Virtual Screen for Ligands of Orphan G Protein-Coupled Receptors

Joel R. Bock, and David A. Gough

J. Chem. Inf. Model., **2005**, 45 (5), 1402-1414• DOI: 10.1021/ci050006d • Publication Date (Web): 30 June 2005

Virtual Screen for Ligands of Orphan G Protein-Coupled Receptors

Joel R. Bock and David A. Gough*

Department of Bioengineering, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412

Received January 5, 2005

This paper describes a virtual screening methodology that generates a ranked list of high-binding small molecule ligands for orphan G protein-coupled receptors (oGPCRs), circumventing the requirement for receptor three-dimensional structure determination. Features representing the receptor are based only on physicochemical properties of primary amino acid sequence, and ligand features use the two-dimensional atomic connection topology and atomic properties. An experimental screen comprised nearly 2 million hypothetical oGPCR−ligand complexes, from which it was observed that the top 1.96% predicted affinity scores corresponded to "highly active" ligands against orphan receptors. Results representing predicted high-scoring novel ligands for many oGPCRs are presented here. Validation of the method was carried out in several ways: (1) A random permutation of the structure−activity relationship of the training data was carried out; by comparing test statistic values of the randomized and nonshuffled data, we conclude that the value obtained with nonshuffled data is unlikely to have been encountered by chance. (2) Biological activities linked to the compounds with high cross-target binding affinity were analyzed using computed log-odds from a structure-based program. This information was correlated with literature citations where GPCR-related pathways or processes were linked to the bioactivity in question. (3) Anecdotal, out-of-sample predictions for nicotinic targets and known ligands were performed, with good accuracy in the low-to-high "active" binding range. (4) An out-of-sample consistency check using the commercial antipsychotic drug olanzapine produced "active" to "highly-active" predicted affinities for all oGPCRs in our study, an observation that is consistent with documented findings of cross-target affinity of this compound for many different GPCRs. It is suggested that this virtual screening approach may be used in support of the functional characterization of oGPCRs by identifying potential cognate ligands. Ultimately, this approach may have implications for pharmaceutical therapies to modulate the activity of faulty or disease-related cellular signaling pathways. In addition to application to cell surface receptors, this approach is a generalized strategy for discovery of small molecules that may bind intracellular enzymes and involve protein−protein interactions.

1. INTRODUCTION

An essential model in cellular signal transduction is the three-component system of discriminator, transducer, and amplifier.[1] In this model, the "discriminator" (a plasma membrane-bound receptor) is activated through the chemical binding of ligands, which may be hormones, neurotransmitters, peptides, or small molecules. The specific context of the signal reception event encodes a message to be conveyed across the membrane into the cellular interior by the "transducer". Finally, the "amplifiers" (effector molecules) boost the chemical signal strength and relay its information to various cytoplasmic or nuclear targets. This regulatory mechanism connects a stimulation or binding event at the cell surface with its consequent intracellular physiological effect.

An important superfamily of cell surface receptors which implement this signal transduction paradigm are the *G protein-coupled receptors* (GPCRs), so-named for their mediation of intracellular heterotrimeric G proteins.[2] The molecular mechanisms underlying the modulation of GPCR-stimulated signaling and the connection to other cellular signaling pathways may be quite elaborate.[3] Defective signaling in cells is often closely linked to disease.[4] Dysfunctional GPCR-mediated signal transduction systems in particular have been shown to play a role in a number of pathological states, including endocrine diseases,[5] cancer,[6,7] retinitis pigmentosa,[8] nephrogenic diabetes insipidus,[9] neurological or psychiatric disorders,[10] asthma and rhinitis,[11] and cardiac disease.[12]

1.1. GPCRs Are Important Drug Targets. G protein-coupled receptors have proven to be excellent targets for pharmaceutical treatment; along with kinases, GPCRs constitute the most widely screened classes of signal transduction targets.[13] Estimates suggest that GPCRs comprise 50−60% of currently marketed drugs, including 30% of the top-selling 100 drugs.[14,15] Beyond the intrinsic association to disease-related signaling pathways, GPCR agonist or antagonist drugs have been therapeutically successful because of their direct activity on the cell surface.[16] In commercial terms, GPCRs will continue to predominate as drug targets, largely because they have been successfully targeted in the past. The tremendous cost to develop new drugs creates pressure to avert risk, motivating the focus on "precedented targets". In the 1990s, 74% of the drug products launched with annual sales exceeding $1 billion were associated with precedented approaches.[17]

* Corresponding author phone: (858)822-3446; fax: (858)534-5722; e-mail: dgough@bioeng.ucsd.edu.

10.1021/ci050006d CCC: $30.25
Published on Web 06/30/2005

Even though GPCRs have been intensely investigated as potential drug targets, their structural and functional diversity[18,19] still present opportunities to develop novel drugs. Analysis of the human genomic sequence suggests there may be 750 human GPCR-encoding genes, of which approximately 160 cannot be functionally characterized either on the basis of sequence homology or by association with known endogenous ligands.[20] These are referred to as *orphan* GPCRs-receptors (oGPCRs) which bind (as yet) unknown ligands.[21,22] The physiological role of oGPCRs can only be elucidated by first identifying cognate peptides or small molecule ligands which modulate their function. Afterward, a significant task remains—specifically, to establish bioactivity in the face of nonspecific GPCR ligand binding and to isolate pathway associations of the ligand binding event given complex second messenger responses.[23] This article addresses the first objective, that is to discriminate small molecule ligands for oGPCRs.

1.2. Contribution of Current Research. Experimental ligand identification strategies have been based upon "reverse pharmacology",[24] in which an oGPCR is cloned and expressed in a cell line and then transfected into tissue extract containing endogenous ligands presumed to bind the receptor with high affinity. Finally, biological and pharmaceutical activity and association of the ligands to pathological states is assessed.[25]

Previous investigators have proposed structure-based virtual screens for ligands, which can be categorized as ligand-based or receptor-based methods (reviewed in ref 26). The ligand-based methods extrapolate from properties of compounds ("pharmacophores") known to bind a target receptor, by searching databases for compounds with similar profiles. This approach does not apply to the present research; the premise here is that high-affinity ligands are unknown. The receptor-based methods use computational docking procedures to bind compounds from a ligand database to the binding site of the receptor of interest. This presupposes that the three-dimensional structure of the receptor is available. For GPCRs, such an approach has limited utility; integral membrane proteins continue to be difficult to crystallize, constraining the analysis to a small number of structurally known GPCRs.[27] GPCRs as a superfamily do not share any overall sequence homology (outside the seven membrane-spanning helices),[3] and in cases of low sequence identity to the nearest known GPCR, ligands for oGPCRs cannot be discovered on the basis of expressed sequence tag (EST) or genomic sequence database homology search.[28] Even where homology to a known GPCR may be assumed, the scoring functions used in computational docking simulations remain imprecise.[29]

We describe here an approach to virtually screen for ligands of orphan G-protein coupled receptors, using bioinformatics. This method is based on a machine learning approach recently introduced by the authors to estimate the binding free energy between a small-molecule ligand and a receptor protein.[30] A distinct advantage of this approach is the simplicity of requisite input data: proteins are described using only physicochemical properties of primary amino acid sequence, and ligand features are based on the two-dimensional connectivity between constituent atoms and atomic properties. In application, large numbers of chemical compounds may be screened against a particular oGPCR sequence, with a ranked list of putative high-affinity ligands generated automatically on output. This screening approach may be used to aid in the functional characterization of oGPCRs by identifying potential cognate ligands, thereby providing clues to direct the therapeutic regulation of important signaling pathways in the cell.

2. METHODS

2.1. Quantitative Receptor Pharmacology. GPCRs are important regulators of central nervous system function in health and disease.[31] Accordingly, in this investigation, a data set of known psychoactive drugs and their associated ligand binding affinities were used to create a discriminative statistical model of ligand−receptor interaction. This model was constructed using a machine learning technique known as *support vector regression*[32,33] which uses empirical examples to learn to approximate a functional mapping (more details are provided in section 2.2). The data examples used in this investigation were derived from the PDSP K_i Database, a public repository containing information on affinities between real or candidate drugs and GPCRs and other receptors found in the central nervous system.[34]

Ligand−receptor affinities used to generate this data set were estimated using a variety of experimental protocols, many of which are described in detail on the PDSP Web site. [The PDSP home page is http://pdsp.cwru.edu.] Data collected during binding assays can be compared across protocols and laboratories by expressing the results in terms of a normalized index of affinity (or, reciprocally, dissociation) for a given ligand−receptor complex. One such expression in common usage is given by the Cheng−Prusoff equation[35] for competitive radioligand binding, given by

$$K_i = IC_{50} * \left(1 + \frac{[L^*]}{K_d}\right) \quad (1)$$

where K_i is the equilibrium dissociation constant for the analyte of interest ([*L*]), IC_{50} is the concentration of ligand displacing 50% of the specific bound labeled ligand [L*], and K_d is the (inverse) affinity of the radioligand for the receptor. K_i represents the equilibrium concentration of unlabeled ligand that would bind half the receptor binding sites in the absence of radioligand or other competitors. A fundamental pharmacological characteristic of the receptor−drug complex, K_i may be used as the basis for evaluating different candidate drugs. Inference of biological activity for a single compound can be made based on the computed value

$$pK_i = -\ln(K_i) \quad (2)$$

To assign degree of bioactivity to pK_i, this investigation followed the convention listed in Table 1.[36] Values $pK_i > 7$ are generally taken to imply high binding affinity.

Alternatively, qualitative comparisons between elements of a group of compounds are possible by their rank-ordering in terms of binding affinity for a given receptor (e.g., see ref 37). The supposition is that the highest-affinity ligands are correlated with efficacy of pharmacological effect, either as agonists or antagonists. This is the approach taken in the present investigation, where we predict and rank the values

Table 1. Relationship between Negative Logarithm of the Dissociation Constant (pK_i) and Biological Activity[a]

pK_i	inferred activity
>7	highly active
6−7	active
5−6	weakly active
<5	inactive

[a] This scheme may be used to infer biological activity of a single ligand−receptor complex or to rank order a library of compounds bound to a receptor in experimental screening. Source: GPCRDB.[36]

of pK_i for a large number of druglike, small molecule ligands in the specific context of a set of orphan G protein-coupled receptors.

