

X ENCONTRO ACADÊMICO

MODELAGEM COMPUTACIONAL

AVALIAÇÃO DE FUNÇÕES EMPÍRICAS DE *SCORING* EM EXPERIMENTOS DE TRIAGEM VIRTUAL EM LARGA ESCALA

Felipe S. S. Pereira

Isabella A. Guedes

Laurent E. Dardenne

felipesp@lncc.br

isabella@lncc.br

dardenne@lncc.br

Grupo de Modelagem Molecular de Sistemas Biológicos, Laboratório Nacional de Computação Científica, LNCC/MCTIC

Av. Getúlio Vargas, 333, Quitandinha, 25651-075, Rio de Janeiro, Petrópolis, Brasil

Resumo. Considerando os promissores resultados na predição de modos de ligação proteína-ligante previamente obtidos pelo programa DockThor, objetiva-se no presente momento o desenvolvimento de novas funções empíricas de scoring para predição da afinidade de ligação proteína-ligante para estudos de triagem virtual em larga escala. Objetivamos neste trabalho, avaliar a preditividade e a classificação de ligantes das funções de scoring recentemente desenvolvidas em nosso grupo de pesquisa. Foram analisadas três proteínas-alvo selecionadas a partir do DUD-E, pertencentes a distintas classes proteicas, associadas a conjuntos de ligantes ativos e decoys (i.e. ligantes de menor afinidade que ativos). Buscamos analisar os resultados das predições utilizando diferentes funções empíricas: (1) genérica (i.e., sem especificidade quanto à classe de proteína ou resultados de docking); (2) classe-proteína específica e (3) derivada de resultados de docking obtidos com o programa DockThor. Os valores das áreas sob as curvas ROC obtidos nas análises (maiores que 0,6) demonstram o potencial das funções e nos guiam para o aprimoramento das mesmas, objetivando a obtenção de um classificador confiável para experimentos de triagem virtual em busca da correta distinção de ligantes a partir das afinidades de ligação preditas.

Palavras-chave: Triagem virtual, Funções empíricas de afinidade receptor-ligante, docking

1 INTRODUÇÃO

A metodologia de *docking* molecular consiste basicamente de: (1) algoritmo de busca, que tem como finalidade a exploração da superfície de energia livre de um complexo receptor-ligante almejando encontrar os modos de ligação mais favoráveis, e (2) funções de *scoring* que objetivam realizar a predição da afinidade de ligação, tanto para ranquear os ligantes de acordo com a afinidade predita quanto classificá-los como ativos ou inativos (Guedes et al., 2014). Dado um conjunto de ligantes, a correta distinção entre ativos e inativos é questão crucial na busca por compostos químicos de interesse farmacológico.

O programa DockThor mostra-se competitivo em estudos comparativos com distintos programas de atracamento molecular, obtendo resultados promissores no que se refere à predição de modos de ligação (de Magalhães et al., 2014). Recentemente, foram desenvolvidas em nosso grupo de pesquisa novas funções empíricas de *scoring* para predição da afinidade de ligação proteína-ligante, baseadas na função de *scoring* original do DockThor (i.e. campo de força MMFF94S) (Guedes, 2016). Essas funções empíricas buscam predizer a energia livre de ligação a partir de uma soma ponderada de termos baseados em propriedades físicas, onde os coeficientes associados aos termos são obtidos através de regressão linear, usando como referência dados de afinidade previamente elucidados experimentalmente (Liu & Wang, 2015).

Experimentos computacionais que abordam um grande número de ligantes (e.g. bancos de dados de compostos químicos) com o objetivo de selecionar ligantes potencialmente ativos no âmbito farmacológico, são denominados experimentos de triagem virtual (do inglês, *virtual screening*), complementares aos estudos experimentais de triagem de novos fármacos. Para validação de ferramentas que atacam o problema supracitado, faz-se necessária a consideração do cenário real de estudos de triagem virtual no planejamento racional de fármacos, onde nos deparamos com uma enorme quantidade de compostos químicos, sendo estes em sua maioria indesejados para um dado estudo (i.e. apresentam baixa afinidade de ligação, logo denominados inativos). Para essa validação são disponibilizados conjuntos de dados que buscam reproduzir esse cenário real, i.e. bancos de dados constituídos por um grande número de ligantes associados a uma lista de proteínas-alvo com importância farmacológica. Neste sentido, é de suma importância avaliar a capacidade de nossas funções empíricas com relação a sua confiabilidade na classificação dos compostos químicos (i.e. ligantes) em ativos ou não.

