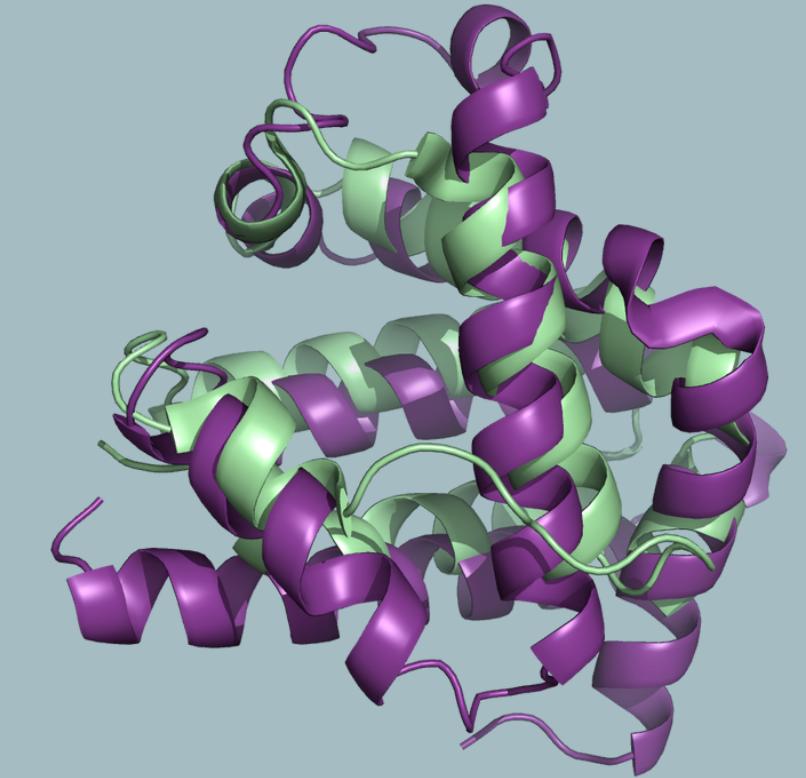
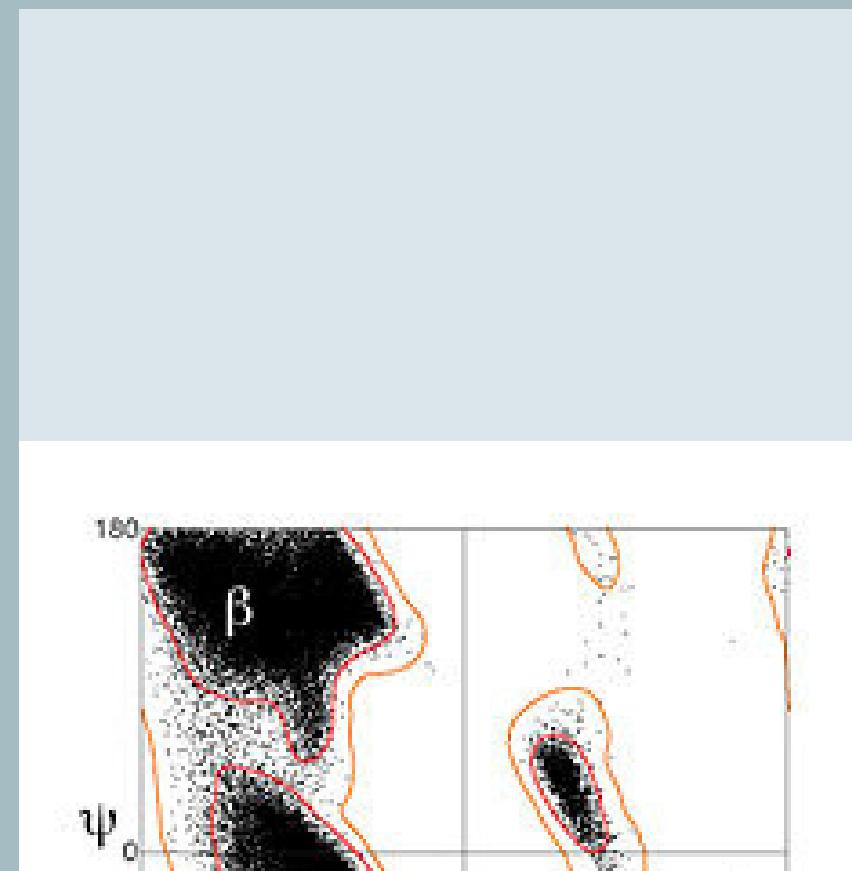
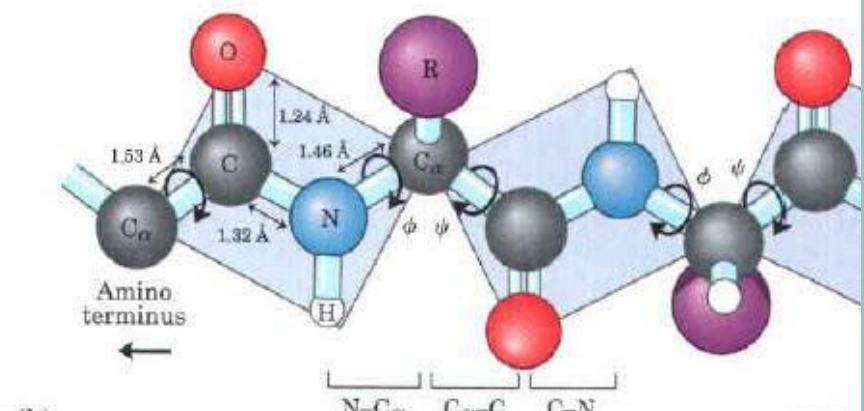
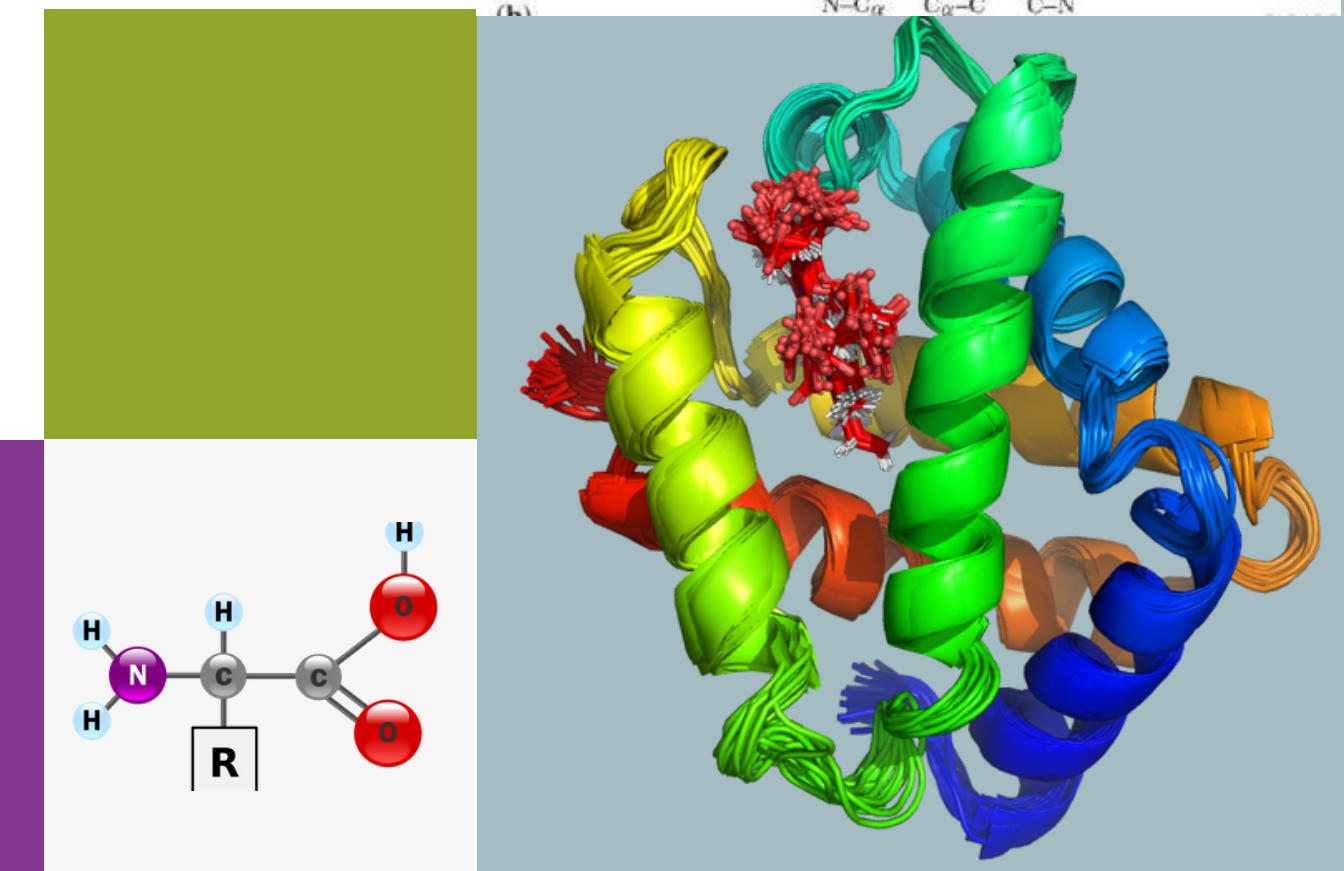
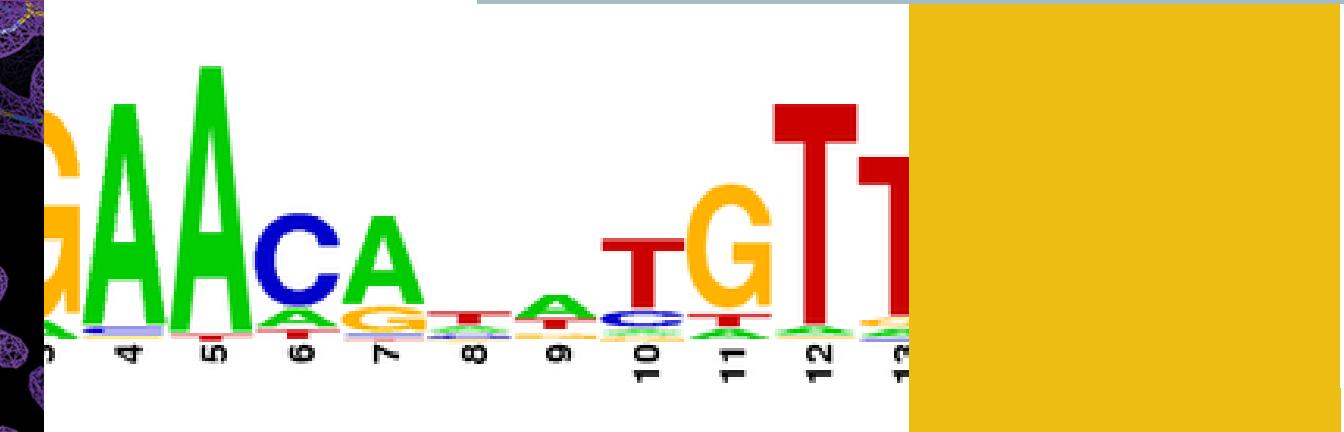
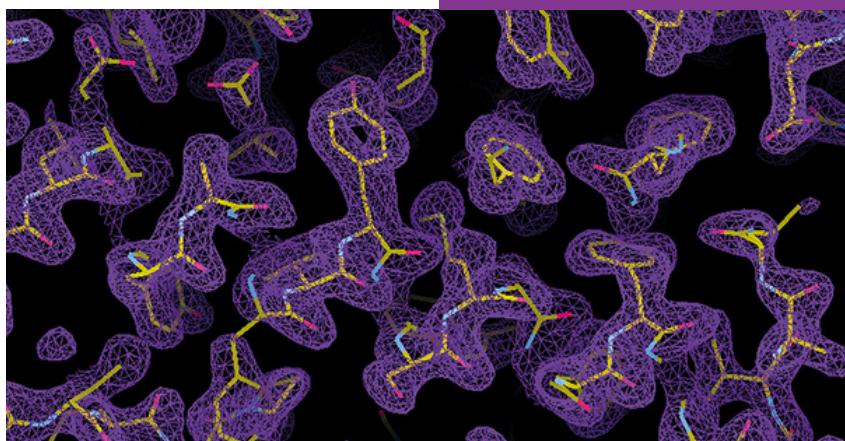
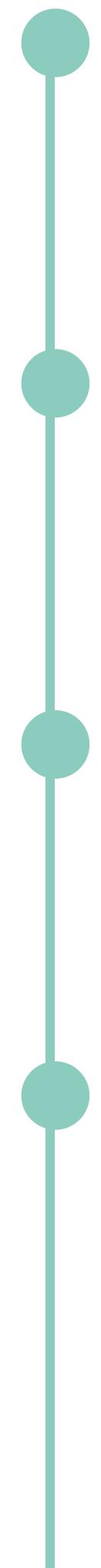


