

**Prediction of G-Protein Coupling Specificity of GPCR:**  
**Abstract**

G-protein coupled receptors (GPCR) are a well-studied superfamily of proteins in pharmacological research as they are the target of approximately 60% of current drugs on the market (Muller, 2000). Coupling with G-proteins, these receptors regulate much of the signal transduction across the cell membrane and therefore the cell's reactions to external stimuli. Abnormalities in GPCR regulation lead to a variety of diseases, such as cancer, cardiovascular disease, Alzheimer's and Parkinson's disease. Identification of the ligands that can activate a particular GPCR can provide information on the causes and symptoms of the disease the GPCR is involved in. However, identification of such ligands requires knowledge of the G-protein coupling specificity of the receptor so that the appropriate G-proteins are in the cell, ready to interact with the receptors once they are activated by a ligand.

Despite the importance of the G-protein coupling specificity prediction problem, there have only been two published studies on the task. Both studies focused on the intracellular domains of the receptor sequences, with the earlier study using n-gram based features (Möller *et al.*, 2001) and the latter using alignment-based features (Cao *et al.*, 2003). Here, we propose a two-step hybrid approach employing the k-nearest-neighbour classifier on both alignment-based and n-gram based features. Preliminary experiments in our study show there is information about coupling specificity within the extracellular and transmembrane regions as well as intracellular regions. Thus, our approach uses whole sequence alignment and n-grams extracted from the whole sequence.

First, BLAST is used to find the  $k_1$  sequences in the training set most similar to the given test sequence. For each G-protein family, if all  $k_1$  training sequences couple to that G-protein family, then the test sequence is predicted to couple to the G-protein family as well. Next, the test sequence is given to a set of binary k-NN classifiers that use unigrams, bigrams, trigrams and tetragrams counts as features. The scores from the binary classifiers are normalized to sum to one. If the normalized score from a binary classifier is above a set threshold  $L$ , then the receptor sequence is predicted to couple to the classifier's associated G-protein family. Thus, a receptor may be predicted to couple with a G-protein family either by the alignment-based classifier or the n-gram based classifier. Using only 81.3% of the training set and the same test set, our coupling specificity predictor is able to outperform Cao *et al.* in precision, recall and F1. These results demonstrate that G-protein coupling specificity is determined by both long-range interactions and local information in the transmembrane, cytoplasmic and extracellular domains of GPCR.

J. Cao, R. Panetta, S. Yue, A. Steyaert, M. Young-Bellido, and S. Ahmad. A naïve Bayes model to predict coupling between seven transmembrane domain receptors and G-proteins. *Bioinformatics*, 19(2): 234-240, 2003.

S. Möller, J. Vilo, and M. D. R. Croning. Prediction of the coupling specificity of G protein coupled receptors to their G proteins. *Bioinformatics*, 17(1):S174-S181, 2001.

Muller, G. Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach. *Current Medical Chemistry*, 7(9): 861-888, 2000.