Elsevier Editorial System(tm) for Journal of Proteomics

Manuscript Draft

Manuscript Number: JPROT-D-15-00495R1

Title: The secrets of Oriental panacea: Panax ginseng

Article Type: Full Length Article

Keywords: panax ginseng proteome; combinatorial peptide ligand libraries; protein-protein interaction network; panax ginseng proteolytic peptides; orbitrap mass spectrometer

Corresponding Author: Dr. Elisa Fasoli, Ph.D

Corresponding Author's Institution: Politecnico di Milano

First Author: Mara Colzani

Order of Authors: Mara Colzani; Alessandra Altomare; Matteo Caliendo; Giancarlo Aldini; Pier Giorgio Righetti; Elisa Fasoli, Ph.D

Abstract: The Panax ginseng root proteome has been investigated via capture with combinatorial peptide ligand libraries (CPLL) at three different pH values. Proteomic characterization by SDS-PAGE and nLC-MS/MS analysis, via LTQ-Orbitrap XL, led to the identification of a total of 207 expressed proteins. This quite large number of identifications was achieved by consulting two different plant databases: Panax ginseng and Arabidopsis thaliana. The major groups of identified proteins were associated to structural species (19.2%), oxidoreductase (19,5%), dehydrogenases (7.6%) and synthases (9.0%). For the first time, an exploration of protein-protein interactions was performed by merging all recognized proteins and building an interactomic map, characterized by 196 nodes and 1554 interactions. Finally a peptidomic analysis was developed combining different in-silico enzymatic digestions to simulate the human gastrointestinal process: from 661 generated peptides, 95 were identified as possible bioactives and in particular 6 of them were characterized by antimicrobial activity. The present report offers new insight for future invesigations focused on elucidation of biological properties of Panax ginseng proteome and peptidome.

Cover Letter

Milano, July 31-2015

To Prof. Jesùs Jorrìn-Novo,

Executive editor, Journal of Proteomics

Dear Editor,

I am pleased to submit to you the following manuscript:

“The secrets of Oriental panacea: Panax ginseng”

by M. Colzani, A. Altomare, M. Caliendo, G. Aldini, P. G. Righetti and myself for consideration for publication in the *Journal of Proteomics*. The manuscript describes the identification of Panax ginseng proteome captured by combinatorial peptide ligand libraries from dried root extracts, focusing attention on biological functions of proteins. Additionally an interactomic map was constructed by exploiting the protein-protein interactions and finally a peptidomic analysis was performed by *in-silico* human gastrointestinal digestion in order to evaluate the peptides bioactivity. In particular, from a list of 95 bioactive peptides, we have focused our attention on 6 peptides characterized by antimicrobial activity. Our study represents the first extensive proteomic and peptidomic investigation of *Panax ginseng* root and we hope that it might be of interest to the readers of Journal of Proteomics.

Since *Journal of Proteomics* focuses on proteomics and peptidomics, we have considered this journal as the most appropriate to submit our manuscript. We will be happy to answer any questions the referees might have concerning this manuscript.

Looking forward to hearing from you soon,

Sincerely yours,

Elisa Fasoli

Ph.D
Politecnico di Milano, Dep. of Chemistry, Materials and Chemical Engineering
Via Mancinelli, 7
20133 Milan, Italy
e-mail: elisa.fasoli@polimi.it
Office Phone: +39-02-23994749

***Detailed Response to Reviewers**

Dear Editor and Reviewer,

Thank you for your comments concerning our manuscript. They are helpful for revising and improving our paper. We have made correction which we hope to meet with the approval: revised portion are marked in red in paper and the answers to reviewer's comments are below.

Reviewer #1: Basically this paper describes the proteome analysis of Panax ginseng roots after an accurate protein extraction and treatment in order to evidence low concentration proteins with a technology that is well known by the authors. Moreover a detailed analysis is given of protein functional interactions as well as the examination (number, structure and function) of peptides after in silico protein breakdown.
An important experimental work is made around protein extraction treatment (three different pHs) and identification, while the rest of the reported data (protein functional analysis, protein interaction and peptide investigation) is an informatics-based treatment of the discovered proteins.
This paper brings an additional important piece of information to the knowledge of Panax ginseng that is considered in certain countries as a promoter of health and wellness. As such it deserves to be considered for publication. This decision is nevertheless dependent on the perfect clarification of few points that are highlighted below.

Page 9: Where the Ginseng powder comes from?

We have re-written this part according to the Reviewer's suggestion, maked in red in the paper. We have received new information about Ginseng powder from laboratory of Aboca s.p.a.

Page 13, section 2.7: It is unclear if this study was performed with the initial protein extract (Raw) or after CPLL treatment. Moreover it is not indicated if the "simulated gastrointestinal digestion" was performed separately for each CPLL eluate protein list. In the absence of this information it is believed the analysis was made globally even if it would have been logical to have that done on each CPLL-desorbed fraction to get a more compelling discussion. Here it is up to the author to either perform complementary investigations or to explain why a global peptidome analysis makes more sense. It may be just a simple question of data presentation; in this respect Table 2 should indicate where each listed protein (column 7 from the left) comes from (eluate pH 2.2, 7.2 or 9.0).

Sorry, this part of the manuscript may be written ambigously.The "simulated gastrointestinal digestion" was performed by using the global protein list. In more details, we created a global protein database by *merging* all the MS tabular report obtained by searching each MS spectra (control and CPLL analysis) against the *Arabidopsis Thaliana* database. We didn't separate data from each CPLL treatment because "unhelpful" for our main purpose: comprehensive bioactive peptides prediction. We were not interested on a detailed bioactivity prediction referred to each pH condition of CPLL extraction. The *functional study* is a different part of the work: it's interesting here to distinguish the main analytical part of the work, aimed at the investigation of the global Panax G. proteome, from the functional analysis aimed at the exploration of bioactivity of global peptidome. For these reasons the final table (Tab. 2) doesn't indicate the sample able to generate the peptide.

Page 14: It is recommended to start the "Result" section by reporting data on protein fractionation. For instance indicate the amount of protein extracted from 3 gr of dry powder, the amount of protein contacted with CPLL and the amount of proteins collected from each pH. As a consequence the authors could remove the sentence lines 271-274 of the "Discussion" section.

Considering Reviewer's suggestion, we haveindicated the amount of protein contacted with CPLLs in the "Result" session. Unfortunately, we can not measure the protein content collected after CPLLs treatment because the presence of SDS, required for an efficient elution, is not compatible with used protein assay. 300mg are the tital amount contacted with CPLLs: considering 3 different pH values of incubation, 100mg of proteins have interacted with 100µl of beads (the minimum protein content, recommended by company for a correct equalisation, is equal or major to 50mg for 100µl of beads)

Page 14, line 230: Comments on the differences/similarities are requested about SDS-PAGE results to assess the interest to use CPLL. In this regard Figure 2 should be reconfigured: the two lanes on the right are "compressed" compared to the panel on the left. Please compress the left panel or stretch the right panel to have them comparable.

We are sorry for our negligence, now detailed comments are added and the Figure 2 is reconfigured.

Page 14, line 249: To assess the interest of the protein treatment at three different pHs it is asked to give a contribution of each of them in terms of protein number (diagram insert within Figure 2?) and/or diversity.

Thank you for the suggestion. We have now inserted the suggested diagramswithin Figure 3 (since it already contained ID numbers) to show protein number and common/different identifications obtained at pH 2.2, 7.2 ad 9.0, updating and completing text in all session (result, discussion and legend).

Discussion: It is recommended to subdivide this part into three sections instead of two: "Protein Identification", "Protein Functional Data" (this should include "Gene Ontology" and "Protein Interaction") and "Analysis of Proteolytic Peptides".

Thank you for the good suggestion. We have subdivided the “Discussion” part into three sections.

Second part of line 368: please correct.

Sorry, this part of manuscript my be written ambigously: we have re-written in a correct way.

Line 416: Use Italics for Panax ginseng.

We are sorry for our negligence, now we have corrected.

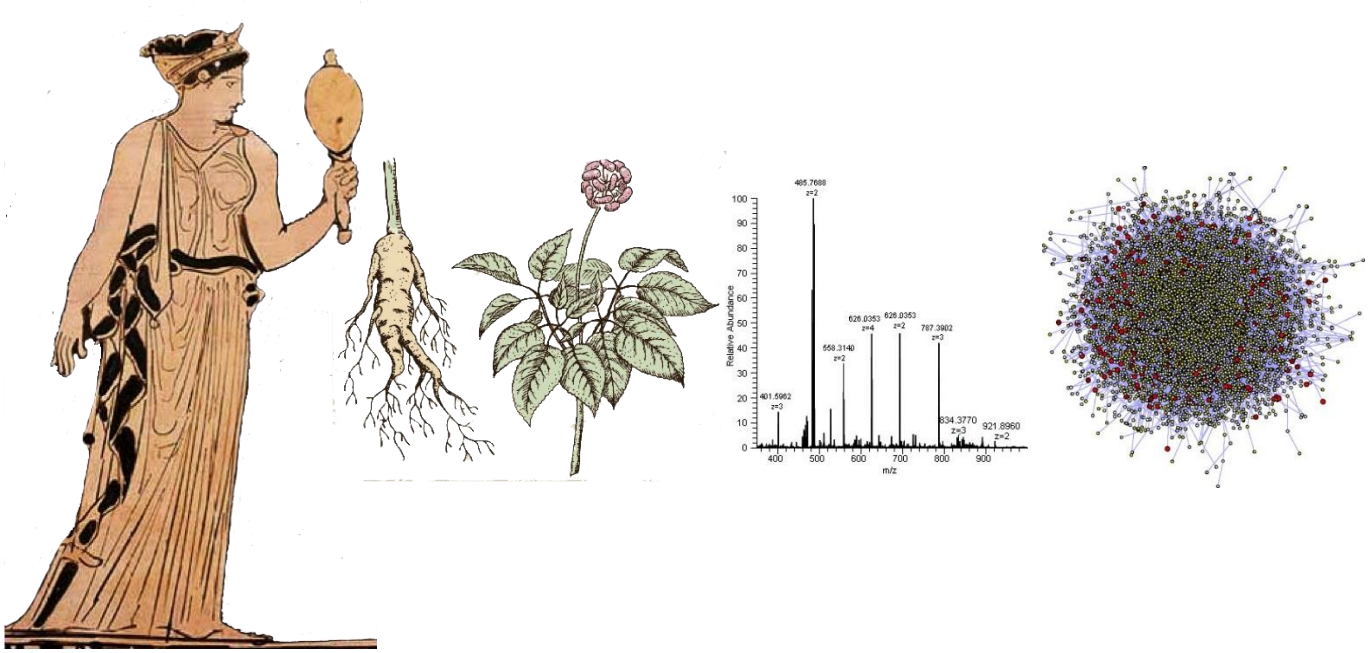
Once again, thank you very much for your comments and suggestions.
Best regards,
Yours sincerely

Elisa Fasoli

***Significance**

Ginseng is a traditional oriental herbal remedy whose use is very diffused in all the world for its numerous pharmacological effects. However, the exact mechanism of action of ginseng components, both ginsenosides and proteins, is still unidentified. So the common use of ginseng requires strict investigations to assess both its efficiency and its safety. Although many reports have been published regarding the pharmacological effects of ginseng, little is known about the biochemical pathways of root. Proteomics analysis could be useful to elucidate the physiological pathways. In this manuscript, an integrated approach to proteomics and peptidomics will usher in exploration of Panax ginseng proteins and proteolytic peptides, obtained by *in-silico* gastrointestinal digestion, characterized by antimicrobial action. The present research would pave the way for better knowledge of metabolic functions connected with ginseng proteome and provide with new information necessary to understand better antimicrobial activity of *Panax* ginseng.