2 OBJETIVOS

O presente trabalho objetivou realizar a avaliação de funções empíricas de *scoring* receptor-ligante em estudos de triagem virtual de ligantes em larga escala. O objetivo principal é avaliar a distinção entre ativos e *decoys* dada pelo ranqueamento dos ligantes através das funções de *scoring*, desde que a correta classificação dos ligantes é necessária para o êxito em experimentos de triagem virtual de fármacos. Desde que funções de *scoring* podem apresentar comportamento variável dada a classe da proteína em estudo, buscamos analisar os resultados utilizando diferentes funções empíricas: (1) genérica (i.e., sem especificidade quanto à classe de proteína ou programa de *docking*); (2) classe-proteína específica e (3) específica para resultados de *docking* obtidos com o programa DockThor. A avaliação das funções empíricas se dá em um contexto onde se considera grandes conjuntos de ligantes, e a consequente classificação dos mesmos em ativos ou *decoys*.

3 METODOLOGIA

Foi selecionado o DUD-E (*Directory of Useful Decoys: Enhanced*) como conjunto de dados para o *benchmarking* dos experimentos de triagem virtual. O conjunto supracitado constitui-se de uma versão reestruturada do DUD (*Directory of Useful Decoys*), publicada no ano de 2006 (Huang et al., 2006). Nesta reestruturação foi desenvolvida uma base de dados com novas funcionalidades e disponibilizada como ferramenta para avaliação e otimização de programas de *docking* molecular, bem como de funções de *scoring*. O DUD-E é constituído por ligantes que possuem dados de afinidade experimentais previamente recuperados a partir da literatura e *decoys*. A classificação adotada para os ligantes, compreende a denominação de ligantes ativos aqueles que possuem valor de afinidade de ligação obtida experimentalmente abaixo de $1\mu\text{M}$, ligantes inativos aqueles que apresentam afinidade de ligação experimental acima de $30\mu\text{M}$, ligantes marginais aqueles que apresentam afinidade de ligação experimental entre $1\mu\text{M}$ e $30\mu\text{M}$ e *decoys*, ligantes que possuem propriedades físico-químicas similares aos ligantes ativos, porém são topologicamente dissimilares. A adoção da estratégia citada para a adição de *decoys* não experimentais (i.e. sem dados de inibição conhecidos) no conjunto, ocorre devido ao pequeno número de ligantes inativos reportados (i.e. *decoys* experimentais) (Mysinger et al., 2012).

Selecionamos proteínas-alvo do DUD-E representantes de três classes proteicas distintas descritas na literatura com potencial farmacológico : (1) Quinase Dependente de Ciclina 2 (CDK2) pertencente a classe das quinases; (2) Receptor Beta-1 de Hormônio Tireoidiano (THB) pertencente a classe de receptores nucleares e (3) Tripsina (TRY1) pertencente a classe das proteases. Foram utilizadas as estruturas dos respectivos alvos descritas no DUD-E, obtidas a partir do *Protein Data Bank*. Encontra-se na Tabela 1 os identificadores das estruturas e os números correspondentes aos ligantes e *decoys* utilizados em cada caso. Devido à pequena quantidade de ligantes inativos (i.e. ligantes com dados de afinidade experimentais reportados na literatura) relacionados aos alvos selecionados para o presente trabalho, apenas o conjunto de *decoys* foi utilizado nos experimentos de triagem virtual. A pequena quantidade de ligantes inativos disponíveis corrobora que dados de afinidade negativos raramente são reportados, como descrito pelos autores do DUD-E (Mysinger et al., 2012).

