

NANOMEDICINA: UMA ABORDAGEM PARA FUTURAS CONSTRUÇÕES DE PROTEÍNAS UTILIZANDO NEUROCOMPUTAÇÃO

Aluno: Thiago da Silva Ribeiro

Orientador: Marco Aurélio C. Pacheco

Co-Orientadores: Reinaldo Bellini Gonçalves, Karla Figueiredo

Introdução

A. Objetivo e importância do presente trabalho

Aminoácidos são pequenos blocos de construção de estruturas corporais denominadas proteínas e estas macromoléculas determinam a vida tal como a conhecemos. É válido ressaltar que a palavra proteína é derivada do grego “protos” que significa “de primordial importância”. A sequência de aminoácidos em uma cadeia protéica determina o seu dobramento[2], o que torna esta proteína especializada em uma determinada função corporal, seja ela a queratina, encontrada no cabelo, ou proteínas como as β -amiloides, que estão relacionadas com o mal de Alzheimer. Desvendar o dobramento destas estruturas torna-se mister para entendermos como combater doenças encontradas nos organismos. Conhecendo a estrutura tridimensional podemos não só desenvolver medicamentos que atuam em uma determinada proteína, como também mudar determinados aminoácidos nesta sequência para tratamento de doenças que são consequência de um dobramento imperfeito na estrutura protéica, tais como Alzheimer e Príon, conhecida como “mal da vaca louca”.

Um enorme desafio precisa ser ultrapassado para que a bionanotecnologia comece a gerar aplicações: prever o dobramento da estrutura da proteína apenas através de sua sequência química, é fundamental.

O problema do dobramento proteico impõe grandes dificuldades, em especial, por dois motivos:

- O primeiro é a magnitude do problema, pois as proteínas possuem em média centenas de aminoácidos. Cada um destes pequenos blocos de construção está ligado aos seus vizinhos através de duas ligações flexíveis, que podem adotar uma série de conformações. Além disso, cada aminoácido tem uma cadeia lateral flexível que pode adotar uma série de conformações. Juntos, esses vários níveis de liberdade definem um espaço incrivelmente grande de conformação que ultrapassa todos os atuais métodos de predição computacional;
- O segundo motivo está no método utilizado para estimar a estabilidade de cada conformação durante um experimento de predição. A estrutura da proteína enovelada possui milhares de contatos internos, cada qual acrescenta um pequeno incremento de estabilização na estrutura como um todo. Muitas moléculas de água são liberadas durante o processo de dobramento. Esta liberação de moléculas de água é uma importante força empurrando as proteínas em direção a uma estrutura dobrada.

Juntos, o grande espaço de busca e os erros acumulados têm dificultado as inúmeras tentativas de prever o dobramento proteico.

O carbono α de um aminoácido possui um papel importante na estrutura proteica. Ao descrever uma proteína (que é uma cadeia de aminoácidos), muitas vezes aproxima-se a localização de cada aminoácido com a localização de seu carbono α .

As estruturas tridimensionais de proteínas são de grande complexidade computacional, visto que são problemas NP-completos difíceis de serem tratados computacional e matematicamente devido a grande explosão combinatória de modos de dobramento entre os aminoácidos.

Para a abordagem do problema propõem-se a utilização de uma estrutura da neurocomputação que tenta imitar o funcionamento do cérebro, através da interação entre neurônios, denominada redes neurais[3].

Este trabalho tem como objetivo realizar um estudo sobre o enovelamento de proteínas através de propriedades geométricas entre carbonos α , utilizando como padrões propriedades físico-químicas dos aminoácidos, modeladas através de redes neurais artificiais. Permitindo assim que sejam elucidadas características que possam ser úteis na descoberta de como as proteínas se enovelam biologicamente.

B. Fundamentação teórica

a. Estruturas

A estrutura primária de uma proteína é a sequência de aminoácidos na cadeia polipeptídica, conforme mostra o esquema da Figura 1. Esta estrutura pode ser pensada como uma descrição completa de todas as ligações covalentes em uma proteína. A sequência de aminoácidos determina a estrutura tridimensional da proteína, que por sua vez determina sua função biológica. A alteração na estrutura primária pode produzir resultados catastróficos (Figura 2).