2.2. Support Vector Regression. The support vector machine (SVM) is a pattern recognition algorithm that may be used for regression estimation of a function f by[32,33]

$$f(\mathbf{x}) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b \qquad (3)$$

where $\mathbf{x} \in \mathbb{R}^d$ are observations, α_i^* and α_i are Lagrange multipliers of the constrained quadratic optimization problem, k is a kernel function measuring the similarity between its arguments, b is the intercept, and l is the number of example data pairs. Usually only a subset of the coefficients α_i^*, α_i are nonzero; the associated training observations $\mathbf{x}_i$ are called the *support vectors*, and their sparsity contributes to the efficient computation of the expansion in eq 3, while providing an analytic upper bound on the generalization error.[38]

In this research, support vector learning was chosen over conventional linear regression techniques for several reasons:

1. A single global optimum is found in training an SVM. Learning is accomplished by solving a quadratic programming problem, for which robust numerical packages exist. This provides for computational tractability on massive data sets,[39] a key advantage for large-scale virtual screening.

2. Using SVM, overfitting is controlled according to structural risk minimization principle.[32] In contrast, ordinary least squares (OLS) is meant to provide the "best fit" to a line relating explanatory variables $\mathbf{x}_i$ to the response variable y. The main problem with OLS is its tendency to overfit when the data are sparse (i.e., # of input attributes > > # observations).

3. SVM works by mapping the input data to a high-dimensional feature space, where a linear regression is performed. This enables an infinite-dimensional representation (using Gaussian kernels). SVM directly addresses the "curse of dimensionality"—as the number of explanatory (input) variables increases, convergence to a smooth estimator becomes exponentially slow in other methods.

4. While OLS works best in the presence of Gaussian noise, real data are corrupted by noise with generally non-Gaussian statistical distributions. Recent results indicate that in SVM application to finite-sample regression problems, only the noise *level* (i.e., variance) and not noise *density* is required for optimal generalization.[40] Noise variance can be estimated from the training data; therefore, SVM achieves good prediction performance without preconceived assumptions about noise density.

2.3. Virtual Screening Approach. In this investigation, the support vector regression algorithm is used to approximate the unknown function $f(\mathbf{x})$ which connects descriptors of known receptor−ligand pairs to their experimentally determined dissociation constants pK_i. This function is then evaluated using data patterns corresponding to uncharacterized oGPCR−ligand pairs, producing predicted values for pK_i. These predictions are sorted, producing a ranked list of chemical compounds most likely to bind to the orphan receptor. The direct prediction of binding energy in a support vector regression framework was recently introduced by the authors,[30] where the target quantity was computed binding free energy of ligand−receptor complexes sampled from a heterogeneous database. In ref 30, experimental results displayed error rates and rank correlation values on par with a number of alternative methods appearing in the computational chemistry literature. The dominant conclusion drawn was that no three-dimensional structural information on either the receptor or small molecule ligand was required to construct an accurate nonparametric regression function.

This observation stimulated our interest in extending the methodology to the prediction of oGPCR−ligand affinities, based on limited actual experimental binding data.

2.3.1. Preparation of Example Data. Descriptive Features. Each ligand−receptor complex was transformed into a vector of numerical descriptor arrays. Attributes distinguishing each complex were selected based upon their presumed salience for learning the target concept—specifically, the mapping from feature space to binding affinity. The physicochemical content and numerical construction of these feature vectors is reviewed in this section.

Target Receptors. Target features comprised numerical values for surface tension, isoelectric point, and accessible surface area attributed to each amino acid comprising the receptor primary structure. [Tables of residue physicochemical properties are widely accessible; one source of such data is The Amino Acid Repository at http://www.imb-jena.de/IMAGE_AA.html.] This scheme encodes physicochemical properties of the primary structure that are likely to influence the thermodynamics of binding.

Chemical Ligands. Ligand features were established using a two-dimensional molecular connectivity matrix to exemplify the arrangement of each compound's constituent atoms in space. For example, at row i and column j, a unit-valued entry is made if the corresponding atoms in the molecule are covalently connected; otherwise the value of that matrix element is zero.

Each ligand's 2-D molecular connection array was supplemented by additional arrays, containing numerical values for fundamental, measurable chemical properties characterizing the atoms comprising the molecule. These properties included the atomic ionization potential energy, the electron affinity, and the atomic density. The rationale followed again was to employ quantities relating to the physics of binding. Separate two-dimensional arrays representing these properties (along with connection topology) were concatenated into a single, wide matrix. The resulting aggregate data matrix was then factorized using the singular value decomposition (SVD).[41] The singular values computed in this factorization are extracted, representing a projection onto one-dimensional space of the essential characteristics of molecular bond

topology, and, it is hypothesized, the spatial distribution of molecular properties important for binding with a receptor. [Burden[42] introduced the idea of computing the eigenvalues of a hydrogen-suppressed molecular bond graph with an atomic number on the diagonal and numbers indicating bond presence and type at off diagonal positions. This matrix was used as a means to group substructures for chemical similarity search.[42]] The output of the SVD is a one-dimensional vector of numbers finally resampled to yield a fixed-length sequence representing each small-molecule compound.

Length-Normalization Procedure. Prior to training the support vector regression, both receptor and chemical compound features must be length-normalized by transformation onto fixed-length arrays. This is an essential step to consistently represent proteins and small molecules of widely varying native length and 2-D structure, respectively.

Next, this vector of floating point numbers is transformed (by interpolation or decimation) onto a fixed-length sequence, an essential step to maintain a consistent physical "meaning" for each transformed vector element across examples.

This section provides a mathematical description of the procedure followed during our numerical experiments.

Let the vector of numbers $\{\mathbf{v}\}^i$, $i \in 1,..., M$ in L-dimensional real space $\mathbb{R}^L$ denote feature i for a given 1-D numerical sequence of length L, where M different features are considered. Lengths of the individual feature vectors $\mathbf{v}$ were normalized by mapping onto a fixed-length interval K, via $\{\mathbf{u}_k\}^i = g(\{\mathbf{v}\}^i)$, where the function g is g: $\mathbb{R}^L \rightarrow \mathbb{R}^K$.

The mapping g is implemented using straightforward linear interpolation.[43] An outline of one strategy for doing this is as follows:

1. Discretize the input and output domains:

$$\xi_{\text{in}} = (1/L)^*\{1,..., L\},\ 0 \le \xi_{\text{in}} \le 1$$

$$\xi_{\text{out}} = (1/K)^*\{1,..., K\},\ 0 \le \xi_{\text{out}} \le 1$$

2. For each element of the output domain $\xi_{\text{out},k}$, find the indices $(j, j+1)$ of the input domain whose corresponding values $\xi_{\text{in},j}, \xi_{\text{in},j+1}$ "bracket" it:

$$\xi_{\text{in},j} \le \xi_{\text{out},k} \le \xi_{\text{in},j+1},\ j \in 1,..., L;\ k \in 1,..., K$$

3. Estimate the local slope m:

$$m \approx (v_{\text{in},j+1} - v_{\text{in},j})/(\xi_{\text{in},j+1} - \xi_{\text{in},j})$$

4. Estimate the value of $y_{\text{out},k}$ at $\xi_{\text{out},k}$ by linear interpolation:

$$u_{\text{out},k} = u_{\text{out},k-1} + \{m^*(\xi_{\text{out},k} - \xi_{\text{out},k-1})\}$$

Note that this procedure as summarized assumes that $K < L$ and should be appropriately modified for the case $K > L$.

In this transformed space, the arc length coordinate ξ_{out} along the sequence now varies as $\xi_{\text{out}} \in [0, 1]$, and each vector $u_{\text{out}} \in \mathbb{R}^K$. The full feature vector for a particular protein A (or small-molecule B) is constructed by concatenation of each feature sequence $\mathbf{u}$. This is written as $\{\varphi_{\text{A}}^+\} = \{\mathbf{u}_k\}^1 \oplus \{\mathbf{u}_k\}^2 \oplus \cdots \oplus \{\mathbf{u}_k\}^M$, where $\mathbf{c} \oplus \mathbf{d}$ indicates simple concatenation of vectors $\mathbf{c}$ and $\mathbf{d}$.

Target−Ligand Complexes. A representation of an interaction pair, $\{\varphi_{\text{AB}}^+\}$, is finally formed by concatenating the feature vectors for target A and small-molecule B, i.e., $\{\varphi_{\text{AB}}^+\} = \{\varphi_{\text{A}}^+\} \oplus \{\varphi_{\text{B}}^+\}$. The vector $\{\varphi_{\text{AB}}^+\}$, along with its associated value pK_{i}, becomes a training example for the SVM. For virtual screening, the label pK_{i} is unknown and is predicted by the regression function.

2.3.2. Example Databases. Training Source Database. To construct training examples, target−ligand complexes were selected from the PDSP K_{i} database introduced in section 2.1. From the nominal K_{i} database comprising over 26 000 records, a useable subset of 9075 complexes was identified, based on our ability to associate amino acid sequences with receptors, and SMILES strings[44] with their cognate ligands, respectively.

Statistical redundancy between training examples in any supervised learning situation may result in unreliable cross-validated estimates of generalization error. To address this issue, highly similar examples were excluded within the training data set according to the following procedure:

1. A similarity matrix $S \in \mathbb{R}^{lxl}$ was created for the $l = 9075$ ligand−target complexes found within PDSP. Each matrix element $s_{i,j}$ expresses the degree of similarity between example feature vectors numbered i and j. Values $s_{i,j}$ were evaluated using an heuristic criterion

$$s_{i,j} = \frac{1}{d}\sum_{k=1}^{d} H_{\text{B}}(|\mathbf{x}_{i,k} - \mathbf{x}_{j,k}| \le \sigma),\ 0 < s_{i,j} \le 1 \qquad (4)$$

where H_{B} is Heaviside's step function with Boolean argument, $\sigma \in \mathbb{R}^d$ is the standard deviation estimate for each attribute, and k denotes a feature. In essence, this equation counts the number of corresponding vector elements in $\mathbf{x}_i$ and $\mathbf{x}_j$ whose values differ by less than one standard error.

