STRUCTURE AND PROPERTIES OF CRUCIFERIN: INVESTIGATION OF HOMOHEXAMERIC CRUCIFERIN EXPRESSED IN ARABIDOPSIS

A Thesis Submitted to the College of Graduate Studies and Research
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Department of Food and Bioproduct Sciences

University of Saskatchewan
Saskatoon

By
Thushan Sanjeewa Withana-Gamage

2013

© Copyright Thushan S. Withana-Gamage, June 2013. All rights reserved.
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 in part should be addressed to:

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

The structure of 11S cruciferin has been solved; however, how the individual subunits contribute to its physico-chemical and functional properties are not well known. The cruciferin isoforms in Arabidopsis thaliana, CRUA, CRUB, and CRUC, were investigated with respect to their molecular structures and the relationship of structural features to the physico-chemical and functional properties of cruciferin using homology modeling and various analytical techniques.

Comparison of these models revealed that hydrophobicity and electrostatic potential distribution on the surface of the CRUC homotrimer had more favorable interfacial, solubility, and thermal properties than those of CRUA or CRUB. Flavor binding and pepsin digestion were associated with hypervariable regions (HVRs) and center core regions, respectively, moreso for CRUA and CRUB homotrimers than for CRUC.

Chemical imaging of a single cell area in wild type (WT) and double-knockout seeds (CRUAbc, CRUaBc, and CRUabC) using synchrotron FT-IR microscopy (amide I band, 1650 cm⁻¹, νC=O) showed that seed storage proteins were concentrated in the cell center and protein storage vacuoles, whereas lipids were closer to the cell wall. Secondary structure components of proteins of double-knockout lines did not show major differences. Changes in protein secondary structure components of pepsin-treated CRUabC (CRUC) mutant were minimal, indicating low enzyme accessibility.

A three-step chromatographic procedure allowed isolation of the hexameric form of cruciferin with high purity (>95%). Fourier transform infrared (FT-IR) and circular dichroism (CD) spectroscopic analysis of the secondary structure of these proteins revealed cruciferins were folded into higher order secondary structures; 44−50% β-sheets and 7−9% α-helices. The relative subunit ratio was approximately 1:3:6 (CRUA:CRUB:CRUC) in the WT cruciferin. The $T_m$ values of purified cruciferin at pH 7.4 ($\mu = 0.0$) were in the order of WT = CRUA = CRUB < CRUC. The order of surface hydrophobicity as determined by ANS (1-anilinonaphthalene-8-sulfonate) probe binding was CRUA > CRUB = WT >> CRUC.
Intrinsic fluorescence studies revealed a compact molecular structure for the CRUC homohexamer compared to the CRUA and CRUB homohexamers. The order of emulsion forming abilities was CRUA = CRUB > WT > CRUC (no emulsion formation) and the order of heat-induced network structure strength was WT > CRUA = CRUB > CRUC (no gel formation). The inability of CRUC to form gels or emulsions may be attributed to its low surface hydrophobicity and molecular compactness. At pH 2.0, CRUC hexamers dissociated into trimers which allowed the formation of an O/W emulsion and heat-induced network structures.

Solubility of cruciferin as a function of pH at low ionic strength gave two minima around pH 4 and 7.4 yielding a “W” shape solubility profile deviating from the typical “U” or “V” shape solubility profile of other 11S globulins. The high ionic strength (µ = 0.5) was not favorable for emulsification, heat-induced gel formation, or solubilization for all cruciferins. Furthermore, the CRUA and CRUB homohexamers exhibited rapid pepsinolysis, while the CRUC homohexamer and WT heterohexamer were digested more slowly.

Although fairly well conserved regions were found in the primary structure of these three cruciferin subunits, differences were found in the hypervariable regions and extended loop regions resulting in slight differences in 3D structures and interactions that occur during association to form superstructures, such as hexamers. These differences were reflected in the physico-chemical and techno-functional properties of hexamers and trimers composed of each subunit. *In silico* predictions for certain functionalities were highly correlated with empirical data from laboratory experiments.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisors, Drs. Janitha P. D. Wanasundara and Xiao Qiu, for their support and guidance throughout my Ph.D. program. Their suggestions and thought-provoking conversations during the project and thesis preparation are greatly appreciated.

I would also like to thank the members of my advisory committee, Drs. Dwayne D. Hegedus, Robert T. Tyler, and Peiqiang Yu, and Graduate Chairs, Drs. Darren Korber, Phyllis J. Shand, and Takuji Tanaka, for their advice, knowledge sharing and assistance in many aspects during my study. My appreciation is also expressed to Dr. Rickey Y. Yada for serving as the External Examiner.

Special thanks are extended to Dr. Hegedus for willing support to the project, thesis preparation, and providing Arabidopsis seed materials. I extend my sincere appreciation to Tara McIntosh, Cathy Coutu, and Brad Hope of Agriculture and Agri-Food Canada, Saskatoon Research Centre and Tor Pedersen and Tim May at Canadian Light Source Inc., Jason Maley at Saskatchewan Structural Science Centre and Phil Dillman at Prairie Diagnostic Services Inc. of the University of Saskatchewan, Saskatoon for their technical assistance and support throughout my study. Drs. Michael Nickerson and Phyllis J. Shand are kindly acknowledged for allowing their laboratory facilities to use in my experiments.

Financial support provided by Agriculture and Agri-Food Canada research projects (RBPI 26 & 1827) and the Dr. S. Roger Rimmer Graduate Scholarship from SaskCanola (Saskatchewan Canola Development Commission) during my graduate program is gratefully acknowledged.

Special thanks go to my parents, my wife Chammi, my lovely little son Pravin, friends, and colleagues for their support and motivation throughout my graduate program.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERMISSION TO USE</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xvi</td>
</tr>
</tbody>
</table>

### 1. INTRODUCTION

1.1 Rationale                                                           | 1    |
1.2 Hypothesis                                                         | 2    |
1.3 Objectives                                                         | 2    |

### 2. LITERATURE REVIEW

2.1 *Brassicaceae* Family Plants                                      | 3    |
2.2 Composition of Brassica Seeds and Meals                          | 3    |
2.3 Storage proteins of Brassica seeds                               | 5    |
2.4 Biosynthesis of Cruciferin                                       | 6    |
2.5 Molecular Structure of 11−12S Globulins                           | 7    |
2.5.1 Primary Structure                                              | 7    |
2.5.2 Secondary Structure                                            | 9    |
2.5.3 Tertiary and Quaternary Structure                              | 9    |
2.6 Arabidopsis and Crucifer SSP                                      | 10   |
2.7 Extraction and Purification of Cruciferin                         | 12   |
2.8 Techniques for Protein Structure Probing                         | 15   |
2.8.1 *In Vitro* Spectroscopic Techniques                            | 15   |
2.8.2 *In Silico* Modeling Techniques                               | 19   |
2.9 Biological, Functional, and Physico-chemical Properties of 11S Globulins | 21 |
2.9.1 Physico-chemical Properties                                    | 22   |
5. STRUCTURE AND PHYSICO-CHEMICAL PROPERTY RELATIONSHIPS OF CRUCIFERIN HOMOHEXAMERS .................................................................................................................. 88

5.1 Abstract ........................................................................................................... 88

5.2 Introduction ...................................................................................................... 89

5.3 Materials and Methods .................................................................................... 90

5.3.1 Plant Materials ............................................................................................. 90

5.3.2 Isolation and Purification of Cruciferin ......................................................... 90

5.3.3 Protein Identification by Electrophoresis .................................................... 91

5.3.4 Microfluidic LabChip Electrophoresis ......................................................... 92

5.3.5 Protein Measurements .................................................................................. 93

5.3.6 Fourier Transform Infrared Spectroscopy (FT-IR) ....................................... 93

5.3.7 Circular Dichroism (CD) Spectroscopy ....................................................... 93

5.3.8 Zeta Potential ............................................................................................... 94

5.3.9 Intrinsic Fluorescence .................................................................................. 94

4.4 Results and Discussion .................................................................................... 71

4.4.1 Confirmation of Lines with Altered SSP Profiles ......................................... 71

4.4.2 IR Evidence for Biopolymer Localization in Parenchyma Cells ................... 73

4.4.3 Chemical Makeup of Seeds ......................................................................... 76

4.4.4 Protein Secondary Structure Features ....................................................... 79

4.4.5 Pepsin-Assisted Degradation of Tissue Proteins ......................................... 83

4.5 Conclusions ..................................................................................................... 87

4.6 Connections to Subsequent Chapters .............................................................. 87

4.3 Generation of Arabidopsis Lines with Altered SSP profiles ............................. 64

4.3.2 Seed Production ............................................................................................ 65

4.3.3 Contents of Lipid and Protein .................................................................... 65

4.3.4 DNA Extraction and Polymerase Chain Reaction (PCR) ............................. 67

4.3.5 One-Dimensional Gel Electrophoresis (1D-PAGE) ....................................... 67

4.3.6 Western Blotting ......................................................................................... 69

4.3.7 Sample Preparation and Synchrotron FT-IR Spectromicroscopy ................. 69

4.3.8 Enzyme Accessibility of Tissue Proteins .................................................... 70

4.3.9 Statistical Analysis ....................................................................................... 70
5.3.10 ANS Binding .................................................................................................................. 95
5.3.11 Thermal Properties ....................................................................................................... 96
5.3.12 Statistical Analysis ....................................................................................................... 96
5.4 Results and Discussion .................................................................................................... 97
5.4.1 Purification of Cruciferin from WT and Mutant Lines ................................................ 97
5.4.2 Cruciferin Subunit Composition .................................................................................. 97
5.4.3 Secondary Structure Components of Cruciferin ......................................................... 100
5.4.4 Electrical Potential of Cruciferins ............................................................................. 102
5.4.5 Intrinsic Fluorescence ................................................................................................. 104
5.4.6 Structural Stability in the Presence of the Co-solvents Urea, Guandine-HCl and NaCl or in Response to pH Change ......................................................................................... 111
5.4.7 Surface Hydrophobic (S0) .......................................................................................... 112
5.4.8 Thermal Transitions and Stability ............................................................................. 114
5.4.9 Contribution of Each Subunit Type to WT Cruciferin .............................................. 118
5.5 Conclusions .................................................................................................................... 118
5.6 Connections to Subsequent Chapters ............................................................................. 118
6. PROPERTIES OF CRUCIFERIN WITH MIXED OR IDENTICAL SUBUNIT COMPOSITIONS: SOLUBILITY, HEAT-INDUCED GELATION, AND PEPsin SUSCEPTIBILITY .............................................................. 120
6.1 Abstract ............................................................................................................................ 120
6.2 Introduction ..................................................................................................................... 121
6.3 Material and Methods ..................................................................................................... 122
6.3.1 Materials and Chemicals ............................................................................................ 122
6.3.2 CD and DSC ............................................................................................................... 122
6.3.3 Solubility .................................................................................................................... 122
6.3.4 Heat-Induced Gel Formation and Small Deformation Rheology of Gels .................. 123
6.3.5 Microstructure Assessment ....................................................................................... 123
6.3.6 3D-Structure Modeling and Analysis ........................................................................ 124
6.3.7 Simulated Gastric Fluid (SGF) Degradation Kinetics ................................................ 124
6.3.8 Statistical Analysis ..................................................................................................... 125
6.4 Results and Discussion ................................................................................................... 125
6.4.1 Change of Particle Size and Secondary Structure with pH........................................ 125
6.4.2 Solubility ................................................................................................................. 127
  6.4.2.1 The Effect of NaCl ......................................................................................... 127
  6.4.2.2 The Effect Temperature ................................................................................. 129
6.4.3 Heat-Induced Gelation ............................................................................................ 130
  6.4.3.1 Without NaCl .................................................................................................. 130
  6.4.3.2 With NaCl ....................................................................................................... 132
  6.4.3.3 At pH 2 without NaCl .................................................................................... 134
6.4.4 Susceptibility to Pepsin .......................................................................................... 140
6.5 Conclusions ................................................................................................................ 141
6.6 Connection to Subsequent Chapter ............................................................................ 141
7.0 PROPERTIES OF CRUCIFERIN HOMOHEXAMERS IN OIL-IN-WATER
  EMULSION FORMATION AND STABILIZATION......................................................... 142
7.1 Abstract ...................................................................................................................... 142
7.2 Introduction ................................................................................................................ 143
7.3 Material and Methods ............................................................................................... 145
  7.3.1 Materials and Chemicals ...................................................................................... 145
  7.3.2 Purification of Homohexameric Cruciferins with Different Subunit Compositions .... 145
  7.3.3 Fatty Acid Profile ............................................................................................... 145
  7.3.4 Emulsion Preparation ......................................................................................... 145
  7.3.5 Estimation of Oil Volume Fraction ($\phi$) .............................................................. 146
  7.3.6 Emulsifying Activity Index (EAI) ....................................................................... 146
  7.3.7 Droplet Size Analysis .......................................................................................... 147
  7.3.8 Interface Protein Load ($I'$) and Percentage of Protein Adsorption ...................... 147
  7.3.9 Microstructure Assessment ............................................................................... 148
  7.3.10 Storage Stability of Emulsions ......................................................................... 149
  7.3.11 Intrinsic Fluorescence, Surface Hydrophobicity, and Structure Modeling .......... 149
  7.3.12 Statistical Analysis ......................................................................................... 149
7.4 Results and Discussion .............................................................................................. 150
  7.4.1 Emulsion Formation ......................................................................................... 150
  7.4.2 Emulsion Stability ............................................................................................. 155
LIST OF TABLES

Table 2.1  Extraction and purification of 11−12S globulin fractions from Brassica seeds ................................................................................................................................................................................................. 16

Table 3.1  Average hydrophobicity index (Hø), isoelectric pH (pI), and molecular mass (Mr) of procruciferin and cruciferin ................................................................. 51

Table 4.1  Relationship of cruciferin gene and proteins and their nomenclature in this study ........................................................................................................................................................................................................ 67

Table 4.2  Primer pairs used for PCR to verify T-DNA insertions in Arabidopsis cruciferin genes ................................................................................................................................................................................................. 68

Table 4.3  Sequence comparison of selected cruciferins of A. thaliana with B. napus .......... 72

Table 4.4  Protein (%N based) and crude oil content of Arabidopsis and B. napus seeds ........................................................................................................................................................................................................ 79

Table 4.5  Secondary structure features (%) of embryo seed storage protein of Arabidopsis and B. napus seeds assessed using synchrotron FT-IR............................................. 80

Table 4.6  Protein structure features (percent) of pepsin-treated Arabidopsis and B. napus seeds ........................................................................................................................................................................................................ 86

Table 5.1  Secondary structure features (%) of purified cruciferins from Arabidopsis ....... 102

Table 5.2  Various fluorescence parameters obtained from intrinsic fluorescence and probe binding methods ................................................................................................................ 106

Table 5.3  Values obtained from Differential scanning calorimetric (DSC) for thermal properties of the cruciferin proteins ............................................................................................................................. 116

Table 6.1  The effect of pH and ionic strength on thermal characteristics of cruciferins .................................................................................................................................................................................................. 137

Table 7.1  Oil volume fraction (φ), EAI, Interface protein load (Γ), and percentage of protein adsorption of the cruciferin emulsions generated from canola oil and triolein ................................................................................................................................................................................................. 151
Table 7.2  Average particle size (µm) of emulsion droplets of canola oil or triolein
formed with cruciferin........................................................................................................... 153

Table 7.3  Predicted number of hydrophobic residues of the molecule and the measured
surface hydrophobicity of cruciferins...................................................................................... 157
LIST OF FIGURES

Figure 2.1  Sequence alignment of five major subunits of soybean glycinin......................... 8
Figure 2.2  Secondary structure features of monomer, trimer, and hexamer of 11S globulins in ribbon diagrams................................................................. 11
Figure 2.3  An illustration of Arabidopsis plant and physical components of the seed........... 13
Figure 2.4  A schematic of chromosome map and primary structure assembly of cruciferin CRUA, CRUB, and CRUC .............................................................. 14
Figure 2.5  Process steps involved in homology or comparative modeling of protein 3D structure.................................................................................................. 20
Figure 3.1  Selection of a template for homology modeling of Arabidopsis cruciferins........ 34
Figure 3.2  Sequence and structure conservation of cruciferin protomers.......................... 37
Figure 3.3  Stereospecific quality evaluation of the cruciferin homology models and procruciferin crystal structure ................................................................. 39
Figure 3.4  Three main loop regions in Arabidopsis cruciferins.......................................... 41
Figure 3.5  Surface characterization of Arabidopsis cruciferin models............................... 43
Figure 3.6  Distribution of hydrophobic and hydrophilic residues of Arabidopsis cruciferins........................................................................................................... 44
Figure 3.7  Ribbon diagram showing the homotrimer constructed with each protomer ...... 47
Figure 3.8  A schematic of barrier formation to coalescence and consequent stabilization of the emulsion by CRUC homotrimer .............................................. 49
Figure 3.9  Surface topography of Arabidopsis CRUA, CRUB, and CRUC homotrimer computed by CASTp serve................................................................. 54
Figure 3.10 Prediction of phosphorylation and proteolytic sites of Arabidopsis CRUA, CRUB, and CRUC homotrimer ................................................................. 58
Figure 4.1  Arabidopsis lines with altered seed storage protein profiles.............................. 66
Figure 4.2  Selection of protein rich areas in Arabidopsis seed tissues and analysis of protein secondary structure components............................................. 75
Figure 4.3 Distribution of proteins (amide I at 1650 cm\(^{-1}\)) and lipids (\(\nu\text{C}=\text{O}\) of lipid ester at 1740 cm\(^{-1}\)) in a single cell area of Arabidopsis ...........................................77

Figure 4.4 Protein and lipid distribution in Arabidopsis and B. napus seeds........................................78

Figure 4.5 Analysis of deconvoluted synchrotron FT-IR spectra ....................................................81

Figure 4.6 PCA of the deconvoluted amide I band of WT and SSP-altered Arabidopsis: biplot of the first two principal components.............................................82

Figure 4.7 Visible microscopic images (scale bar = 100 \(\mu\text{m}\)) and synchrotron FT-IR 2D chemical images of untreated and pepsin-treated seeds of WT .....................84

Figure 4.8 Effect of pepsin treatment on Arabidopsis and B. napus SSP ........................................85

Figure 5.1 Chromatographic separation of Arabidopsis protein extracts ........................................92

Figure 5.2 Electrophoretic separation of isolated cruciferin from Arabidopsis .................................99

Figure 5.3 Cartoon depicting the possibility of disulfide exchange in cruciferins..........................100

Figure 5.4 Secondary structure analysis of purified Arabidopsis cruciferins ................................101

Figure 5.5 Zeta (\(\zeta\)) potential of cruciferins (0.1 mg mL\(^{-1}\)) at different pH...............................103

Figure 5.6 Intrinsic fluorescence of cruciferin and fluorescence quenching due to acrylamide..................................................................................................................107

Figure 5.7 Spatial distributions of Trp residues in CRUA (A), CRUB (B), and CRUC (C) homotrimers........................................................................................................109

Figure 5.8 Close-up view of the 3D structure of cruciferin knockout lines in the vicinity of Trp residues of CRUA, CRUB, and CRUC ...........................................110

Figure 5.9 Tryptophan fluorescence of native cruciferins and denaturation induced by 6 M urea and 6 M guanidine-HCl.................................................................112

Figure 5.10 Intrinsic fluorescence of cruciferin as a function of pH............................................113

Figure 5.11 Thermal properties of cruciferins..................................................................................115

Figure 6.1 Far-UV CD spectra of cruciferin at pH 7.4, pH 2.0, and pH 10.0 ...............................126

Figure 6.2 pH-dependent solubility profiles of cruciferin.................................................................128

Figure 6.3 Dynamic oscillatory measurements of 5.0 % (w/v, at pH 7.4) cruciferin dispersions............................131

Figure 6.4 Dynamic oscillatory measurements of 5.0% (w/v) cruciferin dispersion measured at high ionic strength (\(\mu = 0.5\)) (pH 7.4) ..................................................133
Figure 6.5  Dynamic oscillatory measurements of 5.0% (w/v) cruciferin dispersion (pH 2.0) ................................................................. 135

Figure 6.6  Solvent accessible surfaces maps of CRUA (A), CRUB (B), and CRUC (C) hexameric structures ................................................................. 139

Figure 6.7  Pepsin susceptibility of cruciferins ................................................................. 141

Figure 7.1  Microstructure and stability of O/W emulsions of canola oil and triolein stabilized by WT, CRUA, CRUB, or CRUC cruciferin at pH 7.4 .......... 154

Figure 7.2  Effect of NaCl at pH 7.4 on microstructure and stability of O/W emulsions of canola oil and triolein stabilized by WT, CRUA, CRUB, or CRUC cruciferin ................................................................. 156

Figure 7.3  Effect of pH (2.0) on microstructure and stability of O/W emulsions of canola oil and triolein stabilized by WT, CRUA, CRUB, or CRUC cruciferin ................................................................. 158

Figure 7.4  Effect of heating (90 °C) on microstructure and stability of O/W emulsions of canola oil and triolein stabilized by WT, CRUA, CRUB, or CRUC cruciferin ................................................................. 160

Figure 7.5  Emulsion stability after 20 h of storage ................................................................. 161

Figure 7.6  Hexameric structures and solvent accessible surface, electrostatic surface potential, and hydrophobic surface of CRUA, CRUB, and CRUC molecules ................................................................. 163

Figure 7.7  Schematic representation of the proposed mechanism for cruciferin adsorption at oil-water interface ................................................................. 164
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>ANOVA</td>
<td>One-way analysis of variance</td>
</tr>
<tr>
<td>ANS</td>
<td>1-anilino-8-napthalensulfonate</td>
</tr>
<tr>
<td>ASA</td>
<td>Accessible surface area</td>
</tr>
<tr>
<td>CD</td>
<td>Circular dichroism</td>
</tr>
<tr>
<td>CEC</td>
<td>Cation exchange chromatography</td>
</tr>
<tr>
<td>CLSM</td>
<td>Confocal laser-scanning microscope</td>
</tr>
<tr>
<td>CRUA</td>
<td>Cruciferin A</td>
</tr>
<tr>
<td>CRUB</td>
<td>Cruciferin B</td>
</tr>
<tr>
<td>CRUC</td>
<td>Cruciferin C</td>
</tr>
<tr>
<td>EAI</td>
<td>Emulsion activity index</td>
</tr>
<tr>
<td>ELR</td>
<td>Extended loop region</td>
</tr>
<tr>
<td>FSD</td>
<td>Fourier self deconvolution</td>
</tr>
<tr>
<td>FT-IR</td>
<td>Fourier transform-infrared spectroscopy</td>
</tr>
<tr>
<td>HM</td>
<td>Homology modeling</td>
</tr>
<tr>
<td>HVR</td>
<td>Hypervariable region</td>
</tr>
<tr>
<td>IA</td>
<td>Intra-chain disulfide bond containing</td>
</tr>
<tr>
<td>IE</td>
<td>Inter-chain disulfide bond containing</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>LVR</td>
<td>Linear viscoelastic region</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectroscopy</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>O/W</td>
<td>Oil-in-water</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>pI</td>
<td>Isoelectric pH</td>
</tr>
<tr>
<td>PSV</td>
<td>Protein storage vacuole</td>
</tr>
<tr>
<td>RNAi</td>
<td>Ribonucleic acid interference</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC</td>
<td>Size exclusion chromatography</td>
</tr>
<tr>
<td>SGF</td>
<td>Simulated gastric fluid</td>
</tr>
<tr>
<td>SSP</td>
<td>Seed storage proteins</td>
</tr>
<tr>
<td>T-DNA</td>
<td>Transfer deoxyribonucleic acid</td>
</tr>
<tr>
<td>TAG</td>
<td>Triacylglycerol</td>
</tr>
<tr>
<td>TAIR</td>
<td>The Arabidopsis information resource</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>WT</td>
<td>Wild type</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 Rationale

Oilseeds within the *Cruciferae* or *Brassicaceae* family, including canola and mustard, are major contributors to the Canadian agri-food industry and to the bio-economy. In general, oilseed canola contains 45–50% oil and 20–25% protein. At present, the market for canola products is primarily the oil (80–90% of value), which is used primarily for food, and the residual meal (10–20% of value) that is diverted to animal feed applications. The meal is rich in protein and fiber, but its use has so far not extended beyond animal nutrition. Although considerable information is available, the uses and applications of canola protein have not gained traction compared to the seed oil, and also lag behind proteins from other economically important oilseeds, such as soybean.

Cruciferin comprises a significant portion of the seed storage protein (SSP) of crucifer oilseeds and remains in the meal residue after oil extraction. In *Brassica napus*, cruciferin accounts for up to 60% of the total seed protein (Crouch & Sussex, 1981). Mature cruciferin is a hexameric globulin with a molecular weight between 300–350 kDa. The wild type (WT) *B. napus* cruciferin is composed of heterogenous subunits that are expressed by homologous genes of multigene families (Simon, Tenbarge, Scofield, Finkelstein, & Crouch, 1985; Rödin & Rask, 1990a). Five different subunits, CRU1_BRANA, CRU2_BRANA, CRU3_BRANA, CRU4_BRANA, and CRUA_BRANA, have been identified in *B. napus*, although as many as 18 expressed genes exist (Hegedus, unpublished); therefore, cruciferin isolated from a given cultivated variety of canola will be a mixture of these. In the model crucifer *Arabidopsis thaliana*, the cruciferin protomers CRUA, CRUB, and CRUC are encoded by three paralogous genes; CRUA, CRUB, and CRUC that show 78.0, 84.3, and 83.7% similarity with *B. napus* CRUA_BRANA, CRU4_BRANA, and CRU3_BRANA, respectively. Therefore, studies on Arabidopsis cruciferin subunits would be highly relevant to understanding the properties of *B. napus* cruciferin.
By crossing T-DNA insertion mutants, two of the three-cruferin genes of Arabidopsis can be knocked out and viable plants expressing cruciferin that is composed of a single type of subunit can be generated. These new cruciferin genetic variants produce seeds with uncompromised total storage protein content. This study utilized the seeds of cruciferin WT (CRUABC, composed of CRUA, CRUB, and CRUC) and double-knockout lines; CRUAbc (CRUA only), CRUaBc (CRUB only), and CRUabC (CRUC only) to understand the contribution of each subunit type to the properties of the mature protein in Arabidopsis. Knowledge of the structure and the relationship to physico-chemical properties and functions of cruciferin composed of homogeneous subunits will enable the design of cruciferins with new uses and applications, therefore enhancing the value of the crucifer meal residue.

1.2 Hypothesis

The following hypotheses were formulated for Arabidopsis cruciferin and tested in this study;

(i) Cruciferin protein with minimal molecular heterogeneity (homohexamers of CRUA, CRUB, and CRUC) can be obtained by knocking out two of the three-cruferin genes.

(ii) The cruciferin protein expressed by each gene is different in amino acid composition and length (CRUA, 477 aa; CRUB, 455 aa; CRUC, 535 aa), and this will affect the molecular structure.

(iii) The differences in molecular structure of these cruciferins will affect the physico-chemical and techno-functional properties of these proteins.

1.3 Objectives

(i) To predict the structure of Arabidopsis CRUA, CRUB, and CRUC using a modeling approach and investigate these structure models for physico-chemical and techno-functional properties.

(ii) To obtain protein structure details and physico-chemical properties of CRUA, CRUB, and CRUC expressed in Arabidopsis mutant lines.

(iii) To determine the technologically important functional properties of CRUA, CRUB, and CRUC expressed in Arabidopsis mutant lines.

(iv) To generate possible relationships between observed physico-chemical and functional properties of homohexameric CRUA, CRUB, and CRUC and compare these with wild type cruciferin of Arabidopsis.
2. LITERATURE REVIEW

2.1 Brassicaceae Family Plants

The family Brassicaceae (Cruciferae) in the order Brassicales and the clade eudicot is large and includes many plants of major economic importance, such as the familiar edible vegetables (broccoli, cauliflower, cabbage, and turnip), mustard and oil crops (e.g., oilseed rape, camelina, crambe). Most plants of this family produce seeds rich in oil and also contain glucosinolates in different parts of the plant. There are many edible plant species in this family; Brassica oleracea has many edible variants such as broccoli, Brussels sprouts, cabbage, cauliflower, kale, and kohlrabi; B. rapa is the turnip and B. napobrassica is the Swedish turnip, or rutabaga; B. pekinensis and B. chinensis are Chinese cabbage. The seeds of B. juncea (brown/oriental mustard), B. nigra (black mustard) and Sinapis alba (yellow/white mustard, formerly B. hirta) are condiment mustards. The oilseed canola includes B. rapa (formerly B. campestris), B. napus, and B. juncea (canola quality mustard). Most of these plants are native to Eurasia and were later introduced to North America. In Canada, some weeds of this family including Arabidopsis thaliana (thale cress), Camelina spp, and Lepidium spp. also produce seeds rich in oil and glucosinolates.

2.2 Composition of Brassica Seeds and Meals

The seeds of B. napus contain high amounts of oil (40–45%). Similar to many other seed oils, the B. napus oil is composed mainly of triacylglycerols (>90%) with very low amounts of mono- and diacylglycerol and free fatty acids (<0.5%) (Uppström, 1995). The protein content of the dry seed varies between 20–35%. According to Uppström (1995), metabolic activity of mature Brassica seeds is largely controlled by water-soluble albumin proteins, whereas the salt-soluble globulins comprise storage proteins (60–70% of total protein), while oleosins are the structural proteins of oil bodies. The globulins in the Brassica oilseeds are of two types; napin is a 2S albumin and cruciferin is an 11S globulin. In Brassica oilseeds, cruciferin accounts about 60% and napin about 20% of total seed storage protein (SSP) (Crouch & Sussex, 1981; Höglund, Rödin, Larson, & Rask, 1992).
The carbohydrates of Brassica oilseeds are comprised of insoluble carbohydrates (polysaccharides including 50.0% pectins, 24.1% cellulose, 15.5% amylloids, 6.9% arabinans, 3.5% arabinogalactan, and ~0.1% starch) and soluble carbohydrates (5.93–7.58% mono- and oligosaccharides such as sucrose, raffinose, stachyose) which comprise the majority of the fibre fraction is about 12–30% of the meal (Shahidi, Naczk, & Myhara, 1990; Uppström, 1995). Phenolics and glucosinolates predominate among the minor components. Three major categories of phenolic compounds in rapeseed meal are the free phenolic acids (15%), phenolic esters (80%) and insoluble bound phenolics (~5% of total phenolics) (Shahidi & Naczk, 1990). For the most part, these phenolics are located in the embryo and to a lesser degree in the seed coat (Uppström, 1995). Sinapine is the choline ester of sinapic acid and is the abundant phenolic compound in many oilseeds of Brassicaceae (Uppström, 1995).

Rapeseed or canola is the most commercially successful oilseed in this family and is cultivated primarily for edible vegetable oil production. Once the oil (~40% seed dry weight) is extracted, the remaining residue (60% seed dry weight) contains protein, carbohydrates, and other oil-insoluble components of the seed and is referred to as the meal. In Canada, canola includes three Brassica species, *B. napus*, *B. rapa*, and *B. juncea*, for which the oil fatty acid and meal glucosinolate profile fits the definition of “Seeds of the genus *Brassica* (*Brassica napus*, *Brassica rapa*, or *Brassica juncea*) from which the oil shall contain less than 2% erucic acid in its fatty acid profile and the solid component shall contain less than 30 micromoles of any one or any mixture of 3-butenyl glucosinolate, 4-pentenyl glucosinolate, 2-hydroxy-3 butenylglucosinolate, and 2-hydroxy-4-pentenyl glucosinolate per gram of air-dry, oil-free solid” (Canola Council of Canada, accessed Oct, 2012, http://www.canolacouncil.org/). In the past few years, over 99% of Canadian canola production was *B. napus* because of the introduction of herbicide tolerant traits (CGC, 2010). Historically, rapeseed (*B. napus* L.) or its meal has not been used in human food (Wanasundara, 2011), but widely used as a protein and digestible energy source for livestock, such as dairy cattle, farmed fish, poultry, and swine. The meal contains protein (36%), moisture (12%), residual oil (3.5%), crude fiber (12%), starch-free sugars, and soluble non-starch polysaccharides (15%), ash (6.1%), and glucosinolates (7.2 µmoles g⁻¹) (Canola Council of Canada, accessed Oct, 2012, http://www.canolacouncil.org/). It is also rich in vitamins and essential minerals (Canola Council of Canada, accessed Oct, 2012, http://www.canolacouncil.org/). The amino acid profile of canola meal protein better suited for
animal nutrition than meals from cereal crops. Recent studies on canola protein isolates show favorable amino acid utilization values for 10 to 12 year-old children and adults according to rodent models (Klockeman, Toledo, & Sims, 1997) and high ileal digestibility in human assessments (Bos et al., 2007). Like many other plant protein sources, canola meal is limiting in lysine, but noted for its high level of methionine and cysteine (Canola Council of Canada, accessed Oct, 2012, http://www.canolacouncil.org/; Bos et al., 2007).

2.3 Storage Proteins of Brassica Seeds

Among the three major types of proteins (storage, structural, and biologically active) that accumulate during the seed development (Fukushima, 1991), the storage proteins are the most abundant and degrade during the seed germination to provide C, N, and S for developing seedlings (Derbyshire, Wright, & Boulter, 1976; Shewry, Napier, & Tatham, 1995). The classic Osborne classification (1924) based on solubility separates seed proteins into water soluble albumins, salt soluble globulins, alcohol soluble prolamins, and dilute acid/alkali soluble glutelins obtained from sequential solubilizing steps. In contrast to cereals such as rice, oat, wheat, barley, and rye which are rich with glutelin and prolamin, the oil rich soybean, rapeseed, and peanut, and the legumes (e.g., pea, chickpea, lentil) are rich in globulin and albumins (Fukushima, 1991). The globulins of legumes and crucifers include 11−12S pea legumin, soybean glycinin, rapeseed cruciferin and 7S soybean β-conglycinin, pea vicilin and convicilin. The albumins are 1.7−2S proteins, such as soybean albumin and rapeseed napin. However, when water (no pH adjustment) was used as the first solvent of extraction in the Osborne solvent series, both 11−12S and 2S proteins were extracted from canola meal (Wanasundara, Abeysekara, McIntosh, & Falk, 2012), indicating difficulty in using Osborne classification to identify seed proteins segregated according to more detailed physico-chemical properties.

Similar to the any other eudicots, crucifer seeds accumulate reduced form of C and N as storage proteins. The major storage proteins in Brassica seeds are cruciferin (~60%) and napin (~20%) (Crouch & Sussex, 1981; Höglund, Rödin, Larson, & Rask, 1992). The 11S globular cruciferin belongs to the cupin superfamily of proteins, has an approximate molecular weight of 300−350 kDa, and is rich in Gln and Gly (Wanasundara, 2011). Cruciferin is soluble in water and salt solutions above pH 3.0 (Wanasundara et al., 2012). Napin is a dimeric 2S albumin abundant in Cys, Gln, and Pro, has molecular weight of 12−16 kDa and belongs to the prolamin
superfamily. The oleosins are not SSPs, but may comprise up to 20% of the total proteins, have molecular weights of ~19 kDa and are associated with oil-body membranes (Murphy & Cummins, 1989; Murphy, Cummins, & Kang, 1989). The minor proteins in Brassica seeds include protease inhibitors (Ceciliani et al., 1994), dehydrins (Svensson, Palva, & Welin, 2000) and calcium-dependent calmodulin binding proteins (Neumann, Condron, & Polya, 1996). As the most abundant protein in crucifer oilseeds, detailed knowledge on structure, function, and properties of cruciferin is very important and helps in developing new uses for crucifer seed proteins or the meal.

2.4 Biosynthesis of Cruciferin

In general, multiple homologous genes encode 11S globulin subunits; for example, the A1aB1b, A2B1a, A1bB2, A3B4, and A5A4B3 isomers of soybean glycinin and the CRU1_BRANA, CRU2_BRANA, CRU3_BRANA, CRU4_BRANA, and CRUA_BRANA of rapeseed cruciferin. The precursors of cruciferin are synthesized in the endomembrane system and stored in specialized vacuoles referred to as the protein storage vacuoles (PSV) (Jaing et al., 2001). In general, the 11S globulins are synthesized by membrane-bound ribosomes into endoplasmic reticulum (ER) as a single precursor called the preproglobulin (~50kDa) which is composed of acidic (~30kDa) and basic (~20kDa) chains linked together (Adachi et al., 2003; Gruis, Schulza, & Jung, 2004). This preproglobulin signal peptide is cleaved co-translationally. In the ER, inter- and intra-chain disulfide bonds are formed resulting in an 8–9S trimeric proglobulin (e.g., proglycinin or procruciferin). The trimeric proglobulin is then transported to the PSV by either a Golgi-dependent or Golgi-independent pathway (Wan, Ross, Yang, Hegedus, & Kermode, 2007). Either on route to or within the PSV, proglobulins undergo further proteolytic cleavage at the Asp–Gly bonds by a specific asparaginyl endopeptidase, yielding the mature form of the polypeptides linked through one disulfide bond (Adachi et al., 2003; Gruis et al., 2004). This process facilitates the association of mature trimers into a hexameric complex in the presence of other vacuole assembling enzymes (Dickinson, Hussein, & Nielsen, 1989). These proteins accumulate in the cotyledons and axis during seed development and their levels peak in the mature stage. Some of the seed storage proteins undergo post-translational modifications such as glycosylation and phosphorylation. Generally, cruciferins not glycosylated (Simon et al., 1985), but are highly phosphorylated (Wan et al., 2007). Twenty phosphorylation sites were
identified in Arabidopsis cruciferin, comprising nine Ser, eight Thr, and three Tyr residues located on the inter-chain disulfide containing face (IE face) of the globulin trimer and which may be involved in hexamer formation Wan et al. (2007). When the conditions are suitable for germination, SSP are degraded and serve as a carbon and nitrogen source for the emerging seedling.

2.5 Molecular Structure of 11–12S Globulins

2.5.1 Primary Structure

The 11–12S globulins, including glycinin and cruciferin, belong to the cupin (cupa means small barrel in Latin) protein superfamily (Plietz, Damaschun, Müller, & Schwenke, 1983; Plietz, Drescher, & Damaschun, 1987; Rödin & Rask, 1990a). These proteins contain multiple subunits (six different subunits, protomers or monomers) in their wild type form and have a molecular weight between 300 and 350 kDa. The subunit of 11–12S protein is comprised of a polypeptide chain of 450 to 550 amino acid residues and an assembly of two polypeptides, a heavy \( \alpha \) (a or acidic, \( M_r \) approx. 30 kDa) and a light \( \beta \) (b or basic, \( M_r \) approx. 20 kDa) chain (Dalgalarrondo, Robin, & Azana, 1986).

For soybean glycinin, six variant subunits have been identified and classified into two groups according to their homology; group I- A1aB1b, A2B1a, and A1bB2 and group II- A3B4 and A5A4B3 (Figure 2.1) (Nielsen et al., 1989). In each protomer, two disulfide bonds can be identified. One disulfide bond is within the acidic chain, which is known as intra-chain disulfide bond, and the other bond is between the acidic and basic chain and is known as inter-chain disulfide bond (Figure 2.1). Almost all Cys residues, and therefore all intra- and inter-chain disulfide bonds, are conserved in the five soybean glycinin variants. The secondary structure features of five soybean subunits have minimal differences; however, their hypervariable regions (HVRs) are disordered and consist mainly of repeating stretches of hydrophilic amino acids (usually Glu, Gln, and Gly). For 11–12S globulins, three to six non-conserved HVRs have been reported (Adachi, Yagasaki, Gidamis, Mikami, & Utsumi, 2001; Adachi et al., 2003; Tandang-Silvas et al., 2010; Wright, 1987). For example, group I soybean glycinins have six HVRs, whereas group II subunits have five HVRs. Hypervariable region V of the group I glycinins and HVR IV of the group II glycinins are the most disordered regions.
Figure 2.1 Sequence alignment of the five major subunits of soybean glycinin. The glycinins are classified as Group I (a) and II (b) based on the homology of protomer sequences. Identical and semi-conserved amino acids are highlighted in black and grey, respectively. Secondary structure is displayed above the sequences: helices are shown as cylinders (red: α-helix, and pink: 3_{10} helix); β-strands as arrows with labeling according to Tandang-Silvas et al. (2010); disordered residues are shown as dots (signal peptide in green and HVRs I to VI in magenta) as determined for A1aB1b (Adachi et al., 2001) and A3B4 (Adachi et al., 2003). Cysteine residues are shown in black letters in yellow boxes. Intra-protomer (IA) and inter-protomer (IE) disulfide bonds are indicated by yellow lines connecting cysteine residues (Withana-Gamage & Wanasundara, 2012, reuse with permission from Elsevier-copyright clearance center via RightsLink service).
Subtle changes of the primary sequence, especially in HVRs, can result in substantial differences in physico-chemical and techno-functional properties of the subunit variants (see section 2.9). For rapeseed, mature cruciferin also assumes a hexameric assembly that is composed of at least five different predominant subunits, CRU1, CRU2, CRU3, CRU4, and CRUA that are derived from the genes \textit{BnC1}, \textit{BnC2}, \textit{CRU3}, \textit{CRU4}, and \textit{CRU2/3}, respectively. In \textit{A. thaliana} cv. Columbia, three predominant cruciferin subunits, \textit{CRUA}, \textit{CRUB}, and \textit{CRUC}, are present. Differences in cruciferin subunit structure and composition may affect quaternary structure and in turn the physico-chemical and techno-functional properties of the protein.

### 2.5.2 Secondary Structure

The secondary structure of the 11−12S seed globulins is characterized by $\alpha$+$\beta$ type structures. Most 11−12S globulins, including soybean A3B4/proA1aB1b, pumpkin pro-11S, pea prolegumin and rapeseed procruciferin, have 25 to 27 $\beta$-sheets, 5 to 7 $\alpha$-helices and 3 to 4 3$_{10}$-helices (Tandang-Silvas et al., 2010). The secondary structures of the hypervariable regions (HVRs) are not clearly defined in the crystal structure of 11−12S globulins due to poor resolution of atomic density maps (Adachi et al., 2001, 2003). However, it is believed that the HVRs do not have a higher order secondary structure and consisting mainly of loops (Adachi et al., 2001, 2003; Tandang-Silvas et al., 2010). A model of the procruciferin monomer is provided in Figure 2.2A.

### 2.5.3 Tertiary and Quaternary Structure

In the cruciferin tertiary structure, it is believed that the basic $\beta$-chain of the polypeptide is buried within the molecule and the acidic $\alpha$-chain is more solvent exposed (Job, Rajjou, Lovigny, Belghazi, & Job, 2005). The suprastucture of 11−12S globulins may be a trimer or a hexamer. In the trimer, the face containing the inter-chain disulfide bond is known as the IE face and the face containing the intra-chain disulfide bond is designated as the IA face (Figure 2.2A; Adachi et al., 2001, 2003). Mutation of intra-chain disulfide bond in 11S seed globulins had little or no impact on the trimer assembly (Utsumi, Gidamis, Kanamori, Kang, & Kito, 1993a); however, trimers that are devoid of inter-chain disulfide bond cannot assemble into a hexamer (Jung, Nam, Saalbach, Müntz, & Nielsen, 1997).

The mature 11−12S globulins of rapeseed, soy and sunflower are hexamers. Typically, they are heterohexamers composed of protomers derived from multiple genes (Adachi et al., 2001,
The initial hypothetical model for 11S globulin was based on data obtained from microscopic and small angle X-ray scattering methods for soybean glycinin (Badley et al., 1975). According to this proposed model, cruciferin consists of twelve subunits, six acidic and six basic (each subunit corresponding to 2S dissociation product), which are packed into two identical hexagons, yielding a hollow oblate cylinder with a six-fold axis of symmetry. Using data from high-resolution X-ray scattering and dissociation pattern studies, Plietz and coworkers (1983) proposed a model consisting of six subunits arranged as a triagonal antiprism with dihedral point group symmetry 32 for the 11S globulins of sunflower and rapeseed. Advances in X-ray crystallography have allowed the crystal structure of the soybean proglycinin A3B4 and A1aB1b, (Adachi et al., 2001, 2003), pumpkin pro-11S, pea prolegumin, B. napus procruciferin (Tandang-Silvas et al., 2010) and Prunus dulcis (almonds) Pru du amandin (PDB code 3EHK). In all these 11−12S globulins, the crystal structure based models have highly conserved antiparallel β-sheets folded into two jellyroll β-barrels and two extended α-helix domains (Figure 2.2A).

Each trimer is approximately 95×95×40 Å in size and with subunits in a head-to-tail non-covalent configuration (Figure 2.2B; Tandang-Silvas et al., 2010). The mature 11−12S hexamer is assembled via IE face-to-face piling of two trimers (Figure 2.2C) (Adachi et al., 2003). Non-covalent bonds, including hydrophobic, electrostatic, H-bonds, van der Walls, and H-bonded salt bridges, are primarily involved in the association of two trimers to form the hexamer (Adachi et al., 2001, 2003).

2.6 Arabidopsis and Crucifer SSP

Arabidopsis thaliana (thale cress, mouse-ear cress or Wall cress) is widely regarded as the classic model plant for studying Brassicaceae oilseeds. Basic research in genetics and molecular biology takes advantage of the fully sequenced small genome of Arabidopsis (The Arabidopsis Genome Initiative, 2000) and its relatively short life cycle. Arabidopsis is a small flowering plant that produces a large number of seeds in the seedpods or siliques (Figure 2.3A) converts the integument of the ovule into a seed coat (testa), endosperm (unilayered), and embryo (cotyledons, hypocotyl, and radicle) at the early stages of the seed development (Figure 2.3B). The axile bent embryo occupies a large area of the seed (Figure 2.3B).
Figure 2.2 Structural models of the monomer (A), trimer (B), and hexamer (C) of 11S globulins. Monomer and trimer models are based on *B. napus* procruciferin (PDB code 3KGL) and the hexamer model is from *Prunus dulcis* (almonds) Pru du amandin (PDB code 3EHK). Secondary structure features are distinguished in the ribbon diagrams. Inter-chain and intra-chain disulphide bonds are shown as spheres. Three-fold molecular axis is indicated by a filled triangle in trimer and hexamer. Structures were visualized using PyMOL Molecular Graphics System (Version 1.5.0.4 Schrödinger, LLC).
At later maturation stages of seed development, C and N reserves accumulate in the Arabidopsis seed embryo as proteins and oils, and each account for 30–40% of the dry seed weight (Li, Beisson, & Ohlrogge, 2006; North et al., 2010).

