A comparative study of the structural and physicochemical properties of the major proteins from *Camelina sativa* (L.) Crantz and *Brassica napus* L.

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in the Department of Food and Bioproduct Sciences
University of Saskatchewan
Saskatoon

By
Suneru P. P. Ussetti Mohottalalage
2016

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying, publication, or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of Department of Food and Bioproduct Sciences
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5A8
Canada

ABSTRACT

Camelina sativa (L.) Crantz is a new industrial oilseed crop suitable for the Canadian prairies and which shows potential benefits for the Canadian bio-economy. This study was carried out to identify the major proteins of the oil-free residue (meal) of *C. sativa* (camelina) while investigating their structural and physicochemical properties. Canola (*Brassica napus* L.) was used as the control in the study.

Camelina seeds were treated with Viscozyme® (0.1 mL/g) to remove mucilage. The mucilage free meal contained 51.3% protein (dwb, %N×6.25) which was greater than in canola. Both camelina and canola meals shared similar profiles for polypeptides and amino acids. At acidic pHs, canola meal had higher soluble protein content than did camelina meal, but the opposite was observed when the pH moved toward alkaline. A pH of 4.5 identified as the apparent isoelectric point (pI) of the protein from these two meals, which is presumably a cruciferin-napin complex.

The 11S and 2S proteins of both seed types were isolated and purified using liquid chromatography. The purified 11S protein from camelina and canola contained predominantly cruciferin with minor contamination with non-targeted storage proteins. Of the non-cruciferin contamination of camelina, vicilin (7S) found to be abundant. The purified 2S protein from camelina contained napin and a noticeable amount of late embryogenesis abundant (LEA) protein, whereas non-napin contamination was minor in canola. In camelina, cruciferin, napin and vicilin expressed from eleven, four and six genes, respectively, were identified. The oil body proteins were also isolated and several isoforms of oleosin were found in camelina and canola, whereas putative isoforms of caleosin and steroleosin were found only in camelina.

The structural and physicochemical properties of cruciferin and napin were studied in response to changing pH and temperature. The predominance of β-structure and α-helix content in the 2˚ structure of cruciferin and napin, respectively, was confirmed for both camelina and canola. Cruciferin from camelina and canola exhibited acid-induced structural unfolding at the 3˚ structure level. Cruciferin was not completely unfolded and assumed an intermediate state, plausibly a molten globule. Napin structure was not as sensitive as cruciferin to changing medium pH or an increase in temperature. Cruciferin exhibited high thermal stability (>80˚C) at neutral and alkaline pH, whereas the opposite was observed at acidic pH. Results showed that the cruciferin and napin responded differently to changing pH and temperature. Therefore, conditions of oil extraction and protein recovery from meal may affect these two storage proteins differently.

ACKNOWLEDGEMENTS

I would like to express my humble gratitude to my supervisors Drs. Janitha Wanasundara and Robert Tyler for their continued guidance and encouragement, constructive criticisms, and valuable advices provided throughout the M.Sc. Program. I would like to thank my advisory committee members, Drs. Dwayne Hegedus and Phyllis Shand for the continuous support and expertise provided to progress throughout the project and complete the thesis. Also I sincerely thank Dr. Xiao Qiu, and Dr. Darren Korber who are the current and former chairs, respectively, for my graduate advisory committee for their suggestions and advices provided in many aspects throughout the programme to make this research a success. I would like to extend my thanks to Dr. Rex Newkirk for serving as the external examiner.

I am grateful to the research and administration staff of the Agriculture and Agri-Food Canada (AAFC), Saskatoon Research and Development Center for the opportunity given to work under the Research Affiliate Program (RAP), and providing facilities to conduct this research project. I would like to especially thank Tara McIntosh for the immense support and expertise provided to conduct the experiments. I take this opportunity to thank Cathy Coutu, Myrtle Harrington, Dan Southerland, Dan Hupka, Debbie Anderson, Doug Baldwin and Simarpreet Singh for supporting me in numerous ways to succeed the goals of my thesis research. Also, I am grateful to Dr. Kevin Falk for providing field grown camelina seeds required for this study. I am thankful to Agriculture and Agri-Food Canada research project "Camelina-an integrated Industrial Oilseed Crop Platform for Canada (RBPI 2535) and University of Saskatchewan, Global Institute of Food Security (GIFS) project "Developing Camelina as a Modern Crop Platform" for the funding support for my research and graduate program.

I would like to thank my wife Shwetha Jayani Liyange, my parents Mr. Clinton Susantha Perera and Mrs. Vineetha Vithanage, and my brother Vidyu Suyama Perera for being in my side for the every step of this journey. I especially thank Dr. Manjula Bandara and Mrs. Chandara Bandara from Saskatoon for their continuous encouragement and support. Finally, I thank my fellow graduate students in the department of Food and Bioproduct Sciences and all my friends who were there to support me whenever I was in need.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

1D/1DE	One dimensional/one dimensional electrophoresis
2D/2DE	Two dimensional/two dimensional electrophoresis
AAFC	Agriculture and Agri-Food Canada
ANS	1-anilino-8-napthalensulfonate
BN	*Brassica napus*
CA	Chlorogenic acid
CCC	Canola Council of Canada
CD	Circular dichroism
CEC	Cation exchange column
CRD	Complete randomized design
CRFA	Canadian Renewable Fuel Association
CRU	Cruciferin
CS	*Camelina sativa*
DE	Digestible energy
DH	Double haploid
DSC	Differential scanning calorimetry
FA	Fatty acid
FAD	Fatty acid desaturase
FAME	Fatty acid methyl esters
FWHH	Full width at half height
FT-IR	Fourier transform infrared
G	Genome
GL	Glucosinolate
HIC	Hydrophobic interaction chromatography/ column
I	Intermediate
IA	Intra-chain disulfide bond containing
IE	Inter-chain disulfide bond containing
IEA	International Energy Authority
IEF	Isoelectric focusing

IPCC	Intergovernmental Panel on Climate Change
IPG	Immobilized pH Gradient
LC	Liquid chromatography
LEA	Late embryogenesis abundant
Mb	Mega base
ME	Metabolizable energy
MS	Mass spectrometry
Mt	Megaton
MWCO	Molecular weight cut off
MWM	Molecular weight markers
N	Native
NAP	Napin
NRC	Natural Resources Canada
NTS	Normalize total spectra
OB	Oil bodies
OBP	Oil body protein
OLE	Oleosin
PAGE	Polyacrylamide gel electrophoresis
PSV	Protein Storage Vacuoles
S	Svedberg unit
SDS	Sodium dodecyl sulfate
SEC	Size exclusion column
SMA	Saskatchewan Ministry of Agriculture
TAG	Triacylglycerol
TEM	Transmission electron microscopy
U	Unfolded
Vic	Vicilin

1. INTRODUCTION

The *Brassicaceae* (*Cruciferae*) family contains a number of economically important species which provide edible and industrial oils, leafy and root vegetables, condiments, and fodder for animal feed use. These plant species are cultivated in a variety of climates around the world, including Canada (Najda, 1991; Warwick, 2011; Warwick, Francis & Mulligan, 2013). Canola, the world's second largest oilseed crop (Food and Agriculture Organization (FAO), 2014), was developed in Canada from three species of the *Brassicaceae* family, *i.e*. *Brassica napus*, *Brassica rapa* and *Brassica juncea*. In addition, *Brassica carinata* (Ethiopian mustard), *Camelina sativa* (false flax), and *Crambe abyssinica* have been recognized as potential oilseed crops for the Canadian prairies that could be developed as dedicated industrial crops (Warwick & Gugle, 2003; Genome Prairie, 2014). Their ability to adapt to diverse environments, fatty acid profile, high oil and protein contents, and amenability to genetic modification through plant breeding and recombinant methods are the major factors that have led to extensive development of *Brassica* species as commercial crops. Among the *Brassica* oilseeds, canola is the most economically important to Canada. Canadian canola production in 2015 was 17.2 million tonnes. According to a study carried out in 2013, canola contributed an average of $19.3 billion anually to the Canadian economy (Statistics Canada, 2015; Canola Council of Canada (CCC), 2013).

Oil, protein and fibre are the major components that can be obtained from canola seed. Oil is the most economically valuable component and primarily is used for human consumption or biofuel production. After palm oil and soya bean oil, canola oil is the most widely consumed vegetable oil in the world (FEDIOL, 2016). Through decades of research, scientists have been able to develop canola varieties with unique fatty acid profiles - high in unsaturated fatty acids (~61% monounsaturated and ~32% polyunsaturated), low in saturates (~7%) and low in erucic acid (<2%) - and low in glucosinolates (< 30 µmol/g) (Barthet, 2014; CCC, n.d.-a). Canola meal, which is the by-product of oil extraction, is a rich source of protein and is utilized as animal feed (Tan, Mailer, Blanchard, & Agboola, 2011a; Newkirk, 2015). Moreover, a variety of ready-to-use commercial canola protein products have GRAS (Generally Recognized as Safe) approval (Wanasundara, Siong, Alashi, Pudel, & Blanchard, 2015) and may be used in food formulations.

The oil of *Brassica* oilseeds is one of the major feedstocks for biodiesel production in Canada. Renewable liquid fuels such as ethanol and biodiesel are alternatives to liquid fossil fuel that can circumvent many of the sustainability and environmental issues associated with fossil fuels (Ho, Ngo, & Guo, 2014). The demand for biodiesel in Canada has increased as the government has announced national standards for biofuel utilization with a mandate of blending 2% biodiesel in petroleum diesel and heating oil [Canadian Renewable Fuels Association (CRFA), 2014]. To date, canola oil has been the main raw material for biodiesel production in Canada. Mobilizing canola oil production from edible to biodiesel is not sustainable in the long run because of the limitations it can cause to the food oil supply, cultivated land area and crop diversity. For these reasons, development of alternate oilseed crops dedicated to biodiesel production has been carried out and the Canadian emphasis has been on candidate species in the *Brassicaceae* family *i.e.* Ethiopian mustard (*B. carinata*) and camelina (C. *sativa*). The competitive advantages that camelina possesses as an oilseed crop in the prairies are discussed further in the literature review.

Camelina is a fairly new crop to Canada. It originated in Europe and Central Asia during the Bronze age (Ehrensing & Guy, 2008) and was traditionally used as an oilseed for food, fuel and medicine (Fleenor, 2011). Later, it lost its popularity to high yielding grains and other oilseed crops (Ehrensing & Guy, 2008; Fleenor, 2011). Camelina as a dedicated industrial crop provides an opportunity to expand the oilseed production areas of the Canadian prairies beyond the traditional canola growing areas because of the special adaptation of this plant to low moisture conditions and different soil types. With this re-emergence of camelina as an oilseed, the oil has gained interest for use as a salad oil because of its high alpha-linolenic acid content. Cold-pressed camelina oil has approval from Health Canada and is available in the market for dietary purposes (Health Canada, 2012). To make camelina a sustainable industrial oilseed, finding uses for the oil-free meal and meal components is imperative. Considering the nutritional value of the meal protein, camelina meal has gained approval as an animal feed in the United States and Canada, with some limitations on the inclusion level depending on the animal species. The potential use of camelina meal in a wide array of protein-based industrial products such as adhesives, plastics, gels and emulsifiers also has been suggested, and would enhance the value of camelina seed. Only limited information is available on camelina protein and other non-oil components of the seed. As the major chemical component of the meal, a better understanding of the protein with respect to

types, structural properties, physical and chemical properties, and techno-functionalities would aid the development of strategies for utilization of camelina protein.

The aim of this study was to investigate camelina seed proteins in detail to support scientific understanding that will lead to development of various bioproducts and uses. To achieve this goal, the storage proteins and oil body proteins of camelina grown in Canada were investigated and compared with those of canola (*Brassica napus* L.)

1.1 Hypotheses

The following hypotheses were tested in this study:

1) *C. sativa* contains 11S and 2S and oil body proteins similar to those of *B. napus*;
2) *C. sativa* 11S protein has similar structural and physicochemical properties as *B. napus* 11S protein;
3) *C. sativa* 2S protein has similar structural and physicochemical properties as *B. napus* 2S protein; and
4) The method employed to separate and purify oil body proteins will yield highly purified *C. sativa* and *B. napus* oil body proteins.

These hypotheses will be pursued through the following specific objectives.

1.2 Objectives

1) To investigate ways of obtaining *C. sativa* 11S and 2S storage proteins and oil body proteins with minimum interference from mucilage;
2) To investigate the chemical constituents of the meal, and identify methods and conditions suitable for purification of 11S, 2S and major oil body protein purification;
3) To identify the protein composition of purified 11S, 2S and oil body protein fractions of *C. sativa* and compare with those of *B. napus*;
4) To investigate the secondary and tertiary structural properties of 11S and 2S proteins of *C. sativa* and compare with those of *B. napus*; and
5) To investigate the physicochemical properties (solubility and thermal properties) of 11S and 2S proteins of *C. sativa* and compare with those of *B. napus.*

LITERATURE REVIEW

2.1 Biodiesel and oilseeds

The International Energy Authority (IEA) has predicted that the global energy requirement in 2020 will be 50% more than the world consumption in 2000 (Industry Canada, 2004). Wise and more efficient use of diminishing fossil fuel supplies and a major shift to highly efficient energy sources such as fuel cells, small- to medium-scale distributed co-generation systems and renewable biofuels (biogas, biodiesel, bio-oils and alcohol) appear to be the preferential ways of achieving this global energy goal (Industry Canada, 2004). Sustainable energy production using renewable sources becomes very important in this context because of the highly unstable global energy market and unpredictable, large spikes in oil and natural gas prices. Energy produced from renewable sources has less detrimental impact on the environment (Danyard & Danyard, 2011). In fact, it mitigates the problems associated with global warming due to the high volume of carbon dioxide emission from fossil fuels, encouraging the world to focus more on sustainable energy production using renewable sources. According to the IEA (2012), in 2014 approximately 13% of total global energy consumption would be produced using renewable energy sources and 10% of this share would be from bioenergy. Bioethanol and biodiesel are the two major counterparts of global bioenergy production and widely used for transportation purposes and other industrial applications (Ho, Ngo & Guo, 2014). Ethanol produced using carbohydrate-rich sources and ligno-cellulosic biomass is generally referred to bioethanol, and fatty acid methyl esters (FAME) produced from plant oils and animal fats is categorized as biodiesel. To date, many countries, including Canada, have taken necessary actions to promote biodiesel consumption to address the drawbacks associated with fossil fuel utilization (CRFA, 2014).

The Government of Canada announced its strategy for renewable fuel utilization in 2007 (CRFA, 2014). This encouraged the establishment of the Canadian standards of blending 2% biodiesel in petroleum diesel and heating oil along with 5% bioethanol content in gasoline. As a result, Canada produced 400 ML of biodiesel by the end of 2014 (CRFA, 2014). Other than reducing the detrimental environmental effects, these mandates also ensured the provision of new

market opportunities for agricultural producers and rural communities (Natural Resources Canada (NRC), n.d). The Canadian renewable fuel industry currently contributes $3.5 billion economic benefits to the domestic economy every year (CRFA, 2014). If Canada plans to adhere to these mandates by reducing imports and improving domestic production, it will need to increase biodiesel production by 450% (Danyard & Danyard, 2011). The Canadian mandate for biodiesel assumes that it would replace 5% of the current national diesel consumption (CCC, n.d.-b), which indeed would require a dramatic increase in biodiesel production. The process of converting plant (or animal) fat into biodiesel is called transesterification. In this process, fatty acids of triacylglycerol (TAG) molecules are converted to alkyl (methyl, ethyl or propyl) esters with the help of a catalyst while glycerol is produced as a co-product (Balat & Balat, 2010). Methanol is the more widely used alcohol to provide the alkyl group for this conversion compared to ethanol. Therefore, the term biodiesel mainly refers to fatty acid methyl esters (FAME). Utilizing vegetable oil to produce biodiesel is common, compared to animal fat such as beef tallow, lard, poultry fat or fish oils (Feddern et al., 2011).

The dominant vegetable oil crops used to generate biodiesel include rapeseed or canola (*Brassica napus* and *Brassica rapa*), soybean (*Glycine max*) and oil palm (*Elaeis guineensis*), especially in European Union countries and Canada, the United States, and tropical Asian countries, respectively (Ho, Ngo, & Guo, 2014; Romano & Sorichetti, 2011; CCC, n.d.-b). Apart from these crops, oil from sunflower (*Helianthus annuus*), flax (*Linum usitatissimum*), peanut (*Arachis hypogaea*), safflower (*Carthamus tinctorius*), castor oil (*Ricinus communis*), jatropha (*Jatropha curcas*), Ethiopian mustard (*Brassica carinata*) and microalgae are considered suitable for biodiesel production, as most do not interfere with edible oil supply (Romano & Sorichetti, 2011; Agrisoma Bioscience Inc., 2015). The feedstock oils from these plants differ in terms of the type of fatty acids present in the triglyceride, which determines the degree of saturation/unsaturation. These factors directly influence the production process, quality and cost of biodiesel (Ramos, Fernández, Casas, Rodríguez, & Pérez, 2009). Biodiesel produced using these plant-derived oils is used alone or blended with petroleum-based diesel for diesel ignition engines (Ho, Ngo, & Guo, 2014). Biodiesel struggles to compete with petroleum-based diesel in terms of price competitiveness. To keep the price of biodiesel down, a low price of the starting vegetable oil, cost effective FAME production technologies and also development of high-value coproducts is essential (Ho, Ngo, & Guo, 2014).

Canola is the major oilseed in Canada and the oil is suitable for human consumptions as well as biodiesel production. Dependence on canola oil to reach biodiesel production goals interferes directly with the food oil supply. Canola was developed to produce a characteristic fatty acid (FA) composition, *i.e.* a very low level of saturated FAs (~7%), relatively high level of monounsaturated FA (61%) and an intermediate level of polyunsaturated FAs (32%), with a good balance between omega-6 (21%) and omega-3 (one-third of total polyunsaturated) FAs, which is consistent with nutrition recommendations targeting reduced saturated fat intake (Barthet, 2014; CCC, n.d.-a). Therefore, diverting such nutritionally valuable oil for fuel generation undermines the years of research and financial investment of developing canola as a world leading vegetable oil crop. The Canadian canola industry is confident in providing 80% of the required feedstock by 2015 (CCC, n.d.-b); However, development of a dedicated oil crop for industrial uses is a need that could preserve canola oil for human consumption. Hence Exploitation of alternate fuel biomass is necessary to develop a stable bioeconomy through sustainable energy production.

2.2 *Brassicaceae* species as industrial oilseed crops for Canada

Among many other plant families, the family *Brassicaceae (Cruciferae)* comprise a variety of potential candidates to fit into industrial oilseed platforms in temperate climate regions, including Canada. *Brassica carinata* (Ethiopian mustard), *Camelina sativa* (camelina), *Crambe abyssinica* (crambe) and *Eruca sativa* (eruca) are some of the crucifer plants that fall into this category (Warwick, 2011; Feussner, 2015). These *Brassicaceae* plants provide seed oil, protein-rich seed meal and fibre (Brown, n.d). The seed oils can be used in various industrial applications including a feedstock oil for biodiesel and lubricants, ingredients in cosmetics, and for medicinal uses (Warwick, 2011). Seed meal is the co-product of oil extraction, which can be a source of protein and energy in animal feed formulations and feed stocks for other bioproduct development. Other than oil and protein, glucosinolates and polysaccharide mucilage present in seed also provide some additional economic benefits for these *Brassicaceae* plants (Warwick, 2011). The ligno-cellulosic crop residue also plays an important role apart from the seed chemical traits of these *Brassicaceae* plants. It can be used as biomass for ethanol production (Ballesteros, Oliva, Negro, Manzanares, & Ballesteros, 2004; Petersson, Thomsen, Hauggaard-Nielsen, & Thomsen, 2007) and important organic matter for soil in crop rotations with the ability to suppress some weeds, nematodes and diseases (Clark, 2007). It appears that these plants provide a number of industrially valuable products, making these oilseed species well-suited for industrial oilseed platforms.

Ethiopian mustard grows well under Canadian climatic conditions and the seed contains ~44% oil. The oil is used in producing surfactants, dietary supplements, bioplastics and polymers, cosmetics, gels and films in addition to biojet fuel (Johnson, Falk & Eynck, 2014). The plant shows good resistance to blackleg, lodging and pod shattering, which are common agronomic problems associated with other *Brassicaceae* oilseed crops. High yielding and high oil containing verities of *B. carinata* suitable for the Canadian prairies have been developed by Agriculture and Agri-Food Canada (Agriculture and Agri-Food Canada, 2015). *B. carinata* seed also contains 28% protein, which is beneficial in producing animal feed. It contains higher amount of glucosinolates in the meal where was singrin found to be the predominant group of glucosinolate (Alemayehu & Becker, 2005) accounting for over 95% of total glucosinolate in *B. carinata*; no variation in the glucosinolate profile has been reported (Marquez-Lema, Fernandez-Martinez, Perez-Vich & Velasco, 2009). Reduction of glucosinolate levels in the meal is necessary to improve the feed value and it can be achieved through genetic and breeding (Alemayehu & Becker, 2005). The Canadian Food Inspection Agency (CFIA) has approved *B. carinata* meal for use in grower and finisher beef cattle (Agrisoma Bioscience Inc., 2014), placing this oilseed in a better position in the oilseed value-chain.

Crambe is another *Brassicaceae* crop suitable as an industrial oilseed crop. The seed oil content varies from 30-43% (Lazzeri, Leoni, Conte, & Palmieri, 1994; Wang, Tang, Chu, & Tian, 2000; Elementis Specialties, Inc, 2010; Oilseed Crops, 2016), while seed protein content ranges from 20-26% (Lazzeri, Leoni, Conte, & Palmieri, 1994; Massoura, Vereijken, Kolster, & Derksen, 1998). Crambe oil contains erucic acid (C22:1) which accounts for 55-60% of the total fatty acids (Lazzeri, Leoni, Conte, & Palmieri, 1994; Massoura, Vereijken, Kolster, & Derksen, 1998; Wang, Tang, Chu, & Tian, 2000). Crambe seed oil has a wide array of industrial applications, such as in the production of lubricants, nylon 13-13, paints and coatings, pharmaceuticals, cosmetics, and plastics, and also has potential for biodiesel (Nelson, Grombacheer & Baltensperger, 1993; Lazzeri, Leoni, Conte, & Palmieri, 1994; Endres & Schatz, 2003). The United States Food and Drug Administration (US-FDA) approved crambe meal as a livestock feed in 1981, and allowed an inclusion level of 4.2% in beef cattle diet (Glaser, 1996). Crambe seed contains a high level of glucosinolates, which results in high residual levels of glucosinolates and their breakdown products in the meal. It was found that dehulled, defatted crambe meal contained 116 μmol g^{-1} total glucosinolate content, with vinyl-oxazolidine-thione and isothiocyanates, which are

glucosinolate breakdown products, accounting for 16 mg/g and 8 mg/g, respectively (Yong-Gang, Steg, Smits, & Tamminga, 1994). Therefore, the meal can be especially toxic for monogastric animals whereas ruminant show fair tolerance (Endres & Schatz, 2003). This high level of glucosinolates makes it a potential candidate for a biofumigant or fungicide (Mohiuddin, Qureshi, Nasir, & Khatsi, 1990; Peterson, Cossé, & Coats, 2000). Due to high levels of erucic acid and oil content compared to rapeseed, crambe has the potential to be developed as a valuable industrial oilseed crop; however, crambe is not common in Canada. It was commercially grown in North Dakota, Montana and few other Northern states of US during last few decades (Glaser, 1996).

Eruca, commonly known as arugula, is also a *Brassicaceae* oilseed crop that may fit into the industrial oilseed platform. Eruca is an annual herb commonly grown in the Middle East, India and Pakistan (Flanders & Abdulkarim, 1985), and the seed contains approximately 35% oil which is high in erucic acid (Mumtaz et al., 2012; Sharma, Garg, & Alam, 2014). The oil has proved its potential for use as a feedstock for biodiesel (Mumtaz et al., 2012), lubricants, illuminating agents, surfactants and therapeutic oil (Sharma, Garg, & Alam, 2014). Eruca seed contains 37% protein, which is comparatively high compared to other *Brassicaceae* species (Chakrabarti & Ahmad, 2009). Due to its higher protein content, the meal has a potential to be utilized in various industrial applications similar to the above mentioned crucifer oilseed species.

Although all of these crucifer plants show potential to be developed as industrial oilseed crops for Canada, *Camelina sativa* delivers several benefits over the other species because of the low input cultivation, comparatively high drought tolerance, resistance to cold and diseases common to *Brassicaceae* crops (e.g. blackleg and alternaria blight), ease of genetic manipulation, resistance to genetic contamination, seed chemical components, and wide range of applications for meal by-products (Feussner, 2015). Therefore, over the last several years more focus was on developing camelina as an industrial oilseed crop to support a sustainable bioeconomy in Canada (Derouin, 2014).

2.3 Camelina as an industrial oilseed crop for Canada

2.3.1 Agronomic characteristics

Camelina is a herbaceous annual or winter annual plant which grows to an average height of 30 to 90 cm (Ehrensing & Guy, 2008; Fleenor, 2011). It has arrow-shaped, sharp-pointed leaves 5 to 8 cm in length. The stem of camelina is woody and branched, with either a smooth or hairy surface (Ehrensing & Guy, 2008; Fleenor, 2011). The plant produces small yellow (pale yellow or

greenish yellow) flowers with four petals which are predominantly self-pollinated. Pods of camelina are pear-shaped and contain 8-10 seeds (Saskatchewan Ministry of Agriculture (SMA), n.d). The seeds are small (1000 seeds weights of 0.8 to 2 g) and pale yellow in color. The camelina plant has a shorter lifespan than canola and matures in 85 to 100 days (Ehrensing & Guy, 2008). Because of its shorter lifespan, it can be grown in areas, such as the prairies of western Canada, which are suitable for crops with short vegetation periods (Feussner, 2015). The seeds can be successfully sown in autumn and germinate in early spring at temperatures near 0°C, well in advance of the emergence of weeds (Putnam, Budin, Field, & Breene, 1993), and are therefore successful under the growing conditions that exist on the Canadian prairies. Figure 2.1 depicts the different stages of the camelina life cycle, illustrating some morphological characteristics of the plant.

Camelina is identified as a low input crop which can be grown on marginal lands (Sederoff, 2012). It responds to nitrogen, sulfur and phosphorous fertilizer in a similar way to other *Brassicaceae* crops, such as canola and mustard (SMA, n.d; Ehrensing & Guy, 2008). Camelina shows resistance to common pests of *Brassicaceae*, such as flea beetle and cabbage root fly (Pachagounder, Lamb, & Bodnaryk, 1998; Henderson, Hallett, & Soroka, 2004), and diseases, such as blackleg and alternaria blight (Salisbury, 1987; Conn, Tewari, & Dahiya, 1988). High shatter resistance is an important agronomic characteristics of this plant. Several studies showed that camelina has a high yield potential similar to that of other *Brassicaceae* oilseed species, namely *B. rapa, B. napus, B. juncea, Sinapis alba and C. abyssinica* (Putnam, Budin, Field, & Breene, 1993, Gugel & Falk, 2006; Eynck & Falk, 2013).

2.3.2 Genetics and Breeding

Camelina possesses a hexaploid genome structure (Nguyen et al., 2013; Kagale et al., 2014). According to Kagale et al. (2014), the camelina contains three sub genomes (genome 1/CS-G1, genome 2/CS-G2 and genome 3/CS-G3) due to a whole-genome triplication occurring in a common ancestor. The estimated camelina genome size is 785 Mb with 89, 418 predicted protein-coding genes. The total number of camelina genes is three times as large as *Arabidopsis thaliana* and comparable to that of bread wheat (22 times larger genome than camelina). Although the three sub genomes are similar, CS-G3 appears to have some expression dominance. Most camelina traits are controlled by multiple loci due to the polyploidy nature, making traditional breeding and gene

manipulation approaches comparatively difficult to control desired traits to gain economic benefits (Kagale et al., 2014).

Figure 2.1. Stages of the *C. sativa* life cycle. (a) seeds; (b & c) seedlings; (d) plants before flowering; (e) plants at flowering, (f) flower buds; (g) inflorescence; (h) immature pods; and (i) mature pods. Pictures were captured using a Nikon D7000 camera attached to a AF-S micro Nikkor 105 mm lens. Images were not taken under the same magnification.

Despite the genetic complexity, scientists were able to control economically-important genetic traits of camelina through genetic manipulation by simple *Agrobacterium*-mediated transformation (Lu & Kang, 2008; Liu et al., 2012). Lu and Kang (2008) successfully produced hydroxyl fatty acids in camelina oil via seed specific expression of a castor bean fatty acid hydroxylase gene using *Agrobacterium*-mediated transformation. The same method also was successfully utilized elsewhere to improve plant growth rate and seed yield in camelina by overexpressing *Arabidopsis* purple acid phosphatase 2 (*AtPAP2*) under controlled environmental conditions (Zhang et al., 2012). These transgenic camelina plants have the potential to provide higher seed yield per unit area which results in higher oil yield than non-genetically modified (non-GMO) camelina. Expression of docosahexaenioc acid (DHA, 22:6 ω-3), which is a polyunsaturated fatty acid found in marine organisms, in camelina seed was reported by Petrie et al. (2014). The authors were able to express 12% DHA with a minimum amount of intermediate fatty acids in the oil fraction. Furthermore, they were able to produce significant ω-3:ω-6 ratios without any novel long chain ω-6 products. Moreover, production of industrially important 3-acetyl-1,2-diacyl-*sn*-glycerols (acetyl-TAG) and ω-7 monoenes using the same techniques has also been reported (Nguyen et al., 2013; Liu, 2015). The ability to develop transgenic camelina plants with favorable traits could enhance the value of camelina as an industrial oilseed crop.

Camelina does not outcross with other commonly available *Brassicaceae* species such as *B. napus* and *B. juncea*, hence it is considered a safe crop that provides higher resistance to transgene contamination (Feussner, 2015). According to CFIA (2014), several attempts were made to cross *C. sativa* with species outside the genus. The closest phylogenetic relatives of *C. sativa* were identified as *A. thaliana* and *Arabidopsis lyrata* (Figure 2.2, Hutcheon et al., 2010). Despite this close relationship, no evidence was found to prove that these species are capable of crossing with each other (CFIA, 2014). Attempts to cross *C. sativa* with other *Brassicaceae* species with commercial importance, such as *B. rapa*, *B. napus*, *B. juncea* and *B. nigra*, were unsuccessful. The reason for the failure of crossing these species could be the distant phylogenetic relationship between *C. sativa* and other commercial *Brassicaceae* species (CFIA, 2014). However, the possibility of crossing camelina with other species is important for the development of favorable traits, such as adaptation to severe climatic conditions, especially temperature and moisture stress on the prairies. Crossing camelina within species of the genus *Camelina*, such as *C. microcarpa* and *C. alyssum*, has been successful (CFIA, 2014).

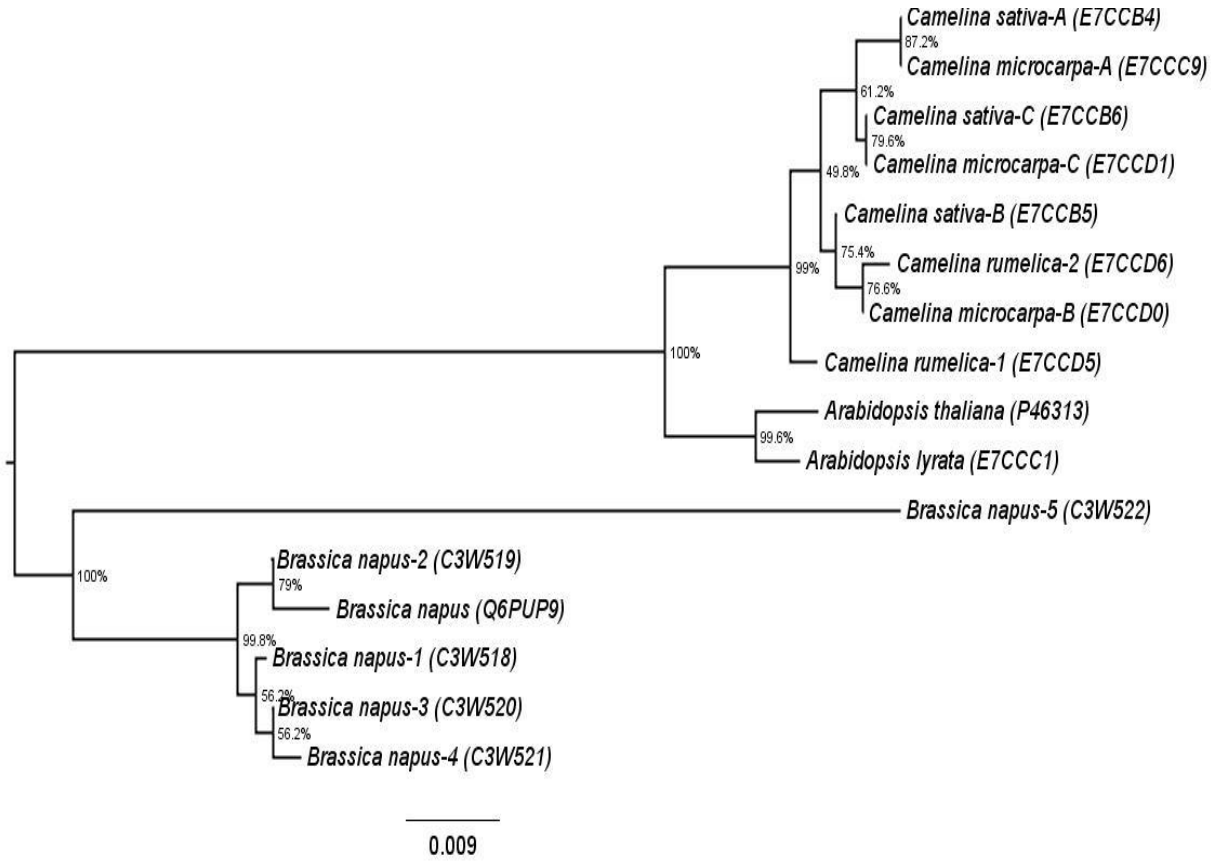

Figure 2.2. Phylogenetic relationship of *C. sativa* with other *Brassicaceae* oilseed crops (*Camelina microcarpa, Camelina rumelica, A thaliana, A. lyrata and B. napus*) based on fatty acid desaturase (FAD 2) (Adapted and re-drawn from Hutcheon, et al., 2010). The amino acid sequences of the proteins expressed from the FAD 2 genes of each species were obtained from the UniProt/ Swiss-Prot (http://www.uniprot.org/) database. The evolutionary history was inferred using the Neighbor-Joining method with the amino acid sequences of FAD 2 proteins in MEGA6 software. The bootstrap values are presented as % at each node.

A study carried out to hybridize *C. sativa* with other North American camelina species has resulted in high fertility in F1 hybrids recovered after crossing with *C. alyssum*, proving that gene flow between *C. sativa* and its wild North American relatives is possible (Séguin-Swartz, Nettleton, Sauder, Warwick, & Gugel, 2013). Several breeding programs in Europe and the US have released successful camelina cultivars over the years. Presumably, 'Calena' is the most successful camelina cultivar in North America and was originally developed in Europe. 'Lindo', 'Ligena' and 'Celine' are other European cultivars that were introduced to North America. Camelina cultivars developed in the US include 'Blaine Creek', 'Suneson' and 'Cheyenne' (Eynck & Falk, 2013)

A number of breeding programs have been initiated in Canada due to the growing interest on camelina production during the last few years. The focus has been on developing camelina cultivars with enhanced traits, such as high oil/ protein content, large seed size, disease resistance and herbicide tolerance, to develop superior cultivars for western Canada (Falk & Klein-Gebbinck, n.d; Eynck & Falk, 2013).

2.3.3 Chemical characteristics of the seed and meal

The chemical composition of the seed was the key for camelina that opened up the doors to an industrial oilseed platform. Camelina seed oil content varies from 32 to 45% (SMA, n.d; Moser, 2010). The unique fatty acid profile is suitable for biodiesel production. Protein, which is the next most abundant constituent in the seed, accounts for 27-32% of the total seed weight (SMA, n.d). Apart from these two major components, camelina seed contains carbohydrates, vitamins, minerals and antioxidative compounds that add value to this oilseed crop.

Several research groups have reported the fatty acid profile of camelina oil (Table 2.1). It contains mono (oleic, C18:1; gondoic, C20:1; and Erucic, C22:1) and polyunsaturated (linoleic, C18:2 and linolenic; C18:3) fatty acids. Linolenic acid (34-39%) appears to be the most significant polyunsaturated fatty acid in camelina oil. The saturated fatty acid content in camelina oil is low compared to that of the unsaturated fatty acids. The fatty acid profile, and the ratios of monounsaturated to polyunsaturated fatty acids and linolenic acid to linoleic acid ratios make camelina oil suitable to be utilized in biofuel and other co-product industries, such as cosmetics, nutraceuticals and oleochemicals (Pecchia, Russo, Brambilla, Reggiani, & Mapelli, 2014). In addition, low levels of free fatty acids (<3%) and the presence of natural antioxidants make camelina oil stable with a long shelf life (Pilgeram et al. 2007; Hrastar, Petrisic, Ogrinc, & Kosir, 2009). Camelina oil biodiesel (fatty acid methyl or ethyl esters) can be produced by several methods (Fröhlich & Rice, 2005; Moser & Vaughn, 2010; Wu & Leung, 2011). Biodiesel produced from camelina oil seems to have similar fuel properties (cold flow properties, oxidative stability, kinematic viscosity, cetane number, etc.) to that of soybean- and canola-based biodiesel (Fröhlich & Rice, 2005; Moser & Vaughn, 2010). Therefore, the oil fraction of camelina can successfully be integrated into biodiesel production, making the plant suitable for an industrial oilseed crop platform.

Table 2.1. Content of fatty acids (% of total fatty acid) in *C. sativa* oil as reported by different research groups.

Fatty acid	Fobert et al., 2008	Newson, 2012	Peiretti & Meineri, 2007	Zubr & Matthaus, 2002
16:0 Palmitic	6	NR	5.7	5.3-5.6
18:0 Stearic	2	NR	2.7	2.3-2.7
18:1 Oleic (ω9)	12	16.7	12.9	14.0-16.9
18:2 Linoleic (ω6)	17	16.9	17.7	13.5-16.5
18:3 Linolenic (ω3)	38	38.1	37.3	34.9-39.7
20:1 Eicosanoic (ω9)	15	16.1	14.4	15.1-15.8
22:1 Erucic (ω9)	4	<5	2.2	2.6-3

NR=Not reported

Nguyen et al. (2013) reported that camelina seed contains storage proteins, mostly cruciferin and napin, equivalent to other *Brassicaceae* species. Other than these two storage proteins, oil body proteins that stabilize oleosomes can be found within camelina seed. The meal obtained from oil extraction is rich in protein and can be used in animal feeds and other bioproducsts development (SMA, n.d; Kim & Netravali, 2012; Reddy, Jin, Chen, Jiang, & Yang, 2012). The amino acid composition of camelina meal is similar to that of canola meal (Table 2.2). Lysine and sulfur containing amino acids, *i.e.* cysteine and methionine, which are important in animal nutrition are close to the levels found in canola. Only the level of glutamic acid shows a marked difference between the two seeds. Therefore, camelina meal has the potential to be an alternative for canola meal and enhancing its value as a sustainable industrial oilseed crop.

According to Zubr (2010), camelna seed contains significant amounts of mucilage (6.7%) and crude fiber (12.8%), which make camelina a potential source of dietary fibre. The amounts of monosaccharides and oligosaccharides in camelina seeds was not high (0.04-0.64%), albeit increased sucrose content (5.5%) was observed. It was found that camelina seed contained an outstanding amount of thiamin (B1), niacin (B3) and pantothenic acid (B5), much higher than in flaxseed and rapeseed. Camelina seed is low in micro-nutients. Among the minerals available, substantial amounts of Fe, Mn and Zn were detected.

Table 2.2. Amino acid profile of *C. sativa* and *B. napus* meal (dry weight basis)

Amino acid	Amount (% of crude protein)		Amino acid	Amount (% of crude protein)	
	Camelina (Zubr, 2003a)	Canola (Newkirk, 2015*)		Camelina (Zuber, 2003a)	Canola (Newkirk, 2015*)
Histidine	2.60	3.85	Phenylalanine	4.19	4.61
Isoleucine	3.96	3.94	Tyrosine	3.04	2.84
Leucine	6.63	7.03	Aspartic acid	8.71	8.24
Lysine	4.95	6.73	Glutamic acid	16.4	20.61
Threonine	4.25	4.85	Glycine	5.44	5.59
Tryptophan	1.15	1.51	Alanine	4.61	4.95
Valine	5.42	5.65	Proline	5.09	6.78
Methionine	1.72	2.20	Serine	5.04	4.44
Cysteine	2.12	2.60	Arginine	8.15	7.52

* calculated from the original values reported on a 12% moisture basis

Camelina seeds contain several compounds that can act as antioxidants; the most prevalent group is tocopherols. Tocopherols are associated with the oil fraction and the total tocopherol content of camelina oil was reported as 806 ppm (Zubr & Matthaus, 2002) and 760 ppm (Abramovič, Butinar, & Nikolič, 2007) which included α-tocopherol, γ-tocopherol, δ-tocopherol and plastochromanol (p-8); neither β-tocopherol nor tocotrienols were detected (Zubr & Matthaus, 2002). The average contents of α-, γ-, and δ-tocopherols and p-8 were reported as 28, 742, 20 and 15 ppm, respectively. Other than tocopherols, the antioxidative components of the seed include phenolics with a total phenolics content of camelina seed reported to be 1536 chlorogenic acid (CA) equivalent/100 g, and 9.1 and 1666 CA equivalents/100g, respectively, in the oil and meal (Terpinc, Polak, Makuc, Ulrih & Abramovič, 2012). Sinapine accounted for a large fraction of camelina polyphenols (Abramovič et al., 2007). Flavonoids, such as rutin, catechin, quercetin and quercetin-3-0-glucoside, and phenolic acids, such as p-hydroxy benzoic and ellagic acids, have

been detected in camelina seed, oil and cake in varying amounts, with the meal retaining most of these components (Terpinc et al., 2012).

Similar to many other *Brassicaceae* plants, glucosinolates (GLs) are found in camelina seed. The total content of GLs in camelina seed ranges from 14-36 μmol/g (Matthaus & Zubr, 2000). The major glucosinolates of camelina are 9-methyl-sulfinyl-nonyl glucosinolate (GL1, glucoarabin), 10-methyl-sulfinyl-decyl glucosinolate (GL2, glucocamelinin) and 11-methyl-sulfinyl-undecyl glucosinolate (GL3) (Schuster & Friedt, 1998). Glucocamelinin seems to be the most abundant glucosinolate, accounting for approximately 65% of the total glucosinolate content (Schuster & Friedt, 1998). In 2010, Health Canada declared camelina oil as safe for human consumption and clearly stated that the glucosinolate content of camelina oil was negligible (Health Canada, 2012).

2.4 Camelina meal and the current situation

When chemical composition is considered, camelina meal appears to be a promising feed source due to the 38 to 43% protein content, 10 to12% residual oil and potential antioxidative compounds (SMA, n.d; Pilgeram et al., 2007). Screw or double-pressing is a more economical way of oil extraction for biofuel production compared to the pre-press solvent extraction, therefore a high residual oil content is to be expected in the meal. Several studies indicate that the inclusion of camelina meal up to 10% in poultry rations increased the unsaturated fatty acid content in animal products, especially the levels of omega-3 fatty acids in the meat and eggs (Cherian, Campbell, & Parker, 2009; Aziza, Quezada, & Cherian, 2010; Kakani et al., 2012). However, inclusion levels over 10% may alter the egg production and quality in negative way, such as lower hen-day egg production, reduction of yolk weight, lower yolk colour and higher potential for lipid oxidation (Cherian et al., 2009). It was also found that incorporation of camelina meal in beef cattle and dairy cow rations enhanced the unsaturated fatty acid composition of the meat and milk, respectively (Halmemies-Beauchet-Filleau et al., 2011; Cappellozza, Cooke, Bohnert, Cherian, & Carroll, 2012). Moreover, the digestible energy (DE) and metabolizable energy (ME) values were improved when growing pigs were fed with corn-soybean meal based diet including 200 g/kg of screw-pressed camelina meal (Kahindi, Woyengo, Thacker, & Nyachoti, 2014). Camelina has also gained attention in the aquaculture industry due to its comparatively high apparent nutrient digestibility, high protein content and the presence of some indispensable amino acids such as methionine, lysine, phenylalanine, threonine, leucine, isoleucine and valine (Hixson, Parrish,

Wells, Winkowski, & Anderson, 2015; Hixson et al., 2015). The high residual oil content which provides significant amounts of poly- and mono-unsaturated fatty acids (oleic, linoleic and linolenic) is an added advantage for camelina meal to be utilized as an aquaculture feed (Hixson et al., 2015). Several studies showed that camelina meal can be successfully incorporated into Atlantic cod (*Gadus morhua*), Atlantic salmon (*Salmo salar*) and Rainbow trout (*Oncorhynchus mykiss*) diets up to 30, 8 and 14%, respectively (Hixson & Parrish, 2014; Hixson, Parrish, Wells, Winkowski, & Anderson, 2015). According these research studies, inclusion of camelina meal in the diets at the above mentioned levels improved the growth performance of these farmed fish. The major limitation that prevented going beyond these inclusion levels was the presence of antinutritive compounds. As the inclusion levels exceeded the limits mentioned above, the feed intake decreased due to the loss of overall palatability, which negatively affected the growth performance of these fish. It was suggested that feeding a protein concentrate rather than direct incorporation of meal or meal treated to remove/reduce aninutritive compounds may allow for an increase the incorporation level of camelina meal in fish diets. The United States Food and Drug Administration (US-FDA) has approved the use of camelina meal as an ingredient for beef cattle, broiler chicken and laying hens up to 10% of the diet, and no more than 2% of swine diet. In 2015, the CFIA approved incorporation of non-solvent extracted camelina meal up to 12% in broiler feeds (SMA, n.d).

The presence of antinutritive compounds in feed components, even in minute amounts, poses adverse effects on animals. The antinutritive compounds of camelina meal are glucosinolates (precursors of isothiocyantes, nitriles and other products), sinapine and phytic acid similar to other *Brassicaceae* seed meals available for animal feed. As shown in Table 2.3, camelina contains lower amounts of sinapine than does canola meal. However, glucosinolate content of camelina appears to be much higher than that of canola meal and also of different types.

Table 2.3. Antinutritive compounds in canola and camelina meals (dry weight basis).

Constituent	Canola (Newkirk, 2015)*	Camelina (Russo & Reggiani, 2012)
Glucosinolates (total)	4.8 µmol/g	18.5 µmol/g
Sinapine	1.13%	0.23%
Phytic acid	2.61%	2.99%

* calculated from the original values reported on 12% moisture basis

Besides being a potential source of animal feed, camelina meal could be used in producing value-added bioproducts such as biodegradable papers, thermoplastics and adhesives (Kim & Netravali, 2012; Reddy et al., 2012; Li, Qi, Sun, Xu, & Wang, 2015). Since camelina meal is a good source of protein, there is a possibility of using the protein in human food applications. Except for the amino acid composition, little information is available on the physicochemical and structural properties of camelina seed proteins which is needed for bioproduct development from the protein fraction. Exploitation of camelina seed proteins for suitable applications is crucial in making it valuable and sustainable as an industrial oilseed platform. Therefore, it is important to obtain a good understanding of proteins in *Brassicaceae* oilseed crops, along with overall knowledge of the different types of protein present in the seed.

2.5 Overview of seed proteins

Seeds are the propagation organ for plants and at the same time they have become important source of nutrients for humans and animals by providing essential macronutrients, including carbohydrates, lipids and proteins. The content of protein in cultivated crops ranges from ~10% (in cereals) to ~40% (in certain legumes and oilseeds) of seed dry weight (Shewry, Napier, & Tatham, 1995). The majority of proteins in eudicot seeds are localized in the protein bodies and termed storage proteins. The storage proteins represent approximately half of the total seed proteins in cereals, but a higher proportion is found in seeds such as soybean (Shewry, 2000) and provide nitrogen required for seed germination (Shewry et al., 1995). Other proteins in mature seeds provide metabolic and/or structural function within the seeds (Shewry et al., 1995). Granule-bound starch synthases, amylase and puroindoline found in maize and wheat (Macdonald & Preiss, 1985; Wall et al., 2010) are good examples for proteins with specific metabolic functions. Oil body proteins (OBPs) create an outer layer of oil bodies and provide structural functions in oil-storing seeds (Lin, Liao, Yang, & Tzen, 2005). However, the storage proteins are considered the most important group of proteins and have much larger contribution to the total protein content of a seed (Shewry et al., 1995). OBPs may be a significant protein in oil-storing seeds because of the abundance of oil bodies in the seed cotyledons, germ and other components.

2.5.1 Types of seed storage proteins

Acoording to Shewry et al. (1995), the seed storage proteins are the first group of proteins that went through identification and characterization due to their abundance and economic

importance. Proteins store up in seeds during embryogenesis and are degraded to provide energy for germination and seedling growth. Seed storage proteins are seldom found responsible for obvious functions, such as enzymatic and structural, other than being a nitrogen sink (Rödin, 1990). The first comprehensive classification for seed storage protein was proposed by Osborne, (1924) and was based on the solubility of proteins in water, diluted salt, aqueous alcohol and weakly acidic or alkaline solutions. According to solubility characteristics in these solvent systems, storage proteins are classified into albumins, globulins, prolamins and glutelins, respectively. Albumins, globulins and prolamins are the major groups of proteins commonly found in seeds and are of economic importance. The terminology in use to identify storage proteins of seeds is quite confusing and several trivial names are used (e.g., legumin for 11S, vicilin and convicilin for 7S, etc.). According to the sedimentation coefficient ($S_{20,w}$) most of the albumins are considered 2S although 1.7-2.2S proteins were reported. Similarly, most of the globulins are 7S and 11S although 7.5S, 8S, and 12S are reported (Shewry, 1995; Miernyk & Hajduch, 2011). Since the glutelin proteins are structurally similar to that of prolamins in wheat, barley and maize or 11S globulins in rice, it is rarely considered as a separate group (Shewry, 2000; Shewry et al., 1995). However, this solubility-based categorization has less relevance to the functions that proteins provide or their genetic relationship in the plant. Also, it is quite difficult to relate these groups of proteins with the protein types characterized according to the sedimentation coefficient ($S_{20,w}$), structure folds (cupins, prolamins) or 3-D structures of protein molecules.

2.5.2 Oil body proteins

Most seeds store lipids mainly as triacylglycerols (TAG) and to a lesser extent sterol esters in small subcellular compartments, called oil bodies (OBs) or oleosomes (Jolivet et al., 2009; Tzen, 2012). Oleosomes consist of a TAG matrix surrounded by a monolayer of phospholipids that is embedded with low molecular mass (~18-25 kDa) proteins known as oleosins (Jason, Cao, Laurent, Ratnayake, & Huang, 1993; Jolivet et al., 2009). Oleosins are the most common protein found in seed oil bodies and account for 75-80% of total oil body proteins (Huang, 1996; Jolivet et al., 2009). According to Tzen, George, & Huang (1992), oleosins prevent coalescence of OBs through steric hindrance and electrostatic repulsion. To date, a wide range of applications have been identified with respect to native, recombinant forms of oleosins or oleosin-fused polypeptides (Roberts, Scott, & Tzen, 2008). Other than the oleosins, there are two minor groups of high molecular mass proteins in oleosomes that are called as caleosins (27 kDa) and steroleosins

(39 or 41 kDa) (Pei-Luen, Guang-Yuh, Co-Shing, & Jason, 2008). Caleosin is known to possesses the ability to bind to calcium ions within the seeds; similar to oleosins, it also plays an important role in stabilizing OBs. On the other hand, no specific role for steroleosin has been experimentally proven apart from sterol dehydrogenase activity. Moreover, number of other potential oil body proteins (OBP) has been screened and some of these proteins are waiting for further verification to confirm that they are genuine OBPs (Tzen, 2012).

2.6 Storage proteins of *Brassicaceae* oilseeds

Two major classes of storage proteins are found in *Brassicaceae* seeds; they are the 11S globulin (legumin type), cruciferin (300-350 kDa), and the 2S albumin, napin (12-16 kDa) (Crouch & Sussex, 1981; Lönnerdal & Janson, 1972). According to the values reported by Malabat, Atterby, Chaudhry, Renard, & Gu´eguen (2003), the cruciferin to napin ratio ranges from 0.6 to 2.0, especially in the European genotypes with low erucic acid and glucosinolates. The storage proteins are usually present in membrane-bound discrete deposits known as protein bodies (PBs). These protein bodies accumulate in protein storage vacuoles (PSVs) which protect stored proteins against cytoplasmic enzymes to prevent early breakdown (Müntz, 1998; Herman & Larkins, 1999; Shewry, 2000). The PSVs assume a globular shape with an approximate diameter of 1.5 to 8 µm (Ashton, 1976). Three morphologically distinct regions called the matrix, crystalloid and globoid can be identified in the PSV. The storage proteins are present in the matrix and crystalloid regions, while phytic acid crystals are present in the globoid region (Weber & Neumann, 1980; Lott, 1980). Numerous small globoids are usually dispersed inside the PSVs of *Brassica* seeds while crystalloids are slightly visible (Kuang, Xiao, McClure & Musgrave, 2000; Gillespie, Rogers, Deery, Dupree, & Rogers, 2005). According to Murphy, Cummins, & Kang (1989), *Brassica* seeds may also contain oleosin up to ~20% of total protein. Some minor proteins, such as rapeseed trypsin inhibitors (18-19 kDa), non-specific lipid transfer protein (~20 kDa) (Østergaard, Højrup, & Knudsen, 1995) and Ca^{2+}dependant–calmodulin binding proteins (Neumann, Condron & Polya, 1996), are also present in *Brassicaceae* seeds.

2.6.1. Cruciferin

Cruciferin, which is the predominant 11S protein in *Brassicaceae*, is categorized within the cupin protein superfamily. Cruciferin has properly arranged primary, secondary, tertiary and quaternary structure levels similar to other 11S seed proteins. The quaternary structure of mature

cruciferin is made of six subunits or protomers. Each protomer is comprised of two polypeptides, an α- (~40 kDa) and a β- (~20) kDa chain that are linked by a disulfide bond. (Shewry, 2000). Cruciferin-like 11S globulins are synthesized on ribosomes attached to endoplasmic reticulum (ER) membrane. They first form as a single precursor called preproglobulin containing an N-terminal signal peptide. The signal peptide is co-transitionally cleaved from pre-proglobulin to form 8-9S trimeric proglobulin (*i.e.* procruciferin) within the ER (Adachi et al., 2003). This trimeric proglobulin is transferred to the PSV via Golgi-dependent or Golgi-independent pathways (Shewry, 2000; Withana-Gamage, 2013) where further proteolytic cleavage at Arg-Gly bond results in acidic (α) and basic (β) chains linked via a single disulfide bond. The mature trimer then assembles into the 11S hexameric form where cleavage of Arg-Gly plays an important role as the trigger point (Dickinson, Hussein & Nielsen, 1998). Post-transitional modifications, such as glycosylation and phosphorylation, sometimes take place during the biosynthesis of storage proteins. Although cruciferin is not glycosylated, it is usually highly phosphorylated (Simon, Tenbarge, Scofield, Finkelstein, & Crouch, 1985; Wan, Ross, Yang, Hegedus, & Kermode, 2007). Biosynthesis of cruciferin is regulated by multiple genes in *Brassicaceae* species. In canola, nine to twelve genes are responsible for cruciferin biosynthesis while four genes eexpress cruciferin in *A. thaliana* cv. Colombia (Wanasundara, 2011; Withana-Gamage, 2013).

As discussed earlier, cruciferin has a hexameric structure composed of six subunits. These subunits may not be the samae since multiple genes are involved in expressing the protein. Canola cruciferin is a collection of protein made up of as many as five different subunits; namely CRU1, CRU2, CRU3, CRUA and CRU5 (Sjödahl, Rödin, & Rask, 1991; Wanasundara, 2011). It is only three different subunits *i.e.* CRUA, CRUB and CRUC that collectively form cruciferin in Arabidopsis. Since the genes responsible for encoding these subunits are different, the primary structure of the subunits has varying number of amino acid residues. In canola, this number ranges from 465 to 509 (Figure 2.3), while in Arabidopsis from 472 to 524 (Wanasundara, 2011; Withana-Gamage, 2013). The α-chain of a subunit in canola contains approximately 254 to 296 amino acids and has a molecular mass of ~30 kDa. Similarly, the β-chain contains approximately 189 to 191 amino acids with a molecular mass of ~20 kDa (Dalgalarrondo, Robin, & Azanza, 1986). Although the α- and β- chains of each subunit vary depending on the number and type of amino acid residues, there is some degree of conservation in particular areas of the sequence (Figure 2.3) (Wanasundara, 2011).

Figure 2.3. Primary structure of canola (*B. napus*) cruciferin subunits (CRU1, CRU2, CRU3, CRU4 and CRUA) showing multiple sequence alignment. The sequence analysis was performed using the multiple sequence alignment option available at *www.uniport.org*. "∗" residues are identical in all sequences, ":" indicates conserved substitution and "." indicates semi-conserved substitutions. Dashes appear for the gaps created for the best alignment. The sequence position of the final residue of the each line of the protein is indicated by the number on the right of the alignment. Pxxxxx in the left alignment indicates the protein identification number in the data base.

Amino acid composition of the α- and β-chains of Arabidopsis cruciferin is also analogous to that of canola. The number of amino acids in the Arabidopsis α-chain ranges approximately from 245 to 310 and in the β-chain it is 186 to 191.

Crystal structure of a protein obtained from X-ray diffraction analysis provide insight into its higher order structural organization. Due to the difficulties associated with obtaining crystals suitable for X-ray diffraction analysis, limited information is available about native cruciferin other than its primary structure. The crystal structural model of rapeseed procruciferin developed by Tandang-Silvas et al. (2010) showed that it possesses around 25 to 27 β-sheets, 7 α-helices and 3 to 4 3_{10}-helices similar to that of A3B4 protomer in soybean glycinin. The A3B4 protomer in soybean glycinin is composed of 27 β-sheets and 7 α-helices folded into two jellyroll β-barrel domains and two extended α-helix domains (Adachi et al., 2003). Figure 2.4A illustrate the secondary structural organization of cruciferin. The tertiary structure of cruciferin is presumed to have the β-chains of the polypeptide buried within the molecule in contrast to α-chains that are exposed more to the solvent environment (Job, Rajjou, Lovingny, Belghazi, & Job, 2005). The quaternary structure of cruciferin is a hexamer and it is made of two trimers. The size of a trimer is around 95×95×40 Å and subunits have a head-to-tail non-covalent orientation (Figure 2.4B; Tandang-Silvas et al., 2010). Each trimer contains IE and IA faces, which refers to the face containing inter-chain disulfide bonds and the face containing intra-chain disulfide bonds, respectively (Figure 2.4A; Adachi et al., 2003). These disulfide bonds play a key role in cruciferin structure. Especially, the trimers having no inter-chain disulfide bonds are not capable in constructing a hexamer (Jung et al., 1997) because two trimers should piled up together via IE face-to-face to form a hexamer (Figure 2.4C). The bonds associated with assembling two trimers together are predominantly non-covalent bonds, such as hydrophobic, electrostatic, hydrogen, van der Walls and hydrogen-bonded salt bridges (Adachi et al., 2003).

2.6.2 Napin

Napin, is the major 2S (sometimes named as 1.7S) protein in *Brassica* seed, and is classified under prolamin superfamily. The mature structure of napin comprises a small/short (~4 kDa) and a large/long (~9 kDa) polypeptide chains (Shewry et al., 1995). Napin synthesis initiates on membrane-bound ribosomes and the prepronapin precursor is first to form in this process. The preproprotein is then translocated into the lumen of ER where the signal peptide is detached to form pronapin with intra-chain disulfide bonds (Ericson et al., 1986).

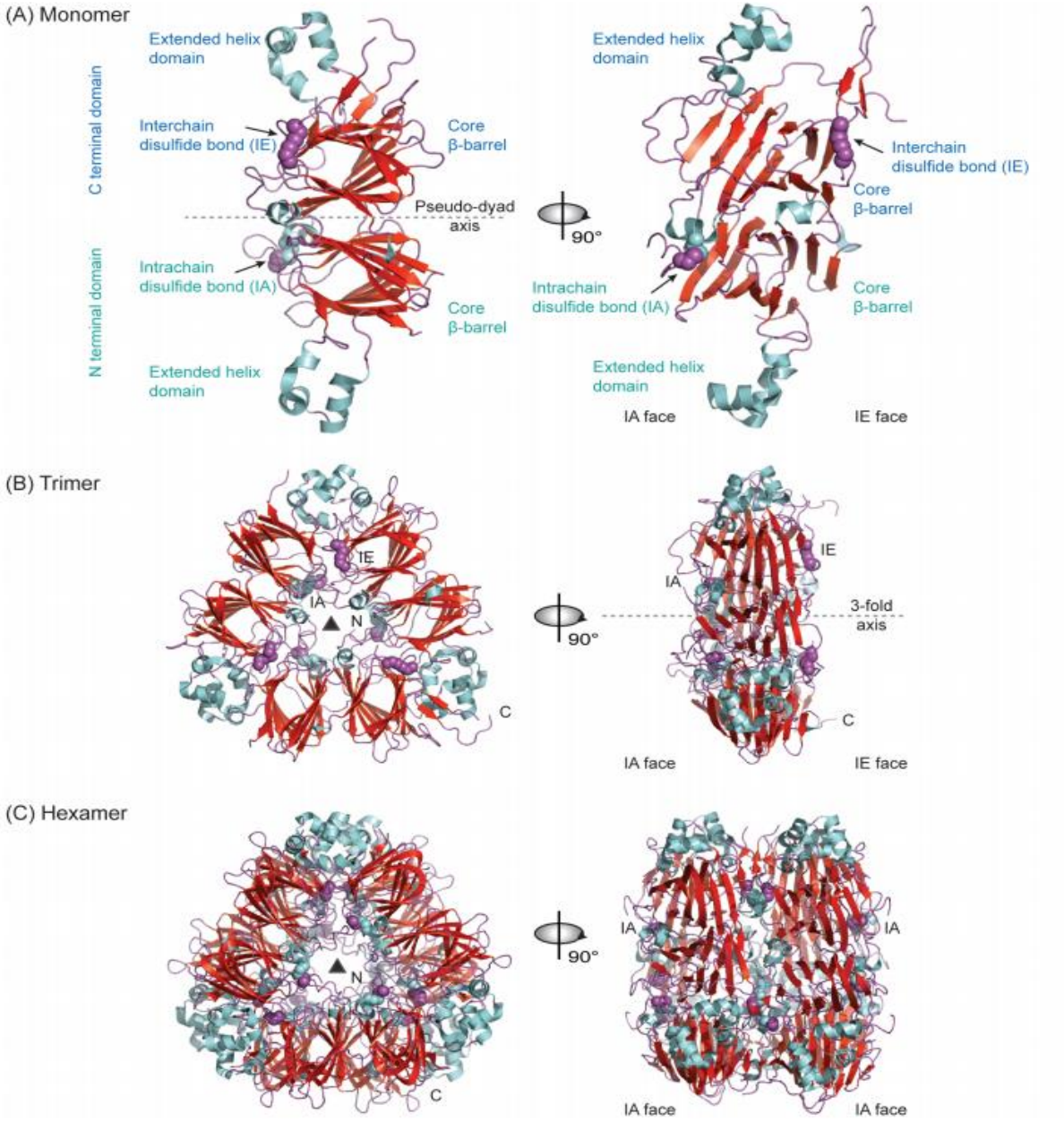

Figure 2.4. Ribbon structure of an 11S globulin molecule: (A) Monomer, (B) Trimer, and (C) Hexamer. A & B structure models are deduced from *B. napus* procruciferin (PDB code 3KGL) and C is from almond (*Prunus dulcis*) pru du amandin (PDB code 3EHK). Ribbon diagram (A) shows the secondary structural details. The filled triangle in the trimer and hexamer indicates the three-fold molecular axis. Adapted from Withana-Gamage (2013) with permission.

The pronapin is later transported to the PSV through the secretory pathway after formation of intra-chain molecular disulfide bonds. Inside the PSV, pronapin undergoes a series of proteolytic cleavages to resulting in mature napin with two subunits, *i.e.* the short chain and long chain (Ericson et al., 1986; Murén, Ek, Björk, & Rask, 1996). The two polypeptide chains are linked together by two inter-chain disulfide bonds; apart, the large chain possesses two intra-chain disulfide bonds between cysteine residues (Rico, Bruix, González, Monsalve, & Rodríguez, 1996), collectively making four disulfide bridges. Expression of napin is regulated by multiple genes in *Brassicaceae* species similar to that of cruciferin. In *A. thaliana*, four genes are responsible in encoding napin while in *B. napus* this number ranges from ten to sixteen (Scofield & Crouch; 1987; Raynal, Depigny, Grellet, & Delseny, 1992).

Several isoforms of napin can be found in each *Brassicaceae* species due to the involvement of multiple genes in expression of the protein. In *B. napus*, six different isoforms, namely Napin-1, Napin-2, Napin-3, Napin-1A, Napin-B and Nap1 have been reported and are available in protein databases such as UniProtKB (http://www.uniprot.org/). Multiple sequence alignment of canola napin isoforms reveals a number of identical and conserved regions (Figure 2.5) showing a high degree of homology between their primary structure.

Schmidt et al. (2004) revealed that the secondary structure of *B. napus* napin contains more α-helices (~48.6 to 59%) and fewer β-sheets (7 to 15%) over wide range of pH (3 to 12) in a mixture of napin isomers in which Napin-3 is predominant. The three dimensional structure of napin (Napin-1A) shows four helical motifs with loop regions that exhibit a simple "up and down" topology. The short chain creates a split helix which is an assembly of two short helices (*i.e.* HIa and HIb) distinct by few amino acid residues. On the other hand, the large chain forms three significant helices, namely HII, HIII and HIV. All the four helices assume right-handed superhelix structure (Figure 2.6; Rico, Bruix, Gonzalez, Monsalve, & Rodriguez, 1996).

