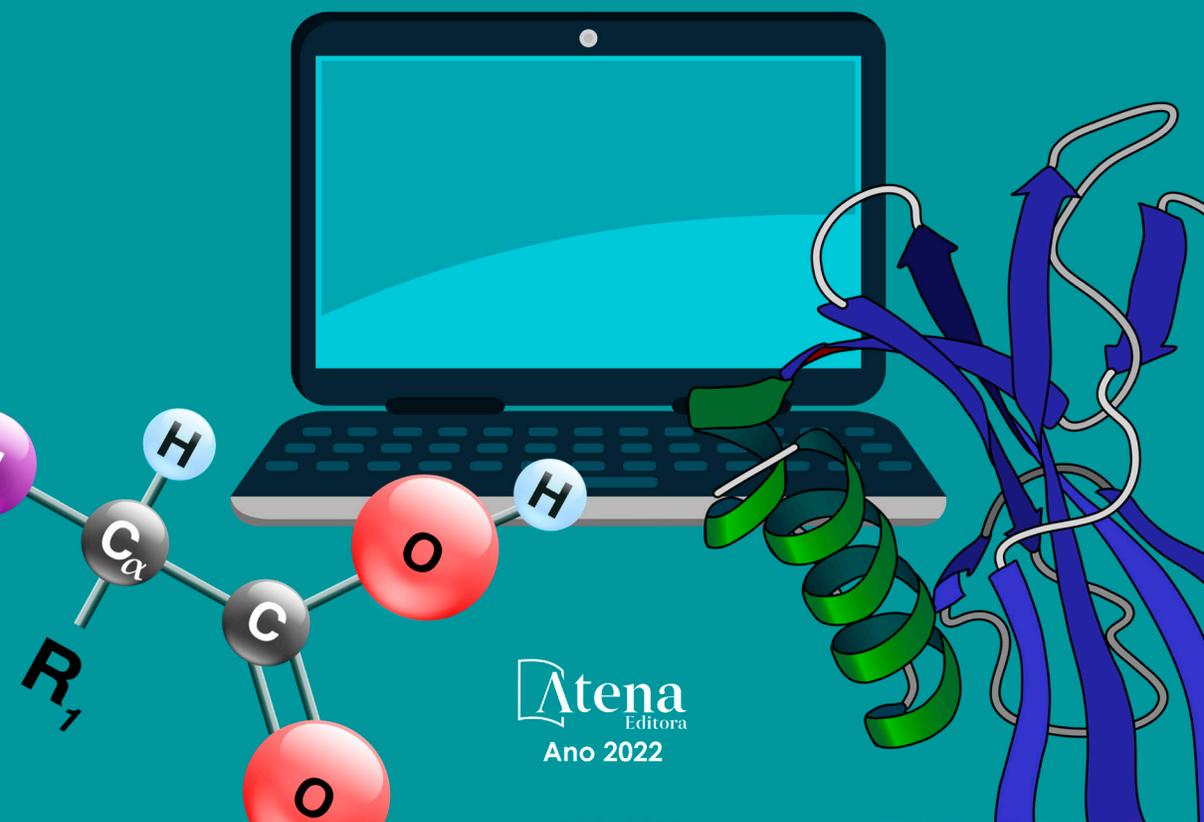


# MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

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<sup>o</sup> Dr<sup>a</sup> Daniela Reis Joaquim de Freitas – Universidade Federal do Piauí  
Prof<sup>o</sup> Dr<sup>a</sup> 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<sup>o</sup> Dr<sup>a</sup> Elizabeth Cordeiro Fernandes – Faculdade Integrada Medicina  
Prof<sup>o</sup> Dr<sup>a</sup> Eleuza Rodrigues Machado – Faculdade Anhanguera de Brasília  
Prof<sup>o</sup> Dr<sup>a</sup> Elane Schwinden Prudêncio – Universidade Federal de Santa Catarina  
Prof<sup>o</sup> Dr<sup>a</sup> Eysler Gonçalves Maia Brasil – Universidade da Integração Internacional da Lusofonia Afro-Brasileira  
Prof. Dr. Ferlando Lima Santos – Universidade Federal do Recôncavo da Bahia  
Prof<sup>o</sup> Dr<sup>a</sup> Fernanda Miguel de Andrade – Universidade Federal de Pernambuco  
Prof. Dr. Fernando Mendes – Instituto Politécnico de Coimbra – Escola Superior de Saúde de Coimbra  
Prof<sup>o</sup> Dr<sup>a</sup> 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<sup>o</sup> Dr<sup>a</sup> 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<sup>o</sup> Dr<sup>a</sup> Juliana Santana de Curcio – Universidade Federal de Goiás  
Prof<sup>o</sup> Dr<sup>a</sup> Lívia do Carmo Silva – Universidade Federal de Goiás  
Prof. Dr. Luís Paulo Souza e Souza – Universidade Federal do Amazonas  
Prof<sup>o</sup> Dr<sup>a</sup> Magnólia de Araújo Campos – Universidade Federal de Campina Grande  
Prof. Dr. Marcus Fernando da Silva Praxedes – Universidade Federal do Recôncavo da Bahia  
Prof<sup>o</sup> Dr<sup>a</sup> Maria Tatiane Gonçalves Sá – Universidade do Estado do Pará  
Prof. Dr. Maurilio Antonio Varavallo – Universidade Federal do Tocantins  
Prof<sup>o</sup> Dr<sup>a</sup> Mylena Andréa Oliveira Torres – Universidade Ceuma  
Prof<sup>o</sup> Dr<sup>a</sup> 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<sup>o</sup> Dr<sup>a</sup> Regiane Luz Carvalho – Centro Universitário das Faculdades Associadas de Ensino  
Prof<sup>o</sup> Dr<sup>a</sup> Renata Mendes de Freitas – Universidade Federal de Juiz de Fora  
Prof<sup>o</sup> Dr<sup>a</sup> Sheyla Mara Silva de Oliveira – Universidade do Estado do Pará  
Prof<sup>o</sup> Dr<sup>a</sup> Suely Lopes de Azevedo – Universidade Federal Fluminense  
