



Published in final edited form as:

Eur J Med Chem. 2010 November ; 45(11): 5406–5419. doi:10.1016/j.ejmech.2010.09.001.

Computational Structure-activity Relationship Analysis of Small-Molecule Agonists for Human Formyl Peptide Receptors

Andrei I. Khlebnikov^{a,*}, Igor A Schepetkin^b, and Mark T. Quinn^{b,*}

^a Department of Chemistry, Altai State Technical University, Barnaul, Russia

^b Department of Veterinary Molecular Biology, Montana State University, Bozeman, Montana

Abstract

N-formyl peptide receptors (FPR) are important in host defense. Because of the potential for FPRs as therapeutic targets, recent efforts have focused on identification of non-peptide agonists for two FPR subtypes, FPR1 and FPR2. Given that a number of specific small molecule agonists have recently been identified, we hypothesized that computational structure-activity relationship (SAR) analysis of these molecules could provide new information regarding molecular features required for activity. We used a training set of 71 compounds, including 10 FPR1-specific agonists, 36 FPR2-specific agonists, and 25 non-active analogs. A sequence of (1) one-way analysis of variance selection, (2) cluster analysis, (3) linear discriminant analysis, and (4) classification tree analysis led to the derivation of SAR rules with high (95.8%) accuracy for correct classification of compounds. These SAR rules revealed key features distinguishing FPR1 versus FPR2 agonists. To verify predictive ability, we evaluated a test set of 17 additional FPR agonists, and found that the majority of these agonists (>94%) were classified correctly as agonists. This study represents the first successful application of classification tree methodology based on atom pairs to SAR analysis of FPR agonists. Importantly, these SAR rules represent a relatively simple classification approach for virtual screening of FPR1/FPR2 agonists.

Keywords

Formyl peptide receptor (FPR); FPR agonists; Atom pairs; Molecular descriptors; Structure-activity relationship analysis

1. Introduction

N-formyl peptides activate phagocytes through G protein-coupled receptors known as formyl peptide receptors (FPR) [1]. FPR1 was the first FPR cloned and encodes a high-affinity receptor for fMLF [2]. Subsequently, it was found that two additional FPRs exist in humans, and these were originally designated as FPR-like 1 (FPRL1; 69% identity to FPR1) and FPR-like 2 (FPRL2; 56% identity to FPR1) [3–5]. Recently, the FPR nomenclature has been revised such that FPRL1 and FPRL2 are now designated as FPR2 and FPR3, respectively [6].

*Corresponding authors: Dr. Andrei I. Khlebnikov, Department of Chemistry, Altai State Technical University, Barnaul 656038, Russia, Phone: +7 3852 245513/522436; Fax: +7 3852 367864, aikhl@chem.org.ru. Dr. Mark T. Quinn, Veterinary Molecular Biology, Montana State University, Bozeman, MT 59717, Phone: 406-994-4707; Fax 406-994-4303, mquinn@montana.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Compared to FPR1, FPR2 exhibits a high level of ligand promiscuity and is activated by numerous chemically unrelated ligands, including synthetic peptides, pathogen-derived peptides, host-derived peptides, and lipids [reviewed in [6]]. In addition to natural peptides and endogenous arachidonic acid metabolites, novel synthetic peptides and several small-molecule non-peptide agonists of FPR1 and FPR2 have recently been reported [7–11]. Indeed, the identification and development of small-molecule ligands represents an ideal approach to analyze FPR structure and function, since such molecules are well defined and can be easily modified for structure-activity relationship (SAR) analysis. Small-molecule agonists can also have advantages over peptides or proteins as potential therapeutics, and they can provide a basis for construction of useful pharmacophore models of FPR1/FPR2 agonists.

Our analysis of novel small-molecule agonists of FPRs showed that individual ring substituents had a significant impact on FPR1/FPR2 agonist activity [12], suggesting that further structure-activity analysis of known FPR agonists could lead to optimization of these lead compounds and identification of improved agonists. In fact, SAR and quantitative SAR (QSAR) models have been instrumental in understanding the molecular mechanism of action of receptor agonists and antagonists, directing their design, and facilitating virtual screening [13–15].

To date, non-computational SAR analysis has been performed for FPR2 agonists with a benzimidazole scaffold [16], FPR1/FPR2 agonists with a pyridazin-3(2H)-one scaffold [17], and pyrazolone-derived FPR2 agonists [9]; however, there are currently no reported computational SAR models for non-peptide FPR1/FPR2 agonists.

Here, we used computational SAR analysis of a large group of FPR1/FPR2 agonists and their non-active analogs. The SAR rules obtained from classification tree analysis, which was based on six atom pair descriptors only, revealed key features that distinguished FPR1 and FPR2 agonists. These studies provide further virtual screening of FPR1/FPR2 agonists and also provide clues to the molecular features required for agonist activity. This is the first application of an atom pair-based approach to a set of FPR agonists with various scaffolds and their analogs.

2. Results and discussion

2.1. Atom pairs and their one-way ANOVA selection

While a variety of molecular parameters can be used in the computational methods for (Q)SAR analysis [18,19], some of these parameters are complex physicochemical or geometrical 3D descriptors whose calculation is associated with difficulties conditioned by molecular flexibility and adequate sampling of conformational space. Conversely, topological indices, or 2D descriptors, obtainable from the structural formula of a compound are very attractive because of their simplicity. A reasonable compromise between ease of interpretation and ease of computation was reported by Carhart et al. [20], who introduced atom pair descriptors as features of the environments of all atoms in the 2D representation of a chemical structure. This approach has been widely used in the context of fragment-based similarity searches and database mining [21–25]. Here, we applied atom pair descriptors to represent the selected molecules. The use of an atom type naming scheme from MM+ force field, as implemented in HyperChem software, is a significant feature of our approach, as it represents more sophisticated atom typing than used in previous studies [20]. This scheme assigns specific names to a given atom, depending on its surroundings in a molecule.

The 2D structures of 71 compounds (training set) that included 10 FPR1-specific agonists, 36 FPR2-specific agonists, and 25 non-active compounds (Table 1, see *Materials and Methods*) represented as a set of HIN files were used by our CHAIN program to generate a table of atom pairs. This program, which finds all possible paths between labeled atoms in a molecular structure, identified 726 unique atom pairs among the 71 compounds. Hence, a matrix

consisting of 71 lines and 726 columns was generated, with each line containing the number of times a given atom pair was present in each molecule. Since this number of columns is too large for SAR analysis, we performed a step-by-step selection of atom pairs to reduce the matrix size. This sequence of steps included one-way ANOVA, linear discriminant analysis, and binary classification tree analysis. These methodologies have a high applicability for variable pre-screening in SAR [26–28] and were also used successfully in our previous studies [22, 23].

For the first selection step, we applied one-way analysis of variance (ANOVA) [29] to select descriptors with significant differences between total and within-class variances. As a result of ANOVA selection, 565 descriptors were filtered out, while the remaining 161 significant atom pairs were retained for further analysis. It is reasonable to compare the distribution of initial and ANOVA-selected descriptors in terms of bond separation (i.e., the number of bonds between atoms in a given atom pair). It should be noted that the relative distribution of initial and ANOVA-selected descriptors, in terms of bond separation, was similar (Figure 1). Although a greater proportion of the original atom pairs was retained after ANOVA selection for pairs with the largest bond separation (i.e., 17–18 bonds), these atom pairs are rare in the data set and have negligible statistical impact on the total SAR analysis. Since the relative distribution of atom pairs by bond separation was not changed after ANOVA, it appears that molecular shape peculiarities do not have a major influence on biological activity of these FPR agonists. In contrast, the dumb-bell shape common to inducers of macrophage tumor necrosis factor- α production led to a significantly higher fraction of “longer” atom pairs among ANOVA-selected descriptors [23].

2.2. Cluster and linear discriminant analyses

The second step of variable selection consisted of finding clusters of highly correlated descriptors. Subsequently a single variable from one cluster can be regarded as independent, whereas the other dependent descriptors of such a cluster can be excluded from further calculations. Using the 161 \times 161 matrix of correlation coefficients for atom pairs selected by ANOVA, we chose 28 clusters of variables (Table 2). Each variable is highly correlated ($r \geq 0.9$) with at least one variable from the same cluster. Descriptors with longer bond separations were taken as representative variables (shown in bold italic in Table 2). Hence, instead of 102 atom pairs included in Table 2, only 28 atom pairs were retained. These variables were combined with the 59 remaining atom pairs that were not highly correlated, resulting in a set of 87 atom pair descriptors selected for further calculations.

The high correlation coefficient between values of descriptors implies that these atom pairs are simultaneously present in most compounds in the data set. This usually occurs when atom pairs are produced by certain molecular features or scaffolds. The features may be very simple. For example, Cluster 14 descriptor *BR_6_CO* corresponds to the presence of a bromine atom and carbonyl carbon separated by 6 bonds. The same substructure also contains another atom pair from this cluster (*BR_7_O1*) corresponding to a carbonyl oxygen and bromine separated by 7 chemical bonds. More populated sets of correlated descriptors are produced by multi-atomic scaffolds. For instance, 11 descriptors of Cluster 2 originate from the 2-(benzimidazol-2-ylsulfanyl)-N-phenylacetamide scaffold common to compound **AG-09/25** and its analogs. Thus, clustering atom pairs according to their mutual correlation not only decreases the number of variables but also provides a rational way to interpret SAR results in terms of chemical features and building blocks, which are much more complex than the atom pairs themselves. It should be noted that lowering the correlation coefficient threshold from 0.9 to 0.8 gave rather heterogeneous clusters of correlated descriptors usually not associated with distinct chemical substructures. Although the composition of several clusters remained the same, some were condensed to larger clusters at $r \geq 0.8$ by inclusion of additional atom pair

descriptors (Figure 2). On the other hand, the adopted threshold of 0.9 for correlation between a given variable and *at least one* variable from the same cluster provides high mutual correlation of *all* variables in this cluster. For example, each pair of descriptors among the 13 variables of Cluster 1 (Table 2) is characterized by an *r* value greater than 0.85.

The 87 atom pairs selected after the two steps described above were used as an input variable set for linear discriminant analysis (LDA). The LDA procedure was applied with the option of “forward stepwise” inclusion of variables, as implemented in STATISTICA 6.0 software. The descriptors were added to the model if their inclusion led to a significant improvement in classification ($p < 0.05$). We found that 17 of the 87 atom pairs were sufficient for good LDA classification of agonists, with 68 of the 71 compounds (95.8%) classified correctly as FPR1, FPR2, or NA (Figure 3A).

The LDA model with 17 atom pairs derived on the third step of variable selection was further simplified after an additional run of LDA with the “best subset search” option. The number of atom pair descriptors was decreased from 17 to 9 without loss of quality of the model (accuracy was the same using either 17 or 9 descriptors). This relatively simple LDA model obtained on the fourth step of variable selection can be expressed by the following three classification functions:

$$\begin{aligned}
 F(\text{NA}) = & - 6.237 + 4.988 \text{ C4_1_N2} \\
 & + 10.364 \text{ C3_3_CO} \\
 & - 1.855 \text{ N2_3_O1} \\
 & + 1.246 \text{ BR_6_C4} \\
 & - 0.039 \text{ C4_6_C4} \\
 & + 0.816 \text{ C4_6_O1} \\
 & - 3.376 \text{ BR_7_O1} \\
 & - 1.033 \text{ C3_9_CA} - 0.27561 \text{ CA_12_O2}
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 F(\text{FPR1}) = & - 12.718 + 3.474 \text{ C4_1_N2} \\
 & + 9.286 \text{ C3_3_CO} \\
 & - 1.355 \text{ N2_3_O1} \\
 & - 7.595 \text{ BR_6_C4} \\
 & - 1.093 \text{ C4_6_C4} \\
 & + 0.289 \text{ C4_6_O1} \\
 & + 1.866 \text{ BR_7_O1} \\
 & - 0.365 \text{ C3_9_CA} + 4.022 \text{ CA_12_O2}
 \end{aligned} \tag{2}$$

$$\begin{aligned}
 F(\text{FPR2}) = & -6.875 + 2.774 \text{C4_1_N2} \\
 & + 3.353 \text{C3_3_CO} \\
 & + 2.942 \text{N2_3_O1} \\
 & - 2.821 \text{BR_6_C4} \\
 & + 1.617 \text{C4_6_C4} \\
 & - 3.260 \text{C4_6_O1} \\
 & + 6.457 \text{BR_7_O1} \\
 & + 0.228 \text{C3_9_CA} + 1.525 \text{CA_12_O2}
 \end{aligned}
 \tag{3}$$

The number of corresponding atom pairs in a given compound should be used as values of descriptors for calculation of functions (1–3). One of the classes (FPR1, FPR2, or NA) is then attributed to the compound according to the maximum value among these functions.