2. Redundant examples were removed, referring to the similarity matrix, using a two-pass algorithm designed for this application. The idea is to eliminate training examples based upon their composite *pattern* and *label* similarities.

(a) The first pass iterates over each row i of S, evaluating the similarity of training vector $\mathbf{x}_i$ to all other vectors $\{\mathbf{x}_j\}$, $j = 1,..., l$, $j \neq i$. Those examples where the similarity to $\mathbf{x}_i$ exceeds a numerical threshold criterion are marked for removal subject to subsequent passes of the algorithm. This investigation used a threshold value 0.98.

(b) For each data vector $\mathbf{x}_i$, the second pass compares its target value y_i to each value $\{y_j\}$, $j = 1,..., l$, $j \neq i$ associated with examples marked as "similar" in the previous pass. The target quantities to be learned by the regression represent binding affinity (pK_{i}), where pK_{i} between respective training instances differed by less than 0.25 logarithm units, the redundant example was excluded from further analysis.

This process removed 3756 redundant observations (41%), leaving a total of 5319 examples for cross-validation training from the preredundancy processed set. The median target value pK_{i} in this set was $\mu^* = 6.32$, with extreme values ranging between −9.8 and +11.

Testing Source Database. Testing examples, forming the basis for the prediction of binding affinities for novel oGPCR complexes, were generated using (i) orphan G protein-coupled receptor sequences found within the Swiss-Prot Protein Knowledgebase[45] and (ii) a "druglike" subset of compounds derived from the National Cancer Institute (NCI) open databases as provided within the Ligand.Info Small-

Molecule Databases[46] [downloadable at http://ligand.info/]. Druglikeness includes bioavailability and pharmacokinetic properties. It has been suggested that a future rational drug design process may filter nondruggable compounds before beginning biological receptor activity screening.[47] From the 69 045 druglike compounds stored in Ligand.Info, 34 753 were selected based on the availability of an unique CAS registry number or NSC accession ID.

The nominal list of orphan receptors contained 135 targets. [We used data found in file "7tmlist.txt" dated June 2, 2004. This list may be accessed at http://www.expasy.org/cgi-bin/lists?7tmrlist.txt.] Many of the orphan receptors represented nearly identical amino acid sequences from different organisms. We analyzed this set of sequences using global, multiple sequence alignment implemented in the program DBClustal,[48] with an *E*-value cutoff of 10^{-40}. This *E*-value was previously used to analyze evolutionary relationships within families of GPCRs.[49] The global alignment produced clusters of sequentially similar receptors; from each, a single archetypical receptor was selected. The resulting set of oGPCRs consisted of 55 targets, for which putative cognate ligands would be identified. These orphan receptors, including their cluster sizes, are summarized in Table 4.

We built feature vectors by connecting the 55 oGPCRs with the 34 753 druglike chemical compounds in our locally constructed database using the methods described above. The resulting set of feature vectors encoding hypothetical oGPCR−ligand complexes ($n = 1\,911\,415$) was processed using the trained support vector regression function of eq 3 to estimate values for their binding affinities.

Standardization of Examples. Attributes in the example databases were mean-corrected and standardized by considering all training and testing data vectors simultaneously as a single matrix of observations. Overall mean and sample standard deviation statistics were calculated for each column (feature) of this matrix; these in turn became normalizing factors that were applied to all data examples.

2.3.3. Model Selection and Validation. Model Selection. The PDSP-derived training examples described in section 2.3.1 were used to develop an optimal support vector regressor. A number of schemes have been proposed in the literature to systematically select support vector machine model parameters; e.g., see refs 50−52. The approach followed here searched a computational grid of parameters of the learning machine, identifying the best parameter set using 10-fold cross-validation. Let us denote target-compound affinity scores using the variable y to simplify notation or

$$y := \mathrm{p}K_\mathrm{i} \tag{5}$$

Each held-out data partition was evaluated by computing the normalized mean squared error (NMSE)

$$\mathrm{NMSE} = \frac{\sum_{i=1}^{l_\mathrm{p}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{l_\mathrm{p}} (y_i - \bar{y})^2} \tag{6}$$

where $\hat{y}_i$ is a predicted value for y_i, $\bar{y}$ is the true mean, and l_p is the number of compounds in the prediction set. Equation 6 shows that an observed value of NMSE = 1 corresponds to simply predicting the mean value of the dependent variable;[53] values less than 1 imply predictive value-added by a particular method. [NMSE is related to the coefficient of determination (R^2) by $R^2 \approx 1 - \mathrm{NMSE}$, suggesting that NMSE be interpreted as a *coefficient of nondetermination*—a measure of the percentage of variance in y that is not explained by the model.]

Table 2. Comparison of Statistics of Predicted Binding Scores

prediction set	μ_y^*	$\hat{y}$ range
test	5.62	[4.50, 8.26]
test(oos)	6.99	[6.19, 7.73]

Data in "test" are the NCI druglike ligand−oGPCR pairs summarized in section 3.1. "Test(oos)" are the out-of-sample complexes formed between olanzapine and the oGPCR targets. The predicted value of pK_i is $\hat{y}$, and its median value is denoted by μ_y^*.

The support vector regression model exhibiting the lowest overall NMSE was selected for the ensuing virtual screen. The optimal model used a Gaussian kernel with parameter values $C = 10$ (train error/margin tradeoff), $\gamma = 0.01$ (inverse kernel width), and $\nu = 0.5$ (solution sparsity). The interested reader should consult ref 33 or ref 38 for a detailed explanation of these hyperparameters and their influence on the support vector machine.

Model Cross-Validation. Average statistics over the ensemble of cross-validation folds may provide an error estimate of the generalization performance of the regression model.[54] The selected model produced a cross-validated prediction error $\mathrm{NMSE}_\mathrm{cv} = 0.57$. This corresponds to a cross-validated predictive $R^2_\mathrm{cv} = 0.43$. The standard and maximum deviation between actual and predicted pK_i were SD = 1.21 and MD = 8.12, respectively.

It is interesting to compare this cross-validated standard error against the method of Gohlke and co-workers,[55] who developed a knowledge-based scoring function ("DrugScore") to predict protein−ligand interactions. Input to DrugScore is 3-D information from X-ray crystallography or statistical mechanics-based computation of the ligand−receptor complex. The presently observed standard error SD = 1.21 log units is smaller than 10 of 14 (71%) of the values presented in ref 55 (see Table 2, p 130). The cross-validated R^2 observed here matched or exceeded 50% of the values in the cited investigation. Importantly, the largest sample size in all data sets analyzed in ref 55 was $n = 71$ complexes, and many of the results presented there represented structurally similar receptor families. In contrast, the present results apply to a 7-fold larger sample size ($n = 531$), and instances were analyzed to eliminate redundancy in the training set. Further, no attempt was made here to remove statistical outliers, which would improve the observed NMSE but reduce the robustness of the model upon generalization.

These observations suggest that the predictive ability of the current approach is competitive with published methods that are based on three-dimensional structural information.

Model Statistical Significance. We examined the question of whether the predictive model evaluated using the R^2_cv statistic could have been obtained by chance. Because this is a virtual screening method, it is important to investigate the discriminative utility of the SVM regression (eq 3),

Table 3. Out-of-Sample Predictions for Known Nicotinic Ligands against a Nicotinic ACh Receptor α4 Subunit Sequence[a]

compound	CAS no.	Δ*y* expt	$\hat{y}$	$y_l < \hat{y} < y_h$?
epibatidine	140111-52-0	[7.07, 11]	5.96	-
ABT-594	198283-73-7	5,10	6.58	√
cytisine	485-35-8	[6.29, 9.9]	6.58	√
nicotine	54-11-5	[7.58, 9.07]	6.11	-
ABT-418	147402-53-7	[7.35, 8.77]	6.02	-
acetylcholine	51-84-3	[7.2, 8.57]	5.95	-
MCC	1918-18-9	[3, 8.42]	6.03	√
lobeline	90-69-7	[7.3, 8.4]	7.00	√*
DMPP	54-77-3	[4.75, 7.97]	7.21	√
DHβE	23255-4-1	[4.22, 7.9]	7.06	√
suberyldicholine	3810-71-8	[5.95, 7.88]	6.28	√
anabasine	494-52-0	[6.6, 7.42]	5.79	-
carbachol	51-83-2	[6.2, 7.16]	5.86	-
mecamylamine	60-40-2	[3, 6.25]	6.12	√
d-tubocurarine	57-95-4	[4.4, 6]	6.77	-
methyllycaconitine	21019-30-7	[5.21, 5.79]	6.98	-
choline	62-49-7	[5, 5.15]	5.66	√ *

[a] The right-most column contains a " √" where the prediction lies within the experimental standard error. Asterisks indicate "hits" in a biological functional sense: (1) lobeline, a potent agonist, is correctly predicted to bind with high activity; (2) choline is a selective agonist of α7 nAChRs (not α4).

constructed with this particular set of molecular descriptors, as compared to a model built from a randomly generated background feature set. The procedure followed here was based upon methods described previously by Ekins and colleagues.[56] The idea is to "shuffle" the structure−activity relationship by random permutation of the binding affinities (eq 2) associated with each ligand−target complex. This creates a randomized data set to use as input for SVM cross-validation training and evaluation. The supposition is that if the optimized SVM regression is not a chance occurrence, the randomized models will manifest a relative substantial degradation in binding energy predictive performance as represented by the test statistic (R^2_{cv}).

For this experiment, the same cross-validation training protocol was applied to the shuffled data set as described above for model selection. Feature vector dimensionality and statistical content were held constant; only the labels were randomly switched. The observed value of the randomized test statistic, averaged over 10 data partitions, was $R^2_{cv,rand} = -0.158$. This is significantly less than was observed using the nonshuffled training data set. In fact, the negative sign indicates that the randomized models trained in this manner had a less predictive value than a trivial guess of the mean binding strength observed on the entire training set.

