Mol. Cells *27*, 629-634, June 30, 2009
DOI/10.1007/s10059-009-0091-2

Molecules and Cells

©2009 KSMCB

A Machine Learning Based Method for the Prediction of G Protein-Coupled Receptor-Binding PDZ Domain Proteins

Hae-Seok Eo, Sungmin Kim[1], Hyeyoung Koo[2], and Won Kim*

G protein-coupled receptors (GPCRs) are part of multi-protein networks called 'receptosomes'. These GPCR interacting proteins (GIPs) in the receptosomes control the targeting, trafficking and signaling of GPCRs. PDZ domain proteins constitute the largest protein family among the GIPs, and the predominant function of the PDZ domain proteins is to assemble signaling pathway components into close proximity by recognition of the last four C-terminal amino acids of GPCRs. We present here a machine learning based approach for the identification of GPCR-binding PDZ domain proteins. In order to characterize the network of interactions between amino acid residues that contribute to the stability of the PDZ domain-ligand complex and to encode the complex into a feature vector, amino acid contact matrices and physicochemical distance matrix were constructed and adopted. This novel machine learning based method displayed high performance for the identification of PDZ domain-ligand interactions and allowed the identification of novel GPCR-PDZ domain protein interactions.

INTRODUCTION

G protein-coupled receptors (GPCRs) are the largest family of cell surface receptors. GPCRs interact with GPCR interacting proteins (GIPs) that are implicated in various functions such as trafficking and targeting of GPCRs into appropriate subcellular compartments, assembly of the receptors into multi-protein networks called 'receptosomes' and fine-tuning of GPCR signaling events (Bockaert et al., 2004; Eo et al., 2007; Hanyaloglu and von Zastrow, 2008; Kreienkamp, 2002). During the last several years, a wide range of membrane associated and intracellular GIPs have been discovered, suggesting that a better understanding of the physiological mechanisms associated with GPCRs and other components may be provide a solid foundation for GPCR-directed pharmacotherapy (Hanyaloglu and von Zastrow, 2008).

PDZ (PSD-95, Dlg and ZO-1) domain proteins constitute the largest family of proteins that is involved in the organization of GPCR signaling complexes (Day and Kobilka, 2006; van Ham and Hendriks, 2003). PDZ domain proteins play a critical role in the targeting of GPCRs to specific membrane compartments and their assembly into receptosomes (Bockaert et al., 2004). Single PDZ domain proteins contain one or multiple copies of the PDZ domains, and structural features allow the proteins to mediate specific protein-protein interactions (Skelton et al., 2003).

A PDZ domain consists of approximately 80 to 100 amino acids, which form six βstrands, βA through βF, and two α helices, αA and αB (Basdevant et al., 2006; Chi et al., 2006). The four C-terminal residues of the PDZ ligand interact with an elongated surface groove located between the βB strand and the αB helix in an antiparallel manner (Hung and Sheng, 2002; Nourry et al., 2003). However, PDZ domains can bind to distinct peptide motifs, and such promiscuity in ligand recognition has been the main obstacle to the development of effective methods for the identification of PDZ domain-ligand interactions (Basdevant et al., 2006; Vaccaro and Dente, 2002). Furthermore, recent studies have shown that the binding specificity, which is primarily determined by the last four residues of the ligand, can also be influenced by residues located further upstream (Kurakin et al., 2007). In this study, we present a novel machine learning based method for the characterization of PDZ domain-ligand interactions and for the identification of novel PDZ domain-GPCR interactions.

