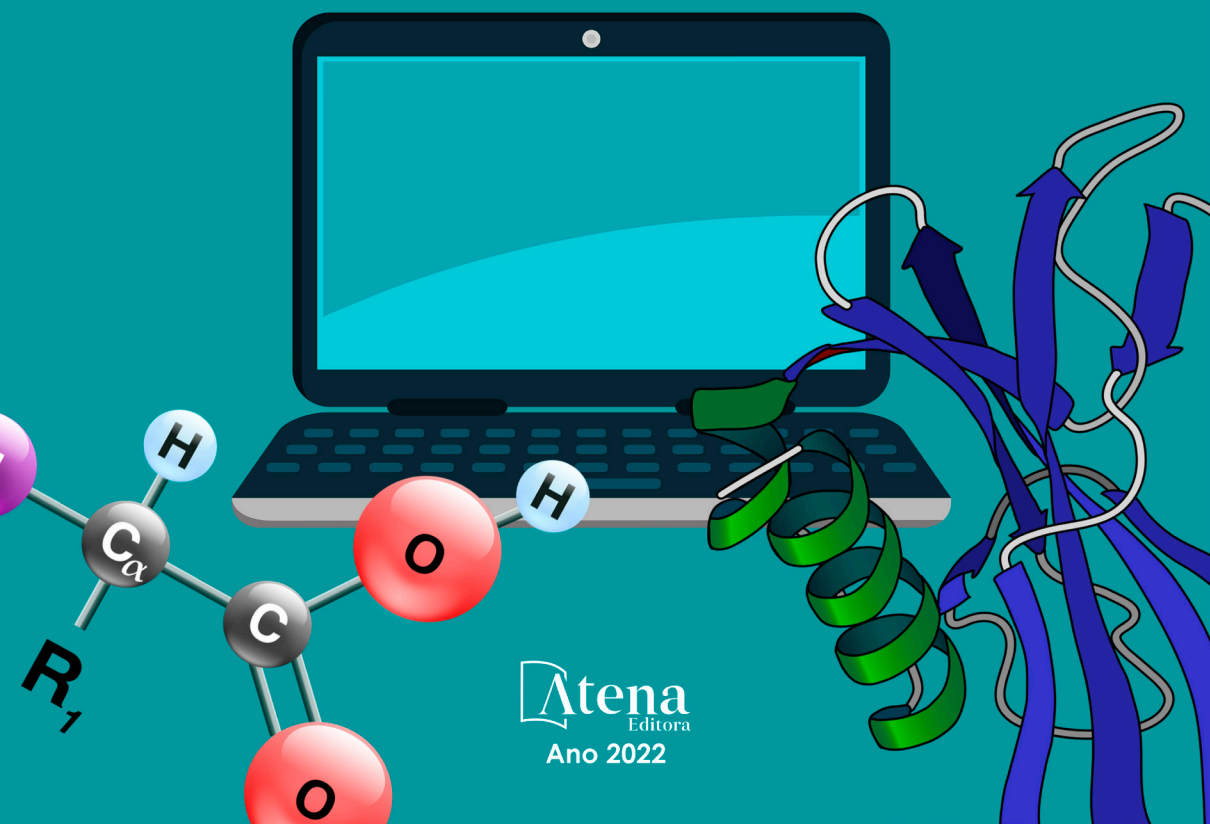


MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

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 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

CAPÍTULO 2

MÉTODOS DE PREDIÇÃO POR HOMOLOGIA

Kauê Santana da Costa

João Marcos Pereira Galúcio

José Rogério de Araújo Silva

A predição por homologia, também conhecida por modelagem comparativa, visa determinar a estrutura da proteína de interesse, utilizando como molde (referência), estruturas de proteínas homólogas previamente elucidadas por métodos experimentais, tais como espectroscopia de Ressonância Magnética Nuclear (RMN) e difração de raio X. Atualmente, é considerada como um dos métodos mais fidedignos de predição da estrutura nativa de proteínas, porque utiliza, como referência, uma estrutura similar, sendo por isso, o método computacional preferido objetiva obter estruturas com maior precisão estrutural, bem como estruturas experimentais estão ausentes em estudos de planejamento *in silico* de fármacos e no estudo do mecanismo catalítico de enzimas (CAVASOTTO; PHATAK, 2009).

