European Journal of Medicinal Chemistry 57 (2012) 468–482

Contents lists available at SciVerse ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

A highly selective structure-based virtual screening model of Palm I allosteric inhibitors of HCV Ns5b polymerase enzyme and its application in the discovery and optimization of new analogues

Amr H. Mahmoud*, Dalal A. Abou El Ella, Mohamed A.H. Ismail, Khaled A.M. Abouzid**

Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Ain Shams University, 6 b abadeer zeitoun, Cairo 11566, Egypt

ARTICLE INFO

Article history:
Received 28 November 2011
Received in revised form
8 April 2012
Accepted 12 April 2012
Available online 25 April 2012

Keywords:
Hypogen
PLIF
GOLD
Knime
Muse
HCV
Virtual screening
Guided docking
Virtual decoys

ABSTRACT

First structure-based activity prediction model of topologically diverse inhibitors of Palm I allosteric site of HCV NS5b polymerase enzyme is reported here. The model is a workflow of structure-based pharmacophore followed by guided docking. The pharmacophore was constructed using a novel procedure which includes PLIF (protein ligand interaction fingerprint), Hypogen, contact-based pharmacophore and shape constraints. The guided docking was tweaked using both a scoring function of high correlation with activity (ChemPLP) and essential pharmacophore features. Statistically, ROC analysis for the workflow, deploying the novel technique of virtual decoys, yielded AUC of 0.947. Experimentally, the model was used to screen Asinex GOLD database yielding a new hit with a different scaffold which was further confirmed by synthesis and biological evaluation.

1. Introduction

Hepatitis C virus (HCV) is one of the major causes of cirrhosis, hepatocellular carcinoma and liver failure that lead to transplantation [1]. Since its identification in 1989 [2], it has been estimated that more than 170 million people worldwide are infected by HCV. The HCV genome encodes one single poly-protein of approximately 3000 amino acids, depending on the HCV genotype. This poly-protein is then proteolytically processed by viral and cellular proteases to produce structural proteins (core, E1, E2 and p7) and nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A, and NS5B) [3]. The NS5B RNA-dependent RNA polymerase is recognized as a key target for therapeutic intervention [4–7] mainly because it does not have a mammalian counterpart and offers a wide range of possibilities for the discovery of new molecular entities as anti-HCV agents. The inhibitors of this target are either nucleoside or non-nucleoside inhibitors. The latter is further classified into active site inhibitors, allosteric site inhibitors and miscellaneous [4]. Mechanistic and structural studies have revealed the existence of multiple allosteric binding sites [8], and in particular two thumb sites (thumb I and II) and three palm pockets (palm I, II and III) have been identified to date. According to the target site, the different inhibitors will be referred to as palm site I NNIs (PSI-NNIs), palm site II NNIs (PSII-NNIs), palm site III NNIs (PSIII-NNIs), thumb site I NNIs (TSI-NNIs) and thumb site II NNIs (TSII-NNIs)(Fig. 1).

Palm I site allosteric site is targeted by wide variety of inhibitors that have topologically diverse scaffolds: acrylic acids, rhodanines, benzothiadiazines, benzothiazines, benzoisothiazole dioxides, proline sulfonamides, acyl pyrollodines, anthranilic acids and benzodiazepines [9]. This site has been explored structurally in literature in order to find out the binding mode mainly [10] and not for virtual screening purpose. Virtual screening on this site was carried out previously adopting the ligand based 3D-QSAR method using GOLPE [11]. Generally, ligand based methods suffer from the common problem of limited applicability domain. They can't retrieve with reliability any ligands having information more than that they were trained on. This triggered us to develop for the first time a multipurpose predictive structure-based computational model for the inhibitors of this site (i.e. palm I). The model was

* Corresponding author. Tel.: +202 01062656276.
** Corresponding author. Tel.: +202 020122165624; fax: +202 25080728.
E-mail addresses: amr.hamed@pharm.asu.edu.eg (A.H. Mahmoud), khaled.abouzid@pharm.asu.edu.eg (K.A.M. Abouzid).

0223-5234/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2012.04.016

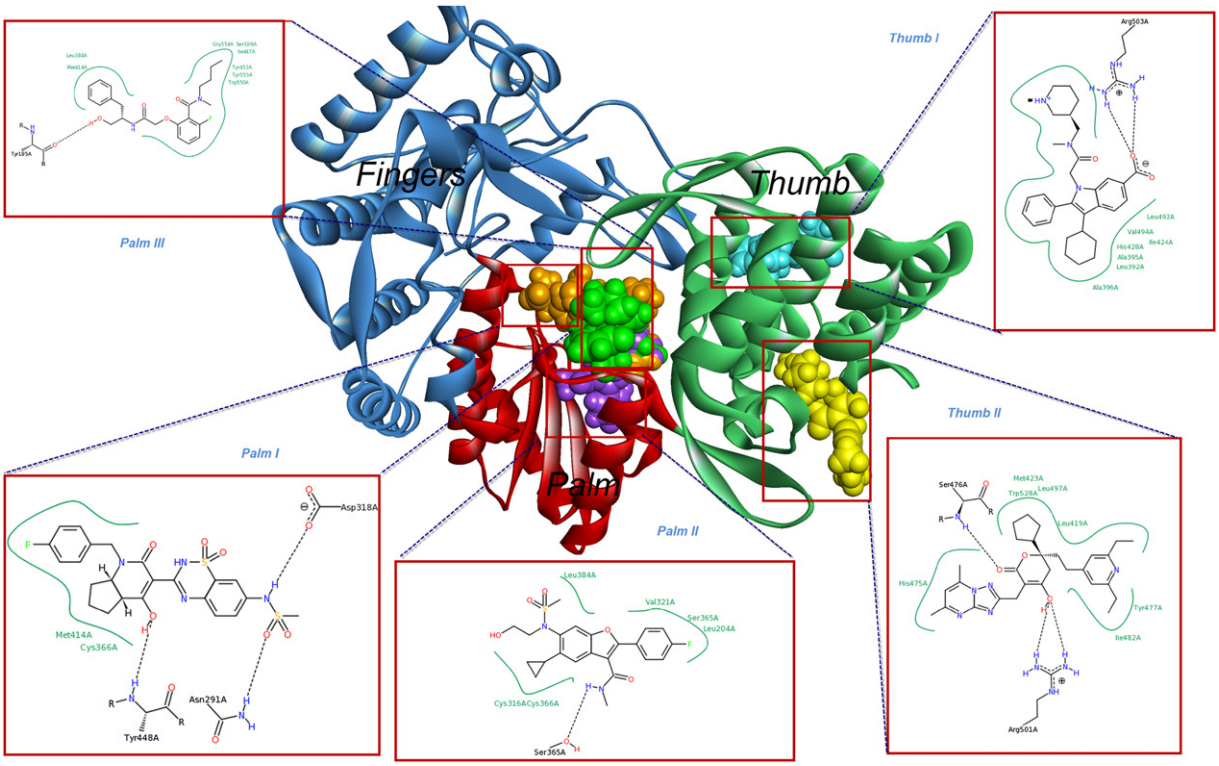


Fig. 1. HCV NS5b different allosteric sites with schematic diagrams of protein ligand interactions of each allosteric site. In this study, we focus on Palm I site. Picture was prepared using Accelrys Discovery Studio 3 and Biosolveit Pose view.

designed while focusing on achieving targets: a-virtual screening enrichment by having superior receiver operator characteristic curves (ROCs) [12] [13]; b-predicting the activity of the inhibitors; c-comprehensive SAR analysis of the site; d-binding mode prediction.

We tried to construct this model using an idea that depends on enhancing the performance of hits finding by directing the search through all the available information related to the palm I site. This information can be translated into the essential chemical features required for the activity and a fit value which correlates well with the activity of the ligands. One tool which can achieve this goal is guided docking [14] which relies on multiple object optimization using genetic [15] or evolutionary algorithm [16]. It uses both types of information we mentioned: the chemical features (pharmacophore restraints) and the correlating fit value which both aim at achieving a balance between favorable interactions with the receptor without deviating from the ideal binding pattern [14]. Moreover, we were aiming at accelerating the process of optimization during docking sampling by initially providing the docking engine with good solutions. In order to achieve this, we preceded docking by structure-based pharmacophore which supplies the docking engine with hits having both good orientation and conformation. In this way, the convergence in the docking algorithm can be achieved rapidly and without losing possible hits due to exceeding the number of iterations during docking sampling. Besides, pharmacophores (even if not structure-based) are normally used before docking in order to rapidly filter ligands which don't have the essential features.

Consequently, we planned the workflow (Fig. 2) to be used as follows: A highly refined structure-based pharmacophore of high selectivity was to be constructed such that the hits found meet the following criteria: 1-Satisify the essential features (those with high probability or propensity in the diverse active inhibitors); 2-Have accurate pose of very small RMSD, thus satisfying the essential interactions with minor need for energy refinement in the actual binding site.

As a second step, knowledge-based docking was to be applied on the hits found by the pharmacophore. It should be guided by two main criteria: 1-Fit value: a scoring function which correlates well with the activity; 2-Soft constraints: provided by a geometrical pharmacophore which is deduced from the above one. The constraints should be applied in order to avoid poses generation which are not satisfying the main features. Thus, we aimed to use docking for refinement of the pharmacophore-provided hits based on the actual ligand–protein interactions. This is unlike pharmacophores which are not based on actual interactions adjustment even if structure-based. In addition, docking is used for the assessment of the hits using a scoring function that correlates with activity and at the same time provide a more accurate binding mode prediction.

Ideally, the docking solutions should not deviate much from that provided by the pharmacophore (i.e. small RMSD deviation) and this was the aim of this workflow: obtaining quick solutions from docking by initially providing it with good seeds as parents for the GA (genetic algorithm) and thus achieving fast convergence. However, this gave rise to an argument of whether docking is actually important or the structure-based pharmacophore is enough and should be followed by just scoring. This was concluded from the results as will be shown.

In order to follow the results sequence, we provided the flow of the protocol illustrated in Fig. 2 and described it briefly as follows:

1 -Genertation of a highly refined hybrid structure-based pharmacophore using the following steps:

A -Superposition of 6 PDB complexes of HCV polymerase inhibitors of allosteric site C and generating a PLIF [17,18] (protein ligand interaction fingerprint) to highlight

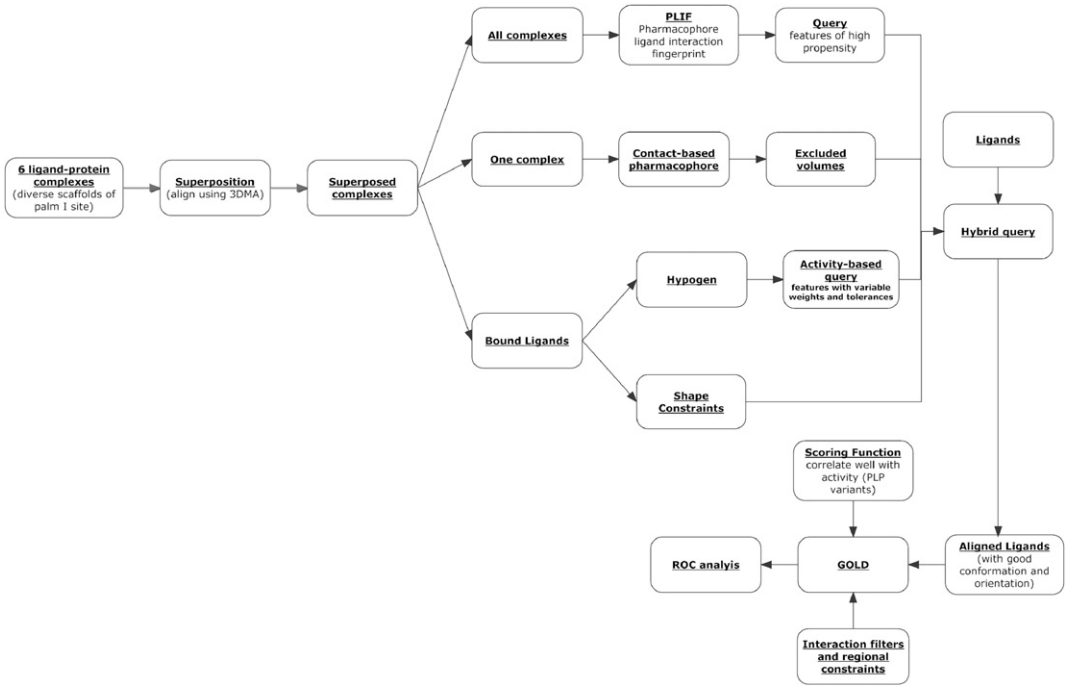


Fig. 2. Workflow used to study Palm I site diverse scaffolds. It consists of structure-based pharmacophore and guided docking modules.

common interactions (those having high probability) between co-crystallized ligands and the proteins.