Table 3 contains calculated classes and results of leave-one-out (LOO) prediction for the entire series of compounds. All three compounds with incorrectly calculated classes (**AG-09/41**, **AG-09/95**, and **AG-09/102**) were inactive, while the LDA model (1–3) and LOO prediction classified them as having FPR2 activity. LOO cross-validation correctly classified 63 of 71 compounds (88.7% accuracy). This can be considered as good quality of prediction, taking into account that the model with 9 descriptors was derived based on 71 molecules in the training set. For the subset of FPR1 agonists, the fraction of correct LOO predictions was expectedly lower (70%) because of the relatively small number of compounds with FPR1 activity included in the series under investigation. This was conditioned by the low number of non-peptide FPR1-specific agonists reported in the literature.

2.3. Classification tree analysis

Although the LDA model is good in terms of fitting and prediction, its use in practice is based on calculation of functions (1–3) and is difficult to interpret. It would be much better to find simple SAR rules which are intuitively understandable and expressed in natural “chemical” language. In recent studies, we exploited a binary classification tree approach to build logical SAR algorithms based on atom pairs for low-molecular weight inhibitors of human neutrophil elastase [22] and non-peptide inducers of TNF- α production [23]. In the present paper, we report the first application of the classification tree methodology based on atom pairs to SAR analysis of compounds with varying scaffolds. Indeed, a multi-scaffold training set would produce a SAR model that is more useful for subsequent virtual screening of potential FPR agonists.

The 9 descriptors involved in the best LDA model (Functions 1–3) were used as the starting variable set for the classification tree algorithm [30] implemented in STATISTICA 6.0. Deriving an optimal tree with cross-validation criteria represented the final stage of our step-by-step variable selection and resulted in a six-branched tree (Figure 4). Thus, 720 of the initial 726 descriptors were filtered out during the selection steps, and the remaining six atom pairs formed a basis for the formulation of simple, logical SAR. Despite the small number of retained variables, the model is characterized by good fitting (i.e., it correctly classified most of the compounds from the training set). Figure 3B shows that 67 of the 71 compounds (94.4%) were accurately recognized by the tree with respect to their activity classes. Three non-active compounds were misclassified as FPR2, while one FPR2 agonist was misclassified as non-active. Detailed information regarding the retained descriptors and terminal nodes of the tree responsible for classification of each compound are shown in Table 3. The low cross-validation

cost (0.141) obtained for the model demonstrates the powerful predictive ability of the classification tree (see also the results below for applying this model to an external data set).

Chemical substructures associated with branches in the classification tree are illustrated in Figure 5. The first branch evaluates molecules for the presence of two or more CA_12_O2 atom pairs. These atom pairs occur when bicoordinated ether oxygen atoms and a benzene ring are separated by 10 or 11 chemical bonds (see examples in Figure 5A and 5B). Molecules containing two or more of these atom pairs move to the right branch of the tree (24 of the 71 compounds evaluated), whereas molecules with less than two CA_12_O2 atom pairs moved to the left branch (47 of the 71 compounds) (see Figure 4).

The first branch on the left evaluates molecules for the presence of BR_7_O1 atom pairs. BR_7_O1 and BR_6_CO are mutually correlated because they are found in compounds with a bromine atom and a carbonyl oxygen separated by 7 bonds (i.e., the topological distance of 6 chemical bonds falls between bromine and the carbonyl carbon atom) (see examples in Figure 5C and 5D). At this branch, molecules with one or more of these atom pairs are designated as FPR2 agonists. Otherwise, they are sent to the next branch associated with the C3_9_CA descriptor. According to the split condition, a compound is designated as an FPR2 agonist in terminal node 9 if it contains five or more C3_9_CA atom pairs. This occurs when several C3-type carbons (MM+ force field notation for sp²-hybridized carbon of non-benzene character) are present with simultaneous presence of an aromatic ring on the opposite side of a molecule (see examples in Figure 5E and 5F). In some molecules, the number of C3_9_CA atom pairs does not exceed 4, despite the occurrence of atoms with C3 and CA types (e.g., see **C-14x** in Figure 5C).

The final branch on the left evaluates remaining molecules for the presence of the N2_3_O1 atom pair, as well as the correlated CO_2_N2 atom pair, which falls within the N2_3_O1 structure (see Figure 5G and 5H). This branch is important for correct classification of FPR2 agonists **AG-09/5** and **AG-09/8** from the training set, since these compounds contain two N2_3_O1 atom pairs and pass to terminal node 13, while the other compounds with less than two N2_3_O1 atom pairs are classified as non-active in terminal node 12 (see Figure 4). Although it might be suggested that this split is not important and could be removed from the tree without significant loss of classification quality, its removal caused a noticeable increase in cross-validation cost from 0.141 to 0.254 for the truncated tree. Thus, such a simplification of the model is not statistically warranted. Moreover, chemical features associated with the N2_3_O1 atom pair were also important for correct classification of several compounds from the test set (see below).

Molecules with two or more CA_12_O2 atom pairs moved to the right branch of the classification tree. The first node of the right branch evaluates molecules for the presence of the C4_6_C4 descriptor, which corresponds to tetrahedral sp³-carbons separated by 6 bonds. Atom pairs of this type can be found in compounds with various scaffolds and is originated by two saturated hydrocarbon moieties located moderate distances from each other (examples are shown in Figure 5I and 5J). The final node in this branch evaluates molecules for the presence of the BR_6_C4 atom pair. This substructure is present in five non-active compounds, three of which contain an *m*-bromophenyl-acetamide fragment (**C-14b**, **AG-09/25**, and **AG-09/72**; see Figure 5K and 5L) and two others containing an N-alkyl-*p*-bromoaniline substructure (**C-14r** and **C-23**). These features can be used in the formulation of “chemical” rules for SAR analysis. For example, movement of bromine from the *meta* to the *para* position of the aromatic ring in a bromo-substituted phenyl-acetamide moiety transformed the non-active **C-14b** into the FPR1 agonist **C-17b**.

Atom pairs from the clusters of correlated variables (Table 2, Figure 2) did not dominate at the nodes of the classification tree, and only N2_3_O1 and BR_7_O1 were involved in the split rules. Additionally, large clusters produced by entire scaffolds did not participate at all in the classification tree. Thus, the classification process does not appear to be biased by large chemical substructures and, therefore, would be useful for evaluation of molecules with various types of chemical scaffolds.

The best approach to validate SAR and QSAR models is to apply them to an independent series of compounds. For this purpose, we evaluated a test set consisting of 17 FPR2-specific or mixed FPR1/FPR2 agonists (Table 4). A matrix of atom pairs was generated using CHAIN program, and six columns of the matrix which correspond to the descriptors important for SAR analysis were taken into account. Values of the 6 descriptors important for SAR analysis used in the classification tree are shown in Table 4 along with the classification results obtained using the binary tree and algorithm from Scheme 1. FPR2-specific agonists **B-25**, **B-35**, and **B-42** were correctly predicted as having FPR2 activity, while most of the mixed-type compounds were classified as either FPR1 (**AG-09/9**, **AG-09/17**, **AG-09/20**, **AG-09/22**, **C-14a**, **C-14e**, **C-14h**, and **C-14n**) or FPR2 (**AG-22**, **B-25**, **B-35**, **B-42**, fMLF, and WKYMVm) agonists. Two members of test set (**AG-09/10** and **1910-5441**) were misclassified as non-active. Note, however, that FPR1 agonist **1910-5441** has relatively lower activity ($EC_{50} \sim 20 \mu\text{M}$) [8] than the other agonists used in our computational SAR analyses. Although oligopeptides were not included in the training set, the peptides fMLF and WKYMVm from the test set were classified correctly as active compounds. Note that these two peptides possess common fragments, e.g. benzyl and 2-methylthioethyl groups. The recognition of molecules by FPRs can also be strongly determined by configuration of chiral centers; however, our atom pair approach does not currently account for molecular chirality and would require introduction of these variables as additional descriptors.

Our simplified SAR model is based on agonists of non-mixed (i.e., “pure” FPR1 or FPR2 agonists), while the test set contained mostly mixed-type compounds. Such a situation was conditioned by the absence of a substantial number of small molecule receptor-specific agonists with relatively high activity reported in the literature. Thus, the aim of this test set was primarily in evaluation of the model for its ability to distinguish active and inactive compounds. Obviously, further discovery of novel specific FPR1 and FPR2 agonists will allow us to expand both training and test sets in order to derive a model with enhanced predictive ability based on atom pair descriptors. On the other hand, we can predict scaffold “specific affinity” of mixed agonists for either FPR1 or FPR2 using this model. For example, mixed-type agonists **C-14a**, **C-14e**, **C-14h**, and **C-14n** [17] and **AG-09/17**, **AG-09/20**, and **AG-09/22** [12] were classified by the model as FPR1 agonists, indicating that 4-benzylpyridazin-3-one and 2-(benzimidazol-2-ylthio)-N-phenylacetamide scaffolds have higher affinity for FPR1. By comparison, the pyrazolone scaffold (agonists **B-25**, **B-35**, **B-42**, and **B-43**) had higher affinity for FPR2. Note, however, that this feature requires the presence of specific agonists with such scaffolds in the training set.

3. Conclusion

Previously, high-throughput screening was used to select unique non-peptide agonists of FPR1 and FPR2 [7–12,16,17]. In the present study, we utilized atom pair descriptors for computational SAR analysis of most active FPR1/FPR2 agonists to further define the features of these molecules critical for agonist activity and to develop a simple, but accurate SAR model for predicting biological activity in future compound screening.

A sequence of ANOVA, cluster analysis, LDA, and classification tree analysis based on the atom pair descriptors led to the derivation of simple SAR rules, despite the large number of

starting variables. The SAR rules obtained from classification tree analysis, which was based on six atom pair descriptors only, revealed that the FPR1 agonists in the series investigated could be characterized by simultaneously satisfying the following conditions: a) presence of a benzene ring and two-coordinated oxygen atoms separated by 10 or 11 bonds (Condition A); b) absence of sp^3 -carbon atoms separated by 6 bonds; and c) absence of *m*-bromophenyl-acetamide and *N*-alkyl-*p*-bromoaniline substructures. A compound can be classified as an FPR2 agonist if it does not match Condition A and satisfies one of the following statements: a) contains a bromine atom and a carbonyl oxygen separated by 7 bonds; b) at least three non-benzene sp^2 -carbons separated by 6 to 9 bonds from benzene ring(s), and at least two of these carbons separated by 7 or 8 bonds from benzene ring(s); or c) at least two *N*-acylhydrazine or α -aminoketone substructures are present. Another type of FPR2 agonist satisfies Condition A and contains sp^3 -carbon atoms separated by 6 bonds. To evaluate predictive ability of the method, we evaluated a test set of 17 FPR agonists. Most, including the two peptides fMLF and WKYVMv, were classified by the derived rules as active agonists. Thus, we provide here the first successful application of the classification tree methodology based on atom pairs for SAR analysis of FPR agonists with various scaffolds.

Good quality and high predictive ability of the SAR model, as well as simplicity and rapidity of calculations associated with the binary tree algorithm, suggest promise in using the classification tree model for large database mining and virtual screening of FPR agonists.

4. Experimental

4.1. Data sets

Training Set—The training set of 71 compounds contained 10 FPR1-specific agonists, 36 FPR2-specific agonists, and 25 non-active compounds, including 2-(benzimidazol-2-ylthio)-*N*-phenylacetamides, *N*-phenethyl-*N'*-phenylureas, piperazines, acetohydrazides, 1-(2-indolylcarbonyl)-4-(1-benzimidazolyl)piperidines, 4-benzylpyridazin-3-ones, arylcarboxylic acid hydrazides, 1-(2-indolylcarbonyl)-3-(1-benzimidazolyl)pyrrolidines and related scaffolds [7–11]. All selected agonists had EC_{50} values in the low micromolar range for their ability to induce intracellular calcium mobilization in cells (RBL-2H3 or HL-60) transfected with human FPR1 or FPR2. All active compounds were evaluated in wild-type cells to verify that the agonists were inactive in non-transfected cells. Names of compounds and their experimentally determined classes are shown in Table 3 (1st and 2nd columns). Structures of all training set compounds are shown in Table 1.

Test Set—The test set consisted of 17 compounds, including 11 compounds with mixed FPR1/FPR2 agonist activity [**AG-22**, **AG-09/9**, **AG-09/10**, **AG-09/17**, **AG-09/20**, **AG-09/22**, **C-14a**, **C-14e**, **C-14h**, **C-14n** [12], and **B-43** [31]] and 3 FPR2-specific agonists (**B-25**, **B-35**, and **B-42**) [9]. Also included in the test set was **1910-5441**, which has been shown to be an FPR1 agonist, although activity for FPR2 was not evaluated [8]. Finally, the test set included two peptide FPR agonists: fMLF and WKYVMv.