O termo homologia é utilizado na Biologia Evolutiva para se referir a duas estruturas que compartilham semelhanças por descenderem de um ancestral comum. Desta forma, a predição por homologia se baseia no pressuposto de que a estrutura terciária é mais

conservada que a estrutura primária (sequência) devido à conservação da função molecular ao longo da evolução. Nesta linha de raciocínio, considerando que duas proteínas pertencentes a mesma família apresentam estrutura mais preservada que a sequência, podemos assumir que caso esta similaridade seja previamente detectada no nível de sequência, então a estrutura tridimensional terá similaridade maior. De modo semelhante, caso a similaridade entre as duas sequências seja pequena, então, a correspondência entre a estrutura de ambas também será baixa.

CONCEITUANDO IDENTIDADE E SIMILARIDADE

Na comparação entre a sequências de aminoácidos de duas proteínas, dois parâmetros são fundamentais: similaridade e identidade. Ambos são obtidos através do alinhamento das sequências. A similaridade corresponde ao percentual de aminoácidos que compartilham entre si as mesmas características físico-químicas, por exemplo, leucina, valina e isoleucina são ambos aminoácidos hidrofóbicos apolares e, portanto, nas matrizes de alinhamento aplicado para sequências de proteínas apresentarão pontuações similares quando ocuparem as mesmas regiões de correspondência, conhecidas como *matches*. Já a identidade corresponde ao percentual de aminoácidos idênticos entre as duas sequências

analisadas (SÁNCHEZ; ŠALI, 1997).

Ambos os conceitos podem ser melhor expressos por meio de equações. A similaridade (S) é expressa pela Equação 1, conforme abaixo:

$$S = \frac{Ls \times 2}{La + Lb} \times 100 \quad (1)$$

Onde:

S = similaridade

Ls = número de aminoácidos similares da sequência.

La e Lb = comprimento da sequência A e B, respectivamente.

A identidade (I) é representada por equação similar, porém o denominador Li corresponde ao número de aminoácidos idênticos da sequência (Equação 2).

$$I = \frac{Li \times 2}{La + Lb} \times 100 \quad (2)$$

Onde:

I = identidade

Li = número de aminoácidos idênticos da sequência.

La e Lb = comprimento da sequência A e B, respectivamente.

Diferentes programas e alinhamento calculam os valores de identidade para sequências de aminoácidos. Entre estes pode-se citar o Muscle (EDGAR, 2004) e o T-Coffee (DI TOMMASO et al., 2011).

ETAPAS ENVOLVIDAS DA MODELAGEM POR HOMOLOGIA

A modelagem por homologia envolve cinco passos consecutivos e complementares que incluem: (1) pesquisa de estruturas homólogas em base de dados; (2) alinhamento da sequência alvo com as sequências de estrutura conhecida (moldes); (3) construção do modelo tridimensional; (4) a otimização estrutural, e por último; (5) a validação do modelo. Uma visão geral do processo de modelagem por homologia é mostrada na figura 3.

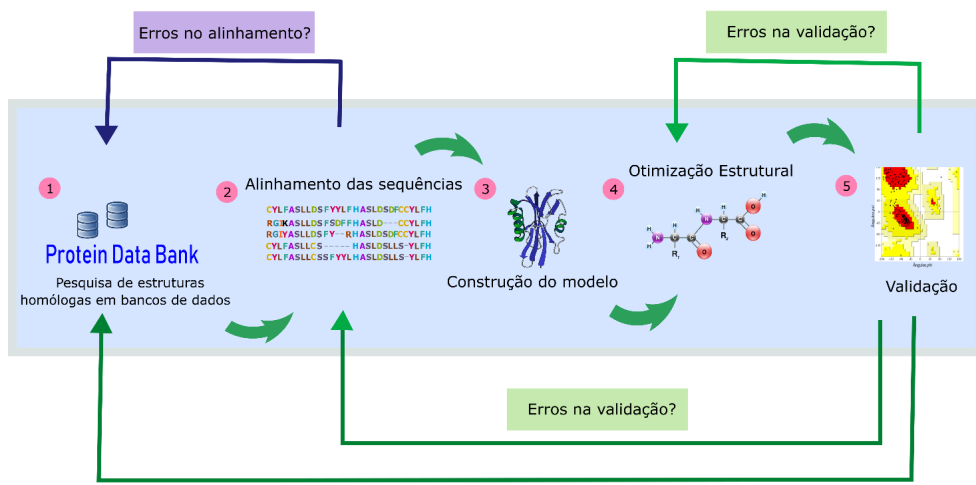


Figura 3. Visão geral do processo de modelagem por homologia. Erros em uma das etapas podem acarretar em inconsistências nos resultados da etapa subsequente, sendo, deste modo, necessário o retorno à uma etapa anterior para a correção.

