

Universidade Estadual Paulista
“Júlio de Mesquita Filho”
Faculdade De Ciências – Campus de Bauru
Programa de Pós-Graduação em Ciência e Tecnologia de Materiais

WILLIAN ELISEU DA SILVA

Simulação Computacional da Estrutura Terciária de Proteínas Através de
Equações Paramétricas

BAURU

2016

Universidade Estadual Paulista
“Júlio de Mesquita Filho”
Faculdade De Ciências – Campus de Bauru
Programa de Pós-Graduação em Ciência e Tecnologia de Materiais

WILLIAN ELISEU DA SILVA

Simulação Computacional da Estrutura Terciária de Proteínas Através de
Equações Paramétricas

Dissertação apresentada como requisito à obtenção de Título de Mestre em Ciência e Tecnologia dos Materiais do Programa de Pós-Graduação em Ciência e Tecnologia dos Materiais, da Universidade Estadual Paulista “Júlio de Mesquita Filho” sob orientação do Prof. Dr. Aguinaldo Robinson de Souza.

BAURU

2016

FICHA CATALOGRÁFICA

Da Silva, Willian Eliseu.
Simulação Computacional da Estrutura Terciária
de Proteínas Através de Equações Paramétricas /
Willian Eliseu da Silva, 2016, 99 folhas.

Orientador: Aguinaldo Robinson de Souza

Dissertação (Mestrado) - Universidade Estadual
Paulista. Faculdade de Ciências, Bauru, 2016

1. Nós. 2. Proteínas. 3. Gráfico de distâncias.
I. Universidade Estadual Paulista. Faculdade de
Ciências. II. Título.



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Bauru



ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE Mestrado DE WILLIAN ELISEU DA SILVA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE MATERIAIS, DA FACULDADE DE CIÊNCIAS.

Aos 31 dias do mês de agosto do ano de 2016, às 14:00 horas, no(a) Sala 1-Seção Téc de Pós-Graduação Faculdade de Ciências/Bauru, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. AGUINALDO ROBINSON DE SOUZA - Orientador(a) do(a) Departamento de Química / Faculdade de Ciências - UNESP/Bauru, Prof. Dr. VALDECIR FARIAS XIMENES do(a) Departamento de Química / Faculdade de Ciências de Bauru, Prof. Dr. ANTONIO CALIRI do(a) Faculdade de Ciências Farmacêuticas / Universidade de São Paulo, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE Mestrado de WILLIAN ELISEU DA SILVA, intitulada **Simulação Computacional da Estrutura Terciária de Proteínas Através de Equações Paramétricas**. Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: A P R O V A D O. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. AGUINALDO ROBINSON DE SOUZA

Prof. Dr. VALDECIR FARIAS XIMENES

Prof. Dr. ANTONIO CALIRI

À minha mãe.

AGRADECIMENTOS

Ao prof. Dr. Aguinaldo Robinson de Souza, pela orientação deste trabalho, pela paciência e pela amizade.

À minha mãe Onícia, pela credibilidade depositada, pelos conselhos, pelas broncas, pois tudo isso me fez ser o que sou hoje.

À minha mulher, Karla, pela companhia, conselhos e compreensão.

Ao programa desta faculdade e à CAPES, pelo apoio financeiro neste período, sem o qual não poderia ter me dedicado ao desenvolvimento da pesquisa.

Muito obrigado.

“No meio da dificuldade encontra-se a oportunidade.”

Albert Einstein

DA SILVA, W. E., “**Simulação Computacional da Estrutura Terciária de Proteínas Através de Equações Paramétricas**”. 2015. 73 páginas. Dissertação (Programa de Pós-Graduação em Ciência e Tecnologia de Materiais). Universidade Estadual Paulista, Bauru, 2016.

RESUMO

Proteínas são polímeros heterogêneos lineares essenciais a todos os organismos vivos. As proteínas apresentam diversas funções nos organismos, tais como: estruturação, catálise, síntese, transferências, entre tantas outras. Neste trabalho foram estudadas as proteínas que apresentam em suas estruturas nós do tipo 3_1 .

Foram calculadas as distâncias entre os carbonos alfa dessas proteínas e com estes valores foram traçados os gráficos de distâncias que permitem uma visualização geral da estrutura da proteína.

Foram realizadas alterações no nó 3_1 padrão para que este se ajustasse à estrutura real da proteína na região do nó.

A partir do gráfico de distâncias foi possível determinar a estrutura secundária presente na proteína, sendo que a alfa hélice apresenta oscilações de aproximadamente 5,5 angstroms.

Com os gráficos de distâncias das proteínas e dos nós matemáticos foi possível comprovar a presença do nó e sua proximidade com a equação proposta.

Palavras chaves: Nós, Proteínas, Gráfico de distâncias.

DA SILVA, W. E., “**Simulação Computacional da Estrutura Terciária de Proteínas Através de Equações Paramétricas**”. 2015. 73 Páginas. Dissertação (Programa de Pós-Graduação em Ciência e Tecnologia de Materiais). Universidade Estadual Paulista, Bauru, 2016.

ABSTRACT

Proteins are linear heterogeneous polymer essential to all living organisms. The proteins have different functions in organisms, such as: structuring, catalysis, synthesis, transfers, among many others. In this work were studied the proteins that presenting in their structure, type 3_1 knots.

The distances between the alpha carbons of those proteins were calculated and with these values were plotted distances graphs that allow general visualization of the protein structure.

Changes were made in the standard knot 3_1 so that it would fit the actual structure of the protein in the node region.

From the distances graph was possible to determine the secondary structure of the protein, wherein the alpha helix presents oscillations proximately 5.5 angstroms.

With distances graph from proteins and from mathematical knots was possible to prove the presence of the node and its proximity to the proposed equation.

Keywords: knots, Proteins, distances graph.

SUMÁRIO

1. INTRODUÇÃO.....	1
1.1. Proteínas.....	1
1.1.1. AMINOÁCIDOS.....	2
1.1.2. Peptídeos	8
1.1.3 Proteínas estudadas	13
1.1.4 Simulação computacional	20
1.2. Equações paramétricas.....	23
1.2.1 Equação da Reta	24
1.2.2 Equação da circunferência	25
1.2.3 Nós matemáticos.....	29
1.4. Softwares utilizados	34
1.4.1 Rasmol	34
1.4.2 Geogebra.....	36
2. OBJETIVOS	38
3. MATERIAIS E MÉTODOS	39
3.1 Materiais	39
3.2 Métodos.....	40
4. RESULTADOS E DISCUSSÃO	44
5. CONCLUSÕES	83
6. REFERÊNCIAS	85

LISTA DE FIGURAS

Figura 1: Convenções para classificação dos átomos de carbono presentes em um aminoácido.	4
Figura 2: Aminoácidos com grupos R apolares.....	4
Figura 3: Aminoácidos com grupos R aromáticos.....	5
Figura 4: Aminoácidos com grupos R polares - não carregados.....	5
Figura 5: Aminoácidos com grupos R carregados positivamente.....	6
Figura 6: Aminoácidos com grupos R carregados negativamente.....	6
Figura 7: Representação dos vinte aminoácidos comuns no diagrama de Venn.	7
Figura 8: Estrutura primária. Sequenciamento de resíduos de aminoácidos (Anidrase Carbônica 1 - Humana).	9
Figura 9: Hélices alfa e Folhas Beta. Cadeia A da proteína hipotética da <i>Methanocaldococcus jannaschii</i> (Código PDB 2EFV).	9
Figura 10: Estrutura terciária da Cadeia A da proteína hipotética 2EFV.	10
Figura 11: Estrutura quaternária: Hemoglobina.	10
Figura 12: Interações que estabilizam as estruturas terciárias e quaternárias.	11
Figura 13: Ligação Peptídica.	12
Figura 14: Anidrase carbônica II.	13
Figura 15: Metiltransferase.	15
Figura 16: Exemplo de reação catalizada pela metiltransferase.	16
Figura 17: Transcarbamilase.....	17
Figura 18: Reação catalisada pela Ornitina Transcarbamilase.	18
Figura 19: S-ADENOSYLMETHIONINE SYNTHETASE.	19
Figura 20: Modelo científico aplicado à simulação.....	21
Figura 21: Segmento de reta AB.	23
Figura 22: Gráfico da função $f(t) = (-2 + 3t, 1 + t)$	25
Figura 23: Representação Geométrica da parametrização da circunferência.	26
Figura 24: Equação paramétrica de uma mola.	27
Figura 25: Esfera.	28
Figura 26: Nó matemático tipo 3_1	29
Figura 27: Nós em proteínas.	31
Figura 28: Nós matemáticos.....	32
Figura 29: Interface do programa RasMol.....	35
Figura 30: Interface do software GeoGebra.....	36
Figura 31: Exemplo de construção com o GeoGebra.....	37
Figura 32: Cálculo de distâncias.....	40
Figura 33: Distância entre dois pontos.	41
Figura 34: Nó 3_1	44
Figura 35: Representação gráfica da equação alterada do nó 3_1	47
Figura 36: Passo 1 – etapa inicial.....	50
Figura 37: Passo 2 - início da abertura do nó.	50
Figura 38: Passo 3 - segundo passo da abertura do nó.....	51
Figura 39: Passo 4 - nó aberto.	51
Figura 40: (a) Representação simplificada da alfa hélice; (b) Representação da alfa hélice com carbonos alfa.	55
Figura 41: Hélices alfa e 3-10.....	57
Figura 42: Sequência de resíduos e estruturas secundárias para a proteína Anidrase Carbônica.....	59

Figura 43: Diagrama de Venn para representação da polaridade da Anidrase carbônica na região do nó.	61
Figura 44: Sequência de resíduos e estruturas secundárias para a proteína Metiltransferase.	64
Figura 45: Diagrama de Venn para representação da polaridade da Metiltransferase.....	65
Figura 46: Sequência de resíduos e estruturas secundárias para a proteína Transcarbamilase.....	68
Figura 47: Diagrama de Venn para representação da polaridade da Transcarbamilase.	69
Figura 48: Sequência de resíduos e estruturas secundárias para a proteína S - Adenosilmetionina Sintetase.....	72
Figura 49: Diagrama de Venn para representação da polaridade da S - Adenosilmetionina Sintetase.	73
Figura 50: Sequência de resíduos e estruturas secundárias para a proteína extraída do escorpião (TsK).	75
Figura 51: Diagrama de Venn para representação da polaridade da TsK.	76
Figura 52: Sequência de resíduos e estruturas secundárias para a proteína hipotética do Methanocaldococcus jannaschii.....	79
Figura 53: Diagrama de Venn para representação da polaridade da proteína hipotética do Methanocaldococcus jannaschii.....	80
Figura 54: Sequência de resíduos e estruturas secundárias da proteína carboxil esterase.....	82

LISTA DE ESQUEMAS

Esquema 1: a) Reação de polimerização do etilenoglicol com ácido tereftálico; b) Reação de condensação entre dois aminoácidos, a ligação amídica aparece em destaque.	8
Esquema 2: Ressonância presente nos aminoácidos.....	12

LISTA DE TABELAS

Tabela 1: Os vinte aminoácidos mais comuns.....	3
Tabela 2: Cálculos das coordenadas e das distâncias do nó 3_1 padrão.	45
Tabela 3: Cálculos das coordenadas e das distâncias do nó 3_1 alterado.	48
Tabela 4: Cálculo da maior distância para uma alfa hélice.	56
Tabela 5: Dados das proteínas analisadas.....	58
Tabela 6: Polaridade da região do nó da Anidrase Carbônica.....	60
Tabela 7: Polaridade da estrutura completa da anidrase carbônica.	60
Tabela 8: Distribuição de resíduos na Anidrase Carbônica na região do nó.	61
Tabela 9: Polaridade da região do nó da Metiltransferase.	65
Tabela 10: Polaridade da estrutura completa da metiltransferase.	65
Tabela 11: Distribuição de resíduos na Metiltransferase na região do nó.	65
Tabela 12: Polaridade da região do nó da transcarbamilase.	69
Tabela 13: Polaridade da estrutura completa da proteína transcarbamilase.....	69
Tabela 14: Distribuição de resíduos na Transcarbamilase.	69
Tabela 15: Polaridade da região do nó da S - Adenosilmetionina Sintetase.....	72
Tabela 16: Polaridade da estrutura completa da proteína s-adenosilmetionina sintetase.....	73
Tabela 17: Distribuição de resíduos na S - Adenosilmetionina Sintetase.	73
Tabela 18: Polaridade da região do nó da TsK.....	76
Tabela 19: Distribuição de resíduos na proteína TsK.....	76
Tabela 20: Polaridade da região do nó da proteína hipotética do Methanocaldococcus jannaschii. 79	
Tabela 21: Distribuição de resíduos na proteína hipotética do Methanocaldococcus jannaschii.	80

1. INTRODUÇÃO

1.1. Proteínas

A funcionalidade das proteínas está estritamente ligada à sua estrutura, com apenas 20 aminoácidos elas se organizam em variados arranjos sequenciais e espaciais para se dividir em classes que são distribuídos em: reguladores, bombas, motores, receptores, tradutores, armazenadores, transportadores, catalizadores entre muitos outros (Lehninger, Nelson, & Cox, 2008).

A estrutura das proteínas está dividida em quatro níveis: estrutura primária, estrutura secundária, estrutura terciária e estrutura quaternária. A estrutura primária das proteínas é formada pelo sequenciamento dos 20 aminoácidos e a quantidade de sequências possíveis pode ser calculada com base na análise da quantidade de resíduos. Por exemplo em uma estrutura com 40 resíduos existem 20^{40} sequências possíveis, isso quer dizer, existem aproximadamente $1,10 \times 10^{52}$ possibilidades para a sequência de aminoácidos nesta estrutura (Voet, Voet, & Pratt, 2008). A estrutura secundária das proteínas é composta de hélices, folhas pregueadas e voltas; alguns exemplos de hélices são: α – hélice (com 3,6 resíduos por volta), π – hélice (com 4,4 resíduos por volta) e γ – hélice (com 5,1 resíduos por volta) (Donohue, 1953). A estrutura terciária de uma proteína descreve o dobramento dos elementos estruturais secundários e especifica as posições de cada átomo na proteína. A estrutura quaternária é obtida pela união de unidades menores de polipeptídios, normalmente as estruturas com mais de 100 kD de massa são constituídas por mais de uma cadeia polipeptídica. Essas subunidades se associam com uma geometria específica e conferem, normalmente, uma nova função para a estrutura.

1.1.1. AMINOÁCIDOS

As funções orgânicas dificilmente são encontradas isoladas na natureza. Elas estão normalmente misturas nas moléculas naturais, e é esta combinação de funções que confere as diferenças nas propriedades físicas e químicas das moléculas naturais. Dentre estas combinações destacam-se os aminoácidos que além de outras funções, apresentam como componentes principais as funções amina e ácido carboxílico. Vinte aminoácidos são mais comuns em proteínas, o primeiro a ser descoberto foi a asparagina, em 1806, quando foi isolada do aspargo pelos químicos *Louis Nicolas Vauquelin* e *Pierre Jean Robiquet*. O último aminoácido descoberto foi a treonina, que foi identificada em 1938 por *William Cumming Rose*. Todos os aminoácidos possuem nomes simples e comuns, e normalmente os nomes derivam do material de onde foram isolados. O nome asparagina deriva de aspargo, a fonte de onde o aminoácido foi isolado, da mesma forma, a tirosina que vem do grego tyros, que significa queijo, e a glicina que também deriva do grego, no caso, glykos, que significa doce (CAMPBELL & FARRELL, 2006).

Todos os vinte aminoácidos comuns são α -aminoácidos. Sua estrutura é composta de um grupo carboxil, um grupo amino e um grupo lateral, ou grupo R. São os grupos laterais que diferem os aminoácidos pois variam em estrutura, tamanho e carga elétrica influenciando as propriedades físicas e químicas destes.

Além dos vinte aminoácidos comuns existem vários outros menos comuns que estão presentes nos organismos vivos (BELL, 1976). Estes podem ser formados por resíduos modificados de proteínas quando estas são sintetizadas, e entram na constituição das proteínas mesmo que em pequena proporção. Os aminoácidos comuns em proteínas são designados por abreviações de três letras e um símbolo, essa denominação pode ser observada na Tabela 1, as abreviações e símbolos são utilizados para representar de forma simplificada a composição e o sequenciamento dos aminoácidos nas proteínas.

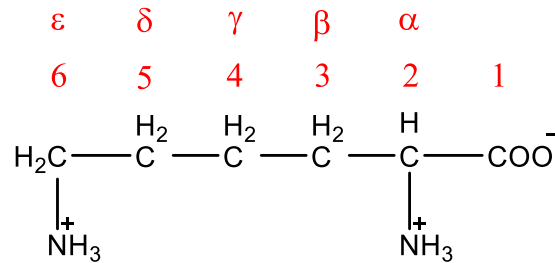
Tabela 1: Os vinte aminoácidos mais comuns

Aminoácido	Abreviação/ Símbolo	Ocorrência em proteínas (%)
Glicina	Gly G	7,2
Alanina	Ala A	7,8
Prolina	Pro P	5,2
Valina	Val V	6,6
Leucina	Leu L	9,1
Isoleucina	Ile I	5,3
Metionina	Met M	2,3
Fenilalanina	Phe F	3,9
Tirosina	Tyr Y	3,2
Triptofano	Trp W	1,4
Serina	Ser S	6,8
Treonina	Thr T	5,9
Cisteína	Cys C	1,9
Asparagina	Asn N	4,3
Glutamina	Gln Q	4,2
Lisina	Lys K	5,9
Histidina	His H	2,3
Arginina	Arg R	5,1
Aspartato	Asp D	5,3
Glutamato	Glu E	6,3

Fonte: (Campbell & Farrell, 2006).

Todos os aminoácidos comuns, com exceção da glicina, apresentam quatro substituintes diferentes no carbono α , este fato faz com que este carbono seja um centro quiral e os aminoácidos apresentem atividade ótica. Para dar nome aos aminoácidos podem ser utilizadas duas formas, uma numérica e a outra com letras do alfabeto grego. Na convenção numérica os átomos são numerados a partir do carbono carboxílico sendo este o número um. Na segunda convenção utiliza-se letras do alfabeto grego, estas são adicionadas a partir do carbono quiral, isto é, o carbono que está ligado à carboxila e obedecem a seguinte sequência: α , β , γ , δ , ϵ ... etc. As duas convenções estão representadas na Figura 1.

Figura 1: Convenções para classificação dos átomos de carbono presentes em um aminoácido.



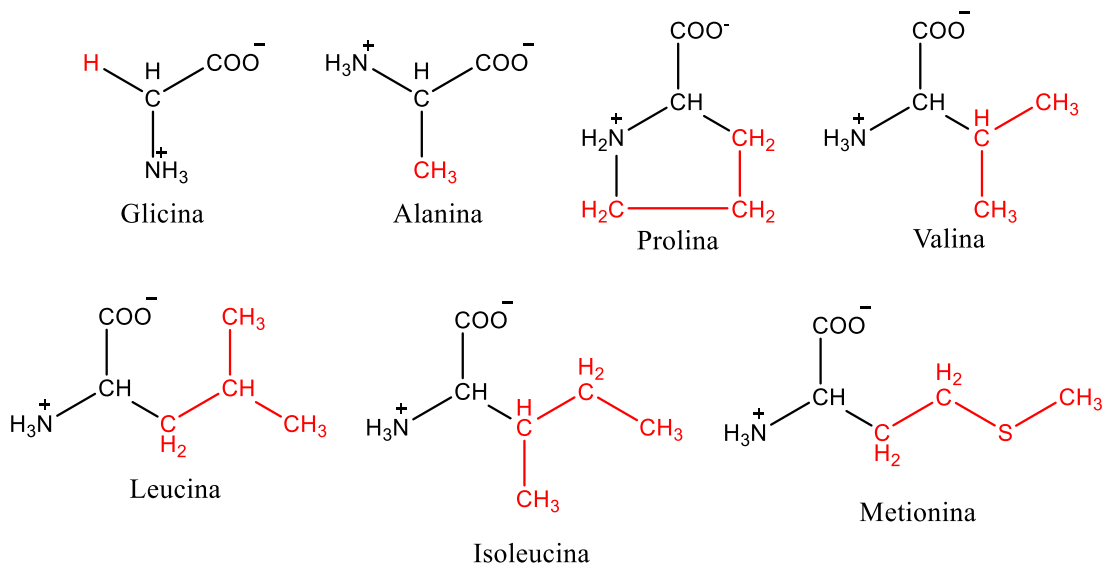
Lisina

Fonte: (Campbell & Farrell, 2006).

A convenção que utiliza letras do alfabeto grego é a mais comum. Pode-se a partir desta convenção classificar os aminoácidos em alfa-aminoácidos, beta-aminoácidos, etc. Os vinte aminoácidos mais comuns nas proteínas são todos alfa-aminoácidos.

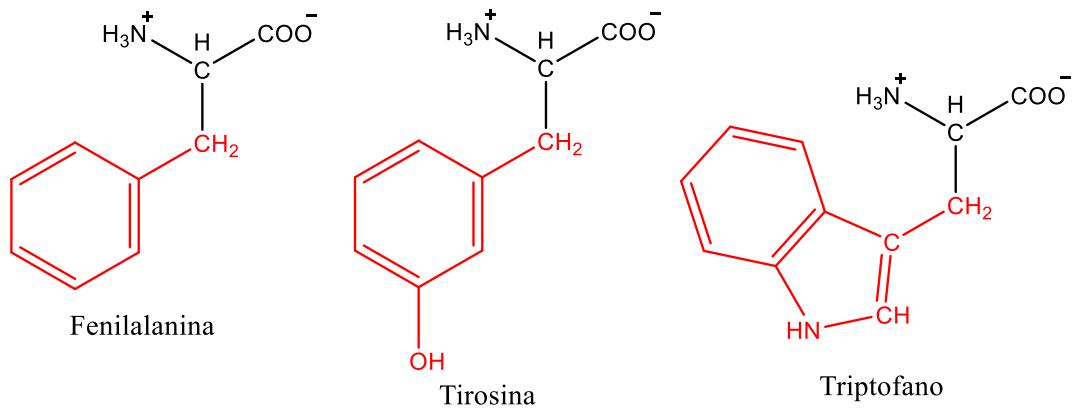
As estruturas moleculares em pH = 7 dos vinte aminoácidos são apresentadas nas Figuras 2, 3, 4, 5 e 6, eles estão separados de acordo com seu grupo R; nas figuras a região em destaque representa o grupo R.

Figura 2: Aminoácidos com grupos R apolares.



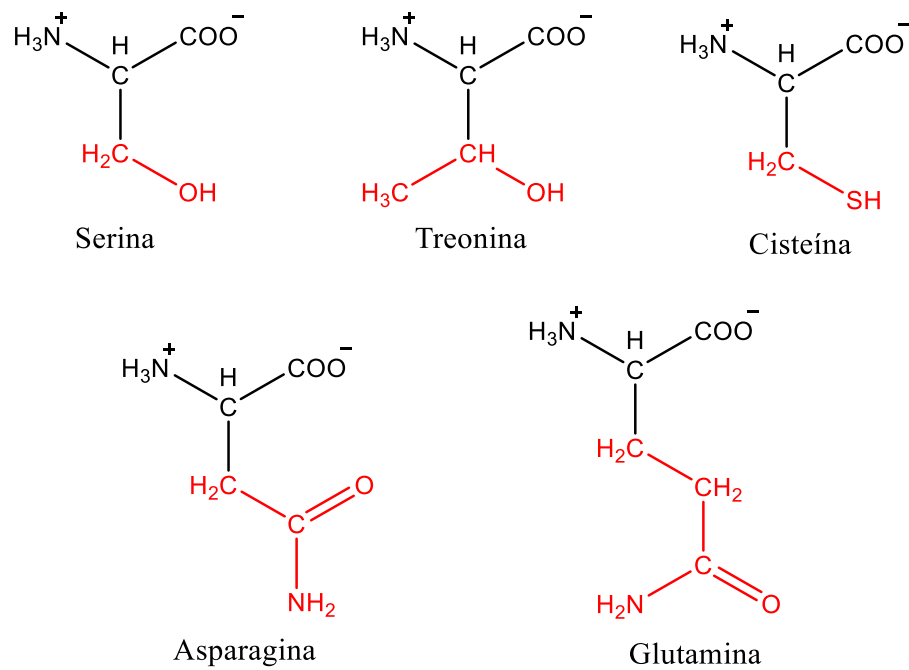
Fonte: (Campbell & Farrell, 2006).

Figura 3: Aminoácidos com grupos R aromáticos.



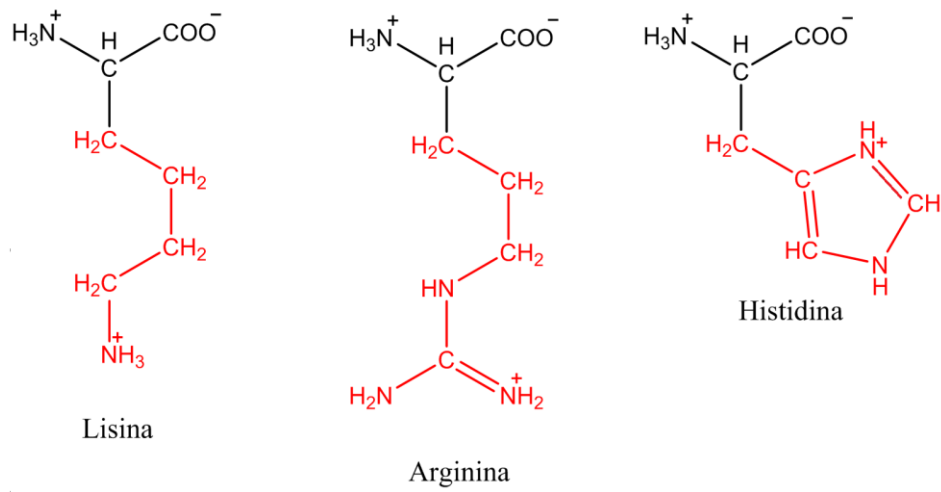
Fonte: (Campbell & Farrell, 2006).

Figura 4: Aminoácidos com grupos R polares - não carregados.



Fonte: (Campbell & Farrell, 2006).

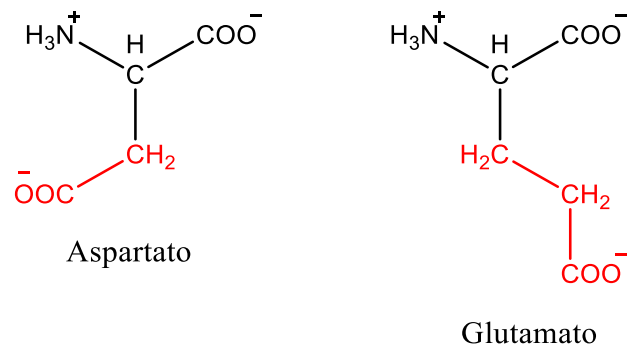
Figura 5: Aminoácidos com grupos R carregados positivamente.



Fonte: (Campbell & Farrell, 2006).

Figura 6: Aminoácidos com grupos R carregados negativamente.