4.2. Structure encoding by atom pairs

For the derivation of SAR models, we used an atom pair representation of molecular structures. Each atom pair was denoted as T1_D_T2, where T1 and T2 are the types of atoms in the pair and D is bond separation, i.e. number of bonds in the shortest path between these atoms in the structural formula. As described previously [22,23], T1 and T2 were defined with symbolic codes used in HyperChem, Version 7 (Hypercube, Inc., Gainesville, FL) for atom type encoding within the MM+ force field. For example, CA, CO, and C4 codes were used for sp^2 -hybridized aromatic, carbonyl, and tetrahedral sp^3 -hybridized carbon atoms, respectively. This approach allows easy generation of atom pairs from the output file containing the

molecular structure (HIN file) built by HyperChem. As atom pairs T1_D_T2 and T2_D_T1 are equivalent, we used a unified definition with lexicographic order of type substrings (i.e., with $T1 \leq T2$).

All 726 unique atom pairs possible for non-hydrogen atoms in the 71 compounds of the training set were generated. This 71×726 data matrix was automatically built by our CHAIN program, based on HIN files created in HyperChem. A matrix element at the intersection of the i^{th} row and j^{th} column was equal to the j^{th} atom pair occurrence in the i^{th} molecule. A similar data matrix was calculated for the 17 compounds in the test set.

4.3. Data processing and derivation of SAR rules

The data matrix for the training set was used as an input for one-way ANOVA [29] implemented in STATISTICA 6.0. From 726 atom pairs, 161 descriptors were selected, which showed significant differences ($p \leq 0.05$) between total and within-class variances. These ANOVA-selected atom pairs were then clustered according to their mutual correlation. Each member of a cluster was tightly correlated ($r \geq 0.9$) with at least one descriptor from the same cluster. All but one variable in each of the 28 clusters were considered dependent and were discarded, while the remaining variable with the highest bond separation was taken as independent in a given cluster. These 28 atom pairs, together with the 59 non-clustered variables (total of 87 descriptors), formed the basis set for further analysis.

Derivation of SAR classification was then performed by the LDA method with the 'Forward Stepwise' option, using the corresponding module of STATISTICA 6.0. The statistical criterion for inclusion or exclusion of descriptors at each step was $p \leq 0.05$. The stepwise LDA allowed selection of 17 significant descriptors from 726 atom pairs generated initially. The LDA procedure was then repeated with the 'Best Subset Search' option on the basis of 17 variables selected in the first LDA run. The best subset consisted of just 9 atom pairs giving the least misclassification error in the LDA model. Starting from 9 variables of the best subset, we developed a binary classification tree model with univariate splits. The classification tree was built with STATISTICA 6.0 using estimated prior probabilities and equal misclassification costs for classes [30,32]. An exhaustive classification and regression tree-style univariate split selection method was used, as described by Breiman et al. [30].

Acknowledgments

This work was supported in part by National Institutes of Health grant P20 RR-020185, National Institutes of Health contract HHSN266200400009C, an equipment grant from the M.J. Murdock Charitable Trust, and the Montana State University Agricultural Experimental Station.

Reference List

1. Le Y, Murphy PM, Wang JM. Trends Immunol 2002;23:541–548. [PubMed: 12401407]
2. Boulay F, Tardiff M, Brouchon L, Vignais P. Biochemistry 1990;29:11123–11133. [PubMed: 2176894]
3. Ye RD, Cavanagh SL, Quehenberger O, Prossnitz ER, Cochrane CG. Biochem Biophys Res Commun 1992;184:582–589. [PubMed: 1374236]
4. Murphy PM, Ozcelik T, Kenney RT, Tiffany HL, McDermott D, Francke U. J Biol Chem 1992;267:7637–7643. [PubMed: 1373134]
5. Bao L, Gerard NP, Eddy RL Jr, Shows TB, Gerard C. Genomics 1992;13:437–440. [PubMed: 1612600]
6. Ye RD, Boulay F, Wang JM, Dahlgren C, Gerard C, Parmentier M, Serhan CN, Murphy PM. Pharmacol Rev 2009;61:119–161. [PubMed: 19498085]
7. Nanamori M, Cheng X, Mei J, Sang H, Xuan Y, Zhou C, Wang MW, Ye RD. Mol Pharm 2004;66:1213–1222.

8. Edwards BS, Bologna C, Young SM, Balakin KV, Prossnitz ER, Savchuck NP, Sklar LA, Oprea TI. *Mol Pharm* 2005;68:1301–1310.
9. Bürlü RW, Xu H, Zou X, Müller K, Golden J, Frohn M, Adlam M, Plant MH, Wong M, McElvaine M, Regal K, Viswanadhan VN, Tagari P, Hungate R. *Bioorg Med Chem Lett* 2006;16:3713–3718. [PubMed: 16697190]
10. Schepetkin IA, Kirpotina LN, Khlebnikov AI, Quinn MT. *Mol Pharm* 2007;71:1061–1074.
11. Schepetkin IA, Kirpotina LN, Tian J, Khlebnikov AI, Ye RD, Quinn MT. *Mol Pharm* 2008;74:392–402.
12. Kirpotina LN, Khlebnikov AI, Schepetkin IA, Ye RD, Rabiet MJ, Jutila MA, Quinn MT. *Mol Pharmacol* 2010;77:159–170. [PubMed: 19903830]
13. Tong W, Welsh WJ, Shi L, Fang H, Perkins R. *Environ Toxicol Chem* 2003;22:1680–1695. [PubMed: 12924570]
14. Andricopulo AD, Montanari CA. *Mini Rev Med Chem* 2005;5:585–593. [PubMed: 15974936]
15. Helguera AM, Combes RD, Gonzalez MP, Cordeiro MN. *Curr Top Med Chem* 2008;8:1628–1655. [PubMed: 19075771]
16. Frohn M, Xu H, Zou X, Chang C, McElvaine M, Plant MH, Wong M, Tagari P, Hungate R, Bürlü RW. *Bioorg Med Chem* 2007;17:6633–6637.
17. Cilibrizzi A, Quinn MT, Kirpotina LN, Schepetkin IA, Holderness J, Ye RD, Rabiet MJ, Biancalani C, Cesari N, Graziano A, Vergelli C, Pieretti S, Dal P, Giovannoni VMP. *J Med Chem* 2009;52:5054–5057.
18. Buttingsrud B, Ryeng E, King RD, Alsberg BK. *J Comput Aided Mol Des* 2006;20:361–373. [PubMed: 17054018]
19. Gute BD, Basak SC. *SAR QSAR Environ Res* 2006;17:37–51. [PubMed: 16513551]
20. Carhart RE, Smith DH, Venkataraghavan R. *J Chem Inf Comput Sci* 1985;25:64–73.
21. Plewczynski D, von GM, Spieser SA, Rychlewski L, Wyrwicz LS, Ginalski K, Koch U. *Comb Chem High Throughput Screen* 2007;10:189–196. [PubMed: 17346118]
22. Khlebnikov AI, Schepetkin IA, Quinn MT. *Bioorg Med Chem* 2008;16:2791–2802. [PubMed: 18234502]
23. Khlebnikov AI, Schepetkin IA, Kirpotina LN, Quinn MT. *Bioorg Med Chem* 2008;16:9302–9312. [PubMed: 18815052]
24. Perez-Nueno VI, Rabal O, Borrell JI, Teixido J. *J Chem Inf Model* 2009;49:1245–1260. [PubMed: 19364101]
25. Yu N, Bakken GA. *J Chem Inf Model* 2009;49:745–755. [PubMed: 19309177]
26. Yan SF, King FJ, He Y, Caldwell JS, Zhou Y. *J Chem Inf Model* 2006;46:2381–2395. [PubMed: 17125181]
27. Li J, Liu H, Yao X, Liu M, Hu Z, Fan B. *Anal Chim Acta* 2007;581:333–342. [PubMed: 17386461]
28. Roncaglioni A, Piclin N, Pintore M, Benfenati E. *SAR QSAR Environ Res* 2008;19:697–733. [PubMed: 19061085]
29. Lindman, HR. *Analysis of variance in complex experimental designs*. W.H. Freeman & Co; San Francisco: 1974.
30. Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. *Classification and regression trees*. Wadsworth & Brooks/Cole Advanced Books & Software; Monterey, CA: 1984.
31. Sogawa Y, Shimizugawa A, Ohyama T, Maeda H, Hirahara K. *J Pharmacol Sci* 2009;111:317–321. [PubMed: 19926937]
32. Loh WY, Shih YS. *Statistica Sinica* 1997;7:815–840.

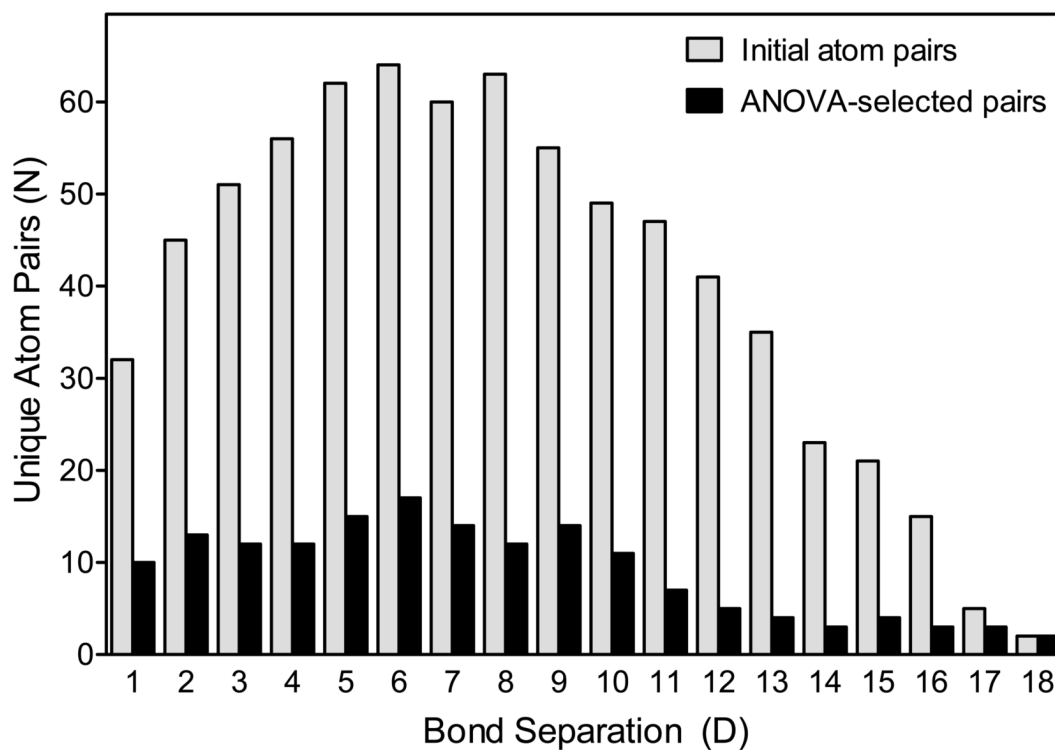


Figure 1.

Comparison of initial and ANOVA-selected atom pairs in compounds **1-71**. The numbers are shown for each of the indicated bond separations initially generated for 71 compounds from Table 1 (light bars). Atom pairs subsequently selected by ANOVA as having significant differences between the three classes of compounds (FPR1 agonists, FPR2 agonists, and non-active compounds) are shown in dark bars.

