BIOINFORMATICS ORIGINAL PAPER Vol. 21 no. 22 2005, pages 4101–4106
doi:10.1093/bioinformatics/bti679

Sequence analysis

Prediction of the coupling specificity of GPCRs to four families of G-proteins using hidden Markov models and artificial neural networks

Nikolaos G. Sgourakis, Pantelis G. Bagos and Stavros J. Hamodrakas*

Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Athens 157 01, Greece

Received on August 16, 2005; revised and accepted on September 12, 2005
Advance Access publication September 20, 2005

ABSTRACT

Motivation: G-protein coupled receptors are a major class of eukaryotic cell-surface receptors. A very important aspect of their function is the specific interaction (coupling) with members of four G-protein families. A single GPCR may interact with members of more than one G-protein families (promiscuous coupling). To date all published methods that predict the coupling specificity of GPCRs are restricted to three main coupling groups $G_{i/o}$, $G_{q/11}$ and G_s, not including $G_{12/13}$-coupled or other promiscuous receptors.

Results: We present a method that combines hidden Markov models and a feed-forward artificial neural network to overcome these limitations, while producing the most accurate predictions currently available. Using an up-to-date curated dataset, our method yields a 94% correct classification rate in a 5-fold cross-validation test. The method predicts also promiscuous coupling preferences, including coupling to $G_{12/13}$, whereas unlike other methods avoids overpredictions (false positives) when non-GPCR sequences are encountered.

Availability: A webserver for academic users is available at http://bioinformatics.biol.uoa.gr/PRED-COUPLE2

Contact: shamodr@cc.uoa.gr

Supplementary information: Results for promiscuous receptors can be found at: http://bioinformatics.biol.uoa.gr/PRED-COUPLE2/tables

INTRODUCTION

G-protein coupled receptors (GPCRs) constitute an extended superfamily of eukaryotic cell surface transmembrane proteins with well-emphasized pharmacological properties (Kristiansen, 2004). This unique class of receptor proteins mediates the actions of various extracellular signals, thus providing an interface between a cell and its environment (Gether, 2000). GPCRs are characterized by a common structural theme of seven transmembrane α-helices arranged in a bundle, presumably with the same spatial arrangement to that found in the only GPCR's solved crystal structure that of rhodopsin (Palczewski *et al*., 2000). According to a commonly used classification scheme (Horn *et al*., 2003), most GPCRs are grouped in classes A–E. Class A is the most widespread and contains rhodopsin-like GPCRs, class B contains the secretin-like GPCRs, while class C the metabotropic glutamate/pheromone receptors. These are the main classes of receptors in animals. From the remaining major classes, the fungal class D comprises pheromone receptors and class E contains cyclic AMP receptors such as those of *Dictyosteliun discoidium*. Other less well characterized classes of GPCRs include the smoothed/frizzled receptors or the chemoreceptors of insects (Hill *et al*., 2002).

The link between an activated GPCR and the cell's physiological responses is a heterotrimeric G-protein ($\alpha\beta\gamma$) located in the interior of the cell that interacts with the activated receptor. Based on sequence similarity among different α subunits that are the main determinants of G-protein coupling specificity and the functionality of the heterotrimers that they participate in, four families of G-proteins are defined in the literature: $G_{i/o}$, $G_{q/11}$, G_s and $G_{12/13}$. The term family is used throughout the manuscript to describe the four classes of G-proteins, in accordance with the gpDB classification scheme (Elefsinioti *et al*., 2004). Members belonging to these four families mediate the actions of the majority of functionally characterized GPCRs. The same receptor can interact with members from more than one family of G-proteins resulting in multiple cellular responses, a phenomenon known as promiscuous coupling (Hermans, 2003). Therefore, the kind of cellular response upon GPCR activation is determined mainly by the selective coupling of GPCRs to members of the four G-protein families.

Despite the large numbers of identified GPCRs in known eukaryotic genomes (Fredriksson and Schioth, 2005), including the human genome, and their profound importance as targets of more than half of the prescribed drugs (Drews, 1996), many GPCRs remain uncharacterized in terms of structure, function and physiology. Thus, there is a need for computer algorithms that predict properties of these orphan GPCRs (Flower and Attwood, 2004). In this study we have focused on the prediction of a very important aspect of GPCR function, that of selective coupling to G-proteins. Such predictions would be very useful in devising experiments to screen orphan receptors for ligands (Ashton *et al*., 2004; Minic *et al*., 2005), since these experiments monitor a specific intracellular response, which is determined by the receptor's coupling specificity.

