Virtual Screening in Hepatitis B Virus Drug Discovery: Current Stateof- the-Art and Future Perspectives

Mohcine Elmessaoudi-Idrissi, Arnaud Blondel, Anass Kettani, Marc P Windisch, Soumaya Benjelloun, Sayeh Ezzikouri

Virtual Screening in Hepatitis B Virus Drug Discovery: Current State–of-the-Art and Future Perspectives

Running Head: *In silico* screening and HBV

Mohcine Elmessaoudi-Idrissi[1,2], Anass Kettani[2], Marc P. Windisch[3], Arnaud Blondel[4], Soumaya Benjelloun[1], Sayeh Ezzikouri [1,*]

[1] Virology Unit, Viral Hepatitis Laboratory, Institut Pasteur du Maroc, Casablanca, Morocco

[2] Laboratoire de Biologie et Santé (URAC34), Département de Biologie, Faculté des Sciences Ben Msik, Université Hassan II de Casablanca, Morocco.

[3] Hepatitis Research Laboratory, Discovery Biology Department, Institut Pasteur Korea, Seongnam-si, Gyeonggi-do, South-Korea.

[4] Unité de Bioinformatique Structurale, CNRS UMR 3528, Département de Biologie Structurale et Chimie, Institut Pasteur, Paris, France.

*Correspondence should be addressed to Dr. Sayeh Ezzikouri Virology Unit, Viral Hepatitis Laboratory, Institut Pasteur du Maroc 1, Place Louis Pasteur, 20360 Casablanca-Morocco. Tél: +212 5 22434470; Fax: +212 5 22260957; E.mail: sayeh.ezzikouri@pasteur.ma

Abstract

Hepatitis B Virus (HBV) is a major global health burden. Interferon alpha and nucleos(t)ide analogues are currently the standard-of-care for chronic HBV infection. However, these antiviral agents have limited efficacy and do not result in a sustained virological response in the majority of infected patients. Virtual Screening (VS) strategies have now a strong impact on drug discovery, the strength of this research field has been corroborated by recent contributions in the development of novel drug candidates which are in clinical trials or which are already available in the clinics. In this context, different VS strategies have been applied to HBV in order to discover novel inhibitors. In this review, we summarize the VS efforts to identify and design novel HBV interventions. We believe that the combination of *in silico* and *in vitro* tools can lead to faster validation of novel drug targets, which could accelerate the HBV drug discovery and development efforts.

Keywords: Ligand-based virtual screening, structure-based virtual screening, quantitative structure-activity relationships, docking, Hepatitis B Virus, inhibitors, workflow.

Introduction

Hepatitis B virus (HBV) is the most prevalent viral liver infection worldwide and poses a significant global public health problem. HBV causes acute and chronic viral hepatitis in humans and is often associated with severe liver diseases, including cirrhosis and hepatocellular carcinoma (HCC) [1]. Although effective HBV vaccines are available, more than 240 million patients are chronically infected worldwide [2,3] and about 600,000 people die every year due to the complications of HBV-related chronic liver diseases. Interferon alpha (IFNa) and nucleos(t)ide analogues are currently the standard-of-care (SOC) for chronic hepatitis B (CHB), although these antiviral agents have limited efficacy and do not result in a sustained virological response (SVR) in most cases[4,5]. Moreover, TAF (Tenofovir alafenamide fumarate), a prodrug of tenofovir has been recently approved by FDA for the treatment of CHB[6-8]. However, long-term treatment of nucleos(t)ide analogues may result in the selection of drug resistant viruses [9,10] . HBV cure is hampered by its covalently closed circular DNA (cccDNA) and chromosomic integrated HBC-DNA[11] , which hides in the hosts' nucleus and evade today's therapies. Therefore, the development of novel therapeutic strategies that interfere with different steps of the viral life cycle including inhibitors of viral entry, polymerase inhibitors, capsid and assembly inhibitors, virus release blockers and inhibitors of cccDNA formation with agents that enhance or activate human response taking into account disease stages, and can be used in combination with existing therapies will lead to complete or functional cure[11,12] In order to address these medical needs, computational methods can be a very useful in the drug discovery process. Use of *in silico* methods in combination with experimental analysis can

help reduce costs and time. Overall, Virtual Screening (VS) is one of a widely used computational method for identification of Lead compounds during target-based drug discovery.

