos seus amigos e colegas;

aos colegas de trabalho do LTIA: Andréa, Clarice, Kátia, Luzinete e Wanessa (As famosas palazzetes)...só mulherada!!! Pela convivência descontraída e as trocas de experiências;

aos colegas do Departamento de Telemática: Renatinho, João, Alessandro e ao grande Valter (Vavá) pessoas que tenho grande admiração e ao Ricardo (Bom menino!) que sempre me ajudou nas dúvidas com o LaTeX e pelo seus ótimos comentários;

aos demais colegas do DT pela ótima convivência;

aos membros e ex-membros da M8-A: Uan's, Bixo, Ricardo, Sir John Erik e aos membros da P2: Cleo, Felipetes e Leandrão;

à FEEC/UNICAMP a ótima estrutura que oferece aos estudantes e pesquisadores;

à CAPES o portal de periódicos eletrônicos, que permite o acesso rápido e eficiente ao conhecimento científico;

à três grandes pensadores que influenciaram a minha maneira de ver o mundo: Karl Marx, Gauss e ultimamente a Darwin;

à duas pessoas que são exemplos de luta e determinação na minha vida: José Miguel de Albuquerque e Maria Judith Holanda (meus pais);

neste caso não mencionarei nomes de irmãos, de sobrinhos ou de outros familiares pois cometeria uma grande injustiça, sendo assim, agradeço a todos da minha família;

e finalmente, a todos aqueles que contribuíram pela minha formação intelectual e crítica.

“Vale a pena lembrar que, embora haja uma grande diferença entre nós no que diz respeito aos fragmentos que conhecemos, somos infinitamente iguais no universo da nossa ignorância.”

Karl Popper, in “As Origens do Conhecimento e da Ignorância”

Resumo

A proposta deste trabalho é apresentar uma abordagem aos processos genéticos e moleculares, utilizando a teoria de comunicações e codificação na modelagem do dogma central da biologia molecular. A partir desta modelagem associamos o código genético a um modulador de um sistema de comunicação. Mais especificamente, tal procedimento consiste em construir uma constelação de sinais a partir dos subgrupos de S_3 e S_4 baseado no código genético. Considerando este método algébrico de construção de sinais, propomos duas possíveis constelações de sinais para o código genético. A representação do código genético em constelações de sinais correlacionadas deu origem à idéia de “constelação de sinais concatenadas”, idéia inovadora na teoria de comunicação e codificação. As constelações de sinais concatenadas possui a propriedade de correção de erros, consistindo de novos conceitos úteis para utilização na teoria da comunicação e codificação. Por outro lado, estas representações do código genético não são únicas pois, até o presente momento, desconhecemos uma álgebra que descreva o código genético juntamente com as suas partições geradas pelos aminoácidos.

Palavras-chave: Aminoácidos. Proteína. DNA e RNA. Código genético. Partição do código genético. Dogma central da biologia molecular. Sistema de comunicação. Constelação de sinais. Grupo de permutações. Isomorfismo. Teorema de Cayley. Teorema de Burnside. Órbita de um conjunto. Representações de grupo. Classes conjugada.

Abstract

The purpose of this work is to present an approach to the genetic and molecular processes by use of the communication and coding theory in modelling the central dogma of the molecular biology. From this modelling we associate the genetic code to a modulator in the communication system. More specifically, such a procedure consists in the construction of a signal constellation by use of the S_3 and S_4 permutation subgroups based on the code genetic. By considering this algebraic method of signal design, we propose two possible signal constellations to the genetic code. The representation of the genetic code as correlated signal constellations provides the idea of “concatenated signal constellation”, an innovative idea in communication and coding theory. The concatenated signal constellations have the property of error-correction, a new concept being introduced. On the other hand, these representations of the genetic code are not unique for currently, we do not know an algebraic structure capable of describing the genetic code together with the partitioning generated by the amino acids.

Key-words: Amino acids. Protein. DNA and RNA. Genetic code. Partition of the genetic code. Central dogma of molecular biology. Communication system. Signal constellation. Permutation groups. Isomorphism. Cayley’s theorem. Burnside’s theorem. Orbit of a set. Group representations. Conjugacy class.

Lista de Figuras

2.1	As células procarióticas apresentam uma organização interna mais simples do que as células eucarióticas. (Lodish et al. 1998)	8
2.2	Estrutura geral dos aminoácidos e logo abaixo uma ligação peptídica. http://www.palaeos.com/Bacteria/Lists/Glossary/Gloss.html	10
2.3	Os diferentes tipos de estruturas que podem representar uma proteína. http://en.wikipedia.org/wiki/Protein_structure	11
2.4	Estrutura de um nucleotídeo. http://bioweb.wku.edu/pix/Pix.htm	14
2.5	Bases nitrogenadas do RNA.	14
2.6	Nucleotídeos do DNA.	15
2.7	Estrutura da dupla hélice do DNA e a complementação entre as bases. http://members.lycos.nl/TheDNApage/dnapix1b.html	16
2.8	Estrutura da molécula de DNA e RNA. Klug, William S.; Cummings, Michael R. <i>Concepts of Genetics</i> , Paramus, NJ, U.S.A. Prentice Hall PTR, 5th Edition, 1997.	17
2.9	Visão geral dos três processos básicos da genética molecular. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition.	20
4.1	Sistema de comunicação e codificação.	37
4.2	Codificador de k dígitos.	38
4.3	Constelações de sinais: (a) BPSK, (b) QPSK e (c) 8-PSK.	41
4.4	Dogma central da biologia.	43
4.5	Dogma central da biologia como um sistema de codificação.	45
4.6	Modulador genético.	48
5.1	Simetrias do triângulo equilátero.	51
5.2	Representação do grupo \mathbb{Z}_2 na reta real \mathbb{R} (bimodal).	52
5.3	Representação de sinais do grupo \mathbb{Z}_3 no \mathbb{R}^2	53

5.4	Representação de sinais do grupo \mathbb{Z}_3 no \mathbb{R}^3	54
5.5	Representação de sinais do grupo S_3 no \mathbb{R}^2	54
5.6	Representação de sinais do grupo S_3 no \mathbb{R}^3	55
5.7	Simetrias do cubo.	56
5.8	Representação de sinais de \mathbb{Z}_4 no \mathbb{R}^2	58
5.9	Representação de sinais do grupo $Z_2 \times Z_2$	59
5.10	Representação de sinais do grupo diedral D_4 no \mathbb{R}^2	63
5.11	Representação de sinais do grupo diedral D_4 no \mathbb{R}^3	63
5.12	Representação de sinais do grupo diedral D_4 no \mathbb{R}^4 (hipercubo).	64
5.13	Representação de sinais do grupo alternante A_4 no \mathbb{R}^2	66
5.14	Representação de sinais do grupo alternante A_4 no \mathbb{R}^3 (tetraedro truncado).	66
5.15	Representação do código genético em uma constelação de sinais. O ponto de tangência entre as subórbitas indica padrões em comum entre os códon.	67
5.16	Representação do código genético em duas constelações de sinais correlacionadas de acordo com os padrões em comum de cada códon.	69

Lista de Tabelas

1.1	21 partições: Código genético particionado de acordo com os 20 tipos de aminoácidos mais os códons de finalização.	4
2.1	Os 20 diferentes tipos de aminoácidos que compõem as proteínas.	9
2.2	Código Genético.	18

Sumário

1	Introdução	1
1.1	Histórico	1
1.1.1	Proposta do trabalho	2
1.1.2	Descrição do problema em aberto	3
1.1.3	Descrição do trabalho	3
2	Revisão de Conceitos de Biologia	5
2.1	A Célula	5
2.1.1	Átomos e moléculas	6
2.1.2	Célula: Unidade fundamental da vida	7
2.1.3	Tipos de célula e organelas celulares	8
2.2	Aminoácidos e Proteínas	9
2.2.1	Importância das proteínas	12
2.3	Ácidos Nucléicos	13
2.3.1	Estrutura da molécula de DNA	13
2.3.2	Estrutura da molécula de RNA	15
2.3.3	O código genético	17
2.4	Dogma Central da Biologia Molecular	19
2.5	Conclusões	21
3	Revisão de Conceitos Matemáticos	22
3.1	Espaços métricos	22
3.2	Grupos	23
3.3	Subgrupos, grupos cíclicos	23
3.4	Classes laterais e o teorema de Lagrange	24
3.5	Homomorfismo e isomorfismo de grupos	25

3.6	Grupos de permutações e o Teorema de Cayley	26
3.7	Ação de grupo	28
3.8	Representação de grupos	32
3.9	Classes conjugadas	33
4	Características e Propriedades dos Sistemas de Comunicação e Genético	35
4.1	Sistema de Comunicação	35
4.2	Exemplos de sinais	36
4.3	Teoria da codificação: uma introdução	36
4.3.1	Códigos corretores de erros	37
4.3.2	Codificador	38
4.3.3	Modulador	39
4.3.4	Canal de transmissão	39
4.3.5	Decodificador	40
4.3.6	Modulação e constelação de sinais	40
4.4	Comunicação biológica	42
4.5	Codificação e o Dogma Central da Genética	43
4.5.1	Codificador Genético	46
4.5.2	Canal Genético	46
4.5.3	Decodificador Genético	47
4.6	Código genético e modulação	48
5	Constelações de Sinais Associadas aos Grupos S_3 e S_4	50
5.1	Grupo de Permutação S_3	50
5.1.1	Subgrupos de S_3	51
5.2	Grupo de Permutação S_4	54
5.2.1	Subgrupos de S_4	56
5.3	Proposta de Constelação de Sinais para o Código Genético	66
6	Conclusões	70
6.1	Perspectivas	71
	Bibliografia	72

Introdução

1.1 Histórico

Em abril de 1953, os pesquisadores James Watson e Francis Crick, por meio de uma publicação na revista “Nature”, elucidaram a estrutura de uma molécula complexa, a molécula de DNA, comparando-a com uma escada de cordas. Essa descoberta foi fundamental para o desenvolvimento da biologia molecular nos anos seguintes. Em 1957, o próprio Francis Crick estabeleceu o **dogma central da biologia**: propôs que a informação gerada no DNA é transcrita para o RNA mensageiro (mRNA) e traduzida em proteína. Um pouco mais tarde, em 1968, Marshall W. Nirenberg, Robert W. Holley e Har Gobind Khorana, ganharam o prêmio nobel de medicina por trabalhar na interpretação do **código genético** e suas funções na síntese de proteína. Linguagem a partir da qual passamos a entender a relação entre os nucleotídeos e os aminoácidos e, portanto, conseguimos entender como uma proteína é produzida.

Sabemos que o código genético consiste de seqüências específicas de nucleotídeos, organizadas linearmente, seqüências que são produzidas a partir do DNA. Cada posição na molécula de DNA pode ser ocupada por um dos 4 nucleotídeos, enquanto que ao longo da molécula de uma proteína podemos ter qualquer um dos 20 aminoácidos ocupando uma posição na estrutura da proteína, variando conforme a funcionalidade da proteína.

Dessa forma, a palavra-código do código genético deverá ter comprimento 3, ou equivalentemente, as 64 possibilidades de combinações dos 4 nucleotídeos.

Desde 1966 o código genético ou código do DNA foi completamente decifrado. Obtendo, assim, resultados importantes como:

1. todos os aminoácidos nas proteínas são codificados por uma combinação de três bases, formando uma tripla chamada **códon**;
2. um aminoácido pode estar associado a mais de um códon (Veja Tabela 1.1) e, por isso, o código é dito **degenerado**;

3. o códon AUG é o único que além de codificar um aminoácido também funciona como um sinalizador de iniciação, informando o início da seqüência formadora da proteína;
4. os códons UAA, UAG, e UGA funcionam como sinalizadores de terminação informando o término da seqüência formadora da proteína, ao invés, de codificar aminoácidos como os códons restantes.

Não se sabe ainda, porque no código genético, existem aminoácidos especificados por seis códons enquanto outros são especificados por dois e somente dois aminoácidos são especificados cada um por um códon. A pergunta que se faz é: há alguma estrutura matemática associada a esta configuração ? ou ainda, será que é possível modelar o código genético ?

A motivação para esta proposta de mestrado surgiu destas perguntas. Sabemos que o código genético é a linguagem que nos permite entender como é realizada a síntese de proteína, ou seja, a informação recebida. Sob este enfoque podemos interpretar o dogma central da biologia como um sistema de comunicação, isto é,

**Fonte de informação (DNA) → Seqüência na saída da fonte (mRNA) →
→ Codificador (ribossomo) → Informação recebida (proteína)**

Em um primeiro momento direcionamos nosso estudo para a linguagem que traduz a informação em uma seqüência de aminoácidos, especificamente para os 64 códons do código genético, peça fundamental para a codificação da informação.

A proposta em consideração é associar o código genético a uma constelação de sinais e para isto ser possível é preciso encontrar no grupo de permutações S_{64} um subgrupo de ordem 64 e a partir deste subgrupo caracterizar a constelação de sinais desejada. Obviamente, não é uma tarefa fácil pois a busca de subgrupos com ordem 64 em S_{64} é um problema combinatorial complexo. Todavia, o teorema de Cayley garante que para um dado valor n todos os grupos de ordem n estão contidos em S_n . Dessa forma, procuramos associar a cada um dos aminoácidos uma constelação de sinais cuja a cardinalidade corresponde à quantidade de códons associados. Até o momento, não temos conhecimento de trabalhos anteriores enfocando esta temática possibilitando, desta maneira, uma nova abordagem para o código genético e desenvolvendo um ferramental matemático para uma futura aplicação na biologia molecular.

1.1.1 Proposta do trabalho

No Capítulo 5 geramos todos os subgrupos de S_3 e S_4 e suas respectivas representações de grupo e a partir disto construímos constelações de sinais associada a cada subgrupo. Importante resultado deste trabalho pois permite abordar e propor um modelo de constelação de sinais para o código genético.

1.1.2 Descrição do problema em aberto

Representaremos o código genético por $X_{64} = \{x_1, \dots, x_{64}\}$, onde cada x_i representa um códon. Temos 61 códons que representam 20 aminoácidos e três códons de terminação (UAA, UAG e UGA). Uma questão curiosa é descobrir se existe alguma estrutura matemática inerente ao conjunto X_{64} com o objetivo de construir uma constelação de sinais associada a X_{64} . Como não sabemos se tal estrutura existe, uma maneira de verificar isso é gerar subgrupos de ordem 64 em S_{64} , e a partir disto, construir constelação de sinais associada a cada subgrupo. Entretanto devido à complexidade algorítmica envolvida, mostramos o procedimento de construção de sinais a partir dos grupos S_3 e S_4 .

Outra questão interessante é a seguinte: dado um conjunto finito X e uma partição de X em k subconjuntos, seria possível encontrar um grupo G e uma ação de G em X , a saber, $\tau : G \times X \rightarrow X$, de tal forma que as classes da relação de equivalência originadas pela ação τ sejam exatamente os k subconjuntos da partição de X ? É importante salientar que este problema pode ser descrito da seguinte maneira: dado um conjunto X e uma ação de grupo de G em X , encontrar o número de partições geradas pela ação τ , (Veja Teorema 7). Acreditamos que o raciocínio descrito anteriormente possa ser generalizado para a resolução da conjectura proposta a saber: Sabemos que o código genético é particionado em 20 tipos de aminoácidos para 61 códons, além dos códons de sinalização. Então, deve-se encontrar um grupo G e uma ação τ de G em X_{64} de tal forma que as classes de equivalência geradas pela ação de τ sejam exatamente iguais as 20 partições (subconjuntos) do código genético, determinadas pelos 20 tipos de aminoácidos, desconsiderando-se os códons de terminação. Na Tabela 1.1, representamos o código genético particionado em 21 partições, onde cada partição é representada por uma cor e cada cor está associada a um dos 20 tipos de aminoácidos. Em particular, a cor vermelha está associada aos códons que determinam o término (finalização) da seqüência formadora da proteína.

1.1.3 Descrição do trabalho

Este trabalho encontra-se dividido em seis capítulos. Os próximos cinco capítulos estão organizados da seguinte maneira:

Capítulo 2: Apresentamos, uma breve revisão dos conceitos de biologia molecular e genética enfatizando o mecanismo de síntese de proteína e o código genético;

Capítulo 3: Discutimos alguns pontos importantes sobre teoria de grupos, particularmente, os grupos de permutações e representação de grupos, parte relevante do desenvolvimento do nosso trabalho. Voltando a nossa atenção para os conceitos de isomorfismos, classes laterais e o teorema de Cayley e Burnside;

Tabela 1.1: 21 partições: Código genético particionado de acordo com os 20 tipos de aminoácidos mais os códons de finalização.

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

Capítulo 4: Fazemos uma breve introdução de um sistema de comunicação e como a teoria de códigos corretores de erros é incorporada ao sistema de comunicação. Descrevemos o dogma central da biologia e associamo-o a um sistema de comunicações. Interpretamos o código genético como sendo o bloco modulador em um sistema de comunicação digital.

Capítulo 5: Neste capítulo classificamos os subgrupos do grupo de permutação S_3 e S_4 e a representação de sinais associada a cada subgrupo. Este procedimento consiste em associar o código genético ao modulador em um sistema de comunicação, mais especificamente, consiste em construir duas constelações de sinais baseadas no código genético;

Capítulo 6: Finalmente, apresentamos as conclusões e propostas para trabalhos futuros.

Revisão de Conceitos de Biologia

O objetivo deste capítulo é apresentar, de maneira sucinta, alguns conceitos biológicos e bioquímicos úteis para o entendimento do mecanismo biológico de síntese protéica e do código genético. Utilizamos neste capítulo as referências (Lodish, Berk, Baltimore, Zipurski & Parnell 1998, Griffiths, Wessler, Lewontin, Gelbart, Suzuki & Miler 2006, Farah 2007, Lewin 2000, Alberts, Raff, Roberts & Walter 2004).

O capítulo está organizado como segue:

- Na Seção 2.1, estabelecemos os conceitos fundamentais necessários ao entendimento deste trabalho. São estes: definições, tipos, funções e compartimentos que compõem as células, bem como suas respectivas funções. Ainda, apresentamos a diferença entre organismos unicelulares e multicelulares. Na Subseção 2.1.1, descrevemos os conceitos de átomos e moléculas. Na Subseção 2.1.2, definimos o conceito de célula, unidade fundamental da vida. Na Subseção 2.1.3, apresentamos os tipos de células e organelas celulares;
- A Seção 2.2 consiste do estudo sobre aminoácidos e proteínas. Mais detalhadamente, na Subseção 2.2.1 descrevemos a importância das proteínas e suas funções;
- Na Seção 2.3, desenvolvemos o estudo sobre ácidos nucleicos (RNA e DNA). Nas Subseções 2.3.1 e 2.3.2, expomos as principais características e diferenças entre as moléculas de DNA e RNA. Na Subseção 2.3.3, descrevemos o código genético;
- Na Seção 2.4, tratamos do dogma central da biologia molecular ressaltando o mecanismo de síntese protéica. Finalmente, na Seção 2.5, relatamos as considerações finais.

2.1 A Célula

Nesta seção, definimos o conceito de célula e organismos unicelulares e multicelulares. Ainda, analisamos os diferentes tipos de células e os elementos que constituem as células, ou seja, as

organelas celulares.

2.1.1 Átomos e moléculas

Tudo o que se encontra na natureza é formado por pequenas partículas denominadas átomos. Como um exemplo, citamos a água, representada por H_2O , ou seja, cada molécula de água é composta por dois átomos do elemento químico hidrogênio ligados a um átomo do elemento químico oxigênio.

Um **elemento químico** é uma substância pura que não pode ser quebrada por processos químicos em partículas mais simples. A tabela periódica contém mais de 100 elementos químicos conhecidos, dentre os quais, aqueles que são encontrados em maior abundância na composição dos seres vivos: hidrogênio, oxigênio, nitrogênio e o carbono.

O **átomo** é a menor unidade de um elemento químico que conserva as suas propriedades físico-químicas. Os átomos ligam-se para formar os compostos químicos. As **moléculas** são as unidades dos compostos químicos. Um exemplo de molécula é a molécula de água, H_2O , que representa a quantidade mínima de átomos de oxigênio e hidrogênio de tal forma que, quando ligados (ligação química), ainda preservam as características do composto químico água. Outros compostos químicos são formados por muitos átomos. Assim, por exemplo, a molécula do composto químico glicose é formada por 6 átomos de carbono, 12 átomos de hidrogênio e 6 átomos de oxigênio.

Uma observação importante é a forma da ligação entre os átomos de uma molécula, ou melhor, como eles estão rearranjados entre si, pois, de acordo com esta conformação espacial dos átomos temos uma representação da estrutura da molécula.

Conforme o tamanho das moléculas em uma célula podemos dividi-las em dois tipos:

1. **Micromoléculas:** são moléculas muito pequenas, formadas por uma pequena seqüência linear de átomos, como é o caso da água, dos sais, açúcares, aminoácidos e ácidos gordurosos. As micromoléculas atuam basicamente na produção e armazenamento de energia na célula.
2. **Macromoléculas:** são moléculas enormes, formadas por uma seqüência linear de milhões de átomos, ou mesmo centenas de milhões de átomos, como é o caso das proteínas e dos ácidos nucléicos, encontrados nos organismos vivos. Desempenham diversas funções nas células e, no caso das proteínas, funcionam como um “sinalizador” ou “catalisador”, protegendo a célula contra corpos estranhos ou acelerando alguma reação química. Além disso, faz parte da estrutura dos organismos vivos. No caso dos ácidos nucléicos as macromoléculas são responsáveis por transmitir e armazenar a informação genética contida

no núcleo. As macromoléculas, tais como as proteínas, são formadas por muitas moléculas menores organizadas linearmente, denominadas **monômeros**, e que, no caso das proteínas, correspondem aos aminoácidos.

