

Tutorial - LiBELa

Alessandro S. Nascimento¹

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

Resumo

Neste tutorial, abordaremos a preparação de um sistema molecular para o docking com o programa LiBELa. Os arquivos usados para este tutorial estarão disponíveis no [GitHub](#).

1 Introdução

Nesta primeira etapa, faremos a '*preparação*' do receptor, isto é, da proteína que na qual faremos o docking, e do ligante de referência, que é uma molécula que tenha sido cristalizada juntamente com a proteína e no sítio que usaremos para o docking. O programa LiBELa usa a estrutura do ligante de referência para guiar o posicionamento inicial do ligante que será docado no sítio ativo. Este posicionamento inicial é posteriormente refinado para que represente uma *pose* em um mínimo de energia de interação. Nesta etapa, usaremos o software [UCSF Chimera](#)[1].

2 Preparação do Receptor e Ligante de Referência

Nesta etapa, vamos fazer a preparação de um arquivo do receptor e um arquivo de um ligante de referência. O objetivo desta etapa é termos, ao final, um arquivo MOL2 para cada uma destas moléculas, com as devidas cargas e tipos atômicos atribuídos.

- Abra o programa UCSF Chimera. Vá em *File* -> *Fetch by ID*. Na janela que se abrirá, selecione a opção PDB, digite o ID *5HZ8* e clique em *Fetch* (Figura 1). Este é o código identificador PDB para uma estrutura da proteína FABP4 [2].

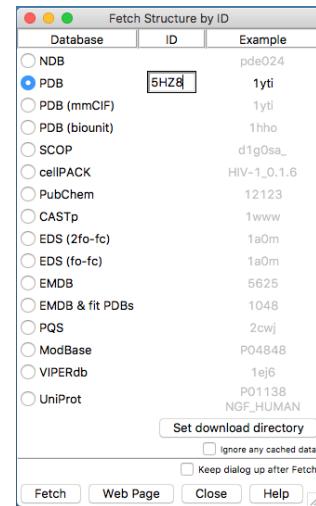


Figura 1: Usando a ferramenta *fetch by ID* no UCSF Chimera.

- A estrutura tem uma única cadeia (único monômero) da proteína FABP4 além do ligante e de moléculas do solvente (água e dimetil sulfóxido, DMSO), como mostrado na Figura 2. Nós vamos excluir o DMSO e as moléculas de solvente que estão longe do sítio ativo para reduzir o nosso *sistema receptor*.

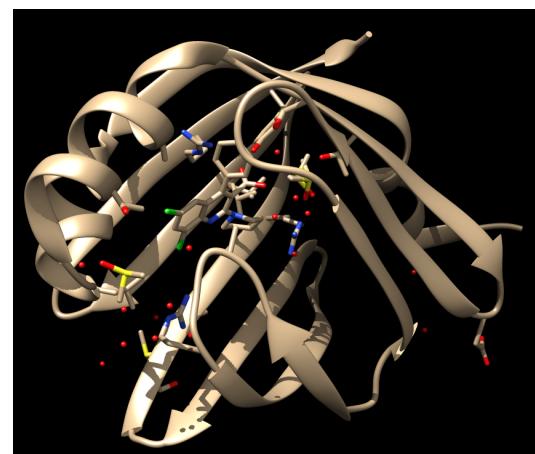


Figura 2: Estrutura da proteína FABP4 no UCSF Chimera.

- Vá até *Favorites* -> *Command Line*. Na linha de comando que aparece na parte inferior

rior da tela do Chimera, digite: `sel :DMS` e aperte ENTER. As moléculas de DMSO ficarão selecionadas. Agora digite: `del sel` para apagar estas moléculas. Outra forma de fazer a mesma coisa seria ir em *Select -> Residue -> DMS* para fazer a seleção e, depois ir em *Actions -> Atoms/Bonds -> delete*.

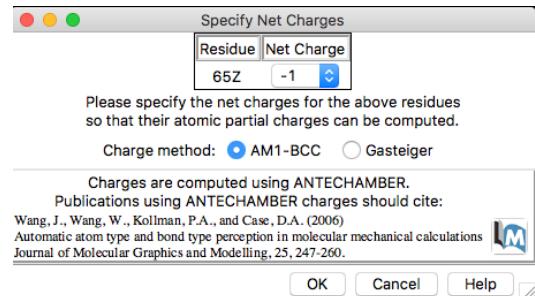


Figura 3: Cálculo das cargas atômicas para o ligante.

