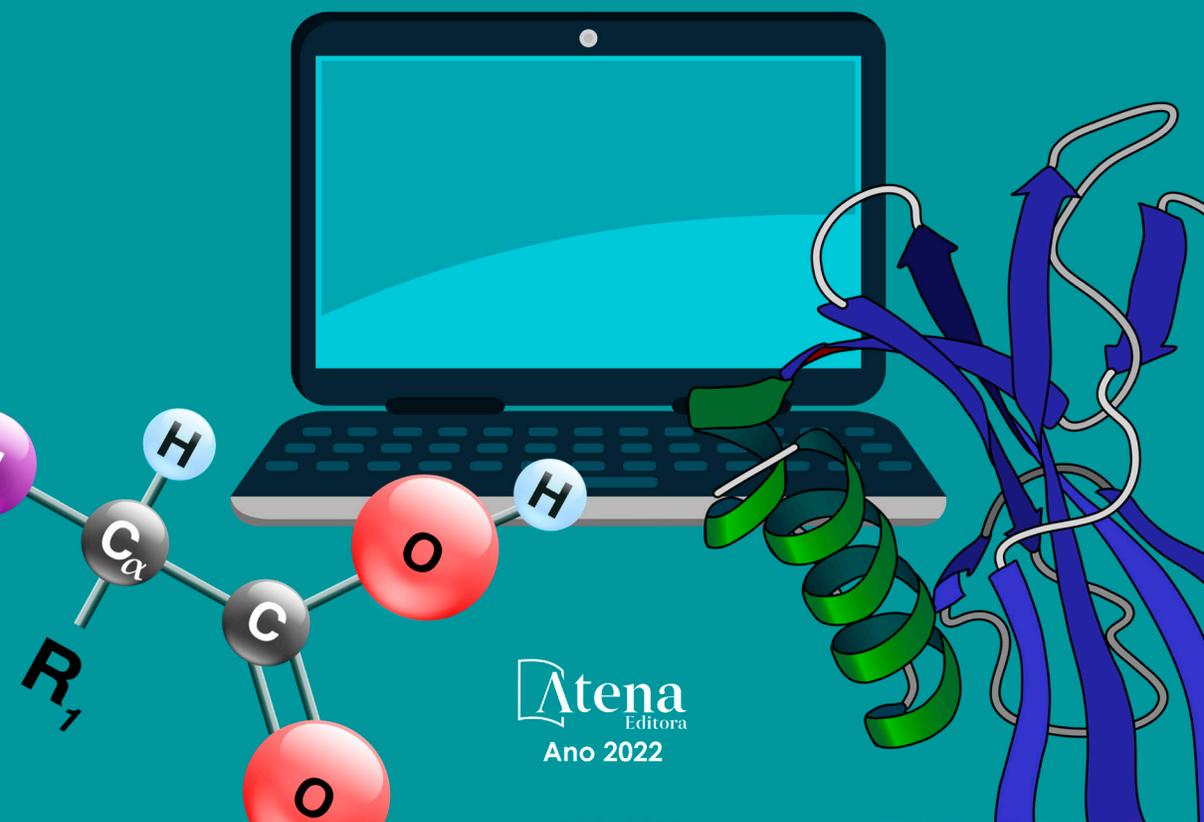


MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

Editado por Kauê Santana e Cláudio Nahum

UM GUIA TEÓRICO E PRÁTICO



Atena
Editora
Ano 2022

MÉTODOS DE MODELAGEM DA ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS

Editado por Kauê Santana e Cláudio Nahum

UM GUIA TEÓRICO E PRÁTICO



Atena
Editora
Ano 2022

Editora chefe

Profª Drª Antonella Carvalho de Oliveira

Editora executiva

Natalia Oliveira

Assistente editorial

Flávia Roberta Barão

Bibliotecária

Janaina Ramos

Projeto gráfico

Bruno Oliveira

Camila Alves de Cremo

Luiza Alves Batista

Natália Sandrini de Azevedo

Imagens da capa

iStock

Edição de arte

Luiza Alves Batista

2022 by Atena Editora

Copyright © Atena Editora

Copyright do texto © 2022 Os autores

Copyright da edição © 2022 Atena Editora

Direitos para esta edição cedidos à Atena Editora pelos autores.

