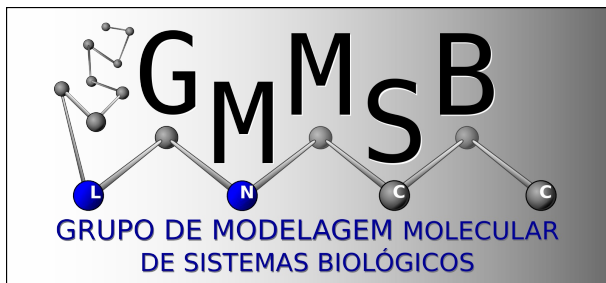


V Escola de Modelagem Molecular em Sistemas Biológicos

LNCC – Petrópolis, 23-27 de Agosto de 2010

Predição de Estrutura 3D de Proteínas por Técnicas de Modelagem Comparativa



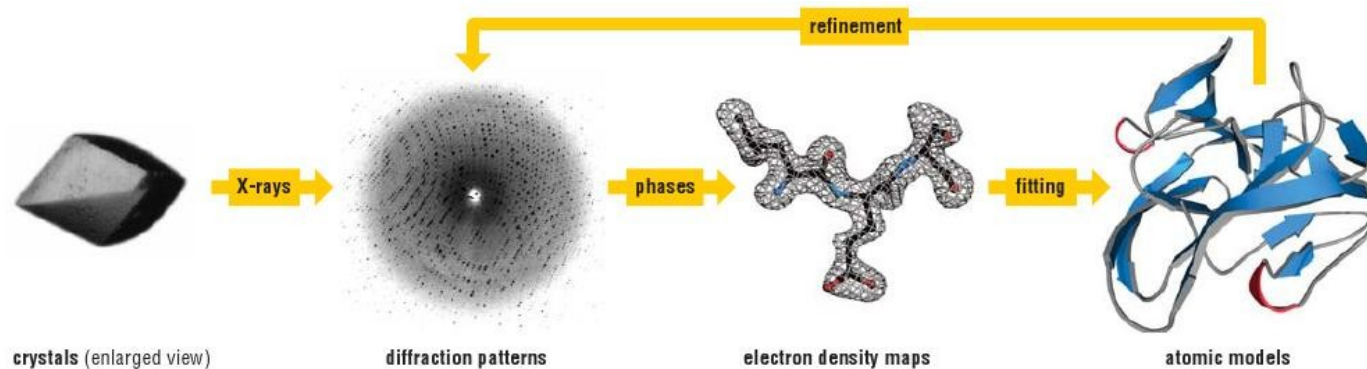
Priscila Vanessa Zabala Capriles Goliatt

Laboratório Nacional de Computação Científica - LNCC/MCT
Av. Getúlio Vargas, 333, Petrópolis, RJ, Brasil
E-mail: capriles@lncc.br

Predição Experimental

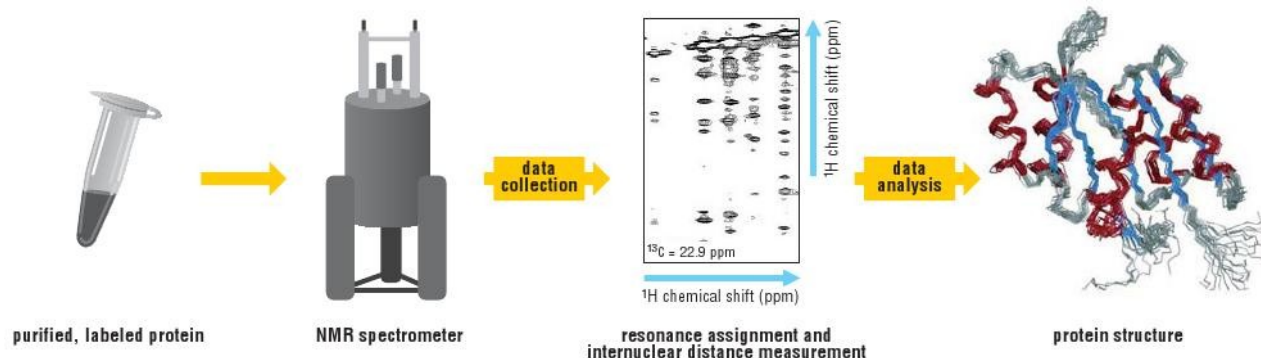
•A estrutura 3D de uma proteína pode ser determinada experimentalmente através da aplicação de técnicas como:

•Difração de raios X (RX):



Fonte: Petsko, G. A., Ringe, D. (2004). Protein Structure and Function. New Science Press Ltd.

•Ressonância Magnética Nuclear (RMN):



Fonte: Petsko, G. A., Ringe, D. (2004). Protein Structure and Function. New Science Press Ltd.

Por quê modelar?

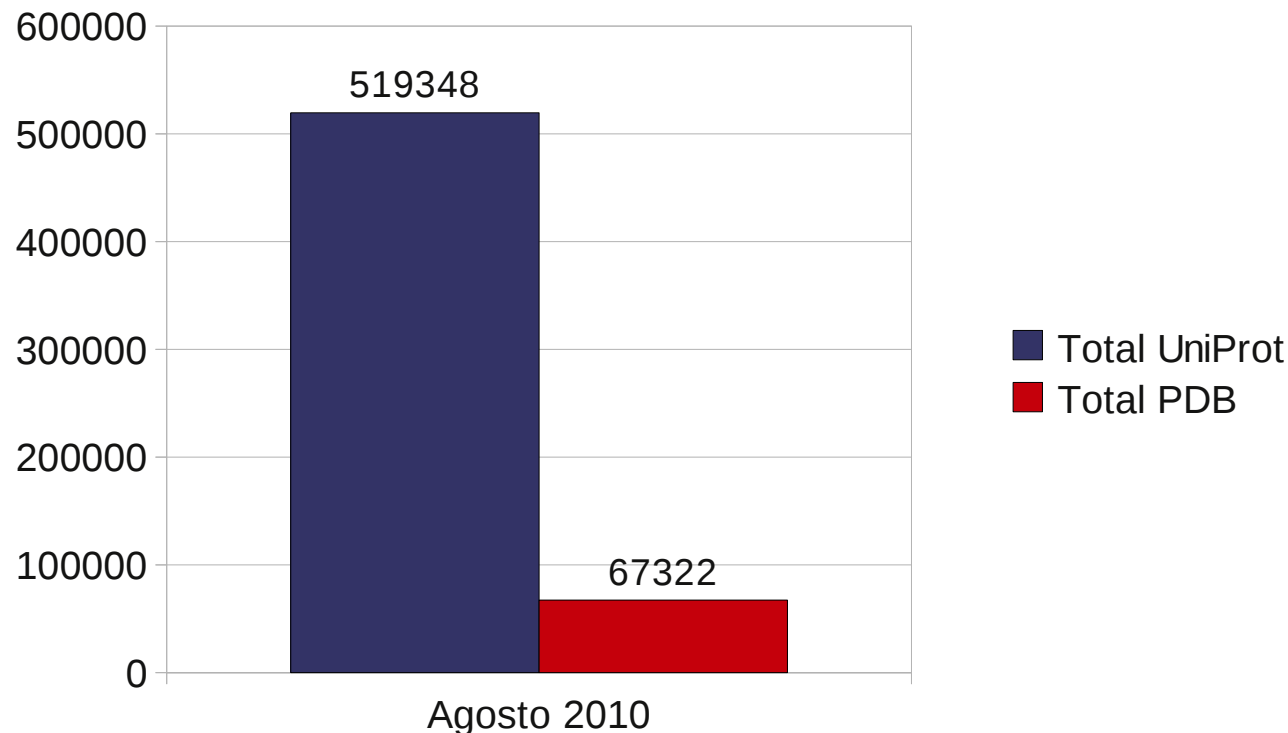
Aumento do número de projetos de seqüenciamento de genomas

+

Limitações da predição experimental de estruturas protéicas.

=

Aumento das técnicas de predição teórica



Predição Teórica de Estruturas 3D

- Modelagem Comparativa:**

Usa como molde a estrutura 3D de outra proteína de referência (resolvida experimentalmente e com coordenadas cartesianas depositadas em banco de dados de estruturas).

- Predição de Enovelamento de Proteínas (*Threading*):**

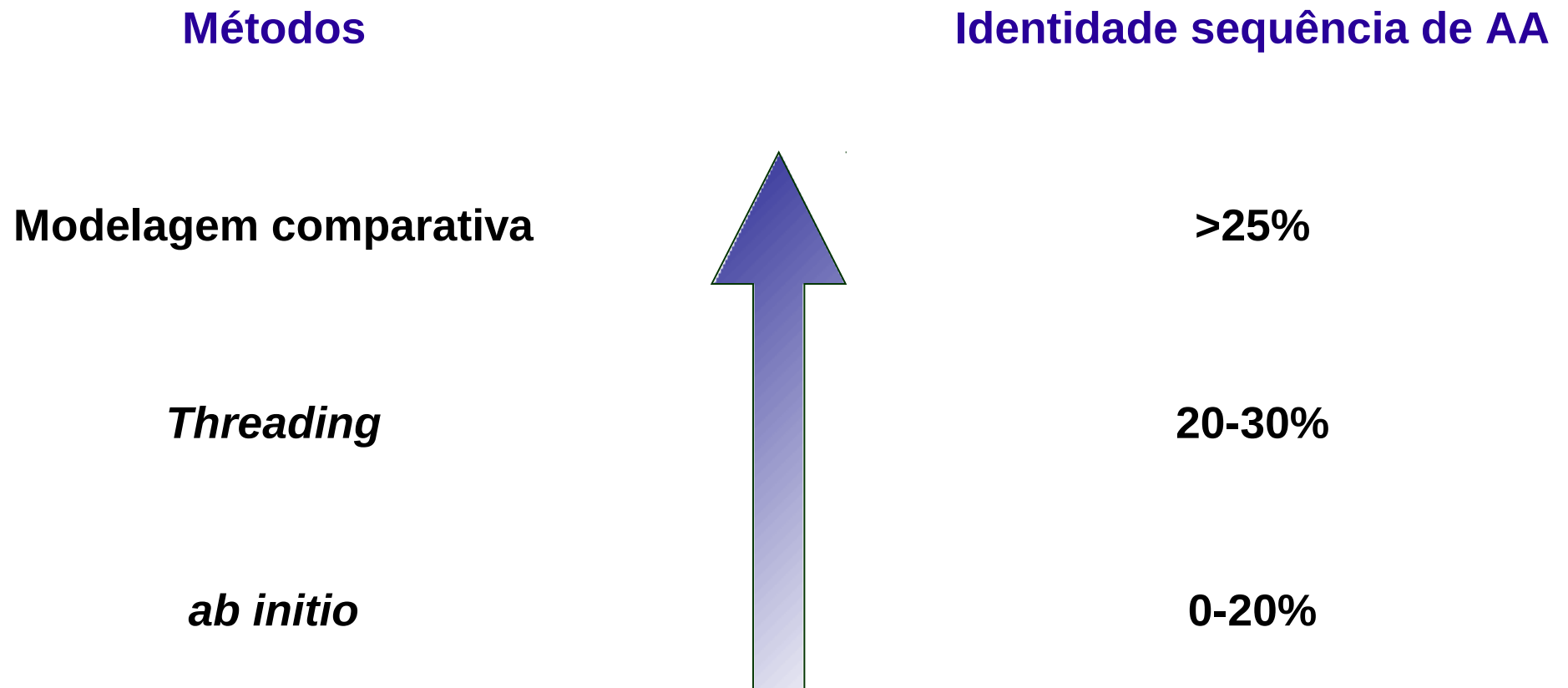
Tenta-se ajustar a estrutura da proteína de interesse aos tipos de enovelamentos de proteínas conhecidas (atualmente mais de 1000 tipos já foram registrados) e depositados em bibliotecas de enovelamentos.

- Predição por Primeiros Princípios (*ab initio*):**

Baseia-se nas propriedades físico-químicas conhecidas de cada aminoácido para a construção de funções de energia. Estas funções são minimizadas por algoritmos que realizam buscas no espaço de conformações que a proteína de interesse possa assumir.

Predição Teórica de Estruturas 3D

✓ Limites dos Métodos de Predição de Estrutura 3D:



Modelagem Comparativa

- O enovelamento protéico permite que aminoácidos seqüencialmente distantes interajam em uma mesma região espacial da proteína, formando distintos padrões conformacionais, dependendo das condições físico-químicas do ambiente.
- O alinhamento estrutural entre proteínas consideradas de uma mesma família, mesmo que seja apenas na região do sítio ativo, pode ser um importante método para se inferir a função da seqüência em estudo ^[3].

A evolução tende a conservar funções que dependem mais diretamente das estruturas 3D que das similaridades entre as seqüências de aminoácidos das proteínas ^[2].

