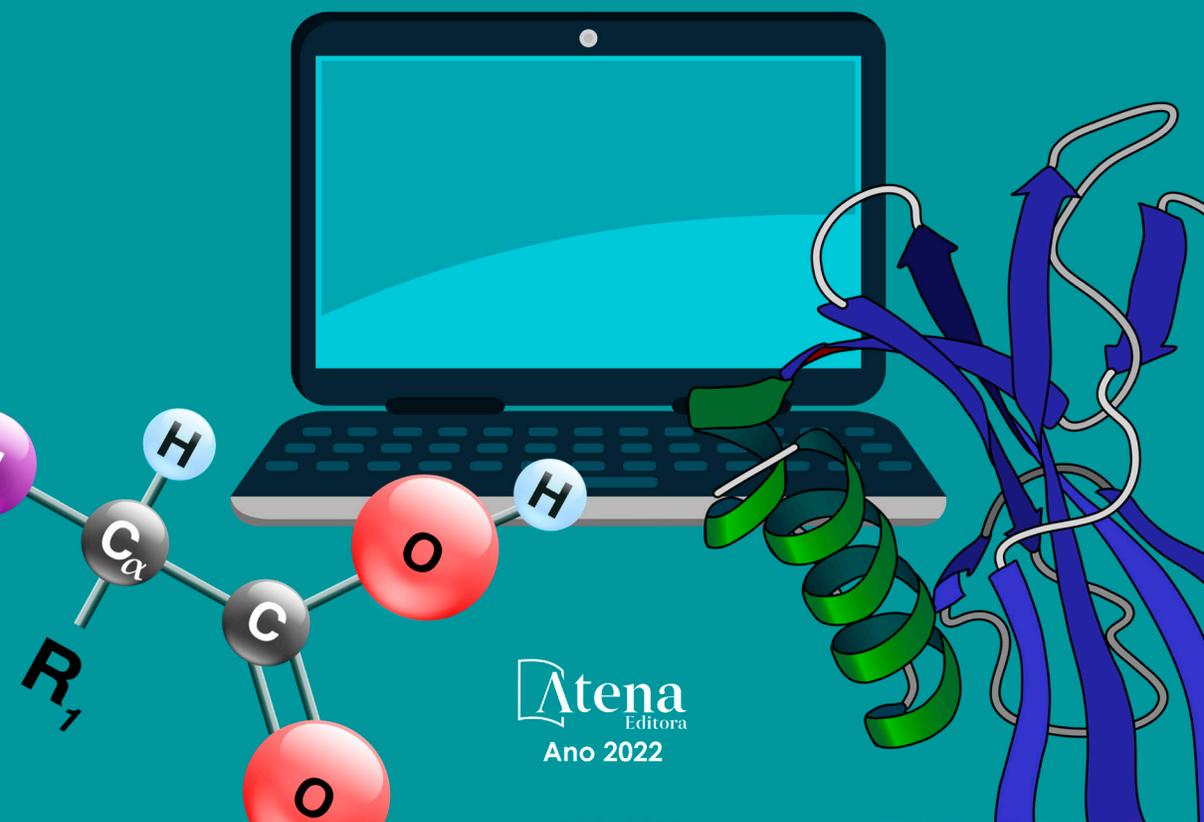


MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

Editado por Kauê Santana e Cláudio Nahum

UM GUIA TEÓRICO E PRÁTICO



Atena
Editora
Ano 2022

MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

Editado por Kauê Santana e Cláudio Nahum

UM GUIA TEÓRICO E PRÁTICO



Atena
Editora
Ano 2022

Editora chefe

Profª Drª Antonella Carvalho de Oliveira

Editora executiva

Natalia Oliveira

Assistente editorial

Flávia Roberta Barão

Bibliotecária

Janaina Ramos

Projeto gráfico

Bruno Oliveira

Camila Alves de Cremo

Luiza Alves Batista

Natália Sandrini de Azevedo

Imagens da capa

iStock

Edição de arte

Luiza Alves Batista

2022 by Atena Editora

Copyright © Atena Editora

Copyright do texto © 2022 Os autores

Copyright da edição © 2022 Atena Editora

Direitos para esta edição cedidos à Atena Editora pelos autores.

Open access publication by Atena Editora



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons. Atribuição-Não-Comercial-Não-Derivativos 4.0 Internacional (CC BY-NC-ND 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores, inclusive não representam necessariamente a posição oficial da Atena Editora. Permitido o *download* da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

A Atena Editora é comprometida em garantir a integridade editorial em todas as etapas do processo de publicação, evitando plágio, dados ou resultados fraudulentos e impedindo que interesses financeiros comprometam os padrões éticos da publicação. Situações suspeitas de má conduta científica serão investigadas sob o mais alto padrão de rigor acadêmico e ético.

Conselho Editorial**Ciências Biológicas e da Saúde**

Profª Drª Aline Silva da Fonte Santa Rosa de Oliveira – Hospital Federal de Bonsucesso

