



Universidade Federal do Espírito Santo  
Programa de Pós-graduação em Biotecnologia



# Modelos Tridimensionais

---

PAOLA SOSSAI AGUIAR

RAQUEL AHNERT AGUIAR

PROF. DR. TEODIANO FREIRE BASTOS FILHO

# Introdução

---

- Rápido avanço na computação científica → Sucesso de estudos em diversas áreas.
- Predição de estruturas tridimensionais:
  - Atracamento proteína-ligante.
  - Desenho racional de fármacos.
  - Desenho de novas proteínas.
  - Elucidação de estruturas.

# Introdução

---

- Esses avanços → Bem estar da sociedade e impactos na área da saúde.
- **Desafio:**
  - Metodologias → ↑ previsibilidade e acurácia.

# Estrutura 3D de proteínas

- Função de proteína → Estrutura tridimensional.

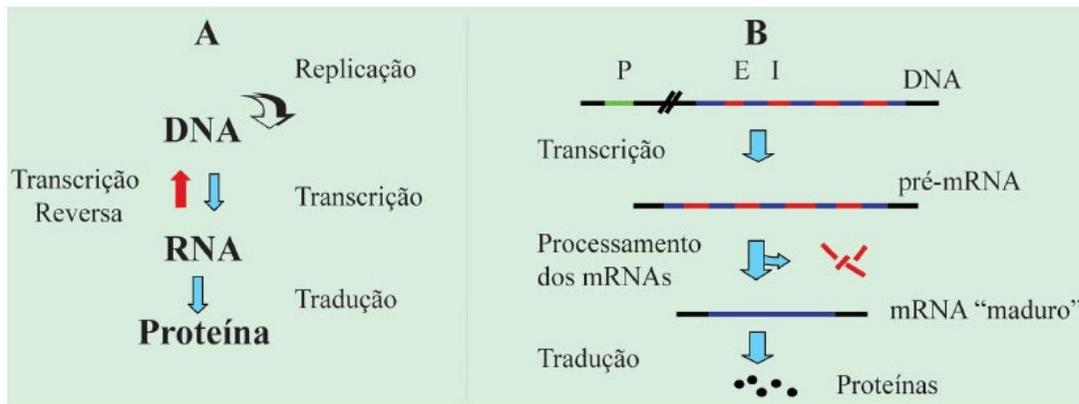


Figura 1 – Dogma central da biologia molecular

Estrutura Nativa

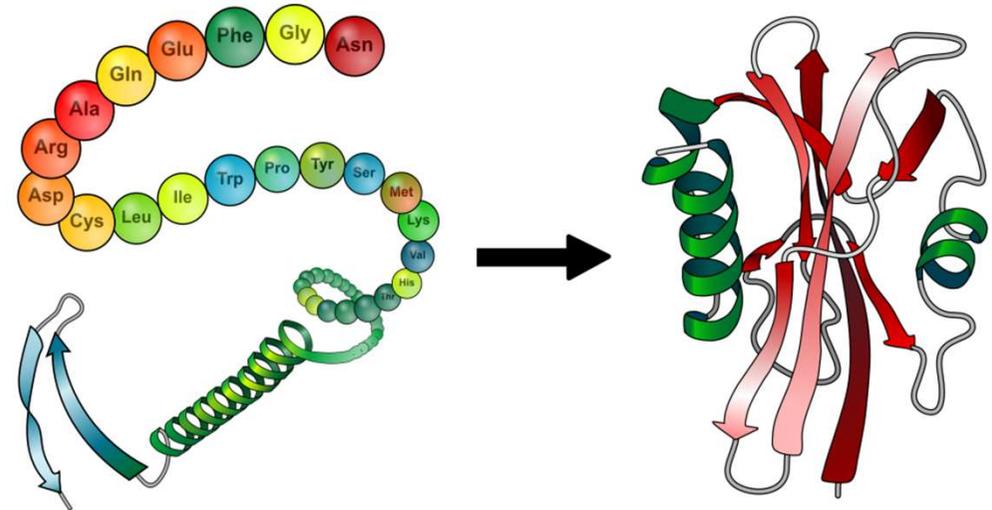


Figura 2 – Estrutura terciária de uma proteína

# Enovelamento de Proteínas

---

- Estudo de enovelamento x Estudo de predição de estruturas.
- Como a cadeia peptídica adota sua estrutura tridimensional?