B -Applying different weights and tolerances to the extracted features according to their importance using activity based pharmacophore (Hypogen) [19].

C -Applying contact-based pharmacophore [20] to extract excluded volumes which represents the regions where the ligand will sterically clash with.

D -Applying shape constraints and adjusting its tolerance to accommodate all ligands.

E -Validation of the pharmacophore by using virtual decoys [21]of high relevance as a more accurate test for ROC analysis.

2 -Selecting a suitable scoring function which can describe the activity well.

3 -Utilizing knowledge-based docking with GA search engine and provide it with the suitable fit value (scoring function selected in the previous step), pharmacophore constraints while keeping the essential structural water molecules.

2. Results

2.1. Pharmacophore generation

As mentioned before, the aim of the structure-based pharmacophore was to act as a fast substitute for docking by initially providing good solutions with good pose accuracy. In addition, it performs the normal pharmacophore function which is reducing the chemical space and enriching the solutions with actives by acting as a pre-filter that screens those ligands not satisfying essential features. This dual functionality was done using a hybrid strategy as follows:

2.1.1. PLIF analysis

The Protein Ligand Interaction Fingerprints (PLIF) tool is a method for summarizing the interactions between ligands and proteins using a fingerprint scheme. Interactions such as hydrogen bonds, ionic interactions and surface contacts are classified according to the residue of origin, and built into a fingerprint scheme. It can be used to obtain the common features shared by group of inhibitors by analyzing the fingerprints. In order to classify interaction fingerprints, each amino acid residue of a protein is broken down into categories, as shown in Fig. 3 where there are 6 types of interactions in which a residue may participate: side chain hydrogen bonds (donor or acceptor), backbone hydrogen bonds (donor or acceptor), ionic interactions, and surface interactions. The most potent of each of these interactions in each category, if any, are considered. The way the fingerprint is calculated is explained in the supplementary data in section 1.

Here in this study, the input data for the PLIF was six X-ray crystal structures of six inhibitors which share the same binding site (Palm I allosteric site of HCV polymerase NS5b genotype 1b) and at the same time have diverse activities and scaffolds. These X-

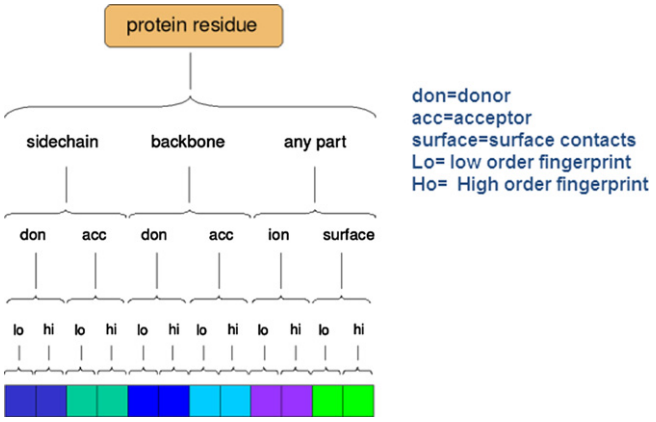


Fig. 3. PLIF (protein ligand interaction fingerprint) way of binning. There are 6 types of interactions in which a residue may participate: side chain hydrogen bonds (donor or acceptor), backbone hydrogen bonds (donor or acceptor), ionic interactions, and surface interactions. Each interaction is further divided into low or high.

Table 1
Non-nucleoside inhibitors of diverse scaffolds acting on Palm I site and their protein data bank codes.

Scaffold type	PDB codes
Acrylic acids	1YVF, 1Z4U
Rhodanines	2AWZ, 2AXO, 2AX1
Benzothiadiazines	2FVC, 2GIQ, 3HHK, 3BSA, 3BSC, 3CDE, 3BR9,3E51, 3CO9, 3CVK, 3H2L, 3H98, 3GYN, 3IGV
Benzothiazines	3CWJ, 3G86, 3H59
Benzoisothiazoles dioxide	3D28, 3D5M, 3H5U, 3H5S, 2GC8, 2JC0, 2JC1
Proline sulfonamides	2GC8
Acylpyrrolodines	2JC0,2JC1
Anthranilic acids	2QE2, 2QE5
Benzodiazepines	3GOL, 3CSO, 3GNV, 3GNW, 3HKW, 3HKY

ray crystals (3HKY, 1YVF, 2GC8, 2GIQ, 2JC0, 2QE5) were obtained from the protein data bank [22] and represent benzodiazepines, acrylic acids, proline sulfonamides, benzothiadiazines, acylpyrrolidines and anthranilic acids respectively. These classes and the PDB codes corresponding to them are shown in (Table 1). Benzothiazines and benzoisothiazoles dioxide based inhibitors are excluded since they are bioisosteres to benzothiadiazines and are likely not to add more information to the structural requirements. We didn't select all the crystal structures and preferred to pick up a complex of each category to simplify method of extraction of common features and avoid over-representation and redundancy as all crystal structures in each category bear the same scaffold.

Briefly, the 6 X-ray structures were superposed and the ligand interaction diagrams (Fig. 4) were created to extract common interactions of high propensity (see section 2 of the supplementary data for full 3D interaction diagrams). According to this analysis, four features were retrieved as common between the scaffolds (as shown in Table 2). These features were mapped on the bound ligands as shown in Fig. 5. Before we give detailed analysis, we saw it is necessary to mention the importance of prior explicit solvation-based molecular dynamics in this case because many of the interactions were not obvious in the bound conformer state. For example, sigma–pi interaction of Met414, hydrogen bonding with Gly449 and generally the binding motifs of hydrophobic interactions were not obvious. Molecular dynamics was carried out previously [10] and results focused on hydrogen bonding only so we replicated the simulation and stressed the other binding motifs (e.g. hydrophobic) in the interaction diagrams (see section 2 in the supplementary data) and mentioned them in Table 2.

2.1.2. Analysis of the binding mode shared by the 6 pdbs

We took 2GIQ benzothiadiazine bound conformer (Fig. 4) as a reference in this analysis and used the four shared features to analyze the binding mode:

1-Hydrophobic feature 1 is represented by Ring A in 2GIQ bound conformer. This feature is essential to bind in the hydrophobic pocket I which consists of amino acid residues Phe415, Met414, Asn411 and Gly410.

2-Hydrophobic feature 2 is represented by hydrophobic tail attached to N1 of Ring B. This feature is essential for binding in the hydrophobic pocket II which consists of amino acid residues Pro197, Leu384 and Cys366 and Tyr415 .The latter showed specific importance in pocket II where it forms pi–pi stacking as a common hydrophobic interaction motif (2GC8 and 2JC0)(see section 2 in the supplementary data for each bound conformer interactions).

3-Two nearby hydrogen bond acceptor features. These features are essential for binding with polar amino acid residues Tyr448,Gly449 and Ser556. The first feature is represented by Oxygen atom of hydroxyl group of 4-hydroxyquinolin-2(1*H*)-one which forms H-bond with Tyr448 directly and indirectly via water bridge with Gly449. The second feature is represented by oxygen atom of sulfonyl group which forms indirect H-bond with Ser556 via water bridge.

The other bound ligands were analyzed the same way and summarized in Table 3 (see section 2 of the supplementary data to visulaize the interactions mentioned).

One important result regarding this analysis is that 2GC8 bound ligand forms hydrogen bond with Gln446 and pi–pi stacking with Tyr415. These 2-amino acids are mutated in genotype 1a where gluatamine is mutated to glutamic acid and tyrosine is mutated to phenylalanine. This hydrogen bonding is absent in case of mutation and this may be the reason why the activity of palm I site inhibitors

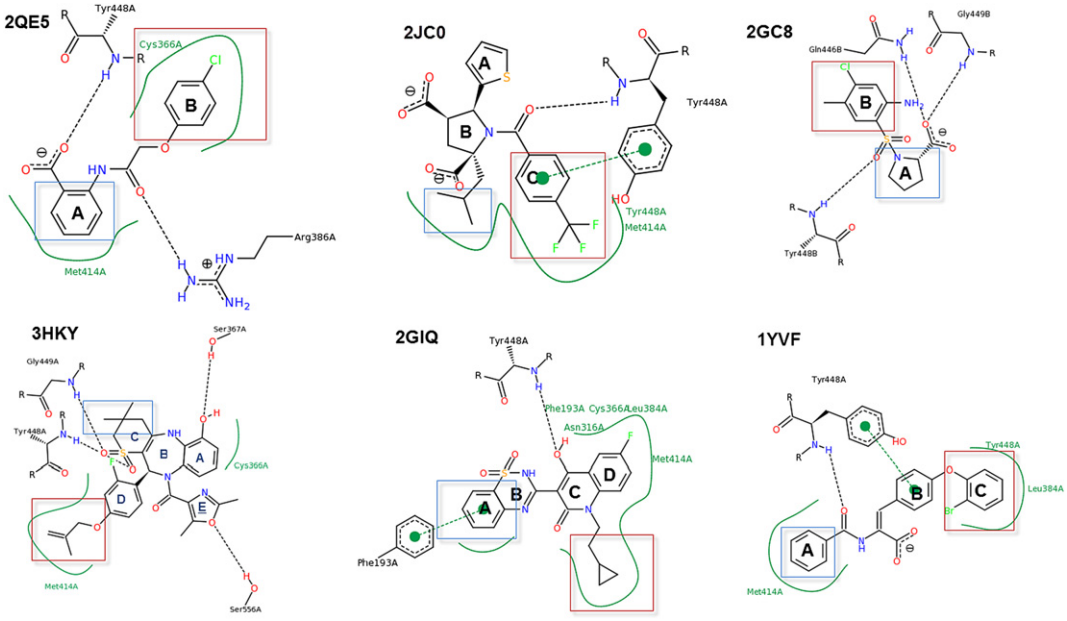


Fig. 4. Schematic diagrams of protein ligand interaction for each scaffold at the palm I site. Red boxes show the features which are responsible for hydrophobic interaction with pocket II while blue boxes show those responsible for hydrophobic interaction with pocket I. Tyrosine448 forms hydrogen bond with all ligands. Met414 is another important amino acid which interacts hydrophobically with the ligand. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Common features of high propensity extracted using PLIF: 2 hydrophobic features and 2 hydrogen bond acceptors.

