University of Catania

DEPARTMENT OF CHEMICAL SCIENCES

Ph.D. Thesis

Proteomic Investigation of chickpea seeds: the case study of the Italian genotype Pascià

Aldo Lanzoni

Tutor

Prof. Vincenzo Cunsolo

Ph.D. Coordinator

Prof. Salvatore Scirè

XXXVIII CYCLE PhD PROGRAMME

INDEX

LIST OF ABBREVIATIONS

AMBIC = Ammonium bicarbonate

CID = Collision-induced dissociation

CPR= Cysteine Proteinase

DF= Dietary Fiber

DFC= Dietary Fiber Content

DTT = Dithiothreitol

ECD = Electron capture dissociation

ER=Endoplasmatic Reticulum

ETD = Electron transfer dissociation

FDR = False Discovery Rate

FT-ICR = Fourier-transform ion cyclotron resonance

HCD = High-energy collision dissociation

IAA = Iodacetamide

LFQ= Label-Free Quantification

MudPIT = Multi-dimensional Protein Identification Technology

MUFA= Monounsaturated Fatty Acids

PFF = Peptide-Fragmentation Fingerprinting

PMF = Peptide Mass Fingerprinting

PSV= Protein Storage Vacuole

PTM = Post-translational modification

PUFA= Polyunsaturated Fatty Acids

SFA= Saturated Fatty Acids

VPE= Vacuolar Processing Enzymes

Introduction

Legumes belong to the family *Leguminosae* and are known as one of the most important groups of plant crops, after cereals, in human nutrition according to FAO classification.[1] This is mainly because they are inexpensive to produce and purchase, and have high nutritional properties and beneficial physiological effects.[2] An example of the most known and cultivated types of legumes include: peas (*Pisum sativum*), chickpeas (*Cicer arietinum* L.), lentils (*Lens culinaris*), soybeans (*Glycine max*), cowpeas (*Vigna unguiculata*), bambara groundnuts (*Vigna subterranea*), pigeon peas (*Cajanus cajan*), jack beans (*Canavalia ensiformis*), lupins (*Lupinus albus*), mung beans (*Vigna radiata*), yellow peas (*Lathyrus aphaca*) and black-eyed peas (*Vigna unguiculata* subsp. *unguiculata*), peanuts (*Arachis hypogaea*), dry beans (*Phaseolus vulgaris* L.), broad beans (*Vicia faba*), and green beans (*Phaseolus vulgaris* L.).[3]

One of the most important characteristics of legumes is that they are drought-tolerant, adaptable to marginal soils and climates for cultivation, and can absorb the essential nutrients like nitrogen from the air and fix it to the soil.[4,5] The capacity of fixing nitrogen makes legumes a very important crop in tropical areas, where most soils do not contain sufficient nitrogen. This makes legume cultivation an inexpensive and simple way for farmers to enrich the soil [6] and reduce their costs for artificial fertilizer.

Nowadays, legumes can also emerge as a suitable crop having prospects for the future, because of increasing levels of pollution, global temperature, and loss of water sources. Climate change is likely to exacerbate regional and global water scarcity to a great extent. It is estimated that the two third of the world population is already experiencing severe water scarcity.[7] Hence, legumes play an important role in providing ecosystem services. Thanks to their ability to fix atmospheric nitrogen, develop deep root systems, require relatively little water, and tolerate poor soils and environmental stresses, legumes represent a valuable option for farmers both in intensive cereal-based systems and in the arid or semi-arid regions where growing conditions are more challenging.[8]

Above all these traits, legumes stand out for their most important characteristic: they are highly nutritious foods with significant benefits for human health. In particular,

Legume seeds are gluten-free and contain many other desirable properties and nutritional benefits, such as the ability to prevent chronic disease (cardiovascular disease, cancer, diabetes, osteoporosis, hypertension), reduce LDL cholesterol, and modulate intestinal microflora.[9] Legumes are also an excellent source of fiber, which contributes to lowering the energy density and reducing the glycaemic response, but they are also a good source of protein. A serving of legumes contains 2–4 g of fibre and 7–8 g of protein. Most beans are very low in fat, generally containing <5% of energy as fat.[10] They contain substantial amounts of the B vitamins as well as the nutritionally important minerals, such as iron, calcium, and potassium. It is the relative proportions of these nutrients, especially protein and fiber, and the composition of the individual components, which, for the most part, determine the nutritional value of legumes.[11]

Regarding the protein fraction, legumes also play an important role in the pursuit of alternative proteins. They possess unusually high protein content compared to other plants, and contain most of the essential amino acids. Therefore, legume-derived proteins present the opportunity to replace some existing proteins from animal sources. Legume-derived proteins are abundant, relatively low-cost, sustainable, not highly allergenic, and widely acceptable. Their techno-functional attributes make them feasible for use in various food systems and applications, in their natural forms as well as their modified forms. Some food products based on legume-derived proteins are already available on the market. Some examples of legume-derived protein incorporation in food systems are gluten-free food,[12] legume-based yogurt-like,[13] and beverages.[14] In light of the considerable potential of legume-derived proteins, research on their use in additional food applications should be expanded.

Taking into account the impressive nutritional composition, legumes play a key role in addressing protein-energy malnutrition and micronutrient deficiencies. Looking ahead, currently, research should focus on the use of modified legume proteins as functional food ingredients, the extraction and analysis of legume proteins for their functional properties, and potential applications in households or industry. Considering the significance of legume proteins across multiple fields, a comprehensive study of these components, the different families they comprise, their biosynthetic pathways, and their unique characteristics and properties is necessary.

Legume Protein Fraction

Legumes, during their growth, accumulate large amounts of proteins, [15] that can be classified based on their sequence, structure, function, conserved motifs, biological activity, solubility, and sedimentation coefficient[16] (refer to Figure 1). In terms of biological activity, proteins can be primarily categorized into two types:

- *Storage proteins,* which constitute approximately 70% of the entire protein fraction, and represent a reserve of free amino acids, as well as of ammonia and carbon.[17]
- *Metabolic proteins* (*enzymatic*, *regulatory, and structural proteins*), including components responsible for the normal cellular functions, such as protease and amylase inhibitors, lectins, lipoxygenase, defense proteins, and many others, but also proteins involved in the synthesis of the storage proteins.[18]

Storage proteins can be further classified as seed storage proteins and vegetative storage proteins. Vegetative storage proteins are components that accumulate in vegetative tissues, such as leaves and stems, depending on the specific plant species. These proteins serve as a temporary reservoir of amino acids for subsequent growth and development phases.[19] They have crucial importance because they determine not only the total protein content of the seed but also its quality for various end uses. Seed storage proteins typically exhibit considerable variation in their specific molecular structures, despite sharing several fundamental characteristics. Firstly, their synthesis occurs at high levels within particular tissues and during defined stages of seed development. Notably, many of these proteins also contain the sulfur-containing amino acids cysteine and methionine, making sufficient sulfur availability, which is a critical requirement for their biosynthesis. In several seed types, distinct groups of storage proteins coexist, some enriched in sulfur amino acids, others deficient, potentially enabling the plant to sustain storage protein production under fluctuating sulfur conditions. Due to their abundance and economic importance, seed storage proteins were among the earliest of all plant proteins to be characterized and classified. However, the detailed study of seed storage proteins dates from the turn of the century, when Osborne (1924) classified them based on their extraction and solubility.[20] Particularly, seed storage proteins could be further

subdivided according to their solubility[21] into four subgroups (Figure 1): globulins (55-60%), soluble in salt solution, glutelins (18-25%), [22] soluble in acid or alkali solution, albumins (8–14%), soluble in water, and prolamines (3-7%),[23-24] soluble in alcohol solution. Glutelins (cereal-like proteins) are soluble in diluted acid or alkali solutions and represent about 18–25% of the total protein fraction. Glutelins contain high levels of methionine and cysteine residues and are therefore important from a nutritional point of view. However, currently, this class of chickpea proteins has not been well-investigated. In legume flours, 2S albumins, prolamins, and globulins are historically classified as seed storage proteins, with the latter constituting the predominant components.[25-26] Nevertheless, it is acknowledged that possible molecular overlaps may exist between globulins and glutelins, because they share structural and evolutionary relationships; however, from this point onward, these two classes of proteins, in the present PhD thesis, will be referred to according to their classical definition, as reported in Osborne's classification.

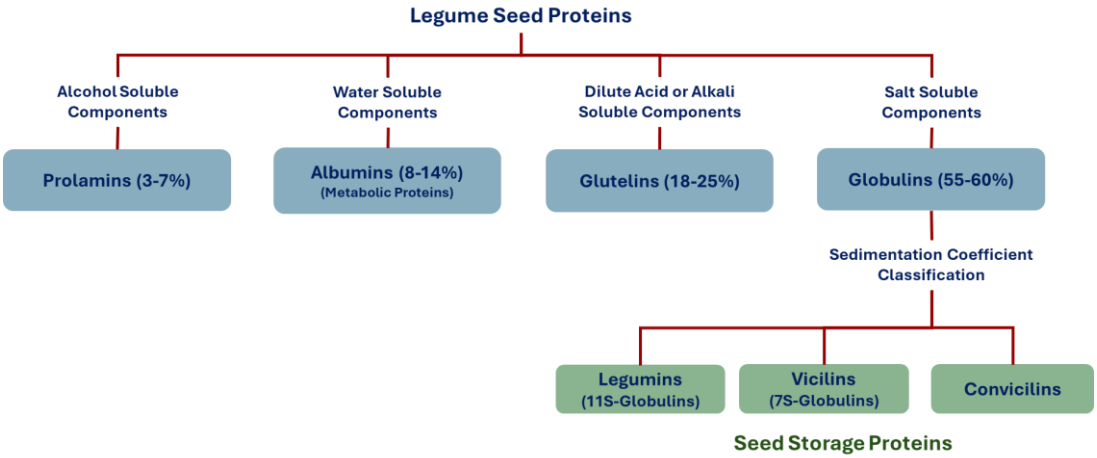


***Figure 1**. Cataloguing of Legume proteins according to their solubility (Osborne's Classification)*

Legume Seeds Storage Proteins: Structures and Biosynthesis

These proteins represent the most abundant proteins in legume seeds and are synthesized by specific biosynthesis pathways.

2S Albumins and Prolamins

2S albumins were initially defined as a group based on their sedimentation coefficient **(**Svedberg Units = S).[27] They are widely distributed in dicot seeds and have been most widely studied in the *Cruciferae*, notably oilseed rape (in which they are called napins) and *Arabidopsis*. They are usually synthesized as single precursor proteins that are proteolytically cleaved with the loss of a linker peptide and short peptides from both the N- and C-termini.[28] This appears to be the most typical 2S albumin structure: similar heterodimeric proteins are present in species like pumpkin,[29] cotton,[30] castor bean,[31] and lupin.[32] The presence of two interchain bonds has been directly demonstrated in 2S albumins from lupin. Variant types of 2S albumin also occur, but, despite differences in their subunit structure and synthesis, all the 2S albumins are compact globular proteins with conserved cysteine residues.

Prolamins are a class of proteins traditionally recognized as a group based on their solubility in alcohol/water mixtures. The name prolamins was used by Osborne (1924) to reflect the high contents of proline and glutamine in cereal proteins.[33] In wheat, the prolamins are called gliadins (based on the name originally proposed in 1819),[34] while in other cereals, they have trivial names based on their Latin generic names: zein in maize (*Zea mays*), hordein in barley (*Hordeum vulgare*), and secalin in rye (*Secale cereale*). The alcohol-insoluble fraction of wheat gluten is called glutenin, and the corresponding fractions from other species, like legumes, are classified as glutelins. Prolamins are deposited in discrete protein bodies in the developing starchy endosperm and have no known function apart from the storage role. Prolamin fractions also consist of many individual proteins, which can usually be classified into a small number of groups or families. Even today, the prolamine class is little studied in organisms such as legumes.[33]

Globulins

Globulins are the most widely distributed group of storage proteins; they are present not only in dicotyledons but also in monocotyledons (including cereals and palms) and fern spores.[35] They are subdivided into two groups based on their sedimentation coefficients (i.e., Svedberg coefficient): the 7S and the 11S. The 11S and 7S

globulins are named legumin-like or legumins and vicilin-like or vicilins, respectively, by the predominant storage globulins in pea *(Pisum sativum* L.) and faba beans *(Vicia faba* L.). Trivial names, which have frequently been derived from the botanical name of the corresponding plant, are in use for many of these storage proteins. Convicilins represent the third minor group of globulins. They show an Mr of about 70 kDa and are often found to be trimers of about 210 kDa constituted by three convicilin molecules, or heteromeric trimers of convicilin and vicilin.

11S legumin-like proteins. 11S legumin-like globulins are typically known to have a molecular weight of 300-400 kDa in their mature form, which consists of six nearly identical subunits, with molecular weights each weighing around 50-60 kDa, that interact noncovalently. Each subunit is composed of two different polypeptide chains, which differ in size (Figure 2). The larger, more hydrophilic one (MW 30-40 kDa) has a weak acidic p*I* and is referred to as α-chain, whereas the smaller, named β-chain (MW 20 kDa) is more hydrophobic and has a strong basic p*I*. These two polypeptide chains are linked together by a disulfide bond between two cysteine residues whose positions are highly conserved in both α- and β-chains. The amino acid sequences of β-chains are more homogeneous than those of the α-chains, which vary considerably in length because of different numbers of repeats in their C-terminal part. [36-37] There is no indication that the legumin holoprotein (a simple protein to which a nonprotein group, such as a carbohydrate or lipid group, is attached) molecules represent homo-oligomers. In the quaternary structure of the protein, two trimers are layered upon and twisted against each other at an angle of 60°. The hydrophobic β-chains are mostly buried inside the protein. The more hydrophilic α-chain is mainly located at the surface and forms loops that extend out of the protein. The major loop is formed by the variable and hydrophilic C-terminal region of the α-chain. Legumin genes and their corresponding mRNAs encode the individual subunits, meaning the initial translation product represents a single subunit. This large subunit has a transient N-terminal signal peptide, and the α- and β-chains are linked by a peptide bond between the C-terminal amino acid residue of the α-chain, which is always an asparagine, and the N-terminal residue of the β-chain, which is almost always a glycine (Figure 2). The disulfide bridge that binds both the α- and β-chain regions of the legumin precursor is already formed when the growing pre-prolegumin is located in the lumen of the endoplasmatic reticulum (ER),

into which is inserted during its biosynthesis. The latter belongs to a specific pair of disulfide bridges that are highly conserved in this type of legume proteins as previously reported in soybean 11S proglycinin by Adachi et al,[38] in which study, they highlight the fact that two disulfide bonds are highly conserved in 11S legumin like proteins, the first is an intra-chain bond between the Cys12 and Cys45 and the second is an inter-chain bond between Cys88 and Cys298. Then, when the signal peptide is co-translationally detached, a new prolegumin is generated. Prolegumin usually assembles into trimers and represents the transport-competent molecular form of legumin precursors. The peptide linkage between α- and β-chains is only proteolytically cleaved after the prolegumin trimers have reached the storage vacuole, where the proteins are stored until the germination of the seeds. Since both chains have already been bridged by the disulfide linkage, α- and β-chains, which are encoded by one gene, remain paired in the subunit. Notably, prolegumins do not appear to be able to assemble directly into hexamers. Cleavage of the Asn-Gly bond is the crucial prerequisite for the transition of the two trimers into one hexamer. This transformation occurs without subunit disassembly and is facilitated by a highly specific single proteolytic cleavage.[39]

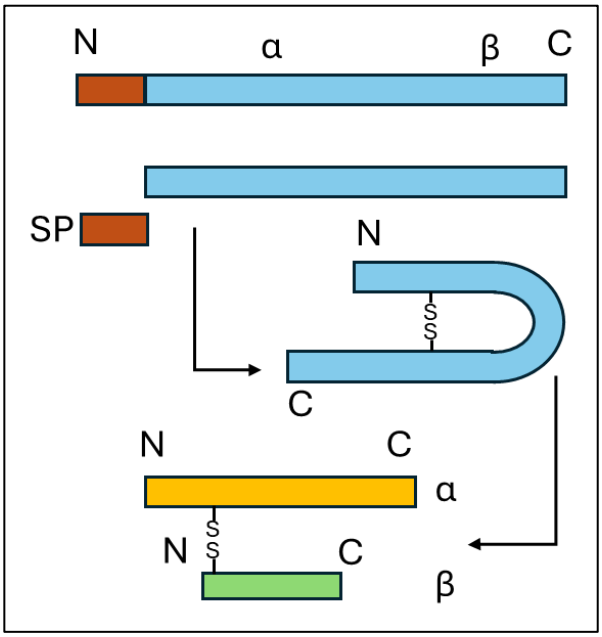


***Figure 2**. Processing of Legumin-like storage proteins by limited proteolysis during seed maturation and the formed polypeptides. N and C indicate the N- and C-terminus of the polypeptides; the transient N-terminal signal peptide (SP) is similarly detached in all cases.*

7S Vicilin-like globulins. 7S vicilin-like globulins are typically trimeric proteins (MW from 150 to 190 kDa) and, unlike other globulins, they are reported in the literature as lacking cysteine residues and hence cannot form disulfide bonds. Their subunit compositions vary considerably due to differences in post-translational modifications, including proteolysis and glycosylation. Homo-oligomeric as well as hetero-oligomeric trimers are known,[40] the latter composed of different ratios of large and small subunits. The large subunits contain a hydrophilic insertion near the N-terminus consisting of repetitive sequences, and have a molecular weight of about 20 kDa. The remaining major C-terminal portion, with a MW of about 50 kDa, exhibits high sequence similarity to the small subunits.[41] The small subunits are polymorphic and differ in two structural characteristics: the degree of glycosylation and the presence of sites for limited proteolysis by trypsin-like enzymes. Some vicilins, such as the major 50 kDa subunit found in *Vicia faba* [42-43] and the 70 kDa convicilin subunit from peas (*Pisum sativum* L.),[44] lack glycosylation and proteolytic cleavage sites entirely. Previous studies also evidenced that the 7S vicilin-like globulins may undergo proteolytic events during seed development while preserving the bicupin structure and trimeric assembly necessary for accumulation and packing into protein storage vacuoles (PSV). Firstly, the primary vicilin translation product contains a transient N-terminal signal sequence which is cleaved co-translationally as the growing polypeptide is directed from membrane-bound polysomes into the lumen of the endoplasmic reticulum (ER). After the signal peptide cleavage, as first described by Gatehouse et al[45] in garden pea (*Pisum sativum L.*), vicilin-like globulins, before germination, may undergo cleavage processes which regard two sites, and produce three lower-mass segments termed as α- (~19 kDa), β- (~14 kDa) and γ-chains (~13-16 kDa) (Figure 3), although also α+β (~36 kDa) and β+γ polypeptides (~32 kDa) have been observed [46]. In this regard, Bourgeois et al.[47] evidenced that some pea cultivars undergo cleavage during seed maturation at both sites, some at only one site, and some not at all. Lately, Wilson et al.[48] have described how a 47 kDa form of 7S vicilin, undergoes limited proteolytic digestion during seed development as previously reported by Gatehouse et al.[45] In particular, the proteolytic processing of this specific pea vicilin at two sites produces three segments termed α, β and γ, is summarized in figure 3 and can follow two different pathways. In one pathway, the intact vicilin is cleaved first at the β/γ junction, specifically at an Asn-Asp bond[45,46] in an exposed loop in the otherwise compact C-terminal cupin

domain.[49] In another pathway, the intact vicilin is cleaved first at the α/β junction, specifically at a Lys-Asp bond.[50] By alignment with the structure of phaseolin (the 7S globulin in the common bean Phaseolus vulgaris L., [51] the α/β junction is located in the linker between the two cupin domains. This pathway is marked by the appearance of a 28 kDa and a 19 kDa intermediate. For cultivars where both pathways occur before germination, vicilin is eventually cut at both sites, so that three shorter polypeptide chains of molecular masses 19 (α), 13 (β), and 12.5/16 kDa (γ) are seen in reducing SDS-PAGE analysis of proteins extracted from developing seeds.[45] Some pea cultivars show evidence of cleavage during seed maturation at both sites, some at only one site, and some not at all. Glycosylation of the respective vicilin subunits takes place in the ER, where trimers are also assembled. Glycosyl-side chains are trimmed and modified in the Golgi apparatus. Phaseolin, for instance, the most well-known 7S globulin from garden bean (*Phaseolus vulgaris L.*), is composed of non-cleavable 50 kDa subunits with different degrees of glycosylation.[52] Like prolegumin, vicilin trimers are sorted into transfer vesicles at the trans-Golgi network and transported into the PSVs, where they are stored until needed.

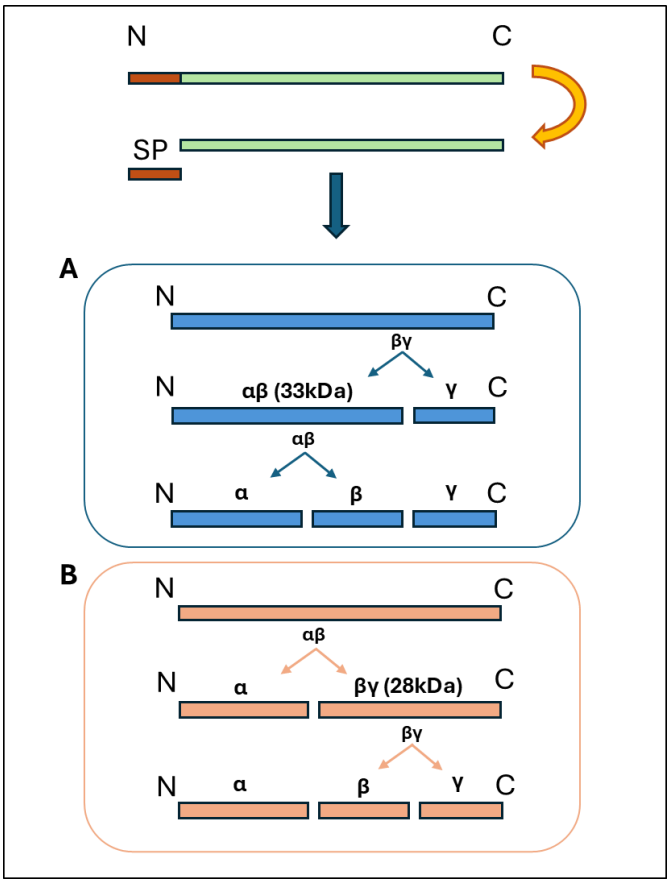


Figure 3. *Schematic diagram of the proteolytic processing pathways of pea vicilin. After the cleavage of the signal peptide (SP), two possible pathways can occur: (A) Initial vicilin cleavage at the β/γ junction, producing the 33 kDa intermediate, followed by cleavage of the 33 kDa intermediate at the α/β junction. (B), initial vicilin cleavage at the α/β junction, producing the 28 kDa intermediate, followed by cleavage of the 28 kDa intermediate at the β/γ junction*

Although 7S- and 11S-like proteins don't seem to share any apparent aminoacidic region similarities in their sequences, they do have similar properties, including the ability to form aggregates and multimeric structures. The 7S globulins are known to form a trimeric structure in their mature form, although they may also undergo reversible aggregation into hexamers depending on the ionic strength.[53] By contrast, the mature form of 11S globulin is a hexameric complex but is initially assembled and transported through the plant system as an intermediate trimer[54]. Therefore, in-depth comparisons have shown that the 11S and 7S globulin subunits are related in structure.[51,55] Such comparisons indicate that the C-terminal chain of the 11S legumins is related to the C-terminal region of the 7S vicilins. Lawrence et al.[51] determined the x-ray crystal structure of 7S phaseolin (French bean 7S globulin) and established sequence homologies between 7S and 11S proteins. The homologies

showed that the 11S sequences, with four major insertions of sequence, can be aligned with 7S sequences. Moreover, Shutov et al.[56] compared the N-terminal and C-terminal domains of eleven different legumins and seven vicilins of several dicot, monocot, and gymnosperm species using multiple alignments. The comparisons using all six possible pairwise combinations revealed that the N-terminal and C-terminal domains of both protein families were similar to each other. These results, together with data obtained from the distribution of variable and conserved regions, on the positions of susceptible sites for proteolytic attack (see the next paragraph) suggest that both protein families share a common single-domain ancestor molecule and lead to the hypothesis that a triplication event has occurred during the evolution of a putative legumin/vicilin ancestor gene. Moreover, the comparison of the intron/exon pattern reveals that at least three out of five intron positions are precisely conserved between the genes of both protein families, further supporting the idea of a common evolutionary origin of recent legumin and vicilin encoding genes.

Glutelins

In pulse crops, the classical distinction between globulins and glutelins based on solubility properties is increasingly regarded as operational rather than molecular, as these storage protein classes display substantial structural and compositional overlap. Globulins and glutelins both belong to the cupin superfamily and share conserved β-barrel domains as well as similar subunit architectures composed of acidic and basic polypeptides linked through disulfide bonds[57]. Proteomic and electrophoretic studies have revealed that a considerable proportion of proteins identified as glutelins in pulses are, in fact, denatured or aggregated globulins rendered alkali-soluble during Osborne fractionation[58,59], while gene sequence analyses indicate high identity between globulin- and glutelin-encoding genes, suggesting a shared evolutionary origin[60]. Immunochemical cross-reactivity and comparable thermal and gelation behaviors further substantiate their molecular relatedness[61]. Although globulins are classically defined as salt-soluble storage proteins, several studies have shown that a subset of globulins in legume seeds exhibits reduced solubility and behaves like glutelins, requiring stronger extraction buffers such as dilute alkali or acid solutions for solubilization. This phenomenon largely arises from protein–protein and protein–

matrix interactions, post-translational modifications, and seed maturation effects that alter the physicochemical environment of the storage proteins. During seed desiccation and maturation, globulins such as legumin (11S) and vicilin (7S) can undergo partial denaturation, aggregation, or covalent cross-linking through disulfide and non-covalent interactions, leading to the formation of insoluble protein complexes[58,59]. These aggregates often become embedded within the protein matrix of the seed storage vacuoles, decreasing their solubility in neutral salt buffers and necessitating extraction with stronger buffers such as dilute NaOH, urea, or SDS, similar to those used for glutelin solubilization[61]. In addition, the amino acid composition and surface hydrophobicity of certain globulin subunits promote aggregation under physiological ionic conditions, where hydrophobic interactions and intermolecular β-sheet formation drive conformational rigidity and reduce solubility[57]. Proteomic analyses in peas, lentils, and chickpeas further confirm that proteins found in the alkali-soluble glutelin fraction share immunological and electrophoretic properties with globulins, indicating that their apparent "glutelin-like" solubility reflects conformational or associative changes rather than true molecular identity[62]. Taken together, the globulin–glutelin continuum in pulse crops reflects a common evolutionary and structural framework, with apparent differences arising primarily from extraction methodology, physicochemical state, and seed maturation rather than intrinsic protein divergence. These structural rearrangements also bear functional implications for protein extraction efficiency, gelation, emulsification, and digestibility, underscoring the biochemical and technological importance of understanding the solubility continuum in the development of pulse-based protein ingredients.

Proteolytic cleavage events of storage proteins in developing and germinating seeds

Different studies demonstrated that legume species vary in how globulins are stored in seeds, undergo proteolytic processing during seed development, while retaining their characteristic bicupin fold and the trimeric or hexameric quaternary structure essential for their stable accumulation within protein storage vacuoles. Proteolytic cleavage plays an important role in the deposition and reactivation of storage proteins

in plant seeds. As reported above, during seed development, precursor polypeptides of both legumins and vicilins undergo limited proteolysis, which processes them into mature subunits of reserve proteins within the storage tissue cells. Proteolytic processing steps are closely related to an intracellular protein transfer through the endomembrane system and to the deposition in the storage vacuole. In germinating seeds, specific endopeptidases trigger storage protein breakdown by limited proteolysis. This event induces changes in the conformation of storage proteins by opening them to attack by additional endo- and exopeptidases, which degrade the protein reserves completely. Proteases responsible for both limited cleavage and complete degradation are synthesized as precursors that also undergo stepwise limited proteolysis when they are formed in cotyledons of developing or germinating seeds. Usually, this is a process that transforms enzymatically inactive proenzymes into active proteases. Multiple cellular compartments are involved in these processing steps, and many proteases are encoded by small multigene families. Different members of these protease families appear to be active at various stages of seed development and germination. Proteolytic processes that contribute to the molecular maturation and the reactivation of storage proteins in plant seeds seem to be controlled by differential expression of members of the protease-encoding gene families.[39]

The processing of globulins pro-polypeptides starts with the signal peptide cleavage. Precursor polypeptides of storage globulins are transported from their site of synthesis in the cytoplasm into the lumen of the endoplasmic reticulum (ER), thereby entering the secretory pathway via the endomembrane system. This transfer is facilitated by an N-terminal signal sequence, which is proteolytically cleaved off after fulfilling its role as a targeting and membrane transfer signal. The single-site cleavage is catalyzed by a signal endopeptidase located in the ER membrane and acting at its inner surface.[63] Following the cleavage of the signal peptide, a pro-polypeptide is typically produced. As far as presently known, these polypeptides do not undergo further limited proteolysis in the ER or Golgi compartment, although there, the polypeptides fold into a conformational state and oligomerize, which makes them competent for intracellular transfer through the endomembrane system and into the PSVs. The next steps in the processing of storage proteins are less well understood and can alter depending on the kind of protein and developmental stage,

both within and between species. The pro-polypeptide of the storage proteins is subjected to limited proteolysis by a specific family of enzymes called *Legumains.* Legumains stands for an Asn-specific cysteine proteinase family of enzymes, which can be divided into two subgroups: animal and plant Legumains. Particularly, "Plants Legumains" is a family of plant cysteine endopeptidases with a cleavage specificity for Asn in the P1 position of peptide bonds, even though Asp-flanked peptide bonds are also cleaved, but with a much lower efficiency. The first plant cysteine proteinase (CPR) to specifically cleave peptide bonds involving the carboxyl group of Asn, i.e., Asn in P1 position, was purified and characterized from cotyledons of vetch (*Vicia sativa* L.) seedlings.[64] This enzyme was named proteinase B. In combination with papain-like CPRs and carboxypeptidases, proteinase B contributes to the breakdown of storage globulins in the protein bodies of cotyledons of growing vetch seedlings.[65] Ten years later, other Asn-specific CPRs were prepared from developing castor bean seeds[66] and cotyledons of maturing soybean[67] at stages when storage protein deposition takes place. *In vitro*, these enzymes catalyze an Asn-specific limited proteolysis of 2S albumin and 11S globulin precursor polypeptides, transforming them into mature proteins. *In vivo*, this process takes place in the protein storage vacuole (PSV). The castor bean proteinase, therefore, was called vacuolar processing enzymes (VPE) and, thereafter, plant Legumains are usually called VPE. As was previously noted for castor bean proVPE, this class of enzymes is processed and activated in the vacuole. ProVPE seems to remain enzymatically inactive in the absence of C-terminal processing; however, cleavage at a specific Asn residue site results in the production of an active intermediate that is then subjected to further autocatalytic processing. The C- and N-terminal propeptides cleave successively to produce the mature active form of VPEs.[68-69] Both VPE and proteinase B show maximum activity under acidic pH conditions similar to those present in the PSV of maturing seeds and in protein bodies during seed germination and seedling growth. Legumains are primarily responsible for the restricted proteolysis that processes precursor polypeptides in a way that is exclusive to Asn and Asp residues. This function is carried out in the vacuole and the cell wall. The most well-studied instance of the structure-function link between a legume and its processing substrate is the processing of 11S globulin pro-polypeptides, or pro-legumin. As previously mentioned, 11S globulins (legumins) are hexameric seed storage proteins. In this conformational state, the 11S proglobulin is transported from the ER through the

secretory pathway towards the PSV. There, the 11S pro-globulin remains assembled in trimers when it is Asn-specifically cleaved into α- and β-legumin chains that remain linked by disulfide bridges. As a result of this processing, two trimers always assemble into an 11S globulin hexamer (Jung et al. 1993, 1998). This conformational state is compatible with 11S globulin deposition in the PSV. The latter finally transforms into the protein bodies of mature seeds. Not only 11S globulins, but also 2S albumins and 7S globulins (vicilins) frequently co-exist with VPE in protein bodies of mature seeds, and thus they can also be subjected to legumains proteolytic events. Like 11S globulin, the mature 7S globulin in growing seeds appears to be shielded against irregular legumain attack by a particular shape. However, despite the fact that it is known that 7S globulin precursors do not undergo Asn-specific proteolytic processing, unlike 11S globulin, some studies reveal that limited proteolysis can still occur.

Chickpea (*Cicer arietinum L.*)

Chickpea (*Cicer arietinum L.*), also called garbanzo bean or Bengal gram, is an Old-World pulse and one of the seven Neolithic founder crops in the Fertile Crescent of the Near East.[70] Nowadays, chickpea is cultivated in over fifty countries across the Indian subcontinent, North Africa, the Middle East, southern Europe, the Americas, and Australia. Globally, it ranks as the third most important pulse crop in terms of production, following dry beans and field peas. There are two distinct types of cultivated chickpea: Desi and Kabuli. The Desi (microsperma) types have pink flowers, anthocyanin pigmentation on stems, and a coloured and thick seed coat. The Kabuli (macrosperma) types have white flowers, lack anthocyanin pigmentation on stems, and have white or beige-coloured seeds with a ram's head shape, a thin seed coat, and a smooth seed surface[71] (Figure 4). Ares, Vittoria, Principe, Reale, Sultano, and Pascià are commercial varieties belonging to the Kabuli type of chickpea and therefore exhibit similar phenotypic and compositional traits[72]. In addition, an intermediate type with pea-shaped seeds of local importance is recognised in India. The seed weight generally ranges from 0,1 to 0,3 g and 0,2 to 0,6 g in the Desi and Kabuli types, respectively.[73] The Desi types account for about 80–85 % of the total

chickpea area and are mostly grown in Asia and Africa[74]. The Kabuli types are largely grown in West Asia, North Africa, North America, and Europe.

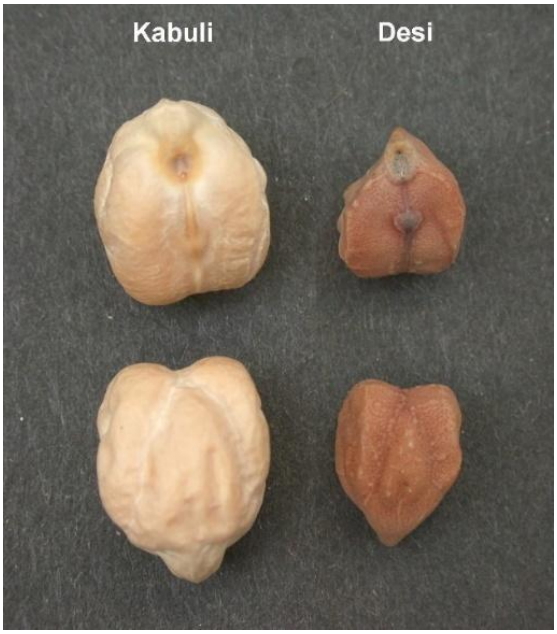


***Figure 4**. An example of Kabuli and Desi kinds of seeds*

Currently, there is a growing demand for chickpeas due to their nutritional value. In the semi-arid tropics, chickpea is an important component of the diets of those individuals who cannot afford animal proteins or those who are vegetarian by choice. Chickpea is a good source of carbohydrates and protein, together constituting about 80 % of the total dry seed mass[75] in comparison with other pulses, but it is also a good source of dietary fibre, fat, vitamins, and minerals.[76] Chickpea grain composition can be categorized as follows:

Carbohydrates. Dietary carbohydrates are classified into two groups: i) available (mono- and disaccharides), which are enzymatically digested in the small intestine, and ii) unavailable (oligosaccharides, resistant starch, non-cellulosic polysaccharides, pectins, hemicelluloses, and cellulose), which are not digested in the small intestine.[77] Chickpea contains monosaccharides (ribose, glucose, galactose, and fructose), disaccharides (sucrose and maltose), and oligosaccharides (stachyose, ciceritol, raffinose, and verbascose). The amount of these fractions varies, though not significantly, between the Desi and Kabuli genotypes. [78]

Dietary Fiber (DF). DF refers to the indigestible portion of plant-based foods that passes through the human small intestine largely unchanged. It consists of polysaccharides, oligosaccharides, lignin, and other plant-derived components.[79] DF can be classified into soluble and insoluble fibres. The soluble fibre is digested slowly in the colon, whereas the insoluble fibre is metabolically inert and aids in bowel movement.[80] The insoluble fiber undergoes fermentation, aiding in the growth of colonic bacteria.[80] The total DF content (DFC) in chickpea is 18–22 g/100 g of raw chickpea seed.[80,81] Soluble and insoluble DFC are about 4–8 and 10–18 g/100 g of raw chickpea seed, respectively. The DFC of chickpea seeds is equal to or higher than that of other pulses such as lentils (Lens culinaris) and dry peas (Pisum sativum).[80] Usually, no significant differences are found in soluble DFC between the Kabuli and Desi types due to the similar proportion of hemicelluloses that constitute a large part (about 55 %) of the total seed DF in the Kabuli and Desi types. [82]

Fat content. The total fat content in raw chickpea seeds varies from 2,70 to 6,48 %. [83,84] Wood & Grusak[76] reported a fat content of 3,40–8,83 and 2,90–7,42 % in Kabuli- and Desi-type chickpea seeds, respectively. Further, even higher levels (3,80–10,20 %) of fat content in chickpea have been reported.[78] The fat content in chickpea (6,04 g/100 g) is higher than that in other pulses such as lentils (1,06 g/100 g), red kidney bean (1,06 g/100 g), mung bean (1,15 g/100 g) and pigeon pea (1·64 g/100 g), and also in cereals such as wheat (1,70 g/100 g) and rice (about 0,60 g/100 g). Globally, chickpea is composed of about 66 % of Polyunsaturated Fatty Acids (PUFA), about 19 % Monounsaturated Fatty Acids (MUFA), and about 15 % of Saturated Fatty Acids (SFA). [78]

Minerals and Vitamins. The different minerals present in raw chickpea seeds on average are about 5,0 mg/100 g of Fe, 4,1 mg/100 g of Zn, 138 mg/100 g of Mg, and 160 mg/100 g of Ca. There were no significant differences between the Kabuli and Desi genotypes except for Ca, with the Desi types having a higher content than the Kabuli types.[85] The amount of total Fe present in chickpea is lower (5,45 mg/100 g) compared with other pulse crops such as lentils (8,60 mg/100 g) and beans (7,48 mg/100 g).[86] Data on other minerals present in chickpea are very limited. Pulses are a good source of vitamins. Chickpea is a relatively inexpensive and good source of folic acid and tocopherols.[87] It is a relatively good source of folic acid, coupled with more modest amounts of water-soluble vitamins such as riboflavin (B2), pantothenic

acid (B5), and pyridoxine (B6), and these levels are similar to or higher than those observed in other pulses. [88]

Protein content. Among various pulse crops, chickpea has been reported to have a higher protein bioavailability.[89,90] The protein content in chickpeas varies notably depending on whether the seeds are whole or dehulled, ranging from 17–22% of the dry weight in whole seeds to 25.3–28.9% after dehulling.[91] Comparison between Kabuli and Desi has been inconsistent, showing significant differences in one instance (241 g/kg in 'Kabuli' v. 217 g/kg in 'Desi')[92] and no differences in another instance (217 g/kg in 'Kabuli' v. 215 g/kg in 'Desi').[93] It has also been reported that chickpea protein quality is better than some pulse crops such as black gram (*Vigna mungo L.*), green gram (*Vigna radiata L.*), and red gram (*Cajanus cajan L.*).[94] Moreover, to date, no significant difference in the protein concentration of raw chickpea seeds compared with some pulses such as lentils, red kidney bean, and white kidney bean has been noticed.[95] The amino acid profiles of chickpea exhibit minor variations in the quantity of a few amino acids, such as lysine, tyrosine, glutamic acid, histidine, and the two combined aromatic amino acids.[96] Generally, sulphur-rich amino acids (methionine and cystine) are limiting in pulses. Commonly consumed food pulses such as chickpea, field pea, green pea, lentils, and common beans are known to have low levels of methionine and cystine residues. There are no significant differences in the amino acid profiles of Kabuli- and Desi-type chickpeas.[97] Like other legumes, chickpea proteins are categorized based on their solubility in different solvents: albumins (soluble in water) that represent about 10-14% of the total protein content, globulins (soluble in salt solutions; 55-60% of the total protein fraction), prolamins (soluble in hydro-alcoholic solutions; 2-3%), and glutelin (soluble in acid/alkali solutions; 18%).[98] These proteins contribute to the nutraceutical properties associated with pulse consumption, but they can also exhibit antinutritional effects, such as hemagglutination, inhibition of digestion, and allergenic responses.[99] Considering the potential impact of chickpea seed proteins on consumer health, including their nutraceutical, antinutritional, and allergenic effects, it is crucial to investigate changes in protein composition influenced by both agronomic and genetic factors. However, up to date, comprehensive studies on the protein fraction of chickpeas are lacking. For all these reasons, it might be useful to

extend the knowledge about the protein fraction of chickpeas by using proteomic approaches based on mass spectrometry and bioinformatic tools.

State of the Art in Chickpea Seed Proteomics

Thanks to its distinctive characteristics and properties, the chickpea has become a central focus for scientific research in recent years, particularly on the characterization of the protein fraction. The release of the chickpea genome in 2013 has enabled high-end molecular studies, particularly in proteomics, which have helped to investigate nutritional features, stress responses, seed composition, and protein functional relationships. Despite insightful information about the composition of storage proteins, stress-related activities, and nutrient resilience that researchers in the field have made possible, significant methodological biases and interpretative gaps remain. The first ever well-documented isolation and characterization of chickpea seed proteins dates back to 1980, when Ganesh Kumar & Venkataraman [100] successfully purified a 10.3S globulin fraction. Their study, published in the Journal of Agricultural and Food Chemistry, marked a milestone in legume protein research and protein isolation. They extracted seed proteins using aqueous buffers and subsequently fractionated them, isolating a distinct 10.3S protein through differential solubility and ultracentrifugation. They described its subunit architecture, sedimentation coefficient, and amino acid composition, offering one of the earliest biochemical profiles of chickpea seeds' storage proteins. In 1983, Franklin Vairinhos & David R. Murray [101] compared seed protein fractions (albumin, globulin) of *C. arietinum*, *C. reticulatum*, and *C. echinospermum* using SDS-PAGE, identifying major subunits typical of legumin and vicilin. This study established the presence of multiple globulin polypeptides (e.g., 46–67 kDa). Later in 1991, Dhawan et al. [102] using SDS-PAGE, quantified macronutrient fractions in six cultivars, confirming that globulins made up approximately 53–60% of total protein, with notable variability in albumin, prolamin, and glutelin fractions. Overall, these works have enabled an early profiling at the protein-family level across species and cultivars, offering foundational data on protein composition diversity, even though no specificity at both peptide and sequence levels was investigated, nor storage proteins' properties in terms of emulsifying, foaming, and solubility capacity.