Drug Discovery Process

Drug design (DD) is a process involving various methods to discover new molecules affecting the function of a specific or yet to be disclosed target [13]. A target is a biological component, such as a protein, which is associated with biological processes involved in the pathology. It can be an enzyme, a receptor or some other protein that responds to an external stimulus [13,14] . There are severaltarget-based and target-free strategies to develop novel drugs. The first ones focus on a target, its identification and validation based on its function and its associations with the disease, in parallel with the very drug design. The second ones focus on the identification of hits, chemicals or biological substances, which can have an effect on a biological function as tested in an integrated assay such as in phenotypic screens (Figure 1). High throughput screening (HTS) and VS techniques are generally used at early steps and are followed by pharmacological studies (preclinical studies) and clinical trials in human, which include four phases from 1 to 4 as mentioned in the Figure 2. The latter, highlights also major steps of conventional approach in drug discovery. Hence, different tools and techniques for the identification of new Lead compounds have been described. The target-free drug design includes HTS, notably phenotypic screens, and combinatorial chemistry. The HTS is a technique for the large-scale screening of a multitude of compound samples against many biological targets using laboratory automation and robotic

systems. As a parallel, combinatorial chemistry enables chemists to potentially synthesize hundreds or thousands of compounds in one reaction instead of preparing only a few ones by conventional methodologies. The latter approach can be exploited in either target-free or target-based DD. The target-based drug design methodology can benefit from the use the VS approach. This review will be focused on the merit of that combination.

Virtual Screening (VS)

VS is the computational alternative to HTS campaigns and refers to the *in silico* methodologies towards the identification of hit or Lead compounds. In the last decades, the term of 'virtual screening' became widely used and the number of publications increased exponentially [14-19]. In particular, several *in silico* screening studies have been successful in the discovery of compounds[15-17] that are now in clinical trials or even commercially available[16-19,]. VS techniques can correspondingly be subdivided in Ligand-Based Virtual Screening (LBVS) and Structure-Based Virtual Screening (SBVS) methods. LBVS is based on exploration of biological information of known active ligands against the target of interest in order to identify new ligands using similarity search, quantitative structure-activity relation-ships (QSAR) and pharmacophore research [20]. Machine learning has also been used in LBVS campaigns[21,22]. Several classification algorithms have been implemented, e.g. Multiple Linear Regression, k-Nearest Neighbors (kNN), Naïve Bayesian Classification, Support Vector Machines (SVM), Artificial Neural Networks (ANN), decision trees and probabilistic neural network (PNN) [21,22].

On the side of SBVS, docking calculation is the main tool, which could to fit compounds from databases into the active site of the target protein structure and rank them using scoring functions [23]. Several tools and packages have been widely used for molecular docking. Among them Autodock[24], Autodock Vina[25], DOCK[26], Glide[27], Gold[28], FlexX[29], ICM[30], eHits[31], LigandFit[32] and some of them were utilized also in the HBV studies (Table 3). These methods have been reported and sufficiently explained in a recent review by Enrico *et al*[35], which proposed a global pipeline for both strategies (SBVS and LBVS) including recommendations and description of different steps of VS from data collection, which include target protein structure selection and compound libraries determination from specialized databases (see Table 4). In addition, preprocessing, Screening (Structure-based and Ligand-based) methods, selectivity and ADME Tox filtering have been also summarized in the same work performed by Enrico and colleagues[33]. However, success stories of VS do not preclude the existence of limitations and challenges. Thus, a recent review has reported severals pitfalls related to VS methods[34]. In fact, many of the pitfalls of QSAR were reported related to the multiple binding modes, data size and variety, incorrect feature selection for pharmacophore design and others issues[34-76]. Moreover, docking methods were also evaluated and different problems were discussed related to prediction of binding pose which can in some cases wrong even if the scores produced are high for resulted hits[34,38-40]. These issues and others have been reported in a recent review summarized a 15 years of research in docking methods[39]. Furthermore, very potent Leads identified by VS techniques are rare and less bioactive comparing to others approaches like HTS[41]. These Limitations are mainly due to computational involvement,

