

Virtual Screening Workshop: Parte III

Isabella Guedes¹ e Alessandro S. Nascimento²

¹ Laboratório Nacional de Computação Científica - LNCC/MCT. Av. Getúlio Vargas 333 - 1A11 - Quitandinha, Petrópolis, RJ, Brazil. 25651-075.

² Instituto de Física de São Carlos, Universidade de São Paulo. Av. Trabalhador São-Carlense, 400. Parque Arnold Schmidt. São Carlos, SP, Brazil. 13566-590.

Resumo

Nesta série de documentos, faremos um tutorial sobre a triagem virtual, passando pela preparação de um alvo, calibração do modelo e avaliação dos resultados. Na sequência, faremos a seleção de moléculas para a triagem, a triagem e a avaliação dos resultados. Estes documentos foram preparados como parte do material da [IX EMMSB 2018](#). Os arquivos gerados nesta prática também estão disponíveis no [GitHub](#).

1 Introdução

Nesta segunda parte, faremos a análise dos resultados da nossa calibração de dados do VS, usando o programa [UCSF Chimera](#) [1].

2 Análise do *Screening* com o UCSF Chimera

- Abra o programa Chimera com a estrutura do receptor e do ligante de referência: `$ chimera ../../ampc_rec.mol2.gz ../../ampc_xtallig.mol2.gz`. Mantenha a linha de comando mostrada, clicando em *Favorites -> Command Line*.
- Na linha de comando, digite `focus #1` para focar a visualização no ligante. Em seguida, `sel #1 za < 5`. Este comando seleciona tudo o que está a menos que 5 Å do ligante. Em seguida, aperte a seta para cima do teclado para expandir a seleção, incluindo todos os resíduos que têm átomos selecionados. Finalmente, `disp sel`.
- Agora vamos analisar efetivamente os resultados do docking. O Chimera tem uma ferramenta muito útil para esta finalidade: Vá em *Tools -> Surface/Binding Analysis -> View-Dock*. Na janela que se abre, selecione o arquivo `AMPC_FDA_dock.mol2.gz`. Quando perguntado sobre o tipo de arquivo, selecione o *Dock 4, 5 or 6* e clique em OK.
- O ViewDock abrirá uma janela para a visualização dos resultados e carregará a primeira molécula no visualizador. Na janela do ViewDock, clique em *Column -> Show -> Energy Score*. As energias de interação do docking serão mostradas para cada molécula que foi docada. Clicando sobre o *Energy Score*, no cabeçalho da tabela, as moléculas serão ordenadas por suas energias de interação e pode-

mos inspecionar uma a uma, bem como analisar as interações que fazem com a enzima β -lactamase.

- Vamos adicionar mais uma informação: Na janela do ViewDock, clique em *HBonds -> AddCount to Entire Receptor*. Na janela que se abre, certifique-se de marcar a opção *Inter-Model*. O chimera irá desenhar uma linha (azul, tipicamente) para cada HB que ele identifica dentro dos critérios de distância e ângulo definidos. O número de HBs também será mostrado na janela do ViewDock.
- Nosso objetivo agora é encontrar moléculas com **boa possibilidade de interação** com o nosso alvo. Como selecionar estas moléculas?
 1. Ela deve ter uma **complementaridade com o sítio ativo**. Uma forma de visualizar melhor a complementaridade é mostrar a enzima como superfície: `$ surf #0`. Podemos deixar a superfície transparente também (*Actions -> Surface -> Transparency*).
 2. Ela deve ter o **máximo de grupos polares fazendo interação com o alvo**. Grupos polares com potencial de interação não satisfeito significam interação mais favorável com o solvente e menos favorável com o alvo, tornando a interação desfavorável.
 3. Ela deve **mimetizar, pelo menos em parte, algumas das interações conhecidas do substrato natural ou de um inibidor conhecido**.

Análise as molécula que foram docadas e tente eleger 10 moléculas que melhor satisfaçam a este critério de escolha. Uma forma de selecionar as moléculas é clicar em *Purged* para as moléculas que você não deseja manter na lista final, de forma que a

lista final tem as demais como *viáveis*. As moléculas marcadas como *purged* não apagadas. Só não são mostradas mais na lista inicial. Outra informação importante, o Chimera permite salvar a *sessão*, isto é, salvar o estado em que se encontra, com moléculas carregadas e análises realizadas, para continuar o trabalho depois: *File -> Save Session As*.

