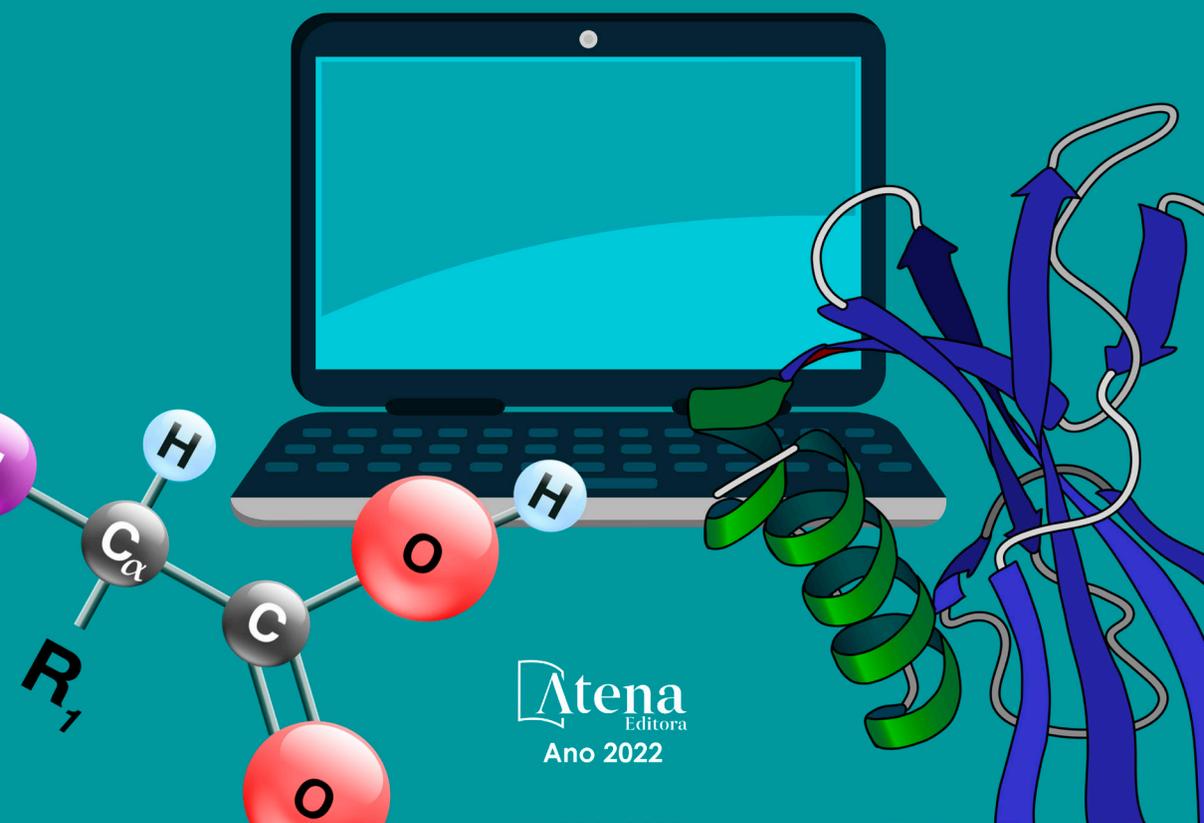


MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

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-NãoDerivativos 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

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

Kauê Santana da Costa

Anderson Henrique Lima e Lima

Alberto Monteiro dos Santos

Apesar de terem evoluído consideravelmente nos últimos anos mostrando resultados satisfatórios nas estruturas obtidas, os métodos de modelagem da estrutura tridimensional de proteínas também apresentam suas limitações, em especial no que diz respeito aos métodos que utilizam a abordagem *threading* e *ab initio*. Este capítulo irá apresentar as principais limitações e discutir perspectivas de desenvolvimento, e algumas novas aplicações que estão sendo abordadas para estes métodos.

ATUAIS LIMITAÇÕES E DESAFIOS

Conciliar um baixo custo computacional com uma simulação que obtenha modelos mais realísticos com baixa resolução e baixa energia mostra-se um grande desafio para a área de predição de estruturas de proteínas. Com relação aos algoritmos *ab initio*, apesar de eles terem evoluído ao longo destes anos – progresso notado durante os eventos do CASP – estes apresentam algumas limitações. Uma delas refere-se ao demorado tempo computacional em relação às abordagens baseadas em

molde, uma vez que a grande parte da sua aplicação, tem sido na obtenção de estruturas que não apresentam qualquer identidade ou não apresentam valores satisfatórios desta com outras sequências de proteínas (BAKER et al., 2006; BAR-ON et al., 2011; KEMEGE et al., 2011).

No entanto, os métodos *ab initio*, que não eram capazes de prever de maneira acurada a topologia de estruturas que continham mais de 100 resíduos (SIMONS; STRAUSS; BAKER, 2001), na última década, nos mostram, com certa precisão, a estrutura de proteínas com 100-120 resíduos de aminoácidos. É importante ressaltar ainda que já foi relatado na literatura, a modelagem bem sucedida de estruturas raras, em que se obtiveram modelos de resolução satisfatória (<2,0 Å de átomos de Ca) (BRADLEY; MISURA; BAKER, 2005). Com relação a isso, o algoritmo aplicado no programa de predição Quark, disponibilizado e mantido pela Universidade de Michigan (EUA) permite sequências com tamanhos maiores de até 200 resíduos de aminoácidos (XU; ZHANG, 2012).

Tomando como base os resultados apresentados nos últimos eventos do CASP, os modelos, algumas vezes, são distantes ou apresentam determinadas deficiências em relação às estruturas nativas, o que requer que inspeções visuais sejam realizadas com o propósito de estabelecer características

importantes no alvo que se deseja obter, pois pelo menos até o momento, as funções de pontuação ou métodos de comparação – como o GDT_TS (do inglês *Global Distance Test Total Score*), largamente utilizados para a seleção de estruturas – não são capazes de reproduzir as decisões humanas na escolha de modelos teóricos próximos à conformação nativa (BEN-DAVID et al., 2009; KRYSHTAFOVYCH; FIDELIS; MOULT, 2014; TAI et al., 2014).