due to biological process intricacy and interactions physical complexity, and the quasi-infinite number of compounds that could be considered. Furthermore, the use of virtual structures in *de novo* compound design may raise the issue of synthesis or access to compounds selected by virtual screening [42]. Despite these challenges, VS is continuously evolving, in terms of methods and strategies to maximize the discovery of novel active chemotypes[43].

In Silico Strategies for HBV Inhibitors Identification

Several studies reported various *in silico* strategies and methods to identify novel HBV inhibitors[15-19,27,29,33,41,49]. All the targets were not subjected to the same research efforts. Nonetheless, for sake of clarity, we classified these studies according to HBV targets in this review. We described VS approaches using LBVS or SBVS, and their associated computational tools, as well as the compounds libraries used.

Reverse Transcriptase (RT) Inhibitors:

The HBV polymerase is composed of four domains, bearing three enzymatic activities. The Terminal Protein (TP), non-conserved spacer, Reverse Transcriptase (RT) and the RNase H domain [44]. The RT domain is an RNA-dependent DNA polymerase and DNA-dependent DNA polymerase, which is crucial for HBV replication. Several studies targeted the HBV RT[45-49] with VS methods. In fact, a recent study proposed a novel workflow to discover new HBV polymerase inhibitors [45]. The pipeline was based on a novel method ChooseLD [50] a protein-ligand-flexible docking method. The fingerprint alignment score (FPAScore) value was used to determine the docking conformation of

the ligand. This method includes matching of chemical descriptors such as fingerprints and calculation of the root mean square deviation (RMSD) of the coordinates of atoms used to define the chemical descriptors. The FPAScore is calculated after optimization of the translation and rotation by a Monte Carlo approach. The Akos Samples database (see Table 4) was chosen, and its 2.2 million compounds were screened *in silico* by docking in the RT active site. Using this *in silico* analysis, 60 candidates showing an anti-viral activity on the HBV RT were identified. Thereby, 30 compounds were pyrimidine-type nucleotides and 30 were purine-type nucleotides. Interestingly, of this selection, 12 compounds were prioritized for virus assays on cells, and 4 displayed detectable activity, 2 without significant cell toxicity. Hence, this approach displayed 15 % hit rate in primary test. Later on, a docking study revealed good interactions between phytocompounds extracted from *Phyllanthus* niruri and the viral polymerase[46]. Nonetheless, *in vitro* tests are crucial to validate these *in silico* predictions. Furthermore, in previous studies the impact of drug resistance mutations on the HBV therapy was evaluated[47–49]. In this context, Ismail and Colleagues adopted a Structure-based VS approach[51]. The followed procedure includes, the homology modelling which was performed based on the PDB[84] structure (ID: 1RTD) of (HIV-1) RT template using MODELLER[52], followed by a quality control using VERIFY3D[53], and PROCHECK[54] of resulted models in order to test HBV drugs adefovir (ADV) (Figure 3A) and entecavir (ETV) (Figure 3B). These have been tested later in a similar study by the same group[47]. Docking of those compounds revealed that the mutation (rtI233V) was unlikely to raise resistance against ADV treatment, as it is located away from the

drug interaction site. Similarly, rtV173L should not confer resistance to ETV. In addition, Molecular Dynamics simulations was performed for the wild type and mutant HBV polymerase/rt in complex with entecavir using GROMACS[55] and the result of RMSD indicated that the cited mutation (rtV173L) should not affect the conformation of the targeted proteins[47]. Taken together, *in silico* analyses have often been a complementary step to validate the dependance between resistance mutation and antiviral action of specific treatment[47,48] or to guide chemical experiments[49] . Moreover, the latter study[49] was performed to investigate the molecular mechanisms underlying the influence of DNA polymerase from different HBV genotypes[48]. The results support the hypothesis that the HBV genotype C polymerase is more sensitive to ADV compared to genotype B. In addition, the sensitivity is related to residue N236 and the polymorphic site 238[48].