In the first decade of the 2000s, different studies highlighted more in-depth the characterization of the globulin and albumin fractions in terms of extraction and functional properties. One of the first works about these protein fractions was carried out by Liu et al[103] who extracted and analyzed protein isolates prepared by water and salt solutions from desi and kabuli chickpea cultivars in order to evaluate relative amounts of albumin and globulin, depending on extraction conditions. Their extractable yields varied with the extraction method and also with chickpea variety. Another interesting work made by Papalamprou et al.[104] was focused on isolating different fractions of chickpea seeds protein using alkaline extraction followed by isoelectric precipitation or ultrafiltration, containing mainly chickpea globulins and albumins, with the globulin fraction dominating over the albumins. The differences in protein composition between the isolates, as well as the impact of extraction conditions, were reflected in their protein solubility, surface hydrophobicity, sulfhydryl group content, thermal properties, and the onset of gelation during heating. Undoubtedly, these studies represent an initial contribution to the analysis and characterization of chickpea protein fractions and legume proteins in general. Although the protein fractions were not investigated at the molecular level, in terms of peptide sequence or post-translational modifications, the methods of extraction, purification, and analysis of physicochemical properties laid the groundwork for a more accurate and detailed investigation of the various protein fractions of this legume.

The first detailed mass spectrometry-based proteomic work on chickpea seeds was published by Chang et al.[105] in 2011. In particular, albumin, globulin, and glutelin fractions were prepared from chickpea seeds using sequential extractions. Molecular characteristics of individual protein fractions were investigated using SDS-PAGE in combination with LC-MS/MS techniques, and the data obtained were finally searched against the *Viridiplantae* database using 'Mascot MS/MS ion search'. SDS-PAGE results revealed that chickpea albumin and globulin fractions showed protein bands with molecular weights related to subunits of legumin (11S globulin) and pea vicilin (7S globulin). Proteomic analysis led to the identification of eighteen tryptic peptides from chickpea globulin fraction that corresponded to chickpea legumin α- and β-subunit (NCBI accession number: gi|6273402), while sixteen tryptic peptides from chickpea were identified as chickpea pro-vicilin precursor. This work

represents the groundwork of proteomic studies on chickpea seeds, demonstrating how tryptic peptides belonging to legumin and vicilin were identified from extracts of the most abundant fractions (i.e., albumins, globulins, and glutelins). On the other hand, since mass spectrometry-based proteomic studies rely heavily on the database used for bioinformatic searches, the absence of a database generated from the genome of the organism under investigation may result in a reduced number of identifications.

This gap was later overcome in 2013, when Varshney et al.[106] published the first draft genome of the chickpea seed. This team resequenced and analyzed 90 cultivated and wild genotypes from ten countries, identifying targets of both breeding-associated genetic sweeps and breeding-associated balancing selection, including traits that distinguish the two main market classes of cultivated chickpea, desi and kabuli. These data comprised an important resource for chickpea improvement through molecular investigation and provided insights into the genome and proteome.

Before 2013, proteomics efforts suffered from limited annotation databases, often relying on homology with other legumes. Post-genome, peptides detected by LC-MS/MS could be mapped directly to chickpea proteins, improving confidence in protein IDs and enabling allele-specific proteomics. The genome also helped uncover splice variants and protein forms that had been missed in earlier work, especially in major storage proteins like legumins and vicilins, and stress-related proteins, which are known to have several isoforms. On top of that, having access to genomic coordinates meant researchers could link seed proteomics data to specific genes, including those tied to allergenicity (like members of the Cupin family) and nutrition (such as sulfur-rich albumins).

More efforts on proteomics studies of chickpea have been made in more recent years, aiming not only at the characterization of proteins but mainly focused on the impact of these proteins on human health (i.e., resistance to digestion, allergenic properties), but just a few of them are focused on the seed protein fraction. In 2016, Singh et al.[107] employed proteomic analysis to map the seed storage protein network in landrace and cultivated chickpea accessions. They separated protein extracts by 2D-PAGE, and the spots obtained were analyzed using mass spectrometry. A total of

600 protein spots were detected in these accessions. Using trypsin in-gel digestion, proteins were identified by matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS), which showed 45% amino acid homology of chickpea seed storage proteins with the homologues of *Arabidopsis thaliana*. This study explored differentially expressed low-abundance proteins and seed storage proteins in chickpea, which extends information on the protein networks and helps better understand their expression patterns at the maturity level.

In the same regard, another interesting work by Ribeiro et al.[108] highlighted an important feature of chickpea protein fraction, specifically the resistance to human digestion. They pointed out that chickpea seed proteins can resist cooking (about 86.9% total protein) and/or in vitro simulated human digestion (15.9% total protein resists soaking, cooking, and digestion with pepsin and pancreatin). An interesting point of this work is that they carried out different MS methodologies, in particular, MALDI-MS/MS and LC-nESI-MS/MS, to identify and characterize proteins resistant to digestion. They also tested the efficiency of several proteases (trypsin, AspN, chymotrypsin, and LysC), in order to achieve a better identification, finding that AspN and trypsin were the most performing enzymes. When analyzed by MALDI-MS/MS, trypsin allowed the identification of at least one protein in 60% of the polypeptide bands of the SDS-PAGE employed for the separation, while AspN allows the identification in 48%. The use of LCnESI-MS/MS allowed the identification of many more proteins/polypeptides from digested seeds (232 vs 17 using trypsin). The majority of the proteins found to be able to resist simulated digestion were members of the 7S vicilin and 11S legumin seed storage protein classes, which are reported to contain bio-active functions. This study represents the first proteomic analysis examining how processing and simulated human gastrointestinal digestion impact the chickpea seed proteome. Chickpeas are known to exhibit both anti-nutritional factors and health-promoting (nutraceutical) properties. Therefore, identifying and characterizing the proteins that remain intact during digestion is essential for understanding the molecular basis of these effects. In this respect, Dadon et al.[109] used a serum pool from paediatric patients allergic to chickpeas to detect IgE-binding proteins from chickpea seeds by immunoassay and immunoblot inhibition assay. Protein samples enriched in chickpea legumin and vicilin were obtained by anion exchange chromatography and were identified by

mass spectrometric analysis. IgE-immunoassays of globulin fractions from chickpeas revealed that vicilin (50 kDa) and the basic subunit of legumin (i.e., the β chain, 20 kDa) were bound by IgE from patient serum. This result allowed to conclude that vicilin and the β basic subunit of legumin are major chickpea allergens. These two works together led to the conclusion that globulins constitute the predominant protein fraction in chickpea seeds, but they also have undesirable properties from a human health perspective, such as low digestibility and, more critically, the capacity to induce allergic responses in susceptible individuals.

Another interesting study that highlights the physicochemical properties of seed storage proteins was conducted by Chang et al.[110] in 2022. They developed alkaline extraction-isoelectric precipitation in conjunction with a modified salt dissolution-precipitation method to produce protein fractions on a large scale. Results showed that the purity of globulin, legumin, and vicilin fractions reached more than 90%, 80%, and 90%, respectively. Subsequently, the structural and functional properties of protein fractions were studied, and it was found that legumin fractions had higher foaming stability, emulsion stability index, and thermal stability compared to vicilin fractions since the former contained more sulfur-containing amino acids, SH and SS groups, and larger MW. Nevertheless, vicilin fractions had higher solubility, foaming capacity, and emulsion activity index compared to legumin fractions because of their lower molecular weight and less rigid structure. These findings could provide useful information not only to pulse growers to breed the varieties with premium characteristics, but also to pulse ingredient manufacturers and consumer companies to fabricate appropriate pulse proteins in light of the application area.

More recently, in 2024, Bose et al.[111] profiled eight chickpea cultivars using two complementary protein extraction solvents, namely urea and water-based extraction. The total protein, starch, and soluble sugar contents significantly differ between cultivars, and they have identified and quantified 2434 and 1809 proteins, respectively, from urea- and water-based extraction solvents using a data-independent acquisition (DIA) approach. Proteome-level variation across cultivars ranged from 9% to 25% for the urea-extracted proteins. Notably, storage protein abundance varied significantly between cultivars, with legumin being the most abundant, followed by vicilin and albumin. Furthermore, 50 common allergens were identified across both extraction methods. Overall, this study contributes to the

establishment of a chickpea pan-proteome resource and offers novel insights into key molecular pathways underlying genotype-specific traits, along with the implementation of an advanced MS-based approach to identify the largest amount of protein ever detected in chickpea cultivars.

In conclusion, current chickpea seed proteomics is hindered by several key limitations. First, the dominance of storage proteins in proteomic datasets, primarily due to methodological biases associated with techniques such as 2D-PAGE, obscures the detection of low-abundance regulatory proteins. Second, the resolution of post-translational modifications (PTMs) and proteoforms remains insufficient in complex seed protein extracts, limiting insights into protein function and regulation. Third, there is a notable scarcity of studies addressing seed-specific responses to abiotic and biotic stresses; most proteomic investigations to date have concentrated on seedling stages, thereby neglecting processes unique to mature seeds. Moreover, the functional validation of proteomic findings remains limited, which constrains the translational potential of these datasets. To advance the field, future research would benefit from the integration of high-throughput shotgun proteomics in order to obtain a much broader identification of the entire proteome of chickpea seeds. Additionally, the targeted quantification of candidate proteins, such as through parallel reaction monitoring (PRM) and Label-Free quantification (LFQ), should be prioritized to enhance sensitivity and reproducibility and to highlight the different expression of certain proteins related to different agronomic and environmental conditions. Moreover, the employment of Top-Down proteomic approaches may provide deeper insights into the molecular processes to which storage proteins are subjected during seed maturation, as well as their impact on the structural configuration of the mature proteins, including PTMs.

MS-based Proteomic approaches

Proteomic investigations typically involve multiple technologies to analyze proteins on a high-throughput dataset, identify proteins, assess protein expression levels, investigate their PTM levels, and examine protein-protein interactions. These experiments entail the simultaneous examination of several hundred or even thousands of protein components and therefore require high-resolution separation steps to simplify the complex mixtures of peptides or proteins that have to be subsequently analyzed by mass spectrometry. These separation steps can be accomplished through techniques such as gel electrophoresis or liquid chromatography. The specific strategy employed in proteomic research depends on the specific objectives. There is no doubt that proteomic studies have gained notable development principally as a consequence of the impressive increase in performance and versatility of the mass spectrometry (MS) instrumentation, which today represents an indispensable tool in proteomics, and which will be briefly discussed in the next paragraph.

Mass Spectrometry Technology

Mass spectrometry (MS) involves ionizing a sample in the gas phase and then separating the resulting ions based on their mass-to-charge ratio (m/z) under high pressure conditions, typically in the range of 10^{-6} to 10^{-7} mbar. As shown in Figure 5, the ion source, mass analyzer, and detector are crucial components of every mass spectrometer. The approach was mostly used for tiny volatile substances in the early decades of the 20th century when J.J. Thomson constructed the first mass spectrometer[112]. The large-scale study of protein and peptide analytes was made possible by the invention in the late 1980s of the two “soft” desorption/ionization MS techniques, the Matrix-Assisted Laser Desorption/Ionization (MALDI) by Franz Hillenkamp, Michael Karas[113]and Tanaka[114], and the Electrospray Ionization (ESI) by John Fenn.[115] MALDI and ESI are capable of ionizing polar and nonvolatile compounds such as proteins, DNA, and carbohydrates without significant analyte fragmentation. The main difference in pressure between MALDI and ESI sources lies in their operating environments. MALDI operates under high vacuum

conditions, while ESI technique works under atmospheric pressure, since ionization occurs in the open air before ions are transferred into the vacuum of the mass spectrometer through differential pumping stages. These two sources of ions have greatly broadened the applicability of MS to biology and revolutionized the analysis of biomolecules.

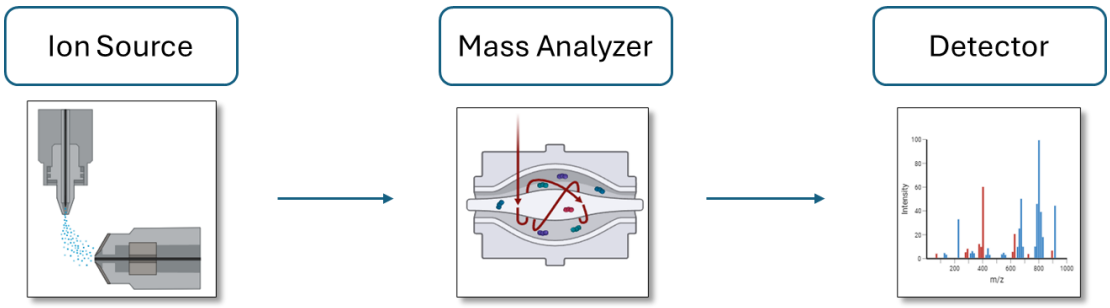


***Figure 5**. Schematic view of the essential parts of a mass spectrometer. Using ESI as Ion Source after the ionization in the Ion Source at atmospheric pressure, the ions continue their path through the mass analyzer in the high vacuum part of the mass spectrometer, where they will be finally detected.*

Particularly, in MALDI, the analyte is mixed with a matrix and deposited on a metal plate where they co-crystallize into a solid deposit. The matrix's energy is increased by a pulsed laser and subsequently transferred to the analyte.[116] Strong gas-phase acids, such as α-cyano-4-hydroxycinnamic acid,[117] 2,5-dihydroxybenzoic acid,[118] and 3,5-dimethoxy-4-hydroxycinnamic acid,[119] are often used as matrices for peptide and protein analysis. The majority of the resultant gas-phase analyte ions are singly charged, making it easier to interpret MALDI mass spectra. MALDI provides quick analysis (1–30 seconds per sample) and is reasonably resistant to sample contaminants such as salts and detergents.

In the ESI, a dissolved analyte is passed through a narrow, charged capillary with a tip positioned close to the counter electrode and the entrance to the mass spectrometer. Charged solvent droplets carrying the analyte are formed at atmospheric pressure, and upon solvent evaporation, the analyte ions are transferred to the gas phase [120], Figure 5). The solvent needs to be volatile for efficient ionization, and involatile salts and detergents negatively impact the quality of ESI spectra. Positive mode ESI is more common than negative mode for peptides and

proteins mass spectrometry. ESI generally produces multiple charge states for each analyte. This makes it harder to interpret the spectra, but also makes it possible to analyze larger molecules with mass spectrometer analyzers with a limited m/z detection range. One important advantage of ESI, when compared to MALDI, is that it makes it possible to couple the mass spectrometer to other techniques that separate proteins and peptides before MS. In this respect, ESI gained immediate popularity because it can be easily coupled online with high-performance separation techniques such as capillary electrophoresis and HPLC, and nowadays represents the most used MS-based method in proteomics. Moreover, the implementation of separation techniques in miniaturized formats coupled online with high-performance mass spectrometers and the development of miniaturized ESI sprayers (nano-ESI) have increased the sensitivity of the technique,[121-122,] as well as reducing the amount of analyte required for complete and routine sequence characterization below the attomole (10^{-18}mol) range.

The gas-phase analyte ions, produced in the Ion Source of the instrument, are passed to the mass analyzer, where the ions are separated based on their mass-to-charge ratio. Usually, the ion passes through a series of electrostatic lenses and filters before reaching the analyzer. The mass analyzer plays a fundamental role in mass spectrometer technology, together with its key parameters such as sensitivity, resolution, mass accuracy, and the ability to generate MS/MS spectra. The mass analyzers commonly used in protein analysis include: quadrupole (Q), Ion Trap (IT), time-of-flight (TOF), Fourier transform ion cyclotron resonance (FTICR), and Orbitrap.[123] These mass analyzers differ considerably in design and performance, each with its strengths and weaknesses. They can be stand-alone or put together in tandem (hybrid instruments), in which each type takes advantage of a different physical property of the ion motion for mass-to-charge-based separation. Briefly, Quadrupole mass analyzers are made of four parallel metal rods that can have hyperbolic or circular shapes. Opposing pairs of rods have the same charge, and both direct current (DC) and radio frequency (RF) alternating current (AC) are applied to the rods. Ions are injected axially into the quadrupole, and their trajectory at a set ratio of voltages depends on the mass-to-charge ratio of the ion. By changing the applied electric potentials, ions of different m/z can be selected to pass the quadrupole. Then, the ions can be detected or passed to another mass analyzer, such

as a TOF or an Orbitrap analyzer. These latter options are very often used, and the quadrupole can be used as a mass filter to select only a subset of ions for further analysis by tandem mass spectrometry. [124]

Ion trap mass analyzers, like quadrupole, are mass analyzers that use electric fields to trap and analyze ions. There are multiple possible geometries of ion traps, such as the cylindrical and the linear geometries. Ions are introduced into the trap and held in place by applying a combination of DC and RF voltages to electrodes arranged in a specific configuration (e.g., a ring electrode with endcap electrodes). [125] Then, the RF voltage is scanned, causing ions to be sequentially ejected from the trap based on their mass-to-charge ratio (m/z). The linear ion trap is built similarly to a quadrupole but with added end electrodes in the axial direction [126]. Ion traps, routinely coupled with an ESI source, represent the unique analyzer that alone allows us to determine either the mass of a given peptide or its sequence by MS/MS analysis. It is robust, sensitive, and moderately inexpensive but shows a relatively low resolution (< 2000) and mass accuracy (100-500 ppm).

As its name suggests, in a time-of-flight (TOF) analyzer, the m/z ratio is determined based on the time it takes the ion to pass the flight tube length to the detector. All the ions are accelerated in the electric field, which leads to the ones carrying the same m/z ratio having the same kinetic energy.[127] The flight tube can be linear or equipped with an ion deflector to prolong the ion path (reflector mode).[128] Stand-alone TOF analyzer shows a mass resolution of up to 20000, and mass accuracy comprised between 5 to 50 ppm. It requires pulsed ionization methods such as MALDI, whereas it cannot be directly interfaced with ESI sources. However, the TOF analyzer can be frequently used in combination with the quadrupole analyzer, i.e., hybrid Q-TOF instruments, or with an ion trap (LTQ-TOF). So, it can also be coupled with ESI because both the front-end quadrupole and the IT work at the vacuum level that is compatible with the ESI source.

The Orbitrap analyzer contains two electrodes, the outer shaped like a barrel and the inner like a spindle.[129] The ions are rotating around the inner electrode, oscillating in the radial direction but also moving in the axial direction (Figure 6). The harmonic movement in the axial direction is dependent on the square root of the mass-to-charge ratio. The image current is used to detect the signal, and the m/z information is

extracted by performing the Fourier transform of that signal. This analyzer, together with the FTICR, allows the measurement of the m/z value of the ions with the highest resolution (~ 500.000) and mass accuracy (< 5ppm) of any analyzer used in MS.

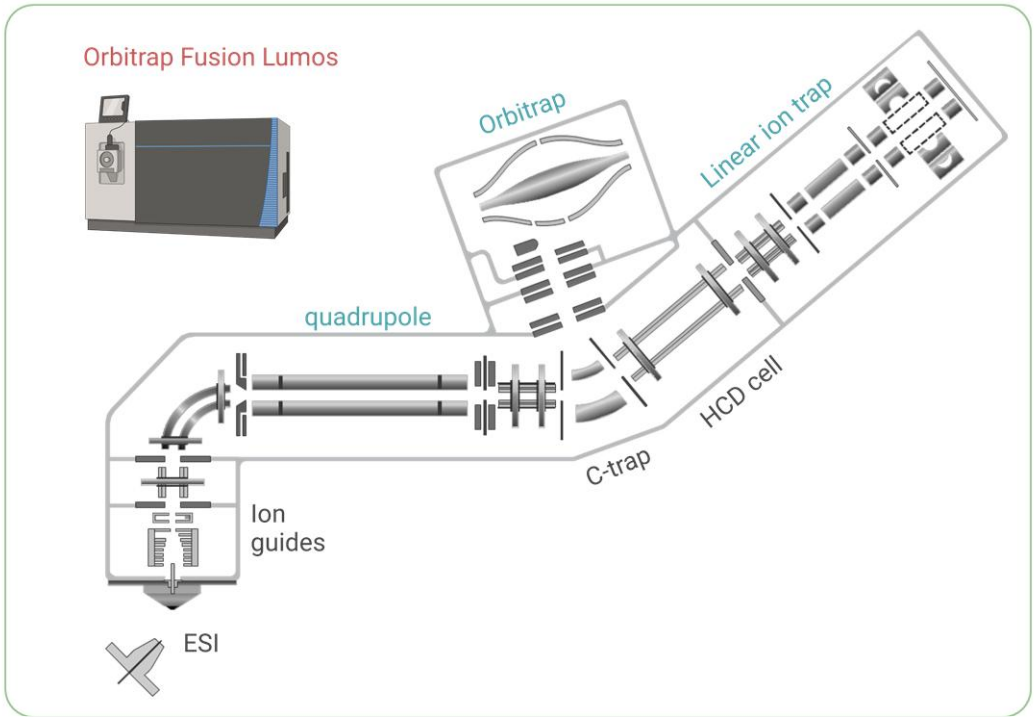


***Figure 6**. Schematic representation of the Orbitrap Fusion Lumos mass spectrometer. Adapted from https://www.thermofisher.com*

Bottom-up vs Top-down Proteomics

There are two commonly used mass spectrometry-based strategies, known as "bottom-up" and "top-down", in proteomics (Figure 7).[130] In bottom-up proteomics, proteins are enzymatically digested into peptides using one or more proteases. The protein digestion is a crucial step in such experiments, and often represents the bottleneck in terms of time consumption. Trypsin is the protease of choice in most bottom-up studies. It cleaves the proteins on the C-terminal side of lysine and arginine residues. The presence of a charged amino acid at the C-terminus makes the tryptic peptides easy to ionize in the positive ion mode, and the peptide size makes them suitable for LC/MS analysis. But, in specific cases, peculiar characteristics of

the target proteins such as amino acid composition or hydrophobicity may require the selection of most suitable proteolytic enzymes or multiple digestion strategies.

Alternative proteases include: LysC, LysN, AspN, GluC, argC, and chymotrypsin. The resulting peptides are then subjected to MS and MS/MS analyses. Bottom-up proteomics is a sensitive and robust technique with a wide range of applications. One major drawback of bottom-up MS is that it doesn´t preserve the proteoform level information. More in detail, bottom-up proteomic strategies are limited in their ability to identify novel protein variants or isoforms due to several inherent constraints. Proteolytic digestion fragments proteins into short peptides, resulting in the loss of information regarding peptide connectivity and complicating the reconstruction of full-length sequences.

The top-down approach involves the MS analysis of intact proteins, with their fragment ions generated within the mass spectrometer, primarily using high-resolution instruments such as FT-ICR and MS Orbitrap, without prior enzymatic digestion.

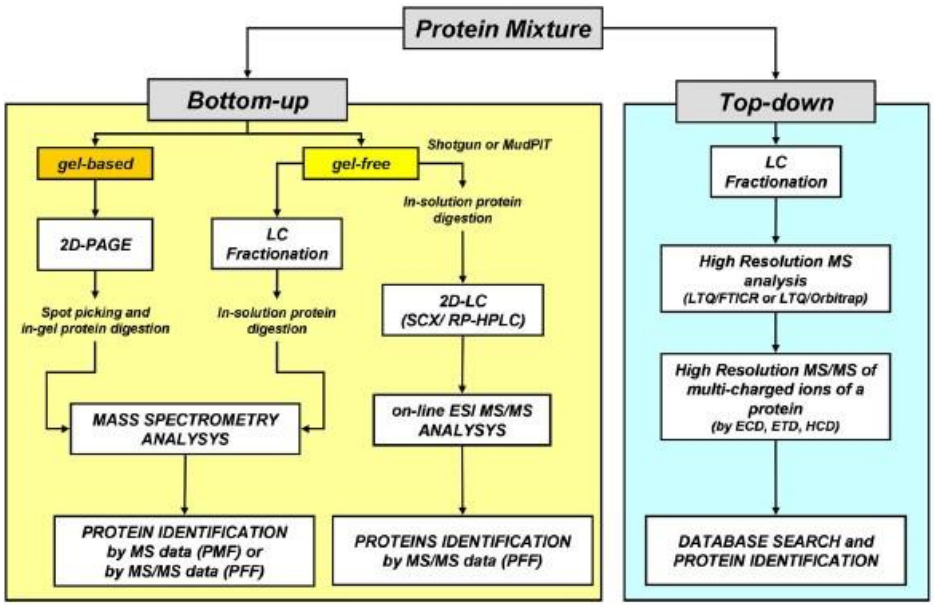


Figure 7. *Overview of both "Bottom-up" and "Top-down" approaches in proteomics*

The bottom-up approach is frequently employed when dealing with complex samples, and it can be further categorized as “gel-free” or "gel-based”. In the gel-free approach, also known as shotgun or MudPIT (Multi-dimensional Protein Identification Technology), the protein mixture is enzymatically digested directly into peptides. Then, these peptides can be separated using bidimensional chromatography. In the first dimension, they are separated based on their charge using a technique called strong cation exchange (SCX), and in the second dimension, they are separated according to their hydrophobic interaction with the stationary phase using a reversed-phase high-performance liquid chromatography (RP-HPLC). Finally, the peptides are transferred into a mass spectrometer, where they are ionized and fragmented in the gas phase to generate MS/MS (MS^2) spectra, which contain all the information to characterize specific peptides. Alternatively, the shotgun approach can also be conducted using one-dimensional liquid chromatography with a slow gradient in columns of 50 cm or more in length.

The “gel-based” approach, instead, typically begins with the initial separation of proteins using 2D electrophoresis; subsequently, the isolated components undergo in-gel enzymatic digestion, followed by LC-MS/MS analysis.

In both approaches, peptide fragmentations can be achieved through various techniques, including CID (collision-induced dissociation), HCD (high-energy collision dissociation), ETD (electron transfer dissociation), and ECD (electron capture dissociation). The most common techniques are CID and HCD, in which selected precursor ions are fragmented by collisions with an inert gas. These collisions primarily lead to the cleavage of peptide bonds, generating numerous fragment ions that are further analyzed. These fragmentation techniques yield distinctive ion series, namely a-, b-, c-, x-, y-, and z-series,[131] based on the type of fragmentation and ions obtained (see Figure 8). CID and HCD predominantly generate b- and y-ions. Conversely, ECD and ETD predominantly produce fragment ions from the c- and z-series, thereby providing complementary information to CID and HCD. The signal pattern of an MS/MS spectrum is strictly related to the molecular structure of the precursor ion. For peptides, the peaks observed in the MS/MS spectrum allow going back to the amino acid sequence of the precursor peptide (*de novo sequencing* strategy).

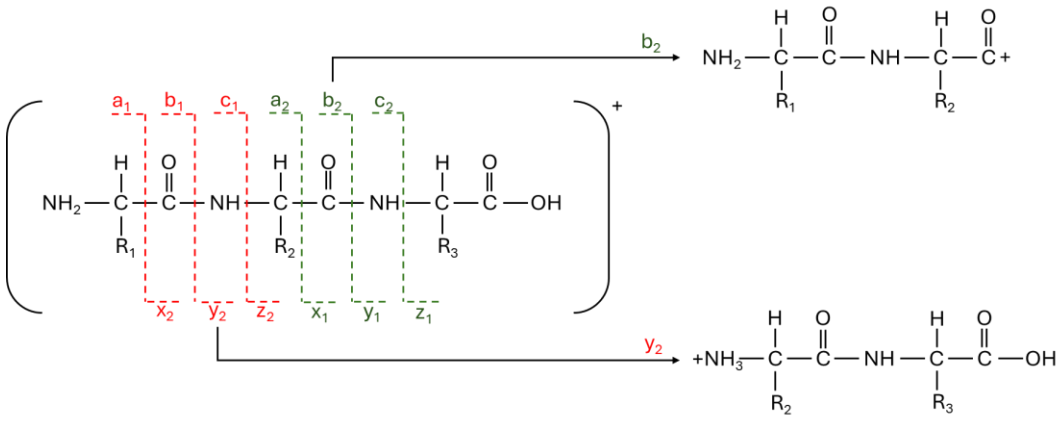


Figure 8. *Schematic overview of the fragmentation of a peptide*

The top-down approach involves the ionization of intact proteins in the gas phase, followed by their direct analysis using MS and tandem MS, without prior enzymatic digestion. In this method, proteins present in complex mixtures are initially fractionated and separated, resulting in either pure individual proteins or less complex protein mixtures. Subsequently, the separated proteins, or simplified protein mixtures, are subjected to offline static infusion or, more commonly, online LC/MS analysis, utilizing mass spectrometers equipped with high-resolution mass measurement capabilities to determine the masses of intact protein ions. The top-down strategy offers the advantage of providing more specific and detailed data. The key to the success of this approach is an efficient fragmentation of intact proteins. Moreover, as above reported, top-down analysis requires a pre-fractionation step of protein samples in order to obtain more simple protein mixtures, or purified proteins. In this respect, separation of proteins represents one of the most important problem for top-down analysis. Chromatographic methods used to separate intact proteins may show poor resolving power for large proteins (>50 kDa). Moreover, intact proteins are often difficult to ionize, fragment, and separate efficiently, which restricts the method primarily to low- and medium-molecular-weight proteins. The technique requires highly specialized instrumentation with high resolving power and sensitivity, making it less accessible and more technically demanding than bottom-up approaches. [132]

However, although it is still less popular than the alternative tandem MS approaches for protein analysis, top-down MS is rapidly advancing. Improvements in sample preparation, proteoforms separation, instrumentation, acquisition methods, and bioinformatics tools are increasing the reproducibility, sensitivity, and the mass range of the technique.

Another emerging proteomic strategy is the "middle-down approach", which faces one of the main limitations of the top-down approach, namely the analysis of proteins too large to be analyzed (mol wt. > 60 kDa). In contrast to the bottom-up approach, the middle-down approach involves the study of large peptides or polypeptides that are obtained through a controlled proteolysis step of the precursor protein. However, the size of these resulting peptides is generally larger than that encountered in the bottom-up approach. In the bottom-up approach, proteolytic peptides typically range from approximately 7 to 20 amino acid residues in length (corresponding to a molecular mass range of 800-2000 Da). In the middle-down approach, proteolytic peptides are generated with lengths greater than about 30-40 amino acid residues, extending up to approximately 100-110 amino acid residues, corresponding to a molecular mass comprised between 4000 Da and 12000 Da.[133] Consequently, the number of proteolytic peptides produced by the middle-down approach is relatively smaller compared to the number of peptides generated by typical bottom-up protocols. This implies that the complexity of a sample obtained through the middle-down approach is lower than that of a bottom-up approach. Consequently, detecting more unique peptides when employing the middle-down approach is more frequent. At the same time, the middle-down approach shows advantages related to the chromatographic analysis of peptides, which are much more efficiently separated than intact proteins, because of their narrower mass distribution, charge, and hydrophobicity.

Overall, top-down and bottom-up approaches represent two complementary approaches. An overview of some of the benefits and drawbacks of top-down and bottom-up MS for protein analysis is shown in Figure 9. Regardless of the chosen proteomic approach, it is crucial to point out that generating MS data is not the final step in proteomics experiments. Instead, it serves as a fundamental prerequisite for conducting the subsequent stage of proteomic studies, which involves protein identification using high-throughput bioinformatics methods. These bioinformatics

techniques play a crucial role in deciphering the huge amount of data obtained from MS analysis and enable the identification of proteins present in the sample.

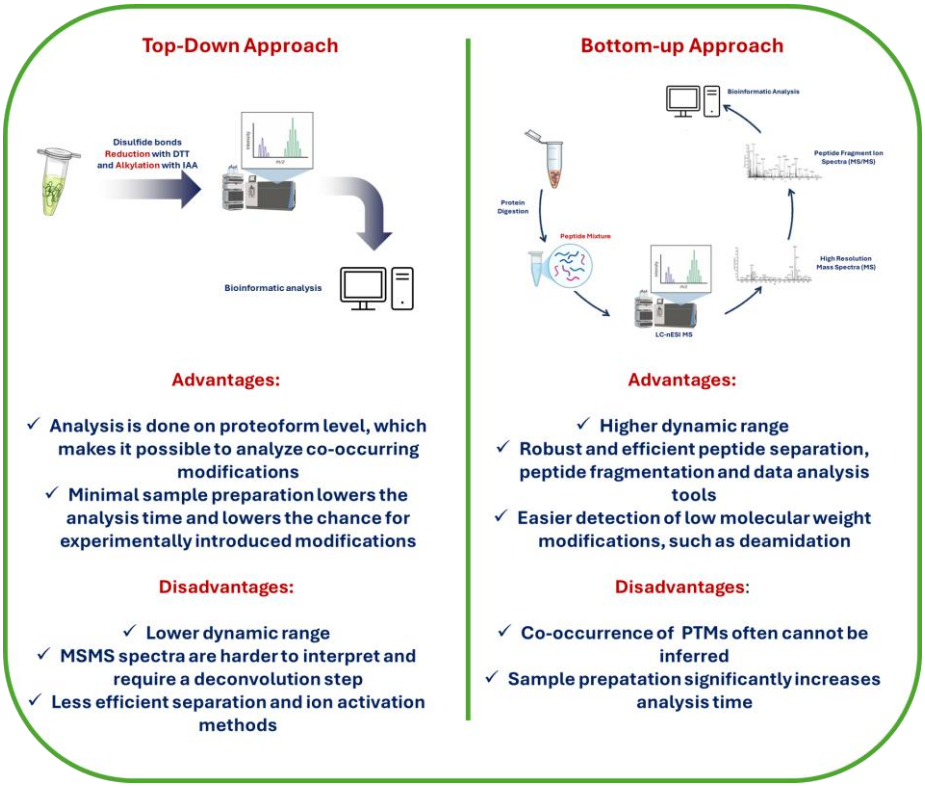


Figure 9. Overview of advantages and disadvantages of 'Bottom-up' and 'Top-Down' approaches

Bioinformatic database search by MS data

The continuous exponential growth of protein databases, together with the improvements in accuracy and resolution of the mass spectrometers, which produce a huge and complex amount of data, requires the indispensable use of dedicated software for the identification and characterization of proteins in proteomics.[134] In general, protein characterization and identification can be achieved by comparing the experimental MS data with those derived from a protein database, employing specialized search engines such as MASCOT, PEAKS, MaxQuant, Ms-Fit, ProFound, Proteome Discoverer, and others. In the bottom-up approach, where a protein is enzymatically digested to generate a set of peptides, protein identification can be accomplished using two strategies: Peptide Mass Fingerprinting (PMF) and Peptide-Fragmentation Fingerprinting (PFF). The concept behind PMF can be described as follows: using a peptide mass fingerprint, an identified protein sequence can be automatically searched using a software tool that reports the best matching

proteins against a proper database. The proteins are ranked based on the number of peptide matches. Advanced scoring algorithms consider factors such as mass accuracy and the percentage of protein sequence coverage, aiming to provide a confidence level for the match. The strategy of PFF is similar to peptide mass fingerprinting. PFF software tools typically generate theoretical MS/MS spectra by theoretical fragmentation of peptides obtained from the in-silico digestion of database proteins. In contrast to peptide maps, which represent an overview of the whole protein, peptide fragmentation provides primary structure information on a single peptide. As *score* is meant to be the statistical amount of protein identification. This score is generated by specific algorithms that assess the agreement between experimental data and *in silico* data obtained through computational analysis. The accuracy of identification depends on various factors, including the number of identified peptides for each protein, the precision of mass measurement, the specificity of the protease used, and the presence of the protein sequence in the database. In cases where a specific protein is not present in the database, it is still possible to identify it by relying on the presence of closely related species in the database (*cross-species protein identification*). These species share a high degree of similarity, allowing for the identification of the target protein.

PEAKS X PRO

An alternative strategy is the use of "*de novo sequencing*" techniques, as employed by software such as PEAKS X PRO (refer to Figures 10 and 11). This software can interpret MS/MS spectra and generate hypothetical peptide sequences without relying on a pre-existing database. Each generated sequence shows a confidence value, indicating the probability of its correctness. Additionally, individual amino acids within the sequence also possess confidence values. PEAKS generates a list of similar sequences arranged from most to least probable. This list is subsequently used to query databases and identify the corresponding protein.

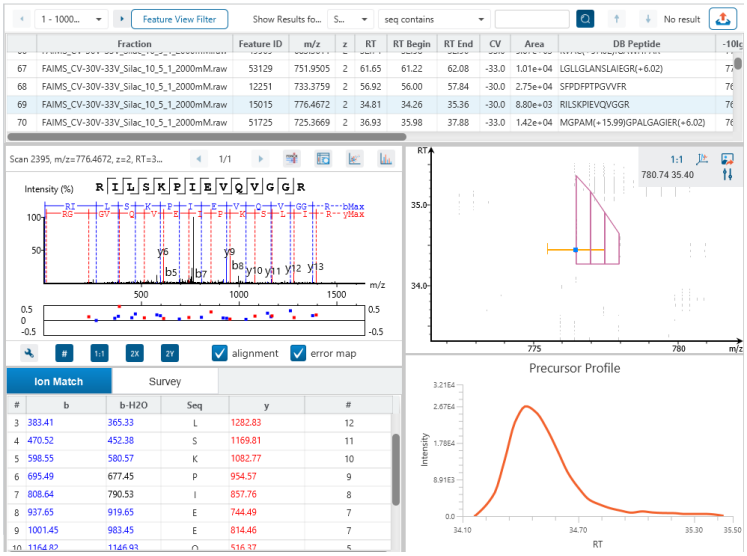


Figure 10. *De novo interpretation of an MS/MS spectrum by PEAKS.*

The *de novo sequencing* method can be advantageous over other approaches, particularly when studying PTMs, searching databases containing homologous sequences for cross-species identification, or analyzing spectra originating from mutated proteins or variants (Integrated Spider Tool). These methods enable the identification of proteins from species that are phylogenetically distant from organisms with completely sequenced genomes.

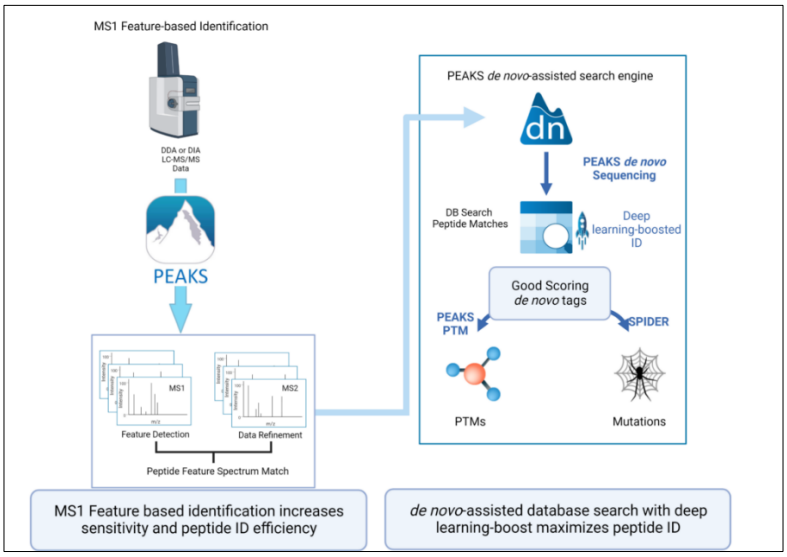


Figure 11*. Flowchart of Peaks X pro tool*

Max Quant

MaxQuant is a software package developed in the computing language C# (C sharp) for protein identification and quantification from proteomics datasets obtained through high-resolution mass spectrometry. It has its peptide database search engine, called Andromeda. A customized multilevel scoring scheme is applied to optimize the identification of peptides for each specific fragmentation technique. MaxQuant also performs a "second peptide" search, specifically looking for signals resulting from co-fragmentation of an additional precursor of the same MS/MS spectrum.

MaxQuant employs a target-decoy strategy to estimate the false-positive identifications, which includes the Posterior Error Probability (PEP), which provides statistical evidence. PEP integrates multiple peptide properties (length, charge, number of modifications, etc.) with the Andromeda score into a single quantity reflecting the quality of a Peptide-Spectrum Match (PSM). This method is similar to a machine-learning approach.[135]

In the MaxQuant interface, once you upload the raw files, the menu "Group-specific parameters" opens a form with several parameters that have to be set to obtain the identification of the protein. In this menu, you can set fixed and variable modifications, cleavage, and enzyme parameters to define your quantification mode (Figure 12). In "Identification and quantification", you must indicate parameters such as FDR and the minimal number of peptides. Figure 13, shows an overview of the computational workflow.

