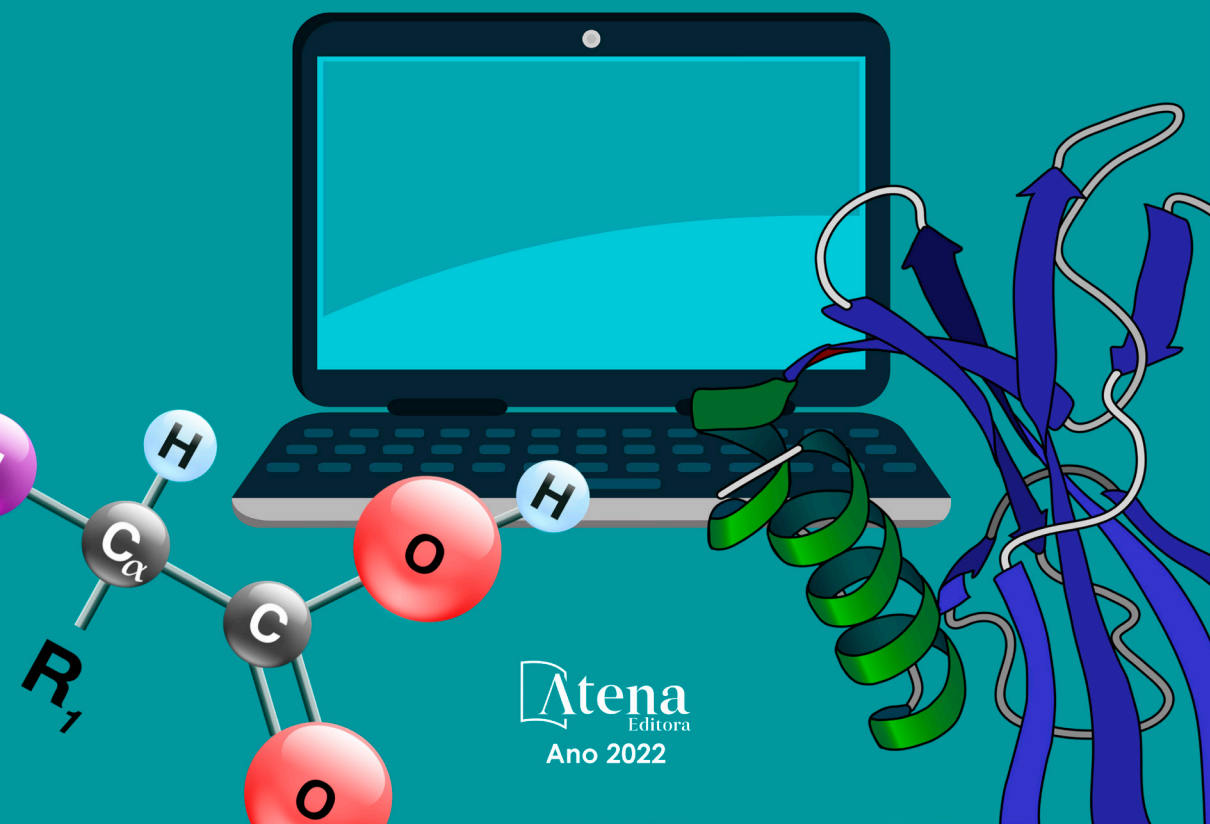


MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

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 discorrem sem 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

Anderson Henrique Lima e Lima

ACESSANDO A SEQUÊNCIA PARA A MODELAGEM

Acesse o base de sequência de proteínas UniProt através do endereço: <https://www.uniprot.org/>

Pesquise no base do UniProt, uma sequência de aminoácidos de seu interesse. No campo “*Sequence*”, copie a sequência dos resíduos de aminoácidos e salve no seu computador em formato Fasta (Figura 24).