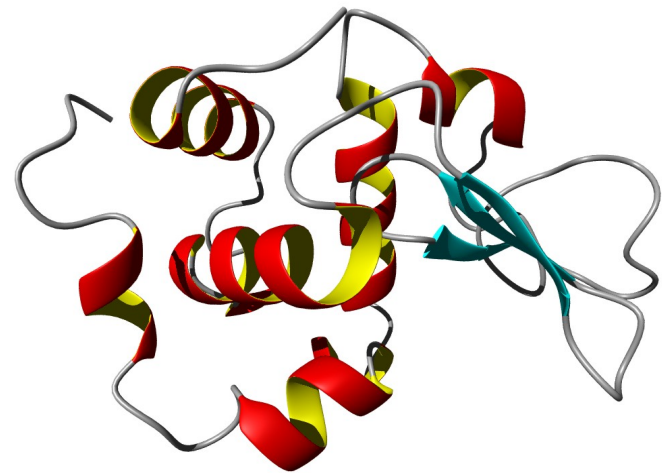
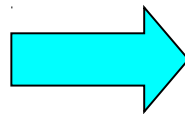
[2] Sánchez, R., et al (2000). *Protein structure modeling for structural genomics*. **Nature Structural Biology**, 7:986-990.

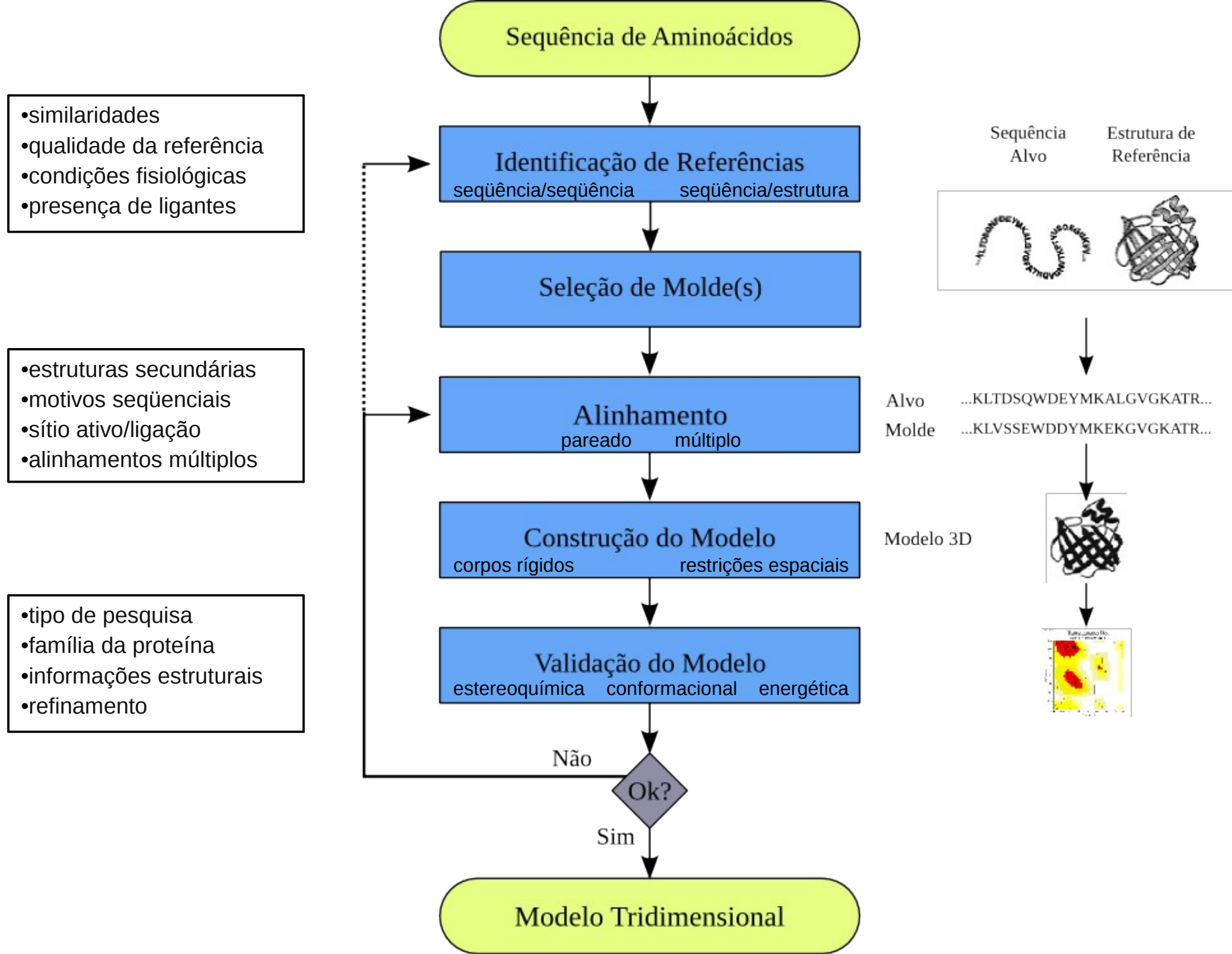
[3] Laskowski, R. A., et al (2003). *From protein structure to biochemical function*. **Journal of Structural and Functional Genomics**, 4:167-177.

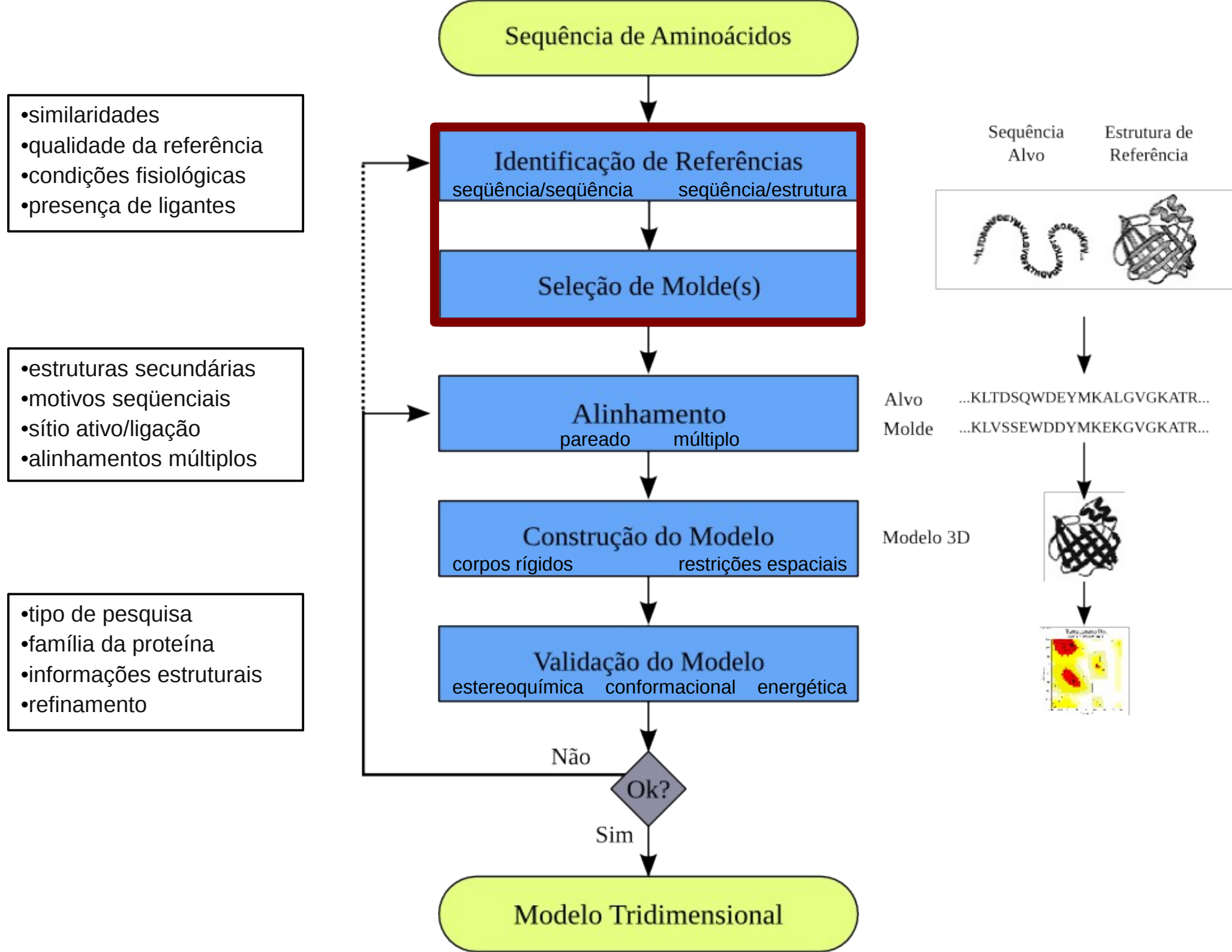
Modelagem Comparativa

- A metodologia se baseia no fato de que duas proteínas com função similar possuem similaridade estrutural.
- Por isso é possível usar estruturas resolvidas experimentalmente como moldes para construção de um modelo estrutural da sequência da proteína de interesse.

MQTVLKKRKKS
GYGYIPDIADIRDFS
YTPEKSVIAALPPKVDLTPSFQVYDQ
GRIGSCTANALAAAIQFERIHDKQSP
EFIPSRLFIYYNERKIEGHVNYDSGA
MIRDGIKVLHKLGVCPKEWPYGDTP
ADQKTEEFPPGAPASKKPSDQCYKEA
RNYKIT







Modelagem Comparativa

1. Identificação de Referências e Seleção de Molde(s):

>Seqüência alvo (formato FASTA)

```
MADNAMTRGSRACYNCGQPGHLSRECPTRPPGVMGDRACYNCGRMGHLSR  
ECPTRPPGVMGDRACYNCGRMGHLSRECPNRPAGGFRGVARGACYHCQQE  
GHLARDCPNAPPGGERACYNCGQTGHTSRACPVK
```

1)Busca por proteínas homólogas no PDB (via BLAST): para uma seqüência ser considerada como candidata a estrutura de referência, é necessário que haja no mínimo 25-30% de identidade entre as seqüências primárias.

2)Estudo estrutural detalhado das moléculas homólogas: busca por domínios, região catalítica, sítios alostéricos, etc.

3)Sobreposição (alinhamento) estrutural: buscar o menor desvio médio quadrático (RMSD) entre as seqüências indicadas como de referência. É importante que se faça uma visualização do alinhamento estrutural (e.g. VMD, SPDBViewer, Pymol).

1.1. Alinhamento Local Entre as Seqüências Alvo e de Referência:






A NEW NEUROTOXIN FROM BUTHUS MARTENSII KARSCH

Chains
E Value

A
2.4E-2



Characteristics

Release Date: 31-Mar-2000 **Exp. Method:** X Ray Diffraction



Classification

Toxin

Compound

Polymer: 1 **Molecule:** PROTEIN (BMK M2) **Chains:** A

Authors

He, X.L., Deng, J.P., Li, H.M., Wang, D.C.

Alignment stats

Length: 47 **Score:** 33.8834bits (76) **E-value:** 0.0237047 **Identities:** 16/47 (34%) **Positives:** 23/47 (49%)
Gaps: 5/47 (11%) *(Blast results help)*

Alignment

	37	50	60	70	78
Query	CVY AENLGEDCENLCKQQA	TDGFCRQP	-----	HCFCTDMPDDYATR	
Sbjct	CVY E C NLC + A G+C+			C+C ++PD+ R	
	CVYECARNEYCNLCTKNGAKSGYCWGCIELPDNVP	IR			


1sn4 **STRUCTURE OF SCORPION NEUROTOXIN BMK M4**

Chains A
E Value 7.5E-2

Characteristics **Release Date:** 17-Nov-1999 **Exp. Method:** X Ray Diffraction

Resolution: 1.30 Å

Classification **Toxin**

Compound **Polymer:** 1 **Molecule:** PROTEIN (NEUROTOXIN BMK M4) **Chains:** A

Authors **He, X.L., Li, H.M., Liu, X.Q., Zeng, Z.H., Wang, D.C.**

Alignment stats **Length:** 49 **Score:** 32.3426bits (72) **E-value:** 0.0749718 **Identities:** 17/49 (35%) **Positives:** 21/49 (43%)
Gaps: 5/49 (10%) *(Blast results help)*

Alignment

	35	40	50	60	70	78	
Query	ELCVY	AENLG	EDCEN	LCKQK	ATDGF	CR-----QPHCF	CTDMPDDYATR
E CVY		E C	LC	A	G+C+		C+C +PDD R
Sbjct	ENC	VYHC	AGNE	GCN	KLCT	DNGAES	GYCQWGGRYGNACWIKLPDDVP

1. Identificação de Referências e Seleção de Molde(s):

1.2. Análise de Estrutura Secundária:

- 1CHZ
- Download Files
 - PDB text
 - PDB (Header) text
 - PDB gz
 - mmCIF text
 - mmCIF (Header) text
 - mmCIF gz
 - PDBML/XML text
 - PDBML/XML (Header) text
 - PDBML/XML gz
 - Structure Factor text
 - Structure Factor gz
 - Biological Unit gz
 - FASTA Sequence
- Download Original Files
- Display Files
- Display Molecule
- Structural Reports
- External Links
- Structure Analysis
 - Geometry
 - RCSB Graphics
 - RCSB Tables
 - MolProbity Ramachandran Plot
- Help

