



BIOINFORMÁTICA DA BIOLOGIA À FLEXIBILIDADE MOLECULAR

CAPÍTULO 2: NÍVEIS DE INFORMAÇÃO BIOLÓGICA

Matéria: Bioinformática

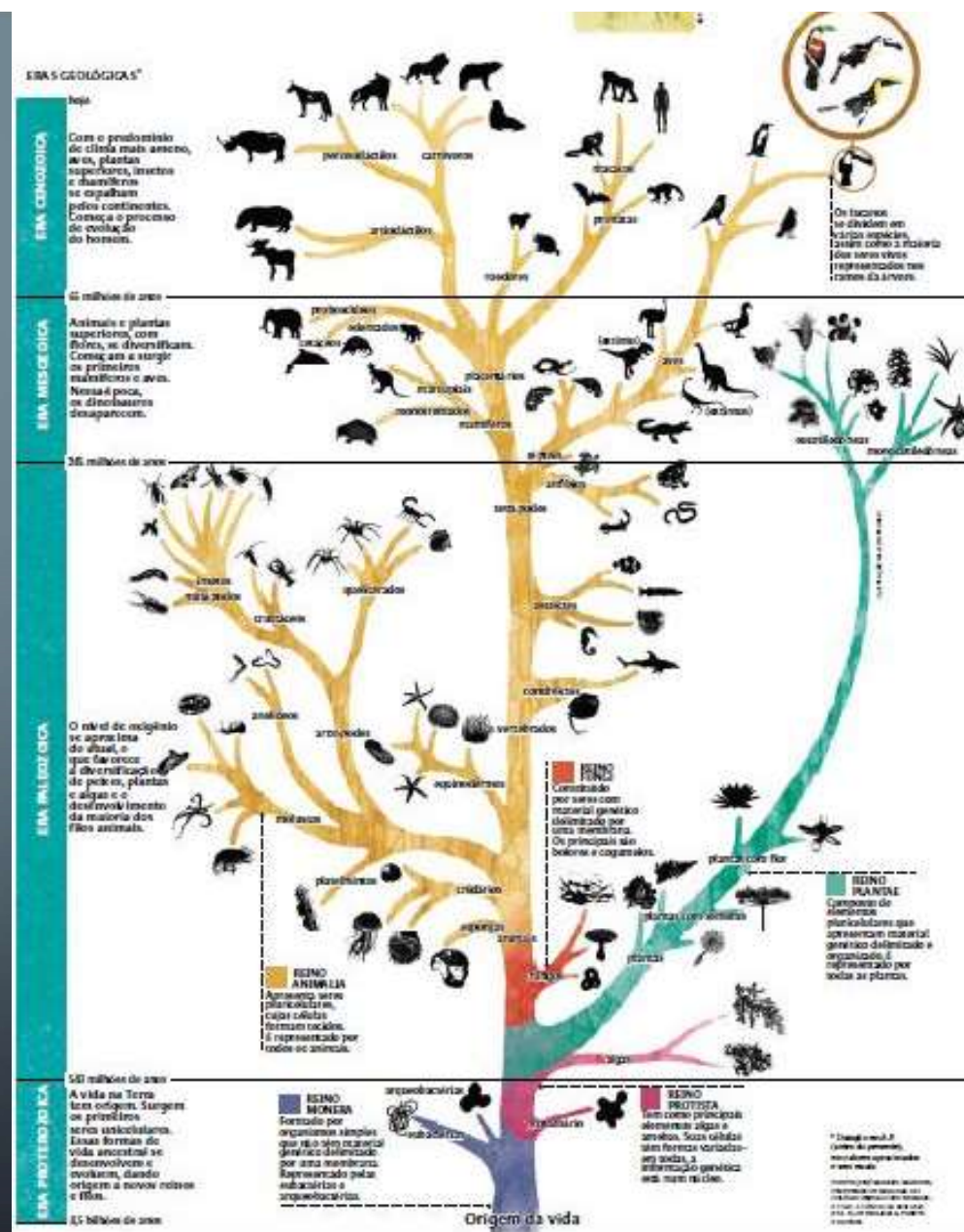
Docente: Prof. Dr. Teodiano Freire Bastos Filho

Discentes: Lucas Bonela, Sheila Luz, Guilherme Crespo

CCCTGTGGAGCCACACCTAGGGTTGGCCA
ATCTACTCCCAGGAGCAGGGAGGGCAGGAG
CCAGGGCTGGGCATAAAAGTCAGGGCAGAG
CCATCTATTGCTTACATTGCTTCTGCACAC
AACTGTGTTCACTAGCAACTCAAACAGACA
CCATGGTGCACCTGACTCTCTGAGGAGAAGT
CTGGCGTTACTGCCCTGTGGGCGAAGTGA
ACCTGGATGAAGTTGGTGGTGGAGCCCTGG
GCAGGTTGGTATCAAGGTTACAAAGACAGGT
TTAAGGAGACCAATAGAACTGGGCATGTG
GAGACAGAGAAGACTCTTGGGTTTCTGATA
GGCACTGACTCTCTCTGCTATTGGTCTAT
TTTCCCACCTTTAGGCTGCTGGTGGTCTAC
CCTTGGACCCAGAGGTTCTTTGAGTCTCTT
GGGGATCTGTCCACTCCTGATGCTGTTATG
GGCAACCTTAAGGTGAAGGCTCATGGCAAG
AAAGTGTCTGGTGGCTTTAGTGATGGCTG
GCTCAGCTGGACAACCTCAAGGGGACCTTT
GCCCACTGAGTGAGCTGCAGCTGTGACAAG
CTGCACGTGGATCCTGAGAATCTCAGGCTG
AGTCTATGGGACCTTGTATTTTCTTCTCC
CCTCTCTTTCTATGTTTAAAGTTCATGTTCAT
AGGAAGGGGAGAAGTAACAGGGTACAGTTT
AGAATGGGAACAGACGAATGATTGCATCA
GTGTGGAAAGTCTCAGGATCGTTTATGTTTC
TTTATTTGCTGTTTCATAACAATGTTTTC
TTTGTGTTAATCTTGGCTTTCTTTTCTTTT
CTTCTCCGCAATTTTACTATTATACATAA
TGCTTAAACATTGTGTATAACAAAGGAAA
TATCTCTGAGATACATTAAAGTAACTTAAAA
AAAACTTTACACAGCTCGCTAGTACATT
ACTATTTGGAATATATGTGTGCTTATTTGC
ATAATCATAATCTCCCTACTTTATTTCTTT
ATAATCATAATCTCCCTACTTTATTTCTTT
TTATTTTAAATTGATACATAATCATTTATAC
ATATTTATGGGTTAAAGTGTAATGTTTAA
TATGTGTACACATATTGACCAATCAGGT
AATTTGCACTTTGTAATTTTAAAAATGCT
TTCTCTTTTAAATATACTTTTGTGTTATC
TTATTTCTAATCTTTCCCTAATCTTTCTT
TTTCAGGGCAATATGATACAATGTATCAT
GCCTCTTTGCACCATTTCAAAGAATAACAG
TGATAATTTCTGGTTAAGGCAATAGCAAT
ATTTCTGCATATAAATATTTCTGCATATAA
ATTGTAACATGATGTAAGAGGTTTCATATTG
CTAATAGCAGCTACAATCCAGCTACCATTC
TGCTTTTATTTATGGTTGGGATAAGGCTG
GATTATCTGAGTCCAAGCTAGGCGCTTTT
GCTAATCATGTTTACACCTCTTATCTTCT
CCCACAGCTCTCTGGGCAACGTCTGGTCTG
TGTGCTGGGCAATACATTTGGCAAGAAAT
CAGCCACAGTGCAGGCTGCCTATCAGAA
AGTGGTGGCTGGTGGCTAATGCCCTGGC
CCACAAGTATCACTAAGCTCGCTTTCTTGC
TGTCGAATTTCTATTAAAGGTTCTTTGTT
CCCTAAGTCCAACCTACTAACTGGGGGATA
TTATGAAGGGCTTGAGCATCTGGATCTTG
CCTAATAAAAAACATTTATTTTCATTGCA
TGAATGATTTAAATATTTCTGAATATTT
ACTAAAAAGGAATGTGGGAGGTCACTGCA
TTTAAACATAAGAAATGATGAGCTGTTT
AAACCTTTGGGAAATACACTATATCTTTAA
CTCATGAAAGAGGTGAGGCTGCAACAG
CTAATGCACATTGGCAACAGCCCTGATGC
CTATGCTTATTCATCCCTCAGAAAGGAT
TCTTGTAGAGGCTGATTTGCAGGTAAAG
TTTGTCTATGCTGATTTTACATTACTTAT
TGTTTATGCTGCTCATGAATGCTTTTTC

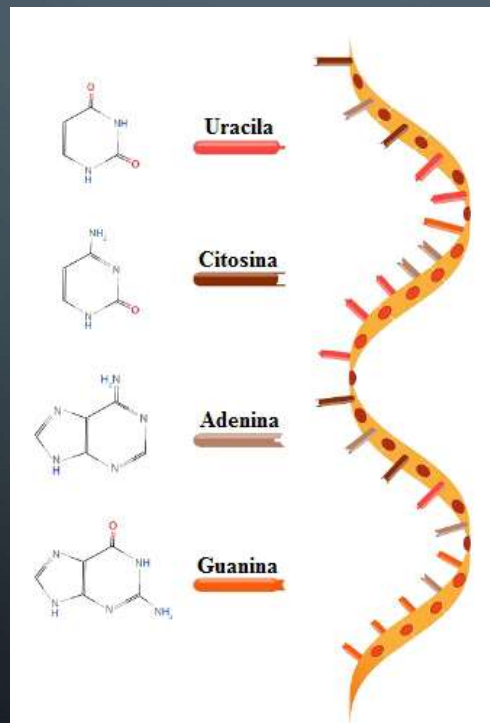
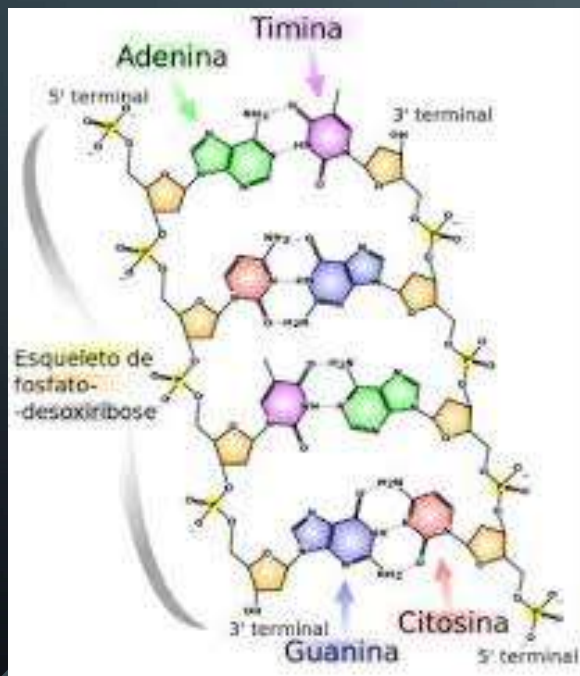
2.1 INTRODUÇÃO

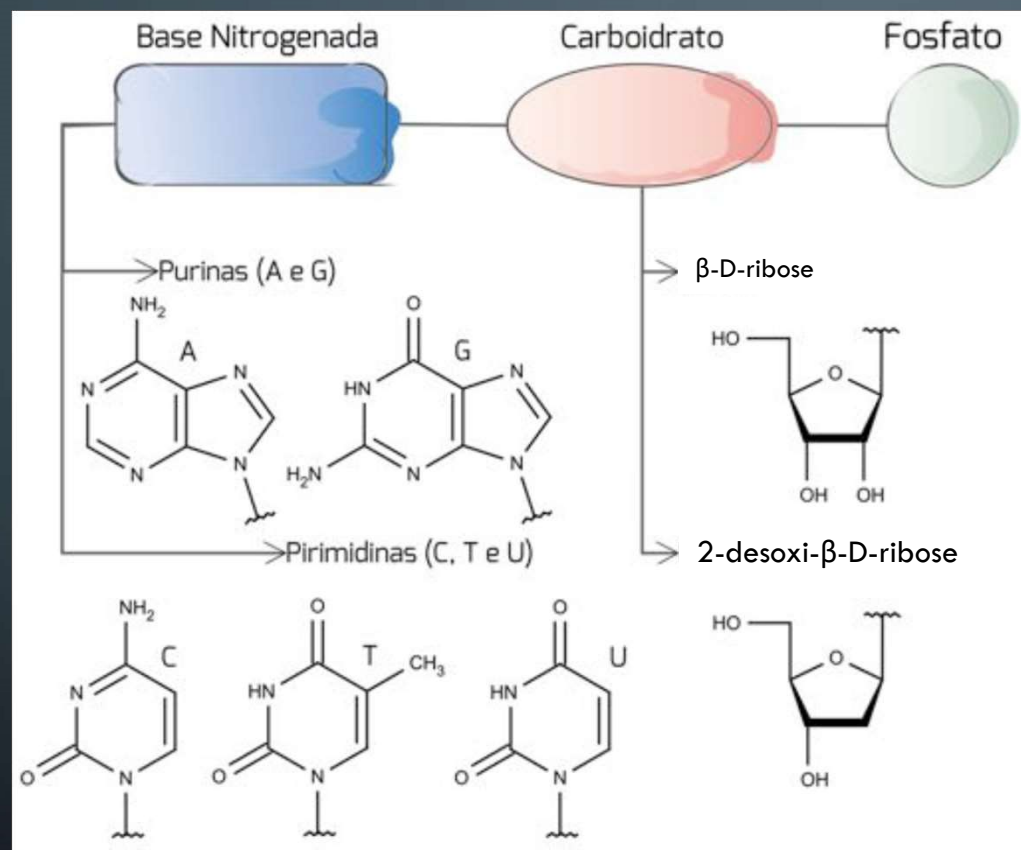
- Toda informação necessária para a formação de um organismo como nós conhecemos está contida em uma molécula linear!
- A bioinformática é capaz de lidar com todas essa biomoléculas, mas se limitando ao custo computacional.



