



**HAL**  
open science

# Caractérisation fonctionnelle des protéines WSCP "Water Soluble Chlorophyll Proteins" du colza

Youssef Bouargalne

► **To cite this version:**

Youssef Bouargalne. Caractérisation fonctionnelle des protéines WSCP "Water Soluble Chlorophyll Proteins" du colza. Sciences agricoles. Université de Rennes, 2022. Français. NNT : 2022REN1B028 . tel-04124772

**HAL Id: tel-04124772**

**<https://theses.hal.science/tel-04124772>**

Submitted on 11 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

ECOLE DOCTORALE N° 600

*Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation*

Spécialité : Biologie et Physiologie végétales

Par

**Youssef BOUARGALNE**

## **Caractérisation fonctionnelle des protéines WSCP « Water-Soluble Chlorophyll Proteins » du colza**

Thèse présentée et soutenue à Rennes, le 10 Juin 2022

Unité de recherche : UMR 1349 Institut de Génétique, Environnement et Protection des Plantes ;

INRAE / Institut Agro Rennes-Angers / Université de Rennes 1

### **Rapporteurs avant soutenance :**

Anne REPELLIN  
Pierre CAROL

Professeur, Université Paris Est Créteil  
Professeur, Université Sorbonne

### **Composition du Jury :**

*David MACHEREL*

*Professeur, Université Angers (Président)*

Anne REPELLIN

Professeur, Université Paris Est Créteil

Pierre CAROL

Professeur, Université Sorbonne

Philippe ETIENNE

Maître de conférences, Université Caen Normandie

Françoise LE CAHEREC

Maître de conférences, Université Rennes 1

Carole DELEU

Maître de conférences, Université Rennes 1 (Directrice de thèse)



## Remerciements

---

Après ces trois années et demi de thèse enrichissantes, il est l'occasion de remercier et d'exprimer ma profonde gratitude à tous ceux qui, de près ou de loin, ont œuvré dans l'élaboration de ce travail en espérant n'avoir oublié personne (si tel est le cas, mes excuses les plus sincères...).

En tout premier lieu, je tiens à remercier vivement ma directrice de thèse Carole DELEU et ma co-encadrante Françoise LE CAHEREC. Je vous suis vraiment reconnaissant pour le pari que vous avez fait sur moi et la confiance que vous m'avez témoignée, pour votre engagement, votre suivi, votre disponibilité, et vos encouragements tout le long de ce projet de thèse. Carole, merci pour tes remarques pertinentes qui vont toujours droit au but, tu m'as souvent orienté sans pour autant me la tracer et c'est l'une des meilleures écoles ! Françoise, merci pour ton attention de tout instant sur mes travaux, ton aide précieuse lors du décorticage des résultats, ton écoute et tes conseils avisés qui ont été prépondérants pour la bonne réussite de cette thèse. Merci de m'avoir soutenue lorsque je proposais de nouvelles manips (*quand celles-ci n'allaient pas dans tous les sens !*). Encore une fois merci à toutes les deux.

Je tiens également à remercier les membres de mon jury de thèse. Je suis très honoré que Anne REPELLIN ainsi que Pierre CAROL aient accepté d'être les rapporteurs de ces travaux. Je remercie également à David MACHEREL et Philippe ETIENNE d'avoir accepté d'être examinateurs de ce travail.

Je remercie les membres de mon comité de thèse, Céline RAGUENES-NICOL, Michèle REISDORF-CREN, Philippe ETIENNE, Philippe SIMIER et Mathieu ROUSSEAU-GUEUTIN, pour leur enthousiasme sur le projet, leur expertise et leurs conseils avisés sur l'évolution de mes recherches.

Mes remerciements les plus sincères s'adressent également à mes collaborateurs : Jean-François HUBERT et Angélique CHÉRON pour m'avoir appris les secrets de la production des protéines recombinantes. Je n'y serai jamais parvenue sans vos astuces et votre machine fantastique (Akta !!). Céline RAGUENES-NICOL, pour m'avoir initié à l'utilisation de la thermophorèse et de m'avoir laissé l'utiliser en autonomie, et aussi de m'avoir aidé pour l'analyse des résultats. Christophe TASCON, pour le temps qu'il a consacré pour les manips de chromatographie. Olivier DELALANDE, pour m'avoir initié au docking moléculaire. David MACHEREL, pour m'avoir accueilli chaleureusement au sein du IRHS et de m'avoir enseigné l'art d'expression transitoire dans les protoplastes. Philippe ETIENNE, pour les lignées surexprimeurs, qui ont été les seules expériences de physiologie pour ma thèse.

Je remercie également tous membres, passé et présents de l'équipe RCA. Merci à tous pour votre accueil, votre gentillesse, les diverses discussions et pour tous ces moments conviviaux. Merci notamment à Patrick (*big boss !*) pour ta gentillesse, ton humour et pour les petites astuces de pêche pour débutant. Merci Younès pour tes conseils, les discussions intéressantes que nous avons pu avoir autour de certaines de mes manips de Bioch, mais aussi pendant les petites balades. Je tiens à remercier plus particulièrement ceux qui ont travaillé de très près, pour l'aide précieux que vous m'avez apporté : Vanessa, un grand merci pour ton énorme soutien en biologie moléculaire et d'avoir toujours pris le temps de répondre à mes demandes. Anne-Marie, je te remercie pour les manips de physiologie. Florian, ces quelques lignes ne suffiront jamais pour te remercier... Merci infiniment d'avoir accepté de travailler avec moi (*ou devrais-je dire le contraire ?*) sur les protéines recombinantes énigmatiques, un énorme merci pour ton soutien, ton humour, pour les moments inoubliables de

rigolades et fou-rires lors des manips qui ne marchaient pas ... ! ces moments ont été très précieux et utiles pour l'aboutissement de ce travail.

Je n'oublie pas, bien sûr, de remercier le LBDF et l'ensemble de ces occupants passés et présents : Amélie, Anaëlle, Anani, Céline, Erwan, Franz, Gautier, Grégoire, Julie, Loeiz, Marianne, Marie D, Marie C, Mathieu, Mathilde, Maud, Maverick, Nadège, Pierre-Nicolas. Merci pour la gentillesse et pour tous ces bons moments partagés, nos discussions parfois scientifiques et souvent un peu moins. Merci Anani pour ton soutien, ton aide, mais aussi pour les bons repas africains. Une série de merci tout particulier à Franz, Loeiz, Marianne et Mathilde, pour les rigolades et les soirées mémorables, mais aussi pour votre soutien et pour m'avoir écouté me plaindre de ma thèse.

Merci à l'ensemble de l'unité IGEPP et à sa directrice Maria Manzanares-Dauleux puis Nesi Nathalie de m'avoir accueillie pour effectuer ma thèse au sein de l'institut. Merci également aux services communs de l'institut (services techniques, administratif, et informatique) qui nous permettent de réaliser les expériences dans de bonnes conditions. Je remercie également l'université Rennes1 et la Région Bretagne d'avoir financé ma thèse.

Mon expérience dans la recherche n'a pas débuté avec ma thèse, mais bel et bien lors de mon stage de fin d'études de master, au sein de FSTT à Tanger (Maroc). Un grand merci à Mohamed NHIRI et Reda BENMRID qui m'ont appris les ficelles de la biochimie, ainsi que la rigueur et l'autonomie lors des expérimentations. J'ai énormément appris à vos côtés. Merci aussi pour vos précieux conseils qui ont pu me servir au cours de cette thèse.

Outside of my biochemical life, je tiens à remercier infiniment toutes les personnes qui m'ont permis d'oublier momentanément la thèse dans des soirées, repas, sorties, voyages ou autres. Merci donc à Arij, Fatiha, Kamal, Maïlis, Marie, Nezha, Wassime, Elvina et Astin (*now it's time for the New Jersey RoadTrip and paragliding!!*). Un grand merci plus particulièrement à mes meilleurs amis " la banda del vino" (Ali, Ismail et Mehdi), toujours là pour le meilleur et pour le pire, vous êtes des amis en or sur qui je peux toujours compter.

Enfin, Je remercie infiniment ma famille pour leur amour inconditionnel et leur soutien sans faille, merci à mes parents, mes grands-parents, mes frères et sœurs et l'ensemble de mes proches que je ne saurai lister ici.

## AVANT-PROPOS

Cette thèse de doctorat s'est déroulée pendant 3 ans et 5 mois, de 1<sup>er</sup> décembre 2018 à 30 avril 2022. Les recherches présentées dans ce manuscrit ont été réalisées au sein de l'unité mixte de recherche : UMR 1349, IGEPP « Institut de Génétique, Environnement et Protection des Plantes », INRAE – Institut Agro Rennes-Angers – Université Rennes 1, sous l'encadrement de Carole DELEU, en qualité de directrice de thèse, et Françoise LE CAHEREC, en qualité d'encadrante de thèse. Ce projet de thèse a été financé pour moitié par l'Université de Rennes 1 et pour moitié par la Région Bretagne.

Le planning de mes travaux de recherche a été fortement impacté suite au confinement lié à l'épidémie de SARS-CoV2 pendant lequel toute expérimentation a dû être interrompue. Le report du travail a concerné non seulement l'acquisition des données brutes, mais également l'ensemble des collaborations qui avaient été initiées. Pour combler ces retards, une prolongation de cinq mois m'a été accordée. Le financement de la prolongation fut assuré par l'Université Rennes 1 et l'unité de recherche IGEPP.

Ma présence à deux congrès internationaux avait été planifiée avec mes encadrantes pour communiquer mes résultats. Il s'agissait pour moi de faire deux présentations orales au sein de ces deux congrès à savoir, « 5<sup>th</sup> Conference on Plant Proteases. 8-10 september 2020 (Ljubljana, Slovenia) » et « 10th International Symposium on Plant Senescence, 2021 ». Dans le contexte exceptionnel lié à la COVID, le 1<sup>er</sup> congrès a été reporté en septembre 2022, date qui va au-delà de la date butoir de ma prolongation de thèse, et le 2<sup>ème</sup> a été annulé. Face à ces conditions exceptionnelles, j'ai essayé de communiquer dans un congrès par visioconférence mais tous les congrès proposés ne présentaient aucun lien de près ou de loin avec ma thématique de recherche. De fait, aucune communication à un congrès international n'a pu se faire pendant ma thèse. Cependant, j'ai pu assister au début de ma thèse au congrès international « 9<sup>th</sup> International Symposium on Plant Senescence », en avril 2019 à Berlin, et j'ai également eu l'opportunité d'exposer mes résultats dans les lab-meeting de mon équipe « Rendement sous Contraintes abiotiques (RCA) » et les journées des doctorants de l'IGEPP et de l'école doctorale EGAAL.

Ce travail a été enrichi par des collaborations externes à l'IGEPP, discussions et interactions avec les membres de l'unité de recherche IGEPP ainsi qu'avec les membres du Comité de Suivi de thèse. Cette thèse devrait donner lieu à 3 articles dont 2 comme premier auteur et 1 comme co-auteur. Cette thèse m'a aussi donné l'opportunité de suivre de nombreuses formations scientifiques que ce soit en éthique de la recherche et qu'en communication scientifique.



# SOMMAIRE

<b>PROBLEMATIQUE ET OBJECTIFS DE RECHERCHE .....</b>	<b>1</b>
<b>ANALYSE BIBLIOGRAPHIQUE .....</b>	<b>4</b>
<b>I. Classification des WSCP, protéines atypiques de liaison à la chlorophylle.....</b>	<b>4</b>
I.1. Complexes WSCPI-chlorophylle photoconvertibles.....	4
I.2. Complexes WSCP II-chlorophylle non-photoconvertibles .....	5
I.2.1. Existence de deux sous-classes de WSCP II .....	5
I.2.2. Vers l'existence d'une troisième sous-classe .....	6
I.2.3. Membres de la famille des inhibiteurs de type Kunitz .....	6
I.2.4. Maturation post-traductionnelle.....	7
I.2.5. Diversité d'identités et de propriétés physico-chimiques .....	7
<b>II. Propriétés fonctionnelles des WSCP II : interaction avec les Chls .....</b>	<b>7</b>
II.1. Approches méthodologiques.....	7
II.1.1. Purification des complexes natifs WSCP II-Chls à partir des feuilles .....	7
II.1.2. Reconstitution <i>in vitro</i> des complexes WSCP II-Chls.....	8
II.2. Structuration des complexes WSCP II-Chls .....	10
II.2.1. Assemblage en tétramères .....	10
II.2.2. Stabilisation du complexe WSCP II-Chl .....	11
II.3. Sélectivité-affinité des WSCP II aux Chls .....	13
II.4. Stœchiométrie des Chls dans les complexes WSCP II-Chls.....	14
II.5. Protection des chlorophylles .....	15
<b>III. Propriétés fonctionnelles des WSCP II : interaction avec les protéases.....</b>	<b>15</b>
<b>IV. Rôles physiologiques des WSCP II .....</b>	<b>17</b>
IV.1. Localisation tissulaire et subcellulaire des WSCP II .....	17
IV.2. Implication des WSCP II dans le métabolisme de la Chl .....	18
IV.3. Implication des WSCP II dans le contrôle de la protéolyse.....	20
IV.3.1. Rôle dans les tissus chlorophylliens en croissance .....	20
IV.3.2. Rôle dans les tissus non chlorophylliens : mort cellulaire programmée et skotomorphogénèse .....	20
IV.4. Implication des WSCP II dans la réponse aux contraintes environnementales .....	21
IV.4.1. Stress biotiques.....	21
IV.4.1. Stress abiotiques.....	22
<b>CHAPITRE I : CARACTERISATION FONCTIONNELLE DE BnD22 IN VITRO .....</b>	<b>24</b>
<b>Presentation .....</b>	<b>24</b>
<b>Article 1: New insights into chlorophyll-WSCP (water-soluble chlorophyll binding proteins) interactions: the case study of BnD22 (<i>Brassica napus</i> drought-induced 22 kDa) .....</b>	<b>26</b>
<b>Article 2 :<i>Brassica napus</i> Drought 22 kD (BnD22), a water-soluble chlorophyll protein acting simultaneously as a cysteine protease inhibitor and chlorophyll-binding protein. ....</b>	<b>37</b>
<b>CHAPITRE II : IDENTIFICATION DE LA FAMILLE DES WSCP II DU COLZA .....</b>	<b>55</b>
<b>Presentation .....</b>	<b>55</b>
<b>Article 3:Genome-wide identification and expression analysis of WSCP II genes in <i>Brassica napus</i> .....</b>	<b>57</b>
<b>CHAPITRE III : DETERMINATION DES PROPRIETES FONCTIONNELLES DES WSCP II DE COLZA A PARTIR DE PROTEINES RECOMBINANTES .....</b>	<b>74</b>
<b>Introduction .....</b>	<b>74</b>
<b>Matériels et méthodes .....</b>	<b>75</b>
<b>I. Clonage et obtention des bactéries transformées.....</b>	<b>75</b>



<b>II. Production et purification des WSCP II recombinantes dans E. coli .....</b>	<b>75</b>
II.1. Expression des protéines recombinantes.....	75
II.2. Extraction des protéines recombinantes.....	76
II.3. Purification des protéines recombinantes .....	76
II.4. Analyse par PAGE-SDS et Western-Blot.....	76
<b>III. Mesure des activités anti- protéolytiques .....</b>	<b>77</b>
<b>IV. Capacités de liaison à la chlorophylle.....</b>	<b>77</b>
IV.1. Retard sur gel ou Electrophoretic Mobility Shift Assay (EMSA) .....	77
IV.2. Détermination de l'affinité des WSCP II aux chlorophylles .....	78
<b>Résultats .....</b>	<b>79</b>
<b>I. Sélection et choix des protéines représentatives des WSCP II du colza .....</b>	<b>79</b>
<b>II. Production et purification des WSCP II recombinantes.....</b>	<b>79</b>
II.1. Expression et degré de solubilité des protéines recombinantes .....	79
II.2. Purification des protéines recombinantes solubles .....	81
<b>III. Evaluation des propriétés fonctionnelles des WSCP II recombinantes.....</b>	<b>82</b>
III.1. Activités anti-protéolytiques .....	82
III.2. Capacité de liaison à la chlorophylle .....	83
<b>Discussion, conclusion et perspectives .....</b>	<b>85</b>
 <b>DISCUSSION GENERALE ET PERSPECTIVES.....</b>	 <b>88</b>
 <b>REFERENCES BIBLIOGRAPHIQUES .....</b>	 <b>95</b>
 <b>ANNEXES .....</b>	 <b>101</b>

## LISTE DES FIGURES

### ANALYSE BIBLIOGRAPHIQUE

- Fig. 1 : Séquence protéique et caractéristiques spectrales et structurales de CaWSCP, WSCPI de *Chenopodium album*.  
Fig. 2 : Séquences et caractéristiques spectrales et structurales de BolWSCP et LvWSCP  
Fig. 3 : Analyse phylogénétique des WSCPII (d'après Prabahar *et al.*, 2020)  
Fig. 4 : Résultats d'analyse par PAGE-SDS de protéines recombinantes WSCPII exprimées dans un système bactérien : exemple de BolWSCP (d'après Takahashi *et al.*, 2012, 2014)  
Fig. 5 : Reconstitution *in vitro* des complexes WSCP-Chl : exemple de BolWSCP-Chla  
Fig. 6 : Structure tridimensionnelle du complexe LvWSCP-Chl de *Lepidium virginicum* (d'après Horigome *et al.*, 2007)  
Fig. 7 : Superposition des structures des complexes BolWSCP-Chl et LvWSCP-Chl (A et B) et focus sur les domaines et acides aminés interagissant avec la Chl (C et D)  
Fig. 8 : Interactions entre les motifs d'affinité et la Chl dans les complexes LvWSCP-Chlb et BolWSCP-Chla (d'après Palm *et al.*, 2018)  
Fig. 9 : Complexes BolWSCP-Chl natifs ou reconstitués, associés à 2 ou 4 Chls (d'après Palm *et al.*, 2017)  
Fig. 10 : Thermo- et photo-stabilité des complexes WSCPII-Chl  
Fig. 11 : Activités anti-protéolytiques des WSCPII  
Fig. 12 : Localisations tissulaires de WSCPII  
Fig. 13 : Localisation subcellulaire des WSCPII  
Fig. 14 : Implication physiologique des WSCPII dans le métabolisme de la Chl (A et B) et le développement des graines (C).  
Fig. 15 : Implication des WSCPII dans la réponse aux contraintes environnementales