HelpStructure SummaryBiology & ChemistryMaterials & MethodsSequence DetailsGeometry

Sequence:Structure Details1CHZ

Chain A (polymer 1)

Description

Chain Type

UniProt reference

Length

SCOP domain assignment

DSSP secondary structure

More annotations

PROTEIN (BMK M2)

polypeptide(L)

P58488

[show this sequence below]

64 residues

d1chza Scorpion toxin: 64 residues

15% helical (1 helices; 10 residues)

42% beta sheet (7 strands; 27 residues)

Select

SCOP

DSSP

PDB

PDB

SCOP

DSSP

PDB

PDB

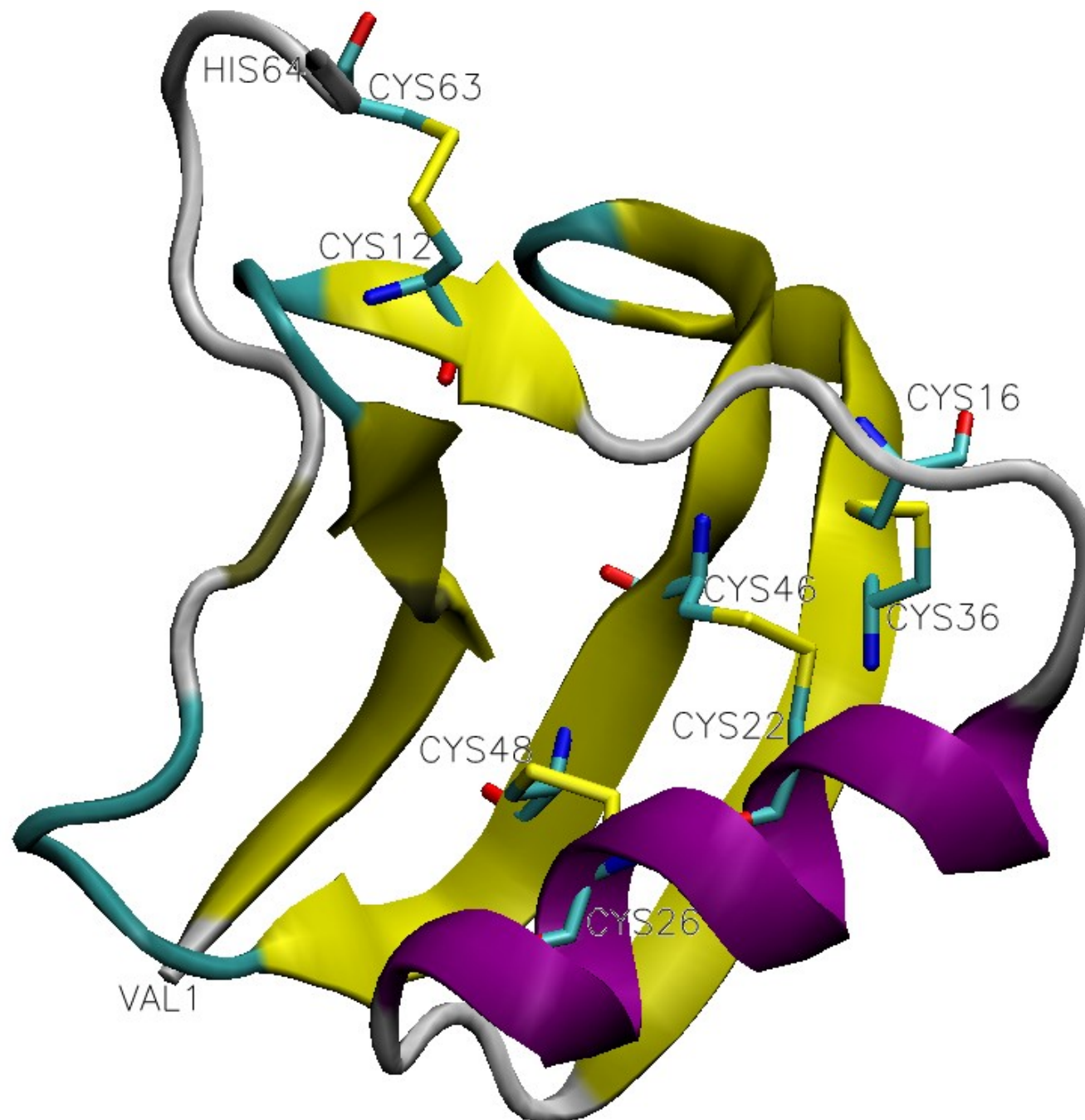
VRDAYIAKPHN¹CVYE¹⁰CARNEY²⁰CNNL²⁵CTKNGAKSGY³⁰CQWSGKYGNG⁴⁰WCIELPDNVPIRVP⁶⁰

GKCH

61 64

1. Identificação de Referências e Seleção de Molde(s):

1.2. Análise das Cisteínas:



Modelagem Comparativa

1. Identificação de Referências e Seleção de Molde(s):

1.3. Alinhamento Estrutural:

Sobreposição (alinhamento) estrutural das proteínas escolhidas como referência para definição de regiões estruturalmente conservadas.

1) Geralmente compostas pelas estruturas secundárias das proteínas:

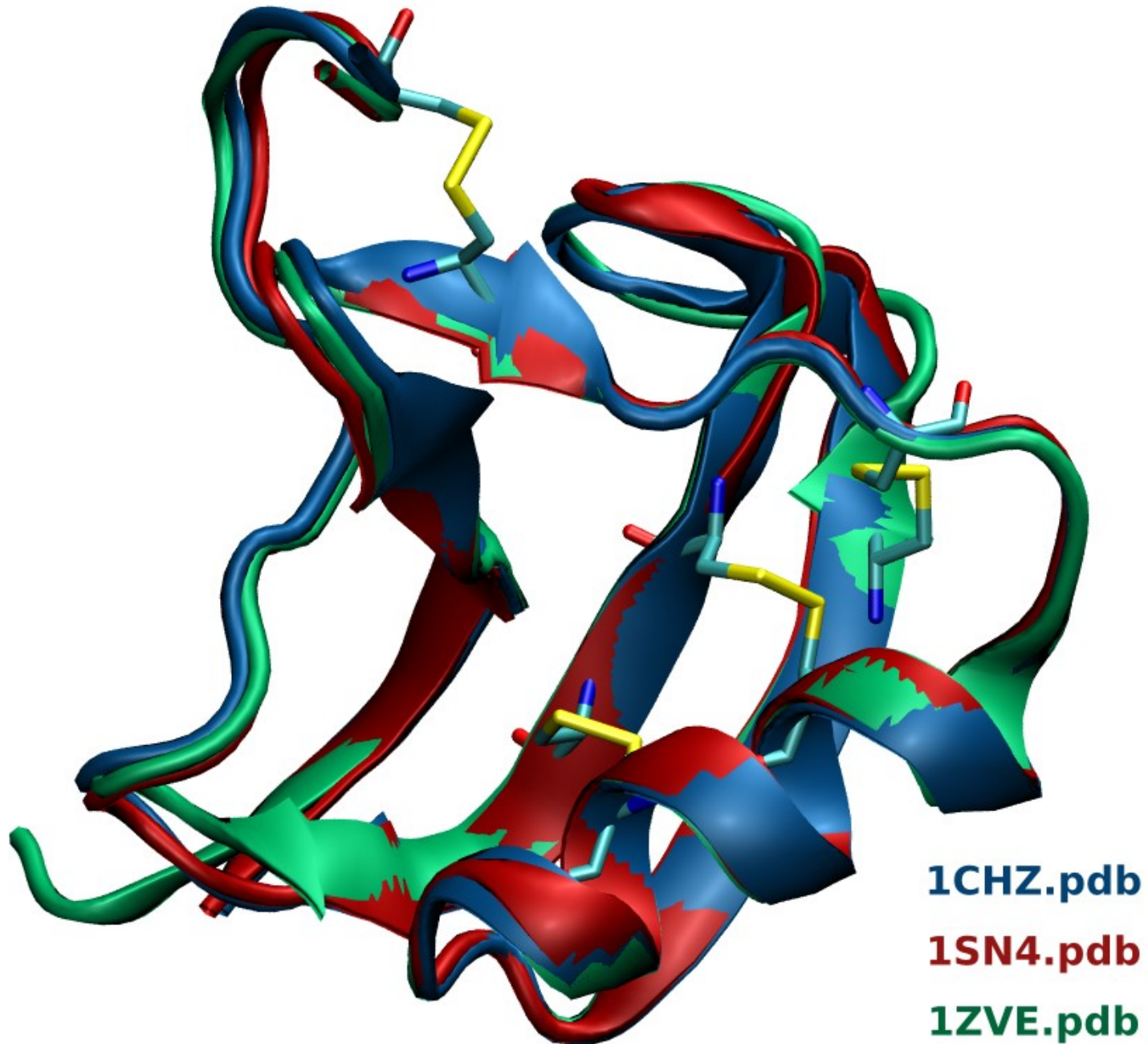
Regiões em que os C α das proteínas de referência encontram-se afastadas a menos de 3,8 Å entre si.

2) Alguns algoritmos de modelagem comparativa (e.g. Swiss-Model, Modeller) possuem internamente algoritmos para predição de estrutura secundária.

3) Existem algoritmos independentes aplicados somente para a predição de estrutura secundária (e.g. Psi-Pred, Jufo).

1. Identificação de Referências e Seleção de Molde(s):

1.3. Alinhamento Estrutural:



Modelagem Comparativa

1. Identificação de Referências e Seleção de Molde(s):

1.3.3. Predição de Estrutura Secundária:

PSIPRED PREDICTION RESULTS

Key

Conf: Confidence (0=low, 9=high)

Pred: Predicted secondary structure (H=helix, E=strand, C=coil)

AA: Target sequence

PSIPRED HFORMAT (PSIPRED V2.6 by David Jones)

Conf: 996122000587777786887802999999999999999985997870737731168999

Pred: CCCCECCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHCCCCCCC

AA: APAAKDWREEGAVTAVKNQGICGSCWAFAAIGNIEGQWFLAGNPLTRLSEQMLVSCDNTN
10 20 30 40 50 60

Conf: 987797989999999985169886376788105478216760378765699422999489

Pred: CCCCCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEEEEEEECCC

AA: SGC GGLSSKAFEWIVQENNGAVYTEDSYYPYHSCIGIKLPCKDSVRTVGATITGHVELPQ
70 80 90 100 110 120

Conf: 999999999985998999987343124884767999897786999999852199889999

Pred: CHHHHHHHHHHHHCCEEEEEECCECCCCCCCCCECCCCCCCCCEEEEEEEEECEEEEEEE

AA: DEAQIAASGAVKGPLSVAVDASSWPPYTGGVLTNCVSKRLSHAVLLVGYNDSSAAVPYWI
130 140 150 160 170 180

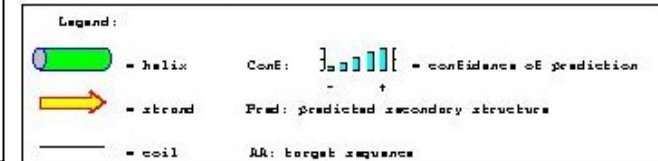
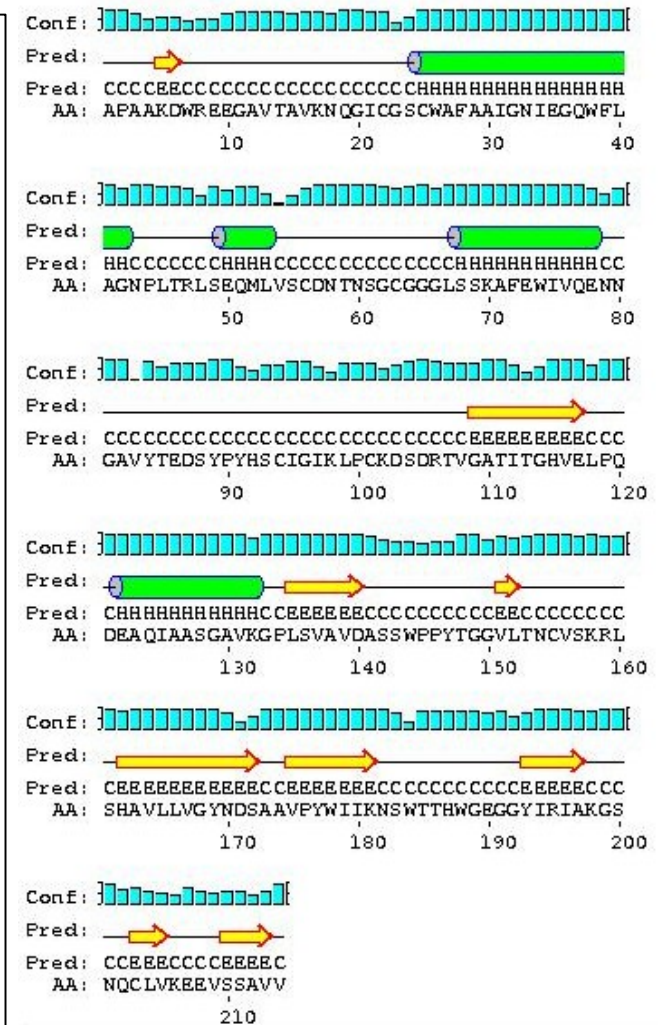
Conf: 8961788785726999769993530052121129