Feature	Complementary site	Corresponding amino acids	Description	Depiction
Hydrophobic feature 1	Pocket I	Consists of residues Phe415, Met414, Asn411 and Gly410, close to the surface of the enzyme.	A hydrophobic shallow and wide s ub-pocket	Depicted by blue dotted lines
Hydrophobic feature 2	Pocket II	Consists of residues Pro197, Leu384 and Cys366	Deep and narrow hydrophobic sub-pocket	Depicted by red dotted lines
Hydrogen bond acceptor feature 1	Polar amino acid residue	Tyr448	Involved in polar interactions with inhibitor	Depicted by purple circle
Hydrogen bond acceptor feature 2	Polar amino acid residue	Gly449	Involved in polar interactions with inhibitor	Depicted by orange circle

against genotype 1a is generally less than 1b [23]. This phenomenon can be extended to explain the decrease of the activity of other scaffolds, for example benzothiadiazines. The above mentioned hydrogen bond can be formed by the benzothiadiazines in case the ligand is shifted slightly upwards away from the reported bound conformer. This was proved by MD (molecular dynamics) study which confirmed the feasibility of formation of this hydrogen bond. As a result, this finding may aid in understanding the effect of mutation of Gln446 which decreases the activity of the compounds acting on this allosteric site [23,24].

2.1.3. Custom feature generation

Based on the above analysis that indicates that there are two hydrophobic features and two hydrogen bond acceptors in common among the 6 inhibitors, we decided to generate a Hypogen-based pharmacophore using these features. The main reason beyond this was to be able to add different weights and tolerances to the common features according to the activity of the inhibitors and thus becoming an activity based query. Moreover, this enables us to use the technology of Accelrys catalyst pharmacophore functionalities like adding shape and excluded volume functionalities, screening using catalyst engine and finally integrates it with gold docking engine which is supported in the Accelrys Discovery studio environment. Some changes were needed however like the need to modify the hydrogen bond acceptor feature. It was customized such that it maps only oxygen type acceptors and encompass only acceptor origin point (deleting both vector direction and projection point). The first criterion which is the mapping oxygen type acceptors only was done to increase selectivity of the pharmacophore where it has been shown through SAR analysis by Tedesco et al. [25] that the amino group (NH2), as an acceptor, showed a drastic drop in activity. The second criterion which is decreasing the number of the mapping points and restricting it to the origin was done to comply with Hypogen module as it was built using 6 ligands only which was not feasible without minimizing the number of matching points in the feature used [19].

2.1.4. Hypogen

Using the information retrieved from the PLIF analysis, Hypogen [19] was used to generate a pharmacophore using the custom-made hydrogen bond acceptor and hydrophobic features while allowing no conformations generation during the run. The co-crystallized ligands were used in their bound conformation and no conformation generation was needed. Hypogen was used as it modifies weights and tolerances of the features such that they are correlated with the activity trend.

Hypogen was selected as we were aiming here to study topologically diverse inhibitors. It was not intended here to study a congeneric series with same scaffold and minor change in substituents (R-variation) which prefers methods like CoMFA and CoMSIA.

Hypogen algorithm produces three cost values during generation of pharmacophore in order to assess its quality. The first cost value is the fixed cost (also known as the ideal cost). It represents the simplest model that fits the data perfectly. The null cost, also known as no correlation cost, represents the highest cost of a pharmacophore with no features. One should expect that significant pharmacophore should have large difference between these two values. The total cost is calculated for each pharmacophore and should be close to the fixed cost value.

These considerations should be taken into account if the Hypogen model was constructed to estimate the activity. However, in this study, it was used to be a part of hybrid structure-based pharmacophore which aims at accurately placing the ligand in the binding site and achieving a good ranking power. Specifically Hypogen was selected to generate the pharmacophore instead of HipHop(which doesn't take activity into consideration) [19] because it can achieve the purpose of accurately placing ligands

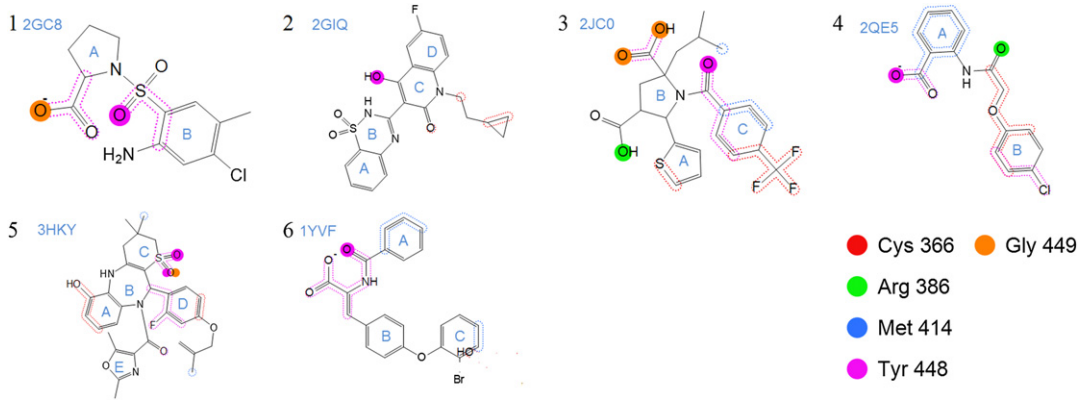


Fig. 5. Applying PLIF on 6 diverse scaffolds of Palm I site. The color-coded amino acids shown in the key are those amino acids interacting with the ligands. The ligands interact with these amino acids by the matching features (have the same color as the interacting amino acids). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Analysis of the 6 bound conformers found in the pdbs chosen to represent Palm I site. Analysis was carried out using features found of high propensity using PLIF analysis.

Scaffold	Hydrophobic feature 1	Hydrophobic feature 2	Two nearby hydrogen bond acceptor features.
Acrylic acid derivative (1YVF)	Ring A of 1YVF forms pi–sigma interaction with Met414	Ring C of 1YVF	O of the carbonyl directly attached to Ring A H-bonds with Tyr448 while carboxylic oxygen H-bonds with Gly449
Proline sulfonamide derivative(2GC8)	Ring A (proline Ring)	Ring B forms pi–pi interaction with Tyr415	O of sulfonamide group H-bonds with Tyr448 while carboxylic oxygen H-bonds with Gly449 and Gln446
Acyl pyrrolidine derivative (2JC0)	Isobutyl group attached to Ring B of 2JC0 bound conformer	Ring C forms pi–pi interaction with Tyr415	O of carbonyl group attached to Ring C H-bonds with Tyr448 while carboxylic oxygen H-bonds with Gly449
Anthranilic acid derivative(2QE5)	Ring A of 2QE5 bound conformer	Ring B of 2QE5 bound conformer	Oxygen of carboxylic group H-bonds with Tyr448
1,5-Benzodiazepine derivative(3HKY)	Ring C and attached methyl group	Hydrophobic tail attached to Ring D	Sulfonyl oxygen forms H-bond with both Gly448 and Tyr449 while Ring E oxygen forms H-bond with Ser556 via water bridge.

effectively. This can be explained as Hypogen was built on the assumption that active molecules should map more features than inactive and mapping should be perfect for most active ligands. In another words, inactivity is attributed to the missing of important feature or, most important in our case, improper orientation of the feature in the space. The Hypogen model can serve this purpose by properly placing the features in the space and assign different weights and tolerances to them. Based on this fact, we selected the best pharmacophore according to the pose accuracy. The best one (we shall call it hypothesis 1 or shortly Hypo1) is shown in Table 4 where it was capable to minimize RMSD as much as possible and was selected for further refinement (Fig. 6). It is worth noting that the pharmacophore has 5 features and not 4 which means that there is an extra feature assigned by Hypogen engine. This extra feature maps to an additional hydrophobic feature found in the most active inhibitor which is the 1,5 benzodiazpine found in 3HKY complex.

The best hypothesis (Hypo1) was modified by changing feature weights manually till reaching a satisfactory RMSD while keeping the features tolerances as they are. The motive beyond the weights selection was the binding pocket analysis. The 2 hydrophobic features were given the largest values (both of them were given 20) since they act as an anchor for the ligands and represent the main bulk of it. Tyr448 role in hydrogen bonding was shown to be more prominent than Gly449 as it forms hydrogen bond directly with the ligand and therefore was given a higher weight (was given a weight of 10) while Gly449 mostly forms hydrogen bond via water bridge (so was given a weight of 5).

Table 5 shows the new modified hypothesis (Hypo2) which indicates a better pose accuracy as shown in Table 6.

Table 4
Hypogen unmodified hypothesis (Hypo1). First table shows 5 features of this pharmacophore and their corresponding weights and tolerances. Second table shows Hypogen different mappings for each bound conformer of the 6 X-ray crystal structures. Fit represents fit value, Est. represents estimated activity and act represents actual activity. Third table shows the three cost values of the selected pharmacophore: Null, fixed and total costs. Correlation value is also shown in the table.

Definition:	Hydrophobic	Hydrophobic	Hydrophobic	Or_1	Or_1
Weights:	1.86857	1.86857	1.86857	1.86857	2.73099
Tolerances:	1.75	1.30	1.45	1.30	1.45

Name	Fit	Cnf/Enan	Mapping	Est	Act	Err	Uncrt
3HKY	7.84	1	+[27 39 1 34 17]	11	10	+1.1	3
2GIQ	7.15	1	+[22 * 4 12 28]	54	50	+1.1	3
1YVF	6.86	1	+[28 23 * 4 17]	110	100	+1.1	3
2JC0	6.34	1	+[2 20 32 11 *]	360	300	+1.2	3
2QE5	5.73	1	+[* 21 8 * 3]	1400	1600	−1.1	3
2GC8	5.53	1	+[8 * * 10 18]	2300	3100	−1.4	3

Null cost: 29.9068 Fixed cost: 50.5808 Total cost: 50.8555 Correl = 0.998492

In our opinion, multiple-parameter optimization can be used to adjust the features weights, tolerances and positions such that it achieves best activity correlation and best pose accuracy at the same time.

2.1.5. Contact-based pharmacophore generation

This technique was developed by Wolber and Langer as a good technique for screening of new compounds instead of computationally expensive docking [20]. The technique was implemented already in ligandscout software [26]. This algorithm extracts information according to certain rules depending on nearby contact residues. It was used here to refine Hypo2 by simply creating excluded volume constraints based on the crystal of the largest inhibitor present (3HKY). Thus, the pharmacophore created above (Hypo2) was just clustered with the excluded volume. Shape constraint was applied in addition to this pharmacophore using the 3HKY bound conformer shape since it was the biggest ligand while decreasing tolerance from the default value 0.5 to 0.3 in order to accommodate all the other ligands. The final form of the pharmacophore (Hypo3) is depicted in (Fig. 7).

The pharmacophore (Hypo3) was employed such that flexible fitting is used and only one feature is allowed to miss. The best mapping was set to the one with the highest fit value and not the one with most features.

To rationalize these additional constraints (excluded volume and shape constraint), we screened a library of highly diverse ligands which is extracted from CAP database 2006 (chemicals available for purchase). This library consists of 9607 diverse ligands. The library was screened using Hypo3 with different features configurations and the results were tabulated in Table 7.

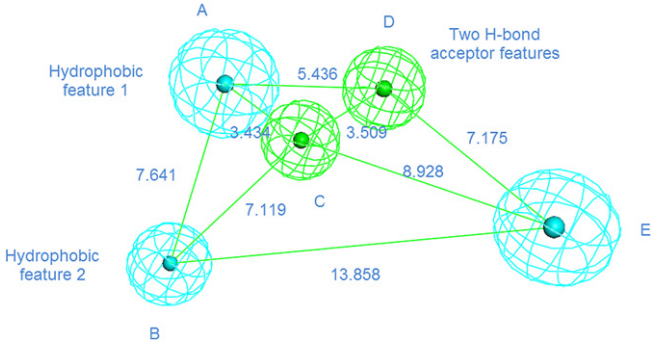


Fig. 6. Hypogen initial unmodified pharmacophore(Hypo1). The cyan-colored features are hydrophobic while the green-colored features represent customized H-bond acceptor features (mainly map oxygen type acceptors). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Different pharmacophore features of Hypogen and their corresponding weights after modification (Hypo2).

