

Virtual high throughput screening (vHTS) - A perspective

Sangeetha Subramaniam¹, Monica Mehrotra² and Dinesh Gupta^{1,*}

¹Structural and Computational Biology Group, International Center for Genetic Engineering and Biotechnology (ICGEB), New Delhi; ²Department of Computer Science, Jamia Millia Islamia University, New Delhi; Dinesh Gupta* - E-mail: dinesh@icgeb.res.in; * Corresponding author

received July 30, 2008; accepted August 30, 2008; published September 08, 2008

Abstract:

With the exponential rise in the number of viable novel drug targets, computational methods are being increasingly applied to accelerate the drug discovery process. Virtual High Throughput Screening (vHTS) is one such established methodology to identify drug candidates from large collection of compound libraries. Although it complements the expensive and time consuming High Throughput Screening (HTS) of compound libraries, vHTS possess inherent challenges. The successful vHTS requires the careful implementation of each phase of computational screening experiment right from target preparation to hit identification and lead optimization. This article discusses some of the important considerations that are imperative for designing a successful vHTS experiment.

Keywords: virtual high throughput screening; receptor based and ligand based screening; homology models; chemical databases; ADME filters; toxicity filters

Background:

The discovery of novel drug targets has increased exponentially in recent years due to advances in genomic and molecular biology techniques. Experimental and computational methods are effectively applied to accelerate the process of lead identification and optimization. The chemical leads are small potential drug like molecules which are capable of modulating the function of the target proteins that are further optimized to act as a therapeutic drug against a targeted disease. Conventional experimental methods like High Throughput Screening (HTS) continue to be the best method for rapid identification of drug leads. HTS identifies lead molecules by performing individual biochemical assays with over millions of compounds. However, the huge cost and time consumed with this technology has led to the integration of cheaper and effective computational methodology namely virtual High Throughput Screening (vHTS). vHTS is a computational screening method which is widely applied to screen insilico collection of compound libraries to check the binding affinity of the target receptor with the library compounds [1]. This is achieved by using a scoring function which computes the complementarity of the target receptor with the compounds. HTS and vHTS are complementary methods [2] and vHTS has been shown to reduce false positives in HTS [3]. Several vHTS strategies have been practiced [4] and the technique is being continuously optimized for better performance.

Methodology:

Based on the availability of structural data, vHTS is carried out using receptor based or ligand based screening methods. Receptor based method involves usage of 3D structure of the target receptors to search for potential candidate compounds that

can modulate the target receptor function. Each of the database compounds is docked into the receptor binding site and best electrostatic fit is predicted. Successful application of receptor based method has been reported in many targets [5]. In case where structural information of the target is unavailable, ligand based method make use of the information provided by known inhibitors. Structures similar to the known inhibitors are identified from chemical databases by variety of methods, including similarity and substructure searching, pharmacophore matching and 3D shape matching. Numerous successful applications of ligand based methods have also been reported [2]. In both the methods, the compounds are ranked using an appropriate scoring function based on either complementarity or similarity and the top ranking compounds are taken to the next step of experimental assays. General workflow of vHTS experiment is shown in Figure 1.

Important considerations for vHTS

Incorporating target flexibility in molecular docking

Proteins undergo conformational change upon ligand binding and due to this rigid protein docking may be inadequate for computational drug designing applications [6, 7]. Sampling the intrinsic flexibility of the protein binding site remains a great challenge in flexible protein docking. Nevertheless this problem is being increasingly addressed by several advanced methods which aim at accurate representation of protein flexibility and its ligand recognition [8]. Most of the attempts to incorporate protein flexibility make use of an ensemble of protein structures. A crystallographic structure of the same target bound to different ligands provides a good starting point to obtain ensemble of protein structures. Ensemble of structures derived

from NMR studies has been used to incorporate protein flexibility in docking studies of HIV-1 protease that has demonstrated good docking accuracy [9]. The study has also shown that docking against NMR structures provide an efficient alternative method for standard docking against individual X-ray crystal structures. In the absence of experimental studies, conformational ensemble generated by Molecular Dynamics (MD) simulations has also been used to dock ligands to flexible receptors [10]. The choice of appropriate docking tool which considers reasonable flexibility in protein conformation enhances the reliability of prediction of protein binding affinity.