BUSCA DE ESTRUTURAS HOMÓLOGAS

No primeiro passo, a seqüência alvo (de interesse) é comparada por meio de alinhamento com outras existentes em base de dados de estruturas experimentais. Atualmente, o RCSB *Protein Data Bank* (<https://www.rcsb.org/>) é o maior e mais abrangente base de dados público com informações estruturais de proteínas, ácidos nucleicos e seus ligantes (BERMAN et al., 2005). Para realizar a pesquisa de estruturas no PDB, pode-se utilizar o *Protein BLAST* (BLASTp sigla do inglês *Basic Local Alignment Search Tool*), um algoritmo de alinhamento simples e local (ALTSCHUL et al., 1997; BENSON et al., 2017) que realiza a busca sistemática em base de dados de seqüências. O BLASTp é disponibilizado para uso *on-line* pelo *National Center for Biotechnology Information* (NCBI) e fornece valores de cobertura, identidade e código de acessos de seqüências que melhor se alinham. A identidade corresponde ao porcentual de resíduos similares entre duas seqüências, sendo elas um dos parâmetros importantes para a seleção de estruturas moldes que serão utilizados na modelagem. Além de identidade satisfatória, devemos avaliar diferentes critérios em conjunto e, de modo geral, as estruturas devem:

1. Pertencer aos organismos filogeneticamente próximos.
2. Pertencer às proteínas que apresentam a mesma função molecular ou mesma família.
3. Se obtidas por meio da técnica de difração de raio X, devem apresentar alta qualidade (resolução baixa, preferencialmente $\leq 2,0 \text{ \AA}$). Valores maiores poderão ser permitidos na ausência de outras estruturas homólogas, porém devido à alta resolução, o modelo poderá apresentar qualidade inferior e, portanto, mais etapas

de otimização estrutural serão necessárias.

4. Estar o mais completas possível, isto é, com poucos ou nenhum fragmento ausente.
5. Apresentar cobertura satisfatória com a sequência alvo ($\geq 90\%$).
6. Apresentar identidade satisfatória no alinhamento ($\geq 30\%$), isto é, quanto maior a semelhança na sequência, maior será a semelhança na estrutura. Não havendo identidade mínima (30%) não é recomendável aplicar a modelagem por homologia.

ALINHAMENTO DAS SEQUÊNCIA ALVO E MOLDE

É realizado, na segunda etapa da modelagem por homologia, o alinhamento entre as sequências. Esta etapa é essencial no processo de modelagem, haja vista que permite a extração do molde informações estruturais que serão utilizadas na criação do modelo tridimensional. Estruturas podem ser modeladas, partindo de um alinhamento simples, isto é, alinhamento em que são consideradas somente duas sequências (sequência alvo e a sequência molde) ou de alinhamento múltiplo, em que mais de duas sequências são comparadas (a sequência alvo, dois ou mais moldes). O alinhamento múltiplo é indicado quando determinadas regiões da proteína-alvo carecem de uma estrutura de referência no molde, sendo este obtido, portanto, de outras proteínas homólogas. Diferentes programas podem ser utilizados no alinhamento das sequências, tais como o ClustalW (CHENNA, 2003), Muscle (EDGAR, 2004) e T-Coffee (DI TOMMASO et al., 2011). O Modeller, por exemplo, utiliza a matriz de similaridade BLOSUM62.

Uma alta identidade entre a sequência da estrutura molde com a sequência de interesse (alvo) é, preferencialmente, desejada, pois maior será a qualidade e, portanto, a acurácia do modelo gerado. Desta forma, modelos de alta identidade (de 95% a 100%), após validados, podem ser satisfatoriamente usados no estudo de mecanismos catalíticos de enzimas, na triagem virtual de fármacos e no planejamento *in silico* de inibidores (WEBB; SALI, 2016).

CONSTRUÇÃO DO MODELO

Na terceira etapa, realizamos a predição da estrutura propriamente, isto é, a construção do modelo tridimensional. Atualmente, vários programas e servidores aplicam abordagem por homologia, entre eles, podemos citar o servidor Swiss-model (BIASINI et al., 2014) e o pacote computacional Modeller (FISER; ŠALI, 2003) que aplicam estratégias diferentes na modelagem comparativa.