In order to infer and to characterize PDZ domain-ligand binding specificity, several studies have attempted to place PDZ domains into discrete functional classes based on the physicochemical properties of specific residues in the PDZ domain or based on the amino acid patterns of their ligands (Bezprozvanny and Maximov, 2001; Songyang et al., 1997). In addition, high-throughput or computational techniques have also been used to facilitate a better understanding of PDZ domain-ligand interactions (Day and Kobilka, 2006; Giallourakis et al., 2006; Reina et al., 2002; Stiffler et al., 2007; Wiedemann et al., 2004). However, most of these studies have not succeeded in presenting an understanding of the specificity of these interactions due to the relatively short length of the ligands that participate in PDZ domain recognition and many false positive interactions

School of Biological Sciences, Seoul National University, Seoul 151-742, Korea, [1]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Korea, [2]Department of Biological Science, Sangji University, Wonju 220-702, Korea
*Correspondence: wonkim@plaza.snu.ac.kr

Received November 13, 2008; revised May 1, 2009; accepted May 12, 2009; published online June 12, 2009

Keywords: G protein-coupled receptor, GPCR interacting protein, machine learning, PDZ domain, support vector machine

have been identified in the studies.

In this study, we present a support vector machine (SVM) classifier that integrates PDZ domain-peptide binding data verified by biological experiments with structural information derived from the crystal structure of the PDZ domain (Doyle et al., 1996). The interaction interfaces are encoded in a set of numerical variables and an SVM classifier is then trained to infer PDZ domain-ligand interactions. With the classifier, we show its performance by the use of a 10-fold cross validation test and illustrate the utility for the identification of novel PDZ domain-GPCR interactions.

MATERIALS AND METHODS

Dataset

The PDZ domain-ligand binding dataset used for constructing the amino acid contact matrices and training the method was obtained from the PDZBase database (Beuming et al., 2005). The PDZBase database contains all the known PDZ domain-ligand interactions that have been identified from *in vivo* or *in vitro* experiments. The PDZ domain-ligand non-binding dataset was also extracted from a previous study (Stiffler et al., 2007) for constructing non-binding contact matrices and training the method. These investigators adopted protein microarrays and then used quantitative fluorescence polarization to verify the interactions of mouse PDZ domains. A total of 348 PDZ domain-ligand binding interactions were obtained from the PDZ Base database, along with the same number for the negative dataset, which was confirmed by the use of the quantitative fluorescence polarization method (Stiffler et al., 2007).

Prediction of PDZ domain-peptide interactions

Application of the machine learning approach for the prediction of PDZ domain-peptide interactions was motivated by the canonical binding mechanism where individual residues in a peptide ligand always target the same position of residues of the PDZ domain-peptide binding pockets (Songyang et al., 1997; Wiedemann et al., 2004). The framework of the machine learning approach to PDZ domain-peptide binding prediction is described in Fig. 1.

Contact matrices

To describe the specificity of PDZ domain-ligand interactions, contact matrices were constructed from the training dataset. The contacting PDZ domain residues for each peptide position were determined based on the solved structure of a PDZ domain-peptide complex (Doyle et al., 1996). Corresponding residues lying on the interaction surface of PDZ domains were deduced by alignment of their sequences with PSD95-3 and AF6 PDZ domains and were labeled (Doyle et al., 1996; Songyang et al., 1997; Wiedemann et al., 2004) (Fig. 1A). The network of interactions between amino acid residues is described in Fig. 1B.

The contact preferences of individual peptide positions with the contacting PDZ domain residues were defined as $\frac{N_i}{T_i}$, where N_i indicates the number of occurrence of any specific pair of residues and T_i is the total number of amino acid contacts of the ith-contact pair in a PDZ domain-peptide binding. Figure 1C indicates the contact preference between the P-1 residue in a peptide and the βC7 residue in a PDZ domain. Thus, a total of 20 contact matrices were constructed between a domain and a peptide (Fig. 1B). The same procedure was performed with the PDZ domain-ligand non-binding dataset (Stiffler et al., 2007) for the construction of non-binding contact matrices (see Dataset). In addition to the contact matrices, a physicochemical distance between each contact pair was calculated based on polarity, molecular volume and the chemical composition of the amino acids (Grantham, 1974) and was normalized in the range of [0, 1] by dividing by the largest distance (Grantham's distance = 215). The normalized physiochemical distances were adopted to describe the cooperative contributions of each residue in a peptide with all of its contacting residues in the PDZ domain. For example, the cooperative contribution of a P-3 residue in a peptide that interacts together with βB3, βB5, βC4 and βC5 in a PDZ domain was calculated by the average of the physicochemical distances between the residues (Fig. 1B) (Altuvia and Margalit, 2004; Ferraro et al., 2006).

