UNIVERSITY OF BELGRADE

Faculty of Chemistry

Tanja M. Bulat

Functional Proteomics of *Arion vulgaris* and *Drosophila melanogaster* by Mass Spectrometry Approach

Doctoral Dissertation

Belgrade, 2017

UNIVERZITET U BEOGRADU

Hemijski fakultet

Tanja M. Bulat

Primena masene spektrometrije u oblasti funkcionalne proteomike na primeru *Arion vulgaris* i *Drosophila melanogaster*

Doktorska disertacija

Beograd, 2017

Mentor:

Prof. dr Marija Gavrović-Jankulović, redovni profesor Hemijskog fakulteta Univerziteta u Beogradu

Članovi komisije:

1. Univ. Prof. dr Gert Lubec, M.D., Department of Pharmaceutical Chemistry, Faculty of Life Sciences, University of Vienna

2. Profesor dr Tanja Ćirković-Veličkovic, redovni profesor Hemijskog fakulteta Univerziteta u Beogradu

3. Profesor dr Zoran Vujčić, redovni profesor Hemijskog fakulteta Univerziteta u Beogradu

Datum odbrane:

Acknowledgement

I would like to express my deepest gratitude to my supervisor Professor Dr. Marija Gavrović-Jankulović and Dr. Jana Aradska for generous contributions of knowledge and experience, valuable comments and unique patience. My deepest gratitude also to Professor Dr. Gert Lubec for his support, the helpful advice and generous assistance in the preparation of this thesis all these years.

I would like to thanks the committee members Professor Dr. Tanja Ćirković-Veličković and Professor Dr. Zoran Vujčić for their useful suggestions and patience during the writing and reading of this dissertation.

I also want to acknowledge my debt to my magnificent colleges and friends Dr. Roman Smidak, MSc Edit Sodorai, Dr. Gansoo Jung, MSc Fernando Sialana for technical advices,help with data handling and casual Fridays. I was a pleasure to work with you. My gratitude likewise goes to all past and present members of Lubec lab team. Very specially, I would like to thanks to Francisco and Ana for friendly support and exciting scientific discussions; to Jiaye Yang for all lunch/ice cream/cinema time.

I would like to thank my friends Ivana, Jelena, Lela, Marko, Dragana, Andrej and Danijela for support and inspiration all these years. This would not have happened without you. All of you have a special role in this journey.

Finally, my unreserved love goes to my family, mother Nada, father Milan, brother Dražen and my husband Vladimir for their unlimited care, love and encouragement.

I would like to dedicate this thesis to my grandparents.

Belgrade, July 2017 Tanja M. Bulat

Abstract

Functional proteomics represent an approach which could provide answers to two major questions: the revelation of biological functions of unknown proteins and the description of cellular mechanisms at the protein level. The Spanish slug, *Arion vulgaris*, is one of the hundred most invasive species in Central Europe. No transcriptomic or proteomic study on *A. vulgaris* has been carried out so far. In this thesis, the first transcriptomic database from adult specimens of *A. vulgaris* is reported. To facilitate and enable proteomics in this non-model organism, a mRNA-derived protein database was constructed for protein identification. A gel-based proteomic approach was used to obtain the first generation of a comprehensive slug mantle proteome. A total of 2128 proteins were unambiguously identified from which 48 proteins represent novel proteins with no significant homology in NCBI non-redundant database. Combined transcriptomic and proteomic analysis revealed an extensive repertoire of novel proteins with a role in innate immunity, including many associated pattern recognition proteins, effector proteins and cytokine-like proteins. The immense and very successful adaptation and spreading of *A. vulgaris* suggests that it developed highly effective mechanisms to deal with infections and natural predators. The number and diversity in gene families encoding lectins point to a complex defense system, probably as a result of adaptation to a pathogen-rich environment. These results are providing a fundamental and important resource for subsequent studies on molluscs, as well as for putative antimicrobial compounds for drug discovery and biomedical applications.

The second part of the thesis is based on a bottom-up approach to analyse an enriched membrane fraction from *Drosophila melanogaster* heads using

multidimensional liquid chromatography (LC) coupled with tandem-mass spectrometry (MS/MS) that relies on complete solubilization and digestion of proteins. Membrane proteins play key roles in several fundamental biological processes such as cell signaling, energy metabolism and transport. Despite the significance, these still remain an under-represented group in proteomics datasets. An enriched membrane fraction was prepared using equilibrium density centrifugation on a discontinuous sucrose gradient, followed by solubilization using the filter-aided sample preparation (FASP), tryptic and sequential chymotrypsin digestion of proteins. Peptides were separated by reversed-phase (RP) LC at high pH in the first dimension and acidic RP-LC in the second dimension coupled directly to an Orbitrap Velos Pro mass spectrometer. A total number of 4812 proteins from 114865 redundant and 38179 distinct peptides corresponding to 4559 genes were identified in the enriched membrane fraction from fly heads. These included brain receptors, transporters and channels that are most important elements as drug targets or are linked to disease.

Apstrakt

Funkcionalna proteomika danas predstavlja metodu koja može da ukaže na odgovor na dva pitanja: koja je biološka funkcija novootkrivenih proteina i da okarakteriše ćelijske procese na proteinskom nivou. *Arion vulgaris* predstavlja jednu od sto najinvaznivijih vrsta u Evropi. Do sada u naučnoj literaturi nisu zabeleženi podaci o iRNK i proteinskom profilu ove vrste. U prvom delu teze obrađena je data problematika gde su prikazani rezultati analize iRNA na osnovu čega je napravljena baza podataka koja je korišćena za identifikaciju proteina. Ukupan broj identifikovanih proteina, dobijenih tripsinskom digestijom u gelu i analizom na masenom spektrometru, iznosi 2128. Od ukupnog broja indentifikovanih proteina, 48 nema homologe u nacionalnom centru za biotehnološke informacije (NCBI) i predstavljaju do sada neopisane proteine. Kombinacijom transkriptoma i proteoma opisan je čitav repertoar kako novih, tako i već poznatih proteina koji imaju ulogu u urođenom imunom sistemu, kao na primer *„associated pattern recognition“*, efektorski i citokinu-slični proteini. Dobijeni rezultati ukazuju na mehanizam borbe protiv infekcija i prirodnih neprijatelja, koji verovatno omogućavaju uspešnu adaptaciju i široku rasprostranjenost ove vrste. Velika raznovrsnost gena koji kodiraju lektine ukazuje na odbrambeni sistem vrste *Arion vulgaris* u sredini koji je bogata patogenima. Rezultati koji su proizašli iz ove teze pružaju fundametalno znanje o proteomskom profilu ove vrste mekušaca, uključujući i 48 novih proteina.

Drugi deo teze je fokusiran na *„gel-free bottom-up“* analizu proteina iz membranskih frakcija glave *Drosophila melanogaster*. Iako je genom *D. melanogaster* sekvencioniran u potpunosti, mali broj membranskih i transmembranskih proteina je

detektovan na proteinskom nivou. Oni predstavljaju značajnu grupu proteina, jer su potencijalna meta za većinu medikamenata, s obzirom da učestvuju u procesima energetskog metabolizma, ćelijske signalizacije i transporta brojnih biomolekula. U svrhu identifikacije membranskih i transmembranskih proteina korišćena je dvodimenzionalna tečna hromatografija u tandemu sa masenom spektrometrijom. Rađena je tripsinska/himotripsinska digestija proteina na membrani (FASP - filter-aided sample preparation). Peptidi su prvo razdvojeni u više frakcija pomoću tečne hromatografije na reverznim fazama pri visokoj vrednosti pH, a zatim je svaka frakcija dodatno razdvojena pomoću tečne hromatografije na reverznim fazama pri niskoj vrednosti pH i direktno analizirana na masenom spektrometru tipa Orbitrap Velos Pro. Broj identifikovanih peptida čija se sekvenca ponavlja u više proteinskih grupa („*redundant peptides"*) iznosi 114865, dok broj identifikovanih peptida koji su prisutni samo u jednoj grupi proteina („*distinct peptides")* iznosi 38179. Ukupan broj identifikovanih proteina iznosi 4812, među kojima je identifikovan veliki broj membranskih i transmembranskih proteina, uključujući receptore, transportere i jonske kanale.

LIST OF ABBREVIATIONS

1D - one dimensional

SDS - sodium dodecyl sulphate

PAGE - polyacrylamide gel electrophoresis

MS - mass spectrometry

2D - two dimensional electrophoresis

Mud PIT - Multidimensional Protein Identification Technology

IEF - isoelectric focusing

FASP - filter aided sample preparation

HPLC - high-performance liquid chromatography

ETD - enhanced trace detection

CID - collision-induced dissociation

HCD - higher-energy collisional dissociation

RNA - ribonucleic acid

mRNA - messenger ribonucleic acid

cDNA - complementary deoxyribonucleic acid

USER - Uracil-Specific Excision Reagent

PCR - Polymerase chain reaction

TABLE OF CONTENTS

LIST OF FIGURES

List of tables

1. INTRODUCTION

1.1. Proteomics

1.1.1. Definition

The word “proteome” is mentioned for the first time in 1995 by Wilkins et al. (1). They have postulated the concept of proteome - which represents the entire set of proteins in cells/organism at a specific time point. Besides the great technological progress of DNA and RNA sequencing, those data are not providing representable bottom line of biological samples. It is known that mRNA is not translated directly, that a journey from gene to protein is not that simple. One gene can encode more than one protein, and those proteins are the ones which define the active life of the cell (2). Hence, it is essential to determine protein expression levels at specific time points or conditions. Proteomics is a field of science whose main interest is to use quantitative protein-level measurement and/or identification of the proteins in order to characterize biological processes and functions (3). Today, we can recognize different fields of proteomic and their application (Figure 1.).

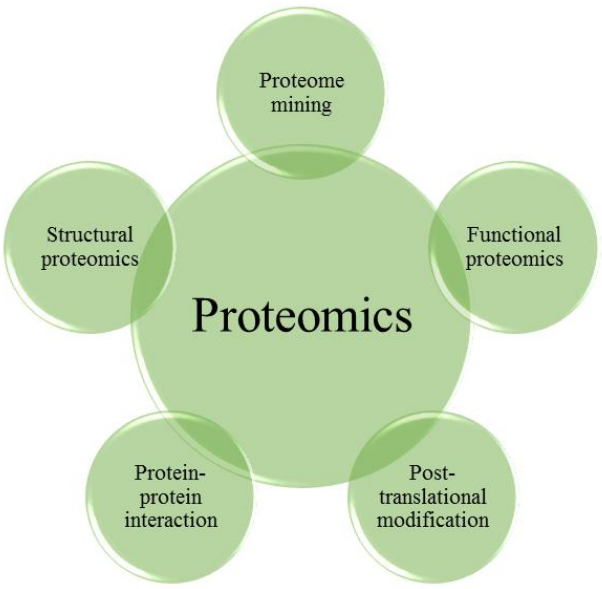


Figure 1. Types of proteomics in biology

The two major areas in the current proteomics investigations are expression and functional proteomics. Expression proteomics aims to measure protein levels (up- and down-regulation, protein-protein interaction), while functional proteomics characterizes protein activities, function, signalling pathways, etc. (4-6). Functional proteomics is "focused to monitor and analyse the spatial and temporal properties of the molecular networks and changes involved in living cells" (7). Also, the goal of this approach is to elucidate the biological function of unknown proteins, and it combines the analysis of mRNA coding and proteomics (8).

In combination with the elegant tools which have been designed to study proteins, such as high quality separation (1D and 2D electrophoresis, chromatography) and characterization by mass spectrometry, proteomics provide insight into many fields – the causes of disorders, identifying early markers of a disease process, isoforms of protein, post-translational modifications, etc. (9)

1.1.2. Sample preparation, protein separation and enrichment

Different types of biological samples are used in proteomic research: cells, tissues and body fluids (blood plasma, cerebrospinal fluid). The sample preparation is the crucial step in protein analysis and it is specific for each type of sample It consists of solubilization, denaturation and reduction, in order to completely break up the interactions between the proteins, and removal of all chemicals which can have a negative effect on the sample separation (10).

Cell culture samples are less challenging for preparation than tissue samples, due to its simple structure. In case of tissue, composition of samples should be considered, based on its origin (lipids in brain, connective tissue in skin, etc.). In case of skin, the best way is to crush the sample in liquid nitrogen, followed by homogenisation in lysis buffer with detergents, reductants and protease/phosphatase inhibitors (11). Mechanical homogenisation (rotor-stator homogenizers, open blade mils) can be combined with ultrasonic homogenisation. Advantage of ultrasonic homogenization is that besides homogenization, it also performs emulsifying, dispersing and suspending of the mixtures(12-14).

During work on this thesis, special attention was given to presence of polysaccharides and lipids in samples for proteomic analysis. Interactions between polysaccharides and sample buffer can increase viscosity of the sample itself, which will make problems during electrophoresis (obstruction of the pores in polyacrylamide gels). If samples are analysed in 2D gels, polysaccharides will interact with ampholytes and cause prolongation of focusing time, as well as streaking visible on 2D gels(11). Moreover, since some polysaccharides are negatively charged, they may form complexes with proteins by electrostatic interaction (15). Removing sugars from samples can be done in many ways: precipitation with trichloroacetic acid (TCA), acetone, ammonium sulphate or commercially available kits.

In case of lipids, especially in the membrane fraction from the brain, they may influence protein solubility, as well as p*I* and molecular weight. Depending on the method for proteomic analysis, removing of lipids can be performed by TCA or acetone precipitation, or by centrifugation of samples with a filter device in buffer which include CHAPS or SDS (11). In case of 2D analysis of membrane proteins, heating in combination with SDS can not be used, since it will influence isolecetrical focusing IEF. The most common way of removing lipids from samples is the combination of Urea/Thiourea/CHAPS with centrifugation cut-off filters (from 3 to 100 kD). Studies have shown that this method not only removes lipids, but also salts. The yield of solubilized proteins is estimated over 90% (11, 16). The method of the sample preparation and separation depends on the ultimate goal and the type of analysis needed.

After isolation of the proteins, by native structure, usually they will be insoluble. In order to enable protein solubilization and expose cleavage places to enzymes, interactions in proteins (disulphide bonds, ionic and hydrophobic interaction) should be broken (17, 18). Depending on the method of separation, different chemicals can be used for this purpose (SDS, Urea/Thiourea, DTT, IAA, etc.).

1.1.2.1. 1D-SDS-PAGE

Polyacrylamide gel electrophoresis (19) is a frequently used technique in biochemistry, molecular biology and biotechnology, in order to separate

macromolecules (proteins or nucleic acid) based on their electrophoretic mobility. The mobility of biomolecules depends on their molecular weight. Sodium dodecyl sulphate (SDS), as anionic detergent, is used for linearization of proteins in the sample. Also, its function is to impart a negative charge on proteins (19, 20). One-dimensional separation is used as a way of prefractionation technique for MS analysis (11). Mixture of the proteins is separated due to different molecular weights. After staining, the whole protein lane is cut into approximately 20 gel slices, and then every slice (one sample) is additionally cut into 3-5 peaces in order to enlarge the reaction surface between enzymes and proteins in gel. It is a very useful and quick method for samples with wide dynamic range, since many projects are limited by the amount of samples and cost/time of procedure and instruments (21). Choosing percentage of gel, better separation can be accomplished for low or high molecular weight. Also, separation of both high and low molecular weight can be accomplished by using gradient gels (21). After digestion and extraction, peptides are analysed on LC-MS. This combination of techniques is called gel-enhanced LC-MS (GeLCMS), and it represents a standard method in the field of proteomics (22). The advantage of SDS-PAGE is removal of low molecular weight impurities (salts, buffers and detergents) from the sample. Moreover, gel serves as a matrix for digestion by proteolytic enzymes (23). The disadvantage of this method is the possibility of traces of SDS, which in combination with low orthogonality can limit sensitivity of MS analysis (22).

1.1.2.2. 2D-IEF-SDS-PAGE

The two-dimensional gel electrophoresis allows higher resolution for the separation of proteins, as compared to 1D-SDS-PAGE. This is due to the introduction of isoelectric focusing followed by separation based on molecular weight on SDS-PAGE. IEF is the first dimension where proteins are separated according to their isoelectric point (p*I*). A protein`s p*I* is defined as a pH at which the total charge of the protein is zero (24). Proteins with a more positive net charge will drift to the cathode, while those with a more negative charge will move to the anode, until they reach their p*I* (Figure 2.) (25). The most significant conditions for a successful IEF-based separation are an unchanged charge of proteins, and that proteins are solubilized during this process (26).

In order to achieve this, combination of neutral chaotropes (urea and thiourea) and electrically neutral detergent (CHAPS) is mostly used (27). The combination of urea/thiourea with reducing agent tributylphosphine (TBP), alkylation agent iodoacetamide (IAA) and protease inhibitors will prevent protein modification, aggregation or precipitation during IEF. Since CHAPS belongs to neutral detergents, it is stable in a wide pH range. Because of this, 4% CHAPS will not interfere with IEF of the samples and it is included in common proteomics practice (28). DTT as a reducing agent is not recommendable for first dimension (IEF), since free-thiol-containing reagents are week acids- it may migrate during focusing. As a result, reoxidation of sulfhydryl group will occur and reduce protein solubility (29). Instead of DTT, phosphines (TBP) are common reducing agents during IEF, not only because they are neutral, but also because they will not react with acrylamide (30). At last but not the least, blocking of protease activity can change major differences in 2D profile. In absence of protease inhibitors, presence of proteins higher than 50 kD will be reduced (28). Finnie and Svensson showed that presence of protease inhibitors during IEF is crucial, especially during rehidratation of IGE strips (31, 32).

A very important condition is a relatively high voltage electric field which will force proteins to travel to their exact p*I* position (33). After IEF, equilibration process with SDS will make proteins mobile in second dimension. Visualization of sample's proteome profile will be possible after staining the gel with the dyes, such as Coomassie, Blue Silver and Silver staining.

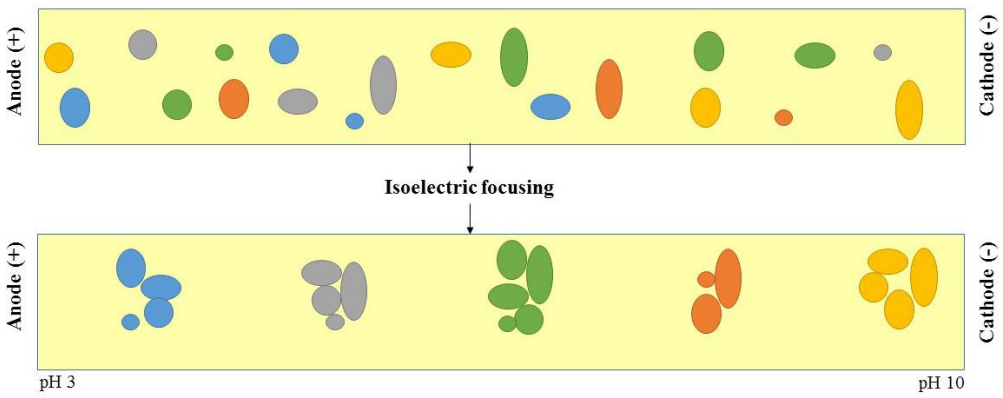


Figure 2. Schematic representation of isoelectric focusing of proteins according to pH

Afterwards, spots of interest can be processed for MS analysis. 2D gel-based proteomics is extensively used for bacterial proteomics (34), micro enzymology (35), immunoproteomics (36, 37), post-translational modifications (38-41).

The most important thing about the sample itself is the dynamic range of protein abundance, since biological samples are complex. By removing the most abundant protein/proteins from the sample (albumin, immunoglobulins from blod), complexity is reduced and sensitivity of analysis is increased (42). Besides this, low abundant proteins could be masked by high abundant proteins on 2-DE profile (11, 43). Removal of protein of interest is possible with affinity chromatography, mimetic ligands and antibodies (44-47). Despite the relatively high costs/time ratio, 2-DE provides insight to proteins as they are in organism, including potential isoforms, PTMs and charges It is estimated that with proper choice of gel percentage, size, and pH gradient, up to 5000 proteins can be resolved (48). Also, when using more sensitive visualization staining (silver staining), only 1 ng of proteins is enough for determination of proteome profile (49). In case of comparative analysis, two-dimensional difference gel electrophoresis (2-DIGE) was developed. Two different samples (control and treated, for example) will be labeled with different cyanine dyes prior to first dimension, but they will be run and analyzed on the same second dimension, which will reduce gel-to-gel variability (50, 51).

1.1.2.3. Membrane enrichment

The plasma membrane is a structure which separates the cell from the cellular environment. It is a dynamic structure, controlling chemical trafficking in and out of the cell (52). The basic structure of the cell membrane is made of lipids, and membrane proteins are carrying out specific functions, depending on the type of the cell (cell adhesion, ion conductivity, signalling etc.). The amount of membrane proteins depends on the cell type (53). The two basic groups of membrane proteins are integral (constantly attached to the membrane) and peripheral membrane proteins (temporarily attached to the membrane) (Figure 3.) (54). It has been valued that 20-30% of the genome is encoding membrane proteins and that they represent the target for over 50%

of all modern medical drugs. Membrane proteins represent a very important part of medical and biology research, because the changes in their structure and function are responsible for the development of several disorders (55-57).

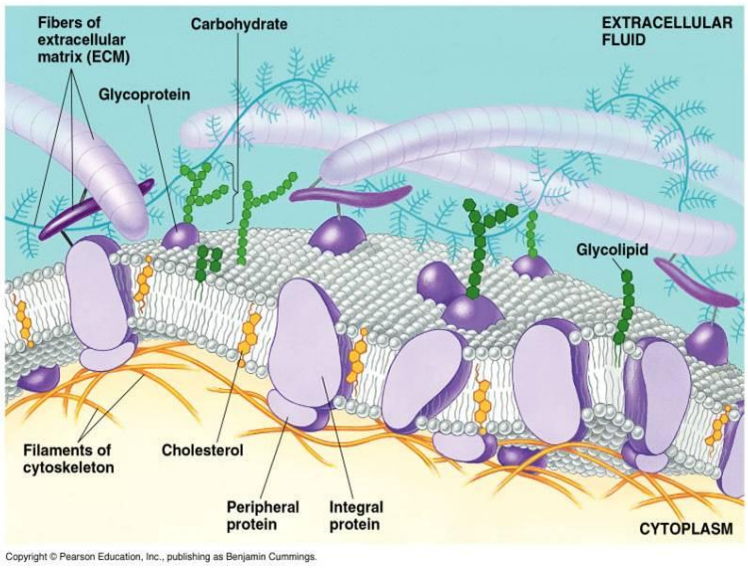


Figure 3. Fluid mosaic model of cell membrane (Picture adopted from Campbell, Neil A.;Reece, Jane B. Biology, 7th Edition)

Analysis of whole cell protein extract remains challenging because of the complexity and of wide dynamic range in abundance of proteins. Subcellular separation of proteins is a common strategy used in order to investigate protein fraction of interest, like membrane proteins which are present in relative low levels in cell (58). According to literature, density gradient centrifugation (e.g., using sucrose, sorbitol, Ficoll, or Percoll), is one of the most common techniques for the separation of plasma membranes (59, 60), mitochondria (61, 62), Golgi and synaptic vesicles (63, 64). There are two types of density gradient centrifugation: rate-zonal and equilibrium. In case of rate-zonal, sucrose is most dense on the bottom of the tube, and then density is gradually decreased to the top of the tube. In equilibrium, two or more layers of different sucrose density are loaded on tube, taking into account that the maximum density of the sucrose is higher than that of the particles of interest (samples) (65).

Besides ultracentrifugation, combination of chloroform and methanol is used in order to enrich membrane fraction. This method is based on hydrophobicity of

transmembrane and membrane associated proteins(66). In addition, Sivars and Tjereld (67) and Everberger et al.(68) describe role of detergents such as n-Dodecyl-β-D-Maltoside (DDM), TritonX-114 and polyethylene glycol (69) Advantages of aqueous two-phase systems are numerous: fast separation, stability of biomolecules, large-scale usage and supplementation of the system with protease/phosphatase (70).

1.1.3. MS based proteomics

1.1.3.1. Proteolytic digestion

Proteolysis is a biochemical process where proteins are disrupted into small peptides and amino acids. This process can be catalysed by enzymes called proteases and they are a very important component of physiological processes including digestion (trypsin, chymotrypsin), blood coagulation (factors VIIa, IXa), immune response (complement factors) and reproduction (acrosin) (71). There are seven known families of proteases, based on catalytic residues: aspartic-, cysteine-, serine-, metallo-, threonine-, glutamic- and asparagine- proteases and each of those families is divided into sub groups (72). In proteomics research the most often used enzymes are serine-proteases (trypsin, chymotrypsin, endoproteinase GluC), and aspartic-protease (pepsin) (Table 1). The enzyme's active site, called the catalytic triad, represents amino acid residues with acid-base-nucleophile motif (73). In case of serine-proteases amino acid residues are histidine, aspartate, and serine. Modifying the electrostatic environment of the serine by increasing nucleophilicity is possible by forming a charge between catalytic triad components. In addition, some enzyme's (trypsin) backbone amide hydrogen atoms of Gly-193 and Ser-195 form an "oxyanion hole" which additionally stabilizes negative charge on the carbonyl oxygen atom of the cleaved amides (Figure 4.) (74).

Table 1. The most common enzymes used in proteomic analysis (75)

HIGHLY SPECIFIC ENDOPROTEASE		
Enzyme	**pH of activity**	**Specificity**
Trypsin	7-8,5	Arg, Lys
Arg-C	7,2-8	Arg
Glu-C	7,5-8,5	Arg, Glu
Lys-C	8-9,5	Lys
Asp-N	6-8,5	Asp, Cysteic acid (Glu)
Prolyl endopeptidase	7-7,5	Pro
LESS SPECIFIC ENDOPROTEASE		
Enzyme	**pH of activity**	**Preference**
Chymotrypsin	7-9	Phe, Trp, Tyr
Pepsin	2-4	Phe, Met, Leu
Proteinase K	6,5-9,5	Aliphatic residue
Thermolysin	7-9	Leu, Ile, Phe, Trp, Met, Val

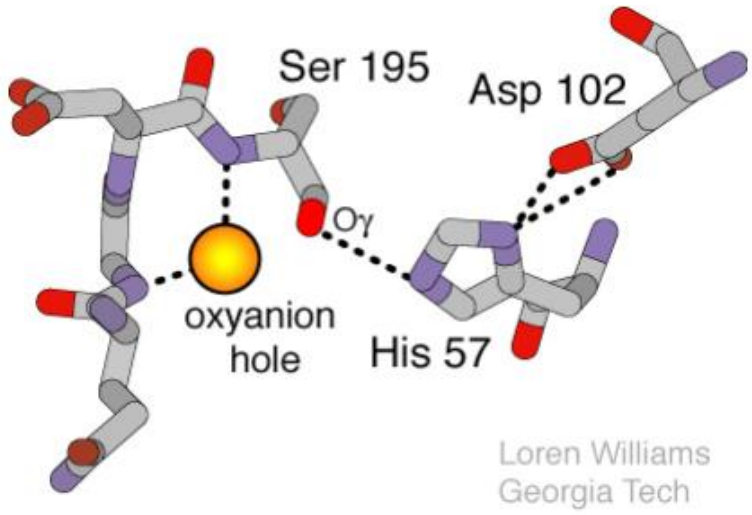


Figure 4. Trypsin active site (The picture is adopted from www.proteopedia.org)

To accomplish the enzyme's maximum access to the protein cleavage sites two important steps are included in sample preparation - reduction and alkylation.

Dithiothreitol (DTT) is the most often used reducing agent. After exposure of samples to DTT solution, cysteine disulphide bonds are reduced into cysteine free sulfhydryl groups. Reduction step is followed by alkylation with iodoacetamide (IAA), which reacts with free sulfhydryl groups of cysteine residues to form S-carboxyamidomethyl-cysteine. In this way sulfhydryl groups cannot be reoxidized (76).

1.1.3.1.1. In-gel digestion

In-gel digestion of samples for MS analysis (Figure 5) is a method which was introduces 1992 by Rosenfeld (77). This method has some important advantages in comparison to gel-free approaches, such as: fractionation at protein level decreases the complexity of the sample, and in some cases allows identification of post/translational modifications, especially with by mass analysis after 2D-IEF-SDS-PAGE (78).

Optimization of digestion conditions for in-gel digestion is crucial in order to enable the reaction between protein and enzyme. Digestion of proteins in polyacrylamide matrix is restricted by enzyme diffusion and concentration, temperature and pH (79). Comasine, as a gel dye, must be completely removed in order to enable enzymatice cleavage and reduce backround during MS analysis of peptides. (80). Extraction of proteins from gel matrix is a multistep process, where in the first step most of peptides are eluted in solution, and an additional step can increase yield of the whole process by only 5-10% (81). Efficiency of extraction depends on many factors, such as thickens of gel, size of gel pores (gel percentage), physico-chemical properties of peptides (e.g. hydrophobicity, p*I*, hydrodynamic radius). It is easier to extract peptides (especially large ones 7-10 kD) from lower percentage of gel than from higher percentage, due to acrylamide pore size (81). Additionally, if it is necessary, gel plug can be submerged in deionized water and freeze. This process will crash pores and ensure recovery of large peptides. Shevchenko and colleges published protocol which suggest combination of formic acid/acetonitrile as extraction solution (82). Formic acid will ensure extraction of hydrophilic peptides, while organic solvent will ensure extraction of hydrophobic peptides from gel plug.

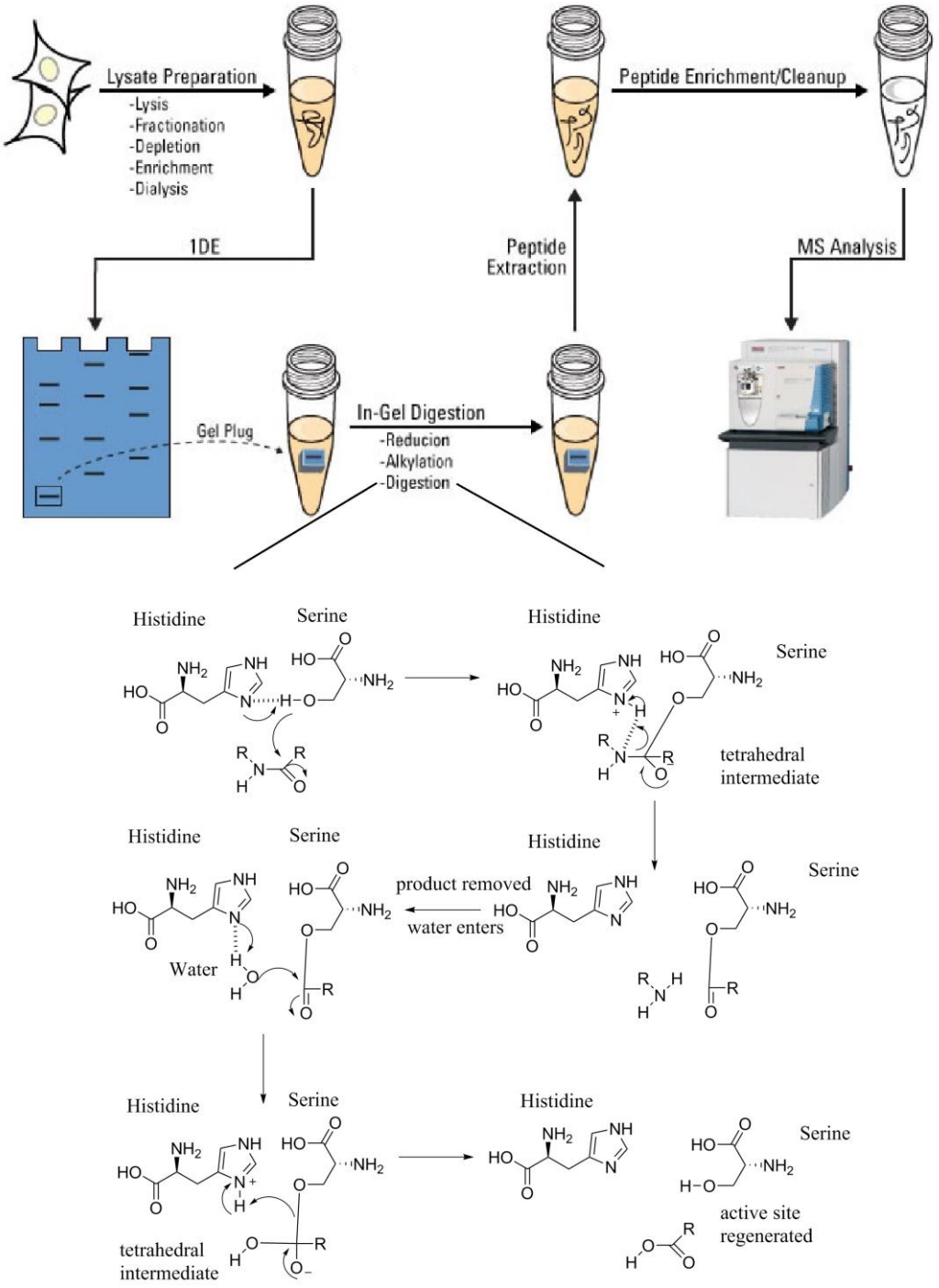


Figure 5. Schematic presentation of in-gel digestion procedure (The picture was adopted and modify from www.sites.psu.edu)

1.1.3.1.2. In-solution digestion

In contrast to the in-gel digestion, in-solution digestion is more efficient, in a sense of producing higher peptide yields and sequence coverage. The digestion protocol is designed in such a way to eliminate salts and detergents which are not compatible with MS analysis, and to keep proteins soluble. Urea is a commonly used denaturant for in-solution digestion (75). Besides this, during the development of the protocol, a great attention has been paid to a non-detergent washing of sample and an elution in low pH range. Usually, in-solution digestion is followed by separation of peptides on HPLC. The most common approach is Multidimensional Protein Identification Technology - Mud PIT strategy (gel-free technique), which enables high resolution separation of peptides and individual identification of complex samples (83).

1.1.3.1.3. FASP

Using gel-free and in-solution digestion enables removal of impurities (detergents, salts, SDS, etc.) and minimizes sample handling. Nevertheless, digestion obstruction and incomplete solubilization of proteins can be potential problems, especially when working with membrane proteins (83). Before usage of SDS, solubilization of membrane proteins was done with formic acid and/or organic solvent (84, 85). Filter-aided sample preparation (FASP) is a universal method of sample preparation for MS analysis. The advantage of this method is that membrane will keep proteins, but not small molecules and peptides. Filter is a "proteomic reactor", which gives us different possibilities for MS sample preparation: detergent removal, buffer exchange, chemical modification (reduction and alkylation) and protein digestion with different enzymes (Figure 6). Numerous citations show that FASP is an appropriate method for proteomic analyses of bacteria, mammalian cells, yeast and paraffin-embedded tissue samples (86-88). Also, it has been successful for membrane proteins, phosphoproteins and N-glycoproteins (89, 90).

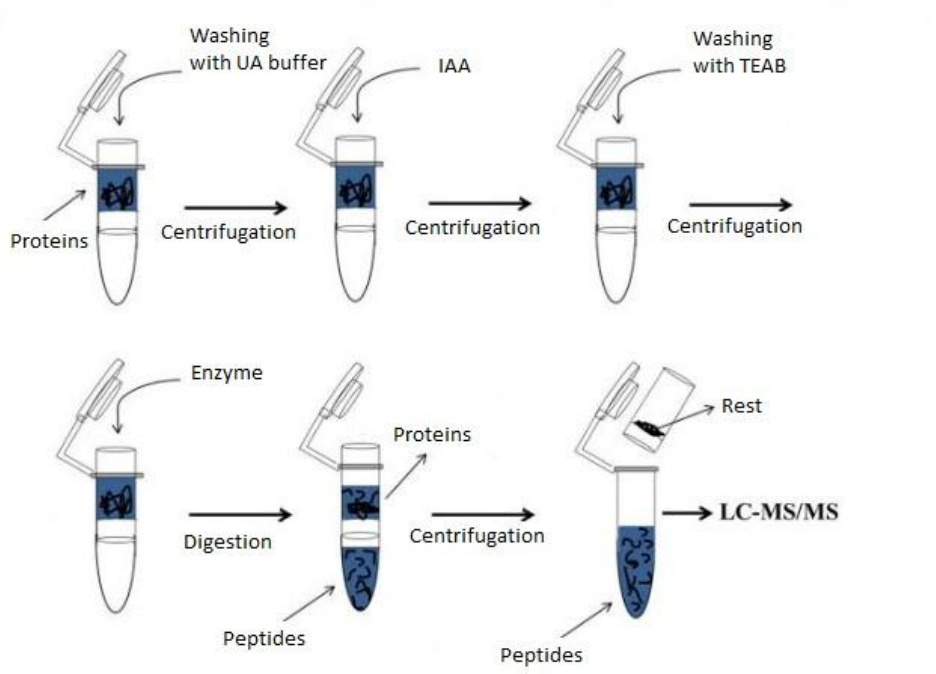


Figure 6. Shematic presentation of Filter-aided sample preparation (91)

1.1.4. HPLC and MS

1.1.4.1. HPLC and pre-fractionation

Chromatography is an analytical technique for separating a mixture of chemical substances into individual components based on different properties of each component. In principle, the separation of components depends on the affinity of the constituents to the stationary phase and their solubility with mobile phase.

High-performance liquid chromatography or high-pressure liquid chromatography (HPLC) utilizes a solid stationary phase with a liquid mobile phase. The operational pressure during the separation is between 50 and 350 bar, which is significantly higher than in regular chromatography. The sample volume is smaller (few microliters), the diameter of the analytical column is 2.1-4.6 mm and it is packed with sorbent particles the size of 2-50 µm. All of those physical characteristics give HPLC a high resolution in separation of mixture of compounds.

In proteomics, the most commonly used HPLC system for the separation of peptides is the reversed-phase liquid chromatography (RPLC). This involves the separation of peptides on the basis of their hydrophobicity. The RPLC is performed on a hydrophobic stationary phase (C18) with an acidic gradient polar mobile phase. The stationary phase consists of a polymeric C18 format (octadodecyl hydrocarbon) integrated via a tri-functional silylation of 2 or 3 silica atoms on the silica gel backbone. This increases the silylation results with a greater column stability, particularly in acidic mobile phase conditions. The hydrophilic peptides pass through the column and are eluted first; this results in the successive elution of hydrophobic peptides by changing the composition of the mobile phase (increasing the percentage of the organic solvent) (Figure 7) (92).

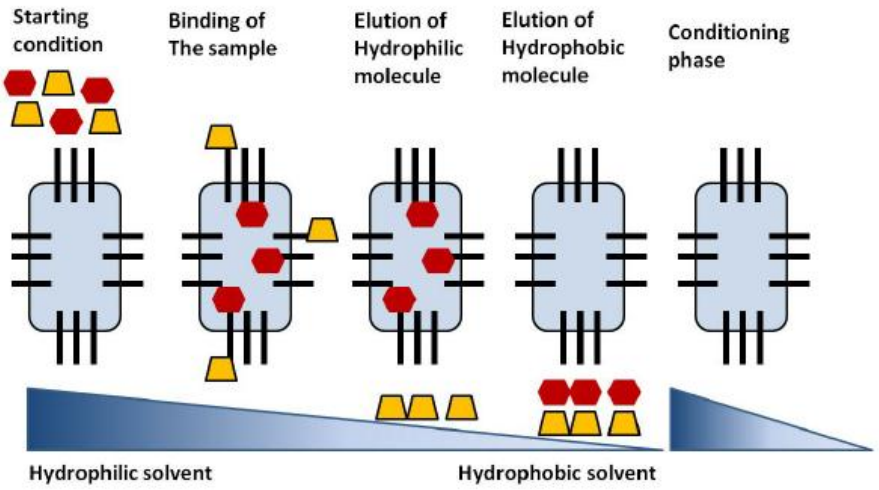


Figure 7. Shematic presentation of reversed-phase chromatography (93)

Given the wide-range of proteins in of biological samples, 2-dimensional peptide-level separation can be performed to reduce sample complexity. This is accomplished by employing two or more methods with different separation selectivities prior to the actual analyses (94). Two-dimensional liquid chromatography, 2D-LC, consists of a liquid chromatography for the pre-fractionation peptides followed by the individual LC MS/MS analyses of the separated fractions. The current separation strategies/modes include: strong cation exchange (SCX), isoelectric focusing (IEF), capillary electrophoresis (CE), capillary isoelectric focusing (CIEF), and mixed mode

pH reverse phase (RP-RP). A strong cation exchange (SCX) was used as a first dimension (95). Traditionally, SCX was the classical peptide pre-fractionation approach. Methods other than SCX were developed as there is a need for sample desalting prior to LCMS analyses leading to sample loss of up to 50%. Further, resolving of peptides in ionic buffers is a concern due to ion suppression of species during MS analyses.

The advent of 2D RPLC/RPLC that utilize buffers that are low in salt or salt-free solutions eventually replaced SCX. MudPIT (2D RPLC/RPLC) is a chromatography-based proteomic technique where a complex peptide sample is separated first into fractions and then every fraction is separately analysed on the RPLC-MS/MS system. Gilar et.al (96) demonstrated that different pH of the mobile phase has a crucial impact on the quality of RPLC on Mud PIT analysis. Peptides contain carboxylic (RCOOH) or basic (RNH_2) functional groups. When the pH of buffer is equal to the isoelectric point of a peptide, the peptide will be in both states: ionized and neutral (Figure 8).

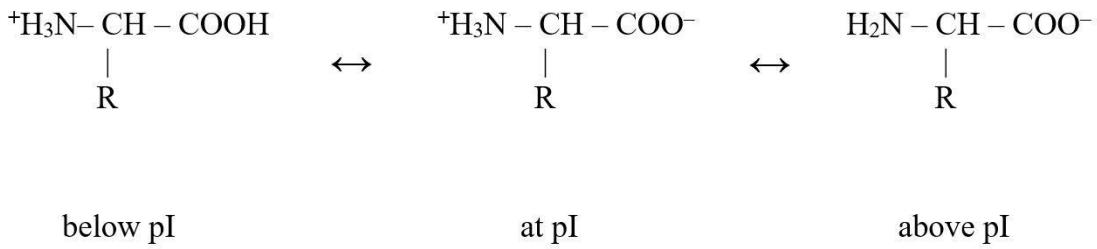


Figure 8. Schematic presentation of ionized and neutral peptide state

The zwitterionic state has poor retention in the reversed phase columns. Peptide retention can be improved by modulating the pH of the aqueous mobile phase. At low pH, positively charged species dominate and at high pH, negatively charged species dominate. In each of these two conditions (below or above its isoelectic p*I*), the analyte acquires an "ion-suppressed state". Elution of the retained peptides is then accomplished by increasing the organic solvents gradient (ACN, MeOH, iPrOH). By

pre-fractionating the peptides at pH 10 followed by the LC MS/MS at pH 2.6, the highest orthogonality was accomplished with C18 (96).

1.1.4.2. Mass spectrometry

Mass spectrometry is an analytical, instrumental technique that can identify and/or quantify ionized chemical species based on mass to charge ratio. The technique was developed by A. J. Dempster (1918) and F.W. Aston (1919). The basic components of a mass spectrometer are an ion source, a mass analyser and a detector (97).

1.1.4.2.1. Ion source

Nowadays, we can distinguish different kinds of ionization. The most used ones in chemical/biochemical research are:

a) Fast atom bombardment (FAB)

b) Chemical ionization (CI)

c) Atmospheric-pressure chemical ionization (APCI)

d) Electrospray ionization (ESI)

e) Matrix-assisted laser desorption/ionization (69)

Charged species are generated by an ion source, travels to a mass analyser of which signals are recorded in a detector. Data processing of the signals provides an output of the mass to charge ratio (m/z, Thomsons) with the relative abundance of each ion in sample (98). In this study, an ESI source was used to ionize the positively charged peptides in liquid to gas phase (ESI/MS). The ESI/MS technique invented by John Fenn paved the way for his Nobel Prize award for chemistry in 2002 in recognition for its development and application in biochemical, biomedical and pharmacological research.

The analysis of complex peptide mixtures is enabled by coupling the MS instrument as a mass detector, and the high separation resolution of the HPLC.

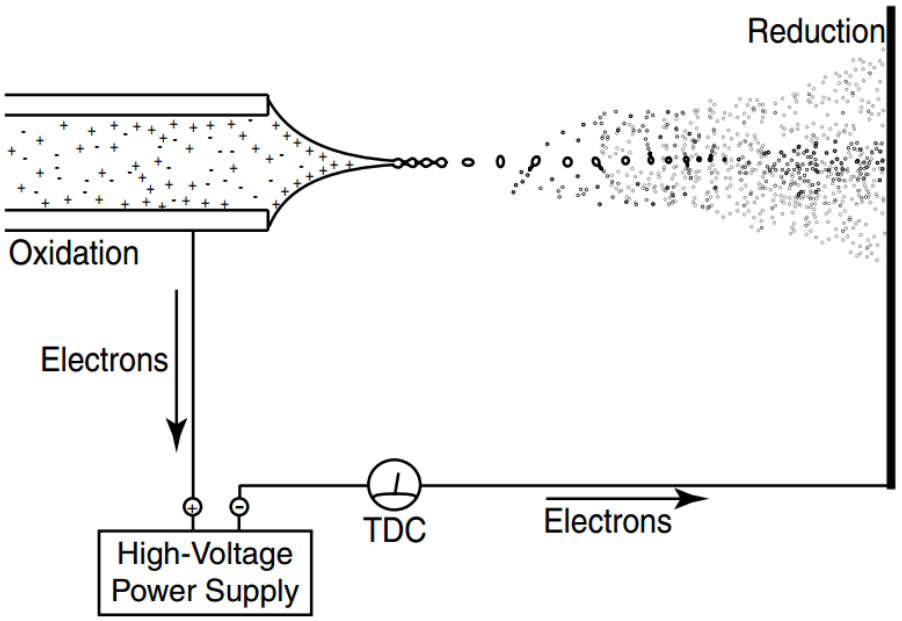


Figure 9. Illustration of main processes of ESI. TDC-total droplet current (I). (Copyright: Paul Kebarle and Udo H. Verkerk, On the Mechanism of Electrospray Ionization Mass Spectrometry (ESIMS))

The curtail events of ESI that occur at atmospheric pressure are presented in Figure 9. Electrical double layer will be formed as a consequence of infiltration of the established electric field into the liquid, due to the dipole moment of the solvent, which will enable formation of the charged jet by a large amount of positive ions (positive ion mode). Gas-phase ions will be formed in the process called Coulomb effect. During the process of evaporation, charged droplets (formed from jet) will become smaller and positive charges will be closer to each other. As a result, Coulombic repulsion of positive charge will destabilize the droplet, and it will emit a jet of smaller positive charged droplets. This process will be repeated, together with a solvent evaporation, until the gas-phase ions are formed (99). Ions are guided by an electrical potential and pressure differences, while solvent vapour will be removed by nitrogen. ″Cleaning-up″ of ions will be performed by electrical field and high vacuum system, in order to accelerate ions to mass analysis chamber.

1.1.4.2.2. Mass analyzer

The first MS and mass analysers were developed during the 1980s, but it was in the past 20 years that several types of mass analysers have been established (Table 2). A mass analyser is the part of the MS which is isolated from the outer surroundings, and it is a combination of electric/magnetic fields. Its function is to capture charged particles and separate them based on m/z ratios. The choice of mass analyser type to be used depends on the application and the desired performance (100).

Table 2. Types of mass analysers

Type of analyser	Symbol	Principe of separation
Electric sector	E or ESA	Kinetic energy
Magnetic sector	B	Momentum
Quadrupole	Q	m/z (trajectory stability)
Ion trap	IT	m/z (resonance frequency)
Time-of-flight	TOF	Velocity (flight time)
Fourier transform ion cyclotron resonance	FTICR	m/z (resonance frequency)
Fourier transform Orbitrap	FR-OT	m/z (resonance frequency)

For the performance of the mass analyser, five main points have to be taken into consideration: mass range limit, analysis speed, transmission, mass accuracy and resolution. Mass range limit is expressed in Thomsons (Th) which is based on *mass-to-charge* ratio. The analysis speed (scan speed) describes the speed rate of mass analyser in specific mass range expressed in mass units per second or mass units per millisecond (u s−1 or u ms−1). The transmission represents the ratio of ions that reach the detector over number of ions that entered the mass analyser. The mass accuracy represents the ratio or difference between $m_{theoretical}$ and $m_{measured}$ expressed in parts per million (ppm) or milimass units (mmu). The resolution or resolving power is the ability of MS to distinguish two ions with close m/z values (100).

In the course of this study, two types of mass analysers were implemented: 3D ion trap and Orbitrap. Using an oscillatory electric field, ion trap stores ions prior to selected analysis.

The first use of modifying ion trap was performed during 1920 and described by Kingdon (101). The 3D ion trap (quadrupole ions traps - QITs) consists of a circular electrode with two ellipsoid caps on top and bottom. The invention of the 3D ion trap awarded Wolfgang Pauli a Nobel Prize in physics in 1989. The Orbitrap mass analyser consists of an outer cylinder-like electrode and a coaxial inner spindle-like electrode that traps ions in an orbital motion around the spindle (102). It was designed and patented by Makarov and was commercialized by Thermo Scientific in 2005. The conversion of the signal into a mass spectra is enabled by Fourier transformation.

Depending on the kind of ion trap and/or combination of other mass analysers (e.g., 2D linear ion traps), one can distinguish LTQ Orbitrap (liner ion trap), Q Exactive (Quadrupole mass filter), as well as MALDI LTQ Orbitrap (103).

1.1.4.2.3. Ion activation

Tandem mass spectrometry (MS/MS or MS^2) implements additional mass analyses after a mass selection/isolation and ion activation/fragmentation. Ion activation can be done based on different mechanisms: collision-induced dissociation (69), electron capture dissociation (ECD), electron transfer dissociation (ETD), negative electron transfer dissociation (NETD), electron-detachment dissociation, charge transfer dissociation (CTD) and a special group of activation called photodissociation (104-106). In proteomics research, the commonly used ion activation techniques are:

a) CID – activation under impact of ion in gas phase and neutral atom or molecule. Higher-energy collisional dissociation (HCD) is a type of CID typically coupled with Orbitrap mass analysers.

b) ETD – activation helped by radical anions (fluoranthene). This kind of activation is used for studying of peptide modification, such as phosphorylation.

Depending on the type of ion activation used, fragmentation occurs at distinct peptide bond positions to generate a highly predictable ions species.

In Figure 10 the nomenclature of the generated fragment ions is shown, as previously described by Roepstorff and Fohlman (107). If a charge of a fragment is maintained on the N terminal, ions will be a, b or c, but if it is located on the C terminal than it will be x, y or z (107, 108). CID activation (i.e., HCD) typically cleaves peptides to generate b and y ions species whereas ETD activation generates c and z fragment ions.

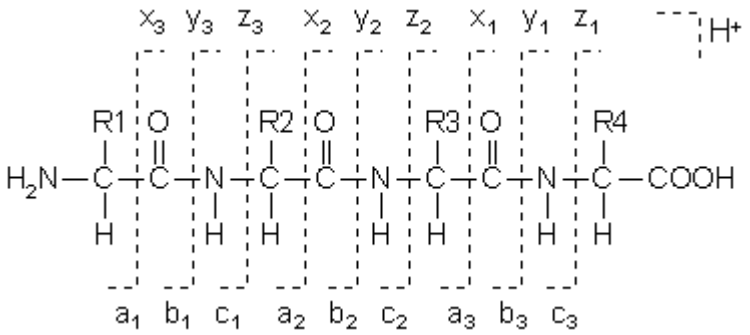


Figure 10. Nomenclature of ions species according to P. Roepstorff and J. Fohlman

1.1.4.2.4. Detector

At the rear end of a mass spectrometer is a detector that measures ion signal intensities proportional to their abundance. Detectors can be divided into two main groups: point ion and array detectors. Point ion detectors measure signals of a single mass in the time point, whereas array detectors detect multiple ion masses simultaneously (109). In both cases, the signals are amplified and processed under different computational platforms for large scale analyses.

1.2. Transcriptomics

1.2.1. Definition, problems and perspectives

Transcriptomics has been a fast evolving field of science during the last years, due to technological development. First concept of transcriptome was

implemented in 1958 with the establishment of central dogma of biology by Francis Crick. He suggested that coding information from DNA is transcribed to RNA. Eventually, it was shown purposed that RNA is used for synthesis of proteins of specific sequence (110). Later on, with the discovery of ribosomal and transfer RNA, and coding (exons) and non-coding sequences (introns), the significance of RNA was elevated to a different level. Today it is known that RNA can be divided into two groups: protein-coding (mRNA) and protein-non-coding RNAs (ncRNAs) (111). Protein-non-coding RNAs are house-keeping RNAs (translation–tRNA; ribosomal–rRNA; small nuclear–snRNA; small nucleolar–sonRNA; guide–gRNA) (112) and regulatory RNAs (antisense–aRNA; CRISPR–crRNA; long noncoding–lnRNA; micro–miRNA; piwi-interacting–piRNA; small interfering–siRNA; short hairpin–shRNA) (113-117). Together, all this molecules shape the future of the cell and organism.

Due to the complexity of biological organisms, one of the first questions was - can we see a correlation between mRNA expression and proteins? Some experiments have shown that the central dogma of molecular biology gives a positive correlation (118-120), while in some cases, there is no significant correlation between mRNA and protein levels (121). Nie and coworkers speculate that the absence of correlation can be caused by biological and/or experimental reasons, such as regulation of translation, half-life differences of mRNA/proteins and methodological limitations (e.g. sample preparation, normalization of data) (122). Eukaryotic polysomes can bind active mRNA and begin a process of protein synthesis, while the mRNA molecule, which is inactive, will be strongly attached to monosomes. There is a constant flow of those two types of mRNA – active and inactive. Besides this, lack of correlation can be a consequence of rapid degradation of mRNA as an answer to different stimuli (123). A combination of proteomics and transcriptomics gives a significant insight in all applied fields of research, such as cancer biology (124), neurosciences (125), ecology, zoology (126) etc. Nevertheless, besides level of mRNA, other processes also have a regulatory function. The most important ones are in proteins, like post-translation modifications (PTMs), which are able to regulate the level of mRNA, protein synthesis and protein activation (127). A large-scale proteomics studies (Mud PIT, tandem MS/MS) in combination with transcriptomics may untangle the regulatory mechanisms and complex signalling networks in living organisms (128-130).

1.2.2. Transcriptomics analysis of model organisms and humans

Collected basic knowledge about biological processes has been expanded thanks to the extensive studies on animal and plant model organisms. The most used ones are *Escherichia coli, Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, Rattus norvegicus, Arabidopsis thaliana, Danio rerio* and *Xenopus laevis* (131). Common in all this cases is that they are well studied and that their genomes have been sequenced, together with genetic material from mitochondria/chloroplast. Transcriptomics analysis of model organisms is a powerful tool used in order to investigate mRNA changes under a special treatment or condition, and how it is different from the already known mRNA status (132, 133). On the other hand, transcriptomics in human pathologically changed tissues have a tremendous impact on collected knowledge about expression profile and/or response to therapy (134, 135). Alternative to model-organisms is a broad spectrum of *in vitro* studies in different types of cell cultures. Besides this, it is possible to have an animal model of a disease. One of the most applicable and widely used is the transcriptomics profile of cancer affected patient (136), where it is relatively easy to collect tissue. Those results may have impact to improve cancer therapy (137).

1.2.3. Transcriptomics analysis of non-model organisms

In contrast to model organisms, non-model organisms are not well studied. With the development of next generation sequencing (138), numerous possibilities are unlocked for the benefit of evolutionary biology, phylogenetics and population genetics (139). Depending on the type of desired analysis, there are several options of sequencing that can be chosen (Figure 11). For example, whole genome sequencing is applicable to simple biological entities such as bacteria, in order to study evolutionary responses (140), while targeted sequencing, like sequencing of ultra-conserved elements, can be used for comparative genomics studies between large species radiations (141, 142). Transcriptomics analysis of non-model organisms is quite challenging, especially if there is no known genomics background of species. Todd et. al. have shown that

biological variance is more expected in field studies of certain species than in *in vitro* experiments (143).

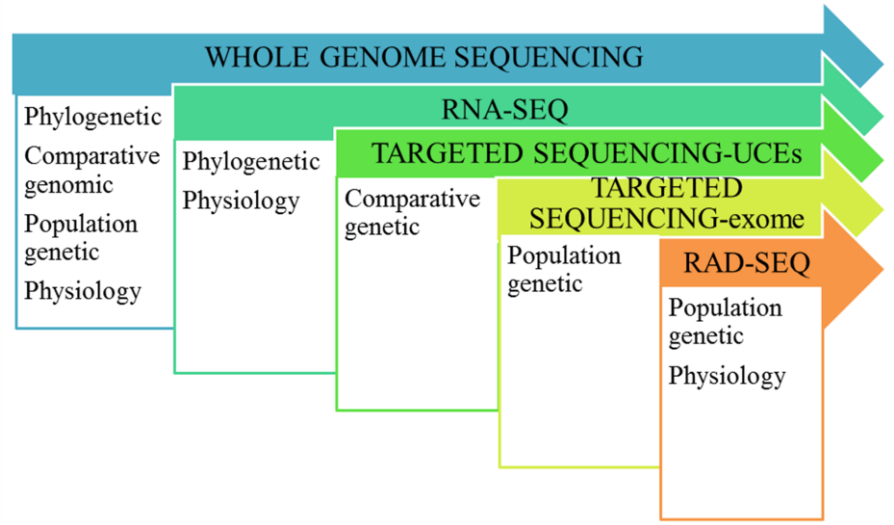


Figure 11. Application of different sequencing techniques depending on requirements of research field (144)

1.3. *Arion vulgaris*

1.3.1. Morphology and phylogeny

Arion vulgaris (Figure 12) was described for the first time by the French naturalist and zoologist Christian Horace Benedict Alfred Moquin-Tandon in 1855. *Arion vulgaris* belongs to phylum Mollusca, class Gastropoda, family Arionidae. Distribution of the family Arionidae is Holarctic and Oriental. The main characteristics of the genus *Arion* (over 40 species) are the absence of internal shell and keel. The animals are 80-150 mm long, brownish or reddish-brown in colour. The caudal mucous pit is located on the tail. The mantle is positioned on the anterior part and covers only one part of the body (145). Like all Pulmonata, *Arion* is a hermaphrodite and the morphology of its reproductive system is a key characteristic to distinguish different species of the genus.

Figure 12. *Arion vulgaris* (*Arion rufus* var. vulgaris Moquin-Tandon, 1855, Arion lusitanicus auct. non Mabille)

1.3.2. Ecology and epidemiology

Habitat of the species *Arion vulgaris* are natural (rivers and lakes, forests or dry valleys, mountains – up to 1700 m), as well as cultivated ecosystems (horticultures) (145). Only a small number of animals can be considered as natural enemies (hedgehogs, domestic ducks, some amphibians, reptiles and beetles) of *Arion vulgaris*. Nevertheless, according to DAISIE (Delivery Alien Invasive Species Inventories for Europe), *Arion vulgaris* was classified among 100 worst pests in Europe in 2006. As pests, slugs are characterized by high reproduction rate, high survival rate, sticky mucus and a large body size for a snail. *Arion vulgaris* can live up to two years. According to the research analysis, it has a huge effect on the economy, ecology and health (146). The economic impact is still not precisely calculated, but it clearly represents a huge problem in private gardens and agriculture. Research published in American Malacological Bulletin, from 2009, reports that costs from alien species invasions in United States, United Kingdom, Australia, South Africa, India, and Brazil have been estimated at US$314 billion per year (147).

Last year's numerous studies have been reporting that *Arion vulgaris* may be a vector for pathogenic species, such as *Clostridium botulinum* (148) and *Listeria monocytogenes* (149). Those findings play a crucial role in agriculture, because they have reported contamination of silage food for cattle. Calculated borderline for the infection is approximately 50 animals per square meter (150). Botulinum neurotoxin

(BoNT A-F) blocks the acetylcholine release in the motor-nerve connection and leads to paralysis and death. *Clostridium botulinum* is very heterogeneous and it is classified into four strains where we can recognize different types of BoNT (C, D - cause of animal botulism; A, B, E and F - human botulism; A and B - bovine botulism outbreaks) (151, 152). *Listeria monocytogenes* is a gram positive bacteria that has an effect on the central nerve system (CNS) and can cause encephalitis. Besides this, it can also cause enteritis, mastitis and septicemia (149). In ruminants, listeriosis is a usual silage associated disease. It has an indirect effect on human listeriosis, due to the fact that fresh milk has been used for production of cheese (153).

In this thesis, we present results from parallel transcriptomics and proteomics analysis of *Arion vulgaris* mantel, in order to identify proteins responsible for its successful survival strategies.

1.4. *Drosophila melanogaster*

Drosophila melanogaster, also known as fruit fly or vinegar fly, belongs to the class Insecta, order Diptera and the family Drosophilidae. It has been used as a model organism for many fields of research, due to the following characteristics - undemanding keeping, presence of sexual dimorphism, 4 pairs of chromosomes, easy and quick reproduction (154). Since 1970s and 1980s, *D. melanogaster* has been a key model organism to uncover gene function, understand development control and genetic origin of disease. Around 75% of human disease genes have homologues in *Drosophila*, such as dysmorphology (155) (single bone in zeugopod, apert syndrome, crouzon syndrome and pfieffer syndrome (156), etc.), cardiac disease (congenital heart disease, venous malformations (157)), neurological disorders (158) (Huntington disease, Parkinson disease (159), Alzheimer disease (160)), cancer (endometrial carcinoma, tuberous sclerosis, melanoma (161)). Explanation for similarity of genetic pathways is that during evolution the control of those processes was intact. Results from investigations of *Drosophila* can be directly applied to vertebrate systems (162).

Including the above mentioned, this invertebrate model system was particularly interesting for the investigation of potential drugs and their mechanisms.

While the genome of *Drosophila* has been decoded and well annotated, 63% of proteins (9124 proteins) were covered by a few proteomics studies (163). 2D analysis of fly heads and brains show approximately 90% of sequence coverage and according to this, heads can be considered for studies of central nervous system (Figure 13) (164). Due to the limitations of 2D electrophoresis, only a few membrane proteins were identified, so studies are still relying on computational predictions.

During work on this thesis, we were focused on membrane proteins, since they are usually drug targets and they have not been successfully investigated until this point. Gel-free bottom-up proteomics approach was used for enriched membrane fraction from fly heads.

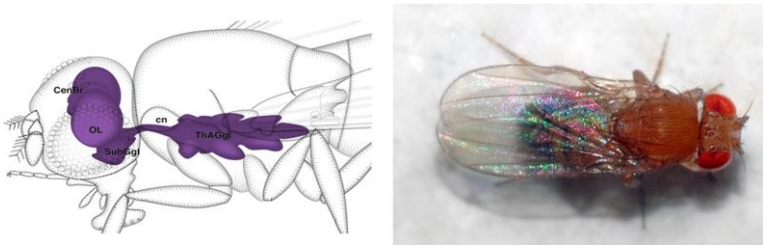


Figure 13. *Drosophila melanogaster*, anatomy of an adult CNS (CenBr- central brain hemispheres; OL- olfactory Lobe; SubGgl- suboesophogael ganglia; cn- cervical connective; ThAGgl- thoracic ganglia; according to Atlas of Drosophila Development by Volker Hartenstein published by Cold Spring Harbor Laboratory Press, 1993) and external appearance

2. AIMS OF THE STUDY

During work in this thesis, we defined the following aims:

a) Whole body transcriptome of *Arion vulgaris*, in order to create an mRNA-derived protein database

b) Proteomic profiling of *Arion vulgaris* mantel using nano-LC-ESI-MS/MS analysis.

b) Identification of membrane associated and transmembrane proteins from *Drosophila melanogaster* heads using gel-free MS/MS approach

3. PUBLISHED PAPERS

Results presented in this thesis are published in two peer-reviewed journals:

1. Bulat T, Smidak R, Sialana FJ, Jung G, Rattei T, Bilban M, et al. *Transcriptomic and Proteomic Analysis of Arion vulgaris--Proteins for Probably Successful Survival Strategies?* PLoS One. 2016;11(3):e0150614

2. Aradska J, Bulat T, Sialana FJ, Birner-Gruenberger R, Erich B, Lubec G. *Gel-free mass spectrometry analysis of Drosophila melanogaster heads*. Proteomics. 2015;15(19):3356-60

4. RESULTS AND DISCUSSION

4.1. *Arion vulgaris* - whole body transcriptome

Since there are no information about genome or transcriptome, we used RNAseq and *de novo* assembly in order to create the *Arion vulgaris* database. Using Illumina HiSeq 2000 technology we managed to create a cDNA library which contains 339 millions of paired end reads. Using Trinity pipeline (165) assembly of *de novo* transcript resulted in 136,406 contigs. 50% of sequences are 971 base pairs (bp) or longer (NP50), while average length was 671 bp. Combining TransDecoder tool with PFAM version 27.0, we identified 53,523 protein-coding sequences (CDS), illustrated in Figure 14.

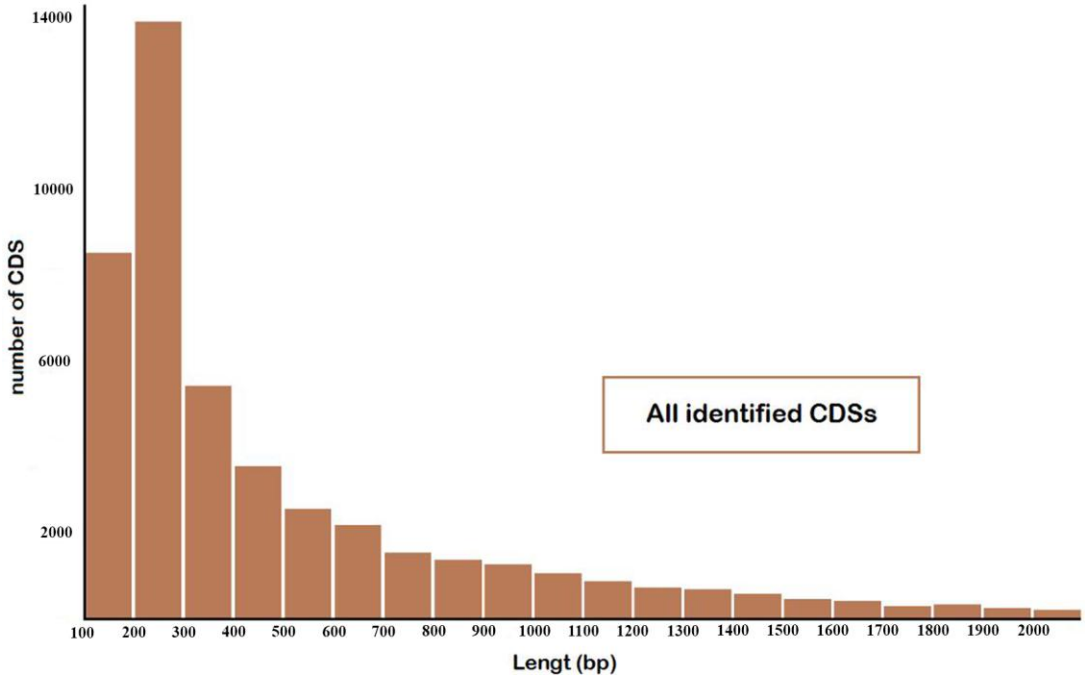


Figure 14. Distribution of 53,523 protein-coding sequences based on length

In order to reach final assembly, all CDS annotated contigs were analysed against the NCBI non-redundant (nr) protein base. In this analysis we used FastAnnotator and cut-off value was set up to 1E-6 (166). Duplicates and substrings were removed from the translated sequence dataset and resulting protein sequences were post-assembled using an in-house script to reduce redundancy. In Figure 15 arrangement of number of CDSs with hits in nr protein data base (A) and without hits in nr protein database (B) is

presented. The final list of CDSs that we published contains 47,451 non-redundant protein sequences. The full list of non-redundant protein sequences is available as supplementary material (Table S1) published in Bulat T. et al. (167). The sequence data has been submitted to the sequence read archive (69) database of GenBank (http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession number PRJEB7891.