Till date, several methods have been developed that perform a similar task (Cao *et al*., 2003; Moller *et al*., 2001; Sgourakis *et al*., 2005; Sreekumar *et al*., 2004; Yabuki *et al*., 2005); however their predictions are limited in three families of G-proteins, not including $G_{12/13}$. $G_{12/13}$ proteins are very important mediators of GPCRs actions, since they are known to couple with many diverse receptors (Riobo and Manning, 2005). Furthermore, all previous methods

*To whom correspondence should be addressed.

Table 1. Cross-validation results for all GPCR/G-protein interactions in the dataset

	Sensitivity	Specificity	CCR
$G_{i/o}$	117/122 (96%)	59/66 (89%)	176/188 (94%)
$G_{q/11}$	76/79 (96%)	102/109 (94%)	178/188 (95%)
G_s	51/56 (91%)	122/132 (92%)	173/188 (92%)
$G_{12/13}$	34/35 (97%)	145/153 (95%)	179/188 (95%)
Total	278/292 (95%)	428/460 (93%)	706/752 (94%)

Predictions for all interactions in the dataset evaluated on a five-fold cross-validation test, including those performed by promiscuous receptors. The set comprises 188 receptors that participate in 292 experimentally determined interactions with G-proteins. The method demonstrates high sensitivity and specificity. CCR, correct classification rate.

have not been designed to predict multiple coupling to more than one family of G-proteins (promiscuous coupling of GPCRs). Our method is novel in overcoming these barriers, while providing accurate predictions. In a 5-fold cross-validation training scheme that includes promiscuous receptors, it yields a correct classification rate of 94% (Table 1). Its efficacy relies on a new refined library of profile hidden Markov models (pHMMs) (Eddy, 1998; Eddy *et al*., 1995) that have been trained to discriminate between distinct GPCR coupling groups, including $G_{12/13}$-coupled receptors, as well as on a feed-forward artificial neural network (ANN) that combines the results of individual profiles in order to produce the final prediction. In addition, all previously published methods, excluding the method published by our group, rely on membrane topology information, which drastically diminishes their accuracy due to limitations in membrane topology prediction algorithms (Moller *et al*., 2001). Finally, we have developed a web server for academic users at the URL http://bioinformatics.biol.uoa.gr/PRED-COUPLE2

METHODS

Datasets used for training and evaluating the method

The dataset that we used to train the pHMMs consists of 158 GPCR sequences with detailed coupling specificity information that includes $G_{12/13}$-coupled promiscuous receptors. This set was constructed as follows. Coupling information for non-promiscuous $G_{i/o}$, $G_{q/11}$ and G_s-coupled GPCRs was derived from Alexander *et al*. (2005). We retrieved 123 such GPCRs. Since our aim was not to exclude $G_{12/13}$-coupled receptors or other promiscuous receptors from the predictions of the method, we enriched the coupling data of this primary dataset (Alexander *et al*., 2005) with information regarding such GPCRs. Detailed coupling information for 65 such receptors was retrieved after an exhaustive search in the literature from review articles (Hermans, 2003; Riobo and Manning, 2005; Wong, 2003) as well as several individual articles that focus on elucidating GPCR coupling specificity. Therefore, the final dataset contains 188 GPCRs. While training the pHMMs, we excluded from this enlarged dataset 30 sequences of promiscuous receptors that have not been shown to couple to members of the $G_{12/13}$ family of G-proteins. This simplification was made in order to avoid multiplicity of sequences among training sets for different coupling groups. However, for training the hidden Markov models (HMMs), a total of 226 GPCR-G-protein experimentally determined interactions (distributed among the 158 GPCRs, since this dataset includes 35 promiscuous receptors, all coupled to $G_{12/13}$ proteins) were used. The distribution of interactions among the four families of G-proteins was 96 $G_{i/o}$, 59 $G_{q/11}$, 36 G_s and 35 $G_{12/13}$.