O GDT_TS é um método métrico que determina a acurácia global de predição do modelo em relação à estrutura nativa experimentalmente obtida, pois mede a porcentagem de átomos de Ca espacialmente alinhados entre ambas as estruturas. É considerado mais robusto que o RMSD na comparação de estruturas. Valores de GDT_TS entre 90-100 representam sucesso na predição. Modelos com valores de 89-60 representam proteínas com similaridade estrutural satisfatória. Modelos com valores de 30-20 representam estruturas aleatórias e sem significância obtidas no processo de modelagem. Durante os eventos do CASP, temos percebido progressos na qualidade dos modelos, por exemplo, no CASP 5, somente um dos cinco alvos estudados se mostrou acima de 60, e os outros com pontuação acima de 40. Em contraste, durante o CASP 10, três estruturas alvo com menos de 120 resíduos mostraram GDT maior que 60 (KINCH et al., 2011).

Métodos de predição *ab initio* que utilizam campos de força físicos aliados à simulação de dinâmica molecular, estão longe de se tornar rotineiros na predição da estrutura de proteínas devido ao elevado custo computacional, embora sucessos de predição já tenham ocorrido com esta abordagem, utilizando super-computadores (NGUYEN et al., 2014). Considerando que a proteína de interesse apresente uma longa cadeia polipeptídica e que o dobramento desta na sua conformação nativa, ao longo da dinâmica molecular, pode requerer escalas superiores ao de nanosegundos, uma capacidade e tempo de processamento maiores serão necessários para a simulação. Recentemente, Voelz et al., (2010) simularam o dobramento da região N-terminal da proteína ribossomal L9 (NTL9) por abordagem *ab initio*, utilizando o sistema de computadores Folding@Home e uma versão modificada do pacote computacional GROMACS escrita para GPUs utilizando um tempo total de simulação 1.52 microssegundos (ms) (VOELZ et al., 2010).

Nesta simulação, utilizaram o campo de força AMBER ff96 e o modelo de solvatação GBSA e iniciaram os cálculos de DM em diferentes temperaturas e com fragmento da proteína em diferentes estados conformacionais (incluindo a forma distendida). No referido estudo, apesar de simularem um fragmento de somente 39 resíduos para a proteína que apresenta um tempo de dobramento de ~1,5 ms obtiveram um RMSD-Ca de 3,1 Å com relação à forma nativa obtida por difração de raios-X. Deste modo, percebemos que os métodos *ab initio* que utilizam simulação de dinâmica molecular aliado ao campos de força com potenciais físicos e modelos de solvatação, podem ser úteis na compreensão do enovelamento de proteínas quando comparados aos métodos de predição baseados em molde, como o *threading* e modelagem por homologia, visto que estes últimos, utilizam

somente estruturas de referência como molde, de onde se extraem um arcabouço estrutural para construção do modelo.

Com relação ao sucesso do desenho de enzimas, embora tenhamos notado avanços interessantes nesta área (KHERSONSKY et al., 2011; RÖTHLISBERGER et al., 2008), as enzimas engenheiradas ainda apresentam pouca eficiência na atividade catalítica quando comparadas àquelas do estado nativo. É por isso que o desenho de enzimas tem sido utilizado em conjunto com a evolução dirigida, como forma de corrigir possíveis erros na estrutura catalítica não previstos pelos métodos computacionais. Neste sentido, avanços no desenvolvimento dos campos de força, na seleção da conformação das alças, ou esqueletos proteicos adequados ao sítio catalítico, ainda serão necessários para aperfeiçoar os métodos *in silico* aplicados ao desenho destas estruturas (BAKER, 2010).

PERSPECTIVAS E DESENVOLVIMENTOS FUTUROS

Apesar das limitações apresentadas no desenho de proteínas, a biologia sintética é, atualmente, um dos mais empolgantes e recentes campos de desenvolvimento científico e encontra-se em crescente expansão. Com relação a isto, os métodos *ab initio* são, sem dúvida, os mais versáteis em aplicações biotecnológicas e a gama de produtos (derivados de proteínas) desenhados por estes métodos não param de crescer. Atualmente, um grande número de trabalhos com este propósito tem sido publicados na literatura, incluindo propósitos e abordagens diferentes, tais como: obtenção novas topologias proteicas para o desenho de vacinas (CORREIA et al., 2014), enzimas desenhadas para exercer novas catálises (ALTHOFF et al., 2012), bem como, nanomateriais com subunidades proteicas capazes de automontagem que são interessantes para fins biotecnológicos (KING et al., 2014).

Com o aumento da qualidade de predição dos métodos *ab initio*, é possível que grupos de pesquisa comecem a adotar estes métodos na sua rotina, podendo se tornar ferramentas não somente auxiliares de predição, mas também concorrentes aos métodos baseados em molde, como a modelagem por homologia que hoje é largamente utilizada no estudo de mutantes e no planejamento *in silico* de fármacos. Com relação ao avanço na qualidade de predição, por exemplo, o I-TASSER tem mostrado resultados convincentes de predição que o colocam em primeiro lugar nos eventos de competição do CASP (do inglês, *Critical Assessment of Techniques for Protein Structure Prediction*), tais como CASP7, CASP8, CASP9, CASP10, CASP11, CASP12, e CASP13 (KRYSHTAFOVYCH et al., 2019). Além disso, recentemente foi reportado que os modelos obtidos pelo algoritmo I-TASSER superam em qualidade as estruturas obtidas por cristalografia por difração de raios-X e, podem desta forma, auxiliar em métodos de triagem virtual baseados em estrutura (SBVS do inglês, *Structure-based Virtual Screening*) (DU et al., 2015; RODRIGUES et al., 2012).

Similarmente, devido ao grande avanço das metodologias de modelagem, é possível que futuramente os métodos *in silico* não baseados em molde, no qual se inclui exclusivamente a predição *ab initio* sejam, gradativamente, mais aplicados com propósitos que requerem detalhes estruturais mais precisos como o planejamento *in silico* de fármacos e no estudo de mecanismos catalítico-enzimáticos.

GLOSSÁRIO DE TERMOS TÉCNICOS E SIGLAS

A

Ab initio: Na Biologia Estrutural, refere-se a uma abordagem de modelagem tridimensional que não depende de estruturas de referência. Neste método, são utilizadas a análise do espaço conformacional unida a uma função de pontuação de energia, a fim de encontrar as estruturas com o mínimo de energia potencial global.