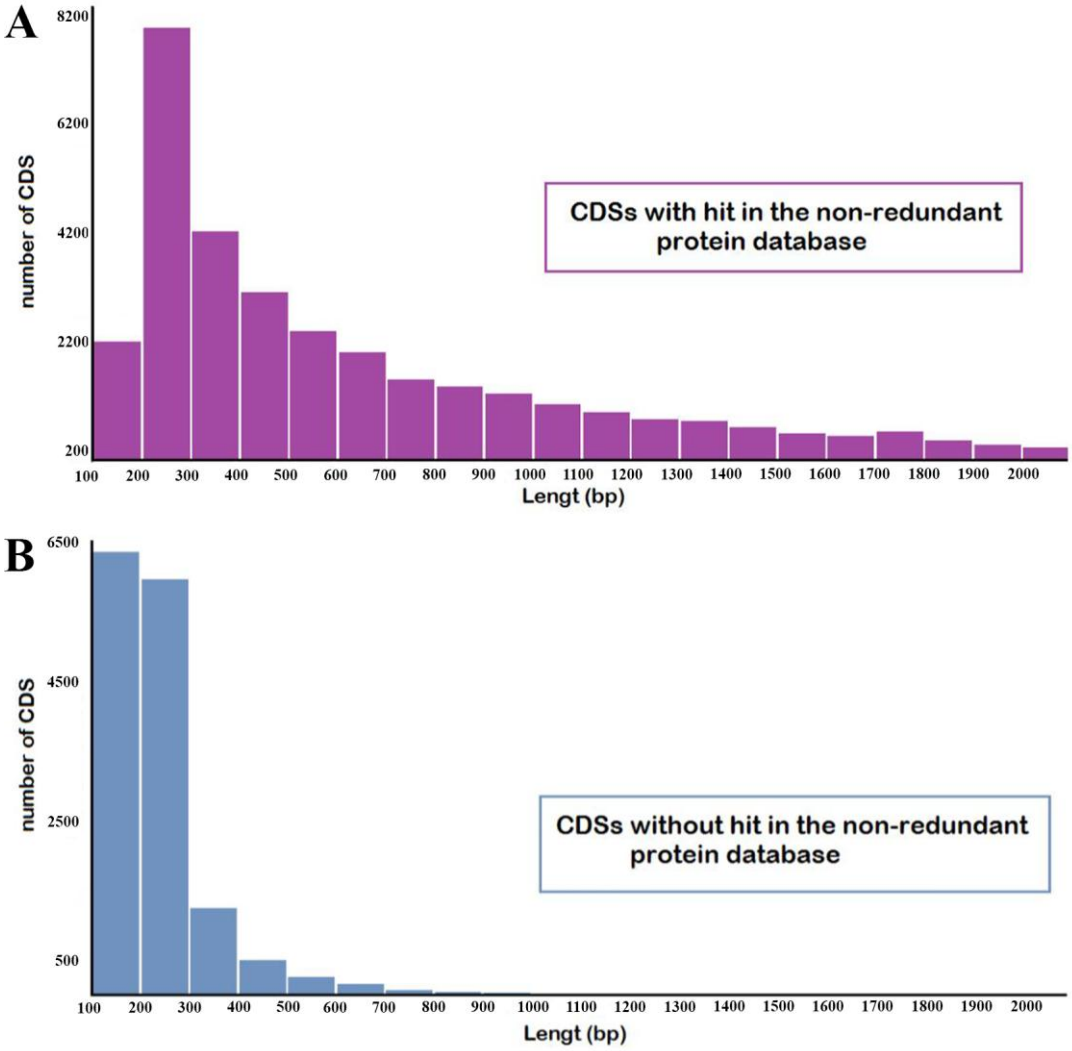


Figure 15. Distribution of number of CDSs with hits in nr database (A) and CDSs without hits in nr database

The analysis of final assembly showed that 1.94 sequences share the same hit against the National Center for Biotechnology Information (NCBI) nr database (Figure 16A), due to sequencing faults and/or possible isoforms of protein. After

comparison with non-redundant data base of the NCBI we obtained the following results: 63.3% of transcriptome, which is 31,463 transcripts, was assigned to at least one protein, while 33.7% (15,988 transcripts) was with no significant hits. The potential reason could be the lack of data for mollusc species.

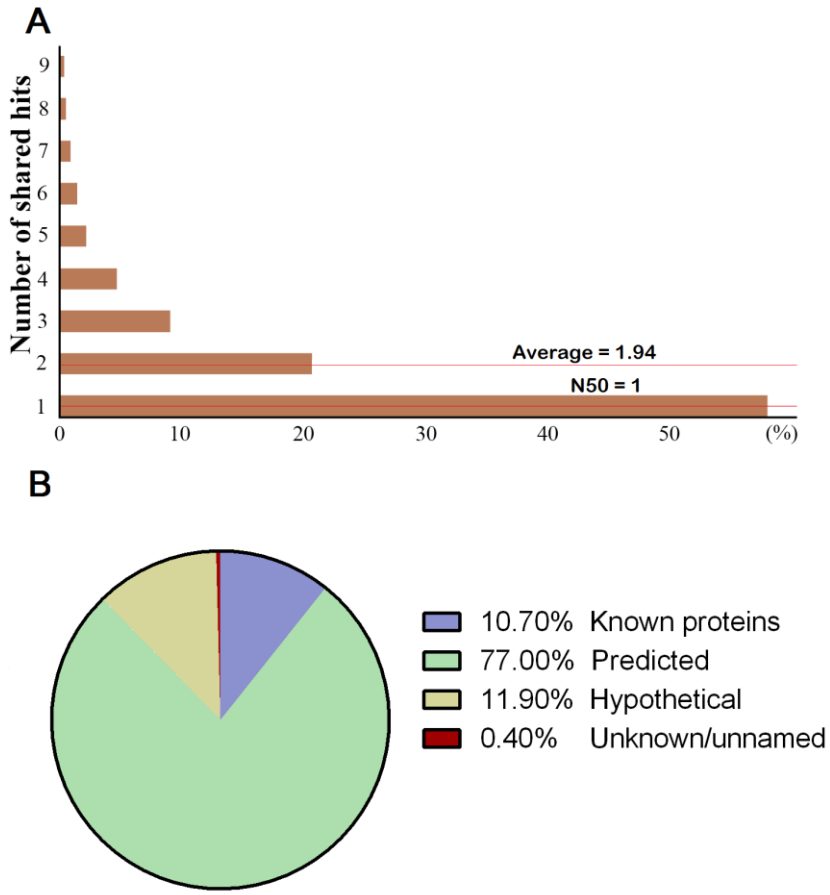


Figure 16. Specification of final assembly of *A.vulgaris* transcripts in comparison with the non-redundant NCBI database (A) and distribution of proteins in comparison with nr NCBI database (B)

Among 31,463 matches, 10.7% (3,380 transcripts) are known proteins, 77% (24,217 transcripts) are predicted, and 11.9% (3,739 transcripts) are hypothetical. Finally, 0.4% (127 transcripts) represents unknown or unnamed proteins (Figure 16B).

From the data that we acquired, one part of sequences, without similarity in the nr database (22,055 transcripts < 300 bp - 46.5% of final assembly), may be

referred as novel genes, ncRNAs or small open reading frames. Kaessmann suggests that novel genes might be a part of a non-coding genome and/or might be fragments of taxonomically ancestral constricted genes (168). Cech and Steitz show that ncRNAs and small open reading frames could be a reason of sequences lacking detectable homology with protein sequences in nr databases (169).

Data mining in genomics is based on functional annotation of individual sequences with potential biological function. According to BLAST2GO tool of FastAnnotator, which is one of the automatic annotation software (170), 39% of transcriptome (18,561 sequences) were functionally annotated with Gene Ontology (GO). The results were grouped into categories "biological process" (33.9%), "molecular function" (28.2%), and "cellular component" (34.8%). The processes with the largest number of sequences in biological process group are: cellular, metabolic, biological regulation, and a response to stimulus. Full list of annotated "biological process" is given in Figure 17.

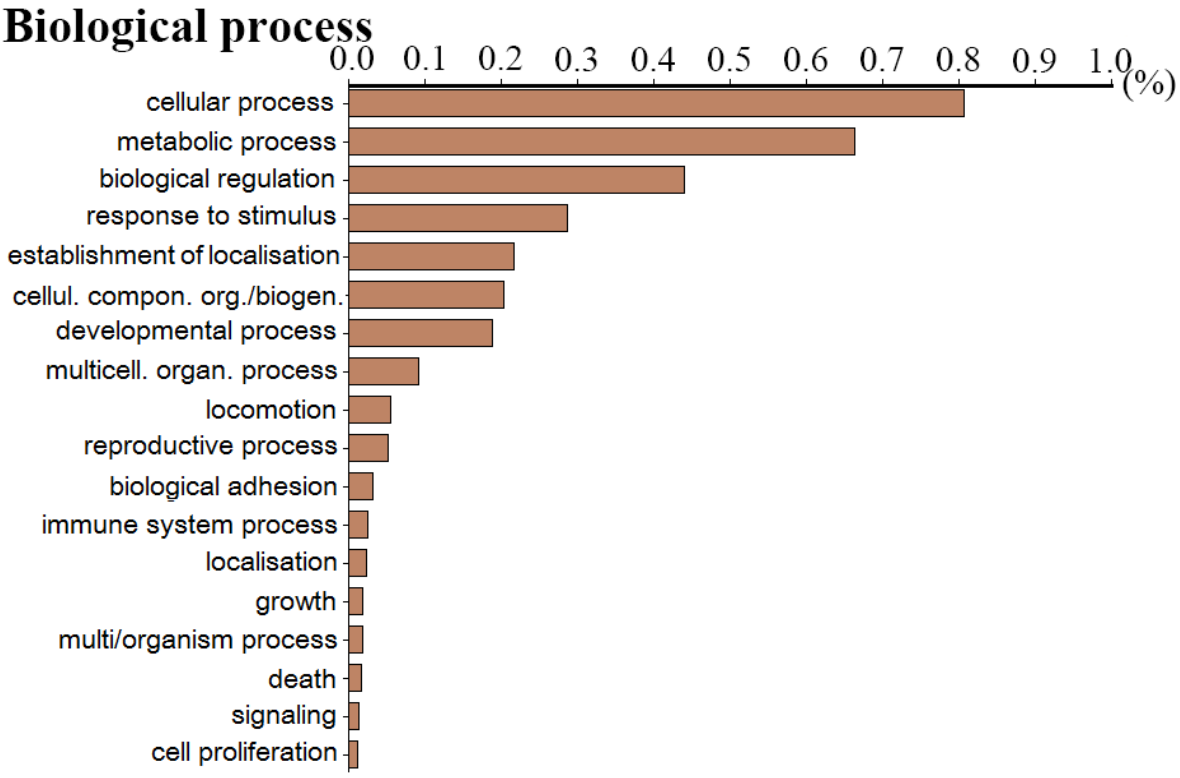


Figure 17. List of annotated biological function according to Gene Ontology (GO) analyses of the *A. vulgaris* transcriptome

The categories which are annotated as "molecular function" and "cellular component" are listed in Figure 18. The highest percentage of GO terms in the „molecular function" group corresponds to binding, catalytic or transporter activity, while in the „cellular component" group these are organelle/organelle part, macromolecular complexes or extracellular regions.

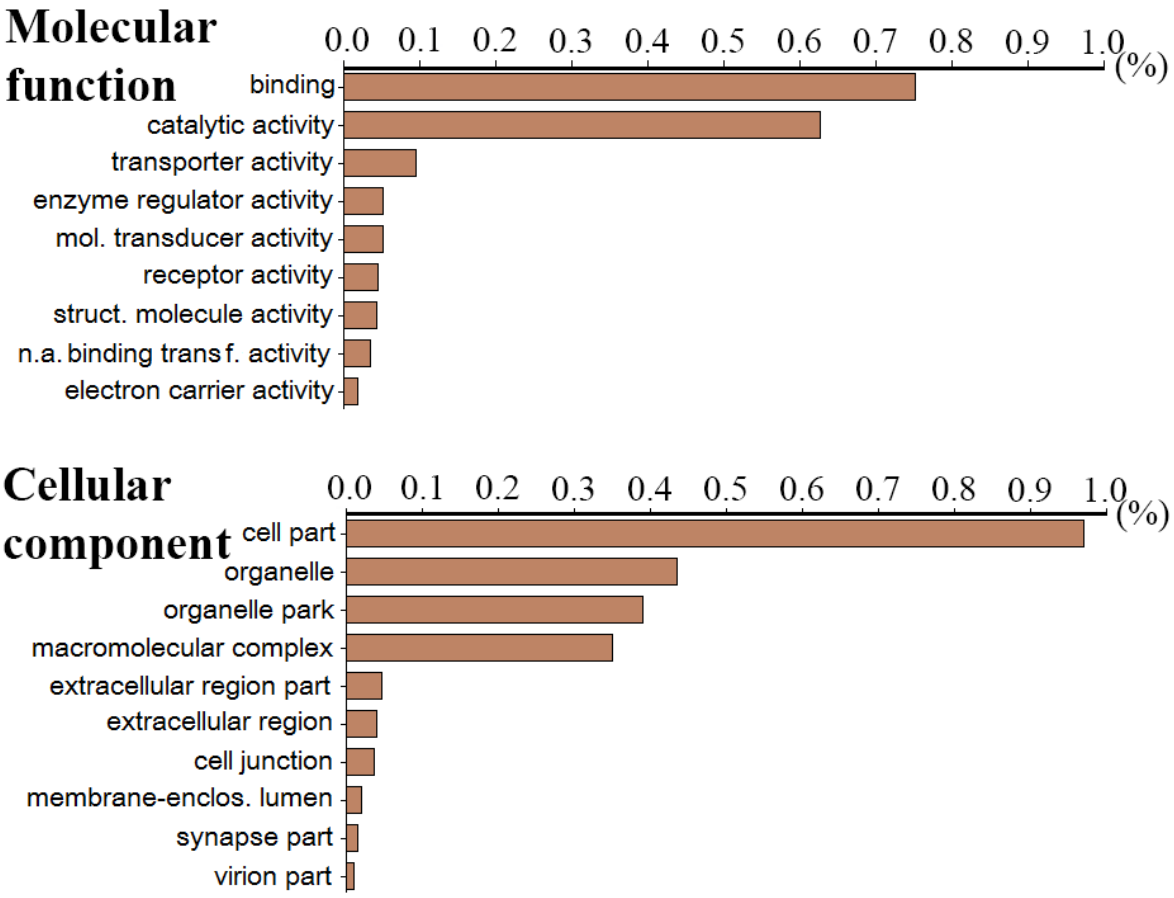


Figure 18. List of annotated molecular function and cellular component according to Gene Ontology (GO) analyses of the *A. vulgaris* transcriptome

A protein domain represents a functional part of a protein, and diverse types of domains can be found in nature. Identification of protein domains can contribute to the understanding of protein function. Many sequence analysis and protein folding algorithms used in computational biology have been built on hidden Markov models (HMMs). This powerful model is now used worldwide (e.g. Human genome project), since laboratories are not able to process the amount of new information so

fast. The Pfam database, created by European Bioinformatics institute contains a large collection of protein families, as well as higher-level groupings of related entries, known as clans. After we ran the transcriptomics data against Pfam database, in 16,994 of entries at least one domain was recognized, and of those entries, 471 sequences was without blastX hit annotation. BlastX is a tool which has a function to compare a nucleotide query sequence translated in all reading frames against a protein sequence database. Three most abundant domains are zinc-finger double domain (zf-H2C2_2), domain of unknown function (DUF4200) and ankyrin repeat (Ank_5). Full list is available as a supplementary material S1 table published in paper by Bulat T. et al. (167), under Pfam numbers 13465, 13863 and 13857. According to output from GO, 1,796 of transcripts were assigned with EC (Enzyme Commission) numbers which specify enzyme-catalysed reactions.

Zinc-finger proteins represent a very structurally diverse group of small proteins with a broad range of functions, such as replication, repair, signalling, cell proliferation and apoptosis. Zinc ion is crucial for the stability of this domain (171). These proteins are usually interaction modules and have a role in binding small molecules, proteins and amino acids. Zinc-finger proteins are divided into few groups according to the type of the domain: Cys2His2, gag-knuckle, treble-clef, zinc ribbon, Zn2/Cys6 and Miscellaneous. Zinc-finger double domain (zf-H2C2_2) belongs to Cys2His2 group and contains an α-helix and an antiparallel β-sheet. Position of zinc ion is organized by two histidine residues and two cysteine residues (Figure 19). These amino acid residues can be present in a protein in various numbers (two, four, six times, etc.).

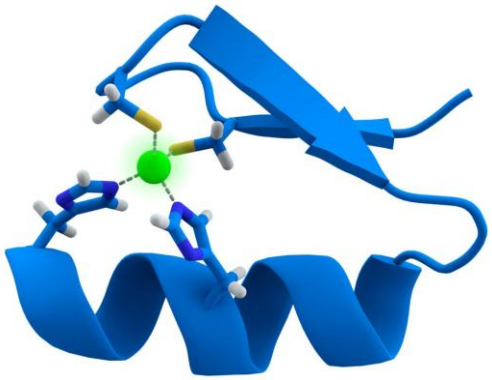


Figure 19. Illustration of zinc-finger domain structure where zinc-ion is stabilize by two histidine residues (dark blue) and two cysteine residues (yellow)

Unlike DUF4200, whose function is unknown, ankyrin repeat domain is found in proteins which have been connected with, among others, transcriptional initiators, ion transporters and signal transducers. Ankyrin domains are among the most common structural motifs within known proteins (172).

Finally, we compared our slug database in blastX in order to check best-match species. Results are presented in Figure 20. 65% of sequences is shared with *Aplysia californica*, 8.7% with *Crassostea gigas* and 2.5% with *Capitella teleta*. More than 80% of the sequences best matched with molluscs species. Some of the transcripts were shared with plants, fungi, bacteria and viruses. These results may also be due to the contamination during sample preparation. Also, it should be taken into consideration that more functional data on these species should be obtained.

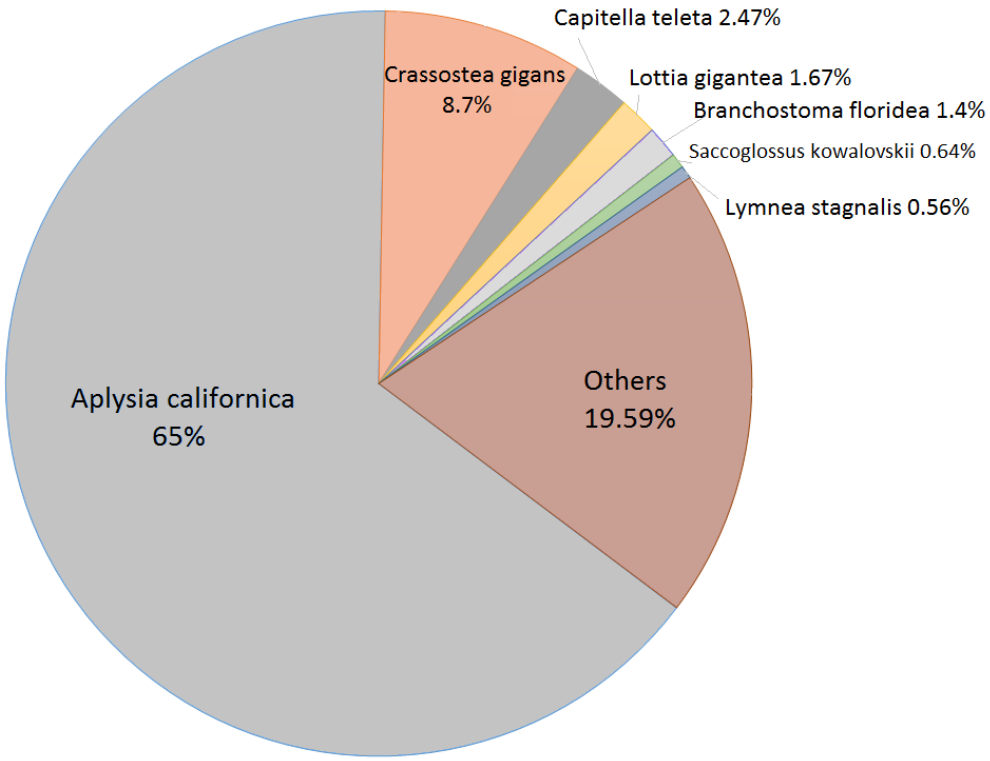


Figure 20. Results of blastX hits against the nr database in comparison with other species

4.2. *Arion vulgaris* mantle proteome

Gel-based proteomics analysis was performed from *A. vulgaris* mantel. Samples were run on 2-DE and 1-DE gel. From 2-DE gel (Figure 21) 356 spots were clearly resolved and trypsin digested. Most of the protein spots in the whole broad range of molecular weight (10-250 kD) are also in the range from pH 5 to pH 9. After running tryptic digested peptides and performing a search against transcriptome-based database, we identified 833 unique proteins, from which 814 had a significant hit in NCBI nr database. Besides this, 19 proteins are with no significant homology in nr NCBI database (Table 3). Some of the novel proteins are detected in unexpected different positions in gel due to possible isoforms and/or posttranslational modifications. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (173) via the PRIDE partner repository with the dataset identifier PXD002078 and 10.6019/PXD002078. Full list of identified proteins is given in Appendix (Table 1).

Due to the limitation of 2-DE based proteomics, we also ran samples on 1-DE gel, where they were digested and analysed on LC/MS-MS with higher resolution, in order to compensate low separation by 1D-SDS-PAGE. The full list of proteins with hits in nr NCBI database is given in Appendix (Table 2). We identified 2011 proteins from the two biological replicates, from which 48 proteins are without a hit on nr NCBI database (Table 4). Total number of identified proteins from 1-D and 2-D gels, with minimum of one unique peptide and distinct peptides, is 2129. Using a target-decoy search strategy, we estimated a false-positive assignment around 1%. Since the mantel is composed of muscle and skin, a high number of peptides were assigned to paramyosin, myosin, actin and arginine kinase, as well as to others cytoskeletal and muscle-related proteins. Arginine kinase has a highly significant role in invertebrates, since its role is in maintenance of ATP levels, which is achieved by the phosphorylation of phosphagens (174).

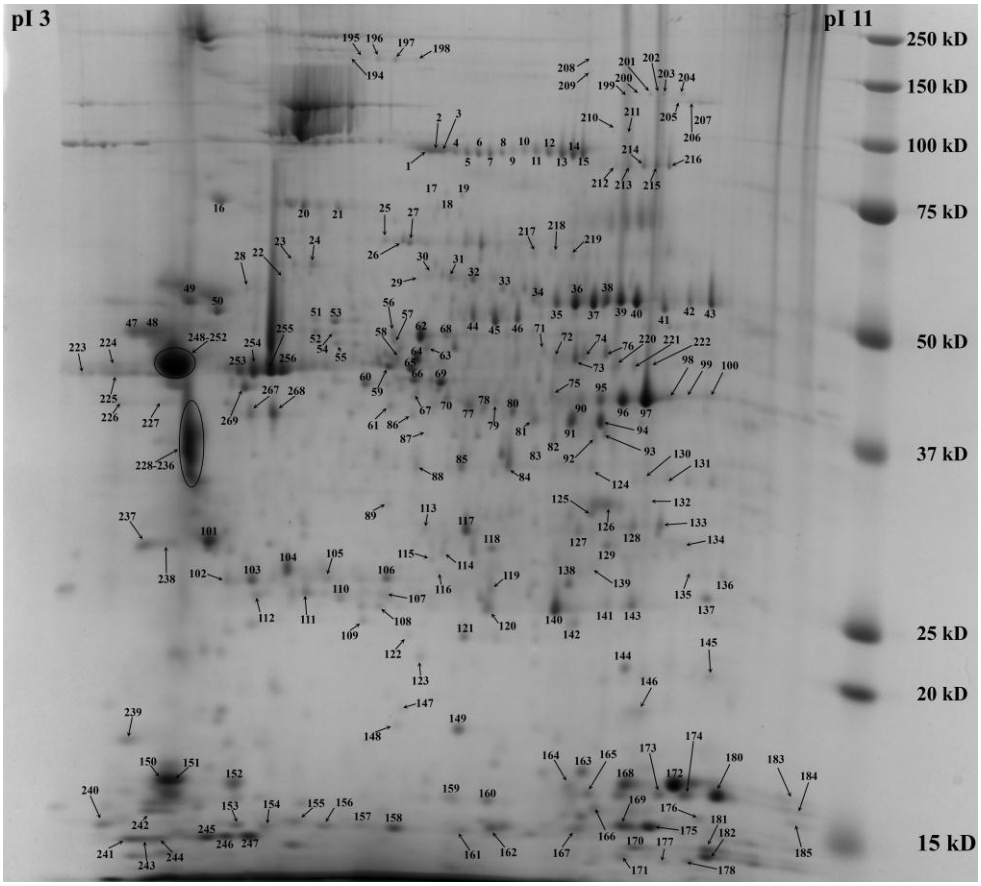


Figure 21. Representative 2-DE gel from *A. vulgaris* mantel proteome study. Numbers represent spots selected for MS analysis

Genus *Arion* belongs to the phylum Mollusca, and it can be considered as the most successful in therms of microbacterial resistance, taking into consideration the fact that molluscs lack adaptive immunity. During their life, they are in contact with many pathogens, including bacteria, fungi, viruses and several lineages of specialized eukaryotic parasites. Body of slugs is protected by mucus-producing epithelium that represents the first physical barrier which defends the organism against pathogens.

Table 3. Detailed information on 19 proteins identified from 2-DE proteomics study without significant hit on NCBI non-redundant database database (acc_embl- EMBL accession number; prot_acc- UniProt accession number; uni_pep- number of unique peptides; prot_cov- protein coverage (%); num_mat- number of assigned peptide matches; num_seq- number of protein sequences; prot_score- MASCOT score)

acc_embl	prot_acc	uni_pep	prot_cov(%)	num_mat	num_seq	prot_score	spot number
HACG01047575	A0A0B7BQP0	26	43	26	122	1681	34,118,180,181,182,183,116,178, 251,252,268,49,86,89,114,102,67,298
HACG01032016	A0A0B7ACZ5	9	51	9	83	1050	240,241,243,244,245,334,338,340
HACG01029986	A0A0B7AA27	5	19	5	29	484	162,228,230,231,232,233,235,236, 229,234,249,250
HACG01019436	A0A0B6ZEL1	5	46	10	53	1374	56,57,63,62
HACG01023202	A0A0B6ZNW9	4	18	4	32	300	16,49,18,105,109,110,111,194,196,19,7,200, 207,211,212,213,98,100,101,103,282,289
HACG01026720	A0A0B7A0U9	4	28	4	56	1389	240,241,242,243,244,245,246
HACG01044712	A0A0B7BE64	4	21	4	7	89	237,238
HACG01005312	A0A0B6Y748	3	20	3	7	95	156,157
HACG01033711	A0A0B7AKN6	3	39	3	10	116	207,213,214,215
HACG01039132	A0A0B7AZH5	3	17	3	20	276	160,163,164,168,183,169,173,174,176,181
HACG01044724	A0A0B7BE79	3	30	3	11	167	231,235,236,232,233,261
HACG01005243	A0A0B6Y7L6	2	7	2	2	53	67
HACG01013119	A0A0B6YUN6	2	8	2	6	130	116,119,139,142
HACG01028842	A0A0B7A4Y5	2	15	2	6	181	244,245,246
HACG01032167	A0A0B7AEI5	2	3	2	2	45	194
HACG01036973	A0A0B7AVD5	2	20	2	3	59	240
HACG01043414	A0A0B7BA86	2	11	2	4	57	156,158,348
HACG01044713	A0A0B7BF10	2	10	2	2	30	102
HACG01019435	A0A0B6ZEY8	1	23	4	5	72	62

Table 4. Detailed information of proteins identified from 1-DE proteomics study without significant hit on NCBI non-redundant database (acc_embl- EMBL accession number; prot_acc- UniProt accesion number; uni_pep- number of unique peptides; prot_cov- protein coverage (%); num_mat- number of assigned peptide matches; num_seq- number of protein sequences; prot_sc- MASCOT score)

acc_embl	prot_acc	uni_pep	prot_cov	num_mat	num_seq	prot_sc
HACG01047575	A0A0B7BQP0	21	73	21	218	4285
HACG01026043	A0A0B6ZWD9	7	17	7	30	766
HACG01026345	A0A0B6ZZP9	7	30	7	30	365
HACG01032167	A0A0B7AEI5	7	20	7	7	96
HACG01044712	A0A0B7BE64	7	40	7	35	727
HACG01044713	A0A0B7BF10	7	30	7	25	443
HACG01026720	A0A0B7A0U9	6	32	6	46	540
HACG01013119	A0A0B6YUN6	5	26	5	15	124
HACG01033711	A0A0B7AKN6	5	62	5	14	282
HACG01036259	A0A0B7APZ4	5	22	5	6	146
HACG01005318	A0A0B6Y7S7	4	16	4	4	137
HACG01023202	A0A0B6ZNW9	4	21	4	32	373
HACG01026996	A0A0B7A183	4	21	4	8	92
HACG01028842	A0A0B7A4Y5	4	29	4	19	447
HACG01033499	A0A0B7AKC1	4	8	4	10	164
HACG01036974	A0A0B7AS42	4	34	4	7	82
HACG01039132	A0A0B7AZH5	4	36	4	17	752
HACG01021092	A0A0B6ZHY7	3	17	3	3	44
HACG01023330	A0A0B6ZR28	3	25	3	4	77
HACG01028841	A0A0B7A6R5	3	32	3	7	116
HACG01030092	A0A0B7A8J9	3	30	3	13	381
HACG01030913	A0A0B7AAF1	3	33	3	6	125
HACG01044700	A0A0B7BHE4	3	19	3	6	117
HACG01044706	A0A0B7BGZ4	3	22	3	8	220
HACG01044707	A0A0B7BFB4	3	29	3	10	274
HACG01019435	A0A0B6ZEY8	3	44	14	37	616
HACG01019436	A0A0B6ZEL1	3	55	15	44	760
HACG01002111	A0A0B6XY56	2	41	2	6	109
HACG01005312	A0A0B6Y748	2	14	2	2	36
HACG01026042	A0A0B6ZX56	2	5	2	4	97
HACG01027525	A0A0B7A2T2	2	50	2	3	48
HACG01030091	A0A0B7AAE4	2	31	2	15	358
HACG01032016	A0A0B7ACZ5	2	15	2	7	100
HACG01033415	A0A0B7AK29	2	30	2	4	39
HACG01035358	A0A0B7AN13	2	12	2	2	75

acc_embl	prot_acc	uni_pep	prot_cov	num_mat	num_seq	prot_sc
HACG01036973	A0A0B7AVD5	2	25	2	4	116
HACG01044893	A0A0B7BFI7	2	2	2	3	32
HACG01047662	A0A0B7BNQ7	2	8	2	2	44
HACG01052636	A0A0B7C2V6	2	33	2	2	37
HACG01044718	A0A0B7BF15	2	37	3	12	207
HACG01044724	A0A0B7BE79	2	50	3	18	271
HACG01044715	A0A0B7BHF6	2	32	5	17	364
HACG01042313	A0A0B7B827	1	8	2	2	33
HACG01044714	A0A0B7BE68	1	12	2	3	88

The modification of immune systems during evolution involves the expansion of particular gene families in given phyla. Interestingly, during the analysis of data that we obtained from a comprehensive transcriptomics and proteomics study, we showed that genes/proteins which are related to immunity are expressed. We identified whide-range of proteins which indicates an unexpected complexity of the immune system with a high grade of pathogen specificity and immune-priming. Some of them may play a role in pattern recognition, as effector proteins or cytokine-like proteins, such as lectins. In addition, we have also detected complement-like proteins, peptidoglycan-recognition proteins (PGRPs), lipopolysaccharide and β1, 3-glucan-binding proteins, fibrinogen-related proteins (FREPs), pore-forming membrane attack/perforin (MACPF) domain proteins and toxin-like proteins.

Lectins are very important in „self-non-self" recognition. Lectins that we detected (C-type lectin family, galectin, malectin, H-type lectins, L-type lectins, Ricin-B lectins and calnexin) contain conserved carbohydrate-recognition domains (CRDs). In total, lectin-associated functions may be attributed to 149 transcripts, 37 of which were unambiguously identified at the protein level. In Table 5 all lectin-like proteins which we detected on transcriptomics and proteomics level are presented, while all detected lectins at transcriptome level are given in Appendix (Table 3.).

C-type lectins are able to bind carbohydrate in a Ca^{2+}-dependent manner, thanks to CRDs which contain two disulphide bridges located in the base of two loops (175). 47 transcripts encoding proteins with C-type lectin CRDs were found and 14 were also identified at the protein level. 17 of 21 complete transcripts have a predicted signal

peptide demonstrating that they are secreted by cells to utilize their functions (Table 5). In Figure 22 we showed an amazing sequence variability within incilarin-like proteins identified in our study, while in Figure 23 we illustrated a domain structure of C-lectin-like molecules found in the transcriptome of *A. vulgaris*. Incilarins belong to C-type lectins. They were isolated from the surface of the body (mucus) of the land slug, *Incilaria fruhstorfer*. This protein was found in the water-soluble fraction and it is known that it possess hemagglutination activity (176).

Table 5. The list of lectin like proteins detected at the transcriptome (T) and protein level (P) with status- complete or incomplete as the entire CDS from the initial ATG to the STOP codon was represented.

prot_acc UniProt	Protein hit description in nr NCBI database	Status	Evidence	Type
A0A0B7AZQ2	PREDICTED: calnexin-like isoform X1 [Aplysia californica]	complete	T,P	calnexin
A0A0B7AWH4	PREDICTED: calnexin-like isoform X2 [Aplysia californica]	complete	T,P	calnexin
A0A0B7A018	incilarin A [Haliotis discus discus]	complete	T,P	C-type lectin
A0A0B6ZB52	Incilarin A [Meghimatium fruhstorferi]	complete	T,P	C-type lectin
A0A0B6ZSS2	Incilarin A [Meghimatium fruhstorferi]	complete	T,P	C-type lectin
A0A0B6ZT36	Incilarin A [Meghimatium fruhstorferi]	complete	T,P	C-type lectin
A0A0B7AGC8	Incilarin A [Meghimatium fruhstorferi]	complete	T,P	C-type lectin
A0A0B6Y7S8	Incilarin A [Meghimatium fruhstorferi]	complete	T,P	C-type lectin
A0A0B6ZSZ0	Aggrecan core protein [Crassostrea gigas]	complete	T,P	C-type lectin
A0A0B6ZQL8	PREDICTED: cell wall protein DAN4-like [Aplysia californica]	complete	T,P	C-type lectin
A0A0B7B8E9	Incilarin A [Meghimatium fruhstorferi]	complete	T,P	C-type lectin
A0A0B7BB09	putative perlucin 4 [Haliotis discus discus]	incomplete	T,P	C-type lectin
A0A0B6ZZL1	Incilarin C [Meghimatium fruhstorferi]	incomplete	T,P	C-type lectin
A0A0B7BCB3	Low affinity immunoglobulin epsilon Fc receptor [Crassostrea gigas]	incomplete	T,P	C-type lectin
A0A0B6Z234	PREDICTED: uncharacterized protein LOC101860995 isoform X1 [Aplysia californica]	incomplete	T,P	C-type lectin
A0A0B7BV24	PREDICTED: uncharacterized protein LOC101848467 [Aplysia californica]	incomplete	T,P	C-type lectin
A0A0B7AWV4	tandem repeat galectin [Biomphalaria glabrata]	complete	T,P	galectin
A0A0B7B7G5	CHAII like protein precursor [Cepaea hortensis]	complete	T,P	H-type lectin
A0A0B6Z3C5	agglutinin [Helix pomatia]	complete	T,P	H-type lectin
A0A0B6Y9P9	agglutinin [Helix pomatia]	complete	T,P	H-type lectin
A0A0B7AA27		complete	T,P	H-type lectin

A0A0B7BA86		complete	T,P	H-type lectin
A0A0B7AEX4	Chain A, Structure Of Helix Pomatia Agglutinin With No Ligands	incomplete	T,P	H-type lectin
A0A0B7ACW2	Chain A, Structure Of Helix Pomatia Agglutinin With No Ligands	incomplete	T,P	H-type lectin
A0A0B7ANH6	PREDICTED: vesicular integral-membrane protein VIP36-like [Aplysia californica]	complete	T,P	L-type lectin
A0A0B7ADM8	PREDICTED: malectin-B-like [Aplysia californica]	incomplete	T,P	malectin
A0A0B6XX19	PREDICTED: malectin-B-like [Aplysia californica]	incomplete	T,P	malectin
A0A0B6ZWR8	PREDICTED: uncharacterized protein LOC101856230 [Aplysia californica]	complete	T,P	C1qDC
A0A0B7A1M5	sialic acid binding lectin [Cepaea hortensis]	complete	T,P	C1qDC
A0A0B7A6S4	sialic acid binding lectin [Cepaea hortensis]	complete	T,P	C1qDC
A0A0B6ZJ78	sialic acid binding lectin [Helix pomatia]	complete	T,P	C1qDC
A0A0B7ABV6	putative C1q domain containing protein MgC1q41 [Mytilus galloprovincialis]	incomplete	T, P	C1qDC
A0A0B6ZN56	sialic acid binding lectin [Haliotis discus discus]	incomplete	T,P	C1qDC
A0A0B7ALE7	sialic acid binding lectin [Haliotis discus discus]	incomplete	T,P	C1qDC
A0A0B7A907	sialic acid binding lectin [Cepaea hortensis]	incomplete	T,P	C1qDC
A0A0B6ZHG1	sialic acid binding lectin [Helix pomatia]	incomplete	T,P	C1qDC
A0A0B7AJC2	PREDICTED: uncharacterized protein LOC101850618 [Aplysia californica]	complete	T,P	C1qDC

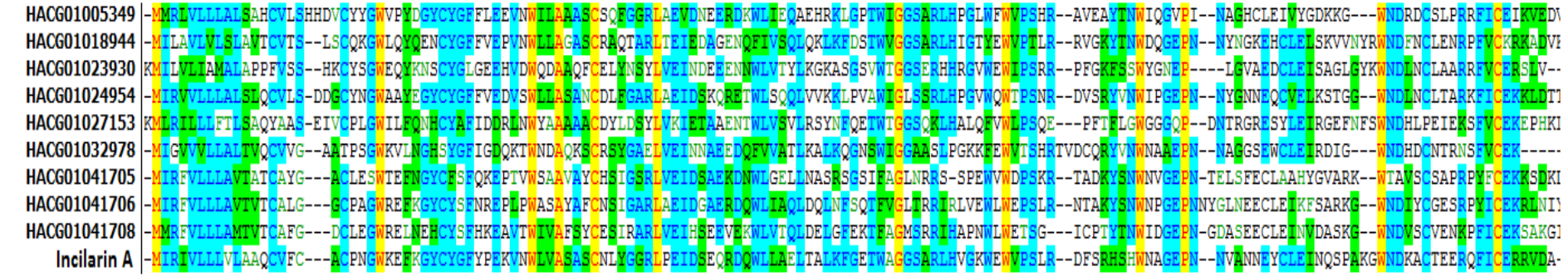


Figure 22. Alignment of amino acid sequences of incilarin-like proteins from *A. vulgaris*

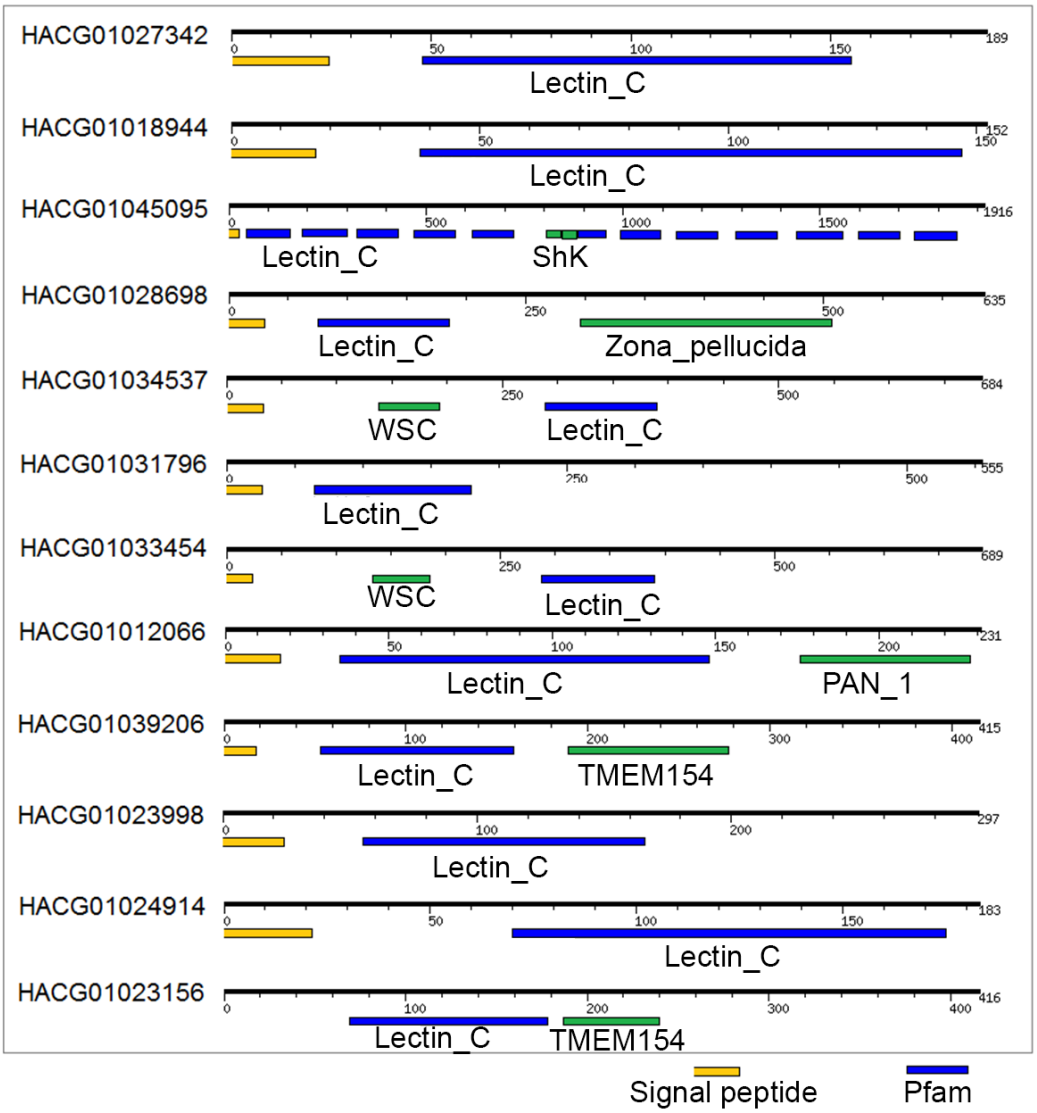


Figure 23. Domain architecture of Lectin_C-like proteins from *A. vulgaris*

Next to Lectin_C, some of the identified peptides contain domains recognized as ShK, WSC, Zona pellucida (ZP), PAN_1 and TMEM154. ShK is a 35 amino acid domain, cross-linked by three disulphide bridges between Cys3-Cys35, Cys12-Cys28, and Cys17-Cys32, which was first identified in a sea anemone *Stichodactyla helianthus* (177). Several independent studies show that the ShK domain blocks the voltage-gated potassium channels - Kv1.1, Kv1.3, Kv1.6, Kv3.2 and KCa3.1 (178-180). WSC (Wall Stress-responsive Component) domain is a cysteine-rich domain, structured by approximately 90 amino acids, and it represents a carbohydrate binding domain. It has been identified in fungi (*Saccharomyces cerevisiae, Aspergillus fumigatus, Aspergillus*

nidulans), mammalias and insects (referring to UniProt database). ZP domain consist of ~260 amino acids and it has an essential role in development, hearing, immunity, and cancer. It was distinguished in eukaryotes, in glycoproteins from secreted mucus (mammalian sperm receptors, transforming growth factor (TGF)-beta receptor III, deleted in malignant brain tumour-1 (DMBT-1) (181)). PAN (APPLE) domains also contains a cysteine rich structure with three or four bridges and it is associated with protein-protein interaction, or protein-carbohydrate interactions (182). The function of TMEM154 domain is still unknown. Experiments performed with mutation of TMEM genes in sheep indicate that it can be involved in a lentivirus immune response (183).

Transcriptomics analysis of *A. vulgaris* shows 4 calnexin proteins, from which 2 were confirmed at the protein level with a highest similarity to *Aplysia californica* homologs (blastp: E = 0.0; Identity = 70–74%) (Table 5). It contains a signal sequence and an additional C-terminal transmembrane helix (Figure 24). This protein belongs to an evolutionarily conserved calnexin/calreticulin-superfamily and it is connected to glycoprotein secretory pathways (184). Calnexin is one of the pattern recognition receptors and it has been characterised as a main protein in shrimp antibacterial immunity response (185).

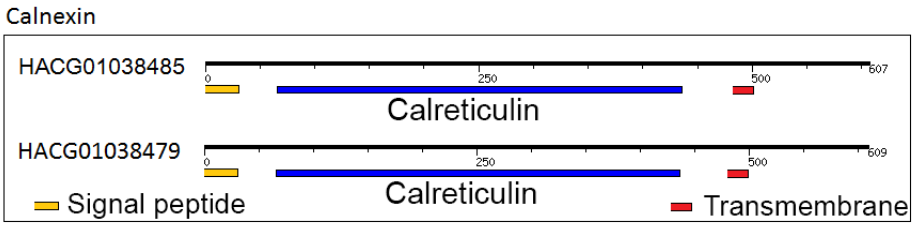


Figure 24. Domain construction of calnexin-like proteins from *A. vulgaris*

The H-type lectins (Figure 25) have a carbohydrate-recognition domain (CRD) structurally similar to *Helix pomatia* agglutinin (HPA) (186). We identified 12 transcripts and from those, 7 were also identified at the protein level (Table 5). These kind of lectins which are detected in *A. vulgaris* share the same size and hexameric arrangement of HPA, thanks to the cysteine residues involved in the intermolecular disulphide bridge construction. HPA function in *H. pomatia* is connected to innate immune system by protecting fertilized eggs from bacteria.

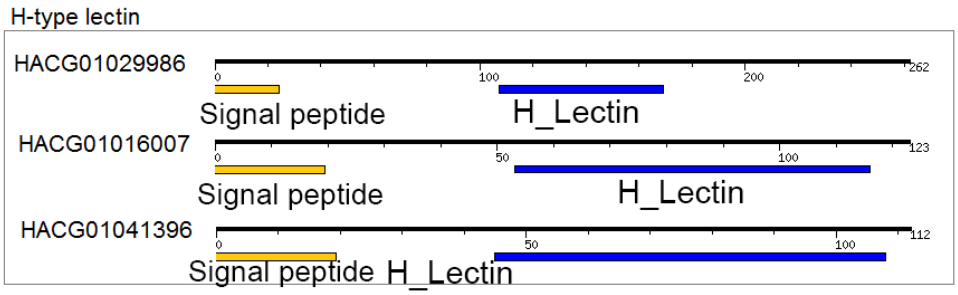


Figure 25. Domain architecture of three H-type lectins of *A. vulgaris*

Additionally, recent studies showed that HPA is able to recognize glycoproteins via O-GlcNAcylation (187). Alignment of amino acid sequences of H-type lectins from *A. vulgaris* with *Helix pomatia* and *Cepaea hortensis* is presented in Appendix (Figure 1).

We reported that ten unigenes coding proteins contain a galectin domain. All of these proteins are the most similar to gastropod galectins (*Aplysia californica*, *Biomphalaria glabrata*, etc. blastp E value ranges from 0 to 2E−32) (Appendix Table 3). Galectins belong to the family of structurally-related lectins which have a binding specificity towards β-galactoside residues. Proteins which contain a galectin domain are tandem-repeat type galectin (Figure 26). The absence of a signal sequence is consistent with other galectins, which are known to be dynamically secreted from diverse tissue via a "nonclassical" secretory pathway (188).

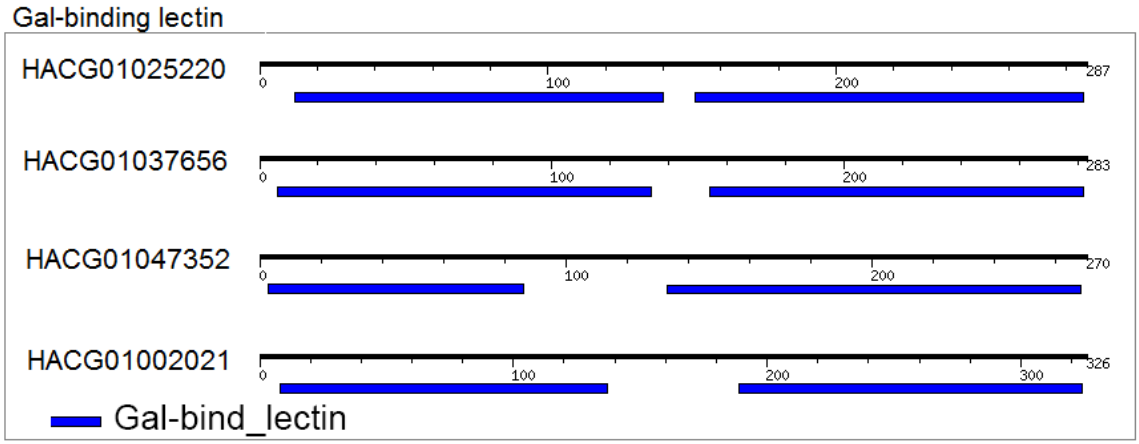


Figure 26. Tandem-repeat structure of galectin like proteins identified in *A. vulgaris*

In total, we detected three L-type lectin sequences. Two of them were at the transcriptome level with the highest similarity to homolog of ERGIC-53 protein from *Littorina littorea* (blastp: E = 0.0; Identity = 57–58%). At the transcriptome and protein level, we detected L-type lectin with VIP36-like homolog from *Aplysia californica*.

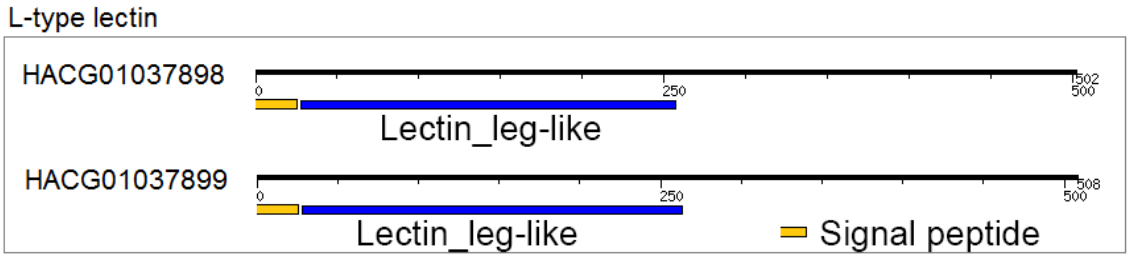


Figure 27. Domain architecture of two L-type lectins of *A. vulgaris*

Both ERGIC-53 and VIP36 are type I transmembrane proteins. ERGIC-53 (ER-Golgi intermediate compartment 53 kDa protein) homologs share the specific architecture with an extracellular signal peptide and a single N-terminal L-type carbohydrate binding site (Figure 27). VIP36 is located in Golgi and post-Golgi portions of the secretory pathway (189).

The R-type lectins belong to a superfamily of proteins which have a CRD structurally similar to the CRD in ricin - toxic, naturally produced lectin in the seeds of *Ricinus communis*. Analysis of the transcriptome predicts 31 unigenes coding proteins of R-type lectins (Table 3. Appendix), which consist of an A chain with enzymatic activity (galactosyltransferase) linked through a disulphide bridge to a B chain with lectin activity (Figure 28). We did not detect them at the protein level probably due to tissue-specificity. Between animal R-type lectins we can distinguish several domains, like mannose receptor (MR) family, EW29, pierisin-1 and pierisin-2. They are identified in the annelid (earthworm) *Lumbricus terrestris* and cabbage butterfly *Pieris rapae*. Pierisin-1 is a 98-kD protein (850 amino acids) that shows extreme toxicity to animal cells in culture (190). Besides this, in *Limulus* horseshoe crab, for example, a coagulation factor G contains two main domains: a central R-type lectin domain and a glucanase-like domain. This protein also has a subunit that is a serine protease, showing that in animals the combination of R-type domain with enzyme is common. The R-type CRD is conserved between animal and bacterial lectins (191).

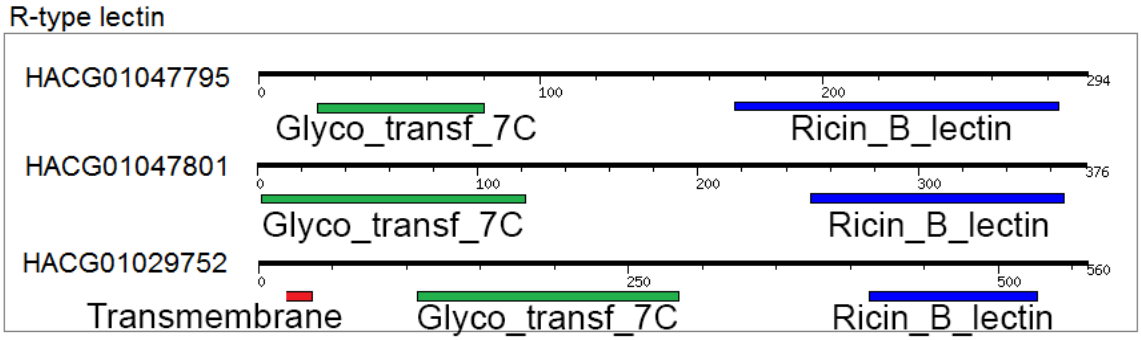


Figure 28. Structure of R-type-like lectin protein from *A. vulgaris*

Malectin is a membrane protein of the ER and it recognises and binds Glc2-N-glycan. At transcriptome and protein level we identified two proteins which have a high similarity with malectin-B-like proteins discovered in *Aplysia californica* (Table 5). The malectin sequence is well conserved in animals (192). The role of malectin in invertebrates is not studied enough, but experimental data from *Xenopus laevis* suggest that malectin could be an important part of the regulatory mechanism in the cleavage process of second glucose from Glc2-N-glycan and in the inhibition of the early entry to the calnexin/calreticulin cycle (CNX/CRT) (193). Taking into consideration that almost all proteins involved in innate and adaptive immune response are glycoproteins (194), and that CNX/CRT cycle ends once glycoproteins reach their native structures, we can postulate that malectin-like proteins detected in *A. vulgaris* can have a significant role in the immune response.

Tachylectins 5A and 5B (TLs-5) are lectins isolated from haemolymph plasma of a horseshoe crab *Tachypleus tridentatus* and they are able to agglutinate human erythrocytes, Gram-positive and Gram-negative bacteria (195). At the transcriptome level we detected three tachylectins like proteins, one TL-5A-like and two TL-5B-like (Table 5). Tachylectin-related proteins have been recognized in numerous organisms (slime molds, sponges, hydroid and bony fish). Their function is correlated with the immune system due to an antibacterial activity: binding to PAMP, phagocytosis of bacteria and binding to lipopolysaccharide (LPS) (196, 197).

Globular C1q domain is the main characteristic of the C1q domain-containing proteins (C1qDC). C1qDC possesses lectin-like properties and has a main role in innate immunity of bivalvia molluscs (198-201). They are involved in several immune responses, such as pathogen recognition (202), microorganism agglutination (203), phagocytosis (204) and mediating cell migration (205). Based on sequence homology and domain similarity, 33 C1qDC sequences from the *A. vulgaris* transcriptome may be classified as members of the C1q family (Table 5 and Appendix Table 3). Ten transcripts have also been identified at the protein level (Table 5). The most of the complete C1qDC sequences revealed a signal peptide or a transmembrane domain on the N-terminus (Figure 29).

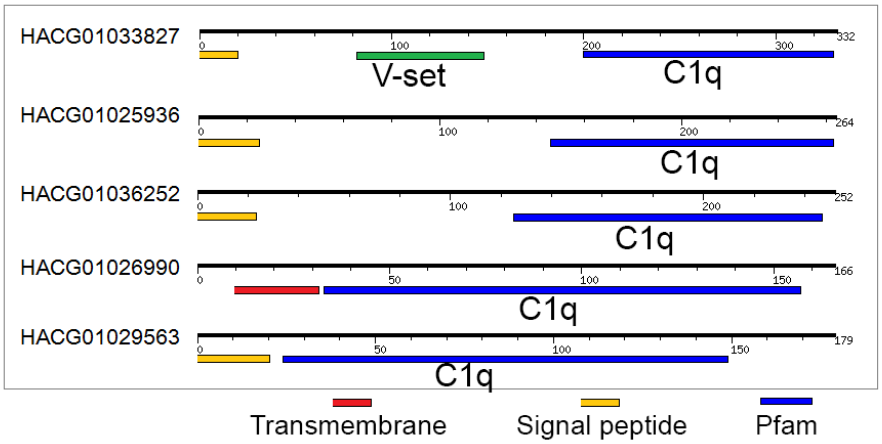


Figure 29. Structure of C1q domain-containing proteins protein from *A. vulgaris*

The C1q domain is involved in the neutralization of viruses, cell adhesion and clearance of apoptotic cells. It is considered as a particularly effective PRD with highly adaptive binding properties, due to the ability of the C1q domain to bind a variety of "self-non-self" ligands, including LPS, virus envelope proteins, phospholipids and some acute-phase proteins (206). Wang et al. suggest that C1qDC proteins could trigger an ancient complement system by the lectin pathway prior to the evolution of immunoglobulins (204).

Besides the C1q domain, C1qDC proteins that we detected, either at protein or transcriptome level, contain homology with V-set and sialic acid-specific binding lectin. V-set are immunoglobulin (Ig)-like domains and can be found in the antibody variable

domain. Sialic acids (Sia) are found mostly at the non-reducing end of oligosaccharide chains on glycoproteins and glycolipids (207), and invertebrate lectins contribute to the innate immune response through binding to Sia on foreign cells (208), like in *Crassostrea virginica* (209), *Tachypleus gigas* (210), *Aphonopelma chalcodes* (211), *Ornithodoros tartakovskyi* and *Ornithodoros tholozani* (212).

During evolution, animals developed different kinds of mechanisms for defence. Some of them have extremely developed escaping skills, while others developed diverse sets of proteins/peptides which they use against predators. Membrane Attack Complex/Perforin (MACPF) consists of proteins with two domains - membrane attack complex (MAC) and perforin (PF), and they belong to pore-forming toxins (PFTs) (213). Recent studies show that MACPF domains are associated with cholesterol-dependent cytolysis (CDCs), as well as bacterial PFTs (defence from viruses), which indicates that their mechanisms of action are similar: cholesterol in the target membrane will trigger a pore-forming process, which will lead to cytolysis (214). MACPF was discovered as a protecting factor of apple snail eggs against natural enemies (215). In our transcriptomics study of *A. vulgaris*, we found a few candidate effector genes which could be a part of biochemical defence system, including a family of potential pore-forming membrane attack/perforin (MACPF) domain proteins. We assume that together with lectins, they can be a part of an innate immune system. Besides them, in comprehensive transcriptomics data we found mRNA, which according to GO annotation, are coding proteins involved in phagocytosis (71 transcript), coagulation (122 transcript), and Toll-like receptors (50 transcript).

Aside from the importance of the immune system for protection of animals, toxins and toxin-like proteins are also essential. Analysis of *A. vulgaris* shows a novel toxin-like protein with significant similarity to agatoxin. Agatoxin is a toxin from spider glands with a characteristic cysteine motive and it acts as a blocker of glutamate-gated ion channels, voltage-gated sodium channels, or voltage-dependent calcium channels (216). The agatoxin-homolog from *A. vulgaris* shares a specific architecture with a signal peptide, pre-peptide sequence and characteristic cysteine motive in the mature peptide (Figure 30). Proteins with characteristic cysteine pattern are able to build disulphide bonds, and we can distinguish three types of cysteine knots - Growth Factor

Cysteine Knot (GFCK), Inhibitor Cysteine Knot (ICK) and the Cyclic Cysteine Knot (CCK) (Figure 31). Most of ion channel blockers, haemolytic agents, or molecules having antiviral and antibacterial activities, are peptides between 26-48 amino acid residues which contain the cysteine knot (217). The stability of different types of cysteine knots, variety of bioactivities and their exceptional structural scaffold can be used for molecular engineering applications and in drug design (218).

The innate immune system is encoded to distinguish "non-self-materials" mediated by a group of pattern recognition proteins (PRPs) or pattern recognition receptors (PRRs), which can recognize and bind to different pathogen-associated molecular patterns (PAMPs) of invasive microorganisms (219, 220). C-type lectins, PGN-recognition proteins (PGRPs) and galectins belong to PRPs group, indicating that they play a key role in primary defence against invasive microorganisms in *A. vulgaris*. In the last two decades key components of innate immune systems of invertebrates have been discovered established and they include coagulation factors, complement factors, protease inhibitors, antimicrobial peptides, Toll receptors, and other humoral factors. Together they protect the host from being invaded by bacterial, fungal, and viral pathogens (208). A lot of these factors were detected in our study (BioProject accession number PRJEB7891; dataset identifier PXD002078) and together with lectin-like proteins they may protect the host from invading bacterial, fungal, and viral pathogens.

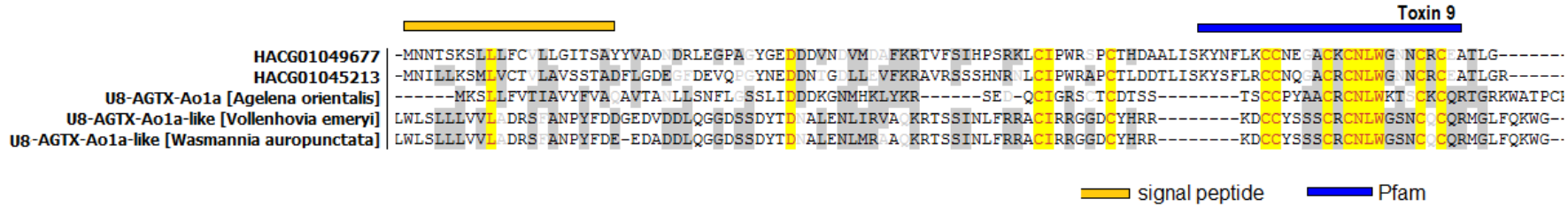
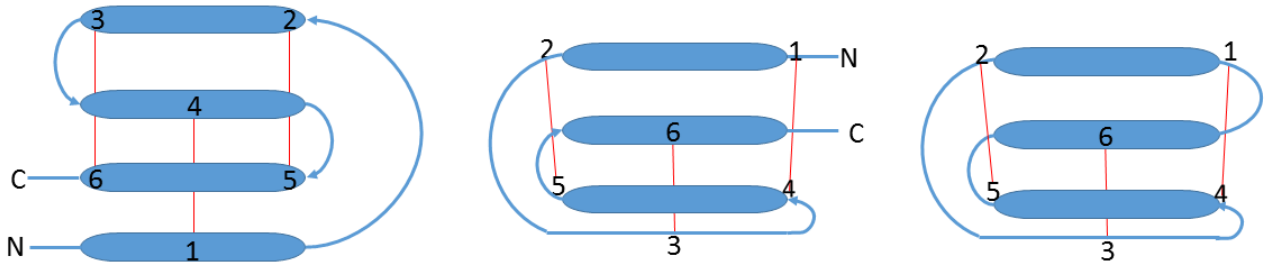


Figure 30. Alignment of amino acid sequences of toxin-like proteins from *A. vulgaris* with U8-agatoxine–Ao1a from *Agelena orientalis*



Growth Factor Cysteine Knot (GFCK)

Inhibitor Cystine Knot (ICK)

Cyclic Cystine Knot (CCK)

Figure 31. Schematic presentation of the three types of cystine's knots. The ellipsoid shapes represent beta-strands, while the cysteine residues are labelled from 1-6 in order from the N- to C-terminus. Red lines represent the disulfide bonds.

4.3. *Drosophila melanogaster* membrane associated and transmembrane proteins

For the first time a comprehensive analysis of an enriched membrane fraction from fly heads, using a gel-free approach, was reported by Aradska et al. (221). The MS proteomics data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD001712 and 10.6019/PXD001712. A total of 114,865 redundant and 38,179 distinct peptides were identified and assigned to 4812 proteins, including splicing variants. Of those, 3920 proteins (81.4%) were identified with more than two distinct peptides, 624 (13%) with two distinct peptides and 268 (5.6%) with two distinct peptides and multiple hits (Figure 32).

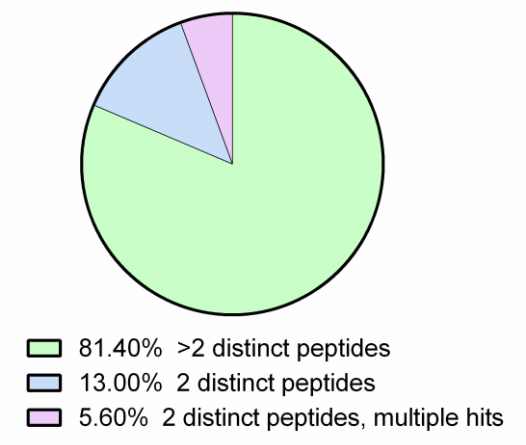


Figure 32. Experimentally identified proteins from membrane fraction of *Drosophila melanogaster*, with respect to the number of distinct peptides

Approximately eight distinct unique peptides were identified per protein. The list of the proteins, arranged according to an exponentially modified protein abundance index (emPAI) (222, 223), is given in Supporting Information Tables S1 (List of experimentally observed proteins) and S2 (List of experimentally observed proteins and peptides) in a paper published in *Proteomics* - Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads (221). All identified proteins correspond to 4559 genes, representing 33,5% of the predicted genes of *Drosophila melanogaster*.

The data from gel-free analysis was classified according to the length of amino acids (Figure 33A) and p*I* value (Figure 33B). Besides this, using the TMHMM algorithm (www.cbs.dtu.dk/services/TMHMM/), transmembrane regions from the identified proteins were predicted (224). 35% of proteins (1679 proteins) were identified as membrane proteins containing at least one transmembrane domain, while 65% (3133 proteins) had no confirmed or predicted transmembrane domain (Figure 34).