Feature extraction

The specificity and stability of domain-ligand interactions have been described by the two assumptions of the independent contribution of each contact residue pair between a domain and a ligand and the cooperative contribution of each peptide position with all of its contacting domain residues (Altuvia and Margalit, 2004; Ferraro et al., 2006). To improve the reliability of the SVM classifier, we combined the two assumptions in the description of the PDZ domain-ligand interaction.

A PDZ domain-peptide complex, which has the 20 contact residue pairs, can be defined as a set of the element f_i, $I \in \{1, ..., 20\}$, where f_i is the relative frequency of the specific amino acid contact of the i-th contact pair ($f_i = \frac{N_i}{T_i}$, see *Contact matrices*). The relative frequency, $f_i(+)$ and $f_i(-)$, of each contact pair can be extracted from the i-th binding and non-binding contact matrices, respectively. If $f_i(+)$ is greater than $f_i(-)$, the i-th contact is more relevant for binding whereas if $f_i(-)$ is greater than $f_i(+)$, the i-th contact is more relevant for non-binding. The i-th contact pair in a PDZ domain-peptide complex can be represented as the numerical code $C(f_i)$, where $C(f_i)$ can be encoded according to its binding significance:

$C(f_i) = 1$ if $f_i(+) > f_i(-)$
$C(f_i) = 0$ if $f_i(+) = f_i(-)$
$C(f_i) = -1$ if $f_i(+) < f_i(-)$

Consequently, a set of 20 contacts in the PDZ domain-ligand binding was transformed into a set of 20 numerical codes that reflect the assumption of an independent contact representation.

The cooperative contributions of the individual peptide position to the stability of the complex were calculated by the average of the physicochemical distances between the contact residues (see *Contact matrices*) and were also added into a feature vector (Fig. 1D).

Performance evaluation

To implement the SVM classifier, we adopted the SVM^{light} package version 6.01 (Joachims, 1999) and tested the classifier with linear, polynomial and radial basis function (RBF) kernels. Subsequently, the classifier was applied to the GPCRs to identify PDZ domain-GPCR interactions.

The performance of the SVM classifier was evaluated by the use of a 10-fold cross validation test. We trained the SVM classifier ten times, each time leaving out one of the subsets from the training. The remaining subset for each time was used to estimate the performance of the trained SVM classifier. The performance quality of the SVM classifier was evaluated by five measurements including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy.