Aminoácido: São os blocos construtores das proteínas. Cada aminoácido é composto por um átomo de carbono quiral denominado de carbono alfa (Ca) que forma ligações com o grupo amina, o carboxila, uma cadeia lateral que variam de acordo com o aminoácido e o hidrogênio.

Alça: Regiões conformacionalmente variáveis da estrutura tridimensional de proteínas que conectam às estruturas secundárias.

Alinhamento: Método computacional de organização de duas ou mais sequências ou estruturas cujo propósito é evidenciar similaridades ou dissimilaridades entre elas.

Algoritmo: Conjunto ordenado e lógico de instruções para execução de determinada tarefa computacional.

B

BLAST: Do inglês *Basic Local Alignment Search Tool*, trata-se de um algoritmo de comparação de sequências utilizado para busca em bases de sequências a fim de gerar alinhamentos locais, a partir de uma sequência de interesse (no inglês, *query sequence*). É a ferramenta de pesquisa padrão do NCBI GenBank.

BLOSUM: Do inglês *Blocks Substitution Matrix*. É um conjunto de matrizes de substituição aplicada às proteínas, as quais as pontuações para cada posição são derivadas de frequências de substituições em blocos de alinhamentos locais em proteínas relacionadas. Cada matriz é adaptada para uma dada distância evolutiva. Exemplo: As matrizes BLOSUM62 derivam de alinhamentos que compartilham 62% de identidade.

C

Campo de força: Conjunto de parâmetros e funções matemáticas que são utilizados para descrever a energia potencial total de um sistema.

CASP: Do inglês, *Critical Assessment of Techniques for Protein Structure Prediction*, são eventos científicos que ocorrem bianualmente em que diferentes métodos e algoritmos são testados levando em consideração a capacidade de prever a estrutura tridimensional de proteínas na sua conformação nativa.

Carbono alfa (Ca): Carbono quiral presente em todos os aminoácidos. Forma quatro grupos de ligações simples: carboxila, amina, hidrogênio e o radical variável que varia de acordo com o aminoácido.

Compilador: Programa utilizado para transformar código fonte escrito em determinada linguagem de programação em um programa executável.

Conformação: Arranjo tridimensional de átomos e de ligações em uma molécula que descrevem sua geometria.

Conformação nativa: Corresponde à conformação final de uma proteína (ou ácido nucleico) na qual esta desempenha sua função biológica. É a conformação funcional.

CPU: Do inglês, *Central Process Unit* (Unidade Central de Processamento). É o componente principal do computador, também é conhecido como processador. A CPU é responsável por calcular e realizar tarefas determinadas pelo usuário.

C-terminal: Porção final da cadeia polipeptídica da proteína que contém um grupo carboxila.

D

Docagem: Do inglês *docking*. É o método computacional que tenta prever a melhor afinidade e complementariedade de duas moléculas quando ligadas a um complexo estável.

Domínio: Em Biologia Estrutural, se trata de regiões ou porções da proteína que exibe sequência conservada evolutivamente. Normalmente, são regiões funcionais da estrutura.

Diagrama de Ribbon: Representação esquemática tridimensional da estrutura de proteínas, em que as alfa-hélices são representadas por fitas enroladas e as beta-folhas, por setas.

DOPE: Do inglês *Discrete Optimized Protein Energy*. Trata-se de uma unidade de energia potencial utilizada pelo programa de modelagem Modeller.

E

Enovelamento: Processo físico no qual uma proteína por meio de interações com o meio ou ação de outras proteínas ou cofatores atingem sua conformação nativa.

Enzima: Proteínas que apresentam função catalítica, isto é, atuam como catalisadoras de reações químicas.

F

FASTA: Formato de representação de sequências biológicas largamente aplicado em bases de dados. É aplicado para a representação de sequências de nucleotídeos e aminoácidos.

Forças de Solvatação: Forças de atração e repulsão causado pelas moléculas dispersas-se em uma solução contra os íons ou outras moléculas que constituem o soluto. Nas proteínas, estas forças normalmente são exercidas pelas moléculas de água.

Função de pontuação de energia: Função pela qual se avalia o estado conformacional de uma proteína por meio de cálculos de energia potencial. As funções variam para cada algoritmo.

G

Gráfico de Ramachandran: Gráfico de dispersão que representa a estrutura de proteínas por meio dos ângulos de torção ϕ e ψ dos aminoácidos.

GUI: Do inglês *Graphical User Interface* (Interface Gráfica ao Usuário). É a interface gráfica que permite a interação do usuário com o sistema operacional.

Genoma: Conjunto haplóide completo de todos os genes e demais sequências presentes em um organismo.

Gap: Na tradução lacuna. Em Bioinformática, se refere às regiões de não correspondência entre duas sequências em alinhamentos e são representadas por traços.

GPU: Do inglês *Graphics Processing Unit*. São unidades de processamento especializado em processar gráficos.

H

Homologia: Relação evolutiva entre duas estruturas em que estas compartilham um ancestral em comum (mesma origem filogenética). Na Biologia Estrutural, duas proteínas são consideradas homólogas quando descendem de um mesmo ancestral.

I

Input: Expressão em inglês que na linguagem da área computacional se refere a todo dado utilizado como entrada para execução de um processo.

Identidade: Porcentagem de caracteres similares entre duas sequências quando alinhadas (excluindo-se os *gaps*).

Iteração: Em computação, se refere à uma série de passos em um algoritmo por meio do qual o processamento de dados é executado repetitivamente até o resultado exceder um limite particular.

L

Ligante: Molécula que se liga a um determinado bioreceptor, podendo agir como um co-fator, substrato ou inibidor.

Loop: ver alça.

M

Matches: Na tradução do inglês *correspondência*. Caracteres idênticos em duas sequências alinhadas que ocupam a mesma coluna.

Metilação: Modificação pós-traducional que correm em proteínas. Corresponde à adição de grupos metil à estrutura.

Mismatches: Na tradução do inglês *não correspondência*. Caracteres diferentes em duas

seqüências alinhadas que ocupam a mesma coluna.

Molde: Estrutura utilizada como referência em métodos de modelagem por homologia. Termo deriva do inglês *template*.

