UNIVERSITÀ DEGLI STUDI DI MILANO

DEPARTMENT OF PHARMACEUTICAL SCIENCES

DOCTORAL SCHOOL IN PHARMACEUTICAL SCIENCES

PhD Cycle the 30th

NUTRITIONAL PEPTIDOMICS: DISCOVERY, QUANTIFICATION, AND FUNCTIONAL ANALYSIS OF PLANT PROTEIN DERIVED PEPTIDES

Sector CHIM/10 - Food chemistry

Dr. Aiello Gilda

R10872

Faculty advisor: Prof. Anna Arnoldi

PhD coordinator: Prof. Giancarlo Aldini

ACADEMIC YEAR

2016/2017

Dedicated to the persons I love

"Everything and every event of the physical universe require, to justify its existence, the use of something else, outside of them. We must therefore resort to something non-physical and supernatural: God"

Davies Paul

Davies, God and the new physics, 1984

Table of Contents

List of Tables

List of Supplementary Tables

List of Figures Page

List of Supplementary Figures

INDEX

Aim of the thesis

Part I. State of the art

Part II. Scientific contributions

Part III. Concluding Remarks

Appendix

Scientific Publications & Communications

Aim of the thesis

Motivations and outlines of the PhD thesis

Recently the increase in human diseases and side effects of drugs has promoted research in food components endowed with biologically active molecules. These bioactive components play a pivotal role in reducing and regulating the onset of such chronic diseases. The great interest seems to be focused on protein and peptides mostly derived from seeds of crop plants, which represent an excellent starting point for therapeutics. Actually, soy and lupin proteins have been largely investigated for their healthy properties in cardiovascular prevention for improving the clinical state of patients suffering of diabetes, hypertension, obesity and hypercolesterolemia. The interest into the consumption of food with potential bioactivities coincides with a clear shift in medicine and biosciences toward prevention of future diseases through adequate food intakes. In particular, the use of food-derived peptides, especially those derived from plants, could be appreciated, since they are recognised for being highly selective, efficacious and, at the same time, relatively safe and well tolerated by humans. The identification of biologically active peptides in food matrices and as well as the investigation of their pharmacological promise is, therefore, a challenging task in food technology.

However, although numerous studies report several modulating activities attributable to natural peptides, the main issue concerns their ability to cross the cell membrane for being active. The matter of bioabsorption has to be faced in order to explore their concrete potential.

Based on these considerations, the main aims of my PhD thesis were focused on the discovery, quantification and functional analysis of plant protein-derived peptides. These objectives were reached by use of different tools, comprising advanced analytical techniques, such as mass spectrometry (MS) coupled to biochemical and bioinformatics tools, in order to improve the understanding of their mechanism of action at cellular level, their intestinal absorption profile as well as to forecast their potential bioactivities.

Briefly, the main aims of the present study were:

- To investigate the bioavailability, the intestinal absorption and the bioactivity of lupin peptides using an *in vitro* epithelial model (Caco-2 cells).
- To evaluate the absorption and the metabolic degradation of soy peptides at Caco-2 cell surface.
- To develop qualitative (untargeted) and quantitative (targeted) LC-MS methods for the characterization of the pool of absorbed peptides (discovery phase) or quantify the most active ones obtaining simultaneously semi-quantitative information about their absorption.
- To investigate the wide proteome of hempseeds including the identification of the minor protein components.
- To produce hydrolysates of bioactive peptides with hypocholesterolemic properties from hempseed proteins using different enzymatic proteases.
- To investigate the functional and descriptive proteome of apricot kernel, focusing the attention on the *in silico* prediction of the potential bioactive peptides released by simulated gastrointestinal digestions.

The thesis contains three parts:

Part I introduces the state of art of bioactive peptides.

In **Chapter 1** an introduction about bioactive peptides, peptidomics and as well as mass spectrometry techniques is provided. Specifically, the underlying biological question concerning the bioavailability and bioabsorption of plant protein-derived peptides, such as *Lupinus* and *Glycine max*, is introduced.

Chapter 2 is dedicated prevalently to the strategies used in the peptidomic and on the technical issues in bioactive peptide discovery.

Part II contains my own scientific contributions.

The first original one is presented in **Chapter 3** with a novel methodologic approach aimed at the investigation of the potentially absorbed peptides derived from lupin proteins through gastrointestinal system. By the use of untargeted MS assay, this contribution explores the pool of lupin-absorbed peptides across the Caco-2 systems and the investigation of their capacity to modulate the cholesterol metabolism through the inhibition of 3-hydroxy-3-methylglutaryl

coenzyme A reductase (HMGCoAR). A further *in silico* docking simulations was applied to hypothesise which peptides may bind more efficiently to the HMGCoAR.

Moving the focus on a different molecular target, i.e. Proprotein convertase subtilisin/kexin type 9 (PCSK9), which it has been recently recognised as a new useful target for hypercholesterolemia treatment, a refined *in silico* model for the inhibition of the PCSK9/LDLR interaction mediated by absorbed-lupin peptides, was developed. Afterwards, a MS-quantitative evaluation of some selected absorbed peptide, identified as the best candidate inhibitors of PCSK9/LDLR binding was performed as reported in **Chapter 4.**

Chapter 5 is dedicated to the investigation of the absorption and the metabolism of soy peptides across Caco-2 cell monolayers. This contribution provides: i) the development of fit-for-purpose analytical methodology for quantifying three synthetic soy peptides with cholesterol-lowering activity through small intestine epithelium, ii) the evaluation of peptide degradation by epithelial peptidases.

Finally, **Chapter 6** and **Chapter 7** are dedicated to proteome and peptidome exploration of hempseeds, respectively. Considering the recent interest in hemp food products and the missing information about the protein composition of this seed, the first objective was to provide a deeper characterization of its proteome. Combinatorial peptide ligand libraries (CPLLs), a well-established technique able to enhance the presence of low or very low abundant proteins, have been adopted for an extensive investigation of the proteome of hempseed flour.

Chapter 7 and 8 present an exploration of potentially bioactive peptides generated from the enzymatic hydrolysis of hempseed proteins using different proteases in order to screen which were the most active ones on the modulation of cholesterol metabolism. The overall objectives of this last research were: i) to identify the peptide composition of hempseed protein hydrolysates by MS, ii) to determine the ability of hempseed protein hydrolysates to reduce the low-density lipoprotein (LDL) cholesterol level through the inhibition of HMGCoAR, focusing the attention mainly on peptic hydrolysate, iv) the prediction of additional biological activities using *in silico* bioinformatics tools. **Chapter 9** presents the proteomic mapping of apricot and the *in silico* release of bioactive peptides derived from the most abundant proteins through a simulated gastrointestinal digestion.

Finally, **Part III** presents the concluding remarks and the impact of this thesis on the field of biopeptides discovery. In addition, the appendix contains the skills acquired during the foreign study experience, with specific reference to very innovative proteomic approaches.

Part I.

State of the art

CHAPTER 1

1 Introduction

1.1 Peptide therapeutics: current status and future directions

More than 7000 naturally occurring peptides, having crucial roles in human physiology, including actions as hormones, neurotransmitters, growth factors, ion channel ligands, or anti-infective, have been recently identified.[1-4] Within the organism, peptides may exert diverse biological roles, most prominently as signalling/regulatory molecules in a broad variety of physiological processes, including defence, immunity, stress, growth, homeostasis, and reproduction.[5] Based on these features, peptides have been of great interest from the point of view of different disciplines, such as pharmaceutical, clinical, functional foods or nutriomics.[6] Given their attractive pharmacological profile and intrinsic properties, natural peptides represent an excellent starting point for the design of novel therapeutics and their specificity has been seen to translate into excellent safety, tolerability, and efficacy profiles in humans. In this scenario, new tendencies in the discovery of drug candidates involve the study of natural bioactive peptides since their accumulation in organs, such like kidney and liver, is very low, resulting in a minimal development of the most severe toxic side-effects.[7] This aspect might also be the primary differentiating factor of peptides compared with traditional small molecules. In addition, therapeutic peptides are typically associated with lower production complexity compared with protein-based biopharmaceuticals and, therefore, the production costs could be lower, generally approaching those of small molecules.

In the face of these positive aspects, however, in the past naturally occurring peptides were often considered not directly suitable for use as convenient therapeutics, because they show intrinsic weaknesses, including poor chemical and physical stability, and short circulating plasma-life. Despite such neglect, a number of recent technological breakthroughs and advances have sparked major interest in their usage as both diagnostic as well as therapeutics. Peptides have greater potential to meet the ever-increasing expectations of new drugs, as they are highly specific to individual protein targets, amenable to site-specific modification and

highly selective. In other words, some attributes originally considered troublesome in drug development may now turn out to be more convenient rather than unfavourable.

1.2 Peptide drug market between food and pharma

In the current pharmaceutical industry, peptides are considered mature compounds addressing unmet medical needs, and accelerating the personalised medicine model.[8] In details, peptides promise to combine the lower production costs of conventional (small molecular chemical) drugs with the high specificity of (the larger) biological entities. The market contains already peptides used as drugs, i.e. LupronTM from Abbott Laboratories is used for the treatment of prostate cancer and more, which achieved global sales of more than 2.3 billion US$ in 2011.[9] In addition, LantusTM from Sanofi (which is really at the border between a peptide drug and a small biopharmaceutical) reached sales of 7.9 billion US$ in 2013. Currently, there are more than 70 US Food and Drug Administration (FDA)-approved peptide medicines on the market and this is expected to grow significantly, with approximately 200 peptide drugs at present in clinical trials and more than 600 therapeutic peptides in preclinical development[10] **(Figure 1.1)**. In terms of value, the global peptide drug market has been predicted to increase from 14.1 billion US$ in 2011 to an estimated 25.4 billion US$ in 2018, with an underlying increase in novel innovative peptide drugs from 8.6 billion US$ in 2011 (60%) to 17.0 billion US$ (66%) in 2018.[11]

The most recent example of a novel peptide drug class is the group of glucagon-like peptide-1 (GLP-1) agonists for the treatment of type 2 diabetes mellitus, which reached total sales of over 2.6 billion US$ in 2013, with VictozaTM, the most prominent member of the class, reaching blockbuster status.

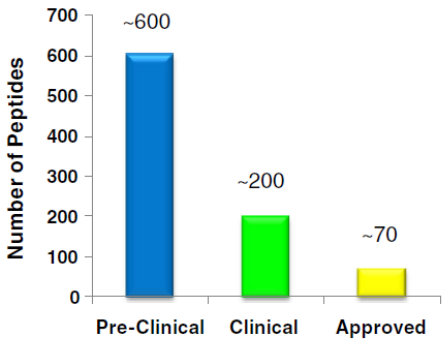


Figure 1.1 - Stage of peptide drugs in discovery and development.[10]

The main disease areas currently driving the therapeutic use of peptide drugs are metabolic disease and oncology. However, increasing attention is directed towards the use of peptide therapeutics in the treatment of diabetes and obesity. That is probably why North America currently represents the largest share of the peptide drug market, with the Asian market expected to have the largest growth. This commercial trend leads to pay attention, consequently, to occurring bioactive peptides mostly derived from plants for their potential pharmaceutical and/or nutraceutical applications.[12] The interest for health-promoting functional food, dietary supplements and pharmaceutical preparations containing bioactive peptides is actually markedly increasing.[13]

Table 1.1 describes commercial products that companies have launched on the market with antihypertensive, antimicrobial and mineral-binding peptides. These industrial-scale applications include from dairy food (Calpis®, Evolus®, Peptide Soup EX), ingredients (Capolac® MM0525, Lacprodan®, Recaldent™, Ameal peptide®) to food supplements (Ameal bp®, PeptACE® Fish Peptide, Vasotensin®). At least, the two milk fermented products, Calpis® and Evolus®, have shown beneficial effects on blood pressure in several rat models and human studies.[14-16]

Table 1.1 Commercial products and ingredients with health or functional claims containing bioactive peptides

Brand name	Type of product	Bioactive peptides	Health/function claims	Manufacturers
Alpha lactalbumin	Whey protein isolate	α-lactalbumin	Helps sleep and memory	Davisco Foods International Inc., USA http://www.daviscofoods.com/specialty/alpha
Ameal bp ®, Ameal peptide ®	Tablets, ingredient	VPP, IPP derived from β-casein and κ-casein	Reduction of blood pressure	Calpis Co., Japan http://www.amealbp.com/
BioZate 1	Hydrolyzed whey protein isolate	β-lactoglobulin fragments	Reduction in blood pressure	Davisco Foods International Inc., USA http://www.daviscofoods.com/specialty/biozate
BioZate 3	Hydrolyzed whey protein isolate	β-lactoglobulin fragments	Peptides to extend shelf life by providing a softer texture over time	Davisco Foods International Inc., USA http://www.daviscofoods.com/specialty/biozate
Calpis/Calpico®	Sour milk	VPP, IPP derived from β-casein and κ-casein	Reduction in blood pressure	Calpis Co., Japan http://www.calpis.net/, http://www.calpico.com/
Capolac ® MM0525	Ingredient	Caseino phosphopeptides	Helps mineral absorption	Arla Foods Ingredients Group P/S, Denmark http://www.arlafoodsingredients.com/products/milkprotein-minerals/milk-minerals/
Evolus ®	Fermented milk drink	VPP, IPP derived from β-casein and κ-casein	Reduction in blood pressure	Valio Ltd., Finland http://ammattilaiset.valio.fi/portal/page/portal/valiocom/R_D/Nutritional_research/evolus18082006094558
HI-NUTE series	Ingredient	Soy protein oligopeptides	Helps prevent obesity and muscle fatigue	Fuji Oil Co. Ltd., Japan http://www.fujioil.co.jp/fujioil_e/product/soy_index3.html

Kotsu Kotsu calcium	Soft Drink	Caseino phosphopeptides	Helps mineral absorption	Asahi Soft Drinks Co. Ltd., Japan https://www.asahiinryo.co.jp/products/tokuho/special_ca/
Lacprodan ® ALPHA-20	Ingredient	α-lactoalbumin (60 %)	Helps reduce peptic ulcers	Arla Foods Ingredients Group P/S, Denmark http://www.arlafoodsingredients.com/news/lacprodan-alpha-20-and-peptic-ulcers/
PeptACE ® fish peptides	Capsules	Bonito-derived peptide (85 %): LKPNM	Hypotensive	Natural Factors, USA http://naturalfactors.com/ca/en/products/detail/2698/peptace
Peptide ACE tablet type	Tablets	Bonito-derived peptide LKPNM	Hypotensive	Nippon Supplement Inc., Japan http://www.nippon-sapuri.com/english/

1.3 Regulation and application of bioactive peptides in food

Since bioactive peptides may be included in foods, several limitations and challenges have been reported for their development, their use and their impact on human nutrition. The main issue concerns the lack of knowledge of mechanism of action on molecular targets. In addition, the high level of complexity of protein hydrolysates generally makes difficult the extensive chemical characterization of their amino acid (AA) sequences even if high performance technologies are used. For these unclear aspects, their regulatory approval in many countries is extremely difficult. To date, in fact, there is no single regulatory legislation. In Europe, Regulation (EC) 1924/2006 covers health claims made for food products and the European Food Safety Authority (EFSA) must review all applications based on restricted scientific proofs. On the contrary, the Food Drug Administration (FDA, USA) and the Ministry of Health, Labour and Welfare (MHLW, Japan) adopt less restrictive regulation. Even if to date, any peptide-food derived ingredients were approved in Europe and the USA, Japanese legislators, already in 1991 enacted a set of "Foods for Specified Health Use" (FOSHU) measures, which regulated the use of food for therapy.[17] Products carrying a FOSHU label are now certified with respect to their pharmaceutical activity;[18] an example is the use of "Fine Rice", rice grains as a treatment of rice-associated atopic dermatitis. The list of category 2 FOSHU products includes soybean-derived peptides which control the level of cholesterol in the blood, while further plant-derived products are represented in categories #4 (blood pressure reduction) and #10 (serum triacylglycerol level and lipid reduction).

The tight regulation required by entities such as EFSA regarding the approval of the use of peptide-based ingredients, highlights the need to investigate either their mechanism of action or their chemical compositions in order to prove scientific evidences on which a dossier should be efficiently based.

1.4 Bioactive Peptides: Definition and Sources

In terms of chemical complexity, peptides fill the niche between typical small molecules and larger proteins. They refers to different short amino acid chains (2–20 AAs, in some cases up to 40 AAs) that, generally, are inactive within the sequence of the parent proteins until protein digestion, fermentation or food processing.[19]

Until last decade, their production was focused almost exclusively on animal protein sources, and in particular milk. The latest researches, on the contrary, opened new possibilities for obtaining peptides from edible plant sources and from alternative and cheap matrices, such as agriculture surplus and wastes, by-product and non-conventional vegetables.

Actually, among the seed of crop plant, soybean is considered a popular source of food proteins, which has been traditionally used in Asia. It offers a complete spectrum of (essential) amino acids. Numerous health claims, at least in part supported by proteomic analyses [20], were made for it.

Cereals (supplying half the world's protein needs) and legumes are both rich sources of proteins with a complementary spectrum of amino acids.[21, 22] In addition, bioactive peptides were found in many other vegetables (algae, edible fungi, garlic, ginkgo biloba seeds, curcuma, sesame, peanut, spinach, sunflower, hempseeds, tubers and others) as the consequence of fermentation or enzymatic hydrolysis.[23] Peas have been analysed for bioactive peptides, most of them using mass spectrometry technologies based on 2-D gels.[24] By using more recent nano-LC based methods further, in part complementary details on the role of bioactive proteins and peptides from these sources can be covered to improve our basic knowledge. Such increasingly growing background information on the composition of bioactive peptides in food may eventually result in the design of efficient strategies to tailor nutritional interventions.

1.5 Beneficial biological functionalities of plant-derived peptides

Among the human health benefits claimed for plant peptides, the reduction in blood cholesterol level, the reduction in blood pressure, antioxidant properties and as well as immunomodulatory activities are reported. In this dissertation only lipid lowering properties are taken into account.

1.5.1 Lipid lowering effects of bioactive protein hydrolysates

An unfavourable profile of blood lipids is an important risk factor for the genesis of various cardiovascular diseases (CVDs). Not surprisingly, treatment for hyperlipidemia-accelerated

diseases often includes the improvement of serum lipid distribution through diet modifications. It is generally known that several dietary proteins can improve blood lipid profiles. To date hypocholesterolemic properties have been reported for soy[25], whey[26] and fish protein[27] capable of altering the plasma profile from atherogenic to cardio protective. Although a number of published studies have already demonstrated potential beneficial effects of pulses on dyslipidemia, more research is needed to achieve a complete picture of their mechanism of action.

The mechanisms of regulation of cholesterol levels are described in detail here. The cellular requirements for cholesterol modulation are mainly satisfied through two mechanisms:

1) Exogenously by cholesterol influx pathways involving cholesterol uptake mediated by various lipoprotein receptors; **2)** endogenously by cholesterol synthesis regulated by the rate-limiting enzyme of this process, HMGCoAR.

At the transcriptional level, cholesterol uptake and biosynthesis are closely regulated through a negative feedback control.

1) The low-density lipoprotein receptor (LDLR) pathway is a negative feedback system that plays decisive roles in the regulation of plasma and intracellular cholesterol homeostasis. LDL Receptors enable the intake of plasma-derived LDL cholesterol and its metabolism. To maintain a cholesterol homeostasis, LDLR expression is tightly regulated by sterol regulatory element-binding protein-2 (SREBP-2) and SREBP cleavage-activating protein (SCAP) in transcriptional level and by proprotein convertase subtilisin/kexin type 9 (PCSK9) in posttranscriptional level. PCSK9 stimulates LDLR degradation, resulting in LDL accumulation in plasma.[28]

2) When there is an accumulation of cholesterol into the cell, the endogenous cholesterol biosynthesis mediated by HMGCoAR is reduced and the number of LDL receptors decreases. Conversely, when there is a depletion of intracellular cholesterol, the activity of HMGCoAR is elevated and there are more numerous LDL receptors on the cellular membrane.

Soybean. Some investigations have shown that 7S globulin (β-conglycinin) component has a major role in the cholesterol-lowering properties of soybean. In particular, a peptic soy protein hydrolysate has been reported to show a stronger serum cholesterol lowering effect than intact soy protein in rats.[29] In details, soy peptides have the potential to modulate cholesterol homeostasis by increasing receptor-mediated LDL uptake in HepG2 cells. A clinical trial [30] has

investigated the expression of LDL-receptor in patients affected by familial hypercholesterolemia, treated with animal proteins or with soy protein with cholesterol added to balance the two diets. The soy protein diet determined a marked LDL-C reduction as well as an 8-fold increase of LDL degradation, whereas there were minimal changes in LDL-C levels or LDL receptor activity after the animal protein diet.

Recently, **LPYPR**, a peptide derived from soy glycinin, was found to endow serum cholesterol-lowering effects in mice following oral administration.[31] **LPYPR** is structurally homologous to enterostatin (**VPDPR**), an endogenous peptide exhibiting hypocholesterolemic effect.[32] Another glycinin-derived peptide, i.e. **IAVPGEVA**, endowed the same activity was identify by Pak.[33]

The regulation of cholesterol homeostasis by peptides has been proposed to be due to the activation of LDL receptors and LDL degradation in liver cells at least *in vitro*. In details, recent paper[34], suggests that **IAVPGEVA**, **IAVPTGVA**, and **LPYP** are able to interfere with the catalytic activity of HMGCoAR modulating the cholesterol metabolism, through the activation of the LDLR-SREBP-2 pathway, increasing the ability of HepG2 cells to uptake the LDL. In addition, these three soy peptides are also able to inhibit the activity of HMGCoAR due to an increase of its phosphorylation level on Ser 872 residue *via* the activation of the AMPK-pathway. In parallel, other evidences have been reported for two other peptides isolated from soybean β-conglycinin, i.e., **YVVNPDNDEN** and **YVVNPDNNEN**, which are known to be also absorbed by human enterocytes.[35] Both of them, similarly to **IAVPGEVA**, **IAVPTGVA**, and **LPYP**, are also able to up-regulate the LDLR protein levels sharing the same mechanism of action.

Lupin. The main storage proteins in lupin are the acidic legumin (α-conglutin) and vicilin (β conglutin), as well as small amounts of γ-conglutin, a basic 7S globulin.[36] To date, only a few studies, all in animal models, have investigated the mechanism of the hypocholesterolemic activity of lupin protein. The reduced plasma TAG levels in hypercholesterolemic rats treated with a white lupin protein isolate was in part due to the down-regulation of the SREBP-1c mRNA in the liver, which led to a reduction of hepatic fatty acid synthase.[37] Another study, however, observed lower hepatic mRNA concentrations of genes involved in fatty acid synthesis, and a parallel up-regulation of genes involved in TAG hydrolysis.[38] A third study showed that lupin-fed rats displayed significantly higher hepatic mRNA levels of SREBP-2, the major transcriptional regulator of intracellular cholesterol levels, and CYP7A1, the rate-limiting enzyme in bile acid biosynthesis.[39]

Along-side *in vivo* studies, recent investigations have been aimed at the elucidation of the mechanism of action of lupin protein hydrolysates on cholesterol reduction. A study on human hepatic HepG2 cells has demonstrated that tryptic and peptic hydrolysates from a total lupin protein extract are able to interfere with the HMGCoAR activity, up-regulating the LDLR and SREBP-2 proteins via the activation of phosphoinositide 3-kinase (PI3K), protein kinase B (Akt), and glycogen synthase kinase-3β (GSK3β) pathways, increasing the LDL-uptake of HepG2 cells.[40] In parallel, in a synergic way, lupin peptides reduce the PCSK9 protein level production and secretion. **Figure 1.2** shows the hypocholesterolemic mechanism of action, mediated by lupin peptides in human hepatic HepG2 cells.

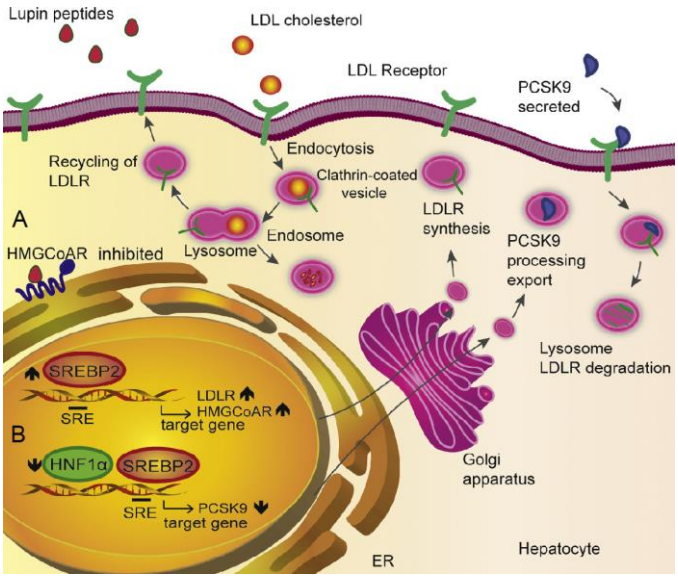


Figure 1.2 Hypocholesterolemic mechanism of action mediated by lupin peptides in human hepatic HepG2 cells. Upon cell penetration, peptic and tryptic lupin peptides act as competitive inhibitors of HMGCoAR, leading to a reduction of intracellular cholesterol synthesis. When intracellular cholesterol level decreases, the transcription factor SREBP-2 is activated and LDLR and HMGCoAR genes are transcripted with subsequent increase of LDLR and HMGCoAR protein levels **(A)**. In parallel, in a synergic way, lupin peptides reduce the PCSK9 protein level production and secretion. In particular, through the reduction of HNF1-alpha protein level, they lead to a decrease of intracellular precursor and mature PCSK9 protein levels. In agreement, the PCSK9 down-regulation is translated in a reduction of HepG2 cell ability to secrete mature PCSK9 in the extracellular medium, with the consequent stabilisation of active LDLR on hepatic cellular membrane **(B)**. For this reason, the distinct modulation of the two pathways leads to hypocholesterolemic effects through an improved and synergic activity of LDLR, which can bind and carry the extracellular LDL in HepG2 cells. [41]

Hempseed. Recently, it has been proved that hempseed peptides endowed important activities, especially at cardiovascular level, as ACE inhibitors (**WYT**, **SVYT**, and **IPAGV**) and as antioxidants (**WVYY** and **PSLPA**).[42] On the contrary, literature reports only a few evidences

on the ability of hempseed proteins to modulate the lipid profile.[43, 44] Some efforts of this thesis have been addressed to fill this gap elucidating the mechanism through which hempseed hydrolysate mediates a cholesterol-lowering effect at HepG2 cells, by molecular and functional investigations on the LDLR-SREBP-2 pathway.

1.6 Bioavailability of plant-derived peptides

Peptides can escape the normal process of digestion and thus can be absorbed intact across the intestinal mucosa. The availability of peptides is of particular importance in the context of their therapeutic efficacy and their physiological effects depend on their ability to reach in an active form their target organs. An experimental assessment of the health benefits of plant-derived peptides is difficult to achieve, since ingested proteins are processed by a variety of proteases and peptidases during their passage through the gastro-intestinal tract. In addition, the activity of these compounds depends on their amino acid composition, bioavailability, interactions with other food components carried in the bloodstream, and on the resistance to the proteases activity. On the other hand, the influence of several different physiological parameters, i.e. blood pressure, insulin and glucose homoeostasis, plasma cholesterol concentrations, and immune function can modulate certain peptide functions. Nevertheless, a body of evidence proves that dietary peptides that survive luminal digestion and endure the brush-border-membrane peptidases hydrolysis can be detected in measurable amounts in the peripheral blood and urine.[45,46] In other words, the processes underlying absorption, distribution, metabolism and excretion (ADME) are important pharmacological parameters to evaluate for the future use of naturally occurring peptide as pharmaceuticals.[47]

1.6.1 Absorption

The ability of peptides to enter the intestinal lymphatic system is affected by their molecular size and structural properties, such as hydrophobicity, which influence their transport route.[48] Generally, the small (di and tri peptides) and large (10–30 amino acids) peptides can cross the intestinal barrier intact and show their activity at the tissue level.[49] However, the potency of bioactive peptides decrease as the chain length increases.[49] Interestingly, small di- and tripeptides are absorbed more rapidly from the small intestine than free amino acids, showing further some stability to proteases.[50]

To date, in order to evaluate the phenomenon of absorption, several cell line models are employed. Most of the current intestinal models are using transformed cell line and among them the Caco-2 being the most widely used.

1.6.1.1 The Caco-2 model

The intestinal absorption model based on Caco-2 cells, originally derived from a human colon adenocarcinoma, is a well-defined *in vitro* system, extensively used to study absorption, metabolism and bioavailability of compounds. According to Biopharmaceutics Classification System (BCS) and FDA approval, Caco-2 cells can be used as a screening method for new drug candidates during drug discovery and development. Despite of being colonic cells, after being cultured, they form a polarized confluent monolayer that shows most of the morphological and functional characteristic of enterocyte-like cells: microvilli on the apical membrane, tight junction between adjacent cells and expression of intestinal enzyme and transporters. In order to mimic *in vivo* conditions, these cells are typically cultured on permeable supports that separate the apical and basolateral compartments, which resemble to the luminal and circulation flux of the intestine, respectively.[51] **Figure 1.3** reports the structural features of Caco-2 monolayer. The proposed Caco-2 absorption model provides information both about permeability in human intestine and the role of metabolism. In addition, measurements derived from permeability experiments with the monolayer show a good correlation with data from human *in vivo* absorption.

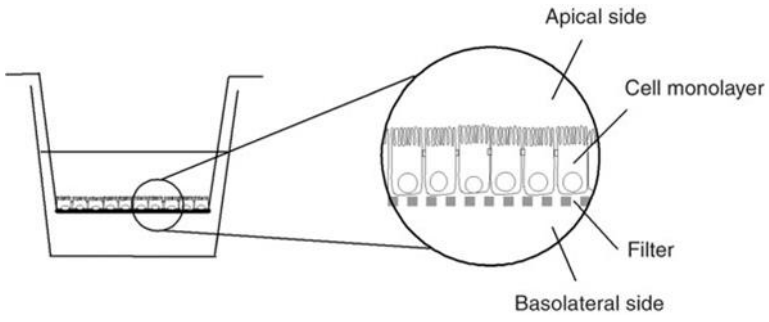


Figure 1.3 - Diagram of a Caco-2 monolayer grown on a permeable filter support.[52]

1.6.2 Metabolism

Peptides are catabolized in free amino acids, used as building blocks for protein synthesis. Once ingested, peptides are subjected to metabolism, commonly due by the action of peptidases that are present in enterocytes brush-border, in plasma, in the endothelium and inside the target

cells. The proteases that characterise the epithelium are classified according to their catalytic site of action (endopeptidases or exopeptidases) and the nature of the catalytic site. Endopeptidases cleave the polypeptide chain at particularly susceptible peptide bonds distributed along the chain, whereas exopeptidases hydrolyse one amino acid (or dipeptide) at a time from either the N-terminus (aminopeptidases) or the C-terminus (carboxypeptidases). The four major classes of endopeptidases are serine proteases, cysteine proteases, aspartic proteases and metallo proteases and are characterised by different pH of actions.

Peptide degradation persists also during their transport through the epithelium, because of the action of hydrolases in cellular cytosol. In several studies, the transport of short-chain peptides across enterocytes occurs at a very low level, indicating that the peptides are largely hydrolyzed by cytosol hydrolase.[53] Small peptides, particularly those carrying a proline-proline dipeptide at their C-terminus, are the least prone to degradation by proteases and peptidases present in the stomach, secreted by pancreas, or contained in the brush border membrane in the small intestine. [54] Whereas the aromatic amino acids such as Try, Phe, and Trp; the positively charged ones (Arg and Lys); and Leu did not undergo any changes. Charge, molecular weight (MW), hydrophobicity and ability to form hydrogen bonds are properties connected with crossing through epithelium, susceptibility to degradative enzymes and consequently bioavailability.

1.6.3 Distribution

Detection of peptides in plasma is technically challenging, because of their low concentrations estimated at picograms per millilitre. Peptides comprised of up to 5–6 residues could be detected in the blood after feeding volunteers with peptide-enriched drinks.[46] However, larger peptides can be found intact in plasma, as in the case of the immunomodulatory lunasin after consumption of soybean.[55] The high bioavailability of lunasin has been also attributed to the simultaneous presence of protease inhibitors, allowing 30% of the peptide to reach the target tissues.[56] The concentration detected in the blood was also proportional to the dose ingested, i.e. **VY** dipeptide was found increased as the ingestion of an enriched drink increased.[57] It has been calculated that the half-life of the dipeptide **VY** was longer than the half-life of the tripeptide **IPP**. To date, the quantification of peptides in plasma has been performed through the application of MS-based techniques. A Multiple Reaction Monitoring (MRM) method was developed to quantify 17 ACE inhibitory peptides (2, 3, and 5 residues) in human plasma collected after ingestion of a peptide-enriched drink.[46] Other authors have also applied a MRM

method to assess the bioavailability of **IPP** and seven other ACE inhibitory peptides contained in a peptide-enriched beverage, and the influence of meal intake on the bioavailability of **IPP**.[58]

CHAPTER 2

2 Technological strategies in the discovery of bioactive peptides

2.1 Peptidomic-based approach for the analysis of hydrolysed proteins

In 2011, the term of peptidomics has been introduced by Schulz[59] to indicate the comprehensive characterization of peptides present in a biological sample. This relatively young discipline has been successfully applied in several areas of research, with particular scientific interest in food nutrition for the discovery and quantification of nutritionally relevant bioactive peptides and their health benefits. A systematic approach to identify biologically and physiologically active peptides, as well as biomarkers, with their pharmacological features is the key of peptidomics. Understanding nature and bioactivity of nutritional peptides means comprehending an important level of environmental regulation of the human genome, because diet is the environmental factor with the most profound life-long influence on health.[60] The peptidomic strategy provides useful information about the presence of bioactive peptides and covers the development of computational tools for: (i) the identification of peptides and their protein precursors, (ii) optimization of hydrolysis for the production of very active peptides, (iii) prediction of protein/peptide structure and bioactive peptide properties, and (iv) quantitative structure-activity relationship (QSAR) modelling.

2.2 Classical downstream processing for discovering of bioactive peptides

Bioactive peptide discovery traditionally consists of a stepwise approach, typically including **(i)** selection of an appropriate protein source of plant origin; **(ii)** proteolysis by enzymatic processing, fermentation or gastrointestinal digestion; **(iii)** initial *in vitro* screening for targeted bioactivity; **(iv)** fractionation of the peptide pool into more defined sub-fractions by chromatography or membrane separation; **(v)** further bioactivity screening followed by identification of the bioactive peptide(s) by mass spectrometry proposing possible structure-function relationships; and **(vi)** production of synthetic peptide to validate bioactivity *in vitro*

and *in vivo*. **Figure 2.1** reports graphically the main steps occurring in the classical discovery processes.

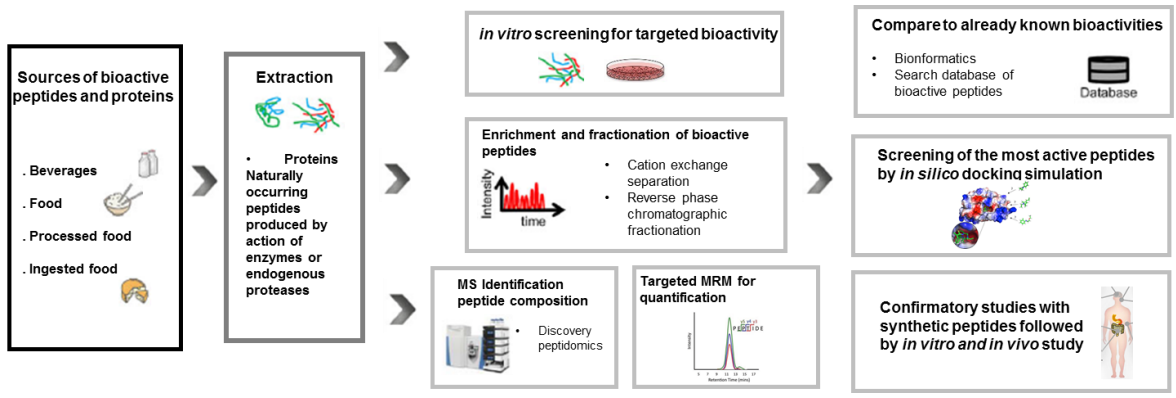


Figure 2.1 - Integrated analytical workflow for peptidomic investigation.
In conventional approaches the peptide composition of nutrients is determined by discovery peptidomics. The release of bioactive sequences is the first step followed by the checking for their biological activities. The large number of peptides contained in the samples, usually, leads their prior separation before identification by mass spectrometry (MS). Each fraction are then tested separately for bioactivity evaluation and further MS characterised.

2.2.1 Release of bioactive peptides by enzymatic hydrolysis

Bioactive peptides can be produced from precursor proteins where they occur as inactive amino acid sequences using the following methods: **(a)** enzymatic hydrolysis by digestive enzymes, **(b)** fermentation of precursor proteins with proteolytic starter cultures, and **(c)** proteolysis by enzymes derived from microorganisms. In many studies, the combination of **(a)** and **(b)** or **(a)** and **(c)** has proven effective in generation of short functional peptides.[61] The choice of proteolytic enzymes or microorganisms used for protein processing has a crucial impact on the composition of the released peptides. However, the most common way to produce bioactive peptides is the use of single or multiple specific or unspecific proteases. Compared to microorganisms, enzymes generally require less time to generate a similar degree of hydrolysis and their reaction can be controlled giving reproducible MW profiles and peptide composition. Additionally, enzymes present substrate specificity which allows the development of protein hydrolysates with well-defined chemical and nutritional characteristics.[62] Critical hydrolysis parameters, such as temperature, pH, aqueous or buffered solution, must be optimized for each protein/substrate and each selected enzyme or combination of them and maintained constant during proteolysis to ensure efficient release of peptides. Moreover, each bioactive peptide

shows different release kinetics during enzymatic hydrolysis. Mostly, larger one appear in early stage of hydrolysis and are then cleaved into smaller peptides showing different bioactivity.[63]

2.2.2 Enrichment and fractionation of bioactive peptides

Protein hydrolysates generally are tested as complex mixtures or as purified fractions. Extensive fractionation procedures using standard chromatographic techniques, such as size-exclusion (SEC), ion-exchange (IEX), affinity or reversed phase HPLC, or the use of membrane ultrafiltration are applied to get simplified peptide pools more amenable to peptide identification. To reduce the complexity of hydrolysates, generally ultrafiltration by membrane is adopted. The use of membrane ultrafiltration produces peptides with specific molecular weight ranges. To obtain short-chain peptides, membranes with low molecular weight (LMW) cut-offs (1-10 kDa) are used for fractionation of the protein hydrolysate to obtain LMW peptides in the permeate and high molecular weight (HMW) peptides (>10 kDa) in the retentate. The LMW peptides have been reported to be more physiologically active as they are readily absorbed into blood circulation and transported to target organs, where they exert their health effects. Some experiments have shown that membrane ultrafiltration and successive chromatographic purifications usually increase potency of the resulting peptides when compared to the crude (unfractionated) protein hydrolysate.[64] In some cases, however, the crude hydrolysate has shown superior bioactivities over the peptide fractions, which could be attributed to synergistic activities of the crude hydrolysate peptides, which are lost upon successive separation.[65,66] The advantage of crude hydrolysates exhibiting stronger physiological activities than purified fractions is that it saves costs, as expensive purification protocols will be no longer necessary in order to obtain potent products. On the other hand, identification of the active peptide structures/sequences for the production of nutraceutical supplements would require successive chromatographic purifications to obtain pure peptide substances.

2.3 The role of MS in biopeptides discovery

Recently, mass spectrometry (MS)-based peptidomic approaches set a new standard in peptide research. The technological development of MS analysers has been fundamental in driving the proteomics research. MS experiments can be addressed directly to measure the molecular mass

of derived peptides (single-stage MS) or used to generate AA-sequence information, from tandem mass spectra (MS/MS or MSn) obtained either by post-source decay (PSD) or, especially in the hybrid arrays, by collision induced dissociation (CID). Very recent technological improvements further enlarge capabilities of tandem mass spectrometry strategies. Electron transfer dissociation (ETD) and electron capture dissociation (ECD) are emerged as new tools in the analytical instrumentation to sequence peptides from complex samples and to characterize their post-translational modifications PTMs.

Hybrid instruments have been designed to combine the capabilities of different mass analysers. The LTQ-Orbitrap system, introduced a few years ago, deserves special mention with respect to comprehensive peptidomics or proteomics analyses. The high resolution (up to 150,000) and mass accuracy, mass-to-charge range up to 6000 and dynamic range greater than 1000 establish the Orbitrap technology as among the most versatile and powerful ones.

Electrospray ionization (ESI) source and matrix-assisted laser desorption/ionization (MALDI) are the most important ionization sources used in peptidomics. Several examples regarding the use of LC-ESI-MS to identify bioactive peptides from vegetables can be found in literature. Sato et al.[67,68] identified, respectively, seven and nine new ACE-inhibitory and anti-hypertensive dipeptides from a wakame hydrolysates and a soy sauce-like seasoning. Hybrids LC-ESI-Q-TOF and LC-ESI-Q-IT were utilized by Puchalska et al.[69] to identify anti-hypertensive peptides after thermolysin digestion of maize and by Soares[70] to identify the major HMGCoAR inhibitory peptides generated by the *in vitro* hydrolysis of *Amaranthus cruentus* protein, respectively.
MALDI ionization is an alternative source used in peptidomic and proteomic investigations, however, the majority of peptidomic studies are conducted using LC-ESI-MS/MS approach. About half of the known bioactive peptides are shorter than 10 amino acids and these peptides fall into the low mass range, where MALDI matrix interference is overwhelming, which severely limits the applicability of MALDI MS in peptidomic studies.[54]

2.3.1 Hypothesis-free discovery and hypothesis-driven targeted approaches in peptidomics

In peptidomics, there are two different strategies aiming at the discovery and/or quantification of potentially bioactive peptides: **hypothesis-free discovery** and **hypothesis-driven targeted** approach, which are complementary.

The first are referred to as shotgun or data-dependent (DDA), widely used for obtaining a comprehensive protein or peptide profile of biological systems in a complex peptide mixture. The second one is referred as targeted MRM, and it used to probe a predefined subset of peptides. **Figure 2.2** shows in details the different principles on which the data acquisition is based.

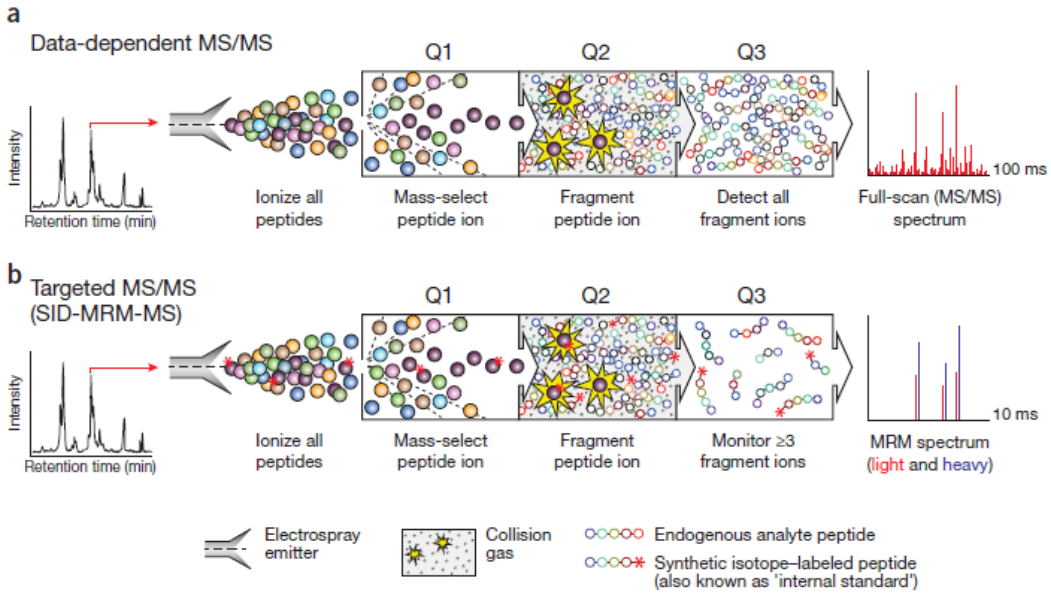


Figure 2.2 - Comparison of conventional data-dependent analysis with targeted MRM-MS. Triple quadrupole mass spectrometer was selected as configuration. **(a)** In a data-dependent MS experiment, digested proteins are loaded on a reversed-phase column attached to a liquid chromatography setup and eluted via electrospray to yield gas-phase ions. At any given point in the chromatographic separation many tens to hundreds of peptides are eluting nearly simultaneously. A full-scan MS spectrum is acquired. The ion observed in the MS spectrum are automatically selected on the basis of their signal intensity (Q1) for fragmentation by collision with inert gas (Q2). The complete array of fragment ions is detected (Q3), which constitutes the full-scan MS/MS spectrum (far right). **(b)** In a MRM-MS analysis, proteotypic peptides are predefined together with their most informative fragment ions. Peptides are selected for fragmentation (Q1 and Q2), and fragment ions are selected for detection (Q3) based on a user-specified list of targeted precursor-fragment pairs (‘transitions’).

Although the DDA approach allows an exploratory data analysis, it suffers of some limitations: its sensitivity is strongly sample dependent and it suffers from inconsistent identification reproducibility across samples due to the biased intrinsic nature of the DDA approach. In addition, the peptide abundance covers very huge orders of magnitude, difficult to investigate only with the shotgun approach. More information are deductible on the most intense peptides deriving from very abundant proteins, but missing values are often observed for peptide deriving from less abundant proteins. The bioactive peptides are characterised by small sequences usually between 2 and 20 amino acids, that can belong, especially the shortest, to

hundreds of proteins. Since they are very small, often they are hidden behind the complex matrix of plant/food. The analysis of food-derived peptide, therefore, requires a different analytical approach, because these entities vary much more in their chemical nature than classical tryptic peptides generated in shotgun proteomics workflow for protein biomarker identification: multiple processing parameters come into play generating not only a large variety in peptide lengths, sequences, and terminal residues, but also a number of peptide modifications. Strategically speaking, the original **hypothesis-free discovery** workflow is being increasingly complemented or followed up by **hypothesis-driven** analysis or candidate-based targeted analysis. A recent review puts the discovery and targeted proteomics approaches into perspective.[71] The **hypothesis-driven targeted** approach (MRM) is able to target several dozen peptides (that need to be known before the measurement) consistently across multiple conditions.[72, 73] By recording the signal of the pre-selected analytes over LC separation, MRM leads to reproducible measurements across a large number of samples with high consistency [74] allowing to increase the sensitivity of the detection, which is an important aspect in the detection of low expressed bioactive peptides.

2.4 Quantitative peptidomics

A better knowledge about the concentration of certain bioactive peptides identified in plant hydrolysates is crucial for future studies of bioavailability as well as to reach more realistic conclusions about the effect that could be expected for an active peptide in the organism.

In fact, the effect and impact of a bioactive peptide in an organic system depend on the amount that is ingested and its ability to reach the bloodstream or the organ of interest. Thus, the quantitation data of identified bioactive peptides is a necessary aim for a better understanding of the effects and mechanisms involved in the action of these compounds.[7]

The most used methodology in the quantitation of known bioactive peptides is currently the MRM, which has been used during decades for the quantitation of therapeutic peptides and biomarker candidates in plasma, serum and urine[75, 76] and recently has been introduced in the area of food science.

2.5 Technical issues in bioactive peptide discovery strategies

The detection of peptides derived from plant proteins requires adaptation of common analytical workflows. Two important parameters, such as peptide length (up to 500 amino acids) and

dynamic range (e.g. twelve orders of magnitude in human blood plasma), are predominant in both fields, proteomics (biomarkers) and peptidomics (bioactives).[71] In both cases the identification and validation is based on the peptide levels and usually it covers different analytical spaces. While proteomics comprises usually peptide molecular weights of approx. 700 to 3000 Da with a dynamic range of twelve orders of magnitude, peptidomics certainly spans a greater peptide length distribution and possibly also a similar dynamic range. The major reason for this greater analytical complexity in nutritional peptidomics is the diverse selectivity and specificity of the food protein processing and digestion level, conferred by multiple *in vitro* and *in vivo* proteases. On the contrary, enzymes with a very selective cut action like trypsin are applied in protein biomarker identification. As a consequence, generated peptides share similar properties, e.g. charge state or length, which reduces the analytical range to be covered. In the case of bioactive peptides, they are released either through processes, such as fermentation or enzymatic hydrolysis,[77] or *in situ* by the host digestive system or gut microbial enzymes or a combination of both.[78] *In vivo*, due to their variable mode of action, endogenous proteases can produce a large variety of peptide lengths, sequences, terminal residues, and chemical modifications.[79]

Whether pre-digested or not, in both cases enzymatic release becomes a very complex and unspecific mechanism involving several enzymes from food processing (e.g. Alcalase, Protamex, Flavorzyme) as well as the digestive system (e.g. pepsin, trypsin, chymotrypsin, elastase)[80] with various activities.

Such difference in the hydrolytic process further impacts analytics, because such peptides have broader physical and chemical properties (e.g. charge state, hydrophobicity, length) and need adapted separation methods and informatics analysis tools as compared to proteomics.

To facilitate the identification of unspecifically digested proteins novel, robust and reliable performing algorithms for bioinformatic analysis of fragment spectra need to be developed. Promising approaches include de novo sequencing[81,82] and protease unspecific database searches.[83] *De novo* peptide sequencing from fragment ion spectra omits the need for creating and matching fragment ion lists generated from protein sequence databases by predefined rules, such as tryptic cleavage, thus enabling the identification of atypically digested peptides. In protease unspecific database search approaches, fragment ion lists for all possible cleavage sites are deduced from protein sequence databases. In this way both peptide termini can be arbitrary positioned in a protein, leading to a much larger search space than by applying only one specific digestion rule, thereby increasing the possibility of false identifications. The other common issue in the peptidomics field regards the ability of MS systems to accurately detect

mostly peptides ≥ 5 amino acid residues in length. In contrast, shorter peptides (2-4 amino acids) have been poorly reported, to date, in the literature. This is possibly due to analytical limitations. The under or over fragmentation of short peptides in tandem MS has made their detection complicated and challenging.[84, 85] These issues are linked to the difficulty to obtain meaningful fragmentation during MS analysis. In addition, short peptides display a low structural diversity, which increases the probability of detecting isobaric fragments within the same sample or samples generated from different protein substrates, making searches against protein databases and the subsequent peptide identification quite complicated. The low level of confidence regarding the identification of these short sequences within dietary hydrolysates is still evident and a MRM atlas is not available as in the case in proteomics for key organism.[86-88] However, the combination of different fractionation approaches, derivatisation methods, front-end analytical separation, *in silico* tools and altered MS settings has proved beneficial in the identification of short peptide sequences within complex mixtures.[89] Definitively, further technology development will be required to provide efficient tools for analysing bioactive peptides of the complex food peptidome.

2.6 Bioinformatics-driven approaches and peptide function prediction

The screening of bioactive peptides from novel substrates using the empirical approach is generally expensive and time-consuming, because it involves the selection of the best proteases that demonstrated the highest potential to liberate the bioactive peptides, confirming later their biological activities through *in vitro* assays. However, this conventional and "empirical" approach has strong limitations; the main drawbacks include the lack of standardised protocols for generation of peptides, the difficulty of systematically checking all hydrolysis parameters (i.e., pH, temperature, time, enzyme to substrate (E/S) ratio, total solids, etc.). These conditions may influence the release of bioactive peptides, as they can modify both substrate conformation and enzyme activity and therefore the accessibility of specific peptide bonds in proteins.[89]

This entire workflow can be simplified by using bioinformatics, which provide a cost-effective strategy through the reduction of the many steps of the traditional workflow. The use of multifactorial design of experiments (DOE) and response surface methodology (RSM) are actually the new approaches used as an optimization tool for biopeptides releasing. [90]

Major contribution of bioinformatics to biopeptides discovery, however, not only concerns the first step of hydrolysis optimization, but is also related to the prediction of biological activity

and the evaluation of food peptides as potential precursor, as observed in the *in silico* analysis of 34 proteins as potential source of DPP-IV inhibitory peptides. [91]

Quantitative structure activity relationship (QSAR) modelling is strongly connected to the prediction of the bioactivity of peptides identified in complex sample and achieves a more target identification of bioactive peptide sequences identified by LC-MS.

Actually, many bioactive peptide databases have been designed to store information such as sequences, MW, activity, reference to published work, IC_{50}, source and more.

All this information, according to the peptidome project, are conveyed together in order to establish the intellectual database for biological and pharmaceutical research by accumulating data of the endogenous peptides in cells, tissue and organism. The most used until now is BioPep containing information on bioactive peptides (n = 2594), major allergens and their epitopes (n = 135) as well as sensory peptides including amino acids (n = 326).[92] **Figure 2.3** represents in a pie chart the major activities described in the literature and available in the BioPep database. ACE inhibitors come first, followed by antibacterial, antioxidative, celiac toxic, inhibitory, opioid, neuropeptides as well as immunomodulating peptides. The hypocholesterolemic activity on the contrary is not taken into consideration.

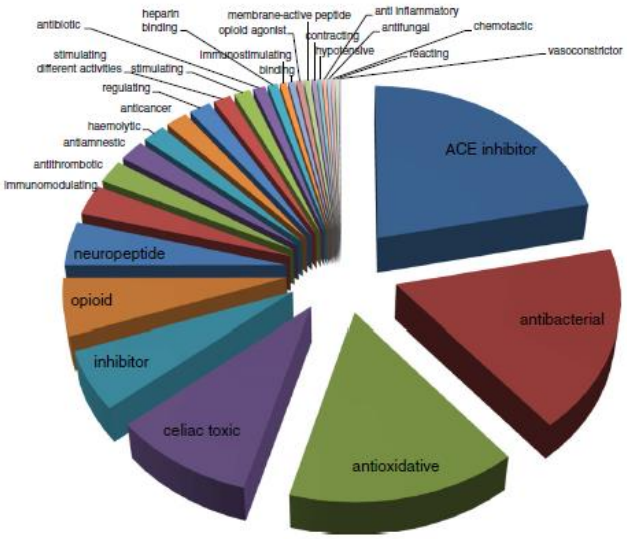


Figure 2.3 - Classification of known peptides based on their reported activity in the BioPep database (n=2594)

References

1. Padhi, A.; Sengupta, M.; Sengupta, S.; Roehm, K. H.; Sonawane, A., Antimicrobial peptides and proteins in mycobacterial therapy: current status and future prospects. *Tuberculosis (Edinb)* **2014,** *94* (4), 363-73.

2. Buchwald, H.; Dorman, R. B.; Rasmus, N. F.; Michalek, V. N.; Landvik, N. M.; Ikramuddin, S., Effects on GLP-1, PYY, and leptin by direct stimulation of terminal ileum and cecum in humans: implications for ileal transposition. *Surg Obes Relat Dis* **2014,** *10* (5), 780-6.

3. Giordano, C.; Marchiò, M.; Timofeeva, E.; Biagini, G., Neuroactive peptides as putative mediators of antiepileptic ketogenic diets. *Front Neurol* **2014,** *5*, 63.

4. Robinson, S. D.; Safavi-Hemami, H.; McIntosh, L. D.; Purcell, A. W.; Norton, R. S.; Papenfuss, A. T., Diversity of conotoxin gene superfamilies in the venomous snail, Conus victoriae. *PLoS One* **2014,** *9* (2), e87648.

5. Hancock, R. E.; Sahl, H. G., Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies. *Nat Biotechnol* **2006,** *24* (12), 1551-7.

6. Li-Chan, E. C. Y., Bioactive peptides and protein hydrolysates: research trends and challenges for application as nutraceuticals and functional food ingredients. *Curr. Opin. Food Sci* **2015,** *1*, 28-37.

7. Mora, L.; Gallego, M.; Reig, M.; Toldrá, F., Challenges in the quantitation of naturally generated bioactive peptides in processed meats. *Trends Food Sci. Technol.* **2017**.

8. Lemes, A. C.; Sala, L.; Ores, J. a. C.; Braga, A. R.; Egea, M. B.; Fernandes, K. F., A Review of the Latest Advances in Encrypted Bioactive Peptides from Protein-Rich Waste. *Int J Mol Sci* **2016,** *17* (6).

9. Kaspar, A. A.; Reichert, J. M., Future directions for peptide therapeutics development. *Drug Discov Today* **2013,** *18* (17-18), 807-17.

10. Sun, L., Peptide-Based Drug Development. *Modern Chemistry and Applications* **2013**.

11. Transparency; Research, M. *Peptide Therapeutics Market: Global Industry Analysis, Size, Share, Growth, Trends and Forecast*; 2012.

12. Pang, G.; Xie, J.; Chen, Q.; Hu, S., How functional foods play critical roles in human health. Food

Sci. Hum. Wellness, 2012; Vol. 1, pp 26-60.

13. Coda, R.; Rizzello, C. G.; Pinto, D.; Gobbetti, M., Selected lactic acid bacteria synthesize antioxidant peptides during sourdough fermentation of cereal flours. *Appl Environ Microbiol* **2012,** *78* (4), 1087-96.

14. Hata, Y.; Yamamoto, M.; Ohni, M.; Nakajima, K.; Nakamura, Y.; Takano, T., Placebo-controlled study of the effect of sour milk on blood pressure in hypertensive subjects. *Am. J. Clin. Nutr.* **1996,** *64* (5), 767-771.

15. Mizushima, S.; Ohshige, K.; Watanabe, J.; Kimura, M.; Kadowaki, T.; Nakamura, Y.; Tochikubo, O.; Ueshima, H., Randomized controlled trial of sour milk on blood pressure in borderline hypertensive men. *Am. J. Hypertens.* **2004,** *17* (8), 701-706.

16. Seppo, L.; Jauhiainen, T.; Poussa, T.; Korpela, R., A fermented milk high in bioactive peptides has a blood pressure-lowering effect in hypertensive subjects. *Am. J. Clin. Nutr.* **2003,** *77* (2), 326-330.

17. Hartmann, R.; Meisel, H., Food-derived peptides with biological activity: from research to food applications. *Curr Opin Biotechnol* **2007,** *18* (2), 163-9.

18. Arai, S.; Yasuoka, A.; Abe, K., Functional food science and food for specified health use policy in Japan: state of the art. *Curr Opin Lipidol* **2008,** *19* (1), 69-73.

19. Kussmann, M.; Van Bladeren, P. J., The Extended Nutrigenomics - Understanding the Interplay between the Genomes of Food, Gut Microbes, and Human Host. *Front Genet* **2011,** *2*, 21.

20. Erickson, B. E., Proteomics data back up soy health claims. *J Proteome Res* **2005,** *4* (2), 219.

21. García, M. C.; Puchalska, P.; Esteve, C.; Marina, M. L., Vegetable foods: a cheap source of proteins and peptides with antihypertensive, antioxidant, and other less occurrence bioactivities. *Talanta* **2013,** *106*, 328-49.

22. Malaguti, M.; Dinelli, G.; Leoncini, E.; Bregola, V.; Bosi, S.; Cicero, A. F.; Hrelia, S., Bioactive peptides in cereals and legumes: agronomical, biochemical and clinical aspects. *Int J Mol Sci* **2014,** *15* (11), 21120-35.

23. Rizzello, C. G.; Tagliazucchi, D.; Babini, E.; Rutella, G. S.; Saa, D. L. T.; Gianotti, A., Bioactive peptides from vegetable food matrices: Research trends and novel biotechnologies for synthesis and recovery. *J. Funct Foods* **2016,** *27*, 549-569.

24. Bourgeois, M.; Jacquin, F.; Savois, V.; Sommerer, N.; Labas, V.; Henry, C.; Burstin, J., Dissecting the proteome of pea mature seeds reveals the phenotypic plasticity of seed protein composition. *Proteomics* **2009,** *9* (2), 254-71.

25. Hori, G.; Wang, M. F.; Chan, Y. C.; Komatsu, T.; Wong, Y.; Chen, T. H.; Yamamoto, K.; Nagaoka, S.; Yamamoto, S., Soy protein hydrolyzate with bound phospholipids reduces serum cholesterol levels in hypercholesterolemic adult male volunteers. *Biosci Biotechnol Biochem* **2001,** *65* (1), 72-8.

26. Nagaoka, S.; Kanamaru, Y.; Kuzuya, Y.; Kojima, T.; Kuwata, T., Comparative-studies on the serum-cholesterol lowering action of whey-protein and soybean protein in rats. *Biosci Biotechnol Biochem* **1992,** *56* (9), 1484-1485.

27. Wergedahl, H.; Liaset, B.; Gudbrandsen, O. A.; Lied, E.; Espe, M.; Muna, Z.; Mørk, S.; Berge, R. K., Fish protein hydrolysate reduces plasma total cholesterol, increases the proportion of HDL cholesterol, and lowers acyl-CoA:cholesterol acyltransferase activity in liver of Zucker rats. *J Nutr* **2004,** *134* (6), 1320-7.

28. Zhang, Y.; Ma, K. L.; Ruan, X. Z.; Liu, B. C., Dysregulation of the Low-Density Lipoprotein Receptor Pathway Is Involved in Lipid Disorder-Mediated Organ Injury. *Int J Biol Sci* **2016,** *12* (5), 569-579.

29. Sugano, M.; Goto, S.; Yamada, Y.; Yoshida, K.; Hashimoto, Y.; Matsuo, T.; Kimoto, M., Cholesterol-lowering activity of various undigested fractions of soybean protein in rats. *J Nutr* **1990,** *120* (9), 977-85.

30. Lovati, M. R.; Manzoni, C.; Canavesi, A.; Sirtori, M.; Vaccarino, V.; Marchi, M.; Gaddi, G.; Sirtori, C. R., Soybean Protein-Diet Increases Low-Density-Lipoprotein Receptor Activity In Mononuclear-Cells From Hypercholesterolemic Patients. *J. Clin. Invest.* **1987,** *80* (5), 1498-1502.

31. Yoshikawa, M.; Fujita, H.; Matoba, N.; Takenaka, Y.; Yamamoto, T.; Yamauchi, R.; Tsuruki, H.; Takahata, K., Bioactive peptides derived from food proteins preventing lifestyle-related diseases. *Biofactors* **2000,** *12* (1-4), 143-6.

32. Takenaka, Y.; Utsumi, S.; Yoshikawa, M., Introduction of enterostatin (VPDPR) and a related sequence into soybean proglycinin A1aB1b subunit by site-directed mutagenesis. *Biosci Biotechnol Biochem* **2000,** *64* (12), 2731-3.

33. Pak, V. V.; Koo, M. S.; Kasymova, T. D.; Kwon, D. Y., Isolation and identification of peptides from soy 11S-globulin with hypocholesterolemic activity. *Chem. Nat. Compd.* **2005,** *41* (6), 710-714.

34. Lammi, C.; Zanoni, C.; Arnoldi, A., IAVPGEVA, IAVPTGVA, and LPYP, three peptides from soy glycinin, modulate cholesterol metabolism in HepG2 cells through the activation of the LDLR-SREBP-2 pathway. *J Funct Foods* **2015,** *14*, 469-478.

35. Amigo-Benavent, M.; Clemente, A.; Caira, S.; Stiuso, P.; Ferranti, P.; del Castillo, M. D., Use of phytochemomics to evaluate the bioavailability and bioactivity of antioxidant peptides of soybean β-conglycinin. *Electrophoresis* **2014,** *35* (11), 1582-9.

36. Duranti, M.; Lovati, M. R.; Dani, V.; Barbiroli, A.; Scarafoni, A.; Castiglioni, S.; Ponzone, C.; Morazzoni, P., The alpha ' subunit from soybean 7S globulin lowers plasma lipids and upregulates liver beta-VLDL receptors in rat feed a hypercholesterolemic diet. *J. Nutr* **2004,** *134* (6), 1334-1339.

37. Spielmann, J.; Shukla, A.; Brandsch, C.; Hirche, F.; Stangl, G. I.; Eder, K., Dietary lupin protein lowers triglyceride concentrations in liver and plasma in rats by reducing hepatic gene expression of sterol regulatory element-binding protein-1c. *Ann Nutr Metab* **2007,** *51* (4), 387-92.

38. Bettzieche, A.; Brandsch, C.; Weisse, K.; Hirche, F.; Eder, K.; Stangl, G. I., Lupin protein influences the expression of hepatic genes involved in fatty acid synthesis and triacylglycerol hydrolysis of adult rats. *Br J Nutr* **2008,** *99* (5), 952-62.

39. Parolini, C.; Rigamonti, E.; Marchesi, M.; Busnelli, M.; Cinquanta, P.; Manzini, S.; Sirtori, C. R.; Chiesa, G., Cholesterol-lowering effect of dietary Lupinus angustifolius proteins in adult rats through regulation of genes involved in cholesterol homeostasis. *Food Chem* **2012,** *132* (3), 1475-1479.

40. Lammi, C.; Zanoni, C.; Scigliuolo, G. M.; D'Amato, A.; Arnoldi, A., Lupin peptides lower low-density lipoprotein (LDL) cholesterol through an up-regulation of the LDL receptor/sterol regulatory element binding protein 2 (SREBP-2) pathway at HepG2 cell line. *J Agric Food Chem* **2014,** *62* (29), 7151-9.

41. Lammi, C.; Zanoni, C.; Calabresi, L.; Arnoldi, A., Lupin protein exerts cholesterol-lowering effects targeting PCSK9: From clinical evidences to elucidation of the in vitro molecular mechanism using HepG2 cells. *J Funct Foods* **2016,** *23*, 230-240.

42. Girgih, A. T.; He, R.; Malomo, S.; Offengenden, M.; Wu, J. P.; Aluko, R. E., Structural and functional characterization of hemp seed (Cannabis sativa L.) protein-derived antioxidant and antihypertensive peptides. *J Funct Foods* **2014,** *6*, 384-394.

43. Gavel, N. T.; Edel, A. L.; Bassett, C. M.; Weber, A. M.; Merchant, M.; Rodriguez-Leyva, D.; Pierce, G. N., The effect of dietary hempseed on atherogenesis and contractile function in aortae from hypercholesterolemic rabbits. *Acta Physiol Hung* **2011,** *98* (3), 273-83.

44. Prociuk, M. A.; Edel, A. L.; Richard, M. N.; Gavel, N. T.; Ander, B. P.; Dupasquier, C. M.; Pierce, G. N., Cholesterol-induced stimulation of platelet aggregation is prevented by a hempseed-enriched diet. *Can J Physiol Pharmacol* **2008,** *86* (4), 153-9.

45. Grimble, G. K., The significance of peptides in clinical nutrition. *Ann Review Nutrition* **1994,** *14*, 419-447.

46. van Platerink, C. J.; Janssen, H. G. M.; Horsten, R.; Haverkamp, J., Quantification of ACE inhibiting peptides in human plasma using high performance liquid chromatography-mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2006,** *830* (1), 151-157.

47. Ménard, S.; Cerf-Bensussan, N.; Heyman, M., Multiple facets of intestinal permeability and epithelial handling of dietary antigens. *Mucosal Immunol* **2010,** *3* (3), 247-59.

48. Shimizu, M.; Tsunogai, M.; Arai, S., Transepithelial transport of oligopeptides in the human intestinal cell, Caco-2. *Peptides* **1997,** *18* (5), 681-7.

49. Roberts, P. R.; Burney, J. D.; Black, K. W.; Zaloga, G. P., Effect of chain length on absorption of biologically active peptides from the gastrointestinal tract. *Digestion* **1999,** *60* (4), 332-7.

50. Webb, K. E., Intestinal absorption of protein hydrolysis products: a review. *J Anim Sci* **1990,** *68* (9), 3011-22.

51. Sambuy, Y.; De Angelis, I.; Ranaldi, G.; Scarino, M. L.; Stammati, A.; Zucco, F., The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biol Toxicol* **2005,** *21* (1), 1-26.

52. Hubatsch, I.; Ragnarsson, E. G. E.; Artursson, P., Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. *Nature Protocols* **2007,** *2* (9), 2111-2119.

53. Wang, B.; Li, B., Effect of molecular weight on the transepithelial transport and peptidase degradation of casein-derived peptides by using Caco-2 cell model. *Food Chem* **2017,** *218*, 1-8.

54. Panchaud, A.; Affolter, M.; Kussmann, M., Mass spectrometry for nutritional peptidomics: How to analyze food bioactives and their health effects. *J Proteomics* **2012,** *75* (12), 3546-59.

55. Dia, V. P.; Torres, S.; De Lumen, B. O.; Erdman, J. W.; De Mejia, E. G., Presence of lunasin in plasma of men after soy protein consumption. *J Agric Food Chem* **2009,** *57* (4), 1260-6.

56. Hsieh, C. C.; Hernández-Ledesma, B.; Jeong, H. J.; Park, J. H.; de Lumen, B. O., Complementary roles in cancer prevention: protease inhibitor makes the cancer preventive peptide lunasin bioavailable. *PLoS One* **2010,** *5* (1), e8890.

57. Matsui, T.; Tamaya, K.; Seki, E.; Osajima, K.; Matsumoto, K.; Kawasaki, T., Val-Tyr as a natural antihypertensive dipeptide can be absorbed into the human circulatory blood system. *Clin Exp Pharmacol Physiol* **2002,** *29* (3), 204-8.

58. Foltz, M.; van Buren, L.; Klaffke, W.; Duchateau, G. S., Modeling of the relationship between dipeptide structure and dipeptide stability, permeability, and ACE inhibitory activity. *J Food Sci* **2009,** *74* (7), H243-51.

59. Schulz-Knappe, P.; Zucht, H. D.; Heine, G.; Jürgens, M.; Hess, R.; Schrader, M., Peptidomics: the comprehensive analysis of peptides in complex biological mixtures. *Comb Chem High Throughput Screen* **2001,** *4* (2), 207-17.

60. Kussmann, M.; Fay, L. B., Nutrigenomics and personalized nutrition: science and concept. *Personalized Medicine* **2008,** *5* (5), 447-455.

61. Korhonen, H.; Pihlanto, A., Food-derived bioactive peptides - Opportunities for designing future foods. *Current Pharmaceutical Design* **2003,** *9* (16), 1297-1308.

62. Castro, H. C.; Abreu, P. A.; Geraldo, R. B.; Martins, R. C.; dos Santos, R.; Loureiro, N. I.; Cabral, L. M.; Rodrigues, C. R., Looking at the proteases from a simple perspective. *J Mol Recognit* **2011,** *24* (2), 165-81.

63. Zhao, Q.; Sannier, F.; Piot, J. M., Kinetics of appearance of four hemorphins from bovine hemoglobin peptic hydrolysates by HPLC coupled with photodiode array detection. *Biochim Biophys Acta* **1996,** *1295* (1), 73-80.

64. Udenigwe, C. C.; Aluko, R. E., Food protein-derived bioactive peptides: production, processing, and potential health benefits. *J Food Sci* **2012,** *77* (1), R11-24.

65. Girgih, A. T.; Udenigwe, C. C.; Aluko, R. E., In Vitro Antioxidant Properties of Hemp Seed (Cannabis sativa L.) Protein Hydrolysate Fractions. *J Am Oil Chem Soc* **2011,** *88* (3), 381-389.

66. Girgih, A. T.; Udenigwe, C. C.; Li, H.; Adebiyi, A. P.; Aluko, R. E., Kinetics of Enzyme Inhibition and Antihypertensive Effects of Hemp Seed (Cannabis sativa L.) Protein Hydrolysates. *J Am Oil Chem Soc* **2011,** *88* (11), 1767-1774.

67. Sato, M.; Hosokawa, T.; Yamaguchi, T.; Nakano, T.; Muramoto, K.; Kahara, T.; Funayama, K.; Kobayashi, A., Angiotensin I-converting enzyme inhibitory peptides derived from wakame (Undaria pinnatifida) and their antihypertensive effect in spontaneously hypertensive rats. *J Agric Food Chem* **2002,** *50* (21), 6245-6252.

68. Nakahara, T.; Sano, A.; Yamaguchi, H.; Sugimoto, K.; Chikata, H.; Kinoshita, E.; Uchida, R., Antihypertensive Effect of Peptide-Enriched Soy Sauce-Like Seasoning and Identification of Its Angiotensin I-Converting Enzyme Inhibitory Substances (vol 58, pg 821, 2010). *J Agric Food Chem* **2010,** *58* (9), 5858-5858.

69. Puchalska, P.; Marina, M. L.; García, M. C., Development of a reversed-phase high-performance liquid chromatography analytical methodology for the determination of antihypertensive peptides in maize crops. *J Chromatogr A* **2012,** *1234*, 64-71.

70. Soares, R. A. M.; Mendonca, S.; de Castro, L. I. A.; Menezes, A.; Areas, J. A. G., Major Peptides from Amaranth (Amaranthus cruentus) Protein Inhibit HMG-CoA Reductase Activity. *Int J Mol Sci.* **2015,** *16* (2), 4150-4160.

71. Domon, B.; Aebersold, R., Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* **2010,** *28* (7), 710-21.

72. Olsen, J. V.; Ong, S. E.; Mann, M., Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **2004,** *3* (6), 608-14.

73. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* **2008,** *4*, 222.

74. Picotti, P.; Clément-Ziza, M.; Lam, H.; Campbell, D. S.; Schmidt, A.; Deutsch, E. W.; Röst, H.; Sun, Z.; Rinner, O.; Reiter, L.; Shen, Q.; Michaelson, J. J.; Frei, A.; Alberti, S.; Kusebauch, U.; Wollscheid, B.; Moritz, R. L.; Beyer, A.; Aebersold, R., A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **2013,** *494* (7436), 266-70.

75. Liu, Y.; Hüttenhain, R.; Surinova, S.; Gillet, L. C.; Mouritsen, J.; Brunner, R.; Navarro, P.; Aebersold, R., Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. *Proteomics* **2013,** *13* (8), 1247-56.

76. Rauh, M., LC-MS/MS for protein and peptide quantification in clinical chemistry. *J Chromatogr B Analyt Technol Biomed Life Sci* **2012,** *883-884*, 59-67.

77. Nakamura, Y.; Yamamoto, N.; Sakai, K.; Okubo, A.; Yamazaki, S.; Takano, T., Purification And Characterization Of Angiotensin I-Converting Enzyme-Inhibitors From Sour Milk. *J Dairy Sci* **1995,** *78* (4), 777-783.

78. De Leo, F.; Panarese, S.; Gallerani, R.; Ceci, L. R., Angiotensin converting enzyme (ACE) inhibitory peptides: production and implementation of functional food. *Curr Pharm Des* **2009,** *15* (31), 3622-43.

79. Kussmann, M.; Affolter, M.; Nagy, K.; Holst, B.; Fay, L. B., Mass spectrometry in nutrition: understanding dietary health effects at the molecular level. *Mass Spectrom Rev* **2007,** *26* (6), 727-50.

80. Vlieghe, P.; Lisowski, V.; Martinez, J.; Khrestchatisky, M., Synthetic therapeutic peptides: science and market. *Drug Discov Today* **2010,** *15* (1-2), 40-56.

81. Frank, A.; Pevzner, P., PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal Chem* **2005,** *77* (4), 964-73.

82. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G. A.; Ma, B., PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* **2012,** *11* (4), M111.010587.

83. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M., Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **2011,** *10* (4), 1794-805.

84. Dallas, D. C.; Guerrero, A.; Parker, E. A.; Robinson, R. C.; Gan, J.; German, J. B.; Barile, D.; Lebrilla, C. B., Current peptidomics: applications, purification, identification, quantification, and functional analysis. *Proteomics* **2015,** *15* (5-6), 1026-38.

85. Lahrichi, S. L.; Affolter, M.; Zolezzi, I. S.; Panchaud, A., Food peptidomics: large scale analysis of small bioactive peptides--a pilot study. *J Proteomics* **2013,** *88*, 83-91.

86. Deutsch, E. W.; Lam, H.; Aebersold, R., PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep* **2008,** *9* (5), 429-34.

87. Picotti, P.; Lam, H.; Campbell, D.; Deutsch, E. W.; Mirzaei, H.; Ranish, J.; Domon, B.; Aebersold, R., A database of mass spectrometric assays for the yeast proteome. *Nat Methods* **2008,** *5* (11), 913-4.

88. Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R., High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* **2010,** *7* (1), 43-6.

89. Nongonierma, A. B.; FitzGerald, R. J., Strategies for the discovery, identification and validation of milk protein-derived bioactive peptides. *Trends Food Sci Technol* **2016,** *50*, 26-43.

90. Quirós, A.; Hernández-Ledesma, B.; Ramos, M.; Martín-Álvarez, P. J.; Recio, I., Short communication: Production of antihypertensive peptide HLPLP by enzymatic hydrolysis: optimization by response surface methodology. *J Dairy Sci* **2012,** *95* (8), 4280-5.

91. Lacroix, I. M. E.; Li-Chan, E. C. Y., Evaluation of the potential of dietary proteins as precursors of dipeptidyl peptidase (DPP)-IV inhibitors by an in silico approach. *J Funct Foods* **2012,** *4* (2), 403-422.

92. Iwaniak, A.; Dziuba, J.; Niklewicz, M., The biopep database - A tool for the in silico method of classification of food proteins as the source of peptides with antihypertensive activity. *Acta Alimentaria* **2005,** *34* (4), 417-425.

Part II.

Scientific contributions

CHAPTER 3

MANUSCRIPT 1

A MULTIDISCIPLINARY INVESTIGATION ON THE BIOAVAILABILITY AND ACTIVITY OF PEPTIDES FROM LUPIN PROTEIN

Carmen Lammi [a], Gilda Aiello [a], Giulio Vistoli [a], Chiara Zanoni [b], Anna Arnoldi [a,*], Yula Sambuy [c], Simonetta Ferruzza [c], Giulia Ranaldi [c]

a) Department of Pharmaceutical Sciences, University of Milan, Milan, Italy

b) Dipartimento Cardiotoracovascolare, A.O. Ospedale Niguarda Cà Granda, Milan, Italy

c) CREA, Food and Nutrition Research Center, Rome, Italy

Abbreviations: **ACE**, angiotensin converting enzyme; **AP**, apical; **APBS**, adaptive Poisson-Boltzmann solver; **BL**, basolateral; **DPP-IV**, dipeptidyl peptidase IV; **HMGCoAR**, 3-hydroxy-3-methylglutaryl coenzyme A reductase; **LDL**, low density lipoprotein; **MD**, Molecular dynamic; **MLPInS**, molecular lipophilic potential interaction score; **PEP**, prolyl endopeptidase; **Plp95**, pairwise linear potential; **SREBP-2**, regulatory element binding proteins 2; **TEER**, trans-epithelial electrical resistance.