Modelagem molecular



Organização estrutural das proteínas

Esquema hierárquico



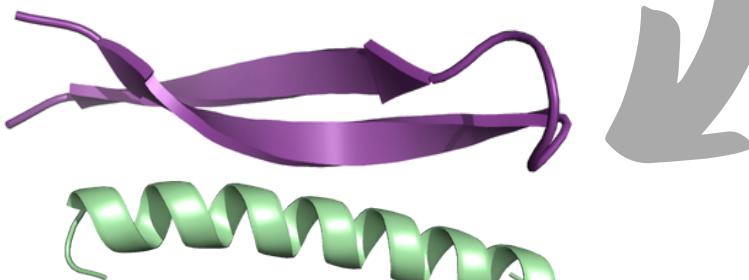
ESTRUTURA PRIMÁRIA

Sequência linear de resíduos de aminoácidos



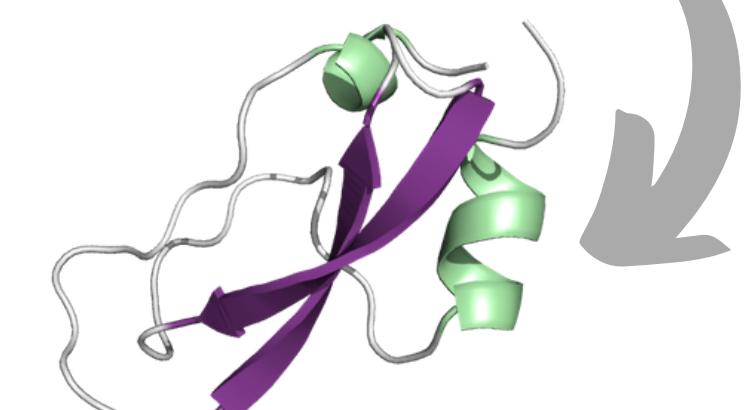
ESTRUTURA SECUNDÁRIA

Motivos formados principalmente graças a ligações de hidrogênio: alfa hélices e folhas beta



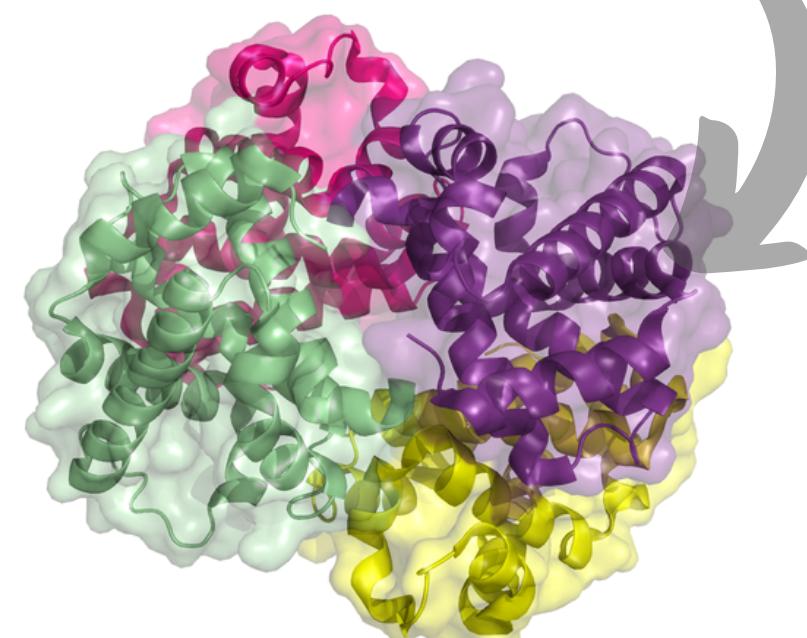
ESTRUTURA TERCIÁRIA

Também chamado enovelamento (*fold*) tridimensional da proteína nativa (funcional)



ESTRUTURA QUATERNÁRIA

Organização da proteína em complexos formados por duas ou mais cadeias de proteínas



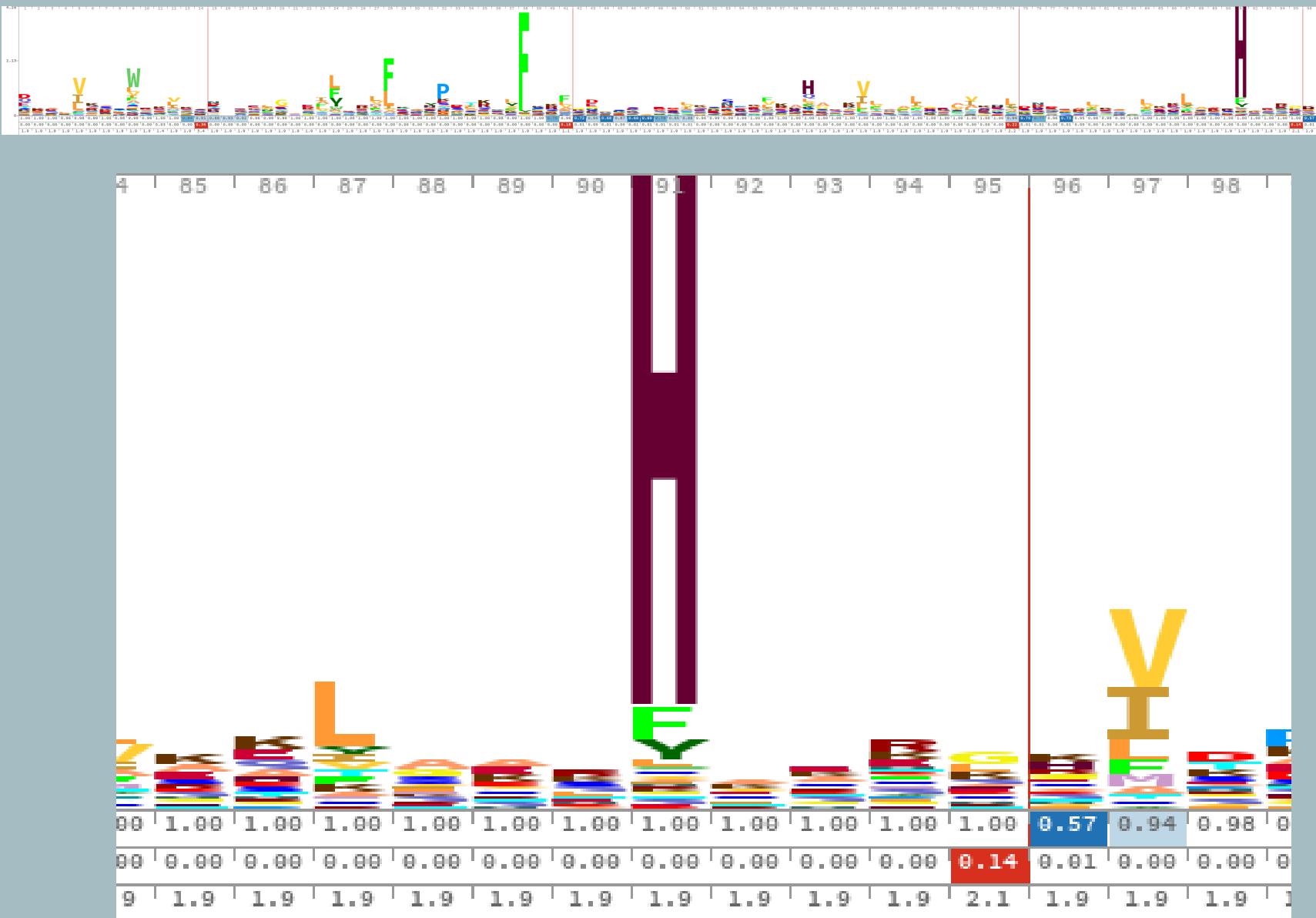
A **sequência** de uma proteína determina sua **estrutura tridimensional** que, por sua vez, determina sua **função**