### CHAPITRE III : DETERMINATION DES PROPRIETES FONCTIONNELLES DES WSCPII DE COLZA A PARTIR DE PROTEINES RECOMBINANTES

- Fig. 1 : Principales étapes pour l'obtention des protéines recombinantes WSCPII de Colza et leur caractérisation  
Fig. 2 : Alignement multiple des protéines WSCPII de colza sélectionnées et BnD22  
Fig. 3 : Induction et solubilité des WSCPII recombinantes entières (avec l'extrémité C-terminale)  
Fig. 4 : Analyse par PAGE-SDS des protéines WSCPII recombinantes entières (avec C-terminal) après purification  
Fig. 5 : Evaluation de l'inhibition des activités protéolytiques de la trypsine, chymotrypsine et papaine par des quantités croissantes de WSCPII recombinantes entières purifiées.  
Fig. 6 : Tests d'oligomérisation des WSCPII recombinantes purifiées Prot1, Prot3, Prot4 et Prot6 entières en absence de Chl (contrôle) ou en présence de la Chla (+Chla) ou de la Chlb (+Chlb)  
Fig. 7 : Caractérisation de la liaison des WSCPII recombinantes entières à la Chl

### DISCUSSION GENERALE ET PERSPECTIVES

- Fig. 1 : Phénotypage des plantules *pSAG12::WSCPI* en présence de mannitol

## LISTE DES TABLEAUX

### ANALYSE BIBLIOGRAPHIQUE

- Tab. 1 : Identités entre les WSCPII identifiées par Prabahar *et al.* (2020)  
Tab. 2 : Caractéristiques physico-chimiques et localisation subcellulaire des WSCPII identifiées par Prabahar *et al.* (2020)

### CHAPITRE III : DETERMINATION DES PROPRIETES FONCTIONNELLES DES WSCPII DE COLZA A PARTIR DE PROTEINES RECOMBINANTES

- Tabl. 1 : Caractéristiques géniques, physico-chimiques et fonctionnelles des 6 WSCPII sélectionnées  
Tabl. 2 : Identité entre les séquences protéiques des 6 WSCPII sélectionnées entières et BnD22  
Tabl. 3 : Conditions de culture optimales produisant le meilleur taux de solubilité pour chacune des 6 WSCPII recombinantes entières ou tronquées

### DISCUSSION GENERALE ET PERSPECTIVES

- Tabl.1 : Propriétés fonctionnelles de WSCPII caractéristiques de la diversité naturelle chez le colza et sites principaux d'expression.

## LISTE DES ABREVIATIONS

AnWSCP: ancestrale WSCP	kDa: kilodalton
AtWSCP: WSCPII d' <i>Arabidopsis thaliana</i>	LHCII : light-harvesting complex II
BAEE : N- $\alpha$ -benzoyl-L-tyrosine ethyl ester	LvWSCP : WSCPII de <i>Lepidium virginicum</i>
BAPNA : N- $\alpha$ -Benzoyl-DL-arginine p-nitroanilide	MST: microScale thermophoresis
BnD22: <i>Brassica napus</i> drought 22 kDa protein	NP40 : Nonidet-P40
BolWSCP : WSCPII de <i>Brassica oleracea</i>	OG : octyl $\beta$ -D-glucopyranoside
BP80 : vacuolar-sorting receptor protein	PAGE-SDS: électrophorèse sur gel d'acrylamide en présence de SDS
BTEE : N- $\alpha$ -benzoyl-L-tyrosine ethyl ester	PCR : polymérase chain reaction
CaWSCP : WSCPI de <i>Chenopodium album</i>	PVDF : polyvinylidène difluoride
CBD : chlorophyll binding domain	qRT-qPCR: quantitative reverse-transcriptase polymérase chain reaction
CD : circular dichroism	RD21: responsive-to-desiccation-21 protein
Chl : chlorophylle	RNA seq : ribonucleic acid sequencing
Chla : chlorophylle a	ROS: reactive oxygen species
Chlb : chlorophylle b	RshWSCP: WSCPII de <i>Raphanus sativus</i>
Chlide : chlorophyllide	RSL: reactive site loop
CP : chlorophyll protein	Ser-IP : inhibiteur de protéases à sérine
Cys-IP : inhibiteur de protéases à cystéine	STI : soybean trypsin inhibitor
DUF538 : domain of unknown function 538	TBS: Tris-buffered saline
ERA : efficacité de remobilisation de l'azote	TX114: Triton X114
ER-bodies : endoplasmic reticulum bodies	WSCP: water soluble chlorophyll proteins
EUA : efficacité d'utilisation de l'azote	WSCP-Chl : complexe WSCP - Chl
FCBP: free Chl(ide)-binding protein	WSCPI: water soluble chlorophyll proteins de classe I
His-Tag : hexa-histidine tag	WSCPII: water soluble chlorophyll proteins de classe II
IP : inhibiteur de protéases	
Kd : constante de dissociation	

## PROBLEMATIQUE ET OBJECTIFS DE RECHERCHE

---

Le colza (*Brassica napus L.*) de la famille des Brassicacées est principalement cultivé pour ses graines afin de produire des huiles alimentaires et industrielles et dont le tourteau résiduel, riche en protéines, est valorisé pour l'alimentation animale. Dans le cadre du changement climatique et de l'évolution de la réglementation concernant les intrants azotés, la culture du colza risque d'être fortement pénalisée dans la mesure où c'est une culture nécessitant d'importants apports azotés pour atteindre un potentiel de rendement satisfaisant. Comme la pérennité de la culture du colza est largement dépendante de sa capacité à maintenir des rendements élevés, il sera nécessaire d'améliorer son efficacité d'utilisation de l'azote (EUA). Ce caractère complexe est subdivisé en trois composantes que sont les efficacités d'absorption, d'assimilation et de remobilisation (ERA) (Avicé et Etienne, 2014). L'absorption et l'assimilation de l'azote ne seraient pas ou peu limitantes chez le colza alors que plusieurs études ont montré que la ERA serait plus pénalisante, notamment au cours du vieillissement des tissus foliaires et de la chute prématurée de feuilles encore riches en azote organique (Malagoli *et al.*, 2005; Gombert *et al.*, 2006; Koeslin-Findeklee *et al.*, 2015). Dans ce contexte, l'ERA constitue un levier d'optimisation de l'EUA du colza particulièrement intéressant et ce, d'autant qu'il existe une variabilité génotypique naturelle de cette composante (Girondé *et al.*, 2015b). Pour répondre à ce défi, il est essentiel de connaître et d'identifier précisément les déterminants moléculaires impliqués dans les processus de remobilisation de l'azote associée à la sénescence naturelle ou induite par des contraintes environnementales. Dans ce contexte, les WSCPII (Water Soluble Chlorophyll Proteins de classe II) ont été identifiées comme des candidats potentiels. Ces protéines, décrites comme des inhibiteurs de protéases (IP) à sérine (Ser) et/ou à cystéine (Cys), sont également capables de se complexer avec les chlorophylles (Chls) et de protéger celles-ci de la photo-dégradation. A l'exception de la WSCPII unique d'*A. thaliana*, aucune étude n'indique clairement qu'une même WSCPII puisse assurer les deux fonctions d'IP et de liaison à la Chl.

Chez le colza, BnD22, la seule WSCPII identifiée à ce jour, présente un gradient d'expression antagoniste à celui des protéases (Ilami *et al.*, 1997; Etienne *et al.*, 2007; Desclos *et al.*, 2008) et est fortement accumulée dans les jeunes feuilles en réponse à des situations de stress tels qu'un stress hydrique (Reviron *et al.*, 1992), une carence en azote (Etienne *et al.*, 2007) ou en soufre (D'Hooghe *et al.*, 2013). BnD22 participerait ainsi au contrôle de la protéolyse et donc à la gestion des pools d'azote organique lors de la sénescence foliaire et en réponse aux contraintes abiotiques. La présence de BnD22 permettrait de maintenir l'intégrité physiologique des jeunes tissus foliaires dans l'attente de conditions plus favorables à la croissance. Avicé et Etienne (2014) proposent également que les WSCPII du colza pourraient permettre de réduire l'asynchronisme existant entre la période de vidage

des organes sources que sont les feuilles sénescentes et le remplissage des graines, et ainsi améliorer l'efficacité de remobilisation de l'azote chez le colza.

Le cœur de ce travail s'est principalement focalisé sur l'*étude fonctionnelle de BnD22* dont les caractéristiques en tant qu'inhibiteur de protéases restent sujettes à controverse et dont la capacité de liaison aux Chls n'a jamais été décrite. La stratégie a consisté à étudier *in vitro* les propriétés IP et de liaison à la Chl de protéines recombinantes BnD22 produites en système bactérien. Afin de mieux cerner le rôle de la modification post-traductionnelle qui consiste en un clivage de l'extrémité C-terminale sur leurs propriétés fonctionnelles, les protéines recombinantes ont été réalisées avec ou sans l'extrémité C-terminale, protéines que nous avons nommées respectivement protéine entière ou protéine tronquée.

- L'aptitude des BnD22 (tronquée ou entière) à se lier à la Chla et/ou la Chlb a été évaluée en regard à leur état d'oligomérisation. Nous avons également développé une nouvelle méthode basée sur les taux de photoblanchiment des complexes BnD22-Chl pour définir précisément la sélectivité de BnD22 aux Chls.

- Les activités anti-protéolytiques de BnD22 ont été suivies *in vitro* contre deux sérine protéases, trypsine et chymotrypsine, et contre une protéase à cystéine, la papaine. Il s'est avéré que BnD22 est une protéine bi-fonctionnelle capable d'inhiber les Cys protéases et de se lier de manière préférentielle à la Chla.

- De plus, nous avons démontré pour la première fois, qu'une WSCPII peut assurer simultanément les deux fonctions. Le pouvoir inhibiteur de protéases de BnD22 est non seulement maintenu, mais il est même optimisé lorsqu'elles sont complexées aux Chls. Par modélisation structurale et docking moléculaire des complexes BnD22-Chl-papaine, nous avons pu définir les interactions moléculaires sous-jacentes. Par ailleurs, la localisation subcellulaire de BnD22 par expression transitoire dans les protoplastes de colza a permis de détecter sa présence dans le réticulum endoplasmique et la vacuole.

Ces travaux ont fait l'objet de 2 articles présentés dans le premier chapitre du manuscrit :

**Article1** : Bouargalne Y., Raguénès-Nicol C., Guilbaud F., Cheron A., Clouet V., Deleu C. and Le Cahérec F. (2022). New insights into chlorophyll-WSCP (water-soluble chlorophyll proteins) interactions: the case study of BnD22 (*Brassica napus* Drought 22 kDa) (Plant Physiology and Biochemistry, sous presse)

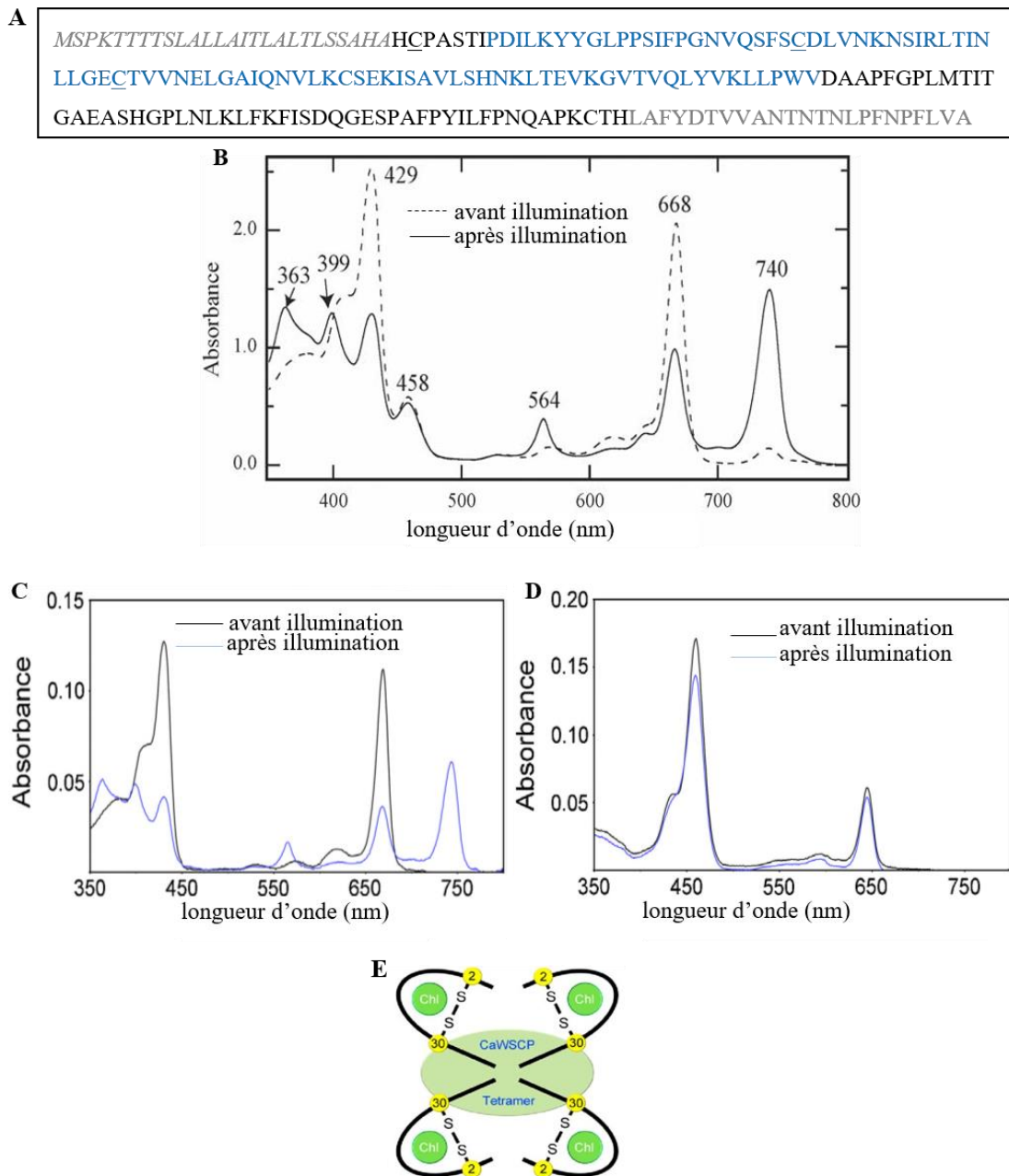
**Article2** : Bouargalne Y., Guilbaud F., Macherel D., Delalande O., Deleu C. and Le Cahérec F. (2022). *Brassica napus* Drought 22 kDa (BnD22), a water-soluble chlorophyll-binding protein acting simultaneously as a cysteine protease inhibitor and chlorophyll-binding protein (en préparation).

Dans le deuxième chapitre, nous avons *identifié l'ensemble des membres de la famille des WSCP II du colza*. En effet, plusieurs études suggéraient la présence d'isoformes WSCP II chez le colza ou le chou ce qui laissait pressentir la présence de familles multigéniques (Ilami *et al.*, 1997; Desclos *et al.*, 2008; Prabahar *et al.*, 2020). La famille des WSCP II n'avait pas encore été étudiée de manière exhaustive chez le colza. Par criblage des génomes, 56 gènes dont 16 pseudogènes ont été identifiés et caractérisés chez *B. napus*, et les espèces parentales *B. rapa* et *B. oleracea*. D'un point de vue phylogénétique, les WSCP II sont divisées en 4 groupes distincts associés à leur préférence de liaison à la Chla ou à la Chlb ou à la perte de cette fonction. Une analyse par qRT-PCR a révélé que les gènes WSCP II de *B. napus* pouvaient présenter des profils d'expression spécifiques aux organes et aux stades de développement. L'identification *in silico* des fonctions IP et de liaison à la Chl des protéines a permis de spécifier les propriétés fonctionnelles supposées au sein de chaque groupe phylogénétique.

Le travail de ce deuxième chapitre sera également présenté sous forme d'un article puisqu'il sera finalisé et soumis très prochainement.

**Article 3 :** Clouet V., Bouargalne Y., Rousseau-Gueutin M., Bouchereau A., Deleu C. and Le Cahérec F. (2022). Genome-wide identification and expression analysis of WSCP II genes in *Brassica napus* (en préparation).

Le troisième chapitre est consacré à la caractérisation de 6 protéines caractéristiques de la variabilité fonctionnelle des WSCP II du colza parmi les 20 protéines existantes. La même stratégie que celle utilisée pour BnD22 a été initiée pour les études préliminaires visant à caractériser leurs propriétés fonctionnelles. La multiplicité des WSCP II chez le colza serait donc associée à une diversité fonctionnelle (IP et/ou liaison à la Chl) et/ou d'expression tissulaire.



**Fig. 1 : Séquence protéique et caractéristiques spectrales et structurales de CaWSCP, WSCPI de *Chenopodium album***

**A)** Séquence protéique (UniProt: K4PX49). La séquence signal est indiquée en gris italique, l'extrémité C-terminale subissant un clivage post-traductionnel en gris et le domaine DUF538 (domain of unknown function) en bleu. Les deux cystéines soulignées sont celles essentielles pour la liaison à la Chl par formation des complexes WSCP1-Chl.