2.2 MACROMOLÉCULAS BIOLÓGICAS

- Biopolímeros: DNA, RNA e Proteínas





ÁCIDOS NUCLEICOS

- $\beta\text{-D-ribose} + \text{Base} = \text{Citidina...}$
- (N) Fosfatos + Base = AMP, ADP ATP
- As bases nitrogenadas possuem estrutura tridimensional plana e invariável (degraus da escada).

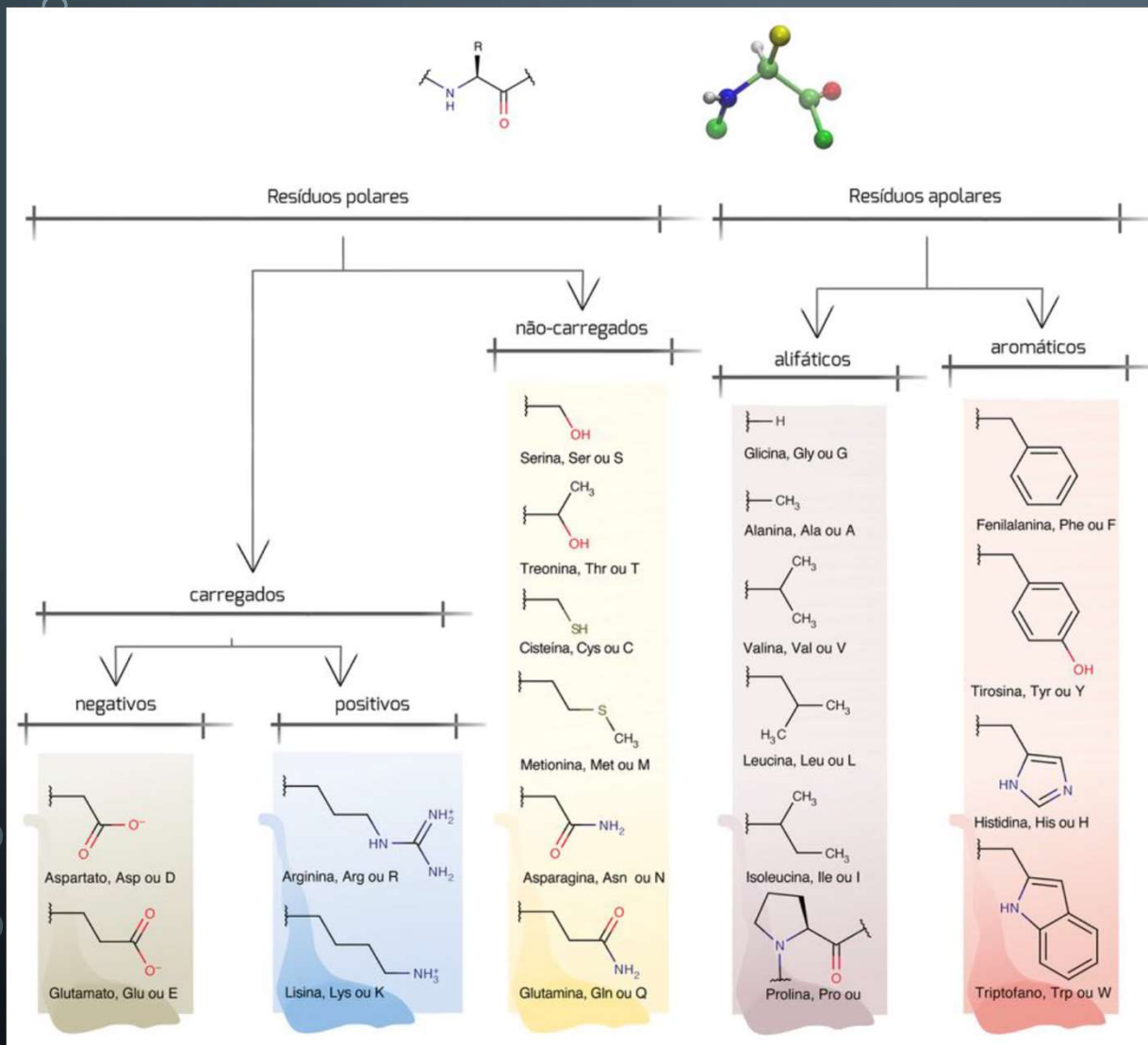
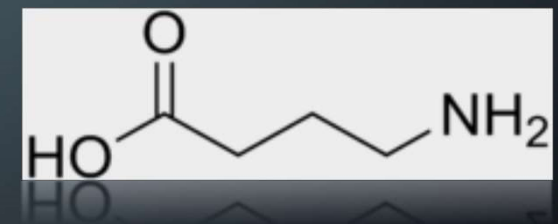
Figura 2-2: Representação esquemática de um nucleotídeo e suas variações na base nitrogenada e no carboidrato.(Adaptada)

PROTEÍNAS

Formadas pela combinação de 20 aminoácidos(AA).

Alguns Aas não são codificados pelo DNA, mas possuem funções no organismo.

→ GABA, GLA



CARBOIDRATOS

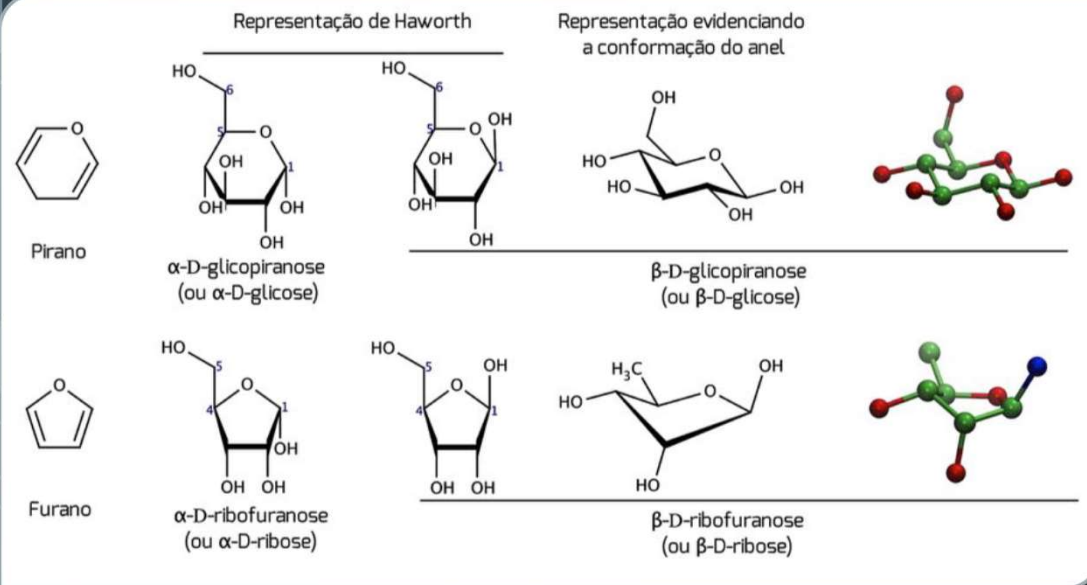


Figura 5-2: Os dois principais grupos de carboidratos envolvem monossacarídeos compostos por anéis de 5 (furanoses) e 6 membros (piranoses).

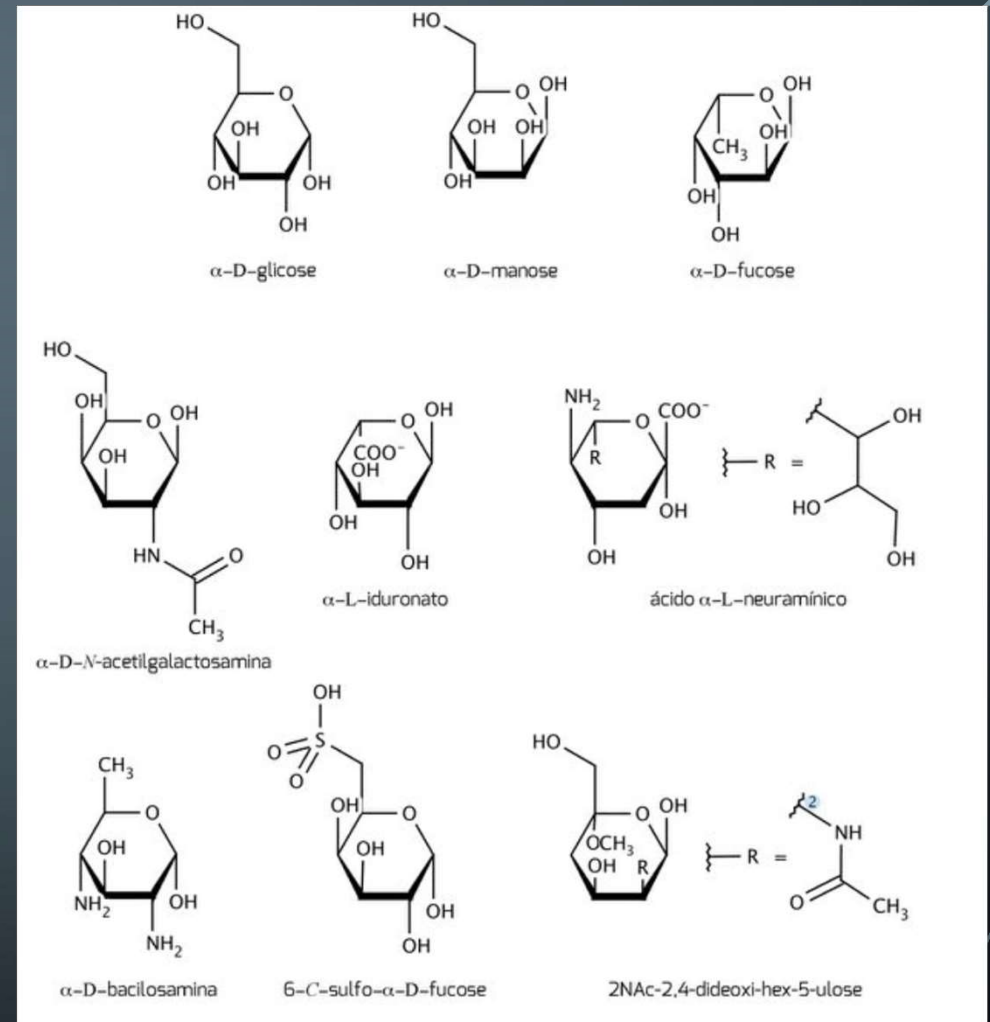
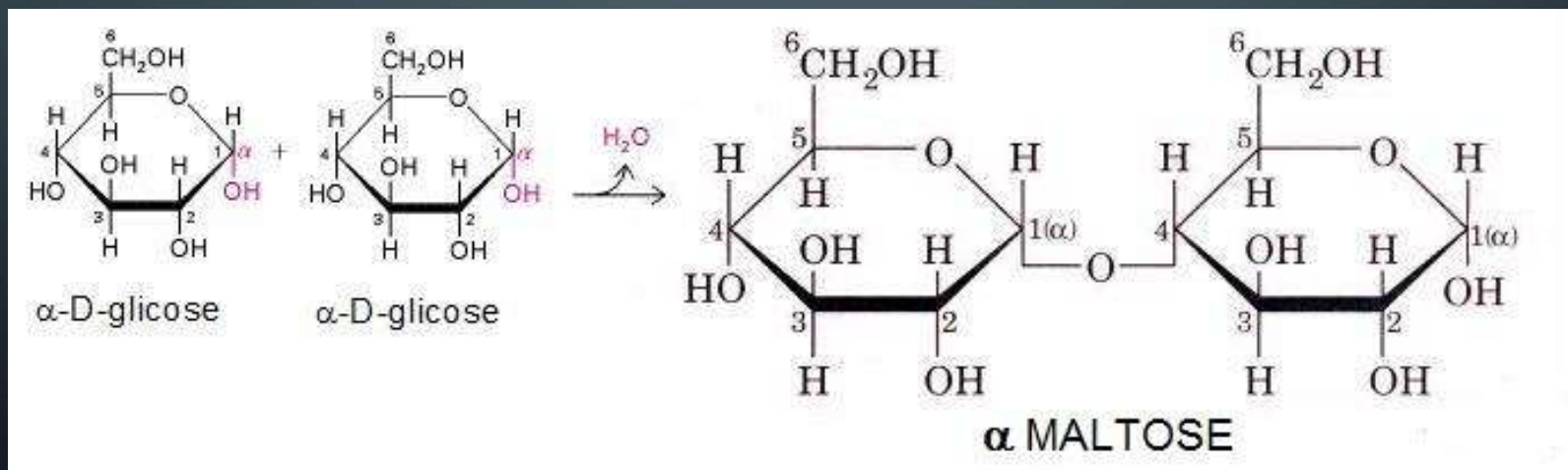