O Modeller é um programa desenvolvido em Fortran 90 e Python pelo grupo do pesquisador Andrej Sali e realiza a predição da estrutura de proteínas por homologia, implementando uma técnica inspirada por ressonância magnética nuclear conhecida como

satisfação das restrições espaciais, pela qual um conjunto de ligações angulares (entre três átomos) e diedrais (entre quatro átomos) são retiradas de uma estrutura molde (usada como referência) obtida experimentalmente e, então, aplicadas na estrutura alvo que será modelada (FISER; ŠALI, 2003). Além destas restrições, o campo de força CHARMM-22 fornece informações sobre os comprimentos e ângulos de ligação dos átomos da estrutura modelada, incorporando uma funcionalidade limitada para previsão *ab initio* das regiões de alças de proteínas, que são frequentemente muito variáveis e, portanto, difíceis de prever por modelagem por homologia (FISER; ŠALI, 2003). Como parâmetro de seleção das estruturas otimizadas, o Modeller aplica o conceito de Energia Descontínua de Proteína Optimizada (Dope, do inglês *Discrete Optimized Protein Energy*), um potencial estatístico obtido da minimização de funções de energia da estrutura que permite selecionar entre os modelos criados o de menor energia. No Modeller, a energia Dope é aplicada por meio de scripts em linguagem *Python*. Além da modelagem, o programa também realiza funções adicionais úteis na modelagem comparativa, tais como, o alinhamento simples (entre duas sequências) ou múltiplo (mais de duas sequências) e a otimização de regiões de alça da estrutura modelada (FISER; ŠALI, 2003).

O Swiss-model é um servidor que realiza a predição e a validação da estrutura de proteínas, além de apresentar uma interface intuitiva e fácil de usar. O servidor aplica a modelagem por homologia utilizando o conceito de corpos rígidos (BIASINI et al., 2014; KIEFER et al., 2009) que, por sua vez, consistem na modelagem da estrutura por partes, sendo que as regiões estruturalmente conservadas da proteína alvo são definidas através de predição das estruturas secundárias. Essas regiões são alinhadas com o molde, levando em conta a média das posições dos carbonos α das regiões estruturalmente conservadas da sequência.

As regiões que não satisfazem as exigências estruturais, geralmente, são porções de alças que, por apresentarem estruturas flexíveis, adquirem diferentes conformações. Informações para a modelagem da cadeia principal, dessas regiões variáveis, são obtidas em bases de dados de estruturas que apresentam conjuntos de alças classificados pelo número de resíduos de aminoácidos e pelo tipo de estruturas secundárias que estas conectam. Após a modelagem das regiões de alças, um modelo inicial do esqueleto peptídico é gerado e, em seguida, procede-se a inserção das cadeias laterais dos resíduos através de busca em bibliotecas de rotâmeros (BIASINI et al., 2014).

OTIMIZAÇÃO ESTRUTURAL

A quarta etapa consiste na otimização estrutural, isto é, o refinamento da estrutura tridimensional. Invariavelmente, após a construção do modelo, há erros esteroquímicos que necessitam de correção, pois influenciam na estabilidade da estrutura e na interpretação de dados estruturais que podem ser obtidos. A otimização compreende a alteração

guiada da estrutura de modo a melhorar sua qualidade estereoquímica e perfil de energia. Além de métodos de minimização de energia, são aplicados métodos de amostragem conformacional que empregam simulações de dinâmica molecular clássica com campos de força específicos (FAN, 2004; RAVAL et al., 2012). Diferentes algoritmos de otimização foram desenvolvidos com esta finalidade, porém uma discussão detalhada está fora do escopo deste trabalho.

VALIDAÇÃO DO MODELO

Após a etapa de otimização estrutural, a quinta e última etapa corresponde da modelagem por homologia corresponde à validação e permite avaliar o nível de qualidade e confiabilidade da estrutura obtida na modelagem. Modelos tridimensionais de alta qualidade, depois de validados, podem revelar dados biológicos relevantes para avaliar a atividade e a função da proteína, bem como realizar estudos sobre sua estrutura. Ao longo dos anos, diferentes ferramentas foram desenvolvidas e auxiliam na validação do modelo fornecendo parâmetros sobre a qualidade da estrutura.