Tabela 1: Estruturas proteicas e as respectivas quantidades de ligantes

Proteína-Alvo	Código PDB	Ativos	<i>Decoys</i>
CDK2 - Quinase Dependente de Ciclina 2	1H00	443	27.830
THB - Receptor Beta-1 de Hormônio Tireoidiano	1Q4X	40	7.652
TRY1 - Tripsina	2AYW	449	25.367

Os experimentos de *docking* para predição do modo de ligação dos ligantes selecionados foram obtidos através do programa DockThor, que utiliza a abordagem de *docking* ligante flexível/receptor rígido, onde todos os graus de liberdade do ligante (translacionais, rotacionais e conformacionais) são explorados e a proteína é considerada rígida. Para encontrar a solução ótima dentre as possíveis conformações de um ligante junto ao sítio ativo de uma proteína-alvo, é utilizado um algoritmo genético (AG) de múltiplos mínimos como método de busca. A função de avaliação de energia é baseada no campo de força clássico MMFF94 (Halgren, 1999).

O DockThor utiliza a metodologia baseada em grade, onde o cálculo da função de energia é realizado a partir de uma malha construída de acordo com o centro do sítio ativo da proteína-alvo de interesse (de Magalhães et al., 2014). Dada a característica do AG do DockThor gerar múltiplas soluções para os modos de ligação dos ligantes, é realizada um agrupamento e ranqueamento para a seleção do modo de ligação de melhor energia observado para cada ligante. Os parâmetros adotados para a execução do programa DockThor consistiram em 24 rodadas do AG, tamanho da população inicial equivalente a 1.000, número de avaliações do AG equivalente a 1.000.000. As dimensões da grade foram definidas em 20 Å para cada coordenada (x, y, z), com discretização de 0,25 Å. Com exceção do número de rodadas do AG, todos os demais parâmetros supracitados são os valores padrões atuais do programa. Experimentos de *redocking* foram realizados a fim de se validar o protocolo de *docking* adotado para o estudo, em cada proteína-alvo. Avaliou-se a partir de um alinhamento estrutural entre a conformação experimental do ligante e a pose predita pelo *docking*, a similaridade entre os modos de ligação utilizando a métrica do RMSD (*Root Mean Square Deviation*), adotando um limiar de 2,0 Å.

Os cálculos do DockThor foram executados no supercomputador Santos Dumont, sistema de computação petaflopica localizado no Laboratório Nacional de Computação Científica (LNCC), nó central (Tier-0) do Sistema Nacional de Processamento de Alto Desempenho (SINAPAD). No presente trabalho, foram realizadas 61.781 execuções do algoritmo para o cálculo dos *dockings* moleculares. Cada execução do algoritmo da versão atual do DockThor leva em média 5 minutos, obtemos 308.405 minutos de cpu/core, equivalente a aproximadamente 214 dias de cpu/core. Estes valores justificam a necessidade de um ambiente de computação de alto desempenho em estudos de triagem virtual de ligantes em larga escala.

A predição das afinidades de ligação foram realizadas utilizando três funções empíricas de *scoring* desenvolvidas em trabalho prévio. As funções utilizadas neste trabalho são classificadas em (1) genérica, (2) classe-específica e (3) genérica derivada de resultados de *docking*. A função de *scoring* genérica foi desenvolvida a partir de treinamento utilizando o conjunto de dados PDBbind v2013, um conjunto de dados composto por estruturas de complexos proteína-ligante de alta qualidade com dados de afinidade de ligação experimental associados, disponível para desenvolvimento e validação de metodologias aplicadas ao *drug design* (Li, 2014). As funções classe-específicas foram obtidas a partir de sub-conjuntos do PDBbind. O desenvolvimento das funções específicas foi realizado a partir do treinamento considerando conjuntos de complexos proteicos que compartilham propriedades em comum. As funções derivadas de *docking* foram desenvolvidas a partir de treinamento utilizando resultados de *redocking* dos complexos do PDBbind, obtidos pelo programa DockThor, visando considerar as conformações proteína-ligante resultantes de *docking* ao invés das conformações experimentais (Guedes, 2016). Os valores de afinidade proteína-ligante foram obtidos utilizando as funções genéricas, genérica derivada de *docking* e classe-proteína específicas, quando da predição dos valores de afinidade do alvo CDK2 (utilizando função empírica quinase-específica) e do alvo TRY (utilizando função empírica protease-específica).