Neste capítulo, mostraremos a predição da estrutura pela abordagem do servidor I-TASSER (do inglês *Iterative Threading ASSEmbly Refinement*) e analisaremos os resultados de predição. O servidor, assim como a versão instalável do programa, foi desenvolvido pelo grupo de Yang Zhang, da Universidade de Michigan (EUA), e ambos empregam um algoritmo de modelagem avançada que utiliza um misto de predição por *threading* com predição *ab initio* de regiões sem cobertura.

Entry **Sequence**

Publications

Feature viewer

Feature table

None

Sequence status: Complete.

A0A3Q8I9F4-1 [UniParc] FASTA Add to basket

< Hide

Length: 121
Mass (Da): 13,360
Last modified: April 10, 2019 - v1
Checksum: 703641A6EE9182BA

BLAST GO

```

10      20      30      40      50
MSAENGNAI VRVSSNKRKF GYDVTKRL HEGYPEVVIS ALGTAIADAV
60      70      80      90     100
SVVELLNQG VVEVKRICTS RAQFDDVRST TDKIEVVVV KSPDFDAIYD
110     120
QQQKREIRAK ARAEADEADD E
    
```

Figura 24. Acesso às sequências de aminoácidos no base UniProt.

PESQUISANDO POR ESTRUTURAS HOMÓLOGAS

Utilizando o *BLASTp* realize uma busca de estruturas homólogas no *Protein Data Bank* que apresentem similaridade $\geq 30\%$ que possam ser utilizadas como molde para a modelagem da estrutura. Após realizar a pesquisa, caso o servidor retorne que não foram encontradas sequências similares na base do PDB (Figura 25) teremos que utilizar uma nova estratégia de modelagem.

The screenshot displays the BLASTp search results interface. On the left, a table lists search details: Job Title (Protein Sequence), RID (19DBUEYP016), Program (Citation), Database (pdb), Query ID (Ic|Query_48512), Description (None), Molecule type (amino acid), Query Length (121), and Other reports. The 'Filter Results' section on the right includes an 'Organism' filter (set to 'only top 20 will appear'), a search box for organism names, and filters for 'Percent Identity', 'E value', and 'Query Coverage'. A yellow banner at the bottom contains a warning icon and the text: 'No significant similarity found. For reasons why, click here'.

Figura 25. Resultado de pesquisa de similaridade realizado no Protein Data Bank usando a ferramenta BLASTp do GenBank.


A abordagem *threading* é indicada para a predição da estrutura quando a similaridade da sequência alvo se mostra menor que 30% e a *ab initio* quando a identidade/similaridade entre ambas é inexistente (0%). No caso acima, como não foram encontradas similaridades suficientes ($< 30\%$), realizaremos a predição usando o servidor I-TASSER.

REALIZANDO A PREDIÇÃO PELO I-TASSER

Acesse o servidor I-TASSER, utilizando o endereço: <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>

O servidor apresenta interface intuitiva e amigável, exibindo na primeira página os campos para preenchimento com as informações necessárias para a modelagem (figura 26, campos indicados pelas setas). Na primeira caixa, deve-se adicionar a sequência de aminoácidos da proteína em formato Fasta, seguido do e-mail de contato no qual os resultados serão enviados (obrigatório), a senha de registro criada no momento do cadastro do usuário (obrigatório) e um código identificador para reconhecimento da submissão (opcional).

Home Research Services Publications People Teaching Jobs



(The server completed predictions for [514429 proteins](#) submitted to the server. [\(The template library was updated on 2014-07-24\)](#))

I-TASSER (Iterative Threading ASSEMBly Refinement) is a hierarchical approach to protein structure and function prediction. It generates full-length atomic models constructed by iterative template-based fragment assembly simulations. Function insights of the 'Zhang-Server' was ranked as the No. 1 server for protein structure prediction in recent community-wide [CASP7](#), [CASP8](#), and [CASP9](#). The server is in active development with the goal to provide the most accurate structural and functional predictions. [\(The template library was updated on 2014-07-24\)](#) and our developers will study and answer the questions accordingly. ([> More about the server ...](#))