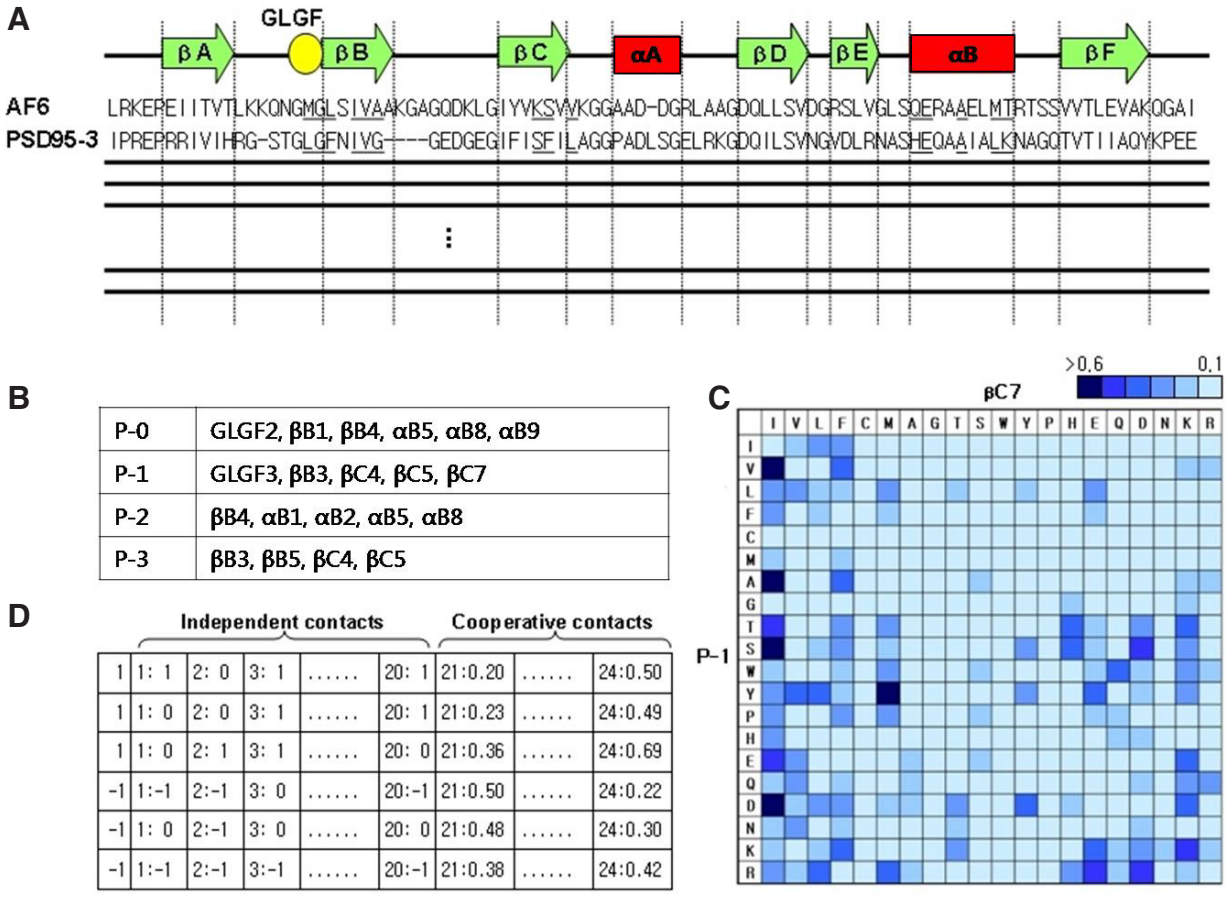

Fig. 1. The computational procedure for encoding a PDZ domain-ligand interaction into a feature vector is presented. (A) Structure-guided sequence alignment of PDZ domains; 14 underlined amino acids, which contact with residues in the ligand peptide, were extracted. (B) Contact combination between ligand and PDZ domain residues; column 1 represents the four carboxyl-terminal positions of a ligand peptide with the P-0,…, P-3 nomenclature as defined by Songyang et al. (1997) and column 2 represents the PDZ residues that forming contacts with the peptide position. (C) A contact matrix between peptide position P-1 and PDZ position βC7; Each amino acid contact frequency is illustrated using color gradients (dark blue shows more frequent contact and light blue shows less frequent contact). (D) Feature vectors extracted from PDZ domain-ligand interactions; the first column contains class labels (1 for the positive class, and -1 for the negative class).

These measurements were defined as the following:

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{FN + TN}$$

$$accuracy\ (\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\ ,$$

where TP, TN, FP and FN were the number of true positives, true negatives, false positives and false negatives, respectively.

