

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM  
INSTITUTO DE CIÊNCIAS EXATAS - ICE  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

*PROBLEMA DE STEINER EUCLIDIANO APLICADO A  
MOLÉCULAS DE INTERESSE BIOLÓGICO*

Alcides de Castro Amorim Neto

MANAUS

2007

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM  
INSTITUTO DE CIÊNCIAS EXATAS - ICE  
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Alcides de Castro Amorim Neto

*PROBLEMA DE STEINER EUCLIDIANO APLICADO A  
MOLÉCULAS DE INTERESSE BIOLÓGICO*

Dissertação apresentada ao Programa de Pós  
Graduação em Matemática da Universidade Fe-  
deral do Amazonas, como requisito parcial para  
obtenção do grau de Mestre em Matemática,  
área de concentração em Otimização.

Orientador: Prof. Dr. Nilomar Vieira de Oliveira

MANAUS  
2007

ALCIDES DE CASTRO AMORIM NETO

PROBLEMA DE STEINER EUCLIDIANO APLICADO A  
MOLÉCULAS DE INTERESSE BIOLÓGICO

Dissertação apresentada ao Programa de Pós  
Graduação em Matemática da Universidade Fe-  
deral do Amazonas, como requisito parcial para  
obtenção do grau de Mestre em Matemática,  
área de concentração em Otimização.

Aprovada em 26 Maio de 2007.

BANCA EXAMINADORA

.....  
Prof. Dr. Nilomar Vieira de Oliveira (orientador)  
Universidade Federal do Amazonas

.....  
Prof. Dr. Roberto Cristovão Mesquita Silva  
Universidade Federal do Amazonas

.....  
Prof. Dr. Celso Carnieri  
Universidade Federal do Paraná

## AGRADECIMENTOS

Agradeço a Deus, por tudo concedido até hoje.

A Minha família em geral: esposa, pais e irmãos, pelo amor, educação e apoio sempre presente.

Agradeço ao meu orientador Prof. Dr. Nilomar Vieira de Oliveira, por sua competência e pelo seu tempo dedicado a mim sempre que necessário.

Agradeço FAPEAM pelo suporte financeiro.

Aos meus colegas do mestrado, em particular aos amigos Samuel e Walter Júnior.

Agradeço a todos os meus professores do Departamento de Matemática da UFAM e aos amigos que me ajudaram.

Finalmente agradeço a banca examinadora pelas sugestões dadas, que contribuíram para a melhoria desta dissertação.

## RESUMO

Um problema antigo e de grande aplicação na Matemática Aplicada é conhecido como problema de Steiner, que consiste na determinação de um ponto que minimize certas distâncias, problema este que foi estudado por outros matemáticos renomados como Fermat e Torricelli. Um dos resultados fundamentais em bioquímica e modelagem molecular é a determinação das Configurações de Energia Mínima (MECs) para estruturas macromoleculares tais como proteínas e DNA. As árvores mínimas de Steiner (SMTs) servem de base para elaboração de algoritmos úteis para modelar estas estruturas. Nesta dissertação, faremos uma revisão bibliográfica sobre o problema de Steiner e verificaremos, através de resultados da literatura, como as SMTs e o valor da razão de Steiner ( $\rho$ ) comparado com as árvores geradoras mínimas MSTs estão correlacionadas com as MECs de uma maneira fisicamente significativa. Uma das observações relevantes é que os átomos de carbono e nitrogênio atuam como pontos de Steiner nas árvores mínimas de Steiner das proteínas.

Palavras-chaves: problema de Steiner, razão de Steiner, proteínas e configuração mínima de energia.

## ABSTRACT

An old problem and of great application in the Applied Mathematics is known as problem of Steiner, with consists of the determination of a point that minimizes certain distances, problem this that was studied by famous mathematical as Fermat and Torricelli. A fundamental result in biochemistry and molecular modelling is the determination of the configurations of minimum energy (MECs) for structures such macromoleculares like proteins and DNA. The Steiner minimal trees (SMT) are seen as a useful algorithm paradigm to model these structures. In this work, we will examine how SMTs and the value of the ratio Steiner ( $\rho$ ) compared with the MSTs are correlated with the energies MECs in a way physically significant. We verified that carbon and nitrogen atoms are Steiner points in the proteins minimal Steiner trees.

Keywords: problem of Steiner, ratio Steiner, protein and minimum configuration of energy.

# Sumário

<b>Introdução</b>	<b>1</b>
<b>1 Problema de Steiner</b>	<b>3</b>
1.1 Histórico do problema de Steiner . . . . .	3
1.2 O Problema de Fermat . . . . .	5
1.2.1 Círculos de Torricelli . . . . .	5
1.2.2 Linhas de Simpson . . . . .	6
1.3 Generalização do Problema de Steiner . . . . .	11
1.4 Definições . . . . .	11
1.5 Propriedades . . . . .	12
1.6 A razão de Steiner no $\mathbb{R}^2$ e no $\mathbb{R}^3$ . . . . .	18
1.7 Árvores Ideais de Steiner no $\mathbb{R}^3$ . . . . .	20
<b>2 Problema de Steiner e as Macromoléculas</b>	<b>22</b>
2.1 Pesquisa Prévia em Proteínas e SMTs . . . . .	22
2.2 Estrutura Fundamental das Proteínas . . . . .	22
2.3 Abordagem do Algoritmo de Steiner . . . . .	23
2.4 Estrutura Primária de Aminoácidos . . . . .	24
2.5 Aminoácidos Perturbados . . . . .	29
2.6 Estrutura Secundária de Aminoácidos . . . . .	32
2.6.1 Colágeno . . . . .	36
2.6.2 Seda . . . . .	38
2.6.3 Miosina . . . . .	38
2.6.4 Protease do HIV - 1 . . . . .	40
<b>3 Considerações Finais</b>	<b>46</b>
<b>Referências Bibliográficas</b>	<b>48</b>

# Introdução

O propósito deste trabalho é mostrar como um problema de otimização discreta, o problema de Steiner euclidiano, pode ser utilizado em Biologia Matemática. A motivação para a escolha do tema é devido a grande aplicabilidade desse problema e de poucas pesquisas realizadas em Biologia Matemática utilizando o problema de Steiner como ferramenta, além da literatura encontrar-se dispersa em vários artigos científicos e livros.

Veremos que a predição espacial é um problema de otimização combinatorial muito complexo, pois 20 diferentes tipos de aminoácidos podem ser arrumados num número de maneira exponencial ao longo de sua rede de estrutura polipeptídica. Podem existir milhares de aminoácidos dentro de uma proteína. Sendo  $m$  o número de possíveis conformações por resíduos de aminoácidos e  $n$  a quantidade de resíduos de aminoácidos na cadeia polipeptídica da proteína, temos aproximadamente que o número total de possíveis conformações é limitada superiormente por  $m^n$  ([1]).

Atualmente, muitas estruturas de proteína conhecidas foram encontradas através de procedimentos experimentais empíricos que usam difração de raios-x ou, como a maioria, ressonância magnética (NMR). Ambas as técnicas empíricas são muito onerosas e levam muito tempo. Um grande avanço seria o desenvolvimento de programas computacionais capazes de gerar e avaliar a estrutura tridimensional a partir da seqüência inicial de aminoácidos.

Tentaremos verificar a importância que as árvores mínimas de Steiner euclidianas apresentam para a geometria dos aminoácidos e a topologia da estrutura secundária de muitas proteínas.



## Estrutura do Trabalho

No primeiro capítulo faremos um histórico sobre o problema de Steiner: como surgiu, seus precursores e suas descobertas com os trabalhos posteriormente desenvolvidos e como estão sendo utilizado atualmente. Também nesse capítulo serão apresentadas as definições e propriedades do problema de Steiner espaço euclidiano bi e tridimensional. Mostraremos os métodos de Fermat-Torricelli e as linhas de Simpson que identificam de forma geométrica a localização do ponto de Steiner. Apresentaremos também os teoremas de Heinen e Maxwell que são de grande importância em nosso estudo, e as propriedades referentes aos números de pontos do conjunto de vértices e os ângulos entre arestas, todos relacionados diretamente ao problema de Steiner.

No capítulo 2, veremos o grande problema em determinar as estruturas das proteínas e de outros modelos moleculares a partir da seqüência primária de aminoácidos. Compreender a estrutura das proteínas é fundamental para o entendimento de suas funções e propriedades, pois as mesmas possuem forma e estrutura interna definidas e essa estrutura evolui de tal forma que a proteína enrola-se no espaço. Entretanto, para compreender a função biológica das proteínas é fundamental predizer suas estruturas tridimensionais a partir da seqüência de aminoácidos, isto é, o chamado "Problema de Enovelamento (dobramento) de Proteínas". Analisaremos, em particular, as estruturas do colágeno, da seda, da miosina e da protease do HIV-1 através de um algoritmo específico.

Finalmente, no capítulo 3, faremos as considerações finais sobre as observações feitas durante a realização do trabalho. Apresentamos as conclusões iniciais e conjecturas que merecem investigações mais minuciosas que servirão também para uma série de problemas que se mostram interessantes para prosseguimento da pesquisa.

# Capítulo 1

## Problema de Steiner

Neste capítulo veremos um breve histórico do problema de Steiner, os métodos para a determinação geométrica do ponto de Steiner, utilizando os círculos de Torricelli e as linhas de Simpson. Também apresentaremos algumas definições e propriedades que serão utilizadas no decorrer da pesquisa.

### 1.1 Histórico do problema de Steiner

Nas diversas áreas do conhecimento, como na engenharia, biologia e outras, surgem com bastante frequência problemas que estão diretamente relacionados a questões de otimização, seja ela contínua ou discreta. Podemos destacar, o problema de minimizar uma soma de distâncias, que é um dos mais antigos no ramo da Matemática, e por conseguinte, foi estudado por matemáticos reconhecidos na História, tais como Euler e Gauss.

Por volta da metade do século XVII, um problema simples foi proposto: "Encontrar um ponto  $P$  que minimiza a soma das distâncias de  $P$  a cada um dos três pontos dados num plano". As soluções para este problema foram estudadas por Fermat, Torricelli e Heinen. Eles deduziram que poderiam surgir duas situações conclusivas: o ponto  $P$  estava na região interior ao triângulo formado pelos três pontos dados (nesse caso, os ângulos em  $P$  determinados pelos segmentos unindo  $P$  aos três pontos são todos  $120^\circ$ ) ou  $P$  é um dos três vértices e o ângulo em  $P$ , formado pelos segmentos juntando  $P$  aos outros dois pontos é maior ou igual a  $120^\circ$ .

No século XIX, um matemático da Universidade de Berlim chamado Jacob Steiner, estudou este problema e o generalizou com a inclusão de um conjunto arbitrariamente grande de pontos no plano. O problema foi generalizado posteriormente para um problema de minimização de caminho com

a seguinte questão: "dados  $n$  pontos num plano, encontrar o comprimento do menor caminho possível, ligando-os por segmentos de retas, podendo-se acrescentar  $k$  pontos, com  $k \leq n - 2$ ". Este problema foi bastante divulgado após o surgimento do livro de Courant e Robbins [2], *What is Mathematics?*, de 1941. Courant e Robbins ligaram o nome de Steiner a este problema generalizado, passando a partir de então a chamar-se "Problema da Árvore Mínima de Steiner". A solução para essa questão que permite o acréscimo de outros pontos, cada um dos quais chamado ponto de Steiner, é mais geral do que o problema original de Fermat para  $n$  pontos dados, e tem uma configuração final de uma "árvore" ao invés de uma "estrela".

Até 1961, apesar de muito estudado, apenas heurísticas eram conhecidas para a solução do problema da árvore mínima de Steiner (ou SMT, do inglês Steiner Minimal Tree). Nesse mesmo ano, Melzak [3] desenvolveu o primeiro algoritmo conhecido para a solução do SMT com um número finito de passos, explorando propriedades das árvores, e baseando-se em construções geométricas com régua e compasso. No entanto, esse algoritmo funciona apenas quando a quantidade de pontos é muito pequena, por ser de ordem exponencial.

Em 1968, Gilbert e Pollak [4], usaram definitivamente o termo "Árvore Mínima de Steiner" para o problema que relaciona a árvore dos pontos dados inicialmente e os pontos de Steiner. Estimularam várias pesquisas através de várias conjecturas interessantes. Os mesmos também tornaram pública a definição de "Razão de Steiner", um quociente entre o comprimento da Árvore Mínima de Steiner e o comprimento da Árvore Geradora Mínima.

Posteriormente, ainda no caso bidimensional, várias aplicações relacionadas ao Problema de Steiner surgiram, destacando ainda mais a importância desse problema. Entre essas aplicações podemos destacar o planejamento de circuitos eletrônicos para rotas VLSI (Very Large Scale Integrated), que precisa utilizar uma definição de distância diferente da usual euclidiana, onde o ganho em performance em soluções ótimas é muito grande.

Até 1992 nenhum algoritmo para tentar encontrar a SMT para o caso tridimensional euclidiano era conhecido. No mesmo ano, W. D. Smith [5] forneceu o primeiro algoritmo para tratar o problema tridimensional euclidiano que é conhecido, cuja performance, bastante satisfatória para quantidade pequenas de pontos, e serviu de base para vários trabalhos que buscavam algoritmos heurísticos para três dimensões que são de grande utilidade na Biologia Matemática.

McGregor Smith e W. Smith [6], em 1995, fizeram uma famosa conjectura sobre uma configuração de pontos chamada de "Sausage", na qual a razão de Steiner no caso tridimensional teria um determinado valor, alcançado para um número infinito de pontos (diferente do caso bidimensional), o qual seria

obtido quando a configuração inicial dos pontos de entrada estivesse disposta em uma estrutura helicoidal, semelhante a uma molécula de DNA.

McGregor Smith e Smith [6] apresentou várias propriedades que foram generalizadas do espaço euclidiano bidimensional para o tridimensional, e ainda métodos algorítmicos que, devido a potencialidade do problema de Steiner, são utilizados em diversas aplicações. Uma das aplicações mais interessantes está relacionada ao problema de Enovelamento de proteínas e DNA, muito importante para vários ramos das Ciências atualmente, como as que buscam a produção de drogas para o tratamentos e curas de doenças como câncer e Aids.

## 1.2 O Problema de Fermat

Também conhecido como problema de Torricelli-Fermat, o problema de Steiner para 3 pontos, pode ser formulado como um problema de Otimização, mais especificamente, um problema de minimização. Foi proposto pela primeira vez no século XVII por Fermat, fazendo o seguinte questionamento: "Dados três pontos no plano, qual é o quarto ponto tal que a soma dos comprimentos de suas distâncias aos três pontos dados seja mínima?"

Apresentaremos duas formas alternativas de encontrarmos a solução para este problema.