2.1.2 Célula: Unidade fundamental da vida

Todo ciclo evolutivo da vida é composto das seguintes fases: nascimento, desenvolvimento, reprodução, envelhecimento e morte. Durante este ciclo recebemos constantemente estímulos, executamos funções, processamos informações e damos origem a novos seres. Semelhantemente, cada célula que forma nosso corpo deve passar por todo o processo do ciclo da vida. A vida é definida por essas capacidades. Os seres humanos e outros organismos multicelulares são compostos por bilhões ou trilhões de células organizadas em estruturas complexas. Porém, existem outros organismos consistindo de uma única célula, denominados organismos **unicelulares**. Exemplos de organismos unicelulares são as bactérias e algumas algas. Mesmo esses organismos mais simples, compostos por uma única célula, estão sujeitos a todas as propriedades da vida, indicando, portanto, que a célula é a unidade fundamental da vida.

A **célula** é definida como sendo a menor parte de um organismo vivo capaz de desenvolver, de forma autônoma, as funções básicas que caracterizam a vida: reprodução e crescimento.

Nos organismos mais complexos, denominados **multicelulares**, o número de células varia em função da complexidade apresentada. No caso humano, por exemplo, estima-se que um indivíduo adulto seja composto por 10 trilhões de células.

As células que participam da formação de um organismo não são todas iguais. Elas podem assumir diferentes formas tornando-se especializadas no desempenho de determinadas funções. Por exemplo, as **hemácias**, que são glóbulos vermelhos do sangue, têm como principal tarefa o transporte de oxigênio no organismo, diferentemente das células hepáticas, ou **hepatócitos**, que participam da metabolização de algumas substâncias, como, por exemplo, o álcool etílico e, além disso, participam da produção da bile. Outras células como, por exemplo, as células nervosas, têm como função conduzir os estímulos nervosos e sensoriais. Estas modificações que ocorrem nas células, a fim de torna-las mais eficientes no desempenho de uma determinada função, acabam por alterar sua forma e conteúdo, resultando na **diferenciação celular**. Essa divisão de trabalho que ocorre nos organismos mais complexos, representa uma grande economia de energia, aumentando muito a eficiência do organismo para realizar suas funções biológicas. Conseqüentemente, as células que se diferenciam para realizar a mesma função formam um **tecido**; os tecidos se organizam em órgãos para formar o indivíduo. Toda essa estrutura deve funcionar de forma extremamente precisa e harmônica, caso contrário, as conseqüências para o indivíduo como um todo podem ser graves e, algumas vezes, tão drásticas que o tornam

incompatível com a vida.

2.1.3 Tipos de célula e organelas celulares

O universo biológico é constituído por dois tipos de células: **procarióticas** e **eucarióticas**. As células procarióticas consistem de um único compartimento fechado, que é delimitado por uma membrana plasmática; essas células não possuem um núcleo definido e apresentam uma organização interna relativamente simples (Veja Figura 2.1 (a)). As células eucarióticas, diferentemente das células procarióticas, contêm um **núcleo** bem definido, delimitado pela membrana nuclear, e o **citoplasma**, região localizada entre a membrana plasmática e o núcleo (Veja Figura 2.1 (b)). No citoplasma há uma grande quantidade de membranas internas que delimitam outros compartimentos altamente organizados, referidos como **organelas celulares**. O meio viscoso que preenche o citoplasma juntamente com as organelas celulares é denominado **citossol**.

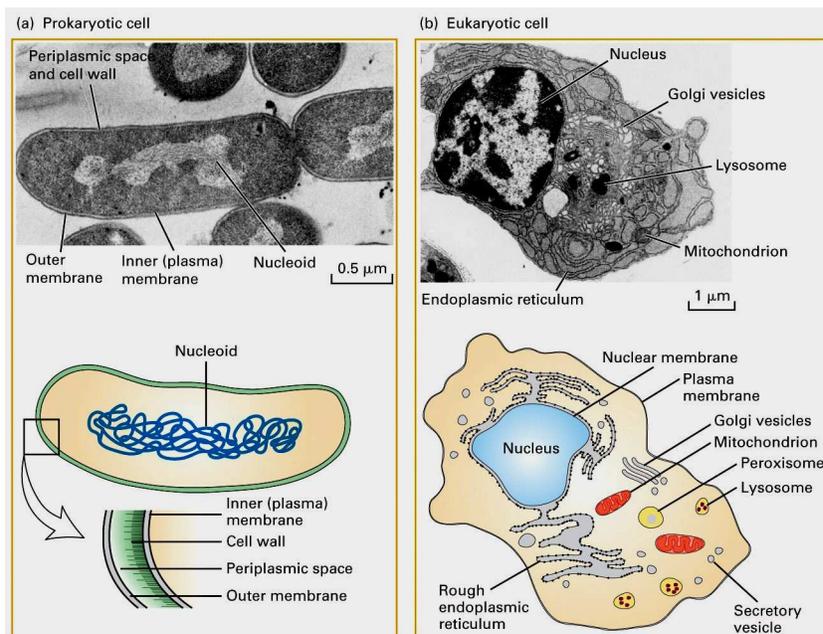


Figura 2.1: As células procarióticas apresentam uma organização interna mais simples do que as células eucarióticas. (Lodish et al. 1998)

Cada uma destas organelas tem uma função bem definida no trabalho geral que a célula deve desenvolver, algumas delas aparecem representadas na Figura 2.1 (b). A célula, como qualquer fábrica, depende de energia para funcionar. A energia é geralmente obtida através da digestão dos alimentos. Como exemplos de organelas citoplasmáticas podemos citar as **mitocôndrias**, responsáveis pela produção de energia, os **lisossomos**, onde os alimentos são digeridos, o **complexo de Golgi**, que concentra e secreta as proteínas para o exterior da célula e os **ribossomos**, organelas nas quais as proteínas são fabricadas e que, geralmente, aparecem

ligadas a um sistema de membranas do citoplasma, o **retículo endoplasmático**. Este processo de fabricação de proteínas é denominado **síntese protéica**. A principal diferença entre as células procarióticas e eucarióticas é a ausência da membrana nuclear nas células procarióticas. Outra diferença importante está na maneira em que a própria molécula de DNA está organizada. Por exemplo, o DNA dos eucariotos está altamente associado com proteínas específicas para formar nucleoproteínas que são reorganizadas para formar cromossomos, enquanto o DNA dos procariotos está livre de tais proteínas estruturais.

As bactérias, vírus, e algumas algas são exemplos de organismos procariotos; as plantas e animais são exemplos de organismos eucariotos.

2.2 Aminoácidos e Proteínas

Os **aminoácidos** (aa), pequenas moléculas que compõem as proteínas, podem ser de 20 tipos diferentes. Cada um destes aminoácidos tem um nome específico, onde tais nomes costumam ser abreviados tanto por três letras quanto por uma única letra do alfabeto escolhida por convenção (Tabela 2.2).

Tabela 2.1: Os 20 diferentes tipos de aminoácidos que compõem as proteínas.

Descrição dos Aminoácidos		
Aminoácidos	Nomes abreviados	Abreviação simplificada
Glicina	Gly	G
Alanina	Ala	A
Valina	Val	V
Leucina	Leu	L
Isoleucina	Ile	I
Serina	Ser	S
Cisteína	Cys	C
Metionina	Met	M
Tirosina	Tyr	Y
Fenilalanina	Phe	F
Tripofano	Trp	W
Histidina	His	H
Arginina	Arg	R
Lisina	Lys	K
Asparto	Asp	D
Glutamato	Glu	E
Asparagina	Asn	N
Glutamina	Gln	Q
Prolina	Pro	P
Treonina	Thr	T

Todos os 20 aminoácidos apresentam a mesma estrutura básica: um átomo de carbono central aparece simultaneamente ligado a um grupo amina, a um grupo carboxila e a um átomo de hidrogênio. O que faz a diferença entre um aminoácido e outro é o quarto grupo lateral ligado ao carbono. O grupo amina tem carga positiva, enquanto o grupo carboxila tem carga negativa. Portanto, a carga final do aminoácido depende da carga presente no grupo lateral (Veja Figura 2.2).

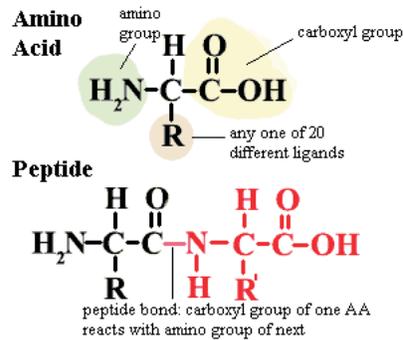


Figura 2.2: Estrutura geral dos aminoácidos e logo abaixo uma ligação peptídica. <http://www.palaeos.com/Bacteria/Lists/Glossary/Gloss.html>

A síntese de uma proteína tem início quando o grupo carboxila de um aminoácido reage com o grupo amina de outro, em uma reação química chamada **ligação peptídica**, na qual é eliminada uma molécula de água (Veja Figura 2.2). O resultado desta reação é um **dipeptídeo**. Observamos que um dipeptídeo ainda possui em uma das pontas um grupo amina e, na outra, um grupo carboxila. O grupo carboxila livre pode reagir com um grupo amina de outro aminoácido para formar um **tripeptídeo**. Esse processo pode se repetir indefinidamente, ou seja, não importa quantos aminoácidos já foram ligados, sempre poderá ser adicionado mais um grupo carboxila livre.

Os aminoácidos ligam-se uns aos outros para formar longas **seqüências polipeptídicas** lineares que contêm uma variedade de centenas, ou mesmo milhões, daqueles 20 tipos diferentes de aminoácidos. Uma proteína típica contém entre 50 e 2000 aminoácidos e, eventualmente, pode ser formada por duas ou mais cadeias polipeptídicas sintetizadas separadamente, que se juntam posteriormente para formar a proteína final.

Outro fato interessante é sobre o rearranjo espacial da seqüência de amino-ácidos. A primeira estrutura, que é a mais simples de todas, é denominada **estrutura primária** e é definida logo após a ligação dos aminoácidos em uma seqüência linear. Após isso, esta longa seqüência polipeptídica, composta de aminoácidos, dobra-se em porções localizadas sobre si mesma, promovendo diversos rearranjos espaciais na seqüência de aminoácidos. Esta nova conformação é denominada **estrutura secundária**. Analogamente, esta seqüência de aminoácidos dobra-se

em porções específicas sobre si mesma, definindo uma nova conformação denominada **estrutura terciária**. Se uma proteína em particular é formada por um complexo de mais de uma seqüência polipeptídica, a estrutura completa é denominada **estrutura quartenária** (Veja Figura 2.3).

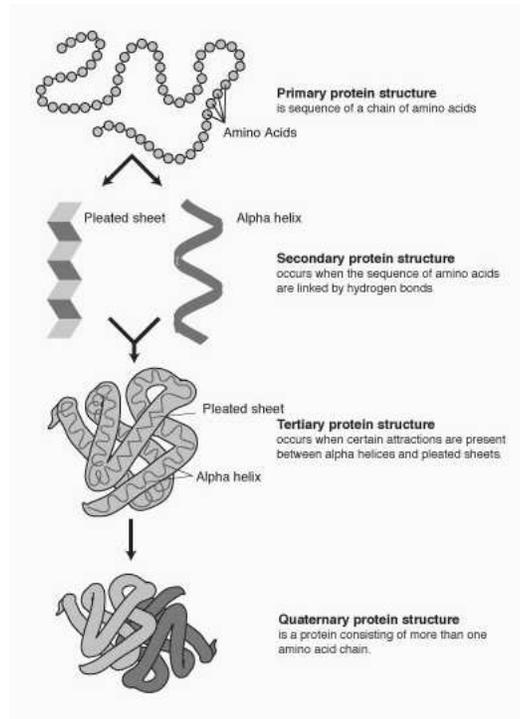


Figura 2.3: Os diferentes tipos de estruturas que podem representar uma proteína. http://en.wikipedia.org/wiki/Protein_structure

Desse modo é possível que um aminoácido em uma determinada posição, pertencente a uma determinada estrutura protéica, fique próximo, por exemplo, de aminoácidos em outras posições, dando uma nova configuração à seqüência de aminoácidos. O modo como uma seqüência polipeptídica será dobrada depende, aparentemente, somente da seqüência de aminoácidos que a compõe. Em outras palavras, a interação entre diferentes grupos laterais dos aminoácidos que formam a cadeia polipeptídica faz com que uma dada seqüência de aminoácidos assumam naturalmente uma forma tridimensional específica.

De um modo geral, a conformação espacial da seqüência de aminoácidos determina as propriedades físico-químicas e a funcionalidade das proteínas.

Percebemos, então, que a **proteína** é formada por um conjunto de aminoácidos seqüencialmente ordenados (rearranjados espacialmente) sendo responsável em realizar a maior parte das tarefas dentro e fora das células. Além disso, incluem enzimas, anticorpos, hormônios, moléculas sinalizadoras, receptoras e estruturais, executando diversas tarefas de acordo com sua forma tridimensional.

Portanto, entre as moléculas conhecidas, as proteínas são estruturalmente as mais complexas e as mais sofisticadas funcionalmente.

2.2.1 Importância das proteínas

Este grupo de macromoléculas constituem a maior parte da matéria seca de uma célula. Quando olhamos uma célula ao microscópio ou analisamos sua atividade elétrica ou bioquímica, estamos, na verdade, observando proteínas. Por outro lado, as proteínas não são meramente os blocos que constituem as células; elas também executam praticamente todas as funções celulares. São as enzimas, por exemplo, que constituem as intrincadas superfícies moleculares numa célula, que promovem suas múltiplas reações químicas. Os anticorpos atuam em nosso organismo sinalizando, ou protegendo o corpo de alguma infecção. Proteínas imensas na membrana plasmática formam canais e bombas que controlam a passagem de pequenas moléculas para dentro e para fora das células. Algumas proteínas carregam mensagens de uma célula para outra (transporte), enquanto outras agem como integradoras de sinais, direcionando o conjunto de sinais da membrana plasmática para o núcleo de células individuais. Outras funcionam ainda como máquinas protéicas com partes móveis: a **topoisomerase** pode desamarrear moléculas de DNA que apresentam nós. Outro grupo de proteínas especializadas agem como, toxinas, movimentação, hormônios, moléculas anticoagulantes, fibras elásticas, etc. Apesar, de toda esta infinidade de funções desempenhadas pelas proteínas podemos dividi-las em dois grandes grupos:

1. **Proteínas estruturais:** Da mesma forma que o tijolo e o ferro compõe a estrutura de uma casa, este tipo de proteína participa da estrutura do organismo. Citamos como exemplos: fibras musculares, fibras do colágeno, cabelo, unha, etc.
2. **Proteínas funcionais:** Atuam no organismo protegendo-o de algum corpo estranho ou transportando moléculas para dentro das células ou regulando e sinalizando em alguns processos bioquímicos, ou promovendo e acelerando uma determinada reação bioquímica, receptora de estímulos internos e externos, etc. Citamos como exemplos: anticorpos, hemoglobina, repressor de lactose, insulina, rodopsina (proteína receptora de luz), etc.

Verificamos, portanto, a multiplicidade de funções biológicas desempenhadas pela proteína e a sua grande importância para entendermos os mecanismos bioquímicos que são realizados nas células. Sendo de principal relevância compreendermos a função realizada pelas proteínas nas células.

2.3 Ácidos Nucléicos

Como vimos, na Subseção 2.1.3, os **ácidos nucleicos** são macromoléculas que contêm toda a informação que determina a seqüência de aminoácidos e, conseqüentemente, a estrutura e a função de todas as proteínas de uma célula. A principal diferença entre os ácidos nucleicos e as proteínas é que as proteínas desempenham um trabalho dentro das células enquanto os ácidos nucleicos são responsáveis por transmitir a informação genética contida no núcleo. As proteínas são como os operários da fábrica, cada uma realizando uma função diferente de acordo com a sua forma tridimensional. Por outro lado, os ácidos nucleicos são o sistema de computadores dessa fábrica, que guardam em seus programas e arquivos inúmeras informações sendo, portanto, capazes de gerenciar o trabalho geral das células, mas eles próprios não realizam nenhuma tarefa.

Os ácidos nucleicos contêm as informações para a produção das proteínas no local e momento adequados. A informação referente a como, quando e onde deve ser produzido cada tipo de proteína está contida no material genético que se localiza dentro do núcleo. Há dois tipos de ácidos nucleicos, quimicamente semelhantes, abreviados por **DNA** e **RNA**, que controlam de forma integrada a síntese de proteínas. Cada proteína que será sintetizada na célula tem sua informação, ou seja, a seqüência de seus aminoácidos arquivada num computador central, o DNA. Por outro lado, sabemos que o DNA se localiza na administração da fábrica, isto é, o núcleo, e a fabricação das proteínas acontece nos ribossomos, organelas do citoplasma. É de se perguntar como a informação responsável em fabricar proteína sai do núcleo e chega ao citoplasma? A Natureza foi responsável em solucionar esta questão. Esta informação especificada para fabricar uma determinada proteína funciona como um arquivo, selecionado no DNA, que é copiado e armazenado em um “disquete” e depois transportado até o citoplasma. Este disquete com as informações específicas é denominado RNA.

Deve ficar claro que toda informação arquivada no DNA está guardada na estrutura de sua molécula.

2.3.1 Estrutura da molécula de DNA

O DNA é um polímero, ou seja, é formado por uma molécula muito longa composta por um grande número de subunidades pequenas que se repetem, os monômeros. Neste caso cada monômero é denominado **nucleotídeo** (Veja Figura 2.4). Cada nucleotídeo é composto por três partes: um **grupo fosfato**, um **açúcar** e uma **base nitrogenada**.

O nome difícil de se pronunciar do DNA, **ácido desoxirribonucleico**, é dado pelo açúcar presente em sua molécula, a **desoxirribose**, formada por um anel de átomos de carbono e oxigênio. Ligado ao átomo de carbono de número 5 do açúcar aparece o grupo fosfato carregado

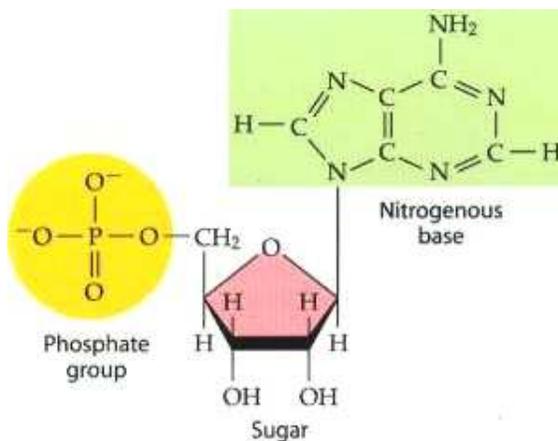


Figura 2.4: Estrutura de um nucleotídeo. <http://bioweb.wku.edu/pix/Pix.htm>

de carga negativa. Isto torna o nucleotídeo, bem como toda a molécula de DNA, negativos.

As bases nitrogenadas são assim chamadas, pois são formadas de anéis de carbono e nitrogênio. Existem quatro tipos de bases no DNA (Veja Figura 2.5): duas delas maiores, **adenina** e **guanina**, são chamadas de **purinas** e são constituídas por um anel duplo de carbono e nitrogênio. As outras menores, **citossina** e **timina**, são compostas por um anel simples e são chamadas de **pirimidinas**. Uma vez que o nome de cada base começa com uma letra diferente, elas podem ser referidas, abreviadamente, por **C**, **T**, **A** e **G**. As bases nitrogenadas são de extrema importância pois são elas que carregam a informação para a formação das proteínas, pois de acordo com a formação das seqüências em que aparecem as letras A, C, G, e T no DNA obtemos um tipo de proteína diferente.

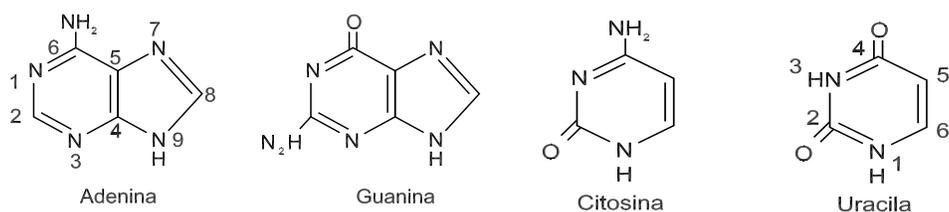


Figura 2.5: Bases nitrogenadas do RNA.

Açúcar mais fosfato são componentes invariáveis nos nucleotídeos e apresentam uma função unicamente estrutural na molécula de DNA. A parte que guarda a informação importante sobre os programas para o processamento das proteínas são as bases nitrogenadas. De acordo com a base presente no DNA, existem quatro tipos de nucleotídeos possíveis e que estão ilustrados na Figura 2.6.

