

# **SAM Database: Banco de Dados para Cadastro e Comparação de Moléculas de Plantas do Semiárido Brasileiro**

**Tiago N. Oliveira, Gécica A. T. Melo, Alex F. Santos, Bruno S. Andrade e Wagner R. A. Soares**

Centro de Pesquisa e Desenvolvimento de Software (CPDS)  
Departamento de Química e Exatas (DQE)  
Universidade Estadual do Sudoeste da Bahia (UESB)  
Rua José Moreira Sobrinho, Jequezinho – Jequié, Bahia, Brasil, CEP – 45206-190  
tn\_oliveira3@hotmail.com, gessicaatm@hotmail.com,  
afsantos@uesb.edu.br, bandrade@uesb.edu.br, wrasoares@uesb.edu.br

***Abstract.** The Brazilian semi-arid region has a great diversity of plants that have molecular compounds with some known medicinal properties. The objective of this work was to develop a database to store these molecules, as well as make comparisons between them to select those with greater pharmacological potential.*

***Resumo.** O semiárido brasileiro tem uma grande diversidade de plantas que têm compostos moleculares com algumas propriedades medicinais conhecidas. Assim, o objetivo deste trabalho foi desenvolver um banco de dados que permite armazenar essas moléculas, bem como fazer comparações entre elas para selecionar aquelas com maior potencial farmacológico.*

## **1. Introdução**

Várias plantas do semiárido brasileiro vêm sendo estudadas nos últimos anos, na tentativa de encontrar novos alvos terapêuticos para uma diversidade de patologias. No entanto, apesar de vários esforços pouco ainda se sabe quanto ao respectivo mecanismo de ação dos seus princípios ativos, ou seja, pouco se conhece onde os compostos naturais atuam no organismo e de que forma eles ocasionam seus terapêuticos efeitos colaterais, adversos e tóxicos.

A Bioinformática e a química computacional (ou quimioinformática) estão entre as áreas de maior crescimento dentro dos processos de descoberta de medicamentos [Bleicher et al 2003]. Essas ferramentas são utilizadas para triagem de moléculas in silico e tornaram-se componentes cruciais para a descoberta de muitas drogas. A seleção das melhores moléculas candidatas a interagir com seus respectivos alvos proteicos (receptores farmacológicos), ocorre de acordo com um padrão de pontuações (Hits) que obedecem a uma série de características para um encaixe perfeito do ligante-receptor [Kitchen et al 2004]. A disponibilização de bancos de dados gratuitos, na Internet, como o ZINC [Irwin e B. K. Shoichet 2005] e o SEA [Keiser et al. 2007], que possuem milhares de moléculas para pesquisa, facilita ainda mais a comparação entre estruturas

recém descobertas em bancada com outras semelhantes que já possuem alguma atividade [Keiser et al. 2007].

A criação de um banco de dados específico para o cadastro e comparação de estruturas de moléculas isoladas de plantas do semiárido brasileiro é uma forma de organizar esses compostos em grupos, de acordo com o seu provável uso medicinal e seus alvos moleculares, facilitando o desenvolvimento de novos fármacos para o tratamento de diversos tipos de doenças. Além disso, esse é o primeiro banco de dados destinado especificamente ao uso de moléculas isoladas de plantas brasileiras obtidas da biodiversidade de um de seus biomas e muitas dessas moléculas não são encontradas em nenhum outro banco de dados no mundo, e diferente das moléculas contidas nos outros banco de dados, a maioria delas não possuem atividade biológica conhecida, dificultando o estudo sobre elas. Assim o desenvolvimento desse trabalho irá agregar mais possibilidades de estudos para os pesquisadores que tem interesse em encontrar moléculas com potencial farmacológico, provenientes do semiárido brasileiro.

## 2. Metodologia

A iniciativa partiu de uma proposta interdisciplinar na qual a participação de docentes e discentes dos cursos de Sistemas de informação, Biologia e Farmácia, da Universidade XXX foram cruciais. A equipe agregou os conhecimentos, na idealização e construção de um software para armazenamento e comparação de moléculas isoladas de plantas do semiárido brasileiro, servindo de ferramenta para pesquisas e seleção de compostos químicos candidatos, no tratamento de patologias humanas.