RESULTS AND DISCUSSION

Model evaluation

The reliability of the SVM classifier, which includes the cooperative effect in the feature vector, was confirmed by the use of 10-fold cross validation test and was compared with an encoding model that included only the independent contacts representation. During the 10-fold cross validation procedures, the kernel functions and parameters were optimized and the polynomial kernel with d = 3 and a regulation parameter of C = 0.0001 was selected. The prediction accuracies of the SVM classifiers are shown in Table 1. The accuracy of the classifier reached 88.93% when only the independent contribution of contact pairs was reflected. However, the accuracy of the SVM classifier increased to 92.39% when the cooperative contributions of the contact residues were added to the description of the PDZ domain-ligand interaction (Table 1).

Identification of novel PDZ domain-GPCR interactions

Encouraged by the performance of the SVM classifier, we decided to apply this method to discover novel PDZ domain-GPCR interactions. To identify novel interactions between neuronal PDZ domain proteins and GPCRs, which play a central role in the organization of signaling complexes and in neuronal synaptic communication (Fang et al., 2003), PDZ domain proteins were extracted from SwissProt (Release 56.2, Boeckmann et al., 2003) based on the following three conditions. The conditions included a description of "*Homo sapiens*" in the OS line, a description of "Synapse" in the KW line and a description

Table 1. The performance of the SVM classifiers to discriminate PDZ domain-ligand binding from non-binding

	Sensitivity	Specificity	PPV[3]	NPV[4]	Accuracy (%)
Classifier model 1[1]	0.925	0.922	0.923	0.925	92.39
Classifier model 2[2]	0.885	0.894	0.893	0.886	88.93

[1]SVM classifier when both the independent and cooperative contribution of contact pairs was used.
[2]SVM classifier when only the independent contributions of contact pairs were used.
[3]Positive predictive value
[4]Negative predictive value

of "PDZ" in the FT line. Using this procedure, 49 PDZ domain sequences from 20 PDZ domain proteins were obtained, but several redundant domains, which have identical 14 contact residues, were excluded from the test (Table 2). To identify GPCRs that interact with the synaptic PDZ domain proteins, all metabotropic glutamate receptors including their splice variants, which regulate the process of communication from a neuron to the target across a synapse, were adopted and their last four amino acid residues were extracted. Two of 12 metabotropic glutamate receptors, which have the same carboxyl-terminal motifs with others, were excluded from the test.

The SVM classifier was tested on the 410 vectors, which were created by combining the PDZ domains and the last four C-terminal amino acids of the above-mentioned GPCRs for the identification of novel PDZ domain-GPCR interactions. From the 410 PDZ domain-GPCR vectors, 57 vectors were positively classified by the SVM classifier, suggesting that the PDZ domain-GPCR pairs could form constitutive complexes (Table 3). The results were compared with the experimentally derived protein-protein interaction data in the Human Protein Reference Database (HPRD) (http://www.hprd.org; Mathivanan et al., 2006) and the protein microarray data (Stiffler et al., 2007). Among the positively classified vectors, the six PDZ domain-GPCR interactions including PICK1 (GRM3 and GRM7) and GRASP (GRM1, GRM2, GRM3 and GRM5) could be confirmed from the HPRD and one false negative interaction between the SHANK3 and GRM3 could be verified from the protein microarray data (Stiffler et al., 2007). In addition to the six interactions, SHAN1-GRM1 and SHAN3-GRM1 interactions were also confirmed from the SwissProt database. As shown in Table 3, our method correctly classified all experimentally determined neuronal PDZ-metabotropic glutamate receptor interactions in the databases and predicted novel PDZ-GPCR interactions.