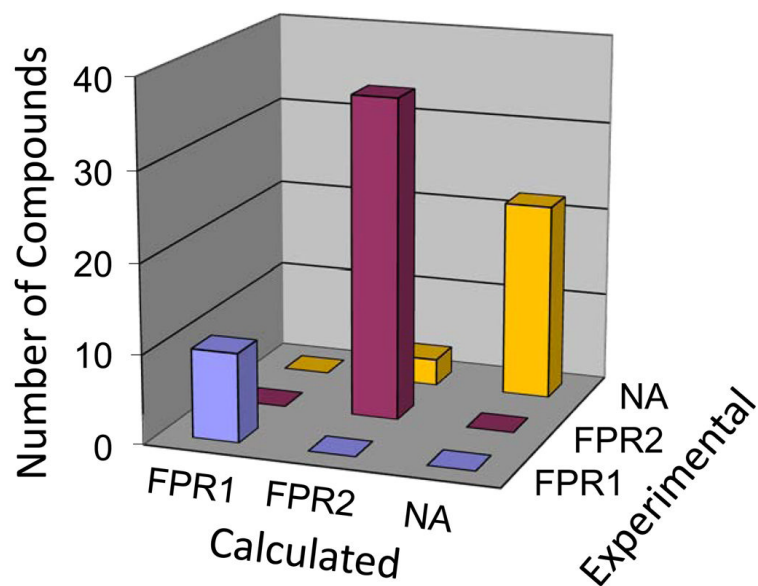
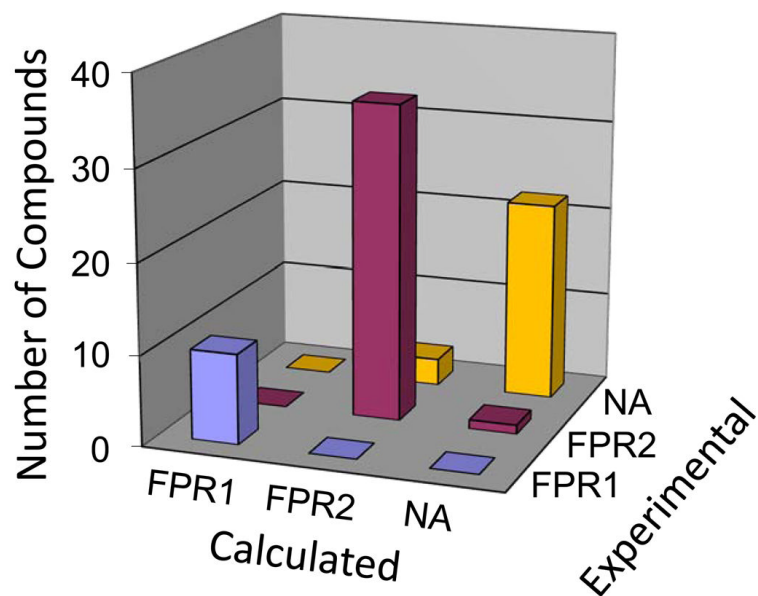
A. LDA**B. Tree**

Figure 3. Classification results of linear discriminant analysis (LDA) (Panel A) and binary classification tree analysis (Panel B) versus experimental classes of compounds investigated. The LDA was based on either 17 or 9 atom pairs from the best subset, and binary classification tree analysis was based on 6 atom pairs.

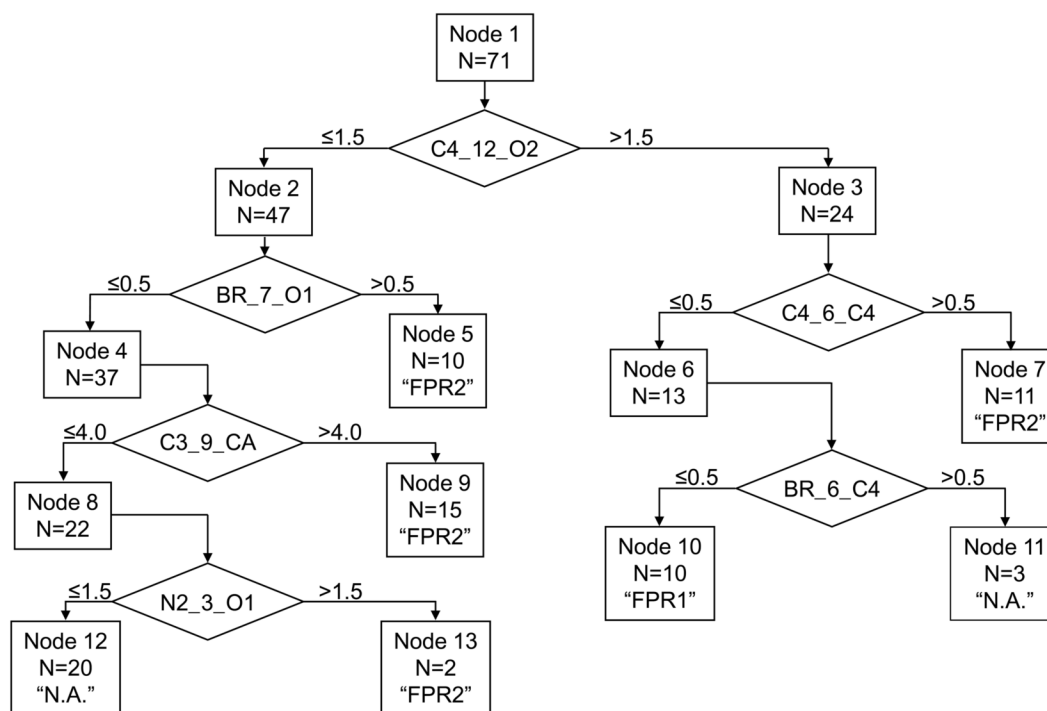


Figure 4. Optimal classification tree for splitting compounds into activity classes. The number of compounds that entered each node is indicated. Terminal nodes correspond to the three activity classes: FPR1-specific agonists, FPR2-specific agonists, or non-active (N.A.).

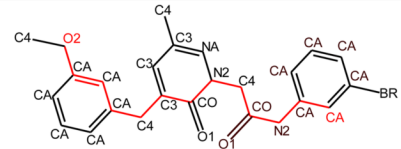
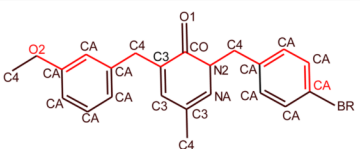
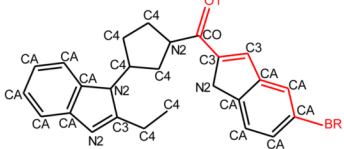
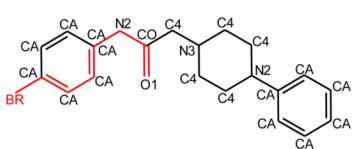
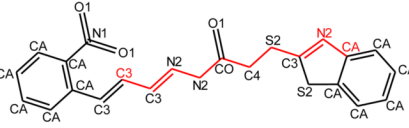
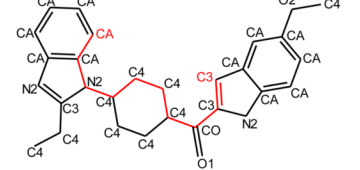
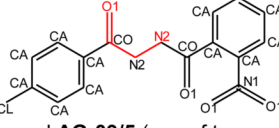
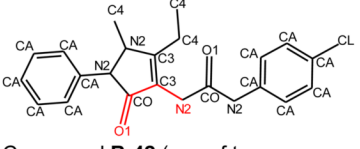
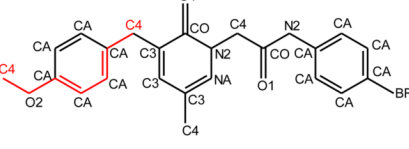
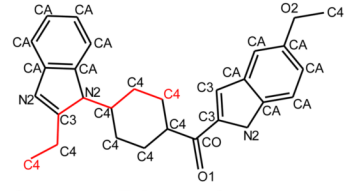

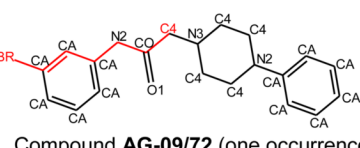
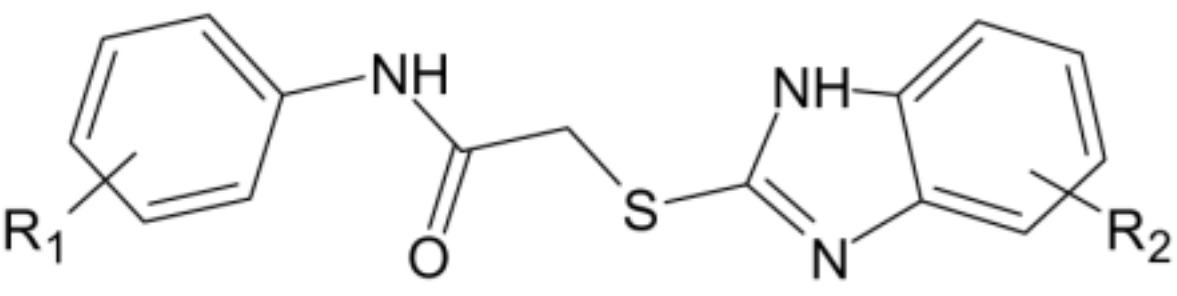
Atom Pair	Example and Number of Occurrences of Atom Pair in Molecule	
CA_12_O2	 <p>A. Compound C-14b (one of two occurrences)</p>	 <p>B. Compound C-18a (one occurrence)</p>
BR_7_O1	 <p>C. Compound F-15 (one occurrence)</p>	 <p>D. Compound AG-09/73 (one occurrence)</p>
C3_9_CA	 <p>E. Compound AG-09/97 (one of five occurrences)</p>	 <p>F. Compound F-6 (one of six occurrences)</p>
N2_3_O1	 <p>G. Compound AG-09/5 (one of two occurrences)</p>	 <p>H. Compound B-42 (one of two occurrences)</p>
C4_6_C4	 <p>I. Compound C-14x (one occurrence)</p>	 <p>J. Compound F-6 (one of two occurrences)</p>
BR_6_C4	 <p>K. Compound C-14b (one occurrence)</p>	 <p>L. Compound AG-09/72 (one occurrence)</p>

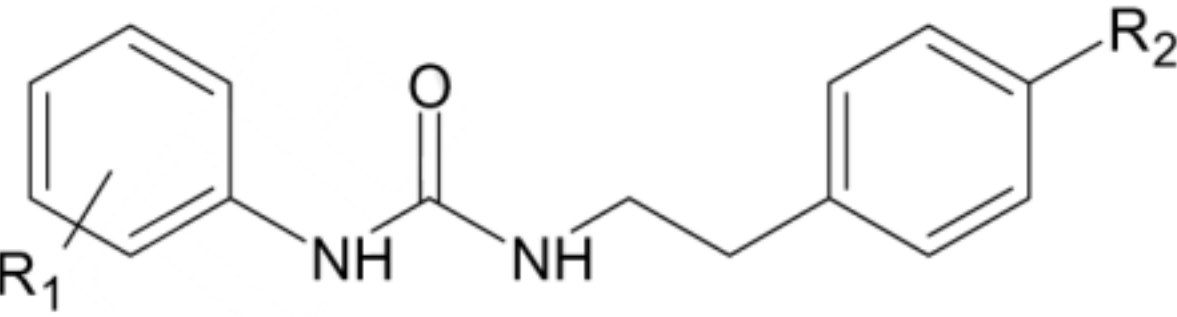
Figure 5.

Examples of atom pair descriptors and their occurrences in structures of selected compounds under investigation. The indicated atom pairs are highlighted in red. Notation of atom types: CA – aromatic carbon; C3 – olefin-type or imino carbon; C4 – tetrahedral sp^3 -hybridized carbon; CO – carbonyl carbon; NA – pyridine nitrogen; N1 – nitrogen in nitro group; N2 – amino-, amido-, or imino-nitrogen; O1 – carbonyl or nitro oxygen; O2 – two-coordinated alcohol or ether-type oxygen; CL – chlorine; BR – bromine.

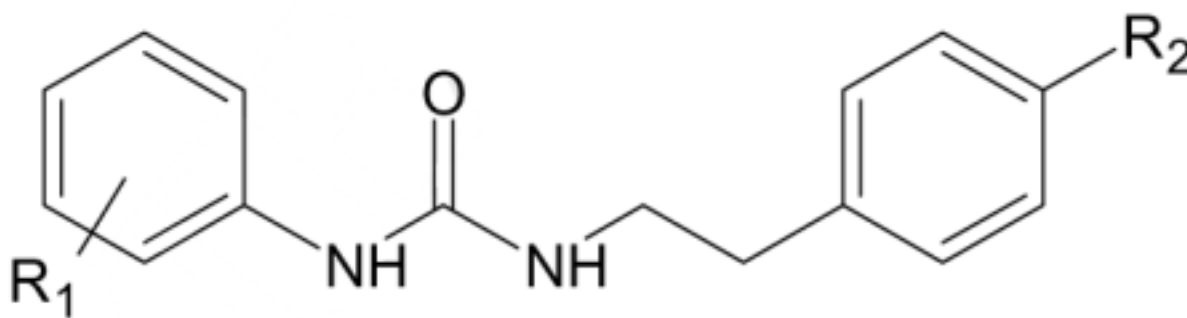
Table 1

Structure and receptor specificity of compounds under investigation

A. Compounds with 2-(benzimidazol-2-ylthio)-N-phenylacetamide scaffold [from [12]].			
			
Compound	R ₁	R ₂	Specificity
AG-09/1	4-NO ₂	5-OMe	FPR1
AG-09/2	4-OEt	5-OMe	FPR1
AG-09/11	H	H	N.A.
AG-09/13	4-OMe	5-OEt	FPR1
AG-09/17	3,4-diOMe	5-OMe	FPR1/FPR2
AG-09/18	H	5-OMe	FPR1
AG-09/19	3-OMe	5-OMe	FPR1
AG-09/20	4-Cl	5-OMe	FPR1/FPR2
AG-09/21	4-Br	5-OMe	FPR1
AG-09/22	2-OMe	5-OMe	FPR1/FPR2
AG-09/25	3-Br	H	N.A.
AG-09/26	3-NO ₂ , 4-Cl	4-Me	N.A.
AG-09/28	4-NO ₂	H	N.A.
AG-09/30	3-CF ₃	5-Me	N.A.