**B-C-D)** Spectres d'absorption avant et après illumination des complexes natifs CaWSCP-Chl extraits des feuilles de *Chenopodium album* (**B**), des complexes CaWSCP reconstitués avec la Chla (**C**) ou la Chlb (**D**) (d'après Hirabayashi *et al.*, 2006 ; Takahashi *et al.*, 2013b).

**E)** Modèle structural du complexe CaWSCP-Chl (Takahashi *et al.*, 2014).

## ANALYSE BIBLIOGRAPHIQUE

---

### I. Classification des WSCP, protéines atypiques de liaison à la chlorophylle

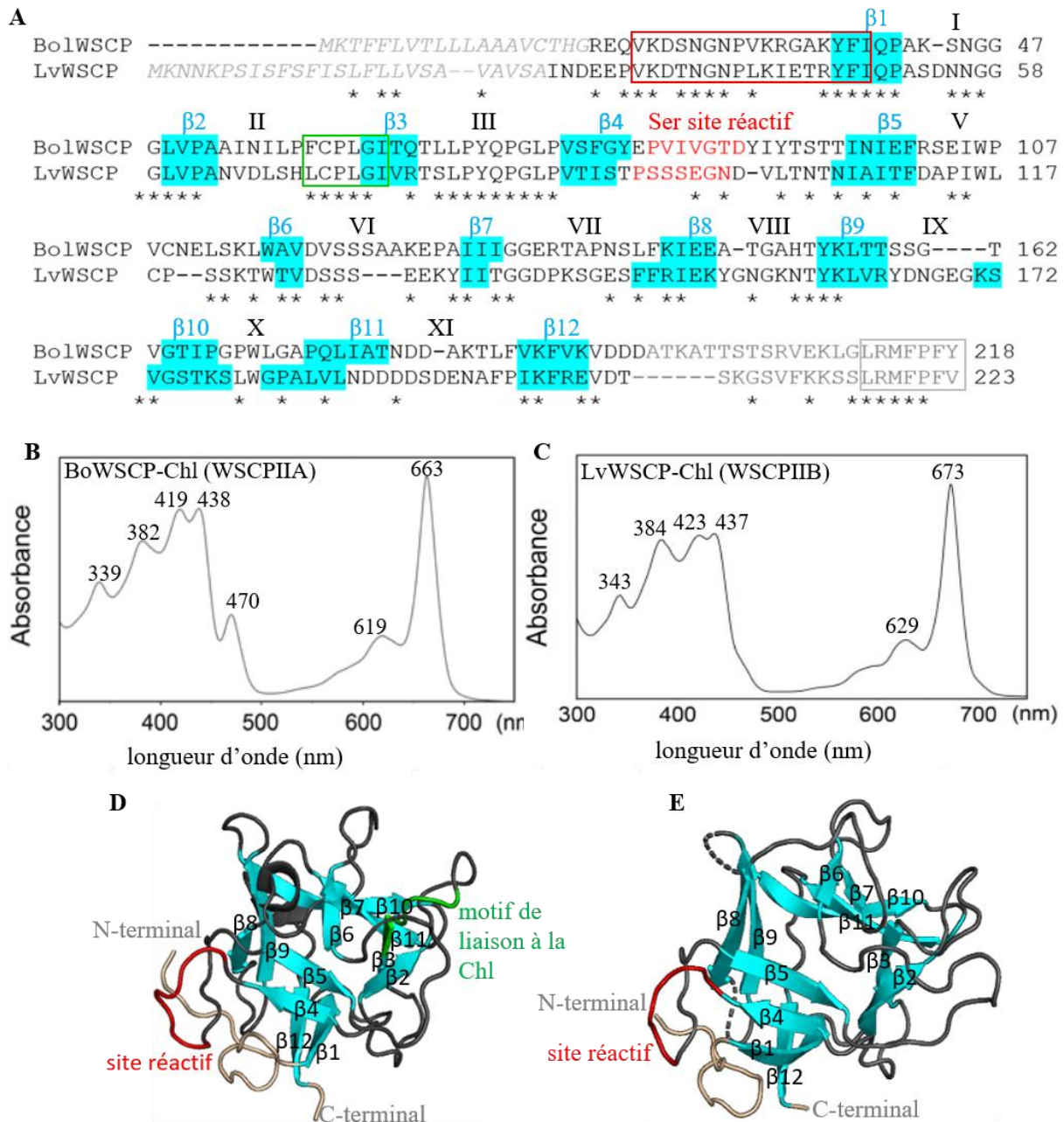
Identifiées pour la première fois dans les années 1960-70 (Yakushiji *et al.*, 1963; Murata *et al.*, 1971), les WSCP « Water-Soluble Chlorophyll Proteins » constituent une famille de protéines de 19-25 kDa non associées aux thylakoïdes qui ont la capacité à se lier à la chlorophylle (Chl), aptitude atypique de par leur caractère hydrosoluble. Leur affinité aux Chla ou Chlb est telle qu'elles sont capables d'extraire ces pigments des membranes de thylakoïdes isolées et de s'assembler en complexes multimériques avec les Chls (Satoh *et al.*, 2001; Maleeva *et al.*, 2019). Les WSCP peuvent également se lier à des dérivés chlorophylliens mais sont incapables de se lier aux caroténoïdes (Satoh *et al.*, 2001; Schmidt *et al.*, 2003; Bektas *et al.*, 2012; Takahashi *et al.*, 2013c, 2014b, 2015a; Palm *et al.*, 2018; Girr *et al.*, 2020; Girr et Paulsen, 2021). Malgré ce lien aux chlorophylles, les WSCP, par ailleurs identifiées que chez un nombre restreint d'espèces comme les Chenopodiaceae, Amaranthaceae, Polygonaceae et Brassicaceae, n'interviennent pas dans le processus photosynthétique.

Sur la base de leurs propriétés photochimiques, les WSCP sont scindées en deux classes : les Chenopodium-type (classe I) et les Brassica-type (classe II) (Satoh *et al.*, 2001). Les protéines des classes I et II présentent très peu de similitude entre elles (Fig. 1A et Fig. 2A) : les WSCPI appartiennent à la famille des protéines DUF538 (domain of unknown function 538) dont les fonctions biologiques sont inconnues (Takahashi *et al.*, 2013c) et les WSCPII sont membres de la famille des IP de type Kunitz (Satoh *et al.*, 2001). Il faut noter que les protéines WSCP sont codées par des gènes sans intron. Quelle que soit la classe, les WSCP sont des protéines très stables face à la dénaturation thermique et aux pH extrêmes (Bektas *et al.*, 2012; Takahashi *et al.*, 2013c; Bednarczyk *et al.*, 2015).

#### I.1. Complexes WSCPI-chlorophylle photoconvertibles

Initialement découvertes chez *Chenopodium album* (Fig. 1, Yakushiji *et al.*, 1963), les WSCPI se retrouvent dans d'autres genres comme *Atriplex*, *Polygonum*, *Chenopodium*, *Amaranthus* (Takamiya, 1973; Goedheer et Vos, 1977) où elles sont préférentiellement localisées dans les tissus en croissance comme les tiges et les feuilles (Takamiya, 1973). Leur fonction physiologique est à ce jour inconnue, mais les WSCPI pourraient être impliquées dans la synthèse et/ou la dégradation de la Chl (Gholizadeh, 2018). Également appelées CP (Chlorophyll Protein), elles se caractérisent par un changement de leur spectre d'absorption après illumination à la lumière blanche, avec une diminution caractéristique de l'absorbance à 688 nm (forme CP688) et l'apparition d'un nouveau pic à 740 nm





**Fig. 2 : Séquences et caractéristiques spectrales et structurales de BolWSCP et LvWSCP**

**A)** Alignement des séquences protéiques de BolWSCP de *B. oleracea* (Q7GDB3) et LvWSCP de *L. virginicum* (O04797). Les brins  $\beta$  (1 à 12) sont surlignés en bleu et les boucles sont positionnées (I à XI) (Horigome *et al.*, 2007; Bednarczyk *et al.*, 2016). La séquence signal est en gris italique et l'extrémité C-terminale subissant un clivage post-traductionnel en gris. Les motifs IP Kunitz et de liaison à la Chl, caractéristiques des WSCP II, sont encadrés respectivement en rouge et en vert. Les résidus en rouge indiquent le site réactif pour l'activité anti-Ser protéolytique. La zone C-terminale hautement conservée chez les WSCP II est encadrée en gris. Les astérisques indiquent les résidus conservés.

**B-C)** Spectres d'absorption de complexes natifs WSCP-Chl purifiés à partir des feuilles (d'après Takahashi *et al.*, 2012, 2013a).

**D-E)** Comparaison de la structure 3D du monomère LvWSCP (D, Horigome *et al.*, 2007) avec celle de STI (soybean trypsin inhibitor) (E, PDB: 1BA7, De Meester *et al.*, 1998). Les codes couleur sont identiques à ceux utilisés en A.

(forme CP740) (Fig. 1B, exemple de CaWSCP de *Chenopodium album*, Yakushiji *et al.*, 1963; Ohtsuki *et al.*, 2007). Dans le complexe WSCP-Chl, après illumination, la Chl passe à un état excité et réagit avec l'oxygène pour former des espèces réactives qui vont oxyder le noyau pyrrolique « B » de la Chl entraînant à terme un changement structural de la Chl et une liaison supplémentaire avec un résidu acide aminé ce qui provoque la photoconversion de la Chl (Hirabayashi *et al.*, 2006). Takahashi *et al.* (2013b) ont montré par reconstitution *in vitro* des complexes CaWSCP-Chl que le processus de photoconversion était spécifique à la Chla (Fig. 1C-D). Au niveau structural, CaWSCP peut subir une modification post-traductionnelle avec un clivage en C-terminal et les deux résidus cystéine (Cys) sont essentiels pour la formation du tétramère de CaWSCP avec 4 Chl (Takahashi *et al.*, 2014a) (Fig. 1A et 1E). Dans ce tétramère, seules trois Chls seraient photo-converties (Takahashi *et al.*, 2013c, 2014a). CaWSCP de *Chenopodium album* est, à ce jour, la seule WSCPI qui a été étudiée de manière plus spécifique pour ses propriétés fonctionnelles.

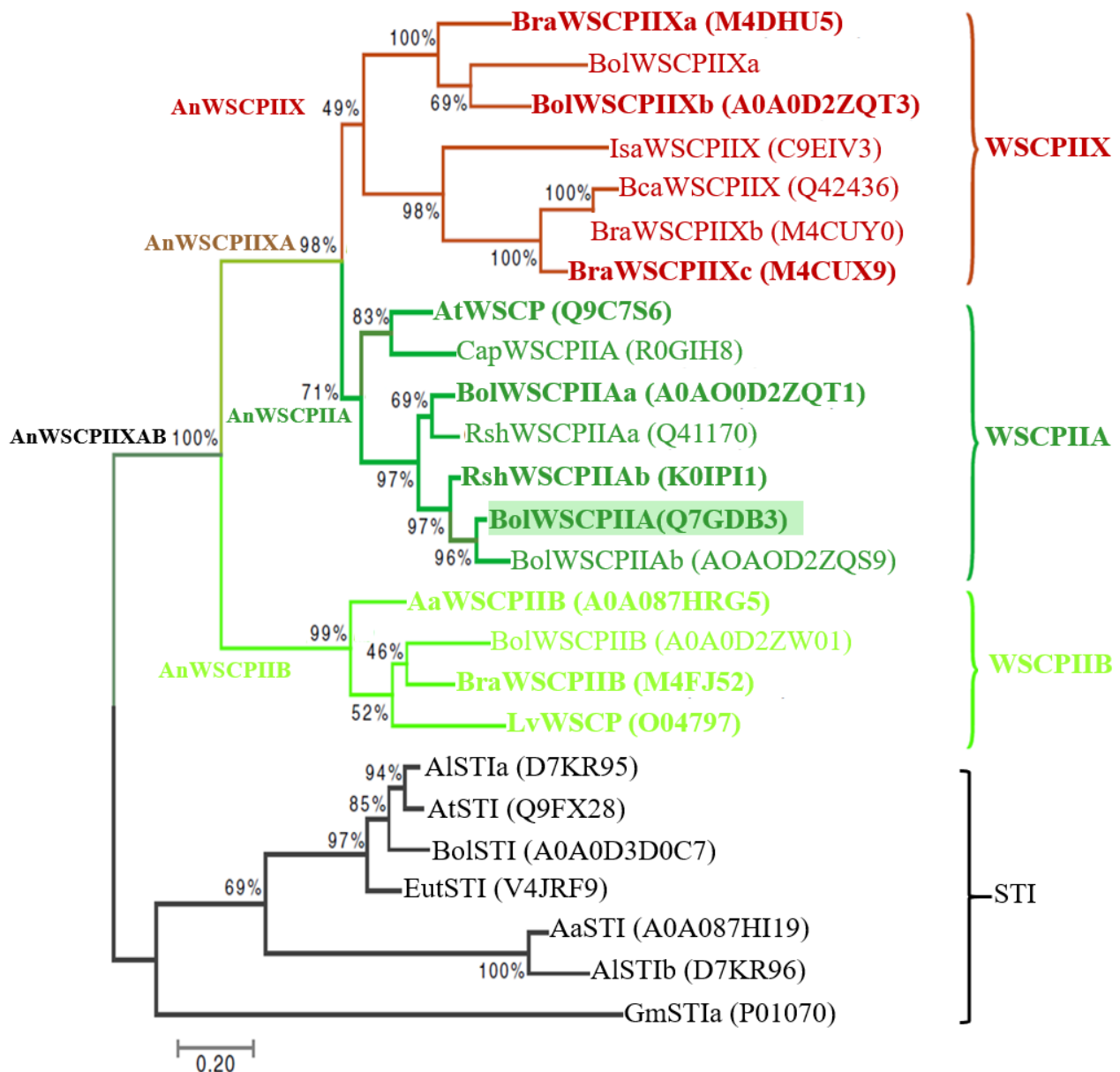
## **I.2. Complexes WSCPII-chlorophylle non-photoconvertibles**

Les WSCPII, qui ont fait l'objet de très nombreuses études, sont au centre de ce travail de thèse. Spécifiques des Brassicacées, les WSCPII possèdent un motif de liaison à la chlorophylle dérivé de celui retrouvé dans les protéines membranaires du photosystème LHCII (Light Harvesting Chlorophyll a/b Binding proteins, [F/Y]-D-P-L-G-L) (Green et Kühlbrandt, 1995; Satoh *et al.*, 2001) (Fig. 2A). En dehors de ce motif, il n'y a pas d'autre similarité de séquence entre WSCPII et LHCII (Satoh *et al.*, 2001). Les WSCPII ne présentent aucun changement spectral après illumination mais des différences dans les spectres et les longueurs maximales d'absorption entre les WSCPII reflètent des différences de teneurs en Chla et Chlb associées aux WSCPII (Fig. 2B-C) (Murata *et al.*, 1971; Shinashi *et al.*, 2000; Satoh *et al.*, 2001; Horigome *et al.*, 2003).

### **I.2.1. Existence de deux sous-classes de WSCPII**

Les WSCPII sont subdivisées en 2 sous-classes selon les ratios Chla/Chlb retrouvés dans les complexes natifs de WSCPII (Satoh *et al.*, 2001) :

- la classe IIA ou Brassica-WSCP forme des complexes enrichis en Chla avec des rapports en Chla/Chlb qui varient entre 6,3 et 10, ce qui est supérieur au ratio Chla/Chlb des feuilles (autour de 3). Les protéines principalement étudiées sont BolWSCP de *Brassica oleracea* (UniProt: Q7GDB3), BnD22 de *Brassica napus* (UniProt: Q43395), AtWSCP d'*Arabidopsis thaliana* (UniProt: Q9C7S6) et RshWSCP de *Raphanus sativus* (UniProt: K0IPI1) (Murata *et al.*, 1971; Murata & Ishikawa, 1981; Ilami *et al.*, 1997; Shinashi *et al.*, 2000; Satoh *et al.*, 2001; Horigome *et al.*, 2003; Halls *et al.*, 2006).
- la classe IIB ou Lepidium-WSCP retrouvée chez *Lepidium virginicum* (LvWSCP, UniProt: O04797) est caractérisée par une proportion plus importante en Chlb avec des rapports Chla/Chlb compris entre 1 et 3,5 (Murata et Ishikawa, 1981). La protéine LvWSCP est la protéine modèle WSCPIIB.



**Fig. 3 : Analyse phylogénétique des WSCPII** (d'après Prabahar *et al.*, 2020)

18 protéines WSCPII identifiées par homologie avec BolWSCPIIA (WSCPII de *B. oleracea*), surlignée en vert, ont été utilisées pour l'analyse phylogénétique avec comme ancrage des inhibiteurs de protéases à sérine de type STI. Les codes UniProt des protéines sont indiqués entre parenthèses. Les protéines, incluant les protéines ancestrales de séquences reconstituées (An), pour lesquelles les capacités de liaison aux Chla et Chlb ont été déterminées, sont indiquées en gras.

Les ratios différents en Chls au sein des complexes des deux sous-classes de WSCPII suggèrent fortement une diversité d'affinité ou de sélectivité des WSCPIIA et des WSCPIIB vis-à-vis des Chls: les WSCPIIA seraient donc plus affines à la Chla alors que les WSCPIIB seraient plus affines à la Chlb.

