



Métodos Avançados para Predição de Estruturas de Proteínas

Priscila Vanessa Zabala Capriles Goliatt

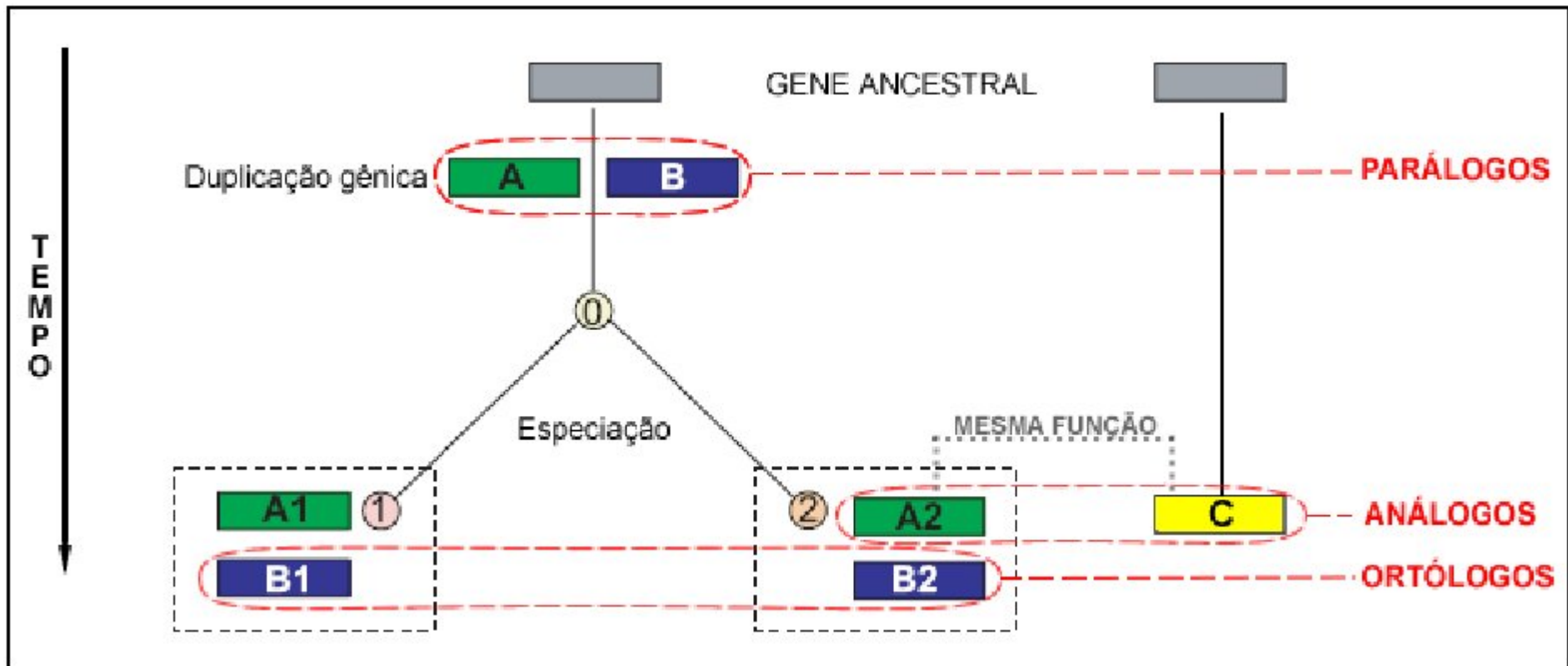
Universidade Federal de Juiz de Fora

(priscilacapriles@gmail.com)

Por que estudar proteínas?

- Obter informações a cerca do papel de cada proteína em um dado organismo.
- Identificar quais proteínas desempenham papéis biológicos semelhantes em organismos distintos:
 - Identificar proteínas homólogas
 - Identificar proteínas análogas
- Identificar proteínas intrínsecas a um dado organismo.
- Buscar por novos alvos moleculares para o tratamento de doenças.
- Auxiliar no desenho racional de fármacos baseado em estrutura.
- Análise filogenética.

Homologia X Analogia



Esquema representativo dos relacionamentos entre ortólogos, parálogos e análogos¹ (Adaptado de Koonin, 2001).

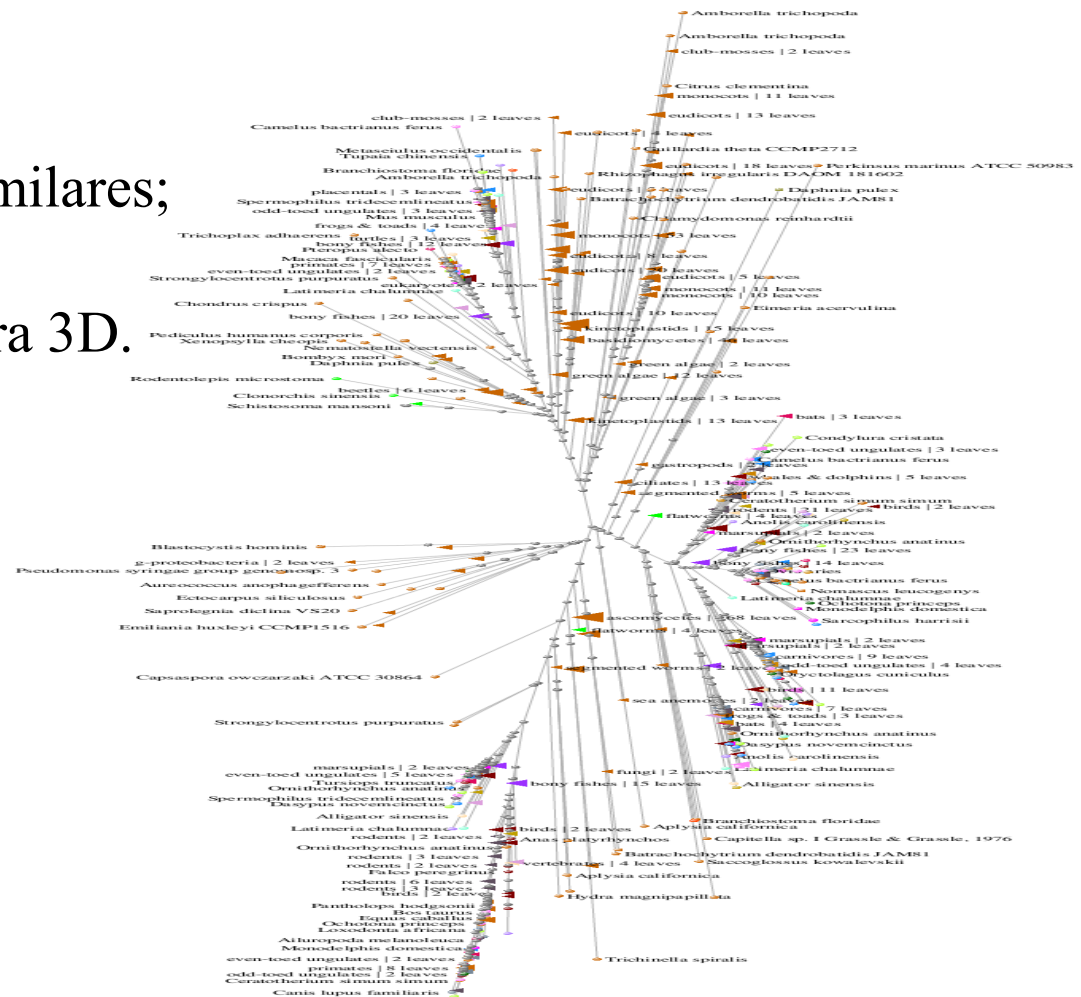
A partir da duplicação gênica de um gene ancestral (em cinza) surgiram os genes A (verde) e B (azul) em uma espécie, sendo denominados **parálogos**. Depois de um evento de especiação, surgem os genes A1, B1 e A2, B2 em 2 espécies diferentes; esses genes são denominados **ortólogos**. Genes que codificam para proteínas que possuem a mesma função, mas não possuem a mesma origem evolutiva, são denominados **análogos** [genes A2 e C (amarelo)].

Similaridade X Função

Lembrando que ...

A alta taxa de similaridade entre duas (ou mais) sequências, sugere:

- possível homologia;
- que podem desempenhar funções similares;
- que podem possuir a mesma estrutura 3D.

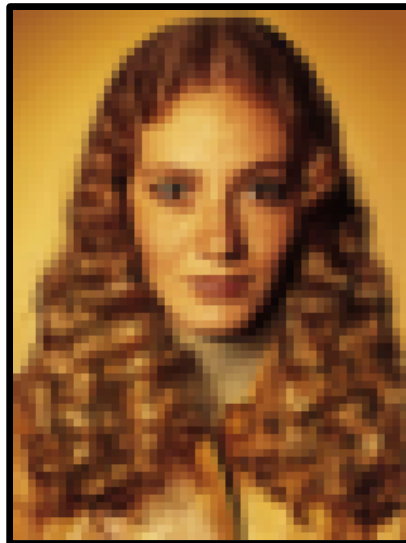
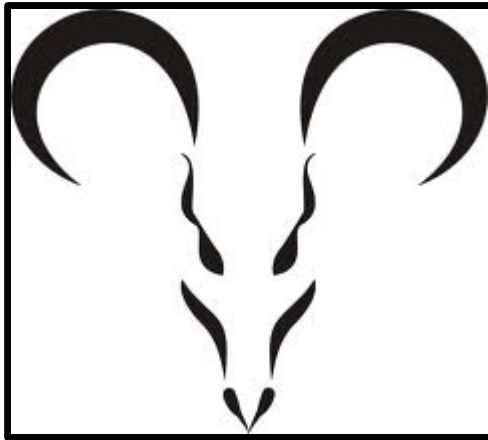


Similaridade X Função

Só que ...

Alvo

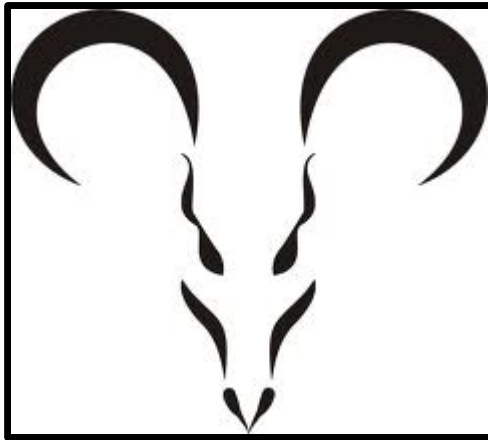
Mais semelhante



Similaridade X Função

Só que ...