Modelagem por homologia: Método de predição da estrutura de proteínas que utiliza estruturas de referência que apresentam similaridade na seqüência devido à ancestralidade comum (homologia).

Motivos: Regiões curtas e conservadas estruturalmente de uma seqüência de proteínas. Geralmente são partes altamente conservadas dos domínios.

N

NGS: Do inglês *Next Generation Sequencing* (Sequenciamento de Nova Geração). Refere-se a todas as metodologias de sequenciamento nascidas após 2005 e incluem, entre outros, os métodos de sequenciamento por ligação, sequenciamento por detecção de prótons, e o pirosequenciamento.

Nucleotídeo: Correspondem aos blocos construtores dos ácidos nucleicos. São constituídos por um fosfato, açúcar e uma base nitrogenada.

N-terminal: Porção inicial da seqüência polipeptídica da proteína que contém o grupo amina. Normalmente, inicia com um resíduo de metionina.

O

Open source: Expressão em inglês que se refere aos programas nos quais a código fonte é aberto, isto é, as informações são livres.

Output: Expressão em inglês que se refere aos dados liberados da execução de um processo computacional.

P

Protein Data Bank: Base de dados que contém a estrutura tridimensional de macromoléculas biológicas elucidadas por métodos experimentais, tais como, cristalografia de raios-X, espectroscopia de Ressonância Magnética Nuclear (RMN) e crio-microscopia eletrônica.

PDB: Formato de arquivo utilizado para a representação da estrutura tridimensional de moléculas. Atualmente é o formato amplamente difundido e interpretado por diferentes programas de análise de estrutura. É representado pela extensão .PDB.

PCR: Do inglês *Polymerase Chain Reaction*, trata-se de uma técnica da biologia molecular que realiza a amplificação de segmentos específicos de DNA em uma mistura complexa, na qual estão presentes também curtos iniciadores (denominados primers) oligonucleotídeos para o segmento de interesse, além de reagentes para síntese de DNA.

R

Refinamento: Corresponde à otimização na estrutura tridimensional de uma macromolécula de modo a corrigir distorções ou erros esterequímicos.

Root Mean Square Deviation (RMSD): Cálculo métrico que exhibe a distância média entre duas proteínas estruturalmente alinhadas. Normalmente, utilizam-se como referência os átomos de carbono α .

Rotâmero: Conformação mais comum da cadeia lateral de resíduos de aminoácidos. São gerados por modificações nos ângulos de rotação.

Ribbon: Ver *Diagrama de Ribbon*.

S

Similaridade: Porcentual de resíduos de aminoácidos similares entre duas proteínas obtidos após o alinhamento de duas sequências. Resíduos similares são aqueles que compartilham características físico-químicas semelhantes.

Simulated Annealing (SA): Na predição estrutural de proteínas, trata-se de um algoritmo que aplica o Metropolis MC para a minimização da estrutura. Este gera perturbações aleatórias nos rotâmeros por meio de ciclos de aquecimento e resfriamento seguindo a distribuição canônica de energia de Boltzman para determinada temperatura com objetivo de se encontrar a menor energia global da estrutura.

Script: Conjunto de códigos escritos em uma linguagem de programação específica que são interpretados por meio de um programa.

SDF: Do inglês, *Structure Data File*. Formato de arquivo de estrutura desenvolvido pela empresa MDL. É representado pelas extensões .sd ou .sdf.

Simulação de Dinâmica Molecular (DM): Método computacional que utiliza a mecânica molecular para descrever o movimento atômico ao longo do tempo. Neste método, átomos e moléculas são descritos como partículas unidas por forças harmônicas ou elásticas, e o movimento atômico é resolvido pelas equações newtonianas.

T

Threading: Método de predição da estrutura de proteínas que reconhece o tipo de dobramento de proteínas homólogas distantes ou sem homologia, mas que conservam estruturas similares devido à convergência evolutiva. Assim como a modelagem por homologia, o *threading* é um método baseado em molde.

TM-score: métrica aplicada pelo programa TM-align utilizado pelo I-TASSER para medir a similaridade estrutural entre duas proteínas.

Tradução: Em Biologia Molecular, se refere ao processo no qual a molécula de RNA mensageiro é interpretada pelo ribossomo para a produção da cadeia polipeptídica da proteína.

V

Virtual Screening: Do inglês, triagem virtual, trata-se de uma abordagem computacional que realiza a triagem/filtragem de compostos com aplicações no desenvolvimento de novos fármacos. Há a abordagem baseada em estrutura e a baseada no ligante.

Z

Z-score: método estatístico que indica a separação de valores em relação à sua contraparte. É representado pela fórmula: $\text{valor} - \text{média} / \text{Desvio padrão}$. Na Biologia Estrutural é utilizado para acessar a significância energética de determinada estrutura com uma amostra de proteínas oriundas de base de dados ou preditas por métodos computacionais.

REFERÊNCIAS

ALTHOFF, E. A. *et al.* Robust design and optimization of retroaldol enzymes. **Protein Science**, v. 21, n. 5, p. 717-26, 2012.

ALTSCHUL, S. F. *et al.* **Gapped BLAST and PSI-BLAST: A new generation of protein database search programs** **Nucleic Acids Research**, 1997.

ANFINSEN, C. B. Principles that govern the folding of protein chains. **Science**, v. 181, n. 4096, p. 223-30, jul., 1973.

BAKER, D. An exciting but challenging road ahead for computational enzyme design. **Protein Science**, v. 19, n. 10, p. 1817-19, out., 2010.

BAKER, M. L. *et al.* Ab initio modeling of the herpesvirus VP26 core domain assessed by cryoEM density. **PLoS Computational Biology**, v. 2, n. 10, p. 1313-24, 2006.

BAR-ON, D. *et al.* Dynamic Conformational Changes in MUNC18 Prevent Syntaxin Binding. **PLoS Computational Biology**, v. 7, n. 3, p. e1001097, 3 mar., 2011.

BAXEVANIS, A. D.; OUELLETTE, B. F. F. **BIOINFORMATICS A Practical Guide to the Analysis of Genes and Proteins SECOND EDITION**. [s.l.: s.n.].

BEN-DAVID, M. *et al.* Assessment of CASP8 structure predictions for template free targets. **Proteins: Structure, Function and Bioinformatics**, v. 77, n. SUPPL. 9, p. 50-65, 2009.