vHTS using Homology models

In the absence of crystal structures, homology models have helped in lead identification process of drug targets like

falcipain-3 [11] and GPCR [12]. Recently, we have used a homology model of a novel drug target of *Plasmodium falciparum* PfHslV [13], an ATP dependant protease and this model has helped in the identification of several hit compounds (*unpublished data*). Database of homology models like ModWeb [14] and organism specific database like plasmodium model database [15] which are being experimentally validated can be a good source to obtain the models of target proteins. It has been shown that models based on templates with greater than 50% sequence identity has yielded 5-fold better enrichment than one with less sequence identity [16]. On the other hand, small structural errors, such as incorrect assignment of few side chain rotamers have been the cause of poor performance. Hence successful vHTS using a homology model depends on the quality of template crystal structures.

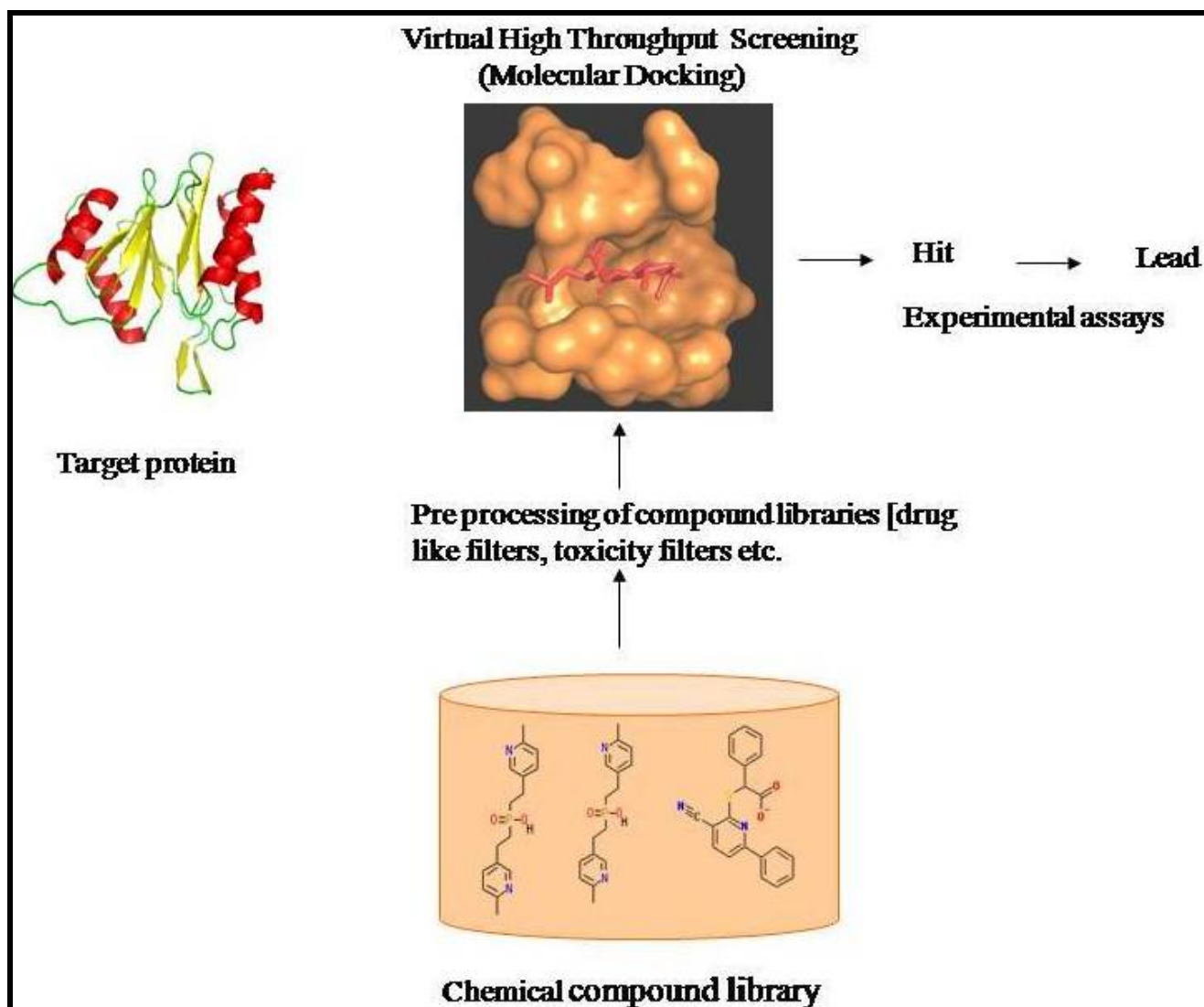


Figure 1: Virtual high throughput screening workflow.

Pre - processing chemical databases

Wide collections of both publicly and commercially available compound libraries are screened in the lead identification process. The compound collection for screening is usually derived from commercial corporate collection or from *in silico* libraries generated by combinatorial methods. Many factors contribute to the selection of an appropriate chemical database for vHTS. Various kinds of filters are being applied to the compound libraries prior to screening in order to: i) increase the probability of finding hits with drug like characteristics (ADME properties) or lead like characteristics (with physico chemical profile of a lead), ii) to ensure the meaningful composition of the library to be screened, there by saving computational demand while screening.

ADME property filters and toxicity filters