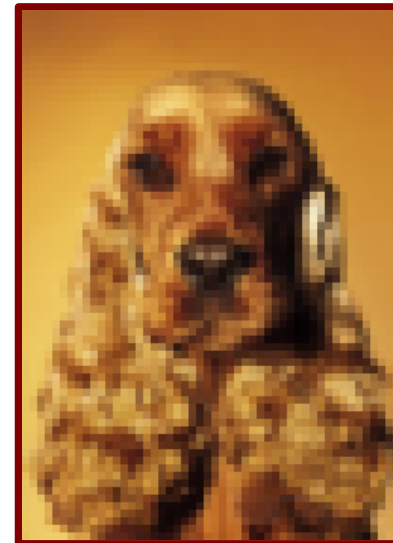
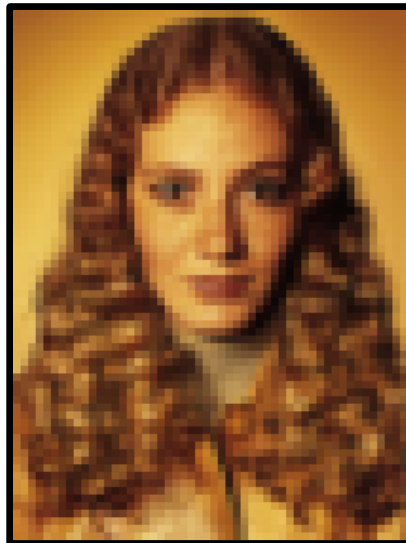
Alvo



Mais semelhante



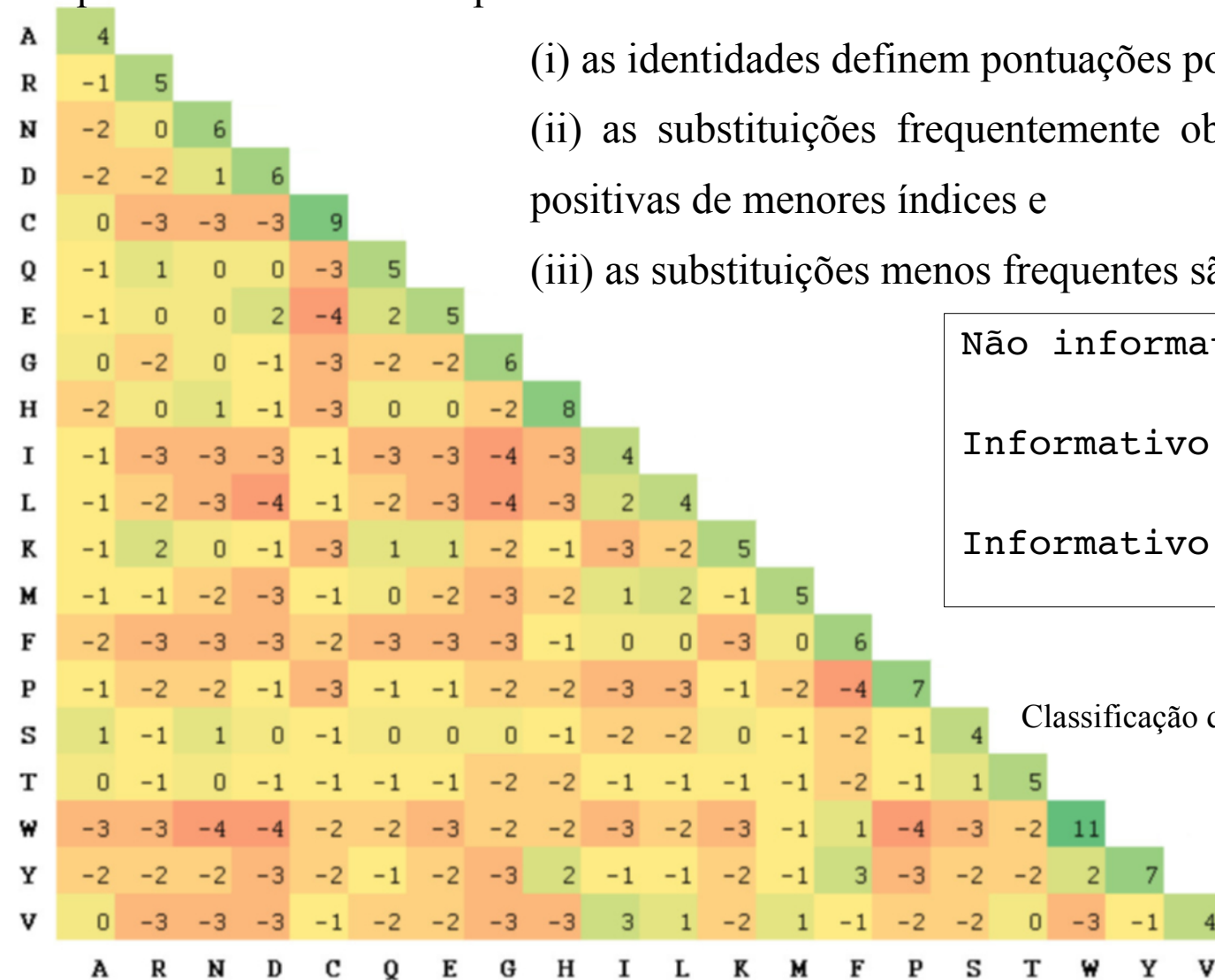
Ok



Xiiiiii !!!

- (i) as identidades definem pontuações positivas de maiores índices,
- (ii) as substituições frequentemente observadas são dadas por pontuações positivas de menores índices e
- (iii) as substituições menos frequentes são pontuadas negativamente.

Classificação dos tipos de alinhamentos entre seqüências.



Alinhamento Local

✓ 1chz     A NEW NEUROTOXIN FROM BUTHUS MARTENSII KARSCH

Chains A
E Value 2.4E-2

Characteristics **Release Date:** 31-Mar-2000 **Exp. Method:** X Ray Diffraction
Resolution: 1.76 Å

Classification **Toxin**

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

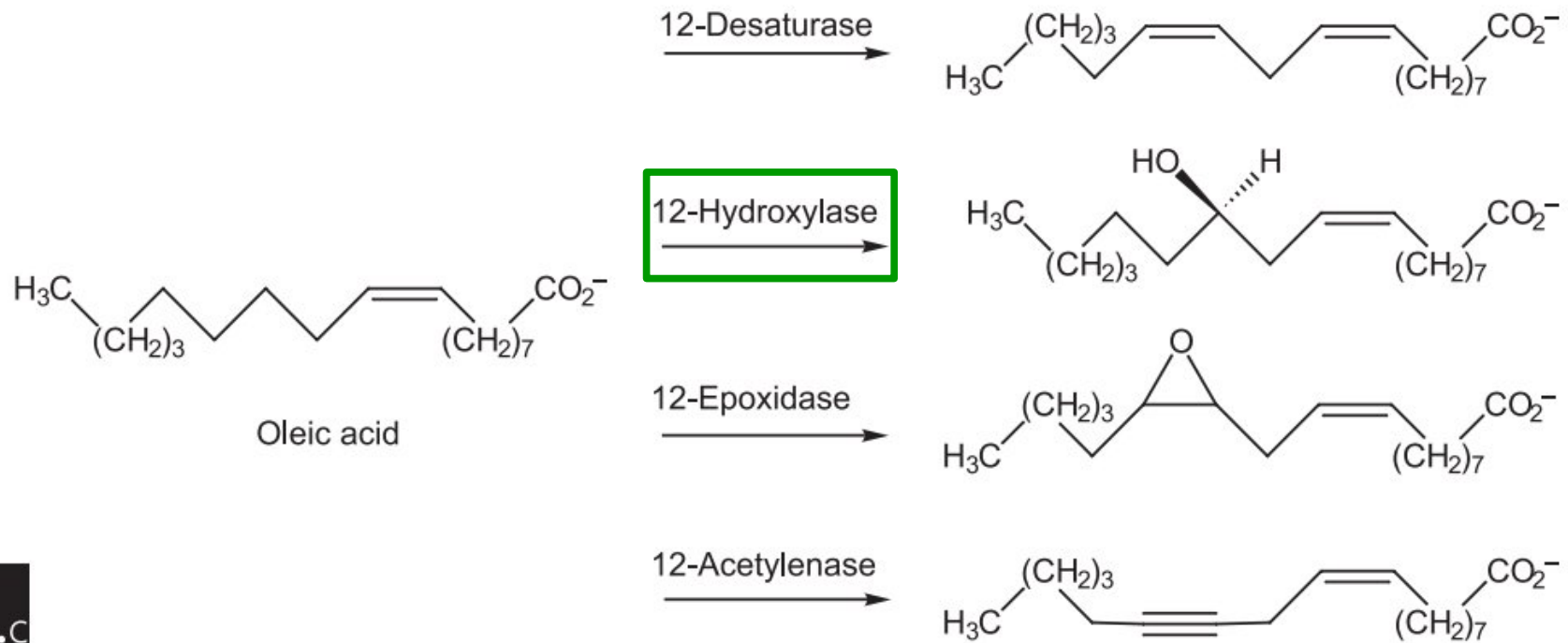
Authors **He, X.L., Deng, J.P., Li, H.M., Wang, D.C.**

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 AENLGEDCENLCKQKATDGF	CRQP	-----	HCFC	TDMPDYATR
	CVY	E C NLC + A G+C+		C+C ++PD+	R
Sbjct	CVYECARNEYCNLCTKNGAKSGYCQWSGKYGN	GCWC	IELPDN	VP	IR

Identidade X Função

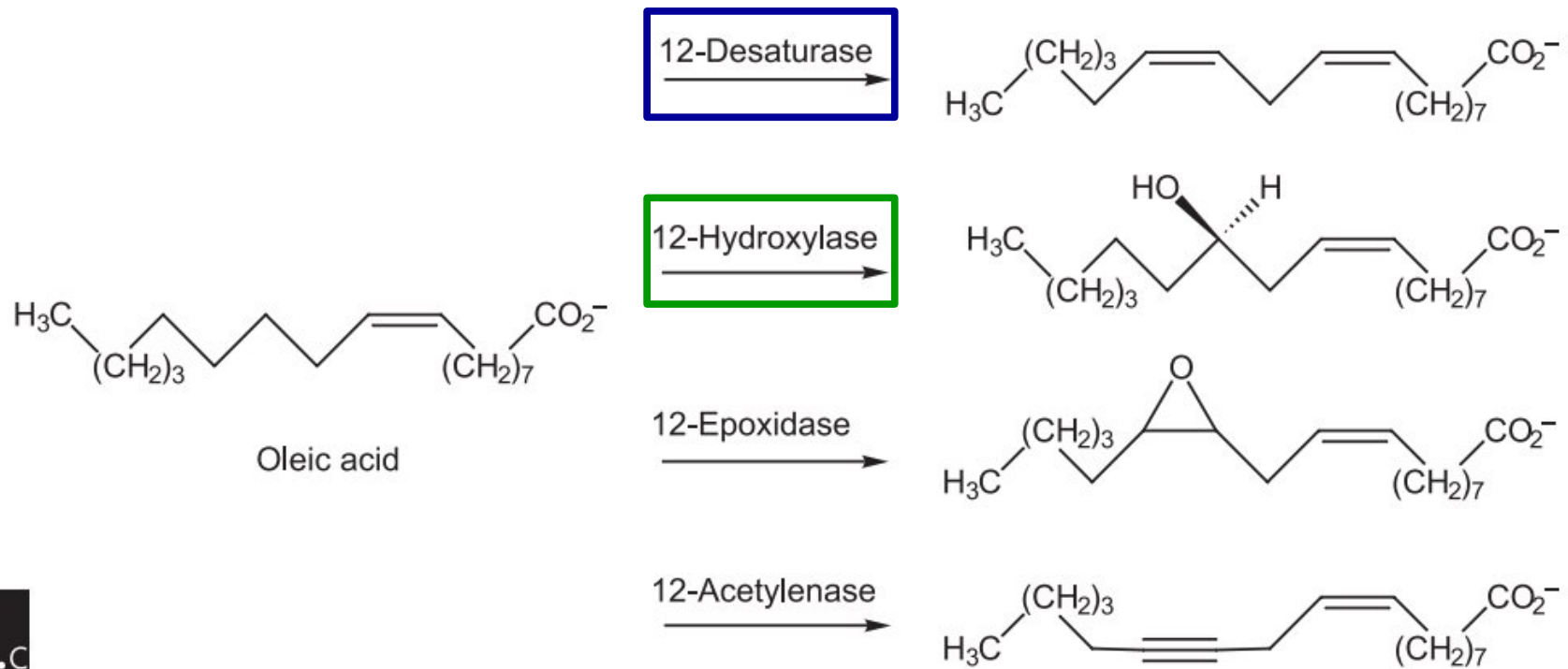


Uma enzima **oleato hidroxilase** de (A) *Lesquerella fendleri*

(A)



Identidade X Função



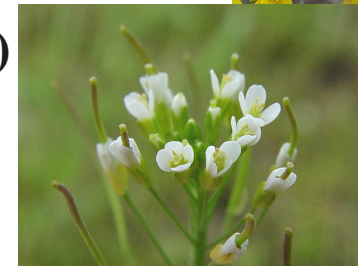
Uma enzima **oleato hidroxilase** de (A) *Lesquerella fendleri* possui 81% de identidade em relação à uma **oleato desaturase** de (B) *Arabidopsis thaliana*,

(A)