Sequência → Estrutura → Função

Conservação

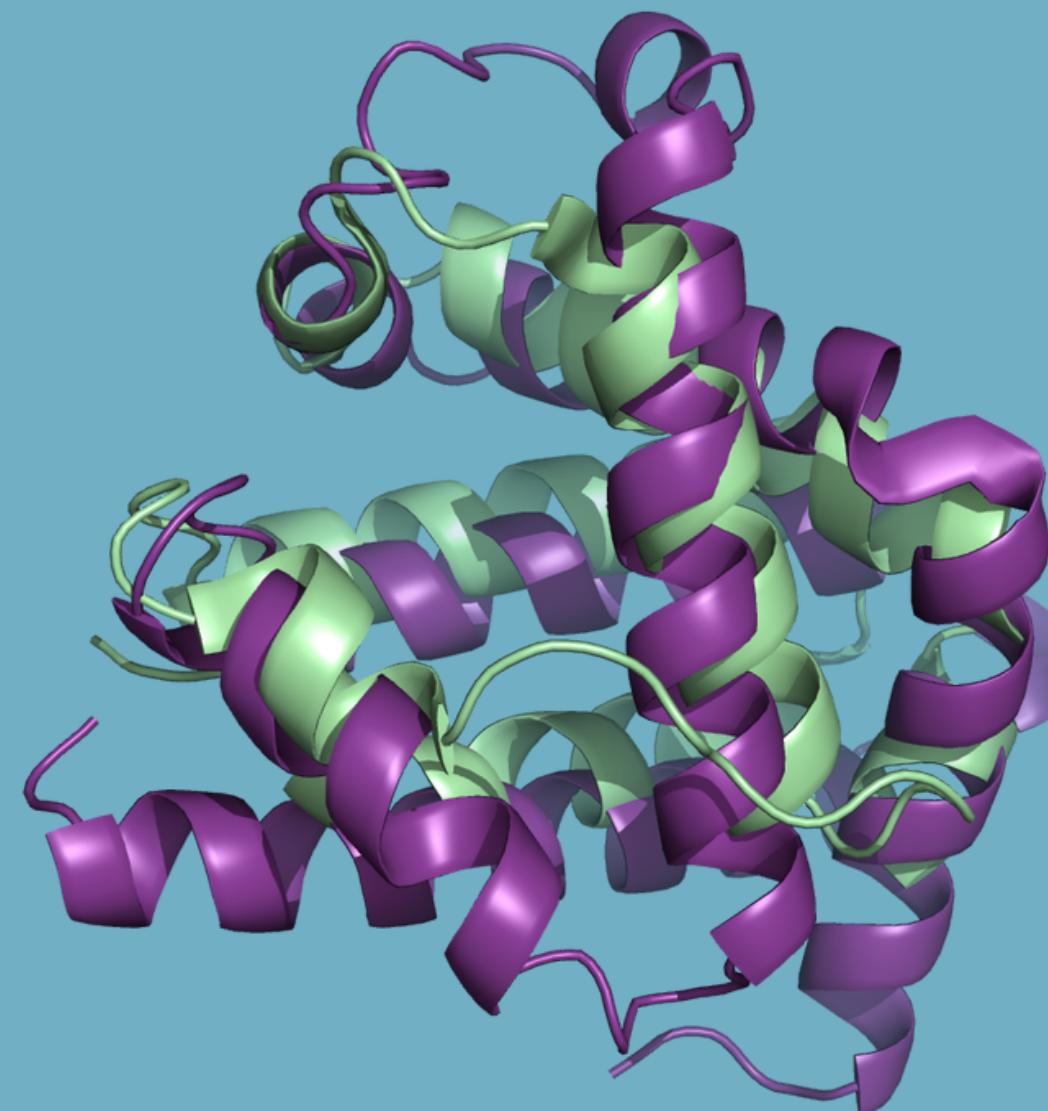
Sequências

Podem ser **bastante variáveis**



Estruturas e Funções

São muito conservados



Homologia, similaridade, identidade...



Identidade

Percentual de resíduos
de aminoácidos
conservados (idênticos)

Calculado por um
alinhamento de
sequências

Similaridade

Percentual de
resíduos e
aminoácidos
considerados
similares
(critério?)

Homologia

Relacionamento
evolutivo

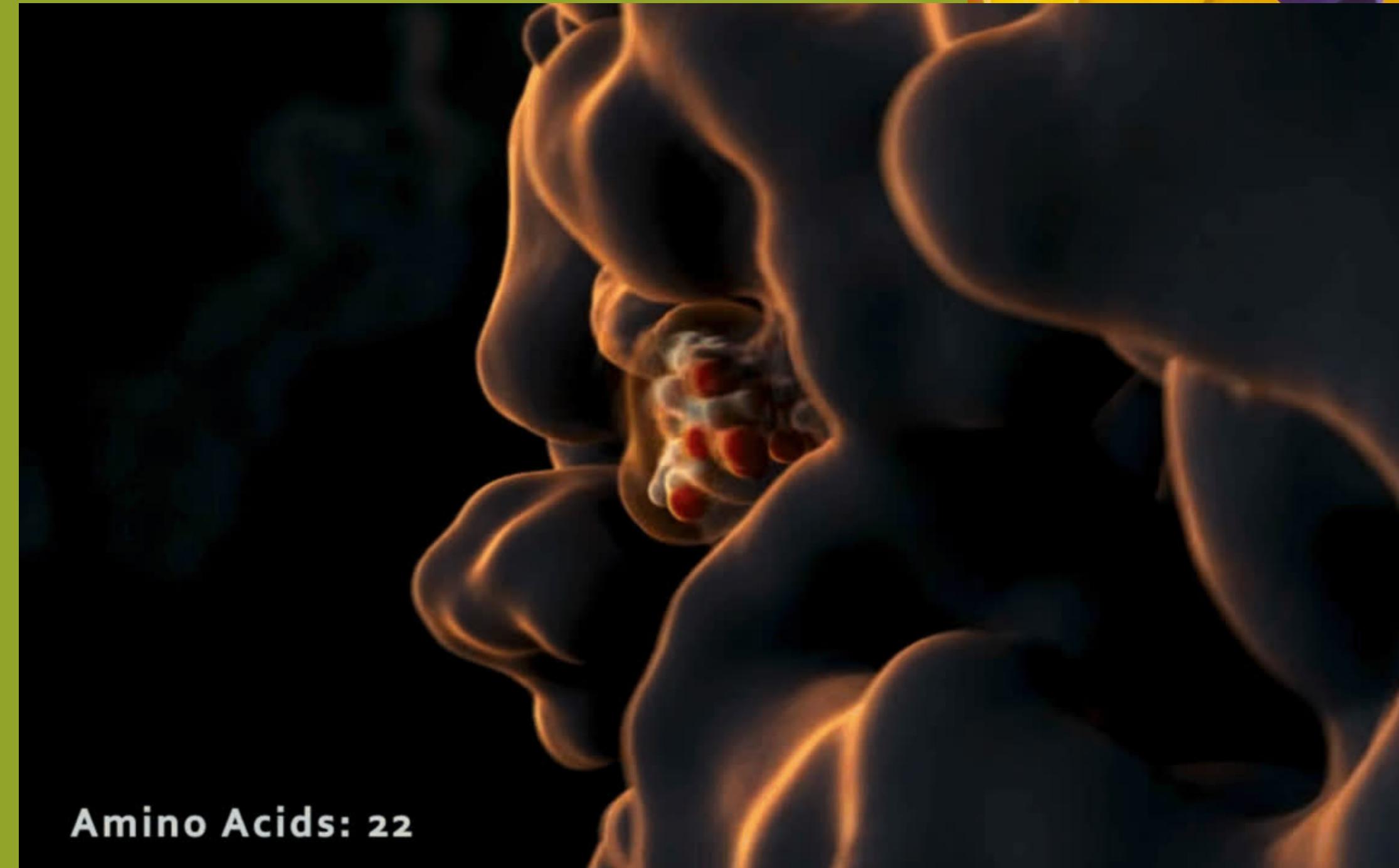
A homologia
existe ou não

Sequências

- 200 milhões de sequências de proteínas conhecidas
 - custo de US\$1.000 para se sequenciar um genoma completo

Estruturas

O processo de enovelamento



Modelagem molecular

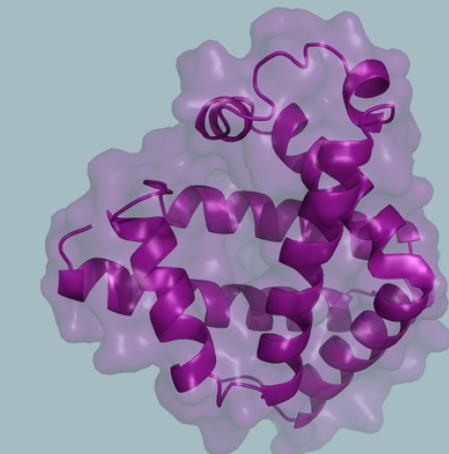
Modelagem por homologia
ou
Modelagem comparativa

Construção de **modelos estruturais atômicos** de uma **sequência alvo** com base em uma **estrutura tridimensional resolvida experimentalmente** (**molde** ou *template*)

Entradas

TWRGAAGVVSFFDAVACHRDHTYYYIAIP

1. sequência alvo

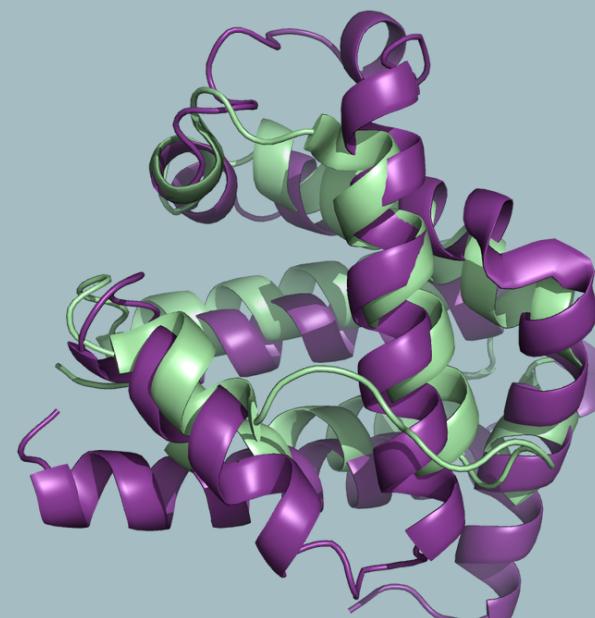


2. estrutura molde (*template*)