In order to improve the efficacy of the ANN in predictions for promiscuous GPCRs, we used all sequences in the extended dataset (188 in total) to train the ANN. Therefore, the 30 promiscuous receptors in the combined dataset, which have not been shown to couple to members of the $G_{12/13}$ family of G-proteins, were included in the training set for the ANN. These 30 receptors participate in 66 experimentally determined interactions with G-proteins, which are divided among the three remaining families of G-proteins as follows: $G_{i/o}$, 26; $G_{q/11}$,20 and G_s, 20. Thus, the dataset used to train the ANN included all 188 GPCR sequences of the enlarged dataset.

All sequences in this study were retrieved from the UniProt (Bairoch *et al*., 2005) database. From the 188 sequences utilized in different training phases of the method, 181 belong to human GPCRs, three belong to mouse receptors and four to viral GPCRs. The mouse GPCRs were used as substitutes for their human homologues that lacked annotation for transmembrane segments, which is needed in the training phase of the method. The viral GPCRs included in the training sets are members of the Chemokine family that have been shown to couple to $G_{12/13}$ proteins (Rosenkilde *et al*., 2001). Therefore, the majority of sequences in the training set are non-homologous (with the exception of the viral receptors).

Training the hidden Markov models

We performed an exhaustive search for profiles that discriminate between distinct coupling groups of GPCRs, in a manner similar to that of a previously published method by our group (Sgourakis *et al*., 2005). Based on the annotation of transmembrane segments found in the UniProt entries (FT lines), we extracted all intracellular sequence regions (the three intracellular loops and the C-terminus), extended by seven amino acid residues towards the membrane. Several experimental studies have shown that these regions interact with G-proteins (Erlenbach *et al*., 2001; Wess, 1993). For each one of the four coupling groups of GPCRs, we constructed multiple alignments of these sequence regions with CLUSTALX (Thompson *et al*., 1994). For all the alignments, we used the BLOSUM 30 series substitution matrices, starting gap opening penalty of 10 and gap extension penalty of 0.10.

Based on alignment column scores produced by CLUSTALX, we isolated low-entropy (high-scoring) alignment blocks from the aforementioned alignments. We then explored these alignment blocks for sub-blocks that could discriminate between different coupling groups of GPCRs. For every possible sub-block of length greater than seven alignment columns, a pHMM was constructed with the *hmmbuild* program of the HMMER software package. The discriminative power of each pHMM was evaluated in a query against all GPCR sequences of the primary dataset, by measuring the Coverage (i.e. percentage of positives that score a lower *E*-value than the lowest *E*-value scoring negative example in the dataset). The highest scoring pHMM for each low entropy block was selected and added in the final library. The result of this exhaustive search was a library of 25 refined pHMMs.

In order to assess the efficacy of discovered pHMMs in discriminating between different GPCR coupling groups we performed a query of all GPCR sequences in the primary dataset against the refined library with the *hmmpfam* program of the HMMER package. In this search, a default cutoff of two was applied to *E*-value results from all profiles. To produce the final output, results from individual profiles that characterize the same coupling group of GPCRs were combined with the QFAST algorithm (Bailey and Gribskov, 1998). Although this approach performs very well with non-promiscuous GPCRs, as we have demonstrated in a previous method (Sgourakis *et al*., 2005), its sensitivity for promiscuous receptors was no more than 76%. However, the refined profiles discovered by this method can discriminate between non-promiscuous GPCRs with high sensitivity and specificity. Therefore, instead of using the QFAST algorithm, we trained a feed-forward back-propagated ANN (Bishop, 1995) that takes as input the scores from individual profiles and produces the final prediction for each one of the four coupling groups.