2.7 Summary

Increasing interest on utilizing biodiesel has promoted *Brassicaceae* oilseed crops in the industrial oilseed platform. Among a number of *Brassicaceae* oilseed crops, *C. sativa* appears to be a sustainable crop for the Canadian prairies due its unique agronomic characteristics, genetic traits and seed chemical composition. In this context, it is essential to develop various applications for

```
P01091|2SS1_BRANA ----------------------------------------------PKCRKEFQQAQHLKAC 16
P01090|2SS2_BRANA MANKLFLVSATLAFFFLLTNASIYRTVVEFDEDDATDSAGPFRIPKCRKEFQQAQHLRAC 60
P80208|2SS3_BRANA -----------------------------------SAGPFRIPKCRKEFQQAQHLRAC 23
P17333|2SS4_BRANA MANKLFLVSATLAFFFLLTNASIYRTIVEVDEDDATNPAGPFRIPKCRKEFQQAQHLKAC 60
P24565|2SSI_BRANA ------------------------------------------QPQKCQREFQQEQHLRAC 18
P27740|2SSB_BRANA MANKLFLVSATLAFFFLLTNASIYRTVVEFDEDDATNPAGPFRIPKCRKEFQQAQHLKAC 60
                                                          **::**** ***:**
P01091|2SS1_BRANA QQWLHKQAMQSGGGPSWTLDGEFDFEDDMEK-QGPQQRPPLHQQYCNELQQEEPLCVCPT 75
P01090|2SS2_BRANA QQWLHKQAMQSGGGPSWTLDGEFDFEDDMENPQGPQQRPPLLQQCCNELHQEEPLCVCPT 120
P80208|2SS3_BRANA QQWLHKQAMQSGSGP-----------------QGPQQRPPLLQQCCNELHQEEPLCVCPT 66
P17333|2SS4_BRANA QQWLHKQAMQSGSGPSWTLDGEFDFEDDMENPQGPQQRPPLLQQCCNELHQEEPLCVCPT 120
P24565|2SSI_BRANA QQWIRQQLAGSP------------------FQSGPQEGPWLREQCCNELYQEDQVCVCPT 60
P27740|2SSB_BRANA QQWLHKQAMQSGSGPSWTLDGEFDFEDDMENPQGPQQRPPLLQQCCNELHQEEPLCVCPT 120
                  ***:::*   *                    .***: * * :* **** **: :*****
P01091|2SS1_BRANA LRGASKAVKQQIQQQEQQQGK--QQMVNRIYQTATHLPKVCNIPQVSVCPFQKTMPGPSY 133
P01090|2SS2_BRANA LKGASKAVKQQIQQQGQQQGK--QQMVSRIYQTATHLPKVCNIPQVSVCPFQKTMPGPSY 178
P80208|2SS3_BRANA LKGASRAVKQQVRQQQGQQGQQLQQVISRIYQTATHLPKVCNIPQVSVCPFQKTMPGPS- 125
P17333|2SS4_BRANA LKGASKAVKQQVRQQQGQQGQQLQQVISRIYQTATHLPKVCNIPQVSVCPFQKTMPGPSY 180
P24565|2SSI_BRANA LKQAAKSVRVQ--------GQHGPFQSTRIYQIAKNLPNVCNMKQIGTCPFIAIPFFP-- 110
P27740|2SSB_BRANA LKGASKAVKQQIQQQGQQQGK--LQMVSRIYQTATHLPKVCKIPQVSVCPFQKTMPGPSY 178
                  *: *:::*: *        *:      .**** *..**:**:: *:..***      *
```

Figure 2.5. Primary structure of *B. napus* napin isoforms (2SS1, 2SS2, 2SS3, 2SS4, 2SSI and 2SSB) showing multiple sequence alignment. The sequence analysis was performed using the multiple sequence alignment option available at *www.uniport.org*. "∗" residues are identical in all sequences, ":" indicates conserved substitution and "." indicates semi-conserved substitutions. Dashes appear for the gaps created for the best alignment. The sequence position of the final residue of the each line of the protein is indicated by the number on the right of the alignment. Pxxxxx in the left alignment indicates the protein identification number in the database.

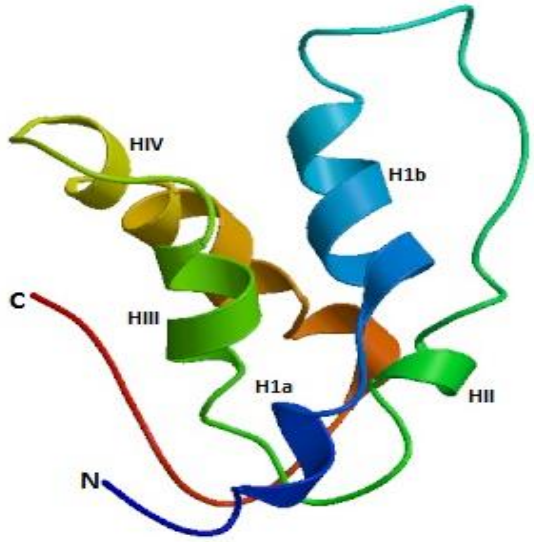

Figure 2.6. A ribbon diagram (secondary structure model) of napin from *B. napus* napin-1A (P24565, 2SSI_BRANA). C and N represents the COOH terminus and the NH_2 terminus, respectively. HIa and HIb represent helix I, whearas HII, HIII and HIV represent Helices II, III and IV, respectively. The P24565 primary sequence was modelled into the 1 sm 7A template with 93% identity using SWISS-MODEL (http://swissmodel.expasy.org/) to generate the schematic ribbon representation of *B. napus* napin.

camelina meal, which is the by-product of the oil extraction process, to enhance its economic importance. Camelina meal is a rich source of protein; therefore, it has a potential to be utilized as an animal feed. Camelina meal has gained approval in the US to be incorporated in feed rations of beef cattle, growing swine, broiler chicken and laying hen, whereas in Canada it is only for broiler chicken. However, the presence of high levels of glucosinolates restricts camelina meal inclusion levels in animal feed rations. To be utilized as a protein source for animal feed and beyond, it is important to obtain a proper understanding of the different types of proteins present in camelina meal and their properties. The 11S globulin (cruciferin) and 2S albumin (napin) are the two major types of seed storage proteins found in seeds of *Brassica* spp. and it can be expected that camelina also possesses the same. In addition, camelina seed may also contains OBPs that may provide potential economic benefits to this emerging oilseed crop. Information available on these camelina seed proteins is limited. Obtaining proper knowledge on protein structure and physicochemical properties is important to predict their functionality for subsequent utilization in suitable industrial applications.

3. MATERIALS AND METHODS

3.1. Preparation of materials

3.1.1 Seed propagation

Camelina sativa (L.) Crantz seed of double haploid line, DH55, used in this study was from a seed increase under greenhouse conditions at Agriculture and Agri-Food Canada (AAFC), Saskatoon Research and Development Centre, Saskatoon, SK. Two seed increases were carried out in separate instances and considered as biological replicates. The first seed increase was from plants grown during November 2012 to February 2013 (winter grown) and the second increase was from plants grown during April 2013 to July 2013 (spring-summer grown). The day and night temperatures of the greenhouse were maintained at 18-22°C and 16-20°C, respectively, during the winter and at 15-20°C and 10-20°C, respectively, during the spring-summer period. The relative humidity varied between 45 and 95% during the growing period and the light intensity in the greenhouse ranged from 200-650 $\mu molm^{-2}s^{-1}$ (MQ-200 Quantum Meter, Apogee Instruments, UT, USA) from November 2012 to July 2013.

Four seeds were planted in a one-gallon pot containing a soil-less mix (moistened) developed by AAFC as the growing medium. Watering was initiated after seedling emergence and an ample supply of water was provided during the growing period until two weeks prior to harvesting. Plants were fertilized from the 5^{th} to 9^{th} weeks after planting using the N:P:K 20:20:20 mix (Plant-Prod®20-20-20 classic, Master Plant-Prod Inc., ON, Canada) at a concentration of 3 g/L. Harvesting was done between 85 to 100 days after planting when plants showed specific harvesting indices. Bunches of camelina balls were hand-picked, threshed and air classified to obtain seeds with less inert material. Later, seeds were stored in a cold room at 4°C during the period of analysis.

Greenhouse-grown canola (*Brassica napus*, double haploid line-DH12075) produced at AAFC was used as the reference material for this study. Seeds were stored in a cold room at 4°C during the period of analysis.

3.1.2 Low-mucilage seed preparation

Based on preliminary experiments, Viscozyme® (Sigma-Aldrich Canada Co., ON, Canada) at a concentration of 0.1 mL/g of seed of was found suitable to reduce seed coat mucilage content of intact camelina seed. First, whole seed was mixed with Milli Q water at a seed-to-liquid ratio of 1:10 (w/v) and the pH of the medium was adjusted to 4.5 with 0.5 M HCl to provide the optimum pH for Viscozyme® activity. The required amount of enzyme then was added and the mixture was stirred for 3h (found to be the optimum incubation time from preliminary studies) at 40°C using an overhead stirrer set at 850 rpm. After stirring, the seed slurry was filtered using a #25 sieve (710 µm mesh size) and washed five times with water to remove residual enzyme. Recovered wet seeds were dried overnight in a forced air oven at 35°C.

3.1.3 Seed meal

Mucilage-reduced seed was extracted with *n*-hexane to remove oil, which was carried out using Swedish tubes with steel balls, similar to oil content determination (section 3.3.3.1). The residue remaining from oil extraction was collected as mucilage and fat-free meal.

3.1.4 Cruciferin and napin

3.1.4.1 Preparation of protein extracts

Assuming the storage proteins of camelina were similar to those of other *Brassica* oilseeds, isolation and purification of cruciferin and napin from camelina seed meal were performed according to the chromatographic separation procedures described by Bérot, Compoint, Larré, Malabat, & Guéguen (2005) with the modification adapted by Wanasundara, Abeysekara, McIntosh, & Falk (2012).

C. sativa and *B. napus* meals were extracted with 50 mM Tris–HCl buffer (containing 750 mM NaCl, 5 mM EDTA and 28 mM sodium bisulfite at pH 8.5) at ambient temperature with a meal-to-solvent ratio of 1:10 (w/v) for 1 h followed by centrifugation at 15 000 × *g* for 10 min. The supernatant was recovered and the remaining pellet was re-extracted under the same conditions. The supernatants were combined and filtered through Whatman #1 filter paper to remove any floating particles. Most of the time fresh extracts were used for protein purification unless they were stored at -20°C. The total nitrogen content of the extracts were determined and used for estimating protein content.

3.1.4.2 Isolation and purification of cruciferin and napin

The protein extract obtained as described above is a mixture of proteins and other components that are soluble under the conditions provided. The first step of the process was to isolate proteins from other components using size exclusion chromatography. The extracts were passed through a Sephadex G-25 Hiprep^{TM} 26/10 desalting column (GE Healthcare Life Science, ON, Canada; mobile phase: 50 mM Tris–HCl, pH 8.5, 1 M NaCl). The resulting protein-containing fraction (identified according to absorbance at 280 nm and SDS-PAGE separation as described in the section 3.3.5.3.2) was then dialyzed using a 2 kDa molecular mass cut-off membrane against deionized water for 48 h at 4°C and then lyophilized. Separation of cruciferin and napin was performed using a cation exchange column (CEC; Resource S column, GE Healthcare Life Science, ON, Canada; mobile phase A: 50 mM Tris–HCl, 5 mM EDTA, 0.3% w/v $NaHSO_3$ at pH 8.5; mobile phase B: 50 mM Tris–HCl containing , 5 mM EDTA, 0.3% w/v $NaHSO_3$, pH 8.5, 1 M NaCl). The unbound protein fraction (cruciferin), which eluted first from the CEC, was separated by a size exclusion chromatography (SEC) column (Sephacryl S-300 Hiprep^{TM} 26/10 high-resolution column, GE Healthcare Life Science, ON, Canada; mobile phase: 50 mM Tris–HCl, pH 8.5, 1 M NaCl) for further purification. The bound protein fraction (napin) from the CEC was then eluted with a NaCl gradient (5 to 35%) and the collected protein was further purified by hydrophobic interaction chromatography (HIC) with a HiTrap Phenyl Sepharose^{TM} 6 Fast Flow column (GE Healthcare Life Science, ON, Canada; mobile phase A: 50 mM Tris–HCl, pH 8.5, mobile phase B: 50 mM Tris–HCl, pH 8.5, 0.85 M Na_2SO_4). The salt in the napin fraction was removed by passing through a Sephadex G-25 Hiprep^{TM} 26/10 desalting column (GE Healthcare Life Science, ON, Canada; mobile phase: Milli Q water). After these purification steps, the resulting cruciferin and napin fractions were dialyzed separately as described above, freeze dried, and stored at -20°C until further use. All the chromatographic separation steps described here were carried out using an ÄKTA Explorer system (Amersham Pharmacia, Uppsala, Sweden) and the elution of protein was monitored by UV absorbance at 280 nm. The proteins in each UV absorbance peak were assessed using SDS-PAGE (section 3.3.5.3.2) to confirm the identity and the purity. The total nitrogen content of isolated protein was determined as in section 3. 3. 5.1.

3.1.4.3 Isolation and purification of napin at low pH

In addition to chromatographic separation, napin was also obtained by low pH extraction as described by Wanasundara & McIntosh (2013) combined with hydrophobic interaction

chromatography. Briefly, napin extraction at pH 3 was performed using meal and Milli Q water at 1:13.5 (w/v) ratio for 50 min while maintaining the pH constant. The protein extract was recovered by centrifuging at 4000 × *g* for 10 min and the supernatant was vacuum-filtered with two Whatman #4 filter papers. The meal was re-extracted at pH 3 at a 1:7 (w/v) meal-to-water ratio with 0.15 M NaCl in the medium and the protein extract was recovered as before and combined with the first. The protein extract was then separated using a 5 kDa molecular mass cut-off membrane and diafiltered to remove salts until the chloride ion concentration of the filtrate was <100 μs/cm. The retentate of the membrane filtration was collected and then freeze-dried. Reconstituted protein (using Milli Q water) was passed through a HIC column as explained in the previous section.

3.1.5 Oil bodies and oil body proteins

Oil bodies were first separated from *C. sativa* and *B. napus* seed and then the proteins on oil bodies (OB) were isolated. Separation of oil bodies was according to Maure et al. (2013) with a few modifications. Seeds were soaked overnight in Milli Q water at 4°C and then ground using a homogenizer (Polytron PT3100 equipped with Generator: PTDA 3020/2, Kinematica Inc., NY, USA) at 10200 rpm for 1 min. The extract was filtered through three layers of cheese cloth to obtain a filtrate devoid of seed particles. Grinding and filtering was repeated two times on the residue retained on the cheese cloth. Filtrates through the cheese cloth were combined. After adding solid sucrose to make a 25% (w/w) sucrose concentration in the filtrate, the pH was adjusted to 11 and centrifuged at 5000 × *g* for 30 min to separate OB as a cream layer that could be seen by the naked eye. The cream layer was separated using a spatula and mixed with 20% (w/v) sucrose solution at pH 11 and then centrifuged to obtain a further cleaned OB layer. The cleaned OB layer was dialyzed against water using a 2 kDa molecular mass cut-off membrane for 24 h at 4°C to remove sucrose. Following centrifugation of the dialysis tube contents, the OB layer was separated and mixed with 1% SDS (w/v) solution at a 10:1 (v/v) OB layer to SDS solution ratio. The mixture was then centrifuged at 7500 × *g* for 30 min to separate oil containing layer from the aqueous medium. The aqueous medium containing separated OB proteins was dialyzed against water using a 2 kDa molecular mass cut-off dialysis tube at room temperature for 24 h. The OB protein fraction was then desalted on the ÄKTA Explorer system by reconstituting the freeze-dried OB protein isolate in Milli Q water and using a HiPrep 26/10 Desalting column (GE Healthcare Life Science, ON, Canada) and water as mobile phase. The fractions associated with excluded protein peaks

(according to UV detection at 280 nm) were collected and freeze-dried. SDS-PAGE and 2D-electrophoresis was performed as described in sections 3.3.4.3.2 and 3.3.4.3.3, respectively.

3.2 Microscopy

3.2.1 Light microscopy

Light microscopy was used to evaluate *C. sativa* seed after Viscozyme® pre-treatment (section 3.1.2). Aliquots of dry seed, both Viscozyme®-treated and untreated, were separately soaked in Milli Q water (1:10, w/v) for 1 h. They were then observed under a digital dissecting microscope (Nikon SMZ 1500, Nikon Canada Inc., ON, Canada) to visualize the swollen seed coat.

3.2.2. Transmission electron microscopy (TEM)

Cut seed of *C. sativa* and *B. napus* were fixed in 3% (v/v) glutaraldehyde in 0.1 M sodium sacadolyte (NaCAC) buffer at pH 7.2, overnight. After rinsing several times with the same buffer, they were fixed in 1% OsO_4 (in 0.1 M NaCAC buffer) for 1 h at ambient temperature (22°C). Samples were rinsed with water, dehydrated using a graded ethanol series, and *enbloc* stained with uranyl acetate as follows: 10 min in 50% (v/v) ethanol, 1 h in saturated uranyl acetate in 70% (v/v) ethanol, 10 min in 95% (v/v) ethanol, and finally three times in 100% ethanol for 30 min. They were then rinsed again three times with propylene oxide for 30 min and subsequently infiltrated with Epon/Araldite (1:2 v/v Epon/Araldite-to-propylene oxide for 30 min, 2:1 v/v Epon/Araldite-to-propylene oxide for 3 h and pure Epon/Araldite overnight). Samples were placed in moulds and fresh Epon/Araldite was added. The samples were then polymerized at 60°C for 24 h and 150 nm sections were obtained using a Reichert-Jung ultra-microtome (Leica Microsystems Inc., ON, Canada). Sections were mounted on a 200 mesh copper grid and imaging was performed using a Hitachi HT7700 transmission electron microscope (Hitachi-High Technologies, Canada Inc., ON, Canada) under 5000 magnification, high contrast field and 80 kV. The images were analyzed using ImageJ Fiji software (Life-Line version, 2014) to calculate oil droplet diameter and number of oil droplets per unit area.

3.3 Chemical analysis

3.3.1 Oil content

The oil content of *C. sativa* and *B. napus* seed were determined according to the modified Swedish tube method of the AOCS (AM 2-93; AOCS, 1997).

3.3.2 Moisture and ash

The moisture content of the meals as well as the seed was measured according to AOAC Official Methods 934.01 (AOAC, 2005a). The ash content of the meal was determined according to AOAC Method 942.05 (AOAC, 2005b).

3.3.3 Total phenolics

The total phenolic content of *C. sativa* and *B. napus* meals was quantified according to the method described by Oomah, Corbe & Balasubramanian (2010) with modifications. The modifications include using 200 mg of meal in 6 mL of 70% (v/v) acetone (1:30 w/v ratio), centrifugation at 10000 rpm for 20 min to recover the supernatant and measuring the absorbance at 326 nm in a microplate reader (Bio-Rad xMark^{TM} Microplate Spectrophotometer, Bio-Rad Laboratories (Canada) Ltd, ON, Canada) to obtain better absorbance for sinapic acid standards and to minimize the interference of acetone. A sinapic acid (Sigma-Aldrich Canada Ltd, ON, Canada) standard curve (0-50 µg/mL) was developed to calculate the total phenolic content and was expressed as mg sinapic acid equivalents/g of meal.

3.3.4 Phytic acid

The phytic acid content of camelina meal was determined according to the method of by Oomah, Blanchard, & Balasubramanian (2008) with modifications. The modifications included utilizing 2% (v/v) HCl to extract phytates from the meals, centrifuging the extracts at 1000 rpm for 20 min, equilibrating an AG-1-X8 anion exchange column (Bio-Rad Laboratories (Canada) Ltd, ON, Canada) using 0.08% (v/v) HCl prior to addition of the extract to the column and mixing 150 µL of the eluate with 50 µL of Wade reagent in the well of a 96-well microplate. Absorbance of the salicylate-Fe (III) complex was monitored at 500 nm using a microplate reader (Bio-Rad xMark^{TM} Microplate Spectrophotometer, Bio-Rad Laboratories (Canada) Ltd, ON, Canada). The concentration of phytic acid was calculated using a standard curve developed with (0-50 µg/mL) sodium phytate (Sigma-Aldrich Canada Ltd, ON,, Canada) and expressed as percent content on as-it-is basis.

3.3.5 Analyses of protein

3.3.5.1 Nitrogen-based protein content

The total nitrogen content (combustion method) of the two meals and seeds was determined according to AOAC Method 990.03 (AOAC, 2005c). To calculate the protein content, a nitrogen-to-protein conversion factor of 6.25 was used.

3.3.5.2 Amino acid profile

The amino acid profile of camelina meal was evaluated according to AOAC Method 994.12 (AOAC, 2005d) and the process is described below.

Acid hydrolysis: Ten milligrams of meal (~5 mg protein; 0.5 mg protein/mg of meal) was added to 2 mL of 6 M HCl with 0.1 % (w/v) phenol and hydrolysed using a microwave digester (Discover SP-D, CEM Corporation, USA). Hydrolysis was performed as a temperature ramp from ambient to 195°C in 5.5 min, holding at 195°C for 10 min under a maximum pressure of 140 psi and maximum power at 300 W. Following hydrolysis and cooling, samples were neutralized with 2.85 mL of 4.2 M NaOH and 125 µL of 20 mM 2-aminobutyric acid internal standard was added and the total volume was adjusted to 10 mL with Milli Q water. A 2 mL aliquot of filtered hydrolysate passed through a 0.45 µm Phenex RC Syringe filter was applied to a C18 column cartridge (Waters Oasis HLB, 3cc, 60 mg extraction cartridges) that was equilibrated with 1.0 mL of acetonitrile followed by 2 mL of Milli Q water. Following sample introduction, hydrolysate was eluted with 2.0 mL of 5% (v/v) acetonitrile in Milli Q water. Both the flow through and the wash were collected and the total volume was adjusted to 5 mL with Milli Q water. Hydrolysed protein samples were then derivatized for major amino acids (except cysteine, methionine and tryptophan) using the method provided in the Waters AccQ-Fluor reagent kit manual.

Performic acid oxidation: Pre-treatment with performic acid converts cysteine to cystic acid and methionine to methionine sulfoxide and prevents degradation during acid hydrolysis. Ten milliliters of freshly prepared performic acid solution (9:1 formic acid: 30% (v/v) hydrogen peroxide) was added to 10 mg meal (~ 5 mg protein, 0.5 mg protein/mg of meal) in a 10 mL hydrolysis tube with a stir bar. The solution was stirred for 15 min and then placed in an ice bath maintained at 4°C for 16 h. Performic acid was decomposed with the addition of 0.085 g of sodium metabisulfite followed by stirring for 20 min. A 250 µL of sample was then transferred into a 10 mL hydrolysis tube with a Teflon liner. Performic acid-treated samples were then hydrolyzed similar to the acid hydrolysis of meal samples after adding 900 µL of 6 M HCl with 0.2 % (w/v)

phenol. Hydrolyzed samples were neutralized with 1.43 mL of 4.2 M NaOH and 50 µL of 20 mM 2-aminobutyric acid internal standard was added and the solution volume was brought up to 5 mL with Milli Q water. Approximately 2 mL of neutralized hydrolysate was filtered through a 0.45 µm Phenex RC Syringe and cleaned through a C18 column similar to the description in Acid Hydrolysis section. Samples were then derivatized before running on HPLC.

Base hydrolysis: Base hydrolysis of samples was required for tryptophan determination. Twenty milligrams of meal (~ 10 mg of protein, 0.5 mg protein/mg of meal) was hydrolyzed with 1 mL of 4.2M NaOH containing 0.1% (v/v) phenol using the same microwave digester. The hydrolysis reaction was carried out under a temperature ramp of ambient to 205°C in 5.5 min, holding at 205°C for 20 min, maximum pressure at 140 psi and maximum power at 300 W. The samples were neutralized with 0.7 mL of 6 M HCl followed by addition of 500 µL of 20 mM 5-methyl tryptophan internal standard and bringing the total volume up to 10 mL with Milli Q water. The neutralized hydrolysate was filtered similar to acid hydrolysed samples and applied to a C18 cartridge that was equilibrated with 1.0 mL of acetonitrile followed by 2 mL of Milli Q water. One mL of the filtered hydrolysate was then applied to the column followed by 5% (v/v) methanol: 5% (v/v) acetonitrile: 90% (v/v) Milli Q water. The eluted fractions were collected and brought up to the volume (1 mL) and used for analysis of tryptophan. Derivatization is not required prior to run on HPLC in this step.

Once the derivatization was complete, samples were run on an Alliance® 2695 HPLC equipped with a multi-wavelength 2475 fluorescence detector (Waters Corporation, USA) to analyze amino acids, except tryptophan (Excitation=250 nm, Emission=395 nm and Gain=15). A gradient elution was carried out using three different eluents; Waters AccQTag Eluent A buffer with 10 times dilution using Milli Q water at pH 5.2, HPLC grade acetonitrile and Milli Q water. Elution of major amino acids were carried out at a 37°C column temperature, maintaining 1.0 mL/min flow rate while it was kept at 30°C and 0.75 mL/min for cysteine and methionine. Derivatization was not required for tryptophan. Instead, 10 µL of final base hydrolysates were mixed with 70 µL of borate buffer from the AccQ-Fluor reagent kit and run on an Alliance® 2695 HPLC system (Excitation=285 nm, Emission=320 nm and Gain=15). The same eluents were used in this case as rest of the amino acids with minor modifications with the gradient. Column temperature and the flow rate was maintained at 37°C and 1.0 mL/min, respectively. The elution

time for major amino acids, cysteine/methionine and tryptophan varied from 100, 90 and 80 min, respectively.

3.3.5.3 Polypeptide profile

3.3.5.3.1 Native polyacrylamide gel electrophoresis (Native-PAGE)

Native-PAGE was performed to confirm the molecular assembly of the purified cruciferin and napin. Proteins were dissolved in 0.1 M phosphate buffer (pH 8.0) containing 0.1 M NaCl to provide 1 mg/mL and 4 mg/mL concentrations of cruciferin and napin, respectively. The samples were centrifuged at 14000 rpm for 10 min and the clear supernatant was loaded onto a 8-25% (%T) gradient gel. The electrophoresis was performed according to the Phastsystem Electrophoresis System-Operating manual. Native-PAGE buffer strips (free from SDS) were used to provide non-denaturing conditions. Non-denaturing protein standards; namely bovine serum albumin (BSA) and urease (Sigma-Aldrich Canada Ltd, ON, Canada) were used as reference molecules to determine the assembly of the purified protein.

3.3.5.3.2 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The polypeptide profile of meal and purified protein was evaluated by SDS-PAGE under non-reducing (SDS extraction buffer without β-mercaptoethanol; β-ME) and reducing (with β-ME) conditions (Laemmli, 1970) using precast 8-25% T gradient gels adapting the protocol of Wanasundara, Abeysekara, McIntosh, & Falk (2012). The samples were prepared in 1.5 mL microcentrifuge tubes using the required amount of SDS extraction buffer (5%, w/v, SDS in 0.05 M Tri-HCl buffer at pH 8). The final concentration of protein in the SDS extract was 1-2 mg/mL. For reducing conditions, the appropriate amount of β-ME was added to achieve a 5% (v/v) concentration. The samples were vigorously mixed in an Eppendorf Thermomixer at 99°C, 1300 rpm for 10 min. The samples were then brought to ambient temperature and centrifuged for 10 min at 1400 rpm. The protein extracts were loaded onto precast gels with molecular weight standards (4.6- 170 kDa, Pageruler^{TM} pre-stained protein ladder, Thermo Fisher Scientific Inc., ON, Canada) and processed according to the Phastsystem Electrophoresis System-Operating manual (Pharmacia PhastSystem Electrophoresis System, GE Healthcare Life Science, ON, Canada). Finally, the gel images were processed to obtain the molecular mass estimation of the polypeptide bands using the ImageQuant TL software (GE Healthcare Life Sciences, ON, Canada).

3.3.5.3.3 Two dimensional gel electrophoresis (2DE)

Purified cruciferin, napin and OB protein were further separated by 2DE under non-reducing conditions. A 1 mg/mL stock solution of cruciferin and napin was prepared and 50 µL of the stock solution was mixed with 100 mL of sample buffer containing 6.7 M urea, 2% (w/v) 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonic acid (CHAPS), 0.5% Bio-Lyte 3-10® (Bio-Rad Laboratories, Inc., ON, Canada) ampholytes (for cruciferin and OB protein) or Zoom® carrier ampholytes (Life Technologies, Inc.) pH 9-11 (for napin), 0.001% bromophenol blue and deionized water. For oil body proteins, the dry proteins (~100 µg) were dissolved in the same sample buffer mixture to completely solubilize the protein. After preparation of protein solutions, an IPG (Immobilized pH Gradient) strip (pH 3-10 for cruciferin and OB protein and pH 9-12 for napin) was re-hydrated in each protein solution overnight at 4°C and focused using a protein IEF cell (Bio-Rad Laboratories, Inc., ON, Canada) for 2 h at 50 V, for 2 h at a voltage gradient from 200-4000 V, and for 9 h at 4000 V. Prior to running the second dimension, the buffer strips were equilibrated for 15 min in buffer 1 (1.8 g urea, 1 mL of 10% SDS, 1.25 mL of 1.5 M Tris, pH 8.8, 1 mL of 100% glycerol and 0.6 mL of Milli Q water) followed by buffer 2 (1.8 g urea, 1 mL of 10% SDS, 1.25 mL of 1.5 M Tris, pH 8.8, 1 mL of 100 % glycerol and 0.6 mL of Milli Q water and 125 mg iodoacetamide). SDS-PAGE was then carried out using 12% T hand-cast polyacrylamide gels for cruciferin and 16 % T for napin and OB protein using Bio-Rad Mini-Protean® tetra cell (Bio-Rad Laboratories, Inc., ON, Canada). Gels were stained for 4 h using 0.1% coomassie blue R-250 staining solution. After destaining for 1-2 h, the stained protein spots were cut and subjected to LC-MS/MS analysis.

3.4. Identification and confirmation of protein

3.4.1 Mass spectroscopic analysis

LC-MS/MS is an effective proteomics tool that can be used to confirm the identity of a protein. The LC-MS/MS analysis for 2D-separated proteins was carried out at the University of Victoria (UVic) Genome BC Proteomic Centre (Victoria, BC). First, the gels were subjected to in-gel trypsin digestion and prepared for LC-MS/MS analysis according to the method described by Parker et al. (2005). LC-MS/MS was then performed on the peptide mixture as described by Senko et al. (2003). The raw files from MS analysis were created using XCalibur 3.0.63 (Thermo Fisher Scientific, ON, Canada) software and analysed with Proteome Discoverer 1.4.0.228 software suite (Thermo Fisher Scientific, ON, Canada). Results from the Proteome Discoverer were then

statistically analyzed using the Scaffold Q+S software package (Proteome Software, Inc., OR, USA). The percentage of Normalized Total Spectra (NTS) of a protein type found in an individual 2DE protein spot was used to express the protein abundancy.

3.4.2 Evolutionary relationship analysis

Phylogenetic relationships were constructed using MEGA6 software (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). The Neighbor-Joining method was used to infer the evolutionary history from the aligned amino acid sequences of the seed proteins of *C. sativa* and *A. thaliana*. The proteins were clustered together using the bootstrap method with 500 replicates and the evolutionary distances were computed using the Poisson correction method. Gaps/missing data for the amino acid sequences were handled using the partial deletion treatment with 95% of site coverage cut-off percentage.

3.5 Spectroscopy for protein structure analysis

3.5.1 Fourier transform infrared (FT-IR) spectroscopy

FT-IR spectroscopy was used to evaluate the secondary structural details (α-helix, β-sheet, β-turns and random structure) of cruciferin and napin. Briefly, purified protein powder (in dry form) was placed on the ATR diamond surface (Agilent Cary 630 ATR-FTIR analyzer, AgilentTechnologies Canada Inc., ON, Canada) and the sample was pressed against the diamond crystal using the attached pressure clamp. The FTIR spectra were recorded with 4 cm^{-1} resolution and a ~30 sec measurement time. The secondary structural details were analyzed using Agilent Resolution Pro, version 5.2.0 software (Agilent Technologies Canada Inc., ON, Canada) and the Fourier Self-Deconvolution method (FSD, Kauppinen, Moffatt, Mantsch, & Cameron, 1981) of the amide I region (1610-1700 nm) was used to quantify percentage of α-helix, β-sheet, β-turns and random structure of each protein.

3.5.2 Circular dichroism (CD) spectroscopy

The secondary structural details (α-helix, β-sheet, β-turns and random structure) of isolated camelina seed proteins were also evaluated using far-UV CD spectra. A protein solution (1 mg/mL of protein) was prepared in 10 mM sodium phosphate buffer at pH 3, 7 and 10. The far-UV spectrum of the protein solution was then obtained at 25°C using a PiStar-180 spectrometer (Applied Photophysics Ltd., Leatherhead, U.K) equipped with a mercury xenon lamp and 0.1 mm quartz cell at 180-260 nm using 6-nm entrance and exit slits. The instrument was calibrated with

0.89 mg/mL d-(+)-10-camphorsulfonic acid (CSA). Four scans per sample were averaged to obtain one spectrum and the baseline was corrected by subtracting the buffer spectrum. The background-corrected spectra were analyzed and the molar ellipticity was calculated using the CDNN 2.1 software package (Applied Photophysics Ltd., Leatherhead, U. K). Near-UV (260-320 nm) CD spectra were also obtained using the method described for far-UV spectra. In this case, samples were introduced to the PiStar-180 spectrometer using a 1 cm quartz cell.

3.5.3 Fluorescence spectroscopy

3.5.3.1 Intrinsic fluorescence

Intrinsic fluorescence of the proteins based on the emission spectra of tryptophan residues was evaluated at different pHs and temperatures. Briefly, the fluorescence emission spectra of protein solutions (50 μg/mL in buffer solution at 20°C) were recorded with an Agilent eclipse fluorescence spectrophotometer (Model G9800A, Agilent Technologies Canada Inc., ON, Canada). 10 mM sodium phosphate buffer (pH 7.4), 10 mM ammonium buffer (pH 10) and 10 mM citrate buffer (pH 3.2) were used to provide different medium pHs. The tryptophan residues of the proteins were excited at 280 nm and emissions scanned from 290-450 nm (5 nm excitation and emission bandwidth, medium PMT voltage and factor 5 smoothing using Savitzky-Golay algorithm). The emission spectrum at each pH level with increasing temperature (22 to 93°C) was recorded and analyzed.

3.5.3.2 Surface hydrophobicity

The anionic fluorescence probe, 8-anilinonaphthalene-1-sulfonic acid (ANS), was used to evaluate surface hydrophobicity of cruciferin and napin as described by Withana-Gamage (2013) with slight modifications, such as mixing 5 μL of the 8 mM ANS stock solution with 1mL of protein solution (0.05−0.25 mg/mL) at each pH level, 10 min incubation in the dark after mixing the two solutions and monitoring fluorescence of the protein-ANS conjugate using a Cary eclipse fluorescence spectrophotometer (Agilent Technologies Canada Inc., ON, Canada) at excitation wavelength 390 nm, and fluorescence emission wavelength 470 nm for cruciferin and 500 nm for napin. To obtain the net fluorescence intensity of protein-ANS conjugate, fluorescence intensity of a protein blank (without ANS) and an ANS blank (without protein) at each concentration was monitored and subtracted from those of protein-ANS conjugate.

3.6 Differential scanning calorimetry (DSC)

The thermal denaturation parameters of purified proteins (denaturation temperature and enthalpy of denaturation) were evaluated using a TA Q2000 differential scanning calorimeter (DSC) (TA Instruments, New Castle, DE, USA). Approximately 20 mg of 5 % (w/v) cruciferin and 10 mg of 10 % (w/v) napin solutions were prepared in 10 mM phosphate buffer (pH 7), 10 mM ammonium buffer (pH 10) and 10 mM citrate buffer (pH 3). They were placed into aluminum liquid pans, hermetically sealed with a Tzero™ press (TA Instruments, New Castle, DE, USA) and subjected to a 30-130°C temperature ramp at a scanning rate of 2°C/min and 5°C/min for cruciferin and napin, respectively, under constant nitrogen purging (flow of 50 mL min-1). A hermetically sealed empty pan was used as a reference and results were analyzed using TA universal analysis 2000 software (TA Instruments).

3.7 Protein solubility

3.7.1 Meal protein

It was expected that the reduced level of mucilage in meal would allow protein to become soluble. This was studied as the solubility of camelina meal protein as a function of medium pH. Camelina meal and Milli Q water were mixed at a 1:20 (w/v) ratio and extraction was performed for 30 min using a Metrohm 906 Tirando Titrator (Metrohm AG, ON, Canada) to maintain a constant medium pH (maximum allowable deviation ± 0.05 units) during the extraction. The slurry was then centrifuged at 3500 × g for 15 min and the liquid portion was filtered under vacuum through two #4 Whatman filter papers. The resulting aqueous extract was analyzed for total nitrogen (combustion based) content to obtain soluble protein percentage and polypeptide profile (SDS-PAGE) to confirm the types of proteins extracted at each pH level.

3.7.2 Purified protein (cruciferin and napin)

The solubility of cruciferin and napin as a function of medium pH was determined using a Pierce® BCA protein assay kit (Thermo Scientific, Tockford, IL, USA). First, separate absorbance correction factors were developed for cruciferin and napin using a bovine serum albumin (BSA) standard and known concentrations of cruciferin and napin (0.025 – 0.1 mg/mL) at pH 3, 7 and 10. A 1 mg aliquot of protein was mixed with 1 mL of buffer and stir for 30 min at 500 rpm. The protein samples were centrifuged at 1500 rpm for 10 min and the soluble cruciferin and napin concentration of the supernatant was determined using BSA as the standard. The absorbance of

cruciferin and napin in the supernatant was adjusted using corresponding correction factor before calculating the concentration from the BSA standard curve.

3.8 Experimental design and statistical analysis

Complete Randomized Design (CRD) was used as the experimental design for the analyses. Winter-grown and spring-summer-grown camelina seed were considered as two biological replicates. Canola seed was used as the control. All the analyses were carried out in triplicate. The results obtained were then analyzed using the General Linear Model (GLM) procedure and Tukey's test was performed as the post-hoc test for mean separation using R statistical software, version 3.2.2 (https://cran.r-project.org/).

4. RESULTS

4.1 Seed and meal composition

4.1.1 Chemical composition of the seed and meal

The average oil contents of *C. sativa* DH55 and *B. napus* DH12075 were 23.3% and 27.0%, respectively, on a dry weight basis (dwb). These values were lower than those of field-grown varieties of the two oilseed types (SMA, n.d, Mag, n.d). When oil-free meal was analysed for protein, phytate and total phenolics (Table 4.1), *C. sativa* showed significantly different ($p<0.05$) values from *B. napus*, but ash content was the same. Camelina meal had higher contents of protein and phytate, but a lower level of phenolic compounds, than did canola meal.

Table 4.1. Contents of protein, ash, phytic acid and total phenolics of *C. sativa* and *B. napus* meals (dry weight basis). Values are presented as mean ± standard error.

Meal sample	Protein, % (% N × 6.25)	Ash, %	Phytate, %	Total phenolics* %
C. sativa	51.3 ± 0.4^{a}	6.8 ± 0.1^{a}	6.1 ± 0.1^{a}	1.0 ± 0.1^{a}
B. napus	42.5 ± 0.6^{b}	6.2 ± 0.1^{a}	4.8 ± 0.1^{b}	2.7 ± 0.1^{b}

Means followed by the same superscript within the same column are not significantly different ($p>0.05$)
* As sinapic acid equivalents

4.1.2 Amino acid and polypeptide profiles

When the amino acid profiles of *C. sativa* and *B. napus* meals (Table 4.2) were compared, the predominance of glutamic acid and the abundance of leucine and lysine among the essential amino acids were common to both crucifers. The total essential amino acid content of camelina meal obtained in this study, and as reported in the literature, was approximately 40% (Table 2.2, Zubr, 2003a), which was comparable to that of canola meal reported in the literature (Newkirk, 2015, Wanasundara et al., 2015). The lysine content, one of the frequently limiting factors in these oilseeds (Russo, 2012) was ~5% in *C. sativa* meal, which was lower than in *B. napus* meal.

Table 4.2. Amino acid composition of *C. sativa* and *B. napus* meal (dry weight basis). Values are presented as mean ± standard error.

Amino acids	% of Total meal protein (w/w)	
	C. sativa	*B. napus*
Histidine (His)	2.55 ± 0.11	2.86 ± 0.03
Isoleucine (Ile)	3.73 ± 0.15	4.66 ± 0.10
Leucine (Leu)	6.46 ± 0.09	7.76 ± 0.14
Lysine (Lys)	5.55 ± 0.03	6.07 ± 0.09
Threonine (Thr)	3.63 ± 0.07	4.73 ± 0.06
Tryptophan (Trp)	1.17 ± 0.04	1.51 ± 0.03
Valine (Val)	5.29 ± 0.15	6.30 ± 0.10
Methionine (Met)	1.89 ± 0.07	2.69 ± 0.05
Cysteine (Cys)	3.12 ± 0.10	4.13 ± 0.08
Phenylalanine (Phe)	4.15 ± 0.07	4.37 ± 0.05
Tyrosine (Tyr)	3.23 ± 0.07	3.24 ± 0.04
Aspartic acid (Asp)[1]	8.83 ± 0.14	8.80 ± 0.02
Glutamic acid (Glu)[1]	18.54 ± 0.25	19.45 ± 0.46
Glycine (Gly)	5.07 ± 0.06	5.63 ± 0.01
Alanine (Ala)	4.39 ± 0.05	4.92 ± 0.03
Proline (Pro)	5.25 ± 0.07	6.17 ± 0.10
Serine (Ser)	4.31 ± 0.08	4.45 ± 0.12
Arginine (Arg)	9.61 ± 0.12	8.40 ± 0.18
Total EAA[2]	40.76	48.33
Total BCAA[3]	15.47	18.71
Total BAA[4]	17.71	17.34
Total NP[5]	32.33	38.38

[1]Includes asparagine and glutamine, respectively
[2]Total essential amino acids (EAA); ∑ His, Ile, Leu, Lys, Thr, Trp, Val, Met, Cys, Phe, Tyr
[3]Total branched chain amino acids (BCAA); ∑ Leu, Ile, Val
[4]Total basic amino acids (BAA); ∑ His, Arg, Lys
[5]Total non-polar (hydrophobic) amino acids (NP); ∑ Phe, Ala, Leu, Met, Ile, Trp, Pro, Val

Although the level of methionine was low in both *C. sativa* and *B. napus* meals compared to most of other essential amino acids, cysteine present in these meals in combination with methionine could contribute to the sulfur amino acid requirement for animal and human nutrition. The total sulfur-containing amino acid content (Met + Cys) of *C. sativa* and *B. napus* meal were 5.0% and 6.8% of protein (Table 4.2), respectively, which is higher than that of soybean meal (~1.3%, Dozier & Hess, 2011). The branched chain amino acid contents of *C. sativa* was lower than that of *B. napus* (Table 4.2), and could be potential sources for nutraceutical use for muscle growth (Shimomura et al., 2006). The total non-polar amino acid content of *C. sativa* meal protein was 32.33%, lower than that of *B. napus* meal protein (38.8%). A similar trend was observed for other lines reported in the literature (Zubr, 2003a; Newkirk, 2015).

4.1.3 Mucilage of camelina and effect of Viscozyme® pre-treatment

The seeds of *C. sativa* were smaller in size than those of *B. napus* with average dimensions of 2.22 mm in length, 1.15 mm in width and approximately 1 mm in thickness. Compared to *B. napus* seeds which were round and ~2 mm in diameter (Riethmuller, Carmody & Walton, 2003; Hellevang, n.d), camelina seeds had an elongated shape (Figure 4.1A). Soaking *C. sativa* seeds in water caused swelling of the mucilage-containing cell layer, creating a halo around the seed (Figure 4.1B). When seeds were soaked in water containing Viscozyme®, a considerable amount of mucilage was removed, to the extent that no halo was created upon rehydration of dried, treated seeds (Figure 4.1C). Seeds maintained their shape but the seed coat surface lost its smooth appearance which was clearly apparent before enzyme treatment (Figure 4.1A). Therefore, soaking and washing of camelina seeds in Viscozyme® caused a definite reduction in seed coat mucilage content and allowed mucilage-reduced (or demucilaged) seed to be obtained. *B. napus* seeds contain little mucilage were not subjected to enzymatic treatment and were used as they were.

4.1.4 Protein and polypeptide profile of meal

Analysis of the meal polypeptide profile under non-reducing conditions (Figure 4.2) showed that polypeptides ranging from 14.0 kDa to 67.1 kDa and 13.9 kDa to 55.0 kDa were present in *C. sativa* and *B. napus* seeds, respectively. The two prominent polypeptide bands in the high molecular weight range, 44.1 kDa and 51.7 kDa for *C. sativa* and 40.9 kDa and 55.0 kDa for *B. napus,* were not observed under reducing conditions (Figure 4.2), indicating the involvement of disulfide bonds in stabilizing the structure of these proteins.

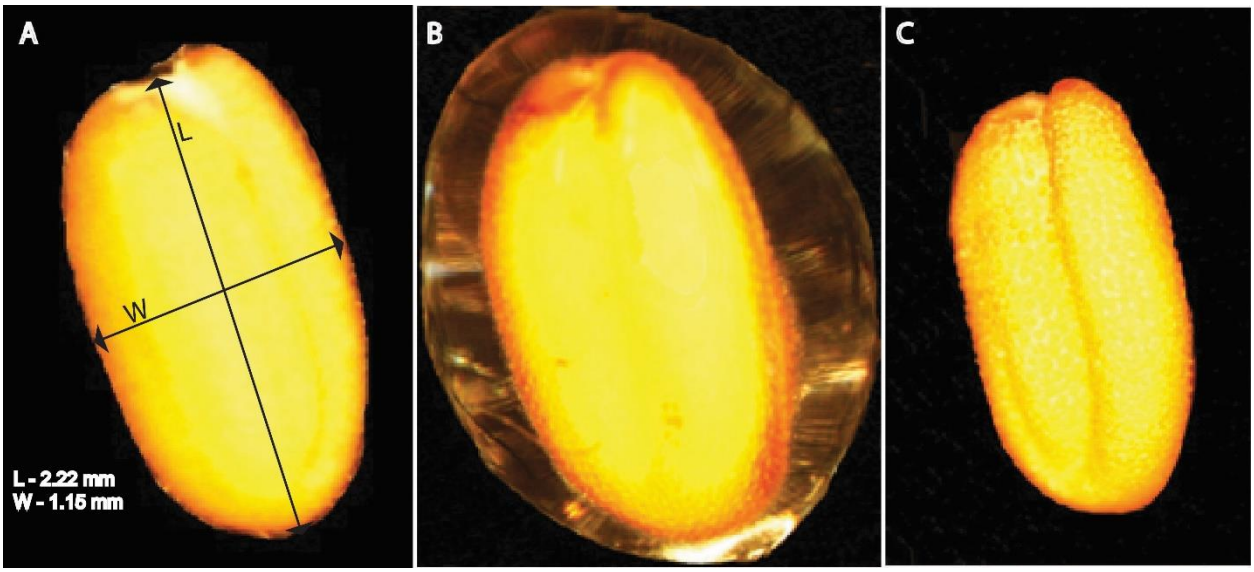

Figure 4.1. Dissecting microscopic images of *C. sativa* seed. (A) untreated dry seed (L=length and W=width); (B) an untreated seed soaked in water for 1h and (C) seed after Viscozyme® treatment, dried and soaked in water for 1h. (Images were captured using a Nikon SMZ 1500 dissecting microscope attached to a Nikon Digital sight DS-5M camera at 2 × zoom range).

In contrast to *B. napus*, *C. sativa* meal exhibited 69.3 kDa and 53.7 kDa bands that were present under both non-reducing and reducing conditions. Polypeptide bands from ~37.0 kDa to 38.5 kDa, which did not disappear due to S-S bond reduction, were present in both meals. In the ~20.0-32.0 kDa region, both *C. sativa* and *B. napus* meal showed several peptide bands under non-reducing conditions. While some polypeptide bands disappeared upon reduction of S-S bonds, while some of the peptide bands became more intensely stained. The few polypeptides that appeared between 15.0 kDa and 20.0 kDa and that were intensely stained became fewer in number when S-S bonds were broken. A distinct band found between 13.9 kDa and 14.0 kDa that was intensely stained in both meal samples under non-reducing conditions completely disappeared upon S-S bond reduction.

Under reducing conditions, *C. sativa* showed a diffuse polypeptide band at 10.4 kDa, whereas *B. napus* showed two bands at 9.6 kDa and 10.9 kDa. *C. sativa* meal samples showed four (16.6/16.4 kDa, 17.5 kDa, 18.4/18.6 kDa and 20.7 kDa), and *B. napus* three (15.4 kDa, 17.5 kDa and 20.4/20.0 kDa) polypeptide bands that did not change due to reducing conditions, indicating that these polypeptides may not contain disulfide bonds and most likely were single polypeptides.

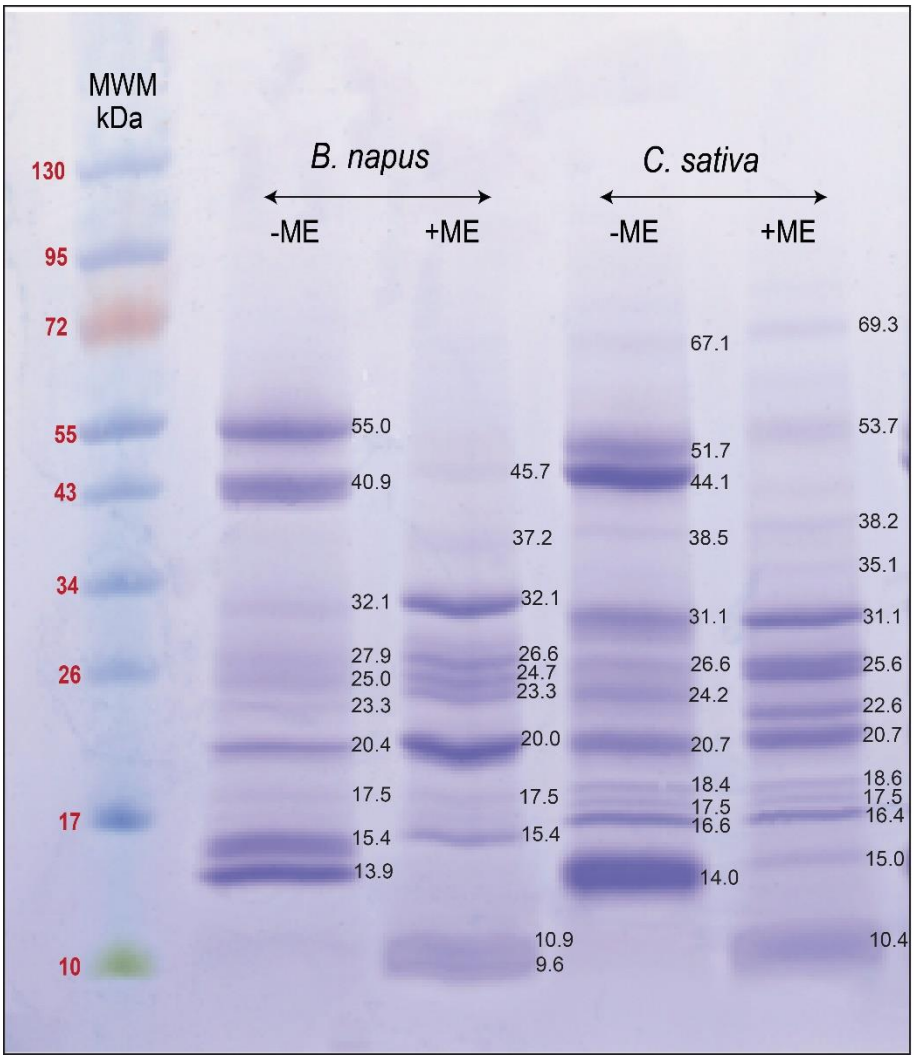

Figure 4.2. Polypeptide profiles of *C. sativa* and *B. napus* meal. Polypeptide profiles under non-reducing (-ME) and reducing (+ME) conditions were separated on an 8-25% gradient precast gel. Estimated molecular masses (kDa) of polypeptide bands are indicated. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

4.1.5 Solubility of protein in meal

The levels of soluble protein in *C. sativa* and *B. napus* meals with changing pH followed similar trends. At pH 4.5, both meals had the lowest level of soluble protein (Figure 4.3A). In a single extraction at room temperature with a meal to solvent ratio of 1:20 (w/v), 43.6% and 39.6% of the meal protein of *C. sativa* and *B. napus*, respectively, was soluble at pH 10, which was the maximum amount of soluble protein observed for these meals. In the pH range of 6.5 to 12, a higher amount of *C. sativa* protein was soluble than was the case for *B. napus;* the opposite was observed between pH 2.0 and 5.5.

When the polypeptide profiles of the soluble protein at each pH were analysed by electrophoresis (Figures 4.3B and 4.3C), it was clear that the solubility of seed protein types was pH dependant. In *C. sativa*, only low molecular weight proteins (<20 kDa) were soluble from pH 2.5 to 6.5, and in *B. napus,* the same was observed between pH 2.5 and 4.5. When the pH reached 5.5, polypeptides larger than 20 kDa (~20-59 kDa) became soluble in *B. napus*; for *C. sativa*, the pH had to be above 8.5 to solubilize these proteins. Comparison of protein levels (Figure 4.3A) with the polypeptide profiles (Figure 4.3B and C) soluble at each pH showed that low soluble protein content corresponded with fewer protein types (bands) in solution. Also, the increase in soluble protein content of camelina above pH 6.5 compared to *B. napus* corresponded with more types of protein found in the soluble fraction.

4.2 Separation and purification of storage proteins

4.2.1 Cruciferin

Separation of *C. sativa* and *B. napus* meal extracts (pH 8.5, 50 mM Tris-HCl buffer) using a size exclusion (desalting) column removed the co-extracted non-protein components, mainly pigments and small molecular weight compounds. The non-retained UV absorbing peak contained protein (Peak 1, Figure 4.4A) and the UV absorbing small molecular weight compounds eluted later. The polypeptide profiles of Peak 1 from *C. sativa* and *B. napus* were similar to those of their meals (Figure 4.4D) and confirmed that only the non-protein compounds were removed during this step. When the resultant protein (Peak 1) was separated on a cation exchange column, the unbound protein peak that eluted first (Peak 2, Figure 4.4B) contained polypeptides in the range of 20-61 kDa (Figure 4.4E).

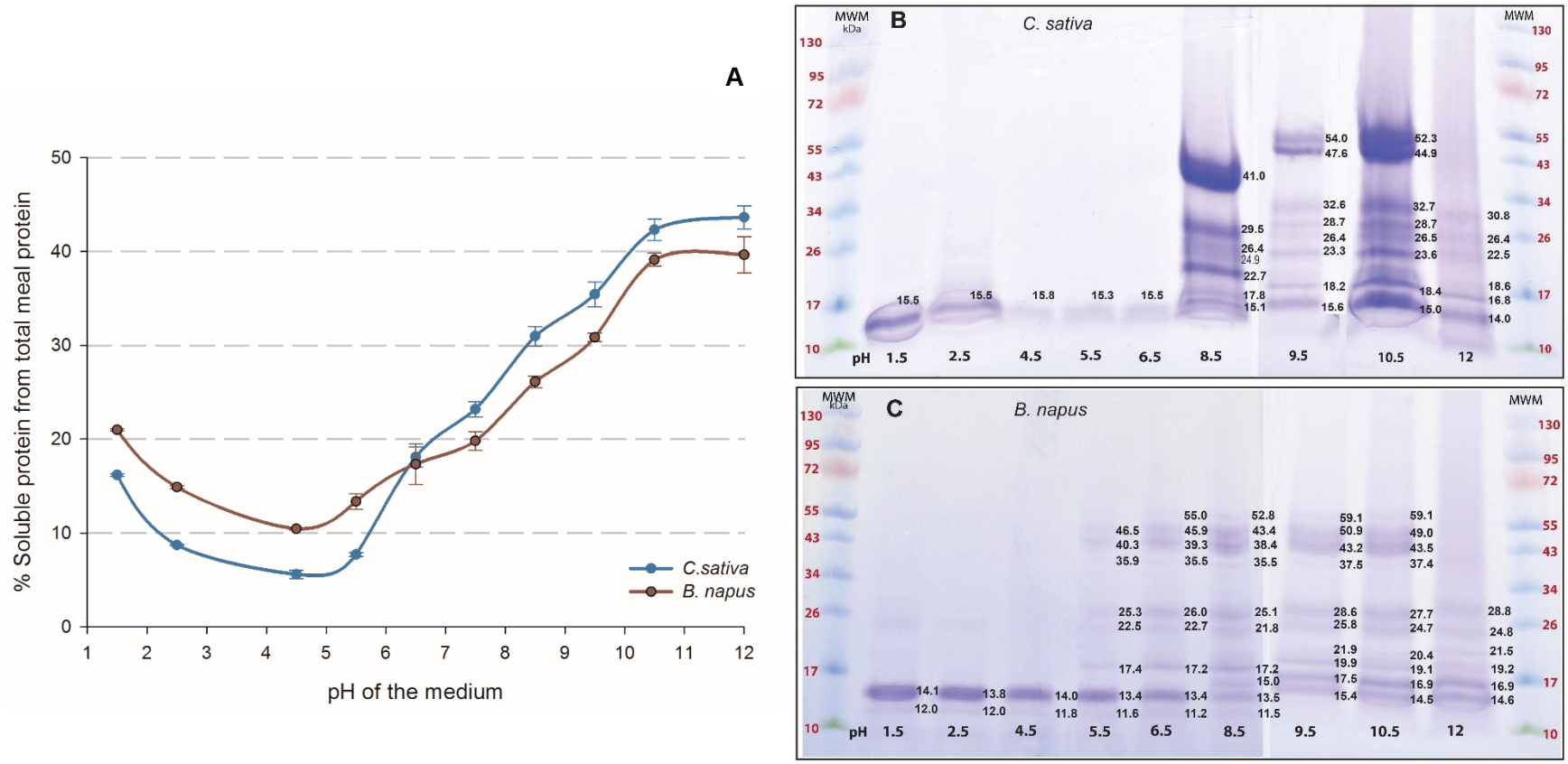

Figure 4.3. Protein solubility pattern and types of polypeptides soluble in *C. sativa* and *B. napus* meal as a function of pH change. (A) Solubility of *C. sativa* and *B. napus* meal protein (% N × 6.25) depending on the pH of the medium; (B) Polypeptide profile of soluble *C. sativa* meal protein different pH levels and (C) Polypeptide profile of soluble *B. napus* meal protein at different pH levels. Estimated molecular masses (kDa) of polypeptide bands of B & C are indicated. The polypeptide profiles are under non- reducing conditions in 8-25% precast gradient gels. (MWM=Molecular weight markers / PageRuler^{TM} Pre-stained Protein Ladder). Each lane contains same level of protein.

The other protein peak (Peak 3, Figure 4.4B) which eluted with the increasing gradient of NaCl in the mobile phase showed polypeptides of 11-17 kDa (Figure 4.4E). It can be noticed that Peak 2 (Figure 4.4E) contained some low molecular mass polypeptides. The aim of the chromatographic purification was to separate cruciferin (high molecular mass, ~55 kDa) from napin (low molecular mass, ~15 kDa). Therefore, a third chromatographic step was performed using a second size-exclusion column to remove low molecular weight polypeptides. The first protein peak eluted in this separation (Peak 4, Figure 4.4C) contained polypeptides in the range of 21-63 kDa (Figure 4.4.F). The second broad protein peak (Peak 5, Figure 4.4C) contained polypeptides of 12-17 kDa (not shown). Therefore, Peak 4, which contained high molecular weight protein, was dialysed and lyophilized to obtain cruciferin.

The polypeptide profiles of cruciferin from *C. sativa* and *B. napus* showed that the proteins were composed of polypeptides in the range of 19.0-59.7 kDa only (Figure 4.5) and they were resolved from the 40.0-59.7 kDa polypeptides when S-S bonds were broken. However, a few faint bands at ~15 kDa and ~11.0 kDa were visible in purified *C. sativa* protein (lanes 1 and 2 in Figure 4.5), whereas purified *B. napus* cruciferin showed a faint band at ~12.5 kDa (lane 4, Figure 4.5). These bands could be trace contaminants of oleosin or napin. For confirmation, 2DE and LC-MS/MS analyses were carried out (Section 4.4). The purified cruciferin contained 100% protein (%N × 6.25) confirming the absence of non-protein components.

Polypeptide bands in the range of ~20.0 kDa to 46.0 kDa under non-reducing conditions, and the ~20.0 kDa to 29.0 kDa range under reducing conditions, were identified as characteristic bands for *C. sativa* cruciferin. Similarly, polypeptide bands in the range of ~19.0 kDa to 60.0 kDa and ~19.0 kDa to 31.0 kDa under non-reducing and reducing conditions, respectively were identified as the characteristic bands for *B. napus* cruciferin. The Peak (Peak 5) obtained from the final size exclusion column contained polypeptides < 15.0 kDa that were characteristic of napin (Monsalve & Rodrigues, 1990; Wanasundara, 2011) (data not shown). Since the amount of protein obtained in these peaks were quite low, further purification of Peak 5 was not performed. A separate napin extraction and purification process was performed to obtain an adequate amount of napin.

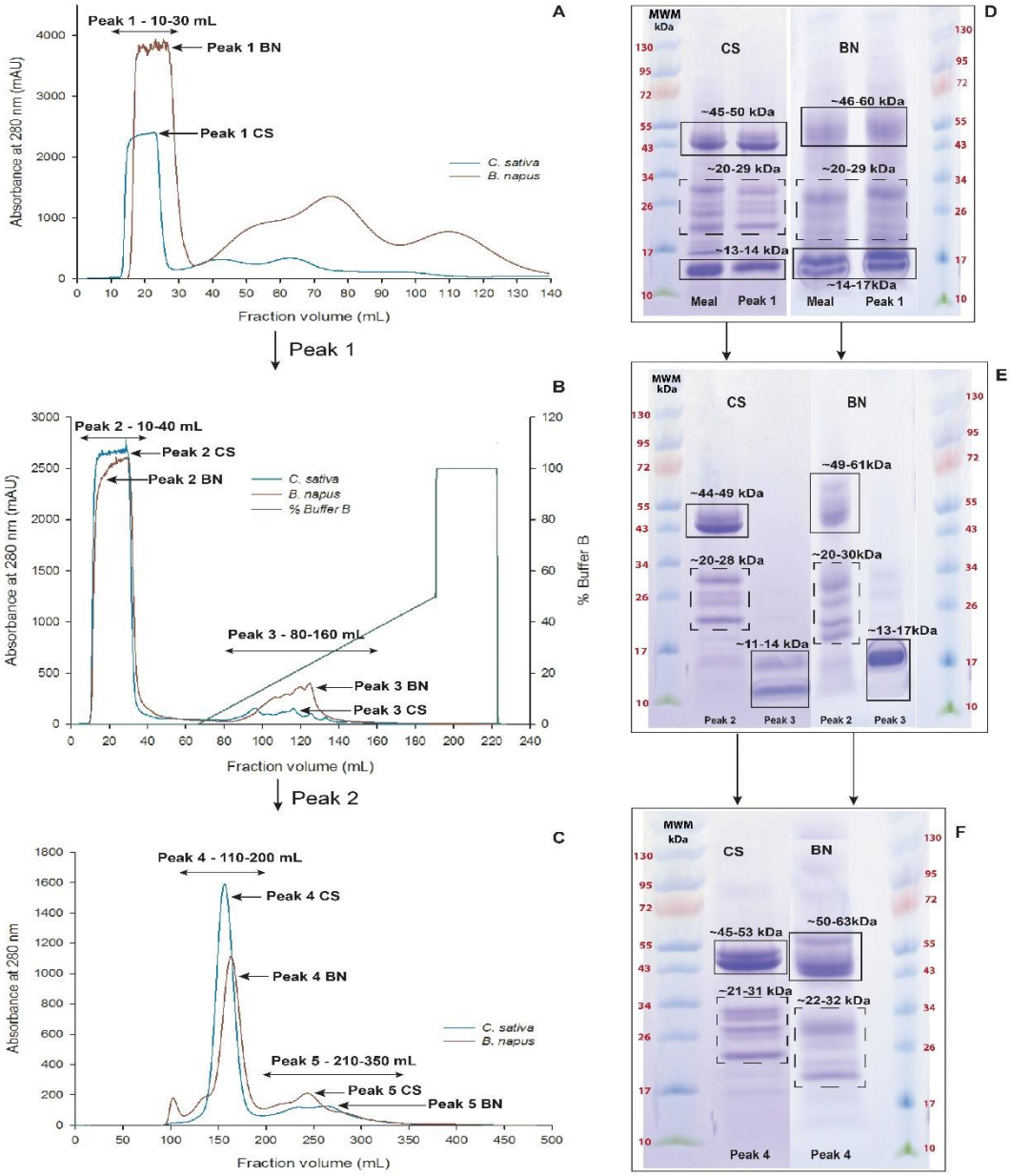

Figure 4.4. Chromatographic purification steps for *C. sativa* (CS) and *B. napus* (BN) cruciferin. (A) Chromatograms of pH 8.5 protein extract separated on a Sephadex G-25 Hiprep^{TM}26/10 desalting column, isocratic elution with buffer 2; (B) Chromatograms of peak 1 CS and peak 1 BN separated on a cation exchange column (CEC; Resource S), gradient elution with buffer A and buffer B; (C) Chromatograms of peak 2 CS and peak 2 BN separated on a Sephacryl S-300 Hiprep^{TM}26/10 high-resolution size exclusion column (SEC S-300), isocratic elution with buffer 2; (D) Polypeptide profiles of CS and BN meal and peak 1 obtained from the desalting column, and (E and F) Polypeptide profiles of peak 2, 3 and peak 4 obtained from CEC and S-300 SEC, respectively. Buffer compositions are as in section 3.1.4.2. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder). Polypeptide profiles were obtained under non-reducing conditions using 8-25% precast gradient gels.