***Graphical Abstract**



The root of well-being, as blessed by Panacea

Highlights

- Investigation of *Panax ginseng* proteome via combinatorial peptide ligand libraries
- Exploration of biological functions of 206 proteins identified via MS analysis
- Building of interactomic map formed by 196 nodes and 1554 interactions
- *In-silico* gastrointestinal digestion has produced 95 possible bioactive peptides
- 6 proteolytic peptides were characterized by antimicrobial activity

The secrets of Oriental panacea: Panax ginseng

Mara Colzani[a], Alessandra Altomare[a], Matteo Caliendo[a], Giancarlo Aldini[a], Pier Giorgio Righetti[b], Elisa Fasoli[*,b]

[a]Department of PharmaceuticalSciences, Università degli Studi di Milano, via Mangiagalli 25, 20133 Milano, Italy

[b] Department of Chemistry, Materials and Chemical Engineering “Giulio Natta”, Polytechnic of Milan, Via Mancinelli 7, 20131 Milan, Italy

Keywords: panax ginseng proteome; combinatorial peptide ligand libraries; protein-protein interaction network; panax ginseng proteolytic peptides;orbitrapmass spectrometer

Corresponding Author:
Elisa Fasoli, Department of Chemistry, Materials and Chemical Engineering “Giulio Natta”, Polytechnic of Milan, Via Mancinelli 7, 20131 Milan, Italy.
Tel. 0039 02 23994749
Fax 0039 02 23993018
E-mail: elisa.fasoli@polimi.it

Abstract

The *Panax ginseng* root proteome has been investigated via capture with combinatorial peptide ligand libraries (CPLL) at three different pH values. Proteomic characterization by SDS-PAGE and nLC-MS/MS analysis, via LTQ-Orbitrap XL,led to the identification of a total of 207 expressed proteins. This quite large number of identifications was achieved by consulting two different plant databases: *Panax ginseng* and *Arabidopsis thaliana*. The major groups of identified proteins were associated to structural species (19.2%), oxidoreductase (19,5%), dehydrogenases (7.6%) and synthases (9.0%). For the first time, an exploration of protein-protein interactions was performed by merging all recognized proteins and building an interactomic map, characterized by 196 nodes and 1554 interactions. Finally a peptidomic analysis was developed combining different *in-silico* enzymatic digestions to simulate the human gastrointestinal process: from 661 generated peptides, 95 were identifiedas possible bioactives and in particular 6 of them were characterized by antimicrobial activity. The presentreport offers new insight for future invesigations focused on elucidation of biological properties of *Panax ginseng*proteome and peptidome.

1. Introduction

Asian ginseng (*Panax ginseng C. A. Meyer)* has a history of herbal use going back over 5,000 years as described in Chinese traditional medicine textbooks. It is one of the most highly regarded of herbal medicines in the Orient, where it has gained a reputation for being able to promote health, general body vigour and also to prolong life [1-3]. The genus name *Panax* is derived from the Greek word meaning "panacea" or "all-healing": the species ginseng is said to mean "wonder of the world". Both terms refer to the medicinal virtues of the plant belonging to *Araliaceae* family. Ginseng is traditionally used as an aid during convalescence and as a prophylatictic to build resistance, to reduce susceptibility to illness and to promote health and longevity [4]. Panax ginseng is a slow-maturing perennial herb native to the mountain forests of northeastern China, Korea and Russia. Seven major species of ginseng are cultivated extensively in China, Japan, Korea, Russia, Canada and Wisconsis in the US. The three mostly studied species are: *Panax ginseng* (Asian ginseng), *Panaxquinquefolius*(American ginseng) and *Panax japonicas* (Japanese ginseng). Ginseng usually starts flowering at its fourth year and the roots take four to six years to reach maturity. Therefore, productivity of *Panax ginseng* can be significantly affected by various environmental factors such as temperature, condition of soil, light intensity, content of water and diseases [5-10]. Ginseng root consists of dried main root, lateral roots and root hairs or “tailings” of *Panax ginseng* C. A. Meyer,used as a tonic to revitalize and replenish vital energy. The bioactivities of ginseng roots, which include hypotensive [11], anticancer [12-14], antioxidant [15], anti-inflammatory activities as well as improving impaired memory [16-18] and anti-cardiovascular diseases [19, 20], are due to the presence of main pharmacologically active components like ginsenosides.To understand these effects, the components of ginseng have been subjected to extensive analysis. Its saponins, known as ginsenosides, are considered to be the main active pharmacological

compunds in P. ginseng. The distribution of ginsenosides varies from species to species [2] and each ginsenoside has different pharmacological effects even if it could produce multiple effects in the same tissue [3].

Although many reports have been published regarding the pharmacological effects of ginsenosides, little is known about biochemical pathways involving different proteins. High-throughput and high-sensitive proteomic techniques, such as gel electrophoresis and mass spectrometry, allow the separation of complex protein mixtures, which makes possible to charcterizeprotein profiles in plant roots. Recently, proteomic studies have been performed on ginseng roots: the first proteomic profiles were provided by Lum et al. [21], subsequently improved by Kim et al. by identifying several high-abundance proteins [22]. Most of highly abundant proteins in ginseng root are root-specific RNase-like proteins that function as vegetative storage proteins for survival in the natural environment [23, 24]. Subsequently, proteomic analysis was used on different varieties of ginseng, such as *Panax ginseng*and*Panaxquinquefolium*[25-27]. However, to date, no systemic research has been reported for a deep investigation focused to identify the entire ginseng root proteome.

Therefore, in this study, a proteomic approach was applied to identify proteinsof*Panax ginseng*C. A.Meyer root, contributing to understand their role in metabolic processes. Previously, a method for sample preparation of dried roots was established [28] and a suitable technology, called combinatorial peptide ligand libraries (CPLLs) was used to capture the entire proteomeby enriching trace proteins and concomitantly reducing the concentration of abundant species [29].Ginseng proteins were identified by nLC-MS/MS, using an Orbitrap mass spectrometer, in order to extensively map the proteome for a consequent exploration of protein functions via Gene Ontology analysis. Moreover, also an interactomic investigation was performed merging all the proteins, by exploiting

the STRING v.9.1 software, which was integrated with the analysis of peptides, obtained after an in-silico human gastrointestinal digestion.

These results provide insight into the proteome and peptidomeof *Panax ginseng* roots and contribute to a more comprehensive understanding of biological functions.

2. Materials and methods

2.1. Chemicals

ProteoMiner™ (combinatorial hexapeptide ligand library beads, CPLL), Laemmli buffer, 40% acrylamide/Bis solution, N,N,N',N'-tetramethylethylenediamine (TEMED), molecular mass standards and electrophoresis apparatus for one-dimensional electrophoresis were from Bio-Rad Laboratories, Inc., Hercules CA. β-mercaptoethanol, dithiothreitol (DTT), ammonium persulfate, 3-[3-cholamidopropyl dimethylammonio]-1-propanosulfonate (CHAPS), acetonitrile (ACN), trifluoroacetic acid (TFA), sodium dodecyl sulphate (SDS), iodoacetamide (IAA), formic acid (FA) and all other chemicals used all along the experimental work were current pure analytical grade products and purchased from Sigma-Aldrich S.r.l, Italy. Water with 0.1% formic acid and acetonitrile with 0.1% formic acid LC/MS grade were also purchased from Sigma Aldrich S.r.l, Italy Complete protease inhibitor cocktail tablets and sequencing grade trypsin were from Roche Diagnostics (Basel, CH).

2.2.Panax ginseng root treatment

Dried powdered roots of *Panax ginseng* C. A.Meyer were used in this experiments. The dry radix was from Martin Bauer group (Vestenbergsgreuth, Germany). The drug was

identified and groundby mill knives by the Phytochemistry Research Laboratory of Abocas.p.a. (Sansepolcro,Arezzo, Italy).

Samples were stored at room temperature until required and they were prepared as described by Wang et al. [28]. Briefly *Panax ginseng* powder (3 g) was washed firstly with 20mL of 10% TCA/ice-cold acetone, secondly with 20 mL of 80% MeOH/H_2O and finally with with 20 mL of 80% ice-cold acetone/H_2O. After drying at 25°C overnight, powder was incubated with 30 mL of extraction buffer (50mMTris-HCl pH=8.0, 200mM DTT, 0.3% SDS, 1 protease inhibitor tablet for 50 mL buffer)overnight at room temperature. After centrifugation at 14000rpm for 30 min, the supernatant was precipitated with 10% TCA solution for 1 h at -20°C. The protein pellet thus obtained was washed with ice-cold acetone to remove contaminants and resuspended into 60 mL buffer for CPLLs incubation (50mMTris-HCl pH=7.2, 50mMNaCl), added with one tablet of proteases inhibitors cocktail.This solution was divided into three, twenty mL fractions: one of them was equilibrated into the same pH 7.2 buffer, the other two fractions were titrated, respectively, to pH 9.0 by addition of NaOH solution and to pH 2.2 by addition of 0.1% TFA and formic acid to mimic reverse phase conditions for the capture of hydrophobic proteins. All fractions were loaded onto 100 µL of CPLL beads at the three different pH values [30] and the capture was performed batch-wise, gently rocking on a rotating platform for 2 hours. After that, the beads were rinsed twice with the incubation buffers, so as to remove any excess of non-adsorbed proteins (see the scheme of Fig. 1). Desorption was implemented by washing the beads twice (each time with 80 µL) with a boiling 4% SDS solution containing 20 mM DTT, 12.5 % (v/v) glycerol, 0.005% (m/v) bromophenol blue, and 62.5 mMTris-HCl (pH 6.8) [31].