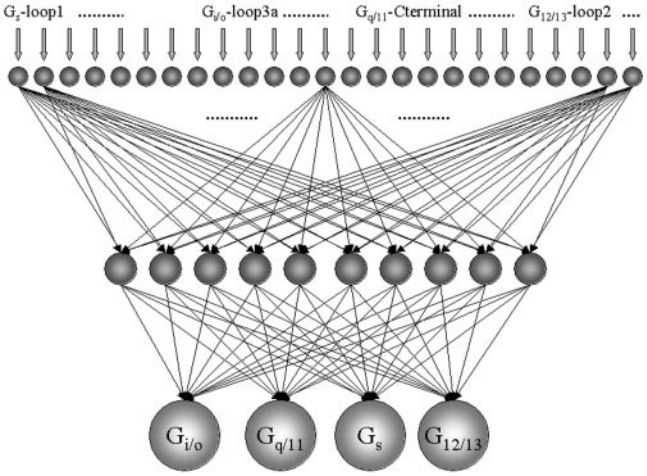

Fig. 1. ANN Architecture. A feed-forward ANN is implemented to produce the final output of the method. Scores obtained from the 25 profile hidden Markov models of the refined library are fed into the network through an equal number of units at the input layer (for simplicity only four scores are depicted as inputs). For instance, G_s-loop1 corresponds to the score of the profile that characterizes the 1st inner loop of G_s-coupled receptors. Ten hidden units intervene between the input and output layers, with maximum connectivity (all-against-all). The final outputs of the network are four numbers, produced by the output units, which correspond to the posterior probability of coupling with each family of G-proteins. For simplicity, most connections between the input and hidden layers have been omitted from the diagram.

Artificial neural network architecture and training

For the development of all ANNs utilized in this method we used the version 3 of the Nevada back propagation ANN simulation platform (http://www.scs.unr.edu/nevprop/)(Goodman, 1996).

The input data for the feed-forward ANN are the 25 scores from individual pHMMs in the refined library (Fig. 1). As output, the network produces the posterior probability that the sequence that produced these scores couples to a specific G-protein family. Thus, the outputs are four numbers ranging from 0 to 1 that correspond to final predictions for each family of G-proteins. After observation of the outputs of all GPCRs in the extended dataset, a safe (empirical) cutoff of 0.3 was applied to discriminate between positive and negative predictions.

A standard feed-forward neural network was used, with a sigmoid transfer function and a single hidden layer of 10 neurons. All possible connections were allowed between the input units and the hidden-layer neurons, as well as between the former neurons and the final output units. The back-propagation algorithm was applied in training the ANN, with random initial weights. The learn rate was set to 0.0001 and the weight decay to -0.001. All networks trained during the development of this method reached the train criterion by the 160th epoch, and produced a C-index (that is the area under the diagram of sensitivity versus specificity) of $\sim$1, demonstrating high correct classification rate in the training set. In order to avoid optimization on the training set (overfitting), we applied the early stopping procedure dividing each time the training set equally between training and validation sets (Bishop, 1995).

Cross validation

The entire procedure described above was repeated five times, by dividing the training dataset of 188 GPCR sequences in five balanced sets. Each of these sets contained approximately the same number of interactions, equally distributed among the four coupling groups. In each round of cross validation we used four sets to train the pHMMs as well as the ANN, while one set was set aside for evaluating the method. As we have already mentioned in the Datasets section, promiscuous receptors that have not been shown to couple to members of the $G_{12/13}$ family of G-proteins were not included in training the pHMMs, but were used instead for training the ANN. Thus, the procedure was repeated five times and the results from each round were added to produce the final evaluation of the method (Table 1).

Improving the specificity of the method against non-GPCR sequences

Although the method described above showed high specificity and sensitivity among GPCRs, it produced many false predictions for non-GPCR sequences. This effect may have been caused by remote similarities of GPCR intracellular loops with loops from other proteins that are not GPCRs, as well as by the small length of these regions resulting in insignificant hits. In a previous method published by our group (Sgourakis *et al*., 2005), this problem was overcome by implementing GPCR profiles from the Pfam database (Bateman *et al*., 2004) to filter non-GPCR sequences in a preceding step. However, many novel putative GPCR sequences are not recognized by any of the Pfam profiles. In order to optimize the method for use against unknown sequences from recently sequenced genomes, we have implemented the QFAST algorithm to combine *E*-values produced by all profiles in the refined library (regardless of their coupling selectivity) and use the output to discriminate between GPCR and non-GPCR sequences. This implementation was found to be very accurate in filtering non-GPCR sequences before proceeding to the neural network and was incorporated in the final method. In addition, the tool that is available through our web server also utilizes pHMMs from the Pfam database to filter the query sequence before executing the main method. Indeed, the library of Pfam profiles has been updated, in comparison with the previous method, to include nine additional profiles that characterize various families of GPCRs.