Keywords: bioactive peptides, Caco-2 cells, HMGCoA reductase, intestinal absorption, *Lupinus albus*, plant protein.

3.0 Abstract

Lupin foods provide useful health benefits. Since tryptic and peptic peptides from lupin protein modulate cholesterol metabolism in HepG2 cells and increase LDL-uptake, this work had the goal of assessing whether these lupin peptides are absorbed by human intestinal Caco-2 cells. Cells were differentiated for 15 days and transport experiments were performed by incubating each lupin peptide mixture from the apical side. After 4 h, basolateral solutions were collected and analysed by HPLC-Chip-MS/MS. Eleven tryptic and eight peptic peptides, deriving from lupin storage proteins, were identified in the basolateral samples. An *in vitro* assay showed that basolateral peptides maintain their capacity to inhibit 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCoAR) activity and *in silico* docking simulations permitted to hypothesise which peptides may bind more efficiently to the HMGCoAR catalytic site. This is the first investigation providing evidence that lupin peptides with specific structures are potentially absorbed in human intestinal cells.

3.1 Introduction

The absorption of nutrients from the small intestinal lumen to blood circulation is mainly performed by enterocytes. An *in vitro* strategy to address the bioavailability problem consists of evaluating the uptake and transport of target compound(s) by human intestinal Caco-2 cells. When these cells are seeded in culture on polycarbonate filters, they undergo spontaneous differentiation leading to the formation of a polarised cell monolayer, coupled by tight junctions and expressing several morphological and functional features of small intestinal enterocytes (Sambuy et al. 2005). In this two-compartment system, the cell monolayer separates the apical (AP) compartment, corresponding to the intestinal lumen, from the basolateral (BL) one, corresponding to the intestinal vascular and lymphatic circulation (Ferruzza et al. 2012, Sambuy et al. 2005).

There is now much interest for bioactive food peptides that may positively influence major body functions, once they are released by digestion from parental proteins and absorbed (Cam and de Mejia 2012). The transport through intestinal cells is a major factor influencing their bioavailability. The hypothesis that small peptides may escape complete digestion and be

transported from intestinal lumen into blood circulation is gaining acceptance, mainly due to numerous studies describing the *in vitro* trans-epithelial transport of bioactive peptides (Miguel et al. 2008, Zhu et al. 2008, Foltz et al. 2007, Regazzo et al. 2010), although its occurrence *in vivo* is still controversial (Miner-Williams, Stevens and Moughan 2014). Differentiated human intestinal Caco-2 cells express carrier-mediated transport systems for amino acids and di- and tri-peptides (Thwaites et al. 1993, Miguel et al. 2008), show transcytosis activity (Heyman et al. 1990), and develop tight junctions involved in the paracellular passive route (Hidalgo, Raub and Borchardt 1989, Hashimoto and Shimizu 1993). In particular, this model has been used for evaluating the bioavailability of antioxidant peptides from soybean β-conglycinin (Amigo-Benavent et al. 2014).

Lupin protein provides various health benefits (Arnoldi et al. 2015a, Duranti et al. 2008), particularly in the area of cholesterol reduction (Sirtori et al. 2012, Bähr et al. 2015), and hypertension (Arnoldi et al. 2015a) or hyperglycaemia prevention (Duranti et al. 2008). We have recently demonstrated that tryptic and peptic peptides from lupin protein are able to interfere with 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCoAR) activity, up-regulating LDL-receptor and sterol regulatory element binding proteins 2 (SREBP-2), and increasing LDL-uptake in HepG2 cells (Lammi et al. 2014b). Moreover, the same peptides inhibit the activity of angiotensin converting enzyme (ACE, EC 3.4.15.1) (Boschin et al. 2014b, Boschin et al. 2014a), an effect possibly involved in the hypotensive effect observed *in vivo*.

From these findings the general question arose whether lupin peptides might be absorbed after food ingestion (Miner-Williams et al. 2014). As a first step to solve this issue, the present work was aimed at assessing whether any tryptic and peptic peptides from lupin protein may be absorbed by Caco-2 cells and transferred to the BL compartment and whether they maintain their hypocholesterolemic activity. The former objective was achieved by conducting transport experiments across differentiated human enterocytes and analysing BL solutions by HPLC-Chip-MS/MS, whereas the latter by using *in vitro* biological assays and bioinformatics tools.

3.2 Materials and methods

3.2.1 Chemicals and reagents

Formic acid, acetonitrile (ACN), sequence grade trypsin, and pepsin were from Sigma–Aldrich (Milan, Italy). LC-grade H_2O (18 MΩ cm) was prepared with a Milli-Q H_2O purification system (Millipore, Bedford, MA, US). Centrifugal filter devices YM-30 (cut off 30 kDa) were

from Amicon Bioseparations (Millipore Corporation, Bedford, USA). Dulbeccos's Modified Eagle Medium (DMEM) was from GIBCO (Thermo Fisher Scientific, Waltham, MA US). Foetal Bovine Serum (FBS) was from Hyclone Laboratories (Logan, UT, US). Stable L-glutamine, 1% non-essential amino acids, and penicillin/streptomycin were from Euroclone (Milan, Italy). HMGCoAR assay Kit, red phenol, and PBS was from Sigma-Aldrich (St. Louis, MO, USA). Polycarbonate filters, 12 mm diameter, 0.4 μm pore diameter were from Transwell Corning Inc. (Lowell, MA, US).

3.2.2 Preparation of the pepsin and trypsin peptide mixtures

The pepsin and trypsin hydrolysates from lupin protein are the same used in a previous paper investigating cholesterol metabolism modulation in HepG2 cells (Lammi et al. 2014b). The procedure for protein extraction from seeds of *Lupinus albus* (cultivar Ares) and the conditions of protein hydrolysis are reported in detail there. In brief, proteins were extracted from defatted lupin flour with 100 mM Tris-HCl/0.5 M NaCl buffer, pH 8.2 for 2 h at 4 °C. After centrifugation at 6,500 *g*, for 20 min at 4 °C, the supernatant was dialysed against 100 mM Tris-HCl buffer pH 8.2 for 24 h at 4 °C. After assessing the protein concentration by Bradford assay, the total protein extract was dissolved in Tris-HCl buffer 100 mM at pH 8, then the pH was adjusted to the optimal hydrolysis conditions for each enzyme (pH = 2 for pepsin and = 8 for trypsin) by adding 1 M NaOH or 1 M HCl. After 18 h incubation and enzyme inactivation, the mixtures were ultra-filtered through 3,000 Da cut-off centrifuge filters (Amicon Ultra-0.5, Millipore, Billerica, MA, USA) at 12,000 *g* for 30 min at 4 °C. The peptide concentration in the permeates was measured according to a literature method (Kito and Ito 2008), based on chelating the peptide bonds by Cu (II) in alkaline media and monitoring the change of absorbance at 330 nm. In brief, the reagent contained 0.6 M sodium citrate, 0.9 M sodium carbonate, and 0.07 M copper sulfate, 2.4 M NaOH at pH 10.6. A solution containing X μL peptide mixture, (500 − X) μL water, 500 μL 6% (w/w) NaOH in water, and 50 μL active reagent was prepared. The reaction was carefully mixed, incubated for 15 min at 20 °C, and then the optical density of the solution was measured at 330 nm. A sterile solution of peptone from casein at 10 mg/mL in water was used as standard for the calibration curve; the assay was linear in the range 100−1000 μg of peptides in cuvette.

3.2.3 Cell culture and differentiation

Caco-2 cells, obtained from INSERM (Paris), were sub-cultured at low density (Natoli et al. 2011). Cells were routinely sub-cultured at 50% density and maintained at 37 °C in a 90% / 10% air / CO_2 atmosphere in DMEM containing 25 mM glucose, 3.7 g/L $NaHCO_3$, 4 mM stable L-glutamine, 1% non-essential amino acids, 100 U/L penicillin, 100 µg/L streptomycin (complete medium), supplemented with 10% heat inactivated foetal bovine serum (FBS) (Hyclone Laboratories, Logan, UT, US). For differentiation, cells were seeded on polycarbonate filters, 12 mm diameter, 0.4 µm pore diameter (Transwell, Corning Inc., Lowell, MA, US) at a $3.5x10^5$ cells/cm^2 density in complete medium supplemented with 10% FBS in both AP and BL compartments for 2 days to allow the formation of a confluent cell monolayer. Starting from day 3 after seeding, cells were transferred to complete medium in both compartments, supplemented with 10% FBS only in BL compartment, and allowed to differentiate for 15 days with regular medium changes three times weekly (Ferruzza et al. 2012).

3.2.4 Trans-epithelial transport of peptic and tryptic lupin peptides

Prior to experiments, the cell monolayer integrity and differentiation were checked by TEER measurement as described in details in the paragraph 2.5. Cells were then washed twice and peptides transport was assayed. In particular, transport experiments were performed in transport buffer solution (137 mM NaCl, 5.36 mM KCl, 1.26 mM $CaCl_2$, 1.1 mM $MgCl_2$, 5.5 mM glucose). In order to reproduce the pH conditions existing *in vivo* in small intestinal mucosa, AP solutions were maintained at pH 6.0 (buffered with 10 mM morpholinoethane sulfonic acid), and BL solutions were maintained at pH 7.4 (buffered with 10 mM N-2-hydroxyethylpiperazine-N-4-butanesulfonic acid). Prior to transport experiments, cells were washed twice with 500 µL PBS with Ca^{++} and Mg^{++}. Peptides transport was assayed by loading the AP compartment with peptic or tryptic lupin peptides (1 mg/mL) in AP transport solution (500 µL) and the BL compartment with the BL transport solution (700 µL). After 4 h incubation, AP and BL media were collected and stored at -80 °C. Five independent transport experiments were performed, each in duplicate.

3.2.5 Cell monolayer integrity evaluation

The trans-epithelial electrical resistance (TEER) was measured at 37 °C using the voltmeter apparatus Millicell (Millipore Co., USA,), immediately before and at the end of the 4 h transport experiments. After incubations, no significant changes in TEER values were observed. At the end of transport experiments, cells were incubated from the AP side with 1 mM phenol-red in PBS with Ca^{++} and Mg^{++} for 1 h at 37 °C. The BL solution was then collected and NaOH (70 µL, 0.1 N) was added before reading the absorbance at 560 nm by a microplate reader Synergy H1 from Biotek (Winooski, VT, USA). Phenol-red passage was quantified using a standard phenol-red curve.

3.2.6 HPLC-Chip-MS/MS analysis

The original peptic and tryptic lupin mixtures given to cells in transport medium and the medium collected at the end of the experiment from AP and BL chambers (500 µL and 700 µL, respectively) were freeze-dried and residues were solubilised in HPLC water (100 µL). Samples were purified by ultrafiltration, using membranes with a 3-kDa molecular weight cut-off (MWCO) (Millipore, USA). Filtered peptide mixtures were analysed on a SL IT mass spectrometer interfaced with a HPLC-Chip Cube source (Agilent Technologies, Palo Alto, CA, USA). The chromatographic chip incorporated a 40 nL enrichment column (Zorbax 300SB-C18, 5 µm pore size), and a 43 mm x 75 µm analytical column packed (Zorbax 300SB-C18, 5 µm pore size). The samples (5 µL), previously acidified with formic acid, were loaded onto the enrichment column at a flow rate 4 µL/min using isocratic 100% C solvent phase (99% water, 1% ACN and 0.1% formic acid). After clean-up, the chip valve was switched to separation and the peptides were eluted into the mass spectrometer at the constant flow rate of 300 nL/min. The LC solvent A was 95% water, 5% ACN, 0.1% formic acid; solvent B was 5% water, 95% ACN, 0.1% formic acid. The nano-pump gradient program was as follows: 5% solvent B (0 min), 50 % solvent B (0-50 min), 95% solvent B (50-60 min) and back to 5% in 10 min. The drying gas temperature was 300 °C, flow rate 3 L/min (nitrogen). Data acquisition occurred in positive ionization mode. Capillary voltage was -1950 V, with endplate offset -500V. Mass spectra were acquired in the mass range from *m/z* 300-2200, with target mass 700 *m/z,* average of 2 spectra, ICC target 30000, and maximum accumulation time 150 ms. LC/MS analysis was performed in data dependent acquisition Auto $MS^{(n)}$ mode.

3.2.7 Mass spectral data elaboration and database searching

Mass spectrometric data were processed using a Spectrum Mill MS Proteomics Workbench (Rev B.04.00, Agilent). The extraction of MS/MS spectra was conducted accepting a minimum sequence length of 3 amino acids and merging scans with same precursor within a mass window of ±0.4 *m/z* in a time frame of ± 5 s. Cysteine carbamidomethylation and methionine oxidation were set as fixed and variable modifications, respectively; trypsin or pepsin were chosen as digestive enzymes; 2 missed cleavage were allowed. MS/MS search was conducted against the subset of *Lupinus* protein sequences (8669 *entries*) downloaded from UNIProtKB (http://www.uniprot.org/). The mass tolerance of parent and fragments of MS/MS data search was set at 2.5 Da for precursor ions and 0.7 for fragment ions respectively. Threshold used for peptide identification score ≥8; Scored Peak Intensity SPI% ≥70%; Local False Discovery Rate ≤ 0.1%. MS/MS spectra were also reprocessed to assign candidate peptides in NCBI database using Mascot search program (http://www.matrixscience.com), as reported in Table S1 and S2 (Supporting Information).

3.2.8 Determination of the P-abs and T-abs peptidic concentration

The peptide concentration in the 100 µL of absorbed BL solution was measured according to a literature method (Levashov, Sutherland, Besenbacher, & Shipovskov, 2009), which is based on chelating the peptide bonds by Cu (II) in alkaline media and monitoring the change of absorbance at 330 nm, following an experimental method (Levashov et al., 2009). Briefly, a solution of X µL peptide mixture, (100 − X) µL water, 95 µL 6% (w/w) NaOH in water, and 9.5 µL of active reagent (containing 0.6 M sodium citrate, 0.9 M sodium carbonate, and 0.07 M copper sulphate, 2.4 M NaOH, pH 10.6) was prepared. The reaction mixture was mixed, incubated for 15 min at 20 °C and then the absorbance was measured at 330 nm using the Synergy H1 fluorescent plate reader (Biotek, Bad Friedrichshall, Germany). As standard for the calibration curve, a sterile solution of peptone from casein at 10 mg/mL in water was used.

3.2.9 HMGCoAR activity assay

BL effluents of Caco-2 cells after 4 h transport were tested for their ability to inhibit HMGCoAR activity. The recovered BL samples (700 µL) were lyophilised and then suspended in 100 µL water to obtain the BL peptic and tryptic solutions at the concentration of 0.25 and

0.32 µg/µL, respectively. The HMGCoAR activity was evaluated using a commercial Kit by Sigma-Aldrich (Milan, Italy), providing HMGCoAR, NADPH, buffer, and substrate solution. The experiments were carried out at 37 °C. Each reaction (200 µL) was prepared adding in the following order: assay buffer (154 µL), BL samples from treatments with peptic or tryptic peptides or from control cells (C) (28 µL), NADPH (4 µL), substrate solution (12 µL), and HMGCoAR (catalytic domain) (2 µL). The samples were mixed and the absorbance at 340 nm read by a microplate reader Synergy H1 from Biotek (Winooski, VT, USA) at time 0 and after 10 min. The HMGCCoA-dependent oxidation of NADPH and sample inhibitory properties were measured by absorbance reduction. Three independent experiments were performed in triplicate. Statistical analyses were carried out by one-way ANOVA (Graphpad Prism 6) followed by Dunnett's test and or *t* test. Values were expressed as means ± SEM; P-values < 0.05 were considered to be significant.

3.2.10 *In silico* docking simulations of BL peptides to the HMGCoAR catalytic site

The selection and refinement of HMGCoAR structure was carried out as already described elsewhere (Lammi et al. 2015b). Since HMGCoAR inhibitors are characterised by a net negative charge, docking simulations were focused on negatively charged BL peptides only. These peptides were modelled by using FUGUE, an on line resource able to generate reliable structures also for relatively short peptides, whose conformations would be, however, hardy predictable by simple conformational analyses (Shi, Blundell and Mizuguchi 2001). Regardless of the utilised template, all peptides were modelled in a folded hairpin-like conformation, which is reasonably suited to be accommodated within the HMGCoAR catalytic pocket. Docking simulations were performed by using the software PLANTS (Protein-Ligand ANT System), which calculates reliable ligand poses by ant colony optimization algorithms (Korb, Stützle and Exner 2009). The search was focused on a 15 Å radius sphere around the optimised pose of the reference soy peptide YVVNPDNDEN, a known HMGCoAR inhibitor (Pak et al. 2008), as computed in a previous study (Lammi et al. 2015b). An energy constrain of -5.0 kcal/mol was imposed to guide ligands in assuming a pose similar to YVVNPDNDEN. For each ligand, 10 poses were computed and scored by using Plp95 function with a search speed equal to 1. The best complexes were minimized by keeping fixed all atoms outside a 15 Å radius sphere around the bound ligand and optimised complexes were finally used to calculate docking scores.

3.2.11 Potential biological activities of available peptides

In order to forecast the potential biological activities of identified peptides, their sequences were submitted to a search using the open access informatic resource BIOPEP (http://www.uwm.edu.pl/biochemia/index.php/pl/biopep).

3.3 Results

3.3.1 Trans-epithelial transport of lupin peptides

Differentiated Caco-2 cells were incubated with peptic and tryptic lupin peptides added to AP compartments at a 1 mg/mL concentration. After 4 h treatment, AP and BL media were collected and submitted to HPLC-Chip-MS/MS analysis. For monitoring cell monolayer permeability and excluding non-specific peptide passage, TEER measurements were taken at both experiment beginning and end. Moreover, phenol-red passage across the monolayer was assayed at the experiment end (Ferruzza et al. 2003). Both assays demonstrated that the incubation with lupin peptides did not affect monolayer permeability (data not shown). Only filters showing TEER values and phenol red passage similar to untreated control cells were considered for peptide transport analysis.

3.3.2 Peptide sequencing by LC-MS/MS

The starting peptic and tryptic peptide mixtures and AP and BL samples taken at the end of transport experiments were analysed by HPLC-Chip-MS/MS. **Figure 3.1** shows exemplary chromatographic profiles of BL peptic and tryptic peptides. The peptide identification was carried out through MS/MS ion search, using the SpectrumMill search engine. The results from peptide sequencing were then subjected to a further evaluation using Mascot search program, consulting the Uniprot_viridiplantae database. Most peptides were sequenced, but no protein hit was assigned. Due to very incomplete lupin protein sequencing, only some peptides were assigned to known lupin proteins.

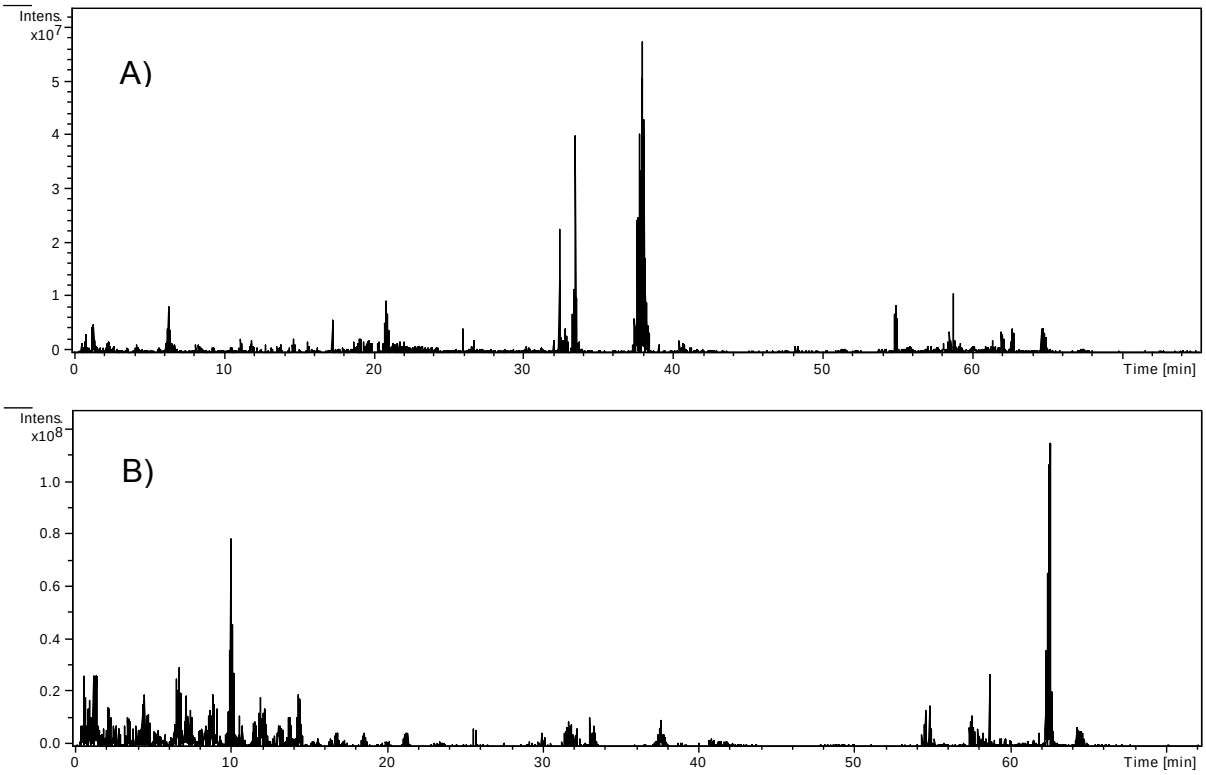


Figure 3.1 - MS/MS total ion current of lupin BL samples:
A) tryptic peptides, B) peptic peptides

3.3.3 Characterisation of bioavailable lupin peptides

Table 3.1 shows a consolidated list of lupin tryptic and peptic peptides detected and identified in starting peptide mixtures and AP and/or BL samples. Among the sixteen tryptic peptides identified in the starting mixture, eleven derive from 7S globulins, i.e. β-Conglutin (Q6EBC1) or Vicilin-like protein (Q53HY0); four peptides derive from 11S globulins, i.e. α-Conglutin (F5B8V8) and Legumin-like protein (Q53I54); and one peptide derives from δ-Conglutin (Q333K7), a main lupin albumin. Among the twelve peptic peptides identified, nine derive from 7S globulins, i.e. β-Conglutin (Q6EBC1) and Vicilin-like protein (Q53HY0), and three from 11S globulins, i.e. α-Conglutin (F5B8V7). All detected peptides have lengths between 7-23 amino acid residues, mass weights in the range 930 - 2720 Da, and contain predominantly acidic and neutral residues.

Out of the sixteen peptides detected in the starting tryptic peptide mixture, nine were detected also in AP samples, whereas eleven were detected also in BL media. Ten peptides (**T1 - T6**, **T9**, **T11, T13**, and **T16**) were detected in all situations (starting peptide mixture, AP and BL

samples). Five peptides (**T7**, **T8**, **T12**, **T14**, and **T15**) were detected only in starting peptide mixtures, perhaps because they were hydrolysed by Caco-2 cells and became undetectable due to the low molecular weights of the resulting fragments. In fact, short peptides are often singly charged, which impairs an efficient fragmentation during MS/MS sequence analysis. One peptide (**T10**) was detected in BL samples, but not in AP ones after the transport experiments. The outcomes of peptic peptide treatments are similar: six peptides were identified in all situations (**P1**, **P2**, **P6 - P8**, and **P12**), four disappeared both in AP and BL samples (**P4**, and **P9 - P11**), and two were detected only in BL samples (**P3** and **P5**).

Table 3.1 LC-ESI-MS/MS based identification of tryptic and peptic peptides from transport experiments

Protein name [a]	Accession no [a]	m/z [b] (charge)	Observed [b] Mr (Da)	Expected [b] Mr (Da)	Start-end [b]	Peptide sequence [b]	Short name	M [c]	AP [c]	BL [c]
Trypsin derived peptides										
β-Conglutin Precursor	Q6EBC1	622.59 (3)	1863.76	1863.85	275-289	R.IILGNEDEQEYEEQR.R	T1	+[d]	+	+
		874.82 (2)	1746.92	1746.84	480-495	R.AVNELTFPGSAEDIER.L	T2	+	+	+
		576.18 (3)	1726.57	1726.00	154-168	R.IVEFQSKPNTLILPK.H	T3	+	+	+
		523.65 (3)	1567.62	1567.83	371-384	K.INEGALLLPHYNSK.A	T4	+	+	+
		520.23 (3)	1556.73	1556.80	169-182	K.HSDADYVLVVLNGR.A	T5	+	+	+
		525.45 (2)	1050.52	1050.50	336-344	R.SNEPIYSNK.Y	T6	+	+	+
β 6-Conglutin	F5B8W4	395.44 (2)	789.58	789.41	391-397	K.SGPFNLR.S	T7	+	nd[e]	nd
Vicilin-like protein	Q53HY0	727.13 (3)	2718.38	2718.97	496-518	R.LIKNQQQSYFANALPQQQQQSEK.E	T8	+	nd	nd
		528.37 (3)	1580.76	1580.75	291-304	R.GQEQSHQDEGVIVR.V	T9	+	+	+
		489.93 (3)	1493.77	1493.66	451-463	R.LLGFGINAYENQR.N	T10	+	nd	+
		732.05 (2)	1462.29	1462.69	483-495	K.ELTFPGSAEDIER.L	T11	+	+	+
α-Conglutin	F5B8V8	715.68 (3)	2145.30	2146.04	46-64	R.IESEGGVTETWNSNKPELR.C	T12	+	nd	nd
Legumin-like protein	Q53I54	843.44 (3)	2526.06	2526.22	41-63	R.LNALEPDNTVQSEAGTIETWNPK.N	T13	+	+	+
		643.62 (3)	1929.75	1929.82	268-282	R.HGREEEEEEEEEDER.R	T14	+	nd	nd
		638.46 (3)	1912.87	1912.96	340-357	R.HNIGESTSPDAYNPQAGR.F	T15	+	nd	nd
δ-Conglutin	Q333K7	934.47 (2)	1866.92	1866.93	117-131	R.QEEQLLEQELENLPR.T	T16	+	+	+
Pepsin derived peptides										
β-Conglutin Precursor	Q6EBC1	844.48 (2)	1686.94	1686.82	349-362	F.YEITPDRNPQVQDL.D	P1	+	+	+
		787.91 (2)	1573.75	1573.65	349-361	F.YEITPDRNPQVQD.L	P2	+	+	+
		783.53 (2)	1564.68	1564.60	239-251	F.YDFYPSSTKDQQS.Y	P3	+	nd	+

		609.83 (2)	1218.65	1217.41	228-238	K.LAIPINNPGYF.Y	P4	+	nd	nd
		555.26 (2)	1107.60	1107.24	164-173	T.LILPKHSDAD.Y	P5	+	nd	+
		532.80 (2)	1063.57	1063.56	204-213	A.LRIPAGSTSY.I	P6	+	+	+
		469.84 (2)	935.43	935.98	484-492	E.LTFPGSAED.I	P7	+	+	+
Vicilin-like protein	Q53HY0	901.98 (2)	1801.86	1801.89	349-363	F.YEITPDRNPQVQDLD.I	P8	+	+	+
		651.86 (2)	1301.72	1301.53	377-387	L.LLPHYNSKAIF.I	P9	+	nd	nd
α 2-Conglutin	F5B8V7	564.34 (2)	1226.68	1226.58	153-163	I.IAIPPGIPYWT.Y	P10	+	nd	nd
		1052.60 (1)	1052.63	1052.62	152-161	D.IIAIPPGIPY.W	P11	+	nd	nd
		1042.52 (1)	1042.56	1042.35	155-163	A.IPPGIPYWT.Y	P12	+	+	+

a) According to "UniProtKB" (http://www.uniprot.org/)
b) The identification of protein parent was performed using SpectrumMill Workbench. The MS/MS data were processed using a mass tolerance of 2.5 Da and 0.7 Da for the precursor and fragment ions, respectively.
c) **M**, starting peptide mixture of tryptic or peptic peptides; **AP**, apical chamber samples; **BL**, basolateral chamber samples.
d) **+**, detected.
e) **nd**, not detected.

3.3.4 Absorbed peptic and tryptic lupin peptides are able to inhibit HMGCoAR activity

Since we have recently shown that tryptic and peptic peptides from lupin protein inhibit the activity of HMGCoAR (Lammi et al. 2014b), in order to verify whether the absorbed peptides are still able to inhibit this enzyme, the BL samples inhibitory capacity was evaluated using the purified HMGCoAR catalytic domain. **Figure 3.2** shows that both peptic and tryptic BL peptides inhibited the HMGCoAR activity in a significant way: BL peptic peptides by 33% at 0.9 μg/μL, whereas tryptic peptides by 20% at 1.1 μg/μL.

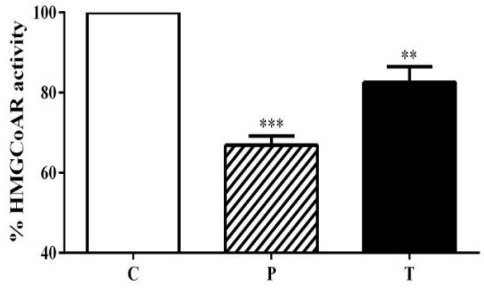


Figure 3.2 - Effect of absorbed peptides on the catalytic domain of HMGCoAR. Bars indicate the effects of peptic and tryptic peptides contained in BL solutions on HMGCoAR activity. The data points represent the averages ± SEM of three independent experiments in triplicate. ** P <0.001 and *** P < 0.0001 *vs* C. P = Peptic peptides, T= Tryptic peptides, and C = BL medium of control samples.

3.3.5 *In silico* evaluation of the capacity of BL peptides to interact with the HMGCoAR catalytic site

Table 3.2 lists the docking scores as computed for minimised complexes of BL peptides within the HMGCoAR catalytic site. It reports pairwise linear potential Plp95 scores, used by the docking program, as well as adaptive Poisson-Boltzmann solver (APBS) scores, and molecular lipophilic potential interaction scores (MLPInS), focused on ionic and apolar interactions, respectively (Vistoli et al. 2010, Baker et al. 2001). Moreover, it reports score values normalised per peptide length, since docking scores tend unavoidably to increase with ligand size, and it includes also score values for the reference peptide YVVNPDNDEN. **Table 3.2** includes also average score values for tryptic and peptic peptides.

The wide ranges of these scores suggest that the putative capacities of simulated peptides to bind the HMGCoAR catalytic pocket greatly differ. Although on average these peptides show

worse docking scores compared to the reference peptide (especially concerning normalized values), some tryptic and peptic peptides have noteworthy score values. In detail, **P5** and **P7** show better normalised Plp95 score values than the reference peptide, while **T1** displays the best APBS score value which is, however, slightly worse than the reference peptide when normalised per peptide length. MLP scores show clear agreements with APBS scores, suggesting that the best complexes are stabilised by a proper balance of both hydrophobic and ionic interactions, the former mostly elicited by ligand apolar residues within N-terminal portion, the latter mainly centred on C-terminus (see below).

Table 3.2. Docking scores for the simulated negatively charged BL peptides as well as the score values normalised per the number of residues and the corresponding score values for the reference peptide YVVNPDNDEN. The corresponding averages for tryptic and peptic peptides are also reported (Plp95 is expressed in kcal/mol, APBS in kJ/mol, MLP is dimensionless**)**

Peptide	residues	Plp95 Score	APBS Score	MLP Score	Plp95 norm	APBS norm	MLP norm
reference	10	-320.55	-125.10	-67.49	-32.06	-12.51	-6.75
T1	15	-366.70	-164.88	-79.54	-24.45	-10.99	-5.30
T2	16	-377.97	-98.86	-72.68	-23.62	-6.18	-4.54
T5	14	-358.65	-54.75	-44.03	-25.62	-3.91	-3.14
T9	14	-369.90	-75.95	-58.24	-26.42	-5.42	-4.16
T11	13	-338.37	-51.13	-34.98	-26.03	-3.93	-2.69
T13	15	-413.98	-92.62	-51.29	-27.60	-6.17	-3.42
T16	15	-327.18	-69.57	-47.72	-21.81	-4.64	-3.18
P1	13	-359.27	-51.87	-49.04	-27.64	-3.99	-3.77
P5	10	-329.43	-109.74	-73.94	-32.94	-10.97	-7.39
P7	9	-338.27	-101.47	-60.96	-37.59	-11.27	-6.77
P8	15	-438.14	-43.79	-41.64	-29.21	-2.92	-2.78
Mean Ts	13.88	-359.16	-91.61	-57.00	-26.40	-6.89	-4.24
Mean Ps	11.75	-366.28	-76.71	-56.40	-31.84	-7.29	-5.18

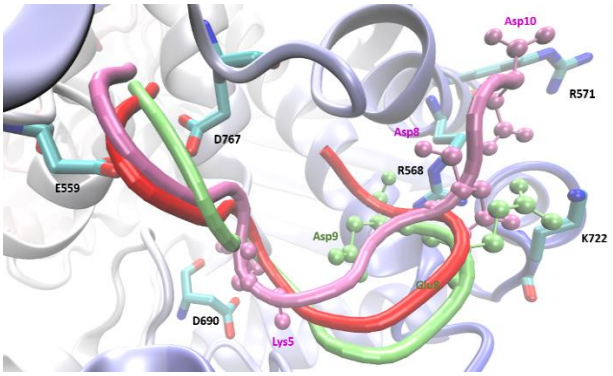


Figure 3.3 - Comparison of computed poses for reference peptide YVVNPDNDEN (tube in red) with P5 (tube in mauve) and P7 (tube in lime). For clarity, hydrogen atoms and apolar residues are not shown.

Figure 3.3 shows that the poses of **P5** (in mauve) and **P7** (in lime) are very similar to that of the reference peptide YVVNPDNDEN (in red), a result easily explainable considering their comparable length and similar sequences characterised by apolar side chains in N-terminus and ionic residues in C-terminus. In detail, the three complexes appear to be stabilised by a set of common interactions, which can be subdivided into three groups. A) The positively charged N-terminus elicits ion-pairs with Glu559 and Asp767, reinforced by H-bonds with surrounding Thr557 and Thr558. B) Negatively charged residues located at C-terminal tail (including the C-terminus itself) are engaged in ionic contacts with Arg568, Arg571, and Lys722. C) Hydrophobic residues located at N-terminus are involved in hydrophobic interactions with apolar residues (e.g., Leu76, Ile536, Leu562, Met655, and Met657, not shown for clarity). Additionally, **P5** includes a central positively charged residue (Lys5), which stabilises a salt bridge with Asp690. Overall, the interaction patterns stabilised by **P5** and **P7**, almost superimposable to that of the reference peptide, suggest that these peptic peptides might have a comparable HMGCoAR inhibitory activity.

Albeit longer, peptide **T1** assumes a pose rather similar to those of peptic ligands due to a very folded conformation stabilised by an intramolecular salt bridge between Glu6 and Arg15 further reinforced by bridging contacts elicited by Asn5. As schematised by **Figure 3.4**, **T1** is able to stabilise an impressive set of ionic contacts, which can explain the best APBS value computed for **T1**. **Figure 3.4** shows that every charged **T1** residue is engaged in ion-pairs further stabilised by a rich set of H-bonds elicited by remaining polar residues. Moreover, apolar residues located at N-terminus are involved in already mentioned hydrophobic contacts.

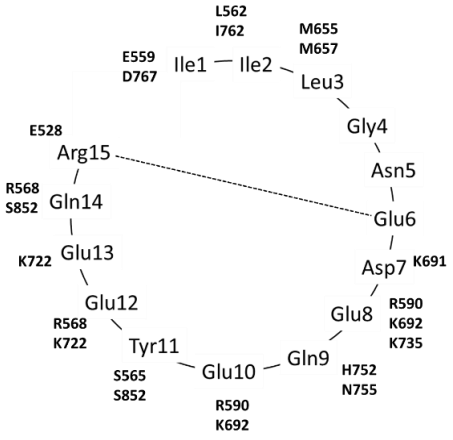


Figure 3.4 - Schematic 2D representation of main contacts stabilizing the putative complex for T1. The dashed line defines the key intramolecular ion-pair stabilizing the folded conformation of the bound peptide.

3.4 Discussions

3.4.1 Absorbed peptides features

The application of the multidisciplinary approach, including cell cultures, proteomics and bioinformatics tools, has allowed achieving a very important goal, since it has demonstrated that eleven tryptic peptides and eight peptic peptides are bioavailable in this *in vitro* model of the small intestine. Considering the overall characteristics of peptides identified in BL samples, they are very heterogeneous being constituted by 9 - 23 amino acid residues, within a 935-2526 Da mass range. Apparently, they do not have any common characteristics, neither length nor charge. They contain predominantly acidic and neutral residues: eleven have a calculated pI < pH 4.5, five have a neutral pI, however, three have a pI > pH 7 **(Table 3.1S**, Supporting information). Acidic peptides are negatively charged, whereas neutral and basic ones have 0 and 1 net charges, respectively; their net charges range from -5 to 1. Longer peptides, in general, appear to be absorbed less than shorter ones, but there are exceptions to this rule, since peptide **T13**, containing 23 amino acid residues, is absorbed, while **T7**, the smallest tryptic peptide, is not absorbed.

The mechanisms of absorption across intestinal epithelium may vary from passive diffusion of hydrophobic molecules across membranes to active transport, either by membrane transporters or vesicular-mediated transcytosis, and finally to permeation across tight junction complexes. Nonspecific permeation between adjacent cells, governed by tight junctions, can be excluded, since phenol red, a paracellular passage marker, did not cross the Caco-2 barrier during transport experiments. The transport of these peptides to the BL compartment takes place with a certain degree of selectivity. A few peptides (**T10, P3,** and **P5**), transported to BL compartments, were no longer detectable in AP media: this may possibly depend on their rapid transport to BL compartment and efficient parallel degradation in the AP medium by brush border peptidases. Peptide **P10** was not absorbed, possibly because it was transformed in **P12**, by removal of N-terminal isoleucine and alanine residues catalysed by dipeptidyl-peptidase IV (DPP-IV; EC 3.4.14.5), a highly specialised aminopeptidase removing dipeptides only from peptides with proline or alanine in the N-terminal penultimate position (Augustyns et al. 2005). **P12**, in turn, was absorbed.

Globulins, accounting for about 90% of total lupin protein (Duranti et al. 2008), are mainly constituted by 7S and 11S globulins. The peptides identified in BL samples derive mostly from

7S globulins, whereas only a few come from 11S globulins, and just one peptide originates from δ-conglutin, a main lupin albumin. Apparently, detected peptides do not reflect the relative abundance of these proteins in lupin seed, since 7S and 11S globulins are normally present in comparable amounts (Duranti et al. 2008), but instead their relative stability. In fact, 7S globulins are much more easily hydrolysed by proteases and more sensible to thermal treatments than 11S globulins (Sirtori et al. 2010) .

3.4.2 Role of absorbed peptides in cholesterol metabolism

Recently, we have demonstrated that tryptic and peptic peptides from lupin protein produce hypocholesterolemic effects through inhibiting HMGCoAR activity in human hepatic HepG2 cells (Lammi et al. 2014b). The consequent reduction of intracellular cholesterol levels leads to activation of SREBP-2, which in turn increases the expression of LDL receptor (LDLR): the LDLR activity increase leads to an improved ability of HepG2 cells to uptake extracellular LDL with a final hypocholesterolemic effect.

A main achievement of the present work is having provided evidence that the peptides transported by Caco-2 cells in the BL compartment maintain their initial capacity of functioning as inhibitors of the HMGCoAR activity. In fact, testing *in vitro* their ability to directly inhibit the activity of HMGCoAR, we have demonstrated that the absorption process did not apparently impair their inhibition capacity. Surprisingly, the activity of BL peptic peptides seemed even improved versus the starting samples (Lammi et al. 2014b), possibly suggesting a selective absorption of the most bioactive components.

The docking simulations of BL peptides with HMGCoAR may possibly explain this finding. In fact, although the different peptide lengths impair a precise comparison, all normalised average values **(Table 3.2)** suggest that peptic peptides should generate more stable complexes with HMGCoAR compared to tryptic ones. This difference can be explained by the positively charged residue in C-terminus of all tryptic peptides, which is unfit within a subpocket surrounded by several positively charged residues and hampers a satisfactory pose for whole peptides. By contrast, most peptic peptides end with an aspartate residue, which is clearly suitable for the positively charged subpocket of HMGCoAR. Finally, the docking results revealed that tryptic and peptic peptides are influenced by different factors. For peptic peptides, the most interesting complexes are afforded by the shortest peptides, which conveniently mimic the reference peptide, while the best complexes for tryptic peptides are generated by those

ligands with a high negative charge, which should counteract the potential repulsive contacts exerted by the unfit C-terminus.

3.4.3 Potential bioactivity according to bioinformatic prediction

BIOPEP (www.uwm.edu.pl/biochemia) (Minkiewicz et al. 2012) is an open access database that permits to hypothesise potential biological activities of peptides depending on the presence of some short amino acid sequences. Unfortunately, it does not take in consideration the inhibition of HGMCoAR activity. The BL peptide sequences were searched in BIOPEP, obtaining the results shown in **Table 3.1S** (Supporting information): the software proposed one or more activities for each peptide, mostly related to cardiovascular disease or diabetes prevention.

All detected peptides are potentially ACE-inhibitors. The active domains LTF, ILP, and LLP occur within peptides **T2**, **T3,** and **T4**, and **P5**, whereas the domain ALEP appears in peptide **T13**. These data are in agreement with the ACE inhibitory activity observed in previous studies (Boschin et al. 2014a, Boschin et al. 2014b). Since ACE inhibitory and antioxidant peptides share some common structural requirements (Yea et al. 2014), some peptides exhibit multifunctional activities, for example peptide **T4** containing the sequences LLPH and PHY. Hydrophobic His-containing peptides often possess antioxidant capacity (Chen, Muramoto and Yamauchi 1995), probably because they could act as metal-ion chelators, active-oxygen quenches, and hydroxy-radical scavengers (Saito et al. 2003). The antioxidant character is also associated with the presence of hydrophobic and aromatic or heteroaromatic amino acids, such as Trp, Tyr, and Phe (Tanzadehpanah, Asoodeh and Chamani 2012), such as in peptides **T5** and **T6**, containing YVL and IY domains.

Another activity frequently proposed by BIOPEP is the ability to function as glucose uptake stimulating peptides, owing to the sequences LIL, IL, II, IV, LL, VL, and LV. Moreover, some peptides might also potentially function as inhibitors of the enzyme DPP‑IV (da Silva Júnior, de Godoy-Matos and Kraemer-Aguiar 2015). Interestingly, some recent antidiabetic drugs inhibit the action of this enzyme (Koska et al. 2015) and the hypoglycaemic activity of lupin protein is supported by numerous experimental evidences (Duranti et al. 2008).

Finally, four peptides **T2**, **T11**, **P7**, and **P12**, containing the sequences PG, were proposed by BIOPEP as potential inhibitors of prolyl endopeptidase (PEP) activity. This highly conserved serine protease is involved in several central nervous system functions (Lafarga, O'Connor and Hayes 2015) and its inhibition has potential applications in the prevention and treatment of

mental disorders. This is a completely unexplored area of potential bioactivity for lupin protein. Although very explorative, the search in BIOPEP has provided some tentative explanation of preceding experimental and clinical results and some hints of new potential areas of functional application of lupin foods.

In conclusion, for the first time this investigation has provided evidence that lupin peptides with specific structures are potentially absorbed in human intestinal cells maintaining their biological activity. These findings may help explaining the beneficial effects observed *in vivo* in animal and human studies (Arnoldi et al. 2015a, Duranti et al. 2008). It is, however, useful to add that a recent literature has shown that γ-conglutin, a specific hypoglycaemic lupin protein, is absorbed intact in Caco-2 cells (Capraro et al. 2011).

SUPPORTING INFORMATION

Table 3.1S Calculated pI's and net charges of BL peptides and their potential bioactivities according to BIOPEP database search

Short name	Peptide sequence	pI [a]	Net Charge [a]	Potential bioactive peptides [b]	Biological functions [b]
T1	IILGNEDEQEYEEQR	3.6	-5	LG, EY	ACE-inhibitor
				IL, II	Glucose uptake stimulating peptide
				EY, II, IL, NE, QE, YE	DPP IV inhibitor
T2	AVNELTFPGSAEDIER	3.8	-3	LTF, TF, PG, GS, IE	ACE inhibitor
				EL	Antioxidative peptide
				AV, NE, LT, TF, PG, AE	DPP IV inhibitor
				PG	Prolyl endopeptidase inhibitor
T3	IVEFQSKPNTLILPK	9.5	1	VE, KP, ILP	ACE inhibitor
				KP	Antioxidative
				VE, FQ, QS, SK, PN, NT, LI, IL, PK	DPP IV inhibitor
				IV, LI, IL	Glucose uptake stimulating peptide
				EF	Hypotensive (Renin inhibitor)
T4	INEGALLLPHYNSK	7.5	1	EG, GA, LLP, PH, HY	ACE inhibitor
				LLPH, PHY	Antioxidative
				IN, NE, EG, AL, LL, LP, HY, YN, SK	DPP IV inhibitor
				LL	Glucose uptake stimulating peptide
				LLL	Stimulating vasoactive substance release
T5	HSDADYVLVVLNGR	5.4	0	DA, LN, NG, GR	ACE inhibitor
				YVL	Antioxidative
				HS, AD, LV, VL, LN, NG	DPP IV inhibitor
				VL, LV	Glucose uptake stimulating peptide
				DY	Ion flow regulating peptide
T6	SNEPIYSNK	6.3	0	IY, NK	ACE inhibitor
				IY	Antioxidative
				NE, EP, PI, YS	DPP IV inhibitor
T9	GQEQSHQDEGVIVR	4.4	-1	GQ, EG, GV, VR	ACE inhibitor
				GQ, QE, QS, SH, QD, EG, VI	DPP IV inhibitor
				IV	Glucose uptake stimulating peptide
				GQ	Neuropeptide inhibitor
T10	LLGFGINAYENQR	6.3	0	LG, GF, FG, GI, AY	ACE inhibitor
				AY	Antioxidative
				LL, GF, GI, IN, NA, AY, YE, NQ	DPP IV inhibitor

				LL	Glucose uptake stimulating peptide
T11	ELTFPGSAEDIER	3.8	-3	LTF, TF, FP, PG, GS, IE	ACE inhibitor
				EL	Antioxidative
				LT, TF, FP, PG, AE	DPP IV inhibitor
				PG	Prolyl endopeptidase inhibitor
T13	LNALEPDNTVQSEAGTIE TWNPK	3.8	-3	LN, ALEP, EA, AG, GT, IE	ACE inhibitor
				TW	Antioxidative
				LN, NA, AL, EP, DN, NT, TV, QS, AG, TI, ET, TW, WN, NP, PK	DPP IV inhibitor
				SE	Stimulating vasoactive substance release
T16	QEEQLLEQELENLPR	3.7	-4	PR	ACE inhibitor
				EL	Antioxidative
				QE, QL, LL, NE	DPP IV inhibitor
				LL	Glucose uptake stimulating peptide
				EE	Stimulating vasoactive substance release
P1	YEITPDRNPQVQD	3.9	-2	EI, PQ	ACE inhibitor
				YE, EI, TP, DR, NP, PQ, QV, VQ, QD	DPP IV inhibitor
P3	YDFYPSSTKDQQS	4.1	-1	FY, YP	ACE inhibitor
				KD	Antioxidative
				YD, YP, PS, TK, DQ, QQ, QS	DPP IV inhibitor
P5	LILPKHSDAD	5.4	0	ILP, DA	ACE inhibitor
				LI, IL, LP, PK, HK, HS, AD	DPP IV inhibitor
				LI, LI	Glucose uptake stimulating peptide
P6	LRIPAGSTSY	9.3	1	IPA, AG, GS, SY	ACE inhibitor
				RI, IPA, AG, TS, SY	DPP IV inhibitor
P7	LTFPGSAED	3.5	-2	LTF, FP, PG, GS	ACE inhibitor
				LT, TF, FP, PG, AE	DPP IV inhibitor
				PG	Prolyl endopeptidase inhibitor
P8	YEITPDRNPQVQDLD	3.7	-3	EI, PQ	ACE inhibitor
				YE, EI, TP, DR, NP, PQ, QV, QD	DPP IV inhibitor
P12	IPPGIPYWT	6.0	0	IPP, PG, GI, IPY, YW	ACE inhibitor
				IP, PG, PPG, GI, PY, YW, WT	DPP IV inhibitor
				PG	Prolyl endopeptidase inhibitor

a) Predicted by the peptide property calculator (http:// www.genscript.com)
b) Bioactive properties obtained by BIOPEP data analysis

Author contributions

CL: ideation & experiment design, Caco-2 absorption experiments, & manuscript writing. GA: mass spectrometry & manuscript writing. GV: *in silico* study and manuscript writing. CZ: HMGCoAR activity inhibition. AA: manuscript writing & grant retrieval. YS and SF: result discussion. GR: part of Caco-2 absorption experiments.

References

Amigo-Benavent, M., Clemente, A., Caira, S., Stiuso, P., Ferranti, P., & del Castillo, M. D. (2014). Use of phytochemomics to evaluate the bioavailability and bioactivity of antioxidant peptides of soybean β-conglycinin. *Electrophoresis, 35*(11), 1582-1589.

Arnoldi, A., Boschin, G., Zanoni, C., & Lammi, C. (2015). The health benefits of sweet lupin seed flours and isolated proteins. *Journal of Functional Foods, 18*, 550-563.

Augustyns, K., Van der Veken, P., Senten, K., & Haemers, A. (2005). The therapeutic potential of inhibitors of dipeptidyl peptidase IV (DPP IV) and related proline-specific dipeptidyl aminopeptidases. *Current Medicinal Chemistry, 12*(8), 971-998.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America, 98*(18), 10037-10041.

Boschin, G., Scigliuolo, G. M., Resta, D., & Arnoldi, A. (2014a). ACE-inhibitory activity of enzymatic protein hydrolysates from lupin and other legumes. *Food Chemistry, 145*, 34-40.

Boschin, G., Scigliuolo, G. M., Resta, D., & Arnoldi, A. (2014b). Optimization of the enzymatic hydrolysis of lupin (Lupinus) proteins for producing ACE-inhibitory peptides. *Journal of Agricultural and Food Chemistry, 62*(8), 1846-1851.

Bähr, M., Fechner, A., Kiehntopf, M., & Jahreis, G. (2015). Consuming a mixed diet enriched with lupin protein beneficially affects plasma lipids in hypercholesterolemic subjects: a randomized controlled trial. *Clinical Nutrition, 34*(1), 7-14.

Cam, A., & de Mejia, E. G. (2012). Role of dietary proteins and peptides in cardiovascular disease. *Molecular Nutrition and Food Research, 56*(1), 53-66.

Capraro, J., Clemente, A., Rubio, L. A., Magni, C., Scarafoni, A., & Duranti, M. (2011). Assessment of the lupin seed glucose-lowering protein intestinal absorption by using in vitro and ex vivo models. *Food Chemistry, 125*(4), 1279-1283.

Chen, H. M., Muramoto, K., & Yamauchi, F. (1995). Structural-Analysis of antioxidative peptides from soybean beta-conglycinin. *Journal of Agricultural and Food Chemistry, 43*(3), 574-578.

Da Silva Júnior, W. S., de Godoy-Matos, A. F., & Kraemer-Aguiar, L. G. (2015). Dipeptidyl Peptidase 4: A New Link between Diabetes Mellitus and Atherosclerosis? *Biomed Res Int, 2015*, 816164.

Duranti, M., Consonni, A., Magni, C., Sessa, F., & Scarafoni, A. (2008). The major proteins of lupin seed: characterisation and molecular properties for use as functional and nutraceutical ingredients. *Trends Food Science and Technology, 19*(12), 624-633.

Ferruzza, S., Rossi, C., Scarino, M. L., & Sambuy, Y. (2012). A protocol for differentiation of human intestinal Caco-2 cells in asymmetric serum-containing medium. *Toxicology in vitro, 26*(8), 1252-1255.

Ferruzza, S., Scarino, M. L., Gambling, L., Natella, F., & Sambuy, Y. (2003). Biphasic effect of iron on human intestinal Caco-2 cells: early effect on tight junction permeability with delayed onset of oxidative cytotoxic damage. *Cell Molecular Biology (Noisy-le-Grand), 49*(1), 89-99.

Foltz, M., Meynen, E. E., Bianco, V., van Platerink, C., Koning, T. M., & Kloek, J. (2007). Angiotensin converting enzyme inhibitory peptides from a lactotripeptide-enriched milk beverage are absorbed intact into the circulation. *Journal of Nutrition, 137*(4), 953-958.

Hashimoto, K., & Shimizu, M. (1993). Epithelial properties of human intestinal Caco-2 cells cultured in a serum-free medium. *Cytotechnology, 13*(3), 175-184.

Heyman, M., Crain-Denoyelle, A. M., Nath, S. K., & Desjeux, J. F. (1990). Quantification of protein transcytosis in the human colon carcinoma cell line CaCo-2. *Journal of Cell Physiology, 143*(2), 391-395.

Hidalgo, I. J., Raub, T. J., & Borchardt, R. T. (1989). Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology, 96*(3), 736-749.

Korb, O., Stützle, T., & Exner, T. E. (2009). Empirical scoring functions for advanced protein-ligand docking with PLANTS. *Journal of Chemical Information and Modeling, 49*(1), 84-96.

Koska, J., Sands, M., Burciu, C., & Reaven, P. (2015). Cardiovascular effects of dipeptidyl peptidase-4 inhibitors in patients with type 2 diabetes. *Diabetes and Vascular Disease Research, 12*(3), 154-163.

Lafarga, T., O'Connor, P., & Hayes, M. (2015). In silico methods to identify meat-derived prolyl endopeptidase inhibitors. *Food Chemistry, 175*, 337-343.

Lammi, C., Zanoni, C., Arnoldi, A., & Vistoli, G. (2015). Two Peptides from Soy beta-Conglycinin Induce a Hypocholesterolemic Effect in HepG2 Cells by a Statin-Like Mechanism: Comparative in Vitro and in Silico Modeling Studies. *Journal of Agricultural and Food Chemistry, 63*(36), 7945-7951.

Lammi, C., Zanoni, C., Scigliuolo, G. M., D'Amato, A., & Arnoldi, A. (2014). Lupin Peptides Lower Low-Density Lipoprotein (LDL) Cholesterol through an Up-regulation of the LDL Receptor/Sterol Regulatory Element Binding Protein 2 (SREBP-2) Pathway at HepG2 Cell Line. *Journal of Agricultural and Food Chemistry, 62*(29), 7151-7159.

Levashov, P. A., Sutherland, D. S., Besenbacher, F., & Shipovskov, S. (2009). A robust method of determination of high concentrations of peptides and proteins. *Analytical Biochemistry, 395*, 111-112.

Miguel, M., Dávalos, A., Manso, M. A., de la Peña, G., Lasunción, M. A., & López-Fandiño, R. (2008). Transepithelial transport across Caco-2 cell monolayers of antihypertensive egg-derived peptides. PepT1-mediated flux of Tyr-Pro-Ile. *Molecular Nutrition and Food Research, 52*(12), 1507-1513.

Miner-Williams, W. M., Stevens, B. R., & Moughan, P. J. (2014). Are intact peptides absorbed from the healthy gut in the adult human? *Nutrition Research Reviews, 27*(2), 308-329.

Minkiewicz, P., Dziuba, J., Darewicz, M., Bucholska, J., & Mogut, D. (2012). Evaluation of In Silico Prediction Possibility of Epitope Sequences Using Experimental Data Concerning Allergenic Food Proteins Summarised in BIOPEP Database. *Polish Journal of Food and Nutrition Sciences, 62*(3), 151-157.

Natoli, M., Leoni, B. D., D'Agnano, I., D'Onofrio, M., Brandi, R., Arisi, I., Zucco, F., & Felsani, A. (2011). Cell growing density affects the structural and functional properties of Caco-2 differentiated monolayer. *Journal of Cell Physiology, 226*(6), 1531-1543.

Pak, V. V., Koo, M., Kim, M. J., Yun, L., & Kwon, D. Y. (2008). Binding effect and design of a competitive inhibitory peptide for HMG-CoA reductase through modeling of an active peptide backbone. *Bioorganic and Medicinal Chemistry, 16*(3), 1309-1318.

Regazzo, D., Mollé, D., Gabai, G., Tomé, D., Dupont, D., Leonil, J., & Boutrou, R. (2010). The (193-209) 17-residues peptide of bovine β-casein is transported through Caco-2 monolayer. *Molecular Nutrition and Food Research, 54*(10), 1428-1435.

Saito, K., Jin, D. H., Ogawa, T., Muramoto, K., Hatakeyama, E., Yasuhara, T., & Nokihara, K. (2003). Antioxidative properties of tripeptide libraries prepared by the combinatorial chemistry. *Journal of Agricultural and Food Chemistry, 51*(12), 3668-3674.

Sambuy, Y., De Angelis, I., Ranaldi, G., Scarino, M. L., Stammati, A., & Zucco, F. (2005). The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biology and Toxicology, 21*(1), 1-26.

Shi, J., Blundell, T. L., & Mizuguchi, K. (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *Journal of Molecular Biology, 310*(1), 243-257.

Sirtori, C. R., Triolo, M., Bosisio, R., Bondioli, A., Calabresi, L., De Vergori, V., Gomaraschi, M., Mombelli, G., Pazzucconi, F., Zacherl, C., & Arnoldi, A. (2012). Hypocholesterolaemic effects of lupin protein and pea protein/fibre combinations in moderately hypercholesterolaemic individuals. *British Journal of Nutrition, 107*(8), 1176-1183.

Sirtori, E., Resta, D., Brambilla, F., Zacherl, C., & Arnoldi, A. (2010). The effects of various processing conditions on a protein isolate from Lupinus angustifolius. *Food Chemistry, 120*(2), 496-504.

Tanzadehpanah, H., Asoodeh, A., & Chamani, J. (2012). An antioxidant peptide derived from Ostrich (Struthio camelus) egg white protein hydrolysates. *Food Research International, 49*(1), 105-111.

Thwaites, D. T., Brown, C. D., Hirst, B. H., & Simmons, N. L. (1993). Transepithelial glycylsarcosine transport in intestinal Caco-2 cells mediated by expression of H(+)-coupled carriers at both apical and basal membranes. *Jornal of Biological Chemistry, 268*(11), 7640-7642.

Vistoli, G., Pedretti, A., Mazzolari, A., & Testa, B. (2010). In silico prediction of human carboxylesterase-1 (hCES1) metabolism combining docking analyses and MD simulations. *Bioorganic and Medicinal Chemistry, 18*(1), 320-329.

Yea, C. S., Ebrahimpour, A., Hamid, A. A., Bakar, J., Muhammad, K., & Saari, N. (2014). Winged bean [Psophorcarpus tetragonolobus (L.) DC] seeds as an underutilised plant source of bifunctional proteolysate and biopeptides. *Food & Function, 5*(5), 1007-1016.

Zhu, X. L., Watanabe, K., Shiraishi, K., Ueki, T., Noda, Y., Matsui, T., & Matsumoto, K. (2008). Identification of ACE-inhibitory peptides in salt-free soy sauce that are transportable across caco-2 cell monolayers. *Peptides, 29*(3), 338-344.

CHAPTER 4

MANUSCRIPT 2

LUPIN PEPTIDES MODULATE THE PROTEIN-PROTEIN INTERACTION OF PCSK9 WITH THE LOW DENSITY LIPOPROTEIN RECEPTOR IN HEPG2 CELLS

Carmen Lammi, Chiara Zanoni, Gilda Aiello, Anna Arnoldi* & Giovanni Grazioso

Department of Pharmaceutical Sciences, University of Milan, Milan, Italy.

4.0 Abstract

Proprotein convertase subtilisin/kexin type 9 (PCSK9) has been recently identified as a new useful target for hypercholesterolemia treatment. This work demonstrates that natural peptides, deriving from the hydrolysis of lupin proteins and absorbable at intestinal level, are able to inhibit the protein-protein interaction between PCSK9 and the low density lipoprotein receptor (LDLR). In order to sort out the best potential inhibitors among these peptides, a refined in silico model of the PCSK9/LDLR interaction was developed. Docking simulation, permitted to select the two best candidates (**T9** and **P5**) among the absorbed peptides that were synthesized and evaluated for their inhibitory activity. In order to quantify the absorbed amount of **P5** and **T9** through the BL compartment, label-free MRM mass spectrometric-based assays were developed. From a functional point of view, the most active was **P5** that induced a concentration dependent inhibition of the PCSK9-LDLR binding, with an IC_{50} value equal to 1.6 ± 0.33 μM. Tested at a 10 μM concentration, this peptide increased by 66 ± 21.4% the ability of HepG2 cells to take up LDL from the extracellular environment.

4.1 Introduction

An increased level of low-density lipoproteins (LDL) predisposes to the development of cardiovascular disease (CVD) and stroke. Currently, the main agents for lowering LDL levels are statins. They reduce the regulatory pool of intracellular cholesterol by competitively inhibiting HMGCoAR, the rate-limiting enzyme of endogenous cholesterol biosynthesis. This in turn activates the LDL receptor (LDLR) transcription, a process under the control of sterol regulatory element binding protein 2 (SREBP-2) (Brown & Goldstein, 2009; Dubuc, Chamberland, Wassef, Davignon, Seidah, Bernier, et al., 2004). However, intolerance to statins produces numerous side effects, such as myopathy with a spectrum of consequences ranging from myalgia to rhabdomyolysis (Wong, Chuang, Wong, Pham, Neff, & Marrett, 2013). To overcome these issues, alternative or adjunctive therapies are thus advisable to use. In this context, proprotein convertase subtilisin/kexin type 9 (PCSK9) has been recently identified as a new useful target for hypercholesterolemia treatment (Seidah & Prat, 2007).

PCSK9 is expressed primarily in liver, kidney, and intestine (Seidah & Prat, 2012) and plays an important role in regulating the degradation of hepatic LDLR receptor (Gencer, Lambert, & Mach, 2015; D. W. Zhang, Lagace, Garuti, Zhao, McDonald, Horton, et al., 2007).

Notably, since PCSK9 and LDLR are co-regulated by SREBP-2 (Ambegaonkar, Tipping, Polis, Tomassini, & Tershakovec, 2014), increased PCSK9 expression in response to statin-induced cellular cholesterol depletion may limit the efficacy of statin treatment (Careskey, Davis, Alborn, Troutt, Cao, & Konrad, 2008; Welder, Zineh, Pacanowski, Troutt, Cao, & Konrad, 2010).

The development of therapies that inhibit PCSK9 function holds promise for improved management of hypercholesterolemia and CVD risk.

In light of these observations and of the fact that circulating PCSK9 may cause the degradation of hepatic LDLR, this protein is considered an attractive target for LDL-cholesterol lowering drugs. The current inhibitors under development are based on different strategies, i.e. monoclonal antibodies (mAbs) (such as evolocumab and alirocumab) (Everett, Smith, & Hiatt, 2015)19, gene silencing compounds (Graham, Lemonidis, Whipple, Subramaniam, Monia, Crooke, et al., 2007), and natural product such as berberine (Dong, Li, Singh, Cao, & Liu, 2015), and peptidomimetics (Shan, Pang, Zhang, Murgolo, Lan, & Hedrick, 2008).

The ability of PCSK9 to mediate LDLR degradation involves protein-protein interactions (PPIs) between LDLR and the PCSK9. In this context, PPIs, which are inherently challenging

small molecule targets, have been successfully inhibited by peptides, which can recapitulate key protein contacts (Ross, Katt, & Hamilton, 2010).

To date, numerous studies have demonstrated the health benefits in the area of hypercholesterolemia prevention provided by lupin peptides. Tryptic and peptic peptides deriving from lupin protein hydrolysis are able to modulate cholesterol metabolism in HepG2 cells by inhibiting HMGCoAR via a statin-like mechanism (Lammi, Zanoni, Scigliuolo, D'Amato, & Arnoldi, 2014). Another investigation in human Caco-2 cells has provided evidence that at least some of these peptic and tryptic peptides may be absorbed in the small intestine, since they are transferred from the apical (AP) to the basolateral (BL) compartment in differentiated Caco-2 cells grown in a two-compartment system (Lammi, Aiello, Vistoli, Zanoni, Arnoldi, Sambuy, et al., 2016). **Table 4.1** shows the sequences of the transferred peptides that here will be referred to as "absorbable lupin peptides".

Finally, another study in mild hypercholesterolemic subjects has demonstrated that the level of circulating PCSK9 is reduced in a statistically significant way after a 4-week consumption of a lupin protein isolate (30 g protein/day) (Lammi, Zanoni, Calabresi, & Arnoldi, 2016). The same paper reported also an investigation in human HepG2 cells showing that lupin peptides reduce the synthesis and secretion of mature PCSK9 as well as the level of hepatic nuclear factor 1-alpha (HNF1-alpha) (Lammi, Zanoni, Calabresi, & Arnoldi, 2016).

In this context, the present work was aimed at evaluating whether "absorbable lupin peptides" may directly inhibit the PCSK9-LDLR PPI. After sorting out the best potential inhibitors among "absorbable peptides" by computational assays, peptic and tryptic peptides were synthesised in order to evaluate their inhibitory activity of the PCSK9-LDLR interaction, and their capability of reducing the LDL uptake by HepG2 cells. In a parallel way, a label-free quantification of the best PCSK9-LDLR PPI peptide inhibitors in BL medium using an MS-based approach was developed.

Table 4.1 Peptides identified in the basolateral compartment of Caco-2 cells grown in two compartment systems

Peptides	Sequence	Uniprot Code	Protein Source	Proteolytic enzyme
T1	IILGNEDEQEYEEQR	Q6EBC1	β2-conglutin	Trypsin
T2	AVNELTFPGSAEDIER	Q6EBC1	β2-conglutin	Trypsin
T3	IVEFQSKPNTLILPK	Q6EBC1	β2-conglutin	Trypsin
T4	INEGALLLPHYNSK	Q6EBC1	β2-conglutin	Trypsin
T5	HSDADYVLVVLNGR	Q6EBC1	β2-conglutin	Trypsin
T6	SNEPIYSNK	Q6EBC1	β2-conglutin	Trypsin
T9	GQEQSHQDEGVIVR	Q53HY0	β1-conglutin	Trypsin

T10	LLGFGINAYENQR	Q53HY0	β1-conglutin	Trypsin
T13	LNALEPDNTVQSEAGTIETWNPK	Q53I54	α-conglutin	Trypsin
T16	QEEQLLEQELENLPR	Q333K7	δ-conglutin	Trypsin
P1	YEITPDRNPQVQDL	Q6EBC1	β2-conglutin	Pepsin
P3	YDFYPSSTKDQQS	Q6EBC1	β2-conglutin	Pepsin
P5	LILPKHSDAD	Q6EBC1	β2-conglutin	Pepsin
P6	LRIPAGSTSY	Q6EBC1	β2-conglutin	Pepsin
P7	LTFPGSAED	Q6EBC1	β2-conglutin	Pepsin

4.2 Methods

4.2.1 Preparation of the pepsin and trypsin peptide mixtures

The peptic (**P-tot**) and tryptic (**T-tot**) hydrolysates from lupin protein are the same used in a previous paper, in which their capability to modulate cholesterol metabolism in HepG2 cells was investigated (Lammi, Zanoni, Scigliuolo, D'Amato, & Arnoldi, 2014). The procedure for protein extraction from *Lupinus albus* (cultivar Ares) seeds, the conditions of enzymatic hydrolysis, and the final composition of the resulting peptide mixtures are reported in detail in the same paper. In the tryptic sample, 12 peptides were assigned to *L. albus* vicilin-like protein (Q53HY0), 10 peptides to *L. albus* beta-conglutin precursor (Q6EBC1), 4 peptides to *L. angustifolius* conglutin beta (B0YJF8)], 4 peptides to *L. angustifolius* conglutin-beta (Q53I55)], and 2 to *L. albus* conglutin-delta seed storage protein precursor (Q99235). In the peptic hydrolysate, 21 peptides were assigned to *L. albus* vicilin-like protein (Q53HY0), 18 peptides to *L. albus* beta-conglutin (Q6EBC1), 7 peptides to *L. angustifolius* conglutin-alpha 3 (F5B8V8], and 8 peptides to *L. albus* conglutin-gamma (Q9FSH9). These mixtures were submitted to absorption experiments in Caco-2 cells grown in two-compartment trans-well systems (Lammi, et al., 2016). The absorbed peptic (**P-abs**) and tryptic peptides (**T-abs**), i.e. those transferred in the basolateral compartments, are shown in **Table 4.1**. The details for the preparation and quantification of the peptides and the analytical procedure for their identification based on mass spectrometry are described in detail in the previous manuscript.

4.2.2 *In vitro* PCSK9-LDLR binding assay

P-tot peptides (1.0 and 2.5 μg/μL), **T-tot** peptides (0.5 and 1.0 μg/μL), **P-abs** (12.5 ng/μL) and **T-abs** peptides (16.0 ng/μL) were tested using the in vitro PCSK9-LDLR binding assay (CycLex Co., Nagano, Japan) following the manufacture instructions. Briefly, plates are pre-coated with a recombinant LDLR-AB domain, which contains the binding site for PCSK9.

Before starting the assay, tested peptides and/or the vehicle were diluted in reaction buffer and added in micro centrifuge tubes. Afterwards, the reaction mixtures were added in each well of the microplate and the reaction was started by adding His-tagged PCSK9 wild type solution (3 μL). The microplate was allowed to incubate for 2 h at room temperature (RT) shaking at 300 rpm on an orbital microplate shaker. Subsequently, wells were washed 4 times with wash buffer. After the last wash, the biotinylated anti-His-tag monoclonal antibody (100 μL) was added and incubated at RT for 1 h shaking at 300 rpm. After incubation, wells were washed for 4 times with wash buffer. After the last wash, 100 μL of HRP-conjugated streptavidin were added and the plate was incubated for 20 min at RT. After incubation, wells were washed 4 times with wash buffer. Finally, the substrate reagent (tetra-methylbenzidine) was added and the plate was incubated for 10 min at RT shaking at ca. 300 rpm. The reaction was stopped with 2.0 N sulfuric acid and the absorbance at 450 nm was measured using the Synergy H1 fluorescent plate reader (Biotek, Bad Friedrichshall, Germany).

4.2.3 *In vitro* PCSK9-LDLR binding assay of synthetic T9 and P5

T9 (GQEQSHQDEGVIVR) and **P5** (LILPKHSDAD) were synthesized by the company PRIMM (Milan, Italy). They were > 95% pure by HPLC.

Synthetic peptides, i.e. **T9** (32–320 μM) and **P5** (0.001–100 μM), were tested using the in vitro PCSK9-LDLR binding assay (CycLex Co., Nagano, Japan) following the manufacture instructions as reported above. The absorbance at 450 nm was measured using the Synergy H1 fluorescent plate reader (Biotek, Bad Friedrichshall, Germany). In particular, for the in vitro screening of the synthetic PCSK9-LDLR inhibitors, **T9** and **P5**, at different concentrations, were added to the appropriate amount of His-tagged PCSK9 in the wells that had been coated with recombinant LDLR-AB domain in a similar fashion as described above, followed by evaluation of inhibitory effect on PCSK9-LDLR interaction by measuring the amount of His-tagged PCSK9 on the wells which is correlated to the absorbance signals at 450 nm, which were measured using the Synergy H1 fluorescent plate reader (Biotek, Bad Friedrichshall, Germany).

4.2.4 Cell line culture

The HepG2 cell line was bought from ATCC (HB-8065, ATCC from LGC Standards, Milan, Italy). HepG2 cells were cultured in DMEM high glucose with stable L-glutamine

supplemented with 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin (complete growth medium) and incubated at 37 °C under 5% CO_2 atmosphere. HepG2 cells were used for no more than 20 passages after thawing, because the increase of the number of passages may change the cell characteristics and impair assay results.

4.2.5 Fluorescent LDL uptake cell based assay

HepG2 cells (3×10^4/well) were seeded in 96-well plates and kept in complete growth medium for 2 d before treatment. The third day, they were treated with **T9** (160 and 320 μM) or **P5** (10 μM), or vehicle (H_2O) for 24 h. At the end of the treatment, the culture medium was replaced with 50 μL/well LDL-DyLight™ 550 working solution (Cayman Chemical Company, Ann Arbor, MI, US). The cells were additionally incubated for 2 h at 37 °C and then the culture medium was aspirated and replaced with PBS (100 μL/well). The degree of LDL uptake was measured using the Synergy H1 fluorescent plate reader from Biotek (excitation and emission wavelengths 540 and 570 nm, respectively).

4.2.6 Statistical analysis of biological assays

Data are presented as mean ± standard error using GraphPad Prism 6 (CA, USA). Statistical analyses were carried out by one-way ANOVA followed by Dunnett's test. *P-values* < 0.05 were considered to be significant.

4.2.7 HPLC-Chip-MRM analysis of peptide T9 and P5 in T-abs and P-abs

The LC-MRM analysis was carried out in positive ionization mode on a SL series Ion Trap mass spectrometer equipped with a nano-ESI source interfaced with a HPLC-Chip Cube source (Agilent Technologies, Palo Alto, CA, USA). The chromatographic separation was performed on a LC-chip containing a 40 nL enrichment column (Zorbax 300SB-C18, 5 μ m pore size), a 43 mm × 75 μ m analytical column packed (Zorbax 300SB-C18, 5 μ m pore size). Both **P-abs** and **T-abs** (1 μL) were loaded onto the enrichment column at a flow rate 4 μL/min for 2 min using isocratic 100% C solvent phase (99% water, 1% ACN and 0.1% FA). After clean-up, the chip valve was switched to separation and trapped peptides were eluted into the mass spectrometer at the constant flow rate of 0.3 μL/min using H_2O, 0.1% HCOOH (A) and ACN, 0.1% HCOOH (B) as elution solvent.

The nano-pump gradient used for **P5** determination was the following: 3% solvent B (0 min), 30% solvent B (0–4 min) and back to 3% in 1 min. The gradient used for **T9** was the following: 0% solvent B (0 min), 60% solvent B (0–8 min). A column re-equilibration time of 2 min was used after each analysis in both peptide assays. The nano-ESI source operated under the following conditions: drying gas temperature 300 °C, flow rate 3 L/min (nitrogen), capillary voltage 1950 V, with endplate offset − 500V. The LC-MS/MS analysis, under MRM condition, were carried out monitoring one diagnostic transition for **P5**, *m/z* 554.8 → *m/z* 770.0, selecting the most intense y-fragment ion, corresponding to ***y7***; and two diagnostic transitions for **T9** *m/z* 528.4 → *m/z* 543.3 and *m/z* 387.1, corresponding to ***y5*** and ***y3*** respectively.

4.2.8 Quantification of T9 in T-abs and P5 in P-abs

The concentration of **T9** and **P5** respectively in **T-abs** and **P-abs** was estimated using external calibration curves. The concentrations of **P5** in the standard solution ranged from 5 to 50 ng/µL, whereas those of **T9** ranged from 10 to 100 ng/mL. The automated software QuantAnalysis (version 1.7) was used to build the calibration curves by integrating at each calibration level (CL) the area of the *m/z* 770.0, the main product ion of double charged *m/z* 554.8 **P5** precursor ion and the area of *m/z* 543.3 and *m/z* 387.1 of the triple charged *m/z* 528.4 **T9.** The equations of the curves were calculated using five calibration points with three set of replicate (*n = 3*).

4.2.9 Analytical parameter evaluation

The accuracy was evaluated by adding known amounts of **P5** and **T9** to BL control samples, respectively. The repeatability of the method was obtained by injecting **P-abs** and **T-abs** samples and the one calibration level three times each. The limit of detection (LOD) and the limit of quantification (LOQ) were calculated by applying the Eqs. $S_{LOD} = S_{RB}+3\ \sigma_{RB}$ and $S_{LOQ} = S_{RB}+10\ \sigma_{RB}$, respectively, following the directives of FDA guide for method validation, where S_{LOD} is the signal at the limit of detection, S_{LOQ} is the limit of quantification, S_{RB} is the signal of blank control, and σ_{RB} is the standard deviation.

4.3 Results

4.3.1 Absorbed peptic and tryptic lupin peptides maintain their capacity to interfere with the PCSK9/ LDLR PPI

As already explained in the introduction, peptic peptides (**P-tot**) and tryptic peptides (**T-tot**) from lupin proteins reduce the synthesis and secretion of mature PCSK9 in HepG2 cells (Lammi, Zanoni, Calabresi, & Arnoldi, 2016). In parallel, other experiments, performed in Caco-2 cells grown in two-compartment systems, have indicated that some of these peptic and tryptic peptides might be absorbed in the small intestine (Lammi, et al., 2016). Precisely, Caco-2 cells have been incubated for 4 h in the AP chamber with 500 µL of a 1.0 µg/µL solution of **P-tot** or **T-tot** peptides and, at the end of treatment, BL solutions (700 µL) have been collected and absorbed peptic and tryptic peptides identified by HPLC-Chip ESI-MS/MS. The BL solutions were then dried and diluted to 100 µL to obtain the solutions **P-abs** and **T-abs** at the concentration of 0.25 and 0.32 µg/µL, respectively. Both solutions were used for the present experiments.

The first experiment had the objective of comparing the ability of total and absorbed peptides to inhibit the PCSK9-LDLR binding. In vitro experiments were performed using a solid-phase binding assay between recombinant His-tagged PCSK9 and recombinant LDLR-AB domain. The results are shown in **Figure 4.1**. **P-tot** peptides reduced the PCSK9-LDLR binding by 22 ± 6.5% and 25 ± 13.7% ($p < 0.001$) at 1.0 and 2.5 µg/µL, respectively (Chart 1A), whereas **T-tot** peptides by 26 ± 3% and 23 ± 2% at 0.5 and 1.0 µ g/µL, respectively. **P-abs** peptides (12.5 ng/µL) and **T-abs** peptides (16.0 ng/µL) inhibited the PCSK9-LDLR interaction by 81 ± 7.6% and 58 ± 4.1% ($p < 0.0001$), respectively, versus the untreated sample (Chart 1B).