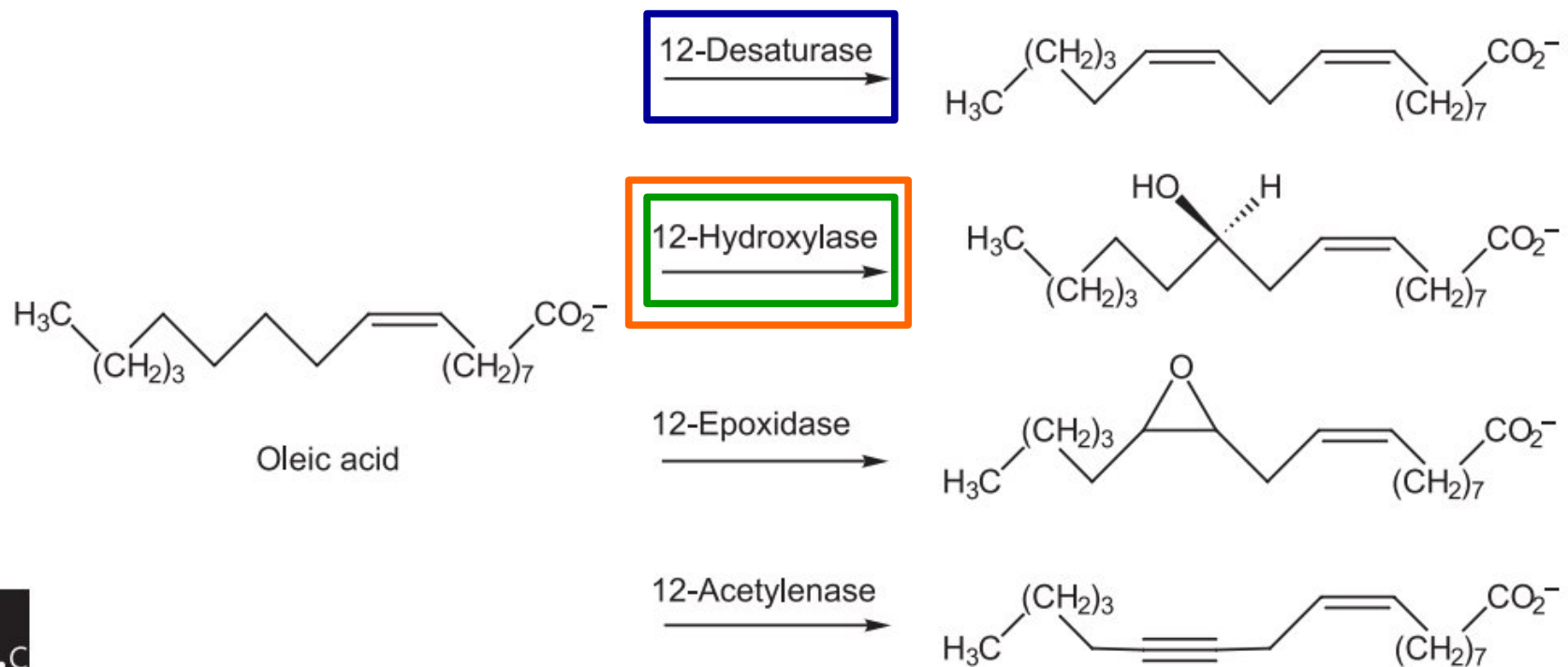


81%

(B)



Identidade X Função



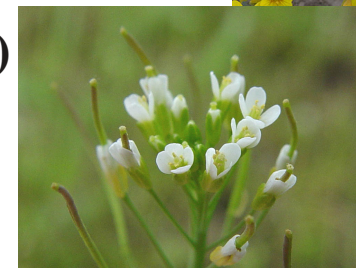
Uma enzima **oleato hidroxilase** de (A) *Lesquerella fendleri* possui 81% de identidade em relação à uma **oleato desaturase** de (B) *Arabidopsis thaliana*, entretanto, possui 71% de identidade em relação à sequência de aminoácidos de uma **oleato hidroxilase** de (C) *Ricinus communis*.

(A)



81%

(B)

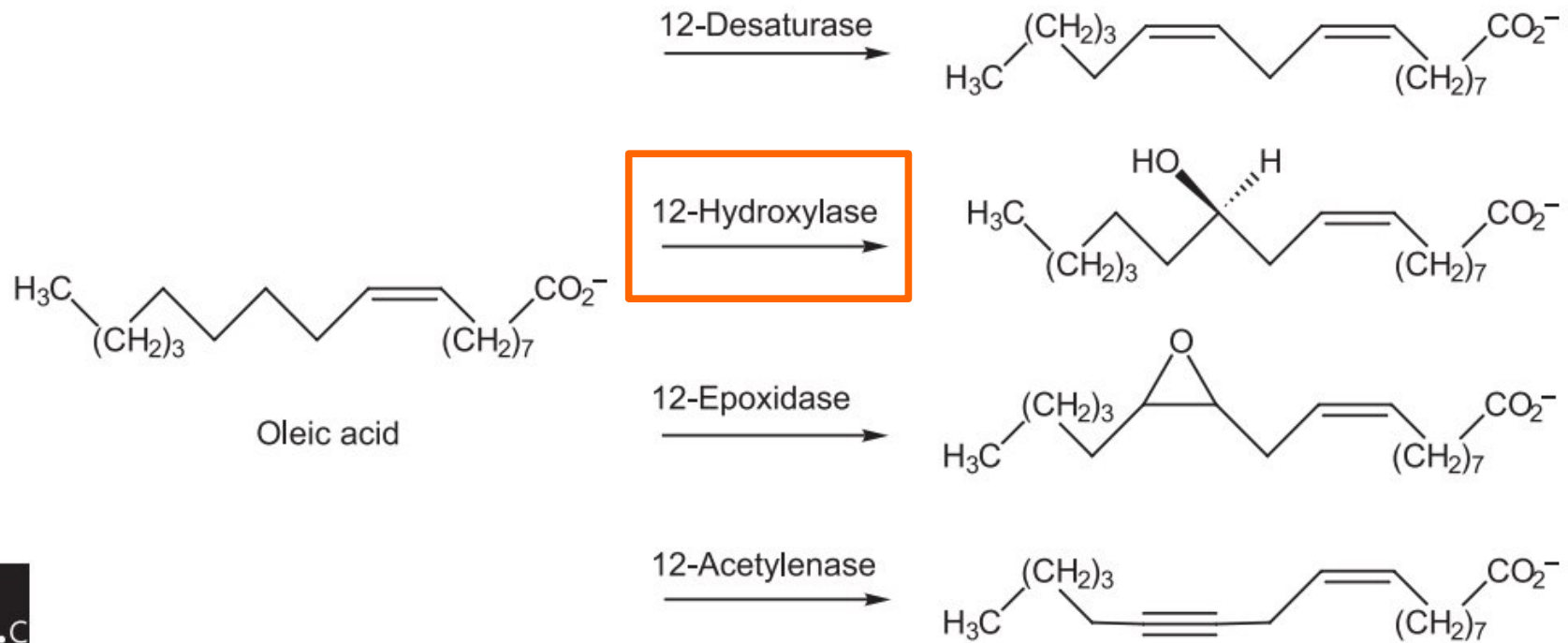


(C)

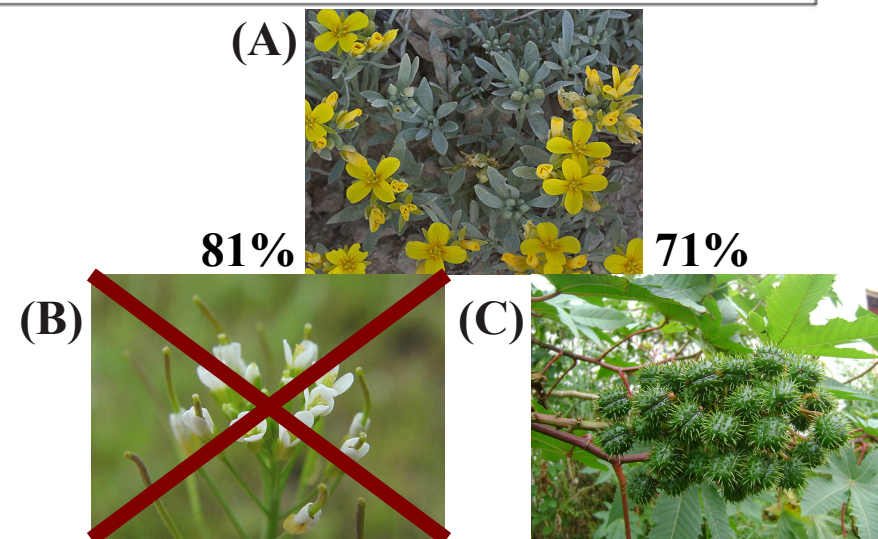


71%

Identidade X Função

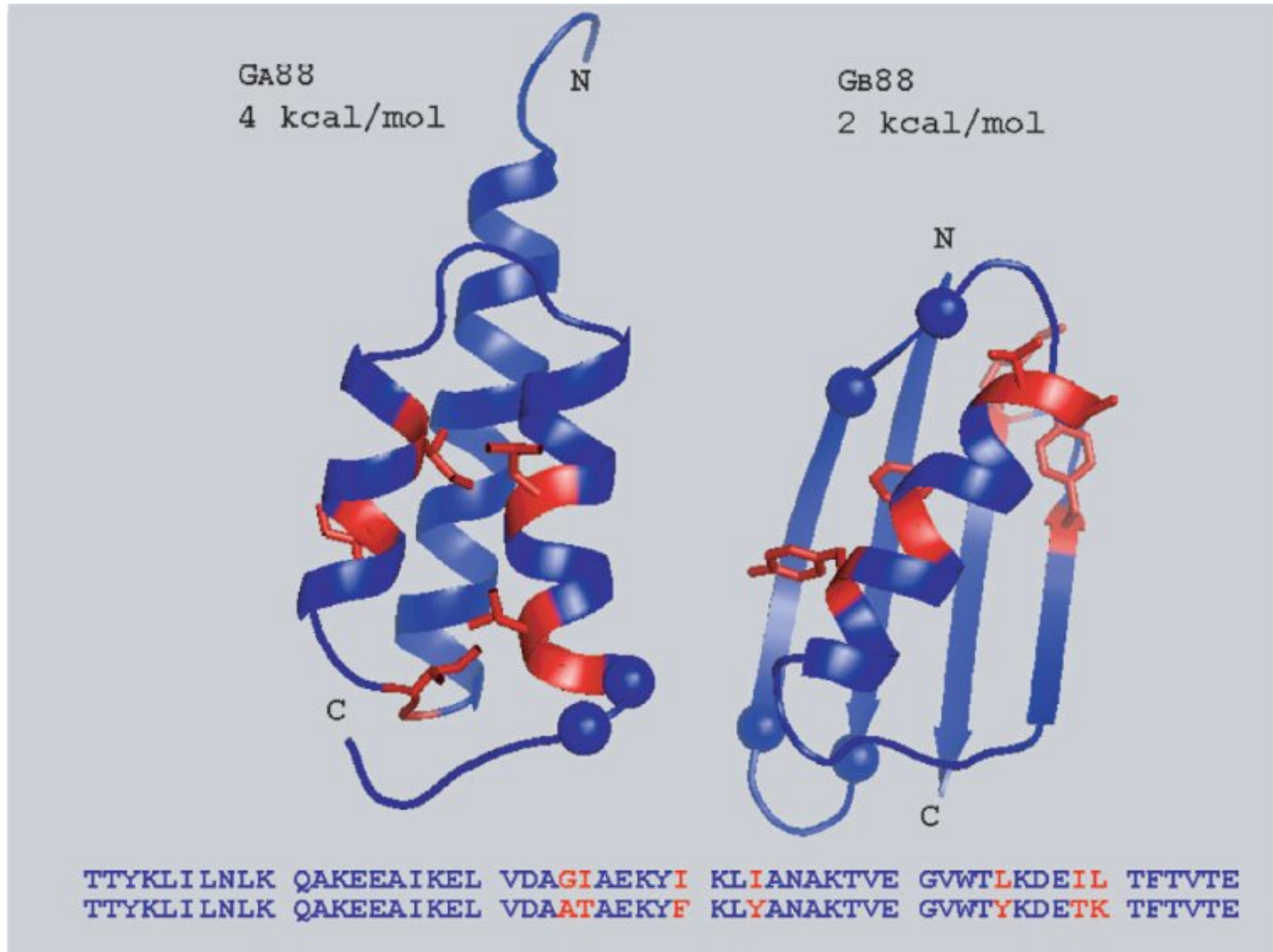


Uma enzima **oleato hidroxilase** de (A) *Lesquerella fendleri* possui 81% de identidade em relação à uma **oleato desaturase** de (B) *Arabidopsis thaliana*, entretanto, possui 71% de identidade em relação à sequência de aminoácidos de uma **oleato hidroxilase** de (C) *Ricinus communis*.