O software atualmente possui as seguintes funcionalidades:

- Cadastro de moléculas e comparação entre uma molécula contida no banco de dados com as outras armazenadas e retorna o grau de similaridade de cada uma em relação à molécula comparada.
- Comparação entre uma molécula contida no banco de dados com as contidas no ChEMBL (ChEMBL-og 2014), um dos principais banco de dados públicos de moléculas com propriedades medicinais conhecidas, mantido pelo Instituto Europeu de Bioinformática (EBI) que disponibiliza web services para integração com outros sistemas.
- Comparação de moléculas que não estejam no banco dados com as armazenadas localmente ou no ChEMBL (ChEMBL-og 2014).
- Classificação das moléculas de acordo com a regra dos cinco de lipinski, a qual é utilizada na avaliação a biodisponibilidade oral teórica de um composto. Essa regra foi definida através de análises feitas com moléculas de um banco de dados com mais de 50.000 substâncias. A regra determina que uma melhor absorção oral está associada a compostos com massa molecular (M.M) menor ou igual a 500, o coeficiente de participação calculado (cLogP) menor ou igual a 5 números de átomos doadores de ligação de hidrogênio (HDB) menor ou igual a 5 e número de átomos aceptores de ligação de hidrogênio (HAB) menor ou igual a 10. Assim as moléculas que obedecem a essa regra estão aptas a serem candidatas a novos fármacos [Gomes et al 2014].

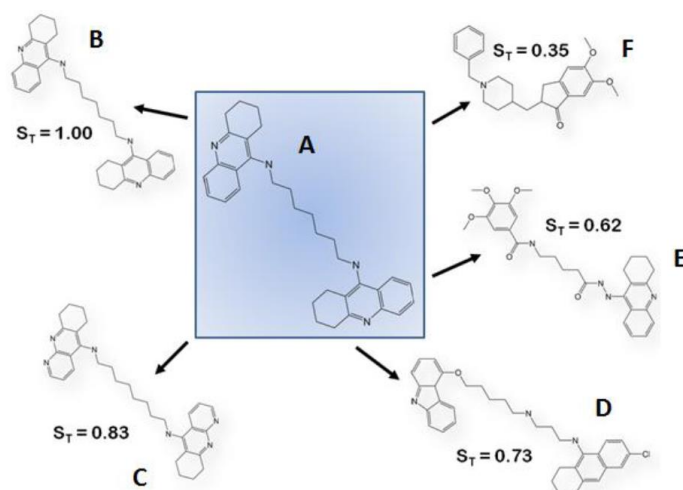
- Pretende-se disponibilizá-lo a comunidade científica da área, assim que seu desenvolvimento estiver completo.

Para o desenvolvimento deste software, foi escolhida a linguagem de programação Java em seu padrão EE (*EnterpriseEdition*) [Deitel 2006]. Além de usar os frameworks JSF e Primefaces, os quais juntos permitem desenvolver aplicações web com interfaces de fácil utilização.

A comparação de moléculas é feita através de técnicas de triagem virtual que representa um dos maiores avanços na área de planejamento de fármacos. Neste trabalho a técnica utilizada é a triagem virtual baseada em ligantes que possui algumas abordagens, ela utiliza moléculas que possuem uma atividade farmacológica conhecida como um alvo inicial para comparação, assim moléculas que possuem uma semelhança estrutural próxima à molécula-alvo tem grande chances de compartilhar atividades farmacológicas semelhantes a ela [Rodrigues et al 2012].

