Journal of Molecular Graphics and Modelling 56 (2015) 20–30

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Exploration of the structural requirements of HIV-protease inhibitors using pharmacophore, virtual screening and molecular docking approaches for lead identification

Md Ataul Islam, Tahir S. Pillay*

Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria and National Health Laboratory Service, Tshwane Academic Division, Pretoria, South Africa

ARTICLE INFO

Article history:
Accepted 30 November 2014
Available online 5 December 2014

Keywords:
HIV protease inhibitors
Pharmacophore
Molecular docking
Virtual screening

ABSTRACT

Pharmacoinformatics approaches are widely used in the field of drug discovery as it saves time, investment and animal sacrifice. In the present study, pharmacore-based virtual screening was adopted to identify potential HIV-protease ligands as anti-HIV agents. Pharmacophore is the 3D orientation and spatial arrangement of functional groups that are critical for binding at the active site cavity. Virtual screening retrieves potential hit molecules from databases based on imposed criteria. A set of 30 compounds were selected with inhibition constant as training set from 129 compounds of dataset set and subsequently the pharmacophore model was developed. The selected best model consists of hydrogen bond acceptor and donor, hydrophobic and aromatic ring, features critical for HIV-protease inhibitors. The model exhibits high correlation ($R = 0.933$), less *rmsd* (1.014), high cross validated correlation coefficient ($Q^2 = 0.872$) among the ten models examined and validated by Fischer's randomization test at 95% confidence level. The acceptable parameters of test set prediction, such as $R^2_{pred} = 0.768$ and $r^2_{m(test)} = 0.711$ suggested that external predictivity of the model was significant. The pharmacophore model was used to perform a virtual screening employing the NCI database. Initial hits were sorted using a number of parameters and finally seven compounds were proposed as potential HIV-protease molecules. One potential HIV-protease ligand is reportedly confirmed as an active agent for anti-HIV screening, validating the current approach. It can be postulated that the pharmacophore model facilitates the selection of novel scaffold of HIV-protease inhibitors and can also allow the design of new chemical entities.

1. Introduction

Acquired immunodeficiency syndrome (AIDS) is an epidemic disease with an estimated two million deaths each year [1] and remains one of the world's most significant public health challenges, predominantly in low- and middle-income countries. The causative agent of the AIDS is human immunodeficiency virus type-1 (HIV-1) [2–5] which is characterized by extensive and dynamic genetic diversity [6] that has implications for the understanding of viral transmission, pathogenesis and diagnosis, and strongly influences strategies for vaccine development. The HIV polyprotein precursor is encoded by relatively simple genomes consisting of *gag*, *pol* and *env* open reading frames. The *gag* gene encodes the structural capsid, nucleocapsid, and matrix protein; *env* undergoes multiple alternative splicing events to regulatory protein; while, *pol* encodes essential viral enzymes necessary for viral replication. The HIV-1 protease receptor (HIV-1 PR) is an aspartyl protease that is required for proteolytic processing of the *gag* and *gag-pol* polyprotein precursors to yield the viral enzyme and structural proteins and is absolutely indispensible for proper viron assembly and maturation [7]. For this reason this protein is one of the major targets for the design of anti-HIV inhibitors [8] for the treatment of AIDS due to its critical role in virus maturation and replication. HIV-1 PR contains a homodimeric C-2 symmetric structure and each monomer contributes one catalytic aspartic residue along with threonine and glycine residues which are flexible and a flap that favours the binding of substrate and inhibitors. The highly active antiretroviral therapy (HAART) and protease inhibitors (PIs) along with reverse-transcriptase inhibitors have resulted in the unprecedented success of HIV/AIDS chemotherapy [9–12]. However owing to the rapid emergence of drug-resistant HIV-1 variants and transmission of these resistant viral strains along with the adverse side

* Corresponding author at: Department of Chemical Pathology, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia, Pretoria 0007, South Africa. Tel.: +27 123192155; fax: +27 123283600.
E-mail address: tspillay@gmail.com (T.S. Pillay).

http://dx.doi.org/10.1016/j.jmgm.2014.11.015
1093-3263/

effects of currently used HIV-1 PIs, are remain critical factors that limiting the clinical effectiveness of HAART [13,14,15]. Numerous groups worldwide have developed HIV-1 protease inhibitors, showing excellent antiviral profiles [16–22]. Up to now, some clinically approved HIV-1 protease inhibitors including atazanavir, indivanir, nelfinavir and sequinavir are available in the market for HIV treatment but they are very peptide-like and have poor bioavailability. Therefore to overcome these problems, there is a need for the development of new PIs with improved activity against drug resistant variants and excellent pharmacokinetic and safety profiles. The pharmacoinformatics approaches including structure activity relationship (SAR), pharmacophore, virtual screening and molecular docking have become pivotal techniques in the pharmaceutical industry for lead discovery. Many groups have applied the pharmacoinformatics approaches to identify inhibitors [23–29] against HIV protease. Hence the current study explores the binding preferences of the inhibitory molecules of HIV protease in terms of space modelling study and virtual screening along with molecular docking.

