

ProtCool 2.0: um gerador de protocolo para ancoragem e simulações de dinâmica molecular em complexos proteína-ligante em versão web

Moisés P. Souza¹, Levi M. Magny¹, Fabiana C. Guedes¹, Carlos H. da Silveira¹

¹Instituto de Ciências Tecnológicas – Universidade Federal de Itajubá (UNIFEI)
Caixa Postal 50 – 37500-903 – Itajubá – MG – Brasil

moises.souza@ifsuldeminas.edu.br, levi_mgy@unifei.edu.br,
fabiana.costaguedes@unifei.edu.br, carlos.silveira@unifei.edu.br

Abstract. *The COVID-19 pandemic made evident the high demand for computational systems that speed up the indication of new drugs. In this sense, understanding the dynamic behavior of biomolecule complexes is essential. However, the preparation of such simulations is highly complex, and their numerous details are not always sufficiently highlighted, compromising their reproducibility. Here, we propose the web tool ProtCool - a protocol generator, focused on the integration between docking and molecular dynamics of protein-ligand complexes. Its use will be demonstrated in the simulation of the interaction between acetylcholinesterase and galantamine, used in the treatment of Alzheimer's.*

Resumo. *A pandemia de COVID-19 deixou evidente a alta demanda por sistemas computacionais que agilizem a indicação de novos fármacos. Nesse sentido, compreender o comportamento dinâmico de complexos biomoleculares é algo fundamental. No entanto, a preparação de tais simulações é de grande complexidade, e seus inúmeros detalhes nem sempre são suficientemente destacados, comprometendo sua reprodutibilidade. Propõe-se aqui a ferramenta web ProtCool – um gerador de protocolos, focado na integração entre ancoragem e dinâmica molecular de complexos proteína-ligantes. Será demonstrado seu uso na simulação da interação entre acetilcolinesterase e galantamina, usada no tratamento do Alzheimer.*

1. Introdução

Em um artigo hoje clássico [McCammon et al, 1977], o laboratório chefiado por Martin Karplus, em Harvard, anunciou uma das primeiras simulações computacionais do comportamento dinâmico de uma proteína globular (BPTI - *Bovine Pancreatic Trypsin Inhibitor*), descrevendo padrões de movimento de seus átomos no tempo. Por essa época, já se sabia que uma modelagem teórica adequada para esse tipo de simulação envolveria as leis da quântica, mas sua aplicação a sistemas biológicos com grandes números de átomos era computacionalmente intratável [Brini et al, 2020]. Uma alternativa foi aproximar as interações quânticas usando “campos de força”, um conjunto de equações empíricas regulando as variações em distâncias e ângulos para ligações covalentes, acrescidas de termos não covalentes para interações de van der Waals, Coulômbicas, ligações de hidrogênio entre outros. Além disso, os sistemas atômicos também deveriam respeitar as leis Newtonianas do movimento e as leis da termodinâmica [Brini et al, 2020]. Foi com base nessas simplificações que Karplus e sua equipe conseguiram operar as primeiras simulações realísticas da BPTI, feito histórico reconhecido posteriormente pelo Nobel de Química de 2013 [Karplus, 2014].

Atualmente, dada a evolução da química computacional e da computação de alto desempenho, é possível realizar simulações fidedignas de milhões de átomos em escalas de micro ou milissegundos contemplando proteínas inteiras ou grandes complexos biomoleculares, em solventes explícitos ou embutidos em bicamadas lipídicas. Compreender o comportamento dinâmico desses sistemas contribui para a elucidação de seus mecanismos de ação, um conhecimento fundamental na descoberta de novos fármacos [Brini et al, 2020].

Se por um lado os enormes avanços da química computacional, em teorias, algoritmos e hardware, resultaram numa maior abrangência, eficiência e acurácia dos sistemas simulados, por outro lado, o planejamento, preparação, execução, análise e reprodução de tais simulações tornaram-se bem mais complexos [Purawat et al, 2017]. As dificuldades advêm principalmente dos inúmeros detalhes, parâmetros e passos exigidos pelos protocolos de simulação, quase sempre insuficientemente relatados na metodologia dos artigos publicados. Tudo isso contribui para tornar mais tortuosa e íngreme a curva de aprendizagem das técnicas e experimentos utilizados em simulações de dinâmica molecular.