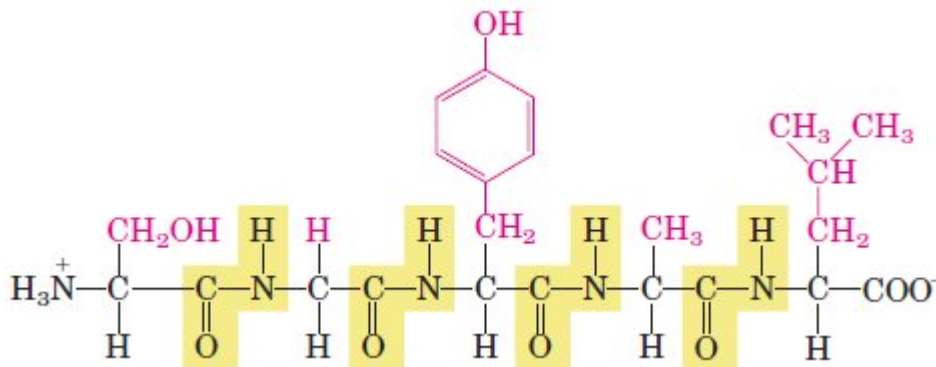


Figura 1. Estrutura primária, determinada pela sequência de aminoácidos. Fonte: Leningher (2002). Princípios de Bioquímica, editora- Sarvier, pag-86



Figura 2. Uma única troca na cadeia de aminoácidos da hemoglobina (substituição de um ácido glutâmico por uma valina) acaba por determinar mudanças na hemácia (anemia falciforme). Fonte: Bettelheim & March (1990) *Introduction to Organic & Biochemistry (International Edition)* Philadelphia: Saunders College Publishing, p301

A estrutura secundária de uma proteína é o arranjo ordenado de aminoácidos nas regiões localizadas de um polipeptídeo. As ligações de hidrogênio desempenham um papel importante na estabilização desses padrões de dobradura. As duas principais estruturas secundárias são a α -hélice (Figura 3) e a folha- β pregueada (Figura 4). Existem outras conformações periódicas, mas estas são as mais estáveis. Uma única proteína pode conter múltiplas estruturas secundárias.

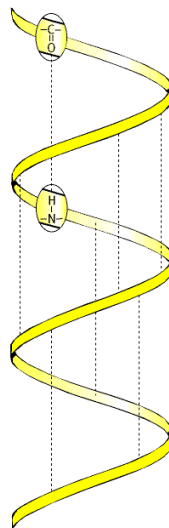


Figura 3. Estrutura de uma α -hélice. Fonte: Elliott, WH. Elliott, DC. (1997) *Biochemistry and Molecular Biology*. Oxford: Oxford University Press. p28

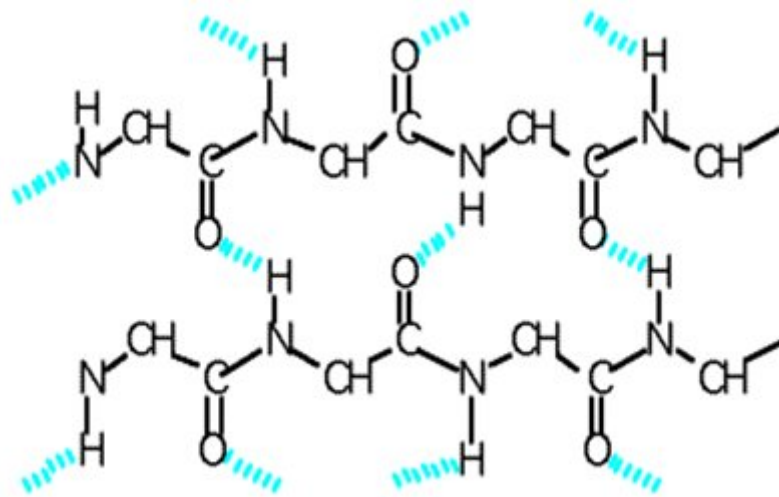


Figura 4. Estrutura de uma folha- β pregueada. Fonte: Elliott, WH. Elliott, DC. (1997) *Biochemistry and Molecular Biology*. Oxford: Oxford University Press. p29