Most important and widely applied filter is the drug-likeness of the compound which represent ADME profile (Adsorption, Distribution, Metabolism and Excretion) as determined by Lipinski's rule of five, which is a set of rules based on molecular weight, lipophilicity and hydrophobicity, H bond donors and H bond acceptors [17]. The compounds which are labeled as drug-like (Figure 2) resembles the existing drug molecules and exhibit the following property cut-off values: $MW \leq 500$, $\text{LogP} \leq 5$, H bond acceptors ≤ 10 and H bond donors ≤ 5 . The compounds exceeding the cut-off values tend to have solubility and permeability problems which would lead to poor oral bioavailability. In addition to the rule of five there are other properties which help in distinguishing drug and non drug-like compounds. The extensions include polar surface area $\leq 140 \text{ \AA}$ and rotatable bonds ≤ 10 . Prediction of toxicity before the synthesis of compounds ensures the removal of compounds with potential toxic effects. A number of *in silico* systems for toxicity prediction are available, which help in the classification of toxic and non-toxic compounds [18]. The above mentioned filters are essential to discard compounds without any drug like characteristics and undesirable toxic effects before screening. Such *in silico* filters are essential to minimize expensive drug failures in the later stage of drug development.

Box 1: Drug-like compounds [21]

Rule of 5:

- $MW \leq 500$
- $\text{ClogP} \leq 5$
- H bond donors ≤ 5
- H bond acceptors ≤ 10

Extensions:

- Polar surface area $\leq 140 \text{ \AA}$ or sum of H-bond donors and acceptors ≤ 12
- Rotatable bonds ≤ 10

Figure 2: Drug-like compounds [21].

Meaningful composition of library

Tautomer enumeration is also an essential step that has been shown to enhance the hit rates in the vHTS [19]. Several studies have shown that specific tautomeric states of molecules are present which interact with the active site residues of the target [20]. Tautomers are often recognized as different structures by computational applications, leading to different results. Hence chemical databases need to be enriched with correct bioactive tautomers to ensure reliable results. Removal of compounds known to interact with anti targets e.g., hERG (*human Ether-a-go-go Related Gene*) channel can be useful [21]. In addition to various factors it is important to confirm the compound availability or the synthetic feasibility of the compound before considering them for vHTS.

