Methods 71 (2015) 4–13

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Exploiting ChEMBL database to identify indole analogs as HCV replication inhibitors

Eleni Vrontaki [a,b], Georgia Melagraki [a,*], Thomas Mavromoustakos [b], Antreas Afantitis [a,*]

[a] Department of Chemoinformatics, NovaMechanics Ltd., Nicosia, Cyprus
[b] Laboratory of Organic Chemistry, Department of Chemistry, University of Athens, Athens 15771, Greece

ARTICLE INFO

Article history:
Received 9 February 2014
Revised 11 March 2014
Accepted 13 March 2014
Available online 27 March 2014

Keywords:
ChEMBL database
HCV replication inhibitors
3D-QSAR CoMSIA
Molecular docking
Virtual Screening

ABSTRACT

Molecular docking, 3D-QSAR CoMSIA and similarity search were combined in a multi-step framework with the ultimate goal to identify potent indole analogs, in the ChEMBL database, as inhibitors of HCV replication. The crystal structure of HCV RNA-dependent RNA polymerase (NS5B GT1b) was utilized and 41 known inhibitors were docked into the enzyme "Palm II" active site. In a second step, the docking pose of each compound was used in a receptor-based alignment for the generation of the CoMSIA fields. A validated 3D-QSAR CoMSIA model was subsequently built to accurately estimate the activity values. The proposed framework gives insight into the structural characteristics that affect the binding and the inhibitory activity of these analogs on HCV polymerase. The obtained *in silico* model was used to predict the activity of novel compounds prior to their synthesis and biological testing, within a Virtual Screening framework. The ChEMBL database was mined to afford compounds containing the indole scaffold that are predicted to possess high activity and thus can be prioritized for biological screening.

1. Introduction

Chronic hepatitis C virus (HCV) is a major disease and constitutes a significant worldwide health concern. Approximately 60–80% of HCV infections evolve in chronic hepatitis that usually leads to liver fibrosis, cirrhosis and hepatocellular carcinoma [1–4]. Globally, up to 3% of the world's population is chronically infected with HCV and 3 to 4 million persons are newly infected every year [5–7]. In the United States, an estimated 2.7 million people are living with chronic hepatitis C [2].

Pegylated alpha Interferon (α-PEG-IFN), alone or in combination with Ribavirin (RBV), is a current effective treatment but only in limited cases since it has serious disadvantages. Thus, the development of new therapeutic strategies to treat HCV-associated hepatitis is a major challenge [8,9]. Due to its serious detrimental effects in the health and its extent of infections, HCV has become the paramount target of antiviral protease inhibitor research, especially the genotype 1.

Hepatitis C Virus (HCV) is a member of the *Flaviviridae* family. It is a small, single-stranded, positive-sense RNA virus with a genome that encodes a single polyprotein of 3010 amino acids with four structural proteins located in the amino terminus and six non-structural proteins encoded in the remainder. Structural proteins made by the HCV include Core protein, E1, E2 and P7, whereas non-structural proteins include NS2, NS3, NS4A, NS4B, NS5A and NS5B. The development of HCV antiviral agents has focused mostly on HCV NS3/4A serine protease and NS5B RNA polymerase [10,11].

NS5B is the RNA-dependent RNA-polymerase and it has the key function of replicating the viral RNA of HCV. Due to its significant action, NS5B has been assigned extensive structural and biochemical characterization [12,13]. Multiple crystallographic structures of HCV NS5B polymerase in several crystalline forms have been determined to date and show that they all have a basic right hand-like structure with fingers, palm and thumb subdomains, similar to common polymerase 3D. Both nucleosides (NIs) and non-nucleoside inhibitors (NNIs) of the most studied target NS5B polymerase for combating HCV have been appeared recently in the literature [14,15].

NIs are modified nucleosides which can be used as substrates at the active site of polymerase to compete with natural nucleosides acting as non-obligate chain terminators. However, the efficiency

Abbreviations: HCV, hepatitis C virus; 3D-QSAR, three-dimensional quantitative structure–activity relationship; CoMSIA, comparative molecular similarity indices analysis; PLS, partial least squares; LOO, leave-one-out; S_{PRESS}, standard error of predictions; ONC, optimal number of components; PRESS, predictive error sum of squares; TSS, total sum of squares; HTS, High Throughput Screening; VS, Virtual Screening.

* Corresponding authors.
E-mail addresses: melagraki@novamechanics.com (G. Melagraki), afantitis@novamechanics.com (A. Afantitis).

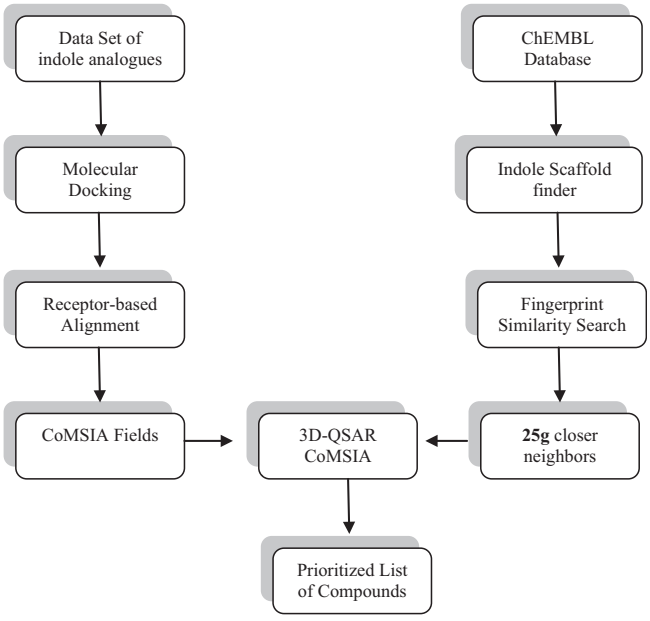
http://dx.doi.org/10.1016/j.ymeth.2014.03.021
1046-2023/

of some nucleoside inhibitors in the clinic has been stained by significant adverse side effects. On the other side, NNIs are chemically diverse small molecules that bind to allosteric sites distinct from the active site. They are non-competitive inhibitors which target the alloenzyme free of substrate and uncomplexed with any other non-structural replicative proteins.