[\[Queue\]](#) [\[Forum\]](#) [\[Download\]](#) [\[Search\]](#) [\[Registration\]](#) [\[Statistics\]](#) [\[Help\]](#)

I-TASSER On-line Server ([View an example of I-TASSER output](#)):

Copy and paste your sequence below ([10, 1500] residues in **FASTA format**). [Click here for a sample input.](#)

```
MSAENGNGNAIVRVSSNKRKFGYVDYTKHRLHEGYPEVVVISALGTAIADAVSVVELLNQGG
VVEVKKICTSRAQFDDVVRSTTTDKIEVVVVVKSDFDAIYDQQQKDREIAKARAEADEADD
E
```

Or upload the sequence from your local computer:

No file selected.

Email: (mandatory, where results will be sent to)

Password: (mandatory, please [click here](#) if you do not have a password)

Figura 26. Campos de preenchimento obrigatório para a modelagem utilizando o servidor I-TASSER.

O servidor também dispõe da opção de incluir uma estrutura cristalográfica como referência e definir restrições para a formação de interações intramoleculares da estrutura para realizar a modelagem (Figura 26). Na caixa *Specify template without alignment*, é possível fazer o upload do arquivo PDB da estrutura que será usada como referência e na caixa *Assign contact/distance restraints* é possível adicionar as restrições referentes às interações entre os resíduos. Segue abaixo um exemplo

```
DIST 12 HG21 50 HB1 8.1
DIST 14 HA 57 1HE 6.2
DIST 21 HB2 43 HD11 4.0
DIST 124 CA 84 CA 17.4
DIST 36 UNK 120 CA 17.4
```

O arquivo também pode especificar o contato entre os resíduos especificando a numeração de cada um:

CONTACT 33 6
 CONTACT 60 29
 CONTACT 37 345
 CONTACT 109 42

▼ **Option I: Assign additional restraints & templates to guide I-TASSER modeling.**
 (Read more explanation on how to add restraints)

- o Assign contact/distance restraints No file selected. Atom-atom restraints [Explanation](#)
- o Specify template without alignment Type a PDB ID [Explanation](#)
- o Specify template without alignment No file selected. Upload a structure [Explanation](#)
- o Specify template with alignment No file selected. Template & alignment [Explanation](#)

Figura 26. Opções de modelagem usando o servidor I-TASSER

ANÁLISE DOS RESULTADOS DE PREDIÇÃO

Os resultados de modelagem do I-TASSER são exibidos em diferentes caixas na página do servidor (Figura 27). Na primeira caixa é exibida a sequência de aminoácidos em formato Fasta da proteína submetida (*Submitted Sequence in FASTA format*), na segunda a predição da estrutura secundária da sequência (*Predicted Secondary Structure*) e na terceira as regiões dos resíduos acessíveis ao solvente (*Predicted Solvent Accessibility*).

Nesta terceira caixa, cada resíduo é indicado abaixo com uma numeração que varia de 0 a 9 que indica o grau de exposição ao solvente, sendo 0 os resíduos menos expostos e 9 os resíduos mais expostos.

Submitted Sequence in FASTA format

```

>Protein
HQKVLVNSIKDRNTIQDNESTLEVTLSKYSTSPYLLEAAHLSFENPVEDEAKICVQELQ
CFKIIITKNSPLPSSII SRREKNDVFLLEAGRAEKIVISRSTPTFNKQTKRVSNWSPNS
LQVFTGKIPKATPELGSSENSASPPRFKTEKMEKVLPTFDKCESNLTVNTS
    
```

Predicted Secondary Structure

	20	40	60	80	100	120	140
Sequence	HQKVLVNSIKDRNTIQDNESTLEVTLSKYSTSPYLLEAAHLSFENPVEDEAKICVQELQCFKIIITKNSPLPSSII SRREKNDVFLLEAGRAEKIVISRSTPTFNKQTKRVSNWSPNSLQVFTGKIPKATPELGSSENSASPPRFKTE						
Prediction	CCCCCCCCCCCCCCCCCCSSSSSSCCCCCHHHHHHHHCCCCCHHHHHHHHHCCCCCHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHH						
Conf. Score	9520024532234335676625677743786267876652058877421246676765236234338989876433223587078898875404267314788623445534553330334303530236752215433366776310001						