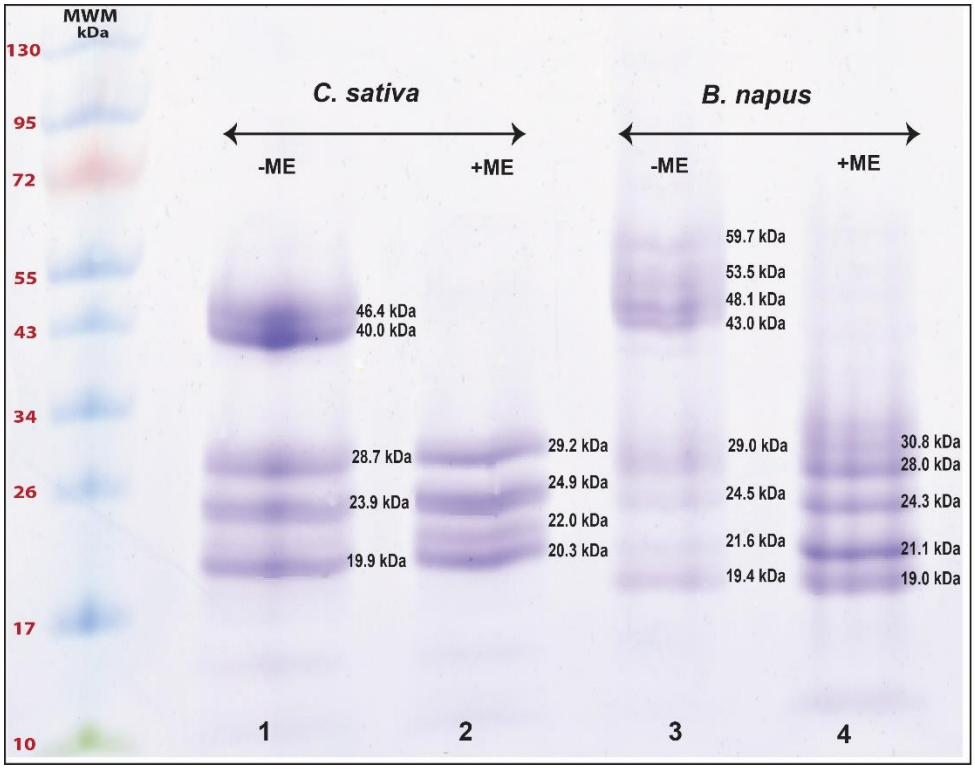

Figure 4.5. Polypeptide profiles of purified cruciferin of *C. sativa* and *B. napus*. Polypeptide profiles are under non-reducing (-ME) and reducing (+ME) conditions in 8-25% gradient recast gel. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

4.2.2 Napin

In a separate extraction, the meal proteins that were soluble at pH 3 and that were retained by a 5 kDa cut-off membrane were considered free of very low molecular weight contaminants. The polypeptide profiles of *C. sativa* and *B. napus* proteins retained by membrane separation predominantly contained polypeptides below 15.0 kDa which contained disulfide bonds (Figure 4.6A and B). These were identified as the characteristic polypeptide bands of napin protein. Moreover, in this protein preparation, *B. napus* exhibited a polypeptide band at 22.0-23.0 kDa, presumably from oleosin or free (dissociated) cruciferin β-chain. Both *C. sativa* and *B. napus* exhibited another polypeptide band at 14.0-15.0 kDa, presumably oleosin. Further separation of these proteins on a hydrophobic interaction column (HIC) produced much cleaner napin (Figure 4.6C) from both species (Figures 4.6D and E). It appears that the HIC separation was not able to remove contaminating polypeptide bands of ~14.0-15.0 kDa and ~22.0-23.0 kDa completely.

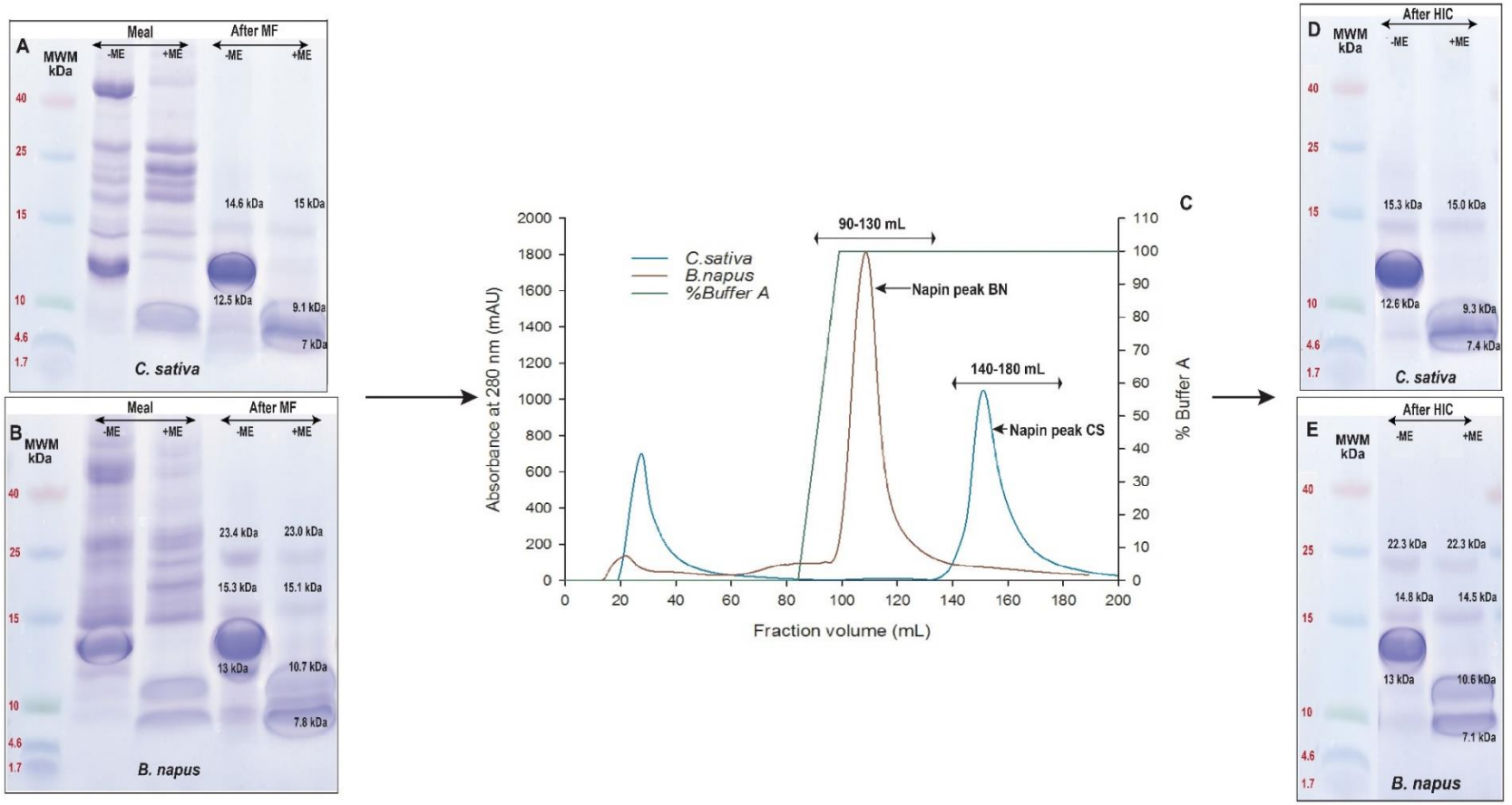

Figure 4.6. Purification of *C. sativa* and *B. napus* napin using membrane filtration (MF) and hydrophobic interaction chromatography (HIC). *C. sativa* (A) and *B. napus* (B) polypeptide profiles of meal and protein after MF; (C) Chromatograms of membrane-separated proteins obtained after HiTrap Phenyl Sepharose™ 6 Fast Flow HIC; (D) *C. sativa* and (E) *B. napus* polypeptide profiles of purified napin after HIC. Polypeptide profiles were obtained under non-reducing (-ME) and reducing (+ME) conditions using 20% homogeneous precast gels. (MWM=Molecular weight markers/Spectra™ Molecular Low-Range protein ladder).

However, the intensities of these bands suggested that they were present in small quantities compared to the polypeptide bands characteristic of napin.

4.2.3 Native-Polyacrylamide Gel Electrophoresis (PAGE) of cruciferin and napin

Native-PAGE provided an idea about the structural conformation of the isolated proteins. This knowledge is important to understand plausible structural changes of cruciferin and napin during the chromatographic purification process. Results obtained from the native-PAGE showed that the protein purification process had not caused dissociation of cruciferin trimeric assembly (can be considered as the quaternary structure) or the subunits (tertiary structure) in either *C. sativa* or *B. napus* (Figures 4.7A and B). However, this process might have caused some degree of dissociation of the hexameric assembly of the cruciferin. The native-PAGE of the purified napin from *C. sativa* and *B. napus* confirmed the monomeric nature of the napin, neither aggregation nor dissociation of the napin was evident (Figures 4.7C and D). It appeared that the purification process performed did not cause marked structural modification in napin from both *C. sativa* and *B. napus*.

4.3 Separation and purification of oil body proteins (OBP)

4.3.1 Microscopic evaluation of oil body ultra-structure

The transmission electron microscopic (TEM) images showed the ultra-structure of cotyledon cells of *C. sativa* and *B. napus* mature seed (Figure 4.8). *C. sativa* contained protein storage vacuoles (PSVs) that were small and fairly uniform in size compared to *B. napus*. Both the PSVs and the oil bodies (OBs) were distributed within the cytoplasm of cotyledon cells and *C. sativa* PSVs had a spherical shape. Within the PSVs, discrete areas were found and they could be the globoids that contain phytic acid crystals (Neumann & Weber, 1978; Lott, 1980; Weber & Neumann, 1980). Compared to *C. sativa*, *B. napus* OBs showed clear morphological differences (Figures 4.9A and B). The number of OBs per unit area was higher in *B. napus* (~5 OBs/μm^2) than in *C. sativa* (~3 OBs/μm^2). The average diameter of *C. sativa* OBs was 0.68 μm, whereas it was 0.43 μm for *B. napus*. The proteins involved in stabilizing oil bodies along with phospholipids are the oil body proteins (OBPs), which may be in the coatings of oil bodies as in Figures 4.9A and B.

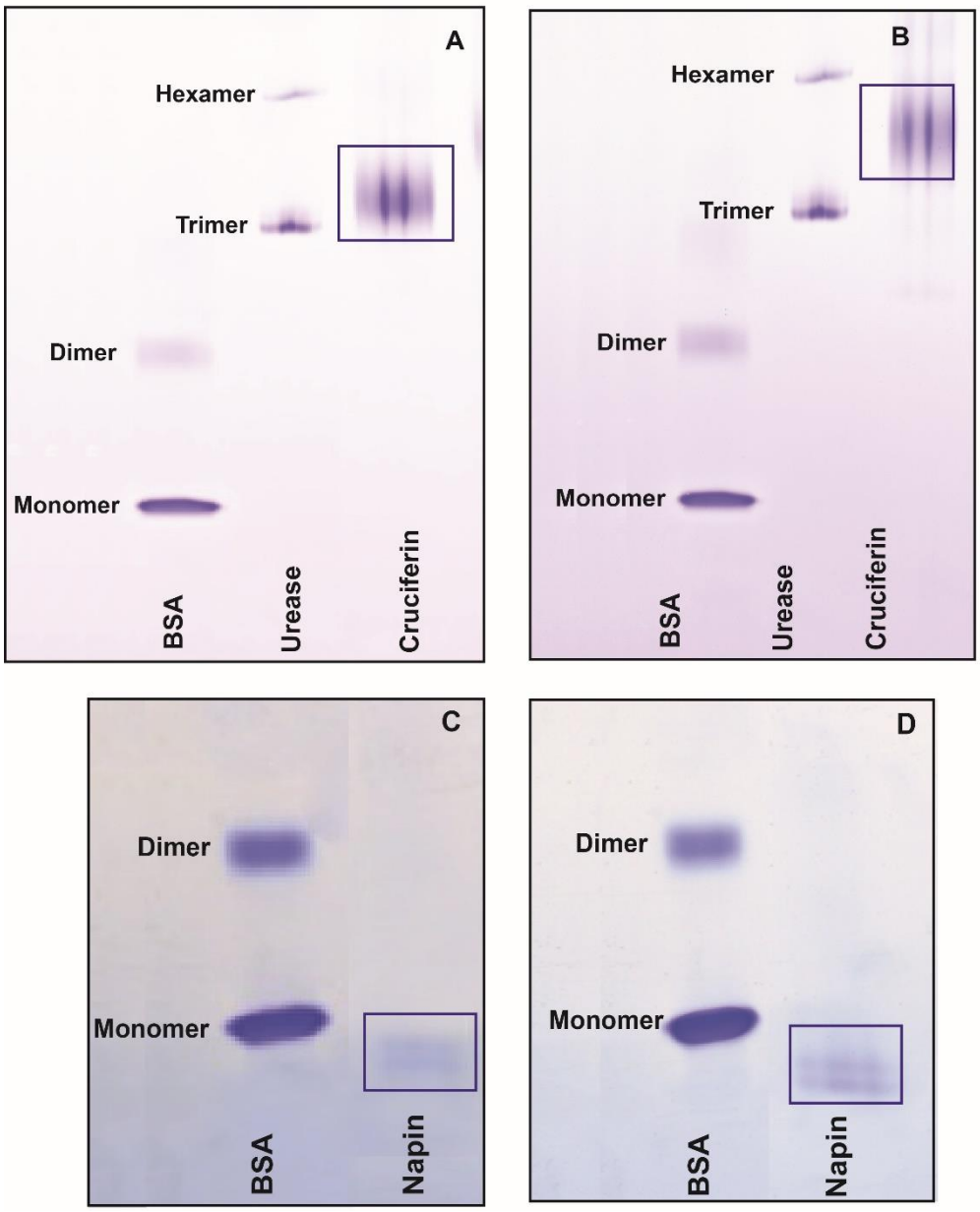

Figure 4.7. Separation of purified cruciferin and napin by native-PAGE. Cruciferin from *C. sativa* (A) *B. napus* (B) and napin from *C. sativa* (C) *B. napus* (D) after native-PAGE separation. Protein levels in each well were the same for the same proteins (1 mg/ mL cruciferin and 4 mg/ mL napin).

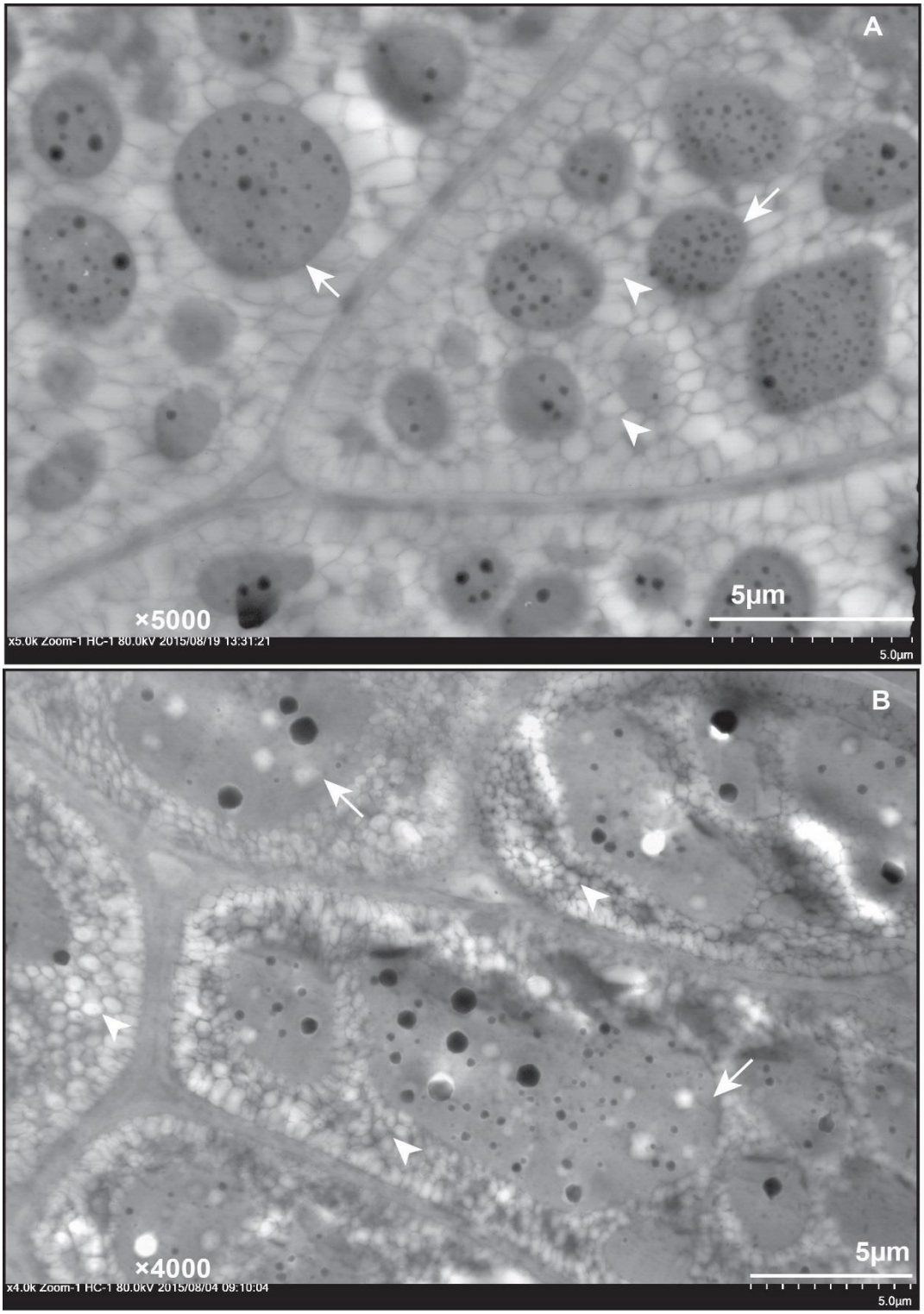

Figure 4.8. Transmission electron microscopic (TEM) images showing the ultra-structure of mature seed cotyledon cells of *C. sativa* and *B. napus*. Cross-sections of the seed along the longitudinal edge of *C. sativa* (A) and *B. napus* (B). Arrows point to PSVs and arrow heads point oil bodies.

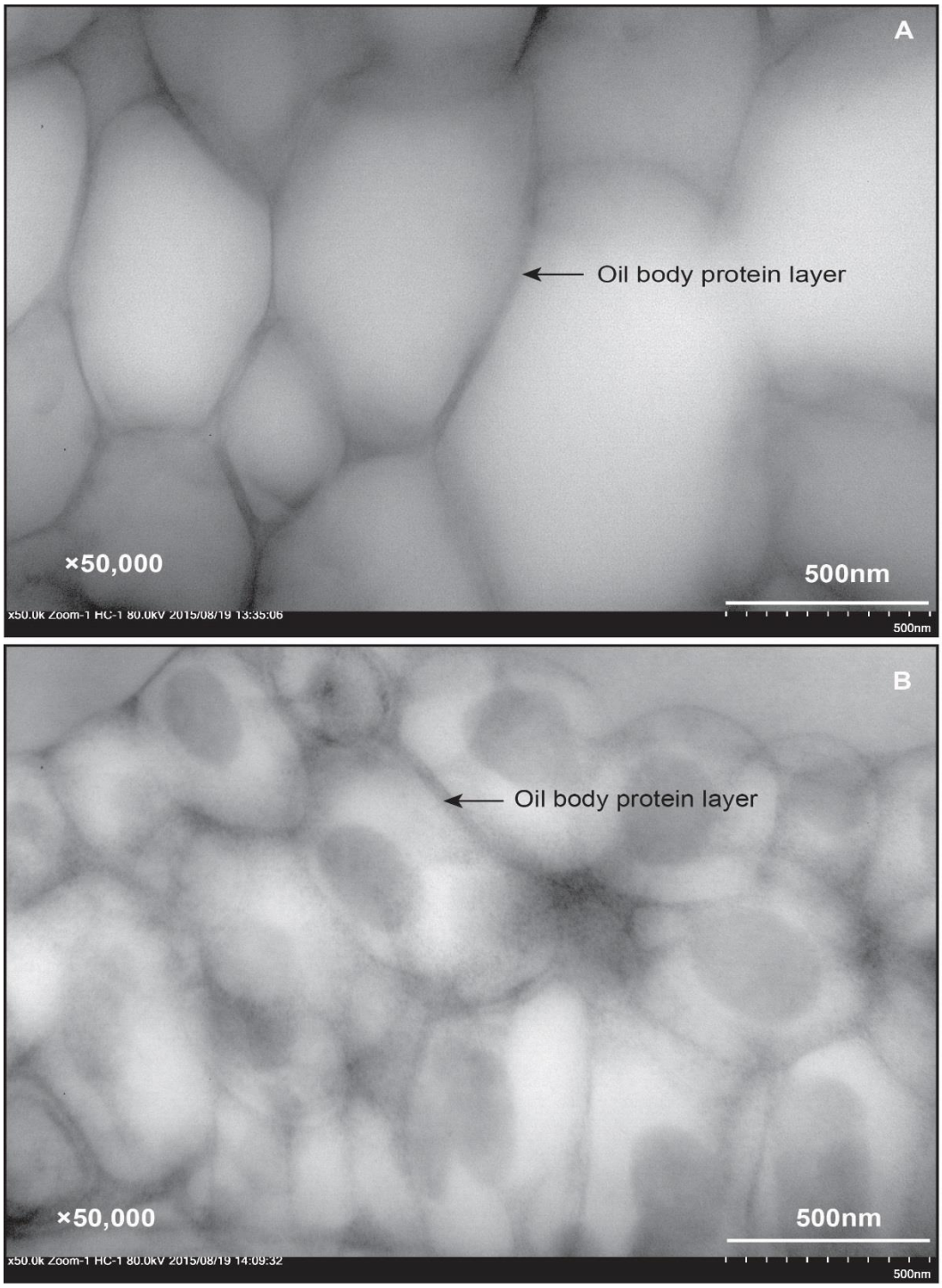

Figure 4.9. Transmission electron microscopic (TEM) images of oil bodies in cotyledon cells from mature *C. sativa* (A) and *B. napus* (B) seed.

4.3.2 Separation and purification of oil body proteins (OBPs)

Although obtaining oil bodies (OBs) from *B. napus* for further studies has been described in the literature, methods available for separating the proteins that are on the surface of the OBs are limited. When OBs obtained from seeds of *C. sativa* were suspended in SDS solution, the protein yielded polypeptide bands with estimated molecular masses of ~ 17, 18, 19 and 26 kDa under non-reducing conditions (Figure 4.10A). Polypeptide bands characteristic of napin indicated possible contamination with the OBPs.

The major polypeptide bands of *B. napus* OBPs were ~15, 17, 19, 23, 28, 34, 36, 62 and 65 kDa under non-reducing conditions; bands characteristic of napin were not obtained (Figure 4.10B). As shown in Figures 4.10A and B, the subnatant of both *C. sativa* and *B. napus* recovered after the washing step with sucrose solution at pH 11 contained polypeptide bands characteristic of storage proteins, *i.e.* cruciferin and napin. This observation confirmed that the washing step with highly alkaline pH adjustment allowed removal of storage proteins that were associated with OBs and OBPs from these seeds. The ability of SDS to displace OBPs was evident, as OBPs were recovered with the addition of SDS.

4.4 Two-dimensional electrophoresis (2DE) and LC-MS/MS analysis of purified proteins: identification and confirmation of identity

4.4.1 Identification and confirmation of cruciferin

Isoelectric focusing followed by separation based on molecular mass confirmed the presence of proteins in the pH range of 3-10 for both *C. sativa* and *B. napus*. On the other hand, no cruciferin isoforms were observed in the pH 9-12 range. Separation of purified cruciferin from *C. sativa* and *B. napus* by 2DE gave 29 and 20 identifiable protein spots, respectively, for each seed type (Figure 4.11A-D); all were in the 17 -55 kDa molecular mass range. Each protein spot on the 2DE gel could be a collection of a number of different proteins that share similar pIs and molecular masses and may contain un-targeted contaminating proteins that were not distinguished by 1DE or 2DE.

The LC-MS/MS analysis carried out on tryptic-digested protein from each of the spots resulted in several matches and only the proteins with 100% probability were considered as the proteins present in the particular spot (Appendix, Table A1).

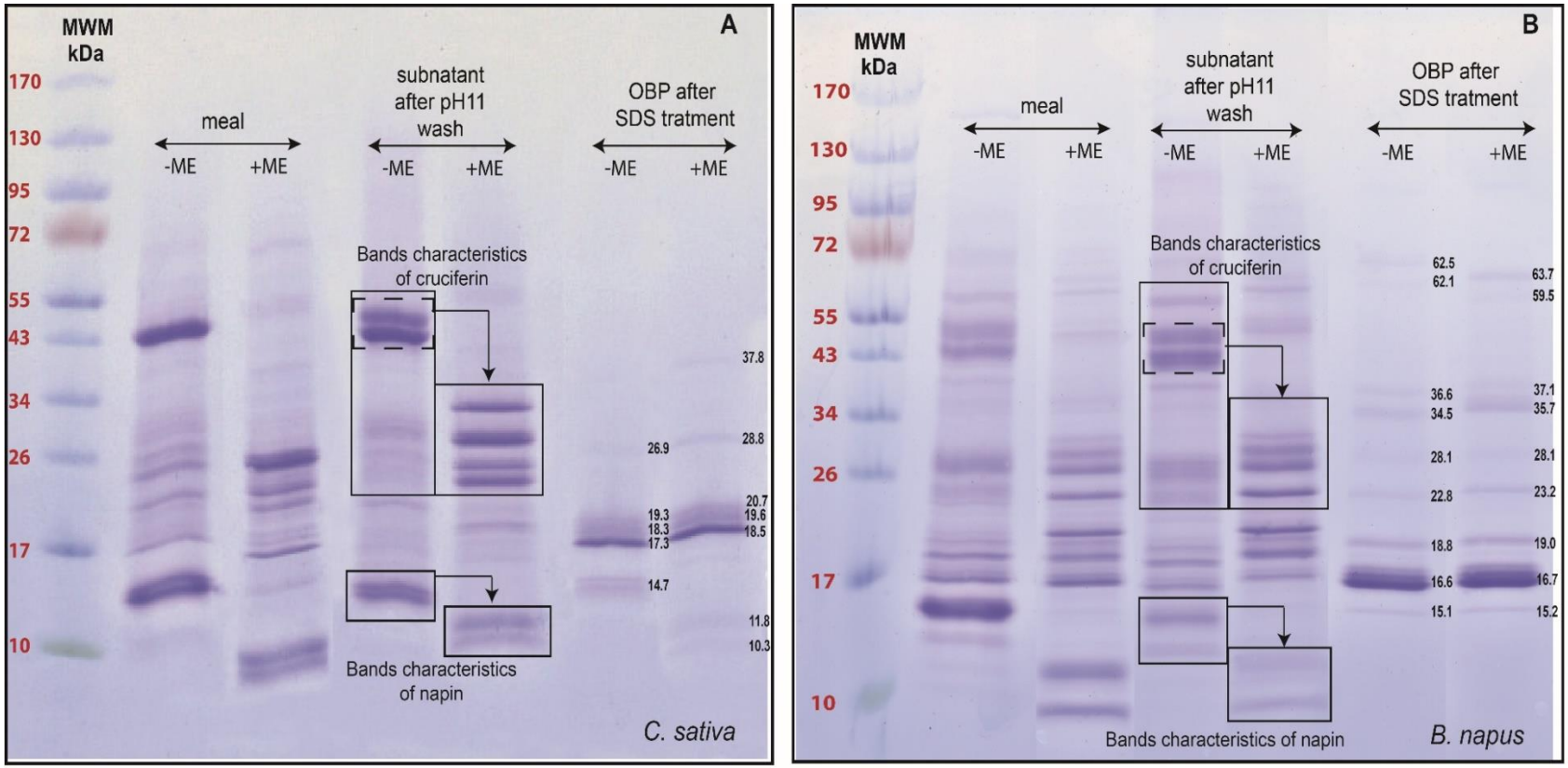

Figure 4.10. Polypeptide profiles of *C. sativa* and *B. napus* seed proteins at different stages of the OBP purification process. (A) *C. sativa* and (B) *B. napus.* Polypeptide profiles under non-reducing (-ME) and reducing (+ME) conditions were separated on 8-25% gradient precast gels. Estimated molecular masses (kDa) are indicated. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

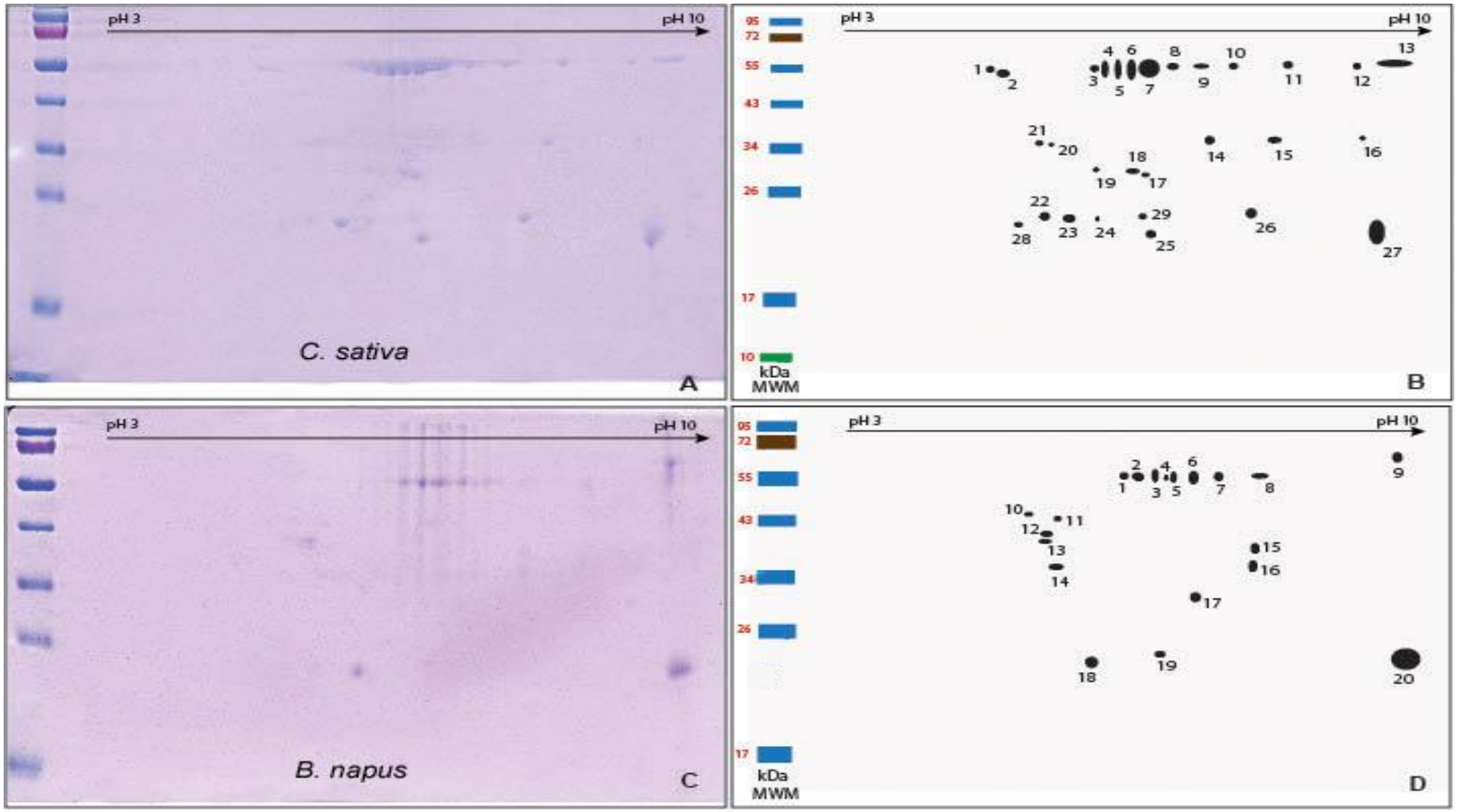

Figure 4.11. Separation of purified cruciferin by 2D electrophoresis under non-reducing conditions. (A and C) *C. sativa* and *B. napus* cruciferin after running the second dimension (SDS-PAGE) using 12% homogeneous hand cast gels; Blue spots indicate the cruciferin isoforms separated by IEF (pH 3 to 10) followed by SDS-PAGE. (B) and (D) are schematic representations of (A) and (C), respectively. Numbers in (B) and (D) represent protein spots visualized in (A) and (C), respectively, based on their staining intensity and subjected to LC-MS/MS analysis. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

.

For *C. sativa* cruciferin, proteins expressed from eleven of the twelve cruciferin genes that are listed in the camelina data base (www.camelinadb.ca) were confirmed (Table 4.3). The cruciferin isoforms were named based on the location of the respective chromosome in genome 1, 2 or 3, and the homology with *A. thaliana* cruciferin. It was noted that three types of cruciferin, *i.e.* CRA, CRB and CRC, which are homologous and closely related to *A. thaliana* CRA, CRB and CRC, respectively, were found in *C. sativa* (Figure 4.12). In addition, *C. sativa* possesses another group of cruciferins named as CRD. The cruciferin encoding gene Csa17g006960 that encodes CRD-1-G1 is the only gene that was not identified in the purified cruciferin from *C. sativa*. The predominant protein found in the purified cruciferin from *C. sativa* was CRA-1-G2.

Besides the twelve cruciferin genes of *C. sativa* listed in Table 4.3, there were six more genes, *i.e.* Csa07g016060, Csa05g038120, Csa19g031870, Csa01g025880, Csa15g039300 and Csa15g039290, had expressed cruciferin-like proteins and they have been identified as the vicilin encoding genes expressing six different vicilins (Table 4.3). These vicilins were named similar manner as did for cruciferins. The phylogenetic relationship (Figure 4.13) shows that vicilin can be categorized into two major classes, *i.e.* Vic1 and Vic2. Although Vic2 was more closely related to two *A. thaliana* vicilin-like proteins (AtPAP85 and AtVCL22, Figure 4.13), AtPAP85 was highly homologous to Vic1, whereas AtVCL22 showed high homology with Vic2.

When the abundance of protein types identified in purified *C. sativa* cruciferin was considered based on the normalized total spectral (NTS) values (Table 4.4), the 2DE separated protein spots of the purified cruciferin contained cruciferin isoforms primarily (89.8-100%). About 0.2-10.2% of vicilin was present in 13 spots. Minor contaminations with napin and other non-storage proteins were also found (Table 4.4). *B. napus* cruciferin was composed of CRU1, CRU2, CRU3 and CRU4 monomers (Appendix, Table A2). When data analysis that was applied similar to the *C. sativa* was applied to *B. napus* purified cruciferin fraction, it was found that >90% cruciferin was present in each and individual 2DE separated protein spot (Table 4.4). Cruciferin of *B. napus* did not result in any matches for vicilin, but contained some minor non-cruciferin protein in which the napin isoform 2SS4 and the late embryogenesis abundant protein (LEA 76) appeared to be predominant. These results showed that the method employed to purify cruciferin resulted predominantly cruciferin for both *C. sativa* and *B. napus*.

Table 4.3. Genes identified that encode cruciferin, vicilin and napin of *C. sativa* and the proposed name for each protein.

Cruciferin gene	Expressed cruciferin*	Vicilin gene	Expressed vicilin **	Napin gene	Expressed napin ***
Csa11g070580	CRA-1-G1	Csa15g039290	Vic1-1-G1	Csa11g017020	Cs2S-1-G1
Csa11g070590	CRA-2-G1	Csa15g039300	Vic1-2-G1	Csa11g017010	Cs2S-2-G1
Csa18g009670	CRA-1-G2	Csa19g031870	Vic1-1-G2	Csa11g017000a	Cs2S-3-G1
Csa17g006950	CRB-1-G1	Csa01g025890	Vic1-1-G3	Csa11g017000b	Cs2S-4-G1
Csa14g004960	CRB-1-G2	Csa01g025885	Vic1-2-G3	Csa12g024730a	Cs2S-1-G3
Csa03g005050	CRB-1-G3	Csa01g025880	Vic1-3-G3	Csa12g024730b	Cs2S-2-G3
Csa11g015240	CRC-1-G1	Csa16g016660	Vic2-1-G1	Csa12g024720a	Cs2S-3-G3
Csa10g014100	CRC-1-G2	Csa07g016060	Vic2-1-G2	Csa12g024720b	Cs2S-4-G3
Csa12g021990	CRC-1-G3	Csa05g038120	Vic2-1-G3		
Csa17g006960	CRD-1-G1				
Csa14g004970	CRD-1-G2				
Csa03g005060	CRD-1-G3				

* Proteins were named based on location of the chromosome in genome G1or G2 or G3 and the homology with *A. thaliana* cruciferins
** Proteins were named based on location of the chromosome in genome G1or G2 or G3 and the homology with *A. thaliana* vicilins
*** Annotations are based on the location of the chromosome in genome G1or G2 or G3.

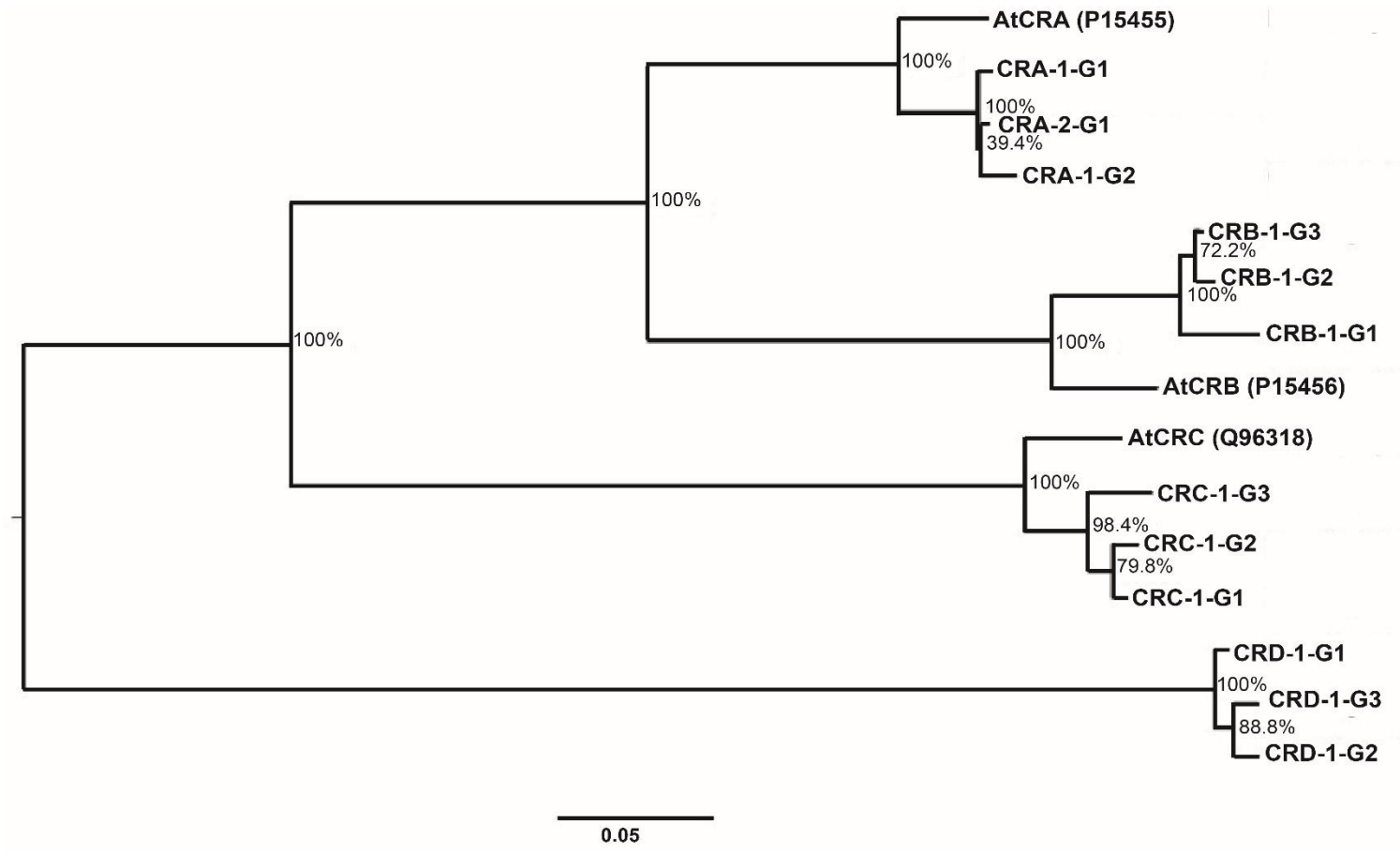

Figure 4.12. Phylogenetic relationship of *C. sativa* and *A. thaliana* cruciferin. The evolutionary history was inferred using the Neighbor-Joining method with the amino acid sequences of *C. sativa* and *A. thaliana* cruciferin in the MEGA6 software. The amino acid sequences of *C. sativa* cruciferin were deduced using the c-DNA sequences of the cruciferin encoding genes available in the camelina genome database (www.camelinadb.ca). The amino acid sequences of *A. thaliana* cruciferin were obtained from the UniProtKB/Swiss-Prot (http://www.uniprot.org/) protein database. The bootstrap values are represented as % at each node.

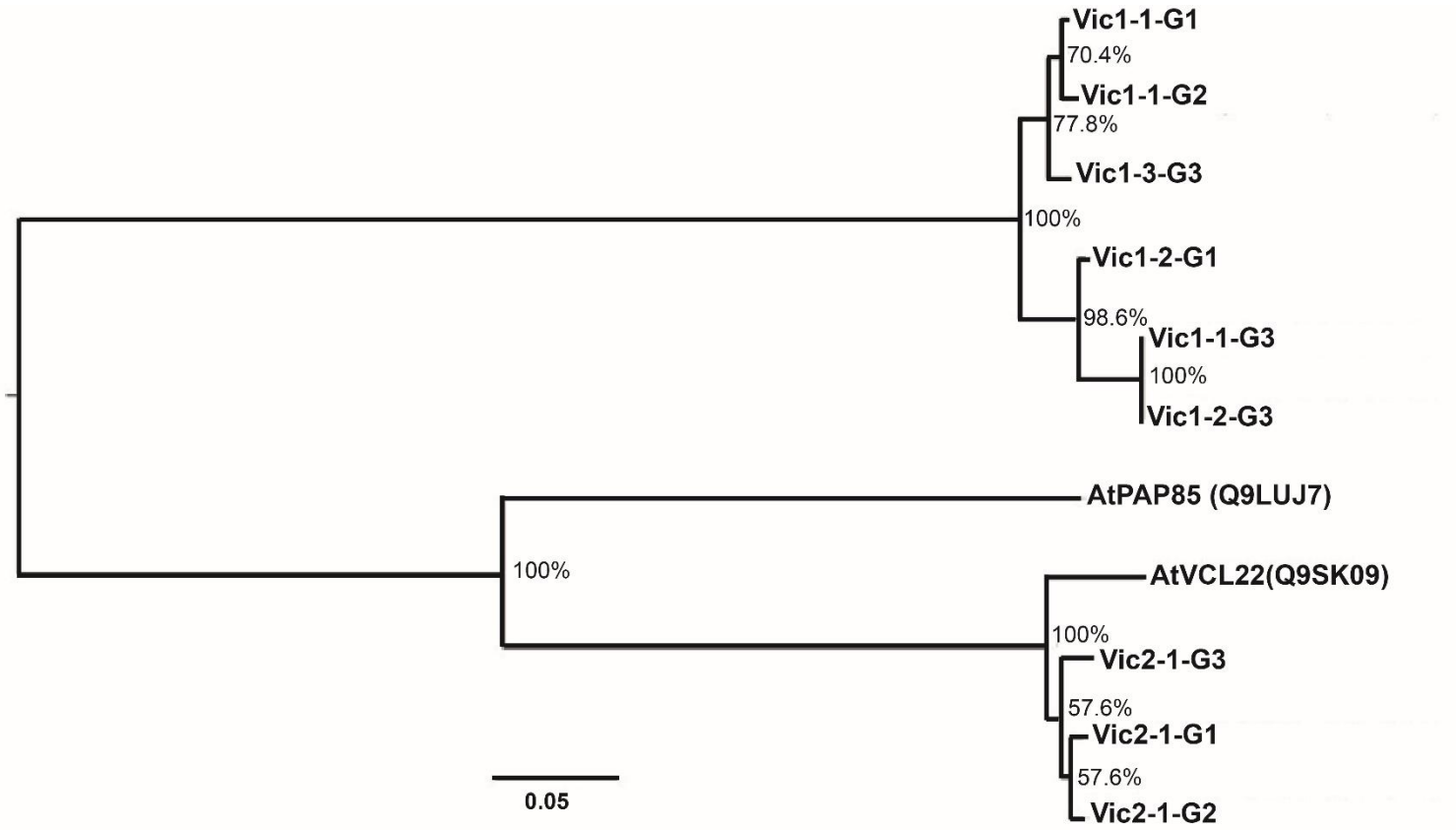

Figure 4.13. Phylogenetic relationship of *C. sativa* and *A. thaliana* vicilin. The evolutionary history was inferred using the Neighbor-Joining method with the amino acid sequences of *C. sativa* and *A. thaliana* vicilin in the MEGA6 software. The amino acid sequences of *C. sativa* cruciferin were deduced using the c-DNA sequences of the vicilin encoding genes available in the camelina genome database (www.camelinadb.ca). The amino acid sequences of *A. thaliana* vicilin were obtained from the UniProtKB/Swiss-Prot (http://www.uniprot.org/) protein database. The bootstrap values are represented as % at each node.

Table 4.4. Abundance of cruciferin, vicilin, napin and other proteins of 2DE separated cruciferin from *C. sativa* and *B. napus* based on normalized total spectral (NTS) values.

C. sativa					*B. napus*				
Spot number*	% cruciferin	% vicilin	% napin	% other protein	Spot number**	% cruciferin	% vicilin	% napin	% other protein
1	100	0	0	0	1	100	0	0	0
2	89.8	10.2	0	0	2	99.2	0	0.8	0
3	100	0	0	0	3	99.1	0	0.6	0.4
4	98.7	1.1	0	0.3	4	99.2	0	0.8	0
5	96.8	1.0	0	2.2	5	99.1	0	0.6	0.3
6	96.6	1.2	0	2.1	6	98.5	0	1.1	0.4
7	100	0	0	0	7	100	0	0	0
8	97.0	1.2	0.1	1.7	8	100	0	0	0
9	99.6	0.4	0	0	9	99.0	0	1.0	0
10	99.8	0.2	0	0	10	100	0	0	0
11	100	0	0	0	11	100	0	0	0
12	100	0	0	0	12	100	0	0	0
13	95.8	1.4	0	2.8	13	96.5	0	3.5	0
14	90.6	9.4	0	0	14	100	0	0	0
15	100	0	0	0	15	97.5	0	0	2.5
16	97.3	2.7	0	0	16	91.3	0	0	8.7
17	100	0	0	0	17	100	0	0	0
18	95.5	4.5	0	0	18	100	0	0	0
19	94.0	5.1	0	0.9	19	98.7	0	0	1.3
20	99.5	0	0	0.5	20	98.6	0	0.5	0.9
21	99.6	0	0	0.4					
22	100	0	0	0					
23	100	0	0	0					
24	100	0	0	0					
25	100	0	0	0					
26	98.8	0	0	1.2					
27	98.0	1.6	0.4						
28	99.6	0	0	0.4					
29	100	0	0	0					

* Refers to the protein spots obtained from 2DE separation of cruciferin from *C. sativa* (29 spots, Figure 4.11B)
** Refers to the protein spots obtained from 2DE separation of cruciferin from *B. napus* (20 spots, Figure 4.11D)

4.4.2 Identification and confirmation of napin

Napin proteins isolated from *C. sativa* and *B. napus* generated 18 and 10 separate spots, respectively (Figure 4.14A and B), which were in the pH range of 9-12 and the molecular mass range of ~12-30 kDa. All the possible proteins that were present in purified napin from *C. sativa* and *B. napus* are listed in Tables A3 and Table A4, respectively, in the Appendix. As was done with cruciferin, only the proteins with 100% probability were considered as the most likely proteins present in a particular spot. Eight different napins from eight napin encoding genes of *C. sativa* were named based on their chromosome location and are listed in Table 4.3. The phylogenetic relationship (Figure 4.15) showed that the napin encoding genes of *C. sativa* and *A. thaliana* are distantly related.

Only five napins, namely. Cs2S-2-G1, Cs2S-2-G3, Cs2S-4-G1, Cs2S-4-G3 and Cs2S-1-G1, (Table 4.3) were identified for the 18 protein spots from *C. sativa* (Figure 4.14A). These napin isoforms accounted for 12.1-100% (Table 4.5) of the proteins, indicating somewhat lesser purity than expected. The remainder of the identified proteins were different isoforms of late embryogenesis abundant (LEA) proteins (Table 4.5 and Table A3, Appendix). It appears that LEA proteins were co-extracted with napin at pH 3 and remained together during membrane separation and the subsequent chromatography step due to hydrophobic interactions. Therefore, it can be assumed that the contaminating polypeptide bands, other than the 9.3 and 7.4 kDa bands, observed in 1DE (Figure 4.6D) may be LEA proteins. In contrast, the LC-MS/MS analysis of napin from *B. napus* showed that nine of eleven 2DE protein spots contained only napin (>99%, Table 4.5) and were comprised of 2SS2, 2SS3, 2SSI, 2SSB and 2SSE isoforms (Table A4, Appendix). Napin from *B. napus* showed contamination with non-napin proteins (only in two protein spots) and different compared to the napin from *C. sativa*. The non-napin proteins from *B. napus* contained mainly the cruciferin (CRU4 > CRU3) and LEA 76 (Table 4.5). Hence, it appears that the contaminating polypeptide bands (Figure 4.6E), other than the 10.6 and 7.1 kDa bands, were from cruciferin and LEA proteins.

4.4.3 Identification and confirmation of oil body proteins

When the OB surface proteins of *C. sativa* were separated by 2DE in the pH ranges of 3-10 and 9-12, each resolved into 10 identifiable protein spots (Figure 4.16). All of them were in the 10.0-55.0 kDa molecular mass range. For *B. napus*, 11 spots (17.0-43.0 kDa) for pH 3 to 10 and 6 spots (17.0-43.0 kDa) for the pH 9-12 range were identified (Figure 4.17).

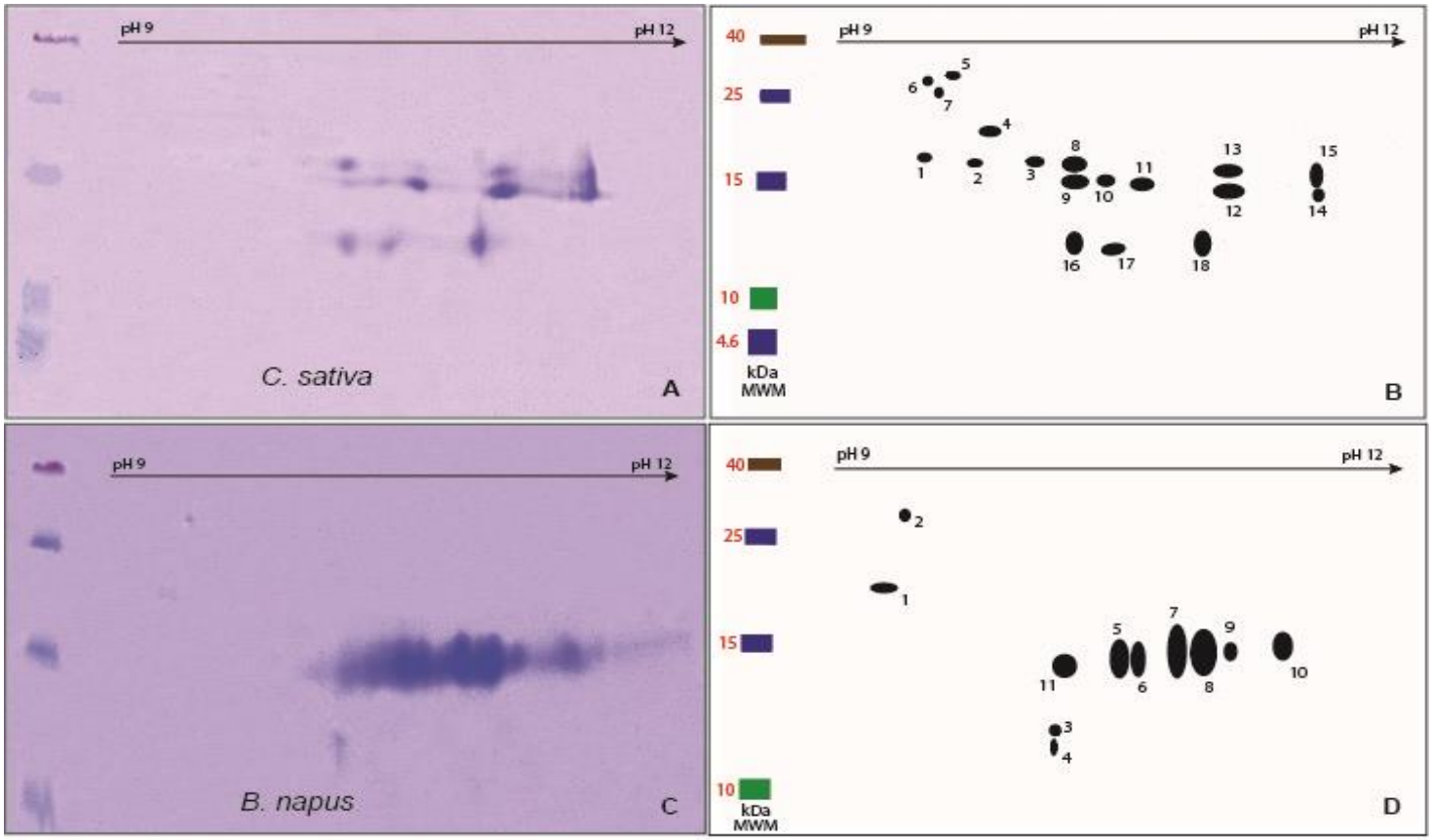

Figure 4.14. Separation of purified napin by 2D electrophoresis under non-reducing conditions. (A and C) *C. sativa* and *B. napus* napin after running the second dimension (SDS-PAGE) using 16% homogeneous hand cast gels. Blue spots indicate the napin isoforms separated by IEF (pH 9 to 12) followed by SDS-PAGE. (B) and (D) are schematic representations of (A) and (C), respectively. Numbers in (B) and (D) represent protein spots visualized in (A) and (C), respectively, based on their staining intensity and subjected to LC-MS/MS analysis. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

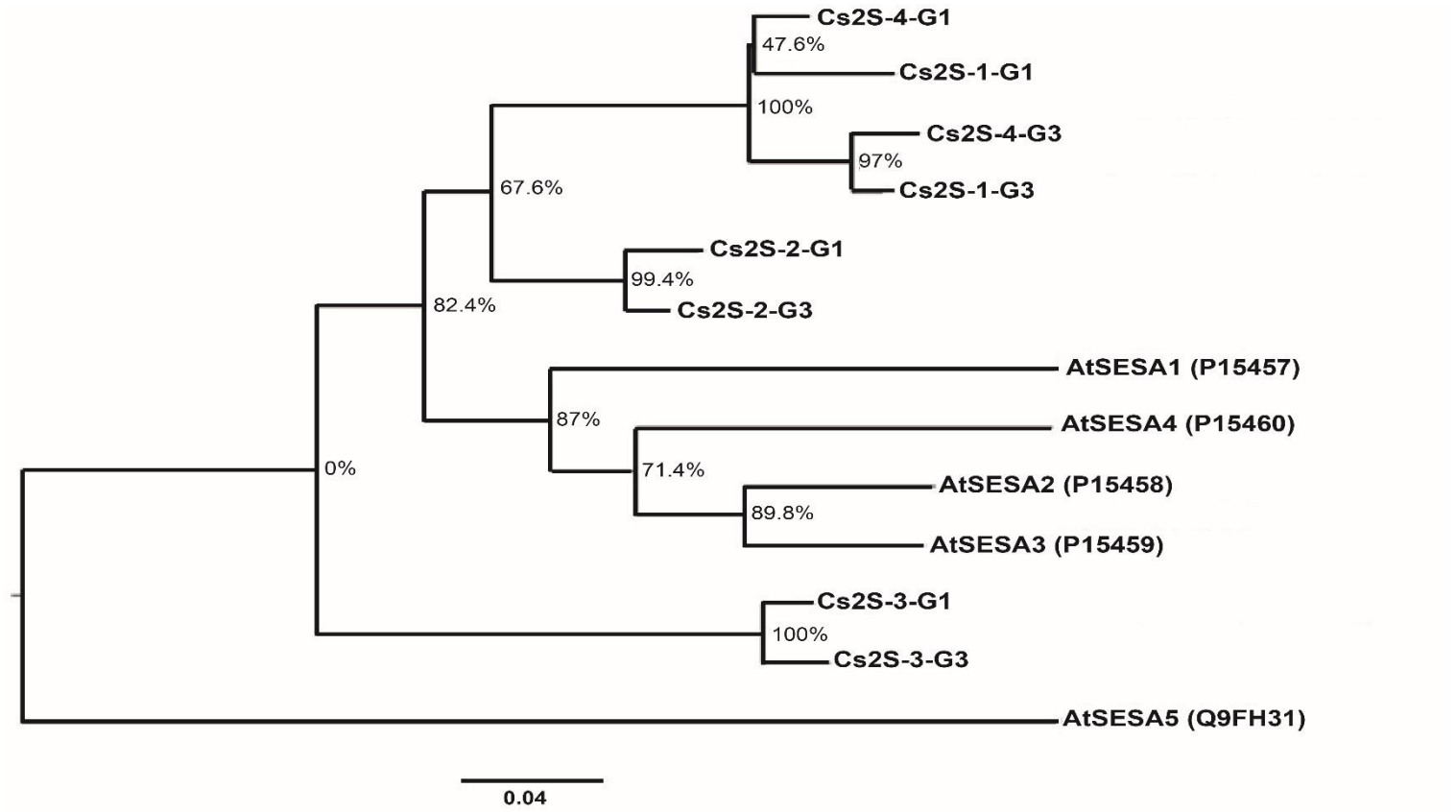

Figure 4.15. Phylogenetic relationship of *C. sativa* and *A. thaliana* napin. The evolutionary history was inferred using the Neighbor-Joining method with the amino acid sequences of *C. sativa* and *A. thaliana* napin in the MEGA6 software. The amino acid sequences of *C. sativa* napin were deduced using the c-DNA sequences of the napin encoding genes available in the camelina genome database (www.camelinadb.ca). The amino acid sequences of *A. thaliana* napin were obtained from the U niProtKB/Swiss-Prot (http://www.uniprot.org/) protein database. The bootstrap values are represented as % at each node.

Table 4.5. Abundance of napin, late embryogenesis abundance (LEA) protein, cruciferin and other proteins of 2DE separated napin from *C. sativa* and *B. napus* based on normalized total spectral (NTS) values.

C. sativa					*B. napus*				
Spot number*	% napin	% LEA	% cruciferin	% other protein	Spot number**	% napin	% LEA	% cruciferin	% other protein
1	41.0	52.3	1.4	5.3	1	0	0	100	0
2	53.1	45.2	0.0	1.7	2	0	91.2	8.8	0
3	98.5	1.5	0.0	0.0	3	100	0	0	0
4	25.5	62.8	0.0	11.7	4	100	0	0	0
5	12.5	87.5	0.0	0.0	5	100	0	0	0
6	12.1	87.9	0.0	0.0	6	100	0	0	0
7	12.1	87.4	0.0	0.4	7	100	0	0	0
8	97.3	0.9	0.0	1.8	8	100	0	0	0
9	97.1	1.0	0.0	1.9	9	100	0	0	0
10	97.4	0.0	0.3	2.3	10	100	0	0	0
11	99.1	0.0	0.0	0.9	11	99.8	0.0	0.2	0
12	99.6	0.4	0.0	0.1					
13	99.6	0.3	0.0	0.1					
14	100.0	0.0	0.0	0.0					
15	99.6	0.4	0.0	0.0					
16	88.9	0.0	0.0	11.1					
17	96.6	0.0	0.0	3.4					
18	99.9	0.0	0.0	0.1					

* Refers to the protein spots obtained from 2DE separation of napin from *C. sativa* (18 spots, Figure 4.14B)
** Refers to the protein spots obtained from 2DE separation of napin from *B. napus* (11 spots, Figure 4.14D)

These results indicated that the presence of proteins with similar molecular weights or pIs, presumably isoforms of OBPs, were present in both species. Separation of *C. sativa* OBPs under 1D SDS-PAGE (Figure 4.10A) showed a few polypeptide bands characteristic of napin. Therefore, napins were expected to be revealed by LC-MS/MS. All possible proteins present on the *C. sativa* OB surface are listed in Table A5 (pH 3-10) and Table A6 (pH 9-12) in the Appendix. The presence of oleosin isoforms encoded from multiple genes, Csa11g019460, Csa12g028090, Csa10g047190, Csa11g082710, Csa02g041750, Csa04g015780, Csa06g008780, were confirmed for the proteins resolved in the pH range of 3-12. Several other genes, Csa00532s200, Csa05g020560, Csa26607s010, Csa03g053840, Csa04g046970 and Csa01g021420, which encode oleosin family proteins, also were identified. The presence of protein encoded from Csa02g057710, which is similar to *A. thaliana* peroxygenase, was evident in the purified OBP sample. *A. thaliana* peroxygenase is known as caleosin (Meesapyodsuk & Qiu, 2011), the next most abundant type of protein found on the OB surface (Tzen, 2012). In addition, several other genes, Csa03g006900, Csa09g069460, Csa05g023090 and Csa07g038560, possibly encoding caleosin, also were identified. The presence of steroleosin was evident only in *C. sativa*. The identified steroleosins have molecular masses of ~38-42 kDa (Tables A5 and A6, Appendix). The analysis showed that a mixture of oleosins, caleosins, napin, cupin family proteins (cruciferin and vicilin), and many other membrane-bound proteins, especially ribosomal proteins, were found in the isolated OBPs of *C. sativa*. The abundance of OBPs were found to be low compared to the other non-OBPs.

The proteins identified for *B. napus* that resolved in the pH 3-12 range were isoforms of oleosins (Tables A7 and A8, Appendix; Figures 4.19A and B). The oleosin isoforms OLES2, OLEO5, OLES1, and OLEO3 accounted for majority of the proteins separated from the *B. napus* OB surface. In addition to oleosin, some amount of cruciferin (CRU1, CRU3 and CRU4) and minor amounts of napin (2SS4 and 2SSE) and myrosinase also were identified, but neither caleosin nor steroleosin was detected. Compared to *C. sativa*, OBPs separated from *B. napus* contained mostly oleosins with minor contamination. Contaminating proteins of the OBP preparation were fewer in number, providing a less complicated mixture which contained only cruciferin and minute amounts of napin and myrosinase. It appears that the non-reducing polypeptide bands observed in the final OBP preparation of *B. napus* (Figure 4.10A) were free α- and β-chains of cruciferin monomers and myrosinase.

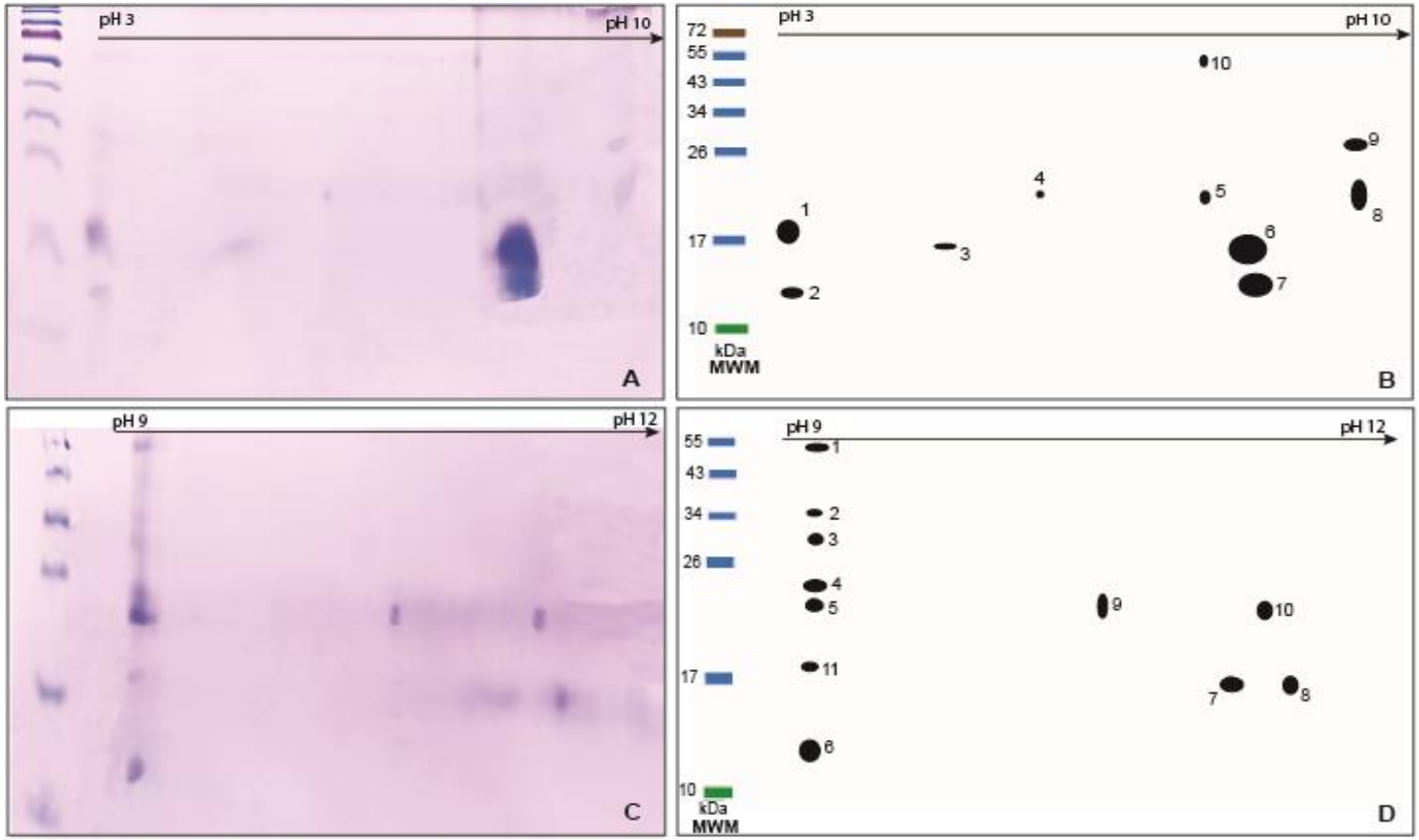

Figure 4.16. Separation of *C. sativa* oil body proteins (OBPs) by 2DE under non-reducing conditions. OBPs separated at pH 3 to 10 (A), and pH 9 to 12 (B) as the first dimension and SDS-PAGE as the second dimension using 14% homogeneous hand cast gels. Blue spots indicate OBPs separated by IEF followed by SDS-PAGE. (B) and (D) are schematic representations of (A) and (C), respectively. Numbers in (B) and (D) represent protein spots visualized in (A) and (C), respectively, based on their staining intensity and subjected to LC-MS/MS analysis. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

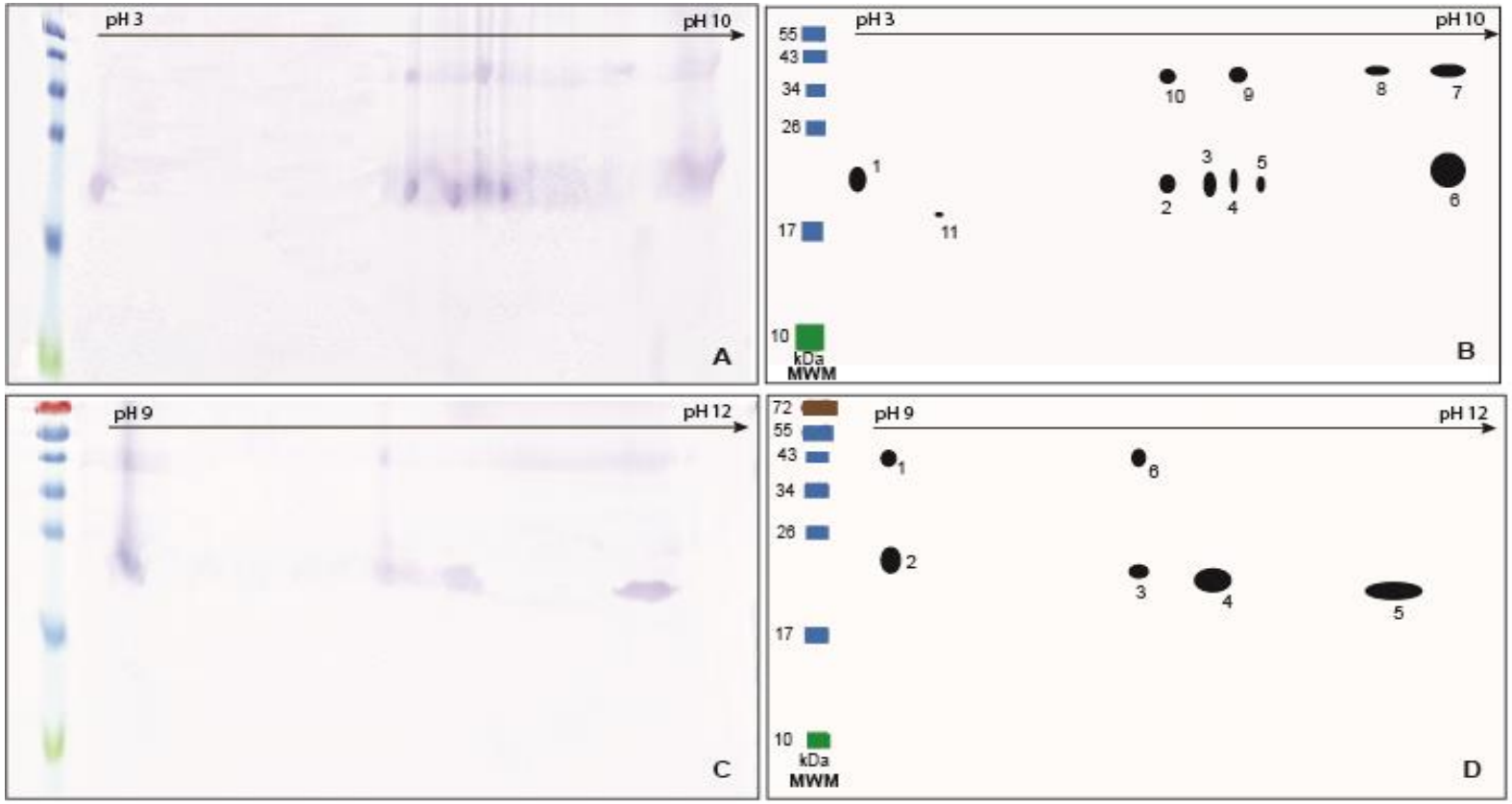

Figure 4.17. Separation of *B. napus* oil body proteins (OBPs) by 2DE under non-reducing conditions. OBP separated at pH 3 to 10 (A), and pH 9 to 12 (B) as the first dimension and SDS-PAGE as the second dimension using 14% homogeneous hand cast gels. Blue spots indicate OBPs separated by IEF followed by SDS-PAGE. (B) and (D) are schematic representations of (A) and (C), respectively. Numbers in (B) and (D) represent protein spots visualized in (A) and (C), respectively based on their staining intensity and subjected to LC-MS/MS analysis. (MWM=Molecular weight markers/ PageRuler^{TM} Pre-stained Protein Ladder).