Prof<sup>o</sup> Dr<sup>a</sup> Vanessa da Fontoura Custódio Monteiro – Universidade do Vale do Sapucaí  
Prof<sup>o</sup> Dr<sup>a</sup> Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa  
Prof<sup>o</sup> Dr<sup>a</sup> Vanessa Bordin Viera – Universidade Federal de Campina Grande  
Prof<sup>o</sup> Dr<sup>a</sup> 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](http://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**

Anderson Henrique Lima e Lima

Kauê Santana da Costa

Alberto Monteiro dos Santos

### MÉTODOS DE PREDIÇÃO POR *THREADING*

O termo *threading* foi inicialmente cunhado em 1992 por David Jones e colaboradores, para a predição da estrutura de proteínas através do reconhecimento de dobramentos (do inglês *protein folding*) semelhantes com ou sem relação evolutiva (JONES; MILLER; THORNTON, 1995; JONES; TAYLOR; THORNTON, 1992). O *threading* mostra-se útil para a predição de estruturas de proteínas que apresentam baixa identidade (valores  $\leq 30\%$ ) com outras sequências disponíveis em bases de dados para a modelagem comparativa, no entanto, seleciona estruturas que compartilham semelhanças entre si (FLOUDAS et al., 2006).

Esta abordagem baseia-se na conjectura que o número de possíveis dobramentos é finito para todas as proteínas existentes e, desta forma, proteínas não relacionadas também podem assumir estruturas similares devido à evolução convergente (WANG, 1998). Um fator determinante para o desenvolvimento deste método foi a descoberta, através da análise espacial da organização dos aminoácidos nas

proteínas, de preferências na interação entre estes, isto é, resíduos hidrofóbicos, por exemplo, podem ser mais suscetíveis à interação com resíduos de mesma natureza hidrofóbica. Esta observação permitiu o desenvolvimento de potenciais de interação explícita, na qual, analisando a estrutura primária, é possível determinar, através de matrizes de pontuação, diferentes características envolvidas em modelar a estrutura da proteína, tais como preferências de interação de pares de resíduos, regiões favoráveis à formação de estruturas secundárias, regiões acessíveis à solvatação, entre outras.<sup>20</sup> Atualmente, diferentes servidores empregam a modelagem de proteínas por *threading* e apresentam uma interface amigável e de fácil utilização, estes incluem os servidores Phyre2 (KELLEY et al., 2015), I-TASSER (ROY; KUCUKURAL; ZHANG, 2010; ZHANG, 2009) e Raptor-X (KÄLLBERG et al., 2012).