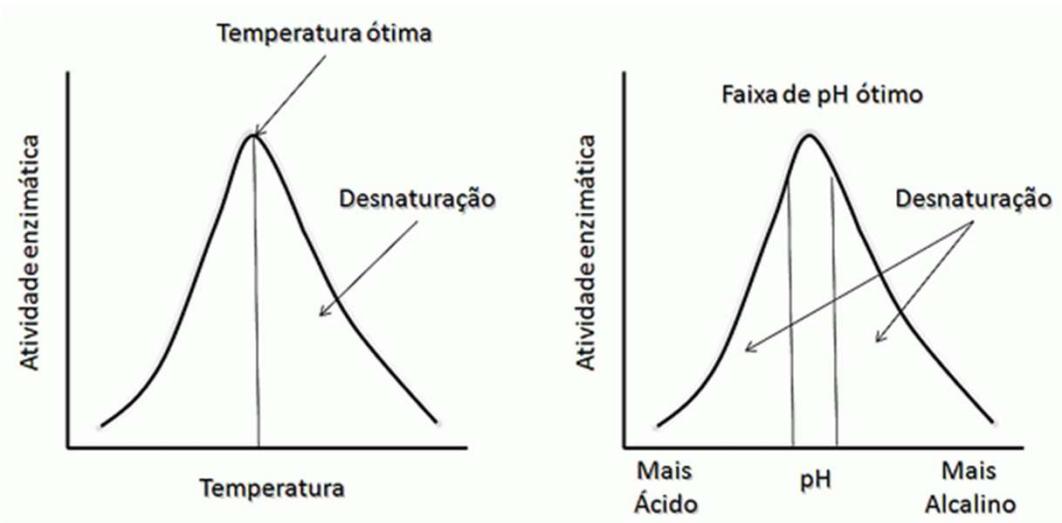


Figura 3 – Comportamento enzimático quanto a variação de pH e Temperatura

Postulado de Anfinsen e a Hipótese termodinâmica

- Única
- Estável
- Cineticamente acessível

Sequência dos aminoácidos determina a estrutura nativa

Enzima: Ribonuclease

B- mercaptoetanol

Uréia e Cloridrato de guanidina

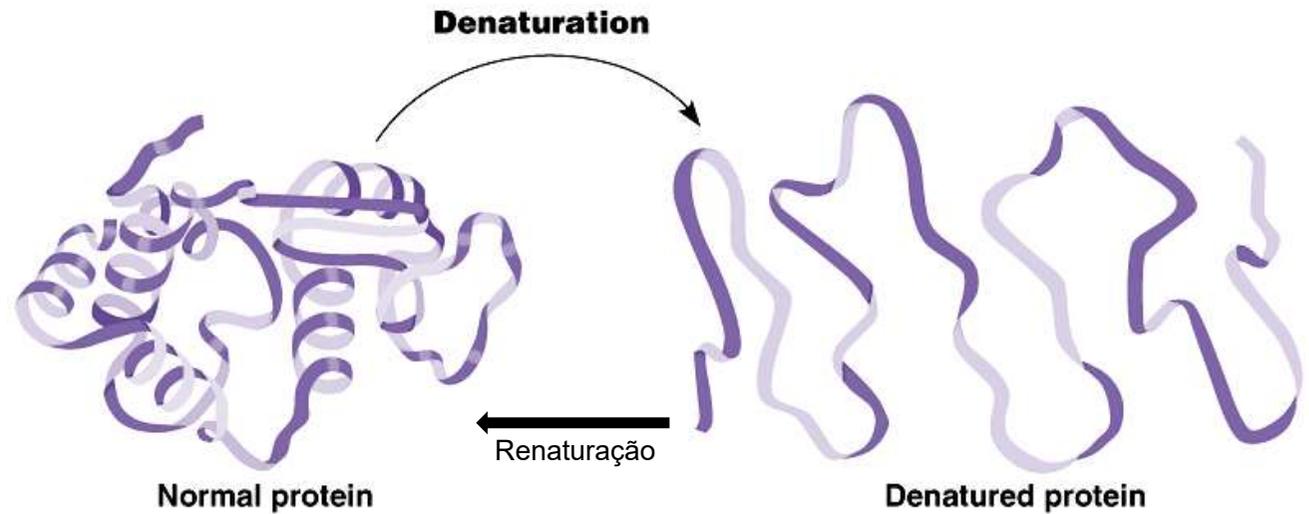


Figura 4 – Processo de desnaturação e renaturação proteica

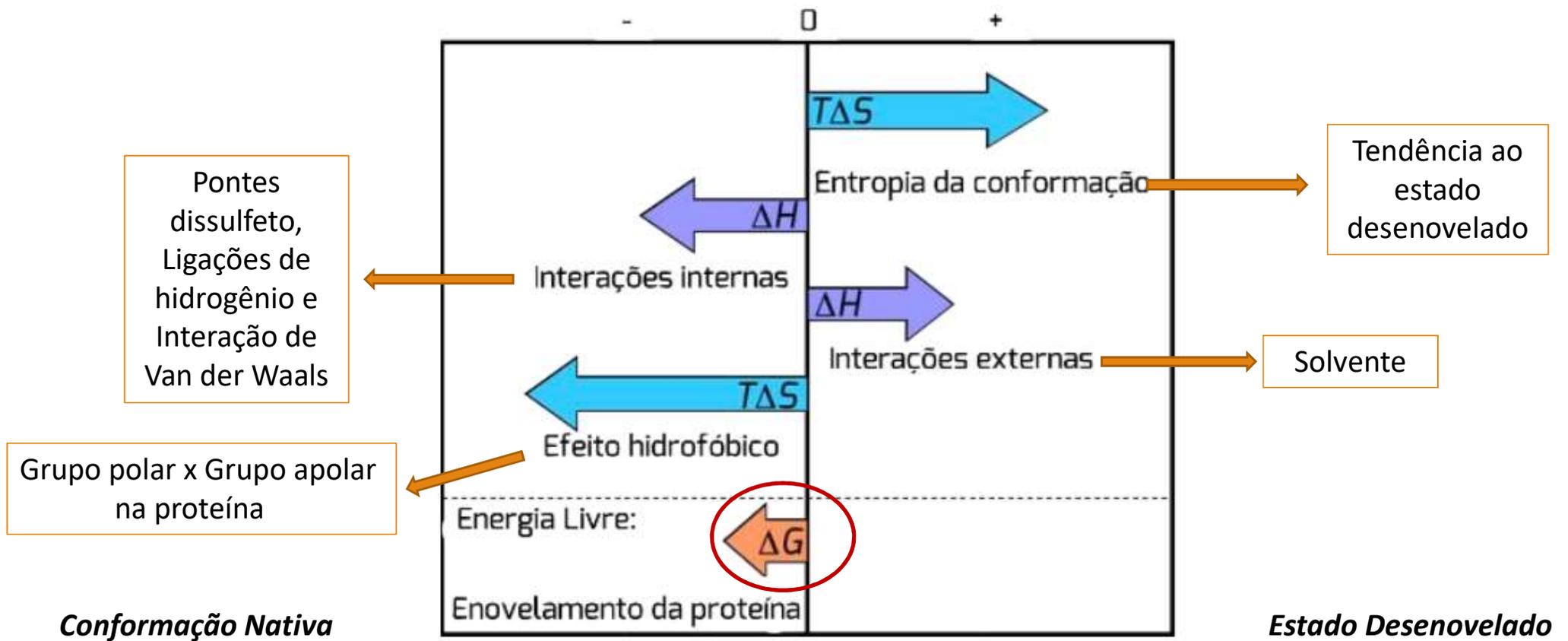


Figura 5 – Conjunto de forças que interferem na estrutura tridimensional de proteínas

# Predição da estrutura

- Determinação experimental -> Principal -> Alto custo -> Podem levar anos



## Métodos computacionais

- Paradoxo de Levinthal.
- Métodos de predição:

Dessa forma, os métodos de predição de estrutura de proteínas apresentam, nas suas metodologias, 3 características em comum.

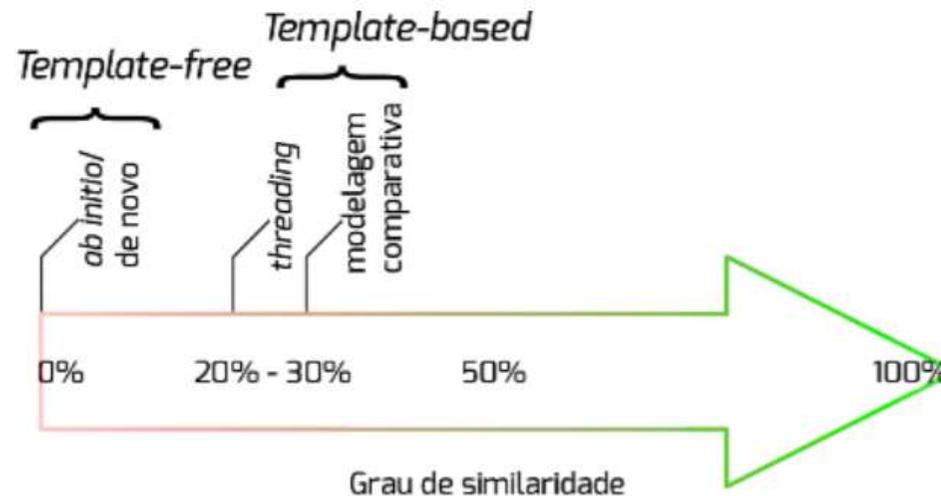


Figura 6 – Métodos de predição da estrutura tridimensional de proteínas

# Representação da estrutura

---

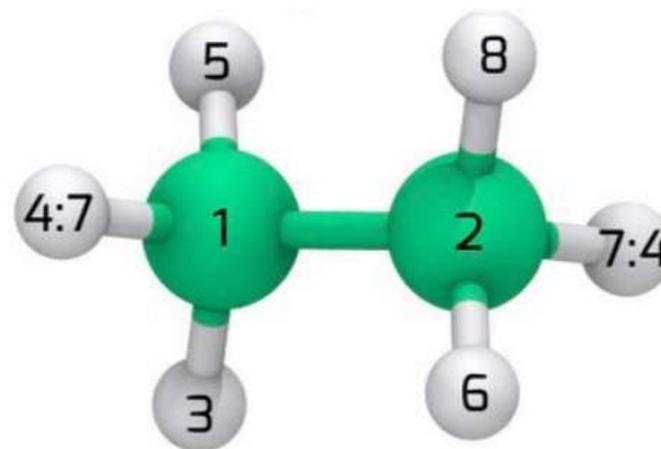