4.5 Structural details of storage proteins of *C. sativa* and *B. napus*

4.5.1 Details of 2˚ structure of cruciferin and napin and the effect of pH

4.5.1.1 Analysis by FT-IR spectroscopy

The amide I band of the FT-IR spectrum provides information on the secondary structural features of a protein. Since the amide I band is a collection of number of peaks, deconvolution allows a quantitative estimation of each secondary structural component that is represented by the IR signal to be obtained (Kong & Yu, 2007). The FT-IR spectra of cruciferin (solid state) obtained from *C. sativa* and *B. napus* (Figure 4.18) did not show clear differences in the deconvoluted amide I band. However, regions characteristic for the $-PO_3$ (970 cm^{-1}), C-O-P (1070 cm^{-1}) and -P=O (1170 cm^{-1}) functional groups showed subtle differences. The deconvoluted amide I band revealed predominant β-sheet structure for cruciferin in both species (Table 4.6). A significantly higher ($P<0.05$) β-sheet content was observed in *B. napus* than in *C. sativa*. The α-helix, β-turn and random structures of cruciferin were not significantly different ($P>0.05$) in the two species.

When the FT-IR spectra generated by napins (solid state) of *C. sativa* and *B. napus* were examined, clear differences were observed in the amide I band and the regions characteristic for the $-PO_3$ (970 cm^{-1}), C-O-P (1070 cm^{-1}) and -P=O (1170 cm^{-1}) functional groups (Figure 4.19). Napin protein is known to have a highly helical secondary structure (Figure 2.6; Rico et al., 1996). Therefore, only α-helix content was calculated. The deconvoluted amide I band of napin showed more α-helical content in *C. sativa* compared to *B. napus* (Table 4.6).

Table 4.6. Secondary structural components (%) of purified cruciferin and napin of *C. sativa* and *B. napus*. Values are presented as mean ± standard error.

Protein	Seed species	*α*-helix	*β*-sheet	*β*-turn	random
Cruciferin	*C. sativa*	10.1 ± 0.3 [a]	43.0 ± 0.9 [a]	19.8 ± 0.8 [a]	3.9 ± 0.6 [a]
	B. napus	9.4 ± 0.4 [a]	45.6 ± 0.1 [b]	20.1 ± 0.4 [a]	3.2 ± 0.5 [a]
Napin	*C. sativa*	32.1 ± 0.4[b]	NA	NA	NA
	B. napus	26.0 ± 0.9[c]	NA	NA	NA

NA=Not Applicable. Means followed by the same superscript within the same column are not significantly different ($p>0.05$)

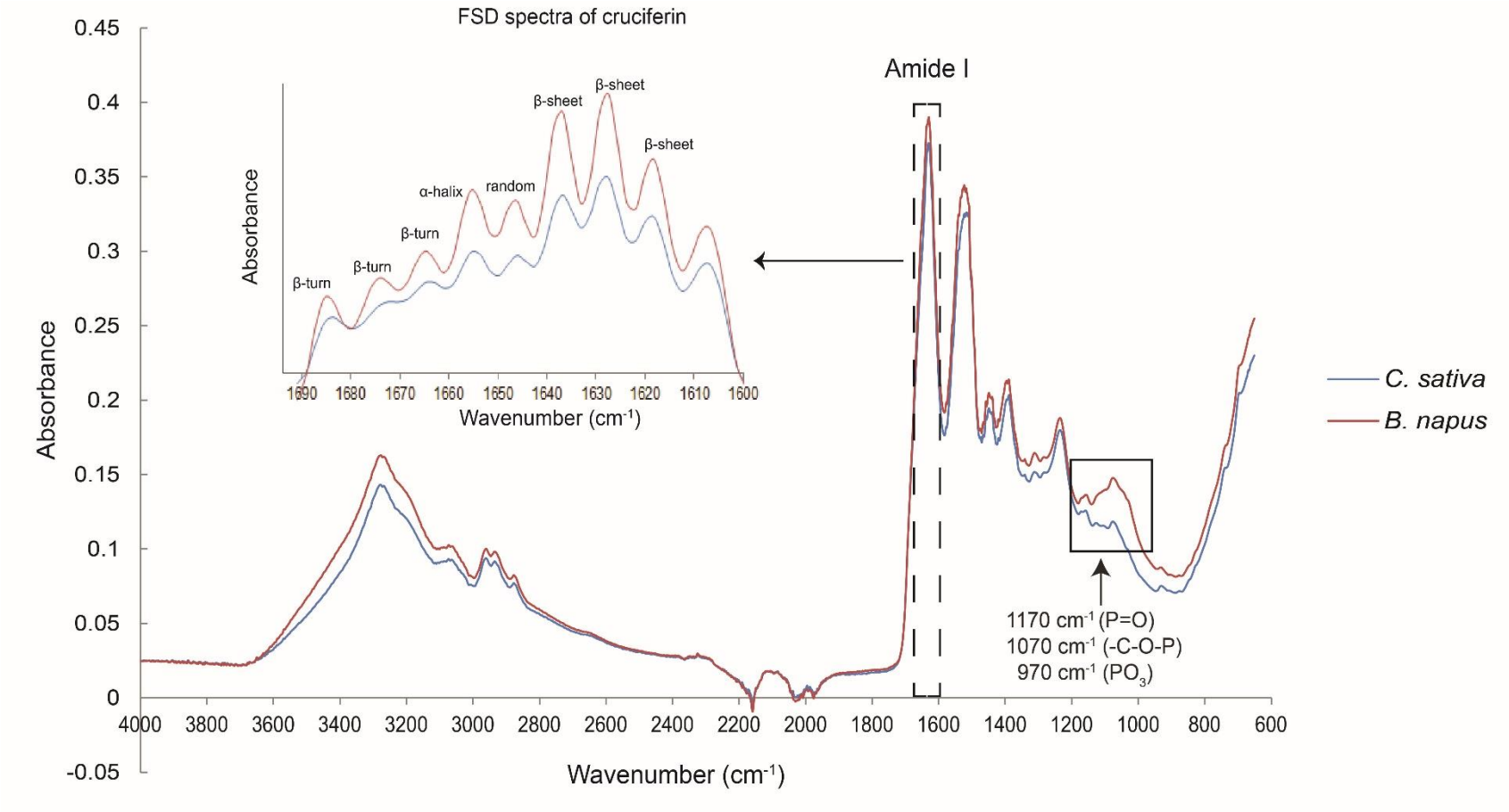

Figure 4.18. FT-IR spectra of cruciferin obtained from *C. sativa* and *B. napus*. Inset: Secondary structural components resolved and identified by Fourier self-deconvoluting of the Amide I region (1600-1690 cm^{-1}). Parameters of Fourier self-deconvolution (FSD) of amide 1 peak: Resolution enhancement factor (K) = 2.5, Full width at half height = 14 cm^{-1} and Apodization filter = Bessel.

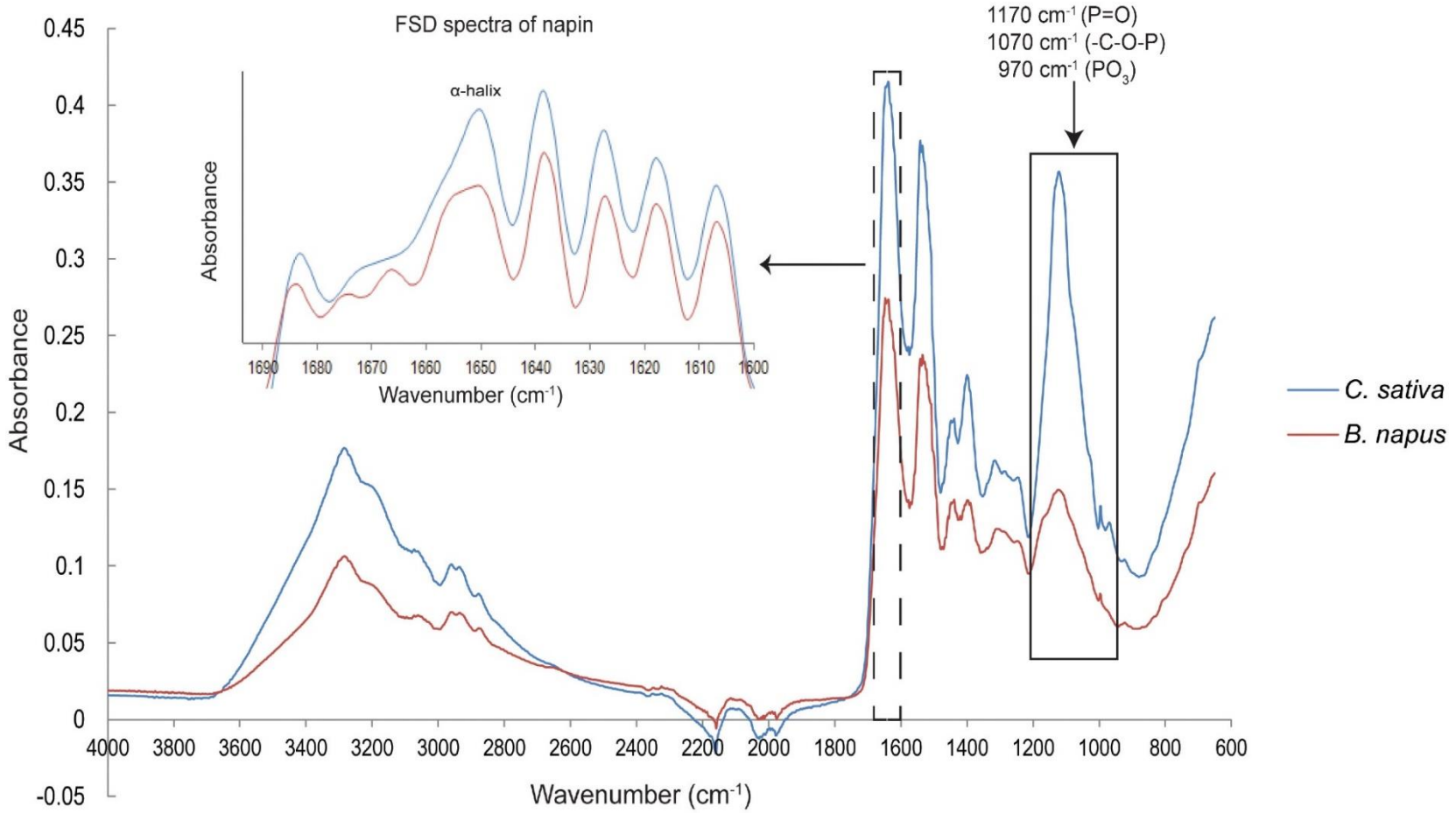

Figure 4.19. FT-IR spectra of napin obtained from *C. sativa* and *B. napus*. Inset: Secondary structure components resolved and identified by Fourier self-deconvoluting of the Amide I region (1600-1690 cm^{-1}). Parameters of Fourier self-deconvolution (FSD) of amide 1 peak Resolution enhancement factor (K) = 2.8, Full width at half height = 18 cm^{-1} and Apodization filter = Bessel.

4.5.1.2 Analysis by far UV-Circular Dichroism (far UV-CD)

The far UV-CD spectra due to the peptide chromophore of cruciferin in solution (Figure 4.20A) provided characteristic spectra for a predominant *β*-structure protein (Kelly, Jess, & Price, 2005). According to Greenfield (2006), CD spectrum of a protein provides characteristic bands for α-helical (negative at ~208 nm and 222 nm, positive at ~195 nm), *β*-sheet (negative at ~218 nm, positive at ~195 nm) and random coil or disordered structure (very low ellipticity above ~210 nm and positive bands near ~195 nm). The changes in the shape and magnitude of these bands indicated possible changes in the secondary structural features that may have occurred at different pHs. The secondary structural features calculated from far UV-CD for different pHs (Table 4.7) clearly indicated that the helical content of *C. sativa* and *B. napus* cruciferin changed considerably depending on the pH. For *C. sativa,* the highest helical content was observed at pH 3 and the lowest at pH 7. The *β*-sheet or *β*-turn content of *C. sativa* cruciferin did not show any significant change with pH. However, the random structure content of *C. sativa* cruciferin showed a significant increase at pH 7 or 10 compared to pH 3. The *β*-sheet content of *B. napus* cruciferin showed an increase when moving from acidic to neutral pH and then decreased as the pH become alkaline. The *β*-turn content of *B. napus* cruciferin was significantly ($p<0.05$) lower at pH 7 compared to the other two pHs, where the values were similar; the random structure content behaved similarly. Compared to pH 3 and 7, the lowest α-helix and total β structure (β-sheet and β-turn) content and the highest random structure content were seen at pH 10 for cruciferin from both *C. sativa* and *B. napus*. These results suggested that the cruciferins of *C. sativa* and *B. napus* went through considerable changes in secondary structure in response to changes in medium pH.

The far UV-CD spectra of napin of both species (Figure 4.20B) at pH 7 showed typical features for an α-helix predominating protein (negative bands at ~222 nm and ~208 nm and a positive band at ~195 nm; Kelly et al., 2005; Greenfield, 2006). As the pH changed to acidic or basic, characteristic features did not change substantially, indicating no or minor changes in secondary structural features in response to changes in medium pH. The α-helix content of both *C. sativa* and *B. napus* napin showed slightly higher levels at pH 7 compared to pH 3 and 10 (Table 4.7), with similar values observed for pH 3 and 10. *C. sativa* napin showed significantly higher random structure content at pH 10 compared to that at pH 3 or 7, whereas no significant difference ($p<0.05$) in the random structure content was observed for *B. napus* napin at any pH. The results suggested that the secondary structure did not change significantly with changes in pH.

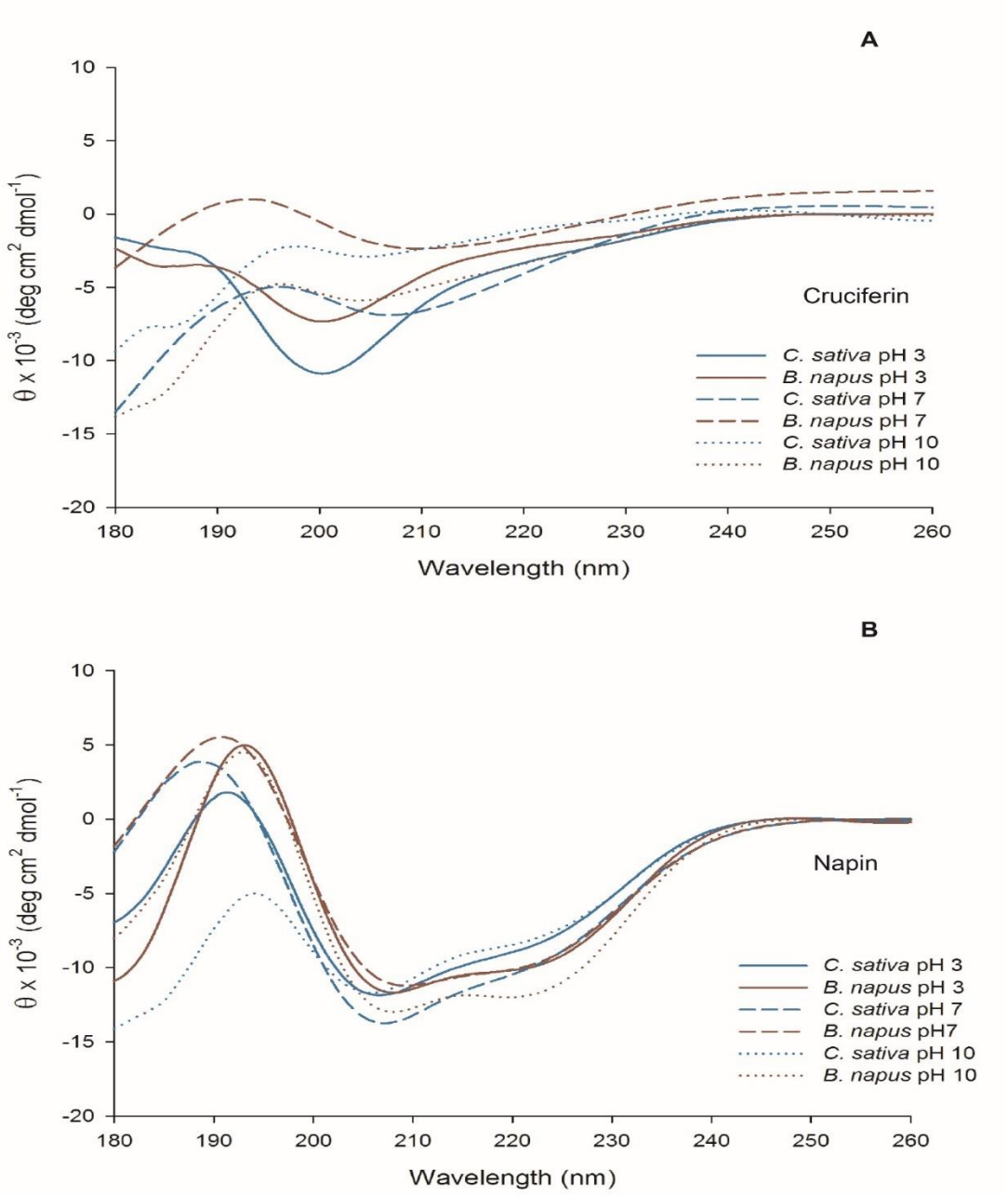

Figure 4.20. Far UV-CD spectra of purified cruciferin and napin when medium pH was 3, 7 or 10. (A) *C. sativa* and *B. napus* cruciferin, and (B) *C. sativa* and *B. napus* napin showing secondary structural changes due to change in medium pH.

Table 4.7. Secondary structural components (%) of purified cruciferin and napin from *C. sativa* and *B. napus* at different pHs. Values are presented as mean ± Standard error.

Protein	pH	Seed species	α-helix	β-sheet	β-turn	random
Cruciferin	3	*C. sativa*	$11.6 \pm 0.3^{a,1}$	$22.6 \pm 1.1^{a,1}$	$28.1 \pm 1.0^{a,1}$	$37.6 \pm 0.5^{a,1}$
		B. napus	$10.7 \pm 1.0^{a,4}$	$25.4 \pm 3.3^{a,2}$	$26.0 \pm 0.7^{a,3}$	$38.0 \pm 3.0^{a,3}$
	7	*C. sativa*	$2.9 \pm 0.2^{b,2}$	$21.1 \pm 0.9^{b,1}$	$24.6 \pm 0.4^{b,2}$	$51.6 \pm 1.2^{b,2}$
		B. napus	$7.6 \pm 0.7^{c,5}$	$39.2 \pm 1.9^{c,3}$	$20.2 \pm 0.9^{c,4}$	$33.1 \pm 1.6^{c,4}$
	10	*C. sativa*	$4.7 \pm 0.6^{d,3}$	$19.8 \pm 1.1^{d,1}$	$24.8 \pm 0.3^{d,2}$	$50.8 \pm 0.6^{d,2}$
		B. napus	$4.8 \pm 0.2^{d,6}$	$18.4 \pm 2.2^{d,4}$	$26.3 \pm 0.6^{d,3}$	$50.5 \pm 1.1^{d,5}$
Napin	3	*C. sativa*	$22.5 \pm 1.2^{a,1}$	NA	NA	$24.9 \pm 1.3^{a,1}$
		B. napus	$24.1 \pm 0.7^{a,3}$	NA	NA	$26.3 \pm 1.3^{a,3}$
	7	*C. sativa*	$27.2 \pm 1.2^{b,2}$	NA	NA	$22.1 \pm 1.3^{b,1}$
		B. napus	$27.5 \pm 1.1^{b,4}$	NA	NA	$26.9 \pm 1.4^{b,3}$
	10	*C. sativa*	$23.6 \pm 0.3^{c,12}$	NA	NA	$28.7 \pm 0.5^{c,2}$
		B. napus	$27.2 \pm 0.7^{c,34}$	NA	NA	$25.4 \pm 0.6^{c,3}$

Means followed by the same superscript within the same column are not significantly different ($p>0.05$); NA= Not Applicable
[a-e] Mean comparison between two species (*C. sativa* and *B. napus*) within the same pH level for each protein type (cruciferin or napin)
[1-6] Mean comparison different pH levels (3, 7 and 10) of the same species for each protein type (cruciferin or napin).

4.5.2 Details of 3˚ structure of cruciferin and napin and the effect of pH

4.5.2.1 Analysis by near UV-Circular Dichroism (near UV-CD)

The CD spectrum in the near UV region (260 to 320 nm) relates to the environment of side chain aromatic amino acid side chains (Kelly et al., 2005) and therefore, provides information about the tertiary structure of cruciferin and napin. The near UV-CD spectra of cruciferin of both *C. sativa* and *B. napus* at pH 7 showed distinct peaks corresponding to phenylalanine (Phe) and tryptophan (Trp) residues (Figure 4.21A). At pH 3, the peaks corresponding to Phe and Trp residues became well resolved, whereas the peak corresponding to Tyr residues were diminished. At alkaline pH, all three peaks were well-resolved and became prominent. These results confirmed that significant changes in the hydrophobic amino acid residue environment had occurred due to changes in the medium pH, especially at pH 3, possibly causing alterations in the tertiary structural conformation of cruciferin.

The napin from both *C. sativa* and *B. napus* showed a distinct peak at pH 7 for Phe residues (Figure 4.21B), but a response for Trp residues was observed only for napin from *B. napus* only. In contrast to cruciferin, napins of both species did not show a peak for Tyr at any of the pHs tested. For both species, the intensities of the Phe and Trp peaks at pH 3, 7 and 10 remained fairly high. Similar to the Phe residues, the signal for Trp residues remained unchanged across the three pHs. This may be an indication that the hydrophobic residue environment of napin did not change drastically with the change in pH.

4.5.2.2 Analysis of surface hydrophobicity (S_0) using ANS fluorescence probe

In the tertiary structure of proteins, the polar and charged amino acid residues are likely to be hydrated, whereas the non-polar residues bond with each other and often form the core of a protein. In globular proteins which are water soluble, these non-polar residues form the hydrophobic core and stabilize the globular folds. Additionally, on the molecular surface, areas or patches of hydrophobic nature exist, and they are very important for interacting with other molecules. Upon unfolding, the hydrophobic amino acid residues buried in the core of the folded protein are exposed to the aqueous environment (Nakai, 1983; Withana-Gamage, 2013).

The results of ANS binding capacity measurements (Table 4.8) showed that the S_0 values of cruciferin were 557.8 and 346.7 for *C. sativa* and *B. napus*, respectively, at pH 7. At pH 3, the S_0 values of the cruciferins were 13 and 27 times as large as at pH 7 for *C. sativa* and *B. napus,* respectively.

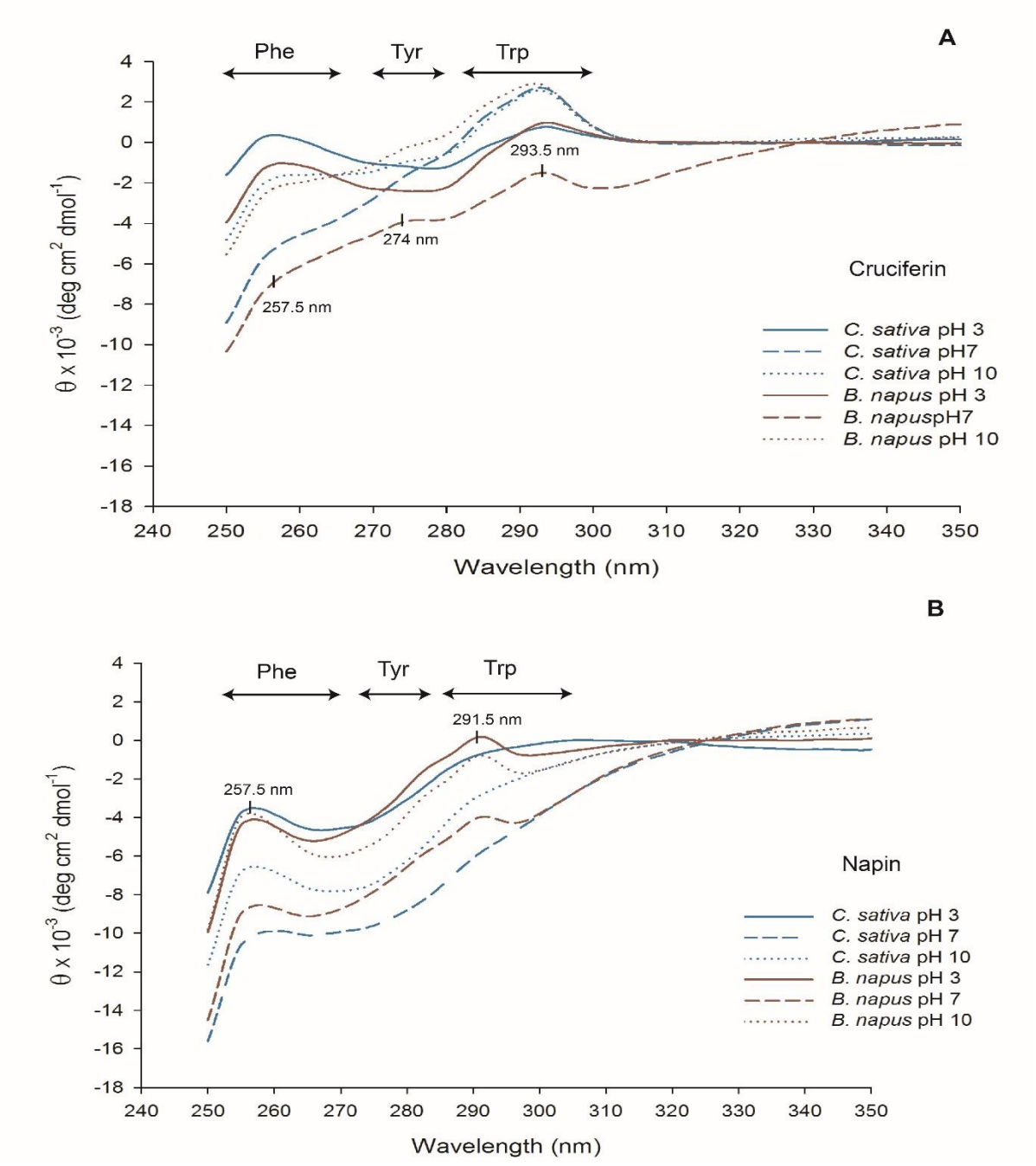

Figure 4.21. Changes in the near UV-CD spectra of purified cruciferin and napin with the changes in the medium pH. (A) *C. sativa* and *B. napus* cruciferin, and (B) *C. sativa* and *B. napus* napin showing peaks corresponding to hydrophobic amino acid residues.

Table 4.8. Surface hydrophobicity of purified cruciferin and napin based on ANS binding capacity. Values are presented as mean ± standard error.

Protein	pH	Species	Surface hydrophobicity (S_0)
Cruciferin	3	*C. sativa*	$7393.1 \pm 32^{a,1}$
		B. napus	$6666.7 \pm 47.2^{b,4}$
	7	*C. sativa*	$557.8 \pm 2.4^{c,2}$
		B. napus	$346.7 \pm 6.4^{d,5}$
	10	*C. sativa*	$266.7 \pm 1.9^{e,3}$
		B. napus	$208.0 \pm 1.3^{e,6}$
Napin	3	*C. sativa*	$363.5 \pm 11.2^{a,1}$
		B. napus	$1239.3 \pm 19.3^{b,3}$
	7	*C. sativa*	$103.5 \pm 3.0^{c,2}$
		B. napus	$103.6 \pm 3.9^{c,2}$
	10	*C. sativa*	$75.5 \pm 1.5^{c,2}$
		B. napus	$150.4 \pm 1.6^{d,4}$

Means followed by same superscript are not significantly different ($p>0.05$)
[a-e] Mean comparisons between two species (*C. sativa* and *B. napus*) within the same pH level for each protein type (cruciferin or napin)
[1-6] Mean comparisons at different pH levels (3, 7 and 10) of the same species for each protein type (cruciferin or napin).

A change in pH to basic (pH 10) resulted in a reduction in S_0 values at neutral pH, and they were smaller than at pH 7 for both *C. sativa* and *B. napus*. Compared to *B. napus*, *C. sativa* cruciferin exhibited significantly higher ($p<0.05$) S_0 values at all three pHs. The results clearly showed that changes in the tertiary structure of cruciferin had occurred at pH 3 compared to pH 7 and pH 10.

The trend of S_0 changes in napin was similar to that of cruciferin for *C. sativa*, but not for *B. napus*. The S_0 values of *C. sativa* napin at pH 7 and 10 were similar. Although *B. napus* napin exhibited a comparatively high S_0 value at pH 3, the values at pH 7 (the lowest) and pH 10 were not as large in magnitude as observed at pH 3. Overall, it can be seen that the pH of the medium, especially acidic pH, had an effect on the tertiary structure of cruciferins and napins from both *C. sativa* and *B. napus*. The changes which occurred at pH 3 were more distinct for both cruciferin and napin when the structural features at pH 7 and 10 are considered. Although the magnitudes of

the S_0 values were slightly different ($p<0.05$), cruciferins from both these species appear to be similar in terms of their tertiary structural features.

4.5.2.3 Analysis of intrinsic fluorescence of Trp residues

Intrinsic fluorescence of Trp residues was evaluated to understand the folding and unfolding behavior of cruciferin and napin. The fluorescence intensity of Trp residues of cruciferin changed with changes in the pH of the medium (Figure 4.22A). When the maximum fluorescence intensity (F_{max}) of cruciferin of *C. sativa* and *B. napus* was plotted against pH, it was clear that a decrease in quantum yield had occurred with a change in pH from alkaline to acidic ($F_{max\ pH10} > F_{max\ pH7} > F_{max\ pH3}$), suggesting that a conformational change had taken place. When the λ_{max} of cruciferin was considered, a red shift (+19 nm and +14 nm for *C. sativa* and *B. napus*, respectively) could be observed with a change in pH from neutral to acidic, but not at alkaline pH (Figure 4.22B). The values of λ_{max} were between 329 nm and 332 nm from cruciferin of both species at pH 7 and 10, whereas it moved to 346 nm (*B. napus*) and 348 nm (*C. sativa*) at pH 3, indicating that Trp residues had a more "polar" environment. The λ_{max} generally increases as protein becomes unfolded (Vivian & Callis, 2001). A better understanding of conformational changes (folding and unfolding) of cruciferin can be obtained from the ratio of the fluorescence intensity (F) at 350 nm to that of 330 nm (F_{350}/F_{330}). The higher the ratio of F_{350}/F_{330}, the more the protein is unfolded (Anonymous, n.d.-b).

Values in Table 4.9 show that the maximum value of F_{350}/F_{330} was obtained at pH 3 for both *C. sativa* and *B. napus*, hence cruciferin may be in a more unfolded state at pH 3 compared to pH 7 and pH 10. The ratios at pH 3 for *C. sativa* and *B. napus* were not significantly different ($P>0.05$), therefore, it can be assumed that the degree of unfolding of cruciferin may be similar for both species. *C. sativa* cruciferin may have exhibited a higher degree of unfolding at pH 10 (0.85 ± 0.02) than at pH 7 (0.71 ± 0.05), but this was not the case for *B. napus* cruciferin where the F_{350}/F_{330} was 0.81 ± 0.01 at both pH 10 and pH 7. The changes observed in cruciferin for the λ_{max} values and F_{350}/F_{330} ratios at pH 3 and pH 10 compared to pH 7 coincided and confirmed the changing hydrophobic residue environment of the molecule at pH 3. Napin did not provide comprehensive and consistent results for intrinsic fluorescence may be due to the low availability of Trp or Tyr residues. As a result, this technique could not be utilized successfully to evaluate changes in tertiary structure of napin.

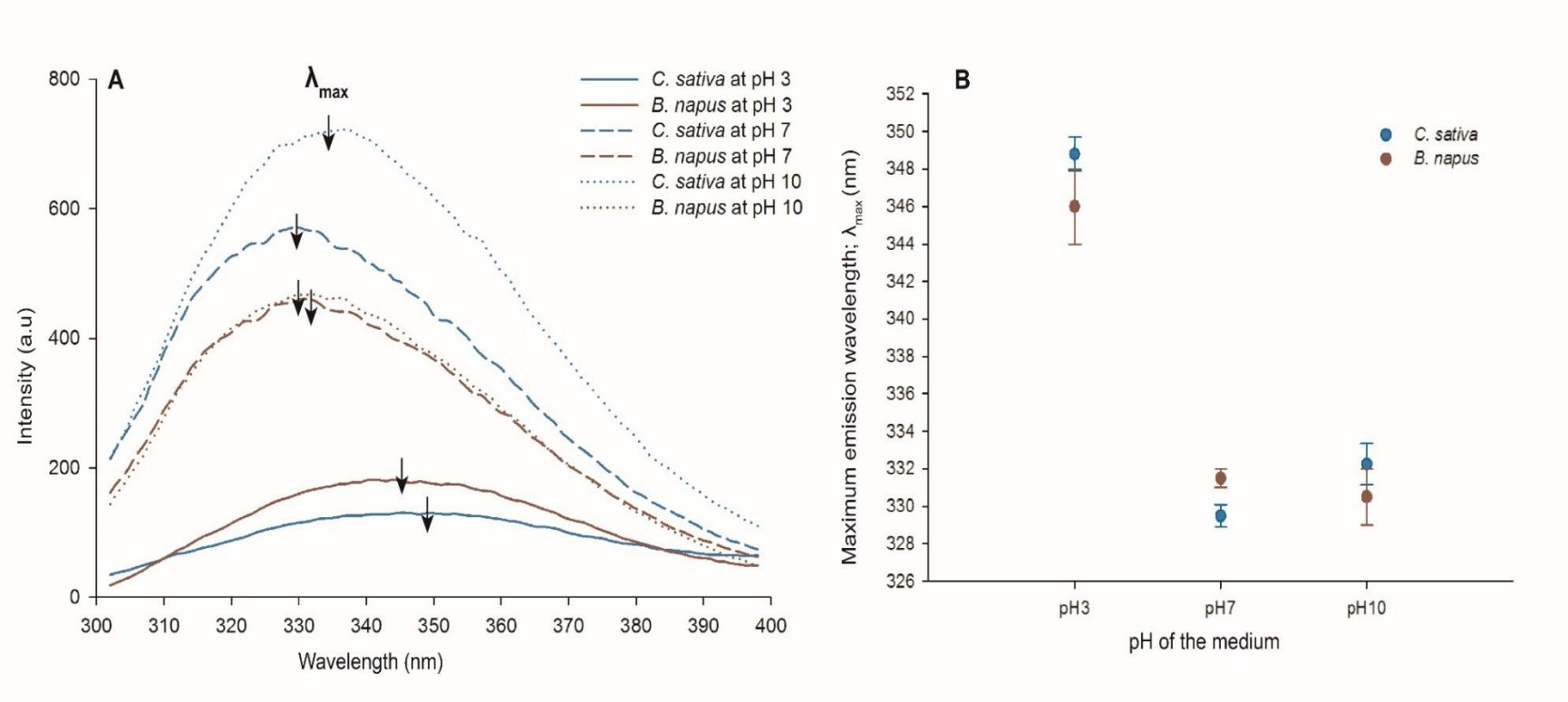

Figure 4.22. Tryptophan fluorescence of purified cruciferin from *C. sativa* and *B. napus* at ambient temperature (22°C). (A) Emission spectra at pH 3, pH 7 and pH 10 and (B) Emission maximum (λ_{max}) at the same pHs. All spectra were recorded at an excitation wavelength of 280 nm.

Table 4.9. Change in F_{350}/F_{330} ratios of *C. sativa* and *B. napus* cruciferin at varying pH. Values are presented as mean ± standard error

pH of the medium	Cruciferin source	
	C. sativa	*B. napus*
3	1.13 ± 0.00	1.11 ± 0.02
7	0.71 ± 0.05	0.81 ± 0.01
10	0.85 ± 0.02	0.81 ± 0.01

4.5.3 Thermal properties of cruciferin

4.5.3.1 Analysis of the effect of pH by differential scanning calorimetry (DSC)

Differential scanning calorimetry (DSC) provides information on the structural stability of a protein in dilute solution as reflected by changes in the partial molar heat capacity at constant pressure. Changes in the heat capacity of a protein showcase its ability to absorb heat. Heating of a dilute protein solution causes the protein molecules to absorb heat energy (Anonymous, n.d.-a). Absorbed heat induces unfolding of the protein molecules over a temperature range characteristic of the protein, while generating an endothermic peak (referred as the denaturation peak) which is detected by the DSC. Integration of the heat capacity of the endothermic peak provides the enthalpy of the unfolding process caused by the endothermic phenomena, such as breaking of H bonds, and exothermic phenomena such as damaging hydrophobic interactions (Anonymous, n.d.-a). Once unfolding is complete, heat absorption decreases. The thermal denaturation of cruciferin and napin evaluated using DSC showed that distinct denaturation peaks were observed for cruciferin from both *C. sativa* and *B. napus* at both pH 7 and 10, but no denaturation peak was observed at pH 3 (Table 4.10). The peak denaturation temperature (T_m) of *C. sativa* at pH 7 (80.6 ± 0.1°C) was significantly different ($P<0.05$) than that at pH 10 (83.1 ± 0.5°C), but the denaturation enthalpy was not. However, both the T_m and the enthalpy of *B. napus* cruciferin did not differ with pH (pH 7 and pH 10). The onset of denaturation of *C. sativa* cruciferin occurred at 60-65°C and ended at 90-95°C. *B. napus* cruciferin exhibited a denaturation onset at 65-70°C, which ended at 95-100°C. At pH 7, the denaturation temperature of *C. sativa* cruciferin was significantly different ($P<0.05$) than that of *B. napus*, but this was not the case at pH 10.

Table 4.10. Thermal denaturation information obtained from DSC analysis of cruciferin from *C. sativa* and *B. napus*. Values are presented as means ± standard error.

pH	Protein	Denaturation Temperature (T_m) (°C)	Enthalpy (J/g)	Onset of peak (°C)	End of peak (°C)
3	*C. sativa*	No peak was observed from 30-130°C for both the species			
	B. napus				
7	*C. sativa*	80.6 ± 0.1^{a}	0.9 ± 0.1^{a}	60-65	90-95
	B. napus	83.2 ± 0.8^{b}	1.1 ± 0.3^{a}	65-70	95-100
10	*C. sativa*	83.1 ± 0.5^{b}	1.0 ± 0.1^{a}	60-65	90-95
	B. napus	84.8 ± 0.2^{b}	0.9 ± 0^{a}	65-70	95-100

Means followed by same superscript within the same column are not significantly different ($p>0.05$)

Napins from both seed species did not exhibit any thermal transition peaks between 30°C and 100°C at any pH (pH 3, 7 or 10) investigated. Above 100°C, the hermetically sealed DSC pans burst and did not provide a convincing denaturation peak for napin. However, the DSC pans containing cruciferin remained stable over the entire temperature ramp from 30°C to 130°C.

4.5.3.2 Analysis of the effect of temperature and pH by intrinsic fluorescence of Trp residues

Intrinsic fluorescence of cruciferin at selected temperatures (T), ambient T (22°C), T of onset of denaturation (60-70°C), average T_m (83°C) and end T of denaturation (95°C), at pH 3, 7 and 10 was investigated. At T_m, both *C. sativa* and *B. napus* exhibited F_{350}/F_{330} values greater than 1 ($F_{350}/F_{330} >1$), showing a high degree of unfolding of cruciferin at that temperature, which proceeded until the end of denaturation temperature (Figures 4.23C-F). Cruciferin from both species exhibited a decrease in the maximum emission intensity and an increase in the maximum emission wavelength (λ_{max}, Figure 4.24), indicating possible structural unfolding events with increasing temperature. On the other hand, the structure of cruciferin from both species were less unfolded ($F_{350}/F_{330} <1$) at the onset of denaturation and could be assumed to be minimally unfolded at ambient temperature at pH 7 or pH 10 (Figures 4.23C-F). At ambient temperature, the F_{350}/F_{330} values for cruciferin of both species were above 1 at pH 3. This was an indication of an unfolded cruciferin structure even before any increase in temperature due to sample heating (Figure 4.23A and B). Moreover, the maximum fluorescence intensity values for cruciferin from both species at pH 3 and $\geq T_m$ at pH 7 and 10 were similar (Figures 4.23A-F). This confirms acid-induced structural unfolding of cruciferin at pH 3.

4.5.4 Solubility properties of cruciferin and napin and the effect of pH

The protein solubilites of cruciferin and napin at pH 3, 7 and 10 were evaluated using the bicinchoninic acid (BCA) assay. At the concentration used for the study (1 mg/mL), both *C. sativa* and *B. napus* cruciferin showed increases in solubility as the pH was increased from 3 to pH 10, with the maximum solubility value at pH 10 (100% for *C. sativa* and 95% for *B. napus,* Figure 4.25A). The solubility of *C. sativa* cruciferin was significantly higher than that of *B. napus* at the pHs studied.

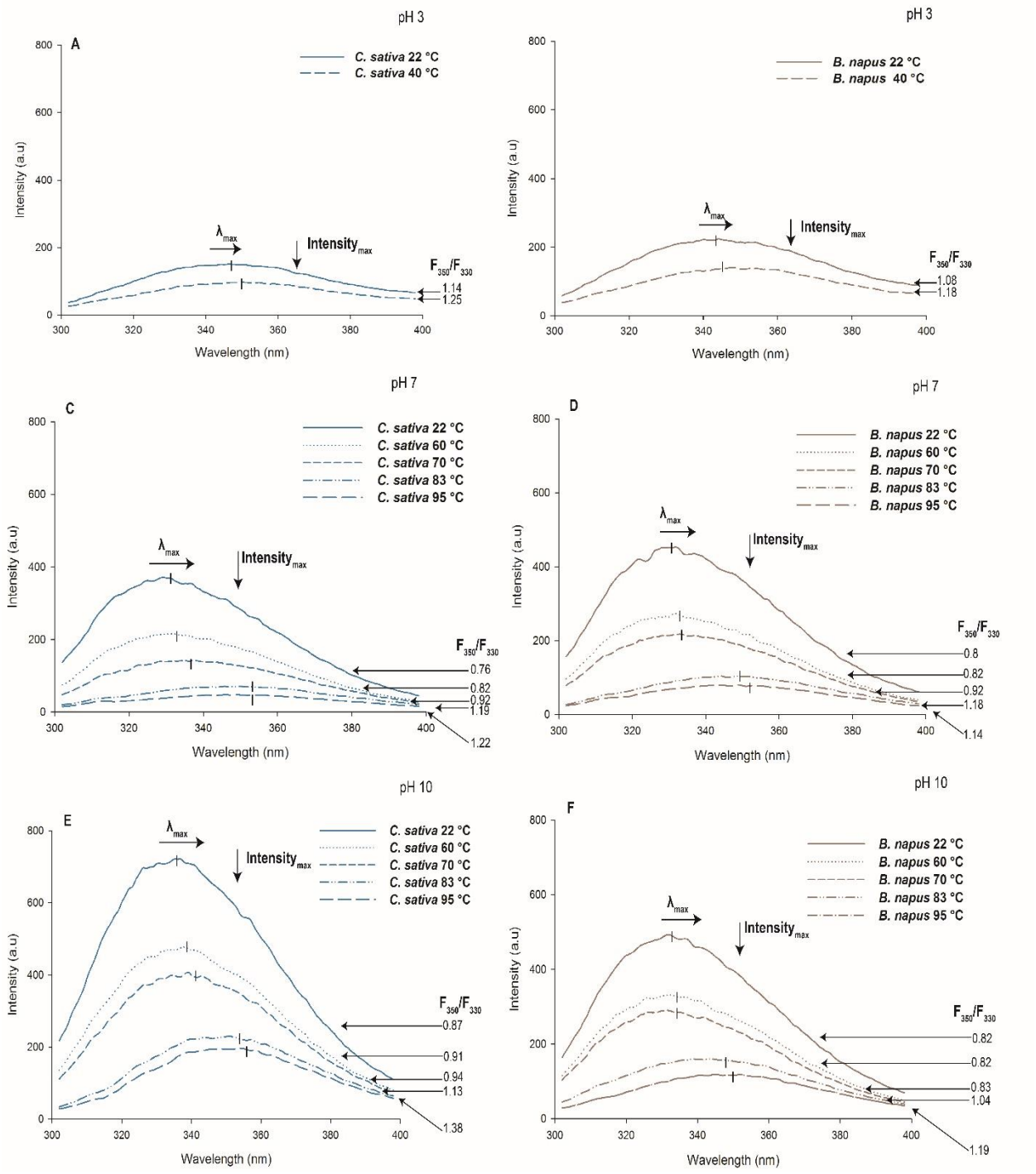

Figure 4.23. Tryptophan fluorescence and $F_{350/330}$ ratio of purified cruciferin from *C. sativa* and *B. napus* at different pHs and temperatures. (A) *C. sativa* at pH 3; (B) *B. napus* at pH 3; (C) *C. sativa* at pH 7; (D) *B. napus* at pH 7; (E) *C. sativa* at pH 10 and (F) *B. napus* at pH 10.

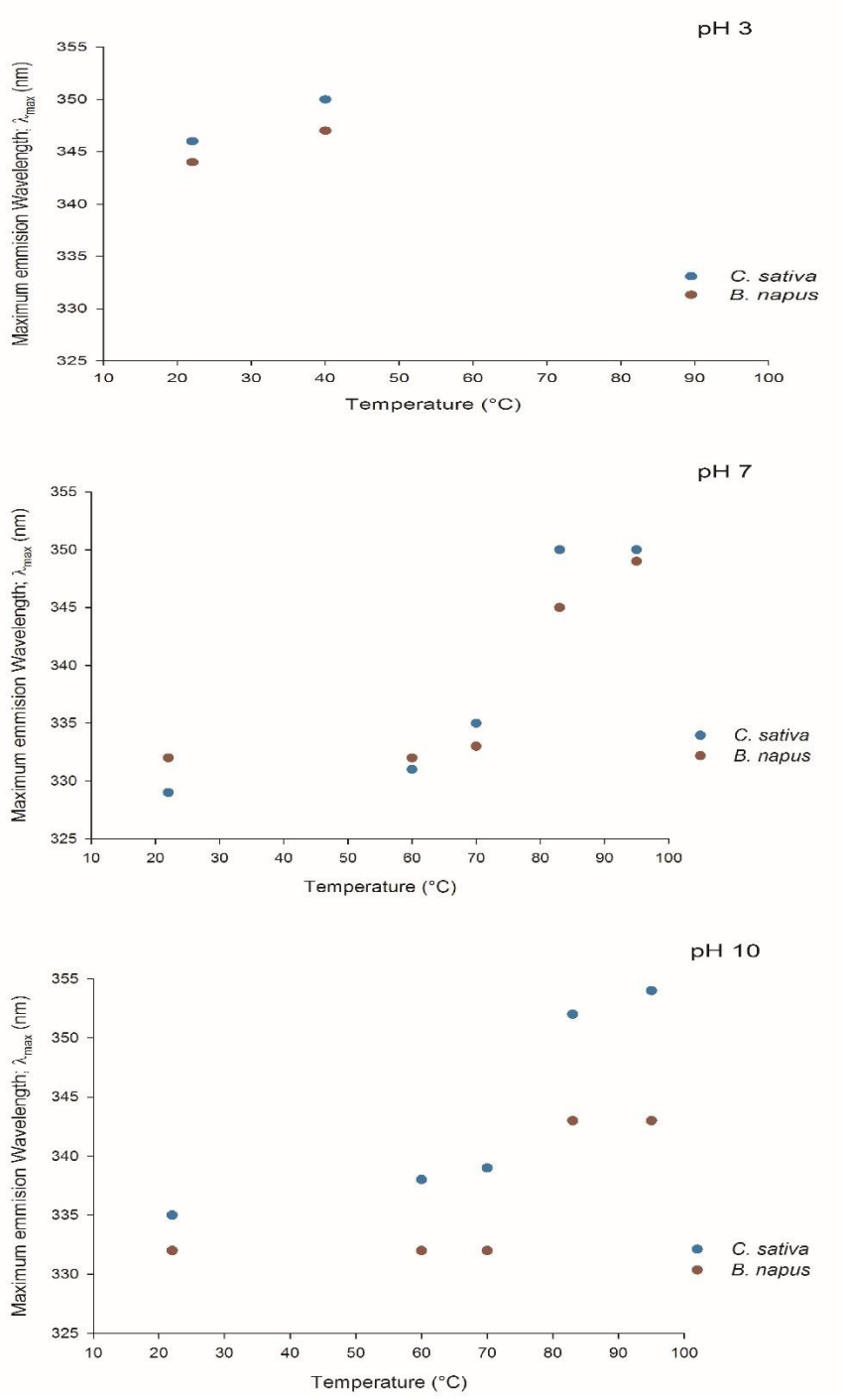

Figure 4.24. Changing pattern of maximum emission wavelength (λ_{max}) of tryptophan fluorescence of *C. sativa* and *B. napus* cruciferin in repose to pH and temperature change. (A) at pH 3; (B) at pH 7 and (C) at pH 10.

Napin from *C. sativa* exhibited a similar trend for the solubility with increasing pH as did cruciferin, but this was not the same for the napin from *B. napus* (Figure 4.25B). At the three pHs, *C. sativa* napin exhibited lower solubility values than did cruciferin, except at pH 10. *B. napus* napin exhibited a decreasing solubility trend from pH 3 to pH 10; pH 10 exhibited the lowest value. Cruciferin and napin from both species exhibited over 70% solubility at pH 3, 7 and 10.

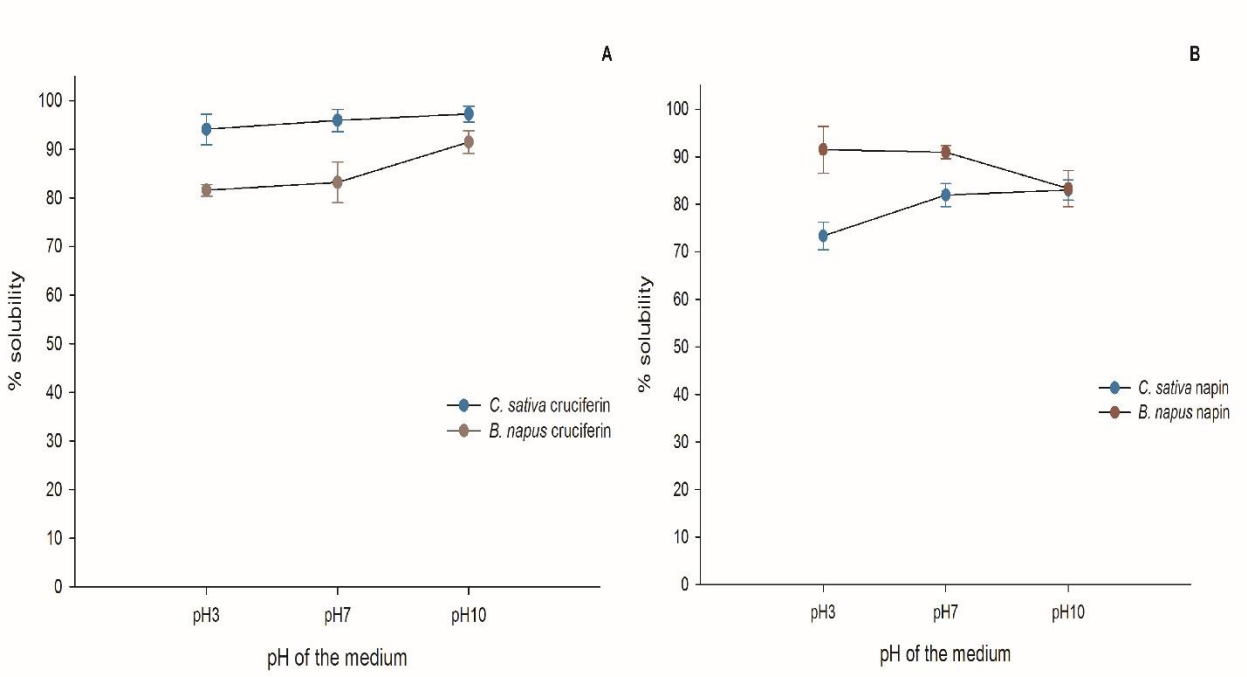

Figure 4.25. Solubility of purified cruciferin and napin from *C. sativa* and *B. napus* in response to changing medium pH. (A) Cruciferin from *C. sativa* and *B. napus* and (B) napin from *C. sativa* and *B. napus*.

5. DISCUSSION

The present study investigated the new cruciferous oilseed *Camelina sativa* (L.) Crantz (camelina) in relation to differences in the chemical constituents of the seed, the components of the storage and oil body proteins, and the structural features of the storage proteins, compared to the well-established oilseed crop *Brassica napus* L. (canola). The responses of the storage proteins to changes in pH and temperature were studied in order to understand some of the changes that camelina and canola proteins may undergo during oil extraction (involving temperatures above 100°C) and protein extraction (involving aqueous conditions and pH changes).

5.1 Comparison of seed and meal composition

5.1.1 Botanical relationship of *C. sativa* and *B. napus*

The principal chemical constituents of *Brassicaceae* oil-containing seeds are oil, protein, cell wall carbohydrates, and secondary metabolites such as aliphatic, indole and aromatic glucosinolates, phytates and phenolic compounds (mainly sinapic acid derivatives) and it was expected that both canola and camelina would contain all of these. The high content of oil and the nutritionally compatible fatty acid profile for human food applications has made canola a very successful vegetable oil crop for the past 35 years. The oil-free meal of canola is a protein and energy source for animal nutrition, with the potential to be developed into a human food protein (Wanasundara et al., 2015). Although camelina is known to prairie farmers as an unwanted plant (weed) in crop fields, a decade of breeding and agronomic research has made it an oilseed crop that is dedicated to providing renewable oil feedstock for industrial purposes, primarily for bio-diesel and bio-lubricants. Camelina oil is rich in linolenic acid (18:3 ω3, Table 2.1), therefore camelina seed is a good source of omega-3 fatty acids. With the approval from Health Canada for camelina oil for human consumption (Health Canada, 2012), cold-pressed camelina oil is available in the Canadian market as an edible oil. Similar to canola, the meal of camelina can be utilized as a protein source in feed rations and other protein-enriched bioproduct development. Camelina meal has been approved for use in poultry, swine and dairy cow rations. (SMA, n.d; Kim & Netravali, 2012; Reddy et al., 2012; Li et al., 2015).

Camelina has a close genetic relationship with *Arabidopsis thaliana* (lineage 1 of *Brassicaceae*) and is distantly related to canola (lineage II) (Kagale et al., 2014). Although camelina and canola are not closely related phylogenetically, they are unique oilseed crops in the *Brassicaceae* family and could be economically analogous. Uses for camelina seed and its components are still at the developmental stage as compared to canola. This study is focused primarily on understanding the types, structure and some key physicochemical properties of the major seed proteins of these two crops, while investigating other related components. The results obtained in this study on chemical composition and protein profiles, types and structural properties support the hypothesis that a close relationship exists between the proteins of the two crops.

5.1.2. Microstructure of cotyledon cells

In canola seed, the cytoplasm of the cotyledons, radical and aleurone layer cells host separate compartments containing oil and protein (Hu et al., 2013). The vacuoles in these cells turn into sub-cellular structures that harbour macromolecules and sequester and inactivate toxic compounds and secondary metabolites (Marty, 1999; Bethke & Jones, 2001). Protein storage vacuoles (PSVs) or protein bodies store protein to be used later as a source of reduced N and are surrounded by a tonoplast membrane. Within PSVs, three distinct regions (matrix, crystalloid and globoid) have been identified and the proteins are stored primarily in the matrix and crystalloid regions, whereas phytic acid crystals are found in the globoid region (Neumann & Weber, 1978; Lott, 1980; Weber & Neumann, 1980). In canola and camelina cotyledon cells, the PSVs contain primarily matrix and globoid region and the crystalline areas cannot be identified distinctly (Unpublished data obtained from transmission electron microscopy and immune electron microscopy). The PSVs of camelina are more even in shape and size compared to those of canola (Figure 4.8). Fewer and uneven PSVs were observed in canola cotyledon cells. However, one cannot generalize because only one genotype of canola was studied. It can be hypothesized that for a smaller seed (compared to canola) to store more protein (Table 4.1), either the size or the number of PSVs should increase. In this genotype of camelina, more PSVs were found, therefore more protein is packed within a cell. Camelina seed and the PSVs are small in size compared to canola (Figures 4.8A and B).

As oil-storing seeds, both canola and camelina have oil stored in cotyledon cells, in compartments called oil bodies (OBs). An OB consists of a triacylglycerol matrix which is surrounded by a phospholipid layer and surface proteins (Huang, 1996; Tzen, 2012). An OB may

be 0.5-2.5 µm in diameter (Huang, 1992; Tzen, Cao, Laurent, Ratnayake, & Huang, 1993; Peng & Tzen, 1998), and in canola, OBs between 0.2 µm and 3.0 µm in diameter have been reported (Katavic, Agrawal, Hajduch, Harris, & Thelen, 2006). The TEM images of canola OBs obtained in the present study (Figures 4.8B and 4.9B) confirmed their spherical shape and average diameter of ~0.43 µm. Interestingly, OBs of camelina have an average diameter of ~ 0.68 µm, which is larger than that of canola OBs, and they are less spherical and more uneven in shape. The other significant difference is that OBs are dispersed uniformly and packed loosely (~3 OBs/μm^2) in the cell matrix of camelina, whereas OBs are squeezed to the edges of the cell and packed densely (~5 OBs/μm^2) in canola. The large protein bodies in canola may have forced dense packing of OB (Figure 4.8B; Hofsten, 1974). Hu et al. (2013) studied the ultra-structure of oil-enriched (64.5% oil content) canola seeds in comparison with regular canola seeds, and noted that most of the cytoplasm area was covered by OBs and the remainder by PSVs. According to this study, the cytoplasm of cotyledon cells harboured nearly 81% of the OBs, whereas it is only 33-38% in low-oil-containing seeds. Therefore, to accommodate more oil, more area in the cotyledon cytoplasm has been used at the expense of protein. This study provides evidence for a cellular basis for the negative relationship between the content of seed protein and oil of canola.

5.1.3 Mucilage of camelina

The term mucilage is used for soluble polysaccharides mainly comprised of pectin and non-pectic components such as cellulose, hemicellulose (xylan, xyloglucan, galactoglucomannan) and arabinogalactans (Western, Skinner, & Haughn, 2000; Pekel et al., 2009; Sun, Tan, Baskin, & Baskin, 2012; Voiniciuc et al., 2015). Mucilage-laden cells in the outermost seed coat are found in several species of the *Brassicaceae* family. Similar to camelina, *Brassicaceae* plants, such as *A. thaliana* and *S. alba*, also possess a mucilage-rich seed coat (Cui, Eskin, Wu, & Ding, 2006; Macquet, Ralet, Kronenberger, Marion-Poll, & North, 2007). Seed coat mucilage is one of the chemical traits of camelina that distinguishes it from canola. Studies on seed coat development in *A. thaliana* (also contains mucilage) revealed that cells of the seed coat epidermis (SCE) synthesize a primary wall and then secondary walls deposit sequentially in two distinct events. Hydration of dry seed causes the mucilage cells to swell and burst, rupturing the primary wall and forming a gelatinous capsule around the seed (Gutterman & Shem-Tov, 1996; Western, et al., 2000; Western, 2012). This swollen mucilage separates into a non-adherent layer that can be easily detached from the seed and an adherent layer that is difficult to remove (Voiniciuc et al., 2015). This was clearly

observed in camelina. Mucilage present in the seed coat of mature camelina rapidly rehydrates as soon as the seed is in contact with water, resulting a swollen, jelly-like layer (a halo) around the seed (Figure 4.1B). Hydrated camelina mucilage is highly viscous. According to Huang and Gutterman (1999) and Huang, Gutterman & Osborne (2004), the primary role of mucilage is increasing the surface area of small-sized seeds to improve uptake of soil water, especially in dry, arid environments, which may be related to low-input cultivation, such as is known for camelina (Eynck & Falk, 2013).

According to Zubr (2010), the mucilage content of camelina is ~ 6.7% of the total seed weight. Mucilage is another economically valuable biopolymer of this oilseed besides oil and protein. The seed coat mucilage of oil-rich seeds, such as yellow mustard (*S. alba*) and flax (*Linum usitatissimum*), is utilized as gums and stabilizers in food systems (Cui et al., 2006). Unlike oil, mucilage and proteins share similar properties. The hydrophilic nature of mucilage makes it highly soluble in aqueous medium, causing co-extraction along with proteins and consequently increases the viscosity of aqueous solutions.

Due to the hydrophilic nature of mucilage, protein extracts obtained from whole ground camelina seed (or oil-free meal) are heavily contaminated with soluble polysaccharides, making it impossible to obtain protein of reasonable purity. In addition, polysaccharides block separation membranes and chromatographic columns. Mucilage contamination with protein can be avoided either by removing or reducing the mucilage content of the seed coat or removing the seed coat. Preliminary experiments carried out by soaking camelina seed in solutions containing $NaHCO_3$ or in enzyme formulations containing pectinase, cellulase or mixed carbohydrases (e.g.,Viscozyme®), showed that Viscozyme® treatment was highly effective in degrading mucilage and made it water soluble. Viscozyme® is an enzyme complex derived from *Aspergillus* spp. that has multiple activities: arabinase, cellulase, β-glucanase, hemicellulase and xylanase (Wanasundara & Shahidi, 1997). It appears that camelina mucilage is complex and different in composition than other seed mucilages, such as those from flax, yellow mustard or chia (Cui, Eskin & Biliaderis, 1993; Marambe, Shand & Wanasundara, 2008; Muñoz, Cobos, Diaz & Aguilera, 2012; Ziolkovska, 2012). However, no detailed compositional analysis is available for camelina. With optimization of the duration of soaking (3h) and the dosage of enzyme (0.1 mL/g of seed) and with vigorous stirring, Viscozyme® pre-treatment as employed in this study was able to remove mucilage from the seed surface (Figure 4.1C). The swollen mucilage layer around the seed was

not clearly visible. Microscopic examination of rehydrated enzyme-treated seed provided convincing evidence of the effectiveness of mucilage removal.

5.1.4. Proteins of meal

5.1.4.1 Protein content and types

When the protein (%N × 6.25) levels of de-oiled seed materials were compared (Table 4.1), the reduced mucilage content may have been the reason that Viscozyme®-treated camelina seed gave a higher value than canola (untreated). Russo (2012) studied the meal protein content of nine camelina genotypes, including Calena and Ligena cultivars grown in a single location in Italy, during the fall and spring for two years, and reported that values in the range of 32.1-36.0%, whereas meal from camelina grown in a single year during the summer under different climatic and soil conditions in Scandinavia was reported to contain 42.5% protein (dwb; Zubr, 2003b). According to Jiang, Caldwell & Falk (2014), the protein content of camelina could vary due to genotype and environmental conditions, which can be observed in the values reported by different authors: 33% protein content (dwb) in camelina seed for early research (1950-1960) conducted in Canada (Plessers, McGregor, Carson, & Nakoneshny, 1962); 39% protein (as-fed basis) in meal from camelina (Calena) cultivated in Canada (Hixon et al., 2015); and 32.4% protein (dwb) in meal from camelina grown in the USA (Li et al., 2014). The camelina meal protein content reported above is comparable to the available values for canola meal (Newkirk, 2015). However, the protein content (51.5% dwb, Table 4.1) obtained for camelina in the present study was much higher than the values reported in the literature.