Open access publication by Atena Editora



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons. Atribuição-Não-Comercial-Não-Derivativos 4.0 Internacional (CC BY-NC-ND 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores, inclusive não representam necessariamente a posição oficial da Atena Editora. Permitido o *download* da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

A Atena Editora é comprometida em garantir a integridade editorial em todas as etapas do processo de publicação, evitando plágio, dados ou resultados fraudulentos e impedindo que interesses financeiros comprometam os padrões éticos da publicação. Situações suspeitas de má conduta científica serão investigadas sob o mais alto padrão de rigor acadêmico e ético.

Conselho Editorial**Ciências Biológicas e da Saúde**

Profª Drª Aline Silva da Fonte Santa Rosa de Oliveira – Hospital Federal de Bonsucesso

Profª Drª Ana Beatriz Duarte Vieira – Universidade de Brasília

Profª Drª Ana Paula Peron – Universidade Tecnológica Federal do Paraná

Prof. Dr. André Ribeiro da Silva – Universidade de Brasília

Profª Drª Anelise Levay Murari – Universidade Federal de Pelotas

Prof. Dr. Benedito Rodrigues da Silva Neto – Universidade Federal de Goiás



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



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

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

Dados Internacionais de Catalogação na Publicação (CIP)

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

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-65-258-0582-5

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

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

CDD 613.282

Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166

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



DECLARAÇÃO DOS AUTORES

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



DECLARAÇÃO DA EDITORA

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



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

Prof. Dr. Kauê Santana

APRESENTAÇÃO

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

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

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

Prof. Kauê Santana
Prof. Cláudio Nahum

PREFÁCIO

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

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

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

Particularmente, o último capítulo do livro é bastante interessante, pois além de mostrar as perspectivas da predição tridimensional, ele discute, de forma geral, as atuais limitações. Os autores 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

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

Anderson Henrique Lima e Lima

Alberto Monteiro dos Santos

Kauê Santana da Costa

Nas seções a seguir, serão discutidas algumas aplicações inteligentes dos métodos que utilizam a abordagem *ab initio* e *threading* em diferentes estudos.

*As aplicações do conhecimento da estrutura de proteínas são inúmeros e incluem: (1) estudo conformacional de enzimas (COSTA et al., 2019; SITTEL; JAIN; STOCK, 2014); (2) análise do mecanismo catalítico enzimático (LAMEIRA et al., 2019; MORAES et al., 2012); (3) predição da função de proteínas; (4) estudo de mutantes (NEVES CRUZ et al., 2019); (5) planejamento *in silico* de fármacos e outros compostos bioativos (DA COSTA et al., 2019; LI et al., 2017) (6) estudo do modo de ligação de inibidores (ZHOU et al., 2013).*

Por utilizarem uma estrutura de referência, os métodos de modelagem comparativa se mostram, de modo geral, mais fiéis à representação estrutural dos modelos em relação àqueles gerados por métodos livres de molde. Deste modo, as estruturas obtidas por estes métodos têm sido aplicadas com sucesso para explicar diferentes questões biológicas, de modo similar aos modelos obtidos por métodos experimentais. Em outros casos, os métodos de modelagem comparativa são aplicados, como ferramentas auxiliares, para corrigir e completar modelos gerados por métodos experimentais.

APLICAÇÕES DOS MÉTODOS *AB INITIO* E *THREADING*

Os métodos *ab initio* se destacam pelas suas aplicações no estudo de enovelamento de proteínas e no desenho de proteínas com novas topologias e atividades catalíticas. O desenho, também chamado de *problema inverso de modelagem de proteínas*, desponta como um campo promissor devido às suas aplicações biotecnológicas e tem mostrado relativo sucesso quando aplicado em conjunto com técnicas experimentais que validam os achados (KHOURY et al., 2014). A figura 24 exibe uma visão geral da aplicação dos métodos *ab initio* na modelagem e no desenho da estrutura de proteínas.