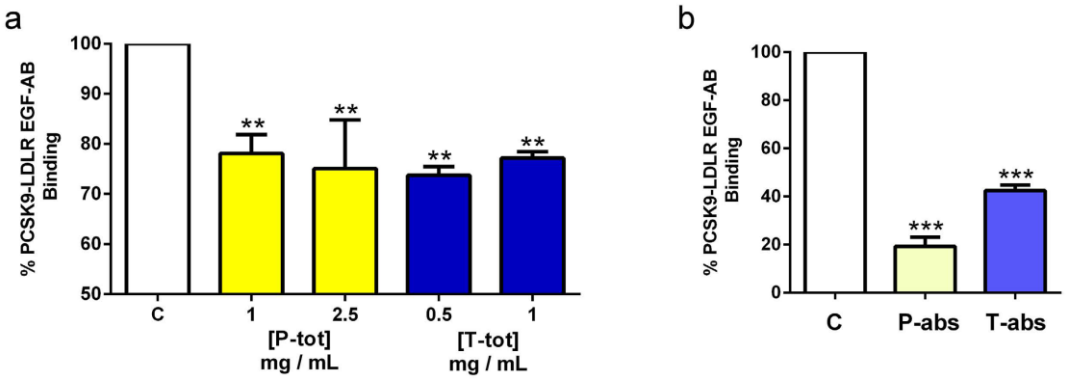


Figure 4.1 - Effects of peptic and tryptic lupin peptides on the PCSK9-LDLR binding inhibition. **(a)** In vitro screening of the inhibitory activity of total lupin peptides, i.e. **P-tot** (1.0 and 2.5 µg/µL) and

T-tot (0.5 and 1.0 µg/µL), on PCSK9-LDLR PPI; and **(b)** of absorbed lupin peptides, i.e. **P-abs** and **T-abs** (diluted 1:20). The commercial assay provides a plate pre-coated with a recombinant LDLR-AB domain, which is the PCSK9 binding site. For screening direct inhibitors of the PCSK9-LDLR interaction, lupin peptides were added to the proper amount of His-tagged PCSK9 in each well coated with LDLR-AB domain. Biotinylated anti-His-tag monoclonal antibody specifically reacted with recombinant His-tagged PCSK9 trapped with LDLR-AB domain immobilized on the microplate surface. The inhibitory effects were analysed measuring the amount of His tagged PCSK9 on the well. In particular, the inhibitory effect, calculated as % PCSK9-LDLR binding inhibition, determines a reduction of absorbance measured at 450 nm using the Synergy H1 fluorescent plate reader from Biotek versus vehicle control. Data points represent averages ± SEM of three independent experiments in duplicate. * * *p < 0.0001 * * p < 0.001 versus C. C, control vehicle.

4.3.2 *In silico* model of the human PCSK9 and lupin peptides

In order to establish which peptides were mainly responsible for the PCSK9/LDLR PPI inhibition, the 3D structure of PCSK9 was modelled. In this sense, all peptides were modelled and used for docking calculations into the LDLR anchor domain of PCSK9. At the end of simulations, the energy minimized structures of small peptides **T1**, **T2**, and **T16** showed helix conformation, **T3**, **T4**, **T5**, **T6**, and **T9** were folded as random coils, **P1**, **P3**, **P5**, **P7**, **T10**, and **P6** showed γ -turn conformations, while peptide **T13** was shaped as β –hairpin **Figure 4.2** show the three-dimensional structures of lupin small-peptides.

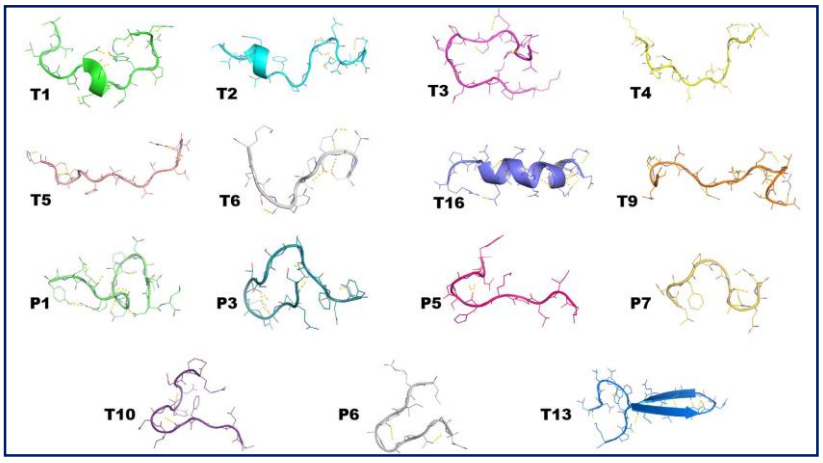


Figure 4.2 - Three-dimensional structure of lupin small-peptides at the end of MD simulations and subsequent energy minimization.

The docking simulation permitted estimating their theoretical binding energy values on PCSK9. The lowest theoretical binding energy values were calculated for **T9** and **P5 (Figure 4.3)**, which were further synthetized (Lammi, Zanoni, Aiello, Arnoldi, & Grazioso, 2016).

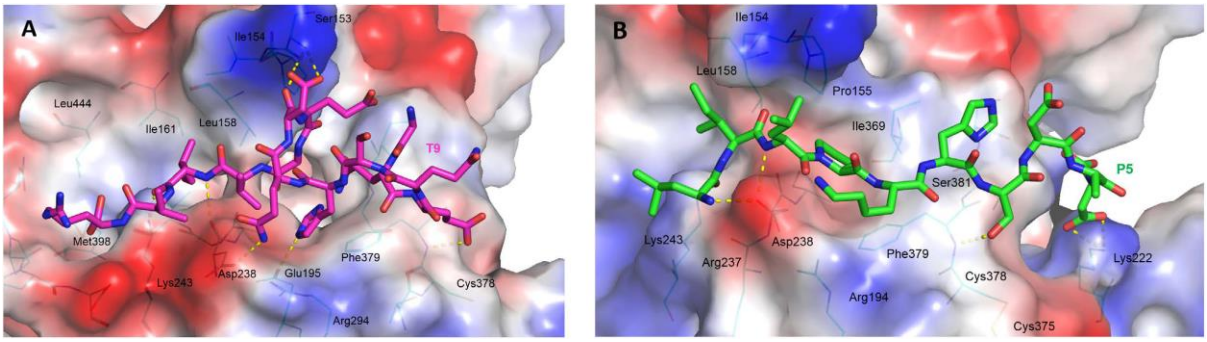


Figure 4.3 - Docking and MD simulations results for T9 (panel A) and P5 (panel B). T9 and **P5** are represented as magenta and green stick models, respectively. The showed conformations of **T9** and **P5** were retrieved minimizing MD frames in which the systems displayed geometrical stability. PCSK9 is displayed as cyan thin stick model and only residues within a radius of 4 Å from docked peptides are shown and labeled in the pictures. PCSK9 surface is colored according to the atoms electrostatic charges (blue for positive and red for negative), suitably calculated by Pymol tools. The materials and method section for the Docking, MD simulations and peptide binding energy estimations are reported in the published paper.

4.3.3 Isolated lupin peptides inhibit the PCSK9-LDLR binding

In order to assess the efficacy of the bioinformatics prediction, **T9** and **P5**, identified as the best potential inhibitors among **T-abs** and **P-abs** peptides, were synthesized and tested for their capacity to inhibit the PCSK9-LDLR binding.

T9 significantly reduced the PCSK9-LDLR binding by 19 ± 8.47%, 33 ± 5.33%, and 53 ± 1.02% ($p < 0.05$ and $p < 0.001$) at 32, 160, and 320 μM concentrations, respectively, versus the untreated sample (**Figure 4.4a**). **P5** (tested in the range 0.001–100 μM) induced a concentration dependent inhibition of the PCSK9-LDLR binding, with an IC_{50} value equal to 1.6 ± 0.33 μM (**Figure 4.4b**).

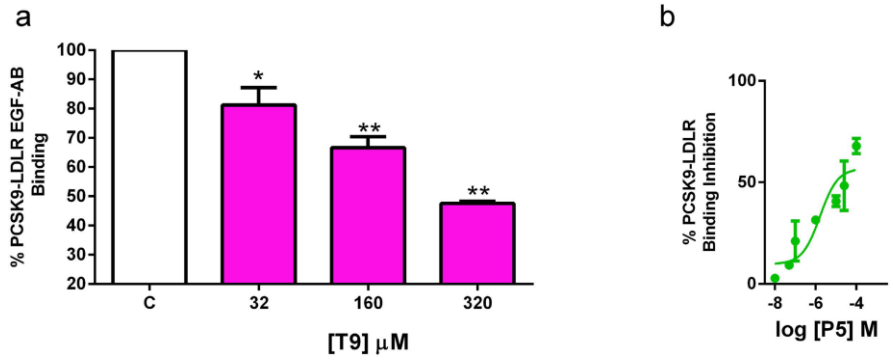


Figure 4.4 - Peptides T9 and P5 inhibit the PCSK9-LDLR binding. Inhibitory effects of **(a) T9** (32, 160, and 320 μM) and **(b) P5** (0.001–100 μM) on the PCSK9-LDLR PPI in vitro. The employed assay provides a plate precoated with a recombinant LDLR-AB domain, which is the PCSK9 binding site. For screening, the peptides were added to proper amount of His-tagged PCSK9 in each well coated with

LDLR-AB domain. Biotinylated anti-His-tag monoclonal antibody specifically reacts with recombinant His-tagged PCSK9 trapped with LDLR-AB domain immobilized on the microplate surface. The inhibitory effects of **T9** and **P5** were analysed measuring the amount of His-tagged PCSK9 on the well. The **T9** and **P5** inhibition, calculated as %PCSK9-LDLR binding inhibition, was measured from the absorbance reduction at 450 nm, using the Synergy H1 fluorescent plate reader (Biotek) versus vehicle (control). **P5** showed a concentration-response behaviour, with an IC_{50} equal to 1.6 μM. Data points represent averages ± SEM of three independent experiments in duplicate. * $p < 0.05$ * * $P < 0.001$ versus C. C, vehicle.

4.3.4 Through the inhibition of PCSK9-LDR PPI, T9 and P5 increase the ability of HepG2 cells to uptake LDL from the extracellular environment

For a deeper insight of the activity of **T9** and **P5**, the change of the functional capability of HepG2 cells to uptake extracellular LDL was investigated after treatment with these peptides by performing fluorescent-LDL uptake experiments. As shown in **Figure 4.5**, each peptide increased the LDL-uptake versus the control in a statistically significant way ($p < 0.001$). In fact, the treatment with **P5** (10 μM) led to an increase of the LDL-uptake by 66 ± 21.4%, whereas that with **T9** (100 μM) increased the LDL-uptake by 55 ± 24.4%, versus the untreated sample. Therefore, both **P5** and **T9** are able to improve the HepG2 ability to absorb extracellular LDL, although their potency is very different, being the latter about ten times less active than the former. These results indicate that the capacity of **T9** and **P5** to impair the binding of secreted PCSK9 to the LDLR stabilizes the active LDLR on cell membrane, leading to an improved ability of hepatic cells to uptake extracellular LDL with a final hypocholesterolemic effect.

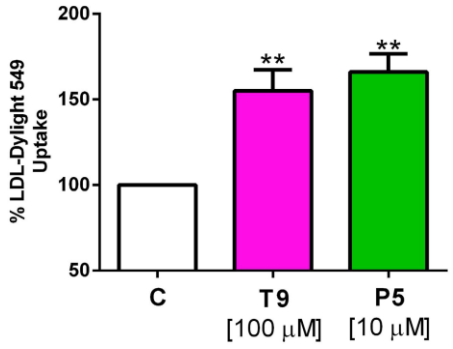


Figure 4.5 - Fluorescent LDL-uptake assay after lupin peptide treatments of HepG2.
Specific fluorescent LDL uptake signal was analysed by Synergy H1 (Biotek). Data points represent averages ± SEM of three independent experiments in triplicate. ** $p< 0.001$ versus C. C, vehicle.

4.3.5 LC-MRM assays for the T9 and P5 quantification

In order to quantify the degree of transport of both peptides **T9** and **P5** across the epithelium barrier, quantitative analysis carried out by using an HPLC-CHIP connected with an ion trap set in MRM mode. Exploiting the high sensitivity and specificity than the data-dependent full scan mode, in which a subset of high signal peptides seen in the first MS stage (MS1) are subjected to the second MS/MS stage (MS2), the MRM mode has been set to monitor few transitions per peptides. High structural specificity was achieved monitoring the transition *m/z* 554.8 → *m/z* 770.0, which represents the most intense product ion among all fragments in MS/MS spectrum of **P5**. Two transitions, *m/z* 528.4→ *m/z* 543.3 and *m/z* 528.4→*m/z* 387.1 were instead used to quantify **T9**. **Figure 4.6** reports the MS2 spectra of **P5** and **T9**, respectively.

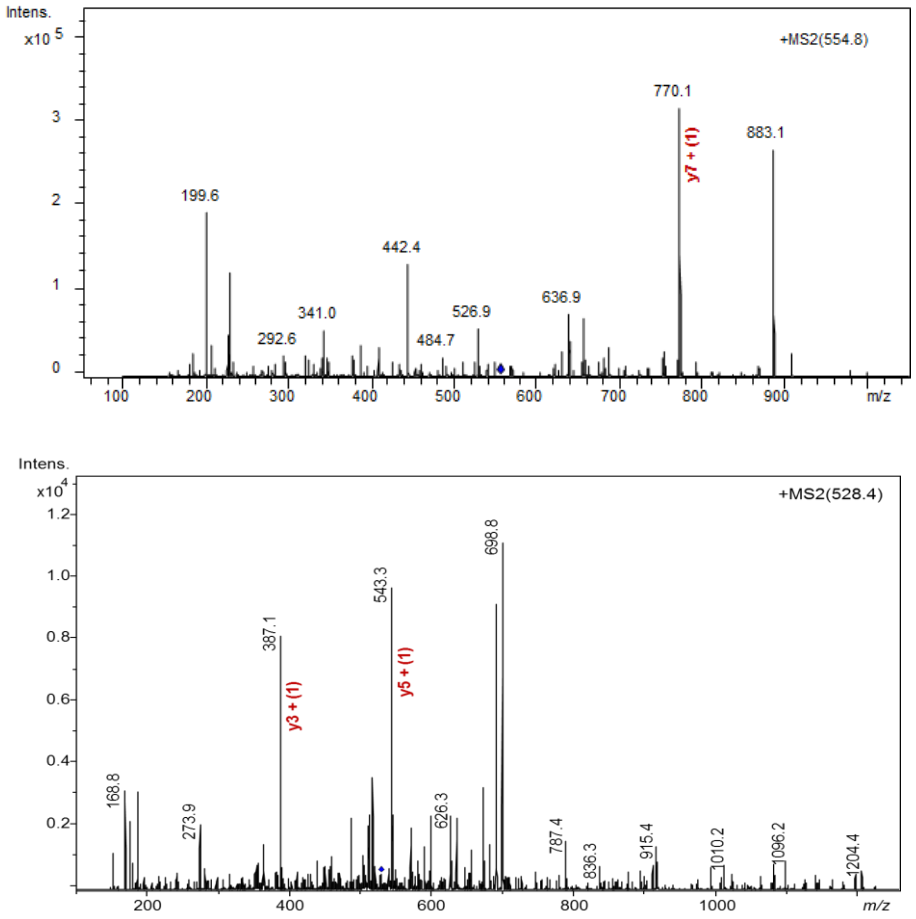


Figure 4.6 - MS/MS spectra of P5, A) and T9, B) respectively.

The CE values reported for all fragments were automatically optimized by the instrument software. QuantAnalysis data package was used to build calibration curves by integrating at

each calibration level the signal of the main product ion in the MS2 spectrum of each precursor ion. **Table 4.2** reports **A)** the details of the calibration curve, **B)** the peptide area values for each replicate analysis of calibration curves. The correspondent area averages, standard deviations, and RSDs% at each calibration levels are reported. By analysing the **P-abs** and **T-abs**, the **P5** concentration in the **P-abs** solution was calculated to be 28.64 ± 0.43 ng/µL, corresponding to 11.5% of the total peptic peptides absorbed by CaCo-2 cells. The relative standard deviation (RSD%) value resulted 1.5%. **T9** concentration in the **T-abs** was 25.02 ± 1.1 ng/mL, corresponding to 0.01% of the total tryptic peptides absorbed by Caco-2.

Table 4.2 Analytical parameters of quantification

A) Description of standard peptide curves for P5 and T9, respectively: concentrations, ng and injection volumes (IV) for each calibration level (CL) loaded on the HPLC-Chip column.

Standard peptide curve P5			
CL	**Concentration (ng/ µL)**	**IV (µL)**	**C (ng)**
1	5	1	5
2	10	1	10
3	20	1	20
4	30	1	30
5	50	1	50
Standard peptide curve T9			
CL	**Concentration (ng/ mL)**	**I.V. (µL)**	**C (ng)**
1	10	1	10
2	35	1	35
3	50	1	50
4	75	1	75
5	100	1	100

B) Peptide area averages (average), standard deviations (DV) and the RSDs% values at each calibration levels (CL) are reported. The RSD% ranged from 1.4 to 18.2%.

P5 CL	C (ng)	Replicates	Area	Average	DV	RSD%
1	5	1	4.46E+05	4.93E+05	8.97E+04	18.2
		2	4.37E+05			
		3	5.97E+05			
2	10	1	1.13E+06	1.22E+06	8.09E+04	6.6
		2	1.28E+06			
		3	1.26E+06			
3	20	1	4.09E+06	4.20E+06	1.95E+05	4.6
		2	4.43E+06			
		3	4.09E+06			
4	30	1	6.34E+06	6.44E+06	8.85E+04	1.4
		2	6.46E+06			
		3	6.51E+06			
5	50	1	1.30E+07	1.20E+07	9.21E+05	7.7
		2	1.18E+07			
		3	1.12E+07			
T9 CL	**C (ng)**	**Replicates**	**Area**	**Average**	**DV**	**RSD%**
1	10	1	1.24E+05	1.47E+05	2.04E+04	13.8
		2	1.58E+05			
		3	1.61E+05			
2	35	1	3.59E+05	3.78E+05	4.40E+04	11.6
		2	4.28E+05			
		3	3.46E+05			
3	50	1	7.15E+05	7.06E+05	8.81E+04	7.8
		2	7.03E+05			
		3	7.00E+05			
4	75	1	8.29E+05	8.40E+05	1.73E+04	2.1
		2	8.60E+05			
		3	8.32E+05			
5	100	1	1.18E+06	1.40E+06	1.84E+05	13.2
		2	1.51E+06			
		3	1.50E+06			

The accuracy of the method was determined prepared by spiking blank matrix fortified with **T9** and **P5** standard peptides. The accuracy was higher than 110% at 8 ng/μL for **P5**, and 97.5% at 75 ng/mL for **T9**. The calculated values of LOQ and LOD, corresponding to 3.92 and 3.71 ng/μL, respectively, confirm the reliability of the quantification method in the case of **P5** and 8.62 and 7.91 ng/mL in the case of **T9**.

4.4 Discussion

The discovery of PCSK9 in 2003 (Abifadel, Varret, Rabès, Allard, Ouguerram, Devillers, et al., 2003) opened a new scenario in the lipid field. In fact, whereas previously statins were considered the best hypocholesterolemic drugs, subsequently PCSK9 inhibitors became another therapeutic option. In this context, numerous studies have focused their attention on mAbs (Everett, Smith, & Hiatt, 2015), gene silencing compounds (Graham, et al., 2007), peptidomimetics (Shan, Pang, Zhang, Murgolo, Lan, & Hedrick, 2008) and berberine (Dong, Li, Singh, Cao, & Liu, 2015). At the best of our knowledge, instead, this is the first paper providing experimental evidence that some peptides deriving from food protein hydrolysis may function as effective inhibitors of the PCSK9-LDLR PPI.

Our investigation was focused on lupin protein, since a very recent paper has shown that in hypercholesterolemic subjects, who had consumed dietary bars containing lupin proteins for a month, the total cholesterol decrease was accompanied by a parallel decrement of circulating PCSK9 (Lammi, Zanoni, Calabresi, & Arnoldi, 2016). Other relevant evidences available before starting the present investigation were the following:
1) peptic and tryptic peptides, obtained digesting a total protein extract from lupin seed with pepsin and trypsin are able to inhibit the interaction of PCSK9 with the LDLR in an in vitro system (Lammi, Zanoni, Calabresi, & Arnoldi, 2016); 2) some of these peptides, identified by HPLC-ESI-MS/MS, are absorbed in a Caco-2 model of the small intestine (Lammi, et al., 2016). These evidences prompted us to investigate whether the absorbed peptides (**P-abs** and **T-abs**) maintained the inhibitory activity of **P-tot** and **T-tot**, i.e. whether at least some bioactive peptides had been absorbed. In this context, the demonstration that the capacity of these peptides to inhibit the PCSK9/LDLR interaction is not impaired by the absorption is certainly a main achievement of this work. The fact that at least some bioactive peptides are absorbed is not obvious, considering that only part of the original peptides were absorbed in the Caco-2 model (Lammi, et al., 2016), since at the end of treatment some were still detectable only in

the AP compartment and other disappeared in both compartments owing to metabolic degradation.

A natural development of this research was the application of in silico tools for potentially sorting out the most bioactive peptides. The first step was the modelling a refined human PCSK9 3D-structure, starting from X-ray available literature data (Shan, Pang, Zhang, Murgolo, Lan, & Hedrick, 2008). Afterward, the peptides 3D-structures were calculated and refined starting from the structures of the proteins to whom they belong. Their conformations are very variable, since three have helix conformations, five γ -coil conformations, one β-hairpin conformation, and six are random coils. Finally, the docking study permitted to estimate their theoretical binding energy values on PCSK9 and therefore to get a classification in which **T9** was the best candidate among tryptic peptides and **P5** the best among peptic peptides. Interestingly, both derive from the hydrolysis of lupin β-conglutin, a 7S vicilin-like storage globulin. **T9** consists of 14 amino acidic residues, whereas **P5** of 10 amino acidic residues.

The following experiments, performed on the synthetic peptides, confirmed that indeed they are able to impair the interaction of PCSK9 with the LDLR. **P5** resulted to be the most active, with an IC_{50} value equal to 1.6 μM, whereas **T9** was much less active, since its IC_{50} falls around 320 μM. This different activity also reflect the ability of both **P5** and **T9** to be absorbed through epithelium barrier, **P5** in fact is much more easily absorbed than **T9** which could be subjected to a metabolic degradation. The experiment on the ability of HepG2 cells to uptake LDL from the extracellular environment confirmed the better efficacy of **P5**.

Our calculations permitted to sort out at least one lupin peptide with a great potential interest as inhibitor of the interaction of PCSK9 with the LDLR, owing to the very low IC_{50} value equal to 1.6 μM. It is useful to observe that Pep2-8 (Ac-TVFTSWEEYLDWV-amide), the best inhibitor singled out by Zhang and co-workers in a recent paper (Y. Zhang, Eigenbrot, Zhou, Shia, Li, Quan, et al., 2014), has an IC_{50} value equal to 0.81 ± 0.08 μM, i.e. only slightly better than that of **P5**. Interestingly, Pep2-8 contains 13 amino acid residues and its C-terminal truncated analogues loose activity, whereas **P5** contains only 10 residues. **P5** is thus one of the shortest peptides ever described in literature endowed with this specific activity.

Finally, it is important to recapitulate the very original approach of our investigation:

1) Detection of a PCSK9 decrease in the plasma of hypercholesterolemic subjects at the end of a dietary intervention study based on dietary bars containing lupin protein.
2) Demonstration that lupin peptides reduce the PCSK9 synthesis and secretion in HepG2 cells;

3) Mass spectrometry based identification of lupin peptides which are absorbed in Caco-2 cells;
4) Demonstration that they retain the original capacity of impairing the PCSK9/LDLR interaction;
5) Classification of their potential activity by docking studies with the PCSK9-LDLR complex and selection of the best candidates;
6) Identification of the very potent peptide **P5**;
7) Development and optimization of an efficient workflow for their quantification in Caco-2 cells;
8) Owing to the unusual process of the identification and its original sequence, indeed **P5** may be a promising candidate for designing new potent inhibitors of the PCSK9/LDLR PPI.

In conclusion, this is the first experimental evidence that a peptide deriving from a plant protein has this specific biological activity.

References

Abifadel, M., Varret, M., Rabès, J. P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derré, A., Villéger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J. M., Luc, G., Moulin, P., Weissenbach, J., Prat, A., Krempf, M., Junien, C., Seidah, N. G., & Boileau, C. (2003). Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet, 34*(2), 154-156.

Ambegaonkar, B. M., Tipping, D., Polis, A. B., Tomassini, J. E., & Tershakovec, A. M. (2014). Achieving goal lipid levels with ezetimibe plus statin add-on or switch therapy compared with doubling the statin dose. A pooled analysis. *Atherosclerosis, 237*(2), 829-837.

Brown, M. S., & Goldstein, J. L. (2009). Cholesterol feedback: from Schoenheimer's bottle to Scap's MELADL (vol 50, pg S15, 2009). *Journal of Lipid Research, 50*(6), 1255-1255.

Careskey, H. E., Davis, R. A., Alborn, W. E., Troutt, J. S., Cao, G., & Konrad, R. J. (2008). Atorvastatin increases human serum levels of proprotein convertase subtilisin/kexin type 9. *J Lipid Res, 49*(2), 394-398.

Dong, B., Li, H., Singh, A. B., Cao, A. Q., & Liu, J. W. (2015). Inhibition of PCSK9 Transcription by Berberine Involves Down-regulation of Hepatic HNF1 alpha Protein Expression through the Ubiquitin-Proteasome Degradation Pathway. *Journal of Biological Chemistry, 290*(7), 4047-4058.

Dubuc, G., Chamberland, A., Wassef, H., Davignon, J., Seidah, N. G., Bernier, L., & Prat, A. (2004). Statins upregulate PCSK9, the gene encoding the proprotein convertase neural apoptosis-regulated convertase-1 implicated in familial hypercholesterolemia. *Arterioscler Thromb Vasc Biol, 24*(8), 1454-1459.

Everett, B. M., Smith, R. J., & Hiatt, W. R. (2015). Reducing LDL with PCSK9 Inhibitors--The Clinical Benefit of Lipid Drugs. *N Engl J Med, 373*(17), 1588-1591.

Gencer, B., Lambert, G., & Mach, F. (2015). PCSK9 inhibitors. *Swiss Med Wkly, 145*, w14094.

Graham, M. J., Lemonidis, K. M., Whipple, C. P., Subramaniam, A., Monia, B. P., Crooke, S. T., & Crooke, R. M. (2007). Antisense inhibition of proprotein convertase subtilisin/kexin type 9 reduces serum LDL in hyperlipidemic mice. *J Lipid Res, 48*(4), 763-767.

Lammi, C., Aiello, G., Vistoli, G., Zanoni, C., Arnoldi, A., Sambuy, Y., Ferruzza, S., & Ranaldi, G. (2016). A multidisciplinary investigation on the bioavailability and activity of peptides from lupin protein. *Journal of Functional Foods, 24*, 297-306.

Lammi, C., Zanoni, C., Aiello, G., Arnoldi, A., & Grazioso, G. (2016). Lupin Peptides Modulate the Protein-Protein Interaction of PCSK9 with the Low Density Lipoprotein Receptor in HepG2 Cells. *Sci Rep, 6*, 29931.

Lammi, C., Zanoni, C., Calabresi, L., & Arnoldi, A. (2016). Lupin protein exerts cholesterol-lowering effects targeting PCSK9: From clinical evidences to elucidation of the in vitro molecular mechanism using HepG2 cells. *Journal of Functional Foods, 23*, 230-240.

Lammi, C., Zanoni, C., Scigliuolo, G. M., D'Amato, A., & Arnoldi, A. (2014). Lupin peptides lower low-density lipoprotein (LDL) cholesterol through an up-regulation of the LDL receptor/sterol regulatory element binding protein 2 (SREBP-2) pathway at HepG2 cell line. *J Agric Food Chem, 62*(29), 7151-7159.

Ross, N. T., Katt, W. P., & Hamilton, A. D. (2010). Synthetic mimetics of protein secondary structure domains. *Philos Trans A Math Phys Eng Sci, 368*(1914), 989-1008.

Seidah, N. G., & Prat, A. (2007). The proprotein convertases are potential targets in the treatment of dyslipidemia. *J Mol Med (Berl), 85*(7), 685-696.

Seidah, N. G., & Prat, A. (2012). The biology and therapeutic targeting of the proprotein convertases. *Nat Rev Drug Discov, 11*(5), 367-383.

Shan, L. X., Pang, L., Zhang, R. M., Murgolo, N. J., Lan, H., & Hedrick, J. A. (2008). PCSK9 binds to multiple receptors and can be functionally inhibited by an EGF-A peptide. *Biochemical and Biophysical Research Communications, 375*(1), 69-73.

Welder, G., Zineh, I., Pacanowski, M. A., Troutt, J. S., Cao, G., & Konrad, R. J. (2010). High-dose atorvastatin causes a rapid sustained increase in human serum PCSK9 and disrupts its correlation with LDL cholesterol. *J Lipid Res, 51*(9), 2714-2721.

Wong, N. D., Chuang, J., Wong, K., Pham, A., Neff, D., & Marrett, E. (2013). Residual dyslipidemia among United States adults treated with lipid modifying therapy (data from National Health and Nutrition Examination Survey 2009-2010). *Am J Cardiol, 112*(3), 373-379.

Zhang, D. W., Lagace, T. A., Garuti, R., Zhao, Z., McDonald, M., Horton, J. D., Cohen, J. C., & Hobbs, H. H. (2007). Binding of proprotein convertase subtilisin/kexin type 9 to epidermal growth factor-like repeat A of low density lipoprotein receptor decreases receptor recycling and increases degradation. *J Biol Chem, 282*(25), 18602-18612.

Zhang, Y., Eigenbrot, C., Zhou, L., Shia, S., Li, W., Quan, C., Tom, J., Moran, P., Di Lello, P., Skelton, N. J., Kong-Beltran, M., Peterson, A., & Kirchhofer, D. (2014). Identification of a small peptide that inhibits PCSK9 protein binding to the low density lipoprotein receptor. *J Biol Chem, 289*(2), 942-955.

CHAPTER 5

MANUSCRIPT 3

ABSORPTION AND METABOLISM OF SOY PEPTIDES ACROSS CACO-2 CELL MONOLAYERS

Gilda Aiello,[1] Giulia Ranaldi,[2] Chiara Zanoni,[1] Simonetta Ferruzza,[2] Yula Sambuy,[2] Anna Arnoldi,[1]* Carmen Lammi[1]

[1]Department of Pharmaceutical Sciences, University of Milan, 20133 Milan, Italy

[2]CREA, Food and Nutrition Research Centre, Rome, Italy

5.0 Abstract

IAVPGEVA, IAVPTGVA, and LPYP, three peptides deriving from glycinin, a main fraction of soy globulins, are known to inhibit the activity of 3-hydroxy-3-methylglutaryl CoA reductase and modulate cholesterol metabolism in HepG2 cells. In order to evaluate whether these peptides can be transported across the epithelium barrier, experiments using human intestinal Caco-2 cell monolayers grown in two-chambers systems were performed. The apical chamber was treated with each single peptide for a time spanning from 15 up to 120 min. The development of a highly sensitive method based on LC-MRM permitted to quantify the peptides in the basolateral chambers and to calculate the apparent permeability coefficients (P_{app}) that were equal to $3.34 \pm 0.06 \times 10^{-8}$, $2.75 \pm 0.08 \times 10^{-8}$, and $1.41 \pm 0.03 \times 10^{-8}$ cm sec^{-1}, for IAVPGEVA, IAVPTGVA, and LPYP, respectively. At the end of experiment, a detailed characterization of the metabolites was also performed either in the apical or in the basolateral chamber.

Keywords: bioactive peptides, Caco-2 cells, HMGCoA reductase, intestinal absorption, mass spectrometry, plant protein, soybean.

Abbreviations used: AP, apical; **BL**, basolateral; **HMGCoAR**, 3-hydroxy-3-methylglutaryl coenzyme A reductase; **LDL**, low density lipoprotein, **LC**, liquid chromatography; **MRM**, multiple reaction monitoring; **TEER**, trans-epithelial electrical resistance.

5.1 Introduction

Cardiovascular diseases are among the main causes of death and disability and hypercholesterolemia is a crucial feature that increases the risk of adverse events. Although drugs, with particular reference to statins, remain the main tool for lowering high cholesterol levels, some functional foods may have a useful preventive role in the presence of a moderate hypercholesterolemia.[1-3] In particular, numerous studies have shown that soybean protein, among the main bioactive components of this bean, is able to reduce total and low-density lipoprotein (LDL) cholesterol levels in highly and mildly hypercholesterolemic subjects,[1, 4, 5] through a positive effect on LDL receptors (LDLR). [6]

The two major components of soybean protein are glycinin, a non-glycosylated 11S globulin (legumin), and β-conglycinin, a glycosylated 7S globulin (vicilin). Since proteins are hydrolyzed during gastrointestinal digestion, some researches have been dedicated to investigate the activity of soy protein hydrolysates. In this context, two peptides, LPYP and IAVPGEVA, have been isolated and characterized after digesting glycinin with trypsin and pepsin, respectively, and were shown to inhibit the activity of 3-hydroxy-3-methylglutaryl CoA reductase (HMGCoAR), a main enzyme in cholesterol biosynthesis.[7] Subsequently, the alignment of IAVPGEVA with the glycinin sequence permitted the identification of another bioactive peptide, i.e. IAVPTGVA.[7] Kinetic experiments showed that these peptides act as competitive inhibitors of HMGCoAR.[8] With a pharmaceutical approach, the same authors have exploited these peptides as leads for the synthesis of more active analogues,[9] without examining their biological effects in detail.

In order to fill this gap, recently our group has performed a series of experiments aimed at identifying the molecular mechanism at the basis of their hypocholesterolemic properties. These studies have confirmed that IAVPGEVA, IAVPTGVA, and LPYP are able to inhibit the activity of HMGCoAR and have demonstrated that the inhibition of this enzyme leads to a favorable modulation of LDLR pathway in human hepatic HepG2 cells. Indeed, the treatment of these cells with each peptide at the concentration of 500 μM, induced the activation of the sterol regulatory element binding proteins 2 (SREBP-2), which lead to an increase of the LDLR

protein levels. Moreover, the up-regulation of the LDLR enhances the ability of HepG2 cells to uptake LDL particles from the extracellular environment producing a final hypocholesterolemic effect.[9] In the meanwhile, another investigation has shown that the same peptides also exert hypoglycemic effects in HepG2 cells.[10]

Intestinal permeability plays a critical role in the oral bioavailability of any active compound. Several studies have investigated the transepithelial transport of antihypertensive peptides derived from different food sources, such as egg or milk.[11-14] A single transport study has investigated the transport across human intestinal Caco-2 cell monolayers of peptides generated by digestion of soy protein.[15] YVVNPDNDEN and YVVNPDNNEN were among the 22 peptides deriving from β-conglycin transferred from the apical (AP) to the basolateral (BL) compartment in this experiment. Interestingly, these peptides induce hypocholesterolemic effects in HepG2 cells by a statin-like mechanism.[16]

Whereas numerous published works are focused on the qualitative identification of absorbable peptides, we chose a quantitative approach combined advanced analytical techniques, such as microfluidic HPLC and targeted mass spectrometry (MS). In particular, the use of multiple reaction monitoring (MRM), coupled with spiking with an internal standard, enable the fast development of highly specific, precise, and robust quantitative analyses.[17-19] Indeed, MRM has emerged as an important analytical tool in pre-clinical and clinical small-molecule drug development and monitoring as well as in nutritional investigations.[20]

The detailed objectives of the present work were the development of a sensitive LC-MRM method for the quantitation of IAVPGEVA, IAVPTGVA, and LPYP, the evaluation of the transepithelial transport of these peptides across the Caco-2 monolayer, and the investigation of their degradation in Caco-2 cells. These absorption data, as well as previously obtained functional data,[16] will be useful for the translation of these findings from *in vitro* systems to *in vivo* models.

5.2 Materials and Methods

5.2.1 Chemicals and reagents. IAVPGEVA, IAVPTGVA, LPYP, and YVVNPDNDEN were synthesized (> 95% purity by HPLC) by PRIMM (Milan, Italy). Dulbecco's Modified Eagle Medium (DMEM) was from GIBCO (Thermo Fisher Scientific, Waltham, MA USA). Fetal Bovine Serum (FBS) was from Hyclone Laboratories (Logan, UT, USA). Stable L-glutamine, 1% non-essential amino acids, and penicillin/streptomycin were from Euroclone (Milan, Italy). Formic acid and acetonitrile (ACN) were from Sigma-Aldrich (Milan, Italy). LC-grade H_2O

(18 MΩ cm) was prepared with a Milli-Q H_2O purification system (Millipore, Bedford, MA, USA).

5.2.2 Cell cultures. Caco-2 cells, obtained from INSERM (Paris), were sub-cultured at low density.[21] Cells were routinely sub-cultured at 50% density and maintained at 37 °C in a 90% / 10% air / CO_2 atmosphere in DMEM containing 25 mM glucose, 3.7 g/L $NaHCO_3$, 4 mM stable L-glutamine, 1% non-essential amino acids, 100 U/L penicillin, 100 µg/L streptomycin (complete medium), supplemented with 10% heat inactivated fetal bovine serum (FBS) (Hyclone Laboratories, Logan, UT, US). For differentiation, cells were seeded on polycarbonate filters, 12 mm diameter, 0.4 µm pore diameter (Transwell, Corning Inc., Lowell, MA, US) at a $3.5x10^5$ cells/cm^2 density in complete medium supplemented with 10% FBS in both AP and BL compartments for 2 days to allow the formation of a confluent cell monolayer. Starting from day 3 after seeding, cells were transferred to complete medium in both compartments, supplemented with 10% FBS only in the BL compartment, and allowed to differentiate for 15 days with regular medium changes three times weekly.[22]

5.2.3 Cell monolayer integrity evaluation. The trans-epithelial electrical resistance (TEER) was measured at 37 °C using the voltmeter apparatus Millicell (Millipore Co., USA,), immediately before and at the end of the 2 h absorption experiments. After incubations, no significant changes in TEER values were observed. At the end of absorption experiments, cells were incubated from the AP side with 1 mM phenol-red in PBS with Ca^{++} and Mg^{++} for 1 h at 37 °C. The BL solutions were then collected and 10 % in volume of 0.1 N NaOH was added before reading the absorbance at 560 nm by a microplate reader Synergy H1 from Biotek (Winooski, VT, USA). Phenol-red passage was quantified using a standard phenol-red curve. Only filters showing TEER values and phenol red passages similar to untreated control cells were considered for peptide transport analysis.

5.2.4 Trans-epithelial absorption of soy peptides. Prior to experiments, the cell monolayer integrity and differentiation were checked by TEER measurement as described in details above. Cells were then washed twice and peptide absorption assayed. Peptide absorption were performed in transport buffer solution (137 mM NaCl, 5.36 mM KCl, 1.26 mM $CaCl_2$, 1.1 mM $MgCl_2$, 5.5 mM glucose). In order to reproduce the pH conditions existing *in vivo* in small intestinal mucosa, AP solutions were maintained at pH 6.0 (buffered with 10 mM

morpholinoethane sulfonic acid), and BL solutions were maintained at pH 7.4 (buffered with 10 mM N-2-hydroxyethylpiperazine-N-4-butanesulfonic acid). Prior to absorption experiments, cells were washed twice with 500 µL PBS with Ca^{++} and Mg^{++}, and cell were loaded from the AP side with each soy peptide (500 µM) dissolved in AP transport solution (500 µl) and the BL compartment contained BL transport solution (700 µL). The plates were incubated for different time periods, i.e. 15, 30, 60, 90 and 120 min at 37 °C, 5% CO_2. At the end of incubation the incubation time, AP and BL media were collected and stored at -80 °C. Three independent experiments were performed, each in duplicate.

5.2.5 LC/ESI-MRM analysis. An aliquot (10 µL) of each BL and AP solution was spiked with the peptide YVVNPDNDEN as an internal standard (IS) and then desalted with C18 resin ZipTip (Millipore Corporation, Bedford, MA, USA). Each sample was lyophilized under vacuum and re-dissolved in 10 µL (0.1% formic acid), vortexing and centrifuging at 6000 *g* thrice. MRM-based analysis of purified peptides was carried out on a SL IT mass spectrometer interfaced with a HPLC Chip Cube source (Agilent Technologies, Palo Alto, CA, USA). Data were processed with MSD Trap control 4.2, and Data analysis 4.2 version (Agilent Technologies). The chromatographic separation was performed using a 1200 HPLC system equipped with a binary pump. The peptide enrichment was performed on a 40 nL enrichment column (Zorbax 300SB-C18, 5 µm pore size), followed by separation on a 43 mm × 75 µm analytical column packed (Zorbax 300SB-C18, 5 µm pore size). Solvent A was 95% LC water, 5% ACN, 0.1% formic acid; solvent B was 5% LC water, 95% ACN, 0.1% formic acid. Peptide enrichment was performed injecting 1 µL at the constant flow rate of 4 µL/min, while the analytical separation was at 300 nL/min. The nanopump gradient program were as follows: IAVPGEVA analysis, 3% solvent B (0 min), 80% solvent B (0–4 min), and back to 5% in 1 min; IAVPTGVA analysis, 3% solvent B (0 min), 60% solvent B (0–6 min), and back to 5% in 2 min; LPYP analysis, 5% solvent B (0 min), 80% solvent B (0–5 min), and back to 5% in 1 min. A post time acquisition of 5 minutes at the initial chromatographic conditions was used after each run. Data acquisition was accomplished in positive ionization mode. Capillary voltage was −1950 V, with endplate offset −500 V, skimmer -40 V, drying gas flow 5 L/min, drying gas temperature 300 °C, nebulizer gas pressure 18 psi, scan range m/z 150–2000, averages of five spectra, ion charge control (ICC) target 30000. The MRM acquisition was set in two segments: IAVPGEVA analysis, the former (0-2.5 min) was used for the IS at *m/z* 589.7 and the latter (2.5 – 5.0 min) for the target peptide precursor ion at *m/z* 755.6; IAVPTGVA analysis, same scheduled time intervals to monitor the IS at *m/z* 589.7 and the target peptide

precursor ion at *m/z* 727.2; LPYP analysis, the former from (0-3.25 min) to monitor the IS at *m/z* 589.7 and the latter (3.25-5.0 min) to monitor the target peptide precursor ion at *m/z* 489.3. A blank was analyzed between every sample to ensure absence of carryover effects. The peptides were identified by comparing their retention times, MS profile and MS/MS fragmentation spectra with those of authentic standards.

5.2.6 Calibration curves and quantification of soy peptides. The standard peptides (IAVPGEVA, IAVPTGVA, and LPYP) in 0.1% formic acid were used to build a calibration curve for each peptide using YVVNPDNDEN as IS. In case of IAVPGEVA, the calibration curve was constructed over a concentration range from 5 to 250 ng/mL adding 37.5 ng/mL of IS; in case of IAVPTGVA, it was constructed over a concentration range from 5 to 1000 ng/mL adding 250 ng/mL of IS, and in case of LPYP over a concentration range from 0.5 to 100 ng/mL adding 37.5 ng/mL of IS. The ratios between the areas of each peptide and the IS were calculated and plotted versus the concentration ratios to build the calibration curves. The equations for the curves were calculated using six calibration points with three set of replicate ($n = 3$) per curve. The data analysis for the standard curves were performed by exporting the peak area ratios from LC/MSD Trap Software 5.3 into the automated software QuantAnalysis (version 1.7).

5.2.7 Accuracy and precision evaluation. To ensure appropriate performance of the developed method, quality control samples were prepared by spiking the standard peptides in a basolateral sample originating from control Caco-2 cells at the concentrations of 0.1, 0.15, 0.1 ng/µL for IAVPGEVA, IAVPTGVA, and LPYP, respectively, followed by sample purification according to the protocol previously described. The IS was spiked into each sample prior to ZipTip desalting. Results were calculated for three independent samples, analyzing each one three times.

5.2.8 Apparent permeability measurement. The apparent permeability coefficients (P_{app}, cm s^{-1}) of the peptides under investigation were calculated using an equation proposed by Stuknyte and coworkers:[23] $P_{app} = dQ/dT \times 1/(S \times C_0)$, where dQ/dT is the amount of the each peptide measured in the acceptor BL chamber as a function of time (µmol sec^{-1}), S is the surface area of the monolayer (here 1.12 cm^2), and C_0 is the initial concentration in the donor AP chamber (µM).

5.2.9 Statistical analysis. All analyses were run in triplicate on each biological replicate. Statistical analysis, including determination of linear regression, average, standard deviation (SD) and coefficient of variance (CV), was performed using Graphpad Prism 6. Values were expressed as means ± SEM.

5.3 Results

5.3.1 Optimization of a HPLC-MRM method for the quantitation of the soy peptides. The first step of the work consisted in the validation of the analytical method based on chip-HPLC ESI-MS/MS. The chromatographic separation was conducted by optimizing a gradient for each peptide. The HPLC–MS/MS analyses, under MRM conditions, were carried out monitoring two most intense diagnostic transitions for each peptide in order to assure the greatest specificity. In details, the MRM transitions for IAVPGEVA were those from the mono-charged precursor ion $[M+H]^+$ (*m/z* 755.6) to the product-ions ***y5*** and ***b7*** with *m/z* 472.1 and 666.3, respectively. The MRM transitions for IAVPTGVA were from the mono-charged precursor ion $[M+H]^+$ (*m/z* 727.4) to the product-ions ***y5*** and ***b7*** with *m/z* 444.1 and 638.3, respectively. The transition of LPYP were from the mono-charged precursor ion $[M+H]^+$ (*m/z* 489.3) to the product-ions ***y2*** and ***b3*** with *m/z* 278.9 and 374.0, respectively. Similarly, the MRM transitions for YVVNPDNDEN were from the double-charged precursor ion $[M+2H]^{2+}$ (*m/z* 589.7) to the single-charged product-ions ***y7*** and ***b3*** with *m/z* 817.3 and 362.0, respectively. The assay was optimized by dividing the methods into several time segments and the specific MRM transitions are monitored in each segmentation according to the corresponding retention time. **Figure 5.1** displays the scheduled chromatograms of each monitored peptides. **Figure 5.2** shows the product ion spectra of IAVPGEVA, IAVPTGVA, LPYP, and YVVNPDNDEN (used as IS).

While building the calibration curves, the peak areas of all monitored transitions from parent to product ions of each peptide were normalized by the peak area of the corresponding MRM transitions of the IS.

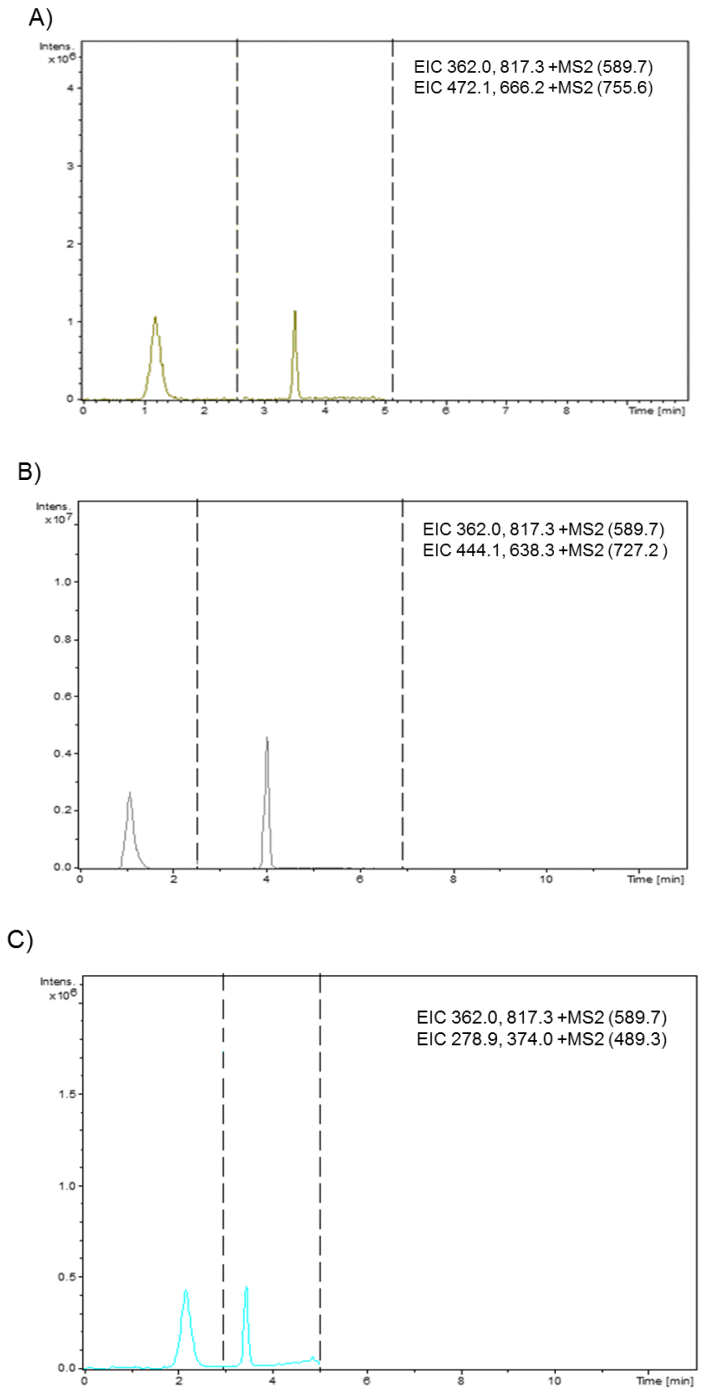


Figure 5.1 - Extract ion chromatograms (EICs) of three soy peptides. A) EIC of IS and IAVPGEVA (t_R 1.25 min, and 3.5 min for IS and IAVPGEVA, respectively), **B)** EIC of IS and IAVPTGVA (t_R 1.15 min, and 4.0 min for IS and IAVPTGVA, respectively), **C)** EIC of IS and LPYP (t_R 2.2 min, and 3.45 min for IS and LPYP, respectively).

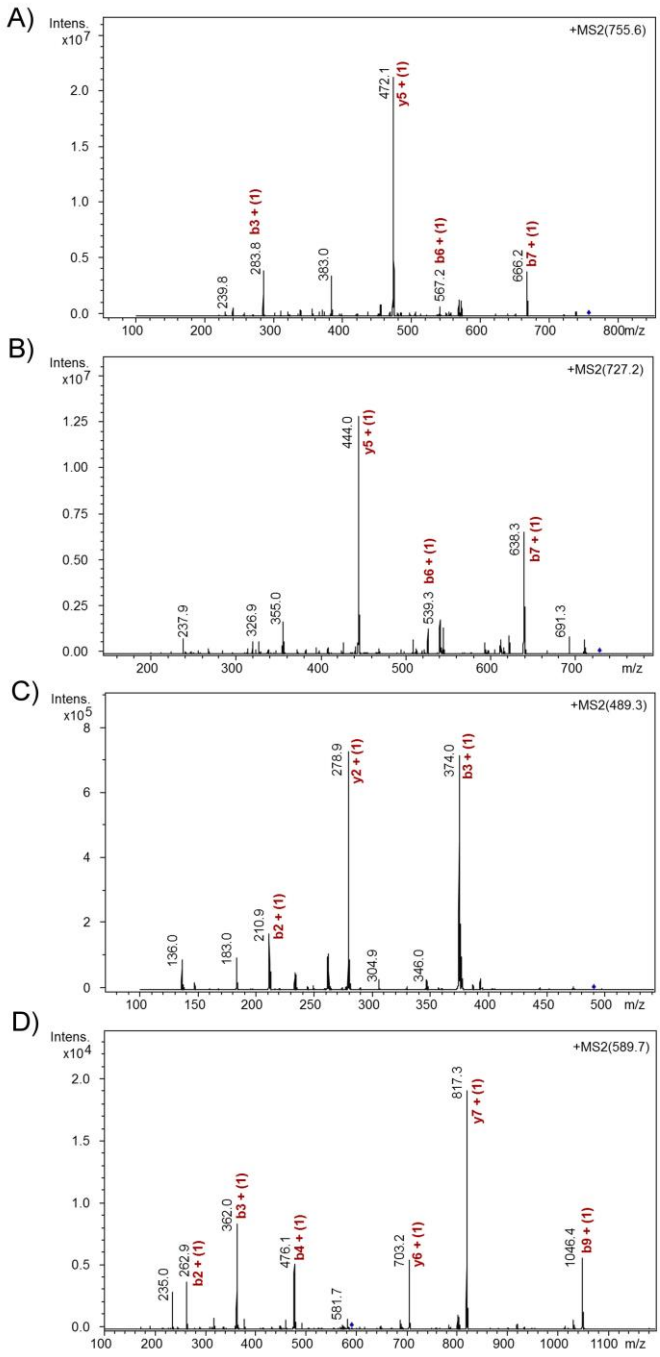


Figure 5.2 – Soy peptide fragment ion spectra. ESI (+) MS/MS product ion spectra of A) IAVPGEVA; B) IAVPTGVA; C) LPYP; and D) IS.

5.3.2 Evaluation of the analytical parameters. The validation parameters of the established LC-MRM method are shown in **Table 5.1**, which reports the concentration ranges, the correlation coefficients, the limits of detection (LOD) and the limits of quantitation (LOQ) for each peptide. The calibration curve was linear over the concentration range from 5 to 250 ng/mL for IAVPGEVA, over the concentration range from 5 to 1000 ng/mL for IAVPTGVA,

and over the concentration range from 0.5 to 100 ng/mL for LPYP. **Figure 5.3** shows the calibration curve of each peptide used for the quantitative evaluation. **Table 5.1S** in Supplementary Information to this work, reports the data on which the curves were built. The correlation coefficient (R^2) was greater than 0.99 for each calibration curve, indicating good linearity.

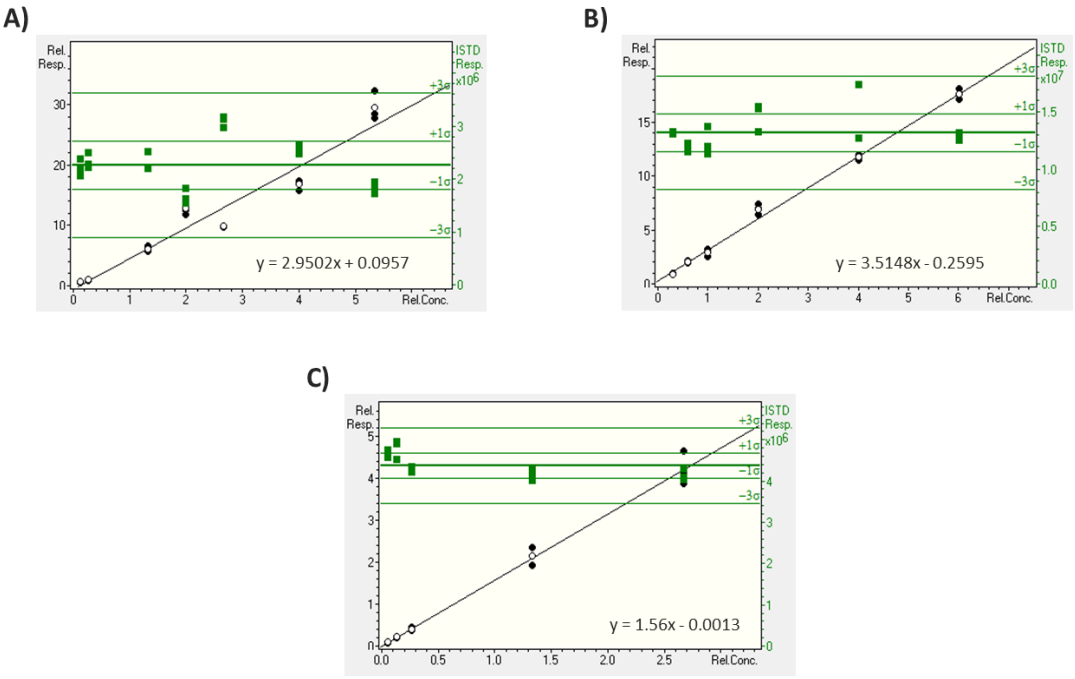


Figure 5.3 – Calibration curves. Plot **A**, **B, C** show the reference peptide curves with the linear equations for IAVPTGVA, IAVPGEVA and LPYP respectively.

The sensitivity of the method was assessed by the LOQ's [signal-to-noise (S/N) =10], which were 3.28 ng/mL, 3.06 ng/mL, and 0.36 ng/mL for IAVPGEVA, IAVPTGVA and LPYP, respectively, and the LOD's (S/N = 3), which were 3.20 ng/mL, 2.94 ng/mL, and 0.22 ng/mL, respectively. The accuracy of the method was evaluated by the analysis of quality control samples prepared as reported in the Material and Method section. Accuracy of 101.4%, 108.4% and 97.0% were obtained for IAVPGEVA, IAVPTGVA and LPYP, respectively, whereas the coefficient of variations (CV%), representing the precision of the LC-MRM method, evaluated by 3 independent analyses, were 4.6 %, 4.2 % and 1.9 %, respectively, reported in **Table 5.1.**

Table 5.1 Calibration range, linearity (R^2) and LOQ-LOD of the LC–ESI-MS/MS method

Peptide	Calibration Range ng/mL	IS ng/mL	Linearity (R^2)[a]	LOQ-LOD ng/mL	CV%	Accuracy%
IAVPTGVA	5 - 1000	37.5	0.993	3.06 - 2.94	4.2	108.4
IAVPGEVA	5 - 250	37.5	0.997	3.28 - 3.20	4.6	101.4
LPYP	0.5 - 100	37.5	0.999	0.36 - 0.22	1.9	97.0

a Correlation coefficient using seven calibration points (n = 3)

5.3.3 Permeation experiments and calculation of apparent permeability coefficients. In order to evaluate the absorption, differentiated Caco-2 cells were incubated adding the AP compartment with each soy peptide (500 µM). The BL media were collected after 15, 30, 60, 90 and 120 min treatments and submitted to LC-MRM analysis, whereas the AP media were analyzed after 120 min. For monitoring cell monolayer integrity and excluding non-specific peptide passage, the TEER values were monitored at the beginning and end of each experiment. Moreover, phenol-red passage across the monolayer was assayed at the experiment end.[24] Both assays demonstrated that the incubation with soy peptides did not affect the monolayer integrity (data not shown).

Specific MRM transitions based on chromatographic retention times (t_R) and MS/MS spectra were used to detect and quantify the absorbed exogenous peptides.

Figure 5.4 shows the time-dependent concentrations of the peptides quantified in the BL chamber. In case of IAVPGEVA, the concentrations were 18.1 ± 1.40 ng/mL at 15 min, 25.6 ± 0.24 ng/mL at 30 min, 31.7 ± 0.08 ng/mL at 60 min, 36.8 ± 0.01 ng/mL at 90 min, and 45.9 ± 1.35 ng/mL at 120 min. In case of IAVPTGVA, the concentrations were 14.3 ± 0.4 ng/mL at 15 min, 20.6 ± 0.2 ng/mL at 30 min, 25.4 ± 0.4 ng/mL at 60 min, 39.3 ± 0.7 ng/mL at 90 min, and 40.4 ± 0.5 ng/mL at 120 min. In case of LPYP, they were 5.0 ± 0.1 ng/mL at 15 min, 7.0 ± 0.02 ng/mL at 30 min, 11.0 ± 0.05 ng/mL at 60 min, 14.3 ± 0.04 ng/mL at 90 min and 15.1 ± 0.53 ng/mL at 120 min. The final values of the transport percentage at 120 min *versus* the amount of each peptide incubated at time 0 in the AP sides were 0.05% for IAVPTGEVA, 0.02% for IAVPTGVA, and 0.009% for LPYP.

In order to investigate the rate of peptide transport from AP to BL side, the P_{app} values of IAVPGEVA, IAVPTGVA, and LPYP were calculated at 15 min using a literature method.[23] As shown in **Figure 5.5**, the peptide IAVPGEVA had the highest P_{app} value equal to 3.34 ± 0.06 × 10^{-8} cm sec^{-1}, followed by that of IAVPTGVA equal to 2.75 ± 0.08 × 10^{-8} cm sec^{-1}, and LPYP equal to 1.41 ± 0.03 × 10^{-8} cm sec^{-1}.

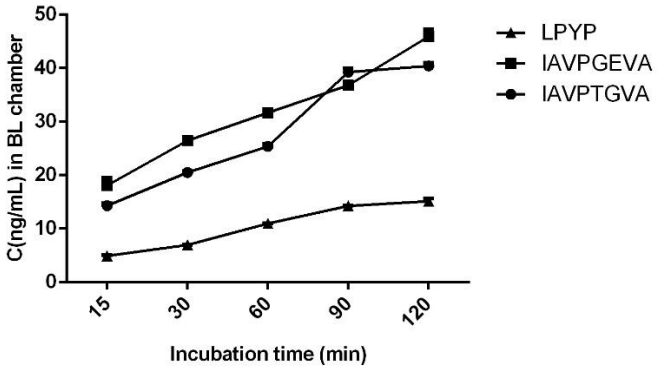


Figure 5.4 - Time-dependent increase of peptide concentrations at BL side. Calculated concentration (ng/mL) of IAVPGEVA, IAVPTGVA, and LPYP upon trans-epithelial absorption across Caco-2 monolayer. Bars represent averages ± SD of three independent experiments (three replicate per sample).

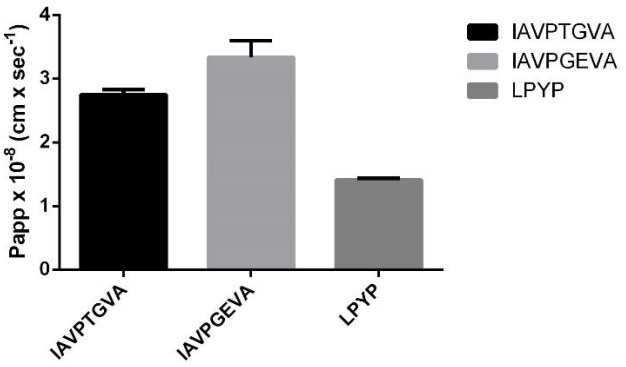


Figure 5.5 - *In vitro* permeability. Bar graphs indicate the P_{app} x 10^{-8} cm sec^{-1} calculated for each peptide. The data points represent the averages ± SEM of three independent experiments in triplicate.

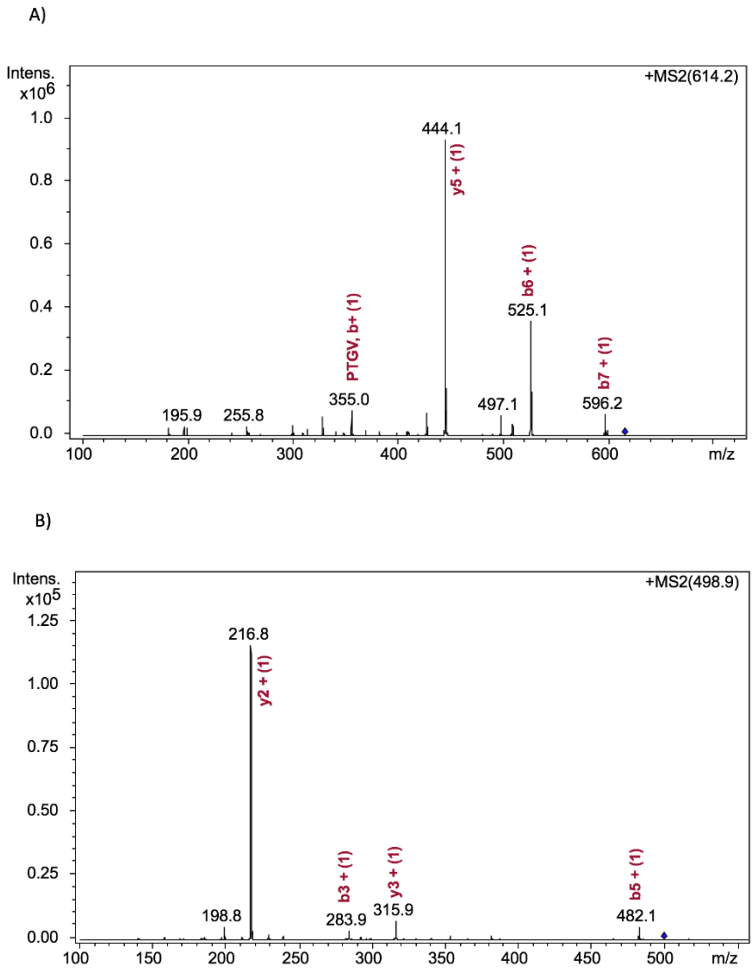
5.3.4 Degradation by brush border peptidases. Under the hypothesis that the low transport rates might be attributed to a competing *in situ* degradation induced by the hydrolytic activity of brush border membrane peptidases, the AP solutions collected at 120 min were analyzed looking for metabolic degradation products. This was possible for IAVPGEVA and IAVPTGVA, whereas in case of LPYP this investigation was impaired by the intrinsic difficulty of obtaining certain attribution of very short unspecific peptides.

The complete sequence coverage assignment of all MS/MS spectra permitted the unambiguous identification of each metabolite (data reported in **Figure 5.6)**.

Figures 5.7A, B display the extracted ion chromatograms of both precursor peptides and their hydrolytic fragments in the AP chamber at 120 min. The susceptibility of these peptides to hydrolysis are very different. In fact, the metabolism of IAVPGEVA produces only one

breakdown fragment (*m/z* 642.3), with a t_R of 1.2 min, deriving from the loss of an isoleucine at the N-terminal. On the opposite, IAVPTGVA produces three breakdown fragments: AVPTGVA (*m/z* 614.2), with a t_R of 1.2 min, again deriving from the loss of an isoleucine at the N-terminal position; IAVPT (*m/z* 498.9), with a t_R of 3.65 min, due to loss of a three amino acids fragment at the C-terminal position; and IAVP (*m/z* 398.7), with t_R = 3.8 min, deriving from the loss of a four amino acids fragment at the C-terminal position. Since the percentages of intact parent peptide were only 19.2% for IAVPGEVA and 12.9% for IAVPTGVA, both peptides had been significantly hydrolyzed by brush border peptidases.

The analyses of the metabolites in the BL solutions at 120 min (**Figures 5.4C, D**) clearly indicated that the breakdown fragments observed in the AP solution had been translocated across the Caco-2 monolayer.



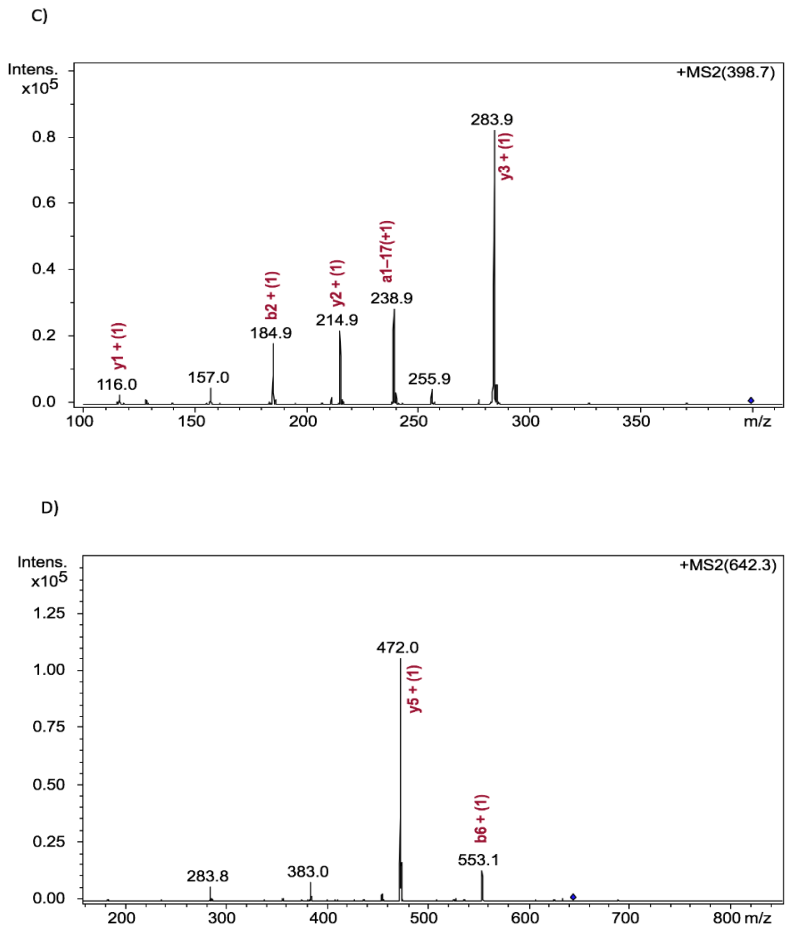


Figure 5.6 - **MS/MS spectra of metabolites.** ESI (+) MS/MS spectrum of metabolite **A)** AVPTGVA (*m/z* 614.2), **B)** IAVPT (*m/z* 498.9) **C)** IAVP (*m/z* 398.7) and **D)** AVPGEVA (*m/z* 642). The sequence ions and internal fragments are indicated in red in the spectrum.

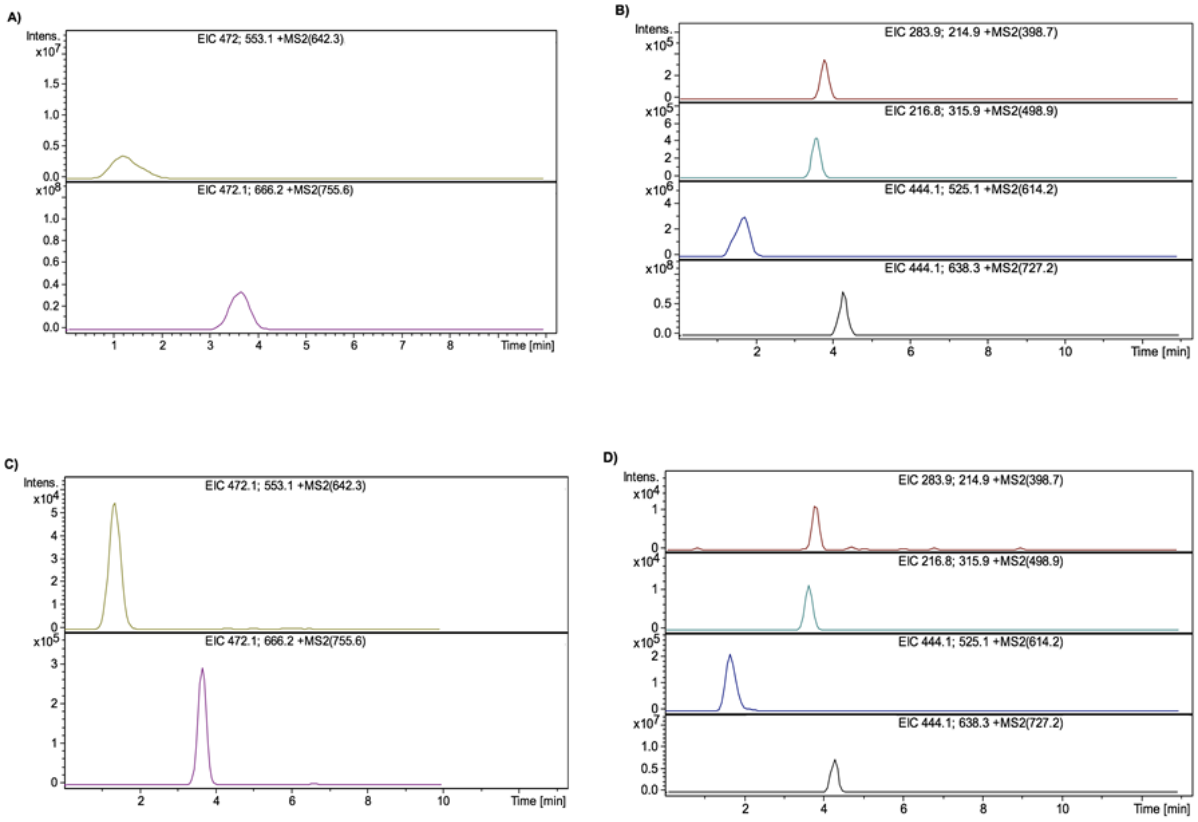


Figure 5.7 - AP and BL Extracted ion chromatogram (EIC) of precursor peptides and hydrolytic fragments after 120 min. A) EIC of AVPGEVA (*m/z* 642. 3) and IAVPGEVA (*m/z* 755.6) and **B)** EIC of IAVP (*m/z* 398.7), IAVPT (*m/z* 498.9), AVPTGVA (*m/z* 614.2) and IAVPTGVA (*m/z* 727.2) at AP side. **C)** EIC of AVPGEVA (*m/z* 642. 3) and IAVPGEVA (*m/z* 755.6) and **D)** EIC of IAVP (*m/z* 398.7), IAVPT (*m/z* 498.9), AVPTGVA (*m/z* 614.2) and IAVPTGVA (*m/z* 727.2) at BL side.

5.3.5 Comparative stability of breakdown fragments through Caco-2 monolayers. With the purpose of a better comprehension of the degradation process, the relative abundances of the parent peptide and its metabolites in the AP solution at 120 min were calculated comparing the Peak Intensity of each peptide. In the case of IAVPGEVA, the parent peptide was equal to 86.5% and the unique breakdown fragment AVPGEVA was 13.5%. In the case of IAVPTGVA, the parent peptide accounted for 91.4%, the most abundant breakdown fragment AVPTGVA accounted for 6.9 %, IAVPT accounted for 1.1%, and IAVP for 0.6 %.

Applying the same considerations to the BL solutions, in case of IAVPGEVA, the percentage of the parent peptide was 82.5% and that of the unique metabolite AVPTGVA was 17.5%, whereas in case of IAVPTGVA, the parent peptide was 83.8% and the breakdown fragments were AVPTGVA equal to 5.8%, IAVP equal to 0.24%, and IAVPT equal to 0.21%.

These results show that while they were formed, the metabolites were also transported at least in part across the cell monolayer. **Figures 5.8A, B** compare the relative abundance of the parent

peptides and their fragments in the AP and BL chambers for IAVPGEVA and IAVPTGVA, respectively. Interestingly, whereas the relative abundance of the fragment AVPTGVA in AP and BL chambers are very similar (6.9% and 5.8%, respectively), the smaller hydrolytic fragments IAVPT and IAVP are much less abundant in the BL chamber than in the AP chamber. This may possibly indicate that the transport of IAVPT and IAVP are slower than that of AVPTGVA.

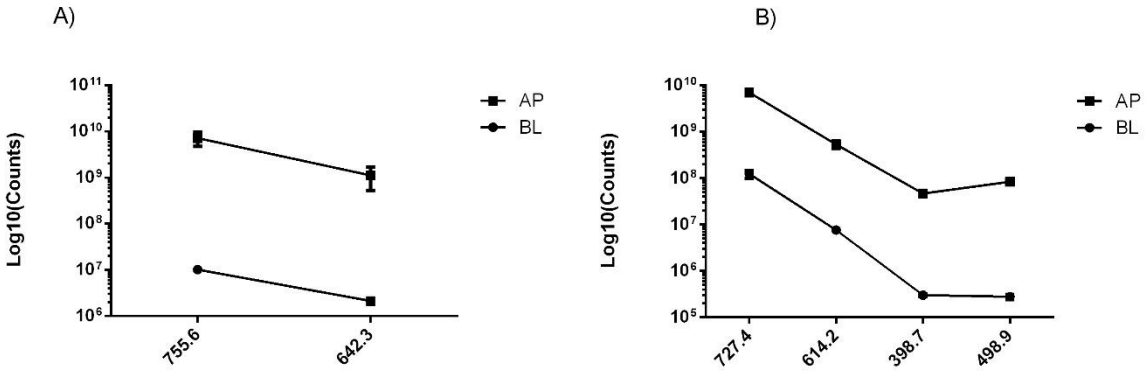


Figure 5.8 - Peptide relative abundance. Peptide spectral intensity of **A)** IAVPGEVA and its fragment AVPGEVA at AP and BL side; **B)** IAVPTGVA and IAVP, IAVPT, AVPTGVA at AP and BL side. The data points represent the averages ± SD of three independent experiments in triplicate.

5.4 Discussion

The intestinal absorption is a critical issue for the bioavailability of any active compound. The intestinal epithelium is composed of a large number of well-differentiated and polarized epithelial cells with microvilli structure on the AP surface.[25] Because of their similarity to intestinal epithelium, Caco-2 cell monolayers are usually selected as cellular model for transport studies. Using this model, it is possible to evaluate the transport of a bioactive compound from the AP side of the cell monolayer, mimicking the lumen, to the BL side, mimicking the circulation flux. The Caco-2 system seems to represent a good model for screening and ranking bioactive peptides with respect to their absorbability.[26]

The soy peptides investigated in this work have interesting hypocholesterolemic[9] and hypoglycemic activities.[10] Our objective was to verify whether they are able to be absorbed by the intestinal model and to obtain quantitative data on the transport across the Caco-2 cell monolayer. For this reason, a selective and specific LC-MRM method was developed that permitted to build a linear regression curve for the quantification of each peptide. Indeed these

soy peptides are transported across the Caco-2 monolayer with a linear trend of absorption during the 120 min of the experiment as clearly suggested by **Figure 5.4**.

Following the biopharmaceutical classification system,[27] the absorption is considered very poor when the P*app* value is lower than 0.2×10^{-6} cm s^{-1}, whereas it is moderate when the P*app* value falls between 0.2 and 2.0×10^{-6} cm s^{-1}, and good when the P*app* value is larger than 2.0×10^{-6} cm s^{-1}. Our data indicate that the absorption of IAVPTGVA, IAVPGEVA, and LPYP is low, since the order of magnitude in all P_{app} values were in the range of 10^{-8} cm sec^{-1}. These values are in accordance to recent transepithelial transport studies of food derived peptides performed in the same model showing very low apparent permeability P_{app} ranging from 10^{-6} to 10^{-9} cm sec^{-1}.[28, 29] The P_{app} values of IAVPGEVA and IAVPTGVA were very similar, being 3.34 ± 0.06 and $2.75 \pm 0.08 \times 10^{-8}$ cm sec^{-1}, respectively; whereas the P_{app} of LPYP resulted to be slower being equal to $1.41 \pm 0.03 \times 10^{-8}$ cm sec^{-1}. The presence of two proline residues may most likely hinder its passage through the epithelium.[30]

Compared to the values suggested by biopharmaceutical classification system, these P_{app} values appear to be rather low. The permeability may be influenced by at least three independent factors: 1) the higher relative tightness of the Caco-2 cell monolayer compared to other more relevant physiological systems able to mimic the intestinal barrier; 2) the cytosolic degradation by intracellular peptidases during the transport across the Caco-2 monolayer; 3) the equilibrium between the absorption process and the degradation of intact peptides by the active endo-peptidases localized on the cellular brush-border.