This internal validation experiment supports the conclusion that the observed value of the test statistic using the nonshuffled data is unlikely to be encountered by chance.

Out-of-Sample Check: Multiple Compounds, One Nicotinic Target. To further explore the generalization potential of the virtual screening method, a spot check was performed by extracting nicotinic targets and their associated ligands from the PDSP data set and predicting the binding strength of certain known substances in the context of a nicotinic receptor. One of the basic premises underlying our approach is the ability to extrapolate beyond the training data set to make novel predictions of ligand binding within different target classes. Neuronal nicotinic acetylcholine receptors (nAChRs), important in a number of central nervous system (CNS)-related functions and diseases,[57] provide an excellent target class for this purpose. Nicotinic receptors consist of five polypeptide subunits surrounding a central ligand-gated ion channel. In the CNS, two α and three β subtypes are observed, and in particular the α4β2 nAChR predominates in high affinity agonist binding. Combinations of different subtypes may facilitate the development of ligands with subtype specificity. Like oGPCRs, nAChRs are interesting targets because their physiological functions are still largely not known.

Selected out-of-sample predictions for known nicotinic ligands against a nicotinic ACh receptor α4 subunit sequence are presented in Table 3. The target receptor is found in the peach-potato aphid *Myzus persicae* (UniProtKB/TrEMBL #Q9U940) but is highly conserved across species. The results shown in Table 3 are anecdotal due to the limited sample available, but certain trends are observable.

The range of experimental binding affinities appear in the column Δ*y* expt; corresponding predictions are listed under $\hat{y}$. Compounds are listed in descending order of highest experimental binding affinity in the range, against targets representing α4β2 receptor subtypes. The column heading $y_l < \hat{y} < y_h$? contains a checkmark where the predicted binding lies within the experimental range.

In absolute terms, certain nicotinic agonists of highest activity (epibatidine, nicotine, ABT-418) are not accurately predicted by the method. Epibatidine is a potent analgesic which binds α4β2 receptors with high affinity.[58] Nicotine and ABT-418 are bioisosteric analogues; it is possible that common structural properties might contribute to their low accuracy binding prediction.

We observe that two known high-affinity binders of α4β2 nAChRs are correctly predicted within the experimental range. These include the analgesic ABT-594[59] and the nicotine-like alkaloid cytisine.[60]

The binding predictions are better for MCC, lobeline, DMPP, DHβE, and suberyldicholine. These compounds are associated with the low to high "active" range ($7 > \mu_y > 6$). The alkaloid lobeline is considered correct in the biological functional sense as (indicated by the asterisk), since the model predicts it to bind with "high activity" according to Table 1.

The last row in the table presents results for choline, a selective agonist of α7 nAChRs; other subtypes, such as α4, are not activated by this substance. We suggest this is a functionally correct prediction, as only a single experimental data point was available, and the selectivity of choline in the α4 subunit context is captured by the model.

While these results are encouraging, they are not as strong as anticipated given the cross-validated performance of the SVM regressor. This may be explainable due to the limited nAChR sample data for validation and/or because of the large experimental error bars. We note that experimental variance for certain compounds may reflect experiments wherein the agonist interaction is measured during the activated state, which occurs with low affinity. We considered another possible explanation for the observed lack of sensitivity for the most potent nicotinic ligands. Our spot check was performed using only the α4 subunit amino acid sequence; the ligand binding pocket in human neuronal nAChRs is know to lie at the interface between α4 and β2 subunits.[61] We tried constructing a model including both subunits;

Table 4. Orphan G Protein-Coupled Receptors Used in the Virtual Screen[a]

no.	Swiss-Prot name	Swiss-Prot accession	description	species	cluster size
1	MAS_HUMAN	P04201	mas proto-oncogene	*H. sapiens*	3
2	MRS_HUMAN	P35410	mas-related MRS (MAS−R)	*H. sapiens*	4
3	CML1_HUMAN	Q99788	chemokine receptorlike 1	*H. sapiens*	3
4	CML2_HUMAN	Q99527	chemokine receptorlike 2	*H. sapiens*	2
5	EBI2_HUMAN	P32249	EBV-induced GPCR 2	*H. sapiens*	2
6	ETB2_HUMAN	O60883	endothelin B receptorlike	*H. sapiens*	2
7	H963_HUMAN	O14626	probable GPCR	*H. sapiens*	6
8	LGR4_HUMAN	Q9BXB1	leucine-rich GPCR 4	*H. sapiens*	8
9	RDC1_HUMAN	P25106	GPCR RDC1 homolog	*H. sapiens*	4
10	GP61_HUMAN	Q9BZJ8	probable GPCR	*H. sapiens*	2
11	GPR1_HUMAN	P46091	probable GPCR	*H. sapiens*	7
12	GPR3_HUMAN	P46089	probable GPCR	*H. sapiens*	7
13	GPR4_HUMAN	P46093	probable GPCR	*H. sapiens*	2
14	GP10_HUMAN	P49683	probable GPCR	*H. sapiens*	5
15	GP15_HUMAN	P49685	probable GPCR	*H. sapiens*	7
16	GP18_HUMAN	Q14330	probable GPCR	*H. sapiens*	1
17	GP19_HUMAN	Q15760	probable GPCR	*H. sapiens*	3
18	GP20_HUMAN	Q99678	probable GPCR	*H. sapiens*	1
19	GP21_HUMAN	Q99679	probable GPCR	*H. sapiens*	1
20	GP22_HUMAN	Q99680	probable GPCR	*H. sapiens*	1
21	GP26_HUMAN	Q8NDV2	probable GPCR	*H. sapiens*	4
22	GP27_HUMAN	Q9NS67	probable GPCR	*H. sapiens*	10
23	GP31_HUMAN	O00270	probable GPCR	*H. sapiens*	4
24	GP32_HUMAN	O75388	probable GPCR	*H. sapiens*	1
25	GP33_MOUSE	O88416	probable GPCR	*M. musculus*	2
26	GP34_HUMAN	Q9UPC5	probable GPCR	*H. sapiens*	3
27	GP35_HUMAN	Q9HC97	probable GPCR	*H. sapiens*	2
28	GP39_HUMAN	O43194	probable GPCR	*H. sapiens*	1
29	GP40_HUMAN	O14842	probable GPCR	*H. sapiens*	1
30	GP41_HUMAN	O14843	probable GPCR	*H. sapiens*	3
31	GP45_HUMAN	Q9Y5Y3	probable GPCR	*H. sapiens*	4
32	GP52_HUMAN	Q9Y2T5	probable GPCR	*H. sapiens*	1
33	GP57_HUMAN	Q9P1P4	probable GPCR	*H. sapiens*	2
34	GP62_HUMAN	Q9BZJ7	probable GPCR	*H. sapiens*	1
35	GP80_HUMAN	Q96P68	probable GPCR	*H. sapiens*	3
36	GP82_HUMAN	Q96P67	probable GPCR	*H. sapiens*	1
37	GP92_HUMAN	Q9H1C0	probable GPCR	*H. sapiens*	1
38	G101_HUMAN	Q96P66	probable GPCR	*H. sapiens*	2
39	G151_HUMAN	Q8TDV0	probable GPCR	*H. sapiens*	3
40	G152_HUMAN	Q8TDT2	probable GPCR	*H. sapiens*	2
41	G160_HUMAN	Q9UJ42	probable GPCR	*H. sapiens*	1
42	G161_HUMAN	Q8N6U8	probable GPCR	*H. sapiens*	1
43	GRE1_BALAM	Q93126	probable GPCR	*B. amphitrite*	3
44	YWO1_CAEEL	Q10904	probable GPCR	*C. elegans*	1
45	YWO4_CAEEL	Q10907	probable GPCR	*C. elegans*	1
46	YS96_CAEEL	Q09965	putative GPCR	*C. elegans*	2
47	YS97_CAEEL	Q09966	putative GPCR	*C. elegans*	1
48	YT66_CAEEL	Q11082	probable GPCR	*C. elegans*	1
49	YKR5_CAEEL	P34311	probable GPCR	*C. elegans*	2
50	YLD1_CAEEL	Q03566	probable GPCR	*C. elegans*	1
51	YYI3_CAEEL	Q18775	probable GPCR	*C. elegans*	1
52	YYO1_CAEEL	Q18904	probable GPCR	*C. elegans*	1
53	YMJC_CAEEL	P34488	putative gpcr	*C. elegans*	1
54	YR13_CAEEL	Q09638	probable GPCR	*C. elegans*	1
55	YN84_CAEEL	Q03613	probable GPCR	*C. elegans*	1

[a] The objective is to find ligands which bind strongly to these receptors, without knowledge of receptor structure in three-dimensional space. oGPCRs taken from the file 7tmrlist.txt dated 2-Jun-2004.

however, the results were not significantly improved for the compounds listed near the top of Table 3.

In the next step of our virtual screening approach, the discriminative model is applied to the task of screening druglike compounds against oGPCRs to find high-affinity binders.

3. RESULTS AND DISCUSSION

This section presents the results of the virtual screen for ligands of orphan G protein-coupled receptors. First, statistics of the calculated binding scores are summarized. Next, we provide some results obtained by analyzing the cross-target binding propensity of certain ligands. Finally, results representing the top-binding compounds found for individual oGPCRs are discussed.