Feature	Weight
Hydrophobic feature A	20
Hydrophobic feature B	20
Hydrophobic feature E	1
H-bond acceptor feature C (with Tyr448)	10
H-bond acceptor feature D (with either Gly449 or Ser556)	5

These results show the selectivity of the pharmacophores (the lower the number of hits retrieved, the more selective the pharmacophore). It is obvious that the common features shared among the ligands of the 6 pdbs are four but these are not enough to attain selectivity. Additionally, one can realize that each ligand has some additional feature beside those four, especially in the most active inhibitors. For example, as mentioned before, Hypogen assigned an additional hydrophobic feature to the pharmacophore which is found in the most active 1,5 benzodiazpine (3HKY). Another example is the hydrogen bond acceptor feature found in benzothiadiazines. It has been shown in the previous studies on this class that there is an extra binding pocket (pocket III) [27]where one of the strategies to increase both potency and the selectivity of this class of inhibitors is achieved by substituting Ring D of the benzothiadiazine with methansulfonamide which can form extra two hydrogen bonds with both Asn291 and Asp316 (see section 3 of the supplementary data for details).

Conclusively, the pharmacophore (Hypo3) was tested so far for both pose accuracy and selectivity. Additionally, as we shall see the pharmacophore has a very good enrichment capability and powerful ability to separate actives from decoys. This was tested in the validation section (see below) and proved to be of very good ROC (AUC = 0.884). The predictive power of the selected model should be, however, considered with care because the number of the ligands used to build the Hypogen model was less than 16 and the difference between fixed and null cost is less than 40 [19]. This was solved by following the pharmacophore with guided docking engine supplied with a better correlating scoring function.

The success of this pharmacophore was based on selecting the best of breed from different algorithms: (1) PLIF is considered a method which can deal with many superposing complexes to retrieve common features according to their propensity and significance; (2) Hypogen can adjust weight, tolerance and position of different features to correlate well with activity; (3) The weight variation and setting the selection criteria to the best fit value instead of number of features; (4) Hypothesis refinement techniques using shape constraints and excluded volumes.

2.2. Guided docking

Hits found by the Hybrid structure-based pharmacophore search described above were subjected to Knowledge-based docking using GOLD [15].As mentioned before, using a structure-based pharmacophore before docking has many advantages: (a) it reduces chemical space; (b) it provides docking with good solutions which have correct orientations and conformations, thus time required for sampling is reduced and convergence is achieved rapidly. Thus, guided docking was applied as a refinement tool for initially good poses that provides a good binding prediction and at the same time an assessment tool that rank the compounds according to activity thus achieving an enhanced enrichment. We didn't suffice with the structure-based pharmacophore only as it doesn't take the actual interactions into consideration and thus can't refine pose inside binding site.

Guided docking and not normal docking was used in this workflow although the hits were coming from the pharmacophore and should be in good orientation. This is because of the of the binding site nature which is wide enough to accommodate ligands in different conformations which may not be suitable to illicit activity. On the contrary, guided docking help to keep the good pose in its place without altering it and restricts its function on the energy refinement. Direct approach [14] of this docking was employed using chemical information explicitly to actively guide the orientation of the ligand during sampling. This was achieved using a scoring function which correlates well with activity while applying constraints which penalize divergence from proper

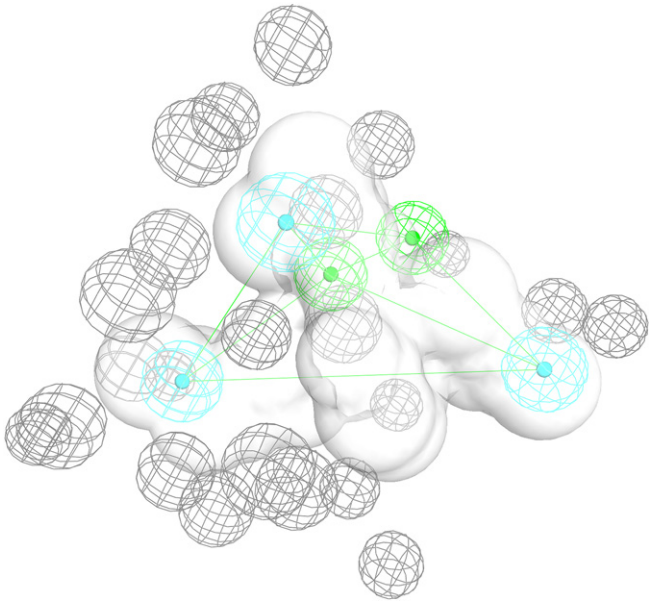


Fig. 7. Structure-based 3D pharmacophore (Hypo3) as an alternative to docking. Gray spheres represent excluded volumes, cyan spheres represent hydrophobic features, and green spheres represent hydrogen bond acceptors of special type (restricted to oxygen type acceptors only). Shape constraint is also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Pose accuracy of different inhibitors using the modified hybrid pharmacophore (Hypo2).

Bound conformer	RMSD[a]
2GC8	1.4
2GIQ	1.2
2JC0	1.3
1YVF	0.9
2QE5	1.1
3HKY	0.7

[a] RMSD = Root Mean Square Distance.

Table 7
The number of hits retrieved from CAP diverse using different configurations for the Hypo3. The selectivity increases as the number of hits decreases.

Features used	Number of compounds retrieved from CAP diverse[a]
5 Features(maximum omitted features = 1)	2337
5 Features (maximum omitted features = 1) and excluded volume	2242
5 Features (maximum omitted features = 1), excluded volume and shape	1601
5 Features (maximum omitted features = 0), excluded volume and shape	431

[a] CAP diverse = Diverse subset of chemical available for purchase.

binding by adding a weighed penalizing energy term to the final score. Thus, sampling of both ligand conformational and orientational space provides a balance between maintaining favorable interactions with receptor without deviating from the appropriate binding mode.

GOLD was chosen as it employs genetic optimization for these multiple parameters at the same time.

Regarding the usage of structure-based pharmacophore before docking despite the fact that docking employs pharmacophore constraints, the docking uses soft constraints which penalize deviation only and will not filter those ligands which are violating the main binding pattern. This was observed during docking optimization runs where it was observed that some ligands can achieve high docking score while not forming many essential interactions. The wide nature of binding site allows many possible interactions (other than essential ones) which can increase the score to the limit it overcomes the penalty and becomes in the top 10% of ranked database. In addition, it was observed that the number of hits missed by docking increases in case of not applying structure-based pharmacophore before it (see the validation study below). Furthermore, the most important reason is that the pharmacophore reduces chemical space rapidly unlike docking which applies pharmacophore while taking interactions into consideration in what is known by multiple parameter optimizations which can takes much more time.

Thus, ligand placement was guided by using:

2.2.1. Scoring function

In our studies regarding Benzothiadiazines (unpublished work); it was found that Moldock score correlates well with activity yielding a q^2 of 0.65. The scoring function, however, was not suitable to guide ligand placement alone but can be used to predict the activity of the ligand when it is accurately placed in the binding site. Thus, the Moldock score applicability was extended in this study to evaluate ligands of diverse scaffolds acting on palm I site. Molegro [16] was used evaluate the bound conformers of the 6 pdbs to find the correlation (Table 8). R^2 was found to be 0.745 which is significant enough to use this docking score for guiding sampling.

Moldock score is a variant of PLP [16]. Since we decided to use GOLD so we tried to use an equivalent scoring function in the software that is based on PLP so we selected CHEMPLP [28] to guide ligand placement. It was checked for its correlation with activity and was found to be true (see validation study below). In general, it was found that Glide score [29], CHEMPLP and Moldock scores [16] are suitable for evaluation of ligands at this site when they are accurately placed (using constraints).

2.2.2. Constraints

Two types of gold constraints were applied, (a) interactions constraints, which allow to define parts of the protein which must form an interaction with the ligand, and (ii) spatial constraints, which allow to define regions in space where some kind of ligand atom must be present. Constraints were applied on the shared interactions pattern only as follows: (1) Regional constraints (hydrophobic spatial constraints) were applied both on pocket I and pocket II; (2) Interaction constraints were applied on Tyr448 and Gly449 using NH of the backbone as H-bond donor feature. Docking was applied using the 3HKY crystal structure while keeping essential water molecules and removing those which clash with other ligands. This was done by aligning all the crystal structures and analyzing bound ligands interactions. In general, it was found that there were some water molecules which clash with Ring D of benzothiadiazine which has no complementary substructures in the other bound conformers. These water molecules are displaceable (not fixed) as proved by experimental data.

Docking was validated by re docking study which confirms capability of GOLD in retrieval of the bound pose. In order to check the reliability of the docking protocol, we performed docking studies on the 6 NS5b polymerase complexes. The docking results were evaluated by the comparison of the predicted docking poses of the ligands and the experimental ones. RMSD between positions of heavy atoms of the ligand in the experimental and calculated structures are taken as a measure of reliability. The best docked conformation was chosen by selecting the conformer with best Chemscore. Random conformer of each bound ligand was docked to prevent bias. Results are shown in Table 9.

In general, the RMSD is very good even when using random conformations because the ligand is guided by pharmacophore which imposes bias in ligand placement. In addition, CHEMPLP proved to be a good evaluator for activity having r^2 of 0.6 which is sufficient for ranking. To confirm that, spearman rank coefficient was measured and was found to be 0.82.

2.3. Virtual decoys and zinc decoys

Virtual decoy [21] dataset was generated in order to evaluate the ability of the pharmacophore model, guided docking and combined workflow to separate active ligands from the decoys. The decoys normally are selected from the zinc database [30–32] such that they are physically similar yet chemically dissimilar to the active ligands. Here in this study, we used in addition the virtual decoys as a more accurate test because they have high TPSA (topological polar surface area) similarity to the active ligand (additional criteria not found in zinc decoys). The virtual DUD of this dataset was created using the strategy adopted by Wallach [21]. A virtual decoy set was generated using Tripos MUSE [33]. This software deploys multiple-parameter optimization technique to generate new ligands. It can generate set of decoys for an active ligand. This method is highly practical and reproducible. It is based on Knime [34] workflow and Evolutionary algorithm which is applied in Tripos Muse [33]. Muse is a de-novo building process which is dependent on Evolutionary Algorithm Inventor (EAI) which was developed by Pearlman group [35]. It is used to generate new chemical entities form a seed set or previous generations. It can be used in conjunction with a single or composite external scoring function. Evolutionary algorithm evolves improved populations of structures by performing modifications on the

Table 8
Moldock scoring of the bound conformers of different PDBs after refinement. The activity of each complex is depicted in minus log units.

Bound conformer	Activity in minus log units	Moldock score
2GC8	−3.49136	−93
2GIQ	−1.69897	−109
2JC0	−2.47712	−122
1YVF	−2	−124
2QE5	−3.21484	−82
3HKY	−1	−170

Table 9
Evaluation of GOLD by docking of the bound ligands in the 6 PDBs (using random conformation) in 3HKY. RMSD is given with respect to the reference and GOLD PLP score is given.

Bound conformer	Activity	Gold.PLP[a]	RMSD[b]
2GC8	−3.49136	−46.0287	1.2
2GIQ	−1.69897	−48.3316	0.43
2JC0	−2.47712	−59.9144	0.34
1YVF	−2	−60.7761	0.6
2QE5	−3.21484	−47.5208	1.4
3HKY	−1	−80.856	0.34

[a] GOLD.PLP = Gold piecewise linear potential scoring function.
[b] RMSD = Root Mean Square Distance.