- A representação tridimensional de uma molécula pode ser dada pela posição geométrica de seus átomos em um sistema de coordenadas cartesianas (x, y, z) ou pelas chamadas coordenadas internas.
- A representação computacional pode ser feita:
  - Todos os átomos.
  - Átomos unidos.
  - Agrupamento de átomos.

### (A) Sistema de coordenadas cartesianas

| NOME   | NATM | ATM | RES | CAD | NRES | COORDX | COORDY | COORDZ | OCUP | BETA | ELEM |
|--------|------|-----|-----|-----|------|--------|--------|--------|------|------|------|
| HETATM | 1    | C1  | ETH | A   | 1    | 3.108  | 0.653  | -8.526 | 1.00 | 0.00 | C2H6 |
| HETATM | 2    | C2  | ETH | A   | 1    | 4.597  | 0.674  | -8.132 | 1.00 | 0.00 | C2H6 |
| HETATM | 3    | 1H1 | ETH | A   | 1    | 2.815  | -0.349 | -8.761 | 1.00 | 0.00 | C2H6 |
| HETATM | 4    | 2H1 | ETH | A   | 1    | 2.517  | 1.015  | -7.711 | 1.00 | 0.00 | C2H6 |
| HETATM | 5    | 3H1 | ETH | A   | 1    | 2.956  | 1.278  | -9.381 | 1.00 | 0.00 | C2H6 |
| HETATM | 6    | 1H2 | ETH | A   | 1    | 4.748  | 0.049  | -7.277 | 1.00 | 0.00 | C2H6 |
| HETATM | 7    | 2H2 | ETH | A   | 1    | 5.187  | 0.312  | -8.947 | 1.00 | 0.00 | C2H6 |
| HETATM | 8    | 3H2 | ETH | A   | 1    | 4.890  | 1.676  | -7.897 | 1.00 | 0.00 | C2H6 |