3.1. Statistics of oGPCR Binding Scores. The experimental results represent predicted pK_i values for $n = 1\,911\,415$ ligand−oGPCR pairs $\mathbf{x}$, output from the trained regression function $f(\mathbf{x})$ (cf. eq 3). The overall distribution of predicted scores had the median value $\mu_y^* = 5.62$, with range $\hat{y} \in [4.50, 8.26]$. Each observation $\hat{y}$ represents an estimate of pK_i corresponding to a novel oGPCR-small

molecule ligand pair in the virtual screen. Visual inspection of the overall histogram of predicted binding scores suggested a non-normal distribution. Normality of the distribution empirical scores was tested using the Cramer-Smirnov-Von-Mises statistic[62]

$$W^2 = \int_0^1 \{S_n(y) - F(y)\}^2 \mathrm{d}F(y) \qquad (7)$$

where $S_n(y)$ is the empirical cumulative distribution function (cdf), $F(y)$ is a theoretical Gaussian cdf, and n is the total sample size. As a result of this test, the null hypothesis [H_0: $S_n(y)$ does not differ from $F(y)$ at significance level $\alpha = 0.05$] was rejected. The observed absence of normality is attributed to a large clustering of the empirical scores around the median value, causing a discontinuous increase in the cdf at this pK_i-value.

The central tendency of the population of predicted affinity scores corresponds to a "weakly active" binding affinity according to the calibration protocol of Table 1. Our interest lies within the "highly active" region $\hat{y} > 7$; predicted affinity scores lying in this region constituted 1.96% (37 407/1 911 415) of all results in our numerical experiments. The methodology therefore screened out 98% of the putative oGPCR−ligand complexes.

A total of 4357 different compounds were represented within the set of high-affinity ligands. This translates to about 12% of the complete set of druglike compounds comprising the virtual screen.

3.2. Cross-Target Analysis of High-Affinity Ligands. Many of the druglike ligands were predicted to bind strongly to more than a single target. We performed cross-target analysis by calculating the average binding score for each ligand across the set of target receptors.

To assign presumptive biological activities to the compounds observed to have the highest average cross-target affinity, we considered the functional annotations found within the online NCI chemical structure database [located at: http://cactus.cit.nih.gov/ncidb2]. The biological activities assigned to structures in the database represent independent predictions output from the program PASS, which computes probabilities based on structure−activity relationships.[63]

An approximate estimate of bioactivity is made as follows. For a given compound, the log-odds ratio that it is associated with biological function F is

$$\mathrm{LOR} = \ln\left(\frac{p(F)}{p(\sim F)}\right) \qquad (8)$$

where $p(F)$ is the probability that F is present, and $p(\sim F)$ is the probability it is not. LOR values indicate the probability that the quantity in the numerator evaluates to "true"; confidence that a given activity is linked with a particular structure increases with the magnitude of positive-valued LOR. For example, a high degree of confidence of bioactivity would be suggested where LOR > 3.0, which can be interpreted as providing greater than 20:1 odds of observing the function (relative to its absence) under the assumptions underlying its prediction by PASS.[63]

A listing of 13 compounds with the strongest generalized (cross-target) binding affinity is presented in Table 5. The table includes values for average predicted pK_i ($\hat{y}$) and possible biological activities where the value of LOR is at least 3.0. The column heading GPCR link? contains a check mark where the experimental literature describe GPCR-related pathways or processes that may be modulated by the compound in question. This modulation might involve activation or inhibition of consequent biological events after binding at the membrane-bound receptor. The authors acknowledge that these functional linkages may be indirect ones; this follows from their role as intermediaries in intracellular signaling circuits. Nonetheless, the literature references provide additional evidence supporting the plausibility of the highly cross-reactive ligands found in Table 5, beyond the computed log-odds ratio that correlates their structure with a particular biological function. A brief perusal of the NCI database should convince the reader that these are highly specific predictions, given the range of possible activities.

Analysis of structural characteristics of these multiple-target binding compounds may eventually provide insight into recurring motifs or pharmacophores correlated with patterns of ligand−receptor affinity. Such information might aid in the design of bioactive compounds for families of receptors based on so-called *molecular fingerprints*,[64] a principal motivation for this study in the long term. Another possible benefit of structural pattern recognition would be to promote the development of combinatorial libraries for lead discovery.[65]

3.3. Out-of-Sample Check: One Compound, Multiple oGPCR Targets. We sought to identify a pharmaceutical agent in widespread commercial use to provide a qualitative check on the consistency of the binding predictions. The antipsychotic drug olanzapine, used in the treatment of schizophrenia [generic name Zyprexa; see: http://pi.lilly.com/us/zyprexa-pi.pdf], was deemed suitable for this purpose because (1) it is not found in the NCI 2D structure repository, thereby representing an out-of-sample data point for prediction, and (2) it is known to promiscuously bind a number of different G protein-coupled receptors with nanomolar affinity.[66] The rationale followed here was that if high-binding affinity between olanzapine and a number of oGPCRs was predicted by the trained SVM, confidence in the generalization potential of the model would increase.

Out-of-sample feature vectors were assembled by conjoining olanzapine with the oGPCR targets in the same manner as described in section 2.3.1, resulting in one feature vector per orphan receptor. These data were virtually screened using the optimal SVM model found during the process detailed in section 2.3.3. The results of this screen are compared to the overall oGPCR predicted binding scores in Table 2. This table shows that the median score observed for olanzapine ($\mu_y^* = 6.99$) lies on the threshold between the "active" and "highly active" activity ranges summarized in Table 1; this score is substantially greater than the observed median for the complete sample of test compounds. In fact, all of the 55 oGPCRs were observed to bind olanzapine strongly (53% "active", 47% "highly active"). The results of the method on this out-of-sample data point are therefore consistent with documented findings of cross-target affinity of this compound for GPCRs.[66]

For reference, the two-dimensional structure of olanzapine is presented in Figure 1.

Table 5. Compounds with High Cross-Target Affinity[a]

CAS no.	ŷ (av)	putative activities	LOR	GPCR link?	ref
24116-23-2	7.59	vasodilator	4.07	√	75
		MAO-B inhibitor	3.58	√	76
		ACh release stimulant	3.49	√	77
		prolactin inhibitor	3.47	√	78
		rhinitis treatment	3.45	√	11
		Ca^{2+} channel antagonist	3.23		
		mediator release inhibitor	3.18		
		antihistamic	3.11	√	79
		antianginal	3.05		
81382-09-4	7.52	antineoplastic	4.60	√	7
		antineoplastic antibiotic	3.02		
40323-42-0	7.46	cardiovascular analeptic	3.10	√	80
17304-96-0	7.45	arrhythmogenic	4.88	√	81
		cardiotonic	4.82		
		analeptic	4.80	√	80
		respiratory analeptic	4.56	√	80
		sodium channel blocker	4.55		
		cardiotoxic	4.46	√	82
		hypertensive	3.89	√	83
		aldosterone antagonist	3.75		
		spasmogenic	3.35		
		diuretic	3.13	√	84
24996-74-5	7.44	squalene epoxidase inhibitor	4.32		
		CNS active muscle relaxant	3.85	√	85
		urokinase inhibitor	3.84		
		sedative	3.79	√	86
		hypertensive	3.41	√	83
		skeletal muscle relaxant	3.12	√	87
		anticonvulsant	3.08	√	88
		muscle relaxant	3.05	√	85,87
		benzodiazepine antagonist	3.02	√	89
35956-47-9	7.44	cholinergic agonist	5.11	√	90
		ACh agonist	4.90	√	91
		ACh muscarinic agonist	4.56	√	91
		ACh M1 receptor agonist	4.00		
		sedative	3.79	√	86
		ACh antagonist	3.65	√	92
		squalene epoxidase inhibitor	3.48		
		spasmolytic, papaverin-like	3.44	√	93
		cystic fibrosis treatment	3.31	√	94
		ACh muscarinic antagonist	3.09	√	95
		arrhythmogenic	3.04	√	81
15093-31-9	7.44	chemopreventive	4.75		
		chemoprotective	3.71	√	96
63362-26-5	7.44	cardiotonic	4.76		
		bronchodilator	4.58	√	97
5408-02-6	7.43	prostaglandin antagonist	4.60	√	98
		spasmolytic	3.11	√	93
		ACh release stimulant	3.06	√	77
		ACh muscarinic antagonist	3.01	√	95
79005-55-3	7.43	insulin promoter	3.34		
		cognition disorders treatment	3.18	√	99
35878-52-5	7.42	cytostatic	4.62	√	100
		antineoplastic	4.59	√	100
		immunosuppressant	3.69	√	101
15569-50-3	7.42	antiamebic	4.83		
		cytostatic	4.62	√	100
		cardiovascular analeptic	4.60	√	80
		expectorant	4.44		
		aldosterone antagonist	3.96		
		antitrichomonal	3.87		
		calcium regulator	3.57	√	102
		parathyroid hormone antagonist	3.51	√	103
		antipsoriatic	3.47	√	104
		dermatologic	3.40	√	104
		antihelminthic	3.02		
6630-45-1	7.41	aromatase inhibitor	4.62	√	105
		male reproductive dysfunction treatment	4.42		
		cannabinoid receptor agonist	4.22	√	106
		antineoplastic	3.93	√	7
		estrone sulfatase inhibitor	3.91		
		neurotrophic factor	3.69	√	107
		cardiovascular analeptic	3.63	√	80
		microtubule formation inhibitor	3.44		
		antimitotic	3.19		
		arrhythmogenic	3.13	√	81
		PDE IV inhibitor	3.04		

[a] CAS no. is the CAS registry identifier. $\hat{y}$ is the average predicted value of pK_i, taken over at least one receptor. Activities and log-odds ratios (LOR) are adapted from the NCI open database. These results suggest that the predicted cross-target binding ligands are plausible in the GPCR context, according to bioactivities attributed to these compounds by independent, structure-based calculations.

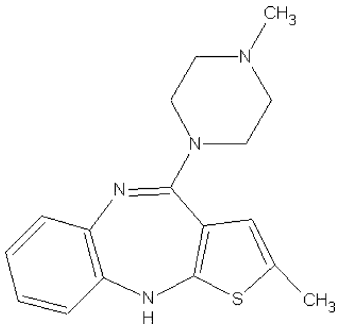

Figure 1. Structure of the antipsychotic olanzapine. Used in the treatment of schizophrenia, this compound binds several GPCRs with nanomolar affinity.[66]

3.4. Top Binding Compounds for oGPCRs. The main results of this research are summarized in Tables 6 and 7, which present the highest-scoring ligands for oGPCRs produced by the virtual screen. Target receptors are identified by number and Swiss-Prot accession, to provide cross-reference to their definition in Table 4. The columns marked "$\#y > 7$" list the number of binding affinity scores predicted to be "highly active" for the corresponding receptor. Parenthetically, this number is shown as a fraction of all scores computed for the corresponding receptor sequence. We have chosen to present only the top three scoring compounds for each oGPCR (shown in column "CAS no."), due to space constraints.