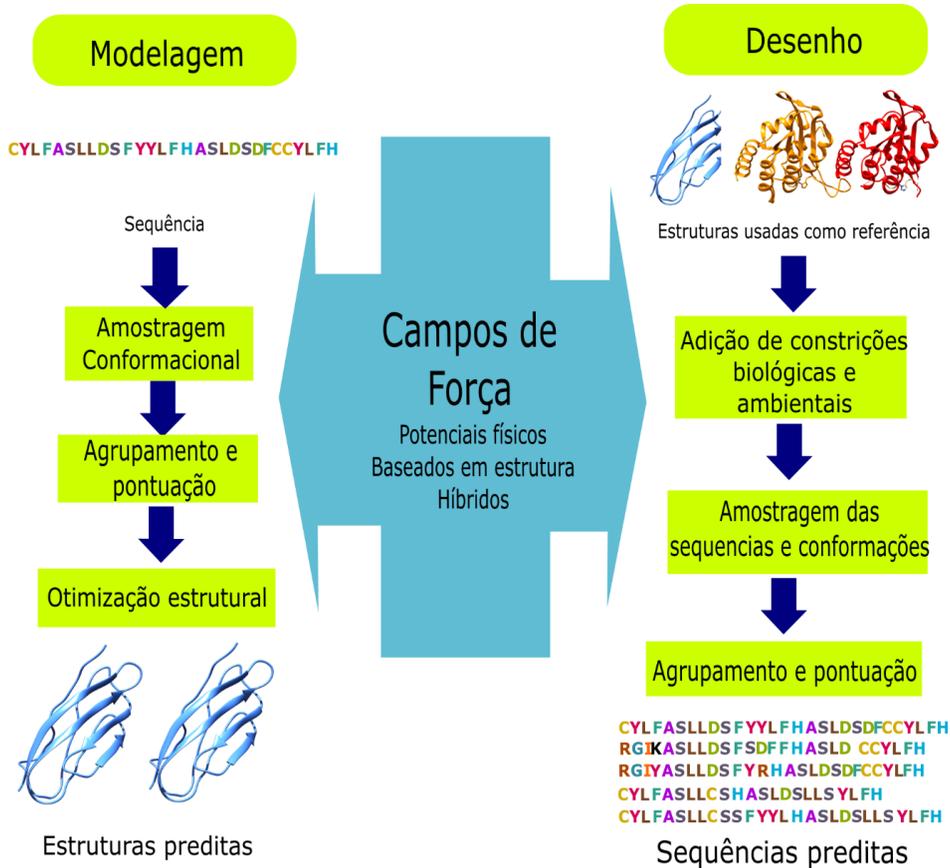


Figura 24. Visão geral do funcionamento e aplicação dos métodos *ab initio* quando usados no desenho e na modelagem de estruturas de proteínas.

DESENHO DE PROTEÍNAS COM NOVAS TOPOLOGIAS

Os métodos *ab initio* são atualmente a única solução para o desenho de novas proteínas (proteínas engenheiradas). Os métodos, portanto, representam um importante passo para o desenvolvimento da biologia sintética através do desenho computacional de proteínas com novas topologias e funções (CORREIA et al., 2014; JIANG et al., 2008; KUHLMAN et al., 2003; RICHTER et al., 2011).

Correia et al. (2014) utilizaram a modelagem *ab initio* para o desenho de uma nova topologia de proteínas, contendo epítomos específicos para a composição de vacinas com o propósito de neutralizar o Vírus Respiratório Sincicial (RSV). Através do Rosetta, os autores utilizaram um protocolo denominado *Fold From Loops* (FFL) que consiste, resumidamente, nos seguintes passos: (1) seleção do motivo proteico funcional e topologia alvo que será unida ao motivo; (2) modelagem *ab initio* para a construção de diferentes esqueletos proteicos que se mostrem compatíveis com a topologia alvo; (3) refinamento das estruturas

obtidas com o objetivo de selecionar aquelas de menor energia para dada conformação; (4) a filtragem dirigida por escolha humana para a correção de falhas remanescentes e seleção dos melhores arcabouços que preservem o epítipo viral.

Neste estudo, os autores produziram pequenos arcabouços proteicos, contendo um novo tipo de dobramento, mas que preservavam um epítipo viral do RSV. Estes oligopeptídeos foram unidos em partículas multivalentes utilizadas na composição de vacinas, e estas, utilizadas na imunização de macacos. As proteínas desenhadas conseguiram, de maneira acurada, mimetizar os epítopos nativos do vírus, sendo capazes desta forma, de induzir uma resposta imune humoral satisfatória nos animais.