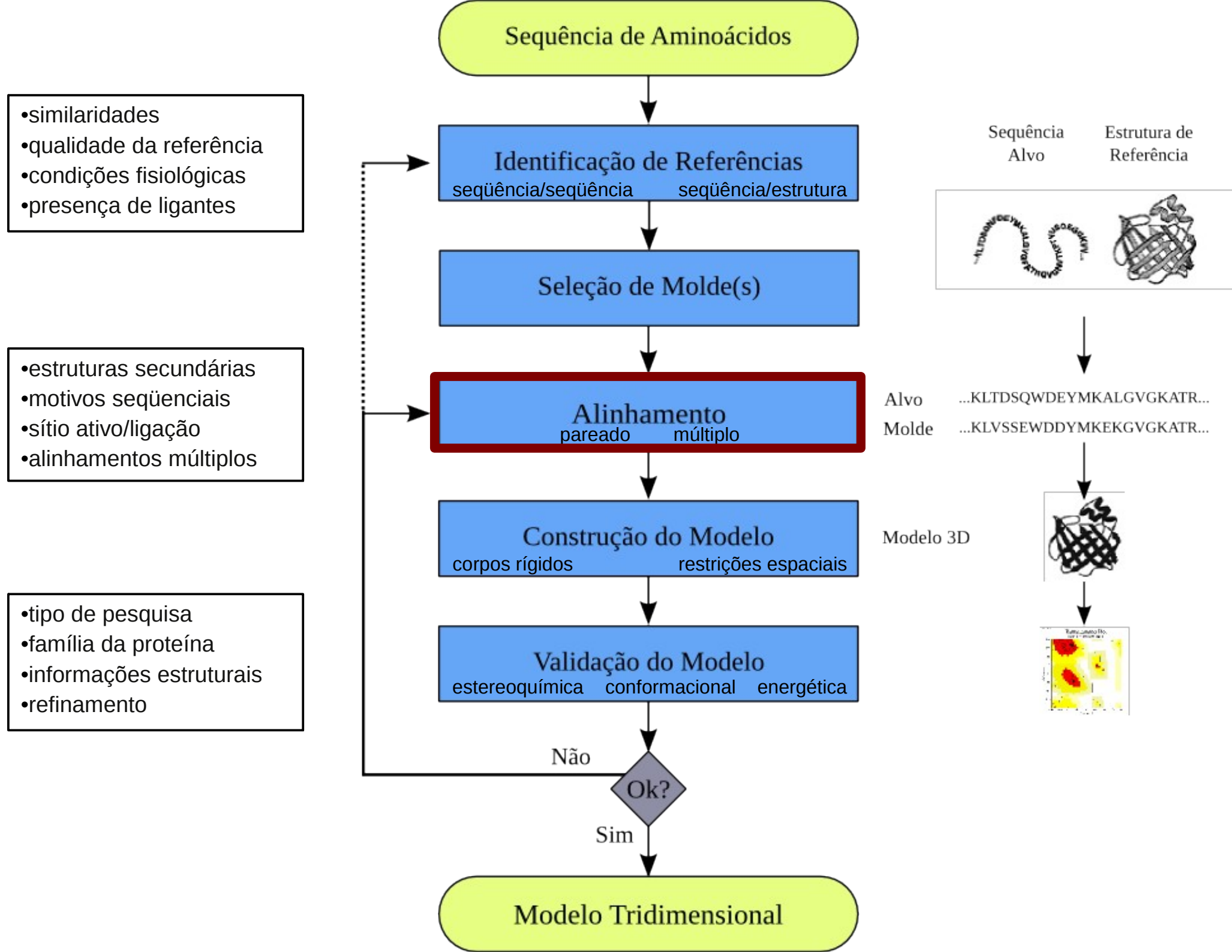
Pred: ECCCCCCCCCEEEEECCCCCEEECCCCCEEEEC

AA: KNSWTTWGGGYIRIAKGSNQCLVKEEVSSAVV
190 200 210

Calculate PostScript, PDF and JPEG graphical output for this result using:

<http://bioinf3.cs.ucl.ac.uk/cgi-bin/psipred/gra/nph-view2.cgi?id=7fd46c0eca549dff.psi>





Modelagem Comparativa

2. Alinhamento Global Entre as Seqüências Alvo e de Referência(s):

- 1) Toma-se a seqüência com o maior número de resíduos.
- 2) Alinha a segunda seqüência ao longo de toda a extensão da primeira, de forma a coincidir o maior número de resíduos idênticos e similares.
- 3) Resulta na abertura de tantos *gaps* quantos forem necessários.
- 4) Atenção para a presença de heteroátomos, águas e ligantes.

2.1. Alinhamento com Única Referência:

```
>P1; ./sequences/1CHZ
structureX: ./sequences/1CHZ      :1      :A:+152 : ::: 1.76:-1.00
VRDAYIAKP-----HNCVYECARNEYC>NNLC
TKNGAKSGYCQWSGKYGNGCWCIELPDNVPPIRVPGKCH. /wwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/A0F0C2
sequence: tr|A0F0C2|A0F0C2 MESMA:      : :      : ::: -1.00:-1.00
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENLC
KQOKATDGFCRQ-----PHCFCTDMPDDYATRPDTPVDPI-M-----
-----
-----*
```

Modelagem Comparativa

2. Alinhamento Global Entre as Seqüências Alvo e de Referência(s):

- 1) Toma-se a seqüência com o maior número de resíduos.
- 2) Alinha a segunda seqüência ao longo de toda a extensão da primeira, de forma a coincidir o maior número de resíduos idênticos e similares.
- 3) Resulta na abertura de tantos *gaps* quantos forem necessários.
- 4) Atenção para a presença de heteroátomos, águas e ligantes.

2.1. Alinhamento com Única Referência:

```
>P1; ./sequences/1CHZ
structureX: ./sequences/1CHZ      :1      :A:+152 : ::: 1.76:-1.00
VRDAYIAKP-----HNCVYECARNEYC>NNLC
TKNGAKSGYCQWSGKYGNGCWCIELPDNVPPIRVPGKCH./wwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/A0F0C2
sequence:tr|A0F0C2|A0F0C2_MESMA:      :      : ::: -1.00:-1.00
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENLC
KQQKATDGFCRQ-----PHCFCTDMPDDYATRPDTPVDPI-M-----
-----*
-----*
```

Modelagem Comparativa

2. Alinhamento Global Entre as Seqüências Alvo e de Referência(s):

- 1) Toma-se a seqüência com o maior número de resíduos.
- 2) Alinha a segunda seqüência ao longo de toda a extensão da primeira, de forma a coincidir o maior número de resíduos idênticos e similares.
- 3) Resulta na abertura de tantos *gaps* quantos forem necessários.
- 4) Atenção para a presença de heteroátomos, águas e ligantes.

2.1. Alinhamento com Única Referência:

```
>P1; ./sequences/1CHZ
structureY: ./sequences/1CHZ .1 :A:+152 : : : 1.76:-1.00
VRDAYIAKP-----HACVYECARNEYC>NNLC
TKNGAKSGYCWSGRYGNCWCIEELPDNVPTRVPGKCH. /wwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/A0F0C2
sequence:tr|A0F0C2|A0F0C2 MESMA: : : : :-1.00:-1.00
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENLC
KQQKATDGFCRQ-----PHCFCTDMPDDYATRPDTPVDPI-M-----
-----
-----*
```

2. Alinhamento Global Entre as Seqüências Alvo e de Referência(s):

2.1. Alinhamento com Múltiplas Referências:

```
>P1; ./sequences/1CHZ
structureX: ./sequences/1CHZ:    1 :A:+152 : ::: 1.76:-1.00
-----VRDAYIAKP---HNCVYECARNEYC>NNL
CTKNGAKSGYCQWSGKYGNGCWCIELPDNVPIRVPGKCH---./wwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/1ZVE
structureX: ./sequences/1ZVE:    1 :A:+169 : ::: 1.70:-1.00
-----NSVRDAYIAGP---HNCVYECARNEYCNDL
CTKNGAKSGYCQWVGKYGNGCWCIELPDNVPIRVPGKCH---/.wwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/1SN4
structureX: ./sequences/1SN4:    1 :A:+186 : ::: 1.30:-1.00
-----VRDAYIAKP---ENCVYHCAGNEGCNKL
CTDNGAESGYCQWGGRYGNACWCIKLPDDVPIRVPGKCH---/.....ww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/A0F0C2
sequence:tr|A0F0C2|A0F0C2_MESMA::::::::::
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENL
CKQQKATDGFCRQP-----HCFCTDMPDDYATR-PDTVDPIM*
```

2. Alinhamento Global Entre as Seqüências Alvo e de Referência(s):

2.1. Alinhamento com Múltiplas Referências:

```
>P1; ./sequences/1CHZ
structureX: ./sequences/1CHZ: 1 :A:+152 : ::: 1.76:-1.00
-----VRDAYIAKP---HNCVYECARNEYCNNL
CTKNGAKSGYCQWSGKYGNGCWCIELPDNVPIRVPGKCH---./wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*
```

```
>P1; ./sequences/1ZVE
structureX: ./sequences/1ZVE: 1 :A:+169 : ::: 1.70:-1.00
-----NSVRDAYIAGP---HNCVYECARNEYCNDL
CTKNGAKSGYCQWVGKYGNGCWCIELPDNVPIRVPGKCH---/.wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*
```

```
>P1; ./sequences/1SN4
structureX: ./sequences/1SN4: 1 :A:+186 : ::: 1.30:-1.00
-----VRDAYIAKP---ENCVYHCAGNEGCNKL
CTDNGAESGYCQWGGRYGNACWCIKLPDDVPIRVPGKCH---/.....ww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*
```

```
>P1; ./sequences/A0F0C2
sequence: tr|A0F0C2|A0F0C2 MESMA: :::::
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENL
CKQQKATDGFCRQP-----HCFCTDMPDDYATR-PDTVDPIM*
```