A molécula de DNA é constituída de duas seqüências (fitas) de nucleotídeos que se enrolam

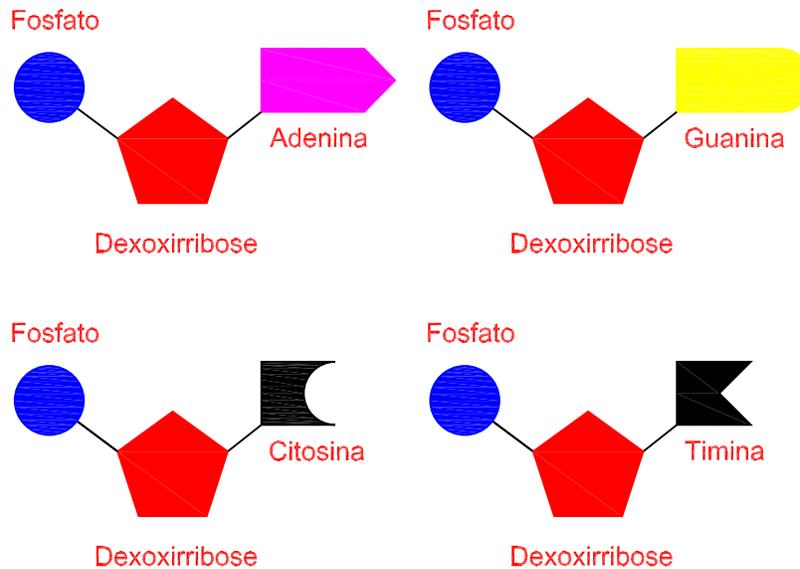


Figura 2.6: Nucleotídeos do DNA.

originando a famosa **dupla hélice**. As fitas são unidas pelos quatro diferentes nucleotídeos, representados por A, T, C e G, tendo suas bases projetadas para fora do esqueleto helicoidal da fita (Veja Figura 2.7). É interessante observar que essas duas fitas dispõem-se de modo antiparalelo, ou seja, cada fita tem direção oposta uma da outra. Cada dupla hélice de DNA é estruturada em uma construção simples sempre que houver um **A** sobre uma das fitas existirá um **T** sobre a outra, e a cada **C** corresponderá um **G** sobre a fita complementar (Figura 2.7). Este pareamento por **complemento** das duas fitas é tão forte que se as fitas complementares forem separadas e colocadas sob condições adequadas de temperatura e concentração salina, elas espontaneamente retomarão a sua estrutura fechada original.

A comparação clássica da estrutura da molécula de DNA com uma escada enrolada sobre si mesma agora faz sentido, os degraus da escada são formados pelos pares de bases complementares **A** \longleftrightarrow **T** e **C** \longleftrightarrow **G** sendo os corrimãos formados pela parte fixa da estrutura do nucleotídeo, isto é, açúcar mais fosfato. Portanto, conhecendo a seqüência de nucleotídeos de uma das fitas é possível deduzir a seqüência complementar. Logo, se temos a seqüência de nucleotídeos **A-C-T-G-C-A** em uma fita, a seqüência complementar é dada por **T-G-A-C-G-T**.

2.3.2 Estrutura da molécula de RNA

O RNA é formado por uma seqüência de nucleotídeos quimicamente semelhante à estrutura do DNA. A diferença estrutural que ocorre no RNA em relação ao DNA é que o açúcar presente é a **ribose** ao invés da desoxirribose, por isso é denominado, **ácido ribonucléico**. A molécula

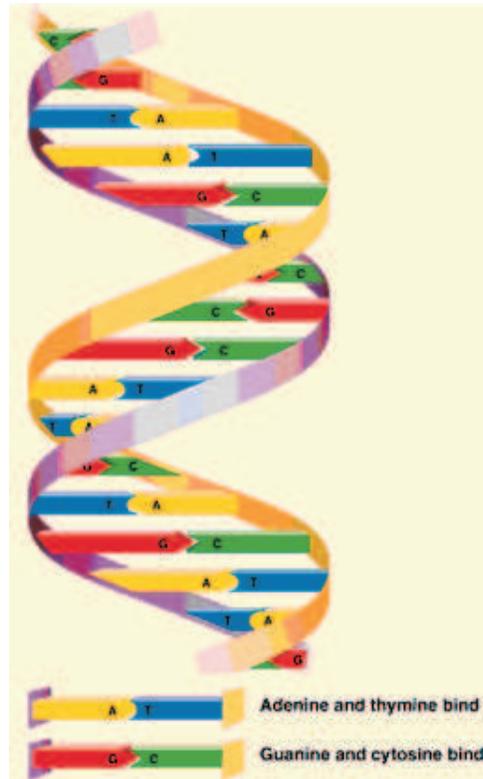


Figura 2.7: Estrutura da dupla hélice do DNA e a complementação entre as bases. <http://members.lycos.nl/TheDNApage/dnapix1b.html>

de ribose difere da molécula de desoxirribose por apresentar um grupo *OH* ligado ao carbono 2' ao invés de um átomo de hidrogênio (Veja Figura 2.8).

Outra diferença é que a base **timina** do DNA é substituída no RNA pela base **uracila**. Logo, as pirimidinas no RNA são a citosina e a uracila e as purinas são as mesmas do DNA. Enquanto que o DNA é formado por um duplo filamento de nucleotídeos o RNA apresenta somente um monofilamento de tamanho bem menor que as fitas do DNA.

As moléculas de RNA são classificadas em três tipos, de acordo com os papéis que desempenham nos processos de transferência de informação. Começamos pela molécula de RNA mais importante, aquela que copia a informação contida no DNA e transporta para o citoplasma. Esse RNA carrega uma mensagem específica do DNA até o citoplasma, por esse motivo é chamado de **mensageiro**, abreviadamente mRNA. Funcionando como um disquete de computador no qual um programa, contendo as etapas e ingredientes para a fabricação de uma proteína específica, pode ser copiado fielmente para facilitar o transporte.

Os outros dois tipos de RNA têm um papel auxiliar na fabricação de proteínas, ou seja, na síntese protéica. O papel do RNA **ribossômico** (rRNA) é somente estrutural: serve como matéria-prima ligando-se às proteínas para a composição dos ribossomos: organelas do cito-

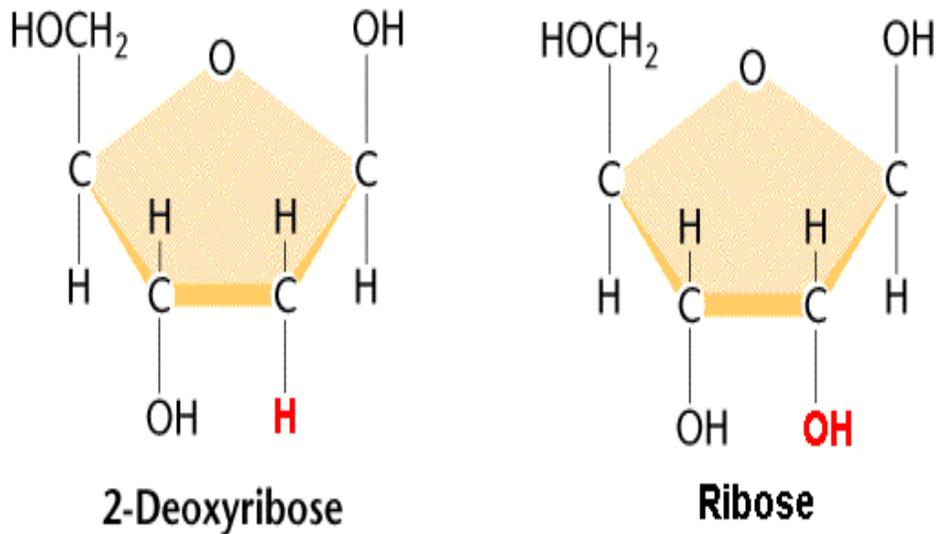


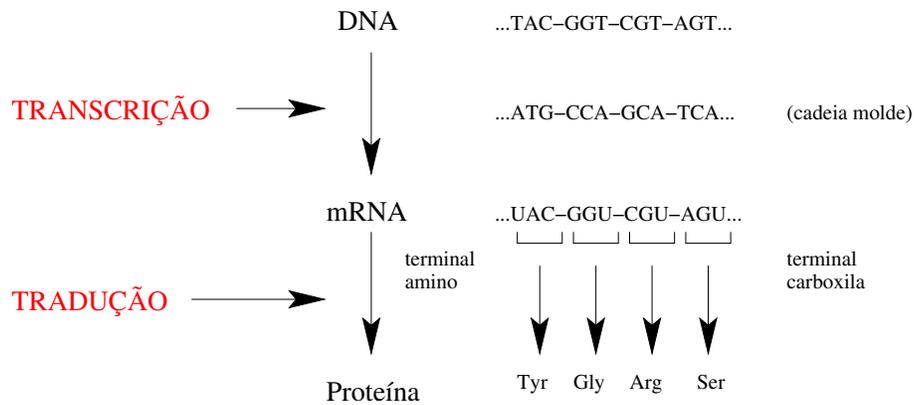
Figura 2.8: Estrutura da molécula de DNA e RNA. Klug, William S.; Cummings, Michael R. *Concepts of Genetics*, Paramus, NJ, U.S.A. Prentice Hall PTR, 5th Edition, 1997.

plasma nas quais ocorrem a síntese protéica. Por outro lado, o RNA **transportador** (tRNA) captura os aminoácidos que se encontram no citoplasma, carregando-os para os ribossomos, responsável em juntar cada aminoácido e formar uma nova seqüência polipeptídica.

2.3.3 O código genético

O DNA presente no núcleo controla toda a síntese de proteínas da célula. Esse controle é efetuado por meio de moléculas de RNA que o DNA fabrica e que passam para o citoplasma para fabricar proteínas. Para sintetizar uma certa proteína, uma determinada porção do DNA fabrica o mRNA, uma cópia semelhante a esta porção. Não é uma cópia igual pois a molécula de mRNA respeita as mesmas regras de ligação entre as bases nitrogenadas, sendo somente a base Timina (T) do DNA substituída pela Uracila (U) no RNA. A seguir exemplificamos como uma certa seqüência de bases presente em uma das fitas do DNA é **transcrita** para o RNA e, observando-se a Tabela 2.2, podemos **traduzir** uma seqüência de informação (código) determinando quais aminoácidos seriam adicionados na molécula de proteína:

Esta relação, exemplificada a seguir, entre a seqüência de base do DNA e a seqüência correspondente de aminoácidos na proteína foi sugerida por Crick em 1957. Esta linguagem que nos permite entender como se dá esta troca de informação entre DNA, RNA e proteína é chamada de **código genético** e é equivalente ao alfabeto de uma língua, com ponto final e inicial



e constando apenas de 4 letras, ou seja, os nucleotídeos A, C, G e U, onde o conjunto de 3 letras especifica uma palavra, tendo um total de 64 palavras, podendo ocorrer palavras com o mesmo sentido, isto é, sinônimas. Cada palavra representa um grupo de 3 nucleotídeos adjacentes (códon) que vão especificar um aminoácido na seqüência polipeptídica. A tabela do código genético estabelece como uma mensagem codificada no DNA é traduzida em proteínas (Tabela 2.2). Algumas propriedades importantes do código genético são explicitadas a seguir:

Tabela 2.2: Código Genético.

1ª base	2ª base				3ª base
	U	C	A	G	
U	UUU } Phe	UCU } Ser	UAU } Tyr	UGU } Cys	U
	UUC } Leu	UCC } Ser	UAC } Tyr	UGC } Cys	C
	UUA } Leu	UCA } Ser	UAA } Fim	UGA } Fim	A
	UUG } Leu	UCG } Ser	UAG } Fim	UGG } Trp	G
C	CUU } Leu	CCU } Pro	CAU } His	CGU } Arg	U
	CUC } Leu	CCC } Pro	CAC } His	CGC } Arg	C
	CUA } Leu	CCA } Pro	CAA } Gln	CGA } Arg	A
	CUG } Leu	CCG } Pro	CAG } Gln	CGG } Arg	G
A	AUU } Ile	ACU } Thr	AAU } Asn	AGU } Ser	U
	AUC } Ile	ACC } Thr	AAC } Asn	AGC } Ser	C
	AUA } Ile	ACA } Thr	AAA } Lys	AGA } Arg	A
	AUG } Met	ACG } Thr	AAG } Lys	AGG } Arg	G
G	GUU } Val	GCU } Ala	GAU } Asp	GGU } Gly	U
	GUC } Val	GCC } Ala	GAC } Asp	GGC } Gly	C
	GUA } Val	GCA } Ala	GAA } Glu	GGA } Gly	A
	GUG } Val	GCG } Ala	GAG } Glu	GGG } Gly	G

1. **Unidade:** Cada códon vai codificar um dos 20 aminoácidos, não existe um códon especificando mais que um aminoácido;

2. **Redundante ou degenerado:** Mais de um códon codifica um aminoácido. Há “códigos sinônimos” pois existem 64 códons para especificar 20 aminoácidos;
3. **Códon de iniciação:** O códon AUG atua tanto como códon de iniciação, sinalizando o início da mensagem codificadora da proteína, como também especifica o aminoácido metionina; Os códons GUG e UGG também podem atuar como códons de iniciação mas é pouco comum;
4. **Códon de finalização:** Os códons UAA, UAG, UGA sinalizam o fim da mensagem codificadora na proteína;
5. **Interpretação:** O código nos indica quais aminoácidos são correspondentes aos códons presentes no mRNA;
6. **Universalidade:** O código genético é dito universal devido ao fato de que cada aminoácido é codificado por um códon em qualquer organismo. Com exceções em mitocôndrias de algumas espécies e no genoma nuclear de alguns protozoários.

2.4 Dogma Central da Biologia Molecular

Nesta seção ao tratarmos do dogma central da biologia molecular, não ficamos limitados ao resumo esquemático $\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Proteína}$. Introduzimos, no entanto, aspectos importantes deste fluxo de informação genética do DNA à proteína, abordamos mecanismos biológicos que elucidam a nossa compreensão deste universo microbiológico.

A descoberta da estrutura do DNA, em 1953, e a subsequente elucidação de como o DNA gerencia a fabricação de proteína em um processo denominado **Dogma Central da Biologia**, foi uma descoberta fundamental para a biologia molecular. No entanto, a representação simplificada não reflete a riqueza das interações ocorridas entre DNA e RNA para a formação de proteína.

Estímulos que vem de dentro ou de fora da célula indicam que está na hora de fabricar uma determinada proteína. A partir daí, o nosso computador central (DNA) seleciona um dos arquivos, armazenado em sua memória, que será copiado no disquete (RNA). A informação está contida na seqüência de bases de uma das fitas que formam a dupla hélice do DNA (fita molde). Para que esta informação chegue ao RNA é necessário que a dupla hélice se abra separando as duas seqüências de polinucleotídeos. O RNA é então formado a partir da fita molde do DNA, copiando a informação de suas bases respeitando a propriedade de complementação e não esquecendo que no RNA temos a uracila (U) no lugar da timina (T). Portanto, a informação contida no DNA é transcrita para o RNA em um processo chamado de **transcrição**.

Curiosamente, não se sabe a razão pela qual nos organismos eucariotos, incluindo o homem, que nas instruções genéticas existem partes que podem ou não conter informação útil para a fabricação de uma nova seqüência polipeptídica. As porções do DNA com a informação necessária para codificar proteínas são chamadas de **éxons** e as porções sem sentido são chamadas de **íntrons**. Pelo menos até agora não foi esclarecido uma funcionalidade para estas porções aparentemente sem utilidade. Logo, quando um arquivo do DNA é selecionado, o RNA copia fielmente tanto os íntrons como os éxons, entretanto, antes do RNA deixar o núcleo, as porções contendo os íntrons são retiradas. Assim, no mRNA maduro, aquele que efetivamente carrega a mensagem para o citoplasma, estão presentes somente as porções que correspondem aos éxons.

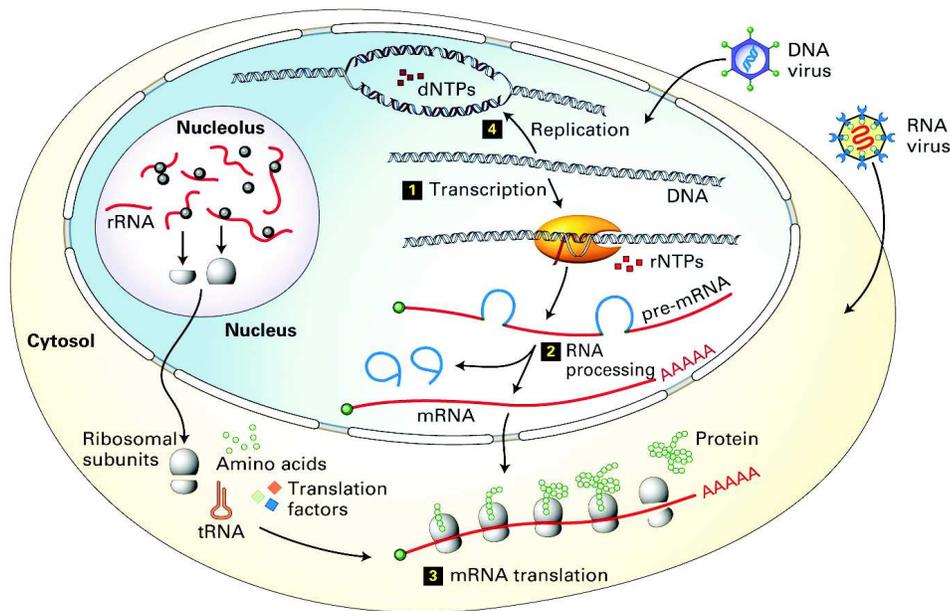


Figura 2.9: Visão geral dos três processos básicos da genética molecular. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

É no citoplasma que ocorre a síntese protéica. As informações que especificam a proteína estão contidas no RNA mensageiro (mRNA) que liga-se ao ribossomo e, à medida que a informação começa a ser traduzida no ribossomo a cadeia polipeptídica vai sendo montada. Este processo é conhecido como **tradução**. O RNA transportador (tRNA), presente no citoplasma, é o responsável em identificar e transportar os aminoácidos corretos para o ribossomo, de acordo com a seqüência de códons presente no mRNA. Os aminoácidos juntam-se por meio de ligações peptídicas para compor a nova seqüência polipeptídica (Veja Figura 2.9). Quando a seqüência polipeptídica está completa, esta desprende-se do ribossomo formando a proteína primária e, conforme a funcionalidade da proteína, dobra-se sobre si mesma até adquirir as conformações necessárias para sua função.

O modelo proposto DNA \rightarrow RNA \rightarrow Proteína, aparentemente simples, porém repleto de minúcias, nos mostra que os mecanismos biológicos vão se tornando cada vez mais complexos e mais ricos em detalhes. Note que o mérito do Dogma Central está, justamente, em poder sintetizar em uma linguagem acessível as principais interações e resultados que ocorrem entre a tríade.

2.5 Conclusões

Neste capítulo, apresentamos os principais conceitos biológicos e bioquímicos e, com isso, mostramos o grau de complexidade que envolve a maquinaria molecular. Mais especificamente exibimos o processo de síntese protéica e a importância do código genético para a codificação da informação, presente no DNA, para a especificação de proteínas.

Iniciamos com a descrição da célula, unidade fundamental da vida, descrevendo suas características, funções e diferenças. Em seguida, apresentamos o conceito de moléculas de uma célula, em especial, a estrutura e função das proteínas. Em seguida discutiremos sobre as características do DNA e RNA e também sobre o **Dogma Central da Biologia**. Este dogma exerce papel relevante na modelagem descrita no Capítulo 4. Finalmente, abordamos o código genético, de fundamental importância para o entendimento da nossa proposta de trabalho, o qual será associado a um modulador.

Revisão de Conceitos Matemáticos

Neste capítulo apresentamos uma breve revisão de geometria: definimos espaços métricos, apresentamos alguns exemplos de métricas, em particular, as métricas: Euclidiana e de Hamming. Abordamos com maiores detalhes as definições e resultados básicos sobre teoria de grupos, parte fundamental no desenvolvimento deste trabalho. Para maiores detalhes consultar as referências (Fraleigh 2003, Armstrong 1988, Garcia & Lequain 2002, Allenby 1993, Gilbert 1976, Firer 2001).

3.1 Espaços métricos

Definição 1 Um conjunto não vazio \mathbb{E} é denominado **espaço métrico** se existe uma função $d : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}$, chamada *métrica*, que satisfaça as condições:

- (i) $d(x, y) \geq 0$ e $d(x, y) = 0$ se, e somente se, $x = y$, $\forall x, y \in \mathbb{E}$;
- (ii) $d(x, y) = d(y, x)$, $\forall x, y \in \mathbb{E}$;
- (iii) $d(x, z) \leq d(x, y) + d(y, z)$, $\forall x, y$ e $z \in \mathbb{E}$.

Denotamos um espaço métrico por (\mathbb{E}, d) .

Exemplo 1 Seja \mathbb{E} um conjunto finito qualquer. Sempre é possível definir uma métrica em \mathbb{E} , chamada **métrica discreta** do tipo;

$$d(x, y) = \begin{cases} 1, & \text{se } x \neq y \\ 0, & \text{se } x = y \end{cases}$$

Exemplo 2 Seja \mathbb{E}^n o espaço n -dimensional, onde, $E^n = E \times \dots \times E$ (n -vezes), definido com a seguinte métrica $d_H(x, y) = \sum_{i=1}^n d(x_i, y_i)$, tal que;

$$d(x_i, y_i) = \begin{cases} 1, & \text{se } x_i \neq y_i \\ 0, & \text{se } x_i = y_i \end{cases}$$

Neste caso a métrica d_H é chamada **métrica de Hamming**.