Identidade X Função

Proteínas com alta taxa de identidade na sequência de aminoácidos (>80%) podem possuir estruturas e funções diferentes.



Fonte: Retirado de Alexander, P. A., et al (2007). *The design and characterization of two proteins with 88% sequence identity but different structure and function.* **PNAS**, 104(29):11963-11968.

Por que prever teoricamente?

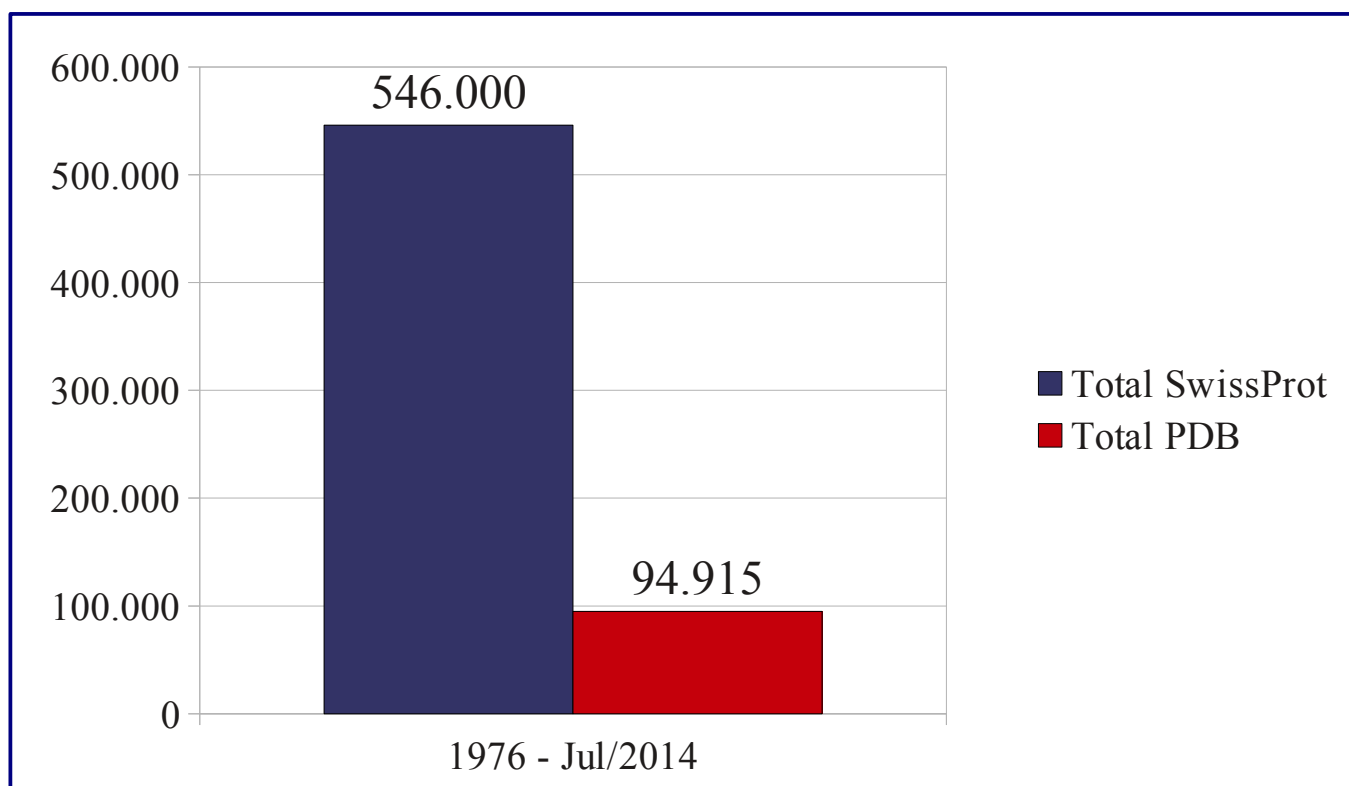
Aumento do número de projetos de sequenciamento de genomas

+

Limitações da predição experimental de estruturas protéicas

=

Aumento das técnicas de predição teórica



Fonte: UniProt: ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/relnotes.txt

PDB: <http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=molType-protein&seqid=100>

Predição Teórica: Metodologias

- Limites dos Métodos de Predição de Estrutura 3D:

Métodos

Identidade na sequência de AA

Modelagem comparativa

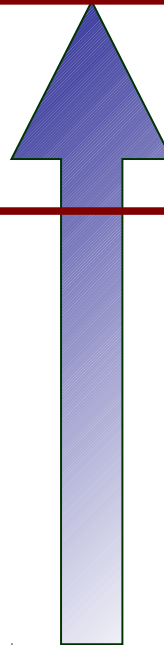
>25%

Threading

20-30%

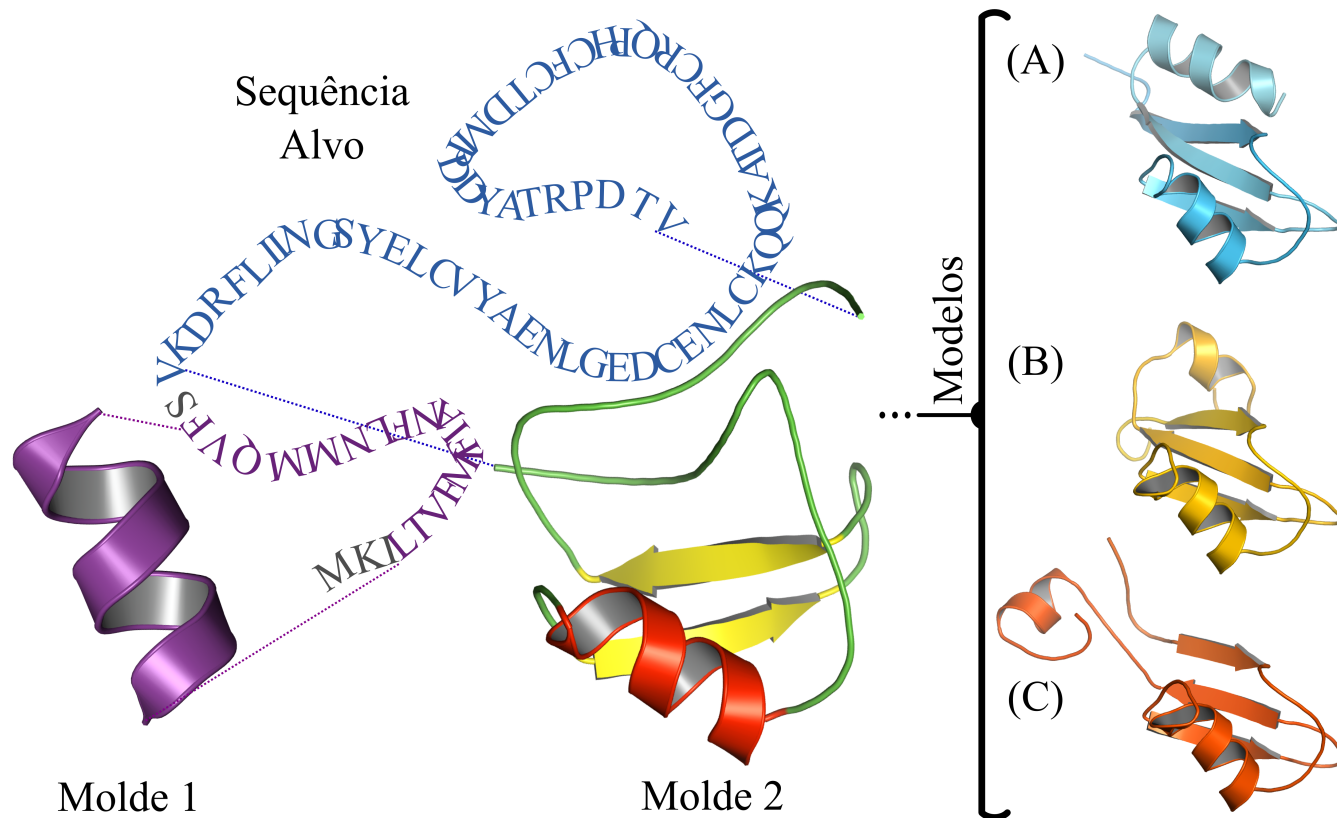
ab initio

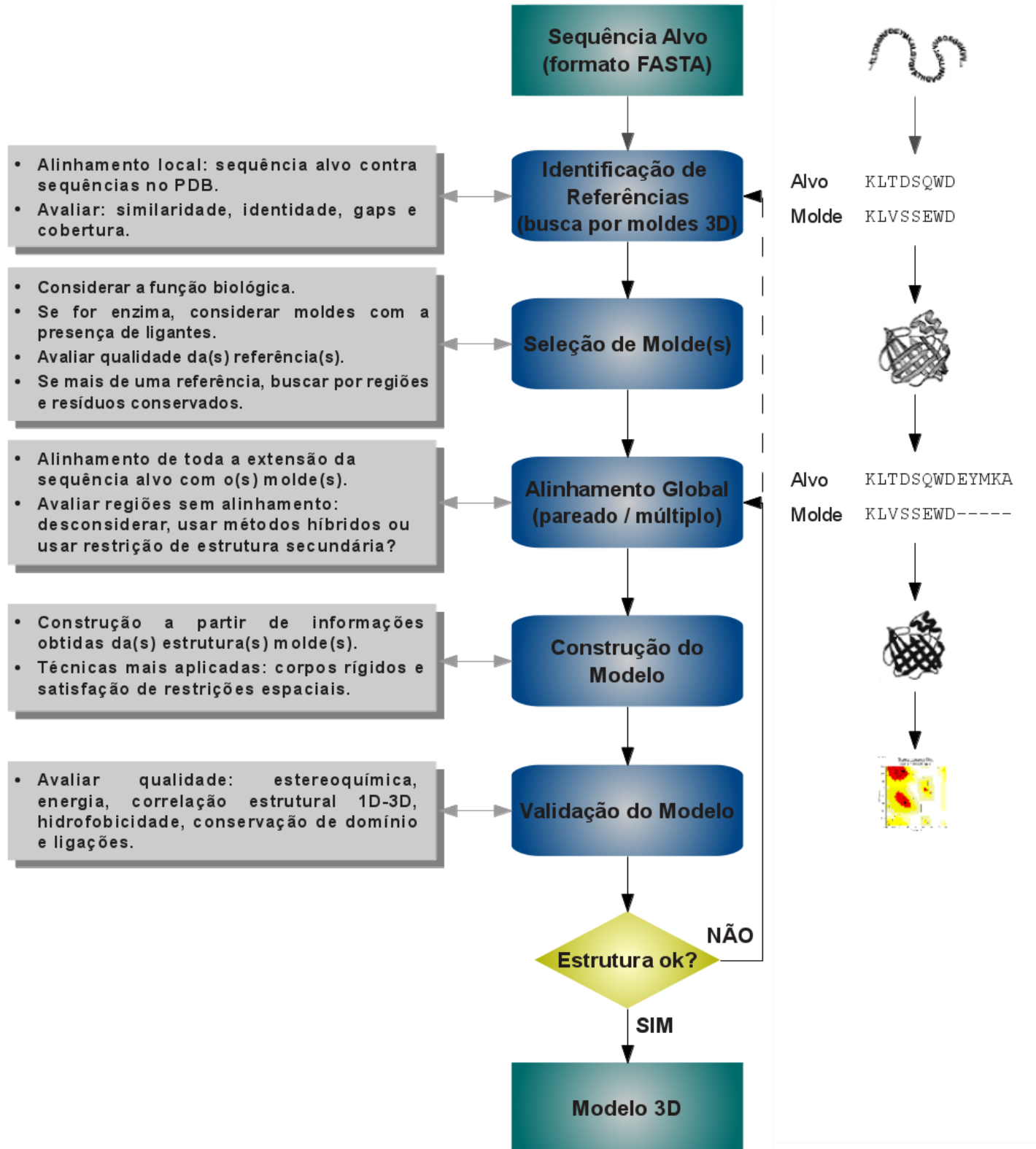
0-20%



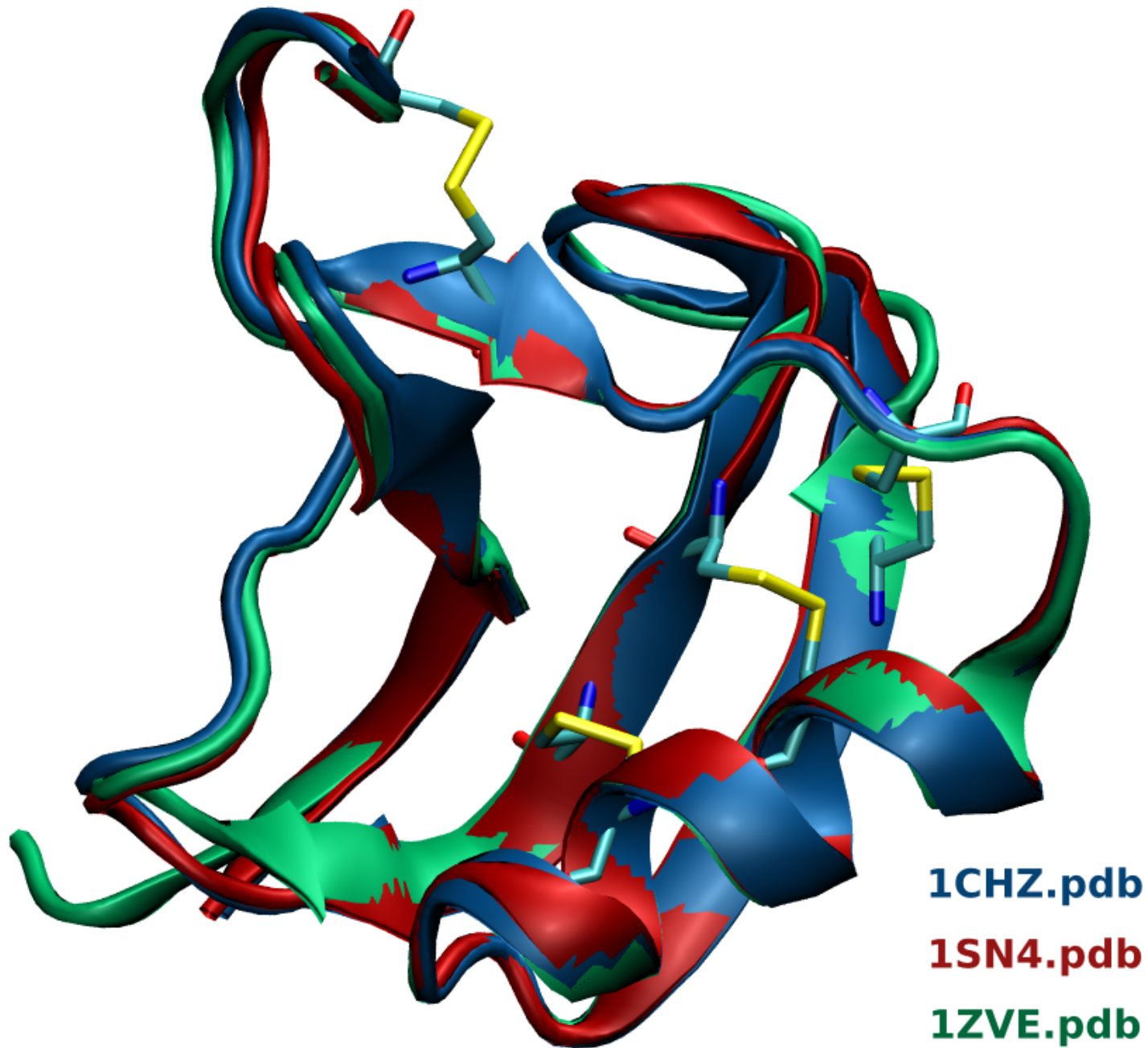
Modelagem Comparativa: Introdução

- A metodologia se baseia no fato de que duas proteínas homólogas 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.





Modelagem Comparativa: Múltiplas Estruturas



Modelagem Comparativa: Alinhamento Estrutural

•O grau de similaridade entre duas estruturas 3D é determinado pelo desvio médio quadrático (*RMSD*, do inglês *Root Mean Square Deviation*). O *RMSD* mede as distâncias interatômicas (em Angstroms), entre o modelo gerado e a estrutura molde e é dado pela seguinte expressão

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \|r_i^A - r_i^B\|^2}$$

onde N é o número total de átomos pareados entre as duas estruturas, r_i é o vetor posição do átomo i , A é a estrutura alvo (modelada) e B é a estrutura molde.

Modelagem Comparativa: Alinhamento Estrutural

- O grau de similaridade entre duas estruturas 3D é determinado pelo desvio médio quadrático (*RMSD*, do inglês *Root Mean Square Deviation*). O *RMSD* mede as distâncias interatômicas (em Angstroms), entre o modelo gerado e a estrutura molde e é dado pela seguinte expressão

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \|r_i^A - r_i^B\|^2}$$

onde N é o número total de átomos pareados entre as duas estruturas, r_i é o vetor posição do átomo i , A é a estrutura alvo (modelada) e B é a estrutura molde.

**Quanto menor o valor do RMSD,
maior a similaridade estrutural!!!**

Modelagem Comparativa: Alinhamento Global

- 1) Toma-se a sequencia com o maior número de resíduos.
- 2) Alinha a segunda sequencia 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.

- 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
KQQKATDGF CRQ-----PHCFCTDMPDDYATRPDTPVDPI-M-----
-----
-----*
```

Modelagem Comparativa: Alinhamento Global

- 1) Toma-se a sequencia com o maior número de resíduos.
- 2) Alinha a segunda sequencia 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.

- Alinhamento com Única Referência:

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

>P1; ./sequences/A0F0C2
sequence:tr|A0F0C2|A0F0C2 MESMA:      : :      : : ::-1.00:-1.00
MKILTVFMIFIANFLNMMQVFSVKDRFLIINGSYELCVYAENLGEDCENLC
KQQKATDGF CRQ-----PHCFCTDMPDDYATRPDTVDPI-M-----
-----
-----*
```

Modelagem Comparativa: Alinhamento Global

- 1) Toma-se a sequencia com o maior número de resíduos.
- 2) Alinha a segunda sequencia 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.

- Alinhamento com Única Referência:

```
>P1; ./sequences/1CHZ
structureY: ./sequences/1CHZ .1 :A:+152 : : : 1.76:-1.00
VRDAYIAK-----HNCVYECARNEYC>NNLC
TKNGAKSGYCQWSGRYCNCGWCIELPDNVPTRVPGKCH. /wwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*

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

- Alinhamento com Múltiplas Referências:

```
>P1;./sequences/1CHZ
structureX:./sequences/1CHZ:      1 :A:+152 : :: 1.76:-1.00
-----VRDAYIAKP---HNCVYECARNEYCNNL
CTKNGAKSGYCQWSGKYGNGCWCIELPDNVPPIRVPGKCH---./wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*
```

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

[illegible]

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

- 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*
```

- Alinhamento com Múltiplas Referências:

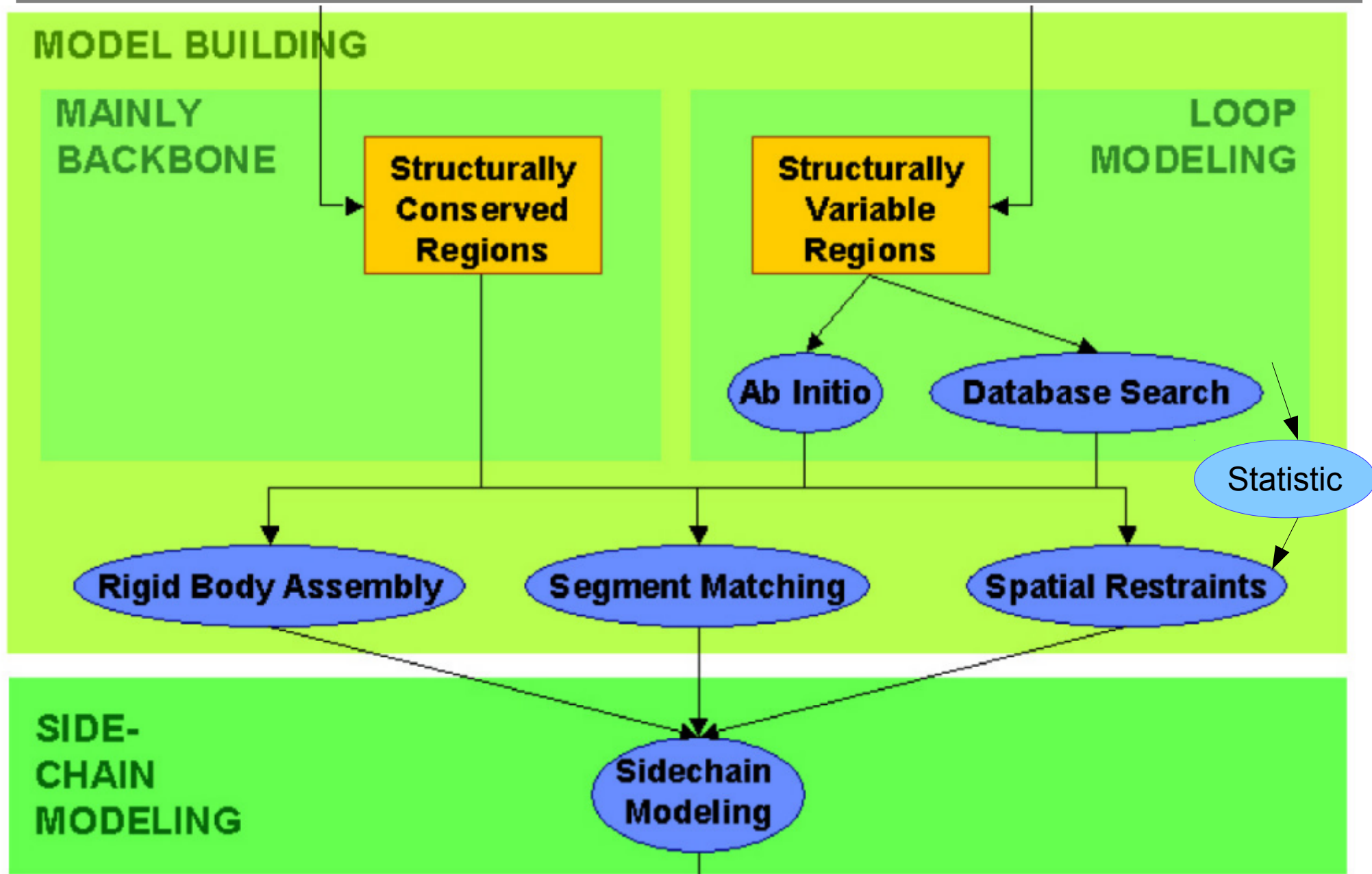
```
>P1; ./sequences/1CHZ
structureX: ./sequences/1CHZ:    1 :A:+152 : ::: 1.76:-1.00
-----VRDAYIAKP---HNCVYECARNEYC>NNL
CTKNGAKSGYCQWSGKYGNGCWCIELPDNVPIRVPGKCH---./wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*
```