### 1.2.1 Círculos de Torricelli

Esse método de solução do problema de Fermat foi descoberto em 1640 por Torricelli, que consiste no seguinte:

Dados três vértices  $A$ ,  $B$  e  $C$  de um triângulo, temos três circunferências  $\lambda_1$ ,  $\lambda_2$  e  $\lambda_3$ , cada uma passando por vértices tomados dois a dois desse triângulo e pontos  $X_i$ ,  $i = 1, 2, 3$ , em cada uma das circunferências formando triângulos equiláteros com estes vértices, conforme figura 1.1. Estas circunferências se interceptam num ponto  $S$  interior ao triângulo  $ABC$  (sob certas condições para os ângulos internos desse triângulo), esse ponto é chamado de *ponto de Steiner*.

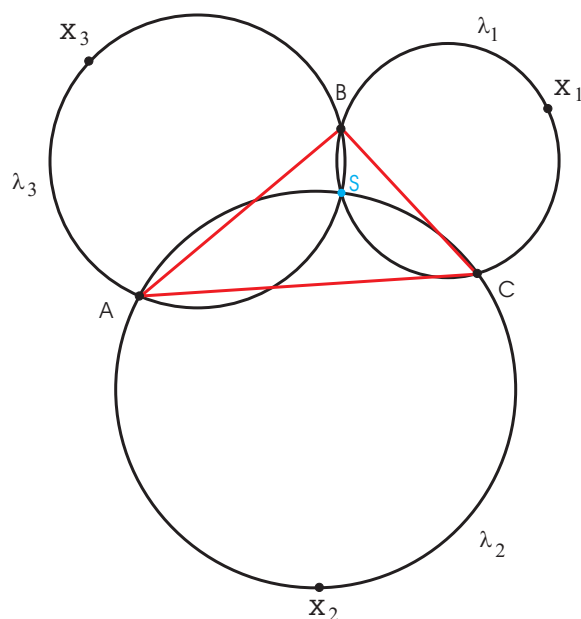


Figura 1.1: O ponto de Steiner como intersecção dos círculos de Torricelli.

### 1.2.2 Linhas de Simpson

Esse método para solução do problema de Fermat foi redescoberto em 1750 por Simpson, pois um aluno de Torricelli chamado Viviane já havia encontrado a solução. O método consiste no seguinte:

Dados três triângulos equiláteros determinados da mesma forma que no método de Torricelli, cujo segmento é obtido pela união dos vértices  $X_i$ ,  $i = 1, 2, 3$  dos triângulos equiláteros com o vértice oposto do triângulo em questão, determina as chamadas *linhas de Simpson*, conforme pode ser observado na figura 1.2. A intersecção das linhas de Simpson também determina a localização do ponto de Steiner.

Os métodos de Torricelli e Simpson resultam num ponto de Steiner somente se os ângulos internos do triângulo  $ABC$  forem menores que  $120^\circ$ , caso contrário nenhum dos métodos funciona, ou seja, o método de Torricelli nos dá um ponto fora do triângulo e as linhas de Simpson não se interceptam.

O teorema de Heinen que apresentaremos na seqüência, encerra a questão para três pontos no plano.

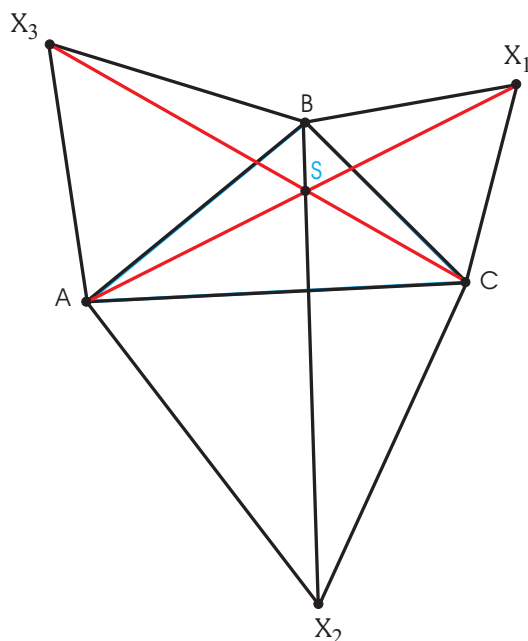


Figura 1.2: O ponto de Steiner como intersecção das linhas de Simpson.

**Teorema 1.2.1. (Heinen)** *Se um Triângulo  $ABC$  tem um ângulo interno maior ou igual a  $120^\circ$ , então o ponto que minimiza a soma das distâncias aos três vértices coincidirá com aquele vértice que determina esse ângulo com os demais. Entretanto, se todos os ângulos internos do triângulo  $ABC$  forem menores que  $120^\circ$  então deve existir um ponto  $S$ , interior à região triangular  $ABC$  dada, ponto esse determinado pelos círculos de Torricelli, cuja distância do ponto  $S$  aos pontos  $ABC$  é mínima. Além disso,*

$$AS + BS + CS = AA' = BB' = CC'$$

onde  $AA'$ ,  $BB'$  e  $CC'$ , são as linhas de Simpson.

**Demonstração.**

A demonstração da segunda parte do teorema será feita considerando que nenhum ângulo interno do triângulo  $ABC$  é maior ou igual a  $120^\circ$ .

Seja  $S$  um ponto qualquer do plano triangular, de acordo com a figura 1.3. Com centro de rotação em  $B$ , gira-se o segmento  $BA$  no sentido anti-horário num ângulo de  $60^\circ$  até chegar na posição  $BC'$ , da mesma forma com o segmento  $BS$  até  $BS'$ . Dessa maneira os triângulos  $ABC'$  e  $BSS'$  são equiláteros.

Como a construção resulta em uma rotação do triângulo  $ABS$  de um

ângulo de  $60^\circ$  em torno de  $B$  até a posição  $C'BS'$ , verifica-se que  $AS = C'S'$  e  $BS = SS'$  e conseqüentemente,

$$AS + BS + CS = C'S' + S'S + SC$$

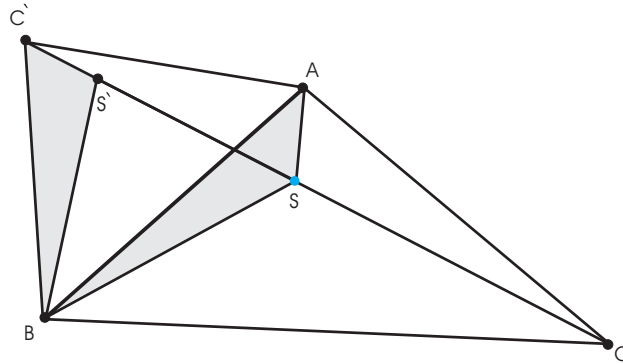


Figura 1.3: Todos os ângulos internos do triângulo  $ABC$  são menores que  $120^\circ$ .

Então, o problema de minimizar  $AS + BS + CS$ , agora consiste em minimizar  $C'S' + S'S + SC$ . Escolhendo de forma conveniente a localização do ponto  $S$ . Observa-se que o ponto  $C'$ , determinado somente pelo triângulo  $ABC$  independe do ponto  $S$ . A nova soma a ser minimizada é o caminho do ponto  $C'$  a  $C$  passando por  $S'$  e  $S$ . Se esse caminho for uma linha reta, então o problema de minimização está resolvido, bem como o comprimento de  $C'$  a  $C$ . E isso pode acontecer se o ponto  $S$  estiver localizado de tal forma que cada um dos ângulos  $\angle ASB$ ,  $\angle ASC$  e  $\angle BSC$  seja  $120^\circ$ . Logo, neste caso, temos:

$$\angle C'S'B = \angle ASB = 120^\circ,$$

$$\angle C'S'B + \angle BS'S = 180^\circ = \angle BSS' + \angle BSC$$

Portanto, temos que  $C'S'SC$  é um segmento de linha reta.

Mostraremos agora que o ponto  $S$  pode realmente ser obtido de forma que os ângulos determinados em  $S$  pelos lados  $AB$ ,  $AC$  e  $BC$  sejam  $120^\circ$ . Para isso, considere os círculos circunscritos aos triângulos equiláteros  $C'AB$  e  $A'BC$ , conforme a figura 1.4.

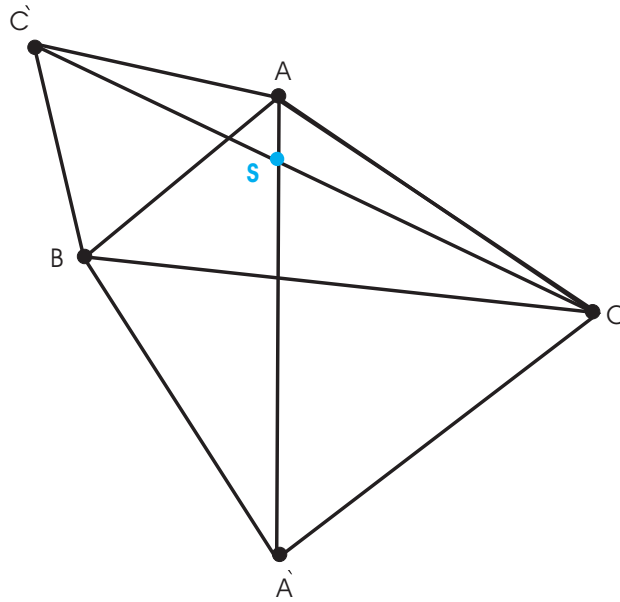


Figura 1.4: O ponto  $S$  é a interseção dos círculos de Torricelli.

Estes círculos se interceptam no ponto  $B$ , obviamente, mas também em outro ponto, e este é o ponto  $S$  que minimiza o comprimento de  $AS + BS + CS$ . Como os quatro pontos  $C'$ ,  $A$ ,  $S$ ,  $B$  estão no mesmo círculo, os ângulos opostos deste quadrilátero são suplementares, e portanto

$$\angle AC'B + \angle ASB = 180^\circ.$$

Mas  $\angle AC'B = 60^\circ$ , então temos  $\angle ASB = 120^\circ$ , como queríamos. Analogamente, os pontos  $A'$ ,  $B$ ,  $C$  e  $S$  estão num mesmo círculo e pelo mesmo argumento da construção anterior  $\angle BSC = 120^\circ$ . Por conseguinte temos que  $\angle ASC = 120^\circ$ , pois os três ângulos em  $S$  devem somar  $360^\circ$ .

Provaremos agora a primeira parte do teorema. No caso de um dos ângulos do triângulo  $ABC$ , digamos  $\angle ABC$ , ser maior ou igual a  $120^\circ$ . No caso que o ângulo  $\angle ABC = 120^\circ$ , por exemplo, os pontos  $C'$ ,  $B$  e  $C$  são colineares. Isto sugere que a localização do ponto  $S$  é o ponto  $B$ . No caso em que  $\angle ABC$  é maior que  $120^\circ$ , a construção usada acima não leva a uma solução do problema.

Recomeçaremos com uma construção diferente para ser analisada e utilizaremos um método análogo ao estudado anteriormente, ou seja, através de rotação de segmentos específicos. Seja  $S$  qualquer ponto, que não seja o ponto  $B$ , no plano do triângulo  $ABC$ . Provaremos que  $AB + BC < AS + BS + CS$ , como afirmado pelo teorema. A figura 1.5 mostra o ponto  $C'$  escolhido na



reta  $CB$ , ou na extensão do segmento  $BC$  para ser mais claro, de forma que  $C'B = AB$ . O ponto  $S'$  é escolhido de forma que o triângulo  $C'S'B$  seja obtido por rotação no sentido anti-horário do triângulo  $ABS$ .

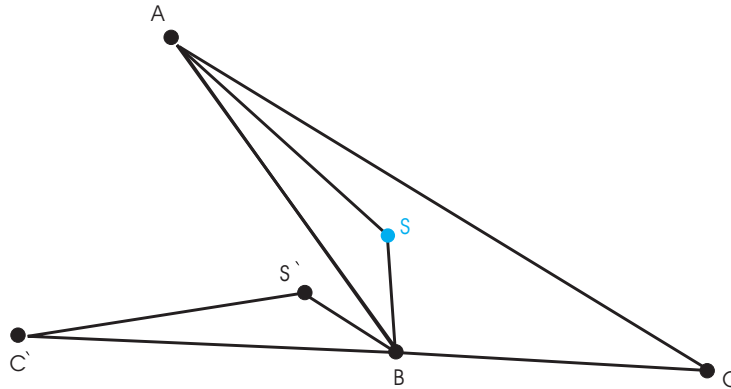


Figura 1.5: O ângulo  $\angle ABC \geq 120^\circ$ .

Podemos observar que

$$\begin{aligned} C'S' &= AS, \\ \angle SBS' &= \angle ABC' \leq 60^\circ \end{aligned}$$

Segue que

$$\begin{aligned} BS &\geq S'S, \\ AS + BS + CS &\geq C'S' + S'S + SC \\ &> C'C = AB + BC \end{aligned}$$

pois não podem estar ambos  $S$  e  $S'$  sobre o segmento  $C'C$  pela construção. Isto conclui a demonstração.

## 1.3 Generalização do Problema de Steiner

Existem várias maneiras de generalizar o problema de Fermat para três pontos. A primeira delas é fazendo a seguinte pergunta: "Qual o ponto  $P$  cuja soma das distâncias de  $n$  pontos, com  $n \geq 3$ ,  $P_1, P_2, \dots, P_n$  a esse ponto é mínima?". O caso  $n = 3$  é o caso que estudamos na seção anterior. Para o caso  $n > 3$  resulta num problema não-linear cuja solução é possível determinar [5]. Podemos ainda atribuir pesos às arestas que conectam os pontos  $P_n$  ao ponto  $P$  procurado. Não trataremos esses casos no presente trabalho.

Outra forma de generalizar o Problema de Fermat é considerar  $n$  pontos no plano (ou no espaço  $d$ -dimensional,  $d \geq 2$ ) e perguntar "qual é a menor árvore que interconecta esses pontos se pudermos acrescentar tantos pontos extras quanto quisermos, desde que esse acréscimo de pontos corresponda a uma diminuição no comprimento total da árvore formada?". Esse problema ficou conhecido na literatura como "problema de Steiner", em homenagem ao célebre geômetra alemão Jacob Steiner, que lecionou na Universidade de Berlin no século XIX.

Foram Courant e Robbins que nomearam dessa forma essa generalização do problema de Fermat no seu livro *What is Mathematics?* de 1941, que ajudou a divulgar o problema devido à popularidade deste livro, no entanto, causando descontentamento em relação ao nome do problema. Alguns reclamam pela prioridade temporal para Fermat, que teria sido o primeiro a propor o problema; outros, que a generalização deveria lembrar Jarnik e Kösler, que em 1934 estudaram o problema de interconectar pontos no plano com comprimento mínimo, analisando o caso em que os  $n$  pontos são os vértices de um polígono regular. Mais especificamente, eles encontraram a árvore mais curta para  $n = 3, 4, 5$  e mostraram que  $n - 1$  lados do polígono regular de  $n$  lados constituem a árvore mais curta quando  $n \geq 13$ .