The transepithelial transport of short-chain peptides has been already demonstrated to be very low in several *in vitro* studies,[28] suggesting that their efflux across the BL membrane in very limited. However, the comparison of the P_{app} values obtained in different models of intestinal absorption points toward an underestimation of P_{app} in the Caco-2 model due to a higher tightness of the Caco-2 cell monolayer compared to the intact mammalian intestinal tissue.[31] Moreover, the variation among the cell-lines used in different labs, the different passage number, protocols, or cell seeding densities may lead to a different tightness of the monolayers resulting in different permeability.[32] For this reason, the comparison between the P_{app} values of these soy peptides and those of other peptides may be controversial.

Literature reports some cases of peptides with good permeability, such as ovalbumin-derived peptides, i.e. YAEERYPIL and YAEER, whose P_{app} were estimated to be 52×10^{-6} cm sec^{-1} and 84×10^{-6} cm sec^{-1}, respectively.[33]

Here, the transport mechanisms were not investigated, since to achieve this goal an *ex vivo* model in Ussing chambers would have been more suitable than Caco-2 cells.[29] However, published data suggest that molecules with a large molecular weight or size are confined to the transcellular pathway.[34] Indeed, these peptides might be transported by a trans-cytotic route, during which the cytosolic peptidase may partially hydrolyze the intact peptides affecting the permeability in the BL compartments making, therefore, difficult the measurement of the real transport rate of intact peptides.

In this context, due to the low P_{app} of these peptides, the second objective of the work was the evaluation the susceptibility of IAVPTGVA and IAVPGEVA to the activities of peptidases, since it is well known that Caco-2 cells express membrane-bound exo- and endopeptidases.[35] The effect of the metabolic degradation was evaluated at maximum incubation time: a remarkable degradation or cellular internalization was observed, since the residual percentages of intact IAVPGEVA and IAVPTGVA were estimated to be only 19.2% and 12.9%, respectively.

These octapeptides differ only for a two-amino acids sequence in position 5-6, i.e. TG and GE. Our results indicate that this structural diversity greatly influences their susceptibility to the metabolic degradation (**Figure 5.7A**, **B)**. Both undergo the loss of an isoleucine from the N-terminal, due to the activity of an amino-exopeptidase,[36] but at diverse rate, since the metabolite released from IAVPGEVA was two-fold more abundant than the metabolite from IAVPTGVA. Interestingly, only IAVPTGVA appeared to be susceptible to the action of endopeptidases, which resulted in the production of more numerous breakdown fragments. The observed differences depend greatly on their structural characteristics: IAVPGEVA is an acidic peptide, endowed of -1 net charge, whereas the net charge of IAVPTGVA is 0. Both contain predominantly hydrophobic residues and consist of two portions,[37] the first of which is relatively rigid, because comprises only hydrophobic amino acids (IAVP), while the second is a relatively flexible, because it includes glutamic acid in IAVPGEVA and threonine IAVPTGVA. The tridimensional structure of these peptides has been investigated by circular dichroism (CD) spectra in water-TFE,[37] revealing a β-turn II conformation for IAVPGEVA and a β-turn I conformation for IAVPTGVA. This structural diversity may be possibly responsible for their different stability towards peptidases.

The second evidence concerns the ability of the breakdown fragments to cross the intestinal barrier to reach the BL side. In particular, both breakdown fragments deriving by the loss of isoleucine are absorbed with a rate that is probably similar to that of their intact precursor, since the relative percentages between the AP and BL compartment are very similar. On the contrary,

the two shortest breakdown fragments IAVPT and IAVP are 5-fold and 2.6-fold less abundant in BL than in the AP side. Up-to-now, several attempts have been carried out in order to establish the correlations between peptide structure and transport,[38] but the complex characteristics of intestinal transport are not yet completely understood as well as the effects of metabolism. However, it is important to consider that the degradation does not determine an automatic activity loss. In particular, a recent study has demonstrated how hydrolytic fragments derived from meat peptides maintain their activity after cellular transport.[39]

In conclusion, this paper provides the first evidence of the transepithelial absorption of three soy peptides across Caco-2 cell monolayers. The MS-based approach allowed detecting soy peptides as intact compound as well as fragments after absorption across the intestinal Caco-2 cell model. Of course, we are aware that further research is needed to clarify the transport mechanism of the parent peptides and derived metabolites.

Acknowledgments

We are indebted to Carlo Sirtori Foundation (Milan, Italy) for having provided part of equipment used in this experimentation.

SUPPLEMENTARY INFORMATION

Table 5.1S Data of the standard curves. Peptide area averages were obtained by averaging area ratios of replicate (Rep. No) for each calibration point (CL). Related standard deviation (SD) and RSD% were reported.

STANDARD PEPTIDE CURVES																
CL	Rep	C (A)	**IAVPGEVA**				C (B)	**IAVPTGVA**				C (C)	**LPYP**			
	No	(ng/mL)	A_{ratio}	Average	SD	RSD %	(ng/mL)	A_{ratio}	Average	SD	RSD %	(ng/mL)	A_{ratio}	Average	SD	RSD %
1	1	5	0.686				5	0.064				0.5	0.036			
	2		0.610	0.593	0.103	17.309		0.056	0.061	0.004	6.736		0.026	0.031	0.008	24.421
	3		0.483					0.062								
2	1	10	0.879				10	0.115				1	0.046			
	2		0.988	0.931	0.055	5.860		0.116	0.112	0.005	4.619		0.051	0.048	0.003	5.356
	3		0.927					0.106					0.047			
3	1	75	6.136				50	0.801				2	0.064			
	2		6.321	6.057	0.311	5.138		0.697	0.757	0.054	7.074		0.083	0.074	0.009	12.622
	3		5.714					0.772					0.076			
4	1	100	9.403				150	2.059				5	0.196			
	2		9.574	9.240	0.439	4.754		2.053	2.009	0.081	4.043		0.202	0.200	0.004	1.954
	3		8.742					1.916					0.203			
5	1	150	13.432				250	3.129				10	0.434			
	2		12.733	13.145	0.366	2.782		2.905	2.824	0.352	12.464		0.384	0.395	0.035	8.921
	3		13.270					2.439					0.366			
6	1	200	18.599				500	6.844				50	2.109			
	2		20.076	18.816	1.168	6.205		7.217	6.774	0.482	7.110		1.878	2.101	0.220	10.453
	3		17.772					6.262					2.317			
7	1	250	24.252				1000	11.288					3.807			
	2		22.653	23.447	0.799	3.410		11.811	11.550	0.370	3.203	100	4.579	4.150	0.393	9.482
	3		23.437										4.063			

CL = calibration points
Rep = number of technical replications
C (A) = concentration of IAVPGEVA
C (B) = concentration of IAVPTGVA
C (C) = concentration of LPYP

A ratio = Area of each replicate / Area of IS
Average = Average area ratio of a calibration point
SD = standard deviation
CV% = Coefficient of variations

References

1. Sirtori, C. R.; Eberini, I.; Arnoldi, A., Hypocholesterolaemic effects of soya proteins: results of recent studies are predictable from the Anderson meta-analysis data. *British Journal of Nutrition* **2007,** *97* (5), 816-822.

2. Zhang, H.; Bartley, G. E.; Jing, W.; Fagerquist, C. K.; Zhong, F.; Yokoyama, W., Peptides identified in soybean protein increase plasma cholesterol in mice on hypercholesterolemic diets. *J Agric Food Chem* **2013,** *61* (35), 8389-95.

3. Cho, S. J.; Juillerat, M. A.; Lee, C. H., Cholesterol lowering mechanism of soybean protein hydrolysate. *J Agric Food Chem* **2007,** *55* (26), 10599-604.

4. Harland, J.; Haffner, T., Systematic review, meta-analysis and regression of randomised controlled trials reporting an association between an intake of circa 25 g soya protein per day and blood cholesterol. *Atherosclerosis* **2008,** *200* (1), 13-27.

5. Jenkins, D. J.; Mirrahimi, A.; Srichaikul, K.; Berryman, C. E.; Wang, L.; Carleton, A.; Abdulnour, S.; Sievenpiper, J. L.; Kendall, C. W.; Kris-Etherton, P. M., Soy protein reduces serum cholesterol by both intrinsic and food displacement mechanisms. *J Nutr* **2010,** *140* (12), 2302S-2311S.

6. Lovati, M. R.; Manzoni, C.; Canavesi, A.; Sirtori, M.; Vaccarino, V.; Marchi, M.; Gaddi, G.; Sirtori, C. R., Soybean protein diet increases low density lipoprotein receptor activity in mononuclear cells from hypercholesterolemic patients. *J Clin Invest* **1987,** *80* (5), 1498-502.

7. Pak, V. V.; Koo, M. S.; Kasymova, T. D.; Kwon, D. Y., Isolation and identification of peptides from soy 11S-globulin with hypocholesterolemic activity. *Chemistry of Natural Compounds* **2005,** *41* (6), 710-714.

8. Pak, V. V.; Koo, M.; Kwon, D. Y.; Yun, L., Design of a highly potent inhibitory peptide acting as a competitive inhibitor of HMG-CoA reductase. *Amino Acids* **2012,** *43* (5), 2015-25.

9. Lammi, C.; Zanoni, C.; Arnoldi, A., IAVPGEVA, IAVPTGVA, and LPYP, three peptides from soy glycinin, modulate cholesterol metabolism in HepG2 cells through the activation of the LDLR-SREBP-2 pathway. *Journal of Functional Foods* **2015,** *14*, 469-478.

10. Lammi, C.; Zanoni, C.; Arnoldi, A., Three peptides from soy glycinin modulate glucose metabolism in human hepatic HepG2 cells. *Int J Mol Sci* **2015,** *16* (11), 27362-70.

11. Bejjani, S.; Wu, J., Transport of IRW, an ovotransferrin-derived antihypertensive peptide, in human intestinal epithelial Caco-2 cells. *J Agric Food Chem* **2013,** *61* (7), 1487-92.

12. Miguel, M.; Dávalos, A.; Manso, M. A.; de la Peña, G.; Lasunción, M. A.; López-Fandiño, R., Transepithelial transport across Caco-2 cell monolayers of antihypertensive egg-derived peptides. PepT1-mediated flux of Tyr-Pro-Ile. *Mol Nutr Food Res* **2008,** *52* (12), 1507-13.

13. Ding, L.; Wang, L. Y.; Zhang, Y.; Liu, J. B., Transport of Antihypertensive Peptide RVPSL, Ovotransferrin 328-332, in Human Intestinal Caco-2 Cell Monolayers. *Journal of Agricultural and Food Chemistry* **2015,** *63* (37), 8143-8150.

14. Vermeirssen, V.; Augustijns, P.; Morel, N.; Van Camp, J.; Opsomer, A.; Verstraete, W., In vitro intestinal transport and antihypertensive activity of ACE inhibitory pea and whey digests. *Int J Food Sci Nutr* **2005,** *56* (6), 415-30.

15. Amigo-Benavent, M.; Clemente, A.; Caira, S.; Stiuso, P.; Ferranti, P.; del Castillo, M. D., Use of phytochemomics to evaluate the bioavailability and bioactivity of antioxidant peptides of soybean β-conglycinin. *Electrophoresis* **2014,** *35* (11), 1582-9.

16. Lammi, C.; Zanoni, C.; Arnoldi, A.; Vistoli, G., Two peptides from soy beta-conglycinin induce a hypocholesterolemic effect in HepG2 cells by a statin-Like mechanism: comparative in vitro and in silico modeling studies. *Journal of Agricultural and Food Chemistry* **2015,** *63* (36), 7945-7951.

17. Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A. J.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H. L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, A.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **2009,** *27* (7), 633-41.

18. Boja, E. S.; Rodriguez, H., Mass spectrometry-based targeted quantitative proteomics: achieving sensitive and reproducible detection of proteins. *Proteomics* **2012,** *12* (8), 1093-110.

19. Picotti, P.; Aebersold, R., Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **2012,** *9* (6), 555-66.

20. Korte, R.; Monneuse, J. M.; Gemrot, E.; Metton, I.; Humpf, H. U.; Brockmeyer, J., New High-Performance Liquid Chromatography Coupled Mass Spectrometry Method for the Detection of Lobster and Shrimp Allergens in Food Samples via Multiple Reaction Monitoring and Multiple Reaction Monitoring Cubed. *J Agric Food Chem* **2016,** *64* (31), 6219-27.

21. Natoli, M.; Leoni, B. D.; D'Agnano, I.; D'Onofrio, M.; Brandi, R.; Arisi, I.; Zucco, F.; Felsani, A., Cell growing density affects the structural and functional properties of Caco-2 differentiated monolayer. *J Cell Physiol* **2011,** *226* (6), 1531-43.

22. Ferruzza, S.; Rossi, C.; Scarino, M. L.; Sambuy, Y., A protocol for differentiation of human intestinal Caco-2 cells in asymmetric serum-containing medium. *Toxicol In Vitro* **2012,** *26* (8), 1252-5.

23. Stuknyte, M.; Maggioni, M.; Cattaneo, S.; De Luca, P.; Fiorilli, A.; Ferraretto, A.; De Noni, I., Release of wheat gluten exorphins A5 and C5 during in vitro gastrointestinal digestion of bread and pasta and their absorption through an in vitro model of intestinal epithelium. *Food Research International* **2015,** *72*, 208-214.

24. Ferruzza, S.; Scarino, M. L.; Gambling, L.; Natella, F.; Sambuy, Y., Biphasic effect of iron on human intestinal Caco-2 cells: early effect on tight junction permeability with delayed onset of oxidative cytotoxic damage. *Cell Mol Biol (Noisy-le-grand)* **2003,** *49* (1), 89-99.

25. Sambuy, Y.; Angelis, I.; Ranaldi, G.; Scarino, M. L.; Stammati, A.; Zucco, F., The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biology and Toxicology* **2005,** *21* (1), 1-26.

26. Lammi, C.; Aiello, G.; Vistoli, G.; Zanoni, C.; Arnoldi, A.; Sambuy, Y.; Ferruzza, S.; Ranaldi, G., A multidisciplinary investigation on the bioavailability and activity of peptides from lupin protein. *J Funct Foods* **2016,** *24*, 297-306.

27. Biganzoli, E.; Cavenaghi, L. A.; Rossi, R.; Brunati, M. C.; Nolli, M. L., Use of a Caco-2 cell culture model for the characterization of intestinal absorption of antibiotics. *Farmaco* **1999,** *54* (9), 594-9.

28. Pauletti, G. M.; Okumu, F. W.; Borchardt, R. T., Effect of size and charge on the passive diffusion of peptides across Caco-2 cell monolayers via the paracellular pathway. *Pharm Res* **1997,** *14* (2), 164-8.

29. Foltz, M.; Cerstiaens, A.; van Meensel, A.; Mols, R.; van der Pijl, P. C.; Duchateau, G. S.; Augustijns, P., The angiotensin converting enzyme inhibitory tripeptides Ile-Pro-Pro and Val-Pro-Pro show increasing permeabilities with increasing physiological relevance of absorption models. *Peptides* **2008,** *29* (8), 1312-20.

30. Daniel, H., Molecular and integrative physiology of intestinal peptide transport. *Annu Rev Physiol* **2004,** *66*, 361-84.

31. Delie, F.; Rubas, W., A human colonic cell line sharing similarities with enterocytes as a model to examine oral absorption: advantages and limitations of the Caco-2 model. *Crit Rev Ther Drug Carrier Syst* **1997,** *14* (3), 221-86.

32. Artursson, P.; Palm, K.; Luthman, K., Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev* **2001,** *46* (1-3), 27-43.

33. Grootaert, C.; Jacobs, G.; Matthijs, B.; Pitart, J.; Baggerman, G.; Possemiers, S.; Van der Saag, H.; Smagghe, G.; Van Camp, J.; Voorspoels, S., Quantification of egg ovalbumin hydrolysate-derived anti-hypertensive peptides in an in vitro model combining luminal digestion with intestinal Caco-2 cell transport. *Food Res Int* **2017,** *99* (Pt 1), 531-541.

34. Kumar, K. K.; Karnati, S.; Reddy, M. B.; Chandramouli, R., Caco-2 cell lines in drug discovery- an updated perspective. *J Basic Clin Pharm* **2010,** *1* (2), 63-9.

35. Li, C.; Zhang, L.; Zhou, L.; Wo, S. K.; Lin, G.; Zuo, Z., Comparison of intestinal absorption and disposition of structurally similar bioactive flavones in Radix Scutellariae. *AAPS J* **2012,** *14* (1), 23-34.

36. Bai, J. P., Distribution of brush-border membrane peptidases along the rat intestine. *Pharm Res* **1994,** *11* (6), 897-900.

37. Pak, V. V.; Koo, M. S.; Kwon, D. Y.; Kasimova, T. D., Conformation analysis of Ile-Ala-Val-Pro peptide and its derivatives by circular dichroism. *Chemistry of Natural Compounds* **2004,** *40* (4), 398-404.

38. Chua, H. L.; Jois, S.; Sim, M. K.; Go, M. L., Transport of angiotensin peptides across the Caco-2 monolayer. *Peptides* **2004,** *25* (8), 1327-38.

39. Gallego, M.; Grootaert, C.; Mora, L.; Aristoy, M. C.; Van Camp, J.; Toldra, F., Transepithelial transport of dry-cured ham peptides with ACE inhibitory activity through a Caco-2 cell monolayer. *Journal of Functional Foods* **2016,** *21*, 388-395.

CHAPTER 6

MANUSCRIPT 4

PROTEOMIC CHARACTERIZATION OF HEMPSEED (*CANNABIS SATIVA* L.)

Gilda Aiello,[1] Elisa Fasoli,[2] Giovanna Boschin,[1] Carmen Lammi,[1] Chiara Zanoni,[1] Attilio Citterio,[2] Anna Arnoldi[1]

1 Department of Pharmaceutical Sciences, University of Milan, via Mangiagalli 25, 20133 Milan, Italy

2 Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Polytechnic of Milan, via Mancinelli 7, 20131 Milan, Italy

6.0 Abstract

This paper presents an investigation on hempseed proteome. The experimental approach, based on combinatorial peptide ligand libraries (CPLLs), SDS-PAGE separation, nLC-ESI-MS/MS identification, and database search, permitted identifying in total 181 expressed proteins. This very large number of identifications was achieved by searching in two databases: *Cannabis sativa* L. (56 gene products identified) and *Arabidopsis thaliana* (125 gene products identified). By performing a protein-protein association network analysis using the STRING software, it was possible to build the first interactomic map of all detected proteins, characterized by 137 nodes and 410 interactions. Finally, a Gene Ontology analysis of the identified species

permitted to classify their molecular functions: the great majority is involved in the seed metabolic processes (41%), responses to stimulus (8%), and biological process (7%).

Biological significance: Hempseed is an underexploited non-legume protein-rich seed. Although its protein is known for its digestibility, essential amino acid composition, and useful techno-functional properties, a comprehensive proteome characterization is still lacking. The objective of this work was to fill this knowledge gap and provide information useful for a better exploitation of this seed.

Keywords: *Cannabis sativa*, hempseed, interactomic maps, foodomics

6.1 Introduction

Plant foods are gaining increasing popularity, owing to the growing public awareness of their better environmental sustainability in respect to animal foods. Numerous seeds are used in human nutrition, but current research is mainly focused on protein-rich ones, such as legumes. They provide high quality proteins as well as useful health benefits, particularly in the area of cardiovascular prevention (Arnoldi, Boschin, Zanoni, & Lammi, 2015; Arnoldi, Zanoni, Lammi, & Boschin, 2015; Bazzano, Thompson, Tees, Nguyen, & Winham, 2011; Lammi, Zanoni, Scigliuolo, D'Amato, & Arnoldi, 2014). In the meanwhile, however, other seeds are deserving increasing attention.

Hempseed, i.e. the seed of the non-drug cultivars of industrial hemp, is certainly an underexploited protein-rich seed. Although its use as human food dates back probably to prehistory, together with the fiber utilization in textiles, the cultivation of this plant has been banned for some decades in most developed countries, owing to the morphological similarity with marijuana. In fact, hemp and marijuana are both strains of the same species *Cannabis sativa* L. (Sawler, Stout, Gardner, Hudson, Vidmar, Butler, et al., 2015), although they are characterized by very different contents of Δ^9-tetrahydrocannabinol (THC). The distinction between these populations, however, is not limited to the genes underlying THC production, since these plants are significantly differentiated at genome-wide level (Sawler, et al., 2015).

In the last years, however, the cultivation of the non-drug cultivars of industrial hemp is legal again in many countries and the global market is rapidly increasing, since well characterized low-THC cultivars are available (THC content < 200 mg/kg) (Oomah, Busson, Godfrey, & Drover, 2002). Many factors favor the reintroduction of the cultivation of this plant: it is a multipurpose crop with numerous applications in different industrial sectors (Vonapartis,

Aubin, Seguin, Mustafa, & Charron, 2015) and it is particularly sustainable, since it rarely requires irrigation and the fast growth reduces the use of herbicides (http://eiha.org/media/2014/10/13-06-European-Hemp-Industry.pdf).

The great interest for hempseed depends on its nutritional content (whole seed): 35.5% oil, 24.8% protein, 20-30% carbohydrates, 27.6% total fiber (5.4% digestible and 22.2% non-digestible fiber) and 5.6% ash (Callaway, 2004). Moreover, the concentration of the main antinutritional factors, such as phytic acid, condensed tannins, and trypsin inhibitors, is low (Russo & Reggiani, 2015). Currently, the main industrial interest is for the oil that has numerous applications either in food or body care products, being rich in polyunsaturated fatty acids (Russo & Reggiani, 2015). In the meanwhile, however, there is an increasing attention for hempseed protein owing to its digestibility (Wang, Tang, Yang, & Gao, 2008), satisfactory essential amino acid composition (Wang, Tang, Yang, & Gao, 2008), and techno-functional properties (Malomo & Aluko, 2015; Malomo, He, & Aluko, 2014; Tang, Ten, Wang, & Yang, 2006). Moreover, hempseed protein has also potential applications in nutraceutics and functional foods, since the treatment with suitable enzymes permits to produce hydrolyzed proteins providing useful health benefits, as hypotensive agents (Girgih, Alashi, He, Malomo, & Aluko, 2014; Girgih, Alashi, He, Malomo, Raj, Netticadan, et al., 2014; Girgih, He, Malomo, Offengenden, Wu, & Aluko, 2014) and antioxidants (Girgih, et al., 2014; Girgih, Udenigwe, & Aluko, 2011, 2013).

Literature indicates that a major protein in hempseed is edestin, a storage protein (Odani, 1998). Edestin, a hexameric 11S protein, is easily digested and contains significant amounts of all essential amino acids (Callaway, 2004), especially sulfur amino acids (Odani, 1998) and arginine (Tang, Ten, Wang, & Yang, 2006).

Hempseed proteome has been investigated only once in the Korean cultivar Cheungsam (Park, Seo, & Lee, 2012): the use of conventional techniques, such as 2-dimensional electrophoresis, mass spectrometry, and search in the nr-NCBI protein database permitted to identify 168 unique protein spots (Park, Seo, & Lee, 2012). However, only one was assigned to *C. sativa*, whereas most of them were assigned to *Medicago sativa*, *Oryza sativa* or other plants. The current improved availability of genome information about *C. sativa* as well as the increasing interest for this seed prompted us to investigate again hempseed proteome. This research was performed on the seeds of the French cultivar Futura.

In order to get as much as possible a comprehensive information, combinatorial peptide ligand libraries (CPLLs) were used for protein equalization (Righetti, Boschetti, Kravchuk, & Fasoli, 2010; Righetti, Fasoli, & Boschetti, 2011). This is a powerful and highly sensitive technique

allowing access to low- and very-low-abundance proteins, which enlarges very much the possibility of protein identification. This technique has been applied, for example, to the analysis of the proteomes of donkey milk (Cunsolo, Muccilli, Fasoli, Saletti, Righetti, & Foti, 2011) and champagne (Cilindre, Fasoli, D'Amato, Liger-Belair, & Righetti, 2014), or to the detection of corn allergens (Fasoli, Pastorello, Farioli, Scibilia, Aldini, Carini, et al., 2009).

6.2 Materials and Methods

6.2.1 Chemicals

ProteoMiner™ (combinatorial hexapeptide ligand library beads, CPLL), Laemmli buffer, 40% acrylamide/bis solution, N,N,N′,N′-tetramethylethane-1,2-diamine (TEMED), molecular mass standards and electrophoresis apparatus for one-dimensional electrophoresis were from Bio-Rad Laboratories Inc. (Hercules, CA, US). β-Mercaptoethanol, dithiothreitol (DTT), ammonium persulfate, 3-[3-cholamidopropyl-dimethylammonio]-1-propanosulfonate (CHAPS), acetonitrile (ACN), trifluoroacetic acid (TFA), sodium dodecyl sulfate (SDS), iodoacetamide (IAA), formic acid (FA) and all other chemicals used in the experimental work were pure analytical grade products, purchased from Sigma-Aldrich Corporation (St Louis, MO, US). Complete protease inhibitor cocktail tablets were from Roche Diagnostics (Basel, CH); modified porcine trypsin from Promega (Madison, WI, USA).

6.2.2 Hempseed proteome extraction and CPLL treatment

Hempseed (*C. sativa* cultivar Futura) were kindly provided by Dr. I. Galasso, Institute of Agricultural Biology and Biotechnology, CNR (Milan, Italy). In order to prepare defatted flour, hempseed were ground with a domestic grinder, added with hexane (1:20 w/v), and kept overnight under gentle stirring. The slurry was centrifuged at 10,000 rpm and 4 °C for 40 min and the organic solvent was then completely removed by filtration and drying. The protein extract (HPE) was obtained from 3 g of defatted flour treating with 30 mL native buffer containing 50 mM Tris-HCl (pH 7.2), 50 mM NaCl, 1% m/v CHAPS and protease inhibitor cocktail under stirring overnight. The slurry was centrifuged at 9,000 rpm for 10 min to separate the insoluble debris from the clarified protein solution. The supernatant was collected and centrifuged at 18,500 rpm for 30 min. The protein was precipitated with 20% TCA in water on ice for 1 h and then centrifuged at 9,000 rpm for 30 min. The resulting pellet was dissolved in 15 mL of 50 mM Tris-HCl (pH 7.2), 50 mM NaCl, and 1% w/v CHAPS and directly submitted

to analysis (raw sample). Protein concentrations were determined using the Bradford assay reagent (Bio-Rad, Laboratories Inc, CA, US) with bovine serum albumin as the standard. The absorbance of the standards and unknown sample was determined at 595 nm and it was estimated to be 8.4 µg/µL. In parallel, this solution was treated with CPLLs for protein equalization. The first protein capture was performed overnight in batch mode at pH 7.2 using 50 µL of ProteoMiner at room temperature under gentle rocking. The second capture was conducted consecutively on the same sample: the treatment with 50 µL of ProteoMiner was performed at pH 2.2 after addition of 0.1% v/v TFA and formic acid. Beads of both captures were recovered by filtering with Micro Bio-Spin chromatographic columns (Bio-Rad) under vacuum and then washed twice with their buffer in order to eliminate unbound protein excesses. The captured proteins from each sample were desorbed (twice, the former time with 50 µL, the latter with 40 µL) with an extraction buffer containing 4% w/v SDS and 20 mM DTT for 15 min, under boiling conditions (Candiano, Dimuccio, Bruschi, Santucci, Gusmano, Boschetti, et al., 2009). The procedure was repeated twice and each of these replicates was submitted to the following procedures.

6.2.3 SDS-PAGE Analysis

Five microliters of each CPLL eluate (labelled as E2.2 and E7.2, respectively) and the untreated sample (raw) dissolved in Laemmli buffer (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue, and 0.125 M Tris–HCl, pH 6.8) were loaded onto the SDS-PAGE gel. The gel was composed by a 4% polyacrylamide stacking gel over a 12% resolving polyacrylamide gel. The cathodic and anodic compartment were filled with Tris–glycine buffer, pH 8.3, containing 0.1%, m/v, SDS. The electrophoresis was conducted at 100 V until the dye front reached the gel bottom. Staining was performed with colloidal Coomassie Blue and destaining with 7% (v/v) acetic acid in water. The three sample lanes (E2.2, E7.2 and Raw) of the SDS-PAGE gels were manually cut in 10 slices along the gel migration path and subjected to the standard procedure of reduction and alkylation, followed by trypsin digestion (Cunsolo, Muccilli, Fasoli, Saletti, Righetti, & Foti, 2011). The recovered peptides were lyophilized and then dissolved in 10 µL water/acetonitrile (99:1 v/v) added with 0.1% FA.

6.2.4 Mass spectrometry and Data Analysis

Mass spectrometry data were acquired on a SL series Ion Trap mass spectrometer interfaced with a HPLC-Chip Cube source (Agilent Technologies, Palo Alto, CA, USA). The chromatographic separations of tryptic digests were performed on a LC-chip containing a 40 nL enrichment column (Zorbax 300SB-C18, 5 µm pore size), a 43 mm x 75 µm analytical column packed (Zorbax 300SB-C18, 5 µm pore size) capillary tubing connections and a nano-electrospray needle. The reconstituted sample (1 µL) was loaded onto the enrichment column at a flow rate 4 µL/min for 2 min using isocratic 100% C solvent phase (99% water, 1% ACN and 0.1% FA). After cleanup, the chip valve was switched to separation and trapped peptides were eluted into the mass spectrometer at the constant flow rate of 0.3 µL/min. The LC solvent A was 95% water, 5% ACN, 0.1% formic acid; solvent B was 5% water, 95% ACN, 0.1% formic acid. The nano-pump gradient program was as follows: 5% solvent B (0 min), 50 % solvent B (0-50 min), 95% solvent B (50-60 min) and back to 5% in 10 min to re-equilibrate the column. The nano-ESI source operated under the following conditions: drying gas temperature 300 °C, flow rate 3 L/min (nitrogen), capillary voltage 1950 V, with endplate offset -500V. Mass spectra were acquired in positive ionization mode, in the mass range from *m/z* 300-2200, with target mass 700 *m/z*, average of 2 spectra, ICC target 30,000 and maximum accumulation time 150 msec. The detection of cationic peptide ions was performed by data dependent acquisition Auto MS(n) mode with a dynamic exclusion set at 2 spectra and released after 0.1 min. Each sample was analysed twice.

The MS/MS spectra was analysed by Spectrum Mill MS Proteomics Workbench (Rev B.04.00, Agilent technology). MS/MS search, concatenated with the reverse one, was conducted against the subset *C. sativ*a and *Arabidopsis thaliana* protein sequences database, downloaded from the National Center for Biotechnology Information (NCBI) containing 531 and 229,367 entries, respectively. In addition, it was conducted also against the UniProtKB database (containing 182 and 83,307 entries respectively for *C. sativa* and *A. thaliana*) including sequences derived from Swiss-Prot and TrEMBL (Translated EMBL Nucleotide Sequence Data Library). The cysteine carbamidomethylation was set as fixed modification, whereas oxidation of methionine and transformation of N-terminal glutamic acid residues in the pyroglutamic acid were included as variable modifications. Trypsin was chosen as digestive enzymes, 2 missed cleavage were allowed; peptide mass tolerance was set to 2.5 Da, fragment mass tolerance to 0.7 Da; Local False Discovery Rate $\leq 0.1\%$. The threshold used for peptide identification was Score Peak Intensity % $\geq$ 70%, accepting only peptides with the difference

between forward and reverse scores ≥ 2. Proteins were considered detected, if they were identified by at least two peptides. The **Table 6.1S (A, B, C)** (Supporting Information) reports a list of all identified proteins, MS/MS scores, sequence coverages, and number of unique peptides.

The software STRING (Search Tool for the Retrieval of Interacting Genes, v.10 web server, http://stringdb.org/) (Jensen, Kuhn, Stark, Chaffron, Creevey, Muller, et al., 2009), set on *A. thaliana* as organism database, was used to generate a functional association network of the potentially interacting proteins. The pathways classification was performed after the automatic enrichment in STRING, based on the information provided by the KEGG-Pathway Database. The assignment of biological processes, cellular components, and molecular functions of identified proteins was performed using the web available software QuickGo (www.ebi.ac.uk/QuickGo).

6.3 Results and Discussion

6.3.1 Protein identification

The workflow adopted for a comprehensive proteome discovery included proteome extraction, pre-fractionation via CPLL technology, protein separation by SDS-PAGE electrophoresis, and characterization by MS analysis. **Figure 6.1** shows the SDS-PAGE profiles of the untreated sample (raw) *versus* those of the CPLL eluates after capture at two different pH values, i.e. pH 7.2 and 2.2 (lanes labelled as E7.2 and E2.2, respectively). The two latter profiles were more resolved, more intense and with improved densities in respect to the raw sample, which showed a series of poorly resolved bands, distributed in the mass regions from 10 to 20 kDa. Therefore, the CPLL treatment had a very positive contribution on protein capture, making more visible the bands in the range 50-75 kDa for eluate E7.2 and those in the range 10-50 kDa for eluate E2.2. The fact that eluate E7.2 exhibits a separation pattern different from eluate E2.2 is not surprising, since the capture efficiency strongly depends on the ionic environment.

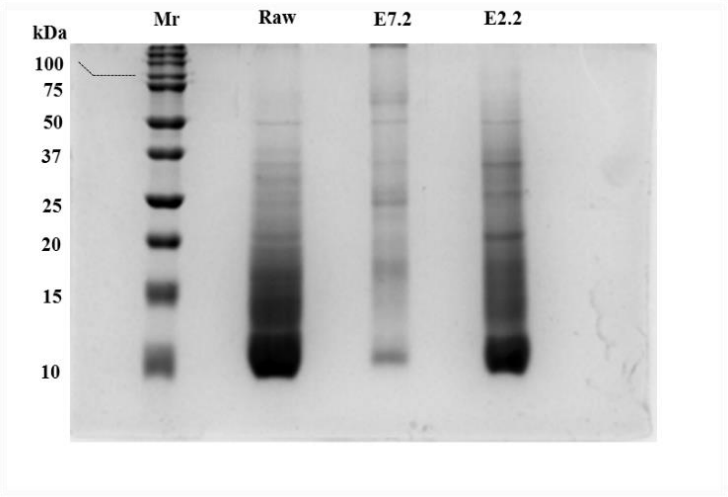


Figure 6.1 - SDS-PAGE profiles of the untreated (raw) sample versus the eluates of CPLL capture at the two pH values (lanes E7.2, E2.2). Each lane was cut into segments, the proteins digested, and submitted to MS analysis. Mr: molecular mass ladder. Staining with micellar Coomassie blue.

The MS Analysis, as reported in **Table 6.1**, reveals the total number of identified proteins using Spectrum Mill search engine against two different database search, i.e. *C. sativa* and *A. thaliana.* Although the *C. sativa* database would be the most appropriate for identifying species-specific gene products, it suffers from the inherent limitation due to an incompletely sequenced DNA (only 531 entries). In order to overcome this inadequacy and to expand the dataset of identified proteins, also the *A. thaliana* database research was performed. The use of this very common database is justified by the phylogenetic proximity of these species: in fact, *C. sativa*, species belonging to the family of *Cannabaceae*, shows a close phylogenetic relationship to *A. thaliana* (*Brassicaceae*) since both belong to the same *Rosidae* subclass. Moreover, the very high number of entries (229,367) contained in NCBI protein sequence database makes *A. thaliana* the species chosen by most plant biologists as a model organism to analyse plant gene expression as well as plant development. In conjunction with CPLL capture, this procedure allowed the identification of 56 species-specific gene products plus other 125 gene products identified by homology to the model species *A. thaliana.* Recognizing in total 181 unique gene products, these findings represent a main improvement in the exploration of the hempseed proteome in respect to the previous investigations. Among the entities identifies through the database of *C. sativa*, this work provided the first experimental detection of the two isoforms of edestin, i.e. Ede1 and Ede2, which had been previously deduced at gene levels (Docimo, Caruso, Ponzoni, Mattana, & Galasso, 2014). Edestin does not appear in the protein entities previously identified (Park, Seo, & Lee, 2012).

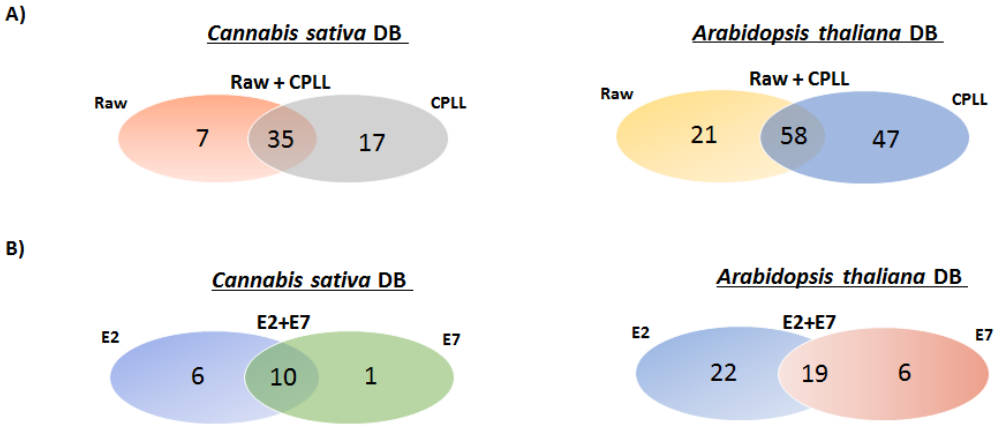


Figure 6.2 - Venn diagrams of the total number of identified species in the raw extract of hempseed versus those detected in all eluates after CPLL treatment considering both database. The IDs in the raw sample consist of 39 unique gene products assigned to the *C. sativa* database, whereas 78 unique gene products were assigned to *A. thaliana* database. The contribution to the total identifications of the eluates from the pH 2.2 and 7.2 CPLL captures are the following: 6 unique proteins were identified at pH 2.2 and 1 at pH 7.2 for *C. sativa*, whereas 22 unique proteins were detected at pH 2.2 and 6 at pH 7.2 for *A. thaliana*.

The relative contribute in terms of MS-recognized proteins carried out by protein equalization is represented by Venn diagrams (**Figure 6.2**), which compare the proteins identified in the combined CPLLs eluates with those of the untreated extracts, considering both databases searches. The efficiency of CPLL treatment allowed the detection of 17 unique proteins of *C. sativa* and 47 of *A. thaliana* after CPLL treatment. This technology permitted a 55% increment in the proteome discovery, attributed to minor proteins of HPEs, unrevealed without the CPLL treatment. The high efficiency is due to the use of a native buffer devoid of ionic and chaotropic detergents, such as SDS or urea, which could prevent the successful interaction between proteins and ligand peptides on the beads. However, it must be underlined that these mild conditions did not allow extracting highly hydrophobic proteins, which may be considered a limitation of this approach. The outstanding result of the equalization is also due to the different pH conditions in which the sample was incubated with CPLLs, initially under physiological condition at pH 7.2 and afterwards at pH 2.2, leading to the identification of 11 and 17 protein species, respectively, using the *C. sativa* database search. The most efficient capture occurred at pH 2.2. This observation implies the extensive hydrophobic interactions occurring at low pH between proteins and the hexapeptide ligands that are very similar to reverse-phase chromatography based interactions.

Table 6.1 Identified proteins from hempseed raw, pH 2.2, and 7.2 samples

N.	Acc. N.	Identified proteins (gene name) [a]	Taxonomy	MW (kDa)	Sample
1	A7IZZ2	(+)-alpha-pinene synthase, (N/A) [b]	*C. sativa*	72.4	E2.2 - E7.2 - Raw
2	A7IZZ1	(-)-limonene synthase (TPS1)	*C. sativa*	72.9	E2.2 - E7.2
3	H9A1V3	Acyl-activating enzyme 1 (N/A)	*C. sativa*	80.5	E2.2 - E7.2
4	H9A1V4	Acyl-activating enzyme 2 (N/A)	*C. sativa*	75.1	E2.2 - Raw
5	H9A1V5	Acyl-activating enzyme 3 (N/A)	*C. sativa*	59.7	Raw
6	H9A1V8	Acyl-activating enzyme 6 (N/A)	*C. sativa*	62.9	Raw
7	H9A1V9	Acyl-activating enzyme 7 (N/A)	*C. sativa*	67.2	E2.2 - Raw
8	H9A1W1	Acyl-activating enzyme 9 (N/A)	*C. sativa*	62.2	E2.2 - Raw
9	H9A1W3	Acyl-activating enzyme 11 partial (N/A)	*C. sativa*	37.2	E2.2 - E7.2
10	H9A8L1	Acyl-activating enzyme 12 (N/A)	*C. sativa*	84.4	E2.2 - Raw
11	H9A8L2	Acyl-activating enzyme 13 (N/A)	*C. sativa*	79.6	E2.2 - Raw
12	H9A8L3	Acyl-activating enzyme 14 (N/A)	*C. sativa*	81.2	Raw
13	H9A8L4	Acyl-activating enzyme 15 (N/A)	*C. sativa*	87.8	E7.2 – Raw
14	A0A0C5APY2	Acetyl-CoA carboxylase carboxyltransferase beta subunit (accD)	*C. sativa*	57.0	E2.2 - E7.2 - Raw
15	A0A0C5ARS5	ATP synthase CF1 beta subunit (atpB)	*C. sativa*	53.8	E2.2
16	E5DK51	Atp1 partial (atp1)	*C. sativa*	37.7	E2.2
17	I6XT51	Betv1-like protein (N/A)	*C. sativa*	17.6	E2.2
18	A0A0E3TJM8	Cannabidiolic acid synthase homolog (CBDAS2)	*C. sativa*	62.7	E2.2 - E7.2
19	C6KI62	Chalcone synthase-like protein 1 (CAN383)	*C. sativa*	43.4	E2.2 - E7.2 - Raw
20	I6WIE9	Chalcone isomerase-like protein (N/A)	*C. sativa*	23.9	E7.2 - Raw
21	A0A0C5ARS7	Chloroplast envelope membrane protein (cemA)	*C. sativa*	27.5	E7.2
22	P00053	Cytochrome c (N/A)	*C. sativa*	12.2	Raw
23	A0A088MFF4	Delta 12 desaturase (FAD2A)	*C. sativa*	44.9	Raw
24[c]	A0A090DLH8, A0A090CXP7	Edestin 1 (ede1A.ede1C), Edestin 1 (ede1B)	*C. sativa*	58.8	E2.2 - E7.2 - Raw
25	A0A090CXP9	Edestin 2 (ede2A)	*C. sativa*	56.0	E2.2 - E7.2 - Raw
26[d]	A0A090DLI7, A0A090CXP8	Edestin 2 (ede2B), Edestin 2 (ede2C)	*C. sativa*	56.0	E2.2 - E7.2 - Raw
27	A0A0C5APZ4	Hypothetical chloroplast RF21 (ycf2)	*C. sativa*	272.8	E2.2 - E7.2 - Raw
28	A0A0C5AUH7	Hypothetical chloroplast RF34 (ycf3)	*C. sativa*	19.7	E7.2 - Raw
29	L0N5C8	Hypothetical protein (N/A)	*C. sativa*	61.7	E2.2 - E7.2
30	E5DKP2	MatR. Partial (matR)	*C. sativa*	60.1	E7.2 - Raw
31	A0A0C5AUE3	Maturase K (matK)	*C. sativa*	61.0	Raw

32	A0A0C5AUJ6	NADH-plastoquinone oxidoreductase subunit 5 (ndhF)	*C. sativa*	86.0	E2.2
33	A0A0C5AS19	NADH-plastoquinone oxidoreductase subunit 7 (ndhH)	*C. sativa*	45.7	E2.2 - E7.2 - Raw
34	E5DL82	NADH dehydrogenase subunit 5 partial (nad5)	*C. sativa*	39.9	E2.2 - E7.2 - Raw
35	V5KWZ6	Phenylalanine ammonia lyase (PAL)	*C. sativa*	77.6	E2.2 - E7.2 - Raw
36	A0A0C5AUH1	Photosystem II CP43 chlorophyll apoprotein (psbC)	*C. sativa*	52.0	E2.2 - E7.2
37	A9XV91	Photosystem II CP47 chlorophyll apoprotein. partial (psbB)	*C. sativa*	53.8	E2.2 - Raw
38	F1LKH7	Polyketide synthase protein (PKSG2)	*C. sativa*	43.1	Raw
39	U6EFF4	Putative LysM domain containing receptor kinase (lyk2)	*C. sativa*	66.6	E2.2 - E7.2 - Raw
40	A0A0C5AS13	Ribosomal protein L2 (rpl2)	*C. sativa*	30.1	E2.2 - E7.2 - Raw
41	A0A0C5AUJ2	Ribosomal protein L16 (rpl16)	*C. sativa*	13.4	E2.2 - E7.2
42	A0A0C5AS20	Ribosomal protein L23 (rpl23)	*C. sativa*	10.7	E2.2 - Raw
43	A0A0C5ART6	Ribosomal protein S3 (chloroplast) (rps3)	*C. sativa*	17.8	E2.2 - E7.2 - Raw
44	E5DLX2	Ribosomal protein S3. partial (mitochondrion) (rps3)	*C. sativa*	57.3	E2.2 - E7.2 - Raw
45	A0A0C5ARS4	Ribosomal protein S4 (rps4)	*C. sativa*	23.7	E2.2 - Raw
46	A0A0C5ARU3	Ribosomal protein S7 (rps7)	*C. sativa*	17.4	E2.2
47	A0A0C5APY7	Ribosomal protein S8 (rps8)	*C. sativa*	15.5	E2.2 - Raw
48	A0A0C5ART4	Ribosomal protein S11 (rps11)	*C. sativa*	15.0	E2.2
49	A0A0C5AUK1	Ribosomal protein S15 (rps15)	*C. sativa*	10.9	E2.2 - E7.2 - Raw
50	A0A0C5ART1	Ribosomal protein S18 (rps18)	*C. sativa*	12.0	E2.2 - E7.2
51	A0A0C5B2I6	Ribulose 1.5-bisphosphate carboxylase/oxygenase large subunit (rbcL)	*C. sativa*	53.3	E2.2 - E7.2 - Raw
52	I0B4C2	Ribulose-1.5-bisphosphate carboxylase/oxygenase large subunit.partial (rbcL)	*C. sativa*	24.4	E2.2 - E7.2
53	A0A0C5AUI9	RNA polymerase alpha subunit (rpoA)	*C. sativa*	38.3	E7.2 – Raw
54	A0A0C5ARQ8	RNA polymerase beta subunit (rpoB)	*C. sativa*	121.6	E2.2 - E7.2
55	A0A0C5ARX9	RNA polymerase beta' subunit (rpoC2)	*C. sativa*	159.4	E2.2 - E7.2 - Raw
56	I1V0C2	Tetrahydrocannabinolic acid synthase. partial (N/A)	*C. sativa*	62.3	E2.2 - Raw
57	Q0WNZ5	5-methyltetrahydropteroyltriglutamate (MS3)	*A. thaliana*	91.0	Raw
58	O23058	A_IG005I10.23 protein (A_IG005I10.23)	*A. thaliana*	88.7	E2.2 - E7.2
59	Q42560	Aconitate hydratase 1 (ACO1)	*A. thaliana*	98.8	E2.2 - Raw
60	C0SVH8	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein (At4g14920)	*A. thaliana*	129.2	E2.2 - Raw
61	P92935	ADP.ATP carrier protein 2 (AATP2)	*A. thaliana*	67.9	E2.2 - E7.2
62	O22241	Arogenate dehydratase 4 (ADT4)	*A. thaliana*	46.4	E7.2
63	Q9M8K5	AT3g06130/F28L1_7 (F28L1.7)	*A. thaliana*	49.4	E2.2 - Raw
64	Q9FG23	At5g06830 (At5g06830)	*A. thaliana*	62.2	Raw
65	B9DG14	AT5G41100 (At5g41100)	*A. thaliana*	65.2	E7.2
66	Q8VZF6	AT5g45560/MFC19_23 (EDR2L)	*A. thaliana*	82.6	E7.2 - Raw

67	F4J394	ATP binding microtubule motor family protein (At3g51150)	*A. thaliana*	120.1	E2.2 - E7.2
68	F4IMB5	ATPase. F1 complex. alpha subunit protein (At2g07698)	*A. thaliana*	86.3	E2.2 - Raw
69	Q8L7G0	Auxin response factor 1 (ARF1)	*A. thaliana*	74.3	Raw
70	Q38810	cDNA-5-encoded protein. partial (N/A)	*A. thaliana*	33.1	E2.2 - E7.2
71	Q9C9U2	Cell division protein kinase (CDKD-1)	*A. thaliana*	45.3	E7.2 - Raw
72	O80891	Cellulose synthase-like protein B4 (CSLB4)	*A. thaliana*	86.2	E2.2
73	Q9SB75	Cellulose-synthase-like C5 (CSLC5)	*A. thaliana*	80.1	E2.2 - Raw
74	Q9LIM5	Chloroplast protein HCF243 (hcf243)	*A. thaliana*	76.8	E2.2
75	O80458	Cis-prenyltransferase (DPS)	*A. thaliana*	35.3	E2.2 - Raw
76	Q8VY07	Clathrin binding protein-like (EPSIN1)	*A. thaliana*	62.7	Raw
77	Q9FKV9	Cyclin-dependent protein kinase-like protein (At5g44290)	*A. thaliana*	73.6	E2.2 - Raw
78	Q9SB89	DEAD-box ATP-dependent RNA helicase 27 (RH27)	*A. thaliana*	72.3	Raw
79	Q8GXD6	DEAD-box ATP-dependent RNA helicase 49 (RH49)	*A. thaliana*	63.4	E7.2
80	Q38K61	Disease resistance protein. Partial (At4g10780)	*A. thaliana*	31.6	E2.2
81	Q19DM3	Disease resistance protein. Partial (At5g44870)	*A. thaliana*	25.7	E2.2
82	X2L4T4	DM2D (N/A)	*A. thaliana*	140.0	Raw
83	Q42572	DNA ligase 1 (LIG1)	*A. thaliana*	94.2	E7.2 - Raw
84	Q9FNI6	DNA repair protein RAD5 protein (At5g22750)	*A. thaliana*	115.6	E7.2
85	F4JRX3	DNA topoisomerase. type IA. core *(*At4g31210)	*A. thaliana*	142.4	E7.2 - Raw
86	F4I5J7	DNAJ heat shock N-terminal domain-containing protein (At1g77020)	*A. thaliana*	43.0	E2.2 - Raw
87	O04478	Endoglucanase 7 (KOR2)	*A. thaliana*	70.1	E2.2 - E7.2 - Raw
88	P42762	Erd1 protein precursor (CLPD)	*A. thaliana*	103.7	E2.2 - Raw
89	O23026	EST gb|T04104 comes from this gene/T1G11.16 protein (T1G11.16)	*A. thaliana*	45.3	E2.2 - Raw
90	Q76E23	Eukaryotic translation initiation factor 4G (EIF4G)	*A. thaliana*	188.7	E2.2 - Raw
91	Q8S905	F15H18.12 (NACK1)	*A. thaliana*	114.6	E2.2 - Raw
92	Q9SSQ2	F6D8.29 (At1g52490)	*A. thaliana*	50.6	Raw
93	Q9SHM9	F7F22.11 (N/A)	*A. thaliana*	70.2	E7.2 - Raw
94	O81490	F9D12.11 protein (F9D12.11)	*A. thaliana*	151.0	E2.2
95	Q9SA52	g5bf (CSP41B)	*A. thaliana*	42.9	E7.2 - Raw
96	Q9T0F7	Glucuronoxylan 4-O-methyltransferase 2 (GXM2)	*A. thaliana*	33.4	E7.2
97	C0Z2I0	Glyceraldehyde-3-phosphate dehydrogenase (At1g13440)	*A. thaliana*	37.6	Raw
98	Q94FQ0	Glycine-rich protein GRP17 (N/A)	*A. thaliana*	53.4	E7.2 - Raw
99	F4IAH9	Glyoxalase I homolog (GLX1)	*A. thaliana*	36.5	E2.2 - Raw
100	Q8S8N9	Golgin candidate 1 (GC1)	*A. thaliana*	79.9	E7.2
101	F4IVL6	Gravitropism defective 2 (GRV2)	*A. thaliana*	281.4	Raw
102	Q56X77	Guanylate kinase (At2g41880)	*A. thaliana*	23.4	E2.2

103	Q9LKR3	Heat shock 70 kDa protein 11 (MED37A)	*A. thaliana*	73.9	E7.2 - Raw
104	O65719	Heat shock 70 kDa protein 4 (MED37C)	*A. thaliana*	71.6	E2.2 - E7.2 - Raw
105	Q9LHA8	Heat shock 70 kDa protein 5 (HSP70-5)	*A. thaliana*	71.5	E7.2 - Raw
106	Q9S9N1	Heat shock protein 60-3A (At3g13860)	*A. thaliana*	71.3	E7.2 - Raw
107	Q93ZM7	Heat shock protein 70-3 (HSP70-3)	*A. thaliana*	60.8	E2.2 - E7.2 - Raw
108	Q56XE8	Hexokinase-4 (At3g20040)	*A. thaliana*	55.3	E2.2 - E7.2
109	Q6NR90	Histone H4 (At1g07660)	*A. thaliana*	11.4	Raw
110	O48701	Hypothetical protein (F3I6.28)	*A. thaliana*	79.0	E2.2 - Raw
111	Q84MA1	Hypothetical protein (At1g06560)	*A. thaliana*	66.4	E2.2
112	Q9ZQQ4	Hypothetical protein (At2g14410)	*A. thaliana*	76.9	E7.2 - Raw
113	Q9S755	Hypothetical protein (F3L24.2)	*A. thaliana*	22.5	E2.2 - E7.2
114	O81893	Inositol-tetrakisphosphate 1-kinase 2 (ITPK2)	*A. thaliana*	44.4	E2.2 - Raw
115	O82762	KH domain-containing protein (At2g25970)	*A. thaliana*	64.7	E2.2 - E7.2
116	G4WT86	Late flowering protein (FRI)	*A. thaliana*	68.8	E2.2 - Raw
117	Q5XF33	Magnesium-chelatase subunit ChlI-2 (CHLI2)	*A. thaliana*	47.0	E2.2 - Raw
118	Q9FNB0	Magnesium-chelatase subunit H (CHLH)	*A. thaliana*	154.0	E2.2 - E7.2 - Raw
119	Q84M98	Meiotic endonuclease 1A (EME1A)	*A. thaliana*	62.1	E2.2 - E7.2
120	Q9SG92	Methylesterase 17 (MES17)	*A. thaliana*	31.2	E2.2
121	Q9FLR3	No apical meristem (NAM) -like protein (NAC080)	*A. thaliana*	37.8	Raw
122	O48689	Paired amphipathic helix repeat-containing protein (F3I6.15)	*A. thaliana*	85.9	E2.2 - Raw
123	O80958	Pentatricopeptide repeat protein (LOJ)	*A. thaliana*	98.5	E2.2 - Raw
124	Q9ZVX5	Pentatricopeptide repeat-containing protein (At2g16880)	*A. thaliana*	84.6	E2.2
125	Q84TI3	PHD-finger TITANIA 2 (OBE4)	*A. thaliana*	132.4	E2.2 - E7.2
126	Q8LFV7	Phosphoglycerate kinase (N/A)	*A. thaliana*	42.2	E2.2 - E7.2 - Raw
127	Q9XIE6	Phospholipid-transporting ATPase 3 (ALA3)	*A. thaliana*	139.2	E2.2
128	Q9M8D3	Phosphoribosylformylglycinamidine synthase (At1g74260)	*A. thaliana*	155.3	E2.2
129	Q8S8D3	Phox (PX) domain-containing protein (At2g25350)	*A. thaliana*	73.1	E7.2 - Raw
130	F4K7T1	P-loop containing nucleoside triphosphate hydrolases superfamily protein (t5g62760)	*A. thaliana*	74.6	E2.2
131	Q0WR59	Probable inactive receptor kinase (At5g10020)	*A. thaliana*	115.7	E2.2 - Raw
132	Q42290	Probable mitochondrial-processing peptidase subunit beta (At3g02090)	*A. thaliana*	59.7	E2.2 - E7.2
133	Q9C533	Probable protein S-acyltransferase 22 (PAT22)	*A. thaliana*	67.0	E7.2 - Raw
134	P59583	Probable WRKY transcription factor 32 (WRKY32)	*A. thaliana*	52.2	E2.2 - Raw
135	O81147	Proteasome subunit alpha type-6-B (PAA2)	*A. thaliana*	27.5	E2.2 - E7.2
136	P0C7Q8	Protein (DA1)	*A. thaliana*	61.5	Raw
137	Q8VZG2	Protein HYPER-SENSITIVITY-RELATED 4 (HSR4)	*A. thaliana*	66.4	Raw
138	O49610	Protein kinase-like (M4E13.90)	*A. thaliana*	52.6	E2.2 - Raw

139	F4J8K6	Protein ribosomal RNA processing 5 (At3g11964)	*A. thaliana*	213.0	Raw
140	Q66GI4	Proteinaceous RNase P 1 (PRORP1)	*A. thaliana*	65.8	E7.2 - Raw
141	Q3EAS6	Putative DNA repair protein recA homolog 4 (At3g32920)	*A. thaliana*	24.5	E2.2 - E7.2
142	Q84W97	Putative kinesin (At2g47500)	*A. thaliana*	109.3	E2.2 - E7.2
143	Q9LR53	Putative protein kinase (At1g03740)	*A. thaliana*	84.9	E2.2 - E7.2
144	Q8VZQ9	Putative protein (REM3)	*A. thaliana*	117.0	E2.2
145	Q9SU68	Putative protein (T17F15.80)	*A. thaliana*	173.3	E2.2
146	Q93YS2	Putative protein phosphatase 2C 51 (At3g63340)	*A. thaliana*	120.9	E2.2 - Raw
147	Q9ZQE9	Putative retroelement pol polyprotein (At2g15650)	*A. thaliana*	156.3	E7.2 - Raw
148	F4IF81	Putative serine/threonine protein kinase (At1g79640)	*A. thaliana*	76.9	E2.2 - E7.2 - Raw
149	Q9SIL1	Putative TNP1-like transposon protein (At2g12390)	*A. thaliana*	110.7	E2.2 - E7.2 - Raw
150	Q8L622	Putative uncharacterized protein (At5g45030)	*A. thaliana*	66.2	E2.2
151	A0MEZ6	Putative uncharacterized protein (N/A)	*A. thaliana*	16.0	E2.2 - Raw
152	F4K4D6	PWWP domain-containing protein (At5g27650)	*A. thaliana*	118.4	E2.2
153	Q9FGN6	Receptor-like kinase MOL1 (LRR-RLK)	*A. thaliana*	100.0	E2.2
154	P17094	Ribosomal protein (ARP1)	*A. thaliana*	44.8	Raw
155	F4KE59	RING/FYVE/PHD zinc finger-containing protein (At5g16680)	*A. thaliana*	143.6	E2.2 - Raw
156	Q9LJX4	RNA binding protein-like (APUM5)	*A. thaliana*	107.7	E2.2
157	Q5D869	RNA polymerase IV largest subunit (NRPE1)	*A. thaliana*	220.4	E2.2 - E7.2
158	B0LC03	RPS2 (rps2)	*A. thaliana*	106.1	E2.2 - Raw
159	F4J3S1	Shugoshin C terminus (At3g10440)	*A. thaliana*	64.5	E2.2 - Raw
160	Q9FWR3	Similar to tRNA-splicing endonuclease positive effector SEN1 (F17F16.1)	*A. thaliana*	242.4	E2.2 - E7.2
161	F4JXC5	Subtilisin-like protease SBT5.4 (SBT5.4)	*A. thaliana*	84.1	E2.2 - Raw
162	O82496	T12H20.15 protein (T12H20.15)	*A. thaliana*	81.5	E2.2 - E7.2 - Raw
163	Q9LQ91	T1N6.6 (T1N6.6)	*A. thaliana*	96.4	E2.2 - Raw
164	Q9LPA7	T32E20.13 (N/A)	*A. thaliana*	156.5	E2.2 - Raw
165	Q9SCZ3	TIR-NBS-LRR class disease resistance protein (F26O13.200)	*A. thaliana*	142.8	E2.2 - E7.2
166	O04251	Transcription coactivator protein (At4g02110)	*A. thaliana*	146.9	E2.2
167	Q9LHE4	Transcription factor-like protein (At3g20010)	*A. thaliana*	116.7	E2.2 - E7.2 - Raw
168	Q9SSP7	Ubiquitin carboxyl-terminal hydrolase-related protein (F6D8.35)	*A. thaliana*	131.5	Raw
169	Q56Y19	Uncharacterized protein (At2g43235)	*A. thaliana*	49.1	E2.2 - E7.2
170	Q8S8L1	Uncharacterized protein (At2g24255)	*A. thaliana*	38.5	E2.2 - Raw
171	Q9FHT3	Uncharacterized protein (At5g37320)	*A. thaliana*	55.3	E2.2
172	F4K4M9	Uncharacterized protein (At5g43230)	*A. thaliana*	95.0	Raw
173	Q9FH23	Uncharacterized protein (N/A)	*A. thaliana*	168.2	E2.2 - Raw
174	Q9LXU4	Uncharacterized protein (T24H18_120)	*A. thaliana*	96.7	E2.2 - E7.2

175	Q9FHW3	Unnamed protein product (N/A)	*A. thaliana*	14.8	E7.2 - Raw
176	P93736	Valine--tRNA ligase (TWN2)	*A. thaliana*	127.7	E2.2 - Raw
177	Q94BQ3	WD and tetratricopeptide repeats protein 1 (At5g10940)	*A. thaliana*	84.9	Raw
178	F4HZB2	WD/BEACH domain protein SPIRRIG (SPI)	*A. thaliana*	404.3	E2.2 - Raw
179	P56785	ycf1 (TIC214)	*A. thaliana*	214.8	Raw
180	O64533	YUP8H12R.20 (YUP8H12R.209	*A. thaliana*	150.3	Raw
181	O49281	Zinc finger protein (F22K20.5)	*A. thaliana*	122.6	E2.2

a) According to "UniProtKB" (http://www.uniprot.org/)
b) N/A: not available
c) Not unique identification of Edestin 1 (ede1A.ede1C) and Edestin 1 (ede1B)
d) Not unique identification of Edestin 2 (ede2B) and Edestin 2 (ede2C)

6.3.2 Protein-protein association network analysis

In order to forecast protein-protein relations among all identified protein species, an interactomic network was obtained using the open-access STRING Software. Many protein–protein associations in STRING are based on systematic genome comparisons focused on conserved genomic neighborhoods, gene fusion events, and co-occurrence of genes across genomes. From a functional perspective, 'association' can mean direct physical binding as well as indirect interaction, such as participation in the same metabolic pathway or cellular process (von Mering, Jensen, Snel, Hooper, Krupp, Foglierini, et al., 2005). Using *A. thaliana* as reference organism, this paper provides the first interactomic map for *C. sativa*, showing in total 137 nodes connected by 410 interactions, some of which are characterized by narrow and dense connections (see **Figure 6.1S** in Supporting Information).

With the aim of grouping the interactions between the different proteins identified, cluster networks were created by using the K-means algorithm, which is included in the STRING website, selecting a value of 5 for the analysis (**Figure 6.3**). Hempseed protein network was characterized by five groups, among which the most compact was integrally constituted by ribosomal proteins (RPS7, RPS8, RPS11, rps15), whereas the others were characterized by genes involved in embryo development, vegetative growth, glucose sensing during plant development (GAPC2, TPS1, ERD1), and the kinase activity (LYK2, GK-1).

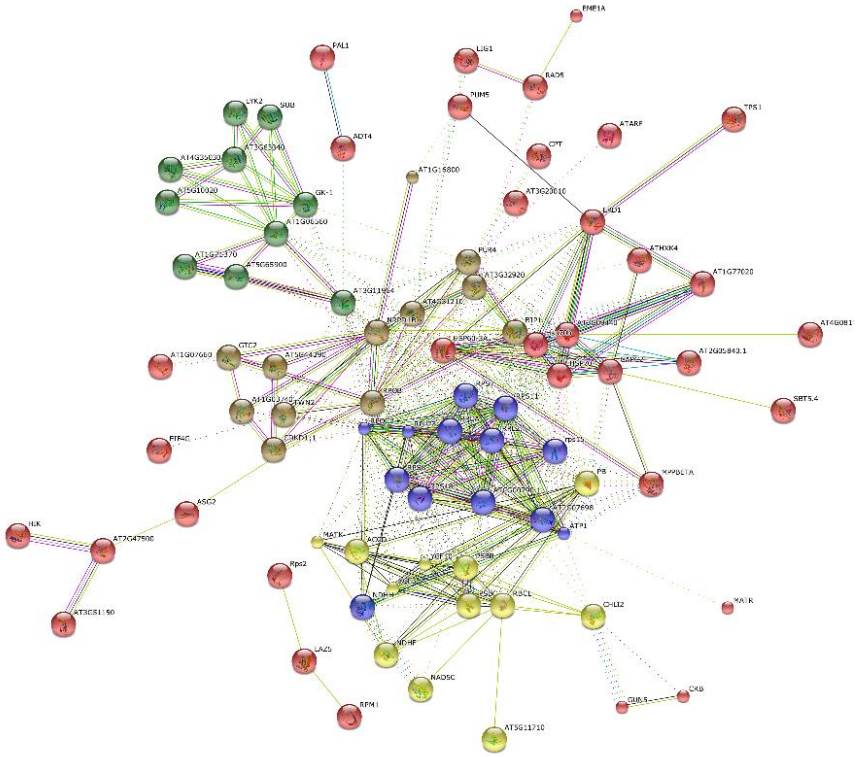


Figure 6.3 - Protein-protein interactomic network of the hempseed obtained by clustering K-MEANS using STRING software. Different coloured edges represent the existence of different types of evidence. A green line indicates neighbourhood evidence; a blue line, concurrence evidence; a purple line, experimental evidence.

Furthermore, in order to investigate biomolecular pathway maps, the KEGG-Pathway Enrichment was obtained by STRING (**Table 6.2**). Confirming our gene ontology results, this procedure revealed that established pathways include energy metabolism, carbohydrate, lipid and nucleotide metabolism, as well as amino acid metabolism and biosynthesis. The richest category is attributable to metabolic processes, representing 67.1% of all analysed proteome dataset. The second most populated one is related to genetic information processes (37.1%), including proteins involved in DNA transcription, translation, and replication and repair activities. Mainly ribosomal proteins, the most expressed proteins in our dataset, belong to this group.

Table 6.2 KEGG-Analysis achieved using STRING software: list of established pathways and number of genes in which they are involved.

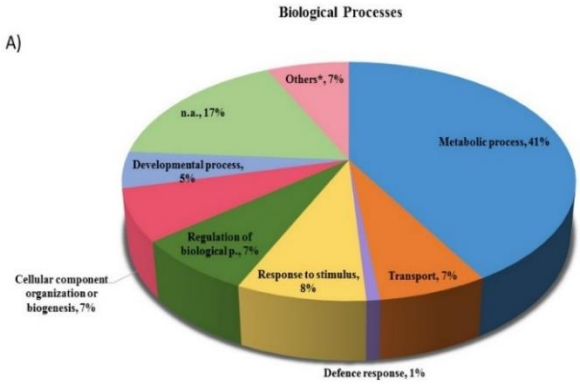
KEGG-Pathways	N. of genes
Metabolic pathways (67.1%)	
Carbohydrate metabolism	
Pyruvate metabolism	2
Starch and sucrose metabolism	2
Glyoxylate and dicarboxylate metabolism	1
Fructose and mannose metabolism	1
Galactose metabolism	1
Glycolysis / Gluconeogenesis	1
Propanoate metabolism	1
Inositol phosphate metabolism	1
Amino sugar and nucleotide sugar metabolism	1
Energy metabolism	
Oxidative phosphorylation	4
Photosynthesis	3
Carbon fixation in photosynthetic organisms	1
Lipid metabolism	
Fatty acid biosynthesis	1
Fatty acid metabolism	1
Nucleotide metabolism	
Purine metabolism	5
Pyrimidine metabolism	3
Amino acid metabolism and biosynthesis	
Cysteine and methionine metabolism	3
Phenylalanine metabolism	1
Phenylalanine, tyrosine and tryptophan biosynthesis	1
Biosynthesis of amino acids	2
Other amino acid metabolism	
Selenocompound metabolism	1
Metabolism of cofactors and vitamins	
Porphyrin and chlorophyll metabolism	2
Metabolism of terpenoids and polyketides	
Terpenoid backbone biosynthesis	1
Biosynthesis of secondary metabolites	
Phenylpropanoid biosynthesis	1
Carbon metabolism	3
Microbial metabolism in diverse environments	3
Genetic information processing (37.1%)	
Transcription	
RNA polymerase	3
Spliceosome	3
Basal transcription factors	1
Translation	
Ribosome	6
RNA transport	2
Aminoacyl-tRNA biosynthesis	1
Replication and repair	
DNA replication	1
Nucleotide excision repair	2
Homologous recombination	1
Mismatch repair	1
Base excision repair	1
Folding, sorting and degradation	
Protein processing in endoplasmic reticulum	3
RNA degradation	1
Environmental Information Processing (2.9%)	

Signal transduction	
Plant hormone signal transduction	1
Phosphatidylinositol signaling system	1
Transport and catabolism	
Endocytosis	4
Enviromental adaptation	
Plant-pathogen interaction	1

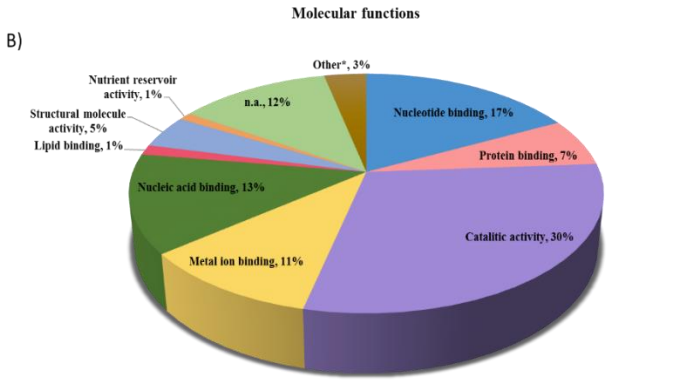
6.3.3 Protein biology

By grouping all identified hempseed proteins on the basis of their supposed molecular function, cellular location and biological processes in which they are involved, Gene Ontology analysis has allowed to ascertain the *C. sativa* proteome functionality. **Figure 6.4** reports the classification of the 181 identified species through Gene Ontology analysis into biological processes (**Figure 6.4A**), molecular functions (**Figure 6.4B**), and cellular components (**Figure 6.4C**). Nevertheless, protein location shows a certain heterogeneity, since 14% of all protein identified is found in the chloroplasts, 15% in the membranes, 10% in the plastids, 7% in the nucleus (**Figure 6.4C**). Proteins involved in glycolysis, such as Phosphoglycerate kinase and Glyceraldehyde-3-phosphate dehydrogenase (GAPDHs) were identified. The latter is a key enzyme during glycolysis and its expression is up-regulated during all four seed germination stages, as recently demonstrated for wheat seeds (He, Zhu, Dong, Zhang, Cheng, Li, et al., 2015). In addition to its role as a cytosolic glycolytic enzyme, GAPDHs has been implicated in embryo development, pollen development, root growth, and mediation of reactive oxygen species response (Guo, Devaiah, Narasimhan, Pan, Zhang, Zhang, et al., 2012). Moreover, 14% of metabolic proteins are involved in translation activities and 12% in oxidation-reduction processes, such as Cytochrome c (CYC_CANSA), NADH-plastoquinone oxidoreductase subunit (ndhF_CANSA), and NADH dehydrogenase (nad5_CANSA). This last is a well-known electron transport oxidoreductase enzyme, which mediates several physiological processes including signalling and regulation of growth, transport activities, and defence against pathogens (Lüthje, Döring, Heuer, Lüthen, & Böttger, 1997). The photosynthetic-protein group includes Photosystem II CP43, CP47 and Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). Some proteins belonging to the category "response to stimulus" and directly involved in stress response were also detected, such as the Heat shock protein family (HSP70 and HSP60). The members of this wide glycoprotein gene family, among which HSP70 is the most conservative, prevent aggregation and assist refolding of non-native proteins under both normal and stress conditions (Frydman, 2001). Although the

production of HSPs has been often described under heat stress, their expression and function in fundamental processes, such as seed development, have been also confirmed (Koo, Park, Kim, Suh, Lee, Lee, et al., 2015). Generally, HSPs are highly expressed in mature and dry seeds, as in the late stages of seed development, whereas they tend to disappear during germination.



*Others: Cell growth (0.4%), Cell differentiation (0.8%), Cell division (1.6%), Cell death (0.4%), Cellular homeostasis (0.4%), Cellular component movement (0.8%), Cell cycle (1.2%), Reproductive process (1.2%)



*Others: Phospholipid binding (0.9%), Transporter activity (0.6%), Cofactor binding (0.9%), Chlorophyll binding (0.6%)

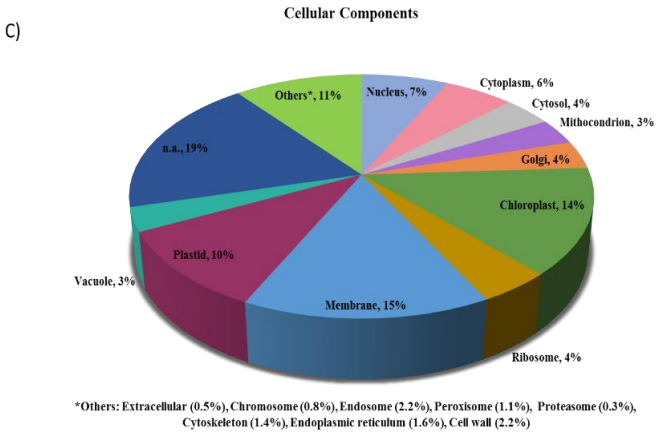


Figure 6.4 - Pie charts of (A) biological processes, (B) molecular functions, and (C) cellular components refer to the 181 unique gene products described in hempseeds.

6.4 Conclusion

Taking advantage of high-throughput technologies, this work has provided very innovative information on hempseed proteome, since it was possible to identify in total 181 unique gene products, 56 specific protein species identified through the *C. sativa* database search and 125 protein species through the *A. thaliana* database search. This represents a main improvement in respect to available literature were all assignments were based on other plant databases (Park, Seo, & Lee, 2012). In addition, it provides the first interactomic map of *C. sativa*, composed by a complex grid formed by 137 nodes connected via 410 interactions, useful to explore proteins relations and to understand the *C. sativa* biological pathways. The relevance of this work is underlined by the increasing interest for including hempseed ingredients in regular foods and, potentially, in functional ones. In fact, recent investigantions have shown that peptides deriving from the gastrointestinal simulated digestion of hempseed proteins are characterized by useful antioxidant (Girgih, He, Malomo, Offengenden, Wu, & Aluko, 2014) and antihypertensive properties (Girgih, Alashi, He, Malomo, & Aluko, 2014).

Acknowledgement: We thank dr. Incoronata Galasso who has provided hempseeds, Sara Mariani for her technical support, and Alpro Foundation for a PostDoc fellowship to CZ.

Conflicts of interests: the Authors declare non conflicts of interest.

SUPPORTING INFORMATION

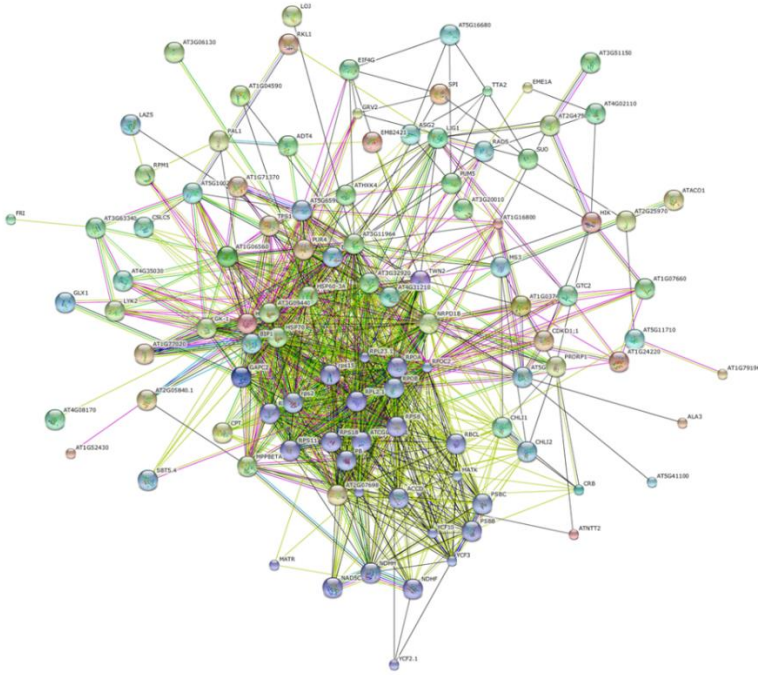


Figure 6.1S - The protein functional interactomic network of the hempseed proteins obtained by STRING software. Different colored edges represent the existence of different types of evidence. A green line indicates neighborhood evidence; a blue line, concurrence evidence; a purple line, experimental evidence.

Table 6.1S A) List of 56 unique gene products identified by *C. sativa* database search.

