Evolutionary based Classification of Fungal Lipases as a Framework for Structure and Function Prediction of Putative Lipases.

Bahareh Behdad

A Thesis

in

The Department

of

Biology

Presented in Partial Fulfillment of the Requirements

For the Degree of Masters of Science (Biology) at

Concordia University

Montreal, Quebec, Canada

March 2006

Your file Votre référence
ISBN: 978-0-494-30076-3
Our file Notre référence
ISBN: 978-0-494-30076-3

ABSTRACT

Evolutionary based Classification of Fungal Lipases as a Framework for Structure and Function Prediction of Putative Lipases.

Bahareh Behdad

Lipases are a family of enzymes which catalyze the hydrolysis of lipids and they exhibit esterase type activities such as phospholipase, lysophospholipase, cutinase and amidase. These enzymes have applications in detergent, the production of oil and fat, baked products, organic synthesis, hard surface cleaning, leather, paper, and cocoa butter. Lipases are versatile and have become a major source of industrial lipolytic enzymes. I performed a comparative analysis of protein sequences of well characterized fungal lipolytic enzymes from SWISS-PROT and developed a classification system to assist structural and functional characterization of newly identified putative lipase gene sequences. The distance-based UPGMA method was used for constructing the phylogenetic tree. Eight sequences of known secondary and tertiary structure were used to predict the secondary structure of similar sequences with unknown structure. Using sixteen different tools from SWISS-PROT and NPS@, the secondary structures of sequences with unknown 3-dimensional structures were predicted. The evolutionary based clustering of lipase protein sequences resulted in seven major families with the largest family being divided into five subfamilies and two single member branches. This analysis allowed us to: (i) perform a comparative study of well-characterized fungal lipases to develop a comprehensive classification system for fungal lipases, (ii) assess

various protein secondary structure prediction tools to select suitable tool(s) for predicting secondary structures of lipases, (iii) predict structural features important for specific function of lipases such as residues forming the catalytic sites, disulfide bonds and salt bridges, and (iv) develop a framework to predict putative function and reaction conditions of newly identified lipase gene sequences.

ACKNOWLEDGEMENT

I would like to gratefully acknowledge the enthusiastic supervision of Dr. Adrian Tsang and Dr. Selvadurai Dayanandan, for giving me the opportunity to work with them. Their guidance and support have been invaluable to me throughout my graduate studies and have helped me to develop skills necessary for a young researcher.

I would also like to thank Dr. Paul Joyce, my committee member who advised me and helped me in various aspects of my research. Professor Reginald K. Storms provided valued criticism of the theory section.

I am grateful to all my fellow graduate students and colleagues, in the Biology departments at Concordia University and Fungal Genomic Center, who have offered me their assistance and for their continued moral support. I would particularly like to thank Mona Hamzeh, Nathalie Broduer, Margo Marszalek, Sandra Chiovitti, Amir R. Kermany for all their friendship and patience.

I would like to thank Concordia University, Genome Canada, Genome Québec, Natural Sciences and Engineering Research Council of Canada (NSERC) for their generous financial support that made this study possible.

A very special thanks goes out to my friend, Farhad Massoudi, without his encouragement and editing assistance, I would not have finished this thesis.

Finally, I am forever indebted to my parents for their understanding, endless patience and encouragement and love throughout my studies. My special thanks to my sisters, Ghazal and Shirin, who have given me enormous support.

DEDICATION

I would like to dedicate this thesis to my loving parents...

TABLE OF CONTENTS

LIST of FIGURES

LIST of TABLES

A. INTRODUCTION

Lipases, a group of enzymes which catalyze hydrolysis of insoluble oil droplets into soluble products, are ubiquitous in nature. They are found in organisms ranging from bacteria and fungi to plants and animals. These enzymes aid in processes such as digestion, membrane phospholipid metabolism, and inflammatory reactions (Tojo *et al.*, 1997). Moreover, these enzymes are commercially important in industries such as detergent, oil and fat, baking, organic synthesis, paper and cocoa butter (Schmid and Verger, 1998; Bornscheuer and Kazlauskas, 1999; Anderson et al., 1998; Clausen *et al.*, 2000; Rubingh, 1998; Jaeger, 1998). Fungal lipases have become a major source of industrial lipolytic enzymes because of their versatility and ease of production. Such new found uses for fungal lipases have spurred growth in patent filings, driving the need for an improved understanding of the structure-function relationships of fungal lipases.

Fungi produce various families of lipolytic enzymes including true lipases (EC 3.1.1.3), carboxylesterases (EC 3.1.1.3), secretory lipases (EC 3.1.1.3), and a variety of phospholipases, namely, phospholipase A1 (EC 3.1.1.32), phospholipase A2 (EC 3.1.1.4), lysophospholipase (EC 3.1.1.5), phospholipase C or 1-phosphatidylinositol-4, 5-bisphosphate phosphodiesterase 1 (EC 3.1.4.11), and phospholipase D (EC 3.1.4.4). The first three classes of enzymes belong to the structure-based super family of α/β-hydrolase's (Ollis et al., 1992), a variety of enzymes whose activities rely mainly on a catalytic triad usually formed by serine, histidine and aspartic acid residues.

Three developments have advanced our knowledge of the structure and function of lipases and esterases: increased availability of gene sequences, biochemical

characterization of lipases, and resolution of numerous crystal structures (Grochulski *et al.,* 1993; Schrag and Cygler, 1993; Cygler *et al.,* 1993).

These classifications are based upon comparative studies of the structure and function of well-characterized enzyme groups. They aid in the identification of unknown structural motifs and also help predict the functions of newly-identified genes. Although a structure and function based classification of bacterial lipases already exists (Arpigny and Jaeger, 1999), no such study is yet available for fungal lipases.

Our objectives are to:

(1) Perform a comparative study of well-characterized fungal lipases to develop a classification system for fungal lipases.

(2) Select the best tools for predicting secondary structure of lipases.

(3) Predict structural features important for function of lipases such as residues forming the catalytic sites, disulfide bonds and salt bridges.

(4) Develop a framework to predict function and reaction conditions of newly-identified lipase gene sequences.

B. LITERATURE REVIEW

B.1 Background Information and Definition of lipases

Lipases were first identified in 1856 by Claude Bernard. They hydrolyse insoluble oil droplets into soluble products. The presence of lipolytic enzymes in blood plasma was first noted by Hahn in 1943. Initially, the enzyme was described as a "clearing factor". Nine years later, Anfinsen and co-workers associated the clearing factor with the presence of lipolytic enzymes (Verger *et al.*, 1984). Since then, there has been a growing interest in both the lipolytic enzymes and their substrates. The lipase enzymes consist of different families showing the same overall structural folding (Ollis *et al.*, 1992; Derewenda *et al.*, 1994a). However, they differ in the versatility of their loop structures that contact the substrate. They also show variable substrate specificities. Lipases, also known as lipolytic enzymes, are capable of hydrolyzing lipid substrates: cutinases, lysophospholipases and enzymes hydrolyzing ester substrates of lipid nature. Lipases, in contrast to esterases, become activated when absorbed in a water/lipid interface. They display low activity with their substrates in a monomeric state (Verger and de Hass, 1976). A "true" lipolytic enzyme exhibits two characteristics (EC 3.1.1.3):

- It should be activated by the presence of an interface. That is, its activity should sharply increase as soon as the triglyceride substrate forms an emulsion (interfacial activation) (Sarda *et al.*, 1958).

- It should contain a "lid" (see below), which consists of hydrophobic residues covering the active site. By movement of the lid, the substrate can enter the cavity for enzyme reaction.

What make lipases so attractive? First, they usually display exquisite chemoselectivity, regioselectivity and stereoselectivity (Rogalska *et al.,* 1993). Secondly, they can readily be produced in high yields from microbial organisms-- fungi and bacteria. Thirdly, the crystal structures of many lipases have been solved, facilitating the design of rational engineering strategies. Finally, they usually do not require cofactors, and do not catalyze side reactions. These properties make lipases the most widely used group of biocatalysts in organic chemistry.

B. 2 Structural studies on lipases

To determine the structure of a protein it is necessary to understand its function and role. The mechanism of catalysis and the molecular nature of interfacial activation was not understood until the 3-D structures of lipases were determined. The lipases from two fungi-- *Aspergillus niger* (Fukumoto *et al.,* 1963) and *Geotrichum candidum* -- (Tsujisaka *et al.,* 1973) were first to be crystallized. These crystals were unstable and of poor quality, likely a result of heterogeneity in the enzyme preparations. Since then, many lipases have been crystallized in a form suitable for high resolution X-ray diffraction studies.

Triacylglycerol lipases are α/β proteins, with a central ß-sheet with the active site serine placed in a loop, in which the nucleophilic residue is essential for catalysis (Cygler *et al.*, 1992). This putative hydrolytic site is covered by a surface loop and is therefore, inaccessible to solvent. Interfacial activation, a property of lipolytic enzymes acting on water-insoluble substrates at water-lipid interfaces, probably involves the movement of this flap region in lipoprotein lipases. This movement changes the surface at the entrance

of the active site, making it more hydrophobic and changing the lipid-binding properties.

The activities of these enzymes rely on a catalytic triad usually formed by three residues following the order Ser-Asp/Glu-His. The active site serine is found in the consensus pentapeptide Gly-X-Ser-X-Gly (X being any amino acid). This sole presence of this motif has identified many serine hydrolases. The active site serine is embedded in a secondary structure element: β-strand-turn-α-helix. This general element was recognized in previous studies (Derewenda and Derewenda, 1991; Schrag *et al.*, 1991). It was found not only in lipases, but also in other hydrolytic enzymes (Schrag *et al.*, 1991; Ollis *et al.*, 1992).

The two glycine residues of the Gly-X-Ser-X-Gly consensus sequence are critical in maintaining the tight bend between the β-strand and the α-helix. These two residues face each other (Figure 1); the distance between their C_α atoms is very short (~ 4.5 Å). Figure 1 shows the comparison of *Rhizomucor miehei* lipase (RML) and *Geotrichum candidum* lipase (GCL), showing the conserved Gly-X-Ser-X-Gly and the short distance between the two Gly residues. Not only should the Gly residue be conserved, but also right after this Gly, there is always a small amino acid such as alanine or another glycine. This keeps the outlook or nature of the β-strand and α-helix unchanged (Cygler *et al.*, 1992). This conserved motif is seen in all known esterase structures: acetyl cholinesterase (Sussman *et al.*, 1991), cutinase (Nicolas *et al.*, 1996) *strepotomyces scabies* esterase (Wei *et al.*, 1995).

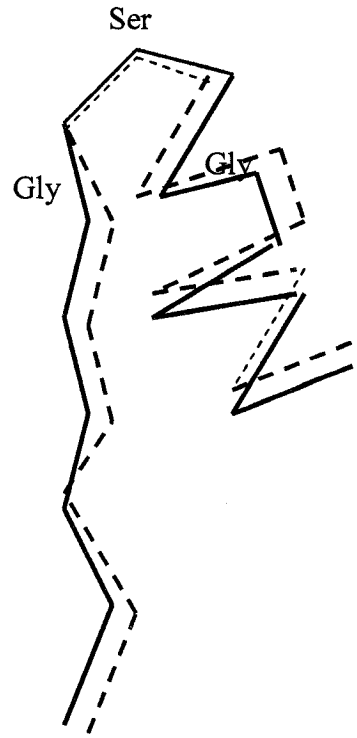


Figure 1: Superposition of strand-serine-helix super secondary motif. GCL (thick line) and RML are shown in dashed line. Red indicates the α-helix, and green indicates the β-strand.

Although the scaffolds and the active sites of lipases and cutinases share common features, accessibility to the active sites in each lipase differs. For example, in RML (*Rhizomucor miehei* lipase) the active site is covered by a short region (flap) and is about 10 Å from the protein surface. A similar structure was seen for human pancreatic lipase (HPL) (Terzyan *et al.*, 2000) with a longer loop and greater distance to the surface (about 14 Å). In fact, a different structure is seen for GCL, where two flap regions from different surfaces cover the active site. Yet no flap region is found in cutinase, and further, the active site is on the surface of the molecule (Martinez *et al.*, 1992). The histidine residue is often part of a special sequence pattern (Svendsen *et al.*, 1995) and mostly found in the C-terminal end of lipases. However, this residue in *Geotrichum candidum* lipase (GCL) comes from the N-terminal part of the sequence. In cutinases, the histidine is located in the same loop as its triad partner. Cutinases have a shorter C-terminus than the above

mentioned lipases (Martinez *et al.*, 1992). Moreover, the Asp/Glu residue would be found in the triacylglycerol lipases between the serine and the histidine residue.

In some lipase-related enzymes, the active site residues are arranged differently. In secretory lipases (discussed later in this study), there are only two catalytic residues, instead of three. Also, in the *Streptomyces scapies* bacteria esterase, the Asp residue is 'replaced' by carbonyl oxygen. Enzymes including lipases, proteases, esterases, peroxidases and lyses are members of the "α/β hydrolase fold" clan (Ollis *et al.*, 1992; Schrag *et al.*, 1997), although not all lipases belong to the same structural family. As an example, the structures of PLA_2, PLA-D, lysophospholipases and phospholipases C differ from triacylglycerol lipases.

B. 3 Function

Because lipases are water-soluble enzymes and function at interfaces, evaluation of substrate specificity is governed by several other factors (Brady *et al.,* 1990). Few lipases show strict substrate specificity. The mode of presentation of the substrate is also important (Brockman, 1984). The same enzyme may show different selectivities when the substrate is added in variable forms, like micelles, monolayer films or dissolved in organic solvents. Classically, types for specificities of lipases are classified by factors such as: a) substrates, i.e., preference for long or short fatty acids; b) positional: including sn1, sn2 and sn3 such that the regioselectivity is rather high for positions sn1 and sn3 so that sn2 is degraded rarely (Figure 2); c) stereospecificity: faster hydrolysis of one primary sn ester as compared to its counterpart and d) combination of all of the

specificities mentioned above (Benjamin and Pandey *et al.*, 1998; Svendsen, 2000).

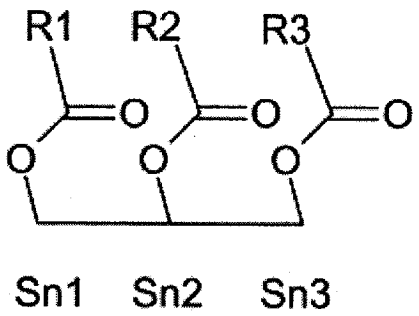


Figure 2: The triacylglycerol type of lipid presented using the *sn* designation for positions of the glycerol moiety (Adopted from Svendsen, 2000).

B. 4 Importance of lipases

Industrial and Biological applications

Lipases have evolved to be efficient catalysts for lipolytic reactions concerning the hydrolysis of ester linkages of mono-, di- and triglycerides in aqueous emulsions. The hydrolysis of these bonds is important in industrial applications:

- Generation of fatty acids from natural oils for the production of soaps
- Removal of oils and fats from fabrics, machinery, hides and waste water
- Production of mono- and diglycerides for food emulsifiers.

Lipases are also used to conduct transesterification reactions in commercial applications, such as the production of cocoa butter substitutes (Sharma *et al.*, 2001). Since lipases have the unique ability to act at oil-water interfaces, their structures are of particular interest. The selectivity of lipases towards the length of the fatty acids or the number and location of unstaurations in the fatty acids is employed to produce high-value fats or oils (Macrae *et al.*, 1983). The high stereoselectivity of lipases is also exploited to synthesize specific compounds-- in particular, enantiomerically pure compounds

containing ester bonds. For thermodynamic reasons, these synthesis reactions must be conducted in environments containing little water. There have been many reports of lipases in organic solvents synthesizing a variety of compounds (Boland *et al.*, 1991), including precursors for biologically active therapeutics, herbicides, and pesticides.

Though lipolytic enzymes are widely distributed in the plant kingdom, knowledge of lipases from plants is limited when compared with those from mammalian systems and micro-organisms. The low abundance of these proteins makes it difficult to purify them in amounts sufficient to get access to amino acid sequence information. During post-germination of oil seed plant (eg. Arabidopsis thaliana), the growth of the seedling is supported mainly by hydrolysis of the oil reserve (Huang *et al.*, 1990). Lipases are involved in the first step of this series of reactions and may control a crucial step in post-germination of seed. Plant lipases play a role in the production of goods such as bread and beer. They may decrease the shelf life of plant-derived foods, due to their participation in the processes that govern spoilage (Mukherjee *et al.*, 1994). Clearly, lipases are widely employed in the food, cosmetic, detergent and pharmaceutical industries.

B. 5 Structural Bioinformatics

The Nomenclature Committee of the International Union of Biochemistry (IUB) has classified hydrolases according to substrates recognized. The term lipase commonly refers to triacylglycerol hydrolases (EC 3.1.1.3). However, as pointed out by the Committee, many of the enzymes hydrolyse a wide variety of ester substrates-- e.g., phospholipases, cholesterol esterase (EC 3.1.1.13), cutinase, amidase and other esterase

type activities that makes the classification somewhat arbitrary (Schmid and Verger, 1998; Bornscheuer and Kazlauskas, 1999). All of these enzymes can be regarded as lipases.

Hydrolases are enzymes that catalyze the hydrolysis of various bonds. Some of these enzymes pose problems because of their wide specificity or substrate compatibility. Thus, deciding whether two preparations described by different authors are the same can be difficult. While the systematic name always includes 'hydrolase', the common name is, in most cases, formed by the name of the substrate with the suffix "ase". It is understood that substrates with this suffix, and no other indicator, indicate a hydrolytic enzyme.

Fungi produce different classes of lipolytic enzymes, including carboxylesterases (EC 3.1.1.-), true lipases (EC 3.1.1.3) and a diverse group of phospholipases A1, A2, B, C and D (Figure 3).

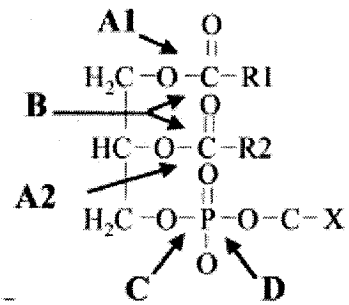
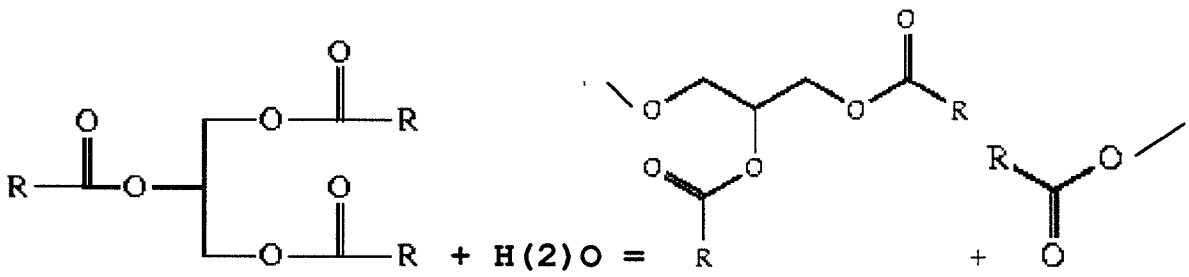


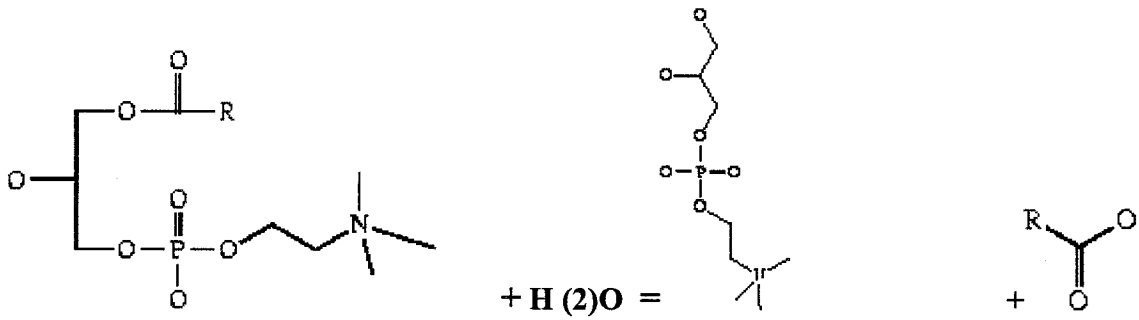
Figure 3: Sites of action of phospholipase A1, A2, B, C and D. The generic phospholipase structure is shown with the corresponding action sites of phospholipases designated by an arrow (adapted from Cox *et al.*, 2001).

Carboxylesterases are enzymes which hydrolyze small ester containing molecules at least partly soluble in water. True lipases are enzymes that hydrolyse triglycerides into free fatty acids and glycerol. Triacylglycerol lipases are present in animals, plants, fungi, and bacteria. They have mainly triacylglycerol activity and are classified in the EC 3.1.1.3 group. The enzyme reaction is shown below by the following equation:



Triacylglycerol **Diacylglycerol** **a × fatty acid anion**

Phospholipids, on the other hand, have an important role in maintaining the structure and function of the cell membrane. Phospholipases generate second messengers, participate in cytotoxity, and hydrolyze phospholipids in the gastrointestinal tract. They assist in regulating cellular functions in both mammalian and plant cells, and lower eukaryotes, including yeast and slime moulds. Phospholipases are a heterogeneous group of enzymes that are able to hydrolyse one or more ester linkages in glycerophospholipids. The actions of phospholipases can result in the destabilization of membranes, cell lyses and release of lipid second messengers (Ghannoum, 2000; Schmiel and Miller, 1999). These enzymes are categorized according to the specificity of the ester link that is cleaved (Figure 3). Five kinds of phospholipase activity have been described, including Phospholipase A1 and A2 (3.1.1.4); phospholipase B or lysophospholipase (3.1.1.5); phospholipase C or phosphatidylinositol-specific phospholipase C (EC 3.1.4.11) and phospholipase D (3.1.1.4). In this study, all phospholipases will be described except phospholipase A1 and A2 which were not experimentally found according to SWISS-PROT data base, but still they will be compared to phospholipase B.



2-lysophosphatidylcholine **glycerophosphocholine** **a × fatty acid anion**

Three-dimensional (3-D) and protein sequences are used to predict aspects of molecular function and evolution. Secondary structure prediction is a base of structure-based sequence analysis. Therefore, finding tools that accurately predict secondary structures are essential (Chen *et al.*, 1999; Klepeis *et al.*, 2003; McGuffin and Jones, 2003). Early studies by Dickerson *et al.* (1976) showed that multiple sequence alignments could be used to predict proteins structures. Zvelebil *et al.* (1987) integrated this theory into an automatic prediction method. These early studies suffered from lack of data and were used only on single sequences. Given the large families of homologous sequences now available and advancements in computing techniques, prediction accuracy is now above 75%.[1] More successful recent methods are PHD (Rost and Sander, 1994) and PSIPRED (Jones, 1999), reporting accuracy above 76%. Frishman and Argos (1997) hit 74.8% using PREDATOR and 79% using PORTER.

The results of the public prediction service (Predict Protein) have been used to determine protein structures (chain tracing in X-ray crystallography), as well as to formulate hypotheses about protein structure and function that guided experiments in molecular biology (Rost, 1998).

[1] Accuracy varies between different proteins (72% ±9%).

Secondary structure prediction tools used in this study are available on the SWISSPROT and NPS@ server. The SWISSPROT database is a manually annotated protein knowledgbase established in 1986 which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure post-translational modifications, etc.). NPS@ is the IBCP (Institute of Biology and Chemistry of Proteins) server contribution to PBIL in Lyon, France, and is a server developing several methods dedicated to protein secondary structure prediction, detection of fuzzy motifs in protein sequences, protein sequence analysis software, and so forth. From the SWISSPROT database, secondary structure tools used in this study include, SOPM (Geourjon and Delage, 1994), SOPMA (Geourjon and Delage, 1995), PSI-PRED (Jones, 1999), PROF. Prediction (Rost, 2001), JUFO (Meiler, 2002), PORTER (Pollastri *et al.*, 2002), NNPREDICTION (Kneller *et al.*, 1990), SSPRO (Baldi and Pollastri, 2003) and from NPS@ server, HNN (Guermeur, 1997), PHD (Rost, 1996), PREDATOR (Argos *et al.*, 1996), SIMPA 96 (Levin et al., 1996), MLRC (Guermeur *et al.*, 1998), DSC (Guermeur *et al.*, 1998) and GORIV (Guermeur, 1996).

B. 5.1 Secondary Structure Prediction tools available from SWISS-PROT

B. 5.1.1 SOPM

Self-optimized method (SOPM) is a server for protein secondary structure prediction and was developed by Geourjon and Deleage in 1994. This method checks against an updated release of the Kabsch and Sander database, Data base of secondary structure prediction (DSSP), comprise protein chains. There are four steps in this method:

(1) Building a sub-database drawn from DSSP using binary comparisons

of all protein sequences.

(2) Taking into account the prediction of structural classes of proteins.

(3) Predicting the secondary structure using an algorithm based on sequence similarity.

(4) Iteratively determining the predictive parameters that optimize the prediction quality on the whole sub-database and applying the final parameters to the query sequence.

This method correctly predicts the secondary structures 69% of amino acids.

B. 5.1.2 SOPMA

SOPM was improved a year later in 1995 by Geourjon and Deleage to predict all the sequences from a set of aligned proteins belonging to the same family. The result was SOPMA, which had 69.5% accuracy.

B. 5.1.3 PSI-PRED

PSIPRED is a straightforward and reliable secondary structure prediction method using the output obtained from PSI-BLAST (Position Specific Iterated - BLAST). It incorporates this output with feed-forward neural networks for analysis. This program achieved the highest accuracy of 80.6% across all 40 submitted target domains with no obvious sequence similarity to structures present in PDB, which placed PSIPRED in first place out of 20 evaluated methods (an earlier version of PSIPRED was also ranked first in CASP3 held in 1998).

B. 5.1.4 PROF. Prediction

This program uses neural network on large data using different PSI-BLAST profiles, using the same concept implemented in PHD program. However, it does use more input data and has a third network layer for more accuracy.

B. 5.1.5 JUFO

JUFO offers a protein secondary structure prediction with 75% accuracy from its primary sequence. A neural network was trained with an amino acid property profile plus the position-based scoring matrix of a blast run (Jones, 1999).

B. 5.1.6 SSPRO

SSPRO (advanced recursive neural network system) is a server for protein secondary structure prediction based on an ensemble of 11 bi-directional recurrent neural networks (BRNNs). The only method that appears to improve prediction accuracy significantly via particular algorithms, as opposed to more divergent profiles, is SSpro. The algorithm to obtain multiple alignments of homologue sequences, based on PSI-BLAST instead of BLAST, is exploited. Experiments on an independent test set yields 78% accuracy (Pollastri *et al.*, 2002).

B. 5.1.7 PORTER

Porter is a new system for protein secondary structure prediction in three classes—it is an evolution of SSPRO. It increases the BRNN's (Bi-directional Recurrent Neural Networks) from 11 to 45. Porter was tested by a rigorous validation procedure

and achieves 79% accuracy (Pollastri *et al.,* 2002).

B. 5.1.8 NNPREDICT

This program predicts the secondary structure type for each residue in an amino acid sequence. Similar to other programs, it uses sequence as input. The output is a secondary structure prediction for each position in the sequence. To increase accuracy, it takes the tertiary class to account for the possible options (none, all-alpha, all-beta, or alpha/beta) for prediction. The best case prediction was 79% for the class of all-alpha proteins.

B. 5.2 Secondary structure prediction tools from NPS@

B. 5.2.1 HNN

The HNN (Hierarchical Neural Network) prediction method is made up of two networks: a sequence-to-structure network and a structure-to-structure network. Its prediction is based only on local information. Thus, the first layer network predicts the secondary structure of the central residue-- this is called the sequence-to-structure network. The second layer network, called the structure-to-structure network, filters the outputs from the first one and produces the final prediction results. To improve the accuracy of this tool, physico-chemical data have been taken into account for the structure-to-structure network.

B.5.2.2. PHD

PHD is neural network system (a sequence-to-structure level and a structure-structure level) to predict secondary structure (PHDsec), relative solvent accessibility (PHDacc) and trans membrane helices (PHDhtm) (Rost and Sander, 1993). The NPS@ server only uses PHDsec. PHDsec focuses on predicting hydrogen bonds. This program uses the BLASTP search for sequences and filters the result by aligning them with CLUSTALW. It uses these results as the input for the neural network.

B. 5.2.3. PREDATOR

PREDATOR is a secondary structure prediction method based on recognition of potential hydrogen-bonded residues in a single amino acid sequence (Frishman and Argos, 1996). The unique feature of this approach involves statistics on residue type occurrences in different classes of beta-bridges to describe interaction of beta-strands. The alpha-helical structures are also recognized on the basis of amino acid occurrences in hydrogen-bonded pairs.

B. 5.2.4. SIMPA96

This secondary structure prediction algorithm assumes that short homologous sequences of amino acids have the same secondary structure tendencies. Comparisons are made between secondary structure predictions from an X-ray database and an empirically determined similarity matrix which assigns a sequence similarity score between any two sequences of seven residues in length. This homologue method had a prediction accuracy of 62.2%.

B. 5.2.5. MLRC

MLRC (Multivariate Linear Regression Combination) is a secondary structure prediction method which combines GOR4, SIMPA96 and SOPMA (Guermeur *et al.*, 1999). It post-processes the outputs of protein secondary structure prediction methods and generates class posterior probability estimates. Experimental results establish that it can increase the recognition rate of methods that provide inhomogeneous scores, even if their individual prediction successes are largely different.

B. 5.2.6. DSC

Discrimination of protein Secondary structure Class (DSC) generates probabilities for helix and strand secondary structural states of each residue in a domain sequence. For input, it uses associated multiple sequence alignment CLUSTALW. This makes the prediction method comprehensible and allows the relative importance of the different sources of information used to be measured (King & Sternberg, 1996). The DSC method from multiply aligned homologous sequences has an overall per residue three-state accuracy of 70.1% accuracy.

B. 5.2.7. GORIV

GOR IV is the fourth version of GOR secondary structure prediction methods based on the information theory (Garnier *et al.*, 1996). GOR IV uses all possible pair frequencies within the window of 17 amino acid residues. GOR IV has a mean accuracy of 64.4% for a three state prediction.

C. METHODS

C. 1 Comparative Study of Fungal Lipases

C.1.1 Multiple Sequence Alignment

Protein sequences selected for this study includes fungal lipases from "true" lipases, lysophospholipase and phospholipase C whose activities have been experimentally determined. Fifty-two such lipases from diverse fungal species were selected by surveying publications on search engines such as SWISS-PROT (http://us.expasy.org/sprot) and NCBI (http://www.ncbi.nlm.nih.gov), and by browsing sequence databases such as GeneBank, and SWISS-PROT. The amino acid sequences of these lipases were aligned using Multiple Sequence Alignment program, CLUSTAL W (Thompson *et al.*, 1994) at http://www.ebi.ac.uk/clustalw. Upon completion of sequence alignment, the data were saved as TEXT-FILE ONLY format prior to transfer to a data exploration program, MacClade 4.03 (Maddison and Maddison, 2002), and then to a phylogenetic tree-building and analysis program, PAUP 4.0, beta 8 (Swofford *et al.*, 2001). All the alignments were visually inspected and manually edited, as needed, in MacClade 4.03 before tree reconstruction. The BLAST tool (Altschul *et al.*, 1997) was also used for sequence similarity searches. The experimentally-determined 3-D structures were retrieved from the Protein Data Bank (Berman *et al.*, 2000). The protein structures were displayed by the program Cn 3-D from NCBI.

C. 1.2 Phylogenetic Analysis and Statistics (Distance Analysis)

UPGMA (Unweighted Pair Group Method with Arithmetic Mean) was used in this study to determine the phylogenetic tree. The aligned amino acid sequences were analyzed using the default options of the program and all characters were weighted equally, gaps treated as missing characters, multi-state taxa interpreted as uncertain, and the distance measure was set to be equal to mean character difference.

C.2 Evaluation of Secondary Structure Prediction Tools for fungal Lipases

C.2.1 Secondary Structure Tools

The best way to assess the accuracy of a method is by carrying out "blind trials". A "blind trial" was used for the "protein Structure Prediction Challenge" meeting (Asilomar, California; 4-6 December 1994), where numerous prediction tools were tested to assess the accuracy of secondary structure prediction from multiple alignments and protein-fold recognition. The input for my analysis was the eight sequences with known X-ray structures (Table 1) and we also carried out "blind trials". For some lipase sequences there have been more than one crystal structure determined[2], thus we screened the Protein Data Bank (July 2005) to select lipase sequences with resolutions of 2.35 Å or better (Table 1). We used fifteen different programs available online from SWISS-PROT (http://us.expasy.org/sprot) and NPS@ http://umber.sbs.man.ac.uk/dbbrowser/bioactivity) to predict the secondary structure of sequences with unknown 3-D structures (Table 2).

[2] For instance, *Candida rugosa* has seven known structures including opened and closed.

Table 1: Sequences with known crystal structures (Information from Protein Data Bank)

PDB#	AC#	EC#	Name	Protein chain	Release Date	Resolution
1thg	P22394	3.1.1.3	Hydrolase(carboxylic esterase)	544 aa	1993	1.80 A
1gz7	P32946	3.1.1.3	Hydrolase	534 aa	2003	1.97 A
1cle	P32947	3.1.1.3	Cholesterol esterase	534 aa	1996	2.00 A
1llf	P32947	3.1.1.3	Hydrolase	534 aa	2003	1.40 A
1dt3	O59952	3.1.1.3	Hydrolase	269 aa	2000	2.69 A
1dt5	O59952	3.1.1.3	Hydrolase	269 aa	2000	2.40 A
1dte	O59952	3.1.1.3	Hydrolase	269 aa	2000	2.35 A
1du4	O59952	3.1.1.3	Hydrolase	269 aa	2000	2.50 A
1ein	O59952	3.1.1.3	Hydrolase	269 aa	2000	3.00 A
1lgy	P61871	3.1.1.3	Hydrolase(carboxylic esterase)	265 aa	1996	2.20 A
1lpm	P20261	3.1.1.3	Hydrolase	534 aa	1995	2.20 A
1lpn	P20261	3.1.1.3	Hydrolase	534 aa	1995	2.20 A
1lpo	P20261	3.1.1.3	Hydrolase	534 aa	1995	2.20 A
1lpp	P20261	3.1.1.3	Hydrolase	534 aa	1995	2.05 A
1trh	P20261	3.1.1.3	Hydrolase(carboxylic esterase)	534 aa	1994	2.10 A
1crl	P20261	3.1.1.3	Hydrolase(carboxylic esterase)	534 aa	1994	2.06 A
1lps	P20261	3.1.1.3	Hydrolase(carboxylic esterase)	534 aa	1995	2.20 A
1tca	P41365	3.1.1.3	Hydrolase(carboxylic esterase)	317 aa	1994	1.55 A
1tcb	P41365	3.1.1.3	Hydrolase(carboxylic esterase)	317 aa	1994	2.10 A
1tcc	P41365	3.1.1.3	Hydrolase(carboxylic esterase)	317 aa	1994	2.50 A
3tgl	P19515	3.1.1.3	Hydrolase(carboxylic esterase)	265 aa	1993	1.90 A
1tgl	P19515	3.1.1.3	Hydrolase(carboxylic esterase)	265 aa	1990	1.90 A
4tgl	P19515	3.1.1.3	Hydrolase(carboxylic esterase)	265 aa	1993	2.60 A
Abbreviations: AC#, Accession number; EC#, Enzyme commission; PDB#, Protein database bank; aa, amino acid;						

These programs classify each residue into three classes (H = α-helix, E/S = strand or β-sheets and the rest known as coils). Each sequence was analyzed separately using each of the above tools (Table 2) and an evaluation performed with respect to a scoring mechanism introduced in this study. A pair-wise similarity score in the range of (0, 1) is computed between all amino acid residues. Score "1" is given to amino acids that have correct matches for the two states, alpha and beta, (similar to crystal structure) and score "0" is given to others.

The tool that predicted secondary structure most similar to the known X-ray structure for each sequence was selected to run for the rest of the homologues sequences with unknown 3-D structure. These selected tool(s) predict secondary structures based on: statistical methods (Chou and Fasman, 1974), physico-chemical (Lim, 1974, Ptitsyn and Finkelstein 1983), sequence patterns, evolutionary conservation, and neural networks.

Table 2: Secondary structure prediction tools used in this study

	Programs	Reference
SWISS-PROT	SOPM	(Geourjon and Delage, 1994)
	SOPMA	(Geourjon and Delage, 1995)
	PSI-PRED	(Jones, 1999)
	Porter	(Pollastri *et al.*, 2002)
	PROF.-Prediction	(Rost. 2001)
	JUFO	(Jens Meiler *et al.*, 2002)
	SSPRO	(Baldi *et al.*, 2003)
	NNPREDICTION	(Kneller *et al.*, 1990)
NPS@	HNN	(Guermeur, 1997)
	GORIV	(Garnier *et al*, 1996)
	SIMPA 96	(Levin et al., 1996)
	PREDATOR	(Argos *et al.*, 1996)
	PHD	(Rost *et al.*, 1994)
	DSC	(King and Stenberg, 1996)
	MLRC	(Guermeur *et al.*, 1998)

C. 2.2 Three-dimensional (3-D) Structure Prediction

Vector Alignment Search Tool (VAST) from NCBI with cross-linking to the Molecular Modeling Data Base (MMDB) and Protein Data Bank (PDB) was used to obtain the 3-D structure and structural neighbors. In order to obtain these 3-D structures we entered the four-digit alphanumeric entry codes. To view the 3-D structures of fungal lipases we should select "view 3-D structure" box. In addition to the 3-D structure of the query, neighbors (similar structures) ranked by similarity will be shown.

D. RESULTS

D.1 Comparative Study of Fungal Lipases

The distance-based UPGMA analysis of 52 selected amino acid sequences of fungal lipases resulted in a single dendrogram (Figure 4), which revealed four major clusters of sequences and a branch with one sequence (P34163). The sequence P34163 is the only known fungal lipase of the super-family "AB hydrolase":

- Cluster 1 consisted of ten sequences of secretary lipases (group A), ten sequences of carboxylesterases (group B) and a single sequence, P54857, which occupied a basal position within this cluster.
- Cluster 2 consisted of six sequences of lipase.
- Cluster 3 consisted of 20 sequences of Phospholipase B.
- Cluster 4 consisted of three sequences of Phospholipase C and a single sequence of Phospholipase D, which was basal within this cluster.

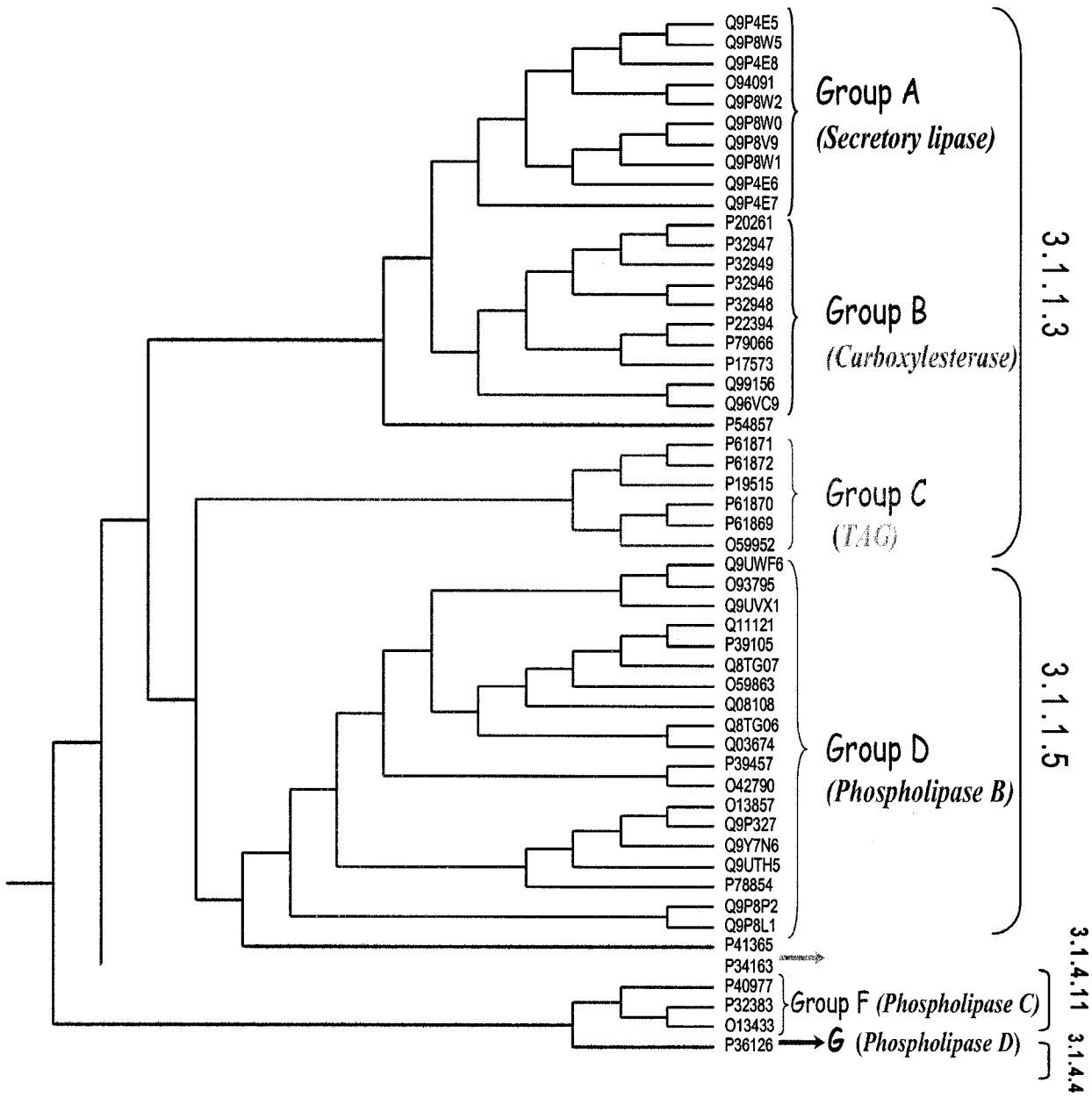


Figure 4: Phylogenetic tree using Neighbor joining method (UPGMA)

Table 3. Families and subfamilies of fungal lipolytic enzymes

Family	Sub-family	AC #	Organism	Properties	Molecular mass	Length	CATALYTIC ACTIVITY
LIP (Secretory lipase)	AI	Q9P4E5	Candida albicans (Yeast	lipase 10	50 kDa	465	TAG+ H2O = diacylglycerol + a carboxylate.
	AI	Q9P8W5	Candida albicans (Yeast)	lipase 2	50 kDa	466	TAG+ H2O = diacylglycerol + a carboxylate.
	AI	O94091	Candida albicans (Yeast)	lipase 6	50 kDa	463	TAG + H2O = diacylglycerol + a carboxylate.
	AI	Q9P8W2	Candida albicans (Yeast)	lipase 1	50 kDa	468	TAG + H2O = diacylglycerol + a carboxylate
	AI	Q9P8W2	Candida albicans (Yeast)	lipase 3	51 kDa	471	TAG+ H2O = diacylglycerol + a carboxylate.
	AII	Q9P8W0	Candida albicans (Yeast)	(lipase 5)	50 kDa	463	TAG+ H2O = diacylglycerol + a carboxylate.
LIP (Secretory lipase)	AII	Q9P8V9	Candida albicans (Yeast)	(lipase 8)	50 kDa	460	TAG + H2O = diacylglycerol + a carboxylate.
	AII	Q9P8W1	Candida albicans (Yeast)	lipase 4	50 kDa	459	TAG + H2O = diacylglycerol + a carboxylate.
	AII	Q9P4E6	Candida albicans (Yeast)	lipase 9	49 kDa	453	TAG + H2O = diacylglycerol + a carboxylate.
	A	Q9P4E7	Candida albicans (Yeast)	lipase 7	49 kDa	426	TAG+ H2O = diacylglycerol + a carboxylate.
Carboxylesterase	BI	P20261	Candida rugosa (Candida cylindracea)	lipase 1	59 kDa	549	TAG + H2O = diacylglycerol + a carboxylate.
	BI	P32947	Candida rugosa (Candida cylindracea)	Lipase 3	59 kDa	549	TAG+ H2O = diacylglycerol + a carboxylate.
	BI	P32949	Candida rugosa (Candida cylindracea)	lipase 5	59 kDa	549	TAG + H2O = diacylglycerol + a carboxylate.