## 1.4 Definições

Nesta seção serão dadas algumas definições referentes ao problema da árvore de Steiner. Lembrando que, dado um conjunto de pontos regulares, ou seja, pontos dados para serem conectados, uma topologia de uma árvore definirá quais pontos devem ter conexões entre si, mas não as posições dos pontos de Steiner, ou seja, os pontos que são adicionados a ela de forma a encurtá-la.

**Definição 1.4.1.** *Uma árvore de Steiner é aquela que, fixada uma topologia, não pode ser encurtada por qualquer perturbação.*

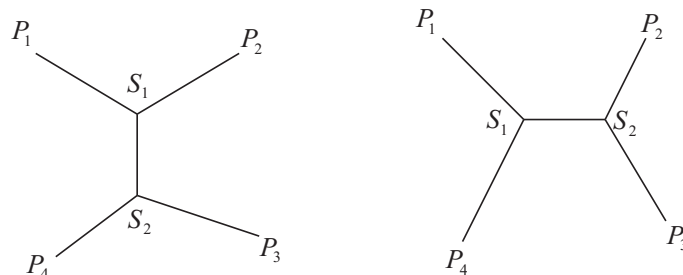


Figura 1.6: *Dois topologias para quatro pontos dados*

**Definição 1.4.2.** *Árvore relativamente mínima é árvore de Steiner que pode ter seu comprimento diminuído pelo acréscimo de pontos de Steiner.*

**Definição 1.4.3.** *Árvore mínima de Steiner (global) é a árvore de Steiner obtida quando todas as possíveis topologias são analisadas.*

**Definição 1.4.4.** *Árvore geradora mínima é a árvore obtida pela conexão de todos os pontos dados, sem formação de ciclos e sem acréscimos de pontos de Steiner.*

**Definição 1.4.5.** *Uma Envoltória Convexa de Steiner para um dado conjunto de "n" pontos é definida como sendo a região do espaço que contém uma árvore mínima de Steiner.*

## 1.5 Propriedades

No problema de Steiner euclidiano bidimensional, existem algumas propriedades que podem ser generalizadas para espaços de dimensão maiores, em particular para o espaço tridimensional, que é de nosso interesse e que serão utilizados posteriormente. Propriedades que serão utilizadas em nosso estudo referente à conformações de macromoléculas, mas especificamente no algoritmo utilizado.

**Proposição 1.5.1. Sobre os ângulos e arestas na árvore de Steiner.**

(i) Numa árvore de Steiner, nenhum par de arestas se encontra com ângulo menor que  $120^\circ$ .

(ii) Duas arestas de uma árvore de Steiner não se interceptam novamente em outro ponto, ou seja, não há loops.

(iii) Cada ponto de Steiner possui exatamente três arestas incidentes, e o ângulo entre quaisquer dois desses segmentos é exatamente  $120^\circ$ .

**Demonstração.** Faremos a demonstração do item (i), e os demais itens são conseqüências do mesmo. Seja  $S$  um ponto de Steiner tal que duas arestas da árvore de Steiner se encontrem nesse ponto formando um ângulo menor que  $120^\circ$ . Considere dois pontos  $A$  e  $B$  sobre estas arestas de tal forma que o comprimento de  $AS$  seja igual ao comprimento de  $BS$ , conforme figura 1.7.

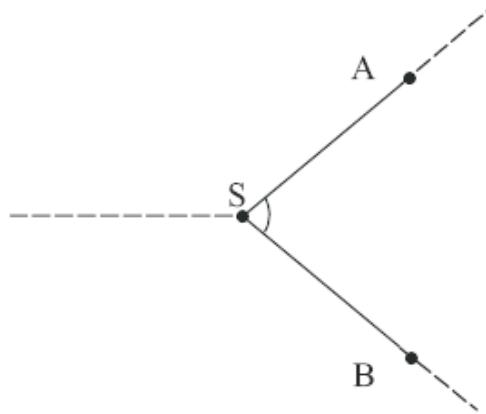


Figura 1.7: O ângulo  $\angle ASB$  tem medida inferior a  $120^\circ$ .

Seja  $S'$  o ponto de intersecção das três linhas de Simpson para o triângulo  $ABS$ . Na árvore de Steiner, se trocarmos  $AS$  e  $BS$  pelos segmentos  $AS'$ ,  $BS'$  e  $SS'$  então obtém-se uma árvore com um comprimento menor ainda, conforme a figura que 1.8. Essa contradição mostra que o ângulo entre duas arestas quaisquer numa árvore de Steiner é, no mínimo,  $120^\circ$ .

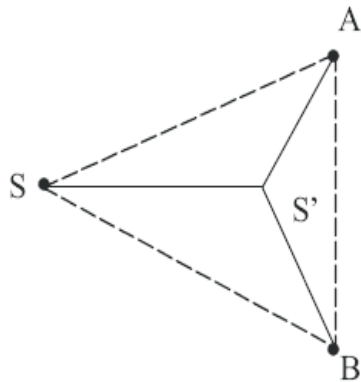


Figura 1.8: O ponto  $S'$  encurta o comprimento da árvore de Steiner

**Proposição 1.5.2. (Sobre o número de pontos de Steiner).** *Cada árvore de Steiner tem um número a mais de pontos em relação ao número de arestas.*

*A demonstração desta proposição será feita no teorema 1.5.1. Basta observar a figura 1.6 da definição 1.4.1 vista anteriormente e em seguida, usamos indução matemática para provar para  $n$  pontos regulares dados.*

**Proposição 1.5.3. (A Envoltória Convexa dos Pontos de Steiner).** *Numa árvore com mínimo relativo, ou seja, possui um ponto que ainda pode ser melhor localizado de tal forma que encurte ainda mais a árvore de Steiner, todos os pontos de Steiner estão dentro de uma envoltória convexa.*

*A demonstração é feita utilizando a proposição 1.5.1, observando figuras 1.7 e 1.8 comentadas anteriormente. Podemos perceber a validade da proposição, pois num conjunto de  $n$  pontos regulares, unindo todos os pontos dados formaremos a MST que por sua vez contém SMT pois está dentro da mesma, assim também com os pontos de Steiner.*

**Lema 1.5.1. (Propriedade Lunar).** *Seja  $UV$  qualquer aresta de uma árvore de Steiner. Seja  $L(U, V)$  a região consistindo de todos os pontos  $P$  satisfazendo:*

$$PU < UV \text{ e } PV < UV$$

*Onde  $L(U, V)$  é uma intersecção de formato Lunar de círculos de raios do comprimento de  $UV$ , centradas nos pontos  $U$  e  $V$ . Então nenhum outro vértice da Árvore Mínima de Steiner pode estar em  $L(U, V)$ .*

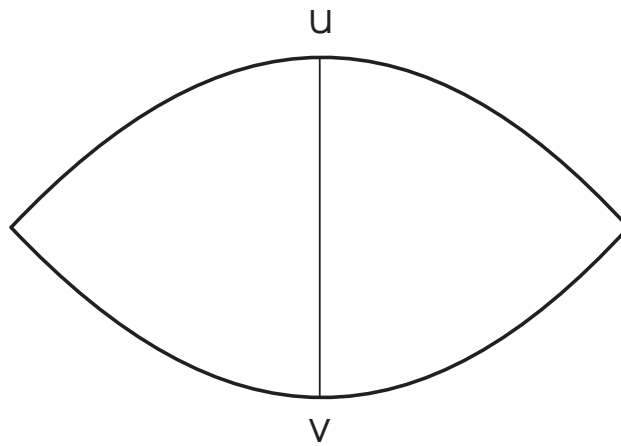


Figura 1.9: Região Lunar

**Demonstração.** *Se  $Q$  fosse um desses vértices, a árvore mínima de Steiner deveria conter ou um caminho de  $Q$  para  $U$  não contendo  $V$ , ou vice-versa. No último caso, por exemplo, a árvore mínima de Steiner pode ser encurtada pela retirada da aresta  $UV$  e pela adição da aresta  $QV$ , uma contradição que pode ser vista na figura 1.9.*

**Lema 1.5.2. (Propriedade da Fatia).** *Seja  $W$  qualquer região aberta com formato de fatia, tendo um ângulo igual ou superior a  $120^\circ$  e não contendo nenhum dos pontos regulares, ou seja, vértices terminais do conjunto dado,. Então  $W$  não contém Pontos de Steiner*

**Demonstração.** Podemos supor, sem perda de generalidades, que  $W$  englobe a região angular de  $-60^\circ$  e  $60^\circ$ , conforme figura 1.10.

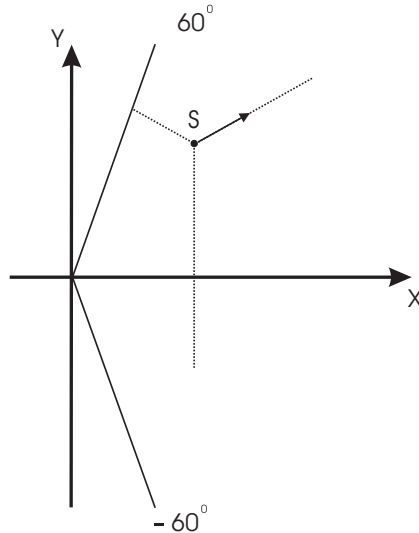


Figura 1.10: Região em forma de fatia

Suponha que exista pontos de Steiner nessa região, dos quais o ponto  $S$  é aquele com maior abscissa. Das três arestas incidentes em  $S$ , uma o deixa na direção  $\pm 60^\circ$  do eixo positivo. Esta não pode sair de  $W$  e, portanto, não pode convergir num ponto regular. Além disso, existiria uma coordenada  $x$  maior que a de  $S$ , contradição.

**Corolário 1.5.1. Envoltória Convexa.** A Envoltória Convexa dos Pontos Regulares é uma Envoltória de Steiner.

**Demonstração.** Cada reta suporte da envoltória convexa dos pontos regulares define uma fatia de  $180^\circ$  livre de pontos regulares.

**Teorema 1.5.1.** Uma árvore de Steiner de  $N$  pontos regulares possui no máximo  $n - 2$  pontos de Steiner.

**Demonstração.** Considere uma árvore de Steiner com  $n$  pontos regulares  $P_1, P_2, \dots, P_n$  e  $K$  pontos de Steiner, temos que o número de arestas incidentes deve ser  $n + k - 1$ . Como cada aresta tem duas extremidades, obtemos  $2(n + k - 1)$  arestas incidentes e o número máximo de arestas deve ser  $\frac{3k+n}{2}$ . Segue

que

$$n + k - 1 \geq \frac{3k + n}{2}$$

e conseqüentemente,

$$k \leq n - 2.$$

**Teorema 1.5.2. (Teorema de Maxwell)**

Tomando vetores unitários apontando para fora de uma árvore mínima relativa nas direções de cada uma das arestas incidentes em  $A_1, A_2, \dots, A_n$  com  $F_i$  denotando a soma das forças unitárias em  $A_i$ . Em termos mecânicos  $F_i$  é a força externa necessária em  $A_i$  para manter as árvores em equilíbrio. O comprimento  $L$  da árvore é dado pela expressão:

$$L = \sum_{i=1}^n A_i \cdot F_i$$

**Demonstração.** Construindo um vetor unitário de força  $f_k$  apontando para fora de cada ponto  $A_k$ , conforme figura 1.11. Considerando todas as arestas da árvore, como existem  $n + s - 1$  arestas, devem existir  $2(n + s - 1)$  vetores para traçar. Depois, consideramos uma soma de produtos internos definidos por

$$T = \sum P_k \cdot f_k$$

Onde a soma é estendida a todos os  $2(n + s - 1)$  pontos das arestas. Os termos de  $T$  podem ser obtidos de seguinte forma: Uma é combinando termos para os quais  $P_k$  é o mesmo ponto. Cada ponto de Steiner aparece em três termos de  $T$ . A contribuição de um ponto de Steiner  $S$  para  $T$  é zero pois as três forças em  $S$  estão em equilíbrio. A contribuição de um dado ponto  $A_i$  para  $T$  é  $A_k \cdot F_k$ . Outra maneira é combinar dois termos associados com as duas extremidades de cada aresta. Se  $P_1 P_2$  é uma aresta de comprimento  $L_{12}$  então:

$$P_2 = P_1 + L_{12} \vec{u}$$

onde  $\vec{u}$  é um vetor unitário. Os termos para esta aresta contribuem  $P_1(-\vec{u}) + P_2 \cdot \vec{u} = L_{12}$ . Então  $T$  é o lado esquerdo da expressão.



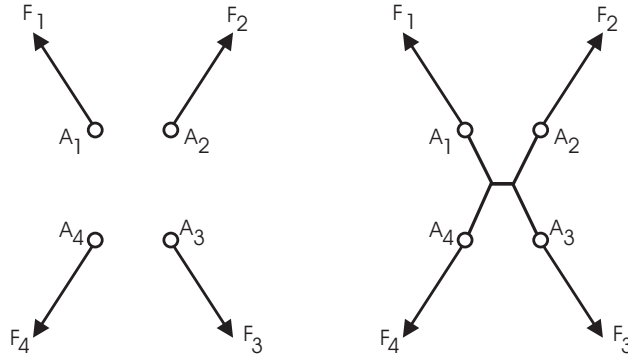


Figura 1.11: *forças unitárias apontando para fora de cada ponto.*

## 1.6 A razão de Steiner no $\mathbb{R}^2$ e no $\mathbb{R}^3$

Definiremos agora a razão de Steiner no caso do espaço de dimensão  $d = 2$ , ou seja, no plano euclidiano e faremos uma breve revisão do problema também no espaço tridimensional, enfatizando algumas aplicações.

Sejam  $\ell(SMT(V))$  o comprimento da árvore mínima de Steiner euclidiana e  $\ell(MST(V))$  o comprimento da árvore mínima geradora, ambas para um conjunto  $V$  de pontos em  $\mathbb{R}^d$ , então  $\rho_d$  é a menor razão de Steiner dentre todos os conjuntos  $V$  na dimensão  $d$ , isto é,

$$\rho_d = \inf_{V \subseteq \mathbb{R}^d} \rho(V)$$

$$\text{onde } \rho(V) = \frac{\ell(SMT(V))}{\ell(MST(V))}.$$

Pode observar facilmente que o valor máximo da razão de Steiner é igual a 1. Mencionaremos agora apenas  $\rho$  e o ambiente será o espaço euclidiano tridimensional.

Podemos interpretar a razão de Steiner de duas maneiras:

Uma é conhecer o quanto ganhamos ao substituir a árvore geradora pela árvore mínima de Steiner. Devido a complexidade computacional em encontrar a árvore mínima de Steiner é mais adequado continuar com a árvore mínima geradora, cuja complexidade na pior das hipóteses é polinomial. De forma resumida, fixadas as topologias e o número de pontos, essa comparação é feita observando a seguinte igualdade:

$$\ell(SMT(V)) = \rho \cdot \ell(MST(V))$$

Assim, pode-se concluir que  $\rho$  representa o decréscimo máximo possível no  $\ell(MST(V))$ , através de acréscimo dos pontos de Steiner.