Accession	Description OS= *C. sativa*	MS/MS Search Score	ΣCoverage	Σ# Peptides	Σ# Unique Peptides	TPSI	Score pH2.2	Coverage pH2.2	# Peptides pH2.2	Score pH7.2	Coverage pH7.2	# Peptides pH7.2	Score RAW	Coverage RAW extract	# Peptides RAW extract	# AAs	calc. pI
A7IZZ2	(+)-alpha-pinene synthase	12.58	6.9	7	7	6.12E+07	6.85	5.20	5	4.80	3.40	2	4.37	4.30	2	615	5.9
A7IZZ1	(-)-limonene synthase	6.35	6.5	2	2	8.51E+06	3.20	4.30	2	3.15	2.20	2				622	6.4
H9A1V3	Acyl-activating enzyme 1	9.41	5.6	5	4	5.31E+07	7.24	5.60	2	7.43	2.70	3				720	6.9
H9A1V4	Acyl-activating enzyme 2	9.33	5.4	4	4	1.24E+07	9.33	5.40	2				5.65	2.20	2	662	6.7
H9A1V5	Acyl-activating enzyme 3	11.25	13.4	5	4	3.22E+08							11.25	5.70	2	543	8.7
H9A1V8	Acyl-activating enzyme 6	11.95	4.9	2	2	4.98E+07							11.95	4.90	2	569	7.9
H9A1V9	Acyl-activating enzyme 7	8.98	4.8	5	3	3.00E+07	5.37	3.10	3				3.61	4.80	2	597	6.2
H9A1W1	Acyl-activating enzyme 9	11.04	5.8	2	2	4.02E+07	5.34	2.10	2				5.70	2.70	2	597	6.2
H9A1W3	Acyl-activating enzyme 11	10.79	5.6	4	2	5.68E+07	7.24	5.60	3	3.55	3.20	4				334	6.5
H9A8L1	Acyl-activating enzyme 12	10.10	3.1	3	3	4.17E+07	5.52	3.10	2				4.58	2.70	2	757	6.9
H9A8L2	Acyl-activating enzyme 13	6.83	3.3	4	3	3.47E+07	3.63	1.9	2				3.20	3.30	2	715	6.4
H9A8L3	Acyl-activating enzyme 14	11.50	4.9	2	2	6.18E+07							11.50	2.70	2	727	6.7
H9A8L4	Acyl-activating enzyme 15	22.34	9.2	8	7	1.32E+07				8.82	1.1	2	13.20	15.70	4	776	5.9
A0A0C5APY2	Acetyl-CoA carboxylase carboxyltransferase beta subunit	11.48	4.6	4	3	5.07E+08	6.38	1.60	2	4.47	1.60	3	5.56	3.00	2	497	4.8
A0A0C5ARS5	ATP synthase CF1 beta subunit	16.83	15.8	4	3	2.56E+08	16.83	15.80	3							498	5.2
E5DK51	Atp1, partial	31.75	15.4	6	4	1.87E+08	31.75	15.40	6							349	7.1
I6XT51	Betv1-like protein	27.16	21.1	11	3	1.09E+09	9.52	7.40	11							161	5.2
A0A0E3TJM8	Cannabidiolic acid synthase homolog	11.94	8.4	2	2	9.94E+06	5.58	4.50	2	6.36	3.80	2				544	8.8
C6KI62	Chalcone synthase-like protein 1	11.17	8.8	3	3	7.27E+07	3.70	2.00	2	5.30	4.50	2	5.79	4.30	2	392	6.5
I6WIE9	Chalcone isomerase-like protein	8.43	11.1	2	2	9.61E+06				3.10	4.60	2	5.33	6.50	2	214	8.2
A0A0C5ARS7	Chloroplast envelope membrane protein	4.38	5.6	2	2	4.68E+06				4.38	5.60	2				231	6.1
P00053	Cytochrome c	7.7	27.9	2	2	7.89E+06							7.7	27.9	2	111	9.8
A0A088MFF4	Delta 12 desaturase	15.28	9.3	4	2	5.49E+08							15.28	9.30	4	387	8.7
A0A090DLH8, A0A090CXP7	Edestin 1 GN=ede1A, ede1C-ede1B	312.36	46.9	407	21	5.69E+10	128.11	17.60	75	137.96	27.00	79	273.67	46.90	217	511	8.8
A0A090CXP9	Edestin 2 GN=ede2A	120.43	22.8	66	17	2.32E+09	44.93	8.90	15	91.66	19.10	26	200.14	32.70	121	491	8.9
A0A090DLI7, A0A090CXP8	Edestin 2 GN=ede2B-ede2C	132.45	25.8	70	18	2.41E+09	44.93	8.90	15	103.68	22.10	29	214.66	35.80	135	491	9.0
A0A0C5APZ4	Hypothetical chloroplast RF21	48.36	6.2	25	17	1.10E+09	26.02	2.90	12	5.02	0.70	7	5.62	0.70	6	2302	8.7
A0A0C5AUH7	Hypothetical chloroplast RF34	12.24	13.6	4	4	2.48E+08				3.64	11.30	2	12.24	13.60	3	168	5.6
L0N5C8	Hypothetical protein	16.57	18.3	6	3	9.46E+07	11.45	18.30	4	5.12	3.10	2				168	5.6
E5DKP2	MatR, partial	16.27	10.2	5	5	3.65E+07				4.27	2.80	2	3.50	4.40	3	535	10.2
A0A0C5AUE3	Maturase K	15.62	11.8	4	4	7.67E+06							4.71	11.80	4	509	9.4
A0A0C5AUJ6	NADH-plastoquinone oxidoreductase subunit 5	8.21	2.7	3	2	5.99E+07	8.21	2.70	3							755	9.1
A0A0C5AS19	NADH-plastoquinone oxidoreductase subunit 7	11.49	12.8	4	3	1.82E+08	3.66	4.00	2	3.25	2.20	2	4.58	6.60	3	393	5.2
E5DL82	NADH dehydrogenase subunit 5, partial	12.45	7.7	3	3	2.16E+08	3.24	7.70	3	6.36	4.10	2	7.45	7.10	2	361	8.5

V5KWZ6	Phenylalanine ammonia lyase	9.47	8.3	3	3	1.14E+07	4.48	3.20	2	4.63	2.90	2	4.39	1.60	2	707	6.0
A0A0C5AUH1	Photosystem II CP43 chlorophyll apoprotein	12.04	8.4	3	3	2.57E+07	5.50	3.30	2	6.54	5.00	2				473	6.6
A9XV91	Photosystem II CP47 chlorophyll apoprotein, partial	4.94	6.3	5	3	3.03E+06	4.10	4.30	3				4.94	6.30	2	488	6.1
F1LKH7	Polyketide synthase protein	5.15	4.1	2	2	1.56E+07							5.15	4.10	3	385	6.0
U6EFF4	Putative LysM domain containing receptor kinase	5.80	2.8	4	2	4.91E+07	6.00	2.80	2	5.80	2.80	2	4.68	3.60	2	599	6.6
A0A0C5AS13	Ribosomal protein L2	7.11	3.6	7	3	1.79E+08	6.58	3.60	2	5.62	3.60	3	6.60	3.60	2	275	10.9
A0A0C5AUJ2	Ribosomal protein L16	5.20	18.4	2	2	8.09E+07	3.01	9.20	2	5.20	18.40	2				119	11.3
A0A0C5AS20	Ribosomal protein L23	20.04	55.9	6	5	2.44E+08	10.66	37.60	4				13.08	36.50	2	93	10.4
A0A0C5ART6	Ribosomal protein S3 (chloroplast)	11.08	23.8	6	5	2.10E+08	5.19	5.00	2	10.51	17.40	3	5.32	2.80	2	155	9.7
E5DLX2	Ribosomal protein S3, partial (mitochondrion)	5.32	2.8	10	8	5.41E+07	3.83	3.40	4	4.91	3.20	5	5.32	2.80	3	494	10.2
A0A0C5ARS4	Ribosomal protein S4	7.15	13.3	4	3	7.66E+07	3.56	4.90	2				3.59	8.40	2	202	10.4
A0A0C5ARU3	Ribosomal protein S7	6.49	23.8	2	2	3.08E+07	6.49	23.80	2							155	11.2
A0A0C5APY7	Ribosomal protein S8	5.50	13.4	2	2	3.68E+07	3.49	6.70	2				5.70	13.40	2	134	11.1
A0A0C5ART4	Ribosomal protein S11	10.26	16.6	2	2	2.40E+07	9.81	16.60	2							138	12.1
A0A0C5AUK1	Ribosomal protein S15	8.03	32.2	7	2	7.74E+06	4.36	11.10	4	3.60	11.10	2	3.67	21.10	2	90	10.7
A0A0C5ART1	Ribosomal protein S18	15.27	30.6	5	2	3.26E+08	9.23	15.80	13	6.04	18.80	2				101	11.9
A0A0C5B2I6	Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit	20.65	6.1	8	6	4.00E+08	9.24	4.40	3	6.47	1.60	3	5.68	2.30	2	475	6.1
I0B4C2	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial	9.77	11.5	7	3	3.96E+08	3.30	7.80	4	6.47	3.70	3				216	7.8
A0A0C5AUI9	RNA polymerase alpha subunit	14.89	19.5	4	4	1.48E+07				6.16	7.45	2	8.30	12.00	2	331	7.7
A0A0C5ARQ8	RNA polymerase beta subunit	17.38	5.7	4	3	9.68E+07				9.48	4.75	3	5.55	4.05	2	1070	8.7
A0A0C5ARX9	RNA polymerase beta' subunit	17.28	4.8	13	9	1.04E+09	10.52	3.00	7	8.48	1.60	2	12.40	2.90	4	1393	9.1
I1V0C2	Tetrahydrocannabinolic acid synthase, partial	11.62	9.9	3	3	2.87E+07	5.58	4.50	2				6.83	5.30	2	545	8.7

Total Protein Spectral Intensity, TPSI

Table 6.1S B) List of 125 proteins identified by the use of *A. thaliana* database search.

Accession	Description OS=*A. thaliana*	MS/MS Search Score	ΣCoverage	Σ# Peptides	Σ# Unique Peptides	TPSI	Score pH2,2	Coverage pH2,2	# Peptides pH2,2	Score pH7.2	Coverage pH7.2	# Peptides pH7.2	Score RAW	Coverage RAW	# Peptides RAW	# AAs	calc. pI
O65719	Heat shock 70 kDa protein 3	88.85	18.3	12	7	3.97E+08	36.70	7.0	3	13.16	2.5	2	52.10	11.1	8	649	4.97
Q9LHA8	Heat shock 70 kDa protein 4	74.67	16.4	11	6	3.37E+08	22.52	5.2	3	13.16	2.5	2	52.20	11.1	8	650	5.14
Q9S9N1	Heat shock 70 kDa protein 5	58.03	9.2	20	4	2.54E+08				13.16	86.7	2	58.03	9.2	19	646	5.30
Q9LKR3	Heat shock 70 kDa protein 11	21.72	4.6	6	2	4.41E+07				21.48	4.6	2	13.40	2.3	4	669	5.08
Q8LFV7	Phosphoglycerate kinase	45.26	14.7	6	3	1.54E+08	24.22	11.4	3	12.19	3.5	2	34.40	7.4	3	401	5.49
Q9LHE4	Transcription factor-like protein	28.86	5.5	7	3	2.81E+08	17.17	2.9	4	11.17	1.1	2	10.91	2.5	2	1047	8.80
Q0WR59	Probable inactive receptor kinase	26.37	6.3	3	3	3.99E+08	8.38	1.8	2				17.99	4.5	3	1048	6.21
Q0WNZ5	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase 3	24.95	3.6	6	2	1.14E+08							24.95	3.6	6	812	8.44
F4HZB2	WD/BEACH domain protein SPIRRIG	23.52	1.4	3	3	3.69E+07	6.15	0.4	2				17.37	1.1	2	3601	5.75
O80891	Cellulose synthase-like protein B4	22.27	8.4	3	3	1.39E+08	22.27	8.4	3							755	7.29
Q9LIM5	Chloroplast protein HCF243	21.45	3.6	6	2	5.91E+08	21.45	3.6	6							684	6.04
O23058	A_IG005I10.23 protein	20.30	5.1	2	2	7.11E+07	9.83	1.9	2	10.47	3.1	2				770	5.65
O81490	F9D12.11 protein	20.10	3.5	3	3	5.73E+07	13.59	2.1	3							1322	9.34
Q5XF33	Magnesium-chelatase subunit ChlI-2	20.09	6.8	2	2	1.19E+08	10.69	4.2	2				9.40	2.6	2	423	5.36
Q9M8D3	Phosphoribosylformylglycinamidine synthase	19.49	3.9	3	3	1.36E+08	19.49	3.9	3							1407	5.21
Q9SHM9	F7F22.11	19.34	6.0	2	2	6.93E+07				10.84	2.7	2	8.50	3.3	2	612	9.64
P59583	Probable WRKY transcription factor 32	18.97	8.1	3	2	1.19E+09	18.97	8.1	2				9.59	2.5	2	466	5.75
Q9SSQ2	F6D8.29	18.35	7.8	5	2	7.23E+08							18.35	7.8	5	423	9.80
Q9SIL1	Putative TNP1-like transposon protein	18.24	2.5	3	2	1.30E+09	8.75	1.4	2	9.49	1.1	2	7.14	1.4	2	984	6.71
Q9SU68	Putative protein	11.07	2.4	2	2	1.08E+07	18.16	2.4	2							1613	5.93
Q9ZQQ4	Hypothetical protein	21.72	5.3	2	2	4.41E+07				8.64	2.3	2	9.45	2.9	2	679	5.71
Q6NR90	Histone H4 - like protein	17.99	21.3	3	2	4.88E+07							17.99	21.3	3	103	11.48
Q66GI4	Proteinaceous RNase P 1	17.93	5.4	2	2	2.70E+07				9.26	1.5	2	8.67	3.9	2	572	9.64
Q9FG23	At5g06830	17.93	4.5	3	2	2.15E+08							17.52	4.5	3	549	4.72
Q84TI3	PHD-finger TITANIA 2	17.89	4.0	5	2	2.28E+09	14.67	4.0	3	10.26	2.2	2				1162	5.87
P56785	ycf1	17.82	1.9	2	2	6.01E+07							17.82	1.9	2	1786	9.85
C0Z2I0	Glyceraldehyde-3-phosphate dehydrogenase	17.81	10.2	2	2	3.09E+07							17.81	10.2	2	343	8.46
O48689	Paired amphipathic helix repeat-containing protein	17.80	5.2	3	2	2.74E+08	8.86	3.3	2				8.94	1.9	3	744	9.83
O04251	Transcription coactivator protein	17.66	3.0	3	2	2.69E+08	17.66	3.0	3							1329	8.75
O80958	Pentatricopeptide repeat protein	17.55	4.3	2	2	7.32E+07	8.64	1.6	2				8.91	2.4	2	867	6.06
F4J8K6	Protein ribosomal RNA processing 5	17.44	1.2	2	2	4.90E+08							17.44	1.2	2	1896	8.75
O82762	KH domain-containing protein	17.40	5.3	2	2	4.46E+07	9.79	2.5	2	7.61	2.8	2				632	5.32
O23026	EST gb|T04104 comes from this gene/T1G11.16 protein	17.38	5.2	3	2	9.94E+07	8.55	5.0	2				8.83	5.3	3	398	9.11
O04478	Endoglucanase 7	17.37	5.4	9	2	4.47E+08	9.72	3.5	3	10.07	3.5	3	16.05	5.4	3	623	9.44
Q38810	cDNA-5-encoded protein, partial	17.37	14.5	2	2	2.35E+07	9.59	5.5	2	7.78	9.0	2				289	9.15
Q9FH23	Uncharacterized protein	9.83	2.1	2	2	2.78E+07	8.42	1.0	2				8.89	1.1	2	1487	5.79
Q9FWR3	Similar to tRNA-splicing endonuclease positive effector SEN1	17.24	1.5	2	2	2.18E+07	8.76	0.8	2	8.48	0.7	2				2142	8.52
Q84M98	Meiotic endonuclease 1A	17.11	5.8	2	2	2.68E+07	6.99	3.8	2	10.12	2.0	2				549	8.43
Q8VZF6	AT5g45560/MFC19_23	17.04	6.6	2	2	9.36E+07				8.84	2.9	2	8.20	3.6	2	719	6.21
Q94FQ0	Glycine-rich protein GRP17	16.95	12.1	2	2	4.38E+07				7.60	5.2	2	9.35	6.9	2	543	10.38

P42762	Erd1 protein precursor	16.63	4.3	2	2	6.59E+07	9.40	1.9	2				7.23	2.4	2	945	5.89
Q9M8K5	AT3g06130/F28L1_7	16.57	9.5	2	2	1.44E+08	7.67	4.6	2				8.90	4.9	2	473	8.34
F4IMB5	ATPase, F1 complex, alpha subunit protein	16.56	4.1	2	2	2.69E+07	10.21	1.2	2				6.35	2.8	2	777	5.43
Q8S8D3	Phox (PX) domain-containing protein	16.36	4.6	2	2	6.67E+07				6.72	2.3	2	9.64	2.3	2	643	5.09
O80458	Cis-prenyltransferase	16.32	8.9	2	2	2.00E+07	6.30	4.0	2				10.02	4.9	2	303	8.57
Q84W97	Putative kinesin	16.21	3.2	3	2	2.11E+08	16.21	3.2	2	6.99	1.4	2				983	8.71
Q9SA52	g5bf	16.17	6.6	2	2	2.35E+07				9.11	4.5	2	7.06	5.0	2	378	8.36
Q9FNB0	Magnesium-chelatase subunit H	20.09	2.0	3	2	1.19E+08	9.65	1.0	2	6.50	0.9	2	9.49	1.0	2	1381	5.80
Q9LXU4	Uncharacterized protein	9.83	4.4	3	2	2.78E+07	6.28	1.7	2	16.01	4.4	2				861	6.68
Q5D869	RNA polymerase IV largest subunit	15.88	2.0	2	2	4.70E+07	8.95	0.9	2	6.93	1.1	2				1976	5.88
Q9LQ91	T1N6.6	15.76	3.4	2	2	3.49E+07	8.73	1.7	2				7.03	1.8	2	832	9.12
Q9SB75	Cellulose-synthase-like C5	15.73	3.4	3	2	1.06E+08	9.26	2.1	2				6.47	1.3	2	692	9.10
Q9C533	Probable protein S-acyltransferase 22	15.69	5.7	3	2	3.81E+08				8.28	2.3	2	7.41	3.4	3	596	9.59
Q8L7G0	Auxin response factor 1	15.66	4.5	2	2	2.78E+08							15.66	4.5	3	665	5.84
Q76E23	Eukaryotic translation initiation factor 4G	15.63	1.5	2	2	5.56E+07	9.51	0.6	2				6.12	0.9	2	1727	7.59
Q9XIE6	Phospholipid-transporting ATPase 3	15.59	2.9	2	2	5.28E+08	15.59	2.9	2							1213	8.30
Q19DM3	Disease resistance protein, partial	15.58	17.5	2	2	1.01E+08	15.58	17.5	2							222	9.16
O81893	Inositol-tetrakisphosphate 1-kinase 2	15.53	7.6	3	2	5.90E+08	6.14	3.8	2				9.39	3.8	2	391	6.15
Q9FHW3	Unnamed protein product	88.85	30.3	2	2	3.97E+08				6.52	15.2	2	8.93	15.1	2	132	6.98
Q9ZQE9	Putative retroelement pol polyprotein	15.42	1.1	2	2	4.52E+07				7.33	0.9	2	8.09	1.1	2	1347	8.46
F4JXC5	Subtilisin-like protease SBT5.4	15.41	6.6	2	2	5.91E+07	8.26	4.0	2				7.15	2.6	2	778	9.24
O82496	T12H20.15 protein	15.41	3.6	4	2	1.26E+08	14.47	3.6	2	9.19	3.1	2	7.10	3.1	2	705	5.53
F4IAH9	Glyoxalase I homolog	15.34	8.3	2	2	5.52E+07	7.96	5.0	2				7.38	3.4	2	322	5.40
O48701	Hypothetical protein]	15.38	5.3	2	2	4.41E+07	9.15	1.3	2				6.23	4.0	2	707	8.36
Q9LR53	Putative protein kinase	15.26	5.5	2	2	5.81E+07	7.94	2.3	2	7.32	3.1	2				740	9.64
P0C7Q8	Protein DA1	15.21	8.4	2	2	2.12E+08							15.21	8.4	2	532	5.98
Q8L622	Putative uncharacterized protein	15.19	7.0	2	2	1.21E+08	15.19	7.0	2							607	5.12
Q9S755	Hypothetical protein	15.18	15.6	2	2	5.07E+07	8.62	9.6	2	6.56	6.0	2				198	9.45
O64533	YUP8H12R.20	15.15	2.7	2	2	1.19E+08							15.15	2.7	3	1325	7.77
Q93ZM7	Heat shock protein 60-3A	15.07	7.6	2	2	2.55E+08				8.26	2.6	2	6.81	5.0	2	572	5.85
Q42290	Probable mitochondrial-processing peptidase subunit beta	15.05	9.5	4	2	2.08E+09	14.87	9.5	2	13.52	4.7	2				535	6.45
Q8S905	F15H18.12	15.05	2.2	2	2	6.51E+07	8.71	0.9	2				6.34	1.4	2	1003	8.57
Q42572	DNA ligase 1	15.17	4.5	2	2	4.09E+08				6.10	1.6	2	9.07	2.9	2	839	9.08
O49281	Zinc finger protein	15.04	2.9	2	2	1.21E+08	15.04	2.9	2							1108	9.19
Q38K61	Disease resistance protein, partial	15.58	12.8	2	2	1.01E+08	15.58	12.8	2							272	6.09
P93736	Valine--tRNA ligase	15.01	3.3	2	2	1.29E+08	6.15	1.5	2				8.95	1.8	2	1115	6.70
F4IVL6	Gravitropism defective 2	15.01	0.9	2	2	8.65E+07							15.01	0.9	2	2554	5.74
Q8VY07	Clathrin binding protein-like	14.98	5.0	2	2	1.70E+07							14.98	5.0	2	577	4.87
Q9SSP7	Ubiquitin carboxyl-terminal hydrolase-related protein	14.95	3.2	2	2	3.05E+07							14.95	3.2	2	1136	5.72
Q9C9U2	Cell division protein kinase	14.93	8.0	2	2	4.26E+08				8.63	4.0	2	6.30	3.8	2	398	9.49
Q9SB89	DEAD-box ATP-dependent RNA helicase 27	14.86	3.1	2	2	2.43E+07							14.86	3.1	2	633	8.99
Q8GXD6	DEAD-box ATP-dependent RNA helicase 49	13.00	4.1	2	2	3.35E+07				13.00	4.1	2				558	9.51
Q9LJX4	RNA binding protein-like	14.83	3.0	2	2	1.01E+08	14.83	3.0	2							962	5.67
F4K7T1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	17.24	4.5	3	2	2.18E+07	14.24	4.5	3							661	6.44
Q56XE8	Hexokinase-4	14.76	3.8	2	2	2.30E+07	8.49	1.6	2	6.27	2.2	2				502	8.53
A0MEZ6	Putative uncharacterized protein	14.67	21.0	2	2	2.54E+07	8.25	9.0	2				6.42	15.0	2	133	6.86
B0LC03	RPS2	14.59	4.7	2	2	7.18E+07	7.64	2.2	2				6.95	2.5	2	909	6.13
B9DG14	AT5G41100	14.58	8.0	2	2	2.34E+07				14.58	8.0	2				586	8.79

F4J3S1	Shugoshin C terminus	14.52	7.3	2	2	3.82E+07	7.98	2.3	2				6.54	4.0	2	572	9.56
Q42560	Aconitate hydratase 1	14.45	4.4	2	2	6.91E+07	6.64	1.9	3				7.81	2.5	2	898	5.98
F4K4M9	Uncharacterized protein	14.37	4.7	2	2	1.75E+07							14.37	4.7	3	848	9.09
F4K4D6	PWWP domain-containing protein	14.29	4.1	2	2	8.59E+07	14.29	4.1	2							1072	6.34
Q9ZVX5	Pentatricopeptide repeat-containing protein	14.20	4.1	2	2	5.66E+07	14.20	4.1	2							743	6.06
O81147	Proteasome subunit alpha type-6-B	14.18	13.0	2	2	3.69E+07	7.27	5.7	2	6.91	7.3	2				246	5.75
Q9FNI6	DNA repair protein RAD5 protein	14.13	3.4	3	2	5.61E+07				14.13	3.4	3				1029	6.41
F4KE59	RING/FYVE/PHD zinc finger-containing protein	14.12	2.5	2	2	9.25E+07	7.82	1.4	2				6.30	1.1	2	1311	6.29
Q8VZG2	Protein HYPER-SENSITIVITY-RELATED 4	14.06	5.0	2	2	2.94E+07							14.06	5.0	2	576	6.30
Q9SCZ3	TIR-NBS-LRR class disease resistance protein	14.01	2.7	3	2	4.54E+07	6.68	1.4	2	7.33	1.3	3				1253	9.08
X2L4T4	DM2D	13.98	4.1	2	2	1.18E+08							13.98	4.1	2	1216	6.18
Q56X77	Guanylate kinase	13.96	15.7	2	2	1.77E+07	13.96	15.7	2							203	5.78
Q8VZQ9	Putative protein	13.92	2.6	2	2	1.44E+08	13.92	2.6	2							1021	9.68
Q94BQ3	WD and tetratricopeptide repeats protein 1	13.78	4.7	2	2	1.45E+08							13.78	4.7	2	757	5.54
Q9FLR3	No apical meristem (NAM) -like protein	13.71	8.8	2	2	4.41E+07							13.71	8.8	2	329	9.17
Q9T0F7	Glucuronoxylan 4-O-methyltransferase 2	13.71	11.7	2	2	3.14E+07				13.71	11.7	2				290	6.76
Q3EAS6	Putative DNA repair protein recA homolog 4	13.70	24.3	2	2	1.86E+09	6.90	11.5	2	6.80	12.8	2				226	9.39
P17094	Ribosomal protein	13.66	11.8	2	2	2.79E+07							13.66	11.8	2	389	10.29
Q9FKV9	Cyclin-dependent protein kinase-like protein	13.62	3.7	3	2	8.73E+07	13.62	3.7	3				6.55	2.0	2	644	9.31
F4I5J7	DNAJ heat shock N-terminal domain-containing protein	13.52	13.2	2	2	2.60E+08	6.92	6.1	2				6.60	7.1	2	379	5.39
F4J394	ATP binding microtubule motor family protein	13.49	3.2	2	2	1.45E+07	6.29	1.4	2	7.20	1.8	2				1054	6.24
Q84MA1	Hypothetical protein	13.49	5.5	2	2	1.84E+08	13.49	5.5	2							599	6.64
Q8S8L1	Uncharacterized protein	13.45	14.8	2	2	6.78E+08	7.38	7.1	2				6.07	7.7	2	336	6.04
Q9FHT3	Uncharacterized protein	13.35	5.7	2	2	1.27E+08	13.35	5.7	3							485	6.00
Q93YS2	Putative protein phosphatase 2C 51	13.27	3.0	2	2	1.05E+08	6.94	1.6	2				6.33	1.4	2	1075	6.17
G4WT86	Late flowering protein	13.04	3.9	2	2	5.81E+07	6.26	2.1	2				6.78	1.8	2	609	9.27
O49610	Protein kinase-like	12.97	10.1	2	2	3.06E+08	12.97	10.1	2							461	6.38
Q9LPA7	T32E20.13	12.91	2.5	2	2	4.15E+07	6.78	1.0	2				6.13	1.5	2	1335	6.15
P92935	ADP,ATP carrier protein 2	12.90	5.5	2	2	1.09E+08	6.10	2.1	3	6.80	3.4	2				618	9.70
C0SVH8	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein	12.73	3.0	2	2	3.90E+07	6.22	2.1	2				6.51	0.9	2	1138	9.16
F4IF81	Putative serine/threonine protein kinase	12.70	5.5	2	2	5.85E+07							12.70	5.5	2	687	5.72
Q56Y19	Uncharacterized protein	12.56	6.4	2	2	7.95E+07	6.09	3.7	2	6.47	2.7	2				437	9.83
Q8S8N9	Golgin candidate 1	12.40	5.4	2	2	1.79E+07				12.40	5.4	2				713	5.18
O22241	Arogenate dehydratase 4	12.33	4.9	2	2	1.19E+07				12.33	4.9	2				424	6.05
Q9FGN6	Receptor-like kinase MOL1	12.32	4.4	2	2	1.05E+07	12.32	4.4	2							895	6.60
F4JRX3	DNA topoisomerase, type IA, core	12.30	1.4	2	2	6.32E+06				6.10	0.9	2	6.20	0.5	2	1284	9.54
Q9SG92	Methylesterase 17	12.19	11.9	2	2	3.68E+07	12.19	11.9	2							276	5.08

Total Protein Spectral Intensity, TPSI

Table 6.1S C) List of identified peptides belonging to *C. sativa* proteins.

N.	Accession N.	Description	SPI(%)	Score	Average Mass (kDa)	Σ# Unique Peptides
1	**A7IZZ2**	**(+)-alpha-pinene synthase**			72.35	7
		SRSRPNCNPIQCSLAKSPSSDTSTIVR	72.6	4.37		
		SITNEALESLQEGHHAACRQGSLMLR	65.7	6.80		
		WFYSGYKPTLKK	61.3	3.78		
		QGSLMLRLADDLGTLSDEmKRGDVPK	50.0	3.98		
		MSLFmYQHGDGHASQDSHSRK	61.6	4.80		
		AEAKWFIEEYEKTQDK	76.3	5.78		
		qGSLMLRLADDLGTLSDEmK	52.2	3.07		
2	**A7IZZ1**	**(-)-limonene synthase**			72.89	2
		TGEFKANISNDIMGALGLYEASFHGKK	66.7	3.20		
		GESILEEARIFTTK	60.0	3.15		
3	**H9A1V3**	**Acyl-activating enzyme 1**			80.45	4
		VTRVVPLSSLPRTATNK	74.4	3.71		
		VKSTNLGALLEKR	60.5	4.05		
		RIPLYSR	69.8	3.38		
		AAADGWSHLDIRK	66.9	4.83		
4	**H9A1V4**	**Acyl-activating enzyme 2**			75.13	4
		LSVERNPGNKmLGRR	75.0	5.81		
		GLKHGDAAPLCDK	64.6	5.15		
		YYQSVIDNmYKEGNKPR	86.3	4.97		
		WAGENNISGDFKSICENPR	63.2	4.36		
5	**H9A1V5**	**Acyl-activating enzyme 3**			59.72	4
		VVPYGIVAQGYGmTETCGIVSMEDIR	77.1	5.55		
		VTHLWVVPPVILALSK	77.8	3.08		
		SPNSSLTENDVKK	85.7	5.70		
		QSDTAALLYSSGTTGmSK	64.0	4.34		
6	**H9A1V8**	**Acyl-activating enzyme 6**			62.88	2
		SVPADGTTmGEIVmR	57.7	4.93		
		SVPADGTTMGEIVmRGNAVmKGYLKNPK	51.1	6.39		
7	**H9A1V9**	**Acyl-activating enzyme 7**			67.22	3
		DRLPHYmAPKTVVFEDLPK	69.3	4.80		
		YALKARQGLNHLAMEEMDVKDSVTMESVK	63.2	3.61		
		VTHFGAAPTVLNMIVNSPK	69.3	5.37		
8	**H9A1W1**	**Acyl-activating enzyme 9**			62.24	2
		MEGAIRCSANYVPLSPISFLKR	59.2	5.70		
		SmGSLPKKNISK	58.4	5.34		
9	**H9A1W3**	**Acyl-activating enzyme 11, partial**			37.16	2
		qAAASNTKLVITLSGFIDK	77.7	7.24		
		TIMSGGAPMGNGAPmGK	51.4	3.24		
10	**H9A8L1**	**Acyl-activating enzyme 12**			84.37	3
		LKISKAKAIFTQDLIVR	70.0	3.33		
		SLDDTmILWRDEGKDDLPVNK	80.3	4.58		
		TILDEMNISFSKPPECILRENFSR	61.6	5.52		
11	**H9A8L2**	**Acyl-activating enzyme 13**			79.64	3
		TSSFEIEHVCDR	98.5	3.51		
		LGTVPSLVKTWKNTQCMKGLDWTK	70.1	3.20		
		NTQCMKGLDWTKIK	58.5	3.63		
12	**H9A8L3**	**Acyl-activating enzyme 14**			81.15	2
		NGYNSEAENLRTKFSK	64.8	5.31		
		AIFTQDFILRGSRKFPLYSR	52.6	6.19		
13	**H9A8L4**	**Acyl-activating enzyme 15**			87.81	7
		DmGVQKGDAVVIYLPmLLELPITmLACAR	69.1	8.19		
		SLMRDGDEYVTRYSRK	80.5	5.01		
		SLILSVRKQIGAFAAPER	53.8	4.72		
		TRSGKIMRR	67.2	4.61		
		HVDSmATmPSGAGK	77.2	3.40		
		TINHLRHVDSmATMPSGAGKIPR	51.6	3.61		
		CWDIVDKYK	68.3	4.42		
14	**A0A0C5APY2**	**Acetyl-CoA carboxylase carboxyltransferase beta subunit**			57.00	3
		HLWVQCENCYGLNYK	74.4	5.10		
		TSTDKSDLTRR	69.4	3.46		
		VIEQTLNK	84.8	6.00		
15	**A0A0C5ARS5**	**ATP synthase CF1 beta subunit**			53.82	3
		GIYPAVDPLDSTSTMLQPR	62.2	7.32		
		ELQDIIAILGLDELSEEDRLTVARAR	68.2	6.56		
		AVAmSATDGLmRGmDVIDTGAPLSVPVGGATLGR	60.5	5.11		
16	**E5DK51**	**Atp1, partial**			37.67	4
		VVDALGVPIDGR	77.7	12.12		
		AVDSLVPIGR	79.6	10.21		
		DNGmHALIIYDDLSK	74.5	5.15		
		TAIAIDTILNQKQmNSR	52.5	4.27		
17	**I6XT51**	**Betv1-like protein**			17.61	3
		AFVLDGDNLVPK	80.7	10.06		
		LVASGDGGNVIK	65.8	9.34		
		ITFGQGVPFK	74.2	7.76		
18	**A0A0E3TJM8**	**Cannabidiolic acid synthase homolog**			62.70	2
		WQNIAYmYEKELLLFTHFITR	52.5	6.36		
		SmGEDLFWAIRGGGGENFGIIAAWK	67.3	5.58		
19	**C6KI62**	**Chalcone synthase-like protein**			43.40	3
		LGKEAAmKAIKEWGQPK	72.5	5.87		
		KEKMADTR	81.0	3.70		
		QDMLVVEIPKLGKEAAmK	62.1	5.30		
20	**I6WIE9**	**Chalcone isomerase-like protein**			23.88	2
		ESREMFGQSmLSLF	61.2	5.33		
		VFGTTSDKIK	85.2	3.10		
21	**A0A0C5ARS7**	**Chloroplast envelope membrane protein**			27.54	2
		WWNTRPSEIFLNDIQEKNILK	65.6	4.15		
		EmIQLIKMDNEDR	79.5	4.38		

22	**P00053**	**Cytochrome c**			12.16	2
		MVFPGLK	55.0	4.11		
		QSGTTAGYSYSAANKNmAVTWZZK	54.1	3.49		
23	**A0A088MFF4**	**Delta 12 desaturase**			44.95	2
		MGAGGRmPEAKSELNGSK	52.6	5.54		
		AVKPILGEYYRLDDTPIVK	66.4	4.94		
	A0A090DLH8;				58,84;	
24	**A0A090CXP7**	**Edestin 1 GN= (ede1A.ede1C); ede1B**			58,86	21
		YLEEAFNVDSETVK	94.1	21.91		
		YLEEAFNVDSETVKR	92.5	14.84		
		VEAEAGLIESWNPNHNQFQCAGVAVVR	94.9	18.99		
		FYLAGNPEDEFEQLRR	95.5	15.85		
		GILGVTFPGCPETFEESQR	88.0	19.14		
		VKGTLDLVSPLR	84.3	14.21		
		FYLAGNPEDEFEQLR	83.6	18.31		
		GTLDLVSPLR	93.3	16.57		
		FLQLSAER	93.9	16.88		
		LQGQNDDRNSIIR	70.3	7.71		
		ADVFTPQAGR	85.6	10.42		
		QGQIVTVPQNHAVVK	80.7	16.65		
		ALPEAVLANAFQISR	94.2	23.26		
		EETVLLTSSTSSR	89.5	16.42		
		ISTVNSYNLPILR	82.2	14.05		
		YNREETVLLTSSTSSR	79.2	10.98		
		TNDNAWVSPLAGR	84.3	12.22		
		qASSDGFEWVSFK	95.2	14.06		
		YTIQQNGLHLPSYTNTPQLVYIVK	64.7	11.40		
		LRENIGDPSR	88.9	5.43		
		RFYLAGNPEDEFEQLR	62.1	9.36		
25	**A0A090CXP9**	**Edestin 2 GN= ede2A**			56.37	17
		VRGEDLQIIAPSR	93.7	18.31		
		ILAESFNVDTELAHK	92.9	13.64		
		ASAQGFEWIAVK	97.2	15.11		
		EGQIFVVPQNFAVVK	89.9	18.01		
		FLQLTAER	91.9	13.01		
		GEDLQIIAPSR	95.5	17.54		
		GIHGAVIPGCPETFER	88.9	23.15		
		ESGEQTPNGNIFSGFDTR	87.0	14.32		
		YGRDEISVFSPSSQQTR	79.5	17.27		
		DEISVFSPSSQQTR	87.7	17.35		
		QNIDRPSQADIFNPR	77.7	12.40		
		KASAQGFEWIAVK	72.3	9.76		
		LQVVDDNGRNVFDGELR	78.7	9.18		
		LQVVDDNGR	92.6	8.70		
		VSAmRAMPDDVLANAFQTSR	58.2	3.77		
		RESGEQTPNGNIFSGFDTR	67.8	10.57		

		GPISYSQASVWGKKYFGMNFK	62.7	3.73		
		SGGHDFEGASYVSKVPFVILDmRNLR	75.5	5.22		
30	**E5DKP2**	**MatR, partial**			60.12	5
		HRLIPIFKEEIDDPK	56.3	6.30		
		STVEFPGTVIREVPPRTTPIQFLR	69.2	3.50		
		TTPIQFLRELEKRLR	72.9	3.77		
		CNNNWAmRDLIK	63.0	5.04		
		qLTKGmSETGSLLDGVQLAETLGTAGVR	80.5	4.93		
31	**A0A0C5AUE3**	**Maturase K**			60.96	4
		FFFSQMISEGFAVVVEIPFSLR	55.1	4.36		
		KSISFFSKSNPR	57.8	4.71		
		SQmLENSFLMDNAmK	81.5	3.58		
		LFLHEYYSWNR	54.3	3.46		
32	**A0A0C5AUJ6**	**NADH-plastoquinone oxidoreductase subunit 5**			85.96	2
		GLAYSTmSQLGYMmLALGmGSYR	77.0	4.32		
		IIDGITNGVGISSFFVGEGIK	68.6	8.21		
33	**A0A0C5AS19**	**NADH-plastoquinone oxidoreductase subunit 7**			45.72	3
		mNIPATRKDLMIVNMGPQHPSMHGVLR	67.9	4.58		
		DVDRERKPEWNDFDYR	57.6	3.66		
		GmEKIAENR	64.8	3.25		
34	**E5DL82**	**NADH dehydrogenase subunit 5, partial**			39.91	3
		LQADKAATKAMPVNR	83.9	7.45		
		LQADKAATKAmPVNRVGDFGLAPGISGR	62.8	5.00		
		NSWISCNmR	56.1	3.60		
35	**V5KWZ6**	**Phenylalanine ammonia lyase**			77.58	3
		HLEENLRHTVKK	61.5	4.39		
		HHPGQIEAAAIMEHILDGSSYVK	74.0	3.41		
		INTLLQGYSGIRFEILEAmAK	84.9	4.63		
36	**A0A0C5AUH1**	**Photosystem II CP43 chlorophyll apoprotein**			52.02	3
		DQRLGANVGSAQGPTGLGKYLmR	57.7	3.86		
		NKmTTILGIHLILLGIGAFLLVFK	53.9	6.54		
		GIDRDLEPVLFmTPLN	62.3	5.50		
37	**A9XV91**	**Photosystem II CP47 chlorophyll apoprotein, partial**			53.81	3
		GGLFRAGSmDNGDGIAVGWLGHPIFKDKEGR	54.4	4.94		
		RMPTFFETFPVVLVDGDGIVR	65.8	4.10		
		AQLGEIFELDR	51.6	3.05		
38	**F1LKH7**	**Polyketide synthase protein**			43.12	2
		VmmYQLGCYGGGTVLR	51.9	5.15		
		SEHMTQLKEKFRK	57.7	4.58		
39	**U6EFF4**	**Putative LysM domain containing receptor kinase**			66.59	2
		DGEGNFRPLKSRTGISR	64.7	6.00		
		GEKAALKKmDMQASNEFLAELR	73.2	4.68		
40	**A0A0C5AS13**	**Ribosomal protein L2**			30.13	3
		KPSTPWGYPALGRRSRK	52.7	4.20		
		IVTIEYDPNR	59.1	6.58		
		KGAVDRQVKSNPR	79.5	4.05		

41	**A0A0C5AUJ2**	**Ribosomal protein L16**			13.37	2
		NARRGGKIWVR	61.1	3.01		
		GGKIWVRIFPDKPVTVRPTETR	51.2	5.20		
42	**A0A0C5AS20**	**Ribosomal protein L23**			10.72	5
		RmGPIMGHTmHYKR	63.6	7.50		
		SIRLLLKNQYTSNVESGSTR	72.3	5.87		
		TEIKHWVELFFGVKVIAmNSHR	65.0	3.30		
		HWVELFFGVKVIAMNSHR	61.9	6.67		
		LLLKNQYTSNVESGSTR	61.5	3.99		
43	**A0A0C5ART6**	**Ribosomal protein S3 (chloroplast)**			17.77	5
		IANPYGHPNILAEFIAGQLKNR	63.7	5.32		
		IANPYGHPNILAEFIAGQLKNRVSFR	62.2	4.04		
		AMKKAIELTEQAGTK	53.5	5.60		
		GIQIQIAGRIDGKEIAR	65.4	3.53		
		LDSSHNLYGIPK	63.6	4.91		
44	**E5DLX2**	**Ribosomal protein S3 partial (mitochondrion)**			57.27	8
		IPYGYNSYLNEVKK	73.1	5.32		
		FIDLGGIGELIKGIEMmIEIILRNR	76.6	5.19		
		HPKYAGVVNDIAFLIENDDSFR	63.1	3.80		
		VGPIGCLHSSDDTEEER	76.2	3.83		
		ICCSGRLEGAEIARTECGK	57.8	3.97		
		TLPAVRPSLNYLVmQYLLNTK	78.3	3.87		
		RLTHFIRLANDLRFAGTTK	54.0	3.50		
		YGYHDRSPSIKKNLSK	91.7	4.91		
45	**A0A0C5ARS4**	**Ribosomal protein S4**			23.68	3
		SKSLVQNYLDSSIR	55.1	3.34		
		DIITVRDEQK	64.7	3.56		
		LGmASTIPQARQLVNHR	66.5	3.59		
46	**A0A0C5ARU3**	**Ribosomal protein S7**			17.43	2
		RVGGSTHQVPIEIGSTQGKALAIR	71.0	3.38		
		NRLVNMLVNRILK	58.0	3.11		
47	**A0A0C5APY7**	**Ribosomal protein S8**			15.52	2
		DTIANILTSIRNADMNKK	53.7	5.50		
		GIMTGREAR	74.7	3.49		
48	**A0A0C5ART4**	**Ribosomal protein S11**			15.05	2
		SGILLSFIRDVTPmPHNGCRPPK	59.2	9.81		
		GVIHVQASFNNTIVTVTDVR	82.1	4.83		
49	**A0A0C5AUK1**	**Ribosomal protein S15**			10.86	2
		EENKGSVVFQILSLTNRIR	75.2	3.67		
		KLSSHLELHR	63.1	5.08		
50	**A0A0C5ART1**	**Ribosomal protein S18**			12.00	2
		ILSSLPFLNNEKQFER	77.6	9.23		
		QFERSESTPRPTGLRTKNK	70.9	6.04		
51	**A0A0C5B2I6**	**Ribulose 1,5-bisphosphate carboxylase/oxygenase large subunit**			53.29	6
		EITLGFVDLLR	67.2	6.03		
		ALRLEDLR	74.3	6.47		
		VALEACVQARNEG	56.5	3.15		
		ALRLEDLR	74.3	6.47		
		AmHAVIDRQK	100.0	3.98		
		EITLGFVDLLRDDFIEKDRSR	80.7	8.15		
52	**I0B4C2**	**Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial**			24.38	2
		QTETKASVGFKAGVKDYK	60.7	3.30		
		LGLSAKNYGRAVYECLRGGLDFTK	78.8	3.02		
53	**A0A0C5AUI9**	**RNA polymerase alpha subunit**			38.32	4
		YIFIDQSELPPR	75.1	3.86		
		NRGYRIKTPNNFQNGSYPIDAIFmPVR	77.6	5.25		
		LYYGRFILAPLMR	72.5	3.41		
		RSNIHTLLDLLNNSQEDLLKIKDFR	56.1	5.50		
54	**A0A0C5ARQ8**	**RNA polymerase beta subunit**			121.58	3
		DILAAADHLIGMKFGMGILDDmNHLK	55.0	5.99		
		HGNKGIVSKILPR	50.9	5.92		
		EILENVCYPEIFLSFLNDKEKKK	56.2	5.47		
55	**A0A0C5ARX9**	**RNA polymerase beta' subunit**			159.43	9
		mEVLmAERAGLVFHNKVIDGTAIKR	60.4	5.10		
		SSGITKYGAIGLHSIFK	70.1	5.05		
		NDPLGSGLISDNRSDCTNPFYSIYSK	70.3	6.43		
		IELKIFSGEIQFPVEmDK	68.3	3.64		
		qEmNPNFRMTDPFNPVHMMSFSGAR	68.1	3.49		
		mTDPFNPVHMmSFSGAR	69.8	3.39		
		NSILAYFDDPQYRRK	57.6	3.12		
		TSTFNFKEKVRK	69.5	4.91		
		AALRGRIDWLK	72.6	3.57		
56	**I1V0C2**	**Tetrahydrocannabinolic acid synthase, partial**			62.32	3
		SGGHDAEGMSYISQVPFVVVDLRNMHSIK	62.6	6.04		
		AGIMYELWYTATWEKQEDNEK	50.3	3.87		
		SmGEDLFWAIRGGGGENFGIIAAWK	67.3	5.58		
57	**Q0WNZ5**	**5-methyltetrahydropteroyltriglutamate**			91.00	2
		YGAGIGPGVYDIHSPR	63.0	10.65		
		AGVTVIQIDEAALR	79.8	14.26		
58	**O23058**	**A_IG005I10,23 protein**			88.70	2
		qASEAEVQPWCIHFDEKDPAPTYR	77.7	10.47		
		ENSVTSDKELVFKYVK	91.9	9.83		
59	**Q42560**	**Aconitate hydratase 1**			98.80	2
		DFNSYGSRRGNDEIMAR	89.1	7.81		
		LTGKLRDGmTATDLVLTVTQMLR	63.7	6.64		
60	**C0SVH8**	**Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger protein**			129.20	2
		TLPEASMPSK	72.4	6.51		
		SCmPKANVTPADTTEPITSFCGKK	76.6	6.22		
61	**P92935**	**ADP,ATP carrier protein 2**			67.90	2
		LGmTPLLAAVYVGALQNIFSK	70.1	6.80		

		GILSLPAKPIGVR	72.0	6.10		
62	O22241	**Arogenate dehydratase 4**				
		FLmLAREPIIPR	70.2	6.22	46.40	2
		FRSTDPTSR	75.8	6.11		
63	Q9M8K5	**AT3g06130/F28L1_7**				
		NGGGGHPQDGKNGGGGGGPNAGK	63.1	8.90	49.40	
		KGQKNGGGGGGGGGGGNSNAPK	78.0	7.67		
64	Q9FG23	**At5g06830**			62.20	2
		RmELQNSLSAIWPK	79.2	10.27		
		LVSELEEKKHR	81.3	7.66		
65	B9DG14	**AT5G41100**			65.20	2
		LPFQERRTQmNLPTVYCEEK	62.6	7.64		
		AENWRQALAISLLDNRIQTLPEK	62.9	6.95		
66	Q8VZF6	**AT5g45560/MFC19_23**			82.60	2
		LLLDWFPmVVWPRDLCYVRYWR	70.6	8.84		
		GSWIVRQSVGSTPCLLGKAVDCNYIR	80.2	68.90		
67	F4J394	**ATP binding microtubule motor family protein**			120.10	2
		DFNHTQRmPAGLDGVNMIK	100.0	7.20		
		DKRIIHLSMEEIEQK	61.5	6.29		
68	F4IMB5	**ATPase, F1 complex, alpha subunit protein**			68.30	2
		AVDSLVPIGR	79.6	10.21		
		AIPNSVKPELLQALKGGLTNER	64.7	6.35		
69	Q8L7G0	**Auxin response factor 1**			74.30	2
		QIRSCTKVHmQGSAVGR	83.0	8.28		
		FKMRFEGEEAPEK	73.1	7.38		
70	Q38810	**cDNA-5-encoded protein, partial**			33.10	2
		SDPDPTPEVQTHTRTK	85.1	9.58		
		VDPIDKIPEDDVTTEPRFGAPSLLQR	71.4	7.78		
71	Q9C9U2	**Cell division protein kinase**			45.30	2
		YRDGRDVLENVLASAR	66.2	8.87		
		SCRDDKVLESIGLLTLVVTHK	71.9	6.08		
72	O80891	**Cellulose synthase-like protein B4**			86.20	3
		SYFLRAVDLTILGLLLSLLLYR	70.4	7.70		
		LDERVHELPPVDmFVTTADPVR	60.4	7.45		
		APFmYFRNSPEAAEGSEFSK	70.3	7.12		
73	Q9SB75	**Cellulose-synthase-like C5**			80.10	2
		LCLGSILTSKIAIWK	70.8	9.26		
		AEVAKWSQK	73.1	6.47		
74	Q9LIM5	**Chloroplast protein HCF243**			76.80	2
		TSGSLKNASAGVLNSPmFGANGGRK	86.2	13.60		
		NASAGVLNSPmFGANGGR	63.7	7.85		
75	O80458	**Cis-prenyltransferase**			35.30	2
		AGLTTSQGHEAGAKR	91.8	10.02		
		FDlmHACKSLVK	76.2	6.30		
76	Q8VY07	**Clathrin binding protein-like**			62.70	2
		KAENIVALLNNKEK	67.4	7.51		
		EVNLKVLKVPEMEQK	77.9	7.47		
77	Q9FKV9	**Cyclin-dependent protein kinase-like protein**			73.60	2
		SHERKLIPPVK	68.8	6.89		
		ARDLTNNKIVALK	82.9	6.73		
78	Q9SB89	**DEAD-box ATP-dependent RNA helicase 27**			72.30	2
		KTEAEILAKGVNLLVATPGR	67.2	8.37		
		TEAEILAKGVNLLVATPGR	69.0	6.49		
79	Q8GXD6	**DEAD-box ATP-dependent RNA helicase 49**			63.40	2
		ISGRSVSLPSIDILLDEKPGWPFLKR	69.1	6.49		
		ATSDFSQENVIGKGGCNEVYR	69.2	6.48		
80	Q38K61	**Disease resistance protein, partial**			31.60	2
		ALQELVLSGCSKLESVPTVVK	64.2	8.52		
		NIAMVNLQDNLKDFSNLK	65.7	7.06		
81	Q19DM3	**Disease resistance protein, partial**			25.70	2
		CLIYDQEQDTKLRLPTmDSLR	61.0	8.15		
		LGSIAGISKLLSLR	78.8	6.89		
82	X2L4T4	**DM2D**			140.00	2
		HQLLVGERDICEVLDDDTTDSRR	67.5	7.17		
		EFPHALDIITGLWLSKSDIEEVPPWVKR	61.7	6.81		
83	Q42572	**DNA ligase 1**			94.20	2
		GSRSTQTmmFKPEPLTVVKVFDTFR	69.1	9.07		
		qNTELGDLGLVAKGSR	62.1	6.10		
84	Q9FNI6	**DNA repair protein RAD5 protein**			115.60	2
		qRMISGALTDQEVRSAR	75.7	8.05		
		FSTKDSGEIGRIPNEWAR	60.5	6.08		
85	F4JRX3	**DNA topoisomerase, type IA, core**			142.40	2
		SALQSPR	69.6	6.20		
		AFGYSFRPFARR	65.3	6.10		
86	F4I5J7	**DNAJ heat shock N-terminal domain-containing protein**			43.00	2
		qAAQELGKRALYLGVPFVAEWVR	72.1	6.92		
		AESEAKRLSDAAFGADMLHTIGYVYTR	72.2	6.60		
87	O04478	**Endoglucanase 7**			70.10	2
		DGLPDVVGGLVGGYYDGGSNVK	63.2	10.07		
		NKVSWRGDSGTK	68.1	7.30		
88	P42762	**Erd1 protein precursor**			103.70	2
		FLPDKAIDLIDEAGSRAR	83.6	9.40		
		VGLKDPDRPIAAMLFCGPTGVGK	74.5	7.23		
89	O23026	**EST gb|T04104 comes from this gene/T1G11,16 protein**			45.30	2
		RAVSSATRNLTAPLASISAAR	72.6	8.83		
		AVSSATRNLTAPLASISAAR	74.2	8.55		
90	Q76E23	**Eukaryotic translation initiation factor 4G**			188.70	2
		VEPPHNLTENR	82.2	9.51		
		VSVKPAGVSEKLGSPK	77.2	6.12		
91	Q8S905	**F15H18,12**			114.60	2

		EEKIVVTVR	83.7	8.71		
		MTIKTPGTPVSKmDR	72.2	6.34		
92	Q9SSQ2	**F6D8,29**			50.60	2
		FRCVSKPWSFIISK	70.3	10.66		
		DVHLPLDLIVEILKKLPTK	66.3	7.69		
93	Q9SHM9	**F7F22,11**			70.20	2
		SSHIVGLIRFIRSMLTR	82.2	10.84		
		SYDKLKIPCSHALVAANSYK	66.4	8.50		
94	O81490	**F9D12,11 protein**			151.00	3
		LLAYAGRGmEDDQAQmR	75.0	7.08		
		MAPAEMAELKK	66.2	6.51		
		VTEVQQRAAVAEEVPSYLR	74.1	6.51		
95	Q9SA52	**g5bf**			42.90	2
		VHQPKGALYVSASSEKK	61.8	9.11		
		GALYVSASSEKKILImGGTR	61.7	7.06		
96	Q9T0F7	**Glucuronoxylan 4-O-methyltransferase 2**			33.40	2
		ASFSSNSTATIRDEYHQK	63.8	7.08		
		ASFGTPFCPADISRRF	65.6	6.63		
97	C0Z2I0	**Glyceraldehyde-3-phosphate dehydrogenase**			37.60	2
		GILGYTEDDVVSTDFVGDNR	64.9	9.75		
		AASFNIIPSSTGAAK	72.1	8.06		
98	Q94FQ0	**Glycine-rich protein GRP17**			53.40	2
		RGMSGSGGGmSGSEGGVSGSEGSMSGGGMSGGSGSKHK	67.2	9.35		
		GMSGSEGGMSGSEGGVSESGMSGSGGGK	62.6	7.60		
99	F4IAH9	**Glyoxalase I homolog**			36.50	2
		GNAYAQIAIGTDDVYK	61.9	7.96		
		TVLVDNKDFLK	100.0	7.38		
100	Q8S8N9	**Golgin candidate 1**			79.90	2
		LARVCAGLSSRLQEIK	71.2	6.24		
		EIHPTDADVQSVLSLPLSVADTK	64.0	6.16		
101	F4IVL6	**Gravitropism defective 2**			281.40	2
		HmSANASVPDNFQR	67.1	8.92		
		LLLNWREFWR	72.0	6.09		
102	Q56X77	**Guanylate kinase**			23.40	2
		MLMKEFPSMFGFSVSHTTR	92.4	7.12		
		SmEmDGVHYHFADK	66.7	6.84		
103	Q9LKR3	**Heat shock 70 kDa protein 11**			73.90	2
		IINEPTAAAIAYGLDK	66.4	13.40		
		DAVVTVPAYFNDAQR	81.8	8.32		
104	O65719	**Heat shock 70 kDa protein 3**			60.80	7
		NAVVTVPAYFNDSQR	85.3	15.67		
		NALENYAYNMR	82.0	14.42		
		IINEPTAAAIAYGLDK	66.4	13.40		
		ATAGDTHLGGEDFDNR	79.4	11.77		
		SINPDEAVAYGAAVQAAILSGEGNEK	80.2	10.61		
		EQVFSTYSDNQPGVLIQVYEGER	81.6	11.31		

		ELESICNPIIAK	81.8	11.08		
105	Q9LHA8	**Heat shock 70 kDa protein 4**			71.60	6
		NAVVTVPAYFNDSQR	85.3	15.67		
		NALENYAYNMR	82.0	14.41		
		IINEPTAAAIAYGLDK	66.4	13.40		
		ATAGDTHLGGEDFDNR	79.4	11.77		
		SINPDEAVAYGAAVQAAILSGEGNEK	81.6	11.31		
		EQIFSTYSDNQPGVLIQVYEGER	61.3	8.11		
106	Q9S9N1	**Heat shock 70 kDa protein 5**			71.50	4
		TTPSYVAFTDTER	85.5	17.19		
		NAVVTVPAYFNDSQR	85.3	15.67		
		IINEPTAAAIAYGLDK	66.4	13.40		
		ATAGDTHLGGEDFDNR	74.7	8.69		
107	Q93ZM7	**Heat shock protein 60-3A**			71.30	2
		VLSKLSSSIGSSTSR	61.3	8.26		
		SVAAGVNVmDLRVGINMAIAAVVSDLKSR	68.7	6.81		
108	Q56XE8	**Hexokinase-4**			55.30	2
		SDKQIMRR	72.2	8.49		
		KVGRDGSGGGR	100.0	6.27		
109	Q6NR90	**Histone H4 - like protein**			11.40	2
		TVTAmDVVYALK	70.2	9.78		
		ISGLIYEETR	74.7	8.21		
110	O48701	**Hypothetical protein**			79.00	2
		LTVIEmRGER	61.5	9.15		
		GSILVRPDQHIAWRAKSGITLDPTLHMR	60.6	6.16		
111	Q84MA1	**Hypothetical protein**			66.40	2
		ALTRPSSYSCIRVNTVK	62.8	7.06		
		GTTLQGLPTDPYYRER	62.3	6.43		
112	Q9ZQQ4	**Hypothetical protein**			76.90	2
		SMGPPPPVTVSPSSGSGGEK	70.0	9.45		
		GmELIQGAIVFIQTEK	60.1	8.64		
113	Q9S755	**Hypothetical protein**			22.50	2
		IICGRPDEKPKMAAImKIIK	64.6	8.62		
		LSGTDmGGGRK	94.9	6.56		
114	O81893	**Inositol-tetrakisphosphate 1-kinase 2**			44.40	2
		IYNRQSmLQGmADLK	84.4	9.39		
		VmRRFSLPNVSNCEK	63.3	6.14		
115	O82762	**KH domain-containing protein**			64.70	2
		RPRVDNGASYDYGDNK	65.9	9.79		
		IPNNKVGLIIGKGGETIK	61.6	7.61		
116	G4WT86	**Late flowering protein**			68.80	2
		ICELmCSKGLR	93.7	6.78		
		AQSPLAFKEAATK	75.4	6.26		
117	Q5XF33	**Magnesium-chelatase subunit ChlI-2**			47.00	2
		GDmVINRAARALAALQGR	79.6	10.69		
		LCLLLNVIDPK	69.9	9.40		

118	**Q9FNB0**	**Magnesium-chelatase subunit H**			154.00	2
		MDAVLVFPSmPEVMR	77.9	9.65		
		KKQGSAGFADSmLK	67.7	6.50		
119	**Q84M98**	**Meiotic endonuclease 1A**			62.10	2
		VQPRYAIAVSK	66.9	10.12		
		RASFQSSLSKDAIEVDSDHEK	77.7	6.99		
120	**Q9SG92**	**Methylesterase 17**			31.20	2
		TLLDRVMKPEQQDAMIRR	80.4	6.16		
		MAEENQEETLELKPSR	63.0	6.03		
121	**Q9FLR3**	**No apical meristem (NAM) -like protein**			37.80	2
		APKGQKTNWVmHEYR	68.2	7.28		
		TAGGKKIPISTLIR	79.2	6.43		
122	**O48689**	**Paired amphipathic helix repeat-containing protein**			85.90	2
		LKCLDSLVFPPmTKK	61.2	8.94		
		FHEFLRLmNDVCDHKIEEANGSAR	64.3	8.86		
123	**O80958**	**Pentatricopeptide repeat protein LOJ**			98.50	2
		ElYNKmVLIGVAGDNVTTQLLmR	75.4	8.91		
		NMDLARSIFSEmLEK	92.5	8.64		
124	**Q9ZVX5**	**Pentatricopeptide repeat-containing protein**			84.60	2
		qKLDmCSKK	73.2	9.26		
		IRDAVVSGGGGWHGQGWLGTGK	72.4	8.67		
125	**Q84TI3**	**PHD-finger TITANIA 2**			132.40	2
		DFYSGSNSWITGLEAGGHDFVETVIR	80.3	10.26		
		DGTITASANKITDTVDEKGDK	69.4	7.63		
126	**Q8LFV7**	**Phosphoglycerate kinase**			42.20	3
		ELDYLVGAVANPK	95.2	17.48		
		GVTTIIGGGDSVAAVEK	84.8	16.92		
		MSHISTGGGASLELIEGKPLPGVLALDEA	69.8	10.86		
127	**Q9XIE6**	**Phospholipid-transporting ATPase 3**			139.20	2
		ADQLEVENELTPQEARSYAISQLPR	90.8	9.10		
		qSVVCRFPDGR	62.8	6.49		
128	**Q9M8D3**	**Phosphoribosylformylglycinamidine synthase**			155.30	3
		TKAVSLRCSAQPNKPK	65.4	7.18		
		qAMLLQRSSmSQLWGSVRMR	61.2	6.28		
		VPLIQESANAELLKAVQTK	64.2	6.03		
129	**Q8S8D3**	**Phox (PX) domain-containing protein**			73.10	2
		VQVSVQSPEGVSTmR	63.9	9.64		
		VNNTNLSMDESSNNR	81.1	6.72		
130	**F4K7T1**	**P-loop containing nucleoside triphosphate hydrolases superfamily protein**			74.60	2
		DLPTTNAEPSKAAGMSR	79.8	8.76		
		qNVEDPVSSGIRPNLK	69.5	8.48		
131	**Q0WR59**	**Probable inactive receptor kinase At5g10020**			115.70	3
		AmPHGNLKPTNIILSSPDNTVR	80.2	9.05		
		ISELWSLNHLNLSSNKFEGGFPSGFR	68.3	8.94		
		SLNLSRNNLEGPIPFRGSR	60.4	8.38		

	Q42290	**Probable mitochondrial-processing peptidase subunit**				
132		**beta**			59.70	2
		EDLQNYIKTHYTASRMVIAAAGAVK	77.0	7.67		
		SSLLLHmDGTSPIAEDIGRQLLTYGR	64.8	7.38		
133	**Q9C533**	**Probable protein S-acyltransferase 22**			67.00	2
		DIKDGCGSATGGAK	63.7	8.28		
		RENPFVGLEASSSFGSSGRR	62.4	7.41		
134	**P59583**	**Probable WRKY transcription factor 32**			52.20	2
		KRHGPPSSmLVAAAAPTSmR	65.1	9.76		
		QSKEALDVGGEKVMESAR	69.3	9.21		
135	**O81147**	**Proteasome subunit alpha type-6-B**			27.50	2
		SLVQQARNEAAEFR	65.5	7.27		
		CDPAGHFYGHKATSAGMK	68.2	6.91		
136	**P0C7Q8**	**Protein DA1**			61.50	2
		WLDAELAAGSTNSNAASSSSSQGLK	66.1	8.46		
		AIALSLLEENQEQTSISGK	63.7	6.75		
137	**Q8VZG2**	**Protein HYPER-SENSITIVITY-RELATED 4**			64.42	2
		TLAMDSDVKTSVmEDLDK	93.5	7.43		
		DLNSTLRSEVR	81.2	6.63		
138	**O49610**	**Protein kinase-like**			52.60	2
		ISGRSVSLPSIDILLDEKPGWPFLKR	69.1	6.97		
		ATSDFSQENVIGKGGCNEVYR	69.2	6.48		
139	**F4J8K6**	**Protein ribosomal RNA processing 5**			213.00	2
		NDSTKSFKPmKKPFK	71.8	10.79		
		SLFEGVLR	80.4	6.65		
140	**Q66GI4**	**Proteinaceous RNase P 1**			65.80	2
		qKLDmCSKK	73.2	9.26		
		IRDAVVSGGGGWHGQGWLGTGK	72.4	8.67		
141	**Q3EAS6**	**Putative DNA repair protein recA homolog 4**			24.50	2
		LVEIYGPEASGKTALALHmLSmLLIR	91.4	6.90		
		qALSLVDTLIQSGSVDVIVVDSVAALVPK	70.3	6.80		
142	**Q84W97**	**Putative kinesin**			109.30	2
		KDDHNSQILDEKMK	75.5	8.26		
		LSSPSSLSTLVRAVLSDK	71.9	7.95		
143	**Q9LR53**	**Putative protein kinase**			84.90	2
		QLLSGLEHCHSRGVLHR	79.7	7.94		
		ADTMDSRHmTAPIDPSWYNPSDSK	62.8	7.32		
144	**Q8VZQ9**	**Putative protein**			117.10	2
		NGVEWPSYVSSSKQR	83.1	7.54		
		TQAPNKSSDDPK	84.6	6.38		
145	**Q9SU68**	**Putative protein**			173.30	2
		VDSVAVEQTPLEGVDDDKK	64.9	10.65		
		GPQRmLGPTTGVSSFTPEGYR	63.1	7.51		
146	**Q93YS2**	**Putative protein phosphatase 2C 51**			120.90	2
		DSLPLHFDDSLPLDIMK	67.2	6.94		
		QGGAPAVFQSPKCPR	75.1	6.33		

147	Q9ZQE9	Putative retroelement pol polyprotein			156.30	2
		EIMNIEMTDKSFKIK	69.6	8.09		
		EImNIEMTDKSFK	77.6	7.33		
148	F4IF81	Putative serine/threonine protein kinase			76.90	2
		KEEDmLAQEKMADGEK	81.6	6.69		
		FKVTSENLDIEKVVAPSPILQK	71.3	6.01		
149	Q9SIL1	Putative TNP1-like transposon protein			110.70	2
		LKDIAKDPNVR	78.0	9.49		
		QVTNIREAKTNQER	79.1	8.75		
150	Q8L622	Putative uncharacterized protein]			66.20	2
		SSGLTTGTImAYALEYNDEK	74.6	8.86		
		LDLRFHHSTSSQSVESAALDLDK	62.8	6.33		
151	A0MEZ6	Putative uncharacterized protein			16.08	2
		DPYYDDLKVAKR	68.9	8.25		
		VAKRAIEQmEmVAMMEGIPK	80.4	6.42		
152	F4K4D6	PWWP domain-containing protein			118.40	2
		APPRAPLSGPLVIAETLGDLK	70.7	7.95		
		mEERAAVLPEHGKSEAmASLKPK	87.9	6.34		
153	Q9FGN6	Receptor-like kinase MOL1			100.00	2
		LmNAGGLmIQNKPK	71.5	6.19		
		LVSVLVACLVSILLMVVAALALYYIR	85.9	6.13		
154	P17094	Ribosomal protein			44.80	2
		DDQTKPCKFTAFmGYKAGMTHIVR	63.1	7.29		
		DVTPmGGFPHYGIVKDDYFMIK	69.2	6.37		
155	F4KE59	RING/FYVE/PHD zinc finger-containing protein			143.60	2
		ELCQTSSPSKHLEKGSSLR	70.2	7.82		
		NTSLPTSNVLPRDR	69.8	6.30		
156	Q9LJX4	RNA binding protein-like			107.70	2
		LELSDIAGR	81.1	7.86		
		FPSRSESAPPSMEGSFAALR	61.3	6.97		
157	Q5D869	RNA polymerase IV largest subunit			220.40	2
		KNSETESDAAAWGSRDK	64.4	8.95		
		ETKLGSGVDFITVDKHTIFSDSR	63.4	6.93		
158	B0LC03	RPS2			106.10	2
		LPFQERRTQmNLPTVYCEEK	62.6	7.64		
		AENWRQALAISLLDNRIQTLPEK	62.9	6.95		
159	F4J3S1	Shugoshin C terminus			64.50	2
		ILAEFNTSKDQLK	66.3	7.98		
		SAmFNIQELGVIQNLNGLPDDQEIAAKAR	61.1	6.54		
160	Q9FWR3	Similar to tRNA-splicing endonuclease positive effector SEN1			242.40	2
		DLPTTNAEPSKAAGMSR	79.8	8.76		
		qNVEDPVSSGIRPNLK	69.5	8.48		
161	F4JXC5	Subtilisin-like protease SBT5,4			84.10	2
		MYSLISAADANVANGNVTDALLCKKGSLDPK	63.3	8.26		
		VDKGmQAAAAGAAGMVLCNDK	74.7	7.15		

162	O82496	T12H20,15 protein			81.50	2
		SRPGSSGAGGSSSmKNMPPSKR	62.1	9.19		
		AFGKSRPGSSGAGGSSSMKNmPPSK	79.4	6.22		
163	Q9LQ91	T1N6,6			96.40	2
		FHQEHNDPNLATSK	60.9	8.73		
		GPVAKSNSFCNTTLK	100.0	7.03		
164	Q9LPA7	T32E20,13			156.50	2
		EAENSRAADEQKK	67.6	6.78		
		KQAGDSTSFVALQEQLQEAQR	82.8	6.13		
165	Q9SCZ3	TIR-NBS-LRR class disease resistance protein			142.80	2
		SLVTLKLTDPSGmSIR	64.5	7.33		
		qPWGVRSIGIWGmPGIGK	74.0	6.68		
166	O04251	Transcription coactivator protein			149.90	2
		MTENVLLQEQRSGSPKQNLSVVPNLR	82.1	10.04		
		LFNSGVVPMEADIR	74.7	7.62		
167	Q9LHE4	Transcription factor-like protein			116.70	2
		LDGTmSLAARDR	81.9	11.17		
		SQFKAYADAGTLSQNYANILLLLLRLR	66.9	10.91		
		RTMVASAFGEEHGGSSATR	81.4	6.78		
168	Q9SSP7	Ubiquitin carboxyl-terminal hydrolase-related protein			131.50	2
		YRDGRDVLENVLASAR	66.2	8.87		
		SCRDDKVLESIGLLTLVVTHK	71.9	6.08		
169	Q56Y19	Uncharacterized protein			49.10	2
		TSIDFVLNACSR	70.6	6.47		
		IDAARmVSAAVAARTR	75.0	6.09		
170	Q8S8L1	Uncharacterized protein			38.50	2
		CCSCSLCLQLVAIHISFVSLVTPK	60.9	7.38		
		WVEVGDLGDCVLLLGNLSCSAKEFPR	62.5	6.07		
171	Q9FHT3	Uncharacterized protein			55.30	2
		ASSPCDSVLGQVLSFK	60.1	7.28		
		KLKISFDEVNIK	75.9	6.07		
172	F4K4M9	Uncharacterized protein			95.00	2
		QSNQPVIVSAEIDQK	70.7	7.93		
		LFHLLNSLGDLmMLPFKMLADKSTR	64.8	6.44		
173	Q9FH23	Uncharacterized protein			168.30	2
		VKVVQAVITLMFLRIQK	71.4	8.89		
		NLSHEAYSSALVKIR	81.5	8.42		
174	Q9LXU4	Uncharacterized protein			96.70	2
		ECTNTPTQLSSHTFR	65.7	8.76		
		LAGNSQALKmATGmADYFYGRVR	65.8	7.25		
175	Q9FHW3	Unnamed protein product			14.80	2
		AVQDAFAEAASSKNGKEIVK	65.0	8.93		
		mLATTNVSTDVASmMLDGTVK	70.1	6.52		
176	P93736	Valine--tRNA ligase			127.70	2
		mISLRFQPSTTAGVLSASVSR	70.7	8.95		

		ASITTPNKLGDCPFCQK	69.4	6.15		
177	**Q94BQ3**	**WD and tetratricopeptide repeats protein 1**			84.90	2
		QASFLGQRGEYIASGSDDGR	64.0	7.20		
		SEANSDSSHDmSRSER	88.8	6.58		
178	**F4HZB2**	**WD/BEACH domain protein SPIRRIG**			404.30	2
		WQNGEISNFQYLmHLNTLAGR	87.1	11.07		
		HVIVRNmLLEMLIDLQVTIK	61.4	6.30		
		INLVEHLLVTLQR	74.3	6.15		
179	**P56785**	**ycf1**			214.80	2
		KLSFFSEPQQEEK	79.1	10.98		
		EETRRIEIAETWDSFLFAQIIR	61.7	6.84		
180	**O64533**	**YUP8H12R,20**			150.30	2
		VLHVSRAMISGAAGSVDALDQAIR	73.8	8.74		
		MQSLVFVDQMRFR	60.8	6.41		
181	**O49281**	**Zinc finger protein**			122.60	2
		YIACGSNYTAAICLHK	83.3	7.73		
		SKIDGWSGGGLSVDASR	69.0	7.31		

References

Arnoldi, A., Bo schin, G., Zanoni, C., & Lammi, C. (2015). The health benefits of sweet lupin seed flours and isolated proteins. *J Funct Foods, 18*, 14.

Arnoldi, A., Zanoni, C., Lammi, C., & Boschin, G. (2015). The role of grain legumes in the prevention of hypercholesterolemia and hypertension. *Critic Rev Plant Sci, 34*(1-3), 144-168.

Bazzano, L. A., Thompson, A. M., Tees, M. T., Nguyen, C. H., & Winham, D. M. (2011). Non-soy legume consumption lowers cholesterol levels: a meta-analysis of randomized controlled trials. *Nutr Metab Cardiovasc Dis, 21*(2), 94-103.

Callaway, J. C. (2004). Hempseed as a nutritional resource: An overview. *Euphytica, 140*(1-2), 65-72.

Candiano, G., Dimuccio, V., Bruschi, M., Santucci, L., Gusmano, R., Boschetti, E., Righetti, P. G., & Ghiggeri, G. M. (2009). Combinatorial peptide ligand libraries for urine proteome analysis: investigation of different elution systems. *Electrophoresis, 30*(14), 2405-2411.

Cilindre, C., Fasoli, E., D'Amato, A., Liger-Belair, G., & Righetti, P. G. (2014). It's time to pop a cork on champagne's proteome! *J Proteomics, 105*, 351-362.

Cunsolo, V., Muccilli, V., Fasoli, E., Saletti, R., Righetti, P. G., & Foti, S. (2011). Poppea's bath liquor: the secret proteome of she-donkey's milk. *J Proteomics, 74*(10), 2083-2099.

Docimo, T., Caruso, I., Ponzoni, E., Mattana, M., & Galasso, I. (2014). Molecular characterization of edestin gene family in Cannabis sativa L. *Plant Physiol Biochem, 84*, 142-148.

Fasoli, E., Pastorello, E. A., Farioli, L., Scibilia, J., Aldini, G., Carini, M., Marocco, A., Boschetti, E., & Righetti, P. G. (2009). Searching for allergens in maize kernels via proteomic tools. *J Proteomics, 72*(3), 501-510.

Frydman, J. (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem, 70*, 603-647.

Girgih, A. T., Alashi, A., He, R., Malomo, S., & Aluko, R. E. (2014). Preventive and treatment effects of a hemp seed (*Cannabis sativa* L.) meal protein hydrolysate against high blood pressure in spontaneously hypertensive rats. *Eur J Nutr, 53*(5), 1237-1246.

Girgih, A. T., Alashi, A. M., He, R., Malomo, S. A., Raj, P., Netticadan, T., & Aluko, R. E. (2014). A novel hemp seed meal protein hydrolysate reduces oxidative stress factors in spontaneously hypertensive rats. *Nutrients, 6*(12), 5652-5666.

Girgih, A. T., He, R., Malomo, S., Offengenden, M., Wu, J., & Aluko, R. E. (2014). Structural and functional characterization of hemp seed (*Cannabis sativa* L.) protein-derived antioxidant and antihypertensive peptides. *J Funct Foods, 6*, 384-394.

Girgih, A. T., Udenigwe, C. C., & Aluko, R. E. (2011). In vitro antioxidant properties of hemp seed (*Cannabis sativa* L.) protein hydrolysate fractions. *J Am Oil Chem Soc, 88*(3), 381-389.

Girgih, A. T., Udenigwe, C. C., & Aluko, R. E. (2013). Reverse-phase HPLC separation of hemp seed (*Cannabis sativa* L.) protein hydrolysate produced peptide fractions with enhanced antioxidant capacity. *Plant Foods Hum Nutr, 68*(1), 39-46.

Guo, L., Devaiah, S. P., Narasimhan, R., Pan, X., Zhang, Y., Zhang, W., & Wang, X. (2012). Cytosolic glyceraldehyde-3-phosphate dehydrogenases interact with phospholipase Dδ to transduce hydrogen peroxide signals in the Arabidopsis response to stress. *Plant Cell, 24*(5), 2200-2212.

He, M., Zhu, C., Dong, K., Zhang, T., Cheng, Z., Li, J., & Yan, Y. (2015). Comparative proteome analysis of embryo and endosperm reveals central differential expression proteins involved in wheat seed germination. *BMC Plant Biol, 15*, 97.

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., & von Mering, C. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res, 37*(Database issue), D412-416.

Koo, H. J., Park, S. M., Kim, K. P., Suh, M. C., Lee, M. O., Lee, S. K., Xinli, X., & Hong, C. B. (2015). Small heat shock proteins can release light dependence of tobacco seed during germination. *Plant Physiol, 167*(3), 1030-1038.

Lammi, C., Zanoni, C., Scigliuolo, G. M., D'Amato, A., & Arnoldi, A. (2014). Lupin peptides lower low-density lipoprotein (LDL) cholesterol through an up-regulation of the LDL receptor/sterol regulatory element binding protein 2 (SREBP-2) pathway at HepG2 cell line. *J Agric Food Chem, 62*(29), 7151-7159.

Lüthje, S., Döring, O., Heuer, S., Lüthen, H., & Böttger, M. (1997). Oxidoreductases in plant plasma membranes. *Biochim Biophys Acta, 1331*(1), 81-102.