One dimensional electrophoresis (1DE) showed that camelina and canola contain several protein types, but there are many similarities between the two seed types. The prominent polypeptide bands of canola meal were characteristic of cruciferin, napin and oleosin (Figure 4.2). Electrophoretic separation of proteins from canola protein storage vacuoles by Nietzel et al. (2013) showed polypeptide bands ranging from 12-60 kDa that were clearly visible and could be identified under non-reducing conditions. Using antibodies specific for the α- and β-chains of cruciferin, and with intact S-S bonds, cruciferin in the molecular weight range of 55-60 kDa, α chain at ~30 kDa, and β chains in the 20-25 kDa range were identified by this research group. Separation of the S-S bonds dissociated the polypeptide bands under reducing conditions, confirming that the α- chain polypeptides were in the 27-33 kDa and the β- chain polypeptides in the 18-20 kDa range. Nietzel et al. (2013) confirmed that the polypeptide bands observed close to

50 kDa and which disappeared when S-S bonds were broken represented unprocessed cruciferin monomers, and the polypeptide bands observed in the 20-35 kDa range under both non-reducing and reducing conditions were free α- and β-chains. A similar polypeptide profile for canola was reported elsewhere (Aluko & McIntosh, 2001; Wanasundara, 2011). Therefore, it can be considered that camelina contains cruciferin similar to that of canola, because of their botanical relationship, *i.e.* their being in the same family *Brassicaceae*.

Although Nietzel et al. (2013) did not investigate further the 12 kDa polypeptide bands, the absence of cruciferin α- and β-chains in this molecular weight range was confirmed in the present study. Hence, the 12 kDa band was not related to cruciferin. In the present study, the low molecular mass polypeptide band of ~14 kDa disappeared under reducing conditions, resulting in two low molecular mass peptide bands of ~9-10 kDa (Figure 4.2) related to napin in both canola and camelina. The molecular mass of canola napin polypeptides was reported to be in the range of 12.5-14.5 kDa (Monsalve & Rodriguez, 1990) and ~10- 4.5 kDa (Gehrig & Biemann, 1996; Gehrig, Krzyzaniak, Barciszewski, & Biemann, 1998). The presence of a single polypeptide band of ~14 kDa and its degradation into two low molecular mass polypeptide bands confirmed the presence of napin in camelina, similar to canola.

The polypeptide bands observed in the molecular mass range of 15-20 kDa in both camelina and canola meal, and that did not change under reducing conditions, indicated no S-S bond involvement in stabilizing these protein molecules (Figure 4.2). It is highly likely that these polypeptides represent oleosin, which is an OB protein. According to Wijesundera et al. (2013), polypeptide bands with a molecular mass of 19-20 kDa were oleosins of canola meal. Therefore, it was assumed that the polypeptide bands evident at ~15-20 kDa in both camelina and canola meal may represent oleosins.

Since similar polypeptide profiles and protein types were identified from 1DE, comparable amino acid (AA) profiles were expected for camelina and canola (Table 4.2). The amino acid profiles of camelina DH55 and canola DH12075 were indeed similar and comparable with values available in literature. Of the amino acids that comprise camelina protein, essential and non-essential amino acids constitute ~40% and 60%, respectively. The essential amino acid content of canola protein was higher (>40%) than that of camelina. Similar trends in the amino acid composition of defatted camelina and canola meal were observed by Li et al. (2014). According to Russo (2012), the AA score of camelina meal protein is close to100, making it a good quality

protein. Lysine is often considered the first limiting essential amino acid in crucifers; the biological value (BV) of camelina meal was reported to be 98, in which lysine was the limiting amino acid. On the other hand, rapeseed meal showed a BV of 100 with a balanced amino acid profile (Russo, 2012). Lysine is affected by commercial oil extraction processes (Newkirk, Classen & Edney, 2003). It was found that the lysine content of canola expeller cake or desolventized toasted meal was generally 9-10% lower than that of seed, as elevated temperature induces lysine participation in Maillard-type reactions (Newkirk et al., 2003). The nutritional significance of camelina meal protein as an animal feed has been reported already by different research groups (Cherian, et al., 2009; Aziza, et al., 2010; Cappellozza, et al., 2012; Hixson & Parrish, 2014; Kahindi et al., 2014; Hixson, Parrish, Wells, Winkowski, & Anderson, 2015, Hixson et al., 2015). So far, incorporation up to 10% in diets for beef cattle and broiler and laying chickens, and 2% in swine diets has been approved in the USA, while incorporation up to 12% in broiler chicken feed has been approved in Canada (SMA, n.d). The content of sulfur-amino acids in camelina protein is comparable to that of canola protein (~5%), which brings a competitive nutritional advantage for these proteins compared to those of legumes (Dozier & Hess, 2011).

Apart from the nutritional benefits as a feed ingredient, there is potential for amino acids to be utilized in other applications, such as in pharmaceuticals, surfactants, amino acid supplements, sweeteners, herbicides and synthetic leathers (Fujimoto, Koiwa, Nagaoka, & Tatsukawa, 1972; Clapés & Infante, 2002; Shimomura et al., 2006; Ivanov, Stoimenova, Obreshkova, & Saso, 2013). For instance, camelina meal contains ~17% glutamic acid and ~15% branched-chain amino acids (Table 4.2) that can be utilized in synthetic leather and protein supplement production, respectively. As an emerging industrial oilseed crop, exploitation of potential diverse applications for protein/amino acids is important for value addition. Knowledge on camelina protein is limited, also it has not been considered for human consumption yet. Extensive research is required to develop camelina to be utilized as a protein source for food applications. Considering the protein profiles and amino acid composition, camelina meal proteins do not deviate much from those of canola, and therefore have the potential for developing products similar to commercial canola protein isolates, such as Supertein™ and Puratein® (http://www.burcon.ca/), and Isolexx™ (http://teutexx.com/).

5.1.4.2 Protein solubility with changes in pH

The polypeptide profiles of camelina and canola showed that the meal is a mixture of several proteins, hence differences in the solubility of meal protein in aqueous solution as a function of pH is to be expected. The *Brassicaceae* oilseed meal proteins exhibited solubility curves with changing pH typical of those of soybean, sunflower and many other seed meal proteins (Berk, 1992; Abeysekara, 2102; Wanasundara & McInstosh, 2013). Figure 4.3A showed that camelina and canola share similar trends of protein solubility with changing pH. The minimum meal protein solubility was observed at pH 4.5 for both species, which is the apparent isoelectric point (pI) of the soluble proteins of camelina and canola meal. Studies, including those by Dendukuri & Diosady (2003) and Marnoch & Diosady (2006) employed this pH to precipitate protein from mustard and other crucifers and referred to it as the isoelectric pH. The studies by Wanasundara et al. (2011) and Wanasundara & McIntosh (2013) reported that the minimum solubility for canola protein occurs between pH 3 and pH 5, where some napin and most of the cruciferin precipitate. The polypeptide profiles of the soluble protein also confirmed that canola napin remained soluble at pH 4.5 (Figure 4.3C). Protein molecules have multiple charges depending on the exposed amino acid residues and attachments, such as lipids, sugars, metal ions and proteins. At the isoelectric pH, the molecular charge assumes neutrality. If all soluble proteins of canola had pIs at 4.5, no protein would remain soluble at this pH. However, the opposite was found in this study. Theoretical pI values for cruciferin and napin are 7.2 and 11.0, respectively, based on amino acid composition, and fairly close values have been obtained experimentally (Schwenke, Schultz, Linow, Gast, & Zirwer, 1980; Crouch, Tenbarge, Simon, & Ferl, 1983). There are no studies available on how the different protein types, cruciferin and napin, exist in PSVs, whether associated as a complex or accumulated independently in the matrix areas of PSVs. The TEM studies did not indicate distinguishable crystalloid areas of PSV. Most likely both cruciferin and napin co-exist in the PSV. Results of IEF (isoelectric focusing, the first dimension of 2DE) in the present study clearly indicated that purified napin from both camelina and canola separates into several isoforms that exhibit pIs between pH 9 and pH 12, confirming that napin has an alkaline pI (Figure 4.14). The cruciferin pI was observed around pH 7, and it was confirmed by 2DE analysis of purified protein (Figure 4.11). With this evidence, it can be hypothesized that cruciferin and some of the napin in the seed of these two species exist in association, and exhibit minimum solubility at pH 4.5. In other words, this proposed cruciferin-napin complex has a pI

around 4.5. Non-complexed napin remains soluble at this pH. However, according to Wanasundara & McIntosh (2013), most of the napin of canola, *B. juncea* and *S. alba* can be solubilized between pH 3 and pH 4 with NaCl or $CaCl_2$ as an additive in the aqueous, acidic medium. This indicates that the cruciferin-napin complex can be dissociated by the addition of neutral salt ions, allowing cruciferin to be complexed with another molecule/salt and maintaining its insolubility under these conditions. Another situation could be cruciferin complexes with components, such as phytic acid, which, depending on the pH, form charge-altered cruciferin-phytate complexes that exhibit pIs near 4.5. Proteins generally exhibit increased solubility as the medium pH moves away from the pI, and the lowest solubility near the pI (Pace, Trevino, Prabhakaran, & Scholtz, 2004). The protein solubility of both meals increased as the pH of the medium moved toward alkaline, with the highest value at pH 12 (Figure 4.3).

5.1.5 Minor constituents of meal

Besides the macromolecules, such as oil and protein, other minor chemical compounds are found in camelina and canola. *Brassicaceae* family plants are known to contain glucosinolates. Canola seed is known to contain glucosinolates, phenolics (sinapine and tannin) and phytates (Russo, 2012; Tan, Mailer, Blanchard, & Agboola, 2011b). These compounds are considered antinutrients and may pose adverse effects on animal and human nutrition, including reduced palatability and lower nutrient digestibility and availability. In addition, the reactive groups of phenolics and glucosinolate-breakdown products are known to associate with proteins, causing undesirable dark colours and functionality changes of protein products from canola (Aider & Barbana, 2011; Tan et al., 2011a; Wanasundara, 2011; Hixson & Parrish, 2014; Hixson, Parrish, Wells, Winkowski, & Anderson, 2015). The phytic acid content of the camelina DH55 line was found to be higher than that of the canola DH12075 line (Table 4.1). A positive correlation was observed between the contents of protein and phytic acid of the meals tested in this study. Since camelina contained a significantly higher amount of protein compared to canola, a higher phytic acid content can be expected in camelina, and *vice versa*. Since both protein and phytic acid are stored in PSVs, the greater number of PSVs in the cotyledon cytoplasm in camelina may relate to a higher globoid number or areas where phytic acid is accumulated. Some evidence for this was found in the camelina seed microstructure assessment (Figure 4.8).

Sinapic acid is the predominant polyphenolic compound found in camelina (Abramovič et al., 2007). Therefore, the total phenolic content of camelina was expressed as mg sinapic acid

equivalent per gram of meal; lower values were reported for camelina meal compared to canola (Table 4.1) and values were comparable with literature values (Table 2.3). The glucosinolates of camelina were not investigated in this study. However, the available literature showed that the total glucosinolate content is higher in camelina than in canola (Table 2.3), which would cause a negative impact on camelina meal utilization in food and feed applications. As a new oilseed, more information on the minor constituents of camelina is very important for devising strategies to lower their levels and improve meal and protein utilization, either by processing or by germplasm enhancement.

5.2 Detailed information on major proteins

5.2.1 Storage proteins

The identity and purity of the purified cruciferin and napin were evaluated using 1DE. The resulting polypeptide bands showed proteins with different molecular weights and provided an idea about the presence of non-target proteins with different molecular weights in the purified cruciferin or napin samples (Figures 4.6D and E). However, it is indeed possible that even polypeptide bands typical for cruciferin and napin also may contain contaminating proteins that might share similar molecular weights and cannot be distinguished by 1DE. Therefore, 2DE followed by LC-MS/MS analysis was performed to obtain insight into purified cruciferin and napin in terms of confirming the presence of cruciferin, napin and their isoforms and possible contaminating proteins, and their abundance.

5.2.1.1 Cruciferin

It is highly unlikely that canola or camelina meal would be used directly as a protein source in food or in protein-enriched bioproduct development. The presence of antinutritonal factors and high fibre content pose negative impacts on protein functionality, organoleptic properties and the nutritional value of the end-products. Therefore, isolation of protein (mainly storage proteins) from the meal is necessary to study them in detail. Processes available in the literature for preparation of canola protein concentrates or isolates, such as alkali extraction followed by isoelectric precipitation and the protein micelle mass (PMM) method as reviewed by Wanasundara et al. (2015), could be adapted to isolate storage protein from camelina meal. These methods can isolate most of the meal protein from non-protein contaminants, but cannot separate cruciferin from napin. The protein separation process proposed by Osborne (1924) allows the obtaining of a salt-soluble

globulin faction (cruciferin and some napin) and a water-soluble albumin fraction (napin and some cruciferin) from canola or camelina meal (Tan et al., 2011a; Li et al., 2014). These protein fractions are mixtures of cruciferin and napin which have quite different structural and chemical properties according to information available in the literature. Obtaining napin and cruciferin separately would allow understanding the proteins better and also would support their utilization in applications where their maximum potential could be obtained. The chromatographic separation and purification process described by Bérot et al. (2005) appears to be a suitable method to overcome problems associated with obtaining purified proteins at large scale.

The chromatographic purification began with protein soluble at pH 8.5 in 50 mM Tris–HCl buffer (Section 3.1.4.1), which would include both napin and cruciferin. The meal protein extracted at pH 8.5 was subsequently passed through three different chromatography columns, namely desalting, cation exchange and size exclusion, to separate and purify cruciferin (Section 3.1.4.2). The desalting column removed co-extracted pigments and other low molecular mass compounds, and separation of cruciferin and napin was via the cation exchange column. The size exclusion column chromatography ensured further purification. Protein soluble at pH 3 was used as the starting protein for further purification to obtain napin (Section 3.1.4.3). These processes provided cruciferin and napin (Figures 4.5 and 4.6, respectively) at 100% protein purity (% N× 6.25) for both camelina and canola. These results confirmed that purified proteins from these processes did not contain any non-protein contaminants.

The hexameric assembly for the native cruciferin was first proposed by Plietz, Damaschun, Muller, & Schwenke (1983) and was later confirmed by Tandang, Adachi & Utsumi (2004) using crystal structure data. Recently, Withana-Gamage (2013) modelled the *A. thaliana* homotrimeric and homohexameric assemblies using homology modelling. On the other hand, octomeric assembly also has been proposed for cruciferin based on microscopic data. Badley et al. (1975) reported that two stacked rings of four subunits each form the cruciferin tertiary structure, and the work of Marcone, Beniac, Harauz, & Yada (1994) on *B. juncea* and *S. alba* globulin supported this assembly model. Recently, Nietzel et al. (2013) used proteins recovered from isolated PSVs of canola and showed that cruciferin may exist as an octomer with two rings of four monomers stacked together. However, consideration of napin in the PSVs was not discussed in any of these studies. In the present study, native-PAGE analysis of purified cruciferin from both camelina and canola revealed that the cruciferin purification process caused disintegration of the hexamer to a

certain extent. However, the trimeric assembly was not harmed (Figures 4.7A and B); therefore, the quaternary structure of cruciferin was conserved. The diffused protein band is an indication of isoforms of the protein (Werner, Winkler & Stabenau, n.d; Chen, Liu, Hsu, Le & Chen, 2004) which was confirmed by LC-MS/MS analyses.

Purified cruciferin fractions from both camelina and canola confirmed the presence of cruciferin isoforms. Hence, the presence of cruciferin in camelina similar to that in canola was confirmed. The polypeptide bands of both species which were tentatively identified as cruciferin were indeed cruciferin. Similar to the 2DE followed by LC-MS/MS results for purified cruciferin in the present study (Section 4.4.1), the presence of cruciferin isoforms was confirmed by Nietzel et al. (2013) who worked with proteins obtained from canola PSVs. Analysis of data from the present LC-MS/MS study enabled assignment of proteins present in each of the 2DE spot of canola and camelina (Figure 4.11, Tables A1 and A2 in the Appendix) using the respective genes in the genomic databases. Proteomics data showed that the method employed for separation and purification of cruciferin was successful in providing pure cruciferin in when calculated based on the normalized total spectral values (NTS) obtained for each cruciferin isoform (Table 4.4). NTS is a parameter calculated to quantify the abundance of a protein present in the tested protein sample in Scaffold 4 proteomic software (Anonymous, 2014) used in this study and calculated at the MS data level, *i.e.* the sample run through a mass spectrometer (Anonymous, n.d.-c). NTS is a spectra counting method that depends on the number of spectra unique to a given protein across multiple experiments along with the normalization process, which provides a comparative abundance across each of the MS sample levels (Anonymous, n.d.-c; McIlwain et al., 2012). The NTS also is a parameter similar to the exponentially modified protein abundance index (emPAI), normalized spectral abundance factor (NSAF) and the distributed normalized abundance factor (dNSAF) used in other proteomics software that calculate the relative abundance of a protein in a tested sample (McIlwain et al., 2012). The percentages were calculated by considering the sum of the NTS values for all cruciferin isoforms and the sum of NTS values of all of the proteins in the entire sample as indicated in Tables A1 and A2 in the Appendix.

The camelina cruciferin sample showed vicilin contamination (Table 4.4). The LC-MS/MS analysis confirmed the presence of nine vicilin or vicilin-like isoforms in purified camelina cruciferin (Section 4.4.1). Vicilin is a 7S trimeric globulin especially found in legumes (Shewry et al., 1995). Both cruciferin and vicilin are members of the cupin super-family, which share a

common domain known as a 'jelly-roll' β-barrel structure (Shewry et al., 1995, Adachi, Takenaka, Gidamis, Mikami, & Utsumi, 2001). Unlike cruciferin, the native 4˚ structure assembly of vicilin is a trimer and it does not contain disulfide bonds due to a lack of cysteine residues (Shewry et al., 1995). The molecular mass of mature vicilin is ~ 150-200 kDa (Shewry, 1998) and therefore, ~ 50 kDa subunits are expected. Gatehouse et al. (1984) showed that a pea vicilin subunit has a molecular mass of 47-50 kDa. Since vicilin does not contain disulfide bonds, a polypeptide band (~50 kDa) that is visualized under non-reducing and reducing conditions should appear in 1DE. The meal polypeptide profile showed two non-reducing bands at ~69 kDa and 53 kDa under reducing conditions, presumably vicilin (Figure 4.2). There was no evidence for such vicilin bands in the 1DE carried out for purified cruciferin (Figure 4.5); however, 2DE followed by LC-MS/MS analysis confirmed the presence of vicilin in the cruciferin sample. The reason could be there was not enough vicilin in the purified cruciferin that could bind and stain with Coomassie Blue. The vicilin content of the meal is minute compared to that of cruciferin or napin, and the quantity obtained after the series of chromatographic purifications was even less. Also, it is possible that these vicilins are post-translationally processed (proteolysis and glycosylation) and give rise to small polypeptides similar to that of pea vicilin (Gatehouse et al., 1984; Casey et al., 1986). These small vicilin fragments may co-exist with free α- and β-chains of cruciferin (Figure 4.5). No evidence was found that vicilin or vicilin-like 7S proteins were present in the purified canola cruciferin in the present study. The existence of 7S proteins in *Brassicaceae* plants have not been reported in the literature (Wanasundara, 2011). However, partial complementary DNA (cDNA) sequences or expressed sequence tags (EST) of *Arabidopsis* which share high homology with pea vicilin and related legume 7S cDNA sequences have been identified (Delseny & Raynal, 1999). Although there is no evidence for expressed 7S proteins in crucifer seeds, at least 1-2 genes in *Arabidopsis* that encode 7S proteins have been discovered (Delseny & Raynal, 1999; Shewry and Casey, 1999a, 1999b) and can be found in proteomic databases, *e.g.* AtPAP85 (Q9LUJ7) and AtVCL22 (Q9SK09) available in UniProtKB/Swiss-Prot (http://www.uniprot.org/). To the best of the author's knowledge, this is the first report of expressed vicilin in camelina, as well as in a *Brassicaceae* oilseed. A minute amount of napin was detected in purified cruciferin from camelina and canola. The napin isoform present in the camelina cruciferin isoform is Cs2S-4-G1, whereas it was 2SS4 in canola.

5.2.1.2 Napin

Conditions that favoured high solubility of napin was suitable for the preparation of protein extracts for purification and minimized non-napin protein contamination (Section 3.1.4.3). Native-PAGE confirmed that the process of obtaining napin did not alter its native conformation (Figures 4.7C and D) in either species. Since SDS is not involved in native-PAGE, the protein conserved its native conformation and its mobility under the electric field was regulated by the ratio of the electric charge to hydrodynamic friction (Arakawa, Philo, Ejima, Tsumoto, & Arisaka, 2006).

Proteomics data analysis of napin confirmed the expression of five napin isoforms in camelina and canola (Section 4.4.2). The presence of napin in camelina similar to that in canola was confirmed. Therefore, the predominant molecular species of purified protein was indeed napin. It was interesting to observe that the 2SS4 napin isoform (identified together in purified cruciferin) was not present in purified napin from canola. On the other hand, the napin isoforms that were present in purified cruciferin from camelina also were identified as a possibility in the purified napin. Therefore, it can be suggested that the 2SS4 napin isoform exists in strong association with cruciferin in canola and the conditions provided were not sufficient to separate them. In napin from camelina, a comparatively high level of contamination with late embryogenesis abundant protein (LEA) was evident (Table 4.5). This means that the conditions suitable for obtaining canola napin were not the best for camelina. The LEA proteins have molecular masses between 10 kDa and 30 kDa and accumulate in seed embryo tissues during the late stage of seed development (Ingram & Bartels, 1996; Hong-Bo, Zong-Suo, & Ming-An, 2005). The role of LEA is to provide protection from environmental stress, especially from dehydration during seed maturation, thereby maintaining seed germination capacity (Goldberg, Baker & Perez-Grau, 1989; Skriver & Mundy, 1990; Hand, Menze, Toner, Boswell, & Moore, 2011). This protein is found in cytoplasm localized in the nuclear region. It is mostly a basic protein with a pI > 7 and has a disordered secondary structure (Filiz, Ozyigit, Tombuloglu, & Koc, 2013; Amara et al., 2014). Although LEA protein is evident in the purified napin from canola, it was to a lesser extent compared to that of napin from camelina. Since LEA is a basic protein similar to napin, it is possible that they exhibit similar solubility and hydrophilic characteristics and co-extract and co-purify. For some reason, camelina contains a significantly higher amount of LEA protein than does canola. The camelina plant is known for its drought tolerance features making it suitable for marginal lands in low soil moisture areas. It appears that the contaminating polypeptide band

observed in Figure 4.6D is possibly the LEA protein from camelina. The proteomics results from canola napin showed that it is contaminated with cruciferin and LEA protein. Therefore, the contaminating bands which appeared in the SDS-PAGE profile (Figure 4.6E) of purified canola napin are presumably dissociated α- or β- chains of CRU 3 and CRU 4 or degraded LEA 76 protein.

5.2.2 Oil body proteins

The TEM images of both camelina and canola seed cotyledon cells (Figures 4.8 and 4.9) clearly showed that OBs are nicely packed without coalescing due to the stabilizing protein-rich membrane. These proteins may have high surface active properties and functionalities that could be exploited to develop oil-water based systems. As discussed in the literature review, oleosin is the most abundant protein and accounts for 75-80% of the oil body proteins in canola and *A. thaliana* (Huang, 1996; Jolivet et al., 2004; Jolivet et al., 2009). The rest mainly consists of caleosin, followed by steroleosin. A number of studies have been carried out to isolate, characterize and study the properties of canola OBPs (Murphy et al., 1989; Katavic et al., 2006; Jolivet et al., 2009; Jolivet et al, 2011). Isolation of OBPs was difficult and required several steps involving floating OBs on density gradients as explained by Tzen, Peng, Cheng, Chen, & Chiu (1997), with modifications unique to each study. The high molarity buffer media contained EDTA, salt, sucrose and urea, and a detergent such as Tween-20 were generally used for dispersing OBs. Afterward, OBPs can be separated in acetone, allowing lipids to be soluble and proteins to precipitate (Katavic et al., 2006, Jolivet et al., 2009). Acetone precipitated the protein from the OB surface in most of these studies, which is a commonly practised method to precipitate or concentrate proteins (Simpson & Beynon, 2010). Acetone and other organic solvents decrease the dielectric constant of the medium, reducing solubility and consequently precipitating the protein (Young, 1994). Organic solvents exhibit an affinity for the hydrophobic surfaces of protein. As a result, organic solvents interrupt the internal hydrogen bonds, causing destabilization of tertiary structure or denaturation of protein along with precipitation (Young, 1994).

One of the objectives of this study was to develop an easy method to isolate OBs from the seed and separate OBPs as intact as possible. Due to the drawbacks associated with OBs isolation using the Tzen et al. (1997) method, an alternative method explained by Maurer et al. (2013) for soybean was adapted. This method involves fewer steps compared to other methods and utilizes water and sucrose with some pH adjustments. The pH is adjusted to 11 to solubilize storage

proteins and sugar is used to change the density of the medium to facilitate separation of OBs into a cream layer. It is also important to obtain intact protein if structure-function studies are to be carried out. Since using acetone to precipitate protein poses a risk of destabilizing the native structure of OBPs, SDS was used instead of acetone. In an emulsion, the oil-water interface always prefers the emulsifier that lowers the interfacial tension the most (McClement, 2005). This phenomenon is called preferential adsorption. The OBPs act as the emulsifier that stabilizes oil droplets and prevents them from coalescing. SDS is known as a low molecular weight emulsifier with good emulsifying capacity (McClement, 2005), better than that of protein. In a situation where both SDS and protein are present, SDS should displace the protein from the OB surface into the medium. It can be assumed that the OBPs recovered from the medium have retained their native properties compared to OBs obtained from acetone precipitation.

The protein content of OBPs was below 30% (%N × 6.25) for both camelina and canola (data not shown). Presumably, SDS is precipitated along with the proteins. Although SDS is more surface active than protein, it did not seem capable of replacing all of the protein on the oil body surface. Therefore, excess SDS that is free in solution may precipitate along with the displaced protein. The isolated proteins were then desalted using a Sephadex G-25 Hiprep^{TM} 26/10 desalting column against Milli Q water in the AKTA explorer system to remove SDS contamination, but this was not successful. The presence of polypeptide bands typical of oleosin was evident in both camelina and canola OBP isolates (Figure 4.10) and was further confirmed by LC-MS/MS analysis (Tables A5-A8 in the Appendix). The presence of some other contaminating proteins also was evident in the polypeptide profiles. A similar polypeptide profile for canola oil body proteins was reported by Katavic et al. (2006) and Jolivet et al. (2006). The proteomics data confirmed that both OBP isolates were contaminated with cruciferin and napin (Tables A5-A8 in the Appendix). In the method used in the present study, storage protein contamination was addressed by the two-step washing of OB layers at pH 11. However, it appears that these washing steps were not adequate to eliminate this contamination. Washing the oil body-containing cream layer one more time at pH 11 might have reduced cruciferin contamination. An additional washing step of the protein isolates at pH 3 may have eliminated contaminating napin. The LC-MS/MS results showed that canola contained more oleosin than did camelina (Section 4.4.3). Therefore, the method should be optimized further to obtain intact OBPs with higher purity, especially from camelina.

Proteomics data also revealed seven different camelina oleosins that are encoded by twelve genes (Section 4.4.3; Tables A5 and A6 in the Appendix). Four different oleosin isoforms, *i.e.* OLES2, OLEO5, OLES1, and OLEO3, also were evident in isolated OBP from canola (Section 4.4.3; Tables A7 and A8 in the Appendix). Katavic et al. (2006) were able to identify three oleosin isoforms, *i.e*. oleosin type 4, 1803528A and oleosin BN-V, from isolated OBPs from canola. The presence of caleosin was evident only in camelina; presumably, the content of caleosin is less in canola compared to camelina. A specific protein type, hydroxysteroid dehydrogenase, was detected in the camelina OBPs. According to Katavic et al. (2006), the hydroxysteroid dehydrogenase is presumably steroleosin, similar to sesame seed steroleosin. The putative steroleosin in camelina has a molecular mass between ~39 kDa and 42 kDa; this value matches molecular mass data reported in the literature (Tzen, 2012). The present study did not provide evidence for steroleosin in canola.

Storage protein contamination is the major issue with respect to isolating OBPs. It is a common problem in this regard and hard to eliminate as the hydrophobic interactions facilitate OBPs and storage protein association (Katavic et al., 2006). Interestingly, isoforms of cruciferin and napin were evident from pH 9-12 (Tables A6 and A8 in the Appendix) and pH 3-12 (Tables A5 and A7 in the Appendix), respectively, in both camelina and canola OBP separations, which was not evident when purified cruciferin and napin were separated using 2DE. The results suggest that cruciferin and napin have close associations with OBPs and move along with them during the separation process. As speculated, addition of SDS was successful in replacing and isolating OBPs from the OB surface. However, more work is needed on a method to minimize contamination to improve the purity of the OBP isolates.

5.3 Structural details of cruciferin and napin and their changes with pH and temperature

Understanding protein structure in relation to its function(s) is a requirement in developing applications for them. Proteins are considered to be in their native folded state (N) based on the conformation adopted under the conditions experienced in their natural environment. The conformation that the protein molecule adopts when it is completely unfolded, which is a highly flexible random coil, is the denatured or unfolded state (U) (Morra, 2006). For a simple globular protein, a two-state-monomeric model has been proposed to elaborate the unfolding mechanism (Equation 5.1). Similarly, the two-state model for a dimeric globular protein can result in two unfolded monomers as shown in Equation 5.2 (Walters, Milam & Clark, 2009). An intermediate

(I) state is possible in the protein unfolding process, and three-state-monomeric and three-state-dimeric models for protein unfolding have been proposed (Equations 5.3-5.5) to explain the unfolding mechanism. In the monomeric model, one intermediate (I) can be found (Equation 5.3), whereas the intermediate can be either a dimer (I_2) or two monomers (2I) in the dimeric model for protein folding, as shown in Equations 5.4 and 5.5, respectively (Walters et al., 2009). These intermediates of the globular proteins retain an appreciable amount of secondary and tertiary structure, and are considered to be in the molten globule state (Holt, 2000, Fink, 2001).

Two-state models for globular protein,

Monomeric: $N \leftrightarrow U$ Eq. 5.1

Dimeric: $N_2 \leftrightarrow 2U$ Eq. 5.2

Three-state models for globular protein,

Monomeric: $N \leftrightarrow I \leftrightarrow U$ Eq. 5.3

Dimeric: $N_2 \leftrightarrow I_2 \leftrightarrow 2U$ Eq. 5.4

Dimeric: $N_2 \leftrightarrow 2I \leftrightarrow 2U$ Eq. 5.5

As far as protein structure is concerned, the three-dimensional arrangement (tertiary structure) of a protein is important as it determines surface properties, such as hydrophobicity, that affect solubility and solubility-associated functional properties, such as emulsification, foaming and gelation (Withana-Gamage, 2013). The secondary structure of a protein greatly influences protein folding (Myers & Oas, 2001; Kwok, Mant, & Hodges, 2002); therefore, is important in determining the final, three-dimensional configuration that produces a specific functionality. Research has shown that secondary structural features can be related to nutritional aspects, such as protein quality, availability, nutrient utilization and digestive behaviour (Yu et al., 2004; Yu, McKinnon, Christensen, & Christensen, 2004). Structural properties are influenced by external factors, such as pH, temperature and pressure; therefore, the functionality of a protein may be affected by processing conditions. In the canola/rapeseed oil extraction process, proteins in the seed denature and are subjected to non-reversible interactions with other constituents, compromising the solubility of the resulting meal protein (Wanasundara, et al., 2015). Therefore,

it is important to understand protein structural features and their alterations in response to changes in processing conditions.

Assuming that the purified cruciferin and napin are in the native state, their tertiary and secondary structure features can be probed. Therefore, the response of these protein molecules, such as changes in molecular conformation to environmental changes, can be understood using the appropriate parameters that describe its structure.

5.3.1 Cruciferin structural features

5.3.1.1 Effect of pH and temperature on cruciferin tertiary structure

As discussed earlier in the literature review (Section 2.6.1), the tertiary structure of cruciferin is presumed to have the β-chains of the polypeptide buried within the molecule, whereas the α-chains are exposed more to the solvent environment. The quaternary structure is a hexamer made up of two trimers. Each trimer contains IE (interchain S-S bond containing) and IA (intrachain S-S bond containing) faces, where two trimers are piled up together via IE face-to-face interaction to form the hexamer (Figure 4B).

The medium pH greatly influenced cruciferin tertiary structure as indicated by the changes in S_0. Cruciferin of both camelina and canola showed maximum S_0 at pH 3; S_0 markedly decreased as the pH moved to neutral and alkaline (Table 4.8). It appears that at a lower pH, more 8-anilinonaphthalene-1-sulfonic acid (ANS) was bound to cruciferin, thereby providing high fluorescence, and *vice versa*. Apenten & Folawiyo (1995) also observed an increasing trend of ANS-binding fluorescence of canola globulin with decreasing pH. The reason for this could be the increasing number of binding sites, or increasing protein-ANS binding affinity due to changes in the surroundings of the binding site; protein unfolding favours ANS binding (Stryer, 1968; Slavik, 1982; Arakawa, Kita & Narhi, 1991). A change in protein S_0 is a positive indication of pH-induced structural change (Korte & Herrman, 1994). The intrinsic fluorescence and near UV-CD data (Figures 4.23 and 4.24) confirmed acid-induced unfolding of cruciferin. Presumably, the hydrophobic residues buried in the core of cruciferin are exposed as it unfolds at pH 3; therefore, the increased affinity to protein-ANS binding provide an increase in the fluorescence intensity. It can be assumed that the environment of protein binding sites is less hydrophobic at pH 10 (Apenten & Folawiyo, 1995) and therefore the lowest S_0 was observed at pH 10.

The cruciferin trimers are stacked together via IE face-to-face interaction to form the hexamer. *In silico* homology modelling of *A. thaliana* cruciferin structure showed that the IE face of the CRUA and CRUB subunits contain more hydrophobic residues than the IA face (Withana-Gamage, 2013). The CRUA and CRUB subunits are 85.3% and 75.8% homologous, respectively, to canola procruciferin (Withana-Gamage, 2013); therefore, similar hydrophobicity in the IE face can be expected. The IE face of the two cruciferin trimers that is occluded in the hexameric assembly might have been exposed due to dissociation, causing the increase in hydrophobicity and ANS binding (Table 4.8). This is why the surface hydrophobicity of canola procruciferin (trimeric) is found to be higher than that of the closely packed mature 11S globulin (hexameric) of *Glycine max* (Tandang-Silvas et al., 2010). It was also shown that soybean glycinin (11S globulin) was mainly present in trimeric complexes (7S) at pH 3.8 (Lakemond, de Jongh, Hessing, Gruppen, & Voragen, 2000). Gueguen, Chevalier, And, & Schaeffer, (1988) showed dissociation of pea legumin into 7S and 3S subunits under acidic conditions. A similar phenomenon also was observed by Jarpa-Parra et al. (2015) with respect to lentil legumin. Legumin exists in its native hexameric conformation at neutral pH with a hydrodynamic radius of 12 nm, which was reduced to 7 nm upon changing the medium to pH 3. Therefore, it seems that the cruciferin hexamer dissociates at pH 3 into trimers. Consequently, protein becomes more hydrophobic and lower in ionized residues, causing aggregation. According to Bhatty, McKenzie & Finlayson (1968), rapeseed globulin dissociates into 2-3S components after dialyzing in 6 M urea, especially in acidic buffers below pH 3.6. Schwenke and Linow (1982) have demonstrated that the cruciferin complex exists as 12S at high ionic strength (≥ 0.5) and dissociates into 7S components when dialysed against water, freeze dried, and reconstituted in weakly alkaline water (pH 8.0). It is assumed that the 7S complex is the trimeric half of the hexamer.

The acid-induced, structural destabilization/unfolding of cruciferin can be explained by the three-state model similar to Equations 5.4 and 5.5. Most likely, the native-hexameric cruciferin (N) is dissociated into two trimers, which is the intermediate (I) of the unfolding process (Equation 5.6, three-state-hexameric model). It is also plausible that at pH 3, the trimer is further disintegrated into corresponding subunits ($F_{350}/F_{330} > 1$), where the trimeric structure of the protein is conserved at pH 7 and pH 10 ($0.81\ F_{350/}\ F_{330}$). In this case, at pH 3, the intermediate of the three-state-hexameric model is the cruciferin monomer (Equation 5.7). The marked increment in S_0 at pH 3 also suggests that the buried β-chain of cruciferin is revealed and the hydrophobic residues (Phe

and Trp) may be exposed to the solvent environment disturbing its tertiary structure (Apenten and Folawiyo, 1995), which was observed in near UV-CD spectra (Figure, 4.23A). Since the secondary structural features were conserved at pH 3 (Table 4.7), it can be assumed that the cruciferin intermediate exists in a molten globule state. With respect to acid-induced denaturation of soybean glycinin (12S), DSC studies by Kim, Kim, Yang, & Kwon, (2004) showed glycinin denaturation occurs at pH 3. The CD spectrum at the pH at which denaturation was observed exhibited conserved secondary structural features with increased α-helix content similar to cruciferin in this study.

Three state models,

Hexameric: N6 (Hexamer) ↔ 2I3 (trimer) ↔ 6U (monomer) Eq. 5.6

Hexameric: N6 (Hexamer) ↔ 6I (monomer) ↔ 6U (monomer) Eq. 5.7

The native state of cruciferin possesses all four levels of structural organization (quaternary, tertiary, secondary and primary). If the intermediate is the trimer assembly (of quaternary structure level) (Equation 5.6), cruciferin still shows all four structural levels. On the other hand, if the intermediate is the monomer of cruciferin (Equation 5.7), it may conserve some tertiary structure, a substantial amount of secondary structure and the primary structure. Cruciferin may only demonstrate the primary structural organization if it is completely unfolded or denatured (U) as illustrated by Equation 5.6 or 5.7, which was not evident in this study.

These structural changes in cruciferin coincide with DSC results. Cruciferin did not show any denaturation peak at pH 3, even at ambient temperature, whereas denaturation at pH 7 and pH 10 was distinct (Table 4.10). The peak denaturation temperatures and enthalpy changes at pH 7 and pH 10 were similar. The peak denaturation temperature of cruciferin from three different canola varieties ranged from 84.6˚C to 86.6˚C at neutral pH, similar to denaturation observed for cruciferin in this study (Table 4.10; Salleh et al., 2002). The results of this study showed that cruciferin has high thermal stability at neutral and alkaline pH, whereas acidic conditions caused loss of structural stability. The binding of ANS (Table 4.8) together with the intrinsic fluorescence of tryptophan residues (Table 4.9 and Figure 4.22) and DSC (Table 4.10) suggest that cruciferin is subject to acid-induced conformational changes, leading to loss of its quaternary and tertiary structure only (Korte & Herrmann, 1994).

5.3.1.2 Cruciferin secondary structure and the effect of pH and temperature

The secondary structure of 11-12S globulins is known to be an α+β-type structure, where the β-type structure is predominant (Withana-Gamage, 2013). The FT-IR spectra of camelina and canola cruciferin did not indicate distinguishable differences (Figure 4.18). The deconvoluted amide I band also confirmed that, except for a subtle difference in β-sheet content (2.6%), other components of cruciferin secondary structure were not different between these two oilseeds (Table 4.6). As proteins with dominant β structure, the β-sheet contents of camelina (43.0%) and canola (45.6%) were similar to the value reported for *A. thaliana* (wild type) cruciferin (44.1%) (Withana-Gamage, 2013). The secondary structural features resolved from the crystal structure of rapeseed procruciferin (25 to 27 β-sheet) (Tandang-Silvas, 2010). A greater β-sheet content could be resulted due to pressure changes during the protein purification and the aggregation of protein during freeze drying. Several studies have shown an effect of pressure on protein secondary structure, especially the content of β-sheet (Mozhaev, Heremans, Frank, Masson, & Balny, 1996; Gao et al., 2005). The content of β-sheet is an indication of protein aggregation (Fink, Seshadri, Khurana, & Oberg, 1999; Shivu et al., 2013). Shivu et al. (2013) showed that protein aggregates exhibited characteristic new β-sheets at lower frequencies in the amide I region, which were not present in the native protein. Therefore, an increase in β-sheets is usually observed compared to native protein. The new β-sheets could be from the strong hydrogen bonds present in intermolecular β-sheets in the protein aggregate (Shivu et al., 2013). The α-helix contents of purified camelina cruciferin (10.1%) and canola (9.4%) were comparable to that of native procruciferin (Tandang-Silvas, 2010) or cruciferin of *A. thaliana* wild type (9.2%; Withana-Gamage, 2013). Therefore, it can be assumed that the secondary structure of camelina cruciferin is similar to that of canola cruciferin.

To the best of the author's knowledge, there is no information available elsewhere on the secondary structure of camelina cruciferin obtained using either FT-IR or CD. However, Li et al. (2014) reported that the deconvoluted amide I region of globulin fractions obtained from camelina meal by the Osborne method showed average peak areas of 1.54 for α-helices and 1.67 for β-sheets. A percentage of these two components with respect to the total area of the amide I region was not reported. The protein types present in the isolated globulin fractions were not identified, hence the composition is not known. It can be assumed that these globulin fractions contained

mostly cruciferin, since cruciferin was reported to be the predominant protein present in the globulin fraction obtained from the Osborne classification (Tan et al., 2011a). The average α-helix to β-sheet ratio of the globulin fractions was 0.94. This revealed that the globulin fraction contained more β-sheet than α-helix. Another study showed that the α-helix to β-sheet ratio of raw camelina seed protein was 1.093 (Peng, Khan, Wang, & Yu, 2014). This study represented all types of proteins inside the seed, and did not focus on the storage proteins.

The secondary structure was sensitive to the pH of the medium (Figure 4.20A and Table 4.7). It is important to note that at pH 3 and pH 10, where cruciferin is moving away from the approximate native pH (pH 7), the secondary structural features were conserved in both species. As discussed earlier, this suggests that cruciferin is not completely unfolded at either pH 3 or at pH 10. The β-sheet content of camelina cruciferin was not significantly affected by the medium pH, whereas it was significantly reduced in canola as the pH moved away from neutral. The decreased or unchanged β-sheet content as the pH moved away from neutral indicates less possibility of aggregation, because it was found that an increase in the content of β-sheets is an indication of protein aggregation (Fink et al., 1999; Shivu et al., 2013). The observed loss of cruciferin solubility at pH 3 (Figure 4.3), therefore, was not related to cruciferin aggregation and cannot be explained by secondary structural changes.

5.3.1.3 Effect of pH and temperature on the solubility of cruciferin

Solubility is an important functional property of a protein governed by its physicochemical properties (net charge) and structural properties (hydrophobicity and conformation), which can be modified by external factors such as pH, ionic strength and temperature (Salleh et al., 2002; Damodaran, 2008; Withana-Gamage, 2013). A uniform distribution of both positive and negative charges promotes formation of aggregates and consequent precipitation, whereas net negative or positive charge increases solubility (Fukuda, Maruyama, Salleh, Mikami, & Utsumi, 2008; Kramer, Shende, Motl, Pace, & Scholtz, 2012). Solubility exhibits a negative correlation with protein surface hydrophobicity (Nakai, 1983). Conformational changes, such as denaturation due to extrinsic factors such as heat or pH, also pose an adverse effect on solubility (Withana-Gamage, 2013). The surface charge of the protein can be manipulated using pH and ionic strength, hence the solubility can be changed by changing these external factors. Due to the structural changes in cruciferin induced by pH, the lowest solubility would be expected at pH 3 (increased hydrophobicity), whereas increased solubility would expected at pH 7 and 10. As expected, the

lowest solubility was observed at pH 3, whereas the highest was observed at pH 10 (Figure 4.25A). However, the solubility of camelina cruciferin appeared to be high even at pH 3 (>90%) compared to that of canola (79%, Figure 4.25A). It appears that acidic-pH-induced denaturation and increased hydrophobicity had less effect on camelina cruciferin. Presumably, surface charge, which is an intrinsic factor, playing a dominant role in determining solubility. Withana-Gamage (2013) showed that at pH 3, *A. thaliana* cruciferin had a positive zeta potential, and the CRUC monomer had a high positive potential of ~30 mV compared to the other cruciferin isoforms. Solubility at pH 3 can be expected due to the positive zeta potential of cruciferin molecules; however, the specific reason for the extremely high solubility is difficult to explain. The zeta potential results suggested that some cruciferin isoforms can demonstrate high positive potential at pH 3. Therefore, it is possible that several isoforms of the twelve cruciferin isoforms identified in camelina (Table 4.3) may have high positive potential at pH 3, which might have led to improved solubility when isolated. Although cruciferin appeared to be soluble at pH 3 (Figure 4.25), cruciferin was not observed when the meal protein was extracted at pH 3 (Figure 4.3). Presumably, the seed coat materials restricted the movement of cruciferin or the proposed cruciferin-napin complex does not exist in purified cruciferin.

5.3.2 Napin structural features

5.3.2.1. Changes in napin structure due to changes in pH and temperature

Napin belongs to the prolamin super family, and has a different and less complex structure than cruciferin (Section 2.6.2, Figure 2.6). Extrinsic factors, such as pH, temperature and pressure, may affect the structure of napin and cause denaturation/unfolding and alter its secondary and tertiary structural conformations. As expected, napin showed increased fluorescence emission (surface hydrophobicity) at pH 3 (Table 4.8). The near UV CD spectra (Figure 4.21B) confirmed acid-induced structural changes in napin, even though it was not as noticeable as in cruciferin. Napin has a monomeric structure and did not demonstrate higher order tertiary or quaternary structure, in contrast to cruciferin. This explains the large difference in surface hydrophobicity values of napin from both camelina and canola, compared to cruciferin. However, camelina napin showed lower S_0 values than did canola, indicating that the camelina napin structure may have different features and, therefore conformational changes compared to that from canola. Moreover, differences in the S_0 values of the same magnitude as at pH 3 were not evident at the other pHs

tested. There is not enough evidence to reach conclusions on napin unfolding or denaturation and subsequent improvement of hydrophobicity as was seen for cruciferin. Presumably, changes in the environment of the hydrophobic binding sites due to changes in pH were the main reason for improved ANS binding at pH 3 and pH 10. The two-state or three-state protein folding/unfolding models (Equations 5.1-5.5) are not adequate to explain pH-induced structural changes in napin, although they are useful for cruciferin. Hence, further investigation is required to understand the napin conformational changes influenced by pH. Napin also did not provide any evidence of a thermal denaturation peak below 100°C at the any of the pHs tested. It appears that napin is highly stable at all of these pHs. In other studies denaturation of napin was observed at 100.3°C and 80°C at pH 6 and 3, respectively (Krzyzaniak, Burova, Haertlé, & Barciszewski, 1998). It also is possible that low pH-induced structural changes caused the loss of thermal stability of napin. The two state model (Equation 5.1) can explain the protein denaturation in this case; however, there was no evidence to confirm the presence of an intermediate during the transition from N to U. The present work confirms that napin structure is highly stable, but medium pH has an affect which cannot be described from the data of this study.

The thermal stability of *Brassica juncea* napin was studied by Jyothi, Sinha, Singh, Surolia, & Appu Rao (2007). The reversible thermal unfolding of napin and consequent aggregation was evident between 26˚C and 80˚C. Two distinct thermal transition peaks at 50.3˚C and 62.7˚C were identified. The napin structure was found to be stable until 74.9˚C, but it started to unfold thereafter resulting in aggregates due to hydrophobic interactions. However, the unfolding was reversible. The study of Jyothi et al. (2007) showed that napin was not thermally denatured up to 80˚C, although structural unfolding was evident. It appears that *B. juncea* napin has high thermal stability, which was also evident with respect to canola napin (Krzyzaniak et al., 1998). The effect of pH on the thermal stability of napin showed that the two transition peaks decreased as pH increased. No denaturation peak was observed with the temperature ramp employed (Jyothi et al., 2007). On the other hand, an irreversible thermal unfolding of canola napin at pH 7 was reported at ~62-63°C (Folawiyo & Apenten, 1997). These studies suggest that napin may exist as one or more intermediates before it unfolds completely and loses its secondary and tertiary structural organization. Neither denaturation peak nor thermal transition peaks similar to *B. juncea* were observed for napin from camelina and canola in the current study. The hermetically sealed pans burst above 100˚C, hence the presence of a peak after 100˚C was not conclusive. Development of

internal pressure could have been the reason for the bursting of the pans above 100˚C; however, this phenomenon was not observed in cruciferin from camelina or canola. When cruciferin was mixed with the buffers, the mixture was more of a slurry, whereas napin was completely soluble in the buffers. This could be the reason that caused pressure build-up in the napin-containing aluminum DSC pans. Many DSC pans cannot stand high internal pressure, resulting in sample leakage and bursting. To overcome the pressure build-up, lids containing holes or crimped DSC pans that do not seal can be used (Gabbott, 2008). However, hermetic sealing is important for water-containing samples; therefore, either of above mentioned solutions would not solve the problem. The best alternative would be using DSC pans that can tolerate high pressure and temperature, such as 'O' ring sealed stainless steel pans or high-pressure capsules (Gabbott, 2008). It is difficult to provide an obvious reason for not observing any thermal transition peak below 100˚C at any of the pHs tested. Presumably, napin has high thermal stability and is not denatured below 100˚C. Therefore, further investigation of the thermal stability of napin is needed with different types of DSC pans and experimental conditions, such as modulated DSC.

5.3.2.2 Secondary structural features of napin and the effect of pH

Napin is known to have a highly helical secondary structure (Figure 4.6; Tan et al., 2011a) similar to that of cytochrome c or myoglobin (Byler & Susi, 1986); therefore, only the α-helical content of napin was calculated by deconvoluting the amide I band of the FT-IR spectrum (Table 4.6). The deconvolution process and algorithm used in the FT-IR data analysis software is a default function that cannot be manipulated by the user. The software allows the user to define parameters, such as the resolution enhancement factor (K), full bandwidth at half height (FWHH) and apodization filter. The user can change these parameters accordingly to obtain the most reliable and comprehensive results for each type of protein of interest. These parameters should be defined with great care to avoid any misinterpretation. Since napin has a highly helical secondary structure, it should be treated differently than cruciferin. Therefore, a different K factor and FWHH were used (Figures 4.20 and 4.21) to deconvolute the napin amide I band as described the Byler & Susi (1986). The deconvoluted amide I band for napin contained peaks that resembled β-sheets (1627-1638 cm^{-1}) and β-turns (1674-1684 cm^{-1}), similar to cruciferin (Figure 4.19), even after adjusting the K factor and FWHH. A similar phenomenon also was evident for hemoglobin, myoglobin and cytochrome c at 1627-1638 cm^{-1} and 1671-1675 cm^{-1} of the amide I region (Byler & Susi, 1986). It is highly unlikely that these proteins contain β-structures; therefore, it is possible that these bands

are related to some segments associated with the short, extended chains attached to helical cylinders (*e.g.* residues 79-84, 98-99 and 150-153 in myoglobin) which were neither β-sheets nor β-turns (Byler & Susi, 1986).

Similar to FT-IR, the deconvolution algorithm used in the CD data analysis software also is beyond the control of the user. Hence, the ability to evaluate an individual protein based on its specific structural features is limited. Previous studies have reported 40-45% helix and 16-20% β-sheet (Schwenke, 1990) and, 25% α-helix and 38% β-sheet (Krzyzaniak et al., 1998) for canola napin using CD analysis. The secondary structure modelled using the primary amino acid sequence (Figure 2.6; Barciszewski, Szymanski, & Haertle, 2000) and the solution structure of 2S albumin (RicC3) from *Ricinus communis* resolved using NMR (Pantoja-Uceda, Bruix, Gimenez-Gallego, Rico, & Santoro, 2003) confirmed the helical napin structure, but not the β-sheet. Therefore, neither β-sheet nor β-turns of napin were taken into account in far-UV CD spectral deconvolution in this study. The α-helix content obtained from FT-IR and CD at pH 7 in this study was similar for camelina and canola. Deconvoluted CD spectra also showed similar α-helix and random structure contents (Table 4.7) for camelina and canola. The results suggest that the napins of both of these species share similar secondary structural features (Tables 4.6 and 4.7). Subtle changes in secondary structure components with changes in medium pH also were evident (Figure 4.20B and Table 4.7). Overall, the napin secondary structure was not greatly influenced by the medium pH.

The secondary structure of a protein can be related to nutritional aspects, such as protein quality, availability, nutrient utilization and digestive behavior (Yu et al., 2004; Yu, McKinnon, Christensen, & Christensen, 2004). It was shown that high β-sheet content might compromise access to gastro-intestinal digestive enzymes, causing low protein value and availability, and further information can be obtained using the β-sheet to α-helix ratio (Yu, 2005). Digestibility and the β-sheet to α-helix ratio demonstrate an inverse relationship. According to this theory, napin (2S albumin) has a greater potential to be highly digestible compared to cruciferin, as it does not contains any β-sheet. However, napins in *Brassicaceae* oilseeds were reported to be resistant to proteolytic digestion by gastric enzymes as the disulfide bonds provide high stability (Abeysekara, 2012). There is not enough evidence to apply this theory to cruciferin digestibility. Most of the studies have been carried out with respect to secondary structural modelling analysis of oilseed meal samples, where a number of different proteins contribute to the individual secondary structural components. It can be assumed that in meal or protein isolates where both cruciferin and

napin are present, napin may improve digestion by reducing the β-sheet to α-helix ratio as it contributes to increased α-helix content in general. However, the trade-off between increased α-helix content from napin vs. resistance to digestion caused by disulfide bonds, as indicated by studies related to the allergenicity of 2S protein and napin, should be considered.

5.3.2.2 Napin solubility

As discussed above, the parameters of structural feature assessment showed that without prominent structural changes with changes in pH, napin structure does not behave in a similar manner as cruciferin. A trend to increasing solubility of camelina napin with pH and an opposite trend in canola were observed (Figure 4.25B). Both species exhibited similar solubility values at pH 10. Camelina napin showed the lowest hydrophobicity at pH 10 and canola napin at pH 7. The increase in solubility of camelina napin at pH 7 and pH 10 can be explained by its reduced hydrophobicity. In contrast, the solubility of canola exhibited an opposite trend. Since the pI of napin is ~11, high solubility away from the pI can be expected; therefore, improved solubility at pH 3 or pH 7 compared to pH 10 can be explained. The highest hydrophobicity values for napin were observed at pH 3. These contrasting observations suggest that other intrinsic factors such as associated non-protein molecules may affect the solubility of napin. Further investigation to understand the solubility of napin is needed.

6. SUMMARY AND CONCLUSIONS

6.1 Summary

Camelina seed coat contains mucilage and it becomes a component of the de-oiled meal. Therefore, removal of mucilage is essential for the recovery of protein with high purity. Treating whole seeds with polysaccharides degrading enzyme, which is Viscozyme®, removes seed coat mucilage and consequently improves protein extraction and recovery. The de-mucilaged camelina meal is a protein-rich plant product similar to canola meal. Camelina contains seed storage proteins cruciferin (11S), napin (2S) and their isoforms, as do canola and many other *Brassicaceae* oilseed crops. The presence of vicilin (7S) protein in camelina meal was confirmed, although vicilin is not commonly found in canola or the *Brassicaceae* model plant, *Arabidopsis thaliana*. Cruciferin expressed from eleven genes, napin expressed from four genes, and vicilin expressed from six genes were identified from a total of twelve, eight and nine genes, respectively. The solubility of camelina and canola meal protein as a function of pH followed similar trends, whereas the lowest solubility (apparent pI) of the cruciferin-napin complex was observed at pH 4.5. Only napin was soluble at acidic pHs (<pH 6.5 and <pH 4.5 in camelina and canola, respectively), whereas cruciferin was more soluble at alkaline pHs. The maximum solubility of meal protein was observed at pH 12 for both oilseeds. The meal protein extracted at pH 8.5 followed by a three-step chromatographic purification process provided cruciferin from both camelina and canola, with minor contamination of non-targeted meal protein. The meal protein extracted at pH 3 followed by diafiltration and chromatographic separation resulted in predominantly napin and noticeable late embryogenesis abundance (LEA) protein from camelina, whereas non-napin proteins were minor in the purified napin form canola.

The secondary and tertiary structural features of cruciferin and napin proteins from camelina and canola were similar. Cruciferin tertiary structure was influenced by the medium pH and temperature, which, consequently, affected its physicochemical properties, such as solubility, thermal and surface properties. Although the tertiary structure of cruciferin unfolded at acidic pH (pH 3), complete denaturation was not evidenced in either camelina or canola. It was confirmed that at pH 3, cruciferin assumes an intermediate state, which is plausibly a molten globule state. This cruciferin structure intermediate could be either a 7S trimer or 2S monomer, but this was not distinguishable from the information gathered from structural analyses. Both camelina and canola

cruciferins possessed high thermal stability (>80˚C) at neutral and alkaline pH, whereas structural stability was lost at acidic pH. A hexameric, three-state unfolding model was suitable to explain the folding/unfolding behavior of cruciferin protein in both camelina and canola. On the other hand, the structures of camelina and canola napin were not affected by the medium pH. The pH-induced structural changes confirmed that napin was not completely denatured at pH 4, 7 or 10. No evidence could be found for an intermediate state of napin, therefore the three-state unfolding model was not appropriate in explaining napin structural changes as it was for cruciferin. Napin did not respond to differential scanning calorimetry (DSC) and intrinsic fluorescence analyses similar to cruciferin; therefore, conclusive information about thermal stability of napin structure could not be obtained. Cruciferin and napin are proteins with distinct structural characteristics, although they co-exist in the protein storage vacuoles (PSVs) of camelina and canola. Therefore, the methods utilized for cruciferin structure probing may not be appropriate for napin structural probing.

The method employed to isolate oil body proteins (OBPs) in this study was successful. The presence of oleosins and their isoforms in the isolated oil body protein fraction was confirmed for both camelina and canola. The presence of putative caleosin and steroleosin isoforms also was confirmed in the camelina oil body protein fraction, whereas only oleosin was evident in canola. Although the method employed for isolating OBPs from camelina and canola seed was successful, it was not capable of removing some non-targeted protein contaminants; therefore, optimization of this method to improve the purity of isolated OBPs is necessary. Additional washing steps applied to the separated cream layer at pH 3 and pH 11 might reduce contamination, especially from storage protein.

6.2 Conclusions

Camelina and canola contain cruciferin and napin, which have more or less similar structural characteristics and stabilities. Therefore, the existing conditions of commercial oil processing may have similar impacts on the quality of the protein in de-oiled camelina and canola meal. It can be expected that any temperature above 85˚C may cause thermal denaturation of cruciferin, whereas napin would experience less structural alterations. Vicilin would not pose a great impact on the composition and properties of camelina protein fraction as it is available in minute quantities. However, vicilin could be a potential candidate for improving the lysine content

of camelina meal through genetic improvement. Improved lysine content in meal would bring a competitive advantage to camelina as an animal feed compared to canola and other *Brassicaceae* oilseed meals. Simultaneously, reducing other antinutrients, such as glucosinolate and phytates, also are essential considerations. In order to obtain protein in concentrated or isolated form, camelina will have to go through the technical hurdles of mucilage reduction or removal if aqueous extraction is involved. This will be an additional constrain that is not found with canola. Since the behaviours of cruciferin and napin in aqueous solutions are pH dependent, careful selection of the pH of extraction would enable higher separation of napin from the seed meal. Proteins from the oil bodies of camelina are another source of protein that should be explored further because of their hydrophobic characteristics that are distinct from those of the storage proteins. Further investigations are needed to understand the technological value of camelina cruciferin, napin and OBPs beyond their nutritional value, particularly for use in other applications. Therefore, properties and behaviours such as surface activity, rheology, interactions with other polymers and the ability to form cross-links need to be studied.

7. REFERENCES CITED

Abeysekara, S. (2012). Solubility, *in vitro* digestibility and allergenicity of *Brassica juncea*, *Brassica napus* and *Sinapis alba* proteins (Master of Science Thesis). University of Saskatchewan, Saskatoon, Canada. Retrieved 2016, from: https://ecommons.usask.ca handle/10388/ETD-2011-11-292

Abramovič, H., Butinar, B., & Nikolič, V. (2007). Changes occurring in phenolic content, tocopherol composition and oxidative stability of *Camelina sativa* oil during storage. *Food Chemistry, 104*, 903-909

Adachi, M., Kanamori, J., Masuda, T., Yagasaki, K., Kitamura, K., Mikami, B., & Utsumi, S. (2003). Crystal structure of soybean 11S globulin: Glycinin A3B4 homohexamer. *Proceedings of the National Academy of Sciences, 100*, 7395-7400

Adachi, M., Takenaka, Y., Gidamis, A. B., Mikami, B., & Utsumi, S. (2001). Crystal structure of soybean proglycinin A1aB1b homotrimer. *Journal of Molecular Biology, 305*, 291-305.

Agriculture and Agri-Food Canada. (2015). Carinata: An oilseed crop ready to lift-off. Retrieved 2016 from: http://www.agr.gc.ca/eng/news/science-of-agricultural-innovation/carinata-an-oilseed-ready-for-lift-off/?id=1389626028671

Agrisoma Bioscience Inc. (2014). CFIA grants Agrisoma and PGF Biofuels approval for *Brassica Carinata* meal. Retrieved 2015, from: http://agrisoma.com/images/pdfs/News%20Releases/Press%20Release%20%20Carinata%20CFIA%20Meal%20Approval%20%206%2025%2014.pdf

Agrisoma Bioscience Inc. (2015). Sustainable energy founded in agriculture. Retrieved 2016, from:http://www1.agric.gov.ab.ca/$Department/deptdocs.nsf/all/crop15187/$FILE/agrisoma-carinata-agronomy-and-marketing-2015.pdf

Aider, M., & Barbana, C. (2011). Canola proteins: composition, extraction, functional properties, bioactivity, applications as a food ingredient and allergenicity - A practical and critical review. *Trends in Food Science & Technology, 22*, 21-39

Alemayehu, A., & Becker, H. (2005). Quantitative genetic analysis of total glucosinolate, oil and protein contents in Ethiopian mustard (*Brassica carinata* A. Braun). *Ethiopian Journal of Sci*ence**,** *28*, 141-150

Aluko, R. E. & McIntosh, T. (2001). Polypeptide profile and functional properties of defatted meals and protein isolates of canola seeds. *Journal of the Science of Food and Agriculture*, 81, 391-396

Amara, I., Zaidi, I., Masmoudi, K., Ludevid, M., Pagès, M., Goday, A., & Brini, F. (2014) Insights into late embryogenesis abundant (LEA) proteins in plants: From structure to the functions. *American Journal of Plant Sciences, 5*, 3440-3455

Anonmous. (n.d.-a). Characterizing protein stability by DSC. Retrieved 2016 from: https://www.bioch.ox.ac.uk/aspsite/services/equipmentbooking/biophysics/CSCnote3.pdf

Anonymous. (n.d.-b). Fluorescence data analysis methods. *Technical Note OPTIM®1000*. Retrieved 2016, from:http://www.isogenlifescience.com/uploads/o4/21/o4210uzD8OIHHtyOEhZEkQ/Optim06v1.0---Fluorescence-data-analysis-methods-ILS.pdf

Anonymous. (n.d.-c). Normalizing spectral counts in Scaffold: Proteome software. Retrieved 2016, from: http://proteome-software.wikispaces.com/file/view/normalizing-spectral-counts-in-scaffold

Anonymous. (2014). Scaffold, version 4.0 user's manual. Retrieved 2016, from: http://proteome-software.wikispaces.com/file/view/Scaffold_4_UG.pdf

AOAC. (2005a). AOAC Official Method 934.01, Loss on drying (Moisture) at 95^{O}C – 100^{O}C for Feeds. In D. G. W. Latimer, Jr. & D. W. Horwitz (Eds.), *Official Methods of Analysis of AOAC International* (18th ed.): AOAC Intl.

AOAC. (2005b). AOAC Official Method 942.05, Ash of animal feed. In D. G. W. Latimer, Jr. & D. W. Horwitz (Eds.), *Official Methods of Analysis of AOAC International* (18th ed.): AOAC Intl.

AOAC. (2005c). AOAC Official Method 990.03, Protein (crude) in animal feed-combustion method. In D. G. W. Latimer, Jr. & D. W. Horwitz (Eds.), *Official Methods of Analysis of AOAC International* (18th ed.): AOAC Intl.

AOAC. (2005d). AOAC Official Method 994.12, Amino acids in feeds. In D. G. W. Latimer, Jr. & D. W. Horwitz (Eds.), *Official Methods of Analysis of AOAC International* (18th ed.): AOAC Intl.

AOCS. (1997). AOCS official method Am 2−93, determination of oil content in oilseeds. In D. E. Firestone (Ed.), *Official Methods and Recommended Practices of the AOCS* (5th ed.): AOCS Press: Champaign, IL.

Apenten, R. K. O., & Folawiyo, Y. L. (1995). The effect of pH on rapeseed globulin (cruciferin) binding to anilinonaphthalene-8-sulfonate. *Journal of Food Biochemistry, 19*(6), 455-465

Arakawa, T., Kita, Y. A., & Narhi, L. O. (1991). Protein-ligand interaction as a method to study surface properties of proteins. *Methods Biochemical Analysis*, 35, 87-125

Arakawa, T., Philo, J. S., Ejima, D., Tsumoto, K., & Arisaka, F. (2006). Aggregation analysis of therapeutic proteins, part 1: General aspects and techniques for assessment. *Bioprocess International, 4*, 42-49

Ashton, F. M. (1976). Mobilization of storage proteins of seeds. *Annual Review of Plant Physiology, 27*, 95-117

Aziza, A. E., Quezada, N., & Cherian, G. (2010). Antioxidative effect of dietary camelina meal in fresh, stored, or cooked broiler chicken meat. *Poultry Science, 89*, 2711-2718

Badley, R. A., Atkinson, D., Hauser, H., Oldani, D., Green, J. P., & Stubbs, J. M. (1975). The structure, physical and chemical properties of the soy bean protein glycinin. *Biochimica et Biophysica Acta (BBA) - Protein Structure, 412*, 214-228

Balat, M., & Balat, H. (2010). Progress in biodiesel processing. *Applied Energy*, 87, 1815-1835

Ballesteros, M., Oliva, J. M., Negro, M. J., Manzanares, P., & Ballesteros, I. (2004). Ethanol from lignocellulosic materials by a simultaneous saccharification and fermentation process (SFS) with *Kluyveromyces marxianus* CECT 10875. *Process Biochemistry*, 39, 1843-1848

Barciszewski, J., Szymanski, M., & Haertle, T. (2000). Minireview: analysis of rape seed napin structure and potential roles of the storage protein. *Journal of Protein Chemistry, 19*, 249-254

Barthet, V. J. (2014). Quality of western Canadian canola 2014. Retrieved 2015, from: http://www.grainscanada.gc.ca/canola/harvest-recolte/2014/hqc14-qrc14-en.pdf

Berk, Z. (1992). Technology of production of edible flours and protein products from soybeans. Retrieved 2016, from: http://www.fao.org/docrep/t0532E/t0532e00.htm#con

Bérot, S., Compoint, J. P., Larré, C., Malabat, C., & Guéguen, J. (2005). Large scale purification of rapeseed proteins (*Brassica napus* L.). *Journal of Chromatography* B, 818(1), 35-42.

