# PROTEIN CLASSIFICATION USING FUNCTIONAL MOTIFS AS SUB-STRUCTURES OF ACTIVE SITE

**AHMAR FARAZ**

*(M. Comp, NUS; B. CS, NUCES)*

**A THESIS SUBMITTED FOR THE DEGREE OF MASTER OF SCIENCE**

**SCHOOL OF COMPUTING**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2005**

## Acknowledgments

I would like to thank David Hsu and Tan Kian Lee for their guidance and support through out the research. They had taken valuable time to read my draft and suggested corrections. I am grateful to them for their sharing of the research ideas and providing insight to research by sharing valuable experiences.

**Table of Contents**

## Summary

Discovering the functionally important regions in the proteins is an important problem in computational biology. Several different techniques have been used for this purpose. Techniques based on the primary structure of the protein perform sequence comparison to identify the functionally important regions. Sequence based techniques cannot be applied for discovering these regions when the sequence similarity fall below a certain threshold. Evolutionary techniques generate consensus sequences to identify the common ancestor of the protein and use the consensus sequences to identify these regions. Structure based techniques consists of many different approaches of using the structural information to identify these regions. Some techniques use the physical properties of the protein complexes while others use the molecular surface for this purpose. The limitations of the structure based techniques are that they fail to use the structure of the active site for identifying the functionally important regions. We present a new technique that uses the structure of the active site for identifying these regions. The new technique explains the binding of the proteins in terms of the structure of the active site.

The new technique is then applied to perform the functional classification of the proteins. In contrast to other structural techniques that are available for functional classification the presented technique explains the reason for functional classification in terms of the structure of the active site. The new technique performs accurate classification of the proteins when compared to the sequence and structure based techniques for the functional classification.

**List of Figures**

**List of Tables**

# 1 Introduction

## 1.1 Motivation

Classification of functionally related proteins shows similarities and differences in their role and behavior during the cellular processes. Proteins with similar functions are involved in the same kind of processes, giving scientists more details about the general relationships that exist between different processes happening inside the cell. Since proteins form the essential parts of the cell of all living organisms, understanding their functions can ultimately lead the way to understand the overall structure and functions of the cellular organelles.

Just like understanding the structure and functions of the cell requires the knowledge of functions and structure of the protein, comprehending the structure and functions of the protein requires the insights of primary, secondary and tertiary structural details as well as the role a group or groups of residues play in determining the function of the protein. Residues are the amino acids that a protein is made up of. These small groups of residues that play an important role in forming either structural details or functional details of the protein are called motifs. Discovering these motifs in different proteins provides insight about their structure and functional details. Functional motifs are group of residues which are directly related to the function of the protein, discovering them in different proteins provides information about their functional linkage. Hence functional motifs can be ultimately used for functional classification of proteins.

Functional classification of the proteins is an important problem in the area of computational biology and it has numerous applications from drug discovery to

understanding the pathways in which proteins are involved. Motif based functional classification can also be used to study the interactions of the protein at a more detailed level. In addition, functional classification can be used to understand the role that the structure of the protein plays in determining its function.

## 1.2 Thesis Contribution

In this thesis, a new structure based technique for discovering functional motifs of a protein is presented. This technique is then utilized to perform structure based functional classification of the protein. This technique presents the solution to the problem of functionally classifying proteins when their structure is known.

Contrary to other structure based techniques that use the molecular surface for identifying the functional motif; the presented technique uses the structural information of the active / interaction sites for this purpose. The main difference between the newly presented technique and other structure based techniques that also use structural information of the active sites is that instead of comparing the overall active site shape or structure of one protein to another, the new technique finds the smaller parts (hence forth referred to as "sub structure") of active site which lie in structural proximity. Structural proximity is defined in terms of specific Euclidean distance threshold between two points in the space. Once these sub structures are detected from the active sites of protein, the technique searches for these sub structures in the active sites of other proteins. These sub structures are functionally important since they are making the active sites of the proteins and same kind of sub structures can be found in the active sites of the proteins having similar functions.

It should be noted that functional motif, functionally important region and sub structure refer to the same thing and will be used interchangeably in the subsequent sections of this thesis.

## 1.3 Organization of the thesis

In the initial section of the thesis, a brief introduction to the techniques of functional motif discovery and general concepts about protein structures and functions are discussed. A brief overview of the functional classification and its importance is also provided. Chapter 2 covers the traditional techniques for functional motif discovery and discusses their limitations. Chapter 3 discusses the new technique and its implementation details. Although currently the technique is used for the discovery of protein functions, it can be applied to lots of different applications that use the structural to functional relationships. Chapter 4 discusses the results which have been achieved using the new technique, it also covers a case study of the G-protein coupled receptors and compares functional classification using bottom up clustering, top down clustering and the newly presented technique. Chapter 5 covers some of these applications where the newly developed technique might be useful in near future. Conclusion and references are also provided in this last section.

# 2 Functional Motifs discovery and Functional Classification

## 2.1 Traditional Computational Techniques for Motif discovery

Researchers in functional genomics area have applied several different kinds of computational techniques to identify the functional sites and to classify proteins based on their functional linkage. We can broadly divide these computational techniques into three major categories, namely Sequence based, Evolutionary relationship based and Structure based.

Sequence based techniques are applied usually when no structural information is present. Although the primary structure of the protein cannot give as much functional detail as the secondary or tertiary structure of the protein, recurring patterns and directly or indirectly related subsequences give some information about the functionally important regions of the proteins. Recurring patterns are the smaller subsequences in the primary structure of the proteins that occur repeatedly. A directly related pattern refers to a subsequence that occurs only if another particular subsequence is present or absent. Similarly indirectly related patterns refer to subsequence that is present when a group of related subsequences are present or absent. The subsequences inside these groups may also be related to each other through direct or indirect relationship.

Sequence based techniques normally use sequence alignments to detect functionally important regions. In the situation where sequence alignment cannot be found, or when the sequence alignments cannot give any particular information about functionally important regions, evolutionary relationships between different proteins are searched. A common ancestor protein can be used to infer functional linkage between two different

proteins. Structural based techniques for functional classification uses the structure related characteristics like the type of domains a particular protein is made up of or a particular patch of the molecular surface that resembles the surface of other protein of known function.

| Type of the technique | Examples |
|---|---|
| Sequence Based | Multiple Sequence Alignment (MSA) [4], Correlated Mutations [5] |
| Evolutionary relationship based | Evolution Trace Method [7], Phylogenetic Profiles [8] |
| Structure based | Geometric Hashing [11], Surface Patch Analysis [10], Fused domains [9] |

**Table 1 Functional motifs discovery techniques**

### *2.1.1 Sequence based Techniques*

Traditionally sequence alignments have been used to identify the homologous sequences [2]. Homologous sequences are derived from a common ancestor but are found in different proteins. The concept of aligning sequence can be extended to multiple sequences by aligning the sequences to the alignment information of two already aligned sequences. Multiple Sequence alignments (MSA) of different proteins can be used to discover the patterns that are conserved in them. In multiple sequence alignment conserved residues are the ones that are present most frequently in the same position in the alignment (with gaps allowed). Examples of conserved residue in multiple sequence alignment are given below:

AA**TLT**AL-

-A**TLT**GVM

LL**TLT**VVM

In the above MSA the conserved pattern is TLT since it is aligned at the same position of all three proteins. Conserved patterns in multiple sequence alignment can infer the functionally important regions of these proteins [3].

According to theory of co-evolution of proteins, if one protein mutates, other proteins that interacts with this mutated protein have to undergo compensatory mutations in order to sustain their functional linkage with this particular protein. Although multiple sequence alignments provide information about the functionally important regions of the proteins, it cannot provide information about the regions that might affect the functionality of the protein by taking part in large number of interactions. Changes in these kind of regions will initiate compensatory mutations in the interacting partners of that particular protein. Residue Correlation Analysis (RCA) is proposed to study these regions [4-5]. Residue correlation analysis consists of two main steps, during the first step the pairs of residue positions are identified where mutations occur in coordinated way. The second step uses these residue positions to identify the regions that are involved in large number of interactions. An example illustrates the concept of coordinated mutation is given below.

Consider the following multiple sequence alignment:

| Before mutation | After mutation |
|---|---|
| LM**S**ALP**G** | LM**D**ALP**K** |
| GM**S**ATV**G** | GM**D**ATV**K** |

Consider the position 3 (from left to right) and position 7; it is obvious that when the residue at position 3 changes, residue at position 7 changes as well, which means that they are related to each other. After finding the residue positions where coordinated mutations are occurring, RCA tries to find out such a contagious set of these positions that change in a coordinated fashion. A set of these contagious positions where mutations occur in coordinated manner represents the functionally important region in the protein.

### *2.1.2 Evolutionary Relationships based techniques*

Though multiple sequence alignments of the related proteins can give information about the functionally important regions, inferring any functional information becomes less possible when the sequence homology falls below 25% [6].

A functional interface of the proteins refers to the part of the tertiary structure of the protein that interacts with the ligands or other proteins during complex formation. Changing the structural details of this functional interface of the protein has severe effects on the functionality of the proteins. The shape similarity of this interface provides information about the similarity of the functions, while shape complementarity of two functional interfaces of different proteins predicts their interactions.

As theory of co-evolution of proteins suggests that the proteins either have to maintain their functional interface or have to undergo compensatory mutations, thus most of the proteins are under constant evolutionary pressure to maintain their functional interfaces.

As a result the residues that are involved in defining the functional interfaces of the proteins undergo fewer mutations than the other residues. [3]

Evolutionary based methods use ancestor linkage to determine the functionally important regions. One such method is Evolutionary Trace (ET) method. Evolutionary trace method uses the concept of coordinated mutations from theory of co-evolution of protein to trace back the functionally important regions. In the absence of any structural and functional information of the protein sequences, the evolutionary trace method partitions the sequences into clusters according to their identity. After consensus sequence is generated for every cluster and then these consensus sequences are aligned to trace their evolution. In the case when a particular consensus sequence cannot be aligned with other sequences, the mutations are considered specific to that particular cluster [7]. The step-by-step details of the ET method are as follows:

*Step 1: Partitioning of the Sequences according to their identity*

| | | |
|---|---|---|
| **LM**S**A**LP**G** | **VTRAG**VM | **TF**G**A**E**RS**V**L** |
| **LM**S**A**LV**G** | **VTRAG**VT | **TF**G**A**E**RS**K**L** |
| **LM**S**A**TV**G** | **VTRAG**VV | **TF**K**A**E**RS**A**L** |
| **LM**V**A**GV**G** | **VTRAG**TK | **TF**G**A**V**RS**V**L** |

*Step 2: Generation of the Consensus Sequences for the above partitions*

| | | |
|---|---|---|
| LM_A__G | VTRAG__ | TF_A_RS_L |

*Step 3: Alignment of the Consensus Sequences*

LM_A__G

VTRAG__

TF_A_RS_L

*Step 4: Results of the alignment*

The following observations can be made from the alignment of the consensus sequences:

Residue A is found at position 4 in most of the proteins.

Mutations at Position 3 (from left to right) and positions 5 and 6 are occurring in many proteins. Mutations at these positions relate to functionally important region.

Mutation at position 7 is specific to particular cluster (cluster 2); it may or may not be related to the functionally important region.

At one glance, it seems that both multiple sequence alignment and evolutionary trace method are similar since both use sequence based representation for identification of functionally important regions. The major difference between these two techniques is that multiple sequence alignment technique performs sequence comparison on a given protein whereas the evolutionary trace technique searches for common ancestor (i.e. concensus sequence) by extracting those residues that are the same and reside at the same position for all the sequences. These concensus sequences are then used for sequence comparison to identify the functionally important regions.

Another technique that uses the evolutionary relationships between proteins to discover functionally important regions is phylogenetic profiles. A phylogenetic profile is a string with n entries (where n represent the number of genome present in the organism), if a protein is transcribed by a particular genome, a value of 1 is given to that entry; otherwise zero value is substituted [8]. The same phylogenetic profile of two homologous proteins infers that they are functionally related to each other.

| | G1 | G2 | G3 | G4 | G5 | G6 | G7 | G8 |
|---|---|---|---|---|---|---|---|---|
| **P1** | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| **P2** | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| **P3** | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 |
| **P4** | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| **P5** | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| **P6** | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| **P7** | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

**Table 2 Phylogenetic Profiles of different proteins**

In the above table P1, P2... P7 represents the different proteins while G1, G2… G8 represents different genomes. As highlighted in gray background, the phylogenetic profiles of P1, P2 and P4 are similar and thus inferring a functional linkage among these proteins.

### *2.1.3 Structure based Techniques*

Few years back, determining the structure of the protein was a complex and tedious task, but now due to the advancements in the techniques of X-ray crystallography and Nuclear Magnetic Resonance (NMR) it has already become a routine task. As we know from basic biochemistry that the structure of the protein has a direct relationship to its function, so recently computational approaches have been devised to use the structural information of the proteins for the discovery of functional motifs and their functional classification.

One of the famous techniques in this regard is the study of fused domains (i.e. domains which exist separately in the other proteins but act as a single domain in other) in proteins. Certain protein families contain fused domains, but these domains exist as stand alone protein in some other protein families. These fused domain proteins are known as composite proteins (also called fusion proteins) while the proteins that only contain part of this fused domain are called component proteins. If the two component proteins of the same composite protein exist, a functional linkage is predicted among the components. Although the technique can be successfully applied to predict indirect functional linkage (indirect refers to functional linkage in a same protein pathway) but fails to predict functional relationships when physical interaction between two proteins are involved. [9].

Other structure based techniques uses the properties of the surface patches of the protein complexes. A patch is defined as the central surface accessible residue with n nearest surface accessible neighbors, where n is the number of the residues that are observed in the interface. Some of the properties studied by these techniques include the solvation potential, residue interface propensity and protrusion index. Solvation potential defines the tendency of the amino acid type for salvation and is normally approximated by the residues solvent access surface area (ASA). Residue interface propensity refers to the fraction of solvent access surface area that an amino acid contribute to the functional interface compared with the fraction of the access surface area it contributes to the molecular surface. Protrusion index defines the absolute value of the extent to which the residue protrudes from the surface of the protein. A detailed description of these properties as been observed in different kind of protein complexes is given below [10].

| Type of Complex | Solvation Potential | Interface Propensity | Access Surface Area (ASA) | Protrusion Index |
|---|---|---|---|---|
| Homodimers | Low | Low | *No specific trend* | High |
| Hetrodimers | High | High | High | *No specific trend* |
| Enzyme-Inhibtor Complex | *No specific trend* | *No specific trend* | High | *No specific trend* |
| Antigen-Antibody Complex | *No specific trend* | Low | Highest | Highest |

**Table 3 Tendencies of protein complexes based on physical properties**

The above table explains the trends of different properties of the patches as observed in the different types of complexes. In case where no specific trend is found, that is for some complexes of the same type the value of the property is high and for some other complexes is low. So the type of the properties found in a particular protein complex can be used to get ideas about its function. The limitation of the above technique is that it can only predict the type of the protein complex with some certainty but it cannot be applied for functional motif discovery in the proteins [10].

A more detailed structural technique works by using the molecular surface comparisons, if a patch of the surface in one protein is found similar to active site surface in another protein, functional linkage is predicted. It uses geometric hashing on the molecular surfaces in three steps, i.e. molecular surface representation, geometric hashing and clustering and extension by reapplying the geometric hashing. In the initial stage of geometric hashing, transformation invariant features (i.e. features that are not changed due to transformation of a molecule surface) are extracted from the protein and saved in the hash table. In the second stage the transformation invariant features are calculated for

the target protein and are used to access the hash table to find the possible instances of the model. In the last stage of geometric hashing, similar transformations are clustered based on the Euclidean distance between them. The limitation of this technique is that it uses the molecular surfaces to search for active sites and functional similarities but it does not take into account the structural information of the active sites [11].

## 2.2 Limitations of the traditional structure based techniques

Three different kinds of structure based techniques are discussed above, the first technique which uses the composite and component proteins can not be used to discover functional motif, the reason is that the technique generates much more false positive results when applied to discover direct physical interaction between the proteins, however the technique works fine for detecting the indirect functional interaction (e.g. two proteins that exists in the same pathway) between different component proteins [9]. The second technique that studies the physical properties of the surface patches is limited to the prediction of type of the protein complex with some certainty [10]. Finally using geometric hashing on the active site surfaces can be used to predict the functionally important regions, but this technique does not consider the structural information of the active sites for the detection of the functionally important regions [11].

## 2.3 Traditional techniques of functional classification

### *2.3.1 Bottom up Clustering using sequence similarity (ProtoMap)*

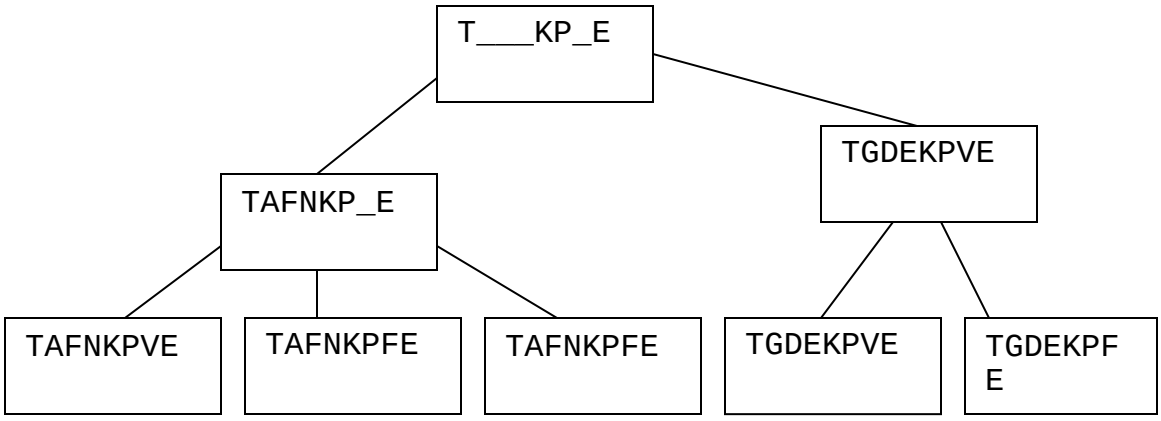
Given a new protein sequence the most common approach to predict its function is based on the pairwise comparisons with the sequences of known function. In the below discussed method [21], the functional classification of the proteins is done based on sequence similarities through the use of bottom up clustering of the protein sequence.

In this procedure the proteins are represented as weighted graph and their sequences are the vertices. The weight of the edge between two sequences corresponds to their degree of similarity. Blosum 50 and Blosum 62 are used as scoring matrices to measure the degree of similarity between two sequences. Related proteins are identified by the strongly connected sets of vertices in the graph. The process is repeated at varying thresholds and the proteins are grouped together into classes and results into forming of a tree that represents a hierarchical organization of all the proteins.

**Figure 1 Bottom up clustering using sequence similarity**

### *2.3.2 Top Down clustering of the proteins*

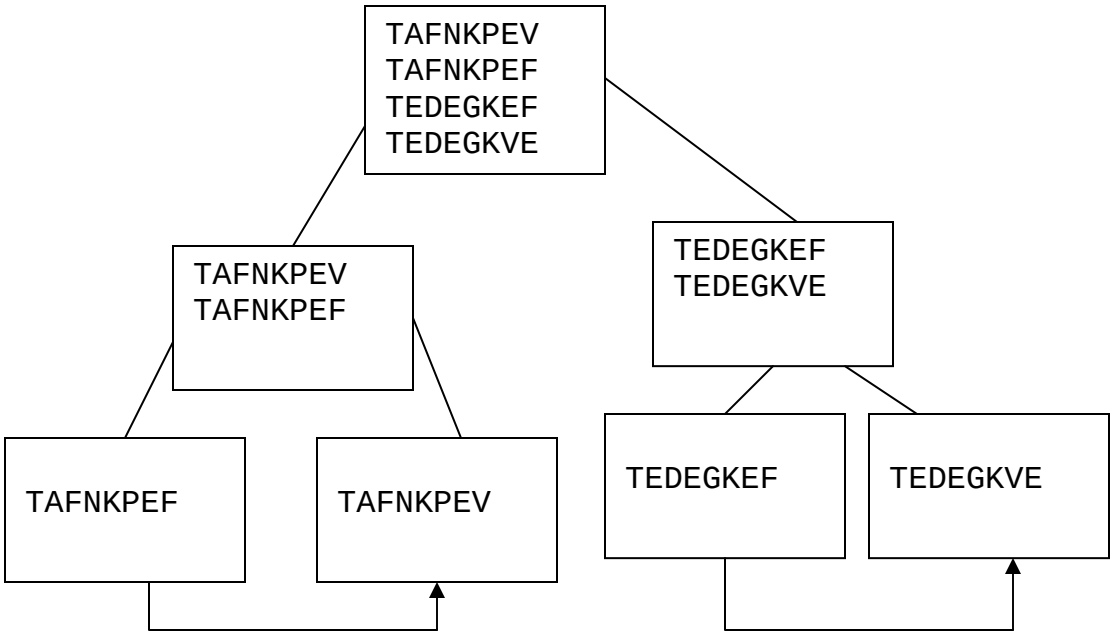
The limitation of the bottom up clustering to classify protein is that only related proteins




can be classified. During evolution non-hierarchical relationships are formed as a result of domain shuffling in the proteins. Also non-hierarchical relationships provide more detailed level insight for the functional classification of the proteins. Since the bottom up clustering method is based on developing a graph to represent hierarchical relationships only, it cannot use the non-hierarchical relationships for functional classification of the proteins.

The top down clustering method begins by putting the proteins into common super family and then splitting the super family into many sub families on the basis of similarity in the sequence. In this way a hierarchical tree structure is obtained. Non-hierarchical relationships are searched by discovering small identical regions among the families and the tree like structure is modified into tree graph like structure in order to capture these similarities at the partial domain level [22].

**Figure 2 Top down clustering of the proteins for functional classification**




## 2.4 Limitations of the primary structure based functional classification schemes

Since the above-mentioned techniques use the primary structure of the proteins for functional classification, and primary structure only provides one-dimensional information of the protein structure, they are not able to classify the remotely related

proteins (i.e. proteins that are different in their primary structure but similar in their three-dimensional structure). The suggested technique in this thesis is based on three-dimensional structure and can study the role of smallest substructures to larger substructures inside the active site.

## 2.5 Structure based functional classification

As mentioned above the primary structure based techniques cannot perform functional classification of the proteins efficiently, thus there is a need for three-dimensional structure based classification. One such classification uses the microenvironment of the active site for this purpose [24].

In this technique the active site is defined as a region within the protein molecule with a surrounding neighborhood of 10Å radius. The spatial distribution of user defined properties like types of atoms, chemical groups, amino acids, secondary structure, charge, polarity etc is calculated in this neighborhood of the active site. The microenvironment of the active site is computed by dividing the volume of the site into concentric shell sub volumes and then calculating the distribution of these properties within each of these sub volumes. These distributions are calculated for both the active and non active sites, and are saved in the places where the distributions of these properties differ significantly between the active site and non active site. These properties are then used to predict the binding tendency of the protein, thus predicting the function of the protein if the structure is known.

## 2.6 Limitations of structure based functional classification

Although structure based functional classification can correctly classify the proteins that are remotely related to each other in their primary structure, it has its own limitations. The structure based technique that uses the microenvironment of the active site to predict binding does not provide any sub structural details of the active site, thus not explaining the reason of binding in terms of the active site. This limitation is overcome in the newly propose technique as it explains the sub structure details of the active site and provides insight to the binding in terms of active site structure.

# 3 Proposed Technique, Problem definition & Solution

As discussed earlier the existing structural technique does not utilize the structure of the active site, so in this thesis, a new technique is proposed which utilizes the structural information of the active sites to detect the functionally important regions.

Instead of comparing the overall structural information of the active sites, the technique compares the small sub structures that are present in the active site. The rationale behind comparing the small sub structures instead of comparing the overall active site is that although during the course of evolution the active sites of the proteins may change, but as the theory of co-evolution of the proteins suggest that the functionally important regions are under evolutionary pressure to retain themselves, so although slight changes can occur in the active sites of the proteins during the evolution but their detail sub structures remain preserved.

## 3.1 Formal Definition of the problem

**Hypothesis:**

If two proteins have the similar sub structures in the active sites, then they will be having the same functions. This hypothesis is constrained within the context of protein binding.

**Formal Definition:**

Formally the problem of finding sub structures from the active site can be defined as follows:

*Find all the possible sub structures containing three or more residues of the active site that come in contact of any fragment of the ligand during the process of docking and lie*

*in structural proximity of each other. Use these sub structures to perform functional classification of proteins.*

Fragment Based Complex Construction technique breaks the ligand into multiple fragments. During the process of docking (i.e. the process of attaching one molecule to another), different fragments of the ligand come into the contact of different residues of the active site on a protein. The ligands are then reconstructed in different ways and an energy function for the ligand is calculated for each reconstruction. The reconstruction that is giving the minimal energy function will be used as the experiment data for this thesis.

Sub structure / functional motif is defined as part of a protein that consists of three or more residues that come into contact of the ligand fragments and at the same time lie in the structural proximity of each other.

**Assumption**: Protein classified manually based on literature reviews and further supported by experimental methods are correct and will be used to verify against the results obtained based on computational methods.