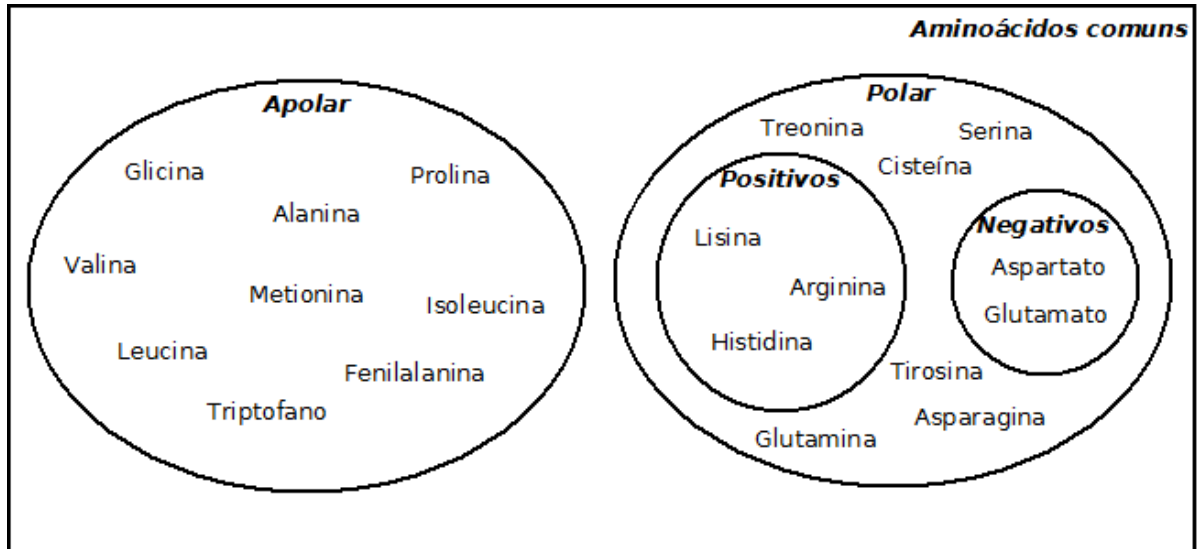
2 - e Grupos R carregados negativamente



Fonte: (Campbell & Farrell, 2006).

Para facilitar a visualização dos aminoácidos e suas características gerais é possível representá-los em um diagrama de Venn:

Figura 7: Representação dos vinte aminoácidos comuns no diagrama de Venn.

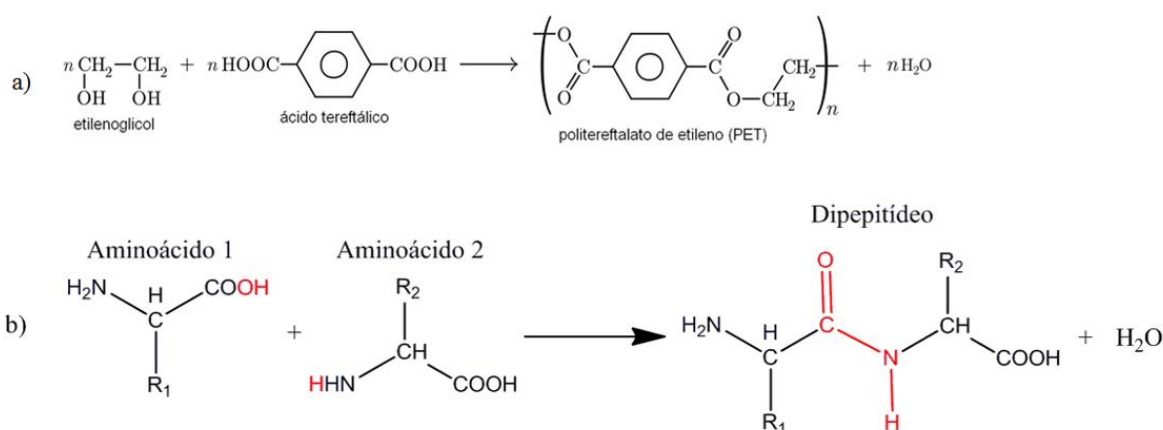


Fonte: Elaborado pelo autor.

1.1.2. Peptídeos

Duas moléculas de aminoácidos podem ser ligadas de forma covalente através da formação de ligação amídica, essa ligação é denominada ligação peptídica. Quando dois aminoácidos reagem através da ligação peptídica forma-se um dipeptídeo. Para a formação do peptídeo entre duas moléculas, é necessário remover uma molécula de água, isto é, ocorre uma reação de condensação. Especificamente a molécula de água é removida pela união do grupo ácido carboxílico de um aminoácido e do grupo amina do outro, de forma prática pode-se dizer que a reação de formação de proteínas é equivalente à reação de formação de polímeros (poliamidas), conforme mostra o Esquema 1. Três aminoácidos podem ser ligados para formar um tripeptídeo, da mesma forma quatro e cinco aminoácidos formam tetrapeptídeo e pentapeptídeo, respectivamente. Quando o número de moléculas é pequeno esse conjunto de peptídeos recebe o nome de oligopeptídeo, quando o número de aminoácidos é grande o termo a ser utilizado é polipeptídeo. Este último termo é, por vezes, utilizado para designar uma proteína, normalmente denominam-se polipeptídeos moléculas que tenham peso molecular abaixo de 10 kD (dez quilo Dalton). Qualquer polipeptídeo que tenha mais que 10 kD de peso molecular, recebe o nome de proteína.

Esquema 1: a) Reação de polimerização do etilenoglicol com ácido tereftálico; b) Reação de condensação entre dois aminoácidos, a ligação amídica aparece em destaque.



Fonte: Elaborado pelo autor.

Não se pode comparar a reação de condensação entre dois aminoácidos com qualquer reação de polimerização, uma vez que boa parte das reações de polimerização não eliminam água e nem unem cadeias através de ligações amídicas. A Figura 8 apresenta a estrutura primária, o sequenciamento de resíduos da proteína anidrase carbônica de código pdb 12CA.

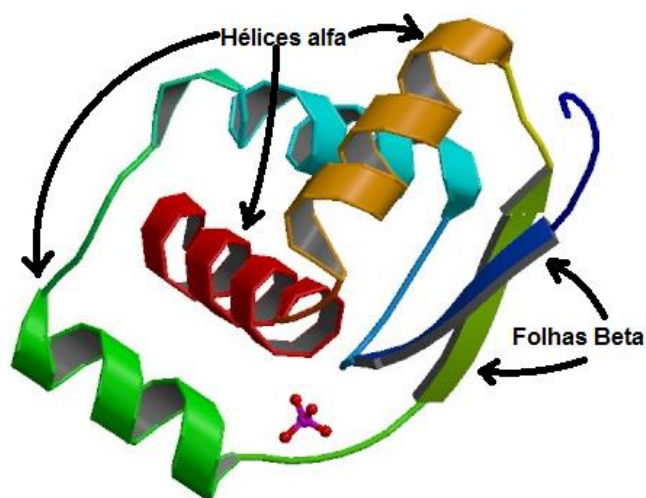
Figura 8: Estrutura primária. Sequenciamento de resíduos de aminoácidos (Anidrase Carbônica 1 - Humana).

10	20	30	40	50
MASPDWGYDD	KNGPEQWSKL	YPIANGNNQS	PVDIKTSETK	HDTSLKPISV
60	70	80	90	100
SYNPATAKEI	INVGHSEFVN	FEDNDNRSVL	KGGPFSDSYR	LFQFHFHWGS
110	120	130	140	150
TNEHGSEHTV	DGVKYSDELH	VAHWNSAKYS	SLAEAASKAD	GLAVIGVLMK
160	170	180	190	200
VGEANPKLQK	VLDALQAIKT	KGKRAPFTNF	DPSTLLPSSL	DFWTYPGSLT
210	220	230	240	250
HPPLYESVTW	IICKESISVS	SEQLAQFRSL	LSNVEGDNAV	PMQHNNRPTQ
260				
PLKGRTVRAS	F			

Fonte: Adaptado de <http://www.uniprot.org/uniprot/P00915>.

A estrutura secundária está exemplificada na Figura 9.

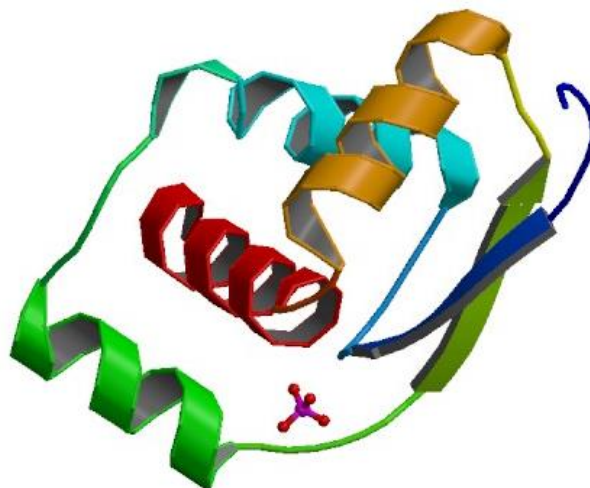
Figura 9: Hélices alfa e Folhas Beta. Cadeia A da proteína hipotética da *Methanocaldococcus jannaschii* (Código PDB 2EFV).



Fonte: Adaptado de <http://www.rcsb.org/pdb/explore/explore.do?structureId=2efv>.

A Figura 10, apresenta a estrutura terciária da proteína hipotética de código pdb 2EFV.

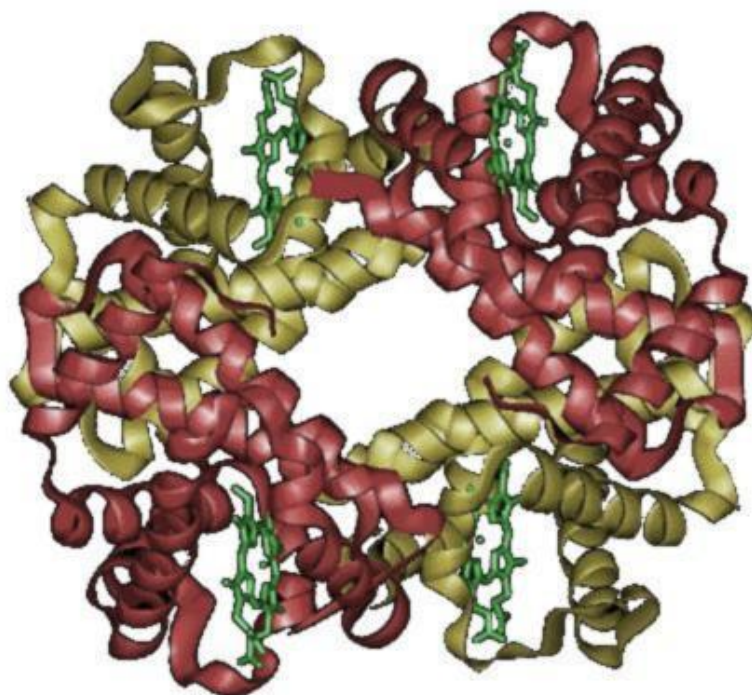
Figura 10: Estrutura terciária da Cadeia A da proteína hipotética 2EFV.



Fonte: <http://www.rcsb.org/pdb/explore/explore.do?structureId=2efv>.

A Figura 11, apresenta a estrutura quaternária da Hemoglobina, que é formada por quatro subunidades.

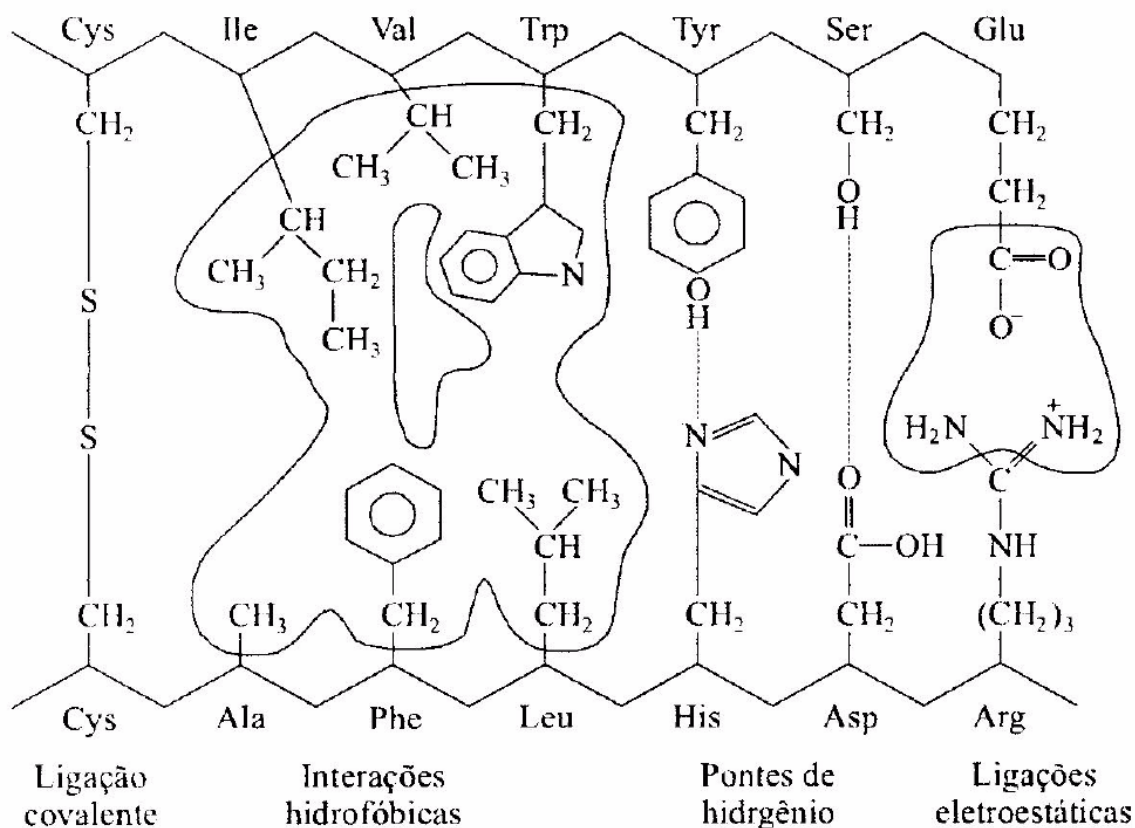
Figura 11: Estrutura quaternária: Hemoglobina.



Fonte: Google imagens.

As estruturas terciária e quaternária são obtidas e estabilizadas por interações que ocorrem ao longo da cadeia (estrutura terciária) e entre as cadeias (estrutura quaternária). Existem muitos tipos de interações químicas que podem ocorrer dentro de uma estrutura terciária, as mais fortes são as ligações covalentes, que ocorrem por exemplo entre dois resíduos de cisteína formando uma ponte de dissulfeto, outras interações são: pontes de hidrogênio, interações eletrostáticas e interações de van der waals entre os grupos R dos resíduos. Um exemplo deste tipo de interação pode ser observado na Figura 12, que apresenta um esquema que representa estas interações.

Figura 12: Interações que estabilizam as estruturas terciárias e quaternárias.



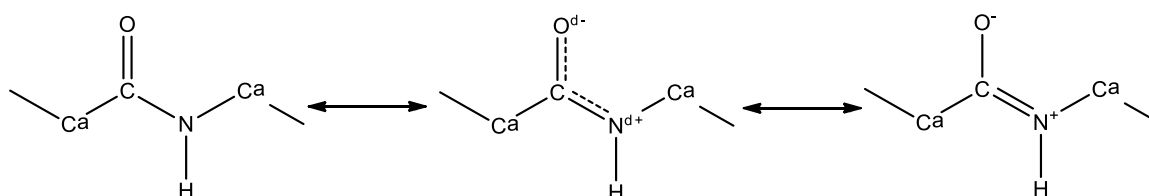
Fonte: Google imagens.

As estruturas quaternárias são formadas por interações entre as cadeias e outros compostos não proteicos. Basicamente a estrutura quaternária pode ser entendida como o arranjo não covalente formado por várias cadeias (estruturas terciárias) polipeptídicas.

O arranjo espacial dos átomos em uma proteína é chamado de conformação. As proteínas, girando em torno das ligações simples podem ter várias conformações, contanto que as ligações covalentes sejam mantidas, isto é, que não haja rompimento de nenhuma ligação em toda a estrutura da proteína. Há duas conformações comuns em proteínas: a conformação em Hélice (as Hélices podem ser: α -Hélice, 3_{10} -Hélice e π -Hélice) e em folha- β (DONOHUE, 1953), (PAULING, COREY & BRANSON, 1951). Essas conformações são mostradas na Figura 9.

Qualquer que seja a conformação que a proteína assuma ela terá apenas o carbono α para girar a estrutura, isto porque a ligação amídica torna a estrutura rígida e planar. Embora a ligação entre o carbono e o nitrogênio tenha a possibilidade de girar, esta opção é descartada devido às ressonâncias que ocorrem entre o oxigênio, o carbono e o nitrogênio, essas ressonâncias estão representadas no Esquema 2.

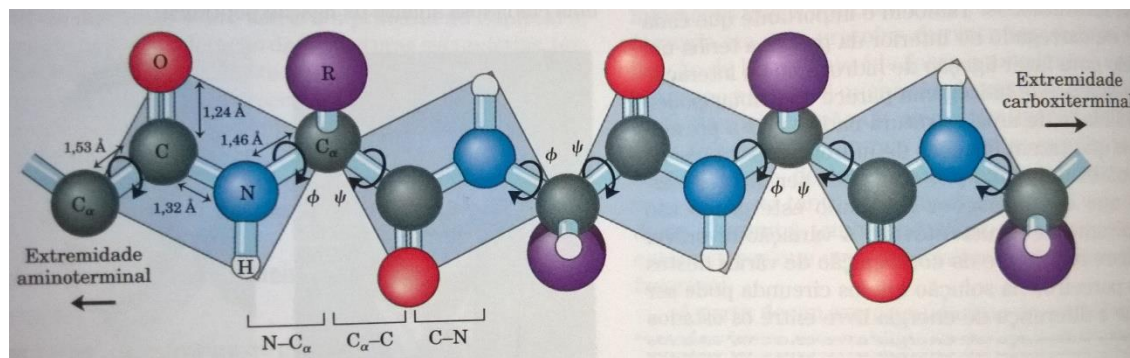
Esquema 2: Ressonância presente nos aminoácidos.



Fonte: Adaptado de (Campbell & Farrell, 2006).

Devido à ressonância, ocorre a formação de um plano inflexível que impede a rotação no Nitrogênio. Dessa forma a estrutura rotaciona apenas no carbono α . A Figura 13 mostra o plano que impede a rotação no Nitrogênio.

Figura 13: Ligação Peptídica.



Fonte: Mary K Campbell, Shawn O Farrell, 2006.

Como a estrutura gira somente no carbono α , as conformações assumidas pela proteína são estabilizadas por interações que ocorrem ao longo da cadeia e são essas interações que mantêm uma ou outra conformação. A conformação de uma proteína está intimamente ligada à sua função biológica, isto é, ela assume uma certa conformação para desempenhar uma determinada função.

1.1.3 Proteínas estudadas

Anidrase Carbônica

A anidrase carbônica é uma proteína que atua como uma enzima. Ela é responsável por catalisar a reação de formação do ácido carbônico a partir do gás carbônico e água. A catálise é realizada por intermédio do íon metálico Zn^{2+} , que tem papel crucial no sítio catalítico. A Figura 14 apresenta a estrutura da anidrase carbônica II, esta atua principalmente no citoplasma.

Figura 14: Anidrase carbônica II.



A reação catalisada depende do pH, a enzima atua acelerando a reação direta ou a reação inversa de modo que esta atinja o equilíbrio mais rapidamente. A anidrase carbônica catalisa a seguinte reação:



Essa reação é importante por dois motivos:

- Controlar o pH do meio, com o auxílio de outros tampões, mantendo em uma faixa considerada ótima que é de 7,4;
- Noventa por cento do CO_2 do sangue é transportado na forma de bicarbonato (ARAZAWA, et. al., 2015).

A variação na faixa de pH é normal no intervalo que vai de 7,38 a 7,42, valores abaixo ou acima caracterizam a acidose e a alcalose respectivamente.

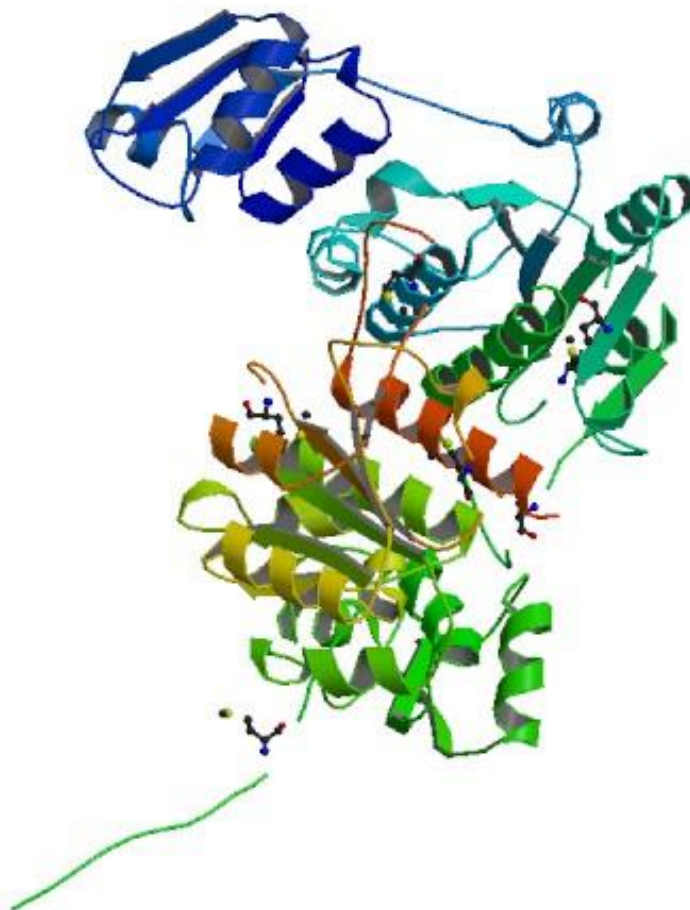
A anidrase carbônica apresenta os seguintes dados:

- Comprimento da cadeia: 255 resíduos;
- Tipo do nó: Nó 3₁
- Comprimento do nó: 228 resíduos;
- Posição do nó na cadeia: resíduo 30 a resíduo 258;

Metiltransferase

A metiltransferase é uma proteína que atua na transferência de grupos metila. A transferência pode ser para resíduos de aminoácidos ou diretamente no DNA (LOWENSON, et. al., 2001). Quando a transferência ocorre em resíduos de aminoácidos a função é a de reparar danos resultantes do desgaste natural das moléculas; quando a transferência ocorre para alguma posição do DNA, a função é a de atuar como ponte entre a estrutura do DNA e alguma proteína. A Figura 15 apresenta a proteína metiltransferase que é encontrada no organismo *Escherichia coli*.

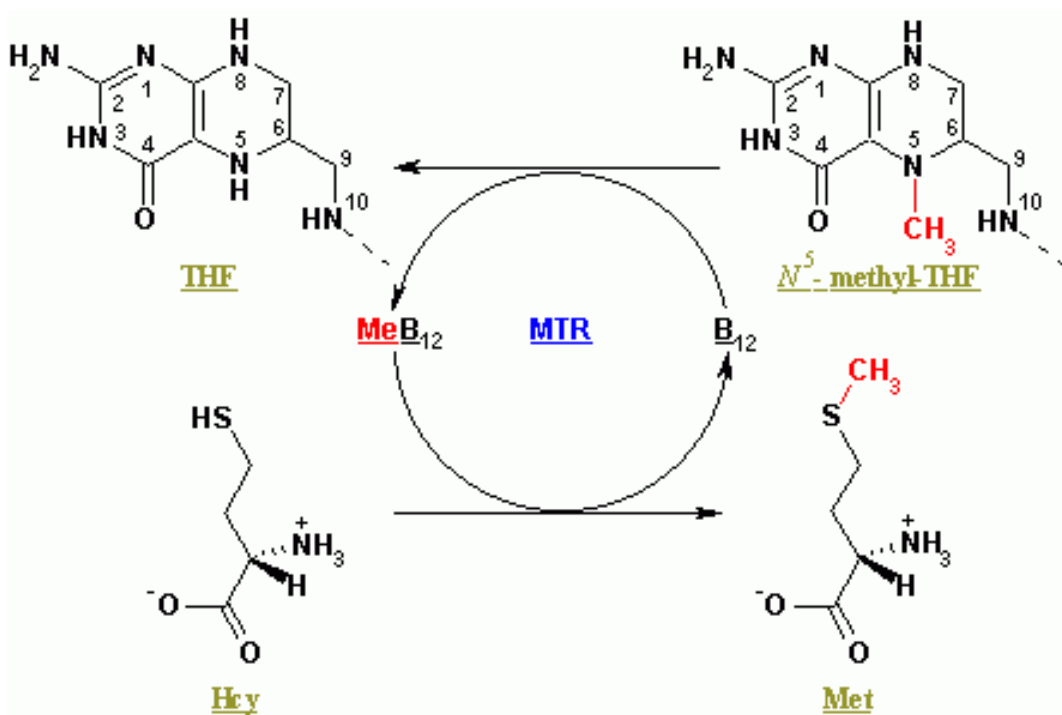
Figura 15: Metiltransferase.



Fonte: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1GZ0>.

A Figura 16 apresenta uma reação de metilação catalisada pela enzima metiltransferase, a transferência ocorre para a regeneração da metionina. A reação transforma metiltetrahydrofolato (mTHF) em tetrahydrofolato (THF), a transferência do grupo metil ocorre por intermédio da vitamina B12 que transporta o grupo metil do THF para a homocisteína (Hcy), neste processo ocorre a metilação da homocisteína regenerando assim a metionina.

Figura 16: Exemplo de reação catalizada pela metiltransferase.



Fonte: Google imagens.

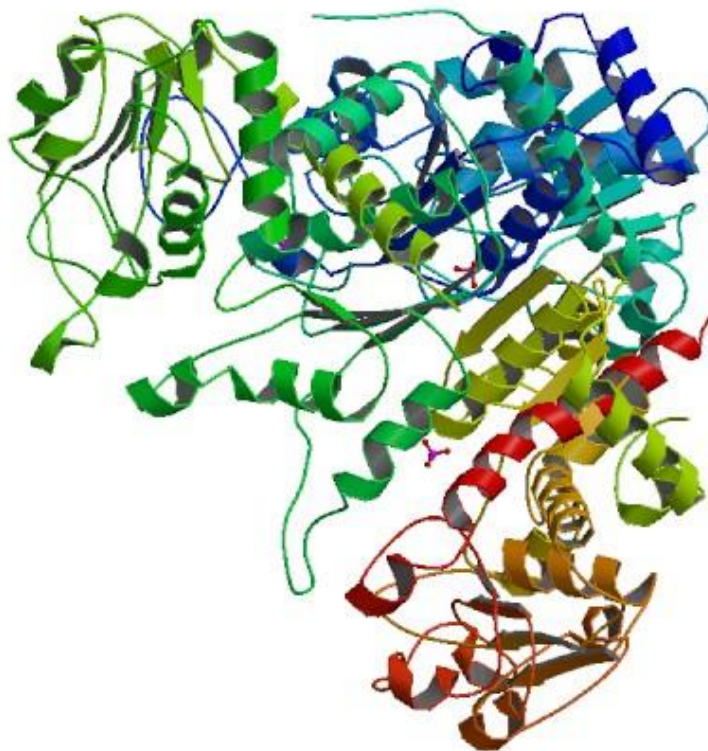
A metiltransferase apresenta os seguintes dados:

- Comprimento da cadeia: 235 resíduos;
- Tipo do nó: Nó 3₁
- Comprimento do nó: 41 resíduos;
- Posição do nó na cadeia: resíduo 173 a resíduo 214;

Transcarbamilase

As Transcarbamilases são uma família de enzimas que catalisam a transferência do grupo carbamoil, da estrutura do Carbamoil Fosfato para um grupo amino ou oxigênio de um outro substrato. As duas classes mais estudadas são a *ATCase* e a *OTCase* que são Asparato Transcarbamilase e *Ornitina* Transcarbamilase respectivamente (SHI, et. al., 2002). A Figura 17 apresenta a estrutura da Transcarbamilase.