RESULTS AND DISCUSSION

The predictions of the method for non-promiscuous GPCRs are very accurate. The coupling specificities for all 123 non-promiscuous receptors in the dataset are predicted with 100% sensitivity and 92% specificity in a 5-fold cross-validation procedure, for all three classes of non-promiscuous receptors. This means that all experimentally determined interactions have been correctly predicted, as well as nine not observed promiscuous interactions (false positives). However, the great advantage of this method is its ability to produce reliable predictions for promiscuous GPCRs also. Since, to the best of the authors' knowledge, all $G_{12/13}$-coupled receptors are promiscuous this case also includes predictions of interactions with $G_{12/13}$ proteins.

The method demonstrates high sensitivity and specificity. In a self-consistency test, correct classifications were obtained for 99% of all GPCR-G-protein interactions in the dataset, while false interactions are overpredicted with a rate of <1%. In order to assess whether the results of the method are dependent on optimization to the training set, a 5-fold cross-validation test was performed, as explicitly described in the Materials and methods section. Correct classification rate measurements from this procedure (Table 1) show an insignificant decrease in comparison with measurements obtained from the self-consistency test; therefore the method is not overfitted to the training set.

Predictions for interactions performed by promiscuous receptors are also accurate, demonstrating a correct classification rate of 85%

in the dataset of all 65 promiscuous receptors used for training the method (detailed results for each GPCR are presented in the Supplementary information, see Abstract). Predictions for interactions with $G_{12/13}$ proteins are predicted with a correct rate of 95%, demonstrating high sensitivity (97%) and specificity (95%). However, it should be noted that promiscuous interactions are being experimentally identified with direct and indirect methods; therefore overpredictions for such receptors could actually be observed using high-resolution methods. Such ambiguous cases of promiscuous receptors are widely observed in the literature (Riobo and Manning, 2005).

An example that demonstrates the validity of our results for promiscuous receptors is that of the human cytomegalovirus–encoded GPCR named US28. This receptor has been found to interact with chemokines and shows similarity to the human chemokine receptors. According to an experimental study by Casarosa *et al*. (2003), US28 shows promiscuous coupling to members of different G-protein families, while its homologue in rat cytomegalovirus R33 couples entirely to $G_{i/o}$. These experimental observations agree with our predictions, despite the fact that R33 was not even included in the training set. Furthermore, our method predicts the experimentally determined coupling specificity ($G_{q/11}$) for the novel GPCR GPRg1 (Matsuo *et al*., 2005). For the same receptor, our method also predicts coupling to $G_{i/o}$ proteins, but with a smaller posterior probability.

When querying a sequence against the refined library there is no need to extract the intracellular regions, since the profiles are very specific. Therefore, this method does not rely on user-supplied membrane topology information or prediction. This feature makes the tool suitable for analysis of orphan GPCRs, such as the *Arabidopsis thaliana* 7TM receptor GCR1, which has been experimentally found to interact with the G-protein α subunit GPA1 (Pandey and Assmann, 2004). This is the only G-protein α subunit identified in the genome of *A.thaliana* (Apone *et al*., 2003; Assmann, 2002). Furthermore, Pandey and Assmann (2004) have shown that this interaction is dependent on the GPCR's intracellular domains. Our method predicts interaction of this receptor only with members of the $G_{i/o}$ family of α subunits. Indeed, in a BLAST (Altschul *et al*., 1997) query of GPA1 against all G-proteins in the gpDB database (Elefsinioti *et al*., 2004), members of the $G_{i/o}$ family dominate the best scoring hits (*E*-value below e^{-63}, lower than any member of the other three families). In fact, the *Arabidopsis* Gα subunit is $\sim$30% identical with mammalian subunits of the $G_{i/o}$ family, while most of this conservation is located in the regions that determine the coupling specificity for the entire G-protein complex (Jones, 2002; Jones and Assmann, 2004). For instance, the N-terminal sequence of the protein, a region shown to interact with the activated receptor (Roginskaya *et al*., 2004), is similar between GPA1 and members of the $G_{i/o}$ family. In addition, experimental data from *Cryptococcus neoformans* show that a homologue of GPA1 inhibits adenylate cyclase (Alspaugh *et al*., 2002), a function commonly attributed to $G_{i/o}$ α subunits. These findings signify the resemblance of GPA1 to members of the $G_{i/o}$ family and imply the evolutionary ancestry of $G_{i/o}$-coupled GPCRs mediated pathways among fungi, plants and animals. Another example comes from the recently identified class of GPCRs PTH_{11} from the pathogenic fungus *Magnaporthe grisea*, one of the largest classes of fungal GPCRs (Kulkarni *et al*., 2005). Using our method, several receptors of this class were also predicted to interact with $G_{i/o}$-like proteins. This class of GPCR-like proteins is one of the most ancient, since there is evidence of its existence in fungi 1210 Mya (Kulkarni *et al*., 2005). However, currently there are no experimental data that indicate physical interaction with G-proteins.