BEN-ZVI, A. P.; GOLOUBINOFF, P. Review: Mechanisms of Disaggregation and Refolding of Stable Protein Aggregates by Molecular Chaperones. **Journal of Structural Biology**, v. 135, n. 2, p. 84-93, ago., 2001.

BENKERT, P.; SCHWEDE, T.; TOSATTO, S. C. QMEANclust: estimation of protein model quality by combining a composite scoring function with structural density information. **BMC structural biology**, v. 9, p. 35, jan., 2009.

BENSON, D. A. *et al.* GenBank. **Nucleic Acids Research**, v. 45, n. D1, p. D37-D42, jan. 2017.

BERMAN, H. M. *et al.* The Protein Data Bank. In: **Structural Bioinformatics**. [s.l.: s.n.]. v. 28p. 181-198.

BIASINI, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, v. 42, n. W1, p. W252–W258, jul., 2014.

BONNEAU, R. *et al.* Rosetta in CASP4: Progress in ab initio protein structure prediction. **Proteins: Structure, Function and Genetics**, v. 45, n. SUPPL. 5, p. 119–126, 2001.

BONNEAU, R. *et al.* De Novo Prediction of Three-dimensional Structures for Major Protein Families. **Journal of Molecular Biology**, v. 322, n. 1, p. 65-78, set., 2002.

BOWIE, J. U.; LÜTHY, R.; EISENBERG, D. A method to identify protein sequences that fold into a known three-dimensional structure. **Science (New York, N.Y.)**, v. 253, n. 5016, p. 164-170, 1991.

- BRADLEY, P.; MISURA, K. M. S.; BAKER, D. Toward high-resolution de novo structure prediction for small proteins. **Science (New York, N.Y.)**, v. 309, n. 5742, p. 1868-71, 2005.
- BROOKS, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. **Journal of Computational Chemistry**, v. 4, n. 2, p. 187-217, 1983.
- BRYANT, S. H.; LAWRENCE, C. E. An empirical energy function for threading protein sequence through the folding motif. **Proteins**, v. 16, n. 1, p. 92-112, 1993.
- CASE, D. A. *et al.* The Amber biomolecular simulation programs. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1668-88, dez., 2005.
- CAVASOTTO, C. N.; PHATAK, S. S. Homology modeling in drug discovery: current trends and applications. **Drug Discovery Today**, v. 14, n. 13-14, p. 676-83, jul., 2009.
- CHENG, J.; BALDI, P. A machine learning information retrieval approach to protein fold recognition. **Bioinformatics**, v. 22, n. 12, p. 1456-63, 2006.
- CHENNA, R. Multiple sequence alignment with the Clustal series of programs. **Nucleic Acids Research**, v. 31, n. 13, p. 3497-3500, jul., 2003.
- CORREIA, B. E. *et al.* Proof of principle for epitope-focused vaccine design. **Nature**, v. 507, n. 7491, p. 201-6, 2014.
- COSTA, C. H. S. *et al.* Computational study of conformational changes in human 3-hydroxy-3-methylglutaryl coenzyme reductase induced by substrate binding. **Journal of Biomolecular Structure and Dynamics**, v. 37, n. 16, p. 4374-83, 23 nov., 2019.
- DA COSTA, K. S. *et al.* Structural analysis of viral infectivity factor of HIV type 1 and its interaction with A3G, EloC and EloB. **PLoS ONE**, v. 9, n. 2, p. e89116, jan., 2014.
- DA COSTA, K. S. *et al.* Structural and evolutionary analyses of Leishmania Alba proteins. **Molecular and Biochemical Parasitology**, v. 217, p. 23-31, 2017.
- DA COSTA, K. S. *et al.* Exploring the Potentiality of Natural Products from Essential Oils as Inhibitors of Odorant-Binding Proteins: A Structure- and Ligand-Based Virtual Screening Approach To Find Novel Mosquito Repellents. **ACS Omega**, v. 4, n. 27, p. 22475-86, 31 dez., 2019.
- DI TOMMASO, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. **Nucleic acids research**, v. 39, n. Web Server issue, p. W13-7, 1 jul., 2011.
- DING, F. *et al.* Ab Initio Folding of Proteins with All-Atom Discrete Molecular Dynamics. **Structure**, v. 16, n. 7, p. 1010-18, 2008.
- DU, H. *et al.* Protein structure prediction provides comparable performance to crystallographic structures in docking-based virtual screening. **Methods**, v. 71, p. 77-84, 2015.
- EDGAR, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792-7, 2004.

- FAN, H. Refinement of homology-based protein structures by molecular dynamics simulation techniques. **Protein Science**, v. 13, n. 1, p. 211-20, 1 jan., 2004.
- FISER, A. Template-based protein structure modeling. **Methods in molecular biology (Clifton, N.J.)**, v. 673, p. 73-94, 2010.
- FISER, A.; ŠALI, A. MODELLER: Generation and Refinement of Homology-Based Protein Structure Models. **Methods in Enzymology**, v. 374, p. 461-91, jan., 2003.
- FLOUDAS, C. A. *et al.* Advances in protein structure prediction and de novo protein design: A review. **Chemical Engineering Science**, v. 61, n. 3, p. 966-88, fev. 2006.
- GIGER, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. **Nature chemical biology**, v. 9, n. 8, p. 494-8, 2013.
- GOPAL, S. M.; KLENIN, K.; WENZEL, W. Template-free protein structure prediction and quality assessment with an all-atom free-energy model. **Proteins**, v. 77, n. 2, p. 330-41, 1 nov., 2009.
- GUO, J. -T. *et al.* PROSPECT-PSPP: an automatic computational pipeline for protein structure prediction. **Nucleic Acids Research**, v. 32, n. Web Server, p. W522–W525, 1 jul., 2004.
- HANSMANN, U. H. E.; OKAMOTO, Y. New Monte Carlo algorithms for protein folding. **Current Opinion in Structural Biology**, v. 9, n. 2, p. 177-83, abr., 1999.
- HAO, M.-H.; SCHERAGAT, H. A. Designing potential energy functions for protein folding. **Current Opinion in Structural Biology**, v. 9, n. 2, p. 184-8, abr., 1999.
- HUANG, Y. J. *et al.* Assessment of template-based protein structure predictions in CASP10. **Proteins: Structure, Function and Bioinformatics**, v. 82, n. SUPPL.2, p. 43-56, 2014.
- IHM, Y. *et al.* Structural Model of the Rev Regulatory Protein from Equine Infectious Anemia Virus. **PLoS ONE**, v. 4, n. 1, p. e4178, 12 jan., 2009.
- JÄCKEL, C.; KAST, P.; HILVERT, D. Protein design by directed evolution. **Annual review of biophysics**, v. 37, p. 153-73, 2008.
- JIANG, L. *et al.* De novo computational design of retro-aldol enzymes. **Science (New York, N.Y.)**, v. 319, n. 5868, p. 1387-91, 2008.
- JONES, D. T. *et al.* Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. **Proteins: Structure, Function and Genetics**, v. 37, n. SUPPL. 3, p. 104-11, 1999.
- JONES, D. T. Protein structure prediction in the postgenomic era. **Current opinion in structural biology**, v. 10, n. 3, p. 371-9, 2000.
- JONES, D. T.; MILLER, R. T.; THORNTON, J. M. Successful protein fold recognition by optimal sequence threading validated by rigorous blind testing. **Proteins: Structure, Function and Genetics**, v. 23, n. 3, p. 387-97, 1995.