Todos os ligantes ativos e *decoys* atracados às respectivas proteínas-alvo correspondentes, foram ranqueados de acordo com as funções de *scoring* e o desempenho das funções foi mensurado através de curvas *Receiver Operating Characteristic* (ROC), que considera a fração de verdadeiros positivos no eixo Y (Sensibilidade, Eq. (1)) contra a fração de falsos positivos no eixo X (Especificidade, Eq. (2)) de acordo com os *scores* preditos. Realizamos a análise comparativa da *area under the curve* (AUC), métrica comumente aplicada na avaliação de modelos

de classificação binária, sendo neste caso a classificação entre ativos e *decoys* baseada em valores de *score*. No cenário ideal visualiza-se um valor AUC-ROC igual a 1,0 i.e. o máximo valor para AUC, onde obtemos uma perfeita separação de ligantes ativos e *decoys*, considerando que, o conjunto de *decoys* é composto por ligantes estruturalmente dissimilares dos ligantes ativos, ou seja, ligantes inativos sem dados de afinidade experimentais relacionados e desfavoráveis a ligação. Valor AUC-ROC equivalente a 0,5 conceitua uma discriminação aleatória dos ligantes e valores abaixo de 0,5 implica em uma metodologia que fornece um ranqueamento melhor para *decoys* do que para ativos, um cenário indesejado (Hawkins, 2008; Kirchmair, 2008).

$$Se = \frac{N \text{ ativos selecionados}}{N \text{ total de ativos}} = \frac{TP}{TP + FN} \quad (1)$$

$$Sp = \frac{N \text{ inativos descartados}}{N \text{ total de inativos}} = \frac{TN}{TN + FP} \quad (2)$$

Os programas DockThor, Dstatistic e as Funções Empíricas de *Scoring* foram desenvolvidos pelo GMMSB/LNCC (de Magalhães et al., 2014; Almeida, 2011; Guedes, 2016).

4 RESULTADOS E DISCUSSÃO

Os resultados dos estudos de *redocking* validaram a metodologia de *docking*, exibindo resultados satisfatórios que demonstram a correta reprodução dos modos de ligação encontrados pelo programa DockThor (Tabela 2). Foram encontrados, para os três sistemas estudados, valores de RMSD abaixo de 2,0 Å quando comparados os modos de ligação preditos e experimentais (Figura 1). Os experimentos de *redocking* são aplicados aos complexos experimentais das proteínas-alvo obtidas do PDB, de forma que validamos a preparação das três estruturas representativas de cada alvo para a triagem virtual em larga escala.

Tabela 2: Valores de RMSD dos estudos de *Redocking*

Complexo Estrutural	RMSD
CDK2 (PDB 1H00)	1,954
THB (PDB 1Q4X)	1,380
TRY1 (PDB 2AYW)	0,912

Os experimentos de *docking* em larga escala executados pelo DockThor resultaram na obtenção da pose (conformação) de menor valor de energia para cada ligante dos conjuntos, i.e. o modo de ligação predito do ligante no sítio ativo da respectiva proteína-alvo que tende a ser o encontrado na natureza. As melhores poses (complexos proteína-ligante) de cada ligante foram utilizadas para a predição dos valores de afinidade de ligação utilizando a função de *scoring* genérica para os três alvos, e utilizando as funções classe-específicas de quinases e proteases, para CDK2 e TRY, respectivamente.

As curvas ROC obtidas nos estudos de triagem virtual do alvo CDK2 podem ser visualizadas na Figura 2. Os valores para AUC obtidos utilizando os três tipos de funções avaliadas