### (B) Sistema de coordenadas internas

| NATM | ATM | BOND | REF1 | ANG   | REF2 | TORC  | REF3 |
|------|-----|------|------|-------|------|-------|------|
| 1    | C   |      |      |       |      |       |      |
| 2    | C   | 1.54 | 1    |       |      |       |      |
| 3    | H   | 1.00 | 1    | 109.5 | 2    |       |      |
| 4    | H   | 1.00 | 2    | 109.5 | 1    | 180.0 | 3    |
| 5    | H   | 1.00 | 1    | 109.5 | 2    | 60.0  | 4    |
| 6    | H   | 1.00 | 2    | 109.5 | 1    | -60.0 | 5    |
| 7    | H   | 1.00 | 1    | 109.5 | 2    | 180.0 | 6    |
| 8    | H   | 1.00 | 2    | 109.5 | 1    | 60.0  | 7    |



Ângulo de torção = os graus de liberdade que irão definir o espaço de conformações.

Figura 7 – Representação tridimensional do etano

- Funções de energia:

- Critério de “qualidade” da estrutura final -> A estrutura deve ter a menor energia total livre.

- Algoritmo de Busca:

- O algoritmo de busca é o componente responsável por gerar a conformação inicial, avaliar sua qualidade usando a função de energia, gerar novas conformações e avaliá-las em um processo iterativo até que algum critério de parada esteja satisfeito.

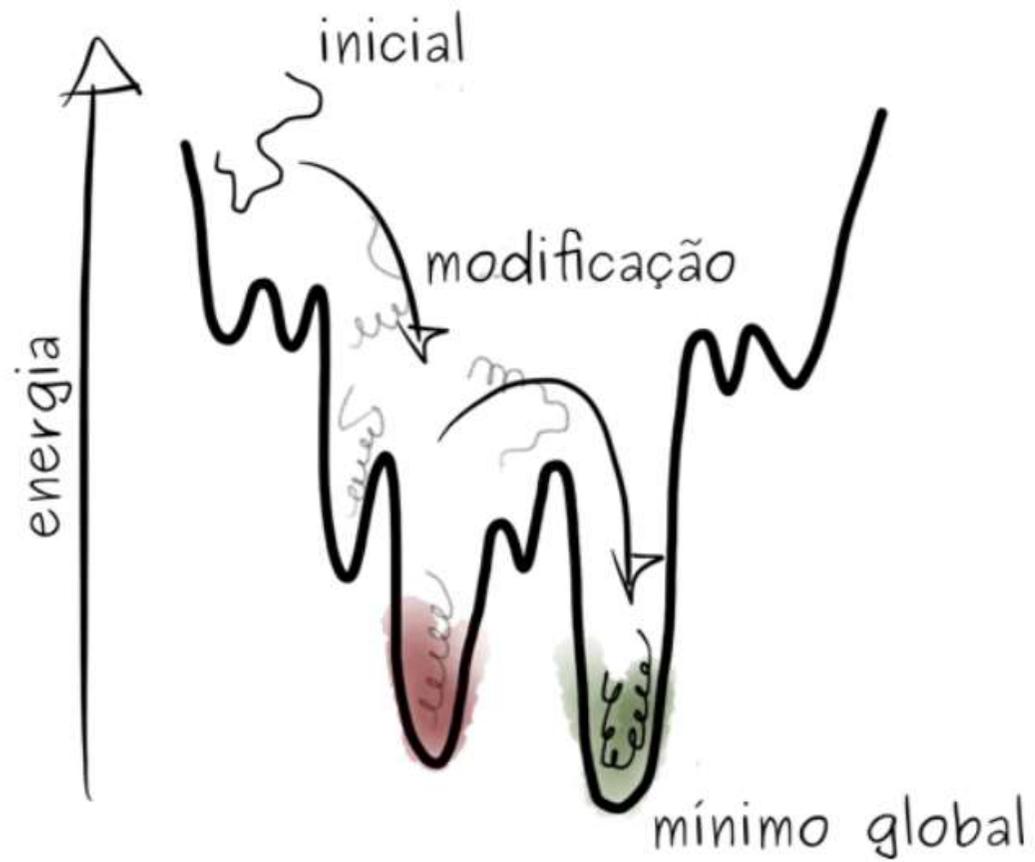
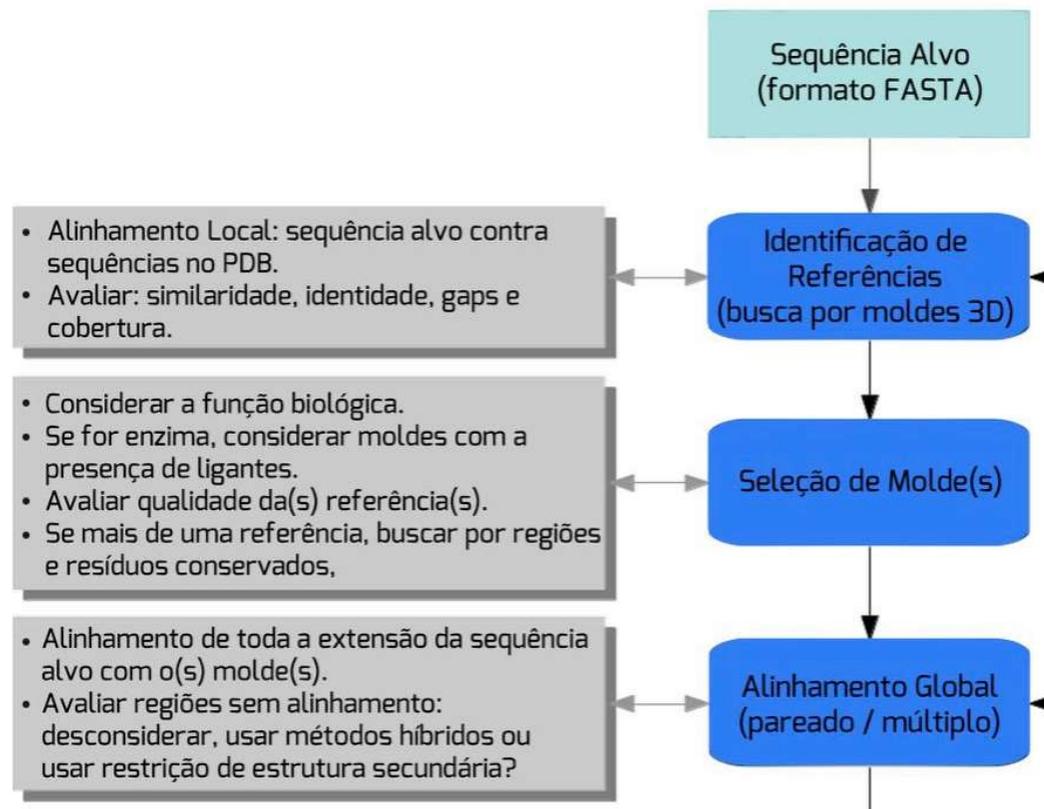