- Agora vamos retirar as moléculas de solvente que estão a uma distância maior que 3 Å do sítio ativo. Para isto, digite na linha de comando: `sel ligand za > 3 & :HOH` e aperte o ENTER. Algumas moléculas de água ficarão selecionadas. Embora o programa não mostre todas as moléculas de água, há 256 moléculas em uma distância superior a 3 Å do sítio ativo. Para excluir as moléculas já selecionadas, digite `del sel` e aperte o ENTER.
- No momento o nosso sistema contém a proteína FABP4, o ligante de nome *65Z* e as moléculas de solvente que fazem parte do sítio ativo. As próximas etapas são: (*i*) adicionar cadeias laterais ausentes ou incompletas, (*ii*) adicionar os átomos de hidrogênio e (*iii*) atribuir cargas atômicas para cada átomo do sistema. Todas estas tarefas podem ser feitas de uma vez com a ferramenta *Dock Prep*.
 - Vá em *Tools -> Structure Editing -> Dock Prep*. Na janela da ferramenta que será mostrada, desmarque a opção *Delete solvent*, uma vez que já deletamos as moléculas de solvente que estão fora do sítio ativo. Clique em *OK*. Ao adicionar os hidrogênios, eu usualmente selecionei a opção *Unspecified (determined by method)* para que o programa utilize os critérios do ambiente para definir os estados de protonação nas histidinas. Na atribuição de cargas, eu usualmente selecionei as cargas *AMBER ff14SB*, com cargas AM1-BCC para os resíduos não-protéticos. Neste caso, o programa percebe a presença do ligante *65Z* e sugere uma carga total -1 para este ligante. O cálculo das cargas atômicas será feito pelo programa sqm, parte do pacote AMBERTOOLS e distribuído juntamente com o Chimera. Clique em *OK* (Figura 3).
 - Salve o arquivo final como *5HZ8.mol2*. Este arquivos, em formato MOL2, contem as posições atômicas, os tipos atômicos para cada átomo do sistema (dentro dos tipos SYBYL) e as cargas atômicas para os átomos da proteína (campo de força AMBER ff14SB) e para o ligante (AM1-BCC). A próxima etapa é dividir este arquivo em dois arquivos: um contendo apenas a proteína e outro contendo apenas o ligante.
 - Vá em *Select -> Residue -> 65Z*. Após ver que o ligante foi selecionado, vá em *Select -> Invert (all models)*. Agora a proteína e o solvente estão selecionados. Vá em *Actions -> Atoms/Bonds -> delete*. Agora permaneceu apenas o ligante. Vá em *File -> Save Mol2...* e salve o arquivo como *5HZ8_lig.mol2*.
 - Agora vamos fazer o inverso: salvar apenas o receptor. Vá em *File -> Close Session*. Em seguida abra novamente o primeiro arquivo que geramos, *5HZ8.mol2*, que contem o receptor e o ligante. Com o sistema aberto, vá em *Select -> Residue -> 65Z*. Em seguida, vá em *Actions -> Atoms/Bonds -> delete*. Agora vamos salvar este novo arquivo sem o ligante. Vá em *File -> Save Mol2 ...* e salve como *5HZ8_rec.mol2*.
 - Estamos quase lá! Para conferir, vamos fechar a sessão e abrir, em seguida, os arquivos *5HZ8_rec.mol2* e o arquivo *5HZ8_lig.mol2* simultaneamente. O resultado deve ser algo similar ao mostrado na Figura 4. Agora podemos fazer uma inspeção mais detalhada das interações entre a proteína e o ligante.

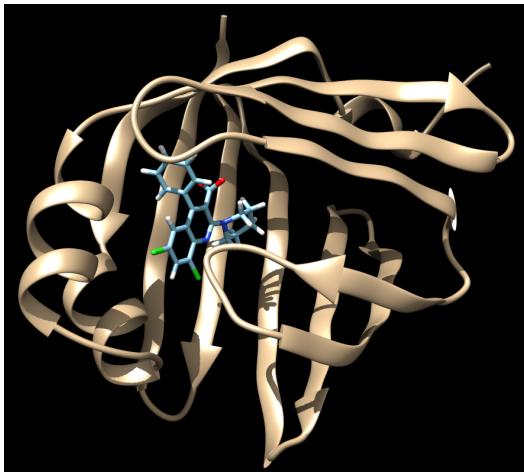


Figura 4: Estrutura da proteína FABP4 e seu ligante no UCSF Chimera.