CONCLUSION

During the past few years, several high-throughput techniques have been developed for the identification of possible domain-ligand interactions and the techniques have produced a large amount of interaction data. However, the interaction data obtained with the use of high-throughput techniques have resulted in a considerable number of false positives and false negatives. Ultimately, in order to understand better the specificity of domain-

Table 2. Details of the neuronal PDZ domain proteins used for the identification of the PDZ domain-GPCR interactions

PDZ domain proteins	Accession no.	PDZ domains[1]
Disks large homolog 1 (DLG1)[3]	Q12959	3
Disks large homolog 2 (DLG2)	Q15700	3
Disks large homolog 4 (DLG4)	P78352	3
Glutamate receptor-interacting protein 1 (GRIP1)	Q9Y3R0	6[2]
SH3 and multiple ankyrin repeat domains protein 1 (SHAN1)[4]	Q9Y566	1
SH3 and multiple ankyrin repeat domains protein 2 (SHAN2)	Q9UPX8	1
SH3 and multiple ankyrin repeat domains protein 3 (SHAN3)	Q9BYB0	1
Delphilin (GRD2I)	A4D2P6	2
Membrane-associated guanylate kinase, WW and PDZ domain-containing protein2 (MAGI2)	Q86UL8	6
Multiple PDZ domain protein (MPDZ)	O75970	13
Lin-7 homolog A (LIN7A)	O14910	1
Lin-7 homolog B (LIN7B)[5]	Q9HAP6	1
Lin-7 homolog C (LIN7C)	Q9NUP9	1
Golgi-associated PDZ and coiled-coil motif-containing protein (GOPC)	Q9HD26	1
Neurabin-1 (NEB1)	Q9ULJ8	1
Neurabin-2 (NEB2)	Q96SB3	1
Regulating synaptic membrane exocytosis protein 1 (RIMS1)	Q86UR5	1
Regulating synaptic membrane exocytosis protein 2 (RIMS2)	Q9UQ26	1
PRKCA-binding protein (PICK1)	Q9NRD5	1
General receptor for phosphoinositides 1-associated scaffold protein (GRASP)	Q7Z6J2	1

[1]The number of PDZ domains
[2]PDZ domain 3, which has an abnormal structure, was excluded.
[3]Each PDZ domain of DLG 1 has identical 14 residues to the PDZ domains of DLG 2 and 4.
[4]PDZ domain of SHAN1 has identical 14 residues to the PDZ domain of SHAN2.
[5]PDZ domain of LIN7B has identical 14 residues to the PDZ domain of LIN7C.

Table 3. A list of predicted PDZ domain-GPCR interactions

Entry name of GPCR	Ligand sequence[1]	Putative interacting PDZ protein( )[2]
GRM1 (Q13255)	SSTL	SHAN1[4], SHAN2, SHAN3[4], GRD2I(2), GRASP[3,4]
GRM1 (Q13255-2)	HVQL	-
GRM2 (Q14416), GRM3 (Q14832)	TSSL	PICK1, GRASP[3,4]
GRM4 (Q14833)	NHAI	SHAN1, SHAN2, SHAN3, MAGI2(2), DLG1(2), DLG2(2), DLG4(2), MPDZ(2, 3, 9, 10, 11, 13), NEB1, NEB2
GRM5 (P41594, P41594-2)	SSSL	PICK1, GRASP[3,4]
GRM7 (Q14831)	NLVI	GRIP1(3,4,5,6), SHAN1, SHAN2, SHAN3, MAGI2(1,2,4,6), MPDZ (1,2,3,5, 6,7,9,10,11,13), GRD2I(1), NEB1, NEB2, GOPC, LIN7B, LIN7C, PICK1[3,4]
GRM7 (Q14831-2)	PPTV	SHAN1, SHAN2, SHAN3, GRD2I(2), MPDZ(12), GRASP
GRM7 (Q14831-3)	QSNL	PICK1
GRM7 (Q14831-4)	HKED	-
GRM7 (Q14831-5)	CNCY	GRIP1(5), MPDZ(9, 11), NEB1, NEB2

[1]the last four C-terminal amino acids of GPCRs
[2]PDZ domains in a PDZ domain protein that interact with GPCR
[3]PDZ domain-ligand interactions verified by HPRD (Mathivanan et al., 2006)
[4]Protein-protein interactions predicted by SwissProt database

ligand interactions and to identify novel domain-ligand interactions, the development of accurate computational methods and experimental validation must be simultaneously performed.

