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-

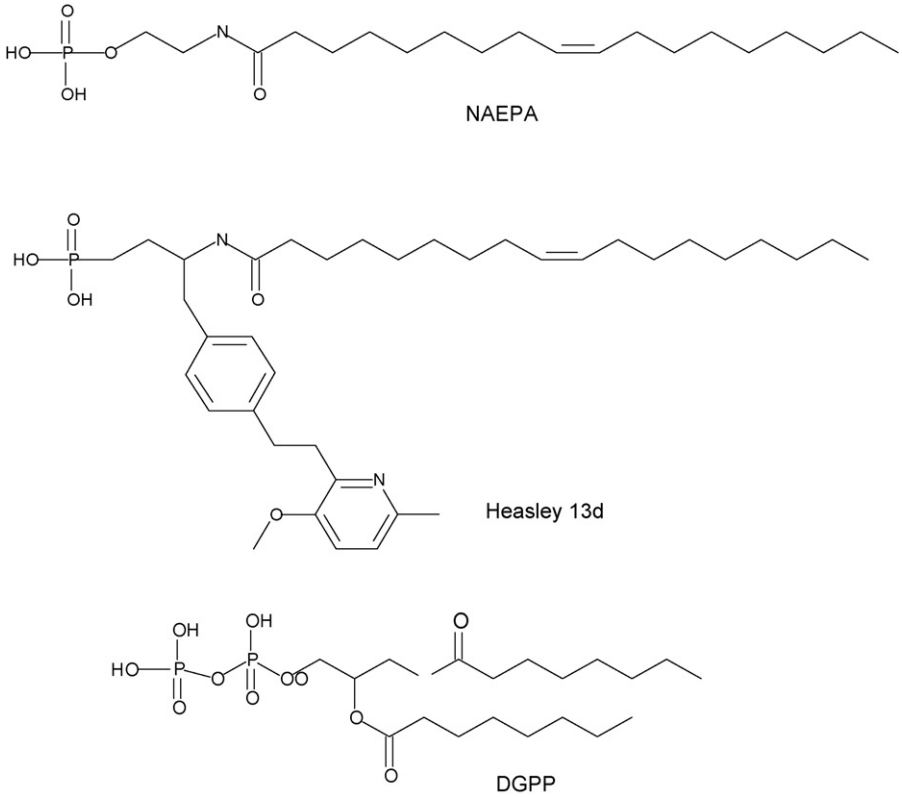

Fig. 1. Previously identified LPA analogs [12,15,17].

able amount of structure activity data on LPA analogues. Fig. 1 shows the structures of some of the identified LPA antagonists. The structure activity relationships (SAR) have been previously reviewed [22,23] and are briefly summarized here. LPA receptor antagonists were developed by making changes to the polar head group and altering the length and number of hydrophobic chains. This is the case with DGPP, an $LPA_{1/3}$ antagonist. DGPP has an additional phosphate attached along with two chains of only eight carbons. Related fatty alcohol phosphates have been synthesized and explored, with 14 carbons producing optimal LPA_3 antagonism [12,13,19]. Synthetic LPA analogs using N-acyl ethanolamine phosphoric acid (NAEPA) as the lead has led to a series of LPA antagonists by making substitutions at the second carbon [15,16,24]. While the majority of the LPA antagonists reported are lipid-like compounds, several non-lipid LPA antagonists have recently been reported (Fig. 2). Ki16425, an $LPA_{1/2/3}$ antagonist, was the first non-lipid antagonist to be reported [18]. Our groups recently identified the LPA_3-selective antagonists NSC 161613 and H2L5747876 [8,21]. These are the only selective non-lipid LPA_3 antagonists reported to date.