A outra maneira para interpretar a razão de Steiner é relacionada à Configuração Mínima de Energia (MEC). Sabe-se que a energia potencial mínima de um sistema ocorre quando o mesmo está em equilíbrio, sendo chamado estável quando sua função potencial não puder ser melhorada para atingir um valor de mínimo ainda menor ao já conseguido. Neste caso dizemos que o sistema atingiu um estado de energia mínima global. No entanto, o sistema pode estar em equilíbrio instável quando a função energia do sistema poder ainda alcançar um valor menor. Quando isso ocorre, diz-se que o sistema está num mínimo relativo. A conexão que existe entre a razão de Steiner e a configuração de energia mínima é que, possivelmente, quanto mais próximo da unidade for a razão de Steiner de certos agregados protéicos, mais estáveis estariam essas determinadas moléculas, ou trechos destas. Ou seja, aumentar a razão de Steiner corresponderia a diminuir a energia potencial da configuração. Portanto, algumas macromoléculas de interesse biológico, como as proteínas, que representariam configurações estáveis, estariam com seus átomos numa configuração dada por árvores de Steiner, ou seja, a razão de Steiner de sua configuração seria aproximadamente igual a unidade, o que representaria um valor de equilíbrio estável.

Em 1968, Gilbert e Pollak observaram que a razão de Steiner era  $\frac{\sqrt{3}}{2}$  para os vértices de um triângulo equilátero e conjecturaram que nenhum conjunto de pontos no plano alcançava uma razão de Steiner inferior. Por vários anos, pesquisadores avançaram à demonstração de duas maneiras: Pelo estudo de conjuntos arbitrários, mas com um número fixo de pontos, para  $n = 3, 4, 5, 6$  [7]; e por outra linha de raciocínio, levantando o limite inferior para a razão de Steiner para conjuntos arbitrários de pontos, com  $\rho \geq 0.5, \geq 0.57, \geq 0.74, \geq 0.80, \geq 0.824$  [7], respectivamente. Finalmente, aproximadamente vinte e dois anos após conjecturado por Gilbert e Pollack, Du e Hwang anunciaram uma prova da conjectura [8]:

$$\rho_2 = \frac{\sqrt{3}}{2}$$

Para o caso tridimensional, a generalização de Gilbert e Pollak se mostrou-se falsa. Smith e Du [8] mostraram que a razão de Steiner para o  $d$ -simplex, que era a configuração imaginada por Gilbert e Pollak como a que deveria possuir a menor razão de Steiner, não era a que apresentava a menor razão de Steiner para o caso  $d$ -dimensional,  $d \geq 3$ . Eles mostraram que uma configuração que chamaram de  $d$ -sausage apresentava valores ainda menores que os  $d$ -simplex, e desta constatação, conjecturaram que a configuração da

"*d-sausage*" seria a que apresentava a menor razão de Steiner em todas as dimensões, com  $d \geq 3$ .

## 1.7 Árvores Ideais de Steiner no $\mathbb{R}^3$

Recentemente, foi conjecturado que a configuração ótima para a minimização  $\rho$  de qualquer conjunto de pontos é uma cadeia infinita de faces justapostas de um tetraedro, também conhecida como tripla hélice ou tira de sausage ([6]). A figura 1.12 ilustra a *R-sausage* para um conjunto de  $N$  vértices. O  $\rho$  da *R-sausage* é o menor valor que encontramos entre todas possíveis configurações. A razão de Steiner da *R-sausage* é uma cota superior da razão de Steiner para conjunto de vértices terminais e

$$\rho = \frac{1}{10}(3\sqrt{3} + \sqrt{7}) \approx 0.784903733\dots$$

O melhor valor de cota inferior conhecida para a razão de Steiner em qualquer dimensão incluindo  $\mathbb{R}^3$  é 0.615 e foi demonstrado por Du (1991). Centenas de milhares de conjunto de pontos foram avaliados para sustentar a conjectura que a *R-sausage* é ótima para o espaço tridimensional e a *R-sausage* pode ser a estrutura ótima para dimensões maiores.([5])

A *R-sausage* é composta de faces comuns regulares do tetraedro ligadas de tal forma que os pontos de Steiner dentro delas formam sua própria hélice e também formam uma topologia de caminhos. Essa configuração da *R-sausage* foi que motivou a pesquisa em proteínas, ácidos nucléicos e outras biomoléculas MECs, porque se a estrutura ideal tridimensional é uma cadeia longa de tetraedros, isso poderia ajudar a explicar porque proteínas e outras estruturas bioquímicas tem longas cadeias. Na verdade, em ciências de materiais, essa tripla hélice é conhecida como Espiral de Bernal porque está relacionada ao estudo de líquidos densos e gases metálicos ([8]). O que não era conhecido até agora por pesquisadores é a relação e propriedades especiais da *R-sausage* para o problema de Steiner.

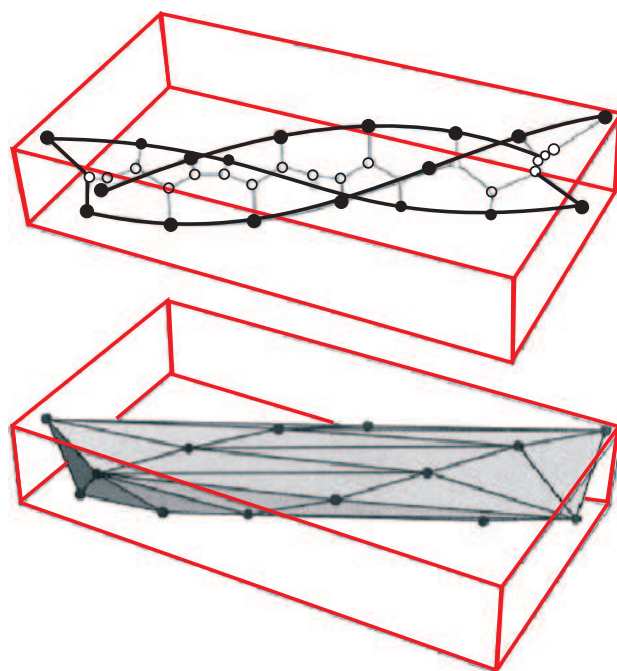


Figura 1.12: *R-sausage*

## Capítulo 2

# Problema de Steiner e as Macromoléculas

### 2.1 Pesquisa Prévia em Proteínas e SMTs

Em estudos recentes de proteínas catalíticas e estruturais, houve tentativa de determinar a efetividade das cotas inferiores a partir das SMTs na predição e avaliação da estrutura de proteínas. Examinaram-se dados ([8]) do conjunto de Banco de Dados de Proteínas (PDB) em colágenos, fibras, actina, miosina, hemoglobina e muitas outras. A diferença entre o estudo prévio e o corrente é que neste estudo foram examinados os aminoácidos individualmente bem como as perturbações nas suas estruturas e na estrutura geométrica das proteínas.

### 2.2 Estrutura Fundamental das Proteínas

Existem quatro importantes tipos de macromoléculas do interesse das Ciências Biológicas: carboidratos, ácidos nucleicos, proteínas e lipídios. Devemos focar a estrutura das proteínas porque é a que possui estrutura geométrica mais próxima possível de ser relacionada com a da *R-sausage*. Mas como será discutido, um tipo similar de análise estrutural pode ser montada para outros tipos de macromoléculas, especialmente DNA e RNA.

Um conteúdo chave em bioquímica hoje, citado na introdução, é prever a estrutura tridimensional de proteínas a partir da seqüência primária de aminoácidos. Basicamente, duas abordagens para este problema foram desenvolvidas ao longo dos anos ([1]): Métodos Homológicos e de Condensação hierárquica.

Os métodos homológicos consistem na comparação de estruturas de proteínas desconhecidas com estrutura de proteínas conhecidas. Basicamente, pesquisadores decompõem uma proteína em suas partes constituintes, analisam as partes separadamente, então comparam as proteínas com aquelas conhecidas no banco de dados.

A outra abordagem para predição está no princípio da construção de blocos fundamentais de estruturas de proteínas, isto é, os aminoácidos e como são unidos no  $\mathbb{R}^3$  de acordo com um conjunto de regras. Esta abordagem para modelagem de proteína nos dá mais poder e flexibilidade em lidar com estruturas de proteínas desconhecidas. Pelo Método da Condensação Hierárquica, visualiza-se o problema identificando e predizendo a estrutura proteica da seguinte maneira: primeiramente, define-se o princípio da construção dos blocos da seqüência de aminoácidos com o primeiro nível. Com a seqüência de aminoácidos conhecida, os tipos de estrutura geométrica ligam-se a essas seqüências, desenvolvendo o nível secundário.

Tipicamente, a geometria da estrutura secundária, pode ser qualquer hélice-alpha, folha de beta ou combinações dessas estruturas. O modo que as estruturas geométricas enovelam-se (ou dobram-se) no espaço é chamado de nível terciário, e, finalmente, o nível quaternário representa a composição final de todas as estruturas secundárias e terciárias dentro da proteína. A abordagem hierárquica é a que será apresentada neste trabalho.

## 2.3 Abordagem do Algoritmo de Steiner

Usamos um algoritmo similar ao de Smith e Smith ([6]) para a avaliação de proteínas. Similar pois fizemos apenas algumas alterações, devido o mesmo não ter uma tela própria e não conseguirmos executá-lo corretamente em nosso sistema operacional. Ele é um algoritmo de otimização *branch-and-bound* como em Smith([5]) para a solução de árvores de Steiner em  $\mathbb{R}^d$ , e a versão usada para resolver o problema em  $\mathbb{R}^3$  busca uma solução ótima. Ela busca a solução ótima adicionando no máximo  $N - 2$  pontos de Steiner de forma que a estrutura seja uma FST, e seu tempo de convergência cresce exponencialmente com o aumento do número de vértices terminais. Avaliamos o tempo de convergência para saber a configuração ótima dos aminoácidos e proteínas, quando foi possível. Porém, a menos que o conjunto de vértices terminais seja  $N \leq 13$  ou tenha uma estrutura topológica especial, não podemos alcançar um ótimo com um tempo de computação razoável.

A função objetivo utilizada para o programa de Steiner é um problema de Otimização, descrito da seguinte maneira:

$$\text{Min } \mathbf{Z} = \sum_{i \in V \cup S} \sum_{j \in S} [(v_{ix} - s_{jx})^2 + (v_{iy} - s_{jy})^2 + (v_{iz} - s_{jz})^2]^{1/2}$$

onde  $V$  é o conjunto de vértices terminais e  $S$  é conjunto dos pontos de Steiner.

Para um caso particular de quatro pontos, que é o objetivo do nosso trabalho, podemos reescrevê-lo da seguinte forma:

$$a = [(v_{ix} - s_{ix})^2 + (v_{iy} - s_{iy})^2 + (v_{iz} - s_{iz})^2]^{1/2}$$

$$b = [(v_{jx} - s_{ix})^2 + (v_{jy} - s_{iy})^2 + (v_{jz} - s_{iz})^2]^{1/2}$$

$$c = [(v_{kx} - s_{jx})^2 + (v_{ky} - s_{jy})^2 + (v_{kz} - s_{jz})^2]^{1/2}$$

$$d = [(v_{lx} - s_{jx})^2 + (v_{ly} - s_{jy})^2 + (v_{lz} - s_{jz})^2]^{1/2}$$

$$e = [(s_{ix} - s_{jx})^2 + (s_{iy} - s_{jy})^2 + (s_{iz} - s_{jz})^2]^{1/2}$$

Podemos observar a complexidade do cálculo a medida que aumenta o número de pontos do conjunto  $V$ .

Na parte experimental deste trabalho, começamos analisando a geometria fundamental da cada aminoácido e examinamos suas propriedades de Steiner, prosseguindo com a modelagem da estrutura secundária das proteínas e perturbações em sua estrutura. Finalmente, buscamos determinar como as propriedades de Steiner transcendem todos os níveis primários e secundários.

## 2.4 Estrutura Primária de Aminoácidos

Primeiramente examinamos os aminoácidos individualmente para ver quais propriedades de Steiner resultam. Depois avaliamos a robustez de diferentes proteínas com respeito a como a razão de Steiner pode ser usada na modelagem de proteínas. A estrutura primária da proteína é composta de diferentes aminoácidos acoplados da "cabeça à cauda". Cada aminoácido tem um átomo de carbono com um R-grupo ligado a ele, um grupo amino na

esquerda e um grupo carboxila na direita, conforme figura 2.1. Cada aminoácido difere por uma cadeia lateral, isto é, seu R-grupo. O átomo de carbono conectado a cadeia lateral é chamado de *carbono- $\alpha$*  e os subsequentes átomos de carbono na cadeia lateral são alfabeticamente ordenados por  $\beta, \gamma, \delta, \dots$ . Um diagrama típico de um aminoácido pode ser visto na figura 2.1.

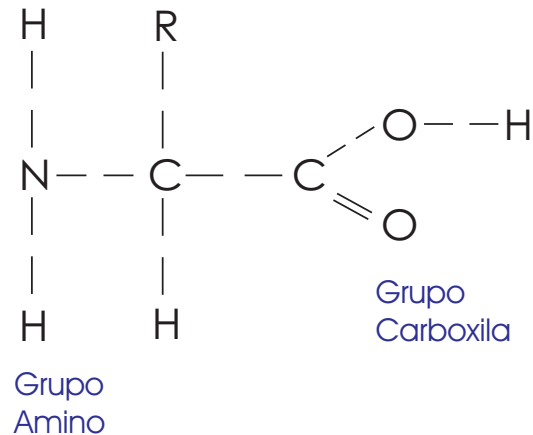


Figura 2.1: Carbono com um R-grupo ligado a ele, um grupo amino na esquerda e um grupo Carboxila na direita

Existem 20 aminoácidos que naturalmente aparecem nas proteínas e são combinados numa variedade de modos para gerar as várias proteínas e estruturas protéicas.

O que foi feito na parte experimental deste capítulo foi simplesmente tomar cada aminoácido e calcular sua árvore de Steiner. Lembremos que na solução para os pontos de Steiner, o algoritmo não sabe a priori, onde os pontos de Steiner deveriam localizar-se. O algoritmo sistematicamente adiciona  $N - 2$  pontos de Steiner para um dado conjunto de vértices terminais  $V$  afim de reduzir o comprimento global da árvore geradora que conecta  $V$ . Eles são alimentados através dos pontos médios das extremidades do conjunto  $V$ , então tenta recolocá-los de acordo com a topologia e informações obtidas da função objetivo do problema da árvore de Steiner para que satisfaçam a propriedade dos  $120^\circ$  das arestas incidentes nos pontos de Steiner. O algoritmo inicia sem qualquer informação sobre onde os pontos de Steiner deveriam estar localizados. Ele termina após analisar todas possíveis topologias e verificar que não pode reduzir o comprimento da SMT além do valor obtido. Como poderemos ver, os átomos de carbono e nitrogênio funcionam como pontos de Steiner nas estruturas de aminoácidos. A tabela 2.1 mostra os resultados que encontramos. Examinamos todos os 20 diferentes aminoáci-



dos disponíveis no site educacional PDB. Veremos que as informações podem ser obtidas a partir do cálculo da árvore de Steiner para cada aminoácido. Testes com diferentes conjuntos de dados revelaram diferentes soluções SMT e MST, ainda resultado topologicamente similares ocorreriam na tabela 2.1.