H:Helix; S:Strand; C:Coil

Predicted Solvent Accessibility

	20	40	60	80	100	120	140
Sequence	HQKVLVNSIKDRNTIQDNESTLEVTLSKYSTSPYLLEAAHLSFENPVEDEAKICVQELQCFKIIITKNSPLPSSII SRREKNDVFLLEAGRAEKIVISRSTPTFNKQTKRVSNWSPNSLQVFTGKIPKATPELGSSENSASPPRFKTE						
Prediction	84433633476444366643313022244433321243344334733465624121442523413364452355324445663232352164543313334444145435624142243234203440353344414347434444441444						

Values range from 0 (buried residue) to 9 (highly exposed residue) Acesse Configurações para ativar o Windows.

Figura 27. Resultados de predição do I-TASSER mostrando a predição da estrutura secundária e as regiões dos resíduos acessíveis ao solvente.

Na quarta caixa é exibido um gráfico do *B-factor* (*Predicted normalized B*) distribuído ao longo da sequência de aminoácidos juntamente com a predição da estrutura secundária

(Figura 28). Os valores de *B-factor* são um importante indicativo da mobilidade e flexibilidade de regiões da estrutura (resíduos e átomos da proteína). No I-TASSER, os valores do gráfico são deduzidos das estruturas das proteínas obtidas do base PDB utilizadas como molde, assim como, dos perfis de sequência derivados dos bases de dados de sequência.

Predicted normalized B-factor

(B-factor is a value to indicate the extent of the inherent thermal mobility of residues/atoms in proteins. In I-TASSER, this value is deduced from threading template proteins from the PDB in combination with the sequence profiles derived from sequence databases. The reported B-factor profile in the figure below corresponds to the normalized B-factor of the target protein, defined by $B=(B-u)/s$, where B is the raw B-factor value, u and s are respectively the mean and standard deviation of the raw B-factors along the sequence. [Click here to read more about predicted normalized B-factor](#))

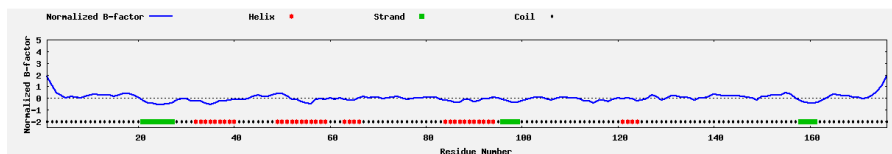


Figura 28. Gráfico do *B-factor* plotado em função da sequência de resíduos de aminoácidos da proteína.

A quinta caixa (Figura 29) exibe o resultado da pesquisa das estruturas experimentais obtidas no PDB utilizadas, como molde para modelagem (*Top 10 threading templates used by I-TASSER*). Na Figura 29, a seta no topo indica a predição da estrutura secundária ao longo da sequência de resíduos, na qual “H” indica as regiões de alfa-hélice, “S” as regiões de beta-folha, e “C” as regiões de alça da estrutura; logo abaixo é exibido o alinhamento *threading* das sequências pesquisadas como molde.

As sete colunas que exibem os resultados indicam da esquerda para a direita: (1) *rank* das melhores estruturas (coluna *Rank*), onde as melhores são exibidas primeiro e são seguidas das demais; (2) os códigos de acesso do PDB das estruturas usadas como moldes, seguido da cadeia (coluna *PDB Hits*); (3) o valor percentual de identidade da sequência dos moldes usados para a região alinhada com a sequência indagada (*query*) (coluna *Iden1*); (4) o valor percentual de identidade da sequência de todas as cadeias das estruturas moldes (*templates*) com a sequência de indagada (*query*); (5) cobertura do alinhamento, que representa o número de resíduos alinhados dividido pelo comprimento da sequência proteína indagada (*query*) (coluna *Cov*); (6) valor de Z-score referente ao alinhamento, sendo que o ideal são valores maiores que 1, pois indicam alinhamento satisfatório; (7) link para *download* das estruturas pesquisadas pelo alinhamento do *threading* utilizado pelo I-TASSER.