As moléculas selecionadas podem ser diferentes para os diferentes alunos (selecionamos o conjunto de 300 moléculas fazendo uma seleção randômica!). Logo, não se preocupe se os seus resultados forem diferentes do resultado do seu vizinho! Usaremos as moléculas selecionadas posteriormente (Parte IV) para uma análise final do resultado do nosso *screening*.

3 Triagem Virtual Sem Receptor?

Sabemos que aproximadamente de 45% dos fármacos no mercado são endereçados a receptores de membranas [2]. Neste contexto, o desafio em se obter moléculas candidatas é aumentado em função da presença minoritária desta classe de proteínas no PDB. Apesar dos esforços recentes que renderam o [Prêmio Nobel de Química em 2012](#) para Lefkowitz e Kobilka (veja [3], para detalhes), a cristalização de proteínas de membrana ainda é um desafio em biologia estrutural. Em muitos casos, portanto, teremos um projeto, um alvo definido e nenhuma estrutura para usar na triagem virtual. O que pode ser feito em situações como esta?

Uma estratégia que pode ser empregada nestas situações é o emprego da triagem baseada na estrutura de um ligante conhecidamente bioativo. A premissa aqui é que *ligantes similares podem se ligar de modo similar a um mesmo receptor biológico*. Todo o trabalho passa a ser então: (i) como definir similaridade entre ligantes; (ii) como usar esta similaridade em triagem virtual.

Há várias ferramentas capazes de realizar este tipo de comparação (e.g., [ROCS](#), [ElectroShape](#), etc). Uma ferramenta computacional que também permite fazer este tipo de busca é o nosso já conhecido [LiBELa](#). Neste contexto, a triagem baseada na estrutura do ligante ativo oferece algumas vantagens:

1. Costuma ser mais rápida que a triagem baseada no receptor;
2. Dados retrospectivos de enriquecimento costumam revelar capacidade de enriquecimento similar ou maior que aquele observado para

os métodos que se baseiam na estrutura do receptor.

3. Dispensa a estrutura do receptor e sua preparação;

Nas seção seguinte, repetiremos a triagem da base de dados FDA usando a estratégia baseada somente no ligante de referência. Vamos ver o que podemos obter como resultado desta análise.

4 Preparando o Ambiente

4.1 Opção 1: Triando todas as moléculas FDA

- Na pasta `EMMSB_VS/VS`, crie uma nova pasta chamada `LB` (*ligand-based*): `$ mkdir LB` e entre nesta pasta.
- Vamos gerar a lista de moléculas que devem ser triadas. Desta vez, vamos usar toda a base de dados:

```
$ for i in `seq 1 2156`; do echo ../mol2/fda.$i.mol2.gz >> multimol.dat; done.
```
- Novamente, vamos gerar um arquivo de input para o `LiBELa`. As mudanças importantes aqui são:
 1. A keyword `rec_mol2` passa a ter o ligante como entrada. Ou seja, não temos receptor, só ligante. Energias errôneas serão calculadas. Mas não estamos mais interessados nelas.
 2. `use_grids no`: Não usaremos grids, uma vez que não temos receptor.
 3. `energy_optimizer none`: Não faz sentido fazer a otimização da energia de interação.
- Com os arquivos na pasta `LB`, vamos executar o `LiBELa`: `$ time McLiBELa.openMP libela.inp`. Esta triagem deve ser bem mais rápida que a triagem baseada na estrutura.

4.2 Opção 2: Triando 300 moléculas do FDA

- Na pasta `EMMSB_VS/VS`, crie uma nova pasta chamada `LB` (*ligand-based*): `$ mkdir LB` e entre nesta pasta.
- Vamos gerar a lista de moléculas que devem ser triadas. Para reduzir o tempo de triagem, vamos usar novamente uma seleção aleatória de 300 moléculas usando o script python que já usamos anteriormente: `$ picker.py`.