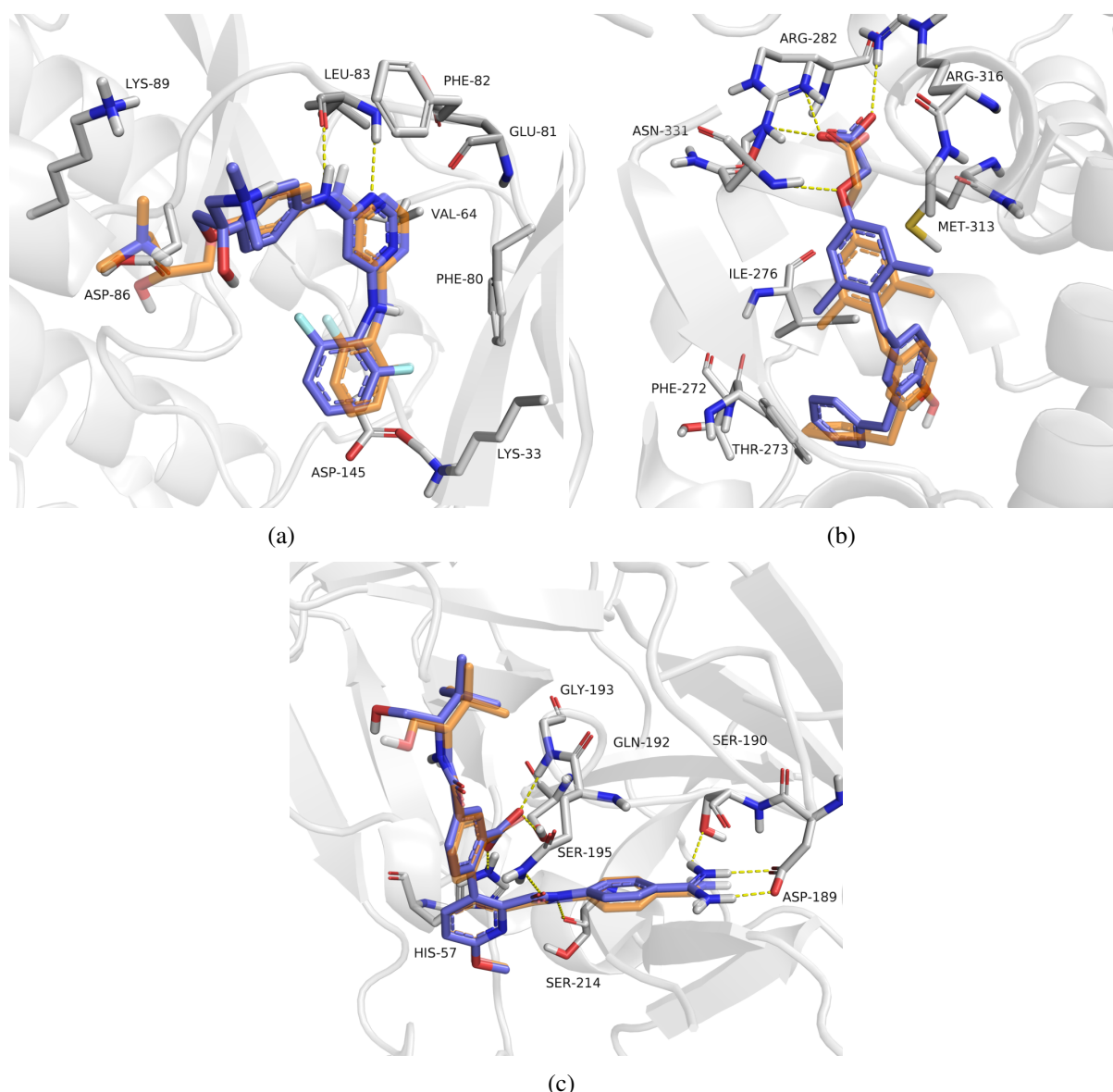


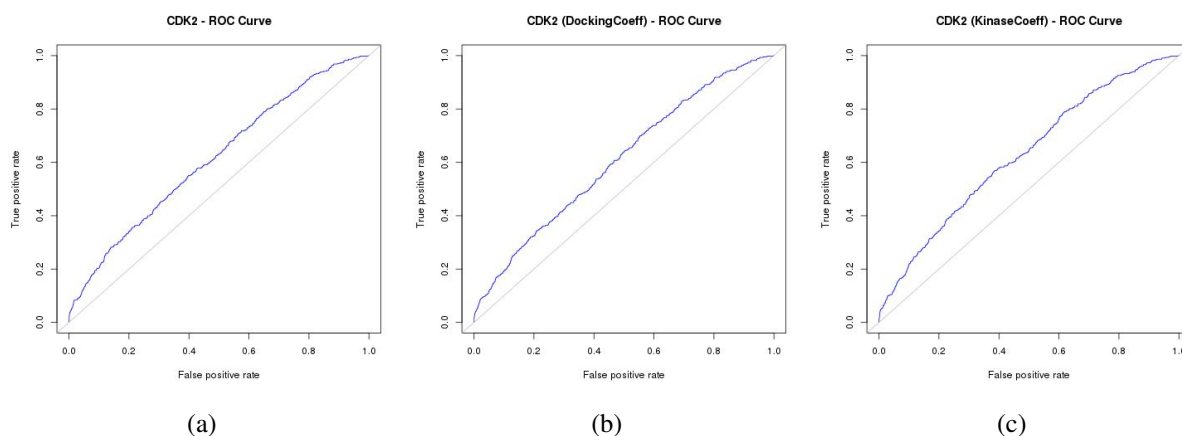
Figura 1: Modos de ligação obtidos no estudo de *redocking*. A conformação do *docking* encontra-se colorida em azul e a conformação do ligante do cristal, colorida em laranja. (a) CDK2, PDB 1H00, (b) THB, PDB 1Q4X, (c) TRY, PDB 2AYW.

(Tabela 3) são bastante similares e demonstram a dificuldade das funções em discriminar ligantes ativos de *decoys*. Entretanto, é possível observar um desempenho um pouco melhor para a função proteína-específica. A dificuldade na discriminação dos ligantes deste alvo pode ser atribuída ao alto grau de flexibilidade observado em proteínas pertencentes a classe das quinases. A qualidade da predição da afinidade de ligação a partir das funções de *scoring* são dependentes da correta predição do modo de ligação e são bem conhecidas as limitações do tratamento da flexibilidade da proteína pelo método *single docking* receptor-rígido (i.e. realizando o atracamento dos ligantes utilizando apenas uma estrutura para representar a proteína-alvo).

Os resultados da triagem virtual dos ligantes do alvo THB (Figura 3) são similares qualitativamente aos obtidos para o alvo CDK2. O sítio ativo das proteínas pertencentes a classe de receptores nucleares, assim como as quinases, também demonstra considerável flexibilidade e

Tabela 3: Valores de *area under the curve* (AUC) das curvas ROC nas triagens virtuais obtidos pelas distintas funções