Rank	PDB Hit	Ident1	Ident2	Cov	Norm. Z-score	Download	Align.
1	6cmk3	0.13	0.15	0.82	1.11	Download	
2	2pft	0.15	0.40	0.92	1.21	Download	
3	8kz3a	0.04	0.10	0.94	1.27	Download	
4	5mc9b	0.04	0.41	0.97	1.08	Download	
5	1tt4a	0.17	0.23	0.69	1.21	Download	
6	2ma5a	0.10	0.17	0.88	1.12	Download	
7	3ica5	0.10	0.44	0.99	1.08	Download	
8	2pfa2	0.12	0.22	0.97	1.10	Download	
9	3jcsn	0.09	0.20	0.96	1.10	Download	
10	6fthd	0.09	0.33	0.98	1.04	Download	

Sec.Str	Seq
CCCCCCCCCCCCCCCCCCCCCCCCSSSSSSSSCCCCCHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC	HQRVLLVNVIKRRTNIQDNRESTLEVTSLSKYSTPFLLEAAASHLSPEHFVEDEAKICVQLQCPKIIITKNSPLPSSIIIRREKKNDFVLEEAAGRAEIVISRSSTSFTHWQTKRVVSWSSFN
SSSLHYAYRGRYLICVHGRK----ELIKFRG--NTCVLALMGHQLTLLLLQFFINRGRGSPAIRHACNDYVQCLSLIIGVGLVMDVINDTFLHY----CLEYGSIE	AQLVAIFGQRI----TDYF--FEELIKVQYTVIPLDLEKFAETL----SELIITIDAEVYFEMLENYR----TFDDELLSIVYVIALELFTFGKAGHSIAEEDAEFF
-----PAPRABAAASAPVATADAAADATASASASVAAATADADASATATIHASASASVAAAGVAVDPAASASASAAAVASASALELTVVAVAAATANT	LLATLGENVYGVAVVFNLPVAVASQKFLNATVENITFESFVWQVYVYVYIACG--LOPDIILFHGDTQIGERGI--LSGQGRQRISVARALVQGVVFLFDGFSAL----E
RSIVYHNIPIKIF-----DSDDLIPYLDQVYI-----IGHLKHIEHERM-----LFRMVFDFP-----FLKIRHDSILK	-----MDAIKKFMQKLIKFNALDRAEQAFADKKAADRSPQLDELVSLQKIKGTEFEDLRYSEALKDAQKLELAEKAT--DAAAVASLMRRIQVVEELDRAQERLA
HQDSSNKIVNLMIGRINVIIEEPESEDLTHLAQKFPILNLIKIDSYVNVKSYI--NDFISLNGAMTVSVRDLIKLCERLDLIFONGINKDQLISCFAGAIGE	LLGGAKVVTTSRSFSK--QTDYVQSIYAKYG--AKSSTLVVFPNQSQKQVLEALIEFYLDALFPFAAEGOIELDSFAHRMLINKNSARGIETFRMHNHGDGSE--ANQL
-----IGAGVANLYWDDRQSELEHQGVAVKARIALEKIIEGVNINTRKAEVKKLVLDLRGSRG--VRMDLASSAKVWVKTSENNNAVSDGSAYS----HDEVAEMRG	GVQQLTKLTKRPRGETEVGATAVAFSSDFRFLFLOGEGFFLKCSLAPAFQFPSPHGSFRHNRLFSAGTDGRHLSMLQAPPLTSLQLSLFLFAVWRKSI----VRLVFAA--GR

Figura 29. Alinhamento das seqüências utilizadas como molde com a seqüência indagada (*query*); e resultados associados, tais como valores de identidade, cobertura e Z-score.