Similar to the other plants in the Brassicaceae family, the major SSPs of Arabidopsis are cruciferin and napin. In A. thaliana cv. Columbia four genes encode cruciferin; AT1G03890.1, AT1G03880.1/CRUB, AT4G28520.1/CRUC, and AT5G44120.3/CRUA located on chromosomes 1, 1, 4, and 5, respectively (http://www.arabidopsis.org/, accessed June 2012, Figure 2.4A). AT1G03890.1 is considered less important than the other three genes as it is expressed at very low levels and contributes little to the total cruciferin complement (Wan et al., 2007). Therefore, only the expression of AT5G44120.3, AT1G03880.1, and AT4G28520.1 is considered significant and these genes encode CRUA, CRUB, and CRUC protomers, respectively.

The cruciferin subunits differ in length and molecular weight; CRUA 472 aa and 52.6 kDa, CRUB 455 aa and 50.6 kDa and CRUC 524 aa and 58.2 kDa (TAIR, 2012) (Figure 2.4B). The calculated isoelectric point (pI) is 7.68 for CRUA and 6.53 for CRUB and CRUC. The intra-chain disulfide bond is located between C36↔C69 in CRUA, C30↔C63 in CRUB and C37↔C70 in CRUC protomers, and the inter-chain disulfide bond is between C112↔C289 in CRUA, 106↔276 in CRUB and C113↔C340 in CRUC. Two free sulfhydryl (–SH) groups are found in each protomer; residues 11 and 293 of CRUA, residues 40 and 280 of CRUB, and residues 21 and 406 of CRUC (Figure 2.4B).

2.7 Extraction and Purification of Cruciferin

Most seed globulins and albumins in oil-free meals can be extracted into aqueous buffers of pH around 8.0 or aqueous solutions of fairly high salt concentration (>0.5 M). Further separation and isolation of 11–12S or 1.7–2S protein can be achieved through ammonium sulfate precipitation (Rao, Urs, & Rao, 1978) or chromatographic means (Bérot, Compoint, Larré, Malabat, & Guéguen, 2005; Bhatty, MacKenzie, & Finlayson, 1968; Dalgalarrondo et al., 1986; Schwenke, Raab, Linow, Platz, & Uhlin, 1981). Bérot and coworkers (2005) describe a comprehensive isolation and purification process for cruciferin and napin using a series of chromatographic steps that can be scaled up. In this protocol, Tris-HCl buffer at pH 8.5 is used initially to extract total seed proteins with a subsequent nanofiltration step to remove most of the co-extracted small molecules, including phenolic components. Size exclusion chromatography
(SEC; Superdex 75) of this extract allows isolation of all proteins. The protein fraction can be further separated into unbound neutral cruciferin and bound napin using cation exchange chromatography (CEC; streamline SP-XL column). Further purification of cruciferin is achieved with another SEC (S300 column) step. Hydrophobic interaction chromatography (HIC; Source 30S column) allows further refinement of the napin fraction.

Figure 2.3 An illustration of an Arabidopsis plant and physical components of the seed. (A) A view of the plant, flower, and silique. (B) Schematic diagram of a seed cross-section. The bulk of the seed is filled with the embryo where most of the triacylglycerol (TAG) and SSPs accumulate.
Figure 2.4 A schematic of chromosome map and primary structure assembly of cruciferin CRUA, CRUB, and CRUC. (A) Locus position of major cruciferin genes, CRUA, CRUB, AT1G02890.2, and CRUC, on the Arabidopsis chromosome map. The cruciferin loci were placed on the physical map using the program Chromosome Map Tool (TAIR). (B) Major component identification of CRUA, CRUB and CRUC primary structure. Color component representation: grey for signal peptides, black for N/C terminal disordered regions, brown for hypervariable region I (HVR-I), orange for extended loop region (ELR), and red for HVR-II. The green arrowhead indicates the internal processing site. Each number indicates the corresponding amino acid residue on the polypeptide chain.

Several researchers have worked on obtaining protein isolates from rapeseed/canola meal, but most the final isolates were not characterized with respect to protein composition. For the few characterized products, the isolates contain both cruciferin and napin (Wanasundara, 2011). In Canada, MCN Bioproducts Inc. (now owned by Bunge Limited) described an aqueous fractionation method to obtain canola protein concentrates, known as CanPro IP insoluble protein concentrate, CanPro soluble protein concentrate, and CanPro FP fiber protein (http://www.canproingredients.ca/, accessed Jan, 2013). BioExx (www.Bioexx.com, accessed Dec, 2012) and Burcon NutraScience (www.burcon.ca, accessed Dec, 2012) have proprietary technology to produce canola protein products, which are mixtures of 11–12S and 2S. Burcon NutraScience has two protein isolates, branded as Puratein® for globulin-rich fraction and Supertin™ for albumin-rich fraction and BioExx produces Isolexx®, which has self-affirmed GRAS (generally recognized as safe) status in the US, with full GRAS notice being claimed. Wanasundara & McIntosh (2008) reported an aqueous extraction process that separates the napin-rich and cruciferin-rich protein fraction from Brassicaceae defatted meal which can also be scaled up. Recently, Nioi, Kapel, Rondags, & Marc (2012) showed that napin can be isolated in highly pure form from commercial rapeseed meal using carefully selected aqueous extraction
conditions. The various methods for cruciferin extraction and purification from different Brassica oilseeds are summarized in Table 2.1.

2.8 Techniques of Protein Structure Probing

2.8.1 In Vitro Spectroscopic Techniques

The intrinsic fluorescence of proteins due to aromatic amino acid residues, Phe, Tyr, and Trp, provides information on the microenvironment of these residues in the globular protein structure. For example, conformation (folding/unfolding), nonpolar binding sites, dynamics of molecular motion, etc. (Chen & Barkley, 1998; Hannemann et al., 2002; Kuznetsova et al., 2005; Staiano et al., 2005; Stryer, 1968) of the protein molecule’s structure can be obtained, though the technique is limited to proteins containing aromatic residues. Fluorescent probes can also be used to investigate hydrophobic regions or patches on the protein’s surface.

The unsaturated nature of the amide bond in the polypeptide backbone can be probed by vibrational spectroscopy (Raman and mid infra-red regions) to investigate protein secondary structure. In general, the infrared (IR) spectroscopy can be applied to molecules with symmetric vibrations of non-polar groups, whereas Raman spectroscopy can be applied to molecules with asymmetric vibrations of non-polar groups. The intrinsic fluorescence due to amino-acids or co-extracted phenolic compounds continues to be a major obstacle for using Raman spectroscopy on plant proteins (Ma, Rout, Chan, & Phillips, 2000), but can be eliminated by the modified technique of Fourier transformed near-infrared (FT-NIR) Raman spectroscopy (Schrader, Hoffman, Simon, & Sawatzki, 1991). IR spectroscopy, especially Fourier transform-infrared (FT-IR) spectroscopy is considered the most versatile vibrational spectroscopic tool for studying the secondary structure of plant proteins. Synchrotron-powered FT-IR (synchrotron FT-IR) has many advantages compared to conventional Globar source generated FT-IR (Duncan & Williams, 1983; Miller & Dumas, 2006; Wetzel, Eilert, Pietrzak, Miller, & Sweat, 1998). Synchrotron light is much more intense (100–1000 times brighter, higher brilliance) than that from a Globar source and allows scanning at a high signal-to-noise ratio.
<table>
<thead>
<tr>
<th>Material</th>
<th>Extraction and Purification</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defatted <em>B. napus</em> (var. Nugget) meal</td>
<td>Extracted with 10% NaCl (1:50, w/v) for 90 min at 5 °C. Separated on a Sephadex G-100 column (5 mM phosphate buffer at pH 8.6 containing 5% w/v NaCl)</td>
<td>Bhatty et al., 1968</td>
</tr>
<tr>
<td>Dehulled defatted <em>B. juncea</em> and <em>B. campestris</em> meal</td>
<td>Extracted with water (1:20, w/v) for 2 h at 28 °C. Separated with 35% (NH₄)₂SO₄ containing 1 M NaCl.</td>
<td>Rao et al., 1978</td>
</tr>
<tr>
<td>Ground defatted <em>B. napus</em> (Sollux) meal</td>
<td>Extracted with 50 mM phosphate buffer (1:10, w/v) at pH 8.0 for 60 min at room temperature containing 1 M NaCl. Separated on a Sephacryl S-300 and DEAE Sephadex A-50 columns.</td>
<td>Schwenke et al., 1981</td>
</tr>
<tr>
<td>Defatted <em>B. napus</em> (Tandem) meal</td>
<td>Extracted with 20 mM borate buffer (1:10, w/v) at pH 8.0 containing 5% NaCl. Separated on a Sephacryl S-300 column.</td>
<td>Dalgalarondo et al., 1986</td>
</tr>
<tr>
<td>Defatted <em>B. napus</em> meal</td>
<td>Extracted with 35 mM phosphate buffer (1:20, w/v) at pH 7.6 containing 1 M NaCl. Precipitated with 10–20% (NH₄)₂SO₄ and separated on a Sephacryl S-200 column.</td>
<td>Salleh et al., 2002</td>
</tr>
<tr>
<td>Rapeseed (Express) meal</td>
<td>Extracted with 50 mM Tris-HCl (1:10, w/v) at pH 8.5 containing 750 mM NaCl. Separated on desalting, cation exchange, and size exclusion (Sephacryl 300) columns.</td>
<td>Bérot, et al., 2005</td>
</tr>
</tbody>
</table>

*modified from Wanasundara (2011).*
The higher intensity of the synchrotron radiation photon beam together with a narrow range of scanning angles of small sample areas (aperture size around 10–20 µm) results in fine ultraspacial resolution (3–10 µm). Moreover, synchrotron FT-IR microspectroscopy provides a non-chemical and minimally destructive approach to obtain ultraspatially resolved details of chemical structure and distribution of biopolymers in seed tissues. Studies on barley (Yu et al., 2003; Yu, McKinnon, Christensen, & Christensen, 2004a), yellow and brown canola (Yu, 2008), corn (Yu, 2005a; Yu, McKinnon, Christensen, & Christensen, 2004b), oats (Damiran & Yu, 2010), soybean (Pietrazak & Miller, 2005) and wheat (Yu, Block, Niu, & Doiron, 2007) demonstrated the ability of synchrotron FT-IR to investigate the distribution and structure of biopolymer components in seed. Chemical imaging of cellular components in biological tissues is possible with synchrotron FT-IR spectromicroscopy (Miller & Dumas, 2006; Yu, 2004; Yu, 2005b) and allows chemical mapping of tissues.

The synchrotron FT-IR scan of plant seed material or purified protein in the 4000–800 cm\(^{-1}\) frequency range includes five prominent, characteristic, in-plane vibrational amide peaks due to H–N–C=O amide bond group (Barth, 2007; Miyazawa, Shimanouchi, & Mizushima, 1956; Susi, 1972). The following bands are recognized and related to protein; the amide A band at 3300 cm\(^{-1}\) due to 100% \(\nu\)N–H, the amide B band at 3071 cm\(^{-1}\) due to 100% \(\delta\)N–H, the amide I band at 1650 cm\(^{-1}\) due to 80% \(\nu\)C=O and 10% \(\nu\)C–N, the amide II band at 1545 cm\(^{-1}\) due to 60% \(\delta\)N–H and 40% \(\nu\)C–N and the amide III band at 1300 cm\(^{-1}\) due to 40% \(\nu\)C–N and 30% \(\delta\)N–H (Barth, 2007; Miyazawa et al., 1956; Susi, 1972). In the study of seed tissues, the amide A and B absorption bands caused by the N–H flexural vibrations are very weak because of the interference from hydrogen-bonded –OH groups of water and carbohydrates making them not useful for assessing proteins. The amide III band (peak baseline 1200–1350 cm\(^{-1}\)) usually overlaps with the absorption bands of ring breathings (\(\nu\)C=O) of the syringyl (4-hydroxy-3,5-dimethoxyphenyl) and guaiacyl (4-hydroxy-3-methoxyphenyl) units of lignin (Sun, Sun, Fowler, & Tomkinson, 2002) and lignin associated hemicellulose (Sun & Hughes, 1999). When these limitations are considered, the strong peaks of amide I and amide II are the most useful for protein structural studies within seed tissues and could be considered for protein diagnostics. Furthermore, integration of the area under the amide I peak (1650 cm\(^{-1}\), integral baseline 1720–1588 cm\(^{-1}\)) and amide II peak (1550 cm\(^{-1}\), integral baseline 1589–1489 cm\(^{-1}\)) or the total amide I and II peaks
(integral baseline 1720–1489 cm\(^{-1}\)) acquired from whole seed scanning can generate chemical maps of protein distribution and concentration in seed micro-compartments.

The IR vibrations of N–H in-plane bending (\(\delta\)N–H) and C–N stretching (\(\nu\)C–N) of amide I and II are highly sensitive to secondary structure folding upon chain packing in the crystalline region (Miyazawa et al., 1956). The intensity of the amide II band is relatively strong (not as strong as amide I), but an absorbance band originating from the \(\nu\)C=O of the lignin aromatic ring (~1516 cm\(^{-1}\); Suarez-Garcia, Martinez-Alonso, & Tascon, 2002) overlaps with it. Considering the interference from other biopolymers, such as polysaccharides, oil and lignin, only a single absorption band of the protein-backbone, i.e., amide I, can be used to probe protein secondary structural features. In protein secondary structure studies, the Fourier self deconvolution (FSD) permits separation of individual component peaks (Kauppinen, Moffatt, Mantsch, & Cameron, 1981) from helical, \(\beta\)-sheet, \(\beta\)-turn, and random structures that overlap within the broad amide I region (Byler & Susi, 1986; Dong, Huang, & Caughey, 1990). The protein secondary structure content of a protein can be calculated by Gaussian curve fitting to these deconvoluted bands (Byler & Susi, 1986).

The circular dichroism spectroscopy (CD) in the far-UV spectral region (190–250 nm) also allows secondary structure details of a protein to be obtained. At these wavelengths, the peptide bond of the protein backbone acts as a chromophore and allows different secondary structure features to give different CD signals. The \(\alpha\)-helix, \(\beta\)-sheet, \(\beta\)-turn, and random structure segments of a protein give characteristic shapes and magnitudes of the UV signal and quantification of each component can be done by comparative modeling with CD spectra of a well-characterized protein for secondary structure.

X-ray crystallography and NMR are key analytical methods for obtaining 3D protein structure details at the atomic level. The molecular structures of A1aB1b and A3B4 from \textit{Glycine max} and Cru2/3a from \textit{B. napus} have been determined using X-ray crystallography at a resolution of 2.80 Å for A1aB1b (Adachi et al., 2001; PDB code: 1FXZ) and 1.90 Å for A3B4 (Adachi et al., 2003; PDB code: 2D5F) and 2.98 Å for Cru2/3a (Tandang-Silvas et al., 2010; PDB code 3KGL). As mentioned earlier, the poor internal order of cruciferin due to subunit heterogeneity diffracts X-ray to high angles and determination of quaternary structure is difficult. The molecular structures of the above mentioned 11S globulins were determined by X-ray crystallography of homogenous protein expressed in systems such as \textit{Escherichia coli} or tobacco leaves (Adachi et al., 2003;
Tandang-Silvas et al., 2010), which are different than the PSV of seeds. When high-resolution structural data cannot be obtained by X-ray crystallography, NMR generates valuable structural information at low-resolution, but is limited to proteins with molecular weights less than 25–30 kDa (Montelione, Zheng, Huang, Gunsalus, & Szyperski, 2000).

2.8.2 In Silico Modeling Techniques

Approaches based on homology modeling (HM) and energy minimization, can be exploited to derive plausible tertiary structures of proteins with known primary sequence as an alternative to an experimentally resolved structure. The *ab initio* or *de novo* technique is a template-independent structure modeling approach and depends on the amino acid sequence and *in silico* folding of proteins to its native state according to the free energy landscape theory using computational algorithms (Bonneau & Baker, 2001). The complexity of this method is compounded exponentially by the number of residues in the chain (i.e., $20^N$, N is the number of amino acid residues in the chain, 20 amino acids occur naturally) (Samish, MacDermaid, Perez-Aguilar, & Saven, 2010). Application of the *ab initio* method is still limited to small protein molecules (<150 amino acid residues) due to the complexity of the calculations required; therefore, application to SSPs that have few several hundreds of amino acid residues is problematic. Considering these limitations, the HM offers an opportunity to understand the structure of SSPs. HM may also allow the design of structural modifications that can achieve desirable functionalities because it is an *in silico* process (Withana-Gamage & Wanasundara, 2012).

HM uses the principle of evolutionary conservation of an unknown protein to a known structure and is also referred to as comparative modeling or knowledge-based modeling. A few protocols are available for homology model building (Bordoli et al., 2009; Kopp & Schwede, 2004; Martí-Renom et al., 2000) and the key steps are depicted in Figure 2.5. The test amino acid sequence to be modeled is called the ‘target sequence’ and can be obtained from an available protein database, such as TAIR or UniProtKB/TrEMBL. An experimentally obtained molecular structure that has an evolutionary close relationship to the target sequence (>30% sequence identity) is called as the ‘template’. The initial step in HM is recognition of the template structure by aligning the target and template sequences. Template structures can be obtained from RCSB Protein Data Bank (http://www.rcsb.org/pdb/home/home.do, accessed June 2012). Two criteria
are considered when selecting an appropriate template: (i) sequence identity between the target and the template sequence (the higher the better), and (ii) the experimental quality (i.e., resolution) of the template (Bordoli et al., 2009).

Figure 2.5 Steps involved in homology or comparative modeling of protein 3D structure (modified from Withana-Gamage & Wanasundara, 2012).

Once a suitable template is chosen, the target sequence is aligned with the template with some degree of manual intervention to minimize misalignments. Generation of the model with backbone and side chains can be completed using programs such as MODELLER (Sali & Blundell, 1993), or web-based servers such as SWISS-MODEL Workspace (http://swissmodel.expasy.org/, accessed July, 2012). The loop regions, for which experimentally determined structures are generally not available, can be generated and optimized in silico using programs like MODELLER (Sali & Blundell, 1993) or ModLoop (Fisher & Sali, 2003). At this
stage, if necessary, the geometries of side chain packing can be further corrected using force-field energy minimization approaches such as GROMOS96 (van Gunsteren et al., 1996) and CHARMM22 (Brooks et al., 1983). Essentially, the stereochemical quality of the modeled structure should to be validated using tools such as PROCHECK (Laskowski, MacArthur, Moss, & Thornton, 1993), Verify3D (Lüthy, Bowie, & Eisenberg, 1992), or ProSA (Wiederstein & Sippl, 2007). The final step in protein structure modeling involves assembly of the modeled protomers to obtain a quaternary structure using existing trimer and hexamer templates of closely related proteins.

Understanding the structure characteristics of a protein is critical in relating biological and physico-chemical functions of a protein and provides enormous opportunities to design enhancements. Nakai and group (Nakai, 1983; Nakai & Li-Chan, 1988; Townsend & Nakai, 1983; Voustinas, Cheung, & Nakai, 1983; Voustinas, Nakai, & Harwalker, 1983) pioneered modeling of structure-function relations to quantify food protein functionalities using QSAR, while others enhanced this approach (Pripp, Isaksson, Stepaniak, Sørhaug, & Ardo, 2005). This work showed that it is possible to predict the functional properties of a protein depending on the properties of its primary amino acid sequence, including the hydrophobicity, side-chain bulk and electrostatic properties of their side groups. To understand how a protein functions in a complex multi-component system, knowledge of amino acid distribution in three-dimensional space is also essential. Three-dimensional structure of protein provides insight into the intramolecular arrangement of amino acid residues and nature of the microenvironment around the residues; therefore, enabling predictions of its physico-chemical properties in a complex environment. Using in silico modeled glycinin subunits, Withana-Gamage & Wanasundara (2012) showed that techno-functional properties, such as solubility, emulsifying, foaming, and gelation of proteins, can be predicted using physico-chemical properties that are directly related to the structural characteristics, such as surface hydrophobic patches, surface electrostatic potential distribution, cavities, amino acid orientation on the surface, and disulfide and free sulfhydryl group distribution.

2.9 Biological, Functional, and Physico-chemical Properties of 11S Globulins

Among the 11–12S globulins from economically important crops, G. max (soybean) glycinin has been the most widely studied for biological, functional, and physico-chemical properties;
however, much less is known about crucifer 11S globulin. The physico-chemical properties of a protein dictate its biological and technologically important functions therefore; a thorough understanding of physico-chemical properties will enable prediction of functionalities and uses of the protein.

2.9.1 Physico-chemical Properties

The physico-chemical properties of a protein are governed by factors intrinsic to the molecule. The protein conformation under the conditions of investigation affects the physico-chemical properties, including molecular charges, isoelectric pH, denaturation temperature, and residue hydrophobic environment.

The surface charge of the protein is determined by the number and type of the residues exposed on the surface. The charge of the protein under given conditions determines aggregation or repulsion of molecules which is critical for protein solubilization. The IE face of soybean A3B4 homotrimer has a slightly positive (basic) charge, whereas the IA face has a slightly negative charge (acidic) (Adachi et al., 2003). Conversely, the surface of Adzuki 7S1, Adzuki 7S3, and mungbean 8Sα show mainly negative charges on both faces (Fukuda, Maruyama, Salleh, Mikami, & Utsumi, 2008). No reports are available on the surface charge of cruciferin molecules. Molecular surface charge responds to the charged species in the environment. Due to their amphiphilic nature, proteins can assume a number of negative and positive charges in a pH known as isoelectric pH.

Thermal stability is one way that a protein’s molecular stability can be explained. The thermal denaturation temperature ($T_m$) is the point at which the protein structure transitions between native and denatured state. The thermal denaturation temperature ($T_m$) of soy glycinin in dilute aqueous solutions (10%, w/v protein) is around 92–96 °C (Jiang, Xiong, Chen, 2010; Maruyama et al., 2004). Similarly, the $T_m$ for pea prolegumin is 69.3 °C, 72.5 °C for pumpkin Pro-11S, and 74.7 °C for procruciferin (Tandang-Silvas et al., 2010), indicating that among the 11–12S seed globulins the thermal stability of the quaternary or tertiary structures vary. Generally, molecular features, such as the number of surface aliphatic residues, abundance of Pro residues, small cavity size, and short loops, are related to the thermal stability of the protein (Tandang-Silvas et al., 2010). When different subunits of soy glycinin are considered, the $T_m$ values are in the order of $A1bB2 \ll A2B1a \leq A5A4B3 < A3B4 \leq A1aB1b$ (Prak et al., 2005). The pro- or unprocessed
form of 11S globulins have been reported to have lower $T_m$ values and heat stability than the mature form (Maruyama et al., 2004). In the dry state or under very low water activity values, the thermal characteristics of phase transition, such as glass transition temperature ($T_G$), can be observed. The hexameric form of soy glycinin has a glass transition temperature of 220 °C which is relevant when obtaining dry purified proteins, especially through freeze or spray drying (Huson et al., 2011).

The intrinsic fluorescence of the protein indicates the microenvironment of hydrophobic aromatic amino acids, such as Tyr, Trp, and Phe. Surface hydrophobicity of glycinin showed different values, as determined at the subunit level using fluorescent-probe binding (1-anilinonaphthalene-8-sulfonate or ANS probe), indicates group I (A1aB1b + A1bB2 + A2B1a) < A3B4 < A5A4B3 (Tezuka, Yagasaki, & Ono, 2004) and hydrophobic column chromatography indicates group I < A5A4B3 < A3B4 (Maruyama et al., 2004) or A5A4B3 < A1aB1b < A3B4 < A1bB2 < A2B1a (Prak et al., 2005). According to the Tandang-Silvas and coworkers (2010), procruciferin and pumpkin Pro-11S show higher surface hydrophobicity than prolegumin and mature A3B4 glycinin. Protein hydrophobicity is categorized into two types, aromatic and aliphatic (Hayakawa & Nakai, 1985). The hydrophobicity of both aromatic and aliphatic groups on the surface of the protein, called effective hydrophobicity, has been correlated with inter-facial tension, solubility, emulsifying activity, foaming ability, and thermal functional properties of several food proteins, such as bovine serum albumin, β-lactoglobulin, and ovalbumin (Nakai, 1983; Voutsinas et al., 1983).

2.9.2 Functional Properties

Functional properties of a protein are the manifestation of physico-chemical properties under given conditions (Kinsella, 1976). Among these functional properties, solubility is the most critical. The solubility of a protein directly influences its interfacial activities (emulsification and foaming) and hydrodynamic properties (viscosity and heat-induced gelation) (Kinsella, 1976). The degree of solubility of a protein mostly depends on the charge and hydrophobicity of the molecule, which can be modified by extrinsic factors (Damodaran, 2008). Net negative charge on both faces of 7–8S trimeric proteins from Adzuki and mungbean cause repulsion between protein molecules in aqueous solution and keep them in highly soluble form (Fukuda, Maruyama, Salleh, Mikami, & Utsumi, 2008). On the other hand, fairly uniform distribution of positive and negative
patches on the α′ homotrimer of β-conglycinin resulted in large aggregates and it precipitated out of solution due to charge-charge interactions (Fukuda et al., 2008). The extent of surface hydrophobicity correlates inversely with the solubility of the protein (Nakai, 1983; Pande, Ghosh, Banerjee, & Pande, 2010). Hayakawa & Nakai (1985) suggested that aromatic hydrophobicity along with net surface charge plays a more significant role in protein solubility than the aliphatic hydrophobicity. However, Maruyama et al. (2004) found no such relationship between aromatic hydrophobicity and solubility of the soybean glycinin subunits A1aB1b, A1bB2, A2B1a, A3B4, and A5A4B3. Among extrinsic factors, pH and ionic strength are the most important parameters that govern protein solubility. Changes in protein conformation due to denaturation (e.g., induced by heat) also change solubility properties. In general, all 11−12S proteins are soluble in high acidic and alkaline solutions. The isoelectric point (pI) is the pH where protein molecule has no net charge and typically has the lowest solubility. Soybean glycinin has pI of 4.64 (Peng, Quass, Dayton, & Allen, 1984) and solubility is low in the pH range of 4.0–7.5 (µ = 0.08) (Salleh et al., 2002). Cruciferin from B. napus has a pI of 7.25 as determined by isoelectric focusing (Schwenke et al., 1981). The solubility pattern of cruciferin shows two minimum solubility points, one is close to pH 4.0 and the other is near 7.5 (Tandang, Adachi, inui, Maruyama, & Utsumi, 2004a; Tandang, Atsuta, Maruyama, Adachi, & Utsumi, 2005). In another study, minimal solubility of cruciferin is reported between pH 3.5 and 5.7 (µ = 0.08) (Salleh et al., 2002), which is lower than the value reported by Schwenke et al. (1981) for pI. Besides the solubility based estimations of isoelectric point for cruciferin and glycinin, pI values based on methods such as isoelectric focusing are currently lacking. Changes in ionic strength affects to solubility differently in different globulins; cruciferin has lower solubility than glycinin at high ionic strength (µ = 0.5), but higher solubility at lower ionic strength (µ = 0.08) (Salleh et al., 2002). Recently, Wanasundara et al. (2012) reported that cruciferin from different Brassica species including B. juncea, B. napus, and S. alba had poor solubility between pH 4 and 8.

The amphiphilic (i.e., hydrophilic and hydrophobic) nature of protein molecules allows them to be good surface-active agents at various interfaces, such as during emulsion (oil/water) and foam (air/water) formation. Interfacial properties of a protein at the oil/water (O/W) interface or the air/water (A/W) interface depend on the hydrophilic and hydrophobic patches on the molecule’s surface which are exposed to the solvent environment. Salleh and coworkers (2002) showed that cruciferin has lower emulsifying ability at pH 6.0 and 9.0 and also at ionic strengths
of 0.08 and 0.50 than that does glycinin. Furthermore, compared to napin, cruciferin is reported to provide stable O/W emulsions (Wu & Muir, 2008). On the contrary, Krause & Schwenke (2001) studied monolayer adsorption of albumin (2S) and globulin (12S) from B. napus using tensiometry, film-pressure, and Langmuir-Blodgett-techniques, and reported that globulin possesses the lowest emulsifying activity. Furthermore, they concluded that the molecular compactness of globulin compared to albumin creates a barrier for molecular relaxation at the interface and therefore it exhibits low surface activity.

Thermal gelation or heat-induced structure formation is an important functional property in products such as heat set-gels, processed meat and cakes. Soy glycinin has been widely investigated to understand the thermal gelation of SSP (Salleh et al., 2002; Jiang, Xiong, & Chen, 2010; Maruyama et al., 2004; Prak et al., 2005). Globular proteins in solution can generate 3D networks through heat-induced changes such as denaturation, dissociation-association, and aggregation. Thermal stability of the protein is a significant consideration when it comes to heat-induced structure formation. Nakamura, Utsumi, & Mori (1984) established a possible mechanism for heat-induced gel formation by soybean glycinin. They postulated that gel formation involves thermal denaturation of proteins followed by soluble aggregate formation and interactions to form a network, i.e., denaturation → aggregation → gelation (Nakamura et al., 1984). Although the exact mechanism for cruciferin gelation has not been elucidated, as an 11S protein it may follow a similar mechanism as proposed for glycinin. According to the work of Salleh et al. (2002), cruciferin forms an opaque and non-elastic gel, whereas glycinin forms a transparent and elastic gel at pH 7.6 ($\mu = 0.5$).

Digestibility of a protein is a measure of accessibility of digestive enzymes and the ability to release nutritionally available amino acids and peptides in a biological system. For a food or feed, the protein digestibility is an important factor affecting its nutritional quality. According to the studies of Tandang et al. (2004a) and Bos et al. (2007), the relatively low digestibility of native cruciferin may reflect inaccessibility of the enzyme to scissile sites. A human feeding study of rapeseed protein (36.8% globulin, 41.0% napin and 2.7% lipid transfer protein) also confirmed a low level of in vivo digestibility compared to soy or pea protein (Bos et al., 2007).
2.9.3 Biological Properties

Biological activities of a protein include bio-catalytic (enzymatic) activities, immunogenic properties, and any of the properties in the domain of the senses (taste, smell). Cruciferin has not been widely investigated for these properties, other than a few studies on enzyme inhibitory activity of key enzymes in the Renin-angiotensin system (related to blood pressure regulation) by peptides generated from proteolysis and for potential allergenicity. Hydrolysates of canola cruciferin and napin exhibit potent ACE inhibitor activity (IC$_{50}$) of 35 and 29 µg mL$^{-1}$, respectively (Wu & Muir, 2008). The cruciferin of *S. alba* (yellow mustard) is identified as a human gastro-intestinal allergen, Sin a 2 (Palomares et al., 2005), but weaker than that of peanut allergens Ara h 1 (7S globulin) (IC$_{50}$: 1.5–10 µg mL$^{-1}$) and Ara h 2 (2S albumin) (IC$_{50}$: 4–15 ng mL$^{-1}$) (Mondoulet et al., 2005).

2.9.4 Modification of 11S Protein for Enhanced Properties

Protein functionality can be changed by altering selected features of the molecular structure and to a certain extent by modifying the chemical composition of the solvent. Denaturation of pure 11S globulins at extremely low or high pH values is well known. The 11S globulins of fava bean, pea, soybean, and rapeseed are over 90% insoluble at low ionic strengths ($\mu = 0.0$–$0.1$) at pH 4.0–6.5, but changing the pH to below 4 or above 8 makes these proteins over 90% soluble (Baniel, Caer, Colas, & Gueguen, 1992; Kimura et al., 2008; Salleh et al., 2002). 11S globulins dissociate into 7S trimers under acidic (~pH 2) conditions (Adachi et al., 2003; Plietz et al., 1983; Sokolova, Kealley, Hanley, Rekas, & Gilbert, 2010), whereas, oligomerization occurs at the extremely high pH to form large (soluble) aggregates (Jiang, Chen, & Xiong, 2009; Sokolova et al., 2010). In addition, grafting of reactive residues or molecules can also modify the conformation as well as the reactivity of the protein. Grafting of acyl residues, such as acetyl or succinyl, onto cruciferin causes swelling and unfolding of the molecule, leading to a much greater surface hydrophobicity and different functionalities (Gerbanowski, Malabat, Rabiller, & Guéguen, 1999). Moreover, phosphorylation of 12S globulin from rapeseed at pH 10–11 at 20 °C changes the surface charges of the molecule leading to increased surface hydrophobicity and generation of a gel-forming biopolymer (Schwenke, Mothes, Dudek, & Görnitz, 2000).

The loop regions of 11S proteins, including the HVRs, are important for biological (e.g., immunogenicity) and physico-chemical properties (Tandang et al., 2005). This was demonstrated
by the significant improvement of emulsifying ability of proglycinin proA1bB1b achieved by removing its hydrophilic HVR (Kim, Kamiya, Sato, Utsumi, & Kito, 1990b) and attaching a more negatively charged extension region (α′ ext.) from β-conglycinin (Tandang et al., 2005). Insertion of extension regions (α, α′, and A4IV) to the C-terminus of glycinin A1bB1 (A1b1bα, A1b1bα′, and A1b1bA4IV, respectively) caused much better emulsifying ability than the native form (Prak et al., 2005). Improved gelling properties have been achieved by disrupting disulfide bonds of proA1bB1 (C12G and C88S) (Utsumi et al., 1993a). Similarly, when procruciferin Cru2/3a had a charged HVR inserted (from ΔIV + A1aIV and ΔIV + A3IV of glycinin), soluble aggregate formation upon heating was enhanced (Tandang, Adachi, & Utsumi, 2004b). Therefore, the solubility of the native procruciferin at 70 °C (µ = 0.08) can be improved by adding variable regions ΔIV + A3IV, ΔII, or C287T/ΔII (Tandang et al., 2004b). The modification of procruciferin by inserting the A1aIV region of soybean glycinin also improved α-chymotrypsin digestibility (Tandang et al., 2004b); therefore, the HVR of 11S proteins may be important to consider in understanding and modifying functional properties.

2.10 Summary

Oilseeds of the Brassicaceae (Cruciferae) family, including canola and mustard, accumulate considerable amounts of protein in addition to oil. Among the SSPs, cruciferin is the most abundant, comprising ~60% of total protein and contributing much to the nutritional and functional attributes the total protein deliver. Cruciferin, which is a large 11S globulin protein composed of six subunits, has not been investigated to the extent that soy glycinin.

Past studies on B. napus cruciferin indicated that it may possess inferior properties that may limit its use and the development of new products compared to soybean glycinin. The relationship between the molecular structure and function of cruciferin has not been resolved at the subunit level. In the wild type cruciferin, the subunits are not identical. Detailed probing of each subunit’s (protomer) primary, secondary and tertiary structure and the assembled quaternary structure will provide insight into parameters and characteristics that lead to the functional properties of the respective subunits. For such studies, it is essential to have cruciferin composed of identical subunits, which is usually achieved through expression in heterologous systems, mainly yeast or bacteria. Mutant Arabidopsis lines producing cruciferin composed of homogenous subunits (i.e., double-knockouts) have been developed by Agriculture and Agri-

27
Food Canada scientists. The seeds of these mutants are expected to accumulate cruciferin composed of same subunit and this protein is suitable to study the structure and properties of homohexameric cruciferin. Understanding of the structure-function relationships of cruciferin homohexamers will provide information for functional enhancement, designing protein modifications, and formulating new uses for cruciferin.

This thesis research was conducted under five studies to test the hypotheses and to meet objectives outlined above.

**Study 1**  *In silico* homology modeling to predict functional properties of cruciferin homotrimer.

**Study 2**  Synchrotron FT-IR analysis of mutant Arabidopsis seeds containing different seed storage proteins.

**Study 3**  Structural characterizations of isolated homohexameric cruciferins from mutant Arabidopsis lines and the effects of pH and salt on structure stability

**Study 4**  Investigation of solubility, heat-induced gelation, and pepsin degradability of cruciferin homohexamers

**Study 5**  Investigation of the emulsifying properties of cruciferin homohexamers in oil-in-water emulsions.

The research results obtained from these studies are organized into five manuscripts as follows.
3. IN SILICO HOMOLOGY MODELING TO PREDICT FUNCTIONAL PROPERTIES OF CRUCIFERIN*

3.1 Abstract

Cruciferin is the major storage protein in Brassicaceae family oilseeds. The predominant cruciferin isoforms in Arabidopsis were investigated using homology modeling (HM) for their molecular structures and functional properties. The structure of Brassica napus procruciferin was used as the template for HM to determine the molecular structures and hypervariable regions.

Hydrophobicity and electrostatic surface potential distribution on the intradisulfide–containing face (IA) and the interdisulfide-containing face (IE) indicated favorable interfacial and solubility properties. More extensive heat-induced structural changes were predicted for the CRUC homotrimer than for the CRUA or CRUB homotrimers. Structural features that facilitate flavor binding and limit proteolytic digestion were more abundant in CRUA and CRUB than in CRUC.

On the basis of these comparative models, structural differences among cruciferin isoforms and their relevance to potential techno-functionalities were identified. This approach to functional property prediction will link protein structure to utilities and will be valuable in designing proteins for targeted applications.


3.2 Introduction

The physical and chemical properties of proteins that influence their functional behavior in multicomponent systems include size, shape, amino acid composition and sequence, net charge, charge distribution, hydrophobicity, hydrophilicity, structural organization (primary, secondary, tertiary, and quaternary structures), molecular flexibility, and rigidity. These molecular parameters influence how proteins respond to the environment (e.g., temperature, pH, pressure, and ions) and interact with other chemical constituents in a complex system such as food (Kinsella, 1976; Damodaran, 2008). Chemical, physical, or enzymatic modification of a protein has been the conventional approach to improve functionality. Understanding the relationship between protein structure and functional properties provides new opportunities to improve protein functionality at the molecular level. Indeed, quantitative structure–activity relationship (QSAR) studies have advanced the understanding of structure–functional property relationships of food proteins (Nakai & Li-Chen, 1988; Nakai, 2004). This approach relies primarily on predicting the functional properties of a protein on the basis of the properties of its component amino acids, that is, the hydrophobicity, side-chain bulk, and electrostatic properties of their side groups. To understand how a protein functions in food, knowing how the amino acids are distributed in three-dimensional space is also necessary. Such information may be obtained from spectroscopic techniques (circular dichroism, fluorescence, Fourier transform-infra red, Raman, and ultraviolet spectroscopy) that provide global information on functional groups, secondary structure, domains, and residues or from methods that provide sequence-specific information (high resolution NMR and X-ray crystallography). For food proteins, tertiary structure information is often limited because of the difficulty in obtaining purified homogeneous proteins. Homology modeling (HM), or comparative modeling, predicts the three-dimensional (3D) structure of a protein using its primary sequence (target) by comparing it to one or more related proteins with known structures (template). This approach is proving to be valuable in discovering new pharmaceuticals (Liebman, 1998) and identifying allergenic proteins (Schein, Ivanchuc, & Braun, 2007).

Cruciferin, an 11S globulin of the cupin (small β-barrel) superfamily, is the most abundant SSP found in crucifers. Cruciferin is a hexamer that is assembled as two trimers comprised of heterogeneous subunits (protomers); for example, in Arabidopsis CRUA, CRUB, and CRUC protomers are found. Each subunit is comprised of two polypeptides, the heavy α-chain (acidic,
254–296 amino acids) and the light β-chain (basic, 189–191 amino acid residues) that are linked with one disulfide bond. Heterogeneity caused by the random assembly of different subunits in the hexamer has made determination of cruciferin quaternary structure difficult.

Understanding the physico-chemical and functional properties of cruciferin is important because of the considerable amount of protein biomass (22–24% seed dry weight is protein and cruciferin is ~60% of total protein) generated from industrial (Brassica carinata and Camelina sativa) and food-grade (B. napus, B. rapa, canola quality B. juncea and Sinapis alba) crucifer oilseed crops. Molecular hydrophobicity, flexibility, and stability are the key features in determining solubility, adhesiveness, film forming ability, thermal denaturation, and interfacial properties in emulsions and foams, which are important techno-functionalities in food and industrial applications of proteins (Kinsella, 1976; Wanasundara, 2011).

Previous studies on the functional properties of cruciferin have been conducted with the procruciferin form of Cru2/3a expressed in E. coli (Tandang et al., 2004b; Tandang et al., 2005). The objective of this study was to develop an in silico approach to identify differences in the structure of cruciferins that could be used to predict the physico-chemical and functional properties of the protein. This study used Arabidopsis cruciferin because this plant provides a model for crucifer oilseeds and has vast amounts of bioinformatics data available. I used HM to obtain molecular structures of cruciferin consisting of homogeneous subunits encoded by each of the three Arabidopsis cruciferin genes. The structural features of the molecule that are critical in determining techno-functional properties were evaluated, and possible functions were predicted using bioinformatics tools and data on other globular proteins.

3.3 Materials and Methods

3.3.1 Sequence Analysis and Phylogenetics

The homologues of full-length sequences of all Arabidopsis cruciferins were obtained from the Arabidopsis Information Resource 9 (TAIR9) protein database. Phylogenetic analysis was performed using PhyML (http://atgc.lirmm.fr/phyml/). Sequence alignments were performed using MUSCLE (http://www.drive5.com/muscle/) and a neighbor-joining and maximum-likelihood phylogenetic tree constructed using LG matrix as the amino acid replacement model. Statistical support for the phylogeny trees is given as posterior probability (Bayesian)/bootstrap values based on 100 replicates.
3.3.2 Protein Homology Modeling

The structure of the cruciferin isoforms was predicted using software available through the SWISS-MODEL server (http://swissmodel.expasy.org/). Full-length amino acid sequences of Arabidopsis cruciferin subunits CRUA, CRUB, and CRUC were examined. The best alignment templates for cruciferin protomers were identified using the “template identification” tool in the SWISS-MODEL program. The following limits were employed: (1) for the InterPro domain, HMMPfam, HMMTigr, ProfileScan, SuperFamily, BlastProDom; (2) for the gapped blast query, threshold expectation value \((E\) value) \(10^{-6}\), protein substitution matrix BLOSUM62 \(1(G)\) \(1(E)\), and the number of alignments truncated to 25; (3) for the iterative profile BLAST, stage 1 = PSI-BLAST profile generation versus NR, iteration 2, \(E\) value \(10^{-4}\), protein substitution matrix BLOSUM62 \(1(G)\) \(1(E)\), SEG filter for low complexity subsequence activated, and stage 2 = PSI-BLAST profile search, search data base ExPDB90, \(E\) value \(10^{-5}\), protein substitution matrix BLOSUM62 \(1(G)\) \(1(E)\), alignment 50; HHSearch software with default parameters was used to detect distantly related templates. The crystal structure of the procruciferin Cru2/3a subunit (PDB; ID 3KGL, chain A), which was obtained from X-ray crystallography, was used as the template and had the best alignment parameters for all three cruciferin protomers, mostly due to the close phylogenetic relationship between species within the Brassicaceae family (Figure 3.1).

The 3D structures of the Arabidopsis cruciferins CRUA, CRUB, and CRUC were modeled on the SWISS-MODEL server using the program SWISS-PDB Viewer to submit the projects (Kaplan & Littlejohn, 2001). Hypervariable regions (HVRs) were constructed using the ModLoop web server (http://modbase.compbio.ucsf.edu/modloop/). The procedure to build the HVRs used the main-chain and side-chain dihedral angles as well as non-bonded atom pair restraints from many known protein structures. This created loops with regular secondary structure elements that were corrected by employing conjugate gradient minimization and molecular dynamics simulation with simulated annealing. Thereafter, theoretical models were submitted to SWISS-PDB Viewer for further energy minimization (GROMOS96 43B1 force field, 200 cycles of steepest descent) to correct the stereochemistry of the models. The energies calculated for the derived models were CRUA = \(-25825.7\) kJ mol\(^{-1}\), CRUB = \(-23834.2\) kJ mol\(^{-1}\), and CRUC = \(-29840.2\) kJ mol\(^{-1}\). Final structures were selected after evaluation of the stereochemistry of the 3D models using \(Z\) score (ProSA: https://prosa.services.came.sbg.ac.at/prosa.php), \(G\) factor values, and Ramachandran phi-psi plots.
generated through PROCHECK (Laskowski, MacArthur, Moss, & Thornton, 1993). The models and loop regions were verified using Verify3D (http://nihserver.mbi.ucla.edu/Verify_3D/) to identify regions of improper folding. Structures were visualized using PyMol (Warren L. DeLano, DeLano Scientific, San Carlos, CA, http://www.pymol.org); UCSF Chimera (http://www.cgl.ucsf.edu/chimera/) and/or visual molecular dynamics (VMD, http://www.ksuiuc.edu/Research/vmd/).

3.3.3 Analysis of Structural Properties
Multiple sequence alignments were developed using ClustalW2 (Larkin et al., 2007). Evolutionary sequence-conservation scores or surface variability were calculated using ConSurf (http://consurf.tau.ac.il/) using a multiple sequence alignment generated using MUSCLE. Phosphorylation sites were predicted using NetPhos version 2.0 (http://www.cbs.dtu.dk/services/NetPhos/) and were mapped onto the 3D models. The cavity volume and active-site entrance (i.e., mouth) area were calculated using CASTp (http://cast.engr.uic.edu/cast/).

3.3.4 Analysis of Surface Properties
Hydrophobic and hydrophilic residues were visualized on the molecular surface of Arabidopsis cruciferin using VMD software. Electrostatic surface potentials of the molecules were calculated using the Adaptive Poisson–Boltzmann Solver (APBS) (Baker, Sept, Joseph, Holst, & McCammon, 2001) plug-in developed by Michael G. Lerner (University of Michigan) for PyMol. The following parameters were used for all calculations: protein dielectric constant of 2, solvent probe radius of 2.0 Å, solvent dielectric constant of 80, and temperature of 310 K. Limited proteolytic sites or nick sites were predicted using NICKPRED (Hubbard, Beynon, & Thornton, 1998) available from Dr. Simon Hubbard (University of Manchester, Manchester, U.K.). The conformational parameters were set at 0.6 for accessibility (relative) with a window size of 12, at 0.0 for protruding index (α-carbon atoms) with a window size of 20, at 0.7 for temperature factor (main chain atoms) with a window size of 4, at 0.4 for Ooi number (Ooi14) with a window size of 10, at 0.2 for secondary structure (CYS = 0.1, C = 1, H = 0.5, E = 0) with a window size of 14, and 0.4 for main-chain hydrogen bonding (outside loop 6) with a window size of 20.
**Figure 3.1** Selection of a template for homology modeling of Arabidopsis cruciferins. (A) Phylogenetic relationship of three Arabidopsis cruciferins with protein sequences that best fit templates according to the gapped-BLAST and PSI-BLAST search. Sequences were aligned with MUSCLE and a neighbor-joining tree and maximum likelihood phylogenies implemented with PhyML. Support for the clades is indicated by posterior probability (Bayesian)/bootstrap values based on 100 replicates and the four families marked in different colors. The respective PDB codes are given next to the taxon name. (B) Alignment parameters of main Arabidopsis CRU protomers (targets) with *B. napus* procruciferin template (3KGL). These results, combined with phylogenic data, suggest that rapeseed procruciferin (Cru2/3a) should be the best template to predict the structure of Arabidopsis cruciferins.
The program considered only the primary specificity at the P1 site (Schechter and Berger notation) when predicting the nick sites for each protease. Enzyme cleavage sites were predicted for pepsin (cleaves at P1 = NOT Val, Ala, Gly), trypsin (cleaves at P1 = Lys, Arg), and chymotrypsin (cleaves at P1 = Trp, Tyr, Phe, Met, and Leu) for cruciferin homotrimers.