Figura 8 – Representação de uma busca usando metaheurística para predição de estrutura de proteína.

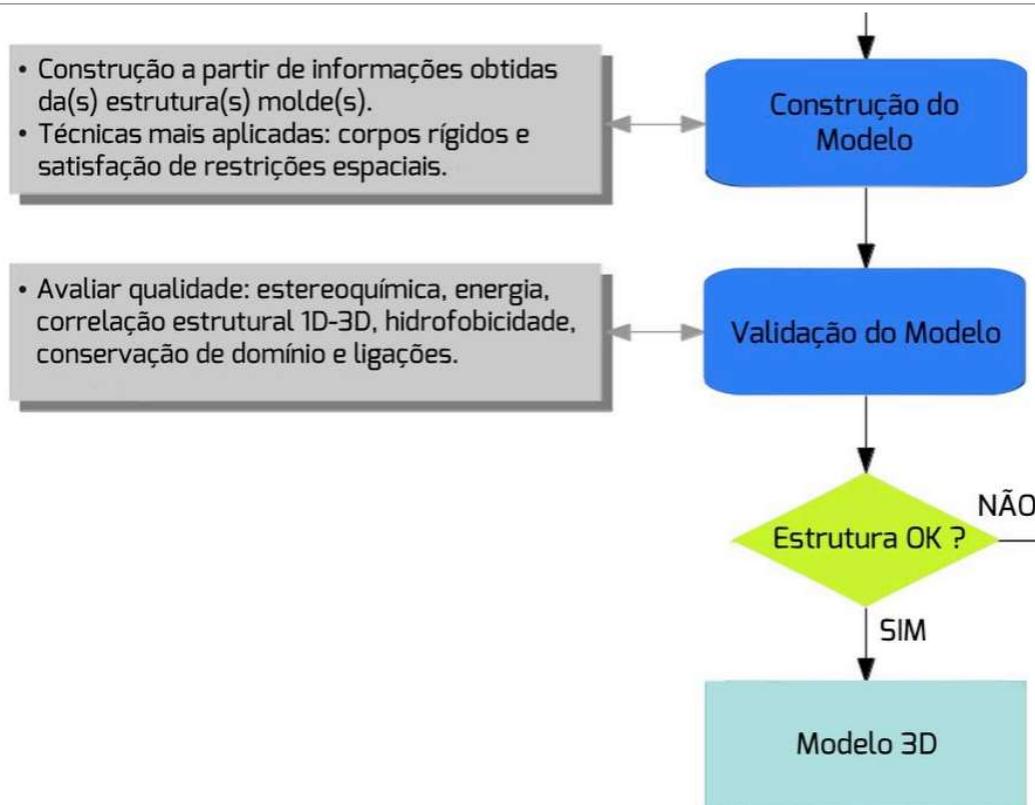


Figura 9 – Fluxograma para predição de estrutura tridimensional

# Modelagem Comparativa



# Modelagem Comparativa



# Predição do enovelamento

---

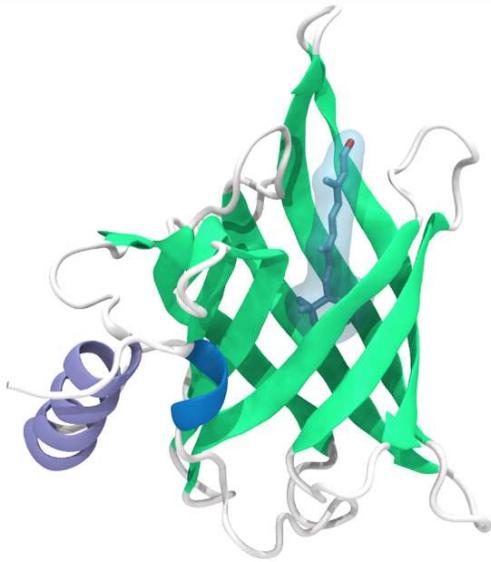
- Threading
- Estrutura 3D é mais conservada que a sequência (identidade entre 20% e 30% da estrutura primaria)
- PDB (Protein Data Bank)
  - CATH (Class, Architecture, Topology, Homology)
  - SCOP (Structural Classification of Proteins)