Nome	SMT	MST	$\rho$	V	$S' \subset S$	$\zeta + n$
Alanina	14.394	14.478	0.9942	13	4	4
Arginina	31.162	31.357	0.9938	27	10	10
Asparagina	19.382	19.481	0.9949	17	6	6
Ácido Aspártico	17.292	17.391	0.9943	15	5	5
Cistina	16.395	16.541	0.9912	14	4	4
Glutamina	23.071	23.186	0.9950	20	7	7
Ácido Glutâmico	20.991	21.096	0.9950	18	6	6
Glicina	10.720	10.775	0.9949	10	3	3
Histidina	23.291	23.518	0.9904	20	8	9
Isoleucina	25.417	25.586	0.9934	22	7	7
Leucina	25.447	25.588	0.9945	22	7	7
Lisina	28.912	29.102	0.9935	25	8	8
Metionina	23.891	23.990	0.9918	20	6	6
Fenilalanina	27.222	27.30	0.9972	23	9	10
Prolina	19.491	19.756	0.9866	17	6	6
Serina	15.696	15.792	0.9940	14	4	4
Treonina	19.359	19.494	0.9931	17	5	5
Triptofano	32.253	32.478	0.9931	27	12	13
Tirosina	28.473	28.562	0.9969	24	10	10
<b>Valina</b>	21.752	21.885	0.9939	19	6	6

Tabela 2.1: *Resultados dos aminoácidos*

Observando as duas colunas do lado direito da tabela, a  $S' \subset S$  coluna representa o subconjunto de pontos de Steiner que corresponde aos átomos de carbono e nitrogênio dados no aminoácido, e a última coluna representa o número de combinações de átomos de carbono e nitrogênio dentro de cada ácido. É bastante notável que o programa Steiner basicamente identifica todos os átomos de carbono e nitrogênio como pontos de Steiner a partir das primeiras propriedades. É interessante também que podemos prever a localização de certos átomos dentro da molécula sem usar Mecânica Quântica.

Na tabela 2.1, os aminoácidos alanina e glicina têm soluções ótimas SMTs. Levouse um tempo computacional razoável para achar o ótimo ( $N = 13$ ) da alanina, para a glicina ( $N = 10$ ) o tempo computacional foi bem melhor. No outro aminoácido, os resultados são cotas superiores porque o algoritmo não convergiu dentro de um tempo computacional razoável. Por exemplo, a cisteína que possui  $N = 14$  átomos, com um teste de 40 minutos, o programa não convergiu para o ótimo. O desenvolvimento de um algoritmo mais eficiente para SMTs ( $N > 13$ ) é crucial para um progresso futuro nessa área.

Alternativamente, foi desenvolvido uma heurística com tempo de convergência polinomial de  $O(N^2)$  [5], todavia queremos buscar a estrutura ótima para proteínas, essa é a razão que escolhemos um algoritmo *branch-and-bound* ([5]). Existem algumas heurísticas teóricas ([1]) que são, supostamente, melhores que o primeiro de ordem  $O(N^2)$ . Todavia não são práticos e nenhuma implementação é conhecida que esteja disponível ou até mesmo melhor que o algoritmo usado. Também o fato de querermos alcançar soluções ótimas, especialmente para um conjunto de aminoácidos com uma pequena quantidade de pontos, vemos que o algoritmo *branch-and-bound* é o mais indicado. O programa vem com um arquivo de geração de saída computacional de tal forma que a solução produzida pela árvore de Steiner pode ser visualizada no Maple.

Com excessão da histidina, fenilalanina e triptofano, o algoritmo identifica todos os átomos de carbono e nitrogênio como pontos de Steiner. Somente para a histidina ele não fez a localização de um único átomo de nitrogênio, e na fenilalanina e triptofano ele falhou na identificação do átomo de carbono. As prováveis razões para isso são a inabilidade do algoritmo de verificar exaustivamente todas as topologias cheias para as árvores de Steiner e a correta localização de todos os pontos de Steiner dentro de um tempo computacional razoável, pois o número de átomos (terminais) da histidina, fenilalanina e triptofano são relativamente grandes. O número de possibilidades de topologias cheias para um conjunto de vértices terminais dado é exponencial em  $N$  [4]. Novamente, assinalamos que estamos tentando encontrar aqui resultados ótimos. Em todos os outros casos, conseguimos identificar exatamente os átomos de carbono e nitrogênio atuando como pontos de Steiner.

Enquanto alguns átomos de carbono e nitrogênio são pontos de Steiner de grau três, muitos deles são degenerados porque seu grau é dois e seus ângulos não são iguais a  $\frac{2\pi}{3}$  rad. Nas ilustrações à esquerda das figuras 2.2 e 2.3 aparecem situações onde os átomos de carbono e nitrogênio são pontos de Steiner degenerados. Em geral, parece que a rede de Steiner resultante age como uma armação estrutural rígida para o aminoácido, bastante parecido com o Teorema de Maxwell. Mostraremos a importância desta estrutura na próxima seção quando fizermos a perturbação dessas estruturas.

Nunca os átomos de oxigênio ou ácidos sulfúricos são identificados como pontos de Steiner. Os hidrogênios são vértices terminais de grau 1, então não podem ser pontos de Steiner.

Uma questão que se abre na pesquisa é: *Existe algum significado para os pontos de Steiner que não correspondem aos átomos dados em  $V$ ?*

Também note o valor surpreendentemente alto do  $\rho$  do aminoácido,  $\rho \approx 1,0$ , que corresponde aos resultados experimentais prévios [8]. Quando consideramos o possível  $\rho$  associado a *R-sausage*, poderíamos pensar que o  $\rho$

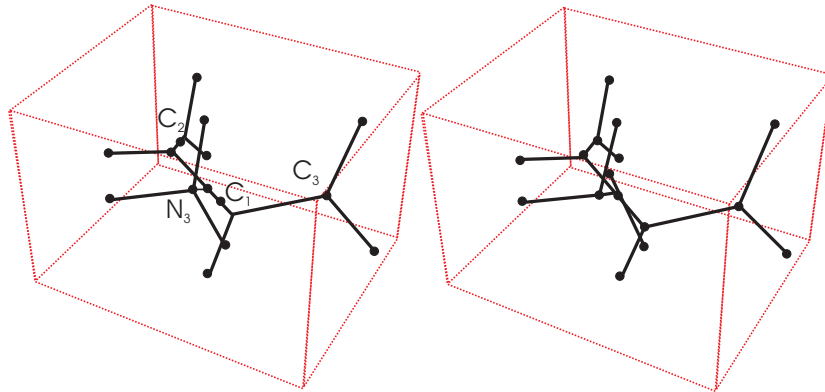


Figura 2.2: *Estrutura original à esquerda e perturbada à direita*

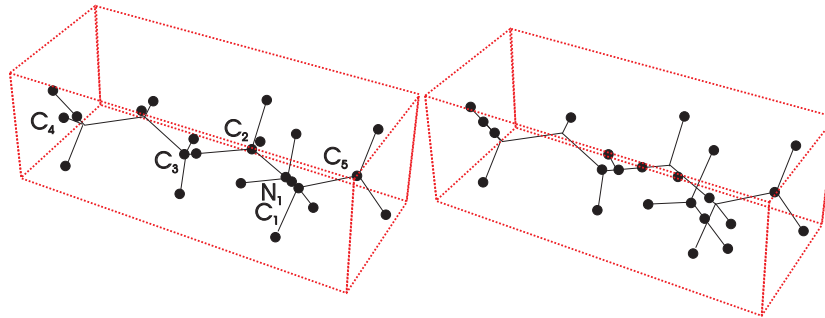


Figura 2.3: *Estrutura original à esquerda e perturbada à direita*

do aminoácido deveria ser próximo de 0.7841903733 em lugar de 1. Por que existe tal diferença no valor do  $\rho$  do aminoácido e da R-sausage? Devido o valor de  $\rho$  ser também notavelmente consistente com o valor de todos os aminoácidos, Isso indica a sua importância na estrutura de uma proteína, como devemos discutir no próximo conjunto de experiências.

Na tabela 2.2 estão ilustradas as saídas computacionais detalhadas do programa de Steiner para o ácido isoleucina, descrevendo a relação entre os átomos de carbono e nitrogênio e os pontos de Steiner.

Quando observamos as saídas computacionais dos aminoácidos, criamos duas colunas e alinhamos aqueles átomos identificados como pontos de Steiner ( $\rightarrow$ ) com aqueles dados de entrada do conjunto  $V$ , de forma que a correspondência seja clara. A indexação dos átomos nas figuras dos aminoácidos corresponde a ordem dos átomos no arquivo PDB. Pelo fato da isoleucina ter 22 átomos, o programa adicionou 20 vértices de Steiner para criar uma

Átomos	Coordenadas Terminais				Pontos de Steiner			
(C) C1:	1.2250	0.3120	0.2470	→	s27	1.2250	0.3120	0.2470
(C) C2:	-0.0650	-0.4640	-0.0120	→	s23	-0.0650	-0.4640	-0.0120
(C) C3:	-0.2480	-1.5220	1.0720	→	s38	-0.2480	-1.5220	1.0720
(C) C4:	-1.2510	0.4990	0.0100	→	s25	-1.2510	0.4990	0.0100
(C) C5:	-2.5400	-0.2760	-0.2480	→	s41	-2.5400	-0.2760	-0.2480
(N) N1:	1.4050	1.3490	-0.8150	→	s32	1.4050	1.3490	-0.8150
(C) C6:	2.3980	-0.6390	0.2250	→	s28	2.3980	-0.6390	0.2250
(O) V8:	3.1030	-0.7610	1.2130		s24	-0.0476	-0.1107	-0.0266
(O) V9:	2.6490	-1.3130	-0.8220		s26	0.9814	0.4214	0.0203
(H) V10:	1.1690	0.7950	1.2200		s29	2.3980	-0.6390	0.2250
(H) V11:	-0.0090	-0.9480	-0.9850		s30	1.3084	0.2917	0.3506
(H) V12:	2.2790	1.8740	-0.6400		s31	-0.0862	-0.6323	-0.0489
(H) V13:	0.6020	2.0010	-0.8000		s33	1.2104	1.3401	-0.7028
(H) V14:	1.4590	0.8890	-1.7410		s34	1.4905	1.3550	-0.8841
(H) V15:	-1.1210	1.2530	-0.7630		s35	-1.0692	0.5461	-0.1261
(H) V16:	-1.3080	0.9820	0.9830		s36	-1.3863	0.5040	0.1192
(H) V17:	-1.1670	-2.0750	0.8880		s37	-0.3553	-1.5083	0.9367
(H) V18:	0.5960	-2.2080	1.0560		s39	-0.1760	-1.5406	1.1596
(H) V19:	-0.3050	-1.0390	2.0450		s40	-2.5394	-1.1073	-0.2096
(H) V20:	-3.3850	0.4090	-0.2320					
(H) V21:	-2.6710	-1.0310	0.5250					
(H) V22:	-2.4840	-0.7600	-1.2210					

Tabela 2.2: *Resultados da Isoleucina*

árvore completa de Steiner (FST), de forma que existisse um total de 42 vértices ao todo. Na realidade, o programa achou duplicados pontos de pontos de Steiner  $S_{38}$  então, somente os únicos 19 pontos de Steiner são mostrados na tabela 2.2. Na realidade, 7 dos 20 vértices de Steiner adicionais que o programa encontrou coincidiram exatamente com os átomos de carbono e nitrogênio. Lembrando que, quando o programa tenta achar os vértices adicionais, ele não conhece, a priori, a localização dos pontos de Steiner. Note na figura 2.4 que os átomos  $C_5$  e  $C_6$  são pontos de Steiner de grau 3 enquanto os outros são degenerados. Deste modo, a presença dos átomos de carbono e nitrogênio dentro do aminoácido agindo como pontos de Steiner servem para fazer  $\rho$  convergir para 1, enquanto na R-sausage os pontos de Steiner no interior dela são necessários, de forma que  $\rho \rightarrow 0.7841903733$ .

## 2.5 Aminoácidos Perturbados

Se uma SMT ótima é gerada com um conjunto de vértices terminais dado  $V$  e pontos de Steiner  $S$ , então criamos um novo conjunto de vértices terminais  $V' = V \cup S$ , e ao resolvê-lo, encontramos  $\rho = 1$ . Isso ocorre porque não existe nenhum outro caminho para melhorar a SMT.

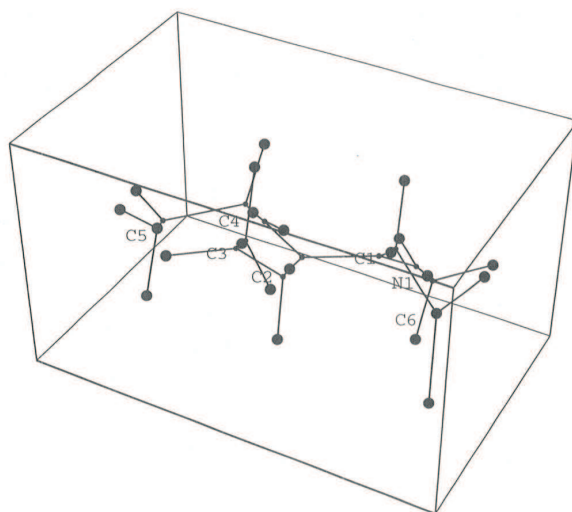


Figura 2.4: *árvore de Steiner da isoleucina*

Para testar a importância dos átomos de carbono e nitrogênio atuando como pontos de Steiner nos aminoácidos, vamos removê-los e verificar o que acontece experimentalmente com o programa de Steiner enquanto perturbamos. O programa vai substituir os átomos de carbono e nitrogênio como pontos de Steiner exatamente no mesmo lugar?

Como podemos observar na tabela 2.3, removendo os átomos de carbono e nitrogênio temos um impacto significativo na razão de Steiner, na maioria dos casos, o  $\rho$  para cada ácido cai para em torno de  $0.90 \pm 2\%$ . Isto assinala o papel significativo dos átomos de carbono e nitrogênio como pontos de Steiner.