Saída

modelo estrutural



Etapas

A modelagem por homologia normamente segue as seguintes etapas:

Seleção dos moldes

Encontrar pelo menos uma estrutura de proteína resolvida experimentalmente que sirva de molde

Construção de alinhamento

Alinhamento múltiplo de sequências com a melhor qualidade possível

- uso de informação estrutural quando possível
 - estruturas secundárias

Modelagem das coordenadas e refinamento

- Definir as coordenadas de cada átomo
- Há vários paradigmas
- Os refinamentos visam melhorar posições desfavoráveis e consideram
 - Características geométricas
 - Cálculos de energia

Avaliação

- Comparação da estrutura com dados estatísticos de estruturas conhecidas
- Cálculos de energia
- Análise dos resíduos expostos e enterrados

Seleção de modelos

Visa identificar modelos com alta similaridade e cobertura

Alinhamentos multiplos de sequências

- Necessidade de se identificar **homólogos remotos**
- Compromisso entre **sensibilidade** e **eficiência**
 - psi-BLAST: *position specific scoring matrix*

Reconhecimento de perfil

- Há diversos métodos que visam inferir um perfil
 - HMM
 - *Fold-recognition*

Estruturas secundárias

Métodos que comparam as estruturas secundárias preditas com as das estruturas resolvidas experimentalmente para escolha dos modelos

Construção de alinhamentos

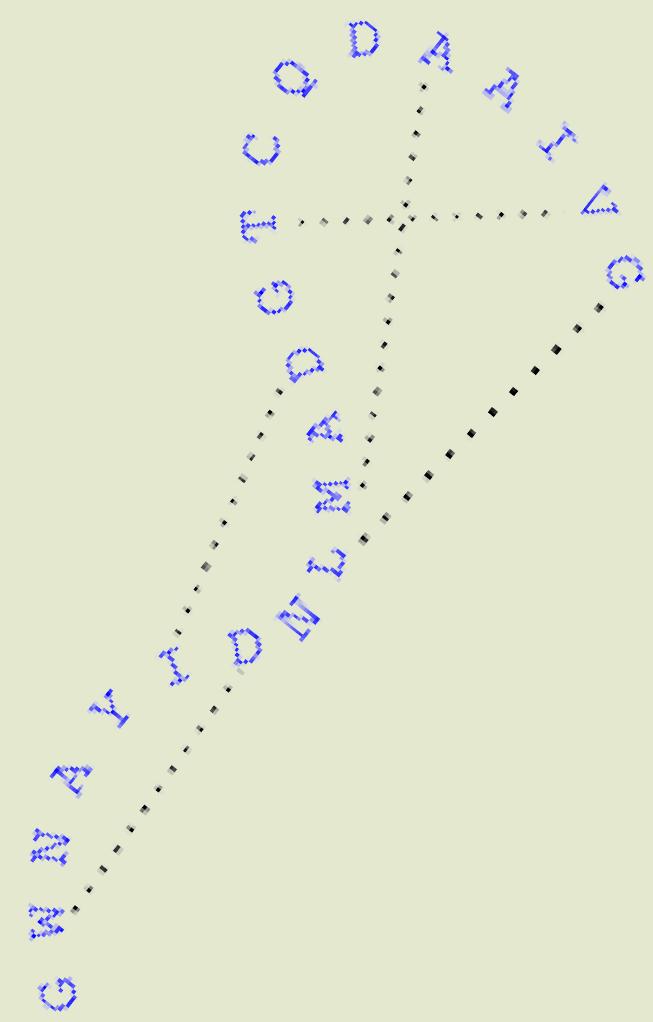
Normalmente é feita na seleção dos moldes
mas pode ser refinado

- A etapa de seleção de modelos realiza o alinhamento mas em **maior escala** e **menor detalhe**
 - O alinhamento pode ser **refinado** com menos **moldes escolhidos**, especialmente tendo em vista as **regiões de maior dissimilaridade**
 - Estruturas secundárias
 - outros padrões

Modelagem das coordenadas

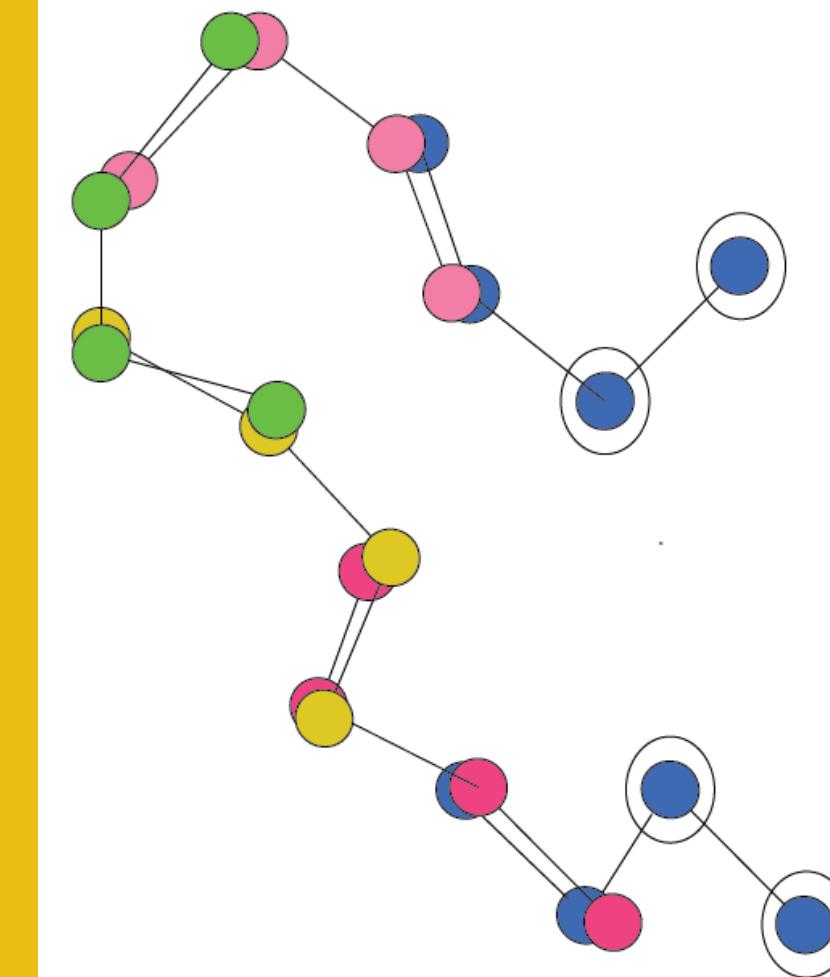
Restrições espaciais

São criadas **restrições** baseadas em **critérios geométricos** e as coordenadas são definidas com base em um **algoritmo de otimização**



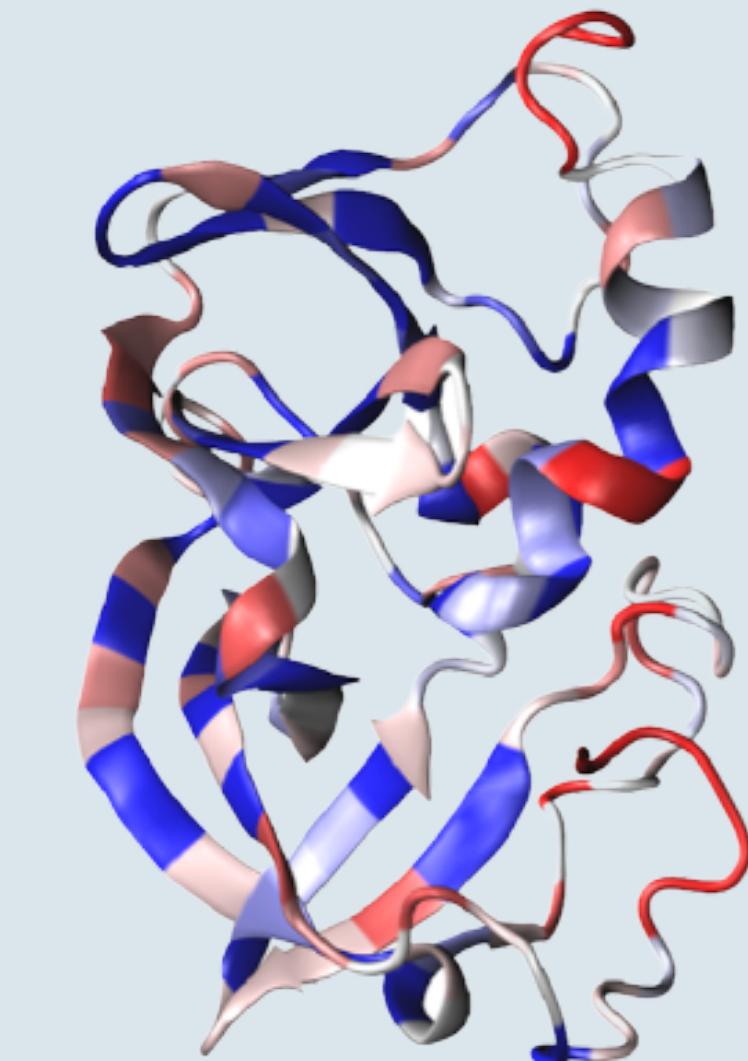
Fragments

- Usa-se modelos mais próximos para construir um **rascunho das posições mais conservadas (core)**
- **Bibliotecas de fragmentos** são usadas para os trechos de baixa identidade