PREDIÇÃO EM LARGA ESCALA DE PROTEÍNAS E ANOTAÇÃO DE GENOMAS

Os métodos *ab initio* podem auxiliar a anotação funcional de matrizes abertas de leitura (ORFs, do inglês *Open Read Frame*) de genes, assim como na determinação da função e mecanismo de ação de proteínas e permitir a análise dos diferentes tipos de dobramento de proteínas em estudos de genômica estrutural (JONES, 2000; JOSHI; JYOTHI, 2003; KIHARA et al., 2002; SIMONS; STRAUSS; BAKER, 2001). Kihara et al. (2002), por exemplo, acessaram, em larga escala, a estrutura de diferentes proteínas expressas pelo genoma *Mycoplasma genitalium*, utilizando para isto, a plataforma de predição Touchstone (KIHARA et al., 2001). Esta plataforma aplica a predição *ab initio* e utiliza restrições obtidas por *threading* como o objetivo de reduzir o espaço de pesquisa para predição da estrutura, além disso aplica como método de pesquisa conformacional e refinamento, o Metrópolis MC.

Neste estudo, foi possível prever a ultraestrutura de 85 pequenas proteínas com tamanho igual ou menor a 150 resíduos de comprimento, sendo que destas 34 mostraram-se possíveis de serem preditas pelo método *threading*, das 51 proteínas restantes, 29 convergiram para cinco agrupamentos, destas, se baseado no conjunto treino, foi possível obter 24 proteínas (~84,8% das 29 proteínas) com dobramento correto e correspondente à forma nativa (KIHARA et al., 2002). Outros estudos utilizam abordagens mistas que unem a modelagem por homologia, *threading* e *ab initio*. Singh et al. (2014), por exemplo, empregaram estes métodos com o objetivo de realizar a anotação baseada em estrutura de genes de *Helicobacter pylori*, para isto os autores utilizaram as plataformas I-TASSER (servidor que realiza um misto de *threading* e *ab initio*), (ZHANG, 2008).

O base de dados ModBase (com modelos obtidos por modelagem comparativa) e Phyre2 (servidor para modelagem por *threading*) (KELLEY et al., 2015), além de servidores de alinhamento e servidores de predição estrutural, tais como: o Psipred (MCGUFFIN; BRYSON; JONES, 2000) para a predição da estrutura secundária, o pGenTHREADER (LOBLEY; SADOWSKI; JONES, 2009) para reconhecimento de dobramento e identificação

de homólogos distantes, e o Fugue (SHI; BLUNDELL; MIZUGUCHI, 2001) para alinhamento também baseado em *threading*. Os autores conseguiram anotar 464 proteínas de um total de 557 analisadas e conhecidas previamente como não caracterizadas (SINGH; GUTTULA; GURUPRASAD, 2014).

PREDIÇÃO DE MUDANÇAS CONFORMACIONAIS

Recentemente, os métodos *ab initio* foram aplicados para prever conformações induzidas por regulação alostérica de determinadas moléculas efectoras. Embora, atualmente, a análise, em detalhes de efeitos alostéricos, possa ser realizada por meio de simulação de dinâmica molecular, esta ferramenta é limitada para alterações conformacionais na escala de nanosegundos. Partindo disso, Kidd et al. (KIDD; BAKER; THOMAS, 2009) utilizaram a metodologia de predição do Rosetta para prever alterações conformacionais em três proteínas alostéricas bem conhecidas (CheY, Integrina α L domínio I, e Ras) com o objetivo de determinar os estados não ligados (sem o ligante) a partir de estruturas ligadas (complexadas ao ligante) (KIDD; BAKER; THOMAS, 2009).

Neste trabalho, os autores utilizaram, como critério de seleção das proteínas do conjunto de teste, aquelas que apresentavam estruturas ligadas e não ligadas no *RCSB Protein Data Bank*, com resolução $\leq 2,5$ Å, sequências com < 200 aminoácidos, e um rearranjo estrutural C α -C α entre as duas formas maior que 3,5 Å. O protocolo de predição consistiu em três principais passos: (1) a geração da diversidade de estrutural através do algoritmo de montagem de fragmentos, o Metrópolis Monte Carlo e utilizando como estrutura de partida, a forma nativa ligada (holo), (2) o refinamento das cadeias laterais de todos os resíduos (3) minimização de todos os átomos explícitos da proteína.

Para seleção, utilizaram um algoritmo de agrupamento e definiram como limiar para as estruturas similares do mesmo agrupamento um RMSD de não mais que 1,0 Å e selecionaram aquelas de menor energia que se encontravam nos agrupamentos com o maior número de estruturas. Além disso, calcularam a energia de interação de pares de resíduos. Os autores especificaram em quais regiões da proteína eram mais susceptíveis de alterações induzidas por efetores alostéricos e obtiveram maior sucesso na predição de alterações conformacionais quando analisaram estruturas secundárias e alças e conseguiram ainda prever com acurácia de 0,3–3,4 Å (C α -RMSD) estruturas de baixa energia próximas à conformação cristalográfica das formas não ligadas.