### **I.2.2. Vers l'existence d'une troisième sous-classe**

Récemment, des analyses phylogénétiques et spectroscopiques menées par Prabahar *et al.* (2020) sur les WSCPII ont conduit à l'identification de nouveaux membres WSCPIIA et WSCPIIB (chez *B. rapa*, *B. oleracea*, *Arabis alpina*, *Capsella rubella* et *Isatis tinctoria*) et d'une nouvelle classe appelée WSCPIIX qui a perdu sa capacité à se lier aux Chls (Fig. 3). D'un point de vue évolutif, les WSCPIIA partagent un ancêtre commun avec les WSCPIIX, alors que les WSCPIIB sont distinctes de ces deux classes. Pour explorer davantage la trajectoire évolutive de la propriété de liaison aux Chls, des archétypes ancestraux (An) des WSCPII : AnIIA, AnIIB, AnIIX, AnIIXA, AnIIXAB ont été générés et testés pour leur capacité et leur affinité d'interaction avec les Chls. Deux événements évolutifs des WSCPII ont été ainsi suggérés. Le premier événement est responsable de l'affinité à la Chla ou à la Chlb comme en témoignent les affinités des archétypes AnIIXA et AnIIB par rapport à leur ancêtre commun AnIIXAB. Seuls les descendants (AnIIA et AnIIB) mais pas leur ancêtre AnIIXAB sont capables de se lier à la Chlb suggérant que cette capacité a été acquise plus tard dans l'évolution. Le deuxième événement évolutif serait responsable de la perte de la capacité de liaison ce qui a donné lieu à la classe IIX. Les propriétés potentielles d'inhibiteur de protéases (IP) n'ont pas été suivies dans cette étude.

### **I.2.3. Membres de la famille des inhibiteurs de type Kunitz**

Les WSCPII sont classées parmi les inhibiteurs de protéases de type Kunitz caractérisés par la présence du motif ([LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM]) (Satoh *et al.*, 2001) (Fig. 2A). Elles sont regroupées dans le sous-groupe IC 3A selon la classification MEROPS (Rawlings *et al.*, 2004). Comme les autres IP de type Kunitz, les WSCPII sont caractérisées par la présence de 12 feuillets  $\beta$  antiparallèles reliés par de longues boucles formant une structure en tonneau (Fig. 2D-E) (Song et Suh, 1998; Khamrui *et al.*, 2005; Horigome *et al.*, 2007). Les IP de type Kunitz sont les plus répandus dans le règne végétal et sont capables d'inhiber les protéases à sérine, à cystéine et à aspartate (Mosolov et Valueva, 2005). Au regard de ces éléments il est facilement envisageable que les WSCPII pourraient exercer une fonction inhibitrice vis-à-vis de certaines protéases.

**Tab. 1 : Identités entre les WSCP II identifiées par Prabahar *et al.* (2020)**

Les pourcentages d'identité (en blanc) et les distances entre paires de séquences (en gris) sont issus d'alignements de WSCP II dont l'activité de liaison aux Chl a été déterminée. BnD22, WSCP IIA de *B. napus*, est indiquée en supplément. Aa, *Arabis alpina*; At, *Arabidopsis thaliana*; Bra, *B. rapa*; Bol, *B. oleracea*; Lv, *Lepidium virginicum*; Rsh, *Raphanus sativus*

		BraWSCP IIXa	BolWSCP IIXb	BraWSCP IIXc	BnD22	AtWSCP	BolWSCP IIAa	RshWSCP IIAb	BolWSCP IIA	AaWSCP IIB	BraWSCP IIB	LvWSCP
WSCP IIX	BraWSCP IIXa		0,45	0,83	0,77	0,70	0,68	0,75	0,75	1,09	1,01	0,99
	BolWSCP IIXb	62,84		0,72	0,69	0,63	0,70	0,72	0,67	0,97	0,98	0,88
	BraWSCP IIXc	41,27	45,79		0,66	0,58	0,72	0,66	0,61	0,89	0,87	0,88
WSCP IIA	BnD22	35,09	38,60	41,52		0,41	0,27	0,16	0,08	0,84	0,77	0,80
	AtWSCP	46,56	50,26	55,14	52,68		0,37	0,43	0,43	0,77	0,77	0,83
	BolWSCP IIAa	46,32	45,79	48,11	61,54	66,13		0,21	0,24	0,82	0,76	0,78
	RshWSCP IIAb	43,39	44,97	49,46	70,18	62,70	78,14		0,14	0,82	0,77	0,78
	BolWSCP IIA	43,39	47,09	52,15	76,15	62,70	75,96	86,59		0,82	0,77	0,77
WSCP IIB	AaWSCP IIB	30,26	34,54	36,73	32,33	41,75	38,78	39,38	39,38		0,36	0,45
	BraWSCP IIB	32,98	34,74	38,02	34,65	42,63	41,67	41,80	41,80	67,39		0,50
	LvWSCP	32,31	36,60	36,22	32,76	38,66	39,29	39,90	40,41	61,96	57,61	

**Tab. 2 : Caractéristiques physico-chimiques et localisation subcellulaire des WSCP II identifiées par Prabahar *et al.* (2020)**

Les WSCP II ciblées sont celles dont l'activité de liaison aux Chl a été déterminée. BnD22, WSCP IIA de *B. napus*, est indiquée en supplément. Les caractéristiques physico-chimiques sont prédites par EXPASY (<https://web.expasy.org/protparam/>) et la localisation subcellulaire par wolf-psortII (<https://www.genscript.com/wolf-psort.html?src=leftbar>). At, *Arabidopsis thaliana*; aa= acide aminé; Aa, *Arabis alpina*; Bra, *B. rapa*; Bol, *B. oleracea*; chlo, chloroplaste; extr, extracellulaire; Lv, *Lepidium virginicum*; Rsh, *Raphanus sativus*

sous-classe	nom de la protéine	propriétés physico-chimiques						localisation sub-cellulaire (kNN score)
		séquence protéique primaire		protéine sans peptide signal				
		longueur (aa)	longueur peptide signal (aa)	longueur (aa)	PM (kDa)	point iso-électrique	index aliphatique	
WSCP IIX	BraWSCP IIXa	213	23	190	20,52	6,55	99,47	extr: 8
	BolWSCP IIXb	218	19	199	21,47	8,13	82,58	extr: 8
	BraWSCP IIXc	226	23	202	22,19	5,28	79,65	extr: 7
WSCP IIA	BnD22	218	19	199	21,52	5,27	88,69	extr: 6
	AtWSCP	215	23	192	20,62	8,81	84,79	extr: 6
	BolWSCP IIAa	222	20	202	21,87	6,92	90,25	extr: 10
	RshWSCP IIAb	222	23	199	21,45	8,47	92,56	extr: 7
	BolWSCP IIA	218	19	199	21,47	7,90	87,24	extr: 6
WSCP IIB	AaWSCP IIB	205	19	186	20,40	5,41	73,28	extr: 4
	BraWSCP IIB	222	23	199	21,83	4,90	72,36	chlo: 3
	LvWSCP	223	27	197	21,50	4,87	73,65	extr: 6

#### **I.2.4. Maturation post-traductionnelle**

Toutes les protéines WSCP<sub>II</sub> possèdent un peptide signal les adressant au réticulum endoplasmique (Fig. 2A, Tab. 2, Satoh *et al.*, 2001) mais leur localisation subcellulaire reste encore sujette à discussion. En plus du peptide signal, comme pour les WSCP<sub>I</sub>, les WSCP<sub>II</sub> possèdent une région C-terminale d'environ 10-22 acides aminés qui serait éliminée par clivage post-traductionnel (Ilami *et al.*, 1997; Takahashi *et al.*, 2012) (Fig. 2A). Cette extrémité C-terminale qui contient une séquence de 7 acides aminés, L[K/R]MFPFY, hautement conservée dans les WSCP<sub>II</sub> (Takahashi *et al.*, 2013b) (Fig. 2A), jouerait un rôle dans la stabilité/solubilité de la protéine sans affecter l'activité des protéines (Takahashi *et al.*, 2014b).

#### **I.2.5. Diversité d'identités et de propriétés physico-chimiques**

Malgré la présence des deux motifs caractéristiques des WSCP<sub>II</sub>, celles-ci présentent une grande diversité avec des identités en acides aminés allant de 30 à 86% et des scores d'alignement de 0,14 à 1,09 (Tab. 1). Les WSCP<sub>IIA</sub> ressemblent plus aux WSCP<sub>IIIX</sub> (35-50% d'identité) qu'aux WSCP<sub>IIIB</sub> (32-40% d'identité). Au sein de chaque groupe phylogénique, une diversité protéique est présente avec au maximum 86% d'identité pour les WSCP<sub>IIA</sub>, et 63-67% pour les WSCP<sub>IIIX</sub> et WSCP<sub>IIIB</sub>. Les WSCP<sub>II</sub> ont des tailles similaires autour de 20-21 kDa et sont extrêmement stables (index aliphatique supérieur à 80%, Tab. 2). Elles présentent une diversité de leurs caractéristiques physico-chimiques avec des protéines acides ou basiques. Il est à noter que la majorité d'entre elles présente un score kNN élevé pour leur adressage au système de sécrétion (Tab. 2).

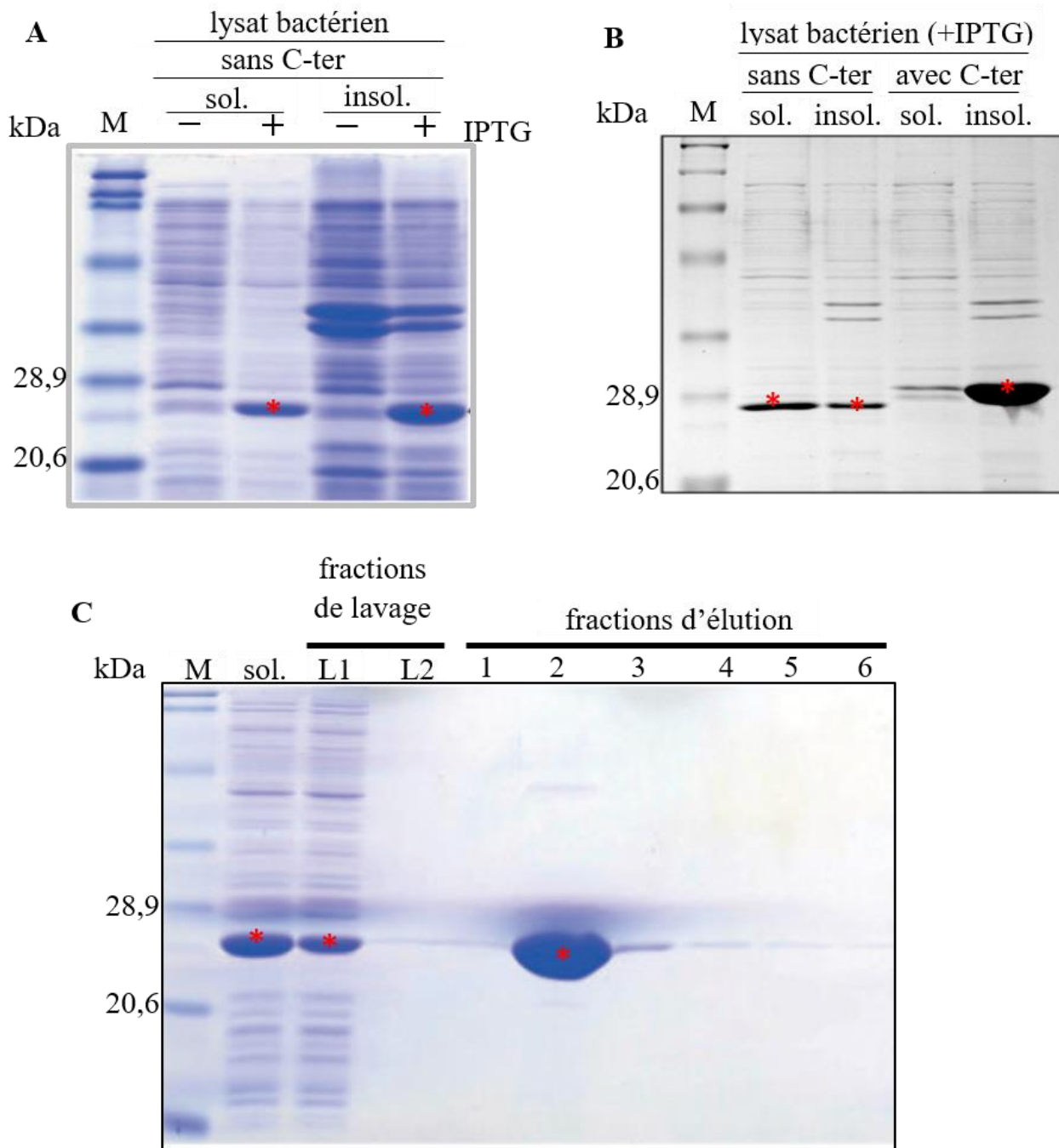
Quelques travaux suggèrent la présence de familles multigéniques WSCP<sub>II</sub> chez les Brassica. Ainsi, dans les feuilles de colza, la purification de BnD22 est concomitante avec d'autres isoformes protéiques de masses moléculaires très proches (Ilami *et al.*, 1997) et plusieurs isoformes protéiques WSCP<sub>II</sub> autour de 19 kD avec des points isoélectriques de 5,0 -5,1 ont été mises en évidence (Desclos *et al.*, 2008). De plus, Prabahar *et al.* (2020) ont identifié plusieurs WSCP<sub>II</sub> de *B. oleracea* et *B. rapa*. Par contre, *A. thaliana* possède une copie unique WSCP<sub>II</sub>.

## **II. Propriétés fonctionnelles des WSCP<sub>II</sub> : interaction avec les Chls**

### **II.1. Approches méthodologiques**

#### **II.1.1. Purification des complexes natifs WSCP<sub>II</sub>-Chls à partir des feuilles**

Certaines WSCP<sub>II</sub>, comme LvWSCP et BolWSCP, sont associées aux Chls lors de leur purification à partir d'extraits de feuilles. Leur purification est obtenue par précipitation au sulfate d'ammonium et chromatographies échangeuses d'ions et d'exclusion de taille (Murata *et al.*, 1971;



**Fig. 4 : Résultats d'analyse par PAGE-SDS de protéines recombinantes WSCP II exprimées dans un système bactérien : exemple de BolWSCP (d'après Takahashi *et al.*, 2012, 2014)**

**A)** Fractions solubles (sol.) et insolubles (insol.) obtenues à partir de lysats de bactéries exprimant la BolWSCP recombinante avec son C-terminal, induite (+) ou non (-) par l'IPTG

**B)** Comparaison des fractions solubles (sol.) et insolubles (insol.) obtenues à partir de lysats de bactéries exprimant la WSCP II recombinante avec ou sans son C-terminal après induction par l'IPTG

**C)** Etapes de purification de la BolWSCP-His tag par chromatographie d'affinité Ni<sup>+</sup> à partir de la fraction soluble : fractions de lavage (L) et d'éluion (1 à 6). L'éluion est effectuée en présence de 0,5 M d'imidazole (analogue structural de l'His)

M = marqueur de taille

Murata & Ishikawa, 1981). La purification des complexes BolWSCP-Chl et LvWSCP-Chl peut être également réalisée par élution des complexes à partir d'un gel obtenu après électrophorèse en conditions non dénaturantes (Takahashi *et al.*, 2012, 2013b). Etant donné le caractère thermostable des WSCP, une étape de dénaturation à 90°C d'extraits protéiques foliaires a également été proposée pour la purification de BolWSCP (Kamimura *et al.*, 1997; Palm *et al.*, 2017). Quel que soit le procédé, les complexes extraits *in vivo* sont hétérogènes avec différentes proportions de Chla et de Chlb qui, de plus, varient selon le mode d'extraction (Murata *et al.*, 1971). Les ratio Chla/Chlb des complexes WSCP-Chl natifs sont estimés en déterminant les teneurs en Chls des complexes après extraction par des solvants organiques (butanol, acétone et éther) (Murata *et al.*, 1971; Palm *et al.*, 2017). Dans ces conditions, les WSCP natives se dissocient des Chls sous forme d'apoprotéines qui, après purification, peuvent être utilisées ultérieurement pour la reconstitution *in vitro* de complexes WSCP-Chls avec des teneurs précises en Chla ou en Chlb ou en mélange Chla+Chlb (Murata *et al.*, 1971).