Na tabela 2.3, as soluções da SMT perturbadas são comparadas com a topologia da árvore de Steiner das estruturas originais da alanina e metionina a esquerda de suas respectivas figuras (2.2 e 2.3). O algoritmo localiza aproximadamente novos pontos de Steiner nos locais abandonados pelos átomos de carbono e nitrogênio. Isto enfatizará mais adiante o papel da árvore e da razão de Steiner como uma cota inferior de todas as topologia da MEC global, porque em todos os casos, quando o aminoácido era reotimizado, uma cota inferior cada vez mais baixa na árvore de Steiner era percebida.

Outro fato importante é que o comprimento real da árvore de Steiner na estrutura perturbada  $SMT'$  é muito próxima em relação ao comprimento da árvore de Steiner onde os átomos de carbono e nitrogênio são incluídos no ácido. Na realidade, em todos os casos, a  $SMT' \leq SMT$  das estruturas inalteradas na tabela 2.3. Esta cota inferior é devida ao Teorema de Maxwell.

A presença dos átomos  $\zeta$  e  $\tau$  aumentará o comprimento da rede global

Nome	SMT	MST	$\rho$	SMT'	MST'	$\rho'$
Alanina	14.394	14.478	0.9942	14.355	16.459	0.8721
Arginina	31.162	31.357	0.9938	29.468	32.132	0.9171
Asparagina	19.382	19.481	0.9949	19.180	20.857	0.9196
Ácido Aspártico	17.292	17.391	0.9943	17.144	18.800	0.9119
Cistina	16.395	16.541	0.9912	16.347	18.425	0.8872
Glutamina	23.071	23.186	0.9950	23.014	25.142	0.9154
Ácido Glutâmico	20.991	21.096	0.9950	20.907	23.908	0.9051
Glicina	10.720	10.775	0.9949	10.708	12.307	0.8701
Histidina	23.291	23.518	0.9904	21.011	22.575	0.9307
Isoleucina	25.417	25.586	0.9934	25.135	28.009	0.8974
Leucina	25.447	25.588	0.9945	25.276	27.812	0.9088
Lisina	28.912	29.102	0.9935	28.674	32.636	0.8786
Metionina	23.891	23.990	0.9918	23.767	26.714	0.8897
Fenilalanina	27.222	27.300	0.9972	26.132	26.969	0.9690
Prolina	19.491	19.756	0.9866	19.062	20.793	0.9168
Serina	15.696	15.792	0.9940	15.660	17.719	0.8838
Treonina	19.359	19.494	0.9931	19.255	21.330	0.9027
Triptofano	32.253	32.478	0.9931	29.774	29.976	0.9993
Tirosina	28.473	28.562	0.9969	26.112	27.698	0.9427
Valina	21.752	21.885	0.9939	21.687	24.365	0.8901

Tabela 2.3: *Resultados dos aminoácidos perturbados*

de Steiner porque as forças de interação nos aminoácidos não são todas uniformes. Somente para a arginina ácida temos que o comprimento da árvore de Steiner perturbada é muito menor que a estrutura original. Na figura 2.4 vemos que existe uma mudança drástica na topologia da SMT a partir da estrutura original até a perturbada.

Os pontos de Steiner da forma como foram gerados nos ácidos sem os átomos de carbono e nitrogênio não são colocados exatamente nas mesmas coordenadas dos átomos de carbono e nitrogênio deslocados. Existem três razões para essa ocorrência:

- Quando os átomos de carbono e nitrogênio são removidos do aminoácido, isto reduz a cardinalidade  $N$  do aminoácido, e quando o programa da árvore de Steiner adiciona  $N - 2$  pontos de Steiner para construir a FST, os  $N - 2$  pontos de Steiner não são necessariamente os mesmos números de átomos de carbono e nitrogênio que foram removidos. Por exemplo, para a alanina,  $N = 13$  com 4 átomos de carbono e nitrogênio. Se removermos os átomos de carbono e nitrogênio,  $N = 9$  então o programa de Steiner tentará adicionar 7 vértices de Steiner invés de 11 quando a Alanina estava inalterada. Portanto, o número total de átomos de Steiner não coincidirá com o número de átomos de carbono e nitrogênio, de forma que os pontos de Steiner que são adicionados provavelmente não coincidam, e isto muda drasticamente a

estrutura da rede, como ocorreu em alguns casos.

• Os átomos de carbono e nitrogênio têm funções importantíssimas para cada aminoácido; são partes integrantes do conjunto de vértices terminais  $V$  e a presença de suas coordenadas são fundamentais para a função da estrutura ácida inteira. Na tabela 2.3, para os ácidos aromáticos histidina, fenilalanina, triptofano e tirosina, a  $MST'$  da estrutura ácida perturbada está abaixo do valor da SMT original, que não faz nenhum sentido. Portanto, removendo os átomos de carbono e nitrogênio fundamentalmente muda o comprimento mínimo da rede estrutural. Isto provavelmente pode ser melhor visto na figura 2.5 abaixo, o aminoácido triptofano. A presença da estrutura anelar afeta muito a árvore de Steiner, porque os átomos de carbono estão na parte central do anel, e sem eles, a árvore de Steiner gravita em torno dos átomos de hidrogênio.

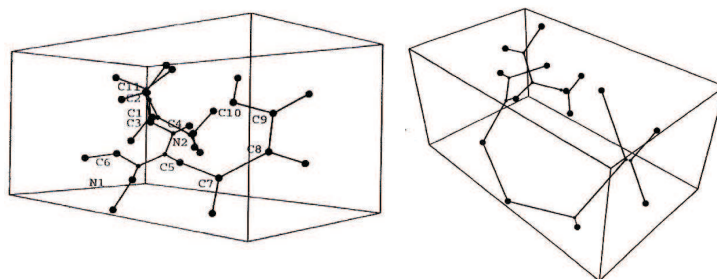


Figura 2.5: *Estrutura original e perturbada*

• As forças nos aminoácidos não são todas uniformes e essas características causam variações no teorema de Maxwell, de forma que as coordenadas de Steiner será ligeiramente diferentes. A proximidade real dos comprimentos de Steiner antes e depois da remoção dos átomos de carbono e nitrogênio atesta essas diferenças na tabela 2.3.

Assim, como foi visto, os átomos de carbono e nitrogênio dentro dos aminoácidos atuam como pontos de Steiner na estrutura química. O que precisamos analisar é o papel de Steiner dos átomos de carbono e nitrogênio na estrutura secundária das proteínas.

## 2.6 Estrutura Secundária de Aminoácidos

A estrutura secundária de uma proteína é a seqüência ligada de grupos peptídicos rígidos; Observe a figura 2.6 abaixo, descrita em ([9]).

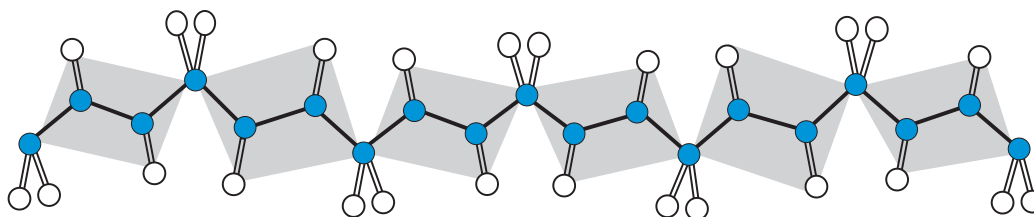


Figura 2.6: *cadeias laterais na rede de amidos-plano na proteína*

Dessa forma, enquanto os aminoácidos são ligados através das ligações peptídicas, uma molécula de água é eliminada. Para que os aminoácidos sejam colocados juntos, os mesmos são conectados por planos de ligações peptídicas, que na realidade são árvores de Steiner planares. A geometria dos planos das ligações peptídicas são de grande importância do ponto de vista de nossa modelagem de proteínas.

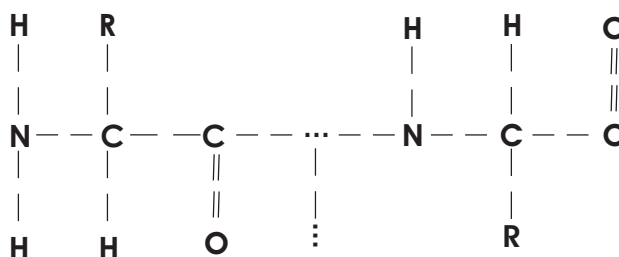


Figura 2.7: *Diagrama químico de amidos-planos na proteína*

Um diagrama químico esquemático aparece na figura 2.7 abaixo. A figura 2.8 mostra, numa visão tridimensional, as possíveis orientações com dois planos amídicos rígidos e dois graus de liberdade baseados nos ângulos de torção  $\phi$  e  $\psi$  que eles mantêm, enquanto a figura 2.6 mostra a típica conjunção dos aminoácidos numa proteína com os planos amídicos e as cadeias laterais. É bastante notável que os planos rígidos são visivelmente similares à solução da SMT para  $N = 4$  no  $\mathbb{R}^2$ . Para um exemplo de  $N = 4$ , veja à direita da figura 1.10, que se refere ao teorema de Maxwell.

Para explorar isto adiante, precisamos ver quais tipos de forças estão funcionando nestas proteínas. Na modelagem de proteínas, a função objetivo energia-potencial freqüentemente usada para medir a MEC é a seguinte, onde  $K_{bi}, K_{\theta_i}, A_{ij}, B_{ij}$  e  $\epsilon$  são pesos ajustáveis ([1]). Esta função objetivo da energia potencial é baseada no modelo de campo de força teórica usado para modelar a maioria das estruturas moleculares. É interessante ver que a função objetivo



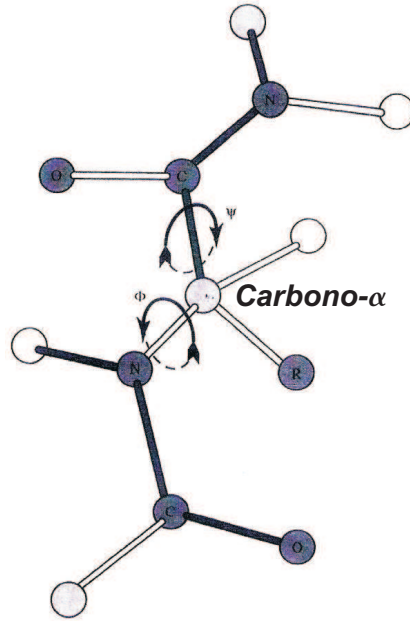


Figura 2.8: Planos amídicos com ângulos de rotação  $\phi$  e  $\psi$

é uma soma de termos não-lineares com pequena interação entre os termos. O fato que a função é separável ainda é discutível.

$$E_{total} = \sum_i K_{b_i} (b_i - b_0)^2 + \sum_i K_{\theta_i} (\theta_i - \theta_0)^2$$

$$+ \sum_i K_{\tau_i} (\cos(3\tau_i - \gamma_0)) + \sum_i \sum_j A_{ij} d_{ij}^{-6} + B_{ij} d_{ij}^{-12} + \sum_i \sum_j V_i V_j / \epsilon d_{ij}$$

ou simplesmente,

$$E_{total} = E_{bs} + E_{ab} + E_{tor} + E_{14vdw} + E_{14\epsilon}$$

onde:

$E_{bs}$ : é a soma de energias que surge a partir do estiragem ou compressão da ligação além do seu comprimento ótimo, e  $b_i$  e  $b_0$  são os comprimentos reais da ligação em equilíbrio.

$E_{ab}$ : é a soma das energias dos ângulos que estão torcidos dos seus valores ótimos, e  $\theta_i$  e  $\theta_0$  são os ângulos da ligação em equilíbrio.

$E_{tor}$ : é a soma das energias de torção que aparecem a partir de rotações sobre cada ângulo diedral respectivo, e  $\tau_i$  e  $\gamma_0$  são ângulos de torção.

$E_{14vdw}$ : é a soma das energias devido as interações de Van der Walls.

$E_{14}\epsilon$ : são interações eletrostáticas.

A geometria resultante da estrutura secundária pode ser uma  $\alpha$ -hélice, uma folha- $\beta$  ou combinações dessas estruturas.

Nesta parte do trabalho, também reexaminamos os resultados para modelagem dos conjuntos de dados do colágeno, seda, miosina e protease do HIV-1, para verificar como eles se comportam às mudanças nas configurações dos átomos da mesma forma que fizemos com os aminoácidos.

Se mudarmos a estrutura atômica do PDB, e perturbarmos os comprimentos das ligações, ângulos de ligação e ângulos de torção, que efeitos ocorrerão na razão de Steiner e sobre os termos individuais da equação de energia potencial, isto é, energia de estiramento, dobramento, torção,  $vdw$ , eletrostática, etc?.

Devido a função objetivo de energia potencial utilizada ser bastante questionável por sua simplicidade, ou seja, existem uma grande quantidade de termos que foram omitidos, resolvemos, após consulta a especialistas na área de modelagem estrutural de proteínas, não repetir a parte experimental para o cálculo da função objetivo energia potencial.

No entanto, apresentamos o procedimento feito nos trabalhos atuais ([1],[8]). O protocolo utilizado descrito nestes artigos foram o seguinte:

Todos os arquivos PDB foram baixados da Internet; os nomes dos arquivos refletem seus arquivos PDB. Para isolar um ou mais aminoácidos específicos, tais como asp-25 usados nos experimentos de protease do HIV-1. Os arquivos PDB foram lidos e editados no Word.

Os arquivos PDB puderam então ser importados para o *Chembuilder* (1997) ou lidos diretamente pelo *Sculpt*, os quais foram os dois programas usados pelos autores para avaliar as energias individuais nas proteínas. Sistemáticamente perturbamos as coordenadas atômicas dos átomos dentro das proteínas pela mudança separadamente: (a) comprimentos das ligações, (b) ângulos das ligações e (c) ângulos diedrais. O *Sculpt* foi usado somente para identificar os aminoácidos no fragmento de sítio ativo que foi de interesse no uso da função *TUG*, que permite mudar a localização de átomos isolados. O *Chembuilder* calcula os valores das energias sem os hidrogênios. Os arquivos de Steiner foram criados a partir da leitura dos arquivos PDB no Microsoft Word e criando um novo documento com as coordenadas desejadas. Todas as figuras (não mostradas aqui) usadas na protease do HIV-1 em arquivos Word foram criados no *Protein Explorer* (2004) que pode ser encontrado na internet e os arquivos PDB foram lidos lá.

## 2.6.1 Colágeno

O colágeno consiste de três cadeias polipeptídicas tecidas juntas em uma estrutura de tripla hélice, muito parecida com a da R-sausage.

O colágeno é composto de 33% glicina, 33% (prolina e hidroxiprolina) e 33% lisina. A repetição é  $(glicina - X - Y)_n$ , onde Y é prolina ou hidroxiprolina. lisina é na posição X. glicina deve ser todo o terceiro resíduo ([9]). Cada cadeia polipeptídica na estrutura do colágeno contém em torno de 1.000 aminoácidos. A hélice do colágeno é completamente estendida, diferentemente da hélice de alfa, é muito mais rígida e flexível.