Figura 7-2: Exemplo da complexidade de possíveis monossacarídeos encontrados na natureza.

- Sua síntese é direcionada pela atividade enzimática.
 - Glicosil transferase



MEMBRANAS

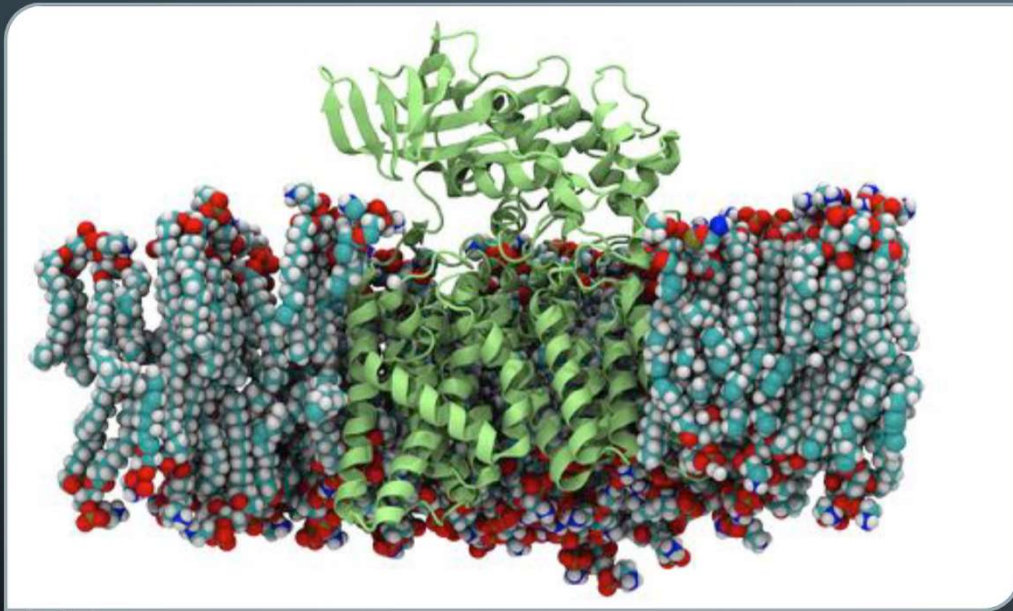
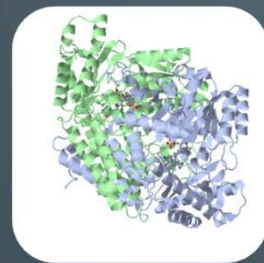
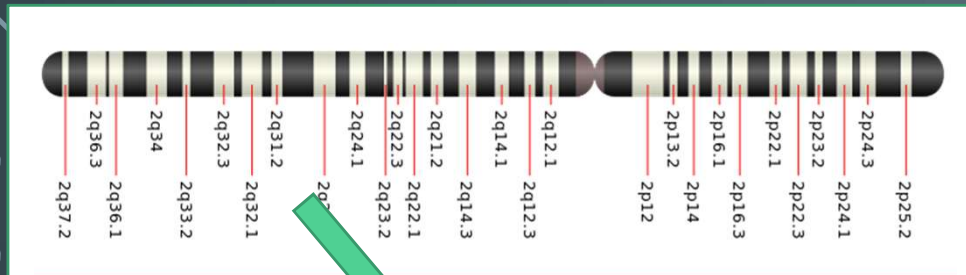
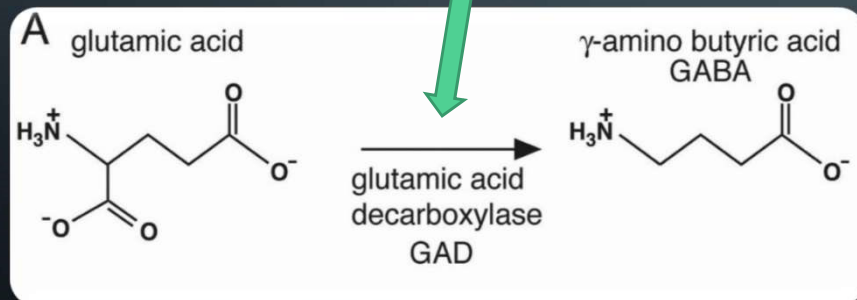


Figura 8-2: Representação de uma membrana POPE (palmitoil oleil fosfatidil etanolamina) contendo a enzima PglB (oligossacaril transferase) de *Campylobacter lari*. Os átomos de oxigênio estão representados em vermelho, os átomos de carbono em verde, os átomos de hidrogênio em branco e nitrogênios em azul. A enzima está representada como cartoon verde.





GAD1



“MICROMOLÉCULAS” BIOLÓGICAS

- Moléculas de baixo peso molecular e que não estão codificadas diretamente no DNA, mas são produzidas por diferentes enzimas.
- Neurotransmissores: GABA, Acetilcolina, Noradrenalina
- Hormônios: GH, Vitamina D

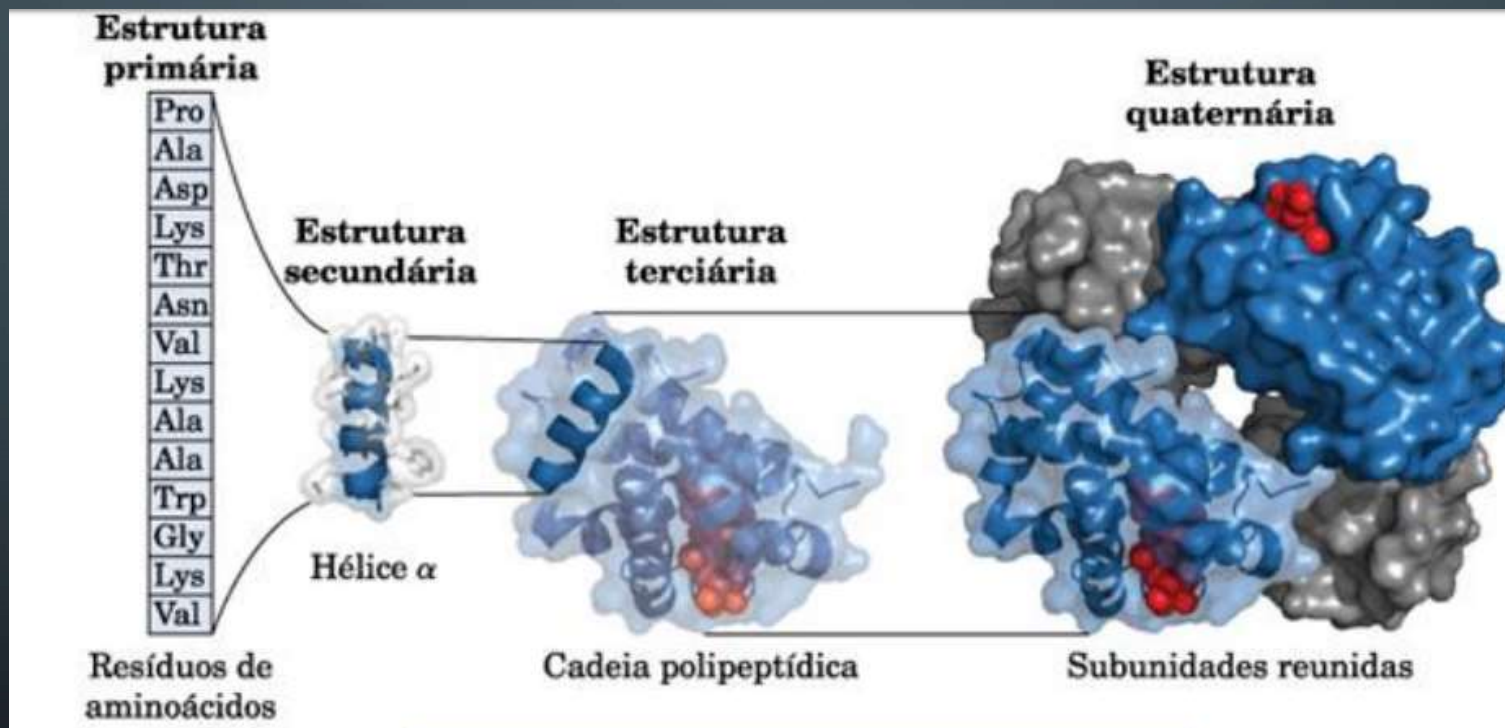


2.3 NÍVEIS DE ORGANIZAÇÃO

CLASSIFICAÇÃO DA ESTRUTURA DE BIOMACROMOLÉCULAS

- Quatro diferentes níveis de complexidade;
- Cada nível traz volume e tipos de informações diferentes, exigindo poder computacional e abordagens distintas;
- Em princípio apresentam nível hierárquico;

CLASSIFICAÇÃO DA ESTRUTURA DE BIOMACROMOLÉCULAS



ESTRUTURA PRIMÁRIA

- Nível inicial de complexidade;
- Padrão de letras ou conjunto de letras que representa a composição do biopolímero;
- Natureza unidimensional (1D);
- Única dimensão descrita é a ordem de aparecimento dos monômeros;

ESTRUTURA PRIMÁRIA

- Em ácidos nucleicos consiste em estrutura de nucleotídeos;
- Para proteínas consiste em uma sequência de AA;
- Para carboidratos consiste em uma sequência de monossacarídeos.

ESTRUTURA PRIMÁRIA

A troca de um resíduo por outro de propriedade distinta pode ser mais facilmente acomodada na estrutura flexível do que nos outros tipos de estruturas secundárias, mais rígido.

DNA:

GGTATAGGCGCTGTTCTTAAGGTGCTAACAAACGGGGT
TACCCGCGTTGATCTCGTGGATAAAACGCAAACGCCA
ACAG

RNA:

GGUAUAGGCGCUGUUCUUAAGGUGCUAACAAACGGG
GUUACCCGCGUUGAUCUCGUGGAUAAAACGCAAAC
GCCAACAG

Aminoácidos:

GIGAVLKVLTTGLPALISWIKRKRQQ

Sequência sacarídica:

α -D-GlcNAc,6S-(1 $\rightarrow$ 3)- β -D-GlcA-(1 $\rightarrow$ 4)- α -D-
GlcN5,3S,6S-(1 $\rightarrow$ 4)- α -L-IdoA,2S-(1 $\rightarrow$ 4)- α -D-
GlcN5,6S

ESTRUTURA PRIMÁRIA

- Menor complexidade;
- Oferece grande volume de informações sobre a forma nativa e as funções das biomoléculas;
- As informações advém da comparação de sequências de nucleotídeos ou AA em busca de padrões específicos associados a determinadas características ou funções;

ESTRUTURA SECUNDÁRIA

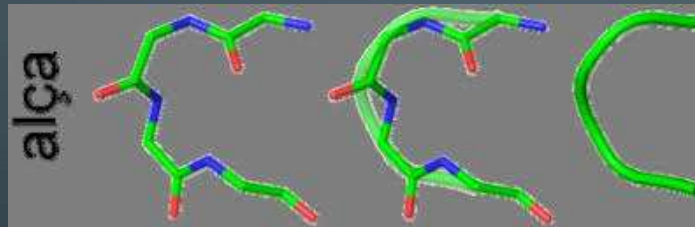
- Surge das interações entre as sequências de monômeros em uma determinada ordem específica e monômeros vizinhos e moléculas solventes;
- Estas interações originam padrões repetitivos de organização espacial.

