

## **Uso de Quimioinformática para o Planejamento de Potenciais Inibidores de Alvos Relacionados com o Vírus SARS-CoV-2**

Objetivo: uso de métodos *in silico* para analisar as possíveis interações entre um conjunto de substâncias sintéticas, disponíveis no catálogo da empresa MERCK do Brasil, e a enzima M<sup>pro</sup>, um dos alvos relacionados com o ciclo de vida do SARS-CoV-2

Journal of Chemical Information and Modeling

Impact Factor 2021: 6.162

Citations 2021:30,162

CiteScore 2021: 7.7

### **MATERIAIS E MÉTODOS**

#### **Estrutura do receptor e ligantes**

A estrutura cristalográfica da enzima M<sup>pro</sup> do SARS-CoV-2 foi obtida no banco de dados *Protein Data Bank* (PDB), cujo código é PDB ID 6Y2F<sup>1</sup>. O conjunto de ligantes usados para esse estudo de triagem virtual foi fornecido pela empresa MERCK do Brasil com o intuito de buscar um potencial inibidor da M<sup>pro</sup><sup>2</sup>.

#### **Simulações de acoplamento molecular (*docking*)**

O acoplamento molecular (*docking*) é um método computacional utilizado para realizar o estudo das possíveis interações entre o potencial inibidor com o sítio ativo da enzima estudada<sup>3</sup>. O programa empregado neste trabalho para as simulações de *docking* é conhecido como GOLD (do inglês, *Genetic Optimization for Ligand Docking*), o qual possibilita que as conformações dos ligantes produzidas sejam categorizadas a partir de um ranqueamento (*scoring*) e, assim, pode-se analisar as interações intermoleculares formadas no sítio de ligação<sup>4</sup>.

Os ligantes selecionados foram otimizados no programa Sybyl 8.1 utilizando-se o método semiempírico PM3 e suas cargas foram calculadas com o módulo MOPAC. O mesmo programa foi utilizado para realizar a preparação da estrutura da enzima SARS-CoV-2 *main protease* (PDB ID: 6Y2F). Para isso, foi empregado o módulo Biopolymer que otimiza os átomos de hidrogênio e determina as cargas atômicas do receptor a partir do método MM94<sup>5</sup>. Após as otimizações das estruturas e dos ligantes as simulações foram realizadas no programa GOLD utilizando a função ChemScore, resíduos Hys163 e Gln189 flexíveis e sítio ativo com um raio de 7 Å<sup>4</sup>.

#### **Propriedades Farmacocinéticas**

Para analisar a substância bioativa (ligante) aspectos como absorção, distribuição, metabolismo, excreção e toxicidade precisam ser investigados<sup>6</sup>. Os cálculos das propriedades farmacocinéticas usando os servidores SwissADME e eMolTox<sup>7,8</sup>.

O *SwissADME* foi o servidor online utilizado para as simulações das propriedades físico-químicas relacionadas à biodisponibilidade oral e é focado nas características de absorção, distribuição, metabolismo e excreção<sup>7</sup>. As propriedades farmacocinéticas e farmacodinâmicas dos ligantes foram avaliadas enfatizando as propriedades “*drug-likeness*”, os filtros de Lipinski (Massa molecular (MW) menor ou igual a 500 D; log P menor que 5; aceptores de ligação de hidrogênio

menor que 10; doadores de ligação de hidrogênio menor que 5)<sup>9</sup>. Estes filtros atrelados às regras de Lipinski respondem às questões como biodisponibilidade, permeabilidade, absorção, metabolismo, excreção e toxicidade, e são considerados propriedades muito importantes já que indicam a viabilidade de moléculas bioativas como potenciais candidatos a fármacos<sup>10</sup>.

O servidor eMolTox, focado na toxicidade de uma molécula, analisa fatores como cardiotoxicidade; mutagenicidade; hepatotoxicidade; citotoxicidade; genotoxicidade; carcinogenicidade; toxicidade mitocondrial; alertas estruturais tóxicos; toxicidade do SNC; nefrotoxicidade; toxicidade respiratória; toxicidade reprodutiva; sensibilização da pele; citocromo P450 e toxicidade oral aguda<sup>8</sup>. Para a estrutura utilizada como *input* em ambos os servidores foi utilizado o programa ChemSpider, o qual utiliza o código SMILES de cada molécula para obter informações sobre propriedades físicas e químicas dos compostos<sup>11</sup>.