Compounds that had been tested for their ability to antagonize LPA action at the LPA_3 receptor, including those shown in Figs. 1 and 2, were collected from the literature in efforts to create a dataset of diverse compounds. The diversity of the dataset determines the applicability domain of our QSAR model [25]. The dataset consists of 36 active and 83 inactive compounds from a diverse array of structural classes. Fig. 2 shows some of the non-lipid compounds for this data set (complete set shown in Supplemental Table 1). Before building a binary QSAR model, the data set was divided into two subsets: a training subset (29 actives, 80 inactives) and a test subset (7 actives, 3 inactives). The compounds were partitioned into these subsets to optimize the diversity of structures in both the training and test subsets. Rational selection of the test set improves the reliability of the QSAR model [26]. The structural diversity in the entire dataset was assessed using the clustering algorithm in MOE. A similarity threshold of 70% yielded 20 clusters (Table 3). Including representatives from all clusters improves the diversity of both the training and test subsets and provides a QSAR model with a broader applicability domain for identifying new antagonists. Compounds in each subset were then ranked using the diverse subset algorithm implemented in MOE. They were ranked based on their similarity to each other. Representative cluster members were extracted for the test subset from the five clusters having more than three compounds. The diversity ranking of each cluster was used to select a compound having a Tanimoto coefficient of 50% or greater similarity to the group. A test subset selected in this fashion is then more reflective of the training subset and provides better validation of the model.

3.2. QSAR parameters

Selection of molecular descriptors is an important step in QSAR modeling. Out of 398 molecular descriptors available in MOE, the 184 2D descriptors formed the basis for calculation of the principal components that were used as input for the binary QSAR model construction. The benefit obtained by selecting only 2D descriptors is that they are independent of the conformation selected for input to the descriptor calculation. This allows calculation of activity probability for new structures very quickly, without requiring a molecular alignment process to ensure applicability of the model to the new structures. The limitation of 2D descriptors is their tendency to represent global characteristics of molecular structures. These descriptors included connectivity indices, molecular weight, and log *P*. Table 1 shows the description and relative weight of the ten molecular descriptors contributing most substantially to the principal components. Partial charge descriptors are strongly represented, reflecting the variability in charge and charge distribution observed among the training set members, which include compounds with zero, one and two ionizable groups. The smoothing parameter and condition limit were both left at default settings. The maximum number of principal components was set at seven.

Table 1
Description and importance of the top 10 molecular descriptors contributing to the principal components utilized in the binary QSAR model for LPA_3 antagonism.

wt%	Name	Description
11.03	Q_PC−	Total negative partial charge
10.50	Q_VSA_POL	Total polar Van der Waals surface area
10.33	Q_PC+	Total positive partial charge
10.31	Q_VSA_PPOS	Total polar positive Van der Waals surface area
9.96	Q_VSA_FPPOS	Fractional polar positive Van der Waals surface area
9.78	Q_VSA_FHYD	Fractional hydrophobic Van der Waals surface area
9.78	Q_VSA_FPOL	Fractional polar Van der Waals surface area
9.54	Q_VSA_NEG	Total negative Van der Waals surface area
9.49	Q_VSA_PNEG	Total polar negative Van der Waals surface area
9.28	Q_VSA_FPNEG	Fractional polar negative Van der Waals surface area

Overfitting of a model typically causes poor performance due to the high potential for chance correlations to occur when the number of adjustable parameters approaches or even exceeds the number of training examples [27,28]. An overfitted model is able to accurately predict the training set but unable to predict activity for new molecules. Here we use principal components analysis as a variable reduction strategy to avoid overfitting.

3.3. Model validation

The quality of the model fit to the training data was assessed using percent accuracy and significance as internal metrics. The predictive ability of our model was validated using an external test set. Table 2 summarizes the validation of our model. Our binary QSAR model has an internal overall accuracy of 82%, with subset accuracies of 91% on inactives and 55% on actives in the training set. The overall accuracy has a significance of 8.4×10^{-5}, indicating a very low probability that a similarly extreme result could have been obtained by chance. The model was similarly accurate when applied to make predictions on our external test set. The predictive test set accuracies were 100% for actives, 43% for inactives, and 60% overall. Analysis of our test set revealed that the four active compounds misclassified as inactive belong to clusters two, eight, and nine. The majority of training set members from these clusters are predominantly inactive compounds, 12 of 18, 11 of 14, and 40 of 48, respectively. In fact, two-thirds of the training set compounds are inactive. These observations suggest that further improvements in accuracy will be obtained by identifying additional compounds to better balance the proportions of active and inactive compounds within each cluster, and in the training set as a whole. These results

Table 2
Accuracy of binary QSAR model in predicting LPA_3 antagonist activity.