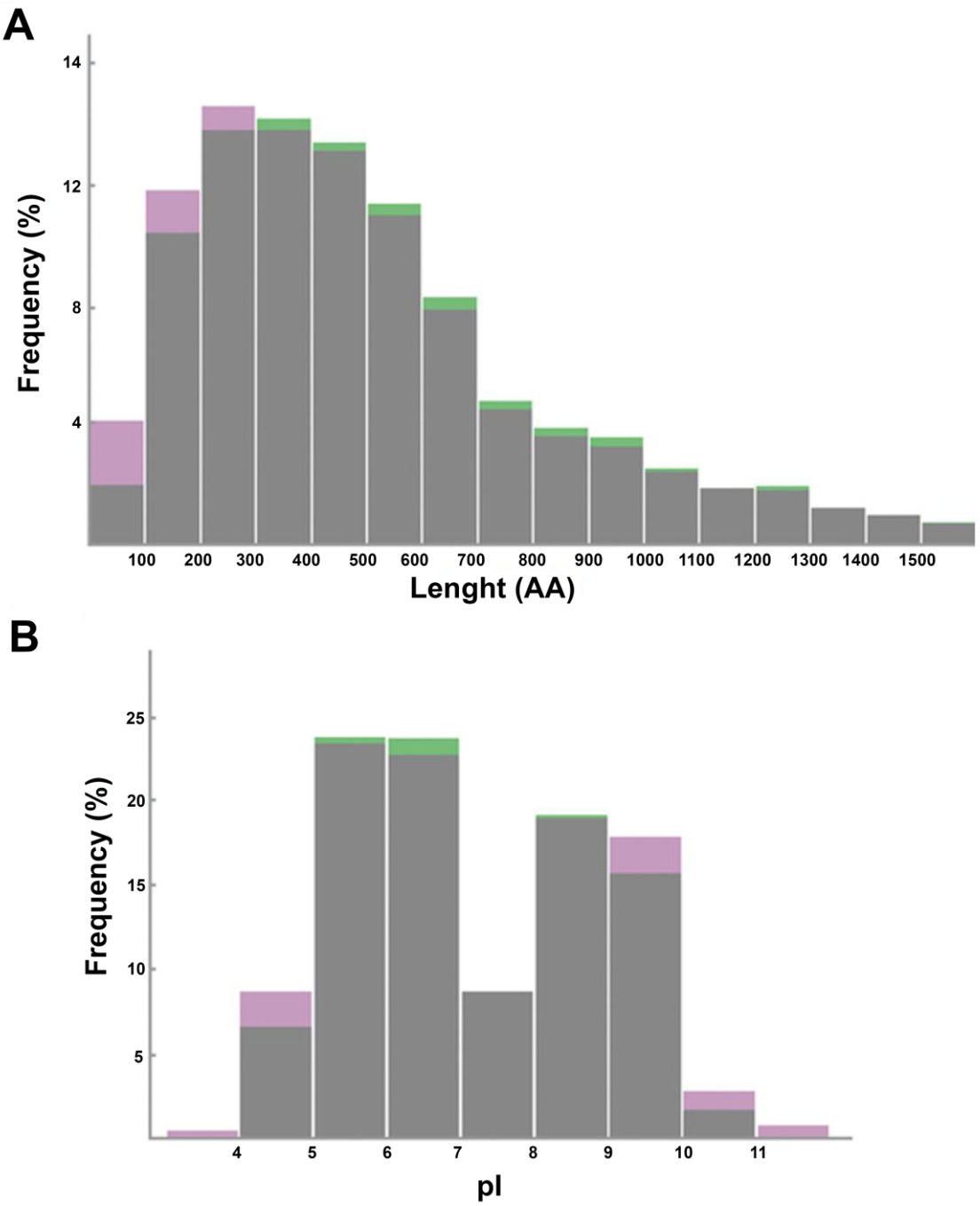


Figure 33. Distributions of length (A) and p*I* (B) of 4812 experimentally identified proteins in comparison to all proteins (20,328). Under-represented areas are shown in pink, over-represented areas in green

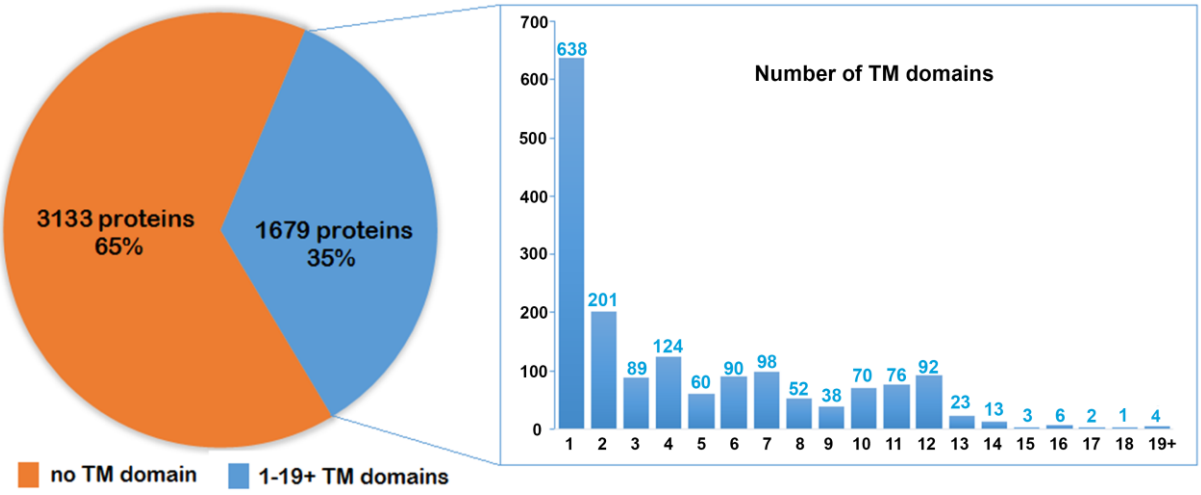


Figure 34. Transmembrane domain distribution among identified proteins

Additionally, the identified proteins were categorized with GO terms, using the AmiGO2 bioinformatics tool (225). AmiGO2 is an open multi-platform software (MATLAB-based) which is used in combination with Gene Ontology (GO), providing information on the ontologies, gene products and annotations, as well as BLAST search, Term Enrichment and GO Slimmer tools (225). For functional analysis, 3565 gene models were considered and an enrichment analysis was performed against *D. melanogaster* background (13 624 gene models). This analysis shows both over-and under-represented GO terms categories. From the data that we obtained, the most under-represented "biological processes" are presented in Figure 35; "molecular functions" and "cellular components" are in Figure 36 and 37, respectively.

Detection of chemical stimuli and sensory perception of chemical stimuli are two of the most under-represented "biological processes". Proteins involved in odorant binding, olfactory receptor activation and taste receptor activation are the most under-represented, according to AmiGo analysis of "molecular functions" (Figure 33). The most under-represented "cellular components" are: neuron projection membrane, dendrite membrane, leading edge membrane, proteins connected with chemosensory system (Figure 34). They represent the part of the genome which is detected on mRNA, but not on protein level. Chemosensory structures in insects are able to sense an extensive range of volatile, soluble chemicals – signals. Those signals are very important for finding and evaluating the quality of food source, as well as identifying

mates and oviposition sites (226). The olfactory systems of *Drosophila* and mammals are exceptionally similar. They involve the same receptors in olfactory receptor neurons (ORNs), which transmit signals from a single glomerulus (*Drosophila*) or the olfactory bulb (mammals) to the higher brain centres - mushroom body and the lateral horn (*Drosophila*), or to the olfactory cortex (mammals) (227). This similarities make *Drosophila* a great model for medical research, and detection of receptors at the protein level using gel-free methods combined with mass spectrometry opens new perspectives in molecular biology.

Figure 35. Under-represented “biological process” according to AmiGO2 bioinformatics tool

In biomedical research, receptors, transporters, and ion channels are extremely important membrane proteins, since they represent the biggest part of drug targets. The list of identified receptors from our study is presented in Tables 6 and 7. Generally, receptors are divided in two major groups: ionotropic and metabotropic, and receptors belonging to both groups are ligand-gated transmembrane proteins. Neurotransmitters which are released into a synaptic cleft can bind to a receptor and start a signalling cascade. Ionotropic receptors, after binding of a ligand (neurotransmitter), will allow an exchange of ions (K^+, Na^+, Cl^-, and Ca^{2+}) between a cell and its extracellular compartment, which triggers a fast response of numerous secondary messengers (228). In contrast, a metabotropic receptor activation initiates a signalling cascade via G-proteins (229).

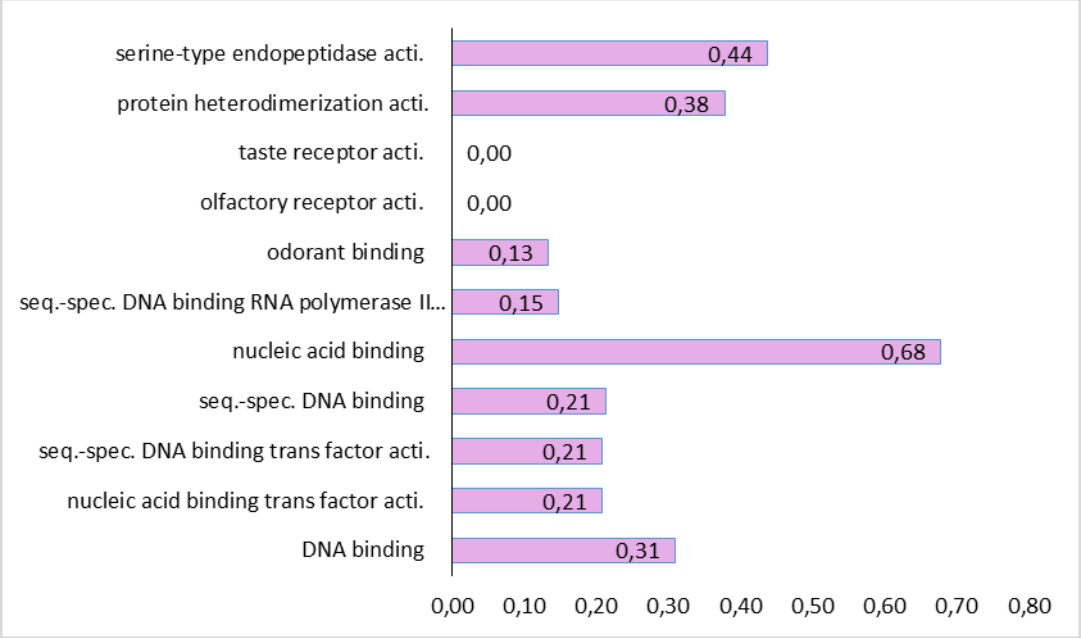


Figure 36. Under-represented "molecular function" according to AmiGO2 bioinformatics tool

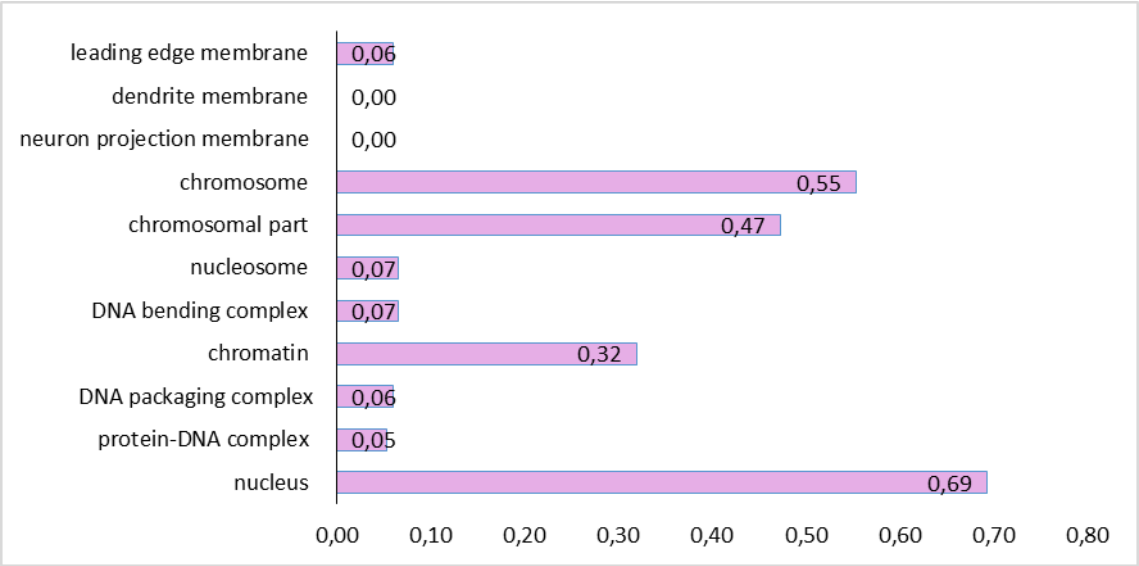


Figure 37. Under-represented "cellular compartment" according to AmiGO2 bioinformatics tool

Ionotropic glutamate receptors are divided into 4 groups: AMPA, 1 kainate, NMDA, and δ receptors, and they play a central role in learning, memory and numerous neurological diseases (230). The most abundant, in regard to emPAI index, are Nmdar1 (0.55) and Nmdar2 (0.41). Nicotinic acetylcholine receptors (nAChRs) are proteins that respond to the neurotransmitter acetylcholine in CNS and peripheral nervous system (PNS) in mammals, whereas in insects it is limited only to the equivalent of CNS. nAChRs are involved in synaptic transmission in the sympathetic/parasympathetic nervous system and as receptors of signals for muscular contraction (231). According to our analysis, α4 (0.29) and β3 are the most abundant nicotinic acetylcholine receptor subunits in *Drosophila* brain (0.28).

Table 6. List of ionotropic receptors detected from gel-free MS analysis of *Drosophila melanogaster* heads

Protein Description	Coverage (%)	emPAI	Unique Peptides
IONOTROPIC RECEPTORS			
iGluR			
Glutamate [NMDA] receptor subunit 1 (Nmdar1)	30,7	0,55	25
NMDA receptor 2, isoform G (Nmdar2)	22,6	0,41	19
Glutamate receptor IA (GluRIA)	11,8	0,15	7
Glutamate receptor IB, isoform B (GluRIB)	6,7	0,08	4
nAChR			
Acetylcholine receptor subunit alpha-like 1 (nAChRalpha1)	12,3	0,21	6
Acetylcholine receptor subunit alpha-like 2 (nAChRalpha2)	12,2	0,14	5
Acetylcholine receptor subunit beta-like 2 (nAChRbeta2)	13,3	0,23	5
Nicotinic acetylcholine receptor alpha4, isoform C (nAChRalpha4)	16,9	0,29	6
Nicotinic acetylcholine receptor alpha5, isoform B (nAChRalpha5)	13	0,21	7
Nicotinic acetylcholine receptor alpha6, isoform H (nAChRalpha6)	9,1	0,17	1
Nicotinic acetylcholine receptor Dalpha7 (nAcRalpha18C)	13,3	0,21	4
Nicotinic acetylcholine receptor beta 3 (Dbeta3) subunit (nAChRbeta3)	18,1	0,28	6
Nicotinic acetylcholine receptor (nicra3)	5,6	0,08	2

Table 7. List of metabotropic receptors detected from gel-free MS analysis of *Drosophila melanogaster* heads

Protein Description	Coverage (%)	emPAI	Unique Peptides
METABOTROPIC RECEPTORS			
mGluR			
Metabotropic glutamate receptor, isoform B (mGluR)	14,3	0,21	12
mAChR			
Muscarinic acetylcholine receptor DM1 (mAChR-A)	7	0,11	4
GABA			
Metabotropic GABA-B receptor subtype 2, isoform A (GABA-B-R2)	4,6	0,07	5
Dopamine			
Dopamine receptor 1 (Dop1R1)	13,3	0,24	5
Isoform 606 of Dopamine D2-like receptor (D2R)	8,7	0,14	4
Dopamine/Ecdysteroid receptor, isoform A (DopEcR)	6,2	0,1	2
Octopamine			
Octopamine receptor beta-3R (Octbeta3R)	1,4	0,03	1
Octopamine-Tyramine receptor, isoform B (Oct-TyrR)	15,1	0,21	7
Serotonin			
Serotonin receptor 1B, isoform C (5-HT1B)	14,7	0,14	4
Adenosine			
Adenosine receptor (69)	2,3	0,04	2

Metabotropic receptors also have neurotransmitters as ligands, but they have longer-lasting effects than ionotropic receptors. Receptors involved in memory are divided into several groups: metabotropic glutamate (mGluR), muscarinic acetylcholine (mAChR), gamma-aminobutyric acid (69), serotonin, as well as receptors for norepinephrine, epinephrine, histamine, dopamine, neuropeptides and endocannabinoids (232). From the data that we obtained, the most abundant receptors from the group of metabotropic ones are dopamine receptor 1 (0.24), mGluR (0.21) and octopamine-tyramine receptor (0.21). Dopamine (3,4-dihydroxyphenethylamine) acts as a neurotransmitter in several well-defined distinct pathways, such as motor control, executive functions, arousal, controlling the release of various hormones, and especially in reward-motivated behaviour. Regulation and metabolism of dopamine is also important in Parkinson's

disease, in case of drug and psychostimulants addiction, psychosis and antipsychotic drugs, attention deficit hyperactivity disorder (ADHD), pain in burning mouth syndrome, fibromyalgia, restless legs syndrome, schizophrenia etc. (233-237).

In invertebrates, the monoamines octopamine (OA) and tyramine (TA) take the roles of epinephrine and norepinephrine, which control behaviour, sleep, learning, memory, fight-or-flight response, as well as aggression. Moreover, the mechanism of regulation trough OA and TA is very similar to mammalian adrenergic system (238). Adenosine and serotonine receptors are also detected. Adenosine receptors modulate neuronal and synaptic function, whereas serotonin receptors are associated with aggression, anxiety, locomotion, memory and learning (239, 240). Analysis of those receptors at the protein level, using *D. melanogaster* as a model system, may improve the knowledge on the above listed processes.

Ion channels are membrane, pore-forming proteins. Their functions include establishing not only a resting membrane potential, but also modulating action potentials in neurons. Generally, they are controlling the flow of ions in cells, regulating various processes, such as contractions in cardiac, skeletal and smooth muscle system, pancreatic beta-cell insulin release, transport of nutrients and ions in epithelial cells (241, 242). It is known that more than 400 genes encode channel subunits. However, these transcripts can also be alternatively spliced. According to literature, mutation in pathology of ion channels is associated with numerous neurologic disorders (myotonic syndromes, periodic paralyses, partial epilepsies, generalized epilepsies, congenital myasthenic syndromes, hyperekplexia, familial erythermialgia, paraneoplastic channelopathies, etc.) Ion channels can be classified according to:

1) Gating
    - Voltage-gated ion channels
    - Ligand-gated ion channel
2) Type of ions
    - Chloride channels

- Potassium channels
- Sodium channels
- Calcium channels
- Protons channels
- Non-selective cation channels

3) Localization

- Plasma membrane channels
- Intracellular (endoplasmic reticulum and mitochondrial) channels

Channels which we have detected by gel-free analysis of *D. melanogaster* heads are listed in Table 8. The most abundant one is Calcium channel flower (1,33), followed by diverse types and isoforms of potassium channels (potassium voltage-gated channel protein shaker, inwardly rectifying potassium channel). Calcium channel flower regulates exo- and endocytosis of the synaptic vesicle (SV) in presynaptic terminals, development of photoreceptor terminals and control of gene transcription (243). Two essential functions of SV are transport of proteins involved in an uptake of neurotransmitter and trafficking of other proteins involved in SV endo/exocytosis and recycling. Nevertheless, voltage-gated Ca^{2+} channels are significant transducers of membrane potential changes (244). Potassium channels play an important role in excitable and non-excitable signalling processes, and are linked to diseases like long-QT syndromes, episodic ataxia/myokymia, familial convulsions, hearing/vestibular diseases, Bartter's syndrome, and familial persistent hyperinsulinemic hypoglycemia of infancy (245). Chloride channels are a divergent group of anion-selective channels and their function is related to the regulation of the excitability of neurones, cell volume regulation, the cell cycle and apoptosis (246). Histamine-gated chloride channels (HisCl α1 and α2) are well studied since they belong to the Cys-loop receptor superfamily of ligand-gated ion channels, and they are also closely related to the mammalian GABA(A) and glycine receptors (GlyR) (247). The first ion channel which was cloned was one of the sodium channels. Their main characteristic is transmission of

depolarization trough the cell networks. Sodium channels play a central role in physiology, because they transmit depolarizing impulses fast throughout cell/cell networks, thus allowing a co-ordination of higher processes like locomotion and cognition (248). At the basic amino acid level, more than 70% of synaptic proteins in *Drosophila* are analogous to mammalian synaptic proteins, which makes *Drosophila* a significant model system for neurobiological studies (249).

A group of proteins that are significant for neurobiological studies is the transporter group. They are integral membrane proteins with a significant role in the transport of ions, proteins and peptides. In contrast to channels, transporters are not open simultaneously to both the extracellular and intracellular environments; either its inner or its outer gate is open. If the movement of a substance across a membrane is against its concentration gradient, they are called active transporters. This process uses chemical energy, such as adenosine triphosphate (ATP), or an electrochemical gradient. On the other hand, transport is also possible without additional energy input. These type of transporters are called facilitated diffusion transporters and they are specific for large polar molecules and charged ions (250). From our analysis, the most abundant ones were amino acid transporter-EAAT1 (1.24) and facilitated trehalose transporter-Tret1-1 (1.02) (Table 9). Glutamate, a dominant neurotransmitter in mammalian CNS, regulates processes in glutamatergic synapses. Besides neurons, glutamates are also present in glial cells (251). EAATs mutations are found in patients with seizures, migraine, cerebellar atrophy, and hemiplegia (252, 253). In addition to that, pathology of EAATs regulation is associated with amyotrophic lateral sclerosis, stroke, epilepsy, schizophrenia, Alzheimer's and Huntington's diseases (254). Trehalose [α-D-glucopyranosyl-(1,1)-α-D-glucopyranoside, Glc(α1–1α)Glc] has numerous functions that distinguish it from other disaccharides. These functions are: protection against stress (desiccation, heat, low temperature, and high/low oxygen)(255-257), as well as prevention of osteoporosis (258), oculopharyngeal muscular dystrophy (259), and an induction of the mammalian target of rapamycin-independent autophagy (260). Besides trehalose transporters, glucose transporters, a huge group of membrane proteins which enable the transport of glucose, were also detected. They are present in all phyla. A large number of glucose transporter isophorms, with specific tissue expression, substrate specificity, transport kinetics and altered function in different physiological conditions,

were detected (261). Neural function is also controlled by choline, which is a precursor of the neurotransmitter acetylcholine (ACh) that regulates a huge number of autonomic, cognitive, and motor functions. Choline transporter (CHT) is related with cholinergic dysfunctions such as Alzheimer's disease, Parkinson's disease, schizophrenia, Huntington's disease and dysautonomia (262). Due to the fact that many functional, morphological and molecular characteristics are conserved in *Drosophila*, it represents an optimal model system for studying the above mentioned conditions (263).

Table 8. List of channels detected from gel-free MS analysis *of Drosophila melanogaster* heads

Protein Description	Coverage (%)	emPAI	Unique Peptides
Ca[2+]-channel protein alpha[[1]] subunit D, isoform H (Ca-alpha1D)	5,9	0,06	8
Ca[2+]-channel protein alpha[[1]] subunit T, isoform C (Ca-alpha1T)	4,4	0,05	7
Ca2+-channel-protein-beta-subunit, isoform G (Ca-beta)	7,2	0,1	2
Ca2+-channel-protein-beta-subunit, isoform P (Ca-beta)	6,7	0,11	1
Calcium channel flower	39,2	1,33	5
Chloride channel protein (CG5284-RB)	22,6	0,38	15
Chloride channel protein (ClC-a-RF)	16,5	0,25	13
Chloride channel protein (ClC-b)	22	0,39	14
Cyclic nucleotide-gated ion channel-like, isoform E (Cngl)	4,6	0,03	4
Eag-like K[+] channel (Elk)	6,6	0,08	5
Histamine-gated chloride channel subunit (ort)	10,1	0,1	5
Histamine-gated chloride channel subunit 1, isoform D (HisCl1)	5,5	0,07	1
I[[h]] channel, isoform E (Ih)	13,2	0,22	7
Inwardly rectifying potassium channel 3, isoform C (Irk3)	12,7	0,16	1
Inwardly rectifying potassium channel 3, isoform D (Irk3)	20,4	0,32	6
Inwardly rectifying potassium channel, isoform C (Irk1)	20,2	0,43	9
Isoform L of Potassium voltage-gated channel protein Shaker (Sh)	28	0,44	1
Isoform L of Potassium voltage-gated channel protein Shaker (Sh)	24,4	0,42	1
Open rectifier K[+] channel 1, isoform D (Ork1)	13	0,21	10
Potassium voltage-gated channel protein Shaker (Sh)	30,4	0,43	1
Small conductance calcium-activated potassium channel, isoform U (SK)	4,9	0,07	4
Sodium channel protein (para)	5,8	0,06	8
Sodium channel protein 60E (NaCP60E)	7,4	0,07	12
Transient receptor potential cation channel protein painless (pain)	21,6	0,45	20
Voltage-dependent calcium channel alpha13 subunit (Ca-alpha1T)	10,4	0,08	1

Table 9. List of transporters detected from gel-free MS analysis of *Drosophila melanogaster* heads

Protein Description	Coverage (%)	emPAI	Unique Peptides
ABC transporter expressed in trachea, isoform C (Atet)	15,4	0,17	7
Adenosine 3~-phospho 5~-phosphosulfate transporter 2 (Papst2)	13,9	0,17	4
Amino acid transmembrane transporter (CG4991)	6,5	0,07	2
Amino acid transporter (Eaat1)	31,9	1,24	16
Amino acid transporter protein (JHI-21)	8,8	0,21	6
CMP-sialic acid/UDP-galactose transporter (Csat)	12,9	0,2	4
ER GDP-fucose transporter, isoform B (Efr)	15,9	0,3	5
Facilitated trehalose transporter (Tret1-1)	28,4	1,02	10
Facilitated trehalose transporter (Tret1-2)	11,7	0,3	2
GDP-fucose transporter 1 (nac)	16	0,26	5
Glucose transporter type 1 (Glut1)	11,8	0,18	4
Glucose transporter type 1 (Glut1)	21,9	0,44	7
High-affinity choline transporter 1 (CG7708)	21,5	0,52	12
Isoform A of Putative inorganic phosphate cotransporter (Picot)	24,2	0,5	12
Major facilitator superfamily transporter 3 (MFS3)	11,5	0,14	4
Na[+]-dependent inorganic phosphate cotransporter (NaPi-T)	11,6	0,27	7
Organic cation transporter protein (Orct)	25	0,58	11
Organic cation transporter-like protein (Orct2)	11,3	0,21	4
Peptide transporter family 1 (yin)	9,4	0,14	4
Putative sodium-dependent multivitamin transporter (CG32669)	11,6	0,2	6
Sodium chloride cotransporter 69, isoform C (Ncc69)	33,6	0,66	2
Transporter (CG10804)	14,7	0,24	8
Transporter (CG13793-RA)	8,3	0,16	4
Transporter (CG13794)	11,8	0,23	5
Transporter (CG13795)	12,3	0,23	5
Transporter (CG15279)	11,6	0,16	6
Transporter (CG1732)	24,7	0,7	19
Transporter (CG3252-RA)	12,2	0,19	7
Transporter (CG5226-RA)	19,6	0,45	13
Transporter (CG5549)(glycin)	21,6	0,29	15
Transporter (DAT)	5,5	0,08	3
Transporter (ine)	15,6	0,24	11
Transporter (List)	11,9	0,15	5
Vesicular glutamate transporter isoform A (Vglut)	21,4	0,39	10

5. CONCLUSIONS

During the work in this thesis, we defined the following aims:

c) Whole body transcriptome of *Arion vulgaris*, in order to create mRNA-derived protein database

d) Proteomic profiling of *Arion vulgaris* mantel using nano-LC-ESI-MS/MS analysis.

b) Identification of membrane associated and transmembrane proteins from *Drosophila melanogaster* heads using gel-free MS/MS approach

5.1. Whole body transcriptome of *Arion vulgaris*

Using Illumina HiSeq 2000 technology we created a cDNA library which contains 339 millions of paired end reads. Using Trinity pipeline we assembled 136,406 contigs of *de novo* transcript, finally resulting in 47,451 protein-coding sequences.

In comparison with non-redundant data base of NCBI we obtained the following results: 63.3% of transcriptome, which is 31,463 transcripts, was assigned to at least one protein, while 33.7% (15,988 transcripts) was with no significant hits.

0.4% (127 transcripts) represents unknown or unnamed proteins.

65% of sequences is shared with *Aplysia californica*, 8.7% with *Crassostea gigas* and 2.5% with *Capitella teleta*. More than 80% of the sequences best matched with mollusc species.

According to Pfam database, in 16,994 entries at least one domain was recognized, and of those entries, 471 sequences were without blastX hit annotation. Three most abundant domains are zinc-finger double domain (zf-H2C2_2), domain of unknown function (DUF4200) and ankyrin repeat (Ank_5).

1,796 of transcripts were assigned with EC (Enzyme Commission) numbers which specify enzyme-catalysed reactions.

5.2. *Arion vulgaris* mantle proteome

We identified 833 unique proteins from 2-DE gel. 814 proteins had a significant hit in NCBI nr database, while 19 proteins are with no significant homology in nr NCBI database.

With combination of 1-DE gel and LC/MS-MS with higher resolution, we identified 2011 proteins, from which 48 proteins are without a hit on nr NCBI database.

Total number of identified proteins from 1-D and 2-D gels, with minimum of one unique peptide and distinct peptides, is 2129.

We identified whide-renge of proteins which indicates an unexpected complexity of the immune system, with a high grade of pathogen specificity and immune-priming: cytokine-like proteins, complement-like proteins, peptidoglycan-recognition proteins (PGRPs), lipopolysaccharide and β1, 3-glucan-binding proteins, fibrinogen-related proteins (FREPs), pore-forming membrane attack/perforin (MACPF) domain proteins and toxin-like proteins.

Lectin-associated functions may be attributed to 149 transcripts, 37 of which were unambiguously identified at the protein level. Lectins that we detected are: C-type lectin family, galectin, malectin, H-type lectins, L-type lectins, Ricin-B lectins and calnexin. They contain conserved carbohydrate-recognition domains (CRDs).

We discovered novel toxin-like proteins with significant similarity to agatoxin in a signal peptide, pre-peptide sequence and characteristic cysteine motive in the mature peptide.

5.3. *Drosophila melanogaster* membrane associated and transmembrane proteins

Using a comprehensive analysis with a gel-free approach of an enriched membrane fraction from fly heads, we identified 114,865 redundant and 38,179 distinct peptides and they were assigned to 4812 proteins, including splicing variants.

3920 proteins (81.4%) were identified with more than two distinct peptides, 624 (13%) with two distinct peptides and 268 (5.6%) with two distinct peptides and multiple hits.

35% of proteins (1679 proteins) were identified as membrane proteins containing at least one transmembrane domain.

According to AmiGo analysis of our data we showed that detection of chemical stimuli and sensory perception of chemical stimuli are two of the most under-represented “biological processes”. Proteins involved in odorant binding, olfactory receptor activation and taste receptor activation are the most under-represented. The most under-represented “cellular components” are: neuron projection membrane, dendrite membrane, leading edge membrane, proteins connected with chemosensory system.

6. EXPERIMENTAL PART

6.1. Chemicals, solutions and buffers

All chemicals and solutions used during experiments were analytical, technical, or LC/MS- grade quality. Details about solutions, chemicals and commercial kits are given in Table 10.

Table 10. List of used chemicals, solutions and commercial kits

Name and abbreviations	Catalog number	Grade quality	Company
4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl) piperazine-N′-(2-ethanesulfonic acid) (HEPES)	H4034	Analytical	Sigma Aldrich
Sucrose	S0389		
Phenylmethanesulfonyl fluoride (PMSF)	P7626		
Ammonium bicarbonate	A6141		
Ammonium-sulfate	A4418		
Tris	252859		
Thiourea	T8656		
Iodoacetamide	I1149		
Ethylenediaminetetraacetic acid (EDTA)	E9884		
Ammonium persulfate (APS)	A3678		
Glycerol	G5516		
Tributylphosphate (TBP)	240494		
Isopropanol	I9516		
N,N,N′,N′-Tetramethylethylenediamine (TEMED)	T9281		
Hydrochloric acid (HCl)	84415	Technical	
Glycine	G8898		
Triethylammonium bicarbonate buffer (TEAB)	17902	LC/MS	
Ammonium formate	17843		

Name and abbreviations	Catalog number	Grade quality	Company
1,4-Dithiothreitol DTT	6908	Analytical	Carl Roth
3-[(3-Cholamidopropyl) dimethylammonio]-1-propansulfonat (CHAPS)	1479		
Urea	2317.1	Technical	
Acrylamide	7906		
Bis-Acrylamide	6755.2		
Acetic acid	6755.2		
Trypsin	V5111	LC/MS	Promega
Chymotrypsin	V1062		
Methanol	M/4062/17	LC/MS	Fisher Scientific
Acetonitrile	A/0638/17		
Formic acid	1.11670.1000	LC/MS	Merck
Triflouracetic acid (TFA)	1.08262.0025		
Isopropanol	1010402500		
Orthophosphoric acid	100573	Technical	
SDS	1610302	Analytical	Bio Rad
Piperazine diacrylamide	161-0202		
Coomassine G-250	1610406	Molecular biology	
Bromophenol blue	1610404		
2-DEclean up kit	1632130		
Protein marker	1610373		
Immobiline DryStrip Gels	17-6003-76	Molecular biology	GE Helthcare
Carrier (3–11pH)	17045601		
Dry cover strip fluid	17-1335-01		
cOmplete protease inhibitor cocktail	11836145001	Molecular biology	Roche
Pierce 660 protein assay	22660	Molecular biology	Thermo Scientific
Colloidal coomassie blue	LC6025	Molecular biology	Novex Life technologies
RNeasy kit	74104	Molecular biology	Qiagen
NEBNext kit	#E7420S/L	Molecular biology	NEB

Following buffers were used during experimental procedures:

1. Urea Buffer

7 M	Urea
2 M	Thiourea
4 %	CHAPS
2 M	Tris pH 8,8
0,5 %	Carrier (3–11pH)
1 %	DTT
1 mM	PMSF
1X	cOmplete protease inhibitor cocktail

2. Homogenization buffer

10 mM	HEPES pH 7,5
300 mM	Sucrose
1X	complete protease inhibitor cocktail

3. Washing buffer

10 mM	HEPES pH 7,5
1X	cOmplete protease inhibitor cocktail

4. SDS extraction buffer

1,5 %	SDS
100 mM	NaCl
20 mM	Tris-HCl pH 6,5
1X	cOmplete protease inhibitor cocktail

5. Urea A buffer (UA buffer)

8 M	Urea
100 mM	Tris-HCl (pH 8,5)

6. Rehydration buffer

8 M	Urea
10 mM	DTT
4 %	CHAPS
0,5 %	Carrier (3–11pH)
0,008 %	Bromphenol Blue

7. Equilibration buffer 1

50 mM	Tris-HCl, pH 8,8
6 M	Urea
30 %	Glycerol
2 %	SDS
1 %	DTT

8. Equilibration buffer 2

50 mM	Tris-HCl, pH 8,8
6 M	Urea
30 %	Glycerol
2 %	SDS
4 %	Iodoacetamide
0,002 %	Bromophenol blue

9. Leammli buffer

150 mM	Tris-HCl pH 6,8
300 mM	DTT
6 %	SDS
0,3 %	Bromophenol blue
30 %	Glycerol

10. SDS running buffer

0,025 M	Tris
0,192 M	Glycine
0,1 %	SDS

11. Fixation buffer

50 %	Methanol
10 %	Acetic acid

12. Blue silver staining solution

12,1 %	Orthophosphoric acid
20 %	Methanol
0,1 g/mL	Ammonium-sulfate
0,0012 g/mL	Coomassine G-250

13. Stage Tip Buffer

0,4 %	Formic acid
2 %	TFA

14. STP Elution Buffer

0,4 % Formic acid
90 % Acetonitrile

15. Solution A for HPLC coupled with amaZon

0,1 % Formic acid

16. Solution B for HPLC coupled with amaZon

0,08 % Formic acid
80 % Acetonitrile

17. Solution A for HPLC coupled with LTQ-OrbitrapVelos

0,4 % Formic acid

18. Solution B for HPLC coupled with LTQ-OrbitrapVelos

0,4 % Formic acid
70 % Methanol
20 % Isopropanol

19. Solution A for HPLC coupled with Thermo Orbitrap Velos Pro

0,01 % Triflouracetic acid

20. Solution B for HPLC coupled with Thermo Orbitrap Velos Pro

0,05 % Triflouracetic acid
80 % Acetonitrile

21. Solution A for HPLC for off-line fractionation

5 % Acetonitrile
20 mM Ammonium formate

22. Solution B for HPLC for off-line fractionation

90 % Acetonitrile
20 mM Ammonium formate

6.2. Apparatus and procedures

6.2.1. Protein isolation and purification

6.2.1.1. *Arion vulgaris*

Arion vulgaris specimens were collected in Woerdern, Austria (48°190′40″ N, 16°120 ′34″ E) in May 2014. In total, 15 specimens were collected, ranging from 7 to 8 g. Species identification was carried out with the assistance of the Zoological Department of Museum of Natural History in Vienna. According to national law no permission is required to collect and use invertebrates. Individuals of *A. vulgaris* were frozen immediately after harvest in liquid nitrogen and stored at -80°. Mantles were dissected directly before protein extraction procedure. Samples were powdered with mortar and pestle in liquid nitrogen and homogenized in urea buffer. Samples were sonicated (Ultra-TurraxT25, Staufen, Germany) using 5 sec pulse, 5 times in row, following incubation on RT for 1 h, and at 4°C for 2h, with vortexing every 15 min. Samples were centrifuged with Eppendorf benchtop centrifuge (18,000 *x g*) at 10°C for 1h. Supernatant was desalted at 1000 *x g* with fresh urea buffer in Amicon Ultra-15 Centrifugal filter unit Ultracel-3 membrane (Merck Millipore). Samples were centrifuged until 500 µl remain.

Protein concertations were estimated with Pierce 660 nm Protein Assay kit. Samples were cleaned using a 2-DE clean up kit (ReadyPrep™ 2-D Cleanup Kit, BioRad), according to the manufacturer's instructions. Briefly, 400 µg of sample was

mixed with precipitating agent No#1, and incubated on ice for 15 min before adding precipitating agent No#2. Pellet was obtained by centrifugation of previously mentioned mixture at 12,000 x *g* for 5 min, following washing with LC/MS water and Washing reagent No#2. Pellet was incubated at -20°C for 30 min, centrifuged at 12,000 x g for 5 min, air dried no more than 5 min and re-suspended by adding an appropriate volume of 2-D rehydration/sample buffer.

.

6.2.1.2. *Drosophila melanogaster*

Drosophila melanogaster heads were isolated from adult flies using a standard freezing protocol (264). Briefly, flies were collected in vials and flash-freezed by immersion of the vials in liquid nitrogen. Heads were isolated with sieve (upper chamber containing largest mesh for collecting bodies while lower chamber was for collecting heads). Sieve, containing flies, was dipped in the liquid nitrogen and shaked vigorously for 1 min. Heads were then ground into a fine powder with a pre-chilled mortar and pestle. Powder was resuspended in ice-cold homogenization buffer. The homogenate was centrifuged for 10 min at 1000 × *g* and the supernatant was centrifuged at 50 000 × *g* for 30 min at 4°C. Subsequently, the pellet was resuspended in 4 mL washing buffer, kept on ice for 30 min and centrifuged at 50 000 × g for 30 min. A total plasma membrane fraction was separated by equilibrium density centrifugation on a discontinuous sucrose gradient. The pellet, resuspended in washing buffer, was layered on top of the sucrose cushion (1 M and 1.25 M sucrose solution) followed by centrifugation at 70 000 × *g* for 2 h at 4 °C. After centrifugation, fractions from the sucrose interface were collected and diluted ten times with washing buffer and subsequently centrifuged for 30 min at 100 000 × *g*. The pellet was stored at 80 °C until use. Membrane protein extraction was carried out with urea buffer according to a previous study with minor modifications (265). Protein quantities were estimated by the Pierce 660 protein assay.

6.2.2. Gel electrophoresis and staining

6.2.2.1. One dimensional electrophoresis (1D)

One dimensional electrophoresis was performed on 10% polyacrylamide gel (10% Acrylamide/bis-acrylamide (29:1), 0.625 M Tris-HCl pH 8.8, 0.1% SDS, 0.1% APS, 0.1% TEMED). 4% resolving gel was used in this experiment (4% Acrylamide/bis-acrylamide (29:1), 0.125 M Tris-HCl pH 6.8, 0.1% SDS, 0.1% APS, 0.1% TEMED). 50 μg of extracted proteins were mixed with Laemmli buffer at a 1:3 volume ratio. During electrophoresis voltage was set up as constant: 30 min 50 V, 30 min 100 V and 1h 150 V.

6.2.2.2. Two dimensional electrophoresis (2D-IEF/SDS-PAGE)

2-DE was performed as reported previously with minor modifications (266-268). 750 μg of proteins were loaded on 18 cm pH 3–11 linear IPG strips (GE Healthcare). IEF was performed in a IPGphore electrophoretic system (GE Healthcare, Uppsala, Sweden) using a protocol with gradually increasing voltage from 200 to 8,000 Vat 4V/min. Prior to second dimension, IPG strips were incubated in 10 ml of equilibration buffer 1 for 15 min with gentle shaking followed by incubation in 10 ml of equilibration buffer 2. The second-dimensional separation was performed on 10–16% gradient SDS-PAGE gels. During electrophoresis voltage was set up as constant: 14 h 50 V, 2 h 100 V and 2 h 150 V. Molecular masses were determined by comparison with precision Bio Rad protein standard markers spanning from 10 to 250 kDa molecular weight range. After fixation for 4 h in fixation buffer, gels were stained over night with the colloidal coomassie blue stain with mild shaking in cold room at +4 °C.

6.2.3. In gel digestion

All spots were excised from 2-DE gels, and digested with trypsin or chymotrypsin. The gel pieces were washed with 50 % Acetonitrile (69) in 10 mM ammonium bicarbonate for 20 min with vortexing and then 20 min with 10 mM

ammonium bicarbonate. These two steps were repeated until the gel was completely distained. 100 % ACN (69) was added for 10 min and gel pieces were dried completely using Eppendorf concentrator for 30 min at 30 °C. Cysteine residues were reduced with 10 mM DTT at 56 °C for 30 min and alkylated with 55 mM Iodoacetamide for 45 min at 22 °C. After washing with 10 mM ammonium bicarbonate and dehydration with 100% ACN, proteins were digested with 12.5 ng/µl trypsin solution buffered in 10 mM ammonium bicarbonate for 16 h (overnight) at 37 °C. The supernatant was transferred to new Eppendorf LoBind 0.5 mL tubes and peptides were extracted with 1 % formic acid and then with 15 % ACN/0.1 % formic acid (269). 20 µL of extracted peptides were analysed by nano-LC-ESI-(CID/ETD)-MS/MS.

6.2.4. Filter aided sample preparation (FASP)

Membrane protein extraction was carried out according to a previous study with minor modifications (265). 300 µg of the sample (1 × 100 µg each) were digested with trypsin using the filter aided sample preparation (FASP) as previously described with minor modifications (83, 270). Samples were mixed in the filter unit with Urea A buffer (UA) buffer and centrifuged at 14 000 × g for 15 min at 22 °C. Any remaining detergent was exchanged by urea with 200 µl UA. The proteins were alkylated by the addition of 100 µl of 50 mM iodoacetamide in UA and incubated for 30 min at 22 °C. Subsequently, excess reagent was removed by 3 × 100 µL UA solution followed by 3×100 µl of 50 mM TEAB buffer. Proteins were digested overnight at 37 °C with trypsin (an enzyme to protein ratio of 1:100 w/w). Tryptic peptides were recovered from the filter with 40 µL of 50 mM TEAB buffer followed by 50 µl of 0.5 M NaCl. The remaining semi-digested material on the FASP filter was washed with 50 µl 50 mM TEAB buffer and further digested with chymotrypsin (an enzyme to protein ratio of 1:10 w/w). Pooled tryptically-digested and sequential trypsin-chymotrypsin digested samples were acidified, desalted and concentrated with customized reversed phase C18 stage tips (271). Bound peptides were eluted with a buffer containing 90 % acetonitrile and 100 mM TEAB and lyophilized in a vacuum concentrator (Eppendorf, Hamburg, Germany).

6.2.5. Solid phase extraction

The peptides obtained after tryptic digestion were cleaned with ultra-microspin column. Briefly, columns were activated with 2 x 200 µl methanol and centrifuged in 100 x *g* for 1 min. After that, they were washed 2 x 400 µl with stage tip buffer. Acidified samples were loaded and washed with 400 µl of stage tip buffer. Peptides were eluted with 200 µl stage tip elution buffer, lyophilized on Eppendorf concentrator and kept at -20 °C, or proceeded immediately for experiments.

6.2.6. Off-line fractionation

Peptides generated in FASP were prefractioned according to protocol that was published by Gilar and coworkers (272). Peptide digests were pooled, purified by solid phase extraction (81) (UltraMicroSpin columns 3–30 µg capacity, Nest Group Inc., Southboro, MA, USA) and separated at pH 10 by reversed-phase liquid chromatography using an Agilent 1200 series HPLC system. Separation was performed at a flow rate of 100 µl/min on a Phenomenex column (150 × 2.0 mm Gemini-NX 3 µm C18 110 Å, Phenomenex) with a 50 min linear gradient from 5–70% (vol/vol) acetonitrile containing 20 mM ammonium formate. Seventy two time-based fractions were collected, acidified, and pooled into 20 or 50 fractions. Peptides were lyophilized in a vacuum centrifuge and reconstituted in 5% (vol/vol) formic acid for on-line LCMS. Peptide abundance per fraction was estimated based on the UV trace and samples were diluted accordingly to avoid overloading the nano-LCMS system.

6.2.7. LC-MS/MS analysis

6.2.7.1. Nano-LC-MS/MS (amaZon speed ETD)

Nano-LC-MS/MS system was used during execution of these experiments. The HPLC used for peptide separation was an Ultimate 3000 system (Dionex Corporation, Sunnyvale, CA) equipped with a PepMap100 C-18 trap column (300 µm ×

5 mm) and PepMap100 C18 analytical column (75 μm × 250 mm). The Amazon speed ETD (Bruker Daltonics, Bremen, Germany) was used to record peptide spectra over the mass range of m/z 400–1,400 and MS/MS spectra in information-dependent data acquisition over the mass range of m/z 100–2,800. Repeatedly, MS spectra were recorded followed by three data-dependent CID MS/MS spectra. An active exclusion of 0.4 min after two spectra was used to detect low abundant peptides. The voltage between ion spray tip and spray shield was set to 1,400 V. Drying nitrogen gas was heated to 150 °C and the flowrate was 3 l/min. The collision energy was set automatically according to the mass and charge state of the peptides chosen for fragmentation. Multiple charged peptides were chosen for MS/MS experiments due to their good fragmentation characteristics. MS/MS spectra were interpreted and peak lists were generated by Data Analysis 4.1 (Bruker Daltonics, Bremen, Germany).

6.2.7.2. Nano-LC-MS/MS (LTQ-Orbitrap Velos)

Samples were analysed on a linear trap quadrupole (LTQ) Orbitrap Velos (Thermo Fisher Scientific, Waltham, MA, USA) coupled to an Agilent 1200 HPLC nano flow system (Agilent Biotechnologies, Palo Alto, CA, USA). Data were acquired using Xcalibur (v2.1.0). The peptide mixture was automatically loaded onto a trap column (Zorbax 300SB-C18 5 μm, 5 × 0.3 mm, Agilent Biotechnologies), and eluted by back-flushing onto a 16 cm fused silica analytical column packed with C18 reversed phase material (ReproSil-Pur 120 C18-AQ, 3 ìm, Dr. Maisch GmbH, Ammerbuch-Entringen, Germany). MS^1 scans were performed from *m/z* 350-1800 at a resolution of 60,000. Using a data-dependent acquisition mode, the 10 most intense precursor ions with a minimum signal threshold of 5,000 from the MS^1 spectrum were isolated (within a 1.2 Da window) and fragmented to obtain the corresponding MS^2 spectrum. The fragment ions were generated in a higher-energy collisional dissociation (HCD) cell at an NCE of 37% and were detected in an Orbitrap mass analyser at a resolution of 7,500. The dynamic exclusion for the selected ions was 60 s and single-charged precursor ions were excluded.

6.2.7.3. Nano Nano-LC-MS/MS (Orbitrap Velos PRO)

Peptide fractions were separated by nanoLC (Dionex Ultimate 3000) equipped with a µ-precolumn (C18, 5 µm, pore size 100 Å, 5 × 0.3 mm, Thermo Fisher Scientific, Vienna, Austria) and an Acclaim PepMap RSLC nanocolumn (C18, 2 µm, 150×0.075 mm, Thermo Fisher Scientific, Vienna, Austria). The sample was ionized in the nanospray source equipped with stainless steel emitters (Thermo Fisher Scientific, Vienna, Austria) and analysed in a Thermo Orbitrap Velos Pro mass spectrometer in positive ion mode by alternating full-scan MS (m/z 380–2000) in the ICR cell and MS/MS by CID of the 20 most intense peaks in the ion trap with dynamic exclusion enabled.

6.3. Transcriptome sequencing and assembly

6.3.1. Isolation of RNA

Isolation of RNA was performed using RNeasy kit (Qiagen, Hilden, Germany). After removal of the gut, two specimens of *Arion vulgaris* were immersed in liquid nitrogen and pulverized with the help of a mortar and pestle and homogenized in QIAzol Lysis Reagent. Adding gDNA Eliminator Solution and chloroform enabled separation of aqueous (RNA partitions) and organic phases/interphase (DNA and proteins) by centrifugation. The upper, aqueous phase was collected, mixed with ethanol, and RNA was purified using RNeasy spin columns. The aqueous phase was mixed with ethanol to provide appropriate binding conditions, and applied to an RNeasy Mini spin column. Total RNA binds to the spin column membrane, while phenols and other contaminants are efficiently washed away. High-quality RNA was then eluted in RNase-free water.

6.3.2. mRNA isolation, fragmentation and priming

The first step in the workflow involves purifying the poly(A)-containing mRNA molecules using two rounds of poly(T)-oligo–attached magnetic beads. Pooled

total RNA was subjected to RNA-Seq following the mRNA sequencing protocol provided by New England Biolabs (NEBNext kit, NEB, Frankfurt, Germany). Briefly, total RNA was diluted with nuclease-free water in a nuclease-free PCR tube and kept on ice, followed by adding NEBNext Oligo d (T) 25 beads and mixing. Solution was on the magnetic rack at room temperature for 2 min and supernatant was removed from tube. These steps were repeated 2 times. After that beads were resuspended in RNA Binding Buffer and 50 µl of total RNA. Tube was incubated for 5 min at 65 °C and then cooled down to 4 °C to denature the RNA and facilitate binding of the poly-A mRNA to the beads following 5 min incubation at room temperature. The tube was on the magnetic rack at room temperature for 2 minutes in order to separate the poly-An mRNA binding beads from the solution. The supernatant was removed, while the pellet was washed 2 times with wash buffer. After adding 50 µl of tris buffer, sample was mixed, incubated at 80 °C for 2 minutes, and then kept at 25 °C to elute the Poly-An mRNA from the beads. Again, 50 µl of RNA Binding Buffer was added to the sample to allow the mRNA to rebind to the beads. And like in previous step, complete mixture was incubated at room temperature for 5 min, placed on magnetic rack for 2 min, after which supernatant was removed. Samples were washed with wash buffer and then with tris buffer. Removing of beads was performed with 15 µl of the first strand synthesis reaction buffer and random primer mix (2X) at 94 °C for 15 minutes.

6.3.3. First strand and second cDNA synthesis

The RNA fragments were copied into first-strand complementary DNA (cDNA) using reverse transcriptase and random primers. 10 µl of mRNA was mixed with Murine RNase Inhibitor, Actinomycin D (0 .1 µg/µl), ProtoScript II Reverse Transcriptase (NEBNext kit, NEB) and nuclease free water according to the manufacturer's instructions. Samples were incubated in a preheated thermal cycler: 10 minutes at 25 °C, 15 minutes at 42 °C, 15 minutes at 70 °C and hold at 4 °C. In order to perform second cDNA synthesis, the following solutions are added to previous mix, according to the manufacturer's instructions: nuclease-free water, second strand synthesis reaction buffer (10X) and second strand synthesis enzyme mix and incubated at 25 °C for 1h, with heated lid set at ≤ 40 °C.

6.3.4. Purifying the double-stranded cDNA

In order to purify double-stranded cDNA, 1.8X Agencourt AMPure XP beads (NEBNext kit, NEB) were used. 144 µl of the beads was mixed with second strand synthesis reaction (~80 µl), followed by incubation for 5 min at room temperature. While DNA targets were attached to beads, supernatant was removed after 5 min incubation of mixture on magnetic rack. Pellet was washed 2 times with 80 % ethanol and air dried for 10 min on magnetic rack. Eluting of DNA target was performed with nuclease-free water. Samples were mixed with nuclease-free water, vortexed, quickly spun and placed on magnetic rack. Supernatant was used for following steps.

6.3.5. End repair/dA-tail of cDNA library and adaptor ligation

The purified double-stranded cDNA was mixed with dA-Tailed cDNA, NEBNext End Repair Reaction Buffer (10X) and NEBNext End Prep Enzyme Mix (NEBNext kit, NEB) according to the manufacturer's instructions. Samples were incubated in a thermal cycler as follows: 30 minutes at 20 °C, 30 minutes at 65 °C and hold at 4 °C. Blunt/TA ligase master mix, diluted NEBNext adaptor and nuclease-free water were mixed and incubated 15 minutes at 20 °C. Nuclease-free water was used to adjust volume to 100 µl, followed by mixing with the sample from the previous step, and incubation on room temperature for 5 min. Supernatant (containing unwanted fragments) was discarded and pellet was washed two times with 80 % ethanol and air dried. DNA target was eluted with nuclease-free water, mixed with 1 X AMPure XP beads, incubated for 5 min and supernatant was separated from beads on magnetic rack. Pellet was washed 2 times with 80 % ethanol and air dried. DNA target was eluted with nuclease-free water and it was proceeded to the next step.

6.3.6. USER excision, PCR library enrichment and purifying

Sample was mixed with NEBNext USER enzyme, NEBNext high-fidelity PCR master mix (2X), Universal PCR primer and index (273) primer according to the

manufacturer’s instructions. PCR cycling conditions were: USER digestion (37 °C, 15 min, and 1 cycle), initial denaturation (98 °C, 30 seconds, and 1 cycle), denaturation (98° C, 10 sec, and 15 cycles), annealing (65 °C, 30 sec, and 15 cycles), extension (72 °C, 30 sec, and 15 cycles), final extension (72 °C, 5 min, 1 cycle). Samples were mixed with AMPure XP beads and incubated 5 min at room temperature. After 5 min at magnetic rack, supernatant was discarded, pellet was washed two times with 80 % ethanol and air dried. DNA target (size distribution of ~200–250 bp) was eluted with nuclease-free water and stored at –20 °C.

6.3.7. Selection of Adaptor-ligated DNA and sequencing

Sample was mixed with AMPure XP beads (NEBNext kit, NEB) according to the manufacturer’s instructions and incubated 5 min at room temperature. Magnetic stand was used to separate the beads (containing unwanted large fragments) from the supernatant (containing DNA). Procedure was repeated, but in this step supernatant was discarded and pellet (containing DNA target) was washed two times with 80 % ethanol and air dried. DNA target was eluted with 10 mM Tris-HCl and transferred to a new PCR tube. Finally, the adaptor-ligated DNA was sequenced for 100 cycles on a HiSeq2000 sequencing system (Illumna, San Diego, USA).

Schematic presentation of preparation of biological material for RNA library for next-generation sequencing is given in Figure 38, according to the manufacturer's instructions (NEBNext kit, NEB, Frankfurt, Germany, https://www.neb.com/products/e7420-nebnext-ultra-directional-rna-library-prep-kit-for-illumina).

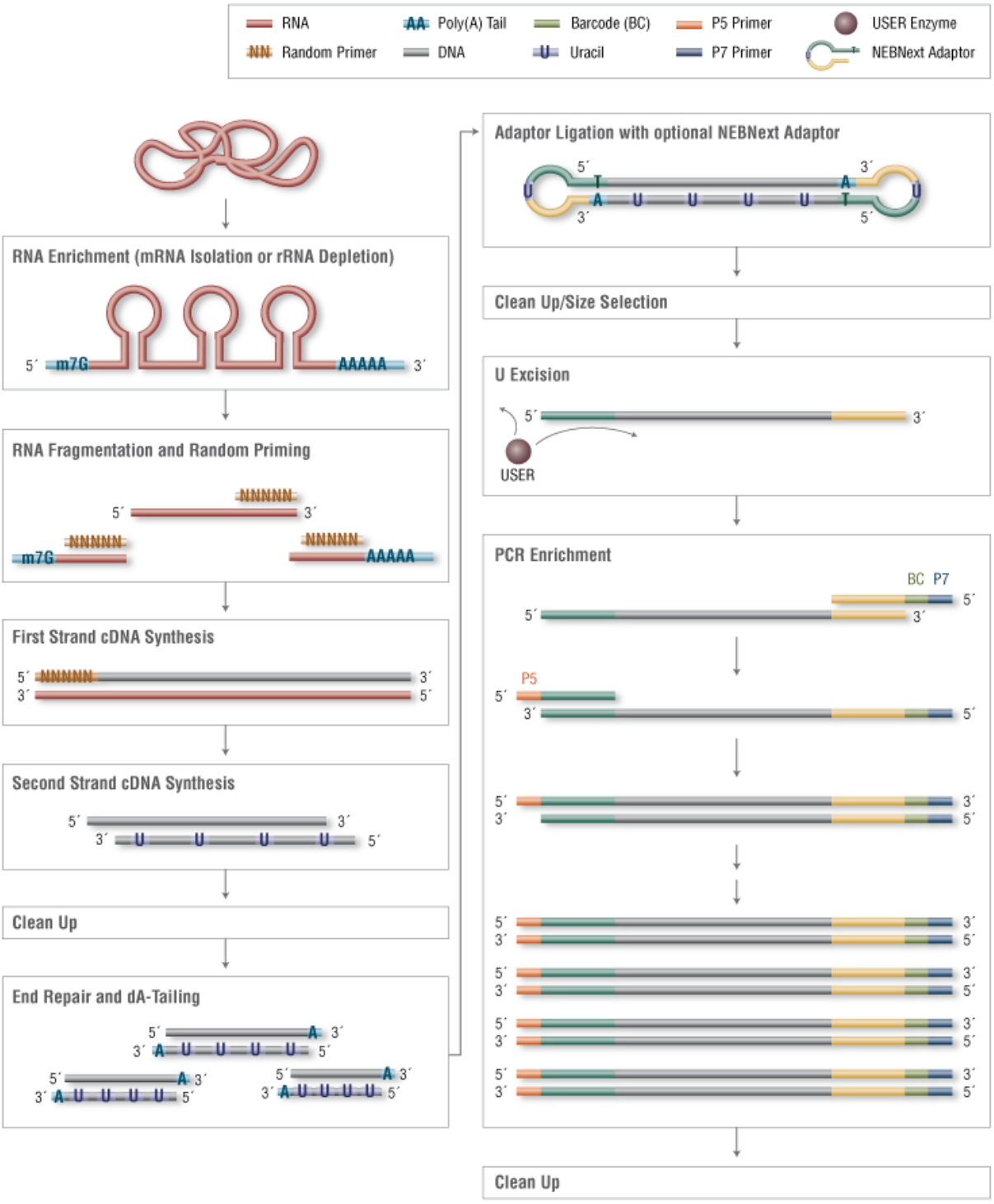


Figure 38. - Workflow for preparation of RNA library

6.4. Bioinformatics and statistics

Approximatelly 346 million reads were generated from *Arion vulgaris* transcriptomics study. These were filtered and trimmed by prinseq-lite using parameters trim_qual_right 30, min_qual_mean 30, min_len 70, trim_tail_right 6, trim_tail_left6, and rRNA sequences were removed using SortMeRNA 1.99 (274). Digital normalization (maximal coverage 30) and transcript assembly were performed using Trinity pipeline (275) (release 2013-02- 25) resulting in 136,406 assembled contigs with average length of 671.04 bp and N50 of 971 bp. CDS annotation was performed using the TransDecoder tool and PFAM version 27.0. Sequence similarities were obtained by blastx searches (cut-off E-value 1E-6) of the contigs and predicted CDS, respectively, against the NCBI non-redundant protein database (version July 2014). The assembled transcripts were further annotated with GO terms, PFAM protein domains and Enzyme Commision (EC) numbers using web platform FastAnnotator (166) with the default search parameters. 36,948 genes were assigned to the records present in NCBI non-redundant database and 22,868 of them to Gene Orthology terms. Searching against Pfam database identified 21,651 of entries to have at least one domain and 2,336 of genes were assigned with EC numbers. Duplicates and substrings were removed from the translated sequence dataset and resulting protein sequences were post-assembled using an in-house script to reduce redundancy. Final assembly containing 47,451 non-redundant protein sequences was used as an expression dataset for MS protein identification.

7. REFERENCES

1. Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphery-Smith I, Hochstrasser DF, et al. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. Biotechnol Genet Eng Rev. 1996;13:19-50.
2. Dhingra V, Gupta M, Andacht T, Fu ZF. New frontiers in proteomics research: a perspective. Int J Pharm. 2005;299(1-2):1-18.
3. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998;19(11):1853-61.
4. Neubauer G, Gottschalk A, Fabrizio P, Seraphin B, Luhrmann R, Mann M. Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. Proc Natl Acad Sci U S A. 1997;94(2):385-90.
5. Shevchenko A, Keller P, Scheiffele P, Mann M, Simons K. Identification of components of trans-Golgi network-derived transport vesicles and detergent-insoluble complexes by nanoelectrospray tandem mass spectrometry. Electrophoresis. 1997;18(14):2591-600.
6. Pandey A, Fernandez MM, Steen H, Blagoev B, Nielsen MM, Roche S, et al. Identification of a novel immunoreceptor tyrosine-based activation motif-containing molecule, STAM2, by mass spectrometry and its involvement in growth factor and cytokine receptor signaling pathways. J Biol Chem. 2000;275(49):38633-9.
7. Godovac-Zimmermann J, Brown LR. Perspectives for mass spectrometry and functional proteomics. Mass Spectrom Rev. 2001;20(1):1-57.
8. Alberts B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. Cell. 1998;92(3):291-4.
9. Celis JE, Gromov P. 2D protein electrophoresis: can it be perfected? Curr Opin Biotech. 1999;10(1):16-21.
10. Rabilloud T. Solubilization of proteins for electrophoretic analyses. Electrophoresis. 1996;17(5):813-29.
11. Bodzon-Kulakowska A, Bierczynska-Krzysik A, Dylag T, Drabik A, Suder P, Noga M, et al. Methods for samples preparation in proteomic research. J Chromatogr B Analyt Technol Biomed Life Sci. 2007;849(1-2):1-31.
12. Martinac A, Filipovic-Grcic J, Perissutti B, Voinovich D, Pavelic Z. Spray-dried chitosan/ethylcellulose microspheres for nasal drug delivery: swelling study and evaluation of in vitro drug release properties. J Microencapsul. 2005;22(5):549-61.
13. Pacheco LF, Carmona-Ribeiro AM. Effects of synthetic lipids on solubilization and colloid stability of hydrophobic drugs. J Colloid Interface Sci. 2003;258(1):146-54.
14. Niemczewski B. Observations of water cavitation intensity under practical ultrasonic cleaning conditions. Ultrason Sonochem. 2007;14(1):13-8.
15. Watkins LK, Bondarenko PV, Barbacci DC, Song S, Cockrill SL, Russell DH, et al. Fast C18 solid-phase desalting/delipidation of the human serum apolipoproteins for matrix-assisted laser desorption ionization and electrospray ionization mass spectrometric analysis. J Chromatogr A. 1999;840(2):183-93.
16. Joo WA, Lee DY, Kim CW. Development of an effective sample preparation method for the proteome analysis of body fluids using 2-D gel electrophoresis. Biosci Biotechnol Biochem. 2003;67(7):1574-7.
17. Rabilloud T. Solubilization of proteins for electrophoretic analyses. Electrophoresis. 1996;17(5):813-29.