2.3. SDS-PAGE analysis and trypsin digestion

TenµLof the above eluates (labelled as 2.2, 7.2 and 9.0) and ten µL of the non-treated sample (CTRL)were loaded onto an SDS-PAGE gel, composed by a 4% polyacrylamide stacking gel (125 mMTris-HCl, pH 6.8, 0.1%, m/v, SDS) over a 12% resolving polyacrylamide gel (in 375 mMTris-HCl, pH 8.8, 0.1%, m/v, SDS buffer). The cathodic compartment was filled with Tris-glycine buffer, pH 8.3, containing 0.1%, m/v, SDS whereasTris buffer, at pH 8.8, was present at the anode. Electrophoresis was performed at 100 V until the dye front reached the bottom of the gel and at 150 V until the end of the separation. Staining and distaining were performed with Colloidal Coomassie Blue and 7% (v/v) acetic acid in water, respectively.For each lane of the gel, subsequent gel pieces in the range of 200-15 kDa were excised to perform in-gel digestion, accordingly to a standardized protocol [32]. Thirteen gel pieces were cut from each sample, rinsed with pure water, destainedfor 10 min with 50% acetonitrile/25 mM ammonium bicarbonate and incubated at 56 °C with 10 mMdithiothreitol in 50 mM ammonium bicarbonate for cysteine reduction. After 60 min incubation, the solution was discarded and the gel pieces were incubated with 55mMiodoacetamide in 50 mM ammonium bicarbonate for 45 min, in the dark, for cysteine alkylation. After discarding the iodoacetamide solution, gel pieces were rinsed with 50 mM ammonium bicarbonate and incubated in 100% acetonitrile for dehydration. Each gel slice was incubated with 1 µg of sequencing grade trypsin (Roche) dissolved in 50 mM ammonium bicarbonate. After overnight digestion at 37°C, the solutions containing the digested peptides were collected. Gel pieces were incubated for 10 min with 30% acetonitrile, 3% trifluoroacetic acid, and the solution was collected and pooled with the initial peptide mixture. Gel pieces were incubated for an additional 10 min with 100% acetonitrile, and the solution was pooled with the previous one. Eluted peptides were evaporated on a vacuum concentrator (Christ) and stored at -20°C.

2.4. Mass spectrometry analysis

The dried peptides were solubilized in 15 µLof 0.1% formic acid. Five µLof sample were injected on a C_{18} column (picoFrit column, C18 HALO, 90Å, 75 µm ID, 2.7 µm, 10.5 cm length, New Objective) by a constant flow rate of 0.4 µL/min delivered by a nano-chromatographic system (UltiMate 3000 RSLCnano System, Thermo Scientific). The separating gradient ramped linearly from 1% acetonitrile to 40% acetonitrile in 30 minutes; the eluting peptides were on-line sprayed in a LTQ-Orbitrap XL mass spectrometer by a nano-ESI source (all Thermo Scientific). Full scan mass spectra were acquired in the Orbitrap cell in the mass range 300 to 1500 *m/z* (positive polarity, profile mode, AGC = 5×10^5). The nine most intense ions (minimum charge state 2+, minimum intensity 10000 cps) were automatically selected and fragmented in the ion trap by collision-induced dissociation (CID). After two subsequent occurrences in less than 30 s, target ions already selected for fragmentation were dynamically excluded for 45 s.

2.5.Protein identification

The MS data were analyzed by the Proteome Discoverer software (v. 1.3.0.339 Thermo), using the Sequest algorithm. The database of proteins belonging to different types of Panax (640 entries, downloaded on 02 July 2014 from Uniprot) and the database of *Arabidopsis thaliana* proteins (31760 entries, downloaded on 18 February 2014 from UniProt) were used for spectra matching. Cysteine carbamidomethylation and methionine oxidation were set as variable modifications. Peptide mass tolerance was set to 10 ppm, fragment mass tolerance to 0.5 Da and maximum number of missed cleavages = 2. The false discovery rate (FDR) for peptide identification was set at 0.05; only proteins identified by at least two peptides were considered as genuine identifications and further analyzed.

2.6. Analysis of identified proteins

In order to describe the functional classes of identified proteins, a Gene Ontology (GO) analysis was performed by using the web available software QuickGo (www.ebi.ac.uk/QuickGo).

An *interactomics* map was built up by means of STRING(Search Tool for the Retrieval of Interacting Genes) v.9.1 software(http://stringdb.org/), set on *Arabidopsis thaliana* as organism database. This is a large database of known and predicted protein-protein interactions. Proteins were represented with nodes and the interactions with continuous lines to represent direct interactions (physical), while indirect ones (functional) were presented by brokenlines. All the edges were supported by at least a reference from the literature or from canonical information stored in the STRING dataset. Cluster networks were created by using the K-means algorithm which is included in the STRING website and a value of 7 was selected for all the analyses. The pathways classification was done after the automaticenrichment in STRING, based on the information provided by the KEGG-Pathway Database.

2.7. Analysis of proteolytic peptides obtained by simulated gastrointestinal digestion

Bioactive peptides encrypted in the Panax proteome were predicted by combining different in-silico enzymatic digestions in order to simulate the human gastrointestinal process: pepsin (stomach) and intestinal enzymes (trypsin, chymotrypsin, elastase, carboxypeptidase A and B and aminopeptidases).In order to perform the simulated gastrointestinal digestion, a global protein database was created by merging all MS tabular report, obtained by searching MS spectra of initial protein extract and all CPLLs eluates, against an*Arabidopsis Thaliana* database.All the digestions were performed *in-silico* using the MS-Digest software which is included in the ProteinProspector v 5.10.1

website (http://prospector.ucsf.edu/prospector/mshome.htm). To evaluate the results, all the potential peptides were ranked by using the PeptideRanker software (http://bioware.ucd.ie/~testing/biowareweb/), using the N-to-1 neural network probability to predict which peptides could be more bioactive. In addition, all the potential bioactive peptides were compared with the CAMP database, which includes known antimicrobial bioactive peptides (http://www.bicnirrh.res.in/antimicrobial/).

3. Results

Figure 1 shows a complete scheme of protein extraction, CPLLs enrichment, proteins separation by SDS-PAGE electrophoresis and their identification by MS analysis. In order to increase the knowledge on the *Panax ginseng* proteome, we have performed the most efficient protocol able to extract proteins from a dried root powder. After testing various extraction procedures reported in literature [21], we have applied the initial steps of the protocol described by Wang et al. [28], which requiredpreliminary washes of dried powder in order to eliminate interferences and contaminants until the plant powder became colorless. The protein extraction was performed by using a sample buffer previously described [22]. The final combination of TCA and acetone is commonly used to precipitate proteins and to remove any remained contaminants. By using a combination of different methods, we were able to extract about 300 mg of soluble proteins from 3 g of sample, quantified by the BioRad protein assay with BSA as standard [33]. The extracted protein quantity was appropriate for incubation with bead-based libraries of combinatorial peptide ligands (CPLLs), a protein enrichment technology which simultaneously dilutes high-abundance proteins and concentrates low-abundance ones [34]. The CPLLs technology was originally optimized for use with human biological samples [35], but it was recently adoptedwith other sample types like

plants and foodstuffs within the recommend total protein range (>50mg of proteins for 100 μL beads volume) [36-38]. So in order to aim to a very large protein discovery, a sample overloading, as well as a capture at three pH values, are recommended.

The CPLLs treatment has contributed to increase protein capture, as seen by the SDS-PAGE profiling in figure 2 where the untreated sample (Raw) is characterized by three major protein bands, while CPLLs beads eluates from the three pH values (lanes E 2.2, E 7.2 and E 9.0) exhibit additional bands, particularly increased in the regions of the gel corresponding to 75-30 KDa and 20-15 KDa. In fact, while the electrophoretic profile of the raw extract has displayed only 3 or 4 evident protein bands, probably corresponding to high-abundance proteins, the CPLLs eluates profiles have revealed more intense and much more resolved protein bands, referred both to high-abundance proteins and to low-abundance ones. The efficiency and the potentiality of CPLLs treatment has been demonstrated by mass spectrometry analysis.

Figure 3 shows the number of proteins identified by matching the experimental spectra to two different databases: *Arabidopsis thaliana* and *Panax ginseng*. The use of two different databases was prompted by the necessity to increase the number of identified proteins. In fact while the *Panax ginseng* database is most specific, it's characterized by a large number of unreviewed entries. Moreover, the lack of complete DNA sequencing cause the incompleteness of the *Panax ginseng*proteome. The choice of *Arabidopsis thaliana* database is due to many reasons: first of all its database is complete and characterized by a higher number of entries than the *Panax ginseng* database (31706 vs. 640); in fact, the full proteome of *Arabidopsis thaliana* is available at the UniProt depository as reference proteome. Secondly, *Arabidopsis thaliana* is commonly used as model organism for studying plant sciences and biology, after its complete genetic mapping and sequencing in 2000. The table displays the IDs obtained in the untreated sample (raw extract) and in all CPLLs eluates: many species are recognized in all

samples so, in order to know the correct number of unique gene products, the last column reports the total distinct IDs without redundancies. The Venn diagrams, in the lower panel of figure 3, compare the identified proteins in CPLLs eluates vs. those recognized in raw extracts considering both types of databases. 13 and 43 more proteins, belonging tothe*Panax ginseng* and the *Arabidopsis thaliana* database respectively, could be identified only after CPLLs treatment. Finally the different contributions of each CPLLs incubation were considered for a deep investigation of Panax proteome: for both databases, CPLLs treatments at different pH values were successful for capture of specific proteins, undetectable applying only the standard protocol. Concerning the *Panax ginseng* database, the most performant incubation condition was at physiological pH, able to capture 8 unique ginseng's proteins. Considering the *Arabidopsos thaliana* database, the beads incubation at pH 9,0 has identified the higher number of specific proteins (27) than other two pH conditions (15 at pH 7,2 and 9 at pH 2,2).

Figure 4 depicts a pie chart of Gene Ontology (GO) analysis applied to all 207 identified species in order to understand /categorize the belonging protein classes.

Figure 5 displays the general workflow for the analysis of the peptides obtained after an *in-silico* human gastrointestinal digestion: the software simulates digestion of pepsin, trypsin, chymotrypsin, elastase, carboxipeptidases (A and B) and aminopeptidases, using the MS-Digest program. All generated peptides (661) were ranked on the basis of their predicted bioactivity by softwarethe PeptideRanker, which was able to select 95 bioactive peptides focusing our attention on their potential antimicrobial action as reported in Figure 6.