Dataset	Number of compounds		Accuracy (%)	Significance (*p*-value)
	Observed	Correct predictions		
Training set				
Active	29	16	55	4.5×10^{-6}*
Inactive	80	73	91	
Overall	109	89	82	8.4×10^{-5}
Test set				
Active	3	3	100	
Inactive	7	3	43	
Overall	10	6	60	

* *p*-Value represents likelihood that BOTH active and inactive accuracies could be achieved by chance.

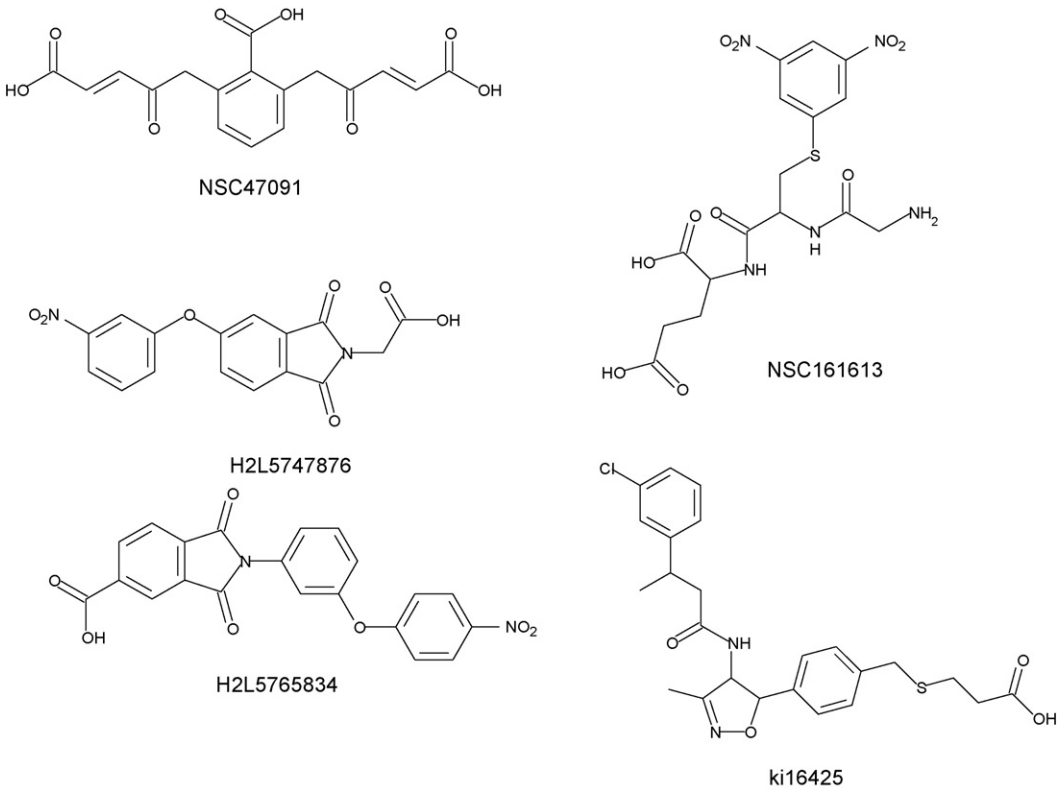

Fig. 2. Representative non-lipid structures in the dataset [8,18,29].

illustrate that predictive accuracy of a QSAR model is dependent on the diversity of the training set and indicate that this model is capable of identifying actives from within the range of structural diversity represented in the training set with 50% accuracy (Table 2).