Carboxylesterase	BI	P32946	Candida rugosa (Candida cylindracea)	lipase 2	59 kDa	548	TAG + H2O = diacylglycerol + a carboxylate.
	BI	P32948	Candida rugosa (Candida cylindracea)	lipase 4	59 kDa	549	TAG + H2O = diacylglycerol + a carboxylate.
	BII	P22394	Geotrichum candidum (Oospora lactis)	(lipase 2)	61 kDa	563	TAG + H2O = diacylglycerol + a carboxylate.
	BII	P79066	Geotrichum fermentans (Trichosporon fermentans)	Lipase1	61 kDa	563	TAG+ H2O = diacylglycerol + a fatty acid anion
	BII	P17573	Geotrichum candidum (Oospora lactis)	Lipase1	61 kDa	563	TAG + H2O = diacylglycerol + a fatty acid anion
	BIII	Q99156	*Yarrowia lipolytica (Candida lipolytica)*	lipase 1	55 kDa	486	TAG + H2O = diacylglycerol + a carboxylate
	BIII	Q96VC9	*Yarrowia lipolytica (Candida lipolytica)*	Lipase 3	56 kDa	498	TAG+ H2O = diacylglycerol + a carboxylate
Lipase_3 (Lipase class 3)	CI	P61871	*Rhizopus niveus*	Lipase, TAG, Lipase II	42 kDa	392	TAG + H2O = diacylglycerol + a carboxylate.
	CI	P61872	*Rhizopus oryzae (Rhizopus delemar)*	Lipase, TAG	42 kDa	392	TAG+ H2O = diacylglycerol + a carboxylate.
	CI	P19515	*Rhizomucor miehei*	Lipase, TAG	40 kDa	363	TAG + H2O = diacylglycerol + a carboxylate.
	CII	P61870	*Penicillium camembertii*	Mono- and diacylglycerol lipase	33 kDa	305	--
	CII	P61869	*Penicillium cyclopium*	Mono- and diacylglycerol lipase	33 kDa	305	--

PLA2_B (Lysophospholipase catalytic domain)	CII	O59952	*Thermomyces lanuginosus (Humicola lanuginosa)*	Lipase, TAG	32 kDa	291	TAG + H2O = diacylglycerol + a carboxylate.
	DI	Q9UWF6	*Candida albicans*	Lysophospholipase 1, phospholipase B1	66 kDa	605	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DI	O93795	*Candida albicans*	Lysophospholipase 2 , phospholipase B 2	67 kDa	608	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DI	Q9UVX1	*Candida albicans*	Lysophospholipase 3 , phospholipase B 3	81 kDa	754	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DII	Q11121	*Torulaspora delbrueckii (Saccharomyces rosei)*	Lysophospholipase , phospholipase B	71 kDa	649	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DII	P39105	*Saccharomyces cerevisiae*	Lysophospholipase , phospholipase B1	72 kDa	664	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DII	Q8TG07	*Candida glabrata (Torulopsis glabrata)*	Lysophospholipase 1, phospholipase B1	72 kDa	659	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DII	O59863	*Kluyveromyces lactis (Yeast)*	lysophospholipase, Phospholipase B	70 kDa	640	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DII	Q08108	*Saccharomyces cerevisiae*	Lysophospholipase 3 , phospholipase B 3	75 kDa	686	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
PLA2_B (Lysophospholipase	DII	Q03674	*Saccharomyces cerevisiae*	Lysophospholipase 2 , phospholipase B 2	75 kDa	706	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DII	Q8TG06	*Candida glabrata (Yeast) (Torulopsis glabrata)*	Lysophospholipase 2 , phospholipase B 2	75 kDa	695	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate

PLA2_B (Lysophospholipase catalytic domain)	DIII	P39457	*Penicillium chrysogenum (Penicillium notatum)*	Lysophospholipase , phospholipase B	66 kDa	612	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DIII	O42790	*Neurospora crassa*	lysophospholipase, Phospholipase B	70 kDa	653	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DIV	O13857	*Schizosaccharomyces pombe*	Putative lysophospholipase, Phospholipase B	69 kDa	624	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DIV	Q9P327	*Schizosaccharomyces pombe*	Phospholipase B	75 kDa	673	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DIV	Q9Y7N6	*Schizosaccharomyces pombe*	Phospholipase B	68 kDa	633	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DIV	Q9UTH5	*Schizosaccharomyces pombe*	Probable lysophospholipase, Phospholipase B	70 kDa	644	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DIV	P78854	*Schizosaccharomyces pombe*	Lysophospholipase 1, phospholipase B1	67 kDa	613	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DV	Q9P8P2	*Cryptococcus neoformans var. grubii*	Phospholipase B, Lysophospholipase	69 kDa	637	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
	DV	Q9P8L1	*Cryptococcus neoformans*	Phospholipase B	69 kDa	637	2-lysophosphatidylcholine + H2O = glycerophosphocholine + a carboxylate
Abhydrolase_1	E	P34163	*Saccharomyces cerevisiae*	-	63 kDa	548	-

PI-PLC-X, PI-PLC-Y	F	P32383	*Saccharomyces cerevisiae (Baker's yeast)*	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase 1, Phosphoinositide phospholipase C	100 kDa	869	1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate + H2O = 1D-myo-inositol 1,4,5-trisphosphate + diacylglycerol
	F	O13433	*Candida albicans*	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase 1, Phosphoinositide phospholipase C	124 kDa	1099	1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate + H2O = 1D-myo-inositol 1,4,5-trisphosphate + diacylglycerol
	F	P40977	*Schizosaccharomyces pombe*	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase 1, Phosphoinositide phospholipase C	102 kDa	899	1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate + H2O = 1D-myo-inositol 1,4,5-trisphosphate + diacylglycerol
PLDc	G	P36126	Saccharomyces cerevisiae (Baker's yeast)	Phospholipase D1, PLD1	160 kDa	1380	phosphatidylcholine + H2O = choline + a phosphatidate
	Other Lipase	P54857	Saccharomyces cerevisiae (Baker's yeast	Lipase 2, Triacylglycerol lipase	38 kDa	326	Triacylglycerol + H2O = diacylglycerol + a carboxylate
	Other Lipase	P41365	Candida antarctica (Yeast) (Trichosporon oryzae)	Lipase B (CALB)	36 kDa	342	Triacylglycerol + H2O = diacylglycerol + a carboxylate.

Abbreviations: AC#, Accession number; EC#, Enzyme commission;

D.1.1 Group A (Secretory lipase)

The group A consists of lipase 1 (O94091), lipase 2 (Q9p8W5), lipase 3 (Q9P8W2), lipase 4 (Q9P8W1), lipase 5 (Q9P8W0), lipase 6 (Q9P4E8), lipase 7(Q9P4E7), lipase 8 (Q9P8V9), lipase 9 (Q9P4E6), and lipase 10 (Q9P4E5) -- all members of a single gene family (Hube *et al.*, 2000). These gene sequences were identified from *Candida albicans* and have a molecular mass of 49-51 kDa. They show high similarity in sequence and the amino acid identities ranged from 36% between lipase (Q9P8V9 and Q9P4E7) to 85% (between Q9P8W0 to Q9P8V9) (Table 4).

Table. 4	Q9P4E5	Q9P8W5	Q9P4E8	O94091	Q9P8W2	Q9P8W0	Q9P8V9	Q9P8W1	Q9P4E6	Q9P4E7
Q9P4E5	0	68%	67%	56%	58%	53%	52%	51%	50%	38%
Q9P8W5		0	65%	58%	57%	52%	52%	50%	50%	39%
Q9P4E8			0	59%	63%	52%	53%	51%	49%	38%
O94091				0	73%	55%	55%	55%	53%	39%
Q9P8W2					0	52%	51%	50%	49%	39%
Q9P8W0						0	85%	81%	76%	36%
Q9P8V9							0	82%	73%	36%
Q9P8W1								0	76%	37%
Q9P4E6									0	36%
Q9P4E7										0

Table 4. Pairwise comparison of lipase1-lipase 10 using the program NCBI (Align two sequences BLAST).

These ten isoenzymes could be further divided in to two subfamilies (A.I and A.II):

Sub-family A.I includes Q9P4E5, Q9P8W5, Q9P4E8, O94091, and Q9P8W2 with more than 56% identity (Table 4) while sub-family A.II includes, Q9P8W0, Q9P8V9, Q9P8W1 and Q9P4E6 which were at least 49% identical to each other (Table 4). Lipase 7 (Q9P4E7) of the "family A" occupied a basal position to A.I and A.II, (Table 3 and Figure 4). The lipases of sub-family A.I have a molecular mass of 50-51

kDa, and contain a sequence of 16 amino acids as a signal peptide. In contrast to the typical catalytic triad (Ser, Asp and His) found in other lipases, the family A.I lipases have a dyad catalytic site with two residues for the catalytic activity at positions 196 (serine) and 344 (histidine). Similar to other lipases, the group A.I lipases have a conserved motif, Gly -X_1- Ser -X_2- Gly around the serine residue. The residues X_1 and X_2 of sub-family A.I lipases are glycine and tyrosine respectively. This sub-family includes conserved *N*-glycosylation site at positions 231 and 319 and at a few more positions as given in Figure 5. They consist of four conserved cysteine residues and a conserved lipase motif which may form disulfide bridges, yielding a similar 3-D structure. However, no crystallography-based structures are available for these protein sequences.

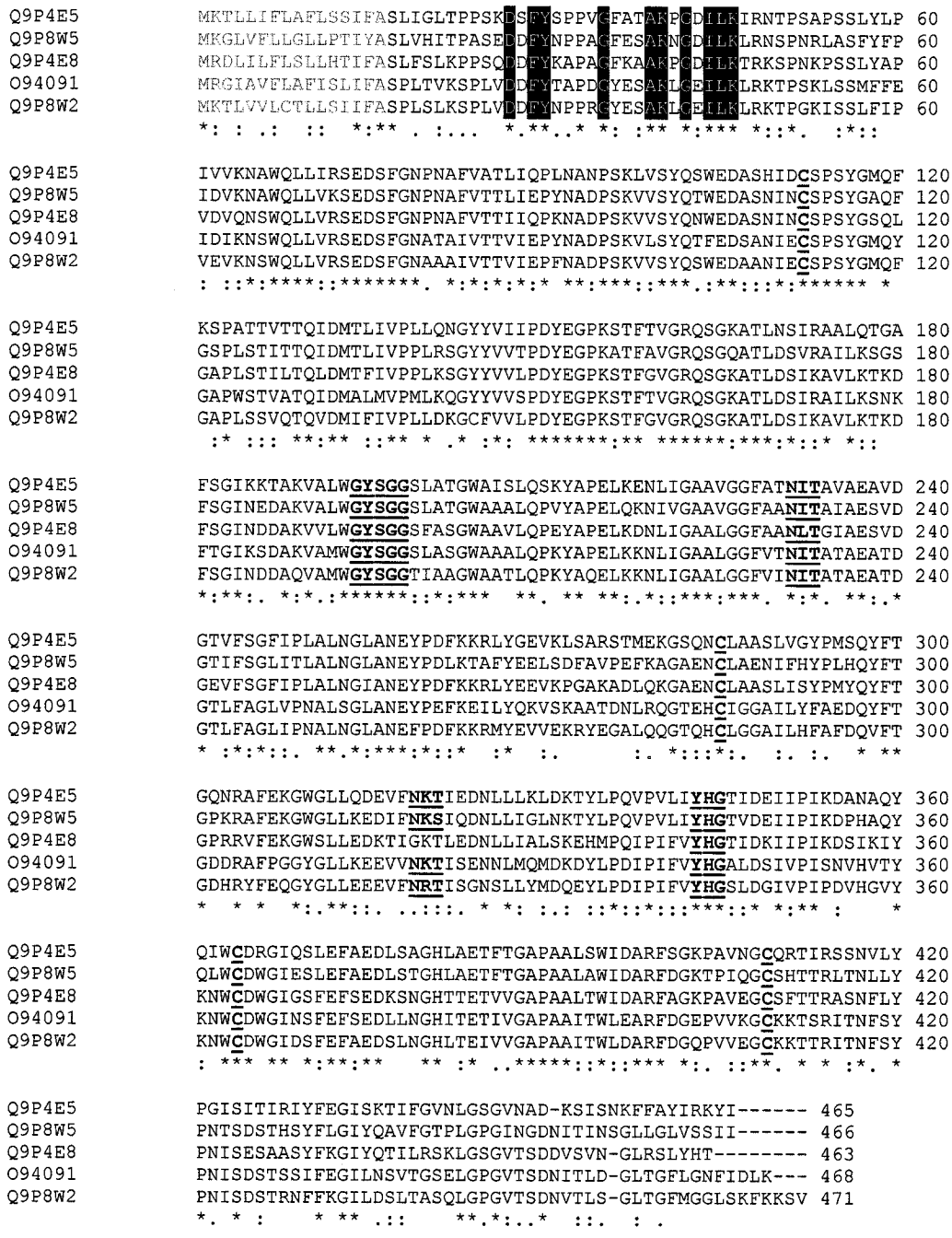


Figure 5. CLUSTALW sequence alignment of Secretory lipase (sub-family A.I). The 16 amino acid signal peptide, the active site (Ser, His) along with the consensus lipase sequence GxSxG, the potential N-glycolysation site and cysteine residues available for disulfide bond formation are shown in this figure.

```
Q9P8W0          MLYLILFLIAPIYAGLIFPTKPSSDPFYNPPKGFENAAVGDILQSRATPKSITGGFTPLK 60
Q9P8V9          MLFLLFLLITPIYAGLIFPTKPSSDPFYNPPKGFEKAAVGDILQSRETPKSITGRFAPLK 60
Q9P8W1          MLFLLFLLVAPIYAGLILPTKPSNDPFYNAPAGFEKAAVGDILQSRKTPKPITGVFVPVK 60
Q9P4E6          MLYLILFLIAPIYAGVLLPTKPSIDPFYNAPEGFKNATVGDILQFRKTPKSITGGFVPLN 60
                **:*:::*::*****:::***** *****.* **::*:****** * ***.*** *.*::

Q9P8W0          IQNSWQLLVRSEDSFGNPNVIVTTVIEPVNADPSKIASYQVFEDAAKADCAPSYALQFGS 120
Q9P8V9          IQNSWQLLVRSEDSFGNPNAIVTTVIEPVNADPSKIASYQVFEDAAKADCAPSYALQFGS 120
Q9P8W1          IQNSWQLLVRSEDSFGNPNVIVTTVMEPFNADPSKLASYQVFEDSAKADCAPSYALQFGS 120
Q9P4E6          VQNSWQFLVRSEDSFGNPNVIVTTVIEPVNADPSKIASYQVSENAARADCAPSYALQFGS 120
                :*****:************.*****:**.******:***** *::*:**************

Q9P8W0          DLTTFVTQAEMYLMAPLLDQGYYVVSPDYEGPKSTFTIGKQSGQAVLNSIRAALKSGKIT 180
Q9P8V9          DLTTFVTQAEMYLMAPLLDQGYYVVSPDYEGPKSTFTIGKQSGQAVLNSIRATLKSSKIT 180
Q9P8W1          DVTTIATQVETYLLAPLLDQGYYVVSPDYEGPKSTFTVGKQSGQAVLNSIRAALKSGKIT 180
Q9P4E6          DVSTLATQAETYLLAPLLDKGYYVVSPDYEGPKSTFTVGKQSGQAVLNSIRASLKSGKIT 180
                *::*:.**.* **:*****:*****************:**************:***.***

Q9P8W0          NIKDDAKVVMWGYSGGSLASGWAAALQPSYAPELGGNLLGAALGGFVTNITATAQATDGT 240
Q9P8V9          NIKEDAKVVMWGYSGGSLASGWAAALQPSYAPELSSSLLGAALGGFVTNITATAQAADGT 240
Q9P8W1          NLAENAKVVMWGYSGGSLASGWAAALQPNYAPELGGNLLGAALGGFVTNITATAEATDGT 240
Q9P4E6          NIAEDAKVLMWGYSGGSLASGWAAALQPDYAPELSRNLLGVALGGFITNVTATVEATDDT 240
                *: ::***:********************.*****. .***.*****:**:***.:*:*.*

Q9P8W0          VFAGIVANALGGVANEYPEFKSILQSDTDKKSVFEEFDGHCLIDGVLNYIGTSFLTGDHK 300
Q9P8V9          VFAGIVANALGGVANEYPEFKSILQSDTDKKSVFDEFDSHCLADGVIDYINTSFLTGDNK 300
Q9P8W1          VFAGIMANALGGVANEYPEFKQILQNDTDKQSVFDQFDNHCLADGVINYIGKHFLSGTNK 300
Q9P4E6          IFAGTAANVLGGIANEYPEFKSILQNDTNKSSIFNKINNHCLTDSFIKYVGARFLTGDNK 300
                :***  **.***:********.***.**:*.*:*::::.*** *..:.*:.  **:* :*

Q9P8W0          IFKTGWDILKNPKIGKVVEDNGLVYQKQLVPKIPVFIYHGSIDQIVPIVDTKKTYQNWCD 360
Q9P8V9          IFKTGWDILKSPTIAKIVEDNGLVYQKQLVPKIPIFVYHGSIDQIVPIVNVKKTYQNWCE 360
Q9P8W1          IFKSGWNILKNPTISKIVEDNGLVYQKQLVPKIPILIYHGAIDQIVPIVNVKKTYQNWCD 360
Q9P4E6          VFKSGWNIFKNLVVSKIVKDNGLVYQKQLIPTIPVFIYHGSMDQISPILNPKKTYQNWCD 360
                :**:**:*:*.  :.*:*:**********:*.**:::***::*** **:: ********:

Q9P8W0          AGISSLEFAEDASNGHLTEAIMGAPAALTWIIDRFDGKQTVSGCQHIQRFSNLEYPNIPS 420
Q9P8V9          GGISSLEFAEDGTNGHLTETVVGAPAALTWIIDRFNGKQTVSGCQHDKRLSNFQYPNISS 420
Q9P8W1          AGIASLEFSEDATNGHITETIVGAPVALTWIINRFNGKQTVSGCQHVKRTSNFEYPNIPP 420
Q9P4E6          AGISSIEFAEDLTNGHFTESIVGAPAALTWIIDRFSNKPPVDGCQHVVRTTNYEYPNVSS 420
                .**:*:**:** :***:**:::***.******:**..* .*.****  * :* :***:..

Q9P8W0          SIANYFKAAMDVVLHLGLGPDVQKDQVSPEGIKKLGSIEMRWL 463
Q9P8V9          SILKYFKVALDTMMSNGLGSDIQKDKITPDDLRKF--LLGGW- 460
Q9P8W1          SILNYFKAALNILIQKGLGPDIQKDQVNPDGLKKISILV---- 459
Q9P4E6          SILDYFKAAMDVVAQQGLGPNIQKDQLEIKSNL---------- 453
                ** .***.*:: :   ***.::***::  ..
```

Figure 6. CLUSTALW sequence alignment of Secretory lipase (sub-family A.II). The 14 amino acid signal peptide, the active site (Ser, His) along with the consensus lipase sequence Gly-x-Ser-x-Gly, the potential N-glycolysation site and cysteine residues available for disulfide bond formation are shown in this figure.

The A.II sub-family is similar to A.I sub-family, but with a few key differences. They include sequences with accession numbers of Q9P8W0, Q9P8V9, Q9P8W1 and Q9P4E6 of lipases. They have fourteen amino acids as the signal peptide, (in contrast to sixteen amino acids in A.I). An *N*-glycosylation site was only conserved at residue 229. The four cysteine residues, (residues 110-281 and 359-404) and the conserved dyad catalytic sites are identical to A.I sub-family (Figure 6).

D.1.2 Group B (Carboxylesterases)

The family B comprises ten lipases from *Ascomycota*, which could be further divided into three sub-families (B.I, B.II and B.III). The sub-family B.I includes five sequences from *Candida rugosa* with accession numbers of P20261, P32947, P32949, P32946 and P32948. All five isozymes have similar structure with a molecular mass of 59 kDa and contain 534 amino acids in their mature protein (main chain). The sequences in the sub-family B.II are from *Geotrichum,* and include sequences with accession numbers of P22394, P79066 and P17573. The molecular mass of these lipases are 61kDa. The sub-family B.III comprises of two sequences from *Yarrowia lipolytica (Candida lipolytica)*; yeast commonly used in genetic studies. The molecular mass of these two sequences (55 kDa) are lower than other sequences in the family B. The sequences in sub-family BIII include sequences with accession numbers of Q99156 and Q96VC9 (Figure 4 and Table 3). Although Q99156 and Q96VC9 are distinct from the rest of the sequences based on the distance dendrogram, they share common properties with the rest of the sequences in this family such as having the same EC number (3.1.1.3) and hydrolyzing triacylglycerol into diacylglycerol and a carboxylate. Unlike other lipases, the catalytic triad in group B (Carboxylesterase) is made up of the amino acids

Ser-Glu-His at positions 223, 355 and 365. The aspartate residue common to most lipases is replaced with glutamic acid in this group. However, in sequence Q99156, the aspartate catalytic residue remains unchanged. The motif Gly-X_1-Ser-X_2-Gly at the serine site (residue 193); (as given in figure 7) is conserved in this family. X_1 in this family is glutamic acid and X_2 is alanine. Family B.I and B.II include two disulfide bonds at regions (74-111) and (282-291) while there is only one disulfide bond found in sub-family B.III (Figure 7). *N*-glycosylation sites were conserved in sub-family B.I at position 365 in all five sequences, but at position 329 sequences P32946 and P32948 has diverged to a hydrophobic amino acid (phenylalanine). Moreover, these lipases show an α/β hydrolase fold, which is a common 3-D fold in several other hydrolases (Ollis *et al.*, 1992).

```
P22394          MVSKSLFLAAAVNLAGVLAQAPRPSLNGNEVISGVLEGKVDTFKGIPFADPPLNDLRFKH 60
P79066          MVSKSLFLAAAVNLAGVLAQAPTAVLNGNEVISGVLEGKVDTFKGIPFADPPLNDLRFKH 60
P17573          MVSKTFFLAAALNVVGTLAQAPTAVLNGNEVISGVLEGKVDTFKGIPFADPPVGDLRFKH 60
P20261          -----MELALALSLIASVAAAPTATLANGDTITGLNAIINEAFLGIPFAEPPVGNLRFKD 55
P32947          -----MKLALALSLIASVAAAPTAKLANGDTITGLNAIINEAFLGIPFAEPPVGNLRFKD 55
P32949          -----MKLALALSLIASVAAAPTATLANGDTITGLNAIINEAFLGIPFAEPPVGNLRFKD 55
P32946          -----MKLCL-LALGAAVAAAPTATLANGDTITGLNAIVNEKFLGIPFAEPPVGTLRFKP 54
P32948          -----MKLALVLSLIVSVAAAPTATLANGDTITGLNAIINEAFLGIPFAQPPVGNLRFKP 55
Q96VC9          ---------MPLELPSLNASIVGNTVQNG---------AVEQFLNIRYADIPG---KFEK 39
Q99156          -----------MSVTSTSLNGTFNGISED---------GIEIFKGIKYANIPYRWAYAER 40
                            : :           :  .           : * .* :*: *      :

P22394          PQPFTGSYQGLKANDFSPACMQLDPGNSLTLLDKALGLAKVIPEEFRGPLYDMAKGTVSM 120
P79066          PQPFTGSYQGLKANDFSPACMQLDPGNSLTLLDKALGLAKVIPEEFRGPLYDMAKGTVSM 120
P17573          PQPFTGSYQGLKANDFSSACMQLDPGNAISLLDKVVGLGKIIPDNLRGPLYDMAQGSVSM 120
P20261          PVPYSGSLDGQKFTSYGPSCMQQNPEGTYEENLPKAALDLVMQS-------KVFEAVSPS 108
P32947          PVPYSGSLNGQKFTSYGPSCMQQNPEGTFEENLGKTALDLVMQS-------KVFQAVLPQ 108
P32949          PVPYRGSLNGQSFTAYGPSCMQQNPEGTYEENLPKVALDLVMQS-------KVFQAVLPN 108
P32946          PVPYSASLNGQQFTSYGPSCMQMNPMGSFEDTLPKNARHLVLQS-------KIFQVVLPN 107
P32948          PVPYSASLNGQKFTSYGPSCMQMNPLGNWDSSLPKAAINSLMQS-------KLFQAVLPN 108
Q96VC9          PVLKN-DWNGAEIDATKVGPVCPQPRTPFNFFSVPDDLWEKVNVD-----------TYQD 87
Q99156          ----IDDYDNGVFDCTQEGMACP-QVLPFDYNIEKGPKEMPFDE---------------- 79
                      . :.       .                .

P22394          NEDCLYLNVFRPAGTKPDAKLPVMVWIYGGAFVYGSSA--AYPGNSYVKESINMGQPVVF 178
P79066          NEDCLYLNVFRPAGTKPDAKLPVMVWIYGGAFVYGSSA--AYPGNSYVKESINMGQPVVF 178
P17573          NEDCLYLNVFRPAGTKPDAKLPVMVWIYGGAFVFGSSA--SYPGNGYVKESVEMGQPVVF 178
P20261          SEDCLTINVVRPPGTKAGANLPVMLWIFGGGFEVGGTS--TFPPAQMITKSIAMGKPIIH 166
P32947          SEDCLTINVVRPPGTKAGANLPVMLWIFGGGFEIGSPT--IFPPAQMVTKSVLMGKPIIH 166
P32949          SEDCLTINVVRPPGTKAGANLPVMLWIFGGGFEIGSPT--IFPPAQMVSKSVLMGKPIIH 166
P32946          DEDCLTINVIRPPGTRASAGLPVMLWIFGGGFELGGSS--LFPGDQMVAKSVLMGKPVIH 165
P32948          GEDCLTINVVRPSGTKPGANLPVMVWIFGGGFEVGGSS--LFPPAQMITASVLMGKPIIH 166
Q96VC9          GLLCDNLIVTRPKGVSANARLPTVVWIHGGSNIEGSIYNLIYEPQFLVAESVRVGKPIVH 147
Q99156          -FECSNLMITRPQGATN---LPVFVWIHGGGNLAGNGYCSDHNPVPFVKHSIVAGRPVLH 135
                   *  : : ** *.     **..:**.**.   *.      .     :  *:  *:*::.

P22394          VSINYRTGPFGFLGGDAITAEGNTNAGLHDQRKGLEWVSDNIANFGGDPDKVMIFGESAG 238
P79066          VSINYRTGPFGFLGGDAITAEGNTNAGLHDQRKGLEWVSDNIANFGGDPDKVMIFGESAG 238
P17573          VSINYRTGPYGFLGGDAITAEGNTNAGLHDQRKGLEWVSDNIANFGGDPDKVMIFGESAG 238
P20261          VSVNYRVSSWGFLAGDEIKAEGSANAGLKDQRLGMQWVADNIAAFGGDPTKVTIFGESAG 226
P32947          VAVNYRVASWGFLAGDDIKAEGSGNAGLKDQRLGMQWVADNIAGFGGDPSKVTIFGESAG 226
P32949          VAVNYRLASFGFLAGPDIKAEGSSNAGLKDQRLGMQWVADNIAGFGGDPSKVTIFGESAG 226
P32946          VSMNYRVASWGFLAGPDIQNEGSGNAGLHDQRLAMQWVADNIAGFGGDPSKVTIYGESAG 225
P32948          VSMNYRVASWGFLAGPDIKAEGSGNAGLHDQRLGLQWVADNIAGFGGDPSKVTIFGESAG 226
Q96VC9          VCIEYRLGLAGFLT-----KNGKGNWGTWDQYTGCQWVNRHIQDFGGDPLNVTLTGESAG 202
Q99156          VMIEYRLSAFGYLAVPDTNGNWVGNWGARDQYTALQWISKHIVEFGGDPSQITIGGESAG 195
                * ::** .  *:*       :   * *  **  . :*:  :*  ***** :: : *****

P22394          AMSVAHQLIAYGGDNTYNGKKLFHSAILQSGGPLPYHDSSSVGPDISYNRFAQYAGCDTS 298
P79066          AMSVAHQLIAYGGDNTYNGKKLFHSAILQSGGPLPYHDSSSVGPDISYNRFAQYAGCDTS 298
P17573          AMSVAHQLVAYGGDNTYNGKQLFHSAILQSGGPLPYFDSTSVGPESAYSRFAQYAGCDAS 298
P20261          SMSVMCHILWNDGDNTYKGKPLFRAGIMQSGAMVPSDAVDGIYGNEIFDLLASNAGCGSA 286
P32947          SMSVLCHLIWNDGDNTYKGKPLFRAGIMQSGAMVPSDPVDGTYGNEIYDLFVSSAGCGSA 286
P32949          SMSVLCHLLWNGGDNTYKGKPLFRAGIMQSGAMVPSDPVDGTYGTQIYDTLVASTGCSSA 286
P32946          SMSTFVHLVWNDGDNTYNGKPLFRAAIMQSGCMVPSDPVDGTYGTEIYNQVVASAGCGSA 285
P32948          SMSVMCQLLWNDGDNTYNGKPLFRAAIMQSGAMVPSDPVDGPYGTQIYDQVVASAGCGSA 286
Q96VC9          SVAVHNMLIK----DSMNGRKLFRNAVMMSGTLETITPQPPKWHARLEEKVAKVTGKEVA 258
Q99156          SIGLHALMVHESMKPKEE--CIIHNVILSSGTMDRMG--TGTISENAFKPIYDGIKTLVG 251
                ::.    ::       . :   :::  :: **                 . .      .

P22394          ASANDTLECLRSKSSSVLHDAQNSYDLKDLFGLLPQFLGFGPRPDGNIIPDAAYELFRSG 358
P79066          ASANDTLECLRSKSSSVLHDAQNSYDLKDLFGLLPQFLGFGPRPDGNIIPDAAYELFRSG 358
P17573          AGDNETLACLRSKSSDVLHSAQNSYDLKDLFGLLPQFLGFGPRPDGNIIPDAAYELYRSG 358
P20261          S---DKLACLRGVSSDTLEDATN--NTPGFLAYSSLRLSYLPRPDGVNITDDMYALVREG 341
P32947          S---DKLACLRSASSDTLLDATN--NTPGFLAYSSLRLSYLPRPDGKNITDDMYKLVRDG 341
P32949          S---NKLACLRGLSTQALLDATN--DTPGFLSYTSLRLSYLPRPDGANITDDMYKLVRDG 341
P32946          S---DKLACLRGLSQDTLYQATS--DTPGVLAYPSLRLSYLPRPDGTFITDDMYALVRDG 340
P32948          S---DKLACLRSISNDKLFQATS--DTPGALAYPSLRLSFLPRPDGTFITDDMFKLVRDG 341
Q96VC9          D--------LASLSDKELLDAQIKLNVAVCMTCD----------DGDFFEPGWKQHLTPD 300
Q99156          D--------INTCSADELLEAQIKAGLDLGFYLQ----------DDFFPPDWRNVR--- 289
                         :   * . * .*    .    :             .  :
```

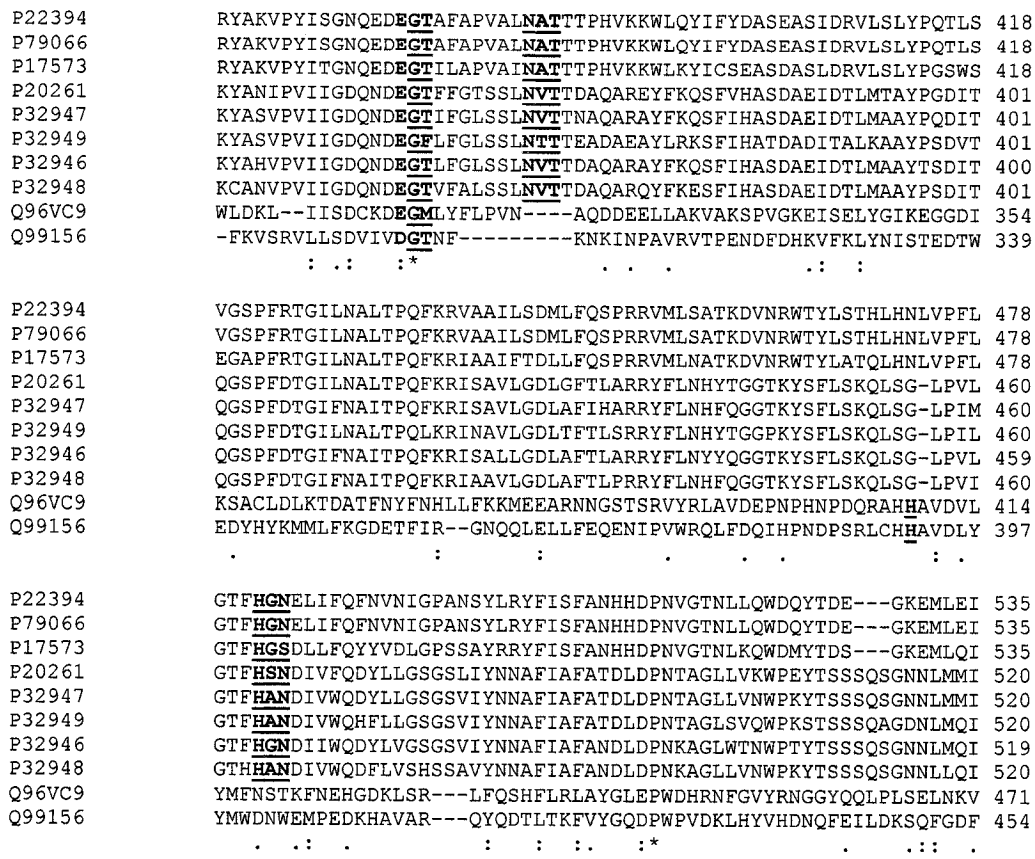


Figure 7: Amino acid sequence alignment carboxylesterase in family B including three sub-families (BI, BII, and BIII). The catalytic residues, N-glycosylation, the conserved proline residue (Pro 85 and phenylalanine residue 326) are shown in this figure.

D.1.3 Group C (AB hydrolase super family/TAG)

Family C comprises six sequences, further divided into two subfamilies (C.I and C.II). Group C.I has three sequences P61871 (*Rhizopus niveus),* P61872 (*Rhizopus oryzae*), and P19515 (*Rhizomucor miehei*). Sequences P61870 (*Penicillium camembertii*), P61869 (*Penicillium cyclopium*), and O59952 (*Thermomyces lanuginosus*) are in sub-family C.II (Figure 4). The 3-D structure for at least one sequence of each sub-family (P19515 and P61871 in sub-family C.I and O59952 in sub-family C.II) is known. The sub-family C.I consists of sequences from two filamentous fungi, *Rhizopus* and *Rhizomucor,* the only members of the phylum *Zygomycota* used in this study. The

lipases produced by these fungi are proteins of 363-392 amino acids with molecular masses of 40-42 kDa. The first known 3-D structure of sub-family C.I is P61871 from *Rhizopus niveus*. The X-ray crystallography structure of P61871 reported by Kohno (1996) showed that this sequence contains a catalytic center with a triad of three amino acids (Ser-Asp-His), similar to serine proteases such as chymotrypsin and subtilisin (Blow *et al.*, 1969). It also contains the consensus sequence (Gly-X_1-Ser-X_2-Gly, with X_1 = histidine and X_2 = leucine) as shown in Figure 8.

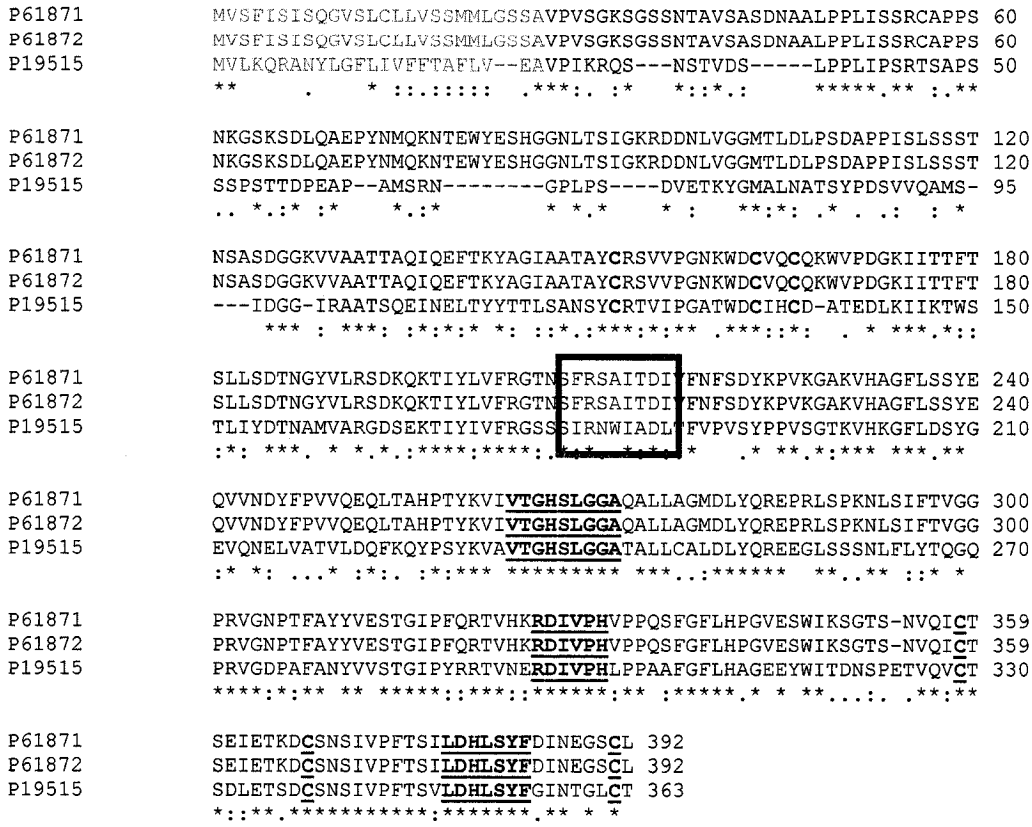


Figure 8: ClustalW alignment, the first 26 potential residue in green are the signal peptide, cystein residues correspond to the sites for disulfide bonds. The red box shows the flap region for P61871 and P19515 and the predicted flap region for P61872. The amino acid residues which belong to the catalytic triad are also shown in red (Brady et al., 1990).

Based on comparative sequence analysis and the tertiary structure of P61871, these three catalytic triad residues interact through hydrogen bonds in their side chains as described in previous studies (Kohno *et al.*, 1999). The structure around the serine active center shown in figure 8 and 9 resembles the structure of lipases known as the α/β hydrolase fold (Ollis *et al.*, 1992) with the consensus pentapeptide of β-ϵ-Ser-α motif (Derewenda *et al.*, 1994 a; Derewenda *et al.*, 1994 b), which forms a sharp turn at the active site connecting the β-sheet and an α-helix lid (Figure 9-A-B). The crystallographic structure of *Rhizomucore miehei* (RmL) triglyceride lipase was resolved by Derewenda *et al.* (1992). The RmL (P19515) molecule falls into the category of parallel α/β domains as defined by Ollis et al (1992) and is similar to P61871 sequence. It contains nine beta strands, and all strands are sequential and connected by either β-hairpin loops or right-handed turns involving the β-α-β motif (3-D structure Figure 9-B). Also, Figure 9 illustrates that the tertiary structures of P61871 and P19515 contain six α-helices. The folding of the polypeptide chain in RmL is stabilized by three disulfide bridges: Cys123-Cys362, Cys134-Cys 137 and Cys329-Cys338.

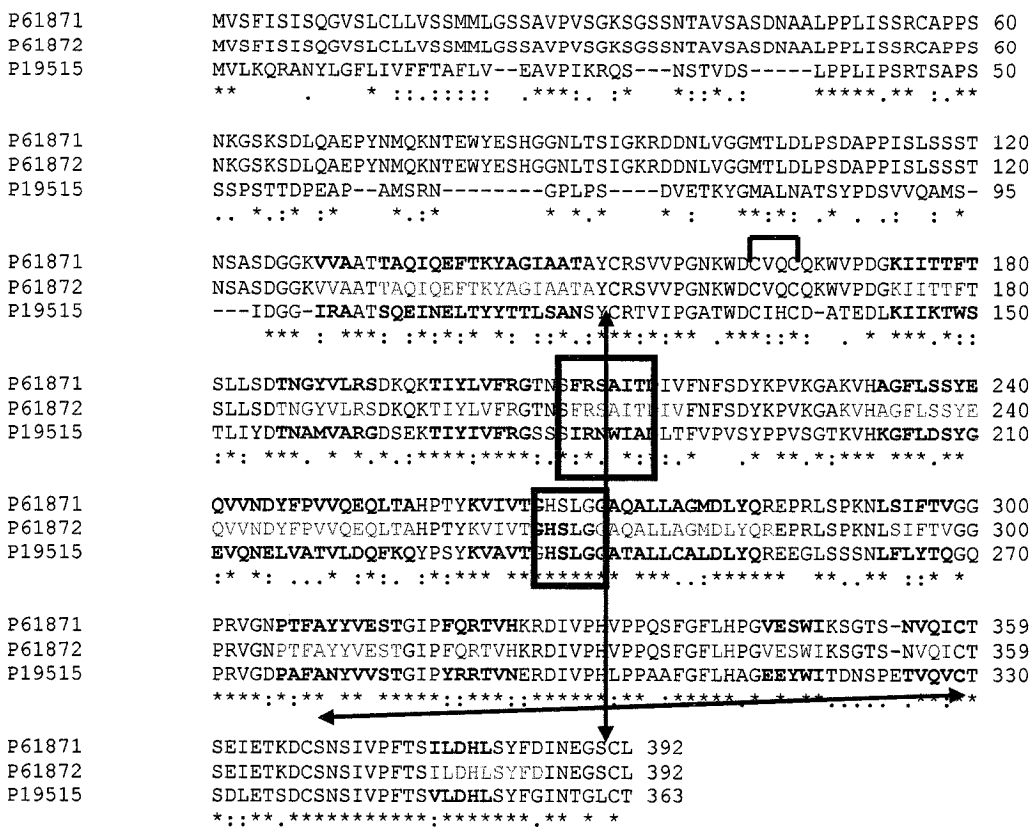


Figure 9-A: Sequence alignment of sub-family CI, the secondary structure based on the crystal structures of sequences P61871 and P19515 are available and its shown in bold "flashy green" for α helix and bold red for β sheet. For P61872 sequence secondary structure were based on PSI-PRED prediction. "Light green" indicate α-helices and orange indicate β sheets. The red box shows the flap region for P61871 and P19515 and the predicted flap region for sequence P61872. The blue arrows show the three disulfide bonds.

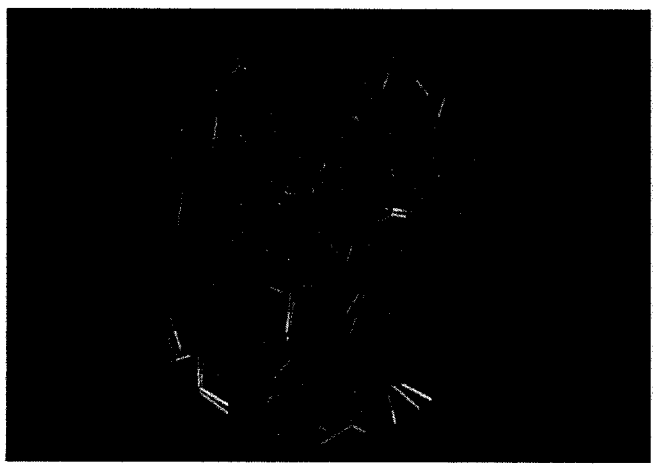


Figure 9-B: The stereo image of comparison between sequences P61871 and P19515 derived from VAST (NCBI). Red lines indicate identical sequences and blue lines indicate similarities.

Lipases of sub-family C.II, P61870 (*Penicillium camembertii)*, P61869 (*Penicillium cyclopium)*, and O59952 (*Thermomyces lanuginosus*), have the smallest molecular masses in the range of 32-33 kDa, compared to the other true lipases. In contrast to the other lipases, members of this sub-family show maximum activity at a wide range of both pH (from 5.5 to 10) and temperature. The catalytic triad of Ser, Asp and His was conserved at positions 171, 225 and 285 for all three sequences. Unlike sub-family C.I, this sub-family has only two disulfide bridges instead of three, and they are close together. As the disulfide bridges in sub-family C.II were very short in length, they may supply only local structural stability (Figure 10).

A.

```
P61870          MRLSFFTALSAVASLGYALPGKLQSRDVSTSELDQFEFWVQYAAASYYEADYTAQVGDKL 60
P61869          MRLSFFTALSAVASLGYALPGKLQSRDVSTSELDQFEFWVQYAAASYYEADYTAQVGDKL 60
O59952          MRSSLVLFF---VSAWTALASPIR-REVSQDLFNQFNLFAQYSAAAYCGKNNDAPAGTNI 56
                ** *:.  :   .*   **.. :: *:** . ::**:::.**:**:*   :  * .* ::

P61870          SCSKGNCPEVEATGATVSYDFSDSTITDTAGYIAVDHTNSAVVLAFRGSYSYRNWVADAT 120
P61869          SCSKGNCPEVEATGATVSYDFSDSTITDTAGYIAVDHTNSAVVLAFRGSYSYRNWVADAT 120
O59952          TCTGNACPEVEKADATFLYSFEDSGVGDVTGFLALDNTNKLIVLSFRGSRSIENWIGNLN 116
                :*: . ***** :.**. *.*.** : *.:*::*:*:**. :**:*** * * .**:. . .

P61870          FVHTNP-GLCDGCLAELGFWSSWKLVRDDIIKELKEVVAQNPNYELVVVGHSLGAAVATL 179
P61869          FVHTNP-GLCDGCLAELGFWSSWKLVRDDIIKELKEVVAQNPNYELVVVGHSLGAAVATL 179
O59952          FDIKEINDICSGCRGHDGFTSSWRSVADTLRQKVEDAVREHPDYRVVFTGHSLGGALATV 176
                *  .:  .:*.** .. ** ***: * * : :::::.* ::*:*.:*. .*****.*:**:

P61870          AATDLRGKGYPSAKLYAYASPRVGNAALAKYITAQ--GNNFRFTHTNDPVPKLPLLSMGY 237
P61869          AATDLRGKGYPSAKLYAYASPRVGNAALAKYITAQ--GNNFRFTHTNDPVPKLPLLSMGY 237
O59952          AGADLRGNGY-DIDVFSYGAPRVGNRAFAEFLTVQTGGTLYRITHTNDIVPRLPPREFGY 235
                *.:****:** . .::::*.:***** *:*:::*.*  *. :*:***** **:**  .:**

P61870          VHVSPEYWITSPNNATVSTSDIKVIDGDVSFDGNTGTGLPLLTDFEAHIWYFVQVDAGKG 297
P61869          VHVSPEYWITSPNNATVSTSDIKVIDGDVSFDGNTGTGLPLLTDFEAHIWYFVQVDAGKG 297
O59952          SHSSPEYWIKSGTLVPVTRNDIVKIEG---IDATGGNNQPNIPDIPAHLWYFGLIGTCL- 291
                 * ******.* . ..*: .**  *:*   :*.. *.. * :.*: **:***  :.:

P61870          PGLPFKRV 305
P61869          PGLPFKRV 305
O59952          -------- 
```

Figure 10-A: Protein sequence alignment of sub-family C.II The blue line shows the disulfide bonds and the red box is the flap region for sequence O59952 and the predicted flap regions for sequences P61870 and P61869.
The secondary structure based on the crystal structure of O59952 is available and is shown in bold "flashy green" for α helix and bold red for β sheet. For P61870 and P61869 sequences secondary structure were based on PORTER prediction. "Light green" indicate α-helices and orange indicate β sheets.

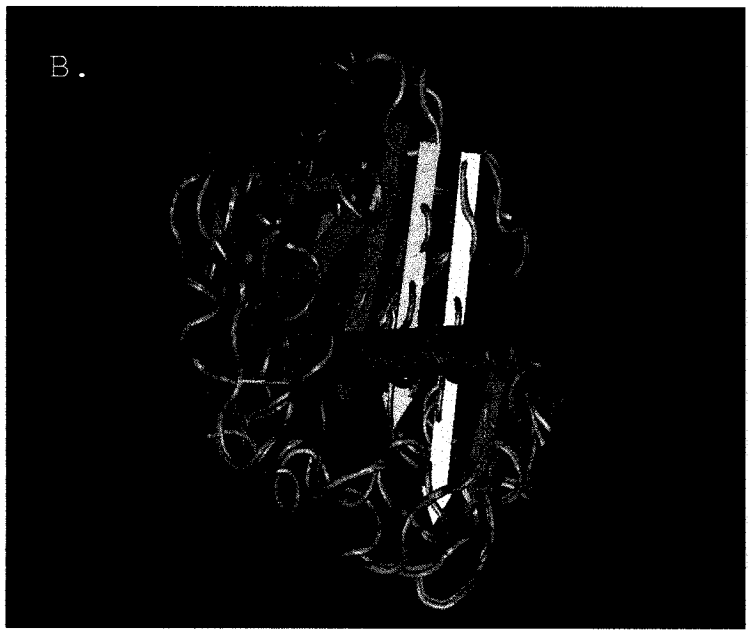


Figure 10-B: The 3D image of O59952 derived from VAST (NCBI).

D.1.4 Group D (Lysophospholipase catalytic domain)

Group D comprises 20 sequences, divided further into five sub-groups (D.I, D.II, D.III, D.IV and D.V). Since no structural information was available for these sequences, they were compared with "human cytosolic phospholipase A2", a known lysophospholipase structure found in humans-- to find conserved motifs, the potential active site and important structural features.