Exemplo 3 Seja \mathbb{R}^n o espaço n -dimensional, definido com a seguinte métrica;

$$d_E(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

Neste caso a métrica d_E é chamada **métrica Euclidiana**.

3.2 Grupos

Definição 2 Um **grupo** é um par ordenado (G, \circ) onde G é um conjunto não vazio munido de uma operação binária \circ , que é uma aplicação

$$\begin{aligned} \circ: G \times G &\longrightarrow G \\ (g, h) &\longmapsto g \circ h \end{aligned}$$

satisfazendo os seguintes axiomas:

- (i) A operação é **associativa**, isto é, $g \circ (h \circ k) = (g \circ h) \circ k$, $\forall g, h, k \in G$;
- (ii) Existe um **elemento neutro**, isto é, $\exists e \in G$ tal que $e \circ g = g \circ e = g$, $\forall g \in G$;
- (iii) Todo elemento possui um **elemento inverso**, isto é, $\forall g \in G$, $\exists h \in G$ tal que $g \circ h = h \circ g = e$.

O grupo é dito **abeliano** ou **comutativo** se:

- (iv) A operação é comutativa, isto é, $g \circ h = h \circ g$, $\forall g, h \in G$.

Observação 1 O elemento neutro e em (ii) é único em G , assim como o elemento inverso h em (iii). Denotamos o elemento inverso de g por g^{-1} . Para simplificar a notação, quando não houver dúvidas, escrevemos também gh ao invés de $g \circ h$.

Definição 3 Seja G um grupo finito. O número de elementos de G é denominado **ordem** de G , denotado por $|G|$. Em geral, para qualquer conjunto finito S , $|S|$ representa a quantidade total de elementos do conjunto S .

3.3 Subgrupos, grupos cíclicos

A seguir, apresentamos a definição de subgrupo de um grupo G . Conceitualmente, entendemos por subgrupo H de G , um subconjunto de G , denotado por $H \subset G$, de tal maneira que H tenha a mesma estrutura de grupo que G . Em outras palavras, a operação nos elementos de H é a mesma que nos elementos de G .

Sabemos que $H \subset G$ e, considere a operação binária \circ em $H \times H$. Como todos os elementos de G satisfazem o Axioma (i), em particular, tal axioma é verdadeiro para todos os elementos de H . Isso induz a seguinte definição:

Definição 4 Dado um grupo G com $H \subseteq G$, H é um **subgrupo** de G (denotamos $H \leq G$) se satisfizer as seguintes condições:

$$(i) \quad h_1 h_2 \in H, \quad \forall h_1, h_2 \in H. \quad (H \text{ é fechado em relação } \circ)$$

$$(ii) \quad \exists e_H \in H, \quad \forall h \in H. \quad (\text{Existência do elemento neutro em } H)$$

$$(iii) \quad h^{-1} \in H, \quad \forall h \in H. \quad (\text{Existência do inverso em } H)$$

Verificamos que a condição (ii) é consequência de (iii), e assim, para demonstrarmos que um dado subconjunto $H \subset G$ é um subgrupo de G é suficiente mostrar que as condições (i) e (iii) são válidas.

Observe que se H é subgrupo de G , então vale a seguinte relação:

$$H \text{ é subgrupo de } G \iff \forall g, h \in H \text{ temos } g * h^{-1} \in H.$$

Note que se H satisfaz as condições de subgrupo, H também satisfaz todos os axiomas de grupo e, portanto, H também é um grupo. A seguir, enunciamos alguns resultados bem conhecidos na teoria de grupos.

Definição 5 Dado um grupo G e um elemento $g \in G$, se

$$G = \{g^n \mid n \in \mathbb{Z}\},$$

então g é um **elemento gerador** de G e o grupo $G = \langle g \rangle$ é dito **cíclico**. Se g é o menor elemento de um grupo G e o subgrupo $\langle g \rangle$ é finito, então a **ordem de** g é a ordem deste subgrupo cíclico $\langle g \rangle$. Quando dizemos que um elemento $g \in G$ é de ordem n , isto equivale a dizer que $g^n = e$, onde, n é o menor inteiro positivo. Caso contrário, dizemos que o elemento g possui **ordem infinita**.

Proposição 1 Se g é um elemento de ordem m em um grupo G , então $H = \{g^r \mid r \in \mathbb{Z}\}$ é um subgrupo de ordem m em G . Tal subgrupo é dito **subgrupo cíclico gerado por** g .

Teorema 1 Seja G um grupo finito de ordem n que possui um elemento g de ordem n . Então G é um grupo cíclico gerado por g .

3.4 Classes laterais e o teorema de Lagrange

Definição 6 Seja G um grupo e H um subgrupo de G . Para cada $g \in G$, a **classe lateral** gH é definida como sendo o conjunto

$$gH = \{gH : h \in H\}.$$

Lema 1 Se H é um subgrupo de G , e $g_1, g_2 \in G$, então

$$g_1H = g_2H \iff g_2^{-1}g_1 \in H.$$

Lema 2 Se H é um subgrupo de G então quaisquer classes laterais de H ou são idênticas, ou são disjuntas. Em símbolos,

$$\forall g_1, g_2 \in H \implies g_1H = g_2H$$

ou

$$g_1H \cap g_2H = \emptyset$$

Definição 7 O *índice* de H em G é definido como sendo o número de classes laterais distintas de H em G , e será denotado por $|G : H|$. Em símbolos,

$$|G : H| = |\{gH : g \in G\}|.$$

Teorema 2 Se H é um subgrupo de um grupo finito G , então

$$|G : H| \times |H| = |G|.$$

Tendo como base as deduções anteriores, podemos demonstrar o Teorema de Lagrange:

Teorema 3 [Teorema de Lagrange] Se H é subgrupo de um grupo finito G , então $|H|$ divide $|G|$. Em particular, para qualquer elemento $g \in G$, $|g|$ divide $|G|$.

3.5 Homomorfismo e isomorfismo de grupos

Definição 8 Um **homomorfismo** de um grupo (G, \cdot) e um grupo $(K, *)$ é uma aplicação $f : G \longrightarrow K$ que preserva as estruturas dos dois grupos, ou seja,

$$f(a \cdot b) = f(a) * f(b), \quad \forall a, b \in G.$$

Definição 9 Um **isomorfismo** é um homomorfismo bijetivo entre os grupos (G, \cdot) e $(K, *)$.

Dizemos que dois grupos G e K são **isomorfos**, denotado por $G \simeq K$, se existe um isomorfismo entre os grupos G e K .

Grupos cíclicos possuem propriedades importantes, a saber,

Teorema 4 Grupos cíclicos de mesma ordem são isomorfos.

Proposição 2 Elementos correspondentes em grupos isomorfos possuem a mesma ordem.

3.6 Grupos de permutações e o Teorema de Cayley

Definição 10 Dado um conjunto não vazio X , uma bijeção $\sigma : X \rightarrow X$ é denominada uma **permutação** do conjunto X .

A coleção de todas as permutações de X forma o grupo de permutações $S(X)$ onde a operação é a composição de funções. Verifica-se facilmente os axiomas de grupo sobre $S(X)$.

Se $\alpha : X \rightarrow X$ e $\beta : X \rightarrow X$ são permutações, a função composta $\alpha\beta : X \rightarrow X$ definida por $\alpha\beta(x) = \alpha(\beta(x))$ também é uma permutação. A composição de funções é associativa e a função identidade e é definida por $e(x) = x \forall x \in X$.

Cada permutação α é uma bijeção e, portanto, possui uma inversa $\alpha^{-1} : X \rightarrow X$ que também é uma permutação; a permutação inversa satisfaz $\alpha^{-1}\alpha = e = \alpha\alpha^{-1}$. Se X é um conjunto infinito então $S(X)$ é um grupo infinito; caso contrário $S(X)$ é um grupo finito. Se X é finito e $|X| = n$, isto é, $X = \{x_1, \dots, x_n\}$, então as permutações desses n elementos são dadas por: n possibilidades de preenchimento para a primeira posição, $n - 1$ possibilidades para a segunda, ..., 2 possibilidades para a $(n - 1)$ -ésima e 1 possibilidade para a n -ésima posição. Assim, pelo princípio multiplicativo, temos $n!$ permutações em $S(X)$, ou seja, $|S(X)| = n!$

Neste trabalho, estamos interessados em estudar permutações quando $X = \{1, 2, 3, \dots, n\}$. Neste caso, representamos $S(X)$ por S_n e o denominamos **grupo simétrico** com n elementos.

De acordo com o Princípio Multiplicativo obtemos o seguinte resultado:

Proposição 3 $|S_n| = n!$

É importante salientar que grupos de permutações de conjuntos que contêm o mesmo número de elementos são isomorfos.

Se $\sigma \in S_n$ e $j \in \{1, 2, \dots, n\}$, a notação $\sigma(j)$ denota o valor da bijeção σ no ponto j . Desta forma, denotamos a permutação σ pelo seguinte diagrama:

$$\sigma = \begin{pmatrix} 1 & \dots & n \\ \sigma(1) & \dots & \sigma(n) \end{pmatrix},$$

exibindo visualmente o domínio e o contradomínio da bijeção σ considerada.

Observação 2 Se $\pi, \sigma \in S_n$ são duas permutações, o produto $\pi \circ \sigma$ é uma permutação obtida aplicando-se primeiramente σ e depois π . Adota-se tal procedimento de acordo com a noção de composição de funções, ou seja, $(\pi \circ \sigma)(x) = \pi(\sigma(x))$.

Definição 11 Uma permutação $\alpha \in S_n$ é chamada **r -ciclo** ou **permutação r -cíclica** se existem elementos distintos $a_1, \dots, a_r \in \{1, \dots, n\}$ tais que

$$\alpha(a_1) = a_2, \alpha(a_2) = a_3, \dots, \alpha(a_{r-1}) = a_r, \alpha(a_r) = a_1,$$

onde $\alpha(j) = j, \forall j \in \{1, \dots, n\} \setminus \{a_1, \dots, a_r\}$; tal r -ciclo será denotado por $(a_1 \dots a_r)$; o número r é chamado de **comprimento** do ciclo. Os 2-ciclos são também chamados **transposições**.

Proposição 4 Um r -ciclo em S_n tem ordem r .

Observação 3 Permutações que não são cíclicas podem ser decompostas em dois ou mais ciclos (veja (Fraleigh 2003)).

Definição 12 Se α é uma permutação no grupo S_n e $j \in \{1, 2, 3, \dots, n\}$, a **órbita** de j em relação a α é o conjunto de todos os elementos distintos de $j, \alpha(j), \alpha^2(j), \alpha^3(j), \dots$.

Observação 4 Pode-se decompor uma permutação entre as suas diferentes órbitas, onde cada órbita está associada a um ciclo.

Exemplo 4 Se a órbita de j é igual $\{a, b, c\}$, onde $a, b, c \in \{1, \dots, n\}$, seu ciclo correspondente é (abc) , isto é, um 3-ciclo.

Se não existem números iguais entre ciclos diferentes, estes ciclos são denominados **disjuntos**. Formalmente,

Definição 13 Seja $\alpha \in S_n$ um r -ciclo e $\beta \in S_n$ um s -ciclo. As permutações α e β são **disjuntas** se nenhum elemento de $\{1, 2, \dots, n\}$ é “movido” por ambas, isto é,

$$\forall j \in \{1, 2, \dots, n\} \implies \alpha(j) = j \text{ ou } \beta(j) = j.$$

A seguir, apresentamos uma proposição fundamental na teoria das permutações:

Proposição 5 Toda permutação pode ser decomposta como um produto de ciclos disjuntos.

O próximo teorema é importante para a fundamentação do nosso trabalho. Este afirma que todo grupo finito é isomorfo a um subgrupo do grupo simétrico.

Teorema 5 [Teorema de Cayley] Todo grupo (G, \cdot) é isomorfo a um subgrupo de seu grupo simétrico $(S(G), \circ)$.

Corolário 1 Se G é um grupo finito de ordem n , então G é isomorfo a um subgrupo de S_n .

3.7 Ação de grupo

Iniciamos esta subseção definindo ação de grupos:

Definição 14 *Seja G um grupo e X um conjunto não vazio. Diz-se que G **age** em X se existe uma aplicação $\tau : G \times X \longrightarrow X$, denotada por $\tau(g, x) = gx, \forall g \in G, x \in X$, satisfazendo as seguintes propriedades:*

- (i) *Para todo $x \in X$, $e_G x = x$, onde e_G é a identidade de G ;*
- (ii) *Para todo $g, h \in G, x \in X$, temos $g(hx) = (gh)x$.*

A condição (i) indica que a ação do elemento identidade de G é sempre trivial, ou seja, fixa todo elemento de X . A condição (ii) relaciona a ação de grupo à sua operação.

Sempre que um grupo age em um conjunto, tem-se uma **ação de grupo** (onde (i) e (ii) são satisfeitas).

Note que os elementos de X e de G não necessariamente são os mesmos.

Proposição 6 *Seja o grupo G agindo sobre o conjunto X , onde $x_1, x_2 \in X$ e $g \in G$. Se $gx_1 = gx_2$, então $x_1 = x_2$.*

Proposição 7 *Seja G um grupo que age no conjunto X . Então, para todo $g \in G$ e $x, y \in X$, temos que*

$$gx = y \iff g^{-1}y = x.$$

Proposição 8 *Um grupo G age sobre o conjunto X se, e somente se, existe um homomorfismo $\gamma : G \longrightarrow S(X)$.*

O estabilizador de um elemento $x \in X$ é um conceito de fundamental importância na teoria de grupos, pois dele decorre o conceito de órbita de x .

Definição 15 *Dado $x \in X$, o **estabilizador** de x , denotado por $Stab(x)$, é o conjunto dos elementos de G que fixam x , ou seja,*

$$Stab(x) = \{g \in G : gx = x\}$$

O estabilizador é um subgrupo:

Proposição 9 *Seja G um grupo que age em X . Então, para todo $x \in X$, $Stab(x)$ é subgrupo de G .*

Definição 16 *O conjunto de todas as imagens de um elemento $x \in X$ sob a ação do grupo G é denominado **órbita** de x em relação a G e definida por*

$$O_x = \{g(x) : g \in G\}$$

A órbita de x é a classe de equivalência de x de acordo com a relação de equivalência em X , tal que,

$$x \equiv y \iff \exists g \in G : y = g(x).$$

Se γ é uma permutação em S_n , o subgrupo gerado por γ age sobre o conjunto $\{1, 2, \dots, n\}$, e esta definição de órbita está de acordo com a definição anteriormente mencionada. O Teorema da órbita-estabilizador é um método alternativo para calcular o número de elementos de um grupo G :

Teorema 6 [*Teorema da órbita-estabilizador*] *Seja G um grupo que age em X . Então para cada $x \in X$,*

$$|O_x| \cdot |Stab(x)| = |G|.$$

Corolário 2 *Seja G um grupo finito que age em um conjunto X . Então, o número de elementos em cada órbita é um divisor da ordem de G .*

Definição 17 *Se G é um grupo agindo em X , define-se o **fixador** de $g \in G$, denotado por $Fix(g)$, como sendo o conjunto de elementos de X fixado por g , ou seja,*

$$Fix(g) = \{x \in X : g(x) = x\}.$$

Muitos problemas combinatoriais podem ser solucionados pela aplicação do Teorema de Burnside:

Teorema 7 [*Teorema de Burnside*] *Seja G um grupo finito agindo em um conjunto finito X . Se N é o número de órbitas de X em relação a G , então*

$$N = \frac{1}{|G|} \sum_{g \in G} |Fix(g)|.$$

Demonstração: Considere o conjunto $S = \{(x, g) \in X \times G : g(x) = x\}$ escrito de duas maneiras diferentes. Observe o esquema a seguir, cujas colunas são indexadas pelos elementos de X e as linhas são indexadas pelos elementos de G . O valor 1 é alocado na posição $g(x)$ se $g(x) = x$, e 0 caso contrário.

- (ii) O grupo G agindo sobre X é o grupo simétrico do triângulo equilátero, ou seja, $S_3 = \{\rho_0, \rho_1, \rho_2, \eta_1, \eta_2, \eta_3\}$, onde, $\rho_0 = i\frac{2\pi}{3}$ corresponde a rotação no sentido anti-horário através da reta que passa verticalmente pelo baricentro do triângulo e η_j corresponde a rotação espacial (reflexão) através da reta que liga o vértice do triângulo ao ponto médio da face oposta;
- (iii) ρ_0 é a identidade, ou seja, rotação de 360° . Neste caso, a identidade leva as 64 possibilidades de rotulamento dos vértices nele mesmo;
- (iv) Para não mudar de letras em relação as rotações de 120° e 240° , os vértices devem ter uma única cor para todos. Logo, temos 4 letras para cada um dos três vértices;
- (v) Para cada reta fixada em η_j , temos que o vértice em que a reta passa está fixo, assim como a letra, o lado oposto ao vértice contém os outros dois vértices com letras idênticas, variando entre as quatro letras dadas. Logo, temos 16 possibilidades de rotulamento dos vértices para cada reflexão η_j ;
- (vi) $|S_3| = 3! = 6$.

Vamos calcular o $|Fix(g)|$ para cada elemento de S_3 . Obtendo o seguinte resultado

$$\begin{aligned} |Fix_{\rho_0}| &= 64, \\ |Fix_{\rho_1}| &= |Fix_{\rho_2}| = 4, \\ |Fix_{\eta_1}| &= |Fix_{\eta_2}| = |Fix_{\eta_3}| = 16. \end{aligned}$$

Logo, utilizando o Teorema 7 temos que

$$\begin{aligned} N &= \frac{1}{|S_3|} \left(\sum_{g \in S_3} |Fix(g)| \right) = \frac{1}{6}(120) \\ N &= 20. \end{aligned}$$

Portanto, temos 20 maneiras distintas de rotular os vértices do triângulo com quatro letras.

Resultado interessante pois o código genético possui 64 elementos associados aos 20 tipos de aminoácidos mais os códons de terminação. Será possível encontrar uma ação τ e um grupo G que agindo no conjunto $X = \{1, 2, \dots, 64\}$, conduza a uma partição $N = 20$, desconsiderando os códons de terminação? Mais intrigante ainda, é verificar se o número de elementos contido em cada órbita gerada é igual ao número de códons que representam cada aminoácido.

3.8 Representação de grupos

Sabemos que o conjunto $GL(V)$ das transformações lineares invertíveis de um espaço vetorial V , com a operação usual de composição, é um grupo. No caso de V ter dimensão finita n , ao escolhermos uma base, podemos identificar canonicamente o conjunto $GL(V)$ com $GL_n(n, F)$, o conjunto das matrizes $n \times n$ de determinante não nulo (sobre o corpo base de V , em nosso caso $F = \mathbb{R}$), onde a operação em $GL(V)$ é compatível com o produto usual de matrizes.

Definição 18 Uma **representação** de um grupo G em $GL(V)$ (por abuso de notação, escrevemos uma representação em V) é um homomorfismo de grupos $\rho : G \longrightarrow GL(V)$.

Definição 19 A **dimensão** da representação ρ é definida como sendo a dimensão do espaço vetorial V .

Estamos interessados somente nas representações sobre o espaço de dimensão finita. A matriz de representação pode ser pensada como uma representação de G sobre o espaço \mathbb{R}^n dos vetores coluna.

Exemplo 6 Sejam o grupo S_n de permutações dos elementos $\{1, \dots, n\}$ e o espaço vetorial V de dimensão n e base $\{e_1, \dots, e_n\}$. A aplicação

$$\begin{aligned} \rho : S_n &\longrightarrow GL(V) \\ \sigma &\longmapsto \rho(\sigma) : \begin{array}{c} V \longrightarrow V \\ \sum_{i=1}^n \alpha_i e_i \longmapsto \sum_{i=1}^n \alpha_i e_{\sigma(i)} \end{array}, \end{aligned}$$

é uma representação. Observe que S_n “age” em V via permutação de coordenadas.

Ilustração 1 Considere o grupo $S_3 = \{e, (123), (132), (23), (13), (12)\}$ e o espaço vetorial \mathbb{R}^3 , cuja base canônica é $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$. Desenvolveremos o procedimento do Exemplo 6 para o elemento (123) , em particular:

$$\begin{aligned} \rho : S_3 &\longrightarrow GL(3, \mathbb{R}) \\ (123) &\longmapsto \rho[(123)] : \begin{array}{c} \mathbb{R}^3 \longrightarrow \mathbb{R}^3 \\ \sum_{i=1}^3 \alpha_i e_i \longmapsto \sum_{i=1}^3 \alpha_i e_{\sigma(i)} \end{array} \end{aligned}$$

obtendo a seguinte representação matricial para este elemento do grupo S_3 :

$$\rho \left[\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \right] ((\alpha_1, \alpha_2, \alpha_3)) = (\alpha_3, \alpha_1, \alpha_2) \longleftrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Deve-se observar, no entanto, a relação entre a permutação ocorrida nos índices dos escalares e sua base de fixação.