Capsid Inhibitors

The HBV capsid is a stable polymer composed of 90 or 120 homo-dimers of the viral core protein. It contains viral DNA and reverse transcriptase and within the capsid viral DNA is synthesized[56].

Several studies targeted the HBV capsid in order to discover novel assembly- or disassembly inhibitors. In a previous *in silico* screening [57], a model formed of three capsid protein dimers and the 0YP cognate ligand was built from the Protein Data Bank (PDB[84]) structure (4G93) with FAMS tool[58]. Then, chooseLD[50] *in silico* screening was performed using 1296 chemical compounds from the AKOS database(see Table 4). As a result, 112 compounds based on FPA (fingerprint alignment) Score values were selected on three models. After classification by similarity in 25 groups, 60

compounds were selected as potential HBV inhibitors. 16 compounds could be purchased and tested *in vitro*, of which 4 interfered with HBV replication without cell toxicity. The four compounds C9, C10, C13, and C16 are shown in Table 2. Based on that study[57] Watanabe and colleagues performed Molecular Dynamics (MD) simulations with the AMBER ff12SB force field to study the structure and dynamics of the capsid at the atomic level and its effect on binding of C13 compound[59] (Table 1 of ref. 26). In fact, successions of simulations under different conditions were performed with free HBV capsids, capsid C13 complexes and the capsid-AT-130 complexes. The results comforted that C13 is a capsid-binding compound and is likely to inhibit the functionality of the HBV capsid. Recently, a molecular docking approach was applied to specific proteins targets (EF3-CaM adenylyl cyclase (1PK0), deoxy-cytidine kinase (2NOA), human nucleoside diphosphate kinase (3FKB), human hepatitis B viral capsid (1QGT), hepatitis B X-interacting protein (3MSH)) with the Magnolol, a phenolic bioactive phytomolecule (Figure 3C), to identify a plausible site of action. In this study, the 2D structure of Magnolol was constructed using ChemDraw Ultra 8.0 (www.cambridgesoft.com) and VLife MDS (www.vlifesciences.com) for 3D conversion, followed by Merck molecular force field (MMFF) for energy minimization using an RMS gradient down to 0.01 kcal/mol/Å. Then, the set of Magnolol conformers was generated based on the selection of rotatable bonds. Before running the docking simulations, information on the active site was collected. GRIP[60] docking was performed with parameters allowing extensive search (number of placements: 50, rotation angle: 10°, ligand wise results: 10, exhaustive method, scoring function:

PLP score). Docking studies revealed that Magnolol has the capability to target more than one key mechanism to act as anti-HBV agent (Table 2). In this study, it was demonstrated that Magnolol binds to all tested proteins EF3-CaM adenylyl cyclase, deoxycytidine kinase, human nucleoside diphosphate kinase, HBV capsid, and hepatitis B X-interacting proteins[61] . For the first target, which allostery could be exploited with original *in silico* methods(ref Laine &al. 2010, PNAS, 107 p11277), comparison showed it shared many interactions with the highly effective adefovir inhibitor co-crystalized with its target (1PK0, Figure 3.E).

HBx Inhibitors

HBx protein is a small polypeptide that is essential for viral infectivity. In addition, HBx protein inactivates negative growth regulators, and it inhibits the expression of tumor suppressor genes[62]. This makes HBx a very promising, but challenging target for it lacks a structure for its entire sequence. Nonetheless, in a previous *in silico* study[63], a complex of a motif of HBX with DDB1 (DNA human Damage Binding Protein) (PDB ID 3I7H) was targeted using 100 ligands docked with Ligbuilder[64] and HEX [65] . Ten ligands were selected by applying optimization of their binding affinity. One ligand passed other tests like binding affinity and drug-like properties. After VS, a protein-ligand docking was performed to the catalytic triad of HBx protein using AUTODOCK vina[35] and the protein-ligand conformations were including hydrogen bonds and the bond lengths were analyzed using Accelrys DS Visualizer (http://accelrys.com/). The results implied that the ligand binds well based on its low free energy. Quantitative structure-activity relationships (QSAR)