Sub-family D.I contains three phospholipase B (PLB1, PLB2 and PLB3) sequences, all from *Candida albicans*. They include sequences with accession numbers of Q9UWF6 (PLB1), O93795 (PLB2) and Q9UVX1 (PLB3). The first two sequences have 605 and 608 amino acids with molecular masses between 66-67 kDa. The third sequence has a longer amino acid sequence (754 amino acids) with a molecular mass of 81 kDa. Sequences Q9UWF6 and O93795 contain 17 hydrophobic amino acids and 19 Q9UVX1 amino acids as signal sequence. These signal sequences are believed to target the proteins to the endoplasmic reticulum for subsequent processing leading to the secretory pathway (Leidich *et al.*, 1998). The sequence Q9UWF6 contains seven Asn-X-Ser/Thr motifs at residues 199, 261, 399, 451, 465, 492, and 573 that could potentially be *N*-glycosylated (Figure 11). One potential tyrosine phosphorylation site, Lys-Ser- Asn-Ile Asp- Val-Ser- Ala-Tyr, is shown in Figure 11 (Leidich *et al.*, 1998).

Figure 11: Amino acid sequence alignment of lysophospholipase sub-family DI. Residues in red indicate the active site, and residues in pink indicate potential *N*-glycosylation sites. The red box shows potential tyrosine phosphorylation site with the Tyr residue shown in a black background.

Sub-family D.II includes three diverse phospholipases B, designated as PLB1, PLB2 and PLB3 from different species (Table 3). *Saccharomyces cerevisiae,* one of the species, has three phospholipases B: PLB1, PLB2, and PLB3. *Candida glabrata* the other species has two phospholipases B: PLB1 and PLB2, with only one phospholipase B (PLB) found from *Torulospora delbrueckii* and *Kluyveromyces*. The phospholipase B encoded by the *Saccharomyces cerevisiae* PLB1 gene catalyzes the deacylation of glycerophospholipids at both the *sn*-1 and *sn*-2 positions. Moreover, the PLB1 gene product exhibits phospholipase B, activity catalyzing the deacylation of phospholipase B and transacylase activity which catalyzes the transfer of an acyl chain from one phospholipase B to another and allocating the resynthesis of glycerophopholipids (Figure12 adopted from Fyrst *et al.*, 1999). Fyrst (1999) identified a gene, PLB2, which is functionally similar to PLB1, but contains no significant transacylase activity. Three highly similar genes (PLB1, PLB2 and PLB3 with >60% identity at the DNA level) that code for enzymes with diverse properties were found to exhibit three activities:

- 1, 2-acylhydrolase activity and diacyhydrolase activity on diacylphospholipids.
- Acylhydrolase activity on monoacylphospholipids.
- Acyltransferase activity on monoacylphospholipids to form the corresponding diacyphospholipids (Sreenivas *et al.*, 1998).

All three forms of the enzyme may be derived from a common protein component by using differential glycosylation (Sreenivas *et al.*, 1998). At their pH optimum (2.5-3.5), substrate preference is similar for PLB1p and PLB2p [PtdSer (phosphatidylserine)> PtdIns (phosphatidylinositol)>>PtdCho (phosphatidylcholine) >PtdEtn

(phosphatidylethanolamine)], whereas PLB3p accepts only PtdIns and PtdSer as substrates (Merkel *et al.*, 1999). Preference of PLB3p for PtdIns and PtdSer at low pH is because of the presence of histidine at the substrate-binding site of the enzyme. PtdSer and PtdIns (which is the most abundant anionic phospholipid of the plasma membrane) are tightly segregated to the internal leaflet of the plasma membrane in most cell types (Williamson *et al.*, 1994; Zwaal *et al.*, 1997). At pH values of 5 and above, the substrate preferences change to PtdCho=PtdEtn for PLB1 and PtdSer=PtdEtn for PLB2 (Merkel *et al.*, 1999).

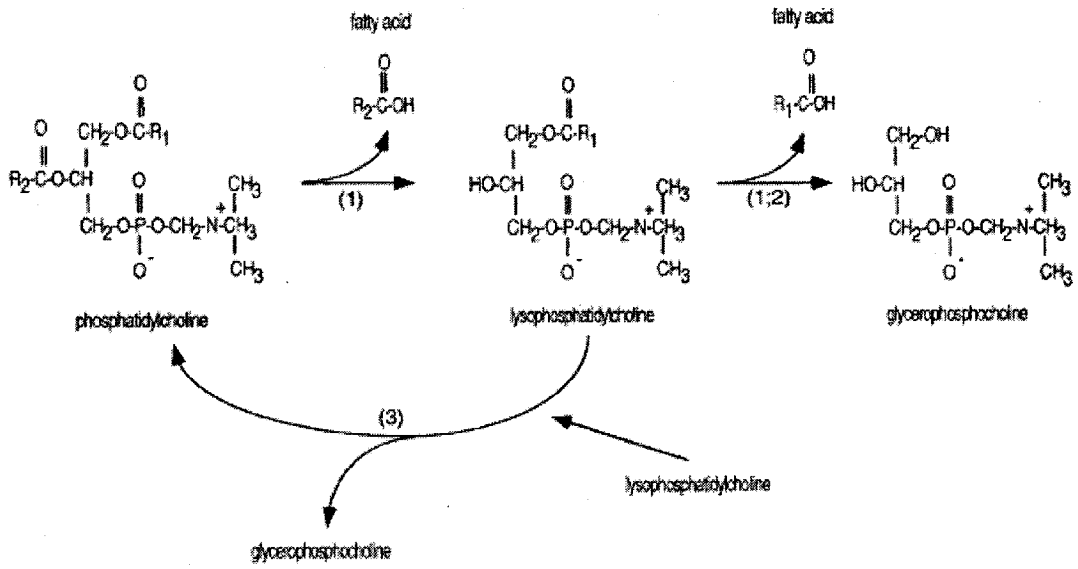


Figure 12: PLB1p activity toward phospholipid substrates containing a choline head group. Phospholipase B (1), lysophospholipase (2) and transacylase (3).

There are nine potential N-glycosidics (N-X-S/T) sites conserved in group D.II (Figure 13). Phospholipase B was reported to be glycosylated with these N-glycosidic linkages. Moreover, the lipase consensus sequence, Gly -X_1- Ser -X_2- Gly and two amino acid residues, Arg 112 and Asp 406, are comparable to catalytically essential residues of

cytosolic phospholipase A2 and are conserved in all sequences in this group (Appendix 2). Since no crystal structure of fungal phospholipase B is available, the functional residues are predicted by comparing to the phospholipases A2, which belong to lysophospholipase catalytic domain family.

```
Q11121          ---MNLKEWLLFSDAVF--FAQGTLAWSPS-NSYTPANVSCD-EDINLIR-QASGPSDNE 52
P39105          ---MKLQS-LLVSAAVLTSLTENVNAWSPN-NSYVPANVTCD-DDINLVR-EASGLSDNE 53
Q8TG07          ---MQLQD-LVVTVSLLAAFNGGVEAWSPT-NSYVPANVTCP-NDINLLR-NATGLSQSE 53
O59863          -MWFLNSVNLLFLVCSVALHLDAVNAWSPT-NGYAPGVVDCD-ENINLVR-KADAVSDDE 56
Q08108          MIRPLCSKIIISYIFAISQFLLAANAWSPT-DSYVPGTVSCP-DDINLVR-EATSISQNE 57
Q03674          ---MQLRN-ILQASSLISGLSLAADSSSTTGDGYAPSIIPCPSDDTSLVR-NASGLSTAE 55
Q8TG06          ----------MQLSVLIASVLAAGAAVDAA--SYTPQNVSCP-DNANFIRNAADGLSPAE 47
                          :     .        : ..   .*.*  : *  :: .::*  * . *  *

Q11121          TEWLKKRDVYTREALRSFLDRATSNFSDSS----LVSQLF-S-NASDIPRIAVACSGGGY 106
P39105          TEWLKKRDAYTKEALHSFLNRATSNFSDTS----LLSTLFGS-NSSNMPKIAVACSGGGY 108
Q8TG07          IDWLKKRDVNTREALESFLKRVTSNFTSNSSASNLIDQLFST-NSSNIPKIGIAASGGGY 112
O59863          ADWLKVRHESTVPALKDFLQRGFKGFTNDTS---IIDKLLAT--QDTAPKVAIACSGGGY 111
Q08108          SAWLEKRNKVTSVALKDFLTRATANFSDSSE---VLSKLFNDGNSENLPKIAVAVSGGGY 114
Q03674          TDWLKKRDAYTKEALHSFLSRATSNFSDTS----LLSTLFSS-NSSNVPKIGIACSGGGY 110
Q8TG06          KEWLKKRDPITRDALQTFLRRAFANVSTEIT-----SALFND--TENVPKLGIAVAGGGY 100
                  **: *.  *  **. ** *   ..:          . *:     .  *::.:* :****

Q11121          RAMLSGAGMLAAMDNRTDGANEHGLGGLLQSTTYLAGLSGGNWLVGTLAWNNWTSVQDIV 166
P39105          RAMLSGAGMLAAMDNRTDGANEHGLGGLLQGATYLAGLSGGNWLTSTLAWNNWTSVQAIV 168
Q8TG07          RAMLSGAGMVSAMDNRTDGANEHGLGGLLQAATYLAGLSGGNWLTTTLSWNNWTSVQDIV 172
O59863          RAMLSGAGMISAMDNRTDGANDHGLGGLLQSSTYLAGLSGGNWLVGTLAYNNWTSVQAII 171
Q08108          RSMLTGAGVLAAMDNRTEGAYEHGLGGLLQSTTYLSGASGGNWLVGTLALNNWTSVQDIL 174
Q03674          RAMLGGAGMIAAMDNRTDGANEHGLGGLLQSSTYLSGLSGGNWLTGTLAWNNWTSVQEIV 170
Q8TG06          RAMFVGAGAFAAMDNRTDGANEHGLGGLLQAATYMAGLSGGNWLTGTLAYNNFTSVQQIL 160
                *:*: *** .:******:** :********.:**::* ******. **: **:**** *:

Q11121          NNMTEDDSIWDISNSIINPGGFMIVTTIKRWDHISDAVEGKQDAGFNVSLTDIWGRALSY 226
P39105          DNTTESNSIWDISHSILTPDGINIFKTGSRWDDISDDVQDKKDAGFNISLADVWGRALAY 228
Q8TG07          DSQDNDSAIWDISHSIVSPGGINIFKTGSRWDHISDAVEDKQKAGFNVSLADVWGRALSY 232
O59863          NNMTDDNSIWDISNSIVNPGGINIFSSISRWDDISDAVEEKKKAGFNTSITDVWGRALSY 231
Q08108          NNMQNDDSIWDLSDSIVTPGGINIFKTAKRWDHISNAVESKQNADYNTSLADIWGRALAY 234
Q03674          DHMSESDSIWNITKSIVNPGGSNLTYTIERWESIVQEVQAKSDAGFNISLSDLWARALSY 230
Q8TG06          EEGDKADAIWNITNSFLNPYDKDFSKTLARWTAIGSQVQGKRDAGFNVTITDLWSRALAY 220
                :   . .:**:::.*::.* .  :  :  **  * . *: * .*.:* :::*:*.***:*

Q11121          NFFPSLYRGGVAYTWSTLRDVEVFQNGEMPFPISVADGRYPGTQIIDLNATVFEFNPFEM 286
P39105          NFWPSLHRGGVGYTWSTLREADVFKNGEMPFPITVADGRYPGTTVINLNATLFEFNPFEM 288
Q8TG07          QFFPTLYRGGVAYLWSDLRESDVFKNAEMPMPISVADGRYPGTAVIDLNSTVFEYSPFEL 292
O59863          NFFPSLDEGGVGYTWNTLRDVDVFKNGEMPFPISVAVGRYPGTQVVNLNATVFEFNPFEM 291
Q08108          NFFPSLNRGGIGLTWSSIRDFPVFQNAEMPFPISVADGRYPGTKVINLNATVFEFNPFEM 294
Q03674          NFFPSLPDAGSALTWSSLRDVDVFKNGEMPLPITVADGRYPGTTVINLNATLFEFTPFEM 290
Q8TG06          GWFPTLPNAGAGLTWSSLRDNEIFMNGEMPMPISVADGRYPGTTVINLNATVFEMTPFEI 280
                 ::*:*  .* .  *. :*:  :* *.***:**:** ****** :::**:*:** .***:

Q11121          GSWDPTLNAFTDVKYLGTKVSNGEPVNKGQCVAGYDNTGFIMGTSSSLFNQFLLQINSTS 346
P39105          GSWDPTLNAFTDVKYLGTNVTNGKPVNKGQCIAGFDNTGFITATSSTLFNQFLLRLNSTD 348
Q8TG07          GSWDPSLSAFTDVQYLGTKVSDGKPAEEGKCIAGFDNVGFLMGTSSTLFNQFLLRINDTS 352
O59863          GSWDYTLHTFTDVRYAGTNVTNGTPNVTGKCVAGFDNTGFVMGTSSSLFNQFLLQLNTTD 351
Q08108          GSWDPSLNSFANVKYLGTNVSNGVPLERGKCTAGFDNAGFIMGTSSTLFNQFLLRINSTH 354
Q03674          GSWDPSLNAFTDVKYLGTNVTNGKPVNKDQCVSGYDNAGFVIATSASLFNEFSLEASTST 350
Q8TG06          GSWDPSLNAFSDIKYLGTQVTDGKPE-TERCINGFDDASFIMGTSSSLFNEFTMSNDSAV 339
                **** :* :*:::::* **:*::* *    :*  *:*:..*: .**::***:* :  . :

Q11121          LPSFIKNLVTGFLDDLSEDEDDIAIYAPNPFKDTSYIQDNFSKSISESDYLYLVDGGEDN 406
P39105          LPSFIANLATDFLEDLSDNSDDIAIYAPNPFKEANFLQKNATSSIIESEYLFLVDGGEDN 408
Q8TG07          IPKFIRNLATHFLKDLSEDYDDIAVYAPNPFRDADYVNNNRSKSLSESEYLFLVDGGEDG 412
O59863          LPSFLYNLLHGFLTDASDDYDDISIWAPNPFYEITNIPSNYSQSISEDDTLYLVDGGEDG 411
```

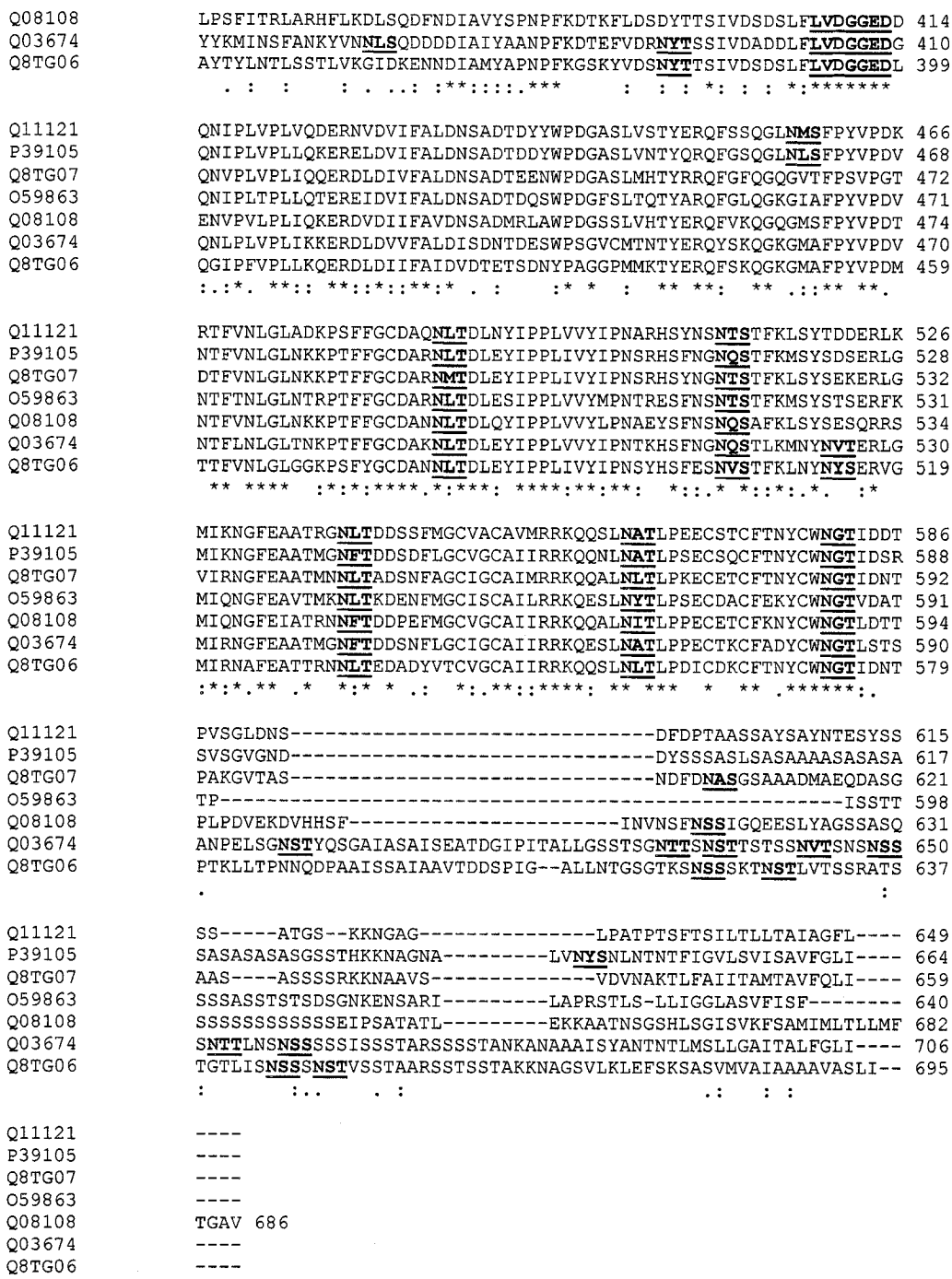


Figure 13: Multiple alignment of sub-family D.II. Residues colored in red show the putative active site residues and residues colored in pink show, potential *N*-glycosylation sites (NXT/S).

Subfamily D.III includes two sequences from *Ascomycota* phylum P39457 (*Penicillium notatum*) and O42790 (*Neurospora crassa).* The primary structure of P39457 (Masuda *et al.,* 1991) having 603 amino acids in the main chain (9 amino acids in the signal peptide) was compatible with the primary structure of O42790 (634 amino acids and 19 amino acids known as signal peptide sequence clustering with it as shown in Figure 14). For example, the *p*I value of both native and modified phospholipase B in this group (D.III) was pH 4.0 (Saito *et al.*, 1991; Kawasaki *et al.*, 1975). Moreover, sequence P39457 (*P. notatum)* contained 63 acidic amino residues in the main chain (44 Asp and 19 Glu) and 41 basic residues (17 Arg, 8 His and 16 Lys), and sequence O42790 had 62 acidic amino acid residues in the main chain (38 Asp, 24 Glu) and 49 basic acids (24 Arg, 1His and 24 Lys).

From the eight cysteine residues in both sequences, two cysteine residues were known to make a disulfide bond. *N*-glycosylation sites (Asn-X-Ser/Thr motifs) of phospholipase B from both sequences were distributed along the whole sequences at fourteen residues 41, 81, 116, 150, 223, 267, 306, 335, 427, 440, 446, 477, 498, 526, 532, 567, and 571 in the sequence from *P. notatum* (Figure 14) and were conserved in ten residues at the sequence O42790 derived from *Neurospora crassa* (Okumura *et al.*, 1981; Masuda *et al.*, 1991).

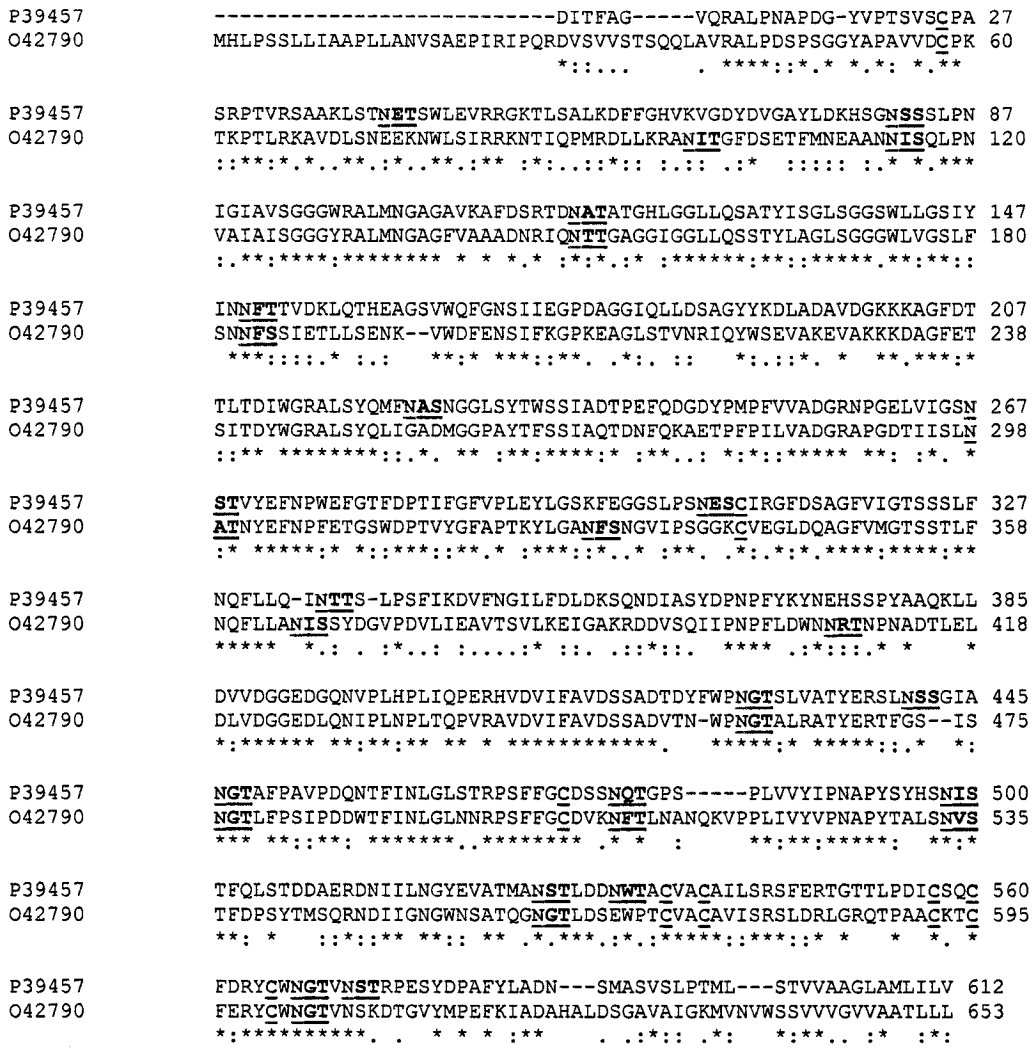


Figure 14: CLUSTALW (1.82) multiple sequence alignment. Residues colored in pink show the *N*-glycosylation sites; residues in blue indicate cysteines. It also shows eight cysteines residue conserved in both sequences that are probably responsible in forming the disulfide bonds.

The subfamily D.IV includes five sequences from the fission yeast, *Schizosaccharomyces pombe.* These five sequences contain sequences with accession numbers O13857 (missing signal peptide), Q9P327, and Q9Y7N6 (19 amino acids as signal peptide) and one lysophospholipase with accession number Q9UTH5 (19 amino acids as potential signal peptide) and sequence P78854 with 21 amino acids as signal peptides. Similar to other phospholipases, these sequences contain the conserved Arg-Ser-Asp triad that may form part of the interfacial recognition site (Figure 15). The consensus motif (Gly -X_1- Ser -X_2- Gly /Ser) similar to true lipases were also conserved in this group but in sequences Q9UTH5 and P78854 the serine residue was replaced by an aspartic acid residue (Figure 15). Moreover, at the third position of the putative catalytic site, the aspartic acid residue was conserved in all five sequences. According to Brenda (A Comprehensive Enzyme Information System; http://www.brenda.uni-koeln.de), the pH value for optimal activity of these five lipases is 2.5 and no activity was detected at neutral and alkaline pHs. These enzymes were not heat-stable and showed a temperature optimum of 35 °C.

```
O13857          ----------------------MRPG-----------------MHDTPLSLMQK-----RE 17
Q9P327          --MYVNYIGLFAFVQISLTLAYPPGRVEISEIYDFEESSSYKGQDIDTSVLYT-----LS 53
Q9Y7N6          --MKLSSFGLFLALQLLPALGLPSR------------IDEVDVSDPELIGLLKPDNVDKP 46
Q9UTH5          -----MYFQSFYFLALLLATAVYGQ-----------------VASPELHSLSR------R 32
P78854          MLFRGLSLWMLFLASCLSALALPAAE-----------------DDGSVKVFKR------- 36
                                                          .     :

O13857          ALAISLSKRDSVGSYAPYNVTCPS-DYMLRPASDG-ISSGEQSFIDKRIPKINTQMRSFI 75
Q9P327          KRKPALVKRSTDASYAPFNVTCSN-DNLLRPASEG-LNEGEQSYINKRISKVNSELRSFI 111
Q9Y7N6          ANSIPLSKRSTSPSYAPYTVACPS-GSLLRPASDG-LSTGEQEFVDKRVSKVNSALESFI 104
Q9UTH5          NWKKPPPFPSTNASYAPVIRSCDSSEIMVNSLPRGELPDLENDFIEKRLSNANEALTTFL 92
P78854          --AKKHSTKQEGPSYAPYYVDCPS-DNIVESLSSNEIPSAESEYLSTRSTIITNTAMKDFL 93
                         .   ****    * .   ::.. . . :   *..::..* .  *  :  *:

O13857          S--NTGLDVDVNSVINDSDGPRLGLAFSGGGLRAMVHGGGVLNAFDSRNGNGSSLAGILQ 133
Q9P327          S--KTGLNVDLDKVVNSSDGPRLGIAFSGGGLRAMVNGGGAFNAFDSRFESDSPLSGLLQ 169
Q9Y7N6          S--KTGLKIDTKSVLNDTDGPRLGIAISGGGFPAMLTGAGAINAFDARNGNTTSLGGILQ 162
Q9UTH5          QSKNTTADLDLSSIVGD-NGPRLGIAVSGGGWRSMLFGGGALAALDSR-SNETTLGGLLQ 150
P78854          R--NANLPGLNADTLSGSEGPSIGIALSGGGLRAMILGSGALSAMDARHDNHTVLTGLLQ 151
                   ::       . :.. :** :*:*.****  :*: *.*.: *:*:*  . : * *:**

O13857          SAMYIAGLSGGSWLVGSVAVNNFANITYLRDNVWNLEHSVFAPHGDNVVENLAYYDDLDD 193
Q9P327          SAMYISGLSGGSWLVGSVAINNFTNITYLRDNVWNLEHSVFAPHGDNVIENLNYYNDLRK 229
Q9Y7N6          SSMYLTGLSGGSWLVGSVAVNNFANITFLHDDVWNLDHSLFAPY-DDAFENFYIYQEWFE 221
```

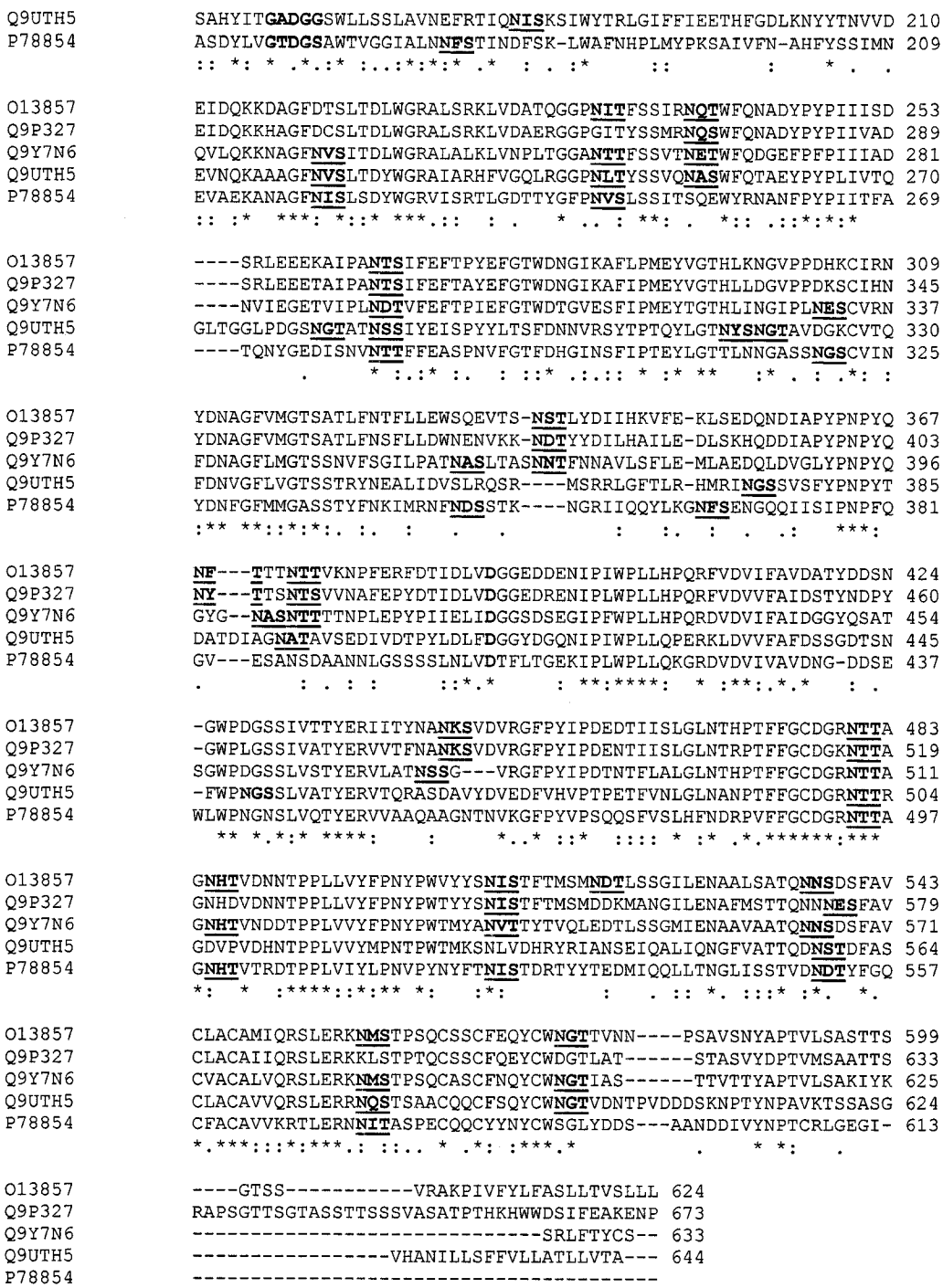


Figure 15: Multiple sequence alignment of sub-family D.IV. The serine and aspartic catalytic site is shown in red. The N-glycolysation site along with the conserved motif N-X-T/S of this carbohydrate is shown in this figure.

Subfamily D.V includes two phospholipase B enzymes from *Cryptococcus neoformans* (Q9P8P2 and Q9P8L1). The extra-cellular phospholipase B1 (PLB1) is known as a virulence factor produced by the pathogenic fungus *Cryptococcus neoformans* (Cox *et al.*, 2001). The serine presented in the pentapeptide Gly-Leu-Ser-Gly-(Gly/Ser) of both sequences in group D.V is similar to the serine presented in the "lipase motif" Gly -X_1- Ser -X_2- Gly (Schrag and Cygler, 1997). The catalytic site and the putative catalytic motif of phospholipases, SerGluGluGluXArgAla (Met/Leu), presented in these two sequences are similar to all other enzymes in family D. A group of hydrophobic amino acids at residues 1-20 is known as a signal peptide and thirteen *N*-glycosylation residues in both sequences have been predicted (Figure 16).

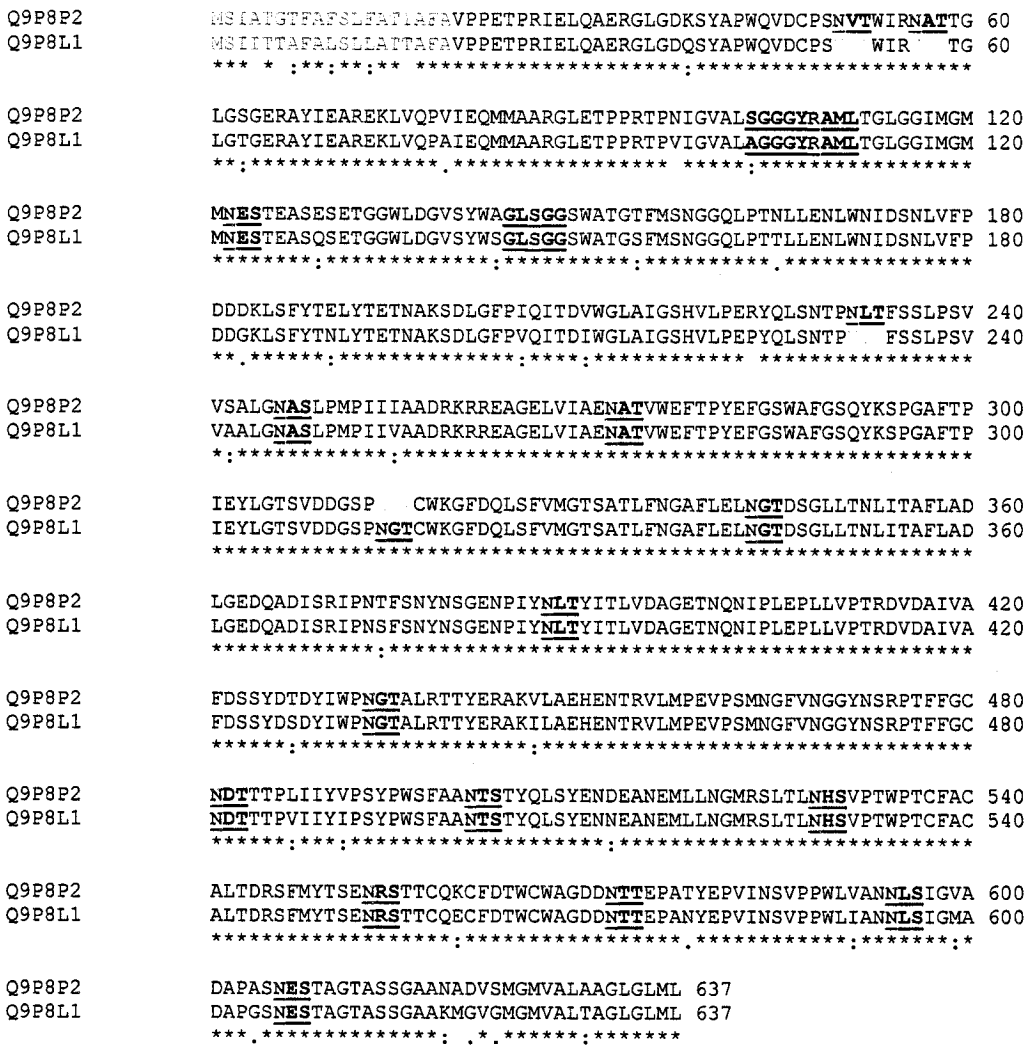


Figure 16: Multiple sequence alignment of sub-family D.V. The *N*-glycolysation sites (NXS/T) and the predicted *N*-glycosylation sites are also shown in light pink. The conserved active-site is shown in red.

D.1.5 Group E (ab-hydrolase)

Branch E which comprises only one sequence (P34163) belongs to the family "ab-hydrolase" (http://www.sanger.ac.uk/Software/Pfam). Among the fifteen ab-hydrolyase sequences in the SWISS-PROT database, only one sequence, P34163, is from fungi (*Saccharomyces cerevisiae).* This sequence has 548 amino acids with a molecular mass of 63 kDa (The sequence is not shown here). This sequence is also called TGL1 (Triglyceride lipase-cholesterol esterase). It has been suggested that TGL1 is a triglyceride-specific lipase on the basis of its homology to lipases from humans and rats, but its enzymatic activity against triacylglycerol has not been demonstrated yet (Abraham *et al.,* 1991).

D.1.6 Group F (Phospholipase C)

The three sequences P32383, O13433, and P40977, clustered in group F, are the only biochemically characterized fungal phosphoinositide phospholipase C (PLC) sequences available at SWISS-PROT. The remaining sequences are of human, rat, mouse and bovine origin. The PLC's are large enzymes with 869-1099 amino acids and molecular weights in the range of 100-124 kDa (Figure 17). Group F includes sequences with an EC number of 3.1.4.11. The digit "4" corresponds to Phosphoric diester hydrolases and "11" corresponds to Phosphoinositide phospholipase C (PI-PLCs). The two most highly conserved domains have been designated as X and Y and form the catalytic core of the molecules (Figure 17). Also, a common feature identified at the C-terminus of the Y region is a C2 domain, which is sometimes known as part of an extended Y domain.


P32383 ---------MTESAIDDQRFNLTKELQRHSCRDQG------------------------ 26
O13433 ----MLESLNRRNSIDSNQADNDNDNDNHSNDELSPSELYYSPSGSPPKSQLLLRKSSSP 56
P40977 MNCEMVTSPESVNLGDSNRAVSPFCLPDCSENAVAQTR----------SKTLDNAALD 48
P32383 --------------KITQKDDALDFISYSSFQSSFNTDQKSANNGSTVRRSIRSIFRRA 71
O13433 SSYSPIKSDLPNIYSHLRSNDSESPPQPSPKQQSSLSSSSSSSSSSNTKSSTTKNIFKKL 116
P40977 LPYVGNRKKSEQDFFKMLSSRDRDAHSTLRKRSNSLSSFLSTKSTSASENKFHGGLNWLS 108
P32383 AELPR--------------VHMGPLTYSHGIN---------------------------- 89
O13433 LRINKSSDNIDESRSIVSNNGGSPMSDSTTVTSTLSTDTAPKRGKSIQRSQILHHTDSDS 176
P40977 LKLNLLLRLQGRMNSARTNTSMNPYSCDSNEN---------------------------- 140
P32383 -------ELVNKKLRKDCDLSTLCRVLQRGIRMIRMTRRRRKFYEFKLINNNGQIIWKDG 142
O13433 LYLENQIELRPEISKSIGNIKIPSIFTNDGMPLLKISHKSKKRILFWIDPSCFKFSWRMA 236
P40977 --LSTLSSVNQNFNSRQLLATIVPESIQNGCSLLRITKKKVRQRKVSLDPISGYLMLDKN 198
P32383 --SKYLELD 149
O13433 NSTTTTTSATTSATTSGLPQGITNTTALSNSAIISTPAIATSAIHRLSITNRTTHEFVLD 296
P40977 --TGKAYKKLCVD 209
P32383 SVKDIRIGDTASTYQEEVDPKRLRSDSKLWIAIIYKVSNKLKALHVVALNELDFNTFLSC 209
O13433 DIKSIYIQNEGSGYREELNISQKLEKNWITIIYFNHKKNSLKSLHLITDNDHDFKKLISA 356
P40977 DIKEIRQGRDARNYREQYKISSENEKR--WFTIIYCADNKLKAMHMISPTLDAHNQWIMA 267
P32383 ICGLVKLRRELMESILLPD---NSQFARIHWQITVSE--------------KEEDEKKDT 252
O13433 IYNLKQLRSQLAKEFLIDLNELDENYVKMLLNKELLAGDNGNVDGNEVDIRKSHKHVREF 416
P40977 LEGLKTYR---LNEFIIGLNLVCHQGEKMIDYSENLN-----------PWEKLEKEQSAQ 313
P32383 LSFADVKKLCDKFHIYVSTGQLLEFFQLADINHNG-------------LLNYFEFEKFIK 299
O13433 LSFNDILKYSKRLNINVNTNHLQQIFDQVLLLSSATTEKPVSTPLFEKGLNFEQFKQFVS 476
P40977 LDLGDVHRMCQMLHLNASMEFLEETFQKADADHSG-------------KLSFEEFQHFVS 360
P32383 ILKNRKEVNMIWSKFTKPPHSHLSFENFFQFLITEQHEQVDRQTAWS--------YFIKY 351
O13433 ILKDRKDLQEIWDSLAQG-KEVLQFDEIKNFIINIQKENFSDDDDNSTINLIFQKYCSND 535
P40977 LLKTRSEIVDIFKEYTSG-SDKMSLEQFRHFLSTSQKARLDSDSIRT--------LYVSF 411
P32383 REPTQLTMGQDGFTKFLK-EQPYLVEVKEEL[redacted] 410
O13433 NGWNKESLNEYLLSSYST-PYREITQTQTNY[redacted] 594
P40977 CSNDDSKMGLIEFTSFLLSPHNSPVVPVI[redacted] 471
P32383 [redacted] 450
O13433 [redacted] 654
P40977 [redacted] 512
P32383 [redacted] 508
O13433 [redacted] 714
P40977 [redacted] 572
P32383 [redacted]KTSEATRGLSVNEP---FPSSFSSSYESANEQELRMKDDSTNSSSATN 565
O13433 [redacted]KTTSFQNLIETENGSFTTSTTTTTTTTTTTTTTATSLSEDNENNKSNSS 774
P40977 [redacted]CSATPLHQFSTDILKVGITDSSTDTTESSELENSELTG---------- 622
P32383 SSSMQRIKRIGLKKHADIINDVSN[redacted] 624
O13433 STSSFIIRRR-KNKSPKIIN[redacted] 833
P40977 ------LRKGKRRMKNIIVQEL[redacted] 676

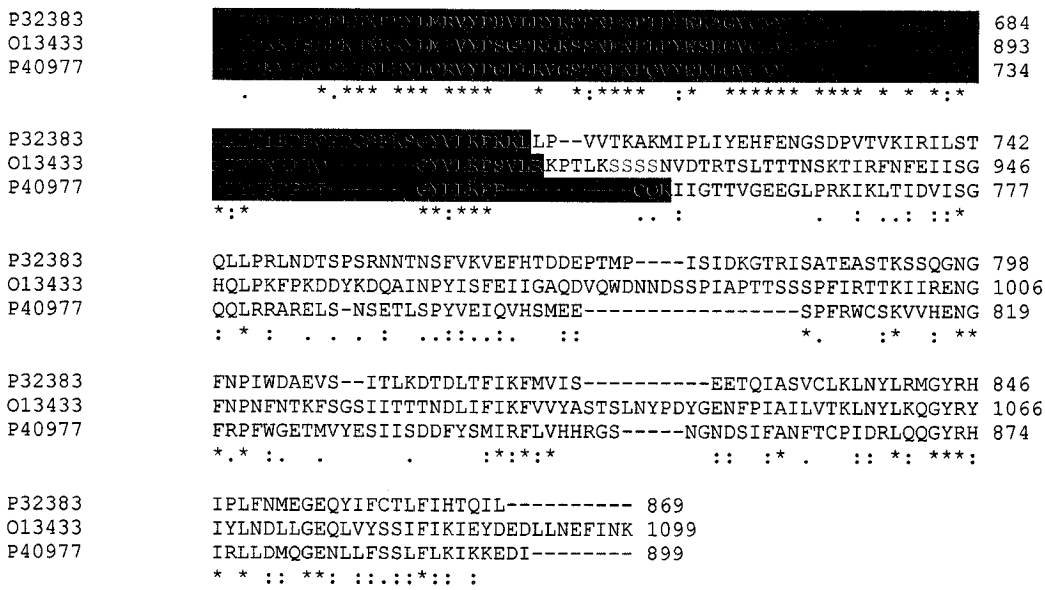


Figure 17: Multiple sequence alignment of sequences in family "F". The [redacted] and [redacted] in phosphoinositide phospholipase C (PI-PLCs) highlighted in green and red, respectively. Residues colored in red indicate the active site.

D.1.7 Group G (Phospholipase D)

The fungal lipase P36126 from *Saccharomyces cerevisiae* (Baker's yeast) is on a branch closely related to group F. This is a long protein (1380 amino acids) with a molecular mass of 160 kDa. P36126 has two domains known in all PLD1 and PLD2 enzymes, including PX and PH domains. The analysis of 3-D structure of several PH domains suggests a structure consisting of two perpendicular anti-parallel β sheets, followed by a carboxyl-terminal α-helix along one end of the barrel (Ferguson *et al.*, 1995; Ferguson *et al.*, 1994). The loops connecting the β-strands differ greatly in length, making the PH domain difficult to detect. The PX domain contains several conserved positively charged and hydrophobic residues characteristic of domains known to bind to specific phosphopeptides or phospholipids.

D.1.8 Other lipases

D.1.8.1. Triacylglycerol lipase 2 (P54857)

The lipase P54857 from *Saccharomyces cerevisiae* with an EC number of 3.1.1.3 does not show the classical consensus core of Gly -X_1- Ser -X_2- Gly but rather displays the Ala-His-Ser-Met-Gly core -- the first Gly is replaced by an alanine residue (Dartois *et al.*, 1992). A BLAST search on EXPASY/SIB shows that this sequence is more similar to the lipases from the bacterium *Pseudomonas* species than fungal lipase (Figure 18). Based on the sequence similarity with the lipase sequence from *Pseudomonas,* the catalytic site (not shown in the SWISS-PROT database) has the conserved serine and glutamic acid residue yet there is no histidine residue found in this sequence. Lipase P54857 has a molecular mass of 35 kDa with the sequence length of 326 amino acid and

shows peak activity at pH 8.0 -- similar to active pH values of group B enzymes, which cluster closer to this group.

Figure 18: Multiple sequence alignment of *Saccharomyces cerevisiae* (P54857) with two bacteria sequences from *Pseudomonas* species (P25275 and P22088).

D.1.8.2. CALB (P41365)

Candida antartica lipase B (CALB) (with a molecular mass of 36 kDa), is among the smallest lipase enzymes known in fungi. Lipase P41365 (*Candida antarctica*), like most other fungal lipases, is from the *Ascomycota* phylum. A BLAST search of this sequence shows striking similarity to lipases from the *Basidiomycota* phylum. This aspect was observable in this case and it was placed near to the sequences Q9P8L1 and Q9P8P2; both of which are sequences from *Basidiomycota* phylum. Lipase P41365 from *Candida antarctica* (3.1.1.3), is similar to all other triacylglycerol lipases at one or more of the three ester bonds of these triacylglycerols substrates (IUBMB, 1992). The amino acid sequence similarity, according to the UPGMA dendrogram, shows no significant homology to other lipase sequences. The catalytic triad, as mentioned by Uppenberg *et al.,* (1994), contains the Ser-Asp-His sequences. Conversely, the consensus sequence found in lipases around the active site serine, Gly -X_1- Ser -X_2- Gly, is not present in lipase P41365. This motif has been changed to Thr-Trp-Ser-Gln-Gly by replacing threonine instead of glycine in the first amino acid, which makes this sequence unique compared to the other lipases. Similar to group A (AB hydrolyze), in this sequence glutamic acid, as a second catalytic site, was replaced by aspartic acid. The histidine residue in this sequence is known as the third active site residue in this catalytic triad.