Utilizando o mesmo procedimento para os elementos restantes do grupo S_3 , chegamos à seguinte representação matricial:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (123), \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (132),$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (12), \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = (13), \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = (23).$$

Exemplo 7 D_{2n} : Considere o grupo diedral $D_n = \langle g, h | g^2 = h^n = e \rangle$. Então, a aplicação

$$g \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$h \mapsto \begin{pmatrix} \cos \frac{2\pi}{n} & \sin \frac{2\pi}{n} \\ -\sin \frac{2\pi}{n} & \cos \frac{2\pi}{n} \end{pmatrix}$$

pode ser estendida (de modo único) a uma representação

$$\rho : D_n \longrightarrow GL(2, \mathbb{R}).$$

3.9 Classes conjugadas

Definição 20 Seja H um subconjunto do grupo G e $g \in G$. O subconjunto $gHg^{-1} = \{ghg^{-1} : h \in H\}$ é denominado **conjugado** de H por g , denotado por H^g . Se H e P são subconjuntos de G , dizemos que H é conjugado de P , em G , se existe um elemento $g \in G$ tal que $gHg^{-1} = P$.

Observação 5 Da definição anterior segue que $P^{g^{-1}} = H$, onde H e P são conjugados em G . Se $H^g = P$, onde H e P são conjuntos unitários, escrevemos $gxg^{-1} = y$ e dizemos que x é conjugado de y em G .

A conjugação determina uma relação de equivalência nos subconjuntos de G , em particular, em todos os subgrupos de G . Neste caso as classes de equivalência são chamadas **classes conjugadas**. Além disso, os elementos que pertencem à mesma classe tem a mesma ordem. A classe conjugada de $h \in G$, denotada por $Cl(h)$, é o conjunto de todos os conjugados de h , isto é,

$$Cl(h) = \{ghg^{-1} : g \in G\}.$$

Exemplo 8 Considere o grupo de permutação S_3 onde os elementos (12) e (23) são conjugados, pois $(13)(12)(13)^{-1} = (23)$. Utilizando o mesmo raciocínio para (23) e (31) , concluímos que os elementos (12) , (23) e (31) pertencem à mesma classe conjugada. Logo, as classes conjugadas de S_3 são

$$\{e\}, \{(23), (13), (12)\} \quad e \quad \{(123), (132)\}.$$

Exemplo 9 Considere os subgrupos G_1 e G_2 , contidos em S_4 , tal que

$$G_1 = \{e, (1243), (14)(23), (1342)\} \quad e \quad G_2 = \{e, (1324), (12)(34), (1423)\}.$$

Sabemos que G_1 e G_2 são conjugados, pois $\exists g \in S_4$ tal que $(123)G_2(123)^{-1} = G_1$, onde, $(123)^{-1} = (132)$. Então, os grupos G_1 e G_2 pertencem à mesma classe conjugada. Como grupos pertencentes à mesma classe conjugada são isomorfos, logo $G_1 \simeq G_2$.

Características e Propriedades dos Sistemas de Comunicação e Genético

O principal objetivo deste capítulo é apresentar uma breve introdução sobre sistemas de comunicações, de maneira que possa ser útil no entendimento da modelagem a ser descrita no Capítulo 5. Iniciamos com o conceito de comunicação e, em seguida, exemplos de sinais e de aplicações dos sistemas de comunicações são descritos. Finalmente, apresentamos o modelo simplificado de sistemas de comunicação em forma de diagrama de blocos, fundamental na modelagem a ser descrita no Capítulo 5, e as conclusões deste capítulo. Para maiores detalhes consultar as referências (Lin & Costello Jr. 2004, Hefez & Villela 2002, Gilbert 1976, May, Rintoul, Johnston, Hart, Watson & Pryor 2003, May, Vouk, Bitzer & Rosnick 2004).

4.1 Sistema de Comunicação

O conceito de transmissão da informação, em telecomunicações, pode ser entendido como sendo o envio da informação de uma lugar para outro, da origem ao destino. Para que se possa realizar uma comunicação, é necessário a utilização de sinais. O **sinal** é um fenômeno físico associado à informação. Em geral, entende-se que um sinal é uma seqüência de estados em um sistema de comunicação que processa uma mensagem. A definição pode mudar de acordo com o contexto em que se está trabalhando.

Uma simplificação de um sistema de telecomunicações pode ser vista da seguinte maneira:

Fonte → **Transmissor** → **Meio + Ruído** → **Receptor** → **Destino**

Em um sistema de comunicação, o transmissor recebe uma mensagem, e a transforma em um sinal, que é enviado pelo sistema de comunicação até o receptor, que processa o sinal e libera uma mensagem. Como um exemplo, o texto "Alô, quem esta falando", ao ser transmitido

via telégrafo, é convertido primeiramente em traços, pontos e pausas, e o telegrafista aciona o dispositivo de telégrafo criando um sinal, que é a tensão elétrica a ser transmitida por um par de fios. Do outro lado, o receptor recebe um sinal, também na forma de uma tensão elétrica em um par de fios. Por fim, o sinal é processado, gerando uma mensagem que se espera ser igual ao texto original transmitido.

O sinal transmitido pelo telegrafista pode sofrer distorções durante a transmissão. Estas distorções podem ser tanto mudanças determinísticas e reversíveis do sinal, quanto a adição de ruídos completamente aleatórios. Estas distorções podem ser intensas o suficiente para modificar o conteúdo da mensagem, originando erros. Então, a teoria da codificação, cuja abordagem será apresentada posteriormente, surgiu na tentativa de solucionar tais problemas.

4.2 Exemplos de sinais

Os sistemas de comunicações podem possuir diversas camadas em que o sinal gerado por uma camada é a mensagem a ser transmitida pela camada inferior. As mensagens e os sinais podem ser tanto discretos quanto contínuos:

- **Contínuo:** São seqüências de valores, normalmente definidos em instantes de tempo periódicos;
- **Discreto:** Possuem estados definidos em qualquer instante de tempo.

Os sistemas de telecomunicações podem ser divididos em:

1. **Sistemas Analógicos:** são aqueles que conservam a forma dos sinais desde a fonte ao destino;
2. **Sistemas Digitais:** são aqueles em que a forma do sinal transmitido pode ser diferente do sinal original.

Os sinais digitais só podem assumir valores discretos, enumeráveis, ou inteiros, normalmente de um conjunto limitado de possíveis valores. De outro modo, os sinais analógicos podem assumir qualquer valor real.

4.3 Teoria da codificação: uma introdução

A teoria da codificação e suas técnicas surgiram da necessidade do mecanismo de correção de erros em um sistema de comunicação. Mecanismos de controle de erro podem ser caracterizados

como corretores de erros na mensagem enviada e de controle de erros na retransmissão, garantindo sua confiabilidade. Nos mecanismos corretores de erros da mensagem enviada supõe-se que os erros irão ocorrer e, portanto, tais mecanismos são capazes de corrigir os erros quando aplicados à mensagem recebida.

As técnicas de controle de erros na retransmissão permitem detectar os erros na mensagem recebida e solicitam a retransmissão da mensagem. O sistema de comunicação ilustrado na Figura 4.1 mostra como a teoria de códigos corretores de erros (CCE) é incorporada ao sistema. A informação digital é codificada pelo codificador de canal e preparada para a transmissão (modulação). A informação é transmitida através de um canal potencialmente ruidoso, e esta, provavelmente poderá ser corrompida pelo ruído. O sinal recebido no receptor é preparado para que o processo de demodulação possa ocorrer, e logo em seguida decodificada em um decodificador de canal.

O processo de decodificação envolve a retirada dos erros introduzidos durante a transmissão. O mecanismo de decodificação é eficiente desde que os erros não excedam a capacidade de correção do código. Enfatizaremos quatro elementos deste sistema de comunicação: o codificador, modulador, canal de transmissão e decodificador.

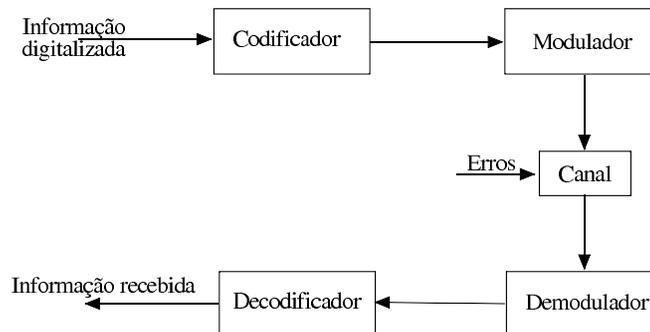


Figura 4.1: Sistema de comunicação e codificação.

4.3.1 Códigos corretores de erros

A teoria da codificação é baseada em corpos finitos, estes são conhecidos como corpos algébricos de Galois. Um corpo finito q -ário, onde q é uma potência de primo, denotado por F_q , contendo q elementos, consiste de um conjunto finito de símbolos munido com duas operações, denotadas por $(+)$ e (\star) ,

No caso em que $q = 2$, tem-se o corpo binário F_2 com as seguintes características:

- Conjunto finito de símbolos $S = \{0, 1\}$;
- Operações módulo-2: adição $(+)$ e multiplicação (\star) ;

- Operações inversas correspondentes.

4.3.2 Codificador

A informação digital consistindo de k dígitos entra no codificador, é codificada e transmitida como uma palavra-código de n dígitos. Na palavra-código introduz-se $n - k$ dígitos de redundância, e os k dígitos de informação são conservados, onde $n > k$. A Figura 4.2 ilustra este esquema mencionado.

A taxa de codificação R é denotada por

$$R = \frac{k}{n},$$

e representa a razão entre os números de dígitos na entrada no codificador e o número de dígitos na saída do codificador. Quanto menor a taxa de codificação maior o grau de redundância $n - k$, ou seja, mais dígitos podem ser utilizados para a correção e controle dos erros.

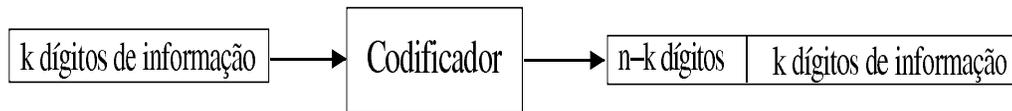


Figura 4.2: Codificador de k dígitos.

O codificador combina os k símbolos de entrada com os $n - k$ símbolos redundantes, baseado em algoritmos determinísticos, embora métodos de codificação aleatórios possam ser utilizados (Shannon & Weaver, 1949).

Para cada seqüência de informação existe uma única palavra-código. Em um alfabeto q -ário (q é potência de primo) o codificador produzirá q^k palavras-código. Como um exemplo, no caso em que $q = 4$ e com $k = 2$, existe um total de 4^2 palavras-código, isto é, dezesseis possibilidades de seqüências de informação.

Definição 21 *O conjunto formado pelas q^k palavras-código de comprimento n é denominado **código de bloco** com parâmetros (n, k) .*

Existem exatamente $q^n - q^k$ seqüências que não fazem parte do conjunto de palavras-código, permitindo, desta maneira, a detecção e correção de erros nas seqüências de n -bits. Verifica-se facilmente que se uma seqüência transmitida de n -bits não tem correspondência com alguma palavra-código do código de bloco assume-se, então, que um ou mais bits foram corrompidos pelo efeito danoso do ruído e das distorções que ocorrem no canal.

A decodificação é responsável por encontrar as possíveis modificações ocorridas na seqüência de n -bits, e posteriormente, recuperar os bits de informação corrompidos. Conseqüentemente,

encontra-se a palavra-código associada à seqüência com erros, introduzidos durante a transmissão no canal.

4.3.3 Modulador

O modulador transforma cada símbolo ou grupo de símbolos codificado em um sinal de duração T segundos, adequado para a transmissão. O sinal é enviado através de um canal de transmissão onde poderá ser corrompido pelo ruído. Os canais de transmissão podem ser linhas telefônicas, rádio de alta frequência (HF), cabos de fibra óptica, discos compactos, links de micro-ondas e satélite, e assim sucessivamente. Cada um destes exemplos está sujeito aos mais diversos tipos de distúrbios e ruídos. Por exemplo, na linha telefônica o distúrbio está associado ao ruído térmico, ruído impulsivo, ou da interferência de canal ou de co-canal. Em discos magnéticos (ou discos compactos) o distúrbio e/ou ruído está associado a sujeira e os defeitos encontrados na superfície observada.

4.3.4 Canal de transmissão

O canal de transmissão é o meio pelo qual a informação é transmitida até chegar ao usuário. No canal, a informação pode ser corrompida através da atenuação, distúrbio, distorção, interferência e ruído. A maneira na qual os símbolos binários (0 ou 1) são corrompidos depende das características do canal de transmissão (Sweeney, 1991).

Entre os tipos de canais tem-se:

1. **Canal sem memória:** A probabilidade de erro do símbolo é estatisticamente independente dos erros ocorridos nos símbolos precedentes;
2. **Canal binário simétrico (BSC):** O canal tem duas entradas e duas saídas para os símbolos binários $x = 0$ e $y = 1$. O canal é simétrico pois a probabilidade de receber 0, tendo sido enviado 1, é a mesma probabilidade de receber 1, tendo sido enviado 0, ou seja, $Prob\{x = 0|y = 1\} = Prob\{x = 1|y = 0\} = \frac{1}{p}$;
3. **Ruído aditivo Gaussiano Branco (AWGN):** Este é um canal sem memória atuando em um sinal modulado;
4. **Canal Rajadas:** Existem períodos de alta taxa de erros separados por períodos de baixa taxa de erros, ou taxa de erro zero;
5. **Canal Composto:** Os erros são uma mistura dos erros de rajadas e erros aleatórios.

4.3.5 Decodificador

O método de decodificação depende do método de codificação utilizado. O objetivo de um sistema de codificação é detectar e corrigir os erros que venham a aparecer.

O decodificador recebe seqüências de informação codificadas podendo estas ter erros ou não. Se não contiverem erros, essas seqüências serão declaradas palavras-código. Se a seqüência recebida foi corrompida durante a transmissão, a seqüência não apresentará correspondência única com as palavras-código. Estes mecanismos são utilizados para a detecção de erros.

Quando a taxa de erro excede a capacidade de correção do código, tem-se duas conseqüências:

1. O decodificador consegue localizar o erro, porém não consegue encontrar uma única solução, ou seja, existe mais de uma palavra-código associada ao erro;
2. O decodificador não consegue detectar o erro, pois a adulteração transformou uma palavra-código em outra palavra-código distinta. Portanto, erros que excedem a capacidade de correção do código não serão processados corretamente.

4.3.6 Modulação e constelação de sinais

O modulador em um sistema de comunicação seleciona um sinal de duração T segundos, adequado para a transmissão de símbolos codificados. No caso binário, o modulador deve gerar dois sinais: $s_1(t)$ para o símbolo codificado como "1" e $s_2(t)$ para o símbolo codificado como "0". Para um canal de banda-larga, a melhor escolha do sinal é tal que

$$s_1(t) = \sqrt{\frac{2E_s}{T}} \cos 2\pi f_0 t, \quad 0 \leq t \leq T$$

$$s_2(t) = -\sqrt{\frac{2E_s}{T}} \cos 2\pi f_0 t, \quad 0 \leq t \leq T,$$

onde a frequência da portadora f_0 é um múltiplo de $1/T$, e E_s é a energia de cada sinal.

Esta forma de modulação é denominada **modulação por chaveamento de fase** (BPSK). Assim, a mudança de fase da portadora ocorre entre 0 e π , de acordo com a saída do codificador.

Uma forma comum de ruído presente em qualquer sistema de comunicação é o ruído aditivo Gaussiano branco (AWGN)[2,6,7]. Se o sinal transmitido é $s(t)$, ($s_1(t)$ ou $s_2(t)$), então o sinal recebido $r(t)$ é dado por

$$r(t) = s(t) + n(t),$$

onde $n(t)$ é uma amostra de um processo aleatório Gaussiano com **densidade espectral de potência** (PSD) N_0 .

Outras formas de ruído estão presentes em muitos sistemas. Em um sistema de comunicação que está sujeito à transmissão de múltiplos-caminhos, o sinal recebido é atenuado em certos intervalos de tempo. Esta perda pode ser modelada como um fator de acumulação do ruído multiplicativo sobre o sinal $s(t)$.

Ao sair do codificador binário a seqüência de l -bits é primeiramente segmentada para depois ser transmitida com $M = 2^l$ sinais no canal de transmissão. Para cada byte (oito bits) temos um símbolo e existem M símbolos distintos. Cada símbolo está associado somente a um dos M sinais, contido no conjunto de sinais S , para transmissão. Cada sinal é uma forma de onda de duração T que resulta em uma modulação M -ária. Um exemplo de modulação M -ária é dado pela modulação com chaveamento de fase M -ária (MPSK), na qual o conjunto de sinais consiste de M símbolos de pulsos senoidais. Estes sinais têm a mesma energia E_s , e porém o espaçamento das M fases são diferentes. Tal conjunto de sinal é dado por

$$s_i(t) = \sqrt{\frac{2E_s}{T}} \cos(2\pi f_0 t + \phi_i), \quad 0 \leq t \leq T,$$

onde $\phi_i = 2\pi(i - 1)/M$, $1 \leq i \leq M$. A modulação envelope-constante é também conhecida por BPSK, pois os sinais têm envelope constante. Para $M = 2$, $M = 4$ e $M = 8$ temos BPSK, 4-PSK (também conhecido como QPSK) e 8-PSK, respectivamente. Estas modulações são freqüentemente usadas em comunicação digital. As constelações de sinais são exibidas na Figura 4.3.

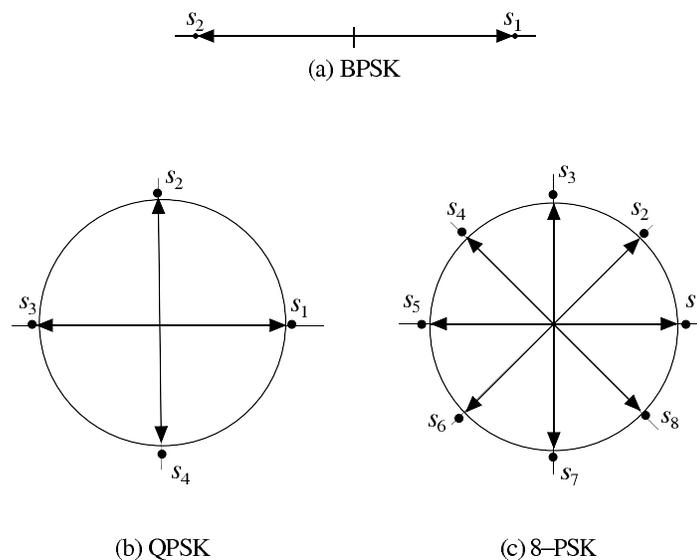


Figura 4.3: Constelações de sinais: (a) BPSK, (b) QPSK e (c) 8-PSK.

As constelações de sinais mencionadas neste capítulo estão relacionadas a grupos finitos

através da relação entre as métricas Euclidiana e de Hamming, ou seja, existe uma relação entre o conjunto de sinais M-PSK e o grupo Z_M .

4.4 Comunicação biológica

Alguns princípios da teoria da informação têm sido utilizados para ampliar e desenvolver conceitos da teoria da codificação e também desenvolver algoritmos de criptografia com a finalidade de que a informação possa ser transmitida com segurança de uma fonte para um receptor, de acordo com a modelagem envolvida no sistema.

O sistema biológico, baseado no Dogma Central da Biologia Molecular, também transmite com segurança suas informações biológicas através de processos bioquímicos, ou seja, replicação, transcrição e tradução, onde o genoma de um organismo é a fonte da informação.

O estudo da capacidade de processamento de sistemas vivos foi despertado novamente em meados de 1980, devido ao aumento de dados genômicos, os quais estimularam o interesse na utilização da teoria da informação nos estudos genômicos (veja(Román-Rolda'n, Bernaola-Galván & Oliver 1996)).

Medidas de informação, tal como entropia, são usadas no reconhecimento de padrões do DNA, na classificação de seqüências genéticas e em vários outros estudos computacionais de seqüências genéticas (veja(Schneider & Mastronard 1996, Schneider 1997)).

Aplicando técnicas da teoria da informação e codificação temos como objetivo estudar e entender o mecanismo de processamento da informação nos sistemas genéticos. Enquanto a teoria da informação analisa processos e seqüências genéticas a teoria da codificação fornece técnicas e métodos para a construção de seqüências genéticas bem como modelar sistemas genéticos.

A aplicação da teoria da codificação para dados genéticos começou no final de 1950 (veja (Hayes 1998)), com o mapeamento do código genético (um códon especifica um amino ácido). Assim, métodos da teoria da codificação vêm sendo utilizados no DNA, como por exemplo, determinar estruturas genéticas, projetar polinucleotídeos e os mais diversos cálculos envolvendo DNA (veja (Loewenstern & Yanilos 1997, Sengupta & Tompa 2002, Kari, Kari & Landweber 1999, May 2002)).

Loewenstern & Yianilos (veja (Loewenstern & Yanilos 1997)) aplicaram métodos de codificação de fonte para seqüências genéticas com o propósito de identificar tais seqüências.

Elebeoba E. May (veja (May 2004)) utilizou a teoria da codificação para determinar quando regiões específicas do DNA, de organismos procariontes, são potencialmente geradoras de proteínas.

Outro problema complexo é propor um algoritmo, baseado para a codificação de proteínas.

Todos esses avanços estão se desenvolvendo em grande parte devido à enorme quantidade de dados genéticos disponíveis hoje e também ao acesso de genomas dos mais variados seres vivos.