Figura 17: Transcarbamilase.



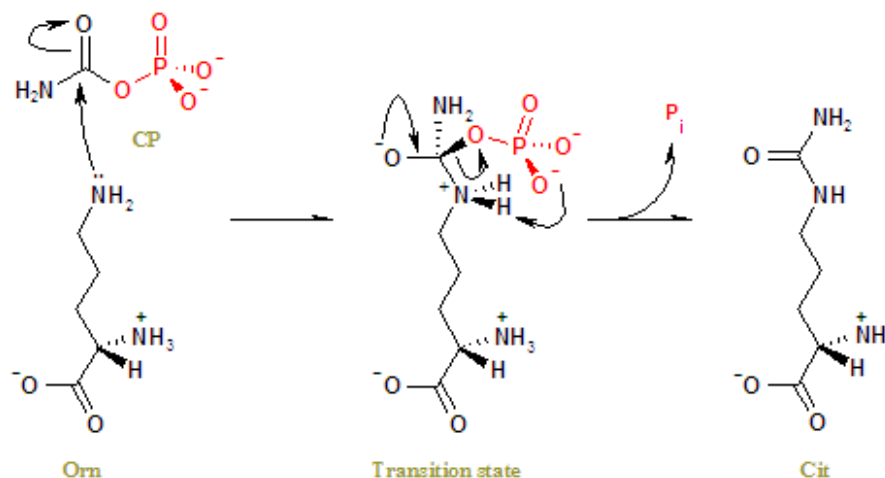
Fonte: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1JS1>.

A reação catalisada pela transcarbamilase é apresentada na Figura 18. A transcarbamilase catalisa a reação entre o carbamoil fosfato e a ornitina que tem como produtos o fosfato e a citrulina que faz parte do ciclo da uréia.

A transcarbamilase apresenta os seguintes dados:

- Comprimento da cadeia: 324 resíduos;
- Tipo do nó: Nó 3₁
- Comprimento do nó: 67 resíduos;
- Posição do nó na cadeia: resíduo 171 a resíduo 238;

Figura 18: Reação catalisada pela Ornitina Transcarbamilase.



Fonte: Google imagens.

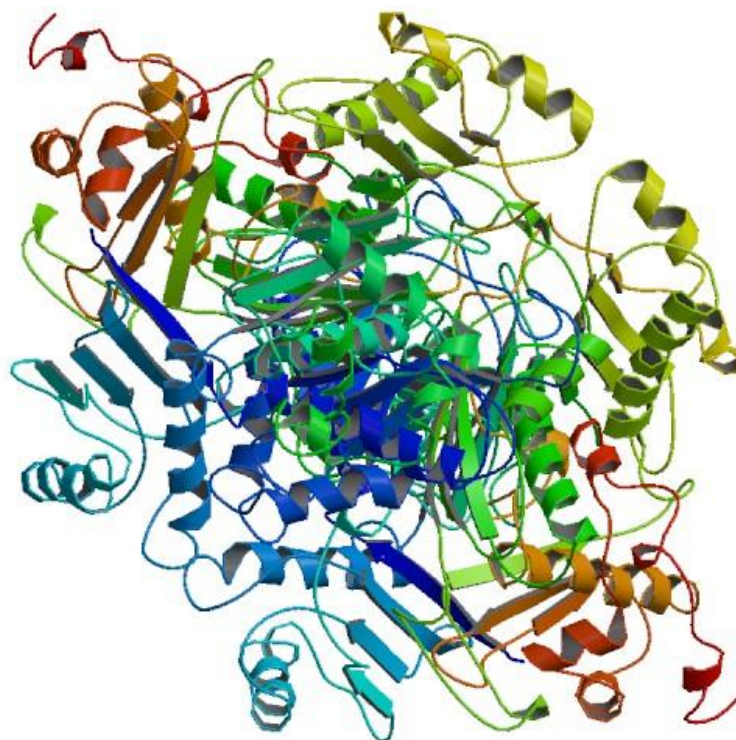
S - Adenosilmetionina Sintetase

A proteína s - adenosilmetionina sintetase é uma enzima responsável por catalisar a síntese da adenosilmetionina (AdoMet). A molécula de AdoMet é um doador de grupos metila que pode atuar tanto em conjunto com outras proteínas para metilação do DNA, quanto em conjunto com a metiltransferase na restauração de outras estruturas (FU, et. al., 1996). A Figura 19 apresenta a estrutura da s - adenosilmetionina sintetase.

A s - adenosilmetionina sintetase apresenta os seguintes dados:

- Comprimento da cadeia: 383 resíduos;
- Tipo do nó: Nó 3_1
- Comprimento do nó: 266 resíduos;
- Posição do nó na cadeia: resíduo 10 a resíduo 276;

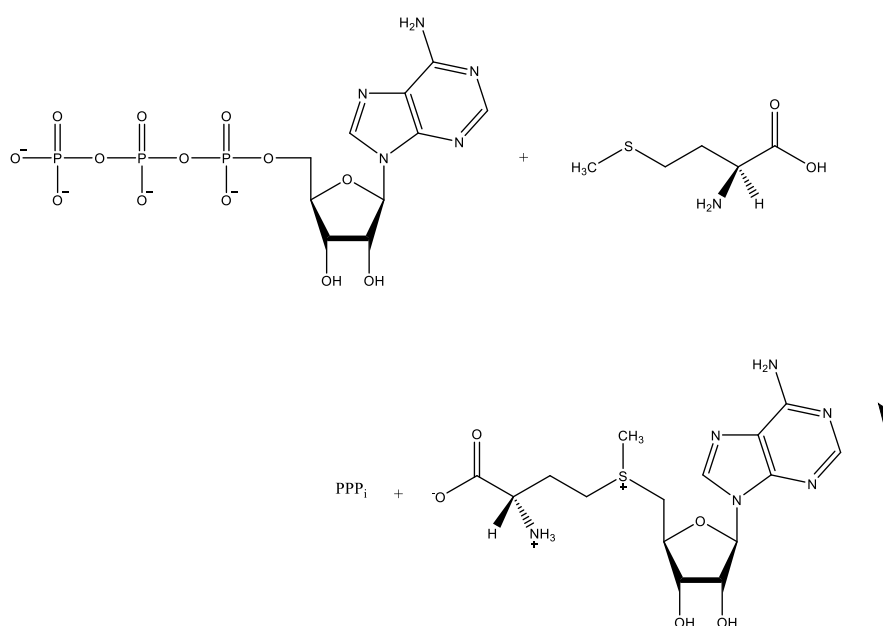
Figura 19: S-ADENOSILMETHIONINE SYNTHETASE.



Fonte: <http://www.rcsb.org/pdb/explore/explore.do?structureId=1FUG>

A reação catalisada pela enzima S-Adenosilmetionina está descrita a seguir:

Esquema 3: Reação de formação da Adenosilmetionina a partir da metionina e uma molécula de ATP.



Fonte: Elaborado pelo autor.

1.1.4 Simulação computacional

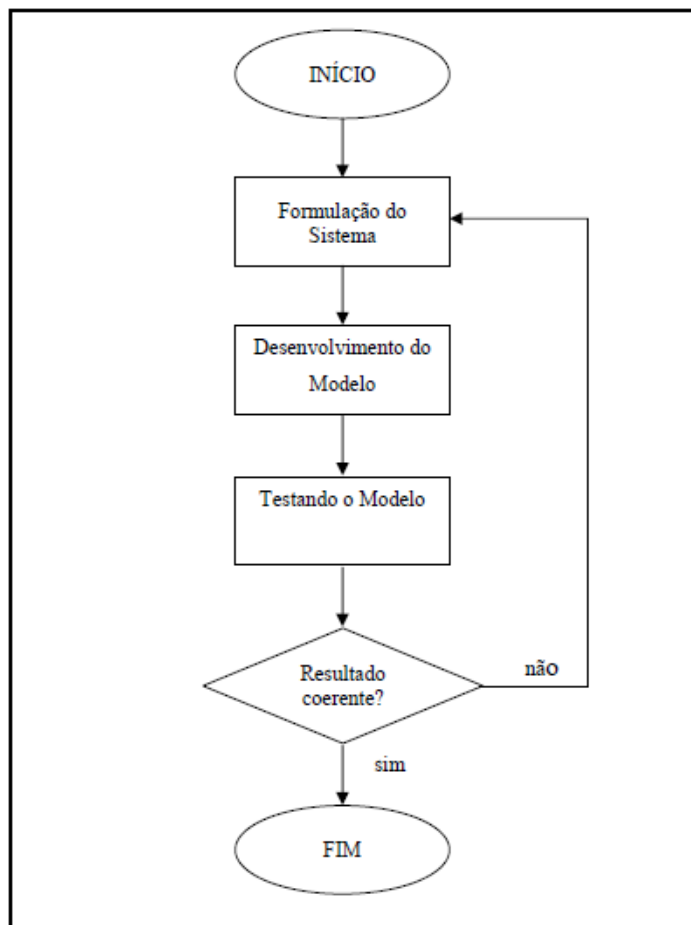
A simulação é uma atividade ou situação que produz uma condição não real, mas tem uma aparência realística e é usada para testar inúmeros fenômenos como, por exemplo o treinamento de pilotos ou a simulação áudio visual do início do universo. Simulação é a imitação de um sistema real modelado em computador para avaliação e melhoria da sua performance, ou seja, simulação é a importação da realidade para um ambiente controlado onde se pode estudar o comportamento do mesmo sob diversas condições sem os riscos físicos e ou grandes custos envolvidos. A simulação envolve a criação de uma história artificial da realidade e com base nesta história artificial são realizadas observações e inferências nas características de operação do sistema real representado.

O processo de simulação segue o método científico:

1. Formular as hipóteses;
2. Preparar o experimento;
3. Testar as hipóteses através do experimento e quanto são válidas as hipóteses através dos resultados obtidos.

A simulação não é uma ferramenta mágica que substitui o trabalho de interpretação humana, mas sim uma ferramenta poderosa capaz de fornecer resultados para análises mais elaboradas a respeito da dinâmica do sistema. Desta forma, permite uma interpretação mais profunda e abrangente do sistema estudado. O modelo utilizado em simulação (HARREL, et. al., 2000) é mostrado na Figura 20.

Figura 20: Modelo científico aplicado à simulação.



Fonte: HARRELL et al. 2000.

CONSTRUÇÃO DO MODELO POR COMPARAÇÃO

Este método é utilizado para construir o modelo tridimensional de uma nova proteína com base no conhecimento da sequência dos aminoácidos da estrutura primária e por comparação com proteínas semelhantes. A estratégia é distribuir as proteínas em famílias e determinar critérios que possam identificar as posições com maior probabilidade de surgimento de divergências com relação à família (GREER, 1990). O processo de modelagem pode ser dividido em alguns passos:

1. *Alinhamento da sequência do alvo com relação às estruturas da família:* Este processo visa mapear cada resíduo da sequência alvo e identificar quais regiões se aproximam das estruturas conhecidas das proteínas de mesma família. Como as proteínas apresentam cadeias com tamanhos diferentes, encontrar semelhanças torna-se um trabalho não trivial. Para contornar o problema dos diferentes tamanhos

utiliza-se métodos de alinhamento resíduo a resíduo. Os resultados são bons quando as proteínas apresentam acima de 40% de semelhança em sua estrutura primária, quaisquer valores abaixo disso geram erros consideráveis nos resultados.

2. *Escolha do núcleo da estrutura*: Uma vez que o alinhamento esteja pronto, é necessário escolher qual região da proteína é semelhante o suficiente, com uma proteína de estrutura já conhecida, para servir como ponto de partida para a construção da estrutura tridimensional. A escolha de um ponto de partida errado compromete toda a estrutura, o que pode gerar uma estrutura totalmente diferente da estrutura real da proteína.

3. *Construção das cadeias laterais*: Alguns algoritmos foram apresentados para a construção das estruturas. Considerando a estrutura principal, os resultados apresentam grande precisão, porém, uma mínima variação na estrutura principal gera erros consideráveis (VÁSQUEZ, 1996).

4. *Construção dos loops*: Depois que a estrutura principal está pronta, algumas regiões pequenas ainda não estão prontas. Estas regiões representam um problema de enovelamento que normalmente é resolvido pelo método *AB INITIO*. Foram relatados alguns algoritmos para a construção dos *loops* que garantem certa precisão na predição da estrutura (JAMES, 1986). Outra dificuldade encontrada na construção dos *loops* é a interação que esta estrutura apresenta com outras ao longo da cadeia (SAMUDRALA, et. al., 1995), erros em regiões vizinhas diminuem a precisão da construção do looping.

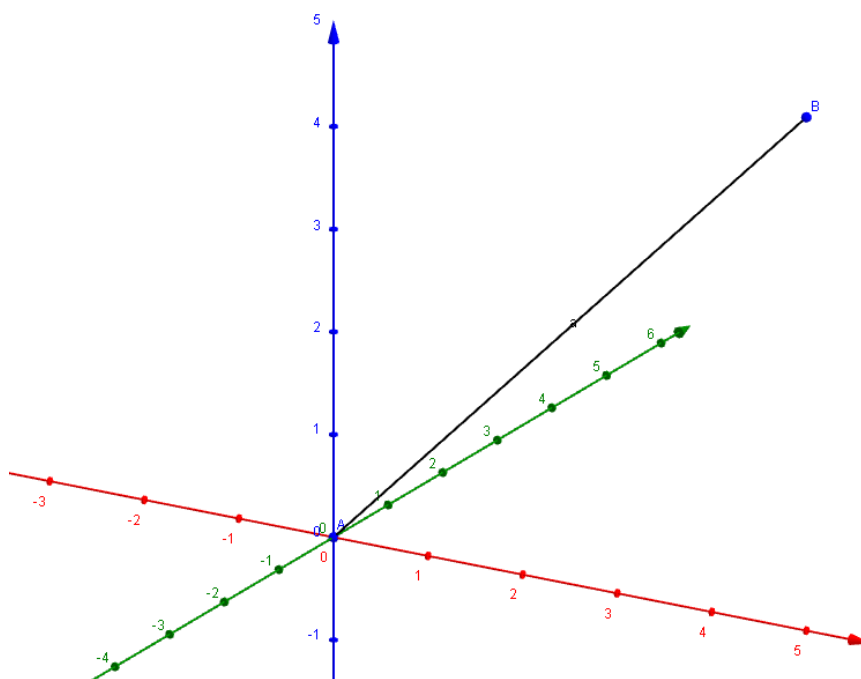
5. *Refinamento dos modelos*: Uma vez que um modelo tenha sido feito, ainda podem ser feitas algumas melhorias. Alguns grupos tentaram minimizar a energia e a dinâmica molecular para reduzir o valor do RMSD (raiz quadrada do desvio médio) dos modelos iniciais com relação às estruturas obtidas experimentalmente.

6. *Estimando a veracidade dos modelos*: Uma grande diferença entre as estruturas e modelos experimentais é que normalmente não existem estimativas de erro que acompanham este último. Ainda que a margem de acerto dos métodos computacionais não rivalize os métodos experimentais, estes são importantes para a predição das funções das proteínas.

1.2. Equações paramétricas

As equações paramétricas são utilizadas em Matemática nas áreas de Cálculo, Geometria Analítica, e principalmente em Geometria Diferencial. A vantagem das equações paramétricas é que elas se prestam a generalizações para espaços de dimensões maiores. Uma reta em três dimensões não tem uma expressão simples na forma explícita ou implícita, pois é escrita como a interseção de dois planos. Por outro lado, as equações paramétricas desta reta são inteiramente análogas, pois derivam da mesma equação vetorial, um exemplo pode ser verificado na Figura 21.

Figura 21: Segmento de reta AB.



Fonte: Elaborado pelo autor.

Seja \overline{AB} um segmento de reta, conforme pode ser observado na Figura 21; a equação que define a reta tridimensional que passa pelos pontos A e B é a combinação de três equações que são definidas para cada eixo. Sejam as coordenadas de A = (0,0,0) e B = (5,0,5), os vetores para cada direção podem ser escritos: $x = x_0 + at$, $y = y_0 + bt$ e $z = z_0 + ct$. Onde x_0 , y_0 e z_0 , são os valores iniciais da abscissa, ordenada e cota, t é um parâmetro, a , b e c são os valores do

vetor nas coordenadas x, y e z . Resumidamente a equação pode ser assim representada:

$$P = P_0 + tv \quad (2)$$

1.2.1 Equação da Reta

Uma equação $y = ax + b$, pode ser “parametrizada”, isto é, pode-se exprimir as variáveis x e y em função de outra variável t , chamada parâmetro. Admitindo-se, $x = t$, a equação $y = ax + b$, toma a forma $f(t) = (t, at + b)$. Em outras palavras, a função agora varia de acordo com o parâmetro t , sendo $x = t$ e $y = at + b$.

Outra forma de se obter uma reta na forma paramétrica é a partir de sua equação vetorial $P = P_0 + tv$, onde $P_0 = (x_0, y_0)$ é um ponto fixo do plano e $v = (a, b)$, o vetor direção da reta. Desta forma se obtém a seguinte equação:

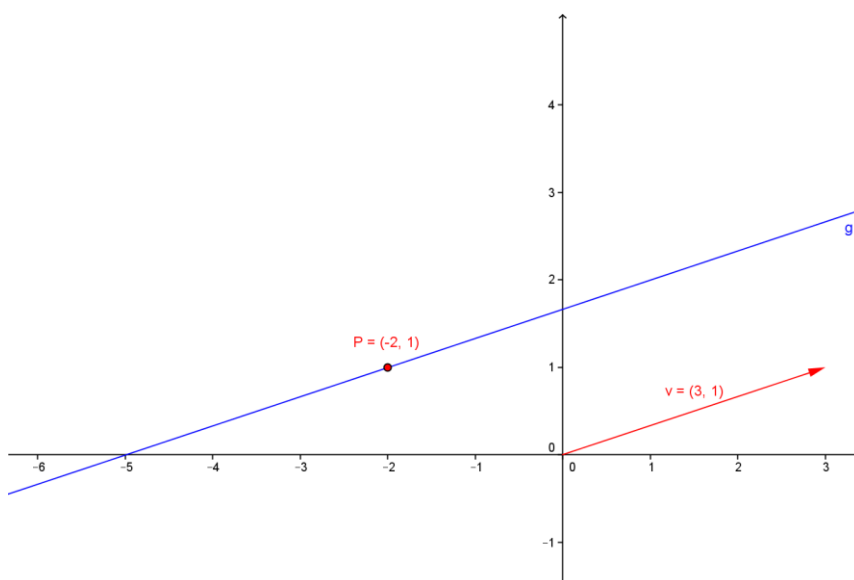
$$f(t) = (x_0 + at, y_0 + bt) \quad (3)$$

Essa mesma equação pode também ser escrita na forma:

$$x = x_0 + at \text{ e } y = y_0 + bt \quad (4)$$

Considerando: $P = (-2, 1)$ e $v = (3, 1)$, é possível construir o gráfico que é apresentado na Figura 22:

Figura 22: Gráfico da função $f(t) = (-2 + 3t, 1 + t)$.



Fonte: Elaborado pelo autor.

Este gráfico apresenta uma reta construída com uma equação em sua forma paramétrica, o vetor v , apresenta o sentido e a direção da reta g que varia com o parâmetro t . O ponto P ocorre quando o parâmetro t assume o valor 0.

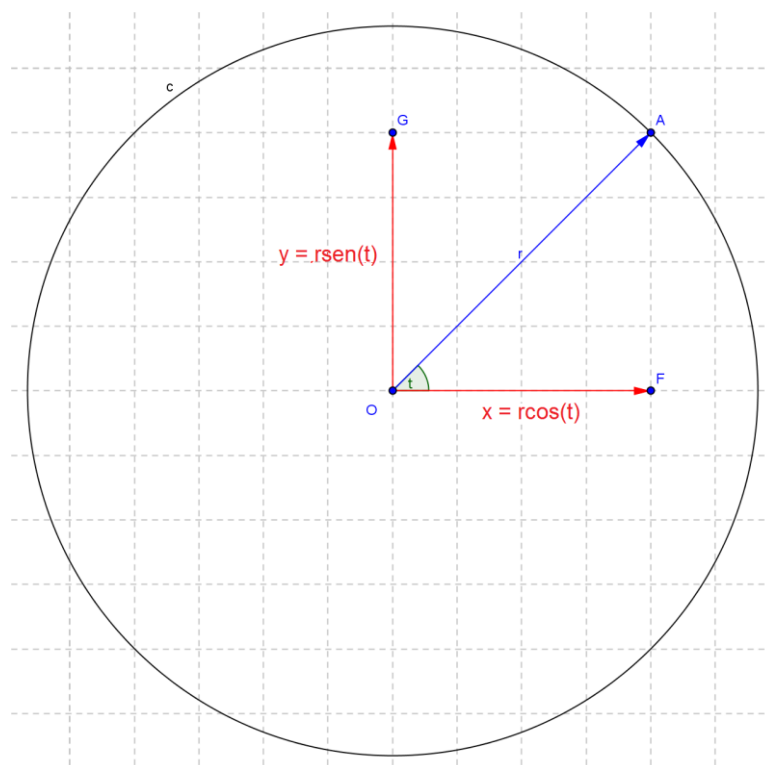
1.2.2 Equação da circunferência

A equação da circunferência poder ser parametrizada a partir de sua equação implícita:

$$x^2 + y^2 = r^2 \quad (5)$$

Pode-se exprimir os valores de x e de y em função de uma variável t , que neste caso representa o ângulo que o vetor r faz com o eixo x , conforme pode ser visto na Figura 23.

Figura 23: Representação Geométrica da parametrização da circunferência.



Fonte: Elaborado pelo autor.

Desta forma a equação inicial pode ser representada em função de t :

$$x^2 + y^2 = (r \cos(t))^2 + (r \sin(t))^2 = r^2 \quad (6)$$

Assim os valores de x e y podem ser obtidos para quaisquer valores que t assumira:

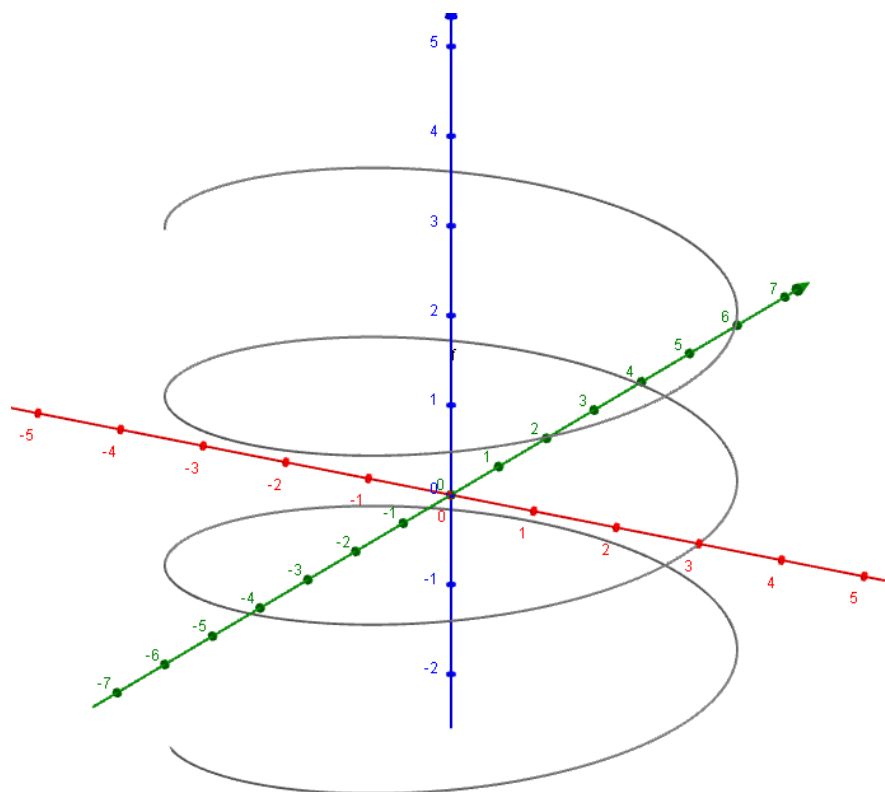
$$C: \begin{cases} x = r \cos(t) \\ y = r \sin(t) \end{cases} \quad (7)$$

A equação geral leva em consideração que a circunferência pode não estar centrada na origem do sistema cartesiano, assim é necessário adicionar as variações:

$$C: \begin{cases} x = x_0 + r \cos(t) \\ y = y_0 + r \sin(t) \end{cases} \quad (8)$$

As paramétricas não se limitam a dois planos, elas também são utilizadas para a construção de retas, curvas e superfícies em três dimensões, conforme podem ser observados nas Figuras 24 e 25.

Figura 24: Equação paramétrica de uma mola.



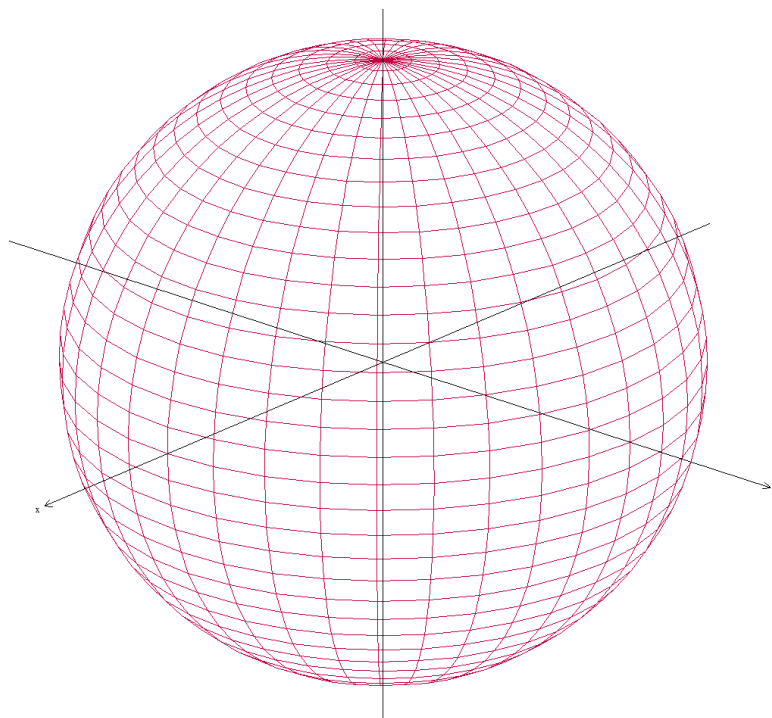
Fonte: Elaborado pelo autor.