D.2 Evaluation of Secondary Structure Prediction Tools for fungal Lipases

D.2.1 Secondary structure and structural feature prediction family B

Using multiple sequence alignments often enables accurate prediction of secondary structures. Hence, we decided to predict secondary structures by analyzing the predictions made by different tools and techniques. To evaluate the different secondary structure prediction tools, we analyzed all of the sequences with known X-ray crystal structures using fifteen tools available on SWISS-PROT and NPS@ separately, and the output of each tool was compared with the X-ray crystallography based structure. The tool that predicted the structure that was the most similar to the crystallography based structure was chosen as the best predictor of a given group of lipases. The sequences with unknown secondary structures were separately analyzed with the chosen tool. Three sequences in sub-family B.I (P20261, P32947, and P32946) have known 3-dimensional structures based on the X-ray crystallography method. The analysis of these sequences suggested that the tool PREDATOR (Argos *et al.*, 1996) had the highest score (highest score = the most similar to the X-ray structure) for P20261 and P32947 (Figure 21) and PSI-PRED had the highest score for P32946 (Figure 21). Since PREDATOR (Argos *et al.*, 1996) was the tool that had the most similar prediction to the X-ray crystal structure, we used this tool to predict the secondary structure for the P32949 sequence (Figure 19) in sub-family B.I and compared it with the other sequences of known secondary structure (P20261 and P32947). The secondary structure of P32948 was best predicted using the PSI-PRED tool (Figure 19).

```
P32947     MKLALALSLIASVAAAPTAKLANGDTITGLNAIINEAFLGIPFAEPPVGNLRFKDPVPYS 60
P32949     MKLALALSLIASVAAAPTAILANGDTITGLNAIINEAFLGIPFAEPPVGNLRFKDPVPYR 60
P20261     MELALALSLIASVAAAPTATLANGDTITGLNAIINEAFLGIPFAEPPVGNLRFKDPVPYS 60
P32948     MKLALVLSLIVSVAAAPTAILANGDTITGLNAIINEAFLGIPFAQPPVGNLRFKPPVPYS 60
P32946     MKLCL-LALGAAVAAAPTATLANGDTITGLNAIVNEKFLGIPFAEPPVGTLRFKPPVPYS 59
           *:*.* *:* .:*******.*************:** *******:****.**** ****

P32947     GSLNGQKFTSYGPSCMQQNPEGTFEENLGKTALDLVMQSKVFQAVLPQSEDCLTINVVRP 120
P32949     GSLNGQSFTAYGPSCMQQNPEGTYEENLPKVALDLVMQSKVFQAVLPNSEDCLTINVVRP 120
P20261     GSLDGQKFTSYGPSCMQQNPEGTYEENLPKAALDLVMQSKVFEAVSPSSEDCLTINVVRP 120
P32948     ASLNGQKFTSYGPSCMQMNPLGNWDSSLPKAAINSLMQSKLFQAVLPNGEDCLTINVVRP 120
P32946     ASLNGQQFTSYGPSCMQMNPMGSFEDTLPKNARHLVLQSKIFQVVLPNDEDCLTINVIRP 119
           .**:**.**:******* ** *.::..* * * . ::***:*:.* *..********:**

P32947     PGTKAGANLPVMLWIFGGGFEIGSPTIFPPAQMVTKSVLMGKPIIHVAVNYRVASWGFLA 180
P32949     PGTKAGANLPVMLWIFGGGFEIGSPTIFPPAQMVSKSVLMGKPIIHVAVNYRLASFGFLA 180
P20261     PGTKAGANLPVMLWIFGGGFEVGGTSTFPPAQMITKSIAMGKPIIHVSVNYRVSSWGFLA 180
P32948     SGTKPGANLPVMVWIFGGGFEVGGSSLFPPAQMITASVLMGKPIIHVSMNYRVASWGFLA 180
P32946     PGTRASAGLPVMLWIFGGGFELGGSSLFPGDQMVAKSVLMGKPVIHVSMNYRVASWGFLA 179
           .**:..*.****:********:*..: **  **:: *: ****:***::***::*:****

P32947     GDDIKAEGSGNAGLKDQRLGMQWVADNIAGFGGDPSKVTIFGESAGSMSVLCHLIWNDGD 240
P32949     GPDIKAEGSSNAGLKDQRLGMQWVADNIAGFGGDPSKVTIFGESAGSMSVLCHLLWNGGD 240
P20261     GDEIKAEGSANAGLKDQRLGMQWVADNIAAFGGDPTKVTIFGESAGSMSVMCHILWNDGD 240
P32948     GPDIKAEGSGNAGLKDQRLGMQWVADNIAGFGGDPSKVTIFGESAGSMSVLCQLLWNDGD 240
P32946     GPDIQNEGSGNAGLHDQRLAMQWVADNIAGFGGDPSKVTIYGESAGSMSTFVHLVWNDGD 239
           * :*: ***.****:****.********.*****:****:*******. ::**.**

P32947     NTYKGKPLFRAGIMQSGAMVPSDPVDGTYGNEIYDLFVSSAGCGSASDKLACLRSASSDT 300
P32949     NTYKGKPLFRAGIMQSGAMVPSDPVDGTYGTQIYDTLVASTGCSSASNKLACLPLS     300
P20261     NTYKGKPLFRAGIMQSGAMVPSDAVDGIYGNEIFDLLASNAGCGSASDKLACLRGVSSDT 300
P32948     NTYNGKPLFRAAIMQSGAMVPSDPVDG             GCGSASDKLACLPSIS     300
P32946     NTYNGKPLFRAAIMQSGCMVPSDPVDGTYGTEIYNQVVASAGCGSASDKLACLPGLSQDT 299
           ***:*******.*****.*****.*** **.:*:: ..:.:**.***:******. * :

P32947     LLDATNNTPGFLAYSSLRLSYLPRPDGKNITDDMYKLVRDGKYASVPVIIGDQNDEGTIF 360
P32949         TNDTPGFLSYTSLRLSYLPRPDGANITD      RDGKYASVPVIIGDQNDEG    360
P20261     LEDATNNTPGFLAYSSLRLSYLPRPDGVNITDDMYALVREGKYANIPVIIGDQNDEGTFF 360
P32948           DTPGALAYPSLRLSFLPRPDGTFITDD     DGKCANVPVIIGDQNDE      360
P32946     LYQATSDTPGVLAYPSLRLSYLPRPDGTFITDDMYALVRDGKYAHVPVIIGDQNDEGTLF 359
           * :**.:*** *:*.*****:******  *****: ***:** * :*********** .*

P32947     GLSSLNVTTNAQARAYFKQSFIHASDAEIDTLMAAYPQDITQGSPFDTGIFNAITPQFKR 420
P32949      LSSLNTTT          FIHATD        AYPSDVTQGSPFDTGILNALTPQ    420
P20261     GTSSLNVTTDAQAREYFKQSFVHASDAEIDTLMTAYPGDITQGSPFDTGILNALTPQFKR 420
P32948        SLNVTT          HAS          SDITQGS   GIFN            420
P32946     GLSSLNVTTDAQARAYFKQSFIHASDAEIDTLMAAYTSDITQGSPFDTGIFNAITPQFKR 419
           . ****.**:*:*. *:::**:**:**:* :* :**. *:**********:**:***:**

P32947     ISAVLGDLAFIHARRYFLNHFQGGTKYSFLSKQLSGLPIMGTFHANDIVWQDYLLGSGSV 480
P32949     INAVLGDLTFTLS          YTGGPKYSFLSKQLSGLPILGTFHANDIVWQHFLLGSGSV 480
P20261     ISAVLGDLGFTLARRYFLNHYTGGTKYSFLSKQLSGLPVLGTFHSNDIVFQDYLLGSGSL 480
P32948                        QGGTKYSFLSKQLSGLPVIGTHHANDIVWQDFLVS    480
P32946     ISALLGDLAFTLARRYFLNYYQGGTKYSFLSKQLSGLPVLGTFHGNDIIWQDYLVGSGSV 479
           * *:**** *  .******:: **.**************::**.*.***::*.:*:. .*

P32947     IYNNAFIAFATDLDPNTAGLLVNWPKYTSSSQSGNNLMMINALGLYTGKDNFRTAGYDAL 540
P32949     IYN          LDPNTAGLSVQWPKSTSSSQAGDNLMQISALGLYTGKDNFR   NAL 540
P20261     IYNNAFIAFATDLDPNTAGLLVKWPEYTSSSQSGNNLMMINALGLYTGKDNFRTAGYDAL 540
P32948                  LDPNKAGLLVNWPKYTSSSQSGNNLLQINALGLYTGK             540
P32946     IYNNAFIAFAUDLDPNKAGLWTNWPTYTSSSQSGNNLMQINGLGLYTGKDNFRFDAYSAL 539
           :*********.*****.*** .:**  *****:*:**: *..***********. .*.**

P32947     MTNPSSFFV 549
P32949     FADPSHFFV 549
P20261     FSNPPSFFV 549
P32948      TNPSSFFV 549
P32946     FSNPPSFFV 548
           :::*. ***
```

Figure 19-A: ClustalW alignment of protein sequences in sub-family B.I. The "green flashy" line shows the salt bridge, the blue line shows the disulfide bonds and the red box is the flap region for sequences P20261 and P32947 and the predicted flap region for sequence P32949. The secondary structures from the crystal structure for sequences of P20261 and P32947 are available and are shown in bold "flashy green" for alpha helix and bold red for Beta sheet. For P32949 the secondary structure was predicted based on the PREDATOR (Argos *et al.*, 1996) method. The α-helices and β-sheets are shown in light-green and orange respectively.

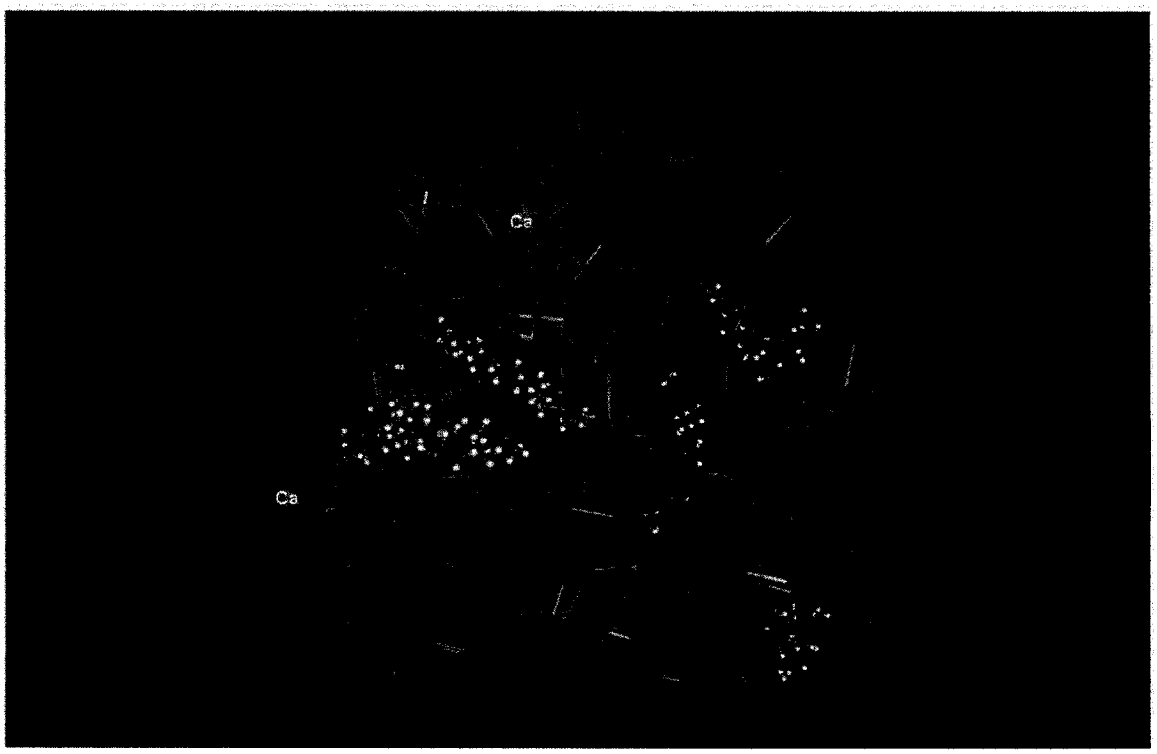


Figure 19-B: The stereo image of comparison between sequences P20261 and P32947 derived from VAST (NCBI).

A.

Figure 20-A. Multiple sequence alignment of the protein sequences in sub-family B.II. The two red boxes show the flap region for sequences P22394 and P17573 and the predicted flap region for sequence P79066. The blue lines show the disulfide bond and the pink line shows the salt bridge between Glu 119 and Arg-309

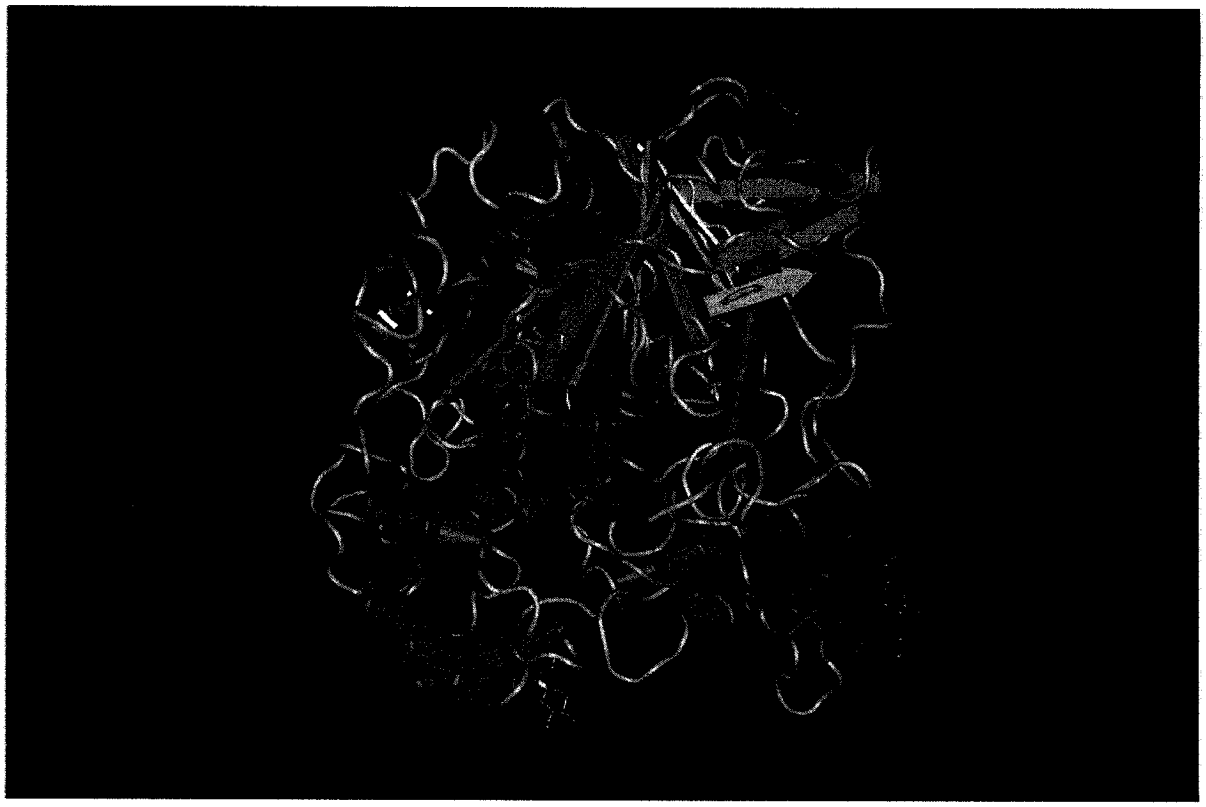
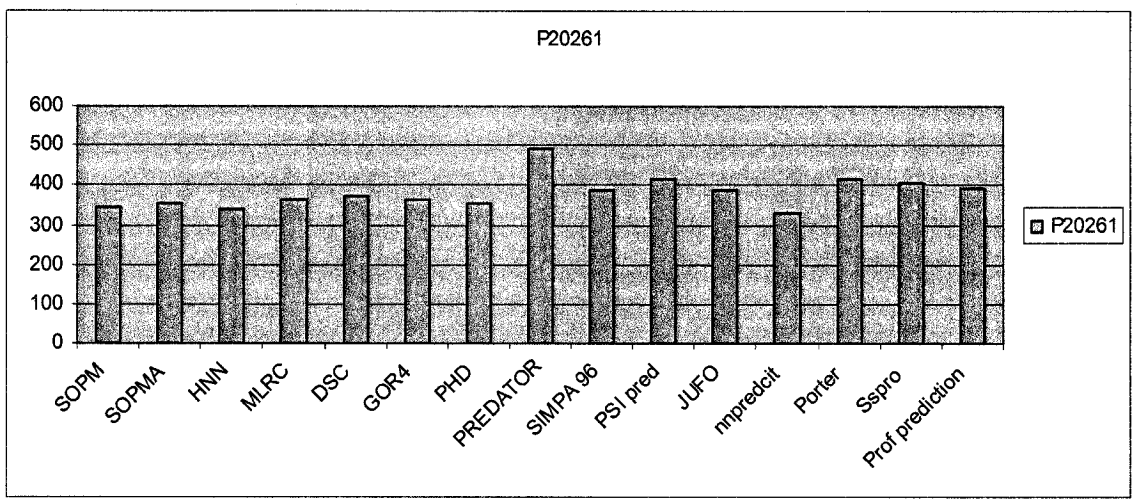
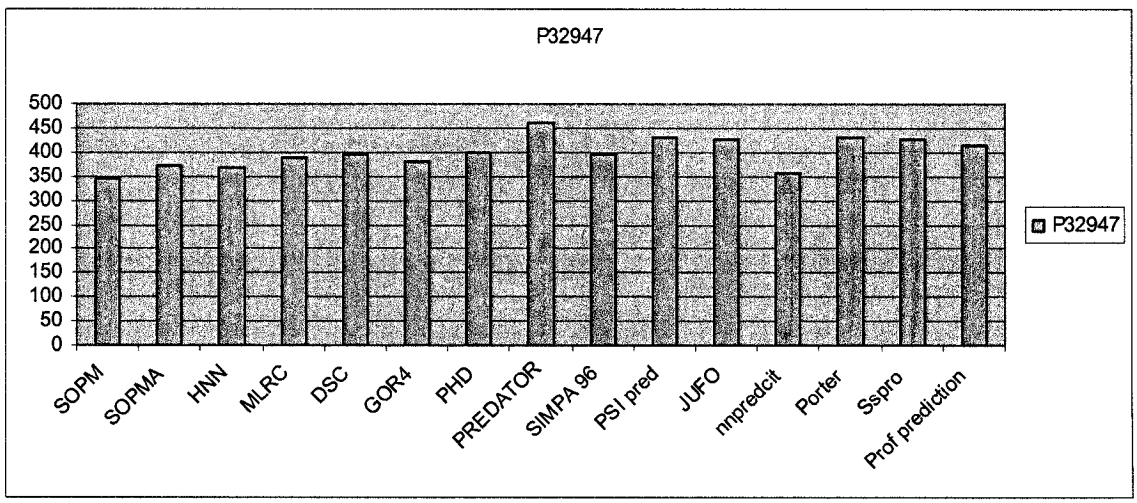


Figure 20-B: 3D image of P22394 (PDB# 1THG) derived from Cn3D 4.1 shows the α-helices and β-sheets in the tertiary structure of this sequence.


P20261
600
500
400
300
200
100
0
SOPM
SOPMA
HNN
MLRC
DSC
GOR4
PHD
PREDATOR
SIMPA 96
PSI pred
JUFO
nnpredcit
Porter
Sspro
Prof prediction
P20261


P32947
500
450
400
350
300
250
200
150
100
50
0
SOPM
SOPMA
HNN
MLRC
DSC
GOR4
PHD
PREDATOR
SIMPA 96
PSI pred
JUFO
nnpredcit
Porter
Sspro
Prof prediction
P32947

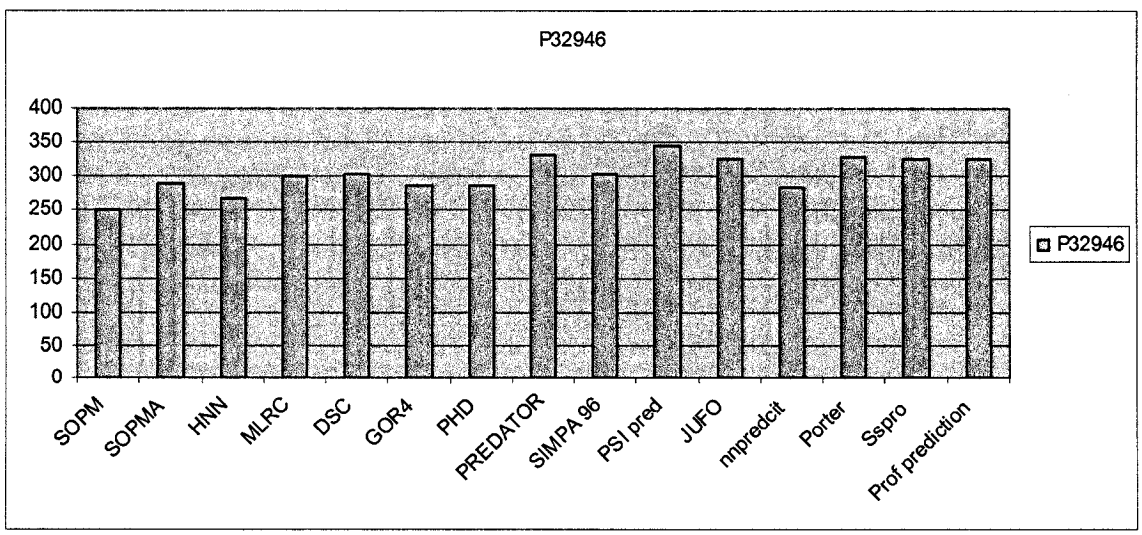

P32946
400
350
300
250
200
150
100
50
0
SOPM
SOPMA
HNN
MLRC
DSC
GOR4
PHD
PREDATOR
SIMPA 96
PSI pred
JUFO
nnpredcit
Porter
Sspro
Prof prediction
P32946

Figure 21: Comparing fifteen SS tools for sequences P20261, P32947 and P32946. Y axis indicates the protein accuracy for individual methods.

For the sub-family B.II, PORTER had the highest score and predicted the alpha helices and beta sheets closer to the structure based on X-ray diffraction studies. Thus, the sequence P79066 was analyzed with the PORTER program to predict its secondary structure (Figure 20). There is no tertiary structure available for any sequences of the sub-family B.III. Using sequence alignment, we predicted not only the secondary structures, but also some important structural features. In group B, which belongs to carboxylase-esterases, the active site includes three residues as the catalytic triad. However, for sequence Q96VC9 in sub-family B.III only two catalytic residues, serine, (at positions 200) and histidine (at positions 400) were predicted by similarity in SWISS-PROT. Moreover, according to multiple sequence alignments (Figure 7) the third residue in the catalytic triad (glutamic acid) was predicted and it perfectly matched with the other sequences in this cluster. Moreover, the catalytic residue, serine, in lipases is shielded from the solvent by one or more loops called the flap; the rearrangement of the flap opens access to the active site for the substrate (Brady *et al.*, 1990; Brozozowski and Thim 1991). The flap region and the salt bridges for sequences P32949 and P32948 in sub-family B.I, and P79066 in sub-family B.II were predicted according to sequence similarities (Figure 19-20).

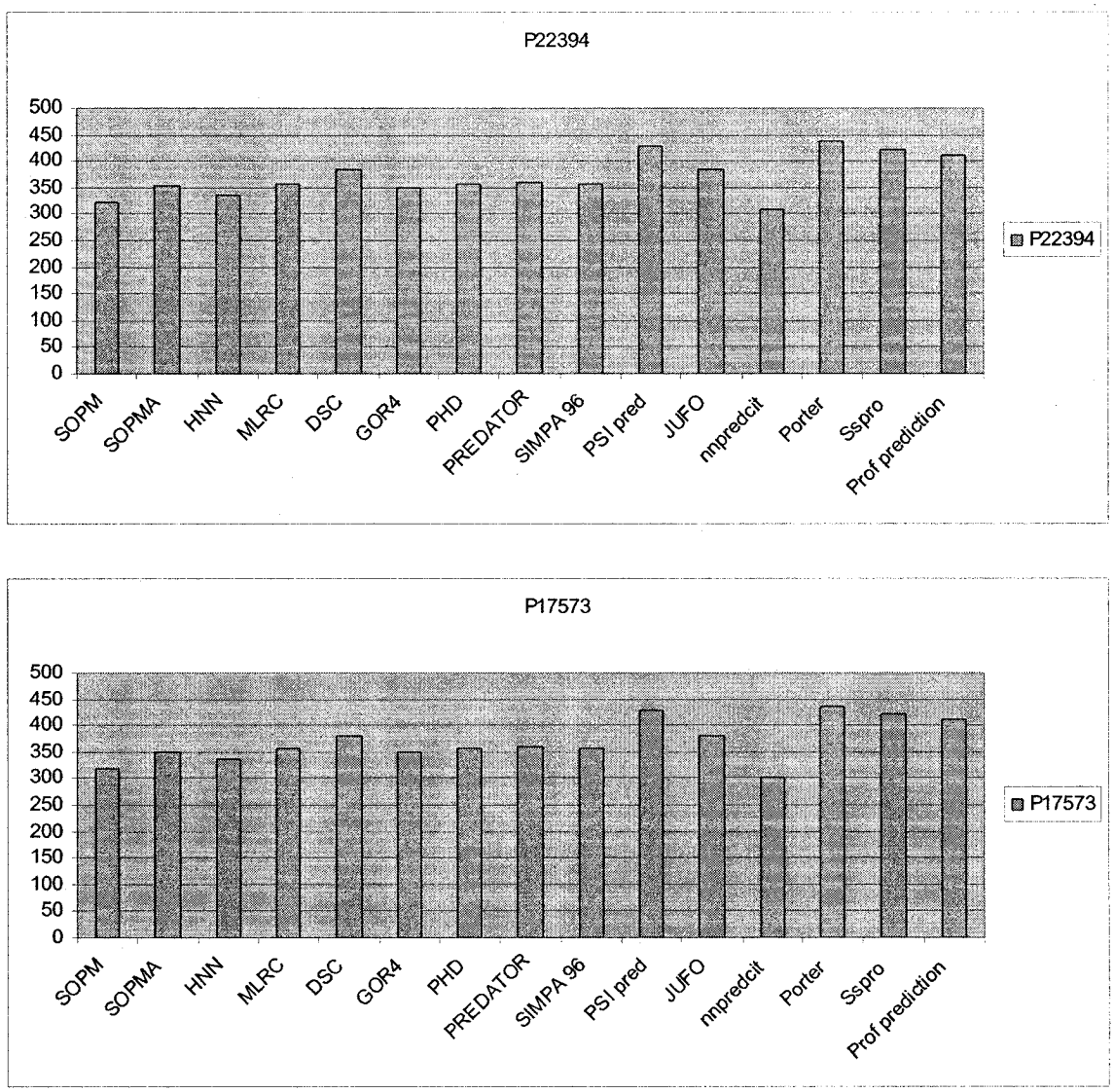


Figure 22: Comparing fifteen secondary structure prediction tools for sequences P22394 and P17573. Y axis indicates the protein accuracy for individual methods.

D.2.2 Secondary structure prediction in Subfamilies C.I and C.II:

For group C, X-ray crystallography based on structures of three sequences (P19515, P61871 and O59952) is available. These sequences were compared separately with the fifteen secondary structure prediction tools (Figure 23). The highest score for structure prediction with sequences of the sub-family C.I and C.II was obtained with SSPRO followed by a slightly lower score with PSI-PRED and PROTER. Thus, SSPRO was chosen to predict the secondary structure for sequences P61872 in sub-family C.I and P61870 and P61869 in sub-family C.II.

D.2.3 Structural feature prediction of group D

There is no known structure available for fungal phospholipase B in group D. Several lines of evidence suggested that human cytosolic phospholipase A2 (cPLA2) may be a member of the α/β hydrolase family. cPLA2 is also similar to fungal phospholipase B. To get a better understanding of the structure of fungal phospholipase B, we compared the human cytosolic phospholipase A2 (cPLA2) structure (which has a known X-ray crystal structure, 1CJY) with all of the fungal sequences in group D (Appendix 2). cPLA2 contain a dyad catalytic motif including serine (Ser-228) as the nucleophilic residue (Sharp *et al.*, 1994) and aspartic acid (Asp-549) located in a deep cleft. This serine is present in a pentapeptide sequence Gly-Leu-Ser-Gly-Ser. A different residue, Arg-200, is required for catalysis, suggesting that cPLA2 may employ a novel mechanism (Dessen *et al.*, 1999). Also the structure reveals a flexible lid that must move to allow substrate access to the active site; thus explaining the interfacial activation of this lipase. A Blast search (Altschul *et al.*, 1997) of the cPLA2 catalytic domain shows low degree of similarity with fungal phospholipase B that includes three glycines (196-

197-198) of the oxy-anion hole. Arg-200 (in sequence Q9Y7N6 was replaced by proline) and the "lipase motif", (which contains Ser-228), that was conserved in all sequences (Appendix 2). Only at Q9UTH5 and P78854 lysophopholipase sequences was the serine in "lipase motif" replaced by aspartic acid. Eight conserved cysteine residues were predicted according to sequence similarities that probably compose the disulfide bonds (Appendix 2). The eight cysteine residues were not aligned with the eight cysteine residues found in cytosolic phospholipase A2.

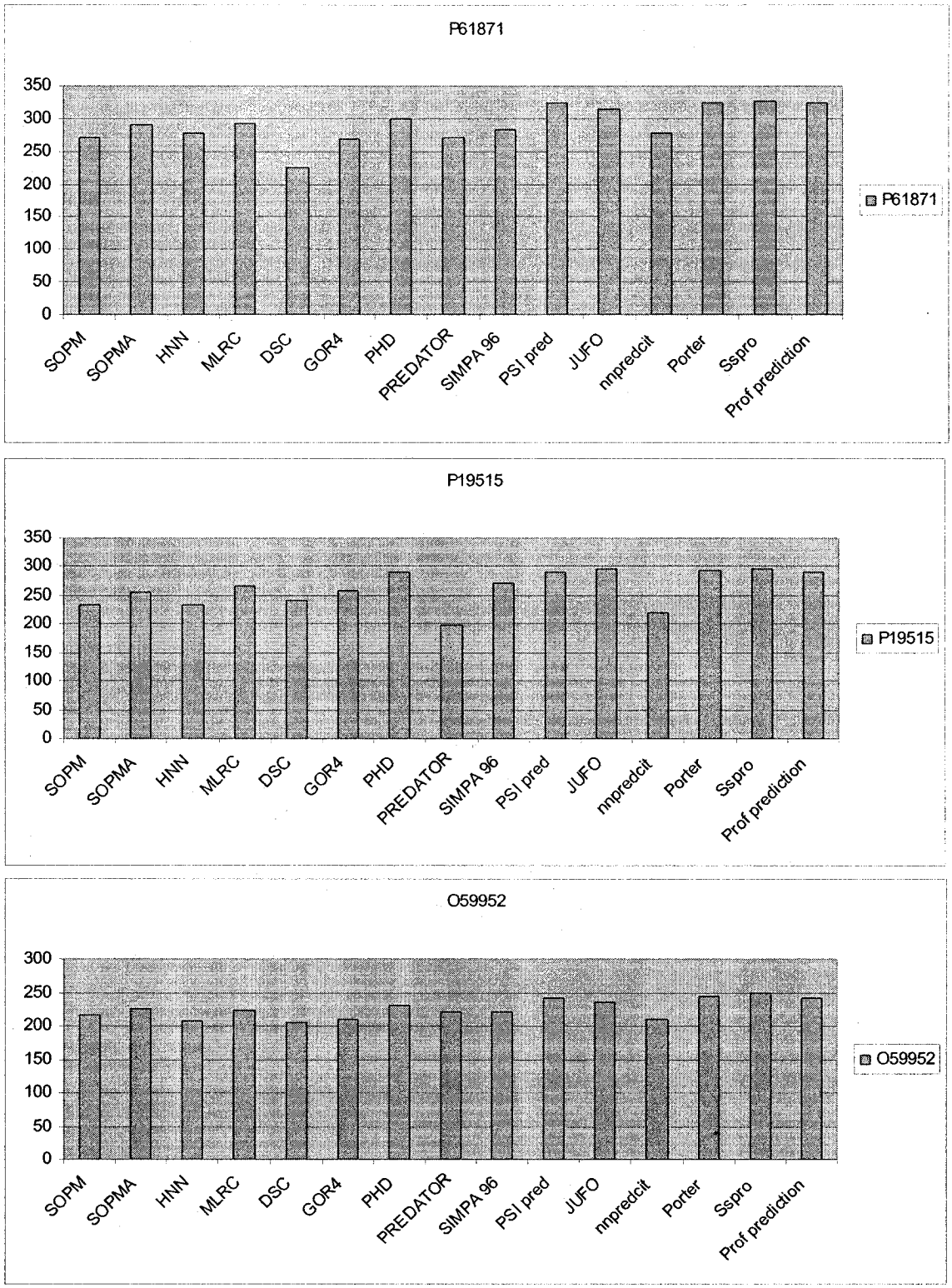


Figure 23: Comparing fifteen secondary structure prediction tools for sequences P61871, P19515 and O59952. Y axis indicates the protein accuracy for individual methods.

E. DISCUSSION

E.1 Comparative Study of Fungal Lipases

Comparative sequence analysis revealed phylogenetic relationships among the members of the lipase gene family. In this study, the 52 sequences of fungal lipases obtained from the SWISS-PROT database were categorized via sequence comparisons, crystal structure, and structure-function relationships. In agreement with the classification system derived from pfam (Protein Family Database), the present set of 52 fungal lipases can be divided into five major groups (A, B, C, D, and F), two additional groups each of which with only one member (E and G), and two sequences, P41365 and P54857, without resemblance to the rest (except in some conserved motifs) (Figure 4). This division is based on the phylogenetic tree (Figure 4) and sequence alignment shown in appendix 1.

E.1.1 Group A (Secretory lipase)

High similarity amongst sequences of group "A" (80% identical amino acid sequences) suggests that they are derived from the same ancestral gene. Amino acid sequences in group "A" contain putative *N*-glycosylation sites and also had four conserved cysteine residues may form disulfide bridges contributing to a similar three-dimensional structure. In addition, all sequences in this group, except lipase 7 (Q9P4E7) contain a putative N-terminal signal sequence. The division of group "A" into two subfamilies is based on sequence similarity. ClustalW (Thompson *et al.*, 1994) alignment shows that sequence divergence in this group happens more at the C-terminal end of the sequence and high similarities can be seen more at the N-terminus and

especially around the catalytic site (Figure 5). The fact that secreted lipases of *Candida albicans* are encoded by a family of genes may indicate that different lipase genes are needed during different stages or types of infection (Hube *et al.*, 2000). The lipase genes in group "A" constitute a large gene-family that may have evolved to adapt to the permanent association of *Candida albicans* with the human or animal host and, therefore, may also have important functions during persistence processes (Hube *et al.*, 2000).

E.1.2 Group B (Carboxylesterases)

Carboxylesterases have been divided into three categories (A, B, and C) on the basis of differential patterns of inhibition by organophosphates (Myers *et al.*, 1988; Krejci *et al.*, 1991; Cygler *et al.*, 1993). In this study group "B" (carboxylesterases) were classified into three subfamilies: BI, BII and BIII, that all belong to the *Ascomycota* taxa. Unlike the other lipases, group "B" lipases have a catalytic triad made up of the amino acids Ser-Glu-His with glutamic acid replacing the usual aspartic acid (usual in serine proteases) (Schrag *et al.*, 1991). The only exception was seen in sequence Q99156, in which the aspartic acid catalytic residue remained unchanged. The catalytic serine is shielded from the solvent by one or more loops called the "flap". The rearrangement of the flap region allows the substrate access to the active site (Brady *et al.*, 1990; Brozozowski *et al.*, 1991). The formation of Gly -X_1- Ser -X_2- Gly in the active site (Figure 7) is conserved in this group. Also, X_1 in this group is glutamic acid and X_2 is alanine. This family includes two disulfide bonds that make the structure more flexible than group "A" (Secretory lipases), which contains four disulfide bonds (Brady *et al.*, 1990).

E.1.3 Group C (AB hydrolase super family/TAG)

Group "C" is known as "true lipases", expression first used by Arpigny and Jaeger (1999) to describe classification of bacterial lipolytic enzymes. This group (group C) is divided into two subfamilies: "C.I" and "C.II". "C.I" consists of three sequences all of which are from the *Zygomycota* taxa. These lipases not only hydrolyze ester bonds of triacylglycerols, but also synthesize ester bonds in transesterification. Transesterification by lipases is particularly useful in industry, for example, in the production of cocoa butter substitutes (Matsuo *et al.*, 1980). The lipases produced by these fungi code for proteins of 363-392 amino acids with a total relative molecular mass of 40-42 kDa. The sequence differences between sequence P19515, and the rest of the *Rhizopus* lipases in sub-family "C.I" could be attributed to amino acid insertion of residues 68-76 and 80-85 in the signal peptide (prepeptide). The catalytic center of the "true lipase" is made up of three residues: Ser-Asp-His and it is responsible for the nucleophilic attack on the ester carbonyl carbon atom (Derewenda *et al.*, 1992). The oxyanion hole which helps to stabilize the growing negative charge on the peptide oxygen through hydrogen bonding also is present in these sequences (Derewenda *et al.*, 1992). This oxyanion hole is responsible for stabilizing the tetrahedral intermediate and the helical "lid". This helical "lid" is responsible for the interfacial activation upon adsorption of the enzyme to an oil-water interface.

Sub-family "C.II" includes three sequences with smaller molecular masses (30-32 kDa). Also sequences in "C.II" members, P61870 and P61869, have unique substrate specificity, in that they are strictly specific to mono and diacylglycerols but not triacylglycerols. Due to the exclusive substrate specificity of these two sequences, they

are classified as E.C. 3.1.1.- (E.C.3.-.-.-, stands for *hydrolases,* and E.C. 3.1.-. - indicating action on *ester bonds* and E.C.3.1.1. - *Carboxylic ester hydrolases*) but have different functions from triacylglycerols or other lipases with E.C. 3.1.1.3 and (Table 3 and Figure 10).

E.1.4 Group D (Lysophospholipase catalytic domain)

Group "D" encompasses 19 sequences of phospholipase B; and is further divided into five subfamilies with more than 45% sequence similarities. According to the specificity of ester linkage that will be cleaved by phospholipases, these sequences will be classified into different groupings including, A1, A2, B, C, and D (Ansell and Hawthorne, 1964; Wang and Dennis, 1999, Ghannoum, 2000). Among this group of multi-functional phospholipases, phospholipase B (PLB) enzymes are perhaps the most poorly understood-- particularly in their biological functions (Kuwabara and Shimooka 1989; Gassama-Diagne *et al.*, 1989; Satio *et al.*, 1991; Witt *et al.*, 1989). Fungi possess a class of highly homologous PLB enzymes that exhibit little sequence similarity to PLB proteins identified thus far in other eukaryotes. The physiological functions of fungal PLB enzymes are largely unknown. The phospholipase B refers to an enzyme that can remove either *Sn*-1 or *Sn*-2 fatty acids from a glycerophospholipid. However, fungi can have a single enzyme that has not only the hydrolase (fatty acid release) activities of phospholipase B, but also lysophospholipase (LPL) and transacylase activities (Cox *et al.*, 2001). The finding of a single enzyme having these multiple and seemingly paradoxical functions was seen before in *Candida albicans*, *Penicillium notatum* and *Saccharomyces cerevisiae* (Saito *et al.*, 1991; Lee *et al.*, 1994; Leidich *et al.*, 1998). Phospholipase B and phospholipase A2 belongs to a family called "lysophospholipase

catalytic domain" which has a C2 domain. The C2 domain is a Ca^{2+}-dependent membrane-targeting module found in many cellular proteins involved in signal transduction or membrane trafficking. The only known 3-D structure for "lysophospholipase catalytic domain" family member is "Human cytosolic phospholipase A2" (cPLA2) (accession code on PDB: 1cjy). In this study all phospholipase B sequences were compared to "Human cytosolic phospholipase A2" sequence to find conserved motifs. All phospholipase B sequences were aligned with "Human cytosolic phospholipase A2" and showed not only a conserved serine corresponding to Ser 228 of cPLA2, but also the consensus sequence Gly-(Leu)-Ser-(Gly)-Ser motif. Moreover, the second probable active site residue (Aspartic acid) was conserved in all sequences. Searching for the histidine as the third catalytic residue showed that none of the 19 histidine residues was enzymatically relevant in the lysophospholipase catalytic domain family (Pickard *et al.*, 1996). Also, Arg200, a different residue, is required for catalysis, suggesting that cPLA2 may employ a novel mechanism (Dessen, 2000). An interfacial activation phenomenon has been shown in both cPLA2s and lysophospholipases that preferentially cleave substrates presented at a membrane interface rather than in monomeric form (Nalefski *et al.*, 1994).

E.1.5 Group E (ab-hydrolase)

Triglyceride lipase-cholesterol esterase (P34163) is the only sequence belonging to the "AB hydrolase superfamily" and branched off from the rest of the true lipases and phospholipases (Figure 4).

E.1.6 Group F (Phospholipase C)

Group F, which includes three sequences from phosphoinositide-specific

phospholipase C enzymes (PI-PLCs) that play a critical role in receptor-linked signaling at the plasma membrane of eukaryotic cell (Singer *et al.*, 1997). In yeast, the PLC1 gene has been shown to be important for growth, but the functional reason for this behavior remains unclear (Singer *et al.*, 1997).

Hydrolysis of inositol phospholipids, particularly phosphatidyl-inositol-4, 5-bisphosphate (PIP_2) by PI-PLCs, results in the formation of secondary messenger molecules, *sn*-1, 2-diacylglycerol (DAG) and inositol 1, 4, 5-triphosphate (IP_3). These are important in regulating many cellular functions (Berridge, 1993). IP_3 causes the release of Ca^{2+} ions from internal reserves and DAG activates protein kinase C (PKC) (Singer *et al.*, 1997). Calcium mobilization and PKC activation are necessary for many cellular activities, including secretion, cell growth and proliferation (Nishizuka, 1992; Mitchell, 1992). PI-PLCs have been categorized into three subfamilies, β, γ and δ with two to four isoforms in each. Each sub-family has two conserved motifs ('X-box' and 'Y-box' important for enzyme activity) that show low similarity among subfamilies. However, they can be distinguished by their primary amino acid structure and molecular size (Rhee *et al.*, 1989). The order of these two regions is always the same (NH_2-X-Y-COOH), but the spacing is variable. There is usually a distance of 50-100 residues between these two regions and the distance in this case is about 70 residues (Singer *et al.*, 1997). The two conserved regions have been shown to be important for the catalytic activity (Singer *et al.*, 1997). At the C-terminus of the Y-box, there is a C2 domain possibly involved in Ca-dependent membrane attachment that was described in the phospholipase B family. By profile analysis, it shows that sequences with significant similarity to the X-box domain occur also in prokaryotic and trypanosome PI-specific

phospholipases C. Apart from this region, the prokaryotic enzymes show no similarity to their eukaryotic counterparts. The smallest PI-PLCs are δ isoenzymes, as they contain only sequences common to all PI-PLCs subfamilies and are frequent to all eukaryotes, from yeast and molds to plants and mammals (Singer *et al.*, 1997).

E.1.7 Group G (Phospholipase D)

P36126 from *Saccharomyces cerevisiae* in group "G" belongs to phospholipase D (PLD) and has the largest molecular mass as compared to the other lipase sequences studied here. PLD, in general, hydrolyzes phospholipids at the terminal phosphorus ester bond, leading to formation of phosphatidic acid (PA) and the free hydrophilic alcohol substituent (Figure 12). PLD1 known as phosphatidylcholine (PC), has been observed in response to a variety of agents including hormones, neurotransmitters, growth factors and phorbol esters. PA has been shown to stimulate DNA synthesis, cell proliferation, and phosphatidylinositol 4, 5-bisphosphate-phospholipase C and phosphatidylinositol phosphate kinase (Wang *et al.*, 1994).

E.2 Evaluation of Secondary Structure Prediction Tools for fungal Lipases

E.2.1 Multiple Sequence Alignment and Secondary Structure prediction for families B and C

Protein secondary structure (SS) prediction is an important stage in the prediction of protein structure and function. Accurate SS information improves the sensitivity of threading methods (Jones, 1999) and is at the core of most *ab initio* methods (Bradley *et al.*, 2003 b) for the prediction of protein structure. Multiple alignment provides much more structural information than a single sequence could (Russell and Sternberg, 1995).

Undoubtedly, the accurate alignment of various sequences is crucial for predicting secondary structures accurately. Recent studies have accurately predicted secondary structures from multiple alignments (Musacchio *et al.*, 1994 and Jenny and Benner, 1994 a). We carried out multiple sequence alignments for each sub-family separately. The results were compared with the best secondary structure prediction tools (section D.2.1). Musacchio *et al.* (1994) and Jenny and Benner (1994 b), used multiple sequence alignment to predict the secondary structure of pleckstrine homology (PH) domain protein. A year later, when the tertiary structure of this protein was determined, Russell and Sternberg (1995) used different secondary structure tools to test the accuracy of these predictions for the PH domain. They showed that multiple sequence alignment prediction was remarkably accurate and also that the PHD method (Rost B., 1996) (a secondary structure prediction tool) had the highest accuracy for this specific protein (Russell and Sternberg, 1995).

In this study we used fifteen different SS prediction tools from which we chose the best tool for structure predictions of unknown lipase sequences. To make this choice we applied our fifteen tools to known homologous sequences and decided which tools provided more accurate results. Comparing all the tools with the lipase sequences shows that the classical methods such as SOMP (Geourjon and Delage, 1994), SOMPA (Geourjon and Delage, 1995), HNN (Guermeur, 1997), MLRC (Guermeur *et al.*, 1998), DSC (King and Stenberg, 1996) and GORIV (Garnier *et al.*, 1996) do not perform well in predicting the secondary structure of these sequences. None of these methods incorporates multiple sequence alignment information. They cannot benefit from any associated improvement in accuracy. The PHD method (Rost B., 1996), on the other

hand, uses neural networks and multiple sequence alignment information, and gives improved results over the classical methods. "PROF. Prediction" (Rost, 2001) showed high accuracy for lipases via an extension of concepts implemented in PHD (Rost, 1996) in that it uses a larger data set as well as a third layer of network. The PREDATOR method (Argos *et al.*, 1996), which relies on only a single protein sequence showed high accuracy for sub-family B.I, but low predictability for other subfamilies (Figure 23). The unique feature of this approach involves database-derived statistics on residue type occurrences in different classes of beta-bridges to delineate interacting beta-strands. PSIPRED (Jones, 1999), PORTER (Pollastri *et al.*, 2002) and SSPRO (Baldi *et al.*, 2003) methods, predict protein secondary structure using the position specific scoring matrices generated by PSI-BLAST (Altschul *et al.*, 1997) and two layers of Bidirectional Recurrent Neural Networks (BRNN). This shows high accuracy for lipase sequences in group C (Baldi *et al.,* 1999). For the sub-family B.II, PORTER had the highest score and predicted the alpha helices and beta sheets closer to the structure based on X-ray diffraction studies. Thus, the sequence P79066 was analyzed with the PORTER program to predict its secondary structure (Figure 20).