4.5 Codificação e o Dogma Central da Genética

Não é óbvia a relação entre os processos de síntese de proteínas e o de codificação e correção de erros. A Figura 4.4 ilustra o Dogma Central da Biologia. A premissa central da genética afirma que os genes são repassados de geração em geração através de seqüências de aminoácidos. Seqüência de três bases de nucleotídeos, denominadas códon, especificam um aminoácido. Existem sessenta e quatro possibilidades de códon e vinte possíveis aminoácidos. Portanto, diferentes códon especificam um mesmo aminoácido. Esta relação aminoácido-códon é conhecida como código genético.

Existem três estágios que transformam a informação contida nos genes em proteínas:

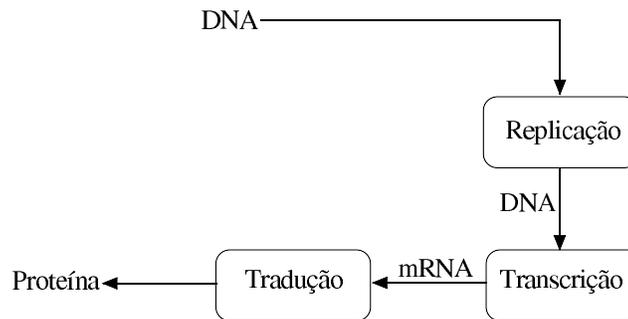


Figura 4.4: Dogma central da biologia.

Estágio 1: Replicação - Processo pelo qual as duas fitas de DNA se separam, onde cada fita serve de molde para sintetizar uma nova fita de DNA complementar à si própria, isto é, uma cópia idêntica da informação genética do DNA original;

Estágio 2: Transcrição - Processo pelo qual um RNA é fabricado a partir de uma das fitas molde do DNA, assim, o RNA contém uma seqüência (fita) complementar à fita molde. Vários tipos de RNA diferentes são gerados pela transcrição. As três principais classes envolvidas na síntese de proteína são:

- RNA mensageiro (mRNA);
- RNA transportador (tRNA);
- RNA ribossômico (rRNA).

Estágio 3: Tradução - Processo pelo qual a mensagem genética gravada no mRNA é traduzida em cadeias polipeptídicas ou proteínas. A cadeia polipeptídica é uma seqüência de aminoácidos. O ribossomo traduz a informação contida no mRNA em proteínas.

O processo de tradução da informação consiste de três fases:

- (i) - **Tradução Iniciação:** A etapa da iniciação é regulada por proteínas adicionais conhecidas como **fatores de iniciação** (IF). O IF-2 age no reconhecimento e ligação do mRNA pela subunidade ribossomal, o IF-3 liga-se a um tRNA iniciador e controla sua entrada no ribossomo, e o IF-1 participa do processo de ligação entre as duas subunidades e controla a capacidade que a subunidade tem de se ligar ao mRNA. Os fatores de iniciação se dissociam do ribossomo antes da fase de alongamento da tradução;
- (ii) - **Tradução Alongamento ou Extensão:** É durante este processo que o ribossomo mais se assemelha a uma fábrica de produção de proteínas. Nesta fase temos somente dois fatores protéicos, chamados de fatores de alongamento, denotados como EF-Tu e EF-G. Esta fase começa quando o aminoacil-tRNA_t liga-se ao sítio A (região onde ocorre a ligação) do ribossomo, se unindo ao mRNA. Esta reação é mediada pelo complexo ternário (composto de tRNA, aminoácido e EF-Tu) utilizando a energia fornecida pelo trifosfato de guanosina (GTP), para maiores detalhes veja (Alberts et al. 2004, pp. 344–350). O outro fator de alongamento EF-G, é responsável pela translocação, isto é, o deslocamento do ribossomo pelo o mRNA;
- (iii) - **Tradução Terminação:** A etapa de terminação determina a conclusão de síntese de proteínas, quando o ribossomo atinge o códon de terminação (UAA, UGA ou UAG). Isto faz com que o sítio A seja ocupado por um fator de terminação RF-1, impedindo a ligação de um outro tRNA. Depois disto ocorre o desprendimento do polipeptídeo, que depende do EF-3, outro fator de terminação.

Observação 6 *Os fatores protéicos considerados estão relacionados com as células procariontes.*

Vários pesquisadores estão desenvolvendo modelos de comunicação para processos genéticos (Román-Rolda'n et al. 1996, May 2002), etc.

Baseado nos trabalhos de May e no Dogma Central da Biologia construímos uma representação do sistema de comunicação tendo como base o processo de informação genética (Veja Figura 4.5). O tRNA representado na figura não é o mesmo durante o processo de decodificação. Para maiores detalhes sobre o modelo de codificação genética consultar referências (May 2002, May 2004).

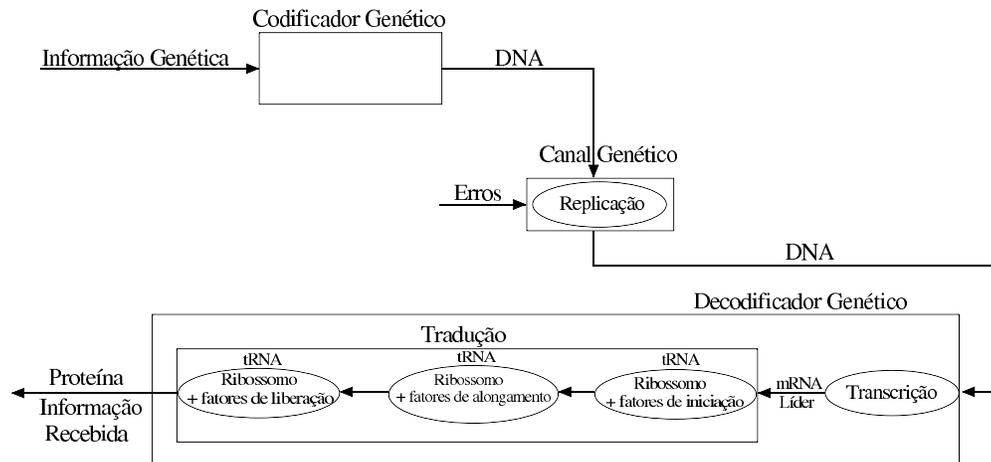


Figura 4.5: Dogma central da biologia como um sistema de codificação.

Assim como no processo de comunicação, o erro também está presente no processo de formação genética, pois qualquer processo que se repete inúmeras vezes está sujeito a algum tipo de perturbação do meio. Embora o DNA se duplique com precisão e tenha um enorme arquivo de informação genética para replicação (para as mais diversas necessidades biológicas, entre elas, a fabricação de proteína) é de se esperar que os erros ocorram. Mesmo dotado de um sistema de segurança que repara os erros que acontecem na cópia de informação genética, os erros podem surgir provocando **mutação genética**.

A mutação nada mais é que uma modificação na informação genética contida no DNA, existem dois tipos de mutação genética:

1. **Mutação Pontual ou Gênica:** São alterações envolvendo um par de bases ou um pequeno número de pares de bases localmente no gene, as mutações pontuais podem ser: substituições de bases, inserções de bases, ou deleções de bases. As mutações pontuais são muitas vezes geradas por erros que ocorrem, especialmente, na hora de transcrever a informação genética do DNA para o RNA sendo menos comum na duplicação do DNA;
2. **Mutação Cromossômica:** São alterações no número de cromossomos de uma célula ou podem ser alterações estruturais, rearranjos do DNA. As alterações estruturais derivam de duplicação, deleção e translocação de cromossomos. As alterações podem ser provocadas durante algumas etapas da divisão celular, durante a segregação de cromossomos, ou durante o momento do crossing over (troca recíproca entre segmentos correspondentes de um mesmo par cromossômico). Em alguns casos as mutações cromossômicas resultam da quebra cromossômica. Se a quebra ocorre dentro de um gene, o resultado é uma perturbação funcional do gene (mutação gênica). Assim, as mutações cromossômicas são acompanhadas de mutações gênicas pois, neste caso, o ponto de quebra foi dentro do gene.

Observação 7 *Não se deve esquecer, no entanto, que a mutação é essencial para a variabilidade genética e para a evolução das espécies.*

A seqüência não replicada do DNA é a saída do codificador genético, com corretor de erros, que adiciona redundância inerente a informação genética com ruído. O ruído na fonte pode ser pensado como a transferência de mutações para os seus descendentes. O canal genético é o processo de replicação do DNA durante o qual os erros são introduzidos dentro da seqüência de nucleotídeos. Utilizando a idéia proposta por May (2002) decodificação e correção de erros ocorrem em três fases, representadas por transcrição, tradução inicial e alongamento mais terminação. Fazendo um paralelo entre o sistema de comunicação “genético” e o sistema de comunicação típico (representado na Figura 4.1) temos o codificador genético, o canal genético e o decodificador genético.

4.5.1 Codificador Genético

Ainda hoje é desconhecido qual mecanismo biológico se comporta como um codificador biológico. Porém, é bastante claro que a redundância está presente na transmissão de mensagens. Por exemplo, o número de bases do DNA no genoma da *E. coli* excede o número de bases necessárias para codificar todas as proteínas produzidas por esta bactéria.

Existem 64 códons dos quais apenas três têm funções específicas. Os 61 códons restantes codificam 20 aminoácidos, os quais são usados para formar proteínas. O número de códons excede o número de aminoácidos, logo, existe informação redundante no código genético.

A redundância também é encontrada nas seqüências de DNA e RNA. A existência de seqüências redundantes como por exemplo, *tandem repeats*, *Shine-Dalgarno*, *TATA box*, de adenosinas entre outras, nos leva a supor que o sistema de comunicação celular utiliza a redundância para proteger a mensagem genética de erros. O tipo e a taxa de codificação usadas para codificar as seqüências de DNA são, até o momento, desconhecidas.

4.5.2 Canal Genético

Assumindo que a seqüência de DNA é a saída do codificador genético, o canal genético é a seqüência de eventos ou a média de eventos que a seqüência de DNA introduz e transmite no decodificador genético. Modelamos o processo de replicação como um canal genético pois, como sabemos, nos canais não ocorrem mudanças no alfabeto da mensagem, porém, são introduzidos erros. Análogo ao canal de comunicação, o canal genético introduz erros e ruídos na mensagem transmitida. De qualquer forma, também é possível que a mensagem de entrada (DNA) no canal genético já venha com alguma alteração, devido a algum outro evento biológico.

Durante a replicação vários tipos de erros podem ocorrer: substituições de bases e inserções ou deleções de bases no DNA. Estes erros (mutações) no código genético podem atenuar, corromper e distorcer o sinal genético, comprometendo, na maioria das vezes, a sobrevivência do organismo. Por outro lado, algumas mutações podem ser inócuas ou, eventualmente, benéficas para o organismo.

4.5.3 Decodificador Genético

Similar aos decodificadores, a transcrição e a tradução são duas fases da decodificação. O trabalho de May se baseia em um sistema de comunicação e faz um paralelo entre os processos de transcrição e tradução e o processo da decodificação. A tradução é o processo pelo qual a informação genética armazenada no RNA é decodificada em seqüências de aminoácidos que formarão a proteína. O ribossomo faz parte do mecanismo da decodificação celular. Uma vez, iniciada a tradução, o ribossomo (em conjunto com o tRNA e outros fatores protéicos) decodificam a mensagem transmitida associando a cada códon da seqüência com o anti-códon carregado pelo tRNA. Durante a fase de alongamento, as palavras-código válidas da seqüência serão traduzidas em proteínas. Analisando as três fases de tradução da proteína, observamos que o ribossomo se comporta como um decodificador, assim como no sistema de comunicação, assumindo papel fundamental na decodificação da informação. Além disso, participa desde o recebimento da seqüência de comprimento fixo n (palavra-código) até sua finalização, isto é, a decodificação da palavra-código em proteínas.

O processo de decodificação biológica pode ser dividido em duas fases: transcrição e tradução, onde a última se divide em três subfases interrelacionadas: iniciação, alongamento e terminação.

Fase 1 - Transcrição: A mensagem é transcrita ou codificada do DNA para o RNA. Assim, o mRNA fica encarregado de transportar a cópia da informação, contida no DNA, para os ribossomos no citoplasma. Troca-se a base U por T;

Fase 2 - Tradução: A mensagem recebida pelo ribossomo através do mRNA é decodificada em proteína. Durante esta fase ainda não fica claro qual o procedimento de decodificação que a subunidade ribossomal emprega.

Observação 8 *Observamos que na tradução o ribossomo juntamente com os fatores protéicos formam um mecanismo complexo de decodificação da informação contida no mRNA. Desta maneira, podemos utilizar a teoria de códigos corretores de erros para classificar seqüências de bases em palavras-código.*

O alfabeto a ser utilizado deve ter uma estrutura algébrica associada. Utilizando a base de paridade e a seqüência de informação contida no RNA, podemos fazer a seguinte associação:

- Adenina (A) = 0;
- Citosina (C) = 1;
- Guanina (G) = 2;
- Timina (T)/Uracila (U) = 3.

As operações utilizadas são a adição e multiplicação módulo-4. As bases no RNA foram definidas de tal modo que a soma das bases, que formam os pares de hidrogênio, conservem a complementariedade entre as mesmas. Em sua tese de doutorado May (May 2002) utilizou a adição e multiplicação módulo-5, pois considerou o pareamento *wobble*, onde uma quinta base nitrogenada foi considerada, a saber, a inosina (I) (associação $I = 0$).

O alfabeto a ser adotado depende do fenômeno biológico e suas restrições impostas ao sistema de comunicação biológico. Desta forma, garante-se maior confiabilidade nos resultados apresentados e uma interpretação mais segura dos dados.

4.6 Código genético e modulação

A proposta deste trabalho é tratar o código genético como um modulador de sinal. O modulador genético se localiza entre o codificador e o canal genético, conforme mostra a Figura 4.6. Construímos as constelações de sinais associadas aos grupos de permutações S_3 e S_4 (Capítulo 5).

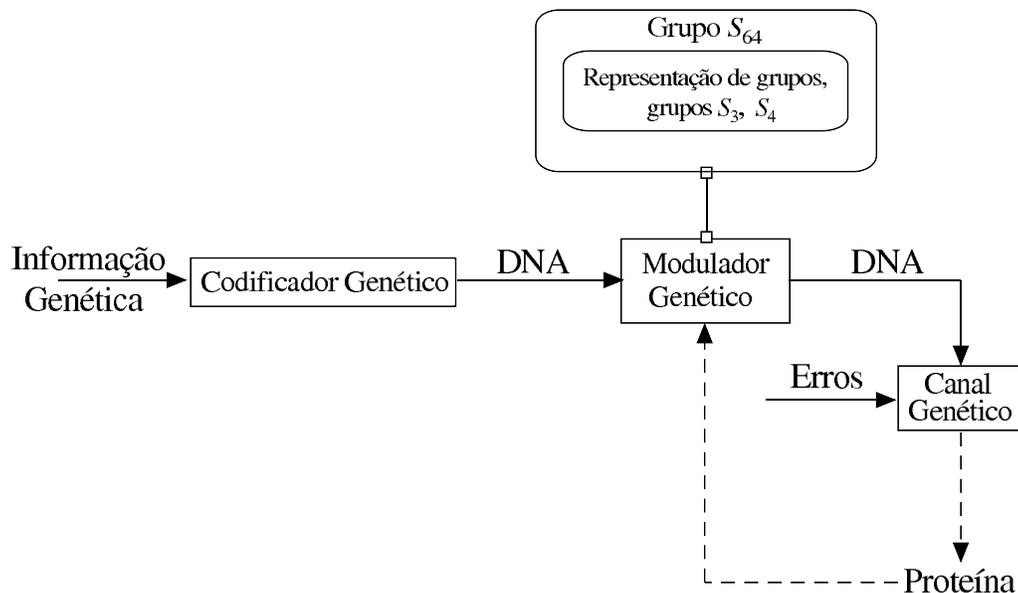


Figura 4.6: Modulador genético.

O Teorema de Cayley garante a existência de um subgrupo H em S_{64} contendo 64 elementos. Portanto, através deste subgrupo, acreditamos que seja possível a construção de uma constelação de sinais associada ao código genético. Para isso, teríamos que associar os 64 elementos do código genético a uma constelação de sinais. Isso explica a necessidade da utilização do grupo S_{64} na construção da constelação de sinais. Biologicamente, falando teríamos uma constelação aminoácidos associada ao código genético.

Sendo assim, conjecturamos a existência de um subgrupo H em S_{64} que descreva as características e propriedades do código genético.

Sabemos que a resolução de tal conjectura não é simples devido a grande dificuldade computacional em gerar e verificar quais dos subgrupos em S_{64} possuem uma estrutura matemática associada ao código genético. Outro fator agravante no processo de busca é que o subgrupo encontrado em S_{64} não pode ser cíclico. Importante salientar que o número de elementos de S_{64} é $64!$.

Outra questão muito trabalhosa e complexa, já mencionada no Capítulo 1, é encontrar uma ação τ de G em X que gere 20 partições (veja Teorema de Burnside, Capítulo 3) em X onde cada partição em X seja equivalente a cada uma das 20 partições do código genético. Importante salientar que cada partição no código genético está associada a um dos 20 tipos de aminoácidos.

Empiricamente, obter uma constelação de “aminoácido” (sinais) associada ao código genético seria muito interessante, pois ficaria mais fácil descrever o processo de codificação realizado pelo DNA no sistema de comunicação biológico e também conferir e analisar a resposta final (proteína), decodificada pelo ribossomo e pelos demais fatores protéicos.

Outro fato importante é que através do modulador genético ficaria mais fácil entender e analisar as perturbações ocorridas antes e depois do canal biológico, e também a verificação da aplicabilidade dos métodos de correção de erros.

Constelações de Sinais Associadas aos Grupos S_3 e S_4

Apresentamos os conceitos de espaço de sinais e de constelações de sinais utilizados neste trabalho.

Um dos objetivos deste capítulo é a representação de constelações de sinais através dos grupos de permutações S_3 e S_4 .

Conjecturamos que tais representações poderão ser utilizadas como parte dos subgrupos em S_{64} . Se tal conjectura for verdadeira, obtém-se um modelo matemático com imediata aplicação ao código genético, descrevendo-o de forma matemática. É claro que tal caracterização seria um avanço relevante, não só para a Biologia Molecular, mas também em todo o processo vital que nos cerca.

Utilizando os conceitos de grupos, classes conjugadas e isomorfismos construiremos constelações de sinais associados aos grupos S_3 e S_4 .

Espaços de sinais e constelações de sinais

Definição 22 *Um **espaço de sinais** é um conjunto discreto de pontos em um espaço métrico (\mathbb{E}, d) em que seja possível realizar uma identificação dos pontos de (\mathbb{E}, d) por sinais.*

Definição 23 *Uma **constelação de sinais** é um subconjunto finito de sinais em um espaço de sinais.*

5.1 Grupo de Permutação S_3

Seja S_3 dado por

$$S_3 = \left\{ \begin{pmatrix} 123 \\ 123 \end{pmatrix}, \begin{pmatrix} 123 \\ 231 \end{pmatrix}, \begin{pmatrix} 123 \\ 312 \end{pmatrix}, \begin{pmatrix} 123 \\ 132 \end{pmatrix}, \begin{pmatrix} 123 \\ 321 \end{pmatrix}, \begin{pmatrix} 123 \\ 213 \end{pmatrix} \right\}.$$

que também pode ser representado na forma de ciclos, isto é,

$$S_3 = \{e, (123), (132), (23), (13), (12)\}.$$

O grupo S_3 é isomorfo ao grupo T das simetrias do triângulo equilátero:

$$T = \{\rho_0, \rho_1, \rho_2, \eta_1, \eta_2, \eta_3\},$$

onde ρ_i é a rotação de $\frac{i2\pi}{3}$ no sentido anti-horário através da reta (eixo) vertical que passa pelo baricentro B do triângulo e η_i é a reflexão através da reta que liga o vértice do triângulo ao ponto médio da aresta oposta. Na Figura 5.1, mostramos o triângulo com vértices 1, 2, 3 e as retas que representam o eixo de rotação.

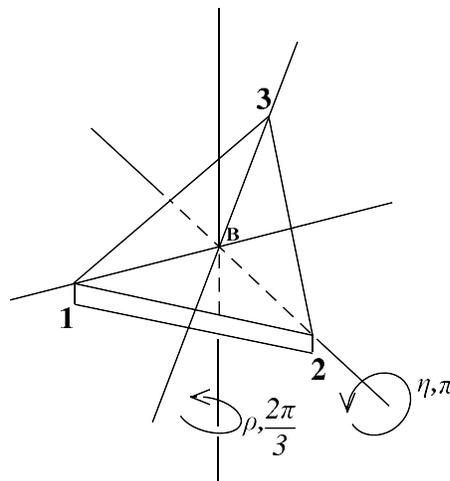


Figura 5.1: Simetrias do triângulo equilátero.

5.1.1 Subgrupos de S_3

Pelo Teorema de Cayley, dado um grupo finito \bar{G} de ordem k , existe um subgrupo G , de ordem k , em S_n tal que G é isomorfo a \bar{G} . Como exemplo, se $\bar{G} = Z_3$, encontraremos um subgrupo $G \subset S_3$ de ordem 3 tal que G é isomorfo a Z_3 . Tendo como base esses conceitos, iremos gerar todos os subgrupos de S_3 e classificá-los de acordo com suas respectivas classes conjugadas.