A Figura 24, apresenta uma equação paramétrica para a representação de uma mola que tem como equação a seguinte expressão:

$$f(t) = \begin{cases} x = 3\cos(t) \\ y = 3\sin(t) \\ z = 0,3(t) \end{cases} \quad (9)$$

A equação 9 tem domínio que varia para t de -10 a 10.

Figura 25: Esfera.



Fonte: Elaborado pelo autor.

A Figura 25, apresenta uma esfera que tem equação paramétrica igual a:

$$f(u, t) = \begin{cases} x = \text{sen}(u) \times \cos(t) \\ y = \text{sen}(u) \times \text{sen}(t) \\ z = \cos(u) \end{cases} \quad (10)$$

O domínio desta equação é de 0 a 2π para t , e 0 a π para u .

1.2.3 Nós matemáticos

Em matemática, nó é definido como sendo uma curva simples fechada no espaço. O espaço normalmente é representado por \mathbb{R}^3 , pois se trata do domínio de uma equação paramétrica que envolve os três eixos: X, Y e Z.

O nó mais comum é a circunferência e em termos matemáticos pode ser representado conforme Equação 7, onde, t é um parâmetro que varia de 0 a 2π radianos.

O nó mais comumente encontrado em proteínas é o 3_1 , e é representado pela seguinte equação:

$$f(t) = \begin{cases} x = \sin t + 2\sin 2t \\ y = \cos t - 2\cos 2t \\ z = -\sin 3t \end{cases} \quad (11)$$

A equação 11, gera um gráfico tridimensional que representa a estrutura do nó do tipo 3_1 . Este nó está representado na Figura 26.

Figura 26: Nó matemático tipo 3_1 .



Fonte: google imagens.

A função do nó nas proteínas ainda não foi esclarecida, mas ao que tudo indica a presença dessas estruturas pode estar relacionada com a estabilidade de certas proteínas.

Um exemplo dessa estabilidade pode ser observada na ubiquitina, que atua na marcação das proteínas para a degradação no lisossomo (PEREIRA, 2014).

Neste processo a ubiquitina é adicionada a uma proteína formando uma “corrente” de ubiquitinas que recebe o nome de proteína poliubiquitinada. Quando no lisossomo a ubiquitina não sofre degradação e sai inteira desse processo, o que não é comum, uma vez que o tamanho médio das partículas que sofreram degradação é de 8 a 12 resíduos. A presença do nó na ubiquitina pode ser a responsável por tal estabilidade.

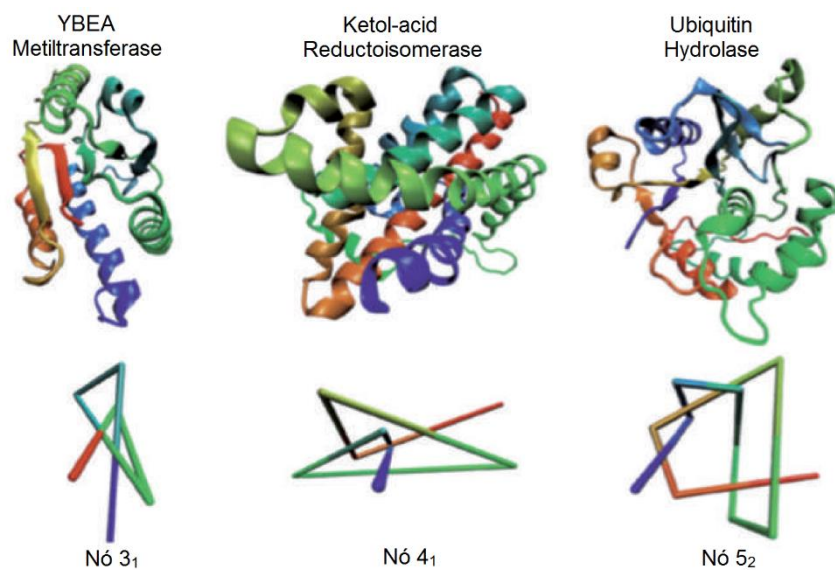
Estudos anteriores revelam que algumas proteínas apresentam nós em suas estruturas. Essas proteínas, antes desconhecidas, trouxeram um novo desafio aos pesquisadores da área: Essas proteínas que apresentam nós, se enovelam normalmente ou este processo ocorre depois de a proteína estar enovelada por um processo externo? A resposta para essa pergunta foi trazida por *Jackson e colaboradores* (MALLAM & JACKSON, 2005). Neste estudo o grupo trabalhou com uma proteína da família da metiltransferase, a YibK. Foram realizados estudos de enovelamento que permitiram concluir que a proteína se enovelava formando nós de forma espontânea.

Alguns estudos realizados por *Todd Yeates e colaboradores* (YEATES, et. al., 2010), mostraram que o enovelamento é mais rápido em proteínas que não apresentam nós em suas estruturas: para duas proteínas quase idênticas o tempo de enovelamento foi cerca de 20 vezes maior para a proteína que continha o nó em sua estrutura. Com base nestes estudos eles justificaram a raridade de se encontrar uma proteína com nó. Outra conclusão que pode ser retirada de seus estudos é que dentre uma infinidade de caminhos possíveis para o enovelamento as proteínas evitam gerar nós em suas estruturas pois este processo é energeticamente desfavorável, demorado e poderia colapsar a estrutura. Foi observado também que os nós são mais comuns em cadeias com mais de 300 aminoácidos pois essa quantidade de resíduos minimiza a energia necessária para a formação do nó além da interação e a probabilidade que a estrutura entre em colapso.

Alguns estudos realizados mostraram que as proteínas enovelam sempre procurando o caminho de menor energia para a estrutura (KUHLMAN, 2003), em alguns casos foram propostos que durante o enovelamento de uma proteína com nó, as voltas componentes dos nós são compostas de peptídeos já prontos que são "encaixados" na estrutura, o que diminui a energia do processo.

Alguns exemplos de nós encontrados em proteínas são apresentados na Figura 27.

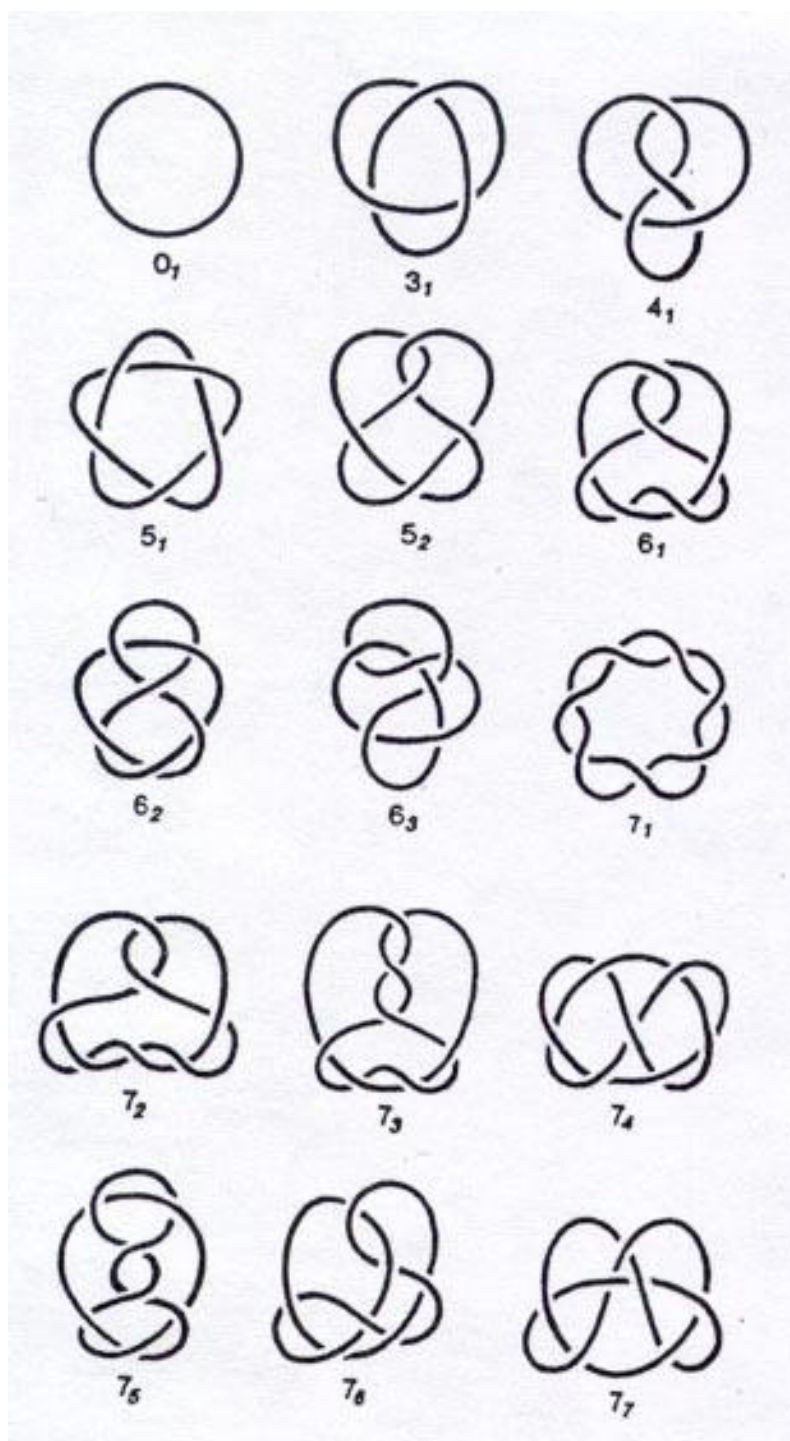
Figura 27: Nós em proteínas



Fonte: Adaptado de nature materials | VOL 10 | FEBRUARY 2011 | www.nature.com/naturematerials, pg 85.

A Figura 27, apresenta as estruturas YBEA Metiltransferase, Ketol-acid Reductoisomerase e Ubiquitin Hydrolase; todas apresentam um nó em suas cadeias. Os nós matemáticos são: 3_1 , 4_1 e 5_2 , e são matematicamente representados como mostrado na Figura 28.

Figura 28: Nós matemáticos



Fonte: Google imagens

Grande parte dos trabalhos desenvolvidos focam seus esforços no entendimento do enovelamento das proteínas, considerando o caminho de menor energia.

Matematicamente os nós são diferentes dos encontrados nas proteínas, isso ocorre devido às interações que ocorrem na estrutura da proteína. Essas interações deformam a proteína, afastando sua distribuição espacial do nó matemático em questão. O principal fator que evidencia esse afastamento é o fato de uma proteína ter começo e fim; matematicamente, os nós são estruturas fechadas sem começo ou fim definidos (Figura 28).

O nó 3_1 é o mais comum em proteínas e também tem sua equação matemática bem definida, conforme apresentada na equação 11. Devido a esses fatores as proteínas com nó 3_1 foram escolhidas para o desenvolvimento deste trabalho.

1.4. Softwares utilizados

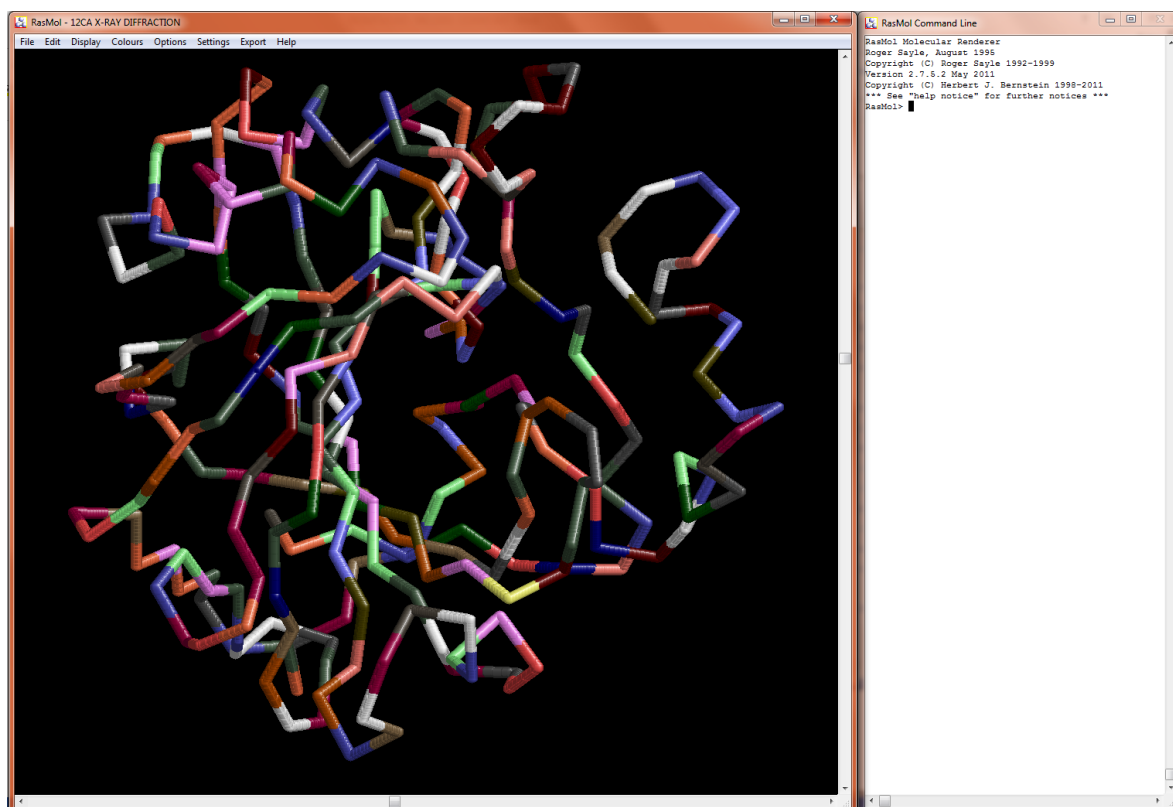
1.4.1 Rasmol

RasMol é um programa para visualização molecular e sua interface permite visualizar separadamente os resíduos de aminoácidos que compõem uma determinada proteína e da mesma forma, é possível mudar a visualização para a atômica, onde pode-se ver os átomos constituintes de toda a estrutura e é possível também observar as conformações presentes na proteína.

Quando executado, o programa exibe duas janelas: uma para a entrada de comandos e outra para a visualização da estrutura, a Figura 29 apresenta a interface do programa. Na janela de visualização há oito menus que consistem em:

- ➔ File: Nesta aba é possível abrir, fechar e imprimir um dado arquivo;
- ➔ Edit: Aqui se encontram os comandos básicos como selecionar, copiar, colar e recortar;
- ➔ Display: É aqui que se seleciona o modo de visualização da estrutura;
- ➔ Colours: Nesta aba é possível escolher os vários tipos de exibições possíveis, em cada modo; a cor é apresentada de forma diferente, permitindo até individualizar os resíduos de aminoácidos presentes na estrutura;
- ➔ Options: Esta aba permite selecionar os átomos que serão mostrados durante a visualização;
- ➔ Settings: Permite escolher quais operações serão realizadas ao clicar na estrutura com os botões direito e esquerdo do mouse;
- ➔ Export: Como o nome sugere, nesta aba, pode-se alterar a extensão do arquivo para vários formatos específicos de outros programas;
- ➔ Help: Nesta última aba, pode-se registrar, visualizar a versão do programa, fazer uma doação para o grupo de desenvolvedores e consultar o manual do usuário.

Figura 29: Interface do programa RasMol.



Fonte: Elaborado pelo autor.

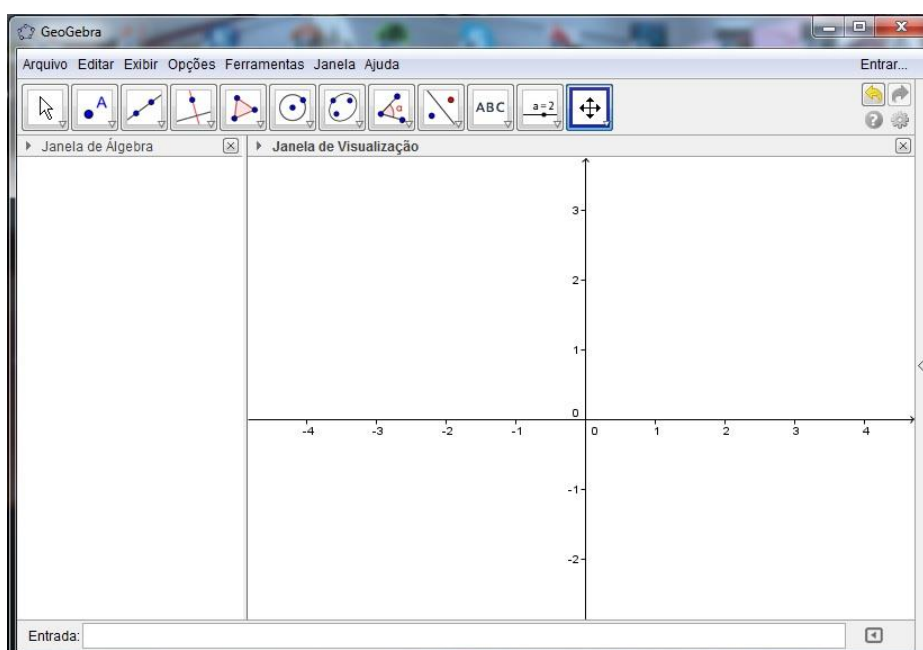
A Figura 29 apresenta a interface do software RasMol sendo apresentada a visualização da proteína 12CA. As configurações utilizadas na exibição foram Display – Backbone e Colours – Shapely, esta configuração de exibição permite visualizar de forma limpa a estrutura da proteína e a cor escolhida apresenta cada resíduo de aminoácido de cor diferente, repetindo as cores conforme os resíduos aparecem na cadeia, o que torna a visualização mais cômoda.

O RasMol é uma ferramenta para a visualização e identificação de estruturas atômicas, moleculares e conformacionais. O software utiliza arquivos de extensão .pdb que podem ser encontrados no banco de dados *Protein Data Bank*.

1.4.2 Geogebra

O GeoGebra é um software de matemática dinâmica que permite trabalhar com: geometria, álgebra, tabelas, gráficos, estatística e cálculo, em uma única ferramenta. Sua distribuição é gratuita e, por ser produzido em linguagem java torna-se um software multi-plataforma, que pode ser executado tanto no windows, como no linux ou Mac-OS. Sua interface é bastante intuitiva, o que permite a um iniciante no programa, realizar tarefas básicas. O software permite que as construções sejam realizadas de duas formas, na primeira utiliza-se o mouse para criar pontos, retas, circunferências, polígonos, gráficos, entre outras tantas possibilidades, na segunda forma, utiliza-se a barra de entrada, que permite executar as mesmas tarefas através de comandos. A Figura 30 mostra a interface do programa.

Figura 30: Interface do software GeoGebra.

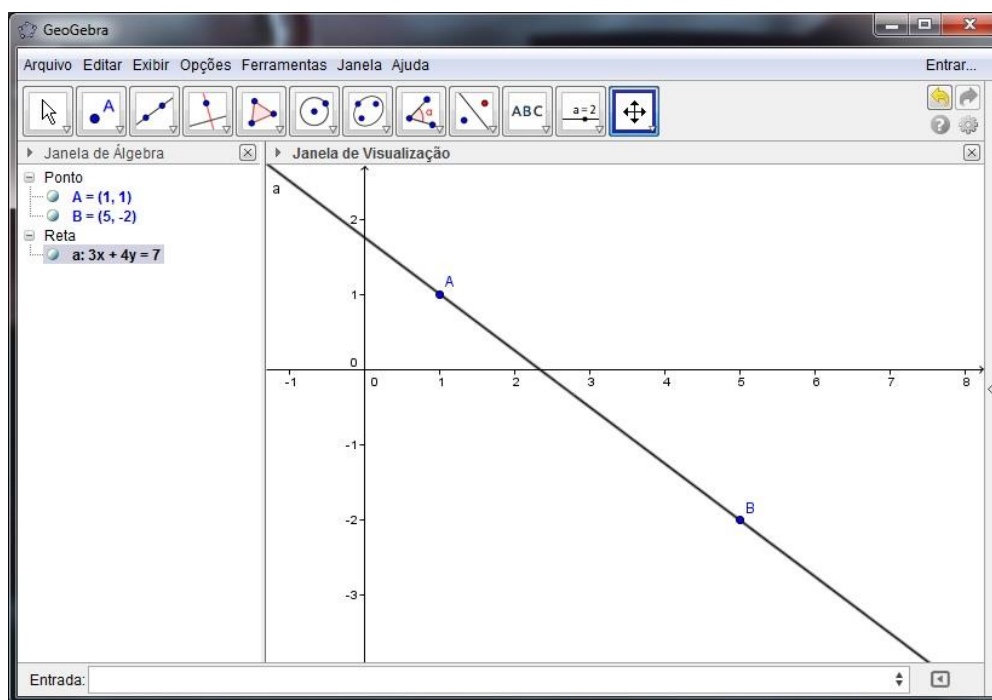


Fonte: Elaborado pelo autor.

As Figuras 21, 22, 23 e 24, são exemplos de trabalhos desenvolvidos com esta ferramenta. Ao ser executado o programa exibe três janelas: a janela de Álgebra, a de Visualização e a de entrada, as construções podem ser realizadas através dos ícones dos atalhos na parte superior da interface do programa. No caso de uma reta que passa por dois pontos, a janela de álgebra mostra as coordenadas

dos pontos e a equação da reta que passa por estes pontos, a janela de visualização neste caso, exibirá a reta e os dois pontos, este exemplo pode ser observado na Figura 31.

Figura 31: Exemplo de construção com o GeoGebra.



Fonte: Elaborado pelo autor.

Todas as construções podem ser realizadas com o mouse de modo que mesmo sem utilizar a janela de entrada o usuário consegue desenvolver diversas atividades no programa.

2. OBJETIVOS

O objetivo deste trabalho foi estudar os nós que algumas proteínas apresentam e propor equações matemáticas que simulem a estrutura do nó da proteína.

Para os cálculos foram utilizadas as equações em forma paramétrica, que permitem trabalhar com três dimensões. Neste trabalho foram utilizadas apenas equações paramétricas com um parâmetro, uma vez que pretendia-se aproximar as estruturas dos nós das proteínas à uma equação matemática.

Aprimorar a curva de distâncias obtida através de equações para que estas se tornassem mais próximas das curvas construídas com as distâncias entre os carbonos alfa da estrutura tridimensional da proteína. Estas curvas foram utilizada como parâmetro para comparação e análise dos resultados.

3. MATERIAIS E MÉTODOS

3.1 Materiais

Os materiais utilizados para a pesquisa foram as análises cristalográficas das proteínas que foram obtidas no *pknnot database*, além de softwares de licença gratuita que foram utilizados para o desenvolvimento deste trabalho.

Os softwares utilizados foram:

- RasMol – para visualização e análise das estruturas das proteínas;
Disponível em: <http://www.openrasmol.org>
- GeoGebra – para desenvolver e alterar as equações;
Disponível em: http://www.geogebra.org/cms/pt_BR/
- Editores de texto – para a contagem dos resíduos de aminoácidos;
- Editores de planilhas – para tratamentos dos dados obtidos;
- Editores de gráficos – para exibir os resultados obtidos;

As proteínas analisadas foram:

- 12CA (Anidrase Carbônica);
- 1GZ0 (Metiltransferase);
- 1JS1 (Transcarbamilase);
- 1FUG (S – Adenosilmetionina Sintetase);
- 1TSK (Neurotoxína isolada do escorpião);
- 2EFV (Proteína hipotética do *Methanocaldococcus jannaschii*);
- 1R1D (Carboxilesterase);

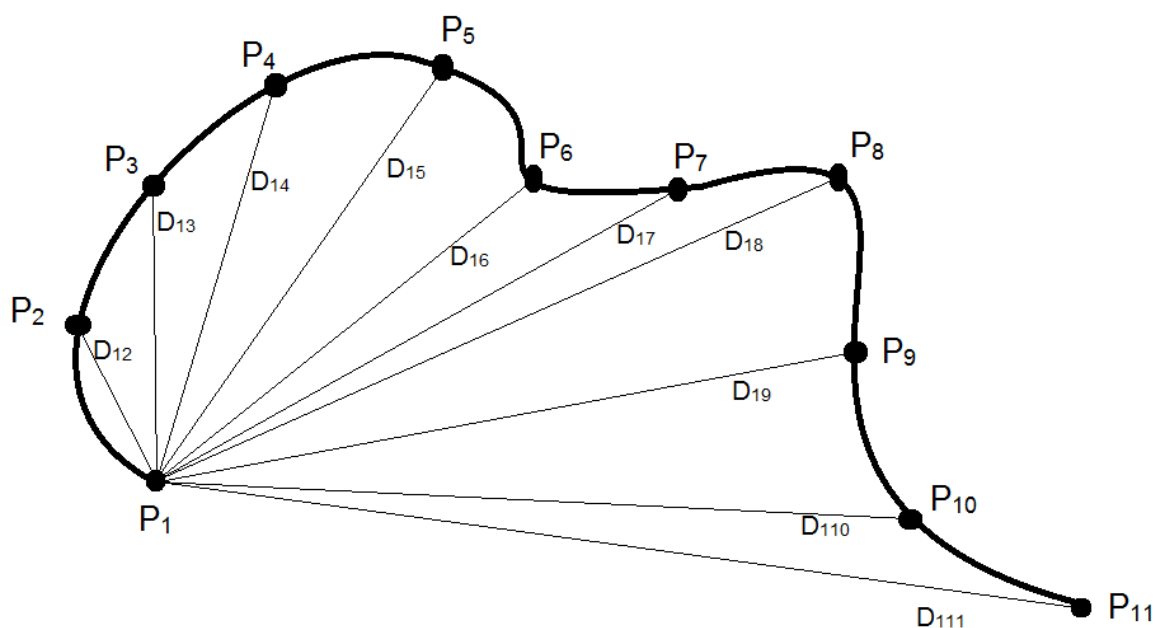
Disponíveis em: <http://www.rcsb.org/pdb/home/home.do>

3.2 Métodos

Primeira parte: Construção do gráfico de distâncias entre os carbonos alfa na região do nó.

A construção do gráfico de distâncias consiste em calcular as distâncias de todos os carbonos alfa da região do nó com relação ao primeiro destes. Este método permite construir um gráfico bidimensional para representar a estrutura da proteína. Os cálculos podem ser realizados de duas formas, na primeira as distâncias são calculadas com relação ao primeiro carbono alfa da região do nó; na segunda, tendo por base todas as coordenadas dos carbonos alfa, é então calculado o ponto médio geométrico, e as distâncias são calculadas a partir deste ponto. A Figura 32 exemplifica o método de cálculo de distâncias entre onze pontos.