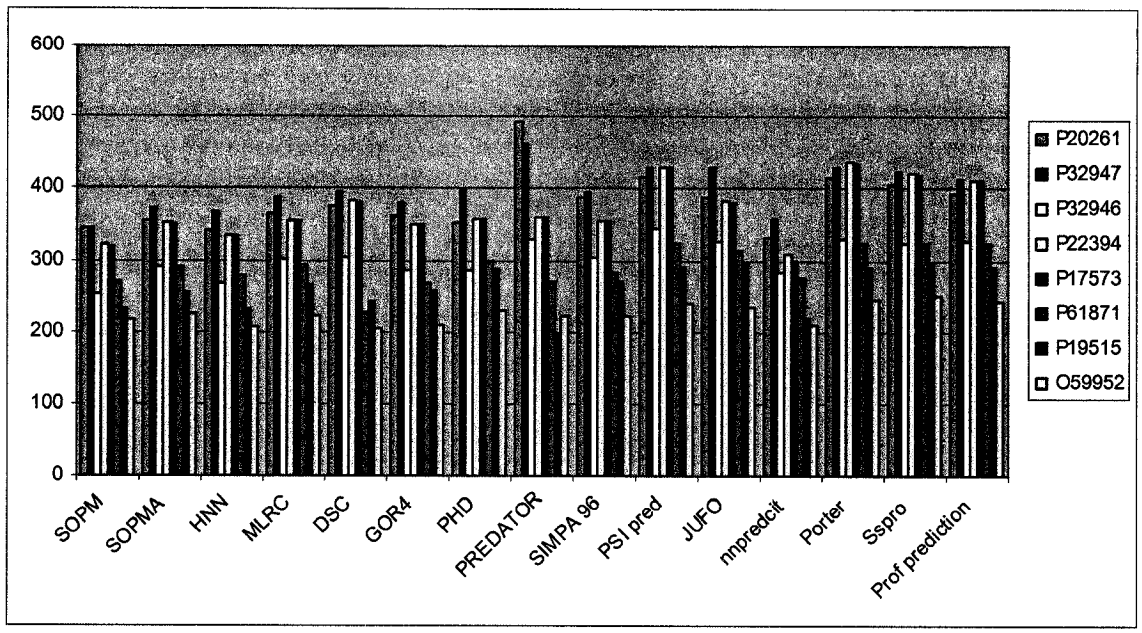


Figure 24: Comparison of fifteen secondary structure prediction tools shows the most accurate prediction for each protein.

The only noticeable errors in prediction of the B.I and B.II subfamilies (Figure 16, 17 and 19) are at the middle of the secondary structural elements, with a few motifs falsely predicted as alpha helices and beta sheets. For sub-family C.I and C.II (Figure 9 and 10), the errors are mostly at the ends of the secondary structure elements, where small errors are to be expected given the variation in accuracy of different secondary structures. Predicting the correct core secondary structure elements in the correct places is the primary point in predicting the three-dimensional fold of proteins (Russell and Sternberg, 1995).

E.3 Structural Feature Prediction

Groups B and C, comprising eight sequences with known 3-dimensional structures, were used as targets to predict the secondary and tertiary structures for unknown sequences. The secondary structures were predicted using variable structural tools as described in the previous section. Structural elements such as cysteine residues,

catalytic motifs and flap regions were predicted according to multiple sequence alignment and compared with sequences with known structural motifs. The flap region, an important feature of the majority of lipases (Schrag *et al.,* 1991), shields the catalytic site from the external environment by loops or helices (variously referred to as "loops", "flaps", or "lids"), which lie over the catalytic triad. The lids are displaced to different extents during the process of the interfacial activation, allowing the lipid substrate to enter the active site. Flap regions, described as responsible for interfacial activation in lipases were predicted for P32949, P32948, P79066, P61872, P61870 and P61869 sequences. In sub-family BI, three sequences (including P20261, P32947 and P32946) have known 3-D structure and two (P32949 and p32948) have unknown structure. Within the three known structures, the following motifs are strictly conserved:

(i) Residues of catalytic triad (Ser-His-Asp)

(ii) Residues forming salt bridges (Arg-37-Glu95 and Glu172-Arg279)

(iii) Cysteine residues involved in formation of disulfide bonds (Cys60-Cys97 and Cys268-Cys277)

(iv) Residues that form the oxy-anion hole upon interaction with the substrate (Gly124-Ala210)

High sequence similarity and high similarity in secondary structure may indicate that the flap region for sequences P32949 and P32948 is between residues 77-108 (Figure 19). Considering this α helix (77-108) region as the flap region for these two sequences, this motif should cover the serine active site. It is probably stabilized by a disulfide bond and a salt bridge similar to that seen for the other three sequences (P20261, P32947 and

P32946) with known structures. The lid structure shown in sequences with known X-ray crystal structures indicates one flap region covering the active site. However, comparisons with sequences in group B.II show that instead of one flap region covering the active site, there are two flap regions coming from two different surface loops, covering the active site in residues 85-94 and 313-326 (Figure 25). Having two flap regions coming from two different surfaces to cover the active site indicates that the active site serine residue is probably buried even deeper in this protein than equivalent serines in the other lipases. This is possibility the reason for the enzyme preference for long chain fatty acids (Schrag *et al.*, 1991).

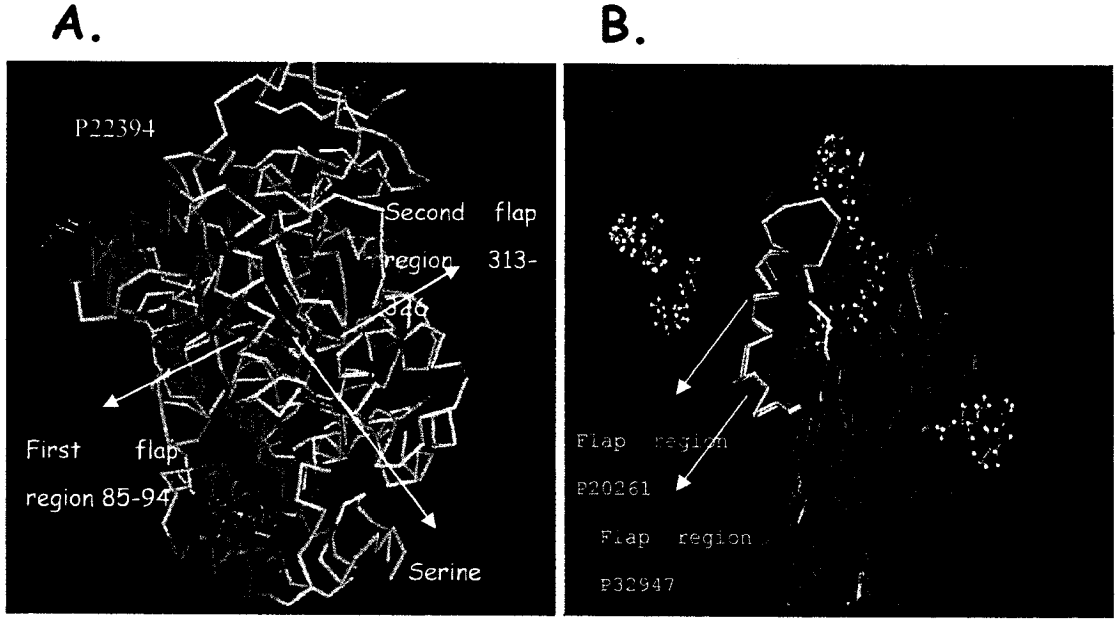


Figure 25: Comparison of Lid structure B. **A.** Two flap regions coming from two different surfaces in sequence P22394 sub-family BIII. **B.** Comparison of two similar flap regions in sequences: P20261 and P32947 (sub-family BI) using VAST derived from NCBI database.

Early reports regarding the substrate specificity of *Geotrichum candidum* lipases (GCLs) point to their preference for long chain fatty acids and for those with cis-9

unsaturated bonds (Alford and Pierce, 1961; Jensen *et al.*, 1965). Also, a study (Bertolini *et al.*,1994) for comparison against a series of triacylglycerol substrates (including: Butyrin C4, Caporin C6, Caprylin C8, Caprin C10, Laurin C12, Myristin C14, Palmitin C16, Stearin C18:0, Triolein C18:1, Linolein C18:2) showed that these enzymes display a higher affinity for Triolein (C18:1) (Figure 26) that has a longer chain.

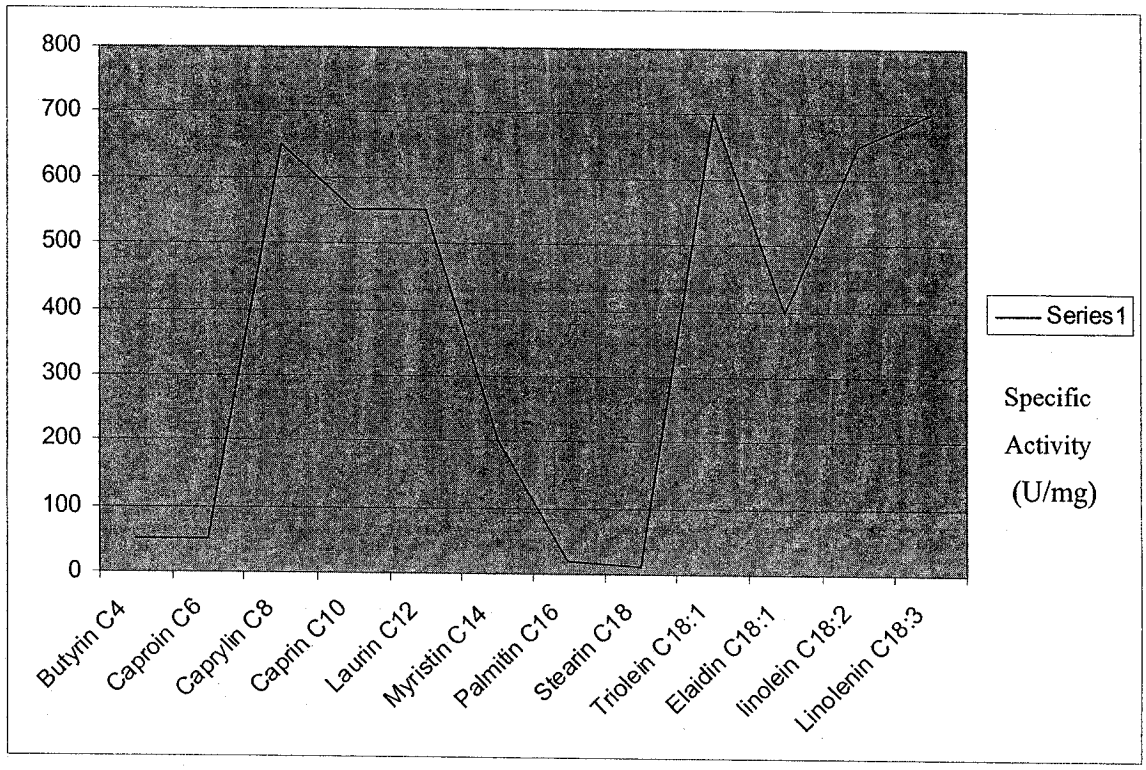


Figure 26: Comparison of different TAG substrates on *Geotrichum candidum* lipases (Adopted from Bertolini *et al.*, 1995)

Moreover, Bertolini (1995) identified the regions of the molecule that were involved in substrate differentiation. The flap region is not involved in discriminating between substrates. It is responsible for interfacial activation in lipases (Brzozowski *et al.*, 1991; van Tilbeurgh *et al.*, 1993; Grochulski *et al.*, 1993).

The flap region of sequence P19515 (*Rhizomucor miehei*) (RmL) was

experimentally proven when a complex of RmL with *n*-hexylphosphonate ethyl ester inhibitor was characterized by X-ray diffraction at 3.0 Å resolution (Brzozowski *et al.*, 1991). It includes only seven residues 179-185 (Figure 9). The same region was predicted for sequence P61871 from the experimentally crystal structure of *Rhizopus niveus* at 2.2 Å (Kohno *et al.*, 1996).

In sub-family CII, the lid structure in O59952 contains two hinge regions: N-terminal (residues 83-84) and C-terminal (residues 91-95). This sequence was aligned with sequences P61870 and P61869 and showed some similarity in their flap regions (Figure 10-A and 10-B). Furthermore, the cysteine residue in all three sequences was conserved and forms the disulfides bonds.

E.4 Structure of CALB (P41365)

As it is shown in Cn3-D view (Uppenberg *et al.*, 1994), the catalytic serine is located in the tight turn between β4 and the following α4 with similar conformation with other lipases. Uppenberg *et al.*, 1995 indicate that the crystal structures of four forms of CALB lack lid/lids. Although the overall structure and serine triad conformation are similar in all lipases, their substrates specificities and degrees of stereoselectivity differ widely (Kazlauskas *et al.*, 1991; Santaniello *et al.*, 1992).

This study agrees with these results by noting that P41365 is the smallest lipase among fungi. It is possible that they lack in some motifs, it is also shown by the UPGMA tree (Figure 4) that this specific triglyceride does not cluster with the rest of the other lipase, but rather, shows a close clustering with lipases that lack such motifs.

F. CONCLUSION

Through comparative and evolutionary analysis of protein sequences of biochemically characterized fungal lipases, I developed a comprehensive classification system for lipases of fungal origin. Using fungal lipases with known secondary and tertiary structure, and a variety of structure prediction software tools, I predicted the putative active sites, disulfide bonds and salt bridges of several lipases. The developed framework phylogenetic tree will serve as an important tool for predicting putative function and reaction conditions of newly identified lipase gene sequences. The results of this study will be invaluable for improving the properties of industrially important lipases through protein engineering.

G. REFERENCES

Abraham N., Miceli M.C., Parnes J.R., and Veillette A. (1991) Enhancement of T-cell responsiveness by the lymphocyte-specific tyrosine protein kinase $p56^{lck}$. *Nature* **350:** 62-66

Alford J. A. and Pierce D. A. (1961) Lipolytic activity of microorganisms at low and intermediate temperatures, *J. Food Sci.* **26:** 518-524.

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

Anderson E.M., Larsson K.M., Kirk O. (1998) One biocatalyst - many applications: The use of *Candida antarctica* B-lipase in organic synthesis, *Biocatal. Biotransform.* **16:** 181-204

Ansell G.B., Hawthorne J.N. (1964) Catabolism. In: Ansell GB, Hawthorne JN (eds) Phospholipids. *Elsevier*, Amsterdam, pp 152–174

Argos. (1996) User's manual. Service Argos, Inc., Landover, Md.

Arpigny, J. L., and Jaeger, K. E. (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J.* **343:** 177-183

Baldi, P. and Pollastri, G. (2003). The principled design of large-scale recursive neural network architectures-DAG-RNNs and the protein structure prediction problem. *Journal of Machine Learning Research*, **4:**575–602.

Baldi P., Brunak S., Frasconi P., Pollastri G., and Soda G. (1999) Bidirectional Dynamics for Protein Secondary Structure Prediction, Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI99), Stockholm, Sweden

Benjamin S., Pandey A. (1998) *Candida rugosa* lipases: molecular biology and versatility in biotechnology. *Yeast* **14**: 1069–1087.

Berman H.M., Bhat T.N., Bourne P.E., Feng Z., Gilliland G. (2000) The Protein Data Bank and the challenge of structural genomics, *Nature Structural Biology* 7: 957 – 959.

Berridge M. J. (1993) Inositol trisphosphate and calcium signaling, *Nature* **361**: 315 - 325

Bertolini M.C., Schrag J.D., Cygler M., Ziomek E., Thomas D.Y., Vernet T. (1995) Expression and characterization of Geotrichum candidum lipase I gene. Comparison of specificity profile with lipase II *Eur. J. Biochem.* **228**: 863-869

Blow D. M., Birktoft J. J. & Hartley B. S (1969) Role of a Buried Acid Group in the Mechanism of Action of Chymotrypsin *Nature* **221**: 337–339

Boland W., Frößl C. and Lorenz M. (1991) Esterolytic and lipolytic enzymes in organic synthesis. *Synthesis* **12:** 1049–1072

Brady L., Brzozowski A.M., Derewenda Z.S., Dodson E.J., Dodson G.G., Tolley S., Turkenburg J.P., Christiansen L., Huge-Jensen B., Norskov L., Thim L., Menge U. (1990) A serine protease triad forms the catalytic centre of a triacylglycerol lipase. *Nature* **343:** 767-770

Bradley P., Chivian D., Meiler J., Misura K., Rohl C., Schief W., Wedemeyer W., Schueler-Furman O., Murphy P., Schonbrun J., Strauss C., Baker D. (2003 b) Rosetta predictions in casp5: successes, failures, and prospects for complete automation. *Proteins*, **53**: 457–468

Bornscheuer U.T., Kazlauskas R.J., (1999) Hydrolysis in Organic Synthesis. *Regio- and Stereoselective Biotransformations,* Wiley-VCH

Brockman HL. (1984) General features of lipolysis: reaction scheme, interfacial structure and experimental approaches. In *Lipases*, pp. 3–46[B Borgström and HL Brockman, editors]. Amsterdam: Elsevier

Brzozowski A.M., and Thim L. (1991) A model for interfacial activation in lipase from the structure of a fungal lipase–inhibitor complex. *Nature* **351**: 491–494

Chen C.C., Singh J.P., and Altman R.B. (1999) Using imperfect secondary structure predictions to improve molecular structure computations. *Bioinformatics*, **15**: 53–65

Chou P. Y. and Fasman G. D. (1974) Prediction of Protein Conformation. *Biochemistry* **13**: 222.

Clausen K., Christensen M.W., Budolfsen G., Spendler T., Deussen H.-J., Svendsen A., Borch K., Patkar S.A., (2000) Lipolytic enzymes in industrial applications, *Biochemie.* **42:** 35–47

Cox G.M., McDade Henry C., Chen Sharon C. A., Tucker Stephanie C.,

Gottfredsson M., Wright L.C., Sorrell T.C., Leidich S.D., Casadevall A., Ghannoum M.A. and Perfect J.R. (2001) Extracellular phospholipase activity is a virulence factor for Cryptococcus neoformans. *Molecular Microbiology* **39**: 166-175

Cygler M., Schrag J.D., Ergan F. (1992) Advances in structural understanding of lipases. *Biotechnol Genet Eng Rev.***10**: 143-84

Cygler M., Schrag J.D., Sussman J.L., Harel M., Silman I., Gentry M.K., Doctor B.P. (1993) Relationship between sequence conservation and three-dimensional structure in a large family of esterases, lipases, and related proteins. *Protein Sci.* **2**: 366-82

Dartois V, Baulard A, Schanck K, Colson C. (1992) Cloning, nucleotide sequence and expression in Escherichia coli of a lipase gene from Bacillus subtilis 168. *Biochim Biophys Acta.* **1131**:253-60.

Derewenda, Z.S. & Derewenda, U. (1991) Relationships among serine hydrolases: evidence for a common motif in triacylglyceride Upases and esterases. *Biochem. cell Biol.* **69**, 842–851

Derewenda Z.S., Derewenda U., Dodson G.G. (1992) The crystal and molecular structure of the Rhizomucor miehei triacylglyceride lipase at 1.9-A resolution. *J. Mol. Biol.* **227**: 818-839

Derewenda U., Swenson L., Green R., Wei Y., Yamaguchi S., Joerger R. D., Haas M. J., and Derewenda Z. S. (1994 a) Current progress in crystallographic studies of new lipases from filamentous fungi. *Prot. Eng.* **7:** 551–557

Derewenda U., Swenson L., Wei Y., Green R., Kobos P. M., Joerger R. D., Haas M. J., and Derewenda Z. S. (1994 b) Conformational labiality of lipases observed in the absence of an oil-water interface: crystallographic studies of enzymes from the fungi *Humicola lanuginosa* and *Rhizopus delemar. J. Lipid Res.* **35:** 524–534

Dessen A., Tang J., Schmidt H., Stahl M., Clark J.D., Seehra J., Somers W.S. (1999) Crystal structure of human cytosolic phospholipase A2 reveals a novel topology and catalytic mechanism. *Cell* **97**:349-60

Dessen A. (2000), Crystal structure of human cytosolic phospholipase A2 reveals a novel topology and catalytic mechanism. *Biochim. Biophys. Acta,* **1488**: 40-47.

Dickerson B.P. (1976). Soil compaction after tree-length skidding in northern Mississippi. *Soil Science Society of America Journal* **40**: 965-966.

Ferguson K.M., Lemmon M.A., Schlessinger J., Sigler P.B. (1994) Crystal structure at 2.2 Å resolution of the pleckstrin homology domain from human dynamin. *Cell* **79:**199–209

Ferguson K.M., Lemmon M.A., Schlessinger J., Sigler P.B. (1995) Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell* **83:**1037–46

Frishman D., Argos P., (1996) Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* **9**:133-142

Frishman D. and Argos P., (1997) Seventy-five percent accuracy in protein secondary structure prediction. *Proteins* **27:** 329 - 335

Fukumoto, J., Iwai, M., Tsujisaka, Y. (1963) Studies on lipase I. Purification and crystallization of a lipase secreted by *Aspergillus niger. J. Gen. Appl. Microbiol.* **9:** 353-361

Fyrst H., Oskouian B., Kuypers F.A., Saba J.D. (1999), The PLB2 Gene of Saccharomyces cerevisiae Confers Resistance to Lysophosphatidylcholine and Encodes a Phospholipase B/Lysophospholipase *Biochemistry* **38**: 5864-5871

Garnier J., Gibrat J.F., Robson B. (1996) GOR secondary structure prediction method version IV Methods in Enzymology R.F. Doolittle Ed. **266:** 540-553.

Gassama-Diagne A., Fauvel J., Chap H. (1989) Purification of a new, calcium-independent, high molecular weight phospholipase A2/lysophospholipase (phospholipase B) from guinea pig intestinal brush-border membrane. *J. Biol. Chem.* **264:** 9470-9475

Geourjon C., Deleage G. (1994) SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng*; **7**:157-164

Geourjon C., Deleage G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*; **11**:681-684

Ghannoum M.A. (2000) Potential role of phospholipases in virulence and fungal pathogenesis. *Clin Micro Rev* **13**: 122–143

Grochulski P., Li Y., Schrag, J.D., Bouthillier F., Smith P., Harrison D., Rubin B., and Cygler M. (1993) Insights into interfacial activation from an open

structure of *Candida rugosa* lipase *J. Biol. Chem.* **268**: 12843-12847.

Guermeur Y., (1997) Combinaison de classifeurs statistiques, Application a' la prediction de structure secondaire des proteines. Ph.D. thesis, University of Lyon.

Guermeur Y., Geourjon C., Gallinari P., & Deleage G., (1999) Improved Performance in Protein Secondary Structure Prediction by Inhomogeneous Score Combination. *Bioinformatics* **15**: 413-421

King R. D. and Sternberg M. J. E. (1996). Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Prot. Sci.* **5**: 2298-2310.

Hahn P. F. (1943) Abolishment of alimentary lipemia following injection of heparin. *Science.* **98:** 19–20

Huang A.H.C. (1990) Enzymes of lipid degradation. In Methods in plant *Biochemistry*, **3:** 219-227 Academic press, London and New York

Hube B., Stehr F., Bossenz M., Mazur A., Kretschmar M. and Schafer W. (2000) Secreted lipases of *Candida albicans*: cloning, characterisation and expression analysis of a new gene family with at least ten members. *Arch. Microbiol.* **174:** 362-374

Jaeger K.E, Reetz M.T. (1998) Microbial lipases from versatile tools for biotechnology, *Trends Biotechnol*, **16**: 396-403

Jensen T., Sundaralingam M. (1965) Refinement of the structure of salicylic acid. *Acta Cryst. 18:* 1053-1058

Jenny T. F., Benner S. A. (1994 a) Evaluating predictions of secondary structure in proteins. *Biochem. Biophys. Res. Comm.* **200**: 149-155

Jenny T. F., Benner S. A. (1994 b) A prediction of the secondary structure of the pleckstrin homology domain. *Proteins. Struct. Funct. Genet.* **20:** 1-3

Jones D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**: 195-202.

Kawasaki M., Lee S. J. and Bersohn R., (1975) Photodissociation of molecular beams of methylene iodide and idoform. *J. Chem. Phys.* **63**: 809.

Kazlauskas R.J., Weissfloch A.N.E., Rappaport A.T., and. Cuccia L.A. (1991). A Rule To Predict Which Enantiomer of a Secondary Alcohol Reacts Faster in

Reactions Catalyzed by Cholesterol Esterase, Lipase from *Pseudomonas cepacia,* and Lipase from *Candida rugosa.* J. Org. Chem. **56:** 2656-2665

Kneller DG, Cohen FE, Langridge R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. *J Mol Biol* **214**:171-182.

Klepeis J.L. and Floudas C.A. (2003) ASTRO-FOLD: A Combinatorial and Global Optimization Framework for Ab Initio Prediction of Three-Dimensional Structures of Proteins from the Amino Acid Sequence. *Biophysical Journal,* **85**: 2119–2146.

Kohno M., Enatsu M., Yoshiizumi M., Kugimiya W. (1999) High-level expression of *Rhizopus niveus* lipase in the yeast *Saccharomyces cerevisiae* and structural properties of the expressed enzyme. *Protein Expr. Purif.* **15**: 327-335.

Kohno M., Funatsu J., Mikami B., Kugimiya W., Matsuo T., Morita Y.; (1996)."The crystal structure of lipase II from Rhizopus niveus at 2.2-A resolution." *J. Biochem.* **120:** 505-510

Krejci E., Duval N., Chatonnet A., Vincens P., Massoulie J. (1991) Aug 1Cholinesterase-like domains in enzymes and structural proteins: functional and evolutionary relationships and identification of a catalytically essential aspartic acid. *Proc Natl Acad Sci* **88**: 6647-51

Kuwabara M., Shimooka H. (1989) Grain size dependence of the critical current density in YBa2Cu3O(x) superconductors. *Applied Physics Letters* **25:** 2781-2783

Lee Y., Schulkin J, Davis M. (1994) Effect of corticosterone on the enhanced acoustic startle reflex by corticotropin-releasing factor. *Brain Res* **666:** 93–99

Leidich S.D., Ibrahim A.S., Fu Y., Koul A., Jessup C., Vitullo J., Fonzi W., Mirbod F., Nakashima S., Nozawa Y., Ghannoum M.A. (1998). Cloning and disruption of caPLB1, a phospholipase B gene involved in the pathogenicity of Candida albicans. *J. Biol. Chem.* **273**: 26078-26086

Levin JM, Robson B, Garnier J. (1996) An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett* **205:** 303-308

Lim, V. I. (1974) Algorithms for prediction of alpha-helical and beta-structural regions in globular proteins. *J. Mol. Biol.* **88**: 817–872.

Macrae A. R. (1983) Extracellular microbial lipases. *Microbial Enzymes and Technology* (W.M. Fogarty, Ed.) 225-250. Applied Science, London

Maddison D. W. Maddison (2002) MacClade version 4.03PPC: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachusetts, USA

Martinez C., Nicolas A., Tilbeurgh H., Egloff Marie-Pierre, Cudrey C., Verger R., and Cambillau C. (1994) Cutinase, a Lipolytic Enzyme with a Preformed Oxyanion Hole. *Biochemistry*, **33:** 83-89

Masuda N., Kitamura N., Saito K. (1991) Primary structure of protein moiety of Penicillium notatum phospholipase B deduced from the cDNA. *Eur. J. Biochem.* **202:** 783-787.

Matsuo Masafumi, Huang Ching-hsien and Huang Laura C. (1980) Modification and identification of glutamate residues at the arginine-recognition site in the catalytic subunit of adenosine 3′:5′-cyclic monophosphate-dependent protein kinase of rabbit skeletal muscle. *Biochem. J.* **187**: 371–379

McGuffin L. J and Jones D. T. (2003) Benchmarking secondary structure prediction for fold recognition. Proteins: *Structure, Function, and Genetics*, **52**: 166–175

Merkel O., Fido M., Mayr J. A., Pr¨uger H., Raab F., Zandonella G., Kohlwein S. D. and Paltauf F. (1999) Characterisation and function in vivo of two novel phospholipases B/lysophospholipases from Saccharomyces cerevisiae. *J. Biol. Chem.* **274:** 28121–28127

Meiler, J., (2001) Secondary structure prediction by JUFO *molecular modeling* **7:** 360

Mukherjee A.B., Miele L, Pattabiraman N. (1994) Phospholipase A2 enzymes: regulation and physiological role. *Biochem Pharmacol.* **48:** 1-10.

Musacchio A., Gibson T, Rice P, Thompson J, Saraste M. (1994) The PH domain:a common piece in the structural patchwork of signaling proteins. *Trends Biochem Sci* **18**: 343-348

Myers M., Richmond R.C., Oakeshott J.G. (1988) On the origins of esterases. *Mol Biol Evol.* Mar; **5**:113-119

Nalefski E.A., Sultzman LA, Martin DM, Kriz RW, Towler PS, Knopf JL, Clark JD. (1994). Delineation of two functionally distinct domains of cytosolic

phospholipase A2, a regulatory Ca2+-dependent lipid binding domain and a Ca2+-independent catalytic domain. *Journal of Biological Chemistry* **269**: 18239-18249.

Nicolas A., Egmond M., Verrips C.T., Longhi S., Cambillau C., Martinez C. (1996) Contribution of cutinase serine 42 side chain to the stabilization of the oxyanion transition state. *Biochemistry* **35:**398-410.

Ollis D.L., Cheah E., Cygler M., Dijkstra B., Frolow F., Franken S.M., Harel M., Remington S.J., Silman I. and Schrag J. (1992) The alpha/beta hydrolase fold, *Protein Engineering*, **5**: 197-211

Pickard G.E., Weber E. Todd, Scott Paul A., Riberdy A.F., and Rea Michael A. (1996) 5HT1B Receptor Agonists Inhibit Light-Induced Phase Shifts of Behavioral Circadian Rhythms and Expression of the Immediate-Early Gene c-fos in the Suprachiasmatic Nucleus, *J. Neurosci.* **16**: 8208-8220

Pollastri G., Przybylski D., Rost B., Baldi P. (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins*, **47**: 228–235.

Ptitsyn, O. B., and Finkelstein, A. V. (1983) Theory of protein secondary structure and algorithm of its prediction. *Biopolymers* **22**: 15–25.

Rogalska E., Cudrey C., Ferrat F., Verger R. (1993) Stereoselective hydrolysis of triglycerides by animal and microbial lipases, *CHIRALITY* **5**: 24-30

Rost B., Sander C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* **232**: 584-99

Rost B., Sander C. (1994) PHD - An automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.*, **10:** 53-60

Rost B., (1996) PHD: predicting 1D protein structure by profile based neural networks *Meth. in Enzym.* **266:** 525-539

Rost B. (1998) Marrying structure and genomics. *Structure* **6:**259-263.

Rost B. (2001) Review: protein secondary structure prediction continues to rise. *J Struct Biol* **134**:204-218.

Rubingh D.N. (1998), Catalytic behavior of lipases under laundry conditions, *Tenside Surf. Det.* **35**: 254-260

Russell R.B. and Sternberg M.J.E. (1995) Protein Structure Prediciton - How good are we? *Current Biology*, **5:** 488-490

Santaniello E., Ferraboschi P., Grisenti P., Manzocchi A. (1992) The Biocatalytic Approach to the Preparation of Enantiomerically Pure Chiral Building Blocks *Chem.Rev.* **92:** 1071–1140.

Sarda L, Desnuelle P. (1958). Action de la lipase pancr′eatique sur les esters en ′emulsion. *Biochim. Biophys. Acta* **30:** 513–21

Saito A., Yano M., Kishimoto N., Nakagahra M., Yoshimura A., Saito K.,. Kuhara S, Ukai Y., Kawase M., Nagamine T., Yoshimura S., Ideta O., Ohsawa R., Hayano Y., Iwata N.and Sugiura M., (1991) Linkage map of restriction fragment length polymorphism loci in rice. *Jpn. J. Breed.* **41**: 665-670

Schmiel D. H., and Miller V. L. (1999) Bacterial phospholipases and pathogenesis, *Microbes and Infection,* **1:** 1103-1112

Schrag J. D., Li Y., Cygler M, Lang D, Burgdorf T, Hecht H.J., Schmid R., Schomburg D., Rydel T.J., Oliver J.D., Strickland L.C, Dunaway C. M., Larson S.B, Day J. and McPherson A. (1997) The open conformation of a *Pseudomonas* lipase **5:** 187-202

Schrag J.D., Li Y., Wu S., Cygler M. (1991) Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum. Nature* **351**:761-764

Schrag J.D., Cygler M. (1993) 1.8 A refined structure of the lipase from *Geotrichum candidum. J. Mol. Biol.* **230**:575-591

Schrag J. D. and Cygler, M. (1997) Lipases and alpha/beta hydrolase fold. *Methods Enzymol.* **284:** 85-107

Schmid R.D. and Verger, R. (1998) Lipases: Interfacial enzymes with attractive applications. *Angew. Chem. Int. Ed. Eng.* **37:** 1609–1633

Sharma R., Chistib Y., Chand B.U.. (2001) Production, purification, characterization, and applications of lipases. *Biotechnology Advances* **19**: 627–662.

Sreenivas, A., Patton, V. J., Bruno, V., Griac, P. and Henry, S. A. (1998) A Role for Phospholipase D (Pld1p) in Growth, Secretion, and Regulation of Membrane Lipid Synthesis in Yeast. *J. Biol. Chem.* **273**: 16635–16638

Sussman J.L., Harel, M., Frolow F., Oefner C., Goldman A., Toker, L. & Silman I. (1991). Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. *Science* **253:** 872–879.

Svendsen A., Borch K., Barfoed M., Nielsen T.B., Gormsen E., Patkar S.A. (1995) Biochemical properties of cloned lipases from the Pseudomonas family. *Biochim Biophys Acta.* **1259:** 9-17

Svendsen A. (2000) Lipase protein engineering. *Biochimica et Biophysica Acta* **1543**: 223-238

Swofford D.L., Sunderland, M.A. Sinauer (2001) PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version 4 Sinauer, Sunderland, Mass

Terzyan S., Wang C.S., Downs D., Hunter B., Zhang X.C., (2000) Crystal structure of the catalytic domain of human bile salt activated lipase. *Protein Sci.* **9:** 1783-90

Thompson J.D., Higgins D.G., Gibson T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-80.

Tilbeurgh h., Egloff M.P., Martinez C., Rugani N., Verger R. and Cambillau C. (1993) Interfacial activation of the lipase–procolipase complex by mixed micelles revealed by X-ray crystallography *Nature* **362**: 814 – 820

Timm D, Salim K, Gout I, Guruprasad L, Waterfield M, Blundell T. (1994) Crystal structure of the pleckstrin homology domain from dynamin. *Nat.Struct. Biol.* **1**:782–88

Tojo H., Ichida T. and Okamoto M. (1997) Purification and Charaterization of a catalytic domain of Rat intestinal Phospholipase B/Lipase associated with brush border membranes. *The journal of Biological Chemistry.* **273:** 2214-2221

Tsujisaka Y., Iwai M., Tominaga Y. (1973) Purification, crystallization and some properties of lipase from *Geotrichum candidum Agr. Biol. Chem.* **37**: 1457-1464

Uppenberg J., Hansen M.T., Patkar S. and Jones T.A. (1994) The sequence, crystal structure determination and refinement of two crystal forms of lipase B from *Candida antarctica. Structure* **2:** 293–308

Uppenberg J., Ohrner N., Norin M., Hult K., Kleywegt G.J., Patkar S., Waagen V., Anthonsen T., Jones T.A. (1995) Crystallographic and molecular-modeling studies of lipase B from *Candida antarctica* reveal a stereospecificity pocket for secondary alcohols. *Biochemistry* **34**: 16838-

Verger, R. (1984) Pancreatic lipase. In: *Lipases* (Borgstrom, B. & Brockman, H. L., eds.), 1st ed., 84-150. Elsevier, Amsterdam, The Netherlands.

Verger R., and G. H. de Haas. (1976) Interfacial enzyme kinetics of lipolysis. *Annu. Rev. Biophys. Bioeng.* **5:** 77-117

Wang, Timothy C. Cardiff Robert D., Zukerberg Lawrence, Lees Emma, Arnold Andrew and Schmidt Emmett V. (1994) Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice *Nature* **369**: 669–671

Wang A. and Dennis E. A. (1999) Mammalian lysophospholipases *Biochim. Biophys. Acta* **1439:** 1–16

Wei Y., Schottel J. L., Derewenda U., Swenson L., P. Shamkant & Derewenda Z.S. (1995) A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nature Structural Biology* **2**: 218 - 223

Witt, W., Hampel, P., B¨ocker, K. and Mertsching, A. (1989) Effects of nucleotides and divalent cations on phospholipase activity in Saccharomyces cerevisiae. *Arch. Microbiol.* **151**, 154–158

Zwaal RFA and Schroit AJ, (1997) Pathophysiological implications of membrane phospholipid asymmetry in blood cells, *Blood* **89:** 1121–1132.

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.