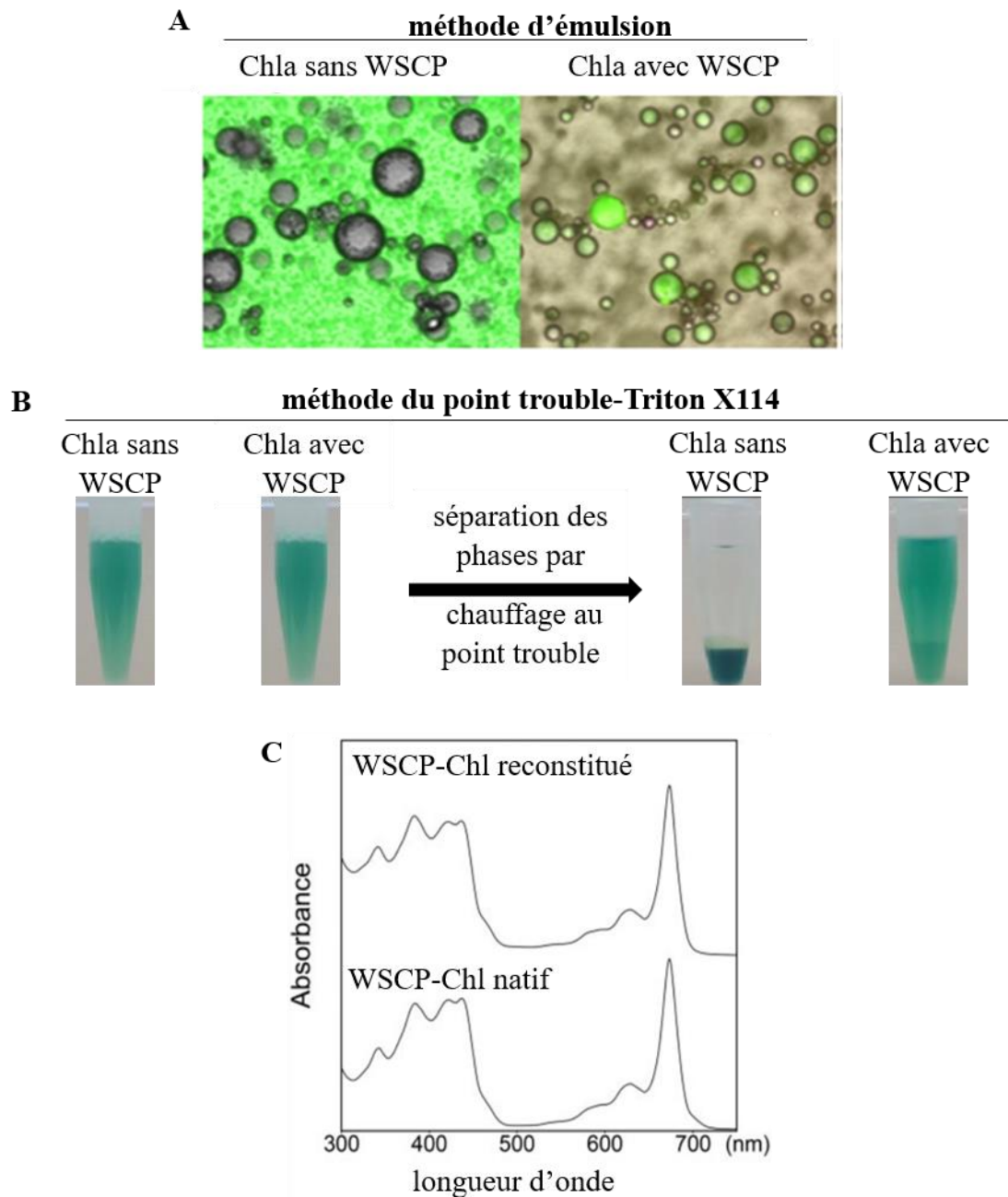
### ***II.1.2. Reconstitution in vitro des complexes WSCP-Chls***

La caractérisation des interactions WSCP-pigment peut être approfondie par reconstitution *in vitro* des complexes WSCP-Chl en jouant sur les proportions en pigments ainsi que sur le ratio WSCP/pigment pendant l'assemblage. Bien que séduisante, cette méthode se heurte à des difficultés d'ordre technique : i) l'obtention à partir des tissus de protéines purifiées sous forme de monomère ou la production de protéines recombinantes fonctionnelles, ii) l'assemblage des complexes *in vitro* étant donné les caractères hydrophobes des pigments et hydrophile des WSCP, iii) l'obtention de complexes dont les propriétés de liaison sont similaires à celles des complexes natifs.

#### ***II.1.2.a Production de protéines recombinantes WSCP***

En substitution des extractions tissulaires des WSCP, les protéines peuvent être produites artificiellement par expression hétérologue sous forme d'apoprotéines monomériques dans un système bactérien (Fig. 4A-B). Les protéines recombinantes sont taguées à leur extrémité N- ou C-terminale le plus souvent avec une étiquette poly-histidine (His-Tag), pour permettre leur purification par chromatographie d'affinité (Satoh *et al.*, 1998; Schmidt *et al.*, 2003; Takahashi *et al.*, 2012) (Fig. 4C). L'un des problèmes majeurs rencontré est la proportion de protéines agrégées dans des corps d'inclusion sous forme insoluble et donc inexploitable sauf par extraction dénaturante suivie d'une étape de renaturation des protéines. A titre d'exemple, environ 70 % des protéines recombinantes BolWSCP exprimées dans *E. coli* sont sous forme insoluble (Takahashi *et al.*, 2012) (Fig. 4A-B). Toutefois, la production de WSCP recombinantes dépourvues de leur extrémité C-terminale clivée lors de leur maturation post-traductionnelle, améliore leur solubilité et leur stabilité sans affecter leur activité (Fig. 4B, Takahashi *et al.*, 2014b). Par ce système de production, le rendement en protéines





**Fig. 5 : Reconstitution *in vitro* des complexes WSCP-Chl : exemple de BolWSCP-Chla**

**A)** Méthode d'émulsion (d'après Bednarczyk *et al.*, 2015). La Chla est ajoutée sur l'émulsion (mélange d'huile-détergent-eau) contenant ou non la WSCP. Les images sont obtenues par microscopie photonique. Les bulles vertes de l'émulsion indiquent la reconstitution de complexes WSCP-Chl.

**B)** Méthode du point trouble (d'après Palm *et al.*, 2018). La WSCP est ajoutée au mélange de reconstitution contenant la Chla et du TritonX114 (TX114). Après incubation, les deux phases sont séparées par chauffage à une température supérieure au point trouble du détergent. La phase supérieure aqueuse contient la WSCP liée ou non à la Chl et la phase inférieure contient le TX114 et la Chl non liée. La couleur verte de la phase supérieure indique la reconstitution des complexes WSCP-Chl.

**C)** Spectres d'absorption des complexes WSCP-Chl natifs purifiés à partir des feuilles, et reconstitués à partir de la protéine recombinante en présence des membranes de thylakoïdes (d'après Takahashi *et al.*, 2012).

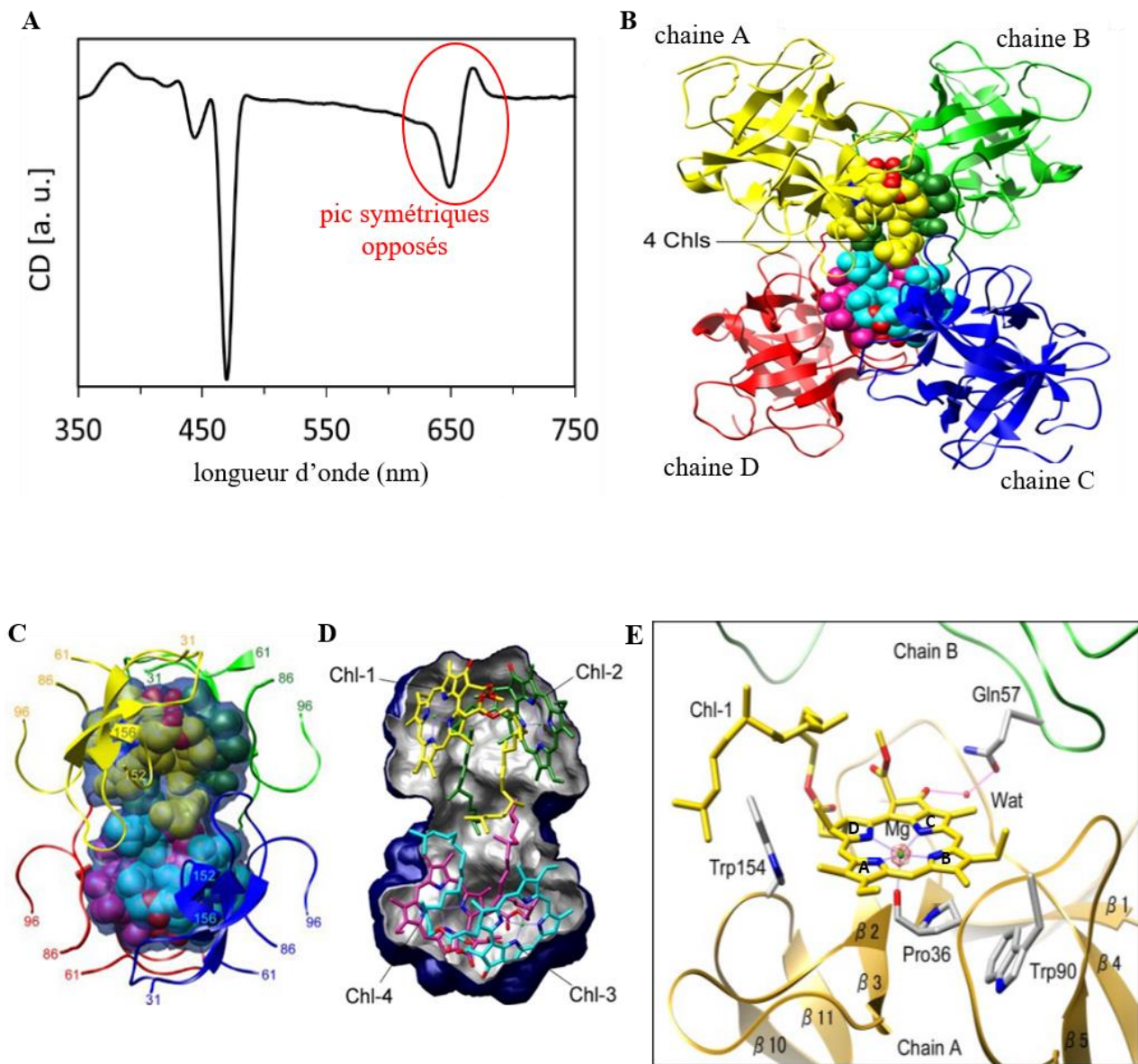
recombinantes purifiées est généralement satisfaisant mais est variable selon la nature des protéines ciblées (Sato *et al.*, 1998; Schmidt *et al.*, 2003; Takahashi *et al.*, 2012, 2014b; Bednarczyk *et al.*, 2015). Plusieurs méthodes de reconstitution *in vitro* des complexes WSCP-II-pigments ont été développées pour optimiser l'assemblage entre les apoprotéines et les pigments (Sato *et al.*, 1998; Bednarczyk *et al.*, 2015; Palm *et al.*, 2017, 2018).

#### **II.1.2.b Assemblage avec des pigments présents dans les membranes des thylakoïdes**

Une des méthodes de reconstitution la plus proche des conditions natives repose sur la capacité des apoprotéines à extraire des pigments membranaires de thylakoïdes préparés à partir d'extraits foliaires (Sato *et al.*, 1998; Hughes *et al.*, 2006). Les complexes ainsi reconstitués ont des proportions Chl<sub>a</sub>/Chl<sub>b</sub> similaires à celles des complexes natifs. Toutefois, en raison de la présence des deux Chls dans les thylakoïdes, cette méthode ne permet pas la formation de complexes homogènes spécifiques Chl<sub>a</sub> ou Chl<sub>b</sub> et est donc inadaptée pour suivre les affinités relatives à chacune des Chls. De plus, la reconstitution de WSCP-II avec d'autres pigments que les Chls est impossible car cette méthode est limitée aux Chl<sub>a</sub> et Chl<sub>b</sub> présentes dans les membranes (Bednarczyk *et al.*, 2015; Palm *et al.*, 2018).

#### **II.1.2.c Assemblage avec des pigments purifiés en présence de solvants organiques, d'émulsion huile-détergent-eau ou de détergents**

La reconstitution de complexes WSCP-II-Chls a été validée en présence de solvants organiques tels que l'éthanol à 20% (Palm *et al.*, 2017) ou le méthanol à 40% (Takahashi *et al.*, 2012), en présence d'émulsion huile-détergent-eau (Bednarczyk *et al.*, 2015 ; Fig. 5A) ou en présence de détergents comme l'octyl glucopyranoside (Schmidt *et al.*, 2003), le lauryl  $\beta$ -D-maltoside (Bektas *et al.*, 2010) ou le Triton X114 (TX114, Fig. 5B, Palm *et al.*, 2018). Les concentrations en solvant ou en détergent sont minimisées dans le milieu réactionnel ce qui peut conduire à l'agrégation partielle des Chls si des concentrations importantes sont nécessaires (Bednarczyk *et al.*, 2015; Palm *et al.*, 2018). La méthode d'émulsion (Fig. 5A) élimine le problème d'agrégation des Chls et permet un rapport surface/volume élevé facilitant le contact entre apoprotéine et pigment mais la préparation de l'émulsion est assez délicate (Bednarczyk *et al.*, 2015). Une méthode alternative a été mise au point par Palm *et al.* (2018) qui repose sur la reconstitution des complexes en présence de Triton X114 et la séparation des phases solubles contenant les Chls liées aux WSCP-II, de la phase organique contenant les Chls libres après incubation à une température supérieure au point trouble du détergent (Fig. 5B). Ces méthodes ont l'avantage de permettre la formation de complexes homogènes soit avec un seul type de Chl, soit en jouant sur les ratio Chl<sub>a</sub>/Chl<sub>b</sub> ou les ratio Chl/WSCP-II, soit avec d'autres pigments tels que les précurseurs ou des dérivés de Chl et permettre ainsi d'estimer les affinités des WSCP-II aux différents pigments (Bednarczyk *et al.*, 2015; Palm *et al.*, 2017, 2018).



**Fig. 6 : Structure tridimensionnelle du complexe LvWSCP-Chl de *Lepidium virginicum* (d'après Horigome *et al.*, 2007)**

**A)** Spectre de dichroïsme circulaire (CD) de LvWSCP-Chl natif avec la présence d'un signal spécifique de deux pics symétriques opposés indiquant l'arrangement global du complexe sous forme de deux dimères (d'après Palm *et al.*, 2018).

**B)** Organisation du tétramère. Les chaînes A, B, C et D forment une cavité hydrophobe centrale dans laquelle sont présentes 4 molécules de chlorophylle (4 Chls) représentées par des sphères de couleur.

**C)** Représentation de la cavité hydrophobe formée par les résidus 31 à 61, 86 à 96 et 152 à 156 de chaque sous-unité.

**D)** Représentation en coupe de la cavité centrale du complexe tétramérique contenant les 4 molécules de chlorophylle. Les 4 Chls (en structure) suivent la symétrie du tétramère en deux dimères contenant chacune une paire de Chls (Chl-1 avec Chl-2 et Chl-3 avec Chl-4).

**E)** Interactions entre le tétramère et la Chl : les acides aminés primordiaux pour l'interaction et les noyaux tétrapyrroliques A, B, C et D sont indiqués.

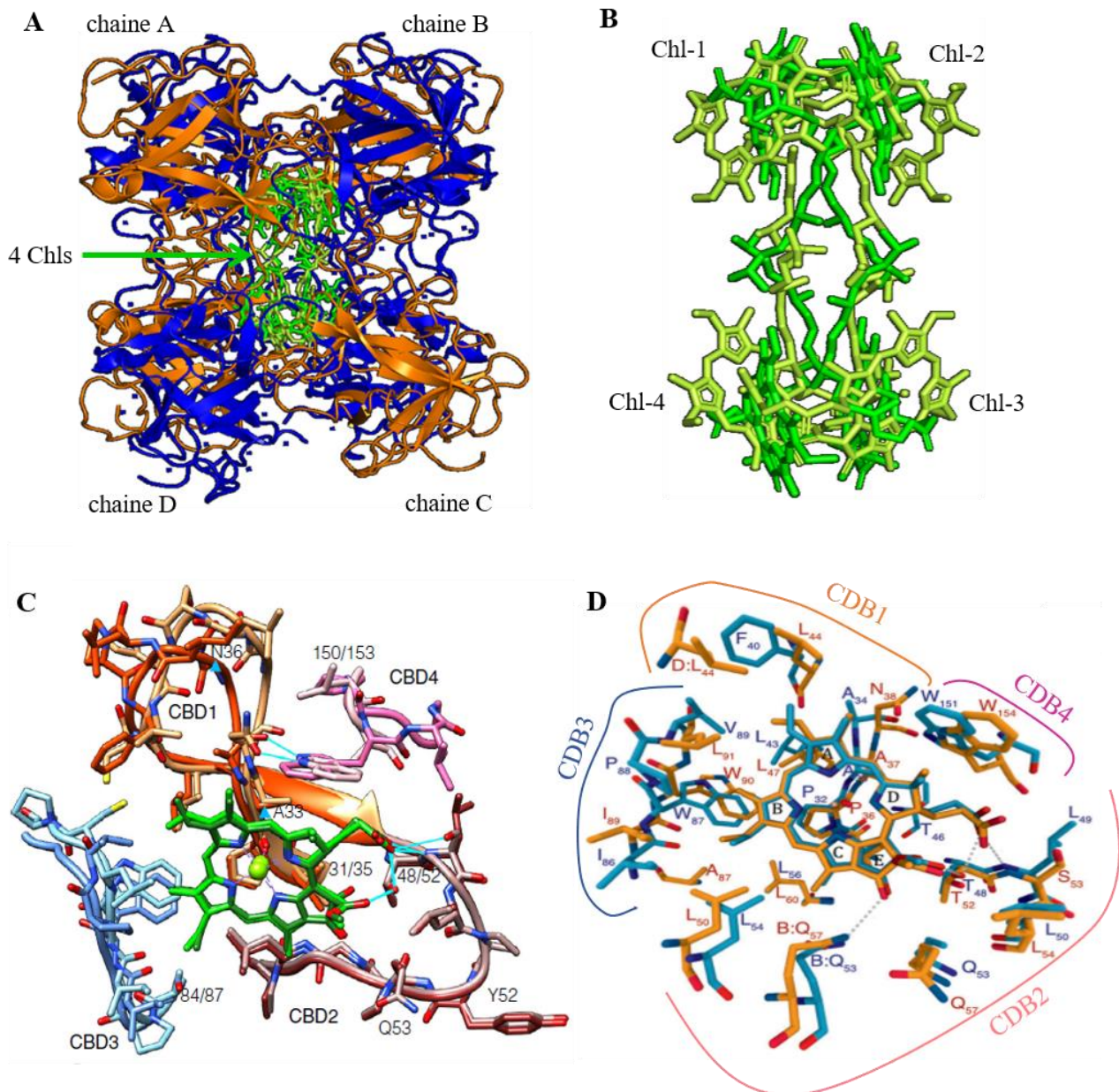
### **II.1.2.d Purification des complexes WSCP<sub>II</sub>-Chls reconstitués**

Pour séparer les complexes WSCP<sub>II</sub>-Chls des apoprotéines monomériques, des étapes de purification sont nécessaires. Celles-ci sont identiques à celles utilisées pour les complexes natifs (chromatographie d'exclusion de taille ou électrophorèse en condition non dénaturante) (Takahashi *et al.*, 2013b; Palm *et al.*, 2017). Les complexes WSCP<sub>II</sub>-Chls obtenus à partir des protéines recombinantes sont très similaires sur le plan biochimique et spectroscopique aux complexes WSCP<sub>II</sub>-Chls natifs purifiés à partir des feuilles (Schmidt *et al.*, 2003; Bektas *et al.*, 2012; Takahashi *et al.*, 2013a; Palm *et al.*, 2017) (Fig. 5C).