Bethke, P. C., & Jones, R. L. (2001). Cell death of barley aleurone protoplasts is mediated by reactive oxygen species. *Plant Journal*, *25*(1), 19-29

Bhatty, R. S., McKenzie, S. L., & Finlayson, A. J. (1968). The proteins of rapeseed (*Brassica napus* L.) soluble in salt solutions. *Canadian Journal of Biochemistry*, *46*, 1191-1197

Brown. J. (n.d). Increased value-added products from *Brassicaceae* oilseed crops. Retrieved 2015, from: http://pnwsteep.wsu.edu/directseed/conf2k4/pdf/2004_Proc_Brown.pdf

Byler, D. M., & Susi, H. (1986). Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers, 25*, 469-487

Canadian Food Inspection Agency (CFIA). (2014). The Biology of *Camelina sativa* (L.) Crantz (Camelina). Retrieved 2016, from: http://www.inspection.gc.ca/plants/plants-with-novel-

traits/applicants/directive-94-08/biology-documents/camelina-sativa-l-/eng/1330971423348/1330971509470

Canadian Renewable Fuels Association (CRFA). (2014). From biofuels to bioeconomy: Evolution & growth. Retrieved 2015, from: http://greenfuels.org/wp-content/uploads/2014/04/EVO-ENG-total.pdf

Canola Council of Canada (CCC). (2013). Economic impact of the canola industry. Retrieved 2016, from: http://www.canolacouncil.org/markets-stats/industry-overview/economic-impact-of-the-canola-industry/

Canola Council of Canada (CCC). (n.d.-a). Health benefits of canola oil. Retrieved 2015, from: http://www.canolacouncil.org/oil-and-meal/canola-oil/health-benefits-of-canola-oil/

Canola Council of Canada (CCC). (n.d.-b). The impact of a canola-based biodiesel industry in Canada. Retrieved 2015, from: http://www.canolacouncil.org/media/509101/7.pdf

Cappellozza, B. I., Cooke, R. F., Bohnert, D. W., Cherian, G., & Carroll, J. A. (2012). Effects of camelina meal supplementation on ruminal forage degradability, performance, and physiological responses of beef cattle. *Journal of Animal Science, 90*, 4042-4054

Casey, R., Domoney, C., & Ellis, N. (1986). Legume storage proteins and their genes. In B.J. Miflin (Ed.), *Oxford Surveys of Plant Molecular and Cell Biology* (pp. 1-95), Oxford University Press, Oxford, United Kingdom

Chakrabarti, M. H., & Ahmad, R. (2009). Investigating possibility of using least desirable edible oil of *Eruca sativa* L. in bio diesel production. *Pakistan Journal of Botany, 41*, 481-487

Chen, Y-C., Liu, W-L., Hsu, H-C. Lee Y. A & Chen, C-S. (2004). Purification and characterization of isoforms of b-N-acetylhexosaminidase from mungbean seedlings. *Botanical Bulletin of Academia Sinica, 45,* 275-283

Cherian, G., Campbell, A., & Parker, T. (2009). Egg quality and lipid composition of eggs from hens fed *Camelina sativa*. *Journal of Applied Poultry Research, 18*, 143-150

Clapés, P., & Infante, R. M. (2002). Amino acid-based surfactants: Enzymatic synthesis, properties and potential applications. *Biocatalysis and Biotransformation, 20,* 215-233

Clark, A. (Ed.). (2007). *Managing cover crops profitably (3rd ed.)*. Sustainable Agriculture Research Education Program, Maryland, USA. Retrieved 206, from: http://www.sare.org/Learning-Center/Books/Managing-Cover-Crops-Profitably-3rd-Edition

Conn, K. L., Tewari, J. P., & Dahiya, J. S. (1988). Resistance to *Alternaria brassicae* and phytoalexin-elicitation in rapeseed and other crucifers. *Plant Science, 56*(1), 21-25

Crouch, M. L., & Sussex, I. M. (1981). Development and storage-protein synthesis in *Brassica napus* L. embryos in vivo and in vitro. *Planta, 153*, 64-74.

Crouch, M. L., Tenbarge, K. M., Simon, A. E., & Ferl, R. (1983). cDNA clones for *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide. *Journal of Molecular and Applied Genetics, 2*, 273-283

Cui W., Eskin N. A. M., Biliaderis C. G. (1993). Chemical and physical properties of yellow mustard (*Sinapis alba* L) mucilage. *Food Chemistry, 46*,169-76

Cui, S. W., Eskin, M. A. N., Wu, Y., & Ding, S. (2006). Synergisms between yellow mustard mucilage and galactomannans and applications in food products-A mini review. *Advances in Colloid and Interface Science, 128–130,* 249-256

Dalgalarrondo, M., Robin, J.-M., & Azanza, J.-L. (1986). Subunit composition of the globulin fraction of rapeseed (*Brassica napus*). *Plant Science, 43*, 115-124.

Damodaran, S. (2008). Amino acids, peptides and proteins. In S. Damodaran, K. L. Parkin & O. R. Fennema (Eds.), *Fennema's Food Chemistry* (pp. 217-330): CRC Press, Boca Raton, FL, USA

Danyard, K., & Danyard, T. (2011). What are the effects of biofuels and bioproducts on the environment, crop and food prices and world hunger? Retrieved 2015, from: http://www.gfo.ca/LinkClick.aspx?fileticket=HKfOeU3cHTI%3D&tabid=139

Delseny, M., & Raynal, M. (1999). Globulin storage proteins in crucifers and non-legume dicotyledonous families. In: P. R. Shewry and R. Casey (Eds.), *Seed Proteins* (pp. 427-451), Kluwer Academic, Amsterdam, Netherland.

Dendukuri, D., & Diosady, L. L. (2003). Oil-free protein isolates from full fat dehulled flour by microfiltration. *Journal of the American Oil Chemists' Society*, *80,* 287-294

Derouin, J. M. (2014). Bioproducts in Canada: A federal perspective. Retrieved 2015, from: www.gtmconference.ca/site/index.php/2014-presentations/doc_download/537-2014-5a-derouin+&cd=2&hl=en&ct=clnk&gl=ca

Dickinson, C. D., Hussein, E. H. A., & Nielsen, N. C. (1989). Role of posttranslational cleavage in glycinin assembly. *Plant Cell1, 1*, 459-469

Dozier, W. A. & Hess, J. B. (2011). Soybean meal quality and analytical techniques, soybean and nutrition, Prof. Hany El-Shemy (Ed.), InTech, DOI: 10.5772/24161. Retrieved 2016, from: http://www.intechopen.com/books/soybean-and-nutrition/soybean-meal-quality-and-analytical-techniques

Ehrensing, D. T., & Guy, S. O. (2008). Camelina. Retrieved 2016, from: http://extension. oregonstate.edu/gilliam/sites/default/files/Camelina_em8953-e.pdf

Elementis Specialties, Inc, (2010). Research and development report: Fancor® Abyssinian oil. Retrieved 2015, from: http://www.essentialingredients.com/pdf/AbyssinianOilResearch Report.pdf

Endres, G., & Schatz, B. (2003). Crambe production. Retrieved 2015, from: https://www. ag. ndsu. edu/pubs/plantsci/crops/a1010.pdf

Ericson, M. L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L. G., & Rask, L. (1986). Structure of the rapeseed 1.7 S storage protein, napin, and its precursor. *Journal of Biolgycal Chemistry, 261*, 14576-14581

Eynck, C., & Falk, K. C. (2013). Camelina (*Camelina sativa*). In B. P. Sing (Ed.), *Biofuel Crops: Production, Physiology and Genetics* (pp. 369-391). CAB International

Falk, K. C., & Klein-Gebbinck, H. (n.d). Camelina breeding at AAFC. Retreived 2013, from:. http://www.agriculture.gov.sk.ca/Default.aspx?DN=8aec8539-4a04-44de-b61e-e2701e3b8029

Feddern, V., Junior, A. C., De Prá, M. C., Abreu, P. G. D., Filho, J. I. D. S., Higarashi, M. M., Sulenta, M., & Coldebella, A. (2011). Animal fat wastes for biodiesel production. In Dr. Margarita Stoytcheva (Ed.), *Biodiesel - Feedstocks and Processing Technologies* (pp.45-70). doi: 10.5772/26691

FEDIOL. (2016). World production data: Vegetable oils world production. Retrieved 2016, from: http://www.fediol.eu/web/world+production+data/1011306087/list1187970075/f1.html

Feussner, I. (2015). Camelina-a promising oilseed crop to contribute to the growing demand for vegetable oils. *European Journal of Lipid Science and Technology, 117*, 271-273

Filiz, E., Ozyigit, I. I., Tombuloglu, H., & Koc, I. (2013). In silico comparative analysis of LEA (Late Embryogenesis Abundant) proteins in *Brachypodium distachyon* L. *Plant Omics, 6*, 433

Fink, A. L. (2001). Molten Globule *eLS*: John Wiley & Sons, Ltd.

Fink, A. L., Seshadri, S., Khurana, R., & Oberg, K. A. (1999). Determination of secondary structure in protein aggregates using attenuated total reflectance FTIR. *Infrared Analysis of Peptides and Proteins* (Vol. 750, pp. 132-144): American Chemical Society

Flanders, A., & Abdulkarim, S. M. (1985). The composition of seed and seed oils of Taramira (*Eruca sativa*). *Journal of the American Oil Chemists' Society, 62*, 1134-1135.

Fleenor, R. A. (2011). Plant guide for camelina (*Camelina sativa*). Retrieved 2015, from: http://plants.usda.gov/plantguide/pdf/pg_casa2.pdf

Fobert, P. R., Smith, M. A., Zou, J., Mietkiewska, E., Keller, W. A., & Taylor, D. C. (2008). Developing Canadian seed oils as industrial feedstocks. *Biofuels, Bioproducts and Biorefining, 2*, 206-214

Folawiyo, Y. L., & Apenten, R. K. O. (1997). The effect of heat- and acid-treatment on the structure of rapeseed albumin (napin). *Food Chemistry, 58*, 237-243

Food and Agriculture Organization (FAO). (2014). Oil crops: Food outlook 2014. Retrieved 2016, from: http://www.fao.org/fileadmin/templates/est/comm_markets_monitoring/Oilcrops/Documents/Food_outlook_oilseeds/food_outlook_may_2014_oilcrops.pdf

Fröhlich, A., & Rice, B. (2005). Evaluation of *Camelina sativa* oil as a feedstock for biodiesel production. *Industrial Crops and Products, 21*(1), 25-31

Fujimoto, Y., Koiwa, Y., Nagaoka, K., & Tatsukawa, K. (1972). Polymer solutions of polyamino acids and the process for their preparation: Google Patents. Retrieved 2016, from: http://www.google.ch/patents/US3671482

Fukuda, T., Maruyama, N., Salleh, M. R., Mikami, B., & Utsumi, S. (2008). Characterization and crystallography of recombinant 7S globulins of Adzuki bean and structure-function relationships with 7S globulins of various crops. *Journal of Agricultural and Food Chemistry, 56*, 4145-4153

Gabbott, P. (2008). Principles and applications of thermal analysis: Blackwell Publishing Ltd.

Gao, C., Taylor, J., Wellner, N., Byaruhanga, Y. B., Parker, M. L., Mills, E. N., & Belton, P. S. (2005). Effect of preparation conditions on protein secondary structure and biofilm formation of kafirin. *Journal of Agriculture and Food Chemistry, 53*, 306-312

Gatehouse, J. A., Croy, R. R. D., & Boulter, D. (1984). The synthesis and structure of pea storage proteins. *CRC Critical Review in Plant Science*, 1, 287-314

Gehrig, P. M., & Biemann, K. (1996). Assignment of the disulfide bonds in napin, a seed storage protein from *Brassica napus*, using matrix-assisted laser desorption ionization mass spectrometry. *Peptide Research, 9*, 308-314

Gehrig, P. M., Krzyzaniak, A., Barciszewski, J., & Biemann, K. (1998). Mass spectrometric amino acid sequencing of a mixture of seed storage proteins (napin) from *Brassica napus* products of a multigene Family. *Biochemistry*, 93:3647-3652

GenomePrairie. (2014). Energizing commercialization of industrial oilseed crops. Retrieved 2016, from: http://www.genomeprairie.ca/project/previous/prairie-gold/

Gillespie, J., Rogers, S. W., Deery, M., Dupree, P., & Rogers, J. C. (2005). A unique family of proteins associated with internalized membranes in protein storage vacuoles of the *Brassicaceae*. *Plant Journal, 41*, 429-441

Glaser, L. L. (1996). Crambe: An economic assessment of the feasibility of providing multiple-peril crop insurance. Retrieved 2015, from: http://www.rma.usda.gov/ pilots/feasible/pdf/ crambe.pdf

Goldberg, R. B., Barker, S. J., & Perez-Grau, L. (1989). Regulation of gene expression during plant embryogenesis. *Cell*, 56, 149-160

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols*, 1, 2876-2890

Gueguen, J., Chevalier, M., And, J. B., & Schaeffer, F. (1988). Dissociation and aggregation of pea legumin induced by pH and ionic strength. *Journal of the Science of Food and Agriculture, 44*, 167-182.

Gugel, R. K., & Falk, K. C. (2006). Agronomic and seed quality evaluation of *Camelina sativa* in western Canada. *Canadian Journal of Plant Science*, *86*, 1047-1058

Gutterman, Y., & Shem-Tov, S. (1996). Structure and function of the mucilaginous seed coats of plantago coronopus inhabiting the negev desert of ISRAEL. *Israel Journal of Plant Sciences, 44*, 125-133

Halmemies-Beauchet-Filleau, A., Kokkonen, T., Lampi, A. M., Toivonen, V., Shingfield, K. J., & Vanhatalo, A. (2011). Effect of plant oils and camelina expeller on milk fatty acid composition in lactating cows fed diets based on red clover silage. *Journal of Dairy Science, 94*, 4413-4430

Hand, S. C., Menze, M. A., Toner, M., Boswell, L., & Moore, D. (2011). LEA proteins during water stress: not just for plants anymore. *Annual Review of Physiology, 73*, 115-134

Health Canada. 2012. Food and nutrition. Camelina oil. Retrieved 2015, from: http://www.hc-sc.gc.ca/fn-an/gmf-agm/appro/camelina-cameline-eng.php

Hellevang, K. (n.d). Canola drying and storage management. Retrieved 2016 from: https://www.ag.ndsu.edu/pubs/plantsci/crops/a1280k.pdf

Henderson, A., Hallett, R., & Soroka, J. (2004). Prefeeding Behavior of the Crucifer Flea Beetle, *Phyllotreta cruciferae*, on host and nonhost Crucifers. *Journal of Insect Behavior*, *17*(1), 17-39

Herman, E. M., & Larkins, B. A. (1999). Protein storage bodies and vacuoles. *The Plant Cell*, *11*, 601-613.

Hixson, S. M., & Parrish, C. C. (2014). Substitution of fish oil with camelina oil and inclusion of camelina meal in diets fed to Atlantic cod (*Gadus morhua*) and their effects on growth, tissue lipid classes, and fatty acids. *Journal of Animal Science, 92*, 1055-1067

Hixson, S. M., Parrish, C. C., Wells, J. S., Winkowski, E. M., & Anderson, D. M. (2015a). Inclusion of camelina meal as a protein source in diets for farmed Atlantic cod *Gadus morhua*. *Aquaculture Research*. doi: 10.1111/are.12710

Hixson, S. M., Parrish, C. C., Wells, J. S., Winkowski, E. M., Anderson, D. M., & Bullerwell, C. N. (2015b). Inclusion of camelina meal as a protein source in diets for farmed salmonids. *Aquaculture Nutrition*. doi: 10.1111/anu.12276

Ho, D. P., Ngo, H. H., & Guo, W. (2014). A mini review on renewable sources for biofuel. *Bioresource Technology, 169,* 742-749

Hofsten, A. (1974). Cellular structure of rapeseed. Retrieved 2016, from: http://gcirc.org/fileadmin/documents/Proceedings/IRC1970Quebec/CO1970QUA01.pdf.

Holt. C. (2000). Molecular basis of whey protein food functionalities. *Australian Journal of Dairy Technology, 55,* 53-55

Hong-Bo, S., Zong-Suo, L., & Ming-An, S. (2005). LEA proteins in higher plants: structure, function, gene expression and regulation. *Colloids Surface B Biointerfaces, 45*, 131-135

Hrastar, R., Petrisic, M. G., Ogrinc, N., & Kosir, I. J. (2009). Fatty acid and stable carbon isotope characterization of *Camelina sativa* oil: implications for authentication. *Journal of Agriculture and Food Chemistry, 57*, 579-585

Hu, Z.-Y., Hua, W., Zhang, L., Deng, L.-B., Wang, X.-F., Liu, G.-H., . . . Wang, H.-Z. (2013). Seed structure characteristics to form ultrahigh oil content in rapeseed. *PLoS ONE, 8,* e62099. doi: 10.1371/journal.pone.0062099

Huang, A. H. C. (1992). Oil bodies and oleosins in seeds. *Annual Review of Plant Physiology and Plant Molecular Biology, 43*, 177-200

Huang, A. H. C. (1996). Oleosins and oil bodies in seeds and other organs. *Plant Physiology, 110,* 1055-1061

Huang, Z., & Gutterman, Y. (1999). Water absorption by mucilaginous achenes of artemisia monosperma: Floating and germination as affected by salt concentrations. *Israel Journal of Plant Sciences, 47*(1), 27-34

Huang, Z., Gutterman, Y., & Osborne, D. J. (2004). Value of the mucilaginous pellicle to seeds of the sand-stabilizing desert woody shrub *Artemisia sphaerocephala* (*Asteraceae*). *Trees, 18,* 669-676

Hutcheon, C., Ditt, R. F., Beilstein, M., Comai, L., Schroeder, J., Goldstein, E., Shewmaker, C. K., Nguyen, T., De Rocher, J., & Kiser, J. (2010). Polyploid genome of *Camelina sativa* revealed by isolation of fatty acid synthesis genes. *BMC Plant Biology, 10*, 233-247

Industry Canada. (2004). Innovation roadmap on bio-based feedstocks, fuels and industrial products. Retrieved 2015, from: http://publications.gc.ca/collections/Collection/Iu44-11-2004E.pdf

Ingram, J. & Bartels, D. (1996). The molecular basis of dehydratation tolerance in plants. *Annual Review of Plant Physiology and Plant Molecular Biology, 47,* 377-403

International Energy Agency (IEA). (2012). World energy outlook 2012: Renewable energy outlook. Retrieved 2015, from: http://www.worldenergyoutlook.org/ media/ weowebsite/ 2012/WEO2012_Renewables.pdf

Ivanov, K., Stoimenova, A., Obreshkova, D., & Saso, L. (2013). Biotechnology in the Production of Pharmaceutical Industry Ingredients: Amino acids. *Biotechnology & Biotechnological Equipment, 27*, 3620-3626

Jarpa-Parra, M., Bamdad, F., Tian, Z., Zeng, H., Temelli, F., & Chen, L. (2015). Impact of pH on molecular structure and surface properties of lentil legumin-like protein and its application as foam stabilizer. *Colloids Surfaces B: Biointerfaces, 132*, 45-53

Jason, T. C. T., Cao, Y.-z., Laurent, P., Ratnayake, C., & Huang, A. H. C. (1993). Lipids, proteins, and structure of seed oil bodies from diverse species. *Plant Physiology, 101*, 267-276

Jiang, Y., Caldwell, C. D., & Falk, K. C. (2014). Camelina seed quality in response to applied nitrogen, genotype and environment. *Canadian Journal of Plant Science, 94*, 971-980

Job, C., Rajjou, L., Lovingny, Y., Belghazi, M., & Job, D. (2005). Patterns of protein oxidation in *Arabidopsis* seeds during germination. *Plant Physiology, 138*, 790-802

Johnson, E., Falk, K., & Eynck, K. (2014). *Brassica carinata* and *Camelina sativa* Retrieved 2015, from: http://www.usask.ca/soilsncrops/conference-proceedings/2014%20pdf/day-2-presentations/07-johnson.pdf.

Jolivet, P., Boulard, C., Bellamy, A., Larré, C., Barre, M., Rogniaux, H., . . . Nesi, N. (2009). Protein composition of oil bodies from mature *Brassica napus* seeds. *Proteomics, 9*, 3268-3284

Jolivet, P., Boulard, C., Bellamy, A., Valot, B., d'Andrea, S., Zivy, M., . . . Chardot, T. (2011). Oil body proteins sequentially accumulate throughout seed development in Brassica napus. *Journal of Plant Physiology, 168*, 2015-2020

Jolivet, P., Roux, E., d'Andrea, S., Davanture, M., Negroni, L., Zivy, M., & Chardot, T. (2004). Protein composition of oil bodies in *Arabidopsis thaliana* ecotype WS. *Plant Physiology and Biochemistry, 42*, 501-509

Jolivet, P., Tailliart, K., Boulard, C., Nesl, N., & Chardot, T. (2006). Purification and protein composition of oil bodies from *Brassica napus* seeds. *OCL, 9*. doi: http://dx.doi.org/ 10.1051/ocl.2006.0050

Jyothi, T. C., Sinha, S., Singh, S. A., Surolia, A., & Appu Rao, A. G. (2007). Napin from *Brassica juncea*: Thermodynamic and structural analysis of stability. *Biochimica Biophysica Acta, 1774*, 907-919

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W. E., Tuteja, R., . . . Parkin, I. A. (2014). The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature Communications*, 5, 3706. doi: 10.1038/ncomms4706

Kahindi, R. K., Woyengo, T. A., Thacker, P. A., & Nyachoti, C. M. (2014). Energy and amino acid digestibility of camelina cake fed to growing pigs. *Animal Feed Science and Technology, 193*, 93-101

Kakani, R., Fowler, J., Haq, A. U., Murphy, E. J., Rosenberger, T. A., Berhow, M., & Bailey, C. A. (2012). Camelina meal increases egg n-3 fatty acid content without altering quality or production in laying hens. *Lipids, 47*, 519-526

Katavic, V., Agrawal, G. K., Hajduch, M., Harris, S. L., & Thelen, J. J. (2006). Protein and lipid composition analysis of oil bodies from two *Brassica napus* cultivars. *Proteomics, 6*, 4586-4598

Kauppinen, J. K., Moffatt, D. J., Mantsch, H. H., & Cameron, D. G. (1981). Fourier self-deconvolution: A method for resolving intrinsically overlapped bands. *Applied Spectroscopy*, 35, 271-276

Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. *Biochimica Biophysica Acta, 1751*, 119-139

Kim, K. S., Kim, S., Yang, H. J., & Kwon, D. Y. (2004). Changes of glycinin conformation due to pH, heat and salt determined by differential scanning calorimetry and circular dichroism. *International Journal of Food Science & Technology*, 39, 385-393

Kim, J. T., & Netravali, A. N. (2012). Non-food application of camelina meal: Development of sustainable and green biodegradable paper-camelina composite sheets and fibers. *Polymer Composites, 33*, 1969-1976

Kong, J., & Yu, S. (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochimica Biophysica Sinica (Shanghai), 39*, 549-559

Korte, T., & Herrmann, A. (1994). pH-dependent binding of the fluorophore bis-ANS to influenza virus reflects the conformational change of hemagglutinin. *European Biophysics Journal*, 23, 105-113

Kramer, Ryan M., Shende, Varad R., Motl, N., Pace, C. N., & Scholtz, J. M. (2012). Toward a molecular understanding of protein solubility: Increased negative surface charge correlates with increased solubility. *Biophysics Journal, 102*, 1907-1915

Krzyzaniak, A., Burova, T., Haertlé, T., & Barciszewski, J. (1998). The structure and properties of Napin-seed storage protein from rape (*Brassica napus* L.). *Food / Nahrung, 42*, 201-204

Kuang, A., Xiao, Y., McClure, G., & Musgrave, M. E. (2000). Influence of microgravity on ultra-structure and storage reserves in seeds of *Brassica rapa*. *Annals of Botany, 85*, 851-859

Kwok, S. C., Mant, C. T., & Hodges, R. S. (2002). Importance of secondary structural specificity determinants in protein folding: Insertion of a native β-sheet sequence into an α-helical coiled-coil. *Protein Science, 11*, 1519-1531

Laemmli, U. K. (1970). Glycine-SDS-PAGE for separation of proteins. *Nature*, 227, 680-685

Lakemond, C. M., de Jongh, H. H., Hessing, M., Gruppen, H., & Voragen, A. G. (2000). Heat denaturation of soy glycinin: influence of pH and ionic strength on molecular structure. *Journal of Agriculture abd Food Chemistry, 48*, 1991-1995

Lazzeri, L., Leoni, O., Conte, L. S., & Palmieri, S. (1994). Papers presented at the 2nd European symposium on industrial crops and products some technological characteristics and potential uses of *Crambe abyssinica* products. *Industrial Crops and Products, 3*(1), 103-112

Li, N., Qi, G., Sun, X. S., Wang, D., Bean, S., & Blackwell, D. (2014). Isolation and characterization of protein fractions isolated from camelina meal. *Transactions of the ASABE, 57*(1), 169-178

Li, N., Qi, G., Sun, X. S., Xu, F., & Wang, D. (2015). Adhesion properties of camelina protein fractions isolated with different methods. *Industrial Crops and Products, 69*, 263-272

Lin, L.-J., Liao, P.-C., Yang, H.-H., & Tzen, J. T. C. (2005). Determination and analyses of the N-termini of oil-body proteins, steroleosin, caleosin and oleosin. *Plant Physiology and Biochemistry, 43*, 770-776

Liu, X., Brost, J., Hutcheon, C., Guilfoil, R., Wilson, A. K., Leung, S., . . . de Rocher, J. (2012). Transformation of the oilseed crop *Camelina sativa* by *Agrobacterium*-mediated floral dip and simple large-scale screening of transformants. *In Vitro Cellular and Developmental Biology - Plant, 48*, 462-468

Lönnerdal, B., & Janson, J. C. (1972). Studies on *Brassica* seed proteins. 1. The low molecular weight proteins in rapeseed. Isolation and characterization. *Biochimca Biophysica Acta, 278*, 175-183

Lott, N. A. (1980). Protein bodies. In N. E. Tolbert (Ed), *The Biochemistry of Plants, Vol.1* (pp.589-623). Academic Press, New York

Lu, C., & Kang, J. (2008). Generation of transgenic plants of a potential oilseed crop *Camelina sativa* by *Agrobacterium*-mediated transformation. *Plant Cell Reports, 27*, 273-278

Macdonald, F. D., & Preiss, J. (1985). Partial purification and characterization of granule-bound starch synthases from normal and waxy maize. *Plant Physiology, 78*, 849-852

Macquet, A., Ralet, M. C., Kronenberger, J., Marion-Poll, A., & North, H. M. (2007). In situ, chemical and macromolecular study of the composition of *Arabidopsis thaliana* seed coat mucilage. *Plant Cell Physiology, 48*, 984-999

Mag, T. (n.d). Canola seed and oil processing. Retrieved 2016, from: http://www.canolacouncil.org/media/515283/canola_seed_and_oil_processing.pdf

Malabat, C., Atterby, H., Chaudhry, Q., Renard, M., & Gu´eguen, J. (2003*).* Genetic variability of rapeseed protein composition. *Proceedings of the 11th International Rapeseed Congress*, 4, 205-208

Marambe, P. W. M. L. H. K., Shand, P. J., & Wanasundara, J. P. D. (2008). An in-vitro investigation of selected biological activities of hydrolysed flaxseed (*Linum usitatissimum* L.) Proteins. *Journal of the American Oil Chemists' Society, 85*, 1155-1164

Marcone, M. F., Beniac, D. R., Harauz, G., & Yada, R. Y. (1994). Quaternary structure and model for the oligomeric seed globulin from *Amaranthus hypochondriacus* K343. *Journal of Agricultural and Food Chemistry, 42*, 2675-2678

Marnoch, R., & Diosady, L. (2006). Production of mustard protein isolate from oriental mustard seed (*Brassica juncea* L.). *Journal of the American Oil Chemists' Society, 83*, 65-69

Marquez-Lema, A., Fernandez-Martinez, J. M., Perez-Vich, b., & Velasco, L. (2009). Inheritance of very high glucosinolate content in Ethiopian mustard seeds. *Plant Breeding, 128*, 278-281

Marty, F. (1999). Plant Vacuoles. *Plant Cell, 11*, 587-599

Matthaus, B., & Zubr, J. (2000). Variability of specific components in Camelina sativa oilseed cakes. *Industrial Crops and Products*, *12*(1), 9-18

Massoura, E., Vereijken, J. M., Kolster, P., & Derksen, J. T. P. (1998). Proteins from *Crambe abyssinica* oilseed. 1. Isolation procedure. *Journal of the American Oil Chemists' Society*, 75, 323-327

Maurer, S., Waschatko, G., Schach, D., Zielbauer, B. I., Dahl, J., Weidner, T., Bonn, M., & Vilgis, T. A. (2013). The role of intact oleosin for stabilization and function of oleosomes. *The Journal of Physical Chemistry B, 117*, 13872-13883

McClements, D. J. (2005). Food emulsions: Principles, practices, and techniques. Florida, US. : CRC Press

McIlwain, S., Mathews, M., Bereman, M. S., Rubel, E. W., MacCoss, M. J., & Noble, W. S. (2012). Estimating relative abundances of proteins from shotgun proteomics data. *BMC Bioinformatics, 13*, 308

Meesapyodsuk, D., & Qiu, X. (2011). A peroxygenase pathway involved in the biosynthesis of epoxy fatty acids in oat. *Plant Physiology, 157(1)*, 454-463

Miernyk, J. A., & Hajduch, M. (2011). Seed proteomics. *Journal of Proteomics, 74*, 389-400

Mohiuddin, S., Qureshi, S. A., Nasir, M. K. A., Khatsi, L.M. (1990). Studies on the repellent activity of some indigenous plant oils against *Tribolium castaneum* (Herbst.). *Pakistan Journal of Scientific and Industrial Research, 33*, 326-328

Monsalve, R. I., & Rodriguez, R. (1990). Purification and Characterization of Proteins from the 2S Fraction from Seeds of the *Brassicaceae* Family. *Journal of Experimental Botany, 41*(1), 89-94

Morra, G. (2006). Role of electrostatics explored with molecular dynamics simulations for protein stability and folding (Doctoral thesis). Freie Universität Berlin, Berlin, Germany. Retrieved 2016, from: http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000001956

Moser, B. R. (2010). Camelina (*Camelina sativa* L.) oil as a biofuels feedstock: Golden opportunity or false hope? *Lipid Technology, 22,* 270-273

Moser, B. R., & Vaughn, S. F. (2010). Evaluation of alkyl esters from *Camelina sativa* oil as biodiesel and as blend components in ultra-low-sulfur diesel fuel. *Bioresource Technology, 101*, 64-653

Mozhaev, V. V., Heremans, K., Frank, J., Masson, P., & Balny, C. (1996). High pressure effects on protein structure and function. *Proteins, 24*(1), 81-91

Mumtaz, M. W., Adnan, A., Mahmood, Z., Mukhtar, H., Danish, M., & Ahmad, Z. (2012). Biodiesel production using *Eruca sativa* oil: optimization and characterization. *Pakistan Journal of Botany,* 44, 1111–1120

Muñoz, L. A., Cobos, A., Diaz, O., & Aguilera, J. M. (2012). Chia seeds: Microstructure, mucilage extraction and hydration. *Journal of Food Engineering, 108*(1), 216-224

Müntz, K. (1998). Deposition of storage proteins. *Plant Molecular Biology, 38*, 77-99

Murén, E., Ek, B., Björk, I., & Rask, L. (1996). Structural comparison of the precursor and the mature form of napin, the 2S storage protein in *Brassica napus. European Journal of Biochemistry, 242*, 214-219

Murphy, D. J., Cummins, I., & Kang, A. S. (1989). Synthesis of the major oil-body membrane protein in developing rapeseed (*Brassica napus*) embryos. Integration with storage-lipid and storage-protein synthesis and implications for the mechanism of oil-body formation. *Biochemical Journal, 258,* 285-293

Myers, J. K., & Oas, T. M. (2001). Preorganized secondary structure as an important determinant of fast protein folding. *Nature Structural Biology, 8*, 552-558

Najda, H. (1991). Forage Brassicas. Retrieved 2016, from: http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/agdex135/$file/128_20-1.pdf?OpenElement

Nakai, S. (1983). Structure-function relationships of food proteins with an emphasis on the importance of protein hydrophobicity. *Journal of Agricultural and Food Chemistry, 37,* 676–680

Natural Resources Canada (NRC). (n.d). Next-generation biofuels Fund. Retrieved 2015, from: http://www.nrcan.gc.ca/energy/alternative-fuels/programs/3643

Nelson, L. A., Grombacheer, A., & Baltensperger, D.D. (1993). G93-1126 Crambe production. *Historical Materials from University of Nebraska-Lincoln Extension*. Paper 776. Retrieved 2015, from: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article =1776& context =extensionhist

Neumann, E. & Weber, D. (1978). Formation of protein bodies in ripening seeds of *Vicia faba*. *Biochemie und Physiologie der Pflanzen (BPP), 173*, 167-180

Neumann, G. M., Condron, R., & Polya, G. M. (1996). Purification and sequencing of yellow mustard seed napin small and large chains that are phosphorylated by plant calcium-dependant protein kinase and are calmodulin antagonists. *Plant Science, 119*, 49-66

Newkirk, R. (2015). Canola meal. Feed industry guide 5th edition. Retrieved 2016, from: http://www.canolacouncil.org/media/516716/2015_canola_meal_feed_industry_guide.pdf

Newkirk, R., Classen, H. L., & Edney, M. J. (2003). Effects of prepress solvent extraction on the nutritional value of canola meal for broiler chickens. *Animal Feed Science and Technology, 104*, 111-119

Newson, W. R. (2012). Protein based plastics from the residuals of industrial oil crops. Introductory paper at the Faculty of Landscape Planning, Horticulture and Agricultural Science, 3, *Swedish University of Agricultural Sciences*, Alnarp, Sweden

Nguyen, H. T., Silva, J. E., Podicheti, R., Macrander, J., Yang, W., Nazarenus, T. J., . . . Cahoon, E. B. (2013). Camelina seed transcriptome: a tool for meal and oil improvement and translational research. *Plant Biotechnology Journal, 11*, 759-769

Nietzel, T.; Dudkina, N. V.; Haase, C.; Denolf, P.;, Semchonok, D. A.; Boekema, E. J.; . . . Sunderhaus, S. (2013). The native structure and composition of the cruciferin complex in *Brassica napus*. *Journal of Biochemistry*, 288, 2238-2245

Oilseed Crops. (2016). Crambe. Retrieved 2016, from: http://www.oilseedcrops.org/crambe/

Oomah, B. D., Blanchard, C., & Balasubramanian, P. (2008). Phytic acid, phytase, minerals, and antioxidant activity in canadian dry bean (*Phaseolus vulgaris* L.) cultivars. *Journal of Agricultural and Food Chemistry, 56*, 11312-11319

Oomah, B. D., Corbe, A., & Balasubramanian, P. (2010). Antioxidant and anti-inflammatory activities of bean (*Phaseolus vulgaris* L.) hulls. *Journal of Agricultural and Food Chemistry, 58*, 8225-8230

Osborne, T. B. (1924). *The vegetable proteins*: London: Longmans green and Co.

Østergaard, J., Højrup, P., & Knudsen, J. (1995). Amino acid sequences of three acyl-binding/lipid-transfer proteins from rape seedlings. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1254*, 169-179

Pace, C. N., Trevino, S., Prabhakaran, E., & Scholtz, J. M. (2004). Protein structure, stability and solubility in water and other solvents. *Philosophical Transaction of the Royal Society of London B Biological Science, 359*, 1225-1234

Pachagounder, P., Lamb, R. J., & Bodnaryk, R. P. (1998). Resistance to the flea beetle *Phyllotreta cruciferae* (coleoptera: chrysomelidae) in false flax, *Camelina sativa* (*Brassicaceae*). *The Canadian Entomologist, 130*, 235-240

Pantoja-Uceda, D., Bruix, M., Gimenez-Gallego, G., Rico, M., & Santoro, J. (2003). Solution structure of RicC3, a 2S albumin storage protein from *Ricinus communis*. *Biochemistry, 42*, 13839-13847

Parker, C. E., Warren, M. R., Loiselle, D. R., Dicheva, N. N., Scarlett, C. O., & Borchers, C. H. (2005). Identification of components of protein complexes. *Methods in Molecular Biology, 301*, 117-151

Pecchia, P., Russo, R., Brambilla, I., Reggiani, R., & Mapelli, S. (2014). Biochemical seed traits of *Camelina sativa* - an emerging oilseed crop for biofuel: Environmental and genetic influences. *Journal of Crop Improvement, 28*, 465-483

Pei-Luen, J., Guang-Yuh, J., Co-Shing, W., & Jason, T. C. T. (2008). A unique caleosin in oil bodies of lily pollen. *Plant & Cell Physiology, 49*, 1390-1390

Peiretti, P. G., & Meineri, G. (2007). Fatty acids, chemical composition and organic matter digestibility of seeds and vegetative parts of false flax (*Camelina sativa* L.) after different lengths of growth. *Animal Feed Science and Technology, 133*, 341-350

Pekel, A. Y., Patterson, P. H., Hulet, R. M., Acar, N., Cravener, T. L., Dowler, D. B., & Hunter, J. M. (2009). Dietary camelina meal versus flaxseed with and without supplemental copper for broiler chickens: live performance and processing yield. *Poultry Science, 88*, 2392-2398

Peng, C.-C., & Tzen, J. T. C. (1998). Analysis of the three essential constituents of oil bodies in developing sesame seeds. *Plant and Cell Physiology, 39*(1), 35-42

Peng, Q., Khan, N. A., Wang, Z., & Yu, P. (2014). Moist and dry heating-induced changes in protein molecular structure, protein subfractions, and nutrient profiles in camelina seeds. *Journal of Dairy Science, 97*(1), 446-457

Peterson, C. J., Cossé, A., & Coats, J. R. (2000). Insecticidal components in the meal of *Crambe abyssinica*. *Journal of Agricultural and Urban Entomology, 17*(1), 27-36

Petersson, A., Thomsen, M. H., Hauggaard-Nielsen, H., & Thomsen, A.-B. (2007). Potential bioethanol and biogas production using lignocellulosic biomass from winter rye, oilseed rape and faba bean. *Biomass and Bioenergy, 31*, 812-819

Petrie, J. R., Shrestha, P., Belide, S., Kennedy, Y., Lester, G., Liu, Q., . . . Singh, S. P. (2014). Metabolic engineering *Camelina sativa* with fish oil-like levels of DHA. *PLoS ONE, 9*(1). doi: 10.1371/journal.pone.0085061

Pilgeram, A. L., Sands, D. C., Boss, D., Dale, N., Wichman, D., Lamb, P., ...Johnson, D. L., & (2007). *Camelina sativa*, a Montana omega-3 and fuel crop. Retrieved 2013, from: http://www.hort.purdue.edu/ newcrop/ncnu07/pdfs/pilgeram129-131.pdf

Plessers, A. G., McGregor, W. G., Carson, R. B., & Nakoneshny, W. (1962). Species trials with oilseed plants: II. Camelina. *Canadian Journal of Plant Science, 42*, 452-459

Plietz, P., Damaschun, G., Muller, J. J., & Schwenke, K. D. (1983). The structure of 11-S globulins from sunflower and rape seed. A small-angle X-ray scattering study. *European Journal of Biochemistry, 130*, 315-320

Putnam, D.H., Budin, J.T., Field, L.A., & Breene, W.M. (1993). Camelina: a promising low input oilseed. In J. Janick, and J.E. Simon (Eds.), *New Crops* (pp. 314-322). Wiley, New York

Ramos, M. J., Fernández, C. M., Casas, A., Rodríguez, L., & Pérez, Á. (2009). Influence of fatty acid composition of raw materials on biodiesel properties. *Bioresource Technology, 100*(1), 261-268

Raynal, M., Depigny, D., Grellet, F., & Delseny, M. (1991). Characterization and evolution of napin encoding genes in radish and related crucifers. *Gene, 99*, 77-86

Reddy, N., Jin, E., Chen, L., Jiang, X., & Yang, Y. (2012). Extraction, characterization of components, and potential thermoplastic applications of camelina meal grafted with vinyl monomers. *Journal of Agricultural and Food Chemistry, 60*, 4872-4879

Rico, M., Bruix, M., Gonzalez, C., Monsalve, R. I., & Rodriguez, R. (1996). 1H NMR assignment and global fold of napin BnIb, a representative 2S albumin seed protein. *Biochemistry, 35*, 15672-15682

Riethmuller, G. P., Carmody, P. C., & Walton, G. H. (2003). Improved canola establishment, yield and oil with large seed on sandplain soil in Western Australia. Retrieved 2016, from: http://www.australianoilseeds.com/__data/assets/pdf_file/0014/4550/Improved_canola_establishment,_yield_and_oil_with_large_seed_in_WA.pdf

Roberts, N. J., Scott, R. W., & Tzen, J. T. C. (2008). Recent biotechnological applications using oleosins. *The Open Biotechnology Journal, 2*, 13-21

Rödin, J. (1990). Studies on the structure and expression of cruciferin, the 12s storage globulin from *Brassica napus* (Oilseed rape). (Doctoral thesis), Swedish University of Agricultural Sciences, Uppsala, Sweden

Romano, S. D., & Sorichetti, P. A. (2011). *Dielectric spectroscopy in biodiesel production and characterization*. Green energy and technology (Vol. Viii), Springer-Verlag London Limited. doi: 10.1007/978-1-84996-519-4_1

Russo, R. (2012). Biochemical characterization of flour from seeds of *camelina sativa* l. (crantz) after chemical extraction of oil (Doctoral thesis). University of Milan, Milan, Italy. Retrieved 2015, from: https://air.unimi. it/bitstream/2434/217472/2/ phd_unimi_R08840. pdf

Russo, R., & Reggiani, R. (2012). Antinutritive compounds in twelve Camelina sativa genotypes. *American Journal of Plant Sciences, 3*, 1408-1412

Salisbury, P.A. (1987). Blackleg resistance in weedy crucifers. *Eucarpia Cruciferae Newsletter, 12*, 90

Salleh, M. R. M., Maruyama, N., Adachi, M., Hontani, N., Saka, S., Kato, N., . . . Utsumi, S. (2002). Comparison of protein chemical and physicochemical properties of rapeseed cruciferin with those of soybean glycinin. *Journal of Agriculture and Food Chemistry, 50*, 7380-7385

Saskatchewan Ministry of Agriculture (SMA). (n.d). Camelina. Retrieved 2015, from: http://publications.gov.sk.ca/documents/20/84138-67a5b5a3-b4fc-402b-9ede-abcebb2b64b8.pdf

Schmidt, I., Renard, D., Rondeau, D., Richomme, P., Popineau, Y., & Axelos, M. A. (2004). Detailed physicochemical characterization of the 2S storage protein from rape (*Brassica napus* L.). *Journal of Agricultur and Food Chemistry, 52*, 5995-6001

Schuster, A., & Friedt, W. (1998). Glucosinolate content and composition as parameters of quality of Camelina seed. *Industrial Crops and Products, 7*, 297-302

Schwenke, K. D. (1990). Structural studies on native and chemically modified storage proteins from rapeseed (*Brassica napus* L.) and related plant proteins. *Nahrung, 34*, 225-240

Schwenke, K. D., & Linow, K. J. (1982). A reversible dissociation of the 12 S globulin from rapeseed (*Brassica napus* L.) depending on ionic strength. *Food / Nahrung*, 26(1), K5-K6.

Schwenke, K. D., Schultz, M., Linow, K. J., Gast, K., & Zirwer, D. (1980). Hydrodynamic and quasi-elastic light scattering studies on the 12S globulin from rapeseed. *International Journal of Peptide and Protein Reaserch, 16*(1), 12-18

Scofield, R. and Crouch, M. J. (1987). Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*. *Journal of Biological Chemistry, 262*, 12202-12208

Sederoff, H. (2012). Feedstock development for jet fuel (or cars). Retrieved 2013, from: http://www.asertti.org/events/webinars/emerging-technologies/2012-11-19-Sederoff.pdf

Séguin-Swartz, G., Nettleton, J. A., Sauder, C., Warwick, S. I., & Gugel, R. K. (2013). Hybridization between *Camelina sativa* (L.) Crantz (false flax) and North American camelina species. *Plant Breeding, 132*, 390-396

Senko, M. W., Remes, P. M., Canterbury, J. D., Mathur, R., Song, Q., Eliuk, S. M., Mullen, C., Earley, L., Hardman, M., Blethrow, J. D., Bui, Huy., Specht, A., Lange, O., Denisov, E., Makarov, A., Horning, S., and Zabrouskov, V. (2003). Novel parallelized quadrupole/linear ion trap/orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Analytical Chemistry, 85*, 11710-11714

Sharma, V., Garg, G., & Alam, A. (2014). Extraction and characterization of industrially valuable oil from *Eruca sativa* (L.) Mill. through FT-IR and GC-MS Analysis. *American Journal of Biological Chemistry*, *2*, 23-28

Shewry, P. R. & Casey, R. (1999a). The 2S Albumin storage Proteins. In P.R. Shewry and R. Casey (Eds.), *Seed Proteins* (pp. 563–586), Kluwer Academic, Amsterdam, Netherland.

Shewry, P. R. & Casey, R. (1999b). Seed proteins. In P.R. Shewry and R. Casey (Eds.), *Seed Proteins* (pp. 1-10), Kluwer Academic, Amsterdam, Netherland

Shewry, P. R. (1995). Plant Storage Proteins. *Biological Reviews, 70*, 375-426

Shewry, P. R. (1998). Manipulation of seed storage proteins. In K. Lindsey (Ed.), *Transgenic Plant Research* (pp. 135–149). Harwood Academic, Amsterdam, Netherland.

Shewry, P. R. (2000). Seed proteins. In M. Black & J. Bewley (Eds.), *Seed technology and its biological basis* (pp. 42-84): CRC press

Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *The Plant Cell, 7*, 945-956

Shimomura, Y., Yamamoto, Y., Bajotto, G., Sato, J., Murakami, T., Shimomura, N., . . . Mawatari, K. (2006). Nutraceutical effects of branched-chain amino acids on skeletal muscle. *Journal of Nutrition, 136*, 529s-532s

Shivu, B., Seshadri, S., Li, J., Oberg, K. A., Uversky, V. N., & Fink, A. L. (2013). Distinct beta-sheet structure in protein aggregates determined by ATR-FTIR spectroscopy. *Biochemistry, 52*, 5176-5183

Simon, A. E., Tenbarge, K. M., Scofield, S. R., Finkelstein, R. R., & Crouch, M. L. (1985). Nucleotide sequence of a cDNA clone of *Brassica napus* 12S storage protein shows homology with legumin from *Pisum sativum*. *Plant Molecular Biology, 5*, 191-201

Simpson, D. M., & Beynon, R. J. (2010). Acetone precipitation of proteins and the modification of peptides. *Journal of Proteome Research, 9*(1), 444-450

Sjödahl, S., Rödin, J., & Rask, L. (1991). Characterization of the 12S globulin complex of *Brassica napus*. Evolutionary relationship to other 11-12S storage globulins. *European Journal Of Biochemistry / FEBS, 196*(3), 617-621

Skriver, K., & Mundy, J. (1990). Gene expression in response to abscisic acid and osmotic stress. *Plant Cell, 2*, 503

Slavik, J. (1982). Anilinonaphthalene sulfonate as a probe of membrane composition and function. *Biochimica Biophysica Acta*, 694(1), 1-25

Statistics Canada. (2015). Production of principle field crops, November 2015. Retrieved 2016, from: http://www.statcan.gc.ca/daily-quotidien/151204/dq151204c-eng.htm

Stryer, L. (1968). Fluorescence spectroscopy of proteins. *Science,* 162, 526-533

Sun, Y., Tan, D. Y., Baskin, C. C., & Baskin, J. M. (2012). Role of mucilage in seed dispersal and germination of the annual ephemeral *Alyssum minus* (*Brassicaceae*). *Australian Journal of Botany, 60*, 439-449

Tamura, K., Stecher G., Peterson D., Filipski A., & Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30, 2725-2729

Tan, S. H., Mailer, R. J., Blanchard, C. L., & Agboola, S. O. (2011a). Canola Proteins for Human Consumption: Extraction, Profile, and Functional Properties. *Journal of Food Science, 76*(1), R16-R28

Tan, S. H., Mailer, R. J., Blanchard, C. L., & Agboola, S. O. (2011b). Extraction and residual antinutritional components in protein fractions of *Sinapis alba* and *Brassica napus* oil-free meals. *17th Australian Research Assembly on Brassicas*, 107-114

Tandang, M. R., Adachi, M., & Utsumi, S. (2004). Cloning and expression of rapeseed procruciferin in *Escherichia coli* and crystallization of the purified recombinant protein. *Biotechnology Letters, 26*, 385-391

Tandang-Silvas, M. R. G., Fukuda, T., Fukuda, C., Prak, K., Cabanos, C., Kimura, A., . . . Maruyama, N. (2010). Conservation and divergence on plant seed 11S globulins based on crystal structures. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1804*(7), 1432-1442

Terpinc, P., Polak, T., Makuc, D., Ulrih, N. P., & Abramovič, H. (2012). The occurrence and characterization of phenolic compounds in *Camelina sativa* seed, cake and oil. *Food Chemistry, 131,* 580-589

Tzen, J. T. C. (2012). Integral proteins in plant oil bodies. *ISRN Botany, 2012*

Tzen, J. T. C., Cao, Y., Laurent, P., Ratnayake, C., & Huang, A. H. C. (1993). Lipids, proteins, and structure of seed oil bodies from diverse species. *Plant Physiology, 101*(1), 267-276

Tzen, J. T. C., George, C. L., & Huang, A. H. C. (1992). Characterization of the charged components and their topology on the surface of plant seed oil bodies. *The journal of biological chemistry, 267*, 15626-15634

Tzen, J. T., Peng, C. C., Cheng, D. J., Chen, E. C., & Chiu, J. M. (1997). A new method for seed oil body purification and examination of oil body integrity following germination. *Journal of Biochemistry, 121*, 762-768

Vivian, J. T., & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys Journal, 80*, 2093-2109

Voiniciuc, C., Schmidt, M. H. W., Berger, A., Yang, B., Ebert, B., Scheller, H. V., . . . Guenl, M. (2015). MUCI10 Produces galactoglucomannan that maintains pectin and cellulose architecture in Arabidopsis seed mucilage. *Plant Physiology*. doi: 10.1104/pp.15.00851

Wall, M. L., Wheeler, H. L., Huebsch, M. P., Smith, J. C., Figeys, D., & Altosaar, I. (2010). The tryptophan-rich domain of puroindoline is directly associated with the starch granule surface as judged by tryptic shaving and mass spectrometry. *Journal of Cereal Science, 52*, 115-120

Walters, J., Milam, S. L., & Clark, A. C. (2009). Practical approaches to protein folding and assembly: spectroscopic strategies in thermodynamics and kinetics. *Methods in enzymology, 455*, 1-39

Wan, L., Ross, A. R., Yang, J., Hegedus, D. D., & Kermode, A. R. (2007). Phosphorylation of the 12 S globulin cruciferin in wild-type and abi1-1 mutant *Arabidopsis thaliana* (thale cress) seeds. *Biochemical Journal, 404*, 247-256

Wanasundara, J. D., Abeysekara, S., McIntosh, T., & Falk, K. (2012). Solubility differences of major storage proteins of *Brassicaceae* oilseeds. *Journal of the American Oil Chemists' Society*, 89(5), 869-881

Wanasundara, J. P. D. & McIntosh, T.C. (2013). Process of aqueous protein extraction from *Brassicaceae* oilseeds. Google Patents. Retrieved 2016, from: http://www.google.com/patents/US8557963

Wanasundara, J. P. D. (2011). Proteins of *Brassicaceae* oilseeds and their potential as a plant protein source. *Critical Reviews in Food Science and Nutrition, 51*, 635-677

Wanasundara, J. P. D., Siong, T., Alashi, A., Pudel, F., and Blanchard, C. (2015). Canola/Rapeseed proteins - Current status. In S. N. Nadathur, J. P. D. Wanasundara & L. Scanlin (Eds.), *Sustainable Protein Sources*, Elsevier Science LTD

Wanasundara, P. K. J. P. D., & Shahidi, F. (1997). Removal of flaxseed mucilage by chemical and enzymatic treatments. *Food Chemistry, 59*(1), 47-55

Wang, Y. P., Tang, J. S., Chu, C. Q., & Tian, J. (2000). A preliminary study on the introduction and cultivation of *Crambe abyssinica* in China, an oil plant for industrial uses. *Industrial Crops and Products*, *12*(1), 47-52

Warwick, S. (2011). *Brassicaceae* in agriculture. In R. Schmidt & I. Bancroft (Eds.), *Genetics and Genomics of the Brassicaceae* (Vol. 9, pp. 33-65): Springer New York. doi. 10.1007/978-1-4419-7118-0_2

Warwick, S. I., & Gugel, R. (2003). Genetic variation in the *Crambe abyssinica-C. hispanica-C. glabrata* complex. *Genetic Resources and Crop Evolution*, 50, 291-305

Warwick, S. I., Francis, A., & Mulligan, G. A. (2013). *Brassicaceae* of Canada. Retrieve 2016 from: http://www.cbif.gc.ca/eng/species-bank/brassicaceae-of-canada/?id=137040326726

Weber, E., & Neumann, D. (1980). Protein bodies, storage organelles in plant seeds. *Biochemie und Physiologie der Pflanzen, 175*, 279-306

Werner, S.; Winkler, U & Stabenau, H. (n.d). Separation of native basic protein by cathodic, discontinuous polyacrylamide gel electrophoresis. *BioRad techical note 2376*. Retrieved 2016, from: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_2376.pdf

Western, T. L. (2012). The sticky tale of seed coat mucilages: production, genetics, and role in seed germination and dispersal. *Seed Science Research, 22*(1), 1-25

Western, T. L., Skinner, D. J., & Haughn, G. W. (2000). Differentiation of mucilage secretory cells of the Arabidopsis seed coat. *Plant Physiology, 122*, 345-356

Wijesundera, C., Boiteau, T., Xu, X., Shen, Z., Watkins, P., & Logan, A. (2013). Stabilization of fish oil-in-water emulsions with oleosin extracted from canola meal. *Journal of Food Science, 78*, C1340-1347. doi: 10.1111/1750-3841.12177

Withana-Gamage, T. (2013). *Structure and properties of cruciferin: investigation of homohexameric cruciferin expressed in Arabidopsis* (Doctoral thesis). University of Saskatchewan, Saskatoon, Canada. Retrieved from: http://hdl.handle.net/10388/ETD-2013-06-1090

Wu, X., & Leung, D. Y. C. (2011). Optimization of biodiesel production from camelina oil using orthogonal experiment. *Applied Energy, 88*, 3615-3624

Yong-Gang, L., Steg, A., Smits, B., & Tamminga, S. (1994). Crambe meal: removal of glucosinolates by heating with additives and water extraction. *Animal Feed Science and Technology, 48*, 273-287

Young, D. R. (1994). Use of an organic solvent for precipitation: Introduction to biochemical engineering term project. Retrieved 2016, from: http://www.rpi.edu/dept/chem-eng/Biotech-Environ/PRECIP/precporg.html

Yu, P. (2005). Protein secondary structures (α-helix and β-sheet) at a cellular level and protein fractions in relation to rumen degradation behaviours of protein: a new approach. *British journal of nutrition, 94*, 655-665

Yu, P., Christensen, D., Christensen, C., Drew, M., Rossnagel, B., & McKinnon, J. (2004). Use of synchrotron FTIR microspectroscopy to identify chemical differences in barley endosperm tissue in relation to rumen degradation characteristics. *Canadian Journal of Animal Science, 84*, 523-527

Yu, P., McKinnon, J. J., Christensen, C. R., & Christensen, D. A. (2004). Using synchrotron-based FTIR microspectroscopy to reveal chemical features of feather protein secondary structure: comparison with other feed protein sources. *Journal of Agriculture and Food Chemistry, 52*, 7353-7361

Zhang, Y., Yu, L., Yung, K. F., Leung, D. Y., Sun, F., & Lim, B. L. (2012). Over-expression of AtPAP2 in *Camelina sativa* leads to faster plant growth and higher seed yield. *Biotechnology for Biofuels*, *5*. doi: 10.1186/1754-6834-5-19

Ziolkovska, A. (2012). Laws of flaxseed mucilage extraction. *Food Hydrocolloids, 26*(1), 197-204

Zubr, J. (2003a). Dietary fatty acids and amino acids of *Camelina sativa* seed. *Journal of Food Quality, 26*, 451-462

Zubr, J. (2003b). Qualitative variation of Camelina sativa seed from different locations. *Industrial Crops and Products, 17*, 161-169

Zubr, J. (2010). Carbohydrates, vitamins and minerals of *Camelina sativa* seed. *Nutrition and Food Science, 40*, 523-531

Zubr, J., & Matthaus, B. (2002). Effects of growth conditions on fatty acids and tocopherols in *Camelina sativa* oil. *Industrial Crops and Products, 15*, 155-162

`

8. APPENDIX

Table A1. LC-MS/MS results showing 100% probable protein matches for the *C. sativa* cruciferin separated by 2D-electrophoresis using pH 3 to 10 IPG strips.