## 3.2 Solution

Ideally, the experiment should be conducted and verified using the fragment based docking method. But since performing fragment based docking between different ligand and proteins is a tedious task, a slightly different way is adopted for the rapid implementation of the technique. Current implementation of the technique uses ligand-

protein complexes that are readily available and extracts the sub structures in the active site from these complexes.

Protein Data Bank has been identified as the source where experiment data is extracted from due to the following reasons:

1. It contains the ligand-protein complexes that exist in nature,
2. It provides the structural details of the protein and ligand-protein complexes,
3. It contains a huge collection of data that makes a good sample for this experiment,
4. Using the readily available structural details reduces the time of development of this technique,
5. It also provides the experimentally verified information of the way a ligand can dock onto the protein [1].

In order to obtain the information from Protein Data Bank (PDB), an application has been implemented based on biopython PDB module. This application extracts the atomic coordinates from the PDB flat file. Another application has also been implemented that filters the PDB and extracts all the ligand-protein complexes. The active sites of ligand-protein complexes are extracted via an interaction system. This interaction system calculates the atomic distances between the molecules of proteins and ligands. The core analysis is then performed on the results of active sites extracted. This is illustrated in the following figure.

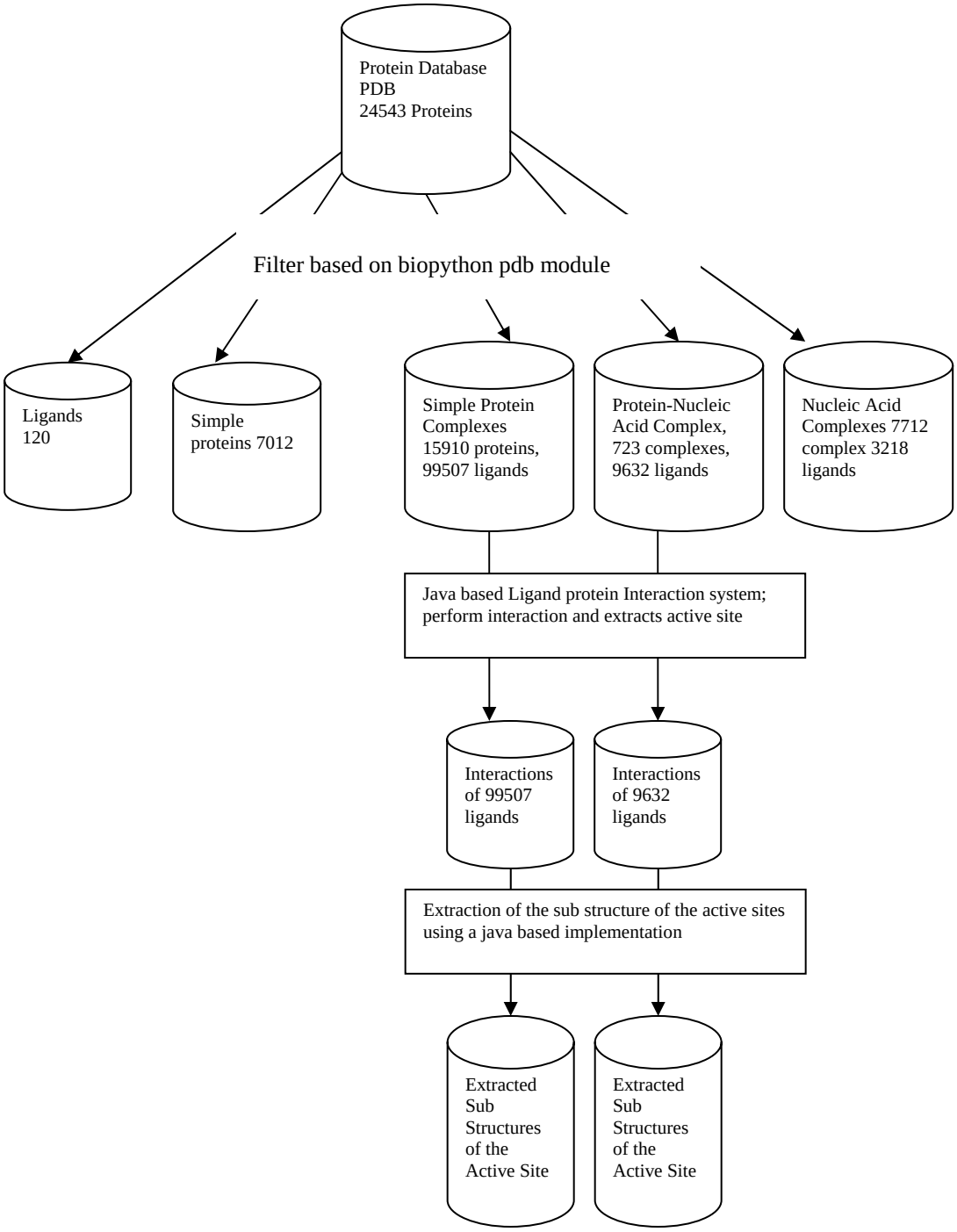



**Figure 3 Distribution of PDB and steps in extracting the active site**

The implementation of the core analysis consists of 2 main stages. During the first stage, protein complexes (i.e. proteins that have already been classified by nature) are analysed to determine their sub structures. The second stage is to perform classification of a new

protein by comparing sub structures of its active sites to the sub structures obtained from stage one of a classified protein.

### *3.2.1 Identify Sub Structures for Classified Proteins*

The first stage is further separated into 3 main steps. During the first step, the ligand-protein interface is analyzed and the residues that are forming contact with ligand fragments are extracted. In the second step, sub structures are extracted from the active site. In the third step, the significant functional motifs for a particular ligand are calculated. All of these steps are described in detail below.

**Step 1: Analysis of Ligand Protein Contact**

In order to find the residues coming in contact with the ligand, Ligand Protein Contact (LPC) analysis is used. Ligand protein analysis calculates the contact surface area between two atoms A and B by placing another atom of the Van der Waals radius double than the radius of A at the center of A. Now if this atom penetrates or touches the atom B then atom B is considered in contact with atom A [17]. As illustrated in the following figures, the one on the left shows that B comes in contact with A whereas the figure on the right shows that B does not come in contact with A.

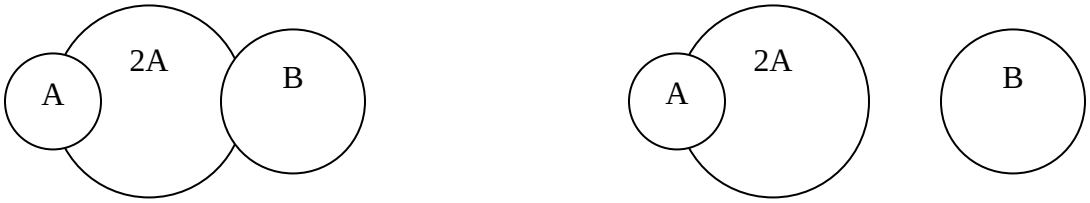



**Figure 4 Ligand Protein Contact Analysis**

Ligand protein contact analysis is performed on all the protein complexes (16633) of the protein database. These complexes contain 109139 ligands in total. An interaction system is developed which performs the interactions of these 109139 ligands with 16633 complexes and store the contact atoms. These contact atoms are then used in the extraction of sub structure from the active site.

For example in the case of RUB binding proteins, interactions are calculated between the ligand RUB and the protein 9rub and the residues that come in contact with the ligand are 111A, 164B, 287B, 288B, 321B, 322B, 323B, 368B, 369B, 391B, 392B where these are the residue numbers of the contacted atoms as specified by the PDB. Each of these residue numbers corresponds to a residue, e.g. GLY, LYS etc.

**Step 2: Extraction of Sub Structures**

Whenever a particular residue comes in contact with the ligand, information about that particular residue and the residues that are in neighborhood and also come in contact with the ligand is saved. In order to calculate the neighborhood, a Euclidean distance threshold is used. A Euclidean distance threshold can hold the value from the range of 7-12 Å. Residues that come in neighborhood are defined as residues with Euclidean distance threshold less than 8.5 Å and 10 Å are used for the experiment. According to this theory, a graph is created that represents residues of the active site that are coming in contact with the ligand and also are in neighborhood of each other.

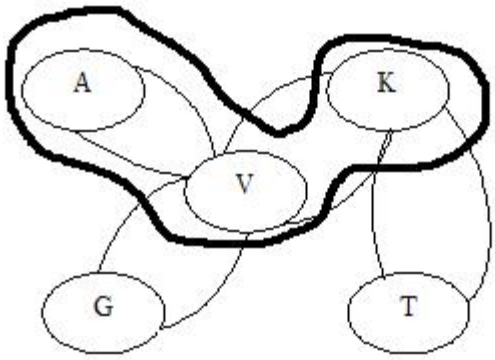



**Figure 5 Sub structure graph**

Figure 5 illustrates five residues that come in contact and are lying in the structural proximity of 8.5 Å or 10Å. One such sub structure that can be extracted from these residues is AVK. Other sub structures that can be extracted are AVG, VKT, KVG, AVGK, GVKT, AVKT and AVCKT. These sub structures are obtained by forming different combinations of the residues that are having edges drawn between them. Since by definition, sub structure is part of a protein that consists of three or more residues that come into contact, AV cannot be used as a sub structure even it has edges drawn between the residues. It should also be noted that AVG and VAG or AGV refer to the same sub structure.

Continuing from the example of RUB binding proteins given in the first stage, a sub structure graph is created for the residues correspond to the residue numbers returned from PDB. The sub structures that are extracted using the sub structure graph are given as follows, where GLY, LYS etc represent residue, and "---" illustrates an edge

connecting two residues that are having Euclidean distance threshold of equal or less than 8.5 Å.

| |
|---|
| GLY --- GLY --- GLY --- GLY --- LYS --- PHE --- SER --- THR --- THR |
| GLY --- GLY --- GLY --- GLY --- LYS --- PHE --- THR --- THR --- TRP |
| GLY --- GLY --- GLY --- LYS --- THR --- TRP |
| GLN --- GLY --- GLY --- GLY --- GLY --- PHE --- SER |

**Table 4 Sub-structures of RUB binding proteins**

**Step 3: Calculation of significant functional motifs for a ligand**

During this step the statistically significant functional motifs are calculated for every ligand. Motifs which are found common in most of the ligand binding classes are removed, also only distinct motifs are used for the calculation. In order to understand the significance of a particular set of functional motifs the criteria of confidence is applied. Those motifs which are having the confidence of more then Y %(where Y can be 40%, 50% or 60%) for the particular ligand binding class are selected.

### *3.2.2 Functional Classification of Proteins*

In the second stage, the significant sub structures / functional motifs that are calculated in first stage for the classified proteins are applied to perform functional classification of the unclassified proteins. This stage includes 3 steps, as described.

**Step1: Identification of Active Sites**

Computed Atlas of Surface Topography of Proteins (CASTP) program is used to determine all the active sites of an unclassified protein. For details on how CASTP works, refer to [23].

**Step2: Extraction of Sub structures**

Similar to step 2 of first stage, every residue and residues that come in the neighbourhood are used to form sub structure graph. Sub structures are then extracted by forming three or more residues that come in the structural proximity. This is conducted for all the active sites identified by CASTP.

**Step3: Classification of Protein**

For all the sub structures extracted in the active sites of the unclassified protein, they are compared against existing sub structures (obtained through stage 1) of all protein complexes. If the sub structures are found similar to X % of the sub structures of protein complexes, then the protein is classified as having the same binding tendencies of those protein complexes.

In order to determine the suitable value for the X % K fold cross validation procedure is applied. The experiment conducted in this thesis uses the value of K =10. This means that the sub structures data (training data) obtained is divided into 10 sets. 9 out of 10 sets of the data are used for training and remaining 1 set of data is used to calculate the suitable value of the X%. This is repeatedly conducted for 10 times, each time rotating 9 sets of data for training and 1 set to calculate the value of the X%. The value found suitable when the confidence value of 50% is used for the sub structure extraction and K fold cross validation is applied for the matching % (i.e. X %) is 45 (i.e. if an unclassified protein contains sub structure that found similar to more than 45% of the sub structures of a protein complex then it is classified as having the same binding tendencies as those of the protein complexes. As it is found that the classification of the proteins can be performed better if the matching percentage is selected as 45%.

Sub structures of unclassified protein are considered as matching to the sub structures of protein complexes as long as the type and number of residues that made up the sub structure tallies. The sequence of these residues does not have to be matched exactly. E.g. "GLY --- GLY --- GLY --- LYS --- THR --- TRP" and "GLY --- LYS --- GLY --- THR --- GLY --- TRP" are considered as matching sub structures because both sub structures are composed of 3 GLY, 1 LYS, 1 THR and 1 TRP residues. Biologically these substructures forms the same functional motif in the active site, so although the internal geometry of the functional motif can be different from each other in terms of the residues, they still form the same motif, as the similarity of the functional motif defines two motifs as similar if their residue composition is similar [18].

Once all the matching sub structures of unclassified protein are identified, the protein is classified according to those protein complexes that have matching sub structures.

The following provides a summary of the entire program in pseudo language:

**Stage 1**

**FilterProteinComplexes**

```
For each record of PDB
If Hetatm is found,
        The record is a protein complex, and contains ligands
        Save this protein record and its atomic details into a file
        Save ligand and its atomic details into a file
```

**ExtractActiveSites**

```
For i=1 to n protein complexes
        For j=1 to m ligands
                Calculate atomic distance between protein complex i and ligand j
                If distance is less than 1Å,
                        Save active site information into a file
```

**ExtractSubStructures**

For each active site file
    For each residue of an active site
        If Euclidean distance threshold of the residue is less than 8.5 Å or 10 Å
            Add this residue to the sub structure graph
    Extract sub structures from the graph by forming three or more residues
    Calculate the confidence for motif
    If Confidence > Y % (Y can take the values like 40%, 50%, 60% etc)
        Save sub structures into a file

**Stage 2**

Given an unclassified protein, use CASTP to extract its active sites and save into a file
For each active site,
    For each residue of an active site
        If Euclidean distance threshold of the residue is less than 8.5 Å or 10 Å
            Add this residue to the sub structure graph
    Extract sub structures from the graph by forming three or more residues
    Save sub structures into a file

Compare each sub structure in the file to sub structures obtained in stage 1
If the substructures match is more then 45%
    Classify the protein similar to the ligand binding class

For Performing Multi-Class classification
Convert the Multi-Class classification problem into series of binary classification problems
Process the binary classification problems and use voting to determine the class membership of the unclassified protein

# 4 Results & discussion

The experiment conducted in this thesis uses a sample of 16642 protein complexes and runs through the three steps in stage 1 to extract a total of 16540 sub structures in active sites. These sub structures can be categorized into 3531 classes of binding proteins. A protein (that is not used in the experiment of stage 1) is selected for each binding protein class from PDB to act as the "unclassified protein". Procedures in stage 2 are conducted on these unclassified proteins to determine the accuracy of protein classification technique proposed in this thesis.

## 4.1 Biotin Binding Proteins

Biotin binding proteins are initially used to test the validation of the above mentioned technique. For this purpose all the biotin binding proteins are extracted from protein data base and the active sites of all the proteins are extracted. Distinct motifs are selected for this purpose and their occurrence in the biotin binding protein is calculated. The following table shows the percentage of occurrence for these motifs

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| LEU --- SER --- THR --- TRP | 88.75 |
| ASN --- ASP --- LEU --- SER | 70.00 |
| ALA --- ALA --- ASN --- GLY --- TRP --- VAL | 62.5 |
| ALA --- ALA --- ASN --- SER --- THR --- TRP | 62.5 |
| ALA --- ALA --- ASN --- SER --- TRP | 63.75 |
| ALA --- LEU --- SER --- THR --- TRP | 82.5 |
| LEU --- SER --- THR --- TRP --- TRP --- TRP | 78.75 |
| ASP --- THR --- TRP --- TRP | 71.25 |
| ASP --- LEU --- THR --- TRP --- TRP | 76.25 |
| ASN --- ASP --- LEU --- TRP --- TRP | 66.25 |
| ASN --- LEU --- SER --- SER --- TYR --- VAL | 57.5 |
| SER --- SER --- TYR | 67.5 |

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| ALA --- ASN --- GLY --- SER --- VAL | 53.75 |
| ALA --- ALA --- ASN --- GLY --- SER --- TRP --- VAL | 55 |

**Table 5 % of existence of distinct functional motifs in Biotin Binding active site**

In order to better understand the significance of these functional motifs for biotin binding, the existence of these functional motifs in non-biotin binding sites is also calculated. For this purpose, the other active sites on the biotin binding proteins are checked. The active sites are calculated using Computed Atlas of Surface Topography of Proteins (CASTP) program [23]. The percentage when 730 non-biotin binding active sites are checked for the occurrence of the above functional motif is given below:

| Functional Motif | Percentage of Occurrence of the Distinct motif in other active sites |
|---|---|
| LEU --- SER --- THR --- TRP | 0.00 |
| ASN --- ASP --- LEU --- SER | 0.13 |
| ALA --- ALA --- ASN --- GLY --- TRP --- VAL | 0.00 |
| ALA --- ALA --- ASN --- SER --- THR --- TRP | 0.00 |
| ALA --- ALA --- ASN --- SER --- TRP | 0.00 |
| ALA --- LEU --- SER --- THR --- TRP | 0.00 |
| LEU --- SER --- THR --- TRP --- TRP --- TRP | 0.00 |
| ASP --- THR --- TRP --- TRP | 0.13 |
| ASP --- LEU --- THR --- TRP --- TRP | 0.13 |
| ASN --- ASP --- LEU --- TRP --- TRP | 0.00 |
| ASN --- LEU --- SER --- SER --- TYR --- VAL | 0.00 |
| SER --- SER --- TYR | 0.13 |
| ALA --- ASN --- GLY --- SER --- VAL | 0.00 |
| ALA --- ALA --- ASN --- GLY --- SER --- TRP --- VAL | 0.00 |

**Table 6 Existence of functional motifs in non-biotin binding active sites**

### *4.1.1 Precision and Recall results*

In order to classify a given protein as biotin binding, the existence of the functional motifs is checked in the active site. Precision and recall are calculated for this purpose.

The precision and recall for the 1swp when the confidence measure of 50% is applied and the distance of 8.5Å is used for the extraction of the sub structures from the active site.

Precision = 12/18 = 0.666

Recall = 12/16= 0.75

When the confidence measure of 50% is applied and the distance of 10Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 14/20 = 0.70

Recall = 14/19 = 0.73

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 16/26 = 0.61

Recall = 16/20= 0.8

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 18/28 = 0.642

Recall = 18/24 = 0.75

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 10/14 = 0.71

Recall = 10/15 = 0.666

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 12/16 = 0.75

Recall = 12/19 = 0.63

## 4.2 Retinal (RET) Binding proteins

Further validation of the classification is done by using retinal binding proteins. For this purpose the distinct motifs are selected in the retinal binding active sites and their percentage of existence is calculated. The percentage of occurrence of the distinct motif in the active sites of the retinal binding protein is given as

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| LEU --- LYS --- THR --- THR | 64.27 |
| MET --- MET --- SER --- THR --- TRP | 62.5 |
| PRO --- TRP --- TRP --- TYR | 67.50 |
| ALA --- MET --- PRO --- TRP --- TYR | 59.10 |
| ALA --- ASP --- LYS --- TRP | 57.95 |

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| ALA --- ASP --- LEU --- LYS --- THR | 51.1 |

**Table 7 % of existence of distinct functional motifs in Retinal Binding active site**

In order to check that these functional motifs are responsible for retinal binding only, the other 605 active sites of retinal binding proteins have been extracted using the CASTP program and checked for the existence of these functional motifs. The percentage of occurrence of these functional motifs is given below:

| Functional Motif | Percentage of Occurrence of the Distinct motif in other active sites |
|---|---|
| LEU --- LYS --- THR --- THR | 0.00 |
| MET --- MET --- SER --- THR --- TRP | 0.00 |
| PRO --- TRP --- TRP --- TYR | 0.00 |
| ALA --- MET --- PRO --- TRP --- TYR | 0.00 |
| ALA --- ASP --- LYS --- TRP | 0.00 |
| ALA --- ASP --- LEU --- LYS --- THR | 0.001 |

**Table 8 Existence of functional motifs in non-retinal binding active sites**

### *4.2.1 Precision and Recall results*

In order to classify a given protein as biotin binding, the existence of the functional motifs is checked in the active site. Precision and recall are calculated for this purpose.

When the confidence measure of 50% is applied and the distance of 8.5Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 5 /7 = 0.714

Recall = 5/6 = 0.83

When the confidence measure of 50% is applied and the distance of 10Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 8/10 = 0.8

Recall = 8/12 =0.666

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 7/11 = 0.64

Recall = 7/8 = 0.875

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 10/13 = 0.77

Recall = 10/14 = 0.72

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 4/5 = 0.8

Recall = 4/6 = 0.666

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 7/8 = 0.875

Recall = 7/11 = 0.64

## 4.3 Pyrroloquinoline Quinone (PQQ) binding proteins

PQQ binding proteins are also used for the validation of the classification method. Distinct motifs are extracted from the active sites of the PQQ binding proteins and their percentage of occurrence is calculated. The percentage of occurrence is given as

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| ARG --- GLU --- GLY --- VAL | 69.56 |
| ARG --- CYS --- GLU --- GLY --- THR --- VAL | 69.56 |
| ARG --- GLU --- THR --- VAL | 69.56 |
| ARG --- SER --- THR --- TRP --- VAL | 60.86 |
| CYS --- CYS --- GLY --- TRP --- VAL | 52.17 |
| ALA --- CYS --- GLU --- GLY --- SER --- THR | 52.17 |
| SER --- THR --- THR --- TRP | 52.17 |

**Table 9 % of existence of distinct functional motifs in PQQ Binding active site**

In order to check that these functional motifs are responsible for PQQ binding only, the other 1605 active sites of PQQ binding proteins have been extracted using the CASTP program and checked for the existence of these functional motifs. The percentage of occurrence of these functional motifs is given below:

| Functional Motif | Percentage of Occurrence of the Distinct motif in other active sites |
|---|---|
| ARG --- GLU --- GLY --- VAL | 0.00 |
| ARG --- CYS --- GLU --- GLY --- THR --- VAL | 0.00 |
| ARG --- GLU --- THR --- VAL | 0.001 |
| ARG --- SER --- THR --- TRP --- VAL | 0.001 |
| CYS --- CYS --- GLY --- TRP --- VAL | 0.00 |
| ALA --- CYS --- GLU --- GLY --- SER --- THR | 0.00 |
| SER --- THR --- THR --- TRP | 0.00 |

**Table 10 Existence of functional motifs in non-PQQ binding active sites**

### *4.3.1 Precision and Recall results*

In order to classify a given protein as biotin binding, the existence of the functional motifs is checked in the active site. Precision and recall are calculated for this purpose.

When the confidence measure of 50% is applied and the distance of 8.5Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 5/7 = 0.72

Recall = 5/8 = 0.63

When the confidence measure of 50% is applied and the distance of 10Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 7/ 11 = 0.64

Recall = 7 / 10 = 0.7

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 5/8 = 0.63

Recall = 6/8 = 0.75

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 9/15 = 0.6

Recall = 9/12 = 0.75

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 4/5 = 0.8

Recall = 4/7 = 0.58

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 5/7 = 0.72

Recall = 5/8 = 0.625

## 4.4 Ornithine (ORN) binding proteins

ORN binding proteins are also used for the validation of the classification method. For this purpose the distinct motifs are extracted from the active site of the ORN binding proteins and their percentage is calculated. The percentage of the distinct motifs for the ORN binding proteins is given as below

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| ALA ---ASP --- GLU --- GLU --- SER --- VAL | 52.94 |
| ALA --- ASP --- GLU --- SER --- VAL | 52.94 |
| ALA --- ASP --- GLU --- LEU --- SER --- VAL | 52.94 |
| ASP --- LEU --- THR --- TYR | 56.86 |
| ASP --- HIS --- TYR | 52.94 |
| ASP --- LEU --- THR | 58.82 |
| ASP --- LEU --- VAL | 62.74 |
| ASP --- HIS --- THR --- TYR | 52.94 |

**Table 11 % of existence of distinct functional motifs in ORN Binding active site**

Although these functional motifs are responsible for ORN binding it might be possible that these motifs generally appear in the other active site also and thus do not have statistical significance for ORN binding, other 348 active sites of the ORN binding proteins have been checked and the percentage of occurrence of these functional motif in other active sites is given below:

| Functional Motif | Percentage of Occurrence of the Distinct motif in other active sites |
|---|---|
| ALA ---ASP --- GLU --- GLU --- SER --- VAL | 0.00 |
| ALA --- ASP --- GLU --- SER --- VAL | 0.00 |
| ALA --- ASP --- GLU --- LEU --- SER --- VAL | 0.002 |
| ASP --- LEU --- THR --- TYR | 0.00 |
| ASP --- HIS --- TYR | 0.00 |
| ASP --- LEU --- THR | 0.002 |
| ASP --- LEU --- VAL | 0.002 |
| ASP --- HIS --- THR --- TYR | 0.00 |

**Table 12 Existence of functional motifs in non-ORN binding active sites**

### *4.4.1 Precision and Recall results*

In order to classify a given protein as biotin binding, the existence of the functional motifs is checked in the active site. Precision and recall are calculated for this purpose.