Profª Drª Ana Beatriz Duarte Vieira – Universidade de Brasília

Profª Drª Ana Paula Peron – Universidade Tecnológica Federal do Paraná

Prof. Dr. André Ribeiro da Silva – Universidade de Brasília

Profª Drª Anelise Levay Murari – Universidade Federal de Pelotas

Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás



Prof. Dr. Cirênio de Almeida Barbosa – Universidade Federal de Ouro Preto
Prof^o Dr^a Daniela Reis Joaquim de Freitas – Universidade Federal do Piauí
Prof^o Dr^a Débora Luana Ribeiro Pessoa – Universidade Federal do Maranhão
Prof. Dr. Douglas Siqueira de Almeida Chaves – Universidade Federal Rural do Rio de Janeiro
Prof. Dr. Edson da Silva – Universidade Federal dos Vales do Jequitinhonha e Mucuri
Prof^o Dr^a Elizabeth Cordeiro Fernandes – Faculdade Integrada Medicina
Prof^o Dr^a Eleuza Rodrigues Machado – Faculdade Anhanguera de Brasília
Prof^o Dr^a Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina
Prof^o Dr^a Eysler Gonçalves Maia Brasil – Universidade da Integração Internacional da Lusofonia Afro-Brasileira
Prof. Dr. Ferlando Lima Santos – Universidade Federal do Recôncavo da Bahia
Prof^o Dr^a Fernanda Miguel de Andrade – Universidade Federal de Pernambuco
Prof. Dr. Fernando Mendes – Instituto Politécnico de Coimbra – Escola Superior de Saúde de Coimbra
Prof^o Dr^a Gabriela Vieira do Amaral – Universidade de Vassouras
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. Helio Franklin Rodrigues de Almeida – Universidade Federal de Rondônia
Prof^o Dr^a Iara Lúcia Tescarollo – Universidade São Francisco
Prof. Dr. Igor Luiz Vieira de Lima Santos – Universidade Federal de Campina Grande
Prof. Dr. Jefferson Thiago Souza – Universidade Estadual do Ceará
Prof. Dr. Jesus Rodrigues Lemos – Universidade Federal do Piauí
Prof. Dr. Jônatas de França Barros – Universidade Federal do Rio Grande do Norte
Prof. Dr. José Aderval Aragão – Universidade Federal de Sergipe
Prof. Dr. José Max Barbosa de Oliveira Junior – Universidade Federal do Oeste do Pará
Prof^o Dr^a Juliana Santana de Curcio – Universidade Federal de Goiás
Prof^o Dr^a Lívia do Carmo Silva – Universidade Federal de Goiás
Prof. Dr. Luís Paulo Souza e Souza – Universidade Federal do Amazonas
Prof^o Dr^a Magnólia de Araújo Campos – Universidade Federal de Campina Grande
Prof. Dr. Marcus Fernando da Silva Praxedes – Universidade Federal do Recôncavo da Bahia
Prof^o Dr^a Maria Tatiane Gonçalves Sá – Universidade do Estado do Pará
Prof. Dr. Maurilio Antonio Varavallo – Universidade Federal do Tocantins
Prof^o Dr^a Mylena Andréa Oliveira Torres – Universidade Ceuma
Prof^o Dr^a Natiéli Piovesan – Instituto Federaci do Rio Grande do Norte
Prof. Dr. Paulo Inada – Universidade Estadual de Maringá
Prof. Dr. Rafael Henrique Silva – Hospital Universitário da Universidade Federal da Grande Dourados
Prof^o Dr^a Regiane Luz Carvalho – Centro Universitário das Faculdades Associadas de Ensino
Prof^o Dr^a Renata Mendes de Freitas – Universidade Federal de Juiz de Fora
Prof^o Dr^a Sheyla Mara Silva de Oliveira – Universidade do Estado do Pará
Prof^o Dr^a Suely Lopes de Azevedo – Universidade Federal Fluminense
Prof^o Dr^a Vanessa da Fontoura Custódio Monteiro – Universidade do Vale do Sapucaí
Prof^o Dr^a Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Prof^o Dr^a Vanessa Bordin Viera – Universidade Federal de Campina Grande
Prof^o Dr^a Welma Emídio da Silva – Universidade Federal Rural de Pernambuco



Métodos de modelagem da estrutura tridimensional de proteínas: um guia teórico e prático

Diagramação: Camila Alves de Cremo
Correção: Maiara Ferreira
Indexação: Amanda Kelly da Costa Veiga
Revisão: RevisAtena
Organizadores: Kauê Santana
Cláudio Nahum

Dados Internacionais de Catalogação na Publicação (CIP)

M593 Métodos de modelagem da estrutura tridimensional de proteínas: um guia teórico e prático / Organizadores Kauê Santana, Cláudio Nahum. – Ponta Grossa - PR: Atena, 2022.

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-65-258-0582-5

DOI: <https://doi.org/10.22533/at.ed.825222810>

1. Proteínas. I. Santana, Kauê (Organizador). II. Nahum, Cláudio (Organizador). III. Título.

CDD 613.282

Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166

Atena Editora
Ponta Grossa – Paraná – Brasil
Telefone: +55 (42) 3323-5493
www.atenaeditora.com.br
contato@atenaeditora.com.br



DECLARAÇÃO DOS AUTORES

Os autores desta obra: 1. Atestam não possuir qualquer interesse comercial que constitua um conflito de interesses em relação ao artigo científico publicado; 2. Declaram que participaram ativamente da construção dos respectivos manuscritos, preferencialmente na: a) Concepção do estudo, e/ou aquisição de dados, e/ou análise e interpretação de dados; b) Elaboração do artigo ou revisão com vistas a tornar o material intelectualmente relevante; c) Aprovação final do manuscrito para submissão.; 3. Certificam que os artigos científicos publicados estão completamente isentos de dados e/ou resultados fraudulentos; 4. Confirmam a citação e a referência correta de todos os dados e de interpretações de dados de outras pesquisas; 5. Reconhecem terem informado todas as fontes de financiamento recebidas para a consecução da pesquisa; 6. Autorizam a edição da obra, que incluem os registros de ficha catalográfica, ISBN, DOI e demais indexadores, projeto visual e criação de capa, diagramação de miolo, assim como lançamento e divulgação da mesma conforme critérios da Atena Editora.



DECLARAÇÃO DA EDITORA

A Atena Editora declara, para os devidos fins de direito, que: 1. A presente publicação constitui apenas transferência temporária dos direitos autorais, direito sobre a publicação, inclusive não constitui responsabilidade solidária na criação dos manuscritos publicados, nos termos previstos na Lei sobre direitos autorais (Lei 9610/98), no art. 184 do Código Penal e no art. 927 do Código Civil; 2. Autoriza e incentiva os autores a assinarem contratos com repositórios institucionais, com fins exclusivos de divulgação da obra, desde que com o devido reconhecimento de autoria e edição e sem qualquer finalidade comercial; 3. Todos os e-book são *open access*, *desta forma* não os comercializa em seu site, sites parceiros, plataformas de *e-commerce*, ou qualquer outro meio virtual ou físico, portanto, está isenta de repasses de direitos autorais aos autores; 4. Todos os membros do conselho editorial são doutores e vinculados a instituições de ensino superior públicas, conforme recomendação da CAPES para obtenção do Qualis livro; 5. Não cede, comercializa ou autoriza a utilização dos nomes e e-mails dos autores, bem como nenhum outro dado dos mesmos, para qualquer finalidade que não o escopo da divulgação desta obra.



Dedico este livro à minha esposa, Lidiane Diniz, pois sem a sua dedicação e amor não seria possível concluir este trabalho.

Prof. Dr. Kauê Santana

APRESENTAÇÃO

A predição da estrutura de proteínas é, sem dúvida, um assunto-chave para quem atua nas áreas de modelagem molecular, bioinformática e biologia estrutural, dada a versatilidade de funções desempenhada por estes biopolímeros em processos biológicos, assim como suas diferentes aplicações farmacêuticas, industriais e biotecnológicas. Com este livro, pretendemos não somente atender às necessidades práticas relacionadas à modelagem e à análise das estruturas tridimensionais de proteínas, conteúdo ainda carente em língua portuguesa, mas também fornecer os principais conceitos necessários para a compreensão dos métodos, suas aplicações, assim como limitações. Focamos o conteúdo e a linguagem do texto para a graduação e devido ao grande número de expressões da língua inglesa presentes na literatura científica, incluímos os termos mais comumente utilizados dessa língua, sempre dando preferência à tradução deles para o português.