# Predição do enovelamento

---

- i) Reconhecimento do tipo de enovelamento: estrutura 2ária, polaridade de cadeias laterais e hidrofobicidade
- ii) Alinhamento entre a sequência de aminoácidos da proteína de interesse e bancos de dados: alinhamentos múltiplos de sequências de estruturas 2árias, matrizes de substituição para cada aminoácido dentro de uma família específica de proteínas e penalização de gaps
- iii) Escolha do(s) melhor(es) molde(s): funções de predição de erro/qualidade, bibliotecas de fragmentos
- iv) Construção do modelo 3D através de técnicas similares às empregadas na modelagem comparativa (Swiss-Model ou Modeller).  
Regiões sem molde: métodos por primeiros princípios (HH-Pred e I-TASSER)
- **Limitações:**
  - **identidade muito baixa, enovelamento da sequência não representado no banco (modelagem comparativa)**
  - **resolução relativamente baixa: posicionamento preciso dos átomos (atracamento)**

# Métodos *de novo*

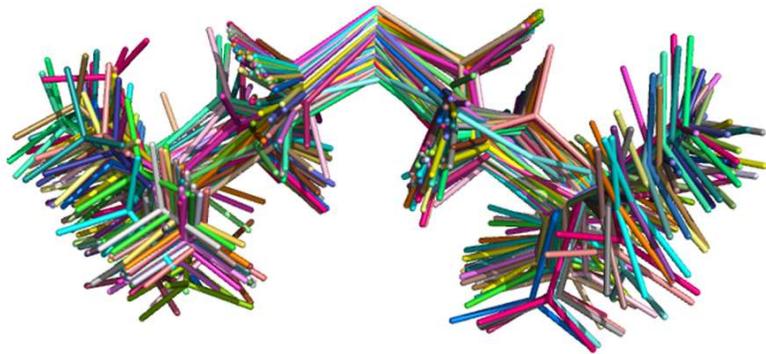


e-book Bioinformática  
www.ufrgs.br/bioinfo/

- **Predição de estruturas 2árias:** estrutura primária, análise de resíduos, interações de longo alcance, alinhamentos múltiplos, informação evolutiva; PSI-BLAST: PSIPRED, DSC, GOR IV, Predator, Prof, PROFphd e Sspro (80% de precisão).
- **Fragmentos de proteínas:** pequenos fragmentos com estruturas similares apresentavam similaridade de sequência, Ambiente físico-químico próprio: **mesma sequência, estruturas diferentes; duas sequências diferentes, mesma estrutura** (lista de candidatos).
- **Modificação da função de energia**

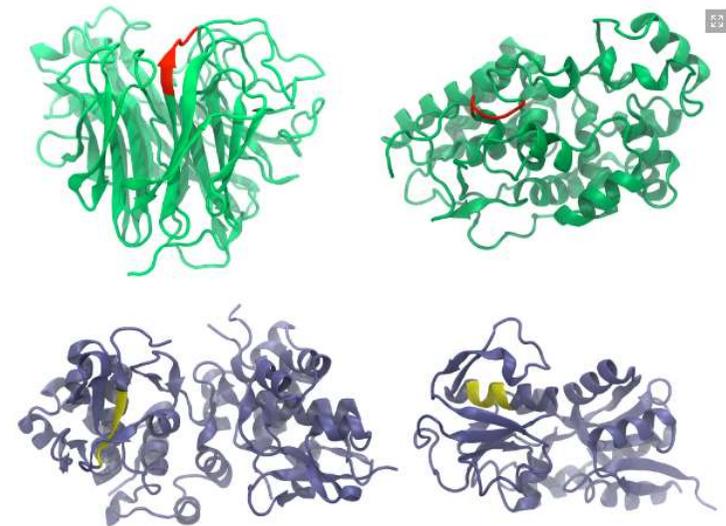
Figura 10: Retinol Binding Protein com o retinol no sítio ativo, código PDB: 1RBP.

# Fragmentos de proteínas:



*e-book* Bioinformática  
[www.ufrgs.br/bioinfo/](http://www.ufrgs.br/bioinfo/)

Figura 11: Fragmentos estruturalmente semelhantes, mas que possuem seqüências de resíduos diferentes.



*e-book* Bioinformática  
[www.ufrgs.br/bioinfo/](http://www.ufrgs.br/bioinfo/)

Figura 12: Fragmentos de proteínas com a mesma seqüência de resíduos que possuem estruturas diferentes. Acima, as proteínas de código PDB 1F8E (fragmento destacado entre os resíduos 243 e 247) e 1BGP (resíduos 63 a 67); abaixo, 1LM5 (2800 a 2804) e 1XS5 (121 a 125).