When the confidence measure of 50% is applied and the distance of 8.5Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 7/10 = 0.7

Recall = 7/9 = 0.78

When the confidence measure of 50% is applied and the distance of 10Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 8/11 = 0.73

Recall = 8/10 = 0.8

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 8/12 = 0.67

Recall = 8/9 = 0.89

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 12/18 = 0.67

Recall = 12/14 = 0.86

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 5/6 = 0.84

Recall = 5/7 = 0.71

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 6/8 = 0.75

Recall = 6/8 = 0.75

## 4.5 Xylose (XLS) binding proteins

XLS binding proteins are also used to validate the classification method. In order to use the XLS binding proteins, distinct motifs are selected from the active sites. The percentage of the distinct motif as found in the active site of the XLS binding proteins is given as below

| Functional Motif | Existence percentage of the Distinct Motif in the active site |
|---|---|
| HIS --- THR --- TRP | 65.55 |
| ASP --- ASP --- GLU | 63.5 |
| GLU --- GLU --- HIS --- LYS --- TRP | 75.33 |
| ASP --- ASP --- GLU --- GLU --- HIS --- LYS | 80 |
| ASP --- GLU --- HIS --- LYS | 70.5 |
| ASP --- HIS --- TRP | 63.33 |
| ASP --- ASP --- GLU --- TRP | 83.33 |
| GLU --- LYS --- THR --- TRP | 83.33 |
| GLU --- GLU --- LYS --- TRP | 83.33 |

**Table 13 % of existence of distinct functional motifs in XLS binding active site**

These functional motifs are then checked in non-XLS binding active sites formed by the XLS binding proteins. The existence of the functional motif in these active sites is given as below:

| Functional Motif | Percentage of Occurrence of the Distinct motif in other active sites |
|---|---|
| HIS --- THR --- TRP | 0.00 |
| ASP --- ASP --- GLU | 0.0008 |
| GLU --- GLU --- HIS --- LYS --- TRP | 0.00 |
| ASP --- ASP --- GLU --- GLU --- HIS --- LYS | 0.00 |
| ASP --- GLU --- HIS --- LYS | 0.00 |
| ASP --- HIS --- TRP | 0.0008 |
| ASP --- ASP --- GLU --- TRP | 0.00 |
| GLU --- LYS --- THR --- TRP | 0.0016 |
| GLU --- GLU --- LYS --- TRP | 0.00 |

**Table 14 Existence of functional motifs in non-XLS binding active sites**

### *4.5.1 Precision and Recall results*

In order to classify a given protein as biotin binding, the existence of the functional motifs is checked in the active site. Precision and recall are calculated for this purpose.

When the confidence measure of 50% is applied and the distance of 8.5Å is used to extract the sub structures from t he active site, the value of the precision and recall are given as

Precision = 8/10 = 0.8

Recall = 8/9 = 0.88

When the confidence measure of 50% is applied and the distance of 10Å is used to extract the sub structures from the active site, the value of the precision and recall are given as

Precision = 12/14 = 0.85

Recall = 12/13 = 0.92

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 12/16 = 0.75

Recall = 12 /13= 0.92

When the confidence measure of 40% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 15/20 = 0.75

Recall = 15/16 = 0.94

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 8.5Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 6/7 = 0.85

Recall = 6/8 = 0.75

When the confidence measure of 60% is applied for the extraction of the sub structures and the distance of 10Å is used to calculate the neighborhood between the residues the precision and the recall are given as

Precision = 7/8 = 0.875

Recall = 7/8= 0.875

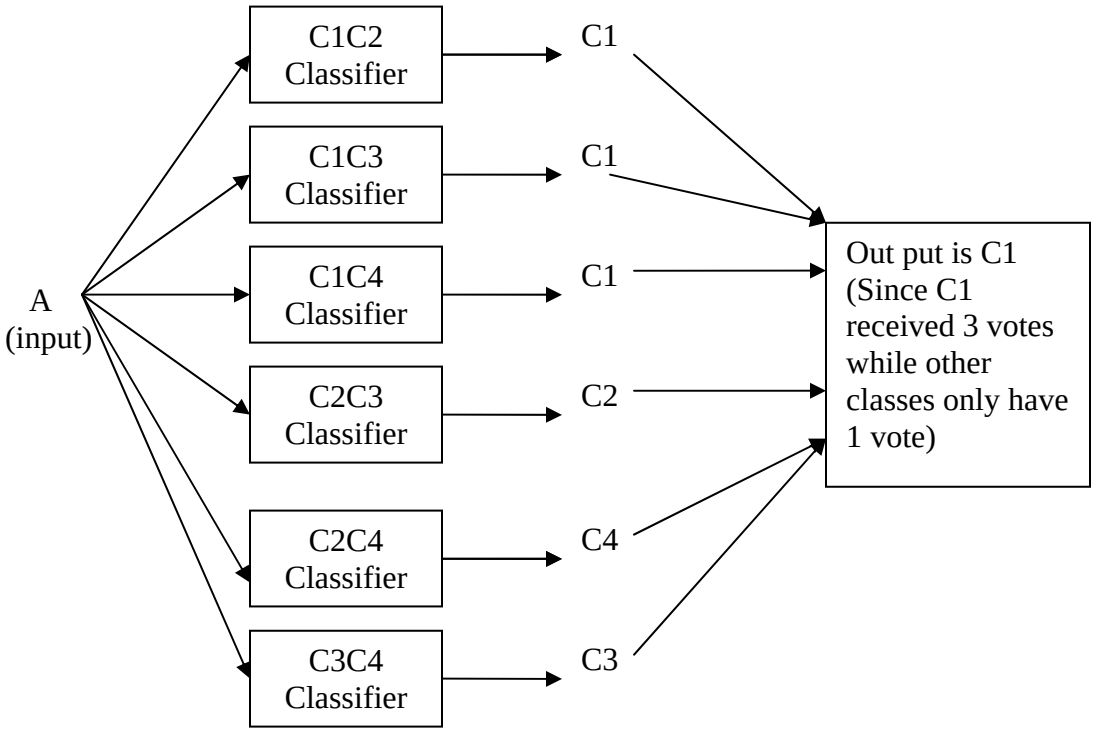
## 4.6 Multi-class Classification of the Proteins

Multi-class classification is inherently a complicated machine learning problem. As the problem discussed in this thesis belongs to this category, further analysis and research is done to successfully predict the class of an unclassified protein.

The most common way to solve the multi-class classification problem is to transform the multi-class problem into series of binary class problems. The common technique tries to find a classifier which can distinguish one class from all the rest of the classes. This technique uses real value in order to determine the class membership. Though this method is applied successfully in some of the machine learning problems it has the limitations of the expressivity. Since this technique assumes that each class can be easily separated from all the rest of the classes.

In case of the problem discussed in this thesis there are 3531 classes of proteins, so in order to perform more accurate multi class classification, a more detailed and expressive approach is used. The approach used to solve the current problem finds a classifier between each pair of the classes and then applies voting scheme to determine the exact class of the protein. This method essentially converts the multi-class classification problem into series of binary classification problem without losing the expressivity. Initially pair-wise binary classifiers are learned between the classes, then for a given problem, the class is determined by calculating the results from these binary classifiers. This is done by applying simple voting scheme. If majority of the binary classifiers classifies the input to a particular class then the input is considered to belong to that particular class[25].

An example with four classes is given below to further explain the concept. Consider there are four classes, C1, C2, C3 and C4. So in order to classify an input, first the pair-wise classifiers are learned. These include a classifier between C1 and C2, C1 and C3, C1 and C4. Similarly the classifiers for the other pair of the classes are learned. Consider an input A is given to the set of these binary pair-wise classifiers. Every pair-wise classifier will try to classify the input to one of the class. The output from these classifiers will then be processed. If the input is classified to one particular class by the majority of the binary classifiers then it will be taken as the original classification of the input. In case of the ties between the classes, they will be resolved in the favor of the bigger class. The following figure explains this process in detail




### *4.6.1 Case for 4 class classification*

In this case biotin binding proteins, retinal binding proteins, PQQ binding protein and XLS binding proteins are selected. Correspondingly the unclassified proteins of the above mentioned classes are selected and they are given as input to these classifiers which can then classify them according to the class. The case for every unclassified in the particular class is given below

#### 4.6.1.1 Classifying “1swp” as biotin binding protein

1swp is given as input to the set of the classifiers for classification. In this case the class biotin receives the majority (i.e. 3) of the votes. So the protein 1swp is classified as biotin binding protein.

#### 4.6.1.2 Classifying “1uaz” as retinal binding protein

1uaz is given as input to the set of the classifiers for classification. In this case the class retinal receives the majority (i.e. 3) of the votes. So the protein 1uaz is classified as retinal binding protein.

#### 4.6.1.3 Classifying “1flg” as PQQ binding protein

1flg is given as input to the set of the classifiers for classification. In this case the class PQQ receives the majority (i.e. 3) of the votes. So the protein 1lfg is classified as PQQ binding protein.

#### 4.6.1.4 Classifying “3xis” as xylose binding protein

3xis is given as input to the set of the classifiers for classification. In this case the class xylose receives the majority (i.e. 3) of the votes. So the protein 3xis is classified as xylose binding protein.

### *4.6.2 Case for 8 class classification*

In this case BTN, RET, PQQ, XLS, ORN, 4MO, LVS and 9PP are selected to perform 8 class classification. For this purpose the following 28 classifiers are leant

(BTN-PQQ classifier), (BTN-RET classifier), (BTN-XLS classifier), (BTN-ORN classifier), (BTN-4MO classifier), (BTN-LVS classifier), (BTN-9PP classifier), (PQQ-RET classifier), (PQQ-XLS classifier), (PQQ-ORN classifier), (PQQ-4MO classifier), (PQQ-LVS classifier), (PQQ-9PP classifier), (RET-XLS classifier), (RET-ORN classifier), (RET-4MO classifier), (RET-LVS classifier), (RET-9PP classifier), ( XLS-ORN classifier), (XLS-4MO classifier), (XLS-LVS classifier), (XLS-9PP classifier), (ORN-4MO classifier), (ORN-LVS classifier), (ORN-9PP classifier), (4MO-LVS classifier) (4MO-9PP classifier) and (LVS-9PP classifier).

4.6.2.1 Classifying “1swp” as biotin binding protein

1swp is given as input to the set of the classifiers for classification. A voting table is created and the votes for every class are calculated. The class which takes the most votes will be selected as corresponding binding class. The voting table for the 1swp is given below

| **Classifier Type** | **Selected Class** | **Vote By Class** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | BTN | 7 | 3 | 3 | 2 | 3 | 4 | 5 | 1 |
| BTN-XLS | BTN | | | | | | | | |
| BTN-ORN | BTN | | | | | | | | |
| BTN-4MO | BTN | | | | | | | | |
| BTN-PQQ | BTN | | | | | | | | |
| BTN-LVS | BTN | | | | | | | | |
| BTN-9PP | BTN | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PQQ-RET | PQQ | | | | | | | | |
| PQQ-XLS | PQQ | | | | | | | | |
| PQQ-ORN | ORN | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | LVS | | | | | | | | |
| PQQ-9PP | PQQ | | | | | | | | |
| RET-XLS | RET | | | | | | | | |
| RET-ORN | RET | | | | | | | | |
| RET-4MO | 4MO | | | | | | | | |
| RET-LVS | LVS | | | | | | | | |
| RET-9PP | RET | | | | | | | | |
| XLS-ORN | ORN | | | | | | | | |
| XLS-4MO | XLS | | | | | | | | |
| XLS-LVS | LVS | | | | | | | | |
| XLS-9PP | XLS | | | | | | | | |
| ORN-4MO | 4MO | | | | | | | | |
| ORN-LVS | ORN | | | | | | | | |
| ORN-9PP | 9PP | | | | | | | | |
| 4MO-LVS | LVS | | | | | | | | |
| 4MO-9PP | 4MO | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class BTN is selected by most of the classifiers (i.e. 7), so 1swp is classified as biotin binding protein.

4.6.2.2 Classifying “1uaz” as retinal binding protein

1uaz is given as input to the set of the classifiers for classification. The voting table for the 1uaz is given below

| **Classifier Type** | **Selected Class** | **Vote By Class** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | RET | 4 | 3 | 6 | 2 | 3 | 4 | 3 | 3 |
| BTN-XLS | BTN | | | | | | | | |
| BTN-ORN | BTN | | | | | | | | |
| BTN-4MO | 4MO | | | | | | | | |
| BTN-PQQ | PQQ | | | | | | | | |
| BTN-LVS | BTN | | | | | | | | |
| BTN-9PP | BTN | | | | | | | | |
| PQQ-RET | RET | | | | | | | | |
| PQQ-XLS | XLS | | | | | | | | |
| PQQ-ORN | PQQ | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | LVS | | | | | | | | |
| PQQ-9PP | PQQ | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| RET-XLS | RET | | | | | | | | |
| RET-ORN | RET | | | | | | | | |
| RET-4MO | 4MO | | | | | | | | |
| RET-LVS | RET | | | | | | | | |
| RET-9PP | RET | | | | | | | | |
| XLS-ORN | ORN | | | | | | | | |
| XLS-4MO | XLS | | | | | | | | |
| XLS-LVS | LVS | | | | | | | | |
| XLS-9PP | 9PP | | | | | | | | |
| ORN-4MO | ORN | | | | | | | | |
| ORN-LVS | ORN | | | | | | | | |
| ORN-9PP | 9PP | | | | | | | | |
| 4MO-LVS | 4MO | | | | | | | | |
| 4MO-9PP | 9PP | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class RET is selected by most of the classifiers (i.e. 6), so 1uaz is classified as retinal binding protein.

#### 4.6.2.3 Classifying “1flg” as PQQ binding protein

1flg is given as input to the set of the classifiers for classification. The voting table for the 1flg is given below

| Classifier Type | Selected Class | Vote By Class | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | RET | 2 | 6 | 3 | 3 | 3 | 3 | 3 | 5 |
| BTN-XLS | BTN | | | | | | | | |
| BTN-ORN | ORN | | | | | | | | |
| BTN-4MO | 4MO | | | | | | | | |
| BTN-PQQ | PQQ | | | | | | | | |
| BTN-LVS | BTN | | | | | | | | |
| BTN-9PP | 9PP | | | | | | | | |
| PQQ-RET | PQQ | | | | | | | | |
| PQQ-XLS | PQQ | | | | | | | | |
| PQQ-ORN | PQQ | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | PQQ | | | | | | | | |
| PQQ-9PP | PQQ | | | | | | | | |
| RET-XLS | XLS | | | | | | | | |
| RET-ORN | RET | | | | | | | | |
| RET-4MO | RET | | | | | | | | |
| RET-LVS | LVS | | | | | | | | |
| RET-9PP | 9PP | | | | | | | | |
| XLS-ORN | XLS | | | | | | | | |
| XLS-4MO | 4MO | | | | | | | | |
| XLS-LVS | XLS | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| XLS-9PP | 9PP | | | | | | | | |
| ORN-4MO | ORN | | | | | | | | |
| ORN-LVS | ORN | | | | | | | | |
| ORN-9PP | 9PP | | | | | | | | |
| 4MO-LVS | LVS | | | | | | | | |
| 4MO-9PP | 9PP | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class PQQ is selected by most of the classifiers (i.e. 6), so 1flg is classified as PQQ binding protein.

4.6.2.4 Classifying “3xis” as xylose binding protein

3xis is given as input to the set of the classifiers for classification. The voting table for the 3xis is given below

| **Classifier Type** | **Selected Class** | **Vote By Class** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | BTN | 4 | 3 | 2 | 7 | 3 | 3 | 3 | 3 |
| BTN-XLS | XLS | | | | | | | | |
| BTN-ORN | BTN | | | | | | | | |
| BTN-4MO | BTN | | | | | | | | |
| BTN-PQQ | PQQ | | | | | | | | |
| BTN-LVS | BTN | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| BTN-9PP | 9PP | | | | | | | | |
| PQQ-RET | PQQ | | | | | | | | |
| PQQ-XLS | XLS | | | | | | | | |
| PQQ-ORN | ORN | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | LVS | | | | | | | | |
| PQQ-9PP | PQQ | | | | | | | | |
| RET-XLS | XLS | | | | | | | | |
| RET-ORN | ORN | | | | | | | | |
| RET-4MO | 4MO | | | | | | | | |
| RET-LVS | RET | | | | | | | | |
| RET-9PP | RET | | | | | | | | |
| XLS-ORN | XLS | | | | | | | | |
| XLS-4MO | XLS | | | | | | | | |
| XLS-LVS | XLS | | | | | | | | |
| XLS-9PP | XLS | | | | | | | | |
| ORN-4MO | ORN | | | | | | | | |
| ORN-LVS | LVS | | | | | | | | |
| ORN-9PP | 9PP | | | | | | | | |
| 4MO-LVS | 4MO | | | | | | | | |
| 4MO-9PP | 9PP | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class XLS is selected by most of the classifiers (i.e. 7), so 3xis is classified as XLS binding protein.

4.6.2.5 Classifying “1vlf” as 4MO binding protein

1vlf is given as input to the set of the classifiers for classification. The voting table for the 1vlf is given below

| **Classifier Type** | **Selected Class** | **Vote By Class** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | BTN | 4 | 3 | 2 | 4 | 3 | 5 | 4 | 3 |
| BTN-XLS | XLS | | | | | | | | |
| BTN-ORN | BTN | | | | | | | | |
| BTN-4MO | BTN | | | | | | | | |
| BTN-PQQ | PQQ | | | | | | | | |
| BTN-LVS | BTN | | | | | | | | |
| BTN-9PP | 9PP | | | | | | | | |
| PQQ-RET | RET | | | | | | | | |
| PQQ-XLS | XLS | | | | | | | | |
| PQQ-ORN | PQQ | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | LVS | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| PQQ-9PP | PQQ | | | | | | | | |
| RET-XLS | RET | | | | | | | | |
| RET-ORN | ORN | | | | | | | | |
| RET-4MO | 4MO | | | | | | | | |
| RET-LVS | LVS | | | | | | | | |
| RET-9PP | 9PP | | | | | | | | |
| XLS-ORN | XLS | | | | | | | | |
| XLS-4MO | 4MO | | | | | | | | |
| XLS-LVS | XLS | | | | | | | | |
| XLS-9PP | 9PP | | | | | | | | |
| ORN-4MO | 4MO | | | | | | | | |
| ORN-LVS | ORN | | | | | | | | |
| ORN-9PP | ORN | | | | | | | | |
| 4MO-LVS | LVS | | | | | | | | |
| 4MO-9PP | 4MO | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class 4MO is selected by most of the classifiers (i.e. 5), so 1vlf is classified as 4MO binding protein

4.6.2.6 Classifying “1kyi” as LVS binding protein

1kyi is given as input to the set of the classifiers for classification. The voting table for the 1kyi is given below

| Classifier Type | Selected Class | Vote By Class | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | BTN | 4 | 3 | 4 | 4 | 1 | 3 | 7 | 2 |
| BTN-XLS | XLS | | | | | | | | |
| BTN-ORN | BTN | | | | | | | | |
| BTN-4MO | BTN | | | | | | | | |
| BTN-PQQ | PQQ | | | | | | | | |
| BTN-LVS | LVS | | | | | | | | |
| BTN-9PP | BTN | | | | | | | | |
| PQQ-RET | RET | | | | | | | | |
| PQQ-XLS | PQQ | | | | | | | | |
| PQQ-ORN | PQQ | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | LVS | | | | | | | | |
| PQQ-9PP | 9PP | | | | | | | | |
| RET-XLS | XLS | | | | | | | | |
| RET-ORN | RET | | | | | | | | |
| RET-4MO | RET | | | | | | | | |
| RET-LVS | LVS | | | | | | | | |
| RET-9PP | RET | | | | | | | | |
| XLS-ORN | XLS | | | | | | | | |
| XLS-4MO | 4MO | | | | | | | | |
| XLS-LVS | LVS | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| XLS-9PP | XLS | | | | | | | | |
| ORN-4MO | ORN | | | | | | | | |
| ORN-LVS | LVS | | | | | | | | |
| ORN-9PP | 9PP | | | | | | | | |
| 4MO-LVS | LVS | | | | | | | | |
| 4MO-9PP | 4MO | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class LVS is selected by most of the classifiers (i.e. 7), so 1kyi is classified as LVS binding protein

4.6.2.7 Classifying “1lvu” as 9PP binding protein

1lvu is given as input to the set of the classifiers for classification. The voting table for the 1lvu is given below

| **Classifier Type** | **Selected Class** | **Vote By Class** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | RET | 2 | 2 | 4 | 3 | 4 | 3 | 4 | 6 |
| BTN-XLS | BTN | | | | | | | | |
| BTN-ORN | ORN | | | | | | | | |
| BTN-4MO | BTN | | | | | | | | |
| BTN-PQQ | PQQ | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| BTN-LVS | LVS | | | | | | | | |
| BTN-9PP | 9PP | | | | | | | | |
| PQQ-RET | RET | | | | | | | | |
| PQQ-XLS | XLS | | | | | | | | |
| PQQ-ORN | PQQ | | | | | | | | |
| PQQ-4MO | 4MO | | | | | | | | |
| PQQ-LVS | LVS | | | | | | | | |
| PQQ-9PP | 9PP | | | | | | | | |
| RET-XLS | RET | | | | | | | | |
| RET-ORN | RET | | | | | | | | |
| RET-4MO | 4MO | | | | | | | | |
| RET-LVS | LVS | | | | | | | | |
| RET-9PP | 9PP | | | | | | | | |
| XLS-ORN | ORN | | | | | | | | |
| XLS-4MO | XLS | | | | | | | | |
| XLS-LVS | XLS | | | | | | | | |
| XLS-9PP | 9PP | | | | | | | | |
| ORN-4MO | ORN | | | | | | | | |
| ORN-LVS | ORN | | | | | | | | |
| ORN-9PP | 9PP | | | | | | | | |
| 4MO-LVS | 4MO | | | | | | | | |
| 4MO-9PP | 9PP | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class 9PP is selected by most of the classifiers (i.e. 6), so 1lvu is classified as 9PP binding protein

4.6.2.8 Classifying “1cs0” as ORN binding protein

1cs0 is given as input to the set of the classifiers for classification. The voting table for the 1cs0 is given below

| **Classifier Type** | **Selected Class** | **Vote By Class** | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | BTN | PQQ | RET | XLS | ORN | 4MO | LVS | 9PP |
| BTN-RET | BTN | 3 | 3 | 2 | 4 | 7 | 2 | 4 | 3 |
| BTN-XLS | XLS | | | | | | | | |
| BTN-ORN | ORN | | | | | | | | |
| BTN-4MO | BTN | | | | | | | | |
| BTN-PQQ | BTN | | | | | | | | |
| BTN-LVS | LVS | | | | | | | | |
| BTN-9PP | 9PP | | | | | | | | |
| PQQ-RET | RET | | | | | | | | |
| PQQ-XLS | PQQ | | | | | | | | |
| PQQ-ORN | ORN | | | | | | | | |
| PQQ-4MO | PQQ | | | | | | | | |
| PQQ-LVS | PQQ | | | | | | | | |
| PQQ-9PP | 9PP | | | | | | | | |
| RET-XLS | XLS | | | | | | | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| RET-ORN | ORN | | | | | | | | |
| RET-4MO | 4MO | | | | | | | | |
| RET-LVS | LVS | | | | | | | | |
| RET-9PP | RET | | | | | | | | |
| XLS-ORN | ORN | | | | | | | | |
| XLS-4MO | 4MO | | | | | | | | |
| XLS-LVS | XLS | | | | | | | | |
| XLS-9PP | XLS | | | | | | | | |
| ORN-4MO | ORN | | | | | | | | |
| ORN-LVS | ORN | | | | | | | | |
| ORN-9PP | ORN | | | | | | | | |
| 4MO-LVS | LVS | | | | | | | | |
| 4MO-9PP | 9PP | | | | | | | | |
| LVS-9PP | LVS | | | | | | | | |

As it can be seen from the voting table that the class ORN is selected by most of the classifiers (i.e. 7), so 1cs0 is classified as ORN binding protein

### *4.6.3 Case for 16 class classification*

In this case 16 different classes are selected i.e. BTN, ORN, 4MO, LVS, 9PP, PQQ, XLS, RET, 2GP, MGN, BOX, FLP, MTX, TDG, NTM and STY. For performing the classification for 16 different classes 120 pair wise classifiers are learnt. The detail of these classifiers is given below

(BTN-ORN), (BTN-4MO), (BTN-LVS), (BTN-9PP), (BTN-PQQ), (BTN-XLS ), (BTN-RET), (BTN-2GP), (BTN-MGN), (BTN-BOX), (BTN-FLP), (BTN-MTX), (BTN-TDG), (BTN-NTM), (BTN-STY), (ORN-4MO), (ORN-LVS), (ORN-9PP), (ORN-PQQ), (ORN-XLS), (ORN-RET), (ORN-2GP), (ORN-MGN), (ORN-BOX), (ORN-FLP), (ORN-MTX),(ORN-TDG),(ORN-NTM),(ORN-STY),(4MO-LVS),(4MO-9PP),(4MO-PQQ),(4MO-XLS),(4MO-RET), (4MO-2GP), (4MO-MGN), (4MO-BOX),(4MO-FLP),(4MO-MTX),(4MO-TDG),(4MO-NTM), (4MO-STY), (LVS-9PP), (LVS-PQQ), (LVS-XLS),(LVS-RET), (LVS-2GP), (LVS-MGN),(LVS-BOX),(LVS-FLP),(LVS-MTX),(LVS-NTM),(LVS-STY),(9PP-PQQ),(9PP-XLS),(9PP-RET), (9PP-2GP), (9PP-MGN), (9PP-BOX), (9PP-FLP),(9PP-MTX), (9PP-TDG),(9PP-NTM),(9PP-STY), (PQQ-XLS), (PQQ-RET), (PQQ-2GP), (PQQ-MGN), (PQQ-BOX), (PQQ-FLP), (PQQ-MTX), (PQQ-TDG), (PQQ-NTM), (PQQ-STY), (XLS-RET), (XLS-2GP),(XLS-MGN),(XLS-BOX), (XLS-FLP),(XLS-MTX), (XLS-TDG),(XLS-NTM),(XLS-STY), (2GP-MGN), (2GP-BOX), (2GP-FLP), (2GP-MTX), (2GP-TDG), (2GP-NTM), (2GP-STY), (MGN-BOX), (MGN-FLP), (MGN-MTX), (MGN-TDG), (MGN-NTM), (MGN-STY), (BOX-FLP), (BOX-MTX), (BOX-TDG), (BOX-NTM), (BOX-STY), (FLP-MTX), (FLP-TDG), (FLP-NTM), (FLP-STY), (MTX-TDG), (MTX-NTM), (MTX-STY), (TDG-NTM), (TDG-STY) and (NTM-STY) classifier.