## **II.2. Structuration des complexes WSCP<sub>II</sub>-Chls**

### **II.2.1. Assemblage en tétramères**

En se basant sur les poids moléculaires des protéines et des complexes WSCP<sub>II</sub>-Chls natifs ou reconstitués *in vitro*, l'état oligomérique des complexes sous forme de tétramères a été déterminé (Kamimura *et al.*, 1997; Satoh *et al.*, 1998; Horigome *et al.*, 2003, 2007; Bednarczyk *et al.*, 2015, 2016; Palm *et al.*, 2018). Les structures tridimensionnelles, résolues par cristallographie aux rayons X des complexes contenant LvWSCP, la protéine modèle des WSCP<sub>II</sub>B ou BolWSCP une WSCP<sub>II</sub>A, associées avec la Chl<sub>a</sub> ou la Chl<sub>b</sub> confirment l'association WSCP<sub>II</sub>-Chl sous forme d'homo-tétramère quelle que soit la classe de WSCP<sub>II</sub> et la nature de la Chl associée (Fig. 6) (Horigome *et al.*, 2007; Bednarczyk *et al.*, 2016; Palm *et al.*, 2018; Agostini *et al.*, 2019). L'arrangement global sous forme de dimères des complexes WSCP<sub>II</sub>-Chls peut être suivi par dichroïsme circulaire (CD) avec la présence d'un signal spécifique (Fig. 6A). Le tétramère résulte donc de l'assemblage de 2 dimères, l'un formé par les sous-unités A et B et l'autre par les sous-unités C et D, chacune des sous-unités étant liée à une molécule de Chl (Fig. 6B). Les résidus acides aminés 31–61, 86–96 et 152–156 de chaque sous-unité forment une cavité centrale hydrophobe dans laquelle les 4 Chls sont empaquetées (Fig. 6C) avec les queues phytol au centre et les noyaux tétrapyrroliques associés deux à deux, face à face dans chaque dimère (Fig. 6D) (Horigome *et al.*, 2007; Palm *et al.*, 2018; Agostini *et al.*, 2019). Les Chls suivent la symétrie du tétramère en formant une structure « open-sandwich » avec un angle d'environ 27° entre les dimères pour la Chl<sub>b</sub> associée à LvWSCP (Hughes *et al.*, 2006; Horigome *et al.*, 2007). Bien que BolWSCP et LvWSCP montrent une capacité de liaison aux Chl<sub>a</sub> ou Chl<sub>b</sub> avec une structure globale similaire (Fig. 7A), il existe des distinctions avec une rotation d'un angle de 60° entre dimères des complexes BolWSCP-Chl<sub>a</sub> et la présence d'acides aminés spécifiques dans l'environnement proche des Chls provoquant une déformation du plan du macrocycle des Chls (Fig. 7B, Agostini *et al.*, 2017; Palm *et al.*, 2018). Il en résulte une différence spectrale caractéristique des deux sous-classes de WSCP<sub>II</sub> avec des longueurs d'onde maximales d'absorption autour de 670 nm



**Fig. 7: Superposition des structures des complexes BolWSCP-Chl et LvWSCP-Chl (A et B) et focus sur les domaines et acides aminés interagissant avec la Chl (C et D).**

A) Superposition des structures globales des complexes BolWSCP-Chl (en bleu) et LvWSCP-Chl (en orange). Les Chls sont représentées en vert.

B) Superposition des 4 Chls du complexe BolWSCP-Chl (en vert foncé) et du complexe LvWSCP-Chl (en vert clair) suivant la symétrie du tétramère.

C) Superposition des 4 domaines de liaison à la Chl des complexes BolWSCP-Chl et LvWSCP-Chl (d'après Prabahar *et al.*, 2020). Le premier résidu acide aminé de chaque CBD est indiqué. Les liaisons hydrogène possibles entre les CBD ou avec la Chl sont présentées en cyan.

D) Superposition des résidus acides aminés en interaction avec la Chl. En orange LvWSCP et en bleu BolWSCP (d'après Bednarczyk *et al.*, 2016). Les CBD sont indiqués avec les mêmes couleurs qu'en A. Les lignes pointillées représentent les liaisons hydrogène.

pour les WSCPIIA et de 660 nm pour les WSCPIIB (Theiss *et al.*, 2007; Bednarczyk *et al.*, 2016).

## **II.2.2. Stabilisation du complexe WSCPII-Chl**

### **II.2.2.a Interaction Chl-protéine dans le tétramère**

Il s'avère que quatre acides aminés sont primordiaux pour l'interaction entre la protéine et le noyau tétrapyrrolique : Pro-36, Gln-57, Trp-90 et Trp-154 (positions dans LvWSP) (Fig. 6E). Ces acides aminés sont conservés pour BolWSCP (Fig. 2A, Fig. 7D). La proline montre une importance toute particulière dans l'interaction avec l'ion  $Mg^{2+}$  au centre du noyau tétrapyrrolique bien que la prévalence de cette interaction soit discutée (Schmidt *et al.*, 2003; Horigome *et al.*, 2007; Girr *et al.*, 2020). Ainsi, des travaux montrent une absence de liaison entre BolWSCP et la phéophytine, un dérivé du catabolisme de la Chl dépourvu de  $Mg^{2+}$  (Schmidt *et al.*, 2003) alors que d'autres études montrent la faisabilité des associations entre WSCPII (BolWSCP et LvWSCP) et la phéophorbide qui est dépourvue à la fois de la chaîne phytol et de l'ion magnésium central, ou encore avec l'hème (Girr *et al.*, 2020). Les deux d'acides aminés aromatiques, Trp-90 et Trp-154, sont en contact avec les C7 (porteur des groupements distincts des Chla et Chlb, annexe 1) et C18 de la Chl par des interactions de van Der Waals (Fig. 6E). Ces interactions seraient responsables du photoblanchiment des Chls dans les complexes après photo-oxydation (Horigome *et al.*, 2007). La Gln-57 interagit avec l'oxygène porté par le C13 du noyau tétrapyrrolique par une liaison hydrogène (Horigome *et al.*, 2007; Bednarczyk *et al.*, 2016; Agostini *et al.*, 2018). Les queues phytol quant à elles génèrent une force hydrophobe avec les sous-unités et les autres queues phytol du tétramère participant ainsi à la stabilisation et à la tétramérisation.

Récemment, Prabahar *et al.* (2020) ont identifié 4 domaines de liaison à la Chl appelés CBD (Chlorophyll Binding Domain) (Fig. 7C-D). Le CBD1 constitué de 17 acides aminés dont la proline liée à l'ion central de la Chl (brins  $\beta$ 1-2 + boucles II et III), a une structure en épingle à cheveux (hairpin). Ce domaine jouerait un rôle dans la différence spectrale entre WSCPIIA et WSCPIIB. Le CBD2 (boucle III + brin  $\beta$ 4), hautement conservé dans les WSCPIIA et WSCPIIB, serait indispensable pour la liaison à la Chl. Le CBD3 (boucle V) pourrait être responsable de la sélectivité de liaison Chla/Chlb et contient le motif d'affinité défini par Palm *et al.* (2018). Le CBD4 de 3 acides aminés contient le Trp responsable de la différence des profils d'absorbance entre WSCPIIA et WSCPIIB (Bednarczyk *et al.*, 2016). Les Gln et Trp identifiés par Horigome *et al.* (2007) se retrouvent respectivement dans les CBD2, 3 et 4.



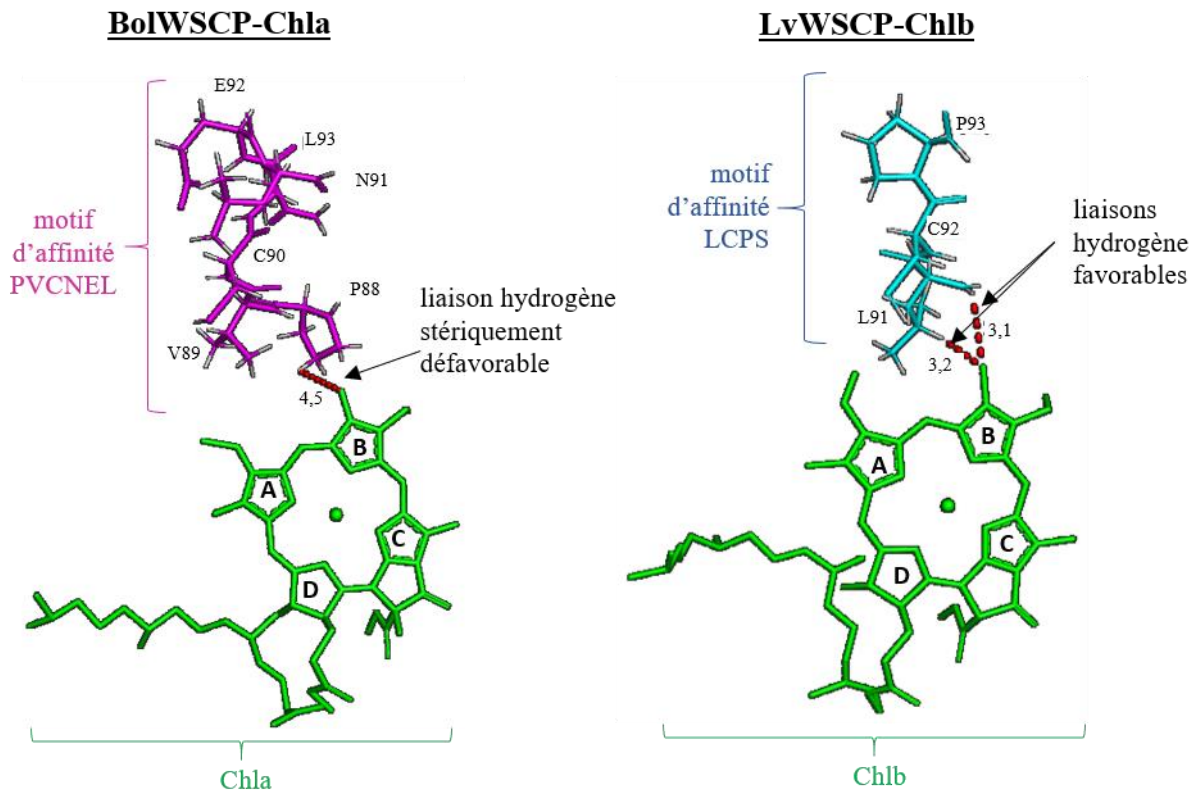
### **II.2.2.b Interaction Chl-protéine : importance des queues phytol et de l'ion $Mg^{2+}$**

Les interactions hydrophobes entre les queues phytol des Chl au centre du complexe participent fortement à l'oligomérisation et à la stabilité du complexe (Agostini *et al.*, 2017). Le rôle obligatoire de ces interactions hydrophobes est encore sujet à controverse. Ainsi, Schmidt *et al.* (2003) ont montré que BolWSCP s'associe avec des précurseurs de Chl dépourvus de queue phytol tels que la Mg-protoporphyrine IX ou la chlorophyllide (Chlide, annexe 2) sans oligomérisation sous forme de tétramère. Des études menées par Agostini *et al.* (2017), Girr *et al.* (2020) et Bednarczyk *et al.* (2021) contredisent ce résultat en montrant la formation de tétramères stables entre WSCPII (BolWSCP et LvWSCP) et la Chlide ou avec des dérivés cataboliques tels que la phéophytine (Chl sans  $Mg^{2+}$ ) ou la phéophorbide (Chl sans queue phytol ni  $Mg^{2+}$ ). Les affinités aux molécules dépourvues de queue phytol ou de  $Mg^{2+}$  sont inférieures à celle de la Chl ce qui indiquerait tout de même l'importance de la chaîne phytol et de l'ion central dans les interactions (Girr *et al.*, 2020). Cependant, le fait que les WSCPII puissent également se lier à l'hème avec un arrangement similaire à celui de la Chl suggèrent l'importance d'un cation central que ce soit le  $Mg^{2+}$  ou le  $Fe^{2+}$  en tant que site d'interaction avec les WSCPII (Pro 36, Fig. 6E) (Girr *et al.*, 2020). Les capacités des WSCPII à se lier à des intermédiaires métaboliques de la Chl ou d'autres molécules tétrapyrroliques doivent également être à prendre en considération dans leur rôle physiologique potentiel.

### **II.2.2.c Interaction protéine-protéine dans le tétramère**

Les CDB1 définis par Prabahar *et al.* (2020) interviennent également dans les interactions entre les sous-unités constitutives du complexe soit à l'interface entre les sous-unités internes de chaque dimère (A:B ou C:D), soit à l'interface entre dimères (A-B:C-D) (Fig. 6 C-D). Plus spécifiquement, le CDB2 interviendrait dans l'assemblage intra-dimère, interface très similaire quelle que soit la nature des WSCPII. La Tyr-52 semble en particulier être primordiale dans ces interactions. Les trois autres domaines, impliqués dans les connexions entre les deux dimères, seraient plus spécifiques selon les deux classes WSCPIIA et WSCPIIB. Plus spécifiquement, les interactions hydrophobes mettant en jeu les Trp-151 et Leu-152 de la sous-unité d'un dimère et Ile-37, Leu-38 et Phe-40 de la sous-unité opposée de l'autre dimère participent à l'interaction inter-dimères et à la stabilisation des complexes WSCPII-Chls.





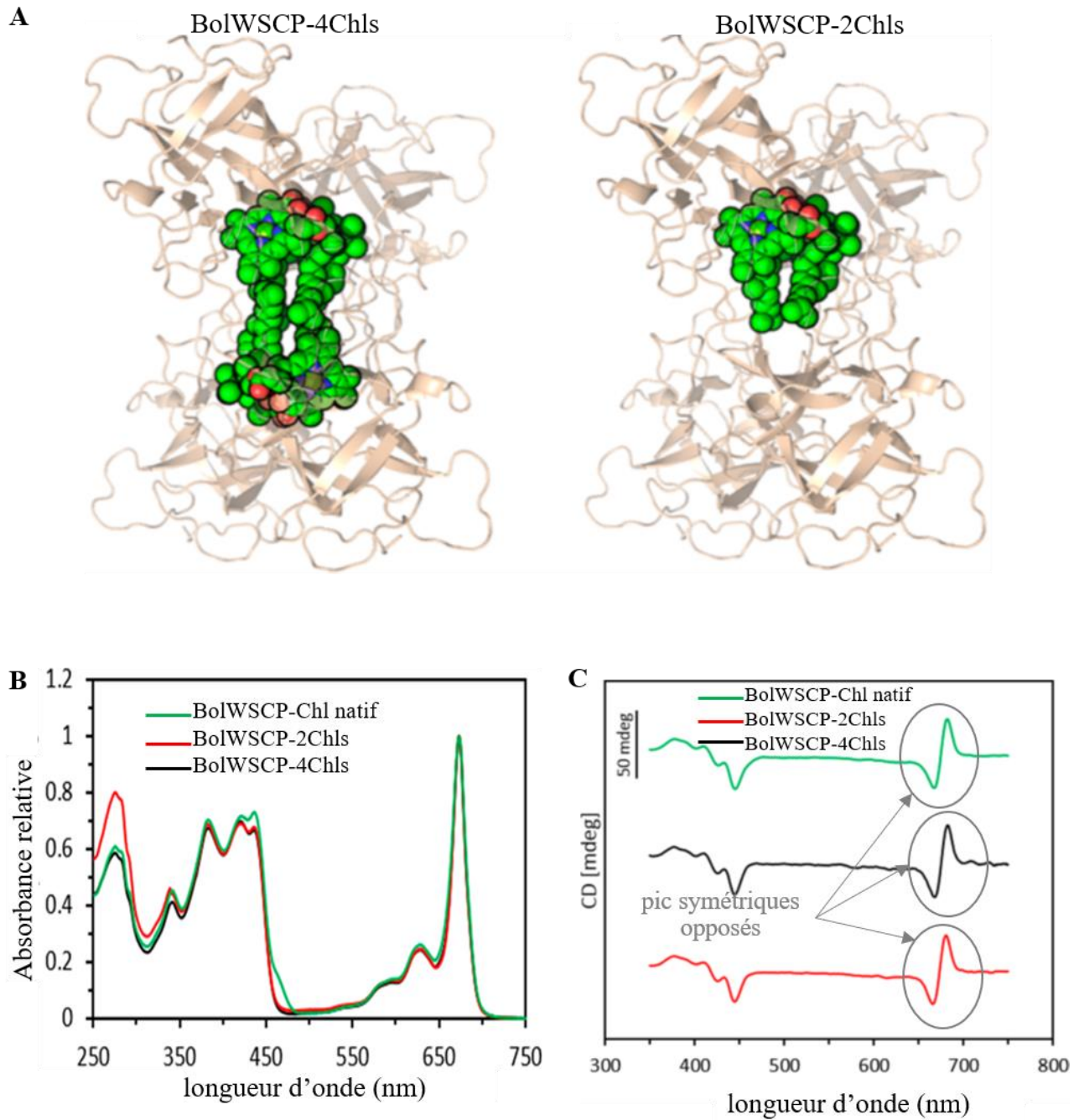
**Fig. 8 : Interactions entre les motifs d'affinité et la Chl dans les complexes LvWSCP-Chlb et BolWSCP-Chla (d'après Palm *et al.*, 2018)**

Les noyaux tétrapyrroliques des chlorophylles sont indiqués par A, B, C et D. La liaison hydrogène entre le carbone C7 du noyau tétrapyrrolique B et le résidu le plus proche est indiquée par des pointillés rouges (distance en Å°).