### Dinâmica Molecular

As simulações de dinâmica molecular utilizam a mecânica clássica que fornece informações sobre o comportamento dinâmico microscópico do sistema em estudo; esse comportamento é calculado em função do tempo para todos os átomos.<sup>12</sup> Sendo assim, as simulações de dinâmica molecular calculam as posições atômicas do complexo em um determinado período de tempo; em geral, as simulações do sistema proteína-alvo + inibidor é realizado em nanosegundos ( $1\text{ns} = 10^{-9}\text{ s}$ )<sup>13</sup>. A preparação da simulação necessita da inserção de um conjunto de potenciais de interações conhecidos como campo de força. Os campos de força são funções de energia potencial empíricas que podem ser implementadas no sistema molecular de interesse, onde é possível calcular a energia potencial total do sistema,  $V(r)$ , a partir da estrutura tridimensional do sistema. A energia potencial  $V(r)$  é descrita como a soma de vários termos de energia, tais como átomos ligados (comprimentos e ângulos de ligação, ângulos diedros) e átomos não ligados (interações de van der Waals e de Coulomb).<sup>14</sup>

Neste trabalho, a primeira etapa da DM, foi o cálculo da carga RESP (do inglês, *Restrained Electrostatic Potential*)<sup>15</sup> do inibidor (molécula 1) mais ativo selecionado na triagem virtual utilizando a metodologia *Hartree-Fock*<sup>16</sup>, implementada no programa Gaussian 09<sup>17</sup>, com a função de base 6-31G\*<sup>40</sup>. Após essa etapa, foi analisada qual a melhor caixa solvente para o alvo proteína principal do SARS-CoV-2, onde essa análise foi realizada utilizando o programa Chimera 1.12<sup>18</sup>. Os melhores parâmetros foram a caixa periódica do tipo octaédrica e com 12 Å de distância entre a proteína principal do SARS-CoV-2 e as paredes da caixa conferindo por análise visual. Após utilizar os parâmetros de solvatação, foram aplicados os seguintes parâmetros:

- I. Preparação da topologia do ligante pelo módulo Antechamber implementado no programa AmberTools12 e com as coordenadas obtidas pelo Gaussian 09<sup>19</sup>;
- II. Inserção do campo de força ff99SB pelo programa Tleap;
- III. Formação do complexo (Proteína principal do SARS-CoV-2 + Molécula) pela união empregando o programa Tleap;
- IV. Cálculo das cargas, sendo que a carga obtida para a proteína principal do SARS-CoV-2 pelo Tleap foi de +2 e foram inseridos 2 íons sódio com o intuito de neutralizar o sistema;
- V. Inserção das coordenadas periódicas (moléculas de água do tipo TIP3P) com distanciamento de 12 Å para as paredes da caixa octaédrica (solvatação explícita);
- VI. Preparação dos *scripts* envolvidos em todos os procedimentos da DM com as seguintes informações: conjunto de configurações e propriedades que serão mantidas constantes no cálculos das equações de Newton (*ensemble*) do tipo isotérmico-isobárico ou NPT (relacionado ao número de partículas, pressão e temperatura constantes, do inglês

*Constant-Temperature, Contant-Pressure ensemble, NPT*), termostato de Langevin (ntt=3) e barostato de Monte Carlo.<sup>19</sup>.

Após a preparação do sistema, o mesmo foi submetido ao cálculo de quatro minimizações, tendo como objetivo a eliminação dos contatos muito próximos entre os átomos e a minimização da energia potencial do sistema (SARS-CoV-2<sup>Mpro</sup> + Molécula ). A etapa de aquecimento foi em 300K e no período de 0,5 ns para controlar a temperatura e ajustar a energia cinética do sistema. O cálculo da etapa de equilíbrio foi em 10 ns, sendo que nessa etapa as propriedades do sistema não são mantidas constantes e a simulação é finalizada com o equilíbrio térmico do sistema. Na obtenção da trajetória de tempo, também chamada de produção, o sistema se movimenta livremente para simular propriedades termodinâmicas e a obtenção de uma conformação de menor energia para o complexo ligante-proteína.

A trajetória de tempo foi realizada em 150 ns, pois trajetórias de DM em longas escalas de tempo também são importantes para oferecer um número maior de informações sobre o enovelamento proteico, já que trajetórias mais longas estão mais próximas da realidade.

A análise do valor de RMSD analisa a estabilidade das trajetórias e o valor de RMSF (flutuações da raiz quadrática média, do inglês, *Root Mean Square Fluctuation*) do esqueleto proteico (*backbone*) indica possíveis flutuações que ocorrem nas posições dos resíduos de aminoácido do SARS-COV-2<sup>Mpro</sup> ao interagir com o ligante.