#### 4.6.3.1 Classifying “1swp” as biotin binding protein

1swp is given as input to the set of the classifiers for classification. A voting table is created and the votes for every class are calculated. The class which takes the most votes will be selected as corresponding binding class. The voting table for the 1swp is given in

Appendix B. As it can be seen from the voting table that the class BTN is selected by most of the classifiers (i.e. 11) so 1swp is classified as biotin binding protein.

#### 4.6.3.2 Classifying "1uaz" as retinal binding protein

1uaz is given as input to the set of the classifiers for classification. The voting table for the 1uaz is given in Appendix B. As it can be seen from the voting table that the class RET is selected by most of the classifiers (i.e.12), so 1uaz is classified as retinal binding protein.

#### 4.6.3.3 Classifying "1flg" as PQQ binding protein

1flg is given as input to the set of the classifiers for classification. The voting table for the 1flg is given in Appendix B. As it can be seen from the voting table that the class PQQ is selected by most of the classifiers (i.e. 12), so 1flg is classified as PQQ binding protein.

#### 4.6.3.4 Classifying "3xis" as XLS binding protein

3xis is given as input to the set of the classifiers for classification. The voting table for the 3xis is given in Appendix B. As it can be seen from the voting table that the class XLS is selected by most of the classifiers (i.e. 11), so 3xis is classified as XLS binding protein.

#### 4.6.3.5 Classifying "1vlf" as 4MO binding protein

1vlf is given as input to the set of the classifiers for classification. The voting table for the 1vlf is given in Appendix B. As it can be seen from the voting table that the class 4MO is selected by most of the classifiers (i.e. 11), so 1vlf is classified as 4MO binding protein.

#### 4.6.3.6 Classifying “1kyi” as LVS binding protein

1kyi is given as input to the set of the classifiers for classification. The voting table for the 1kyi is given in Appendix B. As it can be seen from the voting table that the class LVS is selected by most of the classifiers (i.e. 12), so 1kyi is classified as LVS binding protein.

#### 4.6.3.7 Classifying “1lvu” as 9PP binding protein

1lvu is given as input to the set of the classifiers for classification. The voting table for the 1lvu is given in Appendix B. As it can be seen from the voting table that the class 9PP is selected by most of the classifiers (i.e. 11), so 1lvu is classified as 9PP binding protein.

#### 4.6.3.8 Classifying “1cs0” as ORN binding protein

1cs0 is given as input to the set of the classifiers for classification. The voting table for the 1cs0 is given in Appendix B. As it can be seen from the voting table that the class ORN is selected by most of the classifiers (i.e. 11), so 1cs0 is classified as ORN binding protein.

#### 4.6.3.9 Classifying “1bu4” as 2GP binding protein

1bu4 is given as input to the set of the classifiers for classification. The voting table for the 1bu4 is given in Appendix B. As it can be seen from the voting table that the class 2GP is selected by most of the classifiers (i.e. 12), so 1bu4 is classified as 2GP binding protein.

#### 4.6.3.10 Classifying “1mro” as MGN binding protein

1mro is given as input to the set of the classifiers for classification. The voting table for the 1mro is given in Appendix B. As it can be seen from the voting table that the class MGN is selected by most of the classifiers (i.e. 12), so 1mro is classified as MGN binding protein.

#### 4.6.3.11 Classifying “1ais” as BOX binding protein

1ais is given as input to the set of the classifiers for classification. The voting table for the 1ais is given in Appendix B. As it can be seen from the voting table that the class BOX is selected by most of the classifiers (i.e. 13), so 1ais is classified as BOX binding protein.

#### 4.6.3.12 Classifying “1dvt” as FLP binding protein

1dvt is given as input to the set of the classifiers for classification. The voting table for the 1dvt is given in Appendix B. As it can be seen from the voting table that the class FLP is selected by most of the classifiers (i.e. 13), so 1dvt is classified as FLP binding protein.

#### 4.6.3.13 Classifying “1ddr” as MTX binding protein

1ddr is given as input to the set of the classifiers for classification. The voting table for the 1ddr is given in Appendix B. As it can be seen from the voting table that the class MTX is selected by most of the classifiers (i.e. 14), so 1ddr is classified as MTX binding protein.

#### 4.6.3.14 Classifying “1qap” as NTM binding protein

1qap is given as input to the set of the classifiers for classification. The voting table for the 1qap is given in Appendix B. As it can be seen from the voting table that the class NTM is selected by most of the classifiers (i.e. 13), so 1qap is classified as NTM binding protein.

#### 4.6.3.15 Classifying “1ghx” as STY binding protein

1ghx is given as input to the set of the classifiers for classification. The voting table for the 1ghx is given in Appendix B. As it can be seen from the voting table that the class NTM is selected by most of the classifiers (i.e. 14), so 1ghx is classified as NTM binding protein while in reality 1ghx is STY binding protein. (1ghx cannot be classified by our system correctly)

#### 4.6.3.16 Classifying “1h5t” as TDG binding protein

1h5t is given as input to the set of the classifiers for classification. The voting table for the 1h5t is given in Appendix B. As it can be seen from the voting table that the class NTM and MTX are selected by most of the classifiers (i.e. 11), so 1h5t is classified as NTM and MTX binding protein, while in reality 1h5t binds with TDG. (1h5t cannot be classified by our system correctly).

### *4.6.4 Efficiency of the current method*

Theoretically our multi class classification method converts the M-class problem into M(M-1)/2 learning problems [26]. So in order to make a decision for classification quadratic numbers of the classifiers should be checked. So this method works slower a bit then other classification methods. The advantages of this method is that it is more accurate then the other methods available, as discussed in detail in the below comparison

with the other methods which also try to perform classification. The method maintain high precision and recall when it is applied on the 2,4 and 8 classes. Although the precision of the method is declined to 0.87 when 16 classes are considered, it is much more better then the other methods like top down classification and bottom up classification. A more detail comparison is considered below text using the G coupled receptors.

## 4.7 Comparison of Classification Techniques using G-protein coupled receptors

G-protein coupled receptors (GPCR) becomes part of research focus because of the molecular mechanisms that are involved in the GPCR functions. Many techniques have been devised to classify the GPCR proteins, e.g. top down clustering, bottom up clustering, Gene-Ontology. Due to the lack of functional classification on other common proteins across different techniques, GPCR proteins will be used as a case study for the comparisons of these classification techniques with the technique presented in this thesis.

Gene-ontology is a non-computational classification technique. It does a thorough review across literatures to devise a classification technique based on the properties of proteins, as well as how they are classified by other literatures. For verification purpose on the correctness of its classification, it further performs laboratory experiments to determine the actual protein functions and thus its classification [27]. The classification conducted by Gene-ontology will be used as the point of reference for comparisons of correctness across the other three computational classification techniques, i.e. top down

clustering, bottom up clustering and the presented technique. Precision and recall are used as a measure for

### *4.7.1 Functional classification of class GPCRMGR*

The following diagrams illustrate the functional classification of GPCRMGR class by Gene-ontology, top down clustering, bottom up clustering and presented techniques. Due to space limitations, the sequences of the proteins are not shown in the diagrams though they are used for the classification of the proteins. Instead, their names are presented base on the Swissprot format. The actual sequences of these proteins are given in the appendix attached to the thesis.

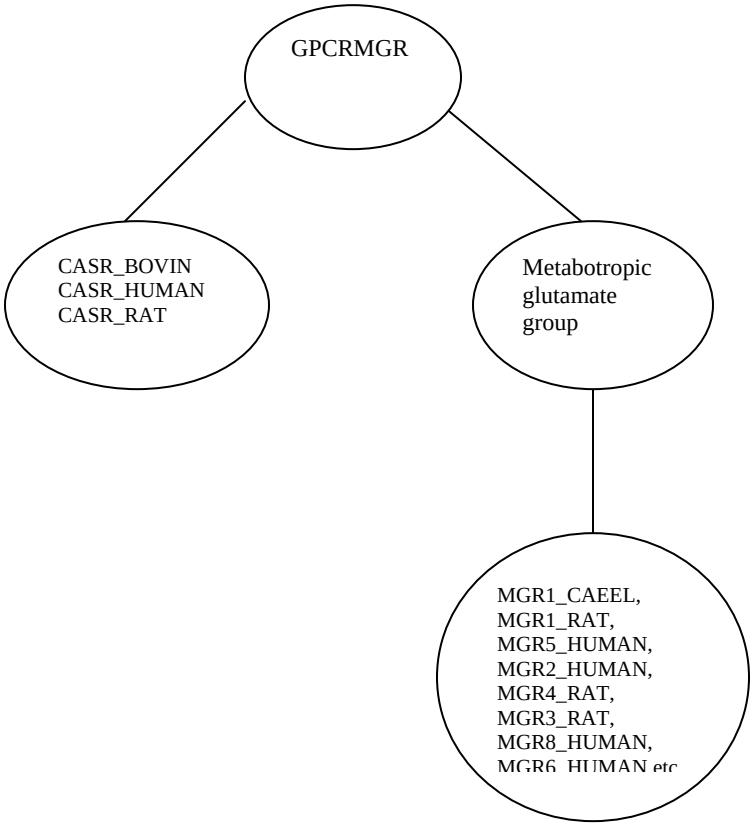



**Figure 6 Functional classification of GPCRMGR by Gene-Ontology**

### *4.7.2 Precision and Recall results for functional classification of class GPCRMGR (Experimental Results)*

| Type of Classification | Precision | Recall |
|---|---|---|
| Top Down Classification 2-Class Classification | 1.0 | 1.0 |
| Top Down Classification 4-Class Classification | 0.5 | 1.0 |
| Top Down Classification 8-Class Classification | 0.25 | 1.0 |
| Top Down Classification 16-Class Classification | 0.2 | 1.0 |
| Bottom up Classification 2-Class Classification | 1.0 | 1.0 |
| Bottom up Classification 4-Class Classification | 0.5 | 1.0 |
| Bottom up Classification 8-Class Classification | 0.25 | 1.0 |
| Bottom up Classification 16-Class Classification | 0.125 | 1.0 |
| Proposed Method 2 Class Classification | 1.0 | 1.0 |
| Proposed Method 4 Class Classification | 1.0 | 1.0 |

| | | |
|---|---|---|
| Proposed Method<br>8 Class Classification | 1.0 | 1.0 |
| Proposed Method<br>16 Class Classification | 0.8 | 1.0 |

## 4.8 Advantages of the presented technique

The presented technique provides the details of the structures of the active site, which are responsible for the binding. Since the technique is based on the structure, it can more accurately perform the classification. Because the new technique only compares the sub structures of the active site instead of calculcating the distributions of the properties in the environment of the active site, it is more faster then other structure-based techniques.

## 4.9 Limitations of the newly presented technique

As the technique is based on structure, in the absence of structural information, no classification can be performed by the technique. In this regards top down clustering and bottom up clustering techniques are advantageous since they can perform classification given only the sequence information though the classification performed by the new structure based technique is more accurate and provide more insight to the functional classification by explaining the sub structure of the active site in detail.

# 5 Future Work

Currently the technique is used to study and analyze the active sites of the proteins and extract the sub structures from this site for discovering the functional motifs. Future work can be done in several aspects of the proteomics. This ranges from motif by motif interaction, functionally important sub sequence discovery and applications in visualization areas. Some details of the applications of the current techniques are given below.

## 5.1 Sub Sequence patterns based Functional Motif discovery

Though advances in structural genomics can now approximate the structure of the protein from its sequence (primary structure), currently these techniques are still in development. A technique with 100% correct results for predicting the tertiary structure of the proteins from the primary structure has yet not been achieved.

In this situation, often the primary structure is used to predict the functional properties of the protein. As the current technique is developed using the information coming from the tertiary structures, currently it is based on the experimental information of the proteins complexes whose structure has already been revealed. So in order to use the technique for the proteins whose tertiary structures has yet not been found, further research work will be done to understand the relationships of these sub structures of the active site to the primary structure of the proteins. Once the sub sequence patterns of these sub structures are found, these sub sequences can be directly used to predict the functional motifs in the proteins and to estimate the functions of the unknown proteins.

## 5.2 Motif by Motif interaction of Proteins

Current studies of protein-protein interaction are done at the domain level. A domain contains multiple structural and functional motifs. In order to study these interactions in more details, a lower level of motif-by-motif interaction can be applied. Currently the technique is using ligands-protein complexes for developing these sub structures, in future the protein-protein complexes can be used to develop a motif-by-motif interaction system.

## 5.3 Structural/Functional Visualizations bridging (Pathway Characterization)

Currently the fields of morphological visualizations that deal with the structural visualizations of the proteins and the physiological visualizations that deal with the biochemical involvement of these proteins in the cellular pathways are two different streams. The morphological visualizations are normally used by the biologists for understanding the structural details of the proteins and their relationships with each other. The physiological visualizations are normally used by the biochemists to study the different chemical processes for the proteins.

Currently there is no interfacing between these two kinds of visualizations, so a biochemist can either look on the functional side in terms of the chemical reactions of the proteins or can see the morphological visualization in terms of protein structures. No visualizations exist which can map the morphological visualizations to the physiological visualizations. The sub structures of the active sites can be clustered according to their functional properties. These clustered sub structure will then act as a bridge between

these two kinds of visualizations. Further more these clustering can be done in multiple layers, thus using these sub structures to represent different cellular pathways.

The basic idea of pathway characterization using these sub structures is given below.

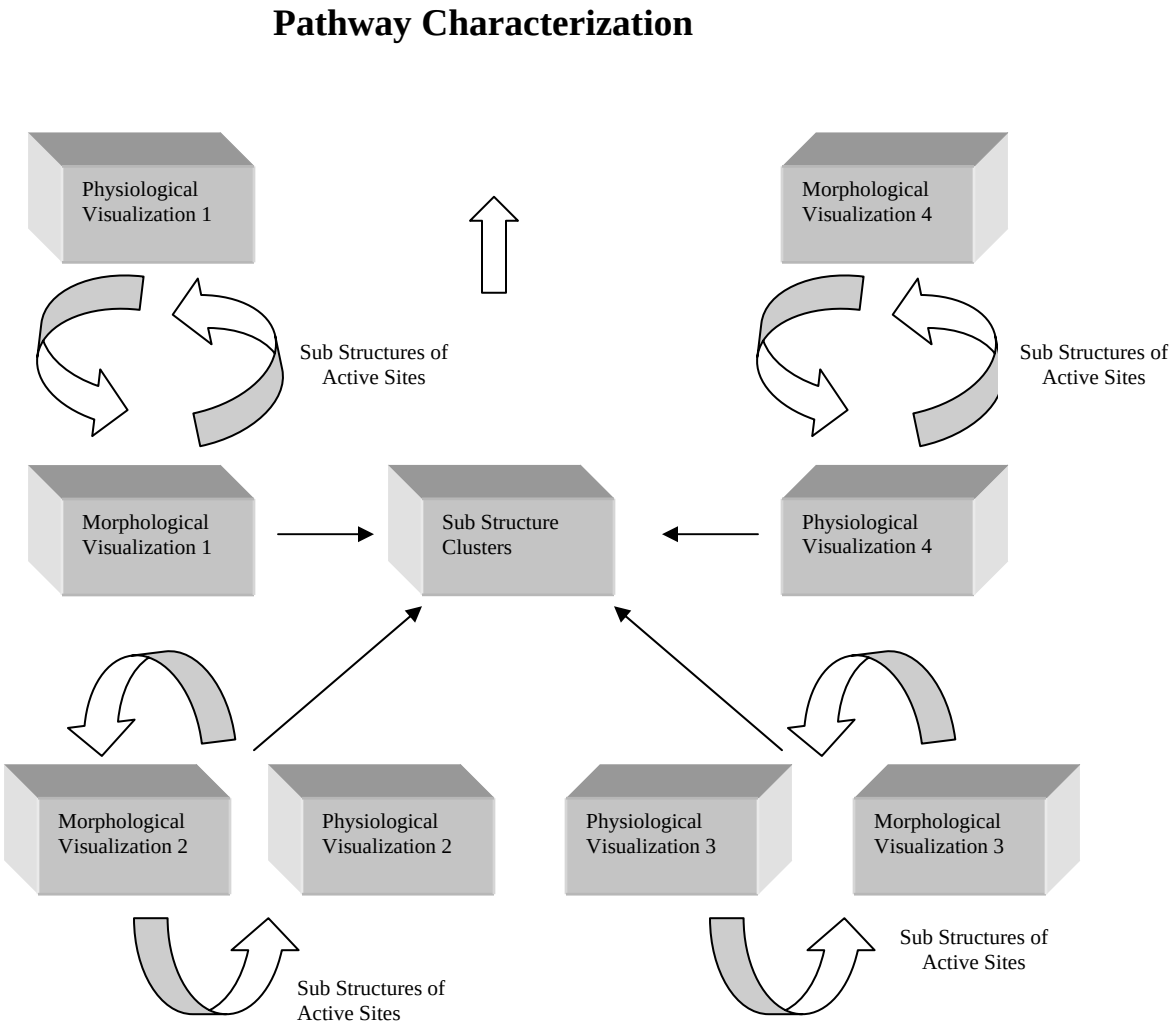



**Figure 7 Pathway characterization by bridging physiological and morphological visualization**

## 5.4 Conclusion

As theory of co-evolution of proteins says that evolutionary pressure is exerted on the functionally important regions, the possibility that the sub structures inside the active sites remain preserved is much more higher the preservation of the whole active site. Biochemistry explains the importance of the tertiary structure of protein for the functional identification and classification, although some aspects of tertiary structure have been used for finding functionally important regions, the small sub structures of the active sites have not been used previously to discover functional motifs. The technique presented in this paper tries to utilize these sub structure inside the active sites to detect functionally important regions in proteins. The advantage of using this technique for predicting the functionally important regions in the proteins is that the technique uses experimental data instead of putative interfaces to extract the substructures from the active site that are functionally important. The technique is successfully applied to perform the functional classification of proteins when the structure of the protein is given. The technique cannot only be used for functional motif discovery but can also be utilized to other problems that lie in the structures to functions relationships domain.

## Bibliography


***1.*** H.M. Berman, J Westbrook, Z Feng, G Gilliland, T.N. Bhat. H Weissig, I.N. Shindyalov, P.E. Bourne: The Protein Data Bank. (2000) *Nucleic Acids Research.* **28**, 235-242.

***2.*** Needleman, S.B. and Wunsch, C.D. (1970) A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *J. Mol. Biol.* **48**: 442-453.

***3.*** Zvelebil, M. J. J. M., Barton, G. J., Taylor, W. R. & Sternberg, M. J. E. (1987). Prediction of protein secondary structure and active sites using the alignment of homologous sequences. *J. Mol. Biol.* **195**, 957-961.

***4.*** Manish C Saraf, Gregory L.Moore & Costas D. Maranas.(2003).Using multiple sequence correlation analysis to characterize functionally important protein regions. *Protein Engineering.* **16**, 397-406.

***5.*** Florencio Pazos, Manuela Helmer-Citterich, Gabriele Ausiello & Alfonso Valencia. (1997). Correlated mutations contain information about protein-protein interaction. *J. Mol. Biol.* **271**, 511-523.

***6.*** Chothia, C & Lesk, A.M.(1986). The relation between the divergence of sequences and structure in proteins. *EMBO J.* **5** 823-826.

***7.*** Lichtarge,O, Bourne, H.R. & Cohen, F. E. (1996). An Evolutionary Trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **257**, 342-358.

***8.*** Matteo Pellegrini, Edward M Marcotte, Micheal J Thompson, David Eisenberg and Todd O Yeates.(1999). Assigning protein functions by comparative genome

analysis: Protein phylogenetic profiles. *Proc. Natl. Acad. Sci, USA.* **96**, 4285-4288.

***9.*** Anton J Enright, Ioannis Illopouios, Nikos C. Kyrpids & Christos A Ouzounis. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*, **402**. 86-90.

***10.*** Susan Jones & Janet M Thornton. (1997). Analysis of Protein-Protein Interaction sites using Surface patches. *J. Mol. Biol.* **272**, 121-132.

***11.*** Maria Rosen, Shuo Liang Lin, Haim Wolfson & Ruth Nussinov. (1998). Molecular shape comparison in searches for active sites and functional similarity. *Protein Engineering.* **11**,263-277.

***12.*** Manfred J. Sippl. (1995). Knowledge based potentials for proteins. *Current Opinion In Structural Biology*. **5-**229-235.

***13.*** Teodoro, M. Philips, G.N.J. & Kavraki, L.E (2001). Molecular Docking, A problem with thousands of degrees of freedom, *IEEE International Conference of Robotics and Automation*, IEEE Press.

***14.*** Eleanor J. Gardiner, Peter Willet and Peter J Artymiuk. (2001). Protein Docking using a Genetic Algorithm, *Proteins: Structure, Function, and Genetics.* **44**, 44-56.

***15.*** Rarey M, Kramer B., Lengaur T. & Klebe G., A fast flexible docking method using an incremental construction Algorithm, *J Mol Biology*, **261**, 470-489.

***16.*** Rarey M, Kramer B. & Lengaur T., Multiple automatic base selection: Protein-ligand docking based on incremental construction with out manual intervention, *Journal of Computer Aided Drug Design*, **11**, 369-384.

***17.*** Sobolev V., Sorokine A., Prilusky J., Abola E.E., Edelman M. (1999) Automated analysis of interatomic contacts in proteins. *Bioinformatics,* **15**, 327-332.

***18.*** Einat AS, Dalit Naor, Haim J. Wolfson & Ruth Nussinov. (1997) Interchanges of spatially neighboring residues in structurally conserved environments. *Protein Engineering*, **10**, 1109-1122.

***19.*** PyMol, http://pymol.sourceforge.net/.

***20.*** Bowie, J. U., Luthy, R. & Eisenberg, D. (1991), A Method to Identify Protein Sequences That Fold into a Known Three-Dimensional Structure, *Science*, **253**, 164-170.

***21.*** G Yona, N Linial, N Tishby and M Linial, A Map of the Protein Space- An automatic hierarchical classification of All protein Sequences, *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology,* Montreal, Canada (June 28- July 1, 1998) pp 212-221

***22.*** A.H Liu and A Califano, Functional Classification of proteins by pattern discovery and top down clustering of primary sequences, *IBM Systems Journal*, issue 40-2.

***23.*** J Liang, H Edelsbrunner, and C Woodward. 1998. Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design. *Protein Science*, 7:1884-1897

***24.*** Steven C Bagley and Russ Altman, Characterizing the microenvironment surrounding protein sites. *Protein Science* 4 : 622-635

***25.*** FrunzKranz J, Round Robin Rule learning, (Technical Report OEFAI-TR-2001-02), Austrian Research Institute for Artificial Intelligence. Austria

***26.*** Friedman J (1996), Another approach to polychotmous classification. (Technical Report) Department of Statistics, Stanford University, Stanford CA.

***27.*** Camon E, Magrane M, Barrell D, Binns D, Fleischmann W, Kersey P, Mulder N, Oinn T, Maslen J, Cox A and Apweiler R**.** The Gene Ontology Annotation (GOA) project: implementation of GO in Swiss-Prot, TrEMBL and InterPro**.***Genome Research (4): 662-672 (2003).*

# Appendix A

CASR_BOVIN 1085 AA. EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID CELL CALCIUM-SENSING RECEPTOR).