- Subgrupos Triviais: $\{e\}, S_3$;
- Subgrupos das Reflexões: $\{e, (23)\}, \{e, (13)\}, \{e, (12)\}$;
- Subgrupos de Rotações: $\{e, (123), (132)\}$.

Classe 1

Esta classe consiste de somente um elemento, ou seja, o grupo trivial $\{e\}$. Sua representação de sinal associada é um ponto na reta real.

Classe 2

Existem três grupos nesta classe. Cada grupo consiste do elemento identidade e de uma transposição:

- Grupo 1: $\{e\}, (23)$;
- Grupo 2: $\{e\}, (31)$;
- Grupo 3: $\{e\}, (12)$.

Sabemos que os Grupos 1, 2 e 3 são isomorfos pois pertencem à mesma classe conjugada. Logo, basta exibir um isomorfismo considerando apenas um destes.

Considere o grupo

$$G = \{e, (23)\} = \{\rho_0, \eta_1\},$$

onde tem-se o seguinte isomorfismo

$$\begin{cases} \mathbb{Z}_2 & \longrightarrow & G \\ \bar{0} & \mapsto & \rho_0 = e \\ \bar{1} & \mapsto & \eta_1 = (23) \end{cases}$$

Sua respectiva representação de sinais está mostrada na Figura 5.2.

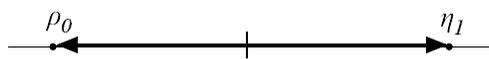


Figura 5.2: Representação do grupo \mathbb{Z}_2 na reta real \mathbb{R} (bimodal).

Classe 3

Existe somente um grupo nesta classe. Este grupo consiste do elemento identidade e de dois 3-ciclo.

- Grupo 1: $\{e\}, (123), (132)$.

Seja G o grupo dado por

$$G = \{e, (123), (132)\} = \{\rho_0, \rho_1, \rho_2\} = T,$$

onde tem-se o seguinte isomorfismo

$$\begin{cases} \mathbb{Z}_3 \longrightarrow G \\ \bar{0} \longmapsto \rho_0 = e \\ \bar{1} \longmapsto \rho_1 = (123) \\ \bar{2} \longmapsto \rho_2 = (132) \end{cases}$$

Considere agora o grupo das matrizes M , ou seja, o grupo das representações matriciais associado ao grupo G :

$$M = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\}$$

Existe um isomorfismo entre M e G , e assim, podemos identificar os elementos de cada grupo da seguinte forma:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (123), \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (132).$$

A Figura 5.3 mostra a representação de sinais associada ao subgrupo G em \mathbb{R}^2 , proporcionando uma associação geométrica ao grupo G .

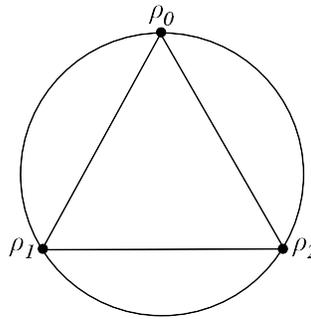


Figura 5.3: Representação de sinais do grupo \mathbb{Z}_3 no \mathbb{R}^2 .

A representação de sinais do grupo \mathbb{Z}_3 no \mathbb{R}^3 é ilustrada na Figura 5.4.

Classe 4

Esta classe possui somente um elemento, a saber, o próprio grupo S_3 . A representação matricial M associada (isomorfa) ao grupo S_3 é

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (123), \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (132),$$

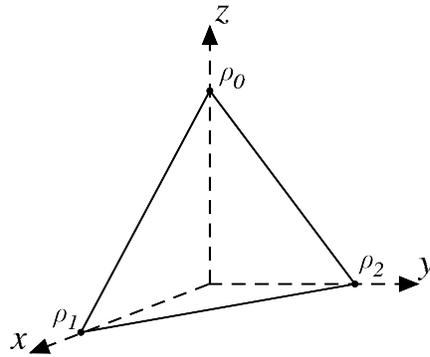


Figura 5.4: Representação de sinais do grupo \mathbb{Z}_3 no \mathbb{R}^3 .

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = (23), \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = (13), \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (12).$$

A Figura 5.5 ilustra a representação de sinais associada ao subgrupo S_3 em \mathbb{R}^2 , proporcionando uma associação geométrica a este grupo.

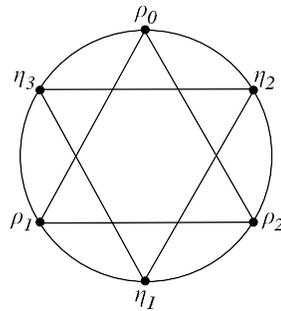


Figura 5.5: Representação de sinais do grupo S_3 no \mathbb{R}^2 .

A representação de sinais do grupo S_3 em \mathbb{R}^3 é ilustrada na Figura 5.6.

Observação 9 *Estamos interessados em subgrupos próprios contendo mais de dois elementos; sendo assim, desconsideraremos as classes cujos elementos (grupos) tenham ordem menor ou igual a dois.*

5.2 Grupo de Permutação S_4

O grupo S_4 é o grupo formado pelas permutações de 4 letras e possui um total de 24 permutações distintas. Cada elemento é denotado por

$$\begin{pmatrix} 1234 \\ abcd \end{pmatrix}.$$

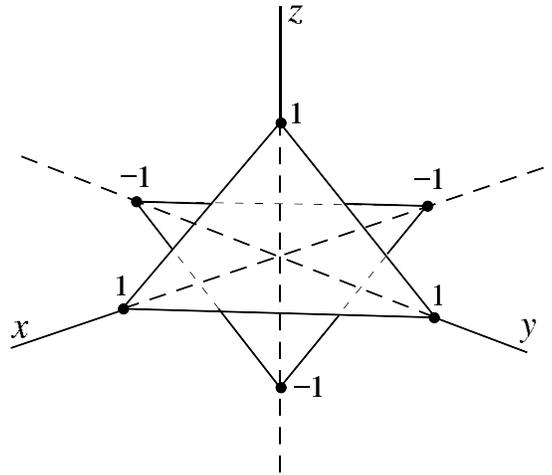


Figura 5.6: Representação de sinais do grupo S_3 no \mathbb{R}^3 .

Esta notação representa a aplicação $f(1) = a, f(2) = b, f(3) = c, f(4) = d$, tal que, $a, b, c, d \in \{1, 2, 3, 4\}$. Os 24 elementos do grupo S_4 são representados por

$$G = \left\{ \begin{array}{l} \left(\begin{array}{c} 1234 \\ 1234 \end{array} \right), \left(\begin{array}{c} 1234 \\ 2341 \end{array} \right), \left(\begin{array}{c} 1234 \\ 2413 \end{array} \right), \left(\begin{array}{c} 1234 \\ 4312 \end{array} \right), \\ \left(\begin{array}{c} 1234 \\ 4123 \end{array} \right), \left(\begin{array}{c} 1234 \\ 3142 \end{array} \right), \left(\begin{array}{c} 1234 \\ 3421 \end{array} \right), \left(\begin{array}{c} 1234 \\ 3412 \end{array} \right), \\ \left(\begin{array}{c} 1234 \\ 4321 \end{array} \right), \left(\begin{array}{c} 1234 \\ 2143 \end{array} \right), \left(\begin{array}{c} 1234 \\ 1243 \end{array} \right), \left(\begin{array}{c} 1234 \\ 1324 \end{array} \right), \\ \left(\begin{array}{c} 1234 \\ 2134 \end{array} \right), \left(\begin{array}{c} 1234 \\ 4231 \end{array} \right), \left(\begin{array}{c} 1234 \\ 3214 \end{array} \right), \left(\begin{array}{c} 1234 \\ 1432 \end{array} \right), \\ \left(\begin{array}{c} 1234 \\ 2314 \end{array} \right), \left(\begin{array}{c} 1234 \\ 3124 \end{array} \right), \left(\begin{array}{c} 1234 \\ 4213 \end{array} \right), \left(\begin{array}{c} 1234 \\ 3241 \end{array} \right), \\ \left(\begin{array}{c} 1234 \\ 1423 \end{array} \right), \left(\begin{array}{c} 1234 \\ 1342 \end{array} \right), \left(\begin{array}{c} 1234 \\ 2431 \end{array} \right), \left(\begin{array}{c} 1234 \\ 4132 \end{array} \right) \end{array} \right\}.$$

Assim, como no caso anterior, o grupo S_4 é isomorfo a um grupo de simetrias, a saber, ao grupo G das simetrias das rotações do cubo, sendo representado por

$$G = \left\{ \begin{array}{l} e, \rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \rho_i^2, \rho_j^2, \rho_k^2, \mu_1, \mu_2, \\ \mu_3, \mu_4, \mu_5, \mu_6, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8 \end{array} \right\} \text{ onde } i, j, k = 1, \dots, 6.$$

As rotações do cubo são ilustradas na Figura 5.7. A figura é composta por três retas (eixos) que geram 24 rotações distintas onde cada rotação representa um elemento do grupo S_4 :

- Grupo 1: $\{e\}, (234), (243)$;
- Grupo 2: $\{e\}, (134), (143)$;
- Grupo 3: $\{e\}, (124), (142)$;
- Grupo 4: $\{e\}, (123), (132)$.

Assim como nos casos anteriores, basta exibirmos um isomorfismo entre um elemento (grupo) da classe conjugada, que denotaremos por G , e o grupo \mathbb{Z}_3 :

$$\begin{cases} \mathbb{Z}_3 \longrightarrow G \\ \bar{0} \longmapsto e \\ \bar{1} \longmapsto (234) \\ \bar{2} \longmapsto (243) \end{cases}$$

A representação matricial associada é dada por

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (234), \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (243).$$

A representação de sinais associada aos subgrupos de S_4 , que pertencem a Classe 1 é a mesma estabelecida para a Classe 3 do grupo S_3 .

Classe 2

Existem três grupos nesta classe. Cada grupo consiste do elemento identidade, duas permutações de 4-ciclo e o produto de duas transposições.

- Grupo 1: $\{e\}, (1234), (13)(24), (1432)$;
- Grupo 2: $\{e\}, (1243), (14)(23), (1342)$;
- Grupo 3: $\{e\}, (1324), (12)(34), (1423)$.

Exibiremos um isomorfismo entre um elemento G da Classe 2 e o grupo \mathbb{Z}_4 . Sem perda de generalidade escolhemos o Grupo 1:

$$\begin{cases} \mathbb{Z}_4 \longrightarrow G \\ \bar{0} \longmapsto e \\ \bar{1} \longmapsto (1234) \\ \bar{2} \longmapsto (13)(24) \\ \bar{3} \longmapsto (1432) \end{cases}$$

A representação matricial associada é dada por

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = e, \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} = (1234),$$

$$\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = (13)(24), \quad \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (1432).$$

A representação de sinais associada aos subgrupos de S_4 , que pertencem a Classe 2 é ilustrada na Figura 5.8.

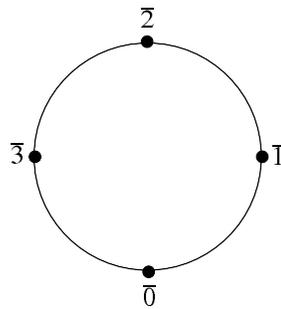


Figura 5.8: Representação de sinais de \mathbb{Z}_4 no \mathbb{R}^2

Classe 3

Existem 3 grupos nesta classe. Cada grupo consiste do elemento identidade, duas transposições e o produto dessas transposições.

- Grupo 1: $\{e\}, (34), (12), (12)(34)$;
- Grupo 2: $\{e\}, (24), (13), (13)(24)$;
- Grupo 3: $\{e\}, (23), (14), (14)(23)$.

Exibimos, a seguir, um isomorfismo entre um elemento G da Classe 3 e o grupo $\mathbb{Z}_2 \times \mathbb{Z}_2$. Assumimos, sem perda de generalidade, que o elemento é o Grupo 1:

$$\begin{cases} \mathbb{Z}_2 \times \mathbb{Z}_2 \longrightarrow G \\ (\bar{0}, \bar{0}) \longmapsto e \\ (\bar{1}, \bar{0}) \longmapsto (34) \\ (\bar{0}, \bar{1}) \longmapsto (12) \\ (\bar{1}, \bar{1}) \longmapsto (12)(34) \end{cases}$$

As respectivas representações matriciais são dadas por

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = e, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (34), \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (12),$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (12)(34).$$

Uma outra representação matricial é dada por

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = e, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = (34), \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (12),$$

$$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} = (12)(34).$$

Em termos computacionais esta forma tem complexidade inferior à anterior. A representação de sinais associada aos subgrupos de S_4 , que pertencem a Classe 3 são exibidas na Figura 5.9.

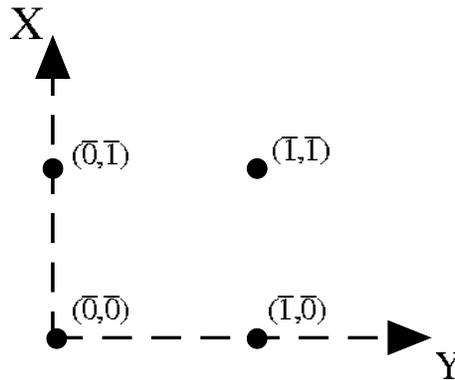


Figura 5.9: Representação de sinais do grupo $Z_2 \times Z_2$.

Classe 4

Esta classe é composta somente por um grupo. Este grupo é conhecido como o grupo de Klein.

- Grupo 1: $\{e\}, (12)(34), (13)(24), (14)(23)$.

O isomorfismo é dado por

$$\left\{ \begin{array}{l} Z_2 \times Z_2 \longrightarrow G \\ (\bar{0}, \bar{0}) \longmapsto e \\ (\bar{1}, \bar{0}) \longmapsto (12)(34) \\ (\bar{0}, \bar{1}) \longmapsto (13)(24) \\ (\bar{1}, \bar{1}) \longmapsto (14)(23) \end{array} \right.$$

O grupo de Klein é isomorfo a qualquer elemento da Classe 3. Sendo assim, a representação de sinais associada é a mesma dos elementos da Classe 3.

Classe 5

Os grupos da Classe 5 são descritos a seguir:

- Grupo 1: $\{e\}, (34), (23), (243), (234), (24)$;
- Grupo 2: $\{e\}, (34), (13), (143), (134), (14)$;
- Grupo 3: $\{e\}, (24), (12), (142), (124), (14)$;
- Grupo 4: $\{e\}, (23), (12), (132), (123), (13)$.

Exibimos um isomorfismo entre um elemento G da Classe 5 e o grupo S_3 , onde consideramos o Grupo 1:

$$\left\{ \begin{array}{l} S_3 \longrightarrow G \\ e \longmapsto e \\ (123) \longmapsto (243) \\ (132) \longmapsto (234) \\ (23) \longmapsto (34) \\ (13) \longmapsto (23) \\ (12) \longmapsto (24) \end{array} \right.$$

Uma representação matricial associada é

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = e, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (243), \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} = (234),$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (34), \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (23), \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = (24).$$

Uma outra representação matricial é dada por

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= e, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (243), \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = (234), \\ - \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= (34), - \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = (23), - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = (24). \end{aligned}$$

Uma possível representação de sinais associada aos subgrupos de S_4 que pertencem a Classe 5 é a mesma dada ao grupo S_3 em \mathbb{R}^2 e em \mathbb{R}^3 .

Classe 6

Os grupos da Classe 6 são descritos a seguir:

- Grupo 1: $\{e\}$, (34), (13)(24), (1423), (1324), (12), (14)(23), (12)(34);
- Grupo 2: $\{e\}$, (24), (12)(34), (1432), (1234), (13), (14)(23), (13)(24);
- Grupo 3: $\{e\}$, (23), (12)(34), (1342), (1243), (14), (13)(24), (14)(23).

Exibimos um isomorfismo entre um elemento G da Classe 6 e o grupo D_4 :

$$\left\{ \begin{array}{l} D_4 \longrightarrow G \\ \rho_0 \longmapsto e \\ \rho_1 \longmapsto (1423) \\ \rho_2 \longmapsto (12)(34) \\ \rho_3 \longmapsto (1324) \\ \mu_1 \longmapsto (34) \\ \mu_2 \longmapsto (12) \\ \delta_1 \longmapsto (14)(23) \\ \delta_2 \longmapsto (13)(24) \end{array} \right.$$

Uma representação matricial é dada por

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} &= e, & \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} &= (1423), \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} &= (12)(34), & \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} &= (1324), \end{aligned}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (34), \quad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (12),$$

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = (14)(23), \quad \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = (13)(24).$$

Uma outra representação matricial é dada por

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = e, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} = (1423),$$

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = (12)(34), \quad \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = (1324),$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} = (34), \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = (12),$$

$$\begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix} = (14)(23), \quad \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = (13)(24),$$

ou ainda,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = e, \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = (1423), \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = (12)(34),$$

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = (1324), \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (34), \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = (12),$$

$$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} = (14)(23), \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = (13)(24).$$

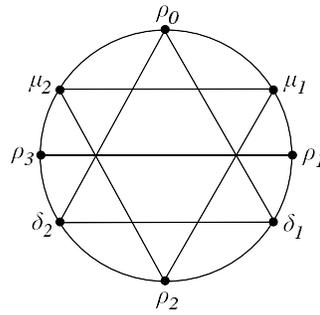


Figura 5.10: Representação de sinais do grupo diedral D_4 no \mathbb{R}^2 .

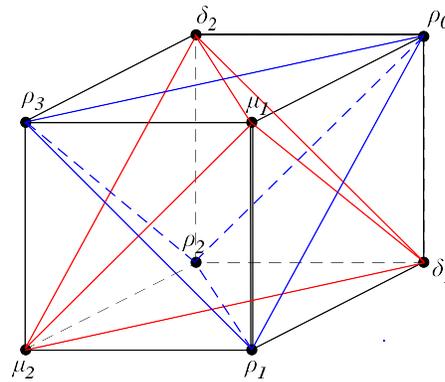


Figura 5.11: Representação de sinais do grupo diedral D_4 no \mathbb{R}^3 .

A representação de sinais associada aos subgrupos de S_4 em \mathbb{R}^2 , que pertencem a Classe 6, é exibida na Figura 5.10.

A representação de sinais no espaço Euclidiano \mathbb{R}^3 consiste de dois tetraedros opostos pelos vértices, inscritos em um cubo, mostrado na Figura 5.11.

No espaço quadridimensional temos o hipercubo projetado no \mathbb{R}^3 , onde no cubo maior se encontra inscrito um tetraedro (em vermelho) e no cubo menor temos inscrito outro tetraedro oposto ao primeiro (em azul), conforme mostra a Figura 5.12.

Finalmente, descrevemos a última classe conjugada.

Classe 7

O grupo da Classe 7 é descrito por

- Grupo 1: $\{e\}, (234), (243), (12)(34), (132), (142), (124), (123), (143), (134), (13)(24), (14)(23)$.

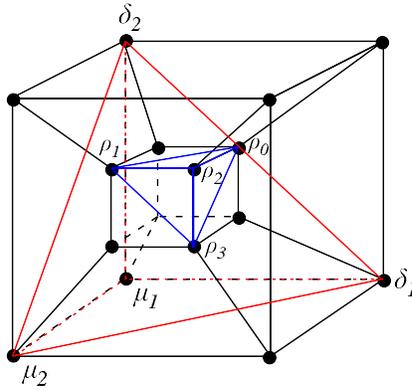


Figura 5.12: Representação de sinais do grupo diedral D_4 no \mathbb{R}^4 (hipercubo).

Este grupo é isomorfo ao grupo $A_4 = \langle a, b \mid a^3 = b^3 = (ab)^2 = 1 \rangle$:

$$\left\{ \begin{array}{ll} A_4 & \longrightarrow G \\ 1 & \longmapsto e \\ a & \longmapsto (142) \\ a^2 & \longmapsto (124) \\ b & \longmapsto (132) \\ b^2 & \longmapsto (123) \\ ab & \longmapsto (13)(24) \\ ab^2 & \longmapsto (234) \\ a^2b & \longmapsto (134) \\ ba & \longmapsto (14)(23) \\ b^2a & \longmapsto (143) \\ ba^2 & \longmapsto (243) \\ ab^2a & \longmapsto (12)(34) \end{array} \right.$$

A representação matricial associada é dada por

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = e, \quad \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = (142),$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = (124), \quad \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (132),$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = (123), \quad \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = (13)(24),$$

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} &= (234), & \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} &= (134), \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} &= (14)(23), & \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} &= (143), \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} &= (243), & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} &= (12)(34). \end{aligned}$$

Uma outra forma de representação é dada por

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= e, & \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} &= (142), \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} &= (124), & \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix} &= (132), \\ \begin{pmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} &= (123), & \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= (13)(24), \\ \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} &= (234), & \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} &= (134), \\ \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} &= (14)(23), & \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} &= (143), \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix} &= (243), & \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} &= (12)(34). \end{aligned}$$

A representação de sinais associada aos subgrupos de S_4 que pertencem a Classe 7 é ilustrada na Figura 5.13.

Agora, a representação em \mathbb{R}^3 consiste de um tetraedro truncado (Veja Figura 5.14).