2. Alinhamento Global Entre as Seqüências Alvo e de Referência(s):

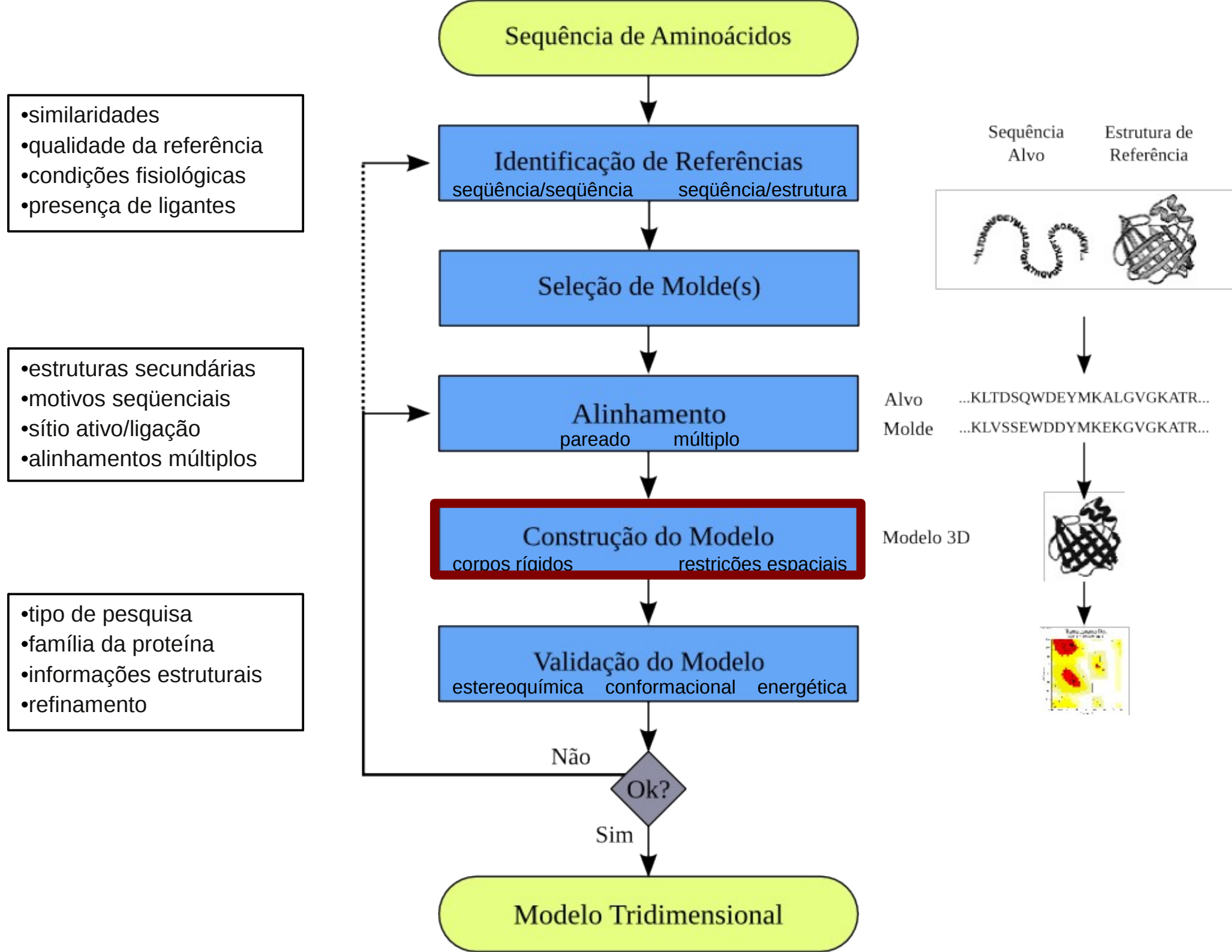
2.1. Alinhamento com Múltiplas Referências:

```
>P1; ./sequences/1CHZ
structureX: ./sequences/1CHZ:    1 :A:+152 : ::: 1.76:-1.00
-----VRDAYIAKP---HNCVYECARNEYCNNL
CTKNGAKSGYCQWSGKYGNGCWCIELPDNVPIRVPGKCH---./wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/1ZVE
structureX: ./sequences/1ZVE:    1 :A:+169 : ::: 1.70:-1.00
-----NSVRDAYIAGP---HNCVYECARNEYCNDL
CTKNGAKSGYCQWVGKYGNGCWCIELPDNVPIRVPGKCH---/.wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

>P1; ./sequences/1SN4
structureX: ./sequences/1SN4:    1 :A:+186 : ::: 1.30:-1.00
-----VRDAYIAKP---ENCVYHCAGNEGCNKL
CTDNGAESGYCQWGGRYGNACWCIKLPDDVPIRVPGKCH---/.....ww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwww*

>P1; ./sequences/A0F0C2
sequence:tr|A0F0C2|A0F0C2_MESMA:::::
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENL
CKQQKATDGFCRQP-----HCFCTDMPDDYATR-PDTVDPIM*
```



Modelagem Comparativa

✓ Algoritmos para Modelagem Comparativa (Corpos Rígidos):

Constrói um modelo por partes baseando-se no fato de que proteínas homólogas possuem regiões estruturalmente conservadas como α -hélices e folhas β :

- 1) Definição das regiões estruturalmente conservadas: através de predição de estruturas secundárias.
- 2) Alinhamento considerando-se a média das posições dos $C\alpha$ das seqüências de aminoácidos das regiões estruturalmente conservadas nas seqüências de referência.
- 3) As regiões que não satisfazem as exigências são chamadas de regiões variáveis (geralmente, regiões de voltas que conectam as regiões conservadas).

Modelagem Comparativa

✓ Algoritmos para Modelagem Comparativa (Corpos Rígidos):

4) A cadeia principal das regiões variáveis: pode ser obtida em bancos de dados de estruturas protéicas, que apresentam conjuntos de voltas determinados através do número de aminoácidos e do tipo de estruturas secundárias que conectam.

5) Inserção das cadeias laterais dos aminoácidos, através de busca em bibliotecas de rotâmeros.

Modelagem Comparativa

✓ Algoritmos para Modelagem Comparativa (Corpos Rígidos):



[v4.0](#)

[Index](#)

[User Guide](#)

[Menus](#)

[Tips & Tricks](#)

[Tutorial](#)

[Download](#)

[Feedback](#)

[Art Gallery](#)

[References](#)

The

Swiss Institute of Bioinformatics

presents

Swiss-Pdb Viewer

DeepView

v4.0

presents

by Nicolas Guex , Alexandre Diemand , Manuel C. Peitsch , & Torsten Schwede

by Nicolas.Guex &
Torsten.Schwede

News: version 4.0.1 has been released on October 7th 2008

- ◆ The Thymine C6 atom is now correctly loaded
- ◆ Fixed the POV-Ray output endless loop on the PC version
- ◆ The occasional crash while saving PDB files (such as 2i37) on the PC has been fixed
- ◆ Updated the address of the Uppsala Electron Density Map Server

Thank you to all users who reported those bugs!

News: version 4.0 has been released on June 27th 2008

- ◆ First of all, and most importantly, PC and Mac versions have been resynchronized.
- ◆ Support for direct login modelling submissions and result retrieval to/from the SwissModel Workspace
- ◆ Enhanced [Import](#) Menu
- ◆ Enhanced user interface with taxonomy support, and new sequence alignment features.
- ◆ New [3D Motif](#) Searching feature
- ◆ Easy access to external user defined scripts directly from the interface
- ◆ Revised Help files accessible from the interface

Modelagem Comparativa

✓ Algoritmos para Modelagem Comparativa (Restrições Espaciais):

- 1) Alinhamento entre a seqüência em estudo e a de referência: extrai-se do molde suas restrições espaciais (distâncias e ângulos) e transfere-as para o modelo.
- 2) O tamanho das ligações bem como seus ângulos preferenciais são obtidos de campos de força (e.g. CHARMM).
- 3) Dados para os átomos não ligados são obtidos por análise estatística de um grupo representativo de estruturas conhecidas (limitação do número de possíveis conformações que o modelo pode assumir).

Modelagem Comparativa

✓ Algoritmos para Modelagem Comparativa (Restrições Espaciais):

4) Principal característica deste método:

- Obtenção empírica das restrições espaciais (funções densidade de probabilidade): A partir de banco de dados contendo informações sobre alinhamentos entre estruturas protéicas de alta resolução.
- Minimização das violações das restrições espaciais: As restrições espaciais e os termos de energia são combinados em uma função objetivo, sendo submetida a métodos de otimização (*e.g.* gradiente conjugado).

Modelagem Comparativa

✓ Algoritmos para Modelagem Comparativa (Restrições Espaciais):

Modeller

Program for Comparative Protein
Structure Modelling by Satisfaction
of Spatial Restraints

```
A I L V G S M P R R D G M E R K D L L K A N V K I F K C Q G A  
V E V C P V D C F Y E G P N F L V I H P D E C I D C A L C E P  
V A C K P E C P V N I I Q G S - - Y A I D A D S C I D C Q G  
G - - I A C G A C K P E C P V N I I Q G S - - I Y A I D A D S
```



About MODELLER

MODELLER is used for homology or comparative modeling of protein three-dimensional structures (1,2). The user provides an alignment of a sequence to be modeled with known related structures and MODELLER automatically calculates a model containing all non-hydrogen atoms. MODELLER implements comparative protein structure modeling by satisfaction of spatial restraints (3,4), and can perform many additional tasks, including de novo modeling of loops in protein structures, optimization of various models of protein structure with respect to a flexibly defined objective function, multiple alignment of protein sequences and/or structures, clustering, searching of sequence databases, comparison of protein structures, etc. MODELLER is [available for download](#) for most Unix/Linux systems, Windows, and Mac.

Several graphical interfaces to MODELLER are [commercially available](#). There are also many other [resources and people using Modeller](#) in graphical or web interfaces or other frameworks.

1. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 200.
2. M.A. Marti-Renom, A. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29, 291-325, 2000.
3. A. Sali & T.L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779-815, 1993.
4. A. Fiser, R.K. Do, & A. Sali. Modeling of loops in protein structures, Protein Science 9. 1753-1773, 2000.

The current release of Modeller is **9v4**, which was released on May 15th, 2008. Modeller is currently maintained by [Ben Webb](#).

About MODELLER
MODELLER News
Download & Installation
Release Notes
Data file downloads
Registration
Accelrys licensing
Discussion Forum
Subscribe
Browse archives
Search archives
Documentation
FAQ
Tutorial
Online manual
Wiki
Developers' Pages
Contact Us

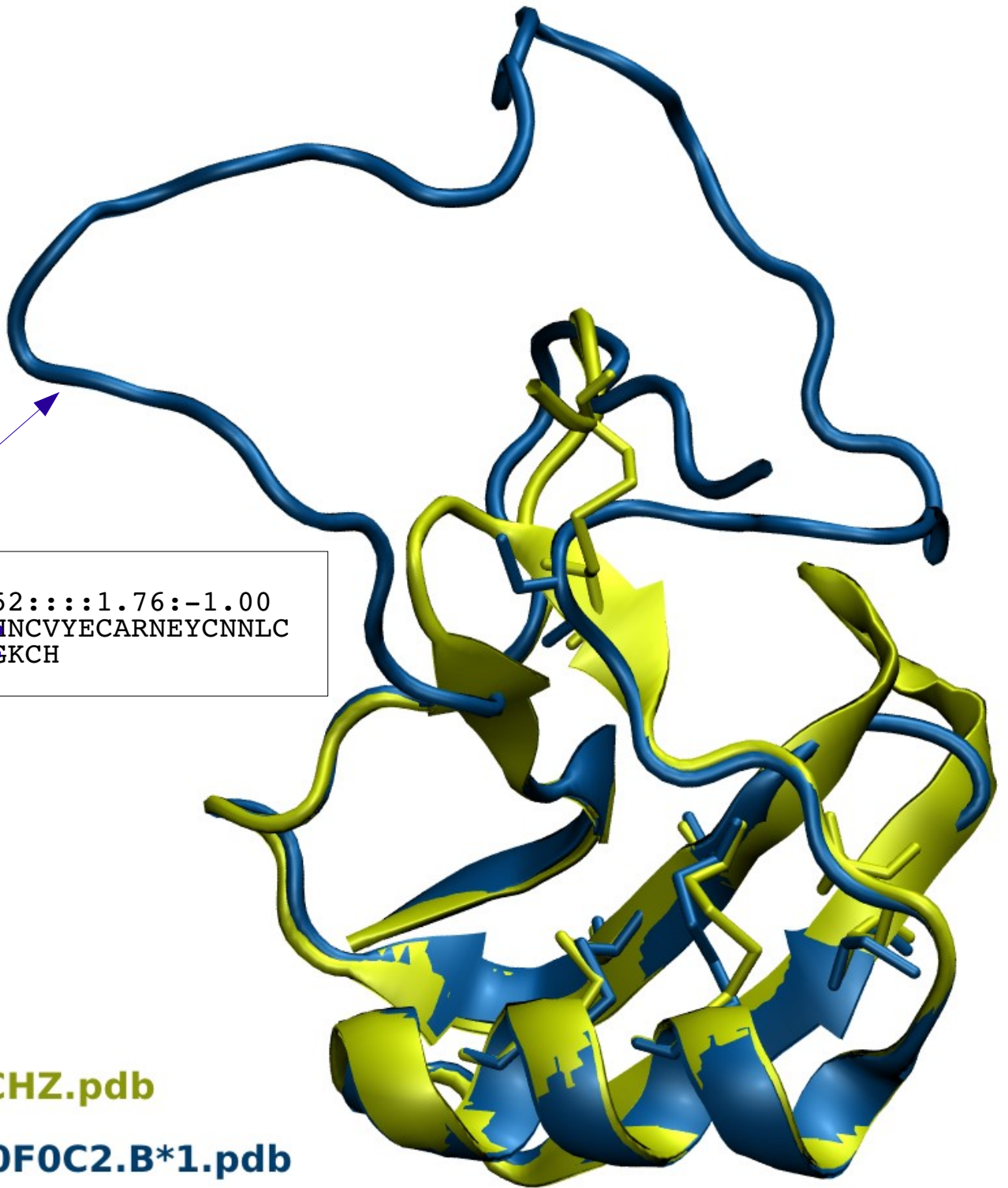
3. Construção do Modelo:

3.1. Baseado em Única Referência

```
>P1;./sequences/1CHZ  
structureX:./sequences/1CHZ:1:A:+152::::1.76:-1.00  
VRDAYIAK-----NLCVYECARNEYCNNLC  
TKNGAKSGYCQWSGRYGNCCWCIELPDNVPTRVPGKCH
```