- No meu caso, o arquivo do receptor foi aberto primeiro e recebeu o identificador 0, enquanto o ligante tem índice 1. Vamos selecionar todos os resíduos do receptor que estejam a pelo menos 3 Å do ligante. Na linha de comando, digite `sel #1 za < 3` e aperte ENTER. Em seguida aperte a tecla da seta para cima do teclado. Esta tecla *expande* a seleção para incluir todos os resíduos que tem pelo menos um átomo dentro da distância limite de 3 Å do ligante. Finalmente, vamos mostrar a seleção com `disp sel`. Para deixar a visualização menos poluída, podemos omitir os hidrogênios apolares: `~disp HC`.

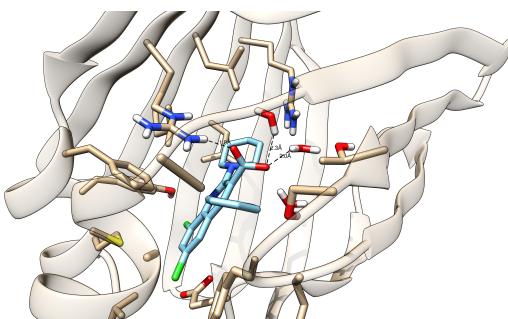


Figura 5: Interações entre o ligante de referência (65Z) e o receptor.

- Analisando as interações presentes, vemos uma interação iônica importante entre a Arg130 e o grupo carboxilato do ligante. Este mesmo grupo interage também com algumas moléculas de água no sítio ativo (Figura 5). Interessante notar também que o sítio é completamente enterrado na estrutura da proteína, isto é, não está exposto ao solvente (Figura 6).

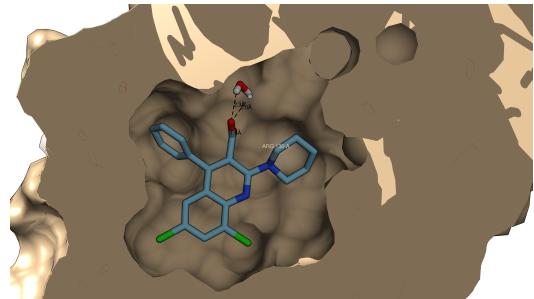


Figura 6: Receptor, mostrado em superfície, e ligante de referência. Um corte foi aplicado à superfície para visualizar o sítio ativo enterrado na estrutura da proteína.

Existem outras formas também automatizadas e até mais precisas para a *preparação do receptor*. Alguns servidores como o [H++](#) ou o [PDB2PQR](#) podem fazer isto.

3 Seleção de Compostos Que Serão Docados

Eu selecionei alguns compostos que são também ligantes da FABP4 e que se encontram em algumas estruturas do PDB. Todos os ligantes foram preparados usando a mesma estratégia da seção anterior e com o UCSF Chimera. Os ligantes são:

- [3FR2_lig](#);
- [3P6H_lig](#);
- [5D4A_lig](#);
- [5EDB_lig](#);
- [5HZ6_lig](#);
- [5HZ8_lig](#);

Os links acima devem fazer o download automático dos arquivos já preparados (hidrogênios adicionados e cargas atômicas calculadas com o método AM1-BCC). Eu tenho o hábito de manter todos os arquivos compactados (*gzipados*) para economizar espaço em disco. O [LiBELa](#) consegue ler e escrever arquivos sem precisar descompactar e o UCSF Chimera também.

4 Docking

Com todos os arquivos prontos, podemos iniciar o *docking* propriamente dito. Para isto, vamos usar o programa [LiBELa](#). No GitHub eu vou deixar instruções mais específicas sobre o download e a instalação do programa em Linux, Windows e MacOS.

- Abra o programa **iMcLiBELa**. Em ambiente Windows, ele deve ter sido adicionado ao Menu de tarefas. No Linux ou MacOs, você pode procurar o binário **iMcLiBELa** na pasta onde o programa foi instalado. Para executar, basta usar o comando `./iMcLiBELa` no terminal (assumindo que você está no diretório que contem o arquivo binário). Se você conseguiu executar o programa, você deve estar visualizando uma tela como a que está mostrada na Figura 7.

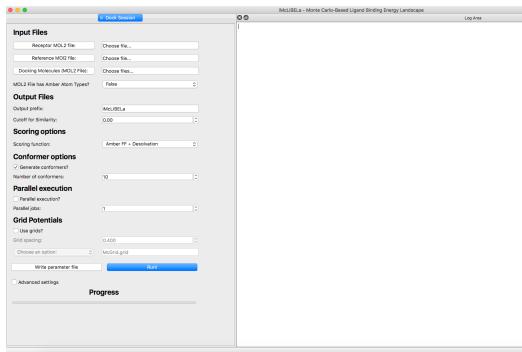


Figura 7: Tela inicial do programa **LiBELa**.