O livro traz uma linguagem acessível, incluindo referências atualizadas sobre os métodos, além de estar ricamente ilustrado, ainda traz, no final, um glossário de termos técnicos que enriquecerão a leitura e a compreensão do leitor.

Temos certeza que será uma importante fonte de consulta para alunos de graduação e pós-graduação, assim como para todos os interessados na área.

Prof. Kauê Santana
Prof. Cláudio Nahum

PREFÁCIO

A obra intitulada *Métodos de Modelagem da Estrutura Tridimensional de Proteínas: Um Guia Teórico-prático*, produzido por Kauê Santana e Cláudio Nahum, faz um apanhado geral das técnicas e abordagens computacionais mais avançadas usadas na predição e estudo das estruturas das proteínas. O livro está organizado de forma que ao leitor, inicialmente, sejam apresentados aos conceitos básicos inerentes aos métodos usados em diversos programas e algoritmos empregados na modelagem. Apesar de o embasamento teórico oferecido ser bastante amplo, a leitura do texto não se torna densa, isso faz com que o leitor, mesmo pertencendo ao público leigo, não tenha dificuldade em avançar no entendimento.

A segunda parte do livro traz um manual aplicado com um detalhamento muito rico e pontual de como proceder para a obtenção de uma determinada estrutura tridimensional. Notadamente, a parte de aplicação do livro mostra tudo o que é necessário para que aqueles que se aventuram nas tarefas de modelar e avaliar as estruturas tridimensionais, sintam-se bastante seguros quanto ao manuseio dos parâmetros e interpretação dos resultados obtidos através deste guia prático.

Na última parte do livro, os autores mostram diversas aplicações da modelagem computacional em distintos contextos da biologia. Vários exemplos são apresentados através de artigos científicos, o que se configura como uma forma de ampliar os horizontes do leitor quanto à vasta gama de possibilidades de aplicação dos métodos de modelagem. Além disso, os exemplos de aplicabilidade servem para instigar os leitores mais audazes a se aventurarem na laboriosa tarefa de predizer a estrutura de uma proteína.

Particularmente, o último capítulo do livro é bastante interessante, pois além de mostrar as perspectivas da predição tridimensional, ele discute, de forma geral, as atuais limitações. Os autores discursem detalhar um método em particular e trazem à tona uma faceta da modelagem que é seu carácter de aproximação. Isso não poderia ser diferente, pois a modelagem computacional é uma inferência que, como tal, nos proporciona uma forma de idealizar uma entidade biológica desconhecida que é a real estrutura da proteína.

Em suma o livro *Métodos de Modelagem da Estrutura Tridimensional de Proteínas: Um Guia Teórico-prático* é um excelente referencial teórico-prático de fácil leitura que está sendo oferecido ao público acadêmico na língua portuguesa.

Prof. PhD. Élcio de Souza Leal
Docente do Instituto de Ciências Biológicas, Universidade Federal do Pará

SUMÁRIO

CAPÍTULO 1..... 1

A ESTRUTURA E O PROBLEMA DE DOBRAMENTO DE PROTEÍNAS

Kauê Santana da Costa

Anderson Henrique Lima e Lima

Jerônimo Lameira

 <https://doi.org/10.22533/at.ed.8252228101>

CAPÍTULO 2..... 7

MÉTODOS DE PREDIÇÃO POR HOMOLOGIA

Kauê Santana da Costa

João Marcos Pereira Galúcio

José Rogério de Araújo Silva

 <https://doi.org/10.22533/at.ed.8252228102>

CAPÍTULO 3..... 16

MODELAGEM DE PROTEÍNAS NO *MODELLER*

Kauê Santana da Costa

João Marcos Pereira Galúcio

 <https://doi.org/10.22533/at.ed.8252228103>

CAPÍTULO 4..... 29

MODELAGEM DE PROTEÍNAS NO *EASYMODELLER*

João Marcos Pereira Galúcio

 <https://doi.org/10.22533/at.ed.8252228104>

CAPÍTULO 5..... 33

MÉTODOS DE PREDIÇÃO *THREADING* E *AB INITIO*

Anderson Henrique Lima e Lima

Kauê Santana da Costa

Alberto Monteiro dos Santos

 <https://doi.org/10.22533/at.ed.8252228105>

CAPÍTULO 6..... 41

ALGORÍTMOS DE MODELAGEM *THREADING* E *AB INITIO*

Anderson Henrique Lima e Lima

Kauê Santana da Costa

Alberto Monteiro dos Santos

 <https://doi.org/10.22533/at.ed.8252228106>

CAPÍTULO 7..... 47

MODELAGEM DE PROTEÍNAS NO *I-TASSER*

Kauê Santana da Costa

Anderson Henrique Lima e Lima

 <https://doi.org/10.22533/at.ed.8252228107>

CAPÍTULO 8..... 53

APLICAÇÕES DA MODELAGEM POR HOMOLOGIA, *AB INITIO* E *THREADING*

Anderson Henrique Lima e Lima

Alberto Monteiro dos Santos

Kauê Santana da Costa

 <https://doi.org/10.22533/at.ed.8252228108>

CAPÍTULO 9..... 60

ATUAIS LIMITAÇÕES E PERSPECTIVAS DOS MÉTODOS *THREADING* E *AB INITIO*

Kauê Santana da Costa

Anderson Henrique Lima e Lima

Alberto Monteiro dos Santos

 <https://doi.org/10.22533/at.ed.8252228109>

GLOSSÁRIO DE TERMOS TÉCNICOS E SIGLAS 64

REFERÊNCIAS 70

SOBRE OS AUTORES 79

Kauê Santana da Costa

João Marcos Pereira Galúcio

O Modeller é um programa desenvolvido em Fortran 90 e *Python* pelo grupo do pesquisador Andrej Šali da Universidade da Califórnia, nos Estados Unidos, e realiza a predição da estrutura de proteínas por homologia. Neste capítulo, mostraremos o passo a passo necessário para a modelagem por homologia.

Programas e bases de dados usados

Modeller

Editor de Texto

UCSF Chimera

RSCB *Protein Data Bank*

As informações deste capítulo podem ser adaptadas para os sistemas Windows e Linux, e é recomendável executar cada uma das etapas em um diretório diferente que contenha os arquivos de input para cada script executado. Para fins de organização, separamos as etapas em diretórios nomeados com os seguintes nomes: *encontrando*, *comparando*, *alinhando*, *construindo* e *otimizando*.

1 | PROCURANDO MOLDES

ESTRUTURAS

Para buscar estruturas homologas as que desejamos modelar, utilizaremos o alinhamento realizado pelo BLASTp. Acesse ao servidor pelo endereço: <https://blast.ncbi.nlm.nih.gov/>.