It is readily seen that this methodology is selective, filtering out all but a very small percentage of the ligand−target complexes presented to the support vector machine. For all orphan G protein-coupled receptors considered, the number of high-scoring virtual "hits" varies from 0 (23.6% or 13/55 cases studied) to 3958 (receptor #16; Swiss-Prot protein *Q*14330[67]).

Notice that a large majority of the top-binding ligands for all of the orphan receptor targets include one or both compounds with CAS registry numbers 24116-32-2 or 81382-09-4. These two ligands were identified in Table 5 as being the most highly cross-reactive, and this is reflected in their frequent appearance in Tables 6 and 7. The first small molecule compound, CAS #24116-32-2, is known by the chemical name 2-cyanoethyl 3-(1-aziridyinyl)propanoate, but

Table 6. oGPCRs and Predicted High-Affinity Ligands[a]

no.	Swiss-Prot accession	$\#\hat{y} > 7$ (%)	CAS no.	$\hat{y}$	no.	Swiss-Prot accession	$\#\hat{y} > 7$ (%)	CAS no.	$\hat{y}$
1	P04201	712	24116-23-2	7.68	16	Q14330	3958	81382-09-4	8.15
		(2.05)	81382-09-4	7.63			(11.39)	24116-23-2	8.11
			727-81-1	7.58				35956-47-9	8.08
2	P35410	2476	24116-23-2	8.26	17	Q15760	138	81382-09-4	7.29
		(7.12)	81382-09-4	8.12			(0.40)	40323-42-0	7.28
			35956-47-9	8.07				24116-23-2	7.26
3	Q99788	494	24116-23-2	7.71	18	Q99678	2	24116-23-2	7.14
		(1.42)	81382-09-4	7.52			(0.01)	81382-09-4	7.03
			57718-77-1	7.47					
4	Q99527	121	24116-23-2	7.52	19	Q99679	1270	24116-23-2	7.86
		(0.35)	63362-26-5	7.30			(3.65)	81382-09-4	7.78
			24996-74-5	7.29				15093-31-9	7.75
5	P32249	0			20	Q99680	88	24116-23-2	7.38
							(0.25)	81382-09-4	7.30
								17304-95-9	7.26
6	O60883	1240	81382-09-4	7.65	21	Q8NDV2	0		
		(3.57)	24116-23-2	7.63					
			6630-44-0	7.59					
7	O14626	17	15093-31-9	7.13	22	Q9NS67	7	24116-23-2	7.11
		(0.05)	81382-09-4	7.12			(0.02)	81382-09-4	7.11
			35956-47-9	7.12				40323-42-0	7.05
8	Q9BXB1	169	24116-23-2	7.32	23	O00270	0		
		(0.49)	40323-42-0	7.27					
			81382-09-4	7.26					
9	P25106	420	24116-23-2	7.53	24	O75388	0		
		(1.21)	81382-09-4	7.49					
			79005-55-3	7.45					
10	Q9BZJ8	365	81382-09-4	7.37	25	O88416	7	81382-09-4	7.13
		(1.05)	40323-42-0	7.32			(0.02)	24116-23-2	7.11
			70492-71-6	7.31				40323-42-0	7.07
11	P46091	1795	24116-23-2	7.80	26	Q9UPC5	2265	24116-23-2	8.22
		(5.16)	17304-96-0	7.72			(6.52)	81382-09-4	8.17
			17304-95-9	7.72				5408-02-6	8.04
12	P46089	3675	24116-23-2	8.38	27	Q9HC97	1060	24116-23-2	7.70
		(10.57)	81382-09-4	8.28			(3.05)	81382-09-4	7.60
			40323-42-0	8.20				40323-42-0	7.57
13	P46093	285	24116-23-2	7.47	28	O43194	214	24116-23-2	7.47
		(0.82)	6630-44-0	7.35			(0.61)	81382-09-4	7.33
			6630-45-1	7.35				24996-74-5	7.31
14	P49683	58	24116-23-2	7.46	29	O14842	850	24116-23-2	7.73
		(0.17)	81382-09-4	7.31			(2.44)	81382-09-4	7.66
			35956-47-9	7.25				40323-42-0	7.58
15	P49685	0			30	O14843	0		

[a] Targets are identified by number and Swiss-Prot accession, providing cross-reference to Table 4. Columns marked "$\#\hat{y} > 7$" list the number of binding scores found "highly active" for the corresponding receptor.

Table 7. oGPCRs and Predicted High-Affinity Ligands[a]

no.	Swiss-Prot accession	#$\hat{y}$ > 7 (%)	CAS no.	$\hat{y}$	no.	Swiss-Prot accession	#$\hat{y}$ > 7 (%)	CAS no.	$\hat{y}$
31	Q9Y5Y3	3011	727-81-1	7.94	42	Q8N6U8	0		
		(8.66)	81382-09-4	7.91	43	Q93126	1415	24116-23-2	7.83
			6630-45-1	7.91			(4.07)	81382-09-4	7.75
32	Q9Y2T5	716	24116-23-2	7.65				40323-42-0	7.69
		(2.06)	81382-09-4	7.64	44	Q10904	11	24116-23-2	7.22
			24996-74-5	7.55			(0.03)	5408-02-6	7.17
33	Q9P1P4	1325	24116-23-2	7.76				24996-74-5	7.14
		(3.81)	81382-09-4	7.71	45	Q10907	0		
			40323-42-0	7.69	46	Q09965	38	727-81-1	7.13
34	Q9BZJ7	282	81382-09-4	7.43			(0.11)	24116-23-2	7.11
		(0.81)	35956-47-9	7.42				17304-96-0	7.09
			24116-23-2	7.41	47	Q09966	117	57718-77-1	7.20
35	Q96P68	817	24116-23-2	7.72			(0.34)	24116-23-2	7.19
		(2.35)	81382-09-4	7.57				35878-52-5	7.19
			17304-96-0	7.53	48	Q11082	0		
36	Q96P67	22	24116-23-2	7.15	49	P34311	53	24116-23-2	7.36
		(0.06)	6630-44-0	7.15			(0.15)	81382-09-4	7.24
			6630-45-1	7.15				35878-52-5	7.16
37	Q9H1C0	15	81382-09-4	7.15	50	Q03566	0		
		(0.04)	24116-23-2	7.11	51	Q18775	635	40323-42-0	7.54
			35956-47-9	7.07			(1.83)	81382-09-4	7.52
38	Q96P66	437	24116-23-2	7.44				24116-23-2	7.48
		(1.26)	81382-09-4	7.41	52	Q18904	0		
			6630-45-1	7.36	53	P34488	2217	24116-23-2	8.05
39	Q8TDV0	0					(6.38)	81382-09-4	8.01
40	Q8TDT2	1205	24116-23-2	7.87				40323-42-0	7.91
		(3.47)	63362-26-5	7.69	54	Q09638	2546	24116-23-2	8.16
			81382-09-4	7.60			(7.33)	81382-09-4	8.09
41	Q9UJ42	872	81382-09-4	7.63				40323-42-0	8.08
		(2.51)	24116-23-2	7.61	55	Q03613	0		
			35956-47-9	7.56					

[a] Targets are identified by number and Swiss-Prot accession, providing cross-reference to Table 4. Columns marked "#$\hat{y}$ > 7" list the number of binding scores found "highly active" for the corresponding receptor.

little information on its pharmaceutical applications is available in public databases (outside of the NCI database). The second compound, CAS #81382-09-4, is a relatively large (mol. wt. 564.6 g/mol), DNA-binding antibiotic and appears to have strong antitumor properties. This compound is known commonly as "saframycin A". Many more strong-affinity ligands were predicted for over 75% of the oGPCRs; their exact numbers can be found in the tables. The structural characteristics of these particular two chemical compounds which contributed to their near-omnipresent cross-target affinity are not clear at the present time. This is a topic for further research.

Although we have chosen to screen orphan receptors mainly from human tissue (cf. Table 4), a great many oGPCR sequences appear to be highly conserved across species. Our intention here was to cluster the target sequences to refine the analysis set such that a single representative from each sequence-based cluster was used. Where a small number of ligands are predicted to bind a particular, conserved target, it would be interesting to employ high-throughput experimental screening techniques and obtain empirical binding data of that target against a complete set of specific ligands as predicted here. The objective would be to ascertain the degree of biological relevance of the predictions. This might lead to understanding of mechanisms of mediation of important signaling pathways, under the hypothesis that conservation implies fundamental functional significance.

4. CONCLUSIONS

GPCRs are widely screened drug targets, due to their close association with disease-related signaling pathways and past record of therapeutic success. Extension of this success to other members of the GPCR superfamily, identified by genomic sequence, has been problematic. Orphan GPCRs (oGPCRs) bind unknown ligands that modulate their function. These ligands, if identified, would offer clues toward understanding (and perhaps ultimately controlling) the physiological function of a receptor. High-throughput screening of compounds against oGPCRs cannot proceed until the target crystal structures have been obtained experimentally; this is a notoriously challenging prerequisite as G protein-coupled receptors are membrane-bound.

In this article, we have presented a virtual screening methodology that circumvents the requirement for receptor three-dimensional structure determination and may be used to directly generate a ranked list of high-binding small molecule ligands for oGPCRs. Perhaps the most compelling advantage of this approach is the simplicity of the requisite input data: proteins are described using only physicochemical properties of primary amino acid sequence, and ligand features are based on the two-dimensional connectivity between constituent atoms and their chemical properties. This virtual screening approach may be used in support of the functional characterization of oGPCRs by identifying potential cognate ligands.