4. Discussion

4.1. Protein identification

The deep proteome identification wasperformed by using two different databases, *Arabidopsis thaliana* and *Panax ginseng*, in order to increase the number of panaxspecies discovered. In fact a total number of 206 proteins were identified by merging all unique IDs obtained considering both databases. Our resultshave substantially improved a previous study of Kim *et al.* [22], who recognized only 17 proteins using MALDI-TOF MS. The low rate of identified species was probably due to lack of ginseng genome DNA sequence database. Also for this reason we have decided to perform a search by considering the databank of plant species mostly used as model organism. In order to increase protein identification, Kim *et al.* have performed a BLAST search of amino acids sequences using the ginseng EST databases, finally recognizing 87 unique gene products. Ourquite largediscovery has been made possible thanks to very sensitive high resolution and high mass accuracy of the Orbitrapmass spectrometer,as well as to the useof CPLLs, whichhave enhanced by more tha 20% the final discovery. Indeed, as reported in the lower Venn diagram of Figure 3, the adopted strategy has beenindispensable to capture 13 more proteins, via identification by the *Panax*db, and 43 more proteins, via the*Arabidospis Thaliana* db. Even if these numbers are not too high, we believe that the CPLLs technology has contributed to proteome knowledge of this perennial plant whose genome is not fully sequenced.Moreover the strategy to use different pHs incubations has contributed to capture specific unique gene products, setting up the most performant conditions for a more efficient interaction between ginseng's proteins and hexapeptides chains.

4.2. Protein functional data

To gain further knowledge of the*Panax ginseng* proteome functionality, the 206 identified proteins were analysed by Gene Ontology, ascertaining their molecular function, and all GO categories were described in a pie chart of Figure 4. The most enriched protein classes were: structural (19.2%) and oxidoreductase (19,5%). The first category was probably a consequence of tissue type considered in this study. In fact roots are a very simple plant tissue, consisting of a central vascular system surrounded by a large storage parenchyma cells on the outside protective layer. So this sample type has justified the discovery of large amount of proteins with structural function. In fact, for example, proteins of cytoskeleton, like tubulin alpha-3 and tubulin beta-4 identified by MS analysis (TBA3_ARATH, TBB4_ARATH), and pectins are involved in the cell wall structure and the changes in structure/chemistry of cell walls directly affect the ripening and senescence of plants [40-41].Moreover oxidoreductases are enzymes able to catalyse the transfer of electrons from one molecule to another one and they are normally involved in different metabolic processes like signalling and regulation of growth [39], transport activities and defense against pathogens [42].Another large percentage of identified proteins belongs to the class of dehydrogenases (7.6%), enzymes able to interconvert alcohols into aldehydes or ketones with the reduction of nicotinamide adenine dinucleotide (NAD^+ to NADH), like our identification NADH dehydrogenase iron-sulphur protein 3 (NDUS3_ARATH). They are expressed at low levels in roots of young plants grown on agar and at a high level in lack of oxygen or water and in low temperatures [43]. In addition, a particular dehydrogenase, the glyceraldehyde 3-phosphate dehydrogenase, identified via MS in our *Panax* extracts (Q6VAL5_PANGI, D0VFU1_PANGI), is involved in glycolysis, a pathway able to break starch into glucose. We speculate that this result could indicate that ginseng mainly absorbs energy as vegetative storage for ginseng root survival, which is destroyed through glycolysis to release the energy required for root maturation in slow-growth

period [24]. Another major functional class of proteins is represented by synthases (9.0%), which catalyzes numerous synthesis processes. For example, the overexpression of Panax ginseng squalene synthase causes a remarkable increase of phytosterols as well as ginsenoside contents [44]. Proteins belonging to the synthesis category such as granule-bond starch synthase I identified in our samples (Q6XY46_9APIA), increase with ginseng root age, indicating that protein synthesis enhances the ginseng maturation at different metabolic levels [9].

In order to investigate protein-protein interactions, a proteome-scale interaction network was created by merging all the proteins identified for the *Panaxginseng*proteome by using STRING software (Figure 7 S3). STRING databases include interactions from published literature describing experimentally studied interactions, as well as those from genome analysis by using several well-established methods based on domain fusion, phylogenetic profiling and gene neighborhood concepts. The whole *Arabidopsis Thaliana* genome was selected as a reference set, correlating the new results with the previous MS analysis. This network obtained represents the first comprehensive interactomics map for the *Panaxginseng* proteome and provides an interestingframework for navigating through the proteome. The topological analysis of this network has demonstrated several sparsely connected sub-networks, including that one corresponding to the ribosomal proteins with highly connected interactions(Figure 7 S3).The KEGG-Pathway Enrichment in STRING has revealed the presence of the canonical pathways: energy metabolism, carbohydrate and lipid metabolism and nucleotide and amino acid metabolism(Table 1).Among them the metabolic processes are the most relevant biological processes, representing the 65.9%of global dataset. KEGG pathway data have confirmed the proteins identifications by MS analysis: most of interactions involved proteins of energy metabolisms, like oxodoreductases recognized previously, of carbohydrate metabolism (glycolysis) such as glyceraldehyde 3-

phosphate dehydrogenase.Considering another pathway map class, the *Genetic Information Processes*has included 27 ribosomal proteins, which represented the top protein class of interactomics map. In conclusion, the investigation of the protein network, realized by using MS- proteomics data, has allowed the global protein–protein interactomicsinspection of *Panaxginseng* proteome, providing a first reference map for further proteomics studies.

4.3. Analysis of proteolytic peptides

In order to explore the possibility to identify bioactive peptides, encrypted in the global*Panax ginseng* proteome, the physiologic enzymatic digestion of proteins was simulated in-silico bysequential hydrolysis with pepsin (stomach) and intestinal enzymes (trypsin, chymotrypsin, elastase, carboxypeptidase A and B and aminopeptidases). Despite restricted parameters such as one missed cleavages and a minimum of 6 residues per peptide, smaller peptides were not considered in the study because their sequence might notbe unique and may belong to several different proteins. This model could not be completely prognostic because it didn't consider a lot of individual physiological factors, like pH variations, peptidase/protein ratio and interactions with other molecules. The predicted peptidomeresultedinto 661 different peptides, as shown in Figure 5, and it was subsequently investigated by using different databases that include known bioactive peptides. By using PeptideRanker (http://bioware.ucd.ie/~testing/ biowareweb/) the complete list of potential bioactive peptides was ranked using the N-to-1 neural network probability, which predicts more bioactive peptides (95). Among them, with a score higher than 0.5 (6–11residues), the majority corresponded to peptides encrypted in the nucleotide/nucleoside phosphate binding proteins(Figure 6).

Antimicrobial peptides (AMPs), characterized by score <−0.251, were identified using the CAMP (Collection of Anti-Microbial Peptides) database (http://www.bicnirrh.res.in/antimicrobial/) and applying the DAC score (Discriminate Analysis Classifier score). Table 2despicts6 potential bioactive peptides with anti-microbial activities. Two peptides belong to traslocon protein of chloroplast membrane, involved in protein precursor import to chloroplasts[45]. RPDRNF is a peptide of chloroplast protein Ycf2previously identified by MS analysis, whose function is still unknown even if it seems to be a ATPase[46]. Other two peptides (QNEWGW and ICCCNKM)derived from proteins involved in cell metabolism: in particular the second one derives from elongation factor 1-alpha recognized in Panax samples and it seems to be involved in protein biosynthesis inside ribosomes. The last peptide, GWNRSW, belongs to alpha-glucan water dikinase 2 protein, an enzyme of PEP-utilizing family.

Although all these potential bioactive peptides need to be validated by further bioactivity assays using corrspondentsynthetic peptides, we believe that the computational methods are useful for a preliminary peptidomic analysis. In fact they are fast and low cost alternatives,able to predict and reduce the number of potential targets to be investigated. In addition, bioinformatics-driven tools provide useful insights not achievable in human or animal model studies.

5. Conclusions

Up to the present the lack of complete knowledge of*Panax ginseng* proteome has contributed to a defective characterizationof ginseng proteinscomposition and functions. Despite numerous efforts to sequence *Panax ginseng*genome [47, 48], most research has focused onginsenosidesinvestigations [49-51]. This work provides new insight into the proteome of the Chinese traditional medical plant root, coupling the power of the Combinatorial Peptide Ligand Library (CPLL) technology,the SDS-PAGE separation

and thehigh resolution nLC-ESI MS/MS. We have successfully identified 55 proteins belonging to *Panax ginseng* database and 152 belonging to *Arabidopsis thaliana* database, for a total of 207 identifications. The GO software has permitted a classification of proteins in order to understand the role and the function of the identified species. The four major classes are represented by oxidoreductase (19.5%), structural compound (19.2%), synthase (9.0%) and dehydrogenase (7.6%). In order to explore proteins interactions useful to understand the *Panax ginseng* biology, an interactomics map was built, thus obtaining a complex grid formed by 196 nodes connected via 1554 interactions. The protein-protein interaction network was simplified in a table, reporting the pathways classification, which has confirmed the previous GOanalysis with a prevalence (65.9%) of proteins involved in metabolic pathways. Our proteomic research has concluded with a novelty in *Panax ginseng*studies: apeptidomic investigation by simulating *in-silico* the enzymatic protein digestion in the human gastrointestinal tract. A total of 661 different peptides were generated, but only 95 could be considered as potentially bioactive, including 6 with a predicted anti-microbial function. This approach could be an useful novelty to evaluate a possible antimicrobial activity expressed by proteolytic peptides, whilenormallyantioxidant and antimicrobialactions are mainly associated to ginsenosides present in *Panax ginseng* root. For example recently fresh root extracts, from 4 years old *Panax ginseng* plants, have been explored for the synthesis of silver and gold nanoparticles, characterized by potentialantimicrobialfunctions against pathogenic microorganisms [52-54].In conclusion our proteomic and peptidomic analysis aimedto provide a deep investigation of *Panax ginseng* components for a better understanding of ginseng metabolism connected with beneficial effects, which have contributed to the diffusion as the most valuable traditional Asian medicine.

Supplementary data

Table S1: Lists of the 152 unique gene products identified by the use of *Arabidopsis thaliana*database, of 55 proteins identified by the use of *Panax ginseng*databaseand of 207 unique identifications considering both searches.

TableS2: Lists of 661 different peptides generated by *in-silico* enzymatic digestion and of 95 potential bioactive peptides.Report of complete annotations about KEGG pathway refered to peptides.

Figure 7 S3: interactomics map realized by STRING v.9.1 software, showing a protein-protein interaction network formed by 196 nodes and 1554 interactions.

Acknowledgments

The authors wish to thank Aboca S.p.a., Sansepolcro (AR), Italy for the supply of *Panax ginseng* root powder and Dr Peter R Baker, from University of California San Francisco (NIGMS Mass Spectrometry Facility), for its support in the peptidomic analysis.

Legends

Figure 1

Scheme of complete treatment protocol of *Panax ginseng* root's powder. The capturewith CPLLs is performed at three pH values, namely 2.2, 7.2 and 9.0.

Figure 2

SDS-PAGE separation of the untreated (Raw) sample vs. the eluates of CPLL capture at the three different pH values (tracks E2.2, E7.2 and E9.0). Each lane has been cut into segments, the proteins digested and sent to MS analysis. Mr: molecular mass ladder. Staining with micellar Coomassie blue.

