

Tutorial – Modelagem Comparativa

Aula Prática – 06/12/2018

A estratégia baseia-se no conhecimento de que a conformação estrutural de uma proteína é mais conservada que sua sequência de aminoácidos durante o processo evolutivo, e que pequenas mudanças na sequência, em geral, resultam em, apenas, sutis modificações na estrutura tridimensional. Se pelo menos uma sequência homóloga para qual a estrutura tridimensional resolvida esteja disponível é encontrada, o método de escolha para predição da estrutura tridimensional de uma proteína-alvo é a própria modelagem comparativa.

Geralmente, o processo de obtenção de um modelo proteico virtual através da execução da estratégia da modelagem comparativa envolve quatro etapas principais (Figura 1), entre elas: (i) identificação dos moldes; (ii) alinhamento das sequências; (iii) construção do modelo e; (iv) validação dos mesmos.

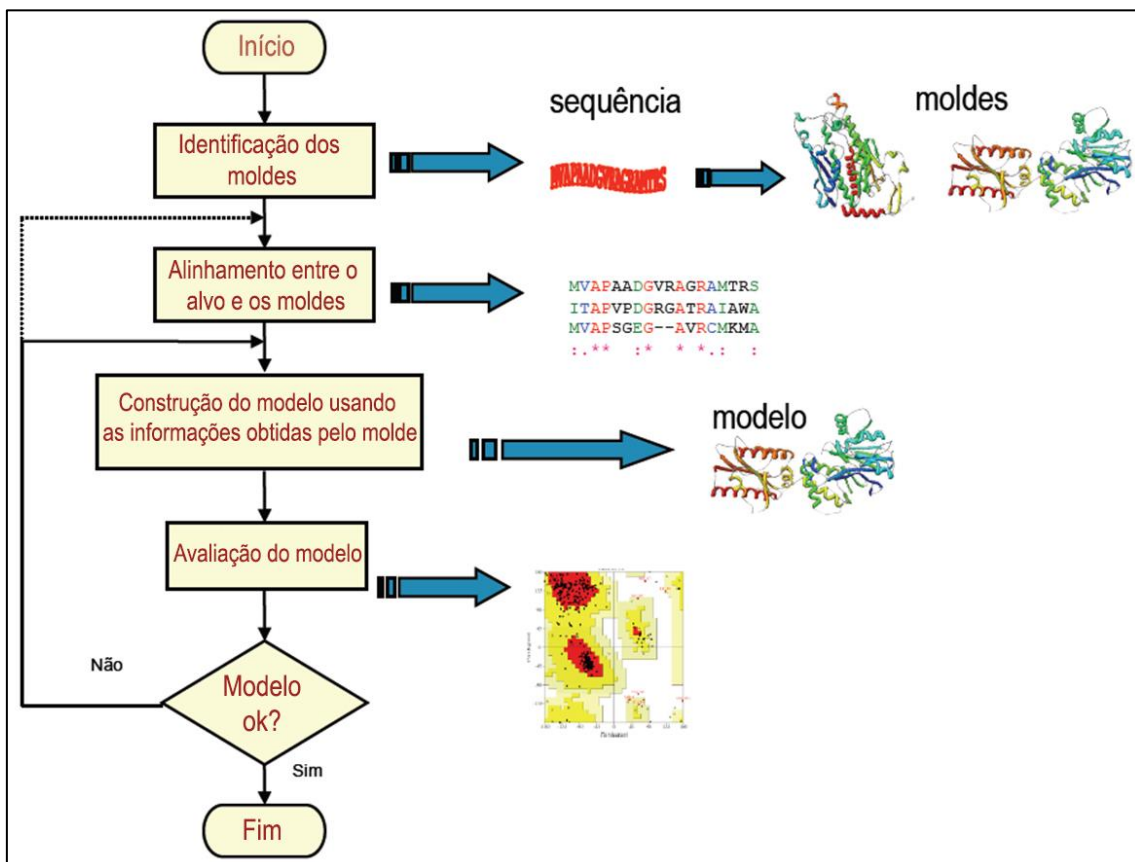


Figura 1: Esquema geral do método teórico da modelagem comparativa. Adaptado de MARTY- RENOM *et al.*, 2000.

I. Identificação dos Moldes

O primeiro passo na modelagem comparativa é a identificação de estruturas tridimensionais resolvidas, que possam atuar como uma base estrutural para a modelagem da sequência-alvo (proteína alvo). Esta identificação pode ser realizada levando-se em consideração vários aspectos, como: conhecimento estrutural, similaridade de função, expressão pelo mesmo grupo de genes, similaridade sequencial ou até correlação evolutiva.

A etapa de busca de proteínas moldes tem como objetivo promover a identificação de proteínas com estruturas tridimensionais resolvidas que se correlacionam com a proteína-alvo, e que possam funcionar como molde (proteínas-molde) para a modelagem dessa proteína com estrutura tridimensional desconhecida. A busca das proteínas moldes pode ser realizada utilizando-se a ferramenta “Basic Local Alignment Search Tool” (BLAST). Nessa busca, que funciona como um processo de triagem, são utilizadas certas ferramentas para a avaliação do grau de similaridade entre fragmentos das sequências (alinhamento local), com o objetivo de distinguir entre similaridades importantes do ponto de vista biológico ou estrutural de similaridades ao acaso e que não apresentam importância significativa. A busca é realizada no banco de dados de proteínas com estruturas resolvidas, Protein Data Bank (PDB). Com a introdução da sequência da proteína-alvo, um algoritmo de busca de similaridade de sequências é, então, aplicado.

II. Alinhamento das sequências

A partir da identificação de uma correlação entre a proteína-alvo e a(s) proteína(s)-molde com estrutura resolvida selecionada(s) na busca, se torna necessária a obtenção de um alinhamento sequencial entre a proteína-alvo e a(s) proteína(s)-molde. O objetivo primordial do alinhamento entre as sequências é o de identificar uma correlação ótima entre os resíduos de aminoácidos de cada sequência. Quando apenas uma sequência-molde é identificada, é realizado apenas um alinhamento “pairwise” (ao par) com a sequência-alvo. Quando várias sequências-molde são identificadas, é necessária a obtenção de um alinhamento múltiplo, em que estarão dispostas todas as sequências-molde selecionadas na busca, bem como a sequência-alvo.

Com respeito à faixa aceitável de identidade sequencial para a execução da modelagem molecular comparativa, a literatura descreve como significativo um valor acima de 30 % de identidade sequencial entre a(s) proteína(s)-molde e a proteína-alvo. Esse valor é dependente do número de resíduos da proteína que será modelada, sendo menos crítico quanto maior o comprimento da sequência, como mostrado na Figura 5.

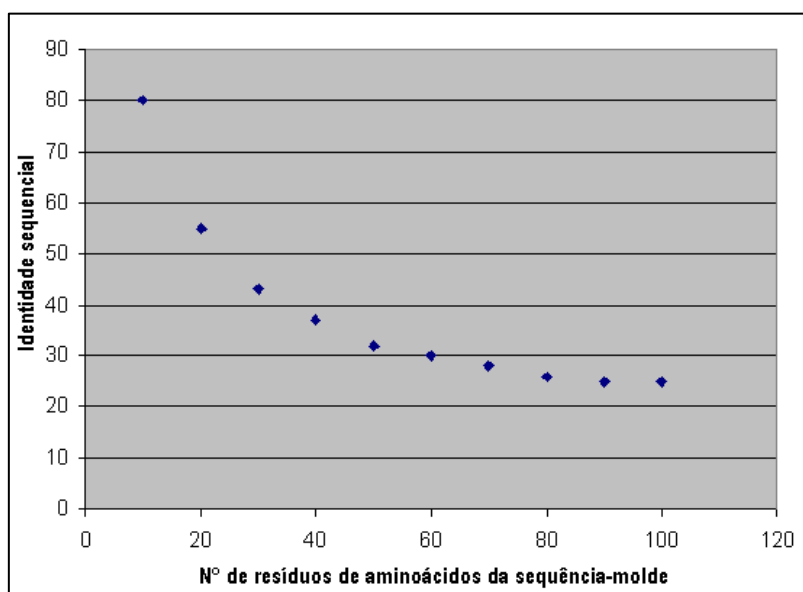


Figura 2: Limiar para realização da modelagem comparativa. O valor significativo da identidade sequencial entre a proteína-molde e a(s) proteína(s)-alvo para a realização do procedimento se torna menos crítico quanto maior o comprimento da sequência-alvo.

III. Construção do Modelo

A localização no espaço tridimensional do maior número possível de átomos da proteína alvo depende do alinhamento. A construção da parte interna da proteína alvo baseia-se na ideia de que a conformação da cadeia principal da estrutura molde pode ser transferida para ela. Existem muitos métodos de modelagem dessas regiões conservadas. Os mais utilizados são (i) corpos rígidos e (ii) restrições espaciais.

III.I. Corpos Rígidos

Este método pode ser utilizado pelo **SWISS-MODEL**, um servidor automatizado na Web, nele alinhamentos múltiplos molde são utilizados para

identificar regiões estruturais conservadas, que são modeladas como corpos rígidos, enquanto que as regiões variáveis são construídas em torno deles. A construção de um modelo ocorre 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 sequências de aminoácidos das regiões estruturalmente conservadas nas sequências de referência.
3. As regiões que não satisfazem as exigências são chamadas de regiões variáveis (geralmente, regiões de voltas que conectam as regiões conservadas).
4. A cadeia principal das regiões variáveis: pode ser obtida em bancos de dados de estruturas proteicas, 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.

III.II. Restrições Espaciais

O software Modeller utiliza restrição espacial para construção da estrutura alvo, em que extrai características geométricas (comprimentos de ligação, ângulos, etc) a partir da estrutura de modelo e construir um modelo de satisfação dessas restrições.

1. Alinhamento entre a sequência 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.

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).

IV. Validação do Modelo

Após a construção do modelo, é necessário identificar possíveis erros relacionados a escolha das estruturas de referência, e ao alinhamento entre sequência-alvo e molde. Caso o modelo seja caracterizado de má qualidade, todo o processo de modelagem deve ser revisto, no intuito de melhorar o alinhamento inicial ou utilizar outros métodos. São avaliadas as qualidades dos modelos, principalmente se houver diferentes orientações referentes aos resíduos do sítio ligante nos modelos gerados. Um grande número de propriedades de diferentes graus de organização estrutural, como: exatidão estereoquímica, qualidade do empacotamento e confiabilidade do enovelamento, são verificadas durante a etapa de validação. Alguns programas podem ser utilizados para validação dos modelos gerados: (i) Procheck, (ii) Whatif e, (iii) Verify3D.

IV.1. PROCHECK (Qualidade Estereoquímica)

O Procheck avalia diversos parâmetros estereoquímicos, de importância fundamental, como os comprimentos de ligação, os ângulos planos, a planaridade dos anéis de cadeias laterais, os ângulos torcionais da cadeia principal, gerando o gráfico de Ramachandran, que é muito útil, pois, define os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis, além de orientar a avaliação da qualidade para modelos teóricos e experimentais de proteínas.