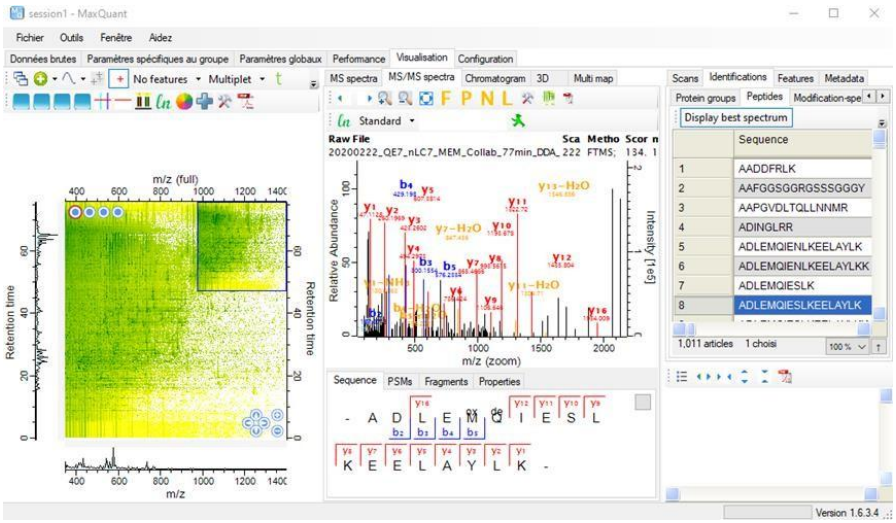


***Figure 12**. Picture of the Max Quant overview tab*

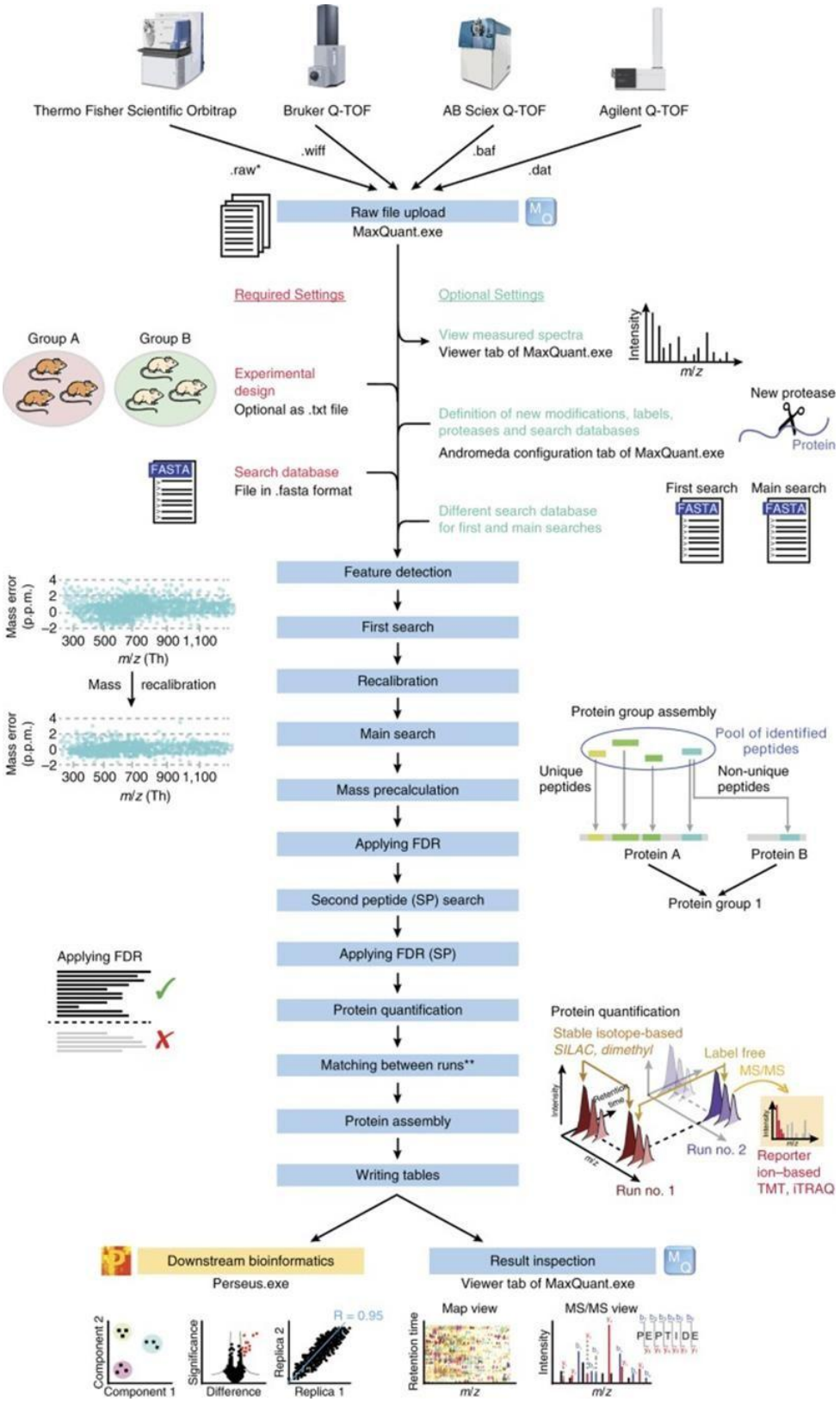


Figure 13. *Overview of the computational workflow* [80].

A similar method, with specific bioinformatics tools, is applied in the *top-down* approach.

MASH Native Explorer: A Software Platform for Native and Top-Down Mass Spectrometry Data Analysis

MASH Native (Mass Analysis Software for Heterogeneous Native Complexes) is an open-source computational platform developed for the processing and interpretation of high-resolution native and denaturing top-down mass spectrometry (MS) data. It is specifically designed to facilitate the structural and compositional analysis of intact proteins, protein complexes, and proteoforms under both native and denaturing conditions. Native mass spectrometry is a powerful analytical approach that allows biomolecules, particularly protein assemblies, to be analyzed under conditions that preserve non-covalent interactions. This technique enables the characterization of quaternary structure, subunit stoichiometry, and ligand binding, providing critical insights into the native conformational states and functional assemblies of biological macromolecules. MASH Native provides a graphical user interface (GUI) that integrates multiple algorithms and tools for comprehensive analysis of both native MS and denaturing top-down MS data. It supports a wide range of experimental workflows and is compatible with various instrument platforms. The software has multiple features:

- Support for Native and Denaturing MS data:

MASH Native is capable of processing datasets obtained under both native conditions (preserving macromolecular structure) and denaturing conditions (facilitating sequence-level analysis of intact proteins via top-down fragmentation).

- Advanced Deconvolution Algorithms:

MASH Native incorporates several deconvolution engines, including UniDec, TopFD, and MS-Deconv, which enable accurate extraction of monoisotopic or average masses from highly charged spectra. These algorithms are essential for resolving heterogeneous protein complexes, post-translational modifications (PTMs), and high-mass proteoforms.

- Proteoform Identification and Database Searching:

The software allows for protein identification by matching experimentally observed mass features with theoretical masses from protein sequence databases, as displayed in Figure 14. It supports integration with external search engines such as ProSight and MS-Align+ to facilitate sequence-resolved identification.

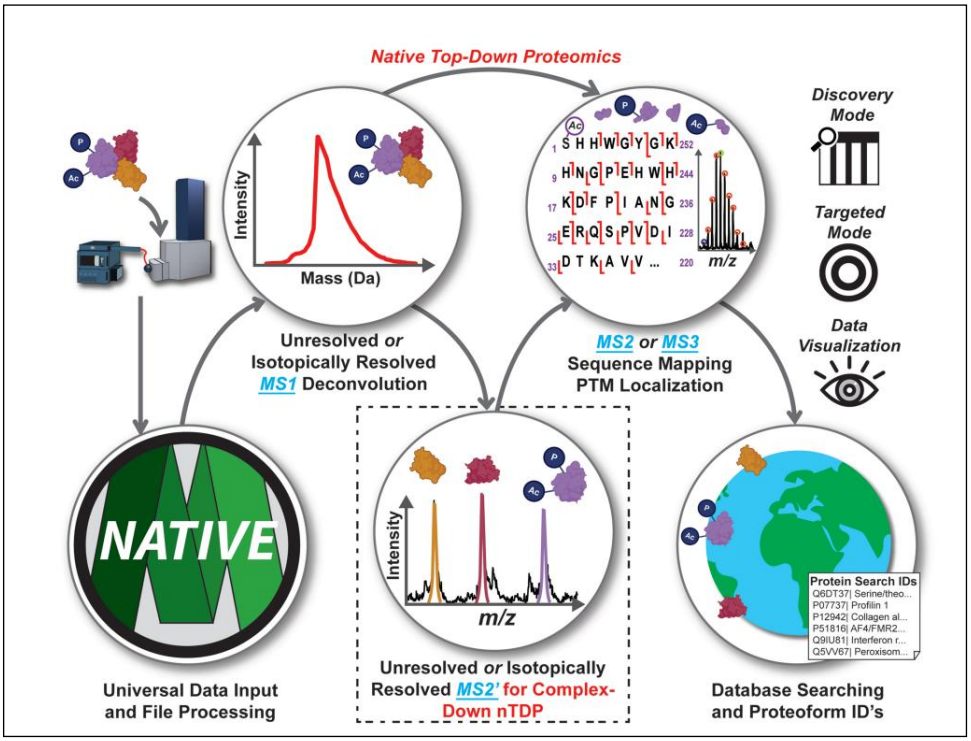


***Figure 14**. MASH Native workflow for identification of nTD and denaturing TD proteomic data* [129]

MASH Native provides a unified software solution for the analysis of a variety of complex nTDP data. As a freely available and universal processing tool, MASH Native is a "one-stop shop" for nTDP data processing that can handle a variety of complex nTDP datasets including isotopically unresolved and isotopically MS1, MS2', MS2, and MS3 in both Discovery and Targeted Modes with database search algorithms as well as data visualization and validation in a user-friendly interface.[136]

ClipsMS (Comprehensive Localization Internal Protein Sequences)

ClipsMS is a bioinformatic tool developed for the analysis of mass spectrometry (MS) data, with particular application in the field of top-down proteomics. The software is designed to process intact protein spectra, facilitating the identification and characterization of cleavage events, post-translational modifications (PTMs), and mass shifts within protein sequences. ClipsMS operates by comparing experimental MS/MS data against a theoretical fragmentation pattern derived from a given protein sequence. The software implements a *customizable scoring algorithm* to evaluate the degree of agreement between observed and theoretical fragment ions, thereby enabling accurate mapping of fragmentation sites and the localization of molecular modifications. One of the key features of ClipsMS is its ability to handle internal fragments and non-canonical cleavage patterns, which are often overlooked by conventional bottom-up proteomic tools. This functionality is crucial for the comprehensive characterization of proteoforms, especially in the context of proteins with complex structural rearrangements, multiple modifications, or non-enzymatic cleavages. The software provides both a graphical and tabular interface, allowing visualization of the alignment of fragment ions with the primary sequence of the protein, as well as exporting detailed reports for downstream analysis. Its ability to detect subtle mass differences and internal fragmentations makes it particularly suitable for studies aiming at a deep understanding of proteoform diversity and structural modifications.[137]

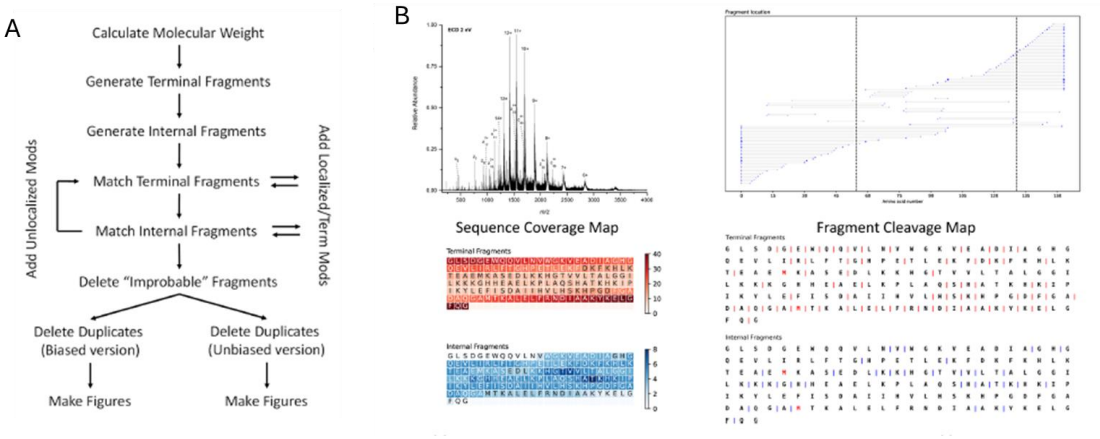


Figure 15. *(A) The workflow of the algorithm and how it matches peaks input by the user. The algorithm calculates all theoretical terminal and internal fragments, matches all peaks, decides which assignments to keep, and automatically generates figures. (B) The figures that the software produces after the analysis include Fragment cleavage maps and a sequence coverage map* [137].

Aim of the Thesis

To date, there is a lack of comprehensive studies focusing on the impact of both agronomic and environmental factors on the protein composition of chickpeas. Therefore, the main purpose of this Ph.D thesis was an improvement of the existing knowledge of the chickpea protein fraction by employing proteomic approaches already applied for other pulses or cereals. Particularly, the first aim of this Ph.D. thesis, during the first year of activity, has been the development of a "shotgun" proteomic method (i.e., the choice of the hydrolysis enzyme and the protein database), to obtain an in-depth characterization of the protein composition of an Italian chickpea genotype, namely Pascià, which is characterized by a rough, yellow and large seed type. Then, this method was also applied to investigate the qualitative and quantitative differences in protein composition between two chickpea samples of the genotype Pascià, a well-known commercial variety of kabuli type chickpea, grown under two different water conditions, namely rainfed and irrigated. The results obtained from this investigation may provide a better understanding of the factors influencing the chickpea protein composition of a specific genotype and may help to establish a connection between protein composition and various aspects, such as agronomic performance and health-related quality of chickpeas. A second goal of the present Ph.D. thesis has been the investigation of the primary structure of legumins and vicilins, of their post-translational products generated by limited proteolytic events occurring during seed maturation, and the accurate identification of the cleavage sites. This aim was achieved by applying an orthogonal approach (i.e., combining a bottom-up and a top-down method). Taking into account these aims, the results of the present Ph.D. thesis are described in three chapters. The first chapter is focused on the results of the in-depth shotgun proteomic characterization of the genotype Pascià grown using two water regimes. The results here reported allowed us to explore the plant's response, at qualitative and quantitative level, to the two different water regimes applied. The second chapter shows the results of a gel-based investigation of the entire protein fraction of this genotype. This approach highlights that, as already observed in other pulse crops, also in chickpea legumins and vicilins undergo to limited proteolytic events producing lower-mass polypeptides. The third chapter reports a detail investigation, obtained by an

orthogonal method, including a shotgun and a top-down approach, of the primary structure of these proteolytic products together the accurate determination of the cleavage sites in the precursor legumins and vicilins. Finally, chapter four describes a proteogenomic approach designed to integrate the proteomic data generated in this work with the chickpea genomic information published in recent years, to corroborate the proteomic findings, including the validation of protein annotation entries. This integrative analysis specifically targets seed storage proteins, which constitute the predominant fraction of chickpea seed proteins.

.

Chapter 1

In-depth Shotgun Proteomic Characterization of Chickpea (*Cicer arietinum* L.). A case study on the genotype Pascià grown using two water regimes

The results of this study are published in the following scientific manuscript:

Di Francesco A., Lanzoni A., De Santis M.A., Pittalà M.G.G., Saletti R., Flagella Z., Cunsolo V. In-depth Shotgun Proteomic characterization of chickpea (Cicer arietinum L.). A case study on the genotype Pascià grown using two water regimes. ACS Agricultural Science & Technology. 2025. Doi:10.1021/acsagscitech.4c00705

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is globally recognized as one of the most critical pulse crops, playing a vital role in global food security due to its high protein and nutritional value. The understanding of seed protein fraction is crucial not only to the seed's quality but also to the plant's metabolic and physiological response mechanisms to environmental conditions. Despite its importance, a deep investigation of chickpea seed proteome, until a few years ago, remained partial due to analytical challenges related to sample complexity and the limited availability of comprehensive protein sequence databases. In this context, the application of proteomic protocols is essential to fill these knowledge gaps. This chapter reports: i) the optimization of the protein digestion protocol aimed to obtain an in-depth characterization of the chickpea seed protein fraction; ii) the results about a comparative proteomic investigation of two samples of chickpea seeds from Pascià genotype grown in an open field under two water regimes, namely rainfed and irrigated. To obtain these two aims, a shotgun proteomic approach was carried out. Particularly, by coupling the high resolution and high mass accuracy of mass spectral data, along with the use of a database restricted only to *Cicer arietinum* entries (reviewed and unreviewed) significantly enlarged the actual knowledge of the seed chickpea proteome by identifying more than 700 novel proteins previously undetected. Moreover, this analytical approach, also including a label-free quantification (LFQ), allowed us to explore the plant’s response, at qualitative and quantitative level, to the two different water regimes applied in this study.

MATERIALS AND METHODS

Chickpea sample collection

Chickpea samples were collected from a field trial conducted under rainfed and deficit irrigation conditions; the details of the experiments and the main preliminary agronomic results are reported in De Santis et al. (2022).[138] In particular, samples of the Pascià genotype coming from the rainfed and irrigated experiment conducted in South Italy (Troia, Apulia Region, Italy) were considered in this study. The field trial was conducted on loam soil (field capacity 34.5%, wilting point 16.4%). Seasonal cumulative precipitations were 45 mm, while irrigated samples received an additional 152 mm. At sowing, both irrigated and rainfed plots were taken to field capacity, while at harvest, volumetric soil moisture was 15.7% (rainfed, below wilting point) and 23.5% (irrigated, above wilting point). Potential evapotranspiration was 408 and 445 mm for rainfed and irrigated plots, respectively, while actual evapotranspiration was 287 and 346 mm. As regards crop nutrition, chickpea samples were fertilized with 40 kg/ha of nitrogen and 100 kg/ha of phosphorus. Rainfed and irrigated chickpea samples achieved maturity at 94 and 100 days after sowing, respectively. An aliquot (500 g) from each experimental plot was collected and ground (laboratory mill) for chemical analysis. The evaluation of the seed proteome was carried out on three biological replicates.

Chemicals

All chemicals were used without further purification because of the highest purity commercially available. Porcine trypsin was purchased from Promega (Madison, WI, USA). Formic acid (FA), ammonium bicarbonate (AMBIC), 1,4-dithiothreitol (DTT), iodoacetamide (IAA), Tris-HCl, hexane, lysozyme from chicken egg white, and EDTA were obtained from Aldrich (St. Louis, Missouri, USA). Water and acetonitrile (ACN) (OPTIMA® LC/MS grade) for LC/MS analyses were purchased from Fisher Scientific (Milan, Italy). RapiGest® SF was provided by Waters Corporation (Milan, Italy).

Defatting and extraction of proteins

Chickpea seeds were ground into powder, and then the flour obtained was defatted using hexane of analytical grade to eliminate the lipid fraction from the chickpea flour by solvent extraction at a ratio of 1:3 weight: volume (flour: solvent) for 1 h under stirring at 200 rpm. Subsequently, the mixture was filtered and dried under an extractor hood at room temperature for 24 hours. The defatting process of the chickpea flour was repeated twice. Then, seed proteins (with the exclusion of prolamins, which require alcohol solutions) from each biological replicate were extracted from the chickpea flour with a dedicated buffer (50 mM Tris–HCl, pH 7.8, 5 mM EDTA, 0.1% 1,4-dithiothreitol). Briefly, 100 mg of ground seed flour was suspended with 1 mL of extraction buffer for 1 h at room temperature with constant stirring, and then centrifuged at 10,000×g for 30 min. The supernatant, containing the extracted proteins (i.e., albumins, globulins, and glutelins), was used to prepare the six biological replicate samples (three for the rainfed condition and three for the irrigated condition) for LC-MS/MS analyses as reported in figure 16.

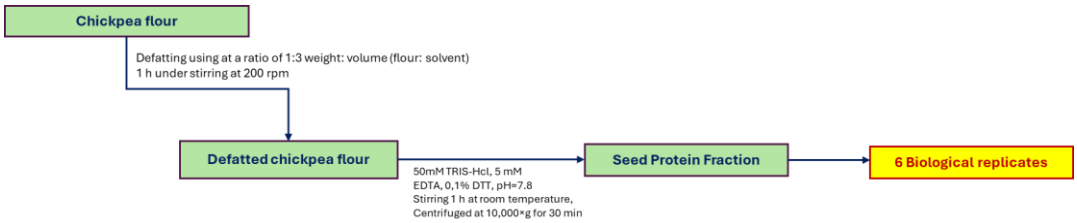


Figure 16. *Workflow of the extraction methods employed for protein extraction.*

The protein concentration in the supernatant was determined using the Qubit Protein Assay kit and the Qubit® 1.0 Fluorometer (ThermoFisher Scientific, Milan, Italy). [139] Despite this, it is fair to point out that we are aware that the extraction protocol could not be totally exhaustive, but it was chosen based on previously published studies that tested the effectiveness of the methodology.

Optimization of the Protein Digestion Protocol

The protein extract obtained by the chickpea flour sample grown under rainfed condition was used to test five different protein hydrolysis protocols. In detail, a solution of RapiGest® 0.1% was added to the protein extract. Then, it was suspended in a solution of ammonium bicarbonate (AMBIC) at pH=8, reduced using DTT (3 hours at 25°C), alkylated with IAA (1 hour at 25°C). Then the sample was freeze-dried, divided into five aliquots and finally enzymatically digested using five different protein hydrolysis protocols. In detail, protein extracts were digested by: i) trypsin, ii) chymotrypsin iii) trypsin plus chymotrypsin, iv) trypsin plus trypsin, and v) trypsin plus Endoproteinase GluC. For the samples digested with trypsin, the material was solubilized in an AMBIC solution at pH 8 and incubated at 37 °C overnight, using an enzyme-to-substrate ratio of 1 µg:50 µg. For the sample treated with trypsin plus an additional trypsin digestion, the material was likewise solubilized in AMBIC at pH 8. Digestion was carried out in two steps: the first incubation was performed at 37 °C for 4 h, after which a fresh aliquot of trypsin was added and the sample was incubated again overnight at 37 °C, maintaining an enzyme-to-substrate ratio of 1:50 for both steps. The sample treated with trypsin plus endoproteinase GluC was also solubilized in AMBIC at pH 8. After the addition of trypsin, the sample was incubated at 37 °C for 4 h; following the addition of GluC, digestion proceeded overnight at 37 °C, with an enzyme-to-substrate ratio of 1:50 for both proteases. The sample treated with trypsin plus chymotrypsin was resuspended in AMBIC at pH 8, incubated with trypsin at an enzyme-to-substrate ratio of 1:50 for 4 h at 37 °C, and subsequently supplemented with chymotrypsin at a 1:20 ratio, followed by overnight incubation at 25 °C. The sample treated with chymotrypsin alone was solubilized in 100 mM Tris-HCl and 10 mM $CaCl_2$ at pH 8, and incubated overnight at 25 °C. Two-step enzymatic digestion was used to ensure greater digestion efficiency and to allow the use of two different digestive enzymes that could increase protein identification.

Then, each enzymatic peptide mixture was analyzed by nUPLC-nESI MS/MS, and MS data used to search against a chickpea protein database (31239 entries from UniProt) to obtain the best protein profiling of the chickpea extracts. The results obtained by the five different protein hydrolysis protocols employed were compared to find the most suitable enzyme of hydrolysis.

LC-MS/MS Analysis of the peptide mixtures

Chicken lysozyme C (0.25 µg) was added as an internal standard for the label-free analysis to 50 µg of each of the six samples. Then, each protein extract solution was reduced, alkylated, and finally, enzymatically digested with porcine trypsin,[140] which provided the best results in terms of protein identification among the digestion protocols tested. Particularly, 39 µg of DTT (3 hours at 20 °C) and then 94 µg of IAA (1 hour in the dark at 20 °C) were used for the reduction and alkylation, respectively. Protein hydrolysis was obtained by porcine trypsin at an enzyme: substrate ratio of 1:50 (overnight at 37 °C). To achieve a final concentration of 25 ng/µL for each sample, and of 0.125 ng/µL for Chicken lysozyme (as an internal standard for the label-free quantification), a 5% aqueous solution of FA was added, resulting in a final volume of 2 mL. Then, a clean-up step was carried out. Particularly, 300 µL of each sample solution was filtered using 0.2 µm regenerated cellulose filters (ex Alltech, SepaChrom, Milan, Italy). MS/MS data were acquired on a ThermoFisher Scientific Orbitrap Fusion Tribrid® (Q-OT-qIT) mass spectrometer (ThermoFisher Scientific, Bremen, Germany). Liquid chromatography was carried out using a ThermoFisher Scientific Dionex UltiMate 3000 RSLCnano system (Sunnyvale, CA). LC-MS/MS analyses were carried out by loading one microliter, corresponding to 25 ng, of peptide mixture onto an Acclaim ®Nano Trap C18 Column (100 µm i.d. × 2 cm, 5 µm particle size, 100 Å). The detailed description of the chromatographic conditions and the MS parameters were the same as in Pinginelli et al. (2023).[141] Mass spectra of all the peptide precursor ions were acquired at 120,000 resolution at m/z 200, in the range m/z 200-1600. MS/MS analyses were performed by HCD fragmentation and rapid-scan MS fragment ion analysis in the linear ion trap (low-resolution MS/MS analysis). To assess the reproducibility of MS data, the three biological replicates of each water regime were subjected to triplicate RP-nHPLC/nESI-MS/MS analyses, giving rise to nine runs for the sample condition. Three blank runs, carried out using the same gradient program, were performed before the first replicate of each sample.

Database Search Analysis

The MS data obtained as above described were processed by the PEAKS X software (https://www.bioinfor.com/de-novo-sequencing/; Bioinformatics Solutions Inc., Waterloo, ON Canada) and searched against a protein database including all the reviewed and unreviewed entries of *Cicer arietinum L.* (chickpea) downloaded from the UniProt database (Swiss-Prot and TrEMBL sections, release February 2023, 31239 entries; https://www.uniprot.org/help/release-statistics). The common Repository of Adventitious Proteins (c-RAP) contaminant database (www.thegpm.org) was included in the database search to identify protein contaminants. All database searches were carried out using the following parameters: i) full tryptic peptides with a maximum of 3 missed cleavage sites; ii) oxidation of methionine, acetylation (N-terminal protein), and transformation of N-terminal glutamine and N-terminal glutamic acid residue in the pyroglutamic acid form as variable modifications; iii) carbamidomethylation of cysteine as a fixed modification. The precursor mass tolerance threshold was 10 ppm, and the max fragment mass error was set to 0.6 Da. Then, the following filters were used to validate the results. In detail, Peptide Spectral Matches (PSMs) were validated using a Target Decoy PSM Validator node based on q-values at a False Discovery Rate (FDR) $\leq$ 0.1%. PEAKS score thresholds for PSMs were set to achieve for each database search, FDR values for PSMs, Peptide sequences, and Proteins identified below the 0.1% value. Moreover, a protein was considered confidently identified if a minimum of three peptides (including at least a unique peptide) were matched, and the sequence coverage was $\geq$ 5%. Last, to produce the final list of the identified proteins in each sample condition, only the proteins identified in at least five out of nine LC-MS/MS technical replicates were considered.

Relative internal protein quantification

For both samples investigated, the relative abundance of each protein identified was calculated. This result was achieved starting from the raw values of the “Sample Area” of each protein reported in the PEAKS panel. The raw protein values of “Sample Area” are calculated by the software PEAKS using the total of all peptide features from unique supporting peptides. The lysozyme from chicken egg white was used as an internal standard to normalize the raw “Sample Area” of each protein.

Then, for each LC-MS run, the relative abundance of each protein (reported as a percentage) was determined. Finally, for each protein, the average of the relative abundance and its standard deviation was calculated (see Supplementary Table S3).

Label-Free Quantification and Technical Validation

Label-free quantification (LFQ) analysis was performed by the PEAKS Q module of the PEAKS X software (https://www.bioinfor.com/de-novo-sequencing/; Bioinformatics Solutions Inc., Waterloo, ON, Canada), which uses ion peak intensity on MS1. PEAKS Q quantified proteins, selecting the three most abundant unique peptides for protein quantification by excluding peptides with both modified and unmodified forms and redundant peptides. Proteins quantified with less than three unique peptides were discarded. The detailed description of the quantification method is reported in Lin et al. (2013).[142] The following parameters were set for label-free quantification analysis: mass error tolerance (i.e., the mass shift between different runs), 10 ppm; retention time (RT) shift tolerance, 3 min (i.e., the maximum elution time range considered for the quantification of an identified peptide between different runs). The only peptide features having the following parameters were considered for the quantification: quality ≥ 20 (a parameter depending on m/z difference, RT difference, isotope distribution, etc.); average area of the MS signal intensity ≥ 107; peptide charge 2, 3, or 4. A protein was considered quantified when: i) at least three of its unique peptides satisfied the parameters above reported; ii) was identified in a minimum of five out of nine LC-MS/MS runs per flour sample; and iii) had a p-value < 0.05 and a significance ≥ 20 (significance method ANOVA; the significance score was calculated as the -10log10 of the significance testing p-value; a significance score threshold of 20 corresponds to a significance testing p-value of 0.01). As reported above, the lysozyme from chicken egg white was used as an internal standard to calculate a normalized factor. For each quantified protein, the normalized abundance was calculated from the raw abundance divided by the normalization factor. Then, to estimate the run-to-run reproducibility across both the technical and biological replicates, the pairwise Pearson correlation coefficients, using the Log10 [protein abundance], were calculated by the PEAKS Q module. Technical replicates determining a Pearson correlation coefficient below 0.85 were

discarded, and the LFQ analysis was carried out using the remaining LC-MS runs (see Supplementary Material). Finally, in the pairwise comparison of rainfed vs. irrigated, the irrigated condition was chosen as reference, and a protein was considered differentially abundant when it showed a fold-change ≤ 0.5 (i.e., it was up-regulated in the irrigated condition) or ≥ 2 (i.e., it was up-regulated in the rainfed condition).

Gene Ontology

Gene Ontology (GO) term enrichment analysis was carried out to statistically find over-represented categories of proteins. GO is a widely adopted bioinformatics framework that provides a standardized vocabulary for describing the roles of genes and gene products across different species. It encompasses three main domains: *biological process*, which refers to the broader cellular or physiological activities in which a gene product is involved; *molecular function*, which captures the specific biochemical activity of the gene product at the molecular level; and *cellular component*, which indicates the subcellular location or complex where the gene product exerts its function. The ontology is structured hierarchically, enabling the annotation of gene functions at varying levels of specificity. GO is essential for functional genomics, supporting the interpretation of large-scale omics data and enabling comparative analyses across organisms. GO analysis was carried out using the open-source "Biological Network Gene Ontology" tool (BiNGO, 3.0.5)[143] as a plugin for Cytoscape 3.9.1.[144] The ontology file in OBO Flat File format 1.4 (release 14-01-2024) and the Cicer arietinum annotations file in GAF format (release 14-01-2024) were downloaded from Gene Ontology (http://geneontology.org/docs/download-ontology/) and Gene Ontology Annotation (https://www.ebi.ac.uk/GOA/proteomes) websites, respectively. The whole Cicer arietinum GOA annotation file was used as a reference set, and GO terms referring to the cellular component terms were searched. The hypergeometric test was selected as a statistical test, and Benjamini & Hochberg False Discovery Rate (FDR) correction as a multiple testing correction. To visualize, by Cytoscape, the over-represented categories after correction, a significance level (p-value) of 0.05 was set. In Cytoscape, yFiles Hierarchical Layout was selected, and finally, the most significant nodes were extracted from the entire network and displayed in the graphs. The GO analysis was separately carried out for the group of proteins identified only in the rainfed sample and the group of proteins exclusively identified in the irrigated sample. Given the current status of GO annotation efforts, a lot of genes don't actually have annotations in the GO hierarchy. The genes for which there is no annotation available are not taken into account in the over-representation analysis, and they're not counted as genes belonging to the test or reference set. The detailed results of the GO analysis are reported in the Supplementary Material.

RESULTS AND DISCUSSION

Choice of the digestion protocol and the protein database

Vicilins, legumins, and other chickpea proteins with metabolic and structural functions have already been shown to resist in vitro simulated digestion.[108] Therefore, their identification by shotgun MS-based methods might be difficult. On the other hand, the benefits of using multiple proteases for the cleavage of proteins before MS analysis have already been reported.[145] Consequently, the first step of this PhD project was to test the efficiency of different protein digestion protocols (i.e. using different proteases) to identify the most suitable one. Particularly, different digestion enzymes were tested: trypsin, trypsin plus trypsin trypsin plus chymotrypsin, and trypsin plus endoproteinase GluC. The peptide mixtures obtained by these digestion protocols were characterized by high-resolution nUPLC-nESI MS/MS analysis. MS data were searched against a protein database including all the reviewed and unreviewed entries of *Cicer arietinum L.* (chickpea) downloaded from the UniProt database.

The results of these different digestion protocols are summarized in Table 1

***Table 1**. Results in terms of proteins identification using the five digestion protocols described in the text.*

Digestion protocol	Proteins Identification
Trypsin	974
Chymotrypsin	27
Trypsin + Chymotrypsin	<10
Trypsin + Endoproteinase GluC	<10
Trypsin + Trypsin	290

Overall, the results showed in Table 1 highlighted that the single-step of trypsin digestion, as also reported by Riberio et al[108], represents, among the digestion protocols here employed, the most suitable enzyme to characterize the protein fraction of seeds chickpeas. So that, trypsin was selected as enzyme of hydrolysis for the proteomic characterization of both chickpea samples.

Qualitative and Internal Quantitative comparison analysis: rainfed vs irrigated condition

The shotgun proteomics approach applied to the protein fractions extracted from chickpea seeds cultivated under two distinct water regimes enabled the confident identification of 974 proteins in the rainfed sample and 1009 proteins in the irrigated sample. A comprehensive list of all identified proteins is provided in Supplementary Table S1, while the corresponding peptide sequences that supported confident protein identification are detailed in Supplementary Table S2. Particularly, Table S1 reports for each identification the UniProtKB Acc. No. and the Protein description, which includes the organism GeneName (GN), the OrganismIdentifier (OX), the ProteinExistence (PE), and the SequenceVersion (SV). Moreover, two additional columns show: i)whether the protein was identified in the rainfed or irrigated sample; and ii) if the protein was already identified in chickpea tissues (leaves, seeds, roots) with the corresponding bibliographic reference.

The results of the internal quantitative analysis, carried out as reported in the Materials and Methods section, are described in detail in the Supplementary Material (Figure S1 and Table S3). This analysis reveals that both the chickpea samples investigated show only about twenty proteins (corresponding to 2% of the total) with a relative abundance above the value of 1% and that all together constitute about 75-80% of the total abundance of the protein fraction. This group represents, in both samples, the most abundant proteins. A second group, comprising roughly fifty proteins (about 5% of the total), shows relative abundance values ranging from 0.1% to 1% and may be categorized as "minor proteins." The majority of the identified proteins, representing approximately 92–93%, exhibit relative abundances below 0.1% and are thus considered the "low-abundance proteins" group. The most abundant components, in both sample conditions, are the same and are mainly constituted, as expected, by storage proteins (i.e., legumins, vicilins, and albumins). However, some metabolic proteins (i.e., NADPH-dependent aldehyde reductase 1-chloroplastic-like, Lipoxygenase, Peroxiredoxin, etc.) belong to the most abundant components. Altogether, these results highlight a similar qualitative distribution of protein composition in both sample conditions.

As illustrated by the Venn diagram in Figure 17, a comparative qualitative analysis of the two protein fractions reveals a total of 1111 distinct protein species identified

across both conditions. A substantial group of proteins (872 components, corresponding to 78.5%) is shared between the rainfed and irrigated samples. Meanwhile, 102 proteins (9.2%) were uniquely detected in the rainfed sample, and 137 proteins (12.3%) were exclusive to the irrigated sample.

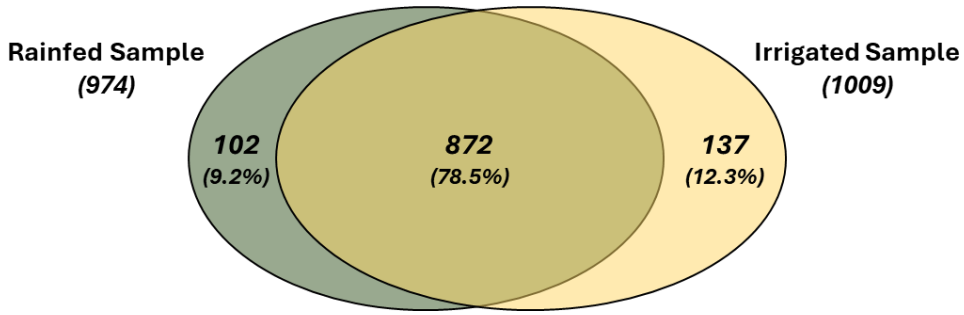


Figure 17*. Venn diagram giving the statistics of the confidentially identified proteins. Contribution to the total identifications as obtained in the rainfed and irrigated samples (the per cent values are also reported). The complete list of the identified proteins is reported in the supplementary Table S1.*

It is worth noting that proteins shared between the two samples encompass all of the most abundant and minor components, as well as a substantial number of low-abundance proteins. In contrast, all proteins exclusively identified in one condition, namely, the 102 unique proteins in the rainfed sample and the 137 unique proteins in the irrigated sample, belong exclusively to the category of low-abundance proteins, collectively accounting for approximately 0.1% of the total protein content in each respective sample.

Label-Free Quantification Analysis

The label-free quantification (LFQ) analysis was initially carried out using eighteen LC-MS runs. To estimate the run-to-run reproducibility, across both the technical and biological replicates, the pairwise Pearson correlation coefficients were also derived by the use of the Log10[protein abundance]. The correlation plots among the technical and biological replicates of the rainfed and irrigated groups are shown in Supplementary Figures S8 and S9. These plots evidenced that replicate A3 of the sample grown in irrigated conditions should be discarded because it showed a weak

correlation with the other replicates (see the details in the Supplementary Material). Consequently, the LFQ analysis was carried out again, but excluding this technical replicate. The scatter plots depicting the Pearson correlation coefficients between technical replicate and biological replicate of this new analysis are reported in the Supplementary Figures S10 and S11 and show an optimum linearity over three orders of magnitude in protein abundance (see the details in the Supplementary Material). In detail, the LFQ analysis highlighted the existence of fourteen differentially abundant proteins (DAPs), which are all more abundant in the sample grown under rainfed conditions. On the contrary, no proteins were up-regulated in the irrigated sample condition. The Protein profile heat map of the LFQ Analysis and the list of the fourteen DAPs are reported in Figure 18 and Table 1, respectively.

Five out of fourteen DAPs are storage proteins, six components are enzymes, and the last three belong to the family of Late Embryogenesis Abundant (LEA) proteins (Figure 19). However, it should be highlighted that the protein profile heat map also shows that for two DAPs, i.e., the NADPH-dependent aldehyde reductase 1 chloroplastic-like (Acc. No. A0A1S2XVM2) and the protein disulfide-isomerase (Acc. No. A0A1S2YBR2), a technical replicate of the rainfed group samples reveals a low reproducibility with respect to the others.

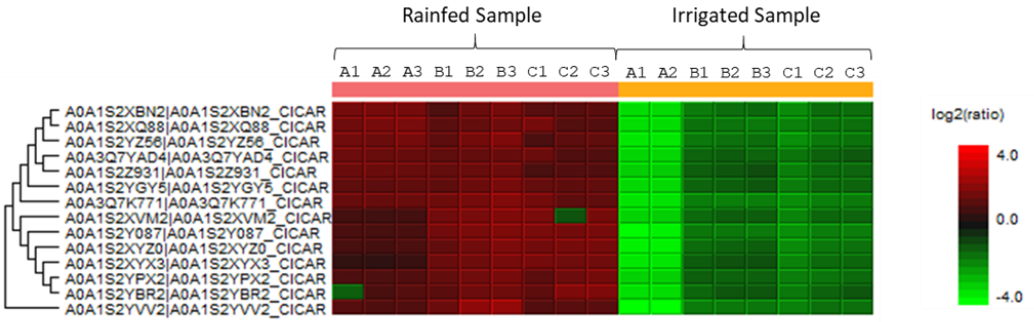


Figure 18. *Protein profile heat map of the Label-Free Quantitative (LFQ) Analysis. Cell color represents the fold change of the Differentially Abundant Proteins (DAPs) observed between the two samples investigated (rainfed vs. irrigated). The irrigated condition was chosen as the reference. Three biological replicates (named A, B, and C) per sample condition were investigated. Each biological replicate was subjected to triplicate RP-nHPLC/nESI-MS/MS analyses (e.g., A1, A2, and A3), giving rise to nine runs for each sample condition. The technical replicate A3 of the sample grown in irrigated conditions was discarded because it showed a weak reproducibility in comparison with the other replicates and, therefore, it wasn't used in the LFQ analysis (see the main text and the Supplementary Material for the details).*

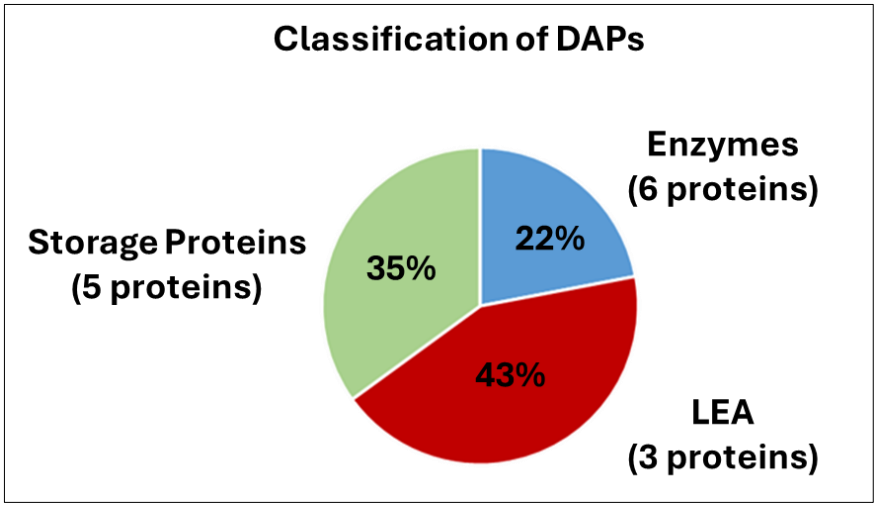


Figure 19. *Pie chart showing the distribution (%) of the differentially Abundant Proteins (DAPs) up-regulated in the rainfed sample***.**

Table 2*. List of the fourteen Differentially Abundant Proteins (DAPs) identified by the label-free quantification analysis carried out between the rainfed and irrigated conditions (rainfed vs irrigated). All the DAPs were up-regulated in the rainfed condition. Columns show the UniProt Accession number, protein description, significance score, number of total matched peptides, number of unique peptides, fold change (rainfed vs. irrigated), p-value, and protein classification (P-class).*

Acc. No.	Description	Significance	Peptides	Unique	Fold Change (rainfed vs irrigated)	p-value	P-class
A0A1S2Y087	vicilin-like	25,20	17	14	3,33	3E-07	storage
A0A3Q7K771	albumin-2-like	30,24	10	10	3,36	6E-09	storage
A0A1S2YVV2	desiccation protectant protein Lea14 homolog	21,71	6	6	3,11	3E-05	LEA
A0A1S2YZ56	Vicilin-like seed storage protein At2g18540	26,35	5	5	3,04	1E-07	storage
A0A1S2XYZ0	provicilin-like	22,26	19	13	2,82	2E-06	storage
A0A1S2XQ88	vicilin-like	27,59	31	5	2,91	8E-08	storage
A0A1S2XBN2	Lipoxygenase	27,35	13	11	2,84	2E-07	Enzyme
A0A1S2YPX2	embryonic protein DC-8-like	25,87	3	3	2,49	9E-07	LEA
A0A1S2XYX3	late embryogenesis abundant protein D-34-like	20,33	5	5	2,41	2E-05	LEA
A0A1S2XVM2	NADPH-dependent aldehyde reductase 1 chloroplastic-like	25,29	12	10	2,46	6E-05	Enzyme
A0A1S2YGY5	Peroxiredoxin	23,11	6	6	2,41	5E-07	Enzyme
A0A1S2YBR2	Protein disulfide-isomerase	20,51	5	5	2,32	1E-04	Enzyme
A0A3Q7YAD4	Lipoxygenase	22,89	9	9	2,35	8E-07	Enzyme
A0A1S2Z931	uncharacterized protein L	24,08	3	3	2,30	1E-06	Enzyme

Advancing the Understanding of Chickpea Seed Proteome: A Comprehensive Update

To date, several studies have focused on the proteomic analysis of chickpea tissues, such as leaves, seeds, and roots, to investigate the effects of various stress conditions, including drought, heat, and salinity, across different chickpea genotypes. Collectively, these investigations have led to the identification of approximately 1136 distinct protein components specifically associated with the *Cicer arietinum* species,[146–151] with about one hundred identified in more than one tissue, and 549 proteins detected in seeds.[149,151] In the present study, using a shotgun approach, a total of 1111 different proteins have been identified (Table S2). In comparison to the 549 proteins of *C. arietinum* species already detected in chickpea seeds, in the present study, only 141 out of the 549 already published components were not identified, whereas 408 were common components. Instead, the largest group of proteins here identified, constituted by 703 components, were not previously reported for chickpea allowed the expansion (Figure 20). Overall, this result clearly demonstrates that, although no alternative or complementary extraction protocols were adopted, nor additional digestion protocols (i.e., with different proteases) were used, the largest group of proteins here identified enabled a substantial expansion of literature data about chickpea seed proteome.

