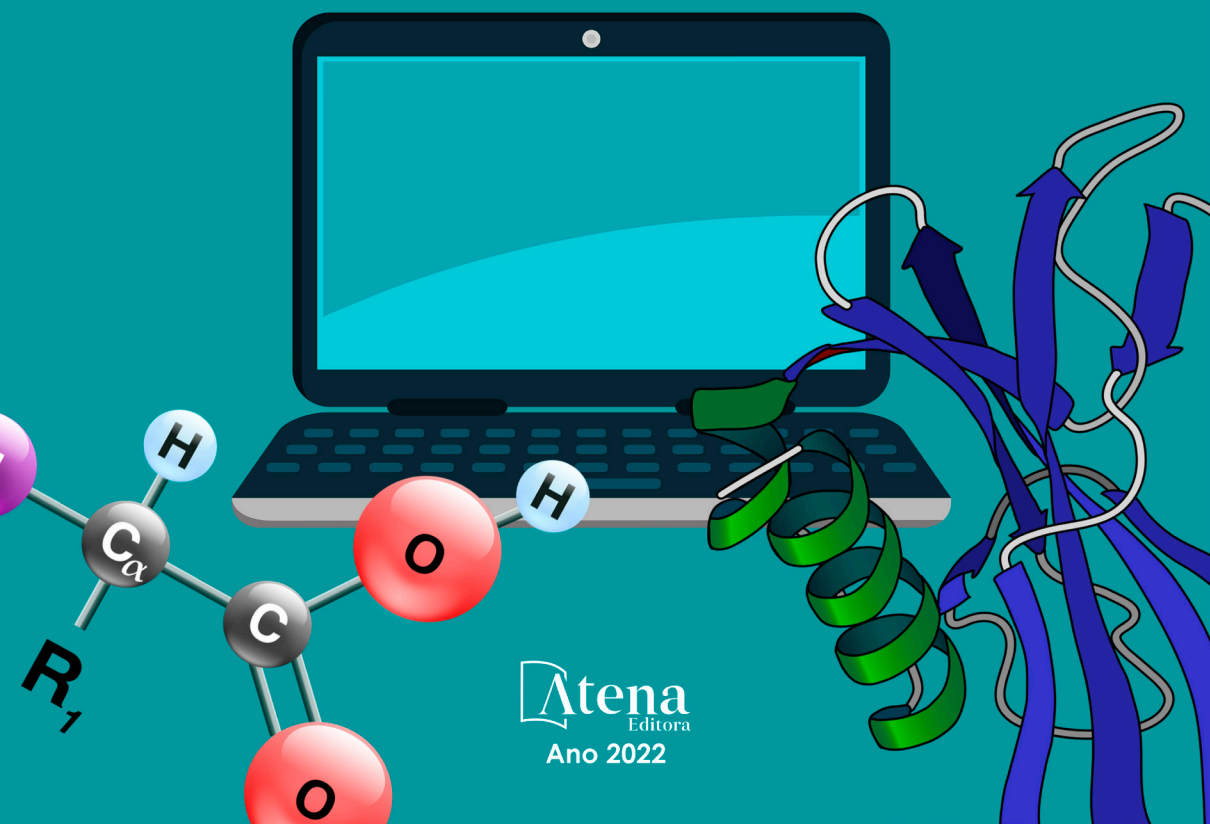


MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

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



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

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

Dados Internacionais de Catalogação na Publicação (CIP)