JONES, D. T.; TAYLOR, W. R.; THORNTON, J. M. A new approach to protein fold recognition. **Nature**, v. 358, n. 6381, p. 86-9, 1992.

JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. **Journal of the American Chemical Society**, v. 118, n. 45, p. 11225-36, jan., 1996.

JOSHI, R. R.; JYOTHI, S. Ab-initio prediction and reliability of protein structural genomics by PROPAINOR algorithm. **Computational Biology and Chemistry**, v. 27, n. 3, p. 241-52, 2003.

KÄLLBERG, M. *et al.* **Template-based protein structure modeling using the RaptorX web server** **Nature Protocols**, 2012.

KAUFMANN, K. W. *et al.* Practically useful: what the Rosetta protein modeling suite can do for you. **Biochemistry**, v. 49, n. 14, p. 2987-98, 13 abr., 2010.

KELLEY, L. A. *et al.* The Phyre2 web portal for protein modeling, prediction and analysis. **Nature protocols**, v. 10, n. 6, p. 845-58, 2015.

KEMEGE, K. E. *et al.* Ab initio structural modeling of and experimental validation for Chlamydia trachomatis protein CT296 reveal structural similarity to Fe(II) 2-Oxoglutarate-dependent enzymes. **Journal of Bacteriology**, v. 193, n. 23, p. 6517-28, 2011.

KHERSONSKY, O. *et al.* Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. **Journal of Molecular Biology**, v. 407, n. 3, p. 391-412, 2011.

KHOURY, G. A. *et al.* Protein folding and de novo protein design for biotechnological applications. **Trends in Biotechnology**, v. 32, n. 2, p. 99-109, 2014.

KIDD, B. A.; BAKER, D.; THOMAS, W. E. Computation of Conformational Coupling in Allosteric Proteins. **PLoS Computational Biology**, v. 5, n. 8, p. e1000484, 28 ago., 2009.

KIEFER, F. *et al.* The SWISS-MODEL Repository and associated resources. **Nucleic Acids Research**, v. 37, n. SUPPL. 1, 2009.

KIHARA, D. *et al.* TOUCHSTONE: an ab initio protein structure prediction method that uses threading-based tertiary restraints. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 18, p. 10125-30, 2001.

KIHARA, D. *et al.* Ab initio protein structure prediction on a genomic scale: application to the Mycoplasma genitalium genome. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 9, p. 5993-8, 2002.

KIM, D. E.; CHIVIAN, D.; BAKER, D. Protein structure prediction and analysis using the Robetta server. **Nucleic Acids Research**, v. 32, n. Web Server, p. W526–W531, 1 jul. 2004.

KINCH, L. *et al.* CASP9 assessment of free modeling target predictions. **Proteins: Structure, Function and Bioinformatics**, v. 79, n. SUPPL. 10, p. 59-73, jan. 2011.

KING, N. P. *et al.* Accurate design of co-assembling multi-component protein nanomaterials. **Nature**, v. 510, n. 7503, p. 103-8, 2014.

KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P. Optimization by simulated annealing. **Science (New York, N.Y.)**, v. 220, n. 4598, p. 671-80, 1983.

KLEPEIS, J. L.; FLOUDAS, C. A. ASTRO-FOLD: a combinatorial and global optimization framework for Ab initio prediction of three-dimensional structures of proteins from the amino acid sequence. **Biophysical journal**, v. 85, n. 4, p. 2119-46, 2003.

KORETKE, K. K. *et al.* Fold recognition using sequence and secondary structure information. **Proteins**, v. Suppl 3, p. 141-8, 1999.

KRYSHTAFOVYCH, A. *et al.* **Protein structure prediction center in CASP8** **Proteins: Structure, Function and Bioinformatics**, 2009.

KRYSHTAFOVYCH, A. *et al.* Critical assessment of methods of protein structure prediction (CASP)—Round XIII. **Proteins: Structure, Function, and Bioinformatics**, v. 87, n. 12, p. 1011-20, 23 dez., 2019.

KRYSHTAFOVYCH, A.; FIDELIS, K.; MOULT, J. CASP10 results compared to those of previous CASP experiments. **Proteins: Structure, Function and Bioinformatics**, v. 82, n. SUPPL.2, p. 164-74, fev., 2014.

KUHLMAN, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. **Science (New York, N.Y.)**, v. 302, n. 5649, p. 1364-68, 2003.

KUHLMAN, B.; BAKER, D. Native protein sequences are close to optimal for their structures. **Proceedings of the National Academy of Sciences of the United States of America**, v. 97, n. 19, p. 10383-8, 12 set., 2000.

LAMEIRA, J. *et al.* Predicting the affinity of halogenated reversible covalent inhibitors through relative binding free energy. **Physical Chemistry Chemical Physics**, v. 21, n. 44, p. 24723-30, 2019.

LASKOWSKI, R. A. *et al.* **PROCHECK: a program to check the stereochemical quality of protein structures**. Disponível em: <<http://scripts.iucr.org/cgi-bin/paper?S0021889892009944>>.