Malomo, S. A., & Aluko, R. E. (2015). A comparative study of the structural and functional properties of isolated hemp seed (Cannabis sativa L.) albumin and globulin fractions. *Food Hydrocol, 43*, 743-752.

Malomo, S. A., He, R., & Aluko, R. E. (2014). Structural and functional properties of hemp seed protein products. *J Food Sci, 79*(8), C1512-C1521.

Odani, S. (1998). Isolation and primary structure of a methionine- and cystine-rich seed protein of Cannabis sativa. *Biosci Biotechnol Biochem, 62*(4), 650-654.

Oomah, B. D., Busson, M., Godfrey, D. V., & Drover, J. C. G. (2002). Characteristics of hemp (Cannabis sativa L.) seed oil. *Food Chemistry, 76*(1), 33-43.

Park, S. K., Seo, J. B., & Lee, M. Y. (2012). Proteomic profiling of hempseed proteins from Cheungsam. *Biochim Biophys Acta, 1824*(2), 374-382.

Righetti, P. G., Boschetti, E., Kravchuk, A. V., & Fasoli, E. (2010). The proteome buccaneers: how to unearth your treasure chest via combinatorial peptide ligand libraries. *Expert Review of Proteomics, 7*(3), 373-385.

Righetti, P. G., Fasoli, E., & Boschetti, E. (2011). Combinatorial peptide ligand libraries: The conquest of the 'hidden proteome' advances at great strides. *Electrophoresis, 32*(9), 960-966.

Russo, R., & Reggiani, R. (2015). Evaluation of protein concentration, amino acid profile and antinutritional compounds in hempseed meal from dioecious and monoecious varieties. *Am. J. Plant Sci., 6*(1), 14-22, 10 pp.

Sawler, J., Stout, J. M., Gardner, K. M., Hudson, D., Vidmar, J., Butler, L., Page, J. E., & Myles, S. (2015). The Genetic Structure of Marijuana and Hemp. *PloS one, 10*(8), e0133292-e0133292.

Tang, C. H., Ten, Z., Wang, X. S., & Yang, X. Q. (2006). Physicochemical and functional properties of hemp (Cannabis sativa L.) protein isolate. *J Agric Food Chem, 54*(23), 8945-8950.

von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., Jouffre, N., Huynen, M. A., & Bork, P. (2005). STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res, 33*(Database issue), D433-437.

Vonapartis, E., Aubin, M.-P., Seguin, P., Mustafa, A. F., & Charron, J.-B. (2015). Seed composition of ten industrial hemp cultivars approved for production in Canada. *J Compos Anal, 39*, 8-12.

Wang, X.-S., Tang, C.-H., Yang, X.-Q., & Gao, W.-R. (2008). Characterization, amino acid composition and in vitro digestibility of hemp (Cannabis sativa L.) proteins. *Food Chemistry, 107*(1), 11-18.

CHAPTER 7

MANUSCRIPT 5

EXPLORATION OF POTENTIALLY BIOACTIVE PEPTIDES GENERATED FROM THE ENZYMATIC HYDROLYSIS OF HEMPSEED PROTEINS

Gilda Aiello, Carmen Lammi, Giovanna Boschin, Chiara Zanoni, Anna Arnoldi*

Department of Pharmaceutical Sciences, University of Milan, 20133 Milan, Italy

PART-I

7.0 Abstract

The seed of industrial hemp is an underexploited protein source. In view of a possible use in functional foods, a hempseed protein concentrate was hydrolyzed with pepsin, trypsin, pancreatin, or a mixture of these enzymes. A detailed peptidomic analysis using data-dependent acquisition showed that the numbers of peptides identified ranged from 90 belonging to 33 parent proteins in the peptic hydrolysate to 9 belonging to 6 proteins in the pancreatin digest. The peptic and tryptic hydrolysates resulted to be the most efficient inhibitors of 3-hydroxymethyl-coenzyme A reductase activity, when tested on the catalytic domain of the enzyme. Using the open access tools PeptideRanker and BIOPEP, a list of potentially bioactive peptides was generated: the alleged activities included the antioxidant property, the glucose uptake stimulating activity, the inhibition of dipeptidyl peptidase-IV (DPP-IV) and of angiotensin converting enzyme I (ACE).

Keywords: bioactive peptides, *Cannabis sativa*, LC-MS/MS, hempseed hydrolysates, HMGCoAR, peptidomics

Abbreviations Used: AAC, amino acid composition; **ACE**, angiotensin converting enzyme I; **ACN**, acetonitrile; **DPP-IV**, dipeptidyl peptidase-IV; **HCA**, hierarchical cluster analysis; HMGCoAR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; **MW**, molecular weight; **MWCO**, molecular weight cut-off; **SDS-PAGE**, sodium dodecyl sulphate – polyacrylamide. **HPC**, protein concentrate; **TPSI**, Total Protein Spectral Intensity.

7.1 Introduction

Currently, there is a growing interest for the production of food protein hydrolysates containing bioactive peptides for potential applications in functional foods. So far, much research has been focused on the use of animal proteins (milk, egg, fish, meat) as raw materials for the production of such bioactive peptides.[1] However, edible plants and mainly their seeds represent cheap and environmentally sustainable protein sources currently investigated for the same purpose.[2]

Certainly, a complete information on peptide sequences is crucially relevant in order to elucidate the correlation between the composition of such hydrolysates and the observed biological activities and to elucidate the molecular mechanisms involved. This knowledge is, therefore, a key factor in the development of applications in nutraceuticals and functional foods. In fact, the specific bioactivity of food peptides against various molecular targets depends on their structural properties, such as amino acid composition, length and physicochemical characteristics of the amino acid side chains, as well as their bulkiness, hydrophobicity, and charge.[3] Given a single starting material, different peptide profiles may be achieved under varying hydrolytic conditions, since enzymatic activity is a function of structural characteristics of the substrate on which the proteases act.[4] To date, the characterization of the relative similarities and differences between such peptide profiles remains largely unstudied. In this context, the advent of peptidomics based on advanced analytical techniques, in particular mass spectrometry, allows to elucidate the full components present in a specific peptide mixture. Indeed, continued advances in tandem MS technologies provide today more and more accurate identifications and quantifications of such peptides.[5] Accordingly, this technology has attracted the attention of food scientists and nutritionists as a promising approach for the characterization of food protein hydrolysates.[6-8]

While proteomics usually comprises molecular weights from approx. 700 to 3000 Da with a dynamic range of twelve orders of magnitude, peptidomics certainly spans over a greater peptide length distribution showing instead a similar dynamic range.[8] Moreover, the high concentration of low molecular weight peptides in food hydrolysates is undoubtedly an analytical challenge to peptidomic analysis, and research in this area is yielding significant results especially for the identification of small peptides.

The seed of industrial hemp, i.e. the non-drug cultivars of *Cannabis sativa*, is certainly an underexploited protein-rich seed.[9] Hempseed proteins are an excellent natural source of highly digestible amino acids when compared to other protein sources, such as borage meal, canola meal, and heated canola meal.[10, 11] Interestingly, recent investigations have demonstrated that peptides produced by enzymatic hydrolysis of hempseed proteins provide several biological activities, including the antihypertensive one[12, 13] and antioxidant one.[12, 14, 15]

In addition, a few literature evidences indicate that the inclusion of hempseed protein in the diet of suitable animal models modulates their lipid profile in a favorable way.[16-18] Since proteins are hydrolyzed during digestion, the activity may be due to specific peptides encrypted in the protein sequences that are released by digestion and absorbed at intestinal level. This has stimulated our interest for assessing whether the enzyme selection and technical conditions may modulate the hypocholesterolemic properties of hempseed peptides.[19] In fact, the bioactivity of food protein hydrolysates depends strictly on these parameters.[20]

Based on these considerations, the overall objective of the present study was to compare the efficiency of some enzymes and/or enzyme combinations in the production of bioactive protein hydrolysates from hempseed. In details, the specific objectives were: (i) the optimization of the release of the peptides from hempseed protein using different enzymes; (ii) the identification and characterization of each hydrolysate by a shotgun MS based approach; iii) the evaluation of the inhibitory activity of each hydrolysate on 3-hydroxy-3-methyl-glutaryl-coenzyme A reductase (HMGCoAR), a key enzyme in cholesterol metabolism; and (iv) the prediction of additional biological activities using *in silico* bioinformatics tools.

7.2 Materials and Methods

7.2.1 Reagents

All chemicals and reagents were of analytical grade. LC-grade H_2O (18 MΩ cm) was prepared with a Milli-Q H_2O purification system (Millipore, Bedford, MA, USA). Acetonitrile (ACN), tris(hydroxymethyl)aminomethane (Tris-HCl), hydrochloric acid (HCl), ammonium

bicarbonate, and HMGCoAR assay Kit were provided by Sigma-Aldrich (St. Louis, MO, USA). Pepsin from porcine gastric mucosa (P7012, lyophilized powder, ≥ 2,500 units/mg protein), trypsin from bovine pancreas (T1426, lyophilized powder, ≥ 10,000 units/mg protein), and pancreatin from porcin pancreas (P1625, powder, 3 x ≥ USP specification) were from Sigma-Aldrich (St. Louis, MO, USA). Bovine serum albumin (BSA) and β-mercaptoethanol were from Thermo Fisher Scientific (Life Techonoly, Milan Italy). Mini-Protean apparatus, precision plus protein standards, Bradford reagent and Coomassie Blue G-250 were purchased from Bio-Rad (Hercules, CA, USA).

7.2.2 Protein concentrate preparation

The seeds of the species *C. sativa* (cultivar Futura) were provided by the Institute of Agricultural Biology and Biotechnology, CNR (Milan, Italy). The hempseed protein concentrate (HPC) was prepared applying the method described previously with some modifications.[21] Briefly, 2 g of defatted hempseed flour were homogenized with 15 mL of 100 mM Tris-HCl/0.5 M NaCl buffer, pH 8.0. The extraction was performed in batch at 4 °C overnight. The solid residue was eliminated by centrifugation at 5,800 *g* for 30 min at 4 °C and the supernatant was dialyzed against 100 mM Tris-HCl buffer, pH 8.0 for 36 h at 4 °C. The protein content of HPC, assessed according to the Bradford method using BSA as standard, was 15.4 mg/mL.

7.2.3 Preparation of the hempseed protein hydrolysates

The HPC, dissolved in 100 mM Tris-HCl/0.5 M NaCl buffer pH 8.0, was hydrolyzed using three single enzymes: i.e. pepsin, trypsin, pancreatin or, in order to mimic the gastrointestinal digestion, a combination of the same enzymes. The peptic hydrolysis was performed adjusting the pH to 2 by adding 1 M HCl to the HPC. The enzyme solution (4 mg/mL in NaCl 30 mM) was added in a 1:50 enzyme/hempseed protein ratio (w/w). The mixture was incubated for 16 h and the enzyme inactivated changing the pH to 7 by adding 1 M NaOH. Tryptic and pancreatic hydrolysis was performed directly in the buffer solution adding trypsin (4 mg/mL in HCl 1 mM) and pancreatin (4 mg/mL in H_2O) in a 1:50 enzyme/HPC ratio (w/w). After 16 h incubation, the digestion was stopped changing the pH to 3 by adding 1 M HCl. The simulated gastrointestinal digestion was initiated by the addition of pepsin [1:20 (w/w) enzyme/hempseed protein ratio] stirring the mixture for 2 h at pH 2. After that, the reaction mixture was adjusted to pH 8.5 with 1 M NaOH followed by the addition of a mixture of trypsin and pancreatin, each at a 1:25 enzyme/HPC (w/w) ratio.

The mixture was incubated at 37 °C for 4 h. The enzymatic reaction was terminated by adjusting the mixture to pH 3 with 1 M HCl. Each digestion was stopped by holding at 95 °C for 10 min to ensure a complete inactivation of residual enzyme activity.

All digestion processes were performed at 37 °C and all obtained hydrolysates were purified separating the undigested proteins, the high molecular-weight polypeptides and the intact enzymes by ultrafiltration through membranes with a 3-kDa molecular weight cut-off (MWCO) (Millipore, USA) at 12,000 *g* for 30 min at 4 °C. Finally, the permeated peptides were collected and stored at -20 °C until used in further experiments.

7.2.4 Evaluation of the percent peptide yield and DH of the hydrolysates

The peptide concentration (µg/µL) of each hydrolysate was determined according to a literature method,[22] which is based on chelating the peptide bonds by Cu(II) in alkaline media and monitoring the change of absorbance at 330 nm according to Lammi *et al*., 2016.[23] The percent peptide yield from the HPC was determined as the ratio between peptide concentration and the protein concentration of the non-hydrolyzed HPC, estimated by Bradford assay. The degree of hydrolysis was determined by the OPA assay, according to Nielsen *et al.*, 2001[24] with some modifications. This assay is based on the formation of an adduct between the α-amino groups of peptides and the OPA reagent. The assay consisted of mixing 200 µL of OPA reagent with 26.6 µL of hydrolysates. After 1.5 min of incubation at 25 °C, the absorbance was measured at 340 nm using the Synergy H1 fluorescent plate reader (Biotek, Bad Friedrichshall, Germany).

7.2.5 Tricine SDS-PAGE Separation

To monitor the efficacy of hydrolysis, Tris-Tricine SDS-PAGE was used following a literature method.[25] A 16.0% resolving gel using 40% acrylamide/bis solution (19:1) and 6 M urea was prepared and overlaid with 5% stacking gel. A fixed volume (500 µL) of each hydrolysate was dried and dissolved in 15 µL of 2X loading buffer containing SDS, β-mercaptoethanol, glycerol and Coomassie G-250 stain. The mixture was heated in a boiling water bath for 5 min, vortexed for 30 sec, and allowed to cool to room temperature. All the 15 µL of the sample were loaded onto the gel. A mixture of proteins (range 26.7 kDa -1.4 kDa, Biorad) was used as a broad range MW marker. The cathodic compartment were filled with Tris–Tricine buffer, pH 8.3, containing 0.1%, m/v SDS, whereas the anodic compartment was filled with Tris-HCl, pH 8.9. Electrophoreses were run on a Mini-Protean II Cell at 100 V until the dye front reached the gel bottom. The resolved

protein bands were stained by immersing the gel in a solution containing 45% methanol, 10% glacial acetic acid, and 0.25% Coomassie Brilliant Blue R-250 for 1.5 h. To visualize the bands, the gel was destained in a solution containing 45% methanol and 10% glacial acetic acid until they were clearly visible.

7.2.6 MS/MS peptide profiling

The peptide solutions (100 µL) were desalted on SepPak C18 cartridge (Thermo Fisher Scientific, Life Technology, Milan Italy) conditioned with MeOH and rinsed with 0.1% FA. Peptides were eluted from the SPE column with 280 µL ACN:H_2O (80:20, v/v) containing 0.1% FA and then dried in a Speed-Vac (Martin Christ). Each sample was reconstituted with 20 µL of a solution of 2% ACN, 0.1% FA, properly diluted, and analyzed on a SL IT mass spectrometer interfaced with a HPLC-Chip Cube source (Agilent Technologies, Palo Alto, CA, USA). Each sample was loaded onto a 40 nL enrichment column (Zorbax 300SB-C18, 5 µm pore size), and separated onto a 43 mm × 75 µm analytical column packed (Zorbax 300SB-C18, 5 µm pore size). Separation was carried out in gradient mode at a flowrate of 300 nL/min. The LC solvent A was 95% water, 5% ACN, 0.1% formic acid; solvent B was 5% water, 95% ACN, 0.1% formic acid. The nano pump gradient program was as follows: 5% solvent B (0 min), 80% solvent B (0–40 min), 95% solvent B (40–45 min), and back to 5% in 5 min. A reconditioning at the initial chromatographic conditions was conducted for 5 minutes. The drying gas temperature was 300 °C, flow rate 3 L/min (nitrogen). Data acquisition occurred in positive ionization mode. Capillary voltage was −1950 V, with endplate offset −500 V. Full scan mass spectra were acquired in the mass range from *m/z* 300 to 2000 Da. LC-MS/MS analysis was performed in data-dependent acquisition AutoMS(n) mode. In order to increase the number of identified peptides, three technical replicates (LC–MS/MS runs) were run for each of the three experimental replicates.

7.2.7 Database searching, protein identification and validation

The MS/MS data were analyzed by Spectrum Mill Proteomics Workbench (Rev B.04.00, Agilent), consulting the *C. sativa* (531 sequences) protein sequences database downloaded from the National Center for Biotechnology Information (NCBI). The enzymes selected were pepsin and trypsin for the analysis of the peptic and tryptic hydrolysates, respectively; whereas none specific cleavage was selected for analyzing the pancreatic and co-digested hydrolysates. Two missed cleavages were allowed to each enzyme used; peptide mass tolerance was set to 1.2 Da and fragment mass

tolerance to 0.9 Da. For quality assignment, a sequence tag lengths > 4 was used. Threshold used for peptide identification score ≥ 6; Scored Peak Intensity SPI% ≥ 70%; autovalidation strategy both in peptide mode and in protein polishing mode was performed using FDR cut-off ≤ 1.2 %. Protein abundance was performed at protein level using Total Protein Spectral Intensity (TPSI) based on the summation of peptide intensities, calculated from extracted ion chromatograms from each precursor ions.

7.2.8 Amino acid composition

The amino acid compositions of the hempseed hydrolysates and the isoelectric points (pI) were determined using ProtParam tool (http://web.expasy.org/protparam/).[26] Hierarchical clustering analysis (HCA) and its visualization were performed using Cluster 3.0 and Java TreeView, respectively. HCA allows the presentation of cluster results in a dendrogram, where the similarity among the samples is determined from the value on the distance axis at which they join in a single cluster (the smaller the distance, the more similar the sample). Euclidean distance was used to calculate the matrix of all samples. The complete linkage method was then used in the assignment of clusters.

7.2.9 HMGCoAR activity assay

The HMGCoAR inhibitory activity of each hydrolysate was evaluated using a commercial assay Kit providing HMGCoAR (catalytic domain), NADPH, assay buffer, and substrate solution. The experiments were carried out at 37 °C following the manufacturer's instructions. Each reaction (200 µL) was prepared by adding the reagents in the following order: 1X assay buffer; 0.2, 0.3, 0.5, and 1.0 mg/mL of co-digested peptides, with 1.0, and 2.0 mg/mL of the peptides digested with pancreatin, 0.1, 0.25, 0.35, 0.5, and 1.0 mg/mL of the peptides digested with pepsin, or 0.2, 0.5, and 1.0 mg/mL of the peptides digested with trypsin or vehicle (C); NADPH (4 µL); substrate solution (12 µL); and finally HMGCoAR (2 µL). Subsequently, the samples were mixed, and the absorbance at 340 nm was read by a microplate reader (Synergy H1 from Biotek, Bad Friedrichshall, Germany) at 0 and 10 min. The HMGCoA-dependent oxidation of NADPH and the inhibition properties of hempseed peptides were measured by the absorbance reduction, which is directly proportional to the enzyme activity.

7.2.10 Profile of potential biological activities and peptide ranking

The potential bioactivities of hempseed peptides were predicted using the open access tool PeptideRanker (http://bioware.ucd.ie/compass/biowareweb/),[27] a web-based tool used to predict the probability of biological activity of peptide sequences. Using N-to-1 neural network probability, PeptideRanker provides peptide scores in the range of 0–1. The maximum scores indicate the most active peptides, whereas the minimum scores denote the least active peptides. Here, only those peptides with a score higher than 0.6 were considered as potentially "bioactive". Subsequently, the lists of best-ranked peptides were submitted to the web-available database BIOPEP (http://www.uwm.edu.pl/biochemia/index.php/pl/biopep/).

7.2.11 Statistical analysis in the HMGCoAR activity assay

Statistical analyses were carried out by One-way ANOVA (Graphpad Prism 6) followed by Dunnett's test. Values were expressed as means ± SD; P-values < 0.05 were considered to be significant.

7.3 Results

7.3.1 Hydrolysis trend, yield, and DH of hempseed hydrolysates

In order to produce protein hydrolysates endowed with potential biological activities, HPC was digested using one enzyme, i.e. pepsin, trypsin, or pancreatin, or a mixture of the same enzymes in order to mimic the gastrointestinal digestion. The highest peptide yield was observed for the pancreatic hydrolysate (43%), followed by the tryptic hydrolysate (24.6%), the co-digested hydrolysate (18.2%) and the peptic one (16%). The DH values were 19.7% for peptic hydrolysate, 46.6% for tryptic, 47.5% for pancreatic, and 34% for codigested. These results indicate a direct correlation between the peptide yields and the DH values, the trends across all hydrolysates are comparable.

In order to monitor the efficiency of the hydrolysis, a tricine-SDS-PAGE was used to resolve the peptide pool composition of each hydrolysates. **Figure 7.1** shows the profile of the molecular weight distribution at the end of digestion. The peptic hydrolysate showed many continuous, intense and unresolved bands in the range from 3.5 to 26.6 kDa, the co-digested hydrolysate displayed another band-rich profile, whereas the tryptic and pancreatic presented only small bands

between 6.5 and 26.6 kDa. The absence of intense bands indicated that these hydrolysates contained mostly very short peptides (MW smaller than 3.5 kDa) that had diffused through the gel. The results of the percent peptide yields and DH were in agreement with these findings: in fact, consistently, yields and DH were larger when the bands were more difficult to visualize on the gels.

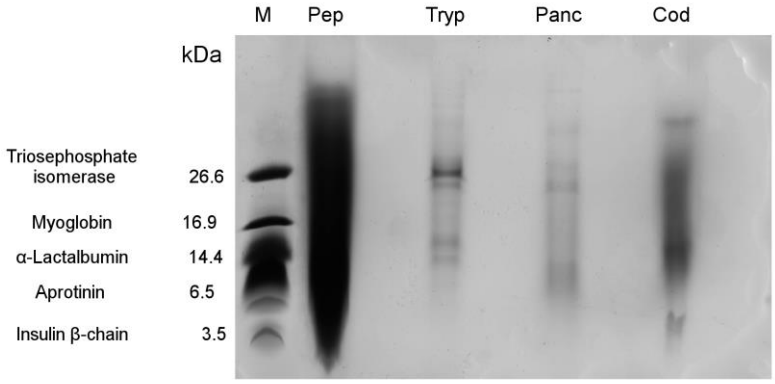


Figure 7.1 - Tricine-SDS-PAGE of hempseed protein hydrolysates. M (marker) 26.6-3.5 kDa. Pep, Tryp, Panc, Cod represent the four hydrolysates.

7.3.2 Chemical characterization of the protein hydrolysates

The characterization of the four hydrolysates was carried out by HPLC-Chip MS/MS analysis. The results are summarized in **Table 7.1S** (Supplementary materials), which reports the identified peptides according to their parent proteins. The chromatograms of each hydrolysate are reported in **Figure 7.1S** (Supplementary materials). The peptic hydrolysate was the richest both in terms of identified peptides and proteins. In fact, it was possible to detect 90 peptides belonging to 33 *C. sativa* proteins, whereas in the codigested sample 62 peptides belonging to 25 proteins were identified. In the tryptic digest, it was possible to detect 25 peptides accounting for 6 proteins, while only 9 peptides deriving from 6 proteins were identified in the pancreatic hydrolysate. Therefore, the composition in terms of proteins and peptides is very specific for each hydrolysate. **Figure 7.2A** shows the percent distribution of the peptides deriving from specific parent proteins in each hydrolysate. The peptic hydrolysate contained peptides derived from numerous proteins, with a small prevalence of the two isoforms of Edestin (6% from Edestin 1 and 6% from Edestin 2), as well as DNA-directed RNA polymerase subunit beta (6%) and Protein Ycf2 (6%). The percentage of peptides deriving

from Edestin increased greatly in the tryptic hydrolysate (40% from Edestin 1 and 24% from Edestin 2). Phenylalanine ammonia-lyase and Ribulose 1,5-bisphosphate carboxylase/oxygenase were instead the most abundant parent proteins in the pancreatic hydrolysate, accounting for 34% and 22%, respectively. On the contrary, Photosystem I P700 chlorophyll (10%), Edestin 2 (8%), NADH-ubiquinone oxidoreductase (6%), and 4-coumarate:CoA ligase (6%) were the parent proteins of most peptides in the co-digested hydrolysate. The Venn diagram (**Figure 7.2B)** highlights the distribution of the parent proteins among the hydrolysates. Only 13 proteins are common to the peptic and the co-digested hydrolysates, whereas none protein is shared by all hydrolysates. These results suggest a very high selectivity of the hydrolytic processes that produced peptide mixtures with different compositions.

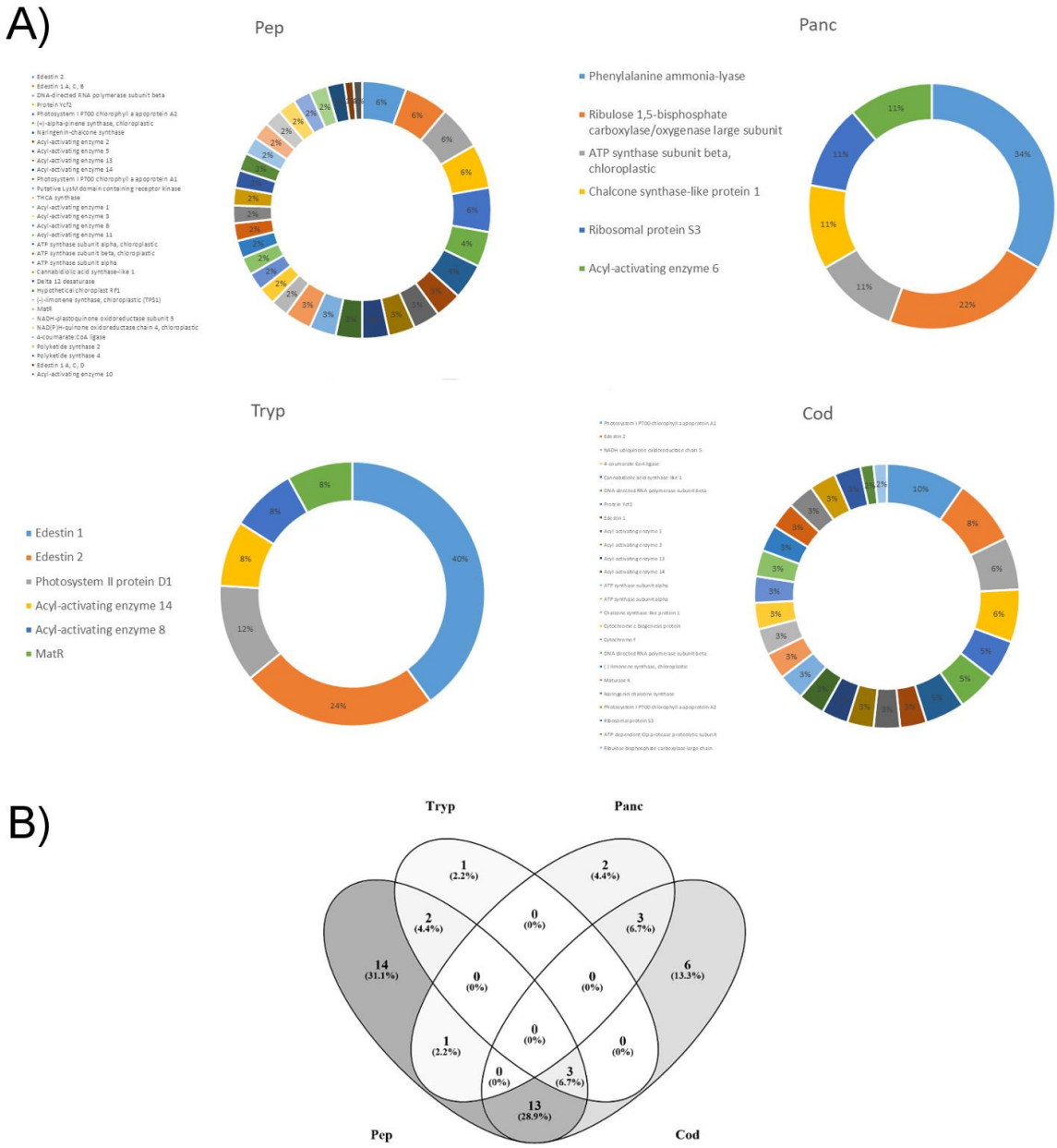


Figure 7.2 - A) Percent distribution of identified peptides according to their parent proteins. B) Venn diagrams of the total number of identified proteins in each hydrolysate.

Based on MS/MS results, the clustering of the molecular weights (MW) distribution of the peptides released after HPC digestion is reported in **Figure 7.3A**. Pepsin hydrolysis produced a high number of peptides that fall into the ranges of 1000-1500 and 2000-2500 Da, whereas the simulated gastrointestinal digestion yielded predominantly peptides in the 1500-2000 Da range. Hierarchical clustering analysis (HCA) was applied to classify all samples according to their amino acid composition (AAC) as reported in **Figure 7.3B.** The tryptic, peptic, and co-digested hydrolysates showed a similar AAC, whereas the pancreatic mixture displayed a different AAC. The amino acid

similarity among all hydrolysates is the driven factory on which the clusters are built. Each step in the clustering process is illustrated by a joint of the tree. Glu, Gly, Pro, Phe, Val, and Try are frequently occurring in the peptic, tryptic, and co-digested hydrolysates, whereas they are poorly expressed in the pancreatic mixture. On the contrary, Met, Arg, Ala, and Cys were the most abundant amino acid residues in the pancreatic mixture.

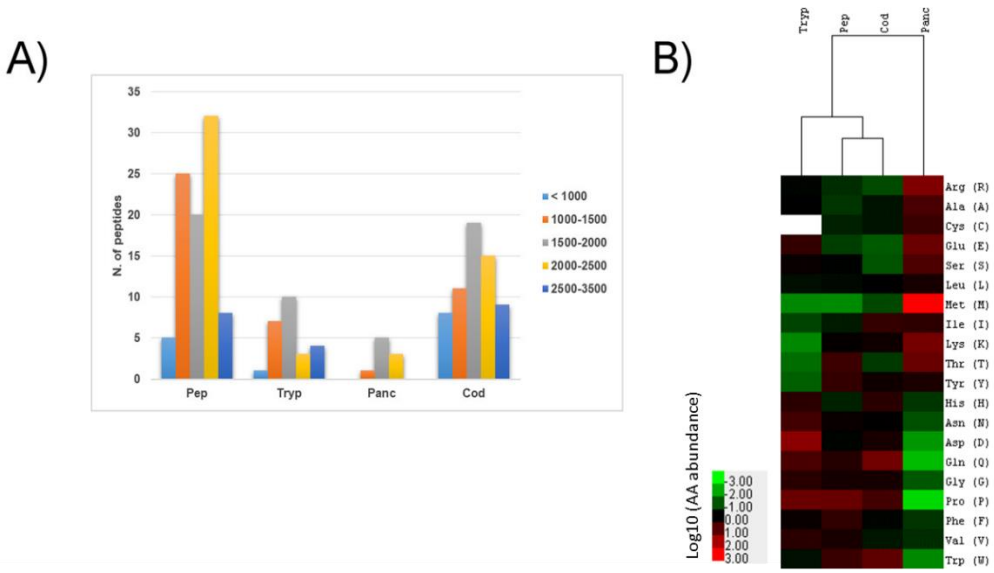


Figure 7.3 - A) MW distribution (in Da) of the identified peptides in each hydrolysate. **B)** Hierarchical clustering analysis (HCA) with dendrogram of amino acid data set composition of each hydrolysate.

7.3.3 Inhibitory effects of the hydrolysates on the HMGCoAR activity

In order to evaluate experimentally the ability of the different hydrolysates to inhibit the activity of HMGCoAR, [23] an *in vitro* assay was performed using the purified catalytic domain of this enzyme. Peptide concentrations ranging from 0.2 to 2.0 mg/mL were tested. **Figure 7.4** shows that, after incubation with the peptic hydrolysate (0.25, 0.5, and 1.0 mg/mL), the HMGCoAR activity was inhibited by 24.5 ± 1.7% (*p<0.001)*, 61.1 ± 0.7% (*p<0.001*), and 80.0 ± 4.0% (*p<0.001*), respectively, *versus* the control. After the incubation with the tryptic hydrolysate (0.2, 0.5, and 1 mg/mL), the HMGCoAR activity was inhibited by 24.6 ± 4.6% ($p<0.05$), 58.4 ± 1.1% ($p<0.001$), 93.3 ± 9.3% ($p<0.001$), respectively, *versus* the control. After incubation with the co-digested hydrolysate (0.2, 0.5, and 1.0 mg/mL), the HMGCoAR activity was inhibited by 16.2 ± 12.6% ($p<0.01$), 50.6 ± 2.3% ($p<0.001$), 47.4 ± 1.5% ($p<0.001$). Finally, after incubation with the pancreatic hydrolysate the HMGCoAR activity was not

significantly inhibited at 1.0 mg/mL, whereas a moderate but significant inhibition by 11.7 ± 6.4% ($p<0.05$) was observed at 2.0 mg/mL.

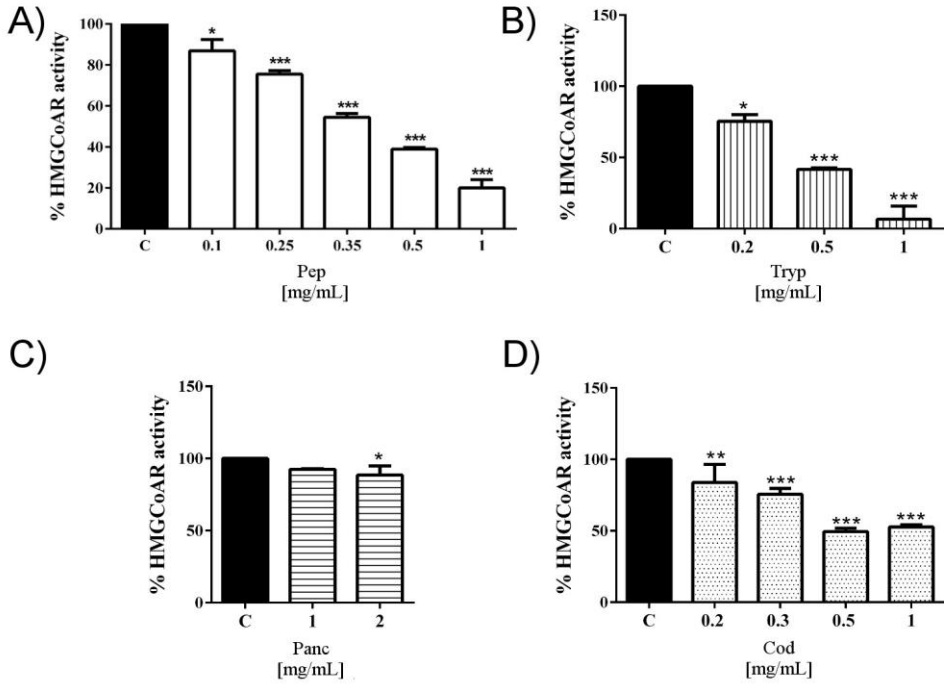


Figure 7.4 - Effect of the hydrolysates on the catalytic domain of HMGCoAR. Bars indicate the effects of each hydrolysate on the HMGCoAR activity at the following concentrations: **(A)** peptic hydrolysate (0.1, 0.25, 0.35, 0.5, and 1.0 mg/mL); **(B)** tryptic hydrolysate (0.2, 0.5, and 1 mg/mL); **(C)** pancreatic hydrolysate (1.0, 2.0 mg/mL); **(D)** co-digested hydrolysate (0.2, 0.5, and 1.0 mg/mL). HMGCoAR, physiologically, catalyzes the four-electron reduction of HMG-CoA to coenzyme A (CoA) and mevalonate (HMG-CoA + 2NADPH + 2H+ > mevalonate + 2NADP+ + CoA-SH). In this assay, the decrease in absorbance at 340 nm, which represents the oxidation of NADPH by the catalytic subunit of HMGCoAR in the presence of the substrate HMG-CoA, was measured spectrophotometrically. Data points represent averages ± SD of three independent experiments in triplicate. (*) $p < 0.05$, (**) $p < 0.001$, and (***) $p < 0.0001$ versus control (C).

7.3.4 Peptide ranking, protein abundance and bioactivity searching

In order to extend the investigation to other potential bioactivities, the peptidome maps were ranked by the tool PeptideRanker. At the end of this process, only those peptides showing score values higher than 0.6 were considered as potentially bioactive. The data reported in **Table 7.1** demonstrate that there is not any correlation between high-scored peptides and the total protein spectral intensity (TPSI) of the proteins from which they are released. In particular, as shown in **Figure 7.5**, a great number of potentially bioactive peptides belong to less abundant proteins. For example, QIQFEGFCRF (score 0.92) derives from DNA-directed RNA polymerase subunit beta,

which is one of the least abundant proteins detected in the hydrolysates. On the contrary, only one peptide derived from Edestin 2, DIFNPRGG (score 0.74), was supposed to be bioactive.

In order to hypothesize their possible bioactivities, the best scored peptides were submitted to BIOPEP search (**Table 7.1**). The alleged biological activities included the inhibition of dipeptidyl peptidase-IV (DPP-IV) and of angiotensin converting enzyme I (ACE), the antioxidant property, and the glucose uptake stimulating activity. Most bioactive peptides were detected in the hydrolysates deriving from the peptic and/or the simulated gastrointestinal digestion. Following the bioinformatic prediction, bioactivities are prevalently provided by short sequences of two or three amino acids included in their structures.

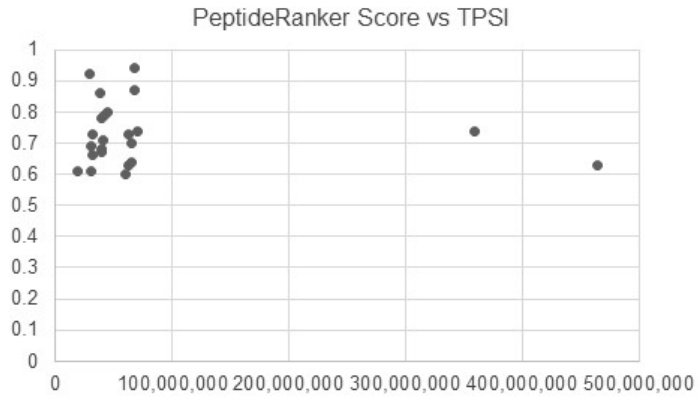


Figure 7.5 - PeptideRanker score of potentially bioactive peptides vs. Total Protein Spectrum Intensity (TPSI) of parent proteins.

Table 7.1 Predicted bioactive peptides by PeptideRanker and BIOPEP

Protein Acc. N.	Peptide sequence	TPSI x 10^6	Enzyme	Score[a]	Potential bioactive peptides[b]	Biological functions[b]
A0A0C5ARZ4	SHLNWVCIFLGFHSFGLYI	67,6	Pep	0.94	GLY	Regulating phospho inositol mechanism peptide
					GF, IF, GL, HL, FG, LG, SF, LN, GLY, IFL	ACE-inhibitor
					HL, LY	Antioxidative
					FL, WV, HL, GL, GF, HS, LN, NW, SF, SH, YI	Dipeptidyl peptidase IV inhibitor
A0A0C5ARQ8	QIQFEGFCRF	29,9	Pep	0.92	RF. GF, EG	ACE-inhibitor
					EG, GF, IQ, QF, QI	Dipeptidyl peptidase IV inhibitor
A0A0C5ARZ4	IPDKANLGFRFP	67,6	Pep	0.87	RF, FP, IP, GF, FR, LG, KA	ACE-inhibitor
					KA, IP, FP, FR, GF, NL	Dipeptidyl peptidase IV inhibitor
A0A0C5B2L0	SSEKGMIATFCCITGLL	38,4	Cod	0.86	IA, GM, GL, KG, TG, EK, TF	ACE inhibitor
					LL	Glucose uptake stimulating peptide
					SE	Stimulating vasoactive substance release
					IA, EK, GL, AT, KG, MI, TF, TG	Dipeptidyl peptidase IV inhibitor
H9A8L3	IPWTQLSPIRCAAESWAHM	44,1	Pep	0.80	IR, LSP, IP, AA, TQ, AH, WA	ACE inhibitor
					AH, PWT, PW, IR	Antioxidative
					IR	Renin inhibitor
					WA	Activating ubiquitin-mediated proteolysis
					IP, SP, WA, AA, WT, AE, AH, ES, IR, PI, PW, QL, SW, TQ	Dipeptidyl peptidase IV inhibitor
C6KI62	PIGISDWNSLFWIVHP	41,9	Cod	0.79	LF, IG, GI, HP	ACE inhibitor
					IV	Glucose uptake stimulating peptide
					HP, SL, WI, WN, GI, PI, VH	Dipeptidyl peptidase IV inhibitor
A0A0C5APZ1	LPDTHGEAHYSTCMLLAGILLKMG	40,2	Pep	0.78	HY, LA, GI, AG, MG, HG, GE, EA, AH, IL, ST	ACE inhibitor
					II, IL	Glucose uptake stimulating peptide
					AH, LK	Antioxidative
					LA	Activating ubiquitin-mediated proteolysis
					LA, LP, LL , AG, AH, GE, GI, HY, IL, MG, ML, TH, YS	Dipeptidyl peptidase IV inhibitor
A0A088MFF4	PGRVLSLFVTLTLGWPLY	69,9	Pep	0.74	PG	Prolyl endopeptidase inhibitor
					LY, LF, PL, GW, GR, LG, PG	ACE inhibitor
					VL	Glucose uptake stimulating peptide
					PG	Peptide regulating the stomach mucosal membrane activity
					LY, WPL	Antioxidative
					WP, SL, PL, GW, LT, PG, TL, VL, VT	Dipeptidyl peptidase IV inhibitor

A0A090CXP8	DIFNPRGG	359,0	Cod	0.74	PR, IF, GG	ACE inhibitor
					NP, FN, GG, RG	Dipeptidyl peptidase IV inhibitor
A6P6W0	RIWGEKYFGKNFNRLVKVK	63,0	Pep	0.73	RL, FGK, IW, VK, FG, GK, WG, GE, NF, KY, EK	ACE inhibitor
					LV	Glucose uptake stimulating peptide
					VKV, WG	Antioxidative
					EK, WG, FN, GE, IW, KV, KY, LV, NF, NR, RI, RL, RL, VK, YF	Dipeptidyl peptidase IV inhibitor
A7IZZ1	VRFEPQFSYFRI	40,9	Pep	0.71	RF, FR, VR, SY, PQ, FEP	ACE inhibitor
					EP, VR, FR, PQ, QF, RI, SY, YF	Dipeptidyl peptidase IV inhibitor
E5DL82	PRNSWISCNMRLNAITL	64,8	Cod	0.70	RL, PR, LN, AI	ACE inhibitor
					WI, LN, MR, NA, NM, RL, RN, SW, TL	Dipeptidyl peptidase IV inhibitor
H9A1V5	NDVKKFIAGQVASFKRL	30,8	Cod	0.69	RL, VK, IA, KR, AG, GQ, SF, KF, FKR	ACE inhibitor
					GQ	Neuropeptide
					KK	Bacterial permease ligand
					KF	Renin inhibitor
					VA, GQ, IA, AG, AS, KF, KK, KR, ND, QV, RL, SF, VK	Dipeptidyl peptidase IV inhibitor
H9A8L2	SPIGGGPEQLVMFVVLKNGY	39,9	Cod	0.68	GP	Prolyl endopeptidase inhibitor
					MF, GY, GP, IG, GG, NG	ACE inhibitor
					GP	Antithrombotic
					VL, LV	Glucose uptake stimulating peptide
					GP	Peptide regulating the stomach mucosal membrane activity
					LK	Antioxidative peptide
					GP, VV, SP, GG, GY, LV, MF, NG, PI, QL, VL, VM	Dipeptidyl peptidase IV inhibitor
H9A8L2	GAVLNIAECCLLPTSYPRKDD	39,9	Pep	0.67	YPR, PR, YP, LLP, IA, GA, SY, IAE, LN, PT, AV, AVL	ACE inhibitor
					VL, LL	Glucose uptake stimulating peptide
					KD	Antioxidative peptide
					LP, LL, YP, GA, IA, AE, AV, LN, PT, RK, SY, TS, VL	Dipeptidyl peptidase IV inhibitor
A0A0E3TIL1	NPRENFLKCFSKHIPNNVA	31,9	Pep	0.66	PR, IP, NF, CF	ACE inhibitor
					LK	Antioxidative peptide
					VA, IP, NP, FL, HI, KH, NF, NN, NV, PN, SK	Dipeptidyl peptidase IV inhibitor
E5DL82	LICILLFIGAVGKS	64,8	Cod	0.64	LF, VG, IG, GA, GK, LLF, AV, IL	ACE inhibitor
					IL, LI, LL	Glucose uptake stimulating peptide
					LL, GA, AV, IL, KS, LI, VG	Dipeptidyl peptidase IV inhibitor
A0A0C5AUJ6	GPTPISALIHAATM	465,0	Pep	0.63	GP	Prolyl endopeptidase inhibitor
					GP, AA, PT, TP, TP	ACE inhibitor
					LI	Glucose uptake stimulating peptide

a. From PeptideRanker

					GP	Peptide regulating the stomach mucosal membrane activity
					GP, HA, TP, AL, AA, AT, IH, LI, PI, PT, TM	Dipeptidyl peptidase IV inhibitor
A6P6W0	GEKYFGKNFNRLVKVKT	63,0	Cod	0.63	RL, FGK, VK, FG, GK, GE, NF, KY, EK	ACE inhibitor
					LV	Glucose uptake stimulating peptide
					VKV	Antioxidative
					EK, FN, GE, KT, KV, KY, LV, NF, NR, RL, VK, YF	Dipeptidyl peptidase IV inhibitor
A0A0C5B2I8	ATGRIVCANCHLANKPVDIEVP	19,4	Cod	0.61	LA, VP, HL, GR, TG, NK, KP, IE, EV	ACE inhibitor
					IV	Glucose uptake stimulating peptide
					HL, KP	Antioxidative
					LA	Ubiquitin-mediated proteolysis activating peptide
					LA, VP, KP, HL, AT, EV, PV, RI, TG, VD	Dipeptidyl peptidase IV inhibitor
H9A1V5	ALSKNSMVKKFNLSSIKYIG	30,8	Pep	0.61	VK, IG, KY, KF, IKY	ACE inhibitor
					KF	Renin inhibitor
					AL, FN, KF, KK, KY, MV, NL, SI, SK, VK, YI	Dipeptidyl peptidase IV inhibitor
E5DKP2	YSIQKVFSAGRLVGGEKGPYSV	60,7	Tryp	0.60	GP	Prolyl endopeptidase inhibitor
					RL, VF, GP, VG, AG, GR, KG, GE, GG, QK, EK, KGP, EKGP	ACE inhibitor
					LV	Glucose uptake stimulating peptide
					GP	Peptide regulating the stomach mucosal membrane activity
					GGE	Antioxidative
					GP, EK, AG, GE, GG, IQ, KG, KV, LV, PY, RL, SI, SV, VF, VG, YS	Dipeptidyl peptidase IV inhibitor

b. From BIOPEP

7.4 Discussion

The first objective of the study was the characterization of the composition of the four hydrolysates, an important step in the pathway to evaluate their potential use as functional ingredients. Having this final goal, all enzymatic digestions were performed avoiding the use of reducing and alkylating agents in order to produce, as far as possible, unmodified and natural peptides. This choice of course partially impaired the hydrolytic efficiency of the enzymes.

All employed enzymes were endopeptidases that, with the exception of trypsin, randomly hydrolyze peptide bonds within the protein sequences, producing peptides differing in amino acid sequences and sizes. The resulting hydrolysates were diverse in terms of composition and percent yield of peptides. Pancreatin gave the highest percent yield: this may be attributed to the endo- and exo-peptidase activities of this enzyme, which increases protein digestion through hydrolysis of more peptide bonds, when compared to enzymes only endowed with an endopeptidase activity.[28] Another high percent yield was observed with trypsin, even if this enzyme is well known for the high selectivity and specificity in site cutting recognition. A higher peptide yield is the expected outcome for increased protein breakdown and a marker of the hydrolytic process efficacy. The consequence of the high hydrolytic efficiency of these enzymes was the relatively small number of peptides that were identified in their hydrolysates, probably due to an extensive production of very short peptides, i.e. di-, tri- and tetra peptides, which are difficult to detect using a data-dependent shotgun approach.

The identified peptides ranged from 7 to 29 amino acid residues, i.e. between 747 Da and 3211 Da, in agreement with the ultrafiltration separation that had been performed with a cut-off of 3 kDa. These values correspond to peptides slightly longer than those reported in other studies after similar digestions of various animal proteins.[29] Possibly, this might be explained by the protease inhibitors present in most plant seeds. In addition, the missing reduction and alkylation of the disulfide bonds reduced the proteolytic activity of each enzyme resulting in longer peptides.

In order to investigate the different features of the hydrolysates, the HCA of the AAC of all identified peptides was employed combining the heat map with a dendrogram. The clustering provides the basis for guiding reasonable enzyme selection in order to produce hydrolysates endowed of specific chemicals features. As shown by **Figure 7.3B**, peptic and co-digested hydrolysates form two very close clusters according to their amino acid similarity. The AAC of the tryptic hydrolysate has also some similarity with the peptic and co-digested hydrolysates, but it falls at a wider distance. On the contrary, the pancreatic hydrolysate is well separated from the others.

The peptic and codigested clusters are near, possibly the effects of pepsin prevails since this enzyme is the first applied in the codigestion. The tryptic cluster falls at a certain distance, since its cut sites are different from those recognized by pepsin. Finally, the pancreatic one is the most distant, since in this case the peptide hydrolysate is generated either by endo- or exopeptidase action.

Apparently, their substantial structural diversities reflect also their different capability of inhibiting the activity of HMGCoAR, a key enzyme in the synthesis of endogenous cholesterol and the main target of statins, which interact with this enzyme as competitive inhibitors.[30] The experiments performed using the purified catalytic domain of this enzyme showed that the tryptic and peptic hydrolysates were the best inhibitors (**Figure 7.4**), whereas the pancreatin hydrolysate was the least active, in line with the clustering provided by the HCA analysis. The pancreatic hydrolysate contains numerous residues of hydrophilic amino acids, such as Glu, Ser, Arg, and Lys, whereas it is completely devoid of Pro and Val, which are abundant in the other hydrolysates. This may be related to a synergistic effect of the hydrophobic peptides present in the mixtures, since the hypocholesterolemic effect is correlated to an increased hydrophobicity.[31]

In a previous paper, we have investigated the interaction of some soy peptides with the catalytic domain of HMGCoAR using *in silico* modeling studies.[32] Medium size peptides, containing 8-10 amino acid residues and characterized by a hydrophobic N-terminus and a negatively charged C-terminus, appeared to be particularly favorable for interacting with HMGCoAR. The negatively charged C-terminal portion is primarily involved in the inhibition by mimicking most of the polar interactions that are clearly seen also in statins,[32] whereas the hydrophobic N-terminal portion is inserted in a deeper and rather polar sub-pocket that corresponds roughly to that harboring the NADPH cofactor. In fact, HMGCoAR contains a second relevant domain that is capable of accepting NADPH: a peptide that prevents this binding impairs the catalytic activity of the enzyme.[32]

In order to predict other potential activities, it was decided to use cost effective and time-saving computer simulated approaches, such as PeptideRanker and BIOPEP. It is important, however, to underline that these tools take into consideration only some possible biological activities, excluding for example the inhibition of HMGCoAR. At the end of the procedure, 22 peptides were postulated to be bioactive. In agreement with the features of BIOPEP, the main proposed activities were the inhibition of dipeptidyl peptidase-IV (DPP-IV) and of angiotensin converting enzyme I (ACE), the antioxidant properties, and the glucose uptake stimulating activity. The forecasted activities are linked to specific short sequences, mostly composed by two or three amino acid

residues, encrypted in their sequences. Specifically, PWT, WPL, and VKV provide antioxidant activities, which could depend on the presence of substantial amounts of hydrophobic, branched-chain, or aromatic amino acid residues. The hydrophobicity is also important to enhance their permeability into the target organs through hydrophobic associations with the cell membrane lipid bilayer, promoting the achievement of potent antioxidant effects.[33] In the meanwhile, the compresence of hydrophobic or aromatic amino acids at the C-terminus as well as of hydrophobic and negative ionizable functions (Thr, Glu, Asp, Ser, Met) positively contributes to an effective ACE inhibition.[1, 34, 35] Specifically, the sequences LSP, FEP, and IAE fall in this category. However, also branched-chain aliphatic amino acids at the C-terminus as well as hydrophobic amino acids at the N-terminus, such as IFL, LLP, AVL, LLF, contribute to high ACE-inhibitory activities.[28, 36] Finally, also the small sequence GP is expected to provide bioavailability and ACE-inhibitory activity owing to its short sequence, hydrophobicity, and theoretical stability to pepsin and trypsin cleavage.[37] The ACE inhibitory activity as well as the antioxidant activity of hempseed hydrolysates are confirmed experimentally by literature.[28]

Another proposed activity is the inhibition of DPP-IV. This enzyme is a new molecular target correlated with the development of type 2 diabetes.[38] The peptides capable of inhibit the DPP-IV activity have in general a hydrophobic character and a length from 2 to 8 amino acids. Very often, they contain a Pro residue in their sequence located at the first, second or third or fourth N-terminal position, which is flanked by Leu, Val, Phe, Ala or Gly.[39, 40] A paper has investigated the DPP-IV inhibitory activity of hempseed protein hydrolysates obtained treating a HPC with different enzymes obtaining a first indication of a moderate activity also on this enzyme.[41]

In conclusion, it seems possible to affirm that there are good prospective that hempseed hydrolysates may be used as multipurpose ingredients in functional foods. Of course, when discussing the bioactivities of food peptides an open issue remains their bioavailability. In case of hempseed peptides, this problem is still to be taken into consideration. However, it is useful to remind that different authors have confirmed the bioavailability of peptides deriving from other food proteins. [23, 42-44]

Author contributions

Experiment ideation and design: GA and CL. Experiments & data analysis: protein hydrolysate preparation & mass spectrometry: GA; inhibition of HMGCoAR activity: CL & CZ. Manuscript writing: GA, CL & AA, figure preparation CZ.

SUPPLEMENTARY INFORMATION

Table 7.1S LC-ESI-MS/MS based identification of hydrolysates of hempseed proteins

Accession n. [a]	Protein Name	Start-end	Sequences	pI[b]	Net Charge [b]	Hydrophobicity % [c]	m/z (Da) (charge)	$[M+H]^+$ (Da)
Peptic								
A0A090CXP8	Edestin 2	176-187	(D)WVYNNGDSPLVL(I)	0.7	-1	50	688.60 (2)	1376.69
		361-370	(V)LYKNGMMAPH(F)	9.7	1.1	50	581.74 (2)	1161.56
		380-403	(I)YVTRGSARLQVVDDNGRNVFDGEL(R)	4.3	-1	33.3	894.16 (3)	2680.34
		435-447	(N)DNAMRNPLAGKVS(A)	10.2	1	46.2	458.55 (3)	1372.70
		235-247	(R)RESGEQTPNGNIF(S)	4.2	-1	23.1	724.68 (2)	1448.67
A0A090CXP7	Edestin 1	450-460	(A)WVSPLAGRTSV(I)	10.7	1	54.6	586.80 (2)	1172.64
		178-187	(L)LDTSNVNNQL(D)	0.7	-1	30	559.84 (2)	1117.55
		279-288	(D)LVSPLRSSQE(H)	6.9	0	40	558.08 (2)	1115.61
		63-73	(L)IESWNPNHNQF(Q)	5.1	-0.9	36.4	693.73 (2)	1385.62
		461-469	(V)IRALPEAVL(A)	6.9	0	77.8	491.16 (2)	981.61
A0A090DLH8	Edestin 1	392-409	(M)YVLRGRARVQVVNHMGQKC(F)	11.2	4	36.8	738.37 (3)	2214.19
H9A1V3	Acyl-activating enzyme 1	625-639	(I)ERVCNEVDDRVFETT(A)	3.8	-3.1	26.7	623.95 (3)	1868.84
		170-190	(G)GYLNSAKNCLNVNSNKKLNDT(M)	9.6	1.9	23.8	789.33 (3)	2367.17
H9A1V4	Acyl-activating enzyme 2	280-298	(H)IFDRVIEELFILHGASIGF(W)	4.3	-1.9	57.9	725.95 (3)	2176.18
		28-44	(Y)RSMYAKDGFPPPIDGLD(C)	4.0	-1	47.1	627.21 (3)	1878.91
		442-453	(G)PPVPNVDVCLES(V)	0.7	-2.1	58.3	442.71 (3)	1325.64
H9A1V5	Acyl-activating enzyme 3	293-312	(L)ALSKNSMVKKFNLSSIKYIG(S)	10.8	4	40	743.33 (3)	2228.25
		360-382	(N)SGSAGMLASGVEAQIVSVDTLKP(L)	3.9	-1	47.8	739.89 (3)	2217.14
H9A1V7	Acyl-activating enzyme 5	253-266	(G)YTWGTAAVGATNVC(L)	3.1	-0.1	42.9	491.20 (3)	1470.67
		497-512	(F)VTLKKGAVRVTVTEKE(I)	10.4	2	37.5	586.93 (3)	1758.05
		54-62	(T)RCLRVASCI(E)	8.8	1.9	44.4	511.27 (2)	1020.55
H9A1W0	Acyl-activating enzyme 8	352-373	(D)QNGSAQLAGVSGEVCIRGPNVT(K)	6.1	-0.1	36.4	738.96 (3)	2214.09
		166-189	(D)VALFLHTSGTTSRPKGVPLTQLNL(A)	11.4	2.1	45.8	850.97 (3)	2550.44
H9A1W2	Acyl-activating enzyme 10	138-154	(Q)NIAAKTSAQFSLIPSVP(S)	9.7	1	58.8	582.30 (3)	1743.96
H9A1W3	Acyl-activating enzyme 11	260-268	(F)EMKKMVELI(E)	7.0	0	55.6	560.8 (2)	1120.61
		8-14	(F)IFRSKLP(D)	11.4	2	57.1	430.58 (2)	860.54

H9A8L2	Acyl-activating enzyme 13	160-180	(P)GAVLNIAECCLLPTSYPRKDD(D)	4.2	-1.1	42.9	760.06 (3)	2278.12
		289-309	(P)LYSRVVEAAPDRVIVLPATGS(N)	6.9	0	57.1	738.54 (3)	2213.23
		535-547	(Y)PDDQACTGEVGLI(P)	0.5	-3.1	38.5	459.05 (3)	1374.62
H9A8L3	Acyl-activating enzyme 14	374-392	(A)IPWTQLSPIRCAAESWAHM(D)	7.1	0	57.9	752.00 (3)	2254.09
		598-615	(I)KRTVGGYFIVQGRADDTM(N)	9.5	1	33.3	672.24 (3)	2014.02
		631-652	(V)CDRADESIVETAAVSVSPVDGG(P)	3.3	-4.1	40.9	744.98 (3)	2234.02
A7IZZ2	(+)-alpha-pinene synthase, chloroplastic	270-283	(I)RAEAKWFIEEYEKT(Q)	4.6	-1	35.7	600.89 (3)	1799.90
		592-606	(G)DGHASQDSHSRKRIS(D)	10.1	1.2	13.3	560.90 (3)	1680.82
		319-336	(H)SELGKNKMVYARDRLVEA(F)	9.4	1	38.9	693.56 (3)	2079.10
		185-201	(I)FNDFKDETGKFKASIKN(D)	9.5	1	29.4	663.93 (3)	1989.01
A0A0C5ARX6	ATP synthase subunit alpha	123-131	(I)STSESRLIE(S)	4.2	-1	22.2	511.23 (2)	1021.52
		134-154	(P)APGIISRRSVYEPLQTGLIAI(D)	9.9	1	52.4	751.82 (3)	2254.29
A0A0C5ARS5	ATP synthase subunit beta	382-405	(G)EEHYETAQRVKQTLQRYKELQDII(A)	5.5	-0.9	25	1007.12 (3)	3018.56
		144-158	(D)TKLSIFETGIKVVDL(L)	6.6	0	46.7	554.85 (3)	1662.97
E5DK51	ATP synthase subunit alpha	151-166	(E)TLYCVYVAIGQKRSTV(A)	9.4	1.9	37.5	620.24 (3)	1857.99
		287-309	(D)VSAYIPTNVISITDGQICLETEL(F)	0.6	-3.1	43.5	846.10 (3)	2536.29
A6P6W0	Cannabidiolic acid synthase-like 1	504-522	(A)RIWGEKYFGKNFNRLVKVK(T)	11.1	5	36.8	794.75 (3)	2382.36
		91-104	(V)SHIQGTILCSKKVG(L)	9.7	2	28.6	491.22 (3)	1470.81
A0A088MFF4	Delta 12 desaturase	179-196	(P)PGRVLSLFVTLTLGWPLY(L)	10.3	1	61.1	677.78 (3)	2032.16
		331-345	(Y)NAMEATKAVKPILGE(Y)	6.6	0	53.3	524.23 (3)	1571.85
A0A0C5ARQ8	RNA polymerase subunit beta	1047-1063	(L)RSLALELNHFLVSEKNF(Q)	7.5	0.1	47.1	672.75 (3)	2017.09
		549-568	(M)QRQAVPLSRSEKCIVGTGLE(S)	8.6	0.9	35	743.36 (3)	2228.18
		743-752	(L)TPQMAKESSY(A)	6.5	0	30	380.81 (3)	1141.52
		358-377	(T)STTLTTTFESFFGLHPLSQV(L)	5.1	-0.9	40	738.8 (3)	2213.11
		17-26	(N)QIQFEGFCRF(I)	6.1	-0.1	40	666.86 (2)	1331.62
A0A0C5AS14	Hypothetical chloroplast RF1	341-355	(Q)ENSKLEILNEKKGVN(Y)	7.1	0	26.7	572.56 (3)	1714.93
		259-279	(T)DVEIETTSETKGTKQEQGGST(E)	3.9	-3	9.5	742.45 (3)	2225.04
A7IZZ1	(-)-limonene synthase, chloroplastic	180-200	(L)RQYGFEVPQEIFNNFKNHKTG(E)	9.4	1.1	28.6	851.36 (3)	2553.26
		349-360	(G)VRFEPQFSYFRI(M)	9.8	1	50	794.79 (2)	1588.83
E5DKP2	MatR	382-400	(G)VQLAETLGTAGVRGPQVSV(L)	6.8	0	47.4	627.62 (3)	1882.04
		242-250	(R)KLAAPLKTH(Y)	10.7	2.1	55.6	489.57 (2)	978.61

A0A0C5AUJ6	NADH-plastoquinone oxidoreductase subunit 5	603-622	(M)DWNWYEFLTNATFSVSIASL(G)	0.6	-2	50	788.88 (3)	2364.12
		256-269	(E)GPTPISALIHAATM(V)	7.8	0.10	64.3	690.19 (2)	1379.74
A0A0C5APZ1	NAD(P)H-quinone oxidoreductase chain 4	234-257	(W)LPDTHGEAHYSTCMLLAGILLKMG(A)	6.1	-0.9	45.8	876.6 (3)	2628.30
		230-238	(P)LHTWLPDTH(G)	6.0	-0.8	44.4	373.88 (3)	1119.56
Q8RVK9	Naringenin-chalcone synthase	353-368	(K)CVEDGLNTTGEGLEWG(V)	0.5	-4.1	25	560.86 (3)	1679.72
		301-324	(W)IAHPGGPAILDQVESKLALKTEKL(R)	7.8	0.1	50	843.79 (3)	2528.45
		236-250	(P)IFELVSAAQTILPDS(D)	0.7	-2	60	535.27 (3)	1603.86
		183-201	(K)GARVLVVCSEITAVTFRGP(N)	8.9	0.9	52.6	678.37 (3)	2032.10
V5KXG5	4-coumarate:CoA ligase	262-281	(G)ATILIMPKFEIGSLLGLIER(Y)	7.1	0	60	738.78 (3)	2214.29
		17-23	(I)IFRSKLP(D)	11.4	2	57.1	430.58 (2)	860.54
F1LKH7	Polyketide synthase 2	371-385	(G)LTVERVVLRSVPINY(-)	9.8	1	53.3	586.93 (3)	1758.03
		303-310	(A)ILDKVEEK(L)	4.3	-1	37.5	487.31 (2)	973.56
F1LKH8	Polyketide synthase 4	2-16	(M)NHLRAEGPASVLAIG(T)	7.4	0.1	53.3	502.22 (3)	1504.82
		256-266	(A)GLIFDLHKDVP(M)	5.0	-0.9	54.6	627.28 (2)	1253.69
A0A0C5APZ4	Protein Ycf2	1630-1650	(P)FSLRLALSLSRGILVIGSIGT(G)	12.1	2	52.4	725.26 (3)	2173.31
		1902-1921	(Q)DHGILFYQIGRAVAQNVLLS(N)	7.8	0.1	50	738.97 (3)	2214.20
		1092-1102	(T)ISPIELQVSNI(F)	0.9	-1	54.6	404.94 (3)	1212.68
		536-547	(S)ENKEIVNIFKII(T)	7.0	0	50	487.41 (3)	1459.85
		143-152	(L)YLPKGKKISE(S)	10.1	2	30	581.39 (2)	1162.68
A0A0C5ARZ4	Photosystem I P700 chlorophyll a apoprotein A1	436-454	(I)SHLNWVCIFLGFHSFGLYI(H)	7.2	0.1	52.6	751.46 (3)	2253.13
		102-122	(W)LSDPTHIGPSAQVVWPIVGQE(I)	3.9	-1.9	52.4	744.39 (3)	2230.15
		561-572	(L)IPDKANLGFRFP(C)	10.1	1	58.3	458.58 (3)	1374.75
A0A0C5APY0	Photosystem I P700 chlorophyll a apoprotein A2	188-207	(S)LAWTGHLVHVAIPGSRGESV(R)	8.0	0.2	50	695.75 (3)	2086.12
		247-257	(T)SQGAGTSILTL(L)	3.4	0	36.4	524.2 (2)	1047.57
		241-258	(S)SHLFGTSQGAGTSILTLL(G)	7.5	0.1	38.9	601.63 (3)	1802.97
		352-369	(H)MYSLPAYAFIAQDFTTQA(A)	0.7	-1	55.6	680.29 (3)	2037.96
		695-708	(R)DKPVALSIVQARLV(G)	10.2	1	64.3	503.33 (3)	1508.92
U6EFF4	Putative LysM domain containing receptor kinase	398-417	(H)LRGSGRDPLTWSSRVQIALD(S)	10.5	1	40	743.19 (3)	2227.20
		125-144	(V)HRVNMFKPTRIPAGSPINVT(V)	12.1	3.1	50	745.35 (3)	2235.22
		115-142	(A)FANLTTEDWVHRVNMFKPTRIPAGSPIN(V)	9.9	1.1	50	1071.23 (3)	3211.65

A0A0E3TIL1	THCA synthase	87-102	(T)PSNNSHIQATILCSKK(V)	10.3	2	31.3	580.67 (3)	1740.91
		29-47	(A)NPRENFLKCFSKHIPNNVA(N)	9.6	2	42.1	743.33 (3)	2228.14
		504-522	(A)RIWGEKYFGKNFNRLVKVK(T)	11.1	5	36.8	794.75 (3)	2382.36
Tryptic								
A0A090CXP7	Edestin 1	245-258	(R)YLEEAFNVDSETVK(R)	3.5	-3	35.7	822.32 (2)	1643.78
		194-208	(R)FYLAGNPEDEFEQLR(R)	3.5	-3	40	914.4 (2)	1827.86
		275-284	(K)GTLDLVSPLR(S)	6.6	0	50	535.76 (2)	1070.62
		417-431	(R)QGQIVTVPQNHAVVK(Q)	9.84	1.1	46.7	540.12 (3)	1617.91
		176-192	(V)SLLDTSNVNNQLDDNPR(R)	3.4	-2	29.4	958.19 (2)	1914.92
		198-208	(A)GNPEDEFEQLR(R)	3.5	-3	27.3	667.24 (2)	1333.60
		485-504	(K)YNREETVLLTSSTSSRREDR(Y)	6.84	0	15	800.87 (3)	2399.19
		56-73	(R)VEAEAGLIESWNPNHNQF(Q)	3.8	-2.9	44.4	685.85 (3)	2054.96
		246-258	(Y)LEEAFNVDSETVK(R)	3.54	-3	38.5	740.98 (2)	1480.72
		150-160	(R)EGDIVAIPAGV(A)	0.7	-2	63.6	520.65 (2)	1040.56
A0A090CXP8	Edestin 2	284-294	(R)GEDLQIIAPSR(I)	3.9	-1	45.5	599.82 (2)	1198.64
		320-334	(R)QNIDRPSQADIFNPR(G)	6.5	0	40	591.18 (3)	1770.89
		397-404	(R)NVFDGELR(E)	3.9	-1	37.5	475.38 (2)	949.47
		256-268	(L)AESFNVDTELAHK(L)	4.3	-1.9	38.5	487.49 (3)	1460.70
		163-176	(K)EGDMVAMPAGVADW(V)	0.48	-3	64.3	724.73(2)	1448.62
		427-439	(F)EWIAVKTNDNAMR(N)	6.9	0	46.2	774.39 (2)	1547.76
A0A0C5ARX1	Photosystem II protein D1	196-222	(H)PFHMLGVAGVFGGSLFSAMHGSLVTSS(L)	8.4	0.2	51.9	898.35 (3)	2693.32
		64-92	(I)REPVSGSLLYGNNIISGAIIPTSAAIGLH(F)	7.51	0.1	48.3	974.32 (3)	2920.59
		332-350	(M)HERNAHNFPLDLAALEVPS(T)	5.1	-1.8	52.6	710.87 (3)	2130.07
H9A8L3	Acyl-activating enzyme 14	708-722	(K)LLRRVLRDQIKHELS(A)	11.12	2.1	40	938.85 (2)	1876.12
		214-240	(T)LKQLREQVISVAKALDAMFSKGDAIAI(D)	9.8	1	55.6	972.27 (3)	2915.64
H9A1W0	Acyl-activating enzyme 8	412-425	(R)IKELINRGGEKISP(I)	9.9	1	35.7	777.94 (2)	1553.90
		424-440	(I)SPIEVDAVLLSHPDISH(A)	4.2	-2.8	52.9	610.51 (3)	1828.95
E5DKP2	MatR	386-411	(A)ETLGTAGVRGPQVSVLWGTVKHIRQG(S)	11.2	2.1	38.5	916.57 (3)	2746.51
		38-59	(F)YSIQKVFSAGRLVGGEKGPYSV(P)	9.9	2	36.4	781.76 (3)	2342.25
Pancreatic								
A0A0C5ARS5	ATP synthase subunit beta	72-89	(N)NRVRAVAMSATDGLMRGM(D)	11.8	2	50	646.18 (3)	1935.97
C6KI62	Chalcone synthase-like protein 1	92-110	(P)SLNTRQDMLVVEIPKLGKE(A)	6.8	0	42.1	724.24 (3)	2170.19
V5KWZ6	Phenylalanine ammonia-lyase	489-503	(S)SRKSAESIEILKLMS(T)	9.7	1	40	846.86 (2)	1691.94

		357-371	(E)VIRFSTKSIEREINS(V)	10.1	1	33.3	890.50 (2)	1778.98
		495-510	(E)SIEILKLMSTTFLVAL(C)	6.6	0	62.5	890.23 (2)	1779.03
E5DLX2	Ribosomal protein S3	318-332	(C)WKLMGKDKVMELIEK(F)	9.6	1	46.7	616.30 (3)	1848.01
A0A0C5B2I6	Ribulose bisphosphate carboxylase large chain	213-233	(G)HYLNATAGTCEEMMKRAIFAR(E)	8.6	1	42.9	805.13 (3)	2413.16
		84-95	(R)CYHIEPVAGEEN(Q)	3.84	-3	33.3	680.69 (2)	1360.58
H9A1V8	Acyl-activating enzyme 6	54-76	(T)YRRCRQFASALTTHSIALGSTVA(V)	10.86	3	39.1	837.39 (3)	2509.31
Co-digested								
A0A090CXP8	Edestin 2	73-80	(Y)WDIQNTED(D)	0.7	-3	25	510.71 (2)	1020.43
		113-120	(F)YVIQGRGI(H)	9.6	1	37.5	453.06 (2)	905.52
		79-90	(T)EDDELHCAGVET(A)	3.3	-5	25	687.88 (2)	1374.55
		329-336	(A)DIFNPRGG(R)	6.7	0	37.5	438.07 (2)	875.44
		449-456	(A)MRAMPDDV(L)	3.7	-1	62.5	467.66 (2)	934.41
A0A090CXP7	Edestin 1	212-221	(G)GRGARFDERI(R)	10.6	1	30	589.35 (2)	1176.62
		69-89	(P)NHNQFQCAGVAVVRYTIQQNG(L)	8.4	1	33.3	801.72 (3)	2404.16
H9A1V3	Acyl-activating enzyme 1	242-255	(H)VDAVVIYLAIVLAG(Y)	0.7	-1	78.6	708.26 (2)	1415.85
		667-678	(T)IDLNQLRLSFNL(G)	6.7	0	50	482.80 (3)	1445.81
H9A1V5	Acyl-activating enzyme 3	395-411	(K)GPNMMQGYFNNPQATKL(T)	9.7	1	41.2	637.40 (3)	1910.89
		497-513	(E)NDVKKFIAGQVASFKRL(R)	10.9	3	47.1	641.19 (3)	1921.10
H9A8L2	Acyl-activating enzyme 13	259-285	(A)PKEIASRLHVSQAKAIFTQDFILRGGR(K)	11.4	3.1	44.4	1013.38 (3)	3038.70
		635-654	(V)SPIGGGPEQLVMFVVLKNGY(D)	6.6	0	50	702.39 (3)	2105.11
H9A8L3	Acyl-activating enzyme 14	311-330	(P)DKVIVLPAIGSNVGIQLREQ(D)	7.0	0	50	717.00 (3)	2149.24
		537-556	(L)VILDEHGNPFPDDQACIGEV(G)	3.25	-5	45	723.31 (3)	2168.00
E5DK51	ATP synthase subunit alpha	271-284	(G)AGSLTALPVIETQA(G)	0.9	-1	57.1	686.28 (2)	1370.75
		90-110	(I)ERKSVHEPMQTGLKAVDSLVP(I)	7.8	0.1	42.9	774.31 (3)	2321.23
A0A0C5ARX6	ATP synthase subunit alpha	149-168	(Q)TGLIAIDSMIPIGRGQRELI(I)	6.6	0	50	718.21 (3)	2153.21
		441-459	(T)GTNGYLDSLEIGQVRKFLL(E)	6.9	0	36.8	708.44 (3)	2123.15
A0A0C5AS06	ATP-dependent Clp protease proteolytic subunit	26-40	(R)LYRERLLFLGQEVDS(E)	4.3	-1	40	919.74 (2)	1837.98
A6P6W0	Cannabidiolic acid synthase-like 1	366-383	(F)KKEILLDRSGGRKAAFSI(K)	10.9	3	38.9	663.77 (3)	1989.16
		507-523	(W)GEKYFGKNFNRLVKVKT(K)	10.76	4	29.4	677.11 (3)	2028.14
		86-101	(T)TPLNVSHIQGTILCSK(K)	8.55	1	37.5	856.32 (2)	1710.92
C6KI62	Chalcone synthase-like protein 1	292-307	(N)PIGISDWNSLFWIVHP(G)	4.9	-0.9	62.5	627.45 (3)	1880.97
		61-76	(R)ICEKTTIKKRHLYLTE(D)	9.6	2	25	659.21(3)	1976.10

A0A0C5B2L0	Cytochrome c biogenesis protein CcsA	40-56	(D)SSEKGMIATFCCITGLL(V)	5.98	-0.1	41.2	591.68 (3)	1773.86
		287-298	(N)SAIVASIGFLII(W)	3.4	0	75	602.8 (2)	1203.74
A0A0C5B2I8	Cytochrome f	203-212	(N)NNVYNATAAG(I)	3.2	0	40	497.42 (2)	994.46
		50-71	(E)ATGRIVCANCHLANKPVDIEVP(Q)	6.9	0	50	774.31 (3)	2320.19
A0A0C5ARX9	DNA-directed RNA polymerase subunit beta"	1235-1241	(T)SKVVVSE(D)	6.6	0	42.9	748.35 (1)	747.43
		790-804	(I)YEIADGINLETLFPQ(D)	0.7	-3	46.7	574.91 (3)	1722.86
A0A0C5ARQ8	DNA-directed RNA polymerase subunit beta	706-720	(I)TNEIPHLEAHLLRNL(D)	6.1	-0.8	46.7	590.83 (3)	1769.97
		122-134	(I)VINQILQSPGIYY(R)	3.55	0	46.2	754.88 (2)	1507.82
		246-263	(K)ELQKKFFQQRCELGRIGR(R)	10.7	2.9	27.8	745.89 (3)	2236.21
A7IZZ1	(-)-limonene synthase, chloroplastic	204-221	(K)ANISNDIMGALGLYEASF(H)	0.7	-2	50	629.08 (3)	1885.90
		197-213	(N)HKTGEFKANISNDIMGA(L)	7.56	0.1	35.3	611.27 (3)	1832.90
A0A0C5AUE3	Maturase K	387-398	(N)VVGQPLSKSTWA(D)	10.1	1	50	637.32 (2)	1272.70
		288-295	(S)ILASKDMP(L)	6.6	0	62.5	438.07 (2)	874.47
Q8RVK9	Naringenin-chalcone synthase	312-331	(D)QVESKLALKTEKLRATRHVL(S)	10.9	3.1	40	774.45 (3)	2320.38
		218-238	(D)GSAALIVGSDPIPEVEKPIFE(L)	3.5	-3	57.1	723.36 (3)	2168.15
E5DL82	NADH-ubiquinone oxidoreductase chain 5	122-131	(I)SCNMRLNAIT(L)	8.6	0.9	40	590.31 (2)	1179.56
		82-100	(P)VNRVGDFGLAPGISGRFTL(F)	10.7	1	47.4	659.21 (3)	1976.07
		132-145	(T)LICILLFIGAVGKS(A)	8.8	0.9	64.3	724.25 (2)	1446.88
		116-132	(A)PRNSWISCNMRLNAITL(I)	11.1	1.9	47.1	663.83 (3)	1989.02
V5KXG5	4-coumarate:CoA ligase	335-360	(P)NVTLGQGYGMTEAGPVLTMSLAFAKE(A)	4.2	-1	46.2	895.43 (3)	2685.33
		7-22	(E)ITKQEQQQDIIFRSKL(P)	9.93	1	31.3	659.18 (3)	1975.10
		291-306	(V)PPIVLAIAKYPDLHKY(D)	9.8	1.1	62.5	613.30 (3)	1838.06
		154-162	(L)VKFVCVDSP(V)	5.9	-0.1	55.6	497.68 (2)	993.51
A0A0C5ARZ4	Photosystem I P700 chlorophyll a apoprotein A1	695-722	(E)LIESIVWAHNKLKVAPATQPRALSIVQG(R)	10.7	2.1	57.1	1013.77 (3)	3039.75
		304-326	(L)VAGHMYRTNWGIGHGIKDILEAH(K)	8.0	0.3	39.1	859.42 (3)	2575.30
		44-61	(T)TWIWNLHADAHDFDSHTS(D)	4.9	-2.7	38.9	718.23 (3)	2152.95
		697-721	(I)ESIVWAHNKLKVAPATQPRALSIVQ(G)	10.7	2.1	56	919.31 (3)	2756.56
		92-117	(H)GARFSNYEAWLSDPTHIGPSAQVVWP(I)	5.17	-0.9	50	962.73 (3)	2885.40
		695-719	(E)LIESIVWAHNKLKVAPATQPRALSI(V)	10.7	2.1	60	919.28 (3)	2755.60
A0A0C5APY0	Photosystem I P700 chlorophyll a apoprotein A2	170-197	(K)NAESRLNHHLSGLFGVSSLAWTGHLVHV(A)	7.7	0.4	42.9	1013.22 (3)	3038.57

		361-369	(F)IAQDFTTQA(A)	0.7	-1	44.4	497.68 (2)	994.48
A0A0C5APZ4	Protein Ycf2	664-678	(S)KKKYLLVLPPIFYEE(N)	9.3	1	53.3	627.21 (3)	1880.09
		1304-1322	(N)LISEISSKCLHNLLLSKEM(I)	7.1	0	42.1	739.27 (3)	2215.18
		1780-1798	(R)CSTRNILVIASTHIPQKVD(P)	9.1	1	42.1	718.35 (3)	2152.15
A0A0C5B2I6	Ribulose bisphosphate carboxylase large chain	84-95	(R)CYHIEPVAGEEN(Q)	3.8	-3	33.3	680.73 (2)	1360.58
5DLX2	Ribosomal protein S3	308-331	(E)QLLGQLRKKCWKLMGKDKVMELIE(K)	10.1	2.9	41.7	963.27 (3)	2887.61
		79-93	(E)SIRLDDREKQNEIRI(W)	6.8	0	26.7	629.40 (3)	1885.03

a) According to “UniProtKB” (http://www.uniprot.org/)
b) According to “Protein Peptide Calculator” (http://pepcalc.com/)
c) According to “Peptide2.0” (http://peptide2.com/

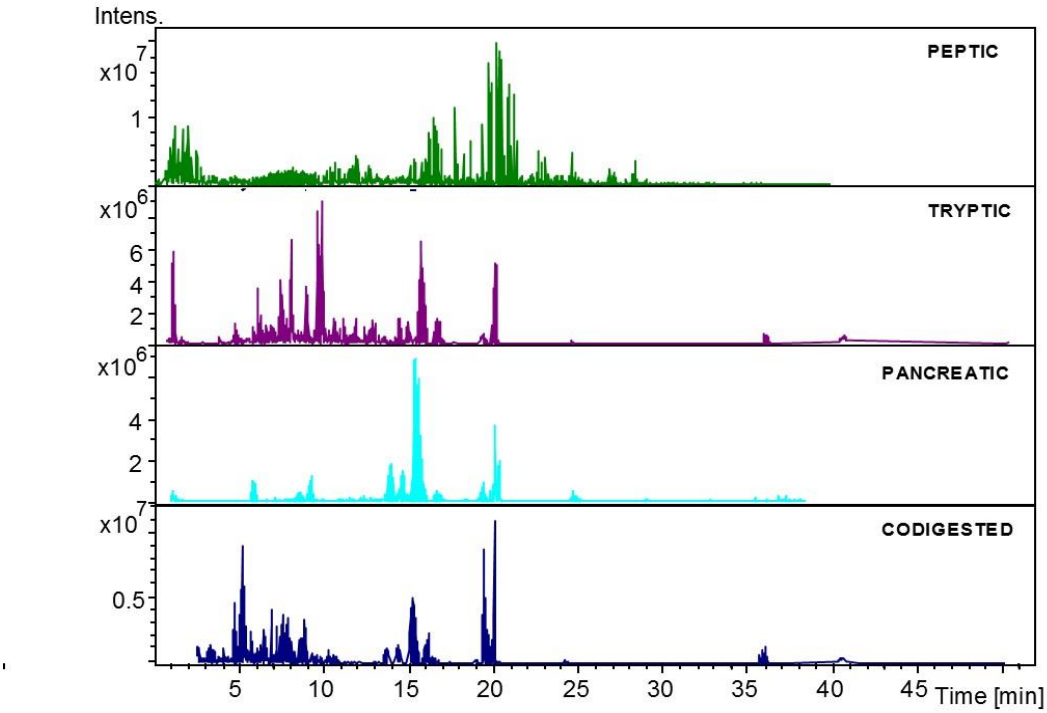


Figure 7.1S - Chromatograms of peptic, tryptic, pancreatic and codigested hydrolysates.