Erros encontrados nas etapas de modelagem por homologia podem, por exemplo, serem devidos:

1. Aos alinhamentos errôneos das sequências, como a existência de muitos *mismatches*, baixa cobertura e identidade.
2. À seleção de estruturas-molde de baixa qualidade ou sem correspondência estrutural (ex.: estruturas-molde resolvidas por cristalografia de raios-X a alta resolução criará modelos de baixa qualidade);
3. À modelagem de regiões sem a estrutura-molde que levam invariavelmente à formação de regiões de alça.

A validação utiliza três principais critérios: (1) correlação das estruturas secundárias e terciárias da proteína de modo a inferir a conservação da sua função; (2) estereoquímica e (3) a energia (WEBB; SALI, 2016).

Na avaliação da correlação da estrutura secundária e terciária da proteína são aplicados métodos de alinhamento estrutural e métricas de comparação, como o desvio da raiz quadrada média das posições atômicas (RMSD do inglês *Root Mean Square Deviation*) calculado através da distância média entre os átomos de duas estruturas sobrepostas. A maioria dos programas utiliza o carbono α (C α), como referência para o cálculo do RMSD, porém outros átomos da cadeia principal podem também ser utilizados como nitrogênio e oxigênio. A equação utilizada para o cálculo do RMSD é mostrada a seguir (Equação 3):

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\|x_i^A - x_i^B\|^2)} \quad (3)$$

Onde:

N = número de átomos comparados.

x_i^A e x_i^B = coordenadas espaciais dos átomos representativos das estruturas A e B, respectivamente.

Para cada valor de RMSD-C α (medido em angstroms, Å), é possível estabelecer um comparativo de quão semelhantes são as duas estruturas em questão – estrutura-molde (do inglês *template*) e a estrutura-alvo modelada (do inglês, *target*) (Tabela 1). Ao se comparar a estrutura molde com a estrutura alvo, os valores de RMSD-C α desejáveis devem ser preferencialmente $\leq 3,0$ Å. Valores maiores que 3,0 Å trazem dúvidas com relação à escolha da estrutura-molde e a qualidade do modelo final. Já estruturas com RMSD $\leq 0,4$ Å mostram que ambas são estruturalmente indistinguíveis, podendo indicar que são conformações da mesma proteína. Através do alinhamento estrutural, é possível analisar a disposição e a conservação de resíduos em sítios catalíticos, motivos proteicos, inferir a geometria de coordenação dos resíduos com íons metálicos, assim como prever a função da proteína, comparando-a com outras estruturas previamente elucidadas por métodos experimentais (BAXEVANIS; OUELLETTE, 2003).

RMSD (Å)	Comparação Estrutural
>12,0	Completamente não relacionadas
7,0	Não relacionadas
5,0	Podem estar estruturalmente relacionadas
3,0	Boa relação estrutural
2,0	Proximamente relacionadas
1,5	Muito proximamente relacionadas
0,8	Muito similares
$\leq 0,4$	Estruturalmente indistinguíveis

Tabela 1. Valores de RMSD-C α e a respectiva interpretação com relação à semelhança estrutural das proteínas.

Fonte: Adaptado de Baxevanis e Ouellette (2003).

Na validação, a avaliação da qualidade estereoquímica do modelo pode ser realizada por diferentes parâmetros estruturais. O gráfico de Ramachandran é, sem dúvida, a ferramenta de avaliação mais popular para este fim, pois permite a representação da proteína por meio de seus ângulos de torção (LOVELL et al., 2003). O gráfico de Ramachandran descreve as proteínas pelos seus ângulos de torção, conhecidos por ϕ

(pronúncia: *phi*) e ψ (*psi*) (Figura 4). O ângulo ϕ é definido com o ângulo diedro presente na ligação N-C α , enquanto que o ângulo ψ é definido como o ângulo diedro presente na ligação C α -C. Cada resíduo, portanto, pode ser definido por um ângulo ϕ e um ψ , quando submetidos num gráfico de dispersão com os ângulos ϕ no eixo X e os ângulos ψ no eixo Y.