Several methods have been reported so far for the prediction of GPCRs coupling specificity to G-proteins. However, they are all limited to the three classes of G-proteins. The methods that implement regular expression patterns (Moller *et al*., 2001), a Naïve Bayes model (Cao *et al*., 2003) and hidden Markov models respectively (Sreekumar *et al*., 2004) rely on some extent on membrane topology information and their limitations have been already described extensively in Sgourakis *et al*. (2005). A previous method developed by our group utilizes refined pHMMs of high discriminative power. However, results from individual hits were combined in a single step by the QFAST algorithm. This is conceptually similar to a minimal ANN with one output unit and no hidden layers. Since interactions between GPCRs and G-proteins are not equally distributed among GPCR intracellular regions, it is clear that an ANN with a hidden layer would provide a much more efficient model. Furthermore, in the current method, we have developed a larger and more efficiently trained library of refined pHMMs that include profiles for $G_{12/13}$ coupled receptors.

Recently, a method that apart from the receptor's sequence requires information of its ligand has been presented (Yabuki *et al*., 2005). Authors claim a correct classification rate of 85%, by combining HMMs and support vector machines (SVMs) to perform the prediction. Although the method is available online, it has some serious disadvantages. First, the requirements of ligand specificity render it inapplicable for cases of putative/orphan GPCRs, where this information is unknown. Second, this method utilizes HMMs to predict the family of a GPCR in a preceding step and then links this prediction with G-protein coupling specificity. However, several cases have been reported for GPCRs that belong to the same subtype, and have demonstrated different coupling specificities (Wess, 1998). Finally, the authors mention using the structure of rhodopsin as template to calculate the boundaries of transmembrane α-helices in the training phase of the method. However, they do not mention details of the prediction algorithm they use. This step is very important for the particular method, since the parameters used in SVMs are dependent on the location of these regions. Thus, the method is bound always to predict seven transmembrane segments (as well as coupling to G-proteins), even for non-transmembrane proteins. Therefore, the method does not control appropriately for the rate of false positives, a fact that renders questions about its applicability as a stand-alone pipeline for large-scale GPCR analysis in sequenced genomes.

Throughout the literature GPCRs are traditionally classified in the four aforementioned coupling groups. However, it should be noted that G-proteins consist of more GPCR interaction groups, than the traditionally defined $G_{i/o}$, $G_{q/11}$, G_s and $G_{12/13}$ families. A study that implemented chimeric G-proteins clearly demonstrated that distinct GPCR coupling groups exist within the $G_{i/o}$ family, as exemplified by 5-HT1A, 5-HT1B Serotonin and M2 muscarinic receptors that couple with G_{i1} but not G_t (Slessareva *et al*., 2003). Furthermore, members of the $G_{q/11}$ family of G-proteins have been showed to mediate the activation of phospholipase C for several G_s- and $G_{i/o}$-coupled receptors (Ho *et al*., 2001). Therefore, classification

of GPCRs according to G-protein coupling to four groups may be considered obsolete. Perhaps the availability of more precise experimental data will allow a classification scheme that focuses at the lowest level.

Furthermore, with the implementation of GPCR-specific profiles from Pfam (Bateman *et al*., 2004), the current method efficiently filters non-GPCR sequences; all sequences in two independent test sets comprising 1133 globular and 1356 transmembrane proteins (Papasaikas *et al*., 2003) not classified as GPCRs were recognized as such by our method. This accomplishment has not been achieved by any of the aforementioned methods that perform a similar task.