In our previous study, we used a workflow consisting of similarity searching, docking, and pharmacophore matching to select compounds. A second test set was selected to evaluate inserting our binary QSAR model in this workflow to improve the selection of compounds. Similarity searching in the ChemBridge database (San Diego, USA) using MACCS fingerprints was performed using two active compounds, H2L5765834 and H2L7724589 (Supplemental Table 1), from clusters two and three with a threshold of 80% similarity based on the Tanimoto coefficient. These two clusters were well populated with both active and inactive compounds. Fourteen compounds predicted to be active by our QSAR model were chosen from a total of 955 similarity search matches for experimental screening (Fig. 3). Selection was based on predicted probability of activity greater than 0.8 with preference given to compounds overlapping in the hitlists for the two similarity search target molecules. This combination of similarity searching and binary QSAR prediction can be completed in just a few minutes, in contrast to our prior

Table 3
Breakdown of 20 clusters into training and test sets for LPA_3 antagonist QSAR model development.

Cluster	Training set		Test set	
	Number of compounds per cluster	Active/inactive	Number of compounds per cluster	Active/inactive
1	5	3/2	1	1/0
2	16	5/11	2	1/1
3	12	8/4	2	2/0
4	1	1/0		
5	1	0/1		
6	1	0/1		
7	2	0/2		
8	12	1/11	2	2/0
9	45	7/38	3	1/2
10	2	2/0		
11	2	0/2		
12	1	0/1		
13	1	0/1		
14	1	0/1		
15	2	1/1		
16	1	0/1		
17	1	1/0		
18	1	0/1		
19	1	0/1		
20	1	0/1		