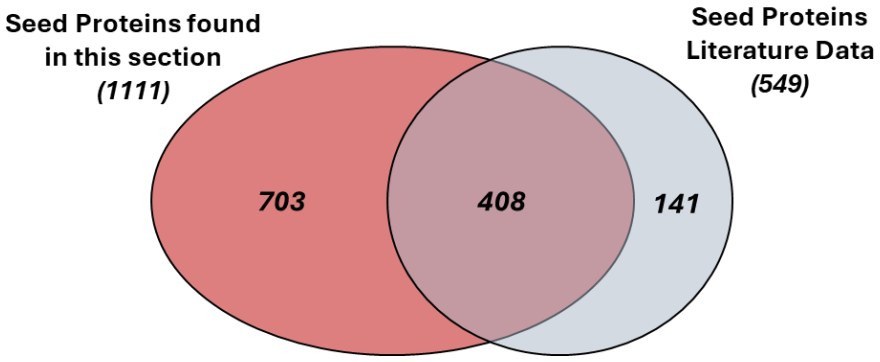


Figure 20. *Venn diagrams giving the statistics about the comparison of the list of proteins found in the present study and the inventory reported protein list from the chickpea seeds investigation, obtained by merging all data from the literature.* [149,151–153]

This proteome discovery is almost certainly related to the joining of important analytic strategies. Primarily, to our knowledge, the present work reports for the first time the application of a shotgun approach to investigate the chickpea seed proteome. Indeed, to date, all the literature data about the chickpea seed proteins have always been obtained by gel-based approaches. Gel-based techniques, as known, have a significant number of limitations; they are time-consuming, costly, insensitive to low-abundant proteins, and in many cases are incapable of separating all components present in a complex protein mixture. Most of these limitations are addressed by shotgun proteomics. Moreover, comparative 2D-gel experiments are usually only aimed at the identification of those proteins having different abundances (i.e., those responsible for the differentially regulated spots detected in the gel) between the samples investigated. Instead, in the shotgun approach, it is essential to know protein identity before quantification, as peptides need to be related to each other and their parent proteins. Therefore, also those proteins that do not show significant differences in abundance between the samples investigated are usually identified and classified, contributing to enlarging our knowledge about the proteome composition. On the other hand, it is clear that both gel-based and gel-free approaches, with their respective advantages and disadvantages, remain complementary techniques and should be used in parallel to get a more complete comprehension of protein expression. This huge protein discovery has also been made possible because of the very high scan rate, very sensitive high resolution, and high mass accuracy of the Orbitrap Fusion Tribrid® mass spectrometer used to obtain mass spectral data. Finally, it cannot be denied the importance of the database used to identify the proteins by searching MS data. Most of the data about the chickpea proteome (particularly, seed proteome) previously published have been obtained using large-scale databases (e.g., using the taxonomy *Viridiplantae*), because the chickpea genome sequence was not known. Therefore, a lot of proteins have been identified by homology-based searching. It has been demonstrated that the choice of the sequence database to use has a strong impact on the results of the search. For instance, if some proteins in the sample are not included in the chosen database, peptides from these unexpected proteins and their corresponding MS spectra may remain unmatched or may be incorrectly matched to other proteins in the database. Moreover, too large sequence databases affect the search sensitivity because both increase the search space and the number of spurious identifications.[154] So, it is

recommended to tailor the database to the species under study when possible. Otherwise, closely related species can also be used to create custom databases. In this respect, the publication of the chickpea genome in 2015 and the consequent update of the protein sequence entries (i.e., 80% of the *Cicer arietinum* protein entries available in UniProt have been published in 2017) represent fundamental steps to increase the accuracy and the confidence of protein identifications, also including the so-called “low-abundance proteins” (i.e. proteins present in traces).

Proteomic Analysis of Chickpea Seeds Under Two Water Regimes: A Case Study

Overall, the seed protein profile of the Pascià genotype grown under two different water regimes (rainfed and irrigated) showed a high degree of similarity, with approximately 78.5% of the identified proteins shared between both conditions. Nevertheless, despite being classified as "low-abundance proteins", 102 and 137 proteins were uniquely detected in the rainfed and irrigated samples, respectively. Furthermore, LFQ analysis revealed that 14 DAPs were significantly up-regulated in the rainfed condition, whereas no proteins showed increased abundance under irrigation.

Unique proteins of the rainfed and irrigated sample conditions

About the proteins exclusively identified in the irrigated sample, the Gene Ontology (GO) Enrichment Analysis evidenced that they mainly play tRNA-ligase and translational regulator activities and are mostly involved in amide and peptide biosynthetic processes (see Supplementary Material). The over-representation of the above-listed GO categories is principally related to a group of Eukaryotic translation initiation factors (eIFs), some ribosomal proteins, and tRNA-ligases (also known as aminoacyl-tRNA synthetases, ARSs), all proteins involved in protein synthesis (i.e., translation process). It is important to evidence that a lot of the proteins exclusively identified in the irrigated sample represent isoforms of proteins also detected in the rainfed sample, but having a different UniProt Accession number (i.e., 40S ribosomal protein S5, S7, and S12; Eukaryotic translation initiation factor 5A, eIF-5A; phenylalanine-tRNA ligase). On the other hand, many of the unique proteins of the irrigated sample do not have the isoform counterpart in the rainfed one (e.g., 60S ribosomal protein L17, L36, P1, and P3; Eukaryotic translation initiation factor subunit 2D, 3H, 3I, 3F, and 4G; threonine-tRNA ligase, and tryptophan-tRNA ligase). ARSs, eIFs, and ribosomal proteins have an essential role in protein synthesis (i.e., the translation process), a fundamental process for plant life, and play a crucial role in the adaptation to energy, developmental, and environmental conditions. Protein synthesis is regulated by the environmental factors that constrain the uptake of CO2 and nutrients, such as high temperature and water potential. Particularly, different studies have unveiled how plants utilize translational control to defend themselves against viruses, regulate translation in response to metabolites, and reversibly adjust translation to a wide variety of environmental parameters.[155] So that these groups of unique proteins of the irrigated sample and lacking in the rainfed one may represent the main response to the greater water supply. In addition, it is interesting to note that different studies highlighted that, in addition to the synthesis of proteins, the Eukaryotic translation initiation factor subunit (eIF4G), a unique protein of the irrigated sample not having the isoform counterpart in the rainfed one, in rice, plays a role in the resistance against Tungro spherical virus.[156] Moreover, among the unique proteins characteristic of the irrigated condition and involved in amide and peptide biosynthetic processes, two enzymes required for branched-chain amino acid (BCAAs) biosynthesis were found: the 3-isopropylmalate dehydrogenase

(IMDH; Acc. No. A0A1S2XB48) and the dihydroxyacid dehydratase (DHAD; Acc. No. A0A1S2XCF6). These two enzymes can increase the amount of BCAAs, which several studies have shown contribute (together with their derivatives) to the growth, defense, and development of the plant's roots.

Instead, the GO Enrichment Analysis of those proteins exclusively identified in the rainfed sample evidenced that most of them belong to ribonucleoprotein complex, proteasome complex, ribosome, and cytoplasm, show oxidoreductase activity, acting on the aldehyde or oxo group of donors, and are involved in primary metabolic processes, and metabolic processes of organic substances (see Supplementary Material). As already observed for the unique proteins of the irrigated sample, a lot of proteins exclusively identified in the rainfed sample are isoforms of proteins also detected in the irrigated one, but with a different UniProt Accession number (i.e. 60S ribosomal protein L13, 40S ribosomal protein S16-like, 26S proteasome non-ATPase regulatory subunit 12 homolog A-like, Delta-1-pyrroline-5-carboxylate synthase, Cysteine synthase). It is interesting to highlight that some of these proteins (e.g. 60S ribosomal protein L13, Acc. No. A0A1S2YZI2; cysteine synthase, Acc. No. A0A1S2YX59; the peptidylprolyl isomerase, Acc. No. A0A1S2XN76) were also identified in the irrigated sample, but were not considered statistically significant identifications (e.g., because not detected at least in five out of nine LC-MS/MS technical replicates) and therefore not included in the final list of identifications. Instead, more interesting are a lot of exclusive protein isoforms of the rainfed sample (e.g., delta-1-pyrroline-5-carboxylate synthase, Acc. No. A0A3Q7XBP2; 26S proteasome non-ATPase regulatory subunit 12, Acc. No. A0A1S2Y987; eukaryotic initiation factor 4A-10-like, Acc. No. A0A1S2XBQ6; glutamine synthetase, Acc. No. A0A1S3DYR2; two lipoxygenases, Acc. Nos. A0A1S2YV39 and A0A1S2YV51) that, in the irrigated sample, were not identified even in the raw (i.e., without filters) results of the database search. This second group of proteins might provide evidence for remodeling of certain biological pathways because these protein isoforms are exclusively expressed under potential drought stress conditions.

Finally, other unique proteins of the rainfed sample that do not have the isoform counterparts in the irrigated sample appear interesting because they play important roles in stress conditions. For instance, among the proteins showing oxidoreductase

activity, acting on the aldehyde or oxo group of donors, it appears interesting the Delta-1-pyrroline-5-carboxylate (P5C) dehydrogenase 12A1 (Acc. No A0A1S2YRI8), an enzyme involved in the second and last step of proline catabolism (i.e. the oxidation of P5C to glutamate) and therefore important in stress conditions, as an example when plants accumulate proline in response to a wide range of biotic and abiotic stresses.[157] Indeed, during recovery from stress, accumulated proline is rapidly oxidized to glutamate, thus serving as a source of nitrogen and energy. Likewise, three unique proteins of the rainfed sample, involved in the primary metabolism of plants, are interesting because they play a role in plant stress response mechanisms: the adenine phosphoribosyl-transferase (Acc. No. A0A1S3E7D2), the homocysteine S-methyltransferase 3-like (Acc. No. A0A1S2Y082), and the Chaperone protein ClpB3, chloroplastic (Acc. No. A0A1S2XN31). The adenine phosphoribosyl-transferase belongs to the phosphoribosyl-transferase (PRTase) family. PRTases catalyze a special glycosyl modification, the phosphoribosylation, which is a key step in the biosynthetic pathways of purine and pyrimidine nucleotides, tryptophan (Try) and histidine (His), and cofactor NAD(P)+ to control the production of these metabolites.[158] In addition, recent studies highlighted that PRTases co-regulate plant abiotic stress response.[159] Homocysteine S-methyltransferase (HMT) converts homocysteine, a non-proteinogenic α-amino acid, to methionine and therefore plays an important role in supplying methionine for the growth and development of plants. Some studies showed that genes encoding homocysteine S-methyltransferases are overexpressed and promote HMT synthesis in response to drought stress,[160] enhancing drought tolerance in plants with less oxidative damage. Finally, the Chaperone protein ClpB3 is a protein involved in establishing heat stress tolerance and aimed to prevent the aggregation of misfolded proteins, and in a recent paper, the up-regulation of ClpB3 was observed during a prolonged drought[161] stress in maize.

Different Abundant Proteins of Rainfed Conditions

The LFQ analysis revealed that, among the proteins shared between the two sample conditions, in the rainfed sample, fourteen proteins were up-regulated, whereas no proteins were up-regulated in the irrigated sample condition. As reported above, the

group of DAPs up-regulated in the rainfed sample condition includes five storage proteins, six enzymes, and three LEA proteins (Table 1). In particular, the five storage DAPs are four vicilins and a 2S-albumin. As reported in the Introduction, vicilins, belonging to the cupin superfamily, can resist cooking and/or in vitro simulated human digestion.[162] Their resistance to boiling might be due to their propensity to form aggregates in response to heat treatments, and to a very stable compact β-barrel structural motif that could render some of the protease cleavage sites inaccessible.[163] Similarly, chickpea 2S-albumin is reported to resist both boiling and gastrointestinal digestion, but also shows immunogenic properties and protease inhibitor activity.[164,165] In a recent investigation, the susceptibility to gastrointestinal digestion of sprouted chickpeas was evaluated.[129] This investigation, confirming previous literature data, revealed that several peptides derived from 7S vicilin and the 2S albumin were resistant to digestion in both chickpea seed flour and chickpea sprout flour. Moreover, recently, a purified 26 kDa protein was recognized as the major allergen of chickpea crude protein extracts using *in vivo* and *in vitro* models.[166] It is interesting to note that the 2S albumin-like protein (Acc. No. A0A3Q7K771) here detected by the LFQ as a component that appears three times more abundant in the rainfed sample, shares the 86.9% identity with the sequence of the 26 kDa protein recognized by Verma et al., [167] and deposited in the COMprehensive Protein Allergen REsource (COMPARE) database (https://comparedatabase.org/). The second group of DAPs more abundant in the rainfed sample, consists of three LEA proteins. This family of proteins is upregulated during dehydration stress and shows a high abundance during the later stages of seed development, giving the seeds the ability to tolerate drought. However, LEA proteins may also represent another class of chickpea allergens as reported by Wangorsch et al.,[166] and Nitride et al.[152] In this respect, to date, the WHO/IUIS database (http://www.allergen.org/), another catalog of allergens, includes as a unique allergenic protein for the chickpea, the Cic a 1.0101, corresponding to a late embryogenesis protein 4.

Altogether, these results reveal that the sample grown under rainfed conditions, with respect to the irrigated condition, shows a greater abundance of proteins having potential allergen properties, and therefore suggest that a balanced water regime might promote a reduction in allergenic potential in chickpeas. Moreover, the LFQ analysis also evidenced that the rainfed conditions up-regulated a group of enzymes,

including two lipoxygenases, a peroxiredoxin, and a protein disulfide-isomerase involved, with different roles, in plant defense mechanisms. Lipoxygenases (LOX) oxidize polyunsaturated fatty acid (PUFA) substrates by a reaction that represents the first enzymatic step for the synthesis of the oxylipins, a group of oxygenated fatty acids, which actively participate in plant defense mechanisms, being implicated in plant signaling against biotic and abiotic stresses.[168] Peroxiredoxins (PRXS) are non-heme, thiol-dependent peroxidases, playing a wide range of functions in antioxidant defense and redox signaling in plants, that accelerate the elimination of H_2O_2, primarily in plants subjected to various stresses. Finally, the disulfide-isomerases (PDIs), involved in the unfolded and refolded protein response, in plants are found to be up-regulated under both biotic and abiotic stresses.

Conclusion

In conclusion, due to the joining of important analytical strategies, such as a shotgun proteomic approach, the high resolution and high mass accuracy of mass spectral data, and the use of a database restricted to the Cicer arietinum species, it was possible not only to enlarge the actual knowledge of the seed chickpea proteome, but also to explore the plant's response to the two different water regimes applied in this study. The results also highlighted that the storage proteins are markedly up-regulated in the rainfed sample, displaying the highest fold-change values. Considering that these proteins represent the most abundant components of the chickpea protein fraction and are already recognized as potential allergens for chickpea-sensitive individuals, a deep characterization of chickpea storage proteins is crucial for elucidating the biological processes that govern their biosynthesis. For these reasons, the next chapter will focus on the characterization of these proteins using bottom-up, gel-based approaches, to determine whether, as already reported in other legumes, chickpea legumins and vicilins also undergo cleavage processes that generate lower-molecular-weight polypeptides. Furthermore, given the importance of these proteins, they will be discussed from a proteogenomic perspective in the fourth chapter of this thesis.. Particularly, this approach is aimed to compare the proteomic data here obtained with the genomic ones, recently published for the chickpea, to validate proteomics data (e.g. validation of protein entries annotation).

Chapter 2

Characterization of Legumins and Vicilins from Chickpea Seed by SDS-PAGE and Mass Spectrometry

The results of this study are published in the following scientific manuscript:

Di Francesco A., De Santis M.A., Lanzoni A., Pittalà M.G.G., Saletti R., Flagella Z., Cunsolo V. Mass Spectrometry Characterization of the SDS-PAGE Protein Profile of Legumins and Vicilins from Chickpea Seed. Foods. 2024 Mar 14;13(6):887. doi: 10.3390/foods13060887. PMID: 38540876; PMCID: PMC10969193

INTRODUCTION

Chickpea (*Cicer arietinum L.*) seed proteins show a lot of functional properties, making this legume an interesting component for the development of protein-enriched foods. Therefore, the detailed characterization of its protein fraction, and particularly of its globulins (i.e., legumins and vicilins), which constitute its most abundant group, is needed.. In this chapter, a detailed characterization of chickpea seed protein fraction was carried out using SDS-PAGE, in-gel protein digestion, high-resolution mass spectrometry, and database searching. Using this strategy, twenty distinct protein bands were excised from the gel for a comprehensive analysis. As expected by previous studies[169,170], the majority of the bands and the identified peptides were related to vicilin and legumin storage proteins. However, proteins involved in metabolic or other functional roles were also detected. Legumins were revealed at 45–65 kDa, as whole subunits with the α- and β-chains linked together by a disulfide bond. But also at lower mass ranges, these proteins were detected (α- and β-chains migrating alone). Interestingly, vicilins, typically reported to be devoid of cysteine residues, were identified across a broader molecular range (65–23 kDa) and in multiple gel bands, indicating a higher degree of heterogeneity than previously recognized. Mass spectrometric data allowed us to determine that chickpea vicilins, although always described as proteins lacking cysteine residues, contain this amino acid residue. Furthermore, MS data suggests that these storage proteins, like legumins, may undergo extensive proteolytic cleavage producing different polypeptides at lower molecular masses. This process fully accounts for the observed complexity within the chickpea seed proteome.

MATERIALS AND METHODS

Chemicals

The chemicals here employed were of the highest purity commercially available and were therefore used without further purification. Tris-HCl, Ethylenediaminetetraacetic acid (EDTA), dithiothreitol (DTT), ammonium bicarbonate (AMBIC), iodoacetamide (IAA), and formic acid (FA) were provided by Aldrich (St. Louis, MI, USA). Modified porcine trypsin was obtained from Promega (Madison, WI, USA). Water and acetonitrile (ACN) (OPTIMA LC-MS grade) for LC-MS analyses were purchased from Fisher Scientific (Milan, Italy).

Extraction of the Protein Fraction from Chickpea Seeds

The chickpea samples of the Pascià genotype were provided by the Department of Agriculture, Food, Natural Resources, and Engineering (DAFNE) of the University of Foggia. The extraction of chickpea grain storage proteins was carried out according to a protocol adapted from the literature,[171-172] and from De Santis et al.[173] Briefly, the protein fraction was obtained by suspending 100 mg of chickpea flour with 1 mL of an extraction buffer (50 mM Tris‑HCl, pH 7.8, 5 mM EDTA, 0.1% 1,4-dithiothreitol) for 1 h at room temperature with constant stirring, and then centrifuging at 10,000× g for 30 min. The extracted total of the soluble proteins (metabolic and storage proteins, with the exclusion of prolamins) was dried in a speed-vacuum (SpeedVac SPD1030 System, Thermo Fisher Scientific, Waltham, MA, USA).

SDS-PAGE Analysis

Proteins were separated via SDS-PAGE according to the protocol reported in De Santis et al.[174] Separations with 12% polyacrylamide gels (SE 600, Hoefer Inc., Holliston, MA, USA) were followed by gel staining with Coomassie BB G250 and were then digitally acquired (Epson Perfection V750 pro, Seiko Epson Corporation, Suwa, Japan). Four replications of the SDS-PAGE were carried out to assess the

reproducibility of the protein profiles. Then, each of the selected bands was manually excised in duplicate from two lanes of the SDS-PAGE. Finally, the duplicate gel pieces were pooled and transferred to a sterilized microcentrifuge tube (volume: 1.5 mL).

In-Gel Protein Digestion

The gel pieces were washed three times with sequential steps of water and acetonitrile. Then, they were subjected to the procedure of reduction and alkylation, followed by trypsin digestion, according to Shevchenko et al.[175] This approach shows few modifications, as previously reported.[176-177] Briefly, to break the disulfide linkages in the proteins, the gel pieces were treated with 50 µL of 10 mM DTT in 50 mM AMBIC (pH 8.3) and incubated for 30 min at 56 ° C. After the removal of the supernatant, the gel pieces were alkylated with 100 µL of 55 mM IAA in 50 mM AMBIC (pH 8.3) and incubated for 30 min at room temperature. After the removal of the supernatant, the gel pieces were washed with a 100 µL solution with a ratio of 50:50 of water and ACN for 10 min, then the pieces were dehydrated with 100 µL of ACN. Finally, the in-gel enzyme digestion was carried out by adding 30 µL of 10 ng/µL of trypsin in ammonium bicarbonate (pH 8.3; 50 mM) for 30 min at 2-4 ° C. After the removal of the supernatant, the gel pieces were incubated with 50 µL of 50 mM AMBIC overnight at 37 ° C. After in-gel digestion, the supernatant was transferred into a clean 1.5 mL tube. The gel pieces were treated with 50 µL of a 5% FA solution and subsequently with 50 µL of ACN to maximize peptide extraction. This procedure was repeated three times. In the end, the total extracts were pooled, joined with the first supernatant, lyophilized, and re-dissolved in 20 µL of 5% FA for mass spectrometry analyses.

Mass Spectrometry Analysis

Mass spectrometry data were obtained using one microliter of the peptide mixture of each sample and acquired using the procedure described in Cucina et al.[178] The analyses were carried out on a Thermo Fisher Scientific Orbitrap Fusion Tribrid (Q-OT-qIT) mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled online with the Thermo Scientific Dionex UltiMate 3000 RSLCnano system

(Sunnyvale, CA, USA) liquid chromatography. The calibration of the MS system was performed using the PierceR LTQ Velos ESI Positive Ion Calibration Solution (Thermo Fisher Scientific), and MS data were acquired using Xcalibur v. 3.0.63 software (Thermo Fisher Scientific). Chromatographic separations were obtained using an Acclaim RNano Trap C18 Column (trapping column; 100 μm i.d. × 2 cm, 5 μm particle size, 100 A) coupled online with a PepMapR RSLC C18 EASY-Spray column (75 μm i.d. × 50 cm, 2 μm particle size, 100 A). MS survey scans of the eluting peptide cation precursors were carried out at a high resolution (120 K resolution @ 200 m/z), whereas the tandem MS was performed via HCD fragmentation and acquired at a low resolution using the linear ion trap as the analyzer. To avoid carryover during nLC-MS/MS analyses, two blank runs were performed between two consecutive samples, using the same sample gradient program.

Database Search Analysis

The MS data obtained were processed by PEAKS Xpro (release 20 October 2020; https://www.bioinfor.com//) *de novo* sequencing software (Bioinformatics Solutions Inc., Waterloo, ON, Canada) and searched against a dedicated protein database, including all the reviewed and unreviewed entries of *Cicer arietinum L.* (chickpeas) downloaded from the UniProt database (Swiss-Prot and TrEMBL sections, released February 2023, 31,239 entries). To identify the protein contaminants, MS data were also searched against the common Repository of Adventitious Proteins (c-RAP) contaminant database (www.thegpm.org). The following parameters were employed: (i) the trypsin was set as proteolytic enzyme; (ii) the number of allowed missed cleavage sites was set to 3; (iii) as variable amino acid modifications, the oxidation of methionine, the acetylation (N-terminal protein), and the transformation of N-terminal glutamine and N-terminal glutamic acid residue in the pyroglutamic acid form were considered; (iv) as a fixed modification, the carbamidomethylation of cysteines was set. The precursor mass tolerance threshold was 10 ppm, and the max fragment mass error was set to 0.6 Da. Peptide Spectral Matches (PSMs) were validated using a Target Decoy PSM Validator node based on q-values at a False Discovery Rate (FDR) ≤ 0.1%. PEAKS score thresholds for PSMs were set to

achieve FDR values below 0.1% for PSMs, Peptide sequences, and Proteins identified from each database search. A protein was considered identified if a minimum of two peptides (including at least one unique peptide) were matched, and the sequence coverage was ≥ 5%. For each SDS slice, a relative internal quantitative analysis was carried out to obtain the relative abundance of each protein identified. This result was obtained using the value of the "Sample Area" of each protein reported in the PEAKS panel of the identified proteins. Protein sample areas were calculated by the PEAKS Xpro (release 20 October 2020; https://www.bioinfor.com//) software using the total of all peptide features from the unique supporting peptides.

RESULTS AND DISCUSSION

MS Characterization of SDS-PAGE Bands

Regarding the SDS-PAGE analysis, twenty gel bands (Figure 21) were excised and subjected to in-gel digestion, as described in the Materials and Methods section. Mass spectrometry (MS) analysis of each band revealed the co-migration of multiple protein components, which were classified into three categories based on their relative abundance derived from MS data: most abundant proteins (typically one to three per band), minor components, and trace proteins. The most abundant proteins almost always showed a relative abundance higher than 30% and up to 95%, although in a few bands, the relative abundances were in the range of 16–28%. On the other hand, minor components usually displayed relative abundances in the range of 4–15%, whereas very minor components (or trace components) showed relative abundances always below 3%. Overall, in nine gel bands, the predominant proteins were storage proteins, specifically, legumins (11S), vicilins (7S), and albumins (2S). Five bands were dominated by metabolic proteins, while the remaining six bands featured a mixture of both metabolic and storage proteins as the most abundant components. In total, approximately 360 distinct proteins were identified through database searches, as detailed in the Materials and Methods section.

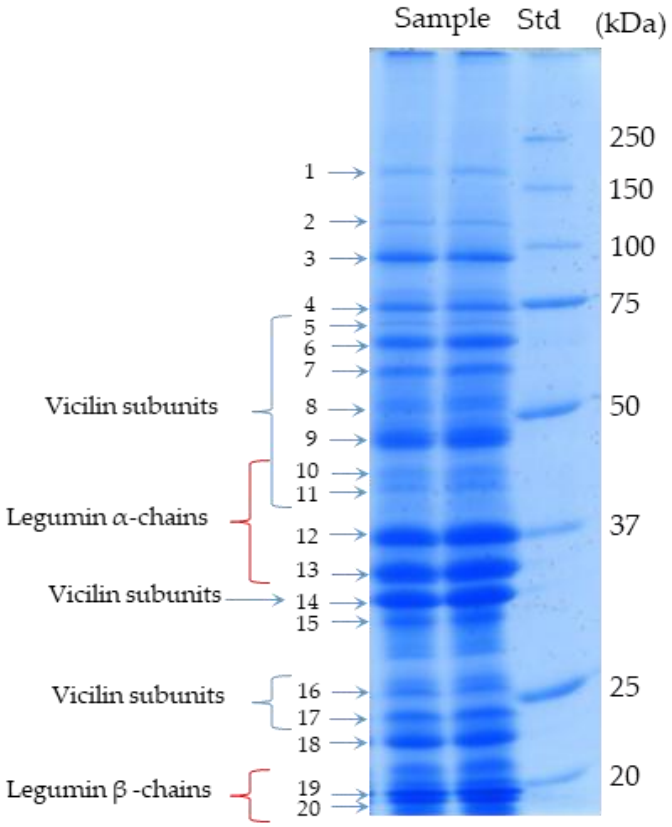


Figure 21. *Two SDS-PAGE lanes of the chickpea seed protein Pascià. Std: Standard protein markers. The twenty bands excised and analyzed via MS are shown by arrows. 11S Legumins and 7S Vicilins containing bands as main components are labelled.*

The characterization of band N.1 (apparent Mr of 150 kDa) revealed the co-migration of 17 different proteins (Table S7). However, the most abundant component (at a relative abundance of 78%) was an LIM domain-containing protein A-like isoform (theoretical Mr 102 kDa, UniProt Accession Number A0A1S2Z626). This protein is involved in biological processes related to the stimuli of abscisic acid (ABA), a plant hormone important in the response to environmental stresses, including drought, soil salinity, cold tolerance, heat stress, etc. The N.2 band (apparent Mr 100 kDa) was related to a ribonuclease (Acc. No. A0A1S2YD23, Mr of 108 kDa; at a relative abundance of 23%), although twenty other minor and trace protein components were also recognized. Two different isoforms of lipoxygenase (Acc. Nos. A0A1S2XBN2, at a relative abundance of 48%, and A0A3Q7YAD4, at a relative abundance of 28%) were found as main components in the band N.3 (apparent Mr 90 kDa), although thirty-six other very low-abundance proteins were identified. Band N.4, focused at an apparent molecular mass of 70 kDa, was related to thirty-seven minor or trace

protein components and a seed biotin-containing protein SBP65 (Acc. No. A0A1S2XET4, Mr of 71 kDa; at a relative abundance of 77%), which represented the most abundant component. This protein belongs to the late embryogenesis abundant (LEA) proteins, a whole family of proteins upregulated during dehydration stress and highly abundant during the later stages of seed development, which gives the seeds the ability to tolerate drought. MS data obtained by the protein in-gel digestion of the bands N.5 (apparent Mr of 65 kDa) and N.6 (apparent Mr of 60 kDa) revealed the co-migration of many proteins: twenty-six in band N.5, and forty-three in band N.6, respectively. But only two were the most abundant components: the alpha-dioxygenase 1 (Acc. No. A0A1S2XH76, Mr of 74 kDa, band N.5, at 53%) and a vicilin-like protein (Acc. No. A0A1S2Y087, Mr of 69 kDa, identified in both bands N.5, at 24%, and N.6, at 95%). Another vicilin-like protein (Acc. No. A0A1S2YZ56, Mr of 82 kDa; at a relative abundance of 34%) was identified as one of the two main components in band N.7 (apparent Mr of 55 kDa), together with a protein disulfide-isomerase (Acc. No. A0A1S2YBR2, Mr of 60 kDa; at a relative abundance of 30%), and 56 other minor or trace protein components. It is important to point out that, despite many studies describing chickpea vicilins as proteins without cysteine residues, three carbamidomethylated cysteine residues were also identified in this vicilin-like protein, and this fact represents the first direct characterization of cysteine residues in chickpea vicilins. Band N.8 (with an apparent Mr of 47 kDa) shows the presence of the vicilin-like protein (Acc. No. A0A1S2Y087) already detected in band N.6, as a main component, and thirty-seven other minor or trace proteins. Analogously, two main proteins were detected in band N. 9 (with an apparent Mr of 44 kDa), together with thirty-five other low-abundant components. The two most abundant components were a provicilin-like protein (Acc. No. A0A1S2XYZ0, Mr of 65 kDa; at a relative abundance of 56%) and a sucrose-binding protein-like (Acc. No. A0A1S2XVJ8, Mr of 54 kDa; at a relative abundance of 32%). Some peptides of the provicilin-like protein carrying carbamidomethyl-cysteine residues were identified and characterized. Two globulins and the alcohol dehydrogenase 1 (Acc. No. A0A1S2YBZ6, Mr of 41 kDa; at a relative abundance of 18%) were the most abundant components in band N.10 (with an apparent Mr of 41 kDa), together with forty-one minor proteins. Particularly, the two globulins were a legumin J (Acc. No. A0A1S2XVG1, Mr of 60 kDa, at a relative abundance of 33%) and a basic 7S globulin-like protein (Acc. No. A0A1S2XV08, Mr of 48 kDa, at a

relative abundance of 29%). The same basic 7S globulin-like (at a relative abundance of 41%), and a different isoform of legumin J (Acc. No. A0A1S3E1N3, Mr of 63 kDa, at a relative abundance of 21%) were identified as the most abundant proteins in band N.11 (with an apparent Mr of 38 kDa), together with thirty-six other minor components. Bands N.12 and 13 (with apparent Mr of 36 and 34 kDa) were mainly determined by legumin proteins, although a lot of minor components were also identified (see Table S7). In detail, a legumin A-like (Acc. No. A0A1S2XSB9, Mr of 59 kDa, at a relative abundance of 81%) was the main component in band N.12, whereas the legumins with the accession numbers A0A3Q7XNW1 (Mr of 56 kDa, at a relative abundance of 60%) and Q9SMJ4 (Mr of 56 kDa, at a relative abundance of 27%), respectively, were identified as the most abundant components in band N.13. The NADPH-dependent aldehyde reductase 1 (Acc. No. A0A1S2YP40, Mr of 32 kDa, at a relative abundance of 38%) and a vicilin-like protein (Acc. No. A0A1S2XQR4, Mr of 52 kDa, at a relative abundance of 28%) were the most abundant components detected in band N. 14 (with an apparent Mr of 32 kDa). The oil body-associated protein (Acc. No. A0A1S2XCR9, Mr of 29 kDa, at a relative abundance of 29%) and the NADPH-dependent aldehyde reductase 1 (Acc. No. A0A1S2XVM2, Mr of 35 kDa, at a relative abundance of 23%) were the most abundant proteins in band N.15 (with an apparent Mr of 30 kDa). The oil body-associated protein belongs to a highly conserved plant group of proteins (i.e., oleosin) that are important for solubilizing seed fats and playing an important role in regulating the biosynthesis, metabolization, and mobilization of lipids during seed maturation and germination. Two group members of the late embryogenesis abundant protein family (i.e., a dehydrin DHN3: Acc. No. A0A1S2Z0P8, Mr of 20 kDa; a LEA protein D-34-like isoform X1: Acc. No. A0A1S3DX03, 22 kDa) and a vicilin-like protein (Acc. No. A0A1S2YZ56, Mr of 82 kDa) were the most abundant proteins identified in band N.16 (with an apparent Mr of 24 kDa). Another vicilin-like protein isoform (Acc. No. A0A1S2XQ88, Mr of 51 kDa, at a relative abundance of 66%) was identified as the most abundant protein in band N.17 (with an apparent Mr of 23 kDa), whereas an albumin-2 (Acc. No. A0A3Q7K771, Mr of 26 kDa, at a relative abundance of 82%) was the main protein responsible in band N.18 (with an apparent Mr of 22 kDa). Finally, two legumins (i.e., legumin-A, Acc. No. A0A1S2XSB9, Mr of 59 kDa; legumin-J, Acc. No. A0A1S2XVG1, Mr of 60 kDa)

and a P24 oleosin (Acc. No. A0A1S2XJM3, Mr of 21 kDa) were identified as the most abundant components in bands (with an apparent Mr of ~20 kDa).

Characterization of legumins and vicilins and their post-translational products

The analysis of the SDS-PAGE bands of grain chickpea protein fraction carried out under reduction conditions shared a similar profile with some results previously reported in the literature, even though some differences have been found. The molecular weight distribution of proteins found in 20 bands ranges from 20 to 130 kDa. Although most of the bands were related to vicilin and legumin storage proteins, many other metabolic functional proteins were also detected. The visual inspection of the SDS-PAGE profile reveals that the most intense bands appear to be mainly focused in three main mass regions, namely the mass ranges of 45–65 kDa, 30–37 kDa, and 20–25 kDa. The mass spectrometry characterization and database search allowed us to verify that these three mass regions were mainly related to six vicilins (Acc. Nos. A0A1S2Y087, A0A1S2YZ56, A0A1S2XV08, A0A1S2XQR4, A0A1S2XQ88, and A0A1S2XYZ0), and five legumin components (Acc. Nos. A0A1S2XSB9, A0A3Q7XNW1, A0A1S2XVG1, A0A1S3E1N3, and Q9SMJ4). However, all these storage proteins were also identified as trace components (i.e., with relative abundances below 3%) spread along almost all of the SDS-PAGE gel. In addition, two other vicilins, with the Acc. Nos. Q304D4 and A0A1S2YKD9 were also identified, although always as components in traces (at relative abundances usually below 2%). The pro-vicilin Q304D4 was spread into almost all the SDS-PAGE slices investigated, and the vicilin A0A1S2YKD9 was instead identified in four bands at lower masses (see Supplementary Figures S23 and S24).

Characterization of Legumins

Focusing on legumins, five different isoforms of this kind of protein were detected in the region between 41kDa (band N.10) and 20 kDa (band N.20), which are the common regions in which it is possible to find the heavy acidic α-chain (molecular masses usually between 36 and 40 kDa) and the light basic β-chain (molecular masses usually about 20kDa), respectively. However, two legumin-like proteins

were also revealed, as minor components, in bands focused at higher mass regions, probably due to migrating as whole subunits (i.e., with the α- and β-chains linked together by a disulfide bond). In this respect, the subunit legumin A-like A0A1S2XSB9 was detected as the most abundant component in bands N.12 (with an apparent mass of 36 kDa) and N.19 (20 kDa), but also as a minor constituent of bands N.2 (100 kDa) and N.7 (55 kDa) (see Supplementary Table S6). The sequence entry of this legumin A-like protein, reported with the unreviewed status by the UniProtKB database, consists of 518 amino acids (Figure 20). The first twenty-one amino acids belong to the signal peptide. Consequently, the entire mature form of this protein is constituted of 497 amino acids (with a Mr of about 57.1 kDa) and, like the other legumins, consists of disulfide-linked acidic (α-chain) and basic (β-chain).

***Figure 22**. Amino acid sequence of the legumin A-like protein reported as an unreviewed (TrEMBL) entry (Acc. No. tr|A0A1S2XSB9) in the UniProtKB database. Peptide signal (AA 1-21), lacking in the mature form of the protein, is reported in bold. The two α- and β-chains are shown by blue and green rectangles, respectively. The cleavage site Asp-Gly (333-334 bond) is indicated by a red arrow. The cysteine residues that are involved in the interchain bond between α- and β-chains are reported in red (107-340), while the cysteine residues (31-64) involved in the intra-chain disulfide bond of the α-chain are reported in bold. The free-thiol cysteine residue (419) of the β-chain is reported in italics and bold. Information about the cleavage site and the disulfide bonds were derived by similarity with the homologue proteins of other legumes (e.g. glycinin from soybean, UniProt Acc. No. P04776).*

By analogy with the homologous legumins of the pea (*Pisum sativum L.*), it can be supposed that the heavy acidic α-chain has 312 amino acids (region Phe22-Asn333, Mr 36.9 kDa) and shows three cysteine residues. In its native form, two out of the three cysteines of the α-chain, namely, Cys31 and Cys64 (reported in red in Figure 22) form an intra-chain disulfide bond, whereas the third one is linked with a cysteine

of the β-chain by an inter-chain disulfide bond, these two cysteine residues are Cys107 and Cys340 (highlighted in bold red in Figure 22). The putative light basic β-chain, which can be generated by a post-translational proteolytic cleavage between the Asn333 and Gly334 of the whole legumin-A, is therefore constituted by 185 amino acids (region Gly334-Ala518 of the entire precursor A0A1S2XSB9) and has a molecular mass of about 20.2 kDa. Moreover, the light basic β-chain shows two cysteine residues; one of them is allowed to form the inter-chain link between α- and β-chains, whereas the second one remains as a free thiol group (Cys 419 in bold italics reported in figure 20). In band N.12, most of the MS/MS data matching this entry were attributable to the α-chain (Figure 25a and Figure S18), whereas few data were associated with the β-chain; on the contrary, the peptides detected in band N.19 were almost exclusively due to the β-chain, and only two MS/MS were related to the α-chain (Figure 23a and Figure S18). In this respect, it is interesting to note that the N-terminal peptide of the β-chain (i.e., GFEETIcm-CTAR) of the A0A1S2XSB9 was also characterized by MS/MS (see Figure 23). On the other hand, the MS data suggest that band N.7 (55 kDa) contains, as a minor constituent, the entire subunit, whereas in band N.2 (100 kDa), the dimeric form of this subunit is probably present as a minor component (Figure 25a and Figure S18). In band N.7, it was also identified the peptide DNGFEETIcm-CTAR (cm-C: carbamidomethylcysteine), which contains the two last C-terminal amino acids (i.e., the aspartic acid and asparagine) of the α-polypeptide linked to the N-terminal peptide of the β-chain (see Figure 20 and Figure S18). This finding suggests the presence of the entire protein in this band. Similarly, the legumin A0A3Q7XNW1 (with a Mr of 54 kDa, as an entire mature protein) was identified as a main component in band N.13 (apparent mass 35 kDa), and as a minor component in bands N.8 (with a mass of 47 kDa) and 19 (20 kDa). The MS data obtained from band N.8 suggest the probable presence of the entire subunit because a lot of peptides spread along the entire sequence length were characterized (Figure 25b and Figure S19). Moreover, as observed in band N.7, the MS data acquired for the band N.8 allowed us to identify the same peptide DNGFEETIcm-CTAR (which is shared between the two legumins A0A3Q7XNW1 and A0A1S2XSB9) containing the C-terminal asparagine of the α-chain and the N-terminal glycine of the β-chain linked together (see Figure S19). Band N.13 was probably related to the α-chain, as almost all the MS/MS matched peptides came from the acidic polypeptide, and only four on their own were due to the β-chain.