A Figura 5 ilustra uma estrutura terciária, que é a estrutura tridimensional de uma dada proteína. Incluem-se nesta descrição a relação espacial das diferentes estruturas secundárias de uma cadeia polipeptídica e como estas estruturas se dobram para formar a estrutura tridimensional da proteína. A estrutura terciária descreve também a relação de domínios diferentes uns dos outros dentro de uma proteína. As interações de diferentes domínios são regidas por diversas forças intermoleculares: ligações de hidrogênio, interações hidrofóbicas, interações eletrostáticas e forças de van der Waals.



Figura 5. Estrutura terciária de uma proteína

Muitas proteínas contêm duas ou mais cadeias polipeptídicas diferentes, estas são mantidas unidas pelas mesmas forças que estabilizam a estrutura terciária de uma proteína. Proteínas com várias cadeias de polipeptídeos são proteínas oligoméricas. A estrutura formada pela interação monômero-monômero em uma proteína oligomérica é conhecida como estrutura quaternária.

b. Forças que atuam no dobramento proteico

Ligação de hidrogênio: São interações que ocorrem entre os íons de hidrogênio e dois ou mais átomos, de forma que o hidrogênio atua como o átomo central entre os átomos com os quais interage. Polipeptídios possuem numerosos doadores e receptores de prótons localizados tanto na cadeia principal quanto na cadeia lateral dos seus aminoácidos. O ambiente no qual a proteína se encontra também possui numerosos doadores e receptores de prótons, uma vez que a mesma se encontra em meio aquoso.

Interações hidrofóbicas: As cadeias laterais dos aminoácidos podem ser classificadas como hidrofílicas ou hidrofóbicas. A interação das diferentes cadeias laterais da proteína com o meio aquoso possui grande participação no processo de dobramento. Portanto, o dobramento de uma proteína é reflexo de um equilíbrio entre as ligações de hidrogênio entre as cadeias laterais hidrofílicas e o meio aquoso, em oposição, a repulsão ao meio pelas cadeias laterais hidrofóbicas. Estas interações ocorrem devido ao efeito denominado efeito hidrofóbico, onde moléculas apolares interagem muito mais entre elas do que com outras moléculas do meio.

Forças eletrostáticas: As forças eletrostáticas são em sua maioria de três tipos: ligação iônica, íon-dipolo, dipolo-dipolo. No dobramento da proteína, ligações iônicas acontecem entre cadeias laterais com cargas opostas, por exemplo, entre arginina e lisina ou entre aspartato e glutamato. As interações íon-dipolo possuem uma participação importante no enovelamento proteico. Estas se referem à interação entre cadeias laterais ionizadas e o momento dipolo da molécula de água. O momento dipolo existente nas cadeias laterais polares também influencia sua interação com a água.

Forças de van der Waals: Embora as forças de van der Waals sejam extremamente fracas, em comparação com as demais forças que participam do dobramento da proteína, é o grande número destas interações que faz a mesma apresentar um papel significativo no enovelamento proteico.

A Figura 6 ilustra a participação destas forças na estrutura da proteína.

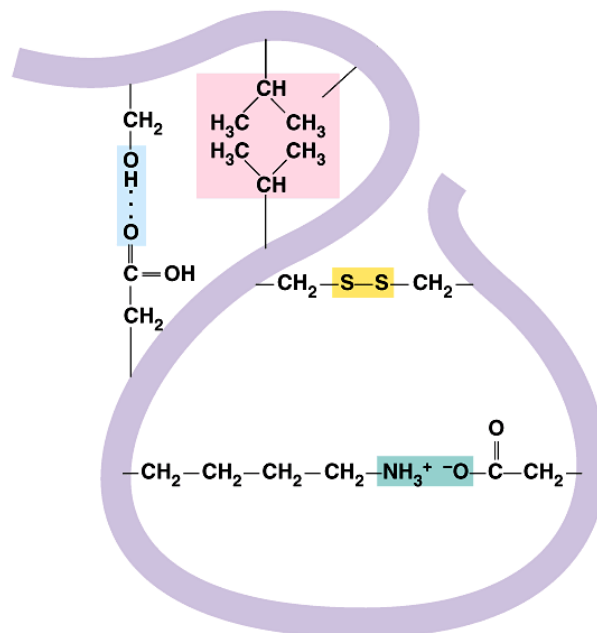


Figura 6. Forças envolvidas no dobramento proteico. Fonte: Freeman, S. (2007) Biological Science. University of Washington: Benjamin Cummings

