# GPCR and G-protein Coupling Selectivity Prediction Based on SVM with Physico-Chemical Parameters

**Makiko Suwa**[1] m-suwa@aist.go.
**Yukimitsu Yabuki** [1,2] yukimitsu-yabuki@is.naist.jp
**Takahiko Muramatsu**[1,3] taka-mu@is.naist.jp
**Takatsugu Hirokawa**[1] t-hirokawa@aist.go.jp
**Hidetoshi Mukai**[4] mukaih@libra.ls.m-kagaku.co.jp

[1] Computational Biology Research Center, 2-43 Aomi, Koto-ku, Tokyo 135-0064, Japan
[2]. Information and Mathematical Science Laboratory, Inc, 2-43-1, Ikebukuro, Toshima-ku, Tokyo, 171-0014, Japan
[3] Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma-shi, Nara 630-0101, Japan
[4]. Mitsubishi Kagaku Institute of Life Science, 11, Minamiootani, Machida-shi, Tokyo, 194-8511, Japan.

**Keywords**: GPCR, G-protein coupling selectivity, SVM, physico-chemical parameters

## 1 Introduction

G-protein coupled receptor (GPCR), which is composed of seven transmembrane helices, plays as the interface of signal transduction. The external stimulation for GPCR, such as binding amine, peptide, hormone, odorant, molecules, induce the coupling with G-protein (Gi/o, Gq/11,Gs, $G_{12/13}$) followed by different kinds of signal transduction to inner cell. About half of distributed drugs is intending to control this GPCR-G protein binding system, and therefore this system is important research target for the development of effective drug. For this purpose, it is necessary to monitor, effectively and comprehensively, of the activation of G protein by identifying ligand combined with GPCR. Since, at present, it is difficult to construct such biochemical experiment system, if the answers for experimental results can be prepared beforehand by using bioinformatics techniques, large progress is brought to G protein related drug design.

Previous works for predicting GPCR-G protein coupling selectivity [1,2,3,4] are using sequence pattern search, statistical models, and HMM representations showed high sensitivity of predictions. However, there are still no works that can predict with both high sensitivity and high specificity. In this work we extracted comprehensively the physico-chemical parameters of each part of ligand, GPCR, G protein, and choose the parameter which has strong correlation with the coupling selectivity of G protein. These parameters were put as a feature vector, used for GPCR classification based on SVM

## 2. Method

We consider that GPCR, and G protein serve as a joint complex, works as one system, therefore, all the physico-chemical features of ligand, the extra-cellular loops, transmembrane helices, cytoplasmic loops correlate with G-protein coupling selectivity. The physico-chemical parameters were extracted from ligand, loop region, and transmembrane helices. The correlation nature of these parameters and G protein selectivity was analyzed, and the parameters with high correlation are picked up, and they were referred as feature vector when classifying it according to SVM. We focus on only Class A family, having 3D structure representative (rhodopsin), and transmembrane helix regions of a sequence were predicted from the sequence alignment with that of rhodopsin.

### 2.1 Dataset

132 amino acid sequence of GPCR(Gi/o binding type: 61sequences、Gq/11 binding type: 47 sequences, Gs binding type: 24sequences) are picked up from SWISS-PROT, TrenBL database with ligand and G-protein information written in TiPS [5] and GPCRDB [6]

### 2.2 Classification by using SVM

The SVM calculation was performed using software packages LIBSVM [7]. From this package we tried several kind of combination of kernel functions and parameter C . The evaluation was performed using various types of

kernel functions such as linear, polynomial, radial basis, sigmoid functions. The SVM was provided with feature vectors obtained from physico-chemical parameters of several regions of GPCRs.

Cross-validation test of n-fold divide randomly known Gi/o Gq/11 and Gs sequence (N sequences) into n equal parts, and consider N (1-1/n) sequence as a training set, and N/n sequence as a test set. This recombination is performed n times. Evaluation of sensitivity, specificity was carried out about distinction of a test set.

## 3.Result

We first evaluated the performance for classifying individual G proteins based on SVM using known GPCR dataset. The Gs coupling predictions showed the highest sensitivity (= 83.33%) and specificity (= 95.24%). when five physico-chemical parameters: (a)The third intercellular (I3) loop length, (b) C terminal loop length, (c) amine profile score and (d)existence of Pro at 170-th position of rhodopsin, (e) total Lys and Arg number at C terminal side of I3 loop, are applied as feature vectors, using RBF kernel functions. On the other hand, the performance of prediction for Gi/o indicated the highest sensitivity (= 91.80%), specificity (=94.91%) and those for Gq/11 indicated the most high sensitivity (= 93.617%) and specificity (= 89.80%), when seven parameters (above five parameters and additional two parameters: (f) Peptide-profile score, (g) The ligand molecular weight) were used with polynomial kernel function.

Since the two cases of Gs prediction and others prediction require different parameter sets and conditions to achieve the best performance, we constructed the hierarchical system, in which Gs coupling is predicted at first stage and Gi/o or Gq/11 are discriminated from sequence which was not judged as to couple with Gs protein (second stage). Applying this hierarchical system to known sequences through 10000 times of 4-fold cross validation, the averaged discrimination sensitivity and specificity became 85% and 88% for Gi/o, 85% and 84% for Gq/11, and 87% and 88% for Gs, respectively. Thus, this system enables us to reduce false positive as compared with the previous works[1,2,3,4]. Of course the SVM methodology is one of the factors for increasing predicting ability, however we think that the physico-chemical parameters we choose here also become the dominant factor. It is expected that the analysis of these parameters lead to understanding the GPCR-G protein coupling mechanism. We hope this system will contribute to predict the function of orphan receptors.

## References


[1]Moller S, Vilo J, Croning MD. Prediction of the coupling specificity of G protein coupled receptors to their G proteins. *Bioinformatics.* **17** Suppl 1:S174-81, 2001.

[2]Cao J, Panetta R, Yue S, Steyaert A, Young-Bellido M, Ahmad S. Related Articles, Links A naive Bayes model to predict coupling between seven transmembrane domain receptors and G-proteins. *Bioinformatics.* 19(2):234-40,2003.

[3]Sreekumar KR, Huang Y, Pausch MH, Gulukota K. Predicting GPCR-G protein coupling using hidden Markov models. *Bioinformatics.* 2004.

[4]Qian B, Soyer OS, Neubig RR, Goldstein RA. Depicting a protein's two faces: GPCR classification by phylogenetic tree-based HMMs. *FEBS Lett.* Nov 6;554(1-2):95-99,2003.[6]Alexander, S., Mathie, A., Peters, J., Mackenzie, G. and Smith, A (2001) *TiPS* Receptor nomenclature supplement.

[5]Alexander, S., Mathie, A., Peters, J., Mackenzie, G. and Smith, A *TiPS Receptor nomenclature supplement.*2001.

[6] Horn, F., Bettler E, Oliveira L, Campagne F, Cohhen, FE., Vriend, G. GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* 31(1):294-297,2003.

[7]Numerical experiments on nu-SVM using LIBSVM can be found inTraining nu-Support Vector Classifiers: Theory and Algorithms. Neural Computation *Neural Computation* 13(9), 2119-2147,2001.