Finally, the opposite sequence coverage pattern observed in band 19 (only three of the MS/MS were related to peptides of the α-chain) suggests the presence of the β-chain. In the same way, the identification of the legumin A0A1S3E1N3 (with a Mr of about 59.6 kDa, as a mature protein) in two different bands, namely bands N.11 (38 kDa) and, as a minor component, band N.19 (20 kDa), can be interpreted. The complementary sequence coverage patterns observed between these two bands (Figure 23c and Figure S20) suggest the migration of the α-chain in band N.11, and the presence of the β-chain in band N.19. The legumin Q9SMJ4 (with a Mr of 53.9 kDa, as a mature protein) was identified in band N.13 (34 kDa). Almost all the MS/MS data were from peptides of the α-chain, whereas only two were related to the basic polypeptide (Figure 25d and Figure S21). Therefore, it can be supposed that the acidic polypeptide migrated into this band. Finally, the legumin J-like A0A1S2XVG1 (with a Mr of 57.7 kDa, in its mature form) was identified in band N.10 (41 kDa) as a main component, and in another three bands (i.e., N.17, 19, and 20, in the mass region of 23–20 kDa), as a minor component. The sequence coverage patterns observed in bands N.19 and 20 (Figure 25e and Figure S22) suggest the presence of the β-polypeptide. On the contrary, the MS/MS data obtained from the other two bands (i.e., N.10, with an apparent mass of 41 kDa, and 17, with 23 kDa) belong to peptides spread along the entire sequence length (Figure 25e and Figure S22). Moreover, it is interesting to note that in both of these two bands, the peptide NGLEETIcm-CSAR, which contains the C-terminal asparagine of the α-chain and the N-terminal glycine of the β-chain linked together, was identified (see Figure S22). Consequently, at this stage, it could only be hypothesized that the migration of this legumin J-like protein in these two bands might be related to a possible degradation that occurred during its extraction and/or post-treatment. However, the existence of lower-mass isoforms, formed through different post-translational proteolytic cleavages, which leave the acidic and basic polypeptides joined together, cannot be excluded. It is important to consider that most of the identifications were carried out with the chickpea reference proteins present in Uniprot, obtained from genomic sequences of a limited number of genotypes (e.g., lines Castellana and ILC3279), whereas here we investigated the genotype Pascià. On the other hand, a polymorphism in chickpea seed traits is known, possibly also in seed storage proteins. In this respect, genotypic variability could help to explain the differences in peptide sequences, also concerning the post-translational cleavage position in

legumins. Finally, post-translational modifications due to environmental conditions could also occur. Particularly, environmental factors, including abiotic stress (such as drought) and nutrient availability, are reported to influence seed protein composition in the chickpea in terms of its accumulation rate and, then, in modifying the proportion of protein fractions, and so affecting the final protein composition. Also, specific protein expressions (e.g., heat shock proteins) may occur, influencing the composition of the proteins.

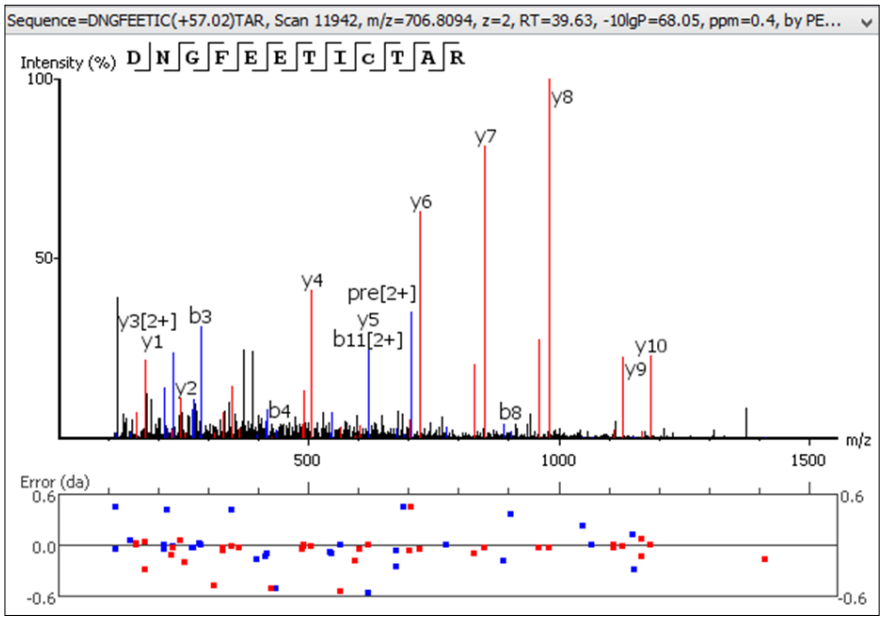


***Figure 23**. Screenshot from the PEAKS X Software of the MS/MS spectrum of the peptide DNGFEETIcTAR. The b and y series identified are highlighted in the Figure.*

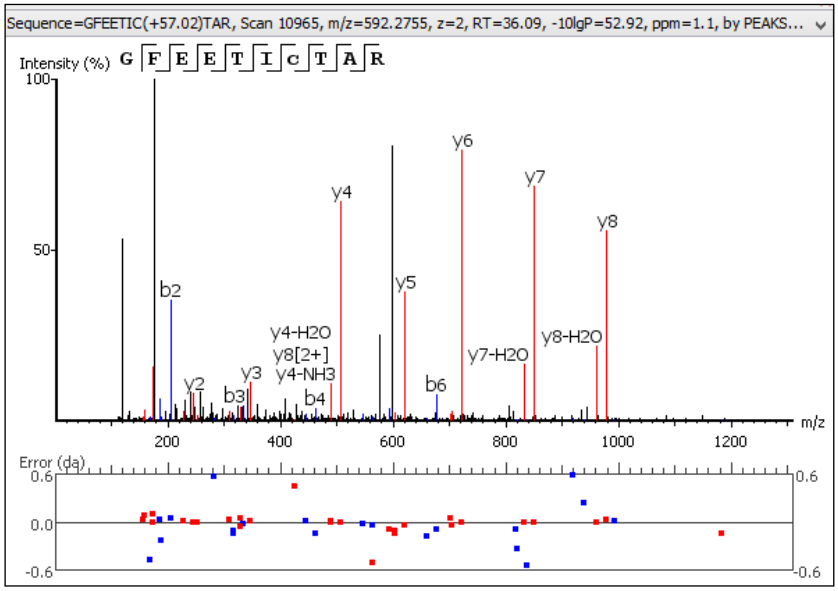


Figure 24. *Screenshot from the PEAKS X Software of the MS/MS spectrum of the peptide NGFEETIcm-CTAR. The b and y series identified are highlighted in the Figure.*

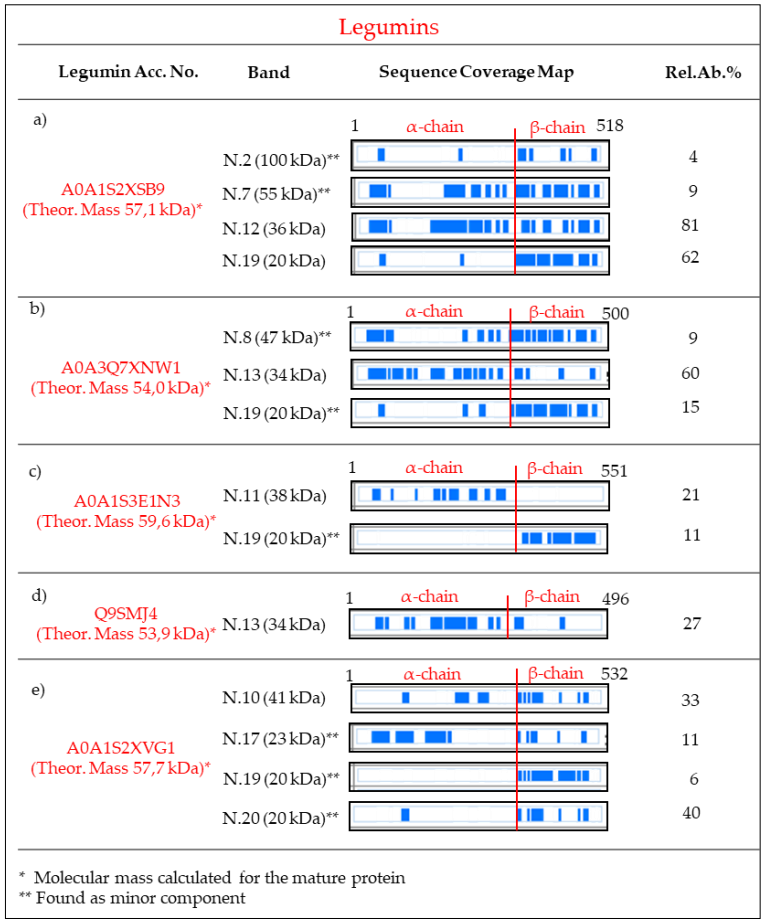


Figure 25. *Protein sequence coverage maps of the legumins identified as minor components in the SDS-PAGE bands (Figure 5). Information about each legumin is reported in subfigures (a–e). Particularly, the coverage is visualized by blue blocks indicating the parts covered by high-confidence peptides (for the detailed sequence coverage pattern, see Supplementary Figures S7–S11). The column Band" refers to the number of the band and the corresponding apparent electrophoretic mass; the column "Rel.Ab.%" shows the percentage of its relative abundance in each band. Finally, the first and last amino acid positions are reported. Amino acid positions are referred to as the pre-propolypeptide sequences (i.e., containing the signal peptides). * Molecular mass calculated for the mature protein. ** Found as a minor component.*

Characterization of Vicilin

About vicilins, firstly, it is interesting to note that the MS data allowed us to ascertain that, although chickpea vicilins were always described as proteins lacking cysteine residues,[179] five out of the six entries of vicilins identified here as the most abundant components carry out this amino acid residue (as reported in the corresponding unreviewed sequence entries from the UniProtKB database). Particularly, seven cysteine residues from four vicilins were characterized (see Supplementary Figures S12–S17). In detail, three out of the eight cysteines for A0A1S2YZ56 (Figure S12), two out of twelve for A0A1S2XV08 (Figure S13), one out of two for A0A1S2XQ88

(Figure S14), and one out of five for A0A1S2XYZ0 (Figure S15) were characterized. On the contrary, the A0A1S2Y087 vicilin shows three cysteine residues, but none were characterized (Figure S16), whereas the A0A1S2XQR4 vicilin does not show cysteines (Figure S17). The observation of the presence of cysteines in the sequence of chickpea vicilins, and particularly of an odd number in the A0A1S2XYZ0 (five cysteine residues) and A0A1S2Y087 (three cysteines) vicilins, suggests the possibility of inter-chain disulfide bonding to form multimeric complexes. In addition, it is interesting to observe that, although these six vicilins show a theoretical molecular mass in the range between 82 and 48 kDa, in the SDS-PAGE they were mainly spread along the mass region between 65 (band N.5) and 23 kDa (band N. 17), and some of them were identified in several bands. The identification of a single vicilin in multiple bands could be related to a possible degradation during protein isolation and post-treatment, but also to indigenous isoforms, generated by post-translational proteolytic events, as already reported in other legumes.[45,50] In this respect, the predicted sequence of vicilin A0A1S2YZ56 has 699 amino acids (no information about the signal peptide is reported in the UniProtKB entry), including ten cysteine residues, and shows a theoretical molecular mass of about 81.9 kDa. However, this protein was detected as a main component in two bands focused at a lower molecular mass than those expected by its putative sequence. In particular, the MS data revealed the presence of peptides related to this protein in the SDS-PAGE region with an apparent mass of 55 kDa (band N.7) and in band N.16 (with an apparent mass of 24 kDa). In particular, in band N.7, peptides flanked between the amino acid positions 289 and 660 were identified, whereas, in band N.16, only peptides between the amino acid positions 102 and 264 were characterized (Figure 26a and Figure S12). Taking into account these results, it should be hypothesized that the vicilin A0A1S2YZ56 undergoes a post-translational proteolytic event, producing two lower-mass vicilin polypeptides. The first one is derived from the C-terminal portion and has a molecular mass of about 50 kDa (band N.7); the second one, instead, is constituted by 250 amino acids of the N-terminal portion of the mature protein and is focused on band N.16. Similarly, the basic 7S vicilin A0A1S2XV08 has 440 amino acids, including the peptide signal constituted by 19 residues. In its mature form, this protein should have a molecular mass of about 46.1 kDa, but it was identified in two bands at lower masses, namely the N.10 (41 kDa) and N.11 (38 kDa). In both these bands, it was identified by a lot of peptides spread

along the full-length sequence, but with the exclusion of the N-terminal fifty amino acids of its mature form (Figure 26b and Figure S13). Consequently, although it cannot be excluded that specifically band N.10 may be related to an entire protein, it can be hypothesized that these two bands were due to an isoform, truncated at the N-terminus.

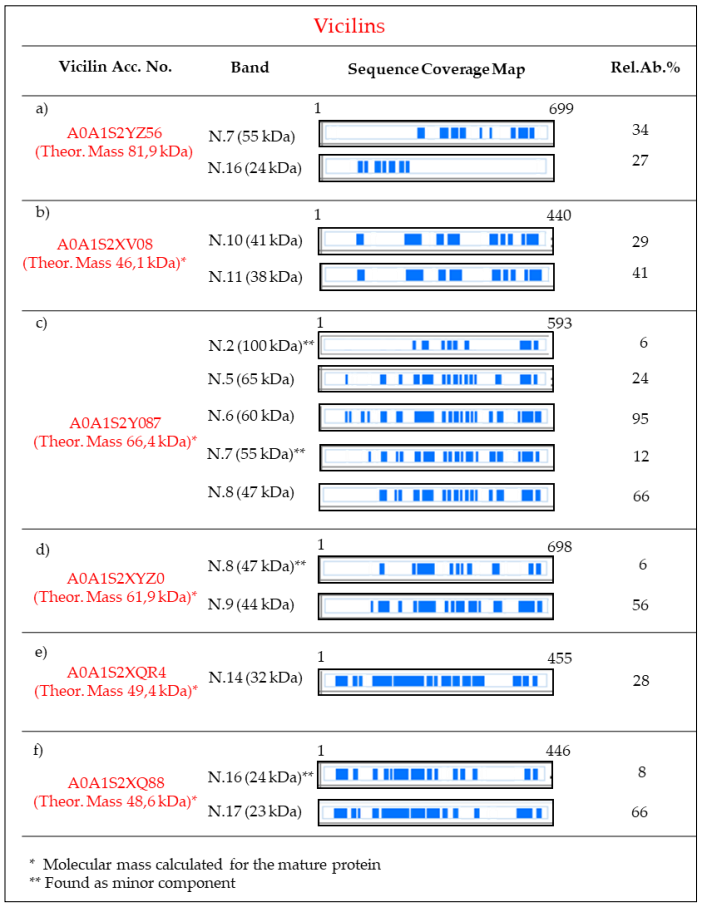


Figure 26. *Protein sequence coverage maps of the vicilins identified as main and, in some cases, minor components in the SDS-PAGE bands (Figure 5). Informations about each legumin is reported in subfigures (a–f). Particularly, the coverage is visualized by blue blocks indicating the parts covered by high-confidence peptides (for the detailed sequence coverage pattern, see Supplementary Figures S1–S6). The column Band"refers to the number of the band and the corresponding apparent electrophoretic mass; the column"Rel.Ab.%" shows the percentage of its relative abundance in each band. Finally, the first and the last amino acid position are reported. Amino acid positions are referred to as the pre-propolypeptide sequences (i.e., containing the signal peptides). * Molecular mass calculated for the mature protein. ** Found as a minor component.*

Also, vicilins A0A1S2Y087 (predicted to have a mature protein Mr of 66.4 kDa) and A0A1S2XYZ0 (predicted to have a mature protein Mr of 61.9 kDa) were identified in multiple bands. Particularly, A0A1S2Y087 was identified as the most abundant protein in bands N.5 (with an apparent mass of 65 kDa), 6 (60 kDa), and 8 (47 kDa), and as a minor constituent in bands N.7 (55 kDa) and 2 (100 kDa) (Figure 26c). The comparison of the sequence coverage patterns observed suggests that in bands N.5 and 6, the full-length sequence was probably focused, whereas bands N.7 and 8 were related to shorter isoforms, lacking about eighty amino acids of the N-terminal part (Figure 26c and Figure S14). Finally, the presence, as a minor component, of this vicilin in band N.2 may be related to its dimeric form. Likewise, in bands N.8 (47 kDa) and 9 (44 kDa), the vicilin A0A1S2XYZ0 (with a predicted Mr of 61.9 kDa) was identified. Taking into account that in both these bands many peptides were characterized as coming from the entire sequence, but that we never detected peptides belonging to the N-terminal part (about eighty amino acids, Figure 26d and Figure S15), it can be supposed that both bands N.8 and 9 are related to an isoform of the vicilin A0A1S2XYZ0, lacking this N-terminal set of amino acids. Finally, both the vicilins A0A1S2XQR4 (with a predicted Mr of 49.4 kDa, as a mature chain) and A0A1S2XQ88 (with a predicted Mr of 48.6 kDa, as a mature protein) were identified only in SDS regions with lower molecular masses than those expected by their putative sequence entries. Particularly, A0A1S2XQR4 was detected in band N.14 (32 kDa), whereas A0A1S2XQ88 was identified in bands N.16 (24 kDa, as a minor component) and 17 (23 kDa, as the most abundant protein). Both these two vicilins were identified, not by peptides coming from delimited regions of the sequence, but by a lot of fragments spread along the entire sequence length (Figure 26 e,f, Figures S16 and S17). Consequently, at this stage, it could only be hypothesized that the identification of these two vicilins in bands at lower molecular masses than the predicted ones might be related to a possible degradation that occurred during their extraction and/or post-treatment. On the other hand, the co-migration of low-mass isoforms, formed through different post-translational proteolytic cleavages, and coming from both the N- and C-terminal portions of the precursor protein, cannot be rejected a priori. Overall, both the presence of cysteines in the sequence of chickpea vicilins and the identification of vicilin fragments at lower molecular masses than those expected by putative sequence entries suggest that, similar to legumins 11S, these storage proteins are first synthesized as pre-

propolypeptides with molecular masses ranging from 50 to 80 kDa. Then, the pre-propolypeptide might undergo proteolytic steps that cut not only the signal peptides but also occur at a midpoint between two cysteine residues, leading to different subunits which remain linked together by disulfide bonds, and, therefore, can be released and detected at lower molecular masses by an SDS-PAGE analysis carried out under reducing conditions. Indeed, as already reported, many vicilins, in some species, undergo extensive proteolytic processing that creates a high degree of heterogeneity in the subunit population.

Conclusion

Overall, the SDS-PAGE analysis coupled with the high-resolution mass spectrometry allowed us to obtain a detailed characterization of legumins and vicilins from chickpea seeds. The most intense bands appeared to be mainly focused in three mass regions, namely, the mass ranges of 45–65 kDa, 30–37 kDa, and 20–25 kDa, and were related to both legumins and vicilin. Both legumin and vicilin proteins were spread along different SDS-PAGE mass regions. The MS data suggest that vicilins, and legumins, as already reported in the pea and other legumes, are firstly synthesized as pre-propolypeptides that may undergo proteolytic steps producing different subunits at lower molecular masses. In order to achieve a precise and comprehensive characterization of the polypeptide subunits derived from legumins and vicilins, confirm the highly specific cleavage sites, and hypothesize the formation of disulfide bonds linking the α- and β- units of legumins, top-down proteomic approaches were carried out in the following and final chapter of this thesis.

Chapter 3

Top-down and Bottom-up MS-based Approaches Uncover the Post-Translational Processes of Legumins and Vicilins in Chickpea (*Cicer arietinum* L.)

INTRODUCTION

Seed storage proteins represent the most important component of legume seeds, as they play crucial roles in both plant development and germination, but also determine the nutritional quality of legumes for human consumption. On the other hand, some studies have highlighted that these proteins are potential allergens. The major storage proteins in chickpea seeds are the 7S and the 11S globulins, known respectively as vicilins and legumins. As observed in other legume species, also chickpea globulins undergo limited proteolysis during seed maturation, a process that contributes to proper folding and assembly of the functional quaternary structures. By similarity with the homologue counterparts, chickpea legumins are processed into two chains, α- and β-, which remain linked by a disulfide bond involving highly conserved cysteine residues. Vicilins are instead cleaved at two internal sites, generating three lower-molecular-weight polypeptides, referred to as α-, β-, and γ-chains. This chapter reports the results of an in-depth structural characterization of the proteolytic processing of chickpea globulins, carried out using an integrated *shotgun* and *top-down* mass spectrometry approach. Overall, the MS data allowed us to confirm that both vicilins and legumins are cleaved at conserved peptide bonds, suggesting the involvement of an asparaginyl endopeptidase (legumain), as previously reported in other plant species. Moreover, the inter-chain disulfide bond linking the α- and β-chain generated by the legumins was hypothesized. For vicilins, two conserved cleavage sites were identified, both containing asparagine or aspartate in the P1 position, again supporting a role for legumain in the maturation process.

MATERIALS AND METHODS

Chemicals

All chemicals were used without further purification because of the highest purity commercially available. Porcine trypsin was purchased from Promega (Madison, WI, USA). Ammonium bicarbonate (AMBIC), cOmplete Protease Inhibitor Cocktail (Hoffmann-La Roche, Basel, CH), 1,4-dithiothreitol (DTT), iodoacetamide (IAA), Tris-HCl, EDTA, formic acid (FA), boric acid, glacial acetic acid, sodium chloride, hexane, and other reagents were analytical grade and purchased from Sigma Aldrich (St. Louis, Missouri, USA). Water and acetonitrile (ACN) were OPTIMA® LC/MS grade. C18 and C4 resins for column packing were purchased from Dr. Maisch GmbH, Beim Brückle, Ammerbuch, Entringen, Germany.

Chickpea sample collection and defatting

The chickpea sample, genotype Pascià, was provided by the Department of Agriculture, Food, Natural Resources and Engineering (DAFNE) of the University of Foggia. Chickpea seeds were ground into powder, and then the flour obtained was defatted using hexane of analytical grade to eliminate the lipid fraction from the chickpea flour by solvent extraction at a ratio of 1:3 weight: volume (flour: solvent) for 1 h under stirring at 200 rpm. Subsequently, the mixture was filtered and dried under an extractor hood at room temperature for 24 hours. The defatting process of the chickpea flour was repeated twice.

Protein Extractions

Extraction of the legumin- and the vicilin-enriched fractions

The legumin and vicilin-enriched fractions were extracted from the defatted chickpea flour as reported by Chang et al.[110] with some modifications (Figure 27). The extraction was carried out in the presence of cOmplete Protease Inhibitor Cocktail (Hoffmann-La Roche, Basel, CH), aimed to protect proteins from aminopeptidases, metalloproteases, and serine, cysteine, and aspartic acid proteases. Particularly, 100 mg of defatted pulse flour was dispersed in 1,5 mL of water, containing 60µL of 25X

stock solution of cOmplete Protease Inhibitor Cocktail. Then, the pH was adjusted to 9.0, using 1.0 M NaOH. The solution was stirred at room temperature for 2 h and then centrifuged at 5600×g for 20 min at 4°C. The residual pellet was discarded, whereas the pH of the supernatant solution was adjusted to 4.5 using 1.0 M HCl to precipitate globulin proteins. This globulin pellet was separated from the supernatant by centrifugation at 5600×g for 5 min at 4°C and re-suspended in 500 µL water by stirring for 2h at room temperature. The pH of the globulin protein solution was neutralized to 7.0 using 0.1 M NaOH, and finally, globulin protein fraction powder was obtained by drying in a speed-vacuum (SpeedVac SPD1030 System, Thermo Fisher Scientific, Waltham, U.S.). The separation of legumin and vicilin fractions from the extracted globulin pellet was obtained as already reported [180,181] with some differences. Particularly, the globulin pellet was dispersed in 800 µL of 0.2 M sodium borate buffer (pH 8.0) containing 0.5 M NaCl, and 60µL of 25X stock solution of cOmplete Protease Inhibitor Cocktail. Afterwards, the solution was stirred for 2h at room temperature. Then, the pH solution was adjusted to 4.5 using 6 M of glacial acetic acid to precipitate the legumin fraction. After centrifugation at 5600×g for 5 min at 4°C, a pellet mainly consisting of legumins was obtained and separated from the supernatant solution, mostly constituted by vicilins. The pH of this supernatant solution was adjusted to 7.0 using 1.0 M NaOH, and desalted by Zeba™ Spin Desalting Columns 7K MWCO (Thermo Fisher Scientific™) to remove borate buffer and salts. The pellet, containing the legumin fraction, was dissolved in a solution at pH 7.0. The protein concentration of legumins and vicilins fractions was determined using the Qubit Protein Assay kit and the Qubit® 1.0 Fluorometer (ThermoFisher Scientific, Milan, Italy). Finally, both protein fractions were dried in a speed vacuum and stored at -80°C until use.

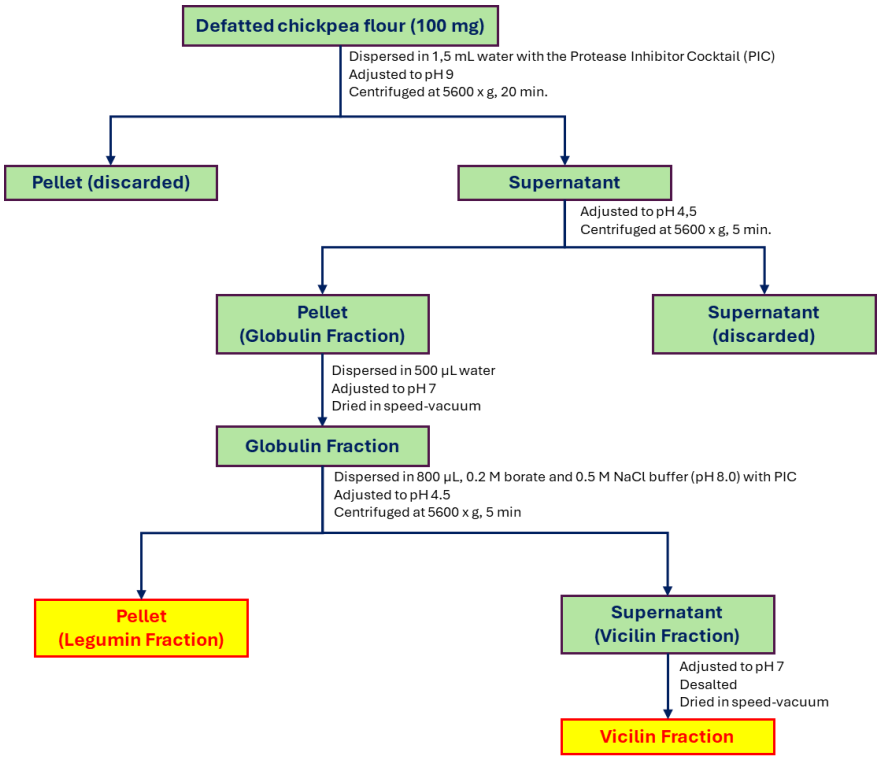


Figure 27. *Flowchart for the preparation of the enriched vicilins and legumins fractions from the defatted chickpea flour*

Protein in-solution hydrolysis

An aliquot corresponding to 50 µg of both enriched-protein fractions of vicilins and legumins was suspended in bicarbonate buffer (50 mM, pH 8), reduced, and alkylated as previously reported by Di Francesco et al.[182] Particularly, about 39 µg of DTT (3 hours at 20 °C) and then 94 µg of IAA (1 h in the dark at 20 °C) were used for the reduction and alkylation, respectively. Then, the solution was split into two aliquots for both shotgun and top-down MS analyses. The aliquot for shotgun analysis was digested by porcine trypsin (Sequencing Grade Modified Trypsin, Porcine, lyophilized, Promega) at an enzyme-substrate ratio of 1:50 (overnight, 37°C). To achieve a final concentration of 25 ng/µL, a 2% aqueous solution of FA was added, resulting in a final volume of 2 mL. The aliquot for top-down analysis was dried in a speed vacuum and finally suspended in 2% FA for top-down MS.

Mass Spectrometry Analyses

Survey scans of analytes (i.e., peptides and proteins) and MS data were acquired using Xcalibur v. 4.6 software (Thermo Fisher Scientific). The calibration of the MS systems was performed by using the Pierce FlexMix Calibration Solution (Thermo Fisher Scientific). The deconvolution of the multi-charged ESI mass spectra of the proteins was obtained using the Xtract® deconvolution algorithm (Thermo Scientific).

Shotgun LC-MS/MS analysis

LC-MS/MS analyses of the peptide mixtures were carried out on a ThermoFisher Scientific Orbitrap Lumos Tribrid® (Q-OT-qIT) mass spectrometer (ThermoFisher Scientific, Bremen, Germany) coupled with a ThermoFisher Scientific EASY-nLC 1000 system (Sunnyvale, CA) liquid chromatography. In particular, 5 µL (corresponding to 125 ng) of the peptide mixture were loaded into a homemade pre-column packed with ReproSil-Pur 120 C18-AQ, (75 µm i. d. x 5 cm, 5 µm particle size, 120 Å)[183] and after separated by a homemade analytical-column packed with Reprosil-Pur 120 C18-AQ (75µm i.d. x 24 cm, 3 µm particle size, 120 Å). Elution was carried out at a flow rate of 300 nL/min by a linear gradient of solvent B (ACN+0.1%FA) in A, 2% to 28% in 55 min, 28% to 40% in 22 min, 40% to 100% in 1 min and finished by holding 100% B for 7 min, 100% to 2% in one minute and re-equilibrating the column at 2% B for 3 min. The eluting peptide cations were converted to gas-phase ions by electrospray ionization using a source voltage of 2.2 kV and introduced into the mass spectrometer through a heated ion transfer tube (275 °C). Survey scans of peptide precursors from 200 to 1700 m/z were performed at 120 K resolution (@ 200 m/z). Tandem MS was performed by isolation at 1.6 Th with the quadrupole, HCD fragmentation with a normalized collision energy of 30, and high-resolution MS analysis of fragments in the Orbitrap (50 K resolution). Only those precursors with charge states 2÷4 and intensity above the threshold of 5000 were sampled for MS2. The dynamic exclusion duration was set to 20 s with a 10 ppm tolerance around the selected precursor and its isotopes. Monoisotopic precursor election was turned on. The instrument was running top speed mode with

3 s cycles, meaning it would continuously perform MS2 events until the list of non-excluded precursors diminished to zero or 3 s, whichever is shorter.

Top-down LC-MS/MS analysis

Top-down analysis of the enriched vicilin protein fraction was carried out on a Thermo Fisher Scientific Orbitrap Fusion Lumos (Q-OT-qIT) mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled online with the Thermo Scientific EASY-nLC 1000 nanosystem (Sunnyvale, CA, USA). In detail, 2 µL (corresponding to 500 ng of protein mixture) of the reduced and alkylated vicilin-enriched fraction was loaded onto a homemade C4 column packed with ReproSil-Pur 100 C4 (100 µm i.d. x 25 cm, 1.8 µm particle size, 100 Å). The proteins were separated by elution at a flow rate of 250 nL/min at room temperature by a linear gradient of solvent B (ACN+0.1%FA) in A, 10% for 1 min, followed by 10 % to 70% in 40 min, 70% to 100% in 5 min. We finished by holding 100% B for 5 min, 100 % to 10% in one minute, and re-equilibrating the column at 2% B for 20 min. The eluting protein cations were converted to gas-phase ions by electrospray ionization using a source fragmentation of 15 eV and a source voltage of 2.2 kV, and finally introduced into the mass spectrometer through a heated ion transfer tube (275 °C). Mass spectra of all the precursor ions were acquired at 120 K resolution at m/z 200, in the range m/z 200-1800. MS/MS analyses were performed by CID fragmentation with a collision energy of 30, and the MS fragment ions were analyzed in the Orbitrap (high-resolution MS/MS analysis) at 60 K resolution in cycle time mode set to 3s.

Top-down analysis of the enriched legumin protein fraction was carried out on Orbitrap Fusion Tribrid (Q-OT-QIT) (ThermoFisher Scientific, Bremen, Germany) coupled online with the UHPLC ThermoFisher Scientific Dionex UltiMate 3000 RSLCnano system (Sunnyvale, CA). In detail, 1 µL of the reduced and alkylated protein fraction, corresponding to 5 ng, was injected into an Easy-Spray PepMap HPLC capillary column C4 (1500 Å, 4 µm, 150 µm x 150 mm). The proteins were separated by elution at a flow rate of 800 nL/min, at 40 °C, by a linear gradient of solvent B (ACN+0.1%FA) in A, from 15% to 65% in 82 minutes, followed by 65% to 95% in 5 minutes, 95% for 10 minutes, then the column was eluted with solvent B from 95% to 15% in 5 min and finally equilibrated at 15% B for 10 min. Eluted

proteins were converted to gas-phase by electrospray ionization, source voltage 1.3 kV, and introduced into the mass spectrometer through a heated (275 °C) ion transfer tube. Mass spectra of all the precursor ions were acquired in a range from 200 to 1800 m/z at a resolution of 120K (@ 200 m/z) using the following parameters: RF lens, 80%; Auto Gain Control: 200000; maximum injection time, 100 ms. MS/MS analyses were performed by HCD fragmentation with a collision energy of 10, 12, and 15, and the MS fragment ions were analyzed in the Orbitrap (high-resolution MS/MS analysis) at 120 K resolution in cycle time mode set to 3s.

Data Analysis

Protein Sequence Handling

Protein sequence handling was performed using the General Protein Mass Analysis for Windows 9.5 (GPMAW) software (Lighthouse data, Odense, Denmark).

Shotgun Data Analysis

The MS data obtained from the shotgun approach were processed by PEAKS Xpro *de novo* sequencing software (Bioinformatics Solutions Inc., Waterloo, ON, Canada; https://www.bioinfor.com//), which provides automated and accurate *de novo* peptide sequencing. The *de novo* amino acid sequences generated by PEAKS were searched against a custom protein database, including only the legumins and vicilins entries (reviewed and unreviewed) of *Cicer arietinum L.* (chickpeas) downloaded from the UniProt database (Swiss-Prot and TrEMBL sections, released May 2025, 31239 entries). The database search was carried out setting trypsin as a proteolytic enzyme, semispecific as digestion mode, and 3 as the maximum number of allowed missed cleavages. In the first step of the database search, carbamidomethylation of cysteines was set as a fixed modification, whereas the oxidation of methionine, the acetylation (N-terminal protein), formylation, and the transformation of N-terminal glutamine and N-terminal glutamic acid residue into the pyroglutamic acid form were considered as variable amino acid modifications. The de-novo amino acid sequences unmatched during the first round of the database search were searched, enabling the PEAKS PTM tool to identify as additional modifications the methionine

dioxidation, cysteine dioxidation, and cysteine trioxidation. Database search was carried out, setting the precursor mass tolerance threshold at 10 ppm and the max fragment mass error at 0.006 Da. Peptide Spectral Matches (PSMs) were validated using a Target Decoy PSM Validator node based on q-values at a False Discovery Rate (FDR) ≤ 1%. PEAKS score thresholds for PSMs were set to achieve FDR values below 1% for PSMs, Peptide sequences, and Proteins identified from each database search. For both the samples investigated by the shotgun approach, the relative abundance of each protein identified was calculated. This result was achieved starting from the raw values of the "Sample Area" of each protein reported in the PEAKS panel. The raw protein values of "Sample Area" are calculated by the software PEAKS using the total of all peptide features from unique supporting peptides. Then, for each LC-MS run, the relative abundance of each protein (reported as a percentage) was achieved to the normalized total areas of all proteins quantified.

Top-down Data Analysis

The Full scan MS and MS/MS data obtained from the top-down analyses were preliminarily processed by the MASH Native software in 'Discovery mode' (Ge Research Group, University of Winsconsin-Madison) and searched against a protein database including all the reviewed and unreviewed entries of *Cicer arietinum L.* (chickpea) of the UniProt database (Swiss-Prot and TrEMBL sections, release February 2023, 31239 entries). This preliminary database search aimed to find what kind of proteins were present in the sample under investigation. The database searches were carried out using the TopFD as a deconvolution algorithm and the TopPIC as a Database Search algorithm. The number of unexpected PTMs was set to 1, the precursor mass tolerance threshold was set to 15 ppm, whereas the max fragment mass error was set to 0.020 Da. Then, to validate the results, a Decoy Database Searching was carried out, with a Spectrum Cutoff value based on E-Value set to 0.01. Finally, a freely available tool coined ClipsMS (Comprehensive localization of Internal Protein Sequences) (https://github.com/loolab2020/ClipsMS-Version-1.0.0),[137] and developed to assign both terminal and internal fragments resulting from a top-down mass spectrometry experiment, was used to in-depth characterize the sequence of each polypeptide detected and identified in the preliminary database search. In detail, the raw MS/MS

spectra obtained by CID fragmentation of polypeptides were first deconvoluted using the Xtract® deconvolution algorithm (Thermo Scientific); then, every deconvoluted mass list was uploaded into the ClipsMS program together with the selected sequence entry to obtain a matched fragments list. The error for fragment matching was set at 10 ppm, and the smallest internal fragment size was set at 5 amino acids. Only "*by"* internal fragments were searched for and assigned, and all terminal fragments were assigned before considering internal fragments. All overlapping internal fragments due to the arrangement and frameshift ambiguity were retained to include all fragmentation propensity possibilities. After matching, all assigned internal fragments were manually validated against the raw MS/MS spectra to ensure: i) these internal fragments were real peaks rather than noise or isotopes and ii) the masses of matched internal fragments did not overlap with terminal fragments.

RESULTS AND DISCUSSION

The legumin-enriched fraction and the vicilin-enriched fraction obtained as described in the "Material and Methods" section of this chapter were investigated by the experimental workflow depicted in Figure 28.

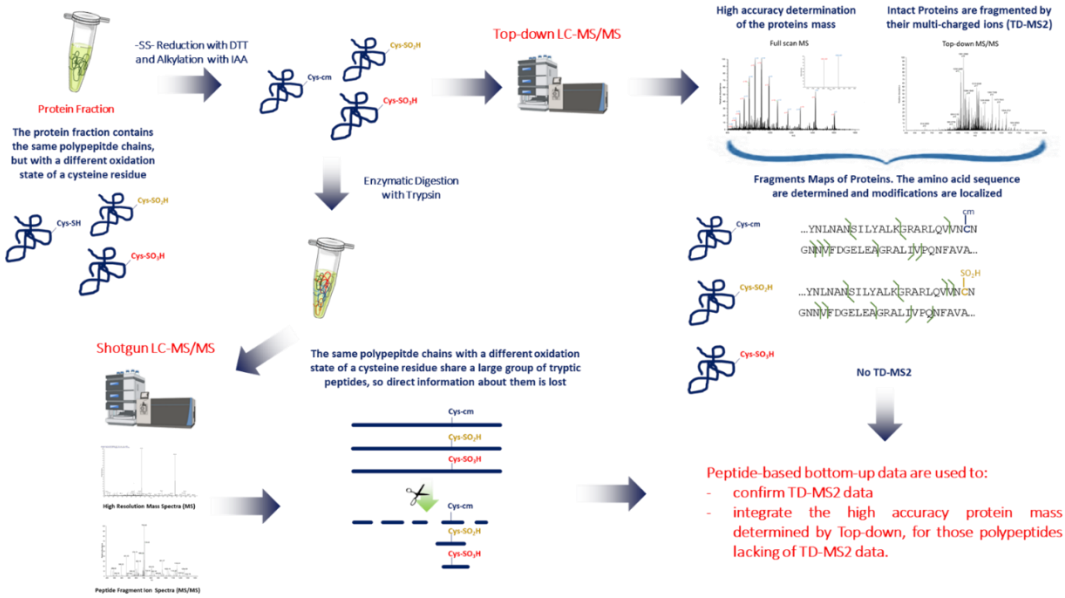


Figure 28. *Workflow applied for the structural characterization of the polypeptides generated by specific post-translational cleavages of legumins and vicilins, occurring during maturation of seed chickpeas. It is reported, as example, how integrating the top-down and the peptide-based bottom-up data can be characterized three polypeptides sharing the same amino acid sequence but showing different oxidatation states of a cysteine residue.*

Characterization of the Legumin-enriched Fraction

Shotgun Analysis

The shotgun proteomic analysis performed on the legumin-enriched fraction reduced and alkylated with IAA, led to the identification of 12 proteins (Table 2).

Table 2. *List of identified proteins in the legumin-enriched fraction, For each polypeptide chain is reported: the UniProt Accession Number, the description as it is reported in UniProt, the relative abundance, the average mass, the peptides and the unique peptides with whom the proteins have been identified.*

UniProt Accession Number	Description	Relative Abundance (%)	Avg. Mass (kDa)	Coverage (%)	Peptides	Unique Peptides
A0A1S2XSB9	Legumin A-like	35,2	59,3	85	181	119
A0A1S2XVG1	Legumin J-like	17,3	60,4	82	132	132
A0A3Q7XNW1	Legumin-like	15,2	56,3	85	180	47
Q9SMJ4	Legumin	8,6	56,2	72	140	24
A0A1S2Y087	Vicilin-like	7,6	69,4	81	125	110
A0A1S2XYZ0	Provicilin-like	4,0	64,6	70	77	62
A0A1S2XQ88	Vicilin-like	3,7	51,1	85	128	34
A0A1S2YZ56	Vicilin-like seed storage protein	2,8	81,9	58	89	89
A0A1S3E1N3	Legumin J	2,6	62,7	69	69	68
A0A1S2XQR4	Vicilin-like	1,9	51,8	83	139	24
Q304D4	Provicilin	0,7	51,4	46	70	8
A0A1S2YKD9	Vicilin-like seed storage protein	0,4	53,1	52	29	29

Among these, five were legumins, while the remaining seven were identified as vicilins. These findings indicate that the stepwise extraction method employed did not achieve complete selectivity for legumins. Nevertheless, internal relative quantification based on mass spectrometry (MS) data (refer to Supplementary Material) revealed that legumins constitute the predominant protein group, accounting for approximately 79% of the total protein content in this fraction, whereas vicilins represent about 21%. Notably, the three most abundant legumins, comprising approximately 68% of the total protein content, correspond to sequences with the following UniProtKB Accession Numbers: A0A1S2XSB9 (35.2% relative abundance), A0A1S2XVG1 (17.3%), and A0A3Q7XNW1 (15.2%). Detailed information on protein identification and peptide matching within the legumin-enriched fraction is provided in Supplementary Table S8.

Top-down MS

To characterize the amino acid sequences of the legumins, the reduced and iodoacetamide (IAA)-alkylated legumin-enriched fraction was further analyzed by liquid chromatography-mass spectrometry (LC-MS) using a top-down approach.

Preliminary processing of the MS data with MASH Native software enabled the identification of numerous polypeptides with experimental molecular masses of approximately 20 kDa. MS data allowed us to relate these components to the C-terminal regions of the three most abundant legumins (i.e., Acc. No. A0A1S2XSB9, A0A1S2XVG1, and A0A3Q7XNW1) previously identified by the shotgun proteomic approach.