3.4 Results and Discussion

3.4.1 Selection of Cruciferin Protomers

The Arabidopsis genome contains four genes encoding cruciferin or cruciferin-like proteins: \textit{AT5G44120} (CRA1), \textit{AT1G03880} (CRU2 or CRB), \textit{AT4G28520} (CRU3 or CRC), and \textit{AT1G03890} (putative cruciferin). Four mRNA splice variants for \textit{AT4G28520} and three splice variants for \textit{AT5G44120} have been detected, resulting in a potential repertoire of nine cruciferin proteins. Phylogenetic analysis revealed that \textit{AT1G03890.1} formed a distinct lineage from cruciferins found in Arabidopsis, \textit{B. napus}, \textit{Sinapis alba}, and \textit{Raphanus sativus} with significant bootstrap support (Appendix Figure A1). The cruciferin derived from \textit{AT1G03890.1} also has an acidic isoelectric point (pI 5.45) that is different from other cruciferin isoforms (pI 6.51–9.08) (Appendix Table A1). Despite these differences, the complete structure of \textit{AT1G03890.1} protein could be modeled (Appendix Figure A1). Wan et al. (2007) reported that in Arabidopsis, the expression of \textit{AT1G03890} was much lower than that of the other cruciferin genes, based on MS analysis of total seed proteins. In the present study, proteins encoded by all splice variants from \textit{AT4G28520} and \textit{AT5G44120} were modeled. Among these, only the protein encoded by \textit{AT4G28520.1} (hereafter referred to as CRUC) was able to generate the complete 3D structure. The proteins encoded by \textit{AT4G28520.2–4} had deletions in the β (basic) subunit, whereas that encoded by \textit{AT4G28520.3} did not have an inter-chain (IE) disulfide bond (Appendix Figure A1 and Table A1). Similarly, deletions in the α (acidic) chain of proteins encoded by the alternatively spliced transcripts of \textit{AT5G44120}, except \textit{AT5G44120.3} (hereafter referred to as CRUA), were observed. Moreover, the protein derived from \textit{AT5G44120.3} is the only product with a signal peptide, which is required for entry into the endomembrane system and subsequent transport to storage vacuoles (Appendix Table A1). A complete model of the protein encoded by \textit{AT1G03880.1} (hereafter referred to as CRUB) could be generated. On the basis of previously published data (Wan et al., 2007) and structural evidence of the present homology modeling study, three protomers, CRUA, CRUB, and CRUC, encoded by the gene models \textit{AT5G44120.3},
AT1G03880.1, and AT4G28520.1, respectively, were selected to construct homotrimers, to study their structural properties, and to predict techno-functional properties.

### 3.4.2 Geometrical Quality of Homology Models

In general, the accuracy of homology models relies mainly on the similarity between the template and the target sequence. In this study, the full-length amino acid sequences of Arabidopsis cruciferin protomers CRUA, CRUB, and CRUC were used for homology modeling. Adequate sequence identity values were observed between the protomers, 85.3% (CRUA), 75.8% (CRUB), and 57.6% (CRUC), and the template procruciferin (Cru2/3a, PDB code 3KGL) from *B. napus* (Figure 3.1B and Figure 3.2A). The low sequence identity between CRUC and Cru2/3a was in accordance with the phylogenetic analysis (Appendix Figure A1). This was primarily due to differences in the hypervariable region I (HVR-I), also known as the disordered region, which contained repeats of Gly and Gln in CRUC (Figure 3.2A).

In earlier studies, the structure of procruciferin HVR-I or HVR-II could not be resolved by crystallography (Tandang-Silvas et al., 2010). In regard to the biological, physiological, or food-related functions, the most important regions may be the loops, including HVRs, because these regions lack defined secondary structure, such as sheets or helices, and most importantly protrude from the compact globulin structure (Adachi, et al., 2003; Tandang-Silvas et al., 2010). Therefore, the present study placed more emphasis on HVRs by reconstruction followed by energy minimization using the GROMOS 96 empirical force field to make further stereochemical corrections to these regions. Ramachandran plots generated for cruciferins confirmed the geometric correctness of the backbone and the percentage of residues in the most favored (core), and additionally allowed regions were 99.1% for CRUA, 99.4% for CRUB, and 100% for CRUC (Figure 3.3A,D). The $G$ factor, calculated by PROCHECK program, indicates stereochemical correctness of a given structure, and the acceptable values are between 0 and −0.5, with the best models displaying values close to 0. Overall, the $G$ factors for the cruciferin models were −0.14 for CRUA, −0.12 for CRUB, and −0.11 for CRUC, suggesting that the models had “normal” torsion angles, covalent geometries (Figure 3.3B,C) and good stereochemical quality. The energy profiles obtained from the Verify3D program were used to validate the correctness of the HVRs in the models.
Figure 3.2 Sequence and structure conservation of cruciferin protomers. (A) Multiple sequence alignment (ClustalW) and secondary structure assignment (PROCHECK) of Arabidopsis cruciferins (AtCRUA, AtCRUB, and AtCRUC) with *B. napus* procruciferin. Identical (black highlight) and conserved (gray highlight) amino acids are indicated. Secondary structure is displayed above the sequences: α-helices (red cylinders), 3_10 helices (pink cylinders), strands (arrows) with the standard labeling for 11S globulins (Tandang-Silvas et al., 2010), disordered regions are shown as dots (signal peptide: green, HVR-I: light violet, HVR-II: dark violet, and N/C terminal disordered regions: light blue); extended loop region (yellow). Green bars indicate residues involved in hexamer formation. Residues omitted in the template coordinates are in purple boxes. Intra-protomer (IA) and inter-protomer (IE) disulfide bonds are indicated by yellow lines connecting cysteine residues. Free and available cysteine residues are shown in black letters in the yellow boxes. The internal proteolytic possessing site is indicated by a blue arrow head. (B) Superimposition of homology models (CRUA: blue, CRUB: green, and CRUC: red) with *B. napus* procruciferin (PDB code 3KGL: violet). Hypervariable region I and II are indicated inside the blue ellipses. (C) Surface variability of Arabidopsis cruciferin models derived from a phylogenetically diverse alignment and mapped using Consurf showing high residue variability in HVRs and conserved areas in the core β-barrels.
This procedure measures the compatibility of a 3D protein model with its primary amino acid sequence (1D), and a compatibility score >0 corresponds to an acceptable side chain environment.

The energy profiles for the cruciferin model structures ranged from 0 to 0.71. The 3D–1D profile score dipped below 0 only at two points in the CRUB model (Ile199 and Leu207) and one point in the CRUC model (Ile252) (Figure 3.3D). The CRUA model 3D–1D score was >0.1 for HVR-I and >0.2 for HVR-II, whereas in the CRUB and CRUC models this value ranged from >0.5 to >0.3 and >0.3 to >0.2, respectively. These values suggest that loops in our homology models are properly constructed. The Z score is an indicator of overall model quality, which is based on the similarity of the model to all NMR, and X-ray crystallographic determined structures available. The Z scores of the cruciferin models were within the range of scores typically found for native proteins of similar size (Figure 3.3C). Superimposition of the modeled structures on procruciferin showed a close backbone relationship (<0.5 Å), which was indicated by a root mean square deviation (RMSD) of 0.27 Å, 0.42 Å, 0.42 Å for the C\text{α} atoms between procruciferin and CRUA, CRUB, and CRUC, respectively (Figure 3.2B and Figure 3.3D).

3.4.3 Comparison of 3D Structure Features

Each modelled protomer had 27 β-sheets, 6 α-helices, and 3 3_{10} helices. An extra-long loop region on the intra-chain disulfide bridge-containing face (IA face) corresponded to the extended loop region (ELR) located in acidic chain (procruciferin, Asn171–Ile191; CRUA, Asn197–Ile217; CRUB, Asn190–Ile210; and CRUC, Asn251–Leu220), which has been reported to be in a loop conformation rather than a disordered region in the procruciferin protomer because of the clear electron density map seen in the crystal structure (Tandang-Silvas et al., 2010). The backbones of CRUA and CRUB deviated about 1.4 Å from CRUC in the ELR. This is due to a deletion in CRUC, as well as substitution of the Trp residue of CRUA or CRUB at positions of 271 and 264, respectively, with Phe residue at corresponding position 323 in CRUC (Figure 3.4).

Because the ELR is situated on the IA face where the intra-chain disulfide bonds of each cruciferin trimer or hexamer are located, any divergence in this region may influence the physiochemical properties of the molecule and therefore its functional properties in a complex food system.
Figure 3.3 Stereospecific quality evaluations of the cruciferin homology models and procruciferin crystal structure. (A) Ramachandran plots generated with the RAMPAGE web server. Amino acid residues of modeled structures are not less than 99.1% in most favored (core) and allowed regions. (B) Plots obtained using Verify3D with a window of 21 residues. Inserted protein structure shows the location of hypervariable region (HVR) I and II of each protomer. The vertical gray bars in the plot correspond to the HVRs found in the protomer domain. (C) The overall model quality, Z score, as predicted using the ProSA web server, indicates the models are within the range of scores typically found for native proteins of similar size. (D) Model quality statistics, root mean square deviation, $G$ factor, Ramachandran scores and Z score, in comparison with the same parameters for procruciferin.
Of the hypervariable regions I and II that are located on the IE face (inter-chain disulfide bridge containing face), HVR-I is directed towards the periphery, whereas HVR-II projects into the core region of the molecule (Figure 3.4A,B,C). The HVR-I of CRUB is the shortest (7 residues), whereas that of CRUC is the longest (60 residues). In CRUA, the HVR-I loop region had two short antiparallel strands (Pro124–Gln134) separated by a small turn (Gly128–Gln129) (Figure 3.4A). Insertion of several GQ repeats has increased the length of HVR-I in CRUC (Figure 3.4A). The secondary structure prediction of this region indicated that, other than a small helix (residues 170–175), it is mainly comprised of loop-like turns, suggesting that it has a flexible structure. Although HVR-II of each protomer did not show any higher order structure, it is believed that the conformation of HVR-I and HVR-II plays a vital role (i.e., steric hindrance) in hexamer formation (Adachi, et al., 2001, 2003; Tandang-Silvas et al., 2010).

During hexamer formation, which occurs via interaction of the IE faces of two trimers in conjunction with proteolytic processing at the $\beta$-cleavage site (Asn–Gly as shown in Figure 3.2A) of HVR-II, the flexible nature of HVR-I helps it to move to the periphery of the molecule, allowing the IE faces to come into close proximity (Adachi, et al., 2001, 2003; Tandang-Silvas et al., 2010). Four highly conserved trimer binding sites were identified among the three cruciferin isoforms (Figure 3.2A). These binding sites are very important for assembly of the hexamer as they contribute to the formation of hydrogen bonds and hydrogen-bonded salt bridges (Adachi, et al., 2001, 2003; Tandang-Silvas et al., 2010).

Residues on the surface of a protein that readily interact with the native environment are less likely to be conserved than residues embedded within the protein or in surface-exposed cavities (Guo, Choe, & Loeb, 2004). This phenomenon is further supported in the present study by showing that the highly variable HVRs and ELRs of all cruciferin protomers are present on the surface, whereas most of the residues contributing to the core cupin domain located towards the channel in the center of the trimer (inside view of Figure 3.2C) are conserved. Residues on the IA face and the periphery of the molecules were less conserved (outside view of Figure 3.2C), suggesting that proteins with different protomer compositions may have different molecular surface properties.
Figure 3.4 Three main loop regions in Arabidopsis cruciferins modeled with ModLoop. (A) Hypervariable region I. CRUC has the longest peptide chain with one helix between residues 170 and 175 located on IE face. CRUA has an antiparallel $\beta$-sheet, whereas CRUB has a simple loop structure. (B) HVR-II. (C) Extended loop region located on the IA face. The backbone of CRUC deviated 2.4 Å from the backbone of its counterparts, CRUA and CRUB, due to a deletion at position 251 and a substitution of Trp by Phe at 252. The substituted residue is shown as a stick model. (D) Location of HVRs and ELR (highlighted areas) on both faces of the molecular surface of cruciferin homotrimeric (CRUA, CRUB, and CRUC).
3.4.4 Surface Hydrophobicity and Related Properties and Protein Structures

The fact that proteins are amphiphilic molecules makes them good surface-active agents at the interface of emulsions (oil/water; O/W) and foams (air/water; A/W). The hydrophobic residues on the protein surface are strongly adsorbed to the interface between water and oil (in emulsion) or air (in foam), causing a pronounced reduction of interfacial or surface tension, which facilitates formation of oil droplets in the emulsion or air bubbles in the foam (Kato, 1991). It is not possible to predict this activity solely on the basis of the polarity of the constituent amino acids. However, it is well-established that the emulsification behavior of globular proteins is influenced by their solubility, surface hydrophobicity, and molecular flexibility, whereas foam-forming ability is mainly governed by surface hydrophobicity, viscosity, and the net charge of the protein (Nakai & Li-Chan, 1988).

In this study, distribution of hydrophobic residues on the molecular surface of cruciferins was compared (Figure 3.5). In general, the IE face contained a higher number of hydrophobic residues than the IA face, and they were more accessible to the external environment (Figure 3.5 and Figure 3.6A). Among the three cruciferins, CRUC had more hydrophobic residues on the IE surface. It was also observed that the IA face of CRUA and CRUB contained more hydrophobic patches than the IA face of CRUC (see Figure 3.5 and Figure 3.6A). In another study, procruciferin from B. napus was reported to have a higher surface hydrophobicity value than other 11S globulins, such as Pisum sativum prolegumin or Glycine max proA1aB1b, proA3B4, and mature A3B4 (Tandang-Silvas et al., 2010). Glycinin, the 12S globulin from soybean (G. max) is considered to have a closely packed globular conformation and exhibits less surface hydrophobicity and molecular flexibility, which limits the foaming and emulsifying properties of the protein (Tandang et al., 2005). The three main loops (HVR-I, HVR-II, and ELR) of cruciferin possessed a large number of polar or hydrophilic groups, relative to other regions of the molecule, and consisted mainly of Gly and Gln residues (Figure 3.6A,B).

The native cruciferin is a hexamer, where two trimers interact face-to-face at the IE surfaces to form a hexamer (Adachi et al., 2003); therefore, the IE face is not available for interactions with the solvent phase. Hexameric CRUA and CRUB contain only a few prominent protruding hydrophobic residues in the ELR (i.e., Try203/Val202 in CRUA and Try197 in CRUB), and these are scattered as doughnut-shaped rings on the IA face. The hydrophobicity profile on the IA face of CRUC was distinctly different from other two (Figure 3.5).
Figure 3.5 Surface characterization of Arabidopsis cruciferin models. Distribution of hydrophilic and hydrophobic residues is represented in purple and yellow, respectively, on white molecular surface models (top row in each set). Molecular surfaces of cruciferin models are colored according to electrostatic potential values ranging from +6 kT/e (blue) and –6 kT/e (red) (bottom row in each set).
Figure 3.6 Distribution of hydrophobic and hydrophilic residues of Arabidopsis cruciferins. (A) Number of hydrophobic residues on different regions of Arabidopsis cruciferin CRUA, CRUB, and CRUC. Surface exposed hydrophobic residues were counted using VMD software. (B) Alignment of cruciferin sequences occurring at the hypervariable and extended loop regions color-coded according to residue hydrophobicity (hydrophilic: purple with intensity increasing with degree, hydrophobic; yellow). Identical (*) and conserved (., and :) residues are indicated below the alignments. (C) Kyte-Doolittle hydropathy plot (Kyte & Doolittle, 1982) for cruciferins. Each amino acid residue had a variable hydropathic index ranging from negative (hydrophilic) to positive (hydrophobic). The plot was created from each value of hydropathic index along the amino acid sequence of Arabidopsis cruciferins (CRUA, CRUB, and CRUC) and B. napus procruciferin. Hyper-variable regions I and II and ELR are denoted by magenta boxes.
This spatial arrangement indicates that CRUA and CRUB hexamers may have stronger interfacial activities than CRUC and may be better emulsifiers for oil-in-water (O/W) interfaces and/or foam formers. The central channel (core region) of the molecule also has hydrophobic patches at the IA face of the CRUA and CRUB homotrimers mainly due to the presence of Phe32 and Trp26 residues, respectively, whereas the CRUC homotrimer has a hydrophilic core surface due to the presence of Gln32, which covers the respective hydrophobic residue Leu33 (also, Leu is smaller compared to the Phe or Trp, Figure 3.5). However, the largely hydrophilic nature of the CRUC homohexamer may cause a considerably different behavior in stabilizing the O/W interface than the other two cruciferin species. According to Adachi and co-worker (2003), slight destabilization of the cruciferin structure by changing ambient pH or ionic strength causes the hexamer to separate into two trimers. Under such circumstances, the buried hydrophobic residues of all the cruciferin isoforms may be exposed, leading to increased surface hydrophobicity with consequent increase in emulsifying and foaming abilities.

In the emulsification or foaming processes, the protein molecule initially diffuses to the contact face between the two phases, where it consequently unfolds and the hydrophobic residues penetrate into the oil or air phase to minimize the free energy of the protein molecule. As a consequence of this unfolding and mobilization, a protein film is formed around the fat globule in an O/W emulsion or an air bubble in foam (Damodaran, 2008; Kato, 1991). Apart from the molecular hydrophobicity, it is believed that the molecular flexibility at the interface is essential for a high-performing emulsifying or foaming agent (Kato, 1991). The inherently poor conformational flexibility and the compact structure of globulins have been reported to underlie the poor emulsifying properties of B. napus cruciferin (Vioque, Sánchez-Vioque, Clemente, Pedroche, & Millán, 2000). Alternatively, cruciferin HVRs may be more mobile due to the lack of ordered secondary structure (Adachi, et al., 2001, 2003; Tandang-Silvas et al., 2010). Addition of a negatively charged region at the C-terminus [A1aIV of glycinin (Maruyama et al., 2004) and extension region of β-conglycinin (Maruyama et al., 1999)] improved emulsion droplet size and the stability of procruciferin (Cru2/3a) (Tandang et al., 2005). Although HVRs have more hydrophilic residues, partial unfolding of these regions and their high flexibility could result in some hydrophobic residues being projected into the solvent phase. These residues include Phe122, Phe126, and Phe135 in HVR-I and Pro267 in HVR-II of CRUA; Phe116, Phe128 in HVR-I and Pro264, Ile267, and Ala268 in HVR-II of CRUB; Pro131, Trp132, Pro142, Trp143,
Trp160, Pro178, and Trp179 in HVR-I and Pro318, Trp324, Pro327, and Pro330 in HVR-II of CRUC (Figure 3.7A). Therefore, it can be assumed that these hydrophobic groups may be pivotal in establishing the interfacial properties of cruciferin A, B, and C protomers. The CRUC protomer has the longest HVRs (60 and 16 residues in HVR-I and HVR-II, respectively), and also contains a considerable number of hydrophobic residues along this segment of polypeptide chain (14 residues) that can orient apolar R-groups to the oil or air phase and increase the interfacial activities required for emulsification or foaming.

The kinetic stability of protein-coated oil droplets or air bubbles in an emulsion or foam is also an important techno-functional property so far as surface-active properties are concerned. It has been proposed that the hydrophobicity of the protein is a major factor for emulsion formation, whereas electrostatic repulsion serves to stabilize the emulsion or foam by preventing coalescence or drainage/burst of the emulsion oil droplets or foam air bubbles, respectively (Kato, 1991). Therefore, I mapped the electrostatic potential on solvent accessible surfaces of the cruciferins using the Poisson–Boltzmann equation, which describes the electrostatic interactions between molecules in ionic solutions and constitutes the most fundamental approach to predict electrostatic effects of a molecule in a solution (Figure 3.5). Similar to the finding of Adachi et al. (2003) for soybean A3B4, the IE faces of all Arabidopsis cruciferin subunits were slightly positive (basic), whereas the IA faces had a slight net negative charge (acidic). On the IE face, the CRUA and CRUC homotrimers have strongly negatively charged patches located around the center of the three-fold axis (i.e., the central channel), whereas the CRUB homotrimer has three negatively charged patches close to the fringe of the molecule with more positive charges around the central channel. A strong negative charge covers the whole surface of the IA face in the CRUC homotrimer, whereas in the CRUB homotrimer the periphery of the molecule is negatively charged and a strong positive charge is found in the middle of the trimer. The electrostatic potential of the IA face of CRUA homotrimer showed that a slightly positive charge is distributed close to the three-fold axis, whereas the rest of the IA face is negatively charged (Figure 3.5). The presence of charged residues in the IA face may result in a net negative or positive molecular charge exposed to the bulk solvent phase. Consequently, in emulsion formation the droplets may have an electrical charge depending on the pH of the aqueous phase.
Figure 3.7 Ribbon diagram showing the homotrimers constructed with each protomer. (A) HVR and ELR for CRUA, CRUB, and CRUC are in red, green, and blue on IE and IA faces, respectively. Functionally important residues are indicated. (B) Enlarged view of central core region at IE face of Arabidopsis CRUA, CRUB, and CRUC.
The overall charge of the emulsion droplet influences interaction between droplets and other charged molecules in the environment (e.g., flavor compounds, minerals, antioxidants, and other biopolymers). Electrostatic repulsion is also important in preventing droplet flocculation of the protein-stabilized emulsion (McClements, 2005).

As discussed earlier, either as a trimer or a hexamer, the IA face of cruciferin protrudes towards the polar solvent phase (i.e., bulk water phase) in an emulsion or a foam. The highly exposed polar residues on the IA faces of the CRUA and CRUB homotrimers (emphasis on Glu208 and Gln209 in CRUA and Lys201 and Gln202 of CRUB of Figure 3.7A) may be involved in forming hydrogen bonds with H₂O molecules; however, predominant hydrophobic patches on the IA face of these homotrimers may override this effect and cause protein aggregation. This condition may be enhanced by the positive and negative electrostatic interactions (charge–charge interactions) with neighboring protein molecules resulting in coalescence and consequent flocculation, or creaming. In contrast, the CRUC homotrimer has a prominent net negative charge on its IA face. In emulsions, this property may provide an effective barrier to coalescence and therefore stabilize the emulsion (as depicted in Figure 3.8).

The electrical charge on the molecule’s periphery may also play a vital role in emulsion and foam stabilization as neighboring protein molecules on the oil droplet or air bubble surfaces can be held together. If the molecule has more hydrophobic residues and unevenly distributed patches of charged residues, such as in the CRUA and CRUC homotrimers, the molecule may engage in abundant electrostatic interactions, creating a cohesive film on the surface of oil droplet or air bubble. The distribution of both the hydrophobic and electrostatic potential on the surface (i.e., IE, IA, and side) of the CRUC homotrimer suggests that it will not only display good emulsifying and foaming ability, but will also be a good stabilizer.

### 3.4.5 Solubility and Protein Structures

Solubility is the most critical functional property of a protein as it is important for other properties, such as emulsification, foaming, and gelation. Therefore, insoluble proteins have limited uses in food (Kinsella, 1976). Solubility depends primarily on the ionic charge and the hydrophobicity of the molecule (Damodaran, 2008).
**Figure 3.8** A schematic of barrier formation to coalescence and consequent stabilization of the emulsion by CRUC homotrimer. (A) Ribbon diagram and surface properties of the cruciferin CRUC homotrimer representation: areas rich in hydrophobic residues (hydrophilic: magenta and hydrophobic: yellow) on the IE face suggests that it may perform well at an oil-water interface. Electrostatic potential (acidic: red and basic: blue) on the IA face illustrates a prominent negative charge. (B) Schematic representation of hypothetical emulsion system formed by the CRUC homotrimer. The negative charge on the surface that is exposed to water may help to stabilize the protein-coated oil droplets in the emulsion system without coalescing. Inset: Hypervariable regions (HVRs) with considerable numbers of hydrophobic residues in the CRUC homotrimer will further intrude the oil and will be tightly adsorbed.
The CRUC homotrimer was determined to have a negative average hydrophobicity value (Kyte & Doolittle, 1982) (Hø) of –0.58 (Table 3.1), suggesting that it is more soluble in aqueous solution than CRUA (Hø = –0.46) or CRUB (Hø = –0.35), the latter two having more hydrophobic patches on their IA faces than that of CRUC (Figure 3.5). While polar groups contact the surrounding water, negatively charged residues (Asp + Glu) on the CRUC homotrimer IA face may contribute to the repulsion of protein molecules (i.e., not favor protein–protein association), which may lead to improved solubility.

Even though the periphery of the CRUC homotrimer contains some buried positive patches, the interaction of the side of the trimer with the IA face of another molecule is less likely to occur due to strong IA–IA repulsion, especially between the freely moving molecules in the bulk solvent phase. Furthermore, disordered regions located on the IE face or inside the hexamer are primarily hydrophilic (Figure 3.6) due to the presence of Gln residues. Therefore, once the hexamer dissociates into trimers to expose these residues, this molecule will be more soluble. The effect of the distribution of electrostatic surface potential on the solubility of 7S globulins from various crops has been reported (Fukuda et al., 2008). The electrostatic potential at the center of the surface of β homotrimer of β-conglycinin is positive, whereas that of the periphery of one surface was negative. On other surfaces, the distribution of negative and positive charges is fairly uniform. For the α' homotrimer of β-conglycinin, negative and positive potentials were closely distributed at the center of the surface of the molecule. In both cases, charge–charge interactions created large aggregates resulting in protein precipitation (Fukuda et al., 2008). Conversely, molecular surface models of Adzuki 7S1, Adzuki 7S3, and mungbean 8Sα showed mainly negative charges on both faces (i.e., repulsion rather than charge–charge interaction) and were reported to be highly soluble (Fukuda et al., 2008).

3.4.6 Potential for Gel Formation

The formation of protein networks is induced by protein unfolding, for example, by heating. This is essential for use in foods because polymer networks hold water, lipids, sugars, flavor molecules, and other ingredients together (Kinsella, 1976). The type and stability of a thermally induced gel formed by protein can be predicted from the hydrophobicity, charge distribution, disulfide/sulfhydryl content, and cavity sizes of the molecules involved (Damodaran, 2008; Shimada & Matsushita, 1980).
Table 3.1. Average hydrophobicity index ($H_0$), isoelectric pH ($pI$), and molecular mass ($Mr$) of procruciferin and cruciferin.

<table>
<thead>
<tr>
<th></th>
<th>Procruciferin</th>
<th>CRUA</th>
<th>CRUB</th>
<th>CRUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formula</td>
<td>$C_{2247}H_{3515}N_{671}O_{696}S_8$</td>
<td>$C_{2318}H_{3628}N_{686}O_{698}S_{10}$</td>
<td>$C_{2234}H_{3507}N_{643}O_{667}S_{16}$</td>
<td>$C_{2546}H_{4002}N_{760}O_{783}S_{14}$</td>
</tr>
<tr>
<td>$Mr^a$ (kDa)</td>
<td>51.3</td>
<td>52.6</td>
<td>50.6</td>
<td>58.2</td>
</tr>
<tr>
<td>Asx + Glx (%)</td>
<td>27.0</td>
<td>25.3</td>
<td>25.1</td>
<td>31.2</td>
</tr>
<tr>
<td>His + Arg + Lys (%)</td>
<td>11.9</td>
<td>12.8</td>
<td>10.8</td>
<td>9.9</td>
</tr>
<tr>
<td>Acidic/Basic</td>
<td>2.3</td>
<td>2.0</td>
<td>2.5</td>
<td>3.2</td>
</tr>
<tr>
<td>$H_0^b$</td>
<td>–0.56</td>
<td>–0.46</td>
<td>–0.35</td>
<td>–0.58</td>
</tr>
<tr>
<td>$pI^c$</td>
<td>6.60</td>
<td>7.68</td>
<td>6.53</td>
<td>6.53</td>
</tr>
<tr>
<td>ASA$^c$ (Å$^2$)</td>
<td>19,339</td>
<td>20,236</td>
<td>20,513</td>
<td>20,849</td>
</tr>
<tr>
<td>Pocket area$^d$ (Å$^3$)</td>
<td>5551.5</td>
<td>7386.1</td>
<td>7020.6</td>
<td>7441.8</td>
</tr>
<tr>
<td>Openings$^e$ (Å$^3$)</td>
<td>727.1 (17)</td>
<td>825 (12)</td>
<td>1324.4 (10)</td>
<td>1422.3 (10)</td>
</tr>
</tbody>
</table>

$^a$ $pI$ and $Mr$ are calculated using primary sequence of the molecule.
$^b$ $H_0$ is the average hydrophobicity index as defined by Kyte and Doolittle (1982).
$^c$ Solvent-accessible surface area (ASA) was calculated for homotrimers by the rolling-ball method with a radius of 1.4 Å.
$^d$ Pocket area indicates the size of the cavities around central core area of homotrimer.
$^e$ The size of the mouth openings of individual pockets and number of openings are in parenthesis.
The presence of <31.5 mol % hydrophobic residues facilitates formation of a translucent protein gel in water (Shimada & Matsushita, 1980). At the molecular level, CRUA, CRUB, and CRUC contain 32.1, 32.9, and 30.6 mol % of the sum of Val, Pro, Leu, Ile, Phe, and Trp, respectively, suggesting CRUA and CRUB may tend to form coagulum type gels, whereas CRUC would form a translucent gel. Unfolding of the native protein structure is an essential step for good network formation in the gelation process. Heat-induced destabilization of tertiary structure can alter surface properties of proteins due to unfolding or exposure of previously buried residues leading to exposure of previously solvent inaccessible sites. During the gelation process, proteins undergo hydrophobic interaction between denatured or unfolded molecules. Higher levels of hydrophobic residues and attractive electrostatic forces (both positive and negative charges present on the IA and IE faces, Figure 3.5) of CRUA and CRUB may facilitate the formation of a thermocoagulum type gel. The comparatively lower number of hydrophilic groups and higher cumulative negative charges on both faces of the CRUC homotrimer may cause it to form thermotranslucent gels. It is believed that translucent gels are much stronger and less prone to syneresis than coagulum type (opaque) gels (Damodaran, 2008).

The strength of thermoset protein gels is also influenced by the free sulfhydryl and disulfide bond content, because these bonds can break and re-form during thermal denaturation (Shimada & Matsushita, 1980). Similar to most other 11S globulins, the cruciferin protomer contains two disulfide bonds; one bond is within the acidic chain, known as the intra-chain disulfide bond (CRUA, C36↔C69; CRUB, C30↔C63; and CRUC, C37↔C70), and the other is between the acidic and basic chain and is referred to as the inter-chain disulfide bond (CRUA, C112↔C289; CRUB, 106↔276; and CRUC, C113↔C340). Each cruciferin protomer contains two free sulfhydryl (–SH) groups: residues 11 and 293 of CRUA, residues 40 and 280 of CRUB, and residues 21 and 406 of CRUC (Figure 3.2A). The first cysteine residue of CRUA and CRUC is located within the signal peptide (Figure 3.2A) and, therefore, does not contribute to the free sulfhydryl pool in the mature protomer; however, CRUB has two –SH groups in the final quaternary protein structure. The sulfhydryl group at Cys140 of CRUB is located in the loop between the F and G strands of the α-chain (Figure 3.2A), which is exposed to the outside, making it possible to form disulfide bridges with other molecules even without protein unfolding. Accessibility of the free –SH group at Cys406 of CRUC, located on the D stand of the β-barrel at the C-terminal, is very limited. The other free sulfhydryl group of CRUA and CRUB, which is
located close to the inter-chain disulfide bond (i.e., Cys293 and Cys280, respectively), can easily undergo intra-molecular sulfhydryl–disulfide exchange reactions (Adachi, Chunying, & Utsumi, 2004). The inter-chain disulfide bond is in a less buried region compared to the intra-chain disulfide bond, and neither is covered by HVRs. Therefore, the inter- and intra-chain disulfide bonds can participate in network formation during gelation. This possibility was suggested by Nakamura et al. (1984), who found that N-ethylmaleimide, a –SH blocking agent, inhibited gel formation of 5% (w/v) glycinin solutions.

The thermal stability of a protein is also very important and has been related to certain features, in particular, the size of cavities, the number of proline residues (Tandang-Silvas et al., 2010; Fukuda et al., 2008), and the length of loop regions (Chakravarty & Varadarajan, 2002). It has been reported that 11S globulins from different sources with smaller cavity sizes, higher proline content, and shorter loops are more stable upon heating (Adachi et al., 2004; Tandang-Silvas et al., 2010). Although the primary structures of ProA1aB1b of soy and pea prolegumin were similar (63% homology), the combined effect of larger cavity size (4492.8 vs 4086.1 Å³), fewer intra-monomer H-bonds (264 vs 295), longer loops (310 vs 290 residues), and fewer proline residues (24 vs 29) contributed to the lower thermal stability of prolegumin ($T_m = 69.3$ °C) than of proA1aB1b ($T_m = 78.1$ °C) (Tandang-Silvas et al., 2010). Small cavity size also reflects efficient packing. Considering the main pocket around the center of the molecule, the CRUC homotrimer showed the largest cavity among the Arabidopsis cruciferins (7441.8 Å³), whereas the CRUB homotrimer had the smallest value (7020.6 Å³) (Table 3.1 and Figure 3.9). Two separate pockets, the main pocket at the IE face (5008.9 Å³) and a minor pocket at the IA face (2011.7 Å³), were predicted for the CRUB homotrimer that were separated by Trp26 residues (three residues from three protomers) (Figure 3.7). These residues form a barrier to the 1.4 Å probe (similar to the size of a water molecule) used for the calculation of this cavity’s properties. In addition, the size of the mouth openings of individual pockets helps to assess the accessibility of a cavity to the outside bulk solution. According these calculations, CRUC had both the largest pocket size and the largest mouth openings (1422.3 Å³), suggesting easy access to the interior of the molecule by the bulk solvent, and therefore less thermal stability would be expected than with the other homotrimers.
Figure 3.9 Surface topography of Arabidopsis CRUA, CRUB, and CRUC homotrimers computed by CASTp serve (http://cast.engr.uic.edu/cast/). Only the prominent cavities are shown. The conical surface-accessible cavity has wide mouth openings at the IE face, while a small opening is present at the IA face. The openings are considerably different in each trimer and may govern access of solvents and other foreign molecules, such as flavorants, to the inside of cruciferin molecule.
The lesser number of proline residues (5.2%) in CRUC than in CRUA or CRUB (5.5%), as well as the longer HVR-I and HVR-II in CRUC (60 and 16 residues, respectively), may further contribute to the ease of thermal destabilization of this homotrimer. Tandang-Silvas et al. (2010) reported that the greater number of proline residues of procruciferin imparts greater heat stability than that of its closest homologue, the pumpkin pro-11S globulin. Despite the factors that contribute to thermal instability (large cavity size, low proline content), CRUC may exhibit high conformational entropy as temperature increases because the high hydrophobicity of the IE face is an endothermic property (Damodaran, 2008).

3.4.7 Ability to Bind Small Molecules

As a macromolecule, proteins provide very limited organoleptic attributes to foods; however, the ability to bind small flavor molecules, such as aldehydes, ketones, ionones, and esters, by proteins can enhance sensory appeal. Most of interactions between protein and flavor molecules are reversible and involve hydrophobic and hydrogen bonding, but physical entrapment in surface cavities is also possible (Damodaran, 2008; Guichard, 2006). Most aroma compounds are hydrophobic. In β-lactoglobulin, most of the flavor molecules bind to the central hydrophobic pocket and another site on the molecule’s surface (Guichard, 2006). Conformational changes in the protein brought about by environmental conditions (e.g., pH, heat) may change the interaction possibilities for various flavor molecules.

In cruciferin, the central channel (core area) of IA face may play a vital role in binding small molecules. With the emphasis on the core area of the IA face, CRUA and CRUB homotrimers may be better carriers of small flavor molecules than the CRUC homotrimer. The core region of CRUA and CRUB has a hydrophobic cavity, mainly due to the Phe32 in CRUA and Trp26 in CRUB that can form hydrophobic interactions with functional groups of flavor molecules (ligands). After binding to the core region by hydrophobic interactions, flavorants with polar head groups, such as carbonyl and hydroxyl groups, may also form hydrogen bonds with the adjacent Arg68 of CRUA and Glu32 of CRUB (Figure 3.7B). Besides the hydrophobic interactions and H-bonds in CRUB, ligands can form an additional H-bond with Trp26, which is not possible in CRUA (i.e., Phe32). The positively charged (electrostatic potential) core regions of the CRUA and CRUB homotrimers may strengthen the interactions; therefore, the cavities in the core regions appear to be well-suited to cargo flavor molecules. Furthermore, the hydrophilic
nature and strong negative charge distribution in the core region of the CRUC homotrimer may provide a thermodynamically unstable environment for flavor molecules, leading to diminished binding affinity.

3.4.8 Prediction of Proteolytic Sites

NICKPRED software was used to study protease hydrolysis of the modeled cruciferins. This program takes into consideration the spatial arrangement of residues in the 3D structure of a protein (PDB data file), making it an ideal method for predicting proteolytic sites in silico. In the gastric phase of digestion in humans, pepsin is the first proteolytic enzyme to act upon proteins, and hydrolysis occurs only in an acidic environment (pH 1.2). The 11–12S globulins (hexamers) also preferentially dissociate into two trimers of 7S under acidic conditions (Adachi et al., 2003; Plietz et al., 1983; Koshiyama, 1972). Therefore, it is reasonable to believe that the hexameric structure of cruciferin will first undergo acid-induced dissociation into two trimers, followed by pepsin hydrolysis. Proteins that do not undergo proteolysis during the gastric phase and proteins that have already broken down into oligopeptides are subjected to trypsin and chymotrypsin hydrolysis in the small intestine under near-neutral pH (pH 6.8). From the NICKPRED analysis, it is clear that almost all top-scoring pepsin nick-sites in the CRUA and CRUB homotrimers are located on the loops containing HVR-I and HVR-II (Figure 3.10). Although CRUC has the longest HVR-I loop, no pepsin cleavage sites were predicted. Instead, all cleavage sites were predicted in the HVR-II area (Figure 3.10) of CRUC. Because HVR-II is located deep inside the hexamer, near complete dissociation of the hexamer would be necessary to allow for pepsin cleavage of CRUC. Thus, digestion of CRUC may be slower than the other two protomers and can progress only upon complete dissociation of the hexamer. The digestive sites in HVR-I of CRUA and CRUB protrude from the rim of the molecule (Figure 3.10), making these molecules more vulnerable to pepsin attack even when the hexamer is only partially destabilized. Among the crucifer SSPs, cruciferin had the highest potential to be degraded by pepsin under in vitro gastro-intestinal conditions (Wanasundara & Abeysekara, 2009), possibly relating to the favorable structural features noted above.

Trypsin cleaves at Lys or Arg, whereas chymotrypsin cleaves at Trp, Tyr, Phe, or Leu (Figure 3.10). In the CRUA homotrimer, two trypsin cleavage sites (Arg125 and Arg134) and three chymotrypsin cleavage sites (Phe122, Phe126, and Phe135) were predicted for the HVR-I loop
and five trypsin (Arg259, Arg263, Arg266, Arg276, and Arg270) and two chymotrypsin (Phe256 and Leu262) cleavage sites in the HVR-II loop. Outside the HVRs, trypsin was predicted to cleave two sites on the ELR (Arg207 and Lys213) and two sites in the α-chain (Leu228 and Leu231 on helix 2 and helix 3, respectively). Interestingly, a high number of cleavage sites was found in the outlying regions (not HVRs) in the CRUB homotrimer: five trypsin sites (Lys222 at the loop between helix 2 and helix 3 domains in α-chain; Lys200, Lys201, and Lys204 on the ELR; and Arg432 on helix 3 in the β-chain) and four chymotrypsin sites (Leu217 and Phe221 on helix 2 in the α-chain; Trp26 in the core region; and Tyr423 on helix 2 in the β-chain). Only three outlier sites, two trypsin sites (Lys266 in ELR and Lys496 on helix 3 domain in β-chain) and one chymotrypsin site (Phe256 in ELR), were identified in the CRUC homotrimer. Hydrolysis in the HVRs by trypsin and chymotrypsin in the intestinal phase becomes less important if these regions are already cleaved by pepsin during the gastric phase. Therefore, attention should be focused on the sites outside these regions (i.e., outliers) to identify potential digestion sites in the intestinal phase. Of the three types of cruciferins studied, CRUB exhibited a greater number of putative cleavage sites both in the α- and β-chain, with most being located on the IA face. Additionally, a nick site at the core (Trp26) may allow further access to the inside of the CRUB molecule, which would accelerate proteolytic digestion. The presence of putative digestive sites all over the CRUB molecule may also increase its digestibility potential compared to the other cruciferins. The low number of nick sites in the regions other than HVRs in CRUC, especially as hexamer, may cause it to be more resistant to trypsin and chymotrypsin digestion. The lower proteolytic susceptibility of CRUC may increase its antigenic potential, especially if the protein consists only of CRUC subunits, as pepsin resistance is a key feature of the sensitization phase of gastrointestinal allergenicity (Mills et al., 2002).

In general, no cleavage sites were predicted to occur on the surface of the cupin (β barrel) fold, suggesting that protein denaturation caused by changes in ionic strength or heating may be required to expose hidden nick sites to accelerate proteolytic degradation of cruciferin.
Figure 3.10 Prediction of phosphorylation and proteolytic sites of Arabidopsis CRUA, CRUB, and CRUC homotrimers. Potential phosphorylation sites are indicated as spheres (Ser: violet, Thr: orange, Tyr: green) as predicted by NetPhos 2.0. The top scoring proteolytic sites (pepsin: red, trypsin: blue, chymotrypsin: yellow) are shown in stick representation. The 3-fold symmetry axis for the trimer view on the left hand side is shown by an arrow.
3.4.9 Factors Influencing Functional Properties of Cruciferins

Phosphorylation of a protein can influence its physiochemical behavior and functional properties in a food system. Surface functional properties and gel-forming ability in the presence of calcium ions has been attributed to phosphorylation of casein, which is useful in cheesemaking. Improved surface hydrophobicity, cross-linking ability, and interfacial properties have been observed upon chemical phosphorylation of *B. napus* 12S protein (Schwenke et al., 2000; Krause, 2002). I used the NetPhos 2.0 server to identify phosphorylation sites originating from serine, threonine, and tyrosine residues (pSer, pThr, and pTyr, respectively) of cruciferin. Residues with a probability score >0.5 were predicted to have potential for phosphorylation. A total of 20 phosphorylated residues were predicted for CRUA (12 pSer, 6 pThr, and 2 pTyr), 14 for CRUB (10 pSer, 1 pThr, and 3 pTyr), and 17 for CRUC (10 pSer, 3 pThr, and 4 pTyr) (Figure 3.10).

In CRUA, six residues (pSer119, pSer120, pSer330, pSer335, pTyr312, and pThr287) were located on the IE face, three (pSer398, pSer439, pSer458, and pThr397) resided on the periphery of the trimer, and none of the phosphorylated residues were exposed to the IA face. These predictions were in accordance with the report of Wan et al. (2007), who used LC-MS/MS analysis to determine that Arabidopsis cruciferin subunits were phosphorylated. Prediction of phosphorylation residues on CRUB indicated that four phosphorylated residues (pSer113, pSer288, pTyr293, and pTyr299) project from the IE face, whereas three phosphoserine residues (pSer426, pSer436, and pSer446) reside at the fringe of the molecule. Similar to CRUA, none of the residues on the IA face were predicted to be phosphorylated. In the CRUC homotrimer, only two phosphorylated residues (pSer321 and pTyr357) were found on the IE face and only one (pSer490) was found at the periphery. Interestingly, four phosphorylated amino acids (pSer53, pSer297, pThr50, and pTyr60) were identified on the IA face of CRUC. *In vitro* phosphorylation of soy SSP, either chemically or enzymatically, was found to improve certain functional properties under slightly acidic conditions, including solubility, gelling, and emulsifying ability (Hirotaka, Taniguchi, Narita, & Kito, 1984; Ross, 1989). Most of the potential phosphorylation sites in CRUA or CRUB are buried at the IE interface in the hexamer, suggesting that ring opening of the hexamer to form two trimers can expose these hidden sites to the surrounding solvent. Along this line, phosphorylation of the IA face of CRUC may enhance its techno-functional properties to a significant extent even without hexamer opening.
So far, the molecular features that affect physico-chemical properties and their related techno-functional properties of cruciferins have been discussed. It is well-known that the properties of proteins are affected and altered by processing treatments and environmental factors applied during food processing and storage. The most important extrinsic factors in foods that cause changes are pH, ionic strength, temperature, pressure, and chemicals. Theoretically, proteins bear a net positive charge at acidic pH and a net negative charge at basic pH with no net charge at the isoelectric pH (pI) of the molecule. The calculated isoelectric points of cruciferins indicated that CRUA has a slightly basic pI of 7.68 (acidic/basic ratio = 0.74), whereas CRUB and CRUC have a slightly acidic pI value of 6.53 (acidic/basic ratio = 0.86), suggesting that the pH of the solvent may have different effects on functional properties, such as solubility, emulsification, and foaming, on CRUA than on CRUB or CRUC (Table 3.1). The surface distribution of acidic and basic amino acids, calculated as electrostatic potential, may be more meaningful than the calculated pI values as far as functional properties are concerned. The electrostatic potential evaluation of CRUA and CRUB shows that they have a similar number of positive and negative charges on their IA faces, whereas the potential of the CRUC IA face is predominantly negative (Figure 3.5). At a pH below the isoelectric point, CRUA and CRUB may have net positive charge on the IA face and have minimum charge–charge interaction resulting in an improvement of properties, such as solubility and emulsification. On the other hand, CRUC has fewer positive patches on the IA face than the other two isomers. At a low pH the overall negative charge will be diminished and become close to neutral, this should allow charge–charge interaction and aggregation of protein molecules is possible. At high pH, the functional properties for all cruciferins will be the opposite of that at low pH. Analysis of charge distribution on the molecular surface may be used to predict how environmental factors should be controlled to optimize functional properties, which is not possible by examining the primary amino acid sequences only. The effect of salts (e.g., NaCl and KCl) on the techno-functional properties of proteins depends on both the type and concentration of salt ions, which can pose electrostatic influences (Damodaran, 2008). Solubility and emulsifying properties of CRUC may decrease to a lesser extent than CRUB or CRUA at high salt concentrations as the neutralization of surface charges on CRUC requires more ions than that of CRUA and CRUB. It is expected that the information obtained from this study will be able to provide directions for genetic improvement of crucifer oilseed proteins for targeted functions in the final food or nonfood product.
3.5 Conclusions

Homology modeling was used to predict the structure of three different Arabidopsis cruciferin isomers (CRUA, CRUB, and CRUC). Cruciferin isomers composed of similar protomers were used in generating structure models to mitigate the complexity of examining protomer combinations. It is clear that the CRUC homotrimer has different structural features resulting from the HVRs and should have different physico-chemical and functional properties from those of the CRUA or CRUB homotrimers. Elucidating the 3D structure of proteins using HM allows exploration of the molecular structural features in detail. Furthermore, one can predict in silico the effect of changing structural features with the aim of enhancing functions and properties that are physico-chemically and physiologically critical. The HM and in silico analysis of molecular properties is possible provided that primary sequence data and sufficient details of the protein are available. Most food proteins originate from a limited number of plants and animals for which a great degree of evolutionary and genomic information is now available. Using information from Arabidopsis, I showed that theoretically, it is possible to identify differences in the tertiary structure features of the various cruciferin protomers with the potential to provide different physico-chemical and functional properties. In addition, this approach can be used to guide empirical laboratory studies on potential modification of protein molecules to alter structural features and enhance functional properties.