18. Wilkins MR. Proteome research : new frontiers in functional genomics. Berlin ; New York: Springer; 1997. xviii, 243 p. p.
19. Schnack HG, van Haren NE, Brouwer RM, van Baal GC, Picchioni M, Weisbrod M, et al. Mapping reliability in multicenter MRI: voxel-based morphometry and cortical thickness. Hum Brain Mapp. 2010;31(12):1967-82.
20. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. P Natl Acad Sci USA. 2009;106(6):1760-5.
21. Dzieciatkowska M, Hill R, Hansen KC. GeLC-MS/MS analysis of complex protein mixtures. Methods Mol Biol. 2014;1156:53-66.
22. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR, 3rd. Protein analysis by shotgun/bottom-up proteomics. Chem Rev. 2013;113(4):2343-94.
23. Piersma SR, Warmoes MO, de Wit M, de Reus I, Knol JC, Jimenez CR. Whole gel processing procedure for GeLC-MS/MS based proteomics. Proteome Sci. 2013;11.
24. Veronesi G, Aude-Garcia C, Kieffer I, Gallon T, Delangle P, Herlin-Boime N, et al. Exposure-dependent Ag+ release from silver nanoparticles and its complexation in AgS2 sites in primary murine macrophages. Nanoscale. 2015;7(16):7323-30.
25. Zilberstein GV, Baskin EM, Bukshpan S, Korol LE. Parallel isoelectric focusing II. Electrophoresis. 2004;25(21-22):3643-51.
26. Gordon JA, Jencks WP. Relationship of Structure to Effectiveness of Denaturing Agents for Proteins. Biochemistry-Us. 1963;2(1):47-&.
27. Rabilloud T, Adessi C, Giraudel A, Lunardi J. Improvement of the solubilization of proteins in two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis. 1997;18(3-4):307-16.
28. Olivieri E, Herbert B, Righetti PG. The effect of protease inhibitors on the two-dimensional electrophoresis pattern of red blood cell membranes. Electrophoresis. 2001;22(3):560-5.
29. Herbert B. Advances in protein solubilisation for two-dimensional electrophoresis. Electrophoresis. 1999;20(4-5):660-3.
30. Molloy MP, Herbert BR, Walsh BJ, Tyler MI, Traini M, Sanchez JC, et al. Extraction of membrane proteins by differential solubilization for separation using two-dimensional gel electrophoresis. Electrophoresis. 1998;19(5):837-44.
31. Finnie C, Svensson B. Proteolysis during the isoelectric focusing step of two-dimensional gel electrophoresis may be a common problem. Anal Biochem. 2002;311(2):182-6.
32. Freeman WM, Hemby SE. Proteomics for protein expression profiling in neuroscience. Neurochemical Research. 2004;29(6):1065-81.
33. Rabilloud T, Lelong C. Two-dimensional gel electrophoresis in proteomics: A tutorial. Journal of Proteomics. 2011;74(10):1829-41.
34. Hecker M, Antelmann H, Buettner K, Bernhardt J. Gel-based proteomics of Gram-positive bacteria: A powerful tool to address physiological questions. Proteomics. 2008;8(23-24):4958-75.
35. Afjehi-Sadat L, Lubec G. Identification of enzymes and activity from two-dimensional gel electrophoresis. Nature Protocols. 2007;2(10):2318-24.
36. Pitarch A, Nombela C, Gil C. Prediction of the Clinical Outcome in Invasive Candidiasis Patients Based on Molecular Fingerprints of Five Anti-Candida Antibodies in Serum. Mol Cell Proteomics. 2011;10(1).
37. Sanchez-Campillo M, Bini L, Comanducci R, Raggiaschi R, Marzocchi B, Pallini V, et al. Identification of immunoreactive proteins of Chlamydia trachomatis by Western blot analysis of a two-dimensional electrophoresis map with patient sera. Electrophoresis. 1999;20(11):2269-79.
38. Ishigami A, Ohsawa T, Hiratsuka M, Taguchi H, Kobayashi S, Saito Y, et al. Abnormal accumulation of citrullinated proteins catalyzed by peptidylarginine

deiminase in hippocampal extracts from patients with Alzheimer's disease. J Neurosci Res. 2005;80(1):120-8.
39. Perluigi M, Poon HF, Hensley K, Pierce WM, Klein JB, Calabrese V, et al. Proteomic analysis of 4-hydroxy-2-nonenal-modified proteins in G93A-SOD1 transgenic mice - A model of familial amyotrophic lateral sclerosis. Free Radical Bio Med. 2005;38(7):960-8.
40. Sobel A, Tashjian AH. Distinct Patterns of Cytoplasmic Protein-Phosphorylation Related to Regulation of Synthesis and Release of Prolactin by Gh Cells. Journal of Biological Chemistry. 1983;258(17):312-24.
41. Bredemeyer AJ, Lewis RM, Malone JP, Davis AE, Gross J, Townsend RR, et al. A proteomic approach for the discovery of protease substrates. P Natl Acad Sci USA. 2004;101(32):11785-90.
42. Patterson SD. Proteomics: evolution of the technology. Biotechniques. 2003;35(3):440-4.
43. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics. 2002;1(11):845-67.
44. Sitnikov D, Chan D, Thibaudeau E, Pinard M, Hunter JM. Protein depletion from blood plasma using a volatile buffer. J Chromatogr B. 2006;832(1):41-6.
45. Govorukhina NI, Keizer-Gunnink A, van der Zee AGJ, de Jong S, de Bruijn HWA, Bischoff R. Sample preparation of human serum for the analysis of tumor markers - Comparison of different approaches for albumin and gamma-globulin depletion. Journal of Chromatography A. 2003;1009(1-2):171-8.
46. Gianazza E, Arnaud P. A general method for fractionation of plasma proteins. Dye-ligand affinity chromatography on immobilized Cibacron blue F3-GA. Biochem J. 1982;201(1):129-36.
47. Fountoulakis M, Juranville JF, Jiang L, Avila D, Roder D, Jakob P, et al. Depletion of the high-abundance plasma proteins. Amino Acids. 2004;27(3-4):249-59.
48. Gorg A, Obermaier C, Boguth G, Harder A, Scheibe B, Wildgruber R, et al. The current state of two-dimensional electrophoresis with immobilized pH gradients. Electrophoresis. 2000;21(6):1037-53.
49. Gorg A, Weiss W, Dunn MJ. Current two-dimensional electrophoresis technology for proteomics. Proteomics. 2004;4(12):3665-85.
50. Lee JE, Lee JY, Kim HR, Shin HY, Lin T, Jin DI. Proteomic Analysis of Bovine Pregnancy-specific Serum Proteins by 2D Fluorescence Difference Gel Electrophoresis. Asian-Australas J Anim Sci. 2015;28(6):788-95.
51. Stasyk T, Morandell S, Bakry R, Feuerstein I, Huck CW, Stecher G, et al. Quantitative detection of phosphoproteins by combination of two-dimensional difference gel electrophoresis and phosphospecific fluorescent staining. Electrophoresis. 2005;26(14):2850-4.
52. Budin I, Devaraj NK. Membrane assembly driven by a biomimetic coupling reaction. J Am Chem Soc. 2012;134(2):751-3.
53. Alberts B. Molecular biology of the cell. 4th ed. New York: Garland Science; 2002. xxxiv, 1548 p. p.
54. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305(3):567-80.
55. Almen MS, Nordstrom KJ, Fredriksson R, Schioth HB. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. BMC Biol. 2009;7:50.
56. von Heijne G. Membrane-protein topology. Nat Rev Mol Cell Biol. 2006;7(12):909-18.

57. Lai X. Reproducible method to enrich membrane proteins with high purity and high yield for an LC-MS/MS approach in quantitative membrane proteomics. Electrophoresis. 2013;34(6):809-17.
58. Zhang ZJ, Peck SC. Simplified enrichment of plasma membrane proteins for proteomic analyses in Arabidopsis thaliana. Proteomics. 2011;11(9):1780-8.
59. Zhang L, Xie J, Wang X, Liu X, Tang X, Cao R, et al. Proteomic analysis of mouse liver plasma membrane: use of differential extraction to enrich hydrophobic membrane proteins. Proteomics. 2005;5(17):4510-24.
60. Zhang LJ, Wang XE, Peng X, Wei YJ, Cao R, Liu Z, et al. Proteomic analysis of low-abundant integral plasma membrane proteins based on gels. Cell Mol Life Sci. 2006;63(15):1790-804.
61. Hunzinger C, Wozny W, Schwall GP, Poznanovic S, Stegmann W, Zengerling H, et al. Comparative profiling of the mammalian mitochondrial proteome: multiple aconitase-2 isoforms including N-formylkynurenine modifications as part of a protein biomarker signature for reactive oxidative species. J Proteome Res. 2006;5(3):625-33.
62. Reifschneider NH, Goto S, Nakamoto H, Takahashi R, Sugawa M, Dencher NA, et al. Defining the mitochondrial proteomes from five rat organs in a physiologically significant context using 2D blue-native/SDS-PAGE. J Proteome Res. 2006;5(5):1117-32.
63. Burre J, Beckhaus T, Schagger H, Corvey C, Hofmann S, Karas M, et al. Analysis of the synaptic vesicle proteome using three gel-based protein separation techniques. Proteomics. 2006;6(23):6250-62.
64. Coughenour HD, Spaulding RS, Thompson CM. The synaptic vesicle proteome: a comparative study in membrane protein identification. Proteomics. 2004;4(10):3141-55.
65. Hinton R, Dobrota M. Density gradient centrifugation. Amsterdam ; New York

New York: North-Holland Pub. Co. ;

sole distributors for the U.S.A. and Canada, Elsevier/North-Holland; 1976. 290 p. p.
66. Ferro M, Seigneurin-Berny D, Rolland N, Chapel A, Salvi D, Garin J, et al. Organic solvent extraction as a versatile procedure to identify hydrophobic chloroplast membrane proteins. Electrophoresis. 2000;21(16):3517-26.
67. Sivars U, Tjerneld F. Mechanisms of phase behaviour and protein partitioning in detergent/polymer aqueous two-phase systems for purification of integral membrane proteins. Bba-Gen Subjects. 2000;1474(2):133-46.
68. Everberg H, Sivars U, Emanuelsson C, Persson C, Englund AK, Haneskog L, et al. Protein pre-fractionation in detergent-polymer aqueous two-phase systems for facilitated proteomic studies of membrane proteins. Journal of Chromatography A. 2004;1029(1-2):113-24.
69. Collaboration CMS, Chatrchyan S, Khachatryan V, Sirunyan AM, Tumasyan A, Adam W, et al. Search for a standard-model-like Higgs boson with a mass in the range 145 to 1000 GeV at the LHC. Eur Phys J C Part Fields. 2013;73(6):2469.
70. Walter H, Sasakawa S. Partition of Proteins in Aqueous Dextran D-Polyethylene Glycol Peg Two-Phase Systems. Federation Proceedings. 1969;28(2):907-&.
71. Polgar L. The catalytic triad of serine peptidases. Cellular and Molecular Life Sciences. 2005;62(19-20):2161-72.
72. Oda K. New families of carboxyl peptidases: serine-carboxyl peptidases and glutamic peptidases. J Biochem. 2012;151(1):13-25.
73. Buller AR, Townsend CA. Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. P Natl Acad Sci USA. 2013;110(8):E653-E61.

74. Dodson G, Wlodawer A. Catalytic triads and their relatives. Trends Biochem Sci. 1998;23(9):347-52.
75. Medzihradszky KF. In-solution digestion of proteins for mass spectrometry. Methods Enzymol. 2005;405:50-65.
76. Gundry RL, White MY, Murray CI, Kane LA, Fu Q, Stanley BA, et al. Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. Curr Protoc Mol Biol. 2009;Chapter 10:Unit10 25.
77. Rosenfeld J, Capdevielle J, Guillemot JC, Ferrara P. In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. Anal Biochem. 1992;203(1):173-9.
78. Shevchenko A, Loboda A, Ens W, Schraven B, Standing KG, Shevchenko A. Archived polyacrylamide gels as a resource for proteome characterization by mass spectrometry. Electrophoresis. 2001;22(6):1194-203.
79. Havlis J, Thomas H, Sebela M, Shevchenko A. Fast-response proteomics by accelerated in-gel digestion of proteins. Anal Chem. 2003;75(6):1300-6.
80. Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem. 1996;68(5):850-8.
81. Speicher KD, Kolbas O, Harper S, Speicher DW. Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. J Biomol Tech. 2000;11(2):74-86.
82. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc. 2006;1(6):2856-60.
83. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nat Methods. 2009;6(5):359-62.
84. Washburn MP, Wolters D, Yates JR, 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol. 2001;19(3):242-7.
85. Blonder J, Xiao Z, Veenstra TD. Proteomic profiling of differentiating osteoblasts. Expert Rev Proteomics. 2006;3(5):483-96.
86. Erde J, Loo RR, Loo JA. Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments. J Proteome Res. 2014;13(4):1885-95.
87. Maurer M, Muller AC, Parapatics K, Pickl WF, Wagner C, Rudashevskaya EL, et al. Comprehensive Comparative and Semiquantitative Proteome of a Very Low Number of Native and Matched Epstein-Barr-Virus-Transformed B Lymphocytes Infiltrating Human Melanoma. Journal of Proteome Research. 2014;13(6):2830-45.
88. Tanca A, Biosa G, Pagnozzi D, Addis MF, Uzzau S. Comparison of detergent-based sample preparation workflows for LTQ-Orbitrap analysis of the Escherichia coli proteome. Proteomics. 2013;13(17):2597-607.
89. Wisniewski JR, Zougman A, Mann M. Combination of FASP and StageTip-Based Fractionation Allows In-Depth Analysis of the Hippocampal Membrane Proteome. Journal of Proteome Research. 2009;8(12):5674-8.
90. Ostasiewicz P, Zielinska DF, Mann M, Wisniewski JR. Proteome, Phosphoproteome, and N-Glycoproteome Are Quantitatively Preserved in Formalin-Fixed Paraffin-Embedded Tissue and Analyzable by High-Resolution Mass Spectrometry. Journal of Proteome Research. 2010;9(7):3688-700.
91. Zhao Q, Fang F, Wu C, Wu Q, Liang Y, Liang Z, et al. imFASP: An integrated approach combining in-situ filter-aided sample pretreatment with microwave-assisted protein digestion for fast and efficient proteome sample preparation. Anal Chim Acta. 2016;912:58-64.

92. Molnar I, Horvath C. Reverse-Phase Chromatography of Polar Biological Substances - Separation of Catechol Compounds by High-Performance Liquid-Chromatography. Clin Chem. 1976;22(9):1497-502.
93. Cuatreca.P. Affinity Chromatography, Principles and Applications. Abstr Pap Am Chem S. 1974:80-.
94. Zhang X, Fang A, Riley CP, Wang M, Regnier FE, Buck C. Multi-dimensional liquid chromatography in proteomics--a review. Anal Chim Acta. 2010;664(2):101-13.
95. Slebos RJ, Brock JW, Winters NF, Stuart SR, Martinez MA, Li M, et al. Evaluation of strong cation exchange versus isoelectric focusing of peptides for multidimensional liquid chromatography-tandem mass spectrometry. J Proteome Res. 2008;7(12):5286-94.
96. Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci. 2005;28(14):1694-703.
97. Sparkman OD. Mass spectrometry desk reference. 1st ed. Pittsburgh, Pa.: Global View Pub.; 2000. xvi, 106 p. p.
98. Ho CS, Lam CW, Chan MH, Cheung RC, Law LK, Lit LC, et al. Electrospray ionisation mass spectrometry: principles and clinical applications. Clin Biochem Rev. 2003;24(1):3-12.
99. Pozniak BP, Cole RB. Current measurements within the electrospray emitter. J Am Soc Mass Spectr. 2007;18(4):737-48.
100. Rapp U. Mass Spectrometry and Proteomics: Principles and Applications. Handbook of Toxicogenomics: Strategies and Applications. 2005:97-114.
101. Kingdon KH. A method for the neutralization of electron space charge by positive ionization at very low gas pressures. Phys Rev. 1923;21(4):408-18.
102. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new mass spectrometer. J Mass Spectrom. 2005;40(4):430-43.
103. McAlister GC, Berggren WT, Griep-Raming J, Horning S, Makarov A, Phanstiel D, et al. A proteomics grade electron transfer dissociation-enabled hybrid linear ion trap-orbitrap mass spectrometer. Journal of Proteome Research. 2008;7(8):3127-36.
104. Wells JM, McLuckey SA. Collision-induced dissociation (CID) of peptides and proteins. Methods Enzymol. 2005;402:148-85.
105. Cooper HJ, Hakansson K, Marshall AG. The role of electron capture dissociation in biomolecular analysis. Mass Spectrom Rev. 2005;24(2):201-22.
106. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A. 2004;101(26):9528-33.
107. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984;11(11):601.
108. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. Anal Chem. 1987;59(21):2621-5.
109. Schilling GD, Andrade FJ, Barnes JHt, Sperline RP, Denton MB, Barinaga CJ, et al. Continuous simultaneous detection in mass spectrometry. Anal Chem. 2007;79(20):7662-8.
110. Kohler RE. The eighth day of creation: Makers of the revolution in biology. Isis. 1997;88(4):730-1.
111. Eddy SR. Non-coding RNA genes and the modern RNA world. Nat Rev Genet. 2001;2(12):919-29.
112. Gesteland RF, Cech TR, Atkins JF. THE RNA WORLD THIRD EDITION The Nature of Modern RNA Suggests a Prebiotic RNA World Preface. Cold Spring Harb Mon. 2006;43:Ix-X.

113. Brantl S. Antisense-RNA regulation and RNA interference. Biochim Biophys Acta. 2002;1575(1-3):15-25.
114. Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science. 2008;321(5891):960-4.
115. Lin SL, Miller JD, Ying SY. Intronic microRNA (miRNA). J Biomed Biotechnol. 2006;2006(4):26818.
116. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94-108.
117. Ahmad K, Henikoff S. Epigenetic consequences of nucleosome dynamics. Cell. 2002;111(3):281-4.
118. Anderson L, Seilhamer J. A comparison of selected mRNA and protein abundances in human liver. Electrophoresis. 1997;18(3-4):533-7.
119. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardia SL, et al. Discordant protein and mRNA expression in lung adenocarcinomas. Mol Cell Proteomics. 2002;1(4):304-13.
120. Mootha VK, Bunkenborg J, Olsen JV, Hjerrild M, Wisniewski JR, Stahl E, et al. Integrated analysis of protein composition, tissue diversity, and gene regulation in mouse mitochondria. Cell. 2003;115(5):629-40.
121. Lichtinghagen R, Musholt PB, Lein M, Romer A, Rudolph B, Kristiansen G, et al. Different mRNA and protein expression of matrix metalloproteinases 2 and 9 and tissue inhibitor of metalloproteinases 1 in benign and malignant prostate tissue. Eur Urol. 2002;42(4):398-406.
122. Nie L, Wu G, Culley DE, Scholten JC, Zhang W. Integrative analysis of transcriptomic and proteomic data: challenges, solutions and applications. Crit Rev Biotechnol. 2007;27(2):63-75.
123. Wu X, Brewer G. The regulation of mRNA stability in mammalian cells: 2.0. Gene. 2012;500(1):10-21.
124. Yeat NC, Lin C, Sager M, Lin J. Cancer proteomics: developments in technology, clinical use and commercialization. Expert Rev Proteomic. 2015;12(4):391-405.
125. Papassotiropoulos A, Fountoulakis M, Dunckley T, Stephan DA, Reiman EM. Genetics, transcriptomics, and proteomics of Alzheimer's disease. J Clin Psychiatry. 2006;67(4):652-70.
126. Wang Y, Yan HL, Wang YP, Yang HL, Wei L, Xiao Y, et al. Proteomics and transcriptome analysis coupled with pharmacological test reveals the diversity of anti-thrombosis proteins from the medicinal insect, Eupolyphaga sinensis. Insect Biochem Molec. 2012;42(8):537-44.
127. Cheng HC, Qi RZ, Paudel H, Zhu HJ. Regulation and function of protein kinases and phosphatases. Enzyme Res. 2011;2011:794089.
128. Brotz-Oesterhelt H, Bandow JE, Labischinski H. Bacterial proteomics and its role in antibacterial drug discovery. Mass Spectrom Rev. 2005;24(4):549-65.
129. Lee JH, Lee DE, Lee BU, Kim HS. Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain. J Bacteriol. 2003;185(18):5442-51.
130. Shimizu T, Shima K, Yoshino K, Yonezawa K, Shimizu T, Hayashi H. Proteome and transcriptome analysis of the virulence genes regulated by the VirR/VirS system in Clostridium perfringens. J Bacteriol. 2002;184(10):2587-94.
131. Fields S, Johnston M. Cell biology. Whither model organism research? Science. 2005;307(5717):1885-6.
132. Seok J, Warren HS, Cuenca AG, Mindrinos MN, Baker HV, Xu W, et al. Genomic responses in mouse models poorly mimic human inflammatory diseases. Proc Natl Acad Sci U S A. 2013;110(9):3507-12.

133. Jubb AW, Young RS, Hume DA, Bickmore WA. Enhancer Turnover Is Associated with a Divergent Transcriptional Response to Glucocorticoid in Mouse and Human Macrophages. J Immunol. 2016;196(2):813-22.
134. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell. 2002;1(2):133-43.
135. Chang JC, Wooten EC, Tsimelzon A, Hilsenbeck SG, Gutierrez MC, Elledge R, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. Lancet. 2003;362(9381):362-9.
136. Zhang J, Abrams Z, Parvin JD, Huang K. Integrative analysis of somatic mutations and transcriptomic data to functionally stratify breast cancer patients. Bmc Genomics. 2016;17 Suppl 7:513.
137. Li H, Zhu Y, Tang X, Li J, Li Y, Zhong Z, et al. Integrated analysis of transcriptome in cancer patient-derived xenografts. PLoS One. 2015;10(5):e0124780.
138. Bings NH. Plasma time-of-flight mass spectrometry as a detector for short transient signals in elemental analysis. Anal Bioanal Chem. 2005;382(4):887-90.
139. da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, et al. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Mar Genomics. 2016.
140. Ziemert N, Lechner A, Wietz M, Millan-Aguinaga N, Chavarria KL, Jensen PR. Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. Proc Natl Acad Sci U S A. 2014;111(12):E1130-9.
141. Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst Biol. 2012;61(5):717-26.
142. McCormack JE, Faircloth BC, Crawford NG, Gowaty PA, Brumfield RT, Glenn TC. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 2012;22(4):746-54.
143. Todd EV, Black MA, Gemmell NJ. The power and promise of RNA-seq in ecology and evolution. Mol Ecol. 2016;25(6):1224-41.
144. da Fonseca RR, Albrechtsen A, Themudo GE, Ramos-Madrigal J, Sibbesen JA, Maretty L, et al. Next-generation biology: Sequencing and data analysis approaches for non-model organisms. Mar Genomics. 2016;30:3-13.
145. Barker GM. The biology of terrestrial molluscs. Wallingford, Oxon, UK ; New York, NY, USA: CABI Pub.; 2001. xiv, 558 p. p.
146. Kałuski T, Kozłowski J, Kozłowska M. Evaluation of damage to seeds and seedlings of different winter wheat cultivars caused by *Deroceras reticulatum* (O.F. Müller, 1774) and *Arion lusitanicus Mabille*, 1868 in laboratory conditions. Folia Malacologica. 2005;13:189-95.
147. Cowie RH, Dillon RT, Robinson DG, Smith JW. Alien non-marine snails and slugs of priority quarantine importance in the United States: A preliminary risk assessment. Am Malacol Bull. 2009;27(1-2):113-32.
148. Gismervik K, Bruheim T, Rorvik LM, Haukeland S, Skaar I. Invasive slug populations (Arion vulgaris) as potential vectors for Clostridium botulinum. Acta Vet Scand. 2014;56.
149. Gismervik K, Aspholm M, Rorvik LM, Bruheim T, Andersen A, Skaar I. Invading slugs (Arion vulgaris) can be vectors for Listeria monocytogenes. Journal of Applied Microbiology. 2015;118(4):809-16.
150. Briner T, Frank T. The palatability of 78 wildflower strip plants to the slug Arion lusitanicus. Ann Appl Biol. 1998;133(1):123-33.
151. Lindstrom M, Myllykoski J, Sivela S, Korkeala H. Clostridium botulinum in Cattle and Dairy Products. Crit Rev Food Sci. 2010;50(4):281-304.

152. Lindstrom M, Korkeala H. Laboratory diagnostics of botulism. Clin Microbiol Rev. 2006;19(2):298-+.
153. Driehuis F. Silage and the safety and quality of dairy foods: a review. Agr Food Sci. 2013;22(1):16-34.
154. Ashburner M, Roote J. Culture of Drosophila: the laboratory setup. CSH Protoc. 2007;2007:pdb ip34.
155. Epstein CJ. The new dysmorphology: application of insights from basic developmental biology to the understanding of human birth defects. Proc Natl Acad Sci U S A. 1995;92(19):8566-73.
156. Campeau P, Mayson P, Parr H, Farrell J, Scriver C. Development of a web log (blog) for the online metabolic and molecular basis of inherited disorders: Facilitating discussions on inborn errors of metabolism. J Inherit Metab Dis. 2006;29:163-.
157. Vikkula M, Boon LM, Carraway KL, 3rd, Calvert JT, Diamonti AJ, Goumnerov B, et al. Vascular dysmorphogenesis caused by an activating mutation in the receptor tyrosine kinase TIE2. Cell. 1996;87(7):1181-90.
158. Marsh JL, Pallos J, Thompson LM. Fly models of Huntington's disease. Hum Mol Genet. 2003;12 Spec No 2:R187-93.
159. Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, et al. Hereditary early-onset Parkinson's disease caused by mutations in PINK1. Science. 2004;304(5674):1158-60.
160. Kopan R, Goate A. Aph-2/Nicastrin: an essential component of gamma-secretase and regulator of Notch signaling and Presenilin localization. Neuron. 2002;33(3):321-4.
161. Sutcliffe JE, Korenjak M, Brehm A. Tumour suppressors - a fly's perspective. Eur J Cancer. 2003;39(10):1355-62.
162. Bier E. Drosophila, the golden bug, emerges as a tool for human genetics. Nat Rev Genet. 2005;6(1):9-23.
163. Brunner E, Ahrens CH, Mohanty S, Baetschmann H, Loevenich S, Potthast F, et al. A high-quality catalog of the Drosophila melanogaster proteome. Nat Biotechnol. 2007;25(5):576-83.
164. Lee TR, Huang SH, Lee CC, Lee HY, Chan HT, Lin KS, et al. Proteome reference map of Drosophila melanogaster head. Proteomics. 2012;12(11):1875-8.
165. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 2011;29(7):644-U130.
166. Chen TW, Gan RCR, Wu TH, Huang PJ, Lee CY, Chen YYM, et al. FastAnnotator- an efficient transcript annotation web tool. Bmc Genomics. 2012;13.
167. Bulat T, Smidak R, Sialana FJ, Jung G, Rattei T, Bilban M, et al. Transcriptomic and Proteomic Analysis of Arion vulgaris--Proteins for Probably Successful Survival Strategies? PLoS One. 2016;11(3):e0150614.
168. Kaessmann H. Origins, evolution, and phenotypic impact of new genes. Genome Res. 2010;20(10):1313-26.
169. Cech TR, Steitz JA. The noncoding RNA revolution-trashing old rules to forge new ones. Cell. 2014;157(1):77-94.
170. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6.
171. Krishna SS, Majumdar I, Grishin NV. Structural classification of zinc fingers: survey and summary. Nucleic Acids Res. 2003;31(2):532-50.
172. Cunha SR, Mohler PJ. Ankyrin protein networks in membrane formation and stabilization. J Cell Mol Med. 2009;13(11-12):4364-76.

173. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol. 2014;32(3):223-6.
174. Voncken F, Gao F, Wadforth C, Harley M, Colasante C. The phosphoarginine energy-buffering system of trypanosoma brucei involves multiple arginine kinase isoforms with different subcellular locations. PLoS One. 2013;8(6):e65908.
175. Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. FEBS J. 2005;272(24):6179-217.
176. Yuasa HJ, Furuta E, Nakamura A, Takagi T. Cloning and sequencing of three C-type lectins from body surface mucus of the land slug, Incilaria fruhstorferi. Comp Biochem Physiol B Biochem Mol Biol. 1998;119(3):479-84.
177. Pennington MW, Lanigan MD, Kalman K, Mahnir VM, Rauer H, McVaugh CT, et al. Role of disulfide bonds in the structure and potassium channel blocking activity of ShK toxin. Biochemistry-Us. 1999;38(44):14549-58.
178. Rauer H, Pennington M, Cahalan M, Chandy KG. Structural conservation of the pores of calcium-activated and voltage-gated potassium channels determined by a sea anemone toxin. J Biol Chem. 1999;274(31):21885-92.
179. Beeton C, Pennington MW, Wulff H, Singh S, Nugent D, Crossley G, et al. Targeting effector memory T cells with a selective peptide inhibitor of Kv1.3 channels for therapy of autoimmune diseases. Mol Pharmacol. 2005;67(4):1369-81.
180. Yan L, Herrington J, Goldberg E, Dulski PM, Bugianesi RM, Slaughter RS, et al. Stichodactyla helianthus peptide, a pharmacological tool for studying Kv3.2 channels. Mol Pharmacol. 2005;67(5):1513-21.
181. Jovine L, Qi H, Williams Z, Litscher E, Wassarman PM. The ZP domain is a conserved module for polymerization of extracellular proteins. Nat Cell Biol. 2002;4(6):457-61.
182. Tordai H, Banyai L, Patthy L. The PAN module: the N-terminal domains of plasminogen and hepatocyte growth factor are homologous with the apple domains of the prekallikrein family and with a novel domain found in numerous nematode proteins. Febs Lett. 1999;461(1-2):63-7.
183. Heaton MP, Clawson ML, Chitko-Mckown CG, Leymaster KA, Smith TP, Harhay GP, et al. Reduced lentivirus susceptibility in sheep with TMEM154 mutations. PLoS Genet. 2012;8(1):e1002467.
184. Lamriben L, Graham JB, Adams BM, Hebert DN. N-Glycan-based ER Molecular Chaperone and Protein Quality Control System: The Calnexin Binding Cycle. Traffic. 2016;17(4):308-26.
185. Zhang Q, Wang XQ, Jiang HS, Jia WM, Zhao XF, Wang JX. Calnexin functions in antibacterial immunity of Marsupenaeus japonicus. Dev Comp Immunol. 2014;46(2):356-63.
186. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, et al. Biochemical and structural analysis of Helix pomatia agglutinin. A hexameric lectin with a novel fold. J Biol Chem. 2006;281(29):20171-80.
187. Rambaruth ND, Greenwell P, Dwek MV. The lectin Helix pomatia agglutinin recognizes O-GlcNAc containing glycoproteins in human breast cancer. Glycobiology. 2012;22(6):839-48.
188. Nickel W. The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. Eur J Biochem. 2003;270(10):2109-19.
189. Fiedler K, Simons K. A Putative Novel Class of Animal Lectins in the Secretory Pathway Homologous to Leguminous Lectins. Cell. 1994;77(5):625-6.
190. Takahashi-Nakaguchi A, Matsumoto Y, Yamamoto M, Iwabuchi K, Totsuka Y, Sugimura T, et al. Demonstration of cytotoxicity against wasps by pierisin-1: a possible defense factor in the cabbage white butterfly. PLoS One. 2013;8(4):e60539.

191. Varki A. Essentials of glycobiology. 2nd ed. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2009. xxix, 784 p. p.
192. Schallus T, Jaeckh C, Feher K, Palma AS, Liu Y, Simpson JC, et al. Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation. Mol Biol Cell. 2008;19(8):3404-14.
193. Deprez P, Gautschi M, Helenius A. More than one glycan is needed for ER glucosidase II to allow entry of glycoproteins into the calnexin/calreticulin cycle. Mol Cell. 2005;19(2):183-95.
194. Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA. Glycosylation and the immune system. Science. 2001;291(5512):2370-6.
195. Gokudan S, Muta T, Tsuda R, Koori K, Kawahara T, Seki N, et al. Horseshoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. Proc Natl Acad Sci U S A. 1999;96(18):10086-91.
196. Saito T, Kawabata S, Hirata M, Iwanaga S. A novel type of limulus lectin-L6. Purification, primary structure, and antibacterial activity. J Biol Chem. 1995;270(24):14493-9.
197. Schroder HC, Ushijima H, Krasko A, Gamulin V, Thakur NL, Diehl-Seifert B, et al. Emergence and disappearance of an immune molecule, an antimicrobial lectin, in basal metazoa. A tachylectin-related protein in the sponge Suberites domuncula. J Biol Chem. 2003;278(35):32810-7.
198. Gerlach D, Schlott B, Schmidt KH. Cloning and expression of a sialic acid-binding lectin from the snail Cepaea hortensis. FEMS Immunol Med Microbiol. 2004;40(3):215-21.
199. Gestal C, Pallavicini A, Venier P, Novoa B, Figueras A. MgC1q, a novel C1q-domain-containing protein involved in the immune response of Mytilus galloprovincialis. Dev Comp Immunol. 2010;34(9):926-34.
200. He X, Zhang Y, Yu F, Yu Z. A novel sialic acid binding lectin with anti-bacterial activity from the Hong Kong oyster (Crassostrea hongkongensis). Fish Shellfish Immunol. 2011;31(6):1247-50.
201. Xu T, Xie J, Li J, Luo M, Ye S, Wu X. Identification of expressed genes in cDNA library of hemocytes from the RLO-challenged oyster, Crassostrea ariakensis Gould with special functional implication of three complement-related fragments (CaC1q1, CaC1q2 and CaC3). Fish Shellfish Immunol. 2012;32(6):1106-16.
202. Yu YH, Huang HQ, Wang Y, Yu YC, Yuan SC, Huang SF, et al. A Novel C1q Family Member of Amphioxus Was Revealed to Have a Partial Function of Vertebrate C1q Molecule. J Immunol. 2008;181(10):7024-32.
203. Kong PF, Zhang HA, Wang LL, Zhou Z, Yang JL, Zhang Y, et al. AiC1qDC-1, a novel gC1q-domain-containing protein from bay scallop Argopecten irradians with fungi agglutinating activity. Developmental and Comparative Immunology. 2010;34(8):837-46.
204. Wang L, Wang L, Zhang H, Zhou Z, Siva VS, Song L. A C1q domain containing protein from scallop Chlamys farreri serving as pattern recognition receptor with heat-aggregated IgG binding activity. PLoS One. 2012;7(8):e43289.
205. Tahtouh M, Croq F, Vizioli J, Sautiere PE, Van Camp C, Salzet M, et al. Evidence for a novel chemotactic C1q domain-containing factor in the leech nerve cord. Mol Immunol. 2009;46(4):523-31.
206. Zhang H, Song L, Li C, Zhao J, Wang H, Qiu L, et al. A novel C1q-domain-containing protein from Zhikong scallop Chlamys farreri with lipopolysaccharide binding activity. Fish Shellfish Immunol. 2008;25(3):281-9.
207. Lehmann F, Tiralongo E, Tiralongo J. Sialic acid-specific lectins: occurrence, specificity and function. Cell Mol Life Sci. 2006;63(12):1331-54.
208. Iwanaga S, Lee BL. Recent advances in the innate immunity of invertebrate animals. J Biochem Mol Biol. 2005;38(2):128-50.

209. Tamplin ML, Fisher WS. Occurrence and characteristics of agglutination of Vibrio cholerae by serum from the eastern oyster, Crassostrea virginica. Appl Environ Microbiol. 1989;55(11):2882-7.
210. Tsuboi I, Matsukawa M, Sato N. Isolation and characterization of a sialic acid-specific lectin from hemolymph of the Southeast Asian horseshoe crab Tachypleus gigas. Biosci Biotechnol Biochem. 1993;57(8):1237-42.
211. Vasta GR, Cohen E. Sialic acid binding lectins in the serum of American spiders of the genus Aphonopelma. Dev Comp Immunol. 1984;8(3):515-22.
212. Grubhoffer L, Kovar V, Rudenko N. Tick lectins: structural and functional properties. Parasitology. 2004;129 Suppl:S113-25.
213. Peitsch MC, Tschopp J. Assembly of macromolecular pores by immune defense systems. Curr Opin Cell Biol. 1991;3(4):710-6.
214. Gonzalez MR, Bischofberger M, Pernot L, van der Goot FG, Freche B. Bacterial pore-forming toxins: the (w)hole story? Cell Mol Life Sci. 2008;65(3):493-507.
215. Dreon MS, Frassa MV, Ceolin M, Ituarte S, Qiu JW, Sun J, et al. Novel animal defenses against predation: a snail egg neurotoxin combining lectin and pore-forming chains that resembles plant defense and bacteria attack toxins. PLoS One. 2013;8(5):e63782.
216. Pringos E, Vignes M, Martinez J, Rolland V. Peptide neurotoxins that affect voltage-gated calcium channels: a close-up on omega-agatoxins. Toxins (Basel). 2011;3(1):17-42.
217. Daly NL, Craik DJ. Bioactive cystine knot proteins. Curr Opin Chem Biol. 2011;15(3):362-8.
218. Craik DJ, Daly NL, Waine C. The cystine knot motif in toxins and implications for drug design. Toxicon. 2001;39(1):43-60.
219. Wang H, Song L, Li C, Zhao J, Zhang H, Ni D, et al. Cloning and characterization of a novel C-type lectin from Zhikong scallop Chlamys farreri. Mol Immunol. 2007;44(5):722-31.
220. Sato S, St-Pierre C, Bhaumik P, Nieminen J. Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). Immunol Rev. 2009;230:172-87.
221. Aradska J, Bulat T, Sialana FJ, Birner-Gruenberger R, Erich B, Lubec G. Gel-free mass spectrometry analysis of Drosophila melanogaster heads. Proteomics. 2015;15(19):3356-60.
222. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics. 2005;4(9):1265-72.
223. Shinoda K, Tomita M, Ishihama Y. emPAI Calc--for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. Bioinformatics. 2010;26(4):576-7.
224. Sonnhammer EL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 1998;6:175-82.
225. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25(2):288-9.
226. Dahanukar A, Hallem EA, Carlson JR. Insect chemoreception. Curr Opin Neurobiol. 2005;15(4):423-30.
227. Stocker RF. The organization of the chemosensory system in Drosophila melanogaster: a review. Cell Tissue Res. 1994;275(1):3-26.

228. Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. Variant Ionotropic Glutamate Receptors as Chemosensory Receptors in Drosophila. Cell. 2009;136(1):149-62.
229. Engelhardt S, Rochais F. G proteins - More than transducers of receptor-generated signals? Circ Res. 2007;100(8):1109-11.
230. Traynelis SF, Wollmuth LP, McBain CJ, Menniti FS, Vance KM, Ogden KK, et al. Glutamate Receptor Ion Channels: Structure, Regulation, and Function. Pharmacol Rev. 2010;62(3):405-96.
231. Yamamoto I. Nicotine to nicotinoids: 1962 to 1997. Abstr Pap Am Chem S. 1997;214:1-Agro.
232. Kandel ER. Principles of neural science. 5th ed. New York: McGraw-Hill; 2013. l, 1709 p. p.
233. Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry. 2008;79(4):368-76.
234. Greenslit N. [Review of: Healy, David. The Creation of psychopharmacology. Cambridge, MA: Harvard University Press, 2002]. Hist Psychiatry. 2004;15(58 Pt 2):227-32.
235. Kimko HC, Cross JT, Abernethy DR. Pharmacokinetics and clinical effectiveness of methylphenidate. Clin Pharmacokinet. 1999;37(6):457-70.
236. Howes OD, Kapur S. The dopamine hypothesis of schizophrenia: version III--the final common pathway. Schizophr Bull. 2009;35(3):549-62.
237. Berridge CW, Devilbiss DM. Psychostimulants as Cognitive Enhancers: The Prefrontal Cortex, Catecholamines, and Attention-Deficit/Hyperactivity Disorder. Biol Psychiat. 2011;69(12):E101-E11.
238. Sharma AM, Pischon T, Hardt S, Kunz I, Luft FC. beta-Adrenergic receptor blockers and weight gain - A systematic analysis. Hypertension. 2001;37(2):250-4.
239. de Boer SF, Koolhaas JM. 5-HT1A and 5-HT1B receptor agonists and aggression: A pharmacological challenge of the serotonin deficiency hypothesis. Eur J Pharmacol. 2005;526(1-3):125-39.
240. Cunha RA, Ferre S, Vaugeois JM, Chen JF. Potential therapeutic interest of adenosine A2A receptors in psychiatric disorders. Curr Pharm Des. 2008;14(15):1512-24.
241. Korogod SM, Kulagina IB. [Dynamical electrical states of heterogeneous populations of ion channels in the membranes of excitable cells]. Fiziol Zh. 2012;58(3):50-9.
242. Camerino DC, Tricarico D, Desaphy JF. Ion channel pharmacology. Neurotherapeutics. 2007;4(2):184-98.
243. Yao CK, Lin YQ, Ly CV, Ohyama T, Haueter CM, Moiseenkova-Bell VY, et al. A synaptic vesicle-associated Ca2+ channel promotes endocytosis and couples exocytosis to endocytosis. Cell. 2009;138(5):947-60.
244. Catterall WA. Voltage-gated calcium channels. Cold Spring Harb Perspect Biol. 2011;3(8):a003947.
245. Shieh CC, Coghlan M, Sullivan JP, Gopalakrishnan M. Potassium channels: molecular defects, diseases, and therapeutic opportunities. Pharmacol Rev. 2000;52(4):557-94.
246. Salzer I, Gantumur E, Yousuf A, Boehm S. Control of sensory neuron excitability by serotonin involves 5HT2C receptors and Ca(2+)-activated chloride channels. Neuropharmacology. 2016;110(Pt A):277-86.
247. Fleck MW, Thomson JL, Hough LB. Histamine-gated ion channels in mammals? Biochem Pharmacol. 2012;83(9):1127-35.
248. Marban E, Yamagishi T, Tomaselli GF. Structure and function of voltage-gated sodium channels. J Physiol-London. 1998;508(3):647-57.

249. Frank CA, Wang X, Collins CA, Rodal AA, Yuan Q, Verstreken P, et al. New approaches for studying synaptic development, function, and plasticity using Drosophila as a model system. J Neurosci. 2013;33(45):17560-8.
250. Wright EM, Hirayama BA, Loo DF. Active sugar transport in health and disease. J Intern Med. 2007;261(1):32-43.
251. Danbolt NC. Glutamate uptake. Prog Neurobiol. 2001;65(1):1-105.
252. Jen JC, Wan J, Palos TP, Howard BD, Baloh RW. Mutation in the glutamate transporter EAAT1 causes episodic ataxia, hemiplegia, and seizures. Neurology. 2005;65(4):529-34.
253. de Vries B, Mamsa H, Stam AH, Wan J, Bakker SL, Vanmolkot KR, et al. Episodic ataxia associated with EAAT1 mutation C186S affecting glutamate reuptake. Arch Neurol. 2009;66(1):97-101.
254. Beart PM, O'Shea RD. Transporters for L-glutamate: An update on their molecular pharmacology and pathological involvement. Brit J Pharmacol. 2007;150(1):5-17.
255. Crowe JH, Crowe LM, Carpenter JF, Wistrom CA. Stabilization of Dry Phospholipid-Bilayers and Proteins by Sugars. Biochem J. 1987;242(1):1-10.
256. Crowe JH, Carpenter JF, Crowe LM. The role of vitrification in anhydrobiosis. Annu Rev Physiol. 1998;60:73-103.
257. Elbein AD, Pan YT, Pastuszak I, Carroll D. New insights on trehalose: a multifunctional molecule. Glycobiology. 2003;13(4):17r-27r.
258. Nishizaki Y, Yoshizane C, Toshimori Y, Arai N, Akamatsu S, Hanaya T, et al. Disaccharide-trehalose inhibits bone resorption in ovariectomized mice. Nutr Res. 2000;20(5):653-64.
259. Davies JE, Sarkar S, Rubinsztein DC. Trehalose reduces aggregate formation and delays pathology in a transgenic mouse model of oculopharyngeal muscular dystrophy. Human Molecular Genetics. 2006;15(1):23-31.
260. Sarkar S, Davies JE, Huang ZB, Tunnacliffe A, Rubinsztein DC. Trehalose, a novel mTOR-independent autophagy enhancer, accelerates the clearance of mutant huntingtin and alpha-synuclein. Journal of Biological Chemistry. 2007;282(8):5641-52.
261. Thorens B. Glucose transporters in the regulation of intestinal, renal, and liver glucose fluxes. Am J Physiol-Gastr L. 1996;270(4):G541-G53.
262. Haga T. Molecular properties of the high-affinity choline transporter CHT1. J Biochem. 2014;156(4):181-94.
263. Murai KK, Van Meyel DJ. Neuron-glial communication at synapses: Insights from vertebrates and invertebrates. Neuroscientist. 2007;13(6):657-66.
264. Jensen K, Sanchez-Garcia J, Williams C, Khare S, Mathur K, Graze RM, et al. Purification of transcripts and metabolites from Drosophila heads. J Vis Exp. 2013(73):e50245.
265. Luche S, Santoni V, Rabilloud T. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. Proteomics. 2003;3(3):249-53.
266. Bae N, Lodl M, Pollak A, Lubec G. Mass spectrometrical analysis of bilin-binding protein from the wing of Hebomoia glaucippe (Linnaeus, 1758) (Lepidoptera: Pieridae). Electrophoresis. 2012;33(12):1787-94.
267. Ahmed KE, Chen WQ, John JP, Kang SU, Lubec G. Complete sequencing of the recombinant granulocyte-colony stimulating factor (filgrastim) and detection of biotinylation by mass spectrometry. Amino Acids. 2010;38(4):1043-9.
268. Chen WQ, Li L, Lubec G. Proteomic differences between white and brown adipocytes. Amino Acids. 2014;46(4):995-1008.
269. Kang SU, Heo S, Lubec G. Mass spectrometric analysis of GABAA receptor subtypes and phosphorylations from mouse hippocampus. Proteomics. 2011;11(11):2171-81.

270. Manza LL, Stamer SL, Ham AJ, Codreanu SG, Liebler DC. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics. 2005;5(7):1742-5.
271. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007;2(8):1896-906.
272. Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. J Sep Sci. 2005;28(14):1694-703.
273. Goldberg ME, ExpertBezancon N, Vuillard L, Rabilloud T. Non-detergent sulphobetaines: A new class of molecules that facilitate in vitro protein renaturation. Fold Des. 1996;1(1):21-7.
274. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28(24):3211-7.
275. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-52.

8. APPENDIX

8.1. Publshed papers

Bulat T, Smidak R, Sialana FJ, Jung G, Rattei T, Bilban M, et al. *Transcriptomic and Proteomic Analysis of Arion vulgaris--Proteins for Probably Successful Survival Strategies?* PLoS One. 2016;11(3):e0150614

Aradska J, Bulat T, Sialana FJ, Birner-Gruenberger R, Erich B, Lubec G. *Gel-free mass spectrometry analysis of Drosophila melanogaster heads*. Proteomics. 2015;15(19):3356-60

PLOS | ONE

RESEARCH ARTICLE

Transcriptomic and Proteomic Analysis of *Arion vulgaris*—Proteins for Probably Successful Survival Strategies?

Tanja Bulat[1], Roman Smidak[1], Fernando J. Sialana[1], Gangsoo Jung[1], Thomas Rattei[2], Martin Bilban[3], Helmut Sattmann[4], Gert Lubec[5]*, Jana Aradska[1]*

1 Department of Pediatrics, Medical University of Vienna, Vienna, Austria, **2** Division of Computational System Biology, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria, **3** Department of Laboratory Medicine and Core Facility Genomics, Medical University of Vienna, Vienna, Austria, **4** Third Zoological Department, Museum of Natural History Vienna, Vienna, Austria, **5** Department of Pharmaceutical Chemistry, University of Vienna, Vienna, Austria

* gert.lubec@univie.ac.at (GL); jana.aradska@gmail.com (JA)

Citation: Bulat T, Smidak R, Sialana FJ, Jung G, Rattei T, Bilban M, et al. (2016) Transcriptomic and Proteomic Analysis of *Arion vulgaris*—Proteins for Probably Successful Survival Strategies? PLoS ONE 11(3): e0150614. doi:10.1371/journal.pone.0150614

Editor: Linsheng Song, Institute of Oceanology, Chinese Academy of Sciences, CHINA

Received: December 3, 2015

Accepted: February 17, 2016

Published: March 17, 2016

Data Availability Statement: The sequence data has been submitted to the sequence read archive (SRA) database of GenBank (http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession number PRJEB7891. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaino et al. 2014) via the PRIDE partner repository with the dataset identifier PXD002078 and 10.6019/PXD002078.

Funding: The authors have no support or funding to report.

Abstract

The Spanish slug, *Arion vulgaris*, is considered one of the hundred most invasive species in Central Europe. The immense and very successful adaptation and spreading of *A. vulgaris* suggest that it developed highly effective mechanisms to deal with infections and natural predators. Current transcriptomic and proteomic studies on gastropods have been restricted mainly to marine and freshwater gastropods. No transcriptomic or proteomic study on *A. vulgaris* has been carried out so far, and in the current study, the first transcriptomic database from adult specimen of *A. vulgaris* is reported. To facilitate and enable proteomics in this non-model organism, a mRNA-derived protein database was constructed for protein identification. A gel-based proteomic approach was used to obtain the first generation of a comprehensive slug mantle proteome. A total of 2128 proteins were unambiguously identified; 48 proteins represent novel proteins with no significant homology in NCBI non-redundant database. Combined transcriptomic and proteomic analysis revealed an extensive repertoire of novel proteins with a role in innate immunity including many associated pattern recognition, effector proteins and cytokine-like proteins. The number and diversity in gene families encoding lectins point to a complex defense system, probably as a result of adaptation to a pathogen-rich environment. These results are providing a fundamental and important resource for subsequent studies on molluscs as well as for putative antimicrobial compounds for drug discovery and biomedical applications.

Introduction

Arion vulgaris as the only terrestrial gastropod, is considered among the 100 worst invasive species in Europe. The spread of invasive species, including slugs, is an increasing problem worldwide with an important economical, ecological, health and social impact. The slug *A. vulgaris* is considered a serious pest, both in agriculture and private gardens. Moreover *A. vulgaris*

Competing Interests: The authors have declared that no competing interests exist.

may act as a vector for some pathogenic bacteria, like *Listeria monocytogenes* [1], *Clostridium botulinum* [2],[3] and host species for some parasites, as *Angiostrongylus vasorum* that can cause serious and potentially fatal disease in dogs and other canids [4]. *A. vulgaris* belongs to a species complex *Arion ater* that can only be distinguished by dissecting their reproductive organs.

Currently, transcriptomic studies on molluscs have been restricted to an ecological framework in the marine and fresh water gastropods. However, no transcriptomic or proteomic study has been carried out so far in *A. vulgaris*. Next-generation sequencing has offered a powerful and cost-efficient technique for the generation of transcriptomic datasets in non-model species using diverse platforms such as the Illumina HiSeq, Roche 454, Pacific Biosystems, and Applied Biosystems SOLiD. Several non-model organisms have been characterized by transcriptome sequencing [5–7], which has provided a better understanding of these species.

Since mRNA expression does not necessarily reflect changes at the protein level, complementary proteomic studies can yield more comprehensive biological insight. Mass spectrometry analysis allows large proteome surveys and comparative analysis, where quantities of hundreds or thousands of proteins can be compared between various conditions. However, the lack of a genomic resource for these animals is the major reasons for restricting proteomic applications.

In the current study, the Illumina HiSeq 2000 platform has been used to generate an *A. vulgaris* transcriptome-based protein database. The transcriptome data generated in this study provide a first comprehensive and valuable genomic resource for future research on this slug. With the objective to obtain insight into the slug proteome gel-based proteomic, analysis from the *A. vulgaris* mantle was performed.

Methods

Collection of Arion vulgaris

A. vulgaris specimens were collected in May 2014 in a location (48° 19′ 40″ N, 16° 12′ 34″ E) in Woerdern, Austria. A total of 15 specimens ranging from 7 to 8 g were collected. Species identification was carried out with the assistance of the Zoological Department of Museum of Natural History in Vienna (H.S.). According to national law no permission is required to collect and use invertebrates.

Transcriptome sequencing and assembly

Isolation of RNA was performed with the RNeasy kit (Qiagen, Hilden, Germany). Following removal of the gut, two specimens of *Arion vulgaris* were immersed in liquid nitrogen and pulverized with the help of a mortar and pestle. Pooled total RNA was subjected to RNA-Seq following the mRNA sequencing protocol provided by New England Biolabs (NEBNext kit, NEB, Frankfurt, Germany). The first step in the workflow involved purifying the poly(A)-containing mRNA molecules using two rounds of poly(T)-oligo–attached magnetic beads. After purification, mRNAs were fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first-strand complementary DNA (cDNA) using reverse transcriptase and random primers. This was followed by second-strand cDNA synthesis using DNA polymerase I and RNase H (NEB, Frankfurt, Germany). These cDNA fragments subsequently went through an end-repair process, the addition of a single A base and ligation of the adaptors. The products were then purified and enriched with 15 cycles of PCR (with a size distribution of ~200–250 bp) to create the final cDNA library. Finally, the adaptor-ligated DNA was sequenced for 100 cycles on a HiSeq2000 sequencing system (Illumna, San Diego, USA) in a paired-end manner according to the manufacturer's

instructions. The generated reads were filtered and trimmed by prinseq-lite using parameters trim_qual_right 30, min_qual_mean 30, min_len 70, trim_tail_right 6, trim_tail_left 6, and rRNA sequences were removed using SortMeRNA 1.99 [8]. Digital normalization (maximal coverage 30) and transcript assembly were performed using Trinity pipeline (release 2013-02-25, [9]).

Transcriptome annotation

Protein-coding sequences were identified in the assembly using the TransDecoder tool and PFAM version 27.0. Sequence similarities were obtained by blastx searches (cut-off E-value 1E-6) of the contigs and predicted CDS, respectively, against the NCBI non-redundant protein database (version July 2014). The assembled transcripts were further annotated with GO terms, PFAM protein domains and Enzyme Commision (EC) numbers using web platform FastAnnotator [10] with the default search parameters. Duplicates and substrings were removed from the translated sequence dataset and resulting protein sequences were post-assembled using an in-house script to reduce redundancy. Final assembly containing 47,451 non-redundant protein sequences were used as an expression dataset for MS protein identification.

Protein extraction

Individuals of *A*. *vulgaris* were frozen immediately after harvest in liquid nitrogen and stored at -80°. Mantles were dissected directly before protein extraction procedure and homogenized in urea buffer (7 M urea, 2 M thiourea, 4% CHAPS, 2 M Tris (pH 8.8), 0.5% Carrier (3–11 pH), 1% DTT, 1 mM EDTA, 1 mM PMSF, Protease Inhibitor C). Protein samples were cleaned using a 2-DE clean up kit (BioRad, Hercules, CA, USA) and protein concentration was estimated by the Pierce 660 kit (ThermoScientific, Rockford, USA).

Two dimensional gel electrophoresis (2-DE) and nano-LC-ESI-MS/MS analysis

2-DE was performed as reported previously with minor modifications [11–13]. 750 μg of proteins were loaded on 18 cm 3–11 linear IPG strips (GE Healthcare). IEF was performed in a IPGphore electrophoretic system (GE Healthcare, Uppsala, Sweden) using a protocol with gradually increasing voltage from 200 to 8,000 V at 4 V/min. Prior to second dimension electrophoresis, IPG strips were incubated in 10 ml of equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 2% SDS and 0.002% bromophenol blue) containing 1% DTT for 15 min with gentle shaking followed by incubation in 10 mL of equilibration buffer with 4% iodoacetamide. The second-dimensional separation was performed on 10–16% gradient SDS-PAGE gels. After fixation for 4 h in 50% methanol and 10% acetic acid, gels were stained overnight with the colloidal coomassie blue (Novex life technologies, Invitrogen, CA). Molecular masses were determined by comparison with precision protein standard markers (#1610373, BioRad, Hercules, CA, USA) spanning the 10 to 250 kDa molecular weight range. All spots were excised from 2-DE gels and digested with trypsin and/or chymotrypsin. The gel pieces were cut into small pieces and washed with 50% ACN in 10mM ammonium bicarbonate for 20 min with vortexing and then 20 min with 10mM ammonium bicarbonate. These two steps were repeated until the gel was completely destained. 100% ACN was added for 10min and gel pieces were dried completely using a SpeedVac concentrator for 30 min at 30°C. Cysteine residues were reduced with 10 mM DTT at 56°C for 30 min and alkylated with 55 mM iodoacetamide for 45 min at 22°C. After washing with 10 mM ammonium bicarbonate and dehydratation with 100% ACN, proteins were digested with 12.5 ng/μL trypsin (Promega,

Mannheim, Germany) solution buffered in 10 mM ammonium bicarbonate for 16 h (overnight) at 37°C. The supernatant was transferred to new LoBind 0.5 mL tubes and peptides were extracted with 1% formic acid and then with 15% ACN/0.1% formic acid [14]. 20 uL of extracted peptides were analyzed by nano-LC-ESI-(CID/ETD)-MS/MS essentially as described previously [14]. The HPLC used was an Ultimate 3000 system (Dionex Corporation, Sunnyvale, CA) equipped with a PepMap100 C-18 trap column (300 μm× 5 mm) and PepMap100 C-18 analytical column (75 μm× 250 mm). The gradient was (A = 0.1% formic acid in water, B = 0.08% formic acid in ACN) 4–30% B from 0 to 105 min, 80% B from 105 to 110 min, 4% B from 110 to 125 min. The Amazon speed ETD (Bruker Daltonics, Bremen, Germany) was used to record peptide spectra over the mass range of m/z 400–1,400, and MS/MS spectra in information-dependent data acquisition over the mass range of m/z 100–2,800. Repeatedly, MS spectra were recorded followed by three data-dependent CID MS/MS spectra. An active exclusion of 0.4 min after two spectra was used to detect low abundant peptides. The voltage between ion spray tip and spray shield was set to 1,400 V. Drying nitrogen gas was heated to 150°C and the flow rate was 3 L/min. The collision energy was set automatically according to the mass and charge state of the peptides chosen for fragmentation. Multiple charged peptides were chosen for MS/MS experiments due to their good fragmentation characteristics. MS/MS spectra were interpreted and peak lists were generated by Data Analysis 4.1 (Bruker Daltonics, Bremen, Germany).

One dimensional (1-D) SDS-PAGE and LC-MS/MS

Although 2-DE followed by LC-MS/MS is a proven method for protein separation and identification, it suffers from poor sensitivity, poor representation of very acidic or basic proteins and low solubility of hydrophobic proteins. Therefore, we also separated proteins by 1-D SDS-PAGE coupled with high resolution LC–MS/MS. 50 μg of extracted proteins were mixed with Laemmli buffer (150 mM Tris HCl pH 6.8, 300 mM DTT, 6% SDS, 0.3% bromophenol blue, 30% glycerol) at a 1:3 volume ratio. Proteins in each sample were separated by 10% SDS-PAGE and stained by Blue silver [15]. After destaining with Milli-Q water, gels were cut into 12 slices with proteins of different mass. Proteins underwent in-gel trypsin digestion as above and were analyzed using a LTQ-Orbitrap Velos (ThermoFisher Scientific, Waltham, MA, USA) coupled with nano-LC (Dionex Ultimate 3000) as described previously [16].

Database search

Data generated from Amazon speed ETD for each of the 2-DE protein spots were analyzed by searching the mRNA-derived *Arion vulgaris* database (47,451 sequences) including 115 commonly observed contaminants with Mascot Search enigma (version 2.4) using the Mascot Daemon interface (Matrix Science, London, UK). Detailed search criteria were used as follows; enzyme: trypsin or chymotrypsin with a maximum of two missing cleavage sites; search mode: MS/MS ion search with decoy database search included; fixed modification: carbamidomethylation (C); variable modification: oxidation (M); search mode: MS/MS ion search with decoy database search included; peptide mass tolerance 0.35 Da; MS/MS mass tolerance ± 0.35 Da.

For the mass spectrometry data generated from the LTQ Orbitrap Velos, the MS raw files were first processed with Proteome Discoverer 1.4 (ThermoFisher Scientific, IL, USA) to generate separate Mascot generic files. The database search was performed using Mascot version 2.4 (MatrixScience, London, UK) against the combined database as mentioned above. The search criteria were 10 ppm for precursor and 0.5 Da for fragments; search mode: MS/MS ion search with decoy database search included; fixed modification: carbamidomethyl (cysteine); variable modification: oxidation (methionine).

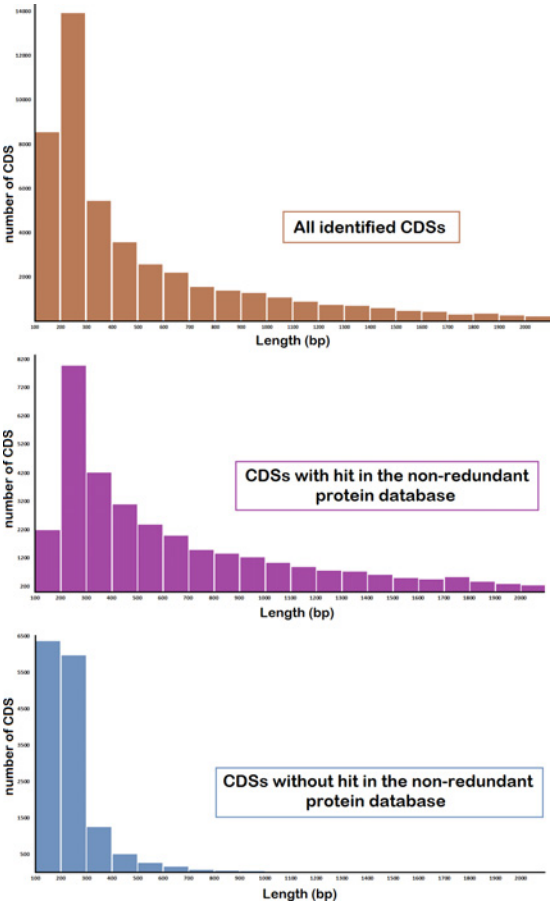


Fig 1. Protein-coding sequences (CDS) distribution showing the majority of sequences in the range from 200 to 300 bp.

doi:10.1371/journal.pone.0150614.g001

Peptide identifications were filtered to a < 1% false discovery rate (FDR) using the target-decoy strategy [17]. Acceptance parameter for protein identification was a minimum of one unique peptide and two distinct peptides.

Results

Transcriptome of whole body

The slug-specific database was constructed using a RNAseq and *de novo* assembly strategy. A cDNA library from the adult slug was generated and using Ilumina HiSeq 2000 technology an approximate number of 339 millions of paired end reads was obtained. *De novo* transcript assembly performed by Trinity pipeline [9] resulted in 136,406 contigs with average lengths of 671.04 bp and N50 of 971 bp, meaning that 50% of the assembled sequences are 971 bp long or longer. 53,523 protein-coding sequences (CDS) were identified using the TransDecoder tool from the Trinity package with PFAM version 27.0. Fig 1 shows the distribution of CDS annotated contig lengths. All CDS annotated contigs of the final assembly were subjected to blastx analysis against the NCBI non-redundant protein database (nr) using the web platform Fastannotator [10] with a cut-off value of 1E-6. Duplicates and substrings were removed from the translated sequence dataset and resulting protein sequences were post-assembled using an in-house script to reduce redundancy. The final assembly contains 47,451 non-redundant protein

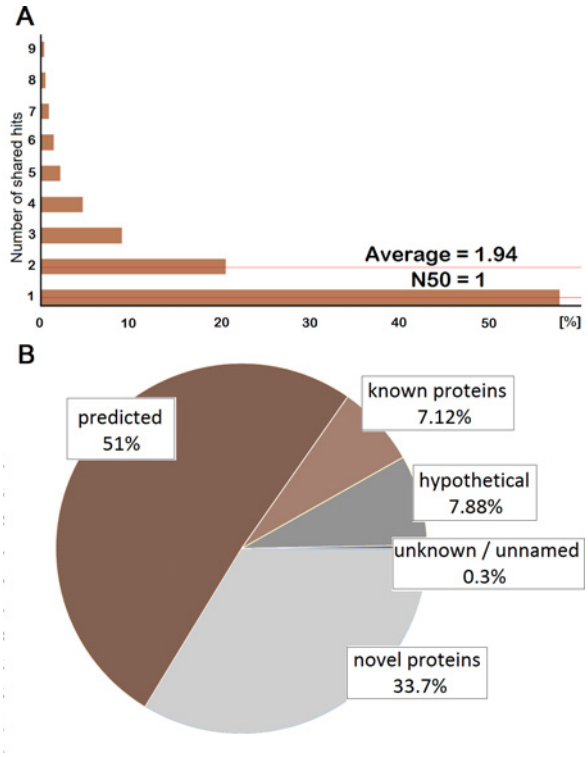


Fig 2. Characteristics of sequences from *A. vulgaris* transcriptome. (a) Distribution of shared top-hits. (b) Classification of top-hits in the current transcriptome.

doi:10.1371/journal.pone.0150614.g002

sequences. On average 1.94 sequences of the final assembly share the same hit against the NCBI nr database (Fig 2A), due to sequencing errors and/or potential isoforms. As a result, 31,463 sequences (66.3%) were assigned to at least one protein and 15,988 (33.7%) transcripts did not have a significant hit in the NCBI nr database probably because of the lack of molecular data of mollusc species. Among 31,463 matches, only 3,380 (10.7%) were known proteins, 24,217 are predicted (77%), 3,739 are hypothetical (11.9%), and 127 are either uncharacterized, unknown or unnamed proteins (0.4%) (Fig 2B). Sequences without significant similarity to sequences present in the nr database are being referred as putative novel genes. These include taxonomically constricted genes derived from ancestral genes or appeared *de novo* from non-coding sequences [18]. The high number of transcripts lacking detectable homology with protein sequences described in the database might be also due to the high number of short transcripts (147–300 bp) in the final assembly. Of the 47,451 transcripts in the final assembly, 46.5% (22,055 transcripts) are shorter than 300 bp (Fig 1). It has been shown that non-coding RNAs [19] as well as small open reading frames are potential sources of sequences lacking detectable homology with protein sequences in nr databases. A high number of putative novel genes was identified but additional genetic as well as proteomic studies are needed to annotate them correctly and this was one of the aims of the current proteomic study described below.

39% (18,561) of all sequences were functionally annotated with Gene Onthology (GO) terms using BLAST2GO tool of Fastannotator [20]. The results were summarized to the categories "biological process" (33.9%), "molecular function" (28.2%), and "cellular component" (34.8%). Within the biological process classification, metabolic processes (66%), biological regulation (43%) and response to stimulus (29%) were the most representative (Fig 3). In the molecular function category, the highest percentage of GO terms corresponded to "binding"

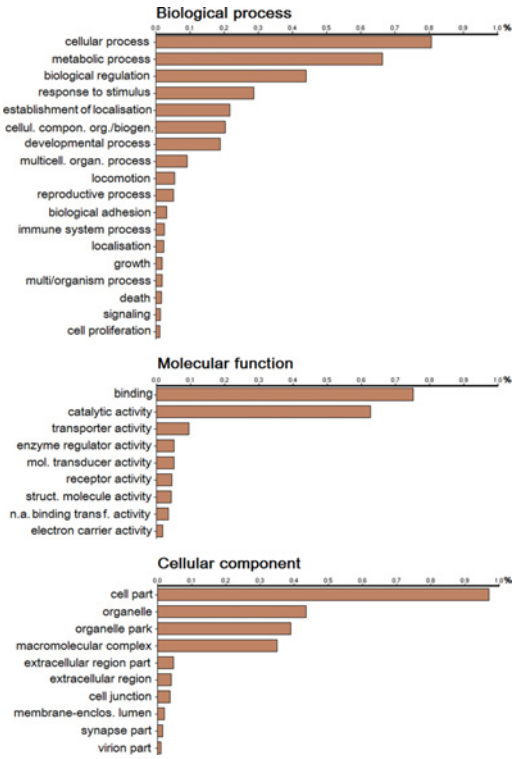


Fig 3. Gene Ontology (GO) analyses of the *A. vulgaris* transcriptome.

doi:10.1371/journal.pone.0150614.g003

(75%), "catalytic activity" (62%) and "transporter activity" (9.8%) (Fig 3). For the cellular component classification, the cluster sizes of "organelle/organelle part", "macromolecular complex" and "extracellular region" were relatively large (Fig 3). Domain analysis against the Pfam database identified 16,994 of entries to have at least one domain of which 471 sequences without blastx hit were further annotated. The zinc-finger double domain zf-H2C2_2 (PF13465) was the most abundant domain followed by domain of unknown function DUF4200 (PF13863) and ankyrin repeat Ank_5 (PF13857) (S1 Table). Ankyrin domains are among the most common structural motifs in known proteins. 1,796 of transcripts were assigned with EC (Enzyme Commission) numbers.

We also analyzed taxonomic distribution of the best-match species of the annotable slug sequences. The highest proportion of the best blastx hits was found for *Aplysia californica* (65%), *Crassostea gigas* (8.7%) and *Capitella teleta* (2.5%) (Fig 4). Despite the limited number of molluscan sequences in the public databases, more than 80% of the sequences best matched to molluscan species. Assignment of transcripts also to chordate, plants, fungi, bacteria and viruses was probably due to more functional data on these species as well as contamination from environmental organisms during sample preparation has to be also considered.