Figura 32: Cálculo de distâncias

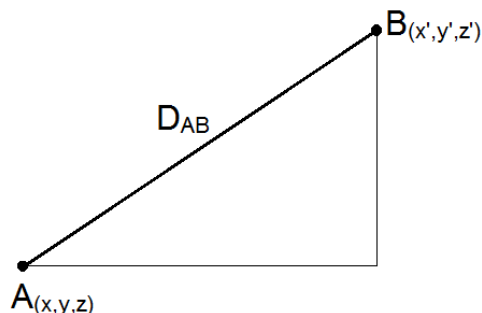


Fonte: Elaborado pelo autor.

A Figura 32 apresenta uma curva com onze pontos P dos quais foram traçadas as distâncias D, com relação ao primeiro ponto. Quando a distância é

calculada em um plano pode-se utilizar o teorema de Pitágoras para calcular a distância entre dois pontos, isso pode ser verificado na Figura 33.

Figura 33: Distância entre dois pontos.



Fonte: Elaborado pelo autor

Na Figura 33, os catetos podem ser obtidos pelas diferenças entre as coordenadas dos eixos x e y, e a distância entre os pontos A e B pode ser calculada pelo teorema de Pitágoras, assim obtemos:

$$D_{A \rightarrow B} = \sqrt{(x' - x)^2 + (y' - y)^2} \quad (12)$$

De forma análoga o cálculo para a distância entre dois pontos não coplanares pode ser realizado pela adição da coordenada do eixo z:

$$D_{A \rightarrow B} = \sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2} \quad (13)$$

Quando a distância é calculada a partir do primeiro carbono alfa, a coordenada de $P_{(x,y,z)}$ permanece constante, variando apenas as coordenadas dos carbonos seguintes P_1 , P_2 , e assim sucessivamente até o último carbono alfa.

O cálculo também pode ser realizado para o ponto médio geométrico, que é calculado como a média aritmética das três coordenadas:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (14)$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (15)$$

$$\bar{z} = \frac{\sum_{i=1}^n z_i}{n} \quad (16)$$

Onde:

\bar{x} , \bar{y} e \bar{z} , são as médias das coordenadas x, y e z de todos os carbonos alfas na região do nó. O ponto obtido (PM) é o ponto médio geométrico que é utilizado como ponto fixo a partir de onde serão calculadas as distâncias.

$$D_{C\alpha \rightarrow PM} = \sqrt{(x' - x_{PM})^2 + (y' - y_{PM})^2 + (z' - z_{PM})^2} \quad (17)$$

O gráfico de distâncias é construído traçando na ordenada a distância em angstroms e na abscissa o número dos carbonos alfa.

Segunda parte: Alteração do nó matemático.

A alteração do nó matemático 3_1 é uma proposta de aproximação deste com a estrutura tridimensional da proteína.

Com a equação matemática, pode-se construir uma curva de distâncias de forma semelhante ao que é feito para a proteína. O processo consiste de dividir em grande quantidade o intervalo por onde a curva varia.

A curva do nó 3_1 está definida para o intervalo de 0 a 2π , dividindo esse intervalo por dez, por exemplo, obteremos $\pi/5$, $2\pi/9$, $\pi/4$, $2\pi/7$, $\pi/3$, ... , 2π . Essas frações de intervalo quando substituídas na equação 11, fornecem os valores das coordenadas x, y e z, gerando no total dez pontos com coordenadas x, y e z. Esses pontos são utilizados para calcular as distâncias de forma análoga à realizada para os carbonos alfa.

As alterações do nó foram realizadas no software GeoGebra que permite a criação de pontos deslizantes. Com este recurso foi possível verificar onde cada parâmetro da equação interferia na construção da curva. Uma vez determinada, a equação foi submetida aos cálculos de distâncias para verificar sua proximidade com a curva de distâncias das proteínas.

4. RESULTADOS E DISCUSSÃO

O gráfico de distâncias foi pensado para permitir a comparação direta entre a estrutura da proteína com a estrutura gerada a partir de uma equação matemática, ele apresenta também a possibilidade de interpretação da estrutura secundária da proteína através da análise das “oscilações curtas”.

Gráficos das Equações:

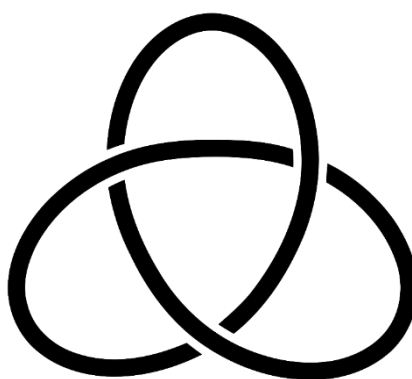
Equação padrão do nó 3_1 :

O nó 3_1 , apresentado na Figura 34, apresenta equação paramétrica igual a:

$$f(t) = \begin{cases} x = \sin t + 2\sin 2t \\ y = \cos t - 2\cos 2t \\ z = -\sin 3t \end{cases} \quad (11)$$

O parâmetro t varia de 0 a 2π .

Figura 34: Nó 3_1 .



Fonte: Google imagens.

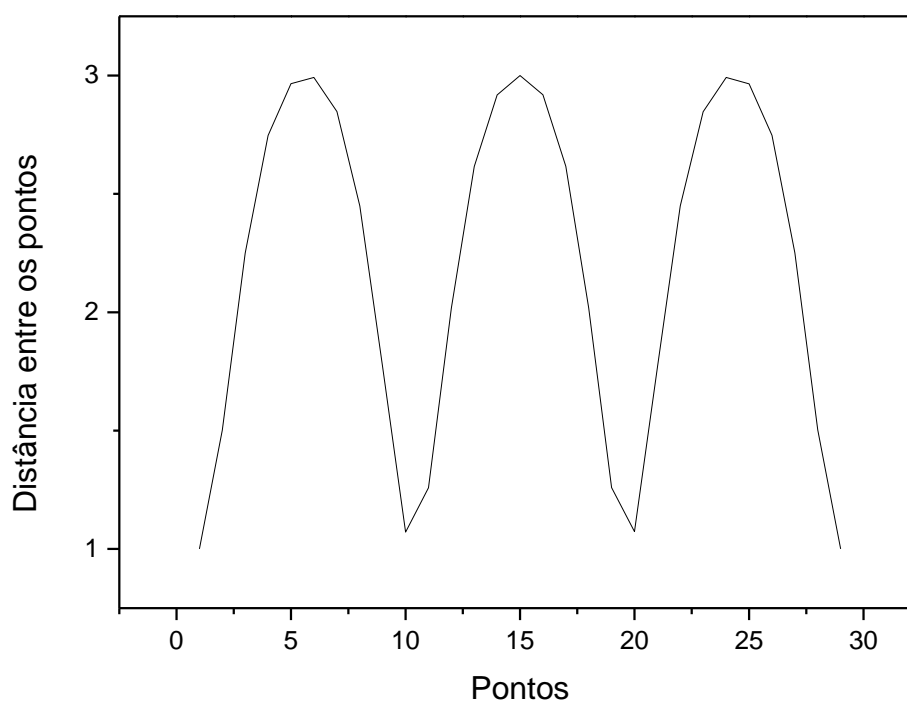
Para o cálculo das coordenadas x , y e z o intervalo foi dividido em 28 partes, os resultados são apresentados na Tabela 2.

Tabela 2: Cálculos das coordenadas e das distâncias do nó 3₁ padrão.

Ponto	x	y	z	Distância
0	0	-1	0	1
0,224	1,088459	-0,82761	-0,62255	1,502418
0,449	1,998229	-0,34547	-0,97506	2,250114
0,673	2,573014	0,33614	-0,90123	2,746925
0,898	2,731579	1,069785	-0,4328	2,965346
1,122	2,464626	1,680869	0,222529	2,991525
1,346	1,844037	2,024156	0,781089	2,847416
1,571	0,999185	1,999796	1	2,448991
1,795	0,10791	1,579947	0,782198	1,76627
2,02	-0,66388	0,811465	0,221337	1,071541
2,244	-1,16803	-0,17847	-0,4339	1,258737
2,468	-1,32641	-1,22509	-0,90046	2,017678
2,693	-1,12945	-2,14868	-0,97479	2,615858
2,917	-0,64575	-2,77649	-0,62394	2,918079
3,142	0,001222	-3	0,001222	3
3,366	0,645267	-2,77685	0,623508	2,918224
3,59	1,129155	-2,14934	0,974666	2,616226
3,815	1,326388	-1,22593	0,900697	2,018282
4,039	1,168313	-0,17934	0,434399	1,259293
4,264	0,661493	0,814747	-0,22372	1,07305
4,488	-0,1072	1,579445	-0,78185	1,765625
4,712	-0,99844	1,99961	-1	2,448536
4,937	-1,84341	2,024298	-0,78144	2,847207
5,161	-2,46424	1,681282	-0,22307	2,991483
5,386	-2,7318	1,067245	0,434999	2,964953
5,61	-2,57332	0,336747	0,900986	2,747211
5,834	-1,99886	-0,34497	0,975185	2,250649
6,059	-1,08931	-0,82733	0,622987	1,503059
6,283	-0,00093	-1	0,000556	1

O Gráfico obtido a partir do cálculo de distâncias pode ser observado no Gráfico 1.

Gráfico 1: Distribuição de distâncias da equação do nó 3₁.



Equação alterada do nó 3₁:

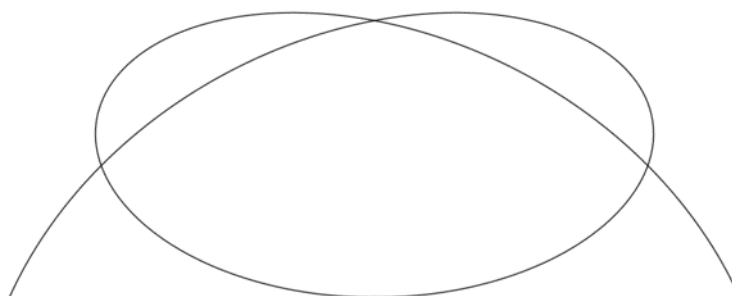
O gráfico do nó matemático foi alterado e analisado conforme a equação:

$$f(t) = \begin{cases} x = \frac{-\text{sen}(t)}{2} + 2 \text{sen}(2t) \\ y = \frac{3\cos(t)}{2} - \frac{3 \cos(2t)}{2} \\ z = -\text{sen}(3t) \end{cases} \quad \text{Domínio : } 0 \leq t \leq \frac{2}{3}\pi; e; \frac{4}{3}\pi \leq t \leq 2\pi \quad (18)$$

A equação proposta tem domínio que varia para t em dois terços com relação à equação x. O domínio não está definido para o intervalo: $\frac{2}{3} \leq \pi \leq \frac{4}{3}$.

A Figura 35 apresenta o nó matemático alterado conforme a equação 10.

Figura 35: Representação gráfica da equação alterada do nó 3₁.



Fonte: Elaborado pelo autor.

Curva padrão para o nó 3₁:

$$f(t) = \begin{cases} x = \text{sen}t + 2\text{sen}2t \\ y = \text{cos}t - 2\text{cos}2t \\ z = -\text{sen}3t \end{cases} \quad \text{Domínio: } 0 \leq t \leq 2\pi \quad (11)$$

A equação 11, começa e termina em um mesmo ponto, o que inviabiliza a comparação com as proteínas, uma vez que suas estruturas não são fechadas. A equação 18, apresenta domínio que vai de zero a 2π , porém não está definida para o intervalo: $\frac{2}{3} \leq \pi \leq \frac{4}{3}$, esse trecho tem a função de aproximar a curva com a estrutura da proteína, o intervalo considerado gera uma curva aberta com início e fim em pontos distintos.

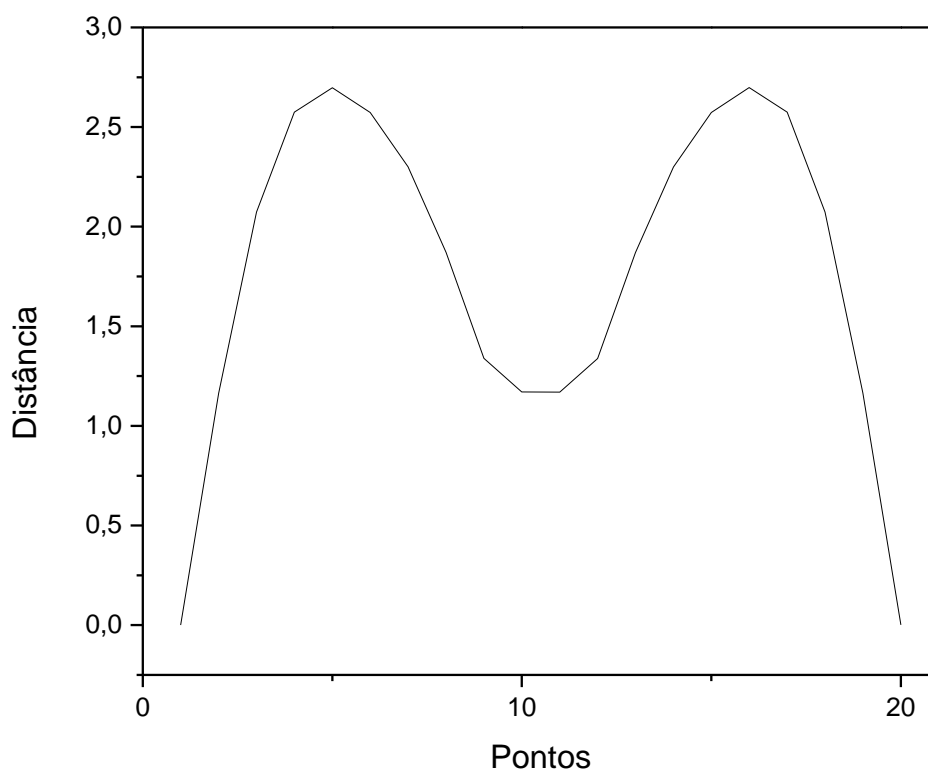
A tabela que apresenta os cálculos para determinação das coordenadas e o cálculo das distâncias pode ser verificada na Tabela 3.

Tabela 3: Cálculos das coordenadas e das distâncias do nó 3₁ alterado.

Ponto	x	y	z	Distância
0	0	0	0	0
0,224	0,977393	0,110552	-0,62255	1,164084
0,449	1,781197	0,416559	-0,97506	2,072904
0,673	2,261346	0,838571	-0,90123	2,574704
0,898	2,340538	1,26972	-0,4328	2,697707
1,122	2,014141	1,586063	0,222529	2,573301
1,346	1,356617	1,685298	0,781089	2,300161
1,571	0,499185	1,499694	1	1,870366
1,795	-0,37958	1,018213	0,782198	1,338906
2,02	-1,11428	0,282913	0,221337	1,170743
2,244	-1,55895	-0,60147	-0,4339	1,72637
2,468	-1,63831	-1,50501	-0,90046	2,399982
2,693	-1,3463	-2,28731	-0,97479	2,827456
2,917	-0,75711	-2,81353	-0,62394	2,979675
3,142	0,001426	-3	0,001222	3
3,366	0,756532	-2,81383	0,623508	2,979724
3,59	1,34592	-2,28786	0,974666	2,827682
3,815	1,638215	-1,50572	0,900697	2,400459
4,039	1,559169	-0,60223	0,434399	1,726962
4,264	1,112067	0,285924	-0,22372	1,169827
4,488	0,380265	1,017701	-0,78185	1,338509
4,712	-0,49844	1,499416	-1	1,869945
4,937	-1,35597	1,685269	-0,78144	2,299875
5,161	-2,01372	1,586247	-0,22307	2,573132
5,386	-2,34101	1,268293	0,434999	2,697798
5,61	-2,26158	0,83894	0,900986	2,574948
5,834	-1,78174	0,416874	0,975185	2,073493
6,059	-0,97815	0,110731	0,622987	1,16497
6,283	-0,00083	7,73E-08	0,000556	0,001002

O Gráfico 2, apresenta a distribuição de distâncias do nó proposto na equação 18.

Gráfico 2: Distribuição de distâncias para a curva da equação 18.



Como pode ser observado, o gráfico 2 difere bastante do gráfico obtido para a equação padrão. A principal diferença é o domínio para o qual as equações estão definidas e uma vez “aberto” a equação pode ser manipulada para se tornar mais representativa com relação aos nós das proteínas. As alterações que foram feitas envolveram somente os coeficientes uma vez que o objetivo era ajustar a equação para torná-la mais real e representativa.

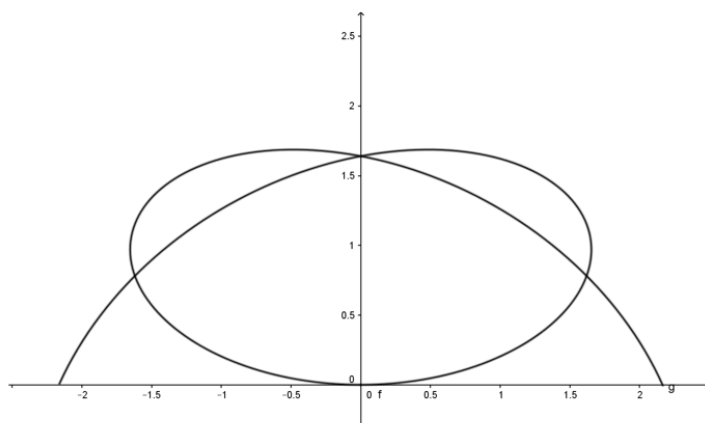
Proposta de enovelamento e/ou desenovelamento

Em termos matemático é possível propor um modelo para o qual a proteína se enovela partindo de uma estrutura aberta, de forma reversa pode ser aplicada para o desenovelamento da cadeia.

Nas Figuras 36, 37, 38 e 39 são apresentadas equações que permitem o acompanhamento do processo de enovelamento/desenovelamento.

Figura 36: Passo 1 – etapa inicial.

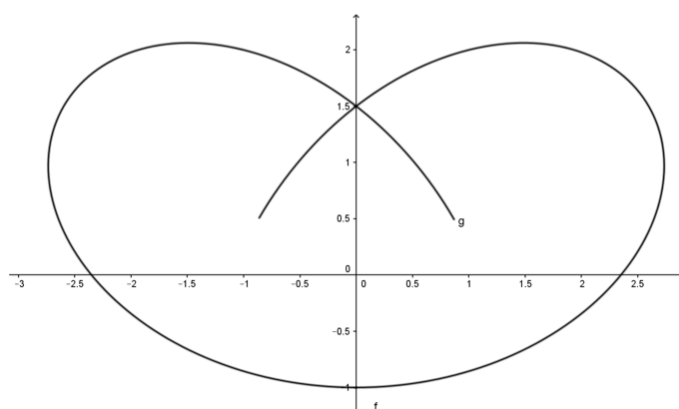
$$f(t) = \begin{cases} x = \frac{-\text{sen}(t)}{2} + 2 \text{sen}(2t) \\ y = \frac{3 \cos(t)}{2} - \frac{3 \cos(2t)}{2} \\ z = -\text{sen}(3t) \end{cases} \quad \text{intervalo: } 0 \leq t \leq \frac{2}{3}\pi; e; \frac{4}{3}\pi \leq t \leq 2\pi$$



Fonte: Elaborado pelo autor.

Figura 37: Passo 2 - início da abertura do nó.

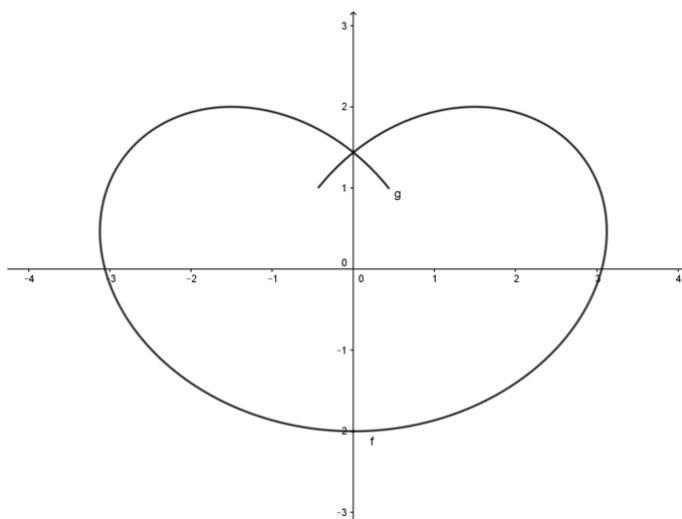
$$f(t) = \begin{cases} x = \text{sen}(t) + 2 \text{sen}(2t) \\ y = \cos(t) - 2 \cos(2t) \\ z = -\text{sen}(3t) \end{cases} \quad \text{intervalo: } 0 \leq t \leq \frac{2}{3}\pi; e; \frac{4}{3}\pi \leq t \leq 2\pi$$



Fonte: Elaborado pelo autor.

Figura 38: Passo 3 - segundo passo da abertura do nó.

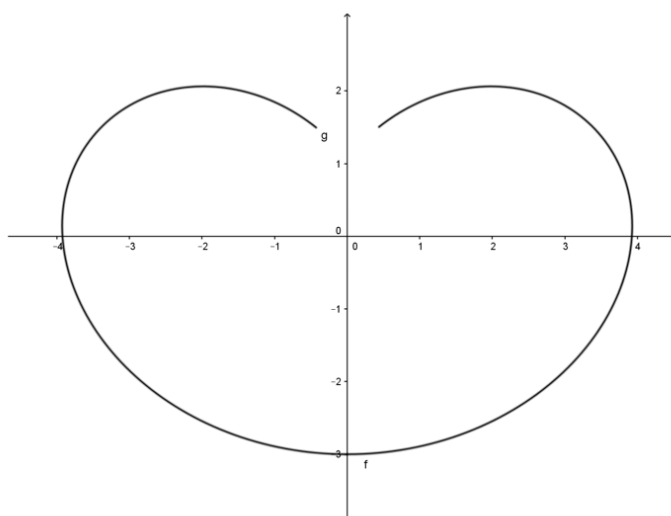
$$f(t) = \begin{cases} x = 1,5 \operatorname{sen}(t) + 2 \operatorname{sen}(2t) \\ y = -2 \cos(2t) \\ z = -\operatorname{sen}(3t) \end{cases} \quad \text{intervalo: } 0 \leq t \leq \frac{2}{3}\pi; e; \frac{4}{3}\pi \leq t \leq 2\pi$$



Fonte: Elaborado pelo autor.

Figura 39: Passo 4 - nó aberto.

$$f(t) = \begin{cases} x = \operatorname{sen}(t) + 2 \operatorname{sen}(2t) \\ y = \cos(t) - 2 \cos(2t) \\ z = -\operatorname{sen}(3t) \end{cases} \quad \text{intervalo: } 0 \leq t \leq \frac{2}{3}\pi; e; \frac{4}{3}\pi \leq t \leq 2\pi$$



Fonte: Elaborado pelo autor.

A sequência apresentada mostra a abertura do nó com a variação dos coeficientes das equações, essa sequência representa o desenovelamento da proteína, o processo de enovelamento é o processo contrário, isto é do passo 4 para o passo 1.

Previsão das estruturas secundárias

As estruturas secundárias das proteínas podem ser identificadas através do gráfico de distâncias, mais especificamente a estrutura em hélice. As estruturas em folhas betas podem ser confundidas com as voltas o que torna a sua identificação um problema. Por se assemelharem a estruturas de molas as hélices são facilmente identificadas no gráfico.

As estruturas em hélices foram inicialmente observadas e identificadas por *Pauling e Corey* que em seus trabalhos verificaram a presença de tais estruturas nas cadeias das proteínas, eles propuseram que a estrutura da hélice era semelhante a estrutura de uma mola e que cada volta continha aproximadamente quatro aminoácidos.

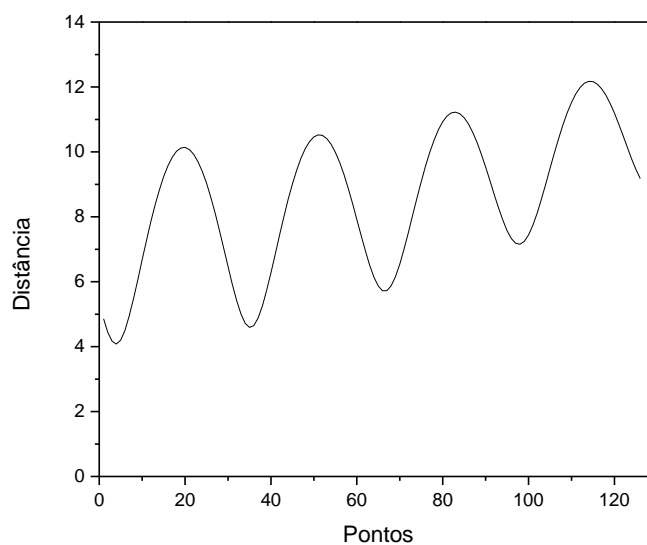
Partindo da ideia de comparar a hélice com uma mola, e tomando como base um ponto de coordenada P (0,0,0) é possível observar, em termos de coordenadas dos carbonos alfas, quatro leituras por volta, isto é, aproximadamente a cada 90 graus existe um carbono alfa.

A tabela de cálculos para a construção do gráfico teve como bases a hélice (helicóide) de equação igual a:

$$f(t) = \begin{cases} x = 3 \cos(t) \\ y = 3 \operatorname{sen}(t) \\ z = 0,3(t) \end{cases} \quad (19)$$

O domínio para esta equação variou de 0 a 8π com relação ao parâmetro t ; o incremento foi de 0,2 e o ponto tomado como referência foi P (5,5,0).

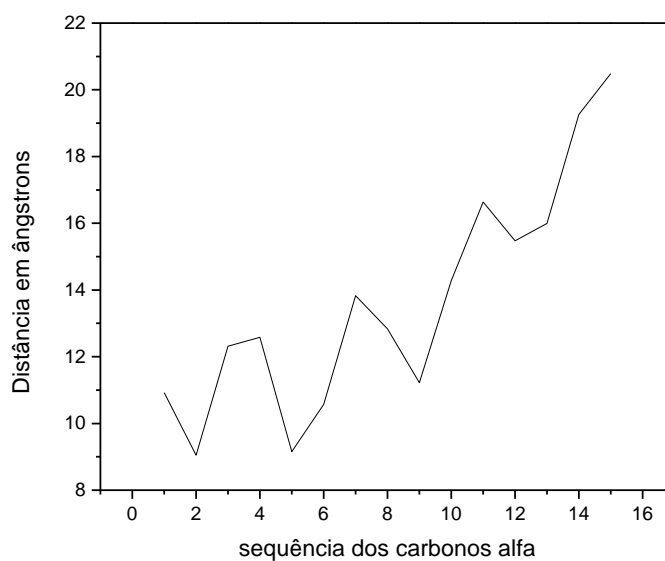
Gráfico 3: Gráfico de distâncias da estrutura tridimensional de uma hélice (helicóide).