Segments

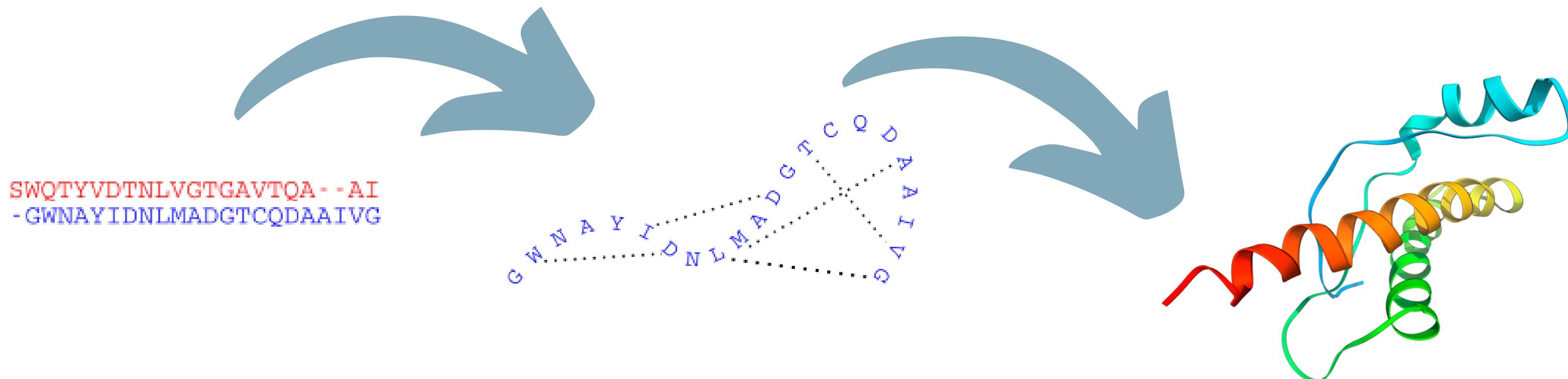
A proteína é segmentada em pequenos trechos que são modelados conforme **diferentes modelos de estruturas** resolvidas experimentalmente



Modelagem comparativa através de restrições espaciais

Modeller

O método mais utilizado
para modelagem molecular

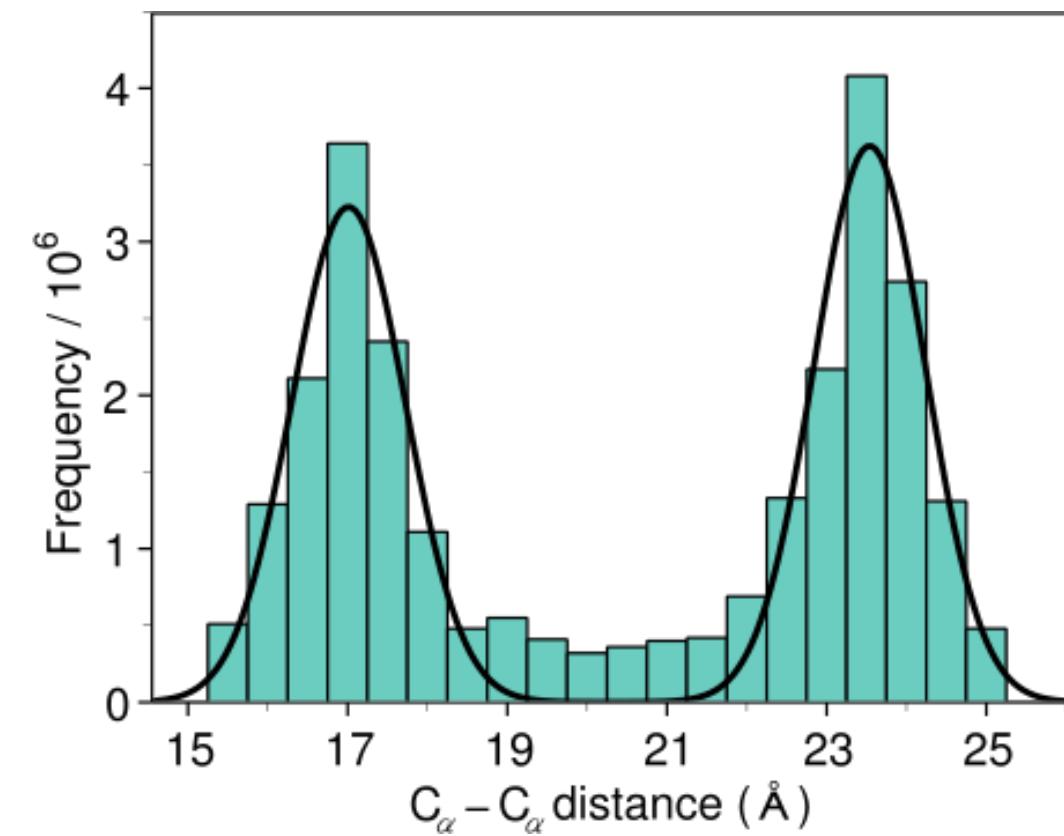


Šali, Andrej, and Tom L. Blundell. "Comparative protein modelling by satisfaction of spatial restraints." Journal of molecular biology 234.3 (1993): 779-815.



Andrej Šali, 1963

Modelagem comparativa através de restrições espaciais

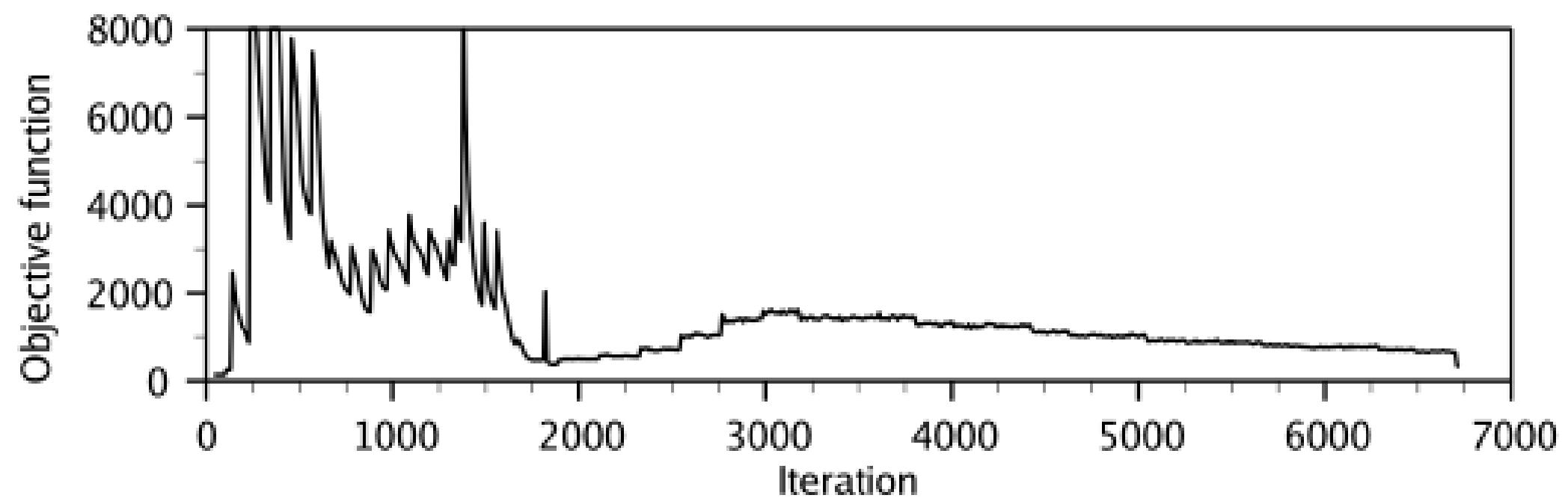


As restrições espaciais são baseadas em:

- Distâncias entre carbonos alfa
- Ângulos diedros

e são calculadas usando estruturas experimentais homólogas

Modeller



Uma função objetivo é calculada baseada em:

- restrições espaciais
- funções de energia do campo de força CHARMM

e são então minimizadas para obter as coordenadas x, y e z dos átomos através de um método numérico como o do gradiente conjugado

Swiss-Model

É um servidor web que disponibiliza:

- Ferramentas para modelagem molecular por homologia
 - Até 2016, usava o algoritmo do Modeller
 - Após 2016, passou a usar um método próprio e híbrido (ProMod3)
- Base de modelos prontos

ferramentas e base de dados

The screenshot shows the SWISS-MODEL website. At the top, there's a header with the logo of the Biozentrum University of Basel, the text "SWISS-MODEL", and links for "Modelling", "Repository", "Tools", "Documentation", "Log In", and "Create Account". Below the header, a blue banner contains the text "SWISS-MODEL" and a brief description: "is a fully automated protein structure homology-modelling server, accessible via the [ExPasy web server](#), or from the program DeepView (Swiss Pdb-Viewer). The purpose of this server is to make protein modelling accessible to all life science researchers worldwide." A "Start Modelling" button is located in this banner. To the right of the banner is a "Repository" section with a search bar and a grid of small protein structure thumbnails. Below the banner is a large image of the SARS-CoV-2 virus. To the right of the image, the text "SARS-CoV-2" is displayed, along with a description: "Severe acute respiratory syndrome coronavirus 2, is a positive-sense, single-stranded RNA coronavirus. It is a contagious virus that causes coronavirus disease 2019 (COVID-19)." It also mentions that the full SARS-CoV-2 proteome was modeled based on the NCBI reference sequence NC_045512 and annotations from UniProt, with results available [here](#). A "Variant of Concern 202012/01" section is also present.