Appendix 1: Multiple sequence alignment of 52 fungal lipase sequences using ClustalW program.

```
52   1392
Q9P4E5     ---------- ---------- ---------- ---------- ----------
Q9P8W5     ---------- ---------- ---------- ---------- ----------
Q9P4E8     ---------- ---------- ---------- ---------- ----------
O94091     ---------- ---------- ---------- ---------- ----------
Q9P8W2     ---------- ---------- ---------- ---------- ----------
Q9P8W0     ---------- ---------- ---------- ---------- ----------
Q9P8V9     ---------- ---------- ---------- ---------- ----------
Q9P8W1     ---------- ---------- ---------- ---------- ----------
Q9P4E6     ---------- ---------- ---------- ---------- ----------
Q9P4E7     ---------- ---------- ---------- ---------- ----------
P20261     ---------- ---------- ---------- ---------- ----------
P32947     ---------- ---------- ---------- ---------- ----------
P32949     ---------- ---------- ---------- ---------- ----------
P32946     ---------- ---------- ---------- ---------- ----------
P32948     ---------- ---------- ---------- ---------- ----------
P22394     ---------- ---------- ---------- ---------- ----------
P79066     ---------- ---------- ---------- ---------- ----------
P17573     ---------- ---------- ---------- ---------- ----------
P61871     ---------- ---------- ---------- ---------- ----------
P61872     ---------- ---------- ---------- ---------- ----------
P19515     ---------- ---------- ---------- ---------- ----------
P61870     ---------- ---------- ---------- ---------- ----------
P61869     ---------- ---------- ---------- ---------- ----------
O59952     ---------- ---------- ---------- ---------- ----------
Q9UWF6     ---------- ---------- ---------- ---------- ----------
O93795     ---------- ---------- ---------- ---------- ----------
Q9UVX1     ---------- ---------- ---------- ---------- ----------
Q11121     ---------- ---------- ---------- ---------- ----------
P39105     ---------- ---------- ---------- ---------- ----------
Q8TG07     ---------- ---------- ---------- ---------- ----------
Q08108     ---------- ---------- ---------- ---------- ----------
O59863     ---------- ---------- ---------- ---------- ----------
Q8TG06     ---------- ---------- ---------- ---------- ----------
Q03674     ---------- ---------- ---------- ---------- ----------
P39457     ---------- ---------- ---------- ---------- ----------
O42790     ---------- ---------- ---------- ---------- ----------
O13857     ---------- ---------- ---------- ---------- ----------
Q9P327     ---------- ---------- ---------- ---------- ----------
Q9Y7N6     ---------- ---------- ---------- ---------- ----------
Q9UTH5     ---------- ---------- ---------- ---------- ----------
P78854     ---------- ---------- ---------- ---------- ----------
Q9P8P2     ---------- ---------- ---------- ---------- ----------
Q9P8L1     ---------- ---------- ---------- ---------- ----------
P40977     MNCEMVTSPE SVNLGDSNRA VSPFCLPDCS ENAVAQTRSK TLDNAALDLP
P32383     ---------- ---------- --------MT ESAIDDQR-- ----------
O13433     ----MLESLN RRNSIDSNQA DNDNDNDNHS NDELSPSELY YSPSGSPPKS
Q99156     ---------- ---------- ---------- ---------- ----------
Q96VC9     ---------- ---------- ---------- ---------- ----------
P36126     -MYSRNENSL FRIHLEYGID EDRLKWSIIR SYKDIKSLHH KLKIVAFQQL
P34163     ---------- ---------- ---------- ---------- ----------
P54857     ---------- ---------- ---------- ---------- ----------
P41365     ---------- ---------- ---------- ---------- ----------

           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
           ---------- ---------- ---------- ---------- ----------
```

```
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- MILHHLLILL
---------- ---------- ---------- ---------- MLVWQSILLF
---------- ---------- ---------M KVNLKLIIGS ILISQAQAIW
---------- ---------- ---------- ---MNLKEWL LFSDAVF---
---------- ---------- ---------- ---MKLQS-L LVSAAVLT-S
---------- ---------- ---------- ---MQLQD-L VVTVSLLA-A
---------- ---------- ---------- -MIRPLCSKI IISYIFAISQ
---------- ---------- ---------- --MWFLNSVN LLFLVCSVAL
---------- ---------- ---------- ---MQLS--V LIASVLAAGA
---------- ---------- ---------- ---MQLRN-I LQASSLISGL
---------- ---------- ---------- ---------- DITFAG----
---------- ---MHLPSSL LIAAPLLANV SAEPIRIPQR DVSVVSTSQQ
---------- ---------- ---------- ---------- -----MRPG-
---------- ---------- ------MYVN YIGLFAFVQI SLTLAYPPGR
---------- ---------- ------MKLS SFGLFLALQL LPALGLPSR-
---------- ---------- ---------M YFQSFYFLAL LLATAVYGQ-
---------- ---------- ----MLFRGL SLWMLFLASC LSALALPAAE
---------- ---------- ------MSIA TGTFAFSLFA TIAFAVPPET
---------- ---------- ------MSII TTAFALSLLA TTAFAVPPET
YVGNRKKSEQ DFFKMLSSRD RDAHSTLRKR SNSLSSFLST KSTSASENKF
---------F NLTKELQRHS CRDQGKITQK DDALDFISYS SFQSSFNTDQ
QLLLRKSSSP SSYSPIKSDL PNIYSHLRSN DSESPPQPSP KQQSSLSSSS
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
TISKLYSDNN RYHSLQLPHF PHYKEMVKER NVMEKKAENK PSSAASAPHT
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- -----IINYC VATSPTN--- ---------- ----------
---------- -----LVGCV LSKSPTN--- ---------- ----------
---------- -----PFDSS GSSSSSDSSP SETGSSGGTF PFDLFGSGSS
---------- -----FAQGT LAWSPSN--- ---------- ----------
---------- -----LTENV NAWSPNN--- ---------- ----------
---------- -----FNGGV EAWSPTN--- ---------- ----------
```

```
---------- -----FLLAA NAWSPTD--- ---------- ----------
---------- -----HLDAV NAWSPTN--- ---------- ----------
---------- -----AVDAA S--------- ---------- ----------
---------- -----SLAAD SSSTTGD--- ---------- ----------
---------- ------VQRA LPNAPDG--- ---------- ----------
---------- -----LAVRA LPDSPSGG-- ---------- ----------
---------- ------MHDT PLSLMQK--- ---------- ----------
VEISEIYDFE ESSSYKGQDI DTSVLYT--- ---------- ----------
---------- -IDEVDVSDP ELIGLLKPDN VD-------- ----------
---------- ------VASP ELHSLSR--- ---------- ----------
---------- -------DDG SVKVFKR--- ---------- ----------
P--------- -----RIELQ AERGLGDK-- ---------- ----------
P--------- -----RIELQ AERGLGDQ-- ---------- ----------
HGGLNWLSLK LNLLLRLQGR MNSARTN--- ---------- ----------
KSANNGSTVR RSIRSIFRRA AELPRVH--- ---------- ----------
SSSSSSNTKS STTKNIFKKL LRINKSSDNI DESRSIVSNN GGSPMSDSTT
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
SENNNNDNGS NITSLETLSS SEISEFNIDN VKMKHLQDLI DEPDDFSQPI
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---MELALAL SLIASVAAAP
---------- ---------- ---------- ---MKLALAL SLIASVAAAP
---------- ---------- ---------- ---MKLALAL SLIASVAAAP
---------- ---------- ---------- ---MKLCL-L ALGAAVAAAP
---------- ---------- ---------- ---MKLALVL SLIVSVAAAP
---------- ---------- --------MV SKSLFLAAAV NLAGVLAQAP
---------- ---------- --------MV SKSLFLAAAV NLAGVLAQAP
---------- ---------- --------MV SKTFFLAAAL NVVGTLAQAP
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
LTQSSSAQAS STKSTSDSAS STDSSLFSSS NSGSSWYQTF LDGDSGDQKT
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- -----REALA ISLSKRDSVG
---------- ---------- ---------- -----LSKRK PALVKRSTDA
---------- ---------- ---------- -----KPANS IPLSKRSTSP
---------- ---------- ---------- ------RNWK KPPPFPSTNA
---------- ---------- ---------- ---------A KKHSTKQEGP
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------T SMNPYSCDSN ENLSTLSSVN QNFNSRQLLA
---------- ---------M GPLTYSHGIN ELVNKKLRKD CDLST-----
VTSTLSTDTA PKRGKSIQRS QILHHTDSDS LYLENQIELR PEISKSIGNI
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
```

```
HLRLERYLRL LNIALCLRPH ANRLFEFYEL SPLGNLLSRE SGFQGKQGYL
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

------MKTL LIFLAFLSSI FAS------- ---------- ----------
------MKGL VFLLGLLPTI YAS------- ---------- ----------
------MRDL ILFLSLLHTI FAS------- ---------- ----------
------MRGI AVFLAFISLI FAS------- ---------- ----------
------MKTL VVLCTLLSII FAS------- ---------- ----------
--------ML YLILFLIAPI YAG------- ---------- ----------
--------ML FLLFLLITPI YAG------- ---------- ----------
--------ML FLLFLLVAPI YAG------- ---------- ----------
--------ML YLILFLIAPI YAG------- ---------- ----------
---------M FVFLALITLT TCL------- ---------- ----------
TATLANGDTI TGLNAIINEA FLG------- ---------- ----------
TAKLANGDTI TGLNAIINEA FLG------- ---------- ----------
TATLANGDTI TGLNAIINEA FLG------- ---------- ----------
TATLANGDTI TGLNAIVNEK FLG------- ---------- ----------
TATLANGDTI TGLNAIINEA FLG------- ---------- ----------
RPSLNGNEVI SGVLEGKVDT FKG------- ---------- ----------
TAVLNGNEVI SGVLEGKVDT FKG------- ---------- ----------
TAVLNGNEVI SGVLEGKVDT FKG------- ---------- ----------
---------- -----MVSFI SIS------- ---------- ----------
---------- -----MVSFI SIS------- ---------- ----------
---------- -----MVLKQ RAN------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
GYAPGPVSCP --SSQLIRSG SQG------- ---------- ----------
LYTPGYVQCP --EGKLTRSS LDG------- ---------- ----------
DYAPFNLTCP --SKKTFIRT ASE------- ---------- ----------
SYTPANVSCD -EDINLIRQA S-G------- ---------- ----------
SYVPANVTCD -DDINLVREA S-G------- ---------- ----------
SYVPANVTCP -NDINLLRNA T-G------- ---------- ----------
SYVPGTVSCP -DDINLVREA T-S------- ---------- ----------
GYAPGVVDCD -ENINLVRKA D-A------- ---------- ----------
-YTPQNVSCP -DNANFIRNA ADG------- ---------- ----------
GYAPSIIPCP SDDTSLVRNA S-G------- ---------- ----------
-YVPTSVSCP --ASRPTVRS AAK------- ---------- ----------
-YAPAVVDCP --KTKPTLRK AVD------- ---------- ----------
SYAPYNVTCP S-DYMLRPAS DG-------- ---------- ----------
SYAPFNVTCS N-DNLLRPAS EG-------- ---------- ----------
SYAPYTVACP S-GSLLRPAS DG-------- ---------- ----------
SYAPVIRSCD SSEIMVNSLP RGE------- ---------- ----------
SYAPYYVDCP S-DNIVESLS SNE------- ---------- ----------
SYAPWQVDCP -SNVTWIRNA TTG------- ---------- ----------
SYAPWQVDCP -SNVTWIRNA TTG------- ---------- ----------
TIVPESIQNG CSLLRITKKK VRQRKVSLDP ---------- ----------
--LCRVLQRG IRMIRMTRRR RKFYEFKLIN ---------- ----------
KIPSIFTNDG MPLLKISHKS KKRILFWIDP SCFKFSWRMA NSTTTTTSAT
---------- ---MSVTSTS LNG------- ---------- ----------
----MPLELP SLNASIVGNT VQN------- ---------- ----------
VIRSTAKAQG WRVSHFGKHA FKDMIDRHTT KWFLVRN--- -----SYLTY
MYFPFLGRLS ITDYIIVVLV YIES------ ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- -------LIG LTPPSKDSFY
---------- ---------- ---------- -------LVH ITPASEDDFY
---------- ---------- ---------- -------LFS LKPPSQDDFY
---------- ---------- ---------- -------PLT VKSPLVDDFY
---------- ---------- ---------- -------PLS LKSPLVDDFY
---------- ---------- ---------- -------LIF PTKPSSDPFY
---------- ---------- ---------- -------LIF PTKPSSDPFY
---------- ---------- ---------- -------LIL PTKPSNDPFY
---------- ---------- ---------- -------VLL PTKPSIDPFY
---------- ---------- ---------- -------QIP LN-PTIDDFY
---------- ---------- ---------- -------IPF AEPPVGNLRF
---------- ---------- ---------- -------IPF AEPPVGNLRF
---------- ---------- ---------- -------IPF AEPPVGNLRF
```

```
---------- ---------- ---------- -------IPF AEPPVGTLRF
---------- ---------- ---------- -------IPF AQPPVGNLRF
---------- ---------- ---------- -------IPF ADPPLNDLRF
---------- ---------- ---------- -------IPF ADPPLNDLRF
---------- ---------- ---------- -------IPF ADPPVGDLRF
---------- ---------- ---------- ---------- -QGVSLCLLV
---------- ---------- ---------- ---------- -QGVSLCLLV
---------- ---------- ---------- ---------- -YLGFLIVFF
---------- ---------- ---------- ---------- -MRLSFFTAL
---------- ---------- ---------- ---------- -MRLSFFTAL
---------- ---------- ---------- ---------- -MRSSLVLFF
---------- ---------- ---------- ---------- -INPNEQSYI
---------- ---------- ---------- ---------- -INSNEKAYI
---------- ---------- ---------- ---------- -LSQQEKDYI
---------- ---------- ---------- ---------- -PSDNETEWL
---------- ---------- ---------- ---------- -LSDNETEWL
---------- ---------- ---------- ---------- -LSQSEIDWL
---------- ---------- ---------- ---------- -ISQNESAWL
---------- ---------- ---------- ---------- -VSDDEADWL
---------- ---------- ---------- ---------- -LSPAEKEWL
---------- ---------- ---------- ---------- -LSTAETDWL
---------- ---------- ---------- ---------- -LSTNETSWL
---------- ---------- ---------- ---------- -LSNEEKNWL
---------- ---------- ---------- ---------- -ISSGEQSFI
---------- ---------- ---------- ---------- -LNEGEQSYI
---------- ---------- ---------- ---------- -LSTGEQEFV
---------- ---------- ---------- ---------- -LPDLENDFI
---------- ---------- ---------- ---------- -IPSAESEYL
---------- ---------- ---------- ---------- -LGSGERAYI
---------- ---------- ---------- ---------- -LGTGERAYI
---------- ---------- ---------I SGYLMLDKNT GKAYKKLCVD
---------- ---------- ---------N NGQIIWKDGS ----KYLELD
TSATTSGLPQ GITNTTALSN SAIISTPAIA TSAIHRLSIT NRTTHEFVLD
---------- ---------- ---------- ---------- TFNGISEDGI
---------- ---------- ---------- --------GA VEQFLNIRYA
VSDLSSTTPL DVFLIDWKFK VRFSGNKNNI LDNENEINWI IHDPNLEIND
---------- ---------- ---------- ---------- -IISSVLKLI
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- --------MK LLSLTGVAGV

SPPVGFATAK PGDILKIRN- --TPSAP--- ---------- ---SSLYLPI
NPPAGFESAK NGDILKLRN- --SPNRL--- ---------- ---ASFYFPI
KAPAGFKAAK PGDILKTRK- --SPNKP--- ---------- ---SSLYAPV
TAPDGYESAK LGEILKLRK- --TPSKL--- ---------- ---SSMFFEI
NPPRGYESAK LGEILKLRK- --TPGKI--- ---------- ---SSLFIPV
NPPKGFENAA VGDILQSRA- --TPKSI--- ---------- ---TGGFTPL
NPPKGFEKAA VGDILQSRE- --TPKSI--- ---------- ---TGRFAPL
NAPAGFEKAA VGDILQSRK- --TPKPI--- ---------- ---TGVFVPV
NAPEGFKNAT VGDILQFRK- --TPKSI--- ---------- ---TGGFVPL
NPPKDLETSQ LGDVLKWRK- --MPFPV--- ---------- ---TSMFVNL
KDPVPYSGSL DGQKFTSYGP SCMQQNP--- ---------- ---EGTYEEN
KDPVPYSGSL NGQKFTSYGP SCMQQNP--- ---------- ---EGTFEEN
KDPVPYRGSL NGQSFTAYGP SCMQQNP--- ---------- ---EGTYEEN
KPPVPYSASL NGQQFTSYGP SCMQMNP--- ---------- ---MGSFEDT
KPPVPYSASL NGQKFTSYGP SCMQMNP--- ---------- ---LGNWDSS
KHPQPFTGSY QGLKANDFSP ACMQLDP--- ---------- ---GNSLTLL
KHPQPFTGSY QGLKANDFSP ACMQLDP--- ---------- ---GNSLTLL
KHPQPFTGSY QGLKANDFSS ACMQLDP--- ---------- ---GNAISLL
SSMMLGSSAV PVSGKSGSSN TAVSASD--- ---------N AALPPLISSR
SSMMLGSSAV PVSGKSGSSN TAVSASD--- ---------N AALPPLISSR
TAFLV--EAV PIKRQS---N STVDS----- ---------- --LPPLIPSR
SAVASLGYAL PGKLQS---- ---------- ---------- ----------
SAVASLGYAL PGKLQS---- ---------- ---------- ----------
---VSAWTAL ASPIR----- ---------- ---------- ----------
NARYPIAKQA LSKFLHN--A NLQNFD---- -VDSFLAH-- --SNPTIGLA
DRRYANAKSE LSRFLHN--A KMVDFD---- -VDGFLN--- --SNPTIGLA
HKRQETTNKN LIDFLSKR-A NLSDFD---- -AKSFINDNA PNHNITIGLS
KKRDVYTREA LRSFLDRATS NFSDSS---- LVSQLF--SN ASDIPRIAVA
KKRDAYTKEA LHSFLNRATS NFSDTS---- LLSTLFG-SN SSNMPKIAVA
KKRDVNTREA LESFLKRVTS NFTSNSSASN LIDQLFS-TN SSNIPKIGIA
EKRNKVTSVA LKDFLTRATA NFSDSS---E VLSKLFNDGN SENLPKIAVA
```

```
KVRHESTVPA LKDFLQRGFK GFTNDTS--- IIDKLLA--T QDTAPKVAIA
KKRDPITRDA LQTFLRRAFA NVS-TE---- ITSALFN--D TENVPKLGIA
KKRDAYTKEA LHSFLSRATS NFSDTS---- LLSTLFSS-N SSNVPKIGIA
EVRRGKTLSA LKDFFGHVKV GDYDVGA--- YLDKHSG--N SSSLPNIGIA
SIRRKNTIQP MRDLLKRANI TGFDSET--- FMNEAAN--N ISQLPNVAIA
DKRIPKINTQ MRSFIS--NT GLDVDVN--- ------SVIN DSDGPRLGLA
NKRISKVNSE LRSFIS--KT GLNVDLD--- ------KVVN SSDGPRLGIA
DKRVSKVNSA LESFIS--KT GLKIDTK--- ------SVLN DTDGPRLGIA
EKRLSNANEA LTTFLQSKNT TADLDLS--- ------SIVG D-NGPRLGIA
STRSTITNTA MKDFLR--NA NLPGLNA--- ------DTLS GSEGPSIGIA
EAREKLVQPV IEQMMAARGL ETPPRT---- ---------- ----PNIGVA
EAREKLVQPA IEQMMAARGL ETPPRT---- ---------- ----PVIGVA
DIKEIRQGRD ARNYREQYKI S--SENEKRW FTIIYCAD-- NKLKAMHMIS
SVKDIRIGDT ASTYQEEVDP KRLRSDSKLW IAIIYKVS-- NKLKALHVVA
DIKSIYIQNE GSGYREELNI S--QKLEKNW ITIIYFNHKK NSLKSLHLIT
EIFKGIKYAN IPYRWAYAER IDDYDNG--- ---------- -----VFDCT
DIPGKFEKPV LKNDWNGAEI DATKVGP--- ---------- -----VCPQP
ELEEFGIEND ANNILDKNGK SKTHQKKSNI SSKLLLLTLE NSERKLKIIC
PQPMINLFEW LINFSTSSDD NTIEEKLR-- ---------S APTIHEMCAI
MKNDNKANDI IIDSVKVPDS YKPPKNP--- ---------- ------IVFC
LATCVAATPL VKRLPSGSDP AFSQPKS--- ---------- --------VL

VVKNAWQLLI RS-------- ---EDSFGNP ---------- ----------
DVKNAWQLLV KS-------- ---EDSFGNP ---------- ----------
DVQNSWQLLV RS-------- ---EDSFGNP ---------- ----------
DIKNSWQLLV RS-------- ---EDSFGNA ---------- ----------
EVKNSWQLLV RS-------- ---EDSFGNA ---------- ----------
KIQNSWQLLV RS-------- ---EDSFGNP ---------- ----------
KIQNSWQLLV RS-------- ---EDSFGNP ---------- ----------
KIQNSWQLLV RS-------- ---EDSFGNP ---------- ----------
NVQNSWQFLV RS-------- ---EDSFGNP ---------- ----------
PISNAWQISV RS-------- ---QDTLNNS ---------- ----------
LPKAALDLVM QS-------K VFEAVSPSSE ---------- ----------
LGKTALDLVM QS-------K VFQAVLPQSE ---------- ----------
LPKVALDLVM QS-------K VFQAVLPNSE ---------- ----------
LPKNARHLVL QS-------K IFQVVLPNDE ---------- ----------
LPKAAINSLM QS-------K LFQAVLPNGE ---------- ----------
DKALGLAKVI PEEFRGPLYD MAKGTVSMNE ---------- ----------
DKALGLAKVI PEEFRGPLYD MAKGTVSMNE ---------- ----------
DKVVGLGKII PDNLRGPLYD MAQGSVSMNE ---------- ----------
CAPPSNKGSK SDLQAEPYNM QKNTEWYESH ---------- ----------
CAPPSNKGSK SDLQAEPYNM QKNTEWYESH ---------- ----------
TSAPSSSPST TDPEAP--AM SRN------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
FSGGGYRAML TGAGEISSLD SRT-KTNTPV ---------- ----------
FSGGGYRAML AGAGELLALD SR--ATNPSV ---------- ----------
FSGGGYRAML AGAGQILGLD GRYEDANKHG ---------- ----------
CSGGGYRAML SGAGMLAAMD NRTDGANEHG ---------- ----------
CSGGGYRAML SGAGMLAAMD NRTDGANEHG ---------- ----------
ASGGGYRAML SGAGMVSAMD NRTDGANEHG ---------- ----------
VSGGGYRSML TGAGVLAAMD NRTEGAYEHG ---------- ----------
CSGGGYRAML SGAGMISAMD NRTDGANDHG ---------- ----------
VAGGGYRAMF VGAGAFAAMD NRTDGANEHG ---------- ----------
CSGGGYRAML GGAGMIAAMD NRTDGANEHG ---------- ----------
VSGGGWRALM NGAGAVKAFD SRTDNATATG H--------- ----------
ISGGGYRALM NGAGFVAAAD NRIQNTTGAG G--------- ----------
FSGGGLRAMV HGGGVLNAFD SRNGNGSSLA ---------- ----------
FSGGGLRAMV NGGGAFNAFD SRFESDSPLS ---------- ----------
ISGGGFPAML TGAGAINAFD ARNGNTTSLG ---------- ----------
VSGGGWRSML FGGGALAALD SR-SNETTLG ---------- ----------
LSGGGLRAMI LGSGALSAMD ARHDNHTVLT ---------- ----------
LSGGGYRAML TGLGGIMGMM NESTEASESE ---------- ----------
LAGGGYRAML TGLGGIMGMM NESTEASQSE ---------- ----------
PTLDAHNQWI MALEGLKTYR LN---EFIIG LNLVCH---- ----------
LNELDFNTFL SCICGLVKLR R----ELMES ILLPDN---- ----------
DNDHDFKKLI SAIYNLKQLR SQLAKEFLID LNELDE---- ----------
QEGMACPQVL PFDYN----- -IEKGPKEMP ---------- ----------
RTPFNFFSVP DDLWEKVNVD TYQDGLLCDN ---------- ----------
KSESSLKQWM SSIIKMSTST PWSKPNRFGS FAPVRTNSFC KFLVDGRDYF
```

```
FDISVEDHLV RTEDNYILTL HRIPPISKNR ---------- ----------
HGLSGFDKLI LIP------- ---------- ---------- ----------
DAGLTCQGAS PS-------- ---------- ---------- ----------

------NAFV ATLIQPLN-- ---------- -ANPSKLVSY QSWEDASHID
------NAFV TTLIEPYN-- ---------- -ADPSKVVSY QTWEDASNIN
------NAFV TTIIQPKN-- ---------- -ADPSKVVSY QNWEDASNIN
------TAIV TTVIEPYN-- ---------- -ADPSKVLSY QTFEDSANIE
------AAIV TTVIEPFN-- ---------- -ADPSKVVSY QSWEDAANIE
------NVIV TTVIEPVN-- ---------- -ADPSKIASY QVFEDAAKAD
------NAIV TTVIEPVN-- ---------- -ADPSKIASY QVFEDAAKAD
------NVIV TTVMEPFN-- ---------- -ADPSKLASY QVFEDSAKAD
------NVIV TTVIEPVN-- ---------- -ADPSKIASY QVSENAARAD
------LAIV ATILEPPN-- ---------- -GDKNKLISH QAFENSPLLS
------DCLT INVVRPPG-- ---------- -TKAGANLPV MLWIFGGGFE
------DCLT INVVRPPG-- ---------- -TKAGANLPV MLWIFGGGFE
------DCLT INVVRPPG-- ---------- -TKAGANLPV MLWIFGGGFE
------DCLT INVIRPPG-- ---------- -TRASAGLPV MLWIFGGGFE
------DCLT INVVRPSG-- ---------- -TKPGANLPV MVWIFGGGFE
------DCLY LNVFRPAG-- ---------- -TKPDAKLPV MVWIYGGAFV
------DCLY LNVFRPAG-- ---------- -TKPDAKLPV MVWIYGGAFV
------DCLY LNVFRPAG-- ---------- -TKPDAKLPV MVWIYGGAFV
------GGNL TSIGKRDD-- ---------- NLVGGMTLDL PSDAP-----
------GGNL TSIGKRDD-- ---------- NLVGGMTLDL PSDAP-----
-------GPL PS----DV-- ---------- ETKYGMALNA TSYPD-----
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
-----LAGIL QASSYIAG-- ---------- LSGGSWLVGS LASNNLNSVD
-----LSGIL QSSSYIVG-- ---------- LSGGSWLVGS LASNDLIPVD
-----LGGLL DSSTYVVG-- ---------- LSGGNWLVGS LALNDWLSVG
-----LGGLL QSTTYLAG-- ---------- LSGGNWLVGT LAWNNWTSVQ
-----LGGLL QGATYLAG-- ---------- LSGGNWLTST LAWNNWTSVQ
-----LGGLL QAATYLAG-- ---------- LSGGNWLTTT LSWNNWTSVQ
-----LGGLL QSTTYLSG-- ---------- ASGGNWLVGT LALNNWTSVQ
-----LGGLL QSSTYLAG-- ---------- LSGGNWLVGT LAYNNWTSVQ
-----LGGLL QAATYMAG-- ---------- LSGGNWLTGT LAYNNFTSVQ
-----LGGLL QSSTYLSG-- ---------- LSGGNWLTGT LAWNNWTSVQ
-----LGGLL QSATYISG-- ---------- LSGGSWLLGS IYINNFTTVD
-----IGGLL QSSTYLAG-- ---------- LSGGGWLVGS LFSNNFSSIE
-------GIL QSAMYIAG-- ---------- LSGGSWLVGS VAVNNFANIT
-------GLL QSAMYISG-- ---------- LSGGSWLVGS VAINNFTNIT
-------GIL QSSMYLTG-- ---------- LSGGSWLVGS VAVNNFANIT
-------GLL QSAHYITG-- ---------- ADGGSWLLSS LAVNEFRTIQ
-------GLL QASDYLVG-- ---------- TDGSAWTVGG IALNNFSTIN
-----TGGWL DGVSYWAG-- ---------- LSGGSWATGT FMSNGGQLPT
-----TGGWL DGVSYWSG-- ---------- LSGGSWATGS FMSNGGQLPT
----QGEKMI DYSENLNP-- ---------W EKLEKEQSAQ LDLGDVHRMC
----SQFARI HWQITVS--- ---------- EKEEDEKKDT LSFADVKKLC
----NYVKML LNKELLAGDN GNVDGNEVDI RKSHKHVREF LSFNDILKYS
-----FDEFE CSNLMITR-- ---------- -PQGATNLPV FVWIHGGGNL
---------- LIVTRPKG-- ---------- -VSANARLPT VVWIHGGSNI
WSLSEALLMA KDVIYIHDWW LSPELYLRRP VKGNQGFRID RMLKSCAEKG
--------FN NKVVYLHHG- --------LL MCSDVWCCNI ERHKNLPFVL
---------- -SVFHLTN-- ---------- ---------- ----------
-------SVS KPILLVPG-- ---------- ---------- ----------

CSPSY----- ---------- ---------- ---------- --GMQFKSPA
CSPSY----- ---------- ---------- ---------- --GAQFGSPL
CSPSY----- ---------- ---------- ---------- --GSQLGAPL
CSPSY----- ---------- ---------- ---------- --GMQYGAPW
CSPSY----- ---------- ---------- ---------- --GMQFGAPL
CAPSY----- ---------- ---------- ---------- --ALQFGSDL
CAPSY----- ---------- ---------- ---------- --ALQFGSDL
CAPSY----- ---------- ---------- ---------- --ALQFGSDV
CAPSY----- ---------- ---------- ---------- --ALQFGSDV
CSPSY----- ---------- ---------- ---------- --SMQVPS-F
VG-------- ---------- ---------- ---------- --GTSTFPPA
IG-------- ---------- ---------- ---------- --SPTIFPPA
IG-------- ---------- ---------- ---------- --SPTIFPPA
LG-------- ---------- ---------- ---------- --GSSLFPGD
```

```
VG--------- ---------- ---------- ---------- --GSSLFPPA
YG--------- ---------- ---------- ---------- --SSAAYPGN
YG--------- ---------- ---------- ---------- --SSAAYPGN
FG--------- ---------- ---------- ---------- --SSASYPGN
---------- ---------- ---------- ---------- -------PIS
---------- ---------- ---------- ---------- -------PIS
---------- ---------- ---------- ---------- -------SVV
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
DMLSQG---- -LWELTHSFL S---Y----- ---------- --YGIEHPIK
QLLREDK--- -LWDIQNSLV A---Y----- ---------- --YGVN-IVR
DIVNGKST-- -IWQLQDSIL N---P----- ---------- --SGMR-IDK
DIVNNMTEDD SIWDISNSII N---P----- ---------- --GGFM-IVT
AIVDNTTESN SIWDISHSIL T---P----- ---------- --DGIN-IFK
DIVDSQDNDS AIWDISHSIV S---P----- ---------- --GGIN-IFK
DILNNMQNDD SIWDLSDSIV T---P----- ---------- --GGIN-IFK
AIINNMTDDN SIWDISNSIV N---P----- ---------- --GGIN-IFS
QILEEGDKAD AIWNITNSFL N---P----- ---------- --YDKD-FSK
EIVDHMSESD SIWNITKSIV N---P----- ---------- --GGSN-LTY
KLQTHEAG-- SVWQFGNSII EGPDA----- ---------- --GGIQ-LLD
TLLSENK--- -VWDFENSIF KGPKE----- ---------- --AGLS-TVN
---------- YLRDNVWNLE HSVFA----- ---------- --PHGDNVVE
---------- YLRDNVWNLE HSVFA----- ---------- --PHGDNVIE
---------- FLHDDVWNLD HSLFA----- ---------- --PY-DDAFE
---------- NISKSIWYTR LGIFF----- ---------- --IEETHFGD
---------- DFSK-LWAFN HPLMY----- ---------- --PKSAIVFN
NLLEN----- -LWNIDSNLV FP--------- ---------- -------DDD
TLLEN----- -LWNIDSNLV FP--------- ---------- -------DDG
QMLHLNASME FLEETFQKAD ADHSG----- --------KL SFEEFQHFVS
DKFHIYVSTG QLLEFFQLAD INHNG----- --------LL NYFEFEKFIK
KRLNINVNTN HLQQIFDQVL LLSSATTEKP VSTPLFEKGL NFEQFKQFVS
AGNG------ ---------- ---------- ---------- --YCSDHNPV
EGSIYN---- ---------- ---------- ---------- ----LIYEPQ
IKIFIVIYRN VGNIVGTDSL WTKHSMLNLH PNIHIIRSPN QWLQNTYFWA
HDLGYD---- ---------- ---------- ---------- --VWMGNNRG
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

TTVTTQIDMT LIVPLLQNGY YVIIPDYEGP KSTFTVGRQS GK--------
STITTQIDMT LIVPPLRSGY YVVTPDYEGP KATFAVGRQS GQ--------
STILTQLDMT FIVPPLKSGY YVVLPDYEGP KSTFGVGRQS GK--------
STVATQIDMA LMVPMLKQGY YVVSPDYEGP KSTFTVGRQS GK--------
SSVQTQVDMI FIVPLLDKGC FVVLPDYEGP KSTFGVGRQS GK--------
TTFVTQAEMY LMAPLLDQGY YVVSPDYEGP KSTFTIGKQS GQ--------
TTFVTQAEMY LMAPLLDQGY YVVSPDYEGP KSTFTIGKQS GQ--------
TTIATQVETY LLAPLLDQGY YVVSPDYEGP KSTFTVGKQS GQ--------
STLATQAETY LLAPLLDKGY YVVSPDYEGP KSTFTVGKQS GQ--------
ETFQIQADLI FISGLLSQGW YVVVPDYEGP NSVFPVGRQS AY--------
QMITKSIAMG KPIIHVSVNY RVSSWGFLAG DEIKAEGSAN AG--------
QMVTKSVLMG KPIIHVAVNY RVASWGFLAG DDIKAEGSGN AG--------
QMVSKSVLMG KPIIHVAVNY RLASFGFLAG PDIKAEGSSN AG--------
QMVAKSVLMG KPVIHVSMNY RVASWGFLAG PDIQNEGSGN AG--------
QMITASVLMG KPIIHVSMNY RVASWGFLAG PDIKAEGSGN AG--------
SYVKESINMG QPVVFVSINY RTGPFGFLGG DAITAEGNTN AG--------
SYVKESINMG QPVVFVSINY RTGPFGFLGG DAITAEGNTN AG--------
GYVKESVEMG QPVVFVSINY RTGPYGFLGG DAITAEGNTN AG--------
LSSSTNSASD GGKVVAATTA QIQEFTKYAG IAATAYCRSV VP--------
LSSSTNSASD GGKVVAATTA QIQEFTKYAG IAATAYCRSV VP--------
QAMS----ID GG-IRAATSQ EINELTYYTT LSANSYCRTV IP--------
---------- ----RDVSTS ELDQFEFWVQ YAAASYYEAD YT--------
---------- ----RDVSTS ELDQFEFWVQ YAAASYYEAD YT--------
---------- ----REVSQD LFNQFNLFAQ YSAAAYCGKN ND--------
QVEEWVNVGN QVASKRNANF NVSLTDIYGR LLSYPLLTNT ED--------
NTAMWGNINL QVQTKQLAGF TVSITDVYGR ALSHQLLTNF DN--------
TIAYYYGLAQ AVQAKEDAGF QTSVTDTWGR ALSYQFFEED DSG-------
TIKRWDHISD AVEGKQDAGF NVSLTDIWGR ALSYNFFPSL YR--------
TGSRWDDISD DVQDKKDAGF NISLADVWGR ALAYNFWPSL HR--------
TGSRWDHISD AVEDKQKAGF NVSLADVWGR ALSYQFFPTL YR--------
TAKRWDHISN AVESKQNADY NTSLADIWGR ALAYNFFPSL NR--------
SISRWDDISD AVEEKKKAGF NTSITDVWGR ALSYNFFPSL DE--------
```

```
TLARWTAIGS QVQGKRDAGF NVTITDLWSR ALAYGWFPTL PN--------
TIERWESIVQ EVQAKSDAGF NISLSDLWAR ALSYNFFPSL PD--------
SAGYYKDLAD AVDGKKKAGF DTTLTDIWGR ALSYQMFNAS NG--------
RIQYWSEVAK EVAKKKDAGF ETSITDYWGR ALSYQLIGAD MG--------
NLAYYDDLDD EIDQKKDAGF DTSLTDLWGR ALSRKLVDAT QG--------
NLNYYNDLRK EIDQKKHAGF DCSLTDLWGR ALSRKLVDAE RG--------
NFYIYQEWFE QVLQKKNAGF NVSITDLWGR ALALKLVNPL TG--------
LKNYYTNVVD EVNQKAAAGF NVSLTDYWGR AIARHFVGQL RG--------
-AHFYSSIMN EVAEKANAGF NISLSDYWGR VISRTLGDTT YG--------
KLSFYTELYT ETNAKSDLGF PIQITDVWGL AIGSHVLPER YQL-------
KLSFYTNLYT ETNAKSDLGF PVQITDIWGL AIGSHVLPEP YQL-------
LLKTRSEIVD IFKEYTSGSD -KMSLEQFRH FLSTSQKARL DSD----SIR
ILKNRKEVNM IWSKFTKPPH SHLSFENFFQ FLITEQHEQV DRQ----TAW
ILKDRKDLQE IWDSLAQGKE -VLQFDEIKN FIINIQKENF SDDDDNSTIN
PFVKHSIVAG RPVLHVMIEY RLSAFGYLAV PDTNGNWVGN WG--------
FLVAESVRVG KPIVHVCIEY RLGLAGFLTK NGKGNWG--- ----------
HHEKFVVIDE TFAFIGGTDL CYGRYDTFEH VLRDDAESLL DQNFPGKDYS
NKYSTAHLNK PPKSNKFWDF SIDEFAFFDI PNSIEFILDI TK--------
--LISNSIVH NMAENFMQDD EDKSDNKYTN LLEIEYWIGV KK--------
--TGTTGPQS FDSNWIPLST QLGYTPCWIS PPPFMLNDTQ VN--------

--------AT LNSIRAALQT GAFSGI-KKT AKVALWGYSG ----GSLATG
--------AT LDSVRAILKS GSFSGI-NED AKVALWGYSG ----GSLATG
--------AT LDSIKAVLKT KDFSGI-NDD AKVVLWGYSG ----GSFASG
--------AT LDSIRAILKS NKFTGI-KSD AKVAMWGYSG ----GSLASG
--------AT LDSIKAVLKT KDFSGI-NDD AQVAMWGYSG ----GTIAAG
--------AV LNSIRAALKS GKITNI-KDD AKVVMWGYSG ----GSLASG
--------AV LNSIRATLKS SKITNI-KED AKVVMWGYSG ----GSLASG
--------AV LNSIRAALKS GKITNL-AEN AKVVMWGYSG ----GSLASG
--------AV LNSIRASLKS GKITNI-AED AKVLMWGYSG ----GSLASG
--------SV LDSIRGTIKF FNSTG--NST VKTALLGYSY ----GAVASL
--------LK DQRLGMQWVA DNIAAFGGDP TKVTIFGESA ----GSMSVM
--------LK DQRLGMQWVA DNIAGFGGDP SKVTIFGESA ----GSMSVL
--------LK DQRLGMQWVA DNIAGFGGDP SKVTIFGESA ----GSMSVL
--------LH DQRLAMQWVA DNIAGFGGDP SKVTIYGESA ----GSMSTF
--------LH DQRLGLQWVA DNIAGFGGDP SKVTIFGESA ----GSMSVM
--------LH DQRKGLEWVS DNIANFGGDP DKVMIFGESA ----GAMSVA
--------LH DQRKGLEWVS DNIANFGGDP DKVMIFGESA ----GAMSVA
--------LH DQRKGLEWVS DNIANFGGDP DKVMIFGESA ----GAMSVA
-------GNK WDCVQCQK-- ----WVPDGK IITTFT---- ----------
-------GNK WDCVQCQK-- ----WVPDGK IITTFT---- ----------
-------GAT WDCIHCD--- ----ATEDLK IIKTWS---- ----------
-------AQV GDKLSCSKGN CPEVEATGAT VSYDFSD--- ----------
-------AQV GDKLSCSKGN CPEVEATGAT VSYDFSD--- ----------
-------APA GTNITCTGNA CPEVEKADAT FLYSFED--- ----------
-------EGD AYLWSDVT-S ASNFQSHQMP FPILISD--- -----GRAPD
-------QGA SFLWSDVT-E TTSFQNNEMP YPILAAL--- -----GREPN
------TGGA NITWSSIR-N LSSFQDHSMP YPIVVAN--- -----GRTPG
-------GGV AYTWSTLR-D VEVFQNGEMP FPISVAD--- -----GRYPG
-------GGV GYTWSTLR-E ADVFKNGEMP FPITVAD--- -----GRYPG
-------GGV AYLWSDLR-E SDVFKNAEMP MPISVAD--- -----GRYPG
-------GGI GLTWSSIR-D FPVFQNAEMP FPISVAD--- -----GRYPG
-------GGV GYTWNTLR-D VDVFKNGEMP FPISVAV--- -----GRYPG
-------AGA GLTWSSLR-D NEIFMNGEMP MPISVAD--- -----GRYPG
-------AGS ALTWSSLR-D VDVFKNGEMP LPITVAD--- -----GRYPG
--------GL SYTWSSIA-D TPEFQDGDYP MPFVVAD--- -----GRNPG
--------GP AYTFSSIA-Q TDNFQKAETP FPILVAD--- -----GRAPG
--------GP NITFSSIR-N QTWFQNADYP YPIIISD--- -----SRLEE
--------GP GITYSSMR-N QSWFQNADYP YPIIVAD--- -----SRLEE
--------GA NTTFSSVT-N ETWFQDGEFP FPIIIAD--- -----NVIEG
--------GP NLTYSSVQ-N ASWFQTAEYP YPLIVTQGLT ----GGLPDG
--------FP NVSLSSIT-S QEWYRNANFP YPIITFA--- -----TQNYG
------SNTP NLTFSSLPSV VSALGNASLP MPIIIAADRK -----RREAG
------SNTP NLTFSSLPSV VAALGNASLP MPIIVAADRK -----RREAG
TLYVSFCSND DSKMGLIEFT SFLLSPHNSP VVPVIQDMS- -------RPL
SYFIKYREPT QLTMGQDGFT KFLKE-QPYL VEVKEELYS- -------KPL
LIFQKYCSND N-GWNKESLN EYLLSSYSTP YREITQTQTN ----YYDYPL
--------AR DQYTALQWIS KHIVEFGGDP SQITIGGES- -----AGSIG
--------TW DQYTGCQWVN RHIQDFGGDP LNVTLTGESA ----GSVAVH
NARIADFHDL DKPFESMYDR KVIPRMPWHD VQMMTLGEPA RDLARHFVQR
-------VDK VICIGFSQGS AQMFAAFSLS EKLNRKVSHF IAIAPAMTPK
```

```
---------F LQSKGCTVIT TKVPGFGSIE ERAMALD--- ----------
---------- -TEYMVNAIT ALYAGSGNNK LPVLTWSQGG ----------

WAISLQSKYA PELKENLIGA AVG--GFATN ITAVAEAVDG TVFSGFIPLA
WAAALQPVYA PELQKNIVGA AVG--GFAAN ITAIAESVDG TIFSGLITLA
WAAVLQPEYA PELKDNLIGA ALG--GFAAN LTGIAESVDG EVFSGFIPLA
WAAALQPKYA PELKKNLIGA ALG--GFVTN ITATAEATDG TLFAGLVPNA
WAATLQPKYA QELKKNLIGA ALG--GFVIN ITATAEATDG TLFAGLIPNA
WAAALQPSYA PELGGNLLGA ALG--GFVTN ITATAQATDG TVFAGIVANA
WAAALQPSYA PELSSSLLGA ALG--GFVTN ITATAQAADG TVFAGIVANA
WAAALQPNYA PELGGNLLGA ALG--GFVTN ITATAEATDG TVFAGIMANA
WAAALQPDYA PELSRNLLGV ALG--GFITN VTATVEATDD TIFAGTAANV
WASIVQPNYA PEL--ELVGA AVG--CTIPN ITAFIEKVDE GPYSGLIVNI
CHILWNDGDN TYKGKPLFRA GIMQSGAMVP SDAVDGIYGN EIFDLLASNA
CHLIWNDGDN TYKGKPLFRA GIMQSGAMVP SDPVDGTYGN EIYDLFVSSA
CHLLWNGGDN TYKGKPLFRA GIMQSGAMVP SDPVDGTYGT QIYDTLVAST
VHLVWNDGDN TYNGKPLFRA AIMQSGCMVP SDPVDGTYGT EIYNQVVASA
CQLLWNDGDN TYNGKPLFRA AIMQSGAMVP SDPVDGPYGT QIYDQVVASA
HQLIAYGGDN TYNGKKLFHS AILQSGGPLP YHDSSSVGPD ISYNRFAQYA
HQLIAYGGDN TYNGKKLFHS AILQSGGPLP YHDSSSVGPD ISYNRFAQYA
HQLVAYGGDN TYNGKQLFHS AILQSGGPLP YFDSTSVGPE SAYSRFAQYA
-------SLL SDTNGYVLRS DKQ-------- -KTIYLVFRG TNS-------
-------SLL SDTNGYVLRS DKQ-------- -KTIYLVFRG TNS-------
-------TLI YDTNAMVARG DSE-------- -KTIYIVFRG SSS-------
-------STI TDTAGYIAVD HTN-------- -SAVVLAFRG SYS-------
-------STI TDTAGYIAVD HTN-------- -SAVVLAFRG SYS-------
-------SGV GDVTGFLALD NTN-------- -KLIVLSFRG SRS-------
TTIINLNSTV IELTPYEFGS WDP-----SL NEFVDTRYLG TKLDNGRP-T
TVLMNFNSTV FELTPYEVGS WDP-----SL RSFVDTKYIG TRLDDGAP-V
TYIINENSTI FEISPYELGS WDP-----SL KSFSNIQYLG SSVNNGNPNN
TQIIDLNATV FEFNPFEMGS WDP-----TL NAFTDVKYLG TKVSNGEPVN
TTVINLNATL FEFNPFEMGS WDP-----TL NAFTDVKYLG TNVTNGKPVN
TAVIDLNSTV FEYSPFELGS WDP-----SL SAFTDVQYLG TKVSDGKPAE
TKVINLNATV FEFNPFEMGS WDP-----SL NSFANVKYLG TNVSNGVPLE
TQVVNLNATV FEFNPFEMGS WDY-----TL HTFTDVRYAG TNVTNGTPNV
TTVINLNATV FEMTPFEIGS WDP-----SL NAFSDIKYLG TQVTDGKPE-
TTVINLNATL FEFTPFEMGS WDP-----SL NAFTDVKYLG TNVTNGKPVN
ELVIGSNSTV YEFNPWEFGT FDP-----TI FGFVPLEYLG SKFEGGSLPS
DTIISLNATN YEFNPFETGS WDP-----TV YGFAPTKYLG ANFSNGVIPS
EKAIPANTSI FEFTPYEFGT WDN-----GI KAFLPMEYVG THLKNGVP-P
ETAIPANTSI FEFTAYEFGT WDN-----GI KAFIPMEYVG THLLDGVP-P
ETVIPLNDTV FEFTPIEFGT WDT-----GV ESFIPMEYTG THLINGIP-L
SNGTATNSSI YEISPYYLTS FDN-----NV RSYTPTQYLG TNYSNGTA-V
EDISNVNTTF FEASPNVFGT FDH-----GI NSFIPTEYLG TTLNNGAS-S
ELVIAENATV WEFTPYEFGS WAFGSQYKSP GAFTPIEYLG TSVDDGSP--
ELVIAENATV WEFTPYEFGS WAFGSQYKSP GAFTPIEYLG TSVDDGSP--
NEYLISSSHN TYLLGKQFGG ESS------I EGYIRSLQRG CKCIEIDCWD
NHYFIASSHN TYLLGKQIAE TPS------V EGYIQVLQQG CRCVEIDIWD
NEYFISSSHN TYLTGRQVAG DSS------V EGYIRTLQRG CRCVEIDIWN
LHALMVHESM KPKEECIIHN VILS------ SGTMDRMGTG TISENAFKPI
NMLIKDSMNG RKLFRNAVMM SGT----LET ITPQPPKWHA RLEEKVAKVT
WNYLLRAKRP SRLTPLLTPP SDLTAEELKS LPMFEILREK STCETQILRS
GLHNRIVDTL AKSSPGFMYL FFG----RKI VLPSAVIWQR TLHPTLFNLC
---AQLQKEV KKIESKDKRH SLN------- --LIAHSMGG LDCRYLICNI
----LVAQWG LTFFPSIRSK VDR------- LMAFAPDYKG TVLAGPLDAL

LNG------- ---------- ---------- ---------- LANEYPDFKK
LNG------- ---------- ---------- ---------- LANEYPDLKT
LNG------- ---------- ---------- ---------- IANEYPDFKK
LSG------- ---------- ---------- ---------- LANEYPEFKE
LNG------- ---------- ---------- ---------- LANEFPDFKK
LGG------- ---------- ---------- ---------- VANEYPEFKS
LGG------- ---------- ---------- ---------- VANEYPEFKS
LGG------- ---------- ---------- ---------- VANEYPEFKQ
LGG------- ---------- ---------- ---------- IANEYPEFKS
FNG------- ---------- ---------- ---------- LANEYRHFRD
GCG------- ---------- ---------- ---------- SAS---DKLA
GCG------- ---------- ---------- ---------- SAS---DKLA
GCS------- ---------- ---------- ---------- SAS---NKLA
GCG------- ---------- ---------- ---------- SAS---DKLA
GCG------- ---------- ---------- ---------- SAS---DKLA
```

```
GCD-------- ---------- ---------- ---------- TSASANDTLE
GCD-------- ---------- ---------- ---------- TSASANDTLE
GCD-------- ---------- ---------- ---------- ASAGDNETLA
---------- ---------- ---------- ---------- FRSAITDIVF
---------- ---------- ---------- ---------- FRSAITDIVF
---------- ---------- ---------- ---------- IRNWIADLTF
---------- ---------- ---------- ---------- VRNWVADATF
---------- ---------- ---------- ---------- VRNWVADATF
---------- ---------- ---------- ---------- IENWIGNLNF
GK-------- ---------- ---------- ---------- CYNGFDNAGF
SKR------- ---------- ---------- ---------- CVNGFDNAGF
TDI------- ---------- ---------- ---------- CVNNFDNAGF
KGQ------- ---------- ---------- ---------- CVAGYDNTGF
KGQ------- ---------- ---------- ---------- CIAGFDNTGF
EGK------- ---------- ---------- ---------- CIAGFDNVGF
RGK------- ---------- ---------- ---------- CTAGFDNAGF
TGK------- ---------- ---------- ---------- CVAGFDNTGF
TER------- ---------- ---------- ---------- CINGFDDASF
KDQ------- ---------- ---------- ---------- CVSGYDNAGF
NES------- ---------- ---------- ---------- CIRGFDSAGF
GGK------- ---------- ---------- ---------- CVEGLDQAGF
DHK------- ---------- ---------- ---------- CIRNYDNAGF
DKS------- ---------- ---------- ---------- CIHNYDNAGF
NES------- ---------- ---------- ---------- CVRNFDNAGF
DGK------- ---------- ---------- ---------- CVTQFDNVGF
NGS------- ---------- ---------- ---------- CVINYDNFGF
NGT------- ---------- ---------- ---------- CWKGFDQLSF
NGT------- ---------- ---------- ---------- CWKGFDQLSF
GPN------- ---------- --GPVVCHGH TFTSMIKFND VIDAIRKYAF
GEN------- ---------- --GPVVCHG- FLTSAIPLKT VIRVIKKYAF
GDSNTTTTTV IGTKDDDDKN EYEPIVNHGR TFTKPISFAN VIRAIKKFAF
YDG------- ---------- ---------- ---------- -------IKT
G--------- ---------- ---------- ---------- ---------K
AGNWSLG--L KETECSIQNA YLKLIEQSEH FIYIENQFFI TSTVWNGTCV
IDIAN----- ---------- ---------- ---------- -KILFNWKSF
KNR------- ---------- ---------- ---------- NYDILSLTTI
AVS------- ---------- ---------- ---------- APSVWQQTTG

RLYGEVKLSA RSTMEKGSQN CLAASLVGYP MSQYFTGQNR AFEKGWGLLQ
AFYEELSDFA VPEFKAGAEN CLAENIFHYP LHQYFTGPKR AFEKGWGLLK
RLYEEVKPGA KADLQKGAEN CLAASLISYP MYQYFTGPRR VFEKGWSLLE
ILYQKVSKAA TDNLRQGTEH CIGGAILYFA EDQYFTGDDR AFPGGYGLLK
RMYEVVEKRY EGALQQGTQH CLGGAILHFA FDQVFTGDHR YFEQGYGLLE
ILQSDTDK-- KSVFEEFDGH CLIDGVLNYI GTSFLTGDHK IFKTGWDILK
ILQSDTDK-- KSVFDEFDSH CLADGVIDYI NTSFLTGDNK IFKTGWDILK
ILQNDTDK-- QSVFDQFDNH CLADGVINYI GKHFLSGTNK IFKSGWNILK
ILQNDTNK-- SSIFNKINNH CLTDSFIKYV GARFLTGDNK VFKSGWNIFK
RLIH------ FGALQPLG-- CLFPICRKFF FQKMIGG--- VYDA--QVLT
CLRGVSSDTL EDATN--NTP GFLAYSSLRL SYLPRPDGVN ITDDMYALVR
CLRSASSDTL LDATN--NTP GFLAYSSLRL SYLPRPDGKN ITDDMYKLVR
CLRGLSTQAL LDATN--DTP GFLSYTSLRL SYLPRPDGAN ITDDMYKLVR
CLRGLSQDTL YQATS--DTP GVLAYPSLRL SYLPRPDGTF ITDDMYALVR
CLRSISNDKL FQATS--DTP GALAYPSLRL SFLPRPDGTF ITDDMFKLVR
CLRSKSSSVL HDAQNSYDLK DLFGLLPQFL GFGPRPDGNI IPDAAYELFR
CLRSKSSSVL HDAQNSYDLK DLFGLLPQFL GFGPRPDGNI IPDAAYELFR
CLRSKSSDVL HSAQNSYDLK DLFGLLPQFL GFGPRPDGNI IPDAAYELYR
NFSDY-KPVK GAKVH----- ---------- ---------- ----------
NFSDY-KPVK GAKVH----- ---------- ---------- ----------
VPVSY-PPVS GTKVH----- ---------- ---------- ----------
VHTNP-GLCD GCLAE----- ---------- ---------- ----------
VHTNP-GLCD GCLAE----- ---------- ---------- ----------
DLKEINDICS GCRGH----- ---------- ---------- ----------
FMGTSSALFN EAVLS-ITEA N-IPSFLKDI IDDILVDPIL KSNIDVSAYN
FMGTSSSLFN IVLQQ-LNNM P-IPPFLKEL ISKFTLDPVE KLNIDIAQYN
IMGTSSSLFN QILLQ-LDNY S-INSIIKMI LEKVLTD-VS DEEYDIAVYE
IMGTSSSLFN QFLLQ-INST S-LPSFIKNL VTGFLDD-LS EDEDDIAIYA
ITATSSTLFN QFLLR-LNST D-LPSFIANL ATDFLED-LS DNSDDIAIYA
LMGTSSTLFN QFLLR-INDT S-IPKFIRNL ATHFLKD-LS EDYDDIAVYA
IMGTSSTLFN QFLLR-INST H-LPSFITRL ARHFLKD-LS QDFNDIAVYS
VMGTSSSLFN QFLLQ-LNTT D-LPSFLYNL LHGFLTD-AS DDYDDISIWA
IMGTSSSLFN EFTMS-NDSA V-AYTYLNTL SSTLVKG-ID KENNDIAMYA
```

```
VIATSASLFN EFSLE-ASTS T-YYKMINSF ANKYVNN-LS QDDDDIAIYA
VIGTSSSLFN QFLLQ-INTT S-LPSFIKDV FNGILFD-LD KSQNDIASYD
VMGTSSTLFN QFLLANISSY DGVPDVLIEA VTSVLKE-IG AKRDDVSQII
VMGTSATLFN TFLLEWSQEV TS-NSTLYDI IHKVFE-KLS EDQNDIAPY-
VMGTSATLFN SFLLDWNENV KK-NDTYYDI LHAILE-DLS KHQDDIAPY-
LMGTSSNVFS GILPATNASL TASNNTFNNA VLSFLE-MLA EDQLDVGLY-
LVGTSSTRYN EALIDVSLRQ SR----MSRR LGFTLR-HMR INGSSVSFY-
MMGASSTYFN KIMRNFNDSS TK----NGRI IQQYLKGNFS ENGQQIISI-
VMGTSATLFN GAFLELNGTD S---GLLTNL ITAFLAD-LG EDQADISRI-
VMGTSATLFN GAFLELNGTD S---GLLTNL ITAFLAD-LG EDQADISRI-
VVSPYPLFIS LEIHCCPDQQ RQMVSYMKQA FGDTLVMKPV TANESVLPSP
ITSPYPLIIS LEINCNKDNQ KLASLIMREV LAEQLYFVGT RTDK--LPSP
IVSPWPLILS LEIHCSPECQ IKVVNILKDI LGENMIIAPI DIDSVILPSP
LVGDINTCSA DELLEAQIKA GLDLGFYLQD DFFPPDWRNV RFKVSRVLLS
EVADLASLSD KELLDAQIKL NVAVCMTCDD GDFFEPGWKQ HLTPDWLDKL
LNKIGDALVD RIVKANQEKK PWKAFILIPL MPGFDSPVDT AEASSLRLIM
NILPRQKIAS YAKLYSTTSV KSIVHWFQIL RSQKFQMFEE SDNMLNSLTR
STPHRGSEMA DYVVD----- ---------- ---------- ----------
SALTTALRNA GGLTQIVPTT N--------- ---------- ----------

D--------- --------EV FNKT----IE DNLLLKLDKT YLPQVPVLIY
E--------- --------DI FNKS----IQ DNLLIGLNKT YLPQVPVLIY
D--------- --------KT IGKT----LE DNLLIALSKE HMPQIPIFVY
E--------- --------EV VNKT----IS ENNLMQMDKD YLPDIPIFVY
E--------- --------EV FNRT----IS GNSLLYMDQE YLPDIPIFVY
N--------- --------PK IGKV----VE DNGLVYQ-KQ LVPKIPVFIY
S--------- --------PT IAKI----VE DNGLVYQ-KQ LVPKIPIFVY
N--------- --------PT ISKI----VE DNGLVYQ-KQ LVPKIPILIY
N--------- --------LV VSKI----VK DNGLVYQ-KQ LIPTIPVFIY
D--------- --------ET IKET----IE INNLLST--R AVPQIPVFLF
EGKYANIPVI IGDQNDEGTF FGTS----SL NVTTDAQARE YFKQSFVHAS
DGKYASVPVI IGDQNDEGTI FGLS----SL NVTTNAQARA YFKQSFIHAS
DGKYASVPVI IGDQNDEGFL FGLS----SL NTTTEADAEA YLRKSFIHAT
DGKYAHVPVI IGDQNDEGTL FGLS----SL NVTTDAQARA YFKQSFIHAS
DGKCANVPVI IGDQNDEGTV FALS----SL NVTTDAQARQ YFKESFIHAS
SGRYAKVPYI SGNQEDEGTA FAPV----AL NATTTPHVKK WLQYIFYDAS
SGRYAKVPYI SGNQEDEGTA FAPV----AL NATTTPHVKK WLQYIFYDAS
SGRYAKVPYI TGNQEDEGTI LAPV----AI NATTTPHVKK WLKYICSEAS
---------- ----AGFLSS YEQV---VND YFPVVQEQLT AHPTYKVIVT
---------- ----AGFLSS YEQV---VND YFPVVQEQLT AHPTYKVIVT
---------- ----KGFLDS YGEV---QNE LVATVLDQFK QYPSYKVAVT
---------- ----LGFWSS WKLV---RDD IIKELKEVVA QNPNYELVVV
---------- ----LGFWSS WKLV---RDD IIKELKEVVA QNPNYELVVV
---------- ----DGFTSS WRSV---ADT LRQKVEDAVR EHPDYRVVFT
PNPFFKS--- ---SGS-NTA ISQS---KNL YLVDGGEDGQ NIPISPLLH-
PNPFHKS--- ---NNS-DTK IAQS---RTL YLADGGEDGQ NVPLLPLIH-
PNPFFGA--- ---DSAGIKS ITTN---DTL YLCDGGEDLQ NVPFYPLIQN
PNPFKDTSY- --IQDNFSKS ISES---DYL YLVDGGEDNQ NIPLVPLVQD
PNPFKEANF- --LQKNATSS IIES---EYL FLVDGGEDNQ NIPLVPLLQK
PNPFRDADY- --VNNNRSKS LSES---EYL FLVDGGEDGQ NVPLVPLIQQ
PNPFKDTKF- --LDSDYTTS IVDS---DSL FLVDGGEDDE NVPVLPLIQK
PNPFYEITN- --IPSNYSQS ISED---DTL YLVDGGEDGQ NIPLTPLLQT
PNPFKGSKY- --VDSNYTTS IVDS---DSL FLVDGGEDLQ GIPFVPLLKQ
ANPFKDTEF- --VDRNYTSS IVDA---DDL FLVDGGEDGQ NLPLVPLIKK
PNPFYKYN-- -----EHSSP YAAQ---KLL DVVDGGEDGQ NVPLHPLIQP
PNPFLDWN-- -----NRTNP NADT---LEL DLVDGGEDLQ NIPLNPLTQP
PNPYQNF--- TTTNTTVKNP FERF---DTI DLVDGGEDDE NIPIWPLLHP
PNPYQNY--- TTSNTSVVNA FEPY---DTI DLVDGGEDRE NIPLWPLLHP
PNPYQGYG-- NASNTTTTNP LEPY---PII ELIDGGSDSE GIPFWPLLHP
PNPYTDATDI AGNATAVSED IVDT---PYL DLFDGGYDGQ NIPIWPLLQP
PNPFQGV--- ESANSDAANN LGSS---SSL NLVDTFLTGE KIPLWPLLQK
PNTFSNYN-- -----SGENP IYNL---TYI TLVDAGETNQ NIPLEPLLVP
PNSFSNYN-- -----SGENP IYNL---TYI TLVDAGETNQ NIPLEPLLVP
EDLLNKILLK VKCSATPLHQ FSTD---ILK VGITDSSTDT TESSELEN--
RELKHKILLK SKKTSEATRG LSVN---EPF PSSFSSSYES ANEQELRMKD
AELKHKFIIK VKKTTSFQNL IETENGSFTT STTTTTTTTT TTTTATSLSE
DVIVDG---- --------TN FKNKIN--PA VRVTPENDFD HKVFKLYNIS
IIS------- --------DC KDEG--MLYF LPVNAQDDEE LLAKVAKSPV
QFQYQSISRG EHSTFSKLKK LNIDP-AQYI QFFSLRKWST FAPNERLITE
PYQIAN---- --------FP TRTN---IKI PILLIYGGID SLVDIDVMKK
---------- ---------- --------LF ENLNALRVSQ KILPICFYQL
```

```
---------- ---------L YSATDEIVQP QVSNSPLDSS YLFNGKNVQA

HGTIDEIIPI KDAN------ ---------- --AQYQIWCD RGIQSLEFAE
HGTVDEIIPI KDPH------ ---------- --AQYQLWCD WGIESLEFAE
HGTIDKIIPI KDSI------ ---------- --KIYKNWCD WGIGSFEFSE
HGALDSIVPI SNVH------ ---------- --VTYKNWCD WGINSFEFSE
HGSLDGIVPI PDVH------ ---------- --GVYKNWCD WGIDSFEFAE
HGSIDQIVPI VDTK------ ---------- --KTYQNWCD AGISSLEFAE
HGSIDQIVPI VNVK------ ---------- --KTYQNWCE GGISSLEFAE
HGAIDQIVPI VNVK------ ---------- --KTYQNWCD AGIASLEFSE
HGSMDQISPI LNPK------ ---------- --KTYQNWCD AGISSIEFAE
HSKFNEMSPF LEIL------ ---------- --KLEKLWCS QLGVNLEIAE
DAEIDTLMTA YPGD------ ---------- --ITQGSPFD TGILNALTPQ
DAEIDTLMAA YPQD------ ---------- --ITQGSPFD TGIFNAITPQ
DADITALKAA YPSD------ ---------- --VTQGSPFD TGILNALTPQ
DAEIDTLMAA YTSD------ ---------- --ITQGSPFD TGIFNAITPQ
DAEIDTLMAA YPSD------ ---------- --ITQGSPFD TGIFNAITPQ
EASIDRVLSL YPQT------ ---------- --LSVGSPFR TGILNALTPQ
EASIDRVLSL YPQT------ ---------- --LSVGSPFR TGILNALTPQ
DASLDRVLSL YPGS------ ---------- --WSEGAPFR TGILNALTPQ
GHSLGGAQAL LAGM------ ---------- --DLYQREPR LSPKNLSIFT
GHSLGGAQAL LAGM------ ---------- --DLYQREPR LSPKNLSIFT
GHSLGGATAL LCAL------ ---------- --DLYQREEG LSSSNLFLYT
GHSLGAAVAT LAAT------ ---------- --DLR----G KGYPSAKLYA
GHSLGAAVAT LAAT------ ---------- --DLR----G KGYPSAKLYA
GHSLGGALAT VAGA------ ---------- --DLR----G NGY-DIDVFS
-RNVSAIFAF DNS------- ---------- --NDVLN-WP DGTSLVKTYE
-RKVSAIFAF DQS------- ---------- --ADKNN-WP DGSALIKTFE
KRGVDVIFAF DNS------- ---------- --ADTNSSWP NGTSIQETYK
ERNVDVIFAL DNS------- ---------- --ADTDYYWP DGASLVSTYE
ERELDVIFAL DNS------- ---------- --ADTDDYWP DGASLVNTYQ
ERDLDIVFAL DNS------- ---------- --ADTEENWP DGASLMHTYR
ERDVDIIFAV DNS------- ---------- --ADMRLAWP DGSSLVHTYE
EREIDVIFAL DNS------- ---------- --ADTDQSWP DGFSLTQTYA
ERDLDIIFAI DVD------- ---------- --TETSDNYP AGGPMMKTYE
ERDLDVVFAL DIS------- ---------- --DNTDESWP SGVCMTNTYE
ERHVDVIFAV DSS------- ---------- --ADTDYFWP NGTSLVATYE
VRAVDVIFAV DSS------- ---------- --ADVTN-WP NGTALRATYE
QRFVDVIFAV DATY------ ---------- --DDSN-GWP DGSSIVTTYE
QRFVDVVFAI DSTY------ ---------- --NDPY-GWP LGSSIVATYE
QRDVDVIFAI DGGY------ ---------- --QSATSGWP DGSSLVSTYE
ERKLDVVFAF DSSG------ ---------- --DTSN-FWP NGSSLVATYE
GRDVDVIVAV DNG------- ---------- --DDSEWLWP NGNSLVQTYE
TRDVDAIVAF DSS------- ---------- --YDTDYIWP NGTALRTTYE
TRDVDAIVAF DSS------- ---------- --YDSDYIWP NGTALRTTYE
----SELTGL RKGK------ ---------- --RRMKNIIV QELQQLAPYA
DSTNSSSATN SSSM------ ---------- --QRIKRIGL KKHADIINDV
DNENNKSNSS STSS------ ---------- --FIIRRRKN KSPKIINELS
TEDTWEDYHY KMML------ ---------- --FKGDETFI RGNQQLELLF
GKEISELYGI KEGG------ ---------- --DIKSACLD LKTDATFNYF
QLYVHAKILI ADDRRCIIGS ANINERSQLG NRDSEVAILI RDTDLIKTKM
NLPFNSVFDV KVDN------ ---------- -YEHLDLIWG KDADTLVIAK
TTAYMKYFNL VTPN------ ---------- --SPKVSYFS YGCSFVPKWY
QAVCGPLFVI DHAG------ ---------- -SLTSQFSYV VGRSALRSTT

DLSAGHLAET FTG------- -----APAAL SWIDARFSGK PAVNGCQRTI
DLSTGHLAET FTG------- -----APAAL AWIDARFDGK TPIQGCSHTT
DKSNGHTTET VVG------- -----APAAL TWIDARFAGK PAVEGCSFTT
DLLNGHITET IVG------- -----APAAI TWLEARFDGE PVVKGCKKTS
DSLNGHLTEI VVG------- -----APAAI TWLDARFDGQ PVVEGCKKTT
DASNGHLTEA IMG------- -----APAAL TWIIDRFDGK QTVSGCQHIQ
DGTNGHLTET VVG------- -----APAAL TWIIDRFNGK QTVSGCQHDK
DATNGHITET IVG------- -----APVAL TWIINRFNGK QTVSGCQHVK
DLTNGHFTES IVG------- -----APAAL TWIIDRFSNK PPVDGCQHVV
DMSYNHMVEA FSG------- -----MPAAI TWIEKRWN-N STLGGCKHVH
FKRISAVLGD LGFTLARRYF LNHYTGGTKY SFLSKQLSG- LPVLGTFHSN
FKRISAVLGD LAFIHARRYF LNHFQGGTKY SFLSKQLSG- LPIMGTFHAN
LKRINAVLGD LTFTLSRRYF LNHYTGGPKY SFLSKQLSG- LPILGTFHAN
FKRISALLGD LAFTLARRYF LNYYQGGTKY SFLSKQLSG- LPVLGTFHGN
FKRIAAVLGD LAFTLPRRYF LNHFQGGTKY SFLSKQLSG- LPVIGTHHAN
FKRVAAILSD MLFQSPRRVM LSATKDVNRW TYLSTHLHNL VPFLGTFHGN
```

```
FKRVAAILSD MLFQSPRRVM LSATKDVNRW TYLSTHLHNL VPFLGTFHGN
FKRIAAIFTD LLFQSPRRVM LNATKDVNRW TYLATQLHNL VPFLGTFHGS
V--------- ----GGPRV- ------GNPT FAYYVESTG- IPFQRTVHKR
V--------- ----GGPRV- ------GNPT FAYYVESTG- IPFQRTVHKR
Q--------- ----GQPRV- ------GDPA FANYVVSTG- IPYRRTVNER
Y--------- ----ASPRV- ------GNAA LAKYITAQ-- GNNFRFTHTN
Y--------- ----ASPRV- ------GNAA LAKYITAQ-- GNNFRFTHTN
Y--------- ----GAPRV- ------GNRA FAEFLTVQTG GTLYRITHTN
R--QFSS--Q GNGIAFPYV- ------PDQY TFRNLNLTSK PTFFGCDAKN
R--QFSS--Q GDGIAFPYV- ------PDQN TFRNTNLTSK PTFFGCDAQN
R--QFSK--Q GKGTPFPFA- ------PDYK TFLDKNMGDK PVFFGCNSSD
R--QFSS--Q GLNMSFPYV- ------PDKR TFVNLGLADK PSFFGCDAQN
R--QFGS--Q GLNLSFPYV- ------PDVN TFVNLGLNKK PTFFGCDARN
R--QFGF--Q GQGVTFPSV- ------PGTD TFVNLGLNKK PTFFGCDARN
R--QFVK--Q GQGMSFPYV- ------PDTN TFVNLGLNKK PTFFGCDANN
R--QFGL--Q GKGIAFPYV- ------PDVN TFTNLGLNTR PTFFGCDARN
R--QFSK--Q GKGMAFPYV- ------PDMT TFVNLGLGGK PSFYGCDANN
R--QYSK--Q GKGMAFPYV- ------PDVN TFLNLGLTNK PTFFGCDAKN
R--SLNSSGI ANGTAFPAV- ------PDQN TFINLGLSTR PSFFGCDSSN
R--TFGS--I SNGTLFPSI- ------PDDW TFINLGLNNR PSFFGCDVKN
RIITYNANKS VDVRGFPYI- ------PDED TIISLGLNTH PTFFGCDGRN
RVVTFNANKS VDVRGFPYI- ------PDEN TIISLGLNTR PTFFGCDGKN
RVLATNSSG- --VRGFPYI- ------PDTN TFLALGLNTH PTFFGCDGRN
RVTQRASDAV YDVEDFVHV- ------PTPE TFVNLGLNAN PTFFGCDGRN
RVVAAQAAGN TNVKGFPYV- ------PSQQ SFVSLHFNDR PVFFGCDGRN
R-AKVLAEHE NTRVLMPEV- ------PSMN GFVNGGYNSR PTFFGCNDTT
R-AKILAEHE NTRVLMPEV- ------PSMN GFVNGGYNSR PTFFGCNDTT
R------SLK FRNFSLPES- ------KTYS HIFSFSER-- TIKKHGKAMV
SNISGIHGIK FRNFSLPES- ------KTIA HCFSLNERKV EYMIKDKHLK
NLGIYTQGIK FRNFSLPES- ------KTFN HCFSLGEKSI NRMIKDDDKK
EQENIPVWRQ LFDQIHPND- ------PSRL CHHAVDLYYM WDNWEMPEDK
NHLLFKKMEE ARNNGSTSR- --------VY RLAVDEPNPH NPDQRAHHAV
NGDDYYAGKF PWELRQRLMR EHLGCDVDLV EFVEKKFERF EKFAAKNYEK
VLRFIEFFNP GNVSVKTNQL LP--SASLVE ELPSTTWKTT HPTHGLSYRT
N--------- ---------- ---------- VFCTPWKIVY ERSKGCPNDG
G--------- ---------- --------QA RSADYGITDC NPLPANDLTP

R--------- ---------- ---SSNVLYP GISIT--IRI YFEGISKTIF
R--------- ---------- ---LTNLLYP NTSDS--THS YFLGIYQAVF
R--------- ---------- ---ASNFLYP NISES--AAS YFKGIYQTIL
R--------- ---------- ---ITNFSYP NISDS--TSS IFEGILNSVT
R--------- ---------- ---ITNFSYP NISDS--TRN FFKGILDSLT
R--------- ---------- ---FSNLEYP NIPSS--IAN YFKAAMDVVL
R--------- ---------- ---LSNFQYP NISSS--ILK YFKVALDTMM
R--------- ---------- ---TSNFEYP NIPPS--ILN YFKAALNILI
R--------- ---------- ---TTNYEYP NVSSS--ILD YFKAAMDVVA
R--------- ---------- ---LSNFEYP GIAPF--LSQ YFRSSLQMVL
D--------- ---------- ---IVFQDYL LGSGSLIYNN AFIAFATDLD
D--------- ---------- ---IVWQDYL LGSGSVIYNN AFIAFATDLD
D--------- ---------- ---IVWQHFL LGSGSVIYNN AFIAFATDLD
D--------- ---------- ---IIWQDYL VGSGSVIYNN AFIAFANDLD
D--------- ---------- ---IVWQDFL VSHSSAVYNN AFIAFANDLD
E--------- ---------- ---LIFQFNV NIGPANSYLR YFISFANHHD
E--------- ---------- ---LIFQFNV NIGPANSYLR YFISFANHHD
D--------- ---------- ---LLFQYYV DLGPSSAYRR YFISFANHHD
---------- ---------- -DIVPHVPPQ SFGFLHPGVE SWIKSGTS-N
---------- ---------- -DIVPHVPPQ SFGFLHPGVE SWIKSGTS-N
---------- ---------- -DIVPHLPPA AFGFLHAGEE YWITDNSPET
---------- ---------- -DPVPKLPLL SMGYVHVSPE YWITSPNN--
---------- ---------- -DPVPKLPLL SMGYVHVSPE YWITSPNN--
---------- ---------- -DIVPRLPPR EFGYSHSSPE YWIKSGTL--
LTSLT----- -----KDIYD VPLVIYLANR PFTYWSNTST FKLTYDDNER
LTSLT----- -----ENIYD VPVVIYLANR PFTYFSNIST FKLKYSDTER
LEDLVAWHEN ---DKINVTD VPLVVYTSNT RMSYNSNFST FKLSYSDQEK
LTDLN----- --------YI PPLVVYIPNA RHSYNSNTST FKLSYTDDER
LTDLE----- --------YI PPLIVYIPNS RHSFNGNQST FKMSYSDSER
MTDLE----- --------YI PPLIVYIPNS RHSYNGNTST FKLSYSEKER
LTDLQ----- --------YI PPLVVYLPNA EYSFNSNQSA FKLSYSESQR
LTDLE----- --------SI PPLVVYMPNT RESFNSNTST FKMSYSTSER
LTDLE----- --------YI PPLIVYIPNS YHSFESNVST FKLNYNYSER
LTDLE----- --------YI PPLVVYIPNT KHSFNGNQST LKMNYNVTER
```

```
QTGPS----- ---------- -PLVVYIPNA PYSYHSNIST FQLSTDDAER
FTLNAN---- -------QKV PPLIVYVPNA PYTALSNVST FDPSYTMSQR
TTAGN----- ---HTVDNNT PPLLVYFPNY PWVYYSNIST FTMSMNDTLS
TTAGN----- ---HDVDNNT PPLLVYFPNY PWTYYSNIST FTMSMDDKMA
TTAGN----- ---HTVNDDT PPLVVYFPNY PWTMYANVTT YTVQLEDTLS
TTRGD----- ---VPVDHNT PPLVVYMPNT PWTMKSNLVD HRYRIANSEI
TTAGN----- ---HTVTRDT PPLVIYLPNV PYNYFTNIST DRTYYTEDMI
---------- ---------- TPLIIYVPSY PWSFAANTST YQLSYENDEA
---------- ---------- TPVIIYIPSY PWSFAANTST YQLSYENNEA
PRLSK----- ---------- HNLRYLCRVY PGPLRVGSTN FNPQVYWRLG
LSLDK----- ---------- HNRRYLMRVY PHVLRYKSSN FNPIPFWKAG
ISLDK----- ---------- HNRRYLMRVY PSGTRLKSSN FNPLPYWSHG
H--------- ---------- --AVARQYQD TLTKFVYGQD PWPVDKLHYV
D--------- ---------- ---VLYMFNS TKFNEHGDKL SRLFQSHFLR
LHTLSKEGDS GNNWSDREMI DSAMIELGYR EIFGCKFSPQ WKSGHGNSVD
HSADR----- ---------- SPLSVQADEA DEVHNADNSR FLRRVFSTSA
---------- ---------- ---LVTINSS KWGEYRGTLK DMDHLDVINW
E--------- ---------- ------QKVA AAALLAPAAA AIVAGPKQNC

GVNLGSGVNA D-KSISNKFF AYIRKYI--- ---------- ----------
GTPLGPGING DNITINSGLL GLVSSII--- ---------- ----------
RSKLGSGVTS DDVSVN-GLR SLYHT----- ---------- ----------
GSELGPGVTS DNITLD-GLT GFLGNFIDLK ---------- ----------
ASQLGPGVTS DNVTLS-GLT GFMGGLSKFK ---------- ----------
HLGLGPDVQK DQVSPE-GIK KLGSIEMRWL ---------- ----------
SNGLGSDIQK DKITPD-DLR KF--LLGGW- ---------- ----------
QKGLGPDIQK DQVNPD-GLK KISILV---- ---------- ----------
QQGLGPNIQK DQLEIK-SNL ---------- ---------- ----------
NNNRYFNNTT R--------- ---------- ---------- ----------
PNTAGLLVKW PEYTSSSQSG NNLMMINALG ---------- ----------
PNTAGLLVNW PKYTSSSQSG NNLMMINALG ---------- ----------
PNTAGLSVQW PKSTSSSQAG DNLMQISALG ---------- ----------
PNKAGLWTNW PTYTSSSQSG NNLMQINGLG ---------- ----------
PNKAGLLVNW PKYTSSSQSG NNLLQINALG ---------- ----------
PNVGTNLLQW DQYTDE---G KEMLEIHMTD ---------- ----------
PNVGTNLLQW DQYTDE---G KEMLEIHMTD ---------- ----------
PNVGTNLKQW DMYTDS---G KEMLQIHMIG ---------- ----------
VQICTSEIET KDCS------ --NSIVP--- ---------- ----------
VQICTSEIET KDCS------ --NSIVP--- ---------- ----------
VQVCTSDLET SDCS------ --NSIVP--- ---------- ----------
ATVSTSDIKV IDGDVSFDGN TGTGLPL--- ---------- ----------
ATVSTSDIKV IDGDVSFDGN TGTGLPL--- ---------- ----------
VPVTRNDIVK IEG---IDAT GGNNQPN--- ---------- ----------
QGMISNGFEI ATRSSGSLDD EWAACVG--- ---------- ----------
QGMISNGYDV ASRLNGKLDN EWAACVG--- ---------- ----------
FGAIRNGFET VTRNNLTDDE NWSTCVG--- ---------- ----------
LKMIKNGFEA ATRGNLTDDS SFMGCVA--- ---------- ----------
LGMIKNGFEA ATMGNFTDDS DFLGCVG--- ---------- ----------
LGVIRNGFEA ATMNNLTADS NFAGCIG--- ---------- ----------
RSMIQNGFEI ATRNNFTDDP EFMGCVG--- ---------- ----------
FKMIQNGFEA VTMKNLTKDE NFMGCIS--- ---------- ----------
VGMIRNAFEA TTRNNLTEDA DYVTCVG--- ---------- ----------
LGMIRNGFEA ATMGNFTDDS NFLGCIG--- ---------- ----------
DNIILNGYEV ATMANSTLDD NWTACVA--- ---------- ----------
NDIIGNGWNS ATQGNGTLDS EWPTCVA--- ---------- ----------
SGILENAALS ATQNN---SD SFAVCLA--- ---------- ----------
NGILENAFMS TTQNN---NE SFAVCLA--- ---------- ----------
SGMIENAAVA ATQNN---SD SFAVCVA--- ---------- ----------
QALIQNGFVA TTQDN---ST DFASCLA--- ---------- ----------
QQLLTNGLIS STVDN---DT YFGQCFA--- ---------- ----------
NEMLLNGMRS LTLNHS--VP TWPTCFA--- ---------- ----------
NEMLLNGMRS LTLNHS--VP TWPTCFA--- ---------- ----------
VQMVALNWQT YDTGLQINDA LFIADPP--- ----TGYLLK PP---------
VQMVATNWQT NDIGQQLNLA MFQILDHQPD GSFKSGYVLK PKKLLPVV--
VQMVATNWQT YDLGQQLNEA LFENKIF--- ----QGYVLK PSVLRKPT--
HDNQFEILDK SQFGDFRNVP ALKFLLG--- ---------- ----------
LAYGLEPWDH RNFGVYRNGG YQQLPLSELN ---------- ----------
DGSTQCGINE KEVGREDENV YEKFFNSVDY GKSSRKRTPL PKHNFASLGL
IDEDNENEHQ DDTEDQIHKE QQRRLSA--- ---------- ----------
KNKLQDDWSK FFRTTTVGEK VDILNFY--- ---------- ----------
EPDLMPYARP FAVGKRTCSG IVTP------ ---------- ----------
```

```
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- -----KSV-- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- -----LYTGK DNFRTAGYDA LFSNPPSFFV ----------
---------- -----LYTGK DNFRTAGYDA LMTNPSSFFV ----------
---------- -----LYTGK DNFRTAGYNA LFADPSHFFV ----------
---------- -----LYTGK DNFRPDAYSA LFSNPPSFFV ----------
---------- -----LYTGK DNFRTAGYDA LFTNPSSFFV ----------
---------- -----NVMRT DDYRIEGISN FETDVNLYG- ----------
---------- -----NVMRT DDYRIEGISN FETDVNLYG- ----------
---------- -----NSMRT DDFRIEGISN FESDVTLFG- ----------
---------- -----FTSIL DHLSYFDINE GSCL------ ----------
---------- -----FTSIL DHLSYFDINE GSCL------ ----------
---------- -----FTSVL DHLSYFGINT GLCT------ ----------
---------- -----LTDFE AHIWYFVQVD AGKGPGLPFK RV--------
---------- -----LTDFE AHIWYFVQVD AGKGPGLPFK RV--------
---------- -----IPDIP AHLWYFGLIG TCL------- ----------
---------- -----CAIIR REQERQGIEQ TEQCKRCFEN YCWDG-----
---------- -----CAIIR REQERLGIEQ TEQCKKCFEN YCWDG-----
---------- -----CAIIR RQQERLGEEQ SDECKKCFQE YCWTGGFKDA
---------- -----CAVMR RKQQSLNATL PEECSTCFTN YCWNG-----
---------- -----CAIIR RKQQNLNATL PSECSQCFTN YCWNG-----
---------- -----CAIMR RKQQALNLTL PKECETCFTN YCWNG-----
---------- -----CAIIR RKQQALNITL PPECETCFKN YCWNG-----
---------- -----CAILR RKQESLNYTL PSECDACFEK YCWNG-----
---------- -----CAIIR RKQQSLNLTL PDICDKCFTN YCWNG-----
---------- -----CAIIR RKQESLNATL PPECTKCFAD YCWNG-----
---------- -----CAILS RSFERTGTTL PDICSQCFDR YCWNG-----
---------- -----CAVIS RSLDRLGRQT PAACKTCFER YCWNG-----
---------- -----CAMIQ RSLERKNMST PSQCSSCFEQ YCWNG-----
---------- -----CAIIQ RSLERKKLST PTQCSSCFQE YCWDG-----
---------- -----CALVQ RSLERKNMST PSQCASCFNQ YCWNG-----
---------- -----CAVVQ RSLERRNQST SAACQQCFSQ YCWNG-----
---------- -----CAVVK RTLERNNITA SPECQQCYYN YCWSG-----
---------- -----CALTD RSFMYTSENR STTCQKCFDT WCWAG-----
---------- -----CALTD RSFMYTSENR STTCQECFDT WCWAG-----
-----CQRII GTTVGEEGLP RKIKLTIDVI SGQQLRRARE LSNS-ETLSP
-TK--AKMIP LIYEHFENGS DPVTVKIRIL STQLLPRLND TSPSRNNTNS
-LKSSSSNVD TRTSLTTTNS KTIRFNFEII SGHQLPKFPK DDYKDQAINP
---------- -------FSA EELGELTKKY TGEGHYTL-- ----------
---------- --KVRPVERY EALSKMDFGQ VGRLSNALSR L---------
TFNHRAGIEN VGIRDHKVLS TDPRLRKNDE HKKEVDGYGP DCWKKESNKK
---------- -----YLESS KDLRQLDANS STTALDALNK E---------
---------- ------LKIT DDLARKGF-- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
```

```
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
----TIYKGE PLGEN----- ---------- ---------- ---------F
----TIYKGE PLGDN----- ---------- ---------- ---------F
ASVSSVSGIS GLAAK----- ---------- ---------- ---------T
----TIDDTP VSGLD----- ---------- ---------- ------NSDF
----TIDSRS VSGVG----- ---------- ---------- ------NDDY
----TIDNTP AKGVT----- ---------- ---------- ------ASND
----TLDTTP LPDVEKD--- ---------- ---------- VHHSFINVNS
----TVDAT- ---------- ---------- ---------- ----------
----TIDNTP TKLLTPN-NQ DPAAISSAIA AVTDDSPIGA LLNTGSG--T
----TLSTSA NPELSGNSTY QSGAIASAIS EATDGIPITA LLGSSTSGNT
----TVNST- ---------- ---------- ---------- ----------
----TVNSK- ---------- ---------- ---------- ----------
----TTVNN- ---------- ---------- ---------- ----------
----TLAT-- ---------- ---------- ---------- ----------
----TIAS-- ---------- ---------- ---------- ----------
----TVDNTP V--------- ---------- ---------- ----------
----LYDDS- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
YVEIQVHSME E--------- --------SP FRWCSKVVHE NGFRPFWGET
FVKVEFHTDD EPTM----PI SIDKGTRISA TEASTKSSQG NGFNPIWDAE
YISFEIIGAQ DVQWDNNDSS PIAPTTSSSP FIRTTKIIRE NGFNPNFNTK
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
FKADATEQLK EWALNSLASK VLDDKEMIKS EIPEGFSNYL PNEKDLEMYL
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
SDDGLTNSAT EYNSNNVAGF ND----GGTS ILKKA----- ----------
SDEGLTTSAA YYNSNNVAGI ND----GGIA LVKRDDLSN- ----------
HTSGGTSSTT QQTSTTTGSS AN----GGSS STGSSSSSKK KNGGDLVNGG
DPTAASSAYS AYNTESYSSS S-----ATGS --KKNGAG-- ----LPATPT
SSSASLSASA AAASASASAS ASASASASGS STHKKNAGNA LVNYSNLNTN
FDNASGSAAA DMAEQDASGA AS----ASSS SRKKNAAVS- ----VDVNAK
FNSSIGQEES LYAGSSASQS SSSSSSSSSS SEIPSATATL EKKAATNSGS
------TPIS STTSSSASST S-----TSDS GNKENSAR-- ----ILAPRS
KSNSSSKTNS TLVTSSRATS TGTLISNSSS NSTVSSTAAR SSTSSTAKKN
TSNSTTSTSS NVTSNSNSSS NTTLNSNSSS -SSISSSTAR SSSSTANKAN
------RPES YDPAFYLADN -------SMA SVSLPTML-- ----------
```

```
------DTGV YMPEFKIADA HA----LDSG AVAIGKMVN- ----------
----PSAVSN YAPTVLSAST TS----GTSS ---------- -VRAKPIVFY
-----STASV YDPTVMSAAT TSRAPSGTTS GTASSTTSSS VASATPTHKH
-----TTVTT YAPTVLSAKI YK-------- ---------- ----------
--DDDSKNPT YNPAVKTSSA SG-------- ---------V HANILLSFFV
---AANDDIV YNPTCRLGEG I--------- ---------- ----------
-DDNTTEPAT YEPVINSVPP WLVANNLSIG VADAPASNES TAGTASSGAA
-DDNTTEPAN YEPVINSVPP WLIANNLSIG MADAPGSNES TAGTASSGAA
MVYESIISDD FYSMIRFLVH HRGSN----- GNDSIFANFT CPIDRLQQGY
VSIT--LKDT DLTFIKFMVI S--------- -EETQIASVC LKLNYLRMGY
FSGSIITTTN DLIFIKFVVY ASTSLNYPDY GENFPIAILV TKLNYLKQGY
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
TDKTVTNRNK WSMLKRICYL QYLSHKLDER KTQRLKKIKD MRRHLSSSTE
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
VPSSIFLVFN SLLGLIIAYL ---------- ---------- ----------
SFTSILTLLT AIAGFL---- ---------- ---------- ----------
TFIGVLSVIS AVFGLI---- ---------- ---------- ----------
TLFAIITAMT AVFQLI---- ---------- ---------- ----------
HLSGISVKFS AMIMLTLLMF TGAV------ ---------- ----------
TLSLLIGGLA SVFISF---- ---------- ---------- ----------
AGSVLKLEFS KSASVMVAIA AAAVASLI-- ---------- ----------
AAAISYANTN TLMSLLGAIT --ALFGLI-- ---------- ----------
--STVVAAGL AMLILV---- ---------- ---------- ----------
VWSSVVVGVV AATLLL---- ---------- ---------- ----------
LFASLLTVSL LL-------- ---------- ---------- ----------
WWDSIFEAKE NP-------- ---------- ---------- ----------
--SRLFTYCS ---------- ---------- ---------- ----------
LLATLLVTA- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
NADVSMGMVA LAAGLGLML- ---------- ---------- ----------
KMGVGMGMVA LTAGLGLML- ---------- ---------- ----------
RHIRLLDMQG ENLLFSSLFL KIKKEDI--- ---------- ----------
RHIPLFNMEG EQYIFCTLFI HTQIL----- ---------- ----------
RYIYLNDLLG EQLVYSSIFI KIEYDEDLLN EFINK----- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
STRNGSNSLP LNEKSNEGES TNVDQDIEGD EYHRLHEDIL KNQELDDGSL
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
```