Os métodos *threading* podem ser subdivididos em dois grupos: *threading* baseado em perfil (em inglês *profile-based threading*) e o *threading* baseado em potenciais de pares (em inglês *pair potentials-based threading*) (KORETKE et al., 1999). Na primeira metodologia, os algoritmos criam a estrutura utilizando, como referência, a topologia de estruturas conhecidas, baseando-se em matrizes de pontuação que definem o “perfil” tridimensional da sequência (ROST; SCHNEIDER; SANDER, 1997). Estas

matrizes podem ser derivadas de diferentes informações, entre as quais, matrizes de distância evolutiva (como a de Dayhoff para proteínas), do alinhamento de seqüências e da predição da estrutura secundária (JONES et al., 1999; ZHOU; ZHOU, 2005). O processo de modelagem utilizado por *threading* baseado em perfil é mostrado esquematicamente na Figura 21. O segundo baseia-se em potenciais de pares de resíduos que informam a probabilidade de determinados resíduos encontrarem-se a certa distância, baseando-se em energias de contato de pares de resíduos e energias hidrofóbicas, sendo alguns derivados de análise estatística de estruturas cristalográficas de proteínas (BRYANT; LAWRENCE, 1993; TAYLOR, 1997).

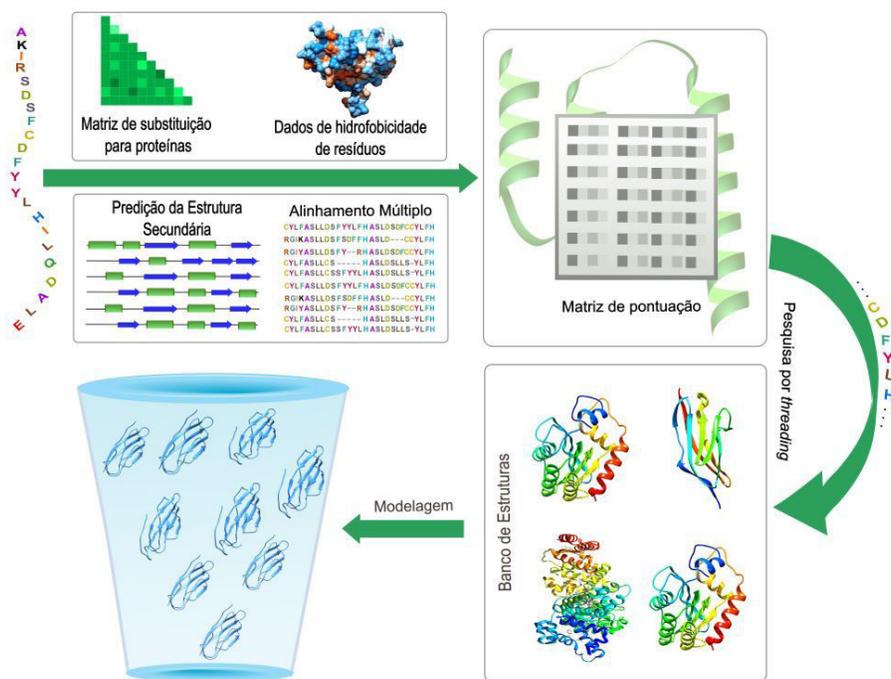


Figura 21. Processo esquemático da modelagem de proteínas pela abordagem *threading* baseado em perfil (em inglês, *profile-based threading*). Etapas de agrupamento estrutural, avaliação e minimização de energia e validação podem estar presentes em determinados algoritmos de modelagem por *threading*