Mesmo quando disponibilizados, em geral os procedimentos de execução das simulações ficam encapsulados em *scripts ad hoc* que tendem a ser de difícil interpretação, adaptação e execução para quem não participou da sua elaboração e uso. Acrescente a isso, complicações advindas de particularidades numéricas e computacionais, em função de diferentes tipos e versões de sistemas operacionais, compiladores, bibliotecas e suas dependências, aparatos de hardware. Sem contar as informações tácitas ou implícitas aos experimentos, em que nem tudo está transparente ou fielmente relatado na metodologia, como outros workflows que foram testados, variações na sequência, se foram replicados e quantas vezes, as combinações e ajustes de parâmetros. Muitas vezes, os próprios autores de um experimento podem perder a capacidade de reproduzi-los tempos depois, uma vez que nem todos os detalhes do que foi feito ficou devidamente registrado, e a memória humana pode falhar em lembrar de todos eles [Purawat et al, 2017].

Outro importante fator a ser considerado é o didático, envolvendo a aprendizagem das simulações por pesquisadores inexperientes ou novatos em química computacional. Uma parte significativa desse processo pode requerer competências e habilidades computacionais mais profundas, ainda não suficientemente dominadas por esse perfil de pesquisadores. Além disso, a complexidade inerente à química computacional pode exigir anos de dedicação para um domínio produtivo e confiável de todos os seus conceitos, detalhes e aplicações. Tudo isso pode tanto desencorajar alguns pesquisadores iniciantes, quanto fazer com que outros, mais ansiosos, subestimem sua complexidade e cometam erros ou enganos nem sempre fáceis de serem detectados [Ribeiro et al, 2016].

No intuito de contribuir com soluções a tais desafios, está sendo proposta a versão web da ferramenta ProtCool – um gerador de protocolos para simulações de dinâmica molecular em complexos proteína-ligante. Em sua versão inaugural, ProtCool foi estrategicamente projetado para focar na integração entre ancoragem molecular e dinâmica molecular do complexo resultante, em função do seu potencial impacto no processo de descoberta de novos fármacos. A pandemia de COVID-19 deixou evidente a

alta demanda por sistemas computacionais que simplifiquem e agilizem a indicação de candidatos a fármacos, sem perda de confiabilidade [Parks et al, 2020].

Dadas as três grandes fases de qualquer simulação de dinâmica molecular entre preparação, execução e análise, o ProtCool, como um gerador de protocolos, coloca ênfase na preparação, considerado o estágio mais complexo de todo o processo [Purawat et al, 2017]. Ainda vem com o bônus de incluir a ancoragem de ligantes durante essa preparação. Mas, é importante salientar que o ProtCool não executa as dinâmicas após preparação nem opera análises dos resultados. Essas duas fases são deixadas para o usuário, conforme a disponibilidade de seus recursos computacionais, conveniência e familiaridade com suas ferramentas de análise.

Como um *expert* na preparação, ProtCool percorre o *workflow* que lhe foi programado, gerando todos os arquivos de configuração necessários para uma execução confiável da dinâmica molecular no aparato computacional do usuário. O *workflow* gera um *log* detalhado de tudo, de modo que pesquisadores podem checar a consistência de cada etapa da preparação. Uma vez que esse *workflow* esteja definido, executado, validado e salvo, todo o processo fica inteiramente registrado no ProtCool. Isso facilita a reprodutibilidade, já que a simulação poderá ser reexecutada a qualquer tempo. Poderá ser reexecutada, inclusive, por outros pesquisadores, em outros laboratórios, desde que os arquivos de preparação sejam devidamente compartilhados. Também facilita a reusabilidade, por conta que a adaptação de protocolos antigos gerados no ProtCool para outros contextos de ligantes e alvos é algo relativamente simples de ser feita.

2. Arquitetura

O ProtCool é uma ferramenta cliente/servidor, compondo no lado cliente uma interface gráfica web, acessada via navegador do usuário. No lado servidor, ficam hospedados os *scripts* que sustentam os *workflows*. Logo, todas as informações são preenchidas e geradas pela API (*backend*), que posteriormente são visualizadas na camada de interface (*frontend*).

2.1. Backend

O *backend* do ProtCool foi construído totalmente na linguagem de programação Python. A escolha dessa linguagem se deu por ela oferecer recursos e bibliotecas robustas para tratamentos de arquivos de texto, vindo ao encontro da necessidade de tratar arquivos extensos e prepará-los para dinâmica molecular. O *backend* é formado por um *script* gerenciador representando o *workflow*, responsável por toda integração dos *scripts* com funções específicas, bem como as suas validações e execuções, garantindo a coesão e a integridade do sistema.