ESTRUTURA SECUNDÁRIA

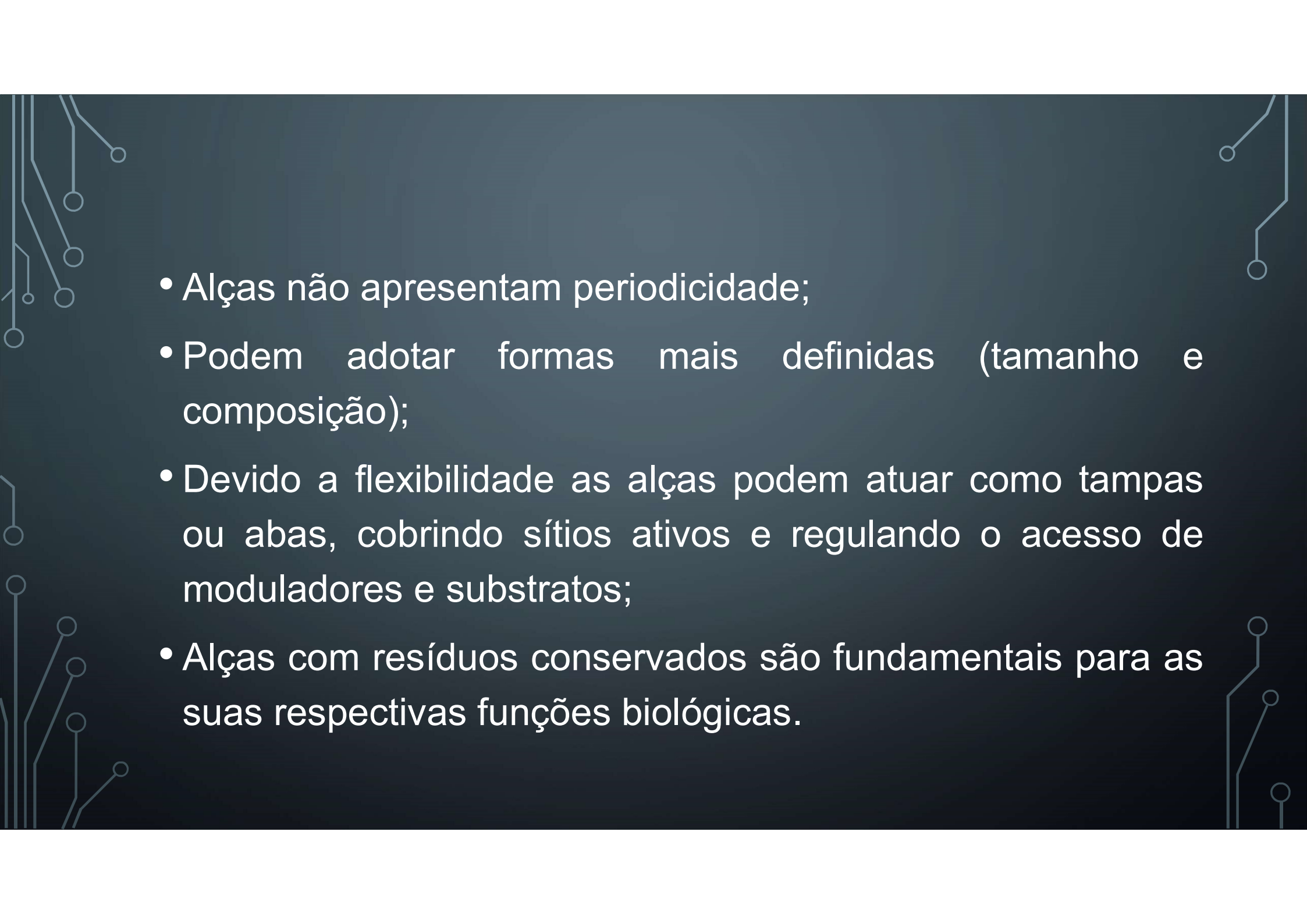
- Estrutura tridimensional;
- Diferentes composições de estrutura primária podem gerar um mesmo tipo de estrutura secundária;
- As mais lembradas são as proteínas, que incluem três grupos de elementos principais (as alças, as hélices e as folhas β).

ESTRUTURA SECUNDÁRIA

- As alças são elementos envolvidos na conexão entre as hélices e as folhas;



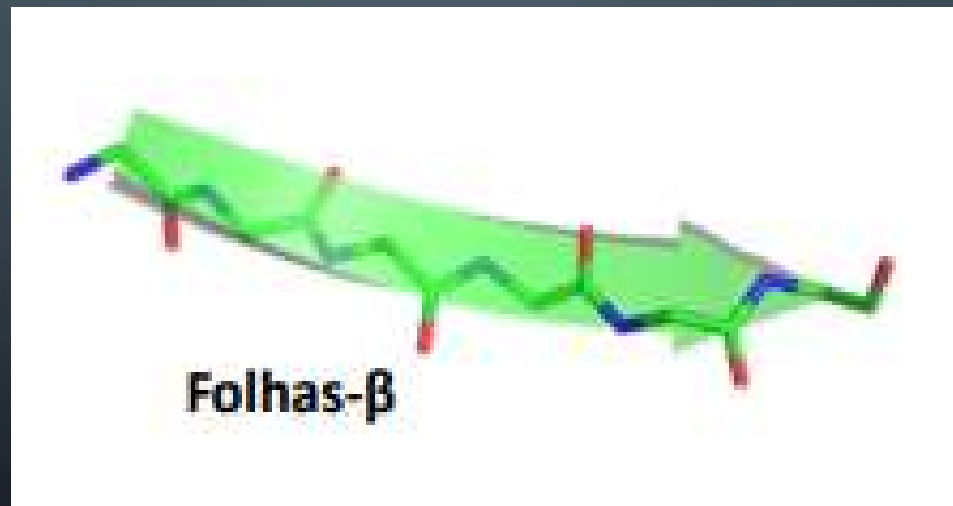
- Tendem a ser mais flexíveis (aumenta conforme o tamanho da alça aumenta);
- São mais susceptíveis a sofrerem mutações (salvo sob pressão evolutiva).

- 
- Alças não apresentam periodicidade;
 - Podem adotar formas mais definidas (tamanho e composição);
 - Devido a flexibilidade as alças podem atuar como tampas ou abas, cobrindo sítios ativos e regulando o acesso de moduladores e substratos;
 - Alças com resíduos conservados são fundamentais para as suas respectivas funções biológicas.

FOLHAS B

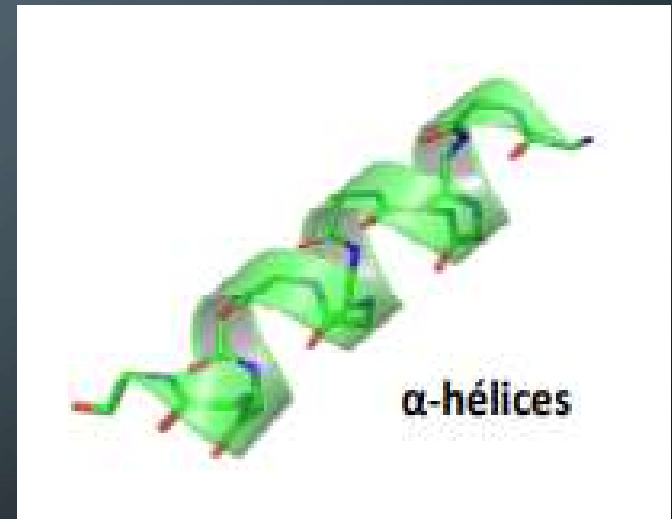
- Formadas por sequências de AA (fita);
- As fitas são quase completamente estendidas e lineares;
- Interação lado a lado ao longo dos seus eixos longitudinais (série de ligação de hidrogênio entre os grupamentos N-H de uma fita e o grupamento C=O da fita vizinha).

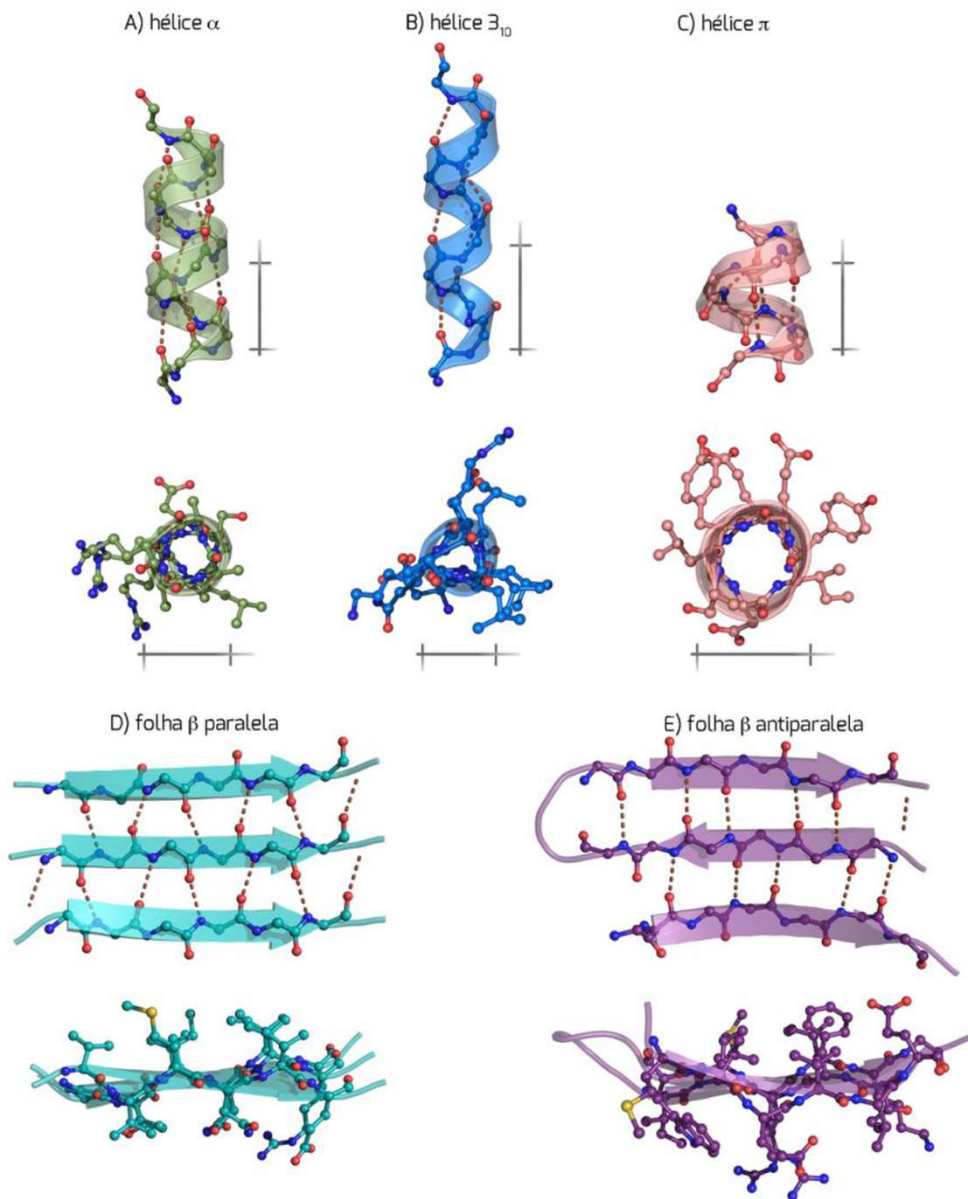
Para que essa organização seja possível, os átomos de $C\alpha$ adotam orientação intercalada, acima e abaixo do plano da folha (Folhas β pregueadas).