O Gráfico 3, apresenta o cálculo de distâncias realizado para a equação 19 e é um gráfico apenas para exemplo e adimensional, o objetivo deste gráfico é apenas mostrar as oscilações que ocorrem quando a partir de um ponto fixo são tomadas as distâncias entre os pontos de uma estrutura em hélice.

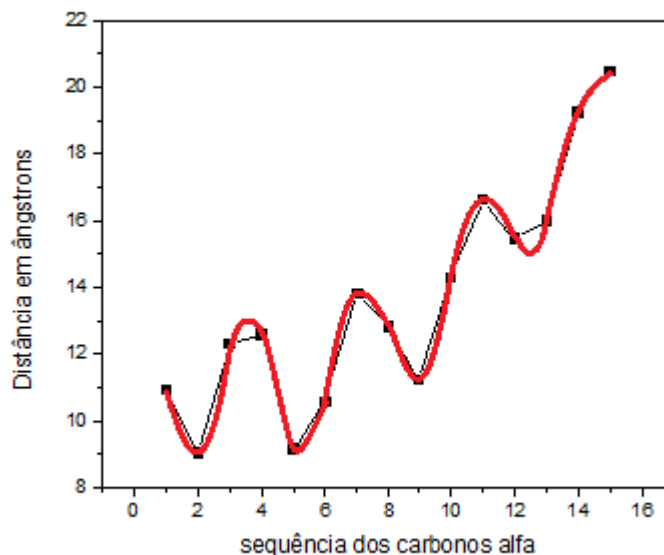
Em adição ao gráfico da helicóide foi construída uma curva de distâncias para uma hélice real da proteína carboxil esterase de código pdb 1R1D, que pode ser observada no gráfico 4.

Gráfico 4: Distribuição de distâncias para uma hélice real da proteína carboxil esterase.



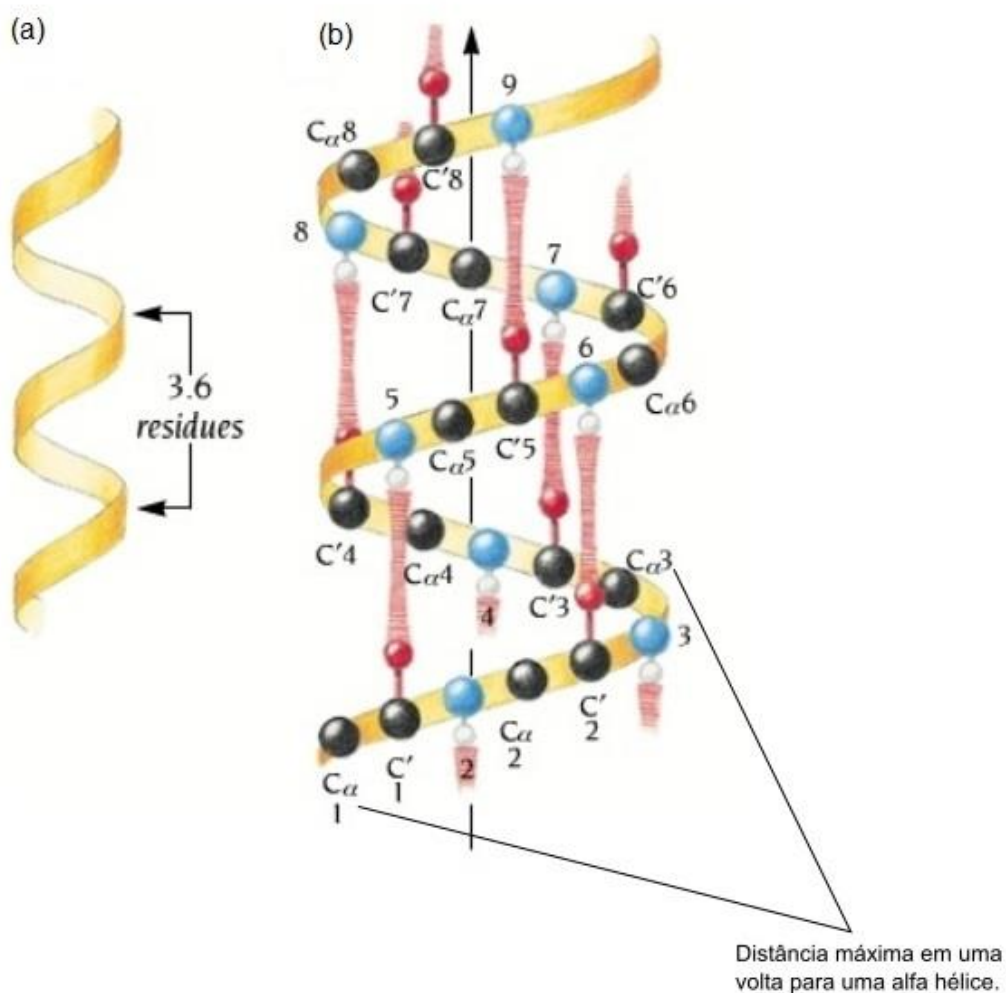
O gráfico 4, permite a comparação direta com o gráfico teórico, contudo em uma tentativa de facilitar a visualização e tornar mais nítida as semelhanças foi feita uma normalização da curva que é apresentada no Gráfico 5.

Gráfico 5: normalização da curva da proteína carboxil esterase.



Embora a forma da hélice seja a apresentada no Gráfico 3, ainda é necessário determinar a distância máxima que uma hélice permite em uma única volta. A forma com que a distância foi calculada é apresentada na Figura 40.

Figura 40: (a) Representação simplificada da alfa hélice; (b) Representação da alfa hélice com carbonos alfa.



Fonte: Google imagens.

O cálculo dessa distância foi realizado para uma hélice da estrutura da proteína anidrase carbônica (12CA), a hélice escolhida foi a quinta hélice que vai do resíduo prolina de número 155 até a isoleucina de número 167. O resultado pode ser observado na tabela a seguir:

Tabela 4: Cálculo da maior distância para uma alfa hélice.

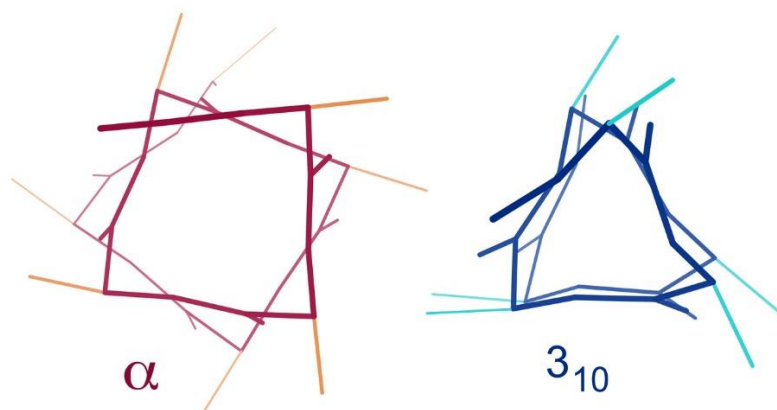
	Dx	Dy	Dz	Distância em Angstroms
D 1-3	4,653	2,711	0,495	5,408
D 2-4	3,410	-2,210	-4,239	5,872
D 3-5	0,105	2,085	-5,157	5,564
D 4-6	3,843	3,878	0,000	5,460
D 5-7	4,465	-3,086	0,565	5,457
D 6-8	0,104	-2,878	-4,662	5,480
D 7-9	1,305	3,753	-4,026	5,657
D 8-10	5,422	1,502	0,989	5,712
D 9-11	4,224	-2,877	-2,120	5,533
D 10-12	1,165	2,460	-4,309	5,097
D 11-13	2,663	4,587	1,908	5,637
Média				5,534
Desvio Padrão				0,199

A Tabela 2, apresenta os valores calculados para a distância máxima que ocorre em uma alfa hélice. Os termos Dx, Dy e Dz, representam as diferenças entre as abscissas, ordenadas e cotas respectivamente. Os termos D 1-3, D 2-4, e todos os demais são as distâncias entre os carbonos alfa, no primeiro caso, D 1-3 é a distância entre os carbonos alfa 1 e 3. A distância foi calculada com base nos trabalhos de Pauling que descobriu que em uma volta da hélice haviam aproximadamente 4 resíduos. Dessa forma os cálculos foram realizados obedecendo uma distância de aproximadamente 180° entre os carbonos alfa. É possível verificar que a distância máxima é de aproximadamente 5,5 Angstroms e que as oscilações dentro deste parâmetro indicam a presença de hélice na estrutura.

A estrutura em hélice 3-10, devido à distribuição dos carbonos alfa, apresentará valor médio para a maior distância em uma volta, menor que o calculado

para a estrutura da alfa hélice. A Figura 41 apresenta uma vista superior de duas hélices.

Figura 41: Hélices alfa e 3-10.



Fonte: Google imagens.

É possível observar na Figura 41 que a hélice 3_{10} é mais regular que a hélice alfa, e que quanto maior for a hélice, mais distorcida será a hélice alfa. Outro fato que pode ser observado é que a distância máxima entre os carbonos alfa em uma hélice é maior para a alfa. A hélice 3_{10} devido a sua regularidade forma, pela vista superior da hélice, um triângulo, o que implica em uma distância máxima menor.

Proteínas

Os dados principais para análise dos resultados podem ser observados na tabela 5. Nela é possível observar o comprimento da cadeia, o tamanho, o início e o fim do nó. Esses dados são importantes para comparação entre os dados apresentados.

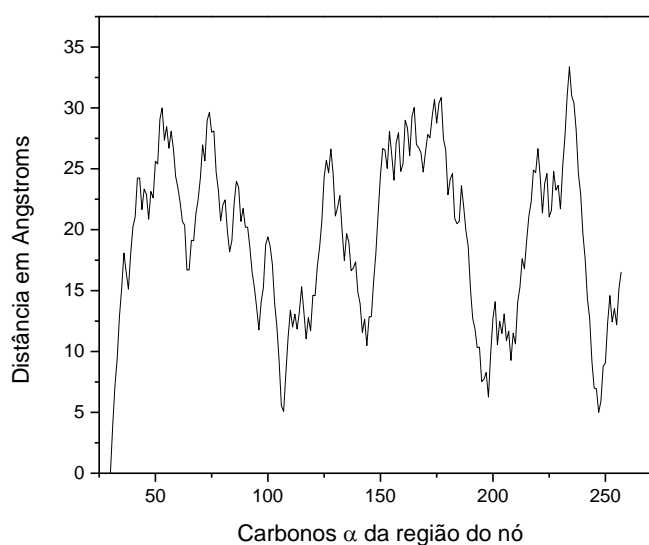
Tabela 5: Dados das proteínas analisadas.

Proteína	Código pdb	Tamanho	Tamanho do nó	Início e fim do nó
Anidrase Carbônica	12CA	255	228	30 – 258
Metiltransferase	1GZ0	235	41	173 – 214
Transcarbamilase	1JS1	324	67	171 – 238
S-adenosilmetionina sintetase	1FUG	383	266	10 – 276
Neurotoxina	1TSK	35	Não identificado	-
Proteína hipotética	2EFV	63	Não identificado	-
Carboxilesterase	1R1D	237	Não se aplica	Não se aplica

Anidrase carbônica

Para a proteína Anidrase Carbônica de código pdb: 12CA, o gráfico de distâncias entre os carbonos alfa na região do nó, com relação ao primeiro carbono alfa, está apresentado no Gráfico 6.

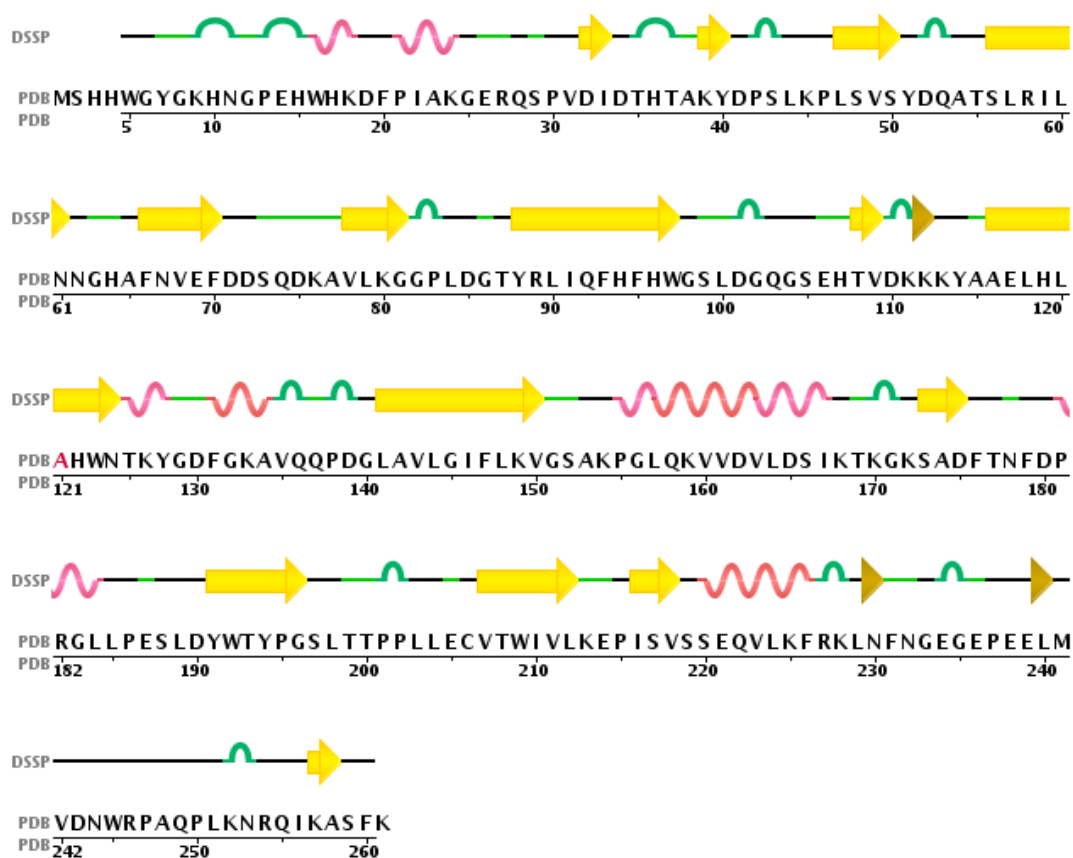
Gráfico 6: Gráfico de distâncias da proteína Anidrase Carbônica (12CA) com relação ao primeiro carbono alfa da região do nó.



Este gráfico não se assemelha ao proposto na equação, isso ocorre devido ao tamanho da região do nó, quanto maior for esta região, mais distante ela se tornará da equação. Isso ocorre devido as interações que ocorrem ao longo da cadeia, essas interações são importantes para manter as conformações na estrutura, porém tais interações deformam a estrutura, afastando-a da equação proposta.

A Figura 42, apresenta o sequenciamento dos resíduos e as estruturas secundárias que estão presentes na cadeia.

Figura 42: Sequência de resíduos e estruturas secundárias para a proteína Anidrase Carbônica.



Legenda

- E: beta strand
- T: turn
- empty: no secondary structure assigned
- G: 3/10-helix
- B: beta bridge
- S: bend
- H: alpha helix

Esta imagem é importante para comparação direta do gráfico de distâncias (Gráfico 2) com a estrutura secundária da proteína. É possível verificar as oscilações curtas nas regiões que apresentam hélices e as longas nas regiões que apresentam folhas betas ou voltas.

Ainda sobre a região do nó da proteína foi realizado um levantamento para verificar a polaridade dos resíduos nesta região, o resultado encontra-se na Tabela 6.

Tabela 6: Polaridade da região do nó da Anidrase Carbônica.

Aminoácidos	Soma	Porcentagem do total
Apolar	112	49,1
Polar Neutro	54	23,7
Polar Positivo	33	14,5
Polar Negativo	29	12,7

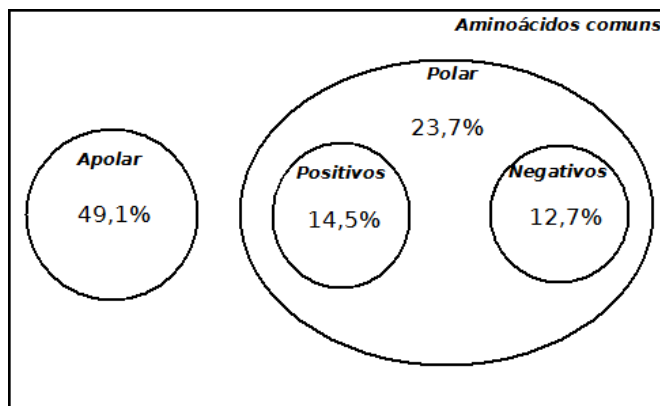
A tabela 7 a seguir mostra os mesmos dados, porém, para a estrutura completa.

Tabela 7: Polaridade da estrutura completa da anidrase carbônica.

Aminoácidos	Soma	Porcentagem do total
Apolar	125	48,1
Polar Neutro	60	23,7
Polar Positivo	43	16,5
Polar Negativo	32	12,3

A Figura 43 apresenta um diagrama de Venn com as porcentagens representadas para facilitar a visualização.

Figura 43: Diagrama de Venn para representação da polaridade da Anidrase carbônica na região do nó.



A contagem dos resíduos está representada na Tabela 8.

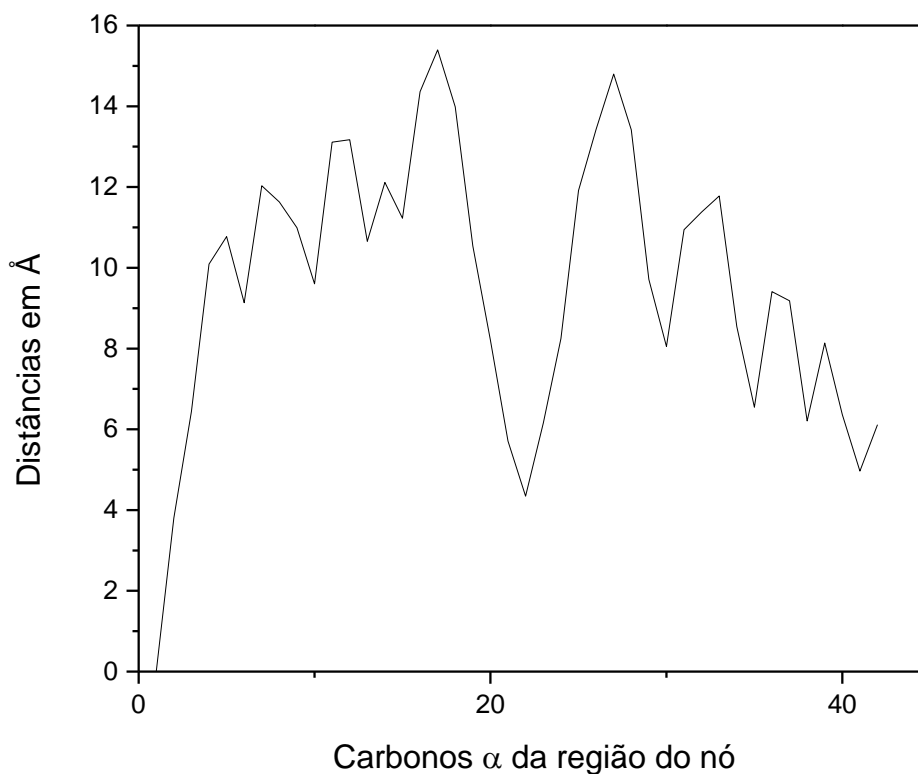
Tabela 8: Distribuição de resíduos na Anidrase Carbônica na região do nó.

Aminoácido	Quantidade
Glicina – gly	18
alanina – ala	13
leucina – leu	26
valina – val	16
isoleucina - ile	8
prolina - pro	15
fenilalanina - phe	10
metionina - met	1
triptofano - trp	5
treonina - thr	12
serina - ser	15
cisteína - cys	1
tirosina - tyr	7
asparagina - asn	9
glutamina - gln	10
lisina – lys	20
arginina - arg	6
histidina - his	7
aspartato - asp	18
glutamato - glu	11

Metiltransferase

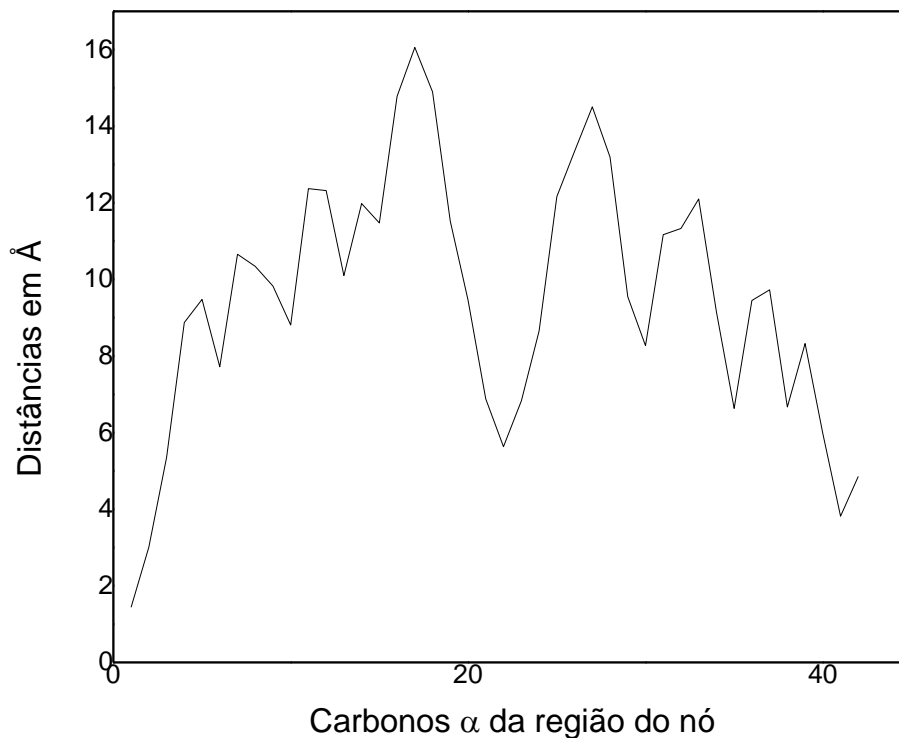
A proteína Metiltransferase de código pdb 1GZ0, foi analisada segundo o cálculo de distâncias dos carbonos alfa com relação ao primeiro carbono alfa e com relação ao ponto médio geométrico, que foi calculado com base nas coordenadas x, y e z, dos carbonos alfa da região do nó, os Gráficos 7 e 8 apresentam essas curvas.

Gráfico 7: Gráfico de distâncias da Proteína Metiltransferase com relação ao primeiro carbono alfa da região do nó.



Este gráfico apresenta a curva obtida pelo cálculo de distâncias com relação ao primeiro carbono alfa da região do nó. Para esta proteína a região do nó compreende apenas 42 resíduos de aminoácidos.

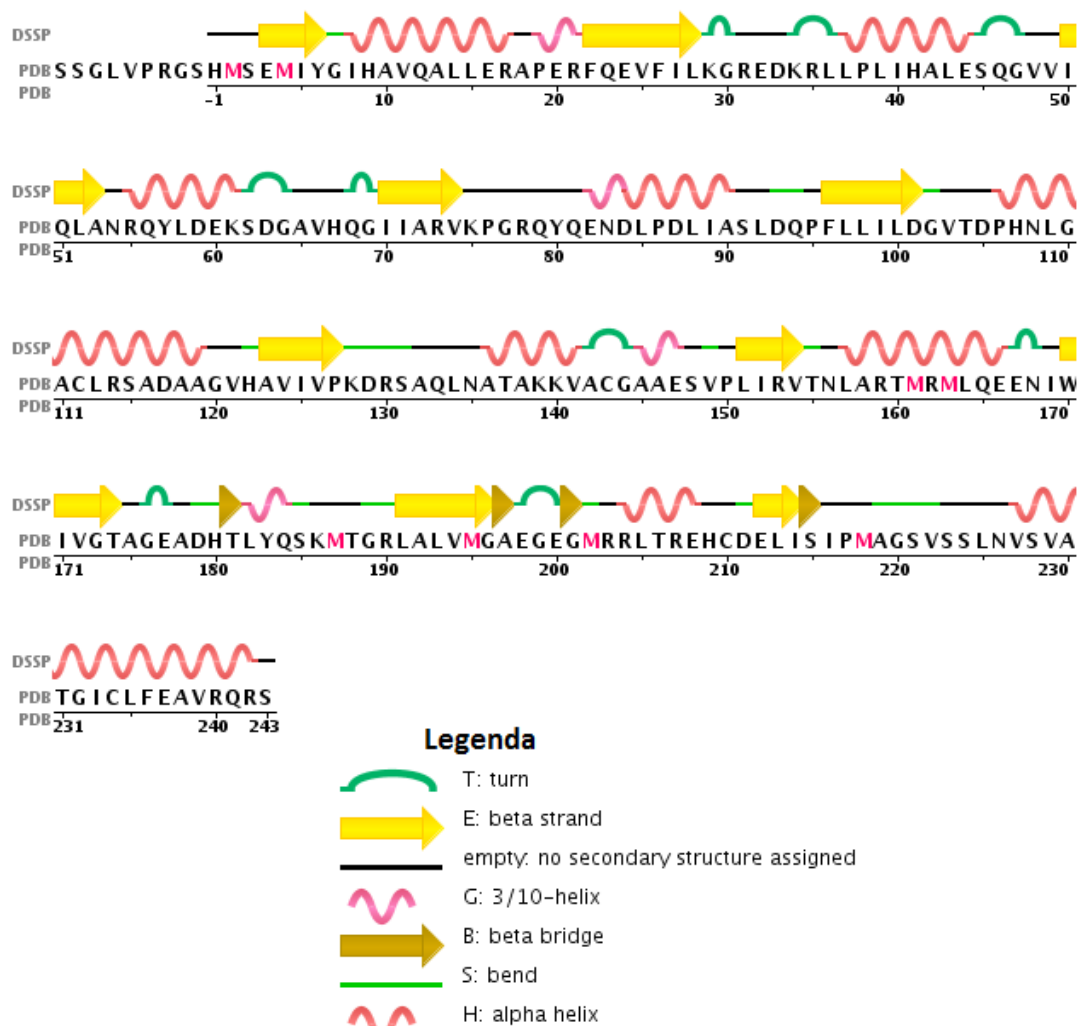
Gráfico 8: Gráfico de distâncias da Proteína Metiltransferase com relação ao ponto médio da região do nó.



Para o caso da Metiltransferase, a variação do gráfico com relação ao primeiro carbono alfa e com relação ao ponto médio foi pequena. Essa pequena diferença só é possível devido ao tamanho reduzido do nó, no caso da Anidrase carbônica onde o nó apresenta 228 resíduos de aminoácidos essa diferença fica muito evidente.

A Figura 35 apresenta o sequenciamento dos resíduos e as estruturas secundárias que estão presentes na cadeia.

Figura 44: Sequência de resíduos e estruturas secundárias para a proteína Metiltransferase.



Fonte: Adaptado de <http://www.rcsb.org>.

A comparação direta com o gráfico de distâncias é possível, porém é necessário ressaltar que a Figura 44 apresenta a cadeia completa da proteína e no gráfico de distâncias foi calculado apenas a região do nó da proteína.