- A interface gráfica do **LiBELa** é muito simples. Na realidade, ela funciona apenas como um *formulário*, onde você define as configurações de como quer realizar o docking. A parte da direita é uma área de log, onde o programa reporta os resultados e o andamento da execução. As mesmas informações também são salvas em um arquivo de texto no diretório de onde o programa foi executado.
- Vamos iniciar selecionando os arquivos do receptor e do ligante de referência. Clique no botão *Receptor MOL2 file*:. Na janela que se abrirá, selecione o arquivo `5HZ8_rec.mol2`. Em seguida, clique em *Reference MOL2 file*: e selecione o arquivo `5HZ8_lig.mol2`. Finalmente, clique em *Docking Molecules (MOL2 File)*:. Segurando o botão *Ctrl*, selecione todos os arquivos que você baixou dos compostos que serão docados.
- Na caixa de seleção para *Scoring Function*, selecione **Amber Softcore**.
- Vá até o final do formulário e marque a opção *Advanced settings*;
- Agora vamos customizar alguns parâmetros para o nosso *docking*. Na caixa **Softcore delta for VDW term**, altere o valor padrão para 2.75.
- Na caixa **Softcore delta for electrostatic term**, altere o valor padrão para 2.50.

- Na opção *Docking timeout*, altere o valor padrão para 20 segundos. Este é o tempo máximo que o otimizador terá para encontrar o mínimo de energia durante o refinamento energético da pose.

- Na opção *Binding energy optimizer*, selecione o algoritmo **global ISRES**. Este algoritmo (*Improved Stochastic Ranking Evolution Strategy*) permite uma estratégia rápida e eficiente de busca por um mínimo global, em muitos casos [3].

- Na seção de confôrmeros, aumente o *Conformers to optimize* para 2;
- Selecione a opção de *Parallel execution* e defina na caixa ao lado o número máximo de *threads* que você quer que sejam executados paralelamente. Para um processador com 4 *threads* (caso de muitos processadores i5, por exemplo), eu sugiro usar 3 *threads*.
- Marque a opção *Use grids* e selecione um *grid spacing* de 0.30. Na caixa abaixo, selecione a opção *Write File*. O programa vai iniciar a execução calculando e salvando os *grids* em um arquivo (McGrid.grid, por default). Estes *grids* contém o potencial de interação eletrostática do receptor, o potencial de interação de van der Waals (Lennard-Jones) e potencial de dessolvatação.
- Clique em *Run!* para iniciar a execução. A tela do programa vai ficar parcialmente congelada durante a execução, mas as informações serão atualizadas na área de log.

5 Análise dos Resultados

Para a análise dos resultados, vamos retornar ao UCSF Chimera.

- No UCSF Chimera, abra a estrutura do receptor e do ligante, `5HZ8_rec.mol2` e `5HZ8_lig.mol2`. Em seguida, vá em *Tools* -> *Surface Binding Analysis* -> *ViewDock*. Na janela que se abrirá, procure no diretório em que o docking foi realizado o arquivo `iMcLiBELa_dock.mol2.gz`. Selecione o tipo de arquivo como **Dock 4, 5 or 6**.
- A janela da ferramenta *ViewDock* mostra a lista dos compostos docados. Ao clicar em cada composto, a estrutura do composto, como proposto pelo modelo do docking, é mostrada na janela principal do UCSF Chimera (Figura 8). Na janela do *ViewDock*, vá em *Column* -> *Show* -> *Energy Score*. A janela vai mostrar a energia (em kcal/mol) calculada para cada ligante que foi docado.

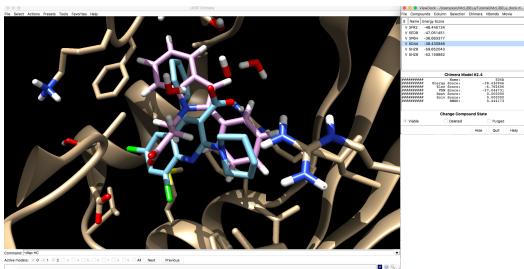


Figura 8: Resultados do Docking com a ferramenta *ViewDock* do Chimera.

- As figuras abaixo mostram os resultados para quatro dos ligantes docados.

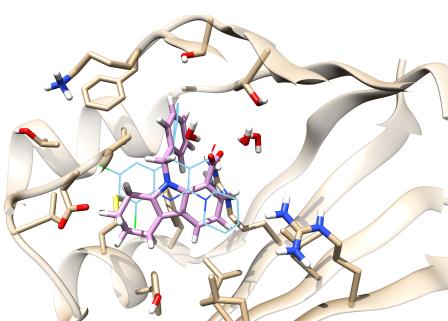


Figura 9: Resultados do Docking para os ligantes 3FR2. O composto docado está mostrado em *sticks* em rosa e o ligante de referência está mostrado em *sticks* em azul.