# Fragmentos de proteínas

- Protein Fragment Generator – Profrager: biblioteca de fragmento + Matriz BLOSUM62
- Limitações:
  - limitada a sequências mais curtas (<150 resíduos) (primeiros princípios)
  - artefatos se a parametrização das funções estatísticas não for feita com cuidado



e-book Bioinformática  
[www.ufrgs.br/bioinfo/](http://www.ufrgs.br/bioinfo/)

Figura 13: Geração de um fragmento de seis resíduos.

# Campos de força estatísticos

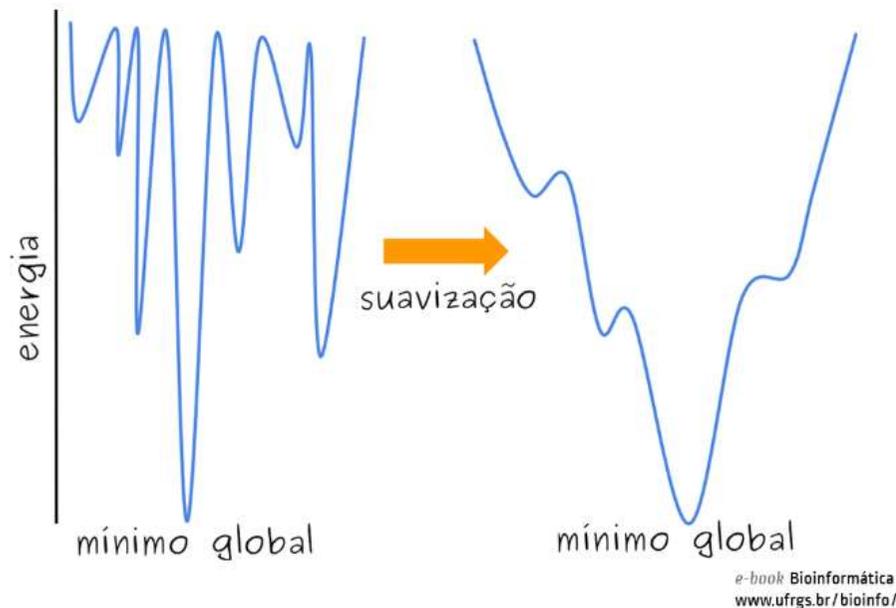


Figura 14: Efeito de suavização da superfície de energia.

- Probabilidade de um par de resíduos possuir uma ligação de hidrogênio.
- Altura de cada ponto representa a energia da conformação, (menor energia estarão no fundo, maior energia no topo)

Conjunto de proteínas teste



conformação de menor energia (ou mínimo global):

conformação nativa (mínimos locais pouco frequentes e com valores de energia distantes do mínimo global)

- **QUARK:** correlação entre a energia e a similaridade estrutural com a conformação nativa é de 0,7 ( 0 - 1)

# Primeiros princípios (*ab initio*)

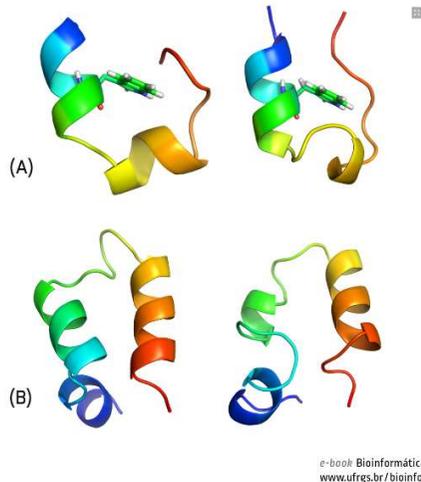


Figura 15: Resultados obtidos com o protocolo *ab initio* do programa GAPP. (A) Trp-cage (PDB1L2Y) com 29 aminoácidos. O modelo na esquerda apresenta um RMSD (do esqueleto peptídico) de 3,04 Å em relação à estrutura experimental na direita. (B) Villin headpiece (PDB1VII) com 36 aminoácidos. O modelo na esquerda apresenta um RMSD de 3,38 Å (do esqueleto peptídico) em relação à estrutura experimental na direita.

- Problema não resolvido na biologia computacional: defasada (velocidade e acurácia) e limitado a proteínas pequenas (menos de 100 resíduos de aminoácidos).
- Refinamento.
- **Não usa informações de estruturas conhecidas: funções de energia** contendo somente termos de significado físico.
- **Dois pressupostos:**
  - todas as informações necessárias sobre a estrutura de uma proteína estão contidas em sua **sequência** de aminoácidos
  - as proteínas enovelam-se para um estado nativo, que se encontra próximo ao mínimo global de **energia livre** (termodinamicamente mais estável).

# Primeiros princípios (*ab initio*)

| Método   | Algoritmo de busca                                 | Função de energia                      |
|----------|--|--|
| GAPF     | Algoritmo genético                                 | GROMOS96 e GAPF-CG                     |
| Profet   | Algoritmo evolucionário                            | OPLSAA, AMBER94, AMBER96, ECEPP e FLEX |
| ProtPred | Algoritmo evolucionário                            | CHARMM (v.27)                          |
| Nicosia  | Algoritmo evolucionário multiobjetivo              | CHARMM (v.27)                          |
| MEAMT    | Algoritmo evolucionário multiobjetivo multitabelas | CHARMM (v.27)                          |

Tabela 1: Exemplo de métodos de predição *ab initio* de estrutura de proteínas.

- Custo computacional:

- computação paralela em larga escala
- modelos de energia e representações simplificadas
- abordagens heurísticas (algoritmos genéticos)
- simulações por dinâmica molecular

- Campo de força empíricos (+ termo de solvatação implícita):

Captura **características físicas e químicas** que impulsionam e estabilizam o enovelamento descrevendo as **interações intramoleculares da proteína e desta com as moléculas de solvente**.