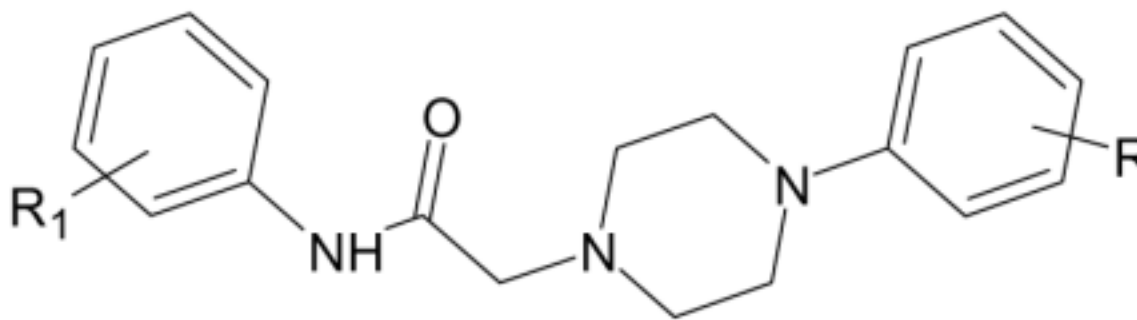
B. Compounds with N-phenethyl-N'-phenylurea scaffold [from [12]].			
			
Compound	R ₁	R ₂	Specificity
AG-26	4-Cl	OMe	FPR2
AG-09/36	H	H	N.A.

B. Compounds with N-phenethyl-N'-phenylurea scaffold [from [12]].



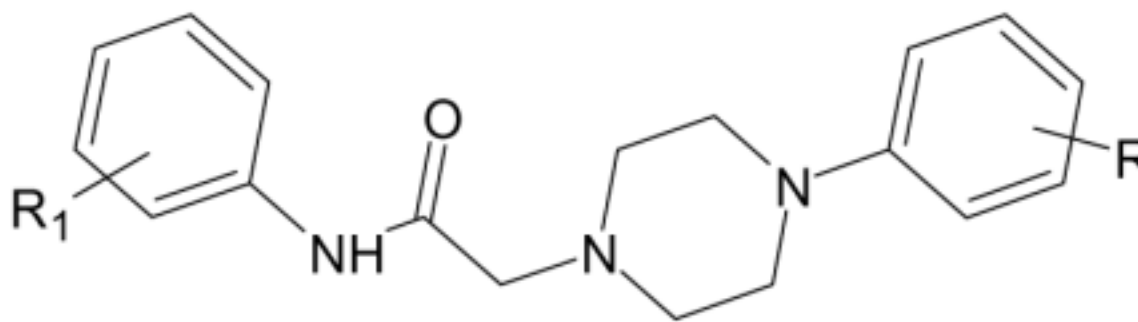
Compound	R ₁	R ₂	Specificity
AG-09/37	H	OMe	FPR2
AG-09/38	4-F	OMe	FPR2
AG-09/41	2,4-diCl	OMe	N.A.
AG-09/42	4-Br	OMe	FPR2
AG-09/45	4-F	Me	N.A.
AG-09/52	2-NO ₂	Cl	N.A.
AG-09/54	3-Me	Cl	N.A.

C. Compounds with piperazine scaffold [from [12]].



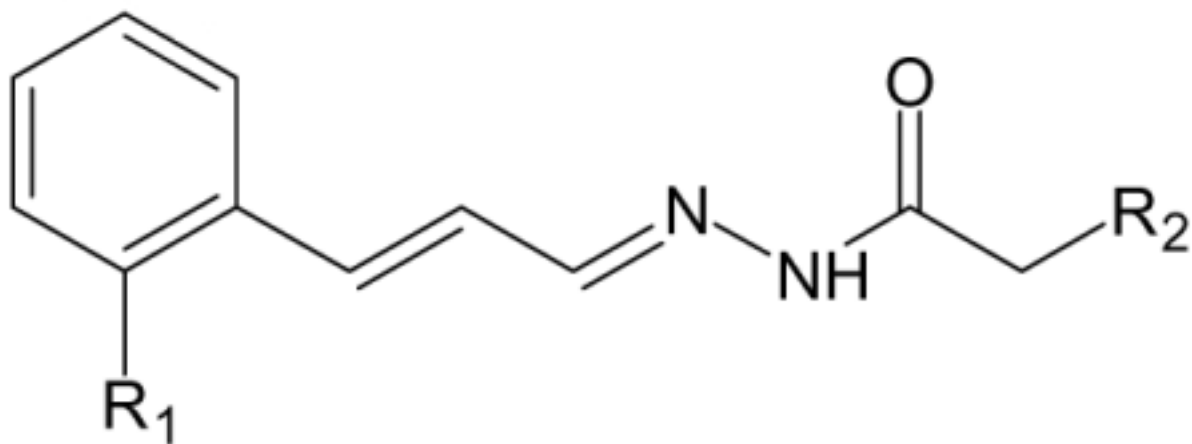
Compound	R ₁	R ₂
AG-09/03	4-Br	4-F
AG-09/04	4-Br	3-Cl
AG-09/55	H	H
AG-09/56	H	2-Cl
AG-09/64	3-Cl,4-F	H
AG-09/71	2-Br	H
AG-09/72	3-Br	H
AG-09/73	4-Br	H
AG-09/74	4-Br	4-Cl
AG-09/75	4-Br	3-Cl, 4-Me
AG-09/76	4-Br	2-Me, 3-Cl

C. Compounds with piperazine scaffold [from [12]].



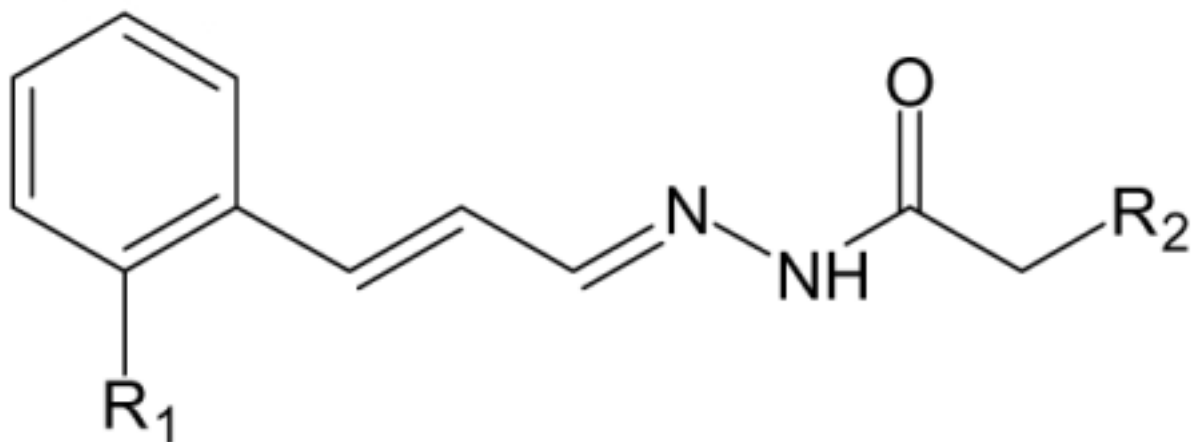
Compound	R ₁	R ₂
AG-09/77	4-Br	2-Me, 5-Cl
AG-09/82	4-Br	3-OMe
AG-09/91		

D. Compounds with acetohydrazide scaffold [from [12]].



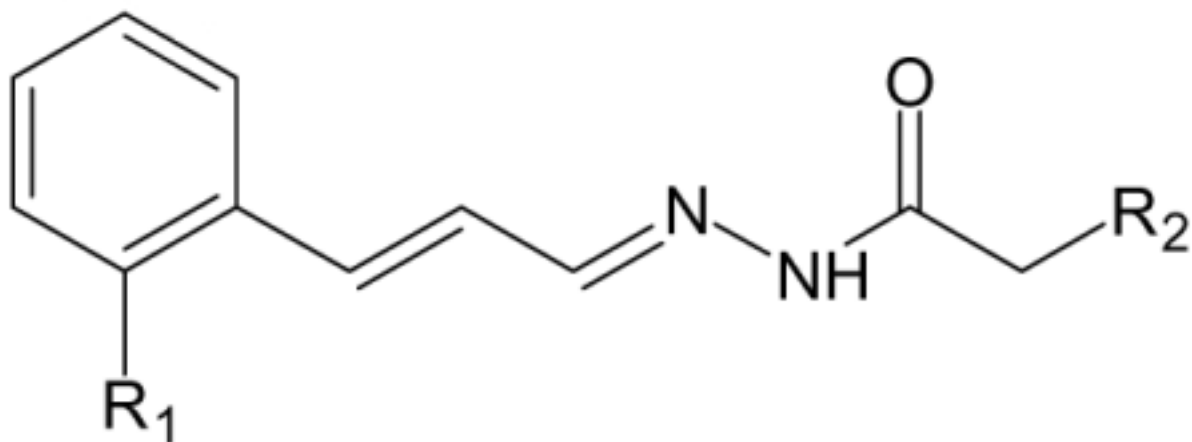
Compound	R ₁	R ₂
AG-09/92	NO ₂	

D. Compounds with acetohydrazone scaffold [from [12]].



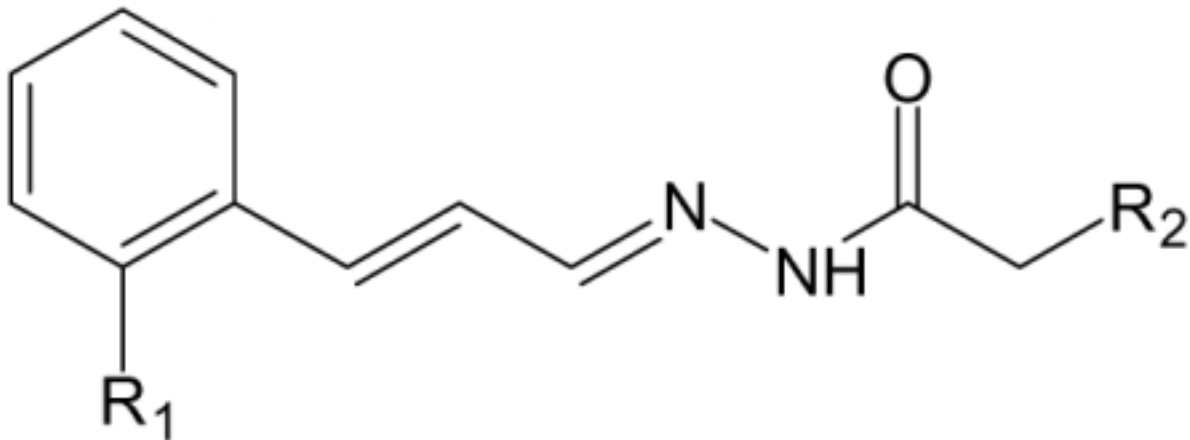
Compound	R_1	R_2
AG-09/93	NO_2	<p>Chemical structure showing a benzene ring substituted with R_2.</p>