Algoritmos que aplicam a predição por *threading* têm se mostrado bons resultados nos eventos CASP e incluem: THREADER (JONES et al., 1999), Phyre2 (KELLEY et al., 2015) e I-TASSER (ROY; KUCUKURAL; ZHANG, 2010; ZHANG, 2009). A literatura, com modelagem de proteínas por *threading*, é extremamente vasta e utiliza diferentes algoritmos, programas e pipelines (IHM et al., 2009; NOCUA et al., 2014; TSANG et al., 2011). Abordaremos a estratégia de predição do I-TASSER, que recentemente apresentou

resultados extremamente satisfatórios nas últimas competições do CASP (ZHANG, 2007, 2009), em especial, o CASP7 ocorrido em 2006, na cidade de Pacific Grove, na Califórnia, Estados Unidos (TRAPANE; LATTMAN, 2007) e o CASP8 na ilha de Sardenia na Itália (KRYSHTAFOVYCH et al., 2009).

## MÉTODOS DE PREDIÇÃO *AB INITIO*

Os métodos de modelagem *ab initio* são usados para predição da estrutura tridimensional, utilizando como ponto de partida, apenas a estrutura primária da proteína alvo. Neste método, são utilizados uma análise do espaço conformacional da estrutura unida a uma função de pontuação de energia a fim de encontrar as estruturas com o mínimo de energia potencial global. Normalmente, os algoritmos que empregam esta abordagem geram um grande número de modelos teóricos com diferentes conformações estruturais e um dos desafios é selecionar o modelo final (GOPAL; KLENIN; WENZEL, 2009).

Esta abordagem se destaca por não depender de estruturas homólogas previamente conhecidas que possam ser utilizadas como referência na modelagem, e emerge do fato de que muitas proteínas ainda não apresentam identidade suficiente ( $\geq 30\%$ ) com seqüências de estruturas depositadas em bases de dados para a utilização de métodos de modelagem comparativa. Os métodos *ab initio* podem ser subdivididos em duas categorias: *ab initio* com informações de base de dados também chamados de baseados em conhecimento (em inglês *knowledge-based*), e *ab initio* sem informações de base de dados (FLOUDAS et al., 2006).

Os métodos *ab initio* baseados em conhecimento não comparam à proteína alvo com a proteína molde, mas sim curtos fragmentos entre ambas, e uma vez que os fragmentos compatíveis são identificados. Estes algoritmos iniciam a montagem dos mesmos para a formação da estrutura-alvo com o auxílio de funções de pontuação de energia e algoritmos de pesquisa conformacional. Os métodos *ab initio* sem informação de base de dados consideram que a conformação de uma proteína depende unicamente da sua estrutura primária. Estes métodos utilizam potenciais físicos para a resolução da estrutura e selecionam aquelas que melhor representam o estado nativo, por meio de algoritmos que empregam pesquisa conformacional baseada nos potenciais. Com relação aos pacotes de programas que podem realizar predição *ab initio* sem informações de base de dados, podemos citar os aplicados à simulação de dinâmica molecular, tais como CHARMM (BROOKS et al., 1983) (*Chemistry at Harvard Molecular Mechanics*), AMBER (*Assisted Model Building with Energy Refinement*), (CASE et al., 2005; SALOMON-FERRER; CASE; WALKER, 2013) e GROMACS (*GRoningen MACHine for Chemical Simulation*) (VAN DER SPOEL et al., 2005). É importante ressaltar que embora estes pacotes tenham sido desenvolvidos para simulação de dinâmica molecular, atualmente, eles podem ser utilizados na elucidação estrutural e para análise do enovelamento de proteínas.

Os métodos *ab initio* consideram que uma proteína pode adquirir diferentes conformações, porém a estrutura de menor energia global corresponde ao seu estado nativo (KUHLMAN; BAKER, 2000). Deste modo, os algoritmos de modelagem, normalmente, geram uma grande quantidade de modelos teóricos com o propósito de se obter diferentes conformações e empregam um tratamento estatístico para selecionar as estruturas que melhor representam a estrutura nativa em relação às demais que são geradas pelo processo de modelagem. Desta forma, é comum programas aplicarem um agrupamento (em inglês *clustering*) das estruturas similares, além de uma análise baseada em centroide, isto é, a seleção das conformações que melhor representam as demais do agrupamento (em inglês *cluster*).