A abordagem da triagem virtual baseada em ligantes utilizada no software é a triagem por similaridade, a busca por similaridade consiste em uma das estratégias de triagem mais simples e com menor gasto computacional, existem outras abordagens mais complexas que trazem resultados com mais precisão, mas requerem maior gasto computacional e são usados em casos mais específicos que não são abordados no trabalho proposto. Na aplicação desta estratégia, como em qualquer busca baseada em ligantes, o ponto de partida corresponde à seleção de uma ou mais moléculas com atividade farmacológica conhecida (denominadas estruturas-alvo ou de referência), assim a lista de moléculas as quais ainda não possuem nenhuma atividade farmacológica, resultante da triagem indicam as mais prováveis de apresentarem características semelhantes à estrutura alvo, tornando-as candidatas a possíveis testes biológicos para verificar se algumas delas terão o mesmo efeito sobre o organismo humano tanto quanto a estruturas-alvo usada na comparação [Rodrigues et al 2012].

O padrão de medida ou coeficiente de similaridade empregado nas comparações é o coeficiente de tanimoto muito usado em técnicas de triagem virtual, ele fornece valores de comparação na faixa de 0 (zero) a 1. O valor 0 significa que não existe similaridade entre as moléculas comparadas, já o valor 1 demonstra 100% de similaridade. A Figura 1 mostra a comparação de similaridade entre uma molécula-alvo e outras estruturas utilizando o coeficiente de tanimoto [Rodrigues et al 2012].



**Figura 1. . Comparação de similaridade entre uma molécula-alvo e outras estruturas utilizando o coeficiente de tanimoto.**

Para realização do cálculo da similaridade entre moléculas e geração do arquivo mol, utilizamos uma biblioteca Open Source índigo desenvolvida pela empresa GGA Softwares Services [GGA Softwares Services 2013]. Essa biblioteca é assim como outras existentes são muito utilizadas em softwares voltados para bioinformática.

Os dados gerados são armazenados em um banco de dados relacional, utilizando-se como Sistema Gerenciador de Banco de Dados (SGBD) o MYSQL Server 5.5.280 [Oracle Corporation 2013].

### 3. Resultados e Discussões

O software foi desenvolvido com intuito de proporcionar aos pesquisadores o armazenamento das informações das moléculas do semiárido brasileiro, podendo assim consultá-las para fins acadêmicos e científicos, além de proporcionar a capacidade de comparações de moléculas na qual o pesquisador pode inserir sua molécula recém-descoberta ou em estudo e comparar com as existentes no banco de dados permitindo que encontre a similaridade entre elas. Desta forma, a estratégia é descobrir possíveis finalidades para o uso desta molécula com base nas suas similares que já possuem atividade.

A Figura 2 exibe o formulário de cadastro de moléculas. Este formulário permite ao usuário inserir os dados básicos da molécula como nome, a planta da qual foi isolada, seu composto, publicações relacionadas entre outras informações. A Figura 3 ilustra as moléculas cadastradas e a possibilidade de compara-las tanto localmente com outras já cadastradas, quanto no ChEMBL (ChEMBL-og 2014).

**Home**

- register Molecule
- molecules Members
- find Molecules

**Compound Name and Classification**

Name:

synonymous:

isolated:

**Representative of Compound**

Canonical Smile:

**Activity Compound**

inhibition:

essays:

**Target Compound**

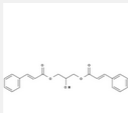
- MEMBRANE RECEPTORS
- ENZYME
- TRANSCRIPTION FACTOR
- CARRIER
- SECRETADO
- ION CHANNEL
- OTHER
- NOT CLASSIFIED

**Related Publications**

**Figura 2. Formulário de Cadastro de Moléculas**

(1 of 3) 1 2 3 12

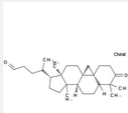
Nome:  
1,3-di-o-cinamoil-glicerol



Delete Edit Download File Mol Seek Chembl Seek SAM

---

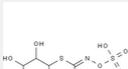
Nome:  
25,26,27-trinor-3-oxocicloartan-24-al



Delete Edit Download File Mol Seek Chembl Seek SAM

---

Nome:  
3-Hidroxiubutil-gliconinsolato



**Figura 3. Tela de Consulta das Moléculas Cadastradas**

Na Figura 4 podemos ver as comparações entre as moléculas. Nesta tela, o usuário pode pesquisar inserindo o canonical *smile* de uma molécula no caso de ter a curiosidade de comparar uma molécula qualquer que não faça parte do banco de dados e assim pesquisar tanto localmente quanto no ChEMBL (ChEMBL-og 2014). A Figura 5 e a Figura 6 exibem os resultado de uma consulta no próprio banco de dados e no ChEMBL (ChEMBL-og 2014), respectivamente.