Proteome of slug mantle

With the objective to gain insight into the slug proteome gel-based proteomic analysis from the *A. vulgaris* mantle was performed. In total, 356 spots corresponding to 833 unique proteins were clearly resolved in 2-DE reproducible gels (S2 Table). Fig 5 shows the representative 2-DE gel covering pH 3–10. Most of the spots were located between pH 5 to 9 over a broad range of MW (10–250 kDa). Among these proteins, 814 were assigned to protein sequence with significant hit in NCBI nr database and 19 represent a proteins with no significant homology in nr

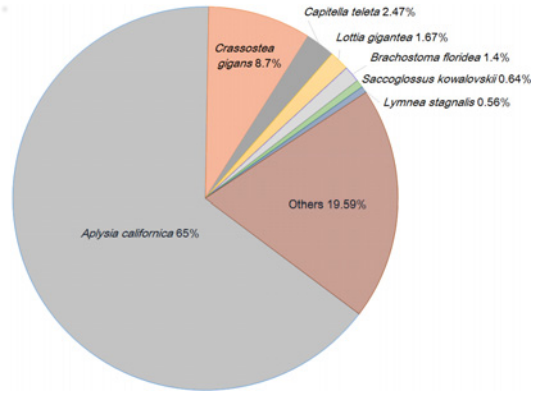


Fig 4. Species distribution of best blastx hits against the nr database.

doi:10.1371/journal.pone.0150614.g004

database. Several putative novel proteins have been identified in several spots representing isoforms with different posttranslational modifications deduced from different position in the gel. Because of the limitation of 2-DE for very small and very large proteins, alkaline proteins and hydrophobic proteins, we also analyzed the protein samples by 1-D SDS-PAGE coupled with LC–MS/MS. 2011 proteins were identified in two biological replicates from 1-D SDS-PAGE (S3 Table). Combining the results from 1-D and 2-D gel separation methods, a total of 2129 proteins were identified with minimum of one unique peptide and two distinct peptides (S4 Table). The overall false-positive assignment was estimated around 1% by a target-decoy search strategy suggesting high quality of our dataset. 1770 identified proteins (83.1%) had matches to Gene Ontology (GO) term annotations and were categorized to molecular function (1647 proteins), cellular component (1337 proteins), and biological process (1596 proteins). Of these, 23.7% have functions that are associated with cytoskeletal protein binding. The mantle, made up of muscle and skin, hence many of the cytoskeletal and muscle-related proteins were observed. The highest number of identified peptides in the combined proteomic approach were observed for paramyosine, myosine, actin and arginine kinase. Arginine kinase belongs to a class of kinases that play a crucial role in invertebrates in the maintenance of ATP levels by the phosphorylation of phosphagens which then serve as a high energy source for rapid ATP replenishment [21]. The GO terms with the largest number of assigned sequences in the biological process category were small molecule metabolic process, catabolic process, biosynthetic process and response to chemical stimulus. 1882 (88.3%) had at least one PFAM domain match and 398 (18.7%) an EC number. The RNA binding motifs RRM_6 and RRM_1 (PF14259 and PF00076) were the most abundant domains followed by Tropomyosin_1 domain (PF12718) and EF hand domains (PF13833, PF13405). (S4 Table).

Combined transcriptomic and proteomic analysis revealed the presence of a high number of sequences sharing carbohydrate-recognition domains, like lectin domains, proteins with C1q domain and toxin-like proteins were present as well (S5 Table).

Discussion

The combined transcriptomic and proteomic approach allowed the identification of a great number of new transcripts and proteins related to metabolic, functional and cellular components. These transcriptomics and proteomics data are by far the most comprehensive among terrestrial gastropods. Moreover, one may speculate that innate immune response-related and defense elements identified in this study may contribute to uncover the success of *A. vulgaris* and indeed, *A.vulgaris* is spreading Europe-wide and is well protected against tentative

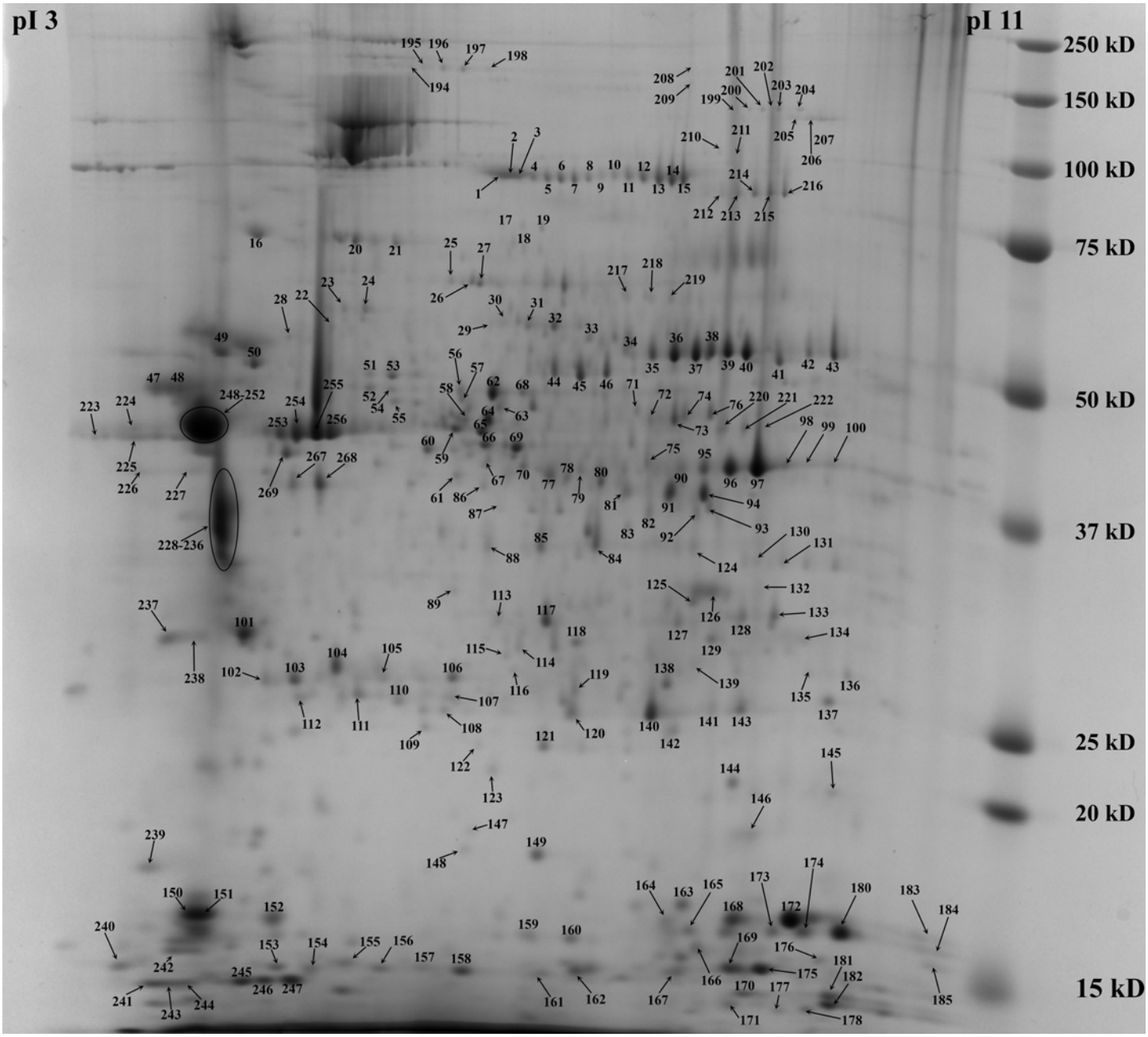


Fig 5. Representative 2-DE gel image from slug mantle showing protein spots selected for mass spectrometry analysis. The identified proteins are listed in S2 Table.

doi:10.1371/journal.pone.0150614.g005

predators and pathogens [22–24]. 2129 *A. vulgaris* mantle proteins were identified through this study, markedly expanding the list of known proteins in terrestrial gastropods, demonstrating the powerfulness of gel-based approach (2-DE and 1-D SDS-PAGE) combined with mass spectrometry identification (LC-ESI-MS/MS). Of the two protein identification approaches used in this study, the 1-D approach has shown markedly higher sensitivity as demonstrated by the significantly higher number of detected proteins. The 2-DE method is useful in separating protein isoforms with different isoelectric point and protein mobility [25].

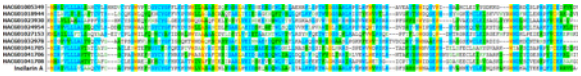


Fig 6. Domain architecture of C-lectin-like molecules found in the transcriptome of *A. vulgaris*.

doi:10.1371/journal.pone.0150614.g006

Slugs belonging to the one of the most successful phyla, Mollusca, lack clear evidence of adaptive immunity. They contend with many pathogens, including bacteria, fungi, viruses and several lineages of specialized eukaryotic parasites, and how they without adaptive immunity effectively and sufficiently defend themselves, is not well understood. The innate immune system in gastropoda is provided by physical barriers (e.g., shell, skin and epithelium), an evolutionary archaic mechanism of molecular "self-non-self" recognition, as well as by a variety of denfense-related factors. The soft, moist slug body protected by a ciliated, mucus-producing epithelium that provides an initial physical barrier to colonization by the pathogens, plays a crucial role in host defense. The diversification of immune systems during evolution involves the expansion of particular gene families in given phyla. Analysis of transcriptomic data from *A. vulgaris* shows a comprehensive repertoire of genes related to innate immunity including many associated pattern recognition, effector proteins and cytokine-like proteins, such as lectins, complement-like proteins, peptidoglycan-recognition proteins (PGRPs), lipopolysaccharide and β1,3-glucan-binding proteins, fibrinogen-related proteins (FREPs), pore-forming membrane attack/perforin (MACPF) domain proteins and toxin-like proteins that indicates an extraordinary complexity of immune system with a high degree of pathogen specificity and immune-priming.

Lectin-like proteins

Lectins play an important role in „self-non-self"recognition and clearance of invaders in gastropods. The slug repertoire of lectin-like proteins bearing conserved carbohydrate-recognition domains (CRDs) is highly diversified, including C-type lectin family, galectin, malectin, H-type lectins, L-type lectins, Ricin-B lectins and calnexin. The diversity of CRDs in lectins may reflect the different functions that the proteins perform. In total, lectin-associated functions may be attributed to 149 transcripts, 37 were unambigously identified at the protein level (S5 Table). These lectin-like proteins show low overlap between species of invertebrates, suggesting relatively rapid evolution of pattern recognition proteins involved in innate immunity. Most of transcripts/proteins described belong to tree lectin families, C-type, R-type and H-type.

C-type lectins

C-type lectins are a superfamily of diverse proteins able to bind specific carbohydrates in a Ca^{2+}-dependent manner. The CRDs contain characteristic double-loop structures stabilized by two highly conserved disulfide bridges located at the bases of the loops [26]. The second loop is structurally and evolutionarily flexible and is involved in Ca^{2+}-dependent carbohydrate binding. In total, 47 transcripts encoding proteins with C-type lectin CRDs have been found and 14 were identified also at the protein level. 17 of 21 complete transcripts have a predicted signal peptide indicating that they are secreted by cells to exert their functions (S5 Table) (Fig 6). Fig 7 shows remarkable sequence variability within incilarin-like proteins identified in our study. Incilarins are C-type lectins originaly isolated from the water-soluble fraction of the body surface mucus of the land slug, *Incilaria fruhstorfer*, possessing hemagglutination activity [27].

R-type lectins

The R-type lectins are members of a superfamily of proteins which contain a carbohydrate-recognition domain (CRD) structurally similar to the CRD in ricin. All identified R-type lectins,

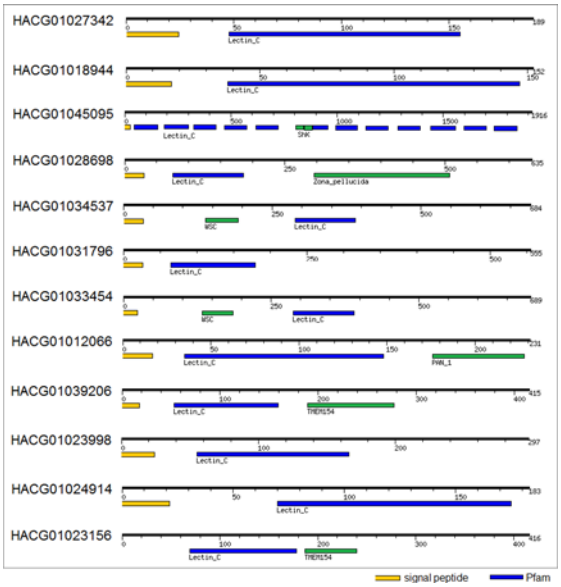


Fig 7. Alignment of amino acid sequences of incilarin-like proteins from *A. vulgaris*.

doi:10.1371/journal.pone.0150614.g007

are chimeric proteins consisting of an A chain with enzymatic activity (galactosyltransferase) linked through a disulfide bridge with a B chain with lectin activity (Fig 8). An analysis of the transcriptome predicts 32 unigenes coding proteins of R-type lectins. R-type lectine-like proteins were not identified at the protein level probably due to tissue-specific expression.

H-type lectins

The H-type lectins are proteins which contain a carbohydrate-recognition domain (CRD) structurally similar to *Helix pomatia* agglutinin (HPA) [28]. It has been shown that HPA is part of the innate immunity system of *H. pomatia* and a component of perivitelline fluid protecting fertilized eggs from bacteria. We identified 12 transcripts, 7 at the protein level. H-type lectins from *A. vulgaris* have a similar size and are predicted to share the same hexameric arrangement achieved by conserved cysteine residues involved in the intermolecular disulfide bridge formation (Figs 8 and 9).

Galectins

Galectins represent a family of structurally-related diverse lectins with carbohydrate binding specificity primarily to β-galactoside residues. Screening of the *A. vulgaris* transcriptome

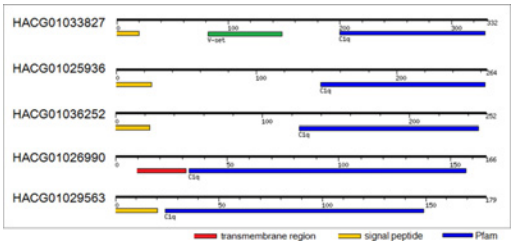


Fig 8. Domain architecture of galectins, H-type lectins, R-type lectins, L-type lectins and calnexin-like proteins found in the transcriptome of *A. vulgaris*.

doi:10.1371/journal.pone.0150614.g008

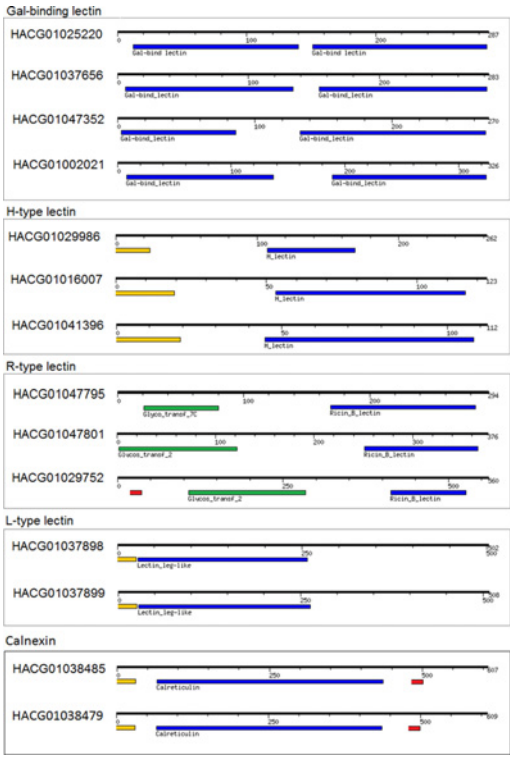


Fig 9. Alignment of amino acid sequences of H-type lectins from *A. vulgaris* with *Helix pomacea* and *Cepaea hortensis*.

doi:10.1371/journal.pone.0150614.g009

revealed that ten unigenes coding proteins share a galectin domain. Galectin domain-containing proteins expressed in *A. vulgaris* are not polymorphic, only a tandem-repeat type galectin has been identified (Fig 8). The absence of a signal sequence is consistent with other galectins, which are known to be actively secreted from diverse tissue via a "nonclassical" secretory pathway [29]. All of these lectins are most similar to gastropod galectins (*Aplysia californica*, *Biomphalaria glabrata*, etc.;blastp E value ranges from 0 to 2E−32).

L-type lectins

A member of L-type (legume-like) lectin family, ERGIC-53 is known as a pattern recognition receptor involved in the immune system of *E. sinensis* [30]. Transcriptomic analysis revealed the presence of two L-type lectin sequences with highest similarity to homolog of ERGIC-53 protein from *Littorina littorea* (blastp: E = 0.0; Identity = 57–58%). ERGIC-53 homologs share the characteristic architecture with an extracellular signal peptide and a single N-terminal L-type carbohydrate binding site (Fig 8).

Calnexin

Members of the evolutionarily conserved calnexin/calreticulin-superfamily bind to oligosaccharides containing terminal glucose residues. Calnexin acts as one of the pattern recognition receptors and has a crucial role in shrimp antibacterial immunity [31]. The search of *A. vulgaris* transcriptomic data revealed four calreticulin-like unigenes coding calnexin proteins with a highest similarity to *Aplysia californica* homologs (blastp: E = 0.0; Identity = 70–74%).

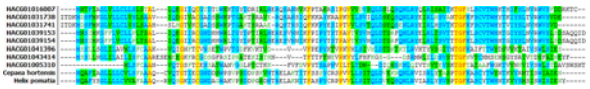


Fig 10. Domain architecture of C1q domain-containing proteins found in the transcriptome of *A. vulgaris*.

doi:10.1371/journal.pone.0150614.g010

Calnexin homologs contain a signal sequence and an additional C-terminal transmembrane helix (Fig 8).

C1q domain-containing proteins

The C1q domain-containing proteins (C1qDC) possessing lectin-like features are a family of proteins characterized by a globular C1q domain [32] regarded as an important player in innate immunity of bivalvia molluscs [33–37]. C1qDC proteins participate in several immune responses, such as pathogen recognition [38], microorganism agglutination [39] and mediating cell migration [40]. In addition, the C1q domain is involved in other immunological processes, such as phagocytosis [41], neutralization of viruses, cell adhesion and clearance of apoptotic cells. It is proposed that C1qDC proteins can activate an ancient complement system by the lectin pathway prior to the evolution of immunoglobulins [41]. The C1q domain has been considered as an extremely efficient pattern recognition domain with highly adaptive binding properties. The extreme versatility of C1q is due to the capability of the C1q domain to bind a variety of self and non-self ligands, including lipopolysaccharides [42], virus envelope proteins, outer membrane proteins from Gram-negative bacteria, phospholipids and some acute-phase proteins.

Based on sequence homology, molecular architecture and domain similarity, 34 C1q domain-containing sequences from the *A. vulgaris* transcriptome may be classified as members of the C1q family (S4 Table). Ten transcripts have been identified also at the protein level. Most of the complete C1qDC sequences displayed a signal peptide or a transmembrane domain on the N-terminus (Fig 10).

Toxin-like proteins

Escaping predation is essential to survival. To reduce predation, organisms have developed a diverse defence mechanisms. A few candidate effector genes were found in the slug transcriptome, including a family of potential pore-forming membrane attack/perforin (MACPF) domain proteins. Pore-forming membrane attack/perforin (MACPF) domain proteins have been described to be involved in the biochemical defence of apple snail eggs against predators [43].

Transcriptomic analysis of *A. vulgaris* revealed a novel toxin-like protein showing significant similarity to agatoxin, expressed in spider glands with a characteristic cysteine motive in the mature peptide. The agatoxin-homolog from *A. vulgaris* shares characteristic architecture with signal peptide, prepeptide sequence and characteristic cysteine motive in the mature peptide (Fig 11).

Fig 11. Alignment of amino acid sequences of toxin-like proteins from *A. vulgaris* with U8-agatoxine–Ao1a from *Agelena orientalis*.

doi:10.1371/journal.pone.0150614.g011

Taken together, one may speculate that the diversity and multitude of proteins with denfensive potential along with already reported defensins (diterpene) of this slug [44] may contribute to success and survival of this terrestrial mollusc.

Data Accessibility

The sequence data has been submitted to the sequence read archive (SRA) database of GenBank (http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession number PRJEB7891.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium [45] via the PRIDE partner repository with the dataset identifier PXD002078 and 10.6019/PXD002078.

Supporting Information

S1 Table.
(XLSX)

S2 Table. List of 814 proteins identified from the mantel of *A. vulgaris* using 2-D electrophoretic separation.
(XLSX)

S3 Table. List of 2011 proteins identified from the mantel of *A. vulgaris* using 1-D electrophoretic separation.
(XLSX)

S4 Table. List of 2129 proteins identified from the mantel of *A. vulgaris*.
(XLSX)

S5 Table. List of lectin-like sequences and toxin-like sequences identified in *A. vulgaris*.
(XLSX)

Acknowledgments

The supervision of F.S. and kind assistance of Dr. Keiryn Bennett, CeMM, Vienna, Austria, is highly appreciated

Author Contributions

Conceived and designed the experiments: JA GL. Performed the experiments: TB FJS GJ. Analyzed the data: RS TR JA. Contributed reagents/materials/analysis tools: MB. Wrote the paper: JA GL. Species identification: HS.

References

1. Gismervik K, Aspholm M, Rorvik LM, Bruheim T, Andersen A, Skaar I (2015) Invading slugs (*Arion vulgaris*) can be vectors for *Listeria monocytogenes*. J Appl Microbiol 118: 809–816. doi: 10.1111/jam.12750 PMID: 25580873
2. Gismervik K, Bruheim T, Rorvik LM, Haukeland S, Skaar I (2014) Invasive slug populations (*Arion vulgaris*) as potential vectors for *Clostridium botulinum*. Acta Vet Scand 56: 65. doi: 10.1186/s13028-014-0065-z PMID: 25277214
3. Stalder GL, Loncaric I, Walzer C (2014) Diversity of enterobacteria including beta-lactamase producing isolates associated with the Spanish slug (*Arion vulgaris*). Sci Total Environ 479–480: 11–16. doi: 10.1016/j.scitotenv.2014.01.103 PMID: 24534696
4. Patel Z, Gill AC, Fox MT, Hermosilla C, Backeljau T, Breugelmans K, et al. (2014) Molecular identification of novel intermediate host species of *Angiostrongylus vasorum* in Greater London. Parasitol Res 113: 4363–4369. doi: 10.1007/s00436-014-4111-6 PMID: 25195057

5. Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M (2011) Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (*Radix balthica*, *Basommatophora*, *Pulmonata*), and a comparison of assembler performance. BMC Genomics 12: 317. doi: 10.1186/1471-2164-12-317 PMID: 21679424
6. Amin S, Prentis PJ, Gilding EK, Pavasovic A (2014) Assembly and annotation of a non-model gastropod (*Nerita melanotragus*) transcriptome: a comparison of de novo assemblers. BMC Res Notes 7: 488. doi: 10.1186/1756-0500-7-488 PMID: 25084827
7. Sun J, Wang M, Wang H, Zhang H, Zhang X, Thiyagarajan V, et al. (2012) De novo assembly of the transcriptome of an invasive snail and its multiple ecological applications. Mol Ecol Resour 12: 1133–1144. doi: 10.1111/1755-0998.12014 PMID: 22994926
8. Kopylova E, Noe L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28: 3211–3217. doi: 10.1093/bioinformatics/bts611 PMID: 23071270
9. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29: 644–652. doi: 10.1038/nbt.1883 PMID: 21572440
10. Chen TW, Gan RC, Wu TH, Huang PJ, Lee CY, Chen YY, et al. (2012) FastAnnotator—an efficient transcript annotation web tool. BMC Genomics 13 Suppl 7: S9. doi: 10.1186/1471-2164-13-S7-S9 PMID: 23281853
11. Bae N, Lodl M, Pollak A, Lubec G (2012) Mass spectrometrical analysis of bilin-binding protein from the wing of *Hebomoia glaucippe* (Linnaeus, 1758) (Lepidoptera: Pieridae). Electrophoresis 33: 1787–1794. doi: 10.1002/elps.201100569 PMID: 22740467
12. Ahmed KE, Chen WQ, John JP, Kang SU, Lubec G (2010) Complete sequencing of the recombinant granulocyte-colony stimulating factor (filgrastim) and detection of biotinylation by mass spectrometry. Amino Acids 38: 1043–1049. doi: 10.1007/s00726-009-0312-1 PMID: 19526310
13. Chen WQ, Li L, Lubec G (2014) Proteomic differences between white and brown adipocytes. Amino Acids 46: 995–1008. doi: 10.1007/s00726-013-1657-z PMID: 24390458
14. Kang SU, Heo S, Lubec G (2011) Mass spectrometric analysis of GABAA receptor subtypes and phosphorylations from mouse hippocampus. Proteomics 11: 2171–2181. doi: 10.1002/pmic.201000374 PMID: 21538884
15. Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, et al. (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis 25: 1327–1333. PMID: 15174055
16. Bennett KL, Funk M, Tschernutter M, Breitwieser FP, Planyavsky M, Ubaida Mohien C, et al. (2011) Proteomic analysis of human cataract aqueous humour: Comparison of one-dimensional gel LCMS with two-dimensional LCMS of unlabelled and iTRAQ(R)-labelled specimens. J Proteomics 74: 151–166. doi: 10.1016/j.jprot.2010.10.002 PMID: 20940065
17. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4: 207–214. PMID: 17327847
18. Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. Genome Res 20: 1313–1326. doi: 10.1101/gr.101386.109 PMID: 20651121
19. Cech TR, Steitz JA (2014) The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157: 77–94. doi: 10.1016/j.cell.2014.03.008 PMID: 24679528
20. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676. PMID: 16081474
21. Voncken F, Gao F, Wadforth C, Harley M, Colasante C (2013) The phosphoarginine energy-buffering system of trypanosoma brucei involves multiple arginine kinase isoforms with different subcellular locations. PLoS One 8: e65908. doi: 10.1371/journal.pone.0065908 PMID: 23776565
22. Pfenninger M, Weigand A, Balint M, Klussmann-Kolb A (2014) Misperceived invasion: the Lusitanian slug (*Arion lusitanicus* auct. non-Mabille or *Arion vulgaris* Moquin-Tandon 1855) is native to Central Europe. Evol Appl 7: 702–713. doi: 10.1111/eva.12177 PMID: 25067951
23. Hatteland BA, Roth S, Andersen A, Kaasa K, Stoa B, T S (2013) Distribution and spread of the invasive slug *Arion vulgaris* MoquinTandon in Norway. Fauna norvegica 32: 13–26.
24. Kozlowski J (2007) The distribution, biology, population dynamics and harmfulness of *Arion lusitanicus* Mabillee, 1868 (Gastropoda: *Pulmonata*: *Arionidae*) in Poland. Journal of Plant Protection Research 47: 219–230.
25. Sun J, Zhang H, Wang H, Heras H, Dreon MS, Ituarte S, et al. (2012) First proteome of the egg perivitelline fluid of a freshwater gastropod with aerial oviposition. J Proteome Res 11: 4240–4248. doi: 10.1021/pr3003613 PMID: 22738194

26. Zelensky AN, Gready JE (2005) The C-type lectin-like domain superfamily. FEBS J 272: 6179–6217. PMID: 16336259

27. Yuasa HJ, Furuta E, Nakamura A, Takagi T (1998) Cloning and sequencing of three C-type lectins from body surface mucus of the land slug, *Incilaria fruhstorferi*. Comp Biochem Physiol B Biochem Mol Biol 119: 479–484. PMID: 9734332

28. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, et al. (2006) Biochemical and structural analysis of *Helix pomatia* agglutinin. A hexameric lectin with a novel fold. J Biol Chem 281: 20171–20180. PMID: 16704980

29. Nickel W (2003) The mystery of nonclassical protein secretion. A current view on cargo proteins and potential export routes. Eur J Biochem 270: 2109–2119. PMID: 12752430

30. Huang Y, Tan JM, Wang Z, Yin SW, Huang X, Wang W, et al. (2014) Cloning and characterization of two different L-type lectin genes from the Chinese mitten crab *Eriocheir sinensis*. Dev Comp Immunol 46: 255–266. doi: 10.1016/j.dci.2014.04.015 PMID: 24796868

31. Zhang Q, Wang XQ, Jiang HS, Jia WM, Zhao XF, Wang JX (2014) Calnexin functions in antibacterial immunity of *Marsupenaeus japonicus*. Dev Comp Immunol 46: 356–363. doi: 10.1016/j.dci.2014.05.009 PMID: 24858031

32. Carland TM, Gerwick L (2010) The C1q domain containing proteins: Where do they come from and what do they do? Dev Comp Immunol 34: 785–790. doi: 10.1016/j.dci.2010.02.014 PMID: 20214925

33. Gerlach D, Schlott B, Schmidt KH (2004) Cloning and expression of a sialic acid-binding lectin from the snail *Cepaea hortensis*. FEMS Immunol Med Microbiol 40: 215–221. PMID: 15039097

34. Allam B, Pales Espinosa E, Tanguy A, Jeffroy F, Le Bris C, Paillard C (2014) Transcriptional changes in Manila clam (*Ruditapes philippinarum*) in response to Brown Ring Disease. Fish Shellfish Immunol 41: 2–11. doi: 10.1016/j.fsi.2014.05.022 PMID: 24882017

35. Gestal C, Pallavicini A, Venier P, Novoa B, Figueras A (2010) MgC1q, a novel C1q-domain-containing protein involved in the immune response of *Mytilus galloprovincialis*. Dev Comp Immunol 34: 926–934. doi: 10.1016/j.dci.2010.02.012 PMID: 20219531

36. He X, Zhang Y, Yu F, Yu Z (2011) A novel sialic acid binding lectin with anti-bacterial activity from the Hong Kong oyster (*Crassostrea hongkongensis*). Fish Shellfish Immunol 31: 1247–1250. doi: 10.1016/j.fsi.2011.08.021 PMID: 21906682

37. Xu T, Xie J, Li J, Luo M, Ye S, Wu X (2012) Identification of expressed genes in cDNA library of hemocytes from the RLO-challenged oyster, *Crassostrea ariakensis* Gould with special functional implication of three complement-related fragments (CaC1q1, CaC1q2 and CaC3). Fish Shellfish Immunol 32: 1106–1116. doi: 10.1016/j.fsi.2012.03.012 PMID: 22484279

38. Yu Y, Huang H, Wang Y, Yu Y, Yuan S, Huang S, et al. (2008) A novel C1q family member of amphioxus was revealed to have a partial function of vertebrate C1q molecule. J Immunol 181: 7024–7032. PMID: 18981122

39. Kong P, Zhang H, Wang L, Zhou Z, Yang J, Zhang Y, et al. (2010) AiC1qDC-1, a novel gC1q-domain-containing protein from bay scallop *Argopecten irradians* with fungi agglutinating activity. Dev Comp Immunol 34: 837–846. doi: 10.1016/j.dci.2010.03.006 PMID: 20346969

40. Tahtouh M, Croq F, Vizioli J, Sautiere PE, Van Camp C, Salzet M, et al. (2009) Evidence for a novel chemotactic C1q domain-containing factor in the leech nerve cord. Mol Immunol 46: 523–531. doi: 10.1016/j.molimm.2008.07.026 PMID: 18952286

41. Wang L, Wang L, Zhang H, Zhou Z, Siva VS, Song L (2012) A C1q domain containing protein from scallop *Chlamys farreri* serving as pattern recognition receptor with heat-aggregated IgG binding activity. PLoS One 7: e43289. doi: 10.1371/journal.pone.0043289 PMID: 22905248

42. Zhang H, Song L, Li C, Zhao J, Wang H, Qiu L, et al. (2008) A novel C1q-domain-containing protein from Zhikong scallop *Chlamys farreri* with lipopolysaccharide binding activity. Fish Shellfish Immunol 25: 281–289. doi: 10.1016/j.fsi.2008.06.003 PMID: 18603000

43. Dreon MS, Frassa MV, Ceolin M, Ituarte S, Qiu JW, Sun J, et al. (2013) Novel animal defenses against predation: a snail egg neurotoxin combining lectin and pore-forming chains that resembles plant defense and bacteria attack toxins. PLoS One 8: e63782. doi: 10.1371/journal.pone.0063782 PMID: 23737950

44. Schroeder FC, Gonzalez A, Eisner T, Meinwald J (1999) Miriamin, a defensive diterpene from the eggs of a land slug (*Arion sp.*). Proc Natl Acad Sci U S A 96: 13620–13625. PMID: 10570122

45. Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D, et al. (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat Biotechnol 32: 223–226. doi: 10.1038/nbt.2839 PMID: 24727771

DOI 10.1002/pmic.201500092 *Proteomics* 2015, *15*, 3356–3360

DATASET BRIEF

Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads

Jana Aradska[1], Tanja Bulat[1], Fernando J. Sialana[1], Ruth Birner-Gruenberger[2], Buchner Erich[3] and Gert Lubec[1]

[1] Department of Pediatrics, Medical University of Vienna, Vienna, Austria
[2] Institute of Pathology, Medical University of Graz, Core Facility Mass Spectrometry, Center for Medical Research, BioTechMed Omics Center Graz, Graz, Austria
[3] Institute for Clinical Neurobiology, University of Würzburg, Würzburg, Germany

Received: March 5, 2015
Revised: May 15, 2015
Accepted: July 20, 2015

Membrane proteins play key roles in several fundamental biological processes such as cell signalling, energy metabolism and transport. Despite the significance, these still remain an underrepresented group in proteomics datasets. Herein, a bottom-up approach to analyse an enriched membrane fraction from *Drosophila melanogaster* heads using multidimensional liquid chromatography (LC) coupled with tandem-mass spectrometry (MS/MS) that relies on complete solubilisation and digestion of proteins, is reported. An enriched membrane fraction was prepared using equilibrium density centrifugation on a discontinuous sucrose gradient, followed by solubilisation using the filter-aided sample preparation (FASP), tryptic and sequential chymotryptic digestion of proteins. Peptides were separated by reversed-phase (RP) LC at high pH in the first dimension and acidic RP-LC in the second dimension coupled directly to an Orbitrap Velos Pro mass spectrometer. A total number of 4812 proteins from 114 865 redundant and 38 179 distinct peptides corresponding to 4559 genes were identified in the enriched membrane fraction from fly heads. These included brain receptors, transporters and channels that are most important elements as drug targets or are linked to disease. Data are available via ProteomeXchange with identifier PXD001712 (http://proteomecentral.proteomexchange.org/dataset/PXD001712).

Keywords:
Animal proteomics / *Drosophila melanogaster* / Mass spectrometry / Membrane proteins

Additional supporting information may be found in the online version of this article at the publisher's web-site

The fruit fly *Drosophila melanogaster* is used as prime model organism for experimental studies of eukaryotic organisms, possessing genetic, anatomic, behavioural, methodological and even economic advantages [1]. A complete genome of *D. melanogaster* has been determined [2] and protein-coding regions are well-annotated. Surprisingly, the human and the drosophila genome are apparently interrelated. The majority of human genes have counterparts in the fly genome and almost 75% of all known human disorder-related genes have fly homologues [3]. This, together with amenability in the genetic manipulations has made *Drosophila* a powerful, simple model to understand human biology, and molecular and cellular mechanisms of human diseases. In the last decade, a few large-scale proteomics analyses were performed to validate gene models and uncover protein expression profiles in different parts of the fly body, cell types, developmental states and cellular fractions [4–7]. So far the most extensive study has covered 63% of the fly proteome by detecting 9124 proteins. This high coverage was achieved by combining sample diversity and applying multiple fractionation strategies [4]. Lee et al. [8] have reported a proteomics dataset from fly heads based on two-dimensional electrophoresis separations. They have shown a high overlap of proteins found in heads and brains, nearly 90% and therefore heads can be considered for analyses of proteins from the central nervous system. Due to

Correspondence: Professor Gert Lubec, Department of Pediatrics, Medical University of Vienna, Waehringer Guertel 18, 1090 Vienna, Austria
E-mail: gert.lubec@meduniwien.ac.at
Fax: +43-1-40400-6065

Colour Online: See the article online to view Fig. 1–3 in colour.

limitations of their two-dimensional gel-based approach only very few membrane proteins were identified in the above-mentioned dataset.

A comprehensive protein coverage of *Drosophila* proteome has not been achieved; membrane proteins still remain under-represented in proteomics studies and there are still gene models relying on computational predictions only. Herein, analysis of an enriched membrane fraction from fly heads through a gel-free approach is reported. The used gel-free bottom-up proteomic approach employs an initial proteolytic digestion of the complex mixture with fractionation at the peptide level using multidimensional liquid chromatography prior to nano-LC separation reducing the complexity of the sample.

D. melanogaster heads were isolated from adult flies using a standard freezing protocol [9]. Heads were then ground into a fine powder with a pre-chilled mortar and pestle. Powder was re-suspended in ice-cold homogenization buffer (10 mM HEPES, pH 7.5, 300 mM sucrose, protease inhibitor (Roche Molecular Biochemicals, Mannheim, Germany)). The homogenate was centrifuged for 10 min at 1000 × *g* and the supernatant was centrifuged at 50 000 × *g* for 30 min at 4°C. Subsequently, the pellet was re-suspended in 4 mL washing buffer (homogenization buffer without sucrose), kept on ice for 30 min and centrifuged at 50 000 × *g* for 30 min. A total plasma membrane fraction was separated by equilibrium density centrifugation on a discontinuous sucrose gradient. The pellet, re-suspended in washing buffer, was layered on top of the sucrose cushion (1 M and 1.25 M sucrose solution) followed by centrifugation at 70 000 × *g* for 2 h at 4°C. After centrifugation, fractions from the sucrose interface were collected and diluted ten times with washing buffer and subsequently centrifuged for 30 min at 100 000 × *g*. The pellet was stored at 80°C until use.

Membrane protein extraction was carried out according to a previous study with minor modifications [10]. Protein quantities were estimated by the Pierce 660 protein assay (ThermoFisher Scientific, IL, USA). 300 μg of the sample (1 × 100 μg each) were digested with trypsin using the filter-aided sample preparation (FASP) as previously described with minor modifications [11, 12]. Samples were mixed in the filter unit with 8 M urea in 100 mM Tris-HCl (pH 8.5) (UA) and centrifuged at 14 000 × *g* for 15 min at 22°C. Any remaining detergent was exchanged by urea with 200 μL UA. The proteins were alkylated by the addition of 100 μL of 50 mM iodoacetamide in UA and incubated for 30 min at 22°C. Subsequently, excess reagent was removed by 3 × 100 μL UA solution followed by 3 × 100 μL of 50 mM TEAB buffer (Sigma-Aldrich, Hamburg, Germany). Proteins were digested overnight at 37°C with trypsin at an enzyme to protein ratio of 1:100 w/w. Tryptic peptides were recovered from the filter with 40 μL of 50 mM TEAB buffer followed by 50 μL of 0.5 M NaCl. The remaining semi-digested material on the FASP filter was washed with 50 μL 50 mM TEAB buffer and further digested with chymotrypsin (enzyme to protein ratio of 1:10 w/w). Pooled tryptically-digested and sequential trypsin–chymotrypsin digested samples were acidified, desalted and concentrated with customised reversed-phase C18 stage tips [13]. Bound peptides were eluted with a buffer containing 90% acetonitrile and 100 mM TEAB and lyophilised in a vacuum concentrator (Eppendorf, Hamburg, Germany). Samples were pooled, peptides were separated by basic reversed-phase liquid chromatography [14] and 50 fractions were collected for trypsin and 20 fractions for the sequential trypsin–chymotrypsin.

Acidified fractions were analysed by LC-MS/MS. Details of the procedure were as described previously with minor modifications [15]. Peptide fractions were separated by nano-LC (Dionex Ultimate 3000) equipped with a μ-precolumn (C18, 5 m, 100 Å, 5 × 0.3 mm) and an Acclaim PepMap RSLC nanocolumn (C18, 2 μm, 100 Å, 150 × 0.075 mm) (Thermo Fisher Scientific, Vienna, Austria). Eight microliters samples were injected and concentrated on the enrichment column for 2 min using 0.05% TFA as isocratic solvent at a flow rate of 20 μL/min. The column was then switched in the nanoflow circuit, and the sample was loaded on the nanocolumn at a flow rate of 200 nL/min and separated using the following gradient: solvent A: water, 0.05% TFA; solvent B: ACN/water 80/20 v/v, 0.05% TFA; 0–4 min: 4 % B; 4– 60 min 4–40% B, 60–95 min: 40–95% B, 95–105 min: 95% B, 105–105.1 min: 95-4% B, 105.1–120 min: 4% B. The sample was ionized in the nanospray source equipped with stainless steel emitters (Thermo Fisher Scientific, Vienna, Austria) and analysed in a Thermo Orbitrap Velos Pro mass spectrometer in positive ion mode by alternating full-scan MS (*m/z* 380–2000) in the ICR cell and MS/MS by CID of the 20 most intense peaks in the ion trap with dynamic exclusion enabled.

LC-MS/MS data were analysed by searching the UniProtKB *Drosophila melanogaster* database (41 965 sequences, download on January 16th, 2015) with Proteome Discoverer 1.4 (Thermo Fisher Scientific, Vienna, Austria) and MASCOT 2.4 (MatrixScience, London, UK). Detailed search criteria were used as follows; enzyme: trypsin or trypchymo with a maximum of two missing cleavage sites; fixed modification: carbamidomethylation (C); variable modification: oxidation (M); search mode: MS/MS ion search with decoy database search included; peptide mass tolerance ± 10 ppm; MS/MS mass tolerance ± 0.5 Da. Using the target-decoy search strategy [16], peptide matches with a false discovery rate (FDR) lower than 1% were filtered. All proteins with a minimum of two distinct peptides and ion score > 20 were accepted.

A total of 114 865 redundant and 38 179 distinct peptides were identified and assigned to 4812 proteins, including splicing variants. Of these, 3920 proteins (81, 4%) were identified with more than two distinct peptides (Fig. 1A). On average eight distinct unique peptides were identified per protein. The list of the identified proteins in order of exponentially modified protein abundance index (emPAI) [17, 18] is given in Supporting Information Tables S1 and S2. Identified proteins correspond to 4559 genes which represent 33, 5% of the predicted genes in *Drosophila melanogaster*. The MS proteomics data have been deposited in the ProteomeXchange

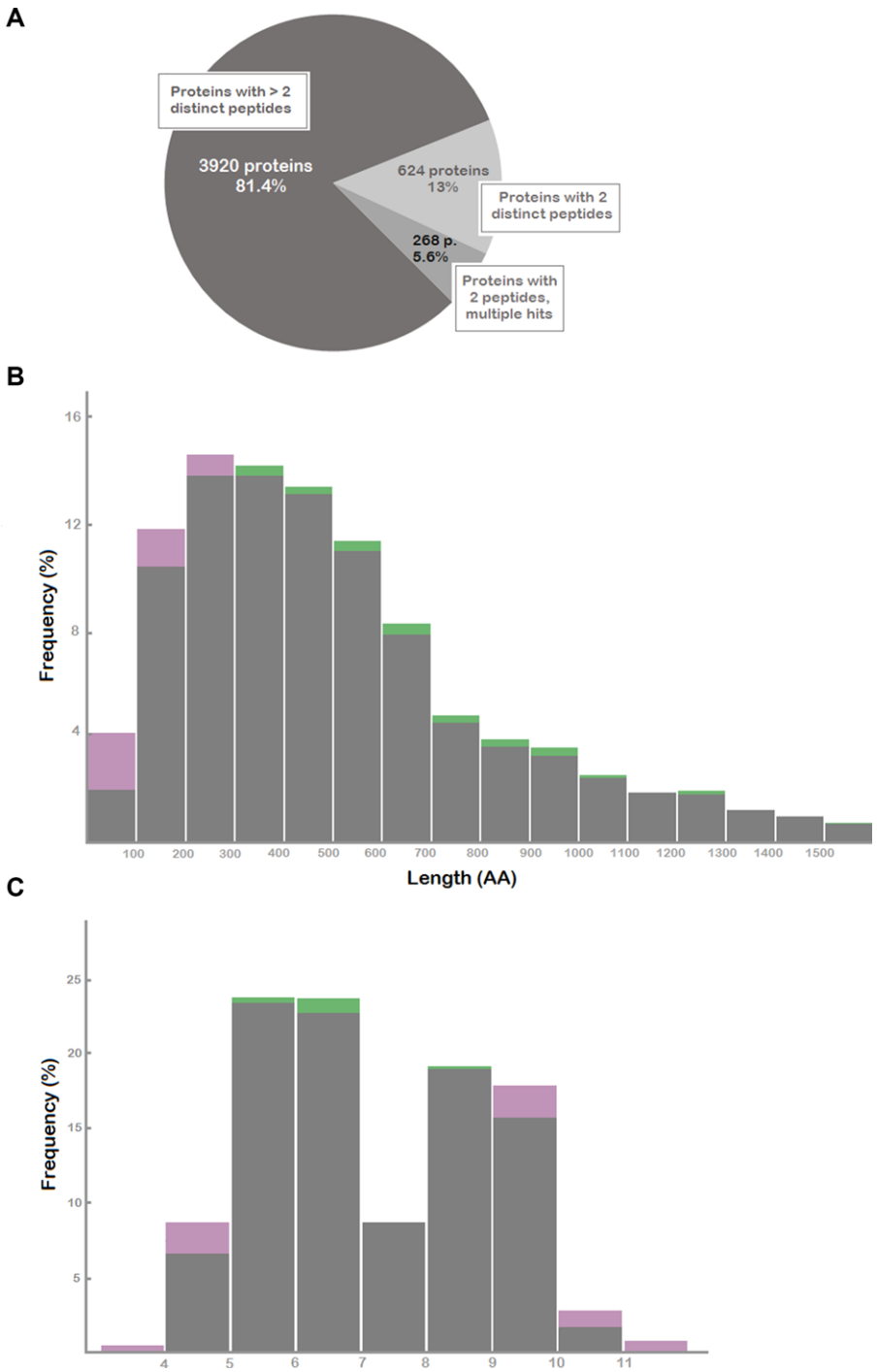


Figure 1. Experimentally identified proteins with respect to the number of distinct peptides (A). Distributions of length (B) and pI (C) of experimentally identified proteins (4812 proteins) in comparison to all proteins (20 328 proteins) in *Drosophila melanogaster* including splice variants. Under-represented areas are shown in pink, over-represented areas in green.

Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [19] with the dataset identifier PXD001712 and 10.6019/PXD001712.The dataset was classified with respect to several aspects (length, isoelectric point, transmembrane domains and functional annotation). Length and pI distribution of identified proteins are shown in Fig. 1B and C. The TMHMM (www.cbs.dtu.dk/services/TMHMM/) algorithm was used to predict transmembrane regions from the identified proteins [20]. 35% of proteins (1679 proteins) were identified as membrane proteins containing at least one transmembrane domain and 65% (3133 proteins) had no confirmed or predicted transmembrane domain (Fig. 2).

The identified proteins were categorized to GO terms using the AmiGO2 bioinformatics Tool [21]. 3565 gene models were considered for functional analysis. An enrichment

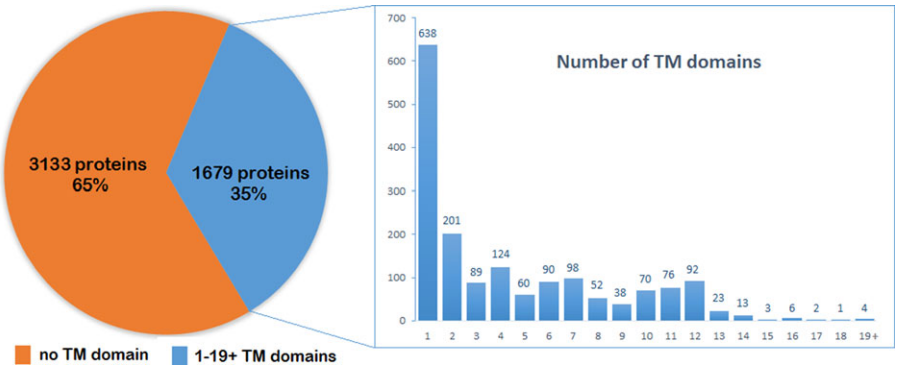


Figure 2. Transmembrane domain distribution among identified proteins.

analysis has been performed against *D. melanogaster* background (13 624 gene models) and GO terms with higher than two-fold enrichment for individual categories are shown in Fig. 3. The list of GO terms, number of associated proteins, fold enrichment and *p*-values for each GO term is given in Supporting Information Table S6. We identified several ionotropic receptors for glutamate and acetylcholine, and metabotropic receptors for acetylcholine, GABA, glutamate, dopamine, octopamine, serotonin and adenosine (Supporting Information Table S3). Based upon the emPAI index we compared the relative abundance of identified receptors. Ionotropic receptors NmdaR1 and NmdaR2, nicotinic acetylcholine receptor subunits α4 and β3; and metabotropic dopamine 1 receptor were the most abundant receptors. The major amino-acid transporters associated with neurotransmitters glutamate, glycine and dopamine were unambiguously identified (Supporting Information Table S4). The most common ion channel families: voltage-gated calcium channels, voltage-gated chloride channels, potassium channels, and voltage-dependent sodium channels were detected and listed in Supporting Information Table S5.

Taken together, a gel-free mass spectrometrical approach using the FASP principle was used to provide a dataset that may be useful for studies on membrane proteins that are

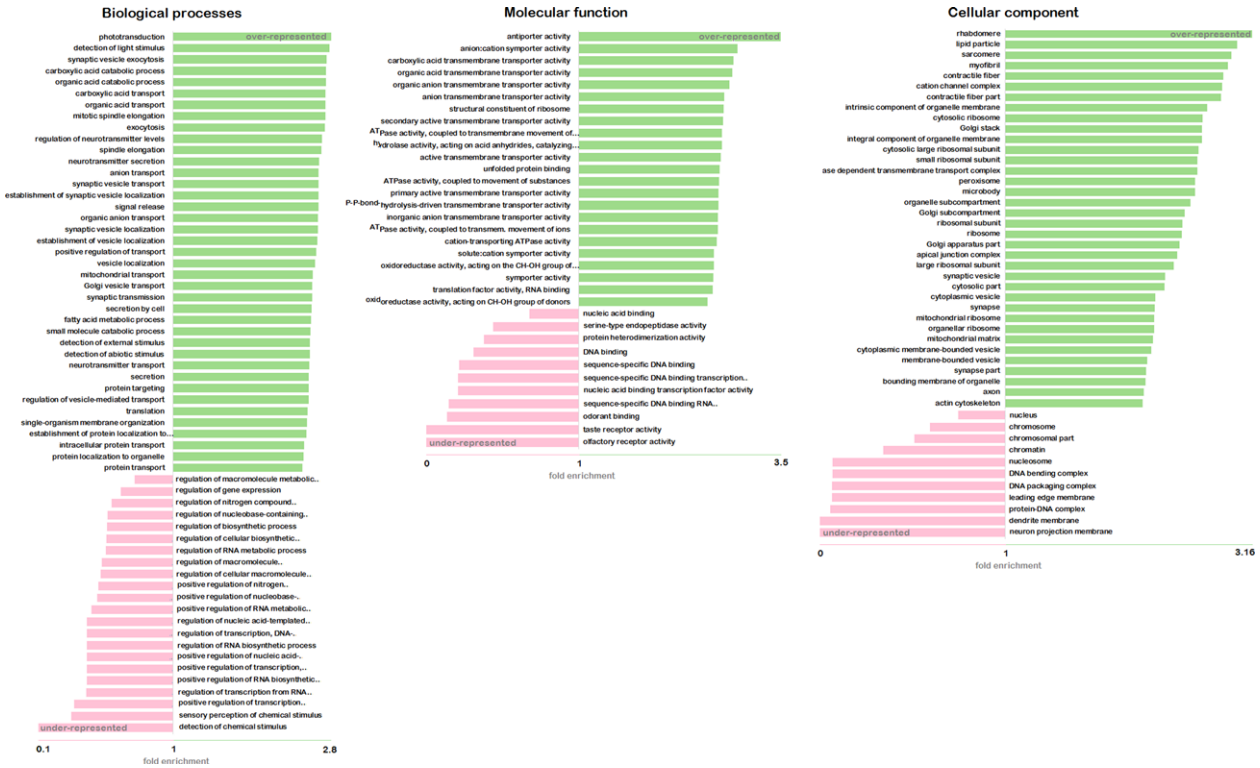


Figure 3. GO term enrichment analysis of experimentally identified proteins. 3565 gene models, considered for functional analysis were compared to 13 624 predicted gene models. The figure shows GO terms that are significantly enriched ($p < 1E-4$) by more than two-fold or are significantly under-represented for all tree functional categories.

mainly representing drug targets and are linked to a large series of human disorders and diseases.

The MS proteomics data in this paper have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository [19]: dataset identifier PXD001712.

The authors have declared no conflict of interest.

References

[1] Prussing, E., Historical trauma: politics of a conceptual framework. *Transcult. Psychiatry* 2014, *51*, 436–458.

[2] Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A. et al., The genome sequence of *Drosophila melanogaster*. *Science* 2000, *287*, 2185–2195.

[3] Bier, E., Drosophila, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genetics* 2005, *6*, 9–23.

[4] Brunner, E., Ahrens, C. H., Mohanty, S., Baetschmann, H. et al., A high-quality catalog of the *Drosophila melanogaster* proteome. *Nat. Biotechnol.* 2007, *25*, 576–583.

[5] Xing, X., Zhang, C., Li, N., Zhai, L. et al., Qualitative and quantitative analysis of the adult *Drosophila melanogaster* proteome. *Proteomics* 2014, *14*, 286–290.

[6] Wasbrough, E. R., Dorus, S., Hester, S., Howard-Murkin, J. et al., The *Drosophila melanogaster* sperm proteome-II (DmSP-II). *J. Proteomics* 2010, *73*, 2171–2185.

[7] Taraszka, J. A., Kurulugama, R., Sowell, R. A., Valentine, S. J. et al., Mapping the proteome of *Drosophila melanogaster*: analysis of embryos and adult heads by LC-IMS-MS methods. *J. Proteome Res.* 2005, *4*, 1223–1237.

[8] Lee, T. R., Huang, S. H., Lee, C. C., Lee, H. Y. et al., Proteome reference map of *Drosophila melanogaster* head. *Proteomics* 2012, *12*, 1875–1878.

[9] Jensen, K., Sanchez-Garcia, J., Williams, C., Khare, S. et al., Purification of transcripts and metabolites from Drosophila heads. *J. Visualized Exp.* 2013, e50245.

[10] Luche, S., Santoni, V., Rabilloud, T., Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics* 2003, *3*, 249–253.

[11] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nat. Methods* 2009, *6*, 359–362.

[12] Manza, L. L., Stamer, S. L., Ham, A. J., Codreanu, S. G., Liebler, D. C., Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* 2005, *5*, 1742–1745.

[13] Rappsilber, J., Mann, M., Ishihama, Y., Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat.Protoc.* 2007, *2*, 1896–1906.

[14] Gilar, M., Olivova, P., Daly, A. E., Gebler, J. C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* 2005, *28*, 1694–1703.

[15] Heo, S., Spoerk, S., Birner-Gruenberger, R., Lubec, G., Gel-based mass spectrometric analysis of hippocampal transmembrane proteins using high resolution LTQ Orbitrap Velos Pro. *Proteomics* 2014, *14*, 2084–2088.

[16] Elias, J. E., Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 2007, *4*, 207–214.

[17] Ishihama, Y., Oda, Y., Tabata, T., Sato, T. et al., Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* 2005, *4*, 1265–1272.

[18] Shinoda, K., Tomita, M., Ishihama, Y., emPAI Calc–for the estimation of protein abundance from large-scale identification data by liquid chromatography-tandem mass spectrometry. *Bioinformatics* 2010, *26*, 576–577.

[19] Vizcaino, J. A., Deutsch, E. W., Wang, R., Csordas, A. et al., ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* 2014, *32*, 223–226.

[20] Sonnhammer, E. L., vonHeijne, G., Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1998, *6*, 175–182.

[21] Carbon, S., Ireland, A., Mungall, C. J., Shu, S. et al., AmiGO: online access to ontology and annotation data. *Bioinformatics* 2009, *25*, 288–289.

8.2. Table 1. List of proteins identified in 2–DE proteomics study (UP-unique peptide number; PC-protein coverage (%); NM-number of assigned peptide matches; NS-number of protein sequences; PS-protein score)

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|318609972|dbj|BAJ61596.1|	96	71	97	2700	44575
gi|524888508|ref|XP_005100824.1|	60	38	78	595	9831
gi|524866478|ref|XP_005090060.1|	57	86	57	667	11570
gi|305689815|gb|ADM64337.1|	48	73	48	965	19654
gi|405975785|gb|EKC40331.1|	45	64	48	342	6259
gi|220067263|gb|ACL79834.1|	44	76	44	566	12445
gi|556096138|gb|ESO84790.1|	43	65	46	676	15337
gi|385145402|emb|CCG28026.1|	42	86	45	2211	30113
gi|127773|sp|P24733.1|MYS_ARGIR	41	39	70	660	11244
gi|147905746|ref|NP_001080566.1|	39	73	47	395	4497
gi|405960426|gb|EKC26351.1|	39	67	39	228	4756
gi|524903383|ref|XP_005107892.1|	39	58	39	256	6023
gi|524916025|ref|XP_005112795.1|	38	69	40	736	16769
gi|73695954|gb|AAZ80783.1|	38	76	38	2339	16103
gi|126697420|gb|ABO26667.1|	37	62	37	165	2543
gi|524888358|ref|XP_005100750.1|	36	62	36	154	3335
gi|524894502|ref|XP_005103748.1|	36	54	39	307	6303
gi|524896814|ref|XP_005104878.1|	35	17	35	128	1273
gi|71726729|gb|AAZ39528.1|	33	61	45	481	7397
gi|524909114|ref|XP_005109609.1|	32	90	33	791	10634
gi|42559558|sp|O97192.1|TPM_HELAS	30	85	40	1540	26058
gi|472824675|dbj|BAN09030.1|	30	55	30	879	14670
gi|556096092|gb|ESO84744.1|	30	44	30	217	4565
gi|556110866|gb|ESO99517.1|	30	27	64	289	5337
gi|829207|emb|CAA39415.1|	30	61	30	335	6356
gi|296837083|gb|ADH59421.1|	28	67	42	1397	26938
gi|443692429|gb|ELT94023.1|	28	71	28	313	6839
gi|524885601|ref|XP_005099406.1|	28	39	28	180	2923
gi|524884921|ref|XP_005099073.1|	27	45	27	576	11029
gi|524895745|ref|XP_005104354.1|	27	55	38	556	11670
gi|524909114|ref|XP_005109609.1|	26	68	27	458	5086
gi|526117385|ref|NP_001267755.1|	26	74	26	376	8603
gi|556097340|gb|ESO85992.1|	26	47	27	343	7708
gi|124233|sp|P22488.2|IFEA_HELAS	25	52	30	304	6401
gi|524875601|ref|XP_005094526.1|	25	48	32	197	4019

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524887592|ref|XP_005100380.1|	25	52	25	143	2627
gi|524883288|ref|XP_005098277.1|	24	42	24	97	1424
gi|524883288|ref|XP_005098277.1|	24	42	24	97	1424
gi|524895745|ref|XP_005104354.1|	24	40	40	301	5740
gi|524911954|ref|XP_005110834.1|	24	43	25	161	2654
gi|524916507|ref|XP_005113030.1|	24	60	28	195	3926
gi|325296989|ref|NP_001191508.1|	23	67	23	102	1713
gi|524885520|ref|XP_005099366.1|	23	80	23	291	2963
gi|524892924|ref|XP_005102985.1|	23	50	23	106	1768
gi|524896814|ref|XP_005104878.1|	23	32	23	136	1278
gi|524912440|ref|XP_005111068.1|	23	67	23	501	7869
gi|556097683|gb|ESO86335.1|	23	63	23	183	2442
gi|524872922|ref|XP_005093216.1|	22	42	22	125	3314
gi|524886817|ref|XP_005100000.1|	22	43	22	108	2123
gi|524895745|ref|XP_005104354.1|	22	29	24	96	2060
gi|524910869|ref|XP_005110308.1|	22	53	22	66	1223
gi|524882748|ref|XP_005098012.1|	21	38	29	90	1701
gi|524900673|ref|XP_005106755.1|	21	64	21	144	2238
gi|524916507|ref|XP_005113030.1|	21	64	26	125	2892
gi|556095567|gb|ESO84220.1|	21	43	22	38	770
gi|524865955|ref|XP_005089803.1|	20	68	20	142	2804
gi|524885898|ref|XP_005099550.1|	20	34	20	44	721
gi|524900516|ref|XP_005106678.1|	20	59	20	404	5064
gi|524911947|ref|XP_005110831.1|	20	50	20	97	1234
gi|524915789|ref|XP_005112685.1|	20	35	20	80	1284
gi|526117385|ref|NP_001267755.1|	20	55	20	194	3066
gi|556110866|gb|ESO99517.1|	20	15	34	150	1602
gi|218683627|gb|ACL00842.1|	19	39	19	31	490
gi|524865670|ref|XP_005089662.1|	19	90	20	397	10161
gi|524883990|ref|XP_005098620.1|	19	56	19	128	2364
gi|524889526|ref|XP_005101318.1|	19	33	19	77	1281
gi|524896814|ref|XP_005104878.1|	19	16	19	61	676
gi|6682319|emb|CAB64662.1|	19	45	28	223	4006
gi|524872674|ref|XP_005093097.1|	18	44	18	62	1156
gi|524882968|ref|XP_005098119.1|	18	36	18	59	1678
gi|524884737|ref|XP_005098987.1|	18	38	18	79	1549
gi|524893715|ref|XP_005103366.1|	18	61	18	103	563
gi|524895745|ref|XP_005104354.1|	18	46	35	249	5457
gi|524899582|ref|XP_005106220.1|	18	56	18	106	2069
gi|524915771|ref|XP_005112676.1|	18	48	18	149	1821
gi|42559558|sp|O97192.1|TPM_HELAS	17	65	29	574	9620

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524871885|ref|XP_005092713.1|	17	43	17	77	575
gi|524895745|ref|XP_005104354.1|	17	46	21	222	4906
gi|526117405|ref|NP_001267760.1|	17	89	23	561	7405
gi|556110866|gb|ESO99517.1|	17	17	26	41	1148
gi|2073142|dbj|BAA19861.1|	16	71	16	138	1786
gi|346987840|gb|AEO51766.1|	16	8	17	23	325
gi|524872618|ref|XP_005093070.1|	16	28	16	30	834
gi|524886261|ref|XP_005099727.1|	16	18	16	100	1693
gi|524886695|ref|XP_005099940.1|	16	43	16	91	1539
gi|524888508|ref|XP_005100824.1|	16	18	41	64	994
gi|524898231|ref|XP_005105567.1|	16	36	16	42	435
gi|524898324|ref|XP_005105612.1|	16	32	16	52	1149
gi|524907580|ref|XP_005108906.1|	16	8	16	26	301
gi|773571|emb|CAA60122.1|	16	29	17	37	725
gi|325197136|ref|NP_001191419.1|	15	84	15	137	2192
gi|346987840|gb|AEO51766.1|	15	22	15	78	1164
gi|51105058|gb|AAT97089.1|	15	31	16	38	290
gi|524871134|ref|XP_005092343.1|	15	41	15	171	4821
gi|524875728|ref|XP_005094588.1|	15	42	15	77	829
gi|524876525|ref|XP_005094978.1|	15	54	19	144	1905
gi|524881606|ref|XP_005097453.1|	15	40	15	23	284
gi|524909838|ref|XP_005109815.1|	15	42	15	72	1527
gi|328933188|gb|AEB70965.1|	14	49	14	86	1454
gi|405953142|gb|EKC20858.1|	14	33	14	81	973
gi|524876525|ref|XP_005094978.1|	14	50	18	134	2002
gi|524878973|ref|XP_005096166.1|	14	39	14	124	2877
gi|524883622|ref|XP_005098438.1|	14	32	14	26	576
gi|524885795|ref|XP_005099501.1|	14	51	22	215	4236
gi|524891584|ref|XP_005102325.1|	14	36	14	79	1813
gi|524911952|ref|XP_005110833.1|	14	60	17	151	1575
gi|524911954|ref|XP_005110834.1|	14	37	16	74	1039
gi|524913044|ref|XP_005111356.1|	14	41	14	59	899
gi|524916317|ref|XP_005112938.1|	14	45	14	85	1262
gi|524917241|ref|XP_005113381.1|	14	36	14	190	3548
gi|556095158|gb|ESO83811.1|	14	31	14	24	467
gi|564743500|gb|AHC02701.1|	14	72	34	1115	21582
gi|321472487|gb|EFX83457.1|	13	21	13	38	516
gi|524865670|ref|XP_005089662.1|	13	62	16	168	4244
gi|524867613|ref|XP_005090615.1|	13	14	13	27	392
gi|524871402|ref|XP_005092474.1|	13	29	13	43	759
gi|524875601|ref|XP_005094526.1|	13	32	21	70	1251

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524880840|ref|XP_005097078.1|	13	40	13	46	1014
gi|524882748|ref|XP_005098012.1|	13	27	18	67	901
gi|524890599|ref|XP_005101843.1|	13	30	13	26	341
gi|524903522|ref|XP_005107960.1|	13	42	13	93	1813
gi|524911950|ref|XP_005110832.1|	13	32	13	32	501
gi|555688717|gb|ESN91949.1|	13	62	13	192	2334
gi|325297040|ref|NP_001191523.1|	12	40	12	43	439
gi|385145402|emb|CCG28026.1|	12	36	12	63	1157
gi|524867573|ref|XP_005090597.1|	12	32	12	117	2361
gi|524873169|ref|XP_005093336.1|	12	38	12	26	625
gi|524876525|ref|XP_005094978.1|	12	39	14	81	1539
gi|524881811|ref|XP_005097555.1|	12	35	12	118	2051
gi|524895342|ref|XP_005104156.1|	12	30	12	23	591
gi|524898259|ref|XP_005105581.1|	12	48	12	106	1802
gi|524907580|ref|XP_005108906.1|	12	6	12	27	317
gi|524907822|ref|XP_005109023.1|	12	43	12	20	313
gi|524913368|ref|XP_005111515.1|	12	37	12	59	1383
gi|524916601|ref|XP_005113074.1|	12	10	12	16	216
gi|576693750|gb|EUB57355.1|	12	51	12	107	1238
gi|74912853|sp|Q6QUW1.1|RGP51_LYMST	12	31	16	84	1496
gi|241273677|ref|XP_002406620.1|	11	26	11	82	1095
gi|325297052|ref|NP_001191530.1|	11	65	35	109	1751
gi|524868522|ref|XP_005091063.1|	11	12	11	19	329
gi|524871732|ref|XP_005092637.1|	11	6	11	52	801
gi|524872616|ref|XP_005093069.1|	11	17	11	27	741
gi|524876391|ref|XP_005094913.1|	11	32	11	45	724
gi|524876525|ref|XP_005094978.1|	11	42	16	83	1709
gi|524876525|ref|XP_005094978.1|	11	44	17	149	2175
gi|524877261|ref|XP_005095336.1|	11	23	11	40	708
gi|524910025|ref|XP_005109900.1|	11	52	11	34	479
gi|524911947|ref|XP_005110831.1|	11	61	14	88	1049
gi|556104876|gb|ESO93528.1|	11	16	14	31	461
gi|556105302|gb|ESO93954.1|	11	43	11	44	836
gi|585719304|ref|XP_006826038.1|	11	6	11	14	147
gi|325297146|ref|NP_001191582.1|	10	46	14	28	435
gi|405972388|gb|EKC37161.1|	10	43	10	11	240
gi|513137512|gb|AGO20698.1|	10	18	10	23	270
gi|524866272|ref|XP_005089958.1|	10	35	10	29	588
gi|524867792|ref|XP_005090703.1|	10	22	10	23	499
gi|524879223|ref|XP_005096289.1|	10	42	10	46	1026