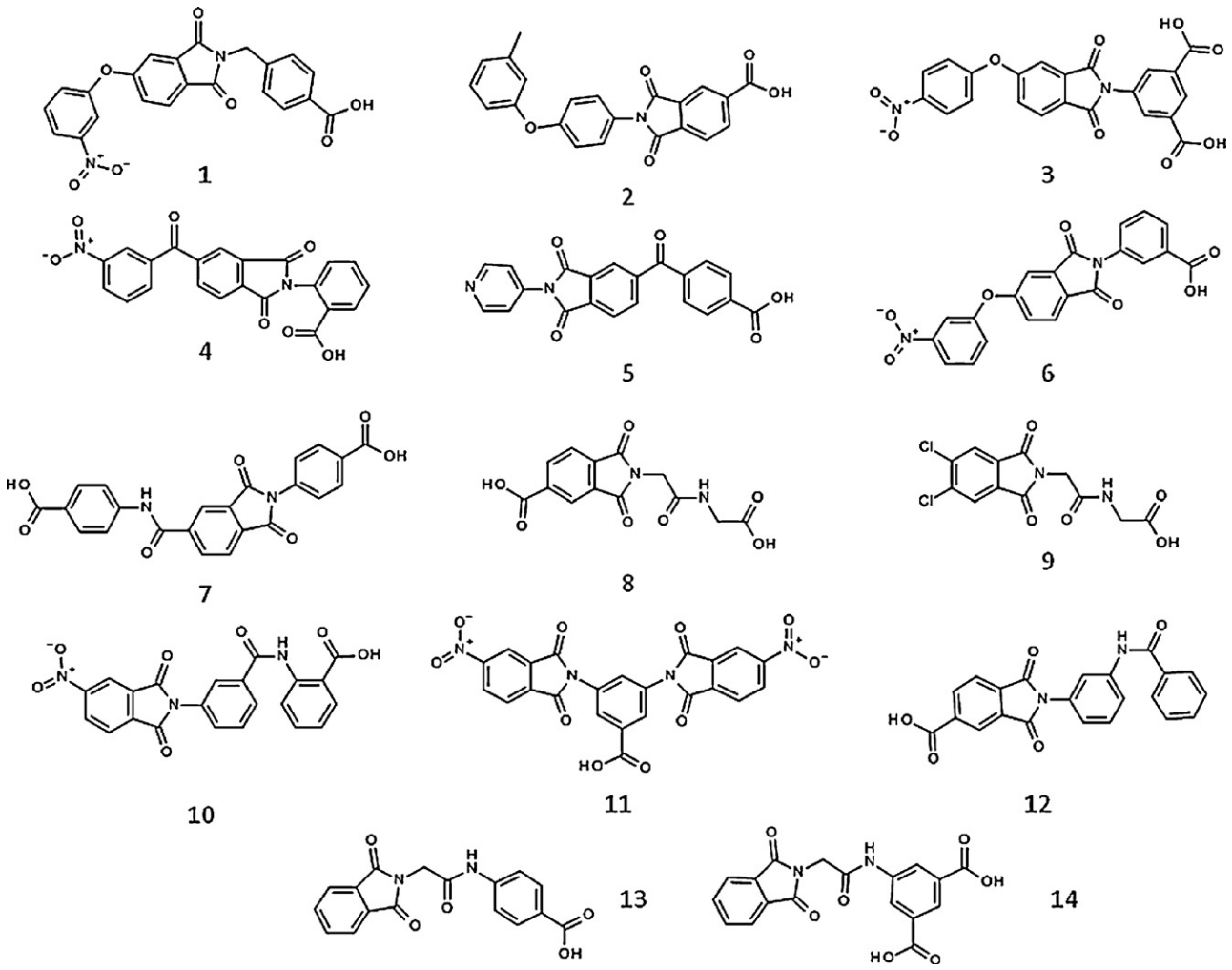

Fig. 3. Chemical structures of tested compounds based on similarity searching and QSAR modeling.

workflow in which similarity search matches would be subjected to flexible docking over the course of days [21]. Thus the primary advantage to this change in workflow is the time savings. Compounds were tested using a screening assay as previously described [21]. Thirteen of fourteen compounds selected were identified as antagonists (Table 4). The most efficacious compounds (**8**, **11** and **14**) identified in this single-dose assay showed about 40% inhibition. The low percent inhibitions are likely due to the relaxed criteria for active compounds set in the model. The addition of a binary QSAR model, therefore provides a filtering step to our virtual screening workflow. Compounds passing the binary QSAR filtering step will then be further analyzed using our structure-based model. This analysis is expected to identify more potent compounds by selecting compounds with favorable interactions. The diversity of our training set affords us the ability to predict the activity for structurally distinct compounds. This will increase the likelihood of identifying novel LPA_3 antagonists.

Table 4
Experimental LPA_3 antagonism assay results for external QSAR-selected test set. Compounds were tested using 10 μM against 200 nM LPA. Results are reported as %inhibition of the LPA response in the absence of test compound.

Tested hits	Hit2lead ID #	%Inhibition
1	5756161	14.32
2	6051643	25.35
3	5224125	25.76
4	5233604	21.32
5	5533659	17.22
6	5160920	36.52
7	5160692	23.00
8	6433264	43.44
9	7097113	36.52
10	5233467	23.40
11	5186588	41.83
12	6634278	23.47
13	5466318	19.03
14	5480696	40.62

4. Conclusion

We successfully generated and validated a binary QSAR model with reliable predictive power for LPA_3 antagonists. The accuracy of the model to predict activity on our external test set further highlights the ability of our model, and indicates that at least 55% of compounds within the applicability domain of the model selected for experimental screening on the basis of classification as active will prove to antagonize LPA action at the LPA_3 receptor. This model will be complementary to research focused on rational design of selective LPA antagonists. The model can be used as a rapid filter to select compounds that will be later subjected to more time-intensive docking studies during our drug design process.

Acknowledgments

This work was supported by grants from the National Institutes of Health (RO1 HL84007 to AP and RO1 CA92160 to GT) and

the Greater Southeast Affiliate of the American Heart Association (0715125B to JF). The MOE program was donated by the Chemical Computing Group and is greatly appreciated.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2010.03.002.

