Journal of Molecular Graphics and Modelling 28 (2010) 828–833

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

2D binary QSAR modeling of LPA_3 receptor antagonism

James I. Fells [a], Ryoko Tsukahara [b], Jianxiong Liu [b], Gabor Tigyi [b], Abby L. Parrill [a,*]

[a] Department of Chemistry and Computational Research on Materials Institute, The University of Memphis, Memphis, TN 38152, United States
[b] Department of Physiology, University of Tennessee Health Science Center, Memphis, TN 38163, United States

ARTICLE INFO

Article history:
Received 25 November 2009
Received in revised form 26 February 2010
Accepted 2 March 2010
Available online 7 March 2010

Keywords:
Binary QSAR
Lysophosphatidic acid receptor
Similarity searching

ABSTRACT

A structurally diverse dataset of 119 compounds was used to develop and validate a 2D binary QSAR model for the LPA_3 receptor. The binary QSAR model was generated using an activity threshold of greater than 15% inhibition at 10 μM. The overall accuracy of the model on the training set was 82%. It had accuracies of 55% for active and 91% for inactive compounds, respectively. The model was validated using an external test set of 10 compounds. The accuracy on the external test set was 60% overall, identifying three out of seven actives and all three inactive compounds. This model was combined with similarity searching to rapidly screen libraries and select 14 candidate LPA_3 antagonists. Experimental assays confirmed 13 of these (93%) met the 15% inhibition threshold defining actives. The successful application of the model to select candidates for screening demonstrates the power of this binary QSAR model to prioritize compound selection for experimental consideration.

1. Introduction

Lysophosphatidic acid (LPA) is a lipid mediator that modulates a host of physiological effects [1] through activation of eight G-protein coupled receptors (LPA_1–LPA_8) [2–5]. Subtype-selective LPA antagonists offer various therapeutic potentials. They could be useful tools to reduce obesity, and treat both ovarian and prostate cancers [1,6,7]. The LPA_3 receptor has a distribution localized to the testis, prostate, heart, brain, lung, and kidney [6]. We have previously reported a structure-based approach to identify selective LPA_3 antagonists [8]. Virtual screening utilizing similarity searching and our structure-based LPA_3 antagonist pharmacophore led to a large set of matches. Flexible docking was utilized to narrow this large set of matches to the most promising subset for experimental screening. Although this strategy led to several nanomolar antagonists, flexible docking is moderately computationally expensive. Additional tools that operate in a more high-throughput fashion need to be developed to aid in the selection of compounds to experimentally screen, particularly as we move toward lead optimization efforts.

In this present work, we have built a binary QSAR model to predict LPA_3 antagonism. The model was trained to predict the probability that a compound will be a member of the active class based on seven principal components derived from 2D molecular descriptors that represent structural features for a collection of 109 training compounds. This model was used as a filtering tool to reduce the large set of hits produced in similarity searching. Binary QSAR is well suited for high-throughput screening and has been used in several screening campaigns [9–11]. The addition of a QSAR model to our virtual screening approach avoids the time expense of flexibly docking a large dataset. This now gives a way to rapidly prioritize potential LPA_3 antagonists for experimental confirmation.

2. Methods

2.1. Data set

The activities of 119 structurally diverse compounds were collected from in-house data and the literature [8,12–19]. The members of the dataset were categorized as active or inactive based upon their inhibitory activity. Compounds exhibiting greater than 15% inhibition at 10 μM were classified as 1 or active. This definition of active was used in this study to ensure a sufficient number of actives in the training set to produce an effective model. All other compounds were classified as 0 or inactive. This criterion divided the training set into 36 active and 83 inactive compounds.

2.2. Training and test set

The 119 member dataset was clustered using the clustering algorithm in the Molecular Operating Environment (MOE) software [20]. The compounds were clustered based on their similarity at a Tanimoto coefficient of 70%. Similarity calculations were performed using the MACCS fingerprint implemented in MOE. A diverse subset selection was performed using MOE

* Corresponding author. Tel.: +1 901 678 4414; fax: +1 901 678 3447.
E-mail address: aparrill@memphis.edu (A.L. Parrill).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.03.002

to produce a training subset and a test subset. The compounds in the individual clusters were ranked based on their similarity to each other. The middle compound in clusters with three or more compounds was extracted and placed in the test set. Ninety-two percent of the compounds were placed in the training set and the remaining compounds were placed in the test set.