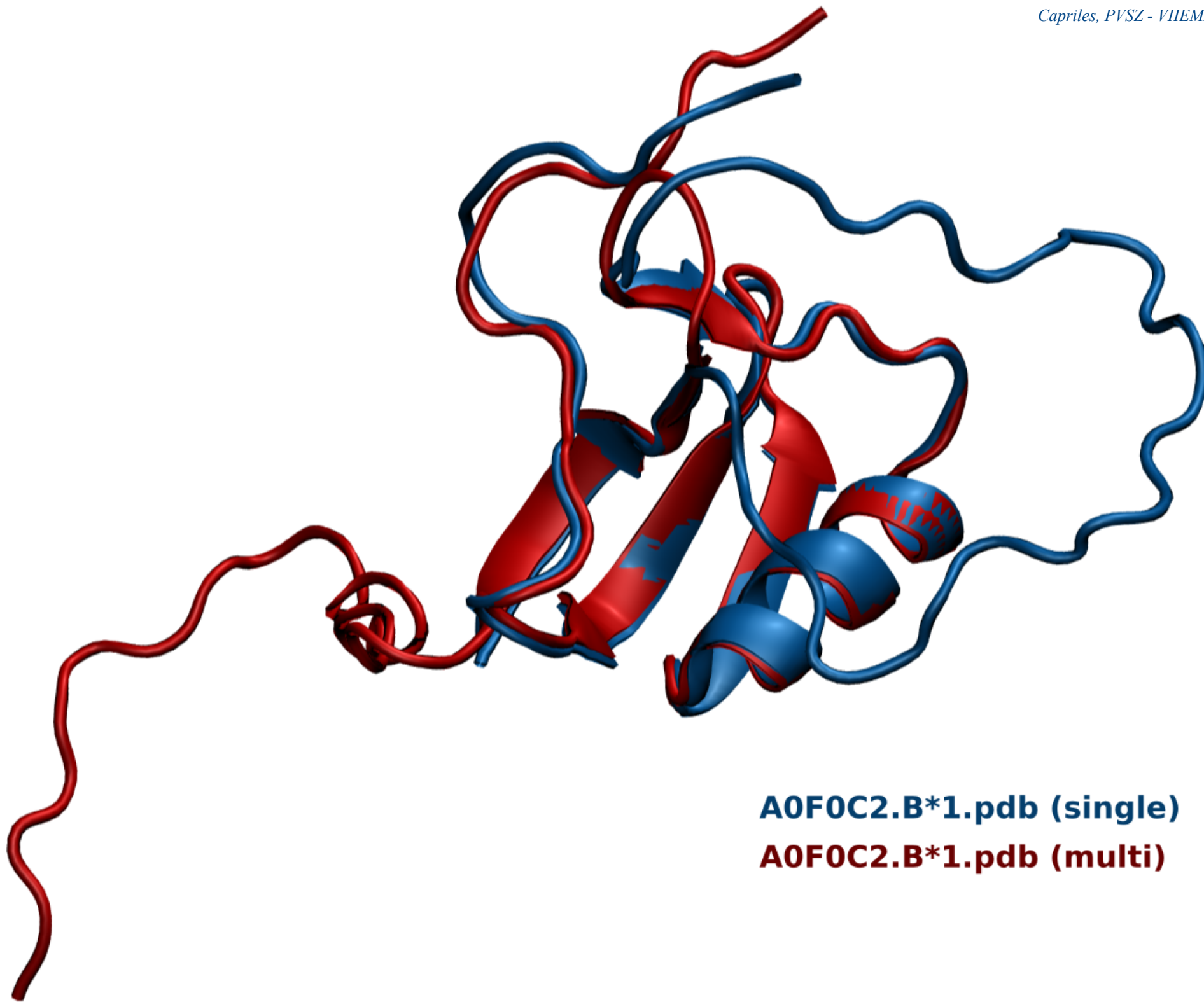
```
>P1; ./sequences/1ZVE
structureX: ./sequences/1ZVE:    1 :A:+169 : ::: 1.70:-1.00
-----NSVRDAYIAGP---HNCVYECARNEYCNDL
CTKNGAKSGYCQWVGKYGNGCWCIELPDNVPIRVPGKCH---/.wwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww
wwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwwww*
```

```
>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*
```

Modelagem Comparativa: Construção do Modelo

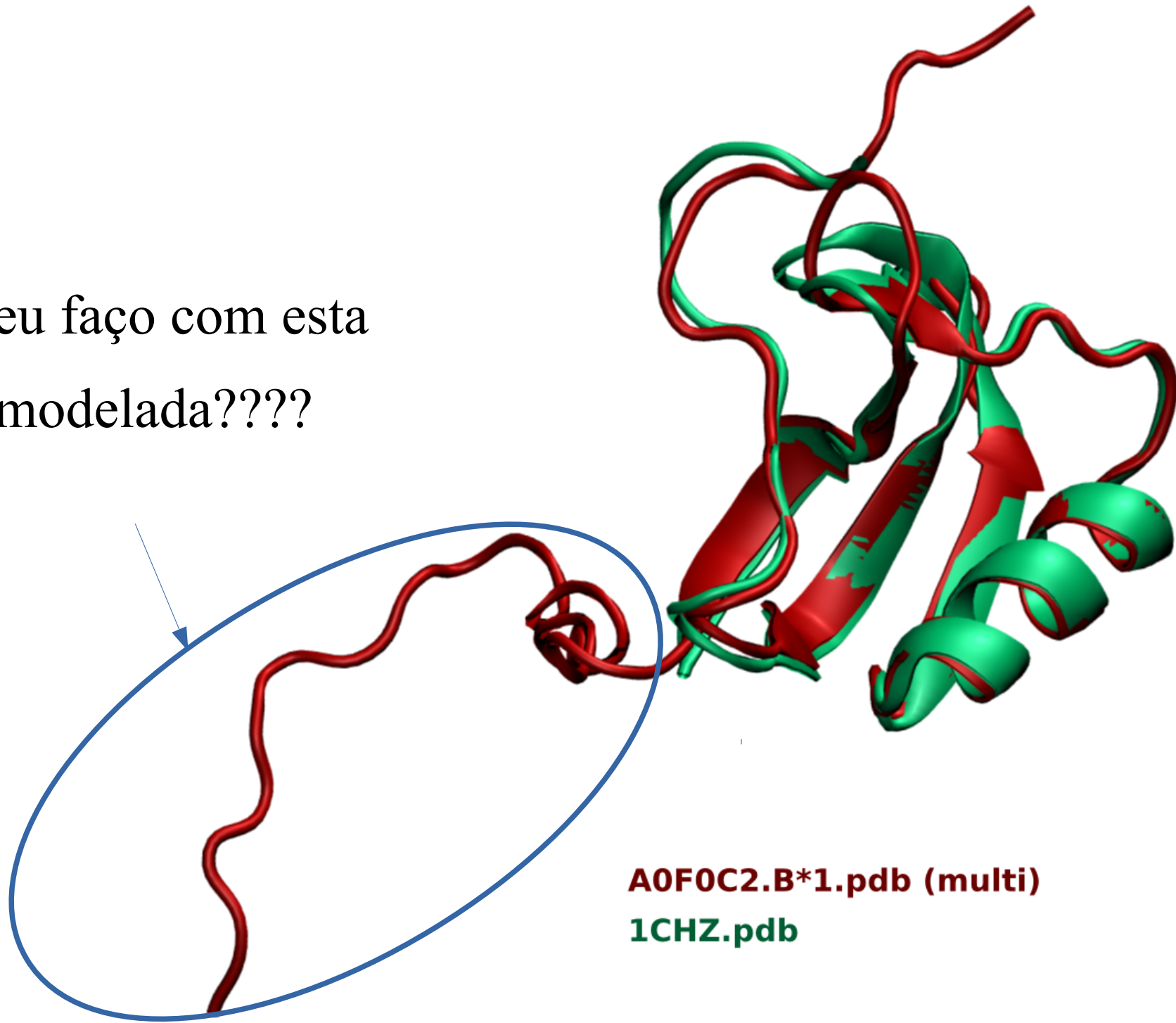




A0F0C2.B*1.pdb (single)

A0F0C2.B*1.pdb (multi)

Mas o que eu faço com esta
região não modelada????



A0F0C2.B*1.pdb (multi)
1CHZ.pdb

• Verificar possíveis erros de montagem.

• Verificar se é uma região de pré-/pró- domínio (*e.g.* SignalP).

• Verificar se é uma porção transmembranar (*e.g.* TMPred, TMHMM2, Phobius).

• Verificar regiões de consenso em preditores de estrutura secundária (*e.g.* PsiPred, Jufo, Porter).

• Verificar contatos entre resíduos (*e.g.* TMhit and TMhhcp)

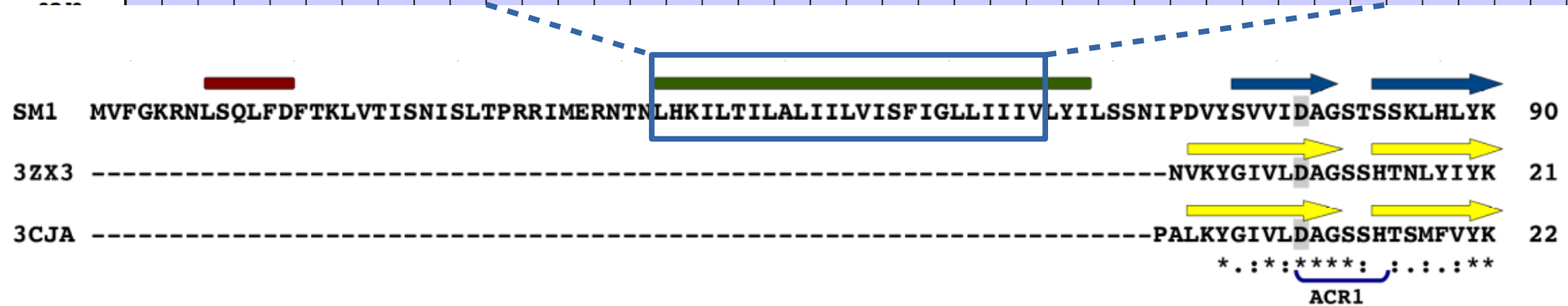
• Realizar a predição estrutural via protocolos híbridos (*e.g. ab initio, threading*).



A0F0C2.B*1.pdb (multi)
1CHZ.pdb

PREDIÇÃO: Peptídeo Sinal: NÃO - Transmembrana: SIM
Estrutura Secundária: HÉLICE

	TM1																																											
Pos Sm1_ali	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66				
Sm1	P	R	R	I	M	E	R	N	T	N	L	H	K	I	L	T	I	L	A	L	I	I	L	V	I	S	F	I	G	L	L	I	I	I	V	L	Y	I	L	S				
Jpred	C	C	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C					
Confidence	7	7	7	7	0	7	6	5	2	4	3	7	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	7	5	0	6	7	7	7					
Psipred	C	C	C	C	C	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C					
Confidence	6	4	4	2	1	1	1	0	2	4	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	9	8	9	8	7	6	6	3	0	3	5				
Jufo	C	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C					
Confidence	0,6	0,5	0,4	0,5	0,5	0,5	0,4	0,4	0,4	0,5	0,6	0,6	0,7	0,7	0,7	0,7	0,6	0,7	0,6	0,6	0,7	0,6	0,7	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,5	0,5	0,5	0,4	0,5	0,6					



												TM2																																			
Pos	Sm1	ali	493	494	495	496	497	498	499			500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530					
Sm1	P	K	V	S	V	S	T	-	-	Q	L	F	V	C	L	F	I	L	L	L	L	L	L	I	I	G	S	I	I	G	L	L	L	T	R	N	F	L	S	K	S						
Jpred	H	H	H	H	H	H	C	-	-	C	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H					
Confidence	7	7	7	7	7	7	7	-	-	7	6	0	6	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	7	4	4	4	5	5	5	5	4	3	1	0	0					
Psipred	H	H	H	C	C	C	H	-	-	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C						
Confidence	9	8	7	4	3	2	0	-	-	3	5	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	8	7	6	6	4	0	3	4	6	8					
Jufo	H	H	H	C	C	C	C	-	-	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	C	C	C						
Confidence	0,7	0,6	0,5	0,4	0,4	0,3	0,4	-	-	0,4	0,5	0,5	0,6	0,6	0,6	0,7	0,6	0,6	0,7	0,6	0,6	0,6	0,6	0,6	0,7	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,6	0,5	0,5	0,5	0,4	0,4	0,5	0,5	0,4	0,4				



PREDIÇÃO: Peptídeo Sinal: NÃO - Transmembrana: SIM
Estrutura Secundária: HÉLICE - Contatos: SIM

The submitted sequence:

>sm1_2

MVFGRNLSQLFDFTKLVTISNISLTPRRIMERNTNLHKILTILALIILVISFIGLLIIIVLYILSSNIP
 DVYSVVIDAGSTSSKLHLYKWIDEPFRSNGKVDEVTNEKLSPGISDYINDTIKAYDTLKPKLLKLTNSLT
 FEQKQHTPIYLAATAGMRLKLIEDPLGSLDLFSVIRQYLKQSGFQIETPNERIRLLYGSEEGLYGWVSVN
 YILGIIKEGKQTNPSDTVGSDDLGGASTQIAFIPKVYSNIPKEKLDFYPLRLYGNDFSVYSHSFLCYGKS
 EFERRVITSIAAASLLQSKIPNPCFLQGYKSDLYNAFEWFSGSCLSGTYVKKTFAEEIFRPPNMNSFSFN
 GTGQPNECVNYILKHFQTKCDHSSCSFNNVFQPAPFGKFMAYSGFSYVMRYLFPNKNTGFTLTEVTDAM
 KFCKKPWKDVAKITKLSQDQGTAKYCFDGLYIITLLKMYGFTTDESWKITITFDSKVNGKSVSWALGYMLD
 QSGHLPSESPKVSSTQLFVCLFILLLLLIIGSIIGLLLTRNFLSKSKKSTSQV

The submitted topology:

[illegible]

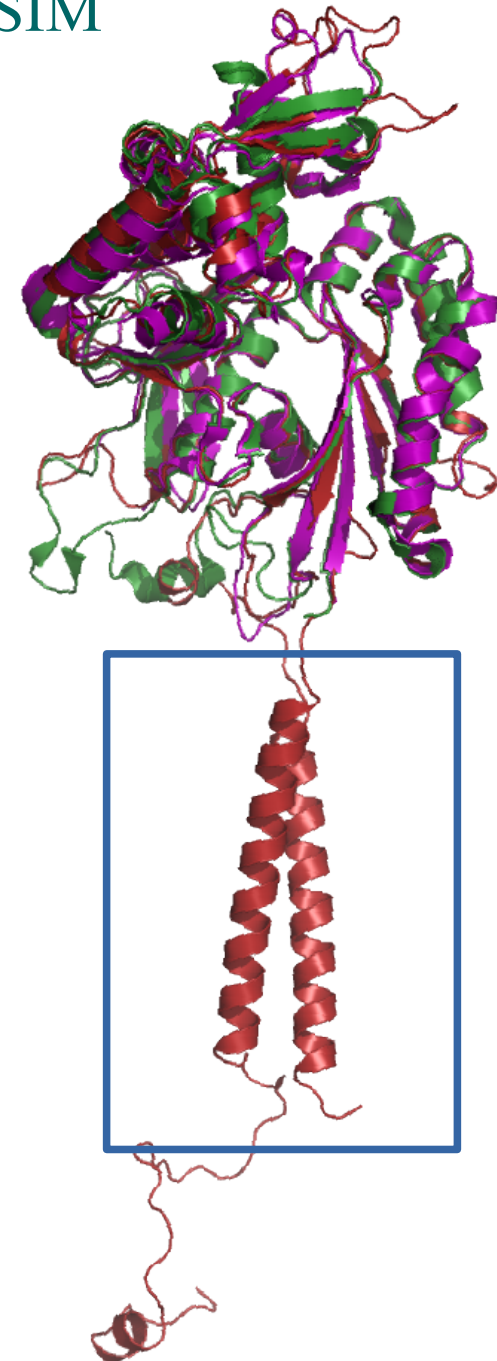
The top L/5 predicted residue contacts are listed in the following table. L is the sum of lengths of the TM segments.

Contact residue pairs predicted by TMhhcp

helix	position	residue	helix	position	residue	score
1	58	I	2	526	G	0.54
1	53	F	2	522	G	0.54
1	60	I	2	526	G	0.54

PREDIÇÃO: Peptídeo Sinal: NÃO - Transmembrana: SIM
Estrutura Secundária: HÉLICE - Contatos: SIM

<u>Region</u>	Prediction			
TM	Phobius	TMHMM	TMpred	SS
TM1/SmATPDase1	40-65	43-65	43-65	35-62
TM2/SmATPDase1	507-529	508-530	510-530	508-537



Pôster 67 (terça-feira):

The 3D models of Ecto-NTPases CD39 (human) and SmNTPDase1 (*Schistosoma mansoni*): A comparative analysis

VSP Nunes¹, P Faria-Pinto², EG Vasconcelos², CCH Borges³,
PVSZ Capriles³

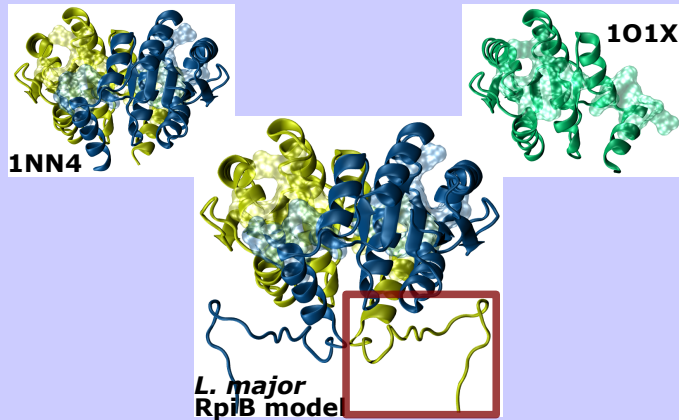
¹Postgraduate Program in Computational Modeling – UFJF/ME,

²Departament of Biochemistry – UFJF/ME,

³Departament of Computer Science – UFJF/ME

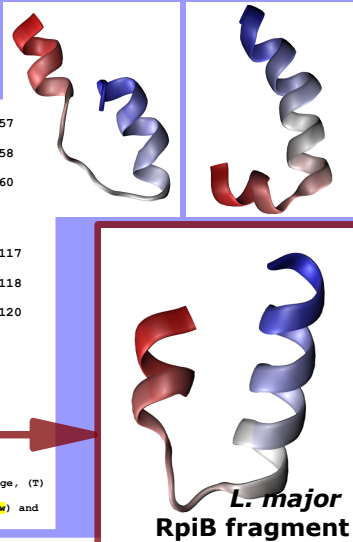
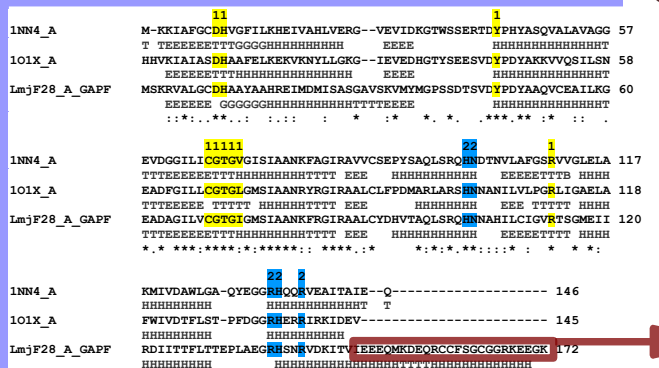
PREDIÇÃO: Peptídeo Sinal+Transmembrana: NÃO - Estrutura Secundária: HÉLICE

Comparative Modeling



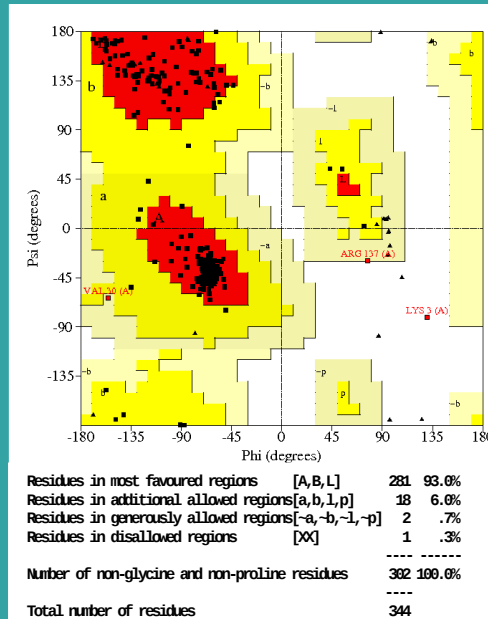
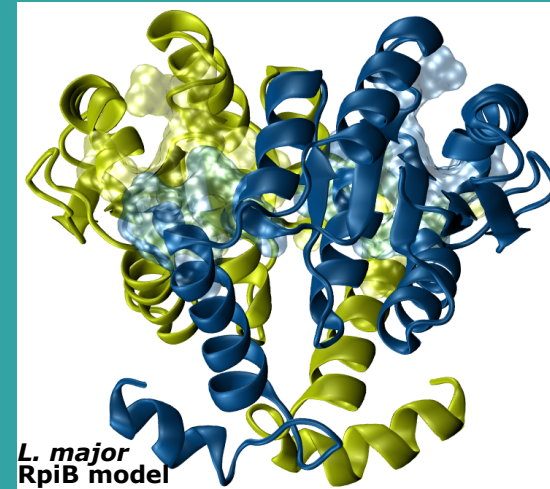
1. Template detection (BLAST)
2. Model construction (Modeller)
3. Quality analysis (Procheck)

Ab initio Prediction



1. Secondary structure prediction
2. Ab initio prediction (GAPF)
3. Clusterization (g_cluster-GROMACS)
4. Secondary structure analysis (Stride)

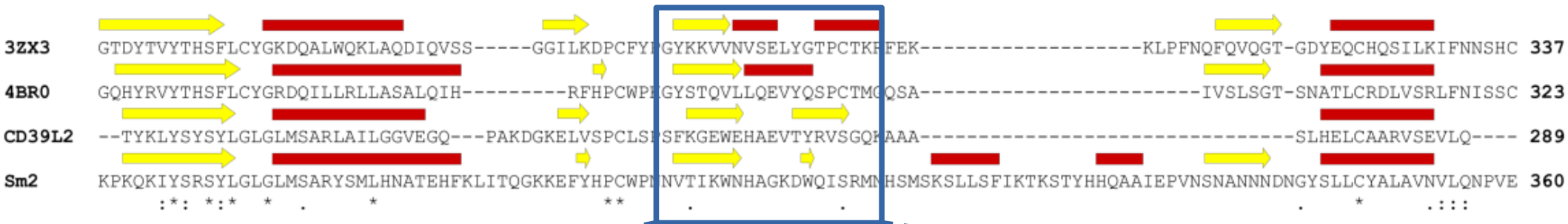
Final Model Construction



1. Model reconstruction (Modeller)
2. Quality analysis (Procheck)
3. Loop optimization (Modeller)
4. Quality analysis (Procheck)