Figure 3

MS data of the identified proteins. Upper panel: table of protein IDs as obtained in a raw extract and all CPLLs eluates by using both*Panax ginseng* and *Arabidopsis thaliana*databases, respectively. The total unique IDswere 152 for *Arabidopsis thaliana* database and 55 for *Panax ginseng*. Lower panel: Venn diagrams comparing the number of identified proteins in all CPLLs eluates with those recognized in the raw extract, considering both databases.In reagard to the *Panax ginseng*db, 37 proteins were found both in untreated and eluted samples, while 13 were specifically of CPLLs eluates. Moreover a quite high number of identifications was specific of each beads capture (e.g. 8 specific proteins for CPLLs treatment at pH 7,2). Also for *Arabidopsis Thaliana*db 81 were common identifications from all samples, while 43 were exclusive proteins of three eluates. The higher number of 27 specific proteins was obtained after beads capture at pH 9,0.

Figure 4

Biological funtional categorization of 207 unique proteins expressed in ginseng dried root powder via GO analysis.The pie chart shows the percentage distribution of reported GO terms.

Figure 5

In-silico simulation scheme of human gastro-intestinal digestion, performed on all proteins identified via the *Arabidopsis thaliana* database. All the listed digestive enzymes generated a total number of 661 peptides.

Figure 6

Representation of 95 proteins releasing bioactive peptides, clusterized by KMeans; bioactive peptides have been ranked on the bases of their aPeptideranker score >0.5 calculated by the corresponding software. All nodes in interactomic map were proteins, showing different activities like antimicrobial or defensive functions.

Table 1 – The KEGG-Pathway enrichment,performed by STRING software, has revealed the presence of canonical pathways: metabolic pathways (65.9%), genetic information processes (27.4%), cellular transport and catabolism (5.8%) and environment adaptation (0.9%).the numbers of involved genes are reported for all subcategories of enriched cell pathways.

KEGG PATHWAY	N° of genes
Metabolicpathways 65.9%	
Energy metabolism	
Carbon fixation in photosynthetic organisms	10
Oxidativephosphorylation	8
Nitrogenmetabolism	3
Photosynthesis	2
Sulfurmetabolism	1
Carbohydrate and lipidmetabolism	
Citratecycle (TCA cycle)	10
Pyruvatemetabolism	8
Glycolysis / Gluconeogenesis	8
Starch and sucrosemetabolism	8
Glyoxylate and dicarboxylatemetabolism	7
Pentosephosphatepathway	6
Butanoatemetabolism	4
Ascorbate and aldaratemetabolism	3
Pentose and glucuronateinterconversions	2
Galactosemetabolism	2
Fructose and mannosemetabolism	1
Sugar metabolism	
Amino sugar and nucleotide sugar metabolism	8
Fatty acid metabolism	
Arachidonic acid metabolism	1
Glycanmetabolism	
N-Glycanbiosynthesis	1
Terpenoidbackbonebiosynthesis	1
Sesquiterpenoidbiosynthesis	1
Otherterpenoidbiosynthesis	
Limonene and pinene degradation	2
Carotenoidbiosynthesis	1
Nucleotide and amino acid metabolism	
Purine metabolism	4
Pyrimidinemetabolism	1
Amino acid metabolism	
Alanine, aspartate and glutamatemetabolism	6
Cysteine and methioninemetabolism	5
beta-Alanine metabolism	3
Phenylalaninemetabolism	4

Arginine and proline metabolism	2
Glycine, serine and threonine metabolism	1
Tyrosinemetabolism	1
Tryptophanmetabolism	1
Other amino acid metabolism	
Glutathionemetabolism	6
Taurine and hypotaurinemetabolism	3
Selenocompoundmetabolism	2
Cyanoamino acid metabolism	1
Metabolism of cofactors and vitamins	
One carbon pool by folate	1
Biosynthesis of secondarymetabolites	
Phenylpropanoidbiosynthesis	4
Stilbenoid, diarylheptanoid and gingerol biosynthesis	3
Flavonoidbiosynthesis	1
Genetic Information Processing 27.4%	
Translation	
Ribosome	27
mRNAsurveillancepathway	3
RNA transport	2
Aminoacyl-tRNAbiosynthesis	1
Transcription	
Spliceosome	5
RNA polymerase	1
Folding, sorting and degradation	
Protein processing in endoplasmic reticulum	11
Proteasome	8
RNA degradation	3
Cellular Processes 5.8%	
Transport and catabolism	
Endocytosis	6
Peroxisome	4
Phagosome	3
Organismal Systems 0.9%	
Environmentaladaptation	
Plant-pathogeninteraction	2

Peptide RankerScore	PeptidesSequence	Peptides Lenght	Class	AMP Probability	Accession Number	ProteinDescription	Protein Family	Function
0,73	NQWKNW	6	AMP	0,989	P56785	Full=Protein TIC 214; AltName: Full=Translocon at the inner envelope membrane of chloroplasts 214;	Belongs to the TIC214 family The Chloroplast Envelope Protein Translocase (CEPT or Tic-Toc) Family P56785 Putative membrane protein ycf1 Q8LPR8 AT5g22640/MDJ22_6 Q7Y1Z4 Putative uncharacterized protein At5g01590 Q8GZ79 Protein TIC 20-I, chloroplastic Q9FMD5 Protein TIC 40, chloroplastic Q8LPR9 Protein TIC110, chloroplastic Q8VY88 Protein LHCP TRANSLOCATION DEFECT Q9FI56 Chaperone protein ClpC1, chloroplastic Q9SXJ7 Chaperone protein ClpC2, chloroplastic	Involved in protein precursor import into chloroplasts. May be part of an intermediate translocation complex acting as a protein-conducting channel at the inner envelope.
0,5	KNNYDF	6	AMP	0,905	P56785	Full=Protein TIC 214; AltName: Full=Translocon at the inner envelope membrane of chloroplasts 214;		
0,73	RPDRNF	6	AMP	0,88	P56786	Protein Ycf2 (chloroplast)	Belongs to the Ycf2 family	This family consists of chloroplast encoded Ycf2, which is around 2000 residues in length. The function of Ycf2 is unknown, though it may be an ATPase. Its retention in reduced chlorplastgeneomes of non-photosynthetic plants, e.g. Epifagusvirginiana (Beechdrops), and transformation experiments in tobacco indicate that it has an essential function which is probably not related to photosynthesis
0,83	QNEWGW	6	AMP	0,774	P25696	Full=Bifunctional enolase 2/transcriptional activator; AltName: Full=2-phospho-D-glycerate hydro-lyase 2; AltName: Full=2-phosphoglycerat	Belongs to the enolase family	Multifunctional enzyme that acts as an enolase involved in the metabolism and as a positive regulator of cold-responsive gene transcription. Binds to the cis-element the gene

						e dehydratase 2;		promoter of STZ/ZAT10, a zinc finger transcriptional represso
0,81	ICCCNKM	7	AMP	0,699	A8MSE8	Elongationfactor 1-alpha	Belongs to the TRAFAC class translation factor GTPase superfamily. Classic translation factor GTPase family. EF-Tu/EF-1A subfamily.	This protein promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes during protein biosynthesis
0,85	GWNRSW	6	AMP	0,603	Q9STV0	Alpha-glucan water dikinase 2	Belongs to the PEP-utilizing enzyme family	Mediates the incorporation of phosphate into alpha-glucan, mostly at the C-6 position of glucose units

Table 2– Itlists 6 peptides characterized by antimicrobial function (AMP), reporting all parameters like PeptideRanker score, AMP probability,peptide sequence and length. Also other parameters, referred to the original proteins, like accession number, protein family and function, are shown.

References

[1] Lee FC. Facts about Ginseng, the Elixir of Life. Hollyn International Corp., Elizabeth, NJ, 1992.

[2] Huang KC.The Pharmacology of Chinese Herbs. CRC Press, Boca Raton, FL, 1999.

[3]A.S. Attele, J.A. Wu, C.-S. Yuan. Ginseng Pharmacology.Biochem. Pharmacol., 1999; 58:1685-1693.

[4]Gillis CN, *Panax ginseng* pharmacology: A nitric oxide link?Biochem. Pharmacol., 1997,54**:** 1–8.

[5] You J., Liu X., Zhang B., Xie Z., Hou Z., Yang Z., Seasonal changes in soil acidity and related properties in ginseng artificial bed soils under a plastic shade. J. Ginseng Res. 2015, 39:81-88.

[6] Kim C., Choo G. C., Cho H. S., Lim J. T., Soil properties of cultivation sites for mountain-cultivated ginseng at local level.J. Ginseng Res. 2015, 39: 76-81.

[7]Kim Y. K., Jeon J. N., jang M. G., Oh J. Y., Kwon W. S., Jung S. K., Yang D. C., Ginsenoside profiles and related gene expression during foliation on *Panax*. J. Ginseng Res. 2014, 38: 66-72.

[8]Lee J. H., Lee J. S., Kwon W. S., Kang J. Y., Lee D. Y., In J. G., Kim Y. S., Seo J., Baeg J. S., Chang I. M., Grainger K. Characteristics of Korean ginseng varieties of Gumpoong, Sunun, Sunpoong, Sunone, Cheongsun and Sunhyang.J. Ginseng Res. 2015, 39: 91-104.

[9] Ma R., Sun L., Chen X., Jiang R., Sun H., Zhao D. Proteomic changes in different growth periods of ginseng roots. Plant Phys. and Biochem. 2013, 67: 20-32.

[10] Nam M. H., Heo E. J., Kim J. Y., Kim S., Kwon K. H., Se J. B., Kwon O., Yoo J. S., Park Y. M. Proteome analysis of the response of Panax ginseng C. A. Meyer leaves to high light: use of electrospray ionization quadrupole-time of flight massa spectrometry and expressed sequence tag data. Proteomics 2003, 3: 2351-2367.