MALYSCCWILLAFSTWCTSAYGPDQRAQKKGDIILGGLFPIHFGVAVKDQDLKSRPESVECIRYNFRGFRWLQ
AMIFAIEEINSSPALLPNMTLGYRIFDTCNTVSKALEATLSFVAQNKIDSLNLDEFCNCSEHIPSTIAVVGAT
GSGISTAVANLLGLFYIPQVSYASSSRLLSNKNQFKSFLRTIPNDEHQATAMADIIEYFRWNWVGTIAADDDY
GRPGIEKFREEAEERDICIDFSELISQYSDEEKIQQVVEVIQNSTAKVIVVFSSGPDLEPLIKEIVRRNITGR
IWLASEAWASSSLIAMPEYFHVVGGTIGFGLKAGQIPGFREFLQKVHPRKSVHNGFAKEFWEETFNCHLQEGA
KGPLPVDTFLRGHEEGGARLSNSPTAFRPLCTGEENISSVETPYMDYTHLRISYNVYLAVYSIAHALQDIYTC
IPGRGLFTNGSCADIKKVEAWQVLKHLRHLNFTSNMGEQVTFDECGDLAGNYSIINWHLSPEDGSIVFKEVGY
YNVYAKKGERLFINDEKILWSGFSREVPFSNCSRDCLAGTRKGIIEGEPTCCFECVECPDGEYSDETDASACD
KCPDDFWSNENHTSCIAKEIEFLSWTEPFGIALTLFAVLGIFLTAFVLGVFIKFRNTPIVKATNRELSYLLLF
SLLCCFSSSLFFIGEPQDWTCRLRQPAFGISFVLCISCILVKTNRVLLVFEAKIPTSFHRKWWGLNLQFLLVF
LCTFMQIVICAIWLNTAPPSSYRNHELEDEIIFITCHEGSLMALGFLIGYTCLLAAICFFFAFKSRKLPENFN
EAKFITFSMLIFFIVWISFIPAYASTYGKFVSAVEVIAILAASFGLLACIFFNKVYIILFKPSRNTIEEVRCS
TAAHAFKVAARATLRRSNVSRQRSSSLGGSTGSTPSSSISSKSNSEDPFPQQQPKRQKQPQPLALSPHNAQQP
QPRPPSTPQPQPQSQQPPRCKQKVIFGSGTVTFSLSFDEPQKTAVAHRNSTHQTSLEAQKNNDALTKHQALLP
LQCGETDSELTSQETGLQGPVGEDHQLEMEDPEEMSPALVVSNSRSFVISGGGSTVTENMLRS

CASR_HUMAN 1078 AA. EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID CELL CALCIUM-SENSING RECEPTOR).

MAFYSCCWVLLALTWHTSAYGPDQRAQKKGDIILGGLFPIHFGVAAKDQDLKSRPESVECIRYNFRGFRWLQA
MIFAIEEINSSPALLPNLTLGYRIFDTCNTVSKALEATLSFVAQNKIDSLNLDEFCNCSEHIPSTIAVVGATG
SGVSTAVANLLGLFYIPQVSYASSSRLLSNKNQFKSFLRTIPNDEHQATAMADIIEYFRWNWVGTIAADDDYG
RPGIEKFREEAEERDICIDFSELISQYSDEEEIQHVVEVIQNSTAKVIVVFSSGPDLEPLIKEIVRRNITGKI
WLASEAWASSSLIAMPQYFHVVGGTIGFALKAGQIPGFREFLKKVHPRKSVHNGFAKEFWEETFNCHLQEGAK
GPLPVDTFLRGHEESGDRFSNSSTAFRPLCTGDENISSVETPYIDYTHLRISYNVYLAVYSIAHALQDIYTCL
PGRGLFTNGSCADIKKVEAWQVLKHLRHLNFTNNMGEQVTFDECGDLVGNYSIINWHLSPEDGSIVFKEVGYY
NVYAKKGERLFINEEKILWSGFSREVPFSNCSRDCLAGTRKGIIEGEPTCCFECVECPDGEYSDETDASACNK
CPDDFWSNENHTSCIAKEIEFLSWTEPFGIALTLFAVLGIFLTAFVLGVFIKFRNTPIVKATNRELSYLLLFS
LLCCFSSSLFFIGEPQDWTCRLRQPAFGISFVLCISCILVKTNRVLLVFEAKIPTSFHRKWWGLNLQFLLVFL
CTFMQIVICVIWLYTAPPSSYRNQELEDEIIFITCHEGSLMALGFLIGYTCLLAAICFFFAFKSRKLPENFNE
AKFITFSMLIFFIVWISFIPAYASTYGKFVSAVEVIAILAASFGLLACIFFNKIYIILFKPSRNTIEEVRCST
AAHAFKVAARATLRRSNVSRKRSSSLGGSTGSTPSSSISSKSNSEDPFPQPERQKQQQPLALTQQEQQQQPLT
LPQQQRSQQQPRCKQKVIFGSGTVTFSLSFDEPQKNAMAHRNSTHQNSLEAQKSSDTLTRHQPLLPLQCGETD
LDLTVQETGLQGPVGGDQRPEVEDPEELSPALVVSSSQSFVISGGGSTVTENVVNS

CASR_RAT 1079 AA. EXTRACELLULAR CALCIUM-SENSING RECEPTOR PRECURSOR (CASR) (PARATHYROID CELL CALCIUM-SENSING RECEPTOR).

MASYSCCLALLALAWHSSAYGPDQRAQKKGDIILGGLFPIHFGVAAKDQDLKSRPESVECIRYNFRG
FRWLQAMIFAIEEINSSPSLLPNMTLGYRIFDTCNTVSKALEATLSFVAQNKIDSLNLDEFCNCSEHIPS
TIAVVGATGSGVSTAVANLLGLFYIPQVSYASSSRLLSNKNQYKSFLRTIPNDEHQATAMADIIEYFR
WNWVGTIAADDDYGRPGIEKFREEAEERDICIDFSELISQYSDEEEIQQVVEVIQNSTAKVIVVFSSGP
DLEPLIKEIVRRNITGRIWLASEAWASSSLIAMPEYFHVVGGTIGFGLKAGQIPGFREFLQKVHPRKSV
HNGFAKEFWEETFNCHLQEGAKGPLPVDTFVRSHEEGGNRLLNSSTAFRPLCTGDENINSVETPYMD
YEHLRISYNVYLAVYSIAHALQDIYTCLPGRGLFTNGSCADIKKVEAWQVLKHLRHLNFTNNMGEQ
VTFDECGDLVGNYSIINWHLSPEDGSIVFKEVGYYNVYAKKGERLFINEEKILWSGFSREVPFSNCSR
DCQAGTRKGIIEGEPTCCFECVECPDGEYSGETDASACDKCPDDFWSNENHTSCIAKEIEFLAWTEPF
GIALTLFAVLGIFLTAFVLGVFIKFRNTPIVKATNRELSYLLLFSLLCCFSSSLFFIGEPQDWTCRLRQP
AFGISFVLCISCILVKTNRVLLVFEAKIPTSFHRKWWGLNLQFLLVFLCTFMQILICIIWLYTAPPSSYR

NHELEDEIIFITCHEGSLMALGSLIGYTCLLAAICFFFAFKSRKLPENFNEAKFITFSMLIFFIVWISFIPA
YASTYGKFVSAVEVIAILAASFGLLACIFFNKVYIILFKPSRNTIEEVRSSTAAHAFKVAARATLRRPNI
SRKRSSSLGGSTGSIPSSSISSKSNSEDRFPQPERQKQQQPLSLTQQEQQQQPLTLHPQQQQQPQQPRC
KQKVIFGSGTVTFSLSFDEPQKNAMAHRNSMRQNSLEAQRSNDTLGRHQALLPLQCADADSEMTIQ
ETGLQGPMVGDHQPEMESSDEMSPALVMSTSRSFVISGGGSSVTENVLHS

MGR1_CAEEL 999 AA. PROBABLE METABOTROPIC GLUTAMATE RECEPTOR MGL-1.

MDKKWSLEQRWLHLLNQQFLDCLNHLFNHYRRLSTFQKPPSIIRHMFSVLALAIQILANVNVVAQTTEAVDLA
PPPKVRQIRIPGDILIGGVFPVHSKSLNGDEPCGEIAETRGVHRVEAMLYALDQINSQNDFLRGYKLGALILD
SCSNPAYALNQSLDFVRDMIGSSEASDYVCLDGSDPNLKKQSQKKNVAAVVGGSYSSVSVQLANLLRLFRIAQ
VSPASTNADLSDKNRFEYFARTVPSDDYQAMAMVEIAVKFKWSYVSLVYSADEYGELGADAFKKEARKKGICI
ALEERIQNKKESFTESINNLVQKLQPEKNVGATVVVLFVGTEYIPDILRYTAERMKLTSGAKKRIIWLASESW
DRNNDKYTAGDNRLAAQGAIVLMLASQKVPSFEEYFMSLHPGTEAFERNKWLRELWQVKYKCEFDTPPGSTAS
RCEDIKQSTEGFNADDKVQFVIDAVYAIAHGLQSMKQAICPDDAIENHWISRYSKQPEICHAMQNIDGSDFYQ
NYLLKVNFTGKTISIFSSFRLSPFSDIVGKRFRFSPQGDGPASYTILTYKPKSMDKKRRMTDDESSPSDYVEI
GHWSENNLTIYEKNLWWDPDHTPVSVCSLPCKIGFRKQLIKDEQCCWACSKCEDYEYLINETHCVGCEQGWWP
TKDRKGCFDLSLSQLKYMRWRSMYSLVPTILAVFGIIATLFVIVVYVIYNETPVVKASGRELSYILLISMIMC
YCMTFVLLSKPSAIVCAIKRTGIGFAFSCLYSAMFVKTNRIFRIFSTRSAQRPRFISPISQVVMTAMLAGVQL
IGSLIWLSVVPPGWRHHYPTRDQVVLTCNVPDHHFLYSLAYDGFLIVLCTTYAVKTRKVPENFNETKFIGFSM
YTTCVVWLSWIFFFFGTGSDFQIQTSSLCISISMSANVALACIFSPKLWIILFEKHKNVRKQEGESMLNKSSR
SLGNCSSRLCANSIDEPNQYTALLTDSTRRRSSRKTSQPTSTSSAHDTFL

MGR1_HUMAN 1194 AA. METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.

MVGLLLFFFPAIFLEVSLLPRSPGRKVLLAGASSQRSVARMDGDVIIGALFSVHHQPPAEKVPERKCGEIREQ
YGIQRVEAMFHTLDKINADPVLLPNITLGSEIRDSCWHSSVALEQSIEFIRDSLISIRDEKDGINRCLPDGQS
LPPGRTKKPIAGVIGPGSSSVAIQVQNLLQLFDIPQIAYSATSIDLSDKTLYKYFLRVVPSDTLQARAMLDIV
KRYNWTYVSAVHTEGNYGESGMDAFKELAAQEGLCIAHSDKIYSNAGEKSFDRLLRKLRERLPKARVVVCFCE
GMTVRGLLSAMRRLGVVGEFSLIGSDGWADRDEVIEGYEVEANGGITIKLQSPEVRSFDDYFLKLRLDTNTRN
PWFPEFWQHRFQCRLPGHLLENPNFKRICTGNESLEENYVQDSKMGFVINAIYAMAHGLQNMHHALCPGHVGL
CDAMKPIDGSKLLDFLIKSSFIGVSGEEVWFDEKGDAPGRYDIMNLQYTEANRYDYVHVGTWHEGVLNIDDYK
IQMNKSGVVRSVCSEPCLKGQIKVIRKGEVSCCWICTACKENEYVQDEFTCKACDLGWWPNADLTGCEPIPVR
YLEWSNIEPIIAIAFSCLGILVTLFVTLIFVLYRDTPVVKSSSRELCYIILAGIFLGYVCPFTLIAKPTTTSC
YLQRLLVGLSSAMCYSALVTKTNRIARILAGSKKKICTRKPRFMSAWAQVIIASILISVQLTLVVTLIIMEPP
MPILSYPSIKEVYLICNTSNLGVVAPLGYNGLLIMSCTYYAFKTRNVPANFNEAKYIAFTMYTTCIIWLAFVP
IYFGSNYKIITTCFAVSLSVTVALGCMFTPKMYIIIAKPERNVRSAFTTSDVVRMHVGDGKLPCRSNTFLNIF
RRKKAGAGNANSNGKSVSWSEPGGGQVPKGQHMWHRLSVHVKTNETACNQTAVIKPLTKSYQGSGKSLTFSDT
STKTLYNVEEEEDAQPIRFSPPGSPSMVVHRRVPSAATTPPLPPHLTAEETPLFLAEPALPKGLPPPLQQQQQ
PPPQQKSLMDQLQGVVSNFSTAIPDFHAVLAGPGGPGNGLRSLYPPPPPPQHLQMLPLQLSTFGEELVSPPAD
DDDDSERFKLLQEYVYEHEREGNTEEDELEEEEEDLQAASKLTPDDSPALTPPSPFRDSVASGSSVPSSPVSE
SVLCTPPNVSYASVILRDYKQSSSTL

MGR1_RAT 1199 AA. METABOTROPIC GLUTAMATE RECEPTOR 1 PRECURSOR.

MVRLLLIFFPMIFLEMSILPRMPDRKVLLAGASSQRSVARMDGDVIIGALFSVHHQPPAEKVPERKCGEIREQ
YGIQRVEAMFHTLDKINADPVLLPNITLGSEIRDSCWHSSVALEQSIEFIRDSLISIRDEKDGLNRCLPDGQT
LPPGRTKKPIAGVIGPGSSSVAIQVQNLLQLFDIPQIAYSATSIDLSDKTLYKYFLRVVPSDTLQARAMLDIV
KRYNWTYVSAVHTEGNYGESGMDAFKELAAQEGLCIAHSDKIYSNAGEKSFDRLLRKLRERLPKARVVVCFCE
GMTVRGLLSAMRRLGVVGEFSLIGSDGWADRDEVIEGYEVEANGGITIKLQSPEVRSFDDYFLKLRLDTNTRN
PWFPEFWQHRFQCRLPGHLLENPNFKKVCTGNESLEENYVQDSKMGFVINAIYAMAHGLQNMHHALCPGHVGL
CDAMKPIDGRKLLDFLIKSSFVGVSGEEVWFDEKGDAPGRYDIMNLQYTEANRYDYVHVGTWHEGVLNIDDYK
IQMNKSGMVRSVCSEPCLKGQIKVIRKGEVSCCWICTACKENEFVQDEFTCRACDLGWWPNAELTGCEPIPVR
YLEWSDIESIIAIAFSCLGILVTLFVTLIFVLYRDTPVVKSSSRELCYIILAGIFLGYVCPFTLIAKPTTTSC
YLQRLLVGLSSAMCYSALVTKTNRIARILAGSKKKICTRKPRFMSAWAQVIIASILISVQLTLVVTLIIMEPP
MPILSYPSIKEVYLICNTSNLGVVAPVGYNGLLIMSCTYYAFKTRNVPANFNEAKYIAFTMYTTCIIWLAFVP

IYFGSNYKIITTCFAVSLSVTVALGCMFTPKMYIIIAKPERNVRSAFTTSDVVRMHVGDGKLPCRSNTFLNIF
RRKKPGAGNANSNGKSVSWSEPGGRQAPKGQHVWQRLSVHVKTNETACNQTAVIKPLTKSYQGSGKSLTFSDA
STKTLYNVEEEDNTPSAHFSPPSSPSMVVHRRGPPVATTPPLPPHLTAEETPLFLADSVIPKGLPPPLPQQQP
QQPPPQQPPQQPKSLMDQLQGVVTNFGSGIPDFHAVLAGPGTPGNSLRSLYPPPPPPQHLQMLPLHLSTFQEE
SISPPGEDIDDDSERFKLLQEFVYEREGNTEEDELEEEEDLPTASKLTPEDSPALTPPSPFRDSVASGSSVPS
SPVSESVLCTPPNVTYASVILRDYKQSSSTL

MGR2_HUMAN 872 AA. METABOTROPIC GLUTAMATE RECEPTOR 2 PRECURSOR.

MGSLLALLALLPLWGAVAEGPAKKVLTLEGDLVLGGLFPVHQKGGPAEDCGPVNEHRGIQRLEAMLFALDRIN
RDPHLLPGVRLGAHILDSCSKDTHALEQALDFVRASLSRGADGSRHICPDGSYATHGDAPTAITGVIGGSYSD
VSIQVANLLRLFQIPQISYASTSAKLSDKSRYDYFARTVPPDFFQAKAMAEILRFFNWTYVSTEASEGDYGET
GIEAFELEARARNICVATSEKVGRAMSRAAFEGVVRALLQKPSARVAVLFTRSEDARELLAASQRLNASFTWV
ASDGWGALESVVAGSEGAAEGAITIELASYPISDFASYFQSLDPWNNSRNPWFREFWEQRFRCSFRQRDCAAH
SLRAVPFEQESKIMFVVNAVYAMAHALHNMHRALCPNTTRLCDAMRPVNGRRLYKDFVLNVKFDAPFRPADTH
NEVRFDRFGDGIGRYNIFTYLRAGSGRYRYQKVGYWAEGLTLDTSLIPWASPSAGPLAASRCSEPCLQNEVKS
VQPGEVCCWLCIPCQPYEYRLDEFTCADCGLGYWPNASLTGCFELPQEYIRWGDAWAVGPVTIACLGALATLF
VLGVFVRHNATPVVKASGRELCYILLGGVFLCYCMTFIFIAKPSTAVCTLRRLGLGTAFSVCYSALLTKTNRI
ARIFGGAREGAQRPRFISPASQVAICLALISGQLLIVVAWLVVEAPGTGKETAPERREVVTLRCNHRDASMLG
SLAYNVLLIALCTLYAFNTRKCPENFNEAKFIGFTMYTTCIIWLALLPIFYVTSSDYRVQTTTMCVSVSLSGS
VVLGCLFAPKLHIILFQPQKNVVSHRAPTSRFGSAAARASSSLGQGSGSQFVPTVCNGREVVDSTTSSL

MGR2_RAT 872 AA. METABOTROPIC GLUTAMATE RECEPTOR 2 PRECURSOR.

MESLLGFLALLLLWGAVAEGPAKKVLTLEGDLVLGGLFPVHQKGGPAEECGPVNEHRGIQRLEAMLFALDRIN
RDPHLLPGVRLGAHILDSCSKDTHALEQALDFVRASLSRGADGSRHICPDGSYATHSDAPTAVTGVIGGSYSD
VSIQVANLLRLFQIPQISYASTSAKLSDKSRYDYFARTVPPDFFQAKAMAEILRFFNWTYVSTVASEGDYGET
GIEAFELEARARNICVATSEKVGRAMSRAAFEGVVRALLQKPSARVAVLFTRSEDARELLAATQRLNASFTWV
ASDGWGALESVVAGSERAAEGAITIELASYPISDFASYFQSLDPWNNSRNPWFREFWEERFHCSFRQRDCAAH
SLRAVPFEQESKIMFVVNAVYAMAHALHNMHRALCPNTTHLCDAMRPVNGRRLYKDFVLNVKFDAPFRPADTD
DEVRFDRFGDGIGRYNIFTYLRAGSGRYRYQKVGYWAEGLTLDTSFIPWASPSAGPLPASRCSEPCLQNEVKS
VQPGEVCCWLCIPCQPYEYRLDEFTCADCGLGYWPNASLTGCFELPQEYIRWGDAWAVGPVTIACLGALATLF
VLGVFVRHNATPVVKASGRELCYILLGGVFLCYCMTFVFIAKPSTAVCTLRRLGLGTAFSVCYSALLTKTNRI
ARIFGGAREGAQRPRFISPASQVAICLALISGQLLIVAAWLVVEAPGTGKETAPERREVVTLRCNHRDASMLG
SLAYNVLLIALCTLYAFKTRKCPENFNEAKFIGFTMYTTCIIWLAFLPIFYVTSSDYRVQTTTMCVSVSLSGS
VVLGCLFAPKLHIILFQPQKNVVSHRAPTSRFGSAAPRASANLGQGSGSQFVPTVCNGREVVDSTTSSL

MGR3_HUMAN 877 AA. METABOTROPIC GLUTAMATE RECEPTOR 3 PRECURSOR.

MLTRLQVLTLALFSKGFLLSLGDHNFLRREIKIEGDLVLGGLFPINEKGTGTEECGRINEDRGIQRLEAMLFA
IDEINKDDYLLPGVKLGVHILDTCSRDTYALEQSLEFVRASLTKVDEAEYMCPDGSYAIQENIPLLIAGVIGG
SYSSVSIQVANLLRLFQIPQISYASTSAKLSDKSRYDYFARTVPPDFYQAKAMAEILRFFNWTYVSTVASEGD
YGETGIEAFEQEARLRNICIATAEKVGRSNIRKSYDSVIRELLQKPNARVVVLFMRSDDSRELIAAASRANAS
FTWVASDGWGAQESIIKGSEHVAYGAITLELASQPVRQFDRYFQSLNPYNNHRNPWFRDFWEQKFQCSLQNKR
NHRRVCDKHLAIDSSNYEQESKIMFVVNAVYAMAHALHKMQRTLCPNTTKLCDAMKILDGKKLYKDYLLKINF
TAPFNPNKDADSIVKFDTFGDGMGRYNVFNFQNVGGKYSYLKVGHWAETLSLDVNSIHWSRNSVPTSQCSDPC
APNEMKNMQPGDVCCWICIPCEPYEYLADEFTCMDCGSGQWPTADLTGCYDLPEDYIRWEDAWAIGPVTIACL
GFMCTCMVVTVFIKHNNTPLVKASGRELCYILLFGVGLSYCMTFFFIAKPSPVICALRRLGLGSSFAICYSAL
LTKTNCIARIFDGVKNGAQRPKFISPSSQVFICLGLILVQIVMVSVWLILEAPGTRRYTLAEKRETVILKCNV
KDSSMLISLTYDVILVILCTVYAFKTRKCPENFNEAKFIGFTMYTTCIIWLAFLPIFYVTSSDYRVQTTTMCI
SVSLSGFVVLGCLFAPKVHIILFQPQKNVVTHRLHLNRFSVSGTGTTYSQSSASTYVPTVCNGREVLDSTTSS
L

MGR3_RAT 879 AA. METABOTROPIC GLUTAMATE RECEPTOR 3 PRECURSOR.

MKMLTRLQILMLALFSKGFLLSLGDHNFMRREIKIEGDLVLGGLFPINEKGTGTEECGRINEDRGIQRLEAML
FAIDEINKDNYLLPGVKLGVHILDTCSRDTYALEQSLEFVRASLTKVDEAEYMCPDGSYAIQENIPLLIAGVI

GGSYSSVSIQVANLLRLFQIPQISYASTSAKLSDKSRYDYFARTVPPDFYQAKAMAEILRFFNWTYVSTVASE
GDYGETGIEAFEQEARLRNICIATAEKVGRSNIRKSYDSVIRELLQKPNARVVVLFMRSDDSRELIAAANRVN
ASFTWVASDGWGAQESIVKGSEHVAYGAITLELASHPVRQFDRYFQSLNPYNNHRNPWFRDFWEQKFQCSLQN
KRNHRQVCDKHLAIDSSNYEQESKIMFVVNAVYAMAHALHKMQRTLCPNTTKLCDAMKILDGKKLYKEYLLKI
NFTAPFNPNKGADSIVKFDTFGDGMGRYNVFNLQQTGGKYSYLKVGHWAETLSLDVDSIHWSRNSVPTSQCSD
PCAPNEMKNMQPGDVCCWICIPCEPYEYLVDEFTCMDCGPGQWPTADLSGCYNLPEDYIKWEDAWAIGPVTIA
CLGFLCTCIVITVFIKHNNTPLVKASGRELCYILLFGVSLSYCMTFFFIAKPSPVICALRRLGLGTSFAICYS
ALLTKTNCIARIFDGVKNGAQRPKFISPSSQVFICLGLILVQIVMVSVWLILETPGTRRYTLPEKRETVILKC
NVKDSSMLISLTYDVVLVILCTVYAFKTRKCPENFNEAKFIGFTMYTTCIIWLAFLPIFYVTSSDYRVQTTTM
CISVSLSGFVVLGCLFAPKVHIVLFQPQKNVVTHRLHLNRFSVSGTATTYSQSSASTYVPTVCNGREVLDSTT
SSL

MGR4_HUMAN 912 AA. METABOTROPIC GLUTAMATE RECEPTOR 4 PRECURSOR.

MPGKRGLGWWWARLPLCLLLSLYGPWMPSSLGKPKGHPHMNSIRIDGDITLGGLFPVHGRGSEGKPCGELKKE
KGIHRLEAMLFALDRINNDPDLLPNITLGARILDTCSRDTHALEQSLTFVQALIEKDGTEVRCGSGGPPIITK
PERVVGVIGASGSSVSIMVANILRLFKIPQISYASTAPDLSDNSRYDFFSRVVPSDTYQAQAMVDIVRALKWN
YVSTVASEGSYGESGVEAFIQKSREDGGVCIAQSVKIPREPKAGEFDKIIRRLLETSNARAVIIFANEDDIRR
VLEAARRANQTGHFFWMGSDSWGSKIAPVLHLEEVAEGAVTILPKRMSVRGFDRYFSSRTLDNNRRNIWFAEF
WEDNFHCKLSRHALKKGSHVKKCTNRERIGQDSAYEQEGKVQFVIDAVYAMGHALHAMHRDLCPGRVGLCPRM
DPVDGTQLLKYIRNVNFSGIAGNPVTFNENGDAPGRYDIYQYQLRNDSAEYKVIGSWTDHLHLRIERMHWPGS
GQQLPRSICSLPCQPGERKKTVKGMPCCWHCEPCTGYQYQVDRYTCKTCPYDMRPTENRTGCRPIPIIKLEWG
SPWAVLPLFLAVVGIAATLFVVITFVRYNDTPIVKASGRELSYVLLAGIFLCYATTFLMIAEPDLGTCSLRRI
FLGLGMSISYAALLTKTNRIYRIFEQGKRSVSAPRFISPASQLAITFSLISLQLLGICVWFVVDPSHSVVDFQ
DQRTLDPRFARGVLKCDISDLSLICLLGYSMLLMVTCTVYAIKTRGVPETFNEAKPIGFTMYTTCIVWLAFIP
IFFGTSQSADKLYIQTTTLTVSVSLSASVSLGMLYMPKVYIILFHPEQNVPKRKRSLKAVVTAATMSNKFTQK
GNFRPNGEAKSELCENLEAPALATKQTYVTYTNHAI

MGR4_RAT 912 AA. METABOTROPIC GLUTAMATE RECEPTOR 4 PRECURSOR.