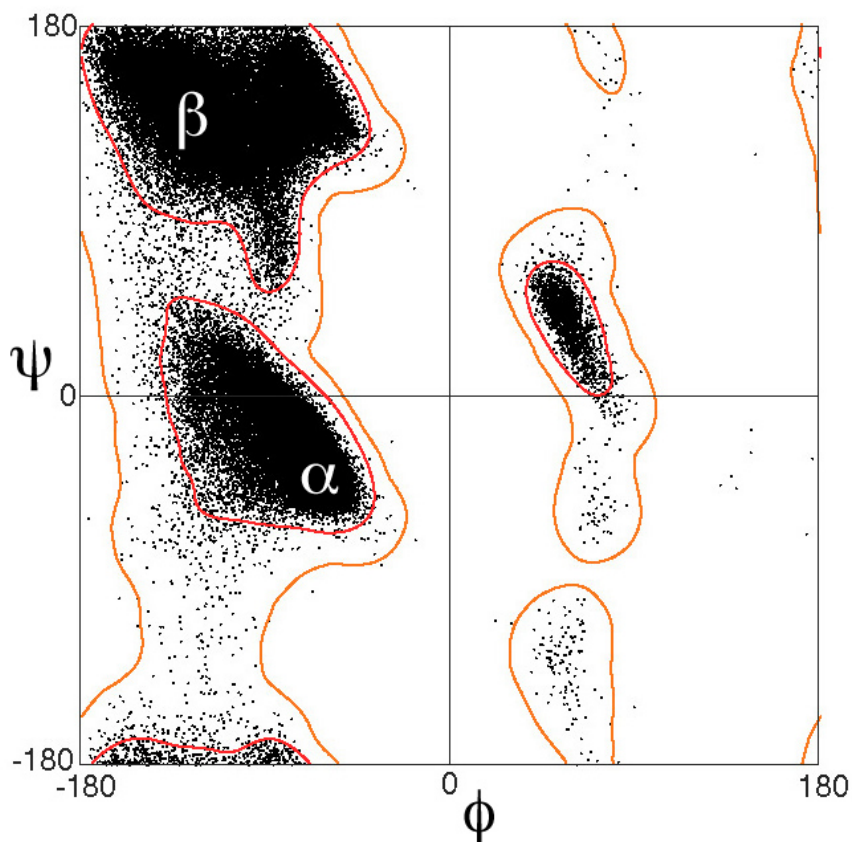


Figura 4. Gráfico de Ramachandran que mostra as regiões aceitáveis para os ângulos de torção ϕ e ψ localizados nas estruturas secundárias (α -hélices e β -folhas). Cada ponto do gráfico representa um resíduo de aminoácido da proteína analisada.

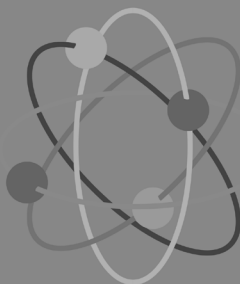
A representação de Ramachandran permite verificar os valores permitidos dos ângulos de torção, indicando quais resíduos se encontram em regiões mais favoráveis ou desfavoráveis dos ângulos, permitindo, assim, prever conflitos estéricos entre as cadeias (LOVELL et al., 2003). As regiões permitidas para cada aminoácido dependem da extensão da cadeia lateral, sendo assim, para a glicina que apresenta menor cadeia lateral são aplicadas menores restrições, visto que ela é formada por um único átomo de hidrogênio, assim como, para a alanina em que a cadeia lateral é formada por um grupo metila. Em

cadeias laterais que envolvem vários átomos, a mobilidade dela é restringida pelo tamanho. As regiões em cinza exibem os ângulos de torção permitidos para as β -folhas (indicados pela letra grega β) e α -hélices (regiões indicadas por α), as regiões demarcadas por linhas tracejadas exibem regiões de repulsão e choque entre diferentes átomos e representam, portanto, resíduos com valores errôneos para ângulos de torção. Entre os programas que geram o gráfico de Ramachandran, podem ser citados o Rampage, o Procheck (LASKOWSKI et al., 1993) e o servidor MolProbity (LOVELL et al., 2003).

Na avaliação da energia são aplicados métodos que aplicam valores normalizados obtidos da minimização de energia. Ao avaliar a energia, objetiva-se verificar a estabilidade da estrutura, isto é, sua propensão de formar interações intra e intermoleculares favoráveis que as tornem estáveis, assim como as conformações com enovelamento adequado. Entre os programas aplicados na avaliação da energia de modelos se destacam o Qmean (BENKERT; SCHWEDE; TOSATTO, 2009), ANOLEA (MELO; FEYTMANS, 1998) e o servidor ProSA (WIEDERSTEIN; SIPPL, 2007).

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.