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

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-65-258-0582-5

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

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

CDD 613.282

Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166

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



DECLARAÇÃO DOS AUTORES

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



DECLARAÇÃO DA EDITORA

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



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

Prof. Dr. Kauê Santana

APRESENTAÇÃO

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

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

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

Prof. Kauê Santana
Prof. Cláudio Nahum

PREFÁCIO

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

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

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

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

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

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

SUMÁRIO


CAPÍTULO 1..... 1

A ESTRUTURA E O PROBLEMA DE DOBRAMENTO DE PROTEÍNAS

Kauê Santana da Costa

Anderson Henrique Lima e Lima

Jerônimo Lameira

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

CAPÍTULO 2..... 7

MÉTODOS DE PREDIÇÃO POR HOMOLOGIA

Kauê Santana da Costa

João Marcos Pereira Galúcio

José Rogério de Araújo Silva

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

CAPÍTULO 3..... 16

MODELAGEM DE PROTEÍNAS NO *MODELLER*

Kauê Santana da Costa


João Marcos Pereira Galúcio

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

CAPÍTULO 4..... 29

MODELAGEM DE PROTEÍNAS NO *EASYMODELLER*

João Marcos Pereira Galúcio

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


CAPÍTULO 5..... 33

MÉTODOS DE PREDIÇÃO *THREADING* E *AB INITIO*

Anderson Henrique Lima e Lima

Kauê Santana da Costa

Alberto Monteiro dos Santos

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


CAPÍTULO 6..... 41

ALGORÍTMOS DE MODELAGEM *THREADING* E *AB INITIO*

Anderson Henrique Lima e Lima

Kauê Santana da Costa

Alberto Monteiro dos Santos

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

CAPÍTULO 7..... 47

MODELAGEM DE PROTEÍNAS NO *I-TASSER*

Kauê Santana da Costa

Anderson Henrique Lima e Lima

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


CAPÍTULO 8..... 53

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

Anderson Henrique Lima e Lima

Alberto Monteiro dos Santos

Kauê Santana da Costa

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

CAPÍTULO 9..... 60

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

Kauê Santana da Costa

Anderson Henrique Lima e Lima

Alberto Monteiro dos Santos

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

GLOSSÁRIO DE TERMOS TÉCNICOS E SIGLAS 64

REFERÊNCIAS 70

SOBRE OS AUTORES 79

ALGORÍTMOS DE MODELAGEM *THREADING* E *AB INITIO*

Anderson Henrique Lima e Lima

Kauê Santana da Costa

Alberto Monteiro dos Santos

Neste capítulo, abordaremos em mais detalhes as estratégias utilizadas por dois principais algoritmos que mostram resultados satisfatórios nos eventos do CASP: o I-TASSER (ZHANG, 2007;5 2009) que utiliza um misto de modelagem *ab initio* e *threading* e o Rosetta (BONNEAU et al., 2001; SIMONS et al., 1999) que emprega a abordagem *ab initio* baseada em conhecimento.

ALGORITMO ROSETTA

O algoritmo Rosetta foi desenvolvido pelo grupo do pesquisador David Baker, da Universidade de Washington (Estados Unidos), e emprega a modelagem *ab initio* baseada em conhecimento que, quando comparada às demais ferramentas usadas para a modelagem de novo, tem mostrado avanços significativos para a predição da estrutura tridimensional de proteínas durante os eventos de competição do CASP (RAMAN et al., 2009; TAI et al., 2014). Devido ao seu sucesso na predição *ab initio* de proteínas, o método empregado tem sido estendido aos outros problemas abordados na modelagem, entre eles: *docking* (em inglês

docking) de proteínas e o refinamento de regiões de alças (ROHL et al., 2004). O programa está disponível na versão instalável e *on-line* (<http://rosetta.bakerlab.org/>) no servidor Robetta (do inglês *Full-chain Protein Structure Prediction Server*) também desenvolvido e mantido pelo mesmo grupo.

A estratégia para predição de proteínas utilizada pelo Rosetta baseia-se no pressuposto de que a distribuição das conformações assumidas por um fragmento pequeno da sequência alvo pode ser aproximada pela distribuição de estruturas adotadas em outras proteínas sem relação evolutiva (BONNEAU et al., 2001; KAUFMANN et al., 2010; SIMONS et al., 1999), sendo assim, os modelos construídos são compostos de fragmentos de baixa energia de estruturas conhecidas. O Rosetta utiliza duas bibliotecas de fragmentos, porém, não homólogas: uma contém sequências de três e a outra de nove resíduos de aminoácidos. A pesquisa de fragmentos para a modelagem ocorre em base de dados de estruturas obtidas por difração de raios-X com resolução de 2,5 Å, ou melhor, e com identidade menor que 50%. Para a pesquisa de fragmentos ideais, também é utilizada a predição da estrutura secundária da proteína-alvo, sendo esta realizada em três servidores: PSIPRED, Sam-T99 e JUFO. Depois de predita a estrutura secundária, o nível de similaridade dos fragmentos em relação à estrutura alvo é calculado para cada posição

(KAUFMANN et al., 2010; ROHL et al., 2004). Atualmente, a forma mais prática para a obtenção das bibliotecas de fragmentos de três e nove resíduos é através do servidor Robetta (KIM; CHIVIAN; BAKER, 2004).