c. Biofísica de proteínas

A estabilidade termodinâmica de uma proteína é medida pela diferença de energia livre entre o estado dobrada e o estado não dobrada. Esta variação de energia possui um efeito profundo sobre a função da proteína. A contribuição energética das forças de dobramento favoráveis, tais como interações hidrofóbicas, ligações de hidrogênio e interações eletrostáticas, são quase compensadas pela penalização entrópica do dobramento.

A conformação nativa de uma proteína representa o estado de menor energia livre do sistema. A energia que estabiliza a estrutura 3D de uma proteína vem do aumento da entropia do solvente em questão.

d. Redes Neurais Artificiais

Redes Neurais Artificiais são técnicas computacionais inspiradas nos neurônios biológicos e na estrutura do cérebro, nas quais o objetivo é obter um sistema com estrutura paralela e eficiente como o cérebro humano (Figura 7a). Uma rede neural possui a capacidade de aprender com a experiência, armazenar dados e gerar conclusões que podem ser mais eficientes do que as conclusões geradas pelo método tradicional, ou que o próprio especialista não tem capacidade de gerar.

A Figura 7b ilustra o neurônio artificial, um processador que possui uma função de ativação que determina o grau de ativação do neurônio. Todas as suas conexões de entrada possuem um peso sináptico que determinam o efeito da informação de entrada nesse processador. O ajuste equilibrado desses pesos é que resultam no aprendizado e no armazenamento de informações da rede neural.

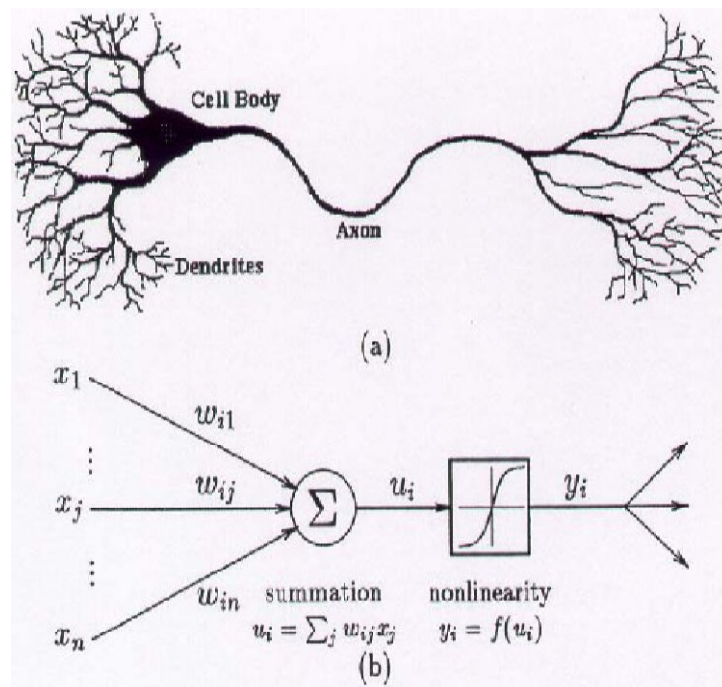


Figura 7. Imagem ilustrativa de um neurônio biológico e um neurônio artificial

O desempenho da rede neural é avaliado de acordo com o valor dos erros obtidos na camada de saída que se referem ao nível de acerto da rede, onde os valores gerados pela rede são comparados com os valores originais do banco de dados. Quanto mais próximo de zero for o erro, melhor o resultado.