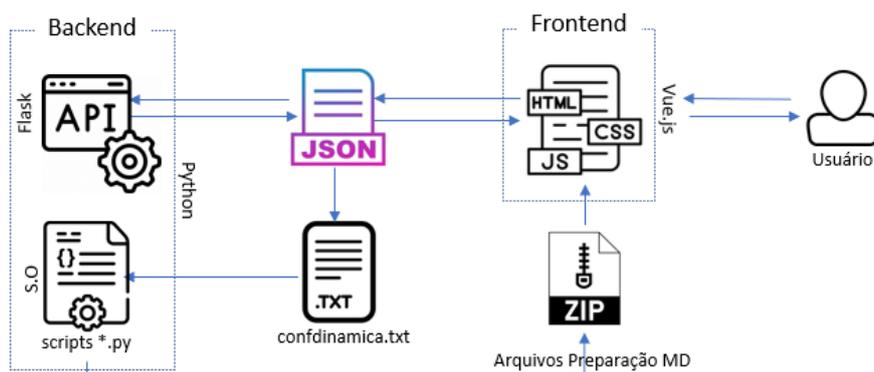


Figura 1. Arquitetura ProtCool. (Fonte: Autores).

2.2. Frontend

O objetivo do *frontend* é estruturar graficamente uma interface que seja intuitiva ao usuário e organize o *workflow*. Cumpre também papel importante no correto preenchimento das informações que alimentarão o *backend*. Nesse sentido, a interface web permite ao pesquisador preencher todos os campos por meio de formulários divididos em etapas como abstração do *workflow* original. Para sua construção foi utilizado o *framework* Javascript Vue.js. Sua escolha deveu-se à sua versatilidade em conciliar um desenvolvimento que adota as melhores práticas com funcionalidades simples de implementação.

3. Funcionalidades / Etapas

As principais funcionalidades do ProtCool confundem-se com as etapas do *workflow*, dos quais destacam-se (uma descrição mais completa das funcionalidades pode ser vista na tese em [GUEDES, 2021]):

- 1- **Informações Gerais:** essa é a etapa inicial, sua função principal é receber os dados iniciais do usuário e a proteína alvo (PDB-ID) desejada recuperada automaticamente do repositório de biomoléculas RSCB (<https://www.rcsb.org>). Por meio dela é permitido receber os parâmetros iniciais do estudo, indicando quais etapas serão processadas e quais parâmetros de proteína serão considerados para a simulação da dinâmica molecular.
- 2- **Ligantes:** é responsável por receber as informações dos ligantes que vão participar do processo de ancoragem, consistindo num posicionamento energeticamente e estruturalmente aceitável do ligante sobre a proteína alvo. Nesta versão do ProtCool, aceitam-se ligantes catalogados na base de dados ZINC (<https://zinc.docking.org/>) ou por *upload* de arquivos formato *.mol2 fornecidos pelo próprio pesquisador.
- 3- **Ancoragem:** visa buscar o encaixe mais perfeito possível entre duas moléculas. É responsável por receber os parâmetros para realização da ancoragem (*docking*), tais como: a posição e tamanho do *pocket* onde o ligante será inserido, a quantidade de *CPU* destinada ao processamento, opção 'Flex' para realizar o *docking* flexível. Nesta versão do ProtCool, podem ser usados os sistemas 'Autodocking Vina' e 'Smina'.
- 4- **Protonação:** faz a inserção de hidrogênios na proteína alvo, definindo parâmetros, tais como: salinidade, tipo de solvente e pH. Nesta versão do

- ProtCool, o sistema de protonação utilizado é o H++ (<http://newbiophysics.cs.vt.edu/H++/>).
- 5- **Solvatação/Ionização:** compreende a inserção de águas e íons explícitos ao redor do complexo proteína-ligante. Entra também com dados referentes à solvatação e ionização que serão utilizados no campo de força. Nesta versão do ProtCool, aceitam-se apenas parâmetros e campos de força da *suite* AMBER (<https://ambermd.org/>).
 - 6- **Simulação:** a última etapa do ProtCool é a simulação, na qual o pesquisador deve informar os parâmetros e configurações necessárias para a execução da simulação da dinâmica molecular em si. Atual versão do ProtCool trabalha com o sistema NAMD (<http://www.ks.uiuc.edu/Research/namd>). Esta etapa pode ser subdividida em outras, conforme desejo do usuário. Em geral, alocam-se uma ou mais etapas para a minimização de energia do sistema, seguida por outras etapas de relaxamento atômico de forma gradual, definindo tempos de simulação para cada uma. Após o relaxamento e equilibração do sistema, vem a simulação da dinâmica molecular per si, por um tempo mais longo.