Polypeptides related to the legumin A0A1S2XSB9

Panel (A) of Figure 29 shows the deconvoluted spectra of four polypeptides having, respectively, monoisotopic molecular masses of 20021.2166, 19989.2364, 19980.1759, and 19964.2195 Da (the corresponding multi-charged ESI mass spectra are reported in the Supplementary Figure S25). The preliminary database search, carried out by the MASH Native software, allowed us to relate these components to the amino acid region flanked between the glycine at position 334 and the asparagine at position 514 of the legumin A-like entry with the Acc. No. A0A1S2XSB9 (Figure 29D). An in-depth characterization by top-down-MS2 (TD-MS2) was achieved only for the polypeptides with M_{mono} 19989.2364 and 19964.2195 Da (see the Supplementary Material for details). The component with M_{mono} 19989.2364 Da was identified as corresponding to the region Gly^{334}-Asn^{514} with Cys^{340} and Cys^{419} in carbamidomethylated form (Figure 29D; Figures S26 and S27, Table S9). The polypeptide with M_{mono} 19964.2195 Da was instead related to the same amino acid region Gly^{334}-Asn^{514}, but carrying Cys^{340} in carbamidomethylated form and Cys^{419} as sulfinic acid (Figures S28 and S29, Table S10). In this respect, it is interesting to note that, by the shotgun approach, the cysteine at position 340 was always identified in the carbamidomethylated form. On the contrary, the cysteine residue at position 419, in some peptides, was identified in the di-oxidized form (i.e., as sulfinic acid, Cys-SO_2H) (Figure S30). For the other two polypeptides at molecular masses of 20021.2166 and 19980.1759 Da, no TD-MS2 data were achieved. However, based upon the accurate intact mass information and taking into account the MS data obtained by the complementary shotgun approach, we could putatively identify these two polypeptides. Particularly, the component having the experimental monoisotopic mass of 20021.2166 Da shows a mass difference of +31.9802 Da respect to the mass of the polypeptide at 19989.2364 Da, which corresponds to the region Gly^{334}-Asn^{514}

with both cysteines in carbamidomethylated form. The observed mass shift could be due to the mono-oxidation (+15.9949 Da) of two out of three methionine residues included in the Gly^{334}-Asn^{514} region (Figure 29D). Otherwise, it can be due to the di-oxidation of a methionine residue to methionine sulfone (+31.9898 Da). Taking into account the results of the shotgun approach, the hypothesis of the di-oxidation of a methionine residue should be excluded because no peptides with this modification were identified in this amino acid region. On the contrary, by the shotgun approach, the peptides Asn^{390}-Asn^{402} and Leu^{468}-Arg^{491} of the legumin A0A1S2XSB9, carrying the methionine residues at positions 392 and 489, respectively, as methionine sulfoxide (+15.9949 Da), were identified (Figure S31). It should be highlighted that the region Asn^{390}-Asn^{402} is common with the legumins with the Acc. No. A0A3Q7XNW1 and Q9SMJ4.

Finally, the last polypeptide related to the region Gly^{334}-Asn^{514} of the legumin entry with the Acc. No. A0A1S2XSB9, and having an experimental monoisotopic mass of 19980.1759 Da, could be assigned to the region Gly^{334}-Asn^{514} carrying the Cys^{340} as carbamidomethyl-cysteine and the Cys^{419} as sulfonic acid (calculated monoisotopic molecular mass 19980.1936 Da). Indeed, by the shotgun analysis, the Cys^{419} was also identified in tri-oxidized form (i.e., as sulfonic acid, Cys-SO_3H) as shown in Figure S32. The results of these polypeptides related to the legumin A-like entry with the Acc. No. A0A1S2XSB9 are summarized in Table 3.

***Table 3**. Top-down results about the polypeptide chains related to the Gly^{334}-Asn^{514} region of the legumin A-like entry with the Acc. No. A0A1S2XSB9. For each polypeptide chain is reported: the amino acid region, the status of two cysteines at positions 340 and 419, the other detected modifications, the experimentally determined monoisotopic mass, the theoretical monoisotopic mass, and the error.*

	AA Region	Cys^{340}	Cys^{419}	Other Modif.	Exp. M_{mono} (Da)	Theor. M_{mono} (Da)	Error (ppm)
1	Gly^{334}-Asn^{514}	CM [a)]	CM	2 Met-ox [b)]	20021.2166	20021.2017	0.7
2	Gly^{334}-Asn^{514}	CM	CM	-	19989.2364	19989.2119	1.2
3	Gly^{334}-Asn^{514}	CM	di-ox [c)]	-	19964.2195	19964.1987	1.4
4	Gly^{334}-Asn^{514}	CM	tri-ox [d)]	-	19980.1759	19980.1936	0.9

a) CM = carbamidomethylation
b) Met-ox = methionine oxidized as methionine sulfoxide
c) Di-ox = Cysteine di-oxidized as sulfinic acid
d) Tri-ox = Cysteine tri-oxidized as sulfonic acid

Polypeptides related to the legumin A0A1S2XVG1

Panel (B) of Figure 29 displays the deconvoluted spectra of four co-eluting polypeptides. The two most abundant species exhibit experimentally determined monoisotopic masses of 20060.4701 and 20131.4791 Da, whereas the two minor components show monoisotopic masses of 20076.4647 and 20147.4786 Da. The corresponding multiply charged ESI mass spectra are provided in Supplementary Figure S33.

Preliminary analyses performed using the MASH Native software enabled us to associate all these polypeptides with the amino acid region spanning from Gly353 to the C-terminal Ala532 of the legumin J-like protein (Accession No. A0A1S2XVG1; see Figure 29, panel E). The components with monoisotopic masses of 20131.4791 and 20147.4786 Da correspond to the Gly353–Ala532 region, while the other two polypeptides (20060.4701 and 20076.4647 Da) correspond to the Gly353–His531 region, both lacking the C-terminal alanine.

This segment contains two cysteine residues located at positions 359 and 438. Assuming both cysteines are carbamidomethylated, the theoretical monoisotopic masses of the Gly353–Ala532 and Gly353–His531 fragments are 20158.4987 Da and 20087.4616 Da, respectively. None of the observed molecular masses precisely matches these theoretical values, suggesting that the discrepancies may arise from post-translational modifications or amino acid substitutions relative to the sequence reported in the database.

Interestingly, shotgun proteomic analyses consistently identified Cys359 in the carbamidomethylated form, while Cys438 was detected in both di-oxidized and tri-oxidized forms (Figures S34a and S34b). The polypeptide with an experimental monoisotopic mass of 20060.4701 Da was identified through MS/MS analysis of the [M+24H]24+ ion at m/z 837.3197 (Supplementary Figures S35 and S36; Table S11). It corresponds to the Gly353–His531 fragment containing Cys359 as carbamidomethyl-cysteine and Cys438 as sulfinic acid. The minor component with a mass of 20076.4647 Da likely represents the same fragment carrying Cys438 as sulfonic acid, although no TD-MS² data were obtained for this peptide.

Similarly, the two remaining polypeptides with experimental masses of 20131.4791 and 20147.4786 Da appear to derive from the Gly353–Ala532 region, each containing different oxidation states of Cys438. The component with a mass of

20131.4791 Da likely contains Cys359 as carbamidomethyl-cysteine and Cys438 as sulfinic acid, while the minor species at 20147.4786 Da is presumed to carry Cys438 as sulfonic acid. The presence of Cys438 as sulfinic acid in the polypeptide with a mass of 20131.4791 Da was confirmed by MS/MS analysis of its [M+24H]24+ ion at m/z 840.3639 (Supplementary Figures S37 and S38; Table S12). Conversely, the assignment of the minor component at 20147.4786 Da remains tentative, as no TD-MS2 data were available.

A summary of the results for the polypeptides associated with the legumin J-like entry (Acc. No. A0A1S2XVG1) is presented in Table 4.

***Table 4**. Top-down results about the polypeptide chains related to the Gly^{353}-Ala^{532} and Gly^{353}-His^{531} regions of the legumin A-like entry with the Acc. No. A0A1S2XVG1. For each polypeptide chain is reported: the amino acid region, the status of two cysteines at positions 359 and 438 of the β-chain, the experimentally determined monoisotopic mass, the theoretical monoisotopic mass, and ΔM (Experimental Molecular Mass – Theoretical Molecular Mass).*

	AA Region	Cys^{359}	Cys^{438}	Exp. M_{mono} (Da)	Theor. M_{mono} (Da)	ΔM (ppm)
1	Gly^{353}- His^{531}	CM [a]	di-ox [b]	20060.4701	20060.4619	0.4
2	Gly^{353}- His^{531}	CM	tri-ox [c]	20076.4647	20076.4568	0.4
3	Gly^{353}- Ala^{532}	CM	di-ox	20131.4791	20131.4990	1.0
4	Gly^{353}- Ala^{532}	CM	tri-ox	20147.4786	20147.4939	1.5

a) CM = carbamidomethylation
b) di-ox = Cysteine di-oxidized as sulfinic acid
c) tri-ox = Cysteine tri-oxidized as sulfonic acid

Polypeptides related to the legumin A0A3Q7XNW1

Panel (C) of Figure 29 displays the deconvoluted spectra of four polypeptides with monoisotopic masses of 20044.2672, 20019.2395, 20035.2252, and 20076.2292 Da, respectively. Preliminary analyses suggest that these components correspond to the legumin-like protein with Accession No. A0A3Q7XNW1 (see Figure 29, panel F). The corresponding multiply charged ESI mass spectra are provided in Supplementary Figure S39. The polypeptides with monoisotopic masses of 20044.2672 and 20019.2395 Da were unequivocally identified through TD-MS^2 analysis of their multiply charged ions (see Supplementary Figures S40–S43 and Tables S13 and S14). The component with a mass of 20044.2672 Da corresponds to the region Gly315–Asn496, containing both cysteine residues at positions 321 and 400 in the carbamidomethylated form. In contrast, the polypeptide with a mass of 20019.2395 Da represents the same region (Gly315–Asn496) but carries Cys321 as

carbamidomethyl-cysteine and Cys400 as sulfinic acid. As previously observed for the legumin A0A1S2XSB9-related peptides, shotgun proteomics consistently identified Cys321 of the legumin A0A3Q7XNW1 protein in the carbamidomethylated form, whereas Cys400 was detected in carbamidomethylated, sulfinic, and sulfonic states (Figure S44). The identification of a peptide containing Cys400 as sulfinic acid supports the assignment of the 20019.2395 Da component, while the presence of the sulfonic form may account for the species at 20035.2252 Da. Although TD-MS² data were not obtained for this latter component, it most likely corresponds to the Gly315–Asn496 fragment carrying Cys321 as carbamidomethyl-cysteine and Cys400 as sulfonic acid.

Similarly, no TD-MS² data were acquired for the polypeptide with a monoisotopic mass of 20076.2292 Da. However, its mass difference of +56.9897 Da relative to the 20019.2395 Da component suggests the occurrence of protein over-alkylation during the iodoacetamide (IAA) alkylation step (ΔM = +57.0215 Da).

A summary of the results obtained for the polypeptides related to the legumin-like entry A0A3Q7XNW1 is presented in Table 5.

Table 5. *Top-down results about the polypeptide chains related to the Gly^{315}-Asn^{496} region of the legumin A-like entry with the Acc. No. A0A3Q7XNW1. For each polypeptide chain is reported: the amino acid region, the status of two cysteines at positions 321 and 400, the experimentally determined monoisotopic mass, the theoretical monoisotopic mass, and ΔM (Experimental Molecular Mass – Theoretical Molecular Mass).*

	AA Region	Cys^{321}	Cys^{400}	Exp. M_{mono} (Da)	Theor. M_{mono} (Da)	ΔM (ppm)
1	Gly^{315}- Asn^{496}	CM [a)]	di-ox [b)]	20019.2395	20019.2377	0.09
2	Gly^{315}- Asn^{496}	CM	tri-ox [c)]	20035.2252	20035.2326	-0.4
3	Gly^{315}- Asn^{496}	CM	CM	20044.2672	20044.2694	-0.1
4	Gly^{315}- Asn^{496}	CM	di-ox	20076.2292 [d)]	20076.2592	-1.5

a) CM = carbamidomethylation
b) di-ox = Cysteine di-oxidized as sulfinic acid
c) tri-ox = Cysteine tri-oxidized as sulfonic acid
d) Carrying an additional iodacetamide molecule (over-alkylation)

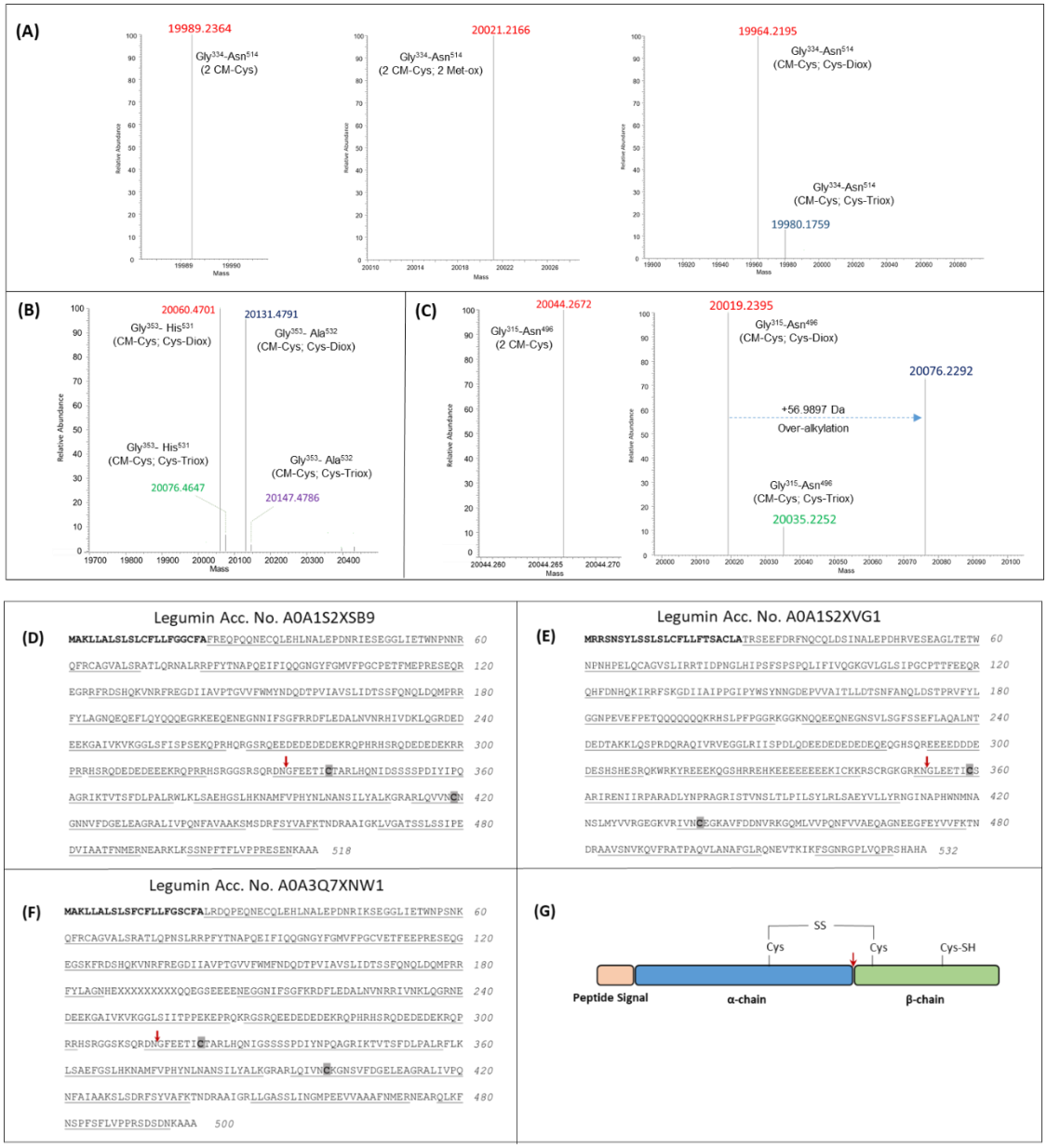


Figure 29. *Panels (A), (B), and (C) report the monoisotopic deconvoluted mass spectra (mass zero-charge) of the post-translational polypeptide products related to the legumins with Acc. No. A0A1S2XSB9, A0A1S2XVG1, and A0A3Q7XNW1, respectively. Panels (D), (E), and (F) show the amino acid sequence of the legumins A0A1S2XSB9, A0A1S2XVG1, and A0A3Q7XNW1, respectively. Peptide signal, lacking in the mature form of the protein, is reported in bold. The sequence characterized via shotgun approach is underlined. The X symbols, reported in the legumin entry A0A3Q7XNW1, refer to unknown amino acid residues. The cleavage site at level of the Asn-Gly bond is indicated by a red arrow. The two cysteine residues of the β-chain are reported in Italics bold and highlighted in grey. Panel (G) shows the schematic picture of the post-translational proteolytic cleavage generating the α- and β-chains, and the putative disulfide inter-chain bond which involves the cysteine located at the N-terminal region of the β-chains. The second cysteine residue of the β-chain, located in the C-terminal cupin domain and not involved in any disulfide bond, is reported with a free thiol group.*

Characterization of the Vicilin-enriched Fraction

Shotgun Analysis

The peptide-based bottom-up proteomics analysis of the the vicilin-enriched fraction allowed the identification of 12 proteins, which correspond to those already identified in the legumin-enriched fraction. The group of vicilins includes seven entries with the UniProt Acc. Nos. A0A1S2XQR4, A0A1S2Y087, A0A1S2YZ56, A0A1S2XYZ0, A0A1S2XQ88, Q304D4, and A0A1S2YKD9. This result confirms that a completely selective separation of these two classes of globulins was not achieved, although a partial enrichment of vicilins (41% of this fraction) was achieved. Protein identifications and matching peptides in the vicilin-enriched fraction are listed in Supplementary Table S15.

Top-down MS of the Vicilin-enriched Fraction

The reduced and IAA-alkylated vicilin-enriched fraction was also analyzed by LC–MS using a top-down approach for intact protein characterization. Applying the same strategy described above for the legumins, several polypeptides corresponding to specific amino acid regions of chickpea vicilins (Accession Nos. A0A1S2XQR4, Q304D4, and A0A1S2XQ88) were detected and characterized.

Panel (A) of Figure 30 displays the deconvoluted spectrum of two co-eluting polypeptides with experimental monoisotopic masses of 13750.0547 and 13996.1054 Da. TD-MS² analysis of the 13750.0547 Da component (see Supplementary Figures S22 and S23, Table S9) identified it as the amino acid region Arg209–Asn329 of the vicilin-like protein A0A1S2XQR4 (Figure 30, panel F). For the component with a monoisotopic mass of 13996.1054 Da, no TD-MS² data were obtained. However, based on its accurate intact mass and peptide-based MS data, this polypeptide was putatively assigned to the region Arg204–Asn327 of the chickpea vicilin Q304D4 (Figure 30, panel G).

Panel (B) of Figure 30 shows the deconvoluted spectrum of two co-eluting polypeptides with monoisotopic masses of 12801.3807 and 12829.3827 Da. Complete structural characterization of the 12801.3807 Da component was achieved by TD-MS² (Supplementary Figures S49 and S50, Table S17), identifying it as the Lys330–Lys443 region of the vicilin-like protein A0A1S2XQR4 (Figure 30, panel

F). Although no TD-MS² data were obtained for the 12829.3827 Da component, its mass, which is 28.002 Da higher than that of the previous peptide, suggests that it corresponds to the same region carrying a formyl modification (theoretical mass shift +27.9949 Da).

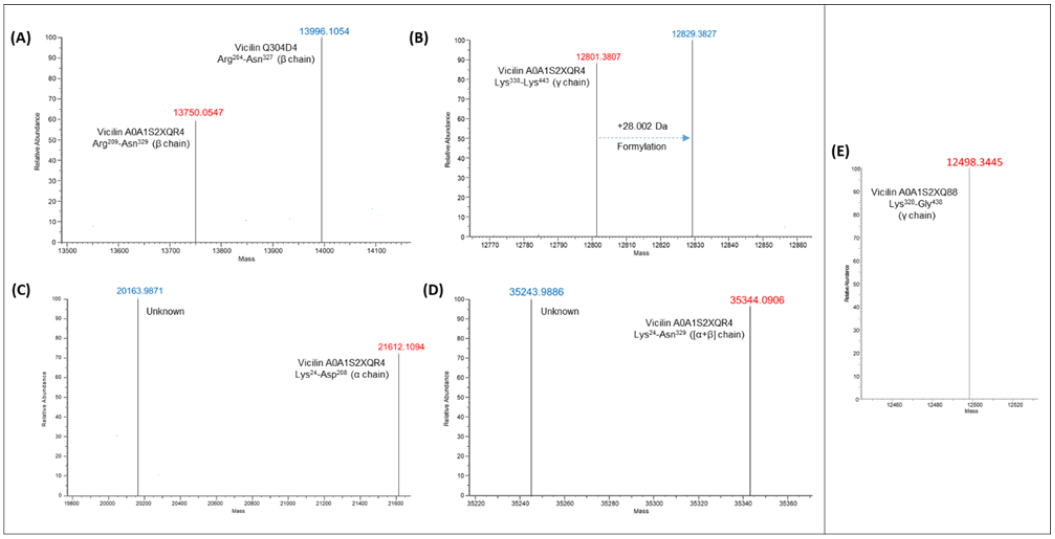
Panels (C) and (D) of Figure 30 display the deconvoluted spectra of two additional polypeptides with monoisotopic masses of 21612.1094 and 35344.0906 Da, respectively. Although TD-MS² spectra were not acquired for these components, their experimental masses match, within an error of approximately 1.5 ppm, the theoretical values of the amino acid regions 24–208 and 24–329, respectively, of the vicilin-like protein A0A1S2XQR4. These MS spectra also reveal two additional co-eluting components with monoisotopic masses of 20163.9871 and 35243.9886 Da, which, however, remain unidentified.

Finally, panel (E) of Figure 30 presents the deconvoluted mass spectrum of a polypeptide with a monoisotopic mass of 12498.3445 Da. TD-MS² analysis (Supplementary Figures S54 and S55, Table S18) identified this polypeptide as the Lys328–Gly438 region of the vicilin A0A1S2XQ88 (Figure 30, panel H).

The corresponding multiply charged ESI mass spectra of all vicilin-related polypeptides are shown in Supplementary Figures S45, S48, S51, S52, and S53. The results of the top-down analysis of the vicilin-enriched fraction are summarized in Table 6.

***Table 6**. The columns report the i) amino acid regions and the putative fragment chains of the vicilin proteins with the Acc. No. A0A1S2XQR4, Q304D4, and A0A1S2XQ88; ii) the experimental and theoretical monoisotopic masses, and relative delta mass in ppm. Amino acid positions are referred to the whole sequence (i.e., also containing the signal peptide).*

AA Region (chain)	Exp. M_{mono} (Da)	Theor. M_{mono} (Da)	ΔM (Exp. – Theor.) (ppm)
Vicilin A0A1S2XQR4			
Lys^{24}-Asp^{208} (α chain)	21612.1094	21612.0902	1.3
Arg^{209}-Asn^{329} (β chain)	13750.0547	13750.0357	2.8
Lys^{24}-Asn^{329} ([α+β] chain)	35344.0906	35344.1153	-0.7
Lys^{330}-Lys^{443} (γ chain)	12801.3807	12801.3926	-1.4
Vicilin Q304D4			
Arg^{204}-Asn^{327} (β chain)	13996.1054	13996.1096	-0.3
Vicilin A0A1S2XQ88			
Lys^{328}-Gly^{438}	12498.3445	12498.3336	0.9



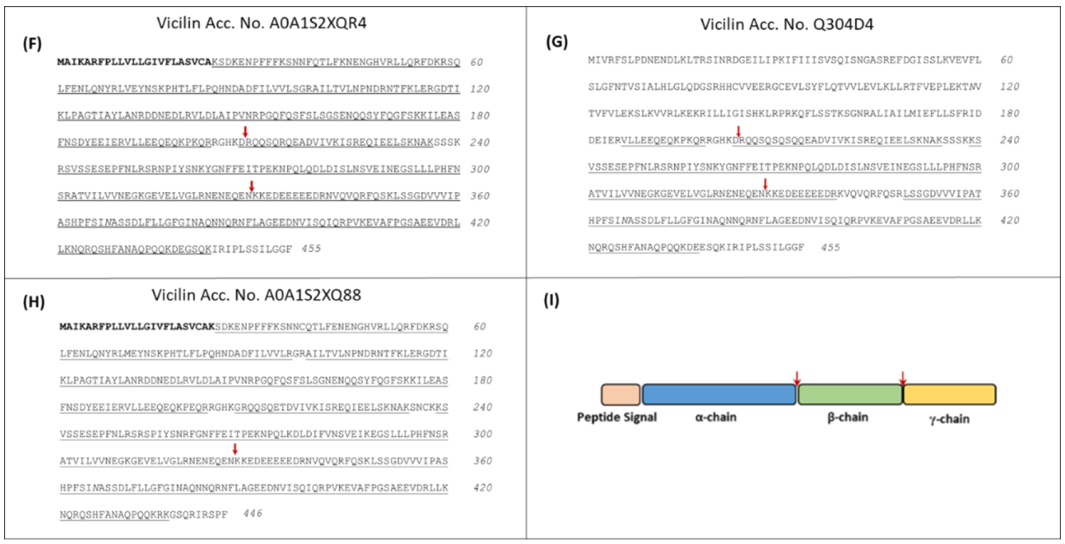


***Figure 29.** Monoisotopic deconvoluted mass spectra (mass zero-charge) of the post-translational polypeptide products coming from precursor vicilins with the Acc. No. A0A1S2XQR4, (panels A-D)*

Q304D4 (A), and A0A1S2XQ88 (E). The corresponding multi-charged ESI-MS are shown in the Supplementary Material. Panels (F), (G), and (H) show the amino acid sequence of the vicilins A0A1S2XQR4, Q304D4, and A0A1S2XQ88, respectively. Peptide signal, lacking in the mature form of the protein, is reported in bold. The lengh of the peptide signal for the vicilin Q304D4 is not reported in UniProt DB. The sequence characterized via shotgun approach is underlined. The cleavage sites at level of α/β and β/γ junctions are indicated by a red arrow.

Discussion and Conclusion

Different previous studies have demonstrated that legume species differ in the extent to which vicilins and legumins stored in seeds undergo proteolytic processing during seed development, while still maintaining the bicupin structural fold and the trimeric or hexameric assembly required for accumulation and packaging into protein storage vacuoles. As noted in the Introduction of this thesis, legumins are proteolytically cleaved into two polypeptide chains, designated as the α- and β-chains, which remain covalently linked via an inter-chain disulfide bond. Similarly, as first reported by Gatehouse et al.,[45] vicilins from garden pea (Pisum sativum L.) may undergo proteolytic cleavage before germination at two specific sites, resulting in three smaller polypeptides referred to as the α-, β-, and γ-chains.[45,46] Bourgeois et al.[47] further observed that the extent of vicilin cleavage during seed maturation varies among pea cultivars: in some, cleavage occurs at both sites; in others, at only one site; and in some cultivars, no cleavage is observed. These differences have been attributed to variations in vicilin amino acid sequences, environmental conditions, and seasonal factors. Similar processing of vicilins during seed maturation has been reported in other legume species, including white lupin (Lupinus albus L.),[184] cocoa bean (Theobroma cacao L.),[185] and lentil (Lens culinaris Medik.). In the present thesis, by integrating a gel-based (descrbed in the previous chapter), a gel-free shotgun approach, and a top-down proteomic method, it was possible to demonstrate that also in chickpea seeds these two classes of storage proteins undergo proteolytic processing during seed development. Integrating these three MS-based approaches an in-depth sequence characterization was achieved for several low-molecular-weight polypeptides generated by these proteolytic events.

β-chain polypeptides derived from legumins

In the legumin-enriched fraction, several polypeptides corresponding to specific C-terminal regions of three legumin entries, UniProt Accession Numbers A0A1S2XSB9, A0A1S2XVG1, and A0A3Q7XNW1, were identified and characterized. The amino acid sequences of these unreviewed (TrEMBL) UniProt entries are shown in Figures 26, 29, and 32. In their mature forms, legumins A0A1S2XSB9 and A0A3Q7XNW1 contain five cysteine residues, whereas A0A1S2XVG1 contains seven. For what concern the legumin entry A0A1S2XSB9, our data revealed a group of polypeptides corresponding to the region Gly^{334}-Asn^{514}, which differ in post-translational modifications (PTMs) of the cysteine residue at position 419. Particularly, this residue was identified as carbamidomethyl-cysteine, sulfinic acid, and sulfonic acid. The existence of polypeptides corresponding to this amino acid region of the legumin entry A0A1S2XSB9 can be due to a proteolytic event that cleaves the precursor legumin protein at the level of the Asn^{333}-Gly^{334} bond. All these polypeptides lack the last four amino acids, namely KAAA, which are shown in the C-termini of the corresponding entry A0A1S2XSB9 deposited in the UniProt DB. In this respect, it is interesting to note that by the shotgun approach, it was possible to obtain the coverage of about 90% of the whole protein sequence, but this amino acid trait has never been characterized. Altogether, these findings suggest that the precursor legumin protein present in the genotype "Pascià" here employed, in comparison with the protein sequence entry deposited in UniProt and taken as reference, has a shorter length and is probably lacking these four C-terminal amino acids. A similar pattern was observed for a group of four polypeptides related to the legumin entry A0A1S2XVG1. Two of these polypeptides span the region Gly353–Ala532, while the other two correspond to Gly353–His531, thus lacking the terminal alanine (Figure 29). As observed for A0A1S2XSB9, these polypeptides also exhibit PTMs at a cysteine residue Cys438 detected in sulfinic and sulfonic acid forms. In all cases, the N-terminal residue was glycine at position 353, suggesting a cleavage event at the Asn352–Gly353 bond in the precursor protein. Finally, three polypeptides related to the region Gly^{315}-Asn^{496} of the legumin entry A0A3Q7XNW1 were detected by the top-down approach. These components differ in the status of the cysteine residue at position 400, which was found as carbamidomethyl-cysteine, sulfinic acid, or sulfonic acid, respectively. These

components could arise from a proteolytic cleavage occurring at the level of the Asn^{314}-Gly^{315} bond of the precursor legumin protein.

Altogether, these findings highlight two interesting aspects worth elaborating upon. Firstly, all the 20kDa polypeptides derived from the legumins here characterized reasonably correspond to the β-chains that, as already observed in other legumes[45] are produced by the post-translational proteolytic cleavage of their legumin precursor. It is interesting to note that this proteolytic cleavage occurs at a single well-conserved Asn-Gly peptide bond of the precursor legumins, as well as for the 11S seed-storage proteins of a wide variety of plant species.[186] It has been demonstrated that this cleavage event is catalysed by an asparaginyl endopeptidase that has an absolute specificity for the asparagine on the N-terminal side of the cleaved peptide bond, but shows a lower specificity for the amino acids on the C-terminal side.[186]

Another interesting finding concerns the identification of different oxidation states of the thiol group in one of the two cysteine residues located within the legumin β-chains characterized in this study. Each β-chain contains two cysteine residues situated at conserved positions: one in the N-terminal region and the other within the central portion of the C-terminal cupin domain (see Figures 26, 29, and 32). Based on structural homology with soybean proglycinin,[38] it has been previously proposed that the inter-chain disulfide bond connecting the α- and β-chains involves the cysteine located in the N-terminal region of the β-chain, while the second cysteine remains as a free thiol. Cysteine residues possess unique physicochemical properties and are known to play diverse functional roles; in addition, they are particularly susceptible to various post-translational modifications (PTMs).[187] Cysteine residues not involved in disulfide bond formation (i.e., free cysteines) possess a thiol group that can undergo deprotonation in response to various interactions with the surrounding environment. This deprotonation results in the formation of a thiolate anion, which significantly enhances the nucleophilicity and reactivity of the cysteine side chain.[188] Particularly, these reactive cysteines can undergo many different oxidation states in response to different redox signals. Briefly, the initial oxidation of a thiol group leads to the formation of sulfenic acid (–SOH), an inherently unstable intermediate that serves as a precursor to various higher oxidation states,[189] including the relatively more stable sulfinic ($–SO_2H$) and sulfonic ($–SO_3H$) acids. Sulfinic and sulfonic acid forms are typically associated with oxidative distress, but while the S-

sulfinylation is a reversible status, the formation of the S-sulfonic, which is the most highly oxidized species of thiol, is completely irreversible. [190]

Results reported in this chapter revealed that the cysteine in the N-terminal portion of the β-chains, here characterized, was always identified as carbamidomethyl cysteine, as a consequence of the sample preparation that included the reactions of reduction and alkylation. Instead, the other cysteine residue of the β-polypeptide was also detected as sulfinic and sulfonic acid. These findings may indirectly suggest that, in chickpea legumins as well, the inter-chain disulfide bond linking the α- and β-chains involves the cysteine residue located in the N-terminal region of the β-polypeptide, while the cysteine within the C-terminal cupin domain likely exists as a free thiol.

α-, β-, and γ-chain polypeptides derived from vicilins

In the enriched fraction of vicilins, although six different vicilins were identified by the shogun approach, the top-down strategy allowed the detection and characterization of a group of polypeptides related only to three chickpea vicilins entries; namely, the entries deposited in the UniProt DB with the Acc. No. A0A1S2XQR4, Q304D4, and A0A1S2XQ88. Figure 35 shows the amino acid sequence of the chickpea vicilin-like protein entry with Acc. No. tr|A0A1S2XQR4. The amino acid region 1-23 represents the peptide signals which is cleaved during the protein maturation process and therefore lacking in the mature protein (amino acid residues 24-455). The mature form does not show any cysteine residue. By the shotgun approach, it was possible to obtain the coverage of the whole protein sequence (412 out of 432 amino acids present in the mature form, 95,4%), except for two tetrapeptides (i.e., RGHK and SSSK) and the last twelve C-terminal amino acids (i.e., IRIPLSSILGGF). The complete list of matched peptides of this vicilin entry is reported in Supplementary Table S8. In the top-down approach, four polypeptides corresponding to the amino acid regions Lys^{24}-Asp^{208}, Arg^{209}-Asn^{329}, Lys^{24}-Asn^{329}, and Lys^{330}-Lys^{443} were detected and characterized. By comparison with the homologous vicilins of the pea, these polypeptides might correspond to the so-called α-, β-, (α+β)-, and γ-chains, respectively, that are generated by a proteolytic event at two bonds. Specifically, the Asp^{208}-Arg^{209} (namely the α/β junction) and the Asn^{329}-Lys^{330} (the β/γ junction) peptide bonds. About these proteolytic cleavage sites, it is

interesting to note that their locations are similar to those found in pea vicilin[48], being the α/β junction situated in the linker region joining the two cupin domains, and the β/γ junction located in an exposed loop in the C-terminal cupin domain (Figure 35). Moreover, the β/γ junction in both pea and chickpea vicilins shows an asparagine residue at position P1 of the peptide bond. Differently, while the α/β junction, in pea vicilin, is located at the level of a Lys-Asp bond, in chickpea corresponds to an Asp-Arg bond. In pea vicilin, it has been reported that the cleavages of these two junctions are catalyzed by two different proteases: a serine protease for the α/β junction, and a legumain for the β/γ one.[48] Although out of the scope of the present work, we could hypothesize that the cleavage of the chickpea vicilin at both junction sites might be carried out by the same enzyme family, namely the legumains. Indeed, legumains are a family of cysteine endopeptidases, involved in legume storage protein-limited proteolysis and degradation,[191] with a cleavage specificity at the carboxylic side of asparagine and aspartic acid peptide bonds. However, a much lower efficiency of legumains is reported for the aspartic acid. In this respect, the MS data here reported indeed suggest that the intact vicilin is cleaved first at the β/γ junction (i.e., the Asn^{329}-Lys^{330} bond) and produces as intermediate the [α+β]-polypeptide, and the γ-polypeptide. Then, the [α+β]-polypeptide undergoes proteolytic cleavage at the α/β junction (Asp^{208}-Arg^{209} bond), leading to the formation of the α- and β-chains. However, it should also be noted that an initial vicilin cleavage at the α/β junction, followed by the cleavage at the β/γ junction, cannot be excluded a priori.

Moreover, it should be highlighted that the last twelve C-terminal amino acids (i.e., I^{444}RIPLSSILGGF^{455}) reported in the database entry A0A1S2XQR4 have never been identified in either of the approaches here adopted. Indeed, the γ-polypeptide shows the lysine at position 443, as a C-terminal amino acid. The absence of these twelve C-terminal amino acids could be related to an extensive post-translational endo-proteolytic processing at the C-terminal portion of the pro-vicilin, as already reported for pea vicilin.[192] Finally, taking into account that vicilin-like proteins from other Fabaceae (e.g., *Pea vulgaris*, *Pisum sativum L.*, *Vicia faba*, and *Glycine max*) have been reported as sparsely glycosylated,[193] and considering that the sequence entry of this vicilin shows a unique consensus Asn-Xaa-Ser/Thr glycosylation site, located at position 368, this post-translational modification was also investigated. However, N-glycosylation was not detected by the methods employed here, but we cannot rule out its existence.

Analogously, the top-down analysis also allowed us to identify two post-translational products related to the vicilins with Acc. No. Q304D4 and A0A1S2XQ88. Figure 37 shows the amino acid sequence of the chickpea vicilin-like protein entry Q304D4. By top-down analysis, a polypeptide with an experimental molecular mass corresponding to the region Arg^{204}-Asn^{327} of this protein was identified. As discussed above for the products of the vicilins A0A1S2XQR4, this polypeptide reasonably corresponds to the β-chain of the vicilin-like protein entry Q304D4. This β-chain is generated by the proteolytic events at level of the bonds Asp^{203}-Arg^{204} (the α/β junction) and the Asn^{327}-Lys^{328} (the β/γ junction), both catalysed b a legumains.

Chapter 4

Cross-Referencing Genomic and Proteomic Data for Storage Proteins in *Cicer arietinum L.*

INTRODUCTION

The post-genomic era has revolutionized the biological and biochemical researches, shifting the focus from simply cataloging genes to understanding the dynamic and functional output of the genome. With the deepening of functional genomics research, the connection and integrity of genes have become increasingly important, and a single-omic approach is often no longer sufficient to fully explain some biological phenomena. At this time, the growing field of ProteoGenomic, a research area that combines areas as genomics, proteomics, and transcriptomics in a multi-omics setup using both mass spectrometry and high-throughput sequencing technologies, could well address these issues. Currently, one of the main goals of this field are to aid genome annotation or to unravel the proteome complexity (identification of proteoforms). Many approaches and tools have been developed to integrate proteome and transcriptome data, particularly in plant biochemistry.[194] Recent developments have already delivered near whole genome sequences for many crops,[195,196] which paved the way to expand to other plant species including chickpea (*Cicer arietinum L.)*. However, the complexity of plant genomes, coupled with post-transcriptional and post-translational modifications (PTMs), often leads to discrepancies between predicted gene models and the actual suite of expressed proteins. Moreover, although ProteoGenomic analyses represent a powerful tool in the context of seed biology, plant storage proteins often share high sequence identity and undergo extensive PTMs proteolytic processing. This aspect may complicate their characterization when relying exclusively on silico prediction (Nesvizhskii, 2014). As reported in the Introduction section, the legume seed protein fraction, including chickpea seeds, is mainly constituted by vicilins (7S globulins) and legumins (11S globulins), a group of storage proteins which represents a primary nitrogen and carbon reserve during the seed germination and seedling establishment.

In this respect, this chapter is aimed to compare the proteomic data here obtained with the genomic ones, recently published for the chickpea, to validate proteomics data (e.g. validation of protein entries annotation). In particular, this cross-referencing analysis is focused on the storage proteins identified in Chapter 1 and listed in Table S2, due to their importance and relevance, and which also represent the most abundant components of the seed protein fraction of chickpeas. This

analysis was carried out querying the chickpea reference genome (Cicar.CDCFrontier_v2.0) deposited in the public database 'NCBI GeneBank'. Particularly, the genomic locus information of all the predicted storage proteins (i.e., legumin-like proteins, vicilin-like proteins, albumins 2S, and glutelins) were searched. Then, this information has been compared with the storage proteins positively identified by our proteomic data. In this respect, it is important to highlight that our MS data were searched against the Reference proteome database of chickpea deposited in the UniProt DB. This Reference proteome database is constituted by almost all unreviewed proteins entries from the TrEMBL section of UniProt. The unreviewed proteins entries are obtained by automatic translations of all coding sequences reported in the EMBL nucleotide database, are computer-annotated and therefore are awaiting manual annotation. So that, this comparative analysis allowed us: i) to check how many genes of storage protein does the chickpea genome harbor; ii) how many of these predicted gene products were detected by our proteomic data; iii) to validate the annotation, as reported in the Reference proteome database, of the storage protein entries detected. This check aims to critically advance the annotation of this key protein class in the *Cicer arietinum* L., thereby providing a valuable resource for future efforts in crop improvement.

DISCUSSION

As reported in the Introduction, the first completion of the chickpea genome sequencing (cultivar CDC Frontier, a kabuli type) was released in 2013 by Varshney et al. Now, the Cicer arietinum (chickpea) genome assembly Cicar.CDCFrontier_v2.0 from International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) is deposited in the NCBI GenBank and contains 29938 genes. 24841 out of these 29938 genes are protein-coding genes. Concerning the group of storage proteins, the reference genome shows 19 genes encoding a single specific gene product, and in particular 4 legumin-like proteins, 9 vicilins-like proteins, 4 albumins, and 2 glutelin-like proteins. The Gene id 101488953 encodes for two legumins (NCBI Reference Sequence Acc. No. Q9SMJ4.1 and XP_073221970.1), which share about 85% of sequence identity and about 90% of

sequence similarity. The complete list of these predicted storage proteins is reported in Table 7, that shows the gene/locus symbol, the GenBank description of the gene product and the NCBI Reference sequence Accession Number. For each of these predicted gene products, this Table also shows the corresponding protein entry in the Cicer arietinum reference proteome database, including the UniProt Acc. No., the description, and finally if it was identified in our proteomic analyses.

In detail, our proteomic approaches allowed the identification of 15 storage proteins, which correspond to predicted gene products of GenBank: six legumin-like proteins Acc. No. A0A1S3E1N3, Q9SMJ4, A0A1S2XSB9, A0A3Q7XNW1, A0A1S2XVG1, and A0A1S2YGT3), seven vicilin-like proteins (Acc. No. Q304D4, A0A1S2XQR4, A0A1S2XQ88, A0A1S2XYZ0, A0A1S2Y087, A0A1S2YKD9, and A0A1S2YZ56), an albumin-like protein (Acc. No. A0A3Q7K771), and a glutelin-like protein (Acc. No. A0A1S2YJV5).