Metodologia

Como uma solução do problema, foi proposto e construído um modelo de rede neural para previsão geométrica do ângulo entre os carbonos α dos aminoácidos. Esta informação permite observar a conformação da cadeia principal da proteína, não levando em consideração as cadeias laterais, visto a grande complexidade computacional de se analisar o problema como um todo.

As estruturas de proteínas utilizadas neste experimento foram obtidas a partir de dados cristalográficos depositados no banco de dados de proteínas (Protein Data Bank)[1]. Após coletados, os dados foram pré-processados. O pré-processamento constituiu de limpeza da base: extração das informações do carbono α e cálculo do ângulo entre os mesmos.

Foram utilizadas 49932 tríades de aminoácidos, divididas em 3 partes: 30549 para treino da rede, 10183 para realizar o *early-stopping* e 9200 para teste. Foram utilizados como atributos de entrada para rede os seguintes parâmetros:

- Polaridade;
- Carga da cadeia lateral;
- Hidrofobia;
- Somatório da massa média dos resíduos;
- Volume médio de resíduos sem contato com o solvente;
- van der Waals.

Desta forma, a rede neural deveria ser capaz de encontrar os valores angulares entre três aminoácidos de uma proteína, dado a sequência de aminoácidos. Este fato seria importante para, futuramente, a construção de um modelo que descreveria a topologia da proteína, dada uma determinada sequência de tríades de aminoácidos.

A configuração de rede neural de melhor desempenho encontrada possui:

- **O algoritmo Backpropagation:** é um algoritmo para treinamento de redes multi-camadas mais difundido. Baseia-se no aprendizado supervisionado por correção de erros. Durante a fase de treinamento deve-se apresentar um conjunto formado pelo par: entrada para a rede e valor desejado para resposta à entrada. A saída será comparada ao valor desejado e será computado o erro global da rede, que influenciará na correção dos pesos no passo de retropropagação;
- **Neurônios da camada escondida:** foram utilizados 14 neurônios;
- **Função de ativação:** tangente hiperbólica;
- **Número de épocas:** 1000;

Resultados

O erro médio encontrado na rede neural para o conjunto de teste de aminoácidos foi de 12.81%. O problema requer uma alta precisão, uma vez que pequenas mudanças na geometria da proteína podem modificar sua função. Mesmo não considerando uma vizinhança que possa interferir no dobramento, a rede neural foi capaz de encontrar os valores angulares previstos próximos ao real (Tabela 1), tornando as cadeias real e prevista com geometria semelhante (Figura 8). Desta forma, este resultado comprova que uso de redes neurais para tratar o problema de enovelamento de proteínas pode proporcionar avanços para o melhor entendimento deste problema.

<i>Real</i>	<i>Previsto</i>	<i>Erro %</i>
94,7710	106,9131	12,812%
96,8904	109,3001	12,808%
130,2468	113,5652	12,808%

Tabela 1. Diferença entre o ângulo real e previsto, considerando o erro percentual médio encontrado

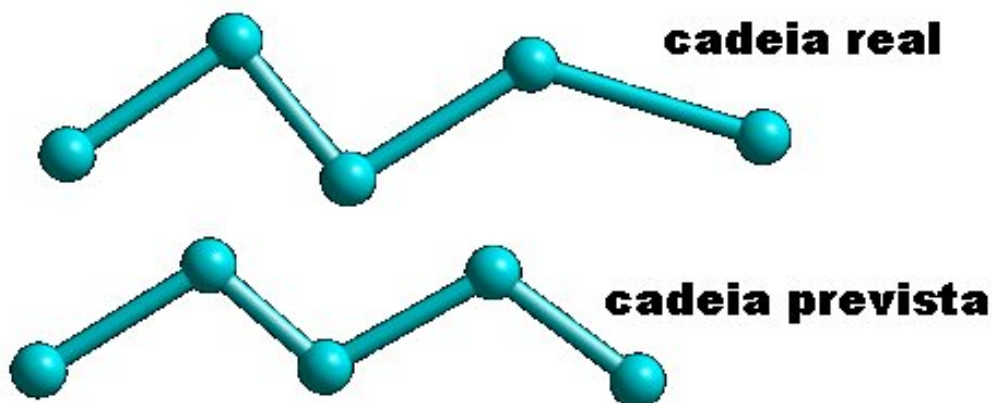


Figura 8. Diferença conformacional entre as cadeias real e prevista, considerando o erro percentual médio encontrado