Proteína-alvo	Funções	AUC
CDK2	Genérica	0,614
	<i>Docking</i>	0,607
	Específica	0,627
THB	Genérica	0,642
	<i>Docking</i>	0,637
TRY	Genérica	0,701
	<i>Docking</i>	0,722
	Específica	0,821

**Figura 2:** Curvas ROC da triagem virtual de ligantes da proteína-alvo Quinase Dependente de Ciclina 2 (CDK2), pertencente a classe das quinases. (a) função genérica, (b) função derivada de *docking* e (c) função quinase-específica.

um comportamento dinâmico, comprovados inclusive pela dificuldade na obtenção de estruturas por cristalografia (Huang et. al., 2010), e devido a isso acreditamos que para a melhoria dos resultados seja necessário tratar a flexibilidade. A classe de proteínas receptores nucleares ainda não dispõe de uma função específica e devido a isso, a análise da triagem virtual do alvo THB foi realizada utilizando apenas as funções genérica e a derivada de *docking*.

Os resultados da triagem da proteína-alvo TRY demonstram bom desempenho no ranqueamento dos ligantes, com valores das AUC-ROC superiores a 0,7 (Tabela 3). As curvas ROC demonstram boa predição das funções (Figura 4), atribuindo um *score* melhor aos ligantes ativos (i.e. verdadeiros positivos) em comparação aos *decoys*. Analisando a classificação realizada ao aplicar as três funções disponíveis, observa-se que a função de *docking* obtém melhores resultados que a genérica e, mais importante, observa-se uma considerável melhora ao utilizarmos a função específica para a classe de proteases.