Os algoritmos de agrupamento utilizam um critério de similaridade estrutural, como a Raiz do Desvio Quadrático Médio (RMSD do inglês *Root Mean Square Deviation*) do carbono alfa (C $\alpha$ ), para a formação dos agrupamentos. Assim, é possível reduzir a quantidade de modelos teóricos gerados e analisar somente as estruturas mais significativas (ROHL et al., 2004; SHORTLE; SIMONS; BAKER, 1998; ZHANG; KOLINSKI; SKOLNICK, 2003; ZHANG; SKOLNICK, 2004). No geral, programas que realizam predição *ab initio* disponibilizam sua ferramenta de agrupamento, contudo, pacotes externos já foram desenvolvidos para esta finalidade como o MaxSub (SIEW et al., 2000).

A figura 22 representa esquematicamente o processo de agrupamento utilizado por programas *ab initio*. Além destes, são realizadas otimizações nas estruturas proteicas, a fim de melhorar a qualidade estereoquímica e o perfil de energia dos modelos obtidos. Devido suas peculiaridades, normalmente, os métodos *ab initio* requerem demasiado tempo de processamento computacional quando comparados aos outros métodos de predição *in silico*.

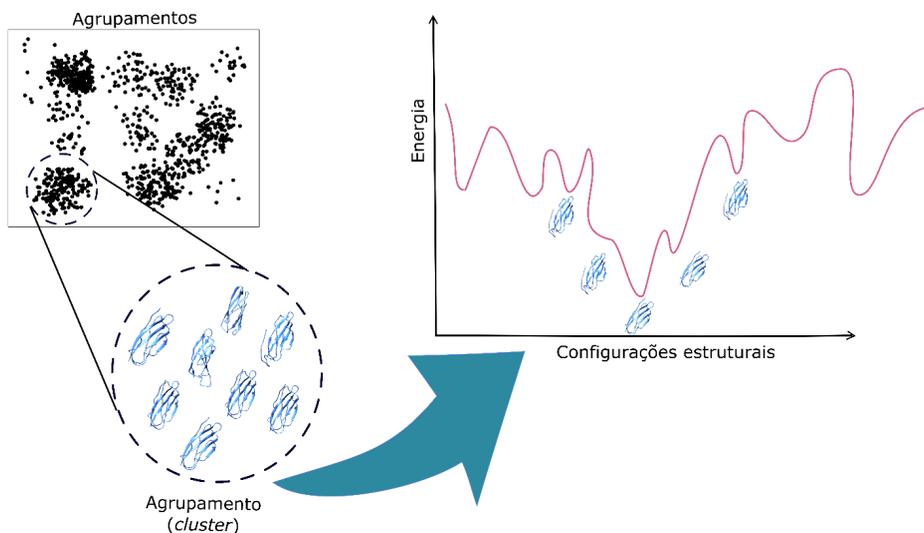


Figura 22. Visão geral do processo de agrupamento utilizado pelos métodos *ab initio*. Modelos de baixa energia global são estruturalmente semelhantes à forma nativa (KUHLMAN; BAKER, 2000).

O sucesso da predição *ab initio* depende de três componentes essenciais para seu funcionamento: (1) uma função de pontuação de energia capaz de discriminar corretamente a forma nativa das demais (denominadas *decoy*); (2) uma representação geométrica da proteína e (3) uma técnica de pesquisa conformacional capaz de discriminar, rapidamente, as formas de baixa energia que representam um estado termodinamicamente estável das outras formas, isto, realizar uma varredura na superfície de energia obtida pela geração das estruturas propostas. Atualmente, há diferentes programas que realizam a modelagem *ab initio* e aplicam diferentes estratégias de pesquisa conformacional e campos de força (KIHARA et al., 2001; KLEPEIS; FLOUDAS, 2003; SIMONS et al., 1999; XU; ZHANG, 2012; ZHANG; KOLINSKI; SKOLNICK, 2003). A tabela 1 sumariza os principais programas. Abordaremos, nas seções seguintes, separadamente cada um dos componentes da predição *ab initio*.