<i>Real</i>	<i>Previsto</i>	<i>Erro %</i>
96,3180	96,3308	0,013%
99,8364	99,8276	0,009%
107,1139	107,1179	0,004%

Tabela 2. Diferença entre o ângulo real e previsto, considerando os menores erros percentuais encontrados

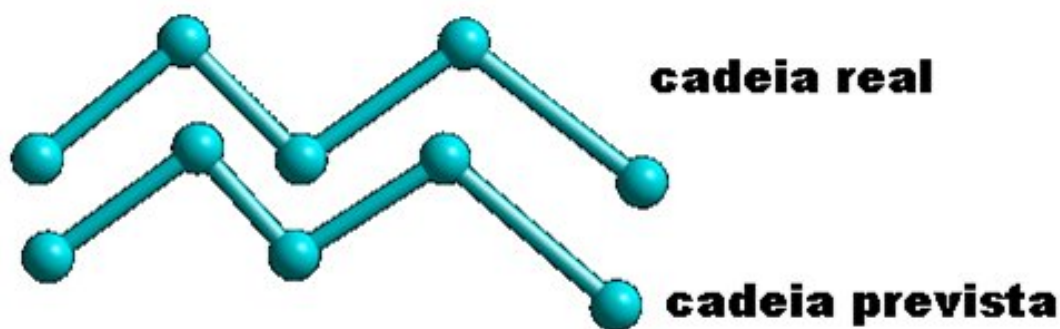


Figura 9. Diferença conformacional entre as cadeias real e prevista, considerando os menores erros percentuais encontrados

Conclusões

O objetivo da pesquisa era encontrar um erro aceitável entre o ângulo real e o previsto pelo modelo. A rede neural conseguiu aprender dentre a variabilidade de dados disponível e assim apresentar resultados satisfatórios. A rede apresentou um erro médio no conjunto de teste de 12.81%, sendo verificado posteriormente que o motivo desta diferença foi devido a uma degenerescência entre os ângulos de uma mesma tríade, o que levou ao erro encontrado.

Para continuidade da pesquisa, propõe-se a construção de uma lista de vizinhos da tríade em questão, a partir da definição de um raio de vizinhança para o carbono α . Isso permitirá o desenvolvimento de um algoritmo apto a calcular os valores angulares de maneira ainda mais eficiente, visto que neste algoritmo serão consideradas interações entre as tríades que modificam os seus ângulos, com o objetivo de eliminar a atual degenerescência encontrada. Também pode-se propor a investigação de novos dados de entrada que irão influenciar na geometria desta cadeia e valores angulares da conformação espacial, diedros.

Referências:

- [1]RESEARCH COLLABORATORY FOR STRUCTURAL BIOINFORMATICS. **An Information Portal to Biological Macromolecular Structures**. Disponível em: <<http://www.pdb.org/pdb/home/home.do>>. Acesso em: 02 de julho de 2010.
- [2]LODISH, H. et al. **Biologia celular e molecular**. 5.ed. Porto Alegre, RS: Artmed, 2005.
- [3]HAYKIN, S. **Redes Neurais - Princípios e prática**. Porto Alegre, RS: Bookman, 2001.
- [4]VELLASCO, M. **Slides do Curso Redes Neurais Artificiais I – Pontifícia Universidade Católica**. Rio de Janeiro, Brasil.
- [5]ROSSI, ALD, BRUNETTO M.A.O.C. **Métodos de Codificação de Proteínas para uso com Redes Neurais Artificiais**. Universidade Estadual de Londrina, Paraná, Brasil.
- [6]SOUTO, A. C. S. **Uso de Redes Neurais Artificiais na Simulação Monte Carlo Aplicado ao Problema de Dobramento de Proteínas**. Dissertação de Mestrado, 2006. Universidade do Vale do Rio dos Sinos, São Leopoldo, RS.