3.6 Connection to Subsequent Chapters

Chapter 3 showed that bioinformatics, HM, and in silico investigation of protein structures, especially tertiary and quaternary levels, can provide insight to the expected physico-chemical properties of the molecules and also to predict potential functionalities. In next four (4, 5, 6, and 7) chapters, genetically defined Arabidopsis plants that produce seeds accumulating single types of cruciferin protomers are used to firmly establish these structure–function relationships. In addition to providing information on homohexameric cruciferin that is not currently available, the data helps to validate the in silico predictions made on Chapter 3 using molecules of similar composition.
4. CHARACTERIZATION OF *ARABIDOPSIS THALIANA* LINES WITH ALTERED SEED STORAGE PROTEIN PROFILES USING SYNCHROTRON POWERED FT-IR SPECTROMICROSCOPY*

4.1 Abstract

*Arabidopsis thaliana* lines that expressed only one cruciferin subunit type (double-knockout; *CRUAbc*, *CRUaBc*, or *CRUabC*) or were devoid of cruciferin (triple T-DNA knockout; *CRU−*) or napin (napin-RNAi) were generated. Seeds of double knockout *CRU* lines accumulated homohexameric cruciferin and contained similar protein levels as the wild type (WT). Chemical imaging of WT and double-knockout seeds using synchrotron FT-IR spectromicroscopy (amide I band, 1650 cm\(^{-1}\), \(\nu_{C=O}\)) showed that proteins were concentrated in the cell center and protein storage vacuoles. Protein secondary structure features of the homohexameric cruciferin lines showed predominant \(\beta\)-sheet content. The napin-RNAi line had lower \(\alpha\)-helix content than WT. Lines entirely devoid of cruciferin had high \(\alpha\)-helix and low \(\beta\)-sheet levels indicating that structurally different proteins compensate for the loss of cruciferin. Lines producing homohexameric CRUC showed minimal changes in protein secondary structure after pepsin treatment indicating low enzyme accessibility. Synchrotron FT-IR provides information on protein secondary structure and changes to the structure within the cell.

4.2 Introduction

Crucifer SSP are predominantly of the 11S globulin and 2S albumin types, with lesser contributions from oleosins, defensins, and late embryogenesis abundant proteins (Wanasundara, 2011). Cruciferin, the most abundant protein, is an 11S hexameric globulin of 300–360 kDa and a member of the cupin superfamily. Cruciferin protomers (subunits) comprise 27 β-strands and 7 helices that are folded into two β-barrel domains and two extended helix domains, each containing two helices (Tandang-Silvas et al., 2010). The second most abundant protein is napin, a 12–16 kDa 2S albumin of the prolamin superfamily. The secondary structure of napin is predominantly α-helix, though some β-sheet structure is also present (Schmidt et al., 2004).

Multiple genes encode cruciferins and napins. For example, in the model crucifer Arabidopsis thaliana cv. Columbia, three paralogous genes AT5G44120.3, AT1G03880.1, and AT4G28520.1 encode CRUA, CRUB, and CRUC cruciferin protomers, respectively (Wan et al., 2007). Each cruciferin protomer is composed of a heavy acidic (α) and a light basic (β) chain linked by a disulfide bond. In WT lines, the resultant hexameric cruciferin is an assortment of these three protomers. Similarly, in Arabidopsis, napin is encoded by multiple (five) genes with the mature protein consisting of two disulfide-linked subunits (Schmidt et al., 2004).

Cruciferin secondary, tertiary, and quaternary structure affects biochemical and techno-functional properties, such as surface active, thermal, and rheological properties. In addition, enzyme accessibility, which is an important consideration in determining the availability of amino acids and peptides upon digestion, as well as safety (i.e., potential immunogenicity) of the protein is related to the protein structure (Yu, 2005c; Yu, Doiron, & Liu, 2008; Todd, Orengo, & Thornton, 2007). The native WT cruciferin has a heterogeneous subunit composition; therefore the ability to extrapolate molecular structure information to functions is limited. Information regarding the contribution of individual cruciferin protomers to the final physico-chemical properties of the hexamer is needed to rationalize protein quality improvement through targeted genetic selection. Whereas single SSP protomers expressed in microbial systems have been examined, these lack post-translational modifications that occur during the normal transport and deposition into protein storage vacuole (PSV) (Tandang et al., 2004; Tandang et al., 2005; Tandang-Silvas et al., 2010). In Chapter 3, the possibility of assembling CRUA, CRUB, and CRUC protomers as homomeric proteins and observing structural and functional differences was presented. In the present study, A. thaliana lines expressing only one type of cruciferin (double-
knockout lines) generated to eliminate the heterogeneity associated with multiple protomer types and lines devoid of either napin (napin-RNAi) or cruciferin (triple-knockout) were studied for SSP structure details at the secondary structure level.

Vibrational spectroscopy in the Raman and mid infrared (IR) regions is very useful in obtaining information on protein secondary structure arrangement using the unsaturated amide bond in the polypeptide backbone. In general, IR spectroscopy applies to molecules with symmetric vibrations of nonpolar groups and Raman spectroscopy applies to molecules with asymmetric vibration of nonpolar groups. The fluorescence from intrinsic amino-acids and the coextracted phenolic compounds interferes with Raman spectroscopy of plant proteins (Ma et al., 2000), but can be eliminated by Fourier transform near-infrared (FT-NIR) Raman spectroscopy (Schrader et al., 1991). Infrared spectroscopy in the mid-IR region is non-destructive and provides label-free fingerprint-like spectra originating from the characteristic vibrational frequencies of the chemical bonds of the molecules of the cell; therefore, it is useful in studying cellular macromolecules.

Synchrotron-powered FT-IR spectromicroscopy was used to examine the secondary structure features of SSP in Arabidopsis lines expressing different cruciferin subunits. As the proteins were in their natural forms and setting within the PSV, modifications or changes that might occur during separation and isolation of the proteins were also eliminated. Furthermore, changes in protein secondary structure were assessed as an indicator of digestive enzyme accessibility directly within seed tissues.

4.3 Materials and Methods

4.3.1 Generation of Arabidopsis Lines with Altered SSP Profiles

To obtain cruciferin composed of identical subunits (i.e., homomeric cruciferin), two of three cruciferin genes in Arabidopsis cv. Columbia were inactivated by genetically combining T-DNA insertions to develop double-knockout lines (Figure 4.1A).

The following T-DNA insertion lines obtained from either the Arabidopsis Biological Resources Centre (www.abrc.osu.edu) or GABI-Kat (www.gabi-kat.de) were used: CRUaBC (cruACRUBCRUC) (SALK 002668, T-DNA inserted in second exon of AT5G44120), CRUAbC (CRUAcrubCRUC) (SALK 045987, T-DNA inserted in fourth exon of AT1G03880) and CRUABc (CRUACRUBcruC) (GK-283D09, T-DNA inserted in first exon of AT4G28520)
(Figure 4.1B). The homozygous double-knockout lines referred to as \textit{CRUAbc} (\textit{CRUAcrubcruC}), \textit{CRUaBc} (\textit{CruACRUBcruC}), and \textit{CRUabC} (\textit{cruAcruBCRUC}) were obtained by conventional crossing. The protein from these double-knockout lines is referred to as CRUA, CRUB, and CRUC homohexamers. A homozygous cruciferin null mutant (i.e., triple-knockout, \textit{cruAcruBcruC or CRUabc or CRU}−) was also generated and the protein of this line is referred to as CRU-. Table 4.1 summarizes the relationship of genes and proteins, including the nomenclature used throughout this study. To generate a line devoid of napin, an RNAi cassette containing nucleotides 328–513 of \textit{AT4G27150} was cloned in the sense and antisense orientations between the 3X35S promoter and the OS terminator of pGSA1252, which separated them with a GUS gene fragment to create the RNAi hairpin structure. This was transformed into \textit{A. thaliana} cv. Columbia, with the BAR gene conferring Bialaphos resistance in transformed plants. This napin-depleted line is referred to as napin-RNAi.

4.3.2 Seed Production

Seeds from the WT, cruciferin-knockout, and napin-RNAi lines, as well as \textit{Brassica napus} (doubled haploid line DH12075) were used. All plants were grown in plastic pots containing a COCO/Sunshine soil mixture (1:1) (Michigan Peat, Houston, TX, USA) consisting of coconut fiber/peat moss/vermiculite (1:3:3, w/w/w) with Osmocote PLUS 15–9–12 controlled release fertilizer (Scotts Co. LLC, Marysville, OH, USA), in a greenhouse under the following conditions: 16 h light (~800 Wm$^{-2}$) at 21 °C and 8 h dark at 16 °C. Plants were bagged at the flowering stage, seeds were harvested at maturity and stored dry at room temperature.

4.3.3 Contents of Lipid and Protein

Protein (based on total N) and lipid contents of the seeds were determined by the combustion method using a Flash EA$^\text{®}$ 1112 N-analyzer (Thermo Fisher Scientific) with a conversion factor of 5.64 g protein/g nitrogen (Lonien & Schwender, 2009) and the AOCS method AM 2-93, respectively. All analyses were performed in triplicate.
Figure 4.1 Arabidopsis lines with altered seed storage protein profiles. (A) An illustration of the expected cruciferin subunit composition of double-knockout (KO) lines. (B) Schematic diagram of AT5G44120.3, AT1G03880.1, and AT4G28520.1 with T-DNA insertion and primer location for T-DNA diagnostic PCR. Black boxes represent exons, white boxes represent untranslated regions (UTR), lines between black boxes represent introns. CRA1-F, CRB-F, and 283D09F2 are forward primers. CRA1-R, CRB-R, and 283D09R2 are reverse primers. LBa1 is a primer specific for the T-DNA left border, and bp is base pairs. Scale bar = 300 bp. (C) T-DNA insertion in the cruciferin gene observed by PCR using gene-specific primers and a T-DNA tag-specific primer suggested by SALK and GABI-Kat. SDS-PAGE of seed extracts and western immunoblotting with a primary polyclonal antiserum raised against Arabidopsis cruciferin α-chain.
Table 4.1 Relationship of cruciferin gene and proteins and their nomenclature in this study

<table>
<thead>
<tr>
<th>Line</th>
<th>Gene</th>
<th>Protein(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild type (WT)</td>
<td>CRUACRUBCRUC or CRUABC</td>
<td>CRUABC</td>
</tr>
<tr>
<td>Double-knockouts</td>
<td>CRUACruBcruC or CRUAbc</td>
<td>CRUA</td>
</tr>
<tr>
<td></td>
<td>CruACRUBCruC or CRUaBc</td>
<td>CRUB</td>
</tr>
<tr>
<td></td>
<td>CruACruBcruC or CRUabC</td>
<td>CRUC</td>
</tr>
<tr>
<td>Triple-knockout</td>
<td>CruACruBcruC or CRUabc or CRU−</td>
<td>CRU−</td>
</tr>
</tbody>
</table>

\(^a\)These acronyms are used as adjective for protomer, trimer, or hexamer of WT and double-knockout proteins (e.g., CRUA protomer, CRUA homotrimer, or CRUA homohexamer).

4.3.4 DNA Extraction and Polymerase Chain Reaction (PCR)

Leaves of 4-week-old plants were used for DNA extraction. Briefly, two leaves were ground in liquid N\(_2\) with 400 \(\mu\)L of DNA extraction buffer (250 mM NaCl, 25 mM EDTA, 0.5% w/v SDS, 200 mM Tris-HCl, pH 8.5) and then centrifuged at 13000g for 2 min. A 300 \(\mu\)L aliquot of the supernatant was transferred to a new microcentrifuge tube, 300 \(\mu\)L of cold isopropanol added, and the tube gently inverted three times and then incubated for 2 min before centrifugation at 13000g for 5 min. The pellet was washed with 70% (v/v) ethanol and resuspended in 50 \(\mu\)L H\(_2\)O. The PCR consisted of 20 ng \(\mu\)L\(^{-1}\) of DNA, 10× PCR buffer, 25 mM MgCl\(_2\), 10 mM dNTPs, 5 \(\mu\)M \(CRUA\), \(CRUB\), or \(CRUC\) forward and reverse primers (Table 4.2), 5 \(\mu\)M SALK LBa1 or GABI LB08409 primer, and 2 units of Taq DNA polymerase. PCR conditions were as follows: 95 °C for 5 min followed by 29 cycles of 93 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min and a final extension at 72 °C for 10 min.

4.3.5 One-Dimensional Gel Electrophoresis (1D PAGE)

Fifteen milligrams of dried seeds were ground in liquid N\(_2\) with 500 \(\mu\)L of extraction buffer consisting of thiourea/urea lysis buffer [7 M urea, 2 M thiourea, 18 mM Tris-HCl, 4% (w/v) CHAPS, 0.2% (v/v) Triton X-100, pH 8.8], 0.34% (w/v) protease inhibitor cocktail (Sigma Steinhein, Germany), 150 mM DNase I, and 10 mM RNase A and incubated at 4 °C for 10 min. Twelve microliters of 100 mM dithiothreitol (DTT) was added and the mixture incubated for another 20 min at 4 °C and then centrifuged at 20000g at 4 °C for 20 min. Finally, the supernatant was collected and frozen at –80 °C if PAGE was not carried out on the same day.
<table>
<thead>
<tr>
<th>Target gene</th>
<th>Locus designation</th>
<th>T-DNA insertion line</th>
<th>Locus primer sequence</th>
<th>T-DNA primer</th>
</tr>
</thead>
</table>
| **CRUA**    | *AT5G44120*       | SALK 002668          | F: 5'-GCTCCGTTGCTACGGTCTC-3'  
R: 5'-GACGGATAGATCCACGGAGGG-3' | CRA1-LB/RB-AGP-002668  
LBa1 |
| **CRUB**    | *AT1G03880*       | SALK 045987          | F: 5'-AACGCGGGCACTCTACGAC-3'  
R: 5'-AACAATCAAATGTGTGCCTTGAAG-3' | CRB-RB/LB-AGP-045987  
LBa1 |
| **CRUC**    | *AT4G28520*       | GABI 283D09          | F: 5'-GATAAATTTAGTTGATGGATAAC-3'  
R: 5'-CCACCTTCTGGTGACATGT-3' | 283D09F2/R2 GABI-LB08409 |
The supernatant was mixed with 10% (v/v) glycerol and Bromophenol Blue (0.05%, w/v) and electrophoresis carried out in a polyacrylamide gel (12.5% separating and 5% stacking) at 125 V for 2 h. The protein bands were stained with Bio-Safe Coomassie Blue (Bio-Rad Laboratories) or transferred to a nitrocellulose membrane for western blotting.

4.3.6 Western Blotting

After 1D PAGE, proteins were transferred onto nitrocellulose membrane with 1× Towbin’s transfer buffer [25 mM Tris-HCl, pH 7.6, 192 mM glycine and 20% (v/v) methanol] at 100 V for 1 h. The membranes were incubated with a solution of 5% (w/v) nonfat skim milk powder (Nestle Carnation, Solon, OH, USA) in 1× TBS buffer (20 mM Tris-HCl and 14 mM NaCl, pH 7.5) overnight with gentle shaking at 4 °C. For cruciferin detection, the membranes were probed with a polyclonal antiserum against recombinant Arabidopsis cruciferin with affinity for the α-subunit in 1× TBST buffer (0.3% Tween 20 in TBS) at a dilution 1:20000 for 1.5 h. The membrane was washed three times for 10 min with 1× TBST buffer and incubated with a secondary antibody (goat anti-rabbit IgG conjugated to horseradish peroxidase; Bio-Rad Laboratories) at a dilution of 1:20000 in 1× TBST buffer containing 5% nonfat dried skim milk powder for 1 h. The membrane was washed three times using 1× TBST buffer and horseradish peroxidase detected using the ECL Plus Western Blotting system (GE Healthcare).

4.3.7 Sample Preparation and Synchrotron FT-IR Spectromicroscopy

Arabidopsis seeds (with or without any treatment) were embedded in optimal cutting temperature (OCT) compound (Sigma Chemical, St. Louis, MO, USA), frozen (−10 °C) and sectioned (~6 µm thickness) using a microtome (Tissue Tech, NJ, USA). Unstained sections were mounted onto BaF₂ windows (13 mm × 1 mm disk, Spectral Systems, NY, USA) for transmission mode synchrotron FT-IR imaging. Fifteen to twenty seeds were sectioned and mounted on a single BaF₂ disk. Mid-IR spectra of the seed cross sections were collected at the Canadian Light Source Inc. (CLSI, University of Saskatchewan, Saskatoon, Canada) on beamline 01B1-1. Microscopic images of seed cross sections were obtained using a charged-couple device (CCD) camera in a Hyperion 3000 confocal microscope using 20× and 36× focusing objectives (Bruker Optik GmbH, Ettlingen, Germany). Spectromicroscopic mapping of seed tissues was performed using a Bruker Optics Vector 70v/S interferometer coupled with a Hyperion 3000 IR confocal
microscope equipped with a 36× objective and a 64 × 64 pixel Focal Plane Array (FPA) detector. The mid-IR probe pulses were generated as synchrotron radiation from the bending magnet with an energy range of 0.070–0.744 eV. Raster scans were collected in mid-IR range of 4000–800 cm⁻¹ with 10 × 10 μm aperture at a resolution of 4 cm⁻¹. The step size of the scanning was 10 μm (1 μm in single cell scanning) in both the x and y orientations for the maps, and 128 interferograms were co-added to a single spectrum. Sections from 10 seeds were analysed.

The amide I band at 1650 cm⁻¹ of the synchrotron FT-IR spectra was integrated (baseline 1720–1588 cm⁻¹) to construct protein chemical images of seed tissues using Opus v6.5 software (Bruker Optik GmbH). For lipid distribution, the carbonyl ester peak at 1747 cm⁻¹ (integral baseline 1786–1720 cm⁻¹) was used. Fourier self deconvolution of the amide I peak was carried out using the same software. Band assignments and quantification of secondary structure elements, α-helix (1659.7 ± 0.7 cm⁻¹), β-sheet (1616.8 ± 0.8, 1627.8 ± 1.3, 1638.4 ± 1.7, and 1694.1 ± 0.2 cm⁻¹), β-turn (1671.0 ± 0.3 and 1682.3 ± 0.3 cm⁻¹), and random structure (1649.1 ± 1.5 cm⁻¹) were made as described previously (Byler & Susi, 1986; Dong, Huang, & Caughey, 1990).

4.3.8 Enzyme Accessibility of Tissue Proteins

Seeds were treated with porcine pepsin (2410 units/mg solid, Sigma Chemical Co.) to assess enzyme accessibility as follows. Seeds were soaked in water for 1–2 h, washed three times to remove most of the seed mucilage and then frozen at −70 °C. Frozen seeds were slightly cracked using a motar and pestle to expose the inner tissues. Simulated gastric fluid (SGF) was prepared by dissolving pepsin in 35 mM NaCl in 0.084 N HCl (pH 2.0) to maintain an enzyme to seed protein ratio of 1:3 (w/w). The cracked seeds were incubated with SGF at 37 °C for 2 h, at which time pepsin digestion was terminated by adjusting the pH to 7.4 with 1 M NaOH. The treated seeds were washed three times with ddH₂O and freeze dried before being used for synchrotron FT-IR analysis as described above.

4.3.9 Statistical Analysis

To quantify secondary structure components and compare integral peak area of the amide I band, 10 spectral data sets were subjected to one-way analysis of variance (ANOVA) using the General Linear Model (GLM) procedure of SAS Version 9.1 (Statistical Analysis Software; SAS
Institute Inc., Cary, NC, USA). If the main effect was significant (\(P < 0.05\)), and mean separation was done by calculating Fisher’s protected least significance difference (LSD). Principal Component Analysis (PCA) on secondary structure data \((n = 10)\) obtained from Fourier self-deconvolution (FSD) was performed using SAS procedure PROC PRINCOMP. A detailed description of the application of PCA method in the FSD amide I components is found in the literature (Yu, 2005c; Yu et al., 2008). The first three principal components (PC1, PC2, and PC3) were extracted based on the criterion that they accounted for at least 10% of the total variance in the data set (SAS manual).

4.4 Results and Discussion

4.4.1 Confirmation of Lines with Altered SSP Profiles

The SSP (cruciferin and napin) from *A. thaliana* are closely related to those from more economically important oilseed species, such as *B. napus* (Inquello, Raymond, & Azanza, 1993; Pang, Pruitt, & Meyerowitz, 1988; Rödin, Ericson, Josefsson, & Rask, 1990; Sjödahl, Rödin, & Rask, 1991). The *A. thaliana* cv. Columbia genome contains four genes that encode cruciferin protomers: *AT5G44120.3* (gene CRUA: syn CRU4 or CRA1), *AT1G03880.1* (gene CRUB: syn CRU3 or CRB), *AT1G03890.1* (gene CRU2), and *AT4G28520.1* (gene CRUC: syn CRU1, CRC, or CRU3) (http://www.arabidopsis.org/ and http://www.uniprot.org, accessed in July 2012). Most transcripts are derived from *AT4G28520, AT5G44120* and less so from *AT1G03880* (White et al., 2000). *AT1G03890* is linked in tandem to *AT1G03880* that encodes CRUB, but the former is very poorly transcribed (http://signal.salk.edu/) and contributes little to the SSP profile. In accordance, MS analysis revealed that the relative contribution of cruciferin isoforms to total seed protein are CRUC > CRUA > CRUB with the cruciferin produced by *AT1G03890* present at very low levels (Wan et al., 2007). In *B. napus*, five cruciferin protomers have been reported: CRUA (gene CRUA syn CRU2/3), CRU4 (gene CRU4), CRU3 (gene CRU1), CRU1 (gene BnC1), and CRU2 (gene BnC2) (http://www.uniprot.org, accessed in July 2012). The Arabidopsis CRUA, CRUB, and CRUC subunits show high degrees of identity (50–84%) and similarity (67–91%) with *B. napus* subunits CRUA (CRU2/3), CRU4, and CRU1 (Table 4.3), indicating that it is reasonable to extrapolate information pertaining to the properties of Arabidopsis cruciferin protomers to those from *B. napus*. 

71
Table 4.3 Sequence comparison of selected cruciferins of *A. thaliana* with *B. napus*

<table>
<thead>
<tr>
<th>Protein</th>
<th>Identity %</th>
<th>Similarity %</th>
<th>Gaps %</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>A thaliana</em> with <em>B. napus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRUA with CRUA (Bn)</td>
<td>78.0</td>
<td>84.0</td>
<td>9.9</td>
</tr>
<tr>
<td>CRUB with CRUA (Bn)</td>
<td>68.9</td>
<td>80.4</td>
<td>8.2</td>
</tr>
<tr>
<td>CRUC with CRUA (Bn)</td>
<td>57.0</td>
<td>71.5</td>
<td>8.7</td>
</tr>
<tr>
<td>CRUA with CRUB (Bn)</td>
<td>71.1</td>
<td>81.4</td>
<td>5.6</td>
</tr>
<tr>
<td>CRUB with CRUB (Bn)</td>
<td>84.3</td>
<td>91.1</td>
<td>2.0</td>
</tr>
<tr>
<td>CRUC with CRUB (Bn)</td>
<td>53.1</td>
<td>68.9</td>
<td>11.4</td>
</tr>
<tr>
<td>CRUA with CRUC (Bn)</td>
<td>53.8</td>
<td>69.4</td>
<td>10.7</td>
</tr>
<tr>
<td>CRUB with CRUC (Bn)</td>
<td>49.9</td>
<td>66.9</td>
<td>13.8</td>
</tr>
<tr>
<td>CRUC with CRUC (Bn)</td>
<td>83.7</td>
<td>89.6</td>
<td>4.4</td>
</tr>
<tr>
<td><em>A thaliana</em> with <em>A. thaliana</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRUA with CRUB</td>
<td>72.5</td>
<td>83.7</td>
<td>4.7</td>
</tr>
<tr>
<td>CRUA with CRUC</td>
<td>54.3</td>
<td>69.4</td>
<td>11.8</td>
</tr>
<tr>
<td>CRUB with CRUC</td>
<td>49.5</td>
<td>66.9</td>
<td>15.1</td>
</tr>
</tbody>
</table>

*a* The possibility of same amino acid residue to be in the same position when two given sequences are aligned.

*b* The possibility of same amino acid residue and/or positive substitution in the same position when two given sequences are aligned.

*c* The degree of spaces introduced in the sequence during the alignment process.

The primary sequences of CRUA, CRUB, and CRUC protomers of Arabidopsis show distinct differences in their hypervariable regions (HVR), but have a highly conserved core structure (Chapter 3). For example, the HVR-1 region of the CRUC protomer contains 61 residues with at least 8 Gln and Gly repeats, whereas those of the CRUA and CRUB protomers contain only 9 and 7 residues with 1 and 2 repeats, respectively. Comparative *in silico* modeling indicated that identical cruciferin subunits could form stable hexameric structures (Chapter 3).

Genetic combination of *CRU* T-DNA insertion mutants yielded double-knockout lines containing a single functional *CRU* gene and a triple-knockout line in which all of the *CRU* genes were inactivated (Figure 4.1B,C). SDS-PAGE analysis showed bands corresponding to the α- and β-chains of the CRUA, CRUB, and CRUC protomers in the seed extracts (Figure 4.1C). The WT profile had three protein bands between 27 and 34 kDa corresponding to the α-chains; the 34 kDa band was more intense and possibly two proteins of similar molecular weight, and two bands at approximately 20 kDa, corresponding to the region where the β-chains migrate. The double-knockout line expressing only *CRUC* produced a predominant 34 kDa and a minor 32 kDa α-chain. The line expressing only *CRUA* gene produced two α-chains of equal abundance, one at
approximately 33.5 kDa and the other at 32 kDa. A single \( \beta \)-chain of 27 kDa was predominant for \( CRUAbc \) and \( CRUabC \). The 33.5 kDa band of CRUA protomer and the 34 kDa band of CRUC protomer likely combine to form the band of greater intensity seen in the WT profile. The line expressing only \( CRUB \) gene produced a single \( \beta \)-chain of approximately 27 kDa and a single \( \alpha \)-chain of 31 kDa. The multiple \( \beta \)-chains seen in the \( CRUAbc \) and \( CRUabC \) genotypes likely derive from alternate proteolytic processing of the protein or post-transcriptional modifications, because previous MS analysis did not detect proteins derived from rare mRNA splice variants that appear in transcript analysis data (Wan et al., 2007). The triple-knockout line had very weak and diffused protein bands of approximately 30 kDa; however, western blot analysis with an antibody having affinity for the cruciferin \( \alpha \)-chain did not detect cruciferin (Figure 4.1C). All the lines, including the triple knockout line, showed a protein band at 21 kDa, which is different from cruciferin \( \beta \)-chain and is most likely oleosin. Napin expression was not affected in the cruciferin altered lines, and the protein separated into large (L) and small (S) chains upon S–S bond reduction (Figure 4.1C).

Five genes encode napin in \( A. \) thaliana cv. Columbia; a single gene at \( AT5G54740 \) and four genes linked in tandem at \( AT4G27140, AT4G27150, AT4G27160, \) and \( AT4G27170 \). It is impractical to combine T-DNA knockouts in the four linked genes through crossing; therefore, RNAi was used to reduce napin gene expression. SDS-PAGE analysis of a line with an especially strong phenotype revealed that little or no napin was produced, whereas the profile of other seed proteins was unaffected (Figure 4.1C). This phenotype in the line has remained stable through at least six generations. Furthermore, in either the cruciferin and napin-deficient lines no detectable differences in seed size, seed germination, or seedling vigor were found compared to the WT (data not shown).

4.4.2 IR Evidence for Biopolymer Localization in Parenchyma Cells

High resolution synchrotron FT-IR techniques have been used for subcellular chemical mapping of single cells, such as hybridoma B cells in mice (Jamin et al., 1998), metal-cyanobacteria (Yee, Benning, Phoenix, & Ferris, 2004), and oral and lung epithelial cells (Diem et al., 2004; Holman, Martin, Blakely, Bjornstad, & McKinney, 2000). Scanning of seed tissue samples (~6 \( \mu \)m thickness) in the mid-IR spectral range from 4000 (\( \lambda = 2.5 \) \( \mu \)m) to 800 cm\(^{-1} \) (\( \lambda = 12.5 \) \( \mu \)m) under transmittance mode allowed information on cellular microstructure to be obtained
that was converted into details of the structural properties of the macromolecules within. This was the first reported use of synchrotron FT-IR analysis of a single cell within an Arabidopsis seed. The confocal microscope coupled to the synchrotron IR light source (flux, $1 \times 10^{14}$ at 10 µm, v/s/0.1%BW at 100 mA) at the CLS is equipped with Schwarzschild objectives with numerical apertures (NA) of 0.4 for 15× and 0.52 for 36× magnifications. The achievable spot size or diffraction-limited spatial resolution for this type of system is 2–10 µm (0.5–1.2λ) (Carr, 2000; Carr, Reffner, & Williams, 1995; Holman, Bechtel, Hao, & Martin, 2010). The single-cell analysis was carried out using an aperture size of 5 × 5 µm with a step size of 1–2 µm in the transmittance mode and scanned with little radiation diffraction, therefore, loss of S/N in the mid-IR range was minimal. Furthermore, the average size of a parenchyma cell in an Arabidopsis embryo is approximately 15 to 20 µm in diameter (Figure 4.2), whereas the average diameter of PSV is 6.2 to 8.2 µm (Shimada et al., 2003), which are much larger than the diffraction-limited spot size used in the mid-IR region.

Protein, oil, cellulose, and lignin distribution was determined in embryonic parenchyma cells of WT Arabidopsis (Figure 4.2). The IR vibrations of N–H in-plane bending ($\delta$N–H) and C–N stretching ($\nu$C–N) of amide I and II (Figure 4.2) are highly sensitive to secondary structure folding upon chain packing in the crystalline region (Miyazawa et al., 1956). The intensity of the amide II band resulted in these cells not being as strong as the amide I band and overlapped with an absorbance band originating from the $\nu$C=O of the lignin aromatic ring (1516 cm⁻¹–shoulder at the left side of amide II band, Figure 4.2D) as reported in other plant tissues (Suarez-Garcia, Martinez-Alonso, & Tascon, 2002). Considering the interference from polysaccharides, oil and lignin, amide I is the sole absorption band from the protein backbone that can be used to probe protein distribution and secondary structure features. According to the chemical images, more protein accumulated in the cell center (Figure 4.2C).

The lipid distribution map was generated using the band at 1740 cm⁻¹ that is due to stretching of carbonyl groups in lipid esters. Lipids accumulated in the cell periphery (Figure 4.2C). This distribution pattern of protein in the middle and lipid bodies at the edge of the cell has been reported for Arabidopsis and other oilseeds (Otegui, Herder, Schulze, Jung, & Staehelind, 2006). The two main cell wall biopolymers, cellulose and lignin, roughly outlined the cell wall (Figure 4.2C).
Figure 4.2 Selection of protein rich areas in Arabidopsis seed tissues and analysis of protein secondary structure components. (A) Visible microscopic image of Arabidopsis WT seed. (B) Close-up view of cotyledon parenchyma cells with arrowheads showing the protein storage vacuoles (PSV). (C) 2D and 3D chemical images of protein (amide I), lipid (νC=O of lipid ester), cellulose (1059 cm\(^{-1}\)), and lignin (1513 cm\(^{-1}\)). Arrowheads show protein-rich area on the protein chemical image. Chemical intensity bars indicate color and scale assigned for IR absorbance, higher value indicates higher biopolymer concentration. (D) FT-IR spectra of the white-circled point on the visible image (C) of PSV, inset shows vibrational modes of amide I and amide II backbone. (E) Fourier self-deconvoluted peaks (dashed grey line) of the amide I band. Labels: (β-sheet)_A; anti-parallel β-sheet, (β-sheet)_P; parallel β-sheet, and SC; side chain.

The protein-rich areas include protein storage vacuoles (PSV) and provide less scattering interference from carbohydrates and oil bodies in the synchrotron FT-IR spectra. Visible images of PSVs in the cells showed that they are globular in shape, and the 2D and 3D chemical images showed protein-rich regions that may contain relatively pure protein (Figure 4.2) and can be used for exploring details of protein structure in its natural setting.

Fourier self deconvolution (FSD) (Kauppinen, Moffatt, Mantsch, & Cameron, 1981) permits separation of individual component peaks of helical, β-sheet, β-turn, and random structures that
overlap within the broad amide I region (Figure 4.2E). Deconvolution of amide I region of WT protein resolved nine individual bands: α-helix at 1659.7 ± 0.7 cm⁻¹, β-sheet at 1616.8 ± 0.8, 1627.8 ± 1.3, 1638.4 ± 1.7, and 1694.1 ± 0.2 cm⁻¹, β-turn at 1671.0 ± 0.3 and 1682.3 ± 0.3 cm⁻¹, and random structure at 1649.1 ± 1.5 cm⁻¹, which were assigned to secondary structure features according to published references (Byler & Susi, 1986; Dong et al., 1990).

Similar to the WT line, chemical imaging of a small area (900 μm²) close to the size of a parenchyma cell in the embryo of double-knockout and napin-RNAi lines showed a high concentration of protein (integral peak area of amide I band at 1650 cm⁻¹) in the mid cell area and a high concentration of lipid (integral peak area of νC=O lipid ester band at 1740 cm⁻¹) near the cell periphery (Figure 4.3). Conversely, the distribution pattern of oil and protein in the cruciferin depleted triple-knockout line indicated that the PSV were much smaller, contained very little protein, and that the remaining seed protein was more evenly distributed throughout the cytoplasm (Figure 4.3). The observations indicated that disruption of any two cruciferin genes or suppression of napin production did not result in an aberrant PSV phenotype. However, complete elimination of cruciferin production caused phenotypic differences in the size and distribution of PSVs. In the cells of B. napus seed, both protein and lipids were concentrated in the middle of the cell. This study clearly demonstrates the capability of using synchrotron FT-IR for probing the distribution and abundance of macromolecules in plant cells.

4.4.3 Chemical Makeup of Seeds

Crucifer seeds contain primarily lipid, protein, and carbohydrates other than starch; however, only the endosperm and embryo store lipid and protein. Signal intensities for lipid and protein in the endosperm and embryo of the SSP-altered lines were calculated using FT-IR spectral data (Figure 4.4). According to this estimation, in WT seed the protein concentration was higher ($P < 0.05$) in the endosperm (55%) than in the embryo (36%), whereas the lipid content was greater in the embryo (10%) than the endosperm (6%). In the SSP-altered lines, the protein content of the endosperm (Figure 4.4A) was lower ($P < 0.05$) (32–43%) than the WT (55%). Similarly, the embryo tissues of the double-knockout lines also showed lower ($P < 0.05$) protein content (31–32%) than that of WT seed (37%).
Figure 4.3 Distribution of proteins (amide I at 1650 cm$^{-1}$) and lipids (νC=O of lipid ester at 1740 cm$^{-1}$) in a single cell area of Arabidopsis (A: WT, B: CRUAbc, C: CRUaBc, D: CRUabC, E: CRU−, F: napin-RNAi) and B. napus (G) seeds. (i) Visible image (black square shows the scanning area); (ii) and (iii) 2D and 3D chemical images of proteins and lipids, respectively.
Figure 4.4 Protein and lipid distribution in Arabidopsis and B. napus seeds. (A) Protein localization in the endosperm and embryo was calculated by integrating the amide I band at peak 1650 cm$^{-1}$ (integral baseline 1720–1588 cm$^{-1}$). (B) Lipid localization in the endosperm and embryo according to the area of lipid ester peak at 1747 cm$^{-1}$ (integral baseline 1786–1720 cm$^{-1}$). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (ANOVA). Error bars, ± SEM ($n = 10$).

As expected, disruption of all three cruciferin genes resulted in markedly reduced (~51% less than WT) SSP accumulation in the embryo, but the accumulated protein content in the endosperm was higher (~40%) than that of double-knockout lines (Figure 4.4A).

When whole seed protein levels were calculated on the basis of total nitrogen content (Dumas combustion method), the cruciferin altered lines were between 25.7 and 30.6% (Table 4.4). The CRUAbc and CRUaBc lines had significantly higher ($P < 0.05$) protein values than WT, whereas the protein content of CRUabC line was comparable ($P > 0.05$) to that of the WT. The triple-knockout line had a protein content of 25.7%, which was 1.4 percentage points lower ($P < 0.05$) than that of the WT type (27.1%) (Table 4.4). However, the protein content calculated using the peak area of the amide I band indicated that the triple-knockout line had at least 50.0% less protein content in the embryo and 20% less protein in the endosperm cells compared to WT (Figure 4.4). The calculation of protein content using synchrotron FT-IR amide I peak is highly selective for protein molecules because the characteristic in-plane vibrational mode of the peptide chain (80% νC=O, 10% νC–N) generates the particular IR signal (Miyazawa et al., 1956; Susi, 1972). The total N-based protein estimation includes protein/peptides, free amino acids, DNA, RNA, and some phospholipids that contain nitrogen, but not necessarily the peptide amide bond. Therefore, the slight change in the total N-based protein value of the cruciferin triple-knockout line compared to WT indicates the possible accumulation of free amino acids or incorrectly assembled proteins. A similar observation was made in SSP-deficient lines of soybean (Schmidt
& Herman, 2008) and common bean (Marsolais et al., 2010), in which rebalancing of proteome via accumulation of free amino acids or sulfur-rich proteins has occurred.

**Table 4.4** Protein (%N based) and crude oil content of Arabidopsis and *B. napus* seeds

<table>
<thead>
<tr>
<th>Line</th>
<th>Protein (%)</th>
<th>Oil (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>27.1 c</td>
<td>40.0 a</td>
</tr>
<tr>
<td>Double-knockouts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRUAbc</td>
<td>30.6 a</td>
<td>36.3 d</td>
</tr>
<tr>
<td>CRUaBc</td>
<td>27.9 b</td>
<td>34.2 e</td>
</tr>
<tr>
<td>CRUabC</td>
<td>26.7 cd</td>
<td>36.5 cd</td>
</tr>
<tr>
<td>CRU–</td>
<td>25.7 e</td>
<td>37.3 c</td>
</tr>
<tr>
<td>Napin-RNAi</td>
<td>26.3 de</td>
<td>38.4 b</td>
</tr>
<tr>
<td>SEM</td>
<td>0.39</td>
<td>0.55</td>
</tr>
<tr>
<td><em>B. napus</em></td>
<td>28.5 ± 0.4</td>
<td>41.8 ± 0.1</td>
</tr>
</tbody>
</table>

*Means of triplicate analysis followed by the same letter within a column do not differ significantly (*P* < 0.05).

**4.4.4 Protein Secondary Structure Features**

The effect of expressed gene composition on the structure of accumulated protein was evaluated by deriving secondary structure features using Fourier self deconvolution (FSD) and Gaussian curve fitting of amide I bands of the respective synchrotron FT-IR spectra (Table 4.5 and Figure 4.5). In the double-knockout lines, genetic alteration of cruciferin expression did not affect napin expression (Figure 4.1C); therefore, it is expected that the secondary structure feature differences observed for these lines in the synchrotron FT-IR study are due to the changes in cruciferin subunit composition.

Interestingly, synchrotron FT-IR *in situ* localization revealed that suppression of napin gene expression (napin-RNAi) did not reduce the total protein level in the embryo, but did so in the endosperm (Figure 4.4A). Examination of an Arabidopsis line expressing an open reading frame encoding a napin-GFP fusion protein driven by a napin promoter revealed that napin accumulation is considerably higher in the endosperm compared to the embryo (data not shown). This preferential accumulation of napin would explain why protein content was reduced only in the endosperm, and not the embryo, in the napin-suppressed line. Seed protein levels based on total N of napin-RNAi line remained fairly close to the value of WT (Table 4.4).
Table 4.5 Secondary structure features (%) of embryo seed storage protein of Arabidopsis and B. napus seeds assessed using synchrotron FT-IR

<table>
<thead>
<tr>
<th>Line</th>
<th>α-helix</th>
<th>β-sheet</th>
<th>turn</th>
<th>random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>15.4</td>
<td>45.3</td>
<td>25.9</td>
<td>13.5</td>
</tr>
<tr>
<td>Double-knockouts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRUAbc</td>
<td>15.8</td>
<td>44.9</td>
<td>26.0</td>
<td>13.3</td>
</tr>
<tr>
<td>CRUaBc</td>
<td>18.1</td>
<td>40.7</td>
<td>27.7</td>
<td>13.6</td>
</tr>
<tr>
<td>CRUabC</td>
<td>15.6</td>
<td>42.2</td>
<td>26.5</td>
<td>15.7</td>
</tr>
<tr>
<td>Napin-RNAi</td>
<td>13.4</td>
<td>44.1</td>
<td>28.2</td>
<td>14.3</td>
</tr>
<tr>
<td>SEM</td>
<td>0.41</td>
<td>0.47</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>B. napus</td>
<td>14.8 ± 1.2</td>
<td>45.0 ± 1.4</td>
<td>27.1 ± 1.6</td>
<td>13.2 ± 0.8</td>
</tr>
</tbody>
</table>

Means of triplicate analysis followed by the same letter within a column do not differ significantly (P < 0.05).

Means of triplicate analysis followed by the same letter within a column do not differ significantly (P < 0.05). WT, wild type; double-knockout lines CRUAbc, CRUaBc, and CRUabC; SEM - standard error of the mean. B. napus data (mean ± SD) are presented for comparison.

The proteins of the WT line showed a high β-sheet content (45.3%) compared to α-helix (15.4%), turns (25.9%) or random (13.5%) structures (Table 4.5), indicating the predominance of cruciferin with its β-barrel structure typical of the cupin superfamily (Tandang-Silvas et al., 2010). Comparison of double-knockout lines with WT line showed that proteins in the CRUAbc line had similar levels (P > 0.05) of secondary structure components (Table 4.5); however, proteins in CRUaBc and CRUabC lines showed some differences, in particular, a low β-sheet content (P < 0.05). The level of α-helix and turns (P < 0.05) in protein from the CRUaBc line was higher than that in the WT. These observed differences in α-helix and β-sheet contents may be related to the level of cruciferin and napin present in the double-knockout lines rather than the structural changes of each protein. However, the homohexameric protein from CRUabC line exhibited significantly (P < 0.05) higher value for random structures than protein from WT, CRUAbc, or CRUaBc lines. This increased random structure in protein in the CRUabC line may be related to the disordered regions arising from its especially long HVR-I region (Chapter 3). The most notable change was observed for the triple-knockout line, with high (P < 0.05) α-helix (24.0%) and low β-sheet (34.4%) contents compared to cruciferin-expressed lines, indicating that the remaining or compensatory proteins deviate from the native cruciferin structure. Furthermore, the slightly higher contents of turns (26.6%) and random coils (14.9%) of this line indicated a less ordered structure for these proteins.
Figure 4.5 Analysis of deconvoluted synchrotron FT-IR spectra. (A) Identification of secondary structure elements in the spectra of Arabidopsis and *B. napus* endosperm tissues, peak labeling is similar to Figure 4.2. (B) Comparison of content of secondary structure elements of embryo tissue proteins of WT and napin-RNAi Arabidopsis lines. *P* < 0.05, **P** < 0.01, ***P** < 0.001 (ANOVA). Error bars, ± SEM. (*n* = 10).

In the napin-RNAi line, embryo proteins showed the lowest α-helix content (13.4%) (Table 4.5) and indicated lost contribution from napin to the helical components. Besides that, the endosperm of this line had a lower (7.8%) (*P* < 0.05) level of α-helix content than the WT (Figure 4.5B), indicating that the endosperm and the adjacent cell layer are lower in protein with helical secondary structure than the WT or others. The increased level of β-sheet content (50 vs 34.4% in WT; Figure 4.5B) of endosperm proteins indicated the reduction of predominantly helical napin in the accumulated protein. Immunolocalization of 2S albumin in developing Arabidopsis seed showed that napin is localized in endothelial cells in the inner integument layer in direct contact with the endosperm (Ondzighi, Christopher, Cho, Chang, & Staehelin, 2008).
Proteins in embryo tissues of the napin-RNAi line had similar secondary structure features as WT.

PCA of the values of deconvoluted amide I band also differentiated the aberrant phenotypes (Figure 4.6). PCA showed clear separation of CRU triple-knockout line from the WT and other mutant lines based on FSD protein secondary structure features (Figure 4.6). The first two components accounted for 79.13% of the total variance, but separation of the double-knockout lines was not possible (Figure 4.6A). The first and third principal components accounted for 51.85 and 20.87%, respectively, of the variability and distinguished the double-knockout lines expressing only CRUA gene from those expressing the CRUB or CRUC gene (Figure 4.6B). The secondary structure differences between the WT and the double-knockout lines or napin-RNAi line proteins did not separate well in PCA analysis (Figure 4.6); however, some degree of clustering was observed due to subtle differences in the dominant secondary structure features of the homohexameric cruciferin assemblies.

**Figure 4.6** PCA of the deconvoluted amide I band of WT and SSP-altered Arabidopsis: Biplot of the first two principal components (A) and first and third principal components (B). The first, second, and third principal components explain 50.70, 26.41, and 22.89%, respectively, of total variance in the protein secondary structure elements.
According to the results of protein distribution and protein secondary structure features of seed storage protein altered seeds, protein accumulation in PSV of double-knockout lines was not much different than in the WT. The differences in secondary structure features indicated the cruciferin core structure is primarily retained in the proteins of these lines. Synchrotron FT-IR revealed subtle differences in the accumulated proteins in these lines, showing the utility of this technique in assessing protein modifications without removing these molecules from tissues.