References

1. Martínez-Maqueda, D.; Miralles, B.; Recio, I.; Hernández-Ledesma, B., Antihypertensive peptides from food proteins: a review. *J. Food Funct.* **2012,** *3* (4), 350-61.

2. Foltz, M.; Meynen, E. E.; Bianco, V.; van Platerink, C.; Koning, T. M.; Kloek, J., Angiotensin converting enzyme inhibitory peptides from a lactotripeptide-enriched milk beverage are absorbed intact into the circulation. *J. Nutr.* **2007,** *137* (4), 953-8.

3. Pripp, A. H.; Isaksson, T.; Stepaniak, L.; Sorhaug, T.; Ardo, Y., Quantitative structure activity relationship modelling of peptides and proteins as a tool in food science. *Trends Food Sci Technol.* **2005,** *16* (11), 484-494.

4. Holton, T. A.; Dillon, E. T.; Robinson, A.; Wynne, K.; Cagney, G.; Shields, D. C., Optimal computational comparison of mass spectrometric peptide profiles of alternative hydrolysates from the same starting material. *Lwt-Food Sci Technol* **2016,** *73*, 296-302.

5. Michalski, A.; Damoc, E.; Hauschild, J. P.; Lange, O.; Wieghaus, A.; Makarov, A.; Nagaraj, N.; Cox, J.; Mann, M.; Horning, S., Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol Cell Proteomics* **2011,** *10* (9), M111.011015.

6. Ibáñez, C.; Simó, C.; García-Cañas, V.; Cifuentes, A.; Castro-Puyana, M., Metabolomics, peptidomics and proteomics applications of capillary electrophoresis-mass spectrometry in Foodomics: a review. *Anal Chim Acta* **2013,** *802*, 1-13.

7. Lahrichi, S. L.; Affolter, M.; Zolezzi, I. S.; Panchaud, A., Food peptidomics: large scale analysis of small bioactive peptides--a pilot study. *J. Proteomics* **2013,** *88*, 83-91.

8. Panchaud, A.; Affolter, M.; Kussmann, M., Mass spectrometry for nutritional peptidomics: How to analyze food bioactives and their health effects. *J. Proteomics* **2012,** *75* (12), 3546-59.

9. Aiello, G.; Fasoli, E.; Boschin, G.; Lammi, C.; Zanoni, C.; Citterio, A.; Arnoldi, A., Proteomic characterization of hempseed (Cannabis sativa L.). *J. Proteomics* **2016,** *147*, 187-196.

10. Malomo, S. A.; He, R.; Aluko, R. E., Structural and functional properties of hemp seed protein products. *J. Food Sci* **2014,** *79* (8), C1512-21.

11. Wang, X.-S.; Tang, C.-H.; Yang, X.-Q.; Gao, W.-R., Characterization, amino acid composition and in vitro digestibility of hemp (Cannabis sativa L.) proteins. *Food Chem.* **2008,** *107* (1), 11-18.

12. Girgih, A. T.; Alashi, A. M.; He, R.; Malomo, S. A.; Raj, P.; Netticadan, T.; Aluko, R. E., A novel hemp seed meal protein hydrolysate reduces oxidative stress factors in spontaneously hypertensive rats. *Nutrients* **2014,** *6* (12), 5652-5666.

13. Girgih, A. T.; He, R.; Aluko, R. E., Kinetics and molecular docking studies of the inhibitions of angiotensin converting enzyme and renin activities by hemp seed (*Cannabis sativa* L.) peptides. *J Agric. Food Chem.* **2014,** *62* (18), 4135-4144.

14. Girgih, A. T.; Udenigwe, C. C.; Aluko, R. E., In vitro antioxidant properties of hemp seed (*Cannabis sativa* L.) protein hydrolysate fractions. *J. Am. Oil. Chem. Soc.* **2011,** *88* (3), 381-389.

15. Girgih, A. T.; Udenigwe, C. C.; Aluko, R. E., Reverse-phase HPLC separation of hemp seed (*Cannabis sativa* L.) protein hydrolysate produced peptide fractions with enhanced antioxidant capacity. *Plant Foods Hum. Nutr.* **2013,** *68* (1), 39-46.

16. Gavel, N. T.; Edel, A. L.; Bassett, C. M. C.; Weber, A. M.; Merchant, M.; Rodriguez-Leyva, D.; Pierce, G. N., The effect of dietary hempseed on atherogenesis and contractile function in aortae from hypercholesterolemic rabbits. *Acta Physiol. Hung.* **2011,** *98* (3), 273-283.

17. Prociuk, M. A.; Edel, A. L.; Gavel, N. T.; Lukas, A.; Pierce, G. N., Influence of dietary hempseed on arrhythmia generation by ischemia/reperfusion. *J. Mol. Cell Cardiol.* **2004,** *37* (1), 275-275.

18. Prociuk, M. A.; Edel, A. L.; Richard, M. N.; Gavel, N. T.; Ander, B. P.; Dupasquier, C. M. C.; Pierce, G. N., Cholesterol-induced stimulation of platelet aggregation is prevented by a hempseed-enriched diet. *Can J Physiol Pharm* **2008,** *86* (4), 153-159.

19. Zanoni, C.; Aiello, G.; Arnoldi, A.; Lammi, C., Hempseed Peptides Exert Hypocholesterolemic Effects with a Statin-Like Mechanism. *J Agric Food Chem* **2017**, 10.1021/acs.jafc.7b02742

20. Ambigaipalan, P.; Al-Khalifa, A. S.; Shahidi, F., Antioxidant and angiotensin I converting enzyme (ACE) inhibitory activities of date seed protein hydrolysates prepared using Alcalase, Flavourzyme and Thermolysin. *J Funct Foods* **2015,** *18*, 1125-1137.

21. Boschin, G.; Scigliuolo, G. M.; Resta, D.; Arnoldi, A., ACE-inhibitory activity of enzymatic protein hydrolysates from lupin and other legumes. *Food Chem* **2014,** *145*, 34-40.

22. Levashov, P. A.; Sutherland, D. S.; Besenbacher, F.; Shipovskov, S., A robust method of determination of high concentrations of peptides and proteins. *Anal Biochem* **2009,** *395* (1), 111-2.

23. Lammi, C.; Aiello, G.; Vistoli, G.; Zanoni, C.; Arnoldi, A.; Sambuy, Y.; Ferruzza, S.; Ranaldi, G., A multidisciplinary investigation on the bioavailability and activity of peptides from lupin protein. *J Funct Foods* **2016,** *24*, 297-306.

24. Nielsen, P. M.; Petersen, D.; Dambmann, C., Improved method for determining food protein degree of hydrolysis. *J Food Sci* **2001,** *66* (5), 642-646.

25. Schägger, H., Tricine-SDS-PAGE. *Nat Protoc* **2006,** *1* (1), 16-22.

26. Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J. C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F., Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **1999,** *112*, 531-52.

27. Mooney, C.; Haslam, N. J.; Pollastri, G.; Shields, D. C., Towards the improved discovery and design of functional peptides: Common features of diverse classes permit generalized prediction of bioactivity. *Plos One* **2012,** *7* (10).

28. Malomo, S. A.; Onuh, J. O.; Girgih, A. T.; Aluko, R. E., Structural and antihypertensive properties of enzymatic hemp seed protein hydrolysates. *Nutrients* **2015,** *7* (9), 7616-32.

29. Bauchart, C.; Morzel, M.; Chambon, C.; Mirand, P. P.; Reynès, C.; Buffière, C.; Rémond, D., Peptides reproducibly released by in vivo digestion of beef meat and trout flesh in pigs. *Br J Nutr* **2007,** *98* (6), 1187-95.

30. Sarver, R. W.; Bills, E.; Bolton, G.; Bratton, L. D.; Caspers, N. L.; Dunbar, J. B.; Harris, M. S.; Hutchings, R. H.; Kennedy, R. M.; Larsen, S. D.; Pavlovsky, A.; Pfefferkorn, J. A.; Bainbridge, G., Thermodynamic and structure guided design of statin based inhibitors of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *J Med Chem* **2008,** *51* (13), 3804-13.

31. Zhong, F.; Zhang, X. M.; Ma, J. G.; Shoemaker, C. F., Fractionation and identification of a novel hypocholesterolemic peptide derived from soy protein Alcalase hydrolysates. *Food Res Int* **2007,** *40* (6), 756-762.

32. Lammi, C.; Zanoni, C.; Arnoldi, A.; Vistoli, G., Two Peptides from Soy β-Conglycinin Induce a Hypocholesterolemic Effect in HepG2 Cells by a Statin-Like Mechanism: Comparative in Vitro and in Silico Modeling Studies. *J Agric Food Chem* **2015,** *63* (36), 7945-51.

33. Sarmadi, B. H.; Ismail, A., Antioxidative peptides from food proteins: a review. *Peptides* **2010,** *31* (10), 1949-56.

34. Li, H.; Aluko, R. E., Identification and inhibitory properties of multifunctional peptides from pea protein hydrolysate. *J Agric Food Chem* **2010,** *58* (21), 11471-6.

35. Kobayashi, Y.; Yamauchi, T.; Katsuda, T.; Yamaji, H.; Katoh, S., Angiotensin-I converting enzyme (ACE) inhibitory mechanism of tripeptides containing aromatic residues. *J Biosci Bioeng* **2008,** *106* (3), 310-2.

36. Huang, G. J.; Lu, T. L.; Chiu, C. S.; Chen, H. J.; Wu, C. H.; Lin, Y. C.; Hsieh, W. T.; Liao, J. C.; Sheu, M. J.; Lin, Y. H., Sweet potato storage root defensin and its tryptic hydrolysates exhibited angiotensin converting enzyme inhibitory activity in vitro. *Bot Stud* **2011,** *52* (3), 257-264.

37. O'Keeffe, M. B.; Norris, R.; Alashi, M. A.; Aluko, R. E.; FitzGerald, R. J., Peptide identification in a porcine gelatin prolyl endoproteinase hydrolysate with angiotensin converting enzyme (ACE) inhibitory and hypotensive activity. *J Funct Foods* **2017,** *34*, 77-88.

38. Bailey, C. J.; Tahrani, A. A.; Barnett, A. H., Future glucose-lowering drugs for type 2 diabetes. *Lancet Diabetes Endocrinol* **2016**.

39. Boots, J.-W. P. Protein hydrolysate enriched in peptides inhibiting DPP-IV and their use. WO2006068480A2, **2006**.

40. Lammi, C.; Zanoni, C.; Arnoldi, A.; Vistoli, G., Peptides derived from soy and lupin protein as Dipeptidyl-Peptidase IV inhibitors: *In vitro* biochemical screening and *in silico* molecular modeling study. *J Agric Food Chem* **2016,** *64* (51), 9601-9606.

41. Nongonierma, A. B.; FitzGerald, R. J., Investigation of the Potential of Hemp, Pea, Rice and Soy Protein Hydrolysates as a Source of Dipeptidyl Peptidase IV (DPP-IV) Inhibitory Peptides. *Food Dig Res Curr Opin* **2015,** *6* (1-3), 19-29.

42. Amigo-Benavent, M.; Clemente, A.; Caira, S.; Stiuso, P.; Ferranti, P.; del Castillo, M. D., Use of phytochemomics to evaluate the bioavailability and bioactivity of antioxidant peptides of soybean β-conglycinin. *Electrophoresis* **2014,** *35* (11), 1582-9.

43. Regazzo, D.; Mollé, D.; Gabai, G.; Tomé, D.; Dupont, D.; Leonil, J.; Boutrou, R., The (193-209) 17-residues peptide of bovine β-casein is transported through Caco-2 monolayer. *Mol Nutr Food Res* **2010,** *54* (10), 1428-35.

44. Miguel, M.; Dávalos, A.; Manso, M. A.; de la Peña, G.; Lasunción, M. A.; López-Fandiño, R., Transepithelial transport across Caco-2 cell monolayers of antihypertensive egg-derived peptides. PepT1-mediated flux of Tyr-Pro-Ile. *Mol Nutr Food Res* **2008,** *52* (12), 1507-13.

PART-II

PEPTIC HEMPSEED HYDROLYSATE EXERTS HYPOCHOLESTEROLEMIC EFFETCTS WITH A STATIN-LIKE MECHANISM

Chiara Zanoni, Gilda Aiello, Anna Arnoldi,* Carmen Lammi

Department of Pharmaceutical Sciences, University of Milan, Milan, Italy

Keywords. Bioactive peptides; *Cannabis sativa*; functional foods; hempseed; plant proteins.

8.0 Abstract

Peptic hempseed protein hydrolysate exerts hypocholesterolemic effects with a statin-like mechanism on HepG2 cells. In the range 0.1-1.0 mg/mL, it inhibits the catalytic activity of 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCoAR) in a dose-dependent manner. Immunoblotting detection showed increments in the protein levels of regulatory element binding proteins 2 (SREBP-2), low-density lipoprotein receptor (LDLR), and HMGCoAR. However, the parallel activation of the phospho-5'-adenosine monophosphate-activated protein kinase (AMPK) pathway, produced an inactivation of HMGCoAR by phosphorylation. The functional ability of HepG2 cells to uptake extracellular LDL was raised by 50.5 ± 2.7%, 221.5 ± 1.6%, and 109 ± 3.5%, respectively, versus the control at 0.25, 0.5, and 1.0 mg/mL concentrations.

8.1 Introduction

Literature indicates that peptides, obtained through hydrolysis of hempseed protein with different enzymes, may function as hypotensive agents [1, 2] and antioxidant.[1, 3, 4] On the contrary, literature reports only a few evidences on the ability of hempseed to modulate the lipid profile.[5-7] The inhibition of cholesterol biosynthesis is the most efficient way to reduce serum

cholesterol levels. Since intracellular cholesterol production is a multistep pathway in which 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGCoAR) mediates the rate-limiting step, this enzyme is an important drug target.

Starting from the hypothesis that most activities observed should depend on specific peptides encrypted in the protein sequences, we decided to investigate the hypocholesterolemic properties of peptic hydrolysates produced as reported in the Part-I of this manuscript, using human hepatic HepG2 cells as model system. More specifically, the present work had the goal to elucidate of the mechanism through which this hydrolysate mediates a cholesterol-lowering effect at HepG2 cells, by molecular and functional investigations on the LDLR-SREBP-2 pathway.

8.2 Materials and Methods

Cell culture conditions. The HepG2 cell line was bought from ATCC (HB-8065, ATCC from LGC Standards, Milan, Italy). The HepG2 cell line was cultured in DMEM high glucose with stable L-glutamine supplemented with 10% FBS, 100 U/mL penicillin, 100 µg/mL streptomycin (complete growth medium) and incubated at 37 °C under 5% CO_2 atmosphere. HepG2 cells were used for no more than 20 passages after thawing, because the increase of the number of passages may change the cell characteristics and impair assay results.

Western blot analysis. A total of 1.5 x 10^5 HepG2 cells/well (24-well plate) were treated with 0.25, 0.5, and 1.0 mg/mL of peptic peptides for 24 h. After each treatment, cells were scraped in 40 µL ice-cold lysis buffer [RIPA buffer + inhibitor cocktail + 1:100 PMSF + 1:100 Na-orthovanadate] and transferred in an ice-cold microcentrifuge tube. After centrifugation at 16,060 *g* for 15 min at 4 °C, the supernatant was recovered and transferred into a new ice-cold tube. Total proteins were quantified by Bradford method and 50 µg of total proteins loaded on a pre-cast 7.5% Sodium Dodecyl Sulfate - Polyacrylamide (SDS-PAGE) gel at 130 V for 45 min. Subsequently, the gel was pre-equilibrated with 0.04% SDS in H_2O for 15 min at room temperature (RT) and transferred to a nitrocellulose membrane (Mini nitrocellulose Transfer Packs) using a Trans-blot Turbo at 1.3 A, 25 V for 7 min. Target proteins, on milk blocked membrane, were detected by primary antibodies as follows: rabbit anti-SREBP-2, rabbit anti-LDLR, anti-HMGCoAR, anti-phospho-AMPK (Thr172), anti-phospho-HMGCoAR (Ser872), and anti-β-actin. Secondary antibodies conjugated with HRP and a chemiluminescent reagent

were used to visualize target proteins and their signal was quantified using the Image Lab Software (Bio-Rad). The internal control β-actin was used to normalize loading variations.

Fluorescent LDL uptake cell based assay. A total of 3 x 10^4 HepG2 cells/well were seeded in 96-well plates and kept in complete growth medium for 2 d before treatment. On the third day, cells were treated with 0.25 mg/mL peptic peptides or vehicle (100 mM Tris) for 24 h. At the end of the treatment periods, the culture medium was replaced with 50 µL/well LDL-DyLight™ 550 working solution. The cells were additionally incubated for 2 h at 37 °C, then the culture medium was aspirated and replaced with PBS 100 µL/well. The degree of LDL uptake was measured using the Synergy H1 fluorescent plate reader from Biotek (excitation and emission wavelengths 540 and 570 nm, respectively).

Statistically Analysis. Statistical analyses were carried out by One-way ANOVA (Graphpad Prism 6) followed by Dunnett's test. Values were expressed as means ± sem; P-values < 0.05 were considered to be significant.

8.3 Results

Preparation and analysis of the peptic hydrolysate from hempseed protein. The hydrolysate preparation is reported in the Part-I **(§ 7.2.2)** of the same manuscript. Regarding the peptic hydrolysate composition, the highest number of detected peptides belongs to the main storage proteins in hempseed: 6 peptides were shared by all isoforms of Edestin 2 (ede2A; 2B; 2C), 6 peptides were shared by the isoforms of Edestin 1 (ede1A,B; 1B) and only 1 by the isoforms of Edestin 1 (ede1A,B; 1D). Numerous peptides belonged to the most heterogeneous protein family identified here, i.e. the acyl-activating enzyme superfamily: 5 unique peptides belonged to DNA-directed RNA polymerase subunit beta, Photosystem I P700 chlorophyll, apoprotein A2, and Protein Ycf2, whereas 4 peptides to (+)-alpha-pinene synthase. This protein is involved in the terpene metabolism and naringenin-chalcone synthase and is better known as flavanone synthase. In total 3 peptides belonged to Putative LysM domain containing receptor kinase and THCA synthase, whereas 2 unique peptides belonged to ATP synthase (alpha and beta subunit), cannabidiolic acid synthase-like 1, Delta 12 desaturase, Hypothetical chloroplast RF1, (-)-limonene synthase, MatR, NADH-plastoquinone oxidoreductase (4 and 5 subunit), 4-coumarate:CoA ligase, and Polyketide synthase family, respectively. (**Table 7.1S Part-I).**

Effects on the LDLR pathway modulation. HepG2 cells were treated with 0.25, 0.5, and 1.0 mg/mL of the hempseed peptic hydrolysate and each sample was investigated with immunoblotting experiments. The treatment induced an up-regulation of the protein level of the N-terminal fragment of SREBP-2 (mature form with a molecular weight of 68 kDa) by 34.8 ± 19% ($p < 0.001$), 29.1 ± 7.7% ($p < 0.001$), and 33.5 ± 6% ($p < 0.001$) at 0.25, 0.5, and 1 mg/mL, respectively, *versus* the control (**Figure 8.1 A-C**). As a consequence, up-regulations of the LDLR and HMGCoAR protein levels were also observed. In particular, as shown in **Figure 8.1 A-C**, after the treatment with the peptic hydrolysate at 0.25, 0.5, and 1.0 mg/mL, the LDLR protein level was increased by 35.6 ± 10.4% ($p <0.01$), 54.7 ± 41.4% ($p< 0.0001$), and 63.0 ± 41.3% ($p< 0.0001$), respectively, *versus* the control, whereas the HMGCoAR protein level was augmented by 32.2 ± 15.6% ($p <0.05$), 54.1 ± 10.4% ($p <0.01$), and 67.7 ± 38.9% ($p< 0.0001$), respectively, *versus* the control (**Figure 8.1 B-C**).

Effects on AMPK pathway activation. Suitable immunoblotting experiments were performed in order to evaluate the effect of the treatment with hempseed peptic hydrolysate on AMPK activation and HMGCoAR inactivation (AMPK substrate). The lysates from treated and untreated HepG2 cells were therefore analyzed using specific antibodies for AMPK phosphorylated at threonine 172 (**Figure 8.1 D**) and for HMGCoAR phosphorylated at serine 872 (AMPK phosphorylation site) (**Figure 8.1 E**). **Figure 8.1 F** shows that treatment with hempseed peptic hydrolysate significantly increased AMPK phosphorylation by 95.2 ± 38.7% (0.50 mg/mL, $p <0.0001$) and by 120.3 ± 18.4% (1 mg/mL, $p <0.0001$) *versus* the control. As a consequence of the AMPK activation, the phosphorylation levels of HMGCoAR were also increased by 67.0 ± 15.8% at 0.5 mg/mL ($p <0.0001$) and 56.0 ± 37.3% at 1 mg/mL ($p<0.0001$), *versus* the control (**Figure 8.1 F**).

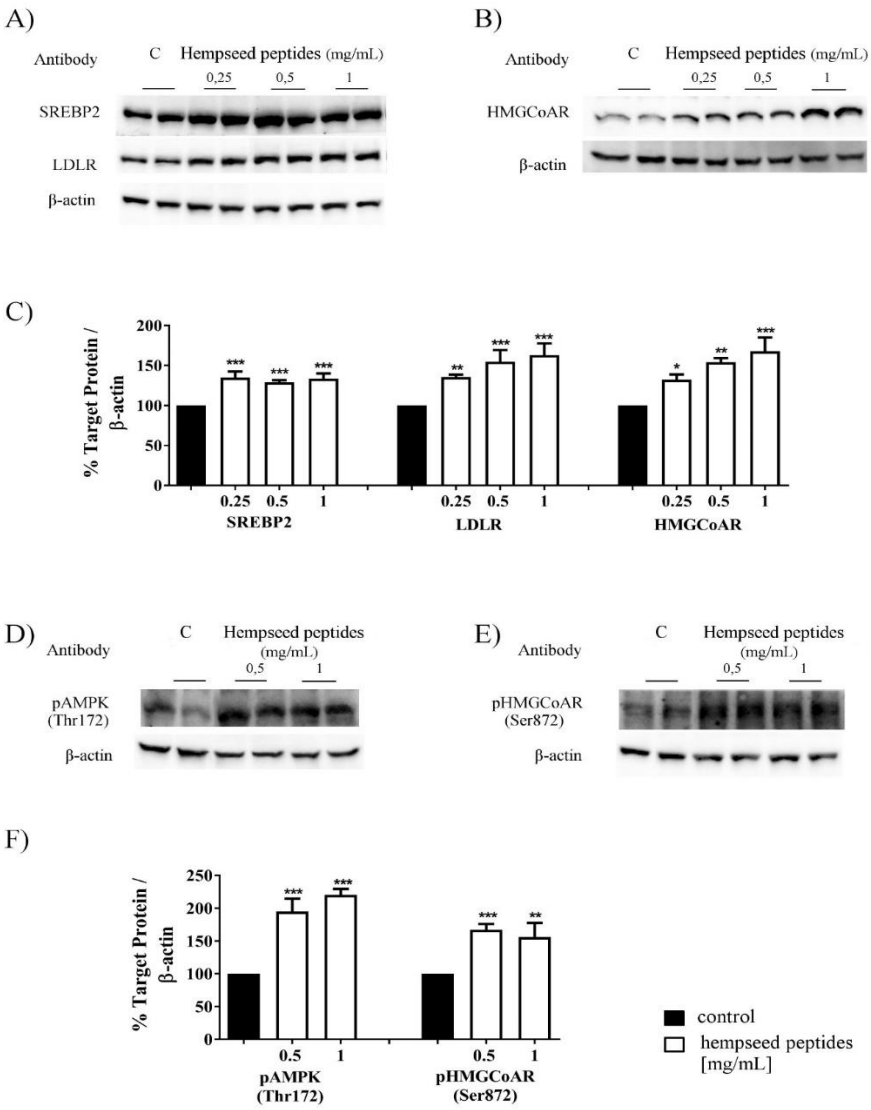


Figure 8.1 - Effects of hempseed peptides on SREBP-2-LDLR pathway. HepG2 cells (1.5×10^5) were treated with peptic hempseed peptides (0.25, 0.5, 1.0 mg/mL) for 24 h. SREBP-2, LDLR, HMGCoAR, phospho-AMPK (Thr172), phospho-HMGCoAR (Ser872), and β-actin immunoblotting signals were detected using specific anti-SREBP-2, anti-LDLR, anti-HMGCoAR, anti-phospho-AMPK (Thr172), anti-phospho-HMGCoAR (Ser872), and anti-β-actin primary antibodies, respectively (**A-B-D-E**). Each protein signal was quantified by ImageLab software (Bio-Rad) and normalized with β-actin signals (**C-F**). Bars represent averages ± SEM of six independent experiments (two duplicates per sample). *(∗) $p < 0.05$*, *(∗∗) $p < 0.001$*, and *(∗∗∗) $p < 0.0001$ versus* control (C).

Modulation of the LDL-uptake in HepG2 cells. The functional ability of HepG2 cells to uptake extracellular LDL after the treatments with the same peptic hydrolysate was investigated by performing fluorescent LDL uptake experiments. As shown in **Figure 8.2**, hempseed peptides increased the LDL-uptake in a statistically significant way *versus* the control. In fact, after treatments with 0.25, 0.5, and 1.0 mg/mL, the LDL-uptake was increased by 50.5± 2.7% ($p < 0.001$), 221.5 ± 1.6% ($p < 0.001$), and 109 ± 3.5% ($p < 0.001$), respectively.

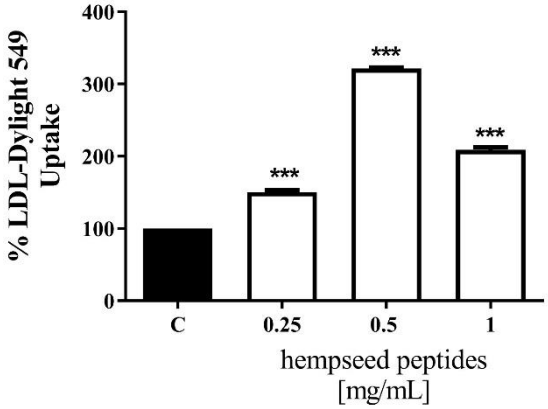


Figure 8.2 - Fluorescent LDL-uptake assay after treatment of HepG2 with hempseed peptides. Cells (3×10^4) were treated with hempseed peptides (0.25, 0.5, 1.0 mg/mL) for 24 h. LDL-Dylight 550 (10 μg/mL) was incubated for an additional 2 h. Excess LDL-Dylight 550 was removed and cells were washed two times with PBS. Specific fluorescent LDL-uptake signal was analyzed by Synergy H1 (Biotek). Data points represent averages ± SEM of three independent experiments in triplicate. *(***) $p < 0.0001$ versus* control (C).

8.4 Discussion

Molecular and cellular investigation of the hypocholesterolemic properties of the peptic hempseed hydrolysate. This paper provides evidence of the hypocholesterolemic effects mediated by the peptic hempseed protein hydrolysate suggesting also a mechanistic explanation. HMGCoAR is the rate-controlling enzyme of cellular cholesterol biosynthesis pathway and therefore it constitutes the target of numerous investigations aimed at lowering the rate of cholesterol biosynthesis.[8-10]

The LDLR expression and the receptor protein localization at cellular membranes are strictly correlated to the intracellular cholesterol biosynthesis pathway. In facts, the transcription of the LDLR and the genes required for cholesterol and fatty acid synthesis are controlled by membrane-bound transcription factors called SREBPs,[11] and intracellular cholesterol acts with a negative feedback inhibition mechanism.[12] The SREBP-2 isoform is responsible for the LDLR and HMGCoAR transcription and the SREBP-2 maturation is regulated by the intracellular cholesterol homeostasis. Thus, the up-regulation of LDLR represents a useful strategy to control plasma LDL cholesterol levels. Our findings demonstrate that hempseed peptides with molecular weights lower than 3000 Da are able to up-regulate the LDLR protein levels through an increase of SREBP-2 protein.

In addition, a detailed investigation of the LDLR pathway revealed that the peptic hydrolysate increases the HMGCoAR protein levels in a significant way *versus* the control (**Figure 8.1 A-C)**. However, this does not mean an increase of cholesterol synthesis, since the peptic hydrolysate is also able to inactivate HMGCoAR, by increasing its phosphorylation mediated by the activation of the AMPK pathway (**Figure 8.1 D-F**). One mechanism of HMGCoAR activity regulation is achieved through phosphorylation of Ser872 by AMPK, which decreases the enzyme.[13] In fact, physiologically HMGCoAR is present in the cell in a non-phosphorylated active form (30%) and a phosphorylated inactive form (70%).[14] By increasing the phosphorylation of AMPK, some natural compounds, such as soy[15] and lupin peptides,[16] are able to produce a direct inhibition of HMGCoAR.[17] Furthermore, also statins are able to activate AMPK,[18] with the consequence of a synergistic inhibition of the HMGCoAR activity. In line with these observations, this work provides evidence according to which a peptic hydrolysate from hempseed protein is not only able to directly inhibit the reductase activity, but also to modulate the HMGCoAR regulation by the activation of the AMPK pathway.

Finally, in agreement with immunoblotting results, the increase of LDLR protein levels leads to an increase of LDL uptake (**Figure 8.2**). The induction of the LDL clearance is strictly correlated to an increase of LDLR protein level. The cholesterol-lowering effects of hempseed peptides in human hepatic HepG2 cells have some similarities with the behaviour of lupin peptides[15]. In facts, also the peptides obtained by the hydrolysis of a total lupin protein extract with pepsin, are able to mediate hypocholesterolaemic effects in the same cells through the activation of the LDLR pathway. Other hypocholesterolemic peptides with a statin-like mechanism may be found in soy [19].

Author Information: Experiment ideation and design: CL and GA. Experiments & data analysis: biological experiments CL & CZ; peptide preparation and identification GA. Figure preparation: CZ and GA. Grant retrieval: AA. Manuscript writing: CL, GA & AA.

References

1. Girgih, A. T.; Alashi, A. M.; He, R.; Malomo, S. A.; Raj, P.; Netticadan, T.; Aluko, R. E., A novel hemp seed meal protein hydrolysate reduces oxidative stress factors in spontaneously hypertensive rats. *Nutrients* **2014,** *6* (12), 5652-5666.

2. Girgih, A. T.; He, R.; Aluko, R. E., Kinetics and molecular docking studies of the inhibitions of angiotensin converting enzyme and renin activities by hemp seed (*Cannabis sativa* L.) peptides. *J Agric Food Chem* **2014,** *62* (18), 4135-4144.

3. Girgih, A. T.; Udenigwe, C. C.; Aluko, R. E., In vitro antioxidant properties of hemp seed (*Cannabis sativa* L.) protein hydrolysate fractions. *J Am Oil Chem Soc* **2011,** *88* (3), 381-389.

4. Girgih, A. T.; Udenigwe, C. C.; Aluko, R. E., Reverse-phase HPLC separation of hemp seed (*Cannabis sativa* L.) protein hydrolysate produced peptide fractions with enhanced antioxidant capacity. *Plant Foods Hum Nutr* **2013,** *68* (1), 39-46.

5. Gavel, N. T.; Edel, A. L.; Bassett, C. M. C.; Weber, A. M.; Merchant, M.; Rodriguez-Leyva, D.; Pierce, G. N., The effect of dietary hempseed on atherogenesis and contractile function in aortae from hypercholesterolemic rabbits. *Acta Physiol Hungarica* **2011,** *98* (3), 273-283.

6. Prociuk, M. A.; Edel, A. L.; Gavel, N. T.; Lukas, A.; Pierce, G. N., Influence of dietary hempseed on arrhythmia generation by ischemia/reperfusion. *J of Mol Cell Cardiol* **2004,** *37* (1), 275-275.

7. Prociuk, M. A.; Edel, A. L.; Richard, M. N.; Gavel, N. T.; Ander, B. P.; Dupasquier, C. M. C.; Pierce, G. N., Cholesterol-induced stimulation of platelet aggregation is prevented by a hempseed-enriched diet. *Can J Physiol Pharmacoly* **2008,** *86* (4), 153-159.

8. Endo, A., The discovery and development of HMG-CoA reductase inhibitors. *J Lipid Res* **1992,** *33* (11), 1569-82.

9. Gencer, B.; Lambert, G.; Mach, F., PCSK9 inhibitors. *Swiss Med Wkly* **2015,** *145*, w14094.

10. Zhang, D. W.; Lagace, T. A.; Garuti, R.; Zhao, Z.; McDonald, M.; Horton, J. D.; Cohen, J. C.; Hobbs, H. H., Binding of proprotein convertase subtilisin/kexin type 9 to epidermal growth factor-like repeat A of low density lipoprotein receptor decreases receptor recycling and increases degradation. *J Biol Chem* **2007,** *282* (25), 18602-12.

11. Horton, J. D.; Goldstein, J. L.; Brown, M. S., SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J Clin Invest* **2002,** *109* (9), 1125-1131.

12. Edwards, P. A.; Tabor, D.; Kast, H. R.; Venkateswaran, A., Regulation of gene expression by SREBP and SCAP. *Biochim Biophys Acta* **2000,** *1529* (1-3), 103-13.

13. Ching, Y. P.; Davies, S. P.; Hardie, D. G., Analysis of the specificity of the AMP-activated protein kinase by site-directed mutagenesis of bacterially expressed 3-hydroxy 3-methylglutaryl-CoA reductase, using a single primer variant of the unique-site-elimination method. *Eur J Biochem* **1996,** *237* (3), 800-8.

14. Pallottini, V.; Martini, C.; Pascolini, A.; Cavallini, G.; Gori, Z.; Bergamini, E.; Incerpi, S.; Trentalance, A., 3-Hydroxy-3-methylglutaryl coenzyme A reductase deregulation and age-related hypercholesterolemia: a new role for ROS. *Mech Ageing Dev* **2005,** *126* (8), 845-51.

15. Lammi, C.; Zanoni, C.; Arnoldi, A., Molecular characterization of the hypocholesterolemic mechanism of action of IAVPGEVA, IAVPTGVA, and LPYP, three peptides from soy glycinin, in HepG2 cells. *J Funct Foods* **2015,** *14*, 469-478.

16. Lammi, C.; Zanoni, C.; Scigliuolo, G. M.; D'Amato, A.; Arnoldi, A., Lupin peptides lower low-density lipoprotein (LDL) cholesterol through an up-regulation of the LDL receptor/sterol regulatory element binding protein 2 (SREBP-2) pathway at HepG2 cell line. *J Agric Food Chem* **2014,** *62* (29), 7151-9.

17. Oliaro-Bosso, S.; Calcio Gaudino, E.; Mantegna, S.; Giraudo, E.; Meda, C.; Viola, F.; Cravotto, G., Regulation of HMGCoA reductase activity by policosanol and octacosadienol, a new synthetic analogue of octacosanol. *Lipids* **2009,** *44* (10), 907-16.

18. Sun, W.; Lee, T. S.; Zhu, M.; Gu, C.; Wang, Y.; Zhu, Y.; Shyy, J. Y., Statins activate AMP-activated protein kinase in vitro and in vivo. *Circulation* **2006,** *114* (24), 2655-62.

19. Lammi, C.; Zanoni, C.; Arnoldi, A.; Vistoli, G., Two peptides from soy beta-conglycinin induce a hypocholesterolemic effect in HepG2 cells by a statin-Like mechanism: comparative in vitro and in silico modeling studies. *J Agric Food Chem* **2015,** *63* (36), 7945-7951.

CHAPTER 9

MANUSCRIPT 6

PROTEOMIC ANALYSIS OF SWEET ALGERIAN APRICOT KERNELS (*PRUNUS ARMENIACA* L.) BY COMBINATORIAL PEPTIDE LIGAND LIBRARIES AND LC-MS/MS

Hamida Ghorab[1ǂ], Carmen Lammi[2], Anna Arnoldi[2*], Zahia Kabouche[1], Gilda Aiello[2ǂ]

1 Université des Frères Mentouri-Constantine, Département de chimie, Laboratoire d'Obtention des Substances Thérapeutiques (LOST), Campus Chaabet-Ersas, 25000 Constantine, Algeria

2 Department of Pharmaceutical Sciences, University of Milan, via Mangiagalli 25, 20133 Milan, Italy

ǂ HG and GA gave equivalent contributions to this work.

Abbreviations used: CPLLs, Combinatorial peptide ligand libraries; **EEP,** eluted enriched proteins; **ACN,** acetonitrile; **FA,** formic acid; **PPI,** protein-protein interaction network; **PDMQ,** Protein Digestion Multi Query.

9.0 Abstract

An investigation on the proteome of the sweet kernel of apricot, based on equalisation with combinatorial peptide ligand libraries (CPLLs), SDS-PAGE, nLC-ESI-MS/MS, and database search, permitted identifying 175 proteins. Gene ontology analysis indicated that their main molecular functions are in nucleotide binding (20.9 %), hydrolase activities (10.6 %), kinase

activities (7 %), and catalytic activity (5.6%). A protein-protein association network analysis using STRING software permitted to build an interactomic map of all detected proteins, characterised by 34 interactions. In order to forecast the potential health benefits deriving from the consumption of these proteins, the two most abundant, i.e. Prunin 1 and 2, were enzymatically digested *in silico* predicting 10 and 14 peptides, respectively. Searching their sequences in the database BIOPEP, it was possible to suggest a variety of bioactivities, including dipeptidyl peptidase-IV (DPP-IV) and angiotensin converting enzyme I (ACE) inhibition, glucose uptake stimulation and antioxidant properties.

Keywords: apricot; combinatorial peptide ligand libraries (CPLL); GO analysis; LC-MS/MS; proteomics; prunin.

9.1 Introduction

Prunus armeniaca L. (synonym *Armeniaca vulgaris* Lam., subclass *Rosidae*, order *Rosales*, family *Rosaceae*, subfamily *Prunoideae*), widely known as "apricot", is an edible plant widely cultivated for its delicious fruits. The term *armeniaca* is supposed to derive from Armenia, where this plant has been grown since ancient times. Apricot is currently cultivated in many countries with suitable climates. Algeria and Italy are among major world producers, with annual yields equal to 269,308 and 247,146 tonnes, respectively (FAO, 2012).

The fruit is a drupe, with a thin downy skin, enclosing a yellowish-orange flesh (mesocarp) and an inner large, compressed, smooth, woody stone, containing the kernel. Although the fruit is certainly the most important product of this plant, also the kernel has some interest, being a rich source of dietary protein, oil, and fibre (Femenia, Rossello, Mulet, & Canellas, 1995).
Its use in human diet is limited, however, since it contains D(-)-mandelonitrile β-gentiobioside, a toxic bitter cyanogenic glycoside generally named amygdalin (Varsha, Akash, Jasmine, Raj, Jain, & Chaudhary, 2012; Yildirim & Askin, 2010). Its concentration is nonetheless variable and also some non-toxic cultivars derived from a sweet population from China are available (Lu, Lu, Gao, Lu, Lu, & Gao, 1994). Bitter kernels may contain up to 5.5% amygdalin (Yildirim & Askin, 2010), while sweet ones less than 1% (Karsavuran, Charehsaz, Celik, Asma, Yakinci, & Aydin, 2014). Whereas bitter kernels are exploited as raw material in cosmetic and pharmaceutical applications, sweet ones may be consumed as nuts (Yildirim &

Askin, 2010). They are currently eaten as roasted salted titbit and appetisers or ground to a flour used in different food formulations (Ozboy-Ozbas, Seker, & Gokbulut, 2010; Seker, Ozboy-Ozbas, Gokbulut, Ozturk, & Koksel, 2009; Seker, Ozboy-Ozbas, Gokbulut, Ozturk, & Koksel, 2010), often together with almond flour. Often, however, they are discarded by food processing industry, although their use would help maximising available resources and might result in generating innovative foods (Ozcan, 2000).

The oil of apricot kernels has been widely characterised (Turan, Topcu, Karabulut, Vural, & Hayaloglu, 2007; Yildirim, Yildirim, Askin, & Kankaya, 2010) and shown to possess hypocholesterolaemic and antioxidant properties (Kutlu, Durmaz, Ates, & Erdogan, 2009). Other components extensively characterised are sugars, fibre and phytochemicals (Turan, Topcu, Karabulut, Vural, & Hayaloglu, 2007; Yildirim, Yildirim, Askin, & Kankaya, 2010). Literature indicates that this kernel has antioxidant (Durmaz & Alpaslan, 2007) and antimicrobial properties (Lee, Ahn, Kwon, Lee, Kwak, & Min, 2014).

The fruit proteome has been investigated during ripening (D'Ambrosio, Arena, Rocco, Verrillo, Novi, Viscosi, et al., 2013), whereas an extensive characterisation of the kernel proteome is still lacking. Only one recent paper has developed a highly specific competitive enzyme-linked immunosorbent assay to detect the addition of apricot kernels in almond products, based on specific proteins differential analysis (Zhang, Wang, Huang, Lai, Du, Liu, et al., 2016).

The first objective of the present work was to get a comprehensive information on kernel proteome: the identification of minor proteins was improved by the use of spin columns packed with combinatorial peptide ligand libraries (CPLLs), a powerful non-depleting tool for discovering low-abundant proteins (Aiello, Fasoli, Boschin, Lammi, Zanoni, Citterio, et al., 2016; Esteve, D'Amato, Marina, García, Citterio, & Righetti, 2012; Righetti, Fasoli, & Boschetti, 2011). The apricot kernel proteome was then elucidated by mass spectrometry (LC-MS/MS) and searching in suitable databases. The second objective was the classification of the detected proteins based on physiological function and localisation using bioinformatic tools. Finally, the third objective was the identification of the potential bioactivities of peptides deriving from the digestion of the most abundant proteins through *in silico* enzymatic digestion and search in the BIOPEP database (Minkiewicz, Dziuba, Iwaniak, Dziuba, & Darewicz, 2008).

9.2 Materials and Methods

9.2.1 Chemicals

All chemicals and reagents were analytical grade. Acetonitrile (ACN), acetic acid, acetone, formic acid (FA), methanol, sodium hydroxide, ammonium bicarbonate (NH_4HCO_3), trichloroacetic acid (TCA), β-mercaptoethanol, dithiothreitol (DTT), iodoacetamide (IAA), urea, sodium chloride (NaCl), trypsin, glycine, tris-(hydroxymethyl)-aminomethane were from Sigma-Aldrich Corporation (St Louis, MO, USA). ProteoMiner™ (CPLL), Laemmli buffer, Precision Plus Protein Standards, 40% acrylamide/bis solution, N,N,N′,N′-tetramethylethylenediamine (TEMED), sodium dodecyl sulphate (SDS), and electrophoresis apparatus for one-dimensional electrophoresis were acquired at Bio-Rad (Hercules, CA, USA).

9.2.2 Treatment of apricot kernels

Apricot kernels were purchased in a local market (N'gaous, Batna, Algeria) in July 2015. The sample was extracted according to a literature method (Vita, Lucarotti, Alpi, Balestrini, Mello, Bachi, et al., 2013) with some modifications. Briefly, 0.5 g of kernels were ground to a fine powder using mortar and pestle cooling in ice. The powder was washed with hexane, then extracted with 12 mL of extraction buffer (1 M Urea, 50 mM Tris-HCl pH 8.0, 1% CHAPS, 60 mM DTT) under stirring for 2 h at 4 °C. The homogenate was centrifuged for 30 min at 8,400 g at 4 °C and extracted proteins (supernatant phase) were precipitated using 13% TCA solution containing 0.007% ß-mercaptoethanol in acetone (keeping at -20°C overnight and then at 4 °C for 2 h). The sample was centrifuged at 8,400 g at 4 °C for 30 min and the pellet was suspended into 5 mL of 50 mM Tris-HCl, pH 8.0, 50 mM NaCl. Part of the supernatant was collected as crude protein extract (**Raw** sample), whereas the other one was submitted to CPLLs incubation (Boschetti & Righetti, 2008). Before the enrichment process, the protein concentration was measured by the Bradford assay using bovine serum albumin as standard. The ProteoMiner enrichment was performed according to the manufacturer protocol. Briefly, the CPLLs spin column containing 50 µL of ProteoMiner beads, stored in a 20% ethanol and 0.5% ACN solution, was conditioned through washing in 1 mL of water followed by 1 mL of phosphate buffered saline (150 mM NaCl, 10 mM NaH_2PO_4). Then, 1 mL of protein solution obtained from the previous extraction was incubated within the column for 2 h at room temperature under gently rocking. In order to remove any excess of non-adsorbed proteins, the CPLL column was washed with 1 mL of PBS buffer for 5 min and then centrifuged at 80 *g*, discarding the solution. To elute bound proteins, the ProteoMiner beads were incubated twice

with 100 µL of rehydrated elution reagent (5% acetic acid, 4 M urea, 1% CHAPS), for 15 min at room temperature and centrifuged for 2 min at 80 *g*. The eluted enriched proteins (**EEP**) were collected in a clean tube and stored at -20 °C. All **EEP** solutions were unified and precipitated using chloroform/methanol to remove the SDS excess. The resulting pellets were mixed with Laemmli buffer. Three independent experiments were performed as biological replicates.

9.2.3 SDS-PAGE analysis

For the SDS–PAGE analysis, 10 µL of **EEP** sample and **Raw** sample were loaded onto a SDS-PAGE gel, composed by a 4% polyacrylamide stacking gel (125 mM Tris-HCl, pH 6.8, 0.1%, m/v, SDS) over a 12% resolving polyacrylamide gel (375 mM Tris-HCl, pH 8.8, 0.1%, m/v, SDS buffer). The cathodic and anodic compartments were filled with Tris-glycine buffer, pH 8.3, containing SDS (0.1%, m/v). Electrophoresis was run at 100 V, until the dye front reached the bottom of the gel and at 150 V until separation end. After electrophoresis, gels were washed and stained/destained using colloidal Coomassie Blue and 7% (v/v) acetic acid in water, respectively. Images were acquired by the GS800 densitometer and analysed by software Quantity One (Bio-Rad). After this, the two gel lanes corresponding to **EEP** and **Raw** were cut into 11 segments along migration path. Each sample was rinsed with pure water and submitted to standard reduction and alkylation procedures. The reducing solution (150 µL of a 1 mg/mL DTT solution in 50 mM NH_4HCO_3) was added to gel pieces and incubated for 20 min at 56 °C; after its removal, gel pieces were added twice to the washing solution [ACN/50 mM NH_4HCO_3 (1:1)]. The alkylating solution (150 µL, 20 mg/mL IAA solution in 50 mM NH_4HCO_3) was added to gel pieces and incubated in the dark for 45 min. After alkylating solution removal, gel pieces were covered by 50 mM NH_4HCO_3 adding 1 µg of trypsin solution in each sample cooling in ice. After overnight incubation at 37 °C, the resulting tryptic peptides were acidified by adding FA (pH < 3) and the solutions centrifuged at 10,000 rpm at 4 °C for 10 min. The supernatants were then lyophilised using a Speed Vac (Martin Christ), suspended in 20 µL water/ACN (98:2), added with 0.1% FA, and submitted to nano-LC-MS/MS analysis.

9.2.4 Nano LC-MS/MS analysis

A SL ion trap mass spectrometer (Agilent Technologies, Palo Alto, CA, USA), interfaced to a Chip-nanospray ion source, was operated in data-dependent acquisition mode with the installed Data Analysis software. The reconstituted peptides (4 µL) were loaded onto an enrichment

column (Zorbax 300SB-C18, 5 µm pore size) at a flow rate of 4 µL/min for 2 min using isocratic 100% solvent phase (99% water, 1% ACN, and 0.1% FA). After clean-up, the chip valve was switched to separation, conducted in a 43 mm x 75 µm analytical column packed (Zorbax 300SB-C18, 5 µm pore size). The separated peptides were eluted into the mass spectrometer at the constant flow rate of 0.3 µL/min. The LC solvent A was 95% water, 5% ACN, 0.1% FA; solvent B was 5% water, 95% ACN, 0.1% FA. The nano-pump gradient program was as follows: 5% solvent B (0-1 min), 50 % solvent B (1-32 min), 95% solvent B (32-35 min) and back to 5% in 5 min to re-equilibrate the column. The nano-ESI source operated under the following conditions: drying gas temperature 300 °C, flow rate 3 L/min (nitrogen), capillary voltage 1950 V, with endplate offset -500V. Mass spectra were acquired in positive ionisation mode, in the mass range from m/z 300-2200, with target mass 700 m/z, average of 2 spectra, ICC target 30,000 and maximum accumulation time 150 msec. Cationic peptide ions detection was performed by data dependent acquisition AutoMS(n) mode with a dynamic exclusion set at 2 spectra and released after 0.1 min. Each sample was analysed twice. All MS/MS spectra of each duplicate were combined and submitted to database search.

9.2.5 Protein identification from MS data

Proteins were identified using Spectrum Mill MS Proteomics Workbench (Rev B.04.00, Agilent Technologies, Palo Alto, CA, USA) against the Uniprot-Plants-Viridiplantae database (3,579,823 sequences) obtained from the Uniprot database (version of January, 2016) (The UniProt Consortium, 2017). The parameters used were as follows: carbamidomethylation of cysteine (+57.02 Da) and oxidation of methionine (+15.99 Da) were set as a fixed and variable modification respectively; trypsin was specified as the proteolytic enzyme; 2 missed cleavages were allowed; peptide mass tolerance was set at 1 Da, fragment mass tolerance at 0.7 Da, and peptide charge at +2 and +3. The thresholds used for peptide identification were peptide Local FDR $\leq$ 1%, Score Peak Intensity% $\geq$ 70%, difference of forward and reverse scores $\geq$ 2. Proteins were considered detected when identified by at least two peptides. Supplementary Table S1 reports a list of all identified proteins, MS/MS scores, sequence coverages, and all amino acid sequences of unique recognised peptides.

9.2.6 Functional categorisation of identified proteins

The identified proteins were categorised based on their cellular localisation, molecular functions or biological processes by using the open access software QuickGo (Binns, Dimmer,

Huntley, Barrell, O'Donovan, & Apweiler, 2009). The pathway mapping was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, Goto, Kawashima, & Nakaya, 2002). The STRING tool (Search Tool for the Retrieval of Interacting Genes) v.10 software (Szklarczyk, Franceschini, Wyder, Forslund, Heller, Huerta-Cepas, et al., 2015), set on *Arabidopsis thaliana* as reference organism, was used to generate a protein-protein interaction network (PPI). A functional analysis aimed at creating an interaction protein map was conducted subjecting the identified proteins (FASTA sequences IDs), detected in all analysed samples (**Raw** and **EEP**) in STRING. In the interactive network output, proteins are represented by nodes and interactions by connecting continuous and discontinuous lines for direct (physical) or indirect (functional) interactions, respectively. Each connection is supported by at least a literature reference or a canonical information stored in STRING dataset. The confidence value (score) was set to 0.7 (high confidence). The pathways classification was performed after the automatic functional enrichment in STRING, based on information provided by KEGG-Pathway Database.

9.2.7 *In silico* simulated gastrointestinal digestion of major storage proteins and potential biological activities of generated peptides

The investigation on bioactive peptides was conducted only on Prunin 1 and Prunin 2, the most abundant proteins in this kernel. A prediction of potential bioactive peptides encrypted in these proteins was obtained by combining different *in-silico* enzymatic digestions using the software tool PDMQ - Protein Digestion Multi Query (Haraszi, Tasi, Juhasz & Makai, 2015) in order to simulate gastrointestinal processes: pepsin (pH 1.3) was the first enzyme followed by trypsin and chymotrypsin. All generated peptides were then ranked using the tool PeptideRanker (Mooney, Haslam, Pollastri, & Shields, 2012) in order to evaluate the quality of these results. For ranking these peptides based on the probability of being bioactive, N-to-1 neural algorithm was used to produce a list of probability scores (Supplementary Table S2b). A score higher than 0.5 was considered as indicative for "bioactivity". After such filtering, the peptide sequences were searched against the tool BIOPEP (Minkiewicz, Dziuba, Iwaniak, Dziuba, & Darewicz, 2008).

9.2.8 Statistically Analysis

The proteomic analyses were conducted on two independent samples, each injected twice. The statistical analysis of the mass spectrometry identification, carried out following a decoy

(reversed) database search, was performed using the MS/MS search option in Spectrum Mill to account for false positives. Peptide scores were compared to those of reversed peptide scores to obtain a delta forward-reverse score. Database Fwd-Rev Score ≥2 and Local False Discovery Rate ≤1% were used.

9.3. Results and Discussion

9.3.1 Characterisation of apricot kernel proteome

In order to obtain an extensive identification of tiny kernel proteins, they were enriched by equalising the most abundant ones using spin columns packed with CPLLs (Aiello, et al., 2016). The workflow adopted included proteome extraction using a strong reducing buffer, pre-fractionation via CPLLs technology, protein separations by SDS-PAGE, and characterisation by MS analysis. **Figure 9.1A** shows the SDS-PAGE profiles of the **Raw** and **EEP** samples. A comparison of these two profiles confirms that CPLLs were very efficient in increasing low-abundant protein detection. The **Raw** lane is characterised by three major protein bands, with molecular weights (MW) equal to 50–75 kDa, 50 kDa and 25kDa, which correspond to Prunin 1 (E3SH28, 63 kDa), Prunin 2 (E3SH29, 53 kDa) and two uncharacterised proteins (M5XS06 and M5XPY4, 25 kDa). Contrariwise, the **EEP** lane exhibits very intense and additional bands, particularly evident in the high MW region corresponding to 100-75 kDa. As revealed by **EEP** eluate profile, the combined use of CPLLs and strong reducing extraction buffer has contributed to increase capture mainly of membrane proteins, as demonstrated by the subsequently mass spectrometry analysis.

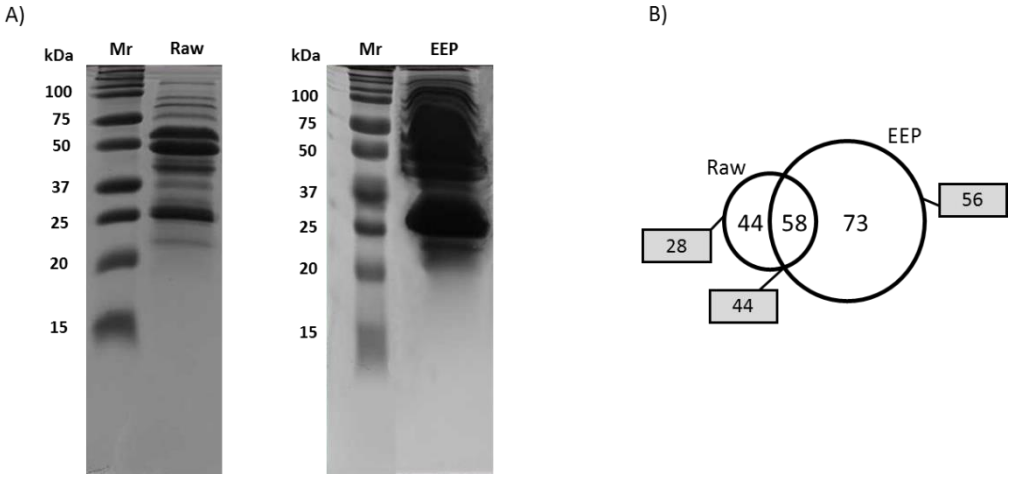


Figure 9.1 - A) SDS-PAGE profiles of **Raw** sample *versus* **EEP** sample; Mr = molecular mass ladder; staining with micellar Coomassie blue. **B)** Venn diagrams of all identified species in **Raw** and **EEP** samples against Uniprot_Viridiplantae database. In total, 44 proteins were identified exclusively in **Raw**, 73 only in **EEP**, and 58 in both samples. The squares indicate the numbers of IDs assigned to *Prunus*.

Proteins from **Raw** and **EEP** samples were reduced, alkylated, digested, and analysed using nano-LC−MS/MS. It was possible identifying 175 unique gene products with at least two matched peptides. The Venn diagram (**Figure 1B**) reports the total IDs in terms of identifications obtained by matching MS/MS data against the Uniprot *Viridiplantae* database. Among identified proteins 44 (25.2 %) and 73 (41.7%) were specific to **Raw** and **EEP** samples, respectively, whereas 58 (33.1%) were identified in both samples (**Table 9.1**). Indeed, the CPLLs capture permitted the additional identification of 73 unique gene products that could not be detected via a conventional solubilisation protocol.

Interestingly, 72% of total IDs was attributable to the *Prunus* genus, specifically three of them were assigned to almond (*Prunus dulcis*), whereas the remaining to peach (*Prunus persica*). Only a few proteins were identified by homology to other plant species, such as *Oryza sativa*, *Setaria italica*, *Triticum aestivum*, *Vitis vinifera*, and *Arabidopsis thaliana*. One of the most recent papers on *Prunus* seed proteome has considered plum and peach (Gonzalez-Garcia, Marina, Garcia, Righetti, & Fasoli, 2016): the analysis permitted to identify 141 unique gene products in plum and 97 in peach, however, only a small percentage of them belonged to the genus *Prunus*, whereas most belonged to *Glycine max*, *Vitis vinifera*, *Zea mays*, and *Populus trichocarpa*.

The most abundant proteins in apricot kernel were Prunin 1 (E3SH28) and Prunin 2 (E3SH29), belonging to the cupin family and associated to 11S globulins. Prunin 1 has been identified as the major component in almond and has been recognised as a main almond allergen (Jin, Albillos, Guo, Howard, Fu, Kothary, et al., 2009). Beyond the nutrient reservoir function, the observed protein profile includes many GTPases involved in vesicle trafficking, cytoskeletal organisation and signal transduction as recently confirmed in peach by another paper (Falchi, Cipriani, Marrazzo, Nonis, Vizzotto, & Ruperti, 2010). We have identified also phosphatidylinositol binding protein, an important lipid binding proteins, playing roles in the stabilisation of membranes, cell wall organisation, seed development and germination (Liu, Zhang, Lu, Zeng, Li, Fu, et al., 2015). In addition, some methyltransferases were recognised. These proteins are connected to the biosynthesis of volatile phenolic derivatives in plants

(Lavid, Wang, Shalit, Guterman, Bar, Beuerle, et al., 2002), as well as involved in the enzymatic methylation of polyphenols, which results in antimicrobial properties.

Table 9.1 Identified proteins in *P. armeniaca* kernels: comparison of **EEP** and **Ra**w compositions [a]

UniProt accession [b]	Taxonomy	Protein description [c]	Summed MS/MS Search score	% aa	Protein MW (kDa)	P unique	EEP	Raw
		Carbon Metabolism						
M5X7A6	PRUPE	Starch synthase, chloroplastic/amyloplastic	11.51	5.4	715.99	2		x
M5VX90	PRUPE	Fructose catabolic process	12.09	10	375.69	2		x
M5WKH0	PRUPE	Tricarboxylic acid cycle	16.65	11.3	469.45	2		x
		Stress Related and Environmental Response Proteins						
M5XL25	PRUPE	Small heat shock protein (HSP20)	27.50	15.5	173.81	2	x	x
O82011	SOLPE	17.7 kDa class I heat shock protein	20.20	16.8	176.85	2	x	x
M5X1C6	PRUPE	Small heat shock protein (HSP20)	15.20	14.5	183.45	2	x	x
M5XPY4	PRUPE	Response to stress	14.74	10	240.60	2	x	x
M5WG38	PRUPE	Response to oxidative stress	12.40	9.7	384.07	2	x	
M5XKD9	PRUPE	NB-ARC domain	14.26	4.9	109.79	2		x
M5WH03	PRUPE	Plant hormone signal transduction	15.68	6.3	767.31	2	x	
M5X2S0	PRUPE	Plant-pathogen interaction	20.35	8	592.80	2	x	x
M5XA26	PRUPE	Pectinesterase	16.68	3.4	640.13	2	x	x
M5W532	PRUPE	Dirigent-like protein	13.43	34	20.78	2		x
A5BMZ0	VITVI	Putative uncharacterized protein	17.80	5.1	135.59	2	x	x
		Oxidoreductive proteins						
D8SML7	SELML	Cytochrome P450	21.63	9.1	538.96	2	x	
M5VT70	PRUPE	Glutamine amidotransferases class-II	15.30	2.5	178.79	2	x	
M5WFA0	PRUPE	Cytochrome P450	13.93	7.4	540.28	2		x
M5VNJ0	PRUPE	Alcohol dehydrogenase GroES-like domain	13.91	14.3	353.18	2		x
M5WZ08	PRUPE	Short chain dehydrogenase	12.65	13.9	375.32	2		x
M5XNX5	PRUPE	Cytochrome P450/ E-class, group I	12.08	11.1	590.05	2		x
		Nucleotide Binding						
M5VPJ6	PRUPE	ABC transporter	12.79	7.6	692.18	2		x
M5X1Y2	PRUPE	Spliceosome	17.45	3.0	136.11	2		x
A0A0K9QEM4	SPIOL	Basic-leucine zipper domain	17.03	5.6	783.79	2		x
B2XWN9	FAGEA	DNA-directed RNA polymerase subunit beta	16.47	2.1	155.64	2	x	
M5XRM5	PRUPE	Phosphoribulokinase	16.66	6.3	104.95	2	x	x
M5X306	PRUPE	Triphosphate hydrolase activity	16.45	1.8	207.33	2	x	
M5VVC5	PRUPE	mRNA surveillance pathway	16.19	1.1	425.57	2	x	x
M0TFM1	MUSAM	Serine/threonine/dual specificity protein kinase	16.31	8.5	561.70	2		x
A0A0D9Y5E4	ORYZ	ATP binding. D-mannose binding lectin	16.55	3	193.23	2	x	x
A9RFQ7	PHYPA	Myb-like DNA-binding domain	16.04	2.4	265.05	2	x	
M5X879	PRUPE	NB-ARC domain/TIR domain/NACHT domain	17.09	2.3	116.47	2	x	x
M5Y288	PRUPE	F-box domain	16.81	8.4	469.80	2	x	x
M5X7F6	PRUPE	NB-ARC domain/Leucine Rich repeats/ATPase domain	16.71	2.8	166.99	2	x	x

M5X3N3	PRUPE	Phosphotransferase enzyme family	14.18	4.8	94.55	2	x	x
M5VXN8	PRUPE	SET domain/Tesmin/TSO1-like CXC domain	14.17	5.2	100.69	2	x	
M5XPK4	PRUPE	Cation transporting ATPase,	14.16	4.3	114.09	2	x	x
M5XIB1	PRUPE	Snf2-ATP coupling, chromatin remodelling complex	16.02	1.3	326.26	2	x	x
M5X9G3	PRUPE	Protein tyrosine kinase/PAN-like domain	15.82	6	832.15	2	x	x
M5WJX8	PRUPE	Auxin response factor	15.78	5.2	80.07	2	x	x
M5WL01	PRUPE	Phospholipid-translocating ATPase	15.69	6.8	81.86	2	x	
M5WQM8	PRUPE	Phosphatidylinositol signaling system	15.09	2.5	204.24	2	x	x
M5WXM8	PRUPE	NAC domain	13.69	4.7	967.98	2	x	x
M5X2H9	PRUPE	tRNA synthetases class	15.46	4.3	91.41	2	x	
M5VX87	PRUPE	Uncharacterized protein	12.51	7	66.15	2		x
A0A078IK57	BRANA	ATPase family	19.07	3.4	130.69	2	x	x
K3XUR4	SETIT	Phosphatidylinositol-4-phosphate 5-Kinase	18.69	1.9	204.14	2	x	x
A8HM74	CHLRE	Kinesin motor domain/K-box region	17.23	10.3	54.06	2	x	
M5X7V1	PRUPE	Protein serine/threonine kinase activity	17.67	8.7	51.44	2		x
V4V803	ROSI	Protein kinase domain/Leucine Rich Repeat	17.13	5.6	100.34	2		x
C1MUM4	MICPC	Helicase conserved C-terminal domain	17.97	5.2	112.94	2	x	
		Transcription factor activity						
A0A0E0EBP5	ORYZ	Transcription factor TFIIB repeat	16.03	6.7	69.96	2	x	
F2DTK8	HORVD	GRAS domain family	15.38	7.4	74.12	2	x	
M5VWQ0	PRUPE	Anaphase-promoting complex subunit 4 WD40 domain	16.16	4.7	124.87	2	x	
M5VYG0	PRUPE	Mediator complex subunit 13 C-terminal	11.04	2.8	206.67	2		x
M5W6G9	PRUPE	Mediator complex subunit MED14	15.85	2	188.73	2	x	x
A0A087SNJ1	AUXPR	RNA-induced silencing complex	17.44	5.7	94.44	2		x
F4IHS2	ARATH	Chromatin structure-remodeling complex protein SYD	15.97	0.7	391.97	2		x
M5XKW1	PRUPE	PB1 domain	13.91	3.6	106.72	2	x	x
A0A087H600	ARAAL	U3 ribonucleoprotein	18.83	7.1	59.37	2	x	x
		Transferase activity						
M5VJV5	PRUPE	Hexosyltransferase	16.42	11.1	43.36	2	x	
M5WDX6	PRUPE	MT-A70	14.99	8.7	72.92	2	x	x
M5W9P4	PRUPE	Glycosyltransferase family	14.70	7.3	59.077	2	x	
M5WEC8	PRUPE	Putative S-adenosyl-L-methionine-dependent methyltransferase	14.62	7.1	70.40	2	x	x
M5XEZ7	PRUPE	Early transcription elongation factor of RNA pol II	13.88	1.9	161.24	2	x	x
M5VUM2	PRUPE	16S rRNA methyltransferase	12.61	11.7	43.57	2		x
M5WAL1	PRUPE	Glycosyltransferase family 29	11.26	7	49.68	2		x
M5WMV1	PRUPE	Ubiquitin ligase	19.31	7.1	60.91	2	x	x
		Transporter activity						
M5X572	PRUPE	ABC_membrane	31.46	8.1	137.73	4	x	x
M5XIX7	PRUPE	Mito_carr	26.46	23.4	34.54	3		x
R0HKY1	BRAS	Substrate-specific transmembrane transporter activity	17.20	10.4	55.13	2		x
M5WAE5	PRUPE	Membrane insertase	14.08	8.1	60.05	2	x	
M5XM22	PRUPE	RNA transport	13.03	14.7	19.06	2	x	x

M5WUT7	PRUPE	ABC2_membrane	19.58	4.3	110.56	2		x
M5WRV7	PRUPE	Voltage gated chloride channel	18.30	5	84.04	2		x
A0A059AD39	EUCGR	ABC transporter	17.45	4.9	136.39	2		x
A0A078G4Y7	BRANA	ABC transporter	15.68	4.6	137.29	2		x
S8CX84	LAMI	Nucleoporin autopeptidase	17.02	5.1	114.76	2		x
M5VYB7	PRUPE	Proton-dependent oligopeptide transporter family	17.33	8.7	63.98	2	x	
M5WDR1	PRUPE	Uncharacterized protein	14.41	7.5	71.64	2	x	
M5X292	PRUPE	Sugar transporter	14.27	6.1	55.94	2	x	
M5X4E2	PRUPE	Endocytosis	15.03	10.9	44.71	2	x	
M5WPM9	PRUPE	ABC transporter	13.64	2.1	165.79	2	x	
M5VMX6	PRUPE	Major Facilitator Superfamily	13.48	13.5	48.16	2		x
M5XKR7	PRUPE	Cation transmembrane transporter activity	17.90	5.6	98.39	2		x
A0A0D2NP98	CHLO	Autophagy-related protein 3	20.15	14.1	39.30	2	x	x
		Kinase activity						
M8BR74	AEGTA	Casein kinase I isoform delta-like protein	15.85	3.6	145.93	2	x	
M5VG74	PRUPE	Protein kinase domain	14.43	3.8	148.66	2		x
M5WQP2	PRUPE	Protein tyrosine kinase	14.91	4.9	114.65	2		x
M5X8G0	PRUPE	Protein kinase domain	13.55	9.4	67.11	2		x
M5X2Q8	PRUPE	Di-glucose binding within endoplasmic reticulum	13.52	6.1	98.79	2		x
M5W5T2	PRUPE	Phosphatidylinositol 3- and 4-kinase	11.66	2	280.37	2		x
		Hydrolase activity						
M5WI23	PRUPE	Aspartic-type endopeptidase activity	31.00	8.8	47.59	2	x	
M5X2H5	PRUPE	Alpha-1,2-Mannosidase	15.15	9.3	70.64	2	x	x
A0A0D2V8W9	GOSRA	Dynamin GTPase effector domain	18.59	7.8	78.06	2	x	
M5XVK6	PRUPE	Lipase/Hydrolase	17.99	8.2	69.94	2		x
M5XBV0	PRUPE	Protein serine/threonine phosphatase activity	15.03	9.7	47.49	2	x	x
M5WYB9	PRUPE	Metal dependent phosphohydrolases	12.94	6.1	80.54	2	x	x
M5VWB9	PRUPE	Exonuclease/phosphatase	13.86	13.1	57.98	2		x
M5XKG8	PRUPE	Hydroxyisourate hydrolase activity	14.68	15.4	36.92	2	x	
A0A0D2RN96	GOSRA	Pectate lyase	15.99	16.6	42.62	2	x	
		Binding activity						
A0A077S2P4	WHEAT	IQ calmodulin-binding motif	20.01	1.5	259.90	2	x	x
V4KUD8	EUTSA	Fatty-acyl-CoA binding	17.32	45	101.54	2	x	x
C1ECC7	MICSR	Zinc ion binding	17.42	6.7	52.82	2	x	x
M5XLQ2	PRUPE	BAH domain /Agenet domain	17.47	8.1	73.47	2		x
M5VWP7	PRUPE	DYW family of nucleic acid deaminases	17.33	5.1	86.51	2		x
M5X7A2	PRUPE	GTPase activity	17.27	4.3	72.88	2		x
M5X8P0	PRUPE	Phosphatidylinositol 3-phosphate binding	14.05	4.5	112.99	2	x	x
M5X715	PRUPE	Glycosyltransferase	13.97	3	110.87	2	x	x
E3W0S0	ROSI	DNA-directed RNA polymerase subunit beta	13.41	4.2	121.82	2		x
M5VPV2	PRUPE	Regulator of chromosome condensation	10.77	4.6	121.38	2		x
M5WQZ5	PRUPE	Chromatin binding	13.31	3	174.52	2	x	x
M5W7U6	PRUPE	DYW family of nucleic acid deaminases	16.60	5.8	75.17	2		x

M5X1G9	PRUPE	Oleosin	21.23	12.8	15.61	2		x
M5XB96	PRUPE	Mitochondrial ribosomal protein subunit L20	17.82	5.8	83.16	2	x	x
M5VWL5	PRUPE	Phosphatidylinositol binding	18.94	5.7	75.78	2		x
M5X9Z6	PRUPE	Proline-binding domain	15.44	3.2	180.67	2	x	
M5WNU7	PRUPE	Plant-specific actin-binding protein	12.94	6.1	80.54	2	x	x
M5W984	PRUPE	Mismatch repair	27.38	7.8	106.68	3	x	x
M5WS62	PRUPE	Uridine kinase	19.15	7.9	53.94	2		x
M5XRX7	PRUPE	RNA binding activity	18.11	2.6	202.43	2		x
M5XS06	PRUPE	CTLH/CRA C-terminal to LisH motif domain	15.82	17.6	24.98	2	x	x
		DNA metabolic process						
M5X0E1	PRUPE	Nucleotide excision repair activity	23.58	6.6	117.19	3	x	
Q5NAA4	ORYSJ	Helicase-like protein	21.38	2.4	188.53	2	x	
		Storage Proteins						
E3SH28	PRUDU	Prunin 1	132.8	25.7	63.39	9	x	x
E3SH29	PRUDU	Prunin 2	96.65	12.5	57.23	6	x	x
M5Y3W2	PRUPE	Nutrient reservoir protein	43.03	5.6	95.67	3	x	x
		Cytoscheleton organization and multicellular component development						
M5XNW4	PRUPE	Microtubule-associated protein 7	13.91	3.6	106.72	2	x	x
I6UTH7	SORBI	Tan1	19.28	5.9	38.34	2	x	x
		Translation activity						
M5WYZ2	PRUPE	Translation initiation factor eIF3	15.77	3.5	185.56	2		x
A0A022Q5F9	ERYGU	Ribosomal protein S18	19.00	14.7	30.513	2	x	
M5Y4G5	PRUPE	Zinc finger PHD-type	18.14	3.7	168.10	2	x	x
		Other						
M5Y6F3	PRUPE	Band_7	17.34	11.3	47.52	2		x
M5VPQ9	PRUPE	HAUS augmin-like complex subunit 3	15.19	6.4	70.02	2	x	x
A9YTJ2	PRUDU	F-box associated interaction domain	14.22	16.6	31.78	2	x	x
M5VVG0	PRUPE	Retrotransposon gag protein	12.45	6	71.06	2	x	x
M5XVF3	PRUPE	Sieve element occlusion (SEO) protein	15.24	6.2	91.56	2		x

a) Proteins with unknown function are listed in Supplementary Table S1 (data available on line as supporting materials)

b) Protein ID, according to the Uniprot database.