analysis has shown the same ligand as best inhibitor candidate for future *in vitro* or *in vivo* validation. In a recent study, a natural compound library of plant-derived products was docked against the predicted structure of HBx protein using LOMET (Local Meta-Threading-Server)[66]. Molegro Virtual Docker (MVD) (http://www.molegro.com) was used for docking against natural compounds (curcumin, oleanolic acid, resveratrol, bilobetin, luteolin, ellagic acid, betulinic acid and rutin). The results put forward Rutin as a candidate, which was subjected to further design and screening against HBx [67]. Among the 20 design analogs, Rutin01 and Rutin08 were fond promising[67]. Yet, those candidates still require *in vitro/in vivo* validation.

Taurocholate cotransporter polypeptide (NTCP) Inhibitors

The human sodium-dependent taurocholate cotransporter polypeptide (NTCP) has recently been identified as a functional receptor for the HBV[68]. In this context, NTCP has been targeted for HBV therapy using an *in silico* analysis [69]. The NTCP protein-protein putative interaction network composed of 340 unique proteins was constructed with systems biology approach using several databases (InAct [70], HPRD[71], HomoMINT [72], BIND[73] , BioGRID[74] and DIP[75]). Structure-based VS was followed to target NTCP. A small molecules library was collected from FDA-approved drugs and ZINC[76] databases. Based on the NTCP_Human PDB[84] structure (32OY) a flexible ligand-docking was performed to a rigid receptor using two successive scoring methods (grid-based and amber). The grid-based identified and ranked the top 200 compounds. Those compounds were ranked using the amber scoring method, which led to the selection of 30 compounds from both

databases. Based on these results five (N1 to N5) newly identified compounds were validated using *in vitro* experiments. Interestingly, the results showed that the compound N4 (azelastine hydrochloride) is compound targeting NCTP.

Ribonucleotide reductase M2 (RRM2) Inhibitors

Human ribonucleotide reductase M2 (RRM2) is the rate-limiting enzyme for DNA synthesis by conversion of ribonucleoside diphosphates (rNDPs) to deoxyribonucleoside diphosphates (dNDPs). The active form of the enzyme is composed of two identical large subunits (RRM1) and two identical small subunits (RRM2 or its homolog RRM2B). A global analysis including *in silico*, *in vitro* and *in vivo* tools were performed for RR small subunit M2 target [77]. A virtual screening based on molecular docking of olasmid selected from CMC(Comprehensive Medicinal Chemistry) database(Table 4) against the RRM2. The PDB[84] crystal structure of RRM2 (PDB ID: 3OLJ) was chosen. Using Glide, a docking algorithm that uses an empirical scoring function [27], the molecular docking revealed osalmid (Figure 3.D) as a potential compound. Furthermore, using stably HBV replicating HepG2.2.15 cells, osalmid displayed very significant inhibition of HBV DNA and cccDNA synthesis activity, higher than that of hydroxyurea. Hence, the new derivative of olasmid, YZ51 (4-cyclopropyl-2-fluoro-N-(4-hydroxyphenyl) benzamide) displayed promising activities and could be a novel class of anti-HBV candidates with potential use for hepatitis B and HBV-related HCC treatment.

Inhibitors of hLa

The human La (hLa) protein is known as an important element, forming a complex with HBV RNA ribonucleoprotein, which promotes HBV replication[78]. A structure-based virtual screening approach was used to discover novel ligands of hLa with anti-HBV activity. A previous study using a SBVS approach based on the structure of hLa protein (PDB ID: 2VOD) was performed. To validate the docking approach, the U-2 nucleotide, which bound most deeply into the RNA groove of the hLa protein, was extracted from the crystal structure of chain A and used to perform docking. 3D structures of compounds from the SPECS Database (Table 4) were downloaded and processed with LigPrep[79] software. Each chemical structure (for small molecules) was docked into the U-2 nucleotide-binding site using GLIDE[27], and the Glide scoring function (G-Score) was used to rank these final poses for all compounds in decreasing order. The binding poses of the top 30 compounds were stored for visual inspection of the docking geometry to avoid unreasonable interactions. Finally, 26 of the 30 compounds were selected for the evaluation of inhibition of HBV and their effects on hLa expression[78]. Among them, HBSC-11, HBSC-15 and HBSC-34 were selected for evaluation (Table 5). One displayed sensitive activities in the micromolar range (HBSC-11).