### **Cálculos MM-PSBA e MM-GBSA**

A partir do resultado selecionado pelo valor de RMSD da simulação de DM, foi possível obter a trajetória mais estável para o complexo e com a utilização dessa trajetória foi possível estimar a energia livre por MM-PBSA e MM-GBSA.<sup>20</sup> Para os cálculos de MM-PBSA e MM-GBSA, foi gerado o arquivo de entrada (*input*) contendo as coordenadas necessárias para os cálculos de energia livre por MM-PBSA e MM-GBSA.<sup>20</sup>. Nesse arquivo são especificados os *frames* iniciais e finais dos trechos mais estáveis da trajetória obtida pela DM, assim como os resíduos de aminoácidos, o inibidor, e mascarar as moléculas de água e os dois íons sódio que neutralizam o sistema; os métodos utilizados para o cálculo da energia de ligação foram MM-GBSA e MM-PBSA; a concentração de sal considerada para os cálculos MM-GBSA foi 0,100 M; a força iônica para os cálculos MM-PBSA foi igual a 0,100 μM; a tensão superficial do sistema foi 0,00542 kcal/mol. Por fim, o cálculo foi realizado e os valores de energia livre de ligação do sistema (SARS-CoV-2<sup>Mpro</sup> + Molécula 1) foram estimados.

### **Referências Bibliográficas**

1. Berman, H. M. *et al*, **The Protein Data Bank archive as an open data resource**. *Journal of Computer-Aided Molecular Design* , n°28, 2014,p. 1009-1014.
2. Jarvis, L., **Biopharmaceuticals - Merck biologics**. *In: Chemical & Engineering News*, n° 84, 2020, p. 10-10.
3. Zhang, L. L. *et al*. **Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors**. *In: Science*, n° 368 , 2020, p. 409-412.
4. Verdonk, M. L. *et al*. **Improved protein-ligand docking using GOLD**. *In: Proteins-Structure Function and Genetics*, n°52,2003, p. 609-623.

5. Sybyl 8.1, Tripos Inc.St . Louis, USA, 2009.
6. Ferreira, R. S.; Glaucius, O.; Andricopulo, A. D., **Integração das técnicas de triagem virtual e triagem biológica automatizada em alta escala: oportunidades e desafios em P&D de fármacos.** In: *Química Nova*, n°34, 2011, p.1770-1778.
7. Daina, A.; Michielin, O.; Zoete, V., **SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules.** In: *Scientific Reports*, n°7, 2017.
8. Ji, C. G. *et al.* **eMolTox: prediction of molecular toxicity with confidence.** In: *Bioinformatics*, n°34, 2018, p. 2508-2509.
9. Lipinski, C. A., **Lead-and drug-like compounds: the rule-of-five revolution.** In: *Drug discovery today: Technologies*, n°1, 2004, p. 337-341.
10. Rodrigues, R. P. *et al.* **Estratégias de triagem virtual no planejamento de fármacos.** In: *Revista Virtual de Química*, n°4, 2012, p.739-776.
11. Pence, H. E.; Williams, A., **ChemSpider: An Online Chemical Information Resource.** In: *Journal of Chemical Education*, n°87, 2010, p. 1123-1124.
12. Namba, A. M.; da Silva, V. B.; Da Silva, C., **Dinâmica molecular: teoria e aplicações em planejamento de fármacos.** *Eclética Química*, n° 33, 2008, p.13-24.
13. Bockmann, R. A.; Gruhmuller, H., **Nanoseconds molecular dynamics simulation of primary mechanical energy transfer steps in F1-ATP synthase.** In: *Abstracts of Papers of the American Chemical Society*, n° 225, 2003, p.519-520.
14. Prigogine, I.; Rice, S. A., **Proteins: a theoretical perspective of dynamics, structure, and thermodynamics.** John Wiley & Sons: 1990; Vol. 86.
15. Stewart, J. P., **Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements.** In: *Journal of Molecular Modeling*, n° 10, 2004, p. 155-164.
16. Echenique, P.; Alonso, J. L., **A mathematical and computational review of Hartree-Fock SCF methods in quantum chemistry.** In: *Molecular Physics*, n°105, 2007, p. 3057-3098.
17. McLean, A.; Chandler, G., **Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z= 11–18.** In: *The Journal of Chemical Physics*, n° 72, 1980, p. 5639.
18. Pettersen, E. F.; *et al.* **UCSF Chimera—a visualization system for exploratory research and analysis.** In: *Journal of computational chemistry*, n° 25, 2004, p.1605-1612.
19. D.A. Case, *et al.* **AMBER 2018**, AMBER: University of California, San

Francisco, USA, 2018.

20. Genheden, S.; Ryde, U., **The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities.** *Expert Opinion on Drug Discovery*, n°10, 2015, p. 449-461.