Os fragmentos selecionados são então utilizados para a construção do modelo que inicia, arbitrariamente, com a proteína na forma completamente distendida. Inicialmente, uma janela de fragmentos de nove resíduos da proteína-alvo é selecionada e, em seguida, um fragmento do base de dados contendo o mesmo tamanho é escolhido dentre as 25 melhores sequências classificadas para esta posição. Depois de selecionado, ocorre a substituição dos ângulos de torção ϕ , ψ e ω do fragmento da proteína-alvo pelos dos ângulos de torção do fragmento de estrutura conhecida e a energia resultante é avaliada a cada nova inserção. O processo de montagem dos fragmentos é realizado por meio de um processo de otimização baseado no algoritmo *Monte Carlo-Simulated annealing*, e utiliza um campo de força baseado em conhecimento que descreve todos os átomos de maneira explícita (em inglês, *all-atom*). Este leva em conta potenciais de Lennard Jones para forças de van der Waals, termos eletrostáticos, potenciais de ligações de hidrogênio, aproximações de interações de solvatação e termos de energia livre interna para a conformação de aminoácidos. Depois de finalizado o processo de montagem, este é reiniciado com janelas de fragmentos da proteína alvo contendo três resíduos (RAMAN et al., 2009; ROHL et al., 2004).

O Rosetta utiliza uma representação do espaço de torção em que o esqueleto proteico é especificado, exclusivamente, por modificações aleatórias dos ângulos de torção ϕ , ψ e ω . Desta forma, os comprimentos e ângulos de ligação de resíduos individuais são mantidos fixos e o espaço cartesiano é gerado pelas coordenadas atômicas dos átomos pesados do esqueleto proteico (ROHL et al., 2004). Esta estratégia, evidentemente, reduz o tempo computacional por diminuir os graus de liberdade da estrutura proteica e, conseqüentemente, o espaço de pesquisa conformacional. O funcionamento do algoritmo *ab initio* empregado pelo Rosetta é mostrado resumidamente na figura 23.