2.3. Principal components

Numerous structural descriptors can be calculated that reflect various aspects of chemical structures and physical properties. In fact, far more descriptors can be calculated than there are observed activities in our training set. In order to avoid overfitting and chance correlations, principal components analysis (PCA) was used as a variable reduction method. PCA offers the benefit of producing a small set of orthogonal (therefore by definition, uncorrelated) variables that represent the maximum variability in the input descriptor set with each added principal component. The input descriptor set for PCA included the184 two-dimensional (2D) molecular descriptors that can be calculated using the MOE descriptor tool. Two-dimensional molecular descriptors are numerical properties calculated from the connection table representation of a molecule. The 184 descriptors can be divided into seven categories including 14 physicochemical property descriptors, 18 subdivided surface area descriptors, 41 atom and bond count descriptors, 16 Kier and Hall connectivity and Kappa shape index descriptors, 33 adjacency and distance matrix descriptors, 12 pharmacophore feature descriptors, and 50 partial charge descriptors which were calculated using the MMFF94x forcefield partial charge assignments. The first seven principal components produced by PCA analysis of these 2D descriptors were used in the binary QSAR model construction process.

2.4. Binary QSAR analysis

Binary QSAR is implemented in the MOE software suite. Binary QSAR is based on a Bayesian inference technique. This method estimates the probability density classifying the compounds as active or inactive. The model is then evaluated using internal validation and an external test set. The quality of the model was assessed using four parameters (1) overall accuracy, (2) accuracy on active compounds, (3) accuracy on inactive compounds, and (4) significance (p-value) representing likelihood that similar accuracies could be obtained by chance.

2.5. Screening assay

The screening assay was performed as previously described [21]. Compounds were purchased from ChemBridge (San Diego, USA). The test compounds were dissolved in methanol and stored as 10 mM stock solutions. Intracellular Ca^{2+} mobilization was measured for LPA receptors stably expressed in RH7777 cells using a FLEXstation II (Molecular Devices, Sunnyvale, CA). Changes in the intracellular Ca^{2+} concentration were measured using a fura-2 dye by determining the ratio of emitted light intensities at 520 nm in response to excitation at 340 and 380 nm. Cells were plated and cultured overnight and transferred to serum-free medium for 6 h prior to assay. Antagonist activity was measured by treating the cells with 200 nM LPA and 10 μM test compound. Results were normalized to 200 nM LPA in the absence of test compound. Determinations were made in triplicate and presented as the average.

3. Results and discussion

3.1. Chemical diversity of LPA_3 dataset

An effective LPA_3 antagonist QSAR model requires a dataset of structure activity data on which to train the model. There is a size-

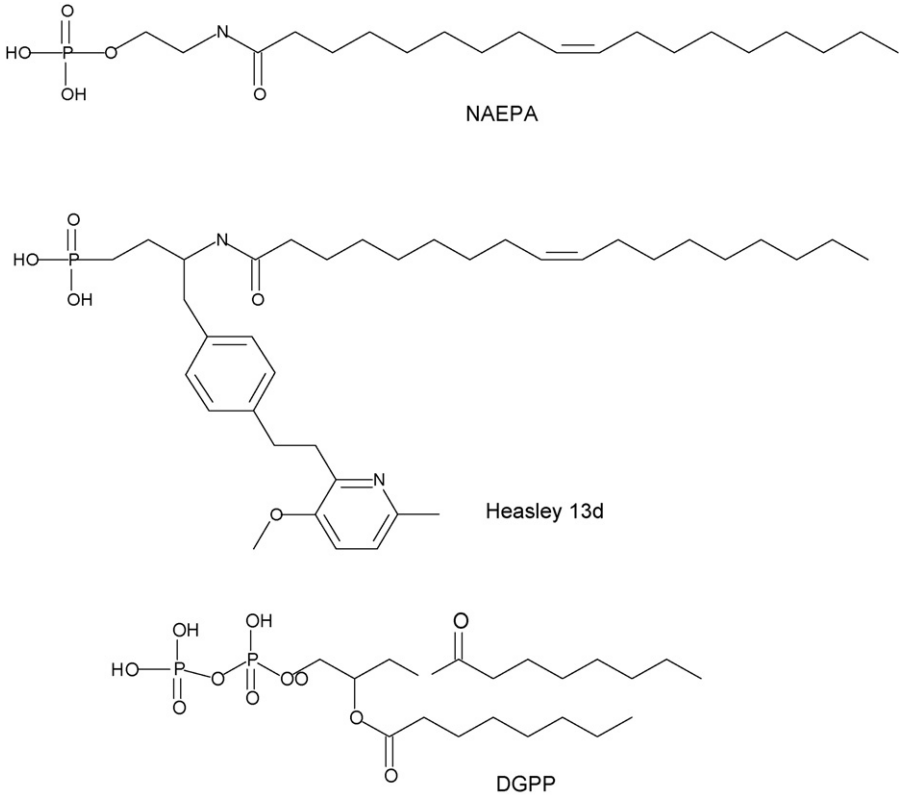

Fig. 1. Previously identified LPA analogs [12,15,17].