564 [11] B.X. Wang, Q.L. Zhou, M. Yang, Y. Wang, Z.Y. Cui, Y.Q. Liu, I. Takashi, Hypo-
565 glycemic activity of ginseng glycopeptides, ActaPharmacol. Sin. 2003, 24: 50-54.
566 [12] J.H. Kang, K.H. Song, J.K. Woo, M.H. Park, M.H. Rhee, C. Choi, S.H. Oh, Ginse-
567 noside Rp1 from Panax ginseng exhibits anti-cancer activity by down- regulation of the
568 IGF-1R/Akt pathway in breast cancer cells, Plant Food Hum. Nutr. 2011, 66: 298-305.
569 [13]Sharma J., Goyal P. K. Chemoprevention of chemical-induced skin cancer by *Panax*
570 *ginseng* root extract. J. Ginseng Res. 2015, 39: 265-273.
571 [14]Kim S. J., Kim A. K., Anti-breast cancer activity of fine black ginseng (*Panax*
572 *ginseng* Meyer) and ginsenoside Rg5. J. Ginseng Res. 2015, 39: 125-134.
573 [15] T. Yokozawa, E. Dong, H. Watnabe, H. Oura, H. Kashiwagi, Increase of active
574 oxygen in rats after nephrectomy is suppressed by ginseng saponin. Phyt- other. Res.
575 1996, 10: 569-572.
576 [16] Y. Wanga, R. Jiangb, G. Lic, Y. Chenb, H. Luob, Y. Gaob, Q. Gao, Structural and
577 enhanced memory activity studies of extracts from Panax ginseng root, Food Chem.
578 2010, 119: 969-973.
579 [17] J.L. Reay, A.B. Scholey, D.O. Kennedy, Panax ginseng (G115) improves aspects
580 of working memory performance and subjective ratings of calmness in healthy young
581 adults, Hum. Psychopharm. Clin. 2010, 25: 462-471.
582 [18] Baek K. S., Hong Y. D., Kim Y., Sung N. Y., Yang S., Lee K. M., Park J. Y., Park J.
583 S., Rho H. S., Shin S. S., Cho J. Y. Anti-inflammatory activity of AP-SF, a ginsenoside-
584 enriched fraction, from Korean ginseng. J. Ginseng Res. 2015, 39: 155-161.
585 [19] Lee C. H., Kim J. H., A review on the medical potentials of ginseng and
586 gensenosides on cardiovascular diseases. J. Ginseng Res. 2014, 38: 161-166.
587 [20] Liang X., Chen X., Liang Q., Zhang H., Hu P., Wang Y., Luo G. Metabonomic study
588 of chinese medicine *shuanglong* formula as an effective treatment for myocardial
589 infarction in rats. J. Prot. Res. 2011, 10: 790-799.

590 [21] J.H. Lum, K.L. Fung, P.Y. Cheung, M.S. Wong, C.H. Lee, F.S. Kwok, M.C. Leung,
591 et al., Proteome of oriental ginseng Panax ginseng C.A. Meyer and the potentialto use it
592 as an identification tool, Proteomics 2002, 2: 1123-1130.
593 [22]Kim SI, Kim JY, Kim EA, Kwon KH, Kim KW, Cho K, Lee JH, Nam MH, Yang DC,
594 Yoo JS, Park YMProteome analysis of hairy root from Panax ginseng C.A. Meyer using
595 peptide fingerprinting, internal sequencing and expressed sequence tag data.
596 Proteomics. 2003, 3: 2379-2392.
597 [23]J.Y. Yoon, B.H. Ha, J.S. Woo, Y.H. Lim, K.H. Kima, Purification and
598 characterizationof a 28-kDa major protein from ginseng root. Comp. Biochem.
599 Physiol.Part B 2002, 132: 551-557.
600 [24] S.I. Kim, S.M. Kweon, E.A. Kim, J.Y. Kim, S. Kim, J.S. Yoo, Y.M. Park,
601 Characterizationof RNase-like major storage protein from the ginseng root byproteomic
602 approach, J. Plant Physiol. 2004, 161: 837-845.
603 [25] M.H. Nam, S.I. Kim, J.R. Liu, D.C. Yang, Y.P. Lim, K.H. Kwon, J.S. Yoo, Y.M.
604 Park,Proteomic analysis of Korean ginseng (Panax ginseng C.A. Meyer). J.
605 Chromatogr. B. 2005, 815: 147-155.
606 [26] L. Sun, X. Lei, R. Ma, R. Jiang, D. Zhao, Y. Wang, Two-dimensional gel
607 electrophoresisanalysis of different parts of Panaxquinquefolius L. root. Afr.
608 J.Biotechnol. 2011, 10: 17023-17029.
609 [27] L. Sun, P. Ma, R. Ma, X. Lei, X. Chen, C. Qi, Two-dimensional
610 electrophoresisanalysis for different parts of Panax ginseng C.A. Meyer root. J.
611 HuazhongNorm. Univ. (Nat. Sci.), 2010, 44: 639-643.
612 [28] Wang W., Vignani R., Scali M., Cresti M., A universal and rapid protocol for
613 proteinextraction from recalcitrant plant tissues forproteomic analysis. Electrophoresis
614 2006, 27: 2782-2786.

615 [29] Fasoli E., Righetti P. G., Proteomics of fruits and beverages,Curr.Opin. Food
616 Science 2015, 4:76–85.
617 [30] Fasoli E, Farinazzo A, Sun CJ, Kravchuk AV, Guerrier L, Fortis F, Boschetti E,
618 Righetti PG. Interaction among proteins and peptide libraries in proteome analysis: pH
619 involvement for a larger capture of species. J Proteomics 2010, 73:733-742.
620 [31] Candiano G, Dimuccio V, Bruschi M, Santucci L, Gusmano R, Boschetti E, Righetti,
621 PG, Ghiggeri GM. Combinatorial peptide ligand libraries for urine proteome analysis:
622 investigation of different elution systems. Electrophoresis 2009, 30:2405-2411.
623 [32] Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T.,
624 Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-
625 electrospray mass spectrometry. Nature 1996, 379, 466– 469.
626 [33] Bradford M. M. A rapid and sensitive method for the quantification of microgram
627 quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 1976,
628 7: 248-254.
629 [34] Righetti PG, Fasoli E, Boschetti E. Combinatorial peptide ligand libraries: the
630 conquest of the 'hidden proteome' advances at great strides.Electrophoresis.
631 2011;32:960-966.
632 [35] U. Restuccia, E. Boschetti, E.Fasoli, F. Fortis, L. Guerrier, A. Bachi, AV. Kravchuk,
633 P.G. Righetti. pI-based fractionation of serum proteomes versus anion exchange after
634 enhancement of low-abundance proteins by means of peptide libraries. J Proteomics.
635 2009, 72:1061-1070.
636 [36] E. Boschetti; L.Bindschedler; C. Tang; E. Fasoli; P.G. Righetti. Combinatorial
637 peptide ligand libraries and plant proteomics: a winning strategy at a price. Journal of
638 Chromatography A, 2009, 1216:1215-1222.
639 [37] Fasoli E, Righetti PG. The peel and pulp of mango fruit: A proteomic samba.
640 BiochimBiophys Acta, 2013;1834:2539-2545.

641 [38] Righetti PG, Esteve C, D'Amato A, Fasoli E, Luisa Marina M, Concepción García
642 M. A sarabande of tropical fruit proteomics: Avocado, banana, and mango.Proteomics
643 2015,15:1639-1645.
644 [39] Robinson N. G., Procter C. M., Connoly E. L., Guerinot M. L. A ferric-chelate
645 reductase for iron uptake from soil. Nature 1999, 397:694-697.
646 [40] Brummell D. A., Harpster M. H. Cell wall metabolism in fruit softening and quality
647 and its manipulation in transgenic plants. Plant Mol. Biol. 2011,47:311-340.
648 [41] Brummell D. A., Cin V. D., Crisosto C. H., Labavitch J.M. Cell wall metabolism
649 during maturation, ripening and senescence of peach fruit. J. Exp. Bot. 2004,55:2029-
650 2039.
651 [42] Bandaranayake P.C.G., Filappova T., Tomilov A., Tomilova N. B., Jamison-
652 McClung D., Ngo Q., Inoue K., Yoder J.I. A single-electron reducing quinone
653 oxidoreductase is necessary to induce haustorium development in the roor parasitic
654 plant *Triphysaria*. The Plant Cell 2010,22:1404-1419.
655 [43] Chung H. J., Ferl R.J. Arabidopsis alcohol dehydrogenase expression in both
656 shoots and roots is conditionated by root growth environment. Plant Phys. 1999,121:
657 429-436.
658 [44] Lee M. H., Jeong J.H., Seo J.W., Shin C.G., Kim Y.S., In J.G., Yang D.C., Yi J.S.,
659 Choi Y.E. Enhanced triterpene and phytosterol biosynthesis in Pnax ginseng
660 overexpressing squalene synthase gene. Plant Cell Physiol. 2004,45:976-984.
661 [45] Drescher A, Ruf S, Calsa T Jr, Carrer H, Bock R. The two largest chloroplast
662 genome-encoded open reading frames of higher plants are essential genes. Plant
663 J. 2000,22:97-104.
664 [46] Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide
665 T, Nakai M. Uncovering the protein translocon at the chloroplast inner envelope
666 membrane. Science. 2013, 339(6119):571-574.

[47] Kim K, Lee SC, Lee J, Kim NH, Jang W, Yang TJ. The complete chloroplast genome sequence of *Panaxquinquefolius* (L.). Mitochondrial DNA. 2015, 10:1-2.

[48] Jayakodi M, Lee SC, Lee YS, Park HS, Kim NH, Jang W, Lee HO, Joh HJ, Yang TJ.Comprehensive analysis of Panax ginseng root transcriptomes. BMC Plant Biol.2015,15:138.

[49] Kim SJ, Kim AK.Anti-breast cancer activity of Fine Black ginseng (Panax ginseng Meyer) and ginsenoside Rg5.J Ginseng Res. 2015, 39:125-134.

[50] Park JY, Choi P, Kim T, Ko H, Kim HK, Kang KS, Ham J.Protective Effects of ProcessedGinsengand Its Active Ginsenosides on Cisplatin-Induced Nephrotoxicity: In Vitro and in Vivo Studies.J Agric Food Chem. 2015, 63:5964-9.

[51] Jun YL, Bae CH, Kim D, Koo S, Kim S.Korean Red Ginseng protects dopaminergic neurons by suppressing the cleavage of p35 to p25 in a Parkinson's disease mouse model.J Ginseng Res. 2015, 39:148-154.

[52] Shin HS, Yu M, Kim M, Choi HS, Kang DH. Renoprotective effect of red ginseng in gentamicin-induced acute kidney injury.Lab Invest. 2014,94:1147-1160.

[53]Singh P, Kim YJ, Wang C, Mathiyalagan R, El-AgamyFarh M, Yang DC.Biogenic silver and gold nanoparticles synthesized using redginsengroot extract, and their applications.Artif Cells NanomedBiotechnol. 2015, 23:1-6.

[54] Singh P, Kim YJ, Wang C, Mathiyalagan R, Yang DC.The development of a green approach for the biosynthesis of silver and gold nanoparticles by usingPanaxginsengroot extract, and their biological applications.Artif Cells NanomedBiotechnol. 2015, 14:1-8.