Virtual Screening (VS) is a computational technique that used in discovery to search libraries of small molecules in order to identify those structures which are most likely to bind to a drug target. It can be viewed as the *in silico* analog of High Throughput Screening (HTS) and refers to the evaluation of properties of different molecular scaffolds, such as activity and has been emerged as a reliable, cost-effective and time-saving technique for the drug discovery. VS has already played a significant role in the discovery of some compounds that are now in the clinic or even on the market.

Virtual Screening techniques used for the identification of HCV inhibitors can be subdivided in two broad categories: Ligand-Based Virtual Screening (LBVS) and Structure-Based Virtual Screening (SBVS) [16–21]. For the first category, biological data are explored to identify known active or inactive compounds that will be used to retrieve other potentially active molecular scaffolds based on similarity measures, common pharmacophore or descriptor values. For the second category, the knowledge about the protein target is employed; the candidate ligands are docked into the target and scoring functions are applied to estimate the likelihood that the ligand will bind to the target with high affinity. Several studies have been reported in the literature with the aim of identifying novel potent HCV inhibitors using different *in silico* protocols and workflows [22–33].

In this study, we describe a multi-step framework by combining molecular docking and 3D-QSAR CoMSIA to understand the structural characteristics that affect the binding of 41 N-substituted indole analogs with HCV NS5B GT1b polymerase serving as receptor [34]. These compounds were synthesized and evaluated for activity in subgenomic HCV replicon inhibition assays, but none of them has been co-crystallized with HCV polymerase up to now. All the compounds were docked into the enzyme active site, "palm II", and the docking pose of each compound was subsequently used in a receptor-based alignment for the generation of the CoMSIA fields. A computational workflow was followed in order to build a 3D-QSAR CoMSIA model and is graphically depicted in Scheme 1. This computational workflow gives insight into the structural characteristics that affect the binding and the inhibitory activity of these analogs on HCV polymerase. The validity of the model was established with new molecules possessing known activity and can serve as a valuable rational design tool for the synthesis of more potent inhibitors.



Scheme 1. Computational workflow used in our studies for identification of new compounds-potent HCV inhibitors.

2. Materials and methods

A computational workflow was has been developed to study potent N-substituted indole analogs as inhibitors of HCV replication and to predict the inhibitory activity of indole analogs included in ChEMBL database. Firstly, a dataset of 41 N-substituted indole analogs was docked into the active site of the HCV polymerase for calculating the docking score. Then, the docking pose of each inhibitor was used in a receptor-based alignment for generating the CoMSIA fields. A 3D-QSAR CoMSIA model was subsequently built to accurately estimate the activity and, finally, indole analogs included in ChEMBL database were prioritized for screening.

2.1. Dataset

For the development of the 3D-QSAR model, the dataset consisted of 41 N-substituted indole analogs (Table 1) [34]. These analogs were all tested for inhibition of HCV replication utilizing a subgenomic HCV replicon assay system carrying a luciferase reporter gene, under the same experimental conditions [34]. The structures of N-substituted indole analogs were built using the SYBYL 8.0 molecular sketcher. All the hydrogen atoms were added and the molecules were energy minimized using Steepest Descent, Powell and Conjugate Gradient algorithms (0.01 kcal/mol Å). Gasteiger–Hückel charges were computed for all molecules after energy minimization.

2.2. Preparation of the receptor

As it is already mentioned, there is no X-ray crystal structure of HCV RNA-Dependent RNA Polymerase co-crystallized with the indole analogs. For this reason, we searched from the RCSB Protein Data Bank for an X-ray crystal structure of HCV NS5B GT1b polymerase co-crystallized with an another small-molecule inhibitor. The most recent and with the higher resolution X-ray (PDB ID: 4KB7) was chosen for the docking [35,36] and was prepared using the Protein Preparation Wizard implementation in Schrödinger suite 2009 [37]. The bond orders were assigned, all the hydrogen atoms were added, the disulfide bonds were assigned and all the water molecules that separated by a distance of more than 5 Å from the active site of the enzyme were deleted. The Epik 2.0 implementation was used to predict ionization and tautomeric states of the ligand het groups [38,39]. The hydrogen-bonding network was optimized by reorienting the hydroxyl groups, amide groups of Asn and Gln residues, and by selecting appropriate states and orientations of the imidazole ring in His residues. Finally, using the "impref utility" and the OPLS_2005 force field [40], the hydrogen atom positions were optimized by keeping all the heavy atoms in place. The prepared structure was saved in PDB format in order to be used in the molecular docking.

2.3. Molecular docking

The Surflex-Dock algorithm in SYBYL 8.0 molecular modeling package was utilized for the molecular docking procedure [41]. Surflex-Dock supports an automated flexible docking procedure