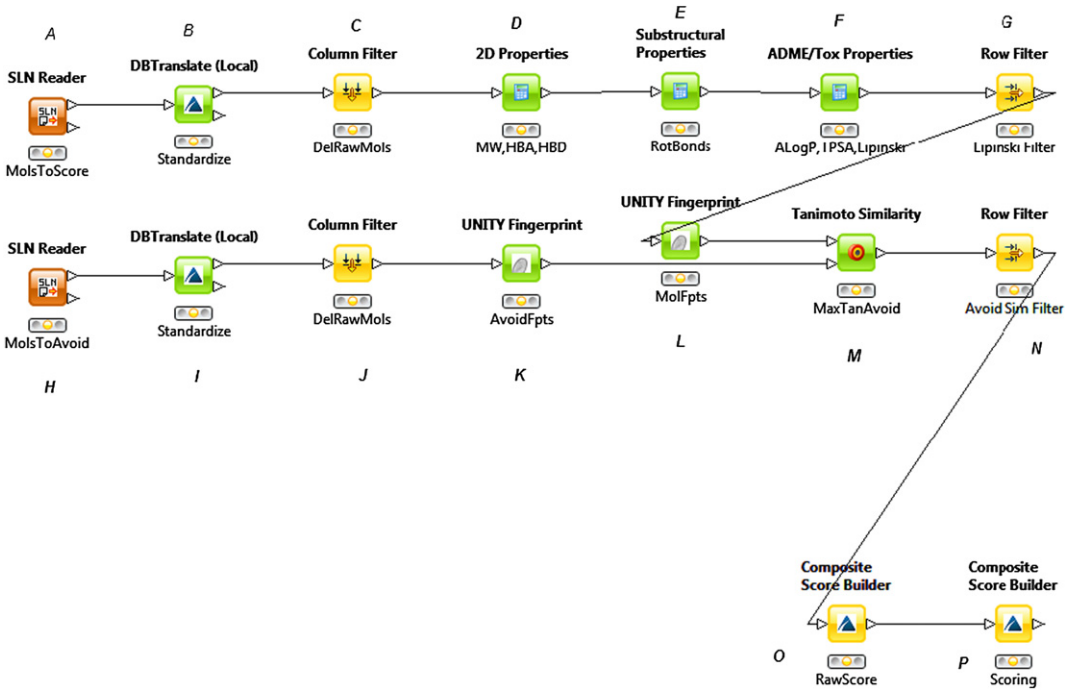


Fig. 8. Workflow for generating virtual decoys. Ligands are generated using evolutionary algorithm in Tripos muse such that they comply with specified MWT, number of hydrogen bond acceptors, number of hydrogen bond donors, number of rotatable bonds and TPSA. The compliance is achieved by applying demerits to any deviation. In order to avoid chemical similarity to the reference, a filter (N) was added which keeps molecules with Tanimoto coefficient less than 0.9.

members of the previous generations (mutations) and swapping successful substructures among different molecules (cross-overs). In this process, molecules evolve via (survival of the fittest) so as to optimize the scoring function. The main advantages of using this protocol described here are: (1) Using the integrated environment of Knime which has many chemistry components from different vendors. In this article, for instance, we used Tripos chemistry nodes; (2) Multiple-parameter optimization uses EA where penalizing the outranged values can be done by Gaussian demerits; (3) Dissimilarity is evaluated in this workflow using the desired fingerprints. Original MACCS can be easily deployed using MOE (Molecular operating environment software) chemistry nodes where Quasar module has fingerprint MACCS node; (4) A fast algorithm which converges quickly. The workflow is depicted in Fig. 8 where it is explained in details the experimental part.

The 6 ligands were used to generate virtual compounds such that they comply with following criteria (i) molecular weight of ±40 Da; (ii) same number of rotational bonds, HBDs (hydrogen bond donors), and HBAs (hydrogen bond acceptors); (iii) cLogP of ±1.0 and (iv) TPSA ±15. The decoys were generated such that each ligand has 36 decoys [31].In addition, zinc decoys (1000 ligands) were used as standard decoys for other screening methods comparison (retrieved from Sybyl folder). Both the virtual decoys and the zinc decoys are supplied as structural data files in the supplementary data.

2.4. Validation

In order to assess workflow performance in screening, it was carried out using the following scenarios: (1) Using GOLD with constraints alone; (2) Using structure-based pharmacophore alone; (3) Using combined workflow by docking hits retrieved by the pharmacophore.

Assessment was done using ROC analysis. ROC curve is a plot of true positives (sensitivity) against false positives (1-selectivity) where selectivity is the rate of false negatives. Therefore, ROC analysis describes sensitivity (Se) as a function of (1 − Sp) where Sp is the specificity [12,13,32].

$$\begin{aligned} \mathrm{Se} &= (\text{number of selected actives})/(\text{total number of actives}) \\ &= \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) \end{aligned}$$

$$\begin{aligned} \mathrm{Sp} &= (\text{number of discarded inactives}) \\ &\quad /(\text{total number of inactives}) = \mathrm{TN}/(\mathrm{TN}+\mathrm{FP}) \end{aligned}$$

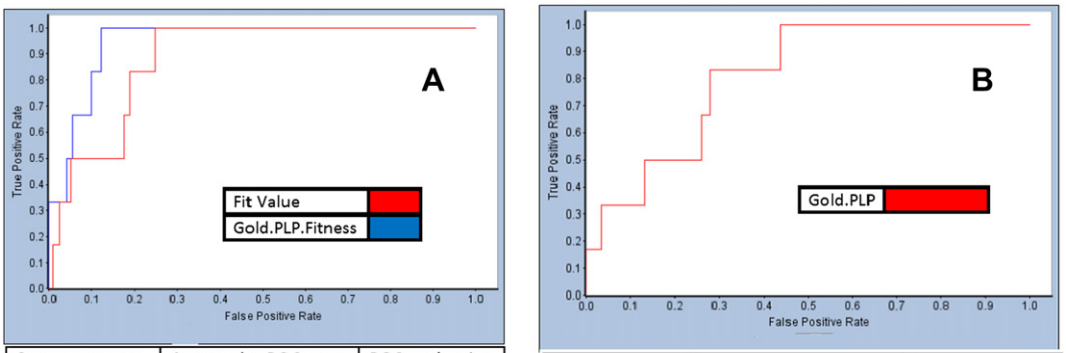
TP is the number of true positives, FN is the false negatives, TN is the true negatives, while FP is the false positives.

Virtual screening protocol is said to have a good discrimination power when most of the active molecules are scored higher than inactive molecules. Actually, an overlap always takes place where some of the actives are scored lower than the decoys. This overlap will lead to the prediction of false positives and false negatives [12,13]. Ratio between actives and decoys will vary according to the score taken as a threshold. That is why we use ROC analysis because it considers all Se and Sp pairs for each score threshold. The protocol is very simple. The ROC curve sets each active score as a threshold, counting the number of decoys within the cutoff and calculating corresponding pair of Se and Sp.

An ideal ROC curve continues as a horizontal straight line to the upper-right corner where all actives and all decoys are retrieved,

Table 10
Comparison of AUC of ROC analysis using pharmacophore alone (Hypo3), guided docking alone and the combined workflow. This was done using either virtual or zinc decoys.

Type of decoys	Pharmacophore alone (Hypo3)	Docking with restraints alone	Combined workflow
Virtual decoys	0.884	0.810	0.947
Zinc decoys	0.851	0.916	0.922



Score property	Area under ROC curve	ROC evaluation
Fitvalue	0.884	Good
Gold.PLP.Fitness	0.947	Excellent

Score property	Area under ROC curve	ROC evaluation
Gold.PLP	0.810	Good

Fig. 9. ROC analysis carried out using virtual decoys. (A) Comparison of ROC curve using pharmacophore (Hypo3) alone (fit value) and combined workflow (Gold.PLP Fitness) using the virtual decoy as dataset. (B) ROC curve using GOLD with constraints only.

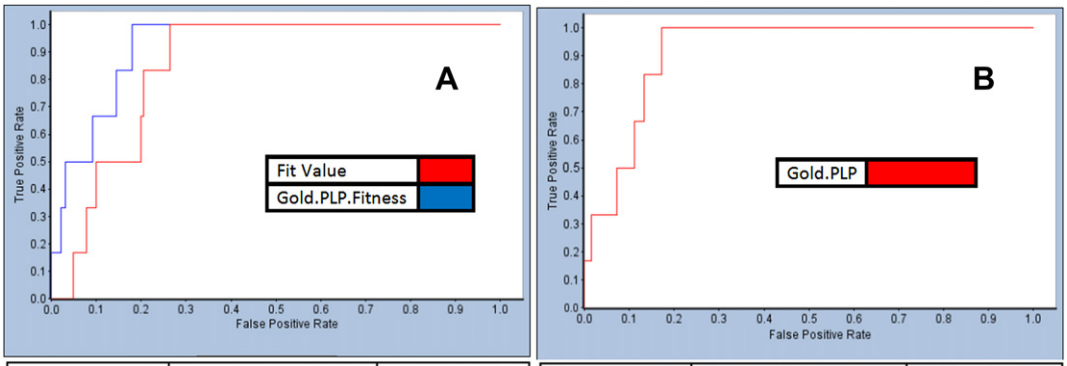
which corresponds to Se = 1 and 1 − Sp = 0. In contrast to that, the ROC curve for a set of actives and decoys with randomly distributed scores tends toward the Se = 1 − Sp line asymptotically with increasing number of actives and decoys.

Area under the ROC curve (AUC)is normally evaluated [13]. In an optimal ROC curve an AUC value of 1 is obtained; however, random distributions cause an AUC value of 0.5. Virtual screening that performs better than a random discrimination of actives and decoys retrieve an AUC value between 0.5 and 1, whereas an AUC value lower than 0.5 represents the unfavorable case of a virtual screening method that has a higher probability to assign the best scores to decoys than to actives [12,13,32].

ROC analysis was carried out using Accelrys Discovery Studio 3 and results are shown in Table 10. ROC curve of the refined pharmacophore (hypo3) using both zinc and virtual decoys shows that the pharmacophore has a good discrimination power. The AUC against virtual decoys was 0.884 and that against zinc was 0.851.This is shown in Fig. 9A and Fig. 10A respectively and represented by fit value. The combined workflow shows an excellent ROC curve with AUC of 0.947 using virtual decoys and 0.922 using zinc (Figs. 9and 10A respectively and represented by GOLD. PLP.Fitness). This was considered excellent for us as the model was built using inhibitors of highly variable activities and was not restricted on highly active ones. Thus, its capability to separate decoys even from moderately active ligands with that superior ROC was a very good result. The superiority of the workflow ROC relies on the fact that it doesn't allow missing of ideal solutions before maximum no of iterations of GA are exceeded. This is done by supplying the GA engine with initial good solutions. The ideality of these solutions is guaranteed by utilizing the structure-based pharmacophore which should succeed in obtaining hits in a good binding pose. The GA function in that case will be restricted on refining the hit according to a correlated fit value.

It was observed that GOLD with constrains performs well when used alone (AUC of 0.916) in case when it was screened against zinc database (Fig. 10B) so further analysis of the results was carried out to assess the performance of using it alone without structure-based pharmacophore.