A quantidade de resíduos na região do nó e a distribuição percentual de polaridade são apresentados nas Tabelas 9, 10, 11 e na Figura 36 que apresenta um diagrama de Venn.

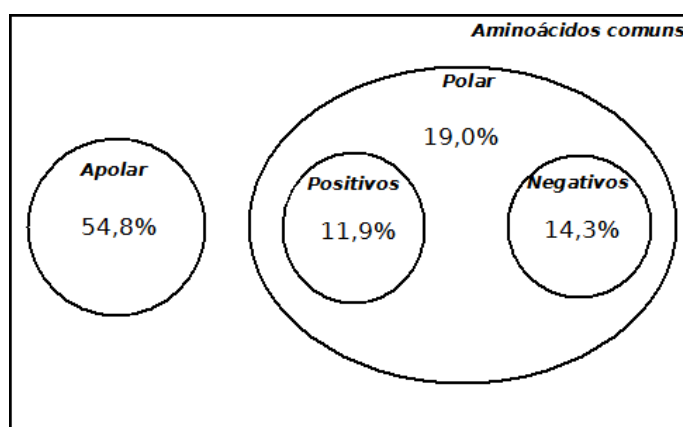
Tabela 9: Polaridade da região do nó da Metiltransferase.

Aminoácidos	Soma	Porcentagem do total
Apolar	23	54,8
Polar neutro	8	19,0
Polar Positivo	5	11,9
Polar Negativo	6	14,3

Tabela 10: Polaridade da estrutura completa da metiltransferase.

Aminoácidos	Soma	Porcentagem do total
Apolar	127	51,8
Polar neutro	54	22,0
Polar Positivo	35	14,3
Polar Negativo	29	11,8

Figura 45: Diagrama de Venn para representação da polaridade da Metiltransferase.



Fonte: Elaborado pelo autor.

Tabela 11: Distribuição de resíduos na Metiltransferase na região do nó.

Aminoácidos	Quantidade
Glicina - gly	6
alanina - ala	4
leucina - leu	4
valina - val	2
isoleucina - ile	2
prolina - pro	0
fenilalanina - phe	0
metionina - met	4
triptofano - trp	1
treonina - thr	3
serina - ser	1
cisteína - cys	0
tirosina - tyr	1

asparagina - asn	1
glutamina - gln	2
lisina - lys	1
arginina - arg	3
histidina - his	1
aspartato - asp	1
glutamato - glu	5

Transcarbamilase

A proteína transcarbamilase de código pdb 1JS1, foi analisada com relação à região do nó. Foram feitas medidas de distâncias com relação ao primeiro carbono alfa e com relação ao ponto médio geométrico. Os dados obtidos foram plotados e são apresentados nos Gráficos 9 e 10.

Gráfico 9: Distribuição de distâncias para a região do nó com relação ao primeiro carbono alfa para a Transcarbamilase (1JS1).

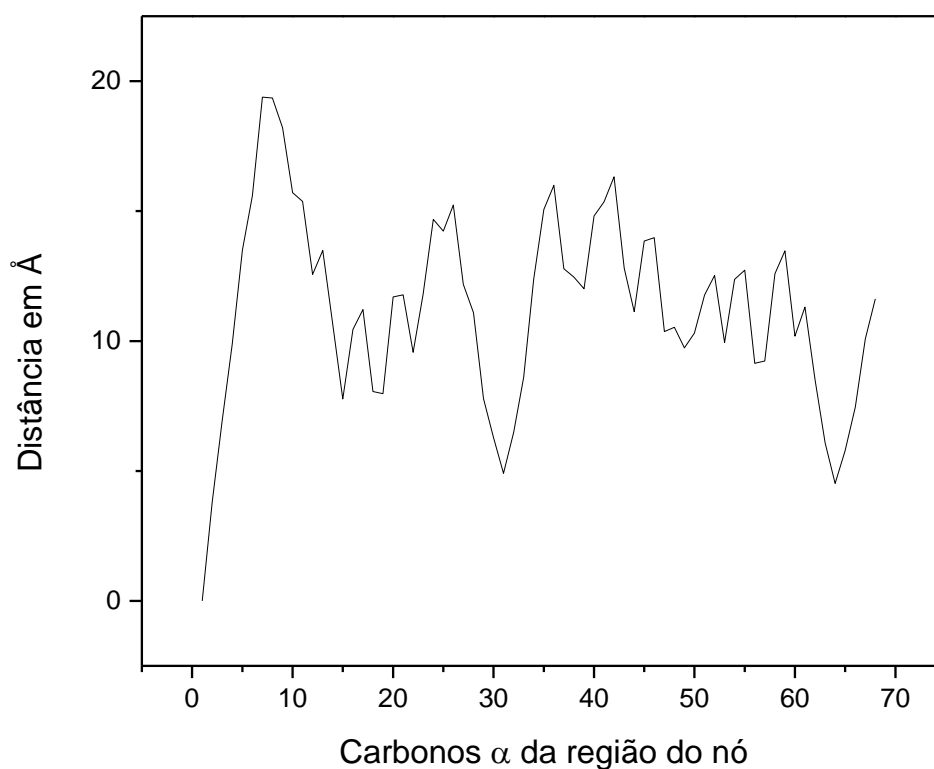
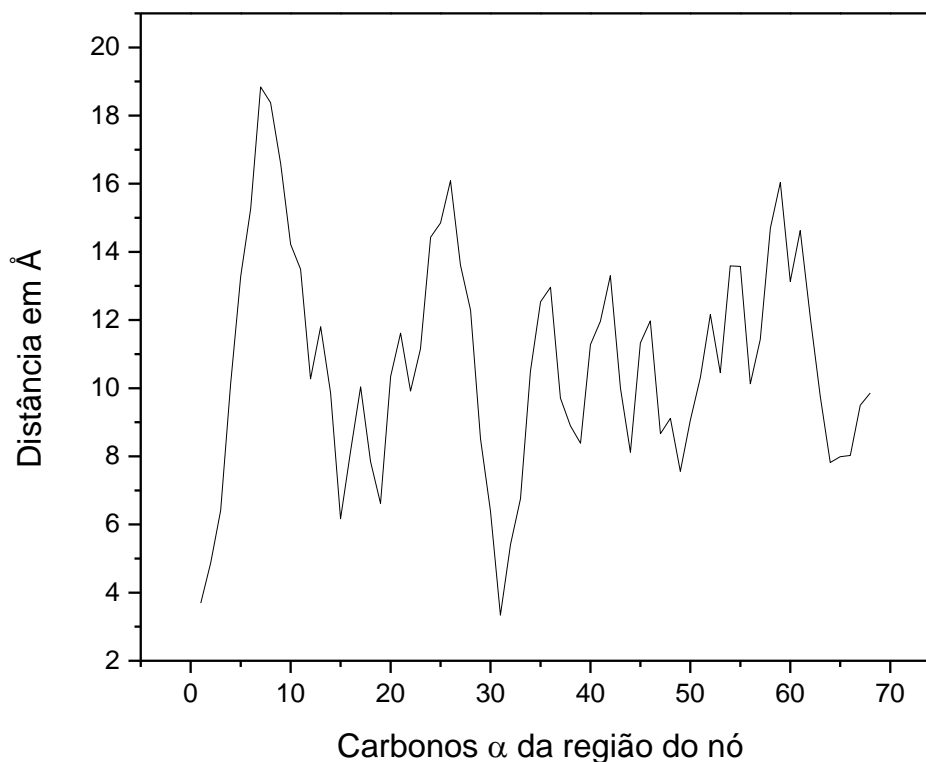


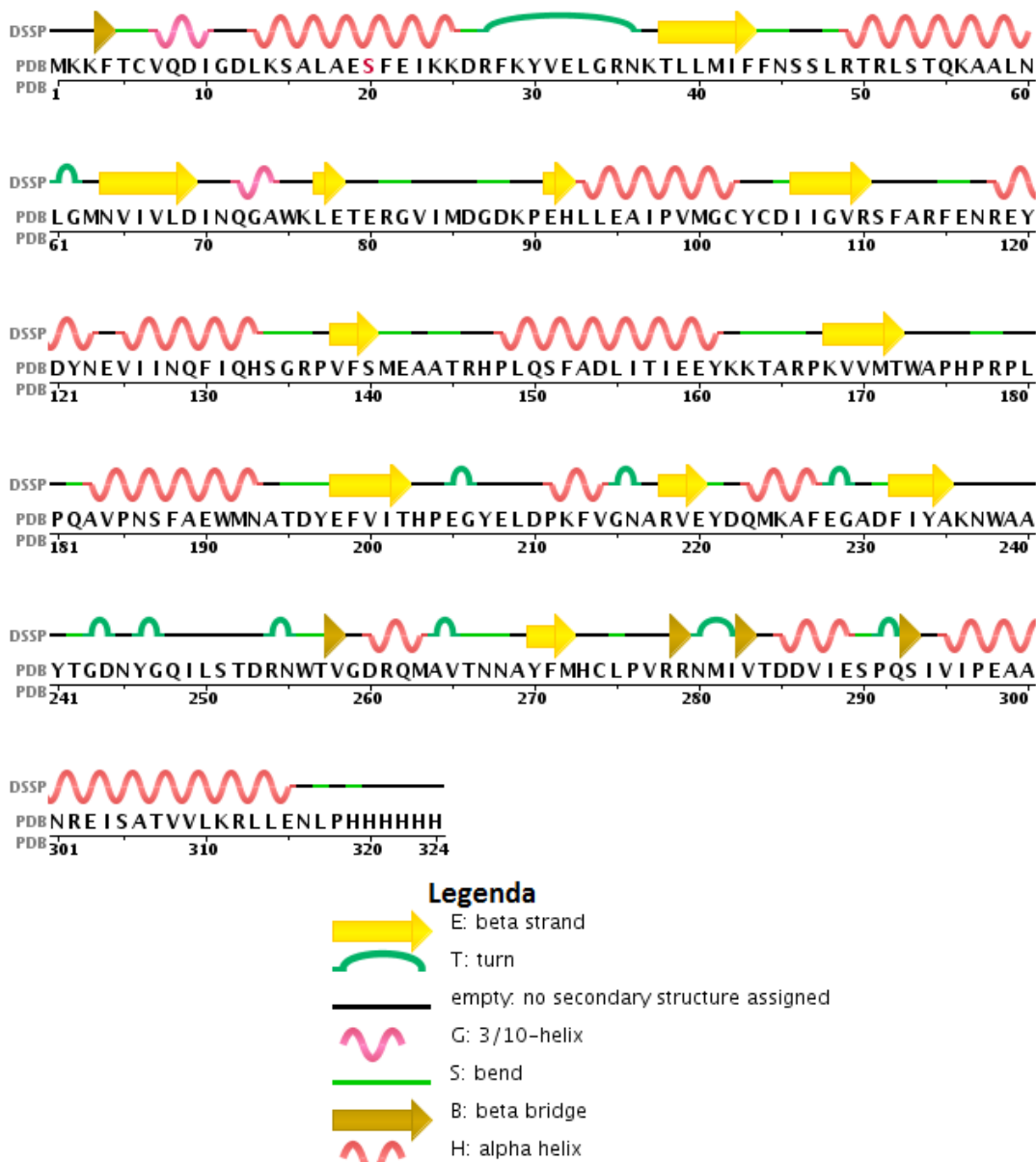
Gráfico 10: Distribuição de distâncias para a região do nó com relação ao ponto médio geométrico para a Transcarbamilase (1JS1).



A estrutura do nó da transcarbamilase apresenta 68 resíduos de aminoácidos, no entanto, o gráfico com relação ao primeiro carbono alfa e com relação ao ponto médio apresentam diferenças, isso evidencia que, a posição com que o nó é observado, influencia no resultado obtido.

A Figura 46 apresenta o sequenciamento dos resíduos e as estruturas secundárias que estão presentes na cadeia.

Figura 46: Sequência de resíduos e estruturas secundárias para a proteína Transcarbamilase.



Fonte: Adaptado de <http://www.rcsb.org>.

Da mesma forma que na Metiltransferase é necessário olhar somente para um trecho da cadeia apresentada na Figura 46, isso por que a estrutura representada é a cadeia completa da proteína Transcarbamilase, e o gráfico de distâncias apresenta apenas a região do nó da estrutura.

A quantidade de resíduos na região do nó e a distribuição percentual de polaridade são apresentados nas Tabelas 12, 13, 14 e na Figura 47 que apresenta um diagrama de Venn.

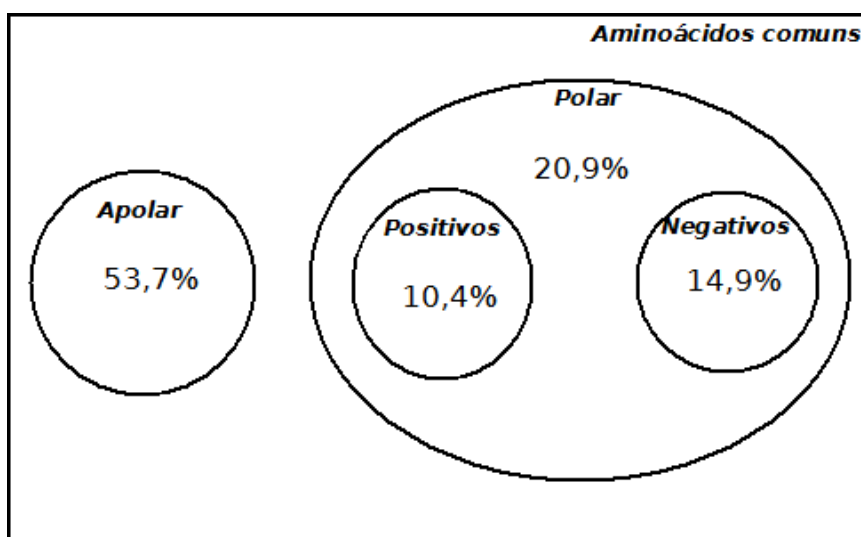
Tabela 12: Polaridade da região do nó da transcarbamilase.

Aminoácidos	Soma	Porcentagem do total
Apolares	36	53,7
Polares Neutros	14	20,9
Polares Positivos	7	10,4
Polares Negativos	10	14,9

Tabela 13: Polaridade da estrutura completa da proteína transcarbamilase.

Aminoácidos	Soma	Porcentagem do total
Apolares	158	48,8
Polares Neutros	77	23,8
Polares Positivos	48	14,8
Polares Negativos	141	12,7

Figura 47: Diagrama de Venn para representação da polaridade da Transcarbamilase.



Fonte: Elaborado pelo autor.

Tabela 14: Distribuição de resíduos na Transcarbamilase.

Aminoácidos	Quantidade
Glicina - gly	3
alanina - ala	8
leucina - leu	2
valina - val	4
isoleucina - ile	2
prolina - pro	7
fenilalanina - phe	5
metionina - met	3
triptofano - trp	2
treonina - thr	3
serina - ser	1

cisteína - cys	0
tirosina - tyr	4
asparagina - asn	4
glutamina - gln	2
lisina - lys	3
arginina - arg	2
histidina - his	2
aspartato - asp	4
glutamato - glu	6

S - Adenosilmetionina Sintetase

A proteína S - Adenosilmetionina Sintetase de código pdb 1FUG, foi analisada com relação à região do nó. Foram feitas medidas de distâncias com relação ao primeiro carbono alfa e com relação ao ponto médio geométrico, esses dados foram plotados e são apresentados nos Gráficos 11 e 12.

Gráfico 11: Distribuição de distâncias para a região do nó com relação ao primeiro carbono alfa para a S - Adenosilmetionina Sintetase (1FUG).

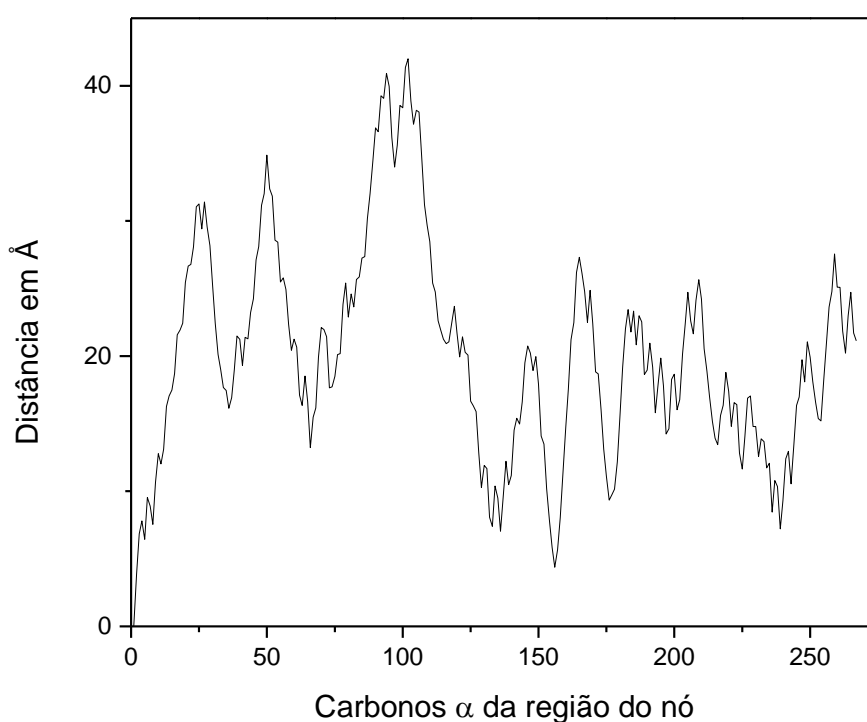
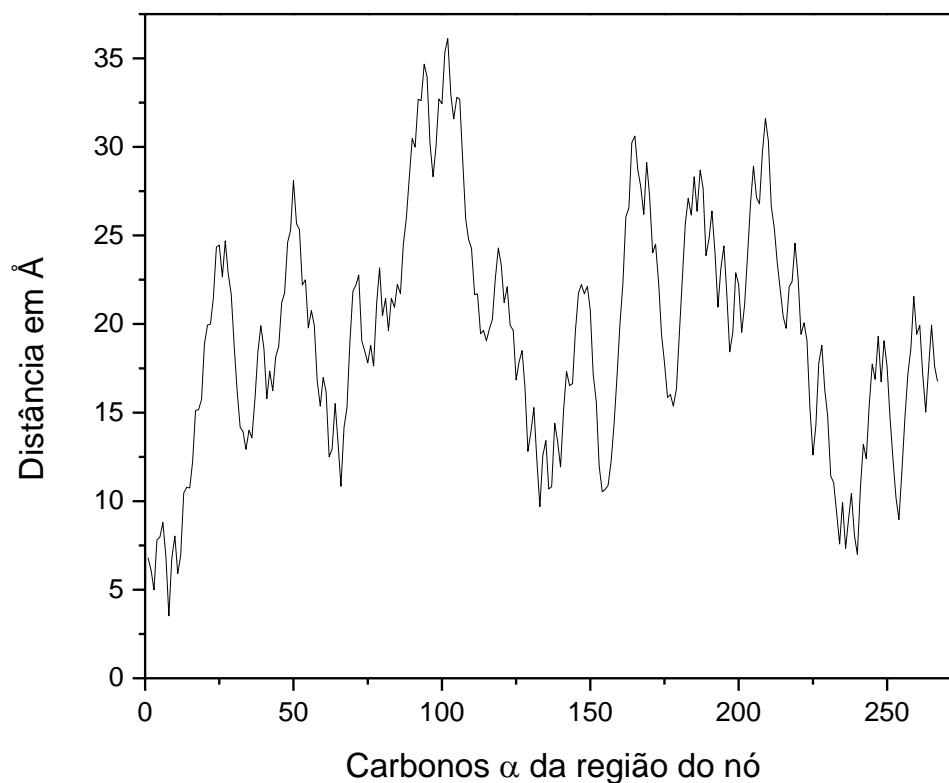
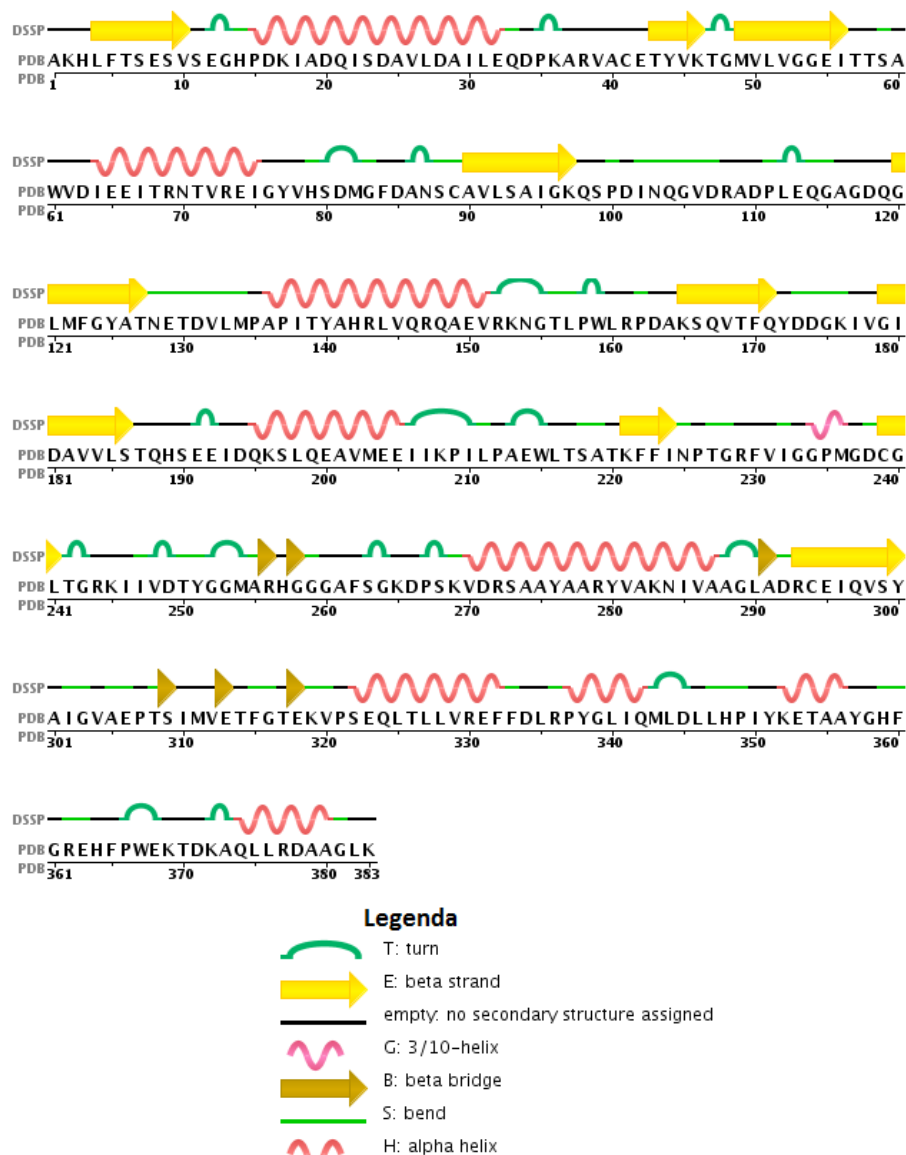


Gráfico 12: Distribuição de distâncias para a região do nó com relação ao ponto médio geométrico para a S - Adenosilmetionina Sintetase (1FUG).



A proteína S - Adenosilmetionina Sintetase apresenta 267 resíduos de aminoácidos na região do nó, esta quantidade representa quase a totalidade dos resíduos presentes em toda a estrutura da proteína. Com essa quantidade elevada de resíduos na região do nó os gráficos 11 e 12 se tornam bastante diferentes. Isso mostra mais uma vez que o ponto de observação da estrutura influencia no resultado obtido, e que quanto maior for a estrutura analisada maior será a diferença no resultado obtido. A Figura 48 apresenta o sequenciamento dos resíduos e as estruturas secundárias que estão presentes na cadeia.

Figura 48: Sequência de resíduos e estruturas secundárias para a proteína S - Adenosilmetionina Sintetase



Fonte: Adaptado de <http://www.rcsb.org>.

A quantidade de resíduos na região do nó e a distribuição percentual de polaridade são apresentados nas Tabelas 15, 16, 17 e na Figura 40 que apresenta um diagrama de Venn.

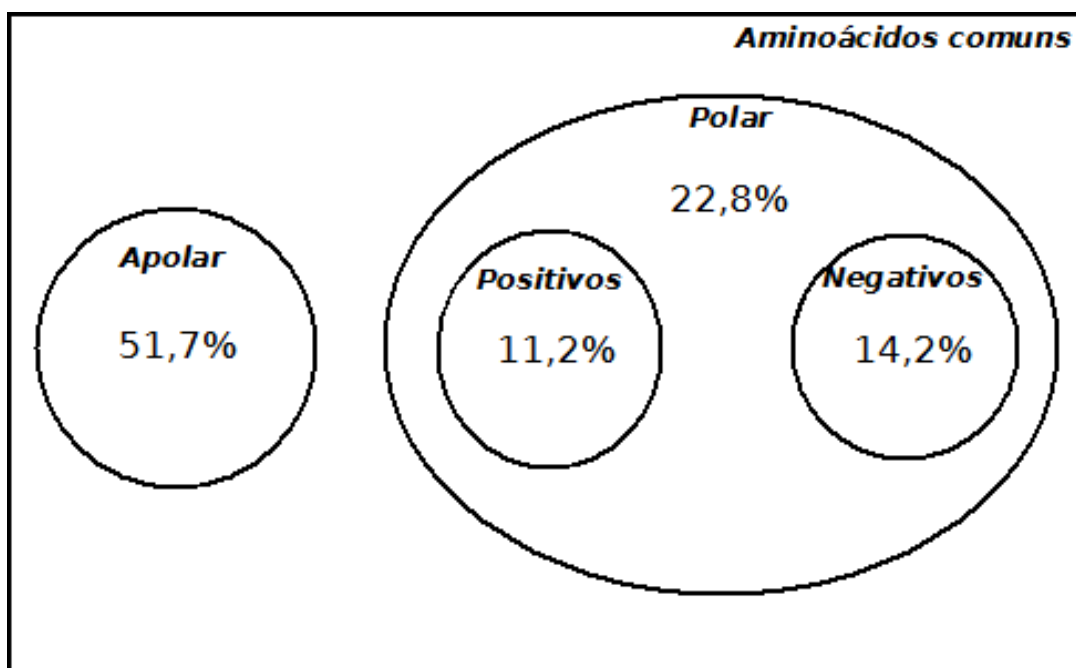
Tabela 15: Polaridade da região do nó da S - Adenosilmetionina Sintetase.

Aminoácidos	Soma	Porcentagem do total
Apolares	138	51,7
Polares Neutros	61	22,8
Polares Positivos	30	11,2
Polares Negativos	38	14,2

Tabela 16: Polaridade da estrutura completa da proteína s-adenosilmetionina sintetase.

Aminoácidos	Soma	Porcentagem do total
Apolares	199	52,0
Polares Neutros	84	21,9
Polares Positivos	47	12,3
Polares Negativos	53	13,8

Figura 49: Diagrama de Venn para representação da polaridade da S - Adenosilmetionina Sintetase.