MSGKGGWAWWWARLPLCLLLSLYAPWVPSSLGKPKGHPHMNSIRIDGDITLGGLFPVHGRGSEGKACGELKKE
KGIHRLEAMLFALDRINNDPDLLPNITLGARILDTCSRDTHALEQSLTFVQALIEKDGTEVRCGSGGPPIITK
PERVVGVIGASGSSVSIMVANILRLFKIPQISYASTAPDLSDNSRYDFFSRVVPSDTYQAQAMVDIVRALKWN
YVSTLASEGSYGESGVEAFIQKSRENGGVCIAQSVKIPREPKTGEFDKIIKRLLETSNARGIIIFANEDDIRR
VLEAARRANQTGHFFWMGSDSWGSKSAPVLRLEEVAEGAVTILPKRMSVRGFDRYFSSRTLDNNRRNIWFAEF
WEDNFHCKLSRHALKKGSHIKKCTNRERIGQDSAYEQEGKVQFVIDAVYAMGHALHAMHRDLCPGRVGLCPRM
DPVDGTQLLKYIRNVNFSGIAGNPVTFNENGDAPGRYDIYQYQLRNGSAEYKVIGSWTDHLHLRIERMQWPGS
GQQLPRSICSLPCQPGERKKTVKGMACCWHCEPCTGYQYQVDRYTCKTCPYDMRPTENRTSCQPIPIVKLEWD
SPWAVLPLFLAVVGIAATLFVVVTFVRYNDTPIVKASGRELSYVLLAGIFLCYATTFLMIAEPDLGTCSLRRI
FLGLGMSISYAALLTKTNRIYRIFEQGKRSVSAPRFISPASQLAITFILISLQLLGICVWFVVDPSHSVVDFQ
DQRTLDPRFARGVLKCDISDLSLICLLGYSMLLMVTCTVYAIKTRGVPETFNEAKPIGFTMYTTCIVWLAFIP
IFFGTSQSADKLYIQTTTLTVSVSLSASVSLGMLYMPKVYIILFHPEQNVPKRKRSLKAVVTAATMSNKFTQK
GNFRPNGEAKSELCENLETPALATKQTYVTYTNHAI

MGR5_HUMAN 1212 AA. METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.

MVLLLILSVLLLKEDVRGSAQSSERRVVAHMPGDIIIGALFSVHHQPTVDKVHERKCGAVREQYGIQRVEAML
HTLERINSDPTLLPNITLGCEIRDSCWHSAVALEQSIEFIRDSLISSEEEEGLVRCVDGSSSSFRSKKPIVGV
IGPGSSSVAIQVQNLLQLFNIPQIAYSATSMDLSDKTLFKYFMRVVPSDAQQARAMVDIVKRYNWTYVSAVHT
EGNYGESGMEAFKDMSAKEGICIAHSYKIYSNAGEQSFDKLLKKLTSHLPKARVVACFCEGMTVRGLLMAMRR
LGLAGEFLLLGSDGWADRYDVTDGYQREAVGGITIKLQSPDVKWFDDYYLKLRPETNHRNPWFQEFWQHRFQC
RLEGFPQENSKYNKTCNSSLTLKTHHVQDSKMGFVINAIYSMAYGLHNMQMSLCPGYAGLCDAMKPIDGRKLL
ESLMKTNFTGVSGDTILFDENGDSPGRYEIMNFKEMGKDYFDYINVGSWDNGELKMDDDEVWSKKSNIIRSVC
SEPCEKGQIKVIRKGEVSCCWTCTPCKENEYVFDEYTCKACQLGSWPTDDLTGCDLIPVQYLRWGDPEPIAAV
VFACLGLLATLFVTVVFIIYRDTPVVKSSSRELCYIILAGICLGYLCTFCLIAKPKQIYCYLQRIGIGLSPAM
SYSALVTKTNRIARILAGSKKKICTKKPRFMSACAQLVIAFILICIQLGIIVALFIMEPPDIMHDYPSIREVY
LICNTTNLGVVTPLGYNGLLILSCTFYAFKTRNVPANFNEAKYIAFTMYTTCIIWLAFVPIYFGSNYKIITMC

FSVSLSATVALGCMFVPKVYIILAKPERNVRSAFTTSTVVRMHVGDGKSSSAASRSSSLVNLWKRRGSSGETL
RYKDRRLAQHKSEIECFTPKGSMGNGGRATMSSSNGKSVTWAQNEKSSRGQHLWQRLSIHINKKENPNQTAVI
KPFPKSTESRGLGAGAGAGGSAGGVGATGGAGCAGAGPGGPESPDAGPKALYDVAEAEEHFPAPARPRSPSPI
STLSHRAGSASRTDDDVPSLHSEPVARSSSSQGSLMEQISSVVTRFTANISELNSMMLSTAAPSPGVGAPLCS
SYLIPKEIQLPTTMTTFAEIQPLPAIEVTGGAQPAAGAQAAGDAARESPAAGPEAAAAKPDLEELVALTPPSP
FRDSVDSGSTTPNSPVSESALCIPSSPKYDTLIIRDYTQSSSSL

MGR5_RAT 1203 AA. METABOTROPIC GLUTAMATE RECEPTOR 5 PRECURSOR.

MVLLLILSVLLLKEDVRGSAQSSERRVVAHMPGDIIIGALFSVHHQPTVDKVHERKCGAVREQYGIQRVEAML
HTLERINSDPTLLPNITLGCEIRDSCWHSAVALEQSIEFIRDSLISSEEEEGLVRCVDGSSSFRSKKPIVGVI
GPGSSSVAIQVQNLLQLFNIPQIAYSATSMDLSDKTLFKYFMRVVPSDAQQARAMVDIVKRYNWTYVSAVHTE
GNYGESGMEAFKDMSAKEGICIAHSYKIYSNAGEQSFDKLLKKLRSHLPKARVVACFCEGMTVRGLLMAMRRL
GLAGEFLLLGSDGWADRYDVTDGYQREAVGGITIKLQSPDVKWFDDYYLKLRPETNLRNPWFQEFWQHRFQCR
LEGFAQENSKYNKTCNSSLTLRTHHVQDSKMGFVINAIYSMAYGLHNMQMSLCPGYAGLCDAMKPIDGRKLLD
SLMKTNFTGVSGDMILFDENGDSPGRYEIMNFKEMGKDYFDYINVGSWDNGELKMDDDEVWSKKNNIIRSVCS
EPCEKGQIKVIRKGEVSCCWTCTPCKENEYVFDEYTCKACQLGSWPTDDLTGCDLIPVQYLRWGDPEPIAAVV
FACLGLLATLFVTVIFIIYRDTPVVKSSSRELCYIILAGICLGYLCTFCLIAKPKQIYCYLQRIGIGLSPAMS
YSALVTKTNRIARILAGSKKKICTKKPRFMSACAQLVIAFILICIQLGIIVALFIMEPPDIMHDYPSIREVYL
ICNTTNLGVVTPLGYNGLLILSCTFYAFKTRNVPANFNEAKYIAFTMYTTCIIWLAFVPIYFGSNYKIITMCF
SVSLSATVALGCMFVPKVYIILAKPERNVRSAFTTSTVVRMHVGDGKSSSAASRSSSLVNLWKRRGSSGETLR
YKDRRLAQHKSEIECFTPKGSMGNGGRATMSSSNGKSVTWAQNEKSTRGQHLWQRLSVHINKKENPNQTAVIK
PFPKSTENRGPGAAAGGGSGPGVAGAGNAGCTATGGPEPPDAGPKALYDVAEAEESFPAAARPRSPSPISTLS
HLAGSAGRTDDDAPSLHSETAARSSSSQGSLMEQISSVVTRFTANISELNSMMLSTAATPGPPGTPICSSYLI
PKEIQLPTTMTTFAEIQPLPAIEVTGGAQGATGVSPAQETPTGAESAPGKPDLEELVALTPPSPFRDSVDSGS
TTPNSPVSESALCIPSSPKYDTLIIRDYTQSSSSL

MGR6_HUMAN 877 AA. METABOTROPIC GLUTAMATE RECEPTOR 6 PRECURSOR.

MARPRRAREPLLVALLPLAWLAQAGLARAAGSVRLAGGLTLGGLFPVHARGAAGRACGPLKKEQGVHRLEAML
YALDRVNADPELLPGVRLGARLLDTCSRDTYALEQALSFVQALIRGRGDGDEVGVRCPGGVPPLRPAPPERVV
AVVGASASSVSIMVANVLRLFAIPQISYASTAPELSDSTRYDFFSRVVPPDSYQAQAMVDIVRALGWNYVSTL
ASEGNYGESGVEAFVQISREAGGVCIAQSIKIPREPKPGEFSKVIRRLMETPNARGIIIFANEDDIRRVLEAA
RQANLTGHFLWVGSDSWGAKTSPILSLEDVAVGAITILPKRASIDGFDQYFMTRSLENNRRNIWFAEFWEENF
NCKLTSSGTQSDDSTRKCTGEERIGRDSTYEQEGKVQFVIDAVYAIAHALHSMHQALCPGHTGLCPAMEPTDG
RMLLQYIRAVRFNGSAGTPVMFNENGDAPGRYDIFQYQATNGSASSGGYQAVGQWAETLRLDVEALQWSGDPH
EVPSSLCSLPCGPGERKKMVKGVPCCWHCEACDGYRFQVDEFTCEACPGDMRPTPNHTGCRPTPVVRLSWSSP
WAAPPLLLAVLGIVATTTVVATFVRYNNTPIVRASGRELSYVLLTGIFLIYAITFLMVAEPGAAVCAARRLFL
GLGTTLSYSALLTKTNRIYRIFEQGKRSVTPPPFISPTSQLVITFSLTSLQVVGMIAWLGARPPHSVIDYEEQ
RTVDPEQARGVLKCDMSDLSLIGCLGYSLLLMVTCTVYAIKARGVPETFNEAKPIGFTMYTTCIIWLAFVPIF
FGTAQSAEKIYIQTTTLTVSLSLSASVSLGMLYVPKTYVILFHPEQNVQKRKRSLKATSTVAAPPKGEDAEAH
K

MGR6_RAT 871 AA. METABOTROPIC GLUTAMATE RECEPTOR 6 PRECURSOR.

MGRLPVLLLWLAWWLSQAGIACGAGSVRLAGGLTLGGLFPVHARGAAGRACGALKKEQGVHRLEAMLYALDRV
NADPELLPGVRLGARLLDTCSRDTYALEQALSFVQALIRGRGDGDEASVRCPGGVPPLRSAPPERVVAVVGAS
ASSVSIMVANVLRLFAIPQISYASTAPELSDSTRYDFFSRVVPPDSYQAQAMVDIVRALGWNYVSTLASEGNY
GESGVEAFVQISREAGGVCIAQSIKIPREPKPGEFHKVIRRLMETPNARGIIIFANEDDIRRVLEATRQANLT
GHFLWVGSDSWGSKISPILNLEEEAVGAITILPKRASIDGFDQYFMTRSLENNRRNIWFAEFWEENFNCKLTS
SGGQSDDSTRKCTGEERIGQDSAYEQEGKVQFVIDAVYAIAHALHSMHQALCPGHTGLCPAMEPTDGRTLLHY
IRAVRFNGSAGTPVMFNENGDAPGRYDIFQYQATNGSASSGGYQAVGQWAEALRLDMEVLRWSGDPHEVPPSQ
CSLPCGPGERKKMVKGVPCCWHCEACDGYRFQVDEFTCEACPGDMRPTPNHTGCRPTPVVRLTWSSPWAALPL
LLAVLGIMATTTIMATFMRHNDTPIVRASGRELSYVLLTGIFLIYAITFLMVAEPCAAICAARRLLLGLGTTL
SYSALLTKTNRIYRIFEQGKRSVTPPPFISPTSQLVITFGLTSLQVVGVIAWLGAQPPHSVIDYEEQRTVDPE
QARGVLKCDMSDLSLIGCLGYSLLLMVTCTVYAIKARGVPETFNEAKPIGFTMYTTCIIWLAFVPIFFGTAQS
AEKIYIQTTTLTVSLSLSASVSLGMLYVPKTYVILFHPEQNVQKRKRSLKKTSTMAAPPQNENAEDAK

MGR7_HUMAN 915 AA. METABOTROPIC GLUTAMATE RECEPTOR 7 PRECURSOR.

MVQLRKLLRVLTLMKFPCCVLEVLLCALAAAARGQEMYAPHSIRIEGDVTLGGLFPVHAKGPSGVPCGDIKRE
NGIHRLEAMLYALDQINSDPNLLPNVTLGARILDTCSRDTYALEQSLTFVQALIQKDTSDVRCTNGEPPVFVK
PEKVVGVIGASGSSVSIMVANILRLFQIPQISYASTAPELSDDRRYDFFSRVVPPDSFQAQAMVDIVKALGWN
YVSTLASEGSYGEKGVESFTQISKEAGGLCIAQSVRIPQERKDRTIDFDRIIKQLLDTPNSRAVVIFANDEDI
KQILAAAKRADQVGHFLWVGSDSWGSKINPLHQHEDIAEGAITIQPKRATVEGFDAYFTSRTLENNRRNVWFA
EYWEENFNCKLTISGSKKEDTDRKCTGQERIGKDSNYEQEGKVQFVIDAVYAMAHALHHMNKDLCADYRGVCP
EMEQAGGKKLLKYIRNVNFNGSAGTPVMFNKNGDAPGRYDIFQYQTTNTSNPGYRLIGQWTDELQLNIEDMQW
GKGVREIPASVCTLPCKPGQRKKTQKGTPCCWTCEPCDGYQYQFDEMTCQHCPYDQRPNENRTGCQDIPIIKL
EWHSPWAVIPVFLAMLGIIATIFVMATFIRYNDTPIVRASGRELSYVLLTGIFLCYIITFLMIAKPDVAVCSF
RRVFLGLGMCISYAALLTKTNRIYRIFEQGKKSVTAPRLISPTSQLAITSSLISVQLLGVFIWFGVDPPNIII
DYDEHKTMNPEQARGVLKCDITDLQIICSLGYSILLMVTCTVYAIKTRGVPENFNEAKPIGFTMYTTCIVWLA
FIPIFFGTAQSAEKLYIQTTTLTISMNLSASVALGMLYMPKVYIIIFHPELNVQKRKRSFKAVVTAATMSSRL
SHKPSDRPNGEAKTELCENVDPNSPAAKKKYVSYNNLVI

MGR7_RAT 915 AA. METABOTROPIC GLUTAMATE RECEPTOR 7 PRECURSOR.

MVQLGKLLRVLTLMKFPCCVLEVLLCVLAAAARGQEMYAPHSIRIEGDVTLGGLFPVHAKGPSGVPCGDIKRE
NGIHRLEAMLYALDQINSDPNLLPNVTLGARILDTCSRDTYALEQSLTFVQALIQKDTSDVRCTNGEPPVFVK
PEKVVGVIGASGSSVSIMVANILRLFQIPQISYASTAPELSDDRRYDFFSRVVPPDSFQAQAMVDIVKALGWN
YVSTLASEGSYGEKGVESFTQISKEAGGLCIAQSVRIPQERKDRTIDFDRIIKQLLDTPNSRAVVIFANDEDI
KQILAAAKRADQVGHFLWVGSDSWGSKINPLHQHEDIAEGAITIQPKRATVEGFDAYFTSRTLENNRRNVWFA
EYWEENFNCKLTISGSKKEDTDRKCTGQERIGKDSNYEQEGKVQFVIDAVYAMAHALHHMNKDLCADYRGVCP
EMEQAGGKKLLKYIRHVNFNGSAGTPVMFNKNGDAPGRYDIFQYQTTNTTNPGYRLIGQWTDELQLNIEDMQW
GKGVREIPSSVCTLPCKPGQRKKTQKGTPCCWTCEPCDGYQYQFDEMTCQHCPYDQRPNENRTGCQNIPIIKL
EWHSPWAVIPVFLAMLGIIATIFVMATFIRYNDTPIVRASGRELSYVLLTGIFLCYIITFLMIAKPDVAVCSF
RRVFLGLGMCISYAALLTKTNRIYRIFEQGKKSVTAPRLISPTSQLAITSSLISVQLLGVFIWFGVDPPNIII
DYDEHKTMNPEQARGVLKCDITDLQIICSLGYSILLMVTCTVYAIKTRGVPENFNEAKPIGFTMYTTCIVWLA
FIPIFFGTAQSAEKLYIQTTTLTISMNLSASVALGMLYMPKVYIIIFHPELNVQKRKRSFKAVVTAATMSSRL
SHKPSDRPNGEAKTELCENVDPNSPAAKKKYVSYNNLVI

MGR8_HUMAN 908 AA. METABOTROPIC GLUTAMATE RECEPTOR 8 PRECURSOR.

MVCEGKRSASCPCFFLLTAKFYWILTMMQRTHSQEYAHSIRVDGDIILGGLFPVHAKGERGVPCGELKKEKGI
HRLEAMLYAIDQINKDPDLLSNITLGVRILDTCSRDTYALEQSLTFVQALIEKDASDVKCANGDPPIFTKPDK
ISGVIGAAASSVSIMVANILRLFKIPQISYASTAPELSDNTRYDFFSRVVPPDSYQAQAMVDIVTALGWNYVS
TLASEGNYGESGVEAFTQISREIGGVCIAQSQKIPREPRPGEFEKIIKRLLETPNARAVIMFANEDDIRRILE
AAKKLNQSGHFLWIGSDSWGSKIAPVYQQEEIAEGAVTILPKRASIDGFDRYFRSRTLANNRRNVWFAEFWEE
NFGCKLGSHGKRNSHIKKCTGLERIARDSSYEQEGKVQFVIDAVYSMAYALHNMHKDLCPGYIGLCPRMSTID
GKELLGYIRAVNFNGSAGTPVTFNENGDAPGRYDIFQYQITNKSTEYKVIGHWTNQLHLKVEDMQWAHREHTH
PASVCSLPCKPGERKKTVKGVPCCWHCERCEGYNYQVDELSCELCPLDQRPNMNRTGCQLIPIIKLEWHSPWA
VVPVFVAILGIIATTFVIVTFVRYNDTPIVRASGRELSYVLLTGIFLCYSITFLMIAAPDTIICSFRRVFLGL
GMCFSYAALLTKTNRIHRIFEQGKKSVTAPKFISPASQLVITFSLISVQLLGVFVWFVVDPPHIIIDYGEQRT
LDPEKARGVLKCDISDLSLICSLGYSILLMVTCTVYANKTRGVPETFNEAKPIGFTMYTTCIIWLAFIPIFFG
TAQSAEKMYIQTTTLTVSMSLSASVSLGMLYMPKVYIIIFHPEQNVQKRKRSFKAVVTAATMQSKLIQKGNDR
PNGEVKSELCESLETNTSSTKTTYISYSNHSI

MGR8_MOUSE 908 AA. METABOTROPIC GLUTAMATE RECEPTOR 8 PRECURSOR.

MVCEGKRSTSCPCFFLLTAKFYWILTMMQRTHSQEYAHSIRLDGDIILGGLFPVHAKGERGVPCGDLKKEKGI
HRLEAMLYAIDQTNKDPDLLSNITLGVRILDTCSRDTYALEQSLTFVQALIEKDASDVKCANGDPPIFTKPDK
ISGVIGAAASSVSIMVANILRLFKIPQISYASTAPELSDNTRYDFFSRVVPPDSYQAQAMVDIVTALGWNYVS
TLASEGNYGESGVEAFTQISREIGGVCIAQSQKIPREPRPGEFEKIIKRLLETPNARAVIMFANEDDIRGILE
AAKKLNQSGHFLWIGSDSWGSKIAPVYQQEEIAEGAVTILPKRASIDGFDRYFRSRTLANNRRNVWFAEFSEG
NFGCKSGSHGKRNSHIKKCTGLERIARDSSYEQEGKVQFVIDAVYSMAYALHNMHKELCPGYIGLCPRMVTID
GKELLGYIRAVNFNGSAGTPVTFNENGDAPGRYDIFQYQINNKSTEYKIIGHWTNQLHLKVEDMQWANREHTH

PASVCSLPCKPGERKKTVKGVPCCWHCGRCEGYNYQVDELSCELCPLDQRPNINRTGCQRIPIIKLEWHSPWA
VVPVLIAILGIIATTFVIVTFVRYNDTPIVRASGRELSYVLLTGIFLCYSITFLMIAAPDTIICSFRRIFLGL
GMCFSYAALLTKTNRIHRIFEQGKKSVTAPKFISPASQLVITFSLISVQLLGVFVWFVVDPPHTIIDYGEQRT
LDPENARGVLKCDISDLSLICSLGYSILLMVTCTVYAIKTRGVPETFNEAKPIGFTMYTTCIIWLAFIPIFFG
TAQSAEKMYIQTTTLTVSMSLSASVSLGMLYMPKVYIIIFHPEQNVQKRKRSFKAVVTAATMQSKLIQKGNDR
PNGEVKSELCESLETNTSSTKTTYISYSDHSI

MGR8_RAT 908 AA. METABOTROPIC GLUTAMATE RECEPTOR 8 PRECURSOR.