4.4.5 Pepsin-Assisted Degradation of Tissue Proteins

Bioaccessibility of seed tissues and their component molecules is important to release nutrients from the seed matrix. The in vitro digestibility of intact starch (e.g., corn, wheat, and potato), lipids, and proteins (almond and peanut) within seeds and cells has been examined using light, scanning electron, and transmission electron microscopy (Chambers et al., 2004; Ellis et al., 2004; Lee et al., 2011; Mandalari et al., 2008). In the case of proteins, release of amino acids and peptides to make them available for absorption in the gastrointestinal tract is a measure of bioavailability or digestibility. Pepsin is the first proteolytic enzyme that starts degrading proteins in a monogastric digestive system; therefore, assessment of pepsin accessibility relates to digestibility and the potential to release small peptides and amino acids of a particular protein-containing substrate. Disruption of enzyme susceptible peptide bonds leads to structural changes in the protein and provides information on protein bio-accessibility, which is linked to nutritional quality (Yu, 2005c; Yu et al., 2008). However, obtaining details about the extent of degradation and quantifying secondary structure are highly unlikely using the microscopic techniques mentioned above.

The protein chemical maps of untreated and pepsin-treated WT and SSP-altered seed tissues were constructed by compared the integrated amide I band (Figures 4.7 and 4.8A). Proteins in the WT, CRUAbc, CRUaBc, and CRU− lines displayed a significant ($P < 0.05$) reduction in the intensity of the integrated amide band, whereas the CRUabC seed proteins did not show such a reduction in pepsin-treated tissues (Figures 4.7 and 4.8A). The greatest depletion of SSP was in the napin-RNAi line (Figure 4.8A). Reduction of protein content in the pepsin-treated tissues indicates the propensity of protein to degradation, because the FT-IR amide I band is related to the intact peptide bond.
Figure 4.7 Visible microscopic images (scale bar = 100 µm) and synchrotron FT-IR 2D chemical images of untreated and pepsin-treated seeds of Arabidopsis WT (A), CRUAbc (B), CRUaBc (C), CRUabC (D), CRU− (triple-knockout) (E), napin-RNAi (F) and B. napus (G) seeds. Integration of peak 1650 cm\(^{-1}\) (amide I) was used to obtain the chemical images. Chemical intensity bars indicate color and scale assigned for IR absorbance; a higher value indicates higher biopolymer concentration.
Figure 4.8 Effect of pepsin treatment on Arabidopsis and B. napus SSP. (A) Comparison of protein concentration calculated by integrating the amide I band. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (ANOVA). Error bars, ± SEM ($n = 10$). (B) Deconvoluted synchrotron FT-IR spectra showing secondary structure elements of pepsin-treated Arabidopsis and B. napus seed. Peak identification and spectra of untreated samples are similar to Figure 4.2.

Although an overall reduction of total protein content was observed due to hydrolysis, details of secondary structure features indicated that the IR absorption peaks of α-helix and β-sheet of FSD spectra responded differently depending on the genetic composition of the line (Figure 4.8B; Table 4.6). The β-sheet component of proteins from WT, triple-knockout, and napin-RNAi lines was reduced in pepsin-treated samples, whereas random structure content increased in all samples except proteins from the CRUabC line (compare Table 4.5 and Table 4.6). When cruciferin and napin structure are concerned, the α-helix organization is more stable to hydrolysis than β-sheet structure (Yu, 2005c; Koppelman et al., 2005), especially when S–S bonds are involved in stabilizing protein structure such as in napin (Mandalari et al., 2008). The napin-RNAi line does not contain napin and showed the greatest reduction of α-helix and β-sheet contents and the highest increase in random structure content of FSD spectra among the lines showing degradation of cruciferin present in the PSV. The presence of napin in the double-knockout lines and the stability of this protein to pepsin-catalyzed hydrolysis may have contributed to the minor changes
observed in $\alpha$-helix content upon pepsin treatment. Therefore, only the changes in $\beta$-sheet content may indicate the propensity of respective cruciferin homohexamer to pepsin-catalyzed degradation. Among the cruciferin altered lines, $CRUabC$ showed the least change in protein secondary structure components after pepsin treatment.

### Table 4.6 Protein structure features (percent) of pepsin-treated Arabidopsis and $B. napus$ seeds

<table>
<thead>
<tr>
<th>Line</th>
<th>$\alpha$-helix</th>
<th>$\beta$-sheet</th>
<th>turn</th>
<th>random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>16.8 $^b$</td>
<td>34.2 $^d$</td>
<td>25.4 $^b$</td>
<td>23.6 $^a$</td>
</tr>
<tr>
<td>Double-knockouts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CRU Abc$</td>
<td>15.4 $^b$</td>
<td>47.5 $^a$</td>
<td>18.3 $^d$</td>
<td>18.8 $^b$</td>
</tr>
<tr>
<td>$CRU aBc$</td>
<td>15.7 $^b$</td>
<td>44.2 $^b$</td>
<td>21.8 $^c$</td>
<td>18.8 $^b$</td>
</tr>
<tr>
<td>$CRU abC$</td>
<td>15.6 $^b$</td>
<td>45.0 $^{ab}$</td>
<td>24.9 $^b$</td>
<td>14.2 $^c$</td>
</tr>
<tr>
<td>$CRU-$</td>
<td>30.2 $^a$</td>
<td>24.5 $^c$</td>
<td>26.8 $^b$</td>
<td>18.5 $^b$</td>
</tr>
<tr>
<td>Napin-RNAi</td>
<td>8.1 $^c$</td>
<td>38.1 $^c$</td>
<td>30.2 $^a$</td>
<td>23.6 $^a$</td>
</tr>
<tr>
<td>SEM</td>
<td>0.89</td>
<td>1.29</td>
<td>0.57</td>
<td>0.56</td>
</tr>
<tr>
<td>$B. napus$</td>
<td>17.4 ± 3.1</td>
<td>42.5 ± 6.7</td>
<td>24.4 ± 4.0</td>
<td>15.7 ± 2.2</td>
</tr>
</tbody>
</table>

$^a$Means of triplicate analysis followed by the same letter within a column do not differ significantly ($P < 0.05$). $^b$WT, wild type; double-knockout lines $CRU Abc$, $CRU aBc$, and $CRU abC$; SEM - standard error of the mean. $B. napus$ data (mean ± SD) are presented for comparison.

The location of cleavage sites in a solvent accessible area of the molecule provides for reasonably good digestibility (Chapter 3). Among the double-knockout lines, the low degradability of protein in the $CRU abC$ line may be related to the structural features of dominant CRUC homohexamer. According to homology modeling and proteolytic site prediction, CRUA and CRUB cruciferin homotrimers have pepsin cleavage sites in both HVR-I and HVR-II, with HVR-I directed toward the trimer periphery and HVR II in the trimer center and on the exposed trimer IE face. Although the HVR-I of the CRUC hexamer structure is long and projects towards solvent, the absence of pepsin cleavage sites could lead to slow hydrolysis. However, dissociation of the hexamer into trimers may allow solvent-accessible pepsin-labile sites of HVR-II of CRUC (Chapter 3). In the PSVs, dissociation of cruciferin hexamers into trimers is less likely, although a pH as low as 2 was provided to facilitate pepsin–catalyzed hydrolysis. These factors may have contributed to the low degradation of the CRUC homohexamer during pepsin treatment.

Even though the protein secondary structure content of $B. napus$ was similar to Arabidopsis (Table 4.5), the former was more resistant to pepsin hydrolysis (Table 4.6). Less ordered protein structure was expected in the triple-knockout line and possibly contributed to the better
degradation of its proteins by pepsin. When all data are considered, pepsin-treated tissues of WT, CRUAbc, CRUaBc, and triple-knockout (CRU−) lines showed that the first three PCs of FSD secondary structure data differentiated from the nontreated counterparts (Appendix Figure A2). Secondary structure assessment of pepsin-treated tissues may be able to predict the propensity of proteins for pepsin hydrolysis.

4.5 Conclusions

The synchrotron FT-IR technique is very useful in mapping biopolymer distribution within a seed and in individual cells. This technique also allows investigation of the secondary structure features of proteins directly within plant tissues without extraction. When proteins with different secondary structure features are present, changes in their composition can be probed using this technique as evidenced in the cruciferin double-knockouts and napin-RNAi lines. A single cell can be used as a unit for identifying changes in protein composition using synchrotron FT-IR; this is especially valuable in evaluating protein structure and composition changers due to alterations in genetic composition. The predominant secondary structure features were not greatly affected in lines expressing only a single cruciferin protomer indicating that normal 11S cruciferin can be obtained from these seeds. Assessment of secondary structure features can be linked to the accessibility of proteins to hydrolytic enzymes, such as pepsin, which are important in assessing the nutritional quality of proteins.

4.6 Connections to Subsequent Chapters

Seed proteins accumulated in Arabidopsis lines expressing only one cruciferin subunit type (double-knockout; CRUAbc, CRUaBc, or CRUabC) showed secondary structure features that were similar to WT in their native storage organelle. To understand the relationships between the structure of these proteins with their physico-chemical and functional properties, proteins isolated from seeds have to be studied. Chapter 5 describes studies conducted to probe structure details of WT heterohexameric and CRUA, CRUB, and CRUC homohexameric protein isolated from Arabidopsis seeds.
5. STRUCTURE AND PHYSICO-CHEMICAL PROPERTY RELATIONSHIPS OF CRUCIFERIN HOMOHEXAMERS*

5.1 Abstract

Heteromeric cruciferin from WT Arabidopsis and homomeric cruciferin CRUA, CRUB, and CRUC composed of identical subunits obtained from double-knockout mutant lines were investigated for their structural and physico-chemical properties. A three-step chromatographic procedure allowed isolation of intact cruciferin hexamers with high purity (>95%). FT-IR and CD analysis of protein secondary structure composition revealed that all cruciferins were folded into higher order structures consisting of 44–50% β-sheets and 7–9% α-helices. The structural and physico-chemical properties of homohexameric CRUC deviated from that of CRUA and CRUB and exhibited a compact, thermostable, and less hydrophobic structure, confirming the predictions made using 3D homology structure models.

5.2 Introduction

Cruciferin is the predominant seed storage protein (SSP) of crucifer oilseeds, such as canola/rapeseed and mustard. As such, it is a very important contributor to the nutritional and functional properties of the crucifer seed protein fraction. Cruciferin is an 11–12S globulin with a hexameric quaternary structure composed of six subunits (protomers). Each protomer consists of two polypeptides, an \( \alpha \) (acidic) and \( \beta \) (basic) chain (Tandang-Silvas et al., 2010). The wild type (WT) cruciferin is a heterogeneous mixture of subunits contributed by families of homologous genes. Therefore, cruciferin isolated from natural crucifers is an undefined mixture of these subunits. In *Brassica napus* L., at least five different cruciferin subunits have been reported, denoted as CRUA (Cru2/3a), CRU1, CRU2, CRU3, and CRU4 (Simon et al., 1985; Rödin et al., 1990b), although transcriptional studies indicate that the genome contains at least 18 cruciferin genes (Hegedus, unpublished). Understanding the structure and structure-related functions of cruciferin protomers will greatly assist in improving the utility of the crucifer protein fraction; however, data obtained from WT cruciferin does not allow assessments of each protomer type, as it is a mixture of several subunit variants.

Current understanding of the structural, physico-chemical, and functional properties of cruciferin is limited to a procruciferin Cru2/3a trimer subunit that was expressed in *Escherichia coli* (Tandang et al., 2004, 2005; Tandang-Silvas et al., 2010). In bacteria, procruciferin is not post-translationally processed by asparaginyl endopeptidase that occurs in plants. Using the glycinn (11S protein from soy bean) protomer, Dickinson et al. (1989) showed that post-translational cleavage is necessary for final assembly of the hexamer in protein storage vacuoles (PSV). Also, the spatial position of hypervariable or disordered region IV (HVR-IV) of the glycinn protomer changes upon post-translational processing (Adachi et al., 2001), indicating that different regions of the molecule are exposed in the mature hexamer compared to the pro-trimeric structure. The HVRs play a vital role in physico-chemical properties of the molecule and therefore the techno-functional properties and bioavailability of amino acids (Maruyama et al., 2004; Tandang et al., 2004, 2005). Furthermore, properties such as electrostatic potential and hydrophobicity in inter-chain (IE) and intra-chain (IA) faces of cruciferin and procruciferin are distinctly different (Adachi et al., 2001; Chapter 3), and may affect the ability of the protein to interact with other proteins and non-protein components. Understanding the structure–function relationships of major crucifer storage proteins at the subunit level will help to improve protein
quality for targeted applications for oilseeds, including food crops such as canola and biofuel crops such as *Brassica carinata* and *Camelina sativa*. To generate this information, we developed mutant Arabidopsis lines in which two of the three genes encoding cruciferin subunits (*AT5G44120.3*, *AT1G03880.1*, and *AT4G28520.1*) were inactivated by T-DNA insertion, yielding lines producing cruciferin composed of a single subunit (CRUA, CRUB, or CRUC). The objective of this study was to probe the structure of these homomeric cruciferins at the secondary and tertiary structure level, and correlate structure information with key physico-chemical properties important for techno-functionalities of these proteins.

5.3 Materials and Methods

5.3.1 Plant Materials

*Arabidopsis thaliana* (ecotype Columbia) cruciferin double-knockout lines (CRUAcrucruC or CRUAbc, cruACRUBcruC or CRuAbc, and cruAcruBcruC or CRuAbC), triple-knockout line (cruAcruBcruC or CRuabc or CRU−), and wild type (WT)(CRUACRUBCRUC or CRUABC) seeds were propagated under controlled greenhouse conditions (16-h ~800 Wm⁻² light at 21 °C and 8-h dark at 16 °C) as described in Study 2 (Chapter 4).

5.3.2 Isolation and Purification of Cruciferin

Isolation and purification of cruciferin from Arabidopsis seeds were as described by Bérot et al. (2005), with the modifications adopted by Wanasundara et al. (2012). Defatting of seed was according to the modified Swedish tube method (AOCS AM 2-93) using hexane as the solvent (3 g of seeds and 25 mL of hexane). Defatted meal recovered by filtering through Whatman #1 filter paper was dried under a fume hood to remove hexane. Protein extracts from defatted meals (1:10, w/v meal to buffer, ambient temperature, 1 h with mixing) were prepared with 50 mM Tris-HCl (pH 8.5), 750 mM NaCl, 5 mM EDTA and 0.3% (w/v) NaHSO₃, containing EDTA-free protease inhibitor cocktail (0.02 tablet mL⁻¹, Roche Diagnostics GmbH, Germany). The supernatant was recovered by centrifugation (15000g for 10 min) and the pellet re-extracted under same conditions. The supernatants were combined and stored at −20 °C.

Seed extract was first passed through a Sephadex G-25 HiPrep™ 26/10 desalting column (2.6 × 10 cm, 53 mL at a protein load of 70 mg protein mL⁻¹ gel with equilibration and elution buffer of 50 mM Tris-HCl containing 1 M NaCl at pH 8.5, 0.2–0.5 mL min⁻¹ isocratic flow rate) to
remove co-extracted pigments and small molecular weight compounds. The depigmented protein fraction (1–4, Figure 5.1A) was dialyzed (2 kDa MWCO) against deionized water for 48 h at 4 °C (3 to 4 water changes) and then lyophilized. The reconstituted protein fraction was then fractionated on a cation exchange column (CEC, Resource S methyl sulfonate attached to polystyrene/divinyl benzene, 1.6 × 9.2 cm, 18.5 mL) equilibrated with elution buffer A [50 mM Tris-HCl, 5 mM EDTA, 0.3% (w/v) NaHSO₃, pH 8.5] at a flow rate of 2.5 mL min⁻¹ and protein load of 25 mg protein mL⁻¹ gel. The unbound fractions 2–9 (Figure 5.1B) were combined, and dialyzed in the same manner as above and then freeze-dried. Absorbed proteins were eluted using buffer B (buffer A containing 1 M NaCl) at a linear gradient of 5 to 35%. The unbound protein was further purified on a Sephacryl S-300 HiPrep™ 26/10 high-resolution column (hydrophilic, allyldextran/bisacrylamide matrix, 2.6 × 60 cm, 320 mL) using elution buffer C consisting of 50 mM Tris-HCl (pH 8.5) and 1 M NaCl at an isocratic flow rate of 1 mL min⁻¹. The most abundant protein peak was collected, dialyzed, lyophilized, and stored at –20 °C until further use. An ÄKTA Explorer medium pressure chromatography system (Amersham Pharmacia, Uppsala, Sweden) was used for all three steps. Protein extract was filtered through a 0.45 µm Whatman GD/X Nylon syringe filter (Whatman Inc., Piscataway, NJ, USA) before being administered into the chromatography system. Elution of protein was monitored as absorbance at 280 nm and the proteins in each UV absorbing peak assessed by SDS-PAGE separation.

5.3.3 Protein Identification by Electrophoresis

SDS-PAGE and native-PAGE were carried out using precast minigels (PhastGel™ gradient 8–25) with PhastGel™ SDS and native buffer strips, respectively. For SDS-PAGE, protein solutions (2 µg protein µL⁻¹) were prepared in 62.5 mM Tris-HCl (pH 6.8) buffer containing 2% (w/v) SDS, 10% (w/v) glycerol, and 0.05% (w/v) bromophenol blue. For native-PAGE, non-reducing buffer [62.5 mM Tris-HCl, pH 6.8 containing 10% (w/v) glycerol] was used. One microliter of protein solution was loaded onto the precast gel with a PhastGel™ sample applicator. Electrophoresis was performed at a constant current of 60 mA per gel for approximately 45 min using a PhastGel™ system. Gels were stained with PhastGel™ Blue R (Coomassie R 350) stain in 20% (v/v) acetic acid and destained in 1:3:6 acetic acid/methanol/water (v/v/v) solution. The approximate molecular masses (Mr) of the separated
polypeptide bands were determined by comparison to molecular weight standards (6.5–200 kDa, Bio-Rad Laboratories, Hercules, CA) on the same gel.

Figure 5.1 Chromatographic separation of Arabidopsis protein extracts. (A) Depigmentation by SEC on a Sephadex G-25 column. Fractions 1–4 are total proteins. (B) Fractionation of the total proteins by CEC on a Resource S column. Unbound protein (mainly cruciferin) eluted in fractions 2–9 and bound fraction (mainly napin) eluted at 35% buffer B. (C) Purification of the CEC cruciferin fraction by SEC on a Sephacryl S-300 column. Inset: SDS-PAGE separation from left to right, MW ladder, peak I, II, III, and IV under non-reducing conditions. (D) Overlay of chromatogram from all samples after the final SEC.

5.3.4 Microfluidic LabChip Electrophoresis

Purified proteins were analysed using a microfluidic, chip-based automated electrophoresis system (Experion™ System, Experion™ Pro260 starter kit, Bio-Rad Laboratories) under non-reducing conditions. A 4 µL aliquot of protein (2 mg protein/mL in 10 mM Tris-HCl, pH 7.4) or Experion™ Pro260 Ladder was mixed with 2 µL of Experion™ Pro260 sample buffer containing
water, glycerol, and lithium dodecyl sulphate, and heated at 95 °C for 3–5 min and then diluted with 84 µL of DEPC (diethylpyrocarbonate)-treated water. The protein standard mixture contained β-mercaptoethanol (3.3%, v/v). The protein separation chip was primed with Experion™ Pro260 Gel containing water, methylurea, and N-(2-hydroxy-1,1-bis(hydroxymethyl)ethyl)glycine in the Experion™ Priming Station. Diluted samples (6 µL) were loaded on to the primed chip and analysed in the Experion™ Pro260 electrophoresis station. The subunit molecular weight and their abundance were determined using the position and integral area of fluorescence peaks relative to the Experion™ Pro260 Ladder proteins.

5.3.5 Protein Measurements

Total nitrogen content of mature seeds was determined as described in Chapter 4.

5.3.6 Fourier Transform Infrared Spectroscopy (FT-IR)

All FT-IR measurements were taken using a Bruker Vertex 70 FT-IR spectrometer (Bruker Optics, Billerica, MA) equipped with a KBr beam splitter in the interferometer and a liquid nitrogen cooled MCT detector. The dry protein powder was mixed with KBr powder (1:125, w/w) using a mortar and pestle and compressed to obtain 13-mm disks. The interferometer was purged with a constant flow of dry nitrogen gas. Each sample was scanned (128 scan) and data acquisition was done at 4 cm⁻¹ resolution. Protein secondary structure analysis was carried out using the Opus 6.5 software package (Bruker Optics, Billerica, MA, USA) according to the Fourier self deconvolution (FSD) method (Kauppinen et al., 1981) and Gaussian curve-fitting of the amide I region (1610–1700 nm) (Byler & Susi, 1986). The percentage of each secondary structure element was calculated from the relative integral area of each fitted curve of the amide I region. Peak assignments of components in the amide I band were as follows; α-helix (1659.1 ± 0.4 cm⁻¹), β-sheet (1616.7 ± 0.4, 1627.8 ± 0.2, 1638.8 ± 0.3, & 1693.3 ± 0.1 cm⁻¹), β-turn (1669.5 ± 0.2 & 1680.9 ± 0.2 cm⁻¹), and random structure (1649.8 ± 0.5 cm⁻¹) (Byler & Susi, 1986; Dong et al., 1990).

5.3.7 Circular Dichroism (CD) Spectroscopy

The far-UV circular dichroism (CD) spectra of protein solutions (2 mg mL⁻¹ in 10 mM sodium phosphate buffer at pH 7.4) were obtained at 25 °C using a PiStar-180 spectrometer (Applied
Photophysics Ltd., Leatherhead, U.K.) equipped with a 75 W mercury xenon lamp and 0.1 mm quartz cell and recorded from 180 to 260 nm using a 0.5-nm step size and 6-nm entrance and exit slits. The instrument was calibrated with 0.89 mg mL\(^{-1}\) d-(+)-10-camphorsulfonic acid (CSA). Ten scans per sample were averaged to obtain one spectrum and the baseline corrected by subtracting buffer spectra. The mean molar ellipticity \([\theta]\) was calculated using equation (5.1),

\[
[\theta](\text{deg cm}^2 \text{ dmol}^{-1}) = 100\langle\theta\rangle M/nCl
\]  

(5.1)

where, \(\theta\) is the measured signal in millidegrees (mdeg), \(M\) is the mean residue molecular weight (g mol\(^{-1}\)) of the protein (generally assumed to be 115), \(n\) is number of amino acid residues of cruciferin protomer, \(C\) is the molar concentration (mg cm\(^{-3}\)), and \(l\) is the path length (cm) of cuvette. Secondary structure of each test protein was estimated using the DicroWeb server (http://dichroweb.cryst.bbk.ac.uk), which employs the CONTIN/LL program with dataset 3 and CD spectra of 37 reference proteins.

5.3.8 Zeta Potential

The electrophoretic mobility of proteins in solution was measured as zeta (\(\zeta\))-potential with changing pH using a Zetasizer Nano ZS90 equipped with a multipurpose MPT-2 autotitrator (Malvern Instrument Ltd., Westborough, MA) at 25 °C. Protein dispersions (0.05%, w/v) prepared in deionized water and filtered through a 0.45 \(\mu\)m Whatman GD/X Nylon syringe filter (Whatman Inc., Piscataway, NJ, USA) were titrated from pH 10 to 2 (by 0.5 units) in a 1.5 mL disposable folded capillary cell (Malvern Instrument Ltd., Westborough, MA, USA) with 0.5 M and/or 0.005 M HCl or 0.01 M NaOH. The pH value of zero \(\zeta\)-potential was considered as the isoelectric point (pI) of the protein.

5.3.9 Intrinsic Fluorescence

The fluorescence emission spectra of protein in solution (50 \(\mu\)g mL\(^{-1}\) in 10 mM Tris-HCl buffer, pH 7.4, 20 °C) were recorded with a FluoroMax\textsuperscript{®}-4 Spectrofluorometer (Horiba Jobin Yvon, Edison, NJ, USA) with the excitation wavelength at 280 nm and emissions scanned from 290–430 nm (5 nm excitation and emission bandwidth).
Acrylamide was used as an aqueous nonionic quencher to characterize the microenvironment of Trp residues. The fluorescence intensity of 800 µL protein (50 µg mL\(^{-1}\)) upon addition of 10 µL aliquots of 5 M acrylamide solution was taken as described above. The classical Stern–Volmer equation (5.2) that describes the relationship between fluorescence intensity and quencher concentration for a dynamic bimolecular collisional quenching system (Stern & Volmer, 1919) was employed to calculate the Stern–Volmer constant \(K_{sv}\) of each protein,

\[
F_0/F = 1 + K_{sv} [Q]
\]

(5.2)

where, \(F_0\), \(F\), \([Q]\) and \(K_{sv}\) are initial fluorescence intensity without quencher, the fluorescence intensity of a given quencher concentration, quencher concentration, and the Stern–Volmer constant, respectively.

The modified Stern–Volmer equation (5.3) was used for calculating the association constant \(K_a\),

\[
F_0/(F - F_0) = 1/f_a K_a [Q] + 1/f_a
\]

(5.3)

where, \(f_a\) is the fractional accessibility of fluorophores. Protein solution (50 µg mL\(^{-1}\)) spectra were obtained with the addition of 6 M urea or 6 M guanidine-HCl and changing the pH from 2 to 10.

### 5.3.10 ANS Binding

Binding of 1-anilino-8-napthalensulfonate (ANS) to the protein was used as a measure of surface hydrophobicity \(S_0\) according to the modified method of Kato & Nakai (1980). ANS (10 µL of 8 mM stock solution) was added to 800 µL of protein solution (0.05–0.25 mg mL\(^{-1}\) in 10 mM phosphate buffer, pH 7.4), incubated for 5 min in the dark, followed by fluorescence emission scanning between 425 and 465 nm (excitation: 390 nm, slit widths: 5 nm). The \(S_0\) value (an index without units) was determined from the initial slope of the linear regression line fitted for the measured fluorescence intensity against protein concentration (mg mL\(^{-1}\)). Binding of ANS was evaluated at 23, 50, 70, and 90 °C with the temperature of the protein solutions (50 µg mL\(^{-1}\)) being maintained using a Peltier temperature controller (Model LFI-3751, Wavelength Electronics Inc., Bozeman, MT, USA).
5.3.11 Thermal Properties

Thermal denaturation parameters of proteins were obtained using a TA Q2000 differential scanning calorimeter (DSC) equipped with Plantinum™ software (TA Instruments, New Castle, DE, USA). Approximately 10 mg of 10% (w/v) protein slurry in 10 mM phosphate buffer (pH 7.4) was placed into aluminum liquid pans, hermetically sealed with Tzero™ press (TA Instruments, New Castle, DE, USA) and subjected to 30–130 °C at a scanning rate of 10 °C min⁻¹ under constant nitrogen purging (flow 50 mL min⁻¹). A hermetically sealed empty pan was used as a reference. Denaturation temperature \((T_m)\), width at half peak height \((\Delta T_{1/2})\), experimental enthalpy change \((\Delta H_{\text{exp}})\), and heat capacity change \((\Delta C_p)\) upon protein unfolding were computed from thermograms using the Universal Analysis 2000, version 4.7A software (TA Instruments-Water LLC). The calorimetric van’t Hoff enthalpy change \([\Delta H_{vH}(T_m)]\) for an oligomeric protein with n monomers, which is 12 for 11S globulins (Koshiyama, Hamano, & Fukushima, 1981), was calculated using equation 5.4, (Privalov & Potekhin, 1986).

\[
\Delta H_{vH}(T_m) = \left[\left(\sqrt{n} + 1\right)^2 R T_m^2 C_p \text{max} \right] / \Delta H_{\text{exp}}(T_m)
\]  

(5.4)

where, R is the gas constant 8.31451 J K⁻¹ mol⁻¹ and \(C_p \text{max}\) is the maximum heat capacity at denaturation. The Gibbs free energy of unfolding of protein can be written as equation 5.5,

\[
\Delta G(T_m) = -RT \ln K_{eq N \leftrightarrow U} = \Delta H_{\text{exp}}(T_m) - T_m \Delta S(T_m)
\]  

(5.5)

where, \(K_{eq N \leftrightarrow U}\) is the equilibrium constant, which is equal to 1 at \(T_m\) when native and unfolded state populations are equal. Therefore, entropy change upon denaturation \((\Delta S_{T_m})\) can be obtained from equation 5.6:

\[
\Delta S(T_m) = \Delta H_{\text{exp}}(T_m) / T_m
\]  

(5.6)

5.3.12 Statistical Analysis

All studies were carried out in triplicate. Data were analyzed using a one-way analysis of variance (ANOVA) using the General Linear Model (GLM) procedure of SAS Version 9.1 (SAS Institute Inc. Cary, NC, USA). Mean separation was carried
out and the least significant difference (LSD) was calculated when the main effect was significant ($P < 0.05$).

5.4 Results and Discussion

The Arabidopsis cruciferin genes AT5G44120, AT1G03880, and AT4G28520, encode the CRUA, CRUB, and CRUC cruciferin protomer (subunit) variants, respectively. These protomers randomly assemble as a heterohexamer and form the quaternary structure of WT cruciferin. The cruciferin mutant lines produce 11S protein composed of a single subunit type CRUA, CRUB, or CRUC (double-knockout lines). A triple-knockout line that does not express any of the cruciferin genes (CRU−) was also generated.

5.4.1 Purification of Cruciferin from WT and Mutant Lines

The first size exclusion chromatographic (SEC, Sephadex G-25) step excluded protein (cruciferin, napin, and other proteins) in the void volume (Figure 5.1A) and separated them from UV-absorbing phenolic compounds (Figure 5.1A). Further separation of the SEC protein peak by CEC resulted in elution of crude cruciferin in the unbound peak (Figure 5.1B). Napin remained bound to the CEC column and can be released later by increasing NaCl concentration (Bérot et al., 2005). The last SEC separation resolved four protein peaks in which the predominant protein peak II contained cruciferin subunit bands of 50–55 kDa, whereas peak I had very little native cruciferin and peaks III and IV contained small molecular weight polypeptides of ~13 kDa (Figure 5.1C and Figure 5.1C inset). Comparison of SEC profiles of all cruciferin types (Figure 5.1D) indicated that the triple-knockout line might contain a very low amount of cruciferin-like proteins. Although the plants of triple-knockout line do not express cruciferin genes, seed N content based on protein level remained comparable with WT (Chapter 4), possibly due to accumulation of non-cruciferin proteins, polypeptides, or free amino acids to compensate for the loss of cruciferin, similar to that reported in barley and soybean (Hansen et al., 2007; Pandurangan et al., 2012). Proteins obtained from this process were >95% pure.

5.4.2 Cruciferin Subunit Composition

According to native-PAGE, undissociated CRUA, CRUB, CRUC, and WT cruciferin had one type of native protein assembly, which separated a fairly close range upon migration. The CRU−
line had proteins that migrated as two different complexes (Figure 5.2A). Western blot analysis with an anti-cruciferin antiserum demonstrated that the proteins in the CRU– line are not cruciferin. This was confirmed by transcriptomic and proteomic analysis that revealed an increase general protein accumulation, including cupin domain proteins similar to the trimeric 7S globulins, which appear to compensate for the loss of cruciferin in this line (Hegedus, unpublished data). The calculated Mr of CRUA, CRUB, and CRUC subunits based on amino acid sequence were 52.6, 50.6, and 58.2 kDa, respectively (Chapter 3). Accordingly, each of the mutant lines appears to have produced a cruciferin complex with six subunits having hexameric native conformation (ca. Mr 303–350 kDa).

SDS-PAGE revealed polypeptides in the molecular mass range of 45 to 60 kDa (Figure 5.1C inset and Figure 5.2B) with intact disulfide bonds between the α and β subunits (α–S-S–β). The absence of protein bands of <20 kDa indicated no contamination of napin in the purified proteins (Bérot et al., 2005). The polypeptide bands of ~20 and ~30 kDa for WT and double-knockout lines are free cruciferin α- and β-chains that arise due to disulphide interchange which may occur during protein sample preparation as observed in Arabidopsis, B. napus and Raphanus sativus, (Inquello et al., 1993) or during intracellular transport (Rödin & Rask, 1990a). The three peaks identified in WT cruciferin by LabChip microfluidic electrophoresis were estimated to have molecular masses (Mr) of 53.5, 49.4, and 57.4 kDa, which can be assigned as CRUA, CRUB, and CRUC, respectively (Figure 5.2C). These values are very close to the calculated Mr of CRUA (52.6 kDa), CRUB (50.6 kDa), and CRUC (58.2 kDa) based on their primary sequences (Chapter 3).

The relative abundance of CRUA, CRUB, and CRUC subunits in WT cruciferin based on the fluorescence peak area was 9.6% (72.8 ng µL⁻¹), 32.8% (249.5 ng µL⁻¹), and 57.6% (438.5 ng µL⁻¹), respectively, and in an approximate ratio of 1:3:6 for CRUA: CRUB: CRUC subunits. On the basis of sequence coverage obtained from MS analysis, CRUC (AT4G28520.1) is the major cruciferin isoform contributing to total seed protein of A. thaliana cv. Columbia (Wan et al., 2007). Cruciferin of the CRUabc double-knockout line showed two partially resolved peaks at 51.6 and 54.6 kDa that gave an average value of 53.1 kDa, which matches with the CRUA peak of WT (Figure 5.2C). This type of partial resolution of protein peaks has been reported for Arabidopsis and pea 11S globulins resulting from disulphide interchange with free –SH groups during sample preparation in SDS-PAGE buffer as described by Rödin & Rask (1990a) and
Inquello et al. (1993) or upon thermal processing (Yamagishi, Miyakawa, Noda, & Yamauchi, 1983). The LabChip microfluidic sample buffer contains lithium dodecyl sulphate which may support disulphide –SH exchange similar to SDS when protein is heated at 95 °C for 2 to 3 min. Examination of molecular structure of cruciferin shows that CRUA and CRUB have free –SH groups in the proximity (9.4 Å apart) of the inter-chain disulfide bond at Cys293 and Cys280, respectively (Figure 5.3), whereas CRUC has no free accessible –SH group close to inter-chain disulfide bond. In the CRUB structure, the relatively long side chain of Thr277 and Met278 residues between the free –SH groups and inter-chain disulfide bond may pose a substantial barrier to disulphide –SH exchange reaction (Figure 5.3).

**Figure 5.2** Electrophoretic separation of isolated cruciferin from Arabidopsis. (A) Native-PAGE showing separation of cruciferins under non-dissociating conditions. All have hexameric forms except the triple-knockout, CRU−. (B) SDS-PAGE of purified cruciferin under non-reducing conditions showing intact S-S bonds in the subunits (α–S–S–β) and free α- and β-chains. (C) LabChip Microfluidic electrophoretic profiles of purified cruciferins under non-reducing conditions showing estimated molecular weights (kDa) of each cruciferin subunit.
Figure 5.3 Cartoon depicting the possibility of disulfide exchange in cruciferins. (A) The spatial arrangement around the inter-chain disulfide bond (C112⟷C289) and Cys293 of CRUA has less steric hindrance (AA, 289CSARC). (B) In CRUB replacement of Ser and Ala of CRUA by Thr and Met (AA, 276CTMRC) creates a barrier for Cys106 to exchange (arrow) with Cys280. There is no Cys residue (replace by Ser) in CRUC at the same position for exchange (AA, 340CSMRS) (described in Chapter 3). Structure illustrations are produced using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

In CRUA, there is no such steric interference with the corresponding amino acids and the short side chains of Ser290 and Ala291 would allow CRUA to undergo intramolecular sulfhydryl–disulfide exchange reaction. The major peaks of homomeric CRUB and CRUC have values of 49.2 and 57.3 kDa, which are fairly close to the corresponding subunits of WT Arabidopsis (Figure 5.2C). The CRU– protein gave three faint peaks at 48.0, 51.8, and 58.5 kDa that did not correspond to cruciferin subunits.

5.4.3 Secondary Structure Components of Cruciferin

FT-IR spectra of pure cruciferins revealed five amide bands; amide A (100% νN–H) at 3302–3307 cm⁻¹, amide B (100% δN–H) at 2931–2933 cm⁻¹, amide I (80% νC=O, 10% νC–N) at 1652–1653 cm⁻¹, amide II (60% δN–H, 40% νC–N) at 1536–1540 cm⁻¹, and amide III (40% νC–N, 30% δN–H, δC=O) at 1238–1241 cm⁻¹ (Figure 5.4A), which were identified according to Miyazawa et al. (1956) and Susi (1972). The strongest amide I band was used for FSD to calculate secondary structure components (Table 5.1).
All of the cruciferin CD spectra (Figure 5.4B) had a broad negative peak at about 208 nm ($\pi \rightarrow \pi^*$) and a positive peak in the 180–195 nm ($\pi \rightarrow \pi^*$) region; the cross-over point from negative to positive was at 200 nm, which is typical for an “$\alpha + \beta$ type” protein (Manavalan & Johnson, 1983). The characteristic negative band at 222 nm ($n \rightarrow \pi^*$) from $\alpha$-helices (Manavalan & Johnson, 1983) was not observed, in accordance with the less predominant $\alpha$-helical structure of cruciferin. Similar CD spectra were reported for native rapeseed cruciferin (84% pure) (Gerbanowski, et al., 1999) and soy glycinin (~90% pure) (Sze, Kshirsagar, Venkatachalam, & Sathe, 2007).

Both FSD of the FT-IR amide I (1610–1700 nm) band and CD analysis showed predominant $\beta$-sheet content. The $\alpha$-helix content of CRUA, CRUB, and CRUC was less than the contents of turn or random structure (Table 5.1). The predominant $\beta$-sheet content confirms previous reports on cruciferin secondary structure (Tandang-Silvas et al., 2010; Chapter 3). The random structure content of these cruciferins determined from CD spectra (25.2–26.2%) was different compared to that from FSD FT-IR (13.6–16.4%, Table 5.1). Among the homohexamers, FT-IR data from CRUC showed lower $\alpha$-helix and $\beta$-sheet contents and higher contents of turns and random structure than CRUA or CRUB (Table 5.1).

**Figure 5.4** Secondary structure analysis of purified Arabidopsis cruciferins. (A) FT-IR spectra of purified cruciferins showing the amide A, B, I–III peaks. Inset: Secondary structure components within the amide I peak (1650 cm$^{-1}$) resolved by FSD. (B) Far-UV CD spectra of cruciferins (pH 7.4) from 180–260 nm showing $\alpha + \beta$ type characteristics; a broad negative band at around 208 nm, a positive peak in the 195–180 nm region and a negative to positive cross-over point (inset) at around 198 nm. [$\theta$] is the mean residue ellipticity.
According to homology modeling (Chapter 3), the long hypervariable region of CRUC (HVR-I, 60 residues) causes the higher degree of random structure in CRUC, and this was confirmed by FT-IR data (Table 5.1). The lower ($P < 0.05$) $\beta$-sheet content (39.9% from FT-IR and 27.9% from CD) and higher $\alpha$-helix content (14.2% from FT-IR and 21.1% from CD) of the proteins remaining in the CRU− line, in combination with electrophoresis data, indicate that the compensatory proteins in the CRU− line are different from cruciferins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\alpha$-helix</th>
<th>$\beta$-sheet</th>
<th>turn</th>
<th>random</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>9.2 $^b$</td>
<td>44.1 $^d$</td>
<td>35.1 $^a$</td>
<td>11.6 $^e$</td>
</tr>
<tr>
<td>CRUA</td>
<td>8.3 $^c$</td>
<td>47.9 $^b$</td>
<td>28.9 $^c$</td>
<td>14.9 $^c$</td>
</tr>
<tr>
<td>CRUB</td>
<td>8.9 $^b$</td>
<td>50.3 $^a$</td>
<td>27.2 $^d$</td>
<td>13.6 $^d$</td>
</tr>
<tr>
<td>CRUC</td>
<td>7.2 $^d$</td>
<td>46.6 $^c$</td>
<td>29.9 $^b$</td>
<td>16.4 $^b$</td>
</tr>
<tr>
<td>CRU−</td>
<td>14.2 $^a$</td>
<td>39.9 $^e$</td>
<td>26.7 $^d$</td>
<td>19.2 $^a$</td>
</tr>
<tr>
<td>SEM</td>
<td>0.35</td>
<td>0.51</td>
<td>0.44</td>
<td>0.38</td>
</tr>
</tbody>
</table>

Secondary structures from far-UV CD spectra

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\alpha$-helix</th>
<th>$\beta$-sheet</th>
<th>turn</th>
<th>random</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>7.9 $^b$</td>
<td>44.6 $^a$</td>
<td>22.0 $^a$</td>
<td>25.6 $^a$</td>
</tr>
<tr>
<td>CRUA</td>
<td>7.4 $^b$</td>
<td>43.8 $^a$</td>
<td>22.9 $^a$</td>
<td>26.2 $^a$</td>
</tr>
<tr>
<td>CRUB</td>
<td>7.7 $^b$</td>
<td>44.6 $^a$</td>
<td>21.5 $^a$</td>
<td>26.1 $^a$</td>
</tr>
<tr>
<td>CRUC</td>
<td>7.8 $^b$</td>
<td>45.3 $^a$</td>
<td>21.9 $^a$</td>
<td>25.2 $^a$</td>
</tr>
<tr>
<td>CRU−</td>
<td>21.1 $^a$</td>
<td>27.9 $^b$</td>
<td>28.8 $^b$</td>
<td>22.4 $^b$</td>
</tr>
<tr>
<td>SEM</td>
<td>1.79</td>
<td>1.90</td>
<td>0.98</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Secondary structures from homology models

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\alpha$-helix</th>
<th>$\beta$-sheet</th>
<th>turn</th>
<th>random</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRUA</td>
<td>12.6</td>
<td>33.8</td>
<td>43.6</td>
<td>10.0</td>
</tr>
<tr>
<td>CRUB</td>
<td>13.0</td>
<td>32.3</td>
<td>48.7</td>
<td>6.0</td>
</tr>
<tr>
<td>CRUC</td>
<td>12.4</td>
<td>28.8</td>
<td>38.8</td>
<td>20.0</td>
</tr>
</tbody>
</table>

$^a$Means followed by the same letter within a column do not differ significantly ($P < 0.05$). SEM, standard error of the mean. Assays were conducted in triplicate.

$^b$Theoretical values were calculated as percentage of the number of residues in corresponding secondary structure to the total residues in the subunit (Chapter 3).

5.4.4 Electrical Potential of Cruciferins

The protein surfaces in an aqueous medium are naturally charged to form an electrical double layer, and the $\zeta$-potential is the potential at the boundary of this layer (surface of hydrodynamic shear). The $\zeta$-potential value of all cruciferins changed from negative to positive when pH was reduced from 10 to 2.0 (Figure 5.5).
Figure 5.5 Zeta (ζ) potential of cruciferins (0.1 mg mL⁻¹) at different pH. The locations and values of the isoelectric points (pI) where ζ = 0 are indicated. Error bars denote ± SD of triplicate measurements.

The proteins exhibited a high potential stability in the pH range of 7.0 to 10 as the ζ-potential remained between −57 and −17 mV (−30 mV considered as the dividing value between stable and unstable colloidal systems). The ζ-potential of CRUB remained well below −30 mV in the basic to neutral pH range with minor fluctuations. In contrast, the CRUC homohexamer, which showed the highest negative potential of −55 mV at pH 9.0, exhibited rapid change in ζ-potential when the pH approached neutrality.

All proteins at acidic pH (3.4 to 4.9), except CRUC, showed zero ζ-potential. During the titration of protein from pH 7.0 to 3.0, the ζ-potential of all cruciferins moved to a positive value with a high +28.0 mV for CRUC and +5.0 to +10.0 mV for CRUA, CRUB, and the WT heterohexamer. The change of ζ-potential beyond pH 3.0 was not drastic and approached zero at
a pH of 2.0. The differences in \( \zeta \)-potential change observed for CRUA, CRUB, and CRUC homohexamers indicated differences in the charges of their surface residues. According to homology models for these three cruciferins, the IA face (solvent exposed) of CRUC showed higher electronegativity that was spread more evenly over the entire surface area than that of CRUA or CRUB (Chapter 3), confirming the observation made on \( \zeta \)-potential change as a function of pH. The surface charge of these three cruciferins at pH ~7.4 is comparable with the ratio of acidic and basic amino acids CRUC (3.2:1) > CRUB (2.5:1) > CRUA (2.0:1) calculated based on primary sequences. When the proteins were titrated beyond the neutral pH to acidic pH, all cruciferins passed through a point where \( \zeta \)-potential becomes zero. The zero net charge or isoelectric point (pI) for CRUA, CRUB, CRUC, and WT cruciferins was at pH 4.9, 3.4, 6.5, and 3.8, respectively. The pI values calculated based on the amino acid composition of the subunits are 7.68 for CRUA, 6.53 for CRUB, and 6.53 for CRUC. Except for CRUC, the pH value of zero \( \zeta \)-potential was far different than the calculated value. No reports are available on experimental pI values for Arabidopsis or B. napus cruciferin. The pI value derived from the pH dependent solubility curve (minimum solubility at its pI) for procruciferin Cru2/3a was at pH ~ 4.2 (\( \mu = 0.08 \)) (Tandang et al., 2005), while the calculated value based on amino acid composition was pH 6.6 (Chapter 3), indicating that theoretical values do not always agree with observed values. The conformation of the protein in the solution affects the extent of residues available for ionization, which in turn reflects charge neutralization by H ions as the pH changes. As such, the 3D structure can cause a considerable difference between calculated and experimental pI values for native proteins (Henriksson, Englund, Johansson, & Lundahl, 1995). Among the cruciferins studied, the pattern of \( \zeta \)-potential as a function of pH indicated that CRUC has more ionizable residues on the solvent accessible surface than do the other cruciferins. The net electrical charge of a protein in given solvent strongly affects their physicochemical properties; therefore, the functionalities they provide.

5.4.5 Intrinsic Fluorescence

Intrinsic fluorescence due to the indole ring of tryptophan depends on the polarity of the surrounding microenvironment (Chen & Barkley, 1998). The number of Trp residues in the primary sequence of the cruciferin variants was considerably different, CRUA with 30, CRUB with 36, and CRUC with 54 (Table 5.2; Appendix Figure A3). However, only considerable
differences in the steady-state fluorescence emission or peak area were observed (Figure 5.6A), indicating a comparatively lower number of solvent-exposed Trp residues in CRUC than those in CRUA and CRUB. CRUA had the lowest \( \lambda_{\text{max}} \) (333 nm) among the cruciferins (Table 5.2), suggesting that a more hydrophobic environment exists around for the Trp indole moieties than in the other cruciferins, possibly buried in hydrophobic pockets (Dufour, Hoa, & Haertlé, 1994).