Conclusion & Future Directions

Several workflows have been reported in this review of known HBV inhibitors with different VS strategies and available compound databases. These approaches produced prioritized list of candidates that are commercially available or that can be synthesized. Many of these *in silico* studies could be validated *in vitro* for different targets such as Reverse transcriptase (RT), capsid

assembly and hLa protein. This is an essential step to identify new antiviral compounds against HBV. In this context, the focus on particular targets like NTCP, cccDNA formation and capsid assembly may produce good results in the future. In this line, authors of a recent review [80]proposed new targets for hepatitis B, which can be studied using *in silico* approach for the discovery of new inhibitors. Moreover, computational approaches have revolutionized the biomedical research, especially in drug discovery with the exponential use of VS as a complementary tool for Lead discovery and optimization. However, computational approaches always have limits related to different factors, which can bias the results and their predictions remain hypothetical until experimental validation. Hence, there are limits for both structure based and ligand-based VS methods. The challenges can be summarized in the algorithm, parameters, and resources used to achieve accuracy and efficiency in molecular docking, including the choice of scoring functions [81] . In effect, issues such as solvent effects, entropic effects, and receptor flexibility are major challenges that must be handled [82]. The recent advanced in proteomics, protein-protein interactions, structural genomics, and also chemistry can be exploited for the development of new methods for lead discovery of new HBV candidates.

List of Abbreviations

DD : Drug Design

HTS: High Throughput Screening

VS: Virtual screening

LBVS : Ligand Based Virtual Screening

SBVS: Structure Based Virtual Screening

kNN: k-Nearest Neighbors

SVM: Support Vector Machines

PNN: Probabilistic Neural Network

RT: Reverse Transcriptase

Legends figures

Figure 1: Lead Compounds Discovery Strategies

Figure 2: Drug Development Process

Figure 3: Chemical Structure of Drugs. (A) Adefovir Chemical Structure. (B) Entecavir Chemical Structure. (C) Magnolol Chemical Structure. (D) Osalmid Chemical Structure. (E) Interaction made by Adefovir with its target (1PK0) and residues forseen to interact with Magnolol. Target secondary structure in cartoon; inhibitor, water molecule and ion in heavy licorices, target interacting residues in light licorice; bond with hydrogen in lines; H-bonds in dashed cylinders; residues found in interaction with Magnolol highlighted by small spheres at heavy atom positions.

REFERENCES

Glossary

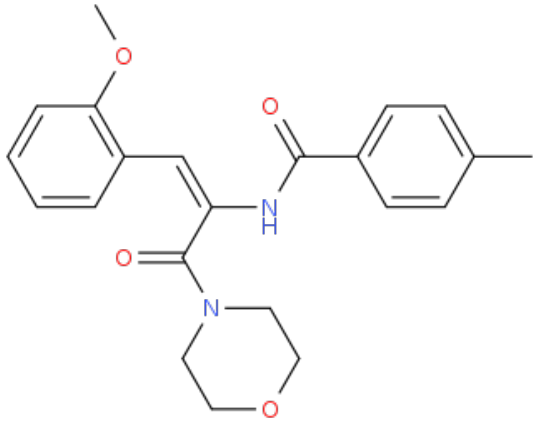
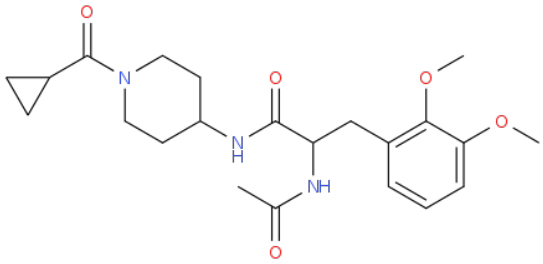
Term	Definition	Reference
Molecular Fingerprints	Molecular fingerprints are a method of encoding the structure of a molecule as binary bits . fingerprints are useful for similarity search between two molecules.	[50]
Pharmacophore	Pharmacophore is a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity	[20]
Machine learning	Machine learning are computational methods which can be subdivided into supervised and unsupervised. In virtual screening, the main role of these methods is the classification of training-set containing known active and known inactive molecules.	[21][22]