A estrutura proteica possui três ângulos principais, Ω , φ e ψ . Como o ângulo Ω é fixo os ângulos φ e ψ são os responsáveis por toda a variação conformacional da cadeia principal, pois tornam a cadeia polipeptídica flexível. Observa-se na Figura 6a, a estrutura proteica e seus ângulos principais. O gráfico de Ramachandran e uma representação gráfica de ângulos φ versus ângulos ψ . Para que se garanta uma maior precisão dos modelos gerados, no mínimo 90% dos ângulos φ e ψ da cadeia principal devem estar na região mais favorável do gráfico de Ramachandran.

Na Figura 6b, observa-se um gráfico de Ramachandran. Os pontos quadrados representam os resíduos de aminoácidos da proteína modelada, as glicinas são representadas por triângulos. As áreas em vermelho compreendem as regiões mais favoráveis para resíduos se encontrarem, em amarelo as regiões favoráveis, em bege as regiões pouco favoráveis e em branco, as regiões desfavoráveis. A região branca e desfavorável para todos os aminoácidos, com exceção da glicina.

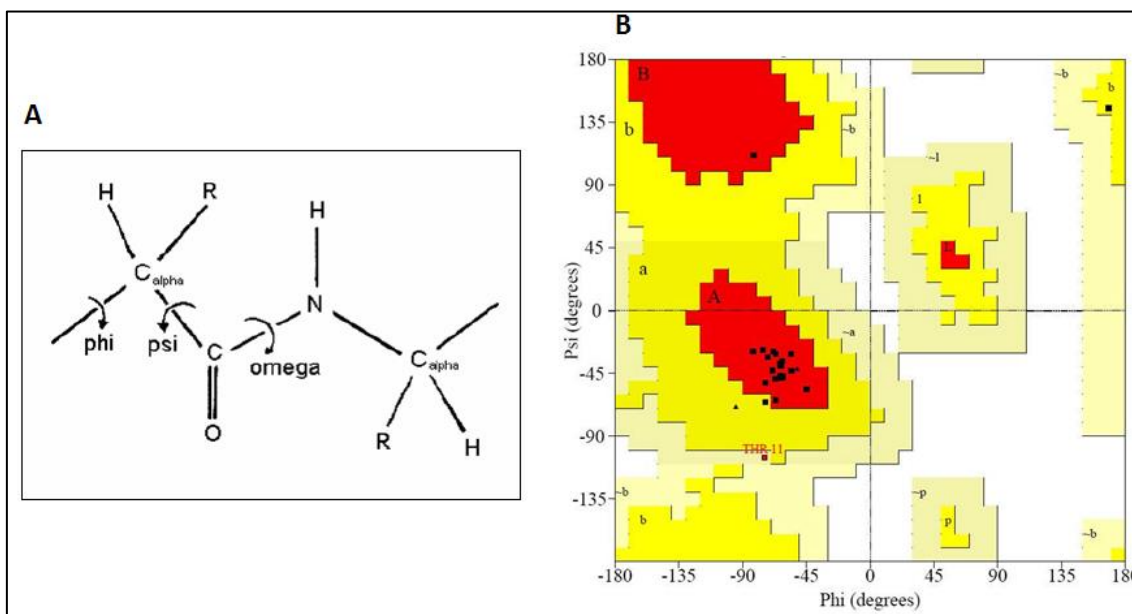


Figura 3: (a) Estrutura proteica e seus três ângulos principais, omega (Ω), phi (Φ) e psi (Ψ). Como o ângulo Ω é fixo os ângulos Φ e Ψ são os responsáveis por toda a variação conformacional da cadeia principal, pois tornam a cadeia polipeptídica flexível. **(b)** Gráfico de Ramachandran gerado pelo Procheck. As regiões representadas em vermelho, são as mais favoráveis, as representadas em amarelo, são as favoráveis, as representadas em bege, são as menos favoráveis e as regiões brancas, são as desfavoráveis.

IV.II. Whatlf (Qualidade do Modelo)

O Whatif (VRIEND, 1990), avalia a qualidade dos contatos atômicos envolvendo os átomos de cada resíduo. É possível comparar a distribuição das posições de átomos em torno de cada resíduo.

IV.III. Verify 3D (Energia de Interação entre Resíduos)

O modelo proteico pode ser avaliado também quanto a qualidade dos ambientes químicos, determinando o nível de confiabilidade do enovelamento proteico do modelo virtual. O Verify 3D pode realizar essa tarefa, determinando os ambientes químicos de cada resíduo do modelo e atribuindo “scores” com referência a uma matriz construída a partir de uma análise estatística envolvendo estruturas de proteínas armazenadas no PDB.

IV. LIMITAÇÕES DO MÉTODO

As etapas mais problemáticas da modelagem de proteínas por homologia são o alinhamento das sequências e a modelagem das regiões de alças. Muitos avanços têm sido feitos nos métodos de alinhamento de sequências. Já a modelagem de alças ainda é um fator limitante, principalmente quando o número de resíduos é grande.

Em alguns casos, cerca de 27% de identidade entre as sequências é suficiente para um bom alinhamento e, conseqüentemente, para um bom modelo. Há casos, porém, em que apesar do alto grau de identidade e do bom alinhamento não se obtém modelos satisfatórios. Cada modelagem é um problema particular e deve-se ter sempre em mente que a construção de um modelo de proteína é uma tarefa que deve ser fundamentada em princípios de evolução molecular para cada grupo específico de proteínas homólogas. Quando as proteínas têm vários domínios, a construção do modelo é um processo ainda mais complexo.

Os erros normalmente encontrados em modelos de proteínas gerados por homologia podem ser divididos em cinco categorias (Figura 7):

1. Posicionamento da cadeia lateral. Quando o grau de identidade entre a estrutura alvo e a estrutura molde é baixo, podem surgir problemas de encaixe de algumas cadeias laterais no corpo principal do modelo, formando regiões interrompidas.
2. Distorções em regiões alinhadas. Divergências entre sequências podem provocar distorções durante a localização dos resíduos da cadeia principal, mesmo que o enovelamento global do modelo seja satisfatório. A utilização de várias proteínas-molde pode reduzir este tipo de erro.
3. Regiões sem molde. Este tipo de erro pode ocorrer em segmentos da proteína alvo que não têm equivalentes na estrutura molde, principalmente em regiões de

alças grandes. A avaliação crítica do alinhamento proposto e do ambiente em que a alça está localizada pode levar à superação deste tipo de erro.

4. Alinhamentos ineficientes. Ocorrem principalmente quando o grau de identidade entre as sequências envolvidas é inferior a 30%. Para reduzi-los, deve-se dar preferência a alinhamentos múltiplos e, se possível, modificar iterativamente as regiões do alinhamento tidas como problemáticas.

5. Escolha errada do molde. Este tipo de erro é mais provável quando o grau de identidade é inferior a 25%. Nestes casos, deve-se buscar parâmetros experimentais e modificar com cautela o alinhamento.

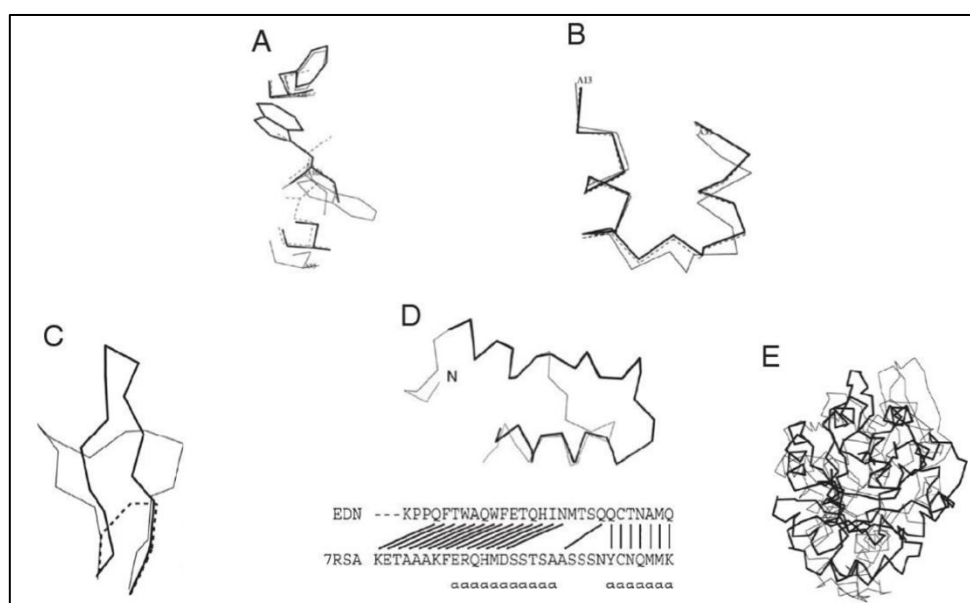


Figura 4: Erros mais comuns na modelagem comparativa: (a) posicionamento da cadeia lateral; (b) distorções em regiões alinhadas; (c) regiões sem molde; (d) alinhamentos ineficientes; (e) escolha errada do molde.

TUTORIAL SWISS-MODEL

Este tutorial está orientado para usuários do sistema operacional Linux e tem por objetivo exemplificar a aplicação do servidor **SWISS-MODEL** para predição de estruturas de proteínas via modelagem comparativa utilizando a técnica de corpos rígidos.

I. **Entre no servidor SWISS-MODEL:** <https://swissmodel.expasy.org/>

Clique em Start Modeling

SWISS-MODEL

Welcome to SWISS-MODEL

SWISS-MODEL 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.

[Start Modelling](#)

Celebrating 25 years of SWISS-MODEL!
Save the date - Thursday 18 October in Basel, more details to follow!

Protein Structure Bioinformatics Group
c/o Prof. Torsten Schwede
Swiss Institute of Bioinformatics
Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel / Switzerland
help-swissmodel@unibas.ch

BIOZENTRUM
The Center for
Molecular Life Sciences

20th Anniversary
Swiss Institute of
Bioinformatics

Existem quatro tipos possíveis de entrada:

Supported Inputs ⓘ

Sequence(s)	▼
Target-Template Alignment	▼
User Template	▼
DeepView Project	▼

1) Sequência de aminoácidos de proteínas ou identificador UniProtKB

A sequência de aminoácidos da proteína alvo pode ser submetida como texto simples ou no formato FASTA.