Concerning the legumins, the two components with the UniProt Acc.No. Q9SMJ4 and A0A3Q7XNW1, correspond to the predicted gene products both encoded by the gene ID 101488953, and showing the gene symbol leg (also classified as leg3 or CLAI). This chickpea gene shares a high percentage of nucleotide sequence identity with the gene LegA of pea,[197] which encodes for a legumin protein classified as legumin-class I or legumin A. Therefore, these two legumins, sharing a sequence similarity of about 90% might be categorized as belonging to the class A. Analogously, the legumin with Acc. No. A0A1S2XSB9, identified by proteomic data, and showing about 75% of sequence identity with the two previous legumins A, is also annotated in both the GenBank and UniProt databases as 'Legumin A-like', and therefore can be categorized as belonging to this class.

In this respect, it is important to highlight that legumin proteins are encoded by a closely related family of genes that, based upon sequence divergence, have been subdivided into two classes: the “major” class I or A, and the “minor” class II or B.[198] The legumin genes of class A and B from pea and the corresponding predicted gene products have been described previously by Lycett et al.[199] They indicated that legumins of class A, show subunits with a higher molecular masses than those of the legumins B. Other studies, carried out by Müntz et al.,[200] and aimed to the characterization of the legumins A and B in Vicia Faba noticed that type A legumins

contain methionine residues, whereas legumins B appear methionine-free. Finally, Gatehouse et al.[201], characterized two very similar genes encoding 'minor' (B-type) legumins, and designated these genes as LegJ and LegK. Although considerable data on storage proteins and their genes are available for many legumes, very little information is available regarding legumin-like proteins of chickpea.

The legumins with the Acc. No. A0A1S3E1N3 and A0A1S2XVG1 are described in the UniProt DB as legumins J, and correspond to the predicted gene products of the genes LOC101489055 (Gene ID 101489055) and LOC101501269 (Gene ID 101501269), respectively. The predicted gene product of the gene LOC101501269 is annotated as legumin J, and corresponds to the annotation reported in UniProt DB. Differently, the predicted gene product of the gene LOC101489055 (i.e. the protein with UniProt Acc.No. A0A1S3E1N3) is described as 'legumin type B', but we cannot establish if this protein may be also categorized as legumin J. Finally, an additional protein with Acc. No. A0A1S2YGT3, whose description in both databases is reported as '11S globulin seed storage protein 2-like', was also identified in our proteomic investigation. This protein shows about 68% of sequence identity and about 80% of sequence similarity with the predicted chickpea Glutelin type-A 2-like (NCBI Acc. No. XP_004504183.1) here not detected. This result confirms that globulins and glutelins share a common evolutionary origin.

Regarding the vicilins protein components, by our proteomic approaches seven vicilin-like proteins were identified (Acc. No. A0A1S2XQR4, Q304D4, A0A1S2XYZ0, A0A1S2Y087, A0A1S2YKD9, A0A1S2YZ56). They correspond to seven out of the eight predicted gene products annotated in GenBank as "vicilin-like proteins".

Moreover, four Basic 7S globulin-like proteins (Acc. No. A0A1S2XV08, A0A1S2XXJ2, A0A1S2XVM3, and A0A1S2XXX1) were identified by proteomic data. The protein entry with the Acc. No. A0A1S2XV08 corresponds to the predicted gene product having the NCBI Acc. No. XP_004494958.1, encoded by the gene LOC101504144. This gene product is described in GenBank as "Basic 7S globulin-like protein". BLAST analysis of this protein entry evidenced that it shares about 35-39 % of sequence identity with the group of chickpea Basic 7S globulin-like proteins discussed below, and that are all annotated in GenBank as "probable aspartic

proteinase GIP2". The cross-referencing analysis of the other three putative Basic 7S globulin-like proteins highlighted that these protein entries correspond to three predicted gene products (NCBI Sequence ID 101509714, Locus LOC101509714; NCBI Sequence ID 101509718, Locus LOC101509718; and NCBI Sequence ID: 101509720, Locus LOC101509720) that, in GenBank, are annotated as "probable aspartic proteinase GIP2". Therefore, the differences observed between the description of the protein entries in UniProt and those reported in the corresponding predicted gene products in GenBank may be related to a mislabeling in the UniProt DB. Finally, an albumin (UniProt Acc. No. A0A3Q7K771) and a glutelin protein (Acc. No. A0A1S2YJV5) were identified by proteomics data. Specifically, these proteins have the description 'Albumin-2-like' and 'Glutelin type-D 1-like' in both UniProt and GenBank databases and correspond to the gene predicted products of the LOC101512722 and LOC101510077, respectively. BLAST analysis, restricted to the Chickpea taxonomy, of the 'Glutelin type-D 1-like' protein revealed that, among the other chickpea proteins, it shares the highest percentage (about 44%) of sequence identity with protein with the "11S globulin seed storage protein 2-like" protein" (Acc. No. A0A1S2YGT3) previously described. BLAST analysis of this Glutelin protein enlarged to the "Viridiplantae" taxonomy, highlighted a very high sequence identity (about 90-92%%) with storage proteins from other legumes, such as the proteins annotated as "Nutrient reservoir protein" from Medicago truncatula or the "Cupin type-1 domain-containing protein" from Trifolium subterraneum. This finding underscores that glutelins, as already reported in rice and soybean, derive from a common ancestral gene and therefore share high sequence identity with classical storage proteins.

***Table** 7. The table lists the Gene IDs and loci reported in GenBank, the corresponding gene product in GenBank, the NCBI Reference Sequence accession number, the UniProt accession number, the UniProt description, whether the protein has been identified, and the final annotation after cross-referencing.*

Locus Symbol/Gene ID	Gene Product Description in GenBank	NCBI Reference Sequence Acc.No.	Uniprot Acc.No	Uniprot description	Identified	Annotation after Cross-reference
LOC101489055	legumin type B	XP_004493780.1	A0A1S3E1N3	Legumin J	x	Legumin J
LEG_CICAR	Legumin	Q9SMJ4.1	Q9SMJ4	Legumin	x	Legumin A
LOC101489278	legumin A-like	XP_004493780.1	A0A1S2XSB9	Legumin A-like	x	Legumin A
LEG	legumin	XP_073221970.1	A0A3Q7XNW1	Legumin-like	x	Legumin A
LOC101501269	legumin J-like	XP_004495100.1	A0A1S2XVG1	Legumin J-like	x	Legumin J
LOC101494720	vicilin-like seed storage protein At2g28490	XP_004487139.1	A0A1S2XCB5	Vicilin-like seed storage protein At2g28490		
LOC101515194	vicilin-like	NP_001296635.1	Q304D4	Provicilin (Vicilin-like)	x	Vicilin-like
LOC101515515	vicilin-like	XP_004493035.1	A0A1S2XQR4	vicilin-like	x	Vicilin-like
LOC101505411	vicilin-like	XP_073222640.1	A0A1S2XQ88	vicilin-like	x	Vicilin-like
LOC101510367	provicilin-like	XP_004496703.1	A0A1S2XYZ0	Provicilin-like	x	Vicilin-like
LOC101510695	vicilin-like	XP_004496704.1	A0A1S2Y087	Vicilin-like	x	Vicilin-like
LOC101502994	vicilin-like seed storage protein At2g28490	XP_004505960.1	A0A1S2YKD9	vicilin-like seed storage protein At2g28490	x	Vicilin-like
LOC101506263	vicilin-like seed storage protein At4g36700	XP_004512222.4	A0A1S2YZ56	Vicilin-like seed storage protein At2g18540	x	Vicilin-like
LOC101488886	11S globulin seed storage protein 2-like	XP_004504184.1	A0A1S2YGT3	11S globulin seed storage protein 2-like	x	Vicilin-like
LOC101504144	basic 7S globulin-like	XP_004494958.1	A0A1S2XV08	SBg7S	x	Vicilin-like
LOC101509384	Albumin-1 D-like	XP_004514273.1	A0A1S2Z3C2	Albumin-1 D-like		\
LOC101507760	Albumin-2-like	XP_004488690.1	A0A1S2XHS6	Albumin-2-like		\
LOC101512722	Albumin-2-like	NP_001351664.1	A0A3Q7K771	Albumin-2-like	x	Albumin-2-like
LOC101505722	Albumin-2-like (Lectin protein C-25)	NP_001296633.1	R9TPI6	Albumin-2-like (Lectin protein C-25)		\
LOC101488538	Glutelin type-A 2-like	XP_004504183.1	A0A1S3EA35	Glutelin type-A 2-like		\
LOC101510077	Glutelin type-D 1-like	XP_004505897.1	A0A1S2YJV5	Glutelin type-D 1-like	x	Glutelin-like
LOC101509714	probable aspartic proteinase GIP2	XP_004495218.1	A0A1S2XXJ2	Basic 7S globulin	x	probable aspartic proteinase GIP2
LOC101495910	probable aspartic proteinase GIP2	XP_004486143.1	A0A1S2XA90	Basic 7S globulin		probable aspartic proteinase GIP2

LOC101499516	probable aspartic proteinase GIP2	XP_004486304.1	A0A1S2XDK5	Basic 7S globulin 2-like		probable aspartic proteinase GIP2
LOC101505205	probable aspartic proteinase GIP2	XP_004515101.1	A0A1S2Z4V7	Basic 7S globulin 2-like		probable aspartic proteinase GIP2
LOC101510034	probable aspartic proteinase GIP2	XP_004495219.1	A0A1S2XVM3	Basic 7S globulin-like	x	probable aspartic proteinase GIP2
LOC101496241	probable aspartic proteinase GIP2	XP_004486144.1	A0A1S2XA89	Basic 7S globulin-like		probable aspartic proteinase GIP2
LOC101509822	probable aspartic proteinase GIP2	XP_004495384.1	A0A1S2XXX1	Basic 7S globulin-like	x	probable aspartic proteinase GIP2
LOC101500045	probable aspartic proteinase GIP2	XP_004508379.1	A0A1S2YQK9	Basic 7S globulin-like		probable aspartic proteinase GIP2

CONCLUSIONS

During the past decade, the interest in the chickpea has increased because it represents an interesting source of proteins for the development of non-gluten, protein-enriched ingredients. However, to date, comprehensive proteomic studies about the protein fraction of chickpea seeds are lacking. In the present Ph.D. thesis, three different goals were reached in order: i) to obtain an in-depth characterization of its protein fraction; ii) to show the biological processes in which the proteins are involved during plant and seed development; and iii) to explore the plant's response (for the Pascià genotype) to two different water regimes. Moreover, for the first time, an in-depth structural characterization of the main storage proteins (i.e. legumins and vicilins) was achieved by joining the advantages of the bottom-up and top-down MS-based approaches.

Firstly, the integration of key analytical strategies, namely, shotgun proteomics, high-resolution mass spectrometry, and a species-specific protein database limited to *Cicer arietinum*, enabled a significant expansion of the current knowledge of the chickpea seed proteome, resulting in the identification of over 700 previously undetected proteins, overcoming the limitations of traditional methods previously used. The key point was the use of a *Cicer arietinum*-specific database for MS data analysis, avoiding common issues associated with broad, nonspecific databases such as the *Viridiplantae* protein database used in earlier research. This analytical strategy, which included label-free quantification, was applied in a comparative case study involving two samples of the chickpea genotype "Pascià" grown under different water regimes (rainfed vs. irrigated). The analysis showed that rainfed conditions led to the overexpression of potentially allergenic proteins and enzymes involved in plant defense mechanisms. In contrast, the irrigated samples exhibited numerous unique proteins, absent under rainfed conditions, primarily related to translational regulation and branched-chain amino acid (BCAA) metabolism, including eukaryotic initiation factors (eIFs), aminoacyl-tRNA synthetases (ARSs), and ribosomal proteins. These proteins may contribute to translational control in response to environmental cues and support root development, defense, and growth. The detailed characterization of chickpea legumins and vicilins was achieved by the use of both gel-based and gel-free bottom-up approaches, integrated with MS top-

down analyses on intact proteins. The gel-based approach allowed us preliminarily to reveal the legumins as whole subunits (α- and β-chains linked together), but also as α- and β-chains migrating alone over the gel lane. Additionally, the gel-based approach indicated that vicilins, much like legumins and as previously observed in pea seeds, are initially synthesized as pre-propolypeptides, which can undergo proteolytic processing to yield subunits of lower molecular weight. In light of these findings, combined bottom-up and top-down proteomic approaches were employed to in-depth investigate the proteolytic products of legumins and vicilins in *Cicer arietinum* (chickpea). The mass spectrometry results reveal that, in chickpea legumins, the cleavage event responsible for generating the α- and β-chains is most likely catalyzed by an asparaginyl endopeptidase. This enzymatic cleavage occurs at a highly conserved Asn–Gly bond within the legumin precursor, in agreement with mechanisms previously described for 11S seed storage proteins in other plant species. Additionally, our results suggest that, as in soybean proglycinins, the inter-chain disulfide bond connecting the α- and β-chains likely involves the cysteine residue located in the N-terminal region of the β-polypeptide. In contrast, the second cysteine residue located in the β-polypeptide, and situated within the C-terminal cupin domain, is probably present in its reduced form, as a free thiol. On the other hand, characterization of the degradation products of vicilin proteins evidences that, like in other legumes, chickpea vicilins show two proteolytic cleavage sites located in conserved amino acid regions, showing an asparagine or an aspartate residue at position P1 of the peptide bond cleaved. Consequently, legumains, a family of cysteine endopeptidases with a cleavage specificity at the carboxylic side of asparagine and aspartic acid peptide bonds, might catalyze these proteolytic events.

To summarize, in this PhD thesis, different aspects of proteomics hereby cited have been addressed. Different methods have been developed and optimized for different kinds of characterization of chickpea seeds' protein fraction. Furthermore, a keen data analysis using different tools and software for proteomic analyses has been established to deeply investigate not only the entire protein fraction, but also to understand the post-translational processes involving the most abundant seed proteins, namely globulins, during the seeds' maturation.

Finally, the proteomic data here obtained for the storage proteins were compared with the genomic ones, recently published for the chickpea. This comparison allowed us to check how many genes of storage protein does the chickpea genome harbor,

and how many of these ones were detected by our proteomic data. Moreover, it was possible to check the annotation, as reported in the Reference proteome database, of the storage protein entries detected. Overall, most of the predicted gene products were detected at protein level and correctly annotated. Annotation discrepancies were evidenced for a group of proteins classified in the Reference Proteome as "Basic 7S globulin-like proteins", but probably belonging to the "aspartic proteinase GIP2" family, as described by genomic data.

Although significant progress has been made, the future directions and emerging applications of proteomics promise to open new and exciting frontiers in protein research.

ACKNOWLEDGMENTS

I would like to thank all the people who have contributed to this thesis.
First of all, my PhD supervisor, Prof. Vincenzo Cunsolo. He helped me along during these three years. I am deeply grateful for your invaluable guidance and support throughout my PhD journey.
I would also like to thank the LSMO group: Prof. Rosaria Saletti and my colleagues Dr. Marinella Pittalà and Dr. Antonella Di Francesco for their availability and kindness. It was a pleasure to work with you!
I would like to acknowledge also the Bio-Nanotech Research and Innovation Tower (BRIT; PON project financed by the Italian Ministry for Education, University and Research, MIUR) of the University of Catania, for the availability of the Thermo Fisher Orbitrap Fusion mass spectrometer used for the analyses.
I kindly acknowledge Prof. Zina Flagella and Dr. Michele A. De Santis of the University of Foggia for providing chickpea samples for the analysis.
I would like to acknowledge Prof. Ole Nørregaard Jensen, who welcomed me at the PRGroup of the University of Southern Denmark (SDU) in Odense. He integrated me as I was part of his group, sustained and trusted me. I would like to thank all the beautiful people I met in Odense for teaching me a lot in the field of proteomics and protein research, in particular, my brazilian brother Victor, for being a brilliant mind to work with and a constant friend to rely on. Thank you for making this path lighter and brighter.
And finally, thanks to my family, which always supports my choices and passions, even if they lead to complicated paths. Thanks to Grazia, who is always my landmark and my boost to keep following my dreams, and she is always on my side during these difficult years of change.

Publications and Conference contributions

List of publications related to the PhD project:

1) Antonella Di Francesco, Michele Andrea De Santis, **Aldo Lanzoni**, Maria Gaetana Giovanna Pittalà, Rosaria Saletti, Zina Flagella, Vincenzo Cunsolo. Mass Spectrometry Characterization of the SDS-PAGE Protein Profile of Legumins and Vicilins from Chickpea Seed. *Foods* (**2024)** Mar 14;13(6):887**. doi**: 10.3390/foods13060887.
2) Antonella Di Francesco, **Aldo Lanzoni**, Michele Andrea De Santis, Maria Gaetana Giovanna Pittalà, Rosaria Saletti, Zina Flagella, Vincenzo Cunsolo. In-Depth Shotgun Proteomic Characterization of Chickpea (*Cicer arietinum L.*): A Case Study on the Genotype Pascià Grown Using Two Water Regimes. *ACS Agricultural Science & Technology* (**2025)** Vol5/Issue7. https://doi.org/10.1021/acsagscitech.4c00705.

List of conference contributions related to the PhD project:

Oral communications (as presenting author):

1) **A. Lanzoni**, A. Di Francesco, M. G. G. Pittalà, R. Saletti, M. A. De Santis, Z. Flagella, V. Cunsolo. "Shotgun Proteomic Analysis of an Italian Chickpea Genotype". 25th – 26th 2023 Bari, 11th May MS Jday. Abstract Book page 31.
2) **A. Lanzoni**, A. Di Francesco, M. G. G. Pittalà, R. Saletti, M. De Santis, Z. Flagella, V. Cunsolo. "Proteomic analysis of an Italian Chickpea Genotype

by MS-methods". ItPA HPS and SePA XVII International Congress, Roma, Italy, November 29th – December 1st, 2023. Abstract Book page. 24

Oral communications (as co-author):

1) Di Francesco, A. Cucina, M. G. G. Pittalà, **A. Lanzoni**, C. Mills, R. Saletti, S. Foti, V. Cunsolo. "*The development of a manually curated database of the metabolic proteins of a Triticum aestivum"* (TriMet_DB). 20th – 22th September 2022 Carlentini, Massa 2022, Società Chimica Italiana, Divisione di Spettrometria di Massa. Abstract book page. 20.
2) V. Cunsolo, A. Di Francesco, M.A. De Santis, **A. Lanzoni**, M.G.G. Pittalà, R. Saletti, Z. Flagella. "Structural Characterization of Legumins and Vicilins from Chickpea Seeds by Mass Spectrometry". SCI 2024 XXVIII National Congress "Elementi di Futuro" Milano 26-30 August 2024. Abstract Book Volume 2 page. 1207.
3) M.G.G. Pittalà, A. Di Francesco, S. Conti-Nibali, S. Reina, V. Cunsolo, A. Cucina, **A. Lanzoni**, S. Foti, V. De Pinto, R. Saletti. "Characterization of disulfide bonds in the rat VDAC3 protein by high resolution mass spectrometry". SCI 2024 XXVIII National Congress "Elementi di Futuro" Milano 26-30 August 2024. Abstract Book Volume 2 page. 1230

Poster communications (as presenting author):

1) **A. Lanzoni**, A. Di Francesco, M. G. G. Pittalà, R. Saletti, M. A. De Santis, Z. Flagella, V. Cunsolo. "MSbased investigation of the protein fraction of Chickpea". 10th – 14th September 2023 Roma, XLI Convegno Nazionale della Divisione di Chimica Organica della Società Chimica Italiana. Abstract book pag. 237.
2) **A. Lanzoni**, A. Di Francesco, M. G. G. Pittalà, R. Saletti, M. A. De Santis, Z. Flagella, V. Cunsolo. "MS-based investigation of chickpea seeds grown under different water conditions" SCI 2024 XXVIII National Congress

"Elementi di Futuro" Milano 26-30 August 2024. Abstract Book Volume 2 page. 1386.

3) **A. Lanzoni**, A. Di Francesco, M. G. G. Pittalà, R. Saletti, V. Cunsolo. "Characterization of Vicilin-like proteins from chickpea seeds by coupling Bottom-up and Top-Down Approaches". ITPA2024 XVIII INTERNATIONAL ANNUAL MEETING, joint with HPS and SEPA, Roma, Italy, November 27th –29th, 2024.

Poster communications (as co-author):

1) Maria G. G. Pittalà, Antonella Di Francesco, Vincenzo Cunsolo, **Aldo Lanzoni**, Loredana Leggio, Greta Paternò, Nunzio Iraci, Rosaria Saletti, N. Iraci. "Proteomic profile of Extracellular Vesicles secreted by Astrocytes using a shotgun proteomic approach and high resolution mass spectrometry". ItPA HPS and SePA XVII International Congress, Roma, Italy, November 29th – December 1st, 2023. Abstract Book page. 69

2) M.G.G. Pittalà, L. Leggio, A. Di Francesco, V. Cunsolo, **A. Lanzoni**, N. Iraci, R. Saletti. "Characterization of the protein fraction by high resolution mass spectrometry reveals specific post-translational modifications in extracellular vesicles secreted by astrocytes". SCI 2024 XXVIII National Congress "Elementi di Futuro" Milano 26-30 August 2024. Abstract Book Volume 2 page. 1420.

3) M. Falcone, L. Leggio, M.G.G. Pittalà, A. Di Francesco, M. Catania, E. Alpi, V. Cunsolo, **A. Lanzoni**, R. Saletti, N. Iraci. "High-resolution mass spectrometry reveals specific post-translational modifications in the proteome of astrocyte-derived extracellular vesicles". Astreocytes Cafè Trieste 10-12 July 2024.

4) A. Di Francesco, A. Spina, A. Lanzoni, M.G.G. Pittalà, R. Saletti, V. Cunsolo. "Comparative proteomic analysis by shotgun approach of two

cultivars of Lupinus angustifolius: Polo and Boregine. SCI 2024 XXVIII National Congress “Elementi di Futuro” Milano 26-30 August 2024. Abstract Book Volume 2 page. 1810

Schools and workshops:

1) “1st Ms Biostats School” organized by “Divisione di Spettrometria di Massa” of Società Chimica Italiana, Dipartimento di Medicina e Chirurgia, University of Milano-Bicocca 20854 Vedano al Lambro (MB) 10-12 May 2023.

Bibliography

(1) Ghelfi, R.; Palmieri, A. Pulses Production in Italy: Trade, Marketing and Policy Issues. *Ital. J. Agron.* **2017**, *12* (3), 891. https://doi.org/10.4081/ija.2017.891.

(2) Moreno-Jiménez, M. R.; Cervantes-Cardoza, V.; Gallegos-Infante, J. A.; González-Laredo, R. F.; Estrella, I.; García-Gasca, T. de J.; Herrera-Carrera, E.; Díaz-Rivas, J. O.; Rocha-Guzmán, N. E. Phenolic Composition Changes of Processed Common Beans: Their Antioxidant and Anti-Inflammatory Effects in Intestinal Cancer Cells. *Food Res. Int.* **2015**, *76*, 79–85. https://doi.org/10.1016/j.foodres.2014.12.003.

(3) Kouris-Blazos, A.; Belski, R. Health Benefits of Legumes and Pulses with a Focus on Australian Sweet Lupins. *Asia Pac. J. Clin. Nutr.* **2016**, *25* (1), 1–17. https://doi.org/10.6133/apjcn.2016.25.1.23.

(4) Muzquiz, M.; Varela, A.; Burbano, C.; Cuadrado, C.; Guillamón, E.; Pedrosa, M. M. Bioactive Compounds in Legumes: Pronutritive and Antinutritive Actions. Implications for Nutrition and Health. *Phytochem. Rev.* **2012**, *11* (2), 227–244. https://doi.org/10.1007/s11101-012-9233-9.

(5) Mubaiwa, J.; Fogliano ,Vincenzo; Chidewe ,Cathrine; and Linnemann, A. R. Hard-to-Cook Phenomenon in Bambara Groundnut (Vigna Subterranea (L.) Verdc.) Processing: Options to Improve Its Role in Providing Food Security. *Food Rev. Int.* **2017**, *33* (2), 167–194. https://doi.org/10.1080/87559129.2016.1149864.

(6) Nieuwenhuis, R.; Nieuwelink, J. Cultivation of Soya and Other Legumes.

(7) Dutta, A.; Trivedi, A.; Nath, C. P.; Gupta, D. S.; Hazra, K. K. A Comprehensive Review on Grain Legumes as Climate-smart Crops: Challenges and Prospects. *Environ. Chall.* **2022**, *7*, 100479. https://doi.org/10.1016/j.envc.2022.100479.

(8) Kumar, N.; Hazra, K. K.; Nath, C. P.; Praharaj, C. S.; Singh, U. Grain Legumes for Resource Conservation and Agricultural Sustainability in South Asia. In *Legumes for Soil Health and Sustainable Management*; Meena, R. S., Das, A., Yadav, G. S., Lal, R., Eds.; Springer: Singapore, 2018; pp 77–107. https://doi.org/10.1007/978-981-13-0253-4_3.

(9) López-Barrios, L.; Gutiérrez-Uribe, J. A.; Serna-Saldívar, S. O. Bioactive Peptides and Hydrolysates from Pulses and Their Potential Use as Functional Ingredients. *J. Food Sci.* **2014**, *79* (3), R273-283. https://doi.org/10.1111/1750-3841.12365.

(10) Messina, M. J. Legumes and Soybeans: Overview of Their Nutritional Profiles and Health Effects. *Am. J. Clin. Nutr.* **1999**, *70* (3 Suppl), 439S-450S. https://doi.org/10.1093/ajcn/70.3.439s.

(11) Rebello, C. J.; Greenway, F. L.; Finley, J. W. A Review of the Nutritional Value of Legumes and Their Effects on Obesity and Its Related Co-Morbidities. *Obes. Rev.* **2014**, *15* (5), 392–407. https://doi.org/10.1111/obr.12144.

(12) Boukid, F.; Rosell, C. M.; Castellari, M. Pea Protein Ingredients: A Mainstream Ingredient to (Re)Formulate Innovative Foods and Beverages. *Trends Food Sci. Technol.* **2021**, *110*, 729–742. https://doi.org/10.1016/j.tifs.2021.02.040.

(13) Hickisch, A.; Beer, R.; Vogel, R. F.; Toelstede, S. Influence of Lupin-Based Milk Alternative Heat Treatment and Exopolysaccharide-Producing Lactic Acid Bacteria on the Physical Characteristics of Lupin-Based Yogurt Alternatives. *Food Res. Int.* **2016**, *84*, 180–188. https://doi.org/10.1016/j.foodres.2016.03.037.

(14) Goldstein, N.; Reifen, R. The Potential of Legume-Derived Proteins in the Food Industry. *Grain Oil Sci. Technol.* **2022**, *5* (4), 167–178. https://doi.org/10.1016/j.gaost.2022.06.002.

(15) Duranti, M. Grain Legume Proteins and Nutraceutical Properties. *Fitoterapia* **2006**, *77* (2), 67–82. https://doi.org/10.1016/j.fitote.2005.11.008.

(16) Mani-López, E.; Palou, E.; López-Malo, A. Legume Proteins, Peptides, Water Extracts, and Crude Protein Extracts as Antifungals for Food Applications. *Trends Food Sci. Technol.* **2021**, *112*, 16–24. https://doi.org/10.1016/j.tifs.2021.03.035.

(17) Turner, L.; Hellens, R. P.; Lee, D.; Ellis, T. H. N. Genetic Aspects of the Organization of Legumin Genes in Pea. *Plant Mol. Biol.* **1993**, *22* (1), 101–112. https://doi.org/10.1007/BF00038999.

(18) *Encyclopedia of Human Nutrition*; 2012.

(19) Fujiwara, T.; Nambara, E.; Yamagishi, K.; Goto, D. B.; Naito, S. Storage Proteins. *Arab. Book* **2002**, *1*, e0020. https://doi.org/10.1199/tab.0020.

(20) The Vegetable Proteins. *Nature* **1924**, *114* (2875), 822–822. https://doi.org/10.1038/114822c0.

(21) Neji, C.; Semwal, J.; Kamani, M. H.; Máthé, E.; Sipos, P. Legume Protein Extracts: The Relevance of Physical Processing in the Context of Structural, Techno-Functional and Nutritional Aspects of Food Development. *Processes* **2022**, *10* (12), 2586. https://doi.org/10.3390/pr10122586.

(22) *Food Chemistry*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009. https://doi.org/10.1007/978-3-540-69934-7.

(23) *Natural Food Additives, Ingredients and Flavourings*; 2012.

(24) Goyal, A. K. *Grain Legumes*; IntechOpen, 2016.

(25) Boye, J.; Zare, F.; Pletch, A. Pulse Proteins: Processing, Characterization, Functional Properties and Applications in Food and Feed. *Food Res. Int.* **2010**, *43* (2), 414–431. https://doi.org/10.1016/j.foodres.2009.09.003.

(26) Stone, A. K.; Nosworthy, M. G.; Chiremba, C.; House, J. D.; Nickerson, M. T. A Comparative Study of the Functionality and Protein Quality of a Variety of Legume and Cereal Flours. *Cereal Chem.* **2019**, *96* (6), 1159–1169. https://doi.org/10.1002/cche.10226.

(27) Youle, R. J.; Huang, A. H. C. Occurrence of Low Molecular Weight and High Cysteine Containing Albumin Storage Proteins in Oilseeds of Diverse Species. *Am. J. Bot.* **1981**, *68* (1), 44–48. https://doi.org/10.2307/2442990.

(28) Ericson, M. L.; Rödin, J.; Lenman, M.; Glimelius, K.; Josefsson, L. G.; Rask, L. Structure of the Rapeseed 1.7 S Storage Protein, Napin, and Its Precursor. *J. Biol. Chem.* **1986**, *261* (31), 14576–14581.

(29) Hara-Hishimura, I.; Takeuchi, Y.; Inoue, K.; Nishimura, M. Vesicle Transport and Processing of the Precursor to 2S Albumin in Pumpkin. *Plant J. Cell Mol. Biol.* **1993**, *4* (5), 793–800. https://doi.org/10.1046/j.1365-313x.1993.04050793.x.

(30) Galau, G. A.; Wang, H. Y.; Hughes, D. W. Cotton Mat5-A (C164) Gene and Mat5-D cDNAs Encoding Methionine-Rich 2S Albumin Storage Proteins. *Plant Physiol.* **1992**, *99* (2), 779–782. https://doi.org/10.1104/pp.99.2.779.

(31) Sharief, F. S.; Li, S. S. Amino Acid Sequence of Small and Large Subunits of Seed Storage Protein from Ricinus Communis. *J. Biol. Chem.* **1982**, *257* (24), 14753–14759. https://doi.org/10.1016/S0021-9258(18)33344-1.

(32) Lilley, G. G.; Inglis, A. S. Amino Acid Sequence of Conglutin δ, a Sulfur-Rich Seed Protein of *Lupinus Angustifolius* L. *FEBS Lett.* **1986**, *195* (1), 235–241. https://doi.org/10.1016/0014-5793(86)80167-3.

(33) Shewry, P. R.; Tatham, A. S. The Prolamin Storage Proteins of Cereal Seeds: Structure and Evolution. *Biochem. J.* **1990**, *267* (1), 1–12.

(34) Shewry, P. R.; Tatham, A. S.; Forde, J.; Kreis, M.; Miflin, B. J. The Classification and Nomenclature of Wheat Gluten Proteins: A Reassessment. *J. Cereal Sci.* **1986**, *4* (2), 97–106. https://doi.org/10.1016/S0733-5210(86)80012-1.

(35) Templeman, T. S.; DeMaggio, A. E.; Stetler, D. A. Biochemistry of Fern Spore Germination: Globulin Storage Proteins in Matteuccia Struthiopteris L. 1. *Plant Physiol.* **1987**, *85* (2), 343–349.

(36) Heim, U.; Schubert, R.; Bäumlein, H.; Wobus, U. The Legumin Gene Family: Structure and Evolutionary Implications of Vicia Faba B-Type Genes and Pseudogenes. *Plant Mol. Biol.* **1989**, *13* (6), 653–663. https://doi.org/10.1007/BF00016020.

(37) Horstmann, C.; Schlesier, B.; Otto, A.; Kostka, S.; Müntz, K. Polymorphism of Legumin Subunits from Field Bean (Vicia Faba L. Var. Minor) and Its Relation to the Corresponding Multigene Family. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **1993**, *86* (7), 867–874. https://doi.org/10.1007/BF00212614.

(38) Adachi, M.; Takenaka, Y.; Gidamis, A. B.; Mikami, B.; Utsumi, S. Crystal Structure of Soybean Proglycinin A1aB1b Homotrimer. *J. Mol. Biol.* **2001**, *305* (2), 291–305. https://doi.org/10.1006/jmbi.2000.4310.

(39) Müntz, K. Proteases and Proteolytic Cleavage of Storage Proteins in Developing and Germinating Dicotyledonous Seeds. *J. Exp. Bot.* **1996**, *47* (5), 605–622. https://doi.org/10.1093/jxb/47.5.605.

(40) Vu Huu Thanh; Kazuo Shibasaki. Beta-Conglycinin from Soybean Proteins. Isolation and Immunological and Physicochemical Properties of the Monomeric Forms. *Biochim. Biophys. Acta BBA - Protein Struct.* **1977**, *490* (2), 370–384. https://doi.org/10.1016/0005-2795(77)90012-5.

(41) Doyle, J. J.; Schuler, M. A.; Godette, W. D.; Zenger, V.; Beachy, R. N.; Slightom, J. L. The Glycosylated Seed Storage Proteins of Glycine Max and Phaseolus Vulgaris. Structural Homologies of Genes and Proteins. *J. Biol. Chem.* **1986**, *261* (20), 9228–9238.

(42) Bassüner, R.; Hai, N. V.; Jung, R.; Saalbach, G.; Müntz, K. The Primary Structure of the Predominating Vicilin Storage Protein Subunit from Field Bean Seeds (Vicia Faba L. Var. Minor Cv. Fribo). *Nucleic Acids Res.* **1987**, *15* (22), 9609. https://doi.org/10.1093/nar/15.22.9609.

(43) Weschke, W.; Bassüner, R.; Van Hai, N.; Czihal, A.; Baümlein, H.; Wobus, U. The Structure of a *Vicia Faba* Vicilin Gene. *Biochem. Physiol. Pflanz.* **1988**, *183* (2), 233–242. https://doi.org/10.1016/S0015-3796(88)80107-0.

(44) Newbigin, E. J.; Delumen, B. O.; Chandler, P. M.; Gould, A.; Blagrove, R. J.; March, J. F.; Kortt, A. A.; Higgins, T. J. Pea Convicilin: Structure and Primary Sequence of the Protein and Expression of a Gene in the Seeds of Transgenic Tobacco. *Planta* **1990**, *180* (4), 461–470. https://doi.org/10.1007/BF02411442.

(45) Gatehouse, J. A.; Lycett, G. W.; Croy, R. R.; Boulter, D. The Post-Translational Proteolysis of the Subunits of Vicilin from Pea (Pisum Sativum L.). *Biochem. J.* **1982**, *207* (3), 629–632. https://doi.org/10.1042/bj2070629.

(46) Gatehouse, J. A.; Lycett, G. W.; Delauney, A. J.; Croy, R. R.; Boulter, D. Sequence Specificity of the Post-Translational Proteolytic Cleavage of Vicilin, a Seed Storage Protein of Pea (Pisum Sativum L.). *Biochem. J.* **1983**, *212* (2), 427–432. https://doi.org/10.1042/bj2120427.

(47) 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), 254–271. https://doi.org/10.1002/pmic.200700903.

(48) Wilson, K. A.; Tan-Wilson, A. Proteases Catalyzing Vicilin Cleavage in Developing Pea (Pisum Sativum L.) Seeds. *J. Plant Physiol.* **2018**, *224–225*, 86–94. https://doi.org/10.1016/j.jplph.2018.03.015.

(49) Shutov, A. D.; Bäumlein, H.; Blattner, F. R.; Müntz, K. Storage and Mobilization as Antagonistic Functional Constraints on Seed Storage Globulin Evolution. *J. Exp. Bot.* **2003**, *54* (388), 1645–1654. https://doi.org/10.1093/jxb/erg165.

(50) Sanchez-Monge, R.; Lopez-Torrejón, G.; Pascual, C. Y.; Varela, J.; Martin-Esteban, M.; Salcedo, G. Vicilin and Convicilin Are Potential Major Allergens from Pea. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* **2004**, *34* (11), 1747–1753. https://doi.org/10.1111/j.1365-2222.2004.02085.x.

(51) Lawrence, M. C.; Izard, T.; Beuchat, M.; Blagrove, R. J.; Colman, P. M. Structure of Phaseolin at 2.2 A Resolution. Implications for a Common Vicilin/Legumin Structure and the Genetic Engineering of Seed Storage Proteins. *J. Mol. Biol.* **1994**, *238* (5), 748–776. https://doi.org/10.1006/jmbi.1994.1333.

(52) Slightom, J. L.; Drong, R. F.; Klassy, R. C.; Hoffman, L. M. Nucleotide Sequences from Phaseolin cDNA Clones: The Major Storage Proteins from Phaseolus Vulgaris Are Encoded by Two Unique Gene Families. *Nucleic Acids Res.* **1985**, *13* (18), 6483–6498. https://doi.org/10.1093/nar/13.18.6483.

(53) Thanh, V. H.; Shibasaki, K. Major Proteins of Soybean Seeds. A Straightforward Fractionation and Their Characterization. *J. Agric. Food Chem.* **1976**, *24* (6), 1117–1121. https://doi.org/10.1021/jf60208a030.

(54) Gatehouse, J. A.; Croy, R. R. D.; Boulter, D.; Shewry, P. R. The Synthesis and Structure of Pea Storage Proteins. *Crit. Rev. Plant Sci.* **1984**. https://doi.org/10.1080/07352688409382182.

(55) Gibbs, P. E.; Strongin, K. B.; McPherson, A. Evolution of Legume Seed Storage Proteins--a Domain Common to Legumins and Vicilins Is Duplicated in Vicilins. *Mol. Biol. Evol.* **1989**, *6* (6), 614–623. https://doi.org/10.1093/oxfordjournals.molbev.a040575.

(56) Shutov, A. D.; Kakhovskaya, I. A.; Braun, H.; Bäumlein, H.; Müntz, K. Legumin-like and Vicilin-like Seed Storage Proteins: Evidence for a Common Single-Domain Ancestral Gene. *J. Mol. Evol.* **1995**, *41* (6), 1057–1069. https://doi.org/10.1007/BF00173187.

(57) Rudakova, A. S.; Cherdivară, A. M.; Wilson, K. A.; Shutov, A. D. Seed Storage Globulins: Origin and Evolution of Primary and Higher Order Structures. *Biochem. Mosc.* **2015**, *80* (10), 1354–1361. https://doi.org/10.1134/S000629791510017X.

(58) Duranti, M.; Gius, C. Legume Seeds: Protein Content and Nutritional Value. *Field Crops Res.* **1997**, *53* (1–3), 31–45. https://doi.org/10.1016/S0378-4290(97)00021-X.

(59) O'Kane, F. E.; Happe, R. P.; Vereijken, J. M.; Gruppen, H.; van Boekel, M. A. J. S. Heat-Induced Gelation of Pea Legumin: Comparison with Soybean Glycinin. *J. Agric. Food Chem.* **2004**, *52* (16), 5071–5078. https://doi.org/10.1021/jf035215h.

(60) Sikorski, Z. E. Chemical and Functional Properties of Food Proteins.

(61) Marcone, M. F. Biochemical and Biophysical Properties of Plant Storage Proteins. *Food Res. Int.* **1999**, *32* (2), 79–92. https://doi.org/10.1016/S0963-9969(99)00061-7.

(62) Shewry, P. R. Plant Storage Proteins. *Biol. Rev.* **1995**, *70* (3), 375–426. https://doi.org/10.1111/j.1469-185X.1995.tb01195.x.

(63) Dalbey, R. E.; Von Heijne, G. Signal Peptidases in Prokaryotes and Eukaryotes--a New Protease Family. *Trends Biochem. Sci.* **1992**, *17* (11), 474–478. https://doi.org/10.1016/0968-0004(92)90492-r.

(64) Shutov, A. D.; Do, N. L.; Vaintraub, I. A. [Purification and partial characterization of protease B from germinating vetch seeds]. *Biokhimiia Mosc. Russ.* **1982**, *47* (5), 814–821.

(65) Fischer, J.; Becker, C.; Hillmer, S.; Horstmann, C.; Neubohn, B.; Schlereth, A.; Senyuk, V.; Shutov, A.; Müntz, K. The Families of Papain- and Legumain-like Cysteine Proteinases from Embryonic Axes and Cotyledons of Vicia Seeds: Developmental Patterns, Intracellular Localization and Functions in Globulin Proteolysis. *Plant Mol. Biol.* **2000**, *43* (1), 83–101. https://doi.org/10.1023/a:1006456615373.

(66) Hara-Nishimura, I.; Inoue, K.; Nishimura, M. A Unique Vacuolar Processing Enzyme Responsible for Conversion of Several Proprotein Precursors into the Mature Forms. *FEBS Lett.* **1991**, *294* (1–2), 89–93. https://doi.org/10.1016/0014-5793(91)81349-d.

(67) Scott, M. P.; Jung, R.; Muntz, K.; Nielsen, N. C. A Protease Responsible for Post-Translational Cleavage of a Conserved Asn-Gly Linkage in Glycinin, the Major Seed Storage Protein of Soybean. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (2), 658–662. https://doi.org/10.1073/pnas.89.2.658.

(68) Hiraiwa, N.; Nishimura, M.; Hara-Nishimura, I. Vacuolar Processing Enzyme Is Self-Catalytically Activated by Sequential Removal of the C-Terminal and N-Terminal Propeptides. *FEBS Lett.* **1999**, *447* (2–3), 213–216. https://doi.org/10.1016/s0014-5793(99)00286-0.

(69) Kuroyanagi, M.; Nishimura, M.; Hara-Nishimura, I. Activation of Arabidopsis Vacuolar Processing Enzyme by Self-Catalytic Removal of an Auto-Inhibitory Domain of the C-Terminal Propeptide. *Plant Cell Physiol.* **2002**, *43* (2), 143–151. https://doi.org/10.1093/pcp/pcf035.

(70) Lev-Yadun, S.; Gopher, A.; Abbo, S. Archaeology. The Cradle of Agriculture. *Science* **2000**, *288* (5471), 1602–1603. https://doi.org/10.1126/science.288.5471.1602.

(71) Moreno, M.-T.; Cubero, J. I. Variation in Cicer Arietinum L. *Euphytica* **1978**, *27* (2), 465–485. https://doi.org/10.1007/BF00043173.

(72) Sellami, M. H.; Lavini, A.; Pulvento, C. Phenotypic and Quality Traits of Chickpea Genotypes under Rainfed Conditions in South Italy. *Agronomy* **2021**, *11* (5), 962. https://doi.org/10.3390/agronomy11050962.