ProMod3

Construção do modelo

Construção de um **modelo inicial** com base nos **moldes** encontrados

Modelagem de *loops*

Usa métodos de modelagem de *loops* para o **fechamento das lacunas** onde os moldes não puderam ser utilizados

Reconstrução das cadeias laterais

Refinamento nos posicionamentos das cadeias laterais com base em **bibliotecas de rotâmeros**

Minimização de energia

Uso de campo de força de **mecânica molecular** para refinamento final do modelo

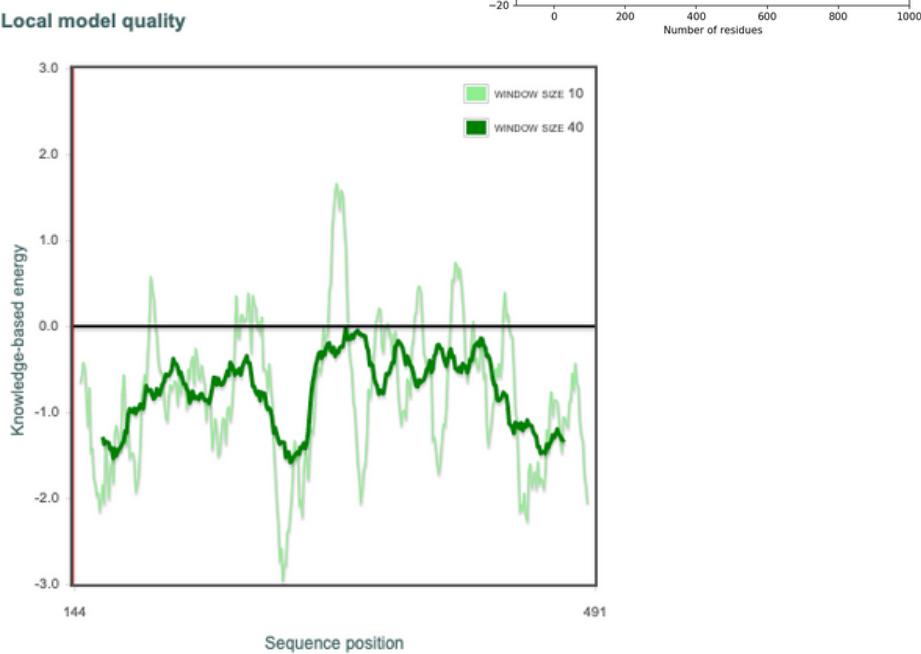
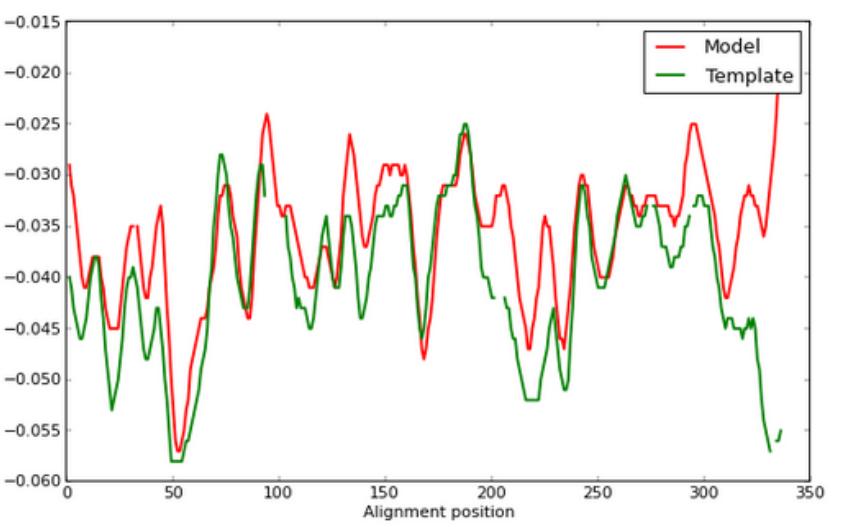
Avaliação

DOPE, 2006

Discrete optimized protein energy

Um **potencial estatístico** dependente da **distância atômica** definido **empiricamente** a partir de estruturas de proteínas nativas conhecidas

Quanto menor, melhor ^a qualidade do modelo



Prosa, 2007

Protein Structure Analysis

Calcula a **energia da estrutura** usando um potencial de **pares atônicos** baseado na distância e em um potencial que captura a **exposição ao solvente de resíduos** de proteína

Quanto menor o z-score, melhor a qualidade do modelo

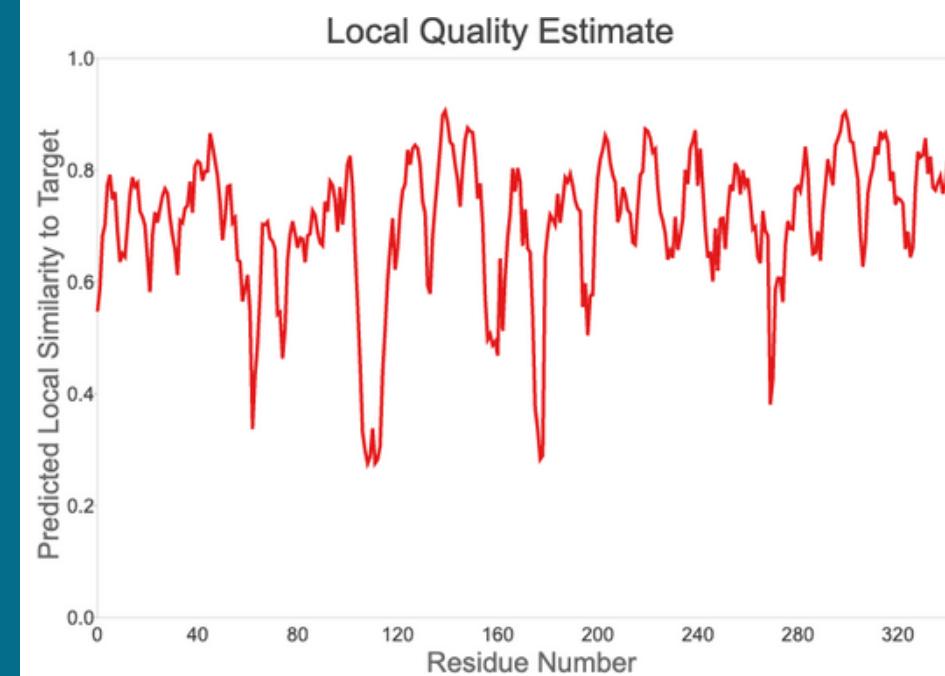
QMEAN, 2011

Qualitative Model Energy ANalysis

Combinação linear de quatro termos

- ângulos torsionais
- estruturas secundárias preditas x observadas
- potencial de solvatação e resíduos enterrados
- interações de longa-distância

Varia no intervalo [0,1], **sendo 1 bom**

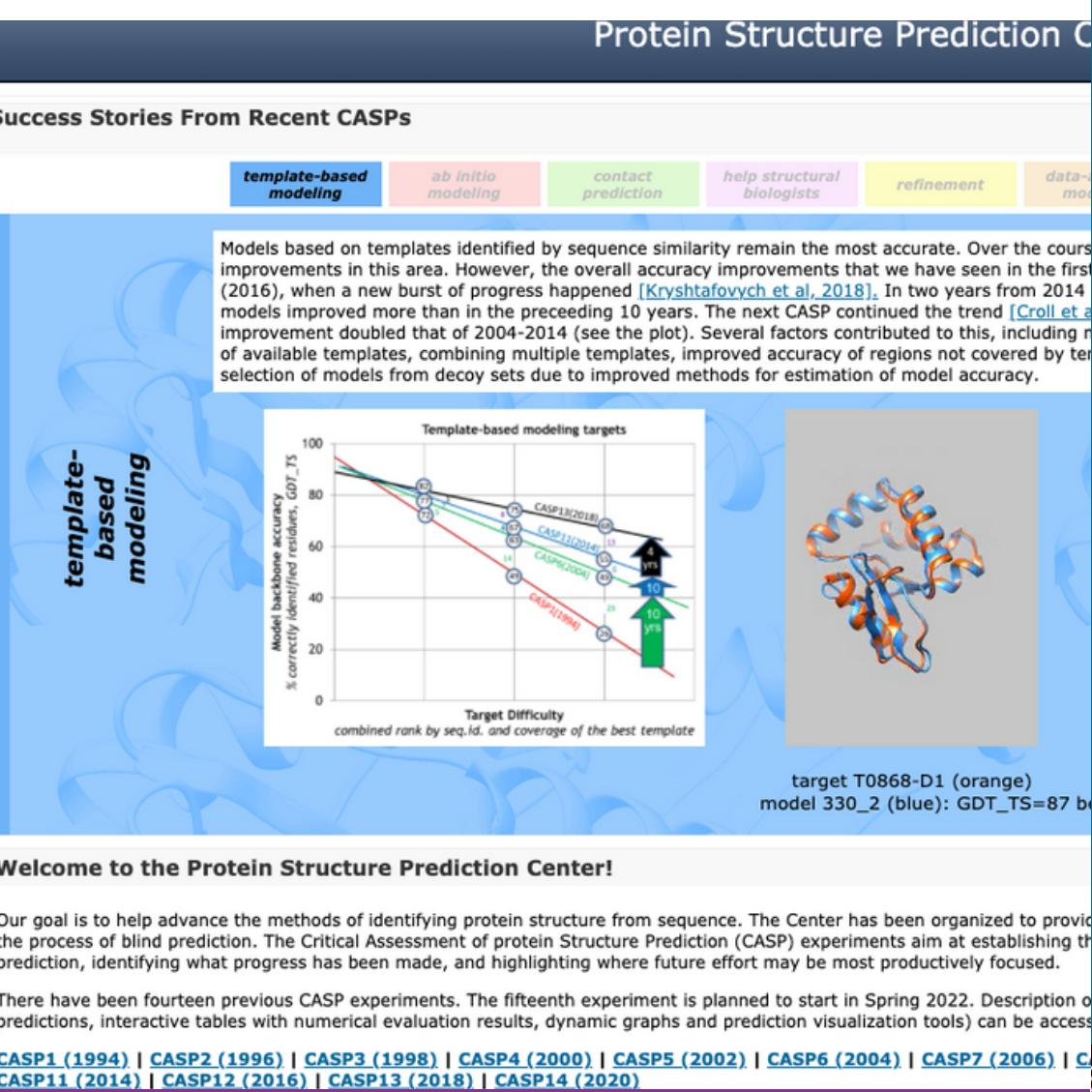


Muito úteis pois fomentam não apenas a avaliação dos métodos do estado-da-arte como ajudam a **melhor compreender** as abordagens mais promissoes para resolução de um determinado problema