Figure

Click here to download Figure: Figures_Colzani_JOP_FinalRev.pptx

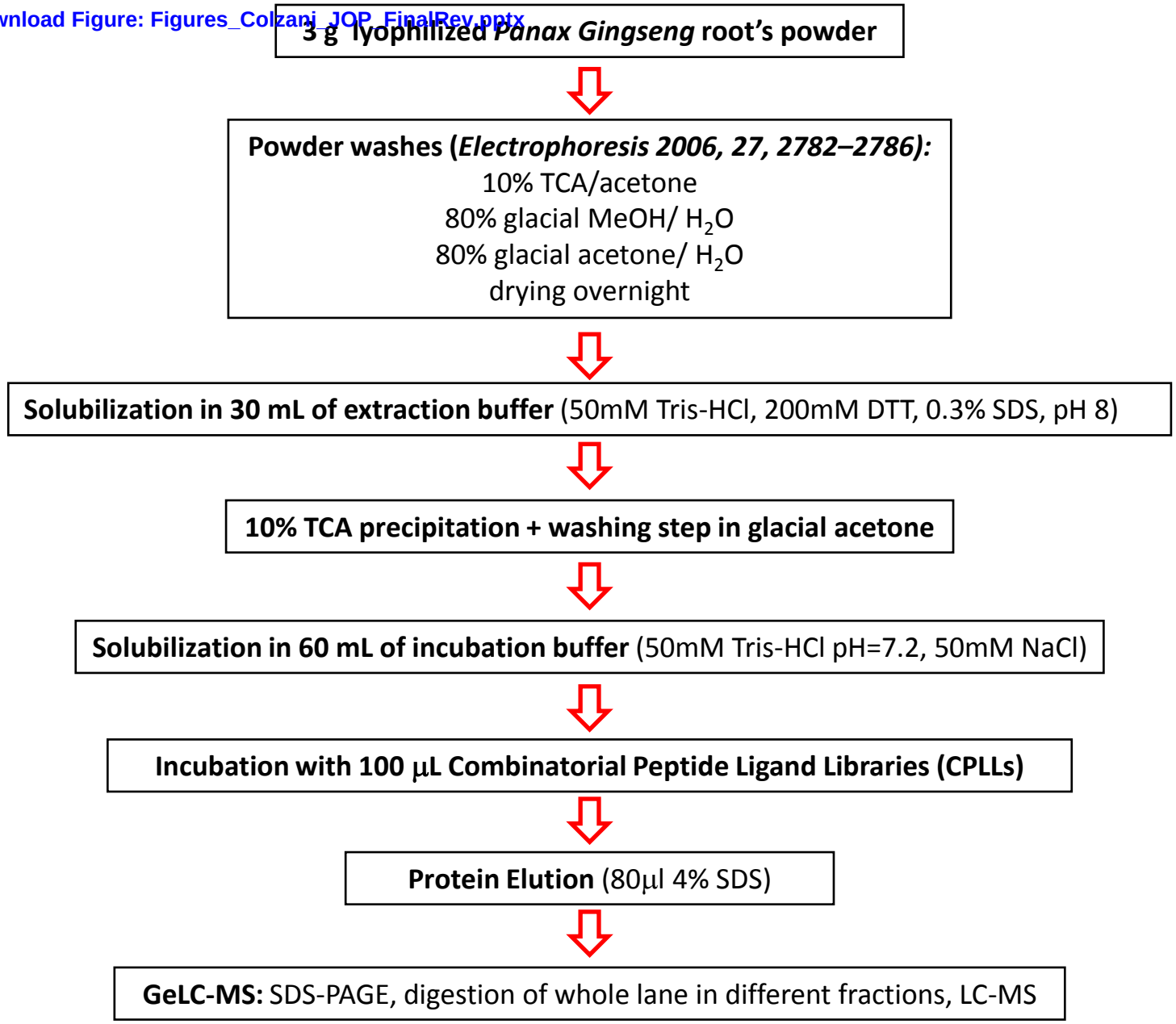


Figure 1 *Colzani et al.*

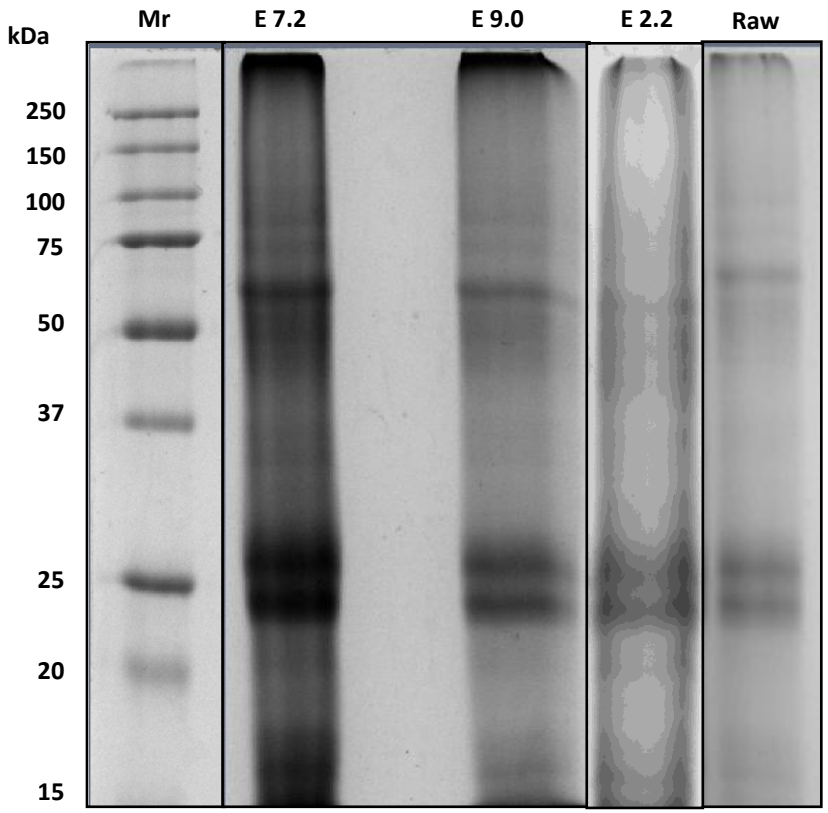


Figure 2 *Colzani et al.*

Database	Database entries	Raw extract	CPLLs pH 7.2	CPLLs pH 2.2	CPLLs pH 9.0	Total ID
Arabidopsis Thaliana	31706	108	83	51	96	**152**
Panax Ginseng	640	42	44	38	31	**55**