Acrylamide, a nonionic neutral quencher, decreases fluorescence emission intensity from Trp residues when it binds to or is in the proximity of a fluorophore. The increase in quencher concentration attenuated fluorescence emission intensity of all cruciferins (Figure 5.6B) and shifted \( \lambda_{\text{max}} \) to lower wavelengths (i.e., hypsochromic shift) (Figure 5.6B inset). The non-linear Stern–Volmer plots (Figure 5.6C) for all cruciferins exhibited an upward concave curvature towards the y-axis, especially at high acrylamide concentrations, indicating that the proteins exhibited both static and dynamic quenching processes (static via formation of a ground state fluorophore–quencher complex and dynamic via collision between fluorophore and quencher molecules) (Lakowicz, 1999). Fluorescence quenching at low acrylamide concentrations up to \( \sim 0.4 \text{ M} \) exhibited a linear \((r^2 > 0.99)\) relationship in the Stern–Volmer plots (Figure 5.6C inset), allowing the Stern–Volmer constant \((K_{sv})\) to be calculated for each protein (Table 5.2). The low \( K_{sv} \) value for the CRUC homohexamer \((6.23 \text{ M}^{-1})\) indicated less exposed Trp residues compared to the other cruciferin variants; therefore, the residues may be buried inside the structure. A less compact protein molecule may allow greater exposure of its fluorophores to the environment than a more compact molecule; therefore, the low \( K_{sv} \) may indirectly indicate a high degree of molecule compactness. The \( K_{sv} \) of CRUB homohexamer \((8.81 \text{ M}^{-1})\) and the WT cruciferin \((9.12 \text{ M}^{-1})\) also indicated less compact molecules compared to the CRUA or CRUC homohexmers. The modified Stern–Volmer plots (Lehrer, 1971) for cruciferins were linear (Figure 5.6D), indicating possible static quenching of heterogeneous Trp residues in cruciferins. Association constants \((K_a)\) calculated using the modified Stern–Volmer plots (Table 5.2) were the lowest \((3.91 \text{ M})\) for the CRUC homohexamer (Figure 5.6D), further suggesting that the Trp residues may be buried within hydrophobic regions (i.e., inter-chain disulfide containing face, IE face) of the hexamer, and may be shielded from the solvent. Such residues may undergo quenching by an aqueous agent when the hexamer is dissociated. Similar to the \( K_{sv} \) values, the highest \( K_a \) value of 13.62 M was observed for CRU−, which further confirmed a more relaxed structure for this protein.
Table 5.2 Various fluorescence parameters obtained from intrinsic fluorescence and probe binding methods

<table>
<thead>
<tr>
<th>Protein</th>
<th>No. of Trp residues&lt;sup&gt;b&lt;/sup&gt;</th>
<th>$\lambda_{\text{max}}$ (nm)</th>
<th>$K_{sv}$ (M&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>$K_a$ (M)</th>
<th>$S_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>-</td>
<td>338 ± 0.0</td>
<td>9.12&lt;sup&gt;b&lt;/sup&gt;</td>
<td>6.91&lt;sup&gt;b&lt;/sup&gt;</td>
<td>446.4&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>CRUA</td>
<td>30</td>
<td>333 ± 0.5</td>
<td>7.00&lt;sup&gt;c&lt;/sup&gt;</td>
<td>5.20&lt;sup&gt;c&lt;/sup&gt;</td>
<td>525.3&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>CRUB</td>
<td>36</td>
<td>345 ± 0.5</td>
<td>8.81&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5.43&lt;sup&gt;c&lt;/sup&gt;</td>
<td>444.5&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>CRUC</td>
<td>54</td>
<td>340 ± 0.0</td>
<td>6.23&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3.91&lt;sup&gt;d&lt;/sup&gt;</td>
<td>282.1&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>CRU−</td>
<td>-</td>
<td>351 ± 1.0</td>
<td>15.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1362&lt;sup&gt;a&lt;/sup&gt;</td>
<td>236.7&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>SEM</td>
<td>-</td>
<td>-</td>
<td>1.05</td>
<td>1.15</td>
<td>36.4</td>
</tr>
</tbody>
</table>

<sup>a</sup>Means followed by the same letter within a column do not differ significantly ($P < 0.05$). SEM, standard error of the mean.

<sup>b</sup>Calculated from the primary amino acid sequence (Chapter 3).

<sup>c</sup>$K_{sv}$: Stern–Volmer quenching constant, $K_a$: Stern–Volmer association constant, $S_0$: surface hydrophobicity.
Figure 5.6 Intrinsic fluorescence of cruciferin and fluorescence quenching due to acrylamide. (A) Fluorescence emission spectra for hexameric cruciferins. (B) Primary plot of fluorescence quenching as a function of acrylamide (C$_3$H$_5$NO) concentration. Inset: Quenching of cruciferin from WT Arabidopsis with the addition of acrylamide. (C) Stern–Volmer plots exhibiting positive deviation from linearity. Inset: Stern–Volmer plots showing linear relationships ($r^2 > 0.99$) at low quencher concentrations that were used to calculate Stern–Volmer quenching constant ($K_{sv}$). (D) Linearity of the modified Stern–Volmer plot indicating static quenching of different accessible populations of fluorophores (i.e., buried and exposed tryptophan residues). The plots allowed calculation of the Stern–Volmer association constant ($K_a$) of exposed tryptophan residues (values are shown in Table 5.2). Excitation was at 280 nm.
The association constant of the CRUA homohexamer did not differ significantly ($P > 0.05$) from that of CRUB, but was lower ($P < 0.05$) than the WT heterohexamer (Table 5.2).

Investigation of 3D structures of CRUA, CRUB, and CRUC homotrimers generated through comparative modeling (Chapter 3) showed that Trp residues of CRUA and CRUB homotrimers were scattered on the IA face ($Trp_{IA}$) and core region ($Trp_{core}$), and no residues were located on IE face (Figure 5.7). However, the Trp residues of CRUC located on both IA ($Trp_{IA}$) and IE ($Trp_{IE}$) faces included four residues (Trp132, Trp143, Trp159, and Trp178) on HVR-I ($Trp_{HVR-I}$, Figure 5.7). The $Trp_{IA}$ residues located on cruciferin hexamer surface may have direct contact with solvent. On the other hand, the $Trp_{core}$, and $Trp_{IE}$ residues are much more buried in the interior of the cruciferin hexamer and may have minimum or no contact with solvent. According to the Trp residues mapped on solvent accessible surface, in CRUA, Trp60 and Trp203, in CRUB, Trp26, Trp54, and Trp196, and in CRUC, Trp61 on IA face are solvent exposed (Figure 5.7 iii). Although HVR-I of CRUC homohexamer is supposed to be on the IE face, some Trp residues such as Trp132 and Trp143 (Figure 5.7C) may be surface exposed due to its length and flexibility, explaining the high $\lambda_{max}$ value (340 nm) (Figure 5.6A).

The intrinsic fluorescence data of cruciferins revealed that the area under the fluorescence peak are in the order of CRUA > CRUB > CRUC (Figure 5.6A inset). The fluorescence of the Trp indole moiety is highly sensitive to its microenvironrment, which can be self-quenched by adjacent amino acids, such as tryptophan, tyrosine, cysteine, and histidine (Chen and Barkley, 1998; Staiano et al., 2005). This intrinsic quenching happens from distinct excited-state energy transfer, through-space or through-bond, from adjacent amino acid side chains (<10 Å) to the $C_\delta_1$, $C_\varepsilon_3$, and $C_\zeta_2$ positions of the tryptophan indole ring (Chen and Barkley, 1998; Staiano et al., 2005). In the microenvironment of Trp60, Trp160, and Trp346 of CRUA has no intrinsic strong quencher in its vicinity (Figure 5.8A). However, Trp203 of CRUA is located close to the Tyr192 (5.2 Å) and efficient energy transfer between the two residues can be expected.

The three tryptophan residues at the center of the CRUB homotrimer (Trp26) are very close to each other (3 Å), and it is in the vicinity (10.8 Å) of the IA disulfide bond (C30$\leftrightarrow$C63), suggesting that Trp26 residue could be greatly quenched by these multiple interactions (Figure 5.8B). Similarly, the four $Trp_{HVR-I}$ residues of CRUC are located very close to each other (Trp143–Trp132, 9.3 Å; Trp132–Trp178, 2.7 Å; Trp178–Trp159, 7.8 Å) (Figure 5.7C); therefore, those residues can effectively transfer their excitation energies among them (Figure 5.8C).
Figure 5.7 Spatial distributions of Trp residues in CRUA (A), CRUB (B), and CRUC (C) homotrimers (each protomer is colored in yellow, red, and blue, respectively). Ribbon representations of cruciferin homotrimers (i and ii). All Trp residues of CRUA and CRUB are located on the IA face ($Trp_{IA}$) and core regions ($Trp_{core}$), no residues are observed on the IE face. In CRUC, Trp residues are located on both the IA and IE faces ($Trp_{IE}$), including four residues on HVR-I ($Trp_{HVR-I}$). HVR-I of CRUC may be located on the IE face (curved arrow); therefore, those $Trp_{HVR-I}$ residues may not be exposed to the solvent. The Trp residues on solvent-accessible surfaces of cruciferins are shown (iii). Three-fold symmetry axis (C3) is shown in arrows. Structure illustrations were produced using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).
Figure 5.8 Close-up view of the 3D structure of cruciferins in the vicinity of Trp160, Trp203, and Trp346 of CRUA (A), Trp26 of CRUB (B), Trp_{HVR-I} residues of CRUC (C), and Trp61 and Trp454 of CruC (D). Nomenclature of atoms in the indole ring is shown in the inset of (A). Large yellow and green spheres are IE disulfide bonds, yellow and purple spheres are IA disulfide bonds and yellow and blue spheres are free –SH groups (in all, yellow spheres are Cys residues). Structure illustrations are produced using PyMOL (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC).

The only surface exposed Trp residue of CRUC (Trp61), which is on the IA face, can also be intra-molecularly quenched by the side chain of Cys407 (Figure 5.8D), which may further contribute to diminishing the CRUC intrinsic fluorescence (Figure 5.7).
5.4.6 Structural Stability in the Presence of the Co-solvents Urea, Guanidine-HCl and NaCl or in Response to pH Change

The changes that occur in the Trp environment due to small organic molecules, such as urea and guanidine-HCl, can be followed by measuring intrinsic fluorescence (Figure 5.9) to provide information on protein structure stability. With the exception of CRUC, the other homohexamers and WT cruciferin showed a $\lambda_{\text{max}}$ redshift (Figure 5.9A, and 5.9C) with a concomitant increase in fluorescence intensity (Figure 5.9B). This indicates that urea caused unfolding to expose Trp residues. Guanidine-HCl caused a drastic decrease (<50%) in fluorescence yield for all cruciferins, other than CRUC. Of all the cruciferins (Figure 5.9A and 5.9B), fluorescence emission intensity of the CRUC homohexamer remained most similar to its native state in the presence of both chaotropic agents and the $\lambda_{\text{max}}$ value remained unchanged with 6 M urea, suggesting that either complete opening of the CRUC homohexamer does not occur easily or that intramolecular quenching of Trp residues happens upon exposure to these agents.

Sodium chloride can stabilize protein structure when present as a co-solvent, but this is dependent on its concentration. The presence of NaCl at concentrations ranging from 0.2 to 1.0 M had a negligible effect on $\lambda_{\text{max}}$. Except for CRUA, the WT, CRUB, and CRUC showed an increase in fluorescence intensity and peak area with NaCl, but there was no difference between the different concentrations (Appendix Figure A4). This indicates that the presence of NaCl up to 1.0 M causes minimal disturbance in the structural stability of these cruciferins at pH 7.4. Moreover, the salt effect (with, $\mu = 0.5$; and without, $\mu = 0.0$) on other pH (2, 4, 6, 8, and 10), was further investigated and no significant salt effect on the intrinsic fluorescence at other pH was found (Appendix Figure A5).

The $\lambda_{\text{max}}$ value showed a red shift for all cruciferins when the pH was reduced from neutral to acidic (Figure 5.10), which increased when the ionic strength was brought up to 0.5 at pH 7.4 (Appendix Figure A5). As the pH was lowered from neutral, the fluorescence intensity initially increased and then decreased for cruciferin, except WT, indicating possible exposure or change of the hydrophobic environment of Trp residues around pH 2.0. Continued increase in fluorescence intensity and $\lambda_{\text{max}}$ for CRUC and CRUA at pH as low as 2.0 indicated that an opposite sequence of events was occurring. Marchone and coworkers (1997) showed that red shift and decrease in fluorescence intensities under acidic conditions for globulin proteins are due to exposure of one or more hydrophobic surfaces, whereas maintaining secondary structure of the...
11S globulin. It is possible that at pH 2.0 the cruciferin hexamers dissociate, exposing the IE face of the trimers and making Trp residues more accessible to the solvent. Increasing the pH to alkaline (up to 10) caused only a minor bathochromic or hypsochromic shift of the emission spectra for all cruciferins, indicating that little conformational alteration is occurring in response to increasing pH.

**Figure 5.9** Tryptophan fluorescence of native cruciferins and denaturation induced by 6 M urea and 6 M guanidine-HCl. (A) Fluorescence emission spectra, (B) peak area, and (C) emission maximum ($\lambda_{\text{max}}$). All spectra were recorded at an excitation wavelength of 280 nm. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

### 5.4.7 Surface Hydrophobicity ($S_0$)

Among the cruciferins, heterohexameric WT and homohexameric CRUB gave similar $S_0$ values, indicating comparable prevalence of ANS anion binding hydrophobic areas exposed to the solvent environment. The CRUA homohexamer showed the highest surface hydrophobicity with $S_0 = 525.3$ (Table 5.2). The lowest $S_0$ value among the homohexamers was for CRUC (282.1), which may be related to the lower number of solvent exposed hydrophobic residues (57 in CRUA, 64 in CRUB, and 31 in CRUC) on the IA face (Table 5.2). In addition, the net negative charge on the CRUC IA face as observed by its $\zeta$-potential ($-42$ mV at pH 7.4) (Figure 5.5) may repulse the anionic ANS molecular probe resulting in further inhibition of ANS binding (Haskard & Li-Chan, 1998).
**Figure 5.10** Intrinsic fluorescence of cruciferin as a function of pH. (A) Emission spectra, (B) fluorescence peak area, (C) maximum wavelength of emission. All spectra were recorded at an excitation wavelength of 280 nm. *P < 0.05, **P < 0.01, ***P < 0.001 (ANOVA). Error bars, ± SD (n = 3).
When CRUA and CRUB $S_0$ values are compared, the $\zeta$-potential at pH 7.4 and number of solvent exposed hydrophobic residues together explain the higher value for CRUA than for CRUB, indicating that the electrostatic charge on the molecule’s surface may be responsible for the difference in the observed $S_0$ values.

Increasing the temperature of the protein solution beyond 70 °C caused an increase in cruciferin surface hydrophobicity (Figure 5.11A), with a more substantial change for CRUA and WT than for CRUB and CRUC. For CRUC, an increase in temperature to 90 °C is not expected to cause a high degree of unfolding or reveal more ANS binding sites (Figure 5.11A) because of the observed thermal stability of this protein (see next section of this study).

### 5.4.8 Thermal Transitions and Stability

The denaturation temperature of the cruciferin variants ranged from 100.2 to 113.5 °C. The CRUC homohexamer exhibited the most thermo-stable structure having a 10 to 13 degree higher $T_m$ value (113.5 °C) than that of WT or the other two homohexamers (Figure 5.11B, Table 5.3). The proteins in the $CRU^-$ line had a $T_m$ value of 100.6 °C. The thermal stability of a protein in aqueous system is the net free-energy change ($\Delta G_{N\rightarrow U}$) due to hydrophobic interactions ($\Delta G_{H\Phi}$) and the difference in conformational entropy ($\Delta G_{\text{conf}}$) upon heat denaturation ($\Delta G_{N\rightarrow U} = \Delta G_{H\Phi} + \Delta G_{\text{conf}}$) (Damodaran, 2008). During hexamer formation, cruciferin trimers interact through hydrophobic residues (Tandang-Silvas et al., 2010). In the CRUC homohexamer, the trimer IE face contains HVR-I and HVR-II, which together comprise the highest number of hydrophobic residues among the cruciferin variants.

The strong hydrophobic interaction between the two trimers supported by the hydrophobic residues of HVR-I and HVR-II may contribute to the high $T_m$ value. Decrease in protein stability due to sulfhydryl (–SH)–disulfide (S–S) interchange during thermal treatment has been reported (Damodaran, 2008; Yamagishi et al., 1983). As discussed earlier, CRUA is more prone to SH–SS interchange than CRUB in which access to Cys residues is sterically hindered by Thr277 and Met278 (Figure 5.3). CRUC has no free –SH group in the vicinity to exchange (Chapter 3). All these may contribute to the observed differences in $T_m$ values among the three cruciferin homomeric variants which were in the increasing order of CRUA < CRUB < CRUC.
Figure 5.11 Thermal properties of cruciferins. (A) Thermal unfolding of cruciferin hexamers as monitored by ANS fluorescence changes. (B) Heat capacity ($C_p$) versus temperature for 10% (w/v) cruciferin in 10 mM phosphate buffer at pH 7.4 ($N_2$ flow of 50 mL min$^{-1}$ with a heating rate of 10 °C min$^{-1}$). Values assigned to the major peaks indicate the denaturation temperature of the hexamer. Orange arrowhead shows the possible denaturation peak for trimeric form of the triple-knockout (CRU−). Inset: DSC parameters at protein denaturation. The dotted line indicates the sigmoidal baseline used for determination of enthalpy ($\Delta H_{\text{exp}} T_m$) associated with change of heat capacity ($\Delta C_p N \rightarrow U$) in transition from the native (N) to unfolded (U) states during protein denaturation. $T_m$, denaturation temperature; $\Delta T_{1/2}$, width at half peak height; $C_p N$, heat capacity of the native state, and; $C_p U$, heat capacity of the unfolded state.

Protein thermal denaturation is a highly cooperative process (unfolding of one part of the protein structure destabilizes the other parts) and the sharpness of the thermal transition peak (width at half peak height, $\Delta T_{1/2}$) is a measure of cooperativeness; a low $\Delta T_{1/2}$ value indicates a highly cooperative unfolding process (Privalov & Khechinashvili, 1974). The $\Delta T_{1/2}$ obtained for CRUA, CRUB, and CRUC homohexamers were 5.7, 6.0, and 5.0 °C, respectively, suggesting a more cooperative denaturation process than WT ($\Delta T_{1/2} = 6.9$ °C) (Table 5.3). The strong interactions that may occur between the IE face of CRUC trimers, as discussed previously, may have contributed to the sharpest peak ($\Delta T_{1/2}$, 5.0 °C) observed for this homohexamer and is indicative of a highly cooperative thermal denaturation process.
Table 5.3 Values obtained from differential scanning calorimetry (DSC) for thermal properties of Arabidopsis cruciferins\textsuperscript{a}

<table>
<thead>
<tr>
<th>Protein</th>
<th>$T_m$ (°C)</th>
<th>$\Delta T_{1/2}$ (°C)</th>
<th>$\Delta C_p \text{N}\rightarrow\text{U}$ (kJ K\textsuperscript{-1} mol\textsuperscript{-1})</th>
<th>$\Delta S (T_m)$ (kJ K\textsuperscript{-1} mol\textsuperscript{-1})</th>
<th>$\Delta H_{\text{exp}} (T_m)$ (kJ mol\textsuperscript{-1})</th>
<th>$\Delta H_{\text{vH}} (T_m)$ (kJ mol\textsuperscript{-1})</th>
<th>$\Delta H_{\text{exp}} (T_m)/\Delta H_{\text{vH}} (T_m)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>100.2\textsuperscript{c}</td>
<td>6.9\textsuperscript{a}</td>
<td>0.49\textsuperscript{a}</td>
<td>0.44\textsuperscript{b}</td>
<td>166.7\textsuperscript{b}</td>
<td>158.6\textsuperscript{ab}</td>
<td>0.96</td>
</tr>
<tr>
<td>CRUA</td>
<td>100.2\textsuperscript{c}</td>
<td>5.7\textsuperscript{c}</td>
<td>0.51\textsuperscript{a}</td>
<td>0.37\textsuperscript{c}</td>
<td>137.2\textsuperscript{c}</td>
<td>139.9\textsuperscript{b}</td>
<td>1.03</td>
</tr>
<tr>
<td>CRUB</td>
<td>102.1\textsuperscript{bc}</td>
<td>6.0\textsuperscript{bc}</td>
<td>0.46\textsuperscript{a}</td>
<td>0.32\textsuperscript{d}</td>
<td>123.3\textsuperscript{d}</td>
<td>128.3\textsuperscript{b}</td>
<td>1.04</td>
</tr>
<tr>
<td>CRUC</td>
<td>113.5\textsuperscript{a}</td>
<td>5.0\textsuperscript{d}</td>
<td>0.36\textsuperscript{ab}</td>
<td>0.47\textsuperscript{a}</td>
<td>181.6\textsuperscript{a}</td>
<td>189.2\textsuperscript{a}</td>
<td>1.04</td>
</tr>
<tr>
<td>CRU-</td>
<td>100.6\textsuperscript{c}</td>
<td>6.3\textsuperscript{b}</td>
<td>0.22\textsuperscript{b}</td>
<td>0.26\textsuperscript{e}</td>
<td>102.8\textsuperscript{e}</td>
<td>16.0\textsuperscript{e}</td>
<td>0.16</td>
</tr>
<tr>
<td>SEM</td>
<td>1.4</td>
<td>0.2</td>
<td>0.04</td>
<td>0.02</td>
<td>7.7</td>
<td>16.3</td>
<td>-</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Means followed by the same letter within a column do not differ significantly ($P < 0.05$). SEM, standard error of the mean. Assays were conducted in triplicate. Denaturation temperature, $T_m$; width at half peak height, $\Delta T_{1/2}$; heat capacity change, $\Delta C_p \text{N}\rightarrow\text{U}$; entropy change, $\Delta S (T_m)$; $\Delta H_{\text{exp}} (T_m)$, experimental enthalpy change; $\Delta H_{\text{vH}} (T_m)$, van’t Hoff enthalpy upon unfolding.
The difference between heat capacities of the native and unfolded states (ΔC_p N→U value) for thermal denaturation of the homohexamers was positive and in the range of 0.36–0.51 kJ K^{-1}mol^{-1} and did not differ significantly (P > 0.05) from the WT heterohexamer (Table 5.3). Enthalpy (ΔH) at the transition temperature (T_m) of cruciferin unfolding was an endothermic process (ΔH_{Tm} > 0) and could be calculated from the area beneath the denaturation peak. According to the enthalpy and entropy calculations (Table 5.3), ΔH_{Tm} values for cruciferins decreased in the order of CRUC > WT > CRUA > CRUB with an opposite pattern being observed for the entropy change (ΔS_{Tm}). Positive values for both enthalpy and entropy changes similar to that observed here generally represent a high degree of hydrophobic interactions in the molecule (Tian et al., 2004). The highest ΔH_{exp(Tm)} value (181.6 kJ mol^{-1}) for the CRUC homohexamer further confirms the compact nature of this molecule, which may occur due to strong hydrophobic forces between two trimers leading to low conformational flexibility. The degree of disordered conformation for the CRUC homohexamer was high when the protein underwent thermal unfolding (ΔS_{Tm}, 0.47 kJ K^{-1}mol^{-1}) compared to the other cruciferin variants (Table 5.3). The unfolding and consequent exposure of the long HVR-I region of CRUC homohexamer to the solvent upon thermal denaturation could be attributed to its high disordering property.

The van’t Hoff enthalpy value [ΔH_vH (T_m), Table 5.3] can be used to describe the transition mode or cooperative unfolding of cruciferins. When a protein undergoes a two-state transition from its native state to an unfolded state (i.e., N→U), the ratio of ΔH_vH (T_m) to ΔH_{exp} (T_m) is unity and a deviation from unity can be caused by either the presence of intermediate molecules or oligomerization of the native or unfolded state (Makhatadze, 1998). Calculated values of ΔH_vH (T_m)/ΔH_{exp} (T_m) for the WT heterohexamer and homohexameric cruciferins were equal to ~1 (Table 5.3); therefore, it can be described as a two-state model, N_{12}↔12U, considering the number of monomers is 12 (α- and β-chains as separate units) for 11S cruciferin (Koshiyama et al., 1981). A two-state model for thermal unfolding of WT cruciferin has been reported and suggests that each monomer in cruciferin acts as a cooperative unit in the denaturation process (Tandang et al., 2004).
5.4.9 Contribution of Each Subunit Type to WT Cruciferin

The present study and previous MS analysis of Wan et al., 2007 support the notion that CRUC is the predominant cruciferin protomer species in *A. thaliana* cv. Columbia. The CRUC homohexamer isolated from the CRUabC (or cruAcrubBCRUC) mutant line had distinct structural and physico-chemical properties compared to the WT heterohexamer or CRUA or CRUB. Although it was expected that the distinct properties of CRUC protomers may dominate in the WT heterohexamer, interestingly this was not observed. Therefore, to elicit distinct surface charge, intrinsic hydrophobicity, surface hydrophobicity, structural stability and thermal denaturation properties, CRUC has to be in its homohexameric composition. The study also indicated that the CRUC protomers interact to form a very stable homohexamer assembly with a compact structure. The physico-chemical properties of CRUA and CRUB homohexamers deviated from CRUC homohexamer in many ways and structural differences must have contributed to this. The divergent properties of the CRUC homohexamer observed here may be moderated when interacting with CRUA and CRUB protomers in WT cruciferin.

5.5 Conclusions

Studying the details of the structure/conformational and physico-chemical properties of homohexameric cruciferin obtained from mutant Arabidopsis lines improved the understanding of the contribution of CRUA, CRUB, and CRUC to these properties, and also to validate the predictions of properties made for these molecules based on *in silico* homology model analysis (Chapter 3). The present study confirmed that the different structural characteristics of the CRUC protomer compared to CRUA and CRUB were reflected in the physico-chemical properties of the respective hexamer and trimer. This study also points out that the secondary structure components of these protomers are only slightly affected by this genetic variation, but that the tertiary and quaternary structure characteristics such as observed through zeta potential, thermal stability, structure stability, intrinsic fluorescence, and surface hydrophobicity, which are directly related to the functions of cruciferin molecule, are very much affected.

5.6 Connections to Subsequent Chapters

The bioinformatics analysis and biochemical characterization of cruciferin the CRUA, CRUB, and CRUC homohexamers and the WT heterohexamer showed that they exhibit distinctive
structural/conformational properties (Chapter 3–5). The next two chapters describe selected functional properties of these three homohexameric cruciferins and compare them with the WT heterohexamer. The properties studied are directly relate to the techno-functionalities of proteins. The next chapter describes solubility, *in vitro* digestibility, and heat-induced gel formation of these cruciferins.
6. PROPERTIES OF CRUCIFERIN WITH MIXED OR IDENTICAL SUBUNIT COMPOSITIONS: SOLUBILITY, HEAT-INDUCED GELATION, AND PEPSIN SUSCEPTIBILITY

6.1 Abstract
This study examined the solubility, heat-induced gel formation ability, and pepsin susceptibility of cruciferin with mixed or identical subunits (CRUA, CRUB, or CRUC). All cruciferin species exhibited minimal solubility at pH 4.0 and slight solubility between pH 7.5 and 8, which was not influenced by the presence of NaCl. Increasing temperature to 90 °C did not affect the solubility of CRUC, but affected the solubility of other cruciferins between pH 4.0 and 8.0. Changes of secondary structure of cruciferins in a range of pH suggested structural alterations of the hexameric assembly might occur at pH 2.0. Near neutral pH was suitable to form heat-induced gels of WT, CRUA, and CRUB species. Although pH 2.0 was suitable for CRUC to form a strong gel, the same pH inhibited CRUB gel formation. High salt content (0.5 M NaCl) did not change the gel formation ability of WT, CRUA, or CRUB, but affected the gel characteristics. All cruciferins were hydrolyzed by pepsin, but the CRUC was degraded at a slower rate than the others. Trimer surface characteristics and the orientation and interaction of the HVR-I region of the hexameric assembly could be main structural factors for these property differences. Although the CRUC subunit showed properties that deviated from other species, the influence of CRUA and CRUB subunits seemed more prominent in the mixed subunit hexamer of WT.
6.2 Introduction

Seed storage proteins (SSP) comprise a significant fraction of the food and feed protein supply. The nutritional and functional characteristics of SSP are the key factors that determine their utility as a protein source. Functional characteristics provided by proteins are important in many applications, including foods, pharmaceuticals, personal care products, and industrial polymers. Nutritional properties of proteins are of particular interest in food, primarily the provision of essential and non-essential amino acids and peptides that is directly related to digestibility. In addition, the immunogenicity of proteins that elicit allergenic reactions is an important consideration of a food protein source.

Heat-induced structure formation of globular proteins is fundamental in stabilizing the dispersed phase in which protein is soluble, and also in developing structure to modulate sensory perception related to texture, taste, stability, and appearance. Protein gels consist mostly of water with the solid phase provided by the protein network. Heat and other forms of energy cause the native structure of globular proteins to undergo conformational and structural changes that expose hydrophobic groups leading to aggregation and non-covalent association of partially denatured protein molecules. These in turn form either a fine-stranded network or an amorphuous particulate structure that settles as a soft solid or gel (Ross-Murphy, 1995). In addition, sulphydryl−disulphide exchanges drive covalent associations in forming protein gels (Clark, Kavanah, & Ross-Murphy, 2001).

The 11S globulin cruciferin is the predominant crucifer SSP. It assumes a hexameric quaternary structure in the mature seed and in WT lines consists of a mixture of different subunits or protomers. Therefore, structural, functional, and physico-chemical data obtained from WT cruciferin cannot be related to individual subunit types. Studies on glycinin, the 11S globulin of G. max, showed that the subunits A1bB2, A2B1a, A5A4B3, A3B4, and A1aB1b possess distinct structural and physico-chemical properties (Maruyama et al., 2004; Prak et al., 2005; Tezuka, Taira, Igarashi, Yagasaki, & Ono, 2000; Tezuka, Yagasaki, & Ono, 2004), which are responsive to manipulation through protein engineering (Adachi et al., 2004; Kim et al., 1990a; Tandang et al., 2005; Utsumi et al., 1993a). The current understanding of cruciferin subunit structure and function is very limited. The solubility of B. napus cruciferin is dependent on ionic strength (Salleh et al., 2002) and it is least soluble at pH values between 4 and 8 (Wanasundara et al., 2012). Improved physico-chemical and functional properties of cruciferin can be achieved by
protein structure engineering (Tandang et al., 2004a,b) or chemical modifications, such as phosphorylation (Schwenke et al., 2000), acylation and sulfamidation of aliphatic and aromatic groups (Gerbanowski et al., 1999). However, it is still not clear what chemical or structural modifications are necessary to improve specific functionalities of cruciferin.

Study 4 examined the solubility, heat-induced gel formation, and pepsin susceptibility of Arabidopsis heterohexameric cruciferin, as well as homohexameric forms composed only of CRUA, CRUB, or CRUC subunits. This information provided insight into the functional contribution of each cruciferin subunit and their relationship to the molecular structure. Because of the high degree of similarity between B. napus and Arabidopsis cruciferin subunits (Chapter 3), the knowledge generated in this study has direct application to canola/oilseed rape cruciferin structure–function properties.

6.3 Materials and Methods
6.3.1 Materials and Chemicals

*Arabidopsis thaliana* cv. Columbia seeds were obtained from mature WT as well as double-knockout lines of CRUAbc (or CRUAcruBcruC), CRUAbc (cruACRUBcruC), and CRUabC (cruAcruBCRUC) (Chapter 4). The development of the lines and isolation of cruciferin protein (>95 % pure), CRUABC, CRUA, CRUB and CRUC, is described in Chapter 5. Analytical grade solvents and reagents were used throughout the study. For confocal microscopy, Fast Green FCF (λex. 633 nm) (≥98% pure) was purchased from Sigma-Aldrich (St Louis, MO). Different buffer systems were used for different pH assays (pH 2.0, 10 mM phosphoric/citric acid–Na2HPO4; pH 4, 10 mM acetate; pH 6, 10 mM citrate buffer; pH 7.4, 10 mM sodium phosphate/Tris-HCl; pH 8, 10 mM Tris-HCl; pH 10.0 10 mM disodium tetraborate buffer). 0.02% NaN3 was added to all buffers.

6.3.2 CD and DSC

Far-UV CD spectra of the protein solutions at pH 2.0, 7.4, and 10.0 were obtained using a PiStar-180 spectrometer (Applied Photophysics Ltd., Leatherhead, UK) according to the method described in Chapter 5. The extent of thermal denaturation of the protein at pH 2.0 and 7.4 in the presence of NaCl (ionic strength 0.5, pH 7.4) was determined using a TA Q2000 differential scanning calorimeter (DSC) (TA Instruments, New Castle, DE, USA).
6.3.3 Solubility

Protein solubility as a function of pH (2–10), ionic strength, and temperature was determined using protein solutions (0.2 mg mL\(^{-1}\)) prepared at different ionic strengths (0.0, 0.2, and 0.5 using NaCl) and at different temperatures (23, 50, 70, and 90 °C) according to the method 46-23 of AACC (1995). The protein samples were then centrifuged at 15000g for 10 min and the concentration of protein in the supernatant was determined using a Pierce\textsuperscript{®} BCA protein assay kit (Thermo Scientific, Tockford, IL, USA). Solubility was expressed as the percentage of protein in the supernatant relative to the total protein content of the solution before centrifugation.

6.3.4 Heat-Induced Gel Formation and Small Deformation Rheology of Gels

Protein dispersions (5.0% w/v, pH 7.4 without salt) were heated and followed by dynamic rheological measurements, using an AR1000 controlled stress rheometer (TA Instruments, New Castle, DE, USA) as described by Sanjeewa et al. (2011). Approximately 300 µL of the sample was applied to the lower plate and a 200-micron gap was set between the lower and upper (parallel plate geometry, diameter 40 mm) plates. A thin layer of paraffin oil was applied to the edge of the parallel plate to prevent any moisture evaporation. The procedure to form heat-induced gels was as follows: conditioning for 2 min at 25 °C, heating from 25 to 95 °C at a rate of 1 °C min\(^{-1}\), holding at 95 °C for 10 min, cooling down to 25 °C at a rate of 1 °C min\(^{-1}\), and then holding for 15 min at 25 °C. Gelation was measured at a constant strain of 0.01, which was within the linear viscoelastic region, at an angular frequency of 1 Hz. The storage modulus (G'), loss modulus (G'\(^{\prime\prime}\)), and apparent viscosity (\(\eta\)) were monitored and the G'–G'\(^{\prime\prime}\) crossover was taken as the onset of gelation. The gel was kept for another 15 min at 25 °C on the stage and then subjected to an angular frequency sweep between 0.1–100 rad s\(^{-1}\) at a constant strain of 0.01. The G' and G'\(^{\prime\prime}\) were recorded as a function of the oscillation frequency. In another experiment, gels were subjected to a strain sweep at a constant frequency of 1 Hz. The linear viscoelastic region (LVR), critical strain (\(\tau_{c}\)), and breaking point were determined from the G'–strain profiles (Eleya & Gunasekaran, 2004). Acid-induced gel formation was followed using protein solutions at pH 2.0 and the salt effect was evaluated by adding 0.5 M NaCl to the same buffer at pH 7.4.
6.3.5 Microstructure Assessment

Protein solutions (1 mL, 5% w/v) were stained with 20 µL of 0.1% (w/v) Fast Green FCF fluorescent dye (for non-covalent labeling) and dispensed into a micro-well in a glass slide surrounded by a press-to-seal rubber gasket spacer (depth 0.5 mm and diameter 20 mm) and then covered with a glass coverslip. Application of a thin layer of nail polish around the spacer and the coverslip provided secondary sealing. The samples were heated in a water bath at 95 °C for 10 min to allow heat-induced gel formation and then cooled overnight at 4 °C. The microstructure of these formed gels was observed using a Zeiss LSM 710 confocal laser-scanning microscope (CLSM) (Microimaging GmbH, Germany) with a 63× (oil immersion, NA 1.4) or 20× (dry, NA 0.8) objective lens using Zen 2011 software (Carl Zeiss, Microimaging GmbH, Germany). The excitation of Fast Green FCF was at 633 nm using a He–Ne laser and the emission of Fast Green FCF was recorded between 638–700 nm. The pinhole was kept at 1 Airy Unit. Gel images were pseudo-colored in red after acquisition.

6.3.6 3D-Structure Modeling and Analysis

Homohexameric structure models were generated by homology modeling using Pru du amandin from Prunus dulcis (PDB ID 3EHK) as a hexameric template and properties were examined as described in Chapter 3.

6.3.7 Simulated Gastric Fluid (SGF) Degradation Kinetics

Degradation of cruciferin using simulated gastric fluid (SGF) was studied according to Astwood, Leach, & Fuchs (1996). The SGF [3.2 mg mL⁻¹ pepsin (2410 U mg⁻¹ solid), 35 mM NaCl, pH 2, 37 °C] was prepared according to the United States Pharmacopoeia (2000). Aliquots (100 µL) of 5 mg mL⁻¹ test protein were mixed with 10 µL SGF (E:P, 1:250 w/w) in separate vials. Protein hydrolysis was terminated after 0.5, 2, 5, 15, 30, 60, and 120 min by adding 35 µL of 0.2 M Na₂CO₃. Proteins from the hydrolyzed and control samples were separated by microfluidic LabChip electrophoresis to examine the amount and profile of the resultant polypeptides according to the procedure described in Chapter 5.
6.3.8 Statistical Analysis

All studies were carried out in triplicate. Data were analyzed by one-way analysis of variance (ANOVA) using the General Linear Model (GLM) procedure of SAS Version 9.1 (SAS Institute Inc. Cary, NC, USA). Mean separation was carried out by calculating the least significant difference (LSD) when the main effect was significant ($P < 0.05$).

6.4 Results and Discussion

The 11S cruciferin of Brassica seeds assumes a hexameric quaternary conformation that is stabilized by electrostatic interaction between two trimers. In Arabidopsis, a similar conformation can be expected and would be composed of three subunits, CRUA, CRUB, and CRUC (Chapter 3). The food or non-food applications of such globular proteins include use in dilute solutions (e.g., protein-rich beverages, enzyme preparations) to aggregated soft solids (e.g., soy tofu, protein based hydrogels). Water is the most common solvent and ionic strength, pH, and temperature are used to modify the solvent environment in many situations.

6.4.1 Change of Secondary Structure with pH

It is well known that 11S globulins dissociate into 7S trimers under acidic (~pH 2) conditions (Adachi et al., 2003; Plietz et al., 1983; Sokolova, Kealley, Hanley, Rekas, & Gilbert, 2010), whereas, oligomerization occurs at the extremely high pH to form large (soluble) aggregates (Jiang, Chen, & Xiong, 2009; Sokolova et al., 2010). The CD spectra of all cruciferins at pH 7.4 showed negative mean residue elipticity with a minimum value around 208 nm that is typical of an $\alpha$-helix conformation (Figure 6.1A-D). From the previous work (Chapter 5), it could be assumed that at pH 7.4 all these cruciferins have intact hexameric assembly. Although the $\beta$-sheet structure remained predominant, some degree of unfolding was evident at pH 2 for all cruciferins. As well, the observed increase in random structure components at pH 2 suggests possible dissociation of the hexameric assembly and exposure of hypervariable regions (HVRs) and extended loop regions (ELRs). At basic pH, the changes in mean residue elipticity indicated that possible conformational transitions had occurred, most likely due to alkali-induced dissociation. According to far UV-CD analysis, CRUC secondary structure changes could be detected only at acidic pH. The majority of the data obtained from CD suggested that all cruciferins failed to
undergo major structural changes at pH 2 and 10, but partial denaturation caused unfolding and changes of the quaternary structure.

**Figure 6.1** Far-UV CD spectra of cruciferin (WT = A; CRUA = B; CRUB = C; and CRUC = D) at pH 7.4, pH 2.0, and pH 10.0. Inserted bar graphs show the percentage values of calculated secondary structures. The minimum at around 208 nm was blue-shifted in acidic pH, whereas the minima were maintained in alkaline pH.

Using a glycinin with only the A3B4 subunit, Sokolova et al. (2010) also showed that acidic pH (2.0) initiates dissociation of the hexameric structure into trimers, whereas an alkaline pH (11.0) facilitates formation of large aggregates. The changes in fluorescence intensity also indicated that Trp residues were exposed as the pH moved towards acidic (Appendix Figure A5). Unchanged intrinsic fluorescence properties of WT and CRUC hexamers at pH 10.0 in the presence of NaCl, but not with CRUA and CRUB, indicated strong hydrophobic interactions within these protein assemblies rather than electrostatic associations as described for 11S globulin (Boxtel, van Beers, Koppelman, van den Broek, & Gruppen, 2006).
6.4.2 Solubility

6.4.2.1 The Effect of NaCl

An insoluble protein has limited applications, because solubility plays a vital role in determining functional properties, such as emulsifying, foaming, and heat-induced gelation abilities, as well as catalytic activities. The solubility pattern of hexameric cruciferins in a pH range of 2.0 to 10.0 exhibited a W-shape pH versus solubility profile with depressed solubility around pH 4.0 and 7.5 (Figure 6.2A–D), which is typical for cruciferin (Tandang et al., 2004a,b; Tandang et al., 2005) and Brassica species meal proteins (Wanasundara et al., 2012). In Chapter 3, I showed that the ζ-potential values of WT, CRUA, and CRUB achieved 0 mV at pH 3.8, 4.9, and 3.4, respectively, and these proteins showed no or the lowest solubility around pH 4.0. The solubility of CRUC was depressed around pH 7.5 and zero ζ-potential was observed at pH 6.5. The solubility of the CRUA homohexamer was much higher above pH 6 than that of the others, and the CRUB homohexamer also showed higher solubility above pH 8. At pH 2, all cruciferins were highly soluble. The presence of 0.5 M NaCl had little effect on solubility above pH 4.0 (Figure 6.2A–D). Similarly, very low solubility was reported for glycinin at acidic pH and high ionic strength conditions (µ = 0.5) (Salleh et al., 2002; Maruyama et al., 2004), which was also reported for cruciferin (Salleh et al., 2002; Tandang et al., 2004a,b; Tandang et al., 2005; Wanasundara et al., 2012). The effect of neutral salts, such as NaCl, on ion-protein interactions for 11S globulins (e.g., glycinin) is concentration dependent. At low NaCl concentrations, they occur through electrostatic interactions between charged groups, while at higher concentrations hydrophobic interactions predominate (Damodaran, 1988). When changes in secondary structure conformation of these cruciferins at different pHs are considered, the complete solubility observed at pH 2.0 suggests that hexamer disassembly is complete.

Using recombinant 7S globulins, Fukuda et al. (2008) showed that electrostatic surface potential is the dominant factor influencing globulin solubility in neutral and weakly basic pH solutions at low ionic strength. Although all cruciferins had highly negative ζ-potentials at pH 7.4, the comparatively lower value for CRUC and zero value at pH 6.5 were reflected in its low solubility at neutral pH. In addition to the charge localization on the cruciferin molecule’s surface, the acidic residues on the HVRs may affect solubility (Salleh et al., 2002; Tandang et al., 2004a,b) through their influence on electrostatic attraction or repulsion.
Figure 6.2 pH-dependent solubility profiles of WT heterohexameric cruciferin (A & E), CRUA (B & F), CRUB (C & G), and CRUC (D & H) homohexamers (0.2 mg mL\(^{-1}\)) under different NaCl concentrations (A–D) and at different temperatures (E–H).
When the number of hydrophobic residues found in the HVR-I, HVR-II and ELR is calculated, CRUC has a higher number (aromatic + aliphatic = 104; aromatic = 36) compared to CRUA (aromatic + aliphatic = 48; aromatic = 12) or CRUB (aromatic + aliphatic = 42; aromatic = 6) (Chapter 3). When the predicted electrostatic potential on the IA face (intradisulfide-containing face) of homomeric cruciferin at pH 7.4 is considered, CRUA shows well-distributed negative charge coverage, CRUB has a positive core area and negative charge distribution at the periphery, while CRUC is covered with negative potential with fewer positively charged residues near the side (Chapter 3). This suggests that at pH values near neutral or above, CRUC tends to aggregate through hydrophobic interactions facilitated through its HVR and ELR. Therefore, the differences in the hydrophobic and charged residues on the molecule and in the extended loops may be related to the solubility differences observed in CRUC compared to the CRUA or CRUB homohexamers. However, the order of ANS surface hydrophobicity at pH 7.4 without NaCl was CRUA > CRUB > CRUC (Chapter 5), which further suggests that fewer hydrophobic binding sites are available on CRUC than the other cruciferins.

6.4.2.2 The Effect of Temperature

The solubility of all cruciferins was unaffected at temperatures up to 70 °C, but at 90 °C, WT, CRUA, and CRUB exhibited very low solubility in the pH range of 4 to 8 (Figure 6.2E–H). Only a slight change in solubility was observed for CRUC at these temperatures and pH range, which may be due to a combination of high denaturation temperature \( T_m = 113.5 \) °C (Chapter 3) and conformational stability of the CRUC molecule in this pH range. The increase in ANS-binding ability at pH 7.4 (without NaCl) at 90 °C for WT (3 fold), CRUA (3 fold), and CRUB (1.8 fold), and the increase in intrinsic fluorescence intensity at 90 °C, suggest that hydrophobic interactions were dominant in the heat-denatured proteins which led to low solubility. The highly hydrophobic IE (interdisulfide containing) face of these cruciferins (WT, CRUA, and CRUB) (Chapter 3), may be exposed due to partial denaturation at 90 °C favoring protein–protein aggregation via hydrophobic interactions. This would promote insoluble aggregate formation observed as reduced solubility in the pH range of 4.0 to 8.0. The CRUC homohexamer has a thermo- and pH-stable structure and conformation and therefore behaves differently from other cruciferins. All the evidence suggests that the CRUC homohexamer deviates from CRUA and CRUB with respect to the effect of changes in pH, NaCl level, and temperature of the solvent,
therefore its contribution to the related properties in WT hexamer properties is expected.