HÉLICES

- Também envolve a realização de ligações de hidrogênio entre os grupos N-H e C=O vizinhos no espaço (não na sequência).
- As interações acontecem com resíduos mais próximos na sequência, entre as voltas da hélice.





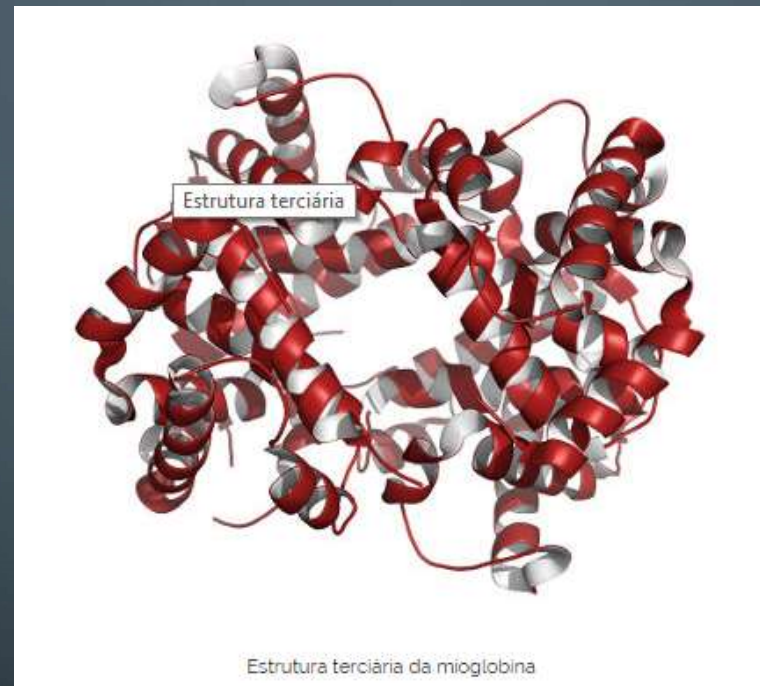
Proteínas, ácidos nucleicos e carboidratos podem apresentar padrões repetitivos de organização espacial, definidos como sequência de monômeros que os constituem.

Figura 10-2: Representação dos tipos mais comuns de estrutura 2ária encontrados em proteínas. Em verde estão as hélices α (A), em azul as hélices 3_{10} (B), em salmão as hélices π (C), em ciano as folhas β paralelas (D) e roxo as antiparalelas (E). As ligações de hidrogênio entre átomos do esqueleto peptídico estão apresentadas como linhas tracejadas em marrom. As estruturas são partes que compõe as proteínas descritas pelos códigos PDB 18D8, 1ABB, 2QD1, 1EE6 e 1PC0, e para cada uma duas diferentes orientações são apresentadas. Note que as cadeias laterais apontam para fora do eixo das hélices e, para as folhas, para cima e para baixo do plano definido pelas fitas.

ESTRUTURA TERCIÁRIA

- Corresponde à montagem dos elementos de estrutura secundária;
- Os elementos de estrutura secundária se organizam em estrutura terciária através de um fenômeno denominado enovelamento;

ESTRUTURA TERCIÁRIA



- O enovelamento é o processo mais estudado para proteínas (principais efetores da informação gênica);
- Envolve aproximação mútua de resíduos hidrofóbicos, que ocasionam a expulsão deste solvente da região central da proteína (colapso hidrofóbico);
- Influenciado por interações covalentes, associadas a modificações co- ou pós-traducionais (a mais abundante na natureza é a glicosilação de proteínas).

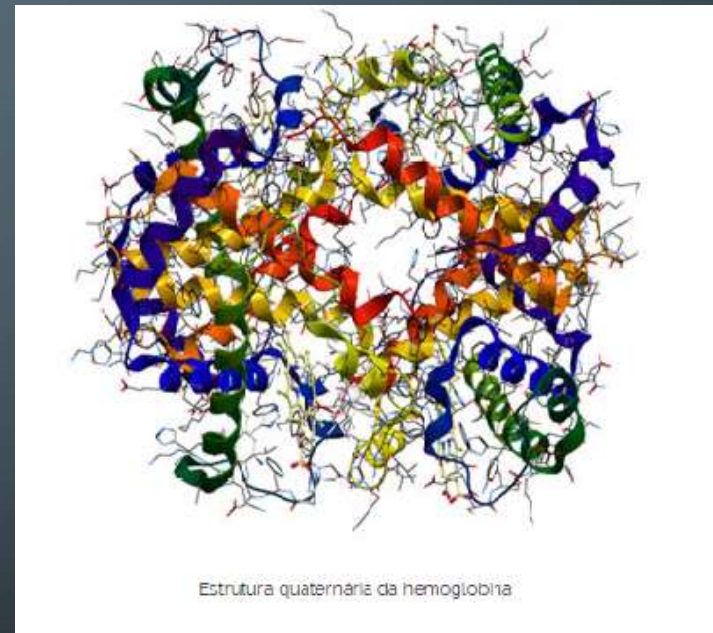
ESTRUTURA QUARTENÁRIA

- Constituída por agregados macromoleculares, principalmente de proteínas;
- Podem adotar estado oligomérico para a realização de determinada função em condições nativas;
- Dímeros (2), trímeros (3), tetrâmeros (4), pentâmeros (5), hexâmeros (6) ou mais.

ESTRUTURA QUARTENÁRIA

Em ácidos nucleicos a estrutura quaternária pode ser observada em complexos entre DNA e proteínas;

O estado oligomérico de proteínas pode ser afetado por diferentes ambientes fisiológicos;

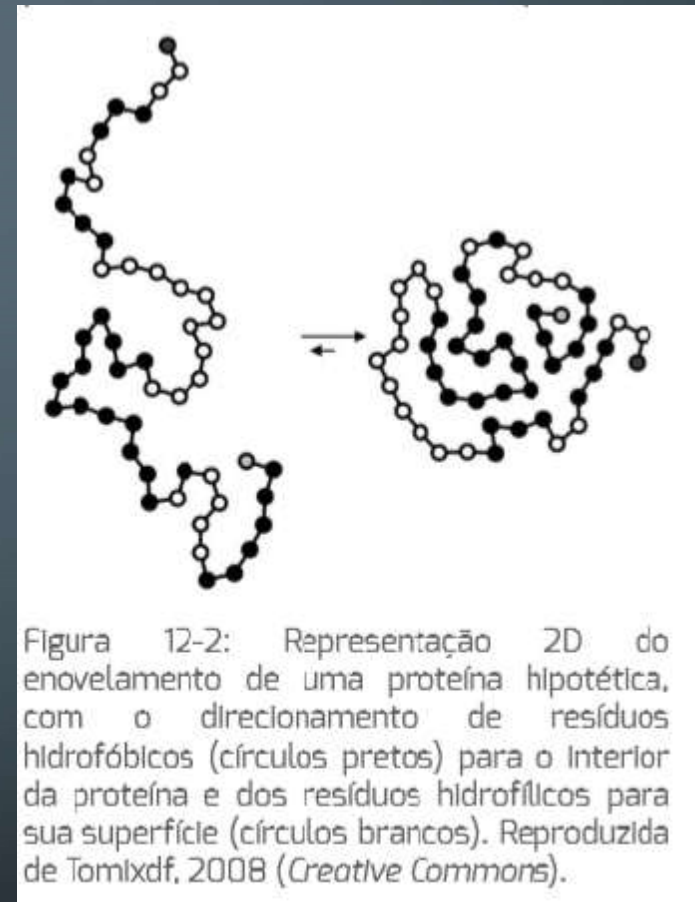


DESCRITORES DE FORMA

- O uso de conceitos de níveis hierárquicos oferecem definições qualitativas, que não abordam nuances ou variações dentro dos níveis;
- Há a necessidade de introduzir medidas quantitativas da forma destes compostos;
- Desta forma é possível calcular precisamente formas associadas a determinados eventos biológicos (como a regulação da expressão gênica) e por conseguinte interferir de forma racional no desenho de novos fármacos capazes de inibirem a expressão deste gene.

As formas tridimensionais dos biopolímeros são definidas pela conectividade entre seus monômeros constituintes (AA, monossacarídeos e bases nitrogenadas).

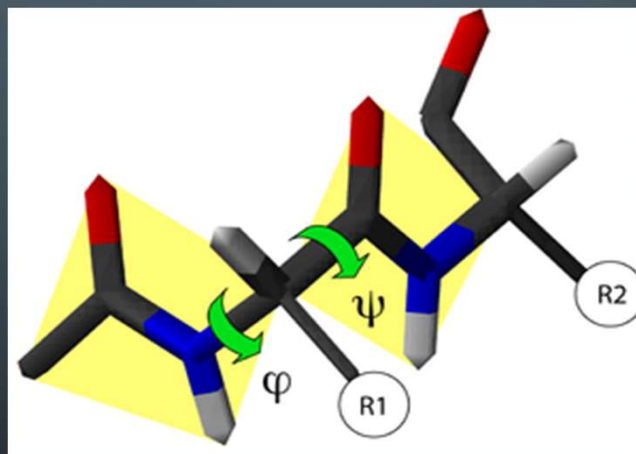
Na cadeia polipeptídica a definição do enovelamento da proteína depende dos ângulos de torção anterior ao carbono alfa, chamado de ângulo ϕ , e do ângulo de torção após o carbono alfa, chamado ângulo ψ .



MAPA DE RAMACHANDRAN (1963)

- Forma de compreender a estrutura de biomacromoléculas;
- Descreve a forma de dois AA vizinhos como fruto dos ângulos de torção ao redor de $C\alpha$, denominados φ e ψ .

A análise da rotação desses ângulos levou à identificação de regiões permitidas, onde não há choques entre os átomos, e regiões não-permitidas, onde há choque entre os átomos;



- Graficando-se para cada resíduo de aminoácido o ângulo ϕ e ψ temos um diagrama bidimensional, onde as regiões permitidas e proibidas são identificáveis;
- Os eixos, ϕ e ψ , no diagrama de Ramachandran, variam de -180° a $+180^\circ$, as regiões no interior das áreas demarcadas no gráfico são regiões permitidas.

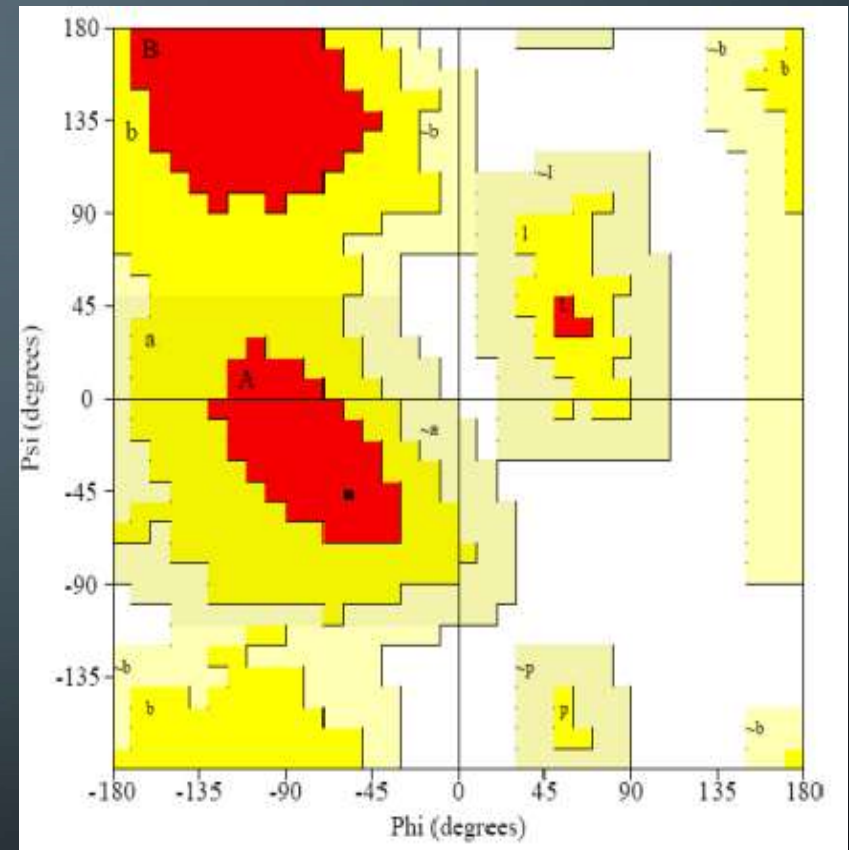
Em 1993 Laskowski e colaboradores elaboraram um programa chamado PROCHECK que define 4 regiões no diagrama de Ramachandran, são elas:

Região permitida (indicada em vermelho);

Região adicionalmente permitida (indicada em amarelo);

Região generosamente permitida (indicada em amarelo claro);

Região proibida (indicada em branco).