Benchmarks e Competições

*Continuously evaluate the accuracy
and reliability of predictions*

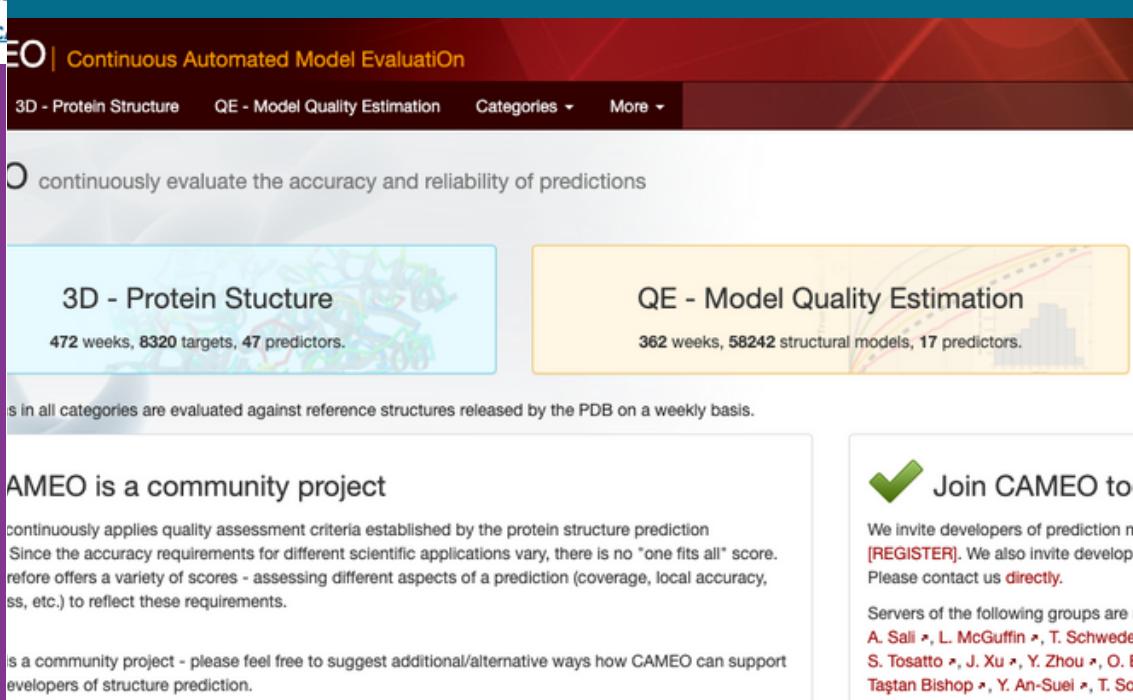
É uma **iniciativa contínua** de avaliação dos modelos teóricos e sua qualidade, de complexos proteicos, entre outros



CASP

*Critical Assessment fo Structure
Prediction*

É uma **competição anual** entre métodos para modelagem por homologia, predição *ab initio*, predição de contatos, entre outras tarefas correlatas



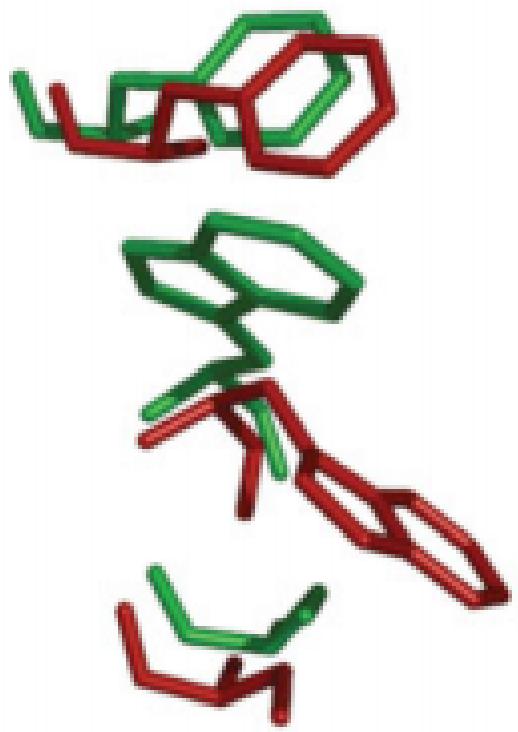
É uma **iniciativa contínua** de avaliação dos modelos teóricos e sua qualidade, de complexos proteicos, entre outros

Erros mais comuns

Maiores limitações

Fonte: Webb, Benjamin, and Andrej Sali.
"Comparative protein structure modeling using
MODELLER." Current protocols in bioinformatics
54.1 (2016): 5-6.

A



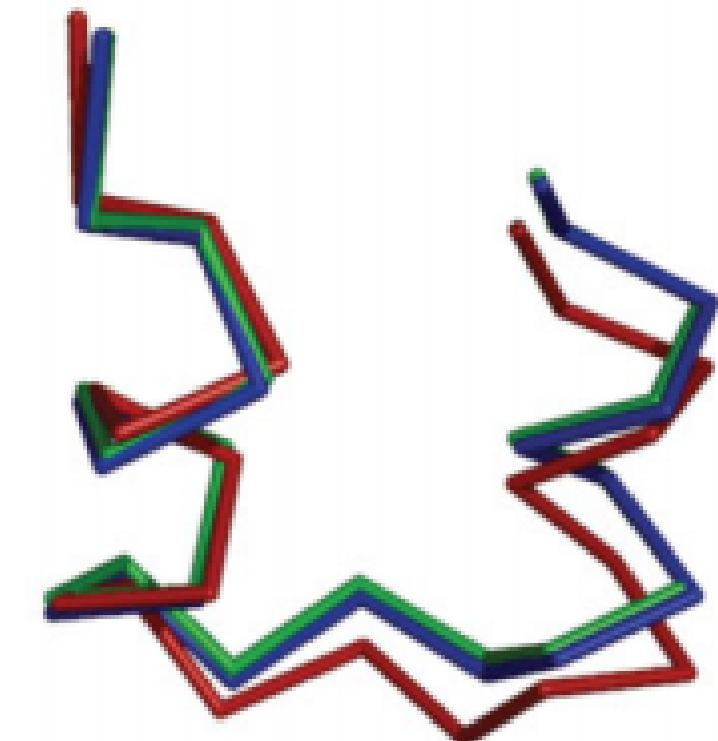
Distorções

Podem acontecer
mesmo em regiões
bem alinhadas

Cadeias laterais

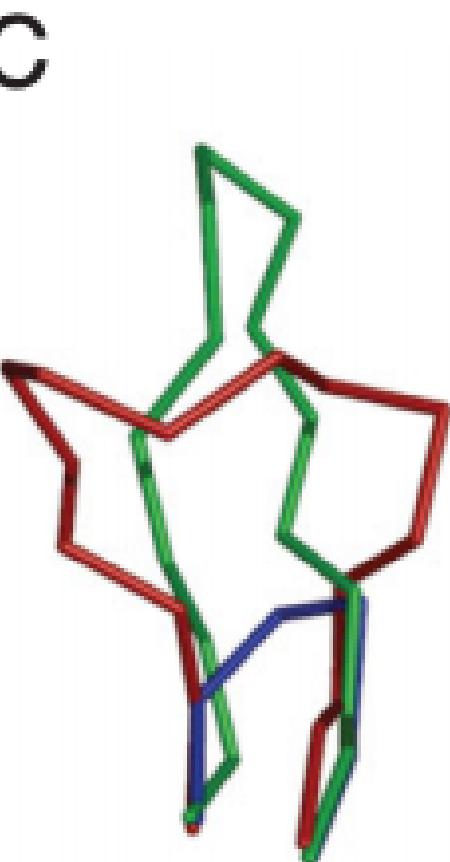
O empacotamento
das cadeias laterais,
visto que o alvo poder
apresentar resíduos
bem diferentes do
molde