O colágeno é a proteína mais abundante nos vertebrados e também aparece em espécies invertebrados, em ossos, pele, tendão, córnea e membrana de porão, e é um ligamento rígido fortemente conectado para transmitir as forças estruturais nesses tecidos ([9]).

Estrutura Original do Colágeno

		ATOM	Coordenadas Atômicas			Coordenadas dos Pontos de Steiner			
SMT=	10.1646								
MST=	10.1980	42	N1	1.6670	0.0000	0.0000	<b>1.3770</b>	<b>1.1610</b>	<b>0.8240</b>
$\rho$ =	0.9967	43	C1	1.3770	1.1610	0.8240	<b>2.3010</b>	<b>1.2150</b>	<b>2.0420</b>
$E_{tot}$ =	741.78	44	C2	2.3010	1.2150	2.0420	1.5519	0.0027	0.0672
stretch=	724.75	45	O1	2.8690	0.1990	2.4400	1.4720	1.2878	0.8565
bend=	6.39	46	H1	0.9660	-0.7160	0.0170	1.2490	1.0314	0.7889
torsion=	5.31	47	H2	1.4940	2.0690	0.2330	2.2646	1.2913	2.0124
vdw=	10.44	48	H3	0.3380	1.1270	1.1530	<b>2.4700</b>	<b>2.3960</b>	<b>2.6520</b>
electro=	0.00	49	N2	2.4700	2.3960	2.6520			
str-bnd=	-4.44	50	C3	3.3400	2.5220	3.8240			
bnd-bnd=	-0.42								
tor-str=	-0.26								

Estrutura Perturbada do Colágeno

		ATOM	Coordenadas Atômicas			Coordenadas dos Pontos de Steiner			
SMT=	8.9208								
MST=	9.5136	1	N1	0.9267	-0.9357	0.3155	0.2328	-0.5458	0.0530
$\rho$ =	0.9377	2	C1	-0.3067	-0.8414	0.1743	<b>0.0264</b>	<b>0.3432</b>	<b>0.2037</b>
$E_{tot}$ =	1004.61	3	C2	0.0264	0.3432	0.2037	0.0611	0.8608	-0.0974
stretch=	773.29	4	O1	1.0962	1.1211	0.6883	0.5382	1.2832	0.0354
bend=	89.39	5	N2	-0.4607	1.0262	-0.6433	-0.0530	-0.8024	-0.0325
torsion=	4.43	6	C3	0.5929	1.7666	-0.2791	<b>-0.3067</b>	<b>-0.8414</b>	<b>0.1743</b>
vdw=	130.06	7	H1	-0.0141	1.2813	-0.7905	0.8385	0.7428	0.0875
electro=	0.00	8	H2	-1.3018	-1.0551	0.5917			
str-bnd=	6.29	9	H3	1.4041	-0.4793	-0.4879			
bnd-bnd=	2.16								
tor-str=	-1.01								

Tabela 2.4: Resultados da estrutura original e perturbada do colágeno

O colágeno é a rede conectiva para transmissão de forças estruturais em tecidos humano e animal e é notável quando comparadas com as proprieda-

des e objetivos matemáticos do problema geométrico de Steiner e do teorema de Maxwell. O colágeno é a implementação natural do problema da árvore de Steiner. Executamos experiências de perturbação. Uma amostra das experiências aparece na tabela 2.4. Nessas experiências, devido  $N \leq 13$ , o algoritmo convergiu para a solução ótima num tempo computacional razoável. Isso pode ser visto na comparação de resultados da tabela 2.4, que o  $\rho$  de Steiner para a estrutura perturbada do colágeno diminuiu de valor e a energia da estrutura perturbada aumentou de valor: de um  $\rho = 0.996725$  com um total de energia MEC = 741.7761 Kcal para um  $\rho = 0.937688$  e um total de energia MEC = 1004.6121 Kcal. A energia de estiramento, curvatura e níveis de energia *Van der Walls* aumentaram no resíduo perturbado. O experimento é baseado na perturbação das coordenadas atômicas dos fragmentos conforme descrito no procedimento. Como o conjunto de vértices terminais é pequeno, os resultados para o algoritmo de Steiner são ótimos.

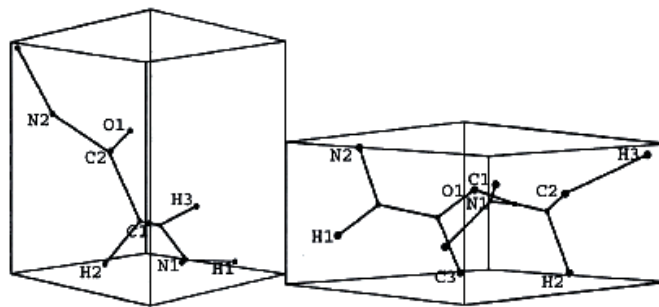


Figura 2.9: *Estrutura do colágeno*

A figura 2.9 e tabela 2.4 mostram o antes e o depois da perturbação do fragmento do Colágeno. A figura 2.9 ilustra os dois conjuntos de dados da tabela 2.4. Pode ser visto que na estrutura perturbada, diferentes pontos de Steiner são necessários e a geometria da árvore de Steiner é bastante diferente da estrutura original. Isto é o esperado, ou seja, que o nível mínimo de energia cresça na estrutura perturbada sob a energia de estiramento e transformações de rotação. Na tabela 2.4, os pontos de Steiner que correspondem aos vértices terminais estão destacados com coordenadas em negrito.

Podemos perguntar por que não aparecem mais átomos de nitrogênio e carbono como pontos de Steiner, mas isto provavelmente é devido ao fato que os átomos foram aleatoriamente selecionados dentro da cadeia polipeptídica do colágeno, e os átomos de carbono e nitrogênio as vezes agem como folhas de vértices na cadeia em vez de verdadeiros pontos de Steiner.

## 2.6.2 Seda

Uma das proteínas mais estudadas, chamada fibroína, que é o material da seda ([9]). É a proteína usada para criar casulos, teias, ninhos, etc. As cadeias peptídicas são organizadas em estruturas antiparalelas  $\beta$  pregueadas, onde as cadeias estendem-se paralelamente até o eixo da fibra. A fibroína contém grandes quantidades de glicina, alanina, tirosina e serina, também apresentam pequenas quantidades de outros aminoácidos. A repetição dos elementos de fibroína são:  $(gli - ser - gli - ala)_n$ . A glicina se estende de um lado da folha enquanto a alanina e a serina do outro. Novamente, como no colágeno, os resultados são baseados em Stanton (2005). A perturbação do comprimento da ligação, dos ângulos de ligação e de torção dos átomos da seda baixa o valor do  $\rho$ , e assim, melhoramentos no comprimento da árvore de Steiner podem ser feitos, contudo essas melhorias no comprimento causam um aumento na função potencial de energia. Os resultados contidos na tabela 2.5 abaixo, são ótimos. Desde modo, novamente, se a proteína no PDB é uma MEC, a razão de Steiner dever ser próxima de  $\approx 1.00$  e os átomos de carbono e nitrogênio devem atuar todos como pontos de Steiner.

Desta forma, as propriedades da árvore de Steiner atuam como uma ótima ferramenta para conformação de proteínas. Para ser mais específico:

- $\rho$  parece variar inversamente com a curva de energia (ângulos de ligação).
- $\rho$  parece ser independente ou afetar insignificamente a energia de estiramento (comprimento da ligação).

## 2.6.3 Miosina

A miosina aparece nos músculos e, juntamente com a proteína actina fornece propriedades catalíticas e estruturais, e mecanismos para acomodar e transferir as forças estruturais dentro dos músculos. Para produzir forças, pontes-transversais de comando dos filamentos de miosina são associados com filamentos de actina, então rotacionam-se levemente para puxar os filamentos através um do outro. Durante a interação miosina-actina, o efeito global é a contração dos músculos. Pelo fato da miosina ser basicamente uma proteína catalítica com uma estrutura globular e hélice de alfa, isso a torna uma interessante proteína a ser examinada do ponto de vista da árvore de Steiner. As duas primeiras proteínas vistas anteriormente, colágeno e seda, são consideradas proteínas estruturais, devido a miosina ser catalítica, poderíamos pensar que ela não necessariamente gozaria das propriedades da árvore de Steiner encontradas anteriormente nas proteínas estruturais colágeno e seda,

mas isso não ocorre, ou seja, ela obedece essas propriedades.

Como pode ser visto através dos resultados experimentais na tabela 2.6 abaixo, perturbações nos átomos da miosina diminuem o valor do  $\rho$  de Steiner e aumenta a energia potencial, exceto quando a energia de estiramento é significativa, como neste caso. Note que a energia de estiramento aumentou significativamente, e também o valor do  $\rho$ . No caso extremo, se todos os átomos de miosina estiverem na mesma linha, então  $\rho \rightarrow 1$ . Note também que o comprimento total da SMT aumenta além da SMT da estrutura original quando as ligações foram estiradas. Ocorre que, mudando o estiramento da ligação, ângulos de ligação e ângulos de torção, todos individualmente e simultaneamente, afetam a razão de Steiner e a equação da energia potencial. Para ser mais específico:

- $\rho$  parece variar inversamente com a curva de energia (ângulos de ligação).
- $\rho$  parece ser independente ou afetar insignificamente a energia de estiramento (comprimento da ligação).
- $\rho$  parece variar inversamente com a energia de *Van der Waals*.
- $\rho$  pode variar inversamente com a energia de torção.

## 2.6.4 Protease do HIV - 1

Finalmente, vamos para nossa última avaliação experimental, envolvendo a protease do HIV-1. A protease do HIV-1 é uma proteína catalítica (enzima) que serve para dividir outras proteínas que são necessárias ao ciclo de vida do vírus HIV. Essa divisão serve para tornar essas proteínas funcionais. Portanto, a protease do HIV é absolutamente essencial para a sobrevivência e proliferação do vírus. Por esta razão, recentemente, a protease do HIV tem sido examinada minuciosamente. Na realidade, a terapia mais popular a base de drogas usada nos Estados Unidos atualmente é a combinação de inibidores de protease e outra importante enzima inibidora, a *transcriptase reversa*. Os inibidores são simplesmente qualquer espécie de substância química ou molécula que liga a proteína diretamente ao sítio catalítico ou outro sítio, interrompendo a função própria da proteína. Ligando qualquer inibidor (ou as moléculas que são envolvidas na reação catalizada da enzima), a conformação da enzima muda devido as forças envolvidas na ligação. Um maior obstáculo para terapia inibidora é a mutação. Uma vez que um indivíduo é tratado com um inibidor de protease, os vírus que tem protease normal (tipo selvagem) são mortos pois a protease não é mais capaz de dividir a outra proteína viral essencial. Porém, existe na população de vírus, mutantes que têm protease que diferem de um a vários aminoácidos. Apesar da diferença, a enzima mutante pode ainda ser funcional. Neste caso, indivíduo tratado com inibidores pode ter se livrado do vírus com a enzima do tipo selvagem, mas não dos vírus com enzimas mutantes funcionais. Então o vírus pode proliferar e continuar causando doenças. Esse fenômeno tem aumentado as tentativas de entender e modelar os diferentes complexos de enzimas/inibidores.

A enzima da protease do HIV-1 é uma proteína dimérica, composta de duas cadeias polipeptídicas idênticas chamada de subunidades. Cada subunidade tem 99 longos resíduos de aminoácidos. O sítio ativo, ou os resíduos que são envolvidos na divisão da proteína, é encontrado no centro da proteína, simetricamente entre as duas cadeias. Na fenda da parte inferior do sítio ativo, estão dois resíduos asp cataliticamente ativos, asp25 e asp125. Neste estudo ([1]), estávamos curiosos em ver como a ligação de diferentes inibidores afetam a conformação do asp25 e asp125, bem como um extenso segmento do sítio ativo. Era esperado que a mudança conformacional induzida pela ligação dos inibidores causassem uma diferença no valor de  $\rho$ .

Da maneira como pode ser visto nas tabelas 2.7 e 2.8 que seguem.

- Para o asp25, quando limitado por diferentes inibidores  $\rho$  não teve variação significativa,  $\approx 0.3\%$ .
- Para os fragmentos do sítio ativo, novamente como no experimento com o asp25,  $\rho$  não teve variação significativa,  $\approx 0.3\%$ .

Também procuraram observar nas simetrias das ligações das cadeias. Lembremos que o sítio ativo está simetricamente colocado entre as duas subunidades. Porém, os inibidores observados são assimétricos e portanto ligam-se às duas subunidades de forma assimétrica. Era de se esperar que a ligação assimétrica causasse uma diferença no  $\rho$  entre as duas subunidades. Foi verificado que:

- Para a cadeia e ligação simétrica,  $\rho$  teve variação entre  $\approx 0.001\%$  e  $\approx 0.2\%$ .

O estudo final traz um olhar para as enzimas mutantes. Uma enzima mutante implica que a seqüência de aminoácidos da proteína mudou um pouco de forma, então muda de certa maneira sua conformação. O objetivo de todos estes estudos é determinar se estas mudanças provocariam ou não alteração no valor de  $\rho$ . Se seria de interesse determinar as energias de ligação dos vários inibidores e observar se existe uma relação com o valor de  $\rho$ .

- Para os mutantes do tipo selvagem,  $\rho$  variou entre  $\approx 0.01\%$  e  $\approx 0.1\%$ .

O que indica uma leve diferença entre as mudanças conformacionais na ligação dos vários inibidores. Mas em todas as protease do HIV estudadas, a razão de Steiner obteve pequena ou nenhuma variação. Isto parece indicar que todos os complexos de protease/inibidores do HIV são estruturas MEC bastante estáveis.



