UNIVERSIDADE FEEVALE

JONAS ELIAS FLESCH

##### CRIAÇÃO DE APLICATIVOS PARA FACILITAR A EXECUÇÃO DE ENSEMBLE DOCKING

(Título Provisório)

###### Anteprojeto de Trabalho de Conclusão

Novo Hamburgo

2014

JONAS ELIAS FLESCH

CRIAÇÃO DE APLICATIVOS PARA FACILITAR A EXECUÇÃO DE ENSEMBLE DOCKING

 (Título Provisório)

Anteprojeto de Trabalho de Conclusão de Curso, apresentado como requisito parcial

à obtenção do grau de Bacharel em

Ciência da Computação pela

Universidade Feevale

Orientador: Ricardo Ferreira de Oliveira

Novo Hamburgo

2014

# RESUMO

O citocromo P450 (ou CYP) é uma das enzimas mais importantes na metabolização de medicamentos. Estes muitas vezes são ativados ao se ligarem as variações do citocromo P450. Como exemplo podemos utilizar a variante CYP2E1 do citocromo P450, que realiza a metabolização do Paracetamol. O papel da Ciência da Computação no estudo das proteínas que será abordado neste trabalho é a simulação computacional da ligação das proteínas com outros compostos. Tendo as formas 3D das proteínas pode-se por exemplo realizar simulações de como uma proteína se liga a um composto, baseando-se na sua forma, o que é conhecido como Docagem Molecular. Está é uma ferramenta importante para a descoberta de novas drogas. Esta ferramenta tenta encaixar a proteína em um ligante de várias maneiras diferentes, gerando uma energia livre para cada encaixe em uma orientação/posição diferente. Quanto menor a energia livre, melhor é a docagem. *Ensemble Docking* é uma técnica que ajuda a trabalhar as duas maiores dificuldades da docagem molecular: o sistema de pontuação e flexibilização de proteínas. Proteínas não são estruturas estáticas, nem os ligantes. A técnica de *Ensemble Docking* utiliza um corpo fixo, pois testes demonstraram que utilizar ligantes flexíveis não apresentam resultados qualitativos significantes na docagem. Para representar um receptor flexível, no Ensemble Docking são utilizadas múltiplas estruturas de proteínas. Uma das ferramentas que será desenvolvida como objetivo deste trabalho deverá atuar juntamente com o programa Gromacs e utilizar as poses geradas como parâmetro para realizar várias docagens no programa Autodock. Cada uma das poses irá gerar uma docagem com uma energia diferente, e as melhores energias deverão ser retornadas ao usuário. Desta maneira o pesquisador poderá simular a docagem de todas as poses geradas pelo Gromacs de forma automatizada. Utilizaremos como caso de aplicação docagens do citocromo CYP2E1.

Palavras-chave: Citocromo P450. Docagem Molecular. Ensemble Docking.

SUMÁRIO

MOTIVAÇÃO ...........................................................................................................................5

OBJETIVOS ..............................................................................................................................8

METODOLOGIA ......................................................................................................................9

CRONOGRAMA ....................................................................................................................10

BIBLIOGRAFIA ....................................................................................................................11

#  MOTIVAÇÃO

O citocromo P450 (ou CYP) é uma das enzimas mais importantes na metabolização de medicamentos. De acordo com Hollenberg (1992, p. 686), o sistema de enzimas do P450 catalisa o metabolismo de uma ampla variedade de compostos naturais e xenobióticos através de reações que requerem NADPH e O2, sendo também capaz de catalizar a hidroxilação dependente de peróxido de substratos na ausência de NADPH e O2. As variantes do P450 catalizam diversas reações químicas incluindo-se monooxigenação, peroxidação, redução, desalquilação, expoxidação e desalogenação (GUENGERICH, 1992, p. 667; HOLLENBERG, 1992, p. 686).