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524879388|ref|XP_005096370.1|	10	33	10	59	1472
gi|524882422|ref|XP_005097855.1|	10	35	10	75	1587
gi|524884878|ref|XP_005099053.1|	10	39	10	118	1913
gi|524885821|ref|XP_005099514.1|	10	33	10	15	276
gi|524887319|ref|XP_005100247.1|	10	26	10	32	592
gi|524891306|ref|XP_005102188.1|	10	35	10	40	723
gi|524894818|ref|XP_005103900.1|	10	48	11	88	1016
gi|524896250|ref|XP_005104602.1|	10	40	10	60	978
gi|524896457|ref|XP_005104703.1|	10	37	10	30	451
gi|524902628|ref|XP_005107543.1|	10	45	10	34	351
gi|524903041|ref|XP_005107738.1|	10	20	10	21	175
gi|524906745|ref|XP_005108595.1|	10	37	10	28	516
gi|524915110|ref|XP_005112361.1|	10	21	10	18	280
gi|556098588|gb|ESO87240.1|	10	38	10	37	723
gi|576693750|gb|EUB57355.1|	10	67	10	60	1229
gi|126697356|gb|ABO26635.1|	9	35	9	58	1175
gi|150251390|gb|ABR68007.1|	9	23	9	95	1958
gi|325296986|ref|NP_001191507.1|	9	12	9	34	621
gi|346987844|gb|AEO51768.1|	9	5	9	35	985
gi|524864482|ref|XP_005089076.1|	9	25	9	26	612
gi|524866272|ref|XP_005089958.1|	9	25	9	21	433
gi|524866912|ref|XP_005090271.1|	9	19	9	48	891
gi|524869485|ref|XP_005091535.1|	9	16	9	46	642
gi|524869987|ref|XP_005091780.1|	9	62	9	26	473
gi|524874769|ref|XP_005094119.1|	9	16	9	19	419
gi|524878388|ref|XP_005095881.1|	9	17	9	19	364
gi|524881984|ref|XP_005097638.1|	9	56	9	24	327
gi|524884630|ref|XP_005098934.1|	9	32	9	47	1055
gi|524885795|ref|XP_005099501.1|	9	63	12	145	2592
gi|524886540|ref|XP_005099863.1|	9	31	9	21	620
gi|524886658|ref|XP_005099922.1|	9	22	9	27	387
gi|524889396|ref|XP_005101256.1|	9	32	9	22	357
gi|524891426|ref|XP_005102246.1|	9	29	9	52	1517
gi|524892644|ref|XP_005102848.1|	9	44	9	124	3499
gi|524893352|ref|XP_005103192.1|	9	28	9	20	430
gi|524901897|ref|XP_005107352.1|	9	37	9	36	791
gi|524907533|ref|XP_005108883.1|	9	32	9	50	1341
gi|524907610|ref|XP_005108921.1|	9	27	9	15	297
gi|524915397|ref|XP_005112499.1|	9	22	9	38	484
gi|54020890|ref|NP_001005689.1|	9	30	9	47	659
gi|556095158|gb|ESO83811.1|	9	21	9	36	971

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|556095567|gb|ESO84220.1|	9	25	9	30	751
gi|556097226|gb|ESO85878.1|	9	24	9	23	462
gi|556099804|gb|ESO88456.1|	9	17	9	15	123
gi|556113964|gb|ESP02616.1|	9	20	9	32	712
gi|994895|gb|AAB34494.1|	9	34	11	70	1300
gi|156399584|ref|XP_001638581.1|	8	57	12	67	1619
gi|2073142|dbj|BAA19861.1|	8	43	8	122	2764
gi|296190822|ref|XP_002743358.1|	8	13	8	15	242
gi|325297144|ref|NP_001191581.1|	8	15	8	17	208
gi|351709298|gb|EHB12217.1|	8	54	8	191	3217
gi|402578741|gb|EJW72694.1|	8	38	8	108	1422
gi|524872618|ref|XP_005093070.1|	8	11	8	13	120
gi|524873600|ref|XP_005093544.1|	8	39	8	23	302
gi|524876525|ref|XP_005094978.1|	8	29	11	42	765
gi|524877031|ref|XP_005095223.1|	8	34	8	36	162
gi|524880725|ref|XP_005097023.1|	8	36	8	28	509
gi|524881984|ref|XP_005097638.1|	8	64	8	60	1510
gi|524882922|ref|XP_005098097.1|	8	31	8	28	633
gi|524885854|ref|XP_005099530.1|	8	36	8	141	2899
gi|524886289|ref|XP_005099741.1|	8	24	8	18	128
gi|524887834|ref|XP_005100500.1|	8	50	8	30	563
gi|524887883|ref|XP_005100524.1|	8	37	8	20	280
gi|524890176|ref|XP_005101637.1|	8	19	8	11	167
gi|524893673|ref|XP_005103345.1|	8	15	8	14	308
gi|524896407|ref|XP_005104678.1|	8	32	8	30	240
gi|524896874|ref|XP_005104908.1|	8	29	8	52	694
gi|524897999|ref|XP_005105454.1|	8	26	8	34	678
gi|524900380|ref|XP_005106610.1|	8	38	8	17	460
gi|524900390|ref|XP_005106615.1|	8	27	10	41	633
gi|524906491|ref|XP_005108543.1|	8	10	8	31	478
gi|524907238|ref|XP_005108742.1|	8	21	8	12	119
gi|524910284|ref|XP_005110023.1|	8	12	8	20	236
gi|524914230|ref|XP_005111933.1|	8	18	8	20	362
gi|541039654|gb|ERG79357.1|	8	21	8	11	79
gi|556103805|gb|ESO92457.1|	8	27	8	17	423
gi|556108834|gb|ESO97486.1|	8	19	8	22	503
gi|126697362|gb|ABO26638.1|	7	46	7	45	567
gi|168252891|gb|ACA23131.1|	7	53	7	16	302
gi|2073142|dbj|BAA19861.1|	7	55	7	190	4407
gi|291235073|ref|XP_002737469.1|	7	24	7	19	253
gi|346987844|gb|AEO51768.1|	7	4	8	9	190

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|353236678|emb|CCA68668.1|	7	24	7	13	266
gi|51105064|gb|AAT97092.1|	7	31	7	28	580
gi|524865020|ref|XP_005089342.1|	7	17	7	15	340
gi|524866304|ref|XP_005089974.1|	7	23	7	12	135
gi|524866627|ref|XP_005090132.1|	7	35	7	18	110
gi|524870833|ref|XP_005092197.1|	7	22	7	11	154
gi|524873892|ref|XP_005093688.1|	7	31	7	11	399
gi|524874065|ref|XP_005093774.1|	7	26	7	18	335
gi|524874374|ref|XP_005093923.1|	7	52	14	108	1696
gi|524874655|ref|XP_005094062.1|	7	6	7	31	499
gi|524876098|ref|XP_005094768.1|	7	17	7	22	531
gi|524876525|ref|XP_005094978.1|	7	30	8	22	344
gi|524877054|ref|XP_005095234.1|	7	28	7	9	144
gi|524881743|ref|XP_005097521.1|	7	34	7	75	1275
gi|524882484|ref|XP_005097885.1|	7	18	7	12	127
gi|524882717|ref|XP_005097997.1|	7	22	7	29	523
gi|524883540|ref|XP_005098397.1|	7	29	7	11	282
gi|524888414|ref|XP_005100778.1|	7	33	7	20	256
gi|524891844|ref|XP_005102454.1|	7	23	7	10	104
gi|524892967|ref|XP_005103005.1|	7	47	7	43	616
gi|524894544|ref|XP_005103769.1|	7	33	7	8	132
gi|524895123|ref|XP_005104049.1|	7	40	7	10	250
gi|524895893|ref|XP_005104427.1|	7	12	7	16	393
gi|524896519|ref|XP_005104733.1|	7	34	9	24	430
gi|524897999|ref|XP_005105454.1|	7	26	7	16	230
gi|524906739|ref|XP_005108592.1|	7	28	7	16	514
gi|524908996|ref|XP_005109585.1|	7	17	7	19	317
gi|524910895|ref|XP_005110321.1|	7	20	7	11	272
gi|524911570|ref|XP_005110646.1|	7	2	7	13	190
gi|524912929|ref|XP_005111300.1|	7	56	7	34	540
gi|524914829|ref|XP_005112224.1|	7	16	7	16	311
gi|524916452|ref|XP_005113003.1|	7	16	7	46	732
gi|524916607|ref|XP_005113077.1|	7	33	7	59	765
gi|529444582|ref|XP_005242003.1|	7	43	7	11	175
gi|54020890|ref|NP_001005689.1|	7	24	7	19	343
gi|546686668|gb|ERL95792.1|	7	55	7	58	735
gi|556098160|gb|ESO86812.1|	7	24	7	20	450
gi|556104876|gb|ESO93528.1|	7	14	13	21	231
gi|556105302|gb|ESO93954.1|	7	43	7	63	1043
gi|556116273|gb|ESP04925.1|	7	35	7	34	311
gi|556116376|gb|ESP05028.1|	7	19	7	20	235

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|557010651|ref|XP_006006018.1|	7	18	7	8	202
gi|620958561|ref|XP_007664363.1|	7	29	7	38	461
gi|150251390|gb|ABR68007.1|	6	12	6	47	406
gi|152032020|gb|ABS28869.1|	6	22	6	86	888
gi|226478902|emb|CAX72946.1|	6	21	7	12	198
gi|325296727|ref|NP_001191598.1|	6	25	6	7	79
gi|346987844|gb|AEO51768.1|	6	15	7	30	372
gi|405952673|gb|EKC20457.1|	6	31	7	35	409
gi|405976077|gb|EKC40598.1|	6	10	6	11	196
gi|421975936|gb|AFX73000.1|	6	21	6	52	499
gi|443708569|gb|ELU03646.1|	6	22	6	18	417
gi|524865616|ref|XP_005089636.1|	6	20	6	13	358
gi|524865670|ref|XP_005089662.1|	6	52	6	117	2465
gi|524866780|ref|XP_005090207.1|	6	18	6	14	357
gi|524869199|ref|XP_005091396.1|	6	21	6	16	265
gi|524871402|ref|XP_005092474.1|	6	17	6	28	475
gi|524872725|ref|XP_005093120.1|	6	16	6	13	253
gi|524874372|ref|XP_005093922.1|	6	14	6	8	108
gi|524876525|ref|XP_005094978.1|	6	53	8	21	243
gi|524876525|ref|XP_005094978.1|	6	42	11	76	1207
gi|524879821|ref|XP_005096582.1|	6	28	6	15	463
gi|524880980|ref|XP_005097146.1|	6	19	6	18	218
gi|524881119|ref|XP_005097215.1|	6	13	6	12	140
gi|524881385|ref|XP_005097344.1|	6	34	6	12	227
gi|524881557|ref|XP_005097429.1|	6	29	6	33	740
gi|524881651|ref|XP_005097475.1|	6	12	6	12	230
gi|524882420|ref|XP_005097854.1|	6	38	7	38	846
gi|524882484|ref|XP_005097885.1|	6	19	6	20	329
gi|524883534|ref|XP_005098394.1|	6	41	6	28	409
gi|524883626|ref|XP_005098440.1|	6	51	6	48	1209
gi|524885581|ref|XP_005099396.1|	6	30	6	13	139
gi|524886950|ref|XP_005100065.1|	6	15	6	12	183
gi|524887258|ref|XP_005100217.1|	6	16	6	11	178
gi|524888067|ref|XP_005100610.1|	6	6	6	23	241
gi|524889482|ref|XP_005101296.1|	6	12	6	16	329
gi|524889720|ref|XP_005101414.1|	6	34	6	16	331
gi|524889992|ref|XP_005101546.1|	6	34	6	21	776
gi|524892640|ref|XP_005102846.1|	6	44	6	13	253
gi|524894013|ref|XP_005103514.1|	6	14	6	15	345
gi|524894778|ref|XP_005103881.1|	6	13	6	10	174
gi|524899715|ref|XP_005106286.1|	6	15	6	12	308

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524901536|ref|XP_005107174.1|	6	14	6	9	95
gi|524901909|ref|XP_005107358.1|	6	36	6	9	130
gi|524902626|ref|XP_005107542.1|	6	31	6	10	154
gi|524903367|ref|XP_005107884.1|	6	19	6	10	213
gi|524907772|ref|XP_005108998.1|	6	12	6	14	349
gi|524908509|ref|XP_005109349.1|	6	38	6	21	438
gi|524909258|ref|XP_005109678.1|	6	20	6	14	307
gi|524911513|ref|XP_005110619.1|	6	33	6	11	194
gi|524912191|ref|XP_005110947.1|	6	59	6	12	189
gi|524913881|ref|XP_005111763.1|	6	11	6	10	173
gi|524914515|ref|XP_005112072.1|	6	28	6	10	75
gi|524916601|ref|XP_005113074.1|	6	6	6	33	564
gi|526117405|ref|NP_001267760.1|	6	68	13	195	1773
gi|555704960|gb|ESO08193.1|	6	53	6	103	2396
gi|556103805|gb|ESO92457.1|	6	22	6	7	102
gi|556105072|gb|ESO93724.1|	6	24	6	12	397
gi|556105072|gb|ESO93724.1|	6	15	6	29	511
gi|556113526|gb|ESP02178.1|	6	11	6	10	145
gi|556113946|gb|ESP02598.1|	6	11	6	9	174
gi|556114460|gb|ESP03112.1|	6	23	6	12	205
gi|1174755|sp|P43689.1|TPM2_BIOGL	5	39	12	63	965
gi|325197162|ref|NP_001191425.1|	5	29	5	5	101
gi|325296839|ref|NP_001191661.1|	5	31	5	6	93
gi|405970776|gb|EKC35652.1|	5	34	5	20	217
gi|42559558|sp|O97192.1|TPM_HELAS	5	51	19	277	4877
gi|431831576|gb|AGA92562.1|	5	39	5	9	193
gi|468860736|gb|AGH32328.1|	5	8	5	7	139
gi|4929136|gb|AAD33872.1|AF141348_1	5	19	5	10	219
gi|524864626|ref|XP_005089148.1|	5	16	5	8	134
gi|524866300|ref|XP_005089972.1|	5	19	5	16	269
gi|524866556|ref|XP_005090099.1|	5	9	5	7	124
gi|524870028|ref|XP_005091800.1|	5	11	5	5	141
gi|524870886|ref|XP_005092223.1|	5	48	5	19	609
gi|524873315|ref|XP_005093407.1|	5	27	5	8	186
gi|524874171|ref|XP_005093825.1|	5	26	5	11	244
gi|524874376|ref|XP_005093924.1|	5	72	8	53	1019
gi|524876523|ref|XP_005094977.1|	5	67	5	15	182
gi|524877040|ref|XP_005095227.1|	5	29	5	21	544
gi|524879616|ref|XP_005096482.1|	5	18	5	7	154
gi|524879821|ref|XP_005096582.1|	5	24	5	14	269
gi|524879933|ref|XP_005096636.1|	5	19	5	8	109

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524880339|ref|XP_005096834.1|	5	16	5	8	163
gi|524880906|ref|XP_005097111.1|	5	20	5	13	279
gi|524881344|ref|XP_005097324.1|	5	18	5	10	191
gi|524882120|ref|XP_005097705.1|	5	16	5	11	176
gi|524882774|ref|XP_005098023.1|	5	20	5	7	146
gi|524883534|ref|XP_005098394.1|	5	34	5	43	1018
gi|524885619|ref|XP_005099415.1|	5	20	5	18	350
gi|524886181|ref|XP_005099688.1|	5	43	9	52	1391
gi|524886263|ref|XP_005099728.1|	5	11	5	8	158
gi|524886299|ref|XP_005099746.1|	5	19	5	23	437
gi|524886483|ref|XP_005099835.1|	5	32	5	13	196
gi|524886570|ref|XP_005099878.1|	5	43	5	28	835
gi|524886963|ref|XP_005100071.1|	5	6	5	6	118
gi|524888229|ref|XP_005100690.1|	5	8	5	10	297
gi|524889251|ref|XP_005101186.1|	5	27	5	17	542
gi|524891178|ref|XP_005102125.1|	5	17	5	16	203
gi|524891634|ref|XP_005102350.1|	5	16	5	7	115
gi|524892341|ref|XP_005102699.1|	5	22	7	33	289
gi|524892819|ref|XP_005102934.1|	5	10	5	6	118
gi|524893098|ref|XP_005103069.1|	5	31	5	23	251
gi|524893793|ref|XP_005103405.1|	5	9	5	8	244
gi|524894261|ref|XP_005103632.1|	5	5	5	6	157
gi|524894818|ref|XP_005103900.1|	5	25	6	26	126
gi|524895639|ref|XP_005104301.1|	5	19	5	6	116
gi|524898799|ref|XP_005105845.1|	5	3	5	5	93
gi|524898960|ref|XP_005105922.1|	5	10	5	9	162
gi|524902421|ref|XP_005107462.1|	5	39	5	13	222
gi|524902628|ref|XP_005107543.1|	5	31	5	7	150
gi|524902929|ref|XP_005107684.1|	5	17	5	15	191
gi|524903247|ref|XP_005107829.1|	5	13	5	22	407
gi|524908412|ref|XP_005109303.1|	5	11	5	6	84
gi|524912853|ref|XP_005111263.1|	5	11	5	5	86
gi|524912929|ref|XP_005111300.1|	5	28	5	12	169
gi|524915771|ref|XP_005112676.1|	5	32	7	37	499
gi|524917285|ref|XP_005113402.1|	5	15	5	14	230
gi|556100834|gb|ESO89486.1|	5	23	5	7	131
gi|556101047|gb|ESO89699.1|	5	4	5	5	94
gi|556113799|gb|ESP02451.1|	5	16	5	10	220
gi|556114398|gb|ESP03050.1|	5	16	5	11	93
gi|556114539|gb|ESP03191.1|	5	12	5	11	224
gi|602626279|ref|XP_007420100.1|	5	42	6	117	2630

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|608090766|gb|EZG30020.1|	5	23	5	17	271
gi|632946707|ref|XP_007888691.1|	5	13	5	16	273
gi|104162064|emb|CAK19329.1|	4	42	4	67	1304
gi|126697398|gb|ABO26656.1|	4	17	4	10	235
gi|150251392|gb|ABR68008.1|	4	8	4	10	202
gi|158997667|gb|ABW86957.1|	4	35	4	51	1030
gi|325296803|ref|NP_001191642.1|	4	32	4	8	126
gi|325296995|ref|NP_001191510.1|	4	38	4	7	102
gi|325297040|ref|NP_001191523.1|	4	11	4	8	128
gi|325504381|emb|CBX41690.1|	4	24	4	17	170
gi|405951515|gb|EKC19421.1|	4	12	4	7	233
gi|405959696|gb|EKC25703.1|	4	13	4	42	672
gi|405972472|gb|EKC37239.1|	4	8	4	8	221
gi|405973356|gb|EKC38075.1|	4	18	4	7	81
gi|4249742|gb|AAD13782.1|	4	11	4	4	81
gi|42559558|sp|O97192.1|TPM_HELAS	4	26	9	19	264
gi|443694956|gb|ELT95974.1|	4	26	4	25	280
gi|4504279|ref|NP_002098.1|	4	19	4	28	131
gi|51105058|gb|AAT97089.1|	4	8	4	6	107
gi|521024335|gb|EPQ06123.1|	4	15	4	14	182
gi|524863988|ref|XP_005088833.1|	4	20	4	6	148
gi|524867613|ref|XP_005090615.1|	4	4	4	8	92
gi|524869393|ref|XP_005091491.1|	4	17	4	13	178
gi|524870300|ref|XP_005091935.1|	4	20	4	7	159
gi|524870848|ref|XP_005092204.1|	4	6	4	6	144
gi|524870945|ref|XP_005092251.1|	4	32	4	9	157
gi|524871178|ref|XP_005092365.1|	4	18	4	7	189
gi|524871180|ref|XP_005092366.1|	4	28	4	9	128
gi|524871774|ref|XP_005092658.1|	4	23	4	8	94
gi|524872775|ref|XP_005093144.1|	4	8	4	7	167
gi|524873604|ref|XP_005093546.1|	4	31	4	11	184
gi|524874374|ref|XP_005093923.1|	4	40	4	8	156
gi|524874452|ref|XP_005093961.1|	4	21	4	13	276
gi|524875207|ref|XP_005094334.1|	4	16	4	6	85
gi|524875325|ref|XP_005094393.1|	4	33	4	7	122
gi|524876601|ref|XP_005095013.1|	4	15	4	8	168
gi|524876605|ref|XP_005095015.1|	4	7	4	6	150
gi|524877129|ref|XP_005095271.1|	4	24	4	11	123
gi|524877675|ref|XP_005095535.1|	4	6	4	16	337
gi|524879957|ref|XP_005096647.1|	4	17	4	10	223
gi|524883640|ref|XP_005098447.1|	4	11	4	20	423

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524883762|ref|XP_005098507.1|	4	13	4	5	88
gi|524884099|ref|XP_005098673.1|	4	6	4	5	93
gi|524884550|ref|XP_005098894.1|	4	12	4	6	98
gi|524886285|ref|XP_005099739.1|	4	42	4	13	120
gi|524886483|ref|XP_005099835.1|	4	30	4	14	182
gi|524888539|ref|XP_005100839.1|	4	9	4	6	102
gi|524890683|ref|XP_005101885.1|	4	3	4	5	78
gi|524892173|ref|XP_005102616.1|	4	15	4	10	177
gi|524892281|ref|XP_005102670.1|	4	4	4	14	285
gi|524895031|ref|XP_005104004.1|	4	9	4	6	109
gi|524895469|ref|XP_005104217.1|	4	13	4	7	97
gi|524896519|ref|XP_005104733.1|	4	23	6	15	153
gi|524896868|ref|XP_005104905.1|	4	18	4	5	151
gi|524899361|ref|XP_005106113.1|	4	8	4	6	74
gi|524900084|ref|XP_005106466.1|	4	10	4	5	148
gi|524900809|ref|XP_005106822.1|	4	14	4	8	146
gi|524900993|ref|XP_005106911.1|	4	16	4	6	116
gi|524901357|ref|XP_005107086.1|	4	30	4	5	73
gi|524901394|ref|XP_005107104.1|	4	23	4	12	567
gi|524906293|ref|XP_005108469.1|	4	24	4	11	242
gi|524907491|ref|XP_005108862.1|	4	15	4	11	407
gi|524907576|ref|XP_005108904.1|	4	24	4	22	366
gi|524910025|ref|XP_005109900.1|	4	31	4	6	116
gi|524910373|ref|XP_005110066.1|	4	10	4	14	250
gi|524910660|ref|XP_005110205.1|	4	9	4	5	92
gi|524910895|ref|XP_005110321.1|	4	13	4	4	90
gi|524911650|ref|XP_005110685.1|	4	12	4	11	256
gi|524913419|ref|XP_005111540.1|	4	3	4	5	91
gi|524913668|ref|XP_005111659.1|	4	18	4	7	164
gi|524914469|ref|XP_005112049.1|	4	17	4	7	188
gi|524915677|ref|XP_005112631.1|	4	23	4	19	312
gi|526117431|ref|NP_001267767.1|	4	31	4	11	161
gi|556093552|gb|ESO82205.1|	4	22	4	18	272
gi|556105117|gb|ESO93769.1|	4	14	5	24	165
gi|556106038|gb|ESO94690.1|	4	11	4	6	146
gi|556110210|gb|ESO98862.1|	4	20	4	6	161
gi|556110891|gb|ESO99542.1|	4	22	4	4	99
gi|556111437|gb|ESP00089.1|	4	18	4	5	60
gi|577029907|gb|AHH81788.1|	4	24	4	8	98
gi|625230674|ref|XP_007606483.1|	4	53	4	10	129
gi|639127132|gb|AIA24462.1|	4	24	4	5	169

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|1174755|sp|P43689.1|TPM2_BIOGL	3	15	3	3	52
gi|126697410|gb|ABO26662.1|	3	18	3	34	727
gi|133874728|gb|ABO40828.1|	3	13	3	4	61
gi|150251390|gb|ABR68007.1|	3	9	3	8	176
gi|150251390|gb|ABR68007.1|	3	6	3	3	86
gi|150251390|gb|ABR68007.1|	3	18	3	3	45
gi|210076789|gb|ACJ06748.1|	3	27	3	10	203
gi|220979902|emb|CAQ64775.1|	3	14	3	4	86
gi|253317417|gb|ACT22633.1|	3	47	6	48	407
gi|260821662|ref|XP_002606151.1|	3	16	5	11	225
gi|260821662|ref|XP_002606151.1|	3	13	5	31	339
gi|325197162|ref|NP_001191425.1|	3	13	3	8	133
gi|325296743|ref|NP_001191607.1|	3	6	3	11	208
gi|346987840|gb|AEO51766.1|	3	13	3	60	1298
gi|374718481|gb|AEZ67032.1|	3	6	3	8	92
gi|405965644|gb|EKC31006.1|	3	14	3	4	89
gi|405965813|gb|EKC31167.1|	3	14	3	7	79
gi|443705209|gb|ELU01864.1|	3	22	3	7	163
gi|443729497|gb|ELU15363.1|	3	7	3	3	87
gi|472824679|dbj|BAN09032.1|	3	39	4	29	223
gi|524864532|ref|XP_005089101.1|	3	10	3	6	111
gi|524864665|ref|XP_005089167.1|	3	18	3	7	92
gi|524865020|ref|XP_005089342.1|	3	33	3	8	270
gi|524870191|ref|XP_005091881.1|	3	11	3	14	328
gi|524870815|ref|XP_005092188.1|	3	11	3	5	117
gi|524871621|ref|XP_005092582.1|	3	8	4	7	107
gi|524871621|ref|XP_005092582.1|	3	14	4	4	59
gi|524872832|ref|XP_005093172.1|	3	8	3	12	294
gi|524872846|ref|XP_005093178.1|	3	25	3	15	219
gi|524872934|ref|XP_005093222.1|	3	23	3	6	134
gi|524873892|ref|XP_005093688.1|	3	33	3	10	383
gi|524874194|ref|XP_005093836.1|	3	9	3	8	168
gi|524874346|ref|XP_005093909.1|	3	19	3	7	92
gi|524875029|ref|XP_005094246.1|	3	20	3	4	132
gi|524875666|ref|XP_005094557.1|	3	33	3	4	66
gi|524876523|ref|XP_005094977.1|	3	40	4	5	74
gi|524876525|ref|XP_005094978.1|	3	33	6	13	221
gi|524877265|ref|XP_005095338.1|	3	8	3	6	101
gi|524877372|ref|XP_005095389.1|	3	20	3	15	126
gi|524877853|ref|XP_005095623.1|	3	23	3	43	726
gi|524879487|ref|XP_005096418.1|	3	17	3	4	66

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524879771|ref|XP_005096557.1|	3	4	3	4	106
gi|524880082|ref|XP_005096709.1|	3	7	3	7	222
gi|524881557|ref|XP_005097429.1|	3	11	3	4	85
gi|524881663|ref|XP_005097481.1|	3	13	3	3	71
gi|524881663|ref|XP_005097481.1|	3	11	3	4	65
gi|524881729|ref|XP_005097514.1|	3	7	3	7	155
gi|524881799|ref|XP_005097549.1|	3	14	3	4	48
gi|524881948|ref|XP_005097620.1|	3	5	3	8	267
gi|524882305|ref|XP_005097797.1|	3	22	3	9	266
gi|524882420|ref|XP_005097854.1|	3	8	4	23	462
gi|524882460|ref|XP_005097873.1|	3	28	3	16	385
gi|524882748|ref|XP_005098012.1|	3	7	3	4	110
gi|524884884|ref|XP_005099056.1|	3	17	3	4	53
gi|524885143|ref|XP_005099182.1|	3	8	3	4	62
gi|524885272|ref|XP_005099245.1|	3	28	3	27	904
gi|524885593|ref|XP_005099402.1|	3	15	3	53	1426
gi|524886568|ref|XP_005099877.1|	3	25	3	7	79
gi|524888237|ref|XP_005100694.1|	3	23	3	3	105
gi|524889096|ref|XP_005101109.1|	3	18	3	9	102
gi|524889992|ref|XP_005101546.1|	3	43	3	13	405
gi|524891741|ref|XP_005102403.1|	3	8	3	6	82
gi|524893468|ref|XP_005103247.1|	3	23	3	4	110
gi|524894608|ref|XP_005103800.1|	3	5	3	6	139
gi|524895637|ref|XP_005104300.1|	3	12	3	11	245
gi|524896746|ref|XP_005104844.1|	3	22	3	6	146
gi|524898165|ref|XP_005105535.1|	3	7	3	3	75
gi|524898449|ref|XP_005105673.1|	3	16	3	9	208
gi|524898531|ref|XP_005105713.1|	3	11	3	8	128
gi|524898670|ref|XP_005105781.1|	3	10	3	3	96
gi|524898743|ref|XP_005105817.1|	3	26	3	5	86
gi|524899675|ref|XP_005106266.1|	3	12	3	3	85
gi|524899675|ref|XP_005106266.1|	3	11	3	7	125
gi|524900353|ref|XP_005106597.1|	3	7	3	5	71
gi|524900386|ref|XP_005106613.1|	3	8	3	3	95
gi|524900516|ref|XP_005106678.1|	3	9	3	11	245
gi|524903383|ref|XP_005107892.1|	3	20	3	7	123
gi|524903438|ref|XP_005107919.1|	3	12	3	3	87
gi|524905999|ref|XP_005108354.1|	3	18	3	8	147
gi|524906745|ref|XP_005108595.1|	3	20	4	7	193
gi|524907151|ref|XP_005108701.1|	3	6	3	4	121
gi|524907406|ref|XP_005108821.1|	3	11	3	4	80

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524907727|ref|XP_005108978.1|	3	11	3	7	204
gi|524907998|ref|XP_005109108.1|	3	10	3	5	88
gi|524909038|ref|XP_005109605.1|	3	10	3	9	272
gi|524910167|ref|XP_005109966.1|	3	11	3	10	197
gi|524910867|ref|XP_005110307.1|	3	9	3	10	204
gi|524911945|ref|XP_005110830.1|	3	24	3	5	79
gi|524912291|ref|XP_005110996.1|	3	31	3	6	150
gi|524912715|ref|XP_005111198.1|	3	4	3	5	62
gi|524912715|ref|XP_005111198.1|	3	4	3	3	62
gi|524912919|ref|XP_005111295.1|	3	6	3	4	67
gi|524913016|ref|XP_005111342.1|	3	14	3	3	83
gi|524913109|ref|XP_005111388.1|	3	16	3	3	59
gi|524914862|ref|XP_005112240.1|	3	13	3	4	70
gi|524915405|ref|XP_005112503.1|	3	21	3	24	598
gi|524915787|ref|XP_005112684.1|	3	26	3	3	101
gi|524915990|ref|XP_005112779.1|	3	11	3	5	96
gi|524916974|ref|XP_005113254.1|	3	22	3	6	126
gi|54020890|ref|NP_001005689.1|	3	11	3	12	146
gi|555930987|emb|CDJ08105.1|	3	35	3	4	82
gi|556094374|gb|ESO83027.1|	3	8	3	4	72
gi|556098443|gb|ESO87095.1|	3	8	3	4	78
gi|556101526|gb|ESO90178.1|	3	9	3	8	149
gi|556106043|gb|ESO94695.1|	3	9	3	7	111
gi|556106190|gb|ESO94842.1|	3	20	5	8	119
gi|556109149|gb|ESO97801.1|	3	16	4	30	460
gi|556110831|gb|ESO99482.1|	3	23	3	21	264
gi|556113071|gb|ESP01723.1|	3	9	3	3	51
gi|556114539|gb|ESP03191.1|	3	3	3	6	88
gi|556116375|gb|ESP05027.1|	3	8	3	8	110
gi|556116442|gb|ESP05094.1|	3	32	3	20	958
gi|583994656|ref|XP_006792469.1|	3	28	3	6	74
gi|585193165|ref|XP_006747803.1|	3	7	3	8	158
gi|608096069|gb|EZG34785.1|	3	17	3	5	102
gi|646720608|gb|KDR22269.1|	3	30	3	6	165
gi|93209532|gb|ABF00124.1|	3	21	3	29	402
gi|115111533|gb|ABI84147.1|	2	6	2	4	77
gi|119568248|gb|EAW47863.1|	2	7	2	2	36
gi|126697398|gb|ABO26656.1|	2	10	2	2	50
gi|149032103|gb|EDL87015.1|	2	11	3	10	185
gi|149627356|ref|XP_001508693.1|	2	14	2	5	115
gi|150251390|gb|ABR68007.1|	2	13	2	4	66

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|206597719|gb|ACI15835.1|	2	22	2	2	67
gi|2073142|dbj|BAA19861.1|	2	10	2	7	134
gi|2073142|dbj|BAA19861.1|	2	11	2	4	81
gi|2073146|dbj|BAA19863.1|	2	27	2	15	502
gi|210076789|gb|ACJ06748.1|	2	16	2	6	142
gi|218683625|gb|ACL00841.1|	2	9	2	2	52
gi|225717556|gb|ACO14624.1|	2	11	2	3	70
gi|260781473|ref|XP_002585834.1|	2	14	2	3	52
gi|325296803|ref|NP_001191642.1|	2	18	2	3	73
gi|325296883|ref|NP_001191473.1|	2	10	2	2	53
gi|325296989|ref|NP_001191508.1|	2	10	2	3	54
gi|325297146|ref|NP_001191582.1|	2	17	6	10	209
gi|333449383|gb|AEF33378.1|	2	17	2	3	104
gi|346987844|gb|AEO51768.1|	2	8	4	7	161
gi|348577231|ref|XP_003474388.1|	2	10	2	2	59
gi|38043955|emb|CAD83837.1|	2	8	2	4	52
gi|38043955|emb|CAD83837.1|	2	11	2	11	132
gi|390349570|ref|XP_782388.3|	2	67	9	33	583
gi|402227989|gb|AFQ35982.1|	2	8	2	2	47
gi|405964935|gb|EKC30373.1|	2	3	2	2	39
gi|405965644|gb|EKC31006.1|	2	9	2	4	58
gi|405967885|gb|EKC33004.1|	2	10	2	2	66
gi|405969794|gb|EKC34745.1|	2	15	2	5	130
gi|405969853|gb|EKC34799.1|	2	4	2	2	46
gi|405972489|gb|EKC37255.1|	2	10	2	2	81
gi|405974628|gb|EKC39257.1|	2	5	2	9	247
gi|405975533|gb|EKC40092.1|	2	10	2	2	92
gi|405978690|gb|EKC43060.1|	2	4	2	2	78
gi|406679464|gb|AFS50734.1|	2	14	2	5	94
gi|42559558|sp|O97192.1|TPM_HELAS	2	19	7	15	259
gi|42560362|gb|AAS20336.1|	2	31	5	209	4192
gi|46276960|gb|AAS86676.1|	2	7	2	3	62
gi|524863982|ref|XP_005088830.1|	2	25	2	3	52
gi|524864066|ref|XP_005088871.1|	2	14	2	5	59
gi|524864582|ref|XP_005089126.1|	2	9	2	2	50
gi|524864622|ref|XP_005089146.1|	2	6	2	8	226
gi|524865813|ref|XP_005089733.1|	2	10	2	12	155
gi|524865891|ref|XP_005089771.1|	2	7	2	6	107
gi|524866978|ref|XP_005090304.1|	2	23	2	2	57
gi|524867549|ref|XP_005090585.1|	2	10	2	2	60
gi|524868395|ref|XP_005091000.1|	2	6	2	2	53

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524868787|ref|XP_005091193.1|	2	7	2	2	62
gi|524870722|ref|XP_005092143.1|	2	10	2	4	82
gi|524871066|ref|XP_005092309.1|	2	5	2	2	44
gi|524872436|ref|XP_005092980.1|	2	5	2	4	111
gi|524872436|ref|XP_005092980.1|	2	6	2	2	126
gi|524872672|ref|XP_005093096.1|	2	19	2	7	100
gi|524872801|ref|XP_005093157.1|	2	8	2	6	83
gi|524872805|ref|XP_005093159.1|	2	9	2	2	56
gi|524873087|ref|XP_005093297.1|	2	6	2	14	203
gi|524873464|ref|XP_005093478.1|	2	13	2	2	31
gi|524873886|ref|XP_005093685.1|	2	3	2	3	88
gi|524873894|ref|XP_005093689.1|	2	28	2	4	71
gi|524874093|ref|XP_005093787.1|	2	12	2	3	59
gi|524874148|ref|XP_005093814.1|	2	11	2	2	54
gi|524874148|ref|XP_005093814.1|	2	7	2	13	125
gi|524874190|ref|XP_005093834.1|	2	6	2	2	62
gi|524874657|ref|XP_005094063.1|	2	2	2	8	192
gi|524874940|ref|XP_005094202.1|	2	12	2	7	85
gi|524875029|ref|XP_005094246.1|	2	15	2	2	86
gi|524875093|ref|XP_005094277.1|	2	21	2	4	111
gi|524875173|ref|XP_005094317.1|	2	4	2	2	60
gi|524875239|ref|XP_005094350.1|	2	5	2	4	82
gi|524875329|ref|XP_005094395.1|	2	26	2	4	128
gi|524876118|ref|XP_005094778.1|	2	10	2	4	111
gi|524876523|ref|XP_005094977.1|	2	22	2	4	68
gi|524876523|ref|XP_005094977.1|	2	31	2	11	183
gi|524876525|ref|XP_005094978.1|	2	14	3	26	467
gi|524876525|ref|XP_005094978.1|	2	51	3	9	272
gi|524876560|ref|XP_005094995.1|	2	3	2	2	87
gi|524876618|ref|XP_005095021.1|	2	8	2	3	95
gi|524876864|ref|XP_005095141.1|	2	6	2	5	126
gi|524877098|ref|XP_005095256.1|	2	12	2	4	45
gi|524877853|ref|XP_005095623.1|	2	10	3	11	291
gi|524878165|ref|XP_005095772.1|	2	5	2	3	97
gi|524879594|ref|XP_005096471.1|	2	6	2	3	43
gi|524879801|ref|XP_005096572.1|	2	6	2	4	75
gi|524881413|ref|XP_005097358.1|	2	8	2	2	44
gi|524882436|ref|XP_005097862.1|	2	5	2	9	207
gi|524882460|ref|XP_005097873.1|	2	19	2	3	72
gi|524882968|ref|XP_005098119.1|	2	29	2	6	148
gi|524883188|ref|XP_005098227.1|	2	12	2	3	89

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524883558|ref|XP_005098406.1|	2	5	2	2	60
gi|524884444|ref|XP_005098842.1|	2	10	2	5	79
gi|524884503|ref|XP_005098871.1|	2	15	2	2	58
gi|524884830|ref|XP_005099031.1|	2	6	2	2	86
gi|524885197|ref|XP_005099209.1|	2	8	2	2	44
gi|524885355|ref|XP_005099286.1|	2	6	2	3	64
gi|524885703|ref|XP_005099456.1|	2	3	2	6	59
gi|524885955|ref|XP_005099578.1|	2	5	2	3	55
gi|524886261|ref|XP_005099727.1|	2	3	2	2	85
gi|524886263|ref|XP_005099728.1|	2	11	2	5	79
gi|524886662|ref|XP_005099924.1|	2	6	2	2	48
gi|524886707|ref|XP_005099946.1|	2	5	2	7	151
gi|524887384|ref|XP_005100278.1|	2	3	2	2	40
gi|524887465|ref|XP_005100318.1|	2	30	2	4	103
gi|524887465|ref|XP_005100318.1|	2	18	2	4	117
gi|524887498|ref|XP_005100334.1|	2	4	2	2	57
gi|524887580|ref|XP_005100374.1|	2	5	2	3	72
gi|524889104|ref|XP_005101113.1|	2	5	2	2	65
gi|524889992|ref|XP_005101546.1|	2	10	2	2	75
gi|524890376|ref|XP_005101735.1|	2	7	2	2	54
gi|524890561|ref|XP_005101826.1|	2	5	2	5	171
gi|524890653|ref|XP_005101870.1|	2	11	2	3	98
gi|524890724|ref|XP_005101905.1|	2	60	18	38	463
gi|524890794|ref|XP_005101939.1|	2	12	2	2	85
gi|524891426|ref|XP_005102246.1|	2	6	2	2	58
gi|524891574|ref|XP_005102320.1|	2	12	2	2	77
gi|524892285|ref|XP_005102672.1|	2	5	2	3	72
gi|524893098|ref|XP_005103069.1|	2	17	2	8	208
gi|524893405|ref|XP_005103218.1|	2	8	2	3	101
gi|524893572|ref|XP_005103295.1|	2	5	2	5	91
gi|524893741|ref|XP_005103379.1|	2	15	2	2	56
gi|524893900|ref|XP_005103458.1|	2	4	2	3	62
gi|524893900|ref|XP_005103458.1|	2	4	2	3	65
gi|524894650|ref|XP_005103820.1|	2	9	2	3	94
gi|524894780|ref|XP_005103882.1|	2	4	2	2	55
gi|524895699|ref|XP_005104331.1|	2	4	2	2	43
gi|524896048|ref|XP_005104503.1|	2	12	2	3	48
gi|524896189|ref|XP_005104572.1|	2	5	2	2	49
gi|524896597|ref|XP_005104771.1|	2	11	2	2	74
gi|524896719|ref|XP_005104831.1|	2	25	2	5	139
gi|524896719|ref|XP_005104831.1|	2	15	2	2	44

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524897390|ref|XP_005105156.1|	2	23	2	2	40
gi|524898060|ref|XP_005105484.1|	2	7	2	2	71
gi|524898265|ref|XP_005105584.1|	2	7	2	2	52
gi|524898815|ref|XP_005105852.1|	2	34	2	4	182
gi|524899012|ref|XP_005105947.1|	2	8	2	2	60
gi|524899021|ref|XP_005105951.1|	2	8	2	4	59
gi|524899091|ref|XP_005105986.1|	2	18	2	2	44
gi|524899882|ref|XP_005106367.1|	2	7	2	2	64
gi|524900256|ref|XP_005106549.1|	2	17	2	4	40
gi|524900270|ref|XP_005106556.1|	2	15	2	20	169
gi|524900390|ref|XP_005106615.1|	2	9	4	10	205
gi|524900699|ref|XP_005106768.1|	2	4	2	7	105
gi|524901519|ref|XP_005107166.1|	2	11	2	3	63
gi|524901804|ref|XP_005107306.1|	2	14	2	5	101
gi|524902448|ref|XP_005107474.1|	2	7	2	3	67
gi|524902931|ref|XP_005107685.1|	2	7	2	2	46
gi|524903201|ref|XP_005107810.1|	2	1	2	2	68
gi|524903262|ref|XP_005107836.1|	2	7	2	2	44
gi|524903293|ref|XP_005107850.1|	2	8	2	7	154
gi|524906293|ref|XP_005108469.1|	2	34	2	4	136
gi|524906293|ref|XP_005108469.1|	2	11	2	7	172
gi|524906745|ref|XP_005108595.1|	2	14	2	2	45
gi|524906745|ref|XP_005108595.1|	2	13	2	2	44
gi|524906745|ref|XP_005108595.1|	2	7	2	2	59
gi|524906932|ref|XP_005108623.1|	2	2	2	3	65
gi|524907576|ref|XP_005108904.1|	2	21	2	21	332
gi|524907590|ref|XP_005108911.1|	2	9	2	2	42
gi|524908282|ref|XP_005109244.1|	2	18	2	4	160
gi|524908284|ref|XP_005109245.1|	2	6	2	2	37
gi|524909838|ref|XP_005109815.1|	2	13	2	3	91
gi|524909848|ref|XP_005109820.1|	2	7	2	2	52
gi|524909938|ref|XP_005109861.1|	2	5	2	2	42
gi|524910284|ref|XP_005110023.1|	2	5	2	3	52
gi|524910373|ref|XP_005110066.1|	2	3	2	3	54
gi|524910383|ref|XP_005110071.1|	2	11	2	3	111
gi|524910825|ref|XP_005110286.1|	2	11	2	8	116
gi|524911477|ref|XP_005110601.1|	2	31	2	2	59
gi|524911866|ref|XP_005110793.1|	2	15	2	3	63
gi|524911952|ref|XP_005110833.1|	2	13	2	2	59
gi|524913773|ref|XP_005111711.1|	2	6	2	2	65
gi|524913818|ref|XP_005111733.1|	2	5	2	6	143

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524914187|ref|XP_005111912.1|	2	6	2	3	94
gi|524915372|ref|XP_005112487.1|	2	10	2	2	42
gi|524916029|ref|XP_005112797.1|	2	7	2	2	89
gi|524916381|ref|XP_005112969.1|	2	9	2	2	42
gi|524916499|ref|XP_005113026.1|	2	7	2	9	226
gi|524916593|ref|XP_005113070.1|	2	9	2	2	44
gi|524917273|ref|XP_005113396.1|	2	11	2	4	152
gi|526117361|ref|NP_001267749.1|	2	13	2	137	1145
gi|556093891|gb|ESO82544.1|	2	9	2	2	42
gi|556095216|gb|ESO83869.1|	2	8	2	5	180
gi|556097528|gb|ESO86180.1|	2	7	2	4	90
gi|556098160|gb|ESO86812.1|	2	7	2	11	226
gi|556100167|gb|ESO88819.1|	2	17	2	7	84
gi|556101117|gb|ESO89769.1|	2	9	2	2	49
gi|556106190|gb|ESO94842.1|	2	8	2	2	74
gi|556106555|gb|ESO95207.1|	2	3	2	2	64
gi|556107537|gb|ESO96189.1|	2	4	2	3	111
gi|556108606|gb|ESO97258.1|	2	17	2	4	82
gi|556111890|gb|ESP00542.1|	2	12	2	2	56
gi|556113049|gb|ESP01701.1|	2	5	2	2	56
gi|556114871|gb|ESP03523.1|	2	23	2	3	73
gi|556115860|gb|ESP04512.1|	2	4	2	5	64
gi|556116519|gb|ESP05171.1|	2	11	2	2	81
gi|556729406|ref|XP_005960416.1|	2	9	2	5	135
gi|568291352|gb|ETN79106.1|	2	13	2	38	164
gi|589107123|ref|XP_006965584.1|	2	18	2	2	64
gi|597867153|gb|EYC16545.1|	2	24	3	26	693
gi|68272049|gb|AAY89300.1|	2	12	2	6	108
gi|773571|emb|CAA60122.1|	2	20	2	3	110
gi|773571|emb|CAA60122.1|	2	7	3	4	130
gi|85753308|gb|AAG49533.3|	2	9	2	13	185
gi|994895|gb|AAB34494.1|	2	10	2	2	81
gi|170596696|ref|XP_001902862.1|	1	17	2	11	257
gi|241151391|ref|XP_002406661.1|	1	14	5	8	209
gi|2827905|gb|AAB99911.1|	1	4	2	5	150
gi|405952673|gb|EKC20457.1|	1	11	2	3	57
gi|51105058|gb|AAT97089.1|	1	11	2	2	75
gi|524876525|ref|XP_005094978.1|	1	13	2	3	58
gi|524885593|ref|XP_005099402.1|	1	15	3	10	349
gi|524906745|ref|XP_005108595.1|	1	7	2	3	74
gi|537705579|dbj|BAN81895.1|	1	30	4	28	503

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|642099995|emb|CDQ77143.1|	1	25	3	3	112

8.3. Table 2. The list of proteins obtained from 1D-SDS-PAGE of *A. vulgaris* mantel

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524911570|ref|XP_005110646.1|	150	48	152	489	9983
gi|318609972|dbj|BAJ61596.1|	112	82	117	3270	88132
gi|524888508|ref|XP_005100824.1|	92	63	116	847	16625
gi|524907580|ref|XP_005108906.1|	76	43	77	167	3386
gi|524896814|ref|XP_005104878.1|	69	50	101	292	6076
gi|346987840|gb|AEO51766.1|	63	44	93	489	10429
gi|524911094|ref|XP_005110416.1|	59	40	59	126	2582
gi|127773|sp|P24733.1|MYS_ARGIR	50	69	176	1022	23056
gi|524907580|ref|XP_005108906.1|	49	28	49	157	4164
gi|556096138|gb|ESO84790.1|	46	74	46	352	7897
gi|524903201|ref|XP_005107810.1|	45	37	70	164	4119
gi|6682319|emb|CAB64662.1|	45	67	68	713	17930
gi|124233|sp|P22488.2|IFEA_HELAS	43	77	59	847	16905
gi|325297144|ref|NP_001191581.1|	43	62	46	159	3236
gi|524896814|ref|XP_005104878.1|	43	38	44	92	1511
gi|524916601|ref|XP_005113074.1|	40	42	45	118	2572
gi|524866478|ref|XP_005090060.1|	37	76	37	455	8975
gi|556110866|gb|ESO99517.1|	37	59	90	456	10538
gi|346987840|gb|AEO51766.1|	36	19	38	105	2142
gi|468860736|gb|AGH32328.1|	35	49	35	119	2372
gi|524868522|ref|XP_005091063.1|	35	42	40	103	2396
gi|556106555|gb|ESO95207.1|	35	43	52	106	1843
gi|346987840|gb|AEO51766.1|	33	44	33	183	3319
gi|524875601|ref|XP_005094526.1|	33	71	43	92	1886
gi|126697420|gb|ABO26667.1|	32	59	32	93	1806
gi|524867613|ref|XP_005090615.1|	32	41	32	42	1015
gi|524906491|ref|XP_005108543.1|	32	51	33	101	1838
gi|220067263|gb|ACL79834.1|	31	86	31	293	6253
gi|472824675|dbj|BAN09030.1|	31	62	31	255	4213
gi|524896814|ref|XP_005104878.1|	31	62	32	118	2214
gi|524894502|ref|XP_005103748.1|	30	57	30	227	5014
gi|556096092|gb|ESO84744.1|	30	46	30	136	2794
gi|150251392|gb|ABR68008.1|	29	51	29	151	3179
gi|218683627|gb|ACL00842.1|	29	65	29	117	3208
gi|556113526|gb|ESP02178.1|	29	62	29	75	1397
gi|73695954|gb|AAZ80783.1|	29	66	29	752	17215
gi|524881119|ref|XP_005097215.1|	28	56	28	90	1496
gi|556106555|gb|ESO95207.1|	28	49	43	113	2377

gi|127773|sp|P24733.1|MYS_ARGIR	27	50	92	659	19586
Hit in the NCBI nr database	**UP**	**PC (%)**	**NM**	**NS**	**PS**
gi|325296743|ref|NP_001191607.1|	27	67	34	104	1852
gi|524869485|ref|XP_005091535.1|	27	57	27	131	2189
gi|524903383|ref|XP_005107892.1|	27	78	28	362	10496
gi|524916025|ref|XP_005112795.1|	27	77	28	309	8062
gi|405960426|gb|EKC26351.1|	26	69	26	121	2932
gi|524888358|ref|XP_005100750.1|	26	56	27	103	2155
gi|524895745|ref|XP_005104354.1|	26	63	35	514	13237
gi|524896814|ref|XP_005104878.1|	26	27	26	73	1752
gi|296837083|gb|ADH59421.1|	25	72	35	441	9017
gi|305689815|gb|ADM64337.1|	25	69	25	341	9279
gi|524887592|ref|XP_005100380.1|	25	72	25	93	1941
gi|556113946|gb|ESP02598.1|	25	46	25	117	2548
gi|385145402|emb|CCG28026.1|	24	80	33	2612	54460
gi|42559558|sp|O97192.1|TPM_HELAS	24	83	41	319	6065
gi|524886261|ref|XP_005099727.1|	24	49	43	322	7707
gi|74912853|sp|Q6QUW1.1|RGP51_LYMST	24	70	31	199	3590
gi|241273677|ref|XP_002406620.1|	23	66	26	73	1343
gi|524881811|ref|XP_005097555.1|	23	73	23	178	4157
gi|556097340|gb|ESO85992.1|	23	47	23	88	2392
gi|556104876|gb|ESO93528.1|	23	76	60	255	5495
gi|556106555|gb|ESO95207.1|	23	35	36	66	1146
gi|585719304|ref|XP_006826038.1|	23	42	69	165	3915
gi|405950185|gb|EKC18187.1|	22	9	22	31	518
gi|524881729|ref|XP_005097514.1|	22	51	22	37	804
gi|524883288|ref|XP_005098277.1|	22	49	22	60	1107
gi|524885898|ref|XP_005099550.1|	22	49	22	30	493
gi|524908996|ref|XP_005109585.1|	22	56	22	86	1619
gi|524913881|ref|XP_005111763.1|	22	47	26	42	966
gi|150251390|gb|ABR68007.1|	21	51	21	130	2865
gi|524884737|ref|XP_005098987.1|	21	57	22	70	1211
gi|524906932|ref|XP_005108623.1|	21	36	27	37	609
gi|556104876|gb|ESO93528.1|	21	66	52	235	5405
Unknown	21	73	21	218	4285
gi|150251390|gb|ABR68007.1|	20	54	20	114	2371
gi|296190822|ref|XP_002743358.1|	20	39	21	52	1118
gi|321472487|gb|EFX83457.1|	20	34	20	55	985
gi|524867613|ref|XP_005090615.1|	20	15	20	35	993
gi|524882748|ref|XP_005098012.1|	20	70	39	196	3973
gi|524885581|ref|XP_005099396.1|	20	71	20	106	1541
gi|346987844|gb|AEO51768.1|	16	36	17	86	1625
gi|524886263|ref|XP_005099728.1|	20	31	20	57	1002

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524892329|ref|XP_005102693.1|	20	34	20	24	564
gi|524893765|ref|XP_005103391.1|	20	42	20	24	403
gi|524894608|ref|XP_005103800.1|	20	36	20	27	391
gi|524896519|ref|XP_005104733.1|	20	75	23	95	1222
gi|524896638|ref|XP_005104791.1|	20	41	20	61	1294
gi|524913881|ref|XP_005111763.1|	20	47	22	81	2223
gi|556097683|gb|ESO86335.1|	20	56	20	77	1512
gi|119630315|gb|EAX09910.1|	19	55	21	93	2079
gi|524867573|ref|XP_005090597.1|	19	66	19	68	1544
gi|524868093|ref|XP_005090853.1|	19	28	19	28	612
gi|524870848|ref|XP_005092204.1|	19	30	19	21	342
gi|524872922|ref|XP_005093216.1|	19	43	20	75	1573
gi|524890683|ref|XP_005101885.1|	19	20	19	25	454
gi|524909114|ref|XP_005109609.1|	19	92	23	592	11227
gi|524916452|ref|XP_005113003.1|	19	55	19	139	1903
gi|526117385|ref|NP_001267755.1|	19	73	20	133	3576
gi|556101476|gb|ESO90128.1|	19	39	26	30	540
gi|773571|emb|CAA60122.1|	19	45	26	36	744
gi|126697362|gb|ABO26638.1|	18	83	18	181	3463
gi|325297052|ref|NP_001191530.1|	18	82	29	351	7647
gi|443692429|gb|ELT94023.1|	18	64	18	112	2336
gi|524878973|ref|XP_005096166.1|	18	68	18	94	2172
gi|524884921|ref|XP_005099073.1|	18	56	18	807	18649
gi|524885520|ref|XP_005099366.1|	18	82	18	188	2297
gi|524889526|ref|XP_005101318.1|	18	44	18	48	1113
gi|524892924|ref|XP_005102985.1|	18	58	18	186	4977
gi|524912440|ref|XP_005111068.1|	18	64	18	307	6517
gi|524916601|ref|XP_005113074.1|	18	25	20	85	2449
gi|526117385|ref|NP_001267755.1|	18	79	19	353	10785
gi|556097528|gb|ESO86180.1|	18	61	18	49	1345
gi|556105302|gb|ESO93954.1|	18	79	21	296	5016
gi|524871134|ref|XP_005092343.1|	17	44	17	58	1385
gi|524880064|ref|XP_005096700.1|	17	47	17	30	620
gi|524886817|ref|XP_005100000.1|	17	43	17	80	1717
gi|524898799|ref|XP_005105845.1|	17	13	17	35	996
gi|524902448|ref|XP_005107474.1|	17	47	17	124	2385
gi|524912715|ref|XP_005111198.1|	17	45	30	66	1607
gi|54020890|ref|NP_001005689.1|	17	66	17	94	1679
gi|241151391|ref|XP_002406661.1|	16	75	23	122	1778
gi|325297040|ref|NP_001191523.1|	16	71	18	71	1048
gi|524907576|ref|XP_005108904.1|	12	64	12	147	3211
gi|524913044|ref|XP_005111356.1|	12	47	12	39	804

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524894602|ref|XP_005103797.1|	12	23	12	14	306
gi|524894780|ref|XP_005103882.1|	12	28	12	27	493
gi|524871930|ref|XP_005092735.1|	16	22	16	33	692
gi|524872578|ref|XP_005093050.1|	16	37	32	62	1549
gi|524874065|ref|XP_005093774.1|	16	69	16	26	674
gi|524877675|ref|XP_005095535.1|	16	26	16	48	1145
gi|524879388|ref|XP_005096370.1|	16	56	16	56	1570
gi|524882748|ref|XP_005098012.1|	16	62	37	152	3098
gi|524885854|ref|XP_005099530.1|	16	73	16	236	3887
gi|524896250|ref|XP_005104602.1|	16	54	16	44	567
gi|524907151|ref|XP_005108701.1|	16	38	16	33	669
gi|524908412|ref|XP_005109303.1|	16	36	16	20	307
gi|524909189|ref|XP_005109645.1|	16	26	16	23	369
gi|524911954|ref|XP_005110834.1|	16	43	19	61	1188
gi|524915110|ref|XP_005112361.1|	16	39	16	45	698
gi|524916507|ref|XP_005113030.1|	16	61	19	65	1077
gi|54020890|ref|NP_001005689.1|	16	59	16	101	2534
gi|555688717|gb|ESN91949.1|	16	79	20	305	6418
gi|556100505|gb|ESO89157.1|	16	42	17	42	758
gi|556101476|gb|ESO90128.1|	16	29	18	46	752
gi|585193165|ref|XP_006747803.1|	16	46	18	99	1856
gi|147905746|ref|NP_001080566.1|	15	84	29	182	3912
gi|405953142|gb|EKC20858.1|	15	47	15	43	711
gi|524866272|ref|XP_005089958.1|	15	55	15	22	319
gi|524867757|ref|XP_005090686.1|	15	29	29	53	1203
gi|524868057|ref|XP_005090835.1|	15	12	15	20	329
gi|524877903|ref|XP_005095647.1|	15	8	15	31	707
gi|524883990|ref|XP_005098620.1|	15	67	15	88	2385
gi|524886695|ref|XP_005099940.1|	15	59	15	39	846
gi|524888067|ref|XP_005100610.1|	15	20	15	44	764
gi|524891424|ref|XP_005102245.1|	15	34	15	29	660
gi|524895953|ref|XP_005104456.1|	15	15	15	22	293
gi|524897671|ref|XP_005105294.1|	15	33	15	56	1426
gi|524897999|ref|XP_005105454.1|	15	54	16	63	1580
gi|524900390|ref|XP_005106615.1|	15	37	15	51	1204
gi|524900673|ref|XP_005106755.1|	15	66	15	96	1921
gi|524906359|ref|XP_005108493.1|	15	37	15	26	540
gi|524909114|ref|XP_005109609.1|	15	89	19	426	7452
gi|526117405|ref|NP_001267760.1|	15	82	15	216	3211
gi|556101047|gb|ESO89699.1|	15	17	22	28	568
gi|556110891|gb|ESO99542.1|	15	68	15	105	1857
gi|1174755|sp|P43689.1|TPM2_BIOGL	14	75	28	84	1265

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|150251390|gb|ABR68007.1|	14	42	15	61	1945
gi|270268071|gb|ACZ65578.1|	14	45	14	30	654
gi|346987844|gb|AEO51768.1|	14	14	15	23	405
gi|405978809|gb|EKC43171.1|	14	32	14	26	634
gi|51105058|gb|AAT97089.1|	14	62	19	39	841
gi|513137512|gb|AGO20698.1|	14	40	14	42	1115
gi|524867763|ref|XP_005090689.1|	14	16	29	34	785
gi|524873169|ref|XP_005093336.1|	14	39	14	24	475
gi|524873538|ref|XP_005093514.1|	14	34	14	21	503
gi|524874194|ref|XP_005093836.1|	14	49	14	29	497
gi|524874655|ref|XP_005094062.1|	14	14	14	53	1275
gi|524875149|ref|XP_005094305.1|	14	27	14	24	431
gi|524875601|ref|XP_005094526.1|	14	42	24	34	770
gi|524876391|ref|XP_005094913.1|	14	61	14	49	775
gi|524878504|ref|XP_005095936.1|	14	84	16	45	751
gi|524879388|ref|XP_005096370.1|	14	50	14	21	435
gi|524879616|ref|XP_005096482.1|	14	63	14	35	883
gi|524880840|ref|XP_005097078.1|	14	52	14	24	381
gi|524883140|ref|XP_005098203.1|	14	31	14	17	408
gi|524883762|ref|XP_005098507.1|	14	55	14	39	796
gi|524884451|ref|XP_005098845.1|	14	45	14	27	420
gi|524886299|ref|XP_005099746.1|	14	64	14	29	520
gi|524892173|ref|XP_005102616.1|	14	49	14	40	635
gi|524892201|ref|XP_005102630.1|	14	20	15	16	311
gi|524892341|ref|XP_005102699.1|	14	72	17	206	3140
gi|524894013|ref|XP_005103514.1|	14	43	16	50	904
gi|524896519|ref|XP_005104733.1|	14	63	18	100	1103
gi|524903522|ref|XP_005107960.1|	14	46	14	75	1603
gi|524907533|ref|XP_005108883.1|	14	46	14	49	861
gi|524911418|ref|XP_005110573.1|	14	21	15	24	361
gi|556101117|gb|ESO89769.1|	14	62	14	43	542
gi|556113049|gb|ESP01701.1|	14	36	14	27	652
gi|556113964|gb|ESP02616.1|	14	31	14	33	602
gi|556114539|gb|ESP03191.1|	14	32	14	29	463
gi|620958561|ref|XP_007664363.1|	14	76	14	61	1074
gi|2073142|dbj|BAA19861.1|	13	62	13	278	4823
gi|325297068|ref|NP_001191538.1|	13	27	14	31	534
gi|325297094|ref|NP_001191555.1|	13	49	13	28	680
gi|42559558|sp|O97192.1|TPM_HELAS	13	67	25	81	1441
gi|464524|sp|Q05974.1|RAB1A_LYMST	13	74	13	62	1166
gi|521024335|gb|EPQ06123.1|	13	50	13	153	2421
gi|524913109|ref|XP_005111388.1|	12	68	12	62	1080

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524864622|ref|XP_005089146.1|	13	36	13	36	939
gi|524871621|ref|XP_005092582.1|	13	53	22	93	1930
gi|524878388|ref|XP_005095881.1|	13	32	13	27	461
gi|524881606|ref|XP_005097453.1|	13	51	13	33	740
gi|524882422|ref|XP_005097855.1|	13	53	13	66	1480
gi|524882436|ref|XP_005097862.1|	13	30	13	26	394
gi|524885248|ref|XP_005099233.1|	13	82	13	27	333
gi|524886261|ref|XP_005099727.1|	13	20	13	23	562
gi|524889482|ref|XP_005101296.1|	13	36	13	48	912
gi|524890176|ref|XP_005101637.1|	13	35	13	22	385
gi|524891426|ref|XP_005102246.1|	13	52	13	34	658
gi|524891679|ref|XP_005102372.1|	13	47	13	47	1023
gi|524897999|ref|XP_005105454.1|	13	45	14	43	1484
gi|524898259|ref|XP_005105581.1|	13	51	13	60	667
gi|524917285|ref|XP_005113402.1|	13	52	13	42	482
gi|556095158|gb|ESO83811.1|	13	42	16	47	1133
gi|556098588|gb|ESO87240.1|	13	53	13	24	472
gi|556106043|gb|ESO94695.1|	13	38	13	17	314
gi|556106306|gb|ESO94958.1|	13	33	13	22	510
gi|556116376|gb|ESP05028.1|	13	44	15	30	628
gi|126697410|gb|ABO26662.1|	12	56	12	207	5349
gi|32169292|emb|CAD89885.1|	12	39	12	26	232
gi|325296847|ref|NP_001191665.1|	12	27	13	18	400
gi|325296855|ref|NP_001191459.1|	12	27	20	25	638
gi|325296989|ref|NP_001191508.1|	12	65	14	43	562
gi|405960204|gb|EKC26145.1|	12	17	12	18	378
gi|405964935|gb|EKC30373.1|	12	26	12	24	491
gi|443708569|gb|ELU03646.1|	12	50	12	26	614
gi|524865670|ref|XP_005089662.1|	12	86	14	259	5953
gi|524865955|ref|XP_005089803.1|	12	55	13	88	1019
gi|524866304|ref|XP_005089974.1|	12	55	12	34	897
gi|524866556|ref|XP_005090099.1|	12	24	12	24	450
gi|524867514|ref|XP_005090568.1|	12	37	12	25	546
gi|524869199|ref|XP_005091396.1|	12	50	12	30	883
gi|524870028|ref|XP_005091800.1|	12	36	12	27	651
gi|524872674|ref|XP_005093097.1|	12	38	12	17	341
gi|524873087|ref|XP_005093297.1|	12	49	12	35	626
gi|524873600|ref|XP_005093544.1|	12	63	12	49	421
gi|524875771|ref|XP_005094609.1|	12	26	12	30	639
gi|524875985|ref|XP_005094714.1|	12	16	12	31	430
gi|524876605|ref|XP_005095015.1|	12	26	12	21	322
gi|524879594|ref|XP_005096471.1|	12	53	12	28	728

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524880339|ref|XP_005096834.1|	12	45	12	24	333
gi|524881417|ref|XP_005097360.1|	12	53	12	38	777
gi|524881980|ref|XP_005097636.1|	12	40	12	14	285
gi|524882968|ref|XP_005098119.1|	12	31	12	18	417
gi|524884737|ref|XP_005098987.1|	12	39	13	33	689
gi|524885197|ref|XP_005099209.1|	12	48	12	14	233
gi|524885795|ref|XP_005099501.1|	12	67	37	239	5797
gi|524886263|ref|XP_005099728.1|	12	34	12	51	2015
gi|524891584|ref|XP_005102325.1|	12	38	12	17	343
gi|524892179|ref|XP_005102619.1|	12	27	12	20	291
gi|524892285|ref|XP_005102672.1|	12	34	13	27	666
gi|524892819|ref|XP_005102934.1|	12	27	12	32	679
gi|524893900|ref|XP_005103458.1|	12	33	15	41	703
gi|524917241|ref|XP_005113381.1|	12	43	12	139	2324
gi|556095158|gb|ESO83811.1|	12	43	16	38	1143
gi|556095567|gb|ESO84220.1|	12	43	13	31	735
gi|556105117|gb|ESO93769.1|	12	66	16	97	1689
gi|556106555|gb|ESO95207.1|	12	39	35	57	1041
gi|556116427|gb|ESP05079.1|	12	50	15	58	1157
gi|564743500|gb|AHC02701.1|	12	81	33	148	2594
gi|773571|emb|CAA60122.1|	12	43	19	21	313
gi|152003985|gb|ABS19816.1|	11	32	20	53	1552
gi|152032020|gb|ABS28869.1|	11	39	11	56	584
gi|168252891|gb|ACA23131.1|	11	66	11	68	1495
gi|291235073|ref|XP_002737469.1|	11	41	11	26	327
gi|325296875|ref|NP_001191469.1|	11	64	11	27	295
gi|325296883|ref|NP_001191473.1|	11	59	11	26	349
gi|325297040|ref|NP_001191523.1|	11	36	13	19	302
gi|351709298|gb|EHB12217.1|	11	53	11	265	4017
gi|405972472|gb|EKC37239.1|	11	27	12	22	486
gi|42559558|sp|O97192.1|TPM_HELAS	11	53	17	41	700
gi|524864482|ref|XP_005089076.1|	11	32	11	13	266
gi|524865670|ref|XP_005089662.1|	11	76	15	137	2808
gi|524866934|ref|XP_005090282.1|	11	12	12	24	452
gi|524867613|ref|XP_005090615.1|	11	19	11	23	578
gi|524870191|ref|XP_005091881.1|	11	45	11	48	770
gi|524870300|ref|XP_005091935.1|	11	56	11	27	433
gi|524870886|ref|XP_005092223.1|	11	67	11	145	3783
gi|524871621|ref|XP_005092582.1|	11	51	22	62	1330
gi|524871885|ref|XP_005092713.1|	11	35	11	20	276
gi|524871930|ref|XP_005092735.1|	11	23	11	20	479
gi|524874055|ref|XP_005093769.1|	11	19	11	15	221