The method predicts ligand binding energy at a given receptor. Receptors bind any number of ligands "promiscuously"—this may in fact be an essential characteristic of all drug action, as receptors develop in evolution to bind endogenous peptides or molecules distinctly different than the man-made compound.[68] Other computational docking and scoring programs have been declared to be incon-

sistent, as each combination of docking and consensus scoring technique varies with a selected target and the physicochemistry of target−ligand interactions.[69] In contrast, the support vector machine approach described here is deterministic in the sense that the trained regression function will produce a consistent output for each ligand−target complex, without appealing to three-dimensional pose or difficult statistical mechanics calculations.

Our experimental screen comprised more than 1.9 million hypothetical oGPCR−ligand complexes, from which we observed that less than 2% of predicted affinity scores corresponded to "highly active" ligands against orphan receptors. This 2% set consisted of 4357 different compounds or about 12% of the complete set of druglike compounds in the virtual screen. In practice, different numerical thresholds or data scaling procedures might be applied to further reduce the set of putative oGCPR ligands under consideration.

Validation of the method was carried out in several ways.

1. We examined the question of whether the predictive model evaluated using the R^2_{cv} statistic could have been obtained by chance. The procedure followed here was to "shuffle" the structure−activity relationship by random permutation of the binding affinities associated with each ligand−target complex. This created a randomized data set to use as input for SVM cross-validation training and evaluation. Feature vector dimensionality and statistical content were held constant; only the labels were randomly switched. The observed value of the randomized test statistic was significantly less than was observed using the nonshuffled training data set. This internal validation experiment supports the conclusion that the observed value of the test statistic using the nonshuffled data is unlikely to have been encountered by chance.

2. Anecdotal predictions for nicotinic targets and known ligands were performed by holding these examples out of the training data set. The highest-binding ligands were not accurately detected by the model; however, a number of correct predictions were observed in the low-to-high "active" range. These results may be explainable due to the limited nAChR sample data for validation and/or because of the large experimental error bars.

3. Possible biological activities linked to the compounds which exhibited high cross-target binding affinity were analyzed using the log-odds of observing particular bioactivities as computed by a structure-based program. This information was in turn correlated with citations to the scientific literature where GPCR-related pathways or processes were found to comport with observance of each bioactivity in question. The results (summarized in Table 5) provide additional evidence to support the plausibility of the method and its predictions.

4. An out-of-sample consistency check using the commercial antipsychotic drug olanzapine produced "active" to "highly-active" predicted affinities for all oGPCRs in our study (see Table 2). This observation is consistent with documented findings of cross-target affinity of this compound for many different GPCRs.[66]

The determination of the biological relevance of drug-oGPCR binding events is a significant challenge. Both in vivo and in vitro studies must be carried out to determine biological function.[70] Given a ranked list of conjectured ligand−oGPCR complexes, the crux is to validate them by experimental ligand binding assays. Once this is done, bioactivity and ultimate association to cellular pathways and cascaded second messenger responses must be performed.[23] Achieving this objective has extraordinary implications for pharmaceutical therapies to modulate or short-circuit faulty or disease-related cellular signaling pathways. Along the way, the problem of specificity, where an activated G protein-coupled receptor has a different role in different pathways, will have to be addressed.[71]

The methodology described here is general and may be applied to other receptor types. Two potential applications of therapeutic importance include design of tyrosine kinase inhibitors[72] or nuclear receptors. In the latter, it may be possible to apply this method to design hormone analogues to bind defective receptors. One only requires access to the amino acid sequence of the modified receptor; the procedures reported here could be easily adapted to provide a sensitive means to investigate small variations in the properties of a ligand (which may be a peptide, for example).[73]

In addition to cell surface receptors, this approach is a generalized strategy for discovery of small molecules which may bind intracellular enzymes and involve protein−protein interactions. Small-molecule mediated inhibition of protein−protein interactions is considered to be the most difficult of these drug design objectives, in part owing to the discrepancy in physical size between small molecule and the targeted protein complex.[74] This approach may provide a means of addressing this problem.

ACKNOWLEDGMENT

This work was funded by a grant from the von Liebig Center for Entrepreneurism and Technology Advancement, University of California San Diego.

REFERENCES AND NOTES

(1) Rodbell, M. *Biosci. Rep.* **1995**, *15*, 117−133.
(2) Gilman, A. *Annu. Rev. Biochem.* **1987**, *56*, 615−649.
(3) Gether, U. *Endocr. Rev.* **2000**, *21*, 90−113.
(4) Hunter, T. *Cell* **2000**, *100*, 113−127.
(5) Farfel, Z.; Bourne, H.; Iiri, T. *New Engl. J. Med.* **1999**, *340*, 1012−1020.
(6) Gutkind, J. *Oncogene* **1998**, *17*, 1331−1342.
(7) Schwindinger, W.; Robishaw, J. *Oncogene* **2001**, *20*, 1653−1660.
(8) Menon, S.; Han, M.; Sakmar, T. *Physiol. Rev.* **2001**, *81*, 1659−1688.
(9) Spiegel, A. *J. Inheritable Metab. Dis.* **1997**, *20*, 113−121.
(10) Rocheville, M.; Lange, D.; Kumar, U.; Patel, S.; Patel, R.; Patel, Y. *Science* **2000**, *288*, 154−157.
(11) Johnson, E.; Druey, K. *J. Allergy Clin. Immunol.* **2002**, *109*, 592−602.
(12) Meij, J. *Mol. Cell. Biochem. J.* **1996**, *157*, 31−38.
(13) Auld, D.; Diller, D.; Ho, K.-K. *Drug Discovery Today* **2002**, *7*, 1206−1213.
(14) Muller, G. *Curr. Med. Chem.* **2000**, *7*, 861−888.
(15) Gasparini, F.; Kuhn, R.; Pin, J.-P. *Curr. Opin. Pharmacol.* **2002**, *2*, 43−49.
(16) Howard, A.; McAllister, G.; Feighner, S.; Liu, Q.; Nargund, R.; Van der Ploeg, L.; Patchett, A. *Trends Pharmacol. Sci.* **2001**, *22*, 132−140.
(17) Ma, P.; Zemmel, R. *Nat. Rev. Drug Discovery* **2002**, *1*, 571−572.
(18) Hamm, H. E. *J. Biol. Chem.* **1998**, *273*, 669−672.
(19) Ji, T.-H.; Grossmann, M.; Ji, I. *J. Biol. Chem.* **1998**, *273*, 17299−172302.
(20) Wise, A.; Gearing, K.; Rees, S. *Drug Discovery Today* **2002**, *7*, 235−246.
(21) Civelli, O.; Nothacker, H.; Saito, Y.; Wang, Z.; Lin, S.; Reinscheid, R. *Trends Neurosci.* **2001**, *24*, 230−237.
(22) Im, D.-S. *Jpn. J. Pharmacol.* **2002**, *90*, 101−106.
(23) Civelli, O. *FEBS Lett.* **1998**, *430*, 55−58.