Obs: aproximações resultam em artefatos, favorece excessivamente hélices.

# Escolhendo o modelo

---

- Protocolos mais usados PSP: executar o algoritmo um grande número de vezes com diferentes sementes para o gerador de números aleatórios -> **Decoys**
- Dois principais passos automáticos de filtragem de decoys:
  - 1) **Filtragem dos decoys pelo valor de energia total.** Investigar apenas os decoys com energia semelhante (até certo valor limite) em relação ao decoy de menor energia.
  - 2) **Agrupamento (clustering) dos decoys:** critério de similaridade (ex: desvio médio até 3 Å). Investigar apenas a estrutura mais representativa de cada grupo (reduzir consideravelmente número de modelos). Aquele grupo que contém a estrutura nativa é, provavelmente o maior grupo (pois os algoritmos conseguem chegar a um nível de refinamento maior – heurísticas genéticas).
- **Inspeção manual** de cada modelo: operador humano (treinamento e alto custo)

# Análise de qualidade

---

- **Conjunto de fatores:** comprimentos de ligação, planaridade das ligações peptídicas, planaridade dos anéis e ângulos de torção nas cadeias principal (ou seja, esqueleto peptídico) e laterais, quiralidade, impedimento estérico, energia e funcional.
  - **Estereoquímica:** aspectos tridimensionais de uma molécula, a fim de se verificar a estabilidade conformacional da mesma. Regiões de tensão angular e torcional, impedimentos estéricos e quiralidades. Análise do gráfico de Ramachandran identifica quais resíduos encontram-se fora das regiões energeticamente favoráveis (Procheck e Molprobit).
  - **Energia:** funções de energia. A análise dos valores normalizados da função (Verify3D, ProSa, QMEAN e PROVE).
  - **Funcional:** envolve a comparação do modelo obtido com aspectos funcionais ou mesmo estruturais (sem resolução atômica) determinados por métodos experimentais. Famílias de proteínas possuem resíduos específicos associados à função. Métodos como dicroísmo circular, infravermelho e modelos tridimensionais RMN: estado conformacional da proteína em meio biológico para validar o modelo obtido.
- **Boa qualidade:** cálculo do desvio médio quadrático (RMSD) do modelo gerado x estrutura molde entre **1 Å e 2 Å**.

# Refinamento dos modelos

---

- **Local:** análise estereoquímica pode-se identificar qual resíduo está violando seus valores limites dentro de sua vizinhança, o que geralmente é resolvido com o reposicionamento de sua cadeia lateral. Em alguns casos, é necessário realizar etapas de otimização somente de regiões de alças, **principalmente de regiões ricas em glicina**. É sempre importante observar **violações causadas por prolinas** nas extremidades de regiões de estruturas em hélice ou folha.
- **Imposição de restrições:** refazer o modelo 3D impondo ao algoritmo restrições de tipo de estrutura 2ária para essas regiões.
- **Dinâmica molecular:** Simulações em solvente explícito ajudam a acomodar a estrutura 3D do modelo melhorando, principalmente, os ângulos  $\phi$  e  $\psi$  de resíduos em regiões desfavoráveis no gráfico de Ramachandran. Estruturas transmembranares (considerar o modelo de membrana a ser empregado e a forma de inserção do modelo 3D da proteína na membrana).

# Aplicações do modelo

---

**Acurácia:** cálculo do desvio médio quadrático (RMSD), que mede as distâncias interatômicas

- Estudos de predição funcional e busca por novos alvos moleculares em organismos patogênicos;
- Planejamento racional de fármacos baseado na estrutura do receptor biológico;
- Estudos de variação conformacional por dinâmica molecular;
- Planejamento de experimentos de mutagênese sítio-dirigida, fornecendo informações sobre possíveis mutações para testar hipóteses funcionais;
- Simulações de interações entre proteínas;
- Auxiliar no refinamento de estruturas resolvidas por cristalografia de raios-X e por experimentos de RMN.

# Referência das Figuras

---

- 1) Disponível em: <https://www.infoenem.com.br/o-dogma-central-da-genetica-e-o-rna/>
- 2) Disponível em: <http://zerpoi.opentronix.com/?p=953>
- 3) Disponível em: <http://maxaug.blogspot.com/2013/02/os-aminoacidos-as-proteinas-e-as.html>
- 4) Disponível em: <https://nutrisaude14.files.wordpress.com/2014/09/aula-02-protec3adnas-2014.pdf>
- 5) Disponível em: <http://cbeb2020.org/Bioinformatica/LivroBioinformatica.pdf> (pág. 152)
- 6) Disponível em: <http://cbeb2020.org/Bioinformatica/LivroBioinformatica.pdf> (pág. 152)
- 7) Disponível em: <http://cbeb2020.org/Bioinformatica/LivroBioinformatica.pdf> (pág. 155)
- 8) Disponível em: <http://cbeb2020.org/Bioinformatica/LivroBioinformatica.pdf> (pág. 157)
- 9) Disponível em: [https://www.ufrgs.br/bioinfo/th\\_gallery/capitulo-7](https://www.ufrgs.br/bioinfo/th_gallery/capitulo-7)

# Referências Bibliográficas

---

1. Bioinformática (da Biologia à Flexibilidade Molecular). Hugo Verli (org), 2014. Disponível em: <https://www.ufrgs.br/bioinfo/ebook/>