Chemical diversity of chemical databases

Chemical database with wide chemical space and structural diversity offers an ideal solution for lead discovery. The basis behind such an analysis is that, similar compounds exhibit similar physico-chemical and biological properties. A chemical database enriched with the representatives of dissimilar compounds or diverse chemical structures shall increase the probability of finding different leads with similar biological activities. It also ensures that database is small but diverse enough to be readily managed. Diversity of chemical databases is evaluated by examining how much the compounds within the library differ in terms of the distribution of their properties. Different kinds of diversity quantification are carried out: Distance based methods, Cell-partitioning methods and clustering methods. At this juncture, it will be useful to mention the study on the major chemical databases demonstrating the diversity presented in these databases and also the overlapping structures in these databases [22]. Such studies provide a view of the diverse nature of the database and helps in selecting an appropriate chemical database.

Conclusion:

vHTS is cost-effective and reliable technique that can be applied to identify potential leads and avoid undesirable compounds that would otherwise result in expensive and time consuming experimental methods. However, vHTS often requires careful preparation of both target and compound library, use of optimal parameters as well as careful analysis of the results. It should be noted that experience and knowledge about the target are very crucial in identifying true positives in such experiments. We have discussed some of the essential considerations in designing vHTS experiment focusing the pre-processing of chemical databases that are used in vHTS.

References:

- [01] B. K. Shoichet, *Nature*, 432: 862 (2004) [PMID: 15602552]
- [02] J. Bajorath, *Nat Rev Drug Discov.*, 1: 882 (2002) [PMID: 12415248]
- [03] J. L. Jenkins *et al.*, *Proteins*, 50: 81 (2003) [PMID: 12471601]
- [04] C. McInnes, *Curr Opin Chem Biol.*, 5: 494 (2007) [PMID: 17936059]

- [05] D. B. Kitchen *et al.*, *Nat Rev Drug Discov.*, 3: 935 (2004) [PMID: 15520816]
- [06] H. A. Carlson, *Curr Opin Chem Biol.*, 6: 447 (2002) [PMID: 12133719]
- [07] S. J. Teague, *Nat Rev Drug Discov.*, 2: 527 (2003) [PMID: 12838268]
- [08] M. L. Teodoro and L. E. Kaviraki, *Curr Pharm Des.*, 9: 1635 (2003) [PMID: 12871062]
- [09] S. Y. Huang and X. Zou, *Protein Sci.*, 16: 43 (2007) [PMID: 17123961]
- [10] C. F. Wong *et al.*, *Proteins*, 61: 850 (2005) [PMID: 16245317]
- [11] P. V. Desai *et al.*, *J. Med. Chem.*, 47: 6609 (2004) [PMID: 15588096]
- [12] C. Bissantz *et al.*, *Proteins*, 50: 5 (2003) [PMID: 12471595]
- [13] G. Ramasamy *et al.*, *Mol Biochem Parasitol.*, 152: 139 (2007) [PMID: 17270290]
- [14] U. Pieper *et al.*, *Nuc. Acid Research*, 34: 291 (2006) [PMID: 16381869]
- [15] G. Ramasamy *et al.*, *Bioinformatics*, 5: 50 (2005) [PMID: 17597852]
- [16] C. Oshiro *et al.*, *J. Med. Chem.*, 47: 764 (2004) [PMID: 14736258]
- [17] C. A. Lipinski *et al.*, *Adv Drug Deliv Rev.*, 46: 3 (2001) [PMID: 11259830]
- [18] A. M. Aronov, *Drug Discov Today*, 10: 149 (2005) [PMID: 15718164]
- [19] J. C. Dearden, *J Comput Aided Mol Des.*, 17: 119 (2003) [PMID: 13677480]
- [20] F. Oellien *et al.*, *J Chem Inf Model*, 46: 2342 (2006) [PMID: 17125178]
- [21] T. H. Keller *et al.*, *Curr Opin Chem Biol.*, 10: 357 (2006) [PMID: 16814592]
- [22] M. Krier *et al.*, *J Chem Inf Model*, 46: 512 (2006) [PMID: 16562979]

Edited by P. Kanguane

Citation: Subramaniam *et al.*, *Bioinformatics* 3(1): 14-17 (2008)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.