(73) Frimpong, A.; Sinha, A.; Tar'an, B.; Warkentin, T. D.; Gossen, B. D.; Chibbar, R. N. Genotype and Growing Environment Influence Chickpea (Cicer Arietinum L.) Seed Composition. *J. Sci. Food Agric.* **2009**, *89* (12), 2052–2063. https://doi.org/10.1002/jsfa.3690.

(74) Pande, S.; Siddique, K. H. M.; Kishore, G. K.; Bayaa, B.; Gaur, P. M.; Gowda, C. L. L.; Bretag, T. W.; Crouch, J. H. Ascochyta Blight of Chickpea (Cicer Arietinum L.): A Review of Biology, Pathogenicity, and Disease Management*. *Aust. J. Agric. Res.* **2005**, *56* (4), 317–332. https://doi.org/10.1071/AR04143.

(75) Chibbar, R. N.; Ambigaipalan, P.; Hoover, R. REVIEW: Molecular Diversity in Pulse Seed Starch and Complex Carbohydrates and Its Role in Human Nutrition and Health. *Cereal Chem.* **2010**, *87* (4), 342–352. https://doi.org/10.1094/CCHEM-87-4-0342.

(76) Wood, J. A.; Grusak, M. A. Nutritional Value of Chickpea. In *Chickpea breeding and management*; CABI Books; 2007; pp 101–142. https://doi.org/10.1079/9781845932138.005.

(77) Chibbar, R. N.; Ganeshan, S.; Båga, M.; Khandelwal, R. L. CARBOHYDRATE METABOLISM. In *Encyclopedia of Grain Science*; Wrigley, C., Ed.; Elsevier: Oxford, 2004; pp 168–179. https://doi.org/10.1016/B0-12-765490-9/00029-X.

(78) Jukanti, A. K.; Gaur, P. M.; Gowda, C. L. L.; Chibbar, R. N. Nutritional Quality and Health Benefits of Chickpea (Cicer Arietinum L.): A Review. *Br. J. Nutr.* **2012**, *108 Suppl 1*, S11-26. https://doi.org/10.1017/S0007114512000797.

(79) Stribling, P.; Ibrahim, F. Dietary Fibre Definition Revisited - The Case of Low Molecular Weight Carbohydrates. *Clin. Nutr. ESPEN* **2023**, *55*, 340–356. https://doi.org/10.1016/j.clnesp.2023.04.014.

(80) Tosh, S. M.; Yada, S. Dietary Fibres in Pulse Seeds and Fractions: Characterization, Functional Attributes, and Applications. *Food Res. Int.* **2010**, *43* (2), 450–460. https://doi.org/10.1016/j.foodres.2009.09.005.

(81) Aguilera, Y.; Martín-Cabrejas, M. A.; Benítez, V.; Mollá, E.; López-Andréu, F. J.; Esteban, R. M. Changes in Carbohydrate Fraction during Dehydration Process of Common Legumes. *J. Food Compos. Anal.* **2009**, *22* (7), 678–683. https://doi.org/10.1016/j.jfca.2009.02.012.

(82) Xiao, S.; Li, Z.; Zhou, K.; Fu, Y. Chemical Composition of Kabuli and Desi Chickpea (Cicer Arietinum L.) Cultivars Grown in Xinjiang, China. *Food Sci. Nutr.* **2023**, *11* (1), 236–248. https://doi.org/10.1002/fsn3.3056.

(83) Kaur, M.; Singh, N.; Sodhi, N. S. Physicochemical, Cooking, Textural and Roasting Characteristics of Chickpea (*Cicer Arietinum* L.) Cultivars. *J. Food Eng.* **2005**, *69* (4), 511–517. https://doi.org/10.1016/j.jfoodeng.2004.09.002.

(84) Alajaji, S. A.; El-Adawy, T. A. Nutritional Composition of Chickpea (*Cicer Arietinum* L.) as Affected by Microwave Cooking and Other Traditional Cooking Methods. *J. Food Compos. Anal.* **2006**, *19* (8), 806–812. https://doi.org/10.1016/j.jfca.2006.03.015.

(85) Ibáñez, M. V.; Rincón, F.; Amaro, M.; Martínez, B. Intrinsic Variability of Mineral Composition of Chickpea (*Cicer Arietinum*, L.). *Food Chem.* **1998**, *63* (1), 55–60. https://doi.org/10.1016/S0308-8146(97)00221-5.

(86) Quinteros, A.; Farré, R.; Lagarda, M. J. Optimization of Iron Speciation (Soluble, Ferrous and Ferric) in Beans, Chickpeas and Lentils. *Food Chem.* **2001**, *75* (3), 365–370. https://doi.org/10.1016/S0308-8146(01)00214-X.

(87) Ciftci, H.; Ozkaya, A.; Cevrimli, B. S.; Bakoglu, A. Levels of Fat-Soluble Vitamins in Some Foods.

(88) Lebiedzińska, A.; Szefer, P. Vitamins B in Grain and Cereal–Grain Food, Soy-Products and Seeds. *Food Chem.* **2006**, *95* (1), 116–122. https://doi.org/10.1016/j.foodchem.2004.12.024.

(89) Yust, M. M.; Pedroche, J.; Girón-Calle, J.; Alaiz, M.; Millán, F.; Vioque, J. Production of Ace Inhibitory Peptides by Digestion of Chickpea Legumin with Alcalase. *Food Chem.* **2003**, *81* (3), 363–369. https://doi.org/10.1016/S0308-8146(02)00431-4.

(90) Sánchez-Vioque, R.; Clemente, A.; Vioque, J.; Bautista, J.; Millán, F. Protein Isolates from Chickpea (*Cicer Arietinum* L.): Chemical Composition, Functional Properties and Protein Characterization. *Food Chem.* **1999**, *64* (2), 237–243. https://doi.org/10.1016/S0308-8146(98)00133-2.

(91) Bibi, N.; Khattak, A. B.; Khattak, G. S. S.; Mehmood, Z.; Ihsanullah, I. Quality and Consumers Acceptability Studies and Their Inter-Relationship of Newly Evolved Desi Type Chickpea Genotypes (Cicer Arietinum L.). Quality Evolution of New Chickpea Genotypes. *Int. J. Food Sci. Technol.* **2007**, *42* (5), 528–534. https://doi.org/10.1111/j.1365-2621.2006.01246.x.

(92) Singh, U.; Jambunathan, R. Studies on Desi and Kabull Chickpea (Cicer Arietinum L.) Cultivars: Levels of Protease Inhibitors, Levels of Polyphenolic Compounds and in Vitro Protein Digestibility. *J. Food Sci.* **1981**, *46* (5), 1364–1367. https://doi.org/10.1111/j.1365-2621.1981.tb04176.x.

(93) Rincón, F.; Martínez, B.; Ibáñez, M. V. Proximate Composition and Antinutritive Substances in Chickpea (Cicer Arietinum L) as Affected by the Biotype Factor. *J. Sci. Food Agric.* **1998**, *78* (3), 382–388. https://doi.org/10.1002/(SICI)1097-0010(199811)78:3%253C382::AID-JSFA128%253E3.0.CO;2-J.

(94) Kaur, M.; Singh, N.; Sodhi, N. S. Physicochemical, Cooking, Textural and Roasting Characteristics of Chickpea (*Cicer Arietinum* L.) Cultivars. *J. Food Eng.* **2005**, *69* (4), 511–517. https://doi.org/10.1016/j.jfoodeng.2004.09.002.

(95) Rehman, Z.; Shah, W. H. Thermal Heat Processing Effects on Antinutrients, Protein and Starch Digestibility of Food Legumes. *Food Chem.* **2005**, *91* (2), 327–331. https://doi.org/10.1016/j.foodchem.2004.06.019.

(96) Iqbal, A.; Khalil, I. A.; Ateeq, N.; Sayyar Khan, M. Nutritional Quality of Important Food Legumes. *Food Chem.* **2006**, *97* (2), 331–335. https://doi.org/10.1016/j.foodchem.2005.05.011.

(97) Wang, X.; Gao, W.; Zhang, J.; Zhang, H.; Li, J.; He, X.; Ma, H. Subunit, Amino Acid Composition and *in Vitro* Digestibility of Protein Isolates from Chinese Kabuli and Desi Chickpea (*Cicer Arietinum* L.) Cultivars. *Food Res. Int.* **2010**, *43* (2), 567–572. https://doi.org/10.1016/j.foodres.2009.07.018.

(98) *Protein Chemistry of Dicotyledonous Grains*. ResearchGate. https://www.researchgate.net/publication/288914789_Protein_Chemistry_of_Dicotyledonous_Grains (accessed 2025-03-04).

(99) Kumari, P.; Rastogi, A.; Yadav, S. Effects of Heat Stress and Molecular Mitigation Approaches in Orphan Legume, Chickpea. *Mol. Biol. Rep.* **2020**, *47* (6), 4659–4670. https://doi.org/10.1007/s11033-020-05358-x.

(100) Kumar, K. G.; Venkataraman, L. V. Chickpea Seed Proteins: Isolation and Characterization of 10.3S Protein. *J. Agric. Food Chem.* **1980**, *28* (3), 524–529. https://doi.org/10.1021/jf60229a018.

(101) *The seed proteins of chick pea: Comparative studies ofCicer arietinum, C. reticulatum andC. echinospermum (Leguminosae) | Plant Systematics and Evolution*. https://link.springer.com/article/10.1007/BF00989600 (accessed 2025-07-07).

(102) Dhawan, K.; Malhotra, S.; Dahiya, B. S.; Singh, D. Seed Protein Fractions and Amino Acid Composition in Gram (Cicer Arietinum). *Plant Foods Hum. Nutr. Dordr. Neth.* **1991**, *41* (3), 225–232. https://doi.org/10.1007/BF02196390.

(103) Liu, L. H.; Hung, T. V.; Bennett, L. Extraction and Characterization of Chickpea (Cicer Arietinum) Albumin and Globulin. *J. Food Sci.* **2008**, *73* (5), C299-305. https://doi.org/10.1111/j.1750-3841.2008.00773.x.

(104) Papalamprou, E. M.; Doxastakis, G. I.; Biliaderis, C. G.; Kiosseoglou, V. Influence of Preparation Methods on Physicochemical and Gelation Properties of Chickpea Protein Isolates. *Food Hydrocoll.* **2009**, *23* (2), 337–343. https://doi.org/10.1016/j.foodhyd.2008.03.006.

(105) Chang, Y.-W.; Alli, I.; Konishi, Y.; Ziomek, E. Characterization of Protein Fractions from Chickpea (*Cicer Arietinum* L.) and Oat (*Avena Sativa* L.) Seeds Using Proteomic Techniques. *Food Res. Int.* **2011**, *44* (9), 3094–3104. https://doi.org/10.1016/j.foodres.2011.08.001.

(106) Varshney, R. K.; Song, C.; Saxena, R. K.; Azam, S.; Yu, S.; Sharpe, A. G.; Cannon, S.; Baek, J.; Rosen, B. D.; Tar'an, B.; Millan, T.; Zhang, X.; Ramsay, L. D.; Iwata, A.; Wang, Y.; Nelson, W.; Farmer, A. D.; Gaur, P. M.; Soderlund, C.; Penmetsa, R. V.; Xu, C.; Bharti, A. K.; He, W.; Winter, P.; Zhao, S.; Hane, J. K.; Carrasquilla-Garcia, N.; Condie, J. A.; Upadhyaya, H. D.; Luo, M.-C.; Thudi, M.; Gowda, C. L. L.; Singh, N. P.; Lichtenzveig, J.; Gali, K. K.; Rubio, J.; Nadarajan, N.; Dolezel, J.; Bansal, K. C.; Xu, X.; Edwards, D.; Zhang, G.; Kahl, G.; Gil, J.; Singh, K. B.; Datta, S. K.; Jackson, S. A.; Wang, J.; Cook, D. R. Draft Genome Sequence of Chickpea (Cicer Arietinum) Provides a Resource for Trait Improvement. *Nat. Biotechnol.* **2013**, *31* (3), 240–246. https://doi.org/10.1038/nbt.2491.

(107) Singh, P. K.; Shrivastava, N.; Chaturvedi, K.; Sharma, B.; Bhagyawant, S. S. Characterization of Seed Storage Proteins from Chickpea Using 2D Electrophoresis Coupled with Mass Spectrometry. *Biochem. Res. Int.* **2016**, *2016*, 1049462. https://doi.org/10.1155/2016/1049462.

(108) Ribeiro, I. C.; Leclercq, C. C.; Simões, N.; Toureiro, A.; Duarte, I.; Freire, J. B.; Chaves, M. M.; Renaut, J.; Pinheiro, C. Identification of Chickpea Seed Proteins Resistant to Simulated in Vitro Human Digestion. *J. Proteomics* **2017**, *169*, 143–152. https://doi.org/10.1016/j.jprot.2017.06.009.

(109) Bar-El Dadon, S.; Pascual, C. Y.; Eshel, D.; Teper-Bamnolker, P.; Ibáñez, M. D. P.; Reifen, R. Vicilin and the Basic Subunit of Legumin Are Putative Chickpea Allergens. *Food Chem.* **2013**, *138* (1), 13–18. https://doi.org/10.1016/j.foodchem.2012.10.031.

(110) Chang, L.; Lan, Y.; Bandillo, N.; Ohm, J.-B.; Chen, B.; Rao, J. Plant Proteins from Green Pea and Chickpea: Extraction, Fractionation, Structural Characterization and Functional Properties. *Food Hydrocoll.* **2022**, *123*, 107165. https://doi.org/10.1016/j.foodhyd.2021.107165.

(111) Bose, U.; Buck, S.; Sirault, X.; Bahmani, M.; Byrne, K.; Stockwell, S.; McWilliam, S.; Colgrave, M.; Juhász, A.; Ral, J.-P. Chickpea Proteome Analysis Reveals Genotype-Dependent Variations Associated with Seed Traits. *J. Agric. Food Chem.* **2024**, *72* (48), 27030–27042. https://doi.org/10.1021/acs.jafc.4c07669.

(112) Thomson, J. J. On the Appearance of Helium and Neon in Vacuum Tubes. *Nature* **1913**, *90* (2259), 645–647. https://doi.org/10.1038/090645d0.

(113) *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons | Analytical Chemistry*. https://pubs.acs.org/doi/10.1021/ac00171a028 (accessed 2025-05-03).

(114) Tanaka, K. The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). *Angew. Chem. Int. Ed.* **2003**, *42* (33), 3860–3870. https://doi.org/10.1002/anie.200300585.

(115) Yamashita, M.; Fenn, J. B. Electrospray Ion Source. Another Variation on the Free-Jet Theme. *J. Phys. Chem.* **1984**, *88* (20), 4451–4459. https://doi.org/10.1021/j150664a002.

(116) Nickerson, J. L.; Baghalabadi, V.; Rajendran, S. R. C. K.; Jakubec, P. J.; Said, H.; McMillen, T. S.; Dang, Z.; Doucette, A. A. Recent Advances in Top-down Proteome Sample Processing Ahead of MS Analysis. *Mass Spectrom. Rev.* **2023**, *42* (2), 457–495. https://doi.org/10.1002/mas.21706.

(117) *α-Cyano-4-hydroxycinnamic acid as a matrix for matrixassisted laser desorption mass spectromtry - Beavis - 1992 - Organic Mass Spectrometry - Wiley Online Library*. https://onlinelibrary.wiley.com/doi/10.1002/oms.1210270217 (accessed 2025-05-15).

(118) Strupat, K.; Karas, M.; Hillenkamp, F. 2,5-Dihydroxybenzoic Acid: A New Matrix for Laser Desorption—Ionization Mass Spectrometry. *Int. J. Mass Spectrom. Ion Process.* **1991**, *111*, 89–102. https://doi.org/10.1016/0168-1176(91)85050-V.

(119) Beavis, R. C.; Chait, B. T. Cinnamic Acid Derivatives as Matrices for Ultraviolet Laser Desorption Mass Spectrometry of Proteins. *Rapid Commun. Mass Spectrom. RCM* **1989**, *3* (12), 432–435. https://doi.org/10.1002/rcm.1290031207.

(120) Wilm, M. Principles of Electrospray Ionization. *Mol. Cell. Proteomics* **2011**, *10* (7), M111.009407. https://doi.org/10.1074/mcp.M111.009407.

(121) Wilm, M.; Mann, M. Analytical Properties of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, *68* (1), 1–8. https://doi.org/10.1021/ac9509519.

(122) Wilm, M. S.; Mann, M. Electrospray and Taylor-Cone Theory, Dole's Beam of Macromolecules at Last? *Int. J. Mass Spectrom. Ion Process.* **1994**, *136* (2), 167–180. https://doi.org/10.1016/0168-1176(94)04024-9.

(123) Sprenger, R. R.; Roepstorff, P. Mass Spectrometry Instrumentation in Proteomics. In *eLS*; John Wiley & Sons, Ltd, 2012. https://doi.org/10.1002/9780470015902.a0006194.pub2.

(124) *The quadrupole mass filter: Basic operating concepts | Journal of Chemical Education*. https://pubs.acs.org/doi/10.1021/ed063p617 (accessed 2025-05-15).

(125) March, R. E. An Introduction to Quadrupole Ion Trap Mass Spectrometry. *J. Mass Spectrom.* **1997**, *32* (4), 351–369. https://doi.org/10.1002/(SICI)1096-9888(199704)32:4%253C351::AID-JMS512%253E3.0.CO;2-Y.

(126) Douglas, D. J.; Frank, A. J.; Mao, D. Linear Ion Traps in Mass Spectrometry. *Mass Spectrom. Rev.* **2005**, *24* (1), 1–29. https://doi.org/10.1002/mas.20004.

(127) Boesl, U. Time-of-Flight Mass Spectrometry: Introduction to the Basics. *Mass Spectrom. Rev.* **2017**, *36* (1), 86–109. https://doi.org/10.1002/mas.21520.

(128) Mamyrin, B. A. Time-of-Flight Mass Spectrometry (Concepts, Achievements, and Prospects). *Int. J. Mass Spectrom.* **2001**, *206* (3), 251–266. https://doi.org/10.1016/S1387-3806(00)00392-4.

(129) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. The Orbitrap: A New Mass Spectrometer. *J. Mass Spectrom. JMS* **2005**, *40* (4), 430–443. https://doi.org/10.1002/jms.856.

(130) Cunsolo, V.; Muccilli, V.; Saletti, R.; Foti, S. Mass Spectrometry in Food Proteomics: A Tutorial. *J. Mass Spectrom. JMS* **2014**, *49* (9), 768–784. https://doi.org/10.1002/jms.3374.

(131) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984**. https://doi.org/10.1002/bms.1200111109.

(132) Wiesner, J.; Premsler, T.; Sickmann, A. Application of Electron Transfer Dissociation (ETD) for the Analysis of Posttranslational Modifications. *Proteomics* **2008**, *8* (21), 4466–4483. https://doi.org/10.1002/pmic.200800329.

(133) Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H. W. P.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89* (6), 3318–3325. https://doi.org/10.1021/acs.analchem.6b03756.

(134) Yates III, J. R. Mass Spectrometry and the Age of the Proteome. *J. Mass Spectrom.* **1998**, *33* (1), 1–19. https://doi.org/10.1002/(SICI)1096-9888(199801)33:1%253C1::AID-JMS624%253E3.0.CO;2-9.

(135) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics. *Nat. Protoc.* **2016**, *11* (12), 2301–2319. https://doi.org/10.1038/nprot.2016.136.

(136) Larson, E. J.; Pergande, M. R.; Moss, M. E.; Rossler, K. J.; Wenger, R. K.; Krichel, B.; Josyer, H.; Melby, J. A.; Roberts, D. S.; Pike, K.; Shi, Z.; Chan, H.-J.; Knight, B.; Rogers, H. T.; Brown, K. A.; Ong, I. M.; Jeong, K.; Marty, M. T.; McIlwain, S. J.; Ge, Y. MASH Native: A Unified Solution for Native Top-down Proteomics Data Processing. *Bioinformatics* **2023**, *39* (6), btad359. https://doi.org/10.1093/bioinformatics/btad359.

(137) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021**, *20* (4), 1928–1935. https://doi.org/10.1021/acs.jproteome.0c00952.

(138) *Water Use Efficiency, Spectral Phenotyping and Protein Composition of Two Chickpea Genotypes Grown in Mediterranean Environments under Different Water and Nitrogen Supply*. https://www.mdpi.com/2077-0472/12/12/2026 (accessed 2025-06-02).

(139) Cucina, A.; Di Francesco, A.; Saletti, R.; Pittalà, M. G. G.; Zilberstein, G.; Zilberstein, S.; Tikhonov, A.; Bublichenko, A. G.; Righetti, P. G.; Foti, S.; Cunsolo, V. Meta-Proteomic Analysis of Two Mammoth's Trunks by EVA Technology and High-Resolution Mass Spectrometry for an Indirect Picture of Their Habitat and the Characterization of the Collagen Type I, Alpha-1 and Alpha-2 Sequence. *Amino Acids* **2022**, *54* (6), 935–954. https://doi.org/10.1007/s00726-022-03160-6.

(140) Di Francesco, A.; Cunsolo, V.; Saletti, R.; Svensson, B.; Muccilli, V.; De Vita, P.; Foti, S. Quantitative Label-Free Comparison of the Metabolic Protein Fraction in Old and Modern Italian Wheat Genotypes by a Shotgun Approach. *Mol. Basel Switz.* **2021**, *26* (9), 2596. https://doi.org/10.3390/molecules26092596.

(141) Punginelli, D.; Catania, V.; Abruscato, G.; Luparello, C.; Vazzana, M.; Mauro, M.; Cunsolo, V.; Saletti, R.; Di Francesco, A.; Arizza, V.; Schillaci, D. New Bioactive Peptides from the Mediterranean Seagrass Posidonia Oceanica (L.) Delile and Their Impact on Antimicrobial Activity and Apoptosis of Human Cancer Cells. *Int. J. Mol. Sci.* **2023**, *24* (6), 5650. https://doi.org/10.3390/ijms24065650.

(142) Lin, H.; He, L.; Ma, B. A Combinatorial Approach to the Peptide Feature Matching Problem for Label-Free Quantification. *Bioinforma. Oxf. Engl.* **2013**, *29* (14), 1768–1775. https://doi.org/10.1093/bioinformatics/btt274.

(143) Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape Plugin to Assess Overrepresentation of Gene Ontology Categories in Biological Networks. *Bioinforma. Oxf. Engl.* **2005**, *21* (16), 3448–3449. https://doi.org/10.1093/bioinformatics/bti551.

(144) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. https://doi.org/10.1101/gr.1239303.

(145) Swaney, D. L.; Wenger, C. D.; Coon, J. J. Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010**, *9* (3), 1323–1329. https://doi.org/10.1021/pr900863u.

(146) Gupta, S.; Mishra, S. K.; Misra, S.; Pandey, V.; Agrawal, L.; Nautiyal, C. S.; Chauhan, P. S. Revealing the Complexity of Protein Abundance in Chickpea Root under Drought-Stress Using a Comparative Proteomics Approach. *Plant Physiol. Biochem. PPB* **2020**, *151*, 88–102. https://doi.org/10.1016/j.plaphy.2020.03.005.

(147) Vessal, S.; Siddique, K. H. M.; Atkins, C. A. Comparative Proteomic Analysis of Genotypic Variation in Germination and Early Seedling Growth of Chickpea under Suboptimal Soil-Water Conditions. *J. Proteome Res.* **2012**, *11* (8), 4289–4307. https://doi.org/10.1021/pr300415w.

(148) Arefian, M.; Vessal, S.; Malekzadeh-Shafaroudi, S.; Siddique, K. H. M.; Bagheri, A. Comparative Proteomics and Gene Expression Analyses Revealed Responsive Proteins and Mechanisms for Salt Tolerance in Chickpea Genotypes. *BMC Plant Biol.* **2019**, *19* (1), 300. https://doi.org/10.1186/s12870-019-1793-z.

(149) Sinha, A.; Haider, T.; Narula, K.; Ghosh, S.; Chakraborty, N.; Chakraborty, S. Integrated Seed Proteome and Phosphoproteome Analyses Reveal Interplay of Nutrient Dynamics, Carbon-Nitrogen Partitioning, and Oxidative Signaling in Chickpea. *Proteomics* **2020**, *20* (8), e1900267. https://doi.org/10.1002/pmic.201900267.

(150) Parankusam, S.; Bhatnagar-Mathur, P.; Sharma, K. K. Heat Responsive Proteome Changes Reveal Molecular Mechanisms Underlying Heat Tolerance in Chickpea. *Environ. Exp. Bot.* **2017**, *141*, 132–144. https://doi.org/10.1016/j.envexpbot.2017.07.007.

(151) Quiroz-Figueroa, F. R.; Monribot-Villanueva, J. L.; Bojórquez-Velázquez, E.; Gómez-Peraza, R. L.; Elizalde-Contreras, J. M.; Bautista-Valle, M. V.; Guerrero-Analco, J. A.; Valdez-Morales, M.; Singh, R. K.; Ruiz-May, E. Proteometabolomic Analysis Reveals Molecular Features Associated with Grain Size and Antioxidant Properties amongst Chickpea (Cicer Arietinum L.) Seeds Genotypes. *Antioxid. Basel Switz.* **2022**, *11* (10), 1850. https://doi.org/10.3390/antiox11101850.

(152) Nitride, C.; Vegarud, G. E.; Comi, I.; Devold, T. G.; Røseth, A.; Marti, A.; Iametti, S.; Mamone, G.; Picariello, G.; Alfieri, F.; Adalgisa Nicolai, M.; Mills, C.; Ferranti, P. Effect of Sprouting on the Proteome of Chickpea Flour and on Its Digestibility by Ex Vivo Gastro-Duodenal Digestion Complemented with Jejunal Brush Border Membrane Enzymes. *Food Res. Int. Ott. Ont* **2022**, *154*, 111012. https://doi.org/10.1016/j.foodres.2022.111012.

(153) Di Francesco, A.; De Santis, M. A.; Lanzoni, A.; Pittalà, M. G. G.; Saletti, R.; Flagella, Z.; Cunsolo, V. Mass Spectrometry Characterization of the SDS-PAGE Protein Profile of Legumins and Vicilins from Chickpea Seed. *Foods* **2024**, *13* (6), 887. https://doi.org/10.3390/foods13060887.

(154) Verheggen, K.; Raeder, H.; Berven, F. S.; Martens, L.; Barsnes, H.; Vaudel, M. Anatomy and Evolution of Database Search Engines-a Central Component of Mass Spectrometry Based Proteomic Workflows. *Mass Spectrom. Rev.* **2020**, *39* (3), 292–306. https://doi.org/10.1002/mas.21543.

(155) Urquidi Camacho, R. A.; Lokdarshi, A.; von Arnim, A. G. Translational Gene Regulation in Plants: A Green New Deal. *Wiley Interdiscip. Rev. RNA* **2020**, *11* (6), e1597. https://doi.org/10.1002/wrna.1597.

(156) Lee, J.-H.; Muhsin, M.; Atienza, G. A.; Kwak, D.-Y.; Kim, S.-M.; De Leon, T. B.; Angeles, E. R.; Coloquio, E.; Kondoh, H.; Satoh, K.; Cabunagan, R. C.; Cabauatan, P. Q.; Kikuchi, S.; Leung, H.; Choi, I.-R. Single Nucleotide Polymorphisms in a Gene for Translation Initiation Factor (eIF4G) of Rice (Oryza Sativa) Associated with Resistance to Rice Tungro Spherical Virus. *Mol. Plant-Microbe Interact. MPMI* **2010**, *23* (1), 29–38. https://doi.org/10.1094/MPMI-23-1-0029.

(157) Verbruggen, N.; Hermans, C. Proline Accumulation in Plants: A Review. *Amino Acids* **2008**, *35* (4), 753–759. https://doi.org/10.1007/s00726-008-0061-6.

(158) Sinha, S. C.; Smith, J. L. The PRT Protein Family. *Curr. Opin. Struct. Biol.* **2001**, *11* (6), 733–739. https://doi.org/10.1016/s0959-440x(01)00274-3.

(159) Liu, Y.; Wu, P.; Li, B.; Wang, W.; Zhu, B. Phosphoribosyltransferases and Their Roles in Plant Development and Abiotic Stress Response. *Int. J. Mol. Sci.* **2023**, *24* (14), 11828. https://doi.org/10.3390/ijms241411828.

(160) Qiu, C.-W.; Ma, Y.; Wang, Q.-Q.; Fu, M.-M.; Li, C.; Wang, Y.; Wu, F. Barley HOMOCYSTEINE METHYLTRANSFERASE 2 Confers Drought Tolerance by Improving Polyamine Metabolism. *Plant Physiol.* **2023**, *193* (1), 389–409. https://doi.org/10.1093/plphys/kiad333.

(161) Danilevskaya, O. N.; Yu, G.; Meng, X.; Xu, J.; Stephenson, E.; Estrada, S.; Chilakamarri, S.; Zastrow-Hayes, G.; Thatcher, S. Developmental and Transcriptional Responses of Maize to Drought Stress under Field Conditions. *Plant Direct* **2019**, *3* (5), e00129. https://doi.org/10.1002/pld3.129.

(162) Tavano, O. L.; Neves, V. A. Isolation, Solubility and *in Vitro* Hydrolysis of Chickpea Vicilin-like Protein. *LWT - Food Sci. Technol.* **2008**, *41* (7), 1244–1251. https://doi.org/10.1016/j.lwt.2007.08.003.

(163) Mills, E. N.; Jenkins, J.; Marigheto, N.; Belton, P. S.; Gunning, A. P.; Morris, V. J. Allergens of the Cupin Superfamily. *Biochem. Soc. Trans.* **2002**, *30* (Pt 6), 925–929. https://doi.org/10.1042/bst0300925.

(164) Vioque, J.; Sánchez-Vioque, R.; Clemente, A.; Pedroche, J.; Bautista, J.; Millán, F. Purification and Partial Characterization of Chickpea 2S Albumin. *J. Agric. Food Chem.* **1999**, *47* (4), 1405–1409. https://doi.org/10.1021/jf980819k.

(165) Clemente, A.; Vioque, J.; Sánchez-Vioque, R.; Pedroche, J.; Bautista, J.; Millán, F. Factors Affecting the in Vitro Protein Digestibility of Chickpea Albumins. *J. Sci. Food Agric.* **2000**, *80* (1), 79–84. https://doi.org/10.1002/(SICI)1097-0010(20000101)80:1%253C79::AID-JSFA487%253E3.0.CO;2-4.

(166) Wangorsch, A.; Kulkarni, A.; Jamin, A.; Spiric, J.; Bräcker, J.; Brockmeyer, J.; Mahler, V.; Blanca-López, N.; Ferrer, M.; Blanca, M.; Torres, M.; Gomez, P.; Bartra, J.; García-Moral, A.; Goikoetxea, M. J.; Vieths, S.; Toda, M.; Zoccatelli, G.; Scheurer, S. Identification and Characterization of IgE-Reactive Proteins and a New Allergen (Cic a 1.01) from Chickpea (Cicer Arietinum). *Mol. Nutr. Food Res.* **2020**, *64* (19), e2000560. https://doi.org/10.1002/mnfr.202000560.

(167) Verma, A. K.; Sharma, A.; Kumar, S.; Gupta, R. K.; Kumar, D.; Gupta, K.; Giridhar, B. H.; Das, M.; Dwivedi, P. D. Purification, Characterization and Allergenicity Assessment of 26kDa Protein, a Major Allergen from Cicer Arietinum. *Mol. Immunol.* **2016**, *74*, 113–124. https://doi.org/10.1016/j.molimm.2016.05.002.

(168) Savchenko, T. V.; Zastrijnaja, O. M.; Klimov, V. V. Oxylipins and Plant Abiotic Stress Resistance. *Biochem. Biokhimiia* **2014**, *79* (4), 362–375. https://doi.org/10.1134/S0006297914040051.

(169) Chang, Y.-W.; Alli, I.; Molina, A. T.; Konishi, Y.; Boye, J. I. Isolation and Characterization of Chickpea (Cicer Arietinum L.) Seed Protein Fractions. *Food Bioprocess Technol.* **2012**, *5* (2), 618–625. https://doi.org/10.1007/s11947-009-0303-y.

(170) Chang, Y.-W.; Alli, I.; Konishi, Y.; Ziomek, E. Characterization of Protein Fractions from Chickpea (*Cicer Arietinum* L.) and Oat (*Avena Sativa* L.) Seeds Using Proteomic Techniques. *Food Res. Int.* **2011**, *44* (9), 3094–3104. https://doi.org/10.1016/j.foodres.2011.08.001.

(171) Nikolić, Z.; Đorđević, V.; Torbica, A.; Mikić, A. Legumes Seed Storage Proteins Characterization by SDS-PAGE and Lab-on-a-Chip Electrophoresis. *J. Food Compos. Anal.* **2012**, *28* (2), 75–80. https://doi.org/10.1016/j.jfca.2012.08.005.

(172) *Pulse Globulins 11S and 7S: Origins, Purification Methods, and Techno-functional Properties | Journal of Agricultural and Food Chemistry*. https://pubs.acs.org/doi/full/10.1021/acs.jafc.2c07507 (accessed 2025-05-19).

(173) De Santis, M. A.; Satriani, A.; De Santis, F.; Flagella, Z. Water Use Efficiency, Spectral Phenotyping and Protein Composition of Two Chickpea Genotypes Grown in Mediterranean Environments under Different Water and Nitrogen Supply. *Agriculture* **2022**, *12* (12), 2026. https://doi.org/10.3390/agriculture12122026.

(174) De Santis, M. A.; Rinaldi, M.; Menga, V.; Codianni, P.; Giuzio, L.; Fares, C.; Flagella, Z. Influence of Organic and Conventional Farming on Grain Yield and Protein Composition of Chickpea Genotypes. *Agronomy* **2021**, *11* (2), 191. https://doi.org/10.3390/agronomy11020191.

(175) Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. Mass Spectrometric Sequencing of Proteins Silver-Stained Polyacrylamide Gels. *Anal. Chem.* **1996**, *68* (5), 850–858. https://doi.org/10.1021/ac950914h.

(176) Cunsolo, V.; Fasoli, E.; Di Francesco, A.; Saletti, R.; Muccilli, V.; Gallina, S.; Righetti, P. G.; Foti, S. Polyphemus, Odysseus and the Ovine Milk Proteome. *J. Proteomics* **2017**, *152*, 58–74. https://doi.org/10.1016/j.jprot.2016.10.007.

(177) Saletti, R.; Reina, S.; Pittalà, M. G. G.; Belfiore, R.; Cunsolo, V.; Messina, A.; De Pinto, V.; Foti, S. High Resolution Mass Spectrometry Characterization of the Oxidation Pattern of Methionine and Cysteine

Residues in Rat Liver Mitochondria Voltage-Dependent Anion Selective Channel 3 (VDAC3). *Biochim. Biophys. Acta Biomembr.* **2017**, *1859* (3), 301–311. https://doi.org/10.1016/j.bbamem.2016.12.003.

(178) Cucina, A.; Cunsolo, V.; Di Francesco, A.; Saletti, R.; Zilberstein, G.; Zilberstein, S.; Tikhonov, A.; Bublichenko, A. G.; Righetti, P. G.; Foti, S. Meta-Proteomic Analysis of the Shandrin Mammoth by EVA Technology and High-Resolution Mass Spectrometry: What Is Its Gut Microbiota Telling Us? *Amino Acids* **2021**, *53* (10), 1507–1521. https://doi.org/10.1007/s00726-021-03061-0.

(179) Grasso, N.; Lynch, N. L.; Arendt, E. K.; O'Mahony, J. A. Chickpea Protein Ingredients: A Review of Composition, Functionality, and Applications. *Compr. Rev. Food Sci. Food Saf.* **2022**, *21* (1), 435–452. https://doi.org/10.1111/1541-4337.12878.

(180) Bamdad, F.; Dokhani, S.; Keramat, J. Functional Assessment and Subunit Constitution of Lentil (Lens Culinaris) Proteins during Germination. *Int J Agric Biol* **2009**, *11* (6).

(181) Rubio, L. A.; Pérez, A.; Ruiz, R.; Guzmán, M. Á.; Aranda-Olmedo, I.; Clemente, A. Characterization of Pea (Pisum Sativum) Seed Protein Fractions. *J. Sci. Food Agric.* **2014**, *94* (2), 280–287. https://doi.org/10.1002/jsfa.6250.

(182) Di Francesco, A.; Cunsolo, V.; Saletti, R.; Svensson, B.; Muccilli, V.; De Vita, P.; Foti, S. Quantitative Label-Free Comparison of the Metabolic Protein Fraction in Old and Modern Italian Wheat Genotypes by a Shotgun Approach. *Molecules* **2021**, *26* (9), 2596. https://doi.org/10.3390/molecules26092596.

(183) Kovalchuk, S. I.; Jensen, O. N.; Rogowska-Wrzesinska, A. FlashPack: Fast and Simple Preparation of Ultrahigh-Performance Capillary Columns for LC-MS. *Mol. Cell. Proteomics MCP* **2019**, *18* (2), 383–390. https://doi.org/10.1074/mcp.TIR118.000953.

(184) Magni, C.; Scarafoni, A.; Herndl, A.; Sessa, F.; Prinsi, B.; Espen, L.; Duranti, M. Combined 2D Electrophoretic Approaches for the Study of White Lupin Mature Seed Storage Proteome. *Phytochemistry* **2007**, *68* (7), 997–1007. https://doi.org/10.1016/j.phytochem.2007.01.003.

(185) Kratzer, U.; Frank, R.; Kalbacher, H.; Biehl, B.; Wöstemeyer, J.; Voigt, J. Subunit Structure of the Vicilin-like Globular Storage Protein of Cocoa Seeds and the Origin of Cocoa- and Chocolate-Specific Aroma Precursors. *Food Chem.* **2009**, *113* (4), 903–913. https://doi.org/10.1016/j.foodchem.2008.08.017.

(186) Jung, R.; Scott, M. P.; Nam, Y. W.; Beaman, T. W.; Bassüner, R.; Saalbach, I.; Müntz, K.; Nielsen, N. C. The Role of Proteolysis in the Processing and Assembly of 11S Seed Globulins. *Plant Cell* **1998**, *10* (3), 343–357. https://doi.org/10.1105/tpc.10.3.343.

(187) Couturier, J.; Chibani, K.; Jacquot, J.-P.; Rouhier, N. Cysteine-Based Redox Regulation and Signaling in Plants. *Front. Plant Sci.* **2013**, *4*, 105. https://doi.org/10.3389/fpls.2013.00105.

(188) Ulrich, K.; Jakob, U. The Role of Thiols in Antioxidant Systems. *Free Radic. Biol. Med.* **2019**, *140*, 14–27. https://doi.org/10.1016/j.freeradbiomed.2019.05.035.

(189) Trost, P.; Fermani, S.; Calvaresi, M.; Zaffagnini, M. Biochemical Basis of Sulphenomics: How Protein Sulphenic Acids May Be Stabilized by the Protein Microenvironment. *Plant Cell Environ.* **2017**, *40* (4), 483–490. https://doi.org/10.1111/pce.12791.

(190) Biteau, B.; Labarre, J.; Toledano, M. B. ATP-Dependent Reduction of Cysteine-Sulphinic Acid by S. Cerevisiae Sulphiredoxin. *Nature* **2003**, *425* (6961), 980–984. https://doi.org/10.1038/nature02075.

(191) Müntz, K.; Blattner, F. R.; Shutov, A. D. Legumains - a Family of Asparagine-Specific Cysteine Endopeptidases Involved in Propolypeptide Processing and Protein Breakdown in Plants. *J. Plant Physiol.* **2002**, *159* (12), 1281–1293. https://doi.org/10.1078/0176-1617-00853.

(192) Lycett, G. W.; Delauney, A. J.; Gatehouse, J. A.; Gilroy, J.; Croy, R. R.; Boulter, D. The Vicilin Gene Family of Pea (Pisum Sativum L.): A Complete cDNA Coding Sequence for Preprovicilin. *Nucleic Acids Res.* **1983**, *11* (8), 2367–2380. https://doi.org/10.1093/nar/11.8.2367.

(193) Shewry, P. R.; Napier, J. A.; Tatham, A. S. Seed Storage Proteins: Structures and Biosynthesis. *Plant Cell* **1995**, *7* (7), 945–956. https://doi.org/10.1105/tpc.7.7.945.

(194) Song, Y.-C.; Das, D.; Zhang, Y.; Chen, M.-X.; Fernie, A. R.; Zhu, F.-Y.; Han, J. Proteogenomics-Based Functional Genome Research: Approaches, Applications, and Perspectives in Plants. *Trends Biotechnol.* **2023**, *41* (12), 1532–1548. https://doi.org/10.1016/j.tibtech.2023.05.010.

(195) Grossmann, J.; Fernández, H.; Chaubey, P. M.; Valdés, A. E.; Gagliardini, V.; Cañal, M. J.; Russo, G.; Grossniklaus, U. Proteogenomic Analysis Greatly Expands the Identification of Proteins Related to Reproduction in the Apogamous Fern Dryopteris Affinis Ssp. Affinis. *Front. Plant Sci.* **2017**, *8*, 336. https://doi.org/10.3389/fpls.2017.00336.

(196) Jenkins, C.; Orsburn, B. The Cannabis Proteome Draft Map Project. *Int. J. Mol. Sci.* **2020**, *21* (3), 965. https://doi.org/10.3390/ijms21030965.

(197) Mandaokar, A. D.; Koundal, K. R. Cloning and Characterization of Legumin Storage Protein Gene of Chickpea (Cicer Arietinum L). *J. Plant Biochem. Biotechnol.* **2000**, *9* (1), 7–12. https://doi.org/10.1007/BF03263075.

(198) *The Biochemistry of Plants*. ScienceDirect. https://www.sciencedirect.com/book/edited-volume/9780080926155/the-biochemistry-of-plants (accessed 2025-12-02).

(199) Lycett, G. W.; Croy, R. R. D.; Shirsat, A. H.; Boulter, D. The Complete Nucleotide Sequence of a Legumin Gene from Pea (Pisum Sativum L.). *Nucleic Acids Res.* **1984**, *12* (11), 4493–4506. https://doi.org/10.1093/nar/12.11.4493.

(200) Müntz, K.; Horstmann, C.; Schlesier, B. Vicia Globulins. In *Seed Proteins*; Shewry, P. R., Casey, R., Eds.; Springer Netherlands: Dordrecht, 1999; pp 259–284. https://doi.org/10.1007/978-94-011-4431-5_12.

(201) Gatehouse, J. A.; Bown, D.; Gilroy, J.; Levasseur, M.; Castleton, J.; Ellis, T. H. Two Genes Encoding "minor" Legumin Polypeptides in Pea (Pisum Sativum L.). Characterization and Complete Sequence of the LegJ Gene. *Biochem. J.* **1988**, *250* (1), 15–24. https://doi.org/10.1042/bj2500015.