MVCEGKRLASCPCFFLLTAKFYWILTMMQRTHSQEYAHSIRVDGDIILGGLFPVHAKGERGVPCGEL
KKEKGIHRLEAMLYAIDQINKDPDLLSNITLGVRILDTCSRDTYALEQSLTFVQALIEKDASDVKCAN
GDPPIFTKPDKISGVIGAAASSVSIMVANILRLFKIPQISYASTAPELSDNTRYDFFSRVVPPDSYQAQA
MVDIVTALGWNYVSTLASEGNYGESGVEAFTQISREIGGVCIAQSQKIPREPRPGEFEKIIKRLLETPN
ARAVIMFANEDDIRRILEAAKKLNQSGHFLWIGSDSWGSKIAPVYQQEEIAEGAVTILPKRASIDGFD
RYFRSRTLANNRRNVWFAEFWEENFGCKLGSHGKRNSHIKKCTGLERIARDSSYEQEGKVQFVIDA
VYSMAYALHNMHKERCPGYIGLCPRMVTIDGKELLGYIRAVNFNGSAGTPVTFNENGDAPGRYDIF
QYQINNKSTEYKIIGHWTNQLHLKVEDMQWANREHTHPASVCSLPCKPGERKKTVKGVPCCWHCE
RCEGYNYQVDELSCELCPLDQRPNINRTGCQRIPIIKLEWHSPWAVVPVFIAILGIIATTFVIVTFVRYN
DTPIVRASGRELSYVLLTGIFLCYSITFLMIAAPDTIICSFRRIFLGLGMCFSYAALLTKTNRIHRIFEQG
KKSVTAPKFISPASQLVITFSLISVQLLGVFVWFVVDPPHTIIDYGEQRTLDPENARGVLKCDISDLSLI
CSLGYSILLMVTCTVYAIKTRGVPETFNEAKPIGFTMYTTCIIWLAFIPIFFGTAQSAEKMYIQTTTLTV
SMSLSASVSLGMLYMPKVYIIIFHPEQNVQKRKRSFKAVVTAATMQSKLIQKGNDRPNGEVKSELCE
SLETNTSSTKTTYISYSNHSI

# Appendix B (16 Class Classification)

***1.*** Classifying “1swp” as biotin binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | ORN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | BTN | 9PP-TDG | TDG |
| 5 | BTN-PQQ | BTN | 9PP-NTM | 9PP |
| 6 | BTN-XLS | BTN | 9PP-STY | STY |
| 7 | BTN-RET | BTN | PQQ-XLS | PQQ |
| 8 | BTN-2GP | BTN | PQQ-RET | PQQ |
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BTN | PQQ-MGN | MGN |
| 11 | BTN-FLP | BTN | PQQ-BOX | BOX |
| 12 | BTN-MTX | BTN | PQQ-FLP | FLP |
| 13 | BTN-TDG | BTN | PQQ-MTX | PQQ |
| 14 | BTN-NTM | NTM | PQQ-TDG | PQQ |
| 15 | BTN-STY | BTN | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | RET |

| | | | | |
|---|---|---|---|---|
| 18 | ORN-9PP | ORN | XLS-2GP | XLS |
| 19 | ORN-PQQ | ORN | XLS-MGN | XLS |
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | 2GP | XLS-MTX | XLS |
| 23 | ORN-MGN | MGN | XLS-TDG | TDG |
| 24 | ORN-BOX | ORN | XLS-NTM | NTM |
| 25 | ORN-FLP | FLP | XLS-STY | XLS |
| 26 | ORN-MTX | MTX | RET-2GP | RET |
| 27 | ORN-TDG | ORN | RET-MGN | RET |
| 28 | ORN-NTM | NTM | RET-BOX | BOX |
| 29 | ORN-STY | ORN | RET-FLP | FLP |
| 30 | 4MO-LVS | 4MO | RET-MTX | RET |
| 31 | 4MO-9PP | 9PP | RET-TDG | TDG |
| 32 | 4MO-PQQ | PQQ | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | RET |
| 34 | 4MO-RET | 4MO | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | MGN | 2GP-FLP | FLP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | FLP | 2GP-TDG | TDG |
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |

| | Type of Classifier | Class Selected | Type of Classifier | Class Selected |
|---|---|---|---|---|
| 41 | 4MO-NTM | 4MO | MGN-BOX | MGN |
| 42 | 4MO-STY | 4MO | MGN-MTX | MTX |
| 43 | LVS-9PP | 9PP | MGN-FLP | FLP |
| 44 | LVS-PQQ | LVS | MGN-TDG | TDG |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | MGN |
| 47 | LVS-2GP | 2GP | BOX-FLP | BOX |
| 48 | LVS-MGN | LVS | BOX-MTX | BOX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | FLP | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | TDG | FLP-MTX | MTX |
| 53 | LVS-NTM | LVS | FLP-TDG | TDG |
| 54 | LVS-STY | LVS | FLP-NTM | NTM |
| 55 | 9PP-PQQ | PQQ | FLP-STY | STY |
| 56 | 9PP-XLS | XLS | MTX-TDG | TDG |
| 57 | 9PP-RET | 9PP | MTX-NTM | MTX |
| 58 | 9PP-2GP | 9PP | MTX-STY | MTX |
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | STY | TDG-STY | STY |

The corresponding voting table for the classes is given as

| Class | Votes | Class | Votes |
|---|---|---|---|
| BTN | 11 | MGN | 7 |
| ORN | 7 | BOX | 8 |
| 4MO | 5 | FLP | 9 |
| 9PP | 6 | LVS | 8 |
| PQQ | 7 | MTX | 9 |
| XLS | 7 | TDG | 9 |
| RET | 7 | NTM | 8 |
| 2GP | 6 | STY | 6 |

***2.*** Classifying “1uaz” as retinal binding protein

| | Type of Classifier | Class Selected | Type of Classifier | Class Selected |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | 9PP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | TDG |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | BTN | 9PP-STY | 9PP |
| 7 | BTN-RET | RET | PQQ-XLS | PQQ |

| | | | | |
|---|---|---|---|---|
| 8 | BTN-2GP | 2GP | PQQ-RET | RET |
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BOX | PQQ-MGN | MGN |
| 11 | BTN-FLP | FLP | PQQ-BOX | BOX |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | TDG | PQQ-MTX | MTX |
| 14 | BTN-NTM | BTN | PQQ-TDG | PQQ |
| 15 | BTN-STY | BTN | PQQ-NTM | NTM |
| 16 | ORN-4MO | 4MO | PQQ-STY | PQQ |
| 17 | ORN-LVS | ORN | XLS-RET | XLS |
| 18 | ORN-9PP | ORN | XLS-2GP | 2GP |
| 19 | ORN-PQQ | PQQ | XLS-MGN | XLS |
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | TDG |
| 24 | ORN-BOX | BOX | XLS-NTM | XLS |
| 25 | ORN-FLP | ORN | XLS-STY | XLS |
| 26 | ORN-MTX | MTX | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | RET |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | ORN | RET-FLP | FLP |
| 30 | 4MO-LVS | LVS | RET-MTX | RET |

| | | | | |
|---|---|---|---|---|
| 31 | 4MO-9PP | 4MO | RET-TDG | RET |
| 32 | 4MO-PQQ | 4MO | RET-NTM | RET |
| 33 | 4MO-XLS | XLS | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | BOX |
| 36 | 4MO-MGN | 4MO | 2GP-FLP | FLP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | 2GP |
| 38 | 4MO-FLP | FLP | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | MTX | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | TDG | 2GP-STY | 2GP |
| 41 | 4MO-NTM | 4MO | MGN-BOX | MGN |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | MGN |
| 44 | LVS-PQQ | PQQ | MGN-TDG | MGN |
| 45 | LVS-XLS | XLS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | MGN |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | MGN | BOX-MTX | BOX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | LVS | BOX-NTM | NTM |
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | TDG | FLP-MTX | FLP |
| 53 | LVS-NTM | LVS | FLP-TDG | FLP |

| | | | | |
|---|---|---|---|---|
| 54 | LVS-STY | STY | FLP-NTM | NTM |
| 55 | 9PP-PQQ | PQQ | FLP-STY | STY |
| 56 | 9PP-XLS | 9PP | MTX-TDG | TDG |
| 57 | 9PP-RET | RET | MTX-NTM | NTM |
| 58 | 9PP-2GP | 9PP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | STY | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 5 | MGN | 8 |
| ORN | 6 | BOX | 8 |
| 4MO | 6 | FLP | 8 |
| 9PP | 6 | LVS | 7 |
| PQQ | 7 | MTX | 8 |
| XLS | 7 | TDG | 8 |
| RET | 12 | NTM | 7 |
| 2GP | 9 | STY | 8 |

***3.*** Classifying “1flg” as PQQ binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | BTN | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | TDG |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | 9PP |
| 6 | BTN-XLS | BTN | 9PP-STY | 9PP |
| 7 | BTN-RET | BTN | PQQ-XLS | PQQ |
| 8 | BTN-2GP | 2GP | PQQ-RET | RET |
| 9 | BTN-MGN | BTN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BTN | PQQ-MGN | MGN |
| 11 | BTN-FLP | FLP | PQQ-BOX | PQQ |
| 12 | BTN-MTX | BTN | PQQ-FLP | PQQ |
| 13 | BTN-TDG | BTN | PQQ-MTX | PQQ |
| 14 | BTN-NTM | NTM | PQQ-TDG | TDG |
| 15 | BTN-STY | STY | PQQ-NTM | PQQ |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | ORN | XLS-RET | RET |

| | | | | |
|---|---|---|---|---|
| 18 | ORN-9PP | 9PP | XLS-2GP | XLS |
| 19 | ORN-PQQ | PQQ | XLS-MGN | XLS |
| 20 | ORN-XLS | XLS | XLS-BOX | XLS |
| 21 | ORN-RET | ORN | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | MGN | XLS-TDG | XLS |
| 24 | ORN-BOX | BOX | XLS-NTM | XLS |
| 25 | ORN-FLP | ORN | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | 2GP |
| 27 | ORN-TDG | TDG | RET-MGN | MGN |
| 28 | ORN-NTM | ORN | RET-BOX | RET |
| 29 | ORN-STY | STY | RET-FLP | RET |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |
| 32 | 4MO-PQQ | PQQ | RET-NTM | RET |
| 33 | 4MO-XLS | 4MO | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | BOX |
| 36 | 4MO-MGN | 4MO | 2GP-FLP | 2GP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | 4MO | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | 4MO | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | TDG | 2GP-STY | STY |

| | | | | |
|---|---|---|---|---|
| 41 | 4MO-NTM | 4MO | MGN-BOX | MGN |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | MGN |
| 44 | LVS-PQQ | PQQ | MGN-TDG | MGN |
| 45 | LVS-XLS | XLS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | MGN |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | LVS | BOX-MTX | BOX |
| 49 | LVS-BOX | BOX | BOX-TDG | TDG |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | LVS | FLP-MTX | FLP |
| 53 | LVS-NTM | NTM | FLP-TDG | FLP |
| 54 | LVS-STY | STY | FLP-NTM | FLP |
| 55 | 9PP-PQQ | PQQ | FLP-STY | STY |
| 56 | 9PP-XLS | 9PP | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |
| 59 | 9PP-MGN | 9PP | TDG-NTM | NTM |
| 60 | NTM-STY | NTM | TDG-STY | TDG |

The corresponding voting table for the classes is given as

| Class | Votes | Class | Votes |
|---|---|---|---|
| BTN | 9 | MGN | 8 |
| ORN | 7 | BOX | 7 |
| 4MO | 6 | FLP | 6 |
| 9PP | 7 | LVS | 6 |
| PQQ | 12 | MTX | 8 |
| XLS | 8 | TDG | 7 |
| RET | 8 | NTM | 5 |
| 2GP | 7 | STY | 9 |

***4.*** Classifying “3xis” as XLS binding protein

| | Type of Classifier | Class Selected | Type of Classifier | Class Selected |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | 9PP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | TDG |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | BTN | 9PP-STY | 9PP |
| 7 | BTN-RET | BTN | PQQ-XLS | XLS |
| 8 | BTN-2GP | BTN | PQQ-RET | PQQ |

| | | | | |
|---|---|---|---|---|
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BTN | PQQ-MGN | PQQ |
| 11 | BTN-FLP | FLP | PQQ-BOX | PQQ |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | BTN | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | TDG |
| 15 | BTN-STY | STY | PQQ-NTM | PQQ |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | XLS |
| 19 | ORN-PQQ | PQQ | XLS-MGN | MGN |
| 20 | ORN-XLS | XLS | XLS-BOX | XLS |
| 21 | ORN-RET | ORN | XLS-FLP | XLS |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | MGN | XLS-TDG | TDG |
| 24 | ORN-BOX | BOX | XLS-NTM | XLS |
| 25 | ORN-FLP | FLP | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | 2GP |
| 27 | ORN-TDG | ORN | RET-MGN | RET |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | ORN | RET-FLP | FLP |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | RET |

| | | | | |
|---|---|---|---|---|
| 32 | 4MO-PQQ | PQQ | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | STY |
| 34 | 4MO-RET | 4MO | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 4MO | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | 4MO | 2GP-FLP | FLP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | 4MO | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | MTX | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | 4MO | 2GP-STY | 2GP |
| 41 | 4MO-NTM | NTM | MGN-BOX | MGN |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | MGN |
| 44 | LVS-PQQ | LVS | MGN-TDG | TDG |
| 45 | LVS-XLS | XLS | MGN-NTM | MGN |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | FLP |
| 48 | LVS-MGN | LVS | BOX-MTX | MTX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | FLP | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | BOX |
| 52 | LVS-TDG | TDG | FLP-MTX | FLP |
| 53 | LVS-NTM | NTM | FLP-TDG | FLP |
| 54 | LVS-STY | STY | FLP-NTM | NTM |

| | | | | |
|---|---|---|---|---|
| 55 | 9PP-PQQ | 9PP | FLP-STY | STY |
| 56 | 9PP-XLS | XLS | MTX-TDG | TDG |
| 57 | 9PP-RET | RET | MTX-NTM | NTM |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | NTM |
| 60 | NTM-STY | STY | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 7 | MGN | 8 |
| ORN | 6 | BOX | 5 |
| 4MO | 6 | FLP | 9 |
| 9PP | 6 | LVS | 8 |
| PQQ | 8 | MTX | 10 |
| XLS | 11 | TDG | 6 |
| RET | 5 | NTM | 9 |
| 2GP | 7 | STY | 9 |

***5.*** Classifying “1vlf” as 4MO binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | ORN | 9PP-BOX | 9PP |

| | | | | |
|---|---|---|---|---|
| 2 | BTN-4MO | 4MO | 9PP-FLP | FLP |
| 3 | BTN-LVS | BTN | 9PP-MTX | MTX |
| 4 | BTN-9PP | BTN | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | BTN | 9PP-NTM | NTM |
| 6 | BTN-XLS | XLS | 9PP-STY | STY |
| 7 | BTN-RET | RET | PQQ-XLS | PQQ |
| 8 | BTN-2GP | BTN | PQQ-RET | PQQ |
| 9 | BTN-MGN | BTN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BOX | PQQ-MGN | MGN |
| 11 | BTN-FLP | FLP | PQQ-BOX | BOX |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | TDG | PQQ-MTX | PQQ |
| 14 | BTN-NTM | BTN | PQQ-TDG | TDG |
| 15 | BTN-STY | BTN | PQQ-NTM | NTM |
| 16 | ORN-4MO | 4MO | PQQ-STY | PQQ |
| 17 | ORN-LVS | ORN | XLS-RET | RET |
| 18 | ORN-9PP | 9PP | XLS-2GP | 2GP |
| 19 | ORN-PQQ | ORN | XLS-MGN | XLS |
| 20 | ORN-XLS | ORN | XLS-BOX | XLS |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | 2GP | XLS-MTX | XLS |
| 23 | ORN-MGN | ORN | XLS-TDG | TDG |
| 24 | ORN-BOX | BOX | XLS-NTM | NTM |

| | | | | |
|---|---|---|---|---|
| 25 | ORN-FLP | ORN | XLS-STY | STY |
| 26 | ORN-MTX | MTX | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | MGN |
| 28 | ORN-NTM | NTM | RET-BOX | BOX |
| 29 | ORN-STY | STY | RET-FLP | FLP |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | RET |
| 32 | 4MO-PQQ | 4MO | RET-NTM | RET |
| 33 | 4MO-XLS | 4MO | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 4MO | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | MGN | 2GP-FLP | FLP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | MTX |
| 38 | 4MO-FLP | 4MO | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |
| 41 | 4MO-NTM | 4MO | MGN-BOX | MGN |
| 42 | 4MO-STY | 4MO | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | LVS | MGN-TDG | TDG |
| 45 | LVS-XLS | XLS | MGN-NTM | MGN |
| 46 | LVS-RET | LVS | MGN-STY | MGN |
| 47 | LVS-2GP | 2GP | BOX-FLP | BOX |

| | | | | |
|---|---|---|---|---|
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | FLP | BOX-NTM | BOX |
| 51 | LVS-MTX | LVS | BOX-STY | BOX |
| 52 | LVS-TDG | TDG | FLP-MTX | MTX |
| 53 | LVS-NTM | LVS | FLP-TDG | FLP |
| 54 | LVS-STY | LVS | FLP-NTM | NTM |
| 55 | 9PP-PQQ | 9PP | FLP-STY | STY |
| 56 | 9PP-XLS | XLS | MTX-TDG | TDG |
| 57 | 9PP-RET | 9PP | MTX-NTM | NTM |
| 58 | 9PP-2GP | 2GP | MTX-STY | MTX |
| 59 | 9PP-MGN | MGN | TDG-NTM | NTM |
| 60 | NTM-STY | STY | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 7 | MGN | 9 |
| ORN | 6 | BOX | 8 |
| 4MO | 11 | FLP | 9 |
| 9PP | 5 | LVS | 8 |
| PQQ | 4 | MTX | 10 |
| XLS | 6 | TDG | 7 |

| | | | |
|---|---|---|---|
| RET | 7 | NTM | 8 |
| 2GP | 7 | STY | 8 |

***6.*** Classifying “1kyi” as LVS binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | BOX |
| 2 | BTN-4MO | 4MO | 9PP-FLP | 9PP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | BTN | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | 9PP |
| 6 | BTN-XLS | XLS | 9PP-STY | STY |
| 7 | BTN-RET | RET | PQQ-XLS | XLS |
| 8 | BTN-2GP | BTN | PQQ-RET | RET |
| 9 | BTN-MGN | MGN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BOX | PQQ-MGN | PQQ |
| 11 | BTN-FLP | BTN | PQQ-BOX | BOX |
| 12 | BTN-MTX | BTN | PQQ-FLP | PQQ |
| 13 | BTN-TDG | TDG | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | TDG |
| 15 | BTN-STY | STY | PQQ-NTM | PQQ |

| | | | | |
|---|---|---|---|---|
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | XLS |
| 18 | ORN-9PP | ORN | XLS-2GP | 2GP |
| 19 | ORN-PQQ | PQQ | XLS-MGN | MGN |
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | ORN | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | XLS |
| 23 | ORN-MGN | ORN | XLS-TDG | XLS |
| 24 | ORN-BOX | BOX | XLS-NTM | XLS |
| 25 | ORN-FLP | FLP | XLS-STY | STY |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | ORN | RET-MGN | MGN |
| 28 | ORN-NTM | ORN | RET-BOX | RET |
| 29 | ORN-STY | ORN | RET-FLP | RET |
| 30 | 4MO-LVS | LVS | RET-MTX | RET |
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |
| 32 | 4MO-PQQ | PQQ | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | RET |
| 34 | 4MO-RET | 4MO | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | MGN | 2GP-FLP | FLP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | FLP | 2GP-TDG | 2GP |

| | | | | |
|---|---|---|---|---|
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |
| 41 | 4MO-NTM | NTM | MGN-BOX | MGN |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | PQQ | MGN-TDG | TDG |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | LVS | MGN-STY | MGN |
| 47 | LVS-2GP | LVS | BOX-FLP | FLP |
| 48 | LVS-MGN | MGN | BOX-MTX | BOX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | LVS | BOX-STY | STY |
| 52 | LVS-TDG | LVS | FLP-MTX | FLP |
| 53 | LVS-NTM | LVS | FLP-TDG | FLP |
| 54 | LVS-STY | STY | FLP-NTM | NTM |
| 55 | 9PP-PQQ | 9PP | FLP-STY | STY |
| 56 | 9PP-XLS | XLS | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | NTM |
| 60 | NTM-STY | NTM | TDG-STY | TDG |

The corresponding voting table for the classes is given as

| Class | Votes | Class | Votes |
|---|---|---|---|
| BTN | 5 | MGN | 8 |
| ORN | 9 | BOX | 9 |
| 4MO | 4 | FLP | 8 |
| 9PP | 4 | LVS | 12 |
| PQQ | 9 | MTX | 7 |
| XLS | 9 | TDG | 5 |
| RET | 8 | NTM | 8 |
| 2GP | 6 | STY | 9 |

***7.*** Classifying “1lvu” as 9PP binding protein

| | Type of Classifier | Class Selected | Type of Classifier | Class Selected |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | 9PP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | 9PP |
| 6 | BTN-XLS | XLS | 9PP-STY | 9PP |
| 7 | BTN-RET | RET | PQQ-XLS | XLS |

| | | | | |
|---|---|---|---|---|
| 8 | BTN-2GP | 2GP | PQQ-RET | RET |
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BOX | PQQ-MGN | MGN |
| 11 | BTN-FLP | BTN | PQQ-BOX | BOX |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | BTN | PQQ-MTX | MTX |
| 14 | BTN-NTM | BTN | PQQ-TDG | PQQ |
| 15 | BTN-STY | STY | PQQ-NTM | PQQ |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | XLS |
| 19 | ORN-PQQ | PQQ | XLS-MGN | MGN |
| 20 | ORN-XLS | ORN | XLS-BOX | BOX |
| 21 | ORN-RET | ORN | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | XLS |
| 23 | ORN-MGN | MGN | XLS-TDG | XLS |
| 24 | ORN-BOX | ORN | XLS-NTM | NTM |
| 25 | ORN-FLP | ORN | XLS-STY | STY |
| 26 | ORN-MTX | MTX | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | MGN |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | STY | RET-FLP | RET |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |

| | | | | |
|---|---|---|---|---|
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |
| 32 | 4MO-PQQ | 4MO | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | RET |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 4MO | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | MGN | 2GP-FLP | 2GP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | 2GP |
| 38 | 4MO-FLP | 4MO | 2GP-TDG | TDG |
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | TDG | 2GP-STY | 2GP |
| 41 | 4MO-NTM | 4MO | MGN-BOX | MGN |
| 42 | 4MO-STY | STY | MGN-MTX | MGN |
| 43 | LVS-9PP | 9PP | MGN-FLP | FLP |
| 44 | LVS-PQQ | PQQ | MGN-TDG | TDG |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | 2GP | BOX-FLP | FLP |
| 48 | LVS-MGN | MGN | BOX-MTX | BOX |
| 49 | LVS-BOX | BOX | BOX-TDG | BOX |
| 50 | LVS-FLP | FLP | BOX-NTM | NTM |
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | TDG | FLP-MTX | MTX |
| 53 | LVS-NTM | NTM | FLP-TDG | FLP |

| | | | | |
|---|---|---|---|---|
| 54 | LVS-STY | STY | FLP-NTM | FLP |
| 55 | 9PP-PQQ | 9PP | FLP-STY | STY |
| 56 | 9PP-XLS | 9PP | MTX-TDG | TDG |
| 57 | 9PP-RET | RET | MTX-NTM | NTM |
| 58 | 9PP-2GP | 9PP | MTX-STY | MTX |
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | NTM | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 5 | MGN | 10 |
| ORN | 6 | BOX | 6 |
| 4MO | 6 | FLP | 7 |
| 9PP | 11 | LVS | 4 |
| PQQ | 6 | MTX | 9 |
| XLS | 7 | TDG | 8 |
| RET | 9 | NTM | 9 |
| 2GP | 8 | STY | 9 |

***8.*** Classifying “1cs0” as ORN binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | ORN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | BTN | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | BTN | 9PP-STY | STY |
| 7 | BTN-RET | BTN | PQQ-XLS | XLS |
| 8 | BTN-2GP | 2GP | PQQ-RET | RET |
| 9 | BTN-MGN | MGN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BTN | PQQ-MGN | MGN |
| 11 | BTN-FLP | BTN | PQQ-BOX | PQQ |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | BTN | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | PQQ |
| 15 | BTN-STY | BTN | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | ORN | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | 2GP |
| 19 | ORN-PQQ | ORN | XLS-MGN | MGN |

| | | | | |
|---|---|---|---|---|
| 20 | ORN-XLS | ORN | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | XLS |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | TDG |
| 24 | ORN-BOX | ORN | XLS-NTM | XLS |
| 25 | ORN-FLP | ORN | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | MGN |
| 28 | ORN-NTM | ORN | RET-BOX | RET |
| 29 | ORN-STY | STY | RET-FLP | FLP |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | RET |
| 32 | 4MO-PQQ | PQQ | RET-NTM | NTM |
| 33 | 4MO-XLS | 4MO | RET-STY | RET |
| 34 | 4MO-RET | RET | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | 4MO | 2GP-FLP | FLP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | 2GP |
| 38 | 4MO-FLP | FLP | 2GP-TDG | TDG |
| 39 | 4MO-MTX | MTX | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | 4MO | 2GP-STY | 2GP |
| 41 | 4MO-NTM | 4MO | MGN-BOX | BOX |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |

| | | | | |
|---|---|---|---|---|
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | LVS | MGN-TDG | MGN |
| 45 | LVS-XLS | XLS | MGN-NTM | MGN |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | BOX | BOX-TDG | BOX |
| 50 | LVS-FLP | FLP | BOX-NTM | BOX |
| 51 | LVS-MTX | LVS | BOX-STY | STY |
| 52 | LVS-TDG | LVS | FLP-MTX | FLP |
| 53 | LVS-NTM | NTM | FLP-TDG | TDG |
| 54 | LVS-STY | LVS | FLP-NTM | FLP |
| 55 | 9PP-PQQ | PQQ | FLP-STY | FLP |
| 56 | 9PP-XLS | 9PP | MTX-TDG | MTX |
| 57 | 9PP-RET | 9PP | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | NTM | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 8 | MGN | 9 |
| ORN | 11 | BOX | 6 |

| | | | |
|---|---|---|---|
| 4MO | 6 | FLP | 10 |
| 9PP | 6 | LVS | 7 |
| PQQ | 7 | MTX | 10 |
| XLS | 6 | TDG | 5 |
| RET | 8 | NTM | 6 |
| 2GP | 8 | STY | 7 |

***9.*** Classifying “1bu4” as 2GP binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | 4MO | 9PP-FLP | FLP |
| 3 | BTN-LVS | BTN | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | BTN | 9PP-NTM | NTM |
| 6 | BTN-XLS | XLS | 9PP-STY | 9PP |
| 7 | BTN-RET | BTN | PQQ-XLS | XLS |
| 8 | BTN-2GP | 2GP | PQQ-RET | PQQ |
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BOX | PQQ-MGN | PQQ |
| 11 | BTN-FLP | BTN | PQQ-BOX | PQQ |
| 12 | BTN-MTX | BTN | PQQ-FLP | FLP |

| | | | | |
|---|---|---|---|---|
| 13 | BTN-TDG | TDG | PQQ-MTX | MTX |
| 14 | BTN-NTM | BTN | PQQ-TDG | PQQ |
| 15 | BTN-STY | STY | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | 2GP |
| 19 | ORN-PQQ | ORN | XLS-MGN | XLS |
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | ORN | XLS-FLP | FLP |
| 22 | ORN-2GP | 2GP | XLS-MTX | XLS |
| 23 | ORN-MGN | MGN | XLS-TDG | TDG |
| 24 | ORN-BOX | BOX | XLS-NTM | XLS |
| 25 | ORN-FLP | ORN | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | 2GP |
| 27 | ORN-TDG | ORN | RET-MGN | MGN |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | STY | RET-FLP | RET |
| 30 | 4MO-LVS | 4MO | RET-MTX | MTX |
| 31 | 4MO-9PP | 9PP | RET-TDG | RET |
| 32 | 4MO-PQQ | 4MO | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | RET |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | 2GP |

| | | | | |
|---|---|---|---|---|
| 36 | 4MO-MGN | 4MO | 2GP-FLP | FLP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | 2GP |
| 38 | 4MO-FLP | FLP | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | 4MO | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | 2GP |
| 41 | 4MO-NTM | 4MO | MGN-BOX | BOX |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | MGN |
| 44 | LVS-PQQ | PQQ | MGN-TDG | MGN |
| 45 | LVS-XLS | XLS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | MGN |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | BOX | BOX-TDG | BOX |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | LVS | FLP-MTX | FLP |
| 53 | LVS-NTM | LVS | FLP-TDG | TDG |
| 54 | LVS-STY | LVS | FLP-NTM | FLP |
| 55 | 9PP-PQQ | 9PP | FLP-STY | STY |
| 56 | 9PP-XLS | XLS | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |

| | | | | |
|---|---|---|---|---|
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | NTM | TDG-STY | TDG |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 7 | MGN | 8 |
| ORN | 6 | BOX | 8 |
| 4MO | 8 | FLP | 7 |
| 9PP | 7 | LVS | 7 |
| PQQ | 6 | MTX | 8 |
| XLS | 11 | TDG | 5 |
| RET | 7 | NTM | 7 |
| 2GP | 12 | STY | 6 |