1CHZ.pdb

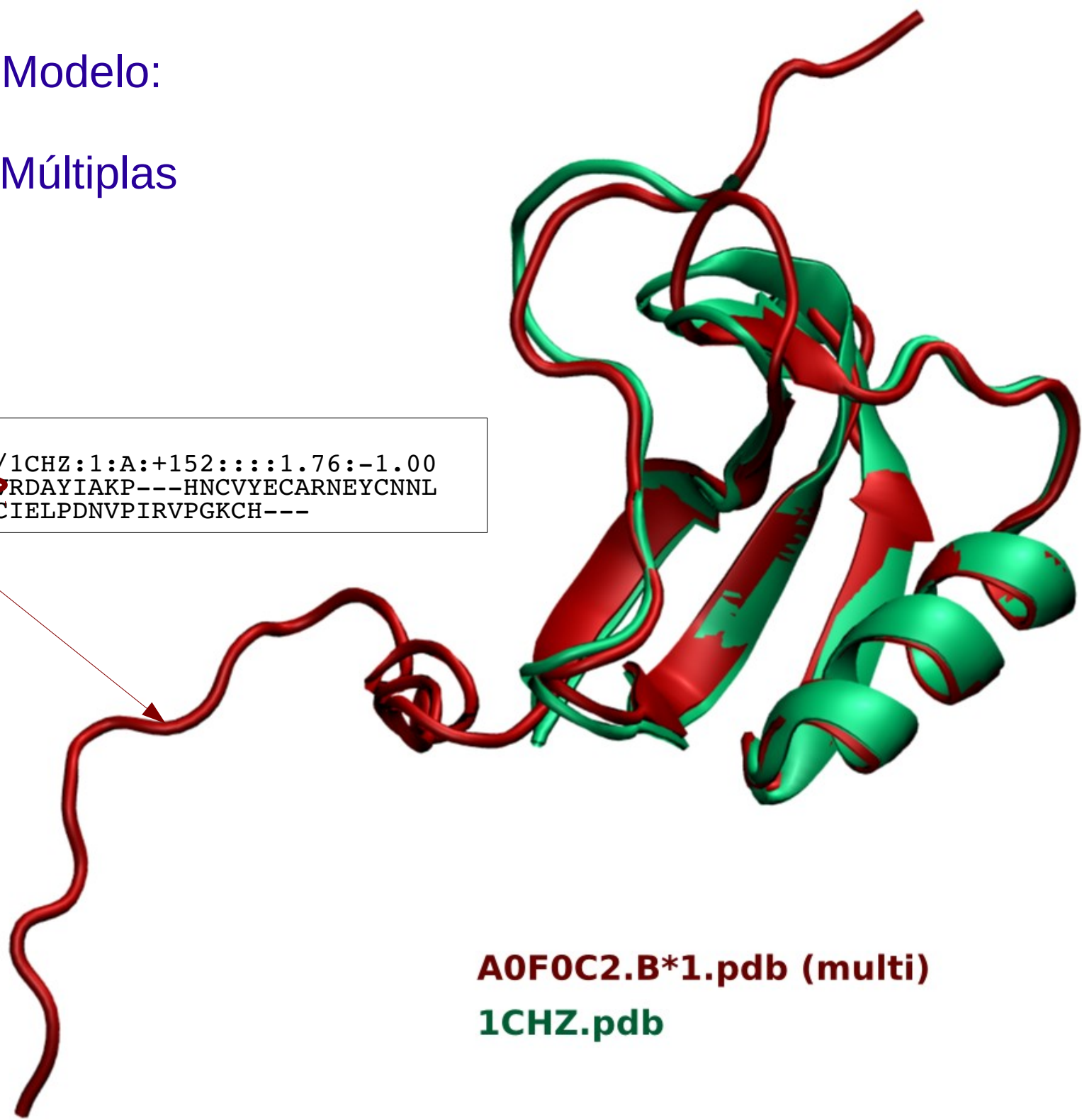
A0F0C2.B*1.pdb



3. Construção do Modelo:

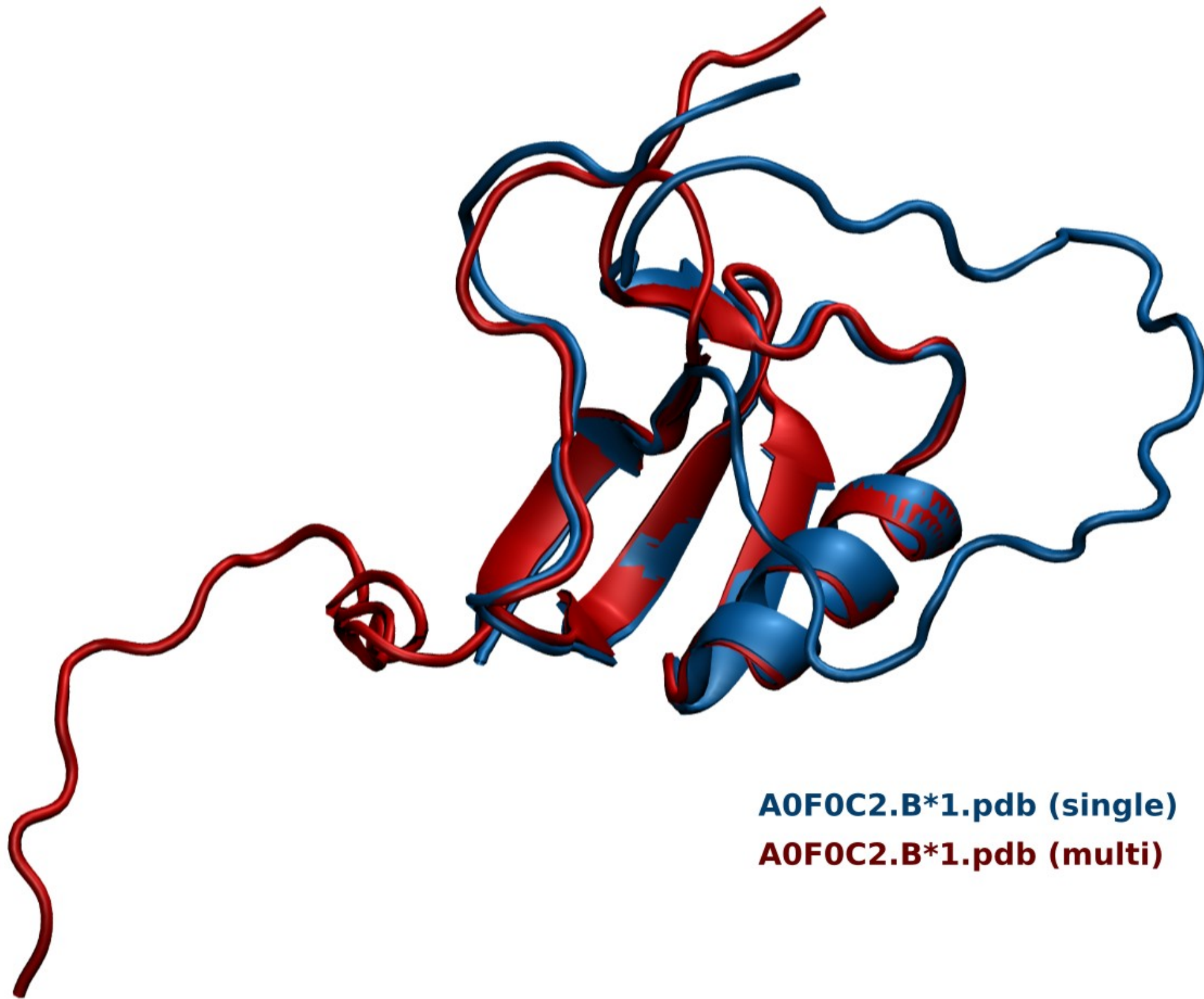
3.2. Baseado em Múltiplas Referências

```
>P1;./sequences/1CHZ  
structureX:./sequences/1CHZ:1:A:+152::::1.76:-1.00  
-----RDAYIAKP---HNCVYECARNEYCNNL  
CTKNGAKSGYCQWSGKYNGCWCIELPDNVPIRVP GKCH---
```



A0F0C2.B*1.pdb (multi)

1CHZ.pdb



A0F0C2.B*1.pdb (single)

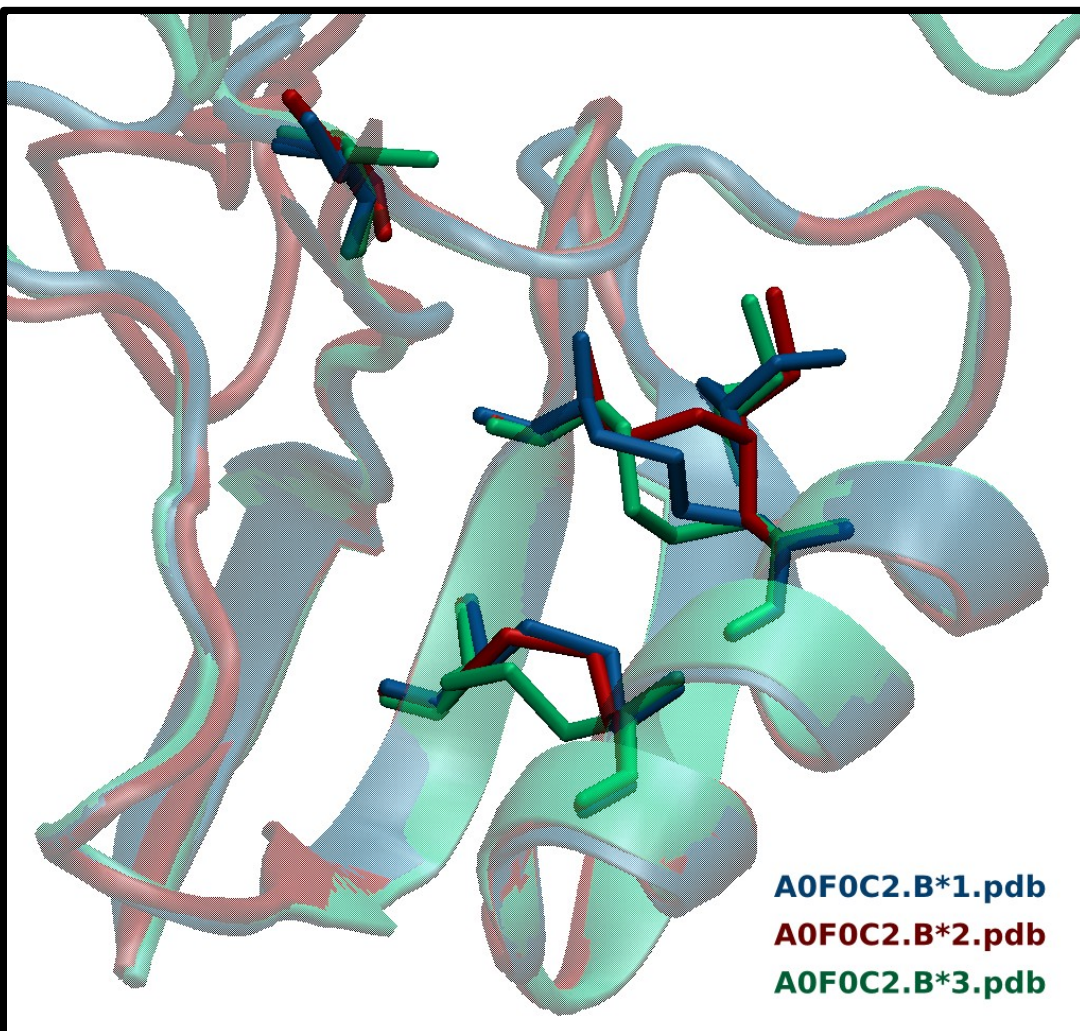
A0F0C2.B*1.pdb (multi)

Modelagem Comparativa

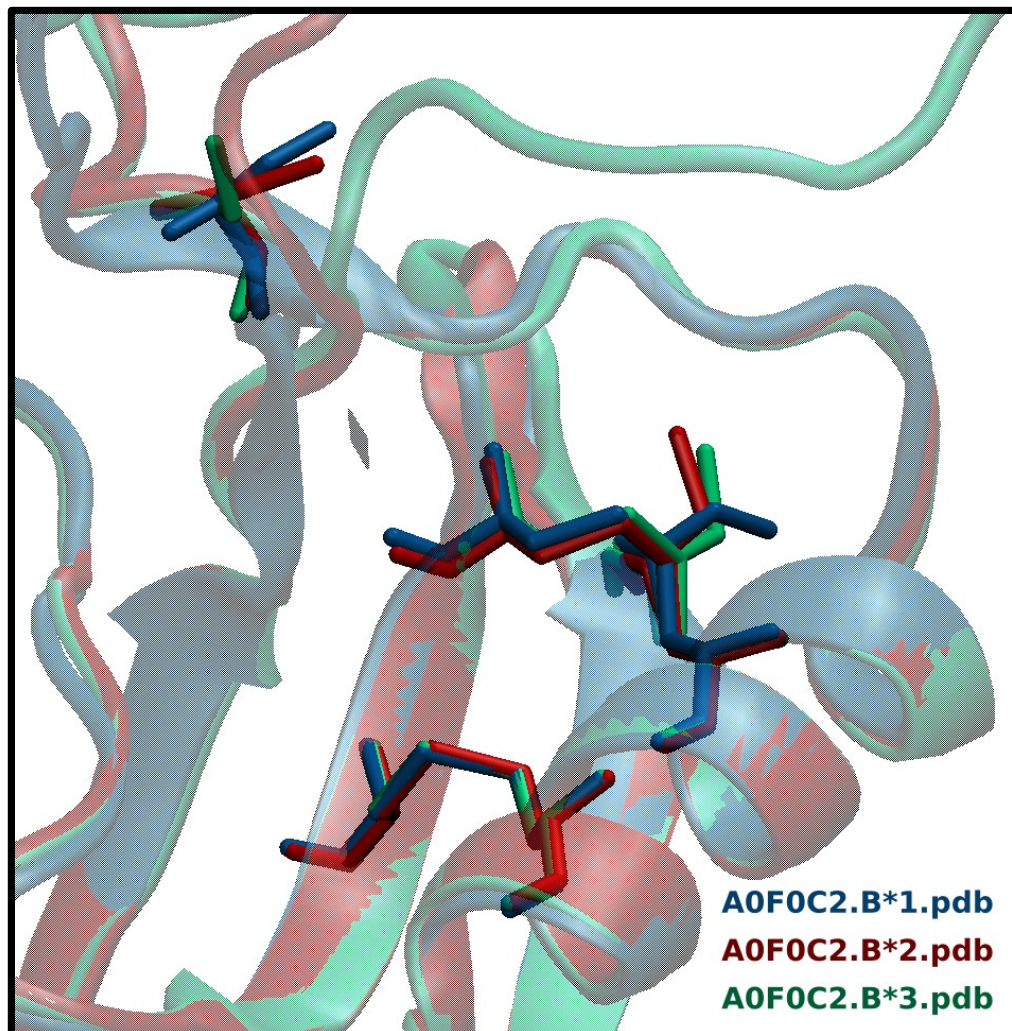
3. Construção do Modelo:

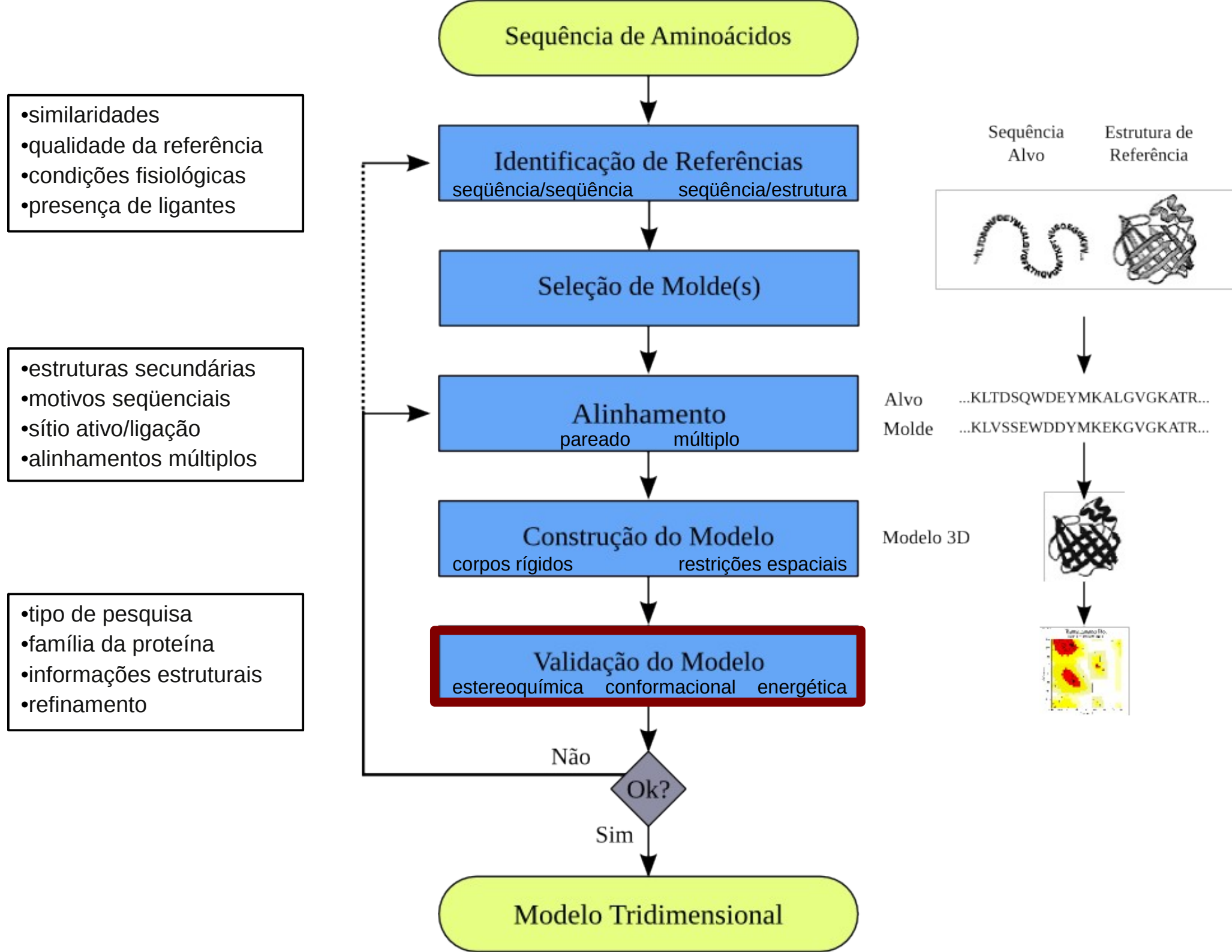
3.3. Construção de Modelos Otimizados:

Modelos sem Otimização



Modelos com Otimização





Modelagem Comparativa

4. Validação do Modelo:

1)Qualidade Estereoquímica



e.g. PROCHECK

2)Qualidade do Modelo



e.g. WHATCHECK

3)Energia de Interação entre os resíduos



e.g. PROSAIL

4)Análise individual de cada resíduo:

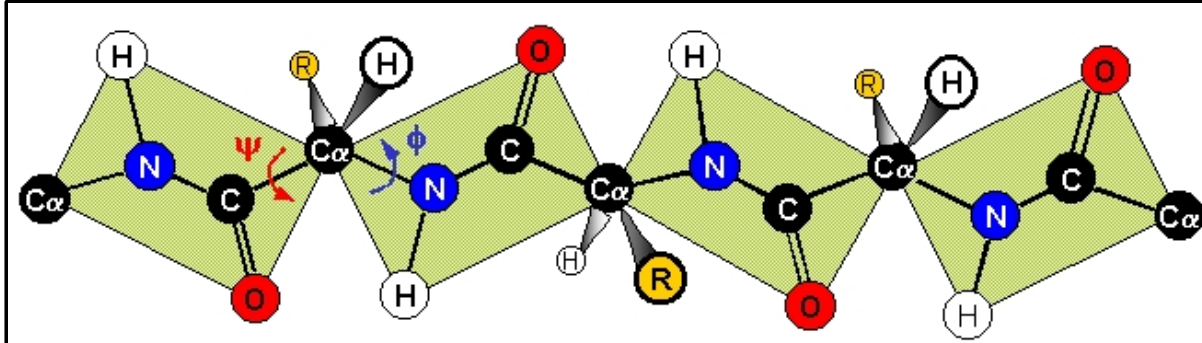
a)Natureza da Estrutura Secundária

b)Grau de exposição ao solvente

c)Hidrofobicidade do resíduo

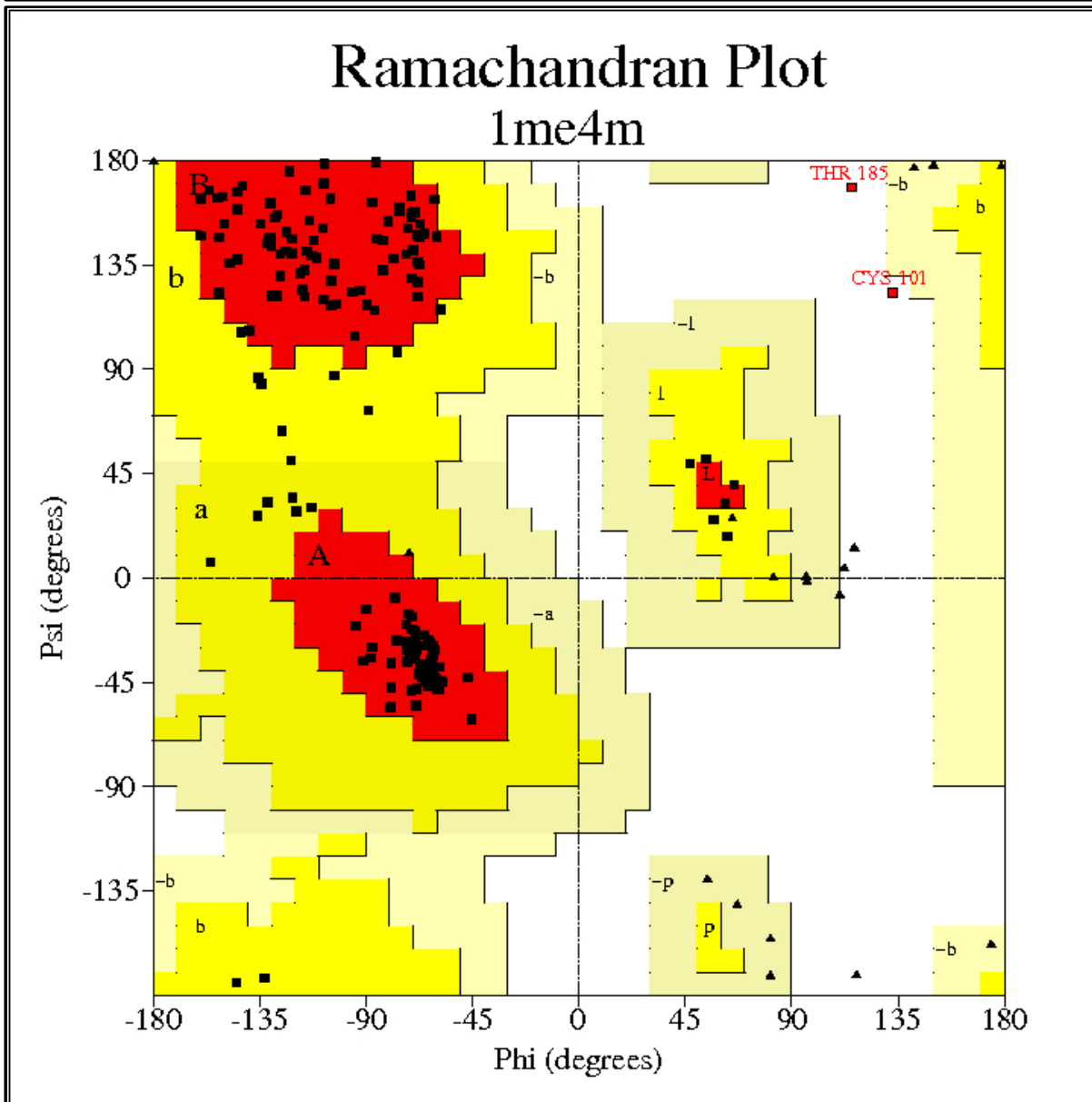
ATENÇÃO: A qualidade do modelo depende da qualidade das estruturas escolhidas como referência (valor de RMSD).

- No gráfico de Ramachandran os resíduos de aminoácidos são representados por quadrados, sendo as glicinas representadas por triângulos.