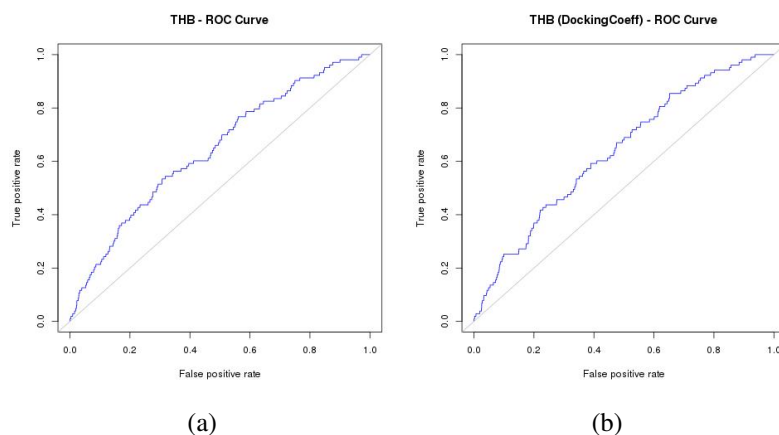


Figura 3: Curvas ROC da triagem virtual de ligantes da proteína-alvo Receptor Beta-1 de Hormônio Tireoidiano (THB), pertencente a classe de receptores nuclear. (a) função genérica e (b) função derivada de *docking*,

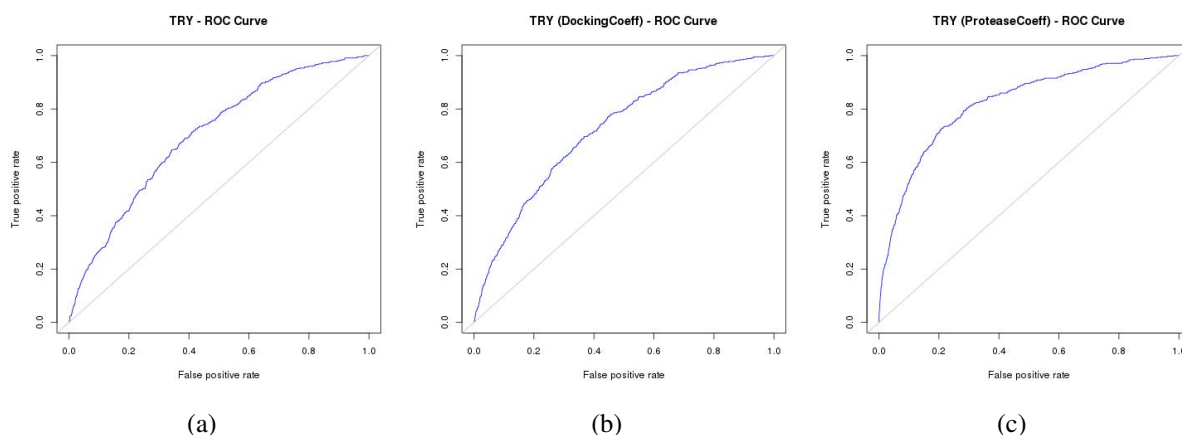


Figura 4: Curvas ROC da triagem virtual de ligantes da proteína-alvo Tripsina (TRY), pertencente a classe das proteases. (a) função genérica, (b) função derivada de *docking* e (c) função protease-específica.

Resultados recuperados da literatura permitem a realização de um comparativo de desempenho junto as funções empíricas de *scoring* desenvolvidas pelo grupo. No trabalho de Zhang, (2014) foi realizada a triagem virtual de alvos selecionados do DUD-E com o objetivo de analisarem o desempenho do programa Vina (Trott, 2010) e da metodologia MM/GBSA. A triagem virtual do alvo CDK2 no estudo citado utilizando o Vina obteve $AUC = 0.684$ e utilizando o MM/GBSA, $AUC = 0.744$. Em análise da proteína-alvo TRY, são demonstrados $AUC = 0.776$ utilizando o programa Vina e $AUC = 0.633$ utilizando o MM/GBSA (Zhang, 2014). Observamos que nossos resultados encontram-se próximos dos visualizados e demonstram um desempenho melhor quando avaliamos a proteína-alvo Tripsina.