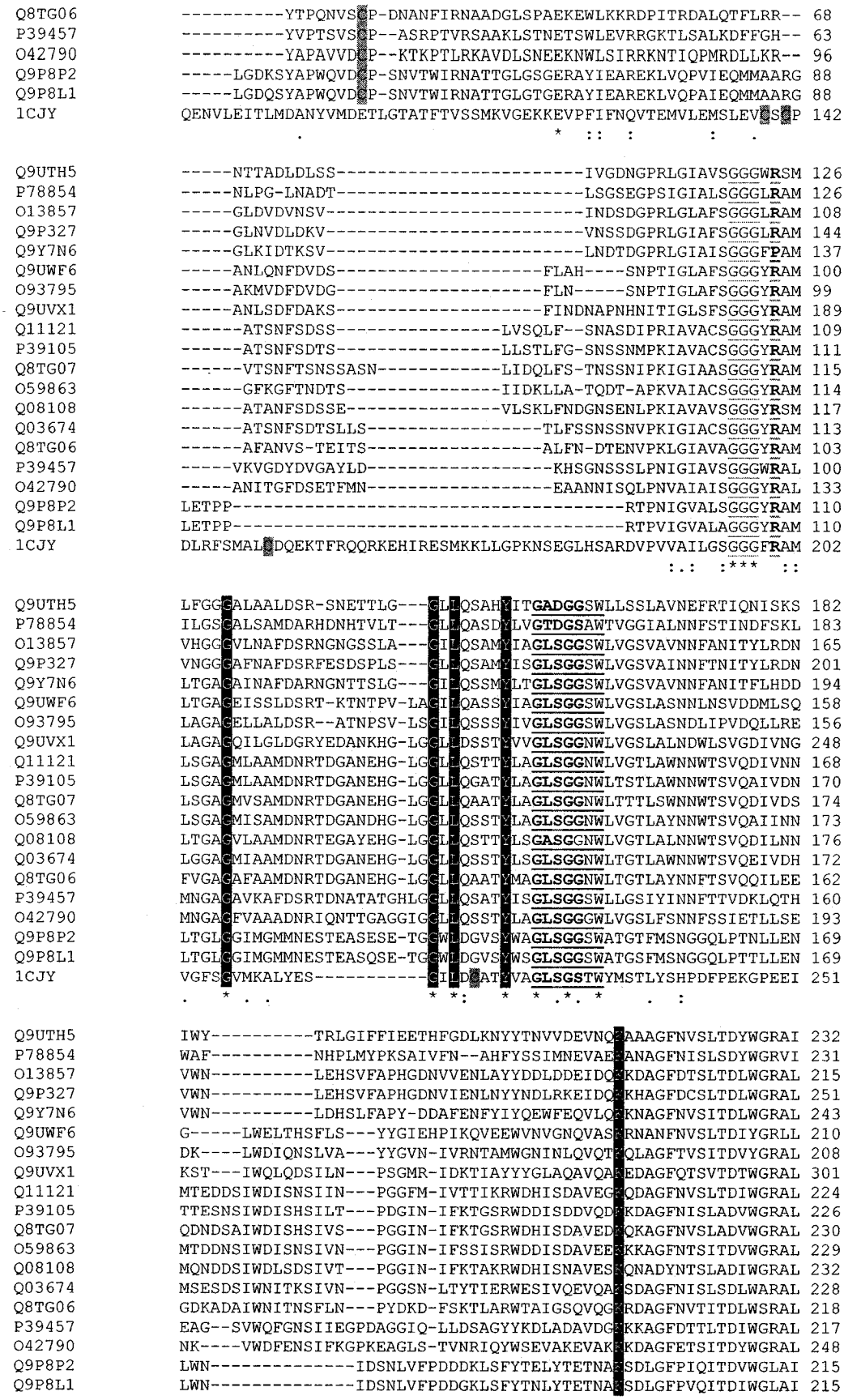
DDLLSQIIPK ITNFNSGEID DAKKEELLKL NFIDPYSFED PLISSFSEGL

```
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
WFTIALRNTL LYKLVFHCQP DNAVQNWKEY GEFTELEQEF QINQEKLIDL
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
```

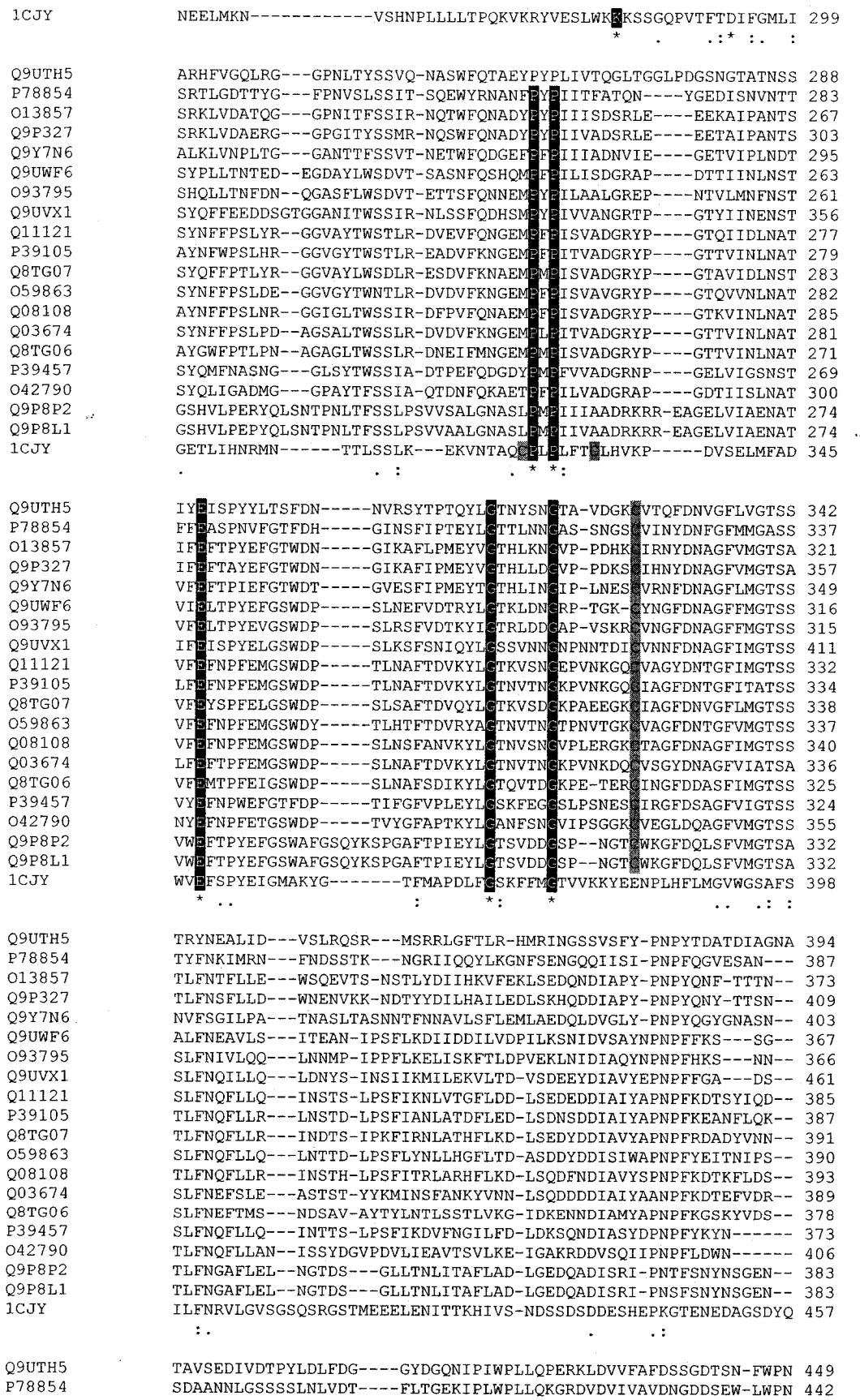
```
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
EAENINSTTT NVVDKDREKE KMRKAAELRM KLSGSLLYGF NQKVFDKHTA
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------
---------- ---------- ---------- ---------- ----------

---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
QRILERIHGH LVIFPTEWLA KEVESRNWIF NSDRLSPMEI YN
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
---------- ---------- ---------- ---------- --
```

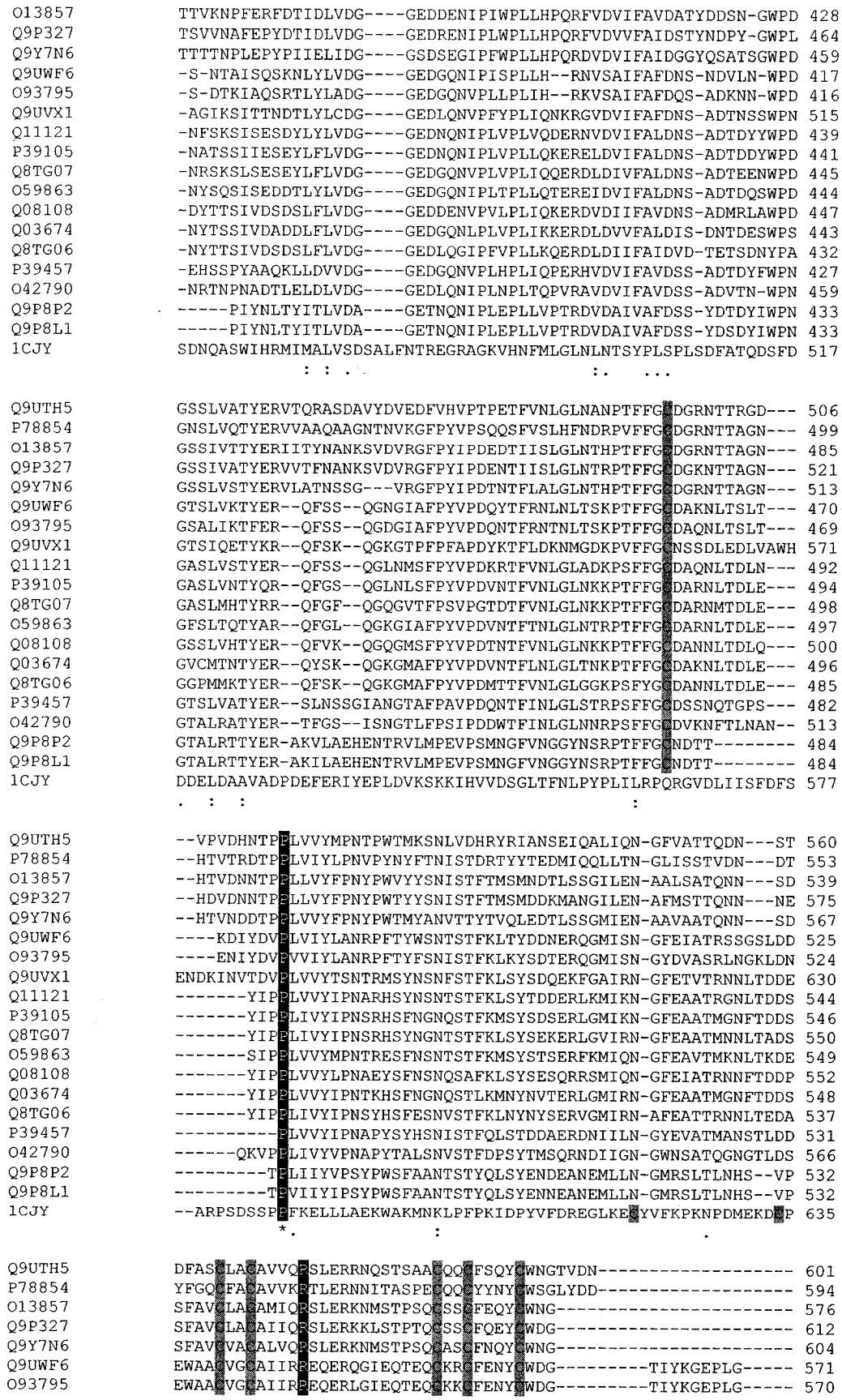
Appendix 2: Multiple sequence alignment of lysophopholipase from fungi with comparison to protein sequences of cytosolic phospholipase A2 from human. The conserved **oxyanion hole (GGG)** is shown in green. The dyad active sites (Ser- Asp) are shown in red. Eight conserved cysteine residues in lysophopholipases were high-lighted in green and were not conserved in cytosolic phospholipase A2. Conserved residues between lysophopholipase and cytosolic phospholipase A2 are shown in black highlights.