D. Compounds with acetohydrazide scaffold [from [12]].



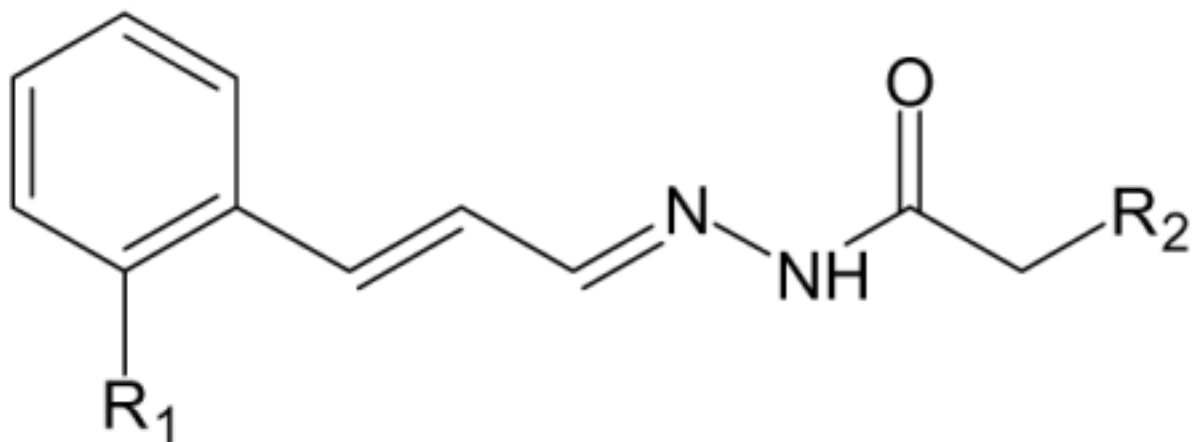
Compound	R_1	R_2
AG-09/95	NO_2	<p>Chemical structure of a 1,3-butadiene chain, which is the R_2 group for compound AG-09/95.</p>

D. Compounds with acetohydrazide scaffold [from [12]].



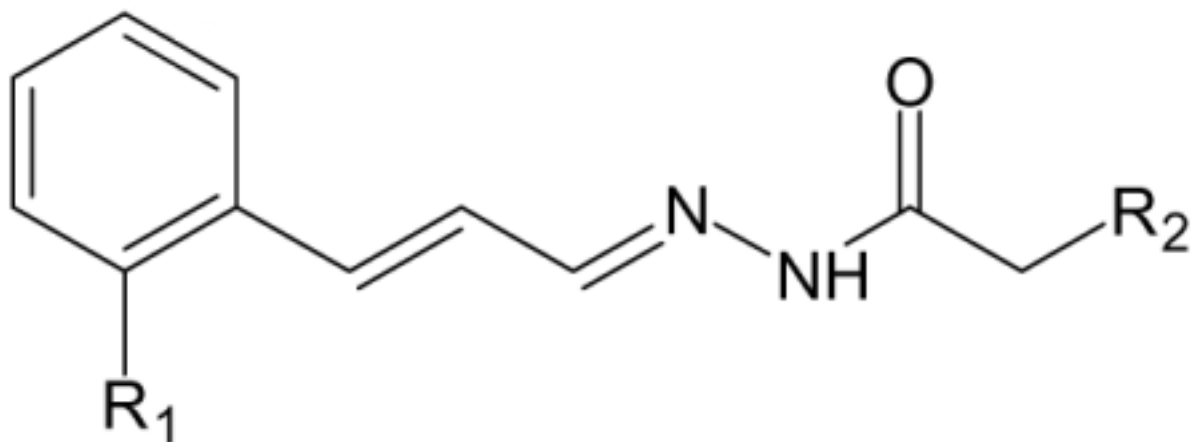
Compound	R ₁	R ₂
AG-09/96	NO ₂	
AG-09/97	NO ₂	

D. Compounds with acetohydrazone scaffold [from [12]].



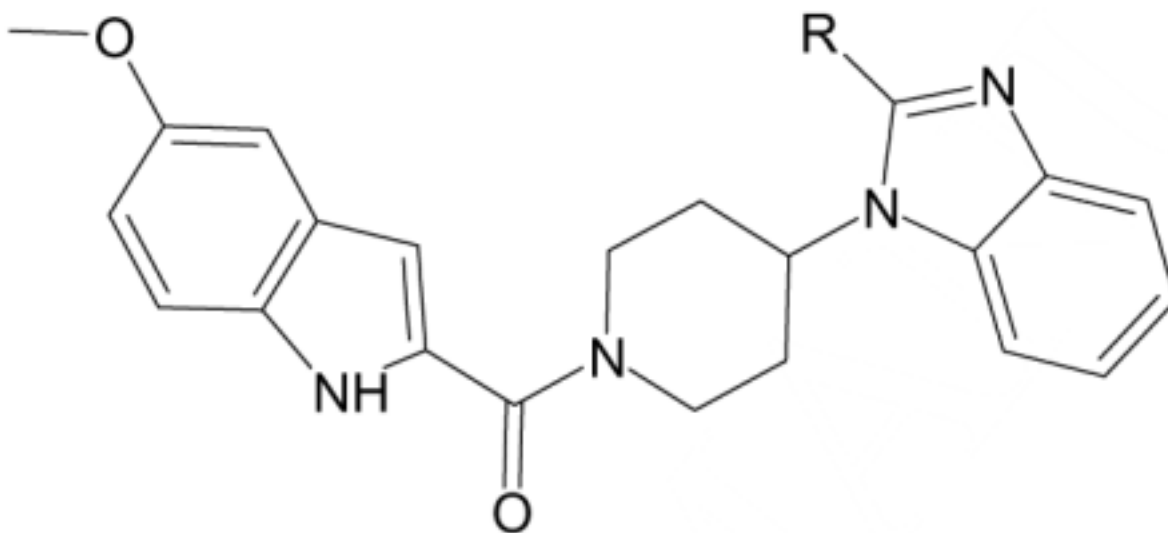
Compound	R ₁	R ₂
AG-09/99	OMe	
AG-09/101		

D. Compounds with acetohydrazide scaffold [from [12]].



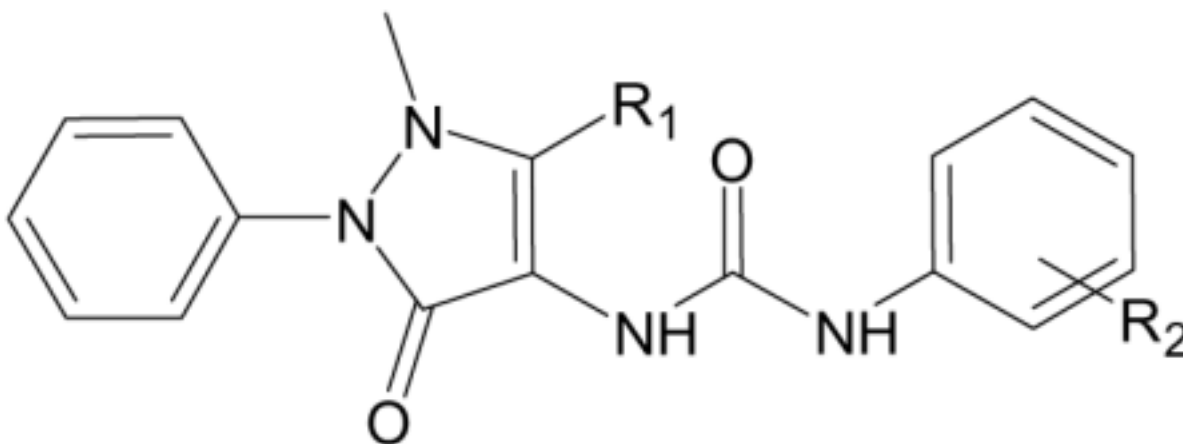
Compound	R_1	R_2
AG-09/102	<p>Structure of R_1 for AG-09/102: A 4-bromophenyl group.</p>	<p>Structure of R_2 for AG-09/102: A complex bicyclic system consisting of a benzene ring fused to a five-membered ring containing a nitrogen atom and two carbonyl groups.</p>
AG-09/7		<p>Structure of R_2 for AG-09/7: A complex bicyclic system consisting of a benzene ring fused to a five-membered ring containing a nitrogen atom and two carbonyl groups, with a nitro group attached to the benzene ring.</p>

E. Specific FPR2 agonists with 1-(2-indolylcarbonyl)-4-(1-benzimidazolyl)piperidine scaffolds [from [16]].



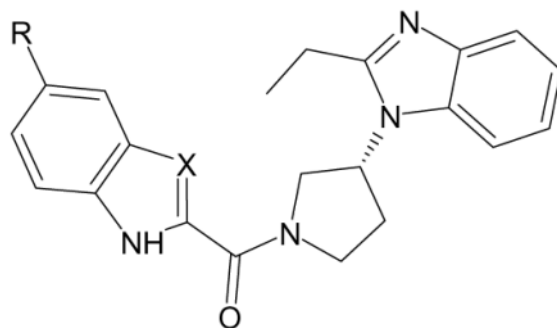
Compound	R	Specificity
F-5	Me	FPR2
F-6	Et	FPR2
F-7	iBu	FPR2
F-8	CF ₃	FPR2
F-9	Ph	FPR2

F. Specific FPR2 agonists with pyrazolone scaffold [from [9,31]].



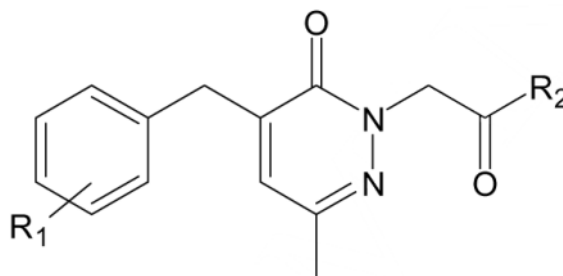
Compound	R ₁	R ₂	Specificity
B-25	Me	4-Br	FPR2
B-35	Me	4-SMe	FPR2
B-42	Et	4-Cl	FPR2
B-43	iPro	4-Cl	FPR1/FPR2

G. Specific FPR2 agonists with 1-(2-indolylcarbonyl)-3-(1-benzimidazolyl)pyrrolidine and related scaffolds [from [16]].



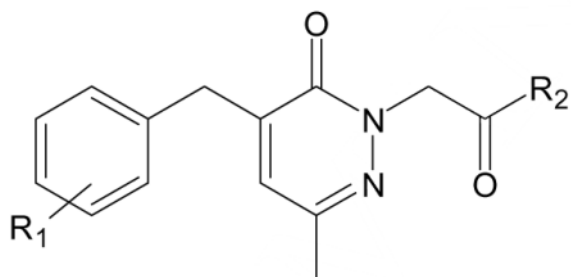
Compound	R	X	Specificity
F-10	OMe	CH	FPR2
F-13	F	CH	FPR2
F-14	Cl	CH	FPR2
F-15	Br	CH	FPR2
F-16	OCF ₃	CH	FPR2
F-17	Et	CH	FPR2
F-18	OMe	N	FPR2

H. Compounds with 4-benzylpyridazin-3-one scaffold [from [17]].

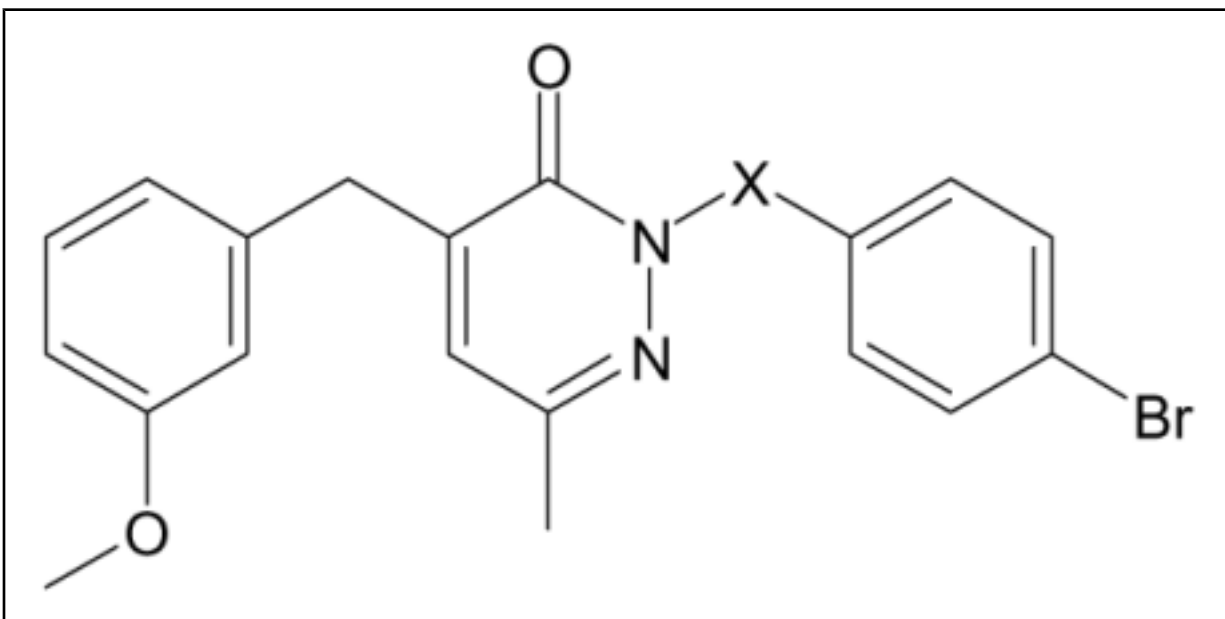


Compound	R1	R2	S
C-14a	3-OMe	NH-C ₆ H ₄ -Br (p)	F
C-14b	3-OMe	NH-C ₆ H ₄ -Br (m)	
C-14e	3-OMe	NH-C ₆ H ₄ -I (p)	F
C-14f	3-OMe	NH-C ₆ H ₄ -F (p)	
C-14h	3-OMe	NH-C ₆ H ₄ -CH ₃ (p)	F