In conclusion, we have developed a method that efficiently classifies GPCRs according to their coupling specificity to the four families of G-proteins, including $G_{12/13}$-coupled receptors. To the best of the authors' knowledge, this is the first published method that performs this task. In addition, the method presented in this study produces reliable predictions for promiscuous receptors as well, a task not performed by any other published method. Another advantage of the method, in comparison with most previously published methods (excluding a method published by our group), is the fact that no membrane topology information is required to perform reliable predictions. We report high sensitivity and specificity, as evaluated by the 5-fold cross-validation procedure. A web server running the application has been developed, and is freely available for non-commercial users.

We expect that predictions from our web server will be useful for GPCR researchers in designing experiments to screen orphan GPCRs for potential ligands, as well as for large-scale Bioinformatics analyses in published proteomes. The challenges in GPCR coupling prediction for the future are more accurate predictions that distinguish between members of the same G-protein family. This task relies mainly on the availability of more precise biochemical data from high-resolution experiments.

Conflict of Interest: none declared.

REFERENCES

Alexander,S.P. *et al.* (2005) 7 TM receptors. *Br. J. Pharmacol.*, **144** (Suppl 1), S4–S62.

Alspaugh,J.A. *et al.* (2002) Adenylyl cyclase functions downstream of the Galpha protein Gpa1 and controls mating and pathogenicity of *Cryptococcus neoformans*. *Eukaryot. Cell*, **1**, 75–84.

Altschul,S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Apone,F. *et al.* (2003) The G-protein-coupled receptor GCR1 regulates DNA synthesis through activation of phosphatidylinositol-specific phospholipase C. *Plant Physiol.*, **133**, 571–579.

Ashton,M. *et al.* (2004) The selection and design of GPCR ligands: from concept to the clinic. *Comb. Chem. High Throughput Screen*, **7**, 441–452.

Assmann,S.M. (2002) Heterotrimeric and unconventional GTP binding proteins in plant cell signaling. *Plant Cell*, **14** (Suppl), S355–S373.

Bailey,T.L. and Gribskov,M. (1998) Combining evidence using *P*-values: application to sequence homology searches. *Bioinformatics*, **14**, 48–54.

Bairoch,A. *et al.* (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **33**, D154–D159.

Bateman,A. *et al.* (2004) The Pfam protein families database. *Nucleic Acids Res.*, **32**, D138–D141.

Bishop,C.M. (1995) *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, England.

Cao,J. *et al.* (2003) A naive Bayes model to predict coupling between seven transmembrane domain receptors and G-proteins. *Bioinformatics*, **19**, 234–240.

Casarosa,P. *et al.* (2003) Constitutive signaling of the human cytomegalovirus-encoded receptor UL33 differs from that of its rat cytomegalovirus homolog R33 by promiscuous activation of G proteins of the Gq, Gi, and Gs classes. *J. Biol. Chem.*, **278**, 50010–50023.

Drews,J. (1996) Genomic sciences and the medicine of tomorrow. *Nat. Biotechnol.*, **14**, 1516–1518.

Eddy,S.R. (1998) Profile hidden Markov models. *Bioinformatics*, **14**, 755–763.

Eddy,S.R. *et al.* (1995) Maximum discrimination hidden Markov models of sequence consensus. *J. Comput. Biol.*, **2**, 9–23.

Elefsinioti,A.L. *et al.* (2004) A database for G proteins and their interaction with GPCRs. *BMC Bioinformatics*, **5**, 208.

Erlenbach,I. *et al.* (2001) Single amino acid substitutions and deletions that alter the G protein coupling properties of the V2 vasopressin receptor identified in yeast by receptor random mutagenesis. *J. Biol. Chem.*, **276**, 29382–29392.

Flower,D.R. and Attwood,T.K. (2004) Integrative bioinformatics for functional genome annotation: trawling for G protein-coupled receptors. *Semin. Cell. Dev. Biol.*, **15**, 693–701.

Fredriksson,R. and Schioth,H.B. (2005) The repertoire of G-protein-coupled receptors in fully sequenced genomes. *Mol. Pharmacol.*, **67**, 1414–1425.

Gether,U. (2000) Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr. Rev.*, **21**, 90–113.