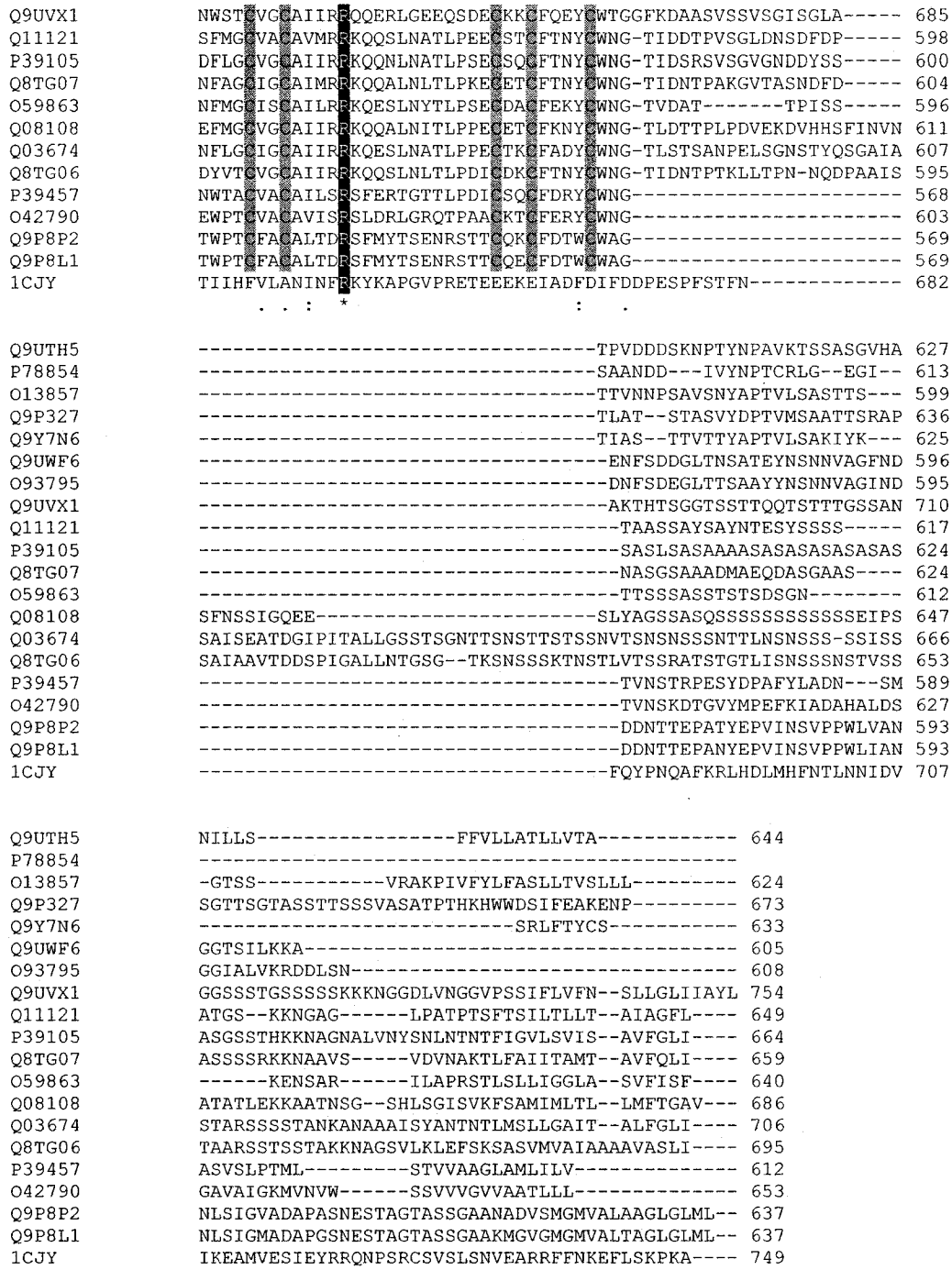
```
Q9UTH5          ---------------------------------MYFQSFYFLALLLATAVYGQVAS---- 23
P78854          ---------------------------------MLFRGLSLWMLFLASCLSALALP---- 23
O13857          ------------------------------------------------------MRPG---- 4
Q9P327          ---------------------------------MYVNYIGLFAFVQISLTLAYPPGRVEI 27
Q9Y7N6          ---------------------------------MKLSSFGLFLALQLLPALGLPSR---- 23
Q9UWF6          ---------------------------------MILHHLLILLIIINYCVATSPTN---- 22
O93795          ---------------------------------MLVWQSILLFLVGCVLSKSPTN---- 22
Q9UVX1          ----------------------MKVNLKLIIGSILISQAQAIWPFDSSGSSSSSDSSPS 37
Q11121          --------------------------MNLKEW-LLFSDAVF--FAQGTLAWSPSN---- 26
P39105          --------------------------MKLQS--LLVSAAVLTSLTENVNAWSPNN---- 27
Q8TG07          --------------------------MQLQD--LVVTVSLLAAFNGGVEAWSPTN---- 27
O59863          ------------------------MWFLNSVNLLFLVCSVALHLDAVNAWSPTN---- 30
Q08108          ------------------------MIRPLCSKIIISYIFAISQFLLAANAWSPTD---- 31
Q03674          --------------------------MQLRN-ILQASSLISGLSLAADSSSTTGD---- 28
Q8TG06          --------------------------MQLS--VLIASVLAAGAAVDAAS---------- 21
P39457          ---------------------------------DITFAG-----VQRALPNAPDG---- 17
O42790          -------MHLPSSLLIAAPLLANVSAEPIRIPQRDVSVVSTSQQLAVRALPDSPSGG--- 50
Q9P8P2          ----------------------MSIATGTFAFSLFATIAFAVPPETPRIELQAERG--- 34
Q9P8L1          ----------------------MSIITTAFALSLLATTAFAVPPETPRIELQAERG--- 34
1CJY            MSFIDPYQHIIVEHQYSHKFTVVVLRATKVTKGAFGDMLDTPDPYVELFISTTPDSR--- 57

Q9UTH5          ------------------------------------PELHSLSRRNWK--------KPP 38
P78854          ------------------------------------AAEDDGSVKVFK--------RAK 38
O13857          -----------------------------------MHDTPLSLMQK-----REALAI 21
Q9P327          -------------------------SEIYDFEESSSYKGQDIDTSVLYT-----LSKRKP 57
Q9Y7N6          ---------------------------------IDEVDVSDPELIGLLKPDNVDKPANSI 50
Q9UWF6          ------------------------------------------------------------
O93795          ------------------------------------------------------------
Q9UVX1          ETGSSGGTFPFDLFGSGSSLTQSSSAQASSTKSTSDSASSTDSSLFSSSNSGSSWYQTFL 97
Q11121          ------------------------------------------------------------
P39105          ------------------------------------------------------------
Q8TG07          ------------------------------------------------------------
O59863          ------------------------------------------------------------
Q08108          ------------------------------------------------------------
Q03674          ------------------------------------------------------------
Q8TG06          ------------------------------------------------------------
P39457          ------------------------------------------------------------
O42790          ------------------------------------------------------------
Q9P8P2          ------------------------------------------------------------
Q9P8L1          ------------------------------------------------------------
1CJY            ---------------------------------KRTRHFNNDINPVWNETFEFILDPN 82

Q9UTH5          PFPSTN--ASYAPVIRSCDSSEIMVNSLPRGELPDLENDFIEKRLSNANEALTTFLQSK- 95
P78854          KHSTKQEGPSYAPYYVDCPS-DNIVESLSSNEIPSAESEYLSTRSTITNTAMKDFLRNA- 96
O13857          SLSKRDSVGSYAPYNVTCPS-DYMLRPASDG-ISSGEQSFIDKRIPKINTQMRSFISNT- 78
Q9P327          ALVKRSTDASYAPFNVTCSN-DNLLRPASEG-LNEGEQSYINKRISKVNSELRSFISKT- 114
Q9Y7N6          PLSKRSTSPSYAPYTVACPS-GSLLRPASDG-LSTGEQEFVDKRVSKVNSALESFISKT- 107
Q9UWF6          ---------GYAPGPVSCP--SSQLIRSGSQGINPNEQSYINARYPIAKQALSKFLHN-- 69
O93795          ---------LYTPGYVQCP--EGKLTRSSLDGINSNEKAYIDRRYANAKSELSRFLHN-- 69
Q9UVX1          DGDSGDQKTDYAPFNLTCP--SKKTFIRTASELSQQEKDYIHKRQETTNKNLIDFLSKR- 154
Q11121          ---------SYTPANVSCD-EDINLIRQAS-GPSDNETEWLKKRDVYTREALRSFLDR-- 73
P39105          ---------SYVPANVTCD-DDINLVREAS-GLSDNETEWLKKRDAYTKEALHSFLNR-- 74
Q8TG07          ---------SYVPANVTCP-NDINLLRNAT-GLSQSEIDWLKKRDVNTREALESFLKR-- 74
O59863          ---------GYAPGVVDCD-ENINLVRKAD-AVSDDEADWLKVRHESTVPALKDFLQR-- 77
Q08108          ---------SYVPGTVSCP-DDINLVREAT-SISQNESAWLEKRNKVTSVALKDFLTR-- 78
Q03674          ---------GYAPSIIPCPSDDTSLVRNAS-GLSTAETDWLKKRDAYTKEALHSFLSR-- 76
```



```
Q8TG06          ----------YTPQNVSCP-DNANFIRNAADGLSPAEKEWLKKRDPITRDALQTFLRR-- 68
P39457          ----------YVPTSVSCP--ASRPTVRSAAKLSTNETSWLEVRRGKTLSALKDFFGH-- 63
O42790          ----------YAPAVVDCP--KTKPTLRKAVDLSNEEKNWLSIRRKNTIQPMRDLLKR-- 96
Q9P8P2          -----LGDKSYAPWQVDCP-SNVTWIRNATTGLGSGERAYIEAREKLVQPVIEQMMAARG 88
Q9P8L1          -----LGDQSYAPWQVDCP-SNVTWIRNATTGLGTGERAYIEAREKLVQPAIEQMMAARG 88
1CJY            QENVLEITLMDANYVMDETLGTATFTVSSMKVGEKKEVPFIFNQVTEMVLEMSLEVCSCP 142
                          .                              *  ::  :       :   .

Q9UTH5          -----NTTADLDLSS-----------------------IVGDNGPRLGIAVSGGGWRSM 126
P78854          -----NLPG-LNADT-----------------------LSGSEGPSIGIALSGGGLRAM 126
O13857          -----GLDVDVNSV------------------------INDSDGPRLGLAFSGGGLRAM 108
Q9P327          -----GLNVDLDKV------------------------VNSSDGPRLGIAFSGGGLRAM 144
Q9Y7N6          -----GLKIDTKSV------------------------LNDTDGPRLGIAISGGGFPAM 137
Q9UWF6          -----ANLQNFDVDS--------------------FLAH----SNPTIGLAFSGGGYRAM 100
O93795          -----AKMVDFDVDG--------------------FLN-----SNPTIGLAFSGGGYRAM 99
Q9UVX1          -----ANLSDFDAKS--------------------FINDNAPNHNITIGLSFSGGGYRAM 189
Q11121          ------ATSNFSDSS-----------------LVSQLF--SNASDIPRIAVACSGGGYRAM 109
P39105          ------ATSNFSDTS-----------------LLSTLFG-SNSSNMPKIAVACSGGGYRAM 111
Q8TG07          ------VTSNFTSNSSASN-------------LIDQLFS-TNSSNIPKIGIAASGGGYRAM 115
O59863          ------GFKGFTNDTS----------------IIDKLLA-TQDT-APKVAIACSGGGYRAM 114
Q08108          ------ATANFSDSSE----------------VLSKLFNDGNSENLPKIAVAVSGGGYRSM 117
Q03674          ------ATSNFSDTSLLS-----------------TLFSSNSSNVPKIGIACSGGGYRAM 113
Q8TG06          ------AFANVS-TEITS-----------------ALFN-DTENVPKLGIAVAGGGYRAM 103
P39457          -----VKVGDYDVGAYLD-----------------KHSGNSSSLPNIGIAVSGGGWRAL 100
O42790          -----ANITGFDSETFMN-----------------EAANNISQLPNVAIAISGGGYRAL 133
Q9P8P2          LETPP-------------------------------------RTPNIGVALSGGGYRAM 110
Q9P8L1          LETPP---------------------------------------RTPVIGVALAGGGYRAM 110
1CJY            DLRFSMALCDQEKTFRQQRKEHIRESMKKLLGPKNSEGLHSARDVPVVAILGSGGGFRAM 202
                                                            :.:  :***  ::

Q9UTH5          LFGGCALAALDSR-SNETTLG---GLLQSAHYITGADGGSWLLSSLAVNEFRTIQNISKS 182
P78854          ILGSCALSAMDARHDNHTVLT---GLLQASDYLVGTDGSAWTVGGIALNNFSTINDFSKL 183
O13857          VHGGCVLNAFDSRNGNGSSLA---GILQSAMYIAGLSGGSWLVGSVAVNNFANITYLRDN 165
Q9P327          VNGGCAFNAFDSRFESDSPLS---GLLQSAMYISGLSGGSWLVGSVAINNFTNITYLRDN 201
Q9Y7N6          LTGACAINAFDARNGNTTSLG---GILQSSMYLTGLSGGSWLVGSVAVNNFANITFLHDD 194
Q9UWF6          LTGAGEISSLDSRT-KTNTPV-LAGILQASSYIAGLSGGSWLVGSLASNNLNSVDDMLSQ 158
O93795          LAGAGELLALDSR--ATNPSV-LSGILQSSSYIVGLSGGSWLVGSLASNDLIPVDQLLRE 156
Q9UVX1          LAGACQILGLDGRYEDANKHG-LGGLLDSSTYVVGLSGGNWLVGSLALNDWLSVGDIVNG 248
Q11121          LSGAGMLAAMDNRTDGANEHG-LGGLLQSTTYLAGLSGGNWLVGTLAWNNWTSVQDIVNN 168
P39105          LSGAGMLAAMDNRTDGANEHG-LGGLLQGATYLAGLSGGNWLTSTLAWNNWTSVQAIVDN 170
Q8TG07          LSGAGMVSAMDNRTDGANEHG-LGGLLQAATYLAGLSGGNWLTTTLSWNNWTSVQDIVDS 174
O59863          LSGAGMISAMDNRTDGANDHG-LGGLLQSSTYLAGLSGGNWLVGTLAYNNWTSVQAIINN 173
Q08108          LTGAGVLAAMDNRTEGAYEHG-LGGLLQSTTYLSGASGGNWLVGTLALNNWTSVQDILNN 176
Q03674          LGGAGMIAAMDNRTDGANEHG-LGGLLQSSTYLSGLSGGNWLTGTLAWNNWTSVQEIVDH 172
Q8TG06          FVGAGAFAAMDNRTDGANEHG-LGGLLQAATYMAGLSGGNWLTGTLAYNNFTSVQQILEE 162
P39457          MNGAGAVKAFDSRTDNATATGHLGGLLQSATYISGLSGGSWLLGSIYINNFTTVDKLQTH 160
O42790          MNGAGFVAAADNRIQNTTGAGGIGGLLQSSTYLAGLSGGGWLVGSLFSNNFSSIETLLSE 193
Q9P8P2          LTGLGGIMGMMNESTEASESE-TGGWLDGVSYWAGLSGGSWATGTFMSNGGQLPTNLLEN 169
Q9P8L1          LTGLGGIMGMMNESTEASQSE-TGGWLDGVSYWSGLSGGSWATGSFMSNGGQLPTTLLEN 169
1CJY            VGFSGVMKALYES----------GILDGATYVAGLSGSTWYMSTLYSHPDFPEKGPEEI 251
                .   * . .                * *:   *  * .*. *    .  :

Q9UTH5          IWY----------TRLGIFFIEETHFGDLKNYYTNVVDEVNQWAAAGFNVSLTDYWGRAI 232
P78854          WAF----------NHPLMYPKSAIVFN--AHFYSSIMNEVAEWANAGFNISLSDYWGRVI 231
O13857          VWN----------LEHSVFAPHGDNVVENLAYYDDLDDEIDQWKDAGFDTSLTDLWGRAL 215
Q9P327          VWN----------LEHSVFAPHGDNVIENLNYYNDLRKEIDQWKHAGFDCSLTDLWGRAL 251
Q9Y7N6          VWN----------LDHSLFAPY-DDAFENFYIYQEWFEQVLQWKNAGFNVSITDLWGRAL 243
Q9UWF6          G-----LWELTHSFLS---YYGIEHPIKQVEEWVNVGNQVASWRNANFNVSLTDIYGRLL 210
O93795          DK----LWDIQNSLVA---YYGVN-IVRNTAMWGNINLQVQTWQLAGFTVSITDVYGRAL 208
Q9UVX1          KST---IWQLQDSILN---PSGMR-IDKTIAYYYGLAQAVQAWEDAGFQTSVTDTWGRAL 301
Q11121          MTEDDSIWDISNSIIN---PGGFM-IVTTIKRWDHISDAVEGWQDAGFNVSLTDIWGRAL 224
P39105          TTESNSIWDISHSILT---PDGIN-IFKTGSRWDDISDDVQDWKDAGFNISLADVWGRAL 226
Q8TG07          QDNDSAIWDISHSIVS---PGGIN-IFKTGSRWDHISDAVEDWQKAGFNVSLADVWGRAL 230
O59863          MTDDNSIWDISNSIVN---PGGIN-IFSSISRWDDISDAVEEWKKAGFNTSITDVWGRAL 229
Q08108          MQNDDSIWDLSDSIVT---PGGIN-IFKTAKRWDHISNAVESWQNADYNTSLADIWGRAL 232
Q03674          MSESDSIWNITKSIVN---PGGSN-LTYTIERWESIVQEVQAWSDAGFNISLSDLWARAL 228
Q8TG06          GDKADAIWNITNSFLN---PYDKD-FSKTLARWTAIGSQVQGWRDAGFNVTITDLWSRAL 218
P39457          EAG--SVWQFGNSIIEGPDAGGIQ-LLDSAGYYKDLADAVDGWKKAGFDTTLTDIWGRAL 217
O42790          NK----VWDFENSIFKGPKEAGLS-TVNRIQYWSEVAKEVAKWKDAGFETSITDYWGRAL 248
Q9P8P2          LWN--------------IDSNLVFPDDDKLSFYTELYTETNAWSDLGFPIQITDVWGLAI 215
Q9P8L1          LWN--------------IDSNLVFPDDGKLSFYTNLYTETNAWSDLGFPVQITDIWGLAI 215
```


1CJY NEELMKN----------VSHNPLLLLTPQKVKRYVESLWKKKSSGQPVTFTDIFGMLI 299
Q9UTH5 ARHFVGQLRG---GPNLTYSSVQ-NASWFQTAEYPYPLIVTQGLTGGLPDGSNGTATNSS 288
P78854 SRTLGDTTYG---FPNVSLSSIT-SQEWYRNANFPYPIITFATQN----YGEDISNVNTT 283
O13857 SRKLVDATQG---GPNITFSSIR-NQTWFQNADYPYPIIISDSRLE----EEKAIPANTS 267
Q9P327 SRKLVDAERG---GPGITYSSMR-NQSWFQNADYPYPIIVADSRLE----EETAIPANTS 303
Q9Y7N6 ALKLVNPLTG---GANTTFSSVT-NETWFQDGEFPFPIIIADNVIE----GETVIPLNDT 295
Q9UWF6 SYPLLTNTED--EGDAYLWSDVT-SASNFQSHQMPFPILISDGRAP----DTTIINLNST 263
O93795 SHQLLTNFDN--QGASFLWSDVT-ETTSFQNNEMPYPILAALGREP----NTVLMNFNST 261
Q9UVX1 SYQFFEEDDSGTGGANITWSSIR-NLSSFQDHSMPYPIVVANGRTP----GTYIINENST 356
Q11121 SYNFFPSLYR--GGVAYTWSTLR-DVEVFQNGEMPFPISVADGRYP----GTQIIDLNAT 277
P39105 AYNFWPSLHR--GGVGYTWSTLR-EADVFKNGEMPFPITVADGRYP----GTTVINLNAT 279
Q8TG07 SYQFFPTLYR--GGVAYLWSDLR-ESDVFKNAEMPMPISVADGRYP----GTAVIDLNST 283
O59863 SYNFFPSLDE--GGVGYTWNTLR-DVDVFKNGEMPFPISVAVGRYP----GTQVVNLNAT 282
Q08108 AYNFFPSLNR--GGIGLTWSSIR-DFPVFQNAEMPFPISVADGRYP----GTKVINLNAT 285
Q03674 SYNFFPSLPD--AGSALTWSSLR-DVDVFKNGEMPLPITVADGRYP----GTTVINLNAT 281
Q8TG06 AYGWFPTLPN--AGAGLTWSSLR-DNEIFMNGEMPMPISVADGRYP----GTTVINLNAT 271
P39457 SYQMFNASNG---GLSYTWSSIA-DTPEFQDGDYPMPFVVADGRNP----GELVIGSNST 269
O42790 SYQLIGADMG---GPAYTFSSIA-QTDNFQKAETPFPILVADGRAP----GDTIISLNAT 300
Q9P8P2 GSHVLPERYQLSNTPNLTFSSLPSVVSALGNASLPMPIIIAADRKRR-EAGELVIAENAT 274
Q9P8L1 GSHVLPEPYQLSNTPNLTFSSLPSVVAALGNASLPMPIIVAADRKRR-EAGELVIAENAT 274
1CJY GETLIHNRMN------TTLSSLK---EKVNTAQCPLPLFTCLHVKP-----DVSELMFAD 345
Q9UTH5 IYPISPYYLTSFDN-----NVRSYTPTQYLGTNYSNGTA-VDGKGVTQFDNVGFLVGTSS 342
P78854 FFPASPNVFGTFDH-----GINSFIPTEYLGTTLNNGAS-SNGSGVINYDNFGFMMGASS 337
O13857 IFPFTPYEFGTWDN-----GIKAFLPMEYVGTHLKNGVP-PDHKGIRNYDNAGFVMGTSA 321
Q9P327 IFPFTAYEFGTWDN-----GIKAFIPMEYVGTHLLDGVP-PDKSGIHNYDNAGFVMGTSA 357
Q9Y7N6 VFPFTPIEFGTWDT-----GVESFIPMEYTGTHLINGIP-LNESGVRNFDNAGFLMGTSS 349
Q9UWF6 VIPLTPYEFGSWDP-----SLNEFVDTRYLGTKLDNGRP-TGK-GYNGFDNAGFFMGTSS 316
O93795 VFPLTPYEVGSWDP-----SLRSFVDTKYIGTRLDDGAP-VSKRGVNGFDNAGFFMGTSS 315
Q9UVX1 IFPISPYELGSWDP-----SLKSFSNIQYLGSSVNNGNPNNTDIGVNNFDNAGFIMGTSS 411
Q11121 VFPFNPFEMGSWDP-----TLNAFTDVKYLGTKVSNGEPVNKGQGVAGYDNTGFIMGTSS 332
P39105 LFPFNPFEMGSWDP-----TLNAFTDVKYLGTNVTNGKPVNKGQGIAGFDNTGFITATSS 334
Q8TG07 VFPYSPFELGSWDP-----SLSAFTDVQYLGTKVSDGKPAEEGKGIAGFDNVGFLMGTSS 338
O59863 VFPFNPFEMGSWDY-----TLHTFTDVRYAGTNVTNGTPNVTGKGVAGFDNTGFVMGTSS 337
Q08108 VFPFNPFEMGSWDP-----SLNSFANVKYLGTNVSNGVPLERGKGTAGFDNAGFIMGTSS 340
Q03674 LFPFTPFEMGSWDP-----SLNAFTDVKYLGTNVTNGKPVNKDQGVSGYDNAGFVIATSA 336
Q8TG06 VFPMTPFEIGSWDP-----SLNAFSDIKYLGTQVTDGKPE-TERGINGFDDASFIMGTSS 325
P39457 VYPFNPWEFGTFDP-----TIFGFVPLEYLGSKFEGGSLPSNESGIRGFDSAGFVIGTSS 324
O42790 NYPFNPFETGSWDP-----TVYGFAPTKYLGANFSNGVIPSGGKGVEGLDQAGFVMGTSS 355
Q9P8P2 VWPFTPYEFGSWAFGSQYKSPGAFTPIEYLGTSVDDGSP--NGTGWKGFDQLSFVMGTSA 332
Q9P8L1 VWPFTPYEFGSWAFGSQYKSPGAFTPIEYLGTSVDDGSP--NGTGWKGFDQLSFVMGTSA 332
1CJY WVPFSPYEIGMAKYG-------TFMAPDLFGSKFFMGTVVKKYEENPLHFLMGVWGSAFS 398
Q9UTH5 TRYNEALID---VSLRQSR---MSRRLGFTLR-HMRINGSSVSFY-PNPYTDATDIAGNA 394
P78854 TYFNKIMRN---FNDSSTK---NGRIIQQYLKGNFSENGQQIISI-PNPFQGVESAN--- 387
O13857 TLFNTFLLE---WSQEVTS-NSTLYDIIHKVFEKLSEDQNDIAPY-PNPYQNF-TTTN-- 373
Q9P327 TLFNSFLLD---WNENVKK-NDTYYDILHAILEDLSKHQDDIAPY-PNPYQNY-TTSN-- 409
Q9Y7N6 NVFSGILPA---TNASLTASNNTFNNAVLSFLEMLAEDQLDVGLY-PNPYQGYGNASN-- 403
Q9UWF6 ALFNEAVLS---ITEAN-IPSFLKDIIDDILVDPILKSNIDVSAYNPNPFFKS---SG-- 367
O93795 SLFNIVLQQ---LNNMP-IPPFLKELISKFTLDPVEKLNIDIAQYNPNPFHKS---NN-- 366
Q9UVX1 SLFNQILLQ---LDNYS-INSIIKMILEKVLTD-VSDEEYDIAVYEPNPFFGA---DS-- 461
Q11121 SLFNQFLLQ---INSTS-LPSFIKNLVTGFLDD-LSEDEDDIAIYAPNPFKDTSYIQD-- 385
P39105 TLFNQFLLR---LNSTD-LPSFIANLATDFLED-LSDNSDDIAIYAPNPFKEANFLQK-- 387
Q8TG07 TLFNQFLLR---INDTS-IPKFIRNLATHFLKD-LSEDYDDIAVYAPNPFRDADYVNN-- 391
O59863 SLFNQFLLQ---LNTTD-LPSFLYNLLHGFLTD-ASDDYDDISIWAPNPFYEITNIPS-- 390
Q08108 TLFNQFLLR---INSTH-LPSFITRLARHFLKD-LSQDFNDIAVYSPNPFKDTKFLDS-- 393
Q03674 SLFNEFSLE---ASTST-YYKMINSFANKYVNN-LSQDDDDIAIYAANPFKDTEFVDR-- 389
Q8TG06 SLFNEFTMS---NDSAV-AYTYLNTLSSTLVKG-IDKENNDIAMYAPNPFKGSKYVDS-- 378
P39457 SLFNQFLLQ---INTTS-LPSFIKDVFNGILFD-LDKSQNDIASYDPNPFYKYN------ 373
O42790 TLFNQFLLAN--ISSYDGVPDVLIEAVTSVLKE-IGAKRDDVSQIIPNPFLDWN------ 406
Q9P8P2 TLFNGAFLEL--NGTDS---GLLTNLITAFLAD-LGEDQADISRI-PNTFSNYNSGEN-- 383
Q9P8L1 TLFNGAFLEL--NGTDS---GLLTNLITAFLAD-LGEDQADISRI-PNSFSNYNSGEN-- 383
1CJY ILFNRVLGVSGSQSRGSTMEEELENITTKHIVS-NDSSDSDDESHEPKGTENEDAGSDYQ 457
Q9UTH5 TAVSEDIVDTPYLDLFDG----GYDGQNIPIWPLLQPERKLDVVFAFDSSGDTSN-FWPN 449
P78854 SDAANNLGSSSSLNLVDT----FLTGEKIPLWPLLQKGRDVDVIVAVDNGDDSEW-LWPN 442




Q9UVX1 NWSTCVGCAIIRRQQERLGEEQSDECKKCFQEYCWTGGFKDAASVSSVSGISGLA----- 685
Q11121 SFMGCVACAVMRRKQQSLNATLPEECSTCFTNYCWNG-TIDDTPVSGLDNSDFDP----- 598
P39105 DFLGCVGCAIIRRKQQNLNATLPSECSQCFTNYCWNG-TIDSRSVSGVGNDDYSS----- 600
Q8TG07 NFAGCIGCAIMRRKQQALNLTLPKECETCFTNYCWNG-TIDNTPAKGVTASNDFD----- 604
O59863 NFMGCISCAILRRKQESLNYTLPSECDACFEKYCWNG-TVDAT-------TPISS----- 596
Q08108 EFMGCVGCAIIRRKQQALNITLPPECETCFKNYCWNG-TLDTTPLPDVEKDVHHSFINVN 611
Q03674 NFLGCIGCAIIRRKQESLNATLPPECTKCFADYCWNG-TLSTSANPELSGNSTYQSGAIA 607
Q8TG06 DYVTCVGCAIIRRKQQSLNLTLPDICDKCFTNYCWNG-TIDNTPTKLLTPN-NQDPAAIS 595
P39457 NWTACVACAILSRSFERTGTTLPDICSQCFDRYCWNG---------------------- 568
O42790 EWPTCVACAVISRSLDRLGRQTPAACKTCFERYCWNG---------------------- 603
Q9P8P2 TWPTCFACALTDRSFMYTSENRSTTCQKCFDTWCWAG---------------------- 569
Q9P8L1 TWPTCFACALTDRSFMYTSENRSTTCQECFDTWCWAG---------------------- 569
1CJY TIIHFVLANINFRKYKAPGVPRETEEEKEIADFDIFDDPESPFSTFN------------- 682
Q9UTH5 ----------TPVDDDSKNPTYNPAVKTSSASGVHA 627
P78854 ----------SAANDD---IVYNPTCRLG--EGI-- 613
O13857 ----------TTVNNPSAVSNYAPTVLSASTTS--- 599
Q9P327 ----------TLAT--STASVYDPTVMSAATTSRAP 636
Q9Y7N6 ----------TIAS--TTVTTYAPTVLSAKIYK--- 625
Q9UWF6 ----------ENFSDDGLTNSATEYNSNNVAGFND 596
O93795 ----------DNFSDEGLTTSAAYYNSNNVAGIND 595
Q9UVX1 ----------AKTHTSGGTSSTTQQTSTTTGSSAN 710
Q11121 ----------TAASSAYSAYNTESYSSSS------ 617
P39105 ----------SASLSASAAAASASASASASASASAS 624
Q8TG07 ----------NASGSAAADMAEQDASGAAS---- 624
O59863 ----------TTSSSASSTSTSDSGN-------- 612
Q08108 SFNSSIGQEE----------SLYAGSSASQSSSSSSSSSSSSEIPS 647
Q03674 SAISEATDGIPITALLGSSTSGNTTSNSTTSTSSNVTSNSNSSSNTTLNSNSSS-SSISS 666
Q8TG06 SAIAAVTDDSPIGALLNTGSG--TKSNSSSKTNSTLVTSSRATSTGTLISNSSSNSTVSS 653
P39457 ----------TVNSTRPESYDPAFYLADN---SM 589
O42790 ----------TVNSKDTGVYMPEFKIADAHALDS 627
Q9P8P2 ----------DDNTTEPATYEPVINSVPPWLVAN 593
Q9P8L1 ----------DDNTTEPANYEPVINSVPPWLIAN 593
1CJY ----------FQYPNQAFKRLHDLMHFNTLNNIDV 707
Q9UTH5 NILLS----------FFVLLATLLVTA---------- 644
P78854 ----------
O13857 -GTSS----------VRAKPIVFYLFASLLTVSLLL--------- 624
Q9P327 SGTTSGTASSTTSSSVASATPTHKHWWDSIFEAKENP--------- 673
Q9Y7N6 ----------SRLFTYCS---------- 633
Q9UWF6 GGTSILKKA---------- 605
O93795 GGIALVKRDDLSN---------- 608
Q9UVX1 GGSSSTGSSSSSKKKNGGDLVNGGVPSSIFLVFN--SLLGLIIAYL 754
Q11121 ATGS--KKNGAG------LPATPTSFTSILTLLT--AIAGFL---- 649
P39105 ASGSSTHKKNAGNALVNYSNLNTNTFIGVLSVIS--AVFGLI---- 664
Q8TG07 ASSSSRKKNAAVS-----VDVNAKTLFAIITAMT--AVFQLI---- 659
O59863 ------KENSAR------ILAPRSTLSLLIGGLA--SVFISF---- 640
Q08108 ATATLEKKAATNSG--SHLSGISVKFSAMIMLTL--LMFTGAV--- 686
Q03674 STARSSSSTANKANAAAISYANTNTLMSLLGAIT--ALFGLI---- 706
Q8TG06 TAARSSTSSTAKKNAGSVLKLEFSKSASVMVAIAAAAVASLI---- 695
P39457 ASVSLPTML---------STVVAAGLAMLILV---------- 612
O42790 GAVAIGKMVNVW------SSVVVGVVAATLLL---------- 653
Q9P8P2 NLSIGVADAPASNESTAGTASSGAANADVSMGMVALAAGLGLML-- 637
Q9P8L1 NLSIGMADAPGSNESTAGTASSGAAKMGVGMGMVALTAGLGLML-- 637
1CJY IKEAMVESIEYRRQNPSRCSVSLSNVEARRFFNKEFLSKPKA---- 749

Appendix 3.				
Sub-family	**AC Number**	**Organism**	**Substrate Specificity**	**Reference**
AI	Q9P4E5	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AI	Q9P8W5	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AI	Q9P4E8	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AI	O94091	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AI	Q9P8W2	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AII	Q9P8W0	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).

Sub-family	AC Number	Organism	Substrate Specificity	Reference
AII	Q9P8V9	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AII	Q9P8W1	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
AII	Q9P4E6	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
A	Q9P4E7	Candida albicans (Yeast)	olive oil [(carbon source) TAG,Tween 40], α-naphthyl palmitate	"Secreted lipases of Candida albicans: cloning, characterisation and expression analysis of a new gene family with at least ten members.",*Arch. Microbiol*. **174**:362-374(2000).
BI	P20261	Candida rugosa (Candida cylindracea)	4-nitrophenyl caprylate, 4-nitrophenyl laurate, sulcatol, tributyrin, triacetin, 2,4,6-trinitrobenzene sulfonic acid	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett*. **501**, 87-91 (2001)
BI	P32947	Candida rugosa (Candida cylindracea)	4-nitrophenyl caprylate, 4-nitrophenyl laurate, sulcatol, tributyrin, triacetin, 2,4,6-trinitrobenzene sulfonic acid	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett*. **501**, 87-91 (2001)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
BI	P32949	Candida rugosa (Candida cylindracea)	4-nitrophenyl caprylate, 4-nitrophenyl laurate, sulcatol, tributyrin, triacetin, 2,4,6-trinitrobenzene sulfonic acid	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
BI	P32946	Candida rugosa (Candida cylindracea)	4-nitrophenyl caprylate, 4-nitrophenyl laurate, sulcatol, tributyrin, triacetin, 2,4,6-trinitrobenzene sulfonic acid	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
BI	P32948	Candida rugosa (Candida cylindracea)	4-nitrophenyl caprylate, 4-nitrophenyl laurate, sulcatol, tributyrin, triacetin, 2,4,6-trinitrobenzene sulfonic acid	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
BII	P22394	Geotrichum candidum (Oospora lactis)	Several simple triglycerides. Long chain fatty acid with cis-9 unsaturated bonds. CaprylinC8, CaprinC10, LaurinC12, TrioleinC18:1, ElaidinC18:1, LinoleinC18:2(opt), LinoleninC18:3	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
BII	P79066	Geotrichum fermentans (Trichosporon fermentans)	Several simple triglycerides. Long chain fatty acid with cis-9 unsaturated bonds. CaprylinC8, CaprinC10, LaurinC12, TrioleinC18:1, ElaidinC18:1, LinoleinC18:2(opt), LinoleninC18:3	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
BII	P17573	Geotrichum candidum (Oospora lactis)	With unsaturated long fatty acyl chain: triolein, trivaccinin, trilinolein, trilinolenin With saturated fatty acyl chain: tributyrin c4, trihexanoin C6, trioctanoin C8, tridecanin C10, trilaurin C12, trimyristin C14, tripalmitin C16, tristearin C18.	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
BIII	Q99156	*Yarrowia lipolytica (Candida lipolytica)*	Several simple triglycerides. Long chain fatty acid with cis-9 unsaturated bonds. CaprylinC8, CaprinC10, LaurinC12, TrioleinC18:1, ElaidinC18:1, LinoleinC18:2(opt), LinoleninC18:3	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
BIII	Q96VC9	*Yarrowia lipolytica (Candida lipolytica)*	Several simple triglycerides. Long chain fatty acid with cis-9 unsaturated bonds. CaprylinC8, CaprinC10, LaurinC12, TrioleinC18:1, ElaidinC18:1, LinoleinC18:2(opt), LinoleninC18:3	Influence of the conformational flexibility on the kinetics and dimerisation process of two Candida rugosa lipase isoenzymes.Pernas, M.A.; Lopez, C.; Rua, M.L.; Hermoso, J.;*FEBS Lett.* **501**, 87-91 (2001)
CI	P61871	*Rhizopus niveus*	Tributyrine, 1,2-didecanoyl-rac-glycerol	Can lipases hydrolyze a peptide bond? Maruyama, T.; Nakajima, M.; Kondo, H.; Kawasaki, K.; Seki, M.; Goto, M.; Enzyme Microb. Technol. 32, 655-657 (2003)
CI	P61872	*Rhizopus oryzae (Rhizopus delemar)*	Tributyrine, 1,2-didecanoyl-rac-glycerol	Can lipases hydrolyze a peptide bond? Maruyama, T.; Nakajima, M.; Kondo, H.; Kawasaki, K.; Seki, M.; Goto, M.; Enzyme Microb. Technol. 32, 655-657 (2003)
CI	P19515	*Rhizomucor miehei*	tributyrine, 4-nitrophenyl decanoate, 1,2-O-dilauryl-rac-glycero-3-glutaric acid resorufin ester , tridecanin, trihexanin, trioctanin	Properties of recombinant Rhizomucor miehei lipase with amino acid substitutions of Phe94 in the substrate binding domain, Oh, S.-W.; Gaskin, D.J.H.; Kwon, D.Y.; Vulfson, E.N.; *Biotechnol. Lett.* 23, 563-568 (2001)
CII	P61870	*Penicillium camembertii*	tributyrine, 4-nitrophenyl decanoate, 1,2-O-dilauryl-rac-glycero-3-glutaric acid resorufin ester , tridecanin, trihexanin, trioctanin	Properties of recombinant Rhizomucor miehei lipase with amino acid substitutions of Phe94 in the substrate binding domain, Oh, S.-W.; Gaskin, D.J.H.; Kwon, D.Y.; Vulfson, E.N.; *Biotechnol. Lett.* 23, 563-568 (2001)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
CII	P61869	*Penicillium cyclopium*	tributyrin,trioctanoin,trihexanin, methyl linoleate, methyl oleate, methyl ricinoleate, polyethylene sorbitan monooleate, trilaurin, trioleoylglycerol, tripalmitin	Biochemical and structural characterization of triacylglycerol lipase from Penicillium cyclopium, Ibrik, A.; Chahinian, H.; Rugani, N.; Sarda, L.; Comeau, L.C.;Lipids **33,** 377-384 (1998)
CII	O59952	*Thermomyces lanuginosus (Humicola lanuginosa)*	tridodecanoin, tributyrin, 4-nitrophenyl caprylate, 1,2-dioleoylglycerol, methyl linoleate, methyl oleate,methyl oleate, trioleoylglycerol	Production, purification, characterization, and applications of lipases, Sharma, R.; Chisti, Y.; Banerjee, U.C.;Biotechnol. Adv. **19**, 627-662 (2001)
DI	Q9UWF6	*Candida albicans*	lysophosphatidylcholine, 1-palmitoyl-sn-glycero-3-phosphocholine	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)
DI	O93795	*Candida albicans*	lysophosphatidylcholine, 1-palmitoyl-sn-glycero-3-phosphocholine	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)
DI	Q9UVX1	*Candida albicans*	lysophosphatidylcholine, 1-palmitoyl-sn-glycero-3-phosphocholine	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)
DII	Q11121	*Torulaspora delbrueckii (Saccharomyces rosei)*	lysophosphatidylcholine, 1-palmitoyl-sn-glycero-3-phosphocholine	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
DII	P39105	*Saccharomyces cerevisiae*	lysophosphatidylcholine, phosphatidylinositol-4,5-bisphosphate, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine PtdSer>PtdIns>>PtdCho.PtdEtn	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)
DII	Q8TG07	*Candida glabrata (Torulopsis glabrata)*	lysophosphatidylcholine, 1-palmitoyl-sn-glycero-3-phosphocholine	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)
DII	O59863	*Kluyveromyces lactis (Yeast)*	lysophosphatidylcholine, phosphatidylcholine	Purification and characterization of lysophospholipase-transacylase of pathogenic fungus Candida albicans, Takahashi, M.; Banno, Y.; Shikano, Y.; Mori, S.; Nozawa, Y.Biochim. Biophys. Acta **1082**, 161-169 (1991)
DII	Q08108	*Saccharomyces cerevisiae*	1-acyl-sn-glycero-3-phosphocholin, 2-acylglycerophosphocholine, phosphatidylinositol, phosphatidylserine, phosphatidylinositol-4,5-bisphosphate,Missing the other 2 from P39105 and Q03674(ptdCho,PtdEtn)	Purification and some properties of soluble phospholipase B from baker's yeast (Saccharomyces cerevisiae), Ichimasa, M.; Shiobara, M.;*Agric. Biol. Chem.* **49**, 1083-1089 (1985)
DII	Q03674	*Saccharomyces cerevisiae*	phosphatidylinositol-4,5-bisphosphate, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine PtdSer>PtdIns>>PtdCho.PtdEtn	Purification and some properties of soluble phospholipase B from baker's yeast (Saccharomyces cerevisiae), Ichimasa, M.; Shiobara, M.;*Agric. Biol. Chem.* **49**, 1083-1089 (1985)
DII	Q8TG06	*Candida glabrata (Yeast) (Torulopsis glabrata)*	phosphatidylinositol-4,5-bisphosphate, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine PtdSer>PtdIns>>PtdCho.PtdEtn	Purification and some properties of soluble phospholipase B from baker's yeast (Saccharomyces cerevisiae), Ichimasa, M.; Shiobara, M.;*Agric. Biol. Chem.* **49**, 1083-1089 (1985)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
DIII	P39457	*Penicillium chrysogenum (Penicillium notatum)*	phosphatidylinositol-4,5-bisphosphate, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine PtdSer>PtdIns>>PtdCho.PtdEtn	Purification and some properties of soluble phospholipase B from baker's yeast (Saccharomyces cerevisiae), Ichimasa, M.; Shiobara, M.; *Agric. Biol. Chem.* **49**, 1083-1089 (1985)
DIII	O42790	*Neurospora crassa*	phosphatidylinositol-4,5-bisphosphate, phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine PtdSer>PtdIns>>PtdCho.PtdEtn	Purification and some properties of soluble phospholipase B from baker's yeast (Saccharomyces cerevisiae), Ichimasa, M.; Shiobara, M.; *Agric. Biol. Chem.* **49**, 1083-1089 (1985)
DIV	O13857	*Schizosaccharomyces pombe*	lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and some properties of phospholipase B from Schizosaccharomyces pombe, Oishi, H.; Tsuda, S.; Watanabe, Y.; Tamai, Y.; *Biosci. Biotechnol. Biochem.* **60**, 1087-1092 (1996)
DIV	Q9P327	*Schizosaccharomyces pombe*	lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and some properties of phospholipase B from Schizosaccharomyces pombe, Oishi, H.; Tsuda, S.; Watanabe, Y.; Tamai, Y.; *Biosci. Biotechnol. Biochem.* **60**, 1087-1092 (1996)
DIV	Q9Y7N6	*Schizosaccharomyces pombe*	lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and some properties of phospholipase B from Schizosaccharomyces pombe, Oishi, H.; Tsuda, S.; Watanabe, Y.; Tamai, Y.; *Biosci. Biotechnol. Biochem.* **60**, 1087-1092 (1996)
DIV	Q9UTH5	*Schizosaccharomyces pombe*	lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and some properties of phospholipase B from Schizosaccharomyces pombe, Oishi, H.; Tsuda, S.; Watanabe, Y.; Tamai, Y.; *Biosci. Biotechnol. Biochem.* **60**, 1087-1092 (1996)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
DIV	P78854	*Schizosaccharomyces pombe*	lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and some properties of phospholipase B from Schizosaccharomyces pombe, Oishi, H.; Tsuda, S.; Watanabe, Y.; Tamai, Y.;*Biosci. Biotechnol. Biochem.* **60**, 1087-1092 (1996)
DV	Q9P8P2	*Cryptococcus neoformans var. grubii*	1,2-dioleoylphosphatidylcholine, 1,2-dioleoylphosphatidylethanolamine, 1,2-dioleoylphosphatidylserine, lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and characterization of secretory phospholipase B, lysophospholipase and lysophospholipase/transacylase from a virulent strain of the pathogenic fungus Cryptococcus neoformans
DV	Q9P8L1	*Cryptococcus neoformans*	1,2-dioleoylphosphatidylcholine, 1,2-dioleoylphosphatidylethanolamine, 1,2-dioleoylphosphatidylserine, lysophosphatidylcholine , lysophosphatidylethanolamine, phosphatidylinositol, phosphatidylserine, lysophosphatidylethanolamine, phosphatidic acid	Purification and characterization of secretory phospholipase B, lysophospholipase and lysophospholipase/transacylase from a virulent strain of the pathogenic fungus Cryptococcus neoformans
E	P34163	*Saccharomyces cerevisiae*	phosphatidylinositol, phosphatidylinositol 4,5-bisphosphate, 1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate	Structure, function, and control of phosphoinositide-specific phospholipase C, Rebecchi, M.J.; Pentyala, S.N.;*Physiol. Rev.* **80**, 1291-1335 (2000)
F	P32383	*Saccharomyces cerevisiae (Baker's yeast)*	phosphatidylinositol, phosphatidylinositol 4,5-bisphosphate, 1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate	Structure, function, and control of phosphoinositide-specific phospholipase C, Rebecchi, M.J.; Pentyala, S.N.;*Physiol. Rev.* **80**, 1291-1335 (2000)

Sub-family	AC Number	Organism	Substrate Specificity	Reference
F	O13433	*Candida albicans*	phosphatidylinositol, phosphatidylinositol 4,5-bisphosphate, 1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate	Structure, function, and control of phosphoinositide-specific phospholipase C, Rebecchi, M.J.; Pentyala, S.N.;*Physiol. Rev.* **80**, 1291-1335 (2000)
F	P40977	*Schizosaccharomyces pombe*	phosphatidylinositol, phosphatidylinositol 4,5-bisphosphate, 1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate	Structure, function, and control of phosphoinositide-specific phospholipase C, Rebecchi, M.J.; Pentyala, S.N.;*Physiol. Rev.* **80**, 1291-1335 (2000)
G	P36126	Saccharomyces cerevisiae (Baker's yeast)	phosphatidylcholine, phosphatidylethanolamin, phosphatidylserine	Molecular and biochemical properties and physiological roles of plant phospholipase D, Pappan, K.; Wang, X.;*Biochim. Biophys. Acta* **1439**, 151-166 (1999)
Other Lipase	P54857	Saccharomyces cerevisiae (Baker's yeast	tributyrin, triglyceride with short chain fatty acids	Molecular and biochemical properties and physiological roles of plant phospholipase D, Pappan, K.; Wang, X.;*Biochim. Biophys. Acta* **1439**, 151-166 (1999)
Other Lipase	P41365	Candida antarctica (Yeast) (Trichosporon oryzae)	Tributyrin	Production, purification, characterization, and applications of lipases, Sharma, R.; Chisti, Y.; Banerjee, U.C.; Biotechnol. Adv. 19, 627-662 (2001)