As reações gerais de monooxigenase típicas do P450 são descritas com a estequiometria que se segue (HOLLENBERG, 1992, p. 686):

$$RH+O\_{2}+NADPH+H^{+} \rightarrow ROH+H\_{2}O+NADP^{+}$$

Na reação indicada, RH é o substrato orgânico que recebe um átomo de oxigênio de uma molécula de *O*2 ao passo que o oxigênio restante forma uma molécula de água. O P450 atua em conjunto com outros dois componentes: a flavoproteína denominada NADPH-CYP450 redutase e a fosfatildilcolina que facilita a transferência de elétrons da NADPH-CYP450 redutase para o CYP (FOYE, 2008, p. 47). Uma característica digna de nota é a falta de especificidade das enzimas da família CYP para com substratos o que as torna capazes de oxidar milhares de compostos hidrofóbicos diferentes convertendo-os em compostos mais hidrofílicos facilitando sua excreção (KARP, 2008, p. 148). As isoformas individualmente são capazes de interagir com uma ampla gama de substratos químicos ocorrendo especificidades de substratos sobrepostas entre CYPs. Esta promiscuidade é vantajosa em termos de aprimoramento da defesa do organismo contra potenciais xenobióticos nocivos não obstante possa levar a efeitos indesejáveis como a rápida eliminação de fármacos e produção de compostos tóxicos (MARÉCHAL et al., 2008, p. 83).

No caso do P450, o interesse em especial é na biotransformação de medicamentos, que muitas vezes são ativados ao se ligarem as variações do citocromo P450. O P450, é considerado inclusive a família mais importante de enzimas envolvidas no metabolismo de drogas, e consequentemente no seu efeito farmacológico e toxicológico. (DE GRAAF; VERMEULEN; FEENSTRA, 2005, p. 2725-2726). Como exemplo podemos utilizar a variante CYP2E1 do citocromo P450, que realiza a metabolização do Paracetamol. Esta variante transforma o composto original do medicamento em N-acetil-p-benzoquinona, que tem é a substância que irá atuar no organismo obtendo-se os efeitos farmacológicos desejados. (LEE, 1996, p. 12063).

O papel da Ciência da Computação no estudo das proteínas que será abordado neste trabalho é a simulação computacional da ligação das proteínas com outros compostos. Com a utilização de simulações computacionais, é possível testar candidatos a medicamentos e sua capacidade de ligar-se com proteínas existentes no ser vivo, antes da realização de testes *in-vitro* e *in-vivo,* a isto se dá o nome de teste *in-silico*. A execução de simulações computacionais pode ser realizada com grande quantidade de candidatos a ligantes, agilizando a descoberta de novos farmacológicos.

Modelos de proteínas constituídas através de várias técnicas, entre elas cristalografia de raios X, basicamente uma forma de microscopia de alta resolução, e RMN, ressonância magnética nuclear. Tendo estes dados digitalizados, é possível realizar uma série de simulações computacionais sobre o comportamentos das proteínas. Tendo as formas 3D das proteínas pode-se por exemplo realizar simulações de como uma proteína se liga a um composto, baseando-se na sua forma, o que é conhecido como Docagem Molecular.

Docagem Molecular é uma ferramenta importante para a descoberta de novas drogas. Esta ferramenta tenta encaixar a proteína em um ligante de várias maneiras diferentes, gerando uma energia livre para cada encaixe em uma orientação/posição diferente (HUANG; ZOU, 2007, p. 399 apud BROOIJMANS; KUNTZ, 2003, p. 335-373). Quanto menor esta energia, mais estável é a conformação.

Uma questão importante no processo de docagem é considerar a flexibilidade não só do ligante, mas do receptor (COHEN, 2010, p. 4). Conforme Najmanovich e colaboradores (NAJMANOVICH et. al., 2000), docagens com receptores rígidos predisseram ligações incorretas para 50-70% dos casos. Por outro lado, os inúmeros graus de liberdade envolvidos na incorporação da flexibilidade do receptor durante a docagem tem posto desafios computacionais de grande monta (B-RAO et. al. 2009, p. 394). Deste modo diferentes abordagens tem sido desenvolvidas para considerar a flexibilidade conformacional do receptor levando em conta estes desafios. Algumas destas são o soft docking, o uso de bibliotecas de rotâmeros de cadeias laterais e estruturas de múltiplos receptores (COHEN, 2010, p. 6). Esta última apresenta grande complexidade computacional. Uma abordagem que considera estruturas de múltiplos receptores mas com um custo moderado é o *ensemble docking* (B-RAO et. al. 2009, p. 395).