References

[1] R. Rivera, J. Chun, Biological effects of lysophospholipids, Rev. Physiol. Biochem. Pharmacol. 160 (2008) 25–46.
[2] A.L. Parrill, D.L. Baker, Autotaxin inhibition: challenges and progress toward novel anti-cancer agents, Anticancer Agents Med. Chem. 8 (2008) 917–923.
[3] S.M. Pasternack, I. von Kügelgen, K.A. Aboud, Y.A. Lee, F. Rüschendorf, K. Voss, A.M. Hillmer, G.J. Molderings, T. Franz, A. Ramirez, P. Nürnberg, M.M. Nöthen, R.C. Betz, G protein-coupled receptor P2Y5 and its ligand LPA are involved in maintenance of human hair growth, Nat. Genet. 40 (2008) 329–334.
[4] M. Murakami, A. Shiraishi, K. Tabata, N. Fujita, Identification of the orphan GPCR, P2Y(10) receptor as the sphingosine-1-phosphate and lysophosphatidic acid receptor, Biochem. Biophys. Res. Commun. 371 (2008) 707–712.
[5] K. Kotarsky, A. Boketoft, J. Bristulf, N. Nilsson, A. Norberg, S. Hansson, R. Sillard, C. Owman, F. Leeb-Lundberg, B. Olde, Lysophosphatidic acids binds to and activates GPR92. A G protein-coupled receptor highly expressed in gastrointestinal lymphocytes, J. Pharmacol. Exp. Ther. (2006) 619–628.
[6] J.W. Choi, C.W. Lee, J. Chun, Biological roles of lysophospholipid receptors revealed by genetic null mice: an update, Biochim. Biophys. Acta 1781 (2008) 531–539.
[7] M.M. Murph, G.H. Nguyen, H. Radhakrishna, G.B. Mills, Sharpening the edges of understanding the structure/function of the LPA1 receptor: expression in cancer and mechanisms of regulation, Biochim. Biophys. Acta 1781 (2008) 547–557.
[8] J.I. Fells, R. Tsukahara, Y. Fujiwara, J. Liu, D.H. Perygin, D.A. Osborne, G. Tigyi, A.L. Parrill, Identification of non-lipid LPA3 antagonists by virtual screening, Bioorg. Med. Chem. 16 (2008) 6207–6217.
[9] I. Ljjaali, F. Petitet, E. Dubus, O. Barberan, A. Michel, Assessing potency of c-Jun N-terminal kinase 3 (JNK3) inhibitors using 2D molecular descriptors and binary QSAR methodology, Bioorg. Med. Chem. 15 (2007) 4256–4264.
[10] K.-M. Thai, G.F. Ecker, A binary QSAR model for classification of hERG potassium channel blockers, Bioorg. Med. Chem. 16 (2008) 4107–4119.
[11] A.L. Parrill, U. Echols, T. Nguyen, T.-C.T. Pham, A. Hoeglund, D.L. Baker, Virtual screening approaches for the identification of non-lipid autotaxin inhibitors, Bioorg. Med. Chem. 16 (2008) 1784–1795.
[12] G.G. Durgam, R. Tsukahara, N. Makarova, M.D. Walker, Y. Fujiwara, K.R. Pigg, D.L. Baker, V.M. Sardar, A.L. Parrill, G. Tigyi, D.D. Miller, Synthesis and pharmacological evaluation of second-generation phosphatidic acid derivatives as lysophosphatidic acid receptor ligands, Bioorg. Med. Chem. Lett. 16 (2006) 633–640.
[13] G.G. Durgam, T. Virag, M.D. Walker, R. Tsukahara, S. Yasuda, K. Liliom, L.A. van Meeteren, W.H. Moolenaar, N. Wilke, W. Siess, G. Tigyi, D.D. Miller, Synthesis, structure–activity relationships, and biological evaluation of fatty alcohol phosphates as lysophosphatidic acid receptor ligands, activators of PPARgamma, and inhibitors of autotaxin, J. Med. Chem. 48 (2005) 4919–4930.