Goodman,P.H. (1996) *NevProp Software, Version 3*. University of Nevada, Reno, NV.

Hermans,E. (2003) Biochemical and pharmacological control of the multiplicity of coupling at G-protein-coupled receptors. *Pharmacol. Ther.*, **99**, 25–44.

Hill,C.A. *et al.* (2002) G protein-coupled receptors in *Anopheles gambiae*. *Science*, **298**, 176–178.

Ho,M.K. *et al.* (2001) Galpha(14) links a variety of G(i)- and G(s)-coupled receptors to the stimulation of phospholipase C. *Br. J. Pharmacol.*, **132**, 1431–1440.

Horn,F. *et al.* (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.*, **31**, 294–297.

Jones,A.M. (2002) G-protein-coupled signaling in *Arabidopsis*. *Curr. Opin. Plant. Biol.*, **5**, 402–407.

Jones,A.M. and Assmann,S.M. (2004) Plants: the latest model system for G-protein research. *EMBO Rep.*, **5**, 572–578.

Kristiansen,K. (2004) Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.*, **103**, 21–80.

Kulkarni,R.D. *et al.* (2005) Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biol.*, **6**, R24.

Matsuo,A. *et al.* (2005) Molecular cloning and characterization of a novel Gq-coupled orphan receptor GPRg1 exclusively expressed in the central nervous system. *Biochem. Biophys. Res. Commun.*, **331**, 363–369.

Minic,J. *et al.* (2005) Yeast system as a screening tool for pharmacological assessment of g protein coupled receptors. *Curr. Med. Chem.*, **12**, 961–969.

Moller,S. *et al.* (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17**, 646–653.

Moller,S. *et al.* (2001) Prediction of the coupling specificity of G protein coupled receptors to their G proteins. *Bioinformatics*, **17** (Suppl 1), S174–S181.

Palczewski,K. *et al.* (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, **289**, 739–745.

Pandey,S. and Assmann,S.M. (2004) The *Arabidopsis* putative G protein-coupled receptor GCR1 interacts with the G protein alpha subunit GPA1 and regulates abscisic acid signaling. *Plant Cell*, **16**, 1616–1632.

Papasaikas,P.K. *et al.* (2003) A novel method for GPCR recognition and family classification from sequence alone using signatures derived from profile hidden Markov models. *SAR QSAR Environ. Res.*, **14**, 413–420.

Riobo,N.A. and Manning,D.R. (2005) Receptors coupled to heterotrimeric G proteins of the G12 family. *Trends Pharmacol Sci.*, **26**, 146–154.

Roginskaya,M. *et al.* (2004) Effects of mutations in the N terminal region of the yeast G protein alpha-subunit Gpa1p on signaling by pheromone receptors. *Mol. Genet. Genomics*, **271**, 237–248.

Rosenkilde,M.M. *et al.* (2001) Virally encoded 7TM receptors. *Oncogene*, **20**, 1582–1593.

Sgourakis,N.G. *et al.* (2005) A method for the prediction of GPCRs coupling specificity to G-proteins using refined profile hidden Markov models. *BMC Bioinformatics*, **6**, 104.

Slessareva,J.E. *et al.* (2003) Closely related G-protein-coupled receptors use multiple and distinct domains on G-protein alpha-subunits for selective coupling. *J. Biol. Chem.*, **278**, 50530–50536.

Sreekumar,K.R. *et al*. (2004) Predicting GPCR-G-protein coupling using hidden Markov models. *Bioinformatics*, **20**, 3490–3499.

Thompson,J.D. *et al*. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*., **22**, 4673–4680.

Wess,J. (1993) Mutational analysis of muscarinic acetylcholine receptors: structural basis of ligand/receptor/G protein interactions. *Life Sci*., **53**, 1447–1463.

Wess,J. (1998) Molecular basis of receptor/G-protein-coupling selectivity. *Pharmacol. Ther*., **80**, 231–264.

Wong,S.K. (2003) G protein selectivity is regulated by multiple intracellular regions of GPCRs. *Neurosignals*, **12**, 1–12.

Yabuki,Y. *et al*. (2005) GRIFFIN: a system for predicting GPCR-G-protein coupling selectivity using a support vector machine and a hidden Markov model. *Nucleic Acids Res*., **33**, W148–W153.