Registered number of molecules- 28

**Home**

register Molecule

molecules Members

find Molecules

seek

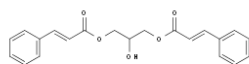
OC(COC(=O)C=Cc1ccccc1)COC(=O)C=Cc1ccccc1

type search :

Smile  Name

(1 of 1) <span style="float: right;">5</span>		
Name	Similarity	Molecule
No records found.		
(1 of 1) <span style="float: right;">5</span>		

Figura 4. Tela de comparação de moléculas.



(1 of 3) 2 3 5

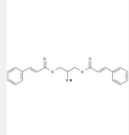
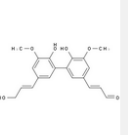
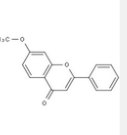

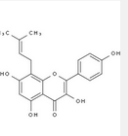
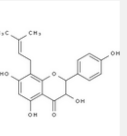
<p style="text-align: center; background-color: #333; color: white; font-weight: bold;">1,3-di-o-cinamoil-glicerol</p> <div style="text-align: center;">  <p>1.0</p> <p><a href="#">p</a></p> </div>	<p style="text-align: center; background-color: #333; color: white; font-weight: bold;">Zeihherol</p> <div style="text-align: center;">  <p>0.4402332305908203</p> <p><a href="#">p</a></p> </div>	<p style="text-align: center; background-color: #333; color: white; font-weight: bold;">7-metoxi flavona</p> <div style="text-align: center;">  <p>0.41863515973091125</p> <p><a href="#">p</a></p> </div>
<p style="text-align: center; background-color: #333; color: white; font-weight: bold;">spermina</p> <div style="text-align: center;">  <p>0.3844984769821167</p> <p><a href="#">p</a></p> </div>	<p style="text-align: center; background-color: #333; color: white; font-weight: bold;">flavona-3-4',5,7-tetrahidroxil-8-prenila</p> <div style="text-align: center;">  <p>0.37919461727142334</p> <p><a href="#">p</a></p> </div>	<p style="text-align: center; background-color: #333; color: white; font-weight: bold;">flavanona-3-ol-4',5,7-tetrahidroxil-8-prenila</p> <div style="text-align: center;">  <p>0.37217193841934204</p> <p><a href="#">p</a></p> </div>

Figura 5. Resultado da comparação no SAM

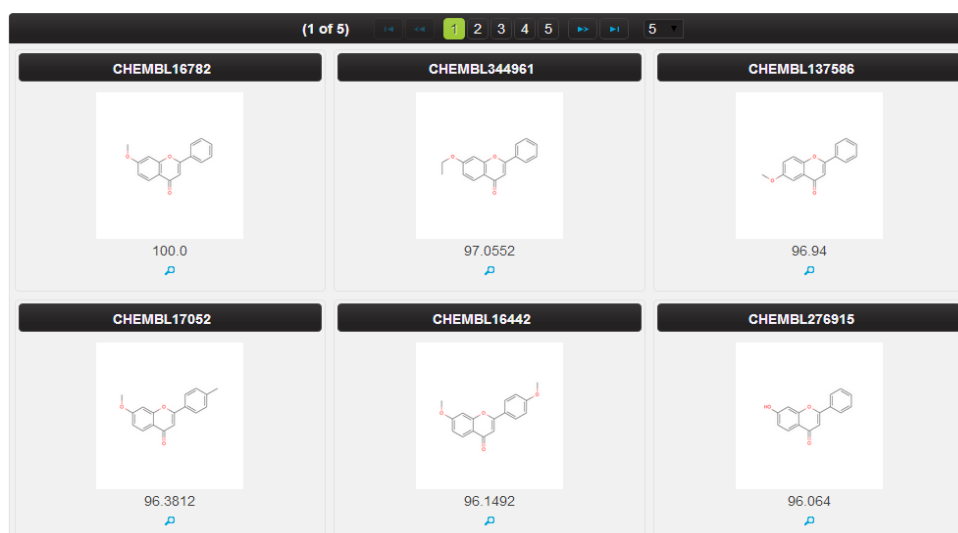
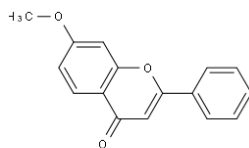