- Ângulos de torção também podem ser aplicados a ácidos nucleicos e carboidratos;
- O número de ângulos de torção são definidos pelas características das ligações entre monômeros (peptídica, glicosídica ou fosfodiéster).

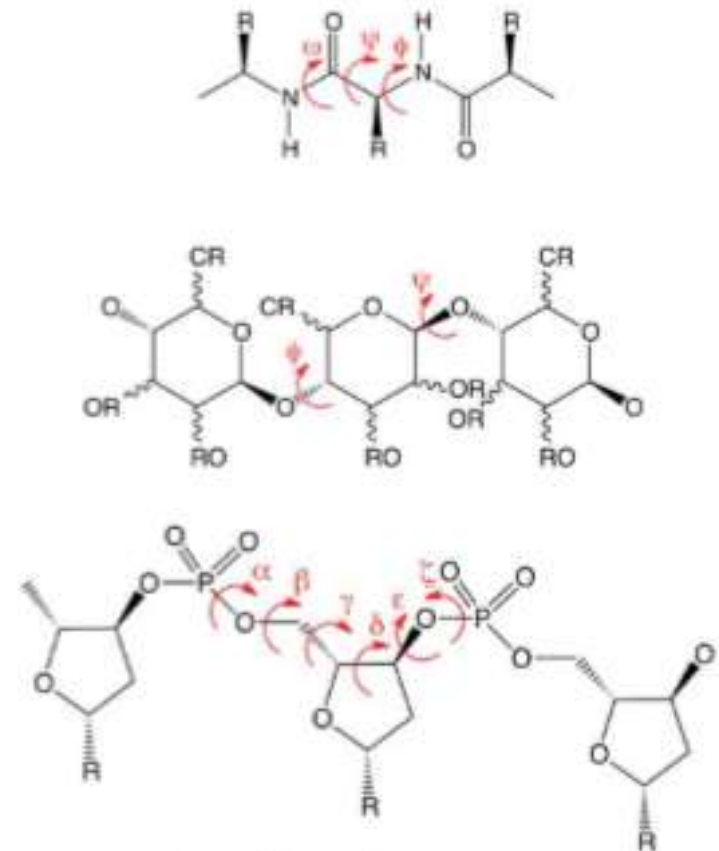


Figura 14-2: Ângulos torsionais para proteínas, carboidratos e ácidos nucleicos ilustrados para, respectivamente, um tripeptídeo, um trissacarídeo e um trinucleotídeo.

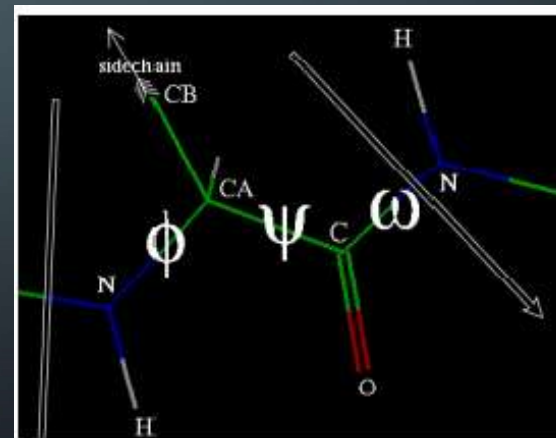
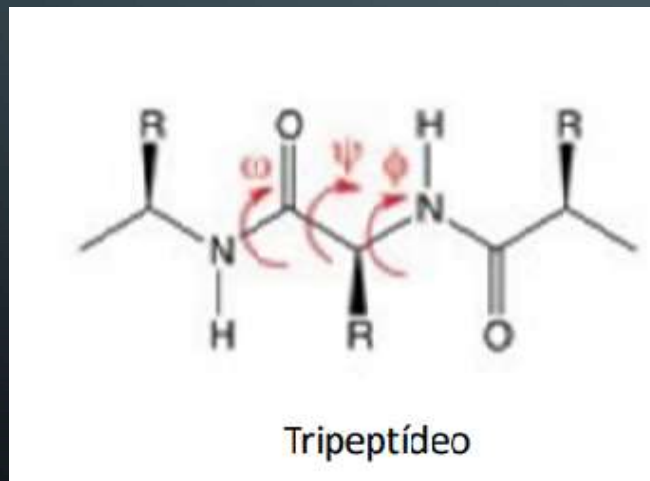
DESCRIÇÃO DE FORMA DE LIGAÇÃO PEPTÍDICA

Descrição da forma de ligação peptídica em proteína
(3 ângulos):

ψ : $C\alpha - C$

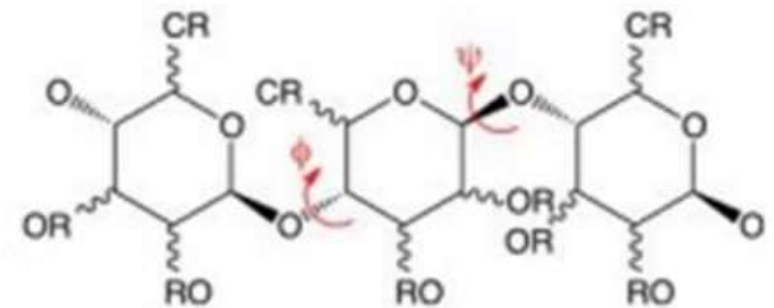
ϕ : $C\alpha - N$ e

ω (grupo amida): $N-H - C=O$.



LIGAÇÃO GLICOSÍDICA

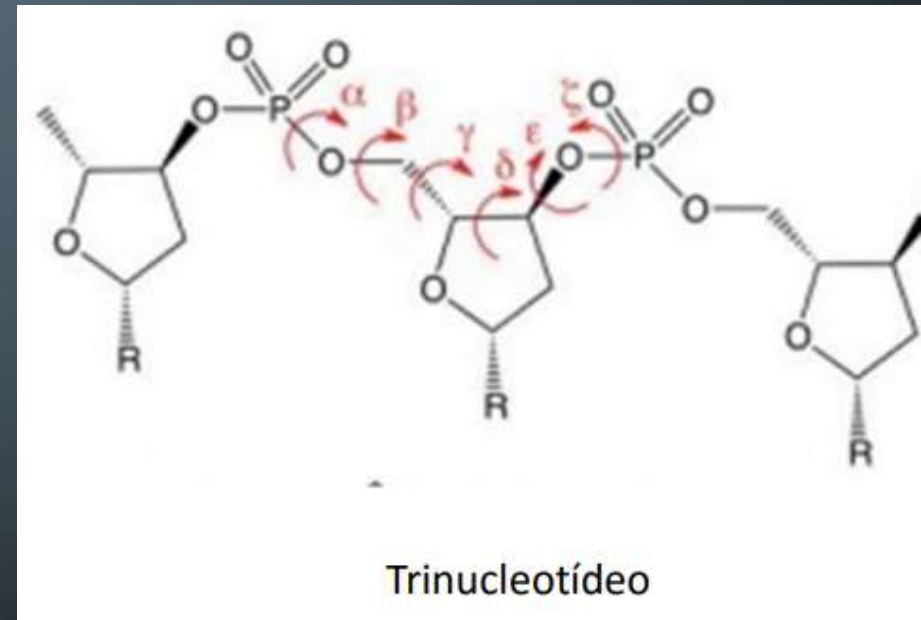
- Descrita por dois ou três ângulos torcionais;
- Dois ângulos : ψ ou φ
- Três ângulos torcionais (ω , quando o átomo de carbono está na posição 6 de piranoses ou na posição 5 de furanoses.