H. Compounds with 4-benzylpyridazin-3-one scaffold [from [17]].



Compound	R1	R2	S
C-14m	3-OMe		
C-14n	3-OMe	NH-C ₆ H ₄ -CF ₃ (p)	F
C-14r	3-OMe	N(CH ₃)-C ₆ H ₄ -Br (p)	
C-14x	4-OMe	NH-C ₆ H ₄ -Br (p)	
C-6c			

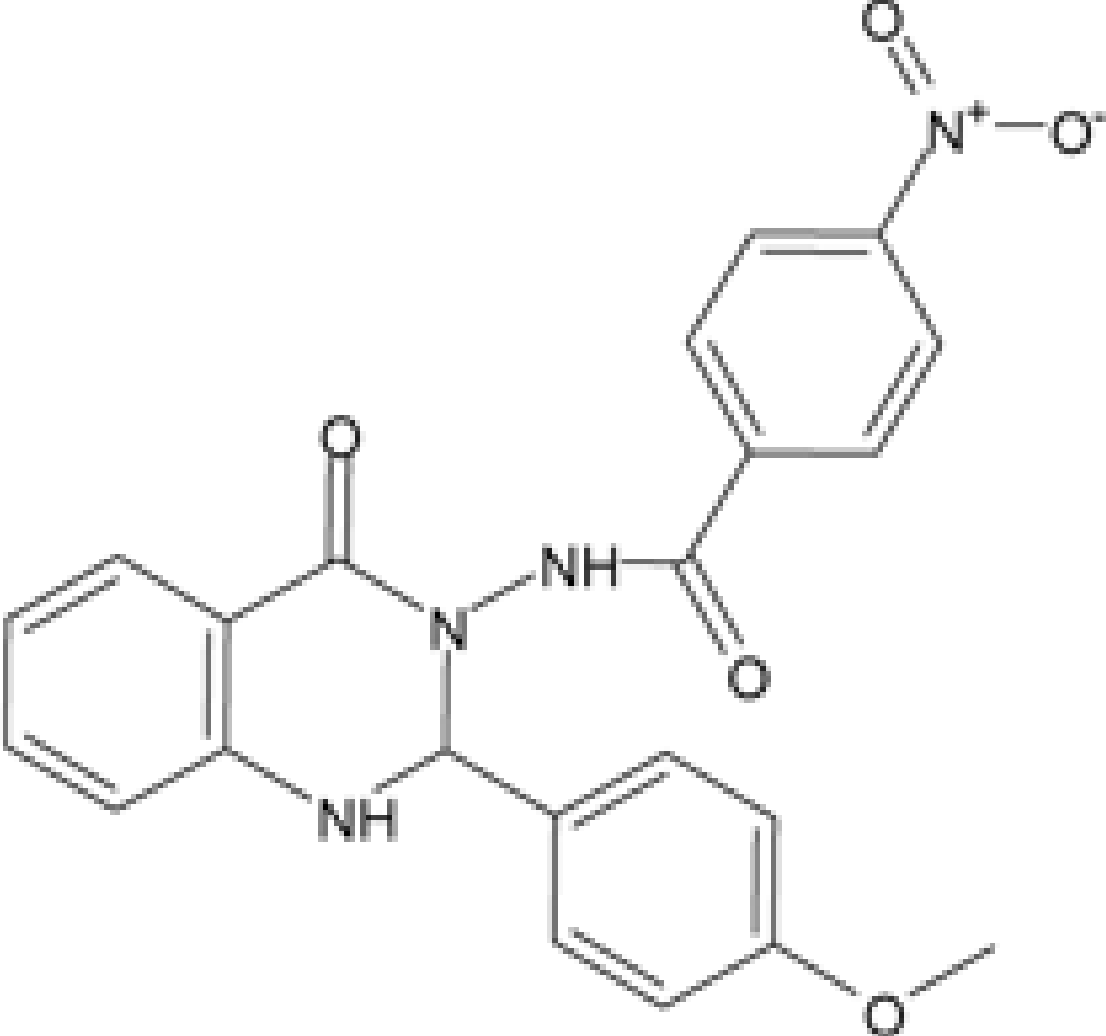
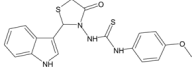
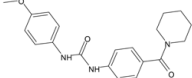
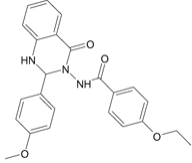
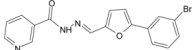


Compound	X	Specificity
C-17b	CH(CH ₃)CONH	FPR1
C-18a	CH ₂	N.A.
C-23	(CH ₂) ₂ NH	N.A.

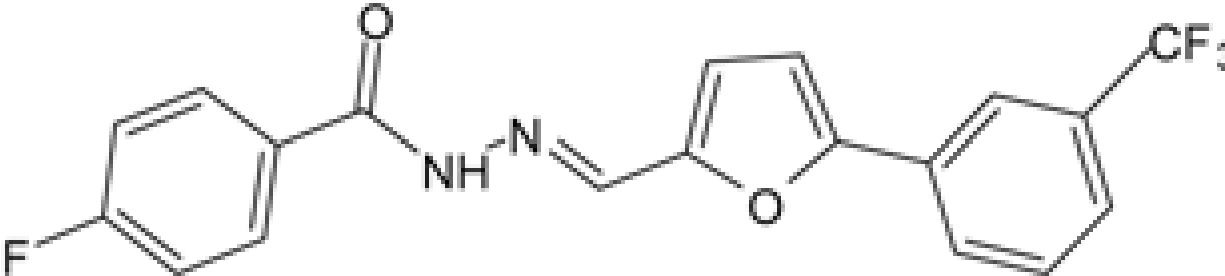
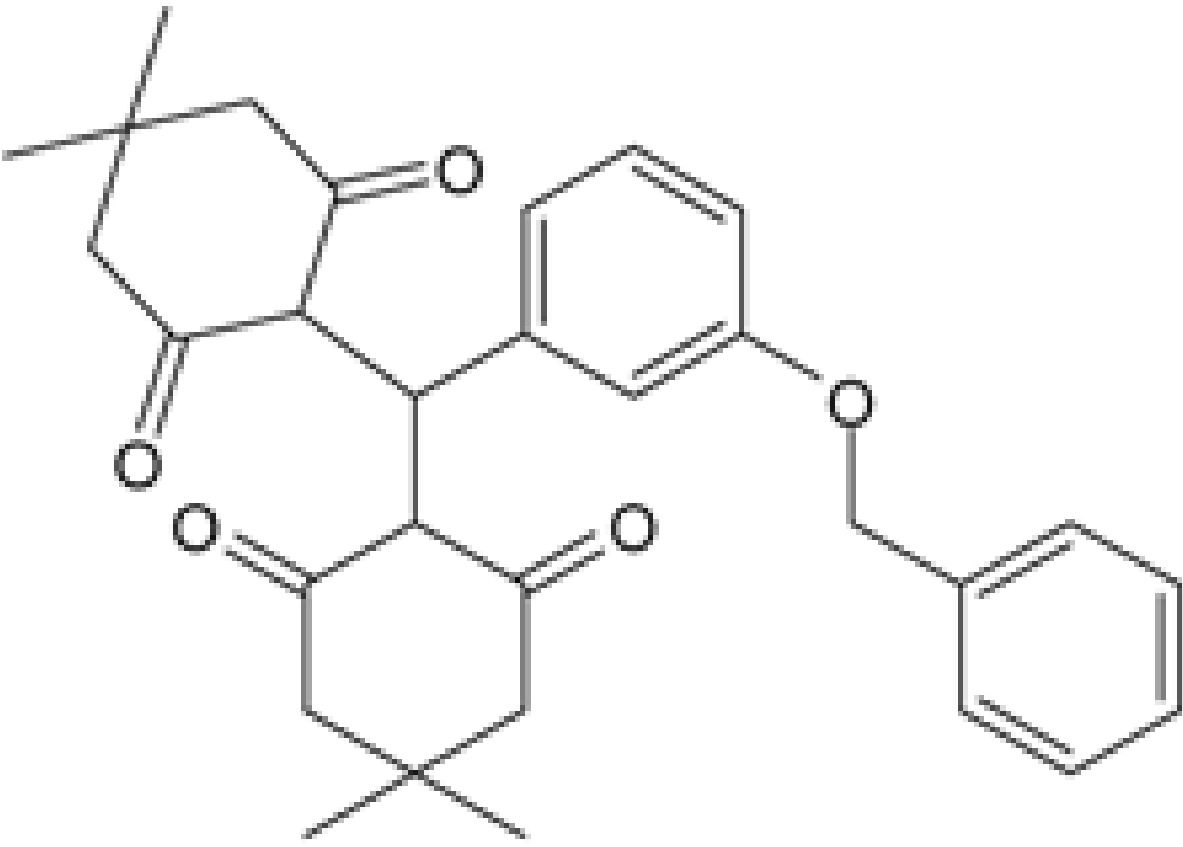
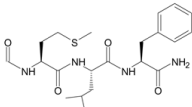
I. Structures of other FPR1/FPR2 agonists.

Compound	Structure
AG-14	
AG-22	
AG-09/5	
AG-09/6	

I. Structures of other FPR1/FPR2 agonists.

Compound	Structure
AG-09/8	 <p>The structure of AG-09/8 is a complex molecule. It features a central benzimidazole ring system. One nitrogen of the benzimidazole is substituted with a 4-nitrophenyl group. The other nitrogen is substituted with a 4-methoxyphenyl group. The 2-position of the benzimidazole ring is substituted with a carbonyl group, which is further substituted with a 4-methoxyphenyl group.</p>
AG-09/9	 <p>The structure of AG-09/9 is a small molecule featuring a benzimidazole core. It has a 4-methoxyphenyl group attached to one of the nitrogens and a 4-methoxyphenyl group attached to the 2-position of the benzimidazole ring.</p>
AG-09/10	 <p>The structure of AG-09/10 is a small molecule featuring a benzimidazole core. It has a 4-methoxyphenyl group attached to one of the nitrogens and a 4-methoxyphenyl group attached to the 2-position of the benzimidazole ring.</p>
Quin-C1	 <p>The structure of Quin-C1 is a small molecule featuring a benzimidazole core. It has a 4-methoxyphenyl group attached to one of the nitrogens and a 4-methoxyphenyl group attached to the 2-position of the benzimidazole ring.</p>
S-1	 <p>The structure of S-1 is a small molecule featuring a benzimidazole core. It has a 4-methoxyphenyl group attached to one of the nitrogens and a 4-methoxyphenyl group attached to the 2-position of the benzimidazole ring.</p>

I. Structures of other FPR1/FPR2 agonists.

Compound	Structure
S-2	
1910-5441	
fMLF	

I. Structures of other FPR1/FPR2 agonists.

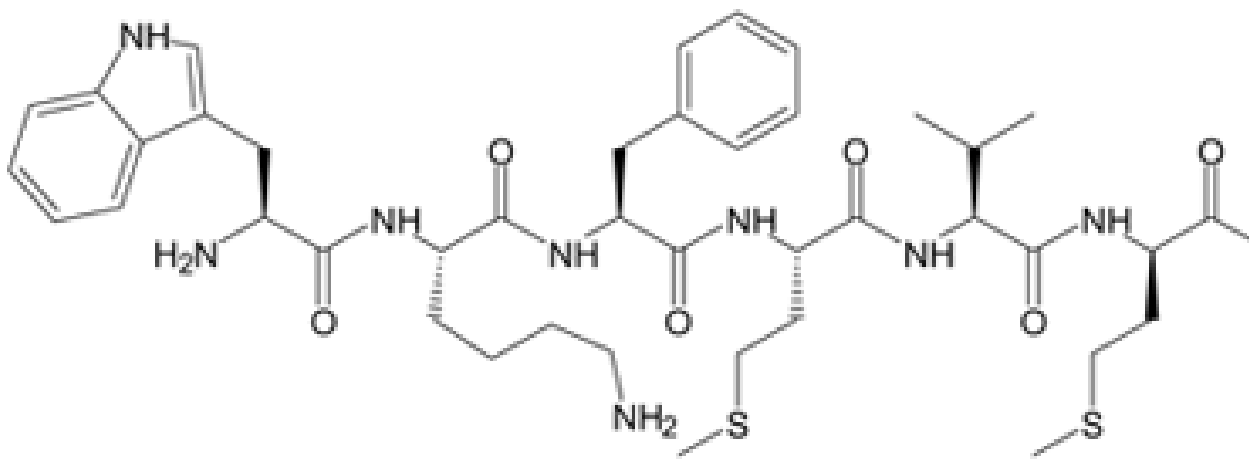
Compound	Structure
WKYMVm	 <chem>CC(C)C[C@@H](CS)C(=O)N[C@@H](Cc1ccccc1)C(=O)N[C@@H](CCCCN)C(=O)N[C@@H](C)C(=O)N[C@@H](Cc2c[nH]c3ccccc23)C(=O)N</chem>