LEVINTHAL, C. Are there pathways for protein folding? **Journal de Chimie Physique et de Physico-Chimie Biologique**, v. 65, p. 44-5, 1968.

LI, K. *et al.* Design and synthesis of novel 2-substituted 11-keto-boswellic acid heterocyclic derivatives as anti-prostate cancer agents with Pin1 inhibition ability. **European Journal of Medicinal Chemistry**, v. 126, p. 910-19, jan., 2017.

LIWO, A. *et al.* United-residue force field for off-lattice protein-structure simulations: III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. **Journal of Computational Chemistry**, fev., 1998.

LIWO, A.; KHALILI, M.; SCHERAGA, H. A. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. **Proceedings of the National Academy of Sciences**, v. 102, n. 7, p. 2362-67, 15 fev., 2005.

LOBLEY, A.; SADOWSKI, M. I.; JONES, D. T. pGenTHREADER and pDomTHREADER: New methods for improved protein fold recognition and superfamily discrimination. **Bioinformatics**, v. 25, n. 14, p. 1761-67, 15 jul., 2009.

LOVELL, S. C. *et al.* Structure validation by C α geometry: phi,psi and C β deviation. **Proteins**, v. 50, n. 3, p. 437-50, 15 fev., 2003.

MAGRANE, M.; CONSORTIUM, U. P. UniProt Knowledgebase: A hub of integrated protein data. **Database**, v. 2011, p. bar009–bar009, 29 mar., 2011.

MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The PSIPRED protein structure prediction server. **Bioinformatics (Oxford, England)**, v. 16, n. 4, p. 404-5, abr., 2000.

MELO, F.; FEYTMANS, E. Assessing protein structures with a non-local atomic interaction energy. **Journal of Molecular Biology**, v. 277, n. 5, p. 1141-52, abr., 1998.

METROPOLIS, N. *et al.* Equation of State Calculations by Fast Computing Machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087-92, 1953.

MORAES, G. *et al.* Homology modeling, molecular dynamics and QM/MM study of the regulatory protein PhoP from *Corynebacterium pseudotuberculosis*. **Journal of molecular modeling**, v. 18, n. 3, p. 1219-27, mar., 2012.

NEVES CRUZ, J. *et al.* Measuring the structural impact of mutations on cytochrome P450 21A2, the major steroid 21-hydroxylase related to congenital adrenal hyperplasia. **Journal of Biomolecular Structure and Dynamics**, p. 1-10, 14 abr., 2019.

NGUYEN, H. *et al.* Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. **Journal of the American Chemical Society**, v. 136, n. 40, p. 13959-62, 8 out., 2014.

NOCUA, P. A. *et al.* Leishmania braziliensis replication protein A subunit 1: molecular modelling, protein expression and analysis of its affinity for both DNA and RNA. **Parasites & Vectors**, v. 7, n. 1, p. 573, 2014.

OKAMOTO, Y. Monte-Carlo Simulated Annealing in Protein Folding. In: **Encyclopedia of Optimization**. Boston, MA: Springer US, 2009. p. 2323-37.

PILLARDY, J. *et al.* Development of Physics-Based Energy Functions that Predict Medium-Resolution Structures for Proteins of the α , β , and α/β Structural Classes. **The Journal of Physical Chemistry B**, v. 105, n. 30, p. 7299-311, ago., 2001.

RAMAN, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. **Proteins**, v. 77, p. 89-99, jan. 2009.

RAVAL, A. *et al.* Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. **Proteins: Structure, Function, and Bioinformatics**, p. n/a-n/a, 2012.

RICHTER, F. *et al.* De novo enzyme design using Rosetta3. **PLoS ONE**, v. 6, n. 5, 2011.

RODRIGUES, R. P. *et al.* Virtual Screening Strategies in Drug Design. **Revista Virtual de Química**, v. 4, n. 6, 2012.

ROHL, C. A *et al.* Protein structure prediction using Rosetta. **Methods in enzymology**, v. 383, n. 2003, p. 66-93, jan. 2004.

ROST, B.; SCHNEIDER, R.; SANDER, C. Protein fold recognition by prediction-based threading. **Journal of molecular biology**, v. 270, n. 3, p. 471-80, 1997.

RÖTHLISBERGER, D. et al. Kemp elimination catalysts by computational enzyme design. **Nature**, v. 453, n. 7192, p. 190-5, 2008.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature Protocols**, v. 5, n. 4, p. 725-38, abr., 2010.

SALOMON-FERRER, R.; CASE, D. A.; WALKER, R. C. An overview of the Amber biomolecular simulation package. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 3, n. 2, p. 198-210, mar., 2013.

SÁNCHEZ, R.; ŠALI, A. Advances in comparative protein-structure modelling. **Current Opinion in Structural Biology**, v. 7, n. 2, p. 206-14, abr., 1997.

SANTOS FILHO, O. A.; ALENCASTRO, R. B. DE; BICCA DE ALENCASTRO, R. Modelagem de proteínas por homologia. **Química Nova**, v. 26, n. 2, p. 253-9, mar., 2003.

SHI, J.; BLUNDELL, T. L.; MIZUGUCHI, K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. **Journal of molecular biology**, v. 310, n. 1, p. 243-57, 2001.

SHORTLE, D.; SIMONS, K. T.; BAKER, D. Clustering of low-energy conformations near the native structures of small proteins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 19, p. 11158-62, 1998.

SIEW, N. *et al.* MaxSub: an automated measure for the assessment of protein structure prediction quality. **Bioinformatics (Oxford, England)**, v. 16, n. 9, p. 776-85, 2000.

SIMONS, K. T. *et al.* Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. **Journal of Molecular Biology**, v. 268, n. 1, p. 209-25, 25 abr., 1997.

SIMONS, K. T. *et al.* Ab initio protein structure prediction of CASP III targets using ROSETTA. **Proteins**, v. Suppl 3, n. SUPPL. 3, p. 171-6, jan. 1999.

SIMONS, K. T.; STRAUSS, C.; BAKER, D. Prospects for ab initio protein structural genomics. **Journal of molecular biology**, v. 306, n. 5, p. 1191-99, 2001.

SINGH, S.; GUTTULA, P. K.; GURUPRASAD, L. Structure Based Annotation of Helicobacter pylori Strain 26695 Proteome. **PLoS ONE**, v. 9, n. 12, p. e115020, 2014.