- Novamente, vamos gerar um arquivo de input para o LiBELa. As mudanças importantes aqui são:
 1. A keyword `rec_mol2` passa a ter o ligante como entrada. Ou seja, não temos receptor, só ligante. Energias errôneas serão calculadas. Mas não estamos mais interessados nelas.
 2. `use_grids no`: Não usaremos grids, uma vez que não temos receptor.
 3. `energy_optimizer none`: Não faz sentido fazer a otimização da energia de interação.
- Com os arquivos na pasta LB, vamos executar o LiBELa: `$ time McLiBELa.openMP libela.inp`. Esta triagem deve ser bem mais rápida que a triagem baseada na estrutura.
- Liste as moléculas em ordem decrescente de similaridade, clicando (duas vezes) sobre o cabeçalho da coluna *RMSD*.
- Inspecione as moléculas com similaridade mais elevada ao nosso ligante de referência (por exemplo, $SI > 0.75$). Lembre-se que:
 1. Na triagem, a estrutura do receptor foi ignorada. Logo, algumas moléculas apresentarão conflito estérico com o receptor.
 2. Aqui, estamos interessados nas moléculas que tenham a melhor similaridade com nosso ligante conhecido, e.g., grupos doadores e aceptores de ligação de hidrogênio em posições equivalentes, além de uma boa sobreposição de volume no espaço.
- Tente identificar uma lista dos 10 melhores candidatos identificados com esta estratégia. Lembre-se de usar a lista de compostos viáveis para separar os melhores compostos dos demais (*purged*).

5 Análise de Resultados

Novamente faremos a análise empregando o programa UCSF Chimera:

- Carregue novamente as estruturas do receptor e do ligante (apesar de não termos usado o receptor, carregaremos para comparação).
- Carregue novamente os resultados do docking (`AMPC_FDA_LB_dock.mol2.gz`) na opção *ViewDock* do Chimera: *Tools -> Surface/Binding Analysis -> ViewDock*.
- Selecione o arquivo `AMPC_FDA_LB_dock.mol2.gz` com a opção *Dock 4, 5 or 6*.
- Na Janela do *ViewDock* que será aberta, clique em *Column -> Show -> RMSD*. Embora o campo esteja marcado como RMSD, a informação mostrada nesta opção é o índice de similaridade (SI) de Hodkin [4]. Este parâmetro assume valores entre -1 a $+1$, sendo -1 uma perfeita anti-similaridade, $+1$, perfeita similaridade e 0 , total falta similaridade. No nosso experimento, estamos buscando as moléculas que tenham a maior similaridade possível (valores mais próximos de 1) em relação à nossa molécula de referência.

Referências

- [1] E F Pettersen, T D Goddard, C C Huang, G S Couch, D M Greenblatt, E C Meng, and T E Ferrin. UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25:1605–1612, 2004.
- [2] Robert Adams, Catherine L Worth, Stefan Guenther, Mathias Dunkel, Robert Lehmann, and Robert Preissner. Binding sites in membrane proteins – Diversity, druggability and prospects. *European Journal of Cell Biology*, 91(4):326–339, 2012.
- [3] Daniel M. Rosenbaum, Søren G. F. Rasmussen, and Brian K. Kobilka. The structure and function of G-protein-coupled receptors. *Nature*, 2009.
- [4] A.C. Good. The calculation of molecular similarity: alternative formulas, data manipulation and graphical display. *Journal of Molecular Graphics*, 10(3):144 – 151, 1992.

6 Apêndice: Arquivo de Input para a Triagem LB

```
# mode

mode                dock
dock_parallel       yes
parallel_jobs        1

# input files

rec_mol2             ../../ampc_xtallig.mol2.gz
lig_mol2             ../../ampc_xtallig.mol2.gz
reflig_mol2          ../../ampc_xtallig.mol2.gz
mol2_aa              no
multifile            ./multimol.dat

# force field parameters

scoring_function     3
dielectric_model      r
diel                  1.0
deltaij               0.0
deltaij_es            0.0
use_grids             no
use_delphi            no
grid_spacing          0.3
grid_box              30.0    30.0    30.0
write_grids           McGrid
solvation_alpha       0.25
solvation_beta        -0.005

# Optimization

search_box            12.0    12.0    12.0
minimization_tolerance 1.0e-6
minimization_delta    1.0e-6
dock_min_tol          1.0e-6
minimization_timeout   30
overlay_optimizer      ln_auglag
energy_optimizer       none
ignore_h              no
deal                  no
elec_scale             1.0
vdw_scale              1.0
sort_by_energy         no

# output

output_prefix         AMPC_FDA_LB
write_mol2            yes

# flexible ligands

generate_conformers   yes
number_of_conformers  10
conformers_to_rank    1
conformer_generator    GA
conformer_min_steps   1000
```