***10.*** Classifying “1mro” as MGN binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | BOX |
| 2 | BTN-4MO | 4MO | 9PP-FLP | FLP |
| 3 | BTN-LVS | BTN | 9PP-MTX | MTX |
| 4 | BTN-9PP | BTN | 9PP-TDG | 9PP |

| | | | | |
|---|---|---|---|---|
| 5 | BTN-PQQ | PQQ | 9PP-NTM | 9PP |
| 6 | BTN-XLS | XLS | 9PP-STY | STY |
| 7 | BTN-RET | BTN | PQQ-XLS | XLS |
| 8 | BTN-2GP | BTN | PQQ-RET | PQQ |
| 9 | BTN-MGN | MGN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BTN | PQQ-MGN | MGN |
| 11 | BTN-FLP | FLP | PQQ-BOX | BOX |
| 12 | BTN-MTX | BTN | PQQ-FLP | FLP |
| 13 | BTN-TDG | BTN | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | PQQ |
| 15 | BTN-STY | STY | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | STY |
| 17 | ORN-LVS | LVS | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | 2GP |
| 19 | ORN-PQQ | PQQ | XLS-MGN | MGN |
| 20 | ORN-XLS | ORN | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | 2GP | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | XLS |
| 24 | ORN-BOX | BOX | XLS-NTM | NTM |
| 25 | ORN-FLP | ORN | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | ORN | RET-MGN | MGN |

| | | | | |
|---|---|---|---|---|
| 28 | ORN-NTM | ORN | RET-BOX | BOX |
| 29 | ORN-STY | STY | RET-FLP | RET |
| 30 | 4MO-LVS | 4MO | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |
| 32 | 4MO-PQQ | PQQ | RET-NTM | RET |
| 33 | 4MO-XLS | 4MO | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | BOX |
| 36 | 4MO-MGN | MGN | 2GP-FLP | 2GP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | 2GP |
| 38 | 4MO-FLP | FLP | 2GP-TDG | TDG |
| 39 | 4MO-MTX | MTX | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | TDG | 2GP-STY | 2GP |
| 41 | 4MO-NTM | 4MO | MGN-BOX | MGN |
| 42 | 4MO-STY | 4MO | MGN-MTX | MGN |
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | LVS | MGN-TDG | MGN |
| 45 | LVS-XLS | XLS | MGN-NTM | MGN |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | BOX | BOX-TDG | TDG |
| 50 | LVS-FLP | FLP | BOX-NTM | BOX |

| | | | | |
|---|---|---|---|---|
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | TDG | FLP-MTX | FLP |
| 53 | LVS-NTM | LVS | FLP-TDG | TDG |
| 54 | LVS-STY | LVS | FLP-NTM | FLP |
| 55 | 9PP-PQQ | 9PP | FLP-STY | FLP |
| 56 | 9PP-XLS | XLS | MTX-TDG | TDG |
| 57 | 9PP-RET | 9PP | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | MTX |
| 59 | 9PP-MGN | MGN | TDG-NTM | NTM |
| 60 | NTM-STY | NTM | TDG-STY | TDG |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 8 | MGN | 12 |
| ORN | 7 | BOX | 9 |
| 4MO | 7 | FLP | 10 |
| 9PP | 5 | LVS | 6 |
| PQQ | 6 | MTX | 8 |
| XLS | 7 | TDG | 8 |
| RET | 6 | NTM | 5 |
| 2GP | 9 | STY | 7 |

***11.*** Classifying “1ais” as BOX binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | BOX |
| 2 | BTN-4MO | BTN | 9PP-FLP | 9PP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | 9PP |
| 6 | BTN-XLS | XLS | 9PP-STY | 9PP |
| 7 | BTN-RET | BTN | PQQ-XLS | PQQ |
| 8 | BTN-2GP | BTN | PQQ-RET | RET |
| 9 | BTN-MGN | BTN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BOX | PQQ-MGN | PQQ |
| 11 | BTN-FLP | FLP | PQQ-BOX | BOX |
| 12 | BTN-MTX | BTN | PQQ-FLP | FLP |
| 13 | BTN-TDG | TDG | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | PQQ |
| 15 | BTN-STY | BTN | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | ORN | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | XLS |
| 19 | ORN-PQQ | ORN | XLS-MGN | MGN |

| | | | | |
|---|---|---|---|---|
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | ORN | XLS-FLP | FLP |
| 22 | ORN-2GP | 2GP | XLS-MTX | XLS |
| 23 | ORN-MGN | ORN | XLS-TDG | TDG |
| 24 | ORN-BOX | BOX | XLS-NTM | NTM |
| 25 | ORN-FLP | FLP | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | MGN |
| 28 | ORN-NTM | NTM | RET-BOX | BOX |
| 29 | ORN-STY | ORN | RET-FLP | RET |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | RET |
| 32 | 4MO-PQQ | 4MO | RET-NTM | RET |
| 33 | 4MO-XLS | XLS | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | BOX |
| 36 | 4MO-MGN | MGN | 2GP-FLP | FLP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | 4MO | 2GP-TDG | TDG |
| 39 | 4MO-MTX | 4MO | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | 4MO | 2GP-STY | 2GP |
| 41 | 4MO-NTM | NTM | MGN-BOX | BOX |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |

| | | | | |
|---|---|---|---|---|
| 43 | LVS-9PP | 9PP | MGN-FLP | MGN |
| 44 | LVS-PQQ | LVS | MGN-TDG | TDG |
| 45 | LVS-XLS | XLS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | MGN |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | LVS | BOX-MTX | MTX |
| 49 | LVS-BOX | BOX | BOX-TDG | TDG |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | BOX |
| 52 | LVS-TDG | LVS | FLP-MTX | FLP |
| 53 | LVS-NTM | LVS | FLP-TDG | TDG |
| 54 | LVS-STY | STY | FLP-NTM | FLP |
| 55 | 9PP-PQQ | 9PP | FLP-STY | STY |
| 56 | 9PP-XLS | XLS | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | NTM |
| 58 | 9PP-2GP | 9PP | MTX-STY | STY |
| 59 | 9PP-MGN | 9PP | TDG-NTM | TDG |
| 60 | NTM-STY | NTM | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 7 | MGN | 5 |
| ORN | 7 | BOX | 13 |

| | | | |
|---|---|---|---|
| 4MO | 5 | FLP | 7 |
| 9PP | 9 | LVS | 9 |
| PQQ | 6 | MTX | 8 |
| XLS | 9 | TDG | 8 |
| RET | 8 | NTM | 8 |
| 2GP | 5 | STY | 6 |

***12.*** Classifying “1dvt” as FLP binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | 9PP | 9PP-TDG | TDG |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | 9PP |
| 6 | BTN-XLS | BTN | 9PP-STY | 9PP |
| 7 | BTN-RET | BTN | PQQ-XLS | XLS |
| 8 | BTN-2GP | 2GP | PQQ-RET | PQQ |
| 9 | BTN-MGN | MGN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BTN | PQQ-MGN | MGN |
| 11 | BTN-FLP | FLP | PQQ-BOX | PQQ |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |

| | | | | |
|---|---|---|---|---|
| 13 | BTN-TDG | BTN | PQQ-MTX | MTX |
| 14 | BTN-NTM | BTN | PQQ-TDG | TDG |
| 15 | BTN-STY | STY | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | RET |
| 18 | ORN-9PP | ORN | XLS-2GP | XLS |
| 19 | ORN-PQQ | PQQ | XLS-MGN | XLS |
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | ORN | XLS-FLP | FLP |
| 22 | ORN-2GP | 2GP | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | XLS |
| 24 | ORN-BOX | ORN | XLS-NTM | XLS |
| 25 | ORN-FLP | FLP | XLS-STY | XLS |
| 26 | ORN-MTX | MTX | RET-2GP | 2GP |
| 27 | ORN-TDG | ORN | RET-MGN | MGN |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | ORN | RET-FLP | FLP |
| 30 | 4MO-LVS | LVS | RET-MTX | RET |
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |
| 32 | 4MO-PQQ | 4MO | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | RET |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 4MO | 2GP-BOX | 2GP |

| | | | | |
|---|---|---|---|---|
| 36 | 4MO-MGN | 4MO | 2GP-FLP | FLP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | FLP | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | 4MO | 2GP-NTM | 2GP |
| 40 | 4MO-TDG | TDG | 2GP-STY | STY |
| 41 | 4MO-NTM | 4MO | MGN-BOX | BOX |
| 42 | 4MO-STY | STY | MGN-MTX | MTX |
| 43 | LVS-9PP | 9PP | MGN-FLP | MGN |
| 44 | LVS-PQQ | PQQ | MGN-TDG | TDG |
| 45 | LVS-XLS | XLS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | FLP |
| 48 | LVS-MGN | LVS | BOX-MTX | BOX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | BOX |
| 52 | LVS-TDG | TDG | FLP-MTX | FLP |
| 53 | LVS-NTM | NTM | FLP-TDG | FLP |
| 54 | LVS-STY | STY | FLP-NTM | FLP |
| 55 | 9PP-PQQ | PQQ | FLP-STY | FLP |
| 56 | 9PP-XLS | 9PP | MTX-TDG | TDG |
| 57 | 9PP-RET | RET | MTX-NTM | NTM |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |

| | | | | |
|---|---|---|---|---|
| 59 | 9PP-MGN | 9PP | TDG-NTM | TDG |
| 60 | NTM-STY | NTM | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 7 | MGN | 4 |
| ORN | 7 | BOX | 7 |
| 4MO | 6 | FLP | 13 |
| 9PP | 7 | LVS | 7 |
| PQQ | 8 | MTX | 8 |
| XLS | 9 | TDG | 8 |
| RET | 7 | NTM | 7 |
| 2GP | 8 | STY | 7 |

***13.*** Classifying “1ddr” as MTX binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | 9PP |

| | | | | |
|---|---|---|---|---|
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | BTN | 9PP-TDG | TDG |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | BTN | 9PP-STY | STY |
| 7 | BTN-RET | RET | PQQ-XLS | PQQ |
| 8 | BTN-2GP | BTN | PQQ-RET | RET |
| 9 | BTN-MGN | BTN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BOX | PQQ-MGN | MGN |
| 11 | BTN-FLP | BTN | PQQ-BOX | BOX |
| 12 | BTN-MTX | MTX | PQQ-FLP | PQQ |
| 13 | BTN-TDG | BTN | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | PQQ |
| 15 | BTN-STY | STY | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | STY |
| 17 | ORN-LVS | LVS | XLS-RET | RET |
| 18 | ORN-9PP | 9PP | XLS-2GP | 2GP |
| 19 | ORN-PQQ | ORN | XLS-MGN | MGN |
| 20 | ORN-XLS | ORN | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | XLS |
| 24 | ORN-BOX | ORN | XLS-NTM | XLS |
| 25 | ORN-FLP | FLP | XLS-STY | XLS |

| | | | | |
|---|---|---|---|---|
| 26 | ORN-MTX | MTX | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | RET |
| 28 | ORN-NTM | ORN | RET-BOX | BOX |
| 29 | ORN-STY | ORN | RET-FLP | RET |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |
| 32 | 4MO-PQQ | 4MO | RET-NTM | NTM |
| 33 | 4MO-XLS | XLS | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | MGN | 2GP-FLP | 2GP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | FLP | 2GP-TDG | TDG |
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |
| 41 | 4MO-NTM | NTM | MGN-BOX | MGN |
| 42 | 4MO-STY | 4MO | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | PQQ | MGN-TDG | TDG |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | LVS | BOX-MTX | MTX |

| | | | | |
|---|---|---|---|---|
| 49 | LVS-BOX | BOX | BOX-TDG | BOX |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | STY |
| 52 | LVS-TDG | LVS | FLP-MTX | MTX |
| 53 | LVS-NTM | LVS | FLP-TDG | FLP |
| 54 | LVS-STY | STY | FLP-NTM | FLP |
| 55 | 9PP-PQQ | PQQ | FLP-STY | FLP |
| 56 | 9PP-XLS | 9PP | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | NTM |
| 60 | NTM-STY | STY | TDG-STY | TDG |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 8 | MGN | 5 |
| ORN | 8 | BOX | 9 |
| 4MO | 4 | FLP | 7 |
| 9PP | 4 | LVS | 10 |
| PQQ | 7 | MTX | 14 |
| XLS | 4 | TDG | 6 |
| RET | 10 | NTM | 8 |

| | | | |
|---|---|---|---|
| 2GP | 6 | STY | 10 |

***14.*** Classifying “1qap” as NTM binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | LVS | 9PP-MTX | MTX |
| 4 | BTN-9PP | BTN | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | XLS | 9PP-STY | 9PP |
| 7 | BTN-RET | BTN | PQQ-XLS | XLS |
| 8 | BTN-2GP | BTN | PQQ-RET | PQQ |
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BOX | PQQ-MGN | MGN |
| 11 | BTN-FLP | BTN | PQQ-BOX | PQQ |
| 12 | BTN-MTX | BTN | PQQ-FLP | PQQ |
| 13 | BTN-TDG | TDG | PQQ-MTX | PQQ |
| 14 | BTN-NTM | NTM | PQQ-TDG | TDG |
| 15 | BTN-STY | BTN | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | ORN | XLS-RET | RET |

| | | | | |
|---|---|---|---|---|
| 18 | ORN-9PP | 9PP | XLS-2GP | XLS |
| 19 | ORN-PQQ | PQQ | XLS-MGN | XLS |
| 20 | ORN-XLS | ORN | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | XLS |
| 22 | ORN-2GP | 2GP | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | XLS |
| 24 | ORN-BOX | ORN | XLS-NTM | NTM |
| 25 | ORN-FLP | FLP | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | ORN | RET-MGN | RET |
| 28 | ORN-NTM | NTM | RET-BOX | BOX |
| 29 | ORN-STY | STY | RET-FLP | FLP |
| 30 | 4MO-LVS | 4MO | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | RET |
| 32 | 4MO-PQQ | 4MO | RET-NTM | RET |
| 33 | 4MO-XLS | XLS | RET-STY | STY |
| 34 | 4MO-RET | RET | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 2GP | 2GP-BOX | BOX |
| 36 | 4MO-MGN | MGN | 2GP-FLP | FLP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | 2GP |
| 38 | 4MO-FLP | FLP | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | 4MO | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |

| | | | | |
|---|---|---|---|---|
| 41 | 4MO-NTM | NTM | MGN-BOX | BOX |
| 42 | 4MO-STY | 4MO | MGN-MTX | MTX |
| 43 | LVS-9PP | LVS | MGN-FLP | MGN |
| 44 | LVS-PQQ | PQQ | MGN-TDG | MGN |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | LVS | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | BOX |
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | LVS | BOX-TDG | BOX |
| 50 | LVS-FLP | FLP | BOX-NTM | NTM |
| 51 | LVS-MTX | MTX | BOX-STY | BOX |
| 52 | LVS-TDG | TDG | FLP-MTX | FLP |
| 53 | LVS-NTM | NTM | FLP-TDG | FLP |
| 54 | LVS-STY | STY | FLP-NTM | NTM |
| 55 | 9PP-PQQ | 9PP | FLP-STY | FLP |
| 56 | 9PP-XLS | XLS | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | NTM |
| 58 | 9PP-2GP | 9PP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | NTM | TDG-STY | STY |

The corresponding voting table for the classes is given as

| Class | Votes | Class | Votes |
|---|---|---|---|
| BTN | 8 | MGN | 8 |
| ORN | 7 | BOX | 9 |
| 4MO | 6 | FLP | 9 |
| 9PP | 6 | LVS | 6 |
| PQQ | 8 | MTX | 7 |
| XLS | 9 | TDG | 4 |
| RET | 8 | NTM | 13 |
| 2GP | 5 | STY | 7 |

***15.*** Classifying “1ghx” as STY binding protein

| | Type of Classifier | Class Selected | Type of Classifier | Class Selected |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | BOX |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | LVS | 9PP-MTX | 9PP |
| 4 | BTN-9PP | 9PP | 9PP-TDG | 9PP |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | BTN | 9PP-STY | 9PP |
| 7 | BTN-RET | RET | PQQ-XLS | PQQ |
| 8 | BTN-2GP | BTN | PQQ-RET | RET |

| | | | | |
|---|---|---|---|---|
| 9 | BTN-MGN | MGN | PQQ-2GP | PQQ |
| 10 | BTN-BOX | BTN | PQQ-MGN | MGN |
| 11 | BTN-FLP | BTN | PQQ-BOX | PQQ |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | TDG | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | PQQ |
| 15 | BTN-STY | STY | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | STY |
| 17 | ORN-LVS | LVS | XLS-RET | RET |
| 18 | ORN-9PP | ORN | XLS-2GP | XLS |
| 19 | ORN-PQQ | ORN | XLS-MGN | MGN |
| 20 | ORN-XLS | XLS | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | ORN | XLS-TDG | XLS |
| 24 | ORN-BOX | BOX | XLS-NTM | NTM |
| 25 | ORN-FLP | ORN | XLS-STY | XLS |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | TDG | RET-MGN | RET |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | ORN | RET-FLP | FLP |
| 30 | 4MO-LVS | LVS | RET-MTX | MTX |
| 31 | 4MO-9PP | 4MO | RET-TDG | TDG |

| | | | | |
|---|---|---|---|---|
| 32 | 4MO-PQQ | PQQ | RET-NTM | NTM |
| 33 | 4MO-XLS | 4MO | RET-STY | RET |
| 34 | 4MO-RET | RET | 2GP-MGN | MGN |
| 35 | 4MO-2GP | 4MO | 2GP-BOX | BOX |
| 36 | 4MO-MGN | 4MO | 2GP-FLP | 2GP |
| 37 | 4MO-BOX | BOX | 2GP-MTX | MTX |
| 38 | 4MO-FLP | 4MO | 2GP-TDG | TDG |
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |
| 41 | 4MO-NTM | NTM | MGN-BOX | BOX |
| 42 | 4MO-STY | 4MO | MGN-MTX | MGN |
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | PQQ | MGN-TDG | TDG |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | LVS | MGN-STY | STY |
| 47 | LVS-2GP | 2GP | BOX-FLP | BOX |
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | LVS | BOX-TDG | TDG |
| 50 | LVS-FLP | FLP | BOX-NTM | NTM |
| 51 | LVS-MTX | LVS | BOX-STY | BOX |
| 52 | LVS-TDG | TDG | FLP-MTX | MTX |
| 53 | LVS-NTM | NTM | FLP-TDG | FLP |
| 54 | LVS-STY | LVS | FLP-NTM | NTM |

| | | | | |
|---|---|---|---|---|
| 55 | 9PP-PQQ | PQQ | FLP-STY | FLP |
| 56 | 9PP-XLS | 9PP | MTX-TDG | TDG |
| 57 | 9PP-RET | RET | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | MTX |
| 59 | 9PP-MGN | MGN | TDG-NTM | NTM |
| 60 | NTM-STY | NTM | TDG-STY | STY |

The corresponding voting table for the classes is given as

| **Class** | **Votes** | **Class** | **Votes** |
|---|---|---|---|
| BTN | 6 | MGN | 7 |
| ORN | 8 | BOX | 8 |
| 4MO | 7 | FLP | 8 |
| 9PP | 5 | LVS | 9 |
| PQQ | 8 | MTX | 10 |
| XLS | 4 | TDG | 8 |
| RET | 10 | NTM | 14 |
| 2GP | 3 | STY | 5 |

***16.*** Classifying “1h5t” as TDG binding protein

| | **Type of Classifier** | **Class Selected** | **Type of Classifier** | **Class Selected** |
|---|---|---|---|---|
| 1 | BTN-ORN | BTN | 9PP-BOX | 9PP |
| 2 | BTN-4MO | BTN | 9PP-FLP | FLP |
| 3 | BTN-LVS | LVS | 9PP-MTX | 9PP |
| 4 | BTN-9PP | BTN | 9PP-TDG | TDG |
| 5 | BTN-PQQ | PQQ | 9PP-NTM | NTM |
| 6 | BTN-XLS | BTN | 9PP-STY | STY |
| 7 | BTN-RET | RET | PQQ-XLS | PQQ |
| 8 | BTN-2GP | BTN | PQQ-RET | RET |
| 9 | BTN-MGN | MGN | PQQ-2GP | 2GP |
| 10 | BTN-BOX | BTN | PQQ-MGN | PQQ |
| 11 | BTN-FLP | BTN | PQQ-BOX | PQQ |
| 12 | BTN-MTX | MTX | PQQ-FLP | FLP |
| 13 | BTN-TDG | TDG | PQQ-MTX | MTX |
| 14 | BTN-NTM | NTM | PQQ-TDG | TDG |
| 15 | BTN-STY | STY | PQQ-NTM | NTM |
| 16 | ORN-4MO | ORN | PQQ-STY | PQQ |
| 17 | ORN-LVS | LVS | XLS-RET | XLS |
| 18 | ORN-9PP | 9PP | XLS-2GP | XLS |

| | | | | |
|---|---|---|---|---|
| 19 | ORN-PQQ | ORN | XLS-MGN | XLS |
| 20 | ORN-XLS | ORN | XLS-BOX | BOX |
| 21 | ORN-RET | RET | XLS-FLP | FLP |
| 22 | ORN-2GP | ORN | XLS-MTX | MTX |
| 23 | ORN-MGN | MGN | XLS-TDG | TDG |
| 24 | ORN-BOX | ORN | XLS-NTM | NTM |
| 25 | ORN-FLP | FLP | XLS-STY | STY |
| 26 | ORN-MTX | ORN | RET-2GP | RET |
| 27 | ORN-TDG | ORN | RET-MGN | RET |
| 28 | ORN-NTM | NTM | RET-BOX | RET |
| 29 | ORN-STY | ORN | RET-FLP | FLP |
| 30 | 4MO-LVS | 4MO | RET-MTX | MTX |
| 31 | 4MO-9PP | 9PP | RET-TDG | TDG |
| 32 | 4MO-PQQ | 4MO | RET-NTM | NTM |
| 33 | 4MO-XLS | 4MO | RET-STY | RET |
| 34 | 4MO-RET | RET | 2GP-MGN | 2GP |
| 35 | 4MO-2GP | 4MO | 2GP-BOX | 2GP |
| 36 | 4MO-MGN | MGN | 2GP-FLP | FLP |
| 37 | 4MO-BOX | 4MO | 2GP-MTX | MTX |
| 38 | 4MO-FLP | FLP | 2GP-TDG | 2GP |
| 39 | 4MO-MTX | MTX | 2GP-NTM | NTM |
| 40 | 4MO-TDG | 4MO | 2GP-STY | STY |
| 41 | 4MO-NTM | NTM | MGN-BOX | BOX |

| | | | | |
|---|---|---|---|---|
| 42 | 4MO-STY | STY | MGN-MTX | MGN |
| 43 | LVS-9PP | LVS | MGN-FLP | FLP |
| 44 | LVS-PQQ | PQQ | MGN-TDG | TDG |
| 45 | LVS-XLS | LVS | MGN-NTM | NTM |
| 46 | LVS-RET | RET | MGN-STY | STY |
| 47 | LVS-2GP | LVS | BOX-FLP | FLP |
| 48 | LVS-MGN | MGN | BOX-MTX | MTX |
| 49 | LVS-BOX | LVS | BOX-TDG | TDG |
| 50 | LVS-FLP | LVS | BOX-NTM | BOX |
| 51 | LVS-MTX | MTX | BOX-STY | BOX |
| 52 | LVS-TDG | LVS | FLP-MTX | MTX |
| 53 | LVS-NTM | NTM | FLP-TDG | TDG |
| 54 | LVS-STY | LVS | FLP-NTM | NTM |
| 55 | 9PP-PQQ | PQQ | FLP-STY | FLP |
| 56 | 9PP-XLS | 9PP | MTX-TDG | MTX |
| 57 | 9PP-RET | RET | MTX-NTM | MTX |
| 58 | 9PP-2GP | 2GP | MTX-STY | STY |
| 59 | 9PP-MGN | MGN | TDG-NTM | TDG |
| 60 | NTM-STY | STY | TDG-STY | TDG |

The corresponding voting table for the classes is given as

| Class | Votes | Class | Votes |
|---|---|---|---|
| BTN | 7 | MGN | 6 |
| ORN | 8 | BOX | 4 |
| 4MO | 6 | FLP | 10 |
| 9PP | 5 | LVS | 9 |
| PQQ | 7 | MTX | 11 |
| XLS | 3 | TDG | 10 |
| RET | 10 | NTM | 11 |
| 2GP | 5 | STY | 8 |