Escolha a opção *Protein* BLAST (BLASTp). Na página que aparece (Figura 5), (1) adicione a sequência de aminoácidos na caixa do quadro *Enter Query Sequence*. São aceitos os códigos de acesso no GenBank e formato FASTA da sequência. Em seguida; (2) selecione o base de dados em que será realizada a pesquisa: no quadro *Choose Search Set* na opção *Database* selecione *Protein Data Bank* e (3) selecione o nome científico da espécie de interesse na opção *Organism*. Mantenha as outras opções inalteradas e clique em BLAST ao final da página.

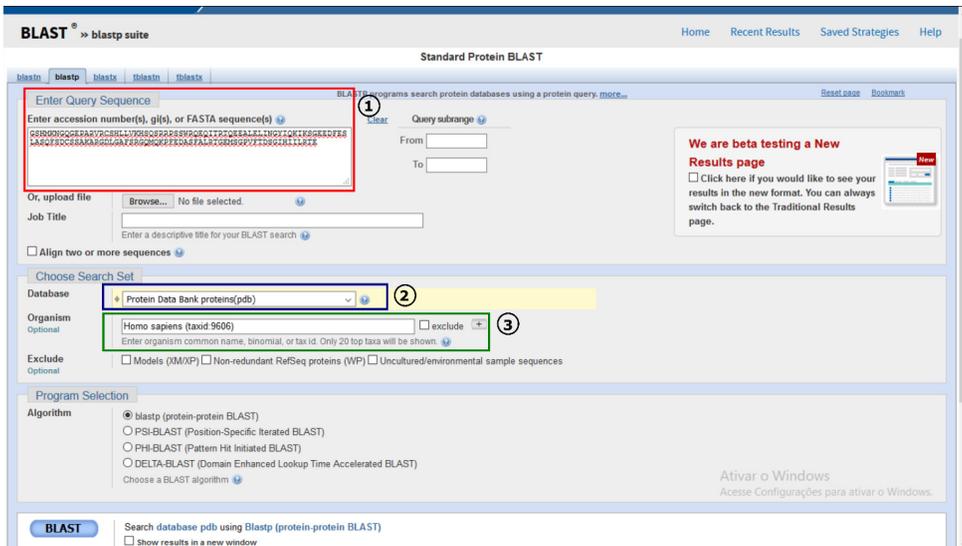


Figura 5. Configurando o BLASTp para pesquisa de seqüências de estruturas homólogas no *Protein Data Bank*.

O resultado da pesquisa do BLASTp (Figura 6) exibe os domínios identificados da proteína (A); a pontuação do alinhamento em uma escala de cores que varia de preto ao vermelho (B), e as descrições de parâmetros obtidos do alinhamento (C). O quadro *Description* contém quatro informações relevantes: cobertura da seqüência em relação ao alvo (C1), *E-value* (C2), identidade (C3) e o código de acesso da seqüência no base de dados pesquisado (C4).

Your search is limited to records that include: Homo sapiens (taxid:9606)

[Edit and Resubmit](#) | [Save Search Strategies](#) | [Formatting options...](#) | [Download](#) | [YouTube](#) [How to read this page](#) | [Blast report description](#) | **NEW** [Click here to see the new BLAST results page](#)

Job title: Protein Sequence

RID: [EK6ZDRAN014](#) (Expires on 05-27 03:47 am)

Query ID: [lcl|Query_349781](#) | Database Name: [pdb](#)
 Description: [None](#) | Description: [PDB protein database](#)
 Molecule type: [amino acid](#) | Program: [BLASTP 2.9.0+](#) | [Citation](#)
 Query Length: [303](#)

Other reports: [Search Summary](#) | [Taxonomy reports](#) | [Distance tree of results](#) | [Multiple alignment](#) | [MSA viewer](#)

Analyze your query with [SmartBLAST](#)

Graphic Summary

Show Conserved Domains

A Putative conserved domains have been detected, click on the image below for detailed results.

B Distribution of the top 100 Blast Hits on 100 subject sequences. Mouse over to see the title, click to show alignments.

C Sequences producing significant alignments: Select [All](#) [None](#) Selected 0

Alignments [Download](#) | [GenPept](#) | [Graphics](#) | [Distance tree of results](#) | [Multiple alignment](#)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/> Chain A, SERINE/THREONINE-PROTEIN KINASE PAK 4 (Homo sapiens)	622	622	100%	0.0	99.67%	2C0D_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 4 (Homo sapiens)	618	618	99%	0.0	99.67%	2QDN_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 4 (Homo sapiens)	617	617	99%	0.0	99.67%	4E1E_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 4 (Homo sapiens)	617	617	99%	0.0	99.67%	5UPL_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 4 (Homo sapiens)	616	616	99%	0.0	99.67%	5UVD_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 4 (Homo sapiens)	616	616	99%	0.0	99.67%	4E1F_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 4 (Homo sapiens)	615	615	99%	0.0	99.67%	4SRU_A
<input type="checkbox"/> Chain A, Protein Fam212a serine/threonine-protein Kinase Pak 4 (Homo sapiens)	615	615	99%	0.0	99.67%	4SRB_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 4 (Homo sapiens)	611	611	99%	0.0	99.00%	4J0J_A
<input type="checkbox"/> Chain A, P21-ACTIVATED KINASE 4 (Homo sapiens)	600	600	96%	0.0	99.66%	2EVA_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 4 (Homo sapiens)	600	600	96%	0.0	99.66%	5UVA_A
<input type="checkbox"/> Chain A, SERINE/THREONINE-PROTEIN KINASE PAK 4 (Homo sapiens)	599	599	96%	0.0	99.32%	2K4Z_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 4 (Homo sapiens)	599	599	96%	0.0	99.66%	4D0V_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 4 (Homo sapiens)	593	593	95%	0.0	99.65%	4DVE_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 7 (Homo sapiens)	519	519	96%	0.0	84.93%	4E87_A
<input type="checkbox"/> Chain A, SERINE/THREONINE-PROTEIN KINASE PAK 4 (Homo sapiens)	461	461	96%	3e-165	74.74%	2C36_A
<input type="checkbox"/> Chain A, Serine/threonine-protein kinase PAK 8 (Homo sapiens)	459	459	95%	3e-165	75.69%	4K87_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 1 (Homo sapiens)	326	326	94%	2e-112	53.50%	4JLI_A
<input type="checkbox"/> Chain A, Serine/threonine-protein Kinase Pak 1 (Homo sapiens)	326	326	94%	2e-112	53.50%	4JLO_A

Ativa o Windows

Figura 6. Parâmetros de alinhamento obtidos pelo BLASTp para a seleção da sequência da melhor estrutura homóloga.