DESENHO DE ENZIMAS CONTENDO NOVOS SÍTIOS CATALÍTICOS

Um campo extremamente promissor e recente na Biologia Estrutural é o desenho computacional de proteínas com sítios ativos capazes de executar novas atividades catalíticas (RICHTER et al., 2011; ZANGHELLINI et al., 2006). O desenho computacional normalmente é referido como um problema inverso de predição, pois inicia pela análise das

estruturas de proteínas conhecidas em que se deseja extrair características interessantes, para então se modelar sequências que possam adquirir a estrutura desejável para estas características. Como diferentes sequências podem adquirir a mesma estrutura, podemos considerar que existe uma redundância no espaço de pesquisa das sequências (KHOURY et al., 2014).

Os métodos *ab initio* são os preferidos nas estratégias de desenho computacional, a exemplo disto, Richter et al. (2011), utilizando a estratégia de modelagem do Rosetta, desenvolveram um protocolo para o desenho de enzimas que catalisam reações químicas novas (RICHTER et al., 2011). O protocolo consiste resumidamente nos seguintes passos: (1) escolha do mecanismo catalítico (utilizando métodos de mecânica quântica) e do arcabouço mínimo do sítio ativo envolvido na reação; (2) a seleção de sítios de um conjunto de esqueletos proteicos existentes que possam representar o sítio ativo idealizado para a catálise da reação química através do módulo RosettaMatch (ZANGHELLINI et al., 2006); (3) desenho do sítio ativo através dos arcabouços proteicos de modo a permitir a complementaridade espacial com o substrato (e seus estados de transição) e de estabilizar as cadeias laterais dos resíduos catalíticos nas suas conformações de ligação; e o último passo consiste na (4) validação e pontuação das sequências obtidas com relação à suas características intrínsecas, tais como número de ligações de hidrogênio e contatos não locais entre resíduos (Figura 25) (RICHTER et al., 2011).

A evolução dirigida pode ser utilizada como técnica complementar ao desenho computacional de proteínas com o objetivo de melhorar a atividade catalítica de enzimas engenheiradas (ALTHOFF et al., 2012; GIGER et al., 2013; KHERSONSKY et al., 2011). A evolução dirigida refere-se às alterações aleatórias na sequência de DNA de determinado gene, induzidas por técnicas como Error-prone PCR e DNA *shuffling*, seguida de clonagem e seleção dos clones recombinantes que exibam o fenótipo desejado. Este processo é repetido de maneira cíclica com o objetivo de incrementar características estruturais de proteínas selecionadas na fase inicial (JÄCKEL; KAST; HILVERT, 2008). No entanto, por ser um processo randômico que não analisa diretamente a estrutura da proteína, técnicas computacionais podem favorecer a análise de determinantes envolvidos no mecanismo de ação molecular e, portanto, auxiliar no desenho racional de proteínas com atividade melhorada.

Similarmente Röthlisberger et al. (2008) desenharam enzimas capazes de catalisar a eliminação Kemp, um modelo de reação que consiste na transferência de próton de um carbono que normalmente é restringida, pois requer enormes barreiras de energia para se processar naturalmente (RÖTHLISBERGER et al., 2008).

Para escolha do mecanismo catalítico, os autores utilizaram cálculos de mecânica quântica do estado de transição para idealizar um possível sítio ativo para a catálise da reação, em seguida, utilizando o algoritmo RosettaMatch, pesquisaram por possíveis esqueletos proteicos que poderiam, em teoria, representar este sítio ativo. Depois de obtidas

as estruturas, todas as proteínas foram analisadas com relação a parâmetros cinéticos, em seguida, o gene da enzima desenhada *in silico* com melhor atividade catalítica (denominada KE07) foi submetido à metagênese e ao DNA *shuffling*, e as formas variantes, em seguida, clonadas em plasmídeos de expressão e transformadas em *E. coli* (RÖTHLISBERGER et al., 2008). As colônias recombinantes foram triadas quanto sua capacidade de hidrolisar o 5-nitrobenzoisoxazol, seguida pela formação dos seus produtos.