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524874414|ref|XP_005093942.1|	11	25	11	31	543
gi|524878451|ref|XP_005095912.1|	11	16	11	16	359
gi|524882484|ref|XP_005097885.1|	11	38	12	28	657
gi|524884293|ref|XP_005098768.1|	11	20	11	21	322
gi|524884630|ref|XP_005098934.1|	11	39	11	45	1118
gi|524885193|ref|XP_005099207.1|	11	22	11	23	489
gi|524887384|ref|XP_005100278.1|	11	13	11	25	446
gi|524890066|ref|XP_005101582.1|	11	28	12	23	506
gi|524890120|ref|XP_005101609.1|	11	39	11	17	201
gi|524893793|ref|XP_005103405.1|	11	30	11	27	649
gi|524895031|ref|XP_005104004.1|	11	27	11	33	659
gi|524895893|ref|XP_005104427.1|	11	22	11	14	284
gi|524896814|ref|XP_005104878.1|	11	21	21	40	768
gi|524897983|ref|XP_005105446.1|	11	30	12	24	428
gi|524898231|ref|XP_005105567.1|	11	31	11	16	291
gi|524898324|ref|XP_005105612.1|	11	31	11	12	214
gi|524901897|ref|XP_005107352.1|	11	55	11	28	646
gi|524901909|ref|XP_005107358.1|	11	54	11	30	679
gi|524906745|ref|XP_005108595.1|	11	63	11	41	894
gi|524907590|ref|XP_005108911.1|	11	59	11	21	439
gi|524908062|ref|XP_005109139.1|	11	59	11	21	559
gi|524909004|ref|XP_005109589.1|	11	61	14	77	907
gi|524909038|ref|XP_005109605.1|	11	34	11	32	717
gi|524909838|ref|XP_005109815.1|	11	34	11	20	479
gi|524910373|ref|XP_005110066.1|	11	28	13	20	392
gi|524911396|ref|XP_005110562.1|	11	44	11	27	611
gi|524911954|ref|XP_005110834.1|	11	41	21	146	2639
gi|524913818|ref|XP_005111733.1|	11	51	13	38	668
gi|524914469|ref|XP_005112049.1|	11	46	11	38	490
gi|524914829|ref|XP_005112224.1|	11	41	11	43	1184
gi|524915789|ref|XP_005112685.1|	11	28	11	19	408
gi|556096567|gb|ESO85219.1|	11	20	11	17	215
gi|556116184|gb|ESP04836.1|	11	54	11	27	729
gi|602626279|ref|XP_007420100.1|	11	75	11	92	1724
gi|126697356|gb|ABO26635.1|	10	55	11	39	535
gi|13177628|gb|AAK14899.1|	10	26	10	17	385
gi|150251390|gb|ABR68007.1|	10	38	10	59	1260
gi|150251390|gb|ABR68007.1|	10	25	10	10	270
gi|2073142|dbj|BAA19861.1|	10	40	10	40	486
gi|328933188|gb|AEB70965.1|	10	37	10	23	503
gi|346987840|gb|AEO51766.1|	10	33	10	29	458
gi|346987844|gb|AEO51768.1|	10	32	12	80	1288

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|383859188|ref|XP_003705078.1|	10	12	10	15	407
gi|405952673|gb|EKC20457.1|	10	59	11	58	706
gi|405962978|gb|EKC28598.1|	10	43	10	14	300
gi|405970776|gb|EKC35652.1|	10	71	10	84	1224
gi|443718889|gb|ELU09307.1|	10	21	10	12	124
gi|524864076|ref|XP_005088876.1|	10	16	10	20	526
gi|524866376|ref|XP_005090010.1|	10	29	10	26	917
gi|524866912|ref|XP_005090271.1|	10	36	10	31	493
gi|524866934|ref|XP_005090282.1|	10	14	11	12	402
gi|524870722|ref|XP_005092143.1|	10	51	10	40	914
gi|524874657|ref|XP_005094063.1|	10	11	10	33	853
gi|524874769|ref|XP_005094119.1|	10	24	10	15	255
gi|524875291|ref|XP_005094376.1|	10	16	10	16	297
gi|524875728|ref|XP_005094588.1|	10	47	10	27	259
gi|524876046|ref|XP_005094743.1|	10	19	10	21	306
gi|524876098|ref|XP_005094768.1|	10	26	10	20	365
gi|524876712|ref|XP_005095067.1|	10	18	10	12	261
gi|524878458|ref|XP_005095915.1|	10	63	10	43	907
gi|524881984|ref|XP_005097638.1|	10	75	10	83	2209
gi|524882420|ref|XP_005097854.1|	10	29	14	21	382
gi|524882717|ref|XP_005097997.1|	10	38	10	31	636
gi|524883534|ref|XP_005098394.1|	10	57	10	31	476
gi|524884406|ref|XP_005098823.1|	10	30	10	15	116
gi|524885703|ref|XP_005099456.1|	10	23	10	17	358
gi|524889104|ref|XP_005101113.1|	10	30	10	19	348
gi|524892967|ref|XP_005103005.1|	10	65	10	172	2717
gi|524893098|ref|XP_005103069.1|	10	83	10	65	1260
gi|524895406|ref|XP_005104186.1|	10	8	10	17	268
gi|524895699|ref|XP_005104331.1|	10	22	10	17	166
gi|524895953|ref|XP_005104456.1|	10	13	10	13	213
gi|524896874|ref|XP_005104908.1|	10	51	10	29	215
gi|524898060|ref|XP_005105484.1|	10	27	10	16	376
gi|524898203|ref|XP_005105554.1|	10	15	10	10	140
gi|524899582|ref|XP_005106220.1|	10	45	10	110	2701
gi|524900516|ref|XP_005106678.1|	10	58	10	186	2701
gi|524900699|ref|XP_005106768.1|	10	36	10	19	413
gi|524901536|ref|XP_005107174.1|	10	28	12	45	878
gi|524908509|ref|XP_005109349.1|	10	55	10	53	745
gi|524909848|ref|XP_005109820.1|	10	49	10	36	455
gi|524910025|ref|XP_005109900.1|	10	89	10	38	982
gi|524910869|ref|XP_005110308.1|	10	36	10	14	287
gi|556095567|gb|ESO84220.1|	10	28	11	28	653

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|556103805|gb|ESO92457.1|	10	48	10	29	691
gi|556107537|gb|ESO96189.1|	10	22	10	24	620
gi|556113071|gb|ESP01723.1|	10	30	10	16	354
gi|556116427|gb|ESP05079.1|	10	43	13	36	696
gi|556116519|gb|ESP05171.1|	10	53	10	34	597
gi|560896664|ref|XP_006174798.1|	10	28	10	20	450
gi|994895|gb|AAB34494.1|	10	40	12	26	557
gi|150251390|gb|ABR68007.1|	9	30	10	25	689
gi|2073142|dbj|BAA19861.1|	9	70	9	126	2816
gi|226478902|emb|CAX72946.1|	9	75	22	234	5014
gi|260821662|ref|XP_002606151.1|	9	59	21	110	2702
gi|325297146|ref|NP_001191582.1|	9	39	9	16	379
gi|346987840|gb|AEO51766.1|	9	32	12	56	1360
gi|405951515|gb|EKC19421.1|	9	35	9	15	414
gi|405959696|gb|EKC25703.1|	9	29	9	14	265
gi|405965813|gb|EKC31167.1|	9	45	9	22	378
gi|405966469|gb|EKC31748.1|	9	45	9	13	385
gi|405970946|gb|EKC35808.1|	9	59	9	17	455
gi|405974628|gb|EKC39257.1|	9	32	9	52	1144
gi|51105064|gb|AAT97092.1|	9	33	9	12	242
gi|524865140|ref|XP_005089402.1|	9	52	9	39	776
gi|524866272|ref|XP_005089958.1|	9	35	9	21	426
gi|524868977|ref|XP_005091286.1|	9	38	9	36	453
gi|524869203|ref|XP_005091398.1|	9	17	9	11	137
gi|524869203|ref|XP_005091398.1|	9	13	9	10	159
gi|524869207|ref|XP_005091400.1|	9	10	10	24	358
gi|524871180|ref|XP_005092366.1|	9	46	9	22	384
gi|524871402|ref|XP_005092474.1|	9	30	10	27	729
gi|524872186|ref|XP_005092857.1|	9	16	9	12	362
gi|524872286|ref|XP_005092906.1|	9	21	9	17	265
gi|524876525|ref|XP_005094978.1|	9	45	15	43	803
gi|524876525|ref|XP_005094978.1|	9	60	18	44	1036
gi|524877795|ref|XP_005095595.1|	9	28	9	20	564
gi|524878165|ref|XP_005095772.1|	9	27	9	12	222
gi|524878167|ref|XP_005095773.1|	9	16	9	19	278
gi|524879801|ref|XP_005096572.1|	9	34	9	16	456
gi|524880507|ref|XP_005096916.1|	9	60	9	23	403
gi|524881016|ref|XP_005097164.1|	9	11	9	12	293
gi|524882305|ref|XP_005097797.1|	9	59	9	86	1955
gi|524882414|ref|XP_005097851.1|	9	45	10	15	105
gi|524882774|ref|XP_005098023.1|	9	33	9	15	387
gi|524883140|ref|XP_005098203.1|	9	24	9	14	288

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524884878|ref|XP_005099053.1|	9	39	9	21	385
gi|524885619|ref|XP_005099415.1|	9	51	10	25	526
gi|524885795|ref|XP_005099501.1|	9	70	16	94	1730
gi|524890066|ref|XP_005101582.1|	9	12	10	13	166
gi|524890599|ref|XP_005101843.1|	9	33	9	11	239
gi|524892644|ref|XP_005102848.1|	9	62	9	52	1179
gi|524893673|ref|XP_005103345.1|	9	18	9	9	220
gi|524895342|ref|XP_005104156.1|	9	30	9	13	239
gi|524895745|ref|XP_005104354.1|	9	73	108	591	13659
gi|524897639|ref|XP_005105279.1|	9	38	9	15	185
gi|524898449|ref|XP_005105673.1|	9	49	9	27	313
gi|524900084|ref|XP_005106466.1|	9	30	9	10	142
gi|524902881|ref|XP_005107662.1|	9	45	9	10	114
gi|524903367|ref|XP_005107884.1|	9	31	9	20	364
gi|524903438|ref|XP_005107919.1|	9	47	9	14	373
gi|524907042|ref|XP_005108650.1|	9	31	9	9	115
gi|524907491|ref|XP_005108862.1|	9	33	9	26	559
gi|524907610|ref|XP_005108921.1|	9	34	9	31	807
gi|524909004|ref|XP_005109589.1|	9	48	12	70	906
gi|524909258|ref|XP_005109678.1|	9	43	9	25	503
gi|524910373|ref|XP_005110066.1|	9	29	10	24	344
gi|524910383|ref|XP_005110071.1|	9	48	9	22	594
gi|524910825|ref|XP_005110286.1|	9	47	9	12	150
gi|524911947|ref|XP_005110831.1|	9	69	16	311	6806
gi|524911947|ref|XP_005110831.1|	9	64	17	57	801
gi|524912191|ref|XP_005110947.1|	9	67	9	20	251
gi|524912929|ref|XP_005111300.1|	9	77	9	105	2318
gi|524912990|ref|XP_005111329.1|	9	45	9	12	241
gi|524913818|ref|XP_005111733.1|	9	31	11	14	261
gi|524915771|ref|XP_005112676.1|	9	43	11	105	1654
gi|524916029|ref|XP_005112797.1|	9	35	9	46	992
gi|524916507|ref|XP_005113030.1|	9	45	12	14	230
gi|556093757|gb|ESO82410.1|	9	31	9	10	139
gi|556093891|gb|ESO82544.1|	9	47	9	10	168
gi|556097339|gb|ESO85991.1|	9	11	9	14	259
gi|556100834|gb|ESO89486.1|	9	42	9	20	468
gi|556101047|gb|ESO89699.1|	9	10	16	26	338
gi|556101476|gb|ESO90128.1|	9	32	17	21	400
gi|556103805|gb|ESO92457.1|	9	36	9	14	231
gi|556107233|gb|ESO95885.1|	9	27	9	33	905
gi|556109519|gb|ESO98171.1|	9	7	9	12	237
gi|556110708|gb|ESO99359.1|	9	35	9	19	297

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|556113321|gb|ESP01973.1|	9	32	9	18	175
gi|556114539|gb|ESP03191.1|	9	12	9	13	252
gi|577029907|gb|AHH81788.1|	9	58	9	19	275
gi|597788484|ref|XP_007257657.1|	9	35	10	17	112
gi|608090766|gb|EZG30020.1|	9	74	9	42	1408
gi|118425914|gb|ABK90856.1|	8	25	8	23	659
gi|150251392|gb|ABR68008.1|	8	37	8	62	1041
gi|158997667|gb|ABW86957.1|	8	47	8	196	3488
gi|2073142|dbj|BAA19861.1|	8	55	8	201	4262
gi|241556129|ref|XP_002399605.1|	8	51	8	19	286
gi|241835856|ref|XP_002415077.1|	8	41	8	15	149
gi|260781473|ref|XP_002585834.1|	8	62	8	37	665
gi|2827905|gb|AAB99911.1|	8	19	10	22	432
gi|325296853|ref|NP_001191458.1|	8	23	9	15	293
gi|325296989|ref|NP_001191508.1|	8	41	10	15	211
gi|326535853|gb|ADZ76593.1|	8	18	8	9	136
gi|346987840|gb|AEO51766.1|	8	34	10	107	2726
gi|375073675|gb|AFA34396.1|	8	58	8	22	355
gi|402227989|gb|AFQ35982.1|	8	47	8	25	512
gi|405957391|gb|EKC23605.1|	8	41	8	8	227
gi|405965644|gb|EKC31006.1|	8	36	9	15	347
gi|405968798|gb|EKC33830.1|	8	30	8	11	282
gi|405969732|gb|EKC34685.1|	8	18	8	17	253
gi|405969853|gb|EKC34799.1|	8	18	8	14	283
gi|405973356|gb|EKC38075.1|	8	43	9	11	186
gi|421975936|gb|AFX73000.1|	8	38	8	19	140
gi|42559558|sp|O97192.1|TPM_HELAS	8	64	23	77	1135
gi|524863988|ref|XP_005088833.1|	8	52	8	13	255
gi|524864118|ref|XP_005088897.1|	8	60	8	11	87
gi|524864532|ref|XP_005089101.1|	8	26	8	14	231
gi|524866585|ref|XP_005090113.1|	8	21	8	9	131
gi|524866627|ref|XP_005090132.1|	8	48	8	13	178
gi|524867048|ref|XP_005090337.1|	8	62	9	19	260
gi|524867757|ref|XP_005090686.1|	8	11	19	26	550
gi|524868003|ref|XP_005090808.1|	8	36	10	44	780
gi|524869549|ref|XP_005091567.1|	8	24	31	57	1150
gi|524869987|ref|XP_005091780.1|	8	55	8	14	234
gi|524871178|ref|XP_005092365.1|	8	37	8	17	207
gi|524871453|ref|XP_005092499.1|	8	23	8	11	128
gi|524871609|ref|XP_005092576.1|	8	14	8	10	260
gi|524872616|ref|XP_005093069.1|	8	21	13	15	250
gi|524876525|ref|XP_005094978.1|	8	55	15	25	531

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524876974|ref|XP_005095195.1|	8	32	8	10	202
gi|524876984|ref|XP_005095200.1|	8	19	8	8	161
gi|524877031|ref|XP_005095223.1|	8	35	8	18	330
gi|524879910|ref|XP_005096625.1|	8	22	8	11	191
gi|524879933|ref|XP_005096636.1|	8	40	8	14	227
gi|524880906|ref|XP_005097111.1|	8	32	8	9	123
gi|524881047|ref|XP_005097179.1|	8	46	8	17	236
gi|524881809|ref|XP_005097554.1|	8	47	8	14	198
gi|524882120|ref|XP_005097705.1|	8	27	8	22	349
gi|524883534|ref|XP_005098394.1|	8	59	8	37	537
gi|524883626|ref|XP_005098440.1|	8	65	8	52	1123
gi|524884884|ref|XP_005099056.1|	8	55	8	19	185
gi|524885593|ref|XP_005099402.1|	8	48	9	34	1083
gi|524886540|ref|XP_005099863.1|	8	33	8	11	306
gi|524886950|ref|XP_005100065.1|	8	21	8	12	135
gi|524886963|ref|XP_005100071.1|	8	12	8	11	177
gi|524887258|ref|XP_005100217.1|	8	22	8	12	173
gi|524888539|ref|XP_005100839.1|	8	21	8	19	464
gi|524889887|ref|XP_005101495.1|	8	34	8	12	252
gi|524890204|ref|XP_005101650.1|	8	44	12	23	398
gi|524890839|ref|XP_005101961.1|	8	39	8	15	150
gi|524890899|ref|XP_005101991.1|	8	15	8	30	508
gi|524891306|ref|XP_005102188.1|	8	27	8	15	369
gi|524891634|ref|XP_005102350.1|	8	38	8	24	346
gi|524892640|ref|XP_005102846.1|	8	59	10	37	934
gi|524892640|ref|XP_005102846.1|	8	80	10	43	929
gi|524894544|ref|XP_005103769.1|	8	38	8	16	130
gi|524896746|ref|XP_005104844.1|	8	56	8	21	106
gi|524899180|ref|XP_005106028.1|	8	29	8	15	371
gi|524900516|ref|XP_005106678.1|	8	29	8	21	222
gi|524900516|ref|XP_005106678.1|	8	43	8	25	684
gi|524901804|ref|XP_005107306.1|	8	74	8	36	814
gi|524901804|ref|XP_005107306.1|	8	71	8	59	1213
gi|524903262|ref|XP_005107836.1|	8	31	8	13	120
gi|524903461|ref|XP_005107930.1|	8	48	8	19	326
gi|524906081|ref|XP_005108389.1|	8	25	8	13	250
gi|524910284|ref|XP_005110023.1|	8	16	8	11	243
gi|524910373|ref|XP_005110066.1|	8	23	10	32	573
gi|524910660|ref|XP_005110205.1|	8	22	8	19	410
gi|524910816|ref|XP_005110282.1|	8	27	20	35	694
gi|524911666|ref|XP_005110693.1|	8	16	8	11	248
gi|524913419|ref|XP_005111540.1|	8	7	9	43	1232

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524913818|ref|XP_005111733.1|	8	30	9	19	396
gi|524914230|ref|XP_005111933.1|	8	19	8	9	127
gi|524915771|ref|XP_005112676.1|	8	44	9	81	1337
gi|524916607|ref|XP_005113077.1|	8	41	8	26	345
gi|556094806|gb|ESO83459.1|	8	27	8	13	227
gi|556104219|gb|ESO92871.1|	8	11	8	8	176
gi|556105302|gb|ESO93954.1|	8	55	8	173	3394
gi|655453820|ref|WP_028837045.1|	8	38	8	16	266
gi|773571|emb|CAA60122.1|	8	16	8	9	153
gi|104162064|emb|CAK19329.1|	7	55	7	207	3529
gi|13177628|gb|AAK14899.1|	7	28	7	11	259
gi|13177628|gb|AAK14899.1|	7	22	7	14	411
gi|232136|sp|P30683.3|GNAO_LYMST	7	19	7	12	101
gi|23394916|gb|AAN31640.1|AF484963_1	7	29	7	12	152
gi|297186099|gb|ADI24337.1|	7	18	7	8	148
gi|325197136|ref|NP_001191419.1|	7	53	7	29	493
gi|325296995|ref|NP_001191510.1|	7	43	7	24	434
gi|325297068|ref|NP_001191538.1|	7	16	9	18	260
gi|348573378|ref|XP_003472468.1|	7	17	7	18	440
gi|38043955|emb|CAD83837.1|	7	53	7	19	236
gi|405967712|gb|EKC32845.1|	7	33	7	7	104
gi|405970416|gb|EKC35324.1|	7	4	7	11	283
gi|524864066|ref|XP_005088871.1|	7	44	7	23	306
gi|524864808|ref|XP_005089238.1|	7	28	7	12	130
gi|524867050|ref|XP_005090338.1|	7	45	8	19	208
gi|524868522|ref|XP_005091063.1|	7	16	15	16	253
gi|524868842|ref|XP_005091219.1|	7	36	7	24	407
gi|524869393|ref|XP_005091491.1|	7	28	7	10	136
gi|524870815|ref|XP_005092188.1|	7	39	7	14	226
gi|524871144|ref|XP_005092348.1|	7	24	7	9	156
gi|524871449|ref|XP_005092497.1|	7	14	7	10	162
gi|524871513|ref|XP_005092529.1|	7	28	8	9	181
gi|524871732|ref|XP_005092637.1|	7	6	7	21	450
gi|524871748|ref|XP_005092645.1|	7	61	7	24	452
gi|524872186|ref|XP_005092857.1|	7	21	8	9	242
gi|524872578|ref|XP_005093050.1|	7	30	23	46	1182
gi|524872801|ref|XP_005093157.1|	7	61	11	37	808
gi|524872832|ref|XP_005093172.1|	7	24	7	10	172
gi|524874093|ref|XP_005093787.1|	7	60	7	20	274
gi|524874093|ref|XP_005093787.1|	7	48	7	27	390
gi|524874372|ref|XP_005093922.1|	7	19	7	9	205
gi|524874414|ref|XP_005093942.1|	7	14	7	11	254

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524875029|ref|XP_005094246.1|	7	58	7	15	299
gi|524876118|ref|XP_005094778.1|	7	50	7	18	617
gi|524877040|ref|XP_005095227.1|	7	43	7	18	508
gi|524877054|ref|XP_005095234.1|	7	22	7	8	183
gi|524877903|ref|XP_005095647.1|	7	6	7	8	61
gi|524878669|ref|XP_005096017.1|	7	40	7	7	147
gi|524878983|ref|XP_005096171.1|	7	40	10	15	117
gi|524879223|ref|XP_005096289.1|	7	32	7	26	428
gi|524879541|ref|XP_005096445.1|	7	24	7	10	184
gi|524880248|ref|XP_005096790.1|	7	38	7	16	105
gi|524880725|ref|XP_005097023.1|	7	31	7	16	191
gi|524881413|ref|XP_005097358.1|	7	31	8	14	93
gi|524881743|ref|XP_005097521.1|	7	32	7	34	671
gi|524882420|ref|XP_005097854.1|	7	63	9	29	343
gi|524882956|ref|XP_005098113.1|	7	20	7	11	235
gi|524883548|ref|XP_005098401.1|	7	33	7	18	405
gi|524883721|ref|XP_005098487.1|	7	27	7	11	128
gi|524884503|ref|XP_005098871.1|	7	46	7	22	200
gi|524885821|ref|XP_005099514.1|	7	44	7	11	196
gi|524886408|ref|XP_005099800.1|	7	25	7	12	135
gi|524886658|ref|XP_005099922.1|	7	20	7	8	195
gi|524886715|ref|XP_005099950.1|	7	13	8	11	186
gi|524886950|ref|XP_005100065.1|	7	31	7	8	171
gi|524887715|ref|XP_005100441.1|	7	17	7	10	128
gi|524887883|ref|XP_005100524.1|	7	38	7	16	200
gi|524888089|ref|XP_005100621.1|	7	25	7	13	213
gi|524889772|ref|XP_005101439.1|	7	28	7	12	297
gi|524890066|ref|XP_005101582.1|	7	18	7	10	122
gi|524890221|ref|XP_005101658.1|	7	14	7	8	121
gi|524890263|ref|XP_005101679.1|	7	14	7	14	255
gi|524890869|ref|XP_005101976.1|	7	22	7	11	206
gi|524891914|ref|XP_005102489.1|	7	39	7	12	59
gi|524893468|ref|XP_005103247.1|	7	68	7	12	198
gi|524893612|ref|XP_005103315.1|	7	42	7	22	363
gi|524893715|ref|XP_005103366.1|	7	34	7	23	192
gi|524893900|ref|XP_005103458.1|	7	29	13	26	355
gi|524894261|ref|XP_005103632.1|	7	10	9	30	355
gi|524894818|ref|XP_005103900.1|	7	33	7	33	285
gi|524895745|ref|XP_005104354.1|	7	57	60	516	14983
gi|524895745|ref|XP_005104354.1|	7	75	81	648	16063
gi|524895745|ref|XP_005104354.1|	7	77	81	605	14610
gi|524896457|ref|XP_005104703.1|	7	35	7	32	567

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524897390|ref|XP_005105156.1|	7	81	7	23	215
gi|524898535|ref|XP_005105715.1|	7	35	8	11	134
gi|524898553|ref|XP_005105724.1|	7	10	7	12	323
gi|524899361|ref|XP_005106113.1|	7	21	7	10	184
gi|524900516|ref|XP_005106678.1|	7	28	7	14	139
gi|524900516|ref|XP_005106678.1|	7	39	8	14	212
gi|524900832|ref|XP_005106833.1|	7	31	11	18	403
gi|524901536|ref|XP_005107174.1|	7	22	9	37	797
gi|524903201|ref|XP_005107810.1|	7	29	47	80	2053
gi|524903293|ref|XP_005107850.1|	7	23	7	14	227
gi|524906293|ref|XP_005108469.1|	7	49	7	9	179
gi|524906745|ref|XP_005108595.1|	7	51	10	28	510
gi|524906932|ref|XP_005108623.1|	7	15	12	13	263
gi|524906932|ref|XP_005108623.1|	7	22	13	14	331
gi|524907047|ref|XP_005108652.1|	7	17	7	12	398
gi|524907590|ref|XP_005108911.1|	7	42	7	20	286
gi|524907772|ref|XP_005108998.1|	7	25	7	9	180
gi|524908561|ref|XP_005109374.1|	7	10	7	13	189
gi|524909434|ref|XP_005109764.1|	7	32	7	12	116
gi|524911952|ref|XP_005110833.1|	7	68	16	88	1738
gi|524912848|ref|XP_005111261.1|	7	29	7	12	150
gi|524913368|ref|XP_005111515.1|	7	42	7	16	374
gi|524913419|ref|XP_005111540.1|	7	7	8	33	1012
gi|524915657|ref|XP_005112621.1|	7	29	7	9	182
gi|524915986|ref|XP_005112777.1|	7	60	7	38	556
gi|524916381|ref|XP_005112969.1|	7	29	7	8	98
gi|524916593|ref|XP_005113070.1|	7	41	8	25	457
gi|556094817|gb|ESO83470.1|	7	9	7	10	137
gi|556097355|gb|ESO86007.1|	7	59	7	13	262
gi|556097653|gb|ESO86305.1|	7	20	7	7	87
gi|556097720|gb|ESO86372.1|	7	40	7	27	370
gi|556098160|gb|ESO86812.1|	7	39	10	26	441
gi|556100505|gb|ESO89157.1|	7	44	7	16	406
gi|556105072|gb|ESO93724.1|	7	14	7	19	431
gi|556111273|gb|ESO99924.1|	7	36	7	28	587
gi|556114460|gb|ESP03112.1|	7	31	7	13	215
gi|556116283|gb|ESP04935.1|	7	17	8	18	339
gi|568301578|gb|ETN87131.1|	7	25	7	7	138
gi|617524151|ref|XP_007544814.1|	7	35	7	15	333
gi|646714599|gb|KDR18512.1|	7	64	7	19	351
gi|74137592|dbj|BAE35828.1|	7	30	7	19	481
gi|126697398|gb|ABO26656.1|	6	29	6	10	148

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|254939725|gb|ACT88125.1|	6	56	6	10	148
gi|325197162|ref|NP_001191425.1|	6	29	6	11	232
gi|325296867|ref|NP_001191465.1|	6	29	6	12	262
gi|348577231|ref|XP_003474388.1|	6	30	6	27	454
gi|353236678|emb|CCA68668.1|	6	20	6	7	131
gi|405970946|gb|EKC35808.1|	6	64	6	30	872
gi|405975785|gb|EKC40331.1|	6	56	32	156	3232
gi|405976077|gb|EKC40598.1|	6	12	6	6	124
gi|42559558|sp|O97192.1|TPM_HELAS	6	57	18	96	1911
gi|42559558|sp|O97192.1|TPM_HELAS	6	73	28	69	876
gi|443733604|gb|ELU17896.1|	6	11	6	9	130
gi|524865020|ref|XP_005089342.1|	6	18	6	6	124
gi|524865020|ref|XP_005089342.1|	6	72	7	11	209
gi|524866113|ref|XP_005089881.1|	6	22	6	6	91
gi|524866188|ref|XP_005089918.1|	6	39	6	16	281
gi|524866538|ref|XP_005090090.1|	6	34	6	16	184
gi|524867549|ref|XP_005090585.1|	6	37	6	21	436
gi|524867701|ref|XP_005090658.1|	6	17	6	8	157
gi|524867763|ref|XP_005090689.1|	6	9	13	21	429
gi|524868003|ref|XP_005090808.1|	6	30	7	11	262
gi|524868892|ref|XP_005091244.1|	6	30	6	16	298
gi|524869453|ref|XP_005091519.1|	6	21	6	8	104
gi|524871619|ref|XP_005092581.1|	6	10	6	13	311
gi|524871674|ref|XP_005092608.1|	6	31	6	20	303
gi|524871774|ref|XP_005092658.1|	6	35	6	10	214
gi|524872460|ref|XP_005092992.1|	6	32	6	9	247
gi|524873900|ref|XP_005093692.1|	6	30	6	10	164
gi|524873980|ref|XP_005093732.1|	6	10	6	6	151
gi|524874093|ref|XP_005093787.1|	6	43	6	12	198
gi|524874148|ref|XP_005093814.1|	6	27	6	62	569
gi|524874969|ref|XP_005094216.1|	6	8	6	7	136
gi|524875029|ref|XP_005094246.1|	6	47	6	18	147
gi|524876286|ref|XP_005094861.1|	6	17	6	7	162
gi|524876396|ref|XP_005094915.1|	6	36	6	12	208
gi|524876525|ref|XP_005094978.1|	6	32	9	16	356
gi|524876525|ref|XP_005094978.1|	6	37	11	27	570
gi|524876525|ref|XP_005094978.1|	6	44	14	31	647
gi|524877265|ref|XP_005095338.1|	6	26	9	14	321
gi|524878037|ref|XP_005095711.1|	6	41	6	7	188
gi|524878114|ref|XP_005095748.1|	6	30	6	9	109
gi|524878247|ref|XP_005095812.1|	6	12	6	6	128
gi|524878596|ref|XP_005095981.1|	6	11	6	8	165

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524879065|ref|XP_005096212.1|	6	18	6	8	173
gi|524879771|ref|XP_005096557.1|	6	25	13	20	426
gi|524879821|ref|XP_005096582.1|	6	22	6	9	230
gi|524881344|ref|XP_005097324.1|	6	23	6	9	178
gi|524881651|ref|XP_005097475.1|	6	19	6	6	173
gi|524881865|ref|XP_005097581.1|	6	16	6	7	158
gi|524881948|ref|XP_005097620.1|	6	8	6	18	550
gi|524881984|ref|XP_005097638.1|	6	44	6	9	211
gi|524882215|ref|XP_005097752.1|	6	56	6	13	211
gi|524882332|ref|XP_005097810.1|	6	27	6	11	135
gi|524882460|ref|XP_005097873.1|	6	54	6	67	1357
gi|524883129|ref|XP_005098198.1|	6	38	6	37	548
gi|524883540|ref|XP_005098397.1|	6	35	6	17	326
gi|524883762|ref|XP_005098507.1|	6	26	6	8	131
gi|524884099|ref|XP_005098673.1|	6	12	6	10	152
gi|524884410|ref|XP_005098825.1|	6	40	6	10	172
gi|524884955|ref|XP_005099089.1|	6	29	6	14	391
gi|524885029|ref|XP_005099126.1|	6	32	6	8	45
gi|524885272|ref|XP_005099245.1|	6	63	6	39	1311
gi|524886028|ref|XP_005099614.1|	6	49	6	14	219
gi|524886707|ref|XP_005099946.1|	6	17	6	12	281
gi|524886950|ref|XP_005100065.1|	6	21	6	18	515
gi|524887138|ref|XP_005100158.1|	6	33	6	11	177
gi|524887169|ref|XP_005100173.1|	6	24	6	7	116
gi|524887214|ref|XP_005100195.1|	6	58	7	16	311
gi|524887319|ref|XP_005100247.1|	6	26	6	11	274
gi|524887394|ref|XP_005100283.1|	6	35	6	16	150
gi|524888756|ref|XP_005100943.1|	6	19	6	8	152
gi|524889720|ref|XP_005101414.1|	6	41	6	10	151
gi|524889734|ref|XP_005101421.1|	6	10	6	9	256
gi|524889992|ref|XP_005101546.1|	6	57	8	21	428
gi|524890794|ref|XP_005101939.1|	6	33	6	8	144
gi|524891363|ref|XP_005102216.1|	6	16	6	8	118
gi|524891574|ref|XP_005102320.1|	6	32	6	9	132
gi|524893281|ref|XP_005103158.1|	6	14	6	9	116
gi|524893352|ref|XP_005103192.1|	6	25	6	14	378
gi|524893513|ref|XP_005103269.1|	6	23	6	8	101
gi|524893741|ref|XP_005103379.1|	6	47	7	24	292
gi|524895469|ref|XP_005104217.1|	6	25	6	12	212
gi|524895639|ref|XP_005104301.1|	6	23	6	14	252
gi|524896057|ref|XP_005104507.1|	6	38	6	7	117
gi|524896189|ref|XP_005104572.1|	6	29	7	8	142

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524896407|ref|XP_005104678.1|	6	36	6	11	220
gi|524898531|ref|XP_005105713.1|	6	29	7	15	338
gi|524898859|ref|XP_005105873.1|	6	15	6	12	132
gi|524898960|ref|XP_005105922.1|	6	18	6	8	207
gi|524899002|ref|XP_005105942.1|	6	38	12	20	401
gi|524899021|ref|XP_005105951.1|	6	22	6	13	140
gi|524899091|ref|XP_005105986.1|	6	53	6	13	151
gi|524899489|ref|XP_005106175.1|	6	20	6	8	145
gi|524899675|ref|XP_005106266.1|	6	24	6	13	296
gi|524900390|ref|XP_005106615.1|	6	19	7	18	375
gi|524900832|ref|XP_005106833.1|	6	24	10	10	165
gi|524901283|ref|XP_005107052.1|	6	73	8	10	226
gi|524901464|ref|XP_005107139.1|	6	38	9	13	173
gi|524902427|ref|XP_005107465.1|	6	30	7	10	284
gi|524902628|ref|XP_005107543.1|	6	50	6	13	319
gi|524902755|ref|XP_005107604.1|	6	25	6	9	179
gi|524903041|ref|XP_005107738.1|	6	40	12	41	779
gi|524903201|ref|XP_005107810.1|	6	34	6	17	552
gi|524906862|ref|XP_005108596.1|	6	74	6	9	133
gi|524906990|ref|XP_005108638.1|	6	23	6	11	94
gi|524907042|ref|XP_005108650.1|	6	26	6	9	101
gi|524907370|ref|XP_005108803.1|	6	26	6	8	169
gi|524908000|ref|XP_005109109.1|	6	39	6	20	396
gi|524910373|ref|XP_005110066.1|	6	22	7	8	198
gi|524910744|ref|XP_005110246.1|	6	16	6	12	286
gi|524910825|ref|XP_005110286.1|	6	40	6	16	347
gi|524910895|ref|XP_005110321.1|	6	23	6	14	429
gi|524911513|ref|XP_005110619.1|	6	33	6	13	218
gi|524911570|ref|XP_005110646.1|	6	10	8	11	258
gi|524911576|ref|XP_005110649.1|	6	32	6	16	230
gi|524911939|ref|XP_005110827.1|	6	26	6	12	288
gi|524911952|ref|XP_005110833.1|	6	60	15	170	3006
gi|524912806|ref|XP_005111241.1|	6	17	6	6	129
gi|524912919|ref|XP_005111295.1|	6	17	6	7	179
gi|524912927|ref|XP_005111299.1|	6	71	6	27	531
gi|524912929|ref|XP_005111300.1|	6	54	6	42	713
gi|524913658|ref|XP_005111654.1|	6	60	6	42	819
gi|524914862|ref|XP_005112240.1|	6	30	6	16	392
gi|524915397|ref|XP_005112499.1|	6	20	6	14	356
gi|524916499|ref|XP_005113026.1|	6	31	6	15	377
gi|541039654|gb|ERG79357.1|	6	21	6	6	219
gi|555930987|emb|CDJ08105.1|	6	78	6	27	350

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|556093552|gb|ESO82205.1|	6	36	6	15	250
gi|556097053|gb|ESO85705.1|	6	25	6	8	289
gi|556098160|gb|ESO86812.1|	6	29	7	13	343
gi|556098160|gb|ESO86812.1|	6	30	7	42	742
gi|556100167|gb|ESO88819.1|	6	68	6	26	531
gi|556101526|gb|ESO90178.1|	6	27	6	13	399
gi|556101906|gb|ESO90558.1|	6	16	6	9	153
gi|556105072|gb|ESO93724.1|	6	38	6	23	505
gi|556109149|gb|ESO97801.1|	6	48	7	11	225
gi|556110831|gb|ESO99482.1|	6	44	6	49	619
gi|556111437|gb|ESP00089.1|	6	41	10	24	461
gi|556114369|gb|ESP03021.1|	6	6	6	8	190
gi|576693750|gb|EUB57355.1|	6	39	6	19	242
gi|72113881|ref|XP_787111.1|	6	82	6	16	438
gi|91077014|ref|XP_966444.1|	6	17	6	10	178
gi|126697408|gb|ABO26661.1|	5	33	5	8	180
gi|147903958|ref|NP_001090965.1|	5	69	5	25	580
gi|150251390|gb|ABR68007.1|	5	46	5	41	1156
gi|2073142|dbj|BAA19861.1|	5	44	5	16	273
gi|239509144|gb|ACR81565.1|	5	56	5	12	182
gi|325296803|ref|NP_001191642.1|	5	45	6	19	231
gi|325296881|ref|NP_001191472.1|	5	28	5	5	72
gi|325296937|ref|NP_001191493.1|	5	36	5	9	162
gi|325296986|ref|NP_001191507.1|	5	9	5	6	84
gi|346987840|gb|AEO51766.1|	5	42	8	22	535
gi|346987844|gb|AEO51768.1|	5	26	6	28	323
gi|350536823|ref|NP_001233149.1|	5	24	5	6	125
gi|351714959|gb|EHB17878.1|	5	17	6	7	193
gi|374093227|ref|NP_001243365.1|	5	63	5	17	304
gi|380022918|ref|XP_003695281.1|	5	44	5	21	398
gi|387157924|ref|NP_001248345.1|	5	23	5	10	71
gi|405959537|gb|EKC25563.1|	5	27	5	12	308
gi|405960428|gb|EKC26353.1|	5	18	5	29	618
gi|405961635|gb|EKC27403.1|	5	5	5	7	150
gi|405967797|gb|EKC32924.1|	5	46	8	34	1232
gi|405968748|gb|EKC33788.1|	5	37	5	14	389
gi|405970417|gb|EKC35325.1|	5	8	5	7	115
gi|405970417|gb|EKC35325.1|	5	13	5	9	109
gi|405972388|gb|EKC37161.1|	5	17	5	7	123
gi|405972489|gb|EKC37255.1|	5	22	5	5	114
gi|431831576|gb|AGA92562.1|	5	46	5	18	478
gi|443690943|gb|ELT92936.1|	5	25	5	6	88

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|46277000|gb|AAS86696.1|	5	32	5	10	213
gi|46277026|gb|AAS86709.1|	5	11	5	7	96
gi|51105058|gb|AAT97089.1|	5	49	6	18	482
gi|524864548|ref|XP_005089109.1|	5	28	5	5	91
gi|524865306|ref|XP_005089483.1|	5	22	5	8	128
gi|524865670|ref|XP_005089662.1|	5	71	5	97	1767
gi|524866300|ref|XP_005089972.1|	5	24	5	16	301
gi|524866733|ref|XP_005090185.1|	5	40	5	10	170
gi|524866733|ref|XP_005090185.1|	5	33	5	7	103
gi|524867753|ref|XP_005090684.1|	5	20	12	16	319
gi|524868037|ref|XP_005090825.1|	5	23	5	10	97
gi|524868530|ref|XP_005091067.1|	5	16	5	6	135
gi|524868604|ref|XP_005091103.1|	5	28	6	8	88
gi|524868651|ref|XP_005091126.1|	5	18	5	5	101
gi|524868983|ref|XP_005091289.1|	5	51	5	15	136
gi|524869485|ref|XP_005091535.1|	5	11	5	5	89
gi|524870008|ref|XP_005091790.1|	5	30	5	9	100
gi|524870146|ref|XP_005091859.1|	5	29	6	7	115
gi|524870945|ref|XP_005092251.1|	5	41	5	13	125
gi|524870955|ref|XP_005092256.1|	5	20	5	9	247
gi|524871172|ref|XP_005092362.1|	5	13	5	7	61
gi|524871463|ref|XP_005092504.1|	5	37	5	12	133
gi|524871621|ref|XP_005092582.1|	5	21	5	12	165
gi|524872186|ref|XP_005092857.1|	5	20	5	5	60
gi|524872436|ref|XP_005092980.1|	5	13	6	6	129
gi|524872618|ref|XP_005093070.1|	5	16	10	14	204
gi|524872622|ref|XP_005093072.1|	5	12	6	6	77
gi|524872725|ref|XP_005093120.1|	5	16	5	10	282
gi|524873045|ref|XP_005093276.1|	5	21	5	19	307
gi|524873315|ref|XP_005093407.1|	5	26	5	5	105
gi|524873378|ref|XP_005093436.1|	5	20	5	12	249
gi|524873738|ref|XP_005093612.1|	5	35	6	6	74
gi|524873892|ref|XP_005093688.1|	5	28	5	10	266
gi|524874171|ref|XP_005093825.1|	5	31	5	10	204
gi|524874182|ref|XP_005093830.1|	5	49	5	10	227
gi|524874452|ref|XP_005093961.1|	5	27	5	15	138
gi|524876470|ref|XP_005094951.1|	5	27	5	10	276
gi|524876525|ref|XP_005094978.1|	5	57	6	23	400
gi|524876560|ref|XP_005094995.1|	5	10	5	8	201
gi|524877129|ref|XP_005095271.1|	5	32	5	20	368
gi|524877853|ref|XP_005095623.1|	5	29	9	20	410
gi|524878217|ref|XP_005095797.1|	5	23	5	5	96

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524878910|ref|XP_005096136.1|	5	21	5	8	214
gi|524879191|ref|XP_005096273.1|	5	15	5	5	91
gi|524879345|ref|XP_005096349.1|	5	20	5	7	171
gi|524880082|ref|XP_005096709.1|	5	28	5	10	114
gi|524881469|ref|XP_005097386.1|	5	18	5	8	173
gi|524881557|ref|XP_005097429.1|	5	56	11	57	1482
gi|524882112|ref|XP_005097701.1|	5	35	5	11	102
gi|524882460|ref|XP_005097873.1|	5	47	5	26	437
gi|524882484|ref|XP_005097885.1|	5	22	6	11	238
gi|524883119|ref|XP_005098193.1|	5	10	5	5	53
gi|524883558|ref|XP_005098406.1|	5	12	5	9	229
gi|524884059|ref|XP_005098653.1|	5	19	5	14	465
gi|524884444|ref|XP_005098842.1|	5	23	5	13	405
gi|524884503|ref|XP_005098871.1|	5	32	5	16	192
gi|524884697|ref|XP_005098967.1|	5	15	5	5	99
gi|524885145|ref|XP_005099183.1|	5	14	5	6	170
gi|524885272|ref|XP_005099245.1|	5	55	5	31	1067
gi|524886570|ref|XP_005099878.1|	5	49	5	20	683
gi|524887214|ref|XP_005100195.1|	5	67	6	13	224
gi|524887250|ref|XP_005100213.1|	5	20	5	7	163
gi|524887580|ref|XP_005100374.1|	5	13	6	12	243
gi|524887834|ref|XP_005100500.1|	5	39	5	15	197
gi|524887886|ref|XP_005100525.1|	5	23	5	14	305
gi|524887980|ref|XP_005100569.1|	5	16	5	6	145
gi|524888015|ref|XP_005100586.1|	5	12	5	5	96
gi|524888025|ref|XP_005100590.1|	5	52	5	8	210
gi|524888089|ref|XP_005100621.1|	5	21	5	7	167
gi|524888414|ref|XP_005100778.1|	5	29	5	11	190
gi|524888480|ref|XP_005100811.1|	5	24	7	8	118
gi|524888807|ref|XP_005100968.1|	5	10	5	13	333
gi|524889199|ref|XP_005101160.1|	5	15	5	9	170
gi|524889332|ref|XP_005101225.1|	5	5	5	5	77
gi|524889992|ref|XP_005101546.1|	5	53	5	9	80
gi|524889992|ref|XP_005101546.1|	5	40	8	14	282
gi|524890204|ref|XP_005101650.1|	5	42	12	22	434
gi|524890370|ref|XP_005101732.1|	5	10	5	8	122
gi|524890590|ref|XP_005101839.1|	5	22	5	5	107
gi|524891178|ref|XP_005102125.1|	5	26	6	10	168
gi|524891741|ref|XP_005102403.1|	5	17	5	7	117
gi|524891896|ref|XP_005102480.1|	5	58	8	31	499
gi|524893405|ref|XP_005103218.1|	5	19	5	5	106
gi|524893771|ref|XP_005103394.1|	5	26	5	9	270

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524894650|ref|XP_005103820.1|	5	18	5	6	176
gi|524894931|ref|XP_005103955.1|	5	26	5	7	66
gi|524894976|ref|XP_005103977.1|	5	52	5	9	97
gi|524896213|ref|XP_005104584.1|	5	4	5	6	208
gi|524896814|ref|XP_005104878.1|	5	30	5	8	196
gi|524896949|ref|XP_005104944.1|	5	33	5	5	58
gi|524897983|ref|XP_005105446.1|	5	26	6	14	254
gi|524898757|ref|XP_005105824.1|	5	14	5	5	102
gi|524899002|ref|XP_005105942.1|	5	32	11	20	496
gi|524899023|ref|XP_005105952.1|	5	26	5	9	182
gi|524900380|ref|XP_005106610.1|	5	31	5	24	475
gi|524900516|ref|XP_005106678.1|	5	32	5	9	172
gi|524900809|ref|XP_005106822.1|	5	16	5	7	122
gi|524900993|ref|XP_005106911.1|	5	20	5	7	202
gi|524902421|ref|XP_005107462.1|	5	39	5	56	1035
gi|524902626|ref|XP_005107542.1|	5	55	10	25	508
gi|524902826|ref|XP_005107636.1|	5	22	5	15	308
gi|524902881|ref|XP_005107662.1|	5	33	5	5	85
gi|524903041|ref|XP_005107738.1|	5	39	15	33	637
gi|524903344|ref|XP_005107874.1|	5	21	5	11	169
gi|524903676|ref|XP_005108033.1|	5	19	5	7	138
gi|524906739|ref|XP_005108592.1|	5	26	6	9	183
gi|524906745|ref|XP_005108595.1|	5	38	8	12	200
gi|524909120|ref|XP_005109612.1|	5	20	5	6	129
gi|524910662|ref|XP_005110206.1|	5	37	5	6	76
gi|524910816|ref|XP_005110282.1|	5	37	18	32	610
gi|524911952|ref|XP_005110833.1|	5	54	9	51	818
gi|524913016|ref|XP_005111342.1|	5	25	5	7	135
gi|524913423|ref|XP_005111542.1|	5	52	5	24	960
gi|524913881|ref|XP_005111763.1|	5	10	5	9	94
gi|524914082|ref|XP_005111860.1|	5	36	6	7	107
gi|524914293|ref|XP_005111963.1|	5	27	5	14	161
gi|524914515|ref|XP_005112072.1|	5	26	5	6	56
gi|524914607|ref|XP_005112116.1|	5	23	5	13	134
gi|524914870|ref|XP_005112244.1|	5	31	5	5	95
gi|524915427|ref|XP_005112513.1|	5	11	5	6	190
gi|524915677|ref|XP_005112631.1|	5	40	5	22	447
gi|524915787|ref|XP_005112684.1|	5	33	5	12	183
gi|524915990|ref|XP_005112779.1|	5	21	5	7	135
gi|524916301|ref|XP_005112930.1|	5	18	5	9	131
gi|524916570|ref|XP_005113059.1|	5	67	9	39	606
gi|524917273|ref|XP_005113396.1|	5	22	5	16	288

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|525007750|ref|XP_005051527.1|	5	26	8	19	359
gi|526117361|ref|NP_001267749.1|	5	59	5	176	3178
gi|526117405|ref|NP_001267760.1|	5	65	10	38	188
gi|54020890|ref|NP_001005689.1|	5	19	5	7	89
gi|546686668|gb|ERL95792.1|	5	39	5	8	212
gi|555706265|gb|ESO09498.1|	5	29	5	9	199
gi|556095216|gb|ESO83869.1|	5	29	5	14	330
gi|556095696|gb|ESO84349.1|	5	19	5	6	77
gi|556097689|gb|ESO86341.1|	5	48	5	12	262
gi|556098443|gb|ESO87095.1|	5	20	5	10	167
gi|556099804|gb|ESO88456.1|	5	13	5	8	172
gi|556101047|gb|ESO89699.1|	5	8	8	8	186
gi|556101476|gb|ESO90128.1|	5	13	8	15	258
gi|556103094|gb|ESO91746.1|	5	36	5	7	164
gi|556104081|gb|ESO92733.1|	5	10	5	7	83
gi|556105440|gb|ESO94092.1|	5	11	5	9	169
gi|556106038|gb|ESO94690.1|	5	17	5	6	146
gi|556106851|gb|ESO95503.1|	5	7	5	6	180
gi|556108834|gb|ESO97486.1|	5	16	5	14	239
gi|556108903|gb|ESO97555.1|	5	26	5	9	191
gi|556116375|gb|ESP05027.1|	5	22	7	11	202
gi|576693750|gb|EUB57355.1|	5	40	5	11	154
gi|632946707|ref|XP_007888691.1|	5	19	5	8	192
gi|632953883|ref|XP_007892665.1|	5	11	5	7	282
gi|639127132|gb|AIA24462.1|	5	39	5	6	164
gi|655453820|ref|WP_028837045.1|	5	18	5	5	85
gi|66361226|pdb|1YX5|B	5	43	5	22	233
gi|71726729|gb|AAZ39528.1|	5	76	57	349	6335
gi|773571|emb|CAA60122.1|	5	47	5	8	203
gi|85362949|gb|ABC69854.1|	5	45	5	36	931
gi|93209532|gb|ABF00124.1|	5	48	5	125	3291
gi|109157294|pdb|2CE6|A	4	33	4	40	789
gi|115111533|gb|ABI84147.1|	4	13	4	5	121
gi|133874728|gb|ABO40828.1|	4	23	4	4	99
gi|156144968|gb|ABU53030.1|	4	22	4	5	60
gi|157690722|tpe|CAL69088.1|	4	41	4	24	311
gi|210076789|gb|ACJ06748.1|	4	40	4	7	86
gi|210076789|gb|ACJ06748.1|	4	32	4	9	220
gi|218683625|gb|ACL00841.1|	4	26	4	17	459
gi|220979902|emb|CAQ64775.1|	4	27	5	7	94
gi|260821662|ref|XP_002606151.1|	4	55	19	78	1609
gi|307192533|gb|EFN75721.1|	4	27	4	12	152

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|325296727|ref|NP_001191598.1|	4	25	4	34	527
gi|325296743|ref|NP_001191607.1|	4	26	11	30	621
gi|325296839|ref|NP_001191661.1|	4	39	4	19	152
gi|325296855|ref|NP_001191459.1|	4	30	6	6	137
gi|325296855|ref|NP_001191459.1|	4	45	8	10	287
gi|325296937|ref|NP_001191493.1|	4	29	4	7	66
gi|325297037|ref|NP_001191522.1|	4	28	4	21	637
gi|325504381|emb|CBX41690.1|	4	24	4	68	1934
gi|333449383|gb|AEF33378.1|	4	35	4	14	398
gi|38043955|emb|CAD83837.1|	4	50	4	39	793
gi|397511018|ref|XP_003825879.1|	4	37	12	73	1732
gi|405950185|gb|EKC18187.1|	4	2	4	4	112
gi|405952673|gb|EKC20457.1|	4	38	6	15	175
gi|405962076|gb|EKC27784.1|	4	10	4	4	55
gi|405966102|gb|EKC31422.1|	4	13	4	4	62
gi|405973516|gb|EKC38224.1|	4	42	10	13	216
gi|410591636|sp|I0CME8.1|HYAL_CONAQ	4	13	4	5	70
gi|443682473|gb|ELT87054.1|	4	22	4	4	34
gi|443686917|gb|ELT90035.1|	4	13	4	4	65
gi|443689529|gb|ELT91902.1|	4	35	4	4	39
gi|443705209|gb|ELU01864.1|	4	34	4	10	208
gi|443720976|gb|ELU10481.1|	4	20	4	4	64
gi|46395316|dbj|BAD16595.1|	4	19	4	4	37
gi|472824679|dbj|BAN09032.1|	4	67	6	76	1264
gi|524864996|ref|XP_005089330.1|	4	19	4	6	63
gi|524866780|ref|XP_005090207.1|	4	15	4	4	129
gi|524866838|ref|XP_005090235.1|	4	31	4	8	330
gi|524866956|ref|XP_005090293.1|	4	14	6	10	221
gi|524867567|ref|XP_005090594.1|	4	19	4	4	48
gi|524867786|ref|XP_005090700.1|	4	9	4	8	256
gi|524867792|ref|XP_005090703.1|	4	9	4	4	95
gi|524868297|ref|XP_005090952.1|	4	49	4	5	127
gi|524868395|ref|XP_005091000.1|	4	11	4	4	59
gi|524868717|ref|XP_005091158.1|	4	5	4	4	96
gi|524868787|ref|XP_005091193.1|	4	14	4	7	143
gi|524868815|ref|XP_005091207.1|	4	25	4	9	152
gi|524869485|ref|XP_005091535.1|	4	33	4	10	261
gi|524869549|ref|XP_005091567.1|	4	23	29	51	1098
gi|524870028|ref|XP_005091800.1|	4	14	4	10	185
gi|524871130|ref|XP_005092341.1|	4	8	4	4	108
gi|524871372|ref|XP_005092459.1|	4	21	4	8	151
gi|524871596|ref|XP_005092570.1|	4	10	4	16	317

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524871621|ref|XP_005092582.1|	4	43	4	6	95
gi|524871621|ref|XP_005092582.1|	4	26	5	12	185
gi|524871930|ref|XP_005092735.1|	4	24	4	10	202
gi|524872284|ref|XP_005092905.1|	4	22	4	9	192
gi|524872340|ref|XP_005092932.1|	4	24	4	4	81
gi|524873738|ref|XP_005093612.1|	4	35	5	6	47
gi|524873824|ref|XP_005093654.1|	4	19	5	7	164
gi|524873848|ref|XP_005093666.1|	4	11	4	4	76
gi|524873900|ref|XP_005093692.1|	4	18	4	5	58
gi|524874148|ref|XP_005093814.1|	4	22	4	9	111
gi|524874374|ref|XP_005093923.1|	4	49	12	31	353
gi|524874572|ref|XP_005094021.1|	4	51	4	9	214
gi|524874629|ref|XP_005094049.1|	4	11	4	4	63
gi|524874940|ref|XP_005094202.1|	4	32	4	7	95
gi|524875329|ref|XP_005094395.1|	4	45	4	42	770
gi|524875359|ref|XP_005094409.1|	4	51	4	16	191
gi|524875599|ref|XP_005094525.1|	4	18	4	10	309
gi|524876286|ref|XP_005094861.1|	4	26	4	8	106
gi|524876525|ref|XP_005094978.1|	4	30	8	10	131
gi|524877261|ref|XP_005095336.1|	4	15	6	13	203
gi|524877602|ref|XP_005095500.1|	4	16	7	9	120
gi|524877849|ref|XP_005095621.1|	4	16	4	4	61
gi|524878382|ref|XP_005095878.1|	4	31	4	5	93
gi|524878669|ref|XP_005096017.1|	4	20	4	4	61
gi|524879077|ref|XP_005096218.1|	4	19	4	4	72
gi|524879183|ref|XP_005096269.1|	4	37	4	6	202
gi|524879769|ref|XP_005096556.1|	4	38	4	8	164
gi|524880507|ref|XP_005096916.1|	4	34	4	8	112
gi|524880608|ref|XP_005096966.1|	4	11	5	5	186
gi|524881385|ref|XP_005097344.1|	4	30	4	5	72
gi|524881557|ref|XP_005097429.1|	4	61	11	37	831
gi|524881799|ref|XP_005097549.1|	4	25	4	6	62
gi|524881948|ref|XP_005097620.1|	4	6	4	8	135
gi|524882007|ref|XP_005097649.1|	4	15	4	4	82
gi|524882402|ref|XP_005097845.1|	4	8	4	4	92
gi|524883562|ref|XP_005098408.1|	4	15	4	5	144
gi|524883640|ref|XP_005098447.1|	4	18	4	11	167
gi|524884087|ref|XP_005098667.1|	4	8	4	6	83
gi|524884624|ref|XP_005098931.1|	4	36	5	15	367
gi|524884995|ref|XP_005099109.1|	4	9	4	4	76
gi|524885217|ref|XP_005099218.1|	4	20	5	7	118
gi|524885248|ref|XP_005099233.1|	4	24	4	6	86

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524885593|ref|XP_005099402.1|	4	35	6	35	1019
gi|524885844|ref|XP_005099525.1|	4	10	4	5	70
gi|524886483|ref|XP_005099835.1|	4	31	4	6	102
gi|524886483|ref|XP_005099835.1|	4	30	4	4	36
gi|524886662|ref|XP_005099924.1|	4	14	4	6	95
gi|524887256|ref|XP_005100216.1|	4	32	4	4	119
gi|524887260|ref|XP_005100218.1|	4	15	4	7	109
gi|524887964|ref|XP_005100561.1|	4	9	4	8	137
gi|524888001|ref|XP_005100579.1|	4	29	4	10	141
gi|524888480|ref|XP_005100811.1|	4	19	6	10	243
gi|524888600|ref|XP_005100869.1|	4	19	4	5	67
gi|524888649|ref|XP_005100893.1|	4	52	4	6	70
gi|524889085|ref|XP_005101104.1|	4	28	4	7	141
gi|524889251|ref|XP_005101186.1|	4	24	4	11	143
gi|524889708|ref|XP_005101408.1|	4	13	4	8	173
gi|524889734|ref|XP_005101421.1|	4	12	4	4	57
gi|524889859|ref|XP_005101481.1|	4	16	4	5	76
gi|524890376|ref|XP_005101735.1|	4	27	4	11	467
gi|524890575|ref|XP_005101832.1|	4	13	4	5	53
gi|524891530|ref|XP_005102298.1|	4	8	4	4	51
gi|524891938|ref|XP_005102501.1|	4	28	4	5	72
gi|524892590|ref|XP_005102822.1|	4	19	6	10	160
gi|524892768|ref|XP_005102909.1|	4	24	4	5	100
gi|524892872|ref|XP_005102960.1|	4	18	4	9	154
gi|524893572|ref|XP_005103295.1|	4	10	4	9	147
gi|524893765|ref|XP_005103391.1|	4	11	4	5	146
gi|524894442|ref|XP_005103718.1|	4	34	4	6	148
gi|524894915|ref|XP_005103947.1|	4	18	4	5	104
gi|524895123|ref|XP_005104049.1|	4	38	9	12	186
gi|524896689|ref|XP_005104816.1|	4	21	4	4	38
gi|524896814|ref|XP_005104878.1|	4	15	5	9	171
gi|524897909|ref|XP_005105409.1|	4	19	4	5	249
gi|524898670|ref|XP_005105781.1|	4	11	4	5	77
gi|524898725|ref|XP_005105808.1|	4	23	4	4	56
gi|524898902|ref|XP_005105894.1|	4	43	4	9	81
gi|524899369|ref|XP_005106117.1|	4	14	4	5	208
gi|524899735|ref|XP_005106295.1|	4	15	4	6	133
gi|524901357|ref|XP_005107086.1|	4	19	4	11	229
gi|524901460|ref|XP_005107137.1|	4	23	4	8	139
gi|524902240|ref|XP_005107439.1|	4	11	4	11	270
gi|524902423|ref|XP_005107463.1|	4	13	4	8	92
gi|524902569|ref|XP_005107527.1|	4	14	4	7	75

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524902628|ref|XP_005107543.1|	4	48	9	27	555
gi|524902664|ref|XP_005107561.1|	4	12	4	4	64
gi|524902931|ref|XP_005107685.1|	4	64	4	11	261
gi|524903184|ref|XP_005107804.1|	4	8	4	6	75
gi|524903434|ref|XP_005107917.1|	4	70	4	5	82
gi|524903799|ref|XP_005108092.1|	4	9	4	5	111
gi|524905758|ref|XP_005108242.1|	4	22	4	7	106
gi|524905999|ref|XP_005108354.1|	4	15	4	7	129
gi|524906293|ref|XP_005108469.1|	4	25	4	4	68
gi|524906745|ref|XP_005108595.1|	4	19	4	4	50
gi|524907181|ref|XP_005108715.1|	4	10	4	7	255
gi|524907228|ref|XP_005108737.1|	4	16	4	4	102
gi|524907238|ref|XP_005108742.1|	4	12	4	6	118
gi|524907406|ref|XP_005108821.1|	4	21	4	6	185
gi|524907580|ref|XP_005108906.1|	4	58	5	13	233
gi|524907664|ref|XP_005108948.1|	4	7	4	4	74
gi|524907822|ref|XP_005109023.1|	4	14	4	6	107
gi|524907998|ref|XP_005109108.1|	4	17	4	5	129
gi|524908916|ref|XP_005109547.1|	4	11	4	6	87
gi|524909155|ref|XP_005109628.1|	4	23	4	7	108
gi|524909236|ref|XP_005109667.1|	4	29	5	11	236
gi|524909272|ref|XP_005109685.1|	4	52	4	4	73
gi|524909283|ref|XP_005109690.1|	4	20	4	8	215
gi|524909958|ref|XP_005109871.1|	4	16	4	4	88
gi|524910389|ref|XP_005110074.1|	4	18	4	6	192
gi|524910744|ref|XP_005110246.1|	4	20	4	4	120
gi|524911947|ref|XP_005110831.1|	4	57	18	373	6727
gi|524911950|ref|XP_005110832.1|	4	53	18	214	4236
gi|524913026|ref|XP_005111347.1|	4	14	5	8	194
gi|524913132|ref|XP_005111399.1|	4	29	4	7	83
gi|524913755|ref|XP_005111702.1|	4	31	4	4	125
gi|524913929|ref|XP_005111786.1|	4	6	4	6	86
gi|524914499|ref|XP_005112064.1|	4	8	4	4	58
gi|524914928|ref|XP_005112273.1|	4	16	4	9	326
gi|524915405|ref|XP_005112503.1|	4	37	4	20	443
gi|529444582|ref|XP_005242003.1|	4	25	4	6	94
gi|544604255|sp|Q3YLA4.2|NCS1_LYMST	4	27	4	6	183
gi|554529929|ref|XP_005859764.1|	4	47	4	7	98
gi|555704960|gb|ESO08193.1|	4	46	4	24	469
gi|556094506|gb|ESO83159.1|	4	21	4	5	74
gi|556094522|gb|ESO83175.1|	4	9	4	5	100
gi|556097226|gb|ESO85878.1|	4	12	4	4	64

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|556097653|gb|ESO86305.1|	4	11	4	5	78
gi|556098530|gb|ESO87182.1|	4	23	4	4	81
gi|556101476|gb|ESO90128.1|	4	10	5	6	159
gi|556102412|gb|ESO91064.1|	4	13	4	15	583
gi|556105643|gb|ESO94295.1|	4	32	4	6	121
gi|556105791|gb|ESO94443.1|	4	6	4	6	103
gi|556106190|gb|ESO94842.1|	4	19	6	16	286
gi|556106190|gb|ESO94842.1|	4	32	9	21	489
gi|556110210|gb|ESO98862.1|	4	22	4	4	61
gi|556110324|gb|ESO98976.1|	4	9	4	4	43
gi|556111239|gb|ESO99890.1|	4	13	4	7	206
gi|556113679|gb|ESP02331.1|	4	27	4	6	113
gi|556114019|gb|ESP02671.1|	4	17	4	4	43
gi|556114028|gb|ESP02680.1|	4	33	4	4	95
gi|556114329|gb|ESP02981.1|	4	28	4	8	109
gi|556116442|gb|ESP05094.1|	4	48	4	11	402
gi|568939963|ref|XP_006505278.1|	4	42	4	10	259
gi|589107123|ref|XP_006965584.1|	4	42	4	9	288
gi|619323847|gb|AHY00642.1|	4	10	4	4	65
gi|646720608|gb|KDR22269.1|	4	45	4	9	212
gi|68272049|gb|AAY89300.1|	4	66	10	53	856
gi|71040619|dbj|BAE16013.1|	4	27	4	16	573
gi|100913266|gb|ABF69533.1|	3	15	3	6	47
gi|1169742|sp|P42577.2|FRIS_LYMST	3	57	3	5	90
gi|126697398|gb|ABO26656.1|	3	19	3	6	128
gi|148236249|ref|NP_001087612.1|	3	31	4	4	52
gi|152003983|gb|ABS19815.1|	3	16	12	19	357
gi|156392725|ref|XP_001636198.1|	3	10	3	3	44
gi|170053486|ref|XP_001862696.1|	3	34	3	4	45
gi|184186107|ref|NP_001116975.1|	3	22	4	6	117
gi|195429214|ref|XP_002062659.1|	3	26	3	7	96
gi|22024001|ref|NP_523708.2|	3	24	3	5	110
gi|2493427|sp|Q27312.1|NADA_APLKU	3	11	3	4	78
gi|253317417|gb|ACT22633.1|	3	49	9	208	4145
gi|260784400|ref|XP_002587255.1|	3	8	3	3	64
gi|260813969|ref|XP_002601688.1|	3	7	3	3	68
gi|260828603|ref|XP_002609252.1|	3	63	10	53	1492
gi|281345481|gb|EFB21065.1|	3	69	4	9	150
gi|29378327|gb|AAO83845.1|AF484088_1	3	19	3	4	131
gi|294896502|ref|XP_002775589.1|	3	43	11	20	344
gi|297186099|gb|ADI24337.1|	3	9	3	9	187
gi|325197162|ref|NP_001191425.1|	3	18	3	5	71