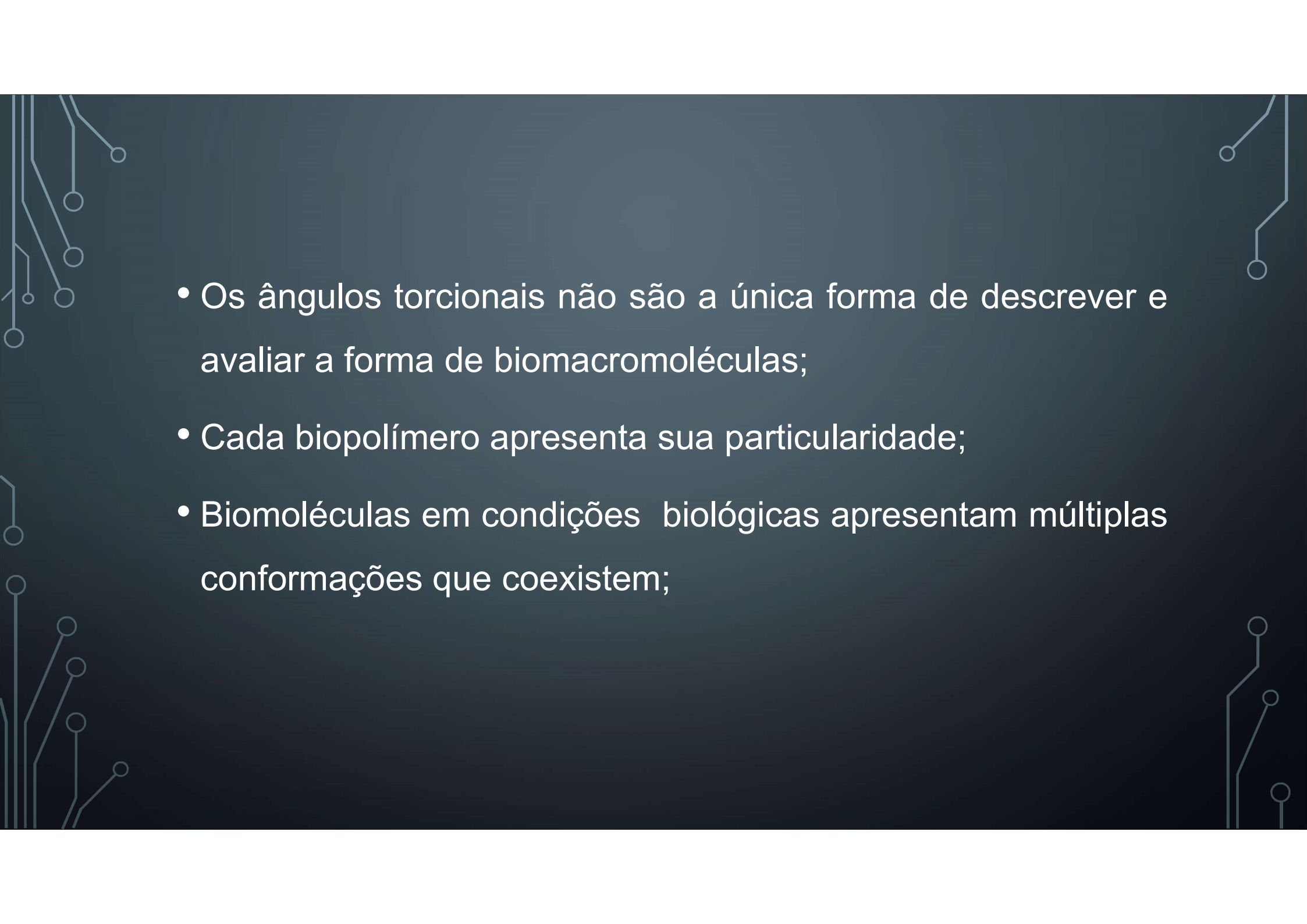


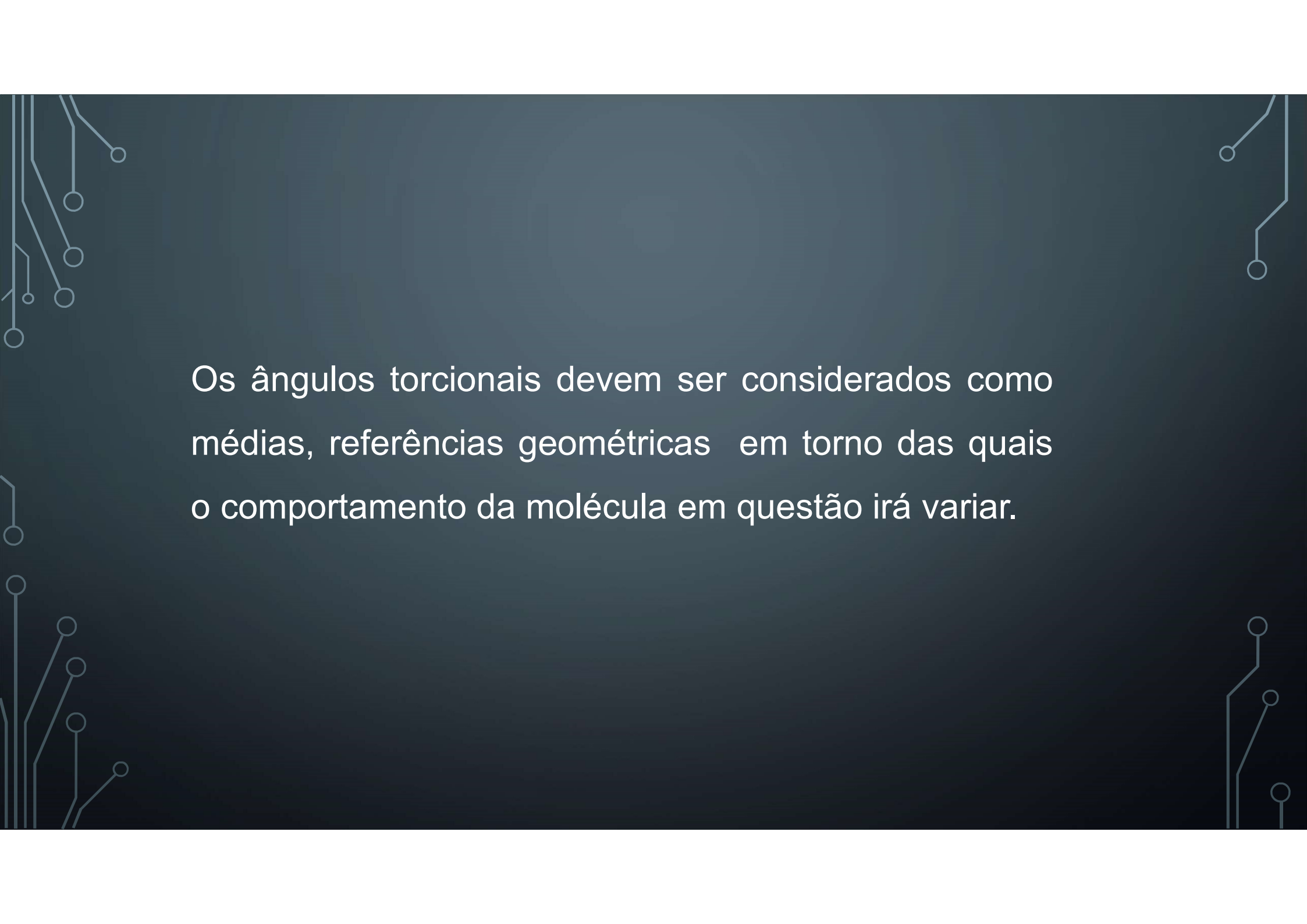
Trissacarídeo

ÁCIDOS NUCLEICOS

- O grupamento fosfato agrega grande flexibilidade;
- Sete ângulos torcionais: α , β , γ (na região 5'), δ (entre os átomos 3' e 4' da pentose), ϵ e ζ (na porção 3') e χ , formado entre o carbono 1' da pentose e a base nitrogenada



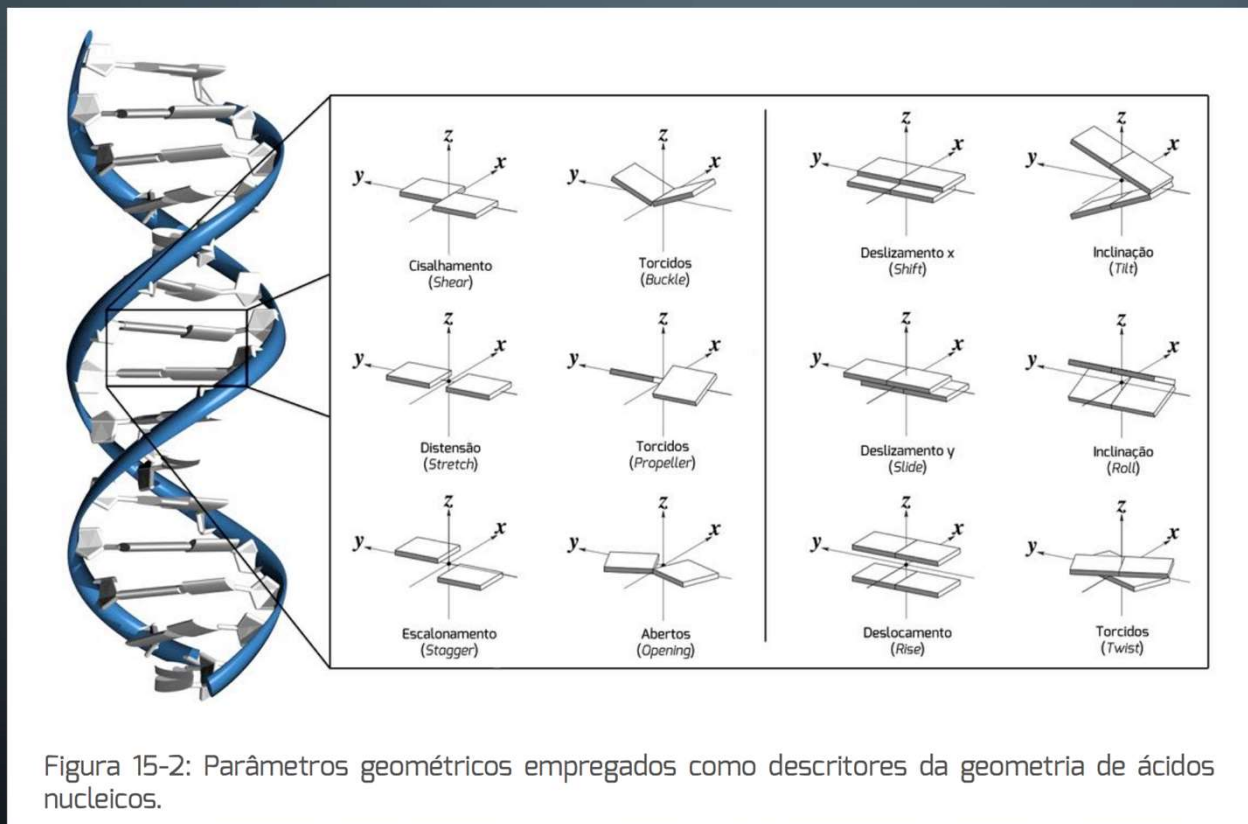
- 
- Os ângulos torcionais não são a única forma de descrever e avaliar a forma de biomacromoléculas;
 - Cada biopolímero apresenta sua particularidade;
 - Biomoléculas em condições biológicas apresentam múltiplas conformações que coexistem;

The background is a dark blue gradient. In the corners, there are decorative white lines that resemble circuit traces or molecular structures, with small circles at the ends of the lines.

Os ângulos torcionais devem ser considerados como médias, referências geométricas em torno das quais o comportamento da molécula em questão irá variar.