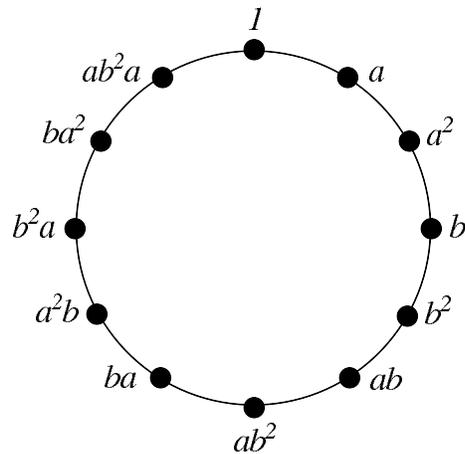


Figura 5.13: Representação de sinais do grupo alternante A_4 no \mathbb{R}^2 .

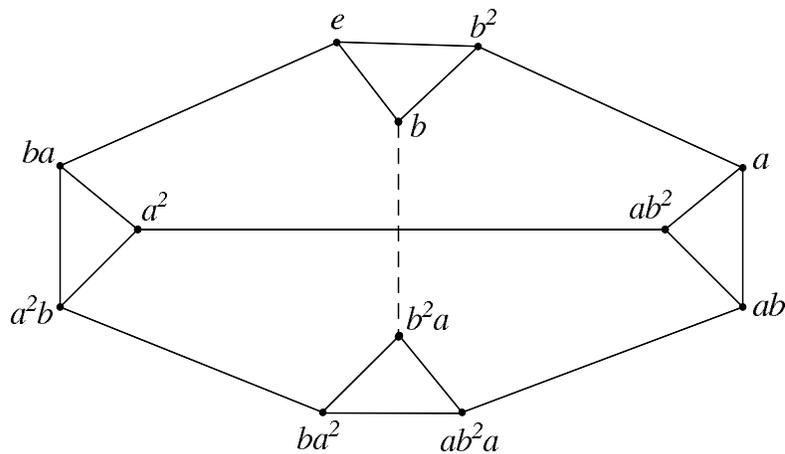


Figura 5.14: Representação de sinais do grupo alternante A_4 no \mathbb{R}^3 (tetraedro truncado).

5.3 Proposta de Constelação de Sinais para o Código Genético

Antes de começar a nossa proposta, comentaremos sobre algumas particularidades do código genético. Como já mencionamos anteriormente, o código genético é degenerado (Veja Capítulo 2, Subseção 2.3.3) e neste código genético temos certos padrões entre as trincas (códon) especificando ou não o mesmo tipo de aminoácido. Por outro lado, este padrão é “quebrado”, pois existem trincas que, mesmo possuindo o mesmo padrão, especificam mais do que um aminoácido. Por exemplo: os códon CAC, CAU, CAA, CAG apresentam o padrão CA, porém os códon CAC e CAU especificam a histidina (His), enquanto que CAA, CAG especificam a glutamina (Gln) e, diferentemente, da treonina (Thr) que é especificada por: ACC, ACU, ACA e ACG. Neste último caso, independentemente da última letra que se agregue ao padrão “AC”, os códon sempre especificarão a treonina.

Apresentamos nesta seção duas possíveis constelações de sinais associadas ao código genético, interpretadas e construídas de maneiras distintas. A primeira representação de sinais, ilustrada na Figura 5.15, é a clássica, ou seja, o código genético está representado em uma única constelação de sinais. Em tal constelação de sinais temos os 20 tipos de aminoácidos, que são representados por pontos coloridos, e os códons de finalização, representados por um ponto em vermelho. Os aminoácidos estão distribuídos em cinco órbitas (circunferências tracejadas), onde o número de códons relacionado a cada aminoácido determina a órbita a qual este pertence. Denotamos por aa_1 a órbita que contém somente aminoácidos especificados por um único có-

$$\begin{array}{lll}
 aa_1 = 1 \text{ códon} & aa_3 = 3 \text{ códons} & aa_6 = 6 \text{ códons} \\
 aa_2 = 2 \text{ códons} & aa_4 = 4 \text{ códons} &
 \end{array}$$

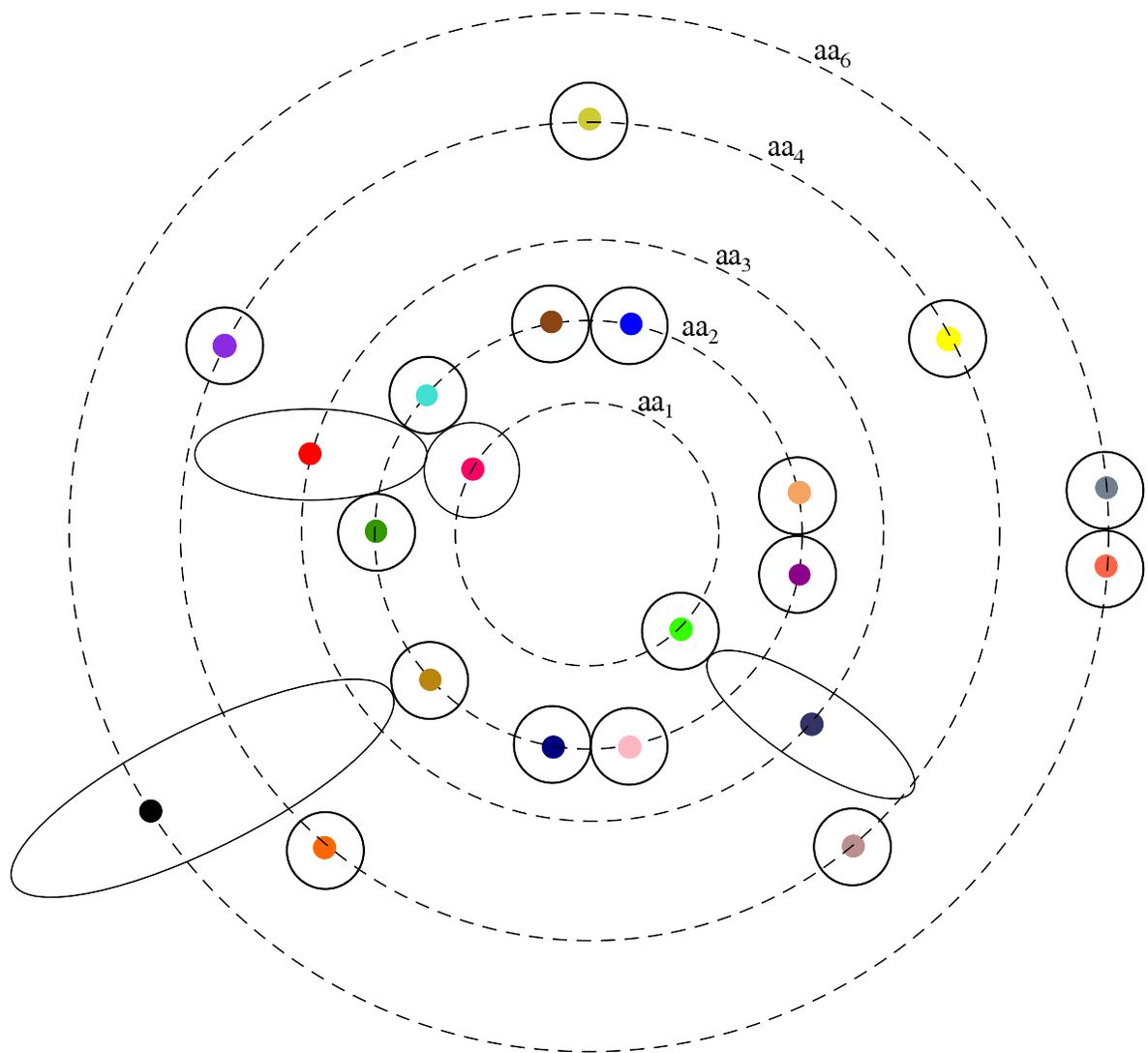


Figura 5.15: Representação do código genético em uma constelação de sinais. O ponto de tangência entre as subórbitas indica padrões em comum entre os códons.

don, aa_2 a órbita que contém aminoácidos especificados por dois códons e assim sucessivamente. Associado a cada aminoácido na correspondente órbita existe uma única subórbita (circunferência ou elipse). As subórbitas são representadas pelos subgrupos de S_3 e S_4 conforme visto nas Subseções 5.1.1 e 5.2.1, inserindo, dessa forma, uma estrutura algébrica para cada subórbita, havendo portanto, a possibilidade da representação das constelações de sinais em dimensões maiores, como, por exemplo, em \mathbb{R}^3 ou \mathbb{R}^4 . Os pontos tangentes às subórbitas indicam padrões em comum entre os códons.

A segunda representação de sinais associada ao código genético (veja Figura 5.16), denominada constelações de sinais correlacionadas, é uma proposta inovadora de se construir constelações de sinais na teoria da comunicação, pois origina a idéia de constelações concatenadas. Acreditamos que, futuramente, possamos gerar constelações concatenadas capazes de corrigir erros, ao contrário das constelações de sinais presentes na literatura. Esta idéia foi fruto do desenvolvimento deste trabalho. Não encontramos na literatura nenhum trabalho que aborde este assunto. Nesta construção os aminoácidos estão distribuídos em duas constelações correlacionadas, contidos em suas referidas órbitas. Neste caso, em particular, a órbita que contém um aminoácido, especificado por um único códon, coincide com o próprio aminoácido e as subórbitas são representadas somente por circunferências. As flechas indicadas na Figura 5.16 representam como os diferentes tipos de aminoácidos mais os códons de terminação estão atrelados através de determinados padrões. Note que se o código genético não apresentasse correlação entre os aminoácidos teríamos duas constelações independentes. O não conhecimento desta correlação implica na indeterminação, pelo menos até o presente momento, de um modelo que possa traduzir efetivamente a estrutura algébrica e geométrica associadas ao código genético.

Os elementos algébricos e geométricos apresentados nos Capítulos 3, 4 e 5, fornecem as condições mínimas para que a proposta seja considerada, isto é, que o código genético possa ser visto em duas dimensões como sendo um conjunto consistindo de uma constelação de sinais (caso clássico) ou de duas constelações de sinais correlacionadas (caso proposto neste trabalho).

Observação 10 *Os aminoácidos considerados na Figura 5.15 não são necessariamente os mesmos da Figura 5.16.*

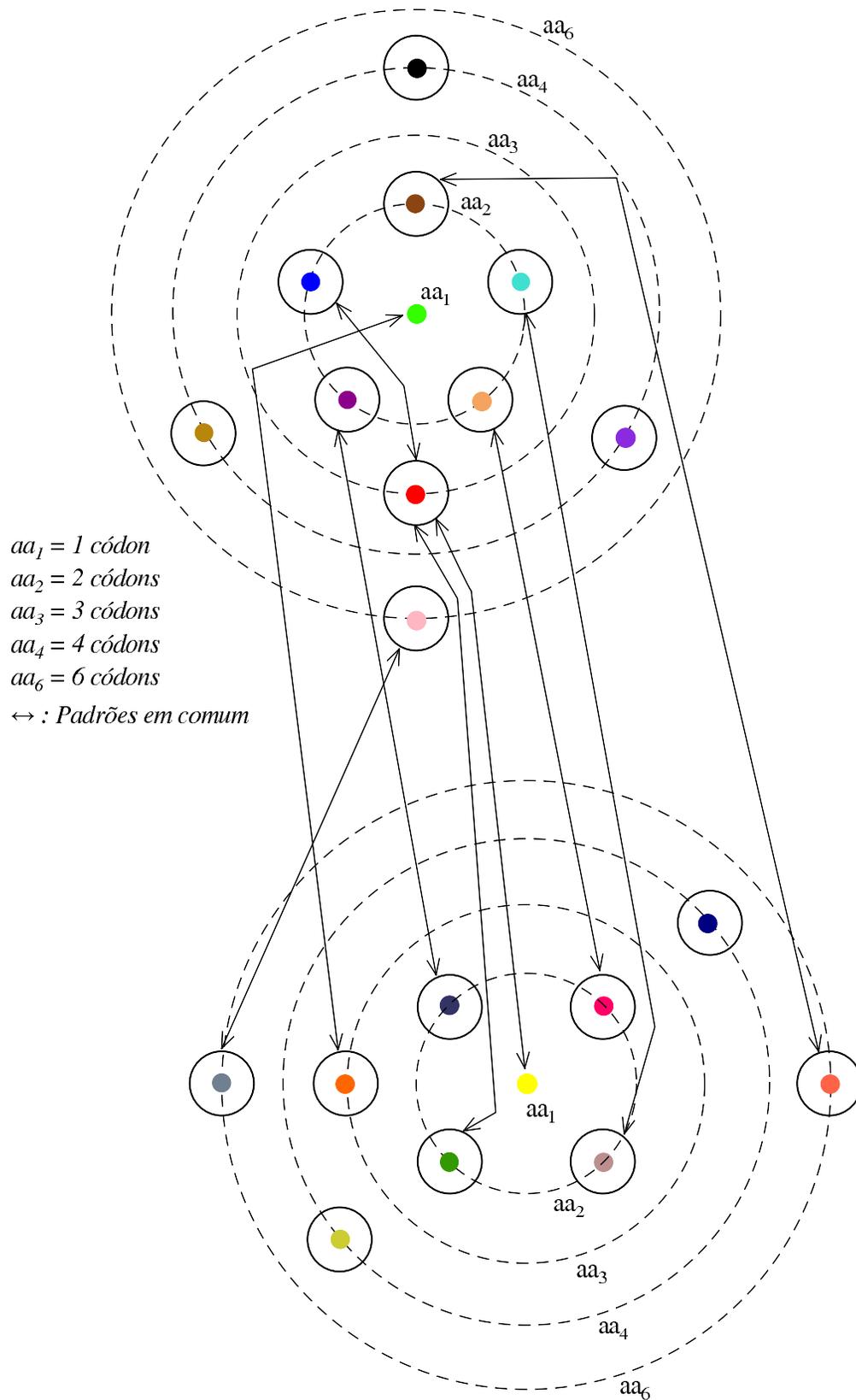


Figura 5.16: Representação do código genético em duas constelações de sinais correlacionadas de acordo com os padrões em comum de cada códon.

Conclusões

A principal proposta deste trabalho foi descrever o código genético como uma constelação de sinais. Devido ao contexto envolvido nesta proposta tivemos que estudar e abordar três assuntos de áreas distintas: biologia molecular, teoria de grupos e teoria de comunicações e codificação, necessárias para o desenvolvimento deste trabalho.

- No Capítulo 1 foi apresentado uma breve introdução histórica sobre o surgimento da estrutura do DNA e a descrição do problema e as suas limitações;
- No Capítulo 2 tratamos de conceitos biológicos importantes para o entendimento da modelagem do dogma central da biologia molecular, descrita no Capítulo 4;
- No Capítulo 3 estudamos as características e propriedades dos grupos, mais especificamente os grupos de permutação S_3 e S_4 : representação e associação geométrica de seus subgrupos;
- No Capítulo 4 apresentamos uma associação do sistema de codificação com o dogma central da biologia molecular e a partir desta associação apresentamos o modulador genético;
- No Capítulo 5 foram apresentadas as constelações de sinais dos subgrupos de S_3 e S_4 , importante resultado pois estas constelações serão associadas as subórbitas dos aminoácidos (mais os códons de finalização) presentes nas duas possíveis representações do código genético como constelações de sinais.
- A introdução deste modelo biológico de comunicação amplia a nossa abordagem na genética molecular, enriquece a nossa visão ao tratar dados e processos biológicos.

6.1 Perspectivas

- Até o presente momento, desconhecemos na literatura uma abordagem utilizando teoria de grupos com este enfoque. Devemos utilizar linguagem de programação, indispensável para o prosseguimento deste trabalho, para gerar e verificar os subgrupos com cardinalidade 64 em S_{64} e comparar com o conjunto de partições geradas pelos aminoácidos no código genético.
- A interpretação e construção de constelações de sinais correlacionadas associadas ao código genético permite-nos propor na teoria da comunicação e codificação a construção de constelações de sinais concatenadas. Acreditamos que futuramente as constelações de sinais concatenadas sejam utilizadas na teoria da comunicação e codificação para correção de erros.
- Futuramente, outro ferramental matemático que pode ser utilizado na abordagem deste problema de modulação genética é a teoria de grafos.

Bibliografia

- Alberts, B., Raff, A. J. M., Roberts, K. & Walter, P. (2004). *Biologia Molecular da Célula*, 4^a edn, ARTMED Editora Ltda.
- Allenby, R. B. J. T. (1993). *Rings, Fields and Groups*, 2^a edn, British Library.
- Armstrong, M. A. (1988). *Groups and Symmetry*, Springer.
- Câmara, C. E. (1995). *Construção de Códigos Esféricos Via a d-Cadeia e a Geometria de Grupos*, PhD thesis, FEEC - Universidade Estadual de Campinas, Campinas - SP - Brasil.
- da Rocha, A. S. L. (2004). *Modelos matemáticos para a previsão de recombinações sítio-específico do DNA.*, Master's thesis, FEEC - Universidade Estadual de Campinas, Campinas - SP - Brasil.
- de Carvalho, E. D. (2001). *Construção e rotulamento de constelações de sinais geometricamente uniformes em espaços euclidianos e hiperbólicos*, PhD thesis, FEEC - Universidade Estadual de Campinas, Campinas - SP - Brasil.
- de Faria, L. C. B. (2004). *Caracterizações topológica, geométrica e algébrica dos produtos da recombinação do dna através dos modelos tangle e frações contínuas.*, Master's thesis, FEEC - Universidade Estadual de Campinas, Campinas - SP - Brasil.
- Farah, S. B. (2007). *DNA Segredos e Mistérios*, 2^a edn, Sarvier.
- Firer, M. (2001). Grupos fuchsianos, Notas de aula, IMMEC-Unicamp.
- Fraleigh, J. B. (2003). *A First Course in Abstract Algebra*, 7^a edn, Greg Tobin.
- Garcia, A. & Lequain, Y. (2002). *Elementos de Álgebra*, IMPA.
- Gilbert, W. J. (1976). *Modern Algebra with Applications*, John Wiley & Sons.

- Griffiths, A. J. F., Wessler, S. R., Lewontin, R. C., Gelbart, W. M., Suzuki, D. T. & Miler, J. H. (2006). *Introdução à Genética*, 8ª edn, Guanabara Koogan S. A.
- Hayes, B. (1998). The invention of the genetic codes, *American Scientist* **86**: 8–14.
- Hefez, A. & Villela, M. L. T. (2002). *Códigos Corretores de Erros*, 1ª edn, IMPA.
- Kari, L., Kari, J. & Landweber, L. F. (1999). Reversible molecular computation in ciliates, *Jewels are Forever, Contributions on Theoretical Computer Science in Honor of Arto Salomaa*, Springer-Verlag, London, UK, pp. 353–363.
- Lewin, B. (2000). *Genes VII*, ARTMED Editora Ltda.
- Lin, S. & Costello Jr., D. J. (2004). *Error Control Coding*, 2 edn, Pearson Prentice Hall.
- Lodish, H., Berk, A., Baltimore, D., Zipurski, L. S. & Parnell, J. (1998). *Molecular Cell Biology*, Scientific American Books.
- Loewenstern, D. & Yanilos, P. N. (1997). Significantly lower entropy estimates for natural dna sequences, *Data Compression Conference*.
- May, E. E. (2002). *Analysis of Coding Theory Based on Models for Initiating Protein Translation in Prokaryotic Organisms*, PhD thesis, Department of Electrical and Computer Engineering - North Carolina.
- May, E. E. (2004). Towards a biological coding theory discipline, *NEW THESIS* **1**(1): 19–38.
- May, E. E., Rintoul, M. D., Johnston, A. M., Hart, W. E., Watson, J. P. & Pryor, R. J. (2003). Detection and reconstruction of error control codes for engineered and biological regulatory systems, *Technical report*, Sandia National Laboratories.
- May, E. E., Vouk, M. A., Bitzer, D. L. & Rosnick, D. I. (2002). A coding theory framework for genetic sequence analysis, *Technical report*, Sandia National Laboratories.
- May, E. E., Vouk, M. A., Bitzer, D. L. & Rosnick, D. I. (2004). Coding theory based models for protein translation initiation in prokaryotic organisms, *BIOSYSTEMS* **76**(1-3): 249–260.
- May, E., Vouk, M. A., Bitzer, D. L. & Rosnick, D. I. (2000). Coding theory based in maximum-likelihood classification of translation initiation regions in escherichia coli k-12, *Biomedical Engineering Society Annual Meeting*.
- Román-Rolda'n, R., Bernaola-Galván, P. & Oliver, J. L. (1996). Application of information theory to DNA sequence analysis: a review, *Pattern Recognition* **29**(7): 1187–1194.

-
- Schneider, T. D. (1991). Theory of molecular machines. i. channel capacity of molecular machines, *Journal of Theoretical Biology* **148**(1): 83–123.
- Schneider, T. D. (1997). Information content of individual genetic sequences, *Journal of Theoretical Biology* **189**(4): 427–441.
- Schneider, T. D. & Mastronard, D. N. (1996). Fast multiple alignment of ungapped dna sequences using information theory and a relaxation method, *Discret Applied Mathematics* **71**: 259–268.
- Sengupta, R. & Tompa, M. (2002). Quality control in manufacturing oligo arrays: A combinatorial design approach, *Journal of Computation Biology* **9**(1): 1–22.
- Silva, A. A. (1996). *Uma Contribuição à Classe dos Códigos Geometricamente Uniformes*, PhD thesis, FEEC - Universidade Estadual de Campinas, Campinas - SP - Brasil.