### II.3. Sélectivité-affinité des WSCPII aux Chls

La comparaison des complexes WSCPIIA-Chls (natif et recombinant) avec ceux de WSCPIIB-Chl montre des différences spectroscopiques claires, en partie dues à leur différence de teneurs en Chla/Chlb (10 pour les complexes natifs BolWSCP et 1,6-1,9 pour LvWSCP). Takahashi *et al.* (2012, 2013a) ont déterminé des quantités équivalentes de Chla et Chlb dans le complexe BolWSCP-Chl si celui-ci est assemblé avec des concentrations équimolaires des deux Chls, soulignant ainsi une absence de sélectivité. Le complexe LvWSCP-Chl par contre montre clairement une sélectivité pour la Chlb. *In vitro*, ces caractéristiques sont retrouvées avec BolWSCP n'ayant aucune préférence claire pour la Chla ou la Chlb et LvWSCP montrant une préférence 16 fois plus importante pour la Chlb par rapport à la Chla (Palm *et al.*, 2018).

Par cristallographie, Palm *et al.* (2018) ont identifié un motif responsable de l'affinité de liaison aux Chls dans la boucle V des WSCPII : LCPS présent dans LvWSCP garant de l'affinité à la Chlb et PVCNEL dans BolWSCP garant de l'affinité à la Chla (Fig. 8). La boucle V contenant le motif d'affinité est en contact étroit avec le carbone C7 du noyau tétrapyrrolique B qui distingue la Chla (groupement méthyle) de la Chlb (groupement formyle) (annexe 1). Dans LvWSCP-Chlb, l'atome d'azote de la leucine (91) de LCPS est suffisamment proche du groupe formyle de la Chlb pour former une liaison hydrogène, alors que dans BolWSCP-Chl les acides aminés de PVCNEL sont trop éloignés. Par conséquent, dans le cas de LvWSCP la Chlb est préférée lorsqu'une liaison hydrogène peut-être formée avec le groupement formyl, tandis que la Chla est préférée lorsque la liaison hydrogène est stériquement défavorable. La BolWSCP ne différencie pas les deux Chl car il n'y a pas de donneur de liaison hydrogène ni à proximité du groupement formyle de la Chlb ni du groupement méthyle de Chla (Palm *et al.*, 2018, 2019; Girr *et al.*, 2020). L'importance des deux motifs a été démontrée par mutagenèse dirigée : l'interconversion des motifs au sein des deux WSCPII suffit à inverser leurs affinités. Plus précisément, le remplacement de la leucine (acide aminé initiateur de la liaison avec le groupement formyle de la Chlb) par une proline dans le motif LCPS de LvWSCP modifie radicalement son profil d'affinité avec une préférence pour Chla d'un facteur d'environ 40 (Palm *et al.*, 2018). Prabahar *et al.* (2020) confirment la présence de ces deux motifs spécifiques aux WSCPIIA ou WSCPIIB (dans le CBD 3) par comparaison de multiples WSCPII avec des protéines ancestrales des différents groupes phylogéniques. Pour l'ensemble de ces travaux, seules des affinités relatives par comparaison entre les deux Chls ont été estimées dans une situation de compétition pour la liaison aux WSCPII. Récemment, Girr *et al.*, (2020) ont expliqué la préférence Chla/Chlb par différentes constantes de dissociation et non pas par différentes de vitesses de réaction de liaison. Ces études restent à approfondir car la limite de détection de la méthode utilisée de dichroïsme circulaire n'a pas permis de définir des constantes de dissociation (Kd) très précises.



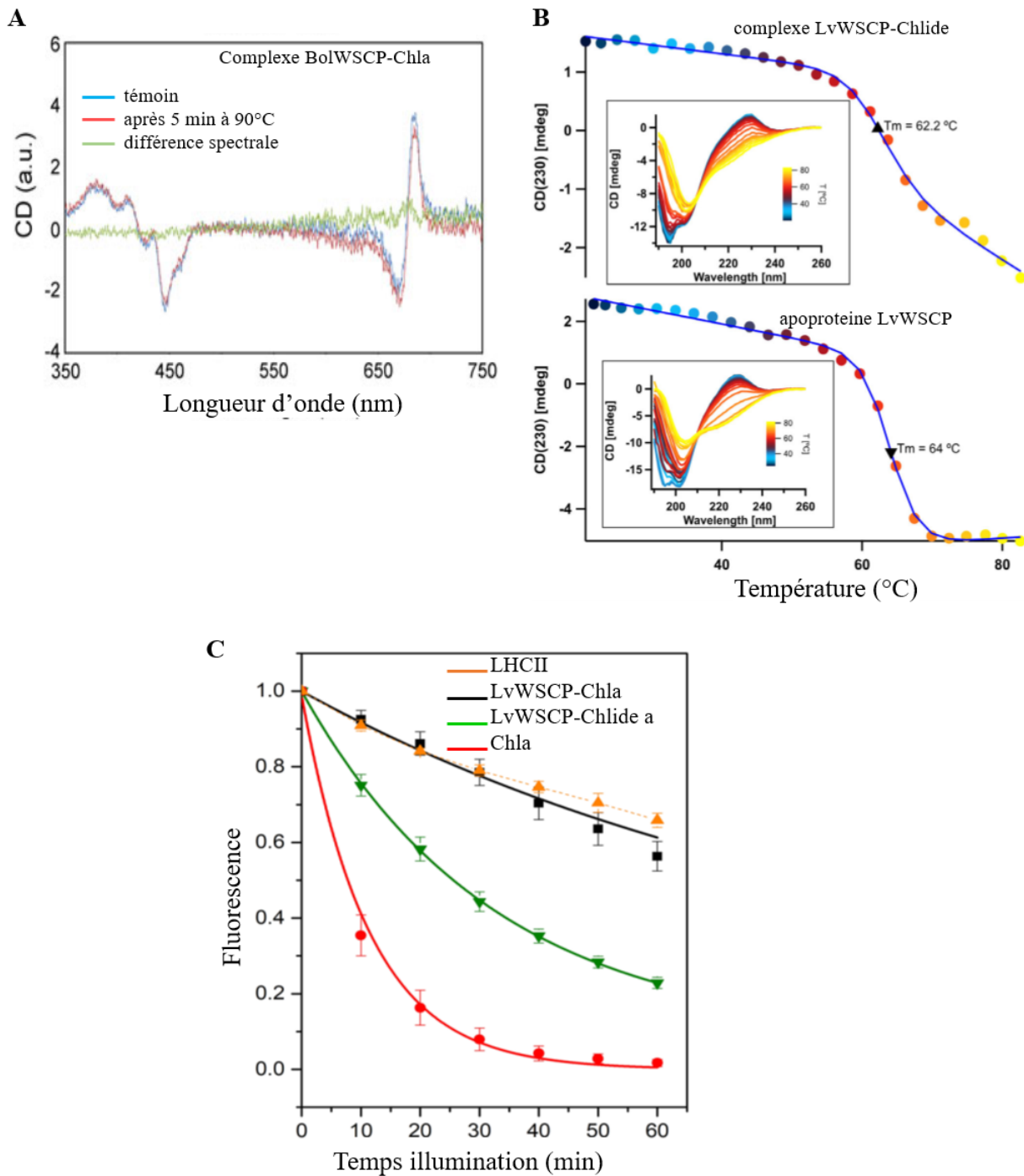
**Fig. 9 : Complexes BolWSCP-Chl natifs ou reconstitués, associés à 2 ou 4 Chls (d'après Palm *et al.*, 2017)**

**A)** Structure du complexe BolWSCP-Chl associé à 4 Chls ou à 2 Chls.

**B-C)** Spectres d'absorption (**B**) et de dichroïsme circulaire (**C**) des complexes BolWSCP-Chl natif (purifié à partir des feuilles), ou reconstitué avec 2 ou 4 Chls. La reconstitution est réalisée en mélangeant les apoprotéines BolWSCP recombinantes avec de la Chla.

#### II.4. Stœchiométrie des Chls dans les complexes WSCPII-Chls

Une fois les complexes formés, certaines études ont rapporté la présence de 1 à 4 Chls pour différentes WSCPII (Kamimura *et al.*, 1997; Satoh *et al.*, 1998; Schmidt *et al.*, 2003). Ces différentes stœchiométries en Chl ont été remises en question par Bednarczyk *et al.* (2015) avec la présence exclusive de 4 Chls dans les complexes constitués de BolWSCP, LvWSCP ou RshWSCP. Récemment, les travaux de Palm *et al.* (2017) ont démontré que le tétramère BolWSCP pouvait fixer 2 ou 4 chls selon le ratio Chl/WSCPII lors de la reconstitution des complexes. A de faibles concentrations en Chl, les complexes se liaient à deux Chls, alors qu'en excès de Chl, les WSCPII seraient saturées en Chl avec les 4 sites occupés. Pour l'état 2 Chls/tétramère, les deux molécules se fixeraient sur le même dimère (Fig. 9A). Les spectres d'absorption et de dichroïsme circulaire des complexes à 2 Chls sont similaires à ceux avec 4 Chls et à ceux des complexes natifs (Fig. 9B-C). Les complexes formés avec 2 Chls possèdent une forte stabilité comparable à celle des complexes avec 4 Chls. Cette stabilité est due à l'interaction entre la queue phytol avec un groupe d'acides aminés hydrophobes constituées par le Trp-151 et la Leu-152 de la sous-unité liant la Chl et les Ile-37, Leu-38 et Phe-40 de la sous-unité opposée de l'autre dimère vide. Bien que les complexes tétramériques à 2 Chl/soient stables, ils sont suffisamment flexibles pour être en capacité de lier deux Chls supplémentaires (Palm *et al.*, 2017). S'il est acté que les WSCPII s'associent aux Chls sous forme de tétramère composé de deux dimères, les étapes d'assemblage du tétramère restent toutefois inconnues. La présence de dimères ou de monomères WSCPII associés à des Chls n'a jamais été observée (Bednarczyk *et al.*, 2015; Palm *et al.*, 2017). En absence de Chl, les WSCPII ne forment pas de tétramères ce qui signifie que la liaison aux Chls (2 ou 4) est un prérequis indispensable à la formation de tétramères stables. De plus, l'absence de monomère-Chl, de dimère-Chl ou même de tétramère avec moins de 2 Chls indique que la fixation de 2 Chl par tétramère est une exigence minimale pour la stabilisation du complexe tétramérique.



**Fig. 10 : Thermo- et photo-stabilité des complexes WSCPII-Chl**

**A)** Spectre de dichroïsme circulaire (CD) des complexes BolWSCP-Chla reconstitués avant et après incubation à 90°C pendant 5 min (d'après Takahashi *et al.*, 2014).

**B)** Courbes de dénaturation thermique de LvWSCP seule ou associée avec la chlorophyllide a (Chlide a) (d'après Bednarczyk *et al.* 2021). Les courbes sont déterminées à partir des spectres CD obtenus sous plusieurs températures.

**C)** Photostabilité des complexes LvWSCP associés avec la Chla ou la Chlide a, du complexe LHCII (purifié à partir des feuilles) et de la Chla seule (d'après Agostini *et al.*, 2017).

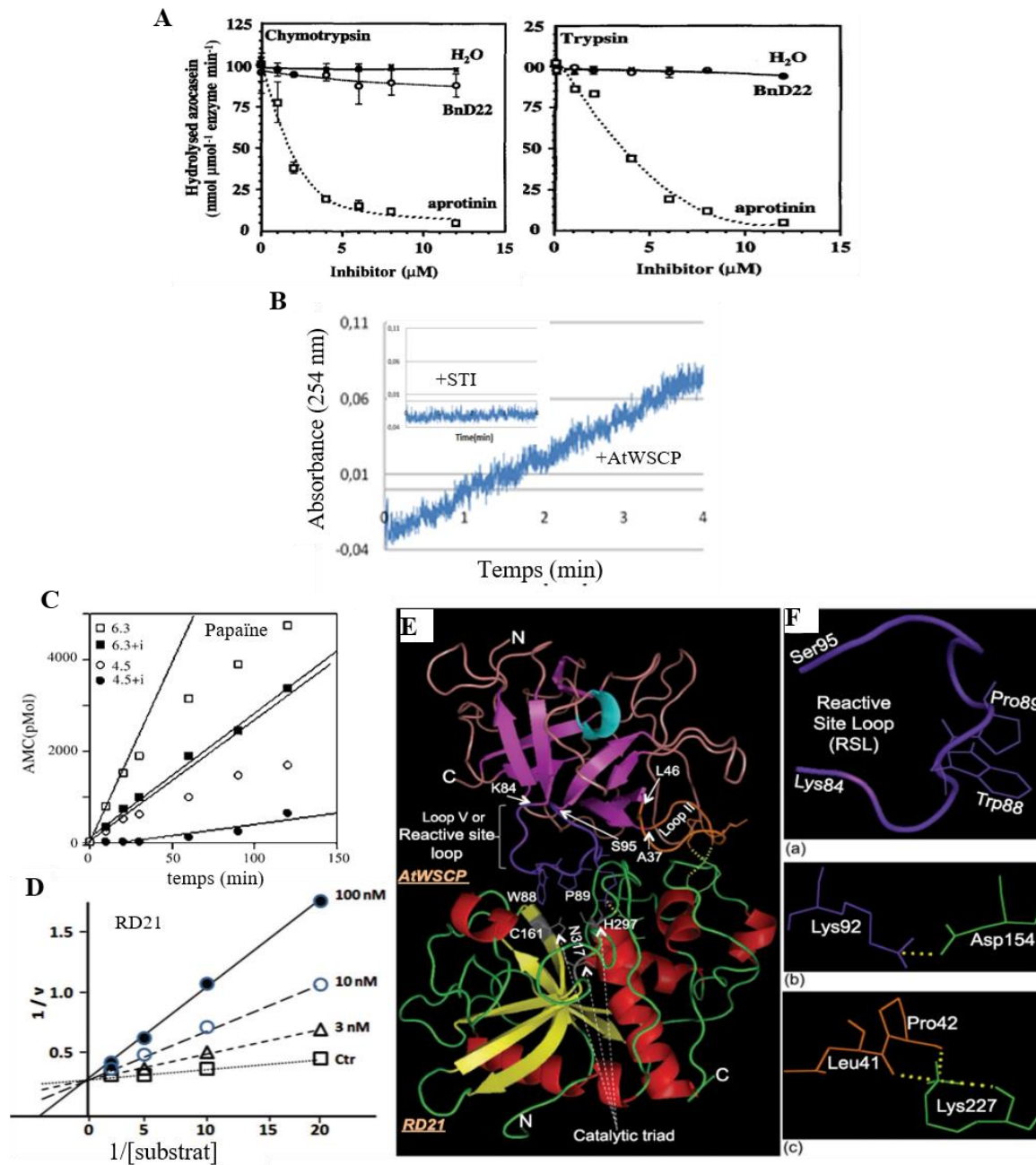
## II.5. Protection des chlorophylles

Les complexes WSCPII-Chl présentent une stabilité vis-à-vis de la dénaturation thermique (Fig. 10 A-B) ainsi qu'une résistance aux pH extrêmes et à la digestion par les protéases (Schmidt *et al.*, 2003; Bektas *et al.*, 2012; Takahashi *et al.*, 2012; Palm *et al.*, 2019). De plus, les Chls liées aux WSCPII sont partiellement protégées de la photo-oxydation lors de l'illumination (Fig. 10C, Schmidt *et al.*, 2003; Agostini *et al.*, 2017; Palm *et al.*, 2019). Bien que le complexe WSCPII-Chl ne contienne aucun caroténoïde qui est un pigment responsable de la photoprotection dans des complexes photosynthétiques, la photostabilité des Chls au sein des complexes est comparable à celle des photosystèmes (Agostini *et al.*, 2017) (Fig. 10C). La photostabilité élevée dans les complexes s'expliquerait par l'empaquetage des Chls dans la cavité centrale limitant fortement leurs interactions avec le milieu environnant et donc le contact entre l'oxygène et le magnésium qui est l'étape initiatrice de production des espèces réactives de l'oxygène (ROS) (Schmidt *et al.*, 2003; Horigome *et al.*, 2007). Dans ce sens, une production moindre de ROS lors de l'illumination des complexes BolWSCP-Chls a été mesurée (Schmidt *et al.*, 2003), mais remise en cause par Agostini *et al.* (2017) qui ont démontré une production efficace de ROS par les complexes LvWSCP-Chl. Pour ces auteurs, la protection des noyaux tétrapyrroliques serait due à la disposition spéciale des queues phytol dans les complexes (Agostini *et al.*, 2017; Palm *et al.*, 2019). Ainsi, les complexes WSCPII-Chlide, molécule dépourvue de queue phytol, offrent une moindre protection vis-à-vis des dommages thermiques et photoniques (Fig. 10B-C). La queue phytol semble jouer un rôle clé dans la photo-protection et la photo-stabilité des Chls dans les complexes WSCPII-Chls (Bektas *et al.*, 2012; Agostini *et al.*, 2017; Palm *et al.*, 2019).

## III. Propriétés fonctionnelles des WSCPII : interaction avec les protéases

En comparaison avec le nombre d'études portant sur la caractérisation de liaison à la Chl, les WSCPII sont peu étudiées pour leur fonction d'inhibiteur de protéases (IP). Seules BnD22, BolWSCP et AtWSCP ont fait l'objet de travaux quant à leur action inhibitrice contre des Ser et des Cys protéases, cibles des IP de type Kunitz.