Spot Number *	Gene name	Annotation**	Molecular Weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS***
1	Csa14g004970.1	CRD-1-G2	50.1	3	5	277	60	1010
1	Csa03g005060.1	CRD-1-G3	50.0	2	4	258	59.6	941
1	Csa18g009670.1	CRA-1-G2	51.9	3	5	83	49.5	303
1	Csa17g006950.1	CRB-1-G1	50.7	2	2	29	36.5	106
1	Csa12g021990.1	CRC-1-G3	55.4	1	1	16	22.8	58
2	Csa14g004970.1	CRD-1-G2	50.1	4	7	211	67.1	999
2	Csa03g005060.1	CRD-1-G3	50.0	2	3	179	59.6	816
2	Csa18g009670.1	CRA-1-G2	51.9	2	2	48	43.7	219
2	Csa07g016060.1	Vic2-1-G2	53.2	1	1	28	20.3	128
2	Csa05g038120.1	Vic2-1-G3	58.8	0	0	25	15.5	114
2	Csa17g006950.1	CRB-1-G1	50.7	0	0	21	28.6	96
3	Csa18g009670.1	CRA-1-G2	51.9	5	15	1043	71	927
3	Csa11g070580.1	CRA-1-G1	52.0	1	1	895	70.5	796
3	Csa17g006950.1	CRB-1-G1	50.7	3	21	407	80.3	362
3	Csa14g004960.1	CRB-1-G2	64.9	2	4	390	58.7	347
3	Csa12g021990.1	CRC-1-G3	55.4	8	13	129	63.2	115
3	Csa10g014100.1	CRC-1-G2	54.0	3	4	91	64.8	81
3	Csa11g015240.1	CRC-1-G1	72.9	4	4	87	44.2	77
3	Csa14g004970.1	CRD-1-G2	50.1	2	3	25	38.2	22
4	Csa18g009670.1	CRA-1-G2	51.9	4	14	2092	80.9	838
4	Csa11g070590.1	CRA-2-G1	52.0	1	2	1779	80.3	713
4	Csa11g070580.1	CRA-1-G1	52.0	1	2	1762	80.3	706
4	Csa14g004960.1	CRB-1-G2	64.9	2	9	823	59.2	330
4	Csa17g006950.1	CRB-1-G1	50.7	8	28	817	81.4	327
4	Csa12g021990.1	CRC-1-G3	55.4	9	23	374	72	150
4	Csa10g014100.1	CRC-1-G2	54.0	5	9	323	72.9	129
4	Csa11g015240.1	CRC-1-G1	72.9	4	9	301	44.2	121
4	Csa14g004970.1	CRD-1-G2	50.1	1	2	68	43.7	27
4	Csa05g038120.1	Vic2-1-G3	58.8	2	3	35	22.1	15
4	Csa19g031870.2	Vic1-1-G2	58.4	2	2	27	21.9	11
4	Csa01g025880.1	Vic1-3-G3	56.0	1	1	19	18.1	8
4	Csa19g023890.1	Lipoamide dehydrogenase 1	61.9	1	1	13	13.6	5
4	Csa15g023000.1	Lipoamide dehydrogenase 1	63.0	1	1	11	13.1	4
4	Csa15g039300.1	Vic1-2-G1	52.4	1	1	6	10.8	2
5	Csa18g009670.1	CRA-1-G2	51.9	5	15	2369	80.2	848
5	Csa11g070590.1	CRA-2-G1	52.0	1	4	2074	79.7	742
5	Csa11g070580.1	CRA-1-G1	52.0	1	3	2041	79.7	731
5	Csa14g004960.1	CRB-1-G2	64.9	4	11	750	72.2	269
5	Csa03g005050.1	CRB-1-G3	49.4	1	3	735	73	263
5	Csa17g006950.1	CRB-1-G1	50.7	8	18	667	74.2	239
5	Csa12g021990.1	CRC-1-G3	55.4	9	24	449	72.2	161
5	Csa10g014100.1	CRC-1-G2	54.0	5	10	363	73.1	130
5	Csa11g015240.1	CRC-1-G1	72.9	4	10	338	44.3	121
5	Csa14g004970.1	CRD-1-G2	50.1	3	4	85	41.7	30
5	Csa09g068650.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	6	67	71.2	24
5	Csa04g042760.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	5	54	71.2	19
5	Csa06g031070.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	3	50	62.9	18
5	Csa19g031870.2	Vic1-1-G2	58.4	4	5	35	37.1	13
5	Csa01g025880.1	Vic1-3-G3	56.0	1	1	28	22.2	10
5	Csa05g038120.1	Vic2-1-G3	58.8	2	3	24	18.1	9
5	Csa03g019850.1	Unknown protein	58.0	2	2	8	8.61	7
5	Csa15g023000.1	Lipoamide dehydrogenase 1	63.0	1	1	10	12.2	4
5	Csa19g023890.1	Lipoamide dehydrogenase 1	61.9	1	1	11	12.7	4
5	Csa15g039300.1	Vic1-2-G1	52.4	1	1	11	10.8	4
5	Csa03g005710.1	Lactate/ malate dehydrogenase family protein	40.4	1	1	7	13.1	3
6	Csa18g009670.1	CRA-1-G2	51.9	5	16	2330	81.3	834
6	Csa11g070590.1	CRA-2-G1	52.0	1	3	2078	89.7	744

`

6	Csa11g070580.1	CRA-1-G1	52.0	1	2	2067	89.7	740
6	Csa14g004960.1	CRB-1-G2	64.9	2	4	679	59.2	243
6	Csa17g006950.1	CRB-1-G1	50.7	6	13	597	65.5	214
6	Csa12g021990.1	CRC-1-G3	55.4	9	24	532	72.2	190
6	Csa10g014100.1	CRC-1-G2	54.0	5	9	447	73.7	160
6	Csa11g015240.1	CRC-1-G1	72.9	5	12	417	47.1	149
6	Csa04g042760.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	5	66	71.2	24
6	Csa09g068650.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	3	64	71.2	23
6	Csa06g031070.1	Iron/ Manganese superoxide dismutase family protein	26.7	1	1	57	48.8	20
6	Csa19g031870.2	Vic1-1-G2	58.4	4	6	43	38.6	15
6	Csa14g004970.1	CRD-1-G2	50.1	2	3	39	30.9	14
6	Csa01g025880.1	Vic1-3-G3	56.0	2	2	36	27.3	13
6	Csa05g038120.1	Vic2-1-G3	58.8	2	3	29	20.2	10
6	Csa03g019850.1	Unknown protein	58.0	1	1	7	6.37	6
6	Csa15g039300.1	Vic1-2-G1	52.4	1	1	7	10.8	3
7	Csa18g009670.1	CRA-1-G2	51.9	1	2	45	40.9	1316
8	Csa18g009670.1	CRA-1-G2	51.9	7	23	2524	93.3	784
8	Csa11g070580.1	CRA-1-G1	52.0	1	4	2454	82.9	762
8	Csa11g070590.1	CRA-2-G1	52.0	1	3	2448	82.9	760
8	Csa14g004960.1	CRB-1-G2	64.9	2	7	745	53.9	231
8	Csa03g005050.1	CRB-1-G3	49.4	1	3	732	65.6	227
8	Csa12g021990.1	CRC-1-G3	55.4	12	29	653	81.5	203
8	Csa17g006950.1	CRB-1-G1	50.7	4	7	573	55.7	178
8	Csa10g014100.1	CRC-1-G2	54.0	6	14	571	73.5	177
8	Csa11g015240.1	CRC-1-G1	72.9	5	13	527	46.6	164
8	Csa19g031870.2	Vic1-1-G2	58.4	6	7	59	41.6	18
8	Csa09g068650.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	3	55	71.2	17
8	Csa04g042760.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	4	50	71.2	16
8	Csa06g031070.1	Iron/ Manganese superoxide dismutase family protein	26.7	1	1	48	48.8	15
8	Csa01g025880.1	Vic1-3-G3	56.0	3	3	45	35	14
8	Csa14g004970.1	CRD-1-G2	50.1	2	3	44	34.4	14
8	Csa05g038120.1	Vic2-1-G3	58.8	2	3	22	18.1	7
8	Csa03g019850.1	Unknown protein	58.0	2	2	7	5.62	7
8	Csa11g017000.1	Cs2S-4-G1	36.6	2	4	13	15.5	4
8	Csa15g039300.1	Vic1-2-G1	52.4	4	4	10	16.9	3
8	Csa12g006190.1	Serine carboxypeptidase 528 family	52.4	2	2	4	8.58	2
8	Csa02g074880.1	Heat shock protein 70 (HSP 70) family	71.3	1	1	7	7.06	2
8	Csa02g039290.1	Hydrooxysteroid dehydrogenase	39.2	2	2	3	4.57	2
9	Csa18g009670.1	CRA-1-G2	51.9	4	14	1377	79.6	1030
9	Csa11g070590.1	CRA-2-G1	52.0	1	2	1266	79.1	947
9	Csa11g070580.1	CRA-1-G1	52.0	1	2	1255	79.1	938
9	Csa14g004960.1	CRB-1-G2	64.9	2	4	241	52	180
9	Csa12g021990.1	CRC-1-G3	55.4	9	14	226	72	169
9	Csa17g006950.1	CRB-1-G1	50.7	2	3	204	44.8	153
9	Csa10g014100.1	CRC-1-G2	54.0	5	7	193	72.9	144
9	Csa11g015240.1	CRC-1-G1	72.9	4	7	164	44.2	123
9	Csa19g031870.2	Vic1-1-G2	58.4	2	2	11	19.6	8
9	Csa01g025880.1	Vic1-3-G3	56.0	1	1	11	19.1	8
10	Csa18g009670.1	CRA-1-G2	51.9	4	14	1091	75.3	1069
10	Csa11g070590.1	CRA-2-G1	52.0	1	1	981	74.8	961
10	Csa11g070580.1	CRA-1-G1	52.0	1	1	977	74.8	957
10	Csa12g021990.1	CRC-1-G3	55.4	9	13	174	72	171
10	Csa14g004960.1	CRB-1-G2	64.9	1	1	158	29	155
10	Csa10g014100.1	CRC-1-G2	54.0	4	6	143	72.9	140
10	Csa17g006950.1	CRB-1-G1	50.7	1	1	142	35.2	139
10	Csa11g015240.1	CRC-1-G1	72.9	4	4	121	44.2	119
10	Csa19g031870.2	Vic1-1-G2	58.4	2	2	9	17.1	9
11	Csa18g009670.1	CRA-1-G2	51.9	4	14	1103	76.8	1128
11	Csa18g009670.1	CRA-1-G2	51.9	4	14	1103	76.8	1128
11	Csa11g070590.1	CRA-2-G1	52.0	1	1	992	76.3	1015
11	Csa11g070580.1	CRA-1-G1	52.0	1	1	982	76.3	1004
11	Csa17g006950.1	CRB-1-G1	50.7	0	0	128	28.6	131
11	Csa10g014100.1	CRC-1-G2	54.0	3	5	121	62.1	124
11	Csa12g021990.1	CRC-1-G3	55.4	6	9	118	60	121
12	Csa18g009670.1	CRA-1-G2	51.9	3	9	440	68.2	1079
12	Csa10g014100.1	CRC-1-G2	54.0	3	4	65	53	159
12	Csa12g021990.1	CRC-1-G3	55.4	4	4	63	50.4	154
12	Csa11g015240.1	CRC-1-G1	72.9	1	1	47	31.5	115
13	Csa18g009670.1	CRA-1-G2	51.9	4	14	1633	80.9	807
13	Csa11g070590.1	CRA-2-G1	52.0	1	2	1470	80.3	726

`

13	Csa11g070580.1	CRA-1-G1	52.0	1	3	1470	80.3	720
13	Csa14g004960.1	CRB-1-G2	64.9	2	4	452	50.7	223
13	Csa17g006950.1	CRB-1-G1	50.7	6	16	443	61.8	219
13	Csa12g021990.1	CRC-1-G3	55.4	9	20	286	72	141
13	Csa10g014100.1	CRC-1-G2	54.0	5	9	246	72.9	122
13	Csa14g004970.1	CRD-1-G2	50.1	4	6	206	66.9	102
13	Csa11g015240.1	CRC-1-G1	72.9	4	8	206	44.2	102
13	Csa03g005060.1	CRD-1-G3	50.0	2	5	188	59.6	93
13	Csa05g038120.1	Vic2-1-G3	58.8	3	4	66	25	33
13	Csa09g068650.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	4	58	71.2	29
13	Csa04g042760.1	Iron/ Manganese superoxide dismutase family protein	26.7	2	5	48	71.2	24
13	Csa17g028510.1	PDI-like-1-1 (Protein disulfide isomarase)	115.5	13	14	44	17.6	22
13	Csa06g031070.1	Iron/ Manganese superoxide dismutase family protein	26.7	1	1	43	48.8	21
13	Csa19g031870.2	Vic1-1-G2	58.4	4	4	29	23.5	14
14	Csa18g009670.1	CRA-1-G2	51.9	3	11	488	54.6	786
14	Csa12g021990.1	CRC-1-G3	55.4	9	12	151	71.5	243
14	Csa10g014100.1	CRC-1-G2	54.0	5	7	143	72.9	230
14	Csa11g015240.1	CRC-1-G1	72.9	4	4	125	44.2	201
14	Csa14g004960.1	CRB-1-G2	64.9	2	3	98	42.5	158
14	Csa17g006950.1	CRB-1-G1	50.7	1	1	82	35.2	132
14	Csa18g009670.1	CRA-1-G2	51.9	3	11	488	54.6	786
14	Csa12g021990.1	CRC-1-G3	55.4	9	12	151	71.5	243
14	Csa10g014100.1	CRC-1-G2	54.0	5	7	143	72.9	230
15	Csa18g009670.1	CRA-1-G2	51.9	3	11	921	58.3	1194
15	Csa11g070590.1	CRA-2-G1	52.0	1	2	813	57.9	1055
15	Csa10g014100.1	CRC-1-G2	54.0	2	2	52	52.2	67
15	Csa12g021990.1	CRC-1-G3	55.4	3	3	48	50.6	62
15	Csa11g015240.1	CRC-1-G1	72.9	1	1	44	28.6	57
16	Csa18g009670.1	CRA-1-G2	51.9	2	9	585	54.8	1027
16	Csa11g070590.1	CRA-2-G1	52.0	1	2	496	57.3	871
16	Csa10g014100.1	CRC-1-G2	54.0	3	5	76	54.7	133
16	Csa12g021990.1	CRC-1-G3	55.4	3	5	74	51	130
16	Csa11g015240.1	CRC-1-G1	72.9	1	1	60	28.6	105
16	Csa19g031870.2	Vic1-1-G2	58.4	3	3	36	25.4	63
17	Csa14g004960.1	CRB-1-G2	64.9	4	15	1204	51.5	1167
17	Csa17g006950.1	CRB-1-G1	50.7	3	7	878	54.6	851
17	Csa18g009670.1	CRA-1-G2	51.9	3	6	247	48.4	239
17	Csa12g021990.1	CRC-1-G3	55.4	4	6	64	50	62
17	Csa10g014100.1	CRC-1-G2	54.0	2	3	61	45.1	59
17	Csa11g015240.1	CRC-1-G1	72.9	1	1	42	27.5	41
18	Csa17g006950.1	CRB-1-G1	50.7	5	17	1402	57	1014
18	Csa14g004960.1	CRB-1-G2	64.9	3	9	1235	56.8	893
18	Csa18g009670.1	CRA-1-G2	51.9	3	7	372	49.5	269
18	Csa12g021990.1	CRC-1-G3	55.4	5	12	105	55.5	76
18	Csa10g014100.1	CRC-1-G2	54.0	3	6	95	54.9	69
18	Csa11g015240.1	CRC-1-G1	72.9	3	3	76	38.8	55
18	Csa01g025880.1	Vic1-3-G3	56.0	3	3	47	33.4	34
18	Csa15g039290.1	Vic1-1-G1	56.0	1	1	45	34.8	33
18	Csa19g031870.2	Vic1-1-G2	58.4	3	3	37	32.4	27
18	Csa15g039300.1	Vic1-2-G1	52.4	5	5	23	21	17
19	Csa17g006950.1	CRB-1-G1	50.7	9	26	954	75.1	901
19	Csa14g004960.1	CRB-1-G2	64.9	3	6	784	55.8	740
19	Csa18g009670.1	CRA-1-G2	51.9	3	7	316	51.8	298
19	Csa12g021990.1	CRC-1-G3	55.4	7	13	130	59.3	123
19	Csa10g014100.1	CRC-1-G2	54.0	3	5	108	54.9	102
19	Csa11g015240.1	CRC-1-G1	72.9	3	3	101	38.8	95
19	Csa15g039290.1	Vic1-1-G1	56.0	1	1	35	31.8	33
19	Csa01g025880.1	Vic1-3-G3	56.0	1	1	34	24.4	32
19	Csa19g031870.2	Vic1-1-G2	58.4	1	1	31	22.7	29
19	Csa15g039300.1	Vic1-2-G1	52.4	5	6	31	24.5	29
19	Csa03g006900.1	Calcium-dependant lipid binding (CaLB domain) family protein	27.3	6	6	12	10.5	11
19	Csa02g065080.1	Actin-12	41.8	2	2	3	7.43	7
19	Csa20g077530.1	Protein of unknown function	27.8	3	3	3	15.4	3
20	Csa18g009670.1	CRA-1-G2	51.9	4	8	378	67.7	898
20	Csa17g006950.1	CRB-1-G1	50.7	4	7	124	54.8	295
20	Csa14g004960.1	CRB-1-G2	64.9	2	2	117	38.7	278
20	Csa12g021990.1	CRC-1-G3	55.4	5	8	81	53.7	193
20	Csa10g014100.1	CRC-1-G2	54.0	1	1	59	50.3	140
20	Csa11g015240.1	CRC-1-G1	72.9	1	1	50	35.8	119
20	Csa18g009420.1	Late embryogenesis abundant protein (LEA) family	34.3	3	3	4	5.59	10

`

21	Csa18g009670.1	CRA-1-G2	51.9	4	7	283	63.2	798
21	Csa14g004960.1	CRB-1-G2	64.9	2	4	115	42.5	324
21	Csa17g006950.1	CRB-1-G1	50.7	3	5	106	50	299
21	Csa12g021990.1	CRC-1-G3	55.4	6	9	81	58.7	228
21	Csa10g014100.1	CRC-1-G2	54.0	2	2	56	49.7	158
21	Csa18g009420.1	Late embryogenesis abundant protein (LEA) family	34.3	2	2	3	2.96	8
22	Csa17g006950.1	CRB-1-G1	50.7	7	22	418	66.6	1026
22	Csa14g004960.1	CRB-1-G2	64.9	2	7	339	54.4	832
22	Csa12g021990.1	CRC-1-G3	55.4	5	7	63	45.7	155
22	Csa18g009670.1	CRA-1-G2	51.9	3	3	60	41.1	147
22	Csa10g014100.1	CRC-1-G2	54.0	2	2	48	39.1	118
22	Csa11g015240.1	CRC-1-G1	72.9	2	2	43	24.3	106
22	Csa14g004970.1	CRD-1-G2	50.1	1	2	11	13	27
23	Csa17g006950.1	CRB-1-G1	50.7	9	34	1150	72.1	1379
23	Csa14g004960.1	CRB-1-G2	64.9	2	4	718	54.1	861
23	Csa12g021990.1	CRC-1-G3	55.4	4	8	71	45.3	85
23	Csa10g014100.1	CRC-1-G2	54.0	3	4	70	50.7	84
23	Csa18g009670.1	CRA-1-G2	51.9	3	6	57	41.1	68
23	Csa11g015240.1	CRC-1-G1	72.9	3	3	53	35.8	64
24	Csa17g006950.1	CRB-1-G1	50.7	8	23	397	66.6	1027
24	Csa14g004960.1	CRB-1-G2	64.9	1	2	278	49.7	719
24	Csa12g021990.1	CRC-1-G3	55.4	4	9	101	42.5	261
24	Csa10g014100.1	CRC-1-G2	54.0	3	4	62	47.4	160
24	Csa11g015240.1	CRC-1-G1	72.9	2	2	52	28.6	135
24	Csa18g009670.1	CRA-1-G2	51.9	3	5	43	35.9	111
25	Csa12g021990.1	CRC-1-G3	55.4	6	17	1307	56.1	1044
25	Csa10g014100.1	CRC-1-G2	54.0	1	2	597	46	477
25	Csa11g015240.1	CRC-1-G1	72.9	1	1	419	25.5	335
25	Csa17g006950.1	CRB-1-G1	50.7	6	7	67	53.1	54
25	Csa14g004960.1	CRB-1-G2	64.9	2	4	67	41.5	54
25	Csa18g009670.1	CRA-1-G2	51.9	1	1	45	38.1	36
26	Csa14g004960.1	CRB-1-G2	64.9	2	9	1083	51	1201
26	Csa17g006950.1	CRB-1-G1	50.7	4	5	619	52	686
26	Csa12g021990.1	CRC-1-G3	55.4	3	8	129	42.7	143
26	Csa10g014100.1	CRC-1-G2	54.0	3	6	122	42.4	135
26	Csa18g009670.1	CRA-1-G2	51.9	2	6	84	35.7	93
26	Csa11g015240.1	CRC-1-G1	72.9	1	3	82	20.8	91
26	Csa04g042760.1	Iron/ Manganese superoxide dismutase family protein	26.7	1	1	26	38.3	29
27	Csa12g021990.1	CRC-1-G3	55.4	9	29	2095	60.8	744
27	Csa10g014100.1	CRC-1-G2	54.0	8	23	1566	61.5	578
27	Csa11g015240.1	CRC-1-G1	72.9	4	11	1267	35.8	468
27	Csa14g004960.1	CRB-1-G2	64.9	2	4	267	52.7	99
27	Csa17g006950.1	CRB-1-G1	50.7	7	16	259	65.9	96
27	Csa18g009670.1	CRA-1-G2	51.9	3	7	103	44.5	38
27	Csa19g031870.2	Vic1-1-G2	58.4	3	3	48	20.4	18
27	Csa15g039300.1	Vic1-2-G1	52.4	5	5	42	10.2	16
27	Csa11g017000.1	Cs2S-4-G1	36.6	2	5	23	17.7	8
28	Csa17g006950.1	CRB-1-G1	50.7	8	23	338	71.4	990
28	Csa14g004960.1	CRB-1-G2	64.9	2	4	245	58.2	718
28	Csa12g021990.1	CRC-1-G3	55.4	4	8	75	55.5	220
28	Csa10g014100.1	CRC-1-G2	54.0	1	2	53	46	155
28	Csa11g015240.1	CRC-1-G1	72.9	1	1	42	25.5	123
28	Csa18g009670.1	CRA-1-G2	51.9	1	1	39	25.2	114
28	Csa14g004970.1	CRD-1-G2	50.1	1	2	26	19.6	76
28	Csa12g037540.1	Late embryogenesis abundant protein (LEA) family	35.1	2	2	3	7.91	9
29	Csa14g004960.1	CRB-1-G2	64.9	2	9	480	60.1	933
29	Csa17g006950.1	CRB-1-G1	50.7	6	12	364	69	708
29	Csa12g021990.1	CRC-1-G3	55.4	5	12	141	59.6	274
29	Csa10g014100.1	CRC-1-G2	54.0	3	5	95	50.5	185
29	Csa18g009670.1	CRA-1-G2	51.9	3	7	86	47.1	167
29	Csa11g015240.1	CRC-1-G1	72.9	1	1	63	25.5	123

* Spot number refers to the protein spots as represented in Figure 4.11B.
** Annotations for *C. sativa* storage protein genes were assigned according to Table 4.3 and the rest is according to the lowest probability obtained for each gene from the BLAST search preformed against the *Arabidopsis thaliana* genome using TAIR 8 database available from (https://www.arabidopsis.org/).
*** NTS – Normalized Total Spectra, is the parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A2. LC-MS/MS results showing 100% probable protein matches for the *B.napus* cruciferin separated by 2D-electrophoresis using pH 3 to 10 IPG strips.

Spot Number *	Protein name	Accession	Molecular Weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS**
1	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	11	17	325	50.6	389
1	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	10	218	53.6	261
1	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	16	28	195	57	233
1	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	3	3	168	22.2	210
2	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	13	25	692	59.8	374
2	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	17	34	452	57	244
2	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	7	15	364	54.4	197
2	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	6	7	345	32.5	186
2	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	3	4	18	20	10
3	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	14	29	917	60.2	412
3	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	16	34	527	57	237
3	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	6	6	460	32.5	207
3	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	8	16	418	56	188
3	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	3	4	14	20	6
3	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	4	4	10	15.7	4
4	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	15	29	526	61.6	450
4	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	6	6	257	32.9	220
4	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	17	29	245	58.1	210
4	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	7	12	238	55	204
4	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	1	2	11	22.2	9
5	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	14	29	808	60.2	417
5	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	18	39	513	58.1	265
5	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	7	8	400	36.9	206
5	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	12	376	54	194
5	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	3	4	13	20	7
5	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	3	3	5	11.8	3
6	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	13	26	640	53.1	359
6	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	18	34	417	58.1	234
6	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	7	15	372	54.4	209
6	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	7	12	338	36.7	190
6	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	3	4	19	20	11
6	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	4	4	7	15.7	4
7	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	14	23	391	59.2	423
7	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	16	26	222	57	240

`

7	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	6	6	195	35.9	211
7	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	8	192	48.5	208
8	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	12	19	239	47.8	309
8	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	9	198	48.3	256
8	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	17	28	195	57	252
8	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	7	7	128	44.4	165
9	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	10	21	1024	62.3	929
9	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	15	25	167	53.1	151
9	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	9	11	94	44.5	85
9	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	6	10	94	36.1	85
9	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	3	3	13	27.8	12
10	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	6	80	46	354
10	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	12	15	67	44.1	296
10	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	6	80	46	354
11	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	8	10	67	35.5	258
11	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	6	64	42	246
11	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	12	13	58	44.1	223
12	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	7	9	69	31.2	296
12	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	11	17	67	37	288
12	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	5	6	60	40.3	258
13	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	14	20	93	49.7	376
13	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	8	11	72	42.2	291
13	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	5	5	54	42.4	218
13	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	2	8	19.4	32
14	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	9	12	189	38.2	449
14	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	2	3	121	23.2	287
14	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	15	22	114	52.5	271
14	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	5	5	54	36. 5	128
15	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	7	14	279	54.8	485
15	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	9	14	124	46.1	216
15	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	7	14	279	54.8	485
15	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	9	14	124	46.1	216
15	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	15	23	121	56.8	210
16	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	10	18	254	53	395
16	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	11	15	124	41.6	193
16	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	12	16	74	43	115
16	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	8	10	43	31.8	67
17	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	21	35	282	54.6	579
17	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	8	11	83	38	170

`

17	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	7	79	48.1	162
18	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	11	23	480	48.2	577
18	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	5	5	100	44.2	120
18	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	14	22	97	49	117
19	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	11	27	414	46.7	630
19	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	15	19	88	52.5	134
19	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	8	85	42.6	129
19	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	4	4	8	15.7	12
20	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	8	14	1620	54.8	575
20	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	25	57	1114	55.1	396
20	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	13	26	468	59.8	166
20	Cruciferin BnC2 OS=*Brassica napus* GN=BnC2 PE=3 SV=2	CRU2_BRANA	54.3	8	11	203	37.9	72
20	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	6	8	22	21.1	8
20	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	3	4	16	20	6
20	Myrosinase OS=*Brassica napus* PE=2 SV=1	MYRO_BRANA	62.7	3	3	9	7.12	3

* Spot number refers to the protein spots as represented in Figure 4.11D.
** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A3. LC-MS/MS results showing 100% probable protein matches for the *C. sativa* pH 3 extracted napin separated by 2D-electrophoresis using pH 9 to 12 IPG strips.

Spot number *	Gene name	Annotation**	molecular weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS***
1	Csa06g037810.1	Late embryogenesis abundant protein family	10.4	6	11	197	70.1	292
1	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	88	21.5	131
1	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	60	27.7	89
1	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	46	40.2	68
1	Csa05g009000.1	LEA domain-containing protein	69.3	3	4	24	12.8	36
1	Csa06g048690.1	LEA domain-containing protein	70.3	2	3	22	12.3	33
1	Csa10g017330.1	CAP160 protein	64.4	6	6	14	14.9	21
1	Csa01g023440.1	endoribonuclease L-PSP family protein	19.6	5	5	11	33.7	16
1	Csa14g004960.1	CRB-1-G2	64.9	3	4	7	10.1	10
1	Csa03g036940.1	Late embryogenesis abundant protein, group 1 protein	13.7	2	2	5	8.73	7
2	Csa06g037810.1	Late embryogenesis abundant protein	10.4	4	7	136	51.5	282
2	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	73	21.5	151
2	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	60	27.7	124
2	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	53	40.2	110
2	Csa03g036940.1	Late embryogenesis abundant protein, group 1 protein	13.7	2	3	12	8.73	25
2	Csa05g009000.1	LEA domain-containing protein	69.3	2	3	10	7.69	21
2	Csa08g057250.1	seed gene 3	22.6	2	2	6	11	12
3	Csa11g017020.1	Cs2S-1-G1	18.8	4	8	95	40.2	336
3	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	81	30.3	293
3	Csa11g017000.1	Cs2S-4-G1	36.6	3	3	78	33.4	276
3	Csa06g037810.1	Late embryogenesis abundant protein	10.4	2	2	4	33	14
4	Csa01g018300.1	(LEA) family protein	25.3	3	4	77	31.4	171
4	Csa05g009000.1	LEA domain-containing protein	69.3	5	6	48	18.6	107
4	Csa06g048690.1	LEA domain-containing protein	70.3	5	5	42	21.3	93
4	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	30	27.7	67
4	Csa17g023640.1	Kunitz family trypsin and protease inhibitor protein	22.0	3	3	28	65.6	62
4	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	20	40.2	44
4	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	18	21.5	40
4	Csa08g057250.1	seed gene 3	22.6	2	2	3	11	7
5	Csa15g020270.1	(LEA) family protein	27.2	1	5	212	47.8	301
5	Csa19g022460.1	(LEA) family protein	23.0	1	1	164	53.3	233
5	Csa01g018300.1	(LEA) family protein	25.3	3	3	86	52.1	122
5	Csa05g009000.1	LEA domain-containing protein	69.3	5	8	74	21.1	105
5	Csa06g048690.1	LEA domain-containing protein	70.3	5	8	67	21.3	95

`

5	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	36	27.7	51
5	Csa11g017000.1	Cs2S-4-G1	36.6	2	2	26	29	37
5	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	24	40.2	34
6	Csa01g018300.1	(LEA) family protein	25.3	4	5	93	41.9	199
6	Csa15g020270.1	(LEA) family protein	27.2	1	2	92	37.6	197
6	Csa05g009000.1	LEA domain-containing protein	69.3	5	8	55	19.1	118
6	Csa06g048690.1	LEA domain-containing protein	70.3	5	7	51	19.4	109
6	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	25	27.7	54
6	Csa11g017020.1	Cs2S-1-G1	18.8	2	2	15	36	32
7	Csa05g009000.1	LEA domain-containing protein	69.3	6	10	111	21.7	247
7	Csa06g048690.1	LEA domain-containing protein	70.3	5	9	104	24.4	231
7	Csa15g020270.1	(LEA) family protein	27.2	1	2	57	37.6	127
7	Csa19g022460.1	(LEA) family protein	23.0	1	1	50	46.3	111
7	Csa01g018300.1	(LEA) family protein	25.3	3	3	38	37.3	84
7	Csa12g024730.1	Cs2S-2-G3	36.1	2	2	19	25.5	42
7	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	18	40.2	40
7	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	13	21.5	29
7	Csa01g023440.1	endoribonuclease L-PSP family protein	19.6	2	2	2	13.9	4
8	Csa11g017000.1	Cs2S-4-G1	36.6	3	3	569	32.8	413
8	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	553	34.1	402
8	Csa11g017020.1	Cs2S-1-G1	18.8	5	11	541	52.4	392
8	Csa12g024720.1	Cs2S-4-G3	36.6	0	0	472	30.6	342
8	Csa06g037810.1	Late embryogenesis abundant protein	10.4	2	4	15	33	11
8	Csa04g046970.1	Pollen Ole e 1 allergen and extensin family protein	19.6	4	5	14	27.8	10
8	Csa08g057250.1	seed gene 3	22.6	3	3	13	20.1	9
8	Csa08g061850.1	Unknown protein	15.0	3	3	10	35	7
8	Csa10g027860.1	Hyaluronan / mRNA binding family	41.5	2	2	4	7.2	3
8	Csa05g009000.1	LEA domain-containing protein	69.3	2	3	12	4.77	3
9	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	364	34.1	404
9	Csa12g024720.1	Cs2S-4-G3	36.6	1	1	352	32.8	382
9	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	364	34.1	404
9	Csa12g024720.1	Cs2S-4-G3	36.6	1	1	352	32.8	382
9	Csa11g017000.1	Cs2S-4-G1	36.6	3	4	309	32.8	336
9	Csa11g017020.1	Cs2S-1-G1	18.8	5	10	287	52.4	312
9	Csa06g037810.1	Late embryogenesis abundant protein	10.4	2	3	14	33	15
9	Csa08g057250.1	seed gene 3	22.6	3	3	10	20.1	11
9	Csa18g023600.1	RNA binding Plectin/S10 domain-containing protein	20.0	2	2	8	14.7	9
10	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	596	37.3	420
10	Csa12g024720.1	Cs2S-4-G3	36.6	3	3	535	43.8	370
10	Csa11g017000.1	Cs2S-4-G1	36.6	3	6	496	36	343
10	Csa11g017020.1	Cs2S-1-G1	18.8	5	11	477	54.3	330

`

10	Csa18g023600.1	RNA binding Plectin/S10 domain-containing protein	20.0	3	3	17	15.2	12
10	Csa04g061210.1	Scorpion toxin-like knottin superfamily protein	9.6	2	2	15	24.7	10
10	Csa08g057250.1	seed gene 3	22.6	4	4	15	27.3	10
10	Csa18g009670.1	CRA-1-G2	51.9	3	3	6	17	4
10	Csa04g061160.1	Trypsin inhibitor protein 2	10.3	2	2	5	18.5	3
11	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	624	36.3	463
11	Csa11g017000.1	Cs2S-4-G1	36.6	2	4	564	25.2	408
11	Csa11g017020.1	Cs2S-1-G1	18.8	4	8	522	52.4	378
11	Csa12g024720.1	Cs2S-4-G3	36.6	1	1	494	32.8	357
11	Csa07g047380.1	Plant defensin 1.2C	8.7	2	2	5	25	9
11	Csa18g023600.1	RNA binding Plectin/S10 domain-containing protein	20.0	2	2	5	14.7	4
11	Csa08g057250.1	seed gene 3	22.6	2	2	3	15.8	2
12	Csa12g024730.1	Cs2S-2-G3	36.1	2	5	1012	36.3	489
12	Csa12g024720.1	Cs2S-4-G3	36.6	1	1	865	35	405
12	Csa11g017000.1	Cs2S-4-G1	36.6	2	5	848	27.4	397
12	Csa11g017020.1	Cs2S-1-G1	18.8	3	5	744	43.3	348
12	Csa19g024650.1	(LEA) family protein	34.3	4	4	12	12.2	6
12	Csa02g005050.1	GLNB1 homolog	24.6	2	2	2	11.1	1
13	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	834	36.3	435
13	Csa11g017020.1	Cs2S-1-G1	18.8	5	15	733	52.4	375
13	Csa11g017000.1	Cs2S-4-G1	36.6	3	6	724	35	370
13	Csa12g024720.1	Cs2S-4-G3	36.6	1	1	696	35	356
13	Csa19g024650.1	(LEA) family protein	34.3	3	3	8	11.2	4
13	Csa06g041400.1	Lipid transfer protein 1	20.2	2	2	3	9.5	2
14	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	850	37.3	507
14	Csa12g024720.1	Cs2S-4-G3	36.6	2	2	772	36	446
14	Csa11g017000.1	Cs2S-4-G1	36.6	4	7	630	37.2	364
14	Csa11g017020.1	Cs2S-1-G1	18.8	4	8	535	45.1	309
15	Csa11g017000.1	Cs2S-4-G1	36.6	4	7	978	38.5	433
15	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	901	36.3	407
15	Csa12g024720.1	Cs2S-4-G3	36.6	2	2	906	37.2	401
15	Csa11g017020.1	Cs2S-1-G1	18.8	4	12	780	43.3	345
15	Csa02g076390.1	Dehydrin family protein	18.1	3	5	10	32.4	4
15	Csa19g024650.1	(LEA) family protein	34.3	2	2	4	6.41	2
16	Csa12g024730.1	Cs2S-2-G3	36.1	2	5	826	31.8	312
16	Csa11g017000.1	Cs2S-4-G1	36.6	4	8	813	33.1	306
16	Csa11g017020.1	Cs2S-1-G1	18.8	4	19	776	43.3	292
16	Csa12g024720.1	Cs2S-4-G3	36.6	2	2	736	38.5	277
16	Csa10g029150.1	Cystatin/monellin superfamily protein	13.0	1	1	112	46.3	42
16	Csa11g033420.1	Cystatin/monellin superfamily protein	12.9	1	1	111	46.7	42

`

16	Csa09048s010.1	oleosin 2	12.8	3	4	14	19.7	30
16	Csa19g048250.1	Nucleolar RNA-binding Nop10p family protein	7.3	5	6	39	59.4	15
16	Csa03g052870.1	1-cysteine peroxiredoxin 1	24.0	3	6	21	15.7	8
16	Csa01g010850.1	Unknown protein	16.9	2	2	10	13.3	4
16	Csa04g061160.1	Trypsin inhibitor protein 2	10.3	2	2	7	18.5	3
16	Csa03g018760.1	chaperonin 10	24.4	2	2	6	19.4	2
16	Csa01g025880.1	Vic1-3-G3	56.0	2	2	4	3.87	2
17	Csa12g024730.1	Cs2S-2-G3	36.1	2	5	833	34.1	372
17	Csa11g017000.1	Cs2S-4-G1	36.6	4	8	806	31.9	359
17	Csa11g017020.1	Cs2S-1-G1	18.8	3	8	672	43.3	299
17	Csa12g024720.1	Cs2S-4-G3	36.6	3	6	647	40.7	288
17	Csa07g047380.1	Plant defensin 1.2C	8.7	2	2	8	25	14
17	Csa10g029150.1	Cystatin/monellin superfamily protein	13.0	1	1	30	46.3	13
17	Csa11g033420.1	Cystatin/monellin superfamily protein	12.9	1	1	26	46.7	12
17	Csa19g048250.1	Nucleolar RNA-binding Nop10p family protein	7.3	3	3	11	43.8	5
17	Csa01g010850.1	Unknown protein	16.9	2	2	6	8	3
18	Csa12g024730.1	Cs2S-2-G3	36.1	3	6	1300	36.3	441
18	Csa12g024720.1	Cs2S-4-G3	36.6	4	5	962	41.6	325
18	Csa11g017020.1	Cs2S-1-G1	18.8	4	17	818	43.3	276
18	Csa11g017000.1	Cs2S-4-G1	36.6	3	5	599	34.1	202
18	Csa06g041400.1	Lipid transfer protein 1	20.2	2	2	3	9.5	1

* Spot number refers to the protein spots as represented in Figure 4.15B.
** Annotations for *C. sativa* storage protein genes were assigned according to Table 4.3 and the rest is according to the lowest probability obtained for each gene from the BLAST search preformed against the *Arabidopsis thaliana* genome using TAIR 8 database available from (https://www.arabidopsis.org/).
*** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A4. LC-MS/MS results showing 100% probable protein matches for the *B.napus* pH 3 extracted napin separated by 2D-electrophoresis using pH 9 to 12 IPG strips.

Spot Number*	Protein name	Accession	Molecular Weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS**
1	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	6	7	22	19.6	272
1	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	9	10	16	22.6	182
2	Late embryogenesis abundant protein 76 OS=*Brassica napus* PE=2 SV=2	LEA76_BRANA	30.4	13	33	382	31.4	479
2	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	13	13	31	47.7	40
2	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	2	3	5	6.88	6
3	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	5	12	289	36.4	1168
3	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	1	3	61	46.4	247
3	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	1	41	28.7	166
4	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	6	13	232	41.8	953
4	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	1	4	59	40	242
4	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	1	39	24.2	160
5	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	4	8	942	56.8	694
5	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	2	914	30.3	674
5	Napin-2 OS=*Brassica napus* PE=2 SV=2	2SS2_BRANA	20.1	2	3	755	35.4	557
5	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	6	349	24.2	257
5	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	2	4	26	19.1	19
6	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	4	8	1157	56.8	769
6	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	4	1111	30.3	739
6	Napin-2 OS=*Brassica napus* PE=2 SV=2	2SS2_BRANA	20.1	2	3	934	35.4	621
6	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	7	430	24.2	286
6	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	2	4	23	19.1	15
7	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	5	11	1493	61.6	892
7	Napin-2 OS=*Brassica napus* PE=2 SV=2	2SS2_BRANA	20.1	2	7	1113	35.4	665
7	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	1	1057	30.3	632
7	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	8	588	24.2	351
7	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	2	4	25	19.1	15
8	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	4	9	1707	56.8	975
8	Napin-2 OS=*Brassica napus* PE=2 SV=2	2SS2_BRANA	20.1	2	3	1256	35.4	717
8	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	4	14	961	29	549
8	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	2	4	643	33.7	367
8	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	2	4	10	19.1	6
9	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	4	7	875	56.8	863
9	Napin-2 OS=*Brassica napus* PE=2 SV=2	2SS2_BRANA	20.1	1	1	674	35.4	665
9	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	3	8	552	26.3	544
9	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	2	429	30.3	423

`

10	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	3	6	791	56.8	674
10	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	4	11	762	29	623
10	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	1	2	335	30.3	274
11	Napin-3 OS=*Brassica napus* PE=1 SV=1	2SS3_BRANA	14.0	3	7	877	56.8	796
11	Napin-2 OS=*Brassica napus* PE=2 SV=2	2SS2_BRANA	20.1	2	3	750	35.4	680
11	Napin-B OS=*Brassica napus* GN=NAPB PE=2 SV=1	2SSB_BRANA	20.1	2	3	744	32	675
11	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	6	332	24.2	301
11	Napin-1A OS=*Brassica napus* PE=1 SV=1	2SSI_BRANA	12.7	2	4	28	19.1	25
11	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	2	2	3	8.39	5

* Spot number refers to the protein spots as represented in Figure 4.15D.
** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A5. LC-MS/MS results showing 100% probable protein matches for the *C. sativa* oil body proteins separated by 2D-electrophoresis using pH 3 to 10 IPG strips.

Spot Number *	Gene name	Annotation**	molecular weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS***
1	Csa12g024730.1	Cs2S-2-G3	36.1	2	5	227	19.4	172
1	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	162	19.2	123
1	Csa11g017020.1	Cs2S-4-G3	18.8	3	3	114	40.9	86
1	Csa03g011110.1	Histone superfamily protein	11.4	8	11	60	38.2	46
1	Csa12g021990.1	CRC-1-G3	55.4	7	8	54	36	41
1	Csa11g015240.1	CRC-1-G1	72.9	1	1	50	19.9	38
1	Csa10g014100.1	CRC-1-G2	54.0	1	1	48	31.1	36
1	Csa18g009670.1	CRA-1-G2	51.9	2	3	41	34.4	33
1	Csa03g059740.1	Histone H2A protein 9	14.3	5	6	37	40.3	28
1	Csa11g070580.1	CRA-1-G1	52.0	1	1	37	34.2	28
1	Csa01g023730.1	Histone H2A 13	13.9	3	5	35	43.9	27
1	Csa03g012310.1	Histone superfamily protein	14.8	1	2	36	41.5	27
1	Csa10g015740.1	Ribosomal protein L14	15.5	6	8	35	52.2	27
1	Csa00532s200.1	Oleosin family protein	21.1	1	1	30	21.3	23
1	Csa03g001760.1	Late embryogenesis abundant protein	16.5	1	1	26	48.3	20
1	Csa19g002660.1	Ribosomal protein S19e family protein	15.8	3	4	26	48.3	20
1	Csa11g017470.1	Seed gene 1	37.8	2	2	25	31.5	19
1	Csa10g016060.1	Seed gene 1	28.0	2	2	24	35.5	18
1	Csa10g029650.1	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	18.3	2	2	22	39.3	17
1	Csa01g011530.1	Ribosomal S17 family protein	15.9	6	8	23	31.4	17
1	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	1	1	20	27.7	15
1	Csa10g007580.1	Ribosomal protein S25 family protein	12.0	2	2	20	42.6	15
1	Csa04g015780.1	Ole-4-G1	20.5	1	1	18	36.8	14
1	Csa12g002060.1	Ribosomal protein S25 family protein	12.1	2	3	18	44.4	14
1	Csa10g047190.1	Ole-2-G2	21.3	4	4	18	26.5	14
1	Csa13g021270.1	Ribosomal protein S5 domain 2-like superfamily protein	16.6	4	5	18	28.8	14
1	Csa10g009990.1	Cytochrome bd ubiquinol oxidase, 14kDa subunit	14.6	5	5	18	40.2	14
1	Csa07g065640.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	92.3	7	7	17	11	13
1	Csa06g008780.1	Ole-4-G1	20.5	1	1	17	36.8	13
1	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	1	1	16	23.4	12
1	Csa02g041750.1	Ole-3-G3	15.0	2	4	12	32.9	9
1	Csa13g044730.1	S18 ribosomal protein	17.6	2	2	12	16.3	9
1	Csa05g007000.1	Ribosomal L38e protein family	10.8	2	3	12	26.9	9
1	Csa08g035240.1	Nucleoside diphosphate kinase family protein	16.4	4	5	12	32.2	9

`

1	Csa03g024410.1	Small nuclear ribonucleoprotein family protein	14.1	2	3	10	14.7	8
1	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	1	1	11	12	8
1	Csa10g001480.1	rotamase CYP 1	18.6	3	5	11	23.7	8
1	Csa01g006420.1	Ribosomal L22e protein family	18.6	3	3	9	19.8	7
1	Csa07g027910.1	Unknown protein	26.4	1	1	9	20.5	7
1	Csa12g028090.1	Ole-1-G3	19.6	2	2	9	18.8	7
1	Csa08g001390.1	Ribosomal protein S10p/S20e family protein	13.7	2	2	8	19.7	6
1	Csa07g051310.1	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	16.7	2	3	7	12.7	5
1	Csa11g019460.1	Ole-1-G1	19.7	1	1	7	21	5
1	Csa02g019830.1	glutathione peroxidase 6	25.6	4	4	7	18.3	5
1	Csa05g067280.1	Unknown protein	36.5	1	1	7	15.1	5
1	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	3	3	6	8.4	5
1	Csa02g070280.1	Embryo-specific protein 3, (ATS3)	21.0	2	2	5	16.4	4
1	Csa00441s380.1	Ribosomal protein S30 family protein	6.9	2	2	5	10.8	4
1	Csa06g016800.1	MD-2-related lipid recognition domain-containing protein	21.5	2	2	5	18.4	4
1	Csa10g010630.1	glutathione peroxidase 7	25.9	2	2	4	10.3	3
1	Csa05g023090.1	Caleosin-related family protein	35.1	2	2	4	5.68	3
1	Csa11g007230.1	LYR family of Fe/S cluster biogenesis protein	13.6	3	3	4	29.9	3
1	Csa11g072130.1	sterol carrier protein 2	13.6	2	2	4	17.1	3
1	Csa04g041530.1	Ribosomal protein L23AB	17.5	2	2	3	14.9	2
1	Csa08g057250.1	Seed gene 3	22.6	2	2	2	9.09	2
1	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	2	2	3	5.93	2
1	Csa02g005590.1	Protein of unknown function, DUF538	16.9	2	2	3	13.5	2
1	Csa05g092580.1	Ribosomal protein S13/S15	17.1	1	1	1	2.28	1
3	Csa12g024730.1	Cs2S-2-G3	36.1	3	6	133	15.6	300
3	Csa03g001760.1	Late embryogenesis abundant protein	16.5	1	2	26	47.7	59
3	Csa03g011110.1	Histone superfamily protein	11.4	10	15	46	40.1	35
3	Csa12g021990.1	CRC-1-G3	55.4	1	1	9	9.15	20
3	Csa10g015740.1	Ribosomal protein L14	15.5	2	2	7	23.9	16
3	Csa04g038900.1	Thioredoxin superfamily protein	24.4	4	4	7	24.1	16
3	Csa02g041750.1	Ole-3-G3	15.0	2	3	6	32.9	14
3	Csa08g057250.1	Seed gene 3	22.6	2	2	4	12.9	9
3	Csa15g023060.1	Adenine nucleotide alpha hydrolases-like superfamily protein	17.8	2	2	4	17.8	9
3	Csa02g005590.1	Protein of unknown function, DUF538	16.9	3	3	4	20	9
3	Csa13g011500.1	actin 7	41.7	1	1	3	7.16	7
4	Csa04g015780.1	Ole-4-G1	20.5	4	7	107	43	115
4	Csa06g008780.1	Ole-4-G2	20.5	4	7	105	43.5	113
4	Csa11g019460.1	Ole-1-G1	19.7	4	6	92	35.4	99
4	Csa12g028090.1	Ole-1-G3	19.6	4	12	113	35.4	99

`

4	Csa10g047190.1	Ole-2-G2	21.3	6	9	76	33.5	82
4	Csa00532s200.1	Oleosin family protein	21.1	2	4	72	32.5	77
4	Csa18g009670.1	CRA-1-G2	51.9	2	3	67	37.2	75
4	Csa17g006950.1	CRB-1-G1	50.7	5	7	63	27.3	68
4	Csa12g021990.1	CRC-1-G3	55.4	7	13	60	31.9	65
4	Csa11g070580.1	CRA-1-G1	52.0	1	1	58	37	62
4	Csa10g014100.1	CRC-1-G2	54.0	1	1	39	24.8	42
4	Csa11g015240.1	CRC-1-G1	72.9	1	2	37	15.3	40
4	Csa11g017470.1	Seed gene 1	37.8	1	1	22	28.1	35
4	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	25	14.3	27
4	Csa14g008800.1	HSP20-like chaperones superfamily protein	18.3	4	6	23	33.5	25
4	Csa15g079170.1	RmlC-like cupins superfamily protein	85.0	4	4	12	6.74	13
4	Csa09g042380.1	HSP20-like chaperones superfamily protein	17.5	1	1	10	23.9	11
4	Csa02g041750.1	Ole-3-G3	15.0	2	2	7	32.9	8
4	Csa08g003200.1	17.6 kDa class II heat shock protein	17.5	3	3	5	17.5	5
4	Csa13g044730.1	S18 ribosomal protein	17.6	1	1	4	12.7	4
4	Csa02g062630.1	temperature-induced lipocalin	21.7	1	1	4	16.9	4
4	Csa04g041530.1	Ribosomal protein L23AB	17.5	2	2	3	14.9	3
4	Csa05g092580.1	Ribosomal protein S13/S15	17.1	1	1	1	2.28	1
5	Csa03g011110.1	Histone superfamily protein	11.4	2	2	4	14.5	104
5	Csa10g014100.1	CRC-1-G2	54.0	5	10	186	46	96
5	Csa12g021990.1	CRC-1-G3	55.4	10	21	182	49	94
5	Csa11g015240.1	CRC-1-G1	72.9	2	4	178	34.3	92
5	Csa04g015780.1	Ole-4-G1	20.5	4	9	146	44	76
5	Csa18g009670.1	CRA-1-G2	51.9	2	8	122	40.4	73
5	Csa10g016060.1	Seed gene 1	28.0	1	1	24	28.2	70
5	Csa06g008780.1	Ole-4-G2	20.5	4	7	134	44.6	69
5	Csa10g047190.1	Ole-2-G2	21.3	10	17	110	43	57
5	Csa11g070580.1	CRA-1-G1	52.0	2	5	102	40.2	53
5	Csa00532s200.1	Oleosin family protein	21.1	2	4	97	34	50
5	Csa12g028090.1	Ole-1-G3	19.6	4	12	105	35.4	49
5	Csa11g019460.1	Ole-1-G1	19.7	4	7	94	35.4	49
5	Csa14g004960.1	CRB-1-G2	64.9	1	2	83	40.1	43
5	Csa26607s010.1	Oleosin family protein	8.3	0	0	73	70.8	38
5	Csa05g035620.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	71	76.3	37
5	Csa16g016260.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	71	76.3	37
5	Csa07g015700.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	72	76.3	37
5	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	61	19.4	32
5	Csa17g006950.1	CRB-1-G1	50.7	1	1	57	25.3	30
5	Csa04g046970.1	Pollen Ole e 1 allergen and extensin family protein	19.6	1	1	27	30.6	16

`

5	Csa11g017020.1	Cs2S-4-G3	18.8	2	2	31	32.3	16
5	Csa05g020560.1	Pollen Ole e 1 allergen and extensin family protein	19.1	2	2	26	47.4	15
5	Csa04g039480.1	Ribosomal protein L11 family protein	18.0	3	6	28	40.4	15
5	Csa10g029650.1	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	18.3	1	1	25	39.3	13
5	Csa11g017470.1	Seed gene 1	37.8	2	2	26	27.2	13
5	Csa11g060020.1	RNA binding Plectin/S10 domain-containing protein	30.6	6	7	25	21	13
5	Csa19g031870.2	Vic1-1-G2	58.4	4	4	24	18.3	12
5	Csa04g041530.1	Ribosomal protein L23AB	17.5	8	9	21	32.5	11
5	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	6	7	21	11.4	11
5	Csa15g020270.1	Late embryogenesis abundant protein (LEA) family protein	27.2	1	1	21	20.4	11
5	Csa01g025880.1	Vic1-3-G3	56.0	2	2	21	15.3	11
5	Csa08g017210.1	Histone superfamily protein	15.9	1	1	20	41.1	10
5	Csa01g011530.1	Ribosomal S17 family protein	15.9	5	6	19	31.4	10
5	Csa10g017160.1	RNA binding Plectin/S10 domain-containing protein	19.6	1	1	18	26.1	9
5	Csa02g048870.1	RNA binding Plectin/S10 domain-containing protein	20.0	1	1	18	21	9
5	Csa06g039870.1	Ribosomal protein L11 family protein	18.0	1	2	17	28.9	9
5	Csa01g018300.1	Late embryogenesis abundant protein (LEA) family protein	25.3	2	2	13	17.4	7
5	Csa02g041750.1	Ole-3-G3	15.0	2	3	12	32.9	6
5	Csa08g057250.1	Seed gene 3	22.6	4	4	11	22	6
5	Csa09g042380.1	HSP20-like chaperones superfamily protein	17.5	2	2	9	26.5	5
5	Csa05g021500.1	arabinogalactan protein 30	27.1	2	3	10	5.06	5
5	Csa10g014190.1	Nucleic acid-binding, OB-fold-like protein	16.7	2	2	6	18.3	3
5	Csa04g035480.1	Translation protein SH3-like family protein	16.9	2	2	6	17.8	3
5	Csa04g029010.1	HSP20-like chaperones superfamily protein	17.7	2	2	6	25.5	3
5	Csa00751s020.1	Ribosomal protein large subunit 16A	20.8	3	3	5	17.7	3
5	Csa17g006960.1	CRD-1-G1	50.5	1	1	4	9.65	2
5	Csa08g014130.1	Ribosomal protein S8e family protein	25.3	2	2	4	13	2
5	Csa03g053840.1	Oleosin family protein	18.4	2	2	4	15.4	2
5	Csa08g055140.1	Ribosomal protein S19 family protein	34.1	2	2	2	15.1	1
5	Csa07g051310.1	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	16.7	2	2	2	12.7	1
5	Csa15g079170.1	RmlC-like cupins superfamily protein	85.0	2	2	2	3.79	1
6	Csa12g024730.1	Cs2S-2-G3	36.1	4	8	240	28.3	417
6	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	224	23	389
6	Csa11g017020.1	Cs2S-4-G3	18.8	3	3	149	45.7	259
6	Csa10g014100.1	CRC-1-G2	54.0	3	3	21	24	36
6	Csa18g009670.1	CRA-1-G2	51.9	2	2	18	18.7	35
6	Csa11g015240.1	CRC-1-G1	72.9	1	1	20	12.7	35
6	Csa12g021990.1	CRC-1-G3	55.4	3	4	19	15.9	33
6	Csa11g070580.1	CRA-1-G1	52.0	1	1	14	18.6	24

`

6	Csa10g009990.1	Cytochrome bd ubiquinol oxidase, 14kDa subunit	14.6	3	3	10	27	17
6	Csa10g047190.1	Ole-2-G2	21.3	2	2	7	10.5	12
6	Csa08g057250.1	Seed gene 3	22.6	3	3	5	12.9	9
6	Csa01g011530.1	Ribosomal S17 family protein	15.9	2	2	4	15.7	7
6	Csa03g011110.1	Histone superfamily protein	11.4	8	9	28	39.5	2
7	Csa12g024730.1	Cs2S-2-G3	36.1	2	7	327	25.2	453
7	Csa12g024720.1	Seed storage albumin 4	36.6	1	2	210	34.1	291
7	Csa11g017000.1	Cs2S-4-G1	36.6	2	2	200	28.4	277
7	Csa11g017020.1	Cs2S-4-G3	18.8	2	3	155	34.8	215
7	Csa12g021990.1	CRC-1-G3	55.4	4	5	25	18.1	35
7	Csa18g009670.1	CRA-1-G2	51.9	1	2	17	15.5	24
7	Csa10g014100.1	CRC-1-G2	54.0	1	1	17	17.4	24
7	Csa05g007000.1	Ribosomal L38e protein family	10.8	2	3	16	26.9	22
7	Csa08g060640.1	non-intrinsic ABC protein 10	7.4	4	4	15	10.8	21
7	Csa11g015240.1	CRC-1-G1	72.9	1	1	13	9.79	18
7	Csa10g015740.1	Ribosomal protein L14	15.5	2	3	11	34.3	15
7	Csa12g028090.1	Ole-1-G3	19.6	2	2	10	18.8	14
7	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	2	2	7	3.94	10
7	Csa02g064030.1	Ribosomal protein S4 (RPS4A) family protein	29.9	2	2	6	10.7	8
7	Csa13g036140.1	Ribosomal protein L6 family	22.0	1	2	6	7.73	8
7	Csa03g002070.1	Protein of unknown function (DUF1138)	9.2	2	2	6	30.6	8
7	Csa08g057250.1	Seed gene 3	22.6	3	3	6	18.2	8
7	Csa19g002520.1	Unknown protein	5.6	2	2	6	59.3	8
7	Csa03g011110.1	Histone superfamily protein	11.4	3	3	5	18.4	7
7	Csa07g047380.1	Plant defensin 1.2C	8.7	2	2	4	25	6
8	Csa04g015780.1	Ole-4-G1	20.5	4	9	160	44	149
8	Csa06g008780.1	Ole-4-G2	20.5	5	9	145	44.6	135
8	Csa10g047190.1	Ole-2-G2	21.3	8	17	118	38	110
8	Csa12g021990.1	CRC-1-G3	55.4	6	12	114	25.8	106
8	Csa18g009670.1	CRA-1-G2	51.9	2	6	89	32.5	93
8	Csa11g015240.1	CRC-1-G1	72.9	2	3	84	15	78
8	Csa10g014100.1	CRC-1-G2	54.0	4	5	84	26.1	78
8	Csa12g028090.1	Ole-1-G3	19.6	4	9	74	40.3	69
8	Csa00532s200.1	Oleosin family protein	21.1	2	4	74	34	69
8	Csa11g070580.1	CRA-1-G1	52.0	1	3	71	32.3	66
8	Csa11g019460.1	Ole-1-G1	19.7	3	5	71	37.6	66
8	Csa26607s010.1	Oleosin family protein	8.3	0	0	57	70.8	53
8	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	28	14.3	26
8	Csa11g017470.1	Seed gene 1	37.8	2	2	29	32.4	23
8	Csa10g016060.1	Seed gene 1	28.0	1	1	25	35.1	23
8	Csa02g057710.1	*Arabidopsis thaliana* Peroxygonase 2	27.9	4	6	18	30.9	17
8	sa04g041530.1	Ribosomal protein L23AB	17.5	6	6	15	31.8	14

`

8	Csa17g006950.1	CRB-1-G1	50.7	1	1	10	10.5	9
8	Csa17g006960.1	CRD-1-G1	50.5	1	1	7	12.3	7
8	Csa02g041750.1	Ole-3-G3	15.0	2	2	4	23.1	4
8	Csa04g035480.1	Translation protein SH3-like family protein	16.9	2	2	3	17.8	3
8	Csa01g011530.1	Ribosomal S17 family protein	15.9	1	1	2	8.57	2
9	Csa11g017470.1	Seed gene 1	37.8	3	7	200	50.3	177
9	Csa10g016060.1	Seed gene 1	28.0	4	5	186	66.5	165
9	Csa18g009670.1	CRA-1-G2	51.9	3	4	72	36.6	73
9	Csa11g070580.1	CRA-1-G1	52.0	2	3	76	34.8	67
9	Csa02g057710.1	*Arabidopsis thaliana* Peroxygonase 2	27.9	10	16	68	41.2	60
9	Csa04g015780.1	Ole-4-G1	20.5	3	4	60	43	53
9	Csa06g008780.1	Ole-4-G2	20.5	2	3	56	41.5	50
9	Csa10g047190.1	Ole-2-G2	21.3	8	10	52	38	46
9	Csa12g028090.1	Ole-1-G3	19.6	4	4	50	34.3	44
9	Csa11g019460.1	Ole-1-G1	19.7	3	4	43	31.5	38
9	Csa11g015240.1	CRC-1-G1	72.9	1	1	41	19.4	36
9	Csa19g031870.2	Vic1-1-G2	58.4	3	4	35	21.9	31
9	Csa01g025880.1	Vic1-3-G3	56.0	3	3	35	21.6	31
9	Csa12g021990.1	CRC-1-G3	55.4	4	5	34	20.7	30
9	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	34	15.9	30
9	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	1	1	32	29.7	28
9	Csa00532s200.1	Oleosin family protein	21.1	2	2	31	25.9	27
9	Csa17g006950.1	CRB-1-G1	50.7	0	0	25	18.6	22
9	Csa15g001200.1	voltage dependent anion channel 1	29.5	2	3	23	39.1	20
9	Csa08g007170.1	voltage dependent anion channel 3	29.3	1	1	23	38	20
9	Csa13g017920.1	voltage dependent anion channel 3	29.4	1	1	22	38	20
9	Csa20g021570.1	voltage dependent anion channel 3	29.3	1	1	20	29.7	18
9	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	4	4	12	19.9	12
9	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	4	4	14	5.39	12
9	Csa07g040360.1	Aquaporin-like superfamily protein	27.9	2	2	8	7.92	7
9	Csa10g049280.1	prohibitin 3	30.4	3	3	8	28.2	7
9	Csa13g011500.1	actin 7	41.7	3	3	7	19.9	6
9	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	1	1	7	26.4	6
9	Csa05g060730.1	Glycoprotein membrane precursor GPI-anchored	21.7	2	2	7	19.8	6
9	Csa08g053790.1	NAD(P)-binding Rossmann-fold superfamily protein	38.6	1	1	6	16.4	5
9	Csa03g011110.1	Histone superfamily protein	11.4	4	4	6	26.3	5
9	Csa03g022390.1	Ribosomal protein L6 family protein	26.2	1	1	6	14.1	5
9	Csa08g014130.1	Ribosomal protein S8e family protein	25.3	2	2	4	13	4
9	Csa02g064030.1	Ribosomal protein S4 (RPS4A) family protein	29.9	3	3	5	10.7	4
9	Csa02g041750.1	Ole-3-G3	15.0	2	3	5	23.1	4
9	Csa08g057250.1	Seed gene 3	22.6	2	2	5	12.9	4

`

9	Csa04g002200.1	Ribosomal protein L30/L7 family protein	19.4	2	2	3	11.7	3
9	Csa03g005870.1	AWPM-19-like family protein	19.7	2	2	3	12.4	3
10	Csa11g015240.1	CRC-1-G1	72.9	4	8	264	40.2	145
10	Csa10g014100.1	CRC-1-G2	54.0	3	6	242	54.5	133
10	Csa12g021990.1	CRC-1-G3	55.4	13	26	237	62.8	130
10	Csa18g009670.1	CRA-1-G2	51.9	3	8	172	47.7	101
10	Csa11g070580.1	CRA-1-G1	52.0	3	5	155	47.4	85
10	Csa10g016060.1	Seed gene 1	28.0	4	4	93	46.5	51
10	Csa11g017470.1	Seed gene 1	37.8	3	4	94	35.2	51
10	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	1	1	70	54.6	45
10	Csa11g082030.1	hydroxysteroid dehydrogenase 1	39.2	1	1	68	54.6	37
10	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	3	4	62	39.3	34
10	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	4	4	61	48.3	34
10	Csa02g026890.1	hydroxysteroid dehydrogenase 5	42.9	3	4	58	39.3	32
10	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	54	19.4	30
10	Csa12g028090.1	Ole-1-G3	19.6	4	7	51	34.3	28
10	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	11	13	47	17.6	26
10	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	46	19.2	25
10	Csa02g057710.1	*Arabidopsis thaliana* Peroxygonase 2	27.9	7	11	43	34.2	24
10	Csa17g006950.1	CRB-1-G1	50.7	0	0	40	24.7	22
10	Csa10g047190.1	Ole-2-G2	21.3	4	6	40	27.5	22
10	Csa00532s200.1	Oleosin family protein	21.1	1	1	40	27.9	22
10	Csa11g019460.1	Ole-1-G1	19.7	3	4	34	31.5	19
10	Csa11g017020.1	Cs2S-4-G3	18.8	3	3	32	40.9	18
10	Csa06g008780.1	Ole-4-G2	20.5	1	1	29	37.8	16
10	Csa04g015780.1	Ole-4-G1	20.5	1	1	29	37.8	16
10	Csa07g027910.1	Unknown protein	26.4	2	3	26	42.6	14
10	Csa10g004530.1	RmlC-like cupins superfamily protein	58.6	2	2	20	16.7	11
10	Csa05g067280.1	Unknown protein	36.5	2	3	20	31.4	11
10	Csa03g004310.1	Eukaryotic aspartyl protease family protein	46.4	2	2	4	5.02	10
10	Csa19g031870.2	Vic1-1-G2	58.4	2	2	15	14.6	8
10	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	6	6	15	20.2	8
10	Csa05g009000.1	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	69.3	3	3	15	16.9	8
10	Csa15g001200.1	voltage dependent anion channel 1	29.5	2	3	13	25.4	7
10	Csa20g009380.1	Eukaryotic aspartyl protease family protein	47.1	1	1	12	14.1	7
10	Csa01g025880.1	Vic1-3-G3	56.0	1	1	13	14.1	7
10	Csa06g048690.1	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	70.3	3	3	13	18.3	7
10	Csa03g011110.1	Histone superfamily protein	11.4	4	4	10	26.3	6
10	Csa08g057250.1	Seed gene 3	22.6	4	4	10	22.5	6

`

10	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	6	6	10	17.1	6
10	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	4	4	9	30	5
10	Csa07g040360.1	Aquaporin-like superfamily protein	27.9	2	2	8	7.92	4
10	Csa03g005870.1	AWPM-19-like family protein	19.7	3	3	8	17.1	4
10	Csa00751s020.1	Ribosomal protein large subunit 16A	20.8	3	3	7	17.7	4
10	Csa02g041750.1	Ole-3-G3	15.0	2	2	5	32.9	3
10	Csa10g001480.1	rotamase CYP 1	18.6	2	3	5	16.8	3
10	Csa07g038560.1	Peroxidase superfamily protein	39.5	2	2	6	7.52	3
10	Csa06g028870.1	Eukaryotic aspartyl protease family protein	45.6	3	3	6	7.29	3
10	Csa12g004770.1	RmlC-like cupins superfamily protein	61.4	2	3	25	16.1	2
10	Csa13g044730.1	S18 ribosomal protein	17.6	1	1	3	8.37	2
10	Csa05g092580.1	Ribosomal protein S13/S15	17.1	1	1	4	2.28	2
10	Csa16g014250.1	winged-helix DNA-binding transcription factor family protein	28.8	2	2	3	8.3	2
10	Csa13g011500.1	actin 7	41.7	1	1	4	11.9	2
10	Csa13g021270.1	Csa10g015740.1	16.6	2	2	3	12.3	2
10	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	1	1	4	14.4	2
10	Csa03g022390.1	Ribosomal protein L6 family protein	26.2	2	2	4	14.1	2
10	Csa00506s140.1	Ribosomal L27e protein family	15.5	1	1	2	8.15	1

* Spot number refers to the protein spots as represented in Figure 4.18B.
** Annotations for *C. sativa* storage protein genes were assigned according to Table 4.3. The annotations for the *C. sativa* oleosins were assigned according to the location of the gene in the sub genome. The rest is according to the lowest probability obtained for each gene from the BLAST search preformed against the *Arabidopsis thaliana* genome using TAIR 8 database available from (https://www.arabidopsis.org/).
*** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A6. LC-MS/MS results showing 100% probable protein matches for the *C. sativa* oil body proteins separated by 2D-electrophoresis using pH 9 to 12 IPG strips.