k-Nearest Neighbors	The *k*-NN algorithm is a simple and intuitive method to predict the class property or rank of a molecule based on nearest training.	[21][22]
Support Vector Machines	SVMs, are supervised machine-learning algorithms for facilitating compound classification, ranking and regression-based property value prediction. In VS SVMs are used to differentiate between actives and inactives compounds	[21][22]
Artificial Neural Networks (ANN)	ANNs are the most popular and widely used In drug discovery for compound classification, QSAR studies, primary VS of compounds, identification of potential drug targets	[21][22]
ChooseLD	ChooseLD (CHOOse biological information Semi-Empirically on the Ligand Docking), which uses simulated annealing for protein–ligand flexible docking	[50]

Table 1: PDB ids of targets

Target Name	Protein Structure ID
EF3-CaM adenylyl cyclase	1PK0
deoxycytidine kinase	2NOA
Human nucleoside diphosphate kinase	3FKB
Human Hepatitis B viral capsid	1QGT
Hepatitis B X-interacting protein	3MSH

Table 2: Chemical nomenclature

Component Name	Chemical Name
 C9	(N-[(E)-1-(2-methoxyphenyl)-3-morpholin-4-yl-3-oxoprop-1-en-2-yl]- 4-methylbenzamide)
 C10	(2-acetamido-N-[1-(cyclopropanecarbonyl) piperidin-4-yl]-3-(2,3-dimethoxyphenyl) propanamide)
C13	(4-(4-acetylpiperazin-1-yl)-N-[1-(cyclopropylamino)-1-oxo-3-phenylpropan-2-yl]-3-nitrobenzamide)

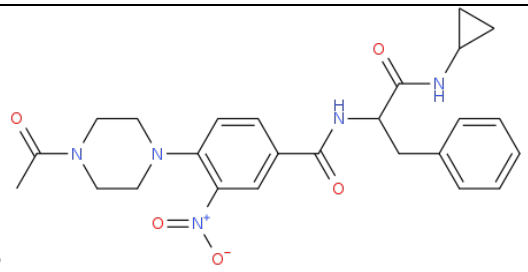
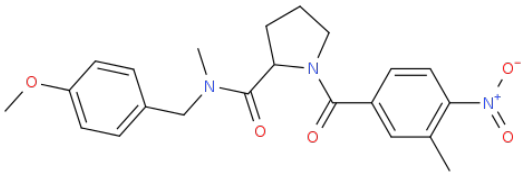
 C13 .	
 C16	N-[(4-methoxyphenyl)methyl]- N-methyl-1- (3-methyl-4-nitrobenzoyl) pyrrolidine-2-carboxamide)

Table 3: List of docking softwares

Program	Principle	Free for academia	Reference	Web link
AutoDock	Monte Carlo & Lamarckian genetic algorithm	Yes	[24]	http://autodock.scripps.edu
Autodock Vina	Iterated local search	Yes	[25]	http://autodock.scripps.edu
Dock	Incremental construction	Yes	[26]	http://dock.compbio.ucsf.edu
Glide	Hybrid	No	[27]	http://www.schrodinger.com
Gold	Genetic algorithm	No	[28]	http://www.ccdc.cam.ac.uk/products/life_sciences/gold
FlexX	Incremental construction	No	[29]	http://www.biosolveit.de/flexx
HEX	First Fourier transform (FFT)-based algorithm	Yes	[65]	http://hexserver.loria.fr/