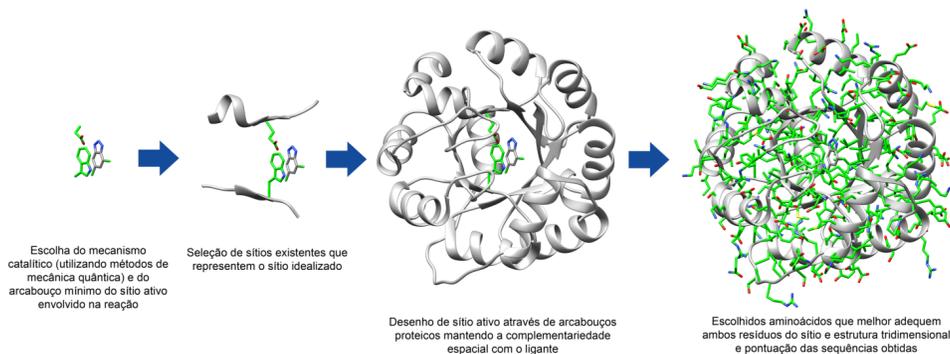


Figura 25. Esquema representativo do desenho computacional de proteínas com novos sítios catalíticos desenvolvido por Richter et al. (2011).

PREDIÇÃO DE ESTRUTURA DE PROTEÍNAS OU SEGMENTOS SEM HOMÓLOGOS

Em alguns casos, nos quais a sequência da proteína-alvo (ou regiões desta) não apresenta qualquer identidade com a sequência de estruturas previamente resolvidas por métodos experimentais e disponíveis em base de dados, os métodos *ab initio* podem ser a solução. Atualmente, esta é uma das mais importantes aplicações dos métodos *ab initio* (BAKER et al., 2006; BAR-ON et al., 2011; KEMEGE et al., 2011).

Em um estudo, realizamos a modelagem *ab initio* para prever a estrutura tridimensional do Fator de Infecção Viral (Vif do inglês, *virion infectivity factor*), uma proteína acessória do Vírus da Imunodeficiência Humana (DA COSTA et al., 2014). Para isto, utilizamos o protocolo *AbinitioRelax* disponível no programa Rosetta (BRADLEY; MISURA; BAKER, 2005) e definimos a relação espacial existente entre as cadeias laterais dos resíduos de cisteína e histidina que compõem um motivo dedo-dezinco essencial para o mecanismo de ação molecular desta proteína. Ao todo, foram gerados 50 mil modelos que, destes 500 foram selecionados e, em seguida, submetidos a uma análise alternativa de centroide e à inspeção visual com relação à formação do motivo e a disposição dos resíduos presentes nos domínios de ligação da proteína.

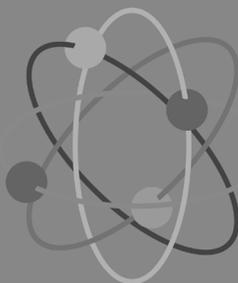
No referido trabalho, conseguimos prever a localização de domínios de ligação e,

além disso, realizamos docagem molecular com outras proteínas, simulação de dinâmica molecular para análise de mudanças conformacionais e cálculos de energia livre de ligação da Vif complexada com as proteínas celulares EloB e EloC. O modelo teórico foi comparado com um fragmento do BC-box do Vif obtido por cristalografia e obtivemos um RMSD satisfatório de 0,904 Å para esta região (STANLEY et al., 2008).

Com relação aos métodos que utilizam a abordagem *threading* para predição da estrutura, estes são recomendados para proteínas ou regiões da cadeia que não apresentam identidade mínima de 30% com sequências utilizadas como molde que permitam a aplicação de métodos de modelagem por homologia. Com relação a estes, recentemente, publicamos um estudo em que analisamos as alterações conformacionais da 3-Hydroxy-3-Methylglutaryl Coenzyme Reductase, uma enzima envolvida na biossíntese do colesterol com o propósito de identificar a relevância da região do domínio Flap para a primeira etapa de reação desempenhada pela enzima. Neste estudo, esta região ausente na cristalografia foi modelada por *threading* utilizando o programa I-TASSER (COSTA et al., 2019) we constructed a model of human HMGR (hHMGR. Em outro estudo, a abordagem *threading* foi utilizada para predizer a estrutura completa da Alba3 de *Leishmania infatum*, uma enzima que apresenta identidade menor que 30% com seus homólogos próximos, mas exibe um domínio Alba envolvido na ligação ao DNA/RNA comum com outras enzimas (DA COSTA et al., 2017).

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.