SIPPL, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. **Journal of molecular biology**, v. 213, n. 4, p. 859-83, 1990.

SITTEL, F.; JAIN, A.; STOCK, G. Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates. **The Journal of Chemical Physics**, v. 141, n. 1, p. 014111, 7 jul., 2014.

SKOLNICK, J. In quest of an empirical potential for protein structure prediction. **Current Opinion in Structural Biology**, v. 16, n. 2, p. 166-71, abr., 2006.

STANLEY, B. J. *et al.* Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. **Journal of virology**, v. 82, n. 17, p. 8656-63, set., 2008.

TAI, C. H. *et al.* Assessment of template-free modeling in CASP10 and ROLL. **Proteins: Structure, Function and Bioinformatics**, v. 82, n. SUPPL.2, p. 57-83, 2014.

TAYLOR, W. R. Multiple sequence threading: an analysis of alignment quality and stability. **Journal of molecular biology**, v. 269, n. 5, p. 902-43, 1997.

THACHUK, C.; SHMYGELSKA, A.; HOOS, H. H. A replica exchange Monte Carlo algorithm for protein folding in the HP model. **BMC bioinformatics**, v. 8, p. 342, 2007.

TRAPANE, T. L.; LATTMAN, E. E. Seventh Meeting on the Critical Assessment of Techniques for Protein Structure Prediction. **Proteins: Structure, Function, and Bioinformatics**, v. 69, n. S8, p. 1-2, 2007.

TSANG, A. *et al.* Francisella tularensis 2-C-Methyl-D-Erythritol 4-Phosphate Cytidyltransferase: Kinetic Characterization and Phosphoregulation. **PLoS ONE**, v. 6, n. 6, p. e20884, 9 jun., 2011.

VAN DER SPOEL, D. *et al.* GROMACS: Fast, flexible, and free. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1701-18, dez., 2005.

VOELZ, V. A. *et al.* Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1-39). **Journal of the American Chemical Society**, v. 132, n. 5, p. 1526-28, 10 fev., 2010.

WANG, Z. X. A re-estimation for the total numbers of protein folds and superfamilies. **Protein engineering**, v. 11, n. 8, p. 621-6, 1998.

WEBB, B.; SALI, A. Comparative protein structure modeling using MODELLER. **Current Protocols in Bioinformatics**, v. 2016, n. 1, p. 5.6.1-5.6.37, set., 2016.

WIEDERSTEIN, M.; SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic acids research**, v. 35, n. Web Server issue, p. W407-10, jul., 2007.

WU, S.; ZHANG, Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information. **Proteins: Structure, Function, and Bioinformatics**, v. 72, n. 2, p. 547-56, 4 fev., 2008.

XU, D.; ZHANG, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. **Proteins: Structure, Function and Bioinformatics**, v. 80, n. 7, p. 1715-35, 2012.

XU, J.; ZHANG, Y. How significant is a protein structure similarity with TM-score = 0.5? **Bioinformatics**, v. 26, n. 7, p. 889-95, 1 abr., 2010.

ZANGHELLINI, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. **Protein science : a publication of the Protein Society**, v. 15, n. 12, p. 2785-94, 2006.

ZHANG, H. A new Hybrid Monte Carlo algorithm for protein potential function test and structure refinement. **Proteins**, v. 34, n. 4, p. 464-71, 1999.

ZHANG, Y. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic Acids Research**, v. 33, n. 7, p. 2302-9, 11 abr., 2005.

ZHANG, Y. Template-based modeling and free modeling by I-TASSER in CASP7. **Proteins: Structure, Function and Genetics**, v. 69, n. SUPPL. 8, p. 108-17, jan., 2007.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC Bioinformatics**, v. 9, n. 1, p. 40, jan. 2008.

ZHANG, Y. I-TASSER: Fully automated protein structure prediction in CASP8. **Proteins: Structure, Function and Bioinformatics**, v. 77, n. SUPPL. 9, p. 100-13, jan., 2009.

ZHANG, Y.; KOLINSKI, A.; SKOLNICK, J. TOUCHSTONE II: a new approach to ab initio protein structure prediction. **Biophysical journal**, v. 85, n. 2, p. 1145-64, 2003.

ZHANG, Y.; SKOLNICK, J. SPICKER: A clustering approach to identify near-native protein folds. **Journal of Computational Chemistry**, v. 25, n. 6, p. 865-71, 2004.

ZHOU, H.; ZHOU, Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. **Proteins: Structure, Function and Genetics**, v. 58, n. 2, p. 321-8, 2005.

ZHOU, M. *et al.* Exploring the binding mode of HIV-1 Vif inhibitors by blind docking, molecular dynamics and MM/GBSA. **RSC Advances**, v. 3, n. 44, p. 22532-43, 2013.

SOBRE OS AUTORES

KAUÊ SANTANA DA COSTA - Laboratório de Simulações e Desenvolvimento de Ferramentas Computacionais. Instituto de Biodiversidade da Universidade Federal do Oeste do Pará.

ALBERTO MONTEIRO DOS SANTOS - Laboratório de Bioinformática do Instituto de Biodiversidade, da Universidade Federal do Oeste do Pará.

ANDERSON HENRIQUE LIMA E LIMA - Laboratório de Planejamento e Desenvolvimento de Fármacos do Instituto de Ciências Exatas e Naturais, da Universidade Federal do Pará.

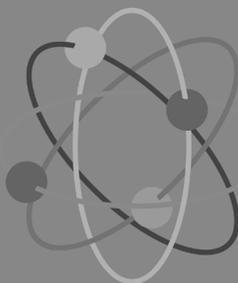
JOSÉ ROGÉRIO DE ARAÚJO SILVA - Laboratório de Planejamento e Desenvolvimento de Fármacos do Instituto de Ciências Exatas e Naturais da Universidade Federal do Pará.

JERÔNIMO LAMEIRA SILVA - Laboratório de Planejamento e Desenvolvimento de Fármacos do Instituto de Ciências Biológicas da Universidade Federal do Pará.

JOÃO MARCOS PEREIRA GALÚCIO - Laboratório de Simulações e Desenvolvimento de Ferramentas Computacionais do Instituto de Biodiversidade da Universidade Federal do Oeste do Pará.

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.