Algoritmo	Campo de Força	Seleção do modelo	Pesquisa Conformacional	Referências
<b>ROSETTA</b>	Físico e baseado em conhecimento	Agrupamento e menor energia potencial	Monte Carlo	(SIMONS et al., 1997)
<b>UNRES</b>	Físico	Agrupamento/ menor energia livre	<i>Conformational space annealing</i>	(LIWO; KHALILI; SCHERAGA, 2005)
<b>I-TASSER</b>	Baseado em conhecimento	Agrupamento e menor energia livre	Monte Carlo	(ZHANG, 2008, 2009)
<b>ASTRO-FOLD</b>	Físico	Menor energia	<i>Conformational space annealing</i> e dinâmica molecular	(KLEPEIS; FLOUDAS, 2003)
<b>QUARK</b>	Baseado em conhecimento	Agrupamento e menor energia livre	Monte Carlo	(XU; ZHANG, 2012)
<b>TOUCHSTONE</b>	Baseado em conhecimento	Agrupamento e menor energia livre	Monte Carlo	(KIHARA et al., 2001; ZHANG; KOLINSKI; SKOLNICK, 2003)
<b>AMBER e CHARMM</b>	Físico	Menor energia livre	Dinâmica molecular	(BROOKS et al., 1983; CASE et al., 2005)

Tabela 1. Resumo dos principais algoritmos utilizados para predição *ab initio* da estrutura de proteínas discriminados pelo tipo de campo de força, método de seleção do modelo, e o algoritmo de pesquisa conformacional. É importante ressaltar que pacotes de simulação de DM também podem ser usados na predição *ab initio* de proteínas.

## REPRESENTAÇÃO GEOMÉTRICA

De modo geral, uma representação geométrica, que leve em consideração todos os átomos da cadeia polipeptídica, demandaria enorme tempo computacional devido ao elevado número de partículas e ângulos de torção necessários para a representação das estruturas. Dessa maneira, uma representação geométrica de proteínas devidamente adotada pode mostrar-se extremamente eficiente para os cálculos de energia potenciais do sistema analisado, sem comprometer o tempo de cálculo computacional.

Alguns algoritmos de predição *ab initio* diminuem os graus de liberdade através de modelos semirreduzidos, considerando somente os átomos pesados da cadeia

polipeptídica (ex.: N, C $\alpha$ , C, O, C $\beta$ ) para representar a cadeia principal e o centro de massa das cadeias laterais (XU; ZHANG, 2012). O algoritmo Rosetta utiliza os átomos pesados da cadeia principal e representa a cadeia lateral pelo C $\beta$  mantendo os ângulos e distâncias de ligação constantes (SIMONS et al., 1997). O algoritmo Touchstone representa a estrutura da cadeia principal pelos átomos de carbono alfa (C $\alpha$ ) com ligações laterais de átomos de carbono beta (C $\beta$ ) e centros de massa da cadeia lateral (KIHARA et al., 2001).

## MÉTODOS DE PESQUISA CONFORMACIONAL

Há três principais métodos utilizados para a pesquisa conformacional da estrutura de proteínas: dinâmica molecular (DM), o método estocástico Metrópolis Monte Carlo (MC) (METROPOLIS et al., 1953) e os algoritmos genéticos. Os métodos de pesquisa conformacional analisam diferentes conformações possíveis que uma proteína pode adquirir de modo a encontrar aquela de menor energia e são cruciais, quando se deseja atingir modelos de ótima resolução atômica e conformação semelhante à forma nativa. É, por este motivo, novos métodos de pesquisa têm sido conduzidos ao longo dos anos, e tem se mostrado um campo ativo de estudo (DING et al., 2008; LIWO et al., 1998; THACHUK; SHMYGELSKA; HOOS, 2007; ZHANG, 1999).

Diferentes algoritmos Metrópolis MC foram desenvolvidos (ZHANG, 1999) ultimately, refine protein structures. The main principle of this algorithm is, in each cycle, a new trial conformation is generated by carrying out a short period of molecular dynamics (MD e existem boas revisões sobre o assunto (HANSMANN; OKAMOTO, 1999). Provavelmente, a versão mais popular de algoritmo MC é o *Simulated Annealing* (SA) (KIRKPATRICK; GELATT; VECCHI, 1983) aplicado com sucesso em programas de predição *ab initio*, em especial, o algoritmo Rosetta (KAUFMANN et al., 2010; ROY; KUCUKURAL; ZHANG, 2010).