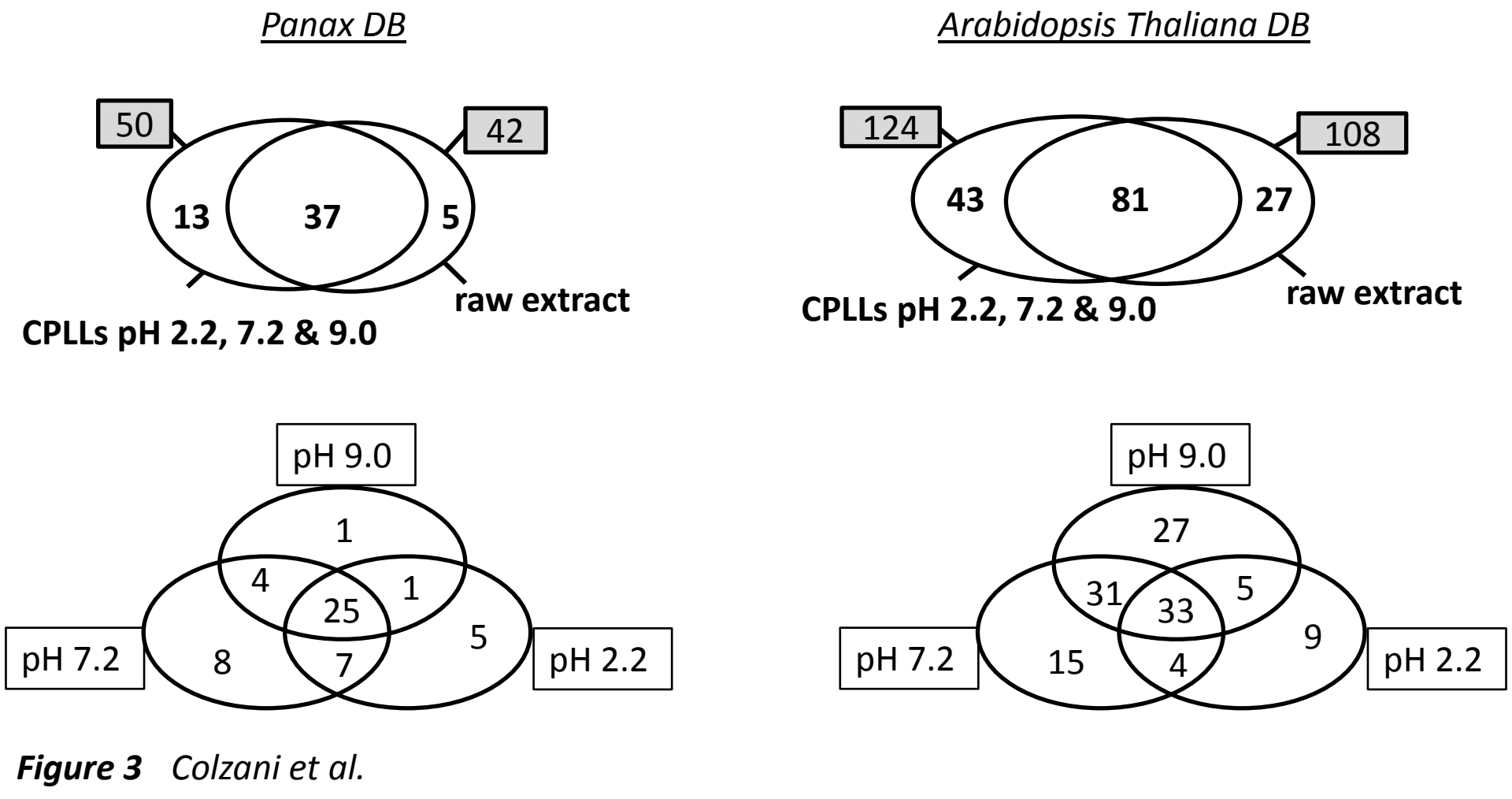


Figure 3 *Colzani et al.*

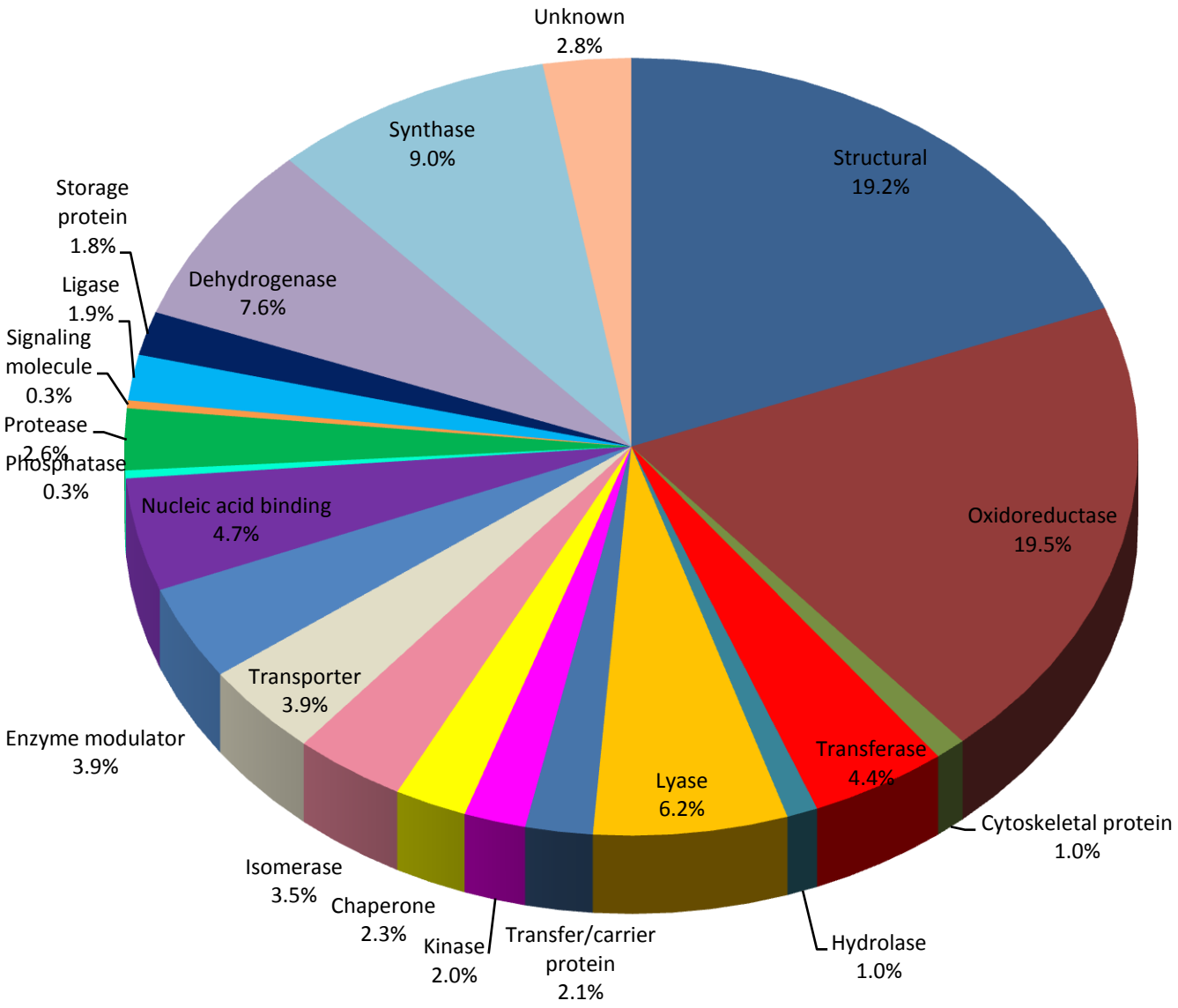


Figure 4 *Colzani et al.*

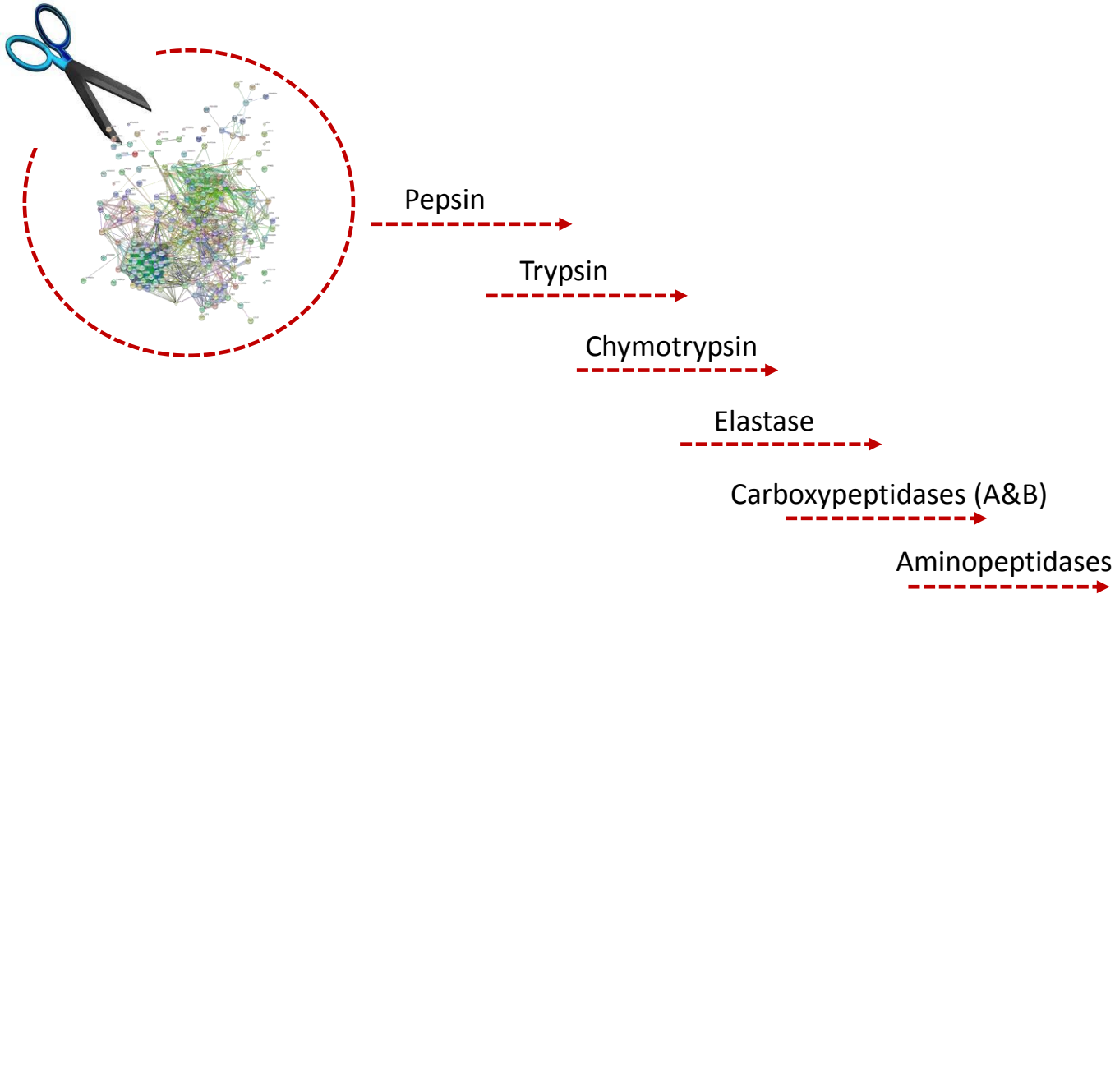


Figure 5 *Colzani et al.*

95 Bioactive Pepides

Score > 0,5 in PeptideRanker
Released from clusterized proteins by *K Means*

Proteins with AMP peptides

Antimicrobial proteins

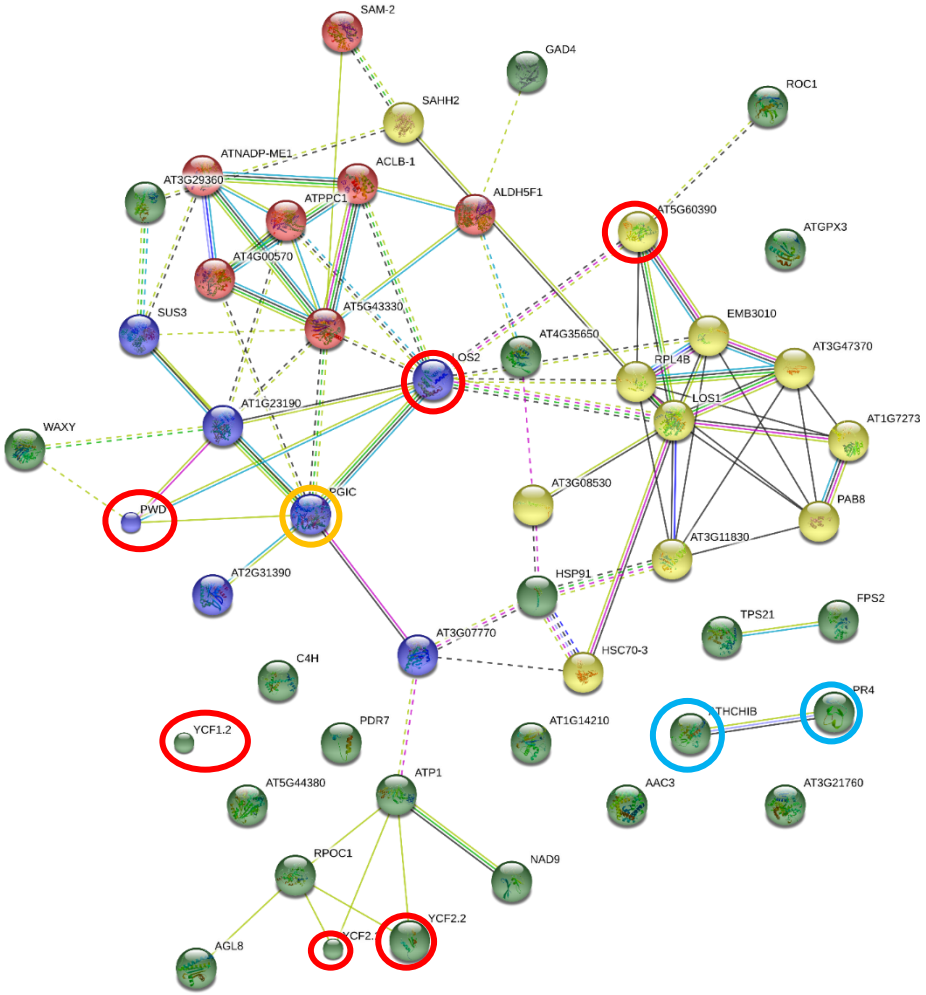


Figure 6 *Colzani et al.*

Supplementary material

Click here to download Supplementary material: S1_Colzani et al.xlsx

Supplementary material

Click here to download Supplementary material: S2_Colzani et al.xlsx

Supplementary material

Click here to download Supplementary material: S3_Figures_Colzani.pptx

***Conflict of Interest**

Click here to download Conflict of Interest: COI_Fasoli.pdf

***Conflict of Interest**

Click here to download Conflict of Interest: COI_Righetti.pdf

***Conflict of Interest**

Click here to download Conflict of Interest: COI_Colzani.pdf

***Conflict of Interest**

Click here to download Conflict of Interest: COI_Altomare.pdf

***Conflict of Interest**

Click here to download Conflict of Interest: COI_Caliendo.pdf

***Conflict of Interest**

Click here to download Conflict of Interest: COI_Aldini.pdf