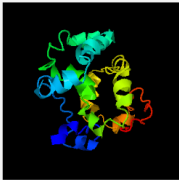
A sexta caixa (*Top 5 final models predicted by I-TASSER*) exibe o resultado da modelagem do servidor, isto é, as estruturas preditas para a seqüência submetida (Figura 30). Nela são exibidas as cinco melhores estruturas com base nos valores de *TM-score* (ver capítulo 6) e *C-score* utilizado para estimar a qualidade dos modelos gerados pelo I-TASSER, sendo calculado com base nos alinhamentos do modelo e nos parâmetros de convergência das simulações de montagem da estrutura. O *C-score* é normalmente encontrado no intervalo que varia entre os valores de -5 a 2, onde quanto maior a pontuação maior será a confiança em relação ao modelo gerado. Esta pontuação tem utilidade semelhante aos valores de similaridade estrutural calculados pelo *TM-score* e RMSD-Ca (ver Capítulo 2 e Capítulo 5) para prever a proximidade da estrutura modelada com o molde gerado. Preferencialmente, deve-se escolher o primeiro modelo, devido aos valores satisfatórios de *C-score* e *TM-score* obtidos na modelagem.

Top 5 final models predicted by I-TASSER

(For each target, I-TASSER simulates a large ensemble of structural conformations, called decoys. To select the final model, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by C-score that is calculated based on the significance of detecting template alignments and the convergence parameters of the structure assembly simulation. C-score is typically in the range of (-5, 2), where a C-score of a higher value signifies a model with a higher confidence and vice-versa. TM-score and RMSD are estimated based on C-score and protein length following the correlation observed between these quantities. Since the top 5 models are ranked by the cluster size, it is possible that the lower-rank models have a higher C-score in some cases. Although the first model has a better quality in most cases, it is also possible that the lower-rank models have a better quality than the higher-rank models in some benchmark tests. If the I-TASSER simulation coverage, it is possible to have less than 5 clusters generated; this is usually an indication that the models have a good quality because of the converged simulations.)

[View About C-score](#)
[Learn structure quality: profile of the top five models](#)

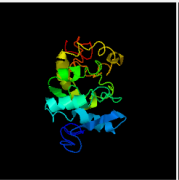
(By right-click on the images, you can export image file or change the configuration, e.g. modifying the background color or setting the spin of your model.)



[Reset to initial orientation](#) Spin On/Off

[Download Model 1](#)

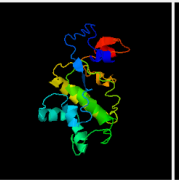
C-score = 4.39 (Read more about C-score)
 Estimated TM-score = 1.21-0.89
 Estimated RMSD = 15.7-3.3 Å



[Reset to initial orientation](#) Spin On/Off

[Download Model 2](#)

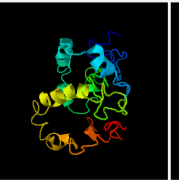
C-score = -4.47



[Reset to initial orientation](#) Spin On/Off

[Download Model 3](#)

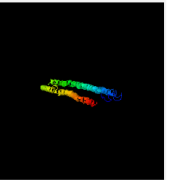
C-score = -4.64



[Reset to initial orientation](#) Spin On/Off

[Download Model 4](#)

C-score = -4.87



[Reset to initial orientation](#) Spin On/Off

[Download Model 5](#)

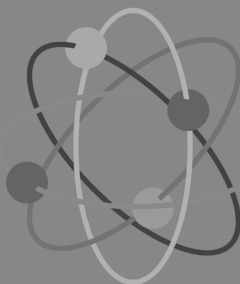
C-score = -3.60

Accesse Configurações para ativar o Windows.

Figura 30. Estruturas modeladas pelo I-TASSER. Por padrão o servidor libera as cinco melhores estruturas, seguida do valor de *C-score* e *TM-score*.

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.