B



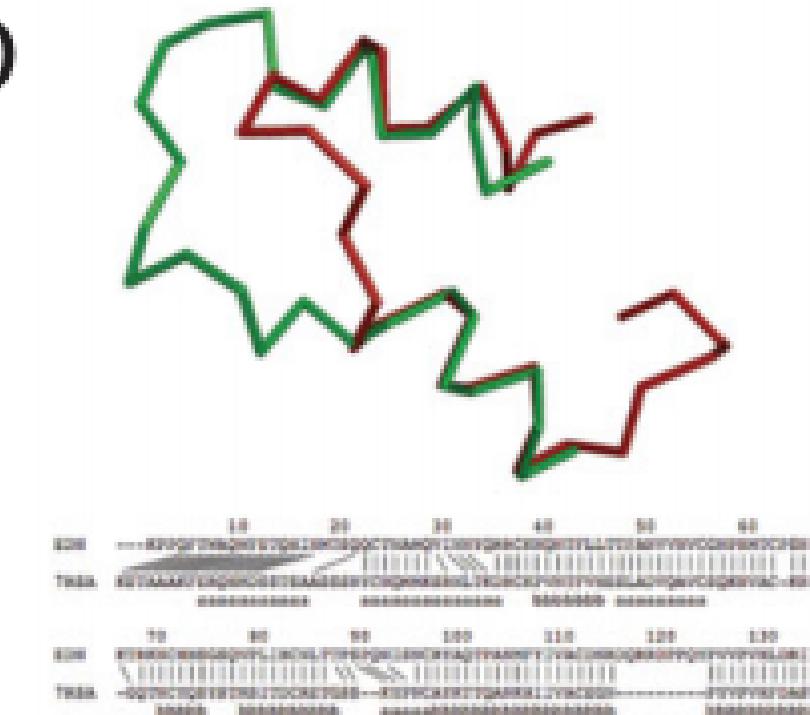
Ausência de molde

Regiões de *gaps*, ou seja, onde o molde não alinha muito bem



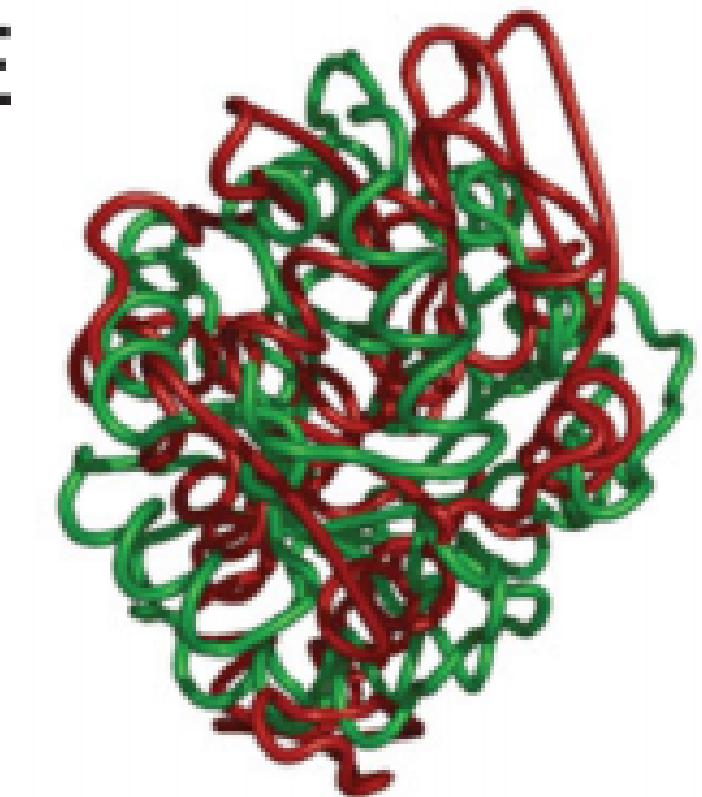
Regiões mal alinhadas

O alinhamento incorreto certamente levará a modelos de qualidade duvidosa



Molde inadequado

O uso de moldes inadequados levará a criação de modelos inúteis

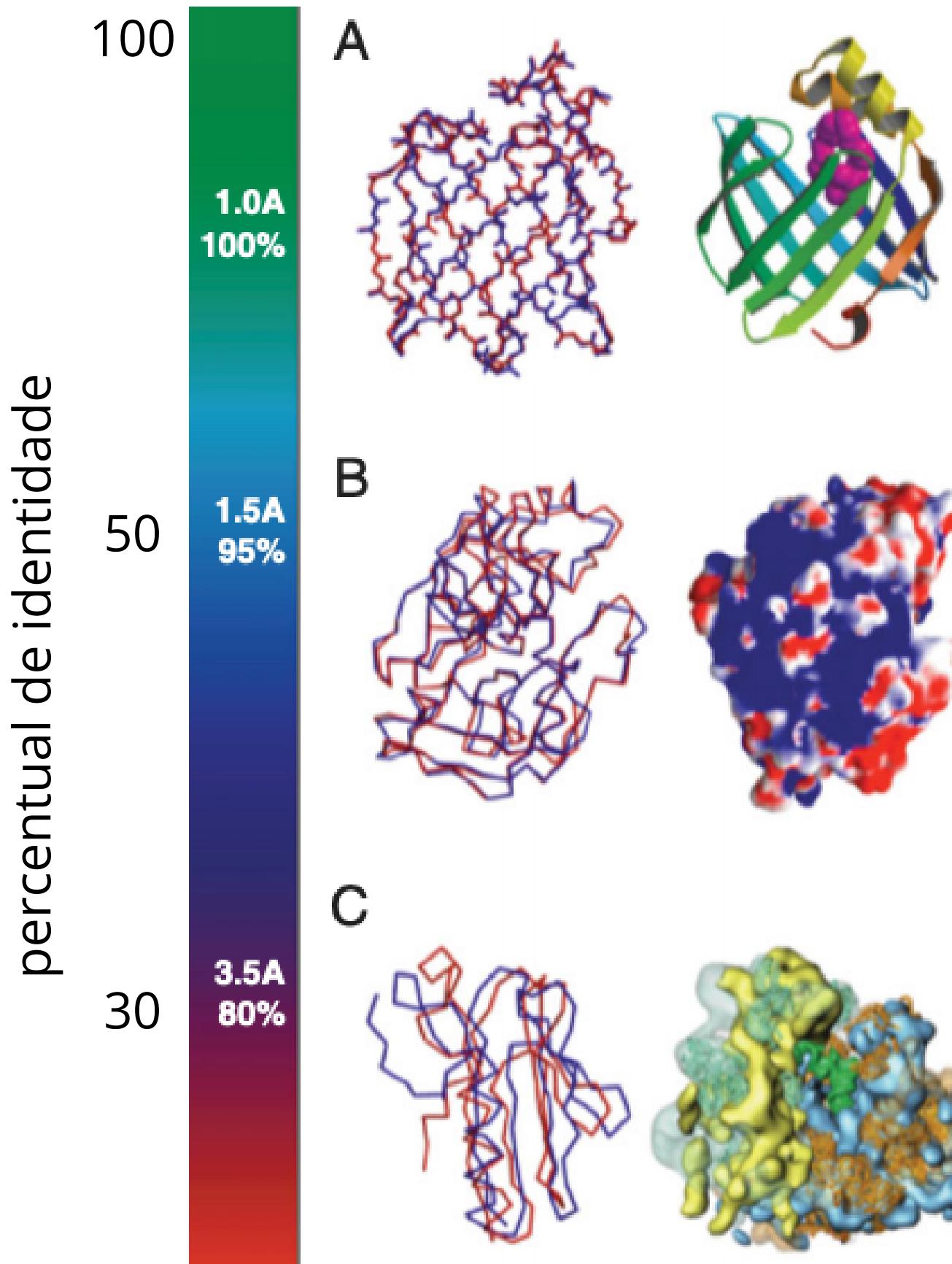


Utilidade dos modelos teóricos

Um modelo 3D não precisa ser absolutamente perfeito para ser útil em biologia, conforme demonstrado pelas aplicações listadas

O tipo de questão que pode ser abordada com um determinado modelo depende de sua precisão

Fonte: Adaptada de Webb, Benjamin, and Andrej Sali. "Comparative protein structure modeling using MODELLER." Current protocols in bioinformatics 54.1 (2016): 5-6.



- Estudo do mecanismo catalítico
- Desenho e melhoramento de ligantes
- *Docking* de macromoléculas e identificação de padrões
- Triagem virtual de pequenas moléculas
- Definição de epítopes de anticorpos
- Suporte a mutações sítio-dirigidas
- Inferência de relacionamentos funcionais
- Identificação de trechos conservados na estrutura das proteínas
- Descoberta de sítios funcionais pela busca de motivos 3D

Modelagem por
segmentos ou
fragmentos



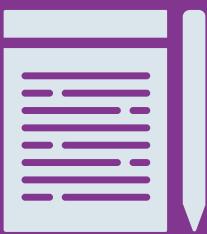
**E se eu não
encontro um
modelo com
mais de 30%
de identidade
com meu alvo?**

Modelagem híbrida

1. **Threading:** identificar proteínas molde de bancos de estruturas experimentais que têm uma estrutura semelhante ou motivo estrutural semelhante a sequência alvo (perfil de não sequência)
 - a. servidor de *meta-threading* que envolve sete programas de *threading*
2. **Montagem estrutural:** os **fragmentos** bem alinhados são usados para modelagem das partes alinhadas e as demais regiões (principalmente *loops* / caudas) são modelados ***ab initio***
 - a. Simulações Monte Carlo para tentar várias possibilidades de modelagem
 - i. várias temperaturas
 - ii. consideram potenciais estatísticos derivados do PDB
 - iii. **restrições espaciais** obtidas das estruturas molde
 - iv. predição de contatos
 - b. Agrupamento dos vários modelos e escolha de centróides
3. Seleção do modelo e refinamentos:
 - a. segunda rodada de *threading* dos centróides
 - b. refinamento do modelo final usando redes de ligações de hidrogênio

I-TASSER

*The Iterative Threading
ASSEMBLY Refinement*



Yang, Jianyi, et al. "The I-TASSER Suite: protein structure and function prediction." Nature methods 12.1 (2015): 7-8.



Zhiping Weng

I-TASSER

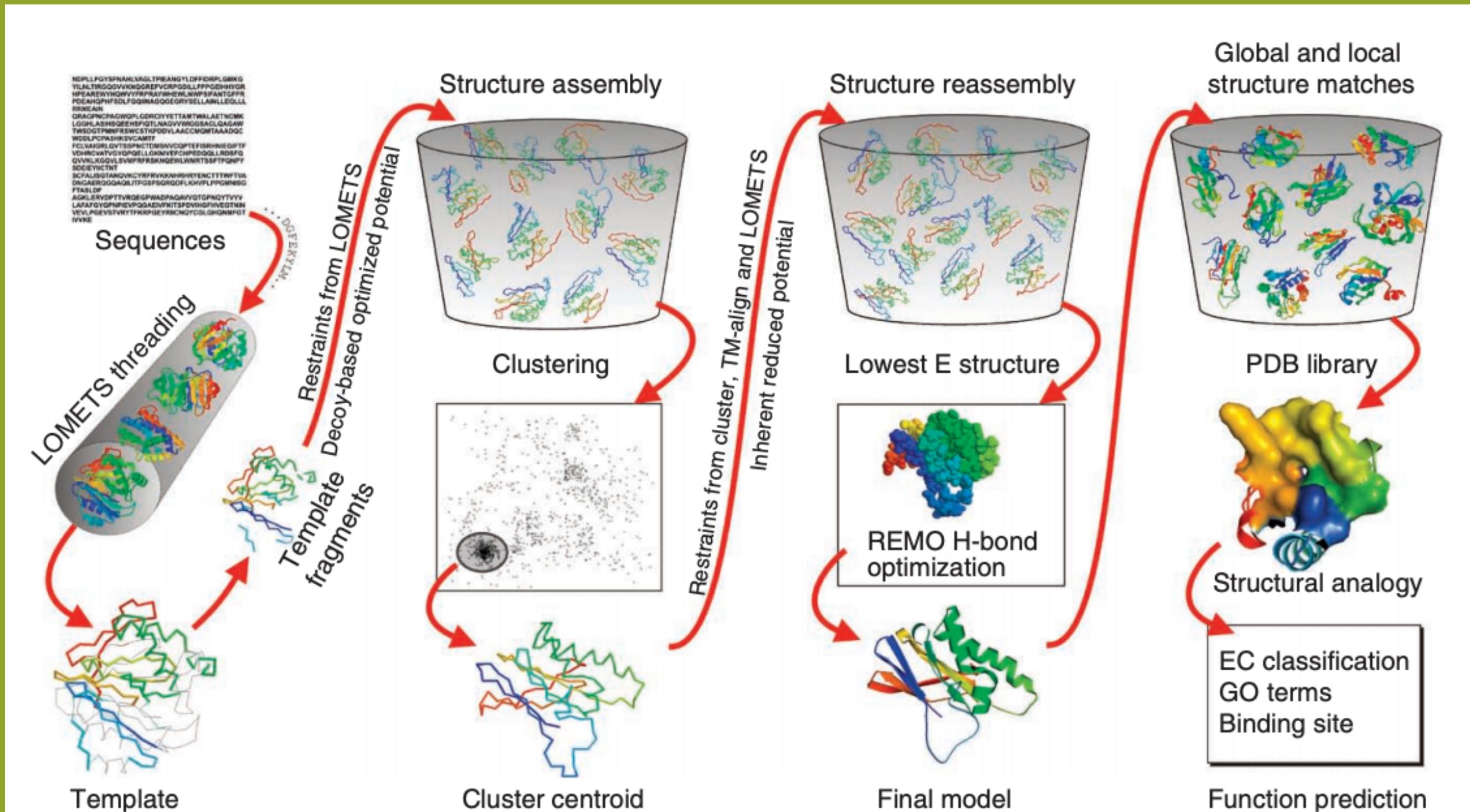


Figure 1 | A schematic representation of the I-TASSER protocol for protein structure and function predictions. The protein chains are colored from blue at the N-terminus to red at the C-terminus.