(24) Libert, F.; Vassart, G.; Parmentier, M. *Curr. Opin. Cell Biol.* **1991**, *3*, 218−223.
(25) Wilson, S.; Bergsma, D.; Chambers, J.; Muir, A.; Fantom, K.; Ellis, C.; Murdock, P.; Herrity, N.; Stadel, J. *Br. J. Pharmacol.* **1998**, *125*, 1387−1392.
(26) Lyne, P. *Drug Discovery Today* **2002**, *7*, 1047−1055.
(27) Ballesteros, J.; Palczewski, K. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 561−574.
(28) Milligan, G. *Biochem. Soc. Trans.* **2002**, *30*, 789−793.
(29) Kubinyi, H. *Drug Discovery Today* **2002**, *7*, 503−504.
(30) Bock, J.; Gough, D. *Mol. Cell. Proteom.* **2002**, *1*, 904−910.
(31) Pangalos, M.; Davies, C.; Davies, C. *Understanding G Protein-Coupled Receptors & Their Role in the CNS;* Oxford University Press: Oxford, U.K., 2003.
(32) Vapnik, V. *The Nature of Statistical Learning Theory;* Springer-Verlag: Heidelberg, Germany, 1995.
(33) Schölkopf, B.; Bartlett, P.; Smola, A.; Williamson, R. Support vector regression with automatic accuracy control. In *Proceedings of the Eighth International Conference on Artificial Neural Networks*; Niklasson, L., Boden, M., Ziemke, T., Eds.; 1998.
(34) Roth, B.; Lopez, E.; Patel, S.; Kroeze, W. *The Neuroscientist* **2000**, *6*, 252−262.
(35) Cheng, Y.; Prusoff, W. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(36) Horn, F.; Bettler, E.; Oliveira, L.; Campagne, F.; Cohen, F.; Vriend, G. *Nucleic Acids Res.* **2003**, *31*, 294−297.
(37) Millan, M.; Brocco, M.; Rivet, J.-M.; Audinot, V.; Newman-Tancredi, A.; Maiofiss, L.; Queriaux, S.; Despaux, N.; Peglion, J.-L.; Dekeyne, A. *J. Pharmacol. Exp. Ther.* **2000**, *292*, 54−66.
(38) Schölkopf, B.; Smola, A.; Williamson, R.; Bartlett, P. *Neural Comput.* **2000**, *12*, 1083−1121.
(39) Mangasarian, O.; Musicant, D. *Massive Support Vector Regression*; Technical Report 99-02; Data Mining Institute, Computer Sciences Department, University of Wisconsin: Madison, WI, 1999.
(40) Cherkassky, V.; Ma, Y. *Neural Networks* **2004**, *17*, 113−126.
(41) Golub, G.; van Loan, C. *Matrix Computations*, 2nd ed.; Johns Hopkins University Press: Baltimore, MD, 1989.
(42) Burden, F. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 225−227.
(43) Kreysig, E. *Advanced Engineering Mathematics*, 5th ed.; Wiley & Sons: New York, 1983.
(44) Weininger, D. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(45) Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.-C.; Estreicher, A.; Gasteiger, E.; Martin, M.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. *Nucleic Acids Res.* **2003**, *31*, 365−370.
(46) von Grotthuss, M.; Pas, J.; Rychlewski, L. *Bioinformatics* **2003**, *19*, 1041−1042.
(47) Lipinski, C. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235−249.
(48) Thompson, J.; Plewniak, F.; Thierry, J.; Poch, O. *Nucleic Acids Res.* **2000**, *28*, 2919−2926.
(49) Graul, R.; Sadee, W. *AAPS Pharm. Sci.* **2001**, *3*, E12.
(50) Chapelle, O.; Vapnik, V. Model selection for support vector machines. In *Advances in Neural Information Processing Systems 12*; Solla, S., Leen, T., Muller, K., Eds.; MIT Press: Cambridge, MA, 2000.
(51) Keerthi, S.; Lin, C.-J. *Neural Comput.* **2003**, *15*, 1667−1689.
(52) Chalimourda, A.; Schölkopf, B.; Smola, A. *Neural Networks* **2004**, *17*, 127−141.
(53) Gershenfeld, N.; Weigend, A. The future of time series: Learning and understanding. In *Time series prediction: Forecasting the future and understanding the past*; Addison-Wesley: Reading, MA, 1993; Vol. XV.
(54) Martin, J.; Hirschberg, D. *Small sample statistics for classification error rates I: Error rate measurements*; Technical Report ICS-TR-96-21; Department of Information and Computer Science, University of California: Irvine, CA, 1996.
(55) Gohlke, H.; Hendlich, M.; Klebe, G. *Perspect. Drug Discovery Des.* **2000**, *20*, 115−144.
(56) Ekins, S.; Crumb, W.; Sarazan, R.; Wikel, J.; Wrighton, S. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427−434.
(57) Gotti, C.; Clementi, F. *Prog. Neurobiol.* **2004**, *74*, 363−396.
(58) Badio, B.; Daly, J. *Mol. Pharmacol.* **199**, *45*, 563−569.
(59) Donnelly-Roberts, D.; Puttfarcken, P.; Kuntzweiler, T.; Briggs, C.; Anderson, D.; Campbell, J.; Piattoni-Kaplan, M.; McKenna, D.; Wasicak, J.; Holladay, M.; Williams, M.; Arneric, S. *Pharmacol. Exptl. Therap.* **1998**, *285*, 777−786.
(60) Papke, R.; Heinemann, S. *Mol. Pharmacol.* **1994**, *45*, 142−149.
(61) Brejc, K.; van Dijk, W.; Klaassen, R.; Schuurmans, M.; van Der Oost, J.; Smit, A.; Sixma, T. *Nature* **2001**, *411*, 269−276.
(62) Eadie, W.; Drijard, D.; James, F.; Roos, M.; Sadoulet, B. *Statistical Methods in Experimental Physics;* North-Holland: Amsterdam, Netherlands, 1971.
(63) Poroikov, V.; Filimonov, D.; Ihlenfeldt, W.; Gloriozova, T.; Lagunin, A.; Borodina, Y.; Stepanchikova, A.; Nicklaus, M. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 228−236.
(64) Greenbaum, D.; Arnold, W.; Lu, F.; Hayrapetian, L.; Baruch, A.; Krumrine, J.; Toba, S.; Chehade, K.; Bromme, D.; Kuntz, I.; Bogyo, M. *Chem. Biol.* **2002**, *9*, 1085−1094.
(65) Klabunde, T.; Hessler, G. *Chembiochem* **2002**, *3*, 928-p44.
(66) Bymaster, F.; Nelson, D.; DeLapp, W.; Falcone, J.; Eckols, K.; Truex, L.; Foreman, M.; Lucaites, V.; Calligaro, D. *Schizophrenia Res.* **1999**, *37*, 107−122.
(67) Gantz, I.; Muraoka, A.; Yang, Y.; Samuelson, L.; Zimmerman, E.; Cook, H.; Yamada, T. *Genomics* **1997**, *42*, 462−4662.
(68) James, L.; Tawfik, D. *Protein Sci.* **2003**, *12*, 2183-2193.
(69) Bissantz, C.; Folkers, G.; Rognan, D. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(70) Gould-Rothberg, B. *Pharmacogenomic J.* **2001**, *1*, 48−58.
(71) Rebois, R.; Allend, B.; Hebert, T. *Drug Discovery Today: Targets* **2004**, *3*,.
(72) Woolfrey, J.; Weston, G. *Curr. Pharm. Des.* **2002**, *8*, 1527−1545.
(73) Habeck, M. *Drug Discovery Today* **2003**, *8*, 236−237.
(74) Gadek, T.; Nicholas, J. *Biochem. Pharmacol.* **2003**, *65*, 1−8.
(75) Eckhart, A.; Ozaki, T.; Tevaearai, H.; Rockman, H.; Koch, W. *Mol. Pharmacol.* **2002**, *61*, 749−758.
(76) Fanciullacci, M.; Alessandri, M.; Del Rosso, A. *Funct. Neurol.* **2000**, *15*, 171−181.
(77) Imperato, A.; Obinu, M.; Casu, M.; Mascia, M.; Carta, G.; Gessa, G. *Eur. J. Pharmacol.* **1996**, *302*, 21−26.
(78) Ben-Jonathan, N.; Hnasko, R. *Endocr. Rev* **2001**, *22*, 724−763.
(79) Bertaccini, G.; Coruzzi, G. *Dig. Dis. Sci.* **1995**, *40*, 2052−2063.
(80) Hinkle, P.; Pekary, A.; Senanayaki, S.; Sattin, A. *Brain Res.* **2002**, *935*, 59−64.
(81) Lefkowitz, R.; Rockman, H.; Koch, W. *Circulation* **2000**, *101*, 1634−1637.
(82) Taglialatela, M.; Timmerman, H.; Annunziato, L. *Trends Pharmacol. Sci.* **2000**, *21*, 52−56.
(83) Gros, R.; Chorazyczewski, J.; ahd JL Benovic, M. M.; Ferguson, S.; Feldman, R. *Hypertension* **2000**, *35*, 38−42.
(84) Brody, T.; Cravchik, A. *J. Cell Biol.* **2000**, *150*, F83−F88.
(85) Honda, M.; Nishida, T.; Ono, H. *Eur. J. Pharmacol.* **2003**, *458*, 91−99.
(86) Andriamampandry, C.; Taleb, O.; Viry, S.; Muller, C.; Humbert, J.; Gobaille, S.; Aunis, D.; Maitre, M. *FASEB J.* **2003**, *17*, 1691−1693.
(87) Tuba, Z.; Maho, S.; Vizi, E. *Curr. Med. Chem.* **2002**, *9*, 1507−1536.
(88) Ng, G. Y. et al. *Mol. Pharmacol.* **2001**, *59*, 144−152.
(89) Breese, G.; Knapp, D.; Overstreet, D. *Neuropsychopharmacology* **2004**, *29*, 470−482.
(90) Kanno, H.; Horikawa, Y.; Hodges, R.; Zoukhri, D.; Shatos, M.; Rios, J.; Dartt, D. *Am. J. Physiol. − Cell Physiol.* **2003**, *284*, C988−C998.
(91) Christopoulos, A.; Grant, M.; Ayoubzadeh, N.; Kim, O.; Sauerberg, P.; Jeppesen, L.; El-Fakahany, E. *Pharmacol. Exp. Ther.* **2001**, *298*, 1260−1268.
(92) Bolden, C.; Cusack, B.; Richelson, E. *Eur. J. Pharmacol.* **1991**, *192*, 205−206.
(93) Lindqvist, S. et al. *Br. J. Pharmacol.* **2002**, *137*, 1134−1142.
(94) Fischer, O.; Hart, S.; Gschwind, A.; Ullrich, A. *Biochem. Soc. Trans.* **2003** *Dec;31(Pt 6)*, 1203−8, **2003**, *31*, 1203−1208.
(95) Williams, F.; Messer, W. *Comp. Biochem. Physiol., Part C: Pharmacol., Toxicol. Endocrinol.* **2004**, *13*, 349−353.
(96) Müller, C. *Curr. Top. Med. Chem.* **2003**, *3*, 445−462.
(97) Drazen, J.; Silverman, E.; Lee, T. *Br. Med. Bull.* **2000**, *56*, 1054−1070.
(98) N Nishigaki, M. N.; Ichikawa, A. *Mol. Pharmacol.* **1996**, *50*, 1031−1037.
(99) Schechter, L.; Dawson, L.; Harder, J. *Curr. Pharm. Des.* **2002**, *8*, 139−145.
(100) Fishman, P.; Bar-Yehuda, S. *Curr. Top. Med. Chem.* **2003**, *3*, 463−469.
(101) Graler, M.; Goetzl, E. *FASEB J.* **2004**, *18*, 551−553.
(102) Sneddon, W.; Magyar, C.; Willick, G.; Syme, C.; Galbiati, F.; Bisello, A.; Friedman, P. *Endocrinology* **2004**, *145*, 2815−2823.
(103) Behar, V.; Bisello, A.; Bitan, G.; Rosenblatt, M.; Chorev, M. *J. Biol. Chem.* **2000**, *275*, 9−17.
(104) Kemeny, L.; Kenderessy, A.; Olasz, E.; Michel, G.; Ruzicka, T.; Farkas, B.; Dobozy, A. *Eur. J. Pharmacol.* **1994**, *258*, 269−272.
(105) Brueggemeier, R.; Richards, J.; Joomprabutra, S.; Bhat, A.; Whetstone, J. *J. Steroid Biochem. Mol. Biol.* **2001**, *79*, 75−84.
(106) McAllister, S.; Hurst, D.; Barnett-Norris, J.; Lynch, D.; Reggio, P.; Abood, M. E. *J. Biol. Chem.* **2004**, *279*, 48024−48037.
(107) Xu, B.; Goulding, E.; Zang, K.; Cepoi, D.; Cone, R.; Jones, K.; Tecott, L.; Reichardt, L. *Nat. Neurosci.* **2003**, *6*, 736−742.

CI050006D