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|325209004|ref|NP_001191451.1|	3	17	3	4	97
gi|325296731|ref|NP_001191601.1|	3	11	3	4	103
gi|325296853|ref|NP_001191458.1|	3	26	4	6	134
gi|325297011|ref|NP_001191514.1|	3	16	4	9	144
gi|38043955|emb|CAD83837.1|	3	32	4	32	910
gi|405950236|gb|EKC18235.1|	3	11	3	3	33
gi|405951631|gb|EKC19528.1|	3	20	3	7	276
gi|405958669|gb|EKC24774.1|	3	12	4	7	190
gi|405965644|gb|EKC31006.1|	3	21	6	7	170
gi|405967885|gb|EKC33004.1|	3	14	3	5	87
gi|405975533|gb|EKC40092.1|	3	14	3	5	65
gi|405976805|gb|EKC41289.1|	3	6	3	4	33
gi|405977577|gb|EKC42020.1|	3	13	3	4	77
gi|405978690|gb|EKC43060.1|	3	7	3	3	118
gi|42560362|gb|AAS20336.1|	3	50	7	438	8101
gi|443694956|gb|ELT95974.1|	3	30	3	5	85
gi|443706726|gb|ELU02640.1|	3	10	3	4	62
gi|443729373|gb|ELU15297.1|	3	19	4	13	96
gi|470643071|ref|XP_004326123.1|	3	82	13	87	2069
gi|47202433|emb|CAF87387.1|	3	48	3	3	32
gi|50593062|gb|AAT79342.1|	3	45	3	14	481
gi|524863982|ref|XP_005088830.1|	3	38	3	13	210
gi|524864475|ref|XP_005089073.1|	3	16	3	3	56
gi|524864482|ref|XP_005089076.1|	3	10	3	6	130
gi|524864620|ref|XP_005089145.1|	3	21	3	7	163
gi|524864626|ref|XP_005089148.1|	3	12	3	5	124
gi|524864665|ref|XP_005089167.1|	3	18	3	16	181
gi|524864840|ref|XP_005089254.1|	3	12	3	3	80
gi|524865955|ref|XP_005089803.1|	3	28	4	11	124
gi|524866007|ref|XP_005089828.1|	3	4	3	4	142
gi|524866201|ref|XP_005089924.1|	3	5	3	3	47
gi|524866665|ref|XP_005090151.1|	3	31	3	8	171
gi|524866836|ref|XP_005090234.1|	3	27	3	7	270
gi|524866978|ref|XP_005090304.1|	3	42	3	10	267
gi|524867383|ref|XP_005090503.1|	3	9	3	3	50
gi|524867640|ref|XP_005090628.1|	3	26	3	6	155
gi|524868832|ref|XP_005091215.1|	3	15	3	4	109
gi|524868942|ref|XP_005091269.1|	3	42	3	3	66
gi|524868994|ref|XP_005091294.1|	3	12	3	6	252
gi|524869145|ref|XP_005091369.1|	3	41	3	9	211
gi|524869207|ref|XP_005091400.1|	3	25	4	6	132
gi|524869284|ref|XP_005091437.1|	3	17	3	11	201

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524869323|ref|XP_005091456.1|	3	9	3	3	70
gi|524869393|ref|XP_005091491.1|	3	28	3	8	187
gi|524869633|ref|XP_005091608.1|	3	11	3	4	77
gi|524870217|ref|XP_005091894.1|	3	19	3	5	111
gi|524870317|ref|XP_005091943.1|	3	7	3	5	90
gi|524870720|ref|XP_005092142.1|	3	14	3	5	248
gi|524870819|ref|XP_005092190.1|	3	6	3	3	53
gi|524870833|ref|XP_005092197.1|	3	12	3	4	134
gi|524870909|ref|XP_005092234.1|	3	11	3	3	55
gi|524871469|ref|XP_005092507.1|	3	7	3	3	41
gi|524871930|ref|XP_005092735.1|	3	25	3	5	85
gi|524872186|ref|XP_005092857.1|	3	10	3	3	137
gi|524872446|ref|XP_005092985.1|	3	16	3	4	77
gi|524872672|ref|XP_005093096.1|	3	55	3	5	99
gi|524872727|ref|XP_005093121.1|	3	28	3	3	58
gi|524872775|ref|XP_005093144.1|	3	6	3	3	62
gi|524872805|ref|XP_005093159.1|	3	18	3	6	140
gi|524872934|ref|XP_005093222.1|	3	19	3	3	53
gi|524873075|ref|XP_005093291.1|	3	21	3	10	184
gi|524873194|ref|XP_005093348.1|	3	6	3	3	67
gi|524873970|ref|XP_005093727.1|	3	43	3	5	91
gi|524873980|ref|XP_005093732.1|	3	9	3	3	58
gi|524874037|ref|XP_005093760.1|	3	31	3	6	108
gi|524874771|ref|XP_005094120.1|	3	15	3	3	101
gi|524874903|ref|XP_005094185.1|	3	4	3	3	66
gi|524875325|ref|XP_005094393.1|	3	26	3	8	115
gi|524875560|ref|XP_005094506.1|	3	14	3	8	169
gi|524876140|ref|XP_005094789.1|	3	14	3	3	62
gi|524876525|ref|XP_005094978.1|	3	16	3	3	124
gi|524876525|ref|XP_005094978.1|	3	35	10	19	374
gi|524876601|ref|XP_005095013.1|	3	14	3	3	72
gi|524876636|ref|XP_005095029.1|	3	12	3	4	50
gi|524876864|ref|XP_005095141.1|	3	13	3	11	207
gi|524877098|ref|XP_005095256.1|	3	32	5	21	197
gi|524877582|ref|XP_005095490.1|	3	10	3	4	78
gi|524877853|ref|XP_005095623.1|	3	41	6	17	283
gi|524878217|ref|XP_005095797.1|	3	14	3	6	114
gi|524878504|ref|XP_005095936.1|	3	85	6	6	153
gi|524878598|ref|XP_005095982.1|	3	17	3	5	94
gi|524878653|ref|XP_005096009.1|	3	37	3	11	188
gi|524878707|ref|XP_005096036.1|	3	9	4	4	47
gi|524878983|ref|XP_005096171.1|	3	32	8	10	96

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524879487|ref|XP_005096418.1|	3	21	3	4	53
gi|524879771|ref|XP_005096557.1|	3	30	10	11	253
gi|524879943|ref|XP_005096641.1|	3	24	3	3	52
gi|524879957|ref|XP_005096647.1|	3	19	3	6	259
gi|524880074|ref|XP_005096705.1|	3	22	3	3	35
gi|524880570|ref|XP_005096947.1|	3	30	3	4	62
gi|524880608|ref|XP_005096966.1|	3	13	4	7	232
gi|524880769|ref|XP_005097043.1|	3	15	3	5	112
gi|524880783|ref|XP_005097050.1|	3	20	3	3	47
gi|524880992|ref|XP_005097152.1|	3	8	3	3	44
gi|524881473|ref|XP_005097388.1|	3	32	4	7	153
gi|524881555|ref|XP_005097428.1|	3	25	5	16	557
gi|524881765|ref|XP_005097532.1|	3	17	3	4	75
gi|524881775|ref|XP_005097537.1|	3	7	3	3	74
gi|524882193|ref|XP_005097741.1|	3	28	3	8	180
gi|524882422|ref|XP_005097855.1|	3	72	6	12	146
gi|524882784|ref|XP_005098028.1|	3	12	3	3	109
gi|524882922|ref|XP_005098097.1|	3	12	3	3	115
gi|524883088|ref|XP_005098178.1|	3	37	9	15	204
gi|524883388|ref|XP_005098325.1|	3	20	3	12	214
gi|524883534|ref|XP_005098394.1|	3	37	3	5	56
gi|524883622|ref|XP_005098438.1|	3	6	3	3	80
gi|524883711|ref|XP_005098482.1|	3	17	3	5	156
gi|524883861|ref|XP_005098556.1|	3	34	5	8	88
gi|524883986|ref|XP_005098618.1|	3	16	3	3	61
gi|524884097|ref|XP_005098672.1|	3	9	3	4	155
gi|524884410|ref|XP_005098825.1|	3	15	3	5	97
gi|524884624|ref|XP_005098931.1|	3	41	4	11	249
gi|524884820|ref|XP_005099026.1|	3	24	3	10	200
gi|524884828|ref|XP_005099030.1|	3	9	3	5	119
gi|524885349|ref|XP_005099283.1|	3	34	5	19	248
gi|524885795|ref|XP_005099501.1|	3	68	34	165	4075
gi|524886285|ref|XP_005099739.1|	3	34	3	6	92
gi|524886568|ref|XP_005099877.1|	3	41	3	8	167
gi|524887166|ref|XP_005100172.1|	3	27	4	7	81
gi|524887212|ref|XP_005100194.1|	3	16	3	3	32
gi|524887216|ref|XP_005100196.1|	3	22	3	3	149
gi|524887218|ref|XP_005100197.1|	3	28	8	10	251
gi|524887218|ref|XP_005100197.1|	3	18	9	11	304
gi|524887498|ref|XP_005100334.1|	3	7	3	4	145
gi|524888025|ref|XP_005100590.1|	3	28	3	5	79
gi|524888089|ref|XP_005100621.1|	3	15	3	4	112

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524888237|ref|XP_005100694.1|	3	23	3	7	214
gi|524888508|ref|XP_005100824.1|	3	62	3	19	271
gi|524888744|ref|XP_005100937.1|	3	13	3	3	88
gi|524889396|ref|XP_005101256.1|	3	10	3	3	39
gi|524889459|ref|XP_005101285.1|	3	9	3	3	86
gi|524889607|ref|XP_005101358.1|	3	13	3	5	66
gi|524889661|ref|XP_005101385.1|	3	10	3	8	117
gi|524889764|ref|XP_005101435.1|	3	33	3	10	124
gi|524890281|ref|XP_005101688.1|	3	18	3	4	107
gi|524890561|ref|XP_005101826.1|	3	7	3	4	68
gi|524890899|ref|XP_005101991.1|	3	11	3	3	98
gi|524891407|ref|XP_005102237.1|	3	10	3	4	63
gi|524891741|ref|XP_005102403.1|	3	14	3	7	121
gi|524891741|ref|XP_005102403.1|	3	11	3	3	36
gi|524892029|ref|XP_005102546.1|	3	6	3	3	90
gi|524892053|ref|XP_005102558.1|	3	29	3	3	78
gi|524892143|ref|XP_005102602.1|	3	7	3	5	101
gi|524892201|ref|XP_005102630.1|	3	49	4	4	57
gi|524892590|ref|XP_005102822.1|	3	14	5	10	201
gi|524893085|ref|XP_005103063.1|	3	15	3	4	97
gi|524893098|ref|XP_005103069.1|	3	22	3	20	284
gi|524893513|ref|XP_005103269.1|	3	5	3	3	73
gi|524893558|ref|XP_005103288.1|	3	10	3	3	73
gi|524893741|ref|XP_005103379.1|	3	30	4	9	112
gi|524894028|ref|XP_005103521.1|	3	10	3	4	100
gi|524894498|ref|XP_005103746.1|	3	13	3	3	42
gi|524894785|ref|XP_005103884.1|	3	4	3	3	64
gi|524894862|ref|XP_005103922.1|	3	13	5	7	106
gi|524895398|ref|XP_005104182.1|	3	9	3	3	60
gi|524895637|ref|XP_005104300.1|	3	12	3	8	135
gi|524895763|ref|XP_005104363.1|	3	16	3	3	58
gi|524896189|ref|XP_005104572.1|	3	13	4	4	82
gi|524896250|ref|XP_005104602.1|	3	14	3	4	139
gi|524896597|ref|XP_005104771.1|	3	20	3	4	80
gi|524896989|ref|XP_005104963.1|	3	18	3	4	64
gi|524897931|ref|XP_005105420.1|	3	6	3	3	50
gi|524898101|ref|XP_005105504.1|	3	10	3	3	80
gi|524898135|ref|XP_005105521.1|	3	2	3	4	94
gi|524898175|ref|XP_005105540.1|	3	7	3	6	179
gi|524898265|ref|XP_005105584.1|	3	19	3	4	112
gi|524898272|ref|XP_005105587.1|	3	24	3	11	177
gi|524898539|ref|XP_005105717.1|	3	14	3	5	91

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524898674|ref|XP_005105783.1|	3	15	3	4	72
gi|524898763|ref|XP_005105827.1|	3	14	3	3	56
gi|524898815|ref|XP_005105852.1|	3	21	3	8	102
gi|524899012|ref|XP_005105947.1|	3	13	3	4	92
gi|524899021|ref|XP_005105951.1|	3	17	3	8	185
gi|524899023|ref|XP_005105952.1|	3	22	3	40	1669
gi|524899511|ref|XP_005106186.1|	3	70	10	10	98
gi|524899675|ref|XP_005106266.1|	3	12	3	3	38
gi|524899675|ref|XP_005106266.1|	3	17	3	3	57
gi|524899892|ref|XP_005106372.1|	3	21	3	3	46
gi|524900236|ref|XP_005106539.1|	3	17	3	4	119
gi|524900270|ref|XP_005106556.1|	3	32	3	8	144
gi|524900353|ref|XP_005106597.1|	3	11	3	4	176
gi|524900538|ref|XP_005106689.1|	3	7	3	4	121
gi|524901394|ref|XP_005107104.1|	3	23	4	10	366
gi|524901619|ref|XP_005107215.1|	3	4	3	3	80
gi|524901901|ref|XP_005107354.1|	3	18	3	6	139
gi|524902714|ref|XP_005107585.1|	3	14	3	3	31
gi|524903041|ref|XP_005107738.1|	3	23	4	13	290
gi|524903247|ref|XP_005107829.1|	3	12	3	4	87
gi|524903608|ref|XP_005108000.1|	3	16	3	8	132
gi|524905947|ref|XP_005108331.1|	3	26	3	6	129
gi|524905999|ref|XP_005108354.1|	3	18	3	7	165
gi|524906745|ref|XP_005108595.1|	3	31	5	10	163
gi|524907032|ref|XP_005108645.1|	3	53	4	6	184
gi|524907580|ref|XP_005108906.1|	3	3	4	5	104
gi|524907590|ref|XP_005108911.1|	3	18	3	8	302
gi|524907727|ref|XP_005108978.1|	3	13	3	3	63
gi|524908048|ref|XP_005109132.1|	3	5	3	4	69
gi|524908261|ref|XP_005109234.1|	3	13	3	5	144
gi|524908261|ref|XP_005109234.1|	3	11	3	6	120
gi|524908282|ref|XP_005109244.1|	3	35	3	14	376
gi|524908297|ref|XP_005109251.1|	3	11	3	3	61
gi|524908577|ref|XP_005109382.1|	3	9	3	3	68
gi|524909270|ref|XP_005109684.1|	3	13	3	4	129
gi|524909457|ref|XP_005109775.1|	3	8	3	5	156
gi|524910500|ref|XP_005110126.1|	3	3	3	3	39
gi|524910662|ref|XP_005110206.1|	3	22	3	3	45
gi|524910831|ref|XP_005110289.1|	3	33	3	6	77
gi|524910950|ref|XP_005110347.1|	3	9	3	4	63
gi|524911110|ref|XP_005110424.1|	3	18	3	5	115
gi|524911477|ref|XP_005110601.1|	3	53	3	8	311

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524911650|ref|XP_005110685.1|	3	11	3	5	173
gi|524912055|ref|XP_005110881.1|	3	7	3	4	112
gi|524912187|ref|XP_005110945.1|	3	10	3	3	73
gi|524912355|ref|XP_005111027.1|	3	15	3	5	84
gi|524912927|ref|XP_005111299.1|	3	43	3	9	198
gi|524913184|ref|XP_005111424.1|	3	20	3	6	158
gi|524913273|ref|XP_005111468.1|	3	26	3	9	286
gi|524913662|ref|XP_005111656.1|	3	14	3	4	86
gi|524913668|ref|XP_005111659.1|	3	20	3	6	153
gi|524913879|ref|XP_005111762.1|	3	7	3	5	173
gi|524913883|ref|XP_005111764.1|	3	25	3	6	58
gi|524914146|ref|XP_005111892.1|	3	5	3	4	67
gi|524914287|ref|XP_005111960.1|	3	28	3	4	112
gi|524914353|ref|XP_005111992.1|	3	13	3	4	85
gi|524914806|ref|XP_005112213.1|	3	7	3	5	75
gi|524914928|ref|XP_005112273.1|	3	10	3	4	79
gi|524915372|ref|XP_005112487.1|	3	17	3	3	60
gi|524915376|ref|XP_005112489.1|	3	15	4	4	96
gi|524916570|ref|XP_005113059.1|	3	62	8	23	262
gi|524916601|ref|XP_005113074.1|	3	22	4	9	100
gi|524916601|ref|XP_005113074.1|	3	24	6	11	254
gi|524916639|ref|XP_005113093.1|	3	20	3	7	159
gi|524916644|ref|XP_005113095.1|	3	13	3	10	164
gi|524916993|ref|XP_005113263.1|	3	8	3	3	96
gi|526117431|ref|NP_001267767.1|	3	31	5	86	1230
gi|556094374|gb|ESO83027.1|	3	13	3	3	74
gi|556097689|gb|ESO86341.1|	3	17	3	3	46
gi|556097805|gb|ESO86457.1|	3	3	3	3	65
gi|556098509|gb|ESO87161.1|	3	7	3	4	156
gi|556098668|gb|ESO87320.1|	3	13	3	5	171
gi|556098954|gb|ESO87606.1|	3	2	3	4	101
gi|556101818|gb|ESO90470.1|	3	9	3	4	57
gi|556102645|gb|ESO91297.1|	3	7	3	7	146
gi|556103976|gb|ESO92628.1|	3	19	3	3	29
gi|556105206|gb|ESO93858.1|	3	20	3	6	120
gi|556106320|gb|ESO94972.1|	3	19	3	5	85
gi|556106462|gb|ESO95114.1|	3	5	3	3	146
gi|556107134|gb|ESO95786.1|	3	10	3	4	76
gi|556107378|gb|ESO96030.1|	3	17	3	4	93
gi|556108215|gb|ESO96867.1|	3	8	3	3	74
gi|556108856|gb|ESO97508.1|	3	8	3	3	56
gi|556108856|gb|ESO97508.1|	3	8	3	4	54

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|556108929|gb|ESO97581.1|	3	34	3	3	38
gi|556109606|gb|ESO98258.1|	3	11	3	8	152
gi|556110532|gb|ESO99183.1|	3	12	3	6	137
gi|556110866|gb|ESO99517.1|	3	53	40	177	4066
gi|556113064|gb|ESP01716.1|	3	12	3	4	127
gi|556113650|gb|ESP02302.1|	3	37	3	3	109
gi|556114587|gb|ESP03239.1|	3	8	3	3	30
gi|556116273|gb|ESP04925.1|	3	24	3	3	72
gi|577029965|gb|AHH81817.1|	3	40	3	5	61
gi|583994656|ref|XP_006792469.1|	3	28	3	32	436
gi|602642710|ref|XP_007428140.1|	3	49	3	6	153
gi|608096069|gb|EZG34785.1|	3	17	3	13	302
gi|632935059|ref|XP_007887607.1|	3	14	3	6	124
gi|635065029|ref|XP_008001282.1|	3	62	4	27	325
gi|6492342|gb|AAF14287.1|AF182339_1	3	7	3	3	63
gi|67005832|gb|AAN31641.3|AF486290_1	3	10	3	3	86
gi|68533593|gb|AAH98579.1|	3	6	3	3	88
gi|72084055|ref|XP_790644.1|	3	11	3	3	92
gi|93209532|gb|ABF00124.1|	3	27	3	79	1692
gi|109157294|pdb|2CE6|A	2	29	2	2	34
gi|110764763|ref|XP_001123191.1|	2	44	2	8	239
gi|1174755|sp|P43689.1|TPM2_BIOGL	2	58	23	33	515
gi|118401762|ref|XP_001033201.1|	2	18	2	4	81
gi|126697272|gb|ABO26593.1|	2	21	3	5	81
gi|126697410|gb|ABO26662.1|	2	16	2	4	102
gi|149032103|gb|EDL87015.1|	2	41	9	53	1271
gi|150251392|gb|ABR68008.1|	2	35	5	9	178
gi|156373963|ref|XP_001629579.1|	2	8	2	2	27
gi|198285569|gb|ACH85323.1|	2	7	2	2	49
gi|198430879|ref|XP_002123222.1|	2	38	2	4	188
gi|2073146|dbj|BAA19863.1|	2	16	2	2	57
gi|220979902|emb|CAQ64775.1|	2	44	7	9	186
gi|260807417|ref|XP_002598505.1|	2	17	2	2	36
gi|260811516|ref|XP_002600468.1|	2	16	2	3	73
gi|27763677|gb|AAO20109.1|	2	9	2	2	40
gi|291234885|ref|XP_002737379.1|	2	29	2	5	100
gi|291242935|ref|XP_002741349.1|	2	20	2	2	30
gi|291245044|ref|XP_002742403.1|	2	8	2	3	65
gi|293595779|gb|ADE45333.1|	2	14	2	3	128
gi|29378329|gb|AAO83846.1|AF484089_1	2	15	2	4	80
gi|29378333|gb|AAO83848.1|AF484091_1	2	27	4	23	845
gi|29378335|gb|AAO83849.1|AF484092_1	2	4	2	2	55

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|297186099|gb|ADI24337.1|	2	15	2	2	50
gi|323134763|gb|ADX31293.1|	2	14	2	4	84
gi|325197143|ref|NP_001191422.1|	2	17	2	4	81
gi|325296789|ref|NP_001191633.1|	2	5	2	2	48
gi|325296853|ref|NP_001191458.1|	2	24	5	5	114
gi|325296855|ref|NP_001191459.1|	2	24	4	5	114
gi|325297015|ref|NP_001191515.1|	2	15	5	5	101
gi|325297068|ref|NP_001191538.1|	2	8	2	2	62
gi|328905044|gb|AEB54797.1|	2	11	2	2	30
gi|340381720|ref|XP_003389369.1|	2	16	2	11	168
gi|346721859|gb|AEO50742.1|	2	14	3	7	118
gi|346987840|gb|AEO51766.1|	2	13	2	8	185
gi|346987844|gb|AEO51768.1|	2	38	65	120	2074
gi|374718481|gb|AEZ67032.1|	2	5	2	4	89
gi|395484809|gb|AFN66649.1|	2	5	2	2	41
gi|397776464|gb|AFO64930.1|	2	27	3	5	110
gi|403310251|emb|CCJ09600.1|	2	6	2	3	70
gi|405952171|gb|EKC20013.1|	2	6	2	3	114
gi|405952830|gb|EKC20593.1|	2	5	2	2	82
gi|405955268|gb|EKC22447.1|	2	3	2	2	48
gi|405957390|gb|EKC23604.1|	2	18	2	3	68
gi|405957680|gb|EKC23874.1|	2	6	2	2	36
gi|405958669|gb|EKC24774.1|	2	8	2	3	60
gi|405960075|gb|EKC26025.1|	2	5	2	5	124
gi|405960381|gb|EKC26308.1|	2	35	3	8	224
gi|405960381|gb|EKC26308.1|	2	35	3	8	160
gi|405960599|gb|EKC26510.1|	2	24	2	2	28
gi|405960672|gb|EKC26570.1|	2	26	2	6	118
gi|405962100|gb|EKC27804.1|	2	4	2	2	34
gi|405970416|gb|EKC35324.1|	2	3	2	3	81
gi|405970417|gb|EKC35325.1|	2	4	2	2	32
gi|405970417|gb|EKC35325.1|	2	6	2	2	47
gi|405972817|gb|EKC37565.1|	2	54	2	6	117
gi|405972881|gb|EKC37628.1|	2	10	2	5	82
gi|405974177|gb|EKC38843.1|	2	12	2	2	58
gi|405977265|gb|EKC41724.1|	2	5	2	2	45
gi|405977876|gb|EKC42303.1|	2	12	2	2	98
gi|42559558|sp|O97192.1|TPM_HELAS	2	24	8	9	130
gi|443720976|gb|ELU10481.1|	2	9	2	2	50
gi|4504279|ref|NP_002098.1|	2	15	3	12	95
gi|459180352|ref|XP_004226627.1|	2	64	18	99	2063
gi|46277000|gb|AAS86696.1|	2	6	2	2	75

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|46277106|gb|AAS86749.1|	2	22	2	3	27
gi|4929136|gb|AAD33872.1|AF141348_1	2	23	4	5	114
gi|51105030|gb|AAT97075.1|	2	17	2	5	77
gi|51105058|gb|AAT97089.1|	2	45	12	19	518
gi|511880705|ref|XP_004760254.1|	2	27	3	3	33
gi|524864102|ref|XP_005088889.1|	2	20	2	2	32
gi|524864130|ref|XP_005088903.1|	2	14	2	6	211
gi|524864189|ref|XP_005088932.1|	2	13	2	3	77
gi|524864189|ref|XP_005088932.1|	2	18	2	2	58
gi|524864582|ref|XP_005089126.1|	2	10	2	3	64
gi|524864630|ref|XP_005089150.1|	2	32	2	7	111
gi|524864630|ref|XP_005089150.1|	2	11	2	2	33
gi|524864893|ref|XP_005089280.1|	2	11	2	3	112
gi|524865136|ref|XP_005089400.1|	2	11	2	3	67
gi|524865144|ref|XP_005089404.1|	2	12	2	2	35
gi|524865614|ref|XP_005089635.1|	2	9	2	2	85
gi|524865813|ref|XP_005089733.1|	2	23	4	7	109
gi|524865869|ref|XP_005089760.1|	2	12	2	2	45
gi|524866201|ref|XP_005089924.1|	2	13	2	3	54
gi|524866264|ref|XP_005089955.1|	2	8	2	2	39
gi|524866798|ref|XP_005090215.1|	2	14	2	2	75
gi|524866978|ref|XP_005090304.1|	2	45	3	8	183
gi|524866999|ref|XP_005090313.1|	2	8	2	2	51
gi|524867173|ref|XP_005090399.1|	2	8	2	2	36
gi|524867407|ref|XP_005090515.1|	2	8	2	2	40
gi|524867567|ref|XP_005090594.1|	2	10	2	2	75
gi|524867755|ref|XP_005090685.1|	2	9	6	6	183
gi|524867763|ref|XP_005090689.1|	2	9	5	5	182
gi|524867882|ref|XP_005090748.1|	2	27	2	7	159
gi|524868005|ref|XP_005090809.1|	2	47	3	8	130
gi|524868127|ref|XP_005090870.1|	2	9	2	2	74
gi|524868280|ref|XP_005090944.1|	2	4	2	2	55
gi|524868395|ref|XP_005091000.1|	2	6	2	2	92
gi|524868850|ref|XP_005091223.1|	2	10	2	3	88
gi|524868888|ref|XP_005091242.1|	2	52	3	5	152
gi|524868962|ref|XP_005091279.1|	2	6	2	2	22
gi|524868983|ref|XP_005091289.1|	2	14	3	3	80
gi|524869296|ref|XP_005091443.1|	2	11	2	3	87
gi|524869517|ref|XP_005091551.1|	2	8	2	4	55
gi|524869549|ref|XP_005091567.1|	2	2	2	2	34
gi|524869798|ref|XP_005091688.1|	2	8	2	3	77
gi|524870116|ref|XP_005091844.1|	2	19	2	5	124

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524870146|ref|XP_005091859.1|	2	28	3	3	67
gi|524870213|ref|XP_005091892.1|	2	35	2	19	365
gi|524870306|ref|XP_005091938.1|	2	4	2	2	30
gi|524870519|ref|XP_005092044.1|	2	5	2	2	42
gi|524870720|ref|XP_005092142.1|	2	9	2	2	69
gi|524870754|ref|XP_005092159.1|	2	19	2	2	42
gi|524870758|ref|XP_005092161.1|	2	8	2	2	56
gi|524871250|ref|XP_005092401.1|	2	27	2	4	96
gi|524871258|ref|XP_005092404.1|	2	5	2	3	87
gi|524871394|ref|XP_005092470.1|	2	14	2	4	87
gi|524871451|ref|XP_005092498.1|	2	6	2	2	60
gi|524871596|ref|XP_005092570.1|	2	2	2	3	133
gi|524871599|ref|XP_005092571.1|	2	7	2	2	66
gi|524871635|ref|XP_005092589.1|	2	9	2	3	99
gi|524872452|ref|XP_005092988.1|	2	18	2	4	78
gi|524872578|ref|XP_005093050.1|	2	4	2	3	107
gi|524872604|ref|XP_005093063.1|	2	6	2	3	60
gi|524872801|ref|XP_005093157.1|	2	51	9	20	345
gi|524872910|ref|XP_005093210.1|	2	11	2	3	116
gi|524873152|ref|XP_005093328.1|	2	4	2	2	45
gi|524873501|ref|XP_005093496.1|	2	3	2	2	50
gi|524873892|ref|XP_005093688.1|	2	37	2	7	311
gi|524873954|ref|XP_005093719.1|	2	6	2	2	67
gi|524874182|ref|XP_005093830.1|	2	14	2	2	47
gi|524874186|ref|XP_005093832.1|	2	10	2	2	38
gi|524874190|ref|XP_005093834.1|	2	7	2	2	45
gi|524874352|ref|XP_005093912.1|	2	3	2	2	70
gi|524874629|ref|XP_005094049.1|	2	12	2	2	59
gi|524874903|ref|XP_005094185.1|	2	5	2	4	57
gi|524875141|ref|XP_005094301.1|	2	16	2	3	70
gi|524875173|ref|XP_005094317.1|	2	6	2	2	51
gi|524875207|ref|XP_005094334.1|	2	12	2	3	121
gi|524875757|ref|XP_005094602.1|	2	3	2	2	48
gi|524876116|ref|XP_005094777.1|	2	15	2	2	49
gi|524876138|ref|XP_005094788.1|	2	5	2	2	40
gi|524876523|ref|XP_005094977.1|	2	47	4	4	84
gi|524876525|ref|XP_005094978.1|	2	32	4	7	187
gi|524876618|ref|XP_005095021.1|	2	8	2	3	51
gi|524876668|ref|XP_005095045.1|	2	13	2	3	79
gi|524876816|ref|XP_005095117.1|	2	6	2	2	74
gi|524876998|ref|XP_005095207.1|	2	4	2	2	37
gi|524877372|ref|XP_005095389.1|	2	30	2	3	48

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524877614|ref|XP_005095506.1|	2	8	4	7	129
gi|524877675|ref|XP_005095535.1|	2	3	2	2	70
gi|524877793|ref|XP_005095594.1|	2	14	2	3	50
gi|524877803|ref|XP_005095598.1|	2	8	2	2	33
gi|524878219|ref|XP_005095798.1|	2	19	2	3	65
gi|524878295|ref|XP_005095836.1|	2	3	2	5	169
gi|524878368|ref|XP_005095871.1|	2	8	2	2	49
gi|524878410|ref|XP_005095892.1|	2	5	2	2	37
gi|524878707|ref|XP_005096036.1|	2	18	3	3	72
gi|524878727|ref|XP_005096045.1|	2	8	2	2	65
gi|524878727|ref|XP_005096045.1|	2	8	2	4	136
gi|524878917|ref|XP_005096139.1|	2	11	2	9	222
gi|524879087|ref|XP_005096223.1|	2	7	2	2	22
gi|524879355|ref|XP_005096354.1|	2	15	2	2	38
gi|524879751|ref|XP_005096547.1|	2	4	2	3	77
gi|524879821|ref|XP_005096582.1|	2	9	2	3	52
gi|524880074|ref|XP_005096705.1|	2	17	2	11	171
gi|524880525|ref|XP_005096925.1|	2	6	2	3	45
gi|524880698|ref|XP_005097010.1|	2	5	2	2	55
gi|524880806|ref|XP_005097061.1|	2	7	2	5	88
gi|524880862|ref|XP_005097089.1|	2	12	2	3	50
gi|524880971|ref|XP_005097142.1|	2	7	2	2	24
gi|524880980|ref|XP_005097146.1|	2	6	2	3	59
gi|524881358|ref|XP_005097331.1|	2	24	2	3	46
gi|524881413|ref|XP_005097358.1|	2	17	3	7	204
gi|524881487|ref|XP_005097395.1|	2	7	2	2	52
gi|524881587|ref|XP_005097444.1|	2	4	2	5	69
gi|524881805|ref|XP_005097552.1|	2	12	2	5	69
gi|524881845|ref|XP_005097572.1|	2	18	2	2	78
gi|524882014|ref|XP_005097652.1|	2	7	2	2	41
gi|524882249|ref|XP_005097769.1|	2	10	2	2	51
gi|524882640|ref|XP_005097961.1|	2	4	2	2	43
gi|524882665|ref|XP_005097972.1|	2	13	2	2	40
gi|524882721|ref|XP_005097999.1|	2	11	2	3	170
gi|524882733|ref|XP_005098005.1|	2	23	2	4	90
gi|524882750|ref|XP_005098013.1|	2	5	2	2	46
gi|524882784|ref|XP_005098028.1|	2	11	2	5	151
gi|524882812|ref|XP_005098042.1|	2	8	2	2	66
gi|524882956|ref|XP_005098113.1|	2	7	2	2	73
gi|524882968|ref|XP_005098119.1|	2	15	2	2	55
gi|524883074|ref|XP_005098171.1|	2	4	2	2	40
gi|524883090|ref|XP_005098179.1|	2	28	8	12	155

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524883672|ref|XP_005098463.1|	2	5	2	3	78
gi|524883770|ref|XP_005098511.1|	2	4	2	2	95
gi|524883861|ref|XP_005098556.1|	2	34	5	5	46
gi|524883909|ref|XP_005098580.1|	2	21	2	5	44
gi|524884059|ref|XP_005098653.1|	2	6	2	5	168
gi|524884099|ref|XP_005098673.1|	2	13	2	2	65
gi|524884550|ref|XP_005098894.1|	2	6	2	2	51
gi|524884737|ref|XP_005098987.1|	2	10	2	9	249
gi|524884830|ref|XP_005099031.1|	2	6	2	3	33
gi|524884921|ref|XP_005099073.1|	2	5	2	2	57
gi|524885143|ref|XP_005099182.1|	2	9	3	3	93
gi|524885349|ref|XP_005099283.1|	2	51	4	10	78
gi|524885451|ref|XP_005099333.1|	2	7	2	2	36
gi|524885518|ref|XP_005099365.1|	2	11	2	2	60
gi|524885623|ref|XP_005099417.1|	2	34	4	5	48
gi|524885625|ref|XP_005099418.1|	2	63	3	5	87
gi|524885654|ref|XP_005099432.1|	2	9	2	2	35
gi|524885681|ref|XP_005099445.1|	2	6	2	2	27
gi|524885955|ref|XP_005099578.1|	2	6	2	2	66
gi|524886190|ref|XP_005099692.1|	2	3	2	3	72
gi|524886289|ref|XP_005099741.1|	2	11	2	2	30
gi|524886313|ref|XP_005099753.1|	2	10	2	6	141
gi|524886402|ref|XP_005099797.1|	2	4	2	3	70
gi|524886408|ref|XP_005099800.1|	2	12	2	4	42
gi|524886760|ref|XP_005099972.1|	2	19	2	3	42
gi|524887166|ref|XP_005100172.1|	2	18	3	4	54
gi|524887278|ref|XP_005100227.1|	2	6	2	2	56
gi|524887465|ref|XP_005100318.1|	2	30	2	2	42
gi|524887733|ref|XP_005100450.1|	2	21	2	2	28
gi|524888476|ref|XP_005100809.1|	2	24	2	3	131
gi|524888516|ref|XP_005100828.1|	2	4	2	2	59
gi|524888693|ref|XP_005100913.1|	2	5	2	3	24
gi|524888709|ref|XP_005100920.1|	2	9	4	4	46
gi|524888713|ref|XP_005100922.1|	2	5	2	2	89
gi|524888919|ref|XP_005101023.1|	2	9	2	2	40
gi|524889160|ref|XP_005101141.1|	2	18	2	4	108
gi|524889160|ref|XP_005101141.1|	2	32	2	3	72
gi|524889486|ref|XP_005101298.1|	2	16	2	2	48
gi|524889574|ref|XP_005101342.1|	2	8	2	2	53
gi|524889734|ref|XP_005101421.1|	2	18	2	2	50
gi|524890096|ref|XP_005101597.1|	2	23	2	4	141
gi|524890208|ref|XP_005101652.1|	2	43	11	16	271

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524890233|ref|XP_005101664.1|	2	3	2	2	54
gi|524890263|ref|XP_005101679.1|	2	4	2	2	37
gi|524890384|ref|XP_005101739.1|	2	18	2	2	53
gi|524890516|ref|XP_005101804.1|	2	12	2	4	63
gi|524890817|ref|XP_005101950.1|	2	13	2	3	59
gi|524890881|ref|XP_005101982.1|	2	12	2	3	83
gi|524890899|ref|XP_005101991.1|	2	19	2	2	53
gi|524891164|ref|XP_005102118.1|	2	17	2	3	116
gi|524891320|ref|XP_005102195.1|	2	54	2	3	72
gi|524891426|ref|XP_005102246.1|	2	10	2	2	26
gi|524891440|ref|XP_005102253.1|	2	14	2	2	101
gi|524891478|ref|XP_005102272.1|	2	10	2	5	88
gi|524891582|ref|XP_005102324.1|	2	23	2	6	66
gi|524892070|ref|XP_005102566.1|	2	27	2	2	97
gi|524892281|ref|XP_005102670.1|	2	4	4	10	360
gi|524892866|ref|XP_005102957.1|	2	11	2	2	40
gi|524892979|ref|XP_005103011.1|	2	34	2	2	34
gi|524893065|ref|XP_005103053.1|	2	16	2	2	80
gi|524893160|ref|XP_005103100.1|	2	17	3	7	57
gi|524893168|ref|XP_005103104.1|	2	33	2	3	154
gi|524893177|ref|XP_005103108.1|	2	4	2	5	123
gi|524893771|ref|XP_005103394.1|	2	7	2	2	42
gi|524893797|ref|XP_005103407.1|	2	3	2	2	37
gi|524893965|ref|XP_005103490.1|	2	12	2	4	204
gi|524894013|ref|XP_005103514.1|	2	21	6	9	137
gi|524894026|ref|XP_005103520.1|	2	6	2	3	114
gi|524894126|ref|XP_005103566.1|	2	10	2	2	114
gi|524894456|ref|XP_005103725.1|	2	12	2	2	44
gi|524894778|ref|XP_005103881.1|	2	4	2	2	39
gi|524894895|ref|XP_005103937.1|	2	24	3	8	154
gi|524894895|ref|XP_005103937.1|	2	32	3	12	314
gi|524894939|ref|XP_005103959.1|	2	13	2	3	63
gi|524895123|ref|XP_005104049.1|	2	27	6	9	164
gi|524895689|ref|XP_005104326.1|	2	15	2	4	37
gi|524895775|ref|XP_005104369.1|	2	9	2	3	72
gi|524895899|ref|XP_005104430.1|	2	17	2	7	115
gi|524896095|ref|XP_005104525.1|	2	13	2	3	57
gi|524896371|ref|XP_005104660.1|	2	9	2	3	62
gi|524896476|ref|XP_005104712.1|	2	16	2	4	61
gi|524896554|ref|XP_005104750.1|	2	16	2	5	118
gi|524896814|ref|XP_005104878.1|	2	15	3	4	77
gi|524897050|ref|XP_005104992.1|	2	8	2	3	147

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524897076|ref|XP_005105005.1|	2	26	2	3	113
gi|524897447|ref|XP_005105184.1|	2	9	2	2	48
gi|524897481|ref|XP_005105201.1|	2	11	2	2	55
gi|524897756|ref|XP_005105335.1|	2	13	2	2	45
gi|524898422|ref|XP_005105660.1|	2	2	2	2	55
gi|524898581|ref|XP_005105738.1|	2	10	2	2	61
gi|524898743|ref|XP_005105817.1|	2	18	2	2	54
gi|524898799|ref|XP_005105845.1|	2	2	2	2	51
gi|524898815|ref|XP_005105852.1|	2	34	2	2	49
gi|524899018|ref|XP_005105950.1|	2	11	2	2	65
gi|524899677|ref|XP_005106267.1|	2	31	2	9	140
gi|524899715|ref|XP_005106286.1|	2	5	2	2	30
gi|524899892|ref|XP_005106372.1|	2	13	2	6	150
gi|524900256|ref|XP_005106549.1|	2	16	2	3	40
gi|524900316|ref|XP_005106579.1|	2	3	2	2	81
gi|524900332|ref|XP_005106587.1|	2	18	2	3	48
gi|524900386|ref|XP_005106613.1|	2	5	2	2	46
gi|524900555|ref|XP_005106697.1|	2	7	2	2	44
gi|524900577|ref|XP_005106708.1|	2	14	2	2	45
gi|524900723|ref|XP_005106780.1|	2	6	2	3	82
gi|524900911|ref|XP_005106871.1|	2	20	2	3	51
gi|524901283|ref|XP_005107052.1|	2	46	5	7	129
gi|524901394|ref|XP_005107104.1|	2	31	5	7	265
gi|524901519|ref|XP_005107166.1|	2	13	2	2	40
gi|524901565|ref|XP_005107188.1|	2	13	2	5	119
gi|524901877|ref|XP_005107342.1|	2	16	2	2	51
gi|524902425|ref|XP_005107464.1|	2	15	3	3	82
gi|524902702|ref|XP_005107579.1|	2	9	2	3	99
gi|524902931|ref|XP_005107685.1|	2	11	2	3	125
gi|524903184|ref|XP_005107804.1|	2	4	2	2	43
gi|524903262|ref|XP_005107836.1|	2	8	2	3	63
gi|524903740|ref|XP_005108064.1|	2	5	2	3	76
gi|524905586|ref|XP_005108170.1|	2	15	2	4	96
gi|524905668|ref|XP_005108199.1|	2	8	2	4	194
gi|524905977|ref|XP_005108344.1|	2	11	2	2	48
gi|524906238|ref|XP_005108452.1|	2	6	2	2	63
gi|524906293|ref|XP_005108469.1|	2	34	2	5	86
gi|524906398|ref|XP_005108509.1|	2	6	2	2	54
gi|524906745|ref|XP_005108595.1|	2	27	5	5	56
gi|524906862|ref|XP_005108596.1|	2	26	2	3	69
gi|524906990|ref|XP_005108638.1|	2	8	2	3	47
gi|524907177|ref|XP_005108713.1|	2	12	2	4	63

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524907372|ref|XP_005108804.1|	2	9	2	3	69
gi|524907491|ref|XP_005108862.1|	2	34	2	2	40
gi|524907580|ref|XP_005108906.1|	2	7	2	3	97
gi|524907580|ref|XP_005108906.1|	2	7	3	3	80
gi|524907699|ref|XP_005108965.1|	2	14	2	3	50
gi|524907836|ref|XP_005109029.1|	2	37	2	4	132
gi|524908062|ref|XP_005109139.1|	2	25	2	3	79
gi|524908579|ref|XP_005109383.1|	2	4	2	2	70
gi|524908585|ref|XP_005109386.1|	2	13	2	2	44
gi|524908782|ref|XP_005109481.1|	2	6	2	2	26
gi|524908892|ref|XP_005109535.1|	2	6	2	2	43
gi|524909938|ref|XP_005109861.1|	2	7	2	2	37
gi|524909980|ref|XP_005109881.1|	2	18	2	3	106
gi|524910025|ref|XP_005109900.1|	2	18	2	4	102
gi|524910197|ref|XP_005109980.1|	2	8	2	3	174
gi|524910546|ref|XP_005110149.1|	2	16	2	7	213
gi|524911076|ref|XP_005110408.1|	2	2	2	2	59
gi|524911122|ref|XP_005110430.1|	2	14	2	5	112
gi|524911542|ref|XP_005110633.1|	2	5	2	5	200
gi|524911570|ref|XP_005110646.1|	2	6	2	2	73
gi|524911866|ref|XP_005110793.1|	2	15	2	9	186
gi|524912187|ref|XP_005110945.1|	2	13	2	3	91
gi|524912191|ref|XP_005110947.1|	2	42	2	2	63
gi|524912369|ref|XP_005111034.1|	2	8	2	3	168
gi|524912387|ref|XP_005111042.1|	2	8	2	2	36
gi|524912628|ref|XP_005111158.1|	2	6	2	4	82
gi|524912696|ref|XP_005111190.1|	2	16	2	4	42
gi|524913352|ref|XP_005111507.1|	2	17	2	4	57
gi|524913662|ref|XP_005111656.1|	2	11	2	3	57
gi|524913685|ref|XP_005111667.1|	2	9	2	3	97
gi|524913834|ref|XP_005111741.1|	2	9	2	2	74
gi|524913836|ref|XP_005111742.1|	2	2	2	2	41
gi|524914201|ref|XP_005111919.1|	2	15	2	2	46
gi|524914370|ref|XP_005112000.1|	2	13	2	3	44
gi|524914697|ref|XP_005112160.1|	2	11	2	2	38
gi|524914776|ref|XP_005112198.1|	2	7	2	2	25
gi|524914926|ref|XP_005112272.1|	2	8	2	4	151
gi|524915376|ref|XP_005112489.1|	2	16	3	4	96
gi|524915538|ref|XP_005112566.1|	2	23	2	4	87
gi|524916317|ref|XP_005112938.1|	2	24	2	4	194
gi|524916317|ref|XP_005112938.1|	2	16	2	6	86
gi|524916381|ref|XP_005112969.1|	2	17	2	3	39

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524916381|ref|XP_005112969.1|	2	16	2	2	52
gi|524916593|ref|XP_005113070.1|	2	42	10	20	434
gi|524916601|ref|XP_005113074.1|	2	34	3	12	259
gi|524916925|ref|XP_005113230.1|	2	5	2	3	46
gi|524917226|ref|XP_005113374.1|	2	7	2	2	26
gi|526117417|ref|NP_001267763.1|	2	4	2	3	53
gi|526117443|ref|NP_001267770.1|	2	9	2	6	125
gi|541950160|ref|XP_005432023.1|	2	40	7	28	557
gi|546679302|gb|ERL89789.1|	2	24	2	2	64
gi|555706265|gb|ESO09498.1|	2	13	2	6	113
gi|556093608|gb|ESO82261.1|	2	10	2	2	50
gi|556093971|gb|ESO82624.1|	2	16	2	5	95
gi|556095217|gb|ESO83870.1|	2	13	2	4	107
gi|556095970|gb|ESO84622.1|	2	11	2	4	140
gi|556097226|gb|ESO85878.1|	2	5	2	2	39
gi|556097346|gb|ESO85998.1|	2	20	2	2	63
gi|556098140|gb|ESO86792.1|	2	38	2	8	193
gi|556098590|gb|ESO87242.1|	2	24	2	3	150
gi|556101476|gb|ESO90128.1|	2	5	3	5	150
gi|556101533|gb|ESO90185.1|	2	17	2	3	59
gi|556101906|gb|ESO90558.1|	2	6	2	6	116
gi|556102306|gb|ESO90958.1|	2	14	2	2	41
gi|556105877|gb|ESO94529.1|	2	25	2	6	172
gi|556106462|gb|ESO95114.1|	2	22	2	3	97
gi|556106462|gb|ESO95114.1|	2	9	2	3	61
gi|556107128|gb|ESO95780.1|	2	20	2	2	44
gi|556108350|gb|ESO97002.1|	2	3	2	3	70
gi|556109892|gb|ESO98544.1|	2	12	2	2	44
gi|556111890|gb|ESP00542.1|	2	13	2	2	58
gi|556112256|gb|ESP00908.1|	2	8	2	2	53
gi|556112262|gb|ESP00914.1|	2	42	2	2	39
gi|556113060|gb|ESP01712.1|	2	2	2	2	32
gi|556114398|gb|ESP03050.1|	2	14	2	4	198
gi|556114716|gb|ESP03368.1|	2	4	2	2	40
gi|556115870|gb|ESP04522.1|	2	7	2	4	115
gi|556116273|gb|ESP04925.1|	2	9	2	2	44
gi|577029901|gb|AHH81785.1|	2	74	3	9	279
gi|585188|sp|P38411.1|GNAQ_LYMST	2	25	2	2	49
gi|632935059|ref|XP_007887607.1|	2	30	2	2	32
gi|632968605|ref|XP_007900618.1|	2	18	2	2	58
gi|637321743|ref|XP_008112910.1|	2	28	2	3	84
gi|67989656|gb|AAY84715.1|	2	7	2	7	293

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|71726733|gb|AAZ39530.1|	2	45	2	13	340
gi|77628181|gb|ABB00056.1|	2	71	3	36	688
gi|85362949|gb|ABC69854.1|	2	28	2	9	206
gi|91079426|ref|XP_967808.1|	2	13	2	5	70
gi|1174755|sp|P43689.1|TPM2_BIOGL	1	14	3	4	34
gi|148724316|gb|ABR08403.1|	1	16	2	7	284
gi|170596696|ref|XP_001902862.1|	1	30	3	21	589
gi|189308122|gb|ACD86945.1|	1	70	7	273	4574
gi|21388656|dbj|BAC00784.1|	1	35	34	37	701
gi|29378333|gb|AAO83848.1|AF484091_1	1	30	4	15	499
gi|294896502|ref|XP_002775589.1|	1	71	5	10	204
gi|302139997|gb|ADK94900.1|	1	86	6	32	523
gi|325296803|ref|NP_001191642.1|	1	27	2	2	55
gi|339247869|ref|XP_003375568.1|	1	19	2	2	60
gi|339247869|ref|XP_003375568.1|	1	62	9	29	525
gi|390349570|ref|XP_782388.3|	1	69	5	21	382
gi|393908216|gb|EFO23038.2|	1	6	2	4	58
gi|399932045|gb|AFP57558.1|	1	35	7	22	540
gi|405970776|gb|EKC35652.1|	1	12	2	2	39
gi|405975785|gb|EKC40331.1|	1	56	32	116	2523
gi|42559558|sp|O97192.1|TPM_HELAS	1	11	4	8	140
gi|42560362|gb|AAS20336.1|	1	54	3	95	2345
gi|524865813|ref|XP_005089733.1|	1	23	4	6	119
gi|524868888|ref|XP_005091242.1|	1	29	2	4	166
gi|524871402|ref|XP_005092474.1|	1	28	9	10	299
gi|524871513|ref|XP_005092529.1|	1	8	2	2	38
gi|524873824|ref|XP_005093654.1|	1	10	2	3	88
gi|524874374|ref|XP_005093923.1|	1	46	5	8	117
gi|524874376|ref|XP_005093924.1|	1	64	7	14	219
gi|524876525|ref|XP_005094978.1|	1	33	5	6	89
gi|524877853|ref|XP_005095623.1|	1	13	2	4	86
gi|524879771|ref|XP_005096557.1|	1	10	3	4	107
gi|524880608|ref|XP_005096966.1|	1	22	2	3	115
gi|524881407|ref|XP_005097355.1|	1	9	2	2	101
gi|524881473|ref|XP_005097388.1|	1	23	2	3	93
gi|524881817|ref|XP_005097558.1|	1	20	9	11	208
gi|524881819|ref|XP_005097559.1|	1	20	9	11	208
gi|524882112|ref|XP_005097701.1|	1	69	13	39	716
gi|524882279|ref|XP_005097784.1|	1	5	2	2	39
gi|524883774|ref|XP_005098513.1|	1	6	2	2	26
gi|524885219|ref|XP_005099219.1|	1	16	3	3	87
gi|524886261|ref|XP_005099727.1|	1	47	41	130	2703

Hit in the NCBI nr database	UP	PC (%)	NM	NS	PS
gi|524887580|ref|XP_005100374.1|	1	10	4	5	69
gi|524888358|ref|XP_005100750.1|	1	4	2	6	83
gi|524890724|ref|XP_005101905.1|	1	84	18	58	1117
gi|524890869|ref|XP_005101976.1|	1	6	2	2	54
gi|524890899|ref|XP_005101991.1|	1	16	2	2	31
gi|524891896|ref|XP_005102480.1|	1	45	6	10	164
gi|524894261|ref|XP_005103632.1|	1	3	3	6	41
gi|524894818|ref|XP_005103900.1|	1	22	2	8	71
gi|524894858|ref|XP_005103920.1|	1	7	3	3	61
gi|524896814|ref|XP_005104878.1|	1	40	78	88	1633
gi|524899004|ref|XP_005105943.1|	1	15	3	13	446
gi|524901033|ref|XP_005106931.1|	1	48	2	7	173
gi|524901281|ref|XP_005107051.1|	1	20	2	2	85
gi|524901464|ref|XP_005107139.1|	1	23	5	6	48
gi|524903383|ref|XP_005107892.1|	1	62	4	7	174
gi|524903383|ref|XP_005107892.1|	1	32	5	18	416
gi|524905758|ref|XP_005108242.1|	1	38	2	2	66
gi|524906745|ref|XP_005108595.1|	1	12	3	3	33
gi|524906745|ref|XP_005108595.1|	1	28	3	5	112
gi|524906862|ref|XP_005108596.1|	1	31	2	7	153
gi|524907032|ref|XP_005108645.1|	1	21	2	6	148
gi|524907032|ref|XP_005108645.1|	1	28	4	6	88
gi|524911570|ref|XP_005110646.1|	1	14	3	4	118
gi|524911945|ref|XP_005110830.1|	1	63	5	15	360
gi|524912496|ref|XP_005111095.1|	1	9	3	3	57
gi|524912496|ref|XP_005111095.1|	1	8	3	3	66
gi|524912621|ref|XP_005111155.1|	1	28	3	3	51
gi|524912715|ref|XP_005111198.1|	1	41	27	41	982
gi|524913881|ref|XP_005111763.1|	1	3	2	5	67
gi|537705579|dbj|BAN81895.1|	1	30	4	8	189
gi|556097681|gb|ESO86333.1|	1	3	2	3	52
gi|556102658|gb|ESO91310.1|	1	10	2	22	457
gi|556729406|ref|XP_005960416.1|	1	61	11	25	747
gi|56792845|gb|AAW30622.1|	1	42	4	5	137
gi|585719304|ref|XP_006826038.1|	1	39	64	121	2420
gi|597867153|gb|EYC16545.1|	1	17	2	5	79
gi|625230674|ref|XP_007606483.1|	1	56	4	15	248
gi|773571|emb|CAA60122.1|	1	7	2	2	109
gi|829207|emb|CAA39415.1|	1	76	39	53	737
gi|85753308|gb|AAG49533.3|	1	20	4	19	264

8.4. Table 3. Lectin-like proteins detected in transcriptomics level (status: I-incomplete, C-complete)

prot_hit	prot_hit_desc	status	evidence	protein
gi|524881819|ref|XP_005097559.1|	PREDICTED: calnexin-like isoform X2 [Aplysia californica]	I	T	calnexin
gi|524881819|ref|XP_005097559.1|	PREDICTED: calnexin-like isoform X2 [Aplysia californica]	I	T	calnexin
gi|2073142|dbj|BAA19861.1|	Incilarin A [Meghimatium fruhstorferi]	C	T	C-type lectin
gi|2073142|dbj|BAA19861.1|	Incilarin A [Meghimatium fruhstorferi]	C	T	C-type lectin
gi|2073146|dbj|BAA19863.1|	Incilarin C [Meghimatium fruhstorferi]	C	T	C-type lectin
gi|405959962|gb|EKC25931.1|	hypothetical protein CGI_10009560 [Crassostrea gigas]	C	T	C-type lectin
gi|524894115|ref|XP_005103561.1|	PREDICTED: uncharacterized protein LOC101845181 [Aplysia californica]	C	T	C-type lectin
gi|524906450|ref|XP_005108528.1|	PREDICTED: secretory phospholipase A2 receptor-like [Aplysia californica]	C	T	C-type lectin
gi|524907181|ref|XP_005108715.1|	PREDICTED: uncharacterized protein LOC101848467 [Aplysia californica]	C	T	C-type lectin
gi|556103033|gb|ESO91685.1|	hypothetical protein LOTGIDRAFT_163414 [Lottia gigantea]	C	T	C-type lectin
gi|556111065|gb|ESO99716.1|	hypothetical protein LOTGIDRAFT_238723 [Lottia gigantea]	C	T	C-type lectin
gi|126697272|gb|ABO26593.1|	putative perlucin 4 [Haliotis discus discus]	I	T	C-type lectin
gi|126697272|gb|ABO26593.1|	putative perlucin 4 [Haliotis discus discus]	I	T	C-type lectin
gi|157694790|gb|ABV66064.1|	type II antifreeze protein [Lates calcarifer]	I	T	C-type lectin
gi|2073142|dbj|BAA19861.1|	Incilarin A [Meghimatium fruhstorferi]	I	T	C-type lectin
gi|405959962|gb|EKC25931.1|	hypothetical protein CGI_10009560 [Crassostrea gigas]	I	T	C-type lectin
gi|405972881|gb|EKC37628.1|	Low affinity immunoglobulin epsilon Fc receptor [Crassostrea gigas]	I	T	C-type lectin

prot_hit	prot_hit_desc	status	evidence	protein
gi|524889950|ref|XP_005101525.1|	PREDICTED: FRAS1-related extracellular matrix protein 1-like [Aplysia californica]	I	T	C-type lectin
gi|524889950|ref|XP_005101525.1|	PREDICTED: FRAS1-related extracellular matrix protein 1-like [Aplysia californica]	I	T	C-type lectin
gi|524890590|ref|XP_005101839.1|	PREDICTED: cell wall protein DAN4-like [Aplysia californica]	I	T	C-type lectin
gi|524894115|ref|XP_005103561.1|	PREDICTED: uncharacterized protein LOC101845181 [Aplysia californica]	I	T	C-type lectin
gi|524895689|ref|XP_005104326.1|	PREDICTED: uncharacterized protein LOC101845349 [Aplysia californica]	I	T	C-type lectin
gi|524907181|ref|XP_005108715.1|	PREDICTED: uncharacterized protein LOC101848467 [Aplysia californica]	I	T	C-type lectin
gi|546241983|gb|AGW83938.1|	vitelline coat lysin M3, partial [Mytilus trossulus]	I	T	C-type lectin
		I	T	C-type lectin
gi|226731904|gb|ACO82046.1|	C-type lectin 13 [Perca flavescens]	I	T	C-type lectin
gi|617468332|ref|XP_007574500.1|	PREDICTED: galactose-specific lectin nattectin-like, partial [Poecilia formosa]	I	T	C-type lectin
gi|524864072|ref|XP_005088874.1|	PREDICTED: uncharacterized protein LOC101856539 [Aplysia californica]	I	T	C-type lectin
gi|524907181|ref|XP_005108715.1|	PREDICTED: uncharacterized protein LOC101848467 [Aplysia californica]	I	T	C-type lectin
gi|556100655|gb|ESO89307.1|	hypothetical protein LOTGIDRAFT_229175 [Lottia gigantea]	C	T	C-type lectin
gi|556097543|gb|ESO86195.1|	hypothetical protein LOTGIDRAFT_235549 [Lottia gigantea]	C	T	C-type lectin
gi|556097543|gb|ESO86195.1|	hypothetical protein LOTGIDRAFT_235549 [Lottia gigantea]	C	T	C-type lectin
gi|524870399|ref|XP_005091984.1|	PREDICTED: macrophage mannose receptor 1-like [Aplysia californica]	I	T	C-type lectin
gi|556103033|gb|ESO91685.1|	hypothetical protein LOTGIDRAFT_163414 [Lottia gigantea]	C	T	C-type lectin
gi|152032020|gb|ABS28869.1|	tandem repeat galectin [Biomphalaria glabrata]	C	T	galectin
gi|524891001|ref|XP_005102040.1|	PREDICTED: galectin-4-like isoform X3 [Aplysia californica]	C	T	galectin

prot_hit	prot_hit_desc	status	evidence	protein
gi|524891001|ref|XP_005102040.1|	PREDICTED: galectin-4-like isoform X3 [Aplysia californica]	C	T	galectin
gi|524891001|ref|XP_005102040.1|	PREDICTED: galectin-4-like isoform X3 [Aplysia californica]	C	T	galectin
gi|402580802|gb|EJW74751.1|	galectin, partial [Wuchereria bancrofti]	I	T	galectin
gi|560121553|emb|CDJ93815.1|	Galectin domain containing protein, partial [Haemonchus contortus]	I	T	galectin
gi|152032020|gb|ABS28869.1|	tandem repeat galectin [Biomphalaria glabrata]	I	T	galectin
gi|568300466|gb|ETN86155.1|	galactoside-binding lectin [Necator americanus]	I	T	galectin
gi|405976166|gb|EKC40683.1|	Cubilin [Crassostrea gigas]	I	T	galectin
gi|85362949|gb|ABC69854.1|	agglutinin [Helix pomatia]	C	T	H-type lectin
gi|109157294|pdb|2CE6|A	Chain A, Structure Of Helix Pomatia Agglutinin With No Ligands	I	T	H-type lectin
gi|85362949|gb|ABC69854.1|	agglutinin [Helix pomatia]	I	T	H-type lectin
gi|85362949|gb|ABC69854.1|	agglutinin [Helix pomatia]	I	T	H-type lectin
		I	T	H-type lectin
gi|405965843|gb|EKC31192.1|	Protein ERGIC-53 [Crassostrea gigas]	C	T	L-type lectin
gi|405965843|gb|EKC31192.1|	Protein ERGIC-53 [Crassostrea gigas]	C	T	L-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|524895404|ref|XP_005104185.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 5-like [Aplysia californica]	I	T	R-type lectin
gi|524896314|ref|XP_005104633.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 2-like [Aplysia californica]	C	T	R-type lectin
gi|556102066|gb|ESO90718.1|	hypothetical protein LOTGIDRAFT_123129, partial [Lottia gigantea]	I	T	R-type lectin
gi|524888837|ref|XP_005100983.1|	PREDICTED: N-acetylgalactosaminyltransferase 7-like isoform X1 [Aplysia californica]	I	T	R-type lectin
gi|524865257|ref|XP_005089459.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 11-like isoform X2 [Aplysia californica]	I	T	R-type lectin

prot_hit	prot_hit_desc	status	evidence	protein
gi|321455342|gb|EFX66478.1|	hypothetical protein DAPPUDRAFT_302681 [Daphnia pulex]	C	T	R-type lectin
gi|405966386|gb|EKC31679.1|	Polypeptide N-acetylgalactosaminyltransferase 5 [Crassostrea gigas]	I	T	R-type lectin
gi|524884259|ref|XP_005098751.1|	PREDICTED: putative polypeptide N-acetylgalactosaminyltransferase 10-like, partial [Aplysia californica]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	I	T	R-type lectin

prot_hit	prot_hit_desc	status	evidence	protein
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	C	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	C	T	R-type lectin
gi|556113252|gb|ESP01904.1|	hypothetical protein LOTGIDRAFT_138892, partial [Lottia gigantea]	C	T	R-type lectin
gi|556093434|gb|ESO82087.1|	hypothetical protein LOTGIDRAFT_198292 [Lottia gigantea]	I	T	R-type lectin
gi|556093434|gb|ESO82087.1|	hypothetical protein LOTGIDRAFT_198292 [Lottia gigantea]	I	T	R-type lectin
gi|524864445|ref|XP_005089059.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 1-like isoform X2 [Aplysia californica]	I	T	R-type lectin
gi|524864445|ref|XP_005089059.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 1-like isoform X2 [Aplysia californica]	I	T	R-type lectin
gi|524864445|ref|XP_005089059.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 1-like isoform X2 [Aplysia californica]	I	T	R-type lectin
gi|524864445|ref|XP_005089059.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 1-like isoform X2 [Aplysia californica]	I	T	R-type lectin
gi|556102066|gb|ESO90718.1|	hypothetical protein LOTGIDRAFT_123129, partial [Lottia gigantea]	I	T	R-type lectin
gi|524890796|ref|XP_005101940.1|	PREDICTED: polypeptide N-acetylgalactosaminyltransferase 5-like [Aplysia californica]	I	T	R-type lectin
gi|556112926|gb|ESP01578.1|	hypothetical protein LOTGIDRAFT_91617, partial [Lottia gigantea]	C	T	R-type lectin
gi|524884463|ref|XP_005098851.1|	PREDICTED: techylectin-5A-like [Aplysia californica]	C	T	techylectin
gi|524914011|ref|XP_005111826.1|	PREDICTED: techylectin-5B-like [Aplysia californica]	I	T	techylectin
gi|524914011|ref|XP_005111826.1|	PREDICTED: techylectin-5B-like [Aplysia californica]	I	T	techylectin
gi|3347852|gb|AAC27743.1|	sialic acid-binding lectin 2 [Limax flavus]	I	T	lectin
gi|3347852|gb|AAC27743.1|	sialic acid-binding lectin 2 [Limax flavus]	I	T	lectin

prot_hit	prot_hit_desc	status	evidence	protein
gi|524900094|ref|XP_005106471.1|	PREDICTED: endoplasmic reticulum lectin 1-like [Aplysia californica]	I	T	lectin
gi|524875712|ref|XP_005094580.1|	PREDICTED: EMILIN-2-like [Aplysia californica]	C	T	C1qDC
gi|583980443|ref|XP_006785537.1|	PREDICTED: complement C1q-like protein 2-like [Neolamprologus brichardi]	C	T	C1qDC
		C	T	C1qDC
gi|325504319|emb|CBX41659.1|	putative C1q domain containing protein MgC1q10 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504319|emb|CBX41659.1|	putative C1q domain containing protein MgC1q10 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504319|emb|CBX41659.1|	putative C1q domain containing protein MgC1q10 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504327|emb|CBX41663.1|	putative C1q domain containing protein MgC1q14 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504381|emb|CBX41690.1|	putative C1q domain containing protein MgC1q41 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504381|emb|CBX41690.1|	putative C1q domain containing protein MgC1q41 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504405|emb|CBX41702.1|	putative C1q domain containing protein MgC1q53 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504405|emb|CBX41702.1|	putative C1q domain containing protein MgC1q53 [Mytilus galloprovincialis]	I	T	C1qDC
gi|325504493|emb|CBX41746.1|	putative C1q domain containing protein MgC1q97 [Mytilus galloprovincialis]	I	T	C1qDC
gi|405955760|gb|EKC22740.1|	Complement C1q subcomponent subunit B [Crassostrea gigas]	I	T	C1qDC
gi|405963338|gb|EKC28921.1|	Collagen alpha-2(VIII) chain [Crassostrea gigas]	I	T	C1qDC
gi|405972817|gb|EKC37565.1|	Caprin-2 [Crassostrea gigas]	I	T	C1qDC
gi|472825914|gb|AGI44588.1|	C1q domain containing protein [Azumapecten farreri]	I	T	C1qDC

prot_hit	prot_hit_desc	status	evidence	protein
gi|472825914|gb|AGI44588.1|	C1q domain containing protein [Azumapecten farreri]	I	T	C1qDC
gi|524875712|ref|XP_005094580.1|	PREDICTED: EMILIN-2-like [Aplysia californica]	I	T	C1qDC
		I	T	C1qDC
gi|126697410|gb|ABO26662.1|	sialic acid binding lectin [Haliotis discus discus]	I	T	C1qDC
gi|38043955|emb|CAD83837.1|	sialic acid binding lectin [Cepaea hortensis]	I	T	C1qDC
gi|38043955|emb|CAD83837.1|	sialic acid binding lectin [Cepaea hortensis]	I	T	C1qDC
gi|93209532|gb|ABF00124.1|	sialic acid binding lectin [Helix pomatia]	I	T	C1qDC
gi|93209532|gb|ABF00124.1|	sialic acid binding lectin [Helix pomatia]	I	T	C1qDC
gi|40218026|gb|AAR82936.1|	macrophage expressed protein [Haliotis rufescens]	C	T	pore-forming membrane attack
gi|524916402|ref|XP_005112978.1|	PREDICTED: uncharacterized protein LOC101845147 [Aplysia californica]	C	T	pore-forming membrane attack
gi|119112553|ref|XP_001237961.1|	AGAP007821-PA [Anopheles gambiae str. PEST]	C	T	toxin-like
gi|193685923|ref|XP_001950560.1|	PREDICTED: uncharacterized protein LOC100163563 isoform X1 [Acyrthosiphon pisum]	C	T	toxin-like

8.5. Figure 1. Alignment of amino acid sequences of H-type lectins from *A. vulgaris* with *Helix pomatia* and *Cepaea hortensis*