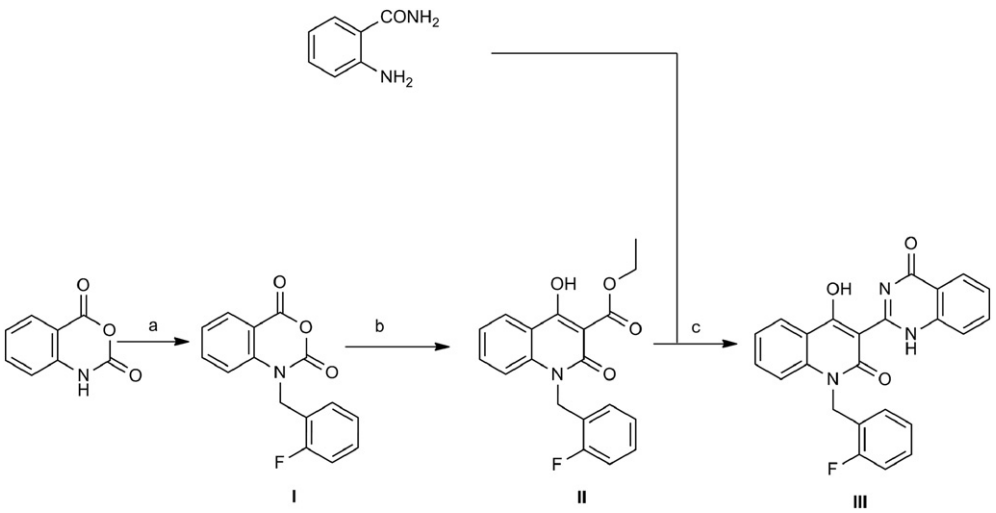
Using Hypo3 to screen Zinc decoys, 429 hits were found. They were all successfully docked after that using GOLD. On the other hand, docking zinc decoys using gold alone retrieved only 391 ligands. This means that GOLD failed to dock some ligands which it successfully did using pharmacophore as a pre-filter. The common ligands which were retrieved by gold alone and the combined workflow were 229 ligands. This means that GOLD missed to dock 200 ligands, which it has already succeeded to do using the workflow. This proposes that GOLD may fail to dock ligands when applying pharmacophore constraints if the ligands were not in a good conformation. This may lead to the loss of ligands which may show good binding. The problem can be solved by increasing



Score property	Area under ROC curve	ROC evaluation
Fitvalue	0.851	Good
Gold.PLP.Fitness	0.922	Excellent

Score property	Area under ROC curve	ROC evaluation
Gold.PLP	0.916	Excellent

Fig. 10. ROC analysis carried out using Zinc decoys. (A) Comparison of ROC curve using pharmacophore (Hypo3) alone (fit value) and the combined workflow (Gold.PLP Fitness) using the virtual decoy as dataset. (B) ROC curve using GOLD with constraints only. Schemes captions.



Scheme 1. Reagents and conditions: (a) *o*-fluorobenzyl bromide,NaH,DMF, 24 h; (b) (i) Diethyl malonate,NaH,DMA, 1 h; (ii) DMA, 120 °C, 17 h; (c) (i)150 °C, 15 min; (ii)KOH,reflux, 3 h; (iii) HCl, pH = 4.

Table 11
Biological evaluation of III and IX using both EC_{50} and CC_{50}.

Compound	CC_{50}[a]	EC_{50}[b]
Modified hit (**III**)	>50 μg/ml	23.7 μg/ml
Optimized hit (**IX**)	>50 μg/ml	2.6 μg/ml

[a] CC_{50} = 50% cytotoxicity concentration.
[b] EC_{50} = 50% effective concentration.

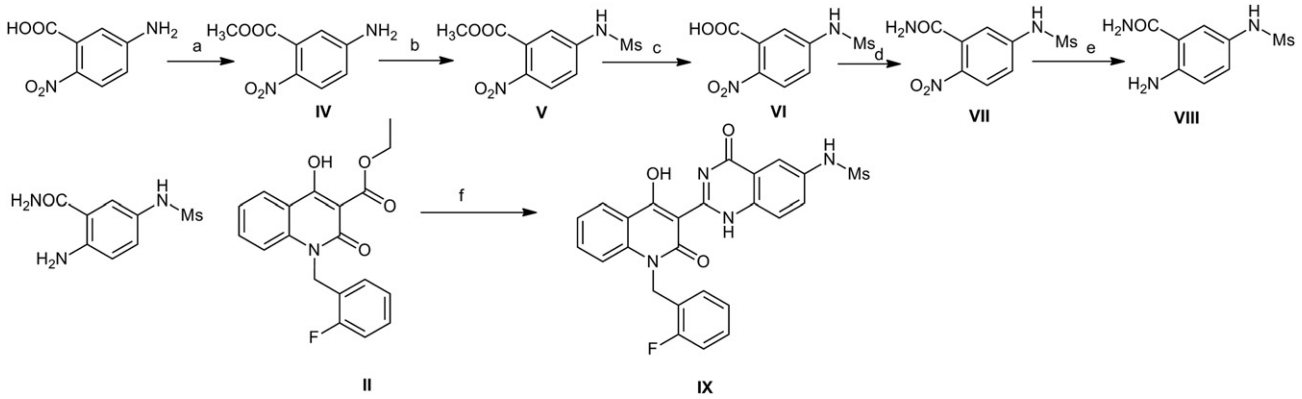
number of iterations of the GA engine but this will increase the time, taking into consideration that we selected the max default value for iteration number which should be slow.

2.5. Virtual screening

Asinex GOLD [36]was screened using the combined workflow. The top 10% of the ranked database was kept. Ligands retrieved were not purchased but assessed chemically using Sylvia score [37]. Ligands with score below 5 (Sylvia score is up to 10 where low scores represent good synthetic feasibility) were further selected and retro synthetically dissected to assess availability of the starting materials. Based on the visual inspection, synthetic feasibility, and non-reported activity, (BAS 00525932) was selected. The IUPAC name of this compound is: 2-(1-Butyl-4-hydroxy-2-oxo-1,2-dihydro-quinolin-3-yl)-1*H*-quinazolin-4-one. One thing which is worth to mention is the presence of the privileged substructure of substituted quinolinone in this hit as it is found in many inhibitors of benzothiadiazine class. This encouraged us to further assess it biologically. Additionally, the ligand found by virtual screening may support the latest discovery of quinoline derivatives lacking sulfonamide moiety at Ring C [38].

2.6. Synthesis and biological evaluation

The hit found via virtual screening of Asinex database was subjected to modification to decrease number of steps required for its synthesis. We preferred N-arylation of the quinolinone to that of N-alkylation. This is because of the synthetic feasibility as will be shown. Regarding the N-alkyl derivative, ethyl esters of 1-alkyl-2-oxo-4 hydroxyquinoline-3-carboxylic acids are synthesized via multistep reaction as reported by Ukraintes et al. [39]. First, ethyl ester of anthranilic acid is alkylated by alkyl halide. Second, N-alkyl ethyl anthranilate was acylated using ethoxymalonyl chloride. Third, the anilide is treated with aqueous KOH. Fourth, potassium salt is treated with acid. N-aryl derivatives, however, can be prepared using readily available diethyl malonate (compared to the expensive lachrymatory ethyl malonyl chloride) and isatioic anhydride in two steps reaction (see supplementary data section 5 for details). The final compound (**III**) can be prepared readily by fusion of Ethyl 1-(2-fluorobenzyl)-4-hydroxy-2-oxo-1,2-dihydroquinoline-3-



Scheme 2. Reagents and conditions : (a) Thionyl chloride, CH_3OH,reflux; (b)Methansulfonyl chloride, pyridine; (c) i-Lithium hydroxide, THF, CH_3OH, H_20, 65 °C, overnight; ii-HCl; (d) i-Thionyl chloride,ii-NH_4OH; (e)Fe/NH_4Cl,CH_3OH, H_20; (f) i-150 °C, 15 min; ii-KOH,reflux,3h; iii-HCl, pH = 4.

carboxylate (**II**) with anthranilamide [40].The general scheme for the compound preparation is depicted in Scheme 1 where (**II**) was synthesized from alkylated isatoic anhydride (I) after fusion with diethyl malonate.

Biological evaluation of the compound was performed on genotype 1b using replicon system [41] and the result is shown in Table 11. It was used as a lead compound for further optimization. One suggestion which was carried out to optimize this ligand is the substitution of 5-position with polar group like methansulfonamide which is known to form H-bond network in Palm I pocket III of benzothiadiazines (Asn291 and Asp318) (see section 3 in supplementary data). This was done according to Scheme 2. It depends on the fusion of 2-amino-5-(methylsulfonamido) benzamide (**VIII**) with Ethyl1-(2-fluorobenzyl)-4-hydroxy-2-oxo-1,2-dihydroquinoline-3-carboxylate (**II**). The first was prepared according to the literature method [42]starting with 5-amino-2-nitrobenzoic acid. The synthesis of this starting material (**VIII**) was based on protecting 5-amino-2-nitrobenzoic acid carboxylic acid by esterification to yield methyl 5-amino-2-nitrobenzoate (**IV**). The ester was alkylated by methansulfonyl chloride to give methyl 5-(methylsulfonamido)-2-nitrobenzoate (**V**) which was further deprotected to give 5-(methylsulfonamido)-2-nitrobenzoic acid (**VI**). The carboxylic acid is then converted to amide 5-(methylsulfonamido)-2-nitrobenzamide (**VII**) and reduced to give 2-amino-5-(methylsulfonamido)benzamide (**VIII**). The final compound (**IX**) potency increased 10 folds where it showed an EC_{50} of 2.6 uM/mL.

3. Conclusion

In this study, we presented a highly selective workflow of structure-based pharmacophore and guided docking that can be used for virtual screening of HCV NS5b polymerase inhibitors of palm I allosteric site.

Regarding the pharmacophore, we presented a novel protocol to generate a highly refined pharmacophore using a combination of different techniques: PLIF, Hypogen, shape constraints and contact-based pharmacophore. The generated pharmacophore was able to define the essential features required for selective activity which are the 2 hydrogen bond acceptors and the three hydrophobic features. It was used as a pre-filter to knowledge-based docking providing it with good solutions as a starting point. In addition, it proved to have very good capability in screening as an alternative to docking. The encompassed PLIF analysis which was based on the refined complexes explained the effect of mutation of Tyr415 and Gln446 on activity of inhibitors of this site.

Regarding docking, PLP proved to be a good scoring function that correlates well with activity of the Palm I site. However, this can be achieved when the pose evaluated is an accurate one. In order to guarantee this accuracy of the ligand placement, the docking engine was guided by pharmacophore during sampling process.

Allover, the workflow of structure-based pharmacophore followed by knowledge-based docking (guided by both a correlating scoring function and pharmacophore constraints) proved to be superior in many aspects: the ROC analysis was excellent, the number of hits missed was decreased, the convergence was very quick and the accuracy of the predicted poses was very high.

The model was validated both statistically and experimentally. Statistically, virtual decoys were employed and proved to be very effective in ROC analysis study. Virtual decoys were generated using a novel knime workflow based on Tripos muse which was given in details. Experimentally, the model was used for screening of the Asinex GOLD database, followed by synthesis of a modified hit and its biological evaluation. Further knowledge-based optimization of the hit based yielded a more potent inhibitor.

4. Experimental data

4.1. PLIF

6 Co-crystallized HCV NS5b Palm I inhibitors were loaded and prepared in MOE (Molecular operating environment software). The complexes were superposed using the Biopolymer protein panel. The ligX algorithm was applied to refine interactions and 2D ligand interaction diagrams were computed for the 6 complexes. A database was created to include the complexes and PLIF was used to compute the interactions and generate fingerprints. After that, a pharmacophore was constructed using query generation panel according to the propensity of the features.