Les IP de type Kunitz sont les plus répandus dans le règne végétal. Des études ultérieures ont révélé les propriétés anti-protéasiques étonnantes de cette famille d'inhibiteurs. En effet, ce sont les seuls inhibiteurs protéiques connus dans le règne végétal capables d'inhiber des protéases appartenant à trois classes différentes (Ser, Cys et Asp protéases) (Mosolov et Valueva, 2005; Rustgi *et al.*, 2018). Les Ser et Cys protéases font partie de la plus grande classe de protéases chez les plantes avec des niveaux d'expression et des localisations spécifiques. Ces protéases sont localisées dans tous les compartiments subcellulaires (Diaz-Mendoza *et al.*, 2016) et participent aux différents processus



**Fig. 11 : Activités anti-protéolytiques des WSCP II**

**A)** Activité de BnD22 contre la trypsin ou la chymotrypsine. BnD22 est purifiée à partir des feuilles de colza et son activité inhibitrice contre la trypsin et chymotrypsine est suivie par hydrolyse de l'azocaséine, substrat des deux protéases (d'après Ilami *et al.*, 1997).

**B)** Activité anti-trypsin d'AtWSCP. L'effet inhibiteur de la protéine recombinante est suivi par hydrolyse de BAEE (N- $\alpha$ -benzoyl-L-tyrosine ethyl ester), substrat de la trypsin. L'inhibiteur de trypsin STI est utilisé comme témoin positif (d'après Boex-Fontvieille *et al.*, 2016).

**C-D)** Activité d'inhibition des Cys protéases (C, papaine et D, RD21) par AtWSCP recombinante. Les activités de la papaine et de RD21 sont suivies respectivement en présence d'AMC (amino-4-méthylcoumarine) et BApNa (N $\alpha$ -benzoyl-dl-arginine p-nitroanilide) en fonction du pH (C) ou des quantités en AtWSCP (D) (d'après Boex-Fontvieille *et al.*, 2015a; Halls *et al.*, 2005).

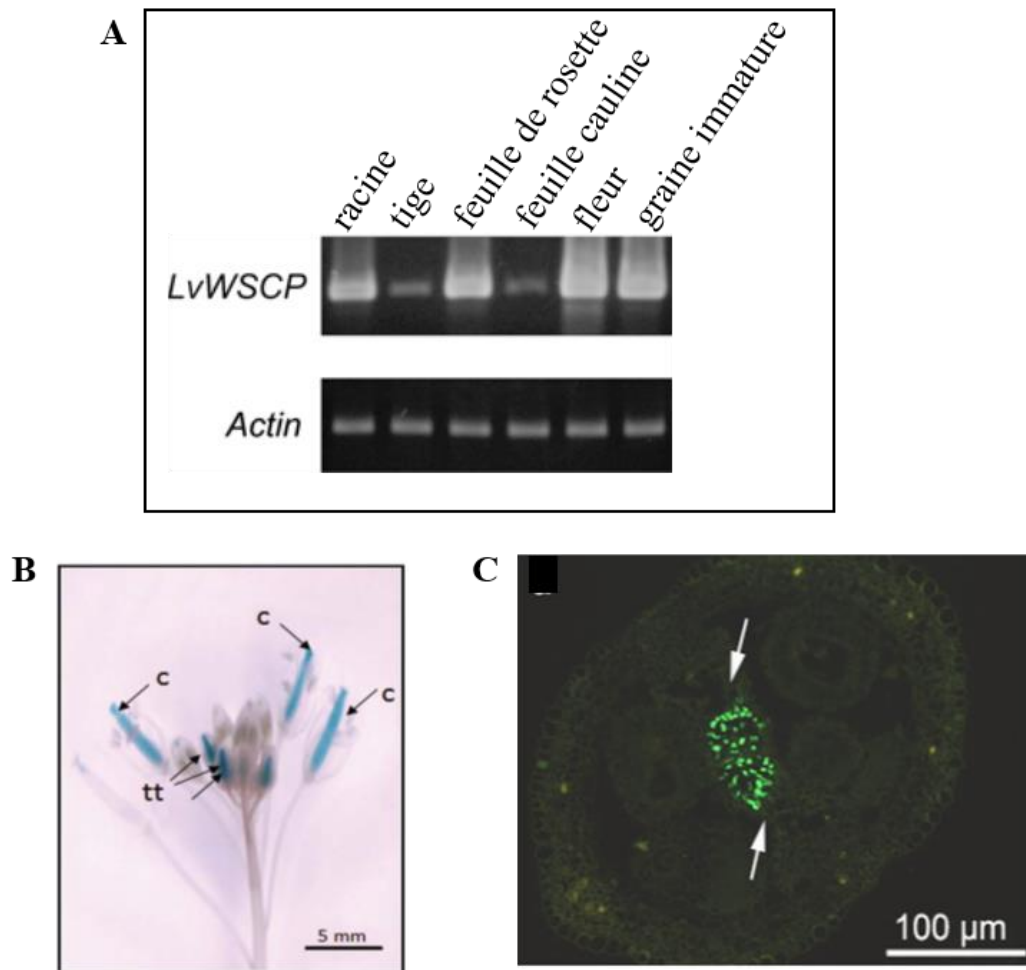
**E-F)** Modèle structural d'interaction entre AtWSCP et RD21 prédit par docking moléculaire (ClusPro). Focus sur les résidus acides aminés en interaction (F) (d'après Boex-Fontvieille *et al.*, 2015).

physiologiques notamment la sénescence et la remobilisation d'azote (Roberts *et al.*, 2012; Girondé *et al.*, 2015a-b; James *et al.*, 2018, 2019)

L'activité inhibitrice envers les Ser protéases se fait selon le mécanisme standard ou mécanisme canonique avec une inhibition de type compétitif réversible (annexe 3). L'inhibiteur se fixe à la place du substrat de l'enzyme avec une très forte affinité (Laskowski Jr. & Kato, 1969; Krowarsch *et al.*, 2003). Le complexe inhibiteur-protéase est alors stabilisé par formation de liaisons hydrogène. L'interaction se fait par le domaine inhibiteur de l'IP ou site réactif localisé à la surface d'une boucle appelée boucle canonique ou « reactive site loop » (RSL). Pour les WSCPII, la boucle RSL serait la boucle IV. La nature de l'acide aminé en position P1 de ce domaine inhibiteur est essentiel pour la spécificité d'interaction avec la protéase cible : Arg/Lys pour les inhibiteurs de trypsine et Ala/Ser pour les inhibiteurs de chymotrypsine (Laskowski & Kato, 1980; Laskowski & Qasim, 2000; Otlewski *et al.*, 2001; De Leo, 2002). Toutefois ces acides aminés sont absents chez AtWSCP, BnD22 et BolWSCP, substitués par une Gly. Une Asn localisée à l'extrémité N-terminale de l'IP joue également un rôle dans l'efficacité de l'inhibition, en stabilisant la boucle canonique (Khamrui *et al.*, 2005). Il semble que les WSCPII ne sont pas ou peu inhibitrices des Ser protéases (Fig. 11A-B). Un faible niveau d'activité anti-protéolytique a été observé pour BnD22 purifiée à partir des feuilles de colza (Fig. 11A, Ilami *et al.*, 1997) ou dans les jeunes feuilles soumises au méthyl jasmonate (Desclos *et al.*, 2008). Par contre, Nishio et Satoh (1997) et Boex-Fontvieille *et al.* (2016) attestent, *in vitro*, d'une absence d'inhibition envers la trypsine pour BnD22, BolWSCP et AtWSCP (Fig. 11B).

L'activité inhibitrice envers les Cys protéases se fait sur un mode d'inhibition compétitive assistée par positionnement sur un exo-site proche du site de liaison du substrat. A ce jour, la capacité d'inhibition des WSCPII envers les Cys protéases a été suivie uniquement pour AtWSCP. Cette protéine se lie *in vitro* à la papaine, protéase de référence des Cys protéases de type granuline (papaine like Cys protease, PLCP) et inhibe son activité (Fig. 11C, Halls *et al.*, 2006; Rustgi *et al.*, 2017). Par le système double hybride, Boex-Fontvieille *et al.* (2015a) ont montré que trois des quatre PLCP d'*A. thaliana*, RD21 (Responsive to Dessication 21), At3g19390 et At5g43060, peuvent interagir avec AtWSCP. *In vitro*, AtWSCP est capable d'inhiber l'activité de RD21 avec une Kd apparente de  $10^{-8}$  M (Fig. 11D, Boex-Fontvieille *et al.*, 2015a). *In vivo*, l'association entre AtWSCP et RD21 est retrouvée dans les plantules étiolées d'*A. thaliana* (Boex-Fontvieille *et al.*, 2015b). L'implication de certains acides aminés dans l'interaction AtWSCP-RD21 a été prédite par visualisation de la structure tridimensionnelle du complexe AtWSCP-RD21 (Fig. 11 E-F). La boucle V s'introduit dans le site actif de RD21 et deux acides aminés (Tyr-88 et Pro-89) interagissent avec deux acides aminés Cys-161 et His-297 de la triade catalytique (Cys-His-Asn) de la protéase pour en bloquer l'activité. De plus, la Lys-92 de la RSL, et les Leu-41 et Pro-42 du motif de liaison à la Chl (boucle II) forment des





**Fig. 12 : Localisations tissulaires de WSCP II**

**A)** Expression du gène *LvWSCP* dans différents tissus chlorophylliens ou non de *L. virginicum* (d'après Takahashi *et al.*, 2013b).

**B)** Expression du promoteur d'*AtWSCP* fusionné au gène rapporteur *GUS* ( $\beta$ -glucoronidase) dans les inflorescences. c, carpelle; tt, tissu de transmission (d'après Boex-Fontvieille *et al.*, 2016).

**C)** Immunolocalisation de la protéine *AtWSCP* dans les siliques d'*A. thaliana*. La présence des protéines est révélée par les anticorps anti-*AtWSCP* couplés à un anticorps secondaire fluorescent sur des coupes transversales des jeunes siliques. Les flèches indiquent le septum décoré (d'après Bektas *et al.*, 2012).

liaisons hydrogène avec RD21 contribuant ainsi à la stabilité du complexe (Boex-Fontvieille *et al.*, 2015b; Rustgi *et al.*, 2017).

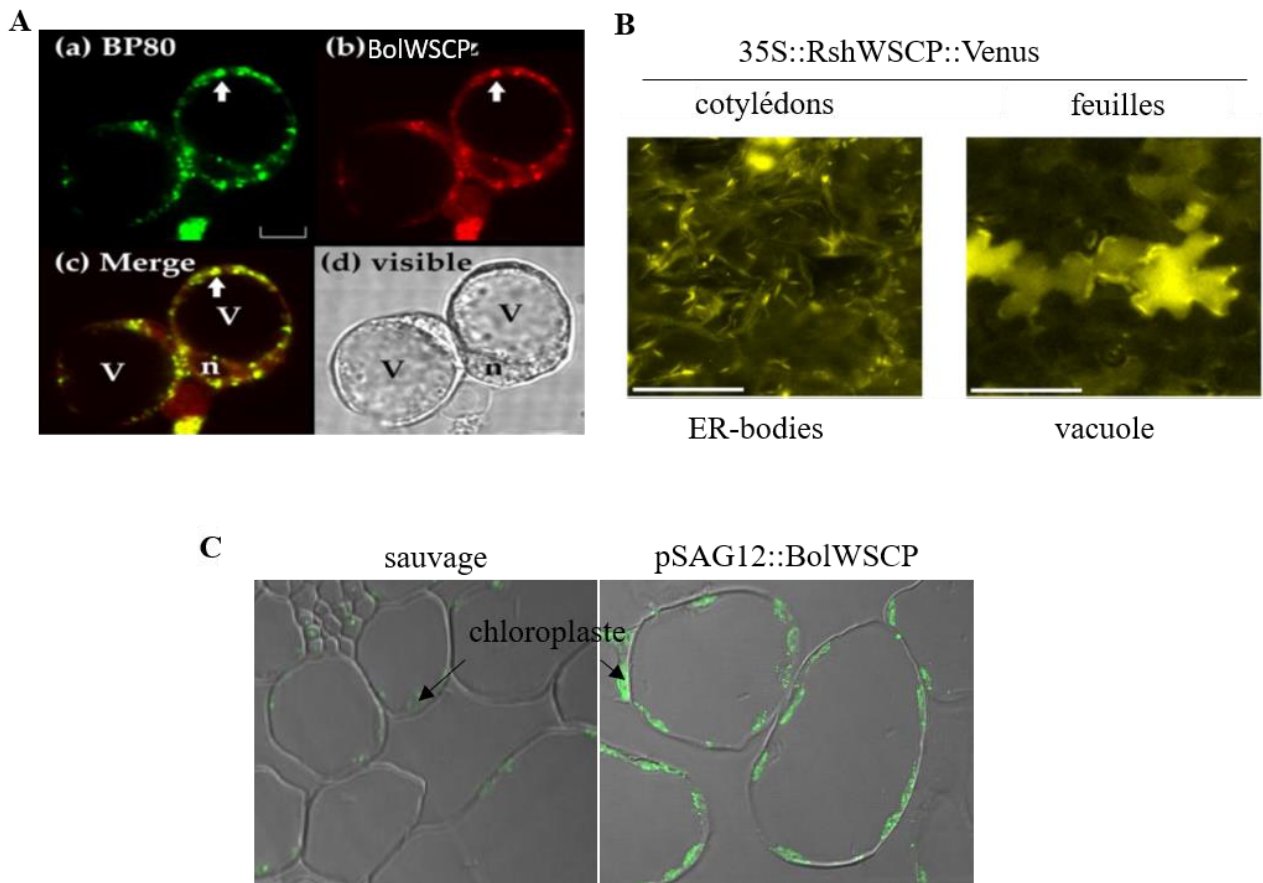
Boex-Fontvieille *et al.* (2015b) suggèrent que l'activité inhibitrice d'AtWSCP contre RD21 est Chl-dépendante. Ainsi, le complexe AtWSCP-RD21 formé dans les plantules étiolées disparaît lors du verdissement des plantules. La liaison d'AtWSCP à la Chl ou à la Chlide lors du passage à la lumière des plantules provoquerait la dissociation avec RD21 car les sites d'interaction avec ces deux molécules sont identiques. AtWSCP est l'unique WSCPII qui a fait l'objet d'étude simultanée des fonctions IP et de liaison à la Chl (Boex-Fontvieille *et al.*, 2015b; Rustgi *et al.*, 2017).

#### IV. Rôles physiologiques des WSCPII

##### IV.1. Localisation tissulaire et subcellulaire des WSCPII

De par leur fonction de liaison aux Chl et leur purification à partir des feuilles, il est attendu que les WSCPII soient exprimées dans les tissus chlorophylliens. BnD22 (Downing *et al.*, 1992; Ilami *et al.*, 1997), BolWSCP (Murata *et al.*, 1971; Nishio & Satoh, 1997; Takahashi *et al.*, 2012) et LvWSCP (Fig. 12A, Murata & Ishikawa, 1981; Takahashi *et al.*, 2013b) sont présentes préférentiellement dans les jeunes feuilles en croissance et RshWSCP a été purifiée à partir des nervures principales (Shinashi *et al.*, 2000). L'expression des WSCPII n'est cependant pas spécifique aux tissus chlorophylliens. Ainsi, LvWSCP est détectée dans les racines, tiges, fleurs, et graines immatures (Fig.12A), avec un niveau de transcription particulièrement élevé dans les jeunes tissus en développement (Takahashi *et al.*, 2013b). RshWSCP est principalement exprimée dans les bourgeons et les fleurs et absente dans les tiges et les feuilles (Takahashi *et al.*, 2013a). AtWSCP n'est pas détectée dans les feuilles (Bektas *et al.*, 2012; Boex-Fontvieille *et al.*, 2015a, 2016a) mais présente dans les inflorescences et plus particulièrement dans le tissu de transmission des pistils, chemin emprunté par les tubes polliniques pour atteindre les ovules (Fig. 12B, Boex-Fontvieille *et al.*, 2015a). Aucune WSCPII n'a été à ce jour détectée dans les graines matures (Bektas *et al.*, 2012; Boex-Fontvieille *et al.*, 2015a, 2016a). Une localisation préférentielle des WSCPII dans les jeunes tissus en croissance est observée et leurs présences tissulaires peuvent évoluer selon le développement de la plante. Ainsi AtWSCP est exprimée dans la crosse apicale des plantules, absente des feuilles, et présente dans les inflorescences puis les siliques (Fig. 12C, Bektas *et al.*, 2012; Boex-Fontvieille *et al.*, 2015a-b, 2016a-b; Rustgi *et al.*, 2017).

La localisation subcellulaire des WSCPII reste encore sujette à débat. Leur capacité à se lier aux Chls sous-entend une localisation chloroplastique et leur rôle d'IP privilégie une localisation vacuolaire sans proscrire leur présence dans les chloroplastes. Aucune WSCPII ne possède de



**Fig. 13 : Localisation subcellulaire des WSCP II**

**A)** Colocalisation de BolWSCP et BP80 (vacuolar-sorting receptor protein) dans les compartiments prévacuolaires de protoplastes de tabac (d'après Halls *et al.*, 2006). Après cotransformation des protoplastes de tabac avec BolWSCP et BP80, les protéines exprimées sont révélées par immunolocalisation à l'aide d'anticorps fluorescents spécifiques de chaque protéine. V, vacuole ; N, noyau.

**B)** Localisation subcellulaire de RshWSCP exprimée dans *A. thaliana* dans les ER bodies et les vacuoles (d'après Takahashi *et al.*, 2015). Après transformation d'*A. thaliana* avec RshWSCP fusionné au gène rapporteur *Venus* sous contrôle du promoteur 35S, la détection de RshWSCP est réalisée par microscopie confocale

**C)** Localisation chloroplastique de BolWSCP dans les feuilles d'*A. thaliana* (d'après James, 2018). Après transformation des lignées RIL232 par BolWSCP sous promoteur *SAG12* (*pSAG12::BolWSCP*), la protéine est détectée par immunolocalisation.