5 CONCLUSÕES

As análises demonstram o potencial de nossas funções empíricas de *scoring* aplicadas a estudos de triagem virtual, onde a classificação precisa de ativos e *decoys* consolida-se como característica mandatória. Os resultados observados apontam para o desenvolvimento de funções

classe-específicas, visto que a classificação ativo/inativo utilizando funções empíricas específicas resultam em um melhor ranqueamento dos ligantes, nos proporcionando um classificador mais confiável.

Acreditamos que a flexibilidade observada em determinadas classes de proteínas pode prejudicar a predição do modo de ligação e logo, interferir na predição da afinidade de ligação que implica diretamente na correta classificação dos ligantes. Uma maneira de superar esta limitação é a utilização da estratégia de *ensemble docking*, que em síntese, consiste em realizar o atracamento de um ligante, junto a mais de uma estrutura do receptor, objetivando representar a flexibilidade do receptor por um conjunto de conformações distintas.

Como continuação deste trabalho, almejamos estender a validação das funções utilizando um maior número de alvos do conjunto DUD-E, abordando outras classes de proteínas. É nossa intenção realizar esta validação utilizando a estratégia de *ensemble docking* e procurar implementar termos adicionais nas funções empíricas de afinidade objetivando uma melhoria no processo de classificação ativo/inativo.

Agradecimentos

Os autores agradecem a estrutura e suporte prestado pela equipe Santos Dumont – SINAPAD, ao fomento concedido pelos órgãos CAPES, CNPq, FAPERJ e a colaboração junto ao INCT-INOVAR.

REFERÊNCIAS

- Almeida, D. M., 2011. *Dockthor: Implementação, Aprimoramento e Validação de um Programa de Docking Receptor-Ligante*. MSc thesis, Laboratório Nacional de Computação Científica / Petrópolis-RJ
- de Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C., & Dardenne, L. E., 2014. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Information Sciences*, vol. 289, p. 206-224.
- Guedes, I. A., de Magalhães, C. S., & Dardenne, L. E., 2014. Receptor–ligand molecular docking. *Biophysical Reviews*, vol. 6, n. 1, p. 75-87.
- Guedes, I. A., 2016. *Development of Empirical Scoring Functions for Predicting Protein-Ligand Binding Affinity*. PhD thesis, Laboratório Nacional de Computação Científica / Petrópolis-RJ.
- Halgren, T. A., 1999. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *Journal of Computational Chemistry*, vol. 20, n. 7, p. 730-748.
- Hawkins, P. C., et al. 2008. How to do an evaluation: pitfalls and traps. *Journal of computer-aided molecular design*, vol. 22, n. 3-4, p. 179-190.
- Huang, N., Shoichet, B. K., & Irwin, J. J., 2006. Benchmarking sets for molecular docking. *Journal of medicinal chemistry*, vol. 49, n. 23, p. 6789-6801.
- Huang, P., Chandra, V., & Rastinejad, F., 2010. Structural overview of the nuclear receptor

superfamily: insights into physiology and therapeutics. *Annual review of physiology*, vol. 72, p. 247.

Kirchmair, J., et al. 2008. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes?. *Journal of computer-aided molecular design*, vol. 22, n. 3-4, p. 213-228.

Li, Y., et al. 2014. Comparative assessment of scoring functions on an updated benchmark: 1. Compilation of the test set. *Journal of chemical information and modeling*, vol. 54, n. 6, p. 1700-1716.

Liu, J., & Wang, R., 2015. Classification of current scoring functions. *Journal of chemical information and modeling*, vol. 55, n. 3, p. 475-482.

Mysinger, M. M., Carchia, M., Irwin, J. J., & Shoichet, B. K., 2012. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *Journal of medicinal chemistry*, vol. 55, n. 14, p. 6582-6594.

Trott, O. & Olson, A. J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry*. vol. 31, p. 455-461

Zhang, et al., 2014. Toward fully automated high performance computing drug discovery: a massively parallel virtual screening pipeline for docking and molecular mechanics/generalized Born surface area rescoring to improve enrichment. *Journal of chemical information and modeling*, vol. 54, n. 1, p. 324-337.