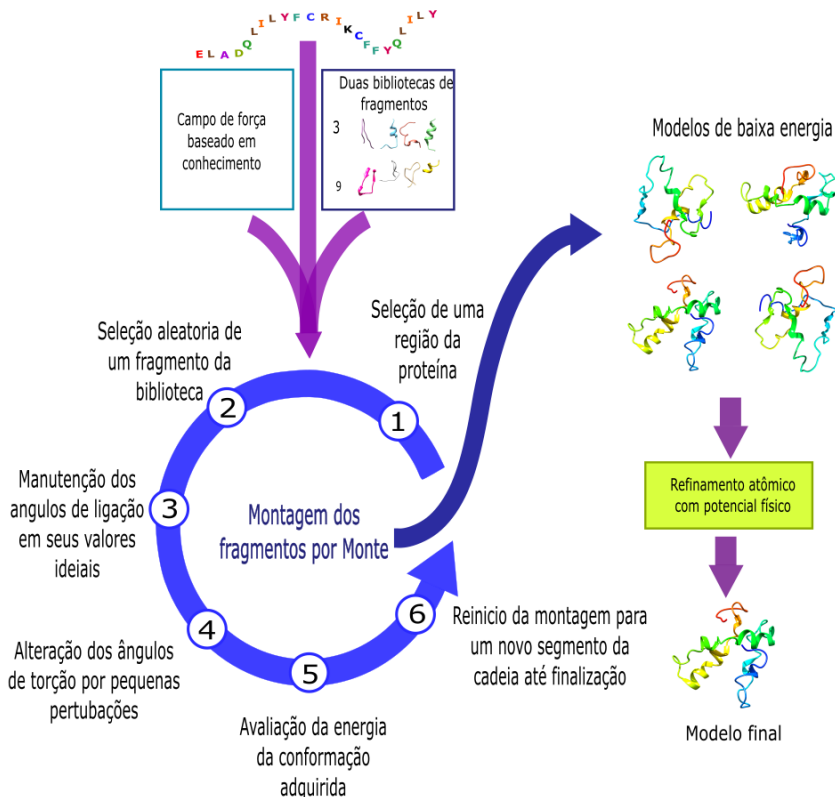


Figura 23. Resumo do funcionamento do algoritmo *ab initio* do Rosetta

A seleção dos modelos no Rosetta é baseada no menor perfil de energia da estrutura resultante e, para isto, se utiliza da função de pontuação de energia (ROHL et al., 2004; SIMONS et al., 1999). A maximização da função de pontuação do Rosetta segue a separação Bayesiana da energia total em componentes que descrevem a vizinhança da estrutura analisada, durante a simulação que emprega Monte Carlo. Baseando-se no fato de que diferentes estruturas apresentam conformações mínimas de energia próximas à estrutura nativa, o Rosetta também agrupa os modelos de acordo com a similaridade estrutural (com $C\alpha$ RMSD $< 5,0 \text{ \AA}$) e seleciona aqueles que aparecem no centro de cada um, isto é, seleciona aquelas estruturas que melhor representam as demais com a menor configuração de energia potencial (centroides), sendo esta, calculada através de todos os termos que compõem a sua função de pontuação de energia (KAUFMANN et al., 2010). Para informações mais detalhadas do protocolo de agrupamento utilizado pelo Rosetta, o leitor pode consultar Shortle et al. (2018) (SHORTLE; SIMONS; BAKER, 1998).

Para obter características de baixa resolução das estruturas nativas e melhor discriminar os modelos nativos dos demais, as conformações de baixa resolução

selecionadas são submetidas a um refinamento da estrutura que aplica um campo de força com funções de potenciais físicos (RAMAN et al., 2009).

O Rosetta utiliza um protocolo denominado *AbinitioRelax* que consiste em duas etapas principais na modelagem: etapa de montagem *ab initio*, dos fragmentos e a etapa de otimização (relaxamento) estrutural (BRADLEY; MISURA; BAKER, 2005). Na primeira etapa, é realizada uma pesquisa de granulação-grossa (do inglês *coarse-grained*) baseada em fragmentos no qual o espaço conformacional da proteína de interesse é pesquisado usando uma função de pontuação de energia “centroide” que permite determinar características, que se adequem à conformações nativas de proteínas (etapa *ab initio*). O segundo passo consiste no refinamento de todos os átomos usando o campo de força explícito detalhado do Rosetta (etapa *relax*). A segunda etapa, como envolve alterações nos ângulos de torção da proteína, demanda maior tempo computacional e também utiliza o protocolo de Minimização Monte Carlo (MCM) que em cada movimento da estrutura emprega: perturbações aleatórias dos ângulos de torção da cadeia principal da molécula (ϕ e ψ) e a otimização combinatória das conformações e rotações da cadeia principal e das cadeias laterais (RAMAN et al., 2009).

ALGORITMO I-TASSER

O algoritmo I-TASSER (do inglês *Iterative Threading ASSEmbly Refinement*) foi desenvolvido pelo grupo de Yang Zhang e consiste num algoritmo de modelagem avançada que emprega, em linhas gerais, um misto de predição por *threading* seguida de refinamento da estrutura proteica e predição *ab initio*. Comparado a outros métodos, tem mostrado resultados convincentes em consecutivos eventos do CASP (ZHANG, 2007, 2009). O programa está atualmente disponível para utilização acadêmica no servidor *on-line* mantido pela Universidade de Michigan (EUA) e na forma instalável compatível com sistemas Linux.

O funcionamento do algoritmo I-TASSER pode ser subdividido em três principais etapas.