6.4.3 Heat-Induced Gelation

6.4.3.1 Without NaCl

When cruciferin dispersions (5% w/w, pH 7.4) were heated from ambient temperature to 95 °C, the storage modulus ($G'$), which indicates elasticity or stiffness of the forming gel network, did not change until the temperature reached 80 °C (Figure 6.3A–C). The loss modulus ($G''$) that measures the viscous component of the developing 3D gel network rapidly increased during the cooling phase, which might be attributed to substantial H-bond and hydrophobic bond formation between protein strands. The increase in $G'$ and $G''$ during cooling is generally ascribed to gel structure strengthening by attractive forces, such as van der Waals interactions and H-bonding between protein strands (Arntfield, Murray, & Ismond, 1990; Huang, Catignani, Foegeding, & Swaisgood, 1994; Eleya & Gunasekaran, 2004). The unchanged moduli values for the CRUC homohexamer indicated that it has poor ability to form heat-induced gel networks (Figure 6.3D). The onset temperature of structure formation or gelation point was determined as the crossover point, where $G' = G''$ (Clark & Ross-Murphy, 1987) was 80.9, 89.9, and 72.1 °C for WT, CRUA, and CRUB cruciferin, respectively. The heterohexameric WT cruciferin formed the strongest gel ($G'_{WT} \approx 1200$ Pa) at the end of the heating and cooling periods (Figure 6.3A), whereas gel strength ($G'$) values for the CRUA and CRUB homohexamers were in the range of 300 to 380 Pa. The CRUC homohexamer did not form a gel upon cooling (Figure 6.3D). When a strain was applied to the gels (strain sweep test), the $G'$ and $G''$ moduli were independent of the strain amplitude in the linear viscoelastic region (LVR) (Figure 3E–G), and further increases in the strain caused strain-thinning of the gels, i.e., Type I behavior (Hyun, Kim, Ahn, & Lee, 2002).

Increasing the strain beyond the LVR released the intermolecular interactions that occurred during network formation due to microstructural anisotropy resulting from the deformation (Figure 6.3E–G). Above the critical strain ($\tau_c$), the strain amplitude at which the storage modulus begins to decrease by 5% from its maximum value, (Shih, Shih, Kim, Liu, & Aksay, 1990) the gel networks are destroyed as the protein strands disentangle and align with the flow field (Hyun et al., 2002). The strain amplitude at which the non-linear behavior began and abruptly decreased the $G'$ and $G''$ values is considered the breaking point of the gel network.
Figure 6.3 Dynamic oscillatory measurements of 5.0 % (w/v) dispersions of WT cruciferin (A and E), CRUA (B and F), CRUB (C and G), and CRUC (D and H) at pH 7.4. (A–D) show time-dependence of the elastic modulus ($G'$) and viscous modulus ($G''$) during heating, holding, and cooling cycles: (i) 2 min conditioning phase, (ii) heating phase (1 °C min$^{-1}$), (iii) 30 min holding phase, (iv) cooling phase (1 °C min$^{-1}$), (v) 15 min holding phase (frequency 1.0 rad s$^{-1}$; strain 1.0 %). Inset: A–D, CLSM images of gels. E–H show storage modulus–strain profiles for the same cruciferin gels (frequency 1.0 rad s$^{-1}$, temperature 25.0 °C). Inset: E–H, storage modulus–frequency sweep at 25 °C.
The $\tau_c$ and the breaking point were 5.8 and 101.0% strain for WT, 4.0 and 91.2% for CRUA, and 4.0 and 138.8% for CRUB, respectively (Figure 6.3E–G). The viscoelastic nature of the gel reflected in the frequency sweep indicated that both moduli were frequency-independent and that the elastic response ($G'$) dominated over the viscous flow ($G''$). This confirmed the transformation of the cruciferin dispersion into an elastic network structure after heating and cooling (Figure 6.3E–G inset). Of the cruciferins, CRUA and CRUB exhibited medium strength gels ($G'_{\text{CRUA}} = 389.2$ Pa and $G'_{\text{CRUB}} = 304.9$ Pa; $\tau_c = 4.0\%$ for both cruciferins) that were more viscoelastic (slight slope in frequency sweep) compared to WT (Figure 6.3B, 6.3C, 6.3F, and 6.3G). Investigation of gel microstructure using CLSM showed an opaque, nonhomogeneous, but well-structured network for WT cruciferin (Figure 6.3A). The gels of CRUA and CRUB were homogeneous, and the smooth structure formation curves during the cooling period, the formation of transparent gels (Figure 6.3B and 6.3C) and the micrographic images suggested that the gels were fine-stranded.

### 6.4.3.2 With NaCl

In the presence of 0.5 M NaCl, gel formation by the WT cruciferin was not affected ($G'_{\text{WT}} = 1196.0$ Pa and $\tau_{c,\text{WT}} = 5.0$); however, the CRUA and CRUB homohexamers showed weakened gelling properties ($G'_{\text{CRUA}} = 256.9$ Pa, $\tau_{c,\text{CRUA}} = 5.1\%$, $G'_{\text{CRUB}} = 283.6$ Pa, $\tau_{c,\text{CRUB}} = 4.0\%$) (Figure 6.4A–C). The breaking point of the WT cruciferin gel was higher (134.3% strain) in the presence of NaCl than without (101.0% strain). The gels of the CRUA and CRUB homohexamers showed lower breaking points (70.6% and 63.4%, respectively) in the presence of NaCl than that of the corresponding gels without NaCl (91.2% and 138.8%, respectively). Furthermore, a less homogenous gel structure was evident for CRUB. The soy 11S protein undergoes partial refolding during the cooling phase of the thermal gelation process when 0.5 M NaCl is present which weakens the gel (Damodaran, 1988). For all cruciferins, NaCl may have stabilized the structure as the $T_m$ value was increased by 3 to 6 °C and the denaturation enthalpy values were approximately doubled (Table 6.1).

The CRUC homohexamer showed weak gel formation when 0.5 M NaCl was added to the dispersion (Figure 6.4D), indicating that a critical number of cross-links had been formed to generate a gel network, which was not the case in the absence of NaCl.
**Figure 6.4** Dynamic oscillatory measurements of 5.0% (w/v) cruciferin dispersions measured at high ionic strength ($\mu = 0.5$, pH 7.4) for WT (A and E), CRUA (B and F), CRUB (C and G), and CRUC (D and H) cruciferins. In A–D the heating conditions were similar to Figure 6.3. Inset: A–D, CLSM images of gels. E–H show storage modulus strain profiles for the same cruciferin gels (frequency 1.0 rad s$^{-1}$, 25.0 °C). Inset: E–H, storage modulus frequency sweep at 25 °C ($\mu = 0.5$, pH 7.4). The CRUC homohexamer made a very weak heat-set gel.
6.4.3.3 At pH 2 without NaCl

The WT cruciferin formed a weak viscoelastic gel ($G''_{WT} = 5.57$ Pa and higher slope in angular frequency sweep) with a more aggregated and less homogenous microstructure (Figure 6.5A) at pH 2 than at neutral pH. In the strain–sweep test, this gel showed high LVR with $\tau_c$ of 7.7% strain with both moduli attaining a peak before the breaking point indicating strong strain overshoot; a Type I behavior (Hyun et al., 2002). Among the three homohexameric cruciferins, CRUA formed the strongest gel ($G'_{CRUA} = 3212$ Pa, $\tau_c$-CRUA = 6.4%, breaking point = 43.2% strain) at pH 2.0. No gel was formed by CRUB at this acidic pH. Interestingly, CRUC was able to form a network structure of medium strength ($G'_{CRUC} = 297$ Pa, $\tau_c$-CRUC = 4.0%, and breaking point = 202.4% strain) (Figure 6.5B–D). The observed $G' >> G''$ and frequency-independent storage moduli indicated that the gels of CRUA and CRUC were predominantly elastic or have solid-like material characteristics. The CRUA and CRUC gels gave loss tangent ($\tan \delta; G''/G'$, degree of lost energy to stored energy per cycle of deformation) values of 4.6 and 3.7, respectively. The denaturation temperatures of all four cruciferins were lower than the values at pH 7.4 (Table 6.1), indicating changes to the hexameric conformation brought about by acid-induced dissociation may help to form stable gels for CRUA and CRUC homohexamers.

Although the gelation point (89.8 °C) was not changed, CRUA formed a strong gel network that exhibited a weak strain overshoot (Figure 6.5F); a Type III behavior (Hyun et al., 2002). This type of overshoot arises from network junctions, which are subject to formation and destruction at the higher strain amplitudes (Sim, Ahn, & Lee, 2003), suggesting the presence of permanent cross-links, such as disulfide bridges, in the CRUA gel at acidic pH. The CRUA gel formed at pH 2.0 was very homogeneous (Figure 6.5B), whereas the CRUB homohexamer formed large protein aggregates which inhibited network formation (Figure 6.5B and G). At pH 2.0, CRUB particles had higher solubility (Figure 6.2C).
Figure 6.5 Dynamic oscillatory measurements of 5.0% (w/v) cruciferin dispersions measured at pH 2.0 for WT (A and E), CRUA (B and F), CRUB (C and G), and CRUC (D and H). In A–D the heating conditions were similar to Figure 6.3. Inset: A–D, CLSM images of gels. E–H show storage modulus strain profiles for same cruciferin gels (frequency 1.0 rad s⁻¹; temperature 25.0 °C). Inset: E–H, storage modulus–frequency sweep at 25 °C at pH 2.0 (µ = 0.0). No heat-set gel structure formation was observed for the CRUB homohexamer.
It is known that high solubility favors heat-induced gel network formation (Damodaran, 2008); however, CRUB particles did not interact to form a gel network (Figure 6.4C and G). In contrast to the observation at pH 7.4, CRUC formed a gel at pH 2.0 (Figure 6.5D) with a crossover point (\(G' \sim 14.5\) Pa and \(G'' \sim 4.7\) Pa) just after heating started and consequently the storage and loss moduli increased upon cooling. The strain-sweep test of the CRUC gel indicated a high elastic performance with \(\tau_c = 4.0\%\) and an extended strain thinning period with a very high breaking point (202.4% of strain). The microstructure of the CRUC gel at pH 2.0 was dense and similar to the WT and CRUB gels formed at pH 7.4 without NaCl.

The heat-induced gels formed from the homomeric cruciferins were different than that of the WT cruciferin containing mixed subunits. The most significant difference was the inability of the CRUC homohexamer to form a gel at neutral pH; however, its’ contribution to WT cruciferin gel formation was somehow negated by the other subunits. Initiation of thermal-induced gel formation of globulin proteins starts with the unfolding of the structure due to heat energy and then association of the unfolded protein to form a straight strand that extends to a 3D network (Nakamura et al., 1984). Most likely, the initial step in 11S hexameric protein gelation is dissociation of the hexameric conformation and generation of trimers which extend into strands upon interaction. At neutral pH, the CRUC homohexamer had a peak denaturation temperature of 113.5 °C, indicating that it has a more thermo-stable structure compared to the WT heterohexamer (100.2 °C) or the two other homomeric cruciferins (CRUA, 100.6 °C and CRUB, 102.7 °C) (Chapter 5).

During the heating phase, unfolding of protein structure facilitates formation of new covalent bonds, such as disulfide and/or disulfide interchanges, needed for new protein strands (Nakamura et al., 1984). Previously, I predicted that the free –SH groups of CRUA (Cys293) and CRUB (Cys140 and Cys280) can readily undergo inter-chain disulfide bond formation, but the free –SH group in CRUC (Cys406) has limited accessibility (Chapter 3). Although I also predicted low thermal stability of CRUC based on the cavity size, length of the loops, and the number of Pro residues in the HVRs based on homology structures, experimental results revealed a remarkably thermo-stable structure that does not unfold easily.
### Table 6.1 The effect of pH and ionic strength on thermal characteristics of cruciferins

<table>
<thead>
<tr>
<th>Protein</th>
<th>At pH 7.4, $\mu = 0.0$</th>
<th>At pH 7.4, $\mu = 0.5$</th>
<th>At pH 2.0, $\mu = 0.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_m$ ($^\circ$C)</td>
<td>$\Delta H$ (kJ mol$^{-1}$)</td>
<td>$T_m$ ($^\circ$C)</td>
</tr>
<tr>
<td>WT</td>
<td>100.2 $^c$</td>
<td>166.7 $^b$</td>
<td>104.1 $^c$</td>
</tr>
<tr>
<td>CRUA</td>
<td>100.2 $^{bc}$</td>
<td>137.2 $^c$</td>
<td>105.8 $^b$</td>
</tr>
<tr>
<td>CRUB</td>
<td>102.1 $^b$</td>
<td>123.3 $^d$</td>
<td>105.0 $^{bc}$</td>
</tr>
<tr>
<td>CRUC</td>
<td>113.5 $^a$</td>
<td>181.6 $^a$</td>
<td>116.0 $^a$</td>
</tr>
<tr>
<td>SEM</td>
<td>1.7</td>
<td>7.7</td>
<td>1.5</td>
</tr>
</tbody>
</table>

$^a$Means of triplicate analysis followed by the same letter within a column do not differ significantly ($P < 0.05$). SEM, standard error of the mean. $T_m$, denaturation temperature and, $\Delta H_{T_m}$, enthalpy change.
To examine this discrepancy further, I modeled the CRUC homohexameric structure using Pru
du amandin from *Prunus dulcis* (PDB ID 3EHK) as a hexameric template. This is the only model
that has complete coordinates for hexamer assembly in the PDB data bank for an 11S protein. In
these models, the orientation of HVR-I (9, 7, and 60 residues for CRUA, CRUB, and CRUC,
respectively) and HVR-II (18, 11, and 16 residues for CRUA, CRUB, and CRUC, respectively)
to the solvent accessible surface is more descriptive (Figure 6.6).

The HVR-I region of each trimer projects out from the accessible molecule surface, positions
about 20–40 Å away from each other in CRUA and CRUB homohexamer, thus there is less
likely to involve H-bonding, electrostatic, and hydrophobic interactions. In CRUC, the HVR-I is
longer and extended providing possibilities for additional interactions between the two trimers for
a compact hexamer structure. These additional interactions may not allow easy dissociation of
trimers leading to the high thermal stability and reduced heat-induced gel formation ability at
neutral pH. The HVR-II is located on the IE face and is less accessible to the solvent. This
analysis supports the hypothesis that trimer formation is necessary for the gelation of hexameric
globular proteins.

The present study also supported the notion that hexameric 11S globulins dissociate and form
trimers in extreme acidic conditions, such as pH 2.0. In the hexameric form, the distinct
properties of the trimer IE face (compared to IA face) are concealed to sustain the hexameric
conformation (Chapter 3; Adachi et al., 2004; Fukuda et al., 2008; Tandang-Silvas et al., 2010).
The IE face of the CRUC trimer has the highest number of hydrophobic patches compared to the
CRUA and CRUB homotrimers (Chapter 3), which may have facilitated cross-linking and
consequent aggregation upon hexamer dissociation. The increased intrinsic fluorescence of the
CRUC homohexamer with increasing ionic strength and change of pH to 2.0 (Appendix Figure
A5) also indicated that hydrophobic residues were exposed on the IE face. The low denaturation
temperature of trimers at pH 2.0 and possible neutralization of the negative side chains of Asp
and Glu in the electrostatic potential map of both IE and IA faces (Chapter 3) may also provide
an opportunity for the hydrophobic groups of the CRUC trimer interfaces to associate.
Interaction of the same Asp and Glu residues on the IA face of CRUC with other negatively charged residues at pH 7.4 may result in strong IA–IA repulsion (Chapter 3) and the presence of NaCl (0.5 M) may allow some degree of charge screening and reinforcement of hydrophobic interactions that may explain the weak gel network formation by CRUC hexamers.

Heat-induced gelation of cruciferin hexamers at pH 7.4 without salt followed the order: WT >> CRUA > CRUB, no gel formation for CRUC. This order in the presence of salt (0.5 M NaCl) was changed to: WT >> CRUB > CRUA >> CRUC. At acidic pH (2.0), the gelation order of the cruciferin trimer was in the order of CRUA >>> CRUC >>> WT, no gelation in CRUB. It is evident that thermal- and acid-induced denaturation and the properties of the gel network for each cruciferin type are different and sensitive to the environmental conditions and co-solvents. Although the solubility and gelation properties of CRUC were different from those of CRUA and CRUB, the influence of CRUC is somewhat neutralized in the mixed subunit WT cruciferin. When observations of the cruciferin behavior are related to an application of low pH and heat-induced gelation, it can be deduced that the contribution of the CRUB subunit was significant to the weak gel formation by WT cruciferin under acidic conditions; however, the inhibitory effect of CRUC on the gel formation was less pronounced in the WT heterohexamer at neutral pH condition.
6.4.4 Susceptibility to Pepsin

Pepsin-catalyzed digestion of cruciferin proteins showed the disappearance of the disulfide-bonded α–S–S–β subunit (49–57 kDa) was observed at 0.5 min for both CRUA and CRUB homohexamers much faster than those for WT (5 min) and CRUC (15 min) hexamers (Figure 6.7A). The decrease in intensity of these bands upon addition of SGF (no pepsin, pH 2.0, 35 mM NaCl, 37 °C) and the appearance of products around 20–37 kDa indicated that dissociation of the S–S bonded subunits starts in the acidic environment (Figure 6.7B). The products of acid dissociation also degraded at different rates upon addition of pepsin: 15 min for WT, 5 min for CRUA, 2 min for CRUB, and 30 min for CRUC. The pepsin-catalyzed degradation fragments of WT (9.2, 9.5, 15.0, 16.1, 16.9, and 17.6 kDa), CRUA (9.2, 9.6, 10.8 kDa), CRUB (9.5, 11.1, 15.0, and 17.0 kDa), and CRUC (9.5, 9.9, 11.4, 15.5, 17.6 kDa) (the main fragments, concentration >1000 ng µL⁻¹, are in bold) were resistant to further digestion even after 2 h of exposure to SGF. According to the predicted proteolytic sites for CRUA, CRUB, and CRUC (Chapter 3), pepsin cleavage sites were found in the both HVR-I and HVR-II of CRUA and CRUB, but only in the HVR-II region of CRUC. Therefore, hexamer ring opening and accessibility of HVR-II which is located deep inside the IE face are critical for pepsin-catalyzed hydrolysis of CRUC. This means the activity of pepsin on CRUC will be observed only after hexamer dissociation under the acidic conditions provided by SGF, and therefore at a slower rate than CRUA or CRUB. Large molecular weight fragments (15.5 and 17.6 kDa) were more abundant among the polypeptides generated from CRUC compared to fragments generated from CRUA and CRUB (9.2 to 11.1 kDa) (Figure 6.7B). This observation can also be linked to the larger number of pepsin cleavage sites found in the HVRs of CRUA and CRUB than in the HVR-II of CRUC. In WT cruciferin, the proteolytic fragments resulting from CRUA, CRUB, and CRUC subunits were evident (Figure 6.7) and the slight delay in digestion may be due to the slow degradation of the CRUC subunit in the heterohexamer. Pepsin-catalyzed degradation of proteins is important in determining the bioavailability of their component essential amino acids (Mandalari et al., 2008), as well as possible immunogenicity of un-degraded proteins (Astwood et al., 1996). This in vitro pepsinolysis study shows that the CRUC subunit eventually degrades and likely does not pose an allergenic concern as immunogenic SSP are generally resistant to pepsin degradation. Also, CRUA and CRUB are readily degraded by pepsin (Mills et al., 2002).
Figure 6.7 Pepsin susceptibility of cruciferins. (A) Degradation kinetics of cruciferin subunits (α–S–S–β). Inset panel shows the molecular bands corresponding to cruciferin subunits as digestion progressed to 120 min. (B) Polypeptide profile of WT, CRUA, CRUB, and CRUC degradation with simulated gastric fluid (SGF) containing pepsin over 2 h. Box shows the protein band corresponding to each intact subunit. Bands in the X region relate to pepsin degradation of cruciferins and bands of Y region are the product of degraded cruciferins.

6.5 Conclusions

The 11S hexameric structure of cruciferins dissociates into trimers at acidic conditions (pH 2.0), whereas oligomeric aggregates form in extreme alkaline conditions (pH 10.0). The heterogeneous WT cruciferin forms a strong heat-induced gel network at neutral pH (7.4, µ = 0.0), whereas CRUC is unable to form a gel due to its compact structure. The acid-induced unfolding of the hexamer into trimers favors heat-induced gel formation of CRUC, but this is not so for the CRUB hexamer. Solubility of cruciferin depends mainly on pH and the effect of NaCl is exerted mostly in acidic pH environments. The effect of co-solvents (H ions and NaCl) and temperature can vary with the structure and conformation stability of the cruciferin molecule. All the cruciferins are susceptible to pepsin cleavage and possess good digestibility characteristics.

6.6 Connection to Subsequent Chapter

In addition to investigation of gelation, solubility, and digestibility properties of cruciferin at the subunit level, the next study explains the assessment of cruciferins in oil-in-water emulsion formation and stabilization.
7. PROPERTIES OF CRUCIFERIN HOMOHEXAMERS ASSOCIATED WITH OIL-IN-WATER EMULSION FORMATION AND STABILIZATION

7.1 Abstract

Cruciferin composed of identical (CRUA, CRUB, and CRUC) or different (WT) subunits were investigated for their ability to generate and stabilize oil-in-water (O/W) emulsions. All the proteins (0.9%, at pH 7.4, \( \mu = 0 \)) except CRUC formed stable canola oil or triolein emulsions with a dispersed volume fraction of 22–23%. CRUC formed a stable emulsion at pH 2.0. A fine emulsion was formed by CRUB at pH 7.4 with droplet sizes of 6.8 and 8.6 \( \mu \text{m} \) in diameter for canola oil and triolein, respectively. Presence of 0.5 M NaCl reduced the level of adsorbed protein and protein load at the interface at pH 7.4, and generated less stable emulsions. Emulsions of CRUA and CRUB (pH 7.4, \( \mu = 0 \), canola oil or triolein) had higher stability than WT up to 15 d after formation. The low solubility, low surface hydrophobicity and compact structure of the CRUC protein may contribute to its inferior emulsifying properties; however, the emulsifying properties were improved by acidic pH-induced dissociation of the hexameric assembly. Heat treating of emulsions near to the denaturation temperatures enhanced the stability, most likely due to the improved rheological properties of the adsorbed proteins.
7.2 Introduction

Protein stabilized oil-in-water (O/W) emulsions are found in food (cream liqueurs, creamers whippable toppings, ice cream mixes) and non-food (cosmetics, pharmaceutics, personal hygiene, hair styling) products in dilute (liquid infant milk formula, homogenized milk, soy milk, vinaigrettes, vaccines) or concentrated (mayonnaises, moisturizing lotions) form. In addition, emulsion systems are found in industrial products, such as lubricants, inks, paints, varnishes, and pesticide carriers. These emulsions are usually formed and stabilized by either small molecular surfactants or macromolecular proteins. In particular, protein-stabilized O/W emulsions are widely found in manufactured foods. The globular proteins of animal (milk and egg proteins) or plant (seed storage proteins) origin are widely found as emulsifiers and stabilizers of O/W emulsions.

The O/W emulsion is a lyophobic colloid dispersion that holds dispersed fine oil droplets in a continuous aqueous phase. Compared to small molecular weight surfactants, proteins are less effective in reducing the surface tension at the O/W interface, but they are better emulsion stabilizers (Damodaran, 2005). When the fine O/W emulsion droplets are formed, a large area of fresh interface is rapidly created and protein molecules orient their hydrophilic and hydrophobic groups at the interface to reduce the interfacial tension. Although proteins contain hydrophilic and hydrophobic groups, their position in the interface, which is related to their folded conformation (tertiary or quaternary), is the key parameter governing the adsorption of protein molecules at the interface. A good emulsifying protein is capable of quickly adsorbing at the oil-water interface, decreasing interfacial tension, and forming a strong viscoelastic film via intermolecular interactions which can withstand mechanical and thermal perturbations (Damodaran, 1997). In addition, protein molecules form a macromolecular barrier at the oil-water interface to protect droplets from coalescing (Dickinson, 1994). The ability to form a cohesive film through protein-protein interactions differentiates protein molecules from low molecular weight surfactants used in emulsions. Once adsorbed at the interface, proteins undergo conformational changes that allow them to form loops which protrude into the bulk phase. These overlapping, protruding chains cause steric repulsion when they approach each other which helps to stabilize the emulsion droplets. The surface activity of proteins and their ability to stabilize emulsions are attributed to their structural properties and the ability of the protein to undergo conformational changes at the oil-water interface. Globular proteins are generally more capable of emulsifying
and stabilizing O/W emulsions due to their amphiphilic character (Dickinson, 2003). The instability of emulsions arises from different mechanisms, such as creaming, Ostwald ripening, flocculation, and coalescence (Damodaran, 2008).

The general understanding on the emulsifying properties of native globulins (e.g., glycinin and cruciferin) is that the structural properties, including compact and rigid molecular structure (i.e., low molecular flexibility), and the low number of exposed hydrophobic patches (Wagner & Guéguen, 1995; Damodaran, 2008) pose restrictions. The emulsifying properties of 11S globulins can be improved by changing environmental factors such as temperature, pH, and ionic strength (Kimura et al., 2008; Peng et al., 1984), modifying molecular components such as reducing S–S bonds (Wagner & Guéguen, 1999), or attaching hydrophobic groups such as acyl residues (Krause, Mothes, & Schwenke, 1996). The acidic chain of glycinin (11S globulin) exhibits better emulsifying and stabilizing abilities for O/W emulsions than the acidic-basic chains in the intact protein (Liu, Lee, & Damodaran, 1999). Modified glycinin obtained by deleting hydrophilic regions of the molecule showed enhanced emulsifying properties due to increased molecular surface hydrophobicity (Kim et al., 1990b). Similarly, addition of a hydrophilic extension region from β-conglycinin α’ type to the C-terminus of the proglycinin A1aB1b subunit (A1aB1bα’) caused significant improvement in both emulsifying activity and stability (Tandang et al., 2005).

Cruciferin, the main storage protein of Brassicaceae seeds, including mustard and canola, has an 11S hexameric globulin structure (α−β)₆ similar to soybean glycinin. Rapeseed cruciferin exhibits better emulsifying properties at pH 7.6 (μ = 0.08) than glycinin, but performed less well at pH 6.0 and 9.0 (μ = 0.5) (Salleh et al., 2002). Similarly to proglycinin A1aB1bα’, addition of an extension region to procruciferin (Cru2/3a) significantly improved the stability to form O/W emulsions compared to its wild type counterpart (Tandang et al., 2005). However, attempts to destabilize the structure by replacing a Cys group (C287T), deleting HVR II (ΔII), replacing a Cys and deleting HVR II (C287T/ΔII), or adding hydrophilic region IV from soybean glycinin A1aB1b and A3B4 subunits (ΔIV+A1aIV and ΔIV+A3IV) were not successful in improving procruciferin emulsifying properties (Tandang et al., 2004). The present study describes some key information on the emulsifying properties of CRUA, CRUB, and CRUC protein composed of identical protomers in comparison with WT cruciferin.
7.3 Materials and Methods

7.3.1 Materials and Chemicals

Mature *Arabidopsis thaliana* cv. Columbia seeds obtained from WT and cruciferin double-knockout lines of *CRUAbc*, *CRUaBc*, and *CRUabC* are described in Chapter 4. Refined, bleached, and deodorized canola oil was obtained from a local supermarket. Triolein [1,2,3-tri(*cis*-9-octadecenoyl)glycerol; >99% purity], Nile Red (λ<sub>ex</sub> 488 nm) (pure >98%) and Fast Green FCF (λ<sub>ex</sub> 633 nm) (pure >85%) were from Sigma-Aldrich (St Louis, MO). In all studies, analytical grade solvents and regents were used. A pH of 2 and 10 was achieved using 10 mM phosphoric/citric acid-Na₂HPO₄ buffer and 10 mM sodium phosphate/Tris-HCl buffer, respectively. All buffers contained 0.02% NaN₃.

7.3.2 Purification of Homohexameric Cruciferins with Different Subunit Compositions

Isolation and purification of heterohexameric cruciferin from seeds of WT and homohexameric CRUA, CRUB, and CRUC from seeds of double-knockout lines were according to Chapter 5. All proteins were >95% pure.

7.3.3 Fatty Acid Profile

Fatty acid composition of the canola oil was determined according to the AOAC Official Method 996.01 (1998). The triolein had a declared purity of >99% and was not analyzed for fatty acid composition.

7.3.4 Emulsion Preparation

Preparation of O/W emulsions was carried out as described by Dalgleish et al. (1999) with some modifications. The oil (canola or triolein) was mixed with the protein dispersion (0.5 mg mL⁻¹, pH 7.4 or pH 2.0, with or without NaCl) at a 1:6 (v:v) ratio of oil to protein (aqueous) solution in 5 mL glass tubes (Pyrex<sup>®</sup> rimless culture tubes, diameter 12 mm, height 75 mm, Sigma-Aldrich, St Louis, MO, USA). The protein solutions prepared with pH 2.0 or 7.4 buffer without NaCl were considered to have zero ionic strength (μ = 0.0) and the solutions containing 0.5 M NaCl were considered as μ = 0.5. The oil and aqueous protein solutions were first homogenized in an ice bath for 1 min with a PT 3100 Polytron homogenizer (30 × 10³ rpm, head diameter 5 mm, couple type B) (Kinematica AG, Littau, Switzerland) and then with an ultra-
sonic homogenizer with a standard horn (S-4000, Misonix Inc., Farmingdale, NY, USA, total energy input 1000 J) for 1 min with a 13 sec run followed by a 2 sec pause (4 times) to obtain O/W emulsions. To study the effect of temperature on emulsion stability, the capped glass tubes containing the emulsions were kept in a water bath maintained at 90 °C for 30 min and then transferred to an ice bath for rapid cooling and stored at ambient temperature.

7.3.5 Estimation of Oil Volume Fraction (ϕ)

The method described by Pearce & Kinsella (1978) with some modifications was used to determine the oil volume fraction (ϕ) of the emulsions. An aliquot (0.5 mL) of freshly prepared emulsion was immediately transferred into a pre-weighed sample pan and dried (120 °C) to achieve a constant weight (~1 h). The ϕ was calculated using droplet phase mass fraction according to equation 7.1 and reported as a percentage.

\[ \phi = \frac{m_{\text{pan}+\text{dm}}-m_{\text{pan}}-C_p(m_{\text{pan}+\text{em}}-m_{\text{pan}+\text{dm}})}{m_{\text{pan}+\text{dm}}-m_{\text{pan}}+(m_{\text{pan}+\text{em}}-m_{\text{pan}+\text{dm}})/(1+C_p)(\rho_p/\rho_o)-C_p} \]  

(7.1)

where, \( m_{\text{pan}} \) is the mass of pan, \( m_{\text{pan}+\text{em}} \) the mass of pan plus emulsion, \( m_{\text{pan}+\text{dm}} \) the mass of pan plus dry matter; \( C_p \) the initial concentration of protein (mg mL\(^{-1}\)); \( \rho_o \) the density of oil (g mL\(^{-1}\)), \( \rho_p \) the density of protein solution (g mL\(^{-1}\)).

7.3.6 Emulsifying Activity Index (EAI)

The EAI of cruciferin-stabilized emulsions was determined from turbidity (\( T \)) measurements. An aliquot (30 \( \mu \)L) of emulsion was taken from the bottom of the tube immediately after preparation and diluted in 0.1% (w/v) SDS at a ratio of 1:17 or 1:33 (v:v). After gentle mixing, the absorbance of the dispersed emulsion was measured at 500 nm using a UV-visible spectrophotometer (Ultrospec 1000, Pharmacia LKB Biochrom, Cambridge, U.K.). The turbidity of an emulsion is given by \( T = 2.303A/l \), where, \( A \) is the absorbance and the \( l \) is the optical path length. The EAI (m\(^2\) g\(^{-1}\)) was determined according to method of Pearce & Kinsella (1978) and modified by Cameron et al. (1991);

\[ EAI = 2T \times df/c_p(1 - \phi) \]  

(7.2)
where, $df$ is the dilution factor, $C_p$ is the initial protein concentration (g mL$^{-1}$), and $\phi$ is the oil volume fraction.

### 7.3.7 Droplet Size Analysis

The size of O/W emulsion droplets was measured by laser light scattering of diluted (6×10$^4$) samples in the same buffer in which the emulsion was prepared. A Mastersizer 2000 (Malvern Instruments, Malvern, UK, measuring range 0.02–2000 µm) equipped with a hydro attachment (Version 5.54 software, Malvern Instruments, Inc.) was employed for taking measurements. The refractive index of water as 1.330 and canola/triolein oil as 1.471, and the scattering theory of Mie (1908) was used in calculating droplet sizes. Average droplet sizes were determined (7.3) in terms of the volume mean diameter,

$$d_{43} = \frac{\sum_{i} n_i \cdot d_i^4}{\sum_{i} n_i \cdot d_i^3}$$  \hspace{1cm} (7.3)

where, $n_i$ is the number of the droplets of diameter $d_i$. The $d_{43}$ is particularly sensitive to slight changes to the size of the emulsion droplets in events such as flocculation and coalescence. Measurements were taken for at least for three samples.

### 7.3.8 Interface Protein Load ($\Gamma$) and Percentage of Protein Adsorption

An aliquot (0.5 mL) of freshly prepared emulsion was centrifuged at 14000g for 25 min at 23°C. A fraction of aqueous (serum) phase was removed and its protein content was determined using the bicinchoninic acid (BCA) method (Smith et al., 1985). The amount of protein adsorbed onto the unit area of the emulsion O/W interface ($\Gamma$, mg m$^{-2}$) was calculated as;

$$\Gamma = (C_{\text{initial}} - C_{\text{serum}}) \times d_{32} / 6\phi$$  \hspace{1cm} (7.4)

where, $C_{\text{initial}}$ is the initial cruciferin protein concentration (kg m$^{-3}$) in the aqueous phase before emulsification, $C_{\text{serum}}$ is the non-adsorbed protein concentration (kg m$^{-3}$) in the aqueous phase after emulsion preparation, $d_{32}$ is the surface weighted mean diameter as determined by light scattering using a Mastersizer 2000 (Malvern Instruments, Malvern, U.K.), and $\phi$ is the oil volume fraction.
The percentage of protein adsorbed onto the O/W interface \( (F_{ads}) \) was calculated as a fraction of initial protein content of the protein solution using equation 7.5,

\[
F_{ads} = C_{initial} - C_{serum}/C_i \times 100\%
\]

(7.5)

where, \( C_{initial} \) and \( C_{serum} \) are the same as in the above equation.

### 7.3.9 Microstructure Assessment

An aliquot (500 µL) of freshly prepared O/W emulsion was carefully mixed with the dye (5 µL, 0.01%, w/v, prepared in the same buffer as the emulsion) without interrupting the emulsion droplets. The oil and protein phases were stained with Nile Red and Fast Green FCF fluorescence dyes, respectively. The emulsion samples were then dispensed into the well of a glass slide prepared with a press-to-seal rubber gasket spacer (0.5 mm deep and 20 mm diameter) and covered with a glass coverslip. To study the effect of temperature, a secondary seal was applied as a thin layer of nail polish around the spacer and the coverslip. When the seal was dry, the slide was covered with aluminum foil prior to heating at 90 °C for 30 min in a water bath.

Microscopic images were acquired 1 h after preparation of the emulsion using a confocal laser-scanning microscope (CLSM) (LSM 710 Carl Zeiss, Microimaging GmbH, Germany) with a 63× (oil immersion, NA 1.4) or 20× (dry, NA 0.8) objective lens and equipped with Zen 2011 software (Carl Zeiss, Microimaging GmbH, Germany). Imaging of Nile Red and Fast Green FCF was done as separate tracks to avoid any interference. Nile Red was excited at 488 nm using an Argon laser and emission was detected between 539 and 638 nm. Fast Green FCF was excited at 633 nm using a He–Ne laser and emission was detected between 638 and 700 nm. The pinhole was kept at 1 Airy Unit. After acquisition, the protein phase in the CLSM images was pseudo-colored red, while the oil phase was pseudo-colored green. The images were deconvoluted using Autodeblur 2.2.2 (Bitplane AG, Zurich, Switzerland) and visualized using Imaris 7.3.1 (Bitplane AG, Zurich, Switzerland). Dim features were made visible by adjusting the gamma component as necessary.
7.3.10 Storage Stability of Emulsions

Freshly prepared O/W emulsions were transferred to 5 mL rimless, Pyrex® glass culture tubes (diameter 9 mm, height 75 mm, Sigma-Aldrich, St Louis, MO, USA), covered with parafilm to prevent evaporation, and stored at ambient temperature without any disturbance. The surface weighted mean diameter ($d_{32}$) of the emulsion was measured 1 h and 20 h after preparation as described in section 7.3.7. The specific surface area (SSA, area per unit mass of emulsion) was calculated from the laser diffraction measurement of $d_{32}$ using equation 7.6,

$$SSA = 6 \sum \frac{V_i}{d_i} / \rho \sum V_i = 6 / \rho_o \cdot d_{32}$$  \hspace{1cm} (7.6)

where, $V_i$ is the relative volume in class $i$ with a mean diameter of $d_i$ and $\rho_o$ is the density of the oil. The extent of stability was determined from the ratio of the observed SSA from the first and second measurement of the emulsion stored at ambient temperature. The general appearance and the stability of emulsion were assessed by taking photographs (Canon EOS DIGITAL REBEL XS, EF 24–105mm f/4L IS USM) after 1 h and each day up to 15 days after preparation.

7.3.11 Intrinsic Fluorescence, Surface Hydrophobicity, and Structure Modeling

The intrinsic fluorescence and surface hydrophobicity ($S_0$, 1-anilino-8-napthalensulfonate as the probe) of the cruciferin protein in solution (at pH 2.0, 7.4, or 10.0) was determined as described in Chapter 5. The three dimensional structures of homohexameric CRUA, CRUB, and CRUC molecules were obtained using homology modeling as described in Chapter 3.

7.3.12 Statistical Analysis

All studies were carried out in triplicate. The data were analyzed by one-way analysis of variance (ANOVA) using the General Linear Model (GLM) procedure of SAS Version 9.1 (SAS Institute Inc. Cary, NC, USA). Mean separation was carried out by calculating the least significant difference (LSD) when the main effect was significant ($P < 0.05$).
7.4 Results and Discussion

7.4.1 Emulsion Formation

The emulsion preparations of this study consisted of 14.3% (v/v) oil and 0.9% (w/v) protein (1:6 v:v ratio of oil to aqueous phase). The emulsions had an oil volume fraction (ϕ) ranging from 21 to 23% with an average value of 22% for both canola oil and triolein (Table 7.1), suggesting that the cruciferin subunit composition, pH, ionic strength, or oil phase differences had little effect on the amount of oil that was emulsified by this protein content.

Among the O/W emulsions found in food, an oil volume faction of 3–4% in milk, 10–30% in cream, 10–15% in coffee whitener and 10–30% in imitation cream (to be aerated) has been reported (Dickinsen & McClements, 1996). High oil volume fractions are reported for products such as mayonnaise (65%), while for salad dressings a minimum of 30% is needed (Pechak, Schwimmer, Borwankar, & Ford, 2003). A large volume fraction of the dispersed phase means that the emulsion droplets are closely packed. The concentration of droplets in an emulsion is an important consideration when determining the cost, appearance, texture, and stability (McClements, 2005) of an emulsion. Processing, such as the extent of homogenization (i.e., energy input), has direct relationship with ϕ (Pearce & Kinsella, 1978). Therefore, ϕ is a quality control measure for the emulsion preparation technique. The adsorbed protein content at the interface (Table 7.1), which is deduced from protein remaining in the continuous (serum) phase, indicated that except for the CRUC homohexamer (18.7% in canola oil and 33.1% in triolein), the other three cruciferins showed high levels (54.9–66.5% in canola oil and 54.6–61.8% in triolein) of adsorbed protein at pH 7.4. The CRUC homohexamer had comparatively lower solubility compared to the CRUA and CRUB homohexamers at this pH (Chapter 6), and may contribute its poor adsorption at the interface. Similarly, the very low interfacial protein load of the CRUC homohexamer at the interface (0.17 mg m⁻² for canola oil and 0.25 mg m⁻² for triolein) is likely responsible for it having the lowest EAI (12.2 and 19.9 for canola and triolein, respectively) among the cruciferins. When other cruciferins are compared, the performance (Γ and EAI of canola and triolein emulsions) of the CRUB homohexamer was comparable to the WT heterohexamer, but they had higher values than CRUA for the parameters determined (Table 7.1).
Table 7.1 Oil volume fraction ($\phi$), EAI, Interface protein load ($\Gamma$), and percentage of protein adsorbed for cruciferin emulsions generated from canola oil and triolein.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Protein</th>
<th>Canola oil-in-water emulsion</th>
<th>Triolein-in-water emulsion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\phi$</td>
<td>% adsorbed</td>
</tr>
<tr>
<td>pH = 7.4, $\mu =$ 0.0</td>
<td>WT</td>
<td>0.22</td>
<td>57.8$^b$</td>
</tr>
<tr>
<td></td>
<td>CRUA</td>
<td>0.21</td>
<td>54.9$^b$</td>
</tr>
<tr>
<td></td>
<td>CRUB</td>
<td>0.22</td>
<td>66.5$^a$</td>
</tr>
<tr>
<td></td>
<td>CRUC</td>
<td>0.22</td>
<td>18.7$^c$</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>-</td>
<td>6.92</td>
</tr>
<tr>
<td>pH = 7.4, $\mu =$ 0.5</td>
<td>WT</td>
<td>0.22</td>
<td>19.3$^a$</td>
</tr>
<tr>
<td></td>
<td>CRUA</td>
<td>0.23</td>
<td>17.4$^a$</td>
</tr>
<tr>
<td></td>
<td>CRUB</td>
<td>0.22</td>
<td>18.1$^a$</td>
</tr>
<tr>
<td></td>
<td>CRUC</td>
<td>0.23</td>
<td>0.2$^b$</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>-</td>
<td>3.00</td>
</tr>
<tr>
<td>pH = 2.0, $\mu =$ 0.0</td>
<td>WT</td>
<td>0.22</td>
<td>65.9$^b$</td>
</tr>
<tr>
<td></td>
<td>CRUA</td>
<td>0.22</td>
<td>61.4$^b$</td>
</tr>
<tr>
<td></td>
<td>CRUB</td>
<td>0.22</td>
<td>60.8$^b$</td>
</tr>
<tr>
<td></td>
<td>CRUC</td>
<td>0.23</td>
<td>78.0$^a$</td>
</tr>
<tr>
<td></td>
<td>SEM</td>
<td>-</td>
<td>2.66</td>
</tr>
</tbody>
</table>

$^a$Means of triplicate analysis followed by the same letter within a column do not differ significantly ($P < 0.05$). SEM, standard error of the mean.
Changing the discontinuous (droplet) phase fatty acid composition from a mixture of $18:1\omega_9\text{cis}\Delta_9$, $18:2\omega_6\text{cis}\Delta_9\text{cis}\Delta_12$, and $18:3\omega_3\text{cis}\Delta_9\text{cis}\Delta_12\text{cis}\Delta_15$ (61.1%, 19.3%, and 9.8%, respectively) in canola oil (Appendix Figure A6) to a primarily $18:1$ (99.9% in triolein) did not change the activity of these proteins in O/W emulsion formation. It is possibly that the slight changes in oil phase polarity due to the presence of unsaturated fatty acids (~30%) are the reason for the lower interfacial protein load and EAI values of canola oil emulsions (Table 7.1). The droplets in the CRUC emulsions were larger (18.9 and 33.3 μm for canola oil and triolein, respectively; Table 7.2) and examination of the emulsion microstructure (Figure 7.1) showed that most of these particles were non-spherical protein aggregates. The droplets in the CRUB emulsions had the smallest volume mean diameter (6.8 and 8.6 for canola oil and triolein, respectively) at pH 7.4, followed by CRUA and WT. According to the size of the droplets, micro-scale emulsions (1–10 μm) had been formed by these proteins.

The inclusion of 0.5 M NaCl lowered the values of adsorbed protein, interfacial protein load, and subsequently EAI of the emulsions (Table 7.1). These changes indicated that the protein layer around the oil droplets was thinner and fewer oil droplets were emulsified. The CRUA emulsion in the presence of NaCl had much larger droplets than the emulsion without NaCl (Table 7.2). Other proteins also showed affected average particle size of droplets in the presence of NaCl (Tables 7.2). Increasing the ionic strength up to 0.5 M with NaCl did not decrease solubility of proteins drastically (Chapter 6), but possible charge screening of protein molecules could have affected the interfacial activity of these proteins. The confocal images of emulsions showed less crowding of the droplets in the presence of NaCl (Figure 7.2A, B, and C) compared to emulsions formed without NaCl (Figure 7.1A, B, and C). The emulsion droplets of CRUB were more uniform than those of the other cruciferins as indicated by the narrow droplet volume mean diameter peak and seen in the micrograph. The CRUC homohexamer exhibits low solubility at pH 7.4 as this pH is close to its pI (6.45) and since NaCl did not improve solubility emulsion formation was negatively affected (Figure 7.2D). The presence of 0.5 M NaCl showed no effect on CRUA and CRUC intrinsic fluorescence (Appendix Figure A7) or molecule surface hydrophobicity (Table 7.3), which further indicated that NaCl does not facilitate emulsion formation. Changing the pH from 7.4 to 2.0 caused a significant increase ($P > 0.05$) in the absorbed protein content at the interface and in turn changed the interfacial protein load and EAI for WT, CRUA, and CRUC emulsions (Table 7.1) of both canola and triolein (Figure 7.3).
Table 7.2 Average particle size (µm) of emulsion droplets of canola oil or triolein formed with cruciferin$^a$.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Canola-in-water emulsion</th>
<th></th>
<th></th>
<th>Triolein-in-water emulsion</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pH = 7.4, µ = 0.0</td>
<td>pH = 7.4, µ = 0.5</td>
<td>pH = 7.4, µ = 0.0</td>
<td>pH = 7.4, µ = 0.0</td>
<td>pH = 7.4, µ = 0.5</td>
<td>pH = 7.4, µ = 0.0</td>
</tr>
<tr>
<td>WT</td>
<td>14.0 $^b$</td>
<td>14.0 $^b$</td>
<td>61.0 $^{bc}$</td>
<td>38.9 $^a$</td>
<td>18.1 $^b$</td>
<td>24.3 $^b$</td>
</tr>
<tr>
<td>CRUA</td>
<td>8.8 $^c$</td>
<td>6.6 $^c$</td>
<td>49.0 $^c$</td>
<td>36.8 $^b$</td>
<td>11.9 $^c$</td>
<td>12.3 $^c$</td>
</tr>
<tr>
<td>CRUB</td>
<td>6.8 $^c$</td>
<td>11.3 $^{bc}$</td>
<td>65.3 $^b$</td>
<td>32.7 $^c$</td>
<td>8.6 $^d$</td>
<td>5.6 $^d$</td>
</tr>
<tr>
<td>CRUC</td>
<td>18.9 $^a$</td>
<td>63.0 $^a$</td>
<td>111.7 $^a$</td>
<td>28.0 $^d$</td>
<td>33.2 $^a$</td>
<td>149.5 $^a$</td>
</tr>
<tr>
<td>SEM</td>
<td>1.68</td>
<td>4.88</td>
<td>0.53</td>
<td>0.70</td>
<td>5.79</td>
<td>7.77</td>
</tr>
</tbody>
</table>

$^a$Means of triplicate analysis followed by the same letter within a column do not differ significantly ($P < 0.05$). SEM, standard error of the mean. Volume mean diameter, $d_{43}$ was determined from laser diffraction Malvern mastersizer.
Figure 7.1 Microstructure and stability of O/W emulsions of canola oil (A–D and A’–D’) and triolein (E–H and E’–H’) stabilized by WT (A, A’, E, E’), CRUA (B, B’, F, F’), CRUB (C, C’, G, G’) or CRUC (D, D’, H, H’) cruciferin at pH 7.4. CLSM images were taken 1 h after emulsion formation, protein is in red and oil is in green (A–H, scale = 20 µm). Droplet size distribution of each emulsion is superimposed on the corresponding micrograph. The magnified images corresponding with the micrograph sections are indicated with an asterisk/s (scale = 2 µm). White arrows indicate possible protein adsorption, yellow arrows indicate possible coalescence of droplets, purple arrows indicate possible protein aggregates, and white arrowheads indicate trapped oil droplets in protein aggregates. Photographs of O/W emulsions up to 15 days (A’–H’) show phase separation and stability. Blue arrow on D’ indicates the oiling-off of CRUC emulsion.
All cruciferin emulsions, except the CRUC triolein emulsion, had larger emulsion droplet volume mean diameters at pH 2.0 than at pH 7.4 ($\mu = 0$) (Table 7.2). The microstructure of the CRUC emulsion droplets showed highly dense protein adsorption/packing at the oil-water interface with spherical oil droplets (Figure 7.3A–D). Low pH values, such as 2.0, brought about distinguishable changes in solubility and heat-induced network formation for CRUC with evidence for disassembly of its hexameric structure (Chapter 6). In forming emulsions at pH 2.0, a similar dissociation of hexamer resulting in cruciferin trimers may have facilitated coating of oil droplets to generate CRUC emulsions.