4.2. Custom features

Accelrys Discovery Studio 3 was used to build up a catalyst type query. Acceptor feature was customized to have oxygen atom as the only type of acceptors. In addition, vector and project were deleted. Map atom tool was used so as to map oxygen atom. The feature was added to catalyst dictionary in order to be used for screening.

4.3. Hypogen

Accelrys Discovery Studio 3 3D-QSAR pharmacophore generation tool was used. Only two features were allowed: H-bond acceptor (custom one) and hydrophobic feature. The variable weight and tolerance options were set to true.

4.4. Contact-based pharmacophore

Accelrys Discovery Studio 3 was used to generate a receptor-ligand pharmacophore. 3HKY was used as a crystal structure. Number of pharmacophores was set to 10 such that minimum features were 4 and maximum features were 6. The excluded volume was extracted from the pharmacophore and clustered with the Hypogen pharmacophore after adjusting its position in the binding site. This was carried out as following: the pharmacophore was mapped to 3HKY bound ligand in the binding site; similar pharmacophore (will call it template pharmacophore) was constructed manually in the binding site using the ligand as a template; pharmacophore comparison was performed using template pharmacophore as a reference and Hypogen pharmacophore as input. Finally, shape constraint was added using shape query and minimum similarity tolerance was set to 0.3.

4.5. Gold

Gold 5 within Accelrys Discovery Studio 3 was deployed. Docking was in 3HKY after removing clashing waters. CHEMPLP was selected as a scoring function. GA settings were adjusted to 100,000 operations. 2 regional hydrophobic features were used as constraints. 2 hydrogen bond acceptor features constraints were applied.

4.6. Muse

Tripos Muse 2 was used to generate 36 ligands as virtual decoy set for each active ligand. The workflow in Fig. 8 is described. The workflow starts by Node-A “MolstoScore” which represents those molecules generated by Tripos muse according to EA. They are read by SLN reader. They are designated by moltoscore to indicate that they will be scored according to the specified criteria. Standardization of structures by filling valances and normalizing aromaticity is done by Node-B. This is done by adding new column having standardized molecules; therefore non-standard molecules are

filtered using Node-C. The standardized molecules are evaluated by calculating their molecular weight, number of hydrogen bond acceptors and donors for the generated ligands using Node D, number of rotatable bonds which is a substructure feature using Node E and finally AlogP, total polar surface area and number of Lipinski violations using ADME/TOX properties calculation in node F. The last parameter (number of Lipinski violations) is calculated in order to control ligands which are de-novo generated where those violating the Lipinski rules can be filtered using Lipinski filter in node G. Regarding the reference ligand (active ligand for which we want to create decoys), its properties should be entered using Muse GUI where one can specify whether to allow decoys to strictly comply with the ranges mentioned before (e.g. MWT of ±40 Da) or expand ranges if no solution can be provided. Controlling the generated ligands to meet criteria of the decoys of the reference ligand can be achieved using this workflow. The chemical dissimilarity of the decoys (which can guarantee that they are not highly similar to the actual reference ligand) is achieved by filtering compounds having Tanimoto coefficients of less than 0.9. This is carried out by standardizing the reference ligand just like the generated ligands were standardized in Nodes (H, I, J). After that, Nodes (K–L) are used to calculate unity fingerprint for both reference and the generated ligands where Tanimoto similarity can be calculated between them using Node M and allowing only ligands with Tanimoto coefficients less than 0.9 to be kept using node N. Regarding the scoring of the generated ligands, it depends on the input we provide in muse which is divided into a range and tolerance of each property to be controlled. This is enough to calculate the score for each molecule using Gaussian normalizer for each number and summing up all normalized scores at the end. For instance, if we set the number of hydrogen bond donors to have range 3–4 with tolerance 1 and weight 0.5, the number will be utilized in the scoring function as follows: If the generated molecule has number of hydrogen bond donors between 3 and 4, the score will be zero (zero is the ideal score here) On the other hand, if the number of hydrogen bond donors is less than 3 or greater than 4,score will be assigned according to the Gaussian bell curve which has spread (S) of 1 (specified by tolerance) and asymptotically approach value −1 (which is the lowest value that can be given for this criterion)(see supplementary data section 4.1 for the curve). This can also be done using Tripos score with avoidance in Tripos muse and this is given in details in section 4.2 supplementary data.

4.7. ROC analysis

ROC analysis was carried out in Discovery Studio. Scores were set to fit value in pharmacophore search while GOLD. PLP fitness was used in docking search and in the workflow. Pose grouping was adjusted to name in docking analysis. Actives were given a control value of 1. Because pharmacophore always filters the ligands, the number of total records was used to avoid false results.

4.8. Virtual screening

Asinex GOLD was converted into catalyst database using fast conformation method. Database was screened using ligand pharmacophore mapping tool and not search 3d database because the first allows adjusting number of missing features. The missing features were allowed to be one while choosing the best hit as the one which fits most features.

4.9. Synthesis

General Methods. Starting materials were either commercially available or prepared as reported previously in the literature, unless otherwise noted. Solvents and reagents were used without further purification. Reactions were monitored by TLC, performed on silica gel glass plates containing F-254 indicator (Merck). Visualization on TLC was achieved by UV. Proton and carbon NMR spectra were recorded on a Bruker ARX-300. Chemical shifts (ä) are reported in ppm downfield from internal TMS standard or from solvent references. Mass spectra were recorded on API-SCIEX 2000. HRMS were recorded using an Agilent MSD-TOF (G1969A) connected to an Agilent 1100 HPLC system. Melting points were determined on a Stuart melting point apparatus (Stuart Scientific, Redhill, UK) and are uncorrected.

4.9.1. 1-(2-Fluoro-benzyl)-1H-benzo[d] [1,3]oxazine-2,4-dione (*I*)

It was prepared according to the literature [43]. To a solution of isatoic anhydride (16.3 g, 10 mmol) in 200 ml of DMF, sodium hydride (4.6 g, 110 mmol, 57% in mineral oil, washed with hexane) was slowly added and the mixture was stirred for 1 h at room temperature. 2-fluorobenzyl bromide (20.79 g, 110 mmol) was added and the reaction was allowed to stir for 18 h at room temperature. About two thirds of the solvent were then evaporated in vacuo and the residue poured into 250 mL of ice/water. A precipitate formed which was filtered off, washed with water and dried. It was recrystallized from DCM. Yield: 60%; mp 151–154 °C as reported [43].

4.9.2. Ethyl 1-(2-fluorobenzyl)-4-hydroxy-2-oxo-1,2-dihydroquinoline -3- carboxylate (*II*)

Neat diethyl malonate (1.2 mL, 8.0 mmol) was added slowly to a suspension of sodium hydride (60 percent in mineral oil, 352 mg, 8.8 mmol) in dimethylacetamide. The mixture was stirred at room temperature until the evolution of hydrogen gas ceased, then the mixture was heated to 90 °C for 30 min and cooled to room temperature. A solution of compound 1-(2-fluoro-benzyl)-1*H*-benzo[*d*] [1,3]oxazine-2,4-dione (8.8 mmol) in dimethylacetamide was added slowly and heated overnight at 110 °C. The mixture was cooled to room temperature, poured into ice water, and acidified by cold 10 percent HCl. The solids formed were filtered, washed several times by water, and dried at room temperature under vacuum. Yield: 80%; ^{1}H NMR (300 MHz, DMSO) $\delta = 8.11$ (dd, $J = 7.3$, 1.4, 1H), 7.70–7.09 (m, 6H), 5.43 (s, 2H), 4.39 (q, $J = 7.5$, 2H), 1.31 (t, $J = 7$, 3H). HRMS-FAB m/z [M^+] calcd for $C_{19}H_{16}FNO_4$: 341.1063; found: 341.1060.

4.9.3. 2-[1-(2-Fluorobenzyl)-4-hydroxy-2-oxo-1,2-dihydroquinolin-3-yl] quinazolin-4(1H)-one (*III*)

Synthesis was carried out using general procedure by Ukrainets [40].

The mixture of 0.01 mol of the ethyl 1-(2-fluorobenzyl)-4-hydroxy-2-oxo-1,2-dihydroquinoline-3-carboxylate and 1.36 g (0.01 mol) of anthranilamide is maintained at 150 °C for 10–15 min. The mixture is cooled prior to the addition of 20 ml of alcohol and the careful trituration. The residue of the 2-carbamylanilide is filtered off, washed with alcohol, and dried. It is recrystallized from DMF.

2-Carbamylanilide (0.01 mol) was mixed with 30 ml of the 10% aqueous solution of KOH and boiled using a reflux condenser for 2.5–3 h. The mixture is cooled and acidified with HC1 to the pH 4. The residue of the quinazolone (I), which was separated out, is filtered off, washed with water, dried, and recrystallized from DMF.Yield:90%.^{1}H NMR (300 MHz, d_6 DMSO) $\delta = 8.13$ (dd, 7.6, 1.6, 2H), 7.64–7.08 (m, 10H), 5.45 (s, 2H). ^{13}C NMR (75 MHz, d_6 DMSO) $\delta =$ 165.34, 164.1, 163.12, 159.76, 158.22, 150.88, 138.28, 136.43, 133.50, 131.68, 129.46, 126.94–125.92 (m), 124.84, 123.19, 118.24, 117.21, 115.86, 115.52, 113.52, 104.48, 47.67. HRMS-FAB m/z [M^+] calcd for $C_{24}H_{16}FN_3O_3$: 413.1176; found: 413.1171.

4.9.4. 2-Amino-5-(methylsulfonamido) benzamide (**VIII**)

It was synthesized according to Chin et al. [42].

4.9.5. N-(2-(1-(2-fluorobenzyl)-4-hydroxy-2-oxo-1,2-dihydroquinolin-3-yl)-4-oxo-1,4-dihydroquinazolin-6-yl) methansulfonamide (**IX**)

Synthesis was carried out using the general procedure applied by Ukrainets [40] just as the previous compound. Yield: 45%, HRMS-FAB *m/z* [M^+] calcd for $C_{25}H_{19}FN_4O_5S$: 506.1060, found: 506, 1054. ^{1}H NMR (300 MHz, DMSO-d_6) δ = 8.11 (dd, J = 7.4, 1.6, 1H), 7.70–7.57 (m, 2H), 7.56–7.44 (m, 2H), 7.38 (m, 2H), 7.23 (m, 1H), 7.15–6.99 (m, 3H), 6.44 (s, 1H), 5.77 (s, 1H), 5.31 (s, 1H). ^{13}C NMR (75 MHz, DMSO-d_6) δ = 165.43, 163.81, 163.02, 159.69, 158.25, 150.76, 140.48, 136.45, 134.65, 131.69, 129.41, 126.94–126.40 (m), 124.83, 123.17, 119.45, 118.35, 117.27, 115.86, 115.52, 115.14, 113.54, 104.47, 47.67, 42.90.

4.10. Biology

One day before addition of test compounds, human hepatoma cells (Huh5.2) containing the hepatitis C virus genotype 1b I389luc-ubi-neo/NS3-3′/5.1 replicon sub-cultured in Dulbecco's modified eagle's medium [DMEM (Cat. N°41965039) supplemented with 10% fetal calf serum (FCS), 1% non-essential amino acids (11140035), 1% penicillin/streptomycin (15140148) and 2% Geneticin (10131027); Invitrogen]. Cell lines were grown for 3–4 days in 75 cm^2 tissue culture flasks (Techno Plastic Products), were harvested and seeded in assay medium (DMEM, 10% FCS, 1% non-essential amino acids, 1% penicillin/streptomycin) at a density of 6500 cells/well (100μl/well) in 96-well tissue culture microtiter plates (Falcon, Beckton Dickinson for evaluation of anti-metabolic effect and CulturPlate, Perkin Elmer for evaluation of antiviral effect). The microtiter plates were incubated overnight (37 °C, 5% CO2, 95–99% relative humidity), yielding a non-confluent cell monolayer. The evaluation of the anti-metabolic as well as the antiviral effect of each compound is performed in parallel. Four-step, 1-to-5 test compound dilution series are prepared. Following assay setup, the microtiter plates are incubated for 72 h (37 °C, 5% CO2, 95–99% relative humidity).