## Experimentos com a Seda

<b>seda<sub>gs</sub></b>		<b>seda<sub>gs</sub> - mod</b>	
estado natural		(somente comprimento da ligação)	
<b>SMT=</b>	11.2564	<b>SMT=</b>	13.1680
<b>MST=</b>	11.2809	<b>MST=</b>	13.1928
<b>ρ=</b>	0,9978	<b>ρ=</b>	0.9981
<b>E<sub>tot</sub>=</b>	149.22	<b>E<sub>tot</sub>=</b>	333.75
<b>stretch=</b>	142.43	<b>stretch=</b>	317.11
<b>bend=</b>	7.89	<b>bend=</b>	7.87
<b>torsion=</b>	3.51	<b>torsion=</b>	3.51
<b>vdw=</b>	1.82	<b>vdw=</b>	-0.15
<b>electro=</b>	0.00	<b>electro=</b>	0.00
<b>str-bnd=</b>	-5.90	<b>str-bnd=</b>	-5.98
<b>bnd-bnd=</b>	-0.39	<b>bnd-bnd=</b>	-0.40
<b>tor-str=</b>	-0.16	<b>tor-str=</b>	-0.16

<b>seda<sub>gs</sub> - mod2</b>		<b>seda<sub>gs</sub> - mod3</b>	
(somente ângulos de ligação)		(somente ângulo de torção)	
<b>SMT=</b>	10.9333	<b>SMT=</b>	11.2438
<b>MST=</b>	11.2807	<b>MST=</b>	11.2791
<b>ρ=</b>	0.9692	<b>ρ=</b>	0.9969
<b>E<sub>tot</sub>=</b>	219.94	<b>E<sub>tot</sub>=</b>	191.37
<b>stretch=</b>	147.19	<b>stretch=</b>	142.49
<b>bend=</b>	60.06	<b>bend=</b>	7.89
<b>torsion=</b>	3.61	<b>torsion=</b>	3.48
<b>vdw=</b>	8.55	<b>vdw=</b>	43.97
<b>electro=</b>	0.00	<b>electro=</b>	0.00
<b>str-bnd=</b>	-0.99	<b>str-bnd=</b>	-5.90
<b>bnd-bnd=</b>	1.69	<b>bnd-bnd=</b>	-0.39
<b>tor-str=</b>	-0.16	<b>tor-str=</b>	-0.18

Tabela 2.5: *Resultados da estrutura da seda original e perturbada*

## Experimentos com a Miosina

<b>1scm82-3</b>		<b>1scm82-3-mod</b>	
estado natural		(somente comprimento da ligação)	
<b>SMT=</b>	14.0359	<b>SMT=</b>	19.0895
<b>MST=</b>	14.1566	<b>MST=</b>	19.2228
<b><math>\rho</math>=</b>	0.9915	<b><math>\rho</math>=</b>	0.9931
<b><math>E_{tot}</math>=</b>	164.9216	<b><math>E_{tot}</math>=</b>	5311.92
<b>stretch=</b>	144.47	<b>stretch=</b>	5275.35
<b>bend=</b>	14.89	<b>bend=</b>	14.81
<b>torsion=</b>	4.92	<b>torsion=</b>	4.97
<b>vdw=</b>	7.45	<b>vdw=</b>	3.02
<b>electro=</b>	0.00	<b>electro=</b>	0.00
<b>str-bnd=</b>	-6.34	<b>str-bnd=</b>	14.24
<b>bnd-bnd=</b>	-0.20	<b>bnd-bnd=</b>	-0.20
<b>tor-str=</b>	-0.27	<b>tor-str=</b>	-0.26

<b>1scm82-3-mod2</b>		<b>1scm82-3-mod3</b>	
(somente ângulos de ligação)		(somente ângulos diedrais)	
<b>SMT=</b>	13.3978	<b>SMT=</b>	14.0204
<b>MST=</b>	14.2030	<b>MST=</b>	14.1565
<b><math>\rho</math>=</b>	0.9433	<b><math>\rho</math>=</b>	0.9904
<b><math>E_{tot}</math>=</b>	324.37	<b><math>E_{tot}</math>=</b>	196.17
<b>stretch=</b>	137.35	<b>stretch=</b>	144.40
<b>bend=</b>	144.10	<b>bend=</b>	14.89
<b>torsion=</b>	4.94	<b>torsion=</b>	5.51
<b>vdw=</b>	40.05	<b>vdw=</b>	38.12
<b>electro=</b>	0.00	<b>electro=</b>	0.00
<b>str-bnd=</b>	-4.97	<b>str-bnd=</b>	-6.34
<b>bnd-bnd=</b>	3.13	<b>bnd-bnd=</b>	-0.20
<b>tor-str=</b>	-0.23	<b>tor-str=</b>	-0.22

Tabela 2.6: *Resultados da estrutura da miosina original e perturbada*

## Protease do HIV-1 sem inibidores

SMT= 9.6945		3PHV Asp25: sem limitação de inibidor							
MST= 9.7235									
p=	0.9970	ATOM	Coordenadas Atômicas			Coordenadas dos Pontos de Steiner			
E <sub>tot</sub> =	165.7153	1	N	-1.4453	1.4033	-1.1451	<b>-0.4288</b>	<b>0.7857</b>	<b>-0.3196</b>
stretch=	159.79	2	C	-0.4288	0.7857	-0.3196	0.9059	1.3927	-0.2877
bend=	7.04	3	C	0.9124	1.4193	-0.2572	-0.2392	0.6847	-0.3329
torsion=	1.80	4	O	1.7517	1.2422	-1.1559	-0.1990	-0.6131	-0.4574
vdw=	2.97	5	C	-0.3118	-0.7183	-0.6369	0.4841	-1.3541	0.4354
electro=	0.00	6	C	0.4742	-1.4618	0.4351	<b>0.4742</b>	<b>-1.4618</b>	<b>0.4351</b>
str-bnd=	-5.42	7	O	1.1872	-0.8308	1.2178			
bnd-bnd=	-0.45	8	O	0.3979	-2.7064	0.5425			
tor-str=	-0.01								

SMT= 9.6247		3PHV Asp25: com limitação do inibidor SDZ283-910							
MST= 9.6475									
p=	0.9976	ATOM	Coordenadas Atômicas			Coordenadas dos Pontos de Steiner			
E <sub>tot</sub> =	202.8486	1	N	-0.2092	-2.5294	-1.8228	<b>-0.8588</b>	<b>-1.2708</b>	<b>-1.7076</b>
stretch=	197.25	2	C	-0.8588	-1.2708	-1.7076	<b>-2.1105</b>	<b>-1.1909</b>	<b>-0.8550</b>
bend=	7.81	3	C	-2.1105	-1.1909	-0.8550	-0.8101	-1.3734	-1.8843
torsion=	2.01	4	O	-3.2259	-1.4534	-1.3067	-1.1326	-0.8122	-3.1369
vdw=	1.82	5	C	-1.1326	-0.8122	-3.1369	<b>-1.4454</b>	<b>0.6714</b>	<b>-3.3016</b>
electro=	0.00	6	C	-1.4454	0.6714	-3.3016	-1.4401	0.6593	-3.3187
str-bnd=	-5.65	7	O	-1.7525	1.3721	-2.3535			
bnd-bnd=	-0.38	8	O	-1.3909	1.1512	-4.4099			
tor-str=	-0.00								

Tabela 2.7: Resultados da estrutura da protease do HIV-1 com e sem inibidores

## Protease do HIV-1 com inibidores

SMT= 9.6548		1HVR Asp25: com limitação do inibidor cíclico não-peptídico Urea Xk263							
MST= 9.6805									
	$\rho$ = 0.9973	ATOM	Coordenadas Atômicas			Coordenadas dos Pontos de Steiner			
$E_{tot}$ =	204.10	1	N	-2.3499	0.4610	-0.8474	<b>-1.2631</b>	<b>-0.3847</b>	<b>0.1335</b>
stretch=	199.05	2	C	-1.2631	0.3847	0.1335	<b>0.1747</b>	<b>1.4590</b>	<b>-0.0521</b>
bend=	7.37	3	C	-0.1747	1.4590	-0.0521	-1.3303	0.2262	-0.0062
torsion=	1.99	4	O	0.7466	1.3481	-0.8589	<b>-0.6887</b>	<b>-1.0240</b>	<b>-0.0115</b>
vdw=	1.96	5	C	-0.6887	-1.0240	-0.0115	<b>0.2039</b>	<b>-1.4282</b>	<b>1.1166</b>
electro=	0.00	6	C	0.2039	-1.4282	1.1166	0.1835	1.4567	1.0937
str-bnd=	-5.87	7	O	0.7986	-0.6007	1.7930			
bnd-bnd=	-0.38	8	O	0.3064	-2.6030	1.3324			
tor-str=	-0.01								

SMT= 9.5923		1AID Asp25: com limitação do inibidor não-peptídico THK							
MST= 9.6209									
	$\rho$ = 0.9970	ATOM	Coordenadas Atômicas			Coordenadas dos Pontos de Steiner			
$E_{tot}$ =	193.5667	1	N	-2.2188	0.0988	-1.2941	<b>1.3826</b>	<b>0.5137</b>	<b>-0.2115</b>
stretch=	187.97	2	C	-1.3826	0.5137	-0.2115	1.5949	-0.2172	1.0806
bend=	7.52	3	C	1.5970	-0.2160	1.0829	1.3187	0.4406	-0.4190
torsion=	1.95	4	O	1.0040	-1.2486	1.2941	-0.0180	0.3633	-0.7401
vdw=	2.12	5	C	-0.0180	0.3633	-0.7401	<b>-1.1380</b>	<b>0.9339</b>	<b>0.0994</b>
electro=	0.00	6	C	-1.1380	0.9339	0.0994	-1.1457	0.9190	0.0471
str-bnd=	-5.61	7	O	0.9750	1.2648	1.2745			
bnd-bnd=	-0.37	8	O	-2.2188	1.0412	-0.4545			
tor-str=	-0.02								

Tabela 2.8: *Resultados da estrutura da protease do HIV-1 com e sem inibidores*

## Capítulo 3

### Considerações Finais

**Conjectura.** Para uma seqüência arbitrária de  $n$ -resíduos de aminoácidos denotadas por  $\{\rho_1, \rho_2, \dots, \rho_n\}$  dentro de uma cadeia protéica, a menor razão de Steiner é limitada inferiormente pela razão de Steiner dos resíduos de aminoácidos na cadeia, isto é,  $\rho_{proteina} \geq \min\{\rho_1, \rho_2, \dots, \rho_n\}$ .

Obviamente, se existem impurezas e outras substâncias desconhecidas agindo na proteína, então talvez a conjectura pode não valer. Idealmente, a propriedade deveria valer devido a natureza de como são reunidos juntos os aminoácidos numa ligação de planos peptídicos rígidos. As razões pelas quais a conjectura deve valer são:

- Existe uma propriedade das árvores de Steiner, ou seja,  $\rho$ , que dá uma medida quantitativa adimensional para cada aminoácido individualmente, bem como da estrutura secundária das proteínas, e é uma importante assinatura para cada aminoácido assim como para cada proteína.

- As árvores de Steiner procuram a rede mais curta possível. No teorema de Maxwell, esta rede mais curta é equivalente à configuração de mínima energia, que é a minimização da energia potencial num sistema com forças uniformes em cada vértice terminal. Infelizmente, as forças não são todas uniformes numa proteína, de forma que a SMT pode resultar em somente uma cota inferior para a topologia MEC. Especialmente nos experimentos com aminoácidos. Contudo, mostramos que este comprimento da cota inferior para cada aminoácido é muito estreito.

- Dos experimentos de aminoácidos neste trabalho, os átomos de carbono e nitrogênio atuam todos como pontos de Steiner. Isto ocorreu nos aminoácidos não-perturbados bem como nos perturbados. Como mostrados na figura 2.4 ao longo de nosso estudo prévio experimental na estrutura secundária ([8]), os planos rígidos das ligações peptídica são árvores de Steiner interconectadas e mantêm uma razão de Steiner constante, de forma que a única

diferença no  $\rho$  para uma proteína é devida às cadeias laterais. Acreditamos que isto é verdade porque nas avaliações das estruturas secundárias, terciárias e quaternárias das proteínas, todos os átomos de carbono e nitrogênio devem atuar como pontos de Steiner. Também parece indicar que, numa seqüência arbitrária de aminoácidos, a razão de Steiner da proteína inteira, não pode ser menor que a razão de Steiner dos resíduos de aminoácidos na cadeia.

Uma pesquisa futura poderá se fixar mais no estudo de sistemas maiores de átomos que realmente incorporam uma boa quantidade de estruturas secundárias ao invés de apenas focar em aminoácidos isolados. Também pretende-se pesquisar o valor da razão de Steiner em outras superfícies buscando confirmar ou até mesmo melhorar o valor que temos até então, que é o da R-sausage. Ainda, contudo, melhoramentos no algoritmo da árvore de Steiner precisam ser feitas para permitir encontrar soluções ótimas para  $N$  grandes. Uma faceta que deve ser valiosa aqui para proteínas é a regularidade da topologia de espinha dorsal e do conhecimento das topologias da estrutura Steiner dentro de cada aminoácido.

Também, precisamos usar ferramentas mais sofisticadas e softwares técnicos para conduzir as perturbações das localizações dos átomos e a medida das energias resultantes. Finalmente, precisamos explorar, como as SMTs podem ser usadas mais diretamente para predizer e avaliar conformações de estruturas de proteínas de baixa energia no problema de enovelamento de proteínas.

Resumidamente, apresentamos uma revisão do problema da árvore de Steiner e sua importância potencial para modelar proteínas. As estruturas primárias dos aminoácidos foram avaliadas com o programa de Steiner, e mostramos claramente que os átomos de carbono e nitrogênio atuam como pontos de Steiner. Verificamos também que quando os átomos de carbono e nitrogênio são removidos da estrutura dos aminoácidos, a razão de Steiner diminuiu drasticamente seu valor, dando assim mais evidências para a importância dos átomos de carbono e nitrogênio na estrutura da árvore de Steiner. Finalmente, na estrutura secundária das proteínas, verificamos com auxílio de dados ([1]), que existe uma correlação próxima entre a razão de Steiner e a função de energia das proteínas. Em geral, com a perturbação da estrutura das proteínas a partir da estrutura original, a razão de Steiner mostrou-se inversamente proporcional ao potencial de energia dentro da proteína. Não estaremos aqui esgotando o assunto abordado mas, apenas iniciando uma pesquisa que acredita-se ter grandes aplicações e importâncias em parceria com a Biologia e, possivelmente com outras áreas.

## Referências Bibliográficas

- [1] STANTON, C., SMITH, J.M., *Steiner Trees and 3 – D Macromolecular Conformation..* Vol. 16, n.04., Informs Journal on Computing, 2004.
- [2] COURANT, R., ROBBINS, H., *What is Mathematics?.* 1 ed., New York, Oxford University Press, 1941.
- [3] MELZAK, Z.A. "On the Problem of Steiner", *Canadian Mathematics Bulletin*, v.4, pp. 143-148, 1961.
- [4] GILBERT, E. N, POLLAK, H.O. "Steiner Minimal Trees", *SIAM J. Appl. Math.*, v. 16, n.1, pp. 1-29, 1968.
- [5] SMITH, W.D. "How to Find Steiner Minimal Trees in Euclidean  $d$ -Space", *Algorithmica*, v. 7, pp. 137-177, 1992.
- [6] SMITH, J.M, SMITH, W.D. "On the Steiner Ratio in 3 - Space", *Journal of Combinatorial Theory, A*, v. 69, pp. 301-332, 1995.
- [7] HWANG, F.K., RICHARDS, D.S., WINTER, P. *The Steiner Tree Problem.* North-Holland, 1992.
- [8] SMITH, J.M., TOPPUR, B. "Euclidean Steiner Minimal Trees, Minimal Energy Configurations, and the Embedding Problem of Weighted Graphs in  $E^3$ ". *Discrete Applied Mathematics*, v.71, pp. 187-215, 1996.
- [9] NELSON, L. D., COX, M.M., *Princípios de Bioquímica.* 3a. edição, editora Sarvier, São Paulo, 1999.