Figura 6. Resultado da comparação no Chembl

#### 4. Conclusões

O crescimento dos bancos de dados informacionais tais como: ZINC [Irwin e B. K. Shoichet 2005] e SEA (Keiser et al. 2007) tem impulsionado o desenvolvimento de ferramentas computacionais com aplicabilidade no direcionamento de ensaios biológicos in vivo e in vitro (novas moléculas isoladas e caracterizadas). Desta forma, a integração destas áreas leva a uma maior agilidade da busca por compostos com potencial farmacológico.

A metodologia empregada destaca-se como um mecanismo de triagem virtual em larga escala, na qual é possível visualizar de forma mais rápida novas oportunidades no que se refere à prospecção in silico de moléculas com ação terapêutica. Percebe-se, que o avanço na incidência e prevalência de doenças, não caminha na mesma velocidade da descoberta e desenvolvimento de novos medicamentos. Apesar da vasta biodiversidade dos biomas brasileiros, a quantidade de estudos para alguns compostos promissores ainda é escassa, portanto a bioinformática pode tornar mais eficiente, o processo de descoberta e desenvolvimento de novos compostos com poder terapêutico.

Sendo assim, a ferramenta SAM Database pode vir a ser de grande auxílio aos pesquisadores que desejam direcionar suas pesquisas em compostos químicos encontrados no semiárido brasileiro, criando uma network para transmissão de conhecimento dentro da química medicinal e farmacologia de produtos naturais.

## 5. Referências

Bleicher, K. H., et al. (2003) "Hit and lead generation: beyond high-throughput screening." *Nat Rev Drug Discov* 2(5): 369-78.

Kitchen, D. B., et al. (2004) "Docking and scoring in virtual screening for drug discovery: methods and applications." *Nat Rev Drug Discov* 3(11): 935-949.

Irwin, J. J. e Shoichet, B. K. (2005) "ZINC-a free database of commercially available compounds for virtual screening." *J Chem Inf Model* 45(1): 177-82.

Keiser, M. J., et al. (2007) "Relating protein pharmacology by ligand chemistry." *Nat Biotechnol* 25(2): 197-206.

Deitel, P. (2006) "Java How to Program", Prentice Hall, 7th edition.

GGA Softwares Services. Indigo Toolkit [internet] 2013 [acesso em 2014 Fev 26]. Disponível em: <http://ggasoftware.com/opensource/indigo>

Oracle Corporation. MySQL 5.5 Reference Manual [internet]. c1997-2013 [acesso em 2014 Fev 26]. Disponível em: <http://dev.mysql.com/doc/relnotes/mysql/5.5/en/index.html>

Rodrigues, R. P. et al. (2012). Estratégias de Triagem Virtual no Planejamento de Fármacos [internet]. 2014 [acesso em 2014 Fev 26]. Disponível em: <http://www.uff.br/RVQ/index.php/rvq/article/viewFile/295/290>.

ChEMBL-og. ChEMBL\_10 Released [internet]. 2014 [acesso em 2014 Fev 26]. Disponível em: <http://chembl.blogspot.com.br/2011/06/chembl-10-released.html>.

Gomes, R. P. et al. (2014). SÍNTESE E AVALIAÇÃO DA ATIVIDADE ANTIMALÁRICA DE COMPOSTOS DERIVADOS DA CURCUMINA [internet]. 2014 [acesso em 2014 Fev 26]. Disponível em: <http://quimicanova.s bq.org.br/qn/No%20Prelo/Artigos/AR13648.pdf>