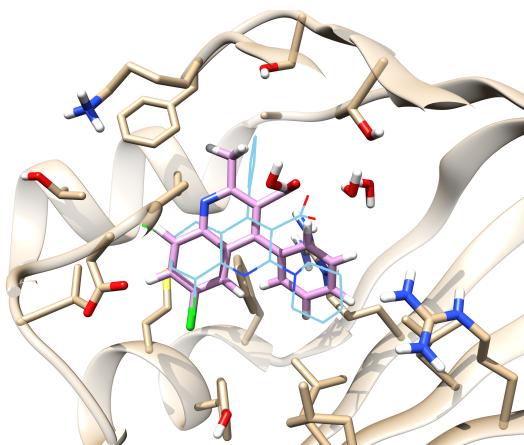


Figura 10: Resultados do Docking para o ligante 5EDB. O composto docado está mostrado em *sticks* em rosa e o ligante de referência está mostrado em *sticks* em azul.

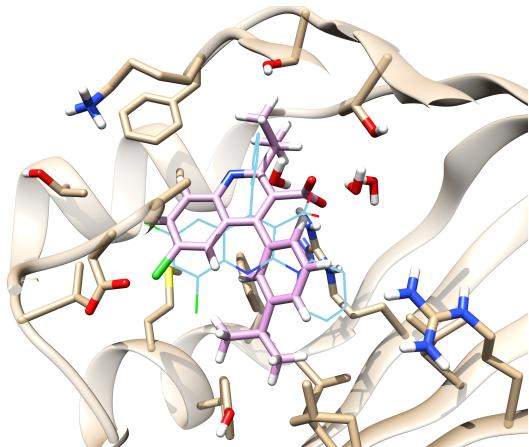


Figura 11: Resultados do Docking para o ligante 5HZ6. O composto docado está mostrado em *sticks* em rosa e o ligante de referência está mostrado em *sticks* em azul.

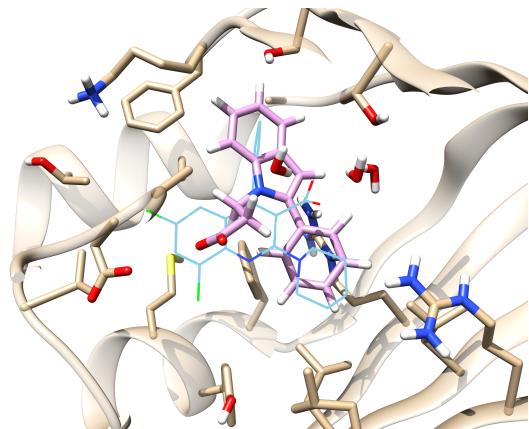


Figura 12: Resultados do Docking para o ligante 5D4A. O composto docado está mostrado em *sticks* em rosa e o ligante de referência está mostrado em *sticks* em azul.

- Os ligantes 3FR2 (Figura 9), 5EDB (Figura 10) e 5HZ6 (Figura 11) mostram boas poses de ligação, similares às poses observadas nas suas respectivas estruturas cristalográficas. Já a estrutura 5D4A (Figura 12) mostra uma conformação *flipada* ou invertida, onde o grupo carboxilato foge da Arg130 com a qual deveria interagir. Algumas das razões que podem ocasionar esta inversão incluem as moléculas de água no sítio ativo (algumas moléculas de água são observadas para alguns ligantes e não observadas para outros, podendo impedir um modo de ligação correto nestes casos); ou mesmo por falha do modelo energético em definir um mínimo claro na conformação correta. Finalmente, o algoritmo de otimização global pode falhar ao buscar exaustivamente o espaço de fase até encontrar a conformação correta. No caso particu-

lar desta proteína, aparentemente a terceira razão parece ser a mais plausível, uma vez que usando outro algoritmo de busca global (DIRECT) conseguimos um modo de ligação mais próximo do esperado.

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] Holger Kühne, Ulrike Obst-Sander, Bernd Kuhn, Aurelia Conte, Simona M. Ceccarelli, Werner Neidhart, Markus G. Rudolph, Giorgio Ottaviani, Rodolfo Gasser, Sung-Sau So, Shirley Li, Xiaolei Zhang, Lin Gao, and Michael Myers. Design and synthesis of selective, dual fatty acid binding protein 4 and 5 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 26(20):5092–5097, 2016.
- [3] T.P. Runarsson and Xin Yao. Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 35(2):233–243, 2005.