4.10.1. Assay protocols

(i) Luciferase assay

For the evaluation of antiviral effects, Viral RNA replication was determined using Renilla luciferase assays (the reporter gene luciferase from the firefly (*Photinus pyralis*) and the coding sequence for ubiquitin were inserted upstream of the neo gene by using standard recombinant DNA technologies). Assay medium is aspirated and the cell monolayers are washed with PBS. The wash buffer is aspirated, 25 μl of Glo Lysis Buffer (Cat. N°. E2661, Promega) is added after which lysis is allowed to proceed for 5 min at room temperature. Subsequently, 50 μl of Luciferase Assay System (Cat. N°. E1501, Promega) is added and the luciferase luminescence signal is quantified immediately (1000 ms integration time/well, Safire^2, Tecan). Relative luminescence units are converted to percentage of untreated controls.

ii) Viability assay

For the evaluation of anti-metabolic effects, the assay medium is aspirated, replaced with 75 μl of a 5% MTS (Promega) solution in phenol red-free medium and incubated for 1.5 h (37 °C, 5% CO2, 95–99% relative humidity). Absorbance is measured at a wavelength of 498 nm (Safire^2, Tecan) and optical densities (OD values) are converted to percentage of untreated controls.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejmech.2012.04.016.

References

[1] D. Lavanchy, The global burden of hepatitis C, Liver International 29 (Suppl. 1) (2009) 74–81.
[2] Q.L. Choo, G. Kuo, A.J. Weiner, L.R. Overby, D.W. Bradley, M. Houghton, Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome, Science 244 (1989) 359–362.
[3] F. Penin, J. Dubuisson, F.A. Rey, D. Moradpour, J.M. Pawlotsky, Structural biology of hepatitis C virus, Hepatology 39 (2004) 5–19.
[4] R.R. Deore, J.W. Chern, NS5B RNA dependent RNA polymerase inhibitors: the promising approach to treat hepatitis C virus infections, Current Medicinal Chemistry 17 (2010) 3806–3826.
[5] F. Legrand-Abravanel, F. Nicot, J. Izopet, New NS5B polymerase inhibitors for hepatitis C, Expert Opinion on Investigational Drugs 19 (2010) 963–975.
[6] H. Li, S.T. Shi, Non-nucleoside inhibitors of hepatitis C virus polymerase: current progress and future challenges, Future Medicinal Chemistry 2 (2010) 121–141.
[7] W.J. Watkins, A.S. Ray, L.S. Chong, HCV NS5B polymerase inhibitors, Current Opinion in Drug Discovery and Development 13 (2010) 441–465.
[8] T.T. Talele, Multiple allosteric pockets of HCV NS5B polymerase and its inhibitors: a structure-based insight, Current Bioactive Compound 4 (2008) 86–109.
[9] M.L. Barreca, N. Iraci, G. Manfroni, V. Cecchetti, Allosteric inhibition of the hepatitis C virus NS5B polymerase: in silico strategies for drug discovery and development, Future Medicinal Chemistry 3 (2011) 1027–1055.
[10] T. Li, M. Froeyen, P. Herdewijn, Insight into ligand selectivity in HCV NS5B polymerase: molecular dynamics simulations, free energy decomposition and docking, Journal of Molecular Modeling 16 (2010) 49–59.
[11] I. Musmuca, A. Caroli, A. Mai, N. Kaushik-Basu, P. Arora, R. Ragno, Combining 3-D quantitative structure-activity relationship with ligand based and structure based alignment procedures for in silico screening of new hepatitis C virus NS5B polymerase inhibitors, Journal of Chemical Information and Modelling 50 (2010) 662–676.
[12] J. Kirchmair, P. Markt, S. Distinto, G. Wolber, T. Langer, 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 22 (2008) 213–228.
[13] N. Triballeau, F. Acher, I. Brabet, J.P. Pin, H.O. Bertrand, Virtual screening workflow development guided by the "receiver operating characteristic" curve approach. Application to high-throughput docking on metabotropic glutamate receptor subtype 4, Journal of Medicinal Chemistry 48 (2005) 2534–2547.
[14] X. Fradera, J. Mestres, Guided docking approaches to structure-based design and screening, Current Topics in Medicinal Chemistry 4 (2004) 687–700.
[15] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, Journal of Molecular Biology 267 (1997) 727–748.
[16] R. Thomsen, M.H. Christensen, MolDock: a new technique for high-accuracy molecular docking, Journal of Medicinal Chemistry 49 (2006) 3315–3321.
[17] T. Sato, T. Honma, S. Yokoyama, Combining machine learning and pharmacophore-based interaction fingerprint for in silico screening, Journal of Chemical Information and Modeling 50 (2009) 170–185.
[18] J.S. Duca, Recent advances on structure-informed drug discovery of cyclin-dependent kinase-2 inhibitors, Future Medicinal Chemistry 1 (2009) 1453–1466.
[19] [In: IUL Biotechnol. Ser., 2000; 2]O.F. Guener (Ed.), Pharmacophore: Perception, Development, and Use in Drug Design, International University Line, 2000.
[20] G. Wolber, T. Langer, LigandScout: 3-d pharmacophores derived from protein-bound ligands and their use as virtual screening filters, Journal of Chemical Information and Modeling 45 (2005) 160–169.
[21] I. Wallach, R. Lilien, Virtual decoy sets for molecular docking benchmarks, Journal of Chemical Information and Modeling 51 (2011) 196–202.
[22] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Research 28 (2000) 235–242.
[23] F. Pauwels, W. Mostmans, L.M. Quirynen, L. van der Helm, C.W. Boutton, A.S. Rueff, E. Cleiren, P. Raboisson, D. Surleraux, O. Nyanguile, K.A. Simmen, Binding-site identification and genotypic profiling of hepatitis C virus polymerase inhibitors, Journal of Virology 81 (2007) 6909–6919.
[24] O. Nyanguile, B. Devogelaere, L. Vijgen, W. Van den Broeck, F. Pauwels, M.D. Cummings, H.L. De Bondt, A.M. Vos, J.M. Berke, O. Lenz, G. Vandercruyssen, K. Vermeiren, W. Mostmans, P. Dehertogh, F. Delouvroy, S. Vendeville, K. VanDyck, K. Dockx, E. Cleiren, P. Raboisson, K.A. Simmen, G.C. Fanning, 1a/1b Subtype profiling of nonnucleoside polymerase inhibitors of hepatitis C virus, Journal of Virology 84 (2010) 2923–2934.
[25] R. Tedesco, A.N. Shaw, R. Bambal, D. Chai, N.O. Concha, M.G. Darcy, D. Dhanak, D.M. Fitch, A. Gates, W.G. Gerhardt, D.L. Halegoua, C. Han, G.A. Hofmann,

V.K. Johnston, A.C. Kaura, N. Liu, R.M. Keenan, J. Lin-Goerke, R.T. Sarisky, K.J. Wiggall, M.N. Zimmerman, K.J. Duffy, 3-(1,1-Dioxo-2*H*-(1,2,4)-benzothiadiazin-3-yl)-4-hydroxy-2(1*H*)-quinolinones, potent inhibitors of hepatitis C virus RNA-Dependent RNA polymerase, Journal of Medicinal Chemistry 49 (2006) 971–983.
[26] S. Boyd, LigandScout, Chemistry World 3 (2006) 69–70.
[27] X. Wang, W. Yang, X. Xu, H. Zhang, Y. Li, Y. Wang, Studies of benzothiadiazine derivatives as hepatitis C virus NS5B polymerase inhibitors using 3D-QSAR, molecular docking and molecular dynamics, Current Medicinal Chemistry 17 (2010) 2788–2803.
[28] O. Korb, T. Stützle, T.E. Exner, Empirical scoring functions for advanced protein–ligand docking with PLANTS, Journal of Chemical Information and Modeling 49 (2009) 84–96.
[29] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, D.T. Mainz, Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes, Journal of Medicinal Chemistry 49 (2006) 6177–6196.
[30] N. Huang, B.K. Shoichet, J.J. Irwin, Benchmarking sets for molecular docking, Journal of Medicinal Chemistry 49 (2006) 6789–6801.
[31] J.J. Irwin, Community benchmarks for virtual screening, Journal of Computer-Aided Molecular Design 22 (2008) 193–199.
[32] J.J. Irwin, B.K. Shoichet, ZINC–a free database of commercially available compounds for virtual screening, Journal of Chemical Information and Modelling 45 (2005) 177–182.
[33] F. Fontaine, M. Pastor, F. Sanz, Incorporating molecular shape into the alignment-free grid-independent descriptors, Journal of Medicinal Chemistry 47 (2004) 2805–2815.
[34] http://www.knime.org/.
[35] M. Feher, Y. Gao, J.C. Baber, W.A. Shirley, J. Saunders, The use of ligand-based de novo design for scaffold hopping and sidechain optimization: two case studies, Bioorganic & Medicinal Chemistry 16 (2008) 422–427.
[36] http://www.asinex.com/Libraries_Gold_Platinum.html.
[37] http://www.molecular-networks.com/products/sylvia.
[38] D.K. Hutchinson, C.A. Flentge, P.L. Donner, R. Wagner, C.J. Maring, W.M. Kati, Y. Liu, S.V. Masse, T. Middleton, H. Mo, D. Montgomery, W.W. Jiang, G. Koev, D.W.A. Beno, K.D. Stewart, V.S. Stoll, A. Molla, D.J. Kempf, Hepatitis C NS5B polymerase inhibitors: functional equivalents for the benzothiadiazine moiety, Bioorganic & Medicinal Chemistry Letters 21 (2011) 1876–1879.
[39] L.V. Ukrainets, O.V. Gorokhova, S.G. Taran, P.A. Bezulyi, A.V. Turov, N.A. Marusenko, O.A. Evtifeeva, 4-hydroxy-2-quinolones. 22.Synthesis and biological properties of 1-alkyl(aryl)-2-oxo-3-carbethoxy-4-hydroxyquinolines and their derivatives, Chemistry of Heterocyclic Compounds 30 (1994) 829–836.
[40] I.V. Ukrainets, S.G. Taran, P.A. Bezuglyi, S.N. Kovalenko, A.V. Turov, N.A. Marusenko, 4-hydroxy-2-quinolones. 18. Synthesis and antithyroid activity of 1-*R*-2-oxo-3-(4-oxo-3*H*-quinazolin-2-yl)-4-hydroxyquinolines, Chemistry of Heterocyclic Compounds 29 (1993) 1044–1047.
[41] G. Yang, M. Huang, Evaluation of compound activity against hepatitis C virus in replicon systems, in: Current Protocols in Pharmacology, John Wiley & Sons, Inc., 2001.
[42] E. Chin, J. Li, A.S.-T. Lui, F.X. Talamas (Eds.), Preparation of Pyrimidine Derivatives as Antiviral Agents, F. Hoffmann-La Roche AG, Switz, 2010, p. 75.
[43] G.E. Hardtmann, G. Koletar, O.R. Pfister, The chemistry of 2*H*-3,1-benzoxazine-2,4(1*H*)dione (isatoic anhydrides) 1. The synthesis of N-substituted 2*H*-3,1-benzoxazine-2,4(1*H*)diones, Journal of Heterocyclic Chemistry 12 (1975) 565–572.