Ferraro et al. (2006) have developed an accurate machine learning based method for inferring SH3 domain specificity. Although their method did not take into account the cooperative contributions in each ligand position of the SH3 domain-ligand complex, the network of interactions between amino acid residues, which was characterized by 27 domain positions and 10 peptide positions, was still sufficient to describe binding specificity. Based on a similar encoding method, Altuvia and Margalit (2004) developed an accurate threading approach to describe MHC-peptide complexes. However, in the case of PDZ domain-ligand interactions, the relatively short ligand sequence, which participates in PDZ domain recognition, has been the main obstacle in the development of a robust classification method.

In this study, we investigated PDZ domain-ligand binding specificity and developed a computational method for the prediction of novel PDZ domain-GPCR interactions. The method was based on the network of interactions between residues that independently and cooperatively contribute to the stability of the PDZ domain-ligand complex. The SVM classifier constructed with the integration of the network of interactions displayed a high level of performance accuracy and the SVM classifier was applied for the prediction of neuronal PDZ domain-GPCR interactions.

The performance results suggest that the classifier, which integrates a suitable encoding of the interaction interfaces of PDZ domain-ligand binding, can be used to provide information about the nature of PDZ domain-ligand interactions and can be effectively used for the identification of novel PDZ domain-GPCR interactions. Further investigations with the use of the SVM classifier for the identification of PDZ domain-GPCR interactions have the potential to provide a steady foundation for GPCR-directed pharmacotheraphy.

ACKNOWLEDGMENTS

This work was supported in part by the second stage of the Brain Korea 21 Project in 2009 and the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-314-C00379).

REFERENCES

Altuvia, Y., and Margalit, H. (2004). A structure-based approach for prediction of MHC-binding peptides. Methods *34*, 454-459.

Basdevant, N., Weinstein, H., and Ceruso, M. (2006). Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study. J. Am. Chem. Soc. *128*, 12766-12777.

Beuming, T., Skrabanek, L., Niv, M.Y., Mukherjee, P., and Weinstein, H. (2005). PDZBase: a protein-protein interaction database for PDZ-domains. Bioinformatics *21*, 827-828.

Bezprozvanny, I., and Maximov, A. (2001). Classification of PDZ domains. FEBS Lett. *509*, 457-462.

Bockaert, J., Roussignol, G., Bécamel, C., Gavarini, S., Joubert, L., Dumuis, A., Fagni, L., and Marin, P. (2004). GPCR-interacting proteins (GIPs): nature and functions. Biochem. Soc. Trans. *32*, 851-855.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. *31*, 365-370.

Chi, C.N., Engström, Å., Gianni, S., Larsson, M., and Jemth, P. (2006). Two conserved residues govern the salt and pH dependencies of the binding reaction of a PDZ domain. J. Biol. Chem. *281*, 36811-36818.

Day, P., and Kobilka, B. (2006). PDZ-domain arrays for identifying components of GPCR signaling complexes. Trends Pharmacol. Sci. *27*, 509-511.

Doyle, D.A., Lee, A., Lewis, J., Kim, E., Sheng, M., and Mackinnon, R. (1996). Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell *85*, 1067-1076.

Eo, H.S., Choi, J.P., Noh, S.J., Hur C.G., and Kim, W. (2007). A combined approach for the classification of G protein-coupled receptors and its application to detect GPCR splice variants. Comput. Biol. Chem. *31*, 246-256.

Fang, M., Tao, Y.X., He, F., Zhang, M., Levine, C.F., Mao, P., Tao, F., Chou, C.L., Sadegh-Nasseri, S., and Johns, R.A. (2003). Synaptic PDZ domain-mediated protein interactions are disrupted by inhalational anesthetics. J. Biol. Chem. *278*, 36669-36675.