Some commercial emulsion products are subjected to thermal fluctuations during processing, such as pasteurization, sterilization, and cooking; therefore, it is also important to study thermal influence on emulsion properties. To assess the effect of temperature on the O/W properties, cruciferin stabilized canola-in-water and triolein-in-water emulsions were held at 90 °C for 30 min and then cooled to room temperature rapidly before analysis of the particle size, microstructure, and stability (Table 7.2 and Figure 7.4). In general, the $d_{43}$ size of canola-in-water emulsion was greater than that of the triolein-in-water emulsion for all cruciferins (Table 7.2). The confocal microscopic images revealed that the canola-water interface has very densely adsorbed layers of protein molecules for WT, CRUA, and CRUB homohexamers, which explains the high laser scattering particle sizes for canola-in-water emulsion. The CRUC homohexamer in the canola-in-water emulsion had the highest particle size (111.7 µm, Table 7.2), reflecting large oil drops resulted from the oiling-off process (Figure 7.4). Thus, heat treatment did not improve the poor ability of CRUC in canola-in-water emulsion formation. On the other hand, particle size and microstructure showed that cruciferin covered a thin layer (~2 µm) around less polar triolein droplets (Table 7.2 and Figure 7.4). ANS binding hydrophobicity confirms the WT, CRUA, and CRUB cruciferin molecules had higher degree of denaturation compared to CRUC (Table 7.3).

### 7.4.2 Emulsion Stability

The stability of cruciferin-stabilized O/W emulsions was assessed as the ratio between specific surface area (SSA) of the oil particles 1 h and 20 h after preparation. As the emulsion ages, formation of intermolecular bonds between molecules at the interface leads to complex events.
Figure 7.2 Effect of NaCl at pH 7.4 on microstructure and stability of O/W emulsions of canola oil (A–D and A’–D’) and triolein (E–H and E’–H’) stabilized by WT (A, A’, E, E’), CRUA (B, B’, F, F’), CRUB (C, C’, G, G’) or CRUC (D, D’, H, H’) cruciferin. CLSM images were taken 1 h after emulsion formation, protein is in red and oil is in green (A–H, scale = 20 µm). Droplet size distribution of each emulsion is superimposed on the corresponding micrograph. The magnified images corresponding to the sections of micrographs are indicated with an asterisk (scale = 2 µm). White arrows indicate possible protein adsorption, yellow arrows indicate possible coalescence of droplets, purple arrows indicate possible protein aggregate formation, and white arrowheads indicate trapped oil droplets in protein aggregates. Photographs of O/W emulsions up to 15 days (A’–H’) show phase separation and stability.
Table 7.3 Predicted number of hydrophobic residues and the measured surface hydrophobicity of cruciferins.

<table>
<thead>
<tr>
<th></th>
<th>Number of hydrophobic residues</th>
<th>Surface hydrophobicity (S₀)</th>
<th>ANS-binding Fluorescence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IA face</td>
<td>On the side</td>
<td>pH 7.4, μ = 0.0</td>
</tr>
<tr>
<td>WT</td>
<td>-</td>
<td>-</td>
<td>446.4 b</td>
</tr>
<tr>
<td>CRUA</td>
<td>57</td>
<td>25</td>
<td>525.3 a</td>
</tr>
<tr>
<td>CRUB</td>
<td>64</td>
<td>32</td>
<td>444.5 b</td>
</tr>
<tr>
<td>CRUC</td>
<td>31</td>
<td>11</td>
<td>282.1 c</td>
</tr>
<tr>
<td>SEM</td>
<td>36.4</td>
<td>9.2</td>
<td>20.7</td>
</tr>
</tbody>
</table>

a Means of triplicate analysis followed by the same letter within a column do not differ significantly (P < 0.05). SEM, standard error of the mean.
b Study 1, Chapter 3.
c From ANS probe binding. d Study 3, Chapter 5.
d Ratio between fluorescence at 23 °C and the molecules heated 90 °C and cooled to 23 °C.
Figure 7.3 Effect of pH (2.0) on the microstructure and stability of O/W emulsions of canola oil (A–D and A′–D′) and triolein (E–H and E′–H′) stabilized by WT (A, A′, E, E′), CRUA (B, B′, F, F′), CRUB (C, C′, G, G′) or CRUC (D, D′, H, H′) cruciferin. CLSM images were taken 1 h after emulsion formation, protein is in red and oil is in green (A–H, scale = 20 µm). Droplet size distribution of each emulsion is superimposed on the corresponding micrograph. The magnified images corresponding to the sections of micrograph are indicated with an asterisk (scale = 2 µm). White arrows indicate possible protein adsorption on oil droplets.
At neutral pH, both CRUA and CRUB homohexamers showed about 85% stability for canola and triolein emulsions after 20 h, while the WT protein emulsions showed much lower (~42%) stability (Figure 7.5). The stability of CRUC emulsions was poor, especially with canola oil. Microscopic observation of emulsion droplets up to 15 days after preparation showed fairly high stability in CRUB stabilized emulsions compared to the other cruciferins (Figure 7.1). Although good emulsion formation characteristics were observed for the CRUA homohexamer and WT heterohexamer, creaming and flocculation was clearly visible after the third day of storage (Figure 7.1, A', B', E', F'). In general, the $d_{43}$ size of canola-in-water emulsions was greater than that of the triolein-in-water emulsion for all cruciferins (Table 7.2). The presence of NaCl at 0.5 M affected the stability of CRUA and CRUB emulsions more so than that of the WT cruciferin emulsion. The effect of NaCl on the emulsion stability was evident after 20 h of the storage (Figure 7.2A'–H').

The emulsions formed at pH 2.0 were more stable than the emulsions at pH 7.4 for all cruciferins (Figure 7.5) and phase separation (or creaming) was evident only after 3 days of storage at ambient temperature (Figure 7.3A'–H'). The thick cream-like emulsion layer consisted of loosely packed droplets and mild agitation was sufficient to disperse the droplets. The density of canola oil (0.914–0.917 g cm$^{-3}$; Przybylski, 1999) and triolein (0.915 g cm$^{-3}$; Fong, Despres, Julien, & Angel, 1988) is fairly similar; therefore, the effect of oil phase density on preventing creaming is the same. Creaming in emulsions can be prevented by reducing droplet size, matching phase densities using weighting agents, or modifying continuous phase rheology by adding thickening agents (McClements, 2005).

A significant improvement in emulsion stability was evident in the heat-treated cruciferin canola oil and triolein emulsions (Figure 7.5). Emulsion stabilization could be due to the increased viscosity of the interfacial protein molecules caused by heat-induced partial unfolding (McClements, 2005) of the proteins in the adsorption layer (Tcholakova, Denkov, Ivanov, & Campbell, 2006). All the cruciferins have high thermal denaturation temperatures in the range of 100.2–113.5 °C (Chapter 5). At 90 °C that is below the $T_m$ of these proteins, <50% of unfolded molecules could be found which might enable changes in the interfacial rheology and elasticity of molecules leading towards stable interfaces (Roth, Murray, & Dickinson, 2000; Tcholakova et al., 2006).
**Figure 7.4** Effect of heating (90 °C) on microstructure and stability of O/W emulsions of canola oil (A–D and A’–D’) and triolein (E–H and E’–H’) stabilized by WT (A, A’, E, E’), CRUA (B, B’, F, F’), CRUB (C, C’, G, G’) or CRUC (D, D’, H, H’) cruciferin. CLSM images were taken 1 h after emulsion formation, protein is in red and oil is in green (A–H, scale = 20 µm). Droplet size distribution of each emulsion is superimposed on the corresponding micrograph. The magnified images corresponding to the micrograph sections are in white boxes (scale = 2 µm). White arrows indicate possible protein adsorption and heat induced protein gel formation around oil droplets.
Figure 7.5 Emulsion stability after 20 h of storage. (A) canola-in-water emulsion, (B) triolein-in-water emulsion under different experimental conditions. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (ANOVA). Error bars are ± SD (n = 3).

It is clear from these results that the homohexameric form of CRUC at neutral pH is not as good an emulsifier as CRUA or CRUB. However, changing the pH to acidic was beneficial in generating and stabilizing O/W emulsions, most likely due to partial unfolding of molecules causing more hydrophobic sites to be exposed and facilitating oil droplet surface adsorption.

All the cruciferins showed 4.0 to 5.7 fold increase in surface hydrophobicity ($S_0$) when the pH was reduced from 7.4 to 2.0 (Table 7.3). Low pH induced structure unfolding was evident from the change in surface hydrophobicity at pH 2.0, but this was not associated with an increase in ionic strength at pH 7.4. The CRUC homotrimer has fewer hydrophobic residues on the solvent exposed IA face according to in silico homology model-based prediction; therefore, the possibility of ANS binding to hydrophobic sites than with the other cruciferins (Chapter 3). In the soybean A5A4B3, A3B4, A1aB1b, A1bB2, and A2B1a subunits, Maruyama and coworkers (2004) reported that the length of hypervariable region (HVR) has a direct relationship to their emulsifying properties, i.e., the longer the length, the better the emulsifying ability of the subunits. The length of the HVR-I + HVR-II + ELR regions of the cruciferins are in the order of CRUC (95) > CRUA (47) > CRUB (38) which does not match with the observed order of emulsifying ability or EAI (CRUB > CRUA > CRUC at pH 7.4 for triolein or canola oil) and the order changes to CRUC > CRUB > CRUA (canola) or CRUB > CRUC > CRUA (triolein) when the pH is changed to 2.0. In the homology-based structure models of cruciferin hexamers, the
HVR-II protrudes outward from the molecule and the extended loop region (ELR) is located on the IA face, whereas HVR-I located on IE face. The presence of large number of Gly and Gln residues in the HVRs of the IA face (Chapter 3) may impose polar/hydrophilic properties and could lower adsorption affinity of hexamers to the oil-water interface. Therefore, not only is the length of the HRV important, but also its location and composition.

When the solubility of the oil phase is considered, triolein may be less soluble, in consequence may be less polar than canola oil because of its low degree of unsaturation (Massey, Gotto, & Powness, 1982; Narayanan & Storch, 1996). It can be expected that a molecule with fewer surface hydrophobic residues, such as homohexameric CRUC, would have less affinity for an oil phase with high polarity; therefore, with a low polar oil phase such as triolein comparatively better emulsification would be expected. Previously, tryptophan residue environment and thermal property analysis showed that the CRUC homohexamer is a more compact and stable molecule than CRUA, CRUB, or WT cruciferin (Chapter 5). Molecules like β-lactoglobulin show a higher degree of conformation rearrangement in a low polarity oil phase than a high polarity oil leading to higher emulsion stability (Zhai et al., 2011). However, the restrictions to conformational change of CRUC upon adsorption to the low polarity surface do not appear to improve its emulsifying properties. In the hexameric cruciferin, the most likely conformation change will be associated with dissociation into the trimer, which involves destabilization of associated hydrophobic interactions (Tandang-Silvas et al., 2010), which could happen at the oil-water interface.

Subsequent unfolding of trimers or the resulting protomers may allow formation of a dense protein layer around the oil droplets and stabilize them. The lower number of hydrophobic residues on the solvent exposed IA face and the HVRs, and the molecular compactness of CRUC homohexamer (Figure 7.6; Table 7.3), allow fewer opportunities to undergo surface adsorbed conformational changes. The geometric orientation of the hexameric globular molecules that adsorb on the oil-water interface is important (Fan, Wang, Zhu, Robillard, & Mark, 2006). If the molecule is oriented at the oil-water interface so the trimer interface interacts, dissociation of trimer assembly could be expected due to strong intermolecular hydrophobic interactions through IE faces. None of the cruciferin species could undergo conformational changes in their structures in 0.5 M NaCl as indicated by no change in intrinsic fluorescence (Figure 7.1A) and surface hydrophobicity (Table 7.3).
Therefore, no improvement in emulsifying properties could be expected. In addition to screening of electrostatic surface charge of protein, salt molecules may promote aggregation of protein-protein molecules and further weakening of O/W emulsion stability.

Figure 7.7 proposes possible scenarios for surface adsorption of hexameric cruciferin at the O/W interface. In the emulsification process, proteins undergo surface adsorption, surface denaturation/conformational change, and cohesive film formation around the oil droplet...
Extrinsic factors, such as low pH, may support cruciferin hexamer dissociation (Figure 7.7; Scheme 1), but this is not so for increasing ionic strength. Once the conformational change occurs, the trimeric cruciferin can undergo cohesive protein film formation through protein-protein interaction as I predicted in the Chapter 3.

**Figure 7.7** Schematic representation of the proposed mechanism for cruciferin adsorption at oil-water interface. Scheme 1: pH-induced dissociation of the cruciferin hexamers and surface adsorption of dissociated cruciferin hexamers at the O/W interface from IE face due to its high hydrophobic residues. Scheme 2: Surface adsorption of native cruciferin via IA face (a) and periphery of the heaxamer (b).

Two possible orientations of native cruciferin hexamer at the O/W interface exist (Scheme 2); either through IA face (Figure 7.7a) or direct contact with the trimer interacting interface (Figure 7.7b). In each orientation different hydrophobic residues are available to interact with hydrophobic oil interface. The surface adsorption may favour quaternary conformation change by
dis-assembling hexameric structure into trimers, but much less probability of this occurrence because there may an energy barrier that should overcome this event to occur. The position of the IA-face would not facilitate trimer formation at the O/W interface due to less number of hydrophobic residues and steric constrains (Figure 7.7a). Depending on the degree of hydrophobicity at trimer interacting interface, hexamer dissociation is possible at O/W interface in periphery of the hexamer (Figure 7.7b). The flexibility of the hexameric assembly may be a key factor that affects surface denaturation; therefore, tight packing and less hydrophobicity of the cruciferin IA face and trimer interacting interface may contribute to poor emulsifying properties of the CRUC homohexamer.

7.5 Conclusions

The subunit composition of cruciferin is an intrinsic factor that governs the emulsifying properties of cruciferin. CRUA and CRUB cruciferins had fairly similar properties with respect to the formation and stabilization of O/W interfaces. Homohexameric cruciferin composed of CRUC subunits showed tight association and a low number hydrophobic patches on the solvent exposed surface, which was unfavourable for O/W interface formation and stabilization. Therefore, conditions that favour dissociation of CRUC hexameric assembly, such as low pH, were necessary for optimum surface adsorption and conformational change at the surface, and cohesive film formation to stabilize the O/W interface.
8. GENERAL DISCUSSION

Five studies described so far provide information on structure, physico-chemical properties, selected technologically valuable functionalities, and their relationships of Arabidopsis cruciferin composed of single protomer types which is not available in scientific literature. This discussion considers all information together in order to understand the contribution of cruciferin subunit type to the structure and properties of the protein and also to the knowledge of 11S seed storage protein in general.

**Cruciferin Structure.** Homology-based protein structure modeling was used to build 3D structures of the Arabidopsis cruciferins CRUA, CRUB, and CRUC in the absence of experimentally established structure models. These proteins are composed of identical subunits expressed by one of the three (CRUA, CRUB, and CRUC) Arabidopsis 11S cruciferin genes and according to homology based modeling, it is possible for them to correctly assemble into homotrimmers and homohexamers. *In silico* assessment of these structure models allowed details on hydrophobic patches, charged residues, phosphorylation sites, specific enzyme cleavage sites, major cavities, and their characteristics that can be found in the protein surface to be obtained. Examination of the details from these proteins and reported for soybean homohexameric glycinin (11S), made it possible to predict physico-chemical properties of CRUA, CRUB, and CRUC such as thermal stability, solubility, interfacial (O/W) properties, and susceptibility to pepsin.

Arabidopsis plants expressing one cruciferin gene were obtained by disrupting two of the three cruciferin genes, therefore the resultant hexameric cruciferin consisted of a single protomer type assembled and deposited in plants. The synchrotron-powered FT-IR imaging of the amide-I band allowed protein distribution mapping of the physical components and organelles of the Arabidopsis seeds. Uninterrupted protein distribution was seen in the seed endosperm, and in the PSVs in the cotyledons where cruciferin is stored. The secondary structure features of the protein in the PSV of CRUA, CRUB, and CRUC containing seeds were not statistically different and confirmed that the core cupin structure is conserved in the homomeric proteins.
Using three chromatographic steps consisting of size exclusion and cation exchange, it was possible to obtain cruciferin of high purity (indicated by total N-based protein value and SDS-PAGE); this is the first report of chromatographically isolated 11S protein from Arabidopsis. Native-PAGE, SDS-PAGE, LabChip microfluidic electrophoresis, and DSC data confirmed the presence of single type of protein structure in this isolated 11S protein of CRUA, CRUB, and CRUC seeds, and it was a hexameric quaternary structure assembly. Both FT-IR and CD analysis of isolated cruciferins confirmed unaltered secondary structure composition consisting of 44–50% β-sheets and 7–9% α-helices that is similar to the reported value for B. napus cruciferin and other 11S seed storage proteins (Adachi et al., 2001; Adachi et al., 2003; Tandang-Silvas et al., 2010).

According to homology models, the uninterrupted 27 β-sheets, 6 α-helices and 3 3_10 helices identified in all three cruciferins further confirmed that the differences contributing to the variation in properties must be attributed to the divergent HVR 1 and II, and the amino acid residues that bring changes to hydrophobic and charged patches of the molecule. The genetic similarity and homology-based modeling of cruciferin CRUA, CRUB, and CRUC homotrimers showed differences on the IA and IE faces, loop regions (HVR-I, HVR-II, and ELR), and core regions. Of the three cruciferins, the longest HVR-I of CRUC (60 residues vs in 9 CRUA and 7 in CRUB) creates steric hindrance and interacting differences between trimer interfaces. This may cause structure dissociation and disintegration differences that eventually reflect in important physico-chemical properties, such as thermal stability and surface hydrophobicity.

Physico-chemical Properties. Using several well-defined structures of food protein molecules, Nakai and group (Nakai et al., 1986; Nakai & Li-Chan, 1988) showed that certain physico-chemical properties, such as surface charge, surface hydrophobicity, and thermal stability are related to the molecular structural features that can be derived from primary, secondary, and tertiary organization. The three cruciferins in solution showed differences in ζ-potential with changes in the medium pH, which is related to the observed charge differences of the surface residues. The ζ-potential at pH 7.4 was in the order of electronegativity as CRUC > CRUB >> CRUA confirming the homology structure based prediction of higher electronegativity on the solvent exposed IA face of CRUC than that of CRUA or CRUB. The highest surface hydrophobicity value ($S_0 = 525.3$) of CRUA homohexamer and the lowest value for CRUC ($S_0 = 282.1$) agreed with the predicted abundance of hydrophobic residues (CRUA 57, CRUB 64, and
CRUC 31) on the IA face. Based on the size of surface cavities, cavity openings, number of proline residues, and the length of HVRs, the trimeric form of CRUC homotrimer was predicted to have low thermal stability compared to CRUA and CRUB homotrimers; however, the high surface hydrophobicity on the IE face of the CRUC homotrimer might provide high conformational entropy which is an endothermic property. The thermal stability of the cruciferin molecule is related to the $T_m$ and was in the range of 100–113.5 °C and in the order of CRUA = CRUB < CRUC with ~13 °C higher denaturation temperature for CRUC homohexamer than other two homohexamers. When CRUC is concerned, the hydrophobic nature of the IE face has a precedence for thermostability over the parameters such as size of pockets, size of cavity opening, number of proline residues, or length of HVRs that have been considered (Chakravarty & Varadarajan, 2002; Fukuda et al., 2008; Tandang-Silvas et al., 2010) for predicting thermal stability.

**Functional Properties.** Solubility of a protein is a key factor in performing and differentiating functionality, such as emulsifying, foaming, and heat-induced network formation, which are important in utilization of a macromolecule. According to the homology structure models, the IE faces of all cruciferins possess a higher number of positively charged patches than negatively charged patches. The IE face is not exposed to solvent in the hexameric form, but should be exposed in the trimeric form, therefore hexamer dissociation may expose more positively charged residues to the solvent and create different solubility properties. The CRUC homohexamer showed lower solubility (~10%) at neutral pH (7.4) compared to CRUA or CRUB (50–60%). Cruciferin is the most abundant protein of Brassica seed meal, therefore it can be expected that isolated cruciferin may show a similar pattern of solubility as demonstrated for B. napus, B. juncea and Sinapis alba seed meal when the pH is changed from basic (pH = 10) to acidic (pH = 2) (Wanasundara et al., 2012). Arabidopsis cruciferins showed two minimum solubility points with one at ~pH 4.0 and the other at 7.4, coinciding with the “W” shape solubility profile with changing pH ($\mu = 0.0$) of other cruciferins. The 11S proteins of legumes, such as glycinin of soy (Prak et al., 2005; Salleh et al., 2002) and legumin of pea (Baniel, Caer, Colas, & Gueguen, 1992), showed “U” or “V” shape solubility profiles with one minimum solubility value in the acidic pH range between 3 and 4. It is not clear why cruciferin has a different solubility pattern, but may associate with its highly phosphorylated groups. Since the solubility pattern has a direct relationship with the net charge of the molecule under different pH conditions (Damodaran,
the \( \zeta \)-potential value of protein observed in this pH range was helpful in understanding solubility pattern. In addition to this, the molecular surface charge, surface hydrophobicity (\( S_0 \)), and hydrophobic residue environment (intrinsic fluorescence) should be considered in understanding the solubility of cruciferins. The sudden drop in solubility of cruciferins at pH 7.4 and increase in solubility at pH 6.0 paralleled the change in intrinsic fluorescence intensities. The insolubility of CRUA at pH 4.0 coincided with the near zero \( \zeta \)-potential. Although not completely insoluble, the lowest solubility value of CRUB and CRUC was also observed around same pH. If the hexameric structure of cruciferin remains upon decrease in pH to 4.0, protonation of \(-\text{COO}^-\) of Asp and Glu residues may neutralize surface charges leading to protein-protein association and aggregation causing reduced solubility and isoelectric precipitation. Considering the same model, at pH 2.0, low pH-induced dissociation of hexamers and repulsion of trimers with charged surfaces (positive charge according to \( \zeta \)-potential values) may have increased the solubility of all cruciferins up to 100%. All cruciferins showed fairly stable solubility performance with changing NaCl concentration for the pH above 5.0, which might be an advantage in some industrial applications. However, when the pH was below 4.0, cruciferin solubility was sensitive to 0.5 M NaCl, most likely due to counteraction of positive charges that led to loss of solubility. Protein charge neutralization is more effective in decreasing solubility, clearly this is the case because an increase in temperature up to 90 °C did not cause a drastic reduction in solubility at pH 2.0, although hexamer opening at pH 2.0 and structure disintegration due to heat was evident.

When heat is used to destabilize protein structure together with pH, cruciferin CRUA, and CRUB were sensitive to pHs above 4.0, which coincided with a drastic increase in surface hydrophobicity. Possible partial opening of the hexamer was evident based on CD, DSC, and intrinsic fluorescence studies, but the heat-induced unfolding and exposure of hidden hydrophobic patches may have caused protein aggregation and precipitation at pHs above 4.0 at 90 °C. The extremely high \( T_m \) (113.5 °C) of the CRUC homohexamer may not have allowed sufficient molecule unfolding leading to further protein aggregation and subsequent lowering of solubility at 90 °C. Although CRUC has a highly stable structure at temperatures above 100 °C, its low solubility around neutral pH may limit its applications where many of the functionalities are required. In general, cruciferins have the unique property of high thermal stability below pH 4.0, showing less aggregation and high solubility, which is unique among 11S seed storage proteins.
With the investigations carried out on functional properties that affect technological use of the cruciferins, explanations for the manifestation of their physico-chemical properties can be found when surface hydrophobicity, surface charge, and flexibility as well as the number of hydrophobic and hydrophilic residues in the HVRs are considered. Heat energy causes the native structure of 11S proteins to undergo conformational changes that expose hydrophobic groups leading to aggregation and non-covalent association of partially denatured protein molecules to form a network capable of entraining water molecules, which becomes a gel (Ross-Murphy, 1995). The order of gel network forming ability was in the order of CRUA > CRUB > CRUC, with no network formation at pH 7.4 for CRUC. The solubility values of cruciferin at pH 7.4 correlated with the order of heat-induced gel formation, CRUA > CRUB >> CRUC. In addition to the very low solubility at pH 7.4, highly thermo-stable structure ($T_m = 113.5 \, ^\circ C$) and very low surface hydrophobic patches ($S_0 = 282.1$) may have caused suppression of protein induced network development during CRUC heating. The sulfhydryl-disulfide exchanges during heating make covalent associations of denatured protein network that lead to gel strengthening (Clark et al., 2001). Therefore, the availability of free –SH groups and S–S bonds in the molecule relates to the bonds that can break and re-form during thermal denaturation (Shimada & Matsushita, 1980). In addition to the high denaturation temperature and low surface hydrophobicity, the absence of free –SH groups in the vicinity of inter- and intra-chain S–S bonds of CRUC that can’t undergo S–S bond interchange reaction may be another significant feature associated with its inability to form a gel network upon increasing thermal energy in the system. Ionic strength increase (0.5 M NaCl) changed the order of gel strength slightly at pH 7.4; CRUB > CRUA >> CRUC. The CRUC homohexamer showed weak gel formation when salt was added to the dispersion, indicating that a critical number of cross-links had been formed to generate a gel network which may also be due to the slight increase in solubility showed at pH 7.4 with 0.5 M NaCl ionic strength. At acidic pH (2.0), the network formation was in the order of CRUA >>> CRUC >>> CRUB. At pH 2.0, CRUA and CRUC homotrimers exhibited strong gel characteristics, but CRUB homotrimer was not able to form a structure upon heating. Changing the pH to acidic (pH 2.0) caused all cruciferins to assume small hydrodynamic radii, most likely dissociated hexamers, which coincided with low denaturation temperatures (83.1–100.0 °C). The hexamer dissociation of CRUC and exposure of the IE face that contains a higher number of hydrophobic patches
compared to the CRUA and CRUB homotrimers, may have facilitated protein interactions leading to cross-linking and consequent aggregation.

The higher surface hydrophobicity and better spatial arrangement of CRUA and CRUB hexamers lead to the prediction that they possess stronger interfacial activities than the CRUC hexamer leading to better interface stability in O/W and/or A/W. For all three cruciferins, the IE face of the trimer structure showed more hydrophobic and less negatively charged residues compared to IA faces. When the higher number of hydrophobic patches on IE face and the comparatively stronger negative charge on the IA face is considered, the trimers or the dissociated products of the CRUC hexamer were predicted to possess good emulsifying and foaming ability, as well as high solubility in contrast to the hexameric form. Experimentally obtained emulsifying activity index (EAI) of cruciferins was in the order of CRUB > CRUA >> CRUC for canola-in-water and triolein-in-water emulsions at near neutral pH demonstrating CRUC as a poor emulsifier at neutral pH compared to other two homohexameric cruciferins.

As stated earlier, the CRUC homohexamer has a comparatively lower surface hydrophobicity and a lower solubility compared to the CRUA and CRUB homohexamers at near neutral pH, and may contribute to its poor adsorption ability at the interface. Addition of NaCl at 0.5 M concentration negatively affected the emulsifying properties (activity and stability). Surface hydrophobicity of cruciferins did not change considerably when ionic strength is increased; indicating hexameric structure was stable under such condition. Similarly, solubility of cruciferins was not affected by high ionic strength at pH 7.4.

At pH 2.0, conditions were favorable for CRUC to stabilize O/W interfaces as evidenced by well-dispersed oil droplets containing emulsions were formed. The emulsions of all three cruciferins formed at pH 2.0 were more stable than the emulsions at pH 7.4 and phase separation (or creaming) was evident only after 3 days of storage at ambient temperature. As mentioned earlier, at pH 2.0, cruciferin hexameric assembly dissociates and generate the trimeric form exposing highly hydrophobic IE face (4.0–5.7 fold surface hydrophobicity increase compared to pH 7.4), resulting in enhanced emulsifying properties of all cruciferins. Predicted emulsifying properties using cruciferin homology models were in accordance with these experimental findings.

_Pepsin Susceptibility._ Changes in the synchrotron FT-IR amide-I signal band of the proteins of CRUAbc, and CRUaBc seeds indicated that they were more susceptible than CRUabC to pepsin-
catalyzed secondary structure disruption. Although it was difficult to attribute this difference to cruciferin subunit composition because of the co-existence of napin in the PSV, this experiment showed that it is possible to assess protein degradability of intact proteins using FT-IR chemical imaging. The in silico predictions indicated possible poor proteolytic cleavage of CRUC. For all cruciferins, the pepsin cleavage sites were predicted only on HVRs; none of the nick sites were found on β-sheets. Further investigation of 3D structures for predicting proteolytic sites revealed that CRUC has a limited number of pepsin nick sites located on HVR-II, which is buried inside the hexamer. In the isolated form, CRUC homohexamer showed an initial delay in pepsin-catalyzed hydrolysis compared to CRUA and CRUB homohexamers. This slow entering into pepsinolysis of CRUC homohexamers may be related to its requirement of low pH-induced dissociation to expose scissile sites for pepsin. Those peptic nick sites predicted using homology models were located mostly in the HVR-II for CRUC, which is located in the IE face and need to be exposed for the initiation of hydrolysis.

Contribution of each Subunit to WT Cruciferin. When all the investigated physico-chemical and functional properties are considered, CRUC homohexamers differed distinctly from CRUA or CRUB homohexamers. The major structural differences identified in these three protomers are related to the HVRs, of which CRUC has long HVR-I. Modeling and assembling of the CRUC hexamer using Pru du amandin from Prunus dulcis (PDB ID 3EHK) as the template clearly showed that the inter-winding of HVR-I allows higher probability of additional interactions between the two trimers than what occurs in homohexamers of CRUA or CRUB. This may result in a stable and compact molecule when the hexamer is composed of only CRUC subunits. In the WT heteromeric cruciferin, the CRUC subunit may associate with two other subunits, either CRUA or CRUB, or both. According to the LabChip capillary electrophoresis analysis, the relative subunit ratio was approximately 1:3:6 (CRUA: CRUB: CRUC, respectively) in the WT cruciferin. Although CRUC was in high proportion among the subunits, interacting with at least one other subunit type in forming the hexamer must have negated the exceptional physico-chemical and functional properties observed in the CRUC homohexamer. For example, low solubility at pH 7.4, poor emulsifying ability, and poor heat-induced gel forming ability were not carried over to the WT cruciferin. Therefore, explanation of physico-chemical and functional properties of WT cruciferin should be more reasonable using combinations of CRUA, CRUB, and CRUC subunits.
Can We Extrapolate This Knowledge to *B. napus*? Canola/rapeseed has three subunits in its WT cruciferin: CRU2/3, CRU4, and CRU1. The phylogenetic analysis suggests that the rapeseed CRU2/3, CRU4, and CRU1 corresponds to CRUA, CRUB, and CRUC subunit, respectively in Arabidopsis. Amino acid sequence comparison indicates that CRUA, CRUB, and CRUC subunits of Arabidopsis share 78, 84, and 83% identity with *B. napus* subunits CRU2/3, CRU4, and CRU1, respectively. Thus, the structural, physico-chemical, and functional properties that have been investigated on homomeric Arabidopsis cruciferins (i.e., CRUA, CRUB, and CRUC homohexamers) can reasonably be extrapolated to corresponding *B. napus* subunits.

Practical Implications. Information of this study on Arabidopsis 11S cruciferin has substantial value in crucifer oilseed protein utilization in designing crucifer protein in the plant as well as in postharvest use of the protein. The wide spread use of crucifer protein is based on the nutritional value, however functions of the proteins are important in interface stabilization (e.g., bio-pesticide emulsions and foams), thermo-plasticization (e.g., bio-plastics, bio-composites), and solid phase interactions (e.g., bio-based adhesives and glues) when the crucifer protein non-food uses are concerned. In traditional prepress solvent extraction of canola, the desolventization-toasting step brings protein meal to ~110 °C (Unger, 1990) that cause significant deterioration of protein functionality. Considering the high thermal stability of CRUC homohexamer, cruciferin consisting of only CRUC subunits may not undergo extensive thermal denaturation as CRUA or CRUB homohexamers, therefore enabling it to retain some functions. In terms of the functionalities required at neutral pH cruciferin consisting of CRUA or CRUB may perform better solubility and interfacial activities than CRUC. In food applications of crucifer proteins as functional molecules, the performance of CRUC homohexamer at low pH values may have a considerable value. The O/W emulsifying activity and stabilization of CRUC homohexamer below pH 4 is important in not only acidic food applications but also when salt is present. Similarly, this observed property of CRUC is significant in protein-based detergent (non-food) applications when salt ions are present in low pH. For the applications such as heat-induced gel structure formation in foods near neutral pH, CRUA or CRUB homohexamers perform better than that of CRUC. These are few examples that information of this study can be utilized in designing crucifer protein expression according to the final use.

Contribution to Knowledge on 11S SSP. The present study tested the hypothesis that the differences in molecular structure features brought out by different amino acid residues and
length of amino acid sequences can affect physico-chemical properties and techno-functionalities of 11S cruciferin and the experimental results supported the hypothesis. The differences of tested cruciferin protomers were mostly found in the HVRs. The length and the consisting amino acid residues of HVRs associate with the flexibility and differences in the interactions in the hexamer formation and stabilization. Wright (1987) first reported that a considerable degree of homology exists between putatively homologous 11S globulins of legumes and non-legume primary sequences but differences restricted to the variable regions or sequence inserts of the α-chain also exist. Such differences were suggested as allowing 11S globulins to accommodate considerable changes in their primary structure, accounting size differences between various 11S globulin subunits. The present work using Arabidopsis 11S cruciferin indicated that the variable regions provide distinguishable physico-chemical properties and functionalities to the 11S protein molecules, especially in those related to the tertiary and quaternary structure levels. The differences in these variable regions of the primary structure affect molecule structure related properties such as thermal stability, surface charge, surface hydrophobicity, proteolytic enzyme susceptibility and pH induced dissociation and the techno-functionalities manifested by these.
9. GENERAL CONCLUSIONS AND FUTURE IMPLICATION

It was hypothesized that: i) the cruciferin protein expressed by each CRU gene has some differences in the number and type of amino acid residues as well as hypervariable regions of the expressed proteins. These differences in protein sequence will reflect in the protein structure and its features; ii) the differences in molecular structure features of these cruciferins will be reflected in the physico-chemical properties and techno-functional properties. To test these hypotheses, five different studies were conducted using A. thaliana cruciferins generated from plants expressing different CRU genes. Conclusions drawn from these studies are as follows.

1. The Arabidopsis double-knockout lines, CRUAbc, CRUaBc, and CRUabC, expressed only one type of CRU gene and accumulated homohexameric cruciferins. The seeds possessed similar level of protein as the seed of WT plant.

2. Synchrotron-powered FT-IR is a suitable technique to monitor protein distribution, protein secondary structure contents, and bioavailability of a protein within plant tissues (including single cell area) without removing them from their native environment.

3. The three-step medium pressure chromatographic procedure allows isolation of the hexameric form of cruciferins with high purity (>95%).

4. Tertiary structure properties (i.e., surface charge, surface hydrophobicity, compactness, etc.) as observed from the measurements of surface hydrophobicity ζ-potential, and intrinsic fluorescence are linked to thermal denaturation parameters, structural stability to pH and salt, and are affected by the subunits composing the cruciferin hexamer.

5. Functional properties, namely solubility, heat-induced network formation, and emulsion properties are related to the physico-chemical properties mentioned in #4.

6. Most of the structural, physico-chemical, and functional properties of CRUA and CRUB homohexamers are comparable to the WT heterohexameric cruciferin.
7. The homohexameric CRUC exhibits a compact, thermostable, and less hydrophobic structure, and possesses unique physico-chemical properties those deviate from those of CRUA and CRUB.

8. The 11S hexameric structure of cruciferins undergoes low pH-induced dissociation to form trimers, which have functionalities different from the hexamers.

9. Generation of the trimeric form due to extrinsic factors (pH 2.0) is highly favorable for CRUC, which enhances its functional properties, specifically solubility, heat-induced structure formation and O/W interface stabilisation.

10. None of the cruciferins is able to perform enhanced functionalities (solubility, emulsion, and gelation properties) at high ionic strength (0.5 M NaCl).

11. All the cruciferins are susceptible to pepsin cleavage and possesses good digestibility characteristics with initial delay in pepsin-catalyzed hydrolysis of CRUC.

12. The predicted physico-chemical properties based on of 3D structure conformation of homology modeling, such as surface hydrophobicity, pepsin digestibility, O/W emulsifying, and heat-induced network formation can reasonably be confirmed by the data gathered from laboratory experiments performed on homomeric cruciferins of Arabidopsis seeds.

**Overall Conclusions:** This study shows that each cruciferin subunit has subtle differences in structural properties at the secondary structure level owing to the predominant conserved secondary structure features. At the tertiary and quaternary structure levels, the differences in non-conserved regions that compose HVR and ELR differentiate the physico-chemical and functional properties elicited by the assembled structures. Of the three-cruciferin subunits that were studied, the CRUC subunit generated distinct characteristics in the trimer and hexamer leading to different property characteristics than other two subunits. Moreover, investigation of such subtle changes in protein primary structure for the differences in higher order protein structures can be done by *in silico* homology modeling. Also it was possible to predict physico-chemical properties of Arabidopsis 11S protein using the homology modeling approach with reasonable accuracy.
Future Studies

1. The present study used cruciferin hexamers composed of similar types of subunits. Since WT cruciferin is a mixture of different combinations of CRUA, CRUB and CRUC, the seeds of single cruciferin gene knockout lines (CRUABc, CRUaBC, and CRUAbC) may provide suitable proteins to study the contribution of two protomer combinations to the properties of cruciferin. This information will enable to understand the behavior and properties of WT cruciferin.

2. The crystal structure of cruciferin expressed in seeds is not available. In the present study homomeric hexameric cruciferin composed of single subunit expressed in crucifer plants was isolated in high degree of purity. The CRUA and CRUB homohexamers have short HVR regions and the CRUB homotrimer has the shortest loop regions. Particle size analysis using DLS (data not reported here) suggests that the CRUB homohexamer has fewer tendencies for protein-protein aggregation (close to a mono disperse solution). Therefore, CRUB will be suitable to obtain crystals for X-ray diffraction studies to obtain cruciferin crystal structure information.

3. The cruciferin structure models developed in this study and their properties obtained from actual experiments allow identifying potential areas of the molecule for modification (genetically) to improve physico-chemical properties and functionalities. The designed cruciferin molecules with enhanced functionalities would help to improve and expand application of Brassicaceae SSP.
10. REFERENCES


Kim, C-S., Kamiya, S., Sato, T., Utsumi, S., & Kito, M. (1990a). High-level expression,
purification and functional properties of soybean proglycinin from *Escherichia coli*. *Agricultural and Biological Chemistry*, 54, 1543–1550.


Appendix Figure A1. Similarities and dissimilarities of closely related cruciferin 12S globulins. (a) Phylogenetic relationship of cruciferin proteins examined by the maximum likelihood method using the PhyML program. The numbers for interior branches refer to the bootstrap values for 100 replications. The scale at the bottom is in units of nucleotide substitutions per site. Cruciferin proteins from *Arabidopsis thaliana* (At, red), *Brassica napus* (Bn, blue), *Sinapis alba* (Sa, green) and *Raphanus sativus* (Rs, orange) were used and data bank (TAIR, NCBI, Gene Bank, or Swiss Prot) accession numbers are given. (b) Homology models of all putative *Arabidopsis* cruciferins. Locations of inter-chain (IE) and intra-chain (IA) disulfide bonds are shown in cyan and black beads, respectively. Missing portion of protomer (i.e. acidic or basic chain, dashed box) or disulfide bonds (dashed circle) is indicated. Alpha and beta chains of each model are represented in contrasting colors. (c) Schematic representation of alignment of proteins arising from each alternatively spliced transcripts of *Arabidopsis* cruciferin genes.
Appendix Figure A2. PCA of the deconvoluted amide I band of synchrotron FT-IR spectra of untreated and pepsin-treated WT (A) and SSP-altered Arabidopsis seeds: CRUAbc (B), CRUaBc (C), CRUabc (D), Cruciferin triple-knockout (CRU–) (E), napin-RNAi (F), and B. napus (G). Biplot of the first two principal components (i) and first and third principal components (ii) shows possible genetic relationship of the seed protein expression.
Appendix Figure A3. Multiple sequence alignment (ClustalW) of Arabidopsis and B. napus cruciferins. Identical (purple highlight and white letters) and conserved (faded purple highlight) amino acids are indicated. Sulfhydryl group containing cysteine residues are shown in black letters in yellow boxes. Secondary structure assignment is similar to the Withana-Gamage et al. (2011). HVR-I and II, hypervariable region I and II, respectively. ELR, extended loop region. The different numbers of aromatic amino acids, tryptophan (red highlight), tyrosine (green highlight), and phenylalanine (blue highlight) can be used as intrinsic fluorescence probe to determine cruciferin structure properties.
Appendix Figure A4. Intrinsic fluorescence of cruciferin as a function of ionic strength. (A) Emission spectra, (B) fluorescence peak area, and (C) maximum emission wavelength. All spectra were recorded at an excitation wavelength of 280 nm. \( *P < 0.05, **P < 0.01, ***P < 0.001 \) (ANOVA). Error bars, ± SD \((n = 3)\).
Appendix Figure A5. Intrinsic fluorescence of cruciferin as a function of pH at $\mu = 0.0$ (black bar) $\mu = 0.5$ (red bar); (A) Fluorescence peak area and (B) maximum wavelength of emission. All spectra recorded at excitation wavelength of 280 nm. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (ANOVA). Error bars, ± SD ($n = 3$).
Appendix Figure A6. Fatty acid composition of commercial canola oil. Canola oil is characterized by a high level of oleic (18:1, ω9, cis-Δ9, 61.1%), linoleic (18:2, ω6, cis-Δ9 cis-Δ12, 19.3%), and linolenic (18:3, ω3, cis-Δ9 cis-Δ12 cis-Δ15, 9.8%) fatty acids.
Appendix Figure A7. Intrinsic fluorescence of cruciferin (50 µg mL\(^{-1}\)) with and without salt in 10 mM phosphate buffer, pH 7.4. (A) Fluorescence intensity, (B) calculated relative peak area, and (C) \(\lambda_{\text{max}}\) of emission peak of tryptophan fluorescence.
### Appendix Table A1. Properties of cruciferins arising from various Arabidopsis gene models

<table>
<thead>
<tr>
<th>Locus (name)</th>
<th>Gene model</th>
<th>Calc. MW (kDa)</th>
<th>Calc. pI</th>
<th>Length</th>
<th>Signal peptide(^a)</th>
<th>Alpha/Beta chain</th>
<th>Disulfide bond (IA/IE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT5G44120(CRA1)</td>
<td>AT5G44120.1</td>
<td>31645.3</td>
<td>9.08</td>
<td>285</td>
<td>none</td>
<td>missing/96–267</td>
<td>both missing</td>
</tr>
<tr>
<td></td>
<td>AT5G44120.2</td>
<td>41031.6</td>
<td>7.13</td>
<td>368</td>
<td>none</td>
<td>missing/179–355</td>
<td>missing/8↔185</td>
</tr>
<tr>
<td></td>
<td>AT5G44120.3 (CRUA)</td>
<td>52594.8</td>
<td>7.68</td>
<td>472</td>
<td>1–24</td>
<td>25–282/283–472</td>
<td>36↔69/112↔289</td>
</tr>
<tr>
<td>AT1G03880 (CRU2)</td>
<td>AT1G03880.1 (CRUB)</td>
<td>50558.0</td>
<td>6.53</td>
<td>455</td>
<td>1–24</td>
<td>25–269/270–455</td>
<td>30↔63/106↔276</td>
</tr>
<tr>
<td>AT4G28520 (CRU3)</td>
<td>AT4G28520.1 (CRUC)</td>
<td>58235.2</td>
<td>6.53</td>
<td>524</td>
<td>1–24</td>
<td>25–346/346-524</td>
<td>37↔70/113↔340</td>
</tr>
<tr>
<td></td>
<td>AT4G28520.2</td>
<td>43723.7</td>
<td>6.51</td>
<td>394</td>
<td>none</td>
<td>31–303/missing</td>
<td>37↔70/113↔340</td>
</tr>
<tr>
<td></td>
<td>AT4G28520.3</td>
<td>50062.2</td>
<td>7.16</td>
<td>453</td>
<td>1–24</td>
<td>25–268/269–453</td>
<td>37↔70/missing</td>
</tr>
<tr>
<td></td>
<td>AT4G28520.4</td>
<td>43878.9</td>
<td>6.94</td>
<td>396</td>
<td>1–24</td>
<td>25–346/missing</td>
<td>37↔70/113↔340</td>
</tr>
<tr>
<td>AT1G03890</td>
<td>AT1G03890.1</td>
<td>49674.5</td>
<td>5.45</td>
<td>451</td>
<td>1–29</td>
<td>30–260/261–446</td>
<td>36↔69/112↔277</td>
</tr>
</tbody>
</table>

\(^a\)Signal peptides were predicted using SignalP 3.0 Server (http://www.cbs.dtu.dk/services/SignalP/)