Table 2Clusters of highly correlated atom pairs ($r \geq 0.9$)

Cluster	Atom pairs
1	C3_1_NA, N2_1_NA, C3_2_NA, C4_2_NA, CO_2_NA, C3_3_NA, NA_3_O1, CA_5_NA, C3_6_O2, CA_6_NA, C3_7_O2, CA_7_NA, CA_8_NA
2	C3_1_S2, C4_1_S2, CO_2_S2, N2_2_S2, CA_3_S2, N2_3_S2, O1_3_S2, CA_4_S2, CA_5_S2, CA_6_S2, CA_7_S2
3	C4_1_CA, C4_2_N2, C4_4_CA, C4_4_N2, C4_5_CA, C4_5_N2, C4_6_CA, C4_7_CA
4	C4_1_CO, C4_2_O1
5	C4_1_O2, CA_1_O2, CA_2_O2, CA_3_O2, CA_4_O2
6	C4_2_C4, C4_3_C4
7	CO_2_N2, N2_3_O1
8	C4_3_CO, C4_4_O1
9	C3_4_C4, C3_5_C4, C3_8_N2
10	CO_4_N2, N2_5_O1
11	CO_4_O2, CA_5_O2, O1_5_O2, CA_6_O2, O2_11_O2, CA_12_CO, O2_12_O2, CA_13_CO, CA_17_CA, CA_17_O2, C4_18_CA, CA_18_O2
12	C4_5_CO, C4_6_O1
13	CO_5_O2, O1_6_O2
14	BR_6_CO, BR_7_O1
15	C3_6_CO, C3_7_O1
16	CO_6_N2, N2_8_N2
17	CO_6_O2, O1_7_O2
18	O2_6_S2, C4_7_S2
19	BR_8_N3, BR_10_C4, BR_11_N2
20	CO_8_O2, O1_9_O2
21	F_8_N2, CO_9_F, F_9_NA, F_10_O1, F_15_O2, C4_16_F
22	NA_8_O2, C4_9_NA
23	O2_8_S2, C4_9_S2, C4_17_C4
24	C4_9_O2, C4_10_C4
25	CO_9_O2, O1_10_O2
26	N2_9_O2, C4_10_N2
27	O2_14_O2, C4_16_O2
28	C4_15_O2, C4_16_C4

Atom pairs with longer bond separations are indicated in bold italic.

Table 3

Experimentally determined, SAR-calculated, and LOO-predicted classes of FPR1/FPR2 agonist activity for FPR1/FPR2 agonists and non-active compounds (training set) and their atom pairs used in binary classification tree analysis

Compound	Determined	LDA		Binary Classification Tree										Terminal Node	Calculated
		Calculated	LOO-predicted	CA_12_O2	C4_6_C4	BR_6_C4	BR_7_O1	C3_9_CA	N2_3_O1						
AG-14	FPR1	FPR1	FPR1	2	0	0	0	0	2	1	10	FPR1			
AG-26	FPR2	FPR2	FPR2	2	1	0	0	0	0	0	7	FPR2			
AG-09/1	FPR1	FPR1	FPR1	2	0	0	0	0	0	0	10	FPR1			
AG-09/2	FPR1	FPR1	FPR1	4	0	0	0	0	0	0	10	FPR1			
AG-09/3	FPR2	FPR2	FPR2	0	0	0	1	0	0	0	5	FPR2			
AG-09/4	FPR2	FPR2	FPR2	0	0	0	1	0	0	0	5	FPR2			
AG-09/5	FPR2	FPR2	FPR2	0	0	0	0	0	0	0	13	FPR2			
AG-09/6	FPR2	FPR2	FPR2	0	0	0	0	0	6	0	9	FPR2			
AG-09/7	FPR2	FPR2	FPR2	0	0	0	0	0	5	1	9	FPR2			
AG-09/8	FPR2	FPR2	FPR2	1	1	0	0	0	0	2	13	FPR2			
AG-09/11	NA	NA	NA	0	0	0	0	0	0	0	12	NA			
AG-09/13	FPR1	FPR1	FPR1	4	0	0	0	0	0	0	10	FPR1			
AG-09/18	FPR1	FPR1	FPR1	2	0	0	0	0	0	0	10	FPR1			
AG-09/19	FPR1	FPR1	FPR1	4	0	0	0	0	0	0	10	FPR1			
AG-09/21	FPR1	FPR1	FPR2	2	0	0	1	0	0	0	10	FPR1			
AG-09/25	NA	NA	NA	0	0	1	0	0	0	0	12	NA			
AG-09/26	NA	NA	NA	0	1	0	0	0	0	0	12	NA			
AG-09/28	NA	NA	NA	0	0	0	0	0	0	0	12	NA			
AG-09/30	NA	NA	NA	0	1	0	0	0	0	0	12	NA			
AG-09/36	NA	NA	NA	0	0	0	0	0	0	0	12	NA			
AG-09/37	FPR2	FPR2	FPR2	2	1	0	0	0	0	0	7	FPR2			
AG-09/38	FPR2	FPR2	FPR2	2	1	0	0	0	0	0	7	FPR2			
AG-09/41	NA	FPR2	FPR2	2	1	0	0	0	0	0	7	FPR2			
AG-09/42	FPR2	FPR2	FPR2	2	1	0	1	0	0	0	7	FPR2			
AG-09/45	NA	NA	NA	0	1	0	0	0	0	0	12	NA			

Compound	Determined	LDA		Binary Classification Tree										Calculated	
		Calculated	LOO-predicted	CA_12_O2	C4_6_C4	BR_6_C4	BR_7_O1	C3_9_CA	N2_3_O1	Terminal Node	Calculated				
AG-09/52	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/54	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/55	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/56	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/64	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/71	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/72	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/73	FPR2	FPR2	FPR2	0	0	0	0	0	1	0	0	0	0	5	FPR2
AG-09/74	FPR2	FPR2	FPR2	0	0	0	0	0	1	0	0	0	0	5	FPR2
AG-09/75	FPR2	FPR2	FPR2	0	2	0	0	0	1	0	0	0	0	5	FPR2
AG-09/76	FPR2	FPR2	FPR2	0	0	0	0	0	1	0	0	0	0	5	FPR2
AG-09/77	FPR2	FPR2	FPR2	0	0	0	0	0	1	0	0	0	0	5	FPR2
AG-09/82	FPR2	FPR2	FPR2	2	2	0	0	0	1	0	0	0	0	7	FPR2
AG-09/91	NA	NA	NA	0	0	0	0	0	0	0	0	0	0	12	NA
AG-09/92	FPR2	FPR2	FPR1	3	1	0	0	0	0	0	1	1	1	7	FPR2
AG-09/93	FPR2	FPR2	FPR2	0	0	0	0	0	1	0	1	1	1	5	FPR2
AG-09/95	NA	FPR2	FPR2	0	0	0	0	0	0	0	0	1	1	12	NA
AG-09/96	FPR2	FPR2	FPR2	0	0	0	0	0	0	0	0	5	2	9	FPR2
AG-09/97	NA	NA	NA	0	0	0	0	0	0	0	0	5	1	9	FPR2
AG-09/99	NA	NA	NA	0	0	0	0	0	0	0	0	3	1	12	NA
AG-09/101	FPR2	FPR2	FPR2	0	1	0	0	0	0	0	0	5	2	9	FPR2
AG-09/102	NA	FPR2	FPR2	0	0	0	0	0	0	0	0	5	1	9	FPR2
C-6c	NA	NA	NA	0	0	0	0	0	0	0	0	1	1	12	NA
C-14b	NA	NA	NA	2	0	0	0	0	1	0	0	4	1	11	NA
C-14f	FPR1	FPR1	FPR2	2	0	0	0	0	0	0	0	4	1	10	FPR1
C-14m	FPR1	FPR1	FPR1	5	0	0	0	0	0	0	0	4	1	10	FPR1
C-14r	NA	NA	FPR1	2	0	0	0	0	1	1	0	4	1	11	NA
C-14x	FPR2	FPR2	FPR2	1	1	0	0	0	1	0	0	4	1	5	FPR2
C-17b	FPR1	FPR1	FPR2	2	0	0	0	0	1	0	0	4	1	10	FPR1

Compound	Determined	LDA		Binary Classification Tree									
		Calculated	LOO-predicted	CA_12_O2	C4_6_C4	BR_6_C4	BR_7_O1	C3_9_CA	N2_3_O1	Terminal Node	Calculated		
C-18a	NA	NA	NA	1	0	0	0	0	0	0	0	12	NA
C-23	NA	NA	NA	2	0	1	0	4	0	0	0	11	NA
F-5	FPR2	FPR2	FPR2	1	0	0	0	6	1	1	1	9	FPR2
F-6	FPR2	FPR2	FPR2	1	2	0	0	6	1	1	1	9	FPR2
F-7	FPR2	FPR2	FPR2	1	6	0	0	6	1	1	1	9	FPR2
F-8	FPR2	FPR2	FPR2	1	0	0	0	6	1	1	1	9	FPR2
F-9	FPR2	FPR2	FPR2	1	0	0	0	9	1	1	1	9	FPR2
F-10	FPR2	FPR2	FPR2	2	1	0	0	5	1	1	7	7	FPR2
F-13	FPR2	FPR2	FPR2	0	1	0	0	5	1	1	9	9	FPR2
F-14	FPR2	FPR2	FPR2	0	1	0	0	5	1	1	9	9	FPR2
F-15	FPR2	FPR2	FPR2	0	1	0	1	5	1	1	5	5	FPR2
F-16	FPR2	FPR2	FPR2	2	1	0	0	5	1	1	7	7	FPR2
F-17	FPR2	FPR2	FPR2	0	1	0	0	5	1	1	9	9	FPR2
F-18	FPR2	FPR2	FPR2	2	1	0	0	3	2	2	7	7	FPR2
S-1	FPR2	FPR2	FPR2	0	0	0	0	3	1	1	12	12	NA
S-2	FPR2	FPR2	FPR2	0	0	0	0	5	1	1	9	9	FPR2
Quin-C1	FPR2	FPR2	FPR2	3	1	0	0	0	2	2	7	7	FPR2

ND, not determined; NA, non active. Incorrect classifications or predictions are indicated in bold italic.

Experimentally determined and predicted classes of FPR1/FPR2 agonists from the test set and their atom pairs used in binary classification tree analysis

Table 4

Compound	Determined	CA_12_O2	C4_6_C4	BR_6_C4	BR_7_O1	C3_9_CA	N2_3_O1	Terminal Node	Predicted
AG-22	FPR1/FPR2	0	1	0	0	8	1	9	FPR2
AG-09/9	FPR1/FPR2	2	0	0	0	3	1	10	FPR1
AG-09/10	FPR1/FPR2	0	0	0	0	0	0	12	NA
AG-09/17	FPR1/FPR2	6	0	0	0	0	0	10	FPR1
AG-09/20	FPR1/FPR2	2	0	0	0	0	0	10	FPR1
AG-09/22	FPR1/FPR2	2	0	0	0	0	0	10	FPR1
B-25	FPR2	0	0	0	1	0	2	5	FPR2
B-35	FPR2	0	0	0	0	0	2	13	FPR2
B-42	FPR2	0	0	0	0	0	2	13	FPR2
B-43	FPR1/FPR2	0	0	0	0	0	2	13	FPR2
C-14a	FPR1/FPR2	2	0	0	1	4	1	10	FPR1
C-14e	FPR1/FPR2	2	0	0	0	4	1	10	FPR1
C-14h	FPR1/FPR2	2	0	0	0	4	1	10	FPR1
C-14n	FPR1/FPR2	2	0	0	0	4	1	10	FPR1
1910-5441	FPR1/ND	0	10	0	0	0	0	12	NA
fMLF	FPR1/FPR2	0	8	0	0	0	3	13	FPR2
WKYMVm	FPR1/FPR2	0	15	0	0	0	5	13	FPR2

ND, not determined; NA, non active.