A pharmacophore defined as ensemble of steric and electronic features that is required to ensure optimal supra-molecular interactions with a specific biological target and to trigger (or block) its biological response [30]. It also can be stated that the pharmacophore concept is based on the kinds of interaction observed in molecular recognition, *i.e.*, hydrogen bonding, charge, and hydrophobic interaction. The pharmacophore features: hydrogen bond acceptor (HBA) and donor (HBD), hydrophobic (H) and aromatic ring (R) were found to be the key features associated with the selectivity and potency of HIV protease inhibitors. The pharmacophore model can be used in virtual screening to identify potential molecules, predict the activity of the newly synthesized compound before animal experiment; or understand the possible mechanism of action [31,32]. In this study, an attempt was made to identify the pharmacophore hypothesis using the *HypoGen* algorithm [33] based on key chemical features of HIV-protease inhibitors with inhibition constant covering a satisfactory wide range of magnitude. The model was validated using several statistical approaches including Fischer's randomization test and test set prediction. The validated model was utilized for the virtual screening to select the virtual hits from structural database. The molecular docking study was also performed to elucidate the binding interactions and preferred orientation of proposed potential molecules. The significance of the work is clearly reflected by the identification of seven potent lead molecules as protease inhibitors. Among these seven potential HIV-protease ligand one compound is reportedly confirmed as an active anti-HIV agent, thus validating the approach.

2. Materials and methods

The pharmacophore space modelling study is one of the most widely used and versatile techniques to discover novel scaffolds for various targets. Mainly, two types of pharmacophore modelling approaches can be used and adopted for searching novel active scaffolds, ligand-based and structure-based. In the present research ligand-based pharmacophore modelling approach was considered for a set of HIV protease inhibitors with inhibitory constant (K_i).

The Discovery Studio 3.5 (DS) [34] was used for the 3D QSAR pharmacophore, virtual screening and molecular docking studies. The DS is commercially available software containing several module packages and widely used in the pharmacoinformatics drug discovery [35–38]. The *3D QSAR Pharmacophore Generation* module enables the use of structure and activity data for a set of potential HIV-protease ligand to create hypotheses. Two algorithms, *HypoGen* and *HipHop* are used for ligand-based pharmacophore modelling. The *HypoGen* allows identification of hypotheses that are common to the 'active' compounds of training set but not present in the 'inactive' compounds, whilst *HipHop* identifies hypotheses present in 'active' compounds only. In the present work the *HypoGen* algorithm was used to generate the hypotheses.

2.1. Dataset

A dataset of 129 HIV protease inhibitors [39–41] were collected from literature. The experimental K_i values were determined by the same group of authors using fluorescence resonance energy transfer (FRET) method. The whole dataset was divided into training and test set compounds for pharmacophore model generation and validation of generated model respectively. The molecules of the dataset have a wide range of K_i, from 0.0008 to 237.8 nM. The whole dataset was divided into three categories based on their activities values; highly active ($K_i < 1.000$ nM, +++), moderately active ($1.000 \leq K_i < 66.000$ nM, ++) and least active ($K_i \geq 66.000$ nM, +). To select the training set for pharmacophore model in DS basic guidelines laid down by Li et al. [42] were followed. The guidelines are (a) molecules should be selected to provide clear and concise information including structure features and activity range; (b) a minimum of 16 diverse molecules for training set should be considered to ensure the statistical significance and avoid chance correlation; (c) the training set must include the most and the least active molecules; (d) the biological activity data of the molecules should have spanned at least four orders of magnitude. Based on the above criteria 30 compounds were selected as training set ($n_{tr} = 30$, Fig. 1) and the remaining 99 compounds (Table S1 in Supplementary file) were considered as test set (n_{ts}) compounds used for assessing the performance of pharmacophore model. The information concerning the structure and the biological activity of test set compounds is provided in the supplementary information, while all the data regarding the training set molecules are reported in Fig. 1. The three-dimensional coordinates of the compounds were generated using the 2D/3D visualizer [34] of DS. For each compound, the geometries were corrected, atoms were typed and energy minimization was performed based on the modified CHARMm force field [43,44]. The various protocols in the molecular modelling package, DS were utilized for 3D-QSAR modelling, virtual screening and molecular docking studies.

Supplementary Table S1 related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2014.11.015.

2.2. Pharmacophore model generation

The pharmacophore model was developed using *3D QSAR Pharmacophore Model Generation* module of DS. The training set molecules were considered to generate conformation by *Cat-Conf* program of the DS software package. The BEST method was applied during generation of multiple acceptable conformations which provides complete and improved coverage of conformational space by performing a rigorous energy minimization and optimizing the conformations in both torsional and Cartesian space using the poling algorithm [45]. The algorithm best conformer generation considers the arrangement in space of chemical features rather than simply the arrangement of atoms [46]. The *Feature mapping* was used to predict the favourable features for the highly active compounds of the dataset. Mapped features were considered as input features for model generation. Followed by the conformer generation, the algorithm also considers chemical features and conformers, and operates in two modes: *HipHop* and *HypoGen*. *HipHop* generates pharmacophore models using active compounds only, whereas *HypoGen* takes activity data into account and uses both active and inactive compounds in an attempt to identify a hypothesis that is common among the active compounds but not in the