Primeiramente, as sequências alvo são pesquisadas através de uma biblioteca de estruturas em formato PDB usando como valor de *cut-off* 70% em relação à identidade da sequência. Em primeiro lugar, o servidor faz uma pesquisa em base de dados não redundante pelo PSI-BLAST com o objetivo de encontrar aquelas que apresentam semelhanças estruturais. Baseando-se nos múltiplos alinhamentos com o seu homólogo, um perfil da sequência é então criado e o alinhamento também é utilizado para prever a estrutura secundária pelo servidor Psipred (MCGUFFIN; BRYSON; JONES, 2000). Obtidos o perfil da proteína e a estrutura secundária, as sequências analisadas são pesquisadas por *threading* em base de estruturas em formato PDB, usando o servidor local Lomets que combina sete programas de pesquisa *threading* que incluem entre outros, o Fugue (SHI; BLUNDELL; MIZUGUCHI, 2001), Prospect (GUO et al., 2004) e o Muster (WU; ZHANG,

2008). Com base nos perfis obtidos por estes servidores, as estruturas de referência são pontuadas e a qualidade do alinhamento é avaliada pelo Z-score que mostra a significância estatística da pontuação de energia (ROY; KUCUKURAL; ZHANG, 2010).

Na segunda etapa, as regiões pesquisadas por *threading* com alta pontuação que se alinham com a sequência alvo são, em seguida, removidas para posterior montagem deles na estrutura da proteína alvo utilizando simulações de Monte Carlo. As regiões da sequência alvo, que não alinham com as sequências pesquisadas por *threading*, são construídas por modelagem *ab initio*. O algoritmo I-TASSER utiliza para a representação da proteína, um modelo reduzido, em que cada resíduo na cadeia principal é especificado pelo carbono α e as cadeias laterais, pelo seu centro de massa. As estruturas modeladas (modelos) de menor energia potencial são agrupadas pelo algoritmo SPIKER e aquelas que melhor representam as demais do agrupamento (denominadas centroides) são selecionadas para posterior otimização.

Na terceira etapa, ocorre um segundo ciclo de modelagem, utilizando como ponto de partida, o centroide, obtido na etapa anterior e, nesta etapa, pretende-se obter estruturas de menor energia. Restrições espaciais que correspondem aos ângulos diedrais do primeiro centroide são extraídas e novamente estruturas experimentais em PDB são pesquisadas pelo programa de alinhamento estrutural TM-align (ZHANG, 2005).

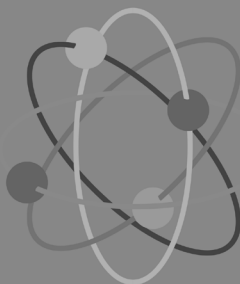
O propósito desta segunda rodada de modelagem é remover possíveis conflitos estéricos da primeira, melhorando assim, a topologia da estrutura. A pontuação TM (em inglês, *TM-score*) é utilizada como critério quantitativo para mensurar a similaridade estrutural de duas proteínas. Valores no intervalo entre $0,5 < TM\text{-score} < 1,0$ exibem estruturas com alta similaridade e, portanto, de predição confiável, enquanto que valores de *TM-score* $< 0,5$ exibem proteínas de topologias diferentes (XU; ZHANG, 2010). O objetivo de propor o *TM-score* é resolver o problema de erros locais nos cálculos de RMSD da estrutura. Como o RMSD é uma distância média de todos os pares de resíduos entre duas estruturas, um erro local, por exemplo, causado pela desorientação numa região de alça (variável do ponto de vista conformacional) aumentará, demasiadamente, o valor do RMSD, embora o alinhamento estrutural global entre ambas as proteínas esteja condizente.

No cálculo do *TM-score*, no entanto, as pequenas variações nas distâncias obtidas do alinhamento entre ambas as estruturas serão melhor consideradas que as variações maiores, o que torna a pontuação por esta abordagem insensível ao erro de modelagem local. Adicional ao *TM-score*, o I-TASSER também utiliza uma pontuação de confiabilidade denominada *C-score* que estima a qualidade dos modelos gerados pelo programa. O *C-score* é calculado com base na significância dos alinhamentos dos moldes da estrutura obtidos pela pesquisa de *threading* (estruturas de referência) e nos parâmetros de convergência das simulações de construção da estrutura nativa. Os valores de *C-score* variam num intervalo de -5 a 2, onde uma pontuação maior indica um modelo de maior confiabilidade e os valores menores indicam um modelo de baixa confiabilidade. De modo

geral, a pontuação do *TM-score* está intimamente correlacionada à pontuação do *C-score*, porque indica que quanto maior a similaridade da estrutura alvo com a estrutura (ou estruturas) utilizada(s) como referência(s) (molde), maior será a confiabilidade de predição do modelo final.

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.