- As áreas em

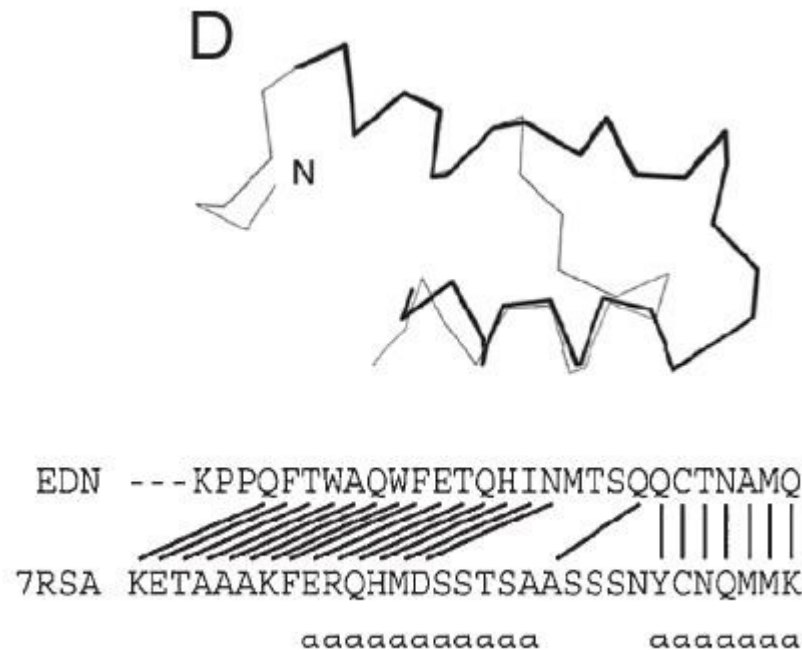
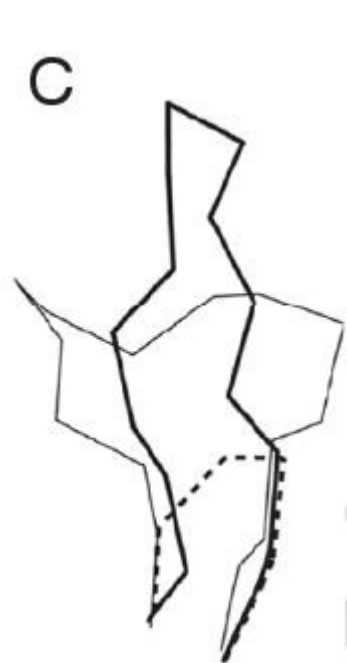
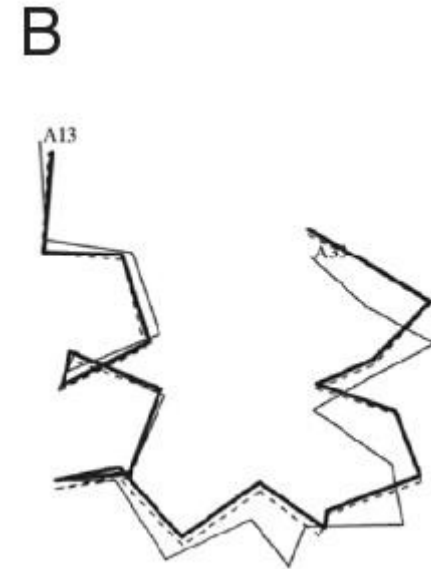
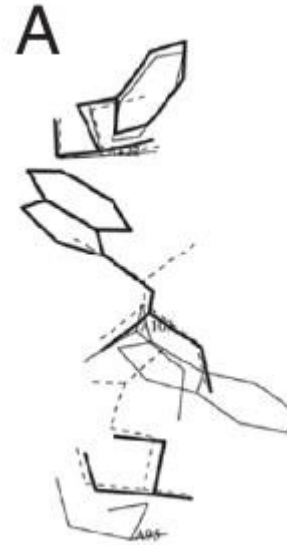
- Vermelho: regiões muito favoráveis
- Amarelo: regiões favoráveis
- Bege: pouco favoráveis
- Branco: desfavoráveis (válida para todos os aminoácidos, com exceção da glicina).



Modelagem Comparativa

Erros mais comuns nos modelos:

- A) Posicionamento da cadeia lateral.
- B) Distorções em regiões alinhadas.
- C) Regiões sem molde.
- D) Alinhamentos ineficientes.
- E) Escolha errada do molde.



Modelagem Comparativa

5. Refinamento do Modelo

1) Minimização de Energia:

- a) Gradiente Conjugado
- b) Máximo Declive
- c) (...)

2) Dinâmica Molecular:

- a) Minimização de Energia
- b) Cadeias Laterais
- c) Águas
- d) Enovelamento

ATENÇÃO: A aplicação do modelo gerado está diretamente relacionada com a sua qualidade.

predição *de novo*

Threading

Modelagem Comparativa

raios-X, RMN

Força
Similaridade Sequencial Insignificante

Porcentagem de Identidade Sequencial

50

100

PRECISÃO DO MODELO

4-8 Å²
~80aa

3,5 Å^o
80%

1,5 Å⁰
95%

1,0 Å^c
100%

E

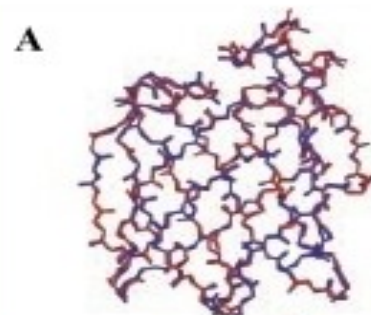
D



C

**B**

A



APLICAÇÕES

Estudios de mecanismos catalíticos

Desenho e melhoramento de ligantes

Docking de macromoléculas

Docking com ligantes pequenos

Predição de epítomos antigênicos

Suporte em estudos de mutagênese sítio-dirigida

Refinamento de estruturas resolvidas por RMN

Correlação funcional a partir de similaridade estrutural

Identificação de regiões conservadas em resíduos de superfície

Detecção de sítios funcionais por busca em bancos de motivos 3D

Modeller Online



ModWeb Server

• [Sali Lab Home](#) • [ModWeb](#) • [ModLoop](#) • [ModBase](#) • [IMP](#) • [ModPipe](#) • [LS-SNP](#) •

[Help](#) • [User Login](#) • [ModBase Datasets for User:Anonymous](#) • [Contact](#) • [News](#) • [Current ModWeb queue](#) • [ModWeb Home](#) •

News:

Additional functionality for registered users:

- Template based modeling
- Access to all user's ModWeb datasets
- Email notification for User's ModBase updates

Developers:

Eswar Narayanan
Ursula Pieper
Ben Webb

Acknowledgements:

David Eramian
Mallur S. Madhusudhan
Marc A. Marti-Renom
Min-Yi Shen
Andrei Gali

ModWeb: A Server for Protein Structure Modeling

Welcome to the new ModWeb ([old version](#))

General information ?

Calculate Models

Redefinir

Name

Email address

Modeller license key ?

(Not necessary for ModBase updates)

Dataset name (optional)

Availability ?

☒ Add to academic dataset

Input data ?

Input protein sequences ?

or upload sequences file ?
(FASTA Format)

Escolher arquivo

Nenhum arquivo selecionado

Calculate Models

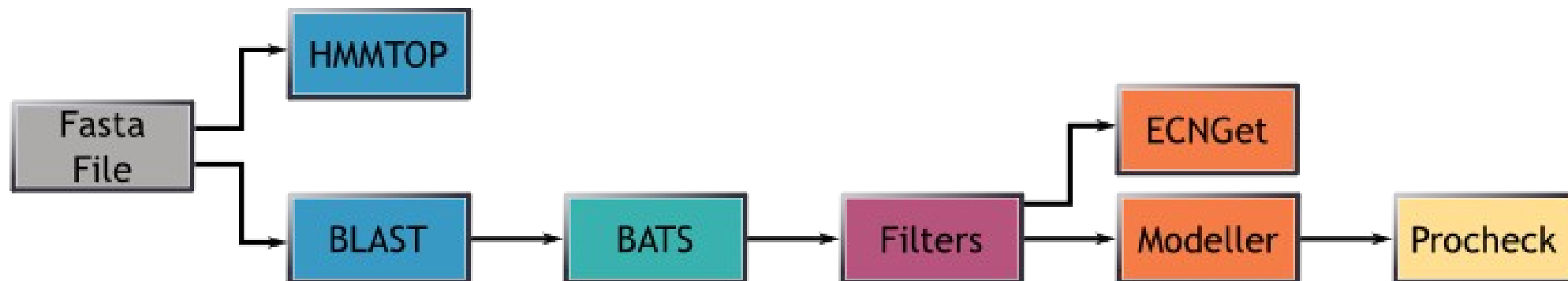
Redefinir

<https://modbase.compbio.ucsf.edu/scgi//modweb.cgi>

Modeller Online



www.mholline.lncc.br



Automatic 3D Comparative Modelling

MHOLLINE

[Home](#)
[New Job Submission](#)
[Manual](#)
[Team](#)
[Contact Us](#)

B Blast
 H Hmmtop
 B Bats
 F Filters
 E ECNGet
 M Modeller
 P Procheck
 S Summary

Name	Date	Status	Results	Summary	Download	Cancel	Delete
job1UBQ	15:31:29 - 2010-08-19	Finished	Output files B H B F E M P Log files B F M Fasta files B H B E M	S	Download	Cancel	Delete

Bioinformática + Biologia Computacional

Bioinformática

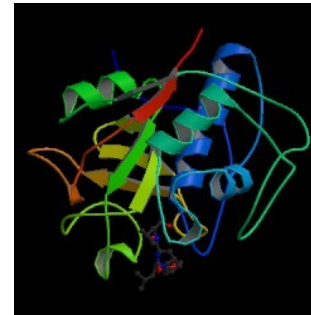
CLUSTAL X (1.83) MULTIPLE SEQUENCE ALIGNMENT
File: Tropinone_Reductase_Lps Date: Tue Oct 25 17:23:13 2005
Page 1 of 1

```
1ae1_A  MLPRTEINMLPRTEINMLPRTEINMLPRTEINMLPRTEIN  48
1ae1_B  MLPRTEINMLPRTEINMLPRTEINMLPRTEINMLPRTEIN  48
7000.m00007 TCTCTCTCTCTC  12
ruler  1.....10.....20.....30.....40.....
```

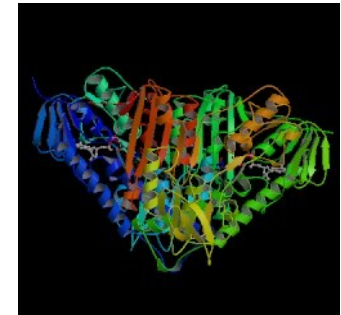


BLAST

Biologia Computacional



Cruzin



*Trypanothione
reductase*

Modelagem Comparativa

Possíveis

Alvos Moleculares



Docking Molecular

**Bancos
de
Ligantes**



2 ps



100 ps



200 ps

Simulação por Dinâmica Molecular

Simulação Computacional

Nome	Tipo	Endereço Eletrônico
------	------	---------------------

S:	servidor
P:	programa

Bancos de Dados

CATH	S	http://www.biochem.ucl.ac.uk/bsm/cath/
GenBank	S	http://www.ncbi.nlm.nih.gov/Genbank/
GeneCensus	S	http://bioinfo.mbb.yale.edu/genome/
MODBASE	S	http://www.salilab.org/modbase/
PALI	S	http://pauling.mbu.iisc.ernet.in/~pali/
PDB	S	http://www.rcsb.org/pdb/
PRESAGE	S	http://presage.berkeley.edu
SCOP	S	http://scop.mrc-lmb.cam.ac.uk/scop/
TrEMBL	S	http://srs.ebi.ac.uk

Busca por Molde

123D	S	http://123d.ncifcrf.gov/123D+.html
BLAST	S	http://www.ncbi.nlm.nih.gov/BLAST/
DALI	S	http://www2.ebi.ac.uk/dali/
FastA	S	http://www.ebi.ac.uk/fasta33/
MATCHMAKER	P	http://bioinformatics.burnham-inst.org
PHD, TOPITS	S	http://cubic.bioc.columbia.edu/predictprotein/
PROFIT	P	http://www.came.sbg.ac.at
THREADER	P	http://bioinf.cs.ucl.ac.uk/threader/threader.html
FRSVR	S	http://fold.doe-mbi.ucla.edu

Alinhamento de Sequência

BCM SERVER	S	http://searchlauncher.bcm.tmc.edu
BLAST2	S	http://www.ncbi.nlm.nih.gov/gorf/bl2.html
BLOCK MAKER	S	http://blocks.fhcrc.org/blocks/blockmkr/make_blocks.html
CLUSTAL	S	http://www2.ebi.ac.uk/clustalw/
FASTA3	S	http://www2.ebi.ac.uk/fasta3/
MULTALIN	S	http://pbil.ibcp.fr

Nome	Tipo	Endereço Eletrônico
------	------	---------------------

S: servidor
P: programa

Construção de Modelos

COMPOSER	P	http://www.tripos.com/
CONGEN	P	http://www.congenomics.com/congen/congen.html
ICM	P	http://www.molsoft.com
InsightII	P	http://www.accelrys.com
MODELLER	P	http://www.salilab.org/modeller/
QUANTA	P	http://www.accelrys.com
SYBYL	P	http://www.tripos.com
SCWRL	P	http://www.fccc.edu/research/labs/dunbrack/scwrl/
SNPWEB	S	http://salilab.org/snpweb-cgi/snpweb.cgi
SWISS-MOD	S	http://www.expasy.org/swissmod/SWISS-MODEL.html
WHAT IF	P	http://www.cmbi.kun.nl/whatif/

Avaliação de Modelos

ANOLEA	S	http://protein.bio.puc.cl/cardex/servers
AQUA	P	http://urchin.bmr.b.wisc.edu/~jorgen/aqua/
BIOTECHg	S	http://biotech.embl-heidelberg.de:8400
ERRAT	S	http://www.doe-mbi.ucla.edu/Services/ERRAT/
PROCHECK	P	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
ProsaII	P	http://www.came.sbg.ac.at
PROVE	S	http://www.ucmb.ulb.ac.be/UCMB/PROVE
SQUID	P	http://www.ysbl.york.ac.uk/~oldfield/squid/
VERIFY3D	S	http://www.doe-mbi.ucla.edu/Services/Verify_3D/
WHATCHECK	P	http://www.sander.embl-heidelberg.de/whatcheck/

Métodos de Avaliação

CASP	S	http://predictioncenter.llnl.gov
CAFASP	S	http://cafasp.bioinfo.pl
EVA	S	http://cubic.bioc.columbia.edu/eva/
LiveBench	S	http://bioinfo.pl/LiveBench/