*Ensemble Docking* é uma técnica que ajuda a trabalhar as duas maiores dificuldades da docagem molecular: o sistema de pontuação e flexibilização de proteínas. Proteínas não são estruturas estáticas, nem os ligantes. A técnica de *Ensemble Docking* utiliza um corpo fixo, pois testes demonstraram que utilizar ligantes flexíveis não apresentam resultados qualitativos significantes na docagem. Para representar um receptor flexível, no Ensemble Docking são utilizadas múltiplas estruturas de proteínas (HUANG; ZOU, 2007, p. 399-401).

Uma das suítes utilizadas para docagem molecular é a Autodock. “Ela é desenhada para prever como pequenas moléculas, como substratos ou candidatos a medicamentos, docam a um receptor com estrutura 3D conhecida” (MORRIS, 2013, tradução nossa).

Uma maneira de realizar-se Ensemble Docking é utilizar o programa Gromacs para realizar a dinâmica molecular, simulando a flexibilidade da proteína. A saída dada pelo Gromacs são várias poses, geradas através da simulação. Cada uma destas poses representa um possível estado da proteína, pois como já foi dito ela não é uma estrutura estática, e sim flexível. Para que o processo de Ensemble Docking esteja completo, esta série de poses será utilizada como entrada para o programa Autodock, que irá simular a docagem para cada uma das posições. Cada uma destas poses poderá se docar, ou encaixar, de maneira melhor ou pior com o ligante.

Uma das ferramentas que será desenvolvida deverá atuar juntamente com o programa Gromacs e utilizar as poses geradas como parâmetro para realizar várias docagens no programa Autodock. Cada uma das poses irá gerar uma docagem com uma energia diferente, e as melhores energias deverão ser retornadas ao usuário. Desta maneira o pesquisador poderá simular a docagem de todas as poses geradas pelo Gromacs de forma automatizada. Utilizaremos como caso de aplicação docagens do citocromo CYP2E1.

OBJETIVOS

Objetivo geral:

Desenvolver um conjunto de programas em Python para facilitar o processo de *Ensemble Docking*. Esta suíte de aplicativos irá realizar chamadas a programas de código fonte aberto já existentes, como por exemplo o Autodock. A suíte desenvolvida também será liberada como software livre, possibilitando melhorias por outros projetos no futuro.

Objetivos específicos:

* Pesquisar e descrever conceitos sobre bioinformática;
* Facilitar o processo de *Ensemble Docking* através de uma ferramenta genérica, que funcione para qualquer proteína e qualquer ligante;
* Esta ferramenta deverá produzir uma série de aplicativos necessária para realizar o processo de Ensemble Docking, do início ao fim, chamando a ferramenta Gromacs para realizar a dinâmica molecular, e o Autodock para realizar a docagem. O programa irá validar a versão destas ferramentas e irá exibir uma mensagem ao usuário caso seja uma versão não homologada.
* Esta ferramenta deverá considerar as melhores docagens, segundo técnicas já conhecidas baseadas na melhor energia, entre todas as poses geradas pela ferramenta que simula a dinâmica molecular;
* Analisar os resultados das docagens realizadas.

# METODOLOGIA

 A metodologia utilizada neste projeto será a pesquisa aplicada. Será realizado ou estudo sobre os conceitos e características da Docagem Molecular, utilizando material bibliográfico e Internet. Será estudada também a técnica Ensemble Docking, que será o principal foco da suíte de aplicativos. É importante um bom levantamento sobre estas técnicas e as ferramentas disponíveis para que os programas desenvolvidos utilizem corretamente os conceitos existentes e gerem o resultado previamente proposto.

 Durante a fase de estudos, será revista a linguagem de programação Python, que será utilizada para o desenvolvimento das ferramentas. Será necessário que seja realizada esta revisão para que o resultado siga os padrões de desenvolvimento da linguagem, gerando um código de fácil entendimento e que possa ser incrementado por projetos futuros. A linguagem Python é gratuita e de código aberto.