Nesta seção, as etapas de criação da biblioteca, a comparação e a seleção da estrutura-molde no Modeller são opcionais, pois os critérios analisados podem ser suficientes para a escolha da estrutura molde. No entanto, aplicaremos os scripts visando a comparação das sequências.

1.1 Criando Biblioteca de Sequências em Formato PIR

Coloque todas as sequências em um único arquivo no formato PIR e, em seguida, remova todos os espaços entre os aminoácidos. No Linux, você poderá utilizar os editores Vim ou Gedit. No Windows, é aconselhável usar o NotePad++.

O formato das sequências deve seguir o mesmo do base original disponibilizado pelo Modeller, sendo que cada linha da sequência polipeptídica deve conter exatos 75 resíduos. O formato de sequências em PIR que são aceitos pelo programa Modeller seguem o seguinte padrão:

Cabeçalho

```
>P1;1cf9A
structureX:1cf9:1:A:753:A:MOL_ID1;MOLECULELYSOZYME;CHAINA;SYNONYML?
```

- Em “structureX:**1cf9**”, o número em negrito corresponde ao código de identificação da sequência no PDB.
- Em “1:A:**753**”, o número em negrito corresponde à quantidade de aminoácidos presente na sequência.

Sequência de resíduos

Cada linha deve conter 75 resíduos e ao final devem terminar com um asterisco (*).

```
MSEAAHVLIITGAAGQIGYILSHWIASGELYGDRQVYLHLLDIPFAMNRLTALTMELEDCAFPHLAGFVATDPK.
AFKIDIDCAFLVASMPLKPGQVRADLISSNSVIFKNTGEYLSKWAKE SVKVLVIGNEPDNTNCEIAMLHAKNLKPE:
FSSL SMLDQNRAYYE VASKLGVDVKDVHDIIVMGNHGESMVADLTQATFTREGKTKQKVVVDVLDHDYVFDTFKK:
GHRAWDILEHRGFTSAASPTKAAIQHMKAWLFGTAPGEVLSMGIPVPEGNEFYGIKPGVVF SFCNVNDRKGGKIHV
EGFRKVNWLREKLDFTKDLFHEKEIALNHLAQGG*
```

1.2 Selecionando as Estruturas Moldes Pesquisadas

Pasta 1: Encontrando

Arquivos de entrada:

TvLDH.ali (sequência da estrutura de interesse, *target*)

pdb_95PIR (biblioteca de sequências)

build_profile.py (script)

Arquivos de saída:

build_profile.ali

build_profile.log

Modificando o script “build_profile.py”

O script *build_profile.py* deve ser editado nos campos mostrados (Figura 7) para realizar a leitura da sequência alvo em formato .ALI (em azul) e da biblioteca em formato PIR (em vermelho). Note que dois arquivos nos retângulos vermelhos aparecem no formato .BIN, no entanto, será necessário modificar somente o nome do arquivo.

```

#-- Prepare the input files

#-- Read in the sequence database
sdb = sequence_db(env)
sdb.read(seq_database_file='pdb_95.pir', seq_database_format='PIR',
         chains_list='ALL', mirmax_db_seq_len=(30, 4000), clean_sequences=True)

#-- Write the sequence database in binary form
sdb.write(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')

#-- Now, read in the binary database
sdb.read(seq_database_file='pdb_95.bin', seq_database_format='BINARY',
         chains_list='ALL')

#-- Read in the target sequence/alignment
aln = alignment(env)
aln.append(file='TvLDH.ali', alignment_format='PIR', align_codes='ALL')

#-- Convert the input sequence/alignment into
# profile format
prf = aln.to_profile()

#-- Scan sequence database to pick up homologous sequences
prf.build(sdb, matrix_offset=-450, rr_file='${LIB}/blosum62.sim.mat',

```

Figura 7. Campos de edição do script *build_profile.py*.

Executando o Modeller (versão 9.21) no Terminal

Execute os comandos abaixo no terminal do sistema. Entre no diretório faça:

\$ cd diretorio_arquivo/execução

No *prompt* do Windows para listar os arquivos de dentro do diretório, faça:

\$ dir

No terminal do Linux, digite:

\$ ls

Execute o Modeller:

\$ mod9.21 build_profile.py

Observações:

Para obter informações de erro de execução, verifique o arquivo de .log gerado em cada etapa.

Este tutorial foi criado para a versão 9.21, no entanto, caso sua versão seja superior ou inferior, como 9.14, utilize o comando mod9.14, etc.

Na pasta em que o arquivo .py foi executado, abra o arquivo *build_profile.log* no editor de texto. Em *HITS FOUND ITERATION*, encontramos informações relacionadas ao alinhamento. São desejáveis as sequências que apresentam maior identidade e menor *E-value* (Figura 8).

1	2	3	4	5	6	7	8	9	10	11			
1cf9A	1	1	37700	753	502	38.27	0.0	2	467	13	488	75	560
1CF9B	1	2	37700	753	502	38.27	0.0	3	467	13	488	75	560
1E93A	1	3	46100	484	502	42.77	0.0	4	467	14	493	2	478
1E93A	1	4	46100	484	502	42.77	0.0	5	467	14	493	2	478
1gg9A	1	5	37600	753	502	38.27	0.0	6	467	13	488	75	560
1GG9B	1	6	37600	753	502	38.27	0.0	7	467	13	488	75	560
1a4eA	1	158	66350	488	502	53.81	0.0	8	474	14	491	4	488
4b1cA	1	1284	53650	499	502	49.15	0.0	9	465	19	492	25	496
1e93A	1	1663	45450	475	502	43.10	0.0	10	461	19	493	4	474
1p80A	1	1736	37650	727	502	38.27	0.0	11	467	13	488	49	534
1d9fA	1	2225	52500	497	502	48.09	0.0	12	466	19	492	24	495
1gweA	1	3988	41150	498	502	40.80	0.0	13	461	19	492	9	481
1q1lA	1	7470	45450	491	502	46.80	0.0	14	428	19	455	9	446
1s18A	1	7917	47900	474	502	44.68	0.0	15	463	19	492	4	473
1sy7A	1	8105	35550	698	502	43.13	0.0	16	349	39	399	27	390
1ye9A	1	9533	21700	223	502	44.09	0.0	17	215	13	230	1	220
1ye9E	1	9534	16100	256	502	34.52	0.0	18	238	245	488	1	252
1m7sA	1	9766	36100	484	502	40.00	0.0	19	413	15	440	1	420

Figura 8. Colunas do arquivo de saída do Modeller (*build_profile.log*).