Descritores de Formas

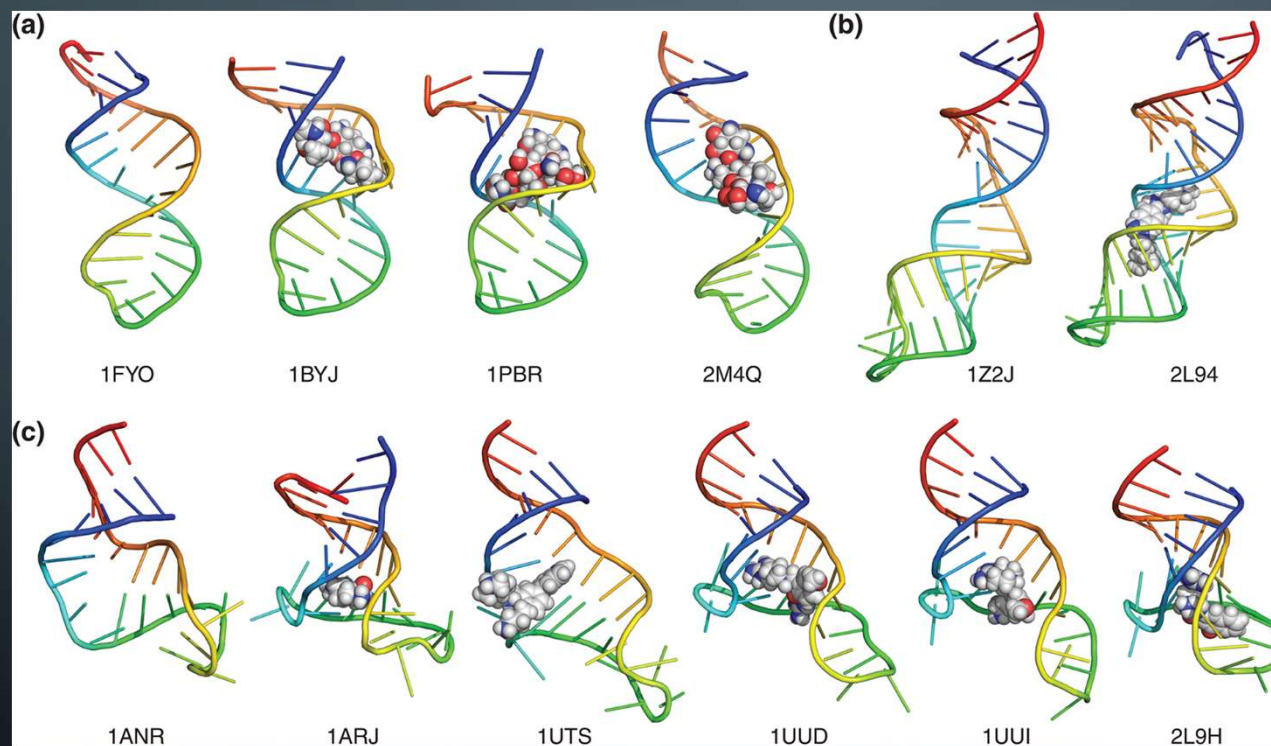
Ácidos Nucleicos



Influência dos pares
De bases na conformação
Adicional aos ângulos
torcionais

Descritores de Formas

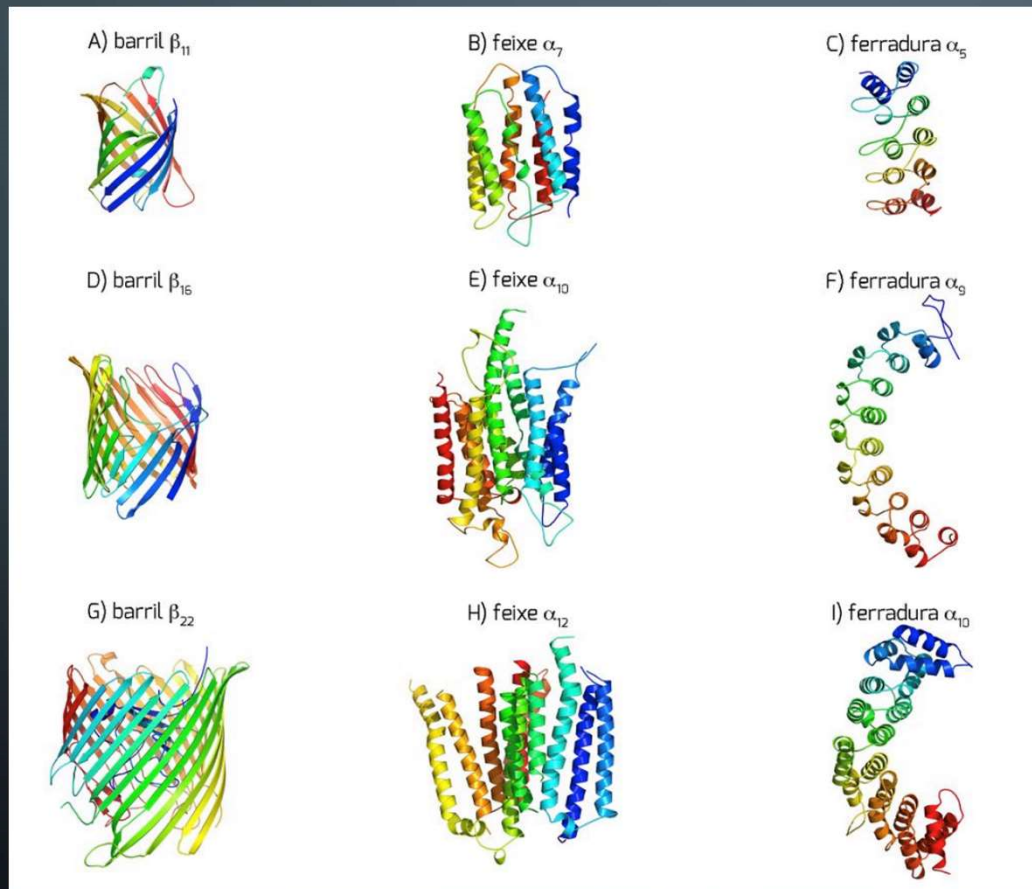
Ácidos Nucleicos



Influência dos pares
De bases na conformação
Adicional aos ângulos
torcionais

Descritores de Formas

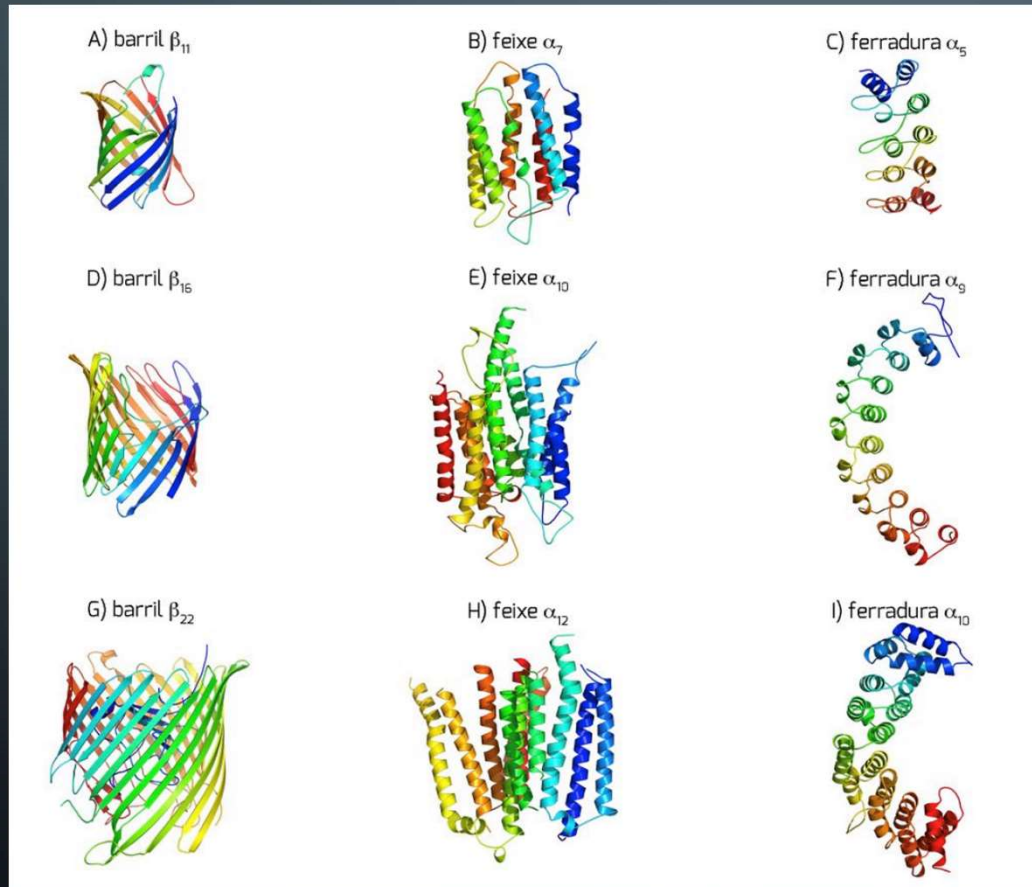
Proteínas



A complexidade esperada para a descrição
De formas das proteínas, dado o N de
Aminoácidos elevado e portanto de
Suas análises combinatórias não se realiza
Na prática

Descritores de Formas

Proteínas



Isso em virtude de fatores tais como
Determinantes estruturais e semelhanças
Físico-químicas, entre outras, que acabam
Determinando conjuntos de formas
Tridimensionais mais ou menos definidos
E finitos

Tais conjuntos de formas são agrupados
Na denominação de motivos (folds)

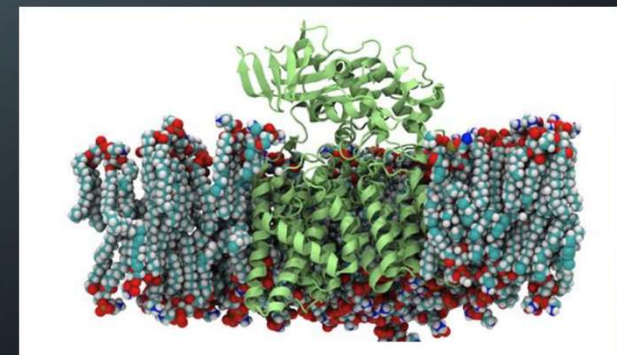
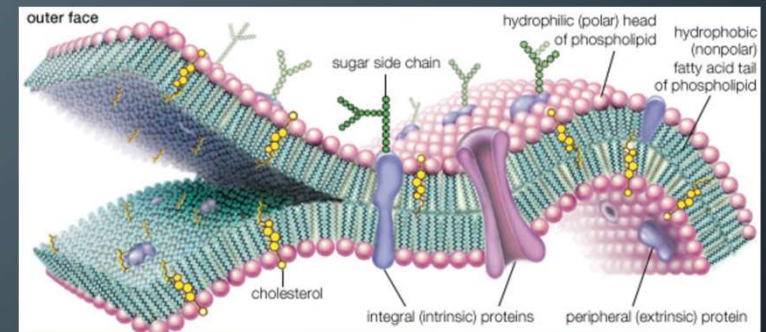
Descritores de Formas

Membranas

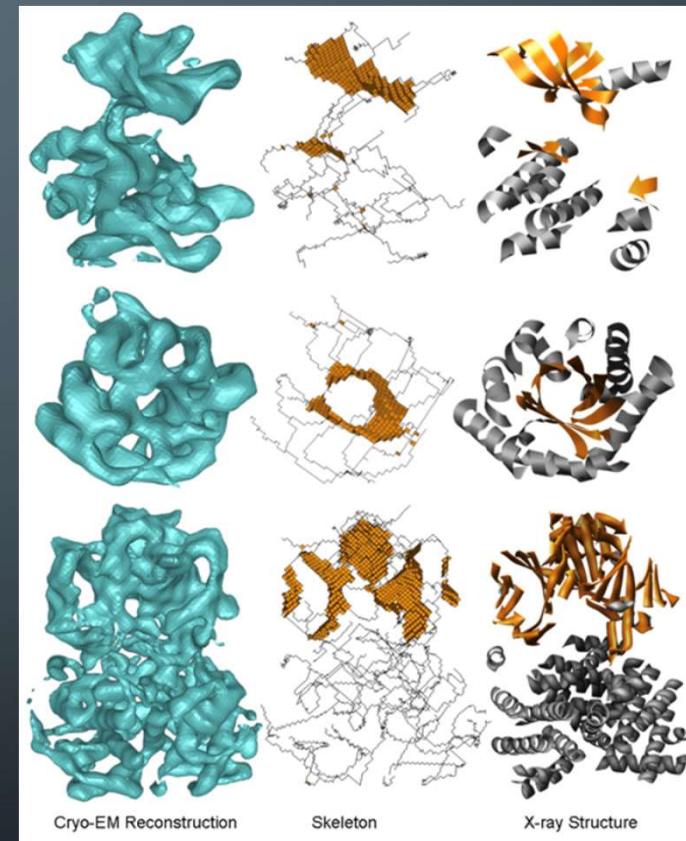
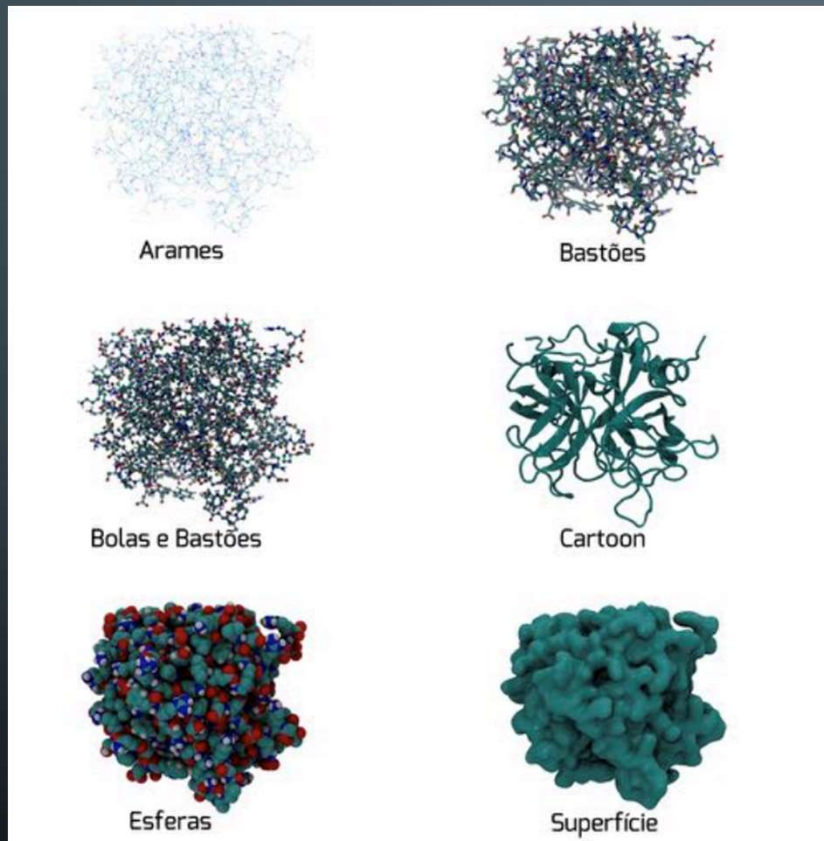
As membranas não são biopolímeros, portanto diferem
Dos itens estudados até o momento.

Além disso, sua estrutura/forma é fluida.

Assim, para o estudo das propriedades de membranas
Biológicas torna-se necessário caracterizá-las
estruturalmente. Isto pode ser feito através de diversas
medidas, tais como a área por lipídeo, espessura da
Membrana e coeficientes de difusão lateral de lipídeos ou
proteínas embebidas na membrana, dentre outros

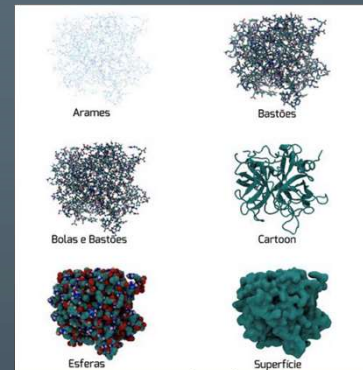


Formas de Visualização

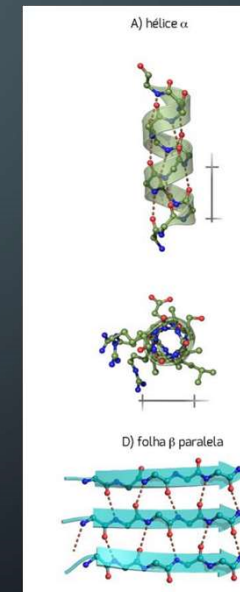


Formas de Visualização

DNA:
GGTATAGGCGCTGTTCTTAAGGTGCTAACAACGGGGT
TACCCGCGTTGATCTCGTGGATAAAACGCAAACGCCA
ACAG



Formas de visualização, contudo, são representações muitas vezes incapazes de descreverem detalhes sobre a molécula em estudo. É difícil distinguir visualmente uma hélice α de uma hélice 3_{10} ou de uma hélice π . Por outro lado, estas hélices podem apresentar deformações importantes, também de difícil visualização. Assim, a combinação de representações visuais, qualitativas, com medidas precisas, quantitativas, da estrutura molecular é uma estratégia bastante útil no estudo de macromoléculas.



Formas de Visualização

https://en.wikipedia.org/wiki/List_of_protein_structure_prediction_software

REFERÊNCIAS

- Imagem da árvore da vida:
<http://paraestudarbiologia.blogspot.com/2013/02/arvore-da-vida-2-anos.html>