Fonte: Elaborado pelo autor

Tabela 17: Distribuição de resíduos na S - Adenosilmetionina Sintetase.

Aminoácidos	Quantidade
Glicina - gly	27
alanina - ala	24
leucina - leu	15
valina - val	22
isoleucina - ile	20
prolina - pro	13
fenilalanina - phe	7
metionina - met	7
triptofano - trp	3
treonina - thr	17
serina - ser	15
cisteína - cys	3
tirosina - tyr	7
asparagina - asn	6

glutamina - gln	13
lisina - lys	13
arginina - arg	12
histidina - his	5
aspartato - asp	22
glutamato - glu	16

Toxina isolada de escorpião

A proteína e também neurotoxina TsK de código pdb 1TSK foi analisada em toda a sua extensão, isso porque não havia dados para dizer o começo ou o fim do nó na estrutura. Foram realizados os cálculos para determinação das distâncias entre os carbonos alfa com relação ao primeiro carbono e com relação ao ponto médio das coordenadas de todos os carbonos alfa da estrutura. Os resultados estão apresentados nos Gráficos 13 e 14, apresentados a seguir.

Gráfico 13: Distribuição de distâncias da proteína TsK com relação ao primeiro carbono alfa.

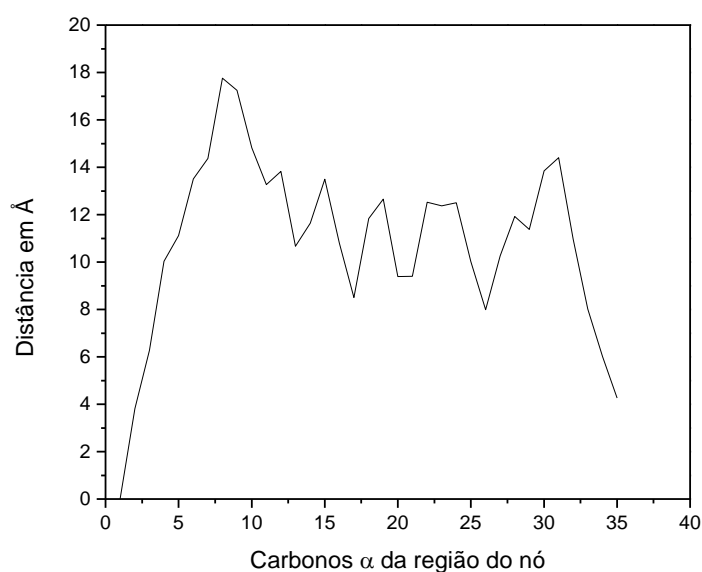
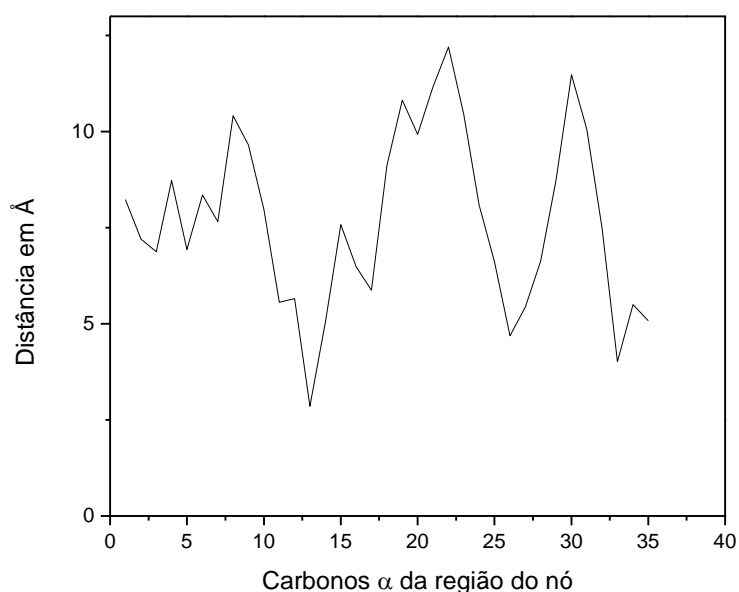


Gráfico 14: Distribuição de distâncias da proteína TsK com relação ao ponto médio geométrico.

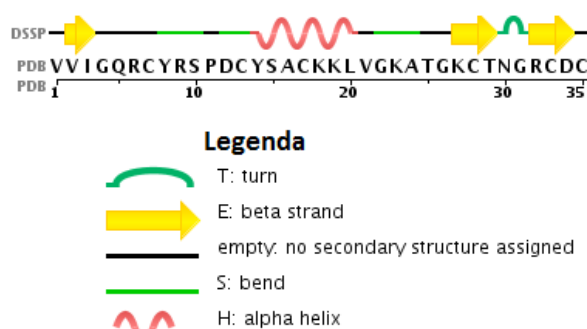


A estrutura da proteína TsK apresentada no gráfico 13 tem características de um nó, porém o trecho da estrutura que vai do carbono alfa 15 ao 30 pode gerar dúvidas na análise do gráfico.

O Gráfico 14 apresenta os dois picos e um vale característicos do nó 3₁, porém, existe um último pico que provavelmente se encontra fora da região do nó desta proteína, desta forma pode-se dizer que a região do nó da estrutura vai de primeiro a aproximadamente vigésimo sexto resíduo da estrutura, uma vez que os gráficos 13 e 14 foram feitos a partir dos cálculos da estrutura completa da proteína.

A Figura 50 apresenta o sequenciamento dos resíduos e as estruturas secundárias que estão presentes na cadeia.

Figura 50: Sequência de resíduos e estruturas secundárias para a proteína extraída do escorpião (TsK).



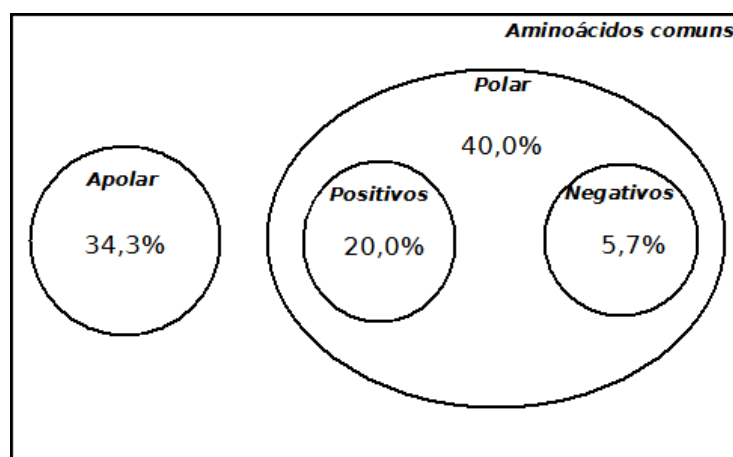
Fonte: Adaptado de <http://www.rcsb.org>.

A quantidade de resíduos na região do nó e a distribuição percentual de polaridade são apresentados nas Tabelas 18, 19 e na Figura 51 que apresenta um diagrama de Venn.

Tabela 18: Polaridade da região do nó da TsK.

Aminoácidos	Soma	Porcentagem do total
Apolares	12	34,3
Polares Neutros	14	40,0
Polares Positivos	7	20,0
Polares Negativos	2	5,7

Figura 51: Diagrama de Venn para representação da polaridade da TsK.



Fonte: Elaborado pelo autor.

Tabela 19: Distribuição de resíduos na proteína TsK.

Aminoácidos	Quantidade
Glicina - gly	4
alanina - ala	2
leucina - leu	1
valina - val	3
isoleucina - ile	1
prolina - pro	1
fenilalanina - phe	0
metionina - met	0
triptofano - trp	0
treonina - thr	2
serina - ser	2
cisteína - cys	6
tirosina - tyr	2
asparagina - asn	1

glutamina - gln	1
lisina - lys	4
arginina - arg	3
histidina - his	0
aspartato - asp	2
glutamato - glu	0

Proteína hipotética do Methanocaldococcus jannaschii

A proteína hipotética de código pdb 2EFV ainda não possui classificação e nem função sua estrutura foi isolada do organismo *Methanocaldococcus jannaschii*. Os Gráficos 15 e 16 apresentam os resultados dos cálculos de distribuição de distâncias com relação ao primeiro e com relação ao ponto médio geométrico, respectivamente.

Gráfico 15: Distribuição de distâncias da proteína hipotética do *Methanocaldococcus jannaschii* com relação ao primeiro carbono alfa.

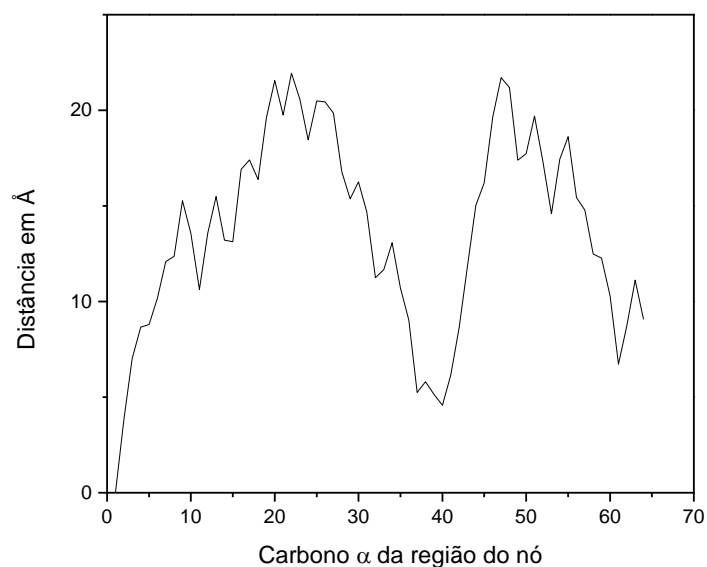
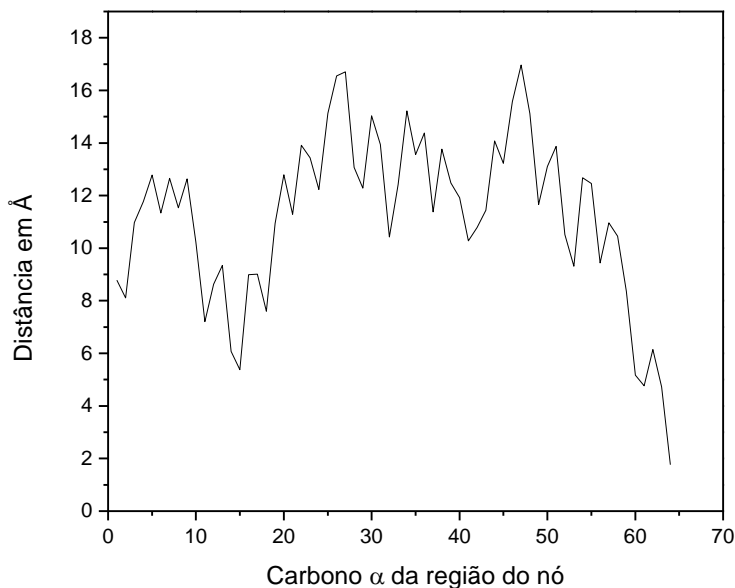


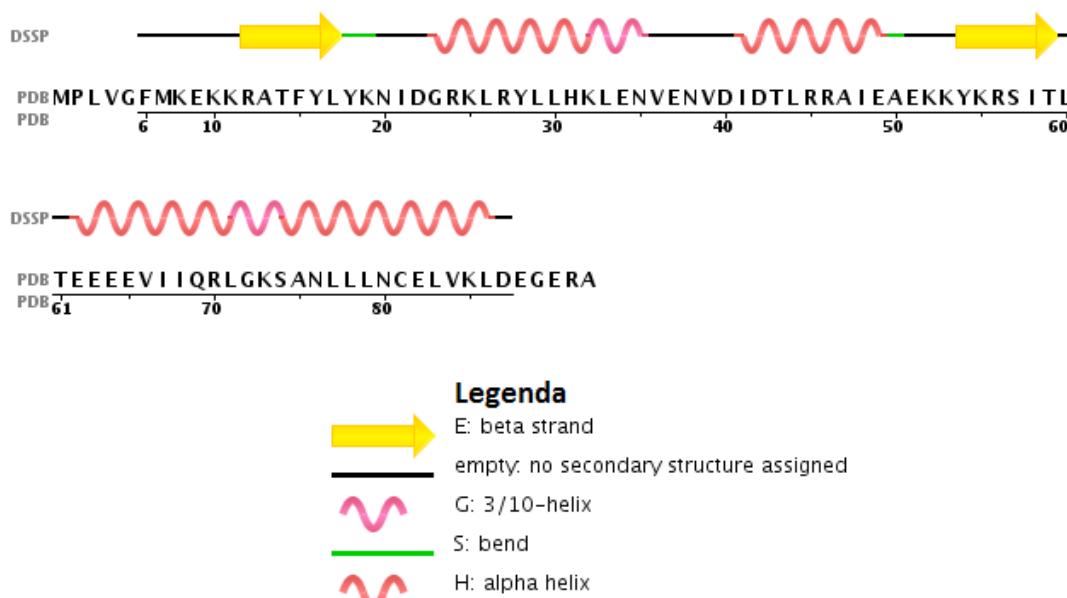
Gráfico 16: Distribuição de distâncias da proteína hipotética do *Methanocaldococcus jannaschii* com relação ao ponto médio geométrico.



O Gráfico 15 permite uma visualização clara de que o nó está presente na proteína, o mesmo não ocorre no Gráfico 16 onde as distâncias foram calculadas com relação ao ponto médio, isso permite concluir que o ponto de referência para o cálculo de distâncias é de extrema importância um vez que o ideal é olhar de um ponto que permita equidistância para os pontos mais afastados da região do nó, sendo assim, nem sempre o primeiro carbono alfa ou o ponto médio serão o ponto ideal para o cálculo de distância, o que dificulta a visualização do nó na estrutura.

A Figura 52 apresenta o sequenciamento dos resíduos e as estruturas secundárias que estão presentes na cadeia.

Figura 52: Sequência de resíduos e estruturas secundárias para a proteína hipotética do *Methanocaldococcus jannaschii*.



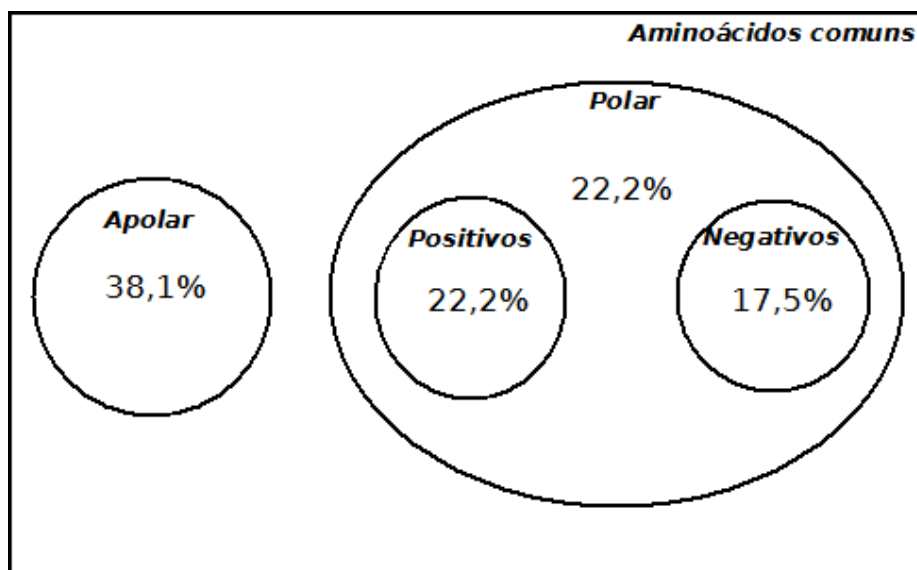
Fonte: Adaptado de <http://www.rcsb.org>.

A quantidade de resíduos na região do nó e a distribuição percentual de polaridade são apresentados nas Tabelas 20, 21 e na Figura 53 que apresenta um diagrama de Venn.

Tabela 20: Polaridade da região do nó da proteína hipotética do *Methanocaldococcus jannaschii*.

Aminoácidos	Soma	% do total
Apolares	24	38,1
Polares Neutros	14	22,2
Polares Positivos	14	22,2
Polares Negativos	11	17,5

Figura 53: Diagrama de Venn para representação da polaridade da proteína hipotética do *Methanocaldococcus jannaschii*.



Fonte: Elaborado pelo autor.

Tabela 21: Distribuição de resíduos na proteína hipotética do *Methanocaldococcus jannaschii*.

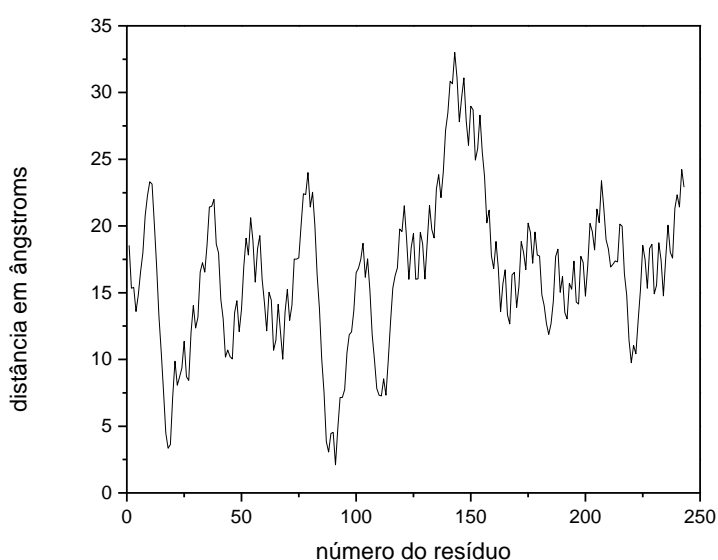
Aminoácido	Quantidade
Glicina - gly	2
alanina - ala	3
leucina - leu	10
valina - val	3
isoleucina - ile	6
prolina - pro	0
fenilalanina - phe	0
metionina - met	0
triptofano - trp	0
treonina - thr	3
serina - ser	2
cisteína - cys	0
tirosina - tyr	4
asparagina - asn	4
glutamina - gln	1
lisina - lys	7
arginina - arg	6
histidina - his	1
aspartato - asp	3
glutamato - glu	8

Como foi observado em todas as proteínas apresentadas, a diferença de polaridade entre a região do nó e a estrutura completa não varia de forma expressiva, esse resultado permite concluir que a estabilidade do nó se dá através das interações que ocorrem ao longo da cadeia e que a estabilidade não está relacionada a diferença de polaridade da região do nó, uma vez que a polaridade é praticamente a mesma que em toda a estrutura.

Carboxil esterase

A proteína carboxil esterase de código pdb 1R1D foi escolhida para comparar os resultados obtidos e validar a equação obtida neste trabalho. Todos os dados apresentados são resultados de proteínas que contém nó em sua estrutura, os gráficos exibem a região do nó de cada uma dessas proteínas. A carboxil esterase não apresenta nó em sua estrutura o que permite verificar se o gráfico de distâncias obtido na equação e nas proteínas com nó são representativos ou não. O Gráfico 17 apresenta a distribuição de distâncias com relação ao ponto médio da cadeia e a Figura 54 apresenta o sequenciamento dos resíduos e as estruturas secundárias presentes na cadeia.

Gráfico 17: Distribuição de distâncias da proteína carboxil esterase.



5. CONCLUSÕES

Neste trabalho de mestrado foram investigadas as estruturas secundárias e terciárias de algumas proteínas. O trabalho foi desenvolvido para aproximar a estrutura real de uma proteína com nó a uma equação matemática que pudesse representá-la. Trabalhos anteriores na área evidenciam que a maioria dos estudos realizados, tem foco no enovelamento da proteína e a energia envolvida no processo. Alguns trabalhos preocupam-se em desvendar os caminhos possíveis para que o enovelamento ocorra. O diferencial desta pesquisa se encontra no fato de tentar explicar a estrutura terciária de uma proteína através do uso de equações matemáticas.

Algumas proteínas apresentam nós em suas estruturas e o nó mais comum é o *trefoil* ou nó 3_1 . Este nó tem uma equação matemática bem definida e foi a partir desta que foram desenvolvidos os estudos realizados neste trabalho. O nó matemático, embora presente na estrutura da proteína, não pode ser comparado diretamente com a estrutura da proteína, isso por que o nó é uma estrutura fechada sem começo ou fim definido, enquanto a proteína é uma estrutura aberta com começo e fim. Desta forma foram necessárias algumas alterações na equação matemática do nó original e no intervalo do domínio considerado para esta nova equação, que foi de $2/3$ com relação ao intervalo completo, assim o novo domínio foi de $0 \rightarrow 2/3\pi$ e $4/3\pi \rightarrow 2\pi$. Este “vazio” no meio do intervalo foi necessário para adaptar a equação à estrutura da proteína.

As equações obtidas foram comparadas com a estrutura real da proteína através dos gráficos de distâncias que foram obtidos através das coordenadas dos carbonos alfas da região do nó e pelo fracionamento do intervalo nas equações. Os resultados mostram que quanto menor for a região do nó da proteína, mais próxima da equação proposta ela será, e para proteínas com regiões maiores, devidos às interações ao longo da cadeia, maior será a diferença com relação à equação.

Com este método foi possível identificar a estrutura secundária das proteínas, mais especificamente a Hélice. Partindo dos estudos desenvolvidos por Pauling foi possível determinar no gráfico de distâncias o comportamento da estrutura em hélice, neste trabalho, foi demonstrado o comportamento para uma alfa hélice real da proteína Anidrase Carbônica (12CA), esta apresentou oscilações de

aproximadamente 5,5 angstroms, o que permite concluir que qualquer oscilação próxima deste valor caracteriza a presença de alfa hélice na estrutura, sendo a hélice 3_{10} com oscilações menores e a π hélice com oscilações maiores. Ainda foi utilizada uma hélice da proteína carboxi esterase (1R1D) para comparação com os resultados teóricos, permitindo a validação da proposta feita para o comportamento da estrutura em hélice no gráfico de distâncias. As estruturas em folhas apresentam grande dificuldade para interpretação pois podem ser facilmente confundidas com as voltas, assim sendo, pode-se apenas dizer que as oscilações longas no gráfico de distâncias são folhas ou voltas, não podendo estas serem identificadas com certeza.

Dentre as proteínas estudadas as que mais se aproximam da equação proposta são a *Metiltransferase*, a *Transcarbamilase*, a *TsK* e a *Proteína hipotética 2EFV*, com 42, 68, 35 e 63 resíduos de aminoácidos na região do nó respectivamente. As proteínas *Anidrase Carbônica* e a *S-Adenosilmetionina sintetase* por serem maiores, 228 e 267 resíduos respectivamente, apresentaram grande variação devido à presença de interações ao longo da cadeia, o que deforma a estrutura com relação ao nó matemático. Pode-se concluir desses valores que quanto maior a região do nó da proteína menor será sua proximidade com a equação proposta devido às interações.

A análise dos diagrama de Venn para a distribuição da polaridade das proteínas na região do nó permite concluir que uma proteína com nó 3_1 apresenta predominantemente a mesma polaridade que a proteína completa. As diferenças observadas são muito pequenas, o que não gera uma diferença resultado expressiva, desta forma, pode-se dizer que a estabilidade do nó está ligada somente ao tamanho e às interações ao longo da cadeia pois a polaridade da região do nó é praticamente a mesma da proteína completa.

6. REFERÊNCIAS

ARAZAWA, D. T., KIMMEL, J. D., FINN, M. C., & FEDERSPIEL, W. J.; **Acidic Sweep Gas With Carbonic Anhydrase Coated Hollow Fiber Membranes Synergistically Accelerates CO₂ Removal From Blood**. Acta Biomaterialia, 25, p. 143-149, 2015.

BELL, A. E.; **Uncommon Amino Acids in Plants**. Febs letters, 64, p. 29 – 35, 1976.

CAMPBELL, M. K., & FARRELL, S. O.; **Bioquímica Básica**. Thomson. 1 edição, v. 1, 2006.

DONOHUE, J.; **Hydrogen Bonded Helical Configurations Of The Polypeptide Chain**. Chemistry, p. 470-478, 1953.

FU, Z., HU, Y., MARKHAM, G. D., & TAKUSAGAWA, F.; **Flexible Loop In The Structure Of S-Adenosylmethionine Synthetase Crystallized In The Tetragonal Modification**; J. Biomol Struct Dyn, p. 727-739, 1996.

GREER, J.; **Comparative Model Building Methods: Application To The Family Of The Mammalian Serine Proteases**. Proteins, 7, p. 317-334, 1990.

HARREL, C. R., GHOSH, B. K., & BOWDEN, R.; Simulation Using ProModel®. McGraw-Hill. 2000.

JAMES, M. J.; **MNG: An Algorithm For Determining The Conformation Of Polypeptide Segments In Proteins By Systematic Search**. Proteins, 1, p. 146-163, 1986.

KING, N. P., JACOBITZ, A. W., SAWAYA, M. R., GOLDSCHMIDT, L. & YEATES, T. O.; **Structure And Folding Of A Designed Knotted Protein**; Proc. Natl Acad. Sci. USA 107, p. 20732–20737, 2010.

KUHLMAN, B. et al.; Science 302, p. 1364–1368, 2003.

LOWENSON, J. D., KIM, E., YOUNG, S. G., & CLARKE, S.; **Limited Accumulation Of Damaged Proteins In L-Isoaspartyl (D-Aspartyl) O-Methyltransferase-Deficient Mice**. *The Journal Of Biological Chemistry*, 276(23), p. 20695-20702, 2001.

MALLAM, A. L. & JACKSON, S. E.; **Folding Studies on a Knotted Protein**; *J. Mol. Biol.* 366, p. 1409 – 1421, 2005.

PAULING, L., COREY, R. B., & BRANSON, H. R.; **The Structure Of Proteins: Two Hydrogen-Bonded Helical Configurations Of The Polypeptide Chain**. *Chemistry*, 37, p. 205-211, 1951.

PEREIRA, M. C. B.; **Bioquímica Das Proteínas: Fundamentos Estruturais e Funcionais**; UFV; p. 169 – 194 ,2014.

SAMUDRALA, R., PEDERSEN, J. T., ZHOU, H., LUO, R., FIDELIS, K., & MOULT, J.; **Confronting The Problem Of Interconnected Structural Changes In Comparative Modeling Of Proteins**. *Proteins*, 23, p. 327-336, 1995.

SHI, D., GALLEGOS, R., DEPONTE, J., MORIZONO, H., YU, X., ALLEWELL, N. M., TUCHMAN, M.; **Crystal Structure Of A Transcarbamylase-Like Protein From The Anaerobic Bacteroides Fragilis At 2.0 Å Resolution**. *J. Mol. Biol*(320), p. 899-908, 2002.

VÀSQUEZ, M.; **Modeling Side Chain Conformation**. *Curr Opin Struct Biol*, 6, p. 217-221, 1996.

VOET, D., VOET, J. G., & PRATT, C. W.; **Fundamentos de Bioquímica: A vida em nível molecular** (2ª ed.). Porto Alegre: Artmed, 2008.