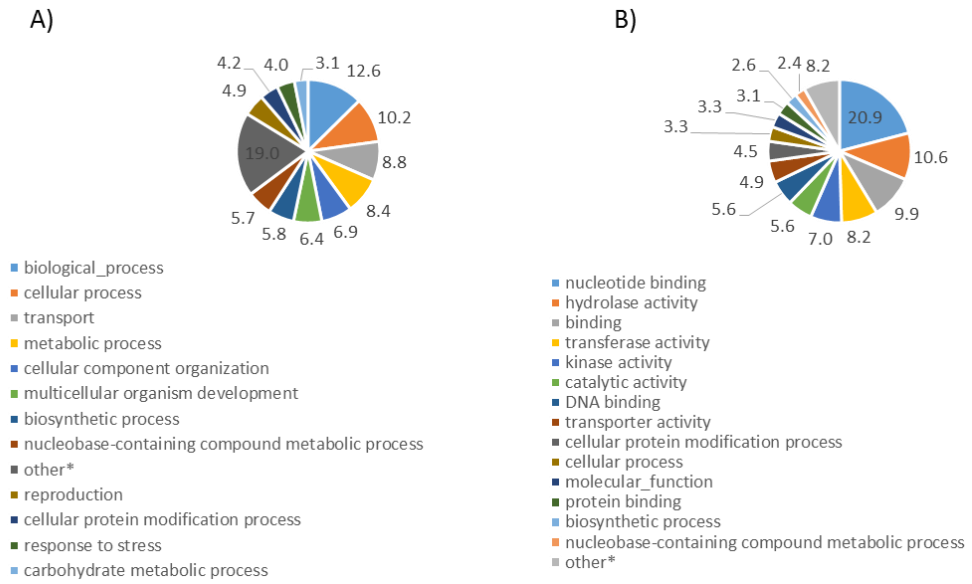
c) Protein description, according to the KEGG databas

9.3.2 Protein functional data analysis

The protein species identified were categorised into specific clusters, using the GO annotation tool, which allows automatic information retrieval from the website available databank. **Figure 9.2** shows the obtained clusters in which they are involved: biological processes (**panel A**), molecular functions (**panel B**), and cellular components (**panel C**).

Focusing the attention on biological processes (**Figure 9.2A**), the largest clusters include proteins involved in biological process regulation (12.6%), cellular processes (10.2%), transport activities (8.8%), metabolic processes (8.4%), cellular component organisation (6.9%) and multicellular organisms development (6.4%). Fewer proteins are involved in cellular processes, such as growth factor, response to extracellular stimulus, and in cell-cell signalling. Regarding molecular functions (**Figure 9.2B**), 20.9% proteins are involved in nucleotide binding, 10.6% in hydrolase activities, 7.0% in kinase activities, 5.6% in catalytic activities, and only 2.6% in biosynthetic processes. As for subcellular localisation (**Figure 9.2C**), most detected IDs are localised in membranes (48.4%), nucleus (10.1%), plastids (9.2%), cytoplasm (6.0%), and cell walls (1.4%).

Figure 9.3 shows a comparison of the GO for molecular function annotation between **EEP** and **Raw**, based on the proteins uniquely identified in each sample. A 47% increment in the proteome discovery, attributed to minor proteins unrevealed without the CPLLs treatment, was observed for binding activity, whereas a 31% enhancement for catalytic activity.



C)

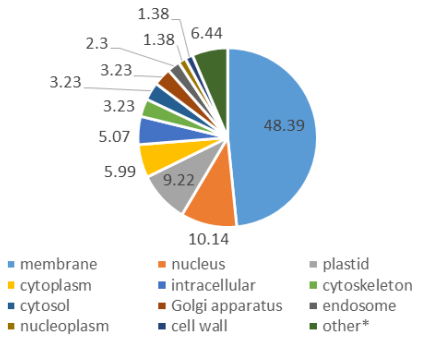


Figure 9.2 - Gene Ontology (GO) analysis of identified gene products. Pie graphs of (A) biological processes, (B) molecular functions, and (C) cellular components show the percentage of proteins in each functional category.

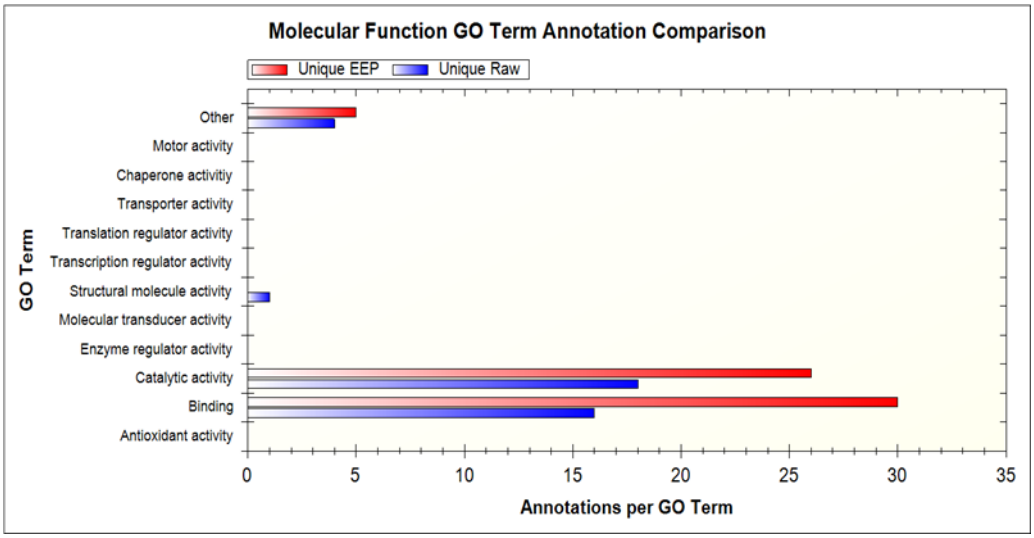


Figure 9.3 - Molecular Function GO Term Annotation Comparison obtained plotting the unique ID entries for Raw and EEP, respectively.

9.3.3 Protein biology

The protein classification based on their functional roles highlighted four main categories, transcription regulators, transmembrane transporters, stress-related proteins, and binding activities, providing a relevant contribution to kernel proteome knowledge during development and maturation.

GRAS domain protein and TFIIB/Zinc finger transcription factor have been identified as transcriptional regulators. Involved in the yield and quality of storage compounds, transcription

factors are considered the main protagonists controlling early seed developments as well as genome-wide epigenetic events (Ikeda, 2012). Associated to well-established TFIIB, the mediator complex family, including MD13, has emerged as the most crucial cofactor in RNAP II-mediated transcriptional events due to its role either in growth or developmental processes or biotic and abiotic stress response (Samanta & Thakur, 2015).

The ATP-binding cassette (ABC) transporter family, responsible of adenosine triphosphate (ATP) hydrolysis, one of the largest protein superfamilies in biology represented in all living organisms, is strictly correlated to active transport of a wide variety of substrates through the mitochondrial outer membranes and cytosol (Rea, 2007). The ABC transporter protein superfamily members share a hydrolytic ability useful in a wide array of functions, including DNA repair and RNA translocation.

The voltage dependent anion channel (VDAC) assumes an important role in the regulation of energetic and metabolic functions (Shoshan-Barmatz & Ben-Hail, 2012). It is involved in Ca^{2+} transport across the mitochondrial outer membrane (Bathori, Csordas, Garcia-Perez, Davies, & Hajnoczky, 2006). The regulation of mitochondrial physiology needs an efficient metabolic exchange systems identified here into the Solute Carrier Protein Family, responsible of sugar-phosphate/phosphate exchange.

Pectinesterase (PME) was identified as the major protein involved in cell wall metabolism. The demethylation of cell wall pectins is mediated by pectin methylesterases, whose activity alters cell walls and mediates various physiological and biochemical processes in plants, including elongation growth, water uptake, and fruit ripening (Peaucelle, Braybrook, & Hofte, 2012).

Several proteins related to stress-response, including 17.7 KDa class I heat shock proteins (HSPs) and NB-ARC domain protein, were also detected. Although HSPs were initially identified as being induced by heat stress, they are overrepresented under a variety of physical and chemical stimuli as well as during plant development and seeds growth (Koo, Park, Kim, Suh, Lee, Lee, et al., 2015; W. Wang, Vinocur, Shoseyov, & Altman, 2004). The possible important roles of HSPs in fruit development and ripening have been recently reported in tomato (Neta-Sharir, Isaacson, Lurie, & Weiss, 2005), apple (A. D. Wang, Tan, Tatsuki, Kasai, Li, Saito, et al., 2009), and apricot (Grimplet, Romieu, Audergon, Marty, Albagnac, Lambert, et al., 2005; Manganaris, Rasori, Bassi, Geuna, Ramina, Tonutti, et al., 2011).

9.3.4 PPI network of apricot kernel

A proteome interactomic map was obtained using the STRING tool for obtaining cross-correlation information. *A. thaliana* was selected as a reference organism, considering the lack of an extensive genome database for *Prunus* specie and the phylogenetic proximity of these species. **Figure 9.4** shows the PPI network for apricot kernel (*p* value = 0.045) calculated by STRING (confidence score value > 0.7). By removing unconnected proteins, the PPI network contains 34 interactions. Proteins with the highest score values are characterised by the highest connectivity in the network. Supplementary **Table 9.1S** lists this latter information and all details, including phylogenetic co-occurrence, genetic neighbourhood and co-expression for each interaction. Among the overall interactions, 14 were endowed high confidence score value (upper to 0.9). The widest interactions involve proteins belonging to RNA-polymerase family protein as well as proteins with transcription regulation and nucleosome positioning activities, such as Chromatin structure-remodelling complex protein SYD.

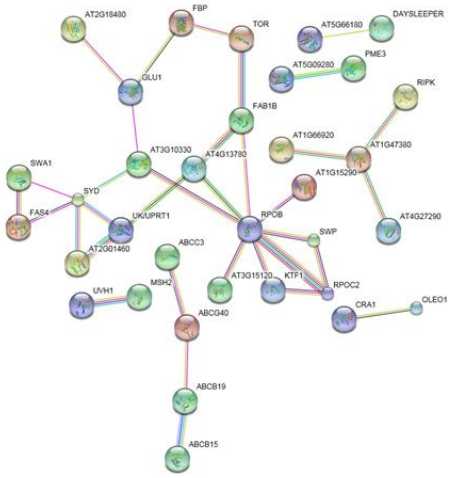


Figure 9.4 - Protein-Protein interaction network obtained by STRING software (p-value = 0.045). Different colored edges represent the existence of different types of evidence. A green line indicates neighborhood evidence; a blue line, gene-co-occurrence; a yellow line, text mining; a purple line, experimental evidence.

9.3.5 Bioactivities of peptides from in silico digestion of prunins

There is now a big interest for bioactive peptides from food sources (Arnoldi, Zanoni, Lammi, & Boschin, 2015) and some previous studies (Gonzalez-Garcia, Marina, & Garcia, 2014; Gonzalez-Garcia, Puchalska, Marina, & Garcia, 2015; Vasquez-Villanueva, Marina, & Garcia,

2015) had given indication that hydrolysates from total protein extracts from kernels of the *Prunus* genus provide antioxidant activities and angiotensin converting enzyme (ACE) inhibition. The final part of the work was, therefore, dedicated to investigate the possible physiological roles of peptides deriving from Prunin 1 (E3SH28) and Prunin 2 (E3SH29), the major storage proteins in apricot kernel. They were subjected to *in silico* digestion by sequential hydrolysis with pepsin, trypsin, and chymotrypsin. Although useful for obtaining a feasible prediction, this model does not exhaustively represent gastrointestinal digestion, because it does not consider several important aspects, such as peptidase activities, pH variations, and microbiota effects. Only peptides derived from 0 missed cleavage and containing at least 4 amino acid residues were further evaluated, whereas smaller peptides (i.e. di- and tripeptides with MW < 400 Da) were not considered, since they cannot unambiguously belong to a single protein.

To optimise the potential bioactive candidates selection, the predicted peptidome map was ranked by the tool PeptideRanker. A score value was assigned to each peptide using N-to-1 neural network probability: peptides showing score values higher than 0.5 were considered to be potentially bioactive. In a probability range from 0 to 1, predicted values closest to 1 indicate a more confident prediction that the candidate resembles a bioactive peptide (Mooney, Haslam, Pollastri, & Shields, 2012). The total number of predicted peptides after *in silico* digestion was 10 for Prunin 1 and 14 for Prunin 2 (**Table 9.2**). Out of these 24 peptides, 7 sequences had high probability to be bioactive, three deriving from Prunin 1 and four from Prunin 2.

Table 9.2 – Predicted bioactive peptides generated from Prunin 1 (E3SH28) and Prunin 2 (E3SH29) by *in silico* digestion. The best candidates (Peptide Ranker score values higher than 0.5) are labelled in bold.

Protein Ids	Peptide	Score	Potential bioactive sequence	Activity
E3SH28				
	GRPR	0.78	GRP, PR, RP, GR	ACE inhibitor
			RP	DPP IV inhibitor
	AGNQG	0.24	AG, GV, QG	ACE inhibitor
			AG, GV, IQ, NQ, QA, QG, QQ, VI	DPP IV inhibitor
	IVPQNH	0.15	VP, PQ,	ACE inhibitor
			IV	Glucose uptake stimulating peptide
			VP, NH, PQ, QN	DPP IV inhibitor
	NLPIL	0.62	IL	Glucose uptake stimulating peptide
			LP, IL, NL,PI	DPP IV inhibitor
	GQNDNR	0.26	GQ	ACE inhibitor
			GQ	Neuropeptide inhibitor
			GQ, DN, ND, NR, QN	DPP IV inhibitor
	QQGEQGRPGQH	0.21	PG,	Prolyl endopeptidase inhibitor

			GRP, RP, GR, GQ, GE, QG, PG	ACE inhibitor
			PG	Antithrombotic
			PG	Peptide regulating the stomach mucosal membrane activity
			GQ	Neuropeptide inhibitor
			GQ, RP, GE, PG, QG, QH, QQ	DPP IV inhibitor
	SPHW	0.91	PH	ACE inhibitor
			PHW	Antioxidant
			SP, HW, PH	DPP IV inhibitor
	SPQNQCQ	0.38	PQ	ACE inhibitor
			SP, NQ, PQ, QN	DPP IV inhibitor
	VSSDH	0.09	VS	DPP IV inhibitor
E3SH29				
	EPDNH	0.16	EP, DN, NH	DPP IV inhibitor
	IPQNH	0.26	IP, PQ	ACE inhibitor
			IP, NH, PQ, QN	DPP IV inhibitor
	QSEAGVT	0.1	AG, GV, EA	ACE inhibitor
			SE	Stimulating vasoactive substance release
			AG, ES, GV, IQ, QS,TE, VT	ACE inhibitor
	NPQGGR	0.57	GR, GG, QG, PQ	ACE inhibitor
			NP, GG, PQ, QG	DPP IV inhibitor
	NPQQQGR	0.31	GR, QG, PQ	ACE inhibitor
			NP, PQ, QG, QQ	DPP IV inhibitor
	NPSDPQF	0.76	PQ	ACE inhibitor
			NP, DP, PQ, PS, QF	DPP IV inhibitor
	GQDDNR	0.27	GQ	ACE inhibitor
			GQ	Neuropeptide inhibitor
			GQ, DN, NR, QD, QG	DPP IV inhibitor
	SATSPPR	0.55	PR, PP	ACE inhibitor
			PP, SP, AT, TS	DPP IV inhibitor
	TPHW	0.75	PH	ACE inhibitor
			PHW	Antioxidant
			TP, HW, PH	DPP IV inhibitor

The complete peptidome map (24 peptides) was then searched by using the BIOPEP database, including information about known bioactive peptides, in order to associate potential bioactivities to their sequences. This suggested numerous potential bioactivities, including dipeptidyl peptidase-IV (DPP-IV) and angiotensin converting enzyme I (ACE) inhibition, antioxidant properties, and glucose uptake stimulating activity. Most predicted peptides exhibited multifunctional activities, since they contain common structural requirements for each activity (Yea, Ebrahimpour, Hamid, Bakar, Muhammad, & Saari, 2014). Our discussion will be restricted to peptides with high PeptideRanker scores.

All seven peptides were predicted as potential inhibitors of DPP-IV activity. This is a new molecular target correlated with type 2 diabetes development. This ubiquitous enzyme has been

shown to cleave and inactivate glucagon-like peptide-1 (GLP-1) and glucose dependent insulinotropic polypeptide (GIP) in the postprandial phase, leading to a loss in their insulinotropic activity (Juillerat-Jeanneret, 2014). In order to exert this activity, a peptide should have a hydrophobic character, a length varying from 2 to 8 amino acid residues, and contain a Pro residue located at the first, second, third, or fourth N-terminal position (Lammi, Zanoni, Arnoldi, & Vistoli, 2016). Based on these considerations, GRPR, NLPIL, NPQGGR, NPSDPQF, SATSPPR, and TPHW seem good candidates as DPP-IV inhibitors. DPP-IV inhibitory peptides have been found in hydrolysates from lupin and soybean protein (Lammi, et al, 2016).

Six peptides were predicted as ACE-inhibitors, i.e. they are potentially hypotensive. The binding of a peptide to the ACE active site is strongly influenced by its C-terminal sequence (Hernandez-Ledesma, Contreras, & Recio, 2011). Structure-activity studies have shown that a positively charged C-terminal residue, such as the ε-amino group of Lys and the guanidine group of Arg, is very important to exert this activity as well as a hydrophobic amino acid residue (aromatic or branched side chain) at least in one of the three C-terminal positions. These peptide motifs are present within peptides GRPR, NPQGGR, NPSDPQF, SATSPPR and TPHW.

Two peptides, SPHW and TPHW, were predicted to be antioxidant for the presence of the fragment *PHW*, which provides this kind of activity. Only peptide NLPIL was predicted as glucose uptake stimulating, owing to the presence of the sequence *IL*. This peptide had been already identified in the sequence of plum kernel proteins (*Prunus domestica* L.) and shown to exert antioxidant capacity *in vitro* (Gonzalez-Garcia, Marina, Garcia, Righetti, & Fasoli, 2016).

An open issue is the stability of these peptides toward hydrolysis by endoproteinases activities that was checked using the open access tool PROSPER (Song, Tan, Perry, Akutsu, Webb, Whisstock, et al., 2012). This process showed that out of these seven peptides only two are susceptible to partial hydrolysis by endoproteinases: the former, NLPIL, might be cleaved by elastase-2, matrix metalloprotease-9 and cathepsine-K, whereas the latter, NPQGGR, only by cathepsine-K.

However, the intestinal phase of digestion is very difficult to model, since small intestinal brush border membrane contains many amino and carboxyl exopeptidases whose action here has not been considered. This is certainly a main limitation of this approach.

9.4. Conclusions

In conclusion, taking advantage of high-throughput technologies, this work has provided innovative information on the proteome of apricot kernel. This represents a major improvement in the knowledge of this food material. Although in general this kernel finds a rare use in human nutrition owing to its bitterness, there are some specific recipes, such as Italian *amaretti*, in which this sensorial characteristic is looked for and carefully modulated precisely through a sapient addition of bitter apricot kernels. Moreover, kernels from sweet varieties may be used without any precaution as roasted salted kernels consumed as titbit or as a flour to be included in different food formulations.

The presence of potentially bioactive peptides encrypted inside its proteins sequences suggests that apricot kernel proteins may offer useful health benefits, increasing even more the interest for this kernel. This fact would help maximising available resources and possibly providing new ingredients to exploit in innovative healthy foods. Finally, considering the challenges in correlating food protein sequences and health, we observe that, in order to facilitate proteome analysis in food chemistry, it would be very useful to develop specific bioinformatics tools including data processing, clustering, dynamics, and integration at various omics levels, and designed to take into account properties, such as taste, technological functions and health promoting features.

Acknowledgments

The authors are grateful to DGRSDT-MESRS (Algeria) for financial support.

Supplemental Table 9.1 STRING Interactions															
#node1	node2	node1_string_internal_id	node2_string_internal_id	node1_external_id	node2_external_id	neighborhood_on_chromosome	gene_fusion	phylogenetic_cooccurrence	homology	coexpression	experimentally_determined_interaction	database_annotated	automated_text mining	combined_score
RPOB	RPOC2	142524	142522	3702.ATCG00190.1	3702.ATCG00170.1	0.642	0.312	0.269	0	0.941	0.649	0	0.926	0.999
UVH1	MSH2	139622	128540	3702.AT5G41150.1	3702.AT3G18524.1	0	0	0	0	0.096	0.951	0.54	0.637	0.991
SWA1	FAS4	126610	118602	3702.AT2G47990.1	3702.AT1G33390.1	0	0	0	0	0.563	0.954	0	0.557	0.99
RPOB	KTF1	142524	136546	3702.ATCG00190.1	3702.AT5G04290.1	0	0	0	0	0.22	0.9	0	0.645	0.969
SYD	FAS4	124495	118602	3702.AT2G28290.1	3702.AT1G33390.1	0	0	0	0	0.191	0.951	0	0	0.958
GLU1	AT3G1033	136529	127614	3702.AT5G04140.2	3702.AT3G10330.1	0	0	0	0	0	0.954	0	0	0.954
SWA1	SYD	126610	124495	3702.AT2G47990.1	3702.AT2G28290.1	0	0	0	0	0	0.951	0	0	0.951
RPOB	SWP	142524	127048	3702.ATCG00190.1	3702.AT3G04740.1	0	0	0	0	0.105	0.929	0	0.186	0.943
RPOB	AT3G1033	142524	127614	3702.ATCG00190.1	3702.AT3G10330.1	0	0	0	0	0.065	0.834	0	0.641	0.939
UK/UPR1	SYD	139592	124495	3702.AT5G40870.1	3702.AT2G28290.1	0	0	0	0	0	0.794	0.243	0.555	0.924
SYD	AT2G0140	124495	122426	3702.AT2G28290.1	3702.AT2G01460.1	0	0	0	0	0	0.794	0.243	0.555	0.924
AT5G0928	PME3	137028	128074	3702.AT5G09280.1	3702.AT3G14310.1	0.189	0	0	0	0	0	0.9	0.147	0.924
AT5G6618	DAYSLE1	142352	129883	3702.AT5G66180.1	3702.AT3G42170.1	0	0	0	0	0	0	0	0.922	0.922
RPOC2	KTF1	142522	136546	3702.ATCG00170.1	3702.AT5G04290.1	0	0	0	0	0.084	0.747	0	0.628	0.906
ABCB19	ABCB15	129678	129624	3702.AT3G28860.1	3702.AT3G28345.1	0	0	0.527	0.94	0	0	0.9	0.69	0.906
UK/UPR1	AT2G0140	139592	122426	3702.AT5G40870.1	3702.AT2G01460.1	0.083	0	0	0	0.064	0.568	0.726	0.145	0.897
AT3G1033	SYD	127614	124495	3702.AT3G10330.1	3702.AT2G28290.1	0	0	0	0	0	0	0.759	0.566	0.89
RPOC2	SWP	142522	127048	3702.ATCG00170.1	3702.AT3G04740.1	0	0	0	0	0.083	0.836	0	0.125	0.856
AT4G1378	FAB1B	133283	128070	3702.AT4G13780.1	3702.AT3G14270.1	0.065	0	0	0	0.066	0.581	0	0.647	0.853
ABCC3	ABCG40	127918	116864	3702.AT3G13080.1	3702.AT1G15520.1	0	0	0	0	0.342	0.099	0	0.719	0.819
TOR	FBP	119335	118929	3702.AT1G50030.1	3702.AT1G43670.1	0	0	0	0	0	0.756	0	0.248	0.808
GLU1	AT2G1848	136529	123467	3702.AT5G04140.2	3702.AT2G18480.1	0	0	0	0	0	0.571	0	0.567	0.806
RPOB	AT3G1512	142524	128162	3702.ATCG00190.1	3702.AT3G15120.1	0	0	0	0	0.067	0.775	0	0.135	0.802
RPOB	AT1G1529	142524	116841	3702.ATCG00190.1	3702.AT1G15290.1	0	0	0	0	0.1	0.766	0	0	0.78
CRA1	OLEO1	139960	134569	3702.AT5G44120.3	3702.AT4G25140.1	0	0	0	0	0.759	0	0	0.101	0.774
GLU1	FBP	136529	118929	3702.AT5G04140.2	3702.AT1G43670.1	0	0	0	0	0.154	0	0	0.727	0.759
ABCB19	ABCG40	129678	116864	3702.AT3G28860.1	3702.AT1G15520.1	0	0	0	0	0	0.09	0	0.735	0.749
UK/UPR1	AT4G1378	139592	133283	3702.AT5G40870.1	3702.AT4G13780.1	0.229	0	0	0	0.137	0	0	0.646	0.743
RIPK	AT1G4738	122827	119066	3702.AT2G05940.1	3702.AT1G47380.1	0.642	0	0	0	0	0.09	0	0.269	0.741
AT4G2729	AT1G4738	134801	119066	3702.AT4G27290.1	3702.AT1G47380.1	0.642	0	0	0	0	0.09	0	0.269	0.741
AT1G6692	AT1G4738	120934	119066	3702.AT1G66920.2	3702.AT1G47380.1	0.642	0	0	0	0	0.09	0	0.269	0.741
FAB1B	TOR	128070	119335	3702.AT3G14270.1	3702.AT1G50030.1	0	0	0	0	0.147	0.224	0.19	0.557	0.73
RPOB	AT4G1378	142524	133283	3702.ATCG00190.1	3702.AT4G13780.1	0.07	0	0	0	0.166	0	0	0.647	0.702
RPOB	FAB1B	142524	128070	3702.ATCG00190.1	3702.AT3G14270.1	0	0	0	0	0	0.361	0	0.552	0.701

References

Aiello, G., Fasoli, E., Boschin, G., Lammi, C., Zanoni, C., Citterio, A., & Arnoldi, A. (2016). Proteomic characterization of hempseed (*Cannabis sativa* L.). *Journal of Proteomics, 147*, 187-196.

Arnoldi, A., Zanoni, C., Lammi, C., & Boschin, G. (2015). The role of grain legumes in the prevention of hypercholesterolemia and hypertension. *Critical Reviews in Plant Sciences, 34*(1-3), 144-168.

Bathori, G., Csordas, G., Garcia-Perez, C., Davies, E., & Hajnoczky, G. (2006). Ca2+-dependent control of the permeability properties of the mitochondrial outer membrane and voltage-dependent anion-selective channel (VDAC). *Journal of Biological Chemistry, 281*(25), 17347-17358.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., & Apweiler, R. (2009). QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics, 25*(22), 3045-3046.

Boschetti, E., & Righetti, P. G. (2008). The ProteoMiner in the proteomic arena: a non-depleting tool for discovering low-abundance species. *Journal of Proteomics, 71*(3), 255-264.

D'Ambrosio, C., Arena, S., Rocco, M., Verrillo, F., Novi, G., Viscosi, V., Marra, M., & Scaloni, A. (2013). Proteomic analysis of apricot fruit during ripening. *Journal of Proteomics, 78*, 39-57.

Durmaz, G., & Alpaslan, M. (2007). Antioxidant properties of roasted apricot (*Prunus armeniaca* L.) kernel. *Food Chemistry, 100*(3), 1177-1181.

Esteve, C., D'Amato, A., Marina, M. L., García, M. C., Citterio, A., & Righetti, P. G. (2012). Identification of olive (*Olea europaea*) seed and pulp proteins by nLC-MS/MS via combinatorial peptide ligand libraries. *Journal of Proteomics, 75*(8), 2396-2403.

Falchi, R., Cipriani, G., Marrazzo, T., Nonis, A., Vizzotto, G., & Ruperti, B. (2010). Identification and differential expression dynamics of peach small GTPases encoding genes during fruit development and ripening. *Journal of Experimental Botany, 61*(10), 2829-2842.

FAO (2012). www.fao.org (2012 data)

Femenia, A., Rossello, C., Mulet, A., & Canellas, J. (1995). Chemical composition of bitter and sweet apricot kernels. *Journal of Agricultural and Food Chemistry, 43*(2), 356-361.

Gonzalez-Garcia, E., Marina, M. L., & Garcia, M. (2014). Plum (*Prunus Domestica* L.) by-product as a new and cheap source of bioactive peptides: Extraction method and peptides characterization. *Journal of Functional Foods, 11*, 428-437.

Gonzalez-Garcia, E., Marina, M. L., Garcia, M. C., Righetti, P. G., & Fasoli, E. (2016). Identification of plum and peach seed proteins by nLC-MS/MS via combinatorial peptide ligand libraries. *Journal of Proteomics, 148*, 105-112.

Gonzalez-Garcia, E., Puchalska, P., Marina, M. L., & Garcia, M. C. (2015). Fractionation and identification of antioxidant and angiotensin-converting enzyme-inhibitory peptides obtained from plum (*Prunus domestica* L.) stones. *Journal of Functional Foods, 19*, 376-384.

Grimplet, J., Romieu, C., Audergon, J. M., Marty, I., Albagnac, G., Lambert, P., Bouchet, J. P., & Terrier, N. (2005). Transcriptomic study of apricot fruit (*Prunus armeniaca*) ripening among 13 006 expressed sequence tags. *Physiologia Plantarum, 125*(3), 281-292.

Haraszi, R., Tasi, C., Juhasz, A., & Makai, S. (2015) PDMQ - Protein Digestion Multi Query software tool to perform in silico digestion of protein/peptide sequences. bioRxiv 014019, doi.org/10.1101/014019

Hernandez-Ledesma, B., Contreras, M. D., & Recio, I. (2011). Antihypertensive peptides: Production, bioavailability and incorporation into foods. *Advances in Colloid and Interface Science, 165*(1), 23-35.

Ikeda, Y. (2012). Plant Imprinted Genes Identified by Genome-wide Approaches and Their Regulatory Mechanisms. *Plant and Cell Physiology, 53*(5), 809-816.

Jin, T. C., Albillos, S. M., Guo, F., Howard, A., Fu, T. J., Kothary, M. H., & Zhang, Y. Z. (2009). Crystal Structure of Prunin-1, a Major Component of the Almond (Prunus dulcis) Allergen Amandin. *Journal of Agricultural and Food Chemistry, 57*(18), 8643-8651.

Juillerat-Jeanneret, L. (2014). Dipeptidyl peptidase IV and its inhibitors: therapeutics for type 2 diabetes and what else? *Journal of Medicinal Chemistry, 57*(6), 2197-2212.

Kanehisa, M., Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic Acids Research, 30*(1), 42-46.

Karsavuran, N., Charehsaz, M., Celik, H., Asma, B. M., Yakinci, C., & Aydin, A. (2014). Amygdalin in bitter and sweet seeds of apricots. *Toxicological and Environmental Chemistry, 96*(10), 1564-1570.

Koo, H. J., Park, S. M., Kim, K. P., Suh, M. C., Lee, M. O., Lee, S. K., Xia, X. L., & Hong, C. B. (2015). Small Heat Shock Proteins Can Release Light Dependence of Tobacco Seed during Germination. *Plant Physiology, 167*(3), 1030-+.

Kutlu, T., Durmaz, G., Ates, B., & Erdogan, A. (2009). Protective effect of dietary apricot kernel oil supplementation on cholesterol levels and antioxidant status of liver in hypercholesteremic rats. *Journal of Food Agriculture & Environment, 7*(3-4), 61-65.

Lammi, C., Zanoni, C., Arnoldi, A., & Vistoli, G. (2016). Peptides Derived from Soy and Lupin Protein as Dipeptidyl-Peptidase IV Inhibitors: In Vitro Biochemical Screening and in Silico Molecular Modeling Study. *Journal of Agricultural and Food Chemistry, 64*(51), 9601-9606.

Lavid, N., Wang, J., Shalit, M., Guterman, I., Bar, E., Beuerle, T., Menda, N., Shafir, S., Zamir, D., Adam, Z., Vainstein, A., Weiss, D., Pichersky, E., & Lewinsohn, E. (2002). O-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. *Plant Physiol, 129*(4), 1899-1907.

Lee, H.-h., Ahn, J.-H., Kwon, A.-R., Lee, E. S., Kwak, J.-H., & Min, Y.-H. (2014). Chemical Composition and Antimicrobial Activity of the Essential Oil of Apricot Seed. *Phytotherapy Research, 28*(12), 1867-1872.

Liu, F., Zhang, X., Lu, C., Zeng, X., Li, Y., Fu, D., & Wu, G. (2015). Non-specific lipid transfer proteins in plants: presenting new advances and an integrated functional analysis. *Journal of Experimental Botany, 66*(19), 5663-5681.

Lu, Y., Lu, Z., Gao, S., Lu, Y. M., Lu, Z. R., & Gao, S. Z. (1994). A study on the evolution relationship and classification of apricots via peroxidase isozyme zymogram analysis. *Acta Agriculturae Boreali-Sinica, 9*(4), 69-74.

Manganaris, G. A., Rasori, A., Bassi, D., Geuna, F., Ramina, A., Tonutti, P., & Bonghi, C. (2011). Comparative transcript profiling of apricot (*Prunus armeniaca* L.) fruit development and on-tree ripening. *Tree Genetics & Genomes, 7*(3), 609-616.

Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M., & Darewicz, M. (2008). BIOPEP database and other programs for processing bioactive peptide sequences. *Journal of AOAC International, 91*(4), 965-980.

Mooney, C., Haslam, N. J., Pollastri, G., & Shields, D. C. (2012). Towards the Improved Discovery and Design of Functional Peptides: Common Features of Diverse Classes Permit Generalized Prediction of Bioactivity. *Plos One, 7*(10).

Neta-Sharir, I., Isaacson, T., Lurie, S., & Weiss, D. (2005). Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. *Plant Cell, 17*(6), 1829-1838.

Ozboy-Ozbas, O., Seker, I. T., & Gokbulut, I. (2010). Effects of resistant starch, apricot kernel flour, and fiber-rich fruit powders on low-fat cookie quality. *Food Science and Biotechnology, 19*(4), 979-986.

Ozcan, M. (2000). Composition of some apricot (*Prunus armeniaca* L.) kernels grown in Turkey. *Acta Alimentaria, 29*(3), 289-293.

Peaucelle, A., Braybrook, S., & Hofte, H. (2012). Cell wall mechanics and growth control in plants: the role of pectins revisited. *Frontiers in Plant Science, 3*.

Rea, P. A. (2007). Plant ATP-Binding cassette transporters. *Annual Review of Plant Biology, 58*, 347-375.

Righetti, P. G., Fasoli, E., & Boschetti, E. (2011). Combinatorial peptide ligand libraries: The conquest of the 'hidden proteome' advances at great strides. *Electrophoresis, 32*(9), 960-966.

Samanta, S., & Thakur, J. K. (2015). Importance of Mediator complex in the regulation and integration of diverse signaling pathways in plants. *Frontiers in Plant Science, 6*.

Seker, I. T., Ozboy-Ozbas, O., Gokbulut, I., Ozturk, S., & Koksel, H. (2009). Effects of fiber-rich apple and apricot powders on cookie quality. *Food Science and Biotechnology, 18*(4), 948-953.

Seker, I. T., Ozboy-Ozbas, O., Gokbulut, I., Ozturk, S., & Koksel, H. (2010). Utilization of apricot kernel flour as fat replacer in cookies. *Journal of Food Processing and Preservation, 34*(1), 15-26.

Shoshan-Barmatz, V., & Ben-Hail, D. (2012). VDAC, a multi-functional mitochondrial protein as a pharmacological target. *Mitochondrion, 12*(1), 24-34.

Song, J., Tan, H., Perry, A. J., Akutsu, T., Webb, G. I., Whisstock, J. C., & Pike, R. N. (2012). PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One, 7*(11), e50300.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & von Mering, C. (2015).

STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research, 43*(Database issue), D447-452.

The UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research, 45*(D1), D158-D169.

Turan, S., Topcu, A., Karabulut, I., Vural, H., & Hayaloglu, A. A. (2007). Fatty acid, triacylglycerol, phytosterol, and tocopherol variations in kernel oil of Malatya apricots from Turkey. *Journal of Agricultural and Food Chemistry, 55*(26), 10787-10794.

Varsha, R., Akash, J., Jasmine, C., Raj, V., Jain, A., & Chaudhary, J. (2012). *Prunus armeniaca* (Apricot): an overview. *Journal of Pharmacy Research, 5*(8), 3964-3966.

Vasquez-Villanueva, R., Marina, M. L., & Garcia, M. C. (2015). Revalorization of a peach (*Prunus persica* (L.) Batsch) byproduct: Extraction and characterization of ACE-inhibitory peptides from peach stones. *Journal of Functional Foods, 18*, 137-146.

Vita, F., Lucarotti, V., Alpi, E., Balestrini, R., Mello, A., Bachi, A., Alessio, M., & Alpi, A. (2013). Proteins from Tuber magnatum Pico fruiting bodies naturally grown in different areas of Italy. *Proteome Science, 11*.

Wang, A. D., Tan, D. M., Tatsuki, M., Kasai, A., Li, T. Z., Saito, H., & Harada, T. (2009). Molecular mechanism of distinct ripening profiles in 'Fuji' apple fruit and its early maturing sports. *Postharvest Biology and Technology, 52*(1), 38-43.

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science, 9*(5), 244-252.

Yea, C. S., Ebrahimpour, A., Hamid, A. A., Bakar, J., Muhammad, K., & Saari, N. (2014). Winged bean [*Psophorcarpus tetragonolobus* (L.) DC] seeds as an underutilised plant source of bifunctional proteolysate and biopeptides. *Food & Function, 5*(5), 1007-1016.

Yildirim, F. A., & Askin, M. A. (2010). Variability of amygdalin content in seeds of sweet and bitter apricot cultivars in Turkey. *African Journal of Biotechnology, 9*(39), 6522-6524.

Yildirim, F. A., Yildirim, A. N., Askin, M. A., & Kankaya, A. (2010). Total oil, fatty acid composition and tocopherol content in kernels of several bitter and sweet apricot (*Prunus armeniaca* Batsch) cultivars from Turkey. *Journal of Food Agriculture & Environment, 8*(3-4), 196-201.

Zhang, S., Wang, S., Huang, J., Lai, X., Du, Y., Liu, X., Li, B., Feng, R., & Yang, G. (2016). High-specificity quantification method for almond-by-products, based on differential proteomic analysis. *Food Chemistry, 194*, 522-528.

Part III.

Concluding Remarks

GENERAL CONCLUSIONS AND PERSPECTIVES

The present thesis offers new insights into the nutritional peptidomics, providing data on the discovery, quantification, and functional analysis of plant protein derived peptides, specifically on lupin and hempseed protein hydrolysates, soy derived peptides, and apricot kernel ones. In the meantime, it provides also an extensive study of the proteomes of hemp and apricot seed either by a functional or a descriptive point of view.

The key achievement of this PhD project was the development of a multidisciplinary strategy aimed at elucidating the composition of plant protein hydrolysates by providing both functional and structural information. The adopted strategy supports and improves the conventional approach currently used for biopeptides discovery by introducing as a specific feature *in vitro* absorption experiments using differentiated cell line such as Caco-2. This was done in order to overcome the expensive and quite long fractionation steps used in the classical downstream approach. In fact, the absorption experiments permits on one side to obtain a rough evaluation of the bioavailability and on the other greatly reduce the sample complexity, since the number of peptides detected after absorption is generally not very high. The following experiments are thus focused on a relatively small list of bioavailable peptides.

Moreover, the high complexity of protein hydrolysates usually makes very difficult the identification of single bioactive peptides. To reach the goal, our approach has adopted bioinformatics tools, such as molecular modelling to investigate the quantitative structure–activity relationship (QSAR) and *in silico* digestion.

As part of the research study, the *in vitro* absorption experiments performed on the monolayer of Caco-2 were optimized to assess cell-penetrating peptides providing the first report on the absorbability of lupin hydrolysates. Biochemical analysis have revealed that the absorption does not impair the HMGCoAR activity, but on the contrary, enriches the concentration of the most active peptides, since the absorbed mixtures are more active than the starting hydrolysates. Among the absorbable peptic peptides, **P5** (**LILPKHSDAD**) deriving from the β-conglutin precursor, was suggested as the best inhibitor of the protein-protein interaction (PPI) between PCSK9 and the LDLR by the molecular modelling study. This was confirmed experimentally, since the IC_{50} value was equal to 1.6±0.33 μM. Quantitative data by the optimization of LC-MRM methods pointed out that **P5** represents the 11.5% of the absorbed

solution. These qualitative and quantitative findings are useful to explain the beneficial effects provided by lupin protein consumption observed *in vivo* in animal and human studies.

With the aim to clarify better the fate of peptides during the gastrointestinal absorption, a study on soy glycinin-derived peptides (**IAVPGEVA**, **IAVPTGVA** and **LPYP**) revealed the existence of an *in situ* competition between the absorption and the metabolic degradation in part due to the action of brush border peptidases, such as DPP-IV expressed at the Caco-2 cell surface, which impair the absorption. However, the degradation products, **AVPGEVA**, **AVPTGVA** and as well as the shortest one, **IAVPT** and **IAVP** were also absorbed. These results open new perspectives aimed at the exploration of mechanism of transports by which these bioactive peptides are internalized. The future verification of the activities provided by the shortest fragments will be of remarkable interest in order to sort out any synergistic effects, which could increase or reduce the beneficial effect.

The proteomics investigations performed on *C. sativa* and *P. armeniaca* seeds contributed not only to mapping the seed proteome composition, but also to set-up an analytical approach useful in the quality control of food processing. The innovative approach of using the Peptide Ligand Library (CPLL) technology, integrated to the state of art techniques for sample preparation, could be applied to the further exploration of other plant species and products. The peptidomics maps of *C. sativa*, obtained through the screening and optimization of different hydrolytic conditions, revealed the possibility to obtain diverse hydrolyses with diversified activities, through the selection of the best proteolytic enzyme. In particular, our findings suggest that using pepsin it is possible to produce peptides able to mediate a hypocholesterolemic effect through a statin-like mechanism.

In conclusion, a growing body of reports on novel peptide sequences and function-structure relationships contributed to the improvement of plant protein and peptide database-platforms. Information obtained from characterizing structural components of plant hydrolysates offers useful technological and functional implications for food ingredient formulation or pharmacological use.

Appendix

ETH-Institute of Molecular Systems Biology

The latest scientific contribution to this thesis comes from the activities carried out at Institute of Molecular Systems Biology in the Aebersold group, where I spent 7 months within my PhD course. The topic of research to which I took part, was quite far from the main researches in line to my project based on food-peptidomics. However, the interest toward the new and high-throughput mass spectrometric techniques, has led me to widen my background focusing my attention on the current analytical aspects for peptide identification and quantitation in complex samples. The use of novel unbiased mass spectrometric techniques, although mainly applied to human proteome researches, has the potential to be adapted also to food component characterization, paying particular attention to both peptides identification and quantification when complex mixtures are handled.

- Recent Advances in Mass Spectrometry -

The most proteomic studies are currently geared toward the discovery or validation of differential protein regulation on a large scale in response to biological perturbations. Mass spectrometry is the method of choice to answer these biological questions. The abilities of MS technologies to identify and quantify thousands of proteins in a single shot already have a broad impact on biology and personalized medicine.[1] The different MS acquisition strategies and the type of information that can be generated through this high-throughput-based approach are actually increased and an implementation of the classical data acquisition modes has been observed in the years.

Untargeted proteomics - shotgun proteomics -

In the past decade, most of the MS based proteomic studies were carried out using **shotgun proteomics** or discovery based which employs **data dependent acquisition (DDA)**. Though widely used, DDA has a number of limitations including instrumental scanning speed,

stochastic selection of ions for fragmentation and poor repeatability [2], a relatively narrow dynamic range and the issues of chimericy (cofragmentation of two or more ions) [3,4] etc. Furthermore, the number of peptides presented in a biological digest may be many times larger than the number of ions that can be sequenced in an experiment despite the recent advances in instrumentation. Due to the bias nature of DDA for the most abundant species, low abundance peptides are unlikely being sequenced in a complex biological sample. Similarly, closely eluted isobaric species, may not be sequenced completely owing to the typical dynamic exclusion setting used in DDA.

Targeted proteomics - SRM -

In contrast to untargeted proteomics, recently, the targeted proteomics using SRM method has experienced a renaissance. Basically this approach is used to monitor a number of selected precursors and fragment transitions of targeted peptides over the LC retention time, yielding a set of chromatographic traces with the retention time and signal intensity for a distinct transition. The selection of the SRM transitions is normally calculated on the basis of the data acquired previously by product ion scanning, repository data in the public databases or based on a series of empirical rules predicting the enzymatic cleavage sites.[5]

Despite the advantages, targeted proteomics has not been the preferred method by many proteomic researchers. Given that it is a targeted approach, a prior knowledge of the targeted proteins in the sample is a requisite. Arguably, up to 6000 transitions can be monitored by an SRM experiment using triggered or intelligent selected reaction monitoring (iSRM) on the latest triple quadrupole mass spectrometer.[6] However, only a relatively small number of proteins (up to 100) are monitored by a typical SRM experiment in practice. The method also requires lengthy and labor intensive development and optimization process.[7] Furthermore, in SRM, the detection of a chromatographic peak, even with all the predicted SRM transitions detected, does not confirm the identity of the peptide. This is because the mass of interfering ions could fall within the tolerance of both quadrupoles and leads to a false positive identification and the data generated will have to be validated with relatively expensive reference or isotopically labeled peptide standards.[8]

To overcome these limitation, while achieving comparable specificity, another targeted approach was recently developed: SWATH-MS.[9]

Data Independent Acquisition (DIA) via SWATH-MS

DIA methods [10] combine certain aspects of both DDA and SRM methods. As with the targeted acquisition mode, the instrument performs continuous time-resolved acquisition of MS/MS signals for each precursor mass region. In contrast to DDA and SRM methods, however, the aim of the DIA mode is to exhaustively acquire MS/MS spectra for any possible precursor mass. Because the current instrumentation is too slow to cover the complete tryptic peptide *m/z* range (typically 300–1,200) with a small precursor isolation window in the time frame of a typical LC separation, the compromise of the DIA method is to use a much wider precursor isolation mass windows for precursor fragmentation as compared with the 1–3 Da used in SRM or DDA. Thus, a mixture of several precursor (peptide) ions are deliberately isolated and fragmented, creating multiplexed MS/MS spectra consisting of fragment ion signals from different peptides. Given sufficient acquisition speed of the instrument and appropriate window sizes, it has become possible to cover the entire informative *m/z* range of a peptide mixture and to generate exhaustive maps of all observable peptide fragments with continuous MS/MS acquisition and sufficient time resolution of the data points.[11] Despite this huge amount of data acquired with this strategy, **SWATH-MS (Sequential Window Acquisition of all Theoretical Mass Spectra)** strategy maintains its roots in the idea of targeted data analysis, whereby extracted ion chromatograms (XIC) of the most intense transitions of a targeted peptide are generated from all corresponding MS2 spectra, producing chromatographic data that is similar to SRM traces. This approach reduces the complexity of the data significantly, facilitating data analysis while retaining the complete fragment ion information of all precursors.

Figure i) reports schematically the different acquisition modes and the resulted data structures obtained by DDA, SRM and DIA approaches.

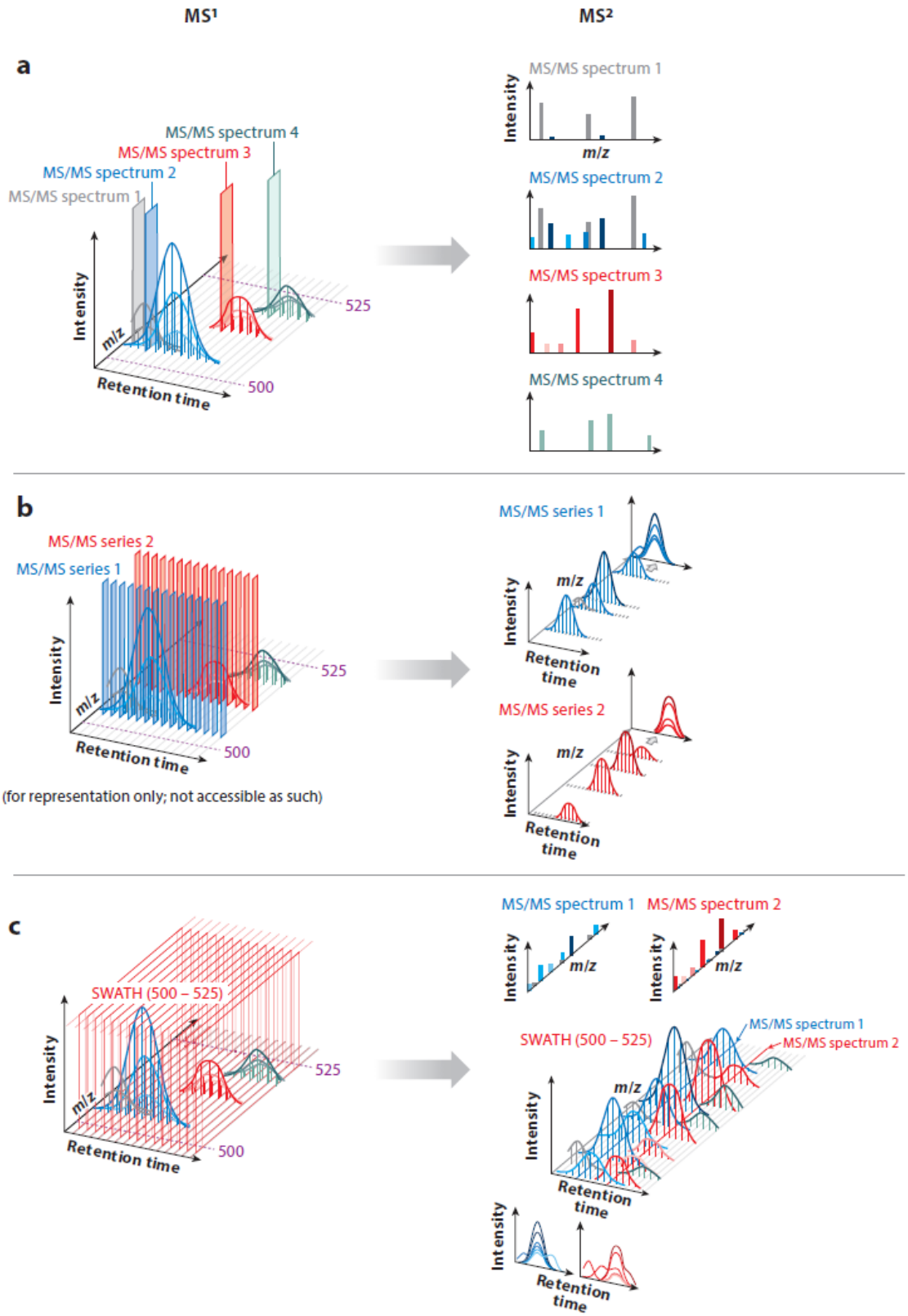

MS¹
MS²
a
MS/MS spectrum 4
MS/MS spectrum 3
MS/MS spectrum 2
MS/MS spectrum 1
Intensity
m/z
Retention time
500
525
MS/MS spectrum 1
MS/MS spectrum 2
MS/MS spectrum 3
MS/MS spectrum 4
b
MS/MS series 2
MS/MS series 1
(for representation only; not accessible as such)
MS/MS series 1
MS/MS series 2
c
SWATH (500 – 525)
MS/MS spectrum 1
MS/MS spectrum 2

Figure i) - Data structure of the LC-MS and LC-MS/MS signals acquired in the three main bottom-up proteomic acquisition strategies: **(a)** data-dependent acquisition (**DDA**), **(b)** targeted selected reaction monitoring (**SRM**), and **(c)** data-independent acquisition (**DIA**).
In **DDA**, or shotgun *(*panel *a)* the instrument is operated in iterative acquisition cycles of intact precursor-level spectra (MS, or MS1) and fragment ion spectra (MS/MS, or MS2). Decisions about which precursors to select for fragmentation are made in real time by the instrument software, according to predefined criteria such as intensity threshold. Typical instrument configurations for DDA workflows include different Orbitrap hybrids or quadrupole time-of-flight (TOF). Peptide identity is derived from the combined information of the precursor mass and the corresponding fragment ion masses (MS/MS), usually following untargeted, spectrum-centric searching strategies.[12] For **SRM** (panel *b)* predetermined fragment ion signals is repeatedly recorded for each predefined peptide over time. The lists of transitions is a prior required. The technique is best applied for consistently quantifying or validating the presence of targeted peptides, rather than discovering new peptides/proteins. Both qualitative and quantitative information is directly derived from the resulting fragment ion chromatographic signals of individual transitions. In **DIA** (panel c, left), the complete precursor mass range is selected for fragmentation in consecutive and contiguous steps (e.g., of 25 Da as shown here for region 500–525 m/z and adjacent regions), repeatedly during the LC separation, regardless of the detection or presumed presence of peptides in that mass-and-time space. For DIA, the time-and-mass-continuous MS/MS recordings can be displayed side-by-side across the LC separation (time), resulting in a three-dimensional MS2 map of the LC-MS/MS data. [11]

In the current thesis these methodology, were used to investigate the pathways of glucose metabolism alterations and tumorigenesis by comprehensively exploration of proteome. In details, SWATH-MS and targeted SRM analysis were used for the evaluation of metabolic and systematic protein profile alteration upon drug perturbation on human colon cancer cell model.

Project

Evaluation of metabolic and systematic drug perturbation on human colon cancer cell model by targeted SRM and SWATH-MS analysis

State of art

Cancer cells differ from their normal counterparts in several aspects including deregulated control of cell proliferation, ability to evade apoptotic cell regulatory signals, and modification of numerous metabolic pathways.[13] Accumulating evidence suggests that malignant transformation is associated with changes that affect several branches of metabolism in particular those associated with bioenergetics.[14]

In fact, according to the Hallmarks of Cancer proposed by Hanahan and Weinberg [15], reprogramming of energy metabolic pathways, such as glycolysis and associated anabolic pathways, is a central feature of tumorigenesis. Cancer-associated metabolic changes including (1) deregulated uptake of glucose and amino acids, (2) use of opportunistic modes of nutrient acquisition, (3) use of glycolysis/TCA cycle intermediates and (4) metabolic interactions with the microenvironment, are crucial to carcinogenesis. In particular, enhanced glycolytic flux leads to an increased synthesis of intermediates for sustaining anabolic pathways, critical for cancer cell growth and survival. In addition, the activation of an alternative pathway for glucose metabolism – the Polyol pathway – is frequently observed in colon cancer cells.[16] Possible mechanisms by which this metabolic alteration may evolve during cancer development have been proposed. The mechanisms include (1) mitochondrial defects, (2) adaptation to hypoxic environment in cancer tissues, (3) oncogenic signals, and (4) abnormal expression of certain metabolic enzymes. In this scenario, the increased dependence of cancer cells on direct glycolytic pathway and/or on its associated pathways provides a biochemical basis for the design of therapeutic strategies using pharmacological inhibition of glycolysis and anti-cancer drugs in order to identify the principal activated routes during tumorigenesis and cancer progression.

Aim of the study

In this study, targeted SRM and SWATH/DIA mass spectrometric approaches were used to investigate the modulation of the polyol pathway and/or glycolysis induced by 1) carbon source-environment perturbations, and 2) therapeutic metabolic inhibitors on the human colon cancer cell line, HT29.

In details, the cellular proteome was probe for:
(1) The effect of glycolytic inhibitors on cellular metabolism in wild-type colon cancer cells.
(2) The effect of different growth conditions: nutrient availability from carbon sources, such high sugar environment (glucose, sorbitol, fructose), could differentially affect the utilization of the polyol pathway in colon cancer cells.
Model-based statistical analysis was employed to identify significantly dysregulated proteins and networks upon metabolic and systematic drug perturbation.

The overall workflow applied is reported in **Figure ii.**

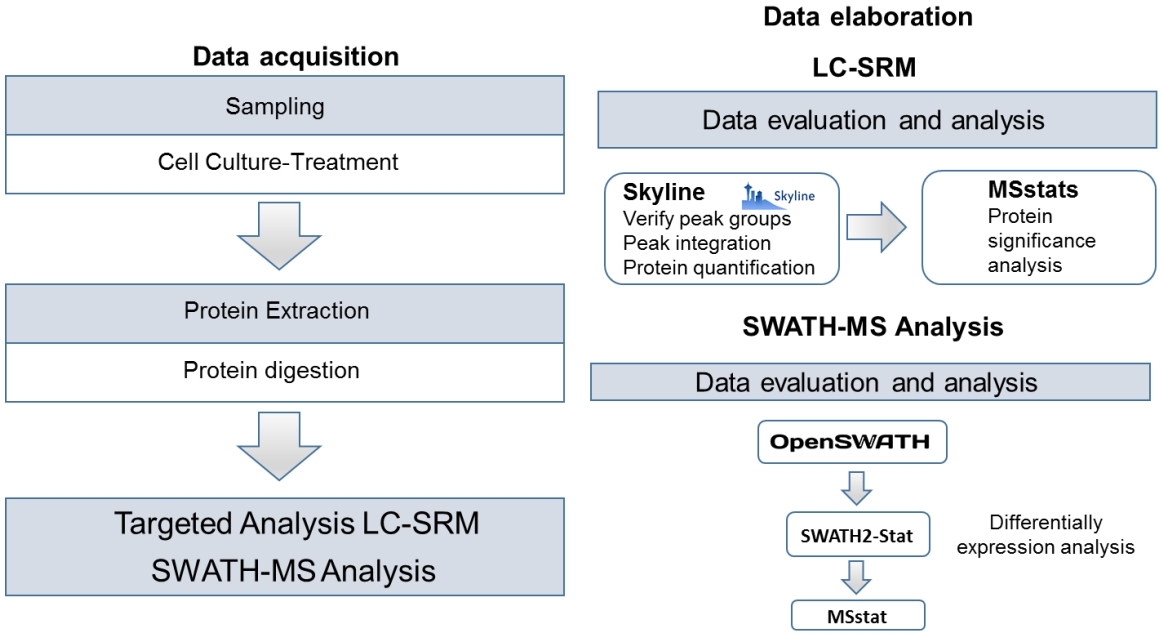


Figure ii) - Methodological approaches adopted.

Methodologies

SRM assay development and time-scheduled SRM analysis. Biologically relevant target proteins involved both in the glycolysis and in the pathways which may be triggered by the inhibition of glycolysis or polyol pathway, were selected. SRM assays were design to quantify 48 proteins covering glycolysis, polyol pathway, tricarboxylic acid cycle, pentose phosphate pathway, hexose metabolism and gluconeogenesis. Moreover, proteins of the different branches of the glycolysis, such as those involved in amino-acid biosynthesis and degradation, in isoprenoid biosynthesis and in the fermentation of pyruvate to lactate were also included in the final set of targeted proteins.

Development and validation of SRM assays to measure abundances of proteins was guided by a "human" spectral library previously generated in-house through a LC-MS/MS analysis. SRM assays were developed following the general high-throughput strategy previously reported.[17] Initially, 4–6 unique peptides ranging from 6 to 20 amino acids in length, containing tryptic ends and no missed cleavages, were chosen for each of the selected proteins. All peptides containing amino acids prone to undergo unspecific reactions (e.g., Met, Trp, Asn and Gln) were generally avoided and only selected when no other options were available. Proteotypic peptides matching to the spectral library were ranked by intensity using Skyline and the most intense transitions consisting of doubly or triply charged precursor ions and singly or doubly charged fragment ions of the *y*- and *b*- ion series, with the restrictions listed above, were experimentally tested in SRM mode to select the most suitable transitions for quantification experiments. Time-scheduled SRM methods was used to monitor the target proteins. In time scheduled SRM mode, the full cycle time is used to detect and quantify peptides expected to elute within a given retention-time window. This restricts the acquisition of defined transitions to a window around the elution time of the corresponding peptide. Retention time scheduling for all transitions was selected based on Skyline scheduling predictions. A window of ± 2.5 min was used to schedule all transitions. Overall, 635 transitions were selected to monitor the selected proteins.

Measurements and data analysis. SRM analysis was conducted on TSQ Vantage instrument equipped with a nano-electrospray ionization source and coupled to a Nano-LC Ultra 2D HPLC system. Q1 and Q3 operated at unit resolution (0.7 *m*/*z* half-maximum peak width) with a cycle

time <2 s. Chromatographic separation of peptides was performed on a frit column (150 mm x 75 µm) made in-house and coupled to a fused silica emitter (100 mm x 75 µm). The columns were packed with reverse-phase C18 material. The peptides were eluted using a linear gradient of acetonitrile/water containing 0.1% formic acid at a flow rate of 300 nL/min. An elution gradient of 5% to 35% acetonitrile over 60 min was used.

Statistical analysis of data from proteome profiling. SRM data was statistically analyzed to detect differences between the different concentrations of inhibitors or sugars tested. The input files for this analysis contained values for Condition, Bioreplicate, and Run, as pre-assigned in Skyline according to the experimental design. All input files were elaborated through MSstats, R-package in R statistical software. All transition intensities were transformed into log2 values.

MSstat

The MSstat analysis has provided the following steps as reported in **Figure iii.**

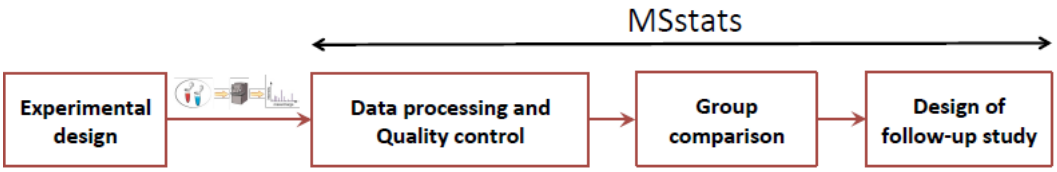


Figure iii) - MSstats pipelines

In details, **Data processing and Quality control** analysis included the Intensity Normalization, the Quality Control plotting, Features selection and Removing Interferences. The last steps concerned the plotting profile in order to visualize the potential source of variation such as Run, Transition, and Conditions. At least, **Group comparison** allowed us to test proteins for differential abundance. **Log2 fold change** was used to provide an estimation of the differences between two conditions. Fold change cut-off of 1.5 was here used. The results are constrained with an **adjusted p-value** cut-off of 0.05. The results are then visualised through Volcano Plots and Heat maps.

SWATH-MS analysis workflow

Automated analysis of SWATH-MS data

In parallel, all digested peptides were measured on the 5600+ TripleToF from Sciex in DIA mode and analysed by OpenSWATH-workflow, an open-source-cross-platform, which allows targeted analysis in an automated mode.

The steps performed by the OpenSWATH software during SWATH data analysis are reported in **Figure iv.**

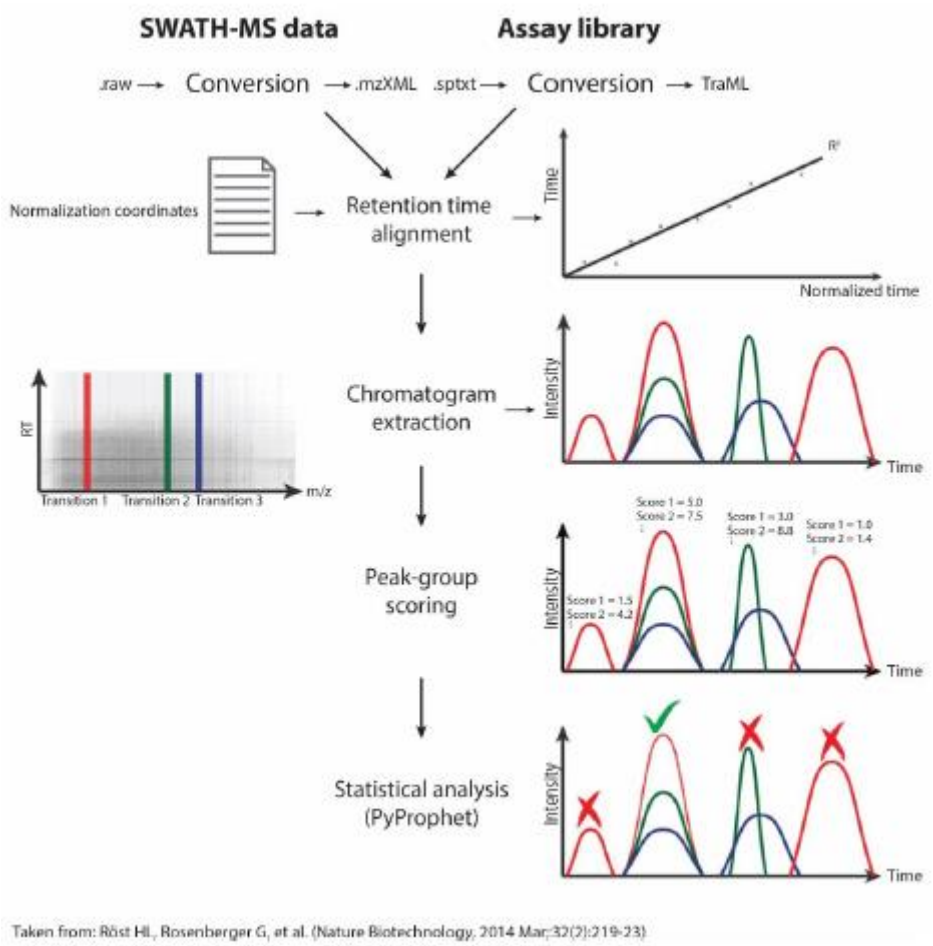


Figure iv) - Steps performed by the OpenSWATH software during SWATH data analysis. The graphic illustrates a peptide precursor with three fragment ion transitions, which are shown in red, green and blue. The graph on the left side shows the three transitions in a mass to charge ratio versus retention time graph. Over the whole retention time a specific mass is extracted. This is a so called transition. The four graphs on the right side illustrate the OpenSWATH analysis step. On the upper part the inputs and the formats of the input files are listed, which are necessary for the OpenSWATH pipeline. For the analysis, the raw data were converted to the mzXML data format. The OpenSWATH analysis required a SWATH assay library. This library was generated in advance from DDA analysis. In the first step the acquired MS data were aligned according to the retention times of the iRT-peptides. In the second step, the transitions for the fragment ions were extracted by making use of the assay library. Afterwards, the

extracted peak-groups were scored by several algorithms. The scores considered among other attributes how well the transitions aligned in terms of retention time, the peak shapes and peak intensities. The PyProphet algorithm weighted the different scores and calculated a single score for each assay. As the library also consisted of decoy peptide sequences further filtering by an FDR was conducted by the OpenSWATH pipeline.

However, in order to conduct a targeted quantitative analysis of **SWATH-MS** data, a spectral assay library, containing all peptides acquired in a sample, was required. The spectral library used as input for peptide queries in the OpenSWATH analysis was a previously published as proteome library containing mass spectrometric coordinates for 10,000+ human proteins built by combining several hundred DDA analyses of various human cells and tissues types. iRT standard peptides (Biognosys) were included in the library for automatic retention time calibration of each different sample set with the ion library retention times. Basically, the complete set of coordinates present in the library describes (i) which peptides of a protein are most representative, i.e. unique and well detectable, (ii) the elution time of these peptides from the LC, (iii) their pre-dominant charge state, (iv) the most abundant fragment ions formed during fragmentation and (v) their relative intensity.

The OpenSWATH pipeline consists of the OpenSWATH software [18], for identifying and extracting quantitative data from targeted peptides within the fragment ion maps and for assessing a statistical assessment of the correct identification of these targeted peptides using the mProphet algorithm.

The algorithm used by OpenSWATH platform were basically summarized in the following steps:

1) Assay libraries importing. **2)** Data conversion. **3)** Retention-time alignment. **4)** Chromatogram extraction and identification of the peak groups. **5)** Peak-group scoring. **6)** Statistical analysis.

Data Conversion. The acquired SWATH-MS data together with the assay library comprise the input data were converted to suitable open file formats (mzML and TraML).

Retention-time alignment. Each run was aligned against a previously determined normalized retention-time space using reference peptides whose mappings to the normalized space was known (for example, spiked-in peptides). Outlier detection is subsequently applied to remove wrongly assigned reference peptides and to evaluate the quality of the alignment.

Chromatogram extraction. Using the *m/z* and retention time information from the assay library, the workflow extracts ion chromatograms from the corresponding MS2 map, producing integrated fragment ion count vs. retention time data.

Peak-group scoring. For the chromatogram extraction scoring, PyProphet [19], an improved version of the mProphet [20] algorithm, was used.
The core algorithm identifies 'peak groups' (that is, positions in the chromatograms where individual fragment traces coelute) and scores them using multiple, orthogonal scores. These scores are based on the elution profiles of the fragment ions, the correspondence of the peak group with the expected retention time and fragment-ion intensity from the assay library, as well as the properties of the full MS/MS spectrum.
PyProphet considered extrinsic scores (e.g. Retention time deviation of the peak-groups compared to the expected retention time of the spiked-in iRT-peptides. Intensity correlation with the intensity of the peak-group in the spectral library) and intrinsic scores (e.g. Co-elution score for precursor and transitions, peak shape score for precursor and fragments, intensity of the precursor and transitions).

Statistical analysis: SWATH2stats and MapDIA. For downstream statistical and quantitative analysis, MapDIA (Model-based Analysis of Quantitative Mass Spectrometry Data in Data Independent Acquisition Mode) tool was applied by the use of SWATH2stats R-package

SWATH2stats package allowed further visualization of the FDR thresholds and to transform the data to a format which was suitable for the mapDIA analysis.[21] The latter provides hierarchical model-based statistical significance analysis for multi-group comparisons under representative experimental designs, providing reliable classification of differentially expressed proteins with accurate control of the false discovery rates.

In details, the mapDIA workflow consists of three major steps: intensity normalization, peptide/ fragment selection, and statistical analysis. First, the fragment-level intensities by total intensity sums as well as by local intensity sums in retention time space were normalized. Second, the outliers were removed and peptides/ fragments that preserved the major quantitative patterns across all samples for each protein were selected. Last, using the selected fragments and peptides, mapDIA performed model-based statistical significance analysis of protein-level differential expression between specified groups of samples.

Based on the methodologies illustrated, the proteome changes upon cellular perturbation either induced by glycolytic inhibitors or by sugar environment alterations, were investigated, however, the institute holds the intellectual property rights, since the results are not jet disclosed.

References

1. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.

2. Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M., Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (14), 5860-5865.

3. Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M., Quantifying the Impact of Chimera MS/MS Spectra on Peptide Identification in Large-Scale Proteomics Studies. *Journal of Proteome Research* **2010,** *9* (8), 4152-4160.

4. Luethy, R.; Kessner, D. E.; Katz, J. E.; McLean, B.; Grothe, R.; Kani, K.; Faca, V.; Pitteri, S.; Hanash, S.; Agus, D. B.; Mallick, P., Precursor-ion mass re-estimation improves peptide identification on hybrid instruments. *Journal of Proteome Research* **2008,** *7* (9), 4031-4039.

5. Colangelo, C. M.; Chung, L. S.; Bruce, C.; Cheung, K. H., Review of software tools for design and analysis of large scale MRM proteomic datasets. *Methods* **2013,** *61* (3), 287-298.

6. Kiyonami, R.; Schoen, A.; Prakash, A.; Peterman, S.; Zabrouskov, V.; Picotti, P.; Aebersold, R.; Huhmer, A.; Domon, B., Increased Selectivity, Analytical Precision, and Throughput in Targeted Proteomics. *Molecular & Cellular Proteomics* **2011,** *10* (2).

7. Mortstedt, H.; Karedal, M. H.; Jonsson, B. A. G.; Lindh, C. H., Screening Method Using Selected Reaction Monitoring for Targeted Proteomics Studies of Nasal Lavage Fluid. *Journal of Proteome Research* **2013,** *12* (1), 234-247.

8. Mirzaei, H.; McBee, J. K.; Watts, J.; Aebersold, R., Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics. *Molecular & Cellular Proteomics* **2008,** *7* (4), 813-823.

9. Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* **2012,** *11* (6).

10. Chapman, J. D.; Goodlett, D. R.; Masselon, C. D., Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom Rev* **2014,** *33* (6), 452-70.

11. Gillet, L. C.; Leitner, A.; Aebersold, R., Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu Rev Anal Chem (Palo Alto Calif)* **2016,** *9* (1), 449-72.

12. Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J., Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol Cell Proteomics* **2015,** *14* (9), 2301-7.

13. Amoedo, N. D.; Obre, E.; Rossignol, R., Drug discovery strategies in the field of tumor energy metabolism: Limitations by metabolic flexibility and metabolic resistance to chemotherapy. *Biochim Biophys Acta* **2017,** *1858* (8), 674-685.

14. Galluzzi, L.; Kepp, O.; Vander Heiden, M. G.; Kroemer, G., Metabolic targets for cancer therapy (vol 12, pg 829, 2013). *Nature Reviews Drug Discovery* **2013,** *12* (12), 965-965.

15. Hanahan, D.; Weinberg, R. A., Hallmarks of cancer: the next generation. *Cell* **2011,** *144* (5), 646-74.

16. Uzozie, A. C.; Selevsek, N.; Wahlander, A.; Nanni, P.; Grossmann, J.; Weber, A.; Buffoli, F.; Marra, G., Targeted Proteomics for Multiplexed Verification of Markers of Colorectal Tumorigenesis. *Mol Cell Proteomics* **2017,** *16* (3), 407-427.

17. Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R., High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* **2010,** *7* (1), 43-6.

18. Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R., OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **2014,** *32* (3), 219-23.

19. Teleman, J., DIANA-algorithmic improvements for analysis of dataindependent acquisition MS data *Bioinformatics* **2015,** *31* (4), 555-562.

20. Reiter, L.; Rinner, O.; Picotti, P.; Hüttenhain, R.; Beck, M.; Brusniak, M. Y.; Hengartner, M. O.; Aebersold, R., mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* **2011,** *8* (5), 430-5.

21. Blattmann, P.; Heusel, M.; Aebersold, R., SWATH2stats: An R/Bioconductor Package to Process and Convert Quantitative SWATH-MS Proteomics Data for Downstream Analysis Tools. *PLoS One* **2016,** *11* (4), e0153160.

Scientific Publications & Communications

Published Article	Aiello G., Lammi C., Boschin G., Zanoni C., Arnoldi A. Exploration of potentially bioactive peptides generated from the enzymatic hydrolysis of hempseed proteins. *J. Agric Food Chem. DOI: 10.1021/acs.jafc.7b03590*
Published Article	Zanoni C., Arnoldi A., Aiello G., Lammi C. Hempseed Peptides Exert Hypocholesterolemic Effects with a Statin-Like Mechanism. *J. Agric. Food Chem.*, **2017**, 65 (40), pp 8829–8838
Published Article	Ghorab H., Lammi C., Arnoldi A., Kabouche Z., Aiello G. Proteomic analysis of sweet algerian apricot kernels (Prunus armeniaca L.) by combinatorial peptide ligand libraries and LC-MS/MS. *Food Chem.* **2018** Jan 15; (239) 935-945
Published Article	Zanoni C., Aiello G., Arnoldi A., Lammi C. Investigations on the hypocholesterolaemic activity of LILPKHSDAD and LTFPGSAED, two peptides from lupin β-conglutin: Focus on LDLR and PCSK9 pathways. *J. Funct. Foods* **2017**, (32), 1-8
Published Article	Ruscica M., Pavanello C., Gandini S., Gomaraschi M., Vitali C., Macchi C., Aiello G., Bosisio R., Calabresi L., Arnoldi A., Sirtori C.R., Magni P. Effect of soy on metabolic syndrome and cardiovascular risk factors: Evidence from a randomized controlled trial. *Nutr Metab Cardiovasc Dis*. **2017**, (27) 1, 35
Published Article	Lammi C., Zanoni C., Aiello G., Arnoldi A., Grazioso G. Lupin Peptides Modulate the Protein-Protein Interaction of PCSK9 with the Low Density Lipoprotein Receptor in HepG2 Cells. Scientific Reports 6:29931 · Scientific Reports 6, Article number: 29931, **2016**
Published Article	Aiello G., Fasoli E., Boschin G., Lammi C., Zanoni C., Citterio A., Arnoldi A. Proteomic characterization of hempseed (*Cannabis Sativa L.*) *J. Proteomics* **2016,** 147, 187-196
Published Article	Lammi C., Aiello G., Vistoli G., C. Zanoni C., Arnoldi A., Sambuy Y., Ferruzza S., Ranaldi G. A multidisciplinary investigation on the bioavailability and activity of peptides from lupin protein. *J. Funct Foods* **2016**, 24:297-306

Published Article	Napoli A., Aiello D., Aiello G., Cappello M.S., Di Donna L., Materazzi S, Fiorillo M., Sindona G. Mass Spectrometry-based protomic approach in *Oenococcus oeni* (*O. oeni*) enological starter" *J Proteome Res* **2014**, 13 (6), 2856-66.
Manuscript in preparation	Aiello G., Ranaldi G., Zanoni C., Ferruzza S., Sambuy Y., Arnoldi A., Lammi C. Absorption and metabolism of soy peptides across Caco-2 cell monolayers.
Abstract in Conference Proceedings	Aiello G., Lammi C., Zanoni C., Boschin G., Arnoldi A. A mass spectrometric assay for the quantitation of peptide LILPKHSDAD from Lupin β-conglutin absorbed through differentiated human enterocytes. The 1st Food Chemistry Conference. Amsterdam (31 October – 2 November **2016**).
Abstract in Conference Proceedings	Aiello G., Fasoli E., Boschin G., Lammi C., Zanoni C., Citterio A., Arnoldi A., Proteomic characterization of hempseed using combinatorial peptide ligand libraries and mass spectrometry. The 1st Food Chemistry Conference. Amsterdam (31 October – 2 November **2016**).
Abstract in Conference Proceedings	Aiello G., Boschin G., Arnoldi A. "Extensive proteome characterization of L. mutabilis using a combined approach based on 2D-electrophoresis and mass spectrometry". XIV Lupin Conference (21-26 **2015** June) Milan.
Oral Communication	Aiello G., Uzozie A.C., Aebersold R. Evaluation of metabolic and systematic drug perturbation on human colon cancer cell model by targeted SRM and SWATH-MS analysis. Proteomics in Drug Discovery. Milan, 11th May **2017**.
Oral Communication	Aiello G., Lammi C., Arnoldi A. Proteomic Investigation of sweet Algerian apricot kernel (Prunus Armeniaca L.) by n-LC-MS/MS coupled to combinatorial peptide ligand libraries. Milan, 25th September **2017**.
Oral Communication	Aiello G. A multidisciplinary investigation for bioavailability, functional analysis and quantification of lupin peptides through Caco-2 cell monolayers. Book of Abstracts of International Meeting on SSPA – 22end Summer School in Pharmaceutical Analysis. Rimini, 18-20 September **2017**.