ICM	Monte Carlo	No	[30]	http://www.molsoft.com/docking.html
LigandFit	Monte Carlo	No	[31]	http://accelrys.com/products/discovery-studio
eHiTS	Incremental construction	No	[32]	http://www.simbiosys.ca/ehits/index.html

Table 4: Databases list

Database	**Type**	**Number of compounds**	**Website**
PubChem	Public	93 million	http://pubchem.ncbi.nlm.nih.gov
ChEMBL	Public	2 million	https://www.ebi.ac.uk/chembldb/index.php
ChemSpider	Public	60 million	http://www.chemspider.com
CoCoCo	Public	7 million	http://cococo.unimore.it/tiki-index.php
ZINC	Public	100 million	http://zinc.docking.org
ChemBridge	Commercial	1.1 million	http://www.chembridge.com

Specs	Commercial	240000	http://www.specs.net
ChemDiv	Commercial	1.5 million	http://www.chemdiv.com
Enamine	Commercial	2.4 million	http://www.enamine.net
ChemNavigator	Commercial	91.5 million	http://www.chemnavigator.com
Akos Samples	Commercial	27 million	http://www.akosgmbh.de/AKosSamples/

Table 5: List of potential anti-HBV compounds

Molecule	Structure	Target	Reference
M8	5′-Deoxy-5-fluorocytidine/245.2/C9H12F1N3O4	RT	[45]
M9	2′-Deoxy-5-fluorocytidine/245.2/C9H12F1N3O4	RT	[45]
M10	2′-Deoxy-5-methylcytidine/241.2/C10H15N3O4	RT	[45]
M14	1-(3-O-Acetyl-2-deoxypentofuranosyl)-5-methyl-2,4(1H,3H)-pyrimidinedione/284.27/C12H16N2O6	RT	[45]
M15	5′-Deoxy-2′,3′-di-O-acetyl-5-fluorocytidine	RT	[45]
M16	2′,3′-Isopropylidineuridine	RT	[45]
M11	S-Adenosylhomocysteine	RT	[45]
M12	5′-N-Ethylcarboxamido-adenosine	RT	[45]
M13	8-Bromoguanosine	RT	[45]
M17	2,3,5-Tri-O-acetyl-1,4-anhydro-2-C-(2,6-dioxo-1,2,3,6-tetrahydro-9H-purin-9-yl)pentitol	RT	[45]
M18	2′,2′,5′-Tri-O-acetyl-2-fluoroadenosine	RT	[45]
M19	(2R,3R,4R,5R)-2-(Acetoxymethy)-5-(6-amino-9 h-purin-9-yl)tetrahydrofuran-3,4-diyldiacetate	RT	[45]
Rutin	2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-3-[(2S,3R,4S,5S,6R)-3,4,5-trihydroxy-6-[[(2R,3R,4R,5R,6S)-3,4,5-trihydroxy-6-methyloxan-2-yl]oxymethyl]oxan-2-yl]oxychromen-4-one		[67]
Rutin 01		HBx	[67]
Rutin 08		HBx	[67]

N1	sodium;(6R,7R)-3-(carbamoyloxymethyl)-7-[[(2Z)-2-(furan-2-yl)-2-methoxyiminoacetyl]amino]-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylate	NTCP	[69]
N2	Z)-but-2-enedioic acid;3-(5,6-dihydrobenzo[b][1]benzazepin-11-yl)-N,N,2-trimethylpropan-1-amine	NTCP	[69]
N3	2-methyl-3-[(E,7R,11R)-3,7,11,15-tetramethylhexadec-2-enyl]naphthalene-1,4-dione	NTCP	[69]
N4	4-[(4-chlorophenyl)methyl]-2-(1-methylazepan-4-yl)phthalazin-1-one;hydrochloride	NTCP	[69]
N5	1-[2-(2,4-dichlorophenyl)-2-[(2,4-dichlorophenyl)methoxy]ethyl]imidazole;nitric acid	NTCP	[69]
HBSC-11		hLA	[78]
HBSC-15		hLA	[78]
HBSC-34		hLA	[78]