[14] D.J. Fischer, N. Nusser, T. Virag, K. Yokoyama, D. Wang, D.L. Baker, D. Bautista, A.L. Parrill, G. Tigyi, Short-chain phosphatidates are subtype-selective antagonists of lysophosphatidic acid receptors, Mol. Pharmacol. 60 (2001) 776–784.
[15] B.H. Heasley, R. Jarosz, K.M. Carter, S.J. Van, K.R. Lynch, T.L. Macdonald, A novel series of 2-pyridyl-containing compounds as lysophosphatidic acid receptor antagonists: development of a nonhydrolyzable LPA3 receptor-selective antagonist, Bioorg. Med. Chem. Lett. 14 (2004) 4069–4074.
[16] B.H. Heasley, R. Jarosz, K.R. Lynch, T.L. Macdonald, Initial structure–activity relationships of lysophosphatidic acid receptor antagonists: discovery of a high-affinity LPA1/LPA3 receptor antagonist, Bioorg. Med. Chem. Lett. 14 (2004) 2735–2740.
[17] K.R. Lynch, T.L. Macdonald, Structure–activity relationships of lysophosphatidic acid analogs, Biochim. Biophys. Acta 1582 (2002) 289–294.
[18] H. Ohta, K. Sato, N. Murata, A. Damirin, E. Malchinkhuu, J. Kon, T. Kimura, M. Tobo, Y. Yamazaki, T. Watanabe, M. Yagi, M. Sato, R. Suzuki, H. Murooka, T. Sakai, T. Nishitoba, D.S. Im, H. Nochi, K. Tamoto, H. Tomura, F. Okajima, Ki16425, a subtype-selective antagonist for EDG-family lysophosphatidic acid receptors, Mol. Pharmacol. 64 (2003) 994–1005.
[19] T. Virag, D.B. Elrod, K. Liliom, V.M. Sardar, A.L. Parrill, K. Yokoyama, G. Durgam, W. Deng, D.D. Miller, G. Tigyi, Fatty alcohol phosphates are subtype-selective agonists and antagonists of lysophosphatidic acid receptors, Mol. Pharmacol. 63 (2003) 1032–1042.
[20] MOE. MOE, version 2008.10. Chemical Computing Group Inc.: Montreal, 2008.
[21] J.I. Fells, R. Tsukahara, J. Liu, G. Tigyi, A.L. Parrill, Structure-based drug design identifies novel LPA3 antagonists, Bioorg. Med. Chem. 17 (2009) 7457–7464.
[22] A.L. Parrill, Lysophospholipid interactions with protein targets, Biochim. Biophys. Acta 1781 (2008) 540–546.
[23] K. Kano, N. Arima, M. Ohgami, J. Aoki, LPA and its analogs-attractive tools for elucidation of LPA biology and drug development, Curr. Med. Chem. 15 (2008) 2122–2131.
[24] C.E. Heise, W.L. Santos, A.M. Schreihofer, B.H. Heasley, Y.V. Mukhin, T.L. Macdonald, K.R. Lynch, Activity of 2-substituted lysophosphatidic acid (LPA) analogs at LPA receptors: discovery of a LPA1/LPA3 receptor antagonist, Mol. Pharmacol. 60 (2001) 1173–1180.
[25] A. Tropsha, A. Golbraikh, Predictive QSAR modeling workflow, model applicability domains, and virtual screening, Curr. Pharm. Des. 13 (2007) 3494–3504.
[26] A. Golbraikh, A. Tropsha, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, Mol. Divers. 5 (2002) 231–243.
[27] D.M. Hawkins, The problem of overfitting, J. Chem. Inf. Comput. Sci. 44 (2004) 1–12.
[28] D.G. Sprous, Fingerprint-based clustering applied to define a QSAR model use radius, J. Mol. Graph. Modell. 27 (2008) 225–232.
[29] H. Zhang, X. Xu, J. Gajewiak, R. Tsukahara, Y. Fujiwara, J. Liu, J.I. Fells, D. Perygin, A.L. Parrill, G. Tigyi, G.D. Prestwich, Dual activity lysophosphatidic acid receptor pan-antagonist/autotaxin inhibitor reduces breast cancer cell migration in vitro and causes tumor regression in vivo, Cancer Res. 69 (2009) 5441–5449.