ProtCool - Dynamic Configs General Info

W
ProtCool

Researcher:
Moises Souza

Email address: moises.souza@ifsuldeminas.edu.br Language: Ingles

Provide a PDB ID:
4EY6

Define workflow steps:

- Prepare
- Protonation
- Prepare NAMD
- Hydrogen
- Molecular Model
- Execution Force Field
- Docking
- Vina
- Smina
- Heteroatom
- Water

Model:
 Model: Not Found

Chains:
 Chain: A Chain: B

Disulfide Bridges:

- CYS A 69 - CYS A 96 CYS A 257 - CYS A 272
- CYS A 409 - CYS A 529 CYS B 69 - CYS B 96
- CYS B 257 - CYS B 272 CYS B 409 - CYS B 529

Figura 2. Tela inicial do ProtCool. (Fonte: Autores).

4. Demonstração

Uma demonstração de uso ProtCool por ser vista em bit.ly/3nNvvSr seu acesso em bit.ly/3ma2mRf. O vídeo mostra a preparação de um protocolo de ancoragem e dinâmica molecular para o complexo envolvendo a proteína acetilcolinesterase (4EY6) com o ligante galantamina (GNT), utilizado no tratamento do Alzheimer [Rocha, 2017]. A galantamina liga-se à acetilcolinesterase, inibindo sua ação e elevando os níveis do neurotransmissor acetilcolina no cérebro, com efeitos positivos na capacidade cognitiva e de memória em pacientes com Alzheimer. Em [Rocha, 2017], foi feita uma simulação de dinâmica molecular desse complexo, cuja etapa preparatória envolveu *scripts ad hoc* dos próprios autores do artigo que culminaram num complexo protocolo de 21 passos. No vídeo, será demonstrado como montar esse mesmo *workflow* utilizando o ProtCool.

5. Considerações Finais

ProtCool, em sua versão inicial, oferece um escopo limitado de opções pré-configuráveis, ainda que estejam disponibilizados as etapas e os parâmetros mais comumente utilizados nos protocolos de dinâmica molecular descritos na literatura. Mais opções poderão ser

integradas à ferramenta conforme a demanda dos usuários. Seja como for, usuários com demandas específicas e com alguma habilidade de programação não teriam dificuldades em customizá-lo de acordo com suas necessidades. Os passos envolvendo os *workflows* produzidos pelo ProtCool são modularizados, de modo que a adequação de alguma etapa tende a ter poucos efeitos colaterais nas demais.

6. Agradecimentos

Um agradecimento especial à Vale SA, pelo apoio a este projeto.

7. Referencias

- Brini, E., Simmerling, C., & Dill, K. (2020). Protein storytelling through physics. *Science*, 370(6520). <https://doi.org/10.1126/science.aaz3041>
- Guedes, F. C. (2021). ProtCool: um Gerador de Protocolos para Ancoragens e Simulações de Dinâmica Molecular em Complexos Proteína-Ligante. [Unifei]. https://repositorio.unifei.edu.br/jspui/bitstream/123456789/2479/1/Tese_2021038.pdf
- Karplus, M. (2014). Development of multiscale models for complex chemical systems: From H+H2 to biomolecules (nobel lecture). *Angewandte Chemie - International Edition*, 53(38), 9992–10005. <https://doi.org/10.1002/anie.201403924>
- McCammon, J. A., Gelin, B. R., & Karplus, M. (1977). Dynamics of folded protein. *Nature*, 267, 585–590. <https://doi.org/10.1038/267585a0>
- Parks, J. M., Ph, D., Smith, J. C., & Ph, D. (2020). How to Discover Antiviral Drugs Quickly. *The New England Journal of Medicine*, 1–4. <https://doi.org/10.1056/NEJMcibr2007042>
- Purawat, S., Jeong, P. U., Malmstrom, R. D., Chan, G. J., Yeung, A. K., Walker, R. C., Altintas, I., & Amaro, R. E. (2017). A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics. *Biophysical Journal*, 112(12), 2469–2474. <https://doi.org/10.1016/j.bpj.2017.04.055>
- Ribeiro, J. V., Bernardi, R. C., Rudack, T., Stone, J. E., Phillips, J. C., Freddolino, P. L., & Schulten, K. (2016). QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. *Scientific Reports*, 6(May), 1–14. <https://doi.org/10.1038/srep26536>
- Rocha, R. E. O. (2017). Estudos de Modelagem Molecular dos Mecanismos de Afinidade Relativa para Quatro Galantamínicos com Potencial Anti-Alzheimer [UFMG]. https://repositorio.ufmg.br/bitstream/1843/BUOS-B43MLZ/1/disserta_o.pdf