- Example of plain text sequence:

```
MVEIVYWSGTGNTTEAMANEIEAAVKAAGADVSVRFEDTNVDDVASKDVILLGCPAMGSE  
ELEDVVEPFPTDLAPKLGKVKVGLFGSYGWGSGEWMDAWKQRTEDTGATVIGTAIVNEM  
PDNAPECKELGEAAAKA
```

- Example of FASTA sequence:

```
>sp|P00321|FLAV_MEGEL Flavodoxin - Megasphaera elsdenii.  
MVEIVYWSGTGNTTEAMANEIEAAVKAAGADVSVRFEDTNVDDVASKDVILLGCPAMGSE  
ELEDVVEPFPTDLAPKLGKVKVGLFGSYGWGSGEWMDAWKQRTEDTGATVIGTAIVNEM  
PDNAPECKELGEAAAKA
```

Se a sequência proteica é depositada no banco de dados UniProtKB (The UniProt Consortium), o identificador UniProtKB da entrada pode ser fornecido como entrada (ou seja, P00321). Nesse caso, o identificador é imediatamente validado e substituído pela sequência correspondente. O botão "Add Hetero

Target" é fornecido para inserir múltiplas sequências de destino representando diferentes subunidades de um hetero-oligômero. As sequências de destino devem ser exclusivas e podem ser enviadas como texto simples, sequências FASTA ou ACs UniProtKB. Se um hetero-oligômero for solicitado, procuramos somente biunits de templates que contenham cadeias conectadas com todas as subunidades desejadas.

2) Alinhamento de template-alvo

Os seguintes formatos são suportados atualmente: FASTA e Clustal.

- Examples for FASTA:

```
>THN_DENCL
KSCCPTTAARNQYNICRLPGTTPRVPVCAALSGCKIISGTGCPPGYRH-
>1crnA
TTCCPSIVARSNFNVCRLPGTPEALCATYTGCIIPGATCPGDYAN-
```

- Examples for Clustal:

```
CLUSTAL W (1.82) multiple sequence alignment

THN_DENCL      KSCCPTTAARNQYNICRLPGTTPRVPVCAALSGCKIISGTGCPPGYRH- 46
1crnA          TTCCPSIVARSNFNVCRLPGTPEALCATYTGCIIPGATCPGDYAN- 46
                .:*** ..* : **: * .. :** :** *...: ** *
```

3) Modelo de usuário

Se o usuário souber a estrutura do modelo a ser usado para modelagem, as coordenadas poderão ser carregadas no formato PDB(*) juntamente com a sequência de proteínas alvo. Modelos oligoméricos são aceitos, e é possível construir heterômeros adicionando múltiplas seqüências alvo à entrada.

Para iniciar um trabalho de modelagem com seu próprio modelo:

- a) Pressione o botão "User Template"
- b) Digite a sequência de destino como normal.
Opcional: para iniciar um projeto hetero, você pode clicar em " Add Hetero Target " para adicionar outra sequência de destino.
- c) Clique no botão "Add Template File ..."
- d) Localize o seu arquivo de coordenadas. Importante: Certifique-se de que não há aminoácidos quimicamente modificados!
- e) Clique em "Build Model".

(*) Um arquivo do tipo PDB contendo as coordenadas da estrutura do modelo. Para mais informações sobre o formato de arquivo PDB, consulte este link. Observe que o formato mmCIF atualmente não é suportado.

4) Projeto DeepView

Arquivos de projeto contendo as estruturas de modelo sobrepostas e o alinhamento entre o destino e o modelo podem ser carregados diretamente no espaço de trabalho SWISS-MODEL. Veja a seção “Project Mode” para mais detalhes.

II. Escolha Sequence e entre a sequência de interesse: Brain lipid-binding protein (BLBP)

> blbp

```
VDAFCATWKLTD SQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRT  
QCTFKNTEINFQLGEEFEETSIDDRNCKSVVRLDGDKLIHVQKWDGKETNCTR  
EIKDGKMVVTLTFGDIVAVRCYEKA*
```

III. Clique em Search for templates

The screenshot shows the SWISS-MODEL web interface. At the top, there is a navigation bar with 'Modelling', 'Repository', 'Tools', 'Documentation', 'Log in', and 'Create Account'. The main content area is titled 'Start a New Modelling Project'. It features a 'Target Sequence(s)' field with two targets: 'VDAFCATWKLTD SQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSIDDRNCKSVVRLDGDKLIHVQKWDGK' (residue 100) and 'ETNCTREIKDGKMVVTLTFGDIVAVRCYEKA' (residue 131). Below the target fields are 'Add Hetero Target' and 'Reset' buttons. The 'Project Title' field is 'Untitled Project' and the 'Email' field is 'Optional'. A red box highlights the 'Search For Templates' button. To the right, there is a 'Supported Inputs' dropdown menu with options: 'Sequence(s)', 'Target-Template Alignment', 'User Template', and 'DeepView Project'. Below the form, there is a disclaimer: 'By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.' The bottom part of the screenshot shows the 'Untitled Project' page with tabs for 'Summary', 'Templates', and 'Models'. The 'Template Results' section shows a progress bar labeled 'Queueing' and a link to bookmark the page: 'https://swissmodel.expasy.org/interactive/FqmQLg/'. At the bottom, the target sequences are displayed again with color coding: blue for the first target and red for the second target.

IV. Avalie o resultado da busca por templates

BIOZENTRUM University of Basel The Center for Molecular Life Sciences SWISS-MODEL Modelling Repository Tools Documentation Log in Create Account

All Projects

Untitled Project Created: today at 18:09

Summary Templates 50 Models

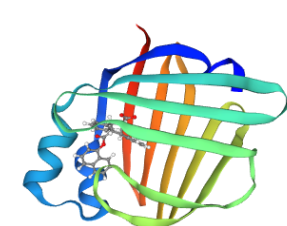
Template Results

Templates Quaternary Structure Sequence Similarity Alignment of Selected Templates More

Sort	Name	Title	Coverage	GMQE	ESQE	Identity	Method	Oligo State	Ligands
<input checked="" type="checkbox"/>	5ura.2.A	Fatty acid-binding protein, brain		0.98	-	87.02	X-ray, 1.9Å	monomer ✓	1 x 8KS
<input type="checkbox"/>	5ura.1.A	Fatty acid-binding protein, brain		0.98	-	87.02	X-ray, 1.9Å	monomer ✓	1 x 8KS
<input type="checkbox"/>	1fdq.2.A	FATTY ACID-BINDING PROTEIN, BRAIN		0.96	-	87.02	X-ray, 2.1Å	monomer ✓	1 x HXA
<input type="checkbox"/>	1fdq.1.A	FATTY ACID-BINDING PROTEIN, BRAIN		0.96	-	87.02	X-ray, 2.1Å	monomer ✓	1 x HXA
<input type="checkbox"/>	1jkk.1.A	BRAIN-TYPE FATTY ACID BINDING PROTEIN		0.96	-	87.02	NMR	monomer ✓	None
<input type="checkbox"/>	1fe3.1.A	FATTY ACID-BINDING PROTEIN, BRAIN		0.95	-	87.02	X-ray, 2.8Å	monomer ✓	1 x OLA
<input type="checkbox"/>	5ura.1.A	Fatty acid-binding protein, brain		0.95	-	87.60	X-ray, 1.9Å	monomer ✓	1 x 8KS
<input type="checkbox"/>	5ura.2.A	Fatty acid-binding protein, brain		0.95	-	87.60	X-ray, 1.9Å	monomer ✓	1 x 8KS

Build Models 1

Clear Selection



BIOZENTRUM University of Basel The Center for Molecular Life Sciences SWISS-MODEL Modelling Repository Tools Documentation Log in Create Account

All Projects

Untitled Project Created: today at 18:09

Summary Templates 50 Models

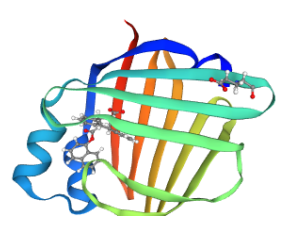
Template Results

Templates Quaternary Structure Sequence Similarity Alignment of Selected Templates More

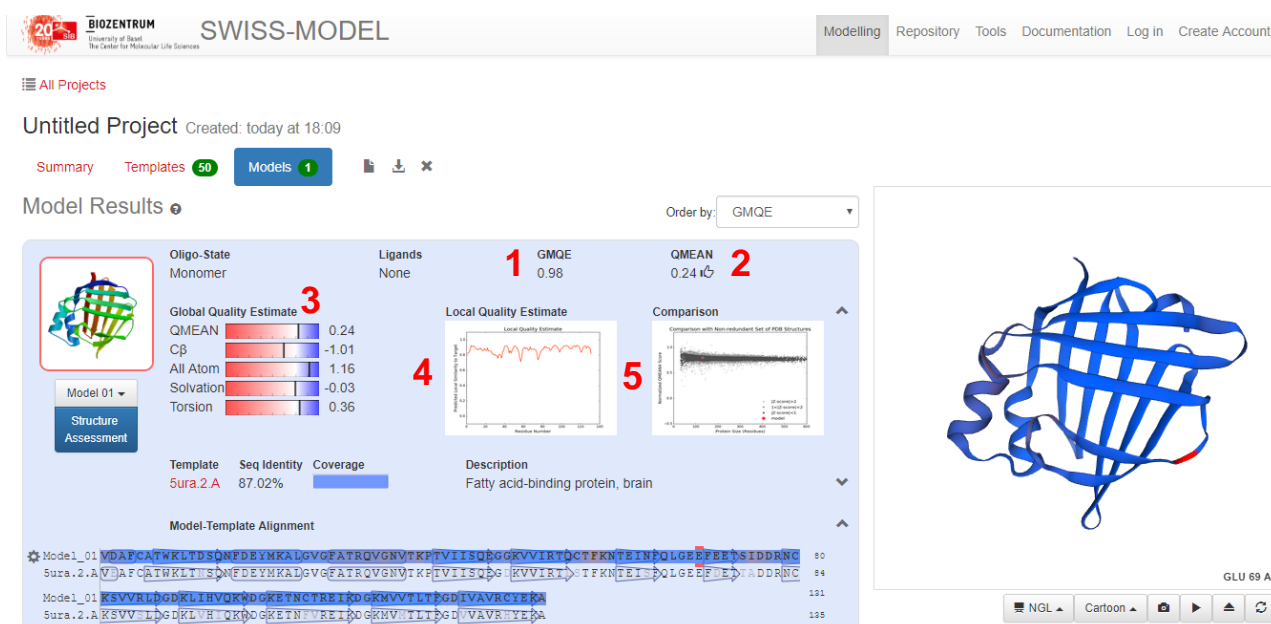
Target:	YDAFCATWKLIDSONFDEYMKALGVGFATROVGNVTKPTVIISQEGGKVVIRIQCTFKNTEINFQIGSEFEFEISIDDRNCKSVVR	85
Suza.2.A	YDAFCATWKLIDSONFDEYMKALGVGFATROVGNVTKPTVIISQEGKVVIRIQCTFKNTEINFQIGSEFEFEISIDDRNCKSVVR	85
Target:	LDGDKLIHVQRWDGKEINCTREIKDGMVILTFGDIVAVRCYERK	101
Suza.2.A	LDGDKLIHVQRWDGKEINCTREIKDGMVILTFGDIVAVRCYERK	101

Build Models 1

Clear Selection



V. Clique em Build Models.



(1) GMQE: O GMQE (Global Model Quality Estimation) é uma estimativa de qualidade que combina propriedades do alinhamento de modelo de destino e o método de pesquisa de modelo. A pontuação resultante do GMQE é expressa como **um número entre 0 e 1**, refletindo a precisão esperada de um modelo construído com esse alinhamento e modelo e a cobertura do alvo. **Números mais altos indicam maior confiabilidade**. Uma vez que um modelo é construído, o GMQE ((1) na figura acima) é atualizado para este caso específico, considerando também a pontuação QMEAN do modelo obtido, a fim de aumentar a confiabilidade da estimativa de qualidade.

(2) QMEAN: QMEAN é um estimador composto baseado em diferentes propriedades geométricas e fornece estimativas de qualidade absoluta globais (isto é, para toda estrutura) e locais (isto é, por resíduo) com base em um único modelo.

O escore Z do QMEAN ((2) na figura acima) fornece uma estimativa do "grau de natalidade" das características estruturais observadas no modelo em escala global. Indica se a pontuação QMEAN do modelo é comparável ao que se esperaria de estruturas experimentais de tamanho similar. Os escores z do QMEAN **em torno de zero indicam boa concordância** entre a estrutura do modelo e estruturas experimentais de tamanho similar. Pontuações de **-4.0 ou abaixo são uma indicação de modelos com baixa qualidade**. Isso também é

destacado por uma mudança do símbolo "**polegar para cima**" para um símbolo "**polegar para baixo**" ao lado da pontuação.

O QMEAN consiste em quatro termos individuais. Os quatro termos individuais das pontuações de qualidade globais da QMEAN também são listados ((**3**) na figura acima). A área branca nos gráficos de barras (valores numéricos próximos de zero) indica que a propriedade é semelhante ao que se esperaria de estruturas experimentais de tamanho semelhante. Valores positivos indicam que os escores do modelo são superiores às estruturas experimentais em média. Números negativos indicam que o modelo pontua abaixo das estruturas experimentais em média. A pontuação Z QMEAN em si é mostrada no topo. Os escores-Z individuais comparam o potencial de interação entre átomos de C β apenas, todos os átomos, o potencial de solvatação e o potencial de ângulo de torção. Para detalhes, por favor recorra à publicação.

(**4**) O gráfico "Qualidade Local" ((**4**) na figura acima) mostra, para cada resíduo do modelo (relatado no eixo x), a semelhança esperada com a estrutura nativa (eixo y). Normalmente, espera-se que os resíduos **com pontuação abaixo de 0,6 sejam de baixa qualidade**. Cadeias de modelos diferentes são mostradas em cores diferentes. Se o modelo for baixado, a pontuação local será informada na coluna Fator B do arquivo PDB. A qualidade local também pode ser visualizada escolhendo o esquema de cores "QMEAN".

(**5**) No gráfico "Comparação" ((**5**) na figura acima), os escores de qualidade do modelo de modelos individuais estão relacionados a pontuações obtidas para estruturas experimentais de tamanho similar. O eixo x mostra o comprimento da proteína (número de resíduos). O eixo y é a pontuação QEMAN normalizada. Cada ponto representa uma estrutura experimental de proteína. Pontos pretos são estruturas experimentais com um escore QMEAN normalizado dentro de 1 desvio padrão da média ($| Z\text{-score} |$ entre 0 e 1), estruturas experimentais com um $| Z\text{-score} |$ entre 1 e 2 são cinza. Estruturas experimentais que estão ainda mais longe da média são cinza claro. O modelo atual é representado como uma estrela vermelha. A média e o desvio padrão das estruturas experimentais em torno da localização x da estrela são a base para calcular a pontuação Z QMEAN do modelo ((**2**) na figura acima), ou seja, quantos desvios padrão das pontuações médias modelo.

VI. Avalie o modelo clicando em Structure Assessment

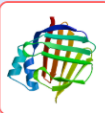
BIOZENTRUM University of Basel The Center for Molecular Life Sciences SWISS-MODEL Modelling Repository Tools Documentation Log in Create Account

All Projects


Untitled Project Created: today at 18:09

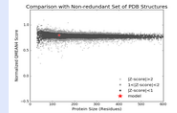
Summary Templates 50 Models 1


Model Results Order by: GMQE


Structure Assessment


Global Quality Estimate
 QMEAN 0.24
 C β -1.01
 All Atom 1.16
 Solvation -0.03
 Torsion 0.36

Local Quality Estimate


Comparison


Template Sura.2.A **Seq Identity** 87.02% **Coverage**  **Description** Fatty acid-binding protein, brain

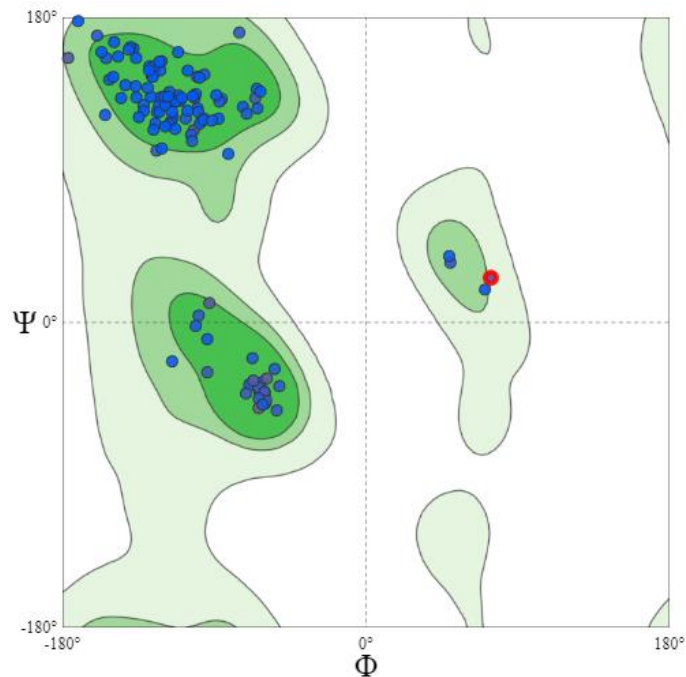
Model-Template Alignment
 Model_01 VDAFCA¹TKLIDSON²FDEYKALGVG³EATROVGNVTK⁴TVIISD⁵GGKV⁶VIRT⁷CT⁸PK⁹TEIN¹⁰POLGE¹¹FEE¹²SIDDR¹³NC¹⁴ 00
 Sura.2.A VDAFCA¹TKHLI²SD³FDEYKALGVG⁴EATROVGNVTK⁵TVIISD⁶GGKV⁷VIRT⁸CT⁹PK¹⁰TEIN¹¹POLGE¹²FEE¹³SIDDR¹⁴NC¹⁵ 04
 Model_01 KSVV¹LD²GD³KLI⁴HVQ⁵K⁶DDG⁷RE⁸NCT⁹RE¹⁰LD¹¹GG¹²KM¹³VV¹⁴TL¹⁵FG¹⁶IV¹⁷AV¹⁸RC¹⁹YE²⁰KA²¹ 101
 Sura.2.A KSVV¹LD²GD³KLV⁴H⁵Q⁶K⁷DDG⁸RE⁹NCT¹⁰RE¹¹LD¹²GG¹³KM¹⁴VV¹⁵TL¹⁶FG¹⁷IV¹⁸AV¹⁹RC²⁰YE²¹KA²² 108



GLU 69 A

NGL Cartoon

Ramachandran Plots



MolProbity Results	
MolProbity Score	1.07
Clash Score	1.45
Ramachandran Favoured	96.9%
Ramachandran Outliers	0.0%
Rotamer Outliers	0.0%
C-Beta Deviations	1 A77 ASP
Bad Bonds	0 / 1048
Bad Angles	5 / 1412 A2 ASP, A59 ASN, A76 ASP, A93 HIS, (A75 ILE-A76 ASP)

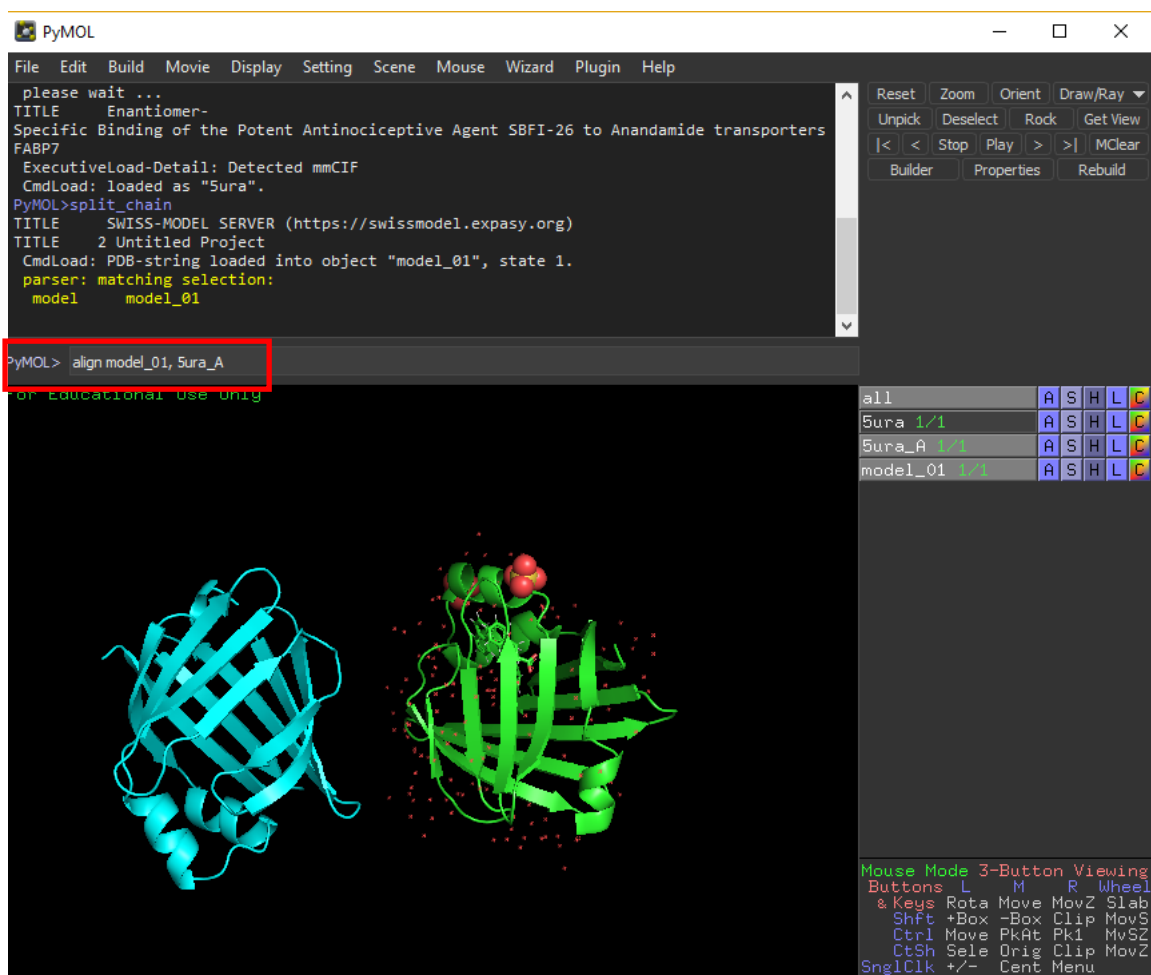
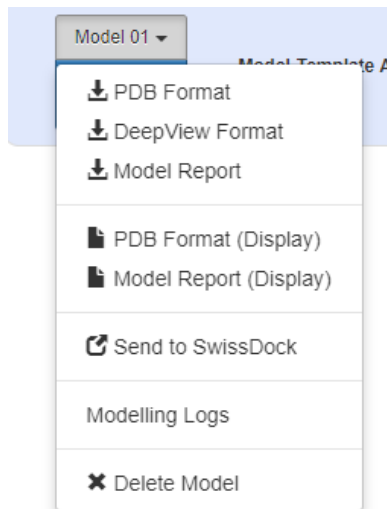
Results obtained using MolProbity in Phenix version 1.13, release 2998

Ramachandran Plots: Um gráfico de Ramachandran é uma maneira de visualizar regiões energeticamente favorecidas para ângulos diedros de backbone contra resíduos de aminoácidos na estrutura da proteína.

MolProbity: É um serviço da Web de validação de estrutura que fornece avaliação da qualidade do modelo nos níveis global e local para proteínas e ácidos nucleicos.

Escore	Comentário	Caso Ideal
MolProbity Score	Índice de qualidade de proteína combinada que reflete a resolução cristalográfica na qual tal qualidade seria esperada.	Tão baixo quanto possível
Clash Score	> 0,45 Å de não-H-bond	Zero
Ramachandran Favoured		> 98%
Ramachandran Outliers	Em resoluções abaixo de 3.0 Å, quaisquer valores discrepantes devem ser considerados erros.	< 0.2%
Rotamer Outliers	Em resoluções abaixo de 3.0 Å, quaisquer valores discrepantes devem ser considerados erros	<1%
C-Beta Deviations	Desvio de > 0,25Å do ideal	Zero
Bad Bonds	> 4 σ desvio do ideal	Zero
Bad Angles	> 4 σ desvio do ideal	Zero
Cis/Twisted Prolines/Non-Prolines	<30 ° do ideal definido como CIS; > 150 ° do ideal definido como Twisted	Zero

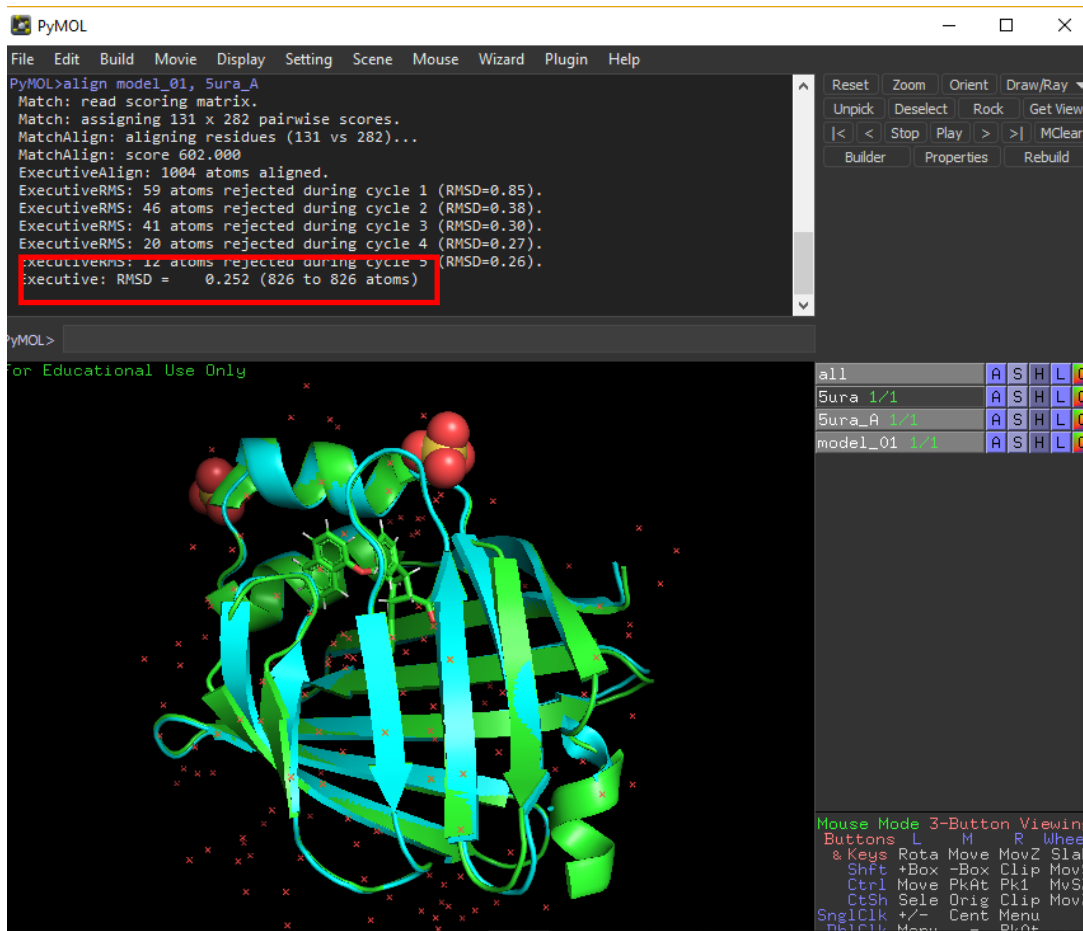
VII. Baixe o modelo. Compare o modelo com a estrutura do template (5URA) escolhido com o pymol.



Utilize o comando de alinhamento do pymol:

`align model_01, 5ura_A`

`align estrutura_a_ser_alinhar, estrutura_referencia`



Exercício:

Faça um documento (Word, libre office, etc) contendo as informações seguintes (pode ser o print das telas):

- 1) Utilize a estrutura 1FDQ como template e crie o modelo pelo SWISS-MODEL.
- 2) Avalie com o Structure Assessment (Gráfico de Ramachandran e informações de score).
- 3) Compare com os valores do modelo anterior onde o template foi a estrutura 5URA. Qual seria o melhor template?
- 4) Baixe o novo modelo (template 1FDQ) e importe no pymol junto com o modelo (template 5URA). Alinhe e crie uma figura no pymol com as possíveis regiões diferentes. Qual foi o valor de RMSD entre elas?

TUTORIAL MODELLER

Este tutorial está orientado para usuários do sistema operacional Linux e tem por objetivo exemplificar a aplicação do software Modeller para predição de estruturas de proteínas via modelagem comparativa utilizando a técnica de restrições espaciais.

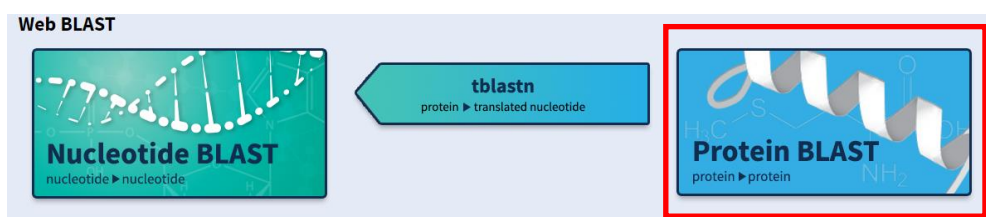
Sequência de interesse: Brain lipid-binding protein (BLBP)

```
> blbp
VDAFCATWKLTD SQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRT
QCTFKNTEINFQLGEEFEETSIDDRNCKSVVRLDGDGLIHVQKWGKETNCTR
EIKDGKMMVTLTFGDIVAVRCYEKA*
```

V.I. IDENTIFICAÇÃO DOS MOLDES

Submeter a sequência ao servidor BLAST para selecionar o(s) melhor(es) moldes (templates) baseados na identidade, cobertura e e-value.

- a) Acessar o site do NCBI: <http://blast.ncbi.nlm.nih.gov>.
- b) Selecionar a ferramenta “protein blast”.



- c) No campo “Enter Query Sequence”, colar a sequência que se deseja modelar (no formato FASTA).
- d) Buscar pela sequência usando o parâmetro: “Database”: Protein Data Bank proteins (pdb).

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```

VDFAFCATHKLTDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSID
DRNCKSVWRLDGDKLIHVQKWDGKETNCTRETKDGMVTLTFGDIVAVRCYEKA*

```

Clear Query subrange

From

To

Or, upload file Nenhum arquivo selecionado

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database **Protein Data Bank proteins(pdb)**

Organism Optional

Exclude Optional Models (XM/XP) Uncultured/environmental sample sequences

Entrez Query Optional [YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search

Program Selection

Algorithm

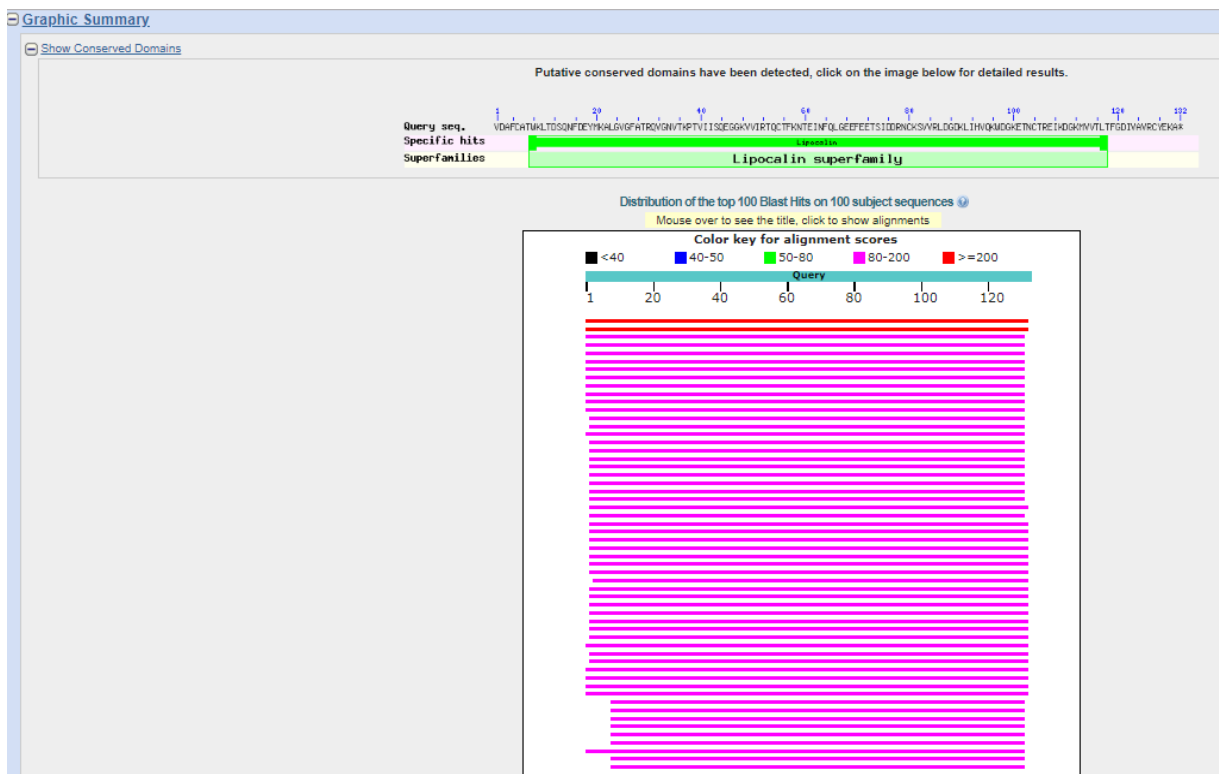
- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST Search database Protein Data Bank proteins(pdb) using Blastp (protein-protein BLAST)

Show results in a new window

e) Analisar os resultados.



Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Chain A, Enantiomer-Specific Binding of the Potent Antinociceptive Agent SBFI-26 to Anandamide transporters FABP7	238	238	99%	2e-82	87%	5URA_A
<input type="checkbox"/> Chain A, Crystal Structure Of Human Brain Fatty Acid Binding Protein	238	238	99%	2e-82	87%	1FDQ_A
<input type="checkbox"/> Chain A, NMR STUDY OF BOVINE HEART FATTY ACID BINDING PROTEIN	185	185	98%	2e-61	66%	1BWY_A
<input type="checkbox"/> Chain A, Human Fabp3 In Complex With 6-chloro-2-methyl-4-phenyl-quinoline-3- Carboxylic Acid	178	178	98%	1e-58	62%	5HZ9_A
<input type="checkbox"/> Chain A, Serial femtosecond X-ray structure of human fatty acid-binding protein type-3 (FABP3) in complex with stearic acid (C18:0) determined using X-ray free-electron laser at	178	178	98%	1e-58	62%	3WXQ_A
<input type="checkbox"/> Chain A, Solution Structure Of Human Heart-Type Fatty Acid Binding Protein	177	177	98%	2e-58	62%	1G5W_A
<input type="checkbox"/> Chain A, High Resolution X-Ray and Neutron diffraction structure of H-FABP	177	177	98%	2e-58	62%	5CF4_A

Alignments

Download ▾ GenPept Graphics ▾ Next Previous Descriptions

Chain A, Enantiomer-Specific Binding of the Potent Antinociceptive Agent SBFI-26 to Anandamide transporters FABP7
Sequence ID: [5URA_A](#) Length: 135 Number of Matches: 1
[See 3 more title\(s\)](#)

Range 1: 5 to 135 GenPept Graphics ▾ Next Match ▴ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
238 bits(607)	2e-82	Compositional matrix adjust.	114/131(87%)	123/131(93%)	0/131(0%)
Query 1	VDAFCATIKLTDSONFDEYKALGVGFATROVGNVTKPTVIIISQEGKVVIRTOCTFKNT				60
Sbjct 5	VAFACATIKLTSQNFDEYKALGVGFATROVGNVTKPTVIIISQEG KVVIRT TFKNT				64
Query 61	E1INFQGEFFFEETSIDDRNCKSVRLDGDKLIHVQKIDGKETNCTREIKDGNVWTLTFG				120
Sbjct 65	E1+FQGEFF+ET+ DDRNCKSVV LDDKLI+H+QKIDGKETN REIKDGNV+TLTFG				124
Query 121	DIIVAVRCYEKA 131				
Sbjct 125	DVAVRHYEKA 135				

Download ▾ GenPept Graphics ▾ Next Previous Descriptions

Chain A, Crystal Structure Of Human Brain Fatty Acid Binding Protein
Sequence ID: [1FDQ_A](#) Length: 131 Number of Matches: 1
[See 3 more title\(s\)](#)

Range 1: 1 to 131 GenPept Graphics ▾ Next Match ▴ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
238 bits(607)	2e-82	Compositional matrix adjust.	114/131(87%)	123/131(93%)	0/131(0%)
Query 1	VDAFCATIKLTDSONFDEYKALGVGFATROVGNVTKPTVIIISQEGKVVIRTOCTFKNT				60
Sbjct 1	VAFACATIKLTSQNFDEYKALGVGFATROVGNVTKPTVIIISQEG KVVIRT TFKNT				60
Query 61	E1INFQGEFFFEETSIDDRNCKSVRLDGDKLIHVQKIDGKETNCTREIKDGNVWTLTFG				120
Sbjct 61	E1+FQGEFF+ET+ DDRNCKSVV LDDKLI+H+QKIDGKETN REIKDGNV+TLTFG				120
Query 121	DIIVAVRCYEKA 131				
Sbjct 121	DVAVRHYEKA 131				

Questions/commen

Buscar o template escolhido no Protein Data Bank:

- Acessar o site do PDB: <http://www.rcsb.org/pdb/home/home.do>
- Buscar o template através do código PDB encontrado na etapa anterior.
- Fazer o download do arquivo PDB e FASTA: "Download Files" – FASTA Sequence e PDF File (text).

RCSB PDB 141616 Biological Macromolecular Structures Enabling Breakthroughs in Research and Education

Search by PDB ID, author, macromolecule, sequence, or ligands

Structure Summary | 3D View | Annotations | Sequence | Sequence Similarity | Structure Similarity | Experiment

Biological Assembly 1

1FDQ

CRYSTAL STRUCTURE OF HUMAN BRAIN FATTY ACID

DOI: 10.2210/pdb1FDQ/pdb

Classification: [LIPID BINDING PROTEIN](#)

Organism(s): [Homo sapiens](#)

Deposited: 2000-07-20 Released: 2001-07-25

Deposition Author(s): [Balendiran, G.K.](#)

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 2.1 Å

R-Value Free: 0.257

R-Value Work: 0.180

wwPDB Validation

Metric

Clashscore

Ramachandran outliers

Sidechain outliers

Biological Assembly 1

Biological Assembly 2

1.2%

8.7%

Worse

Better

FASTA Sequence

PDB Format

PDB Format (gz)

PDBx/mmCIF Format

PDBx/mmCIF Format (gz)

PDBML/XML Format (gz)

Display Files

Download Files

<https://www.rcsb.org/pdb/download/downloadFastFiles.do?structureIdList=1FDQ&compressionType=uncompressed>

V. II. ALINHAMENTO ENTRE O ALVO E O MOLDE

Realizar o alinhamento da sequência alvo com o molde correspondente utilizando o Modeller.

a) Três arquivos são necessários:

1. Sequência alvo em formato PIR (blbp.seq):

```
>P1;blbp
sequence:blbp: :::::::
VDAFCATWKLTDSONFDEYMKALGVGFATRQVGNVTKPTVVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETSIDDRNCKSVV
RLDGDGLIHVQKWDGKETNCTREIKDGKMMVTLTFGDIVAVRCYEKA*
```

Utilize o **servidor**:

<http://sequenceconversion.bugaco.com/converter/biology/sequences/>

Sequence conversion

Provided by bugaco.com

Ad closed by Google

[Report this ad](#) [Why this ad?](#)

Ad closed by Google

[Report this ad](#) [Why this ad?](#)

Convert from:

fasta

to:

pir

Alphabet: None DNA RNA Protein Nucleotide

[Escolher arquivo](#) Nenhum arquivo selecionado

Format	About format
abi	Reads the ABI "Sanger" capillary sequence traces files, including the PHRED quality scores for the base calls. This allows ABI to FASTQ conversion. Note each ABI file contains one and only one sequence (so there is no point in indexing the file).

Ou **Script** em linguagem **python** (converter_pir.py) para converter:

```

from modeller import *
e = environ()
a = alignment(e, file='blbp.fasta', alignment_format='FASTA')
a.write(file='blbp.seq', alignment_format='PIR')

```

2. Arquivo PDB da estrutura molde (1FDQ.pdb):

```

ATOM      3  C   VAL  A   1      23.003  19.471   0.304   1.00  25.86      C
ATOM      4  O   VAL  A   1      21.862  19.576   0.769   1.00  25.86      O
ATOM      5  CB  VAL  A   1      24.110  17.249   0.671   1.00  25.86      C
ATOM      6  CG1 VAL  A   1      25.469  17.823   1.095   1.00  25.86      C
ATOM      7  CG2 VAL  A   1      23.176  17.070   1.863   1.00  25.86      C
ATOM      8  N   GLU  A   2      23.897  20.454   0.368   1.00  25.86      N
ATOM      9  CA  GLU  A   2      23.538  21.731   0.960   1.00  25.86      C
ATOM     10  C   GLU  A   2      22.438  22.326   0.064   1.00  25.86      C
ATOM     11  O   GLU  A   2      21.552  23.045   0.530   1.00  25.86      O
ATOM     12  CB  GLU  A   2      24.761  22.661   1.023   1.00  25.86      C
ATOM     13  CG  GLU  A   2      25.969  22.052   1.743   1.00  25.86      C
ATOM     14  CD  GLU  A   2      26.801  21.136   0.837   1.00  25.86      C
ATOM     15  OE1 GLU  A   2      27.292  20.089   1.335   1.00  25.86      O
ATOM     16  OE2 GLU  A   2      26.975  21.476  -0.367   1.00  25.86      O
ATOM     17  N   ALA  A   3      22.487  22.001  -1.221   1.00  25.86      N
ATOM     18  CA  ALA  A   3      21.481  22.485  -2.162   1.00  25.86      C
ATOM     19  C   ALA  A   3      20.055  22.043  -1.779   1.00  25.86      C
ATOM     20  O   ALA  A   3      19.080  22.765  -2.040   1.00  25.86      O
ATOM     21  CB  ALA  A   3      21.817  21.995  -3.576   1.00  25.86      C
ATOM     22  N   PHE  A   4      19.935  20.867  -1.163   1.00  25.86      N
ATOM     23  CA  PHE  A   4      18.630  20.316  -0.752   1.00  25.86      C
ATOM     24  C   PHE  A   4      18.246  20.692   0.692   1.00  25.86      C
ATOM     25  O   PHE  A   4      17.152  20.362   1.169   1.00  25.86      O
ATOM     26  CB  PHE  A   4      18.663  18.787  -0.848   1.00  25.86      C
ATOM     27  CG  PHE  A   4      18.986  18.262  -2.214   1.00  25.86      C
ATOM     28  CD1 PHE  A   4      18.007  18.196  -3.196   1.00  25.86      C
ATOM     29  CD2 PHE  A   4      20.272  17.823  -2.520   1.00  25.86      C
ATOM     30  CE1 PHE  A   4      18.296  17.699  -4.459   1.00  25.86      C
ATOM     31  CE2 PHE  A   4      20.569  17.325  -3.788   1.00  25.86      C
ATOM     32  CZ  PHE  A   4      19.575  17.263  -4.760   1.00  25.86      C
-----

```

3. Script em linguagem python para alinhamento (align.py):

```

# Example for: alignment.align()
# This will read two sequences, align them, and write the alignment
# to a file:
log.verbose()
env = environ()
aln = alignment(env)
mdl = model(env, file='1FDQ', model_segment=('FIRST:A', 'LAST:A'))
aln.append_model(mdl, align_codes='1FDQ')
aln.append(file='blbp.seq', align_codes=('blbp'))
# The as1.sim.mat similarity matrix is used by default:
aln.align(gap_penalties_1d=(-600, -400))
aln.write(file='blbp-1FDQ.ali', alignment_format='PIR')
aln.write(file='blbp-1FDQ.pap', alignment_format='PAP')

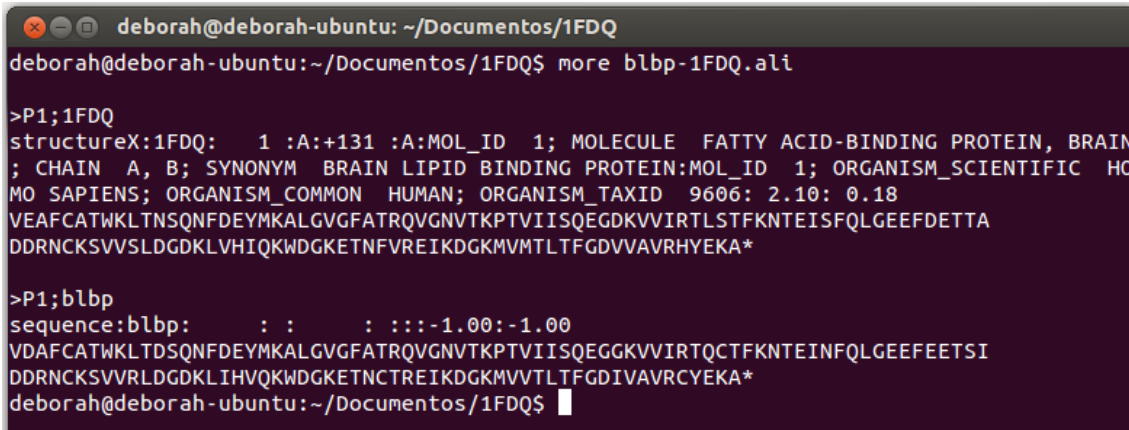
```

b) Através do Terminal rodar o script align.py no Modeller:

\$ mod9.12 align.py

c) Dois arquivos são gerados:

a. blbp-1FDQ.ali – arquivo com o alinhamento para próxima etapa.

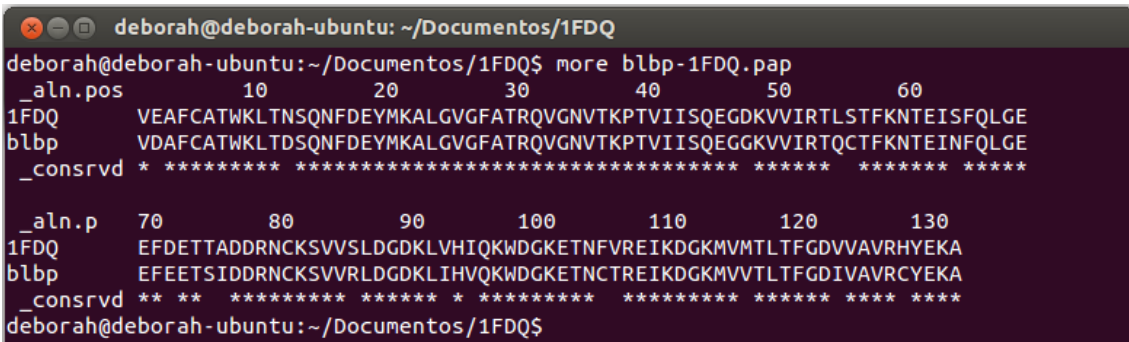


```
deborah@deborah-ubuntu: ~/Documentos/1FDQ
deborah@deborah-ubuntu:~/Documentos/1FDQ$ more blbp-1FDQ.ali

>P1;1FDQ
structureX:1FDQ:  1 :A:+131 :A:MOL_ID 1; MOLECULE  FATTY ACID-BINDING PROTEIN, BRAIN
; CHAIN  A, B; SYNONYM  BRAIN LIPID BINDING PROTEIN:MOL_ID 1; ORGANISM_SCIENTIFIC HO
MO SAPIENS; ORGANISM_COMMON  HUMAN; ORGANISM_TAXID  9606: 2.10: 0.18
VEAFCATWKL TNSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGDKVVIRT LSTFKNTEISFQLGEEFDETTA
DDRNCKSVVSLDGDGKLVHIQKWDGKETNFVREIKDGMVMTLTFGDVVAVRHYEKA*

>P1;blbp
sequence:blbp:      : :      : :::-1.00:-1.00
VDAFCATWKL TDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGEEFEETS I
DDRNCKSVVRLDGDGKLIHVQKWDGKETNCTREIKDGMVVTLTFGDIVAVRHYEKA*
deborah@deborah-ubuntu:~/Documentos/1FDQ$
```

b. blbp-1FDQ.pap – arquivo com alinhamento em que mostra os resíduos conservados.



```
deborah@deborah-ubuntu: ~/Documentos/1FDQ
deborah@deborah-ubuntu:~/Documentos/1FDQ$ more blbp-1FDQ.pap
 _aln.pos    10      20      30      40      50      60
1FDQ        VEAFCATWKL TNSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGDKVVIRT LSTFKNTEISFQLGE
blbp        VDAFCATWKL TDSQNFDEYMKALGVGFATRQVGNVTKPTVIISQEGGKVVIRTQCTFKNTEINFQLGE
 _consvrd   * ***** ***** ***** ***** ***** ***** *****

 _aln.p     70      80      90      100     110     120     130
1FDQ        EFDETTADDRNCKSVVSLDGDGKLVHIQKWDGKETNFVREIKDGMVMTLTFGDVVAVRHYEKA
blbp        EFEETSIDDRNCKSVVRLDGDGKLIHVQKWDGKETNCTREIKDGMVVTLTFGDIVAVRHYEKA
 _consvrd   ** ** ***** ***** * ***** ***** ***** ***** *****
deborah@deborah-ubuntu:~/Documentos/1FDQ$
```

III. CONSTRUÇÃO DO MODELO

Construção do modelo utilizando o resultado do alinhamento (blbp-1FDQ.ali).

a) Três arquivos são necessários:

1. Alinhamento entre sequência de interesse e estrutura molde (blbp-1FDQ.ali).
2. Arquivo PDB da estrutura molde (1FDQ.pdb).

ATOM	3	C	VAL	A	1	23.003	19.471	0.304	1.00	25.86	C
ATOM	4	O	VAL	A	1	21.862	19.576	0.769	1.00	25.86	O
ATOM	5	CB	VAL	A	1	24.110	17.249	0.671	1.00	25.86	C
ATOM	6	CG1	VAL	A	1	25.469	17.823	1.095	1.00	25.86	C
ATOM	7	CG2	VAL	A	1	23.176	17.070	1.863	1.00	25.86	C
ATOM	8	N	GLU	A	2	23.897	20.454	0.368	1.00	25.86	N
ATOM	9	CA	GLU	A	2	23.538	21.731	0.960	1.00	25.86	C
ATOM	10	C	GLU	A	2	22.438	22.326	0.064	1.00	25.86	C
ATOM	11	O	GLU	A	2	21.552	23.045	0.530	1.00	25.86	O
ATOM	12	CB	GLU	A	2	24.761	22.661	1.023	1.00	25.86	C
ATOM	13	CG	GLU	A	2	25.969	22.052	1.743	1.00	25.86	C
ATOM	14	CD	GLU	A	2	26.801	21.136	0.837	1.00	25.86	C
ATOM	15	OE1	GLU	A	2	27.292	20.089	1.335	1.00	25.86	O
ATOM	16	OE2	GLU	A	2	26.975	21.476	-0.367	1.00	25.86	O
ATOM	17	N	ALA	A	3	22.487	22.001	-1.221	1.00	25.86	N
ATOM	18	CA	ALA	A	3	21.481	22.485	-2.162	1.00	25.86	C
ATOM	19	C	ALA	A	3	20.055	22.043	-1.779	1.00	25.86	C
ATOM	20	O	ALA	A	3	19.080	22.765	-2.040	1.00	25.86	O
ATOM	21	CB	ALA	A	3	21.817	21.995	-3.576	1.00	25.86	C
ATOM	22	N	PHE	A	4	19.935	20.867	-1.163	1.00	25.86	N
ATOM	23	CA	PHE	A	4	18.630	20.316	-0.752	1.00	25.86	C
ATOM	24	C	PHE	A	4	18.246	20.692	0.692	1.00	25.86	C
ATOM	25	O	PHE	A	4	17.152	20.362	1.169	1.00	25.86	O
ATOM	26	CB	PHE	A	4	18.663	18.787	-0.848	1.00	25.86	C
ATOM	27	CG	PHE	A	4	18.986	18.262	-2.214	1.00	25.86	C
ATOM	28	CD1	PHE	A	4	18.007	18.196	-3.196	1.00	25.86	C
ATOM	29	CD2	PHE	A	4	20.272	17.823	-2.520	1.00	25.86	C
ATOM	30	CE1	PHE	A	4	18.296	17.699	-4.459	1.00	25.86	C
ATOM	31	CE2	PHE	A	4	20.569	17.325	-3.788	1.00	25.86	C
ATOM	32	CZ	PHE	A	4	19.575	17.263	-4.760	1.00	25.86	C
-----	--	--	-----	-	-	-----	-----	-----	-----	-----	--

3. Script em linguagem python para construção do modelo (model.py).

```
# Homology modelling by the automodel class
from modeller.automodel import * # Load the automodel class
log.verbose() # request verbose output
env = environ() # create a new MODELLER environment
# directories for input atom files
env.io.atom_files_directory = './:../atom_files'
a = automodel(env,
  alnfile = 'blbp-1FDQ.ali', # alignment filename
  knowns = '1FDQ', # codes of the templates
  sequence = 'blbp') # code of the target
a.starting_model = 1 # index of the first model
a.ending_model = 10 # index of the last model
# (determines how many models to calculate)
a.make() # do the actual homology modelling
```

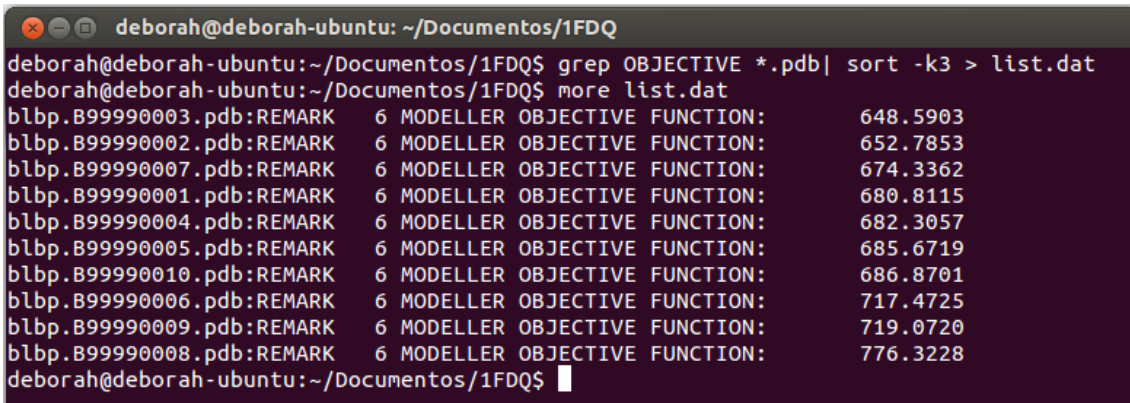
b) Através do Terminal rodar o script model.py no Modeller:

```
$ mod9.12 model.py
```

c) Dez modelos foram gerados. Deve-se escolher aquele com menor função objetiva. Através do terminal digitar o seguinte comando:

```
$ grep OBJECTIVE *.pdb| sort -k3 > list.dat
```

```
$ more list.dat
```



```
deborah@deborah-ubuntu: ~/Documentos/1FDQ
deborah@deborah-ubuntu:~/Documentos/1FDQ$ grep OBJECTIVE *.pdb| sort -k3 > list.dat
deborah@deborah-ubuntu:~/Documentos/1FDQ$ more list.dat
blbp.B99990003.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      648.5903
blbp.B99990002.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      652.7853
blbp.B99990007.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      674.3362
blbp.B99990001.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      680.8115
blbp.B99990004.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      682.3057
blbp.B99990005.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      685.6719
blbp.B99990010.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      686.8701
blbp.B99990006.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      717.4725
blbp.B99990009.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      719.0720
blbp.B99990008.pdb:REMARK      6 MODELLER OBJECTIVE FUNCTION:      776.3228
deborah@deborah-ubuntu:~/Documentos/1FDQ$
```

d) Verificar o modelo gerado com menor função objetiva e visualizar a estrutura pelo PyMOL.

```
$ pymol blbp.B99990003.pdb
```

The PyMOL Molecular Graphics System

File Edit Build Movie Display Setting Scene Mouse Wizard Plugin Help

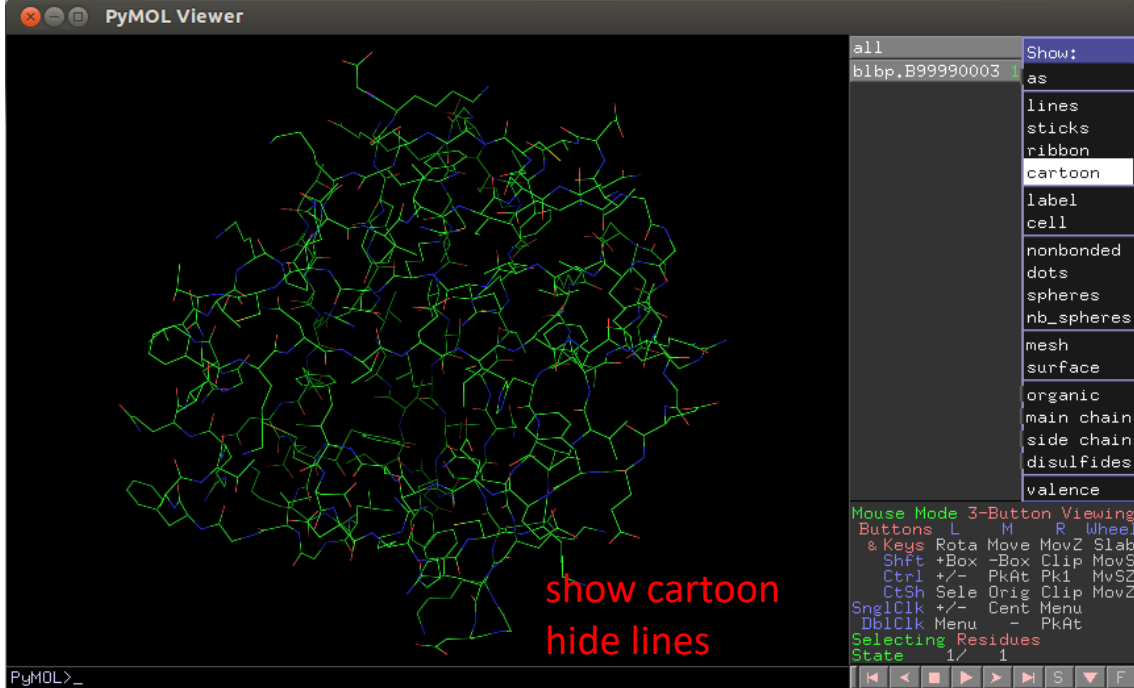
```

plugins
  builtin . import (mod_name)
  File "/usr/lib/python2.7/dist-packages/pmgTk/startup/autodock.py", line 48, in <mo
  dule>
    from numpy import *
  ImportError: No module named numpy
  Error: unable to initialize plugin 'autodock'.
PyMOL>

```

Reset Zoom Orient Draw Ray
 Unpick Deselect Rock Get View
 < < Stop Play > > MClear
 Command Builder Volume
 Rebuild Abort

PyMOL Viewer

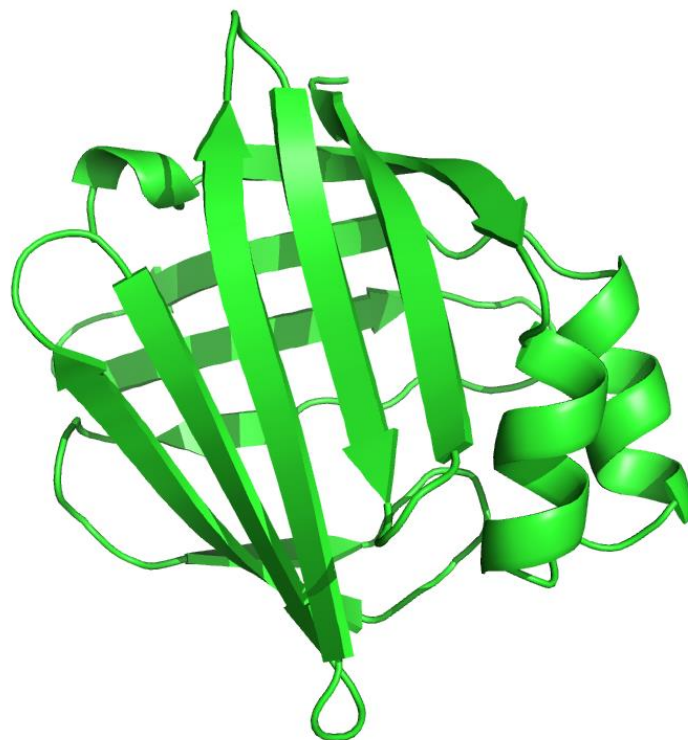


all Show:
 blbp.B99990003 as
 lines
 sticks
 ribbon
 cartoon
 label
 cell
 nonbonded
 dots
 spheres
 nb_spheres
 mesh
 surface
 organic
 main chain
 side chain
 disulfides
 valence

Mouse Mode 3-Button Viewing
 Buttons L M R Wheel
 & Keys Rota Move MovZ Slab
 Shift +Box -Box Clip MovS
 Ctrl +/- PkAt Pk1 MvS
 CtrSh Sele Orig Clip MovN
 SnglClk +/- Cent Menu
 DblClk Menu - PkAt
 Selecting Residues
 State 1/ 1

show cartoon
 hide lines

PyMOL>_



IV. VALIDAÇÃO DO MODELO

Avaliar a qualidade do modelo gerado utilizando o servidor SAVES.

a) Acessar o servidor SAVES: <http://servicesn.mbi.ucla.edu/SAVES/>

[WHATCHECK](#) • [PROCHECK](#) • [ERRAT](#) • [Verify3D](#) • [PROVE](#) • [CRYST](#) • [pdbU](#) (?)

Upload your structure file here

Note: Only submit PDB files, not FASTA

Nenhum arquivo selecionado

<input type="button" value="Verify 3D"/>	<input checked="" type="checkbox"/>
<input type="button" value="ERRAT"/>	<input checked="" type="checkbox"/>
<input type="button" value="Prove"/>	<input checked="" type="checkbox"/>
<input type="button" value="PROCHECK"/>	<input checked="" type="checkbox"/>
<input type="button" value="WHATCHECK"/>	<input checked="" type="checkbox"/>
<input type="button" value="CRYST"/>	<input type="checkbox"/>

[verify3d](#) [errata](#) [prove](#) [procheck](#) [whatcheck](#)
[?](#)

b) No campo “Escolher arquivo” selecionar o arquivo PDB com menor função objetiva.

c) Rodar todos os programas – Run all programs:

1. Procheck
2. What_check
3. Verify_3D
4. Errat
5. Prove

d) Analisar Resultados.

VI. BIBLIOGRAFIA CONSULTADA

BOWIE J. U.; LÜTHY R.; EISENBERG D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 12;253(5016):164-70, 1991.

COLOVOS C.; YEATES T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 2(9):1511-9, 1993.

DOS SANTOS, A. D. Curso de Verão 2014. Modelagem Molecular. 2014.

LASKOWSKI R. A.; MACARTHUR M. W.; MOSS D. S.; THORNTON J. M., PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.*, 26, 283-291, 1993.

LÜTHY R.; BOWIE J. U.; Eisenberg D. Assessment of protein models with three dimensional profiles. *Nature*. 5;356(6364):83-5. 1992.

MODELLER, A programar for Protein Structure Modeling. Release 9.14, r10167. <https://salilab.org/modeller/manual/>. Acessado em 12/01/2014.

SANTOS FILHO, Osvaldo Andrade and ALENCASTRO, Ricardo Bicca de. Modelagem de proteínas por homologia. *Quím. Nova* [online]. 2003, vol.26, n.2, pp. 253-259. ISSN 0100-4042. <http://dx.doi.org/10.1590/S0100-40422003000200019>.

SILVA, V. B.; SILVA, C. H. T. P. Modelagem molecular de proteínas-alvo por homologia estrutural. *Revista Eletrônica de Farmácia*. Vol IV (1), 15-26, 2007.

VERLI, H. Bioinformática da Biologia à flexibilidade molecular. Ed. Porto Alegre, 282 p. , 2014.