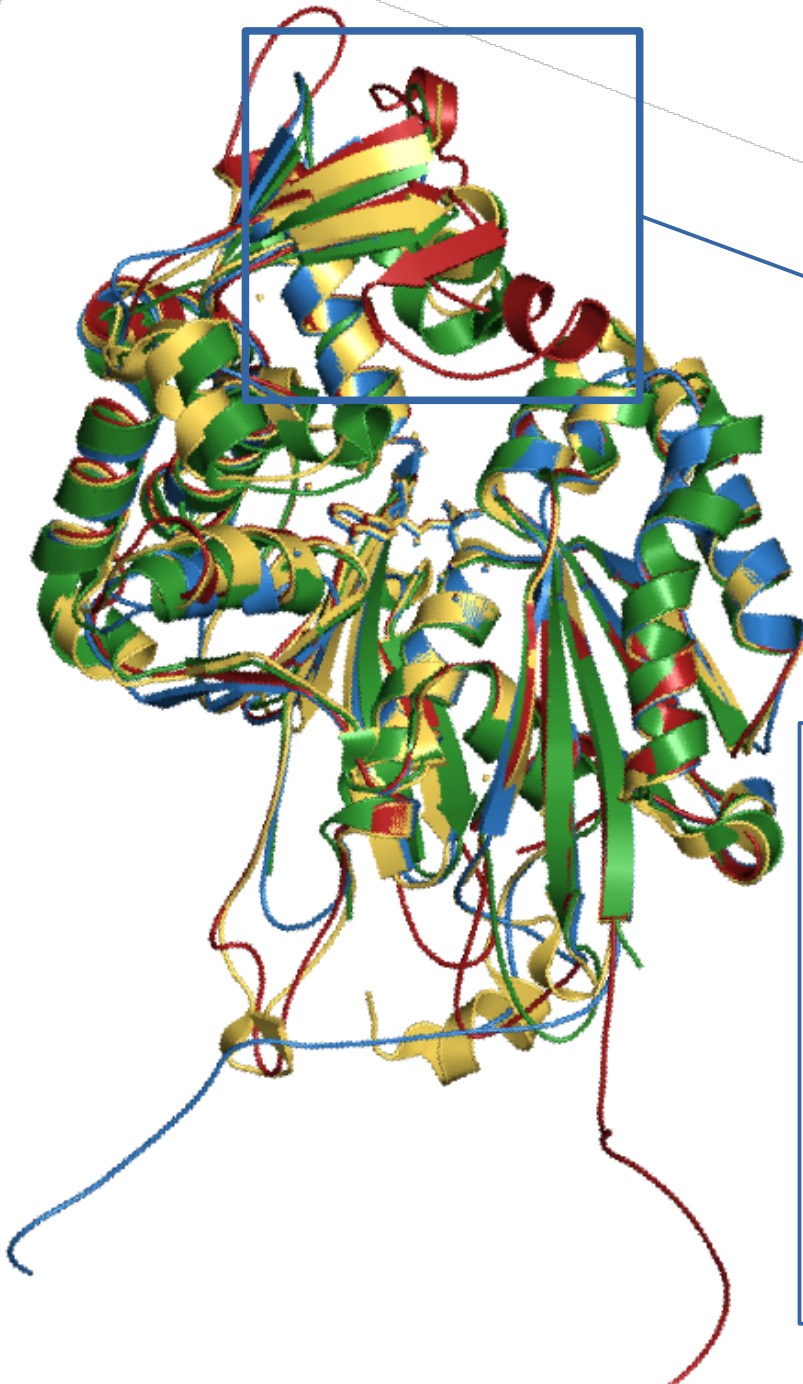
Fonte: Batista, L. P. (2011). Modelagem Molecular da Ribose-5-fosfato isomerase de *Leishmania major* e de *Homo sapiens*: Predição de Estruturas e Estudos de Atracamento Molecular. Dissertação de Mestrado - LNCC.

O que fazer com regiões de baixa identidade/similaridade na estrutura?



SM2	N	V	T	I	K	W	N	H	A	G	K	D	W	Q	I	S	R	M	N
JUFO	C	E	E	E	E	E	E	E	C	C	C	E	E	E	E	C	C	C	C
PROF	0,483	0,404	0,448	0,546	0,547	0,543	0,474	0,401	0,474	0,509	0,433	0,445	0,464	0,445	0,398	0,456	0,509	0,52	0,539
JPRED	C	C	C	E	E	E	E	C	C	C	E	E	E	E	E	C	C	C	C
PSIPRED	7	7	3	5	8	8	6	0	6	7	0	1	5	4	0	6	7	7	7
SABLE	C	C	C	E	E	E	E	C	C	C	C	E	E	E	E	C	C	C	C
PSSPRED	7	5	4	5	6	7	6	4	6	6	5	5	5	4	6	7	7	7	7
APSSP2P	C	C	E	E	E	E	E	E	C	C	E	E	E	E	E	C	C	C	C
PREDATOR	9	5	0	6	7	9	8	5	6	6	3	6	6	5	3	1	5	7	7
3zx3_a	C	C	C	C	E	E	E	E	C	C	E	E	E	E	E	C	C	C	C
STRIDE	0,9	0,8	0,7	0,5	0,9	1	0,9	0,9	0,7	0,7	0,7	0,8	0,9	0,6	0,6	0,8	0,8	0,8	0,8
4bqz	C	E	E	E	E	E	C	C	C	C	C	C	E	E	E	E	E	C	C
STRIDE	0	0	0,732	0,746	0	0,479	0,514	0,753	0,711	0,753	0,67	0,639	0,478	0,489	0	0	0,445	0,608	0,738
3zx3_a	G	Y	K	K	V	V	N	V	S	E	L	Y	G	T	P	C	T	K	R
STRIDE	C	E	E	E	E	E	H	H	H	H	C	C	C	H	H	H	H	H	H
4bqz	G	Y	S	T	Q	V	L	L	Q	E	V	Y	Q	S	P	C	T	M	G
STRIDE	C	E	E	E	E	E	E	H	H	H	H	H	H	C	C	C	C	C	C

O que fazer com regiões de baixa identidade/similaridade na estrutura?



Modelagem: Modeller

Metodologia:

Otimização de loop+

Restrição de estrutura secundária+

Restrição de distância+

Imposição de contatos

Pôster 68 (quarta-feira):

Comparative Modeling of SmATPDase2 (Sm2) and

HsNTPDase6 (CD39L2)

+

**Structural comparative analysis of secreted NTPDase models
of *Schistosoma mansoni* and *Homo sapiens***

Vinicius Carius de Souza , Vinicius Schmitz Nunes, Eveline Gomes

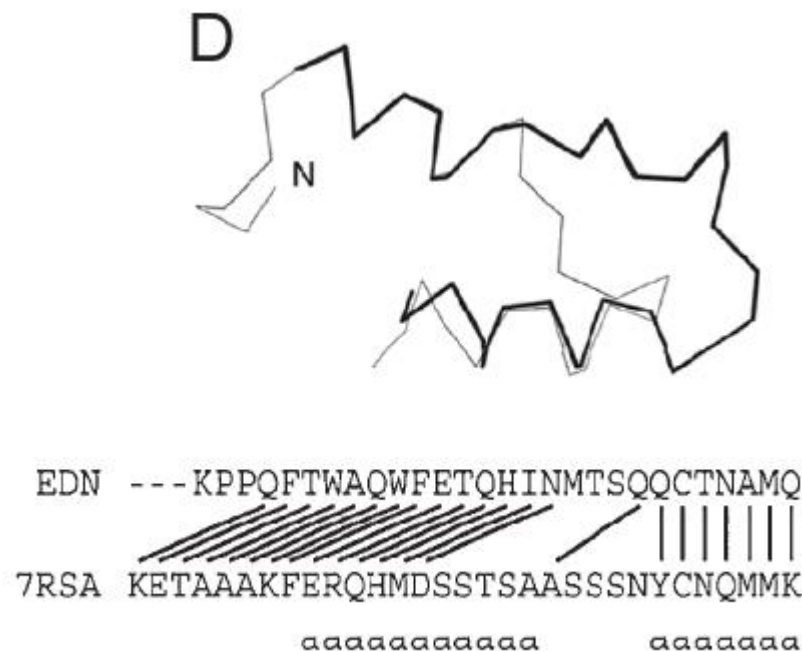
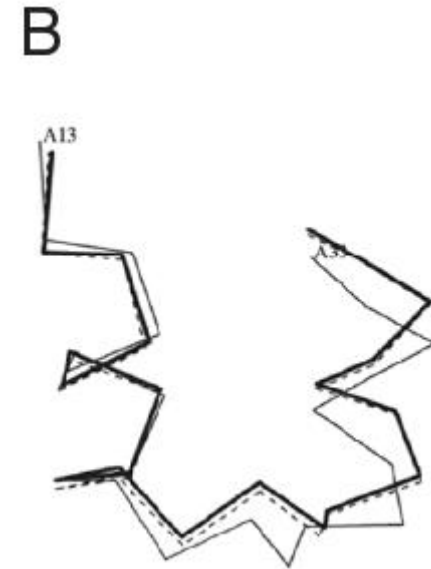
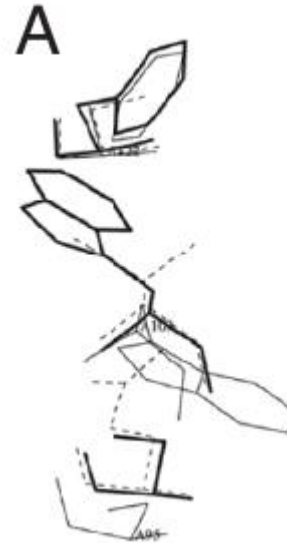
Vasconcelos, Priscila Faria-Pinto, and Priscila V. S. Z. Capriles

Lecture Notes in Computer Science (2014) (Accepted)

Modelagem Comparativa: Validação do Modelo

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.



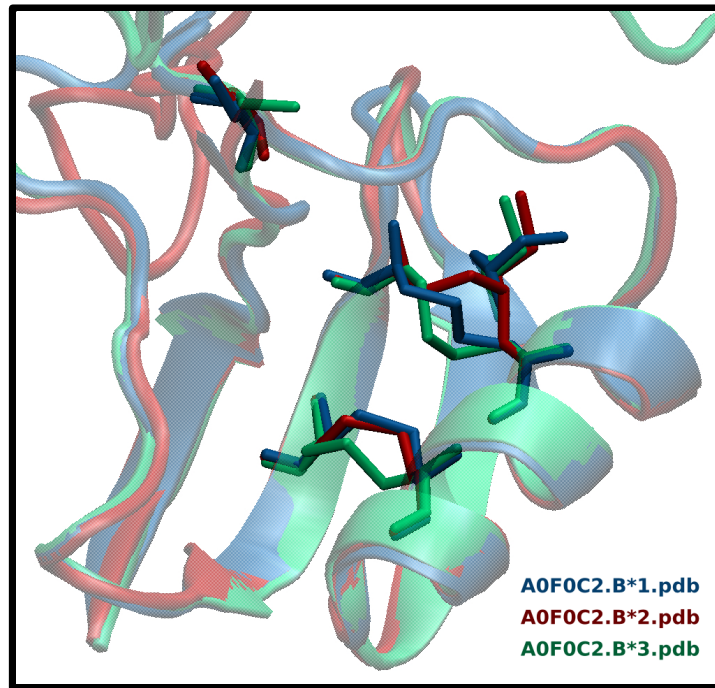
Modelagem Comparativa: Refinamento do Modelo

1) Minimização de Energia:

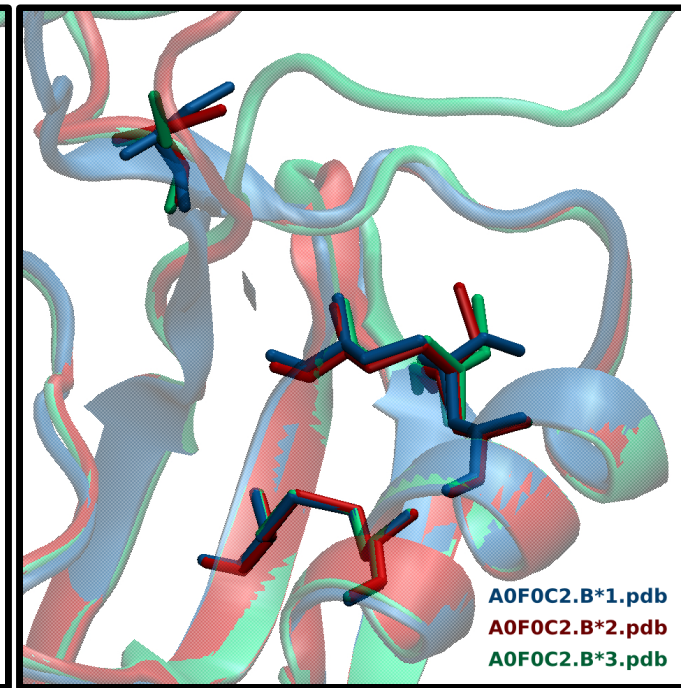
- a) Gradiente Conjugado
- b) Máximo Declive
- c) (...)

2) Dinâmica Molecular:

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



Modelos sem Otimização



Modelos com Otimização

Modelagem Comparativa: 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

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



Fonte: Esquema baseado em Baker, D. e Sali, A. (2001). *Nature*, 294(93): 93-96.

Modeller Online

[Sali Lab Home](#) [ModWeb](#) [ModLoop](#) [ModBase](#) [ModEval](#) [PCSS](#) [FoXS](#) [IMP](#) [ModPipe](#)

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

ModWeb: A Server for Protein Structure Modeling

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
Andrej Sali

Please address inquiries to:
modweb@salilab.org

ModWeb version SVN.r1515

General information ?

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

Model selection criteria ?

☒ Best scoring model

☒ Longest well scoring model

Fold assignment methods ?

Fast (Seq-Prf) ▼

☒ Upload models to ModBase

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

Modeller Online



Automatic 3D
Comparative Modelling

[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



(Site: www.mhonline.lncc.br)