O SA é um algoritmo de otimização global que aplica o Metrópolis MC para gerar as conformações da proteína, seguindo a distribuição canônica de energia de Boltzman para determinada temperatura. Este algoritmo inicia a geração de perturbações aleatórias nas conformações, através do aumento da temperatura seguida das diminuições progressivas, e mantém as melhores soluções conformacionais baseadas no mínimo de energia da estrutura. Em resumo, objetiva encontrar as combinações de rotâmeros que ocupam o mínimo de energia global, de acordo com uma função de pontuação de energia (OKAMOTO, 2009). Os algoritmos genéticos são representados principalmente pela técnica de *Conformational space annealing* (CSA) implementados no ASTRO-FOLD (KLEPEIS; FLOUDAS, 2003) e UNRES (LIWO et al., 1998).

## CAMPOS DE FORÇA E FUNÇÃO DE PONTUAÇÃO DE ENERGIA

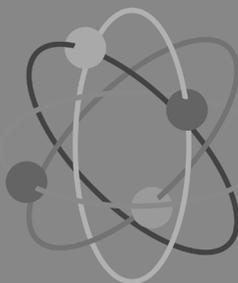
Os campos de força são um conjunto de parâmetros físico-químicos e de funções matemáticas utilizadas para descrever a energia potencial de um sistema, estes contêm informações utilizadas pela função de pontuação de energia (do inglês *energy score function*) que irão determinar se a estrutura do modelo final obtida se mostra fidedigna para representar a conformação nativa da proteína. Uma função de pontuação de energia (também referida por função de mérito) deve, portanto, auxiliar o método de pesquisa conformacional a discriminar corretamente as estruturas semelhantes à forma nativa daquelas não enoveladas ou mal enoveladas geradas no processo de modelagem, sendo desta forma, úteis para avaliar a qualidade dos modelos gerados. Baseando-se na utilização ou não de informações estatisticamente extraídas de estruturas resolvidas experimentalmente, os campos de força são divididos em duas categorias: baseados em conhecimento e campos de força físicos. Atualmente, os algoritmos *ab initio* aplicam um (KLEPEIS; FLOUDAS, 2003) ou ambos na predição da estrutura (RAMAN et al., 2009; ROHL et al., 2004).

Métodos de predição, que empregam campos de força físicos aliado aos métodos de pesquisa conformacional, podem ser considerados estritamente *ab initio*, por não utilizarem nenhuma informação proveniente de estruturas proteicas experimentalmente resolvidas. A utilização destes campos de força de potenciais físicos surge da hipótese termodinâmica de Anfinsen, pois tentam resolvê-la baseando-se na obtenção da sua menor energia livre global (PILLARDY et al., 2001). Com relação aos campos de força físico, podemos citar os já bem conhecidos OPLS (JORGENSEN; MAXWELL; TIRADO-RIVES, 1996), AMBER, (CASE et al., 2005) e CHARMM (BROOKS et al., 1983). Quando associados aos modelos de solvatação e à simulação de dinâmica molecular, estes campos de força podem se tornar eficientes não somente na elucidação da estrutura tridimensional, mas também no estudo do enovelamento de proteínas (VOELZ et al., 2010).

Com relação aos campos de força baseados em conhecimento, eles contêm termos de energia empíricos derivados de estruturas de proteínas resolvidas experimentalmente (SIPPL, 1990). Em alguns casos, estes campos abrangem potenciais de contato de pares de resíduos e propensões da sequência em desenvolver determinada estrutura secundária (HAO; SCHERAGAT, 1999; SKOLNICK, 2006). O algoritmo Rosetta, por exemplo, (SIMONS et al., 1999) aplica um campo de força baseado em conhecimento unido à montagem de fragmentos de proteínas de estrutura conhecida e, no final, depois de obtidos os modelos, estes são submetidos ao refinamento atômico através de um potencial físico (RAMAN et al., 2009).

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.