As colunas mais importantes correspondem à 1ª, 3ª, 5ª e 7ª e 8ª que correspondem, respectivamente, ao código de identificação das sequências no PDB, ordem de disposição das sequências no base PIR, ao número de resíduos da sequência, valor de identidade em relação à sequência alvo e aos valores de *E-value* (valores ideais iguais ou próximos de 0.0).

1.3 Comparando as Estruturas Homólogas

Esta etapa consiste em comparar os modelos (em formato .PDB) selecionados no passo anterior de acordo com os valores de identidade. Acesse o *Protein Data Bank* (<http://www.rcsb.org>) e faça o *download* de quatro modelos selecionados para serem comparados de acordo com os parâmetros analisados no arquivo .log e adicione à pasta juntamente com os scripts.

Modificando o script "compare.py"

No script *compare.py*, o nome do arquivo (azul) e a cadeia (vermelho) de cada arquivo .PDB devem ser especificados. Certifique-se que escreveu corretamente o nome dos arquivos. É importante atentar que o programa, utiliza sintaxe rígida e, portanto, irá diferenciar letras maiúsculas e minúsculas.

```
env = environ()
aln = alignment(env)
for (pdb, chain) in (('1d9f', 'A'), ('1q9w', 'A'), ('2xq1', 'A'),
                   ('4BLC', 'A'), ('1A4E', 'A')):
    m = model(env, file=pdb, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(m, atom_files=pdb, align_codes=pdb+chain)
```

Pasta 2: Comparando
Arquivos de entrada:
 melhores moldes em formato PDB
compare.py (Script)
Arquivo de saída:
compare.log
family.mat

Arquivo de saída compare.log

No arquivo *compare.log* em *Weighted pair-group average clustering based on a distance matrix*, verificamos a identidade da sequência alvo em relação às sequências das estruturas que serão usadas como moldes (verde), assim como a resolução (azul) da estrutura obtida por raios-X (Figura 9). Os modelos ideais devem apresentar alta identidade e baixa resolução.

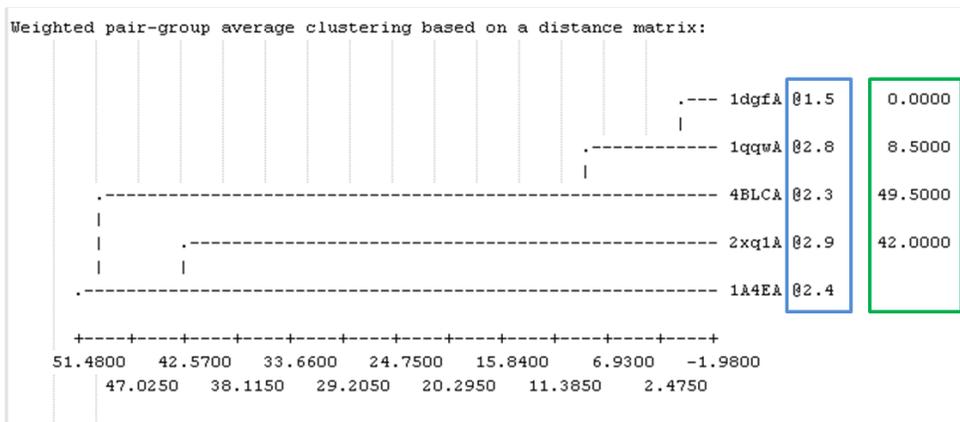


Figura 9. Arquivo de saída *compare.log* do Modeller.

2 | ALINHAMENTO DAS SEQUÊNCIAS

Pasta 3: Alinhando
Arquivos de entrada:
TvLDH.ali (sequência alvo)
1A4E.PDB (sequência molde)
align2d.py (script)
Arquivos de saída:
TvLDH-1bdmA.ali
TvLDH-1bdmA.pap

Modificando o script align2d.py

1. Indique o nome do arquivo em formato PDB (retângulo vermelho) e a cadeia (círculos) em que será realizado o alinhamento.

```
mdl = model(env, file='lbdm', model_segment=('FIRST', 'LAST'))
```

2. Modifique nome do arquivo da sequência alvo TvLDH.ali (azul), assim como o código especificado dentro do mesmo (verde).

```
aln.append(file='TvLDH.ali', align_codes='TvLDH')
```

3. Modifique o nome do arquivo .PDB que será usado como molde (verde) e o nome que será colocado no arquivo gerado (azul).

```
aln.append_model(mdl, align_codes='lbdmA', atom_files='lbdm.pdb')
```

Os arquivos de saída do programa podem ou não serem alterados (fica a critério do usuário).

```
aln.write(file='TvLDH-lbdmA.ali', alignment_format='PIR')
aln.write(file='TvLDH-lbdmA.pap', alignment_format='PAP')
```

Formato do arquivo ALI da sequência alvo

```
>P1;Q8X220
```

```
sequence:Q8X220:::0.00: 0.00
```

```
MALLTAETFRLLQFNKRRRLRRPYPRKALLCYQLTPQNGSTPTRGYFENKKKCHAEICFINEIKSMGL
DETQCYQVTCYLTWSPCSSCAWELVDFIKAHDHLNLRIFASRLYHWCKPQQDGLRLLCGSQVPVE
VMGFPEFADCWENFVDHEKPLSFNPNYKMLEELDKNRSRAIKRRLDRIKIPGVRAQGRYMDILCDAEV*
```

Um alinhamento desejável deve exibir poucos *gaps* (lacunas), e muitos *matches* (correspondência) entre os aminoácidos comparados das duas sequências. A Figura 10 exibe um alinhamento ideal entre a sequência molde (referência) e alvo (de interesse).

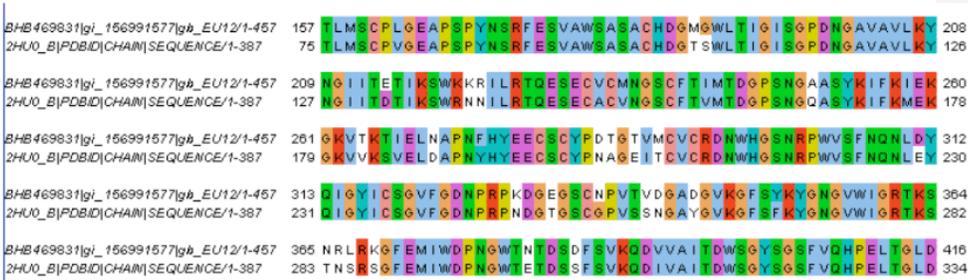


Figura 10. Alinhamento desejável entre a sequência molde e alvo (de interesse). Poucos ou nenhum *gap* e muitos *matches*. Aminoácidos similares, como a isoleucina (I) e a valina (V), assim como os idênticos estão destacados com a mesma cor no alinhamento. Neste exemplo, usou-se o programa MUSCLE.