Ferraro, E., Via, A., Ausiello, G., and Helmer-Citterich, M. (2006). A

novel structure-based encoding for machine-learning applied to the inference of SH3 domain specificity. Bioinformatics *22*, 2333-2339.

Giallourakis, C., Cao, Z., Green, T., Wachtel, H., Xie, X., Lopez-Illasaca, M., Daly, M., Rioux, J., and Xavier, R. (2006). A molecular-properties-based approach to understanding PDZ domain proteins and PDZ ligands. Genome Res. *16*, 1056-1072.

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. Science *185*, 862-864.

Hanyaloglu, A.C., and von Zastrow, M. (2008). Regulation of GPCRs by endocytic membrane trafficking and its potential implications. Annu. Rev. Pharmacol. Toxicol. *48*, 537-568.

Hung, A.Y., and Sheng, M. (2002). PDZ domains: Structural modules for protein complex assembly. J. Biol. Chem. *277*, 5699-5702.

Joachims, T. (1999). Making large-scale SVM learning practical. Advances in kernel methods-support vector machines (MIT-Press).

Kreienkamp, H.J. (2002). Organisation of G-protein-coupled receptor signaling complexes by scaffolding proteins. Curr. Opin. Pharmacol. *2*, 504-514.

Kurakin, A., Swistowski, A., Wu, S.C., and Bredesen, D.E. (2007). The PDZ domain as a complex adaptive system. PLoS ONE *2*, e953.

Mathivanan, S., Periaswamy, B., Gandhi, T.K., Kandasamy, K., Suresh, S., Mohmood, R., Ramachandra, Y.L., and Pandey, A. (2006). An evaluation of human protein-protein interaction data in the public domain. BMC Bioinformatics *7*, S19.

Nourry, C., Grant, S.G., and Borg, J.P. (2003). PDZ domain proteins: Plug and play!. Sci. STKE. *179*, RE7.

Reina, J., Lacroix, E., Hobson, S.D., Fernandez-Ballester, G., Rybin, V., Schwab, M.S., Serrano, L., and Gonzalez, C. (2002). Computer-aided design of a PDZ domain to recognize new target sequences. Nat. Struct. Biol. *9*, 621-627.

Skelton, N.J., Koehler, M.F., Zobel, K., Wong, W.L., Yeh, S., Pisabarro, M.T., Yin, J.P., Lasky, L.A., and Sidhu, S.S. (2003). Origins of PDZ domain ligand specificity. J. Biol. Chem. *278*, 7645-7654.

Songyang, Z., Fanning, A.S., Fu, C., Xu, J., Marfatia, S.M., Chishti, A.H., Crompton, A., Chan, A.C., Anderson, J.M., and Cantley, L.C. (1997). Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science *275*, 73-77.

Stiffler, M.A., Chen, J.R., Grantcharova, V.P., Lei, Y., Fuchs, D., Allen, J.E., Zaslavskaia, L.A., and MacBeath, G. (2007). PDZ domain binding selectivity is optimized across the mouse proteome. Science *317*, 364-369.

Vaccaro, P., and Dente, L. (2002). PDZ domains: troubles in classification. FEBS Lett. *512*, 345-349.

van Ham, M., and Hendriks, W. (2003). PDZ domains-glue and guide. Mol. Biol. Rep. *30*, 69-82.

Wiedemann, U., Boisguerin, P., Leben, R., Leitner, D., Krause, G., Moelling, K., Volkmer-Engert, R., and Oschkinat, H. (2004). Quantification of PDZ domain specificity, prediction of ligand affinity and rational design of super-binding peptides. J. Mol. Biol. *343*, 703-718.

Zhang, L., Shao, C., Zheng, D., and Gao, Y. (2006). An integrated machine learning system to computationally screen protein databases for protein binding peptide ligands. Mol. Cell. Proteomics *5*, 1224-1232.