 Como resultado dos estudos realizados, será redigido o Trabalho de Conclusão I, que será entregue no segundo semestre de 2014.

 No primeiro semestre de 2015 a primeira atividade será a implementação dos programas propostos. Esta suíte de aplicativos deverá seguir as regras definidas no levantamento realizado no semestre anterior, e cobrir os objetivos específicos para os quais foi concebida. Ao final do desenvolvimento de cada programa, serão realizados testes práticos a fim de validar a solução e testar o programa.

 Ao final da implementação, deverá ser gerada uma documentação prática e sucinta que acompanhará a suíte de aplicativos, a fim de permitir que um usuário interessado em utilizá-la possa entender quais as entradas e saídas esperadas de cada programa. Esta documentação deverá especificar quais as dependências do programa, como por exemplo o Autodock, qual o seu fim, quais são os parâmetros de entrada e a saída esperada.

 Os resultados obtidos com a execução da ferramenta serão coletados e documentados no Trabalho de Conclusão II, que será entregue ao final do primeiro semestre de 2015.

# CRONOGRAMA

Trabalho de Conclusão I

|  |  |
| --- | --- |
| Etapa  | Meses |
| Ago | Set | Out | Nov |
| Definição do objetivo e orientador do projeto | X |  |  |  |
| Entrega do Anteprojeto |  | X |  |  |
| Pesquisa bibliográfica |  | X | X |  |
| Redação do TC-I |  | X | X |  |
| Entrega do TC-I |  |  |  | X |

Trabalho de Conclusão II

|  |  |
| --- | --- |
| Etapa  | Meses |
| Mar | Abr | Mai | Jun |
| Implementação da suíte de aplicativos | X | X |  |  |
| Experimentos práticos e testes |  | X |  |  |
| Realizar a documentação dos programas |  | X | X |  |
| Entrega e apresentação do TC-II |  |  |  | X |

# BIBLIOGRAFIA

|  |
| --- |
|  |

B-RAO, Chandrika; SUBRAMANIAN, Jyothi; SHARMA, Somesh D. Managing protein flexibility in docking and its applications. **Drug discovery today**, v. 14, n. 7, p. 394-400, 2009.

COHEN, Elisângela Machado Leal. Um estudo do efeito da flexibilidade explícita da enzima InhA de M. tuberculosis na docagem molecular dos inibidores etionamida, triclosano e isoniazida-pentacionoferrato II. 2010.

DE GRAAF, Chris; VERMEULEN, Nico PE; FEENSTRA, K. Anton. Cytochrome P450 in silico: an integrative modeling approach. **Journal of medicinal chemistry**, v. 48, n. 8, p. 2725-2755, 2005.

FOYE, William O. **Foye's principles of medicinal chemistry**. Lippincott Williams & Wilkins, 2008.

GUENGERICH, F. P. Cytochrome P450: advances and prospects. **The FASEB journal**, v. 6, n. 2, p. 667-668, 1992.

HOLLENBERG, P. F. Mechanisms of cytochrome P450 and peroxidase-catalyzed xenobiotic metabolism. **The FASEB journal**, v. 6, n. 2, p. 686-694, 1992.

HUANG, Sheng‐You; ZOU, Xiaoqin. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking.**Proteins: Structure, Function, and Bioinformatics**, v. 66, n. 2, p. 399-421, 2007.

KARP, Gerald; PRUITT, Nancy L. **Cell and molecular biology: concepts and experiments**. J. Wiley, 2008.

LEE, Susanna ST et al. Role of CYP2E1 in the hepatotoxicity of acetaminophen. **Journal of Biological Chemistry**, v. 271, n. 20, p. 12063-12067, 1996.

MARÉCHAL, J.‐D. et al. Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6‐drug interactions. **British journal of pharmacology**, v. 153, n. S1, p. S82-S89, 2008.

MORRIS, Garret M. **What is AutoDock?**. Disponível em: <http://autodock.scripps.edu/>. Acesso em: 01 set. 2013.

NAJMANOVICH, Rafael et al. Side‐chain flexibility in proteins upon ligand binding. **Proteins: Structure, Function, and Bioinformatics**, v. 39, n. 3, p. 261-268, 2000.