Spot number *	Gene name	Annotation**	molecular weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS***
1	Csa18g009670.1	CRA-1-G2	51.9	5	13	351	61.9	83
1	Csa12g021990.1	CRC-1-G3	55.4	14	31	327	67.7	77
1	Csa11g070580.1	CRA-1-G1	52.0	4	9	311	57.1	73
1	Csa11g015240.1	CRC-1-G1	72.9	5	11	275	42.2	65
1	Csa10g014100.1	CRC-1-G2	54.0	4	6	269	59.2	63
1	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	6	10	173	60.7	41
1	Csa02g026890.1	hydroxysteroid dehydrogenase 5	42.9	4	7	158	57.6	37
1	Csa14g004960.1	CRB-1-G2	64.9	1	3	147	39.4	35
1	Csa17g006950.1	CRB-1-G1	50.7	8	16	147	53.9	35
1	Csa11g017470.1	Seed gene 1	37.8	3	5	143	46.6	34
1	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	2	3	145	68	34
1	Csa03g005050.1	CRB-1-G3	49.4	1	1	139	51.9	33
1	Csa11g082030.1	hydroxysteroid dehydrogenase 1	39.2	1	1	139	68	33
1	Csa10g016060.1	Seed gene 1	28.0	4	4	135	61.6	32
1	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	5	5	124	68	29
1	Csa17g006960.1	CRD-1-G1	50.5	1	1	90	38.8	21
1	Csa14g004970.1	CRD-1-G2	50.1	1	1	87	39.1	20
1	Csa06g008780.1	Ole-4-G2	20.5	3	4	81	40.4	19
1	Csa04g015780.1	Ole-4-G1	20.5	2	2	80	39.9	19
1	Csa12g028090.1	Ole-1-G3	19.6	4	8	75	35.4	18
1	Csa02g057710.1	*Arabidopsis thaliana* peroxygonase 2	27.9	8	13	68	39.1	16
1	Csa10g047190.1	Ole-2-G2	21.3	5	8	65	26.5	15
1	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	63	23.6	15
1	Csa11g019460.1	Ole-1-G1	19.7	2	3	64	32.6	15
1	Csa07g027910.1	Unknown protein	26.4	3	4	61	55.7	14
1	Csa05g067280.1	Unknown protein	36.5	3	5	56	46.2	13
1	Csa20g009380.1	Eukaryotic aspartyl protease family protein	47.1	3	4	57	33	13
1	Csa12g024720.1	Cs2S-4-G3	36.6	2	2	52	32.5	12
1	Csa08g057430.1	Eukaryotic aspartyl protease family protein	74.8	2	2	53	17.3	12
1	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	11	14	46	17.7	11
1	Csa00532s200.1	Oleosin family protein	21.1	1	2	45	25.4	11
1	Csa11g017000.1	Cs2S-4-G1	36.6	2	2	41	26.8	10
1	Csa15g021280.1	nitrile specifier protein 1	51.7	2	3	37	35.7	9
1	Csa05g038120.1	Vic2-1-G3	58.8	1	1	40	18.5	9
1	Csa07g016060.1	Vic2-1-G2	53.2	1	1	38	21.5	9
1	Csa01g016910.1	alpha/beta-Hydrolases superfamily protein	62.9	4	4	34	34.1	8
1	Csa15g016520.1	heat shock protein 70	71.2	3	3	32	19.1	8

`

1	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	14	15	32	38.5	8
1	Csa03g019850.1	heat shock protein 70B	58.2	6	8	35	20.4	8
1	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	11	13	34	37.7	8
1	Csa01g004900.1	glyceraldehyde-3-phosphate dehydrogenase C subunit 1	36.7	0	0	34	47.6	8
1	Csa01g019120.1	nitrile specifier protein 4	59.5	1	1	35	27.1	8
1	Csa14g014760.1	aspartic proteinase A1	54.5	1	2	33	27.7	8
1	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	32	31.7	8
1	Csa19g021040.1	alpha/beta-Hydrolases superfamily protein	61.9	2	2	31	30.1	7
1	Csa03g015610.1	aspartic proteinase A1	54.5	1	1	30	25.6	7
1	Csa19g031870.2	Vic1-1-G2	58.4	4	4	30	26.8	7
1	Csa13g011500.1	actin 7	41.7	6	7	30	42.2	7
1	Csa11g098630.1	calnexin 1	60.4	0	0	29	27.9	7
1	Csa02g074880.1	Heat shock protein 70 (Hsp 70) family protein	71.3	1	1	26	13.8	6
1	Csa19g053880.1	lysm domain GPI-anchored protein 2 precursor	39.1	3	4	24	23.4	6
1	Csa01g025880.1	Vic1-3-G3	56.0	2	2	24	21.6	6
1	Csa12g004770.1	RmlC-like cupins superfamily protein	61.4	2	3	27	14.3	6
1	Csa08g057250.1	Seed gene 3	22.6	5	6	22	29.7	5
1	Csa10g004530.1	RmlC-like cupins superfamily protein	58.6	2	2	23	14.8	5
1	Csa16g016260.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	20	60.2	5
1	Csa07g015700.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	20	60.2	5
1	Csa01g001580.1	Insulinase (Peptidase family M16) protein	63.7	7	8	16	17.7	4
1	Csa03g005870.1	AWPM-19-like family protein	19.7	3	4	18	17.1	4
1	Csa07g040360.1	Aquaporin-like superfamily protein	27.9	3	4	15	12.5	4
1	Csa19g002660.1	Ribosomal protein S19e family protein	15.8	3	3	16	46.9	4
1	Csa08g055540.1	ATP synthase alpha/beta family protein	138.8	7	7	16	21	4
1	Csa18g023590.1	heat shock protein 90.1	81.1	6	6	19	13.6	4
1	Csa02g024850.1	UBX domain-containing protein	54.3	8	8	19	23.1	4
1	Csa03g058960.1	Insulinase (Peptidase family M16) protein	54.3	7	7	18	20.9	4
1	Csa02g053380.1	Heat shock protein 70 (Hsp 70) family protein	73.0	2	2	15	7.19	4
1	Csa15g001200.1	voltage dependent anion channel 1	29.5	2	3	15	34.1	4
1	Csa05g092580.1	Ribosomal protein S13/S15	17.1	3	3	11	7.41	3
1	Csa04g029010.1	HSP20-like chaperones superfamily protein	17.7	3	3	11	34.4	3
1	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	4	4	14	30	3
1	Csa04g038980.1	late embryogenesis abundant protein, putative / LEA protein, putative	49.2	5	5	9	16.4	3
1	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	2	2	12	26.9	3
1	Csa15g023060.1	Adenine nucleotide alpha hydrolases-like superfamily protein	17.8	4	5	12	42.3	3

`

1	Csa04g050030.1	Enolase	52.4	3	3	11	23	3
1	Csa03g022390.1	Ribosomal protein L6 family protein	26.2	4	4	14	25.2	3
1	Csa01g021740.1	RAB GTPase homolog G3F	23.1	3	3	13	31.4	3
1	Csa02g041750.1	Ole-3-G3	15.0	2	4	13	32.9	3
1	Csa15g039300.1	Vic1-2-G1	52.4	4	4	12	16.2	3
1	Csa10g028320.1	RAB GTPase homolog B1C	54.3	2	2	12	16.3	3
1	Csa10g016280.1	heat shock protein 81-2	42.2	4	4	12	21	3
1	Csa13g036140.1	Ribosomal protein L6 family	22.0	6	7	12	43.3	3
1	Csa09g093790.1	phosphoglycerate kinase	42.1	5	5	11	19.5	3
1	Csa05g023090.1	Caleosin-related family protein	35.1	2	2	12	11.4	3
1	Csa03g055130.1	RAB GTPase homolog G3E	23.2	2	2	12	26.6	3
1	Csa04g049610.1	Ribosomal protein S11 family protein	18.0	1	1	13	44	3
1	Csa07g038560.1	Peroxidase superfamily protein	39.5	6	6	11	19.5	3
1	Csa15g021350.1	mitochondrial processing peptidase alpha subunit	54.4	2	2	10	10.8	2
1	Csa04g049690.1	Tubulin/FtsZ family protein	53.4	3	3	9	11.5	2
1	Csa15g076270.1	Senescence/dehydration-associated protein-related	48.4	5	5	9	12	2
1	Csa13g014240.1	Protein of unknown function, DUF642	40.0	4	4	7	14.1	2
1	Csa19g021970.1	Protein phosphatase 2C family protein	31.5	5	5	7	23.9	2
1	Csa09g042380.1	HSP20-like chaperones superfamily protein	17.5	2	2	7	26.5	2
1	Csa08g053790.1	NAD(P)-binding Rossmann-fold superfamily protein	38.6	1	1	9	22	2
1	Csa19g023290.1	nitrile specifier protein 1	55.2	3	3	9	11.5	2
1	Csa01g009210.1	ADP/ATP carrier 1	41.2	2	2	7	10.1	2
1	Csa05g017950.1	Ribosomal protein S11 family protein	16.3	1	1	10	43.7	2
1	Csa20g079430.1	3-ketoacyl-acyl carrier protein synthase I	61.5	4	4	8	9.38	2
1	Csa11g088190.1	Dihydrolipoamide succinyltransferase	49.7	1	1	9	9.57	2
1	Csa02g057460.2	Dihydrolipoamide succinyltransferase	50.0	1	1	9	9.52	2
1	Csa05g009000.1	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	69.3	2	2	7	10.2	2
1	Csa02g039360.1	hydroxysteroid dehydrogenase 6	37.7	4	4	7	12	2
1	Csa09g069460.1	Calcium-dependent phosphotriesterase superfamily protein	121.4	4	4	8	6.86	2
1	Csa01g021620.1	Ribosomal protein L7Ae/L30e/S12e/Gadd45 family protein	12.3	4	4	9	51.8	2
1	Csa10g022860.1	RAB GTPase homolog E1B	52.5	3	3	8	8.88	2
1	Csa03g011110.1	Histone superfamily protein	11.4	4	4	8	26.3	2
1	Csa09g078880.1	heat shock protein 101	101.2	1	1	10	10.9	2
1	Csa14g027540.1	general regulatory factor 10	28.9	2	2	9	13.6	2
1	Csa08g002670.1	aspartate aminotransferase 3	49.1	2	3	9	7.27	2
1	Csa05g007000.1	Ribosomal L38e protein family	10.8	3	3	7	39.8	2

`

1	Csa03g013000.1	Translation elongation factor EF1B, gamma chain	52.7	3	3	7	20	2
1	Csa02g035680.1	ATP synthase subunit 1	55.1	3	3	7	6.71	2
1	Csa11g072000.1	thioredoxin 3	12.9	1	1	9	48.7	2
1	Csa19g006400.1	Ribosomal protein S5 domain 2-like superfamily protein	16.6	2	2	3	19.9	1
1	Csa02g019830.1	glutathione peroxidase 6	25.6	2	2	3	8.7	1
1	Csa02g001470.1	plasma membrane intrinsic protein 1	30.7	3	3	5	11.8	1
1	Csa08g012270.1	aspartate aminotransferase 2	44.3	2	2	5	8.89	1
1	Csa07g057850.1	D-mannose binding lectin protein with Apple-like carbohydrate-binding domain	49.0	2	2	3	4.54	1
1	Csa10g001480.1	rotamase CYP 1	18.6	2	3	6	22	1
1	Csa07g065640.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	92.3	2	2	5	3.74	1
1	Csa04g031030.1	aldehyde dehydrogenase 2B4	58.5	2	2	5	5.21	1
1	Csa15g002430.1	Eukaryotic aspartyl protease family protein	52.8	4	4	5	12.3	1
1	Csa00506s140.1	Ribosomal L27e protein family	15.5	2	2	4	14.1	1
1	Csa10g011360.1	O-Glycosyl hydrolases family 17 protein	52.8	4	4	5	15.3	1
1	Csa01g009920.1	Ribosomal protein L10 family protein	34.3	1	1	3	7.48	1
1	Csa10g007070.1	plasma membrane intrinsic protein 3	29.8	2	2	3	10.4	1
1	Csa08g060050.1	binding to TOMV RNA 1L (long form)	33.9	3	3	4	10.8	1
1	Csa05g029590.1	Ribosomal protein S3 family protein	27.1	3	3	6	13.2	1
1	Csa03g006900.1	Calcium-dependent lipid-binding (CaLB domain) family proein	27.3	2	2	3	3.28	1
1	Csa10g015740.1	3-ketoacyl-acyl carrier protein synthase I	15.5	3	3	5	22.4	1
1	Csa08g006450.1	RING domain ligase2	51.2	2	2	3	6.03	1
1	Csa08g014130.1	Ribosomal protein S8e family protein	25.3	3	3	4	19.7	1
1	Csa11g070810.1	GroES-like zinc-binding dehydrogenase family protein	40.8	3	3	5	13.7	1
1	Csa17g010640.1	UDP-glucosyl transferase 71C3	52.9	2	2	5	4.62	1
1	Csa13g044730.1	S18 ribosomal protein	17.6	2	2	6	16.3	1
1	Csa02g064030.1	Ribosomal protein S4 (RPS4A) family protein	29.9	4	4	4	17.9	1
1	Csa10g032860.1	serine hydroxymethyltransferase 4	51.9	2	2	4	6.37	1
1	Csa00751s020.1	Ribosomal protein large subunit 16A	20.8	3	3	6	17.7	1
1	Csa07g048690.1	alcohol dehydrogenase 1	41.2	2	2	3	5.01	1
1	Csa07g039460.1	40s ribosomal protein SA	36.1	2	2	4	7.07	1
1	Csa02g062670.1	O-Glycosyl hydrolases family 17 protein	52.5	2	2	4	5.44	1
1	Csa08g002300.1	DEAD/DEAH box RNA helicase family protein	48.3	2	2	4	6.25	1
1	Csa01g044670.1	hexokinase 2	53.7	2	2	3	4.58	1
1	Csa13g033910.1	glucoside glucohydrolase 2	24.8	2	2	3	8.92	1
1	Csa11g055440.1	Saccharopine dehydrogenase	51.0	3	3	5	10.8	1

`

1	Csa10g015250.1	SPFH/Band 7/PHB domain-containing membrane-associated protein family	44.9	2	2	3	6.1	1
1	Csa10g018860.1	Chaperone protein htpG family protein	150.5	3	3	4	4.74	1
1	Csa04g066330.1	ADP-ribosylation factor A1E	20.6	4	4	6	31.5	1
1	Csa13g021270.1	Ribosomal protein S5 domain 2-like superfamily protein	16.6	3	3	6	21.9	1
1	Csa06g053650.1	ubiquitin 6	9.4	2	2	6	33.3	1
1	Csa10g007460.1	Ribosomal protein S3Ae	29.7	2	2	3	12.6	1
1	Csa02g040650.1	CLPC homologue 1	103.3	4	4	6	3.26	1
1	Csa02g005250.1	secretion-associated RAS super family 2	22.0	3	3	4	27.9	1
1	Csa09g096900.1	Saposin-like aspartyl protease family protein	55.7	2	2	2	5.45	0
1	Csa08g015690.1	NAD+ ADP-ribosyltransferases	91.7	2	2	2	3.93	0
1	Csa04g056210.1	Ribosomal protein S5 family protein	30.7	2	2	2	9.72	0
2	Csa12g021990.1	CRC-1-G3	55.4	12	22	218	60.6	69
2	Csa11g015240.1	CRC-1-G1	72.9	4	8	210	43.4	67
2	Csa10g014100.1	CRC-1-G2	54.0	3	5	204	60.2	65
2	Csa11g070580.1	CRA-1-G1	52.0	3	5	168	50.9	53
2	Csa18g009670.1	CRA-1-G2	51.9	4	10	186	55.7	53
2	Csa11g017470.1	Seed gene 1	37.8	3	5	131	41	42
2	Csa10g016060.1	Seed gene 1	28.0	4	4	125	54.3	40
2	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	2	2	107	66.3	34
2	Csa06g008780.1	Ole-4-G2	20.5	3	5	105	41.5	33
2	Csa11g082030.1	hydroxysteroid dehydrogenase 1	39.2	1	1	103	61.4	33
2	Csa04g015780.1	Ole-4-G1	20.5	3	5	102	43.5	32
2	Csa12g028090.1	Ole-1-G3	19.6	4	9	101	35.4	32
2	Csa10g047190.1	Ole-2-G2	21.3	5	11	92	28	29
2	Csa02g026890.1	hydroxysteroid dehydrogenase 5	42.9	4	5	89	52.9	28
2	Csa11g019460.1	Ole-4-G1	19.7	4	6	84	35.4	27
2	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	4	4	86	55.1	27
2	Csa14g004960.1	CRB-1-G2	64.9	1	2	75	35.5	24
2	Csa17g006950.1	CRB-1-G1	50.7	3	4	70	45.9	22
2	Csa00532s200.1	Oleosin family protein	21.1	2	4	64	32.5	20
2	Csa15g001200.1	voltage dependent anion channel 1	29.5	3	5	63	46.7	20
2	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	58	19.4	18
2	Csa03g017210.1	glyceraldehyde-3-phosphate dehydrogenase C2	36.9	1	1	53	49.4	17
2	Csa02g057710.1	*Arabidopsis thaliana* peroxigonase 2	27.9	7	11	53	34.2	17
2	Csa01g001220.1	voltage dependent anion channel 1	29.6	2	3	54	46.7	17
2	Csa01g004900.1	glyceraldehyde-3-phosphate dehydrogenase C subunit 1	36.7	1	1	52	49.7	17
2	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	10	11	41	16	13
2	Csa19g031870.2	Vic1-1-G2	58.4	6	6	40	35.1	13

`

2	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	11	12	40	35.6	13
2	Csa01g025880.1	Vic1-3-G3	56.0	3	3	35	28.1	11
2	Csa11g017020.1	Cs2S-1-G1	18.8	1	1	31	23.2	10
2	Csa14g004970.1	CRD-1-G2	50.1	1	1	30	22.7	10
2	Csa17g006960.1	CRD-1-G1	50.5	1	1	32	22.6	10
2	Csa13g017920.1	voltage dependent anion channel 3	29.4	1	1	27	30.4	9
2	Csa07g027910.1	Unknown protein	26.4	2	4	28	42.6	9
2	Csa12g004770.1	RmlC-like cupins superfamily protein	61.4	2	3	26	16.1	8
2	Csa10g049280.1	prohibitin 3	30.4	4	5	24	37.5	8
2	Csa03g022390.1	Ribosomal protein L6 family protein	26.2	5	5	26	39.3	8
2	Csa10g004530.1	RmlC-like cupins superfamily protein	58.6	3	3	22	16.7	7
2	Csa20g021570.1	voltage dependent anion channel 3	29.3	1	1	23	29.7	7
2	Csa02g041750.1	Ole-3-G3	15.0	2	6	21	32.9	7
2	Csa07g044310.1	Ribosomal protein L6 family protein	26.1	2	2	21	39.5	7
2	Csa05g067280.1	Unknown protein	36.5	2	2	19	31.4	6
2	Csa19g031730.1	Seed maturation protein	26.7	1	1	20	36.6	6
2	Csa07g044330.1	Ribosomal protein L6 family protein	25.9	1	1	19	34.3	6
2	Csa01g018300.1	Late embryogenesis abundant protein (LEA) family protein	25.3	3	3	15	25.8	5
2	Csa13g011500.1	actin 7	41.7	3	3	17	22.8	5
2	Csa01g025740.1	Seed maturation protein	26.8	1	1	17	36.6	5
2	Csa15g016520.1	heat shock protein 70	71.2	1	1	15	12.3	5
2	Csa09g096900.1	Saposin-like aspartyl protease family protein	55.7	3	4	15	7.59	5
2	Csa08g053790.1	NAD(P)-binding Rossmann-fold superfamily protein	38.6	2	2	16	26.5	5
2	Csa03g019850.1	heat shock protein 70B	58.2	2	3	17	11.8	5
2	Csa08g014130.1	Ribosomal protein S8e family protein	25.3	5	6	15	29.1	5
2	Csa07g038560.1	Peroxidase superfamily protein	39.5	7	7	16	22.6	5
2	Csa04g030420.1	malate dehydrogenase	42.6	4	4	13	19.4	4
2	Csa10g044580.1	general regulatory factor 3	29.0	2	3	13	18.1	4
2	Csa03g060460.1	Lactate/malate dehydrogenase family protein	90.7	3	3	14	18.2	4
2	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	4	4	12	30	4
2	Csa13g013490.1	NAD(P)-binding Rossmann-fold superfamily protein	31.1	1	1	13	21.4	4
2	Csa03g005870.1	AWPM-19-like family protein	19.7	3	4	13	17.1	4
2	Csa07g040360.1	Aquaporin-like superfamily protein	27.9	2	3	12	7.92	4
2	Csa17g006930.1	prohibitin 2	47.7	1	1	12	18.2	4
2	Csa08g009040.1	binding partner of acd11 1	27.1	1	1	8	20	3
2	Csa03g011110.1	Histone superfamily protein	11.4	4	4	9	26.3	3
2	Csa13g001470.1	Aldolase-type TIM barrel family protein	49.1	4	4	8	10.3	3
2	Csa05g007000.1	Ribosomal L38e protein family	10.8	2	2	8	26.9	3

`

2	Csa08g057250.1	Seed gene 3	22.6	4	4	11	22.5	3
2	Csa13g036140.1	Ribosomal protein L6 family	22.0	4	5	10	30.4	3
2	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	2	2	10	19	3
2	Csa03g050970.1	gamma carbonic anhydrase 2	30.1	4	5	11	23	3
2	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	7	7	11	19.2	3
2	Csa03g015520.1	glyoxalase I homolog	32.0	5	5	10	21.5	3
2	Csa07g039460.1	40s ribosomal protein SA	36.1	5	5	11	24.2	3
2	Csa18g005520.1	Protein of unknown function (DUF1264)	27.8	3	3	9	23.9	3
2	Csa18g010920.1	thioredoxin 3	12.9	1	1	9	48.7	3
2	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	3	4	100	52.6	3
2	Csa15g039300.1	Vic1-2-G1	52.4	3	3	10	13.9	3
2	Csa02g039360.1	hydroxysteroid dehydrogenase 6	37.7	4	4	8	12.9	3
2	Csa20g024770.1	binding partner of acd11 1	27.2	1	1	8	19.2	3
2	Csa11g072000.1	thioredoxin 3	12.9	1	1	10	48.7	3
2	Csa16g043300.1	dehydrin LEA	21.6	2	3	5	32.7	2
2	Csa19g021970.1	Protein phosphatase 2C family protein	31.5	4	4	6	15.9	2
2	Csa03g033410.1	Translation elongation factor EF1B/ribosomal protein S6 family protein	28.5	4	4	7	20.6	2
2	Csa04g041530.1	Ribosomal protein L23AB	17.5	2	2	6	14.9	2
2	Csa02g004530.1	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	24.5	2	2	7	17.1	2
2	Csa01g020210.1	Late embryogenesis abundant protein (LEA) family protein	34.1	4	4	6	10.9	2
2	Csa04g029010.1	HSP20-like chaperones superfamily protein	17.7	2	2	7	28	2
2	Csa03g060150.1	Nucleotide-diphospho-sugar transferases superfamily protein	32.6	3	4	5	10.9	2
2	Csa00511s040.1	fibrillin	35.1	2	2	6	12.1	2
2	Csa01g009920.1	Ribosomal protein L10 family protein	34.3	2	2	5	14.3	2
2	Csa18g022410.1	SPFH/Band 7/PHB domain-containing membrane-associated protein family	32.3	4	4	6	18.9	2
2	Csa03g006900.1	Calcium-dependent lipid-binding (CaLB domain) family protein	27.3	3	3	6	4.92	2
2	Csa04g038850.2	Aldolase superfamily protein	38.5	4	4	7	19.5	2
2	Csa11g070810.1	GroES-like zinc-binding dehydrogenase family protein	40.8	3	3	5	12.1	2
2	Csa10g016280.1	heat shock protein 81-2	42.2	1	1	7	11.1	2
2	Csa07g052870.1	peroxisomal NAD-malate dehydrogenase 1	37.3	4	4	7	16.7	2
2	Csa03g053840.1	Oleosin family protein	18.4	1	1	2	7.1	1
2	Csa08g005830.1	nascent polypeptide-associated complex subunit alpha-like protein 3	22.0	2	2	4	14.9	1
2	Csa01g006090.1	NAD(P)-binding Rossmann-fold superfamily protein	31.3	2	2	4	6.92	1
2	Csa04g066330.1	ADP-ribosylation factor A1E	20.6	2	2	3	11.6	1

`

2	Csa08g002670.1	aspartate aminotransferase 3	49.1	2	2	3	5.29	1
2	Csa02g019830.1	glutathione peroxidase 6	25.6	1	1	2	5.22	1
2	Csa04g051630.1	SOUL heme-binding family protein	25.1	2	2	4	3.36	1
2	Csa07g059780.1	Cystathionine beta-synthase (CBS) family protein	34.4	2	2	4	7.52	1
2	Csa08g012270.1	aspartate aminotransferase 2	44.3	2	2	3	5.93	1
2	Csa11g072210.1	proteasome alpha subunit F1	30.4	2	2	3	10.8	1
2	Csa00630s010.1	NAD(P)-binding Rossmann-fold superfamily protein	24.9	2	2	3	10.7	1
2	Csa10g001480.1	rotamase CYP 1	18.6	2	2	3	16.8	1
2	Csa15g023060.1	Adenine nucleotide alpha hydrolases-like superfamily protein	17.8	2	2	4	17.8	1
2	Csa08g055540.1	ATP synthase alpha/beta family protein	138.8	1	1	4	3.78	1
2	Csa04g056210.1	Ribosomal protein S5 family protein	30.7	2	2	3	9.03	1
2	Csa06g025420.1	20S proteasome alpha subunit PAD1	27.3	2	2	4	10.9	1
2	Csa08g063330.1	NAD(P)-linked oxidoreductase superfamily protein	36.6	2	3	4	8.39	1
2	Csa05g029590.1	Ribosomal protein S3 family protein	27.1	2	2	3	8.4	1
2	Csa13g044730.1	S18 ribosomal protein	17.6	2	2	3	11.6	1
2	Csa00751s020.1	ribosomal protein large subunit 16A	20.8	2	2	3	13.3	1
2	Csa05g087710.1	SPFH/Band 7/PHB domain-containing membrane-associated protein family	31.3	2	2	4	8.77	1
2	Csa02g072050.1	gamma carbonic anhydrase like 1	27.6	1	1	2	4.74	1
2	Csa02g035680.1	ATP synthase subunit 1	55.1	2	2	3	4.34	1
3	Csa11g017470.1	Seed gene 1	37.8	3	7	255	51.2	74
3	Csa10g016060.1	Seed gene 1	28.0	4	7	240	67.8	70
3	Csa11g070580.1	CRA-1-G1	52.0	3	4	181	55.3	52
3	Csa18g009670.1	CRA-1-G2	51.9	5	9	200	57.4	52
3	Csa06g008780.1	Ole-4-G2	20.5	3	6	119	41.5	34
3	Csa14g004960.1	CRB-1-G2	64.9	2	5	115	31.9	33
3	Csa04g015780.1	Ole-4-G1	20.5	2	3	114	40.9	33
3	Csa12g021990.1	CRC-1-G3	55.4	11	17	110	61.6	32
3	Csa03g005050.1	CRB-1-G3	49.4	1	2	107	40.9	31
3	Csa17g006950.1	CRB-1-G1	50.7	3	6	104	35.8	30
3	Csa10g047190.1	Ole-2-G2	21.3	6	11	100	33.5	29
3	Csa12g028090.1	Ole-1-G3	19.6	4	9	100	35.4	29
3	Csa11g015240.1	CRC-1-G1	72.9	2	3	97	32.3	28
3	Csa10g014100.1	CRC-1-G2	54.0	2	4	96	45.1	28
3	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	1	1	76	52	22
3	Csa11g082030.1	hydroxysteroid dehydrogenase 1	39.2	1	1	76	52	22
3	Csa02g057710.1	*Arabidopsis thaliana* peroxigonase 2	27.9	8	15	72	39.1	21
3	Csa08g007170.1	voltage dependent anion channel 3	29.3	2	4	71	59.4	21
3	Csa15g001200.1	voltage dependent anion channel 1	29.5	3	5	72	46.7	21

`

3	Csa08g057250.1	Seed gene 3	22.6	6	14	70	33.5	20
3	Csa13g017920.1	voltage dependent anion channel 3	29.4	2	4	68	59.4	20
3	Csa19g031870.2	Vic1-1-G2	58.4	6	7	66	35.1	19
3	Csa11g019460.1	Ole-4-G1	19.7	2	3	67	32.6	19
3	Csa01g001220.1	voltage dependent anion channel 1	29.6	2	3	62	46.7	18
3	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	63	23.6	18
3	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	3	3	63	47.7	18
3	Csa01g025880.1	Vic1-3-G3	56.0	5	6	59	33.2	17
3	Csa12g024720.1	Cs2S-4-G3	36.6	1	1	52	32.5	15
3	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	4	5	53	57.9	15
3	Csa20g021570.1	voltage dependent anion channel 3	29.3	5	6	53	60.5	15
3	Csa00532s200.1	Oleosin family protein	21.1	1	2	48	22.3	14
3	Csa03g052870.1	1-cysteine peroxiredoxin 1	24.0	2	2	44	49.5	13
3	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	2	3	42	40.8	12
3	Csa02g075160.1	voltage dependent anion channel 2	29.7	1	1	42	55.1	12
3	Csa18g039940.1	voltage dependent anion channel 2	33.4	2	2	38	49.5	11
3	Csa11g017000.1	Cs2S-4-G1	36.6	2	2	39	26.8	11
3	Csa02g026890.1	hydroxysteroid dehydrogenase 5	42.9	2	3	38	40.8	11
3	Csa10g025440.1	RAB GTPase homolog A1D	23.9	4	4	33	48.6	10
3	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	10	10	35	37.7	10
3	Csa02g072050.1	gamma carbonic anhydrase like 1	27.6	9	9	35	45.1	10
3	Csa11g017020.1	Cs2S-1-G1	18.8	3	4	36	31.7	10
3	Csa10g049280.1	prohibitin 3	30.4	5	7	29	42.2	8
3	Csa18g005520.1	Protein of unknown function (DUF1264)	27.8	6	7	26	50.6	8
3	Csa15g039300.1	Vic1-2-G1	52.4	9	10	29	29.4	8
3	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	10	10	26	16.8	8
3	Csa17g006960.1	CRD-1-G1	50.5	1	1	26	27.2	8
3	Csa14g020590.1	proteasome subunit PAB1	25.7	8	9	27	45.1	8
3	Csa19g022460.1	Late embryogenesis abundant protein (LEA) family protein	23.0	1	1	29	41.1	8
3	Csa11g103350.1	voltage dependent anion channel 2	29.7	2	2	44	55.1	8
3	Csa08g053790.1	NAD(P)-binding Rossmann-fold superfamily protein	38.6	2	2	25	38	7
3	Csa07g027910.1	Unknown protein	26.4	2	3	23	33.6	7
3	Csa11g066400.1	RAB GTPase homolog A1C	23.9	2	2	24	39.8	7
3	Csa01g018300.1	Late embryogenesis abundant protein (LEA) family protein	25.3	2	2	24	37.3	7
3	Csa13g011500.1	actin 7	41.7	5	5	21	35	6
3	Csa13g013490.1	NAD(P)-binding Rossmann-fold superfamily protein	31.1	1	1	22	32.5	6
3	Csa03g005870.1	AWPM-19-like family protein	19.7	3	4	19	17.1	6
3	Csa05g067280.1	Unknown protein	36.5	2	3	22	29.9	6

`

3	Csa10g028320.1	RAB GTPase homolog B1C	54.3	3	4	19	21	6
3	Csa01g004900.1	glyceraldehyde-3-phosphate dehydrogenase C subunit 1	36.7	0	0	20	38.4	6
3	Csa12g004770.1	RmlC-like cupins superfamily protein	61.4	3	4	22	20	6
3	Csa05g060730.1	Glycoprotei membrane precursor GPI-anchored	21.7	2	2	21	26.2	6
3	Csa07g040360.1	Aquaporin-like superfamily protein	27.9	3	4	16	12.5	5
3	Csa01g027980.1	Plastid-lipid associated protein PAP / fibrillin family protein	29.8	5	5	18	26.1	5
3	Csa10g004530.1	RmlC-like cupins superfamily protein	58.6	2	2	17	16.7	5
3	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	6	6	13	20.9	4
3	Csa02g041750.1	Ole-3-G3	15.0	2	4	15	32.9	4
3	Csa03g006900.1	Calcium-dependent lipid-binding (CaLB domain) family protein	27.3	4	4	13	6.69	4
3	Csa03g019850.1	heat shock protein 70B	58.2	2	2	13	10.1	4
3	Csa19g031320.1	20S proteasome alpha subunit C1	27.4	4	4	13	17.6	4
3	Csa20g068880.1	20S proteasome subunit PAA2	27.3	4	4	14	30.9	4
3	Csa08g054170.1	RAB GTPase homolog A4A	24.9	2	2	13	21.1	4
3	Csa11g044960.1	proteasome alpha subunit A1	32.4	3	3	11	20.3	3
3	Csa02g004530.1	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	24.5	5	5	11	34.6	3
3	Csa04g029010.1	HSP20-like chaperones superfamily protein	17.7	3	3	9	36.9	3
3	Csa03g050970.1	gamma carbonic anhydrase 2	30.1	11	13	31	46	3
3	Csa15g016350.1	cystatin B	25.9	2	2	3	15.2	3
3	Csa02g070290.1	Embryo-specific protein 3, (ATS3)	23.0	3	3	11	26.8	3
3	Csa04g049020.1	ubiquitin 7	14.7	2	2	5	16.2	3
3	Csa13g036140.1	Ribosomal protein L6 family	22.0	8	9	12	57.7	3
3	Csa03g011110.1	Histone superfamily protein	11.4	4	4	10	26.3	3
3	Csa05g042070.1	20S proteasome alpha subunit G1	32.3	5	5	12	20.3	3
3	Csa05g030650.1	glutathione S-transferase PHI 9	23.9	2	3	9	7.95	3
3	Csa01g021740.1	RAB GTPase homolog G3F	23.1	1	1	12	21.8	3
3	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	2	2	7	17.7	2
3	Csa09g042380.1	HSP20-like chaperones superfamily protein	17.5	2	2	8	26.5	2
3	Csa09g096900.1	Saposin-like aspartyl protease family protein	55.7	2	3	8	5.45	2
3	Csa15g002220.1	Ribosomal protein S7e family protein	22.1	3	3	6	25.3	2
3	Csa15g023060.1	Adenine nucleotide alpha hydrolases-like superfamily protein	17.8	3	3	7	28.8	2
3	Csa06g025990.1	Stress induced protein	27.9	3	3	6	15.7	2
3	Csa03g023370.1	dehydroascorbate reductase	23.7	3	3	7	15	2
3	Csa18g023590.1	heat shock protein 90.1	81.1	1	1	8	5.25	2
3	Csa07g052280.1	Lipase/lipooxygenase, PLAT/LH2 family protein	20.3	2	2	6	7.97	2
3	Csa05g009000.1	late embryogenesis abundant domain-containing protein / LEA	69.3	1	1	8	6.15	2

`

		domain-containing protein						
3	Csa07g039460.1	40s ribosomal protein SA	36.1	3	3	8	13.1	2
3	Csa10g016280.1	heat shock protein 81-2	42.2	1	1	7	11.4	2
3	Csa05g063230.1	20S proteasome beta subunit G1	27.6	4	5	8	27.8	2
3	Csa14g014760.1	aspartic proteinase A1	54.5	1	1	6	13	2
3	Csa10g016690.1	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	26.5	2	2	4	10.5	1
3	Csa17g098210.1	Glycoprotein membrane precursor GPI-anchored	21.0	2	2	20	29.9	1
3	Csa04g051630.1	SOUL heme-binding family protein	25.1	1	1	2	1.68	1
3	Csa02g037960.1	arginosuccinate synthase family	77.6	2	2	5	3.85	1
3	Csa01g011890.1	manganese superoxide dismutase 1	27.7	2	2	4	14.7	1
3	Csa19g021970.1	Protein phosphatase 2C family protein	31.5	3	3	4	10.7	1
3	Csa04g015050.1	Cytochrome C1 family	38.7	2	2	4	4.2	1
3	Csa10g002470.1	Lipase/lipooxygenase, PLAT/LH2 family protein	20.1	2	2	5	14.9	1
3	Csa06g025420.1	20S proteasome alpha subunit PAD1	27.3	3	3	4	15.7	1
3	Csa01g007030.1	Glycoprotein membrane precursor GPI-anchored	22.7	2	2	3	13.3	1
3	Csa04g041110.1	Papain family cysteine protease	33.9	2	2	3	7.78	1
3	Csa04g041700.1	triosephosphate isomerase	27.2	3	3	5	14.6	1
3	Csa13g044730.1	S18 ribosomal protein	17.6	2	2	3	11.6	1
3	Csa00751s020.1	Ribosomal protein large subunit 16A	20.8	2	2	3	13.3	1
3	Csa02g005250.1	secretion-associated RAS super family 2	22.0	2	2	3	17.1	1
3	Csa04g060640.1	Ribosomal protein L3 family protein	29.4	2	2	4	10.3	1
3	Csa03g053840.1	Oleosin family protein	18.4	2	2	4	15.4	1
3	Csa16g043300.1	dehydrin LEA	21.6	2	2	4	32.7	1
3	Csa01g006090.1	NAD(P)-binding Rossmann-fold superfamily protein	31.3	2	2	4	6.92	1
3	Csa01g016850.1	20S proteasome alpha subunit E2	26.0	2	2	3	11.8	1
3	Csa18g012400.1	B-cell receptor-associated 31-like	24.6	2	2	3	8.72	1
3	Csa05g007000.1	Ribosomal L38e protein family	10.8	2	2	2	26.9	1
3	Csa13g018730.1	NAD(P)-binding Rossmann-fold superfamily protein	29.2	3	3	4	15.8	1
4	Csa04g015780.1	Ole-4-G1	20.5	5	18	501	45.1	139
4	Csa06g008780.1	Ole-4-G2	20.5	6	13	468	49.2	130
4	Csa10g047190.1	Ole-2-G2	21.3	11	23	200	43	56
4	Csa12g028090.1	Ole-1-G3	19.6	4	15	196	35.4	54
4	Csa11g019460.1	Ole-1-G1	19.7	4	7	156	35.4	43
4	Csa12g021990.1	CRC-1-G3	55.4	13	21	118	62.8	33
4	Csa11g070580.1	CRA-1-G1	52.0	3	7	109	44	30
4	Csa18g009670.1	CRA-1-G2	51.9	3	8	124	44.3	30
4	Csa10g014100.1	CRC-1-G2	54.0	2	4	99	45.1	28
4	Csa11g015240.1	CRC-1-G1	72.9	2	4	93	32.3	26

`

4	Csa17g006950.1	CRB-1-G1	50.7	6	12	89	50.4	25
4	Csa00532s200.1	Oleosin family protein	21.1	2	4	91	32.5	25
4	Csa11g017470.1	Seed gene 1	37.8	3	4	81	38.9	23
4	Csa14g004960.1	CRB-1-G2	64.9	1	1	82	39.6	23
4	Csa10g016060.1	Seed gene 1	28.0	4	4	84	51.4	23
4	Csa26607s010.1	Oleosin family protein	8.3	1	1	77	72.2	21
4	Csa02g057710.1	*Arabidopsis thaliana* peroxigonase 2	27.9	7	15	67	34.2	19
4	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	1	1	60	48.9	17
4	Csa16g016260.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	61	60.2	17
4	Csa11g082030.1	hydroxysteroid dehydrogenase 1	39.2	1	1	60	48.9	17
4	Csa07g015700.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	61	60.2	17
4	Csa05g035620.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	61	60.2	17
4	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	60	19.4	17
4	Csa08g057250.1	Seed gene 3	22.6	6	12	58	33.5	16
4	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	4	4	56	48.9	16
4	Csa10g028320.1	RAB GTPase homolog B1C	54.3	7	9	58	34.8	16
4	Csa12g007580.1	GTP-binding 2	23.1	5	7	46	75.4	13
4	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	46	19.2	13
4	Csa13g036140.1	Ribosomal protein L6 family	22.0	12	15	42	66.5	12
4	Csa03g055130.1	RAB GTPase homolog G3E	23.2	6	7	39	65.7	11
4	Csa06g050950.1	Ras-related small GTP-binding family protein	23.2	5	5	40	53.4	11
4	Csa01g021740.1	RAB GTPase homolog G3F	23.1	5	6	35	48.3	10
4	Csa05g067280.1	Unknown protein	36.5	3	6	31	34.4	9
4	Csa02g041750.1	Ole-3-G3	15.0	2	7	31	32.9	9
4	Csa04g029010.1	HSP20-like chaperones superfamily protein	17.7	4	6	27	42.7	8
4	Csa11g031130.1	RAB GTPase homolog 1C	22.3	2	2	28	50	8
4	Csa13g012020.1	RAB GTPase homolog H1E	23.2	3	3	30	44	8
4	Csa11g017020.1	Cs2S-1-G1	18.8	2	2	28	31.7	8
4	Csa15g001200.1	voltage dependent anion channel 1	29.5	2	2	28	38.8	8
4	Csa08g007170.1	voltage dependent anion channel 3	29.3	1	2	28	37	8
4	Csa19g031870.2	Vic1-1-G2	58.4	3	3	28	22.7	8
4	Csa13g017920.1	voltage dependent anion channel 3	29.4	2	4	30	47.1	8
4	Csa07g027910.1	Unknown protein	26.4	2	5	26	36.5	7
4	Csa01g025880.1	Vic1-3-G3	56.0	1	1	24	17.3	7
4	Csa02g076390.1	Dehydrin family protein	18.1	3	8	21	29.1	6
4	Csa18g002640.1		22.3	1	1	22	41.6	6
4	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	3	3	23	47.7	6
4	Csa03g005870.1	AWPM-19-like family protein	19.7	4	5	23	19.8	6
4	Csa20g021570.1	voltage dependent anion channel 3	29.3	1	1	22	29.3	6

`

4	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	6	6	20	9.07	6
4	Csa03g001060.1	RAS 5	22.6	1	1	22	27.6	6
4	Csa05g060730.1	Glycoprotein membrane precursor GPI-anchored	21.7	2	2	19	26.2	5
4	Csa06g054270.1	glutathione S-transferase phi 8	24.2	1	1	18	28.6	5
4	Csa11g003030.1	RAB GTPase homolog H1C	23.4	2	2	19	27.8	5
4	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	2	2	19	23.8	5
4	Csa15g039300.1	Vic1-2-G1	52.4	4	5	19	18.6	5
4	Csa12g004770.1	RmlC-like cupins superfamily protein	61.4	2	3	19	18.1	5
4	Csa19g021730.1	RAB GTPase homolog A1G	24.2	2	2	13	35	4
4	Csa03g052870.1	1-cysteine peroxiredoxin 1	24.0	2	2	16	39.4	4
4	Csa09g042380.1	HSP20-like chaperones superfamily protein	17.5	3	3	14	35.5	4
4	Csa06g025990.1	Stress induced protein	27.9	6	8	15	26	4
4	Csa18g005520.1	Protein of unknown function (DUF1264)	27.8	3	4	14	30.8	4
4	Csa18g042170.1	Dehydrin family protein	17.9	2	4	13	26.7	4
4	Csa17g070710.2	RAB GTPASE HOMOLOG B18	23.5	4	4	15	34.4	4
4	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	7	7	14	24.3	4
4	Csa17g006960.1	CRD-1-G1	50.5	1	1	11	12.3	4
4	Csa10g004530.1	RmlC-like cupins superfamily protein	58.6	2	2	16	16.7	4
4	Csa15g079170.1	RmlC-like cupins superfamily protein	85.0	3	4	15	6.6	4
4	Csa13g011500.1	actin 7	41.7	3	3	11	19.6	3
4	Csa11g027670.1		21.9	3	3	11	34.5	3
4	Csa15g002220.1	Ribosomal protein S7e family protein	22.1	3	3	12	24.7	3
4	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	5	5	11	14.1	3
4	Csa05g016580.1	Ribosomal protein 5B	28.0	2	2	11	8.7	3
4	Csa11g012100.1	N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein	25.2	4	4	10	18.5	3
4	Csa13g016800.1	translocon-associated protein beta (TRAPB) family protein	21.1	3	5	10	32.3	3
4	Csa04g011860.1	N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein	34.0	5	5	9	22.1	3
4	Csa10g016280.1	heat shock protein 81-2	42.2	3	3	12	18.5	3
4	Csa02g070290.1	Embryo-specific protein 3, (ATS3)	23.0	2	2	10	20.2	3
4	Csa03g011110.1	Histone superfamily protein	11.4	4	4	10	26.3	3
4	Csa14g027940.1	RAB GTPase homolog G3B	23.2	2	2	6	15	2
4	Csa08g003200.1	17.6 kDa class II heat shock protein	17.5	4	4	6	34.4	2
4	Csa10g017890.1		32.6	3	3	7	14.6	2
4	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	3	3	8	23.8	2
4	Csa08g005300.1	Ubiquinol-cytochrome C reductase iron-sulfur subunit	29.9	1	1	6	14.5	2
4	Csa07g040360.1	Aquaporin-like superfamily protein	27.9	2	2	7	7.92	2
4	Csa03g019850.1	heat shock protein 70B	58.2	2	2	8	10.3	2

`

4	Csa10g049280.1	prohibitin 3	30.4	2	2	6	19.1	2
4	Csa11g060020.1	RNA binding	30.6	2	2	8	9.96	2
4	Csa02g005250.1	Plectin/S10 domain-containing protein secretion-associated	22.0	3	3	8	27.9	2
4	Csa10g015740.1	RAS super family 2	15.5	3	3	8	40.3	2
4	Csa01g013010.1	3-ketoacyl-acyl carrier protein synthase I	21.7	2	2	6	11.2	2
4	Csa14g004970.1	Adenine nucleotide alpha hydrolases-like superfamily protein	50.1	1	1	8	12.4	2
4	Csa05g009000.1	CRD-1-G2	69.3	2	2	8	11.4	2
4	Csa03g059660.1	late embryogenesis abundant domain-containing protein / LEA domain-containing protein	18.2	2	2	8	33.7	2
4	Csa03g053840.1	Late embryogenesis abundant protein (LEA) family protein	18.4	2	2	6	15.4	2
4	Csa07g004060.1	Oleosin family protein	24.6	2	2	6	9.87	2
4	Csa08g002830.1	N-terminal nucleophile aminohydrolases (Ntn hydrolases) superfamily protein	23.1	2	2	5	16.9	1
4	Csa05g095530.1	FUNCTIONS IN: molecular_function unknown	30.8	4	4	6	8.02	1
4	Csa04g041530.1	Chalcone-flavanone isomerase family protein	17.5	4	4	8	28.6	1
4	Csa01g011890.1	Ribosomal protein L23AB	27.7	2	2	3	14.7	1
4	Csa19g021970.1	manganese superoxide dismutase 1	31.5	2	2	3	7.61	1
4	Csa15g023060.1	Protein phosphatase 2C family protein	17.8	2	2	3	17.8	1
4	Csa19g058160.1	Adenine nucleotide alpha hydrolases-like superfamily protein	91.3	3	3	5	4.86	1
4	Csa13g044730.1	RmlC-like cupins superfamily protein	17.6	2	2	3	11.6	1
4	Csa13g047350.1	S18 ribosomal protein	22.0	2	2	4	11.9	1
4	Csa00751s020.1	HSP20-like chaperones superfamily protein	20.8	2	2	4	13.3	1
4	Csa01g011530.1	Ribosomal protein large subunit 16A	15.9	2	3	4	15.7	1
4	Csa13g006790.1	Ribosomal S17 family protein	28.7	2	2	2	13.4	1
5	Csa12g028090.1	Nuclear transport factor 2 (NTF2) family protein	19.6	4	18	278	35.4	92
5	Csa04g015780.1	Ole-1-G3	20.5	5	15	250	45.1	83
5	Csa06g008780.1	Ole-4-G1	20.5	5	11	229	45.1	76
5	Csa11g019460.1	Ole-4-G2	19.7	4	9	217	35.4	72
5	Csa12g021990.1	Ole-1-G1	55.4	9	22	187	50	62
5	Csa10g014100.1	CRC-1-G3	54.0	5	9	152	44.9	50
5	Csa11g070580.1	CRC-1-G2	52.0	4	8	148	56	49
5	Csa18g009670.1	CRA-1-G1	51.9	3	9	169	51.8	49
5	Csa00532s200.1	CRA-1-G2	21.1	2	4	135	34	45
5	Csa10g047190.1	Oleosin family protein	21.3	9	18	134	39.5	44
5	Csa11g015240.1	Ole-2-G2	72.9	2	5	126	31	42
5	Csa26607s010.1	CRC-1-G1	8.3	1	1	111	83.3	37
5	Csa14g004960.1	Oleosin family protein	64.9	2	3	103	47.8	34
5	Csa17g006950.1	CRB-1-G2	50.7	6	11	94	48.9	31
		CRB-1-G1						

`

5	Csa05g035620.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	71	71.5	23
5	Csa16g016260.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	71	71.5	23
5	Csa07g015700.1	Pathogenesis-related thaumatin superfamily protein	26.7	1	1	71	71.5	23
5	Csa11g017470.1	Seed gene 1	37.8	3	3	62	38.6	21
5	Csa10g016060.1	Seed gene 1	28.0	3	3	61	51	20
5	Csa18g021370.1	hydroxysteroid dehydrogenase 1	39.1	1	1	41	44.6	14
5	Csa11g082030.1	hydroxysteroid dehydrogenase 1	39.2	1	1	39	44.6	13
5	Csa07g014960.1	HSP20-like chaperones superfamily protein	18.3	2	3	36	48.1	12
5	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	37	19.4	12
5	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	2	2	32	40.3	11
5	Csa04g029010.1	HSP20-like chaperones superfamily protein	17.7	5	7	33	48.4	11
5	Csa02g041750.1	Ole-3-G3	15.0	2	6	34	32.9	11
5	Csa16g015480.1	HSP20-like chaperones superfamily protein	17.9	2	2	32	49.7	11
5	Csa09g042380.1	HSP20-like chaperones superfamily protein	17.5	3	4	27	40.6	9
5	Csa11g031130.1	RAB GTPase homolog 1C	22.3	2	2	28	57.9	9
5	Csa10g029650.1	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	18.3	1	1	28	39.3	9
5	Csa17g006960.1	CRD-1-G1	50.5	1	1	23	21.7	9
5	Csa02g057710.1	*Arabidopsis thaliana* peroxygonase 2	27.9	5	8	25	30.9	8
5	Csa14g008800.1	HSP20-like chaperones superfamily protein	18.3	3	5	23	32.3	8
5	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	21	19.2	7
5	Csa14g004970.1	CRD-1-G2	50.1	1	1	21	21.9	7
5	Csa04g038130.1	ATP synthase D chain, mitochondrial	19.5	3	4	18	41.7	6
5	Csa17g070710.2	RAB GTPASE HOMOLOG B18	23.5	5	5	17	39.6	6
5	Csa13g047050.1	hydroxysteroid dehydrogenase 5	42.8	1	1	17	17.8	6
5	Csa04g041530.1	Ribosomal protein L23AB	17.5	5	5	15	29.2	5
5	Csa07g027910.1	Unknown protein	26.4	2	2	16	33.2	5
5	Csa10g001480.1	rotamase CYP 1	18.6	3	4	14	23.7	5
5	Csa01g025880.1	Vic1-3-G3	56.0	1	1	14	15.1	5
5	Csa04g039480.1	Ribosomal protein L11 family protein	18.0	2	3	15	31.9	5
5	Csa01g011530.1	Ribosomal S17 family protein	15.9	4	5	16	24.3	5
5	Csa15g079170.1	RmlC-like cupins superfamily protein	85.0	3	4	16	6.6	5
5	Csa01g021740.1	RAB GTPase homolog G3F	23.1	4	4	14	34.9	5
5	Csa11g060020.1	RNA binding Plectin/S10 domain-containing protein	30.6	4	4	14	17.7	5
5	Csa08g057250.1	Seed gene 3	22.6	4	4	12	22.5	4
5	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	2	2	12	35.6	4
5	Csa05g067280.1	Unknown protein	36.5	1	1	12	15.1	4
5	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	5	5	11	20.2	4

`

5	Csa12g004770.1	RmlC-like cupins superfamily protein	61.4	2	3	11	12.6	4
5	Csa02g070610.1	HVA22 homologue B	18.7	3	3	12	32.7	4
5	Csa15g001200.1	voltage dependent anion channel 1	29.5	2	2	13	30.4	4
5	Csa13g011500.1	actin 7	41.7	2	2	9	18.8	3
5	Csa19g007120.1	Ribosomal protein S24e family protein	15.4	4	4	8	30.8	3
5	Csa01g001220.1	voltage dependent anion channel 1	29.6	1	1	8	25.4	3
5	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	3	3	8	19.2	3
5	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	5	5	9	13.6	3
5	Csa19g058160.1	RmlC-like cupins superfamily protein	91.3	3	3	8	5.91	3
5	Csa12g007580.1	GTP-binding 2	23.1	1	1	10	33.6	3
5	Csa07g051800.1	cold, circadian rhythm, and rna binding 2	16.7	2	2	8	21.4	3
5	Csa13g044730.1	S18 ribosomal protein	17.6	2	2	9	16.3	3
5	Csa03g053840.1	Oleosin family protein	18.4	2	3	8	15.4	3
5	Csa03g011110.1	Histone superfamily protein	11.4	3	3	9	19.7	3
5	Csa08g001990.1	Cystathionine beta-synthase (CBS) family protein	22.9	3	3	10	23.2	3
5	Csa13g016800.1	translocon-associated protein beta (TRAPB) family protein	21.1	3	4	8	32.3	3
5	Csa11g099340.1	HVA22 homologue B	18.7	2	2	8	32.7	3
5	Csa14g027940.1	RAB GTPase homolog G3B	23.2	2	2	6	15	2
5	Csa08g003200.1	17.6 kDa class II heat shock protein	17.5	3	3	7	33.8	2
5	Csa08g062980.1	PEBP (phosphatidylethanolamine-binding protein) family protein	17.9	4	4	6	38.9	2
5	Csa08g005830.1	nascent polypeptide-associated complex subunit alpha-like protein 3	22.0	2	2	6	13.4	2
5	Csa15g023060.1	Adenine nucleotide alpha hydrolases-like superfamily protein	17.8	3	3	5	28.8	2
5	Csa00506s140.1	Ribosomal L27e protein family	15.5	2	2	5	14.1	2
5	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	4	4	5	7.1	2
5	Csa05g094290.1	rotamase CYP 4	48.1	2	2	6	7.98	2
5	Csa10g015740.1	3-ketoacyl-acyl carrier protein synthase I	15.5	3	3	6	31.3	2
5	Csa15g039300.1	Vic1-2-G1	52.4	3	3	5	13	2
5	Csa13g047350.1	HSP20-like chaperones superfamily protein	22.0	3	3	5	12.9	2
5	Csa00751s020.1	Ribosomal protein large subunit 16A	20.8	3	3	6	17.7	2
5	Csa10g004530.1	RmlC-like cupins superfamily protein	58.6	1	1	6	9.04	2
5	Csa01g007060.1	Peroxiredoxin IIF	21.5	4	4	6	28.4	2
5	Csa04g066330.1	ADP-ribosylation factor A1E	20.6	3	3	7	24.3	2
5	Csa07g065640.1	P-loop containing nucleoside triphosphate hydrolases superfamily protein	92.3	2	2	5	3.14	2
5	Csa08g055140.1	Ribosomal protein S19 family protein	34.1	3	3	7	15.8	2
5	Csa02g019830.1	glutathione peroxidase 6	25.6	2	2	4	8.7	1
5	Csa03g011430.1	Expressed protein	16.2	2	2	3	30.1	1

`

5	Csa01g041670.1	RmlC-like cupins superfamily protein	85.2	2	2	11	4.76	1
5	Csa03g005870.1	AWPM-19-like family protein	19.7	2	3	4	13.4	1
5	Csa02g005590.1	Protein of unknown function, DUF538	16.9	2	2	3	13.5	1
5	Csa04g035480.1	Translation protein SH3-like family protein	16.9	2	2	4	17.8	1
5	Csa03g059660.1	Late embryogenesis abundant protein (LEA) family protein	18.2	2	2	4	27.2	1
5	Csa02g062630.1	temperature-induced lipocalin	21.7	1	1	2	16.9	1
5	Csa07g061190.1	copper ion binding	11.9	2	2	4	25.9	1
6	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	37	14.3	148
6	Csa12g021990.1	CRC-1-G3	55.4	4	5	20	21.5	80
6	Csa10g014100.1	CRC-1-G2	54.0	1	1	18	19.3	72
6	Csa10g015740.1	3-ketoacyl-acyl carrier protein synthase I	15.5	3	5	18	36.6	72
6	Csa08g057250.1	Seed gene 3	22.6	4	4	14	22	56
6	Csa03g059740.1	Histone H2A protein 9	14.3	3	4	13	29.9	52
6	Csa00532s200.1	Oleosin family protein	21.1	1	1	11	19.3	44
6	Csa05g007000.1	Ribosomal L38e protein family	10.8	2	3	10	26.9	40
6	Csa12g028090.1	Ole-1-G3	19.6	3	3	10	28.2	40
6	Csa10g047190.1	Ole-2-G2	21.3	2	2	9	20	36
6	Csa03g024410.1	Small nuclear ribonucleoprotein family protein	14.1	2	3	7	14.7	28
6	Csa19g002520.1	Unknown protein	5.6	2	2	6	59.3	24
6	Csa11g072130.1	sterol carrier protein 2	13.6	2	2	3	17.1	12
6	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	2	2	3	2.89	12
7	Csa12g024730.1	Cs2S-2-G3	36.1	2	5	87	20.4	234
7	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	75	20.5	202
7	Csa11g017020.1	Cs2S-1-G1	18.8	4	5	60	42.7	162
7	Csa03g011110.1	Histone superfamily protein	11.4	8	9	30	40.1	81
7	Csa00532s200.1	Oleosin family protein	21.1	1	1	22	22.3	59
7	Csa10g047190.1	Ole-2-G2	21.3	3	4	21	27	57
7	Csa06g008780.1	Ole-4-G2	20.5	2	3	20	33.2	54
7	Csa04g015780.1	Ole-4-G1	20.5	1	1	19	32.1	51
7	Csa11g015240.1	CRC-1-G1	72.9	1	1	15	12.7	40
7	Csa12g021990.1	CRC-1-G3	55.4	2	3	14	11	38
7	Csa12g028090.1	Ole-1-G3	19.6	2	2	12	18.8	32
7	Csa18g009670.1	CRA-1-G2	51.9	1	1	9	14.2	24
7	Csa02g041750.1	Ole-3-G3	15.0	2	3	8	32.9	22
7	Csa10g007580.1	Ribosomal protein S25 family protein	12.0	2	2	8	24.1	22
7	Csa04g038960.1	Ribosomal protein L24e family protein	18.6	2	2	4	6.34	11
8	Csa12g024730.1	Cs2S-2-G3	36.1	2	5	180	21	477
8	Csa11g017000.1	Cs2S-4-G1	36.6	1	1	143	21.1	379
8	Csa11g017020.1	Cs2S-1-G1	18.8	3	3	113	43.9	299
8	Csa03g011110.1	Histone superfamily protein	11.4	6	8	27	34.9	71

`

8	Csa06g008780.1	Ole-4-G2	20.5	1	1	11	32.1	29
8	Csa12g028090.1	Ole-1-G3	19.6	2	2	11	9.39	29
8	Csa02g041750.1	Ole-3-G3	15.0	2	3	6	32.9	16
8	Csa10g047190.1	Ole-2-G2	21.3	1	1	4	15	11
8	Csa12g021990.1	CRC-1-G3	55.4	2	2	4	7.93	11
9	Csa04g015780.1	Ole-4-G1	20.5	5	16	241	45.1	200
9	Csa06g008780.1	Ole-4-G2	20.5	5	10	224	45.1	186
9	Csa10g047190.1	Ole-2-G2	21.3	10	22	167	43	139
9	Csa12g028090.1	Ole-1-G3	19.6	4	15	162	35.4	134
9	Csa00532s200.1	Oleosin family protein	21.1	2	3	86	34	71
9	Csa26607s010.1	Oleosin family protein	8.3	0	0	68	70.8	37
9	Csa11g015240.1	CRC-1-G1	72.9	1	1	43	17.3	36
9	Csa10g014100.1	CRC-1-G2	54.0	3	3	39	30.2	32
9	Csa12g024730.1	Cs2S-2-G3	36.1	2	4	35	14.3	29
9	Csa12g021990.1	CRC-1-G3	55.4	6	7	32	30.3	27
9	Csa05g020560.1	Pollen Ole e 1 allergen and extensin family protein	19.1	2	3	21	46.9	20
9	Csa02g041750.1	Ole-3-G3	15.0	2	6	20	32.9	17
9	Csa04g041530.1	Ribosomal protein L23AB	17.5	5	6	19	31.2	16
9	Csa11g060020.1	RNA binding Plectin/S10 domain-containing protein	30.6	5	5	18	20.7	15
9	Csa18g009670.1	CRA-1-G2	51.9	2	2	20	23.4	14
9	Csa11g019460.1	Ole-1-G1	19.7	4	10	128	35.4	12
9	Csa11g070580.1	CRA-1-G1	52.0	1	1	15	23.3	12
9	Csa05g021500.1	arabinogalactan protein 30	27.1	2	4	13	5.06	11
9	Csa03g053840.1	Oleosin family protein	18.4	3	4	10	27.2	8
9	Csa00751s020.1	Ribosomal protein large subunit 16A	20.8	2	2	5	12.2	4
9	Csa03g011110.1	Histone superfamily protein	11.4	2	2	4	14.5	3
9	Csa05g092580.1	Ribosomal protein S13/S15	17.1	1	1	1	2.28	1
10	Csa04g015780.1	Ole-4-G1	20.5	5	15	179	43.5	197
10	Csa06g008780.1	Ole-4-G2	20.5	5	9	171	43.5	188
10	Csa12g028090.1	Ole-1-G3	19.6	4	13	137	35.4	151
10	Csa10g047190.1	Ole-2-G2	21.3	9	16	136	43	150
10	Csa11g019460.1	Ole-1-G1	19.7	4	8	110	35.4	121
10	Csa00532s200.1	Oleosin family protein	21.1	2	4	82	32.5	90
10	Csa12g024730.1	Cs2S-2-G3	36.1	2	3	36	14.3	40
10	Csa11g017470.1	Seed gene 1	37.8	2	2	28	25	31
10	Csa02g057710.1	*Arabidopsis thaliana* peroxigonase 2	27.9	4	7	25	25.1	28
10	Csa10g016060.1	Seed gene 1	28.0	2	2	30	33.1	16
10	Csa12g021990.1	CRC-1-G3	55.4	2	3	12	4.47	13
10	Csa02g041750.1	Ole-3-G3	15.0	2	3	11	32.9	12
10	Csa04g041530.1	Ribosomal protein L23AB	17.5	3	3	10	22.1	11
10	Csa03g053840.1	Oleosin family protein	18.4	3	3	8	27.2	9

`

10	Csa03g011110.1	Histone superfamily protein	11.4	2	2	4	14.5	4
10	Csa01g021420.1	Oleosin family protein	18.1	2	2	2	13.2	2
11	Csa12g024730.1	Cs2S-2-G3	36.1	3	6	100	22.3	89
11	Csa18g009670.1	CRA-1-G2	51.9	3	4	78	52	69
11	Csa11g070580.1	CRA-1-G1	52.0	3	4	78	54.5	69
11	Csa11g015240.1	CRC-1-G1	72.9	1	2	66	22.8	59
11	Csa12g021990.1	CRC-1-G3	55.4	8	12	63	39.6	56
11	Csa10g014100.1	CRC-1-G2	54.0	2	2	63	35.2	56
11	Csa06g008780.1	Ole-4-G2	20.5	2	3	49	41.5	44
11	Csa04g015780.1	Ole-4-G1	20.5	2	2	50	41.5	44
11	Csa11g017020.1	Cs2S-1-G1	18.8	1	1	47	28	42
11	Csa03g011110.1	Histone superfamily protein	11.4	10	13	39	40.1	35
11	Csa00532s200.1	Oleosin family protein	21.1	2	2	39	31.5	35
11	Csa17g006950.1	CRB-1-G1	50.7	0	0	37	30.3	33
11	Csa03g001760.1	Late embryogenesis abundant protein	16.5	1	2	32	57.6	28
11	Csa10g029650.1	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	18.3	1	1	28	45.1	25
11	Csa02g041750.1	Ole-3-G3	15.0	3	9	27	36.4	24
11	Csa10g047190.1	Ole-2-G2	21.3	4	5	27	32.5	24
11	Csa12g053140.1	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	18.4	1	1	27	45.7	24
11	Csa12g028090.1	Ole-1-G3	19.6	4	5	27	34.3	24
11	Csa03g012310.1	Histone superfamily protein	14.8	2	4	26	41.5	23
11	Csa11g017470.1	Seed gene 1	37.8	2	2	25	28.4	22
11	Csa17g001940.1	Late embryogenesis abundant protein	16.6	1	1	24	41.1	21
11	Csa10g015740.1	3-ketoacyl-acyl carrier protein synthase I	15.5	6	8	24	51.5	21
11	Csa10g016060.1	Seed gene 1	28.0	1	1	24	29.8	21
11	Csa11g019460.1	Ole-1-G1	19.7	2	3	21	31.5	19
11	Csa08g057250.1	Seed gene 3	22.6	5	5	20	26.3	18
11	Csa10g001480.1	rotamase CYP 1	18.6	2	3	14	15	12
11	Csa07g051310.1	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	16.7	2	3	13	12.7	12
11	Csa01g023730.1	Histone H2A 13	13.9	4	7	32	43.9	11
11	Csa02g039290.1	hydroxysteroid dehydrogenase 1	39.2	1	1	12	18.9	11
11	Csa03g059740.1	Histone H2A protein 9	14.3	2	2	12	29.9	11
11	Csa01g011530.1	Ribosomal S17 family protein	15.9	2	3	10	15.7	9
11	Csa10g007580.1	Ribosomal protein S25 family protein	12.0	2	2	8	24.1	7
11	Csa02g067290.1	GTP binding Elongation factor Tu family protein	49.5	4	4	8	7.49	7
11	Csa11g082710.1	Ole-3-G1	15.1	2	2	7	25.2	6
11	Csa15g023060.1	Adenine nucleotide alpha hydrolases-like superfamily protein	17.8	3	4	7	31.3	6

`

11	Csa15g001200.1	voltage dependent anion channel 1	29.5	2	3	7	22.8	6
11	Csa14g053080.1	1-cysteine peroxiredoxin 1	24.0	1	1	6	15.3	5
11	Csa10g009990.1	Cytochrome bd ubiquinol oxidase, 14kDa subunit	14.6	2	2	6	17.2	5
11	Csa03g061590.1	Eukaryotic translation initiation factor 4A1	46.8	3	3	6	8.67	5
11	Csa14g009180.1	Ribosomal protein S8 family protein	14.8	2	2	5	8.46	4
11	Csa08g001390.1	Ribosomal protein S10p/S20e family protein	13.7	2	2	5	19.7	4
11	Csa05g060740.1	NAD(P)-binding Rossmann-fold superfamily protein	36.9	2	2	4	5.93	4
11	Csa02g005590.1	Protein of unknown function, DUF538	16.9	2	2	5	13.5	4
11	Csa14g009030.1	dessication-induced 1VOC superfamily protein	15.3677	2	2	3	24.1	3
11	Csa02g005250.1	secretion-associated RAS super family 2	21.9876	2	2	3	21.4	3

* Spot number refers to the protein spots as represented in Figure 4.18D.
** Annotations for *C. sativa* storage protein genes were assigned according to Table 4.3. The annotations for the *C. sativa* oleosins were assigned according to the location of the gene in the sub genome. The rest is according to the lowest probability obtained for each gene from the BLAST search preformed against the *Arabidopsis thaliana* genome using TAIR 8 database available from (https://www.arabidopsis.org/).
*** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A7. LC-MS/MS results showing 100% probable protein matches for the *B. napus* oil body proteins separated by 2D-electrophoresis using pH 3 to 10 IPG strips.

Spot number*	Protein name	Protein accession numbers	Molecular weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS**
1	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	7	51	22.6	48
1	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	4	4	20	15.1	19
1	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	7	9	30	18.9	28
1	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	9	21	189	46.3	178
1	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	14	79	35	74
1	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	5	8	58	27.5	55
1	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	8	22	101	43.6	95
1	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	2	12	22.2	11
2	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	12	54	39.3	66
2	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	7	20	65	43.6	80
2	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	3	3	7	9.46	9
2	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	4	5	25	15.9	31
2	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	6	9	51	29	63
2	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	25	197	48.4	242
3	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	2	2	8	10.2	8
3	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	7	13	66	35.2	64
3	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	9	22	106	44.6	104
3	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	28	243	48.4	237
3	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	3	3	10	9.46	10
3	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	16	77	39.3	75
3	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	2	7	22.2	7
4	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	6	11	53	29	68
4	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	7	21	80	43.6	103

`

4	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	7	14	53	39.3	68
4	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	25	200	48.4	257
5	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	2	2	14	10.2	15
5	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	4	5	15	12.5	16
5	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	2	2	6	12	7
5	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	8	25	110	44.6	120
5	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	7	13	66	35.2	72
5	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	7	15	73	39.3	80
5	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	9	27	202	46.3	220
6	No result was obtained							
7	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	7	7	31	24.2	36
7	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	5	6	41	38.3	48
7	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	3	4	20	13.9	24
7	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	3	4	16	13.8	19
7	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	3	12	22.2	16
7	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	3	5	14.5	6
7	Myrosinase OS=*Brassica napus* PE=2 SV=1	MYRO_BRANA	62.7	4	4	9	7.48	11
7	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	5	12	55	43.6	65
7	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	8	17	100	46.3	118
7	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	5	8	38	29	45
8	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	10	12	50	32.8	90
8	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	2	3	11	21.3	20
8	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	3	3	9	13.9	16
8	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	3	6	21	33.3	38
8	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	9	11	41	34.8	74
8	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	1	1	2	9.7	4
8	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	6	8	46	45.7	83

`

8	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	3	4	10	18.7	18
8	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	3	15	22.2	31
9	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	9	11	46	34	36
9	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	3	6	19.4	5
9	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	8	21	80	43.6	62
9	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	5	10	53	29	41
9	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	3	13	27.2	12
9	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	10	55	39.3	43
9	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	9	22	146	46.3	113
9	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	7	7	25	29	19
9	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	6	7	31	21.5	24
10	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	8	8	29	28.7	23
10	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	1	1	3	14.5	2
10	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	9	16	91	48.4	73
10	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	6	6	25	29.2	20
10	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	4	7	34	25.4	27
10	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	7	33	39.3	26
10	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	7	17	56	43.6	45
10	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	5	5	20	21.1	16

* Spot number refers to the protein spots as represented in Figure 4.19B.
** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.

`

Table A8. LC-MS/MS results showing 100% probable protein matches for the *B. napus* oil body proteins separated by 2D-electrophoresis using pH 9 to 12 IPG strips.

Spot number*	Protein name	Protein accession numbers	Molecular weight (kDa)	Exclusive unique peptide count	Exclusive unique spectrum count	Total spectrum count	% coverage	NTS**
1	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	3	11	22.2	16
1	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	5	5	19	23.9	27
1	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	21	105	48.4	151
1	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	9	11	36	28.7	52
1	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	3	7	19.4	10
1	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	6	7	25	24.7	36
1	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	8	30	39.3	43
1	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	6	13	40	43.6	58
1	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	6	11	48	29	69
2	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	6	6	28	30.1	26
2	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	7	13	58	35.2	54
2	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	11	15	44	24.5	41
2	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	7	8	39	28.2	36
2	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	2	2	6	16.7	6
2	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	7	16	69	39.3	64
2	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	28	207	48.4	193
2	Napin embryo-specific OS=*Brassica napus* PE=2 SV=1	2SSE_BRANA	21.0	2	3	5	14.5	5
2	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	7	21	92	43.6	86
3	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	5	5	29	19.1	33
3	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	7	13	65	39.3	73
3	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	1	1	8	20	9
3	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	3	6	25	11.4	28
3	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	9	24	213	46.3	239
3	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	3	3	7	8.4	8
3	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	7	23	100	43.6	112
3	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	6	11	60	29	67

`

4	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	2	2	5	10.2	7
4	Napin OS=*Brassica napus* GN=NAP1 PE=2 SV=1	2SS4_BRANA	20.3	1	1	7	14.4	10
4	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	25	178	48.4	260
4	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	7	22	81	43.6	118
4	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	7	12	53	39.3	77
4	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	6	10	51	29	75
4	Cruciferin BnC1 OS=*Brassica napus* GN=BnC1 PE=3 SV=2	CRU1_BRANA	53.8	4	4	11	13.1	16
4	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	2	2	3	6.9	4
5	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	3	4	19	12.8	22
5	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	11	27	216	48.9	255
5	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	7	14	65	35.2	77
5	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	2	2	4	5.4	5
5	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	9	28	116	44.6	137
5	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	8	15	75	40.4	88
6	Cruciferin CRU1 OS=*Brassica napus* GN=CRU1 PE=3 SV=1	CRU3_BRANA	56.5	2	2	3	4.5	4
6	Oleosin S2-2 OS=*Brassica napus* GN=S2 PE=1 SV=1	OLES2_BRANA	19.9	10	19	113	48.4	148
6	Oleosin Bn-III OS=*Brassica napus* PE=2 SV=1	OLEO3_BRANA	21.5	8	19	68	44.6	89
6	Cruciferin CRU4 OS=*Brassica napus* GN=CRU4 PE=1 SV=1	CRU4_BRANA	51.4	2	2	6	8.6	8
6	Oleosin S1-2 OS=*Brassica napus* GN=S1 PE=1 SV=1	OLES1_BRANA	20.7	5	9	37	29	48
6	Oleosin Bn-V (Fragment) OS=*Brassica napus* PE=2 SV=1	OLEO5_BRANA	20.3	6	11	45	39.3	55

* Spot number refers to the protein spots as represented in Figure 4.19D.
** NTS – Normalized Total Spectra. The parameter used to calculate the protein abundance in Scaffold 4 proteomic software. The NTS values are only comparable within an individual spot.