3 | CONSTRUINDO O MODELO

Pasta 4: Construindo

Arquivos de entrada:

TvLDH-1a4e.ali (alinhamento)

TvLDH.ali (sequência alvo)

Melhor modelo escolhido em formato .PDB

Model-single.py (script)

Arquivos de saída:

São gerados vários arquivos em formato .PDB

Modificando o script *model-single.py*

Modifique o nome do arquivo de alinhamento (.ALI)

```
a = automodel(env, alnfile='TvLDH-1bdmA.ali',
```

Indique o nome da sequência “alvo” (vermelho) e o nome da sequência “molde” (azul). O nome de ambas está no arquivo de alinhamento gerado no passo anterior (.ALI).

```
knowns='1bdmA', sequence='TvLDH',
```

Indique quantos modelos serão gerados na execução do programa (círculo vermelho).

```
a.starting_model = 1  
a.ending_model = 5
```

Observação: o nome deve ser o mesmo encontrado no alinhamento.

4 | OTIMIZANDO O MODELO

A otimização das alças no Modeller é realizada pelo script *loop_refine.py*. Para saber qual região deve ser otimizada submeta seu melhor modelo no servidor *Swiss-Model* (<http://swissmodel.expasy.org/>) e veja no gráfico do QMEAN as regiões de alta energia.

Através do Chimera podemos identificar as regiões de alça da sequência para discriminá-las no *script* de otimização *loop_refine.py*. É importante ressaltar que o sítio catalítico de enzimas, de modo geral, é conservado e, deste modo, deve-se atentar ao realizar a alteração das conformações para não alterar a posição dos resíduos nestas regiões.

Pasta 5: Otimização das Alças

Arquivos de entrada:

loop_refine.py (script)

Modelo selecionado em formato .PDB

Arquivos de saída:

Arquivos no formato .PDB (modelos)

loop_refine.log

Editando o *script loop_refine.py*

1. Adicione o intervalo dos resíduos que serão otimizados (vermelho):

```
def select_loop_atoms(self):  
    # 10 residue insertion  
    return selection(self.residue_range('273', '283'))
```

2. Adicione o nome do melhor modelo gerado (ou pelo script “*multi-model.py*” ou pelo “*AutoModelMulty.py*”) (azul) e o nome da sequência alvo que será gerado (verde):

```
m = MyLoop(env,  
           inimodel='TvLDH-mult.pdb', # initial model of the target  
           sequence='TvLDH')        # code of the target
```

Avaliando os Modelos

Os modelos gerados serão avaliados de acordo com a pontuação Dope (do inglês *Discrete Optimized Protein Energy*), um potencial estatístico usado para avaliar modelos gerados por modelagem comparativa no Modeller. Quanto menor a pontuação em valores negativos, melhor o modelo.

A análise da energia de cada um dos modelos da proteína alvo é mostrado pelo script *model_energies.py*. Através deste script, selecionaremos o modelo com menor energia através dos valores obtidos para o Dope.

Editando o arquivo *model_energies.py*

Adicione o intervalo dos modelos gerados (em azul) durante o passo 4 (construindo modelos). O símbolo de porcentagem “%” significa que serão lidos todos os arquivos de *TvLDH.BL%04d0001.PDB*.

```
log.verbose() # request verbose output  
env = environ()  
env.libs.topology.read(file='${LIB}/top_heav.lib') # read topology  
env.libs.parameters.read(file='${LIB}/par.lib') # read parameters  
  
for i in range(1, 11):  
    # read model file  
    code = 'TvLDH.BL%04d0001.pdb' % i  
    mdl = complete_pdb(env, code)  
    s = selection(mdl)  
    s.assess_dope(output='ENERGY_PROFILE_NO_REPORT', file='TvLDH.profile',  
                 normalize_profile=True, smoothing_window=15)
```

Deve-se escolher um dos modelos gerados que exibe a menor energia Dope (azul).

Abra o arquivo *loop_refine.log*.

```
<< end of ENERGY.  
DOPE score : -50066.785156  
openf__224_> Open TvLDH.BL00080001.pdb
```

ATENÇÃO: Não pertence ao modelo 08, e sim ao 07

Analizando as Regiões de Alça pelo Chimera

Abra o arquivo .PDB da estrutura usando o programa UCSF Chimera (Figura 11). O modelo será exibido em diagrama de Ribbon (1) e para colori-lo, vá na barra de menu e faça:

Action > *Color*

Pra visualizar a sequência de aminoácidos que formam β - folhas e α -hélices, faça:

Tools > *Sequence* > *Sequence*.

As estruturas secundárias são mostradas nos quadros coloridos e alças não são destacadas (2). Para verificar a numeração dos resíduos da sequência, na caixa que abrir, faça:

Numberings > *Overall alignment*

1	11	21	31
modelo_homolog...db (#0) chain A 289	GSSPQREPRV	SHEGFRAALQLVVD	FGDPRS
51	61	71	81
modelo_homolog...db (#0) chain A 339	IVRS	SGKLVAVK	KMDLRKQQRRELLFNEVVI
101	111	121	13
modelo_homolog...db (#0) chain A 439	RD	IKSDSITLLTHDGRV	KLSDFGFC
151	161	171	18
modelo_homolog...db (#0) chain A 489	RLPYGPEVD	IWSLGI	IMVLEMVDGEP
201	211	221	23
251	261	271	28
modelo_homolog...db (#0) chain A 539	HKVSPSLKGF	LDRLLV	VRDPAQRATAAELLKH

Figura 11. Visualização do modelo tridimensional pelo UCSF Chimera. As regiões em verde exibem as β -folhas em amarelo as α -hélices. Passando o *mouse* sobre os resíduos, vemos a posição de cada um no lado inferior direito da tela.

5 | VALIDANDO O MODELO

Usaremos o servidor suíço *Swiss-model* mantido pela Universidade de Bazel para a validação do modelos gerado nas etapas anteriores. Acesse o servidor pelo endereço abaixo:

<https://swissmodel.expasy.org/assess>

Na página, faça o upload do arquivo .PDB gerado nas etapas anteriores e nomeie a atividade no campo *Structure File*, (1) clicando em *Upload Coordinate File* (Figura 12). Em seguida, (2) clique em *Start Assessment* para executar a análise da estrutura.

Structure Assessment

Start a new Structure Assessment Project

Structure File: 1 + Upload Coordinate File... modelo_homologia.pdb ✓

Project Title (Optional): Modelagem por homologia

Email (Optional): Optional

2 Start Assessment Reset

Ativar o Windows
Acesse Configurações para ativar o Windows.

Swiss Institute of Bioinformatics Schwede Group Contact Us Terms of use [Back to the Top](#)

Figura 12. Página inicial do Swiss-model exibe as regiões para *upload* das estruturas que serão analisadas.

Após a modelagem, o gráfico de Ramachandran da estrutura tridimensional é exibida do lado esquerdo e a estrutura em diagrama de Ribbon do lado direito. Clicando nos pontos referentes a cada resíduo, o painel contendo a estrutura é atualizado para exibição de detalhes estruturais.

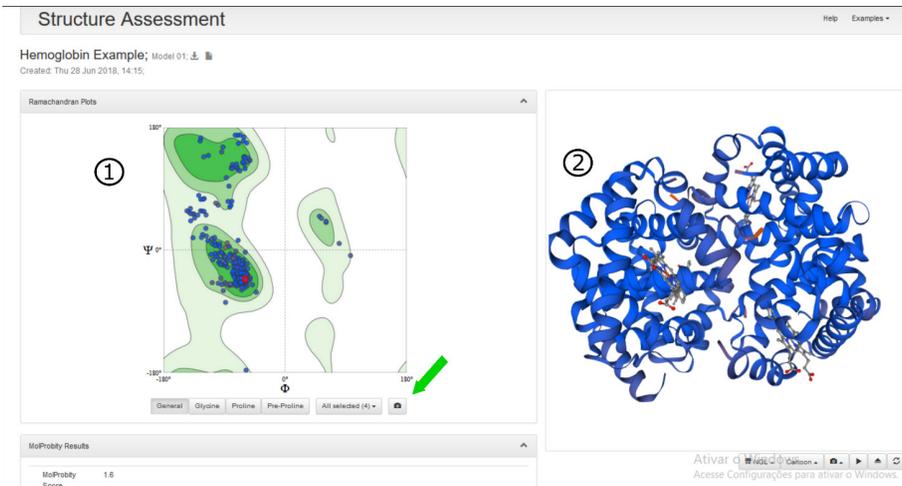


Figura 13. Página de resultados do servidor Swiss-model exibindo em (1) o gráfico de Ramachandran gerado e em (2) a estrutura da proteína representada pelo diagrama de *Ribbon*. A seta verde exibe a região de *download* do gráfico em formato de imagem (.png).

Na mesma página são exibidos dados estereoquímicos e de energia da estrutura (Figura 14).

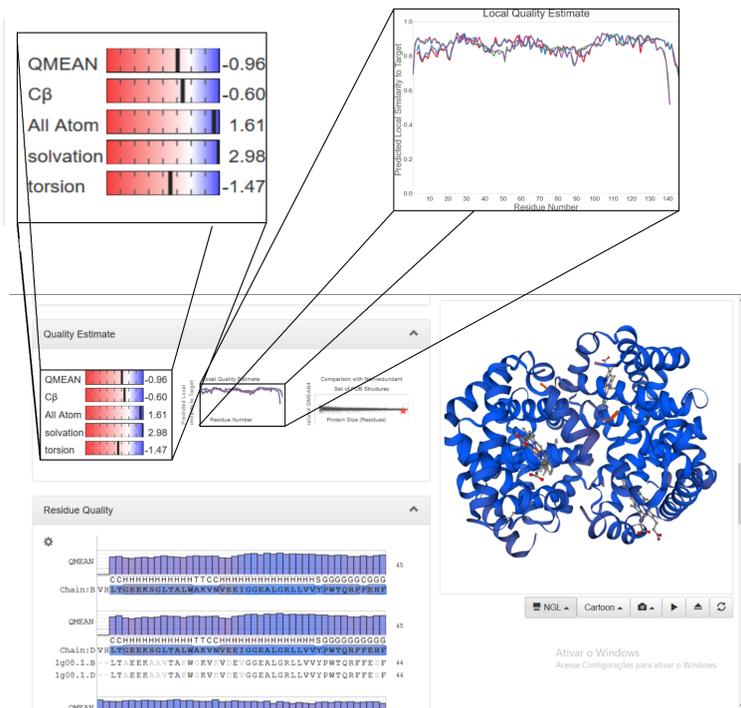


Figura 14. Avaliação da energia e estereoquímica do modelo. No caso acima, analisamos a estrutura da hemoglobina (formada por quatro cadeias). Quatro plotagens são mostradas no gráfico, sendo uma para cada cadeia.

A predição da estrutura de proteínas é, sem dúvida, um assunto chave para todos que atuam nas áreas de modelagem molecular, bioinformática e biologia estrutural, dada a versatilidade de funções desempenhada por estes biopolímeros em processos biológicos, assim como suas diferentes aplicações farmacêuticas, industriais e biotecnológicas. Com este livro, pretende-se não somente atender às necessidades práticas relacionadas à modelagem e análise das estruturas tridimensionais de proteínas, conteúdo ainda carente em língua portuguesa, mas também fornecer os principais conceitos necessários para a compreensão dos métodos, suas aplicações, assim como limitações. O livro traz um conteúdo e linguagem adaptada para alunos de graduação e, devido ao grande número de expressões originárias do inglês presentes na literatura científica, são incluídos os termos mais comumente utilizados na língua de origem, sempre dando preferência à tradução destes para o português.

O livro traz linguagem acessível, inclui referências atualizadas sobre os métodos, se encontra ricamente ilustrado e traz ao final um glossário de termos técnicos que enriquecerão a leitura e compreensão do leitor. Esta obra é uma importante fonte de consulta de alunos de graduação e pós-graduação de cursos relacionados às áreas de ciências Biológicas, Médicas, Exatas e Naturais, assim como, para todos interessados no assunto.



A predição da estrutura de proteínas é, sem dúvida, um assunto chave para todos que atuam nas áreas de modelagem molecular, bioinformática e biologia estrutural, dada a versatilidade de funções desempenhada por estes biopolímeros em processos biológicos, assim como suas diferentes aplicações farmacêuticas, industriais e biotecnológicas. Com este livro, pretende-se não somente atender às necessidades práticas relacionadas à modelagem e análise das estruturas tridimensionais de proteínas, conteúdo ainda carente em língua portuguesa, mas também fornecer os principais conceitos necessários para a compreensão dos métodos, suas aplicações, assim como limitações. O livro traz um conteúdo e linguagem adaptada para alunos de graduação e, devido ao grande número de expressões originárias do inglês presentes na literatura científica, são incluídos os termos mais comumente utilizados na língua de origem, sempre dando preferência à tradução destes para o português.

O livro traz linguagem acessível, inclui referências atualizadas sobre os métodos, se encontra ricamente ilustrado e traz ao final um glossário de termos técnicos que enriquecerão a leitura e compreensão do leitor. Esta obra é uma importante fonte de consulta de alunos de graduação e pós-graduação de cursos relacionados às áreas de ciências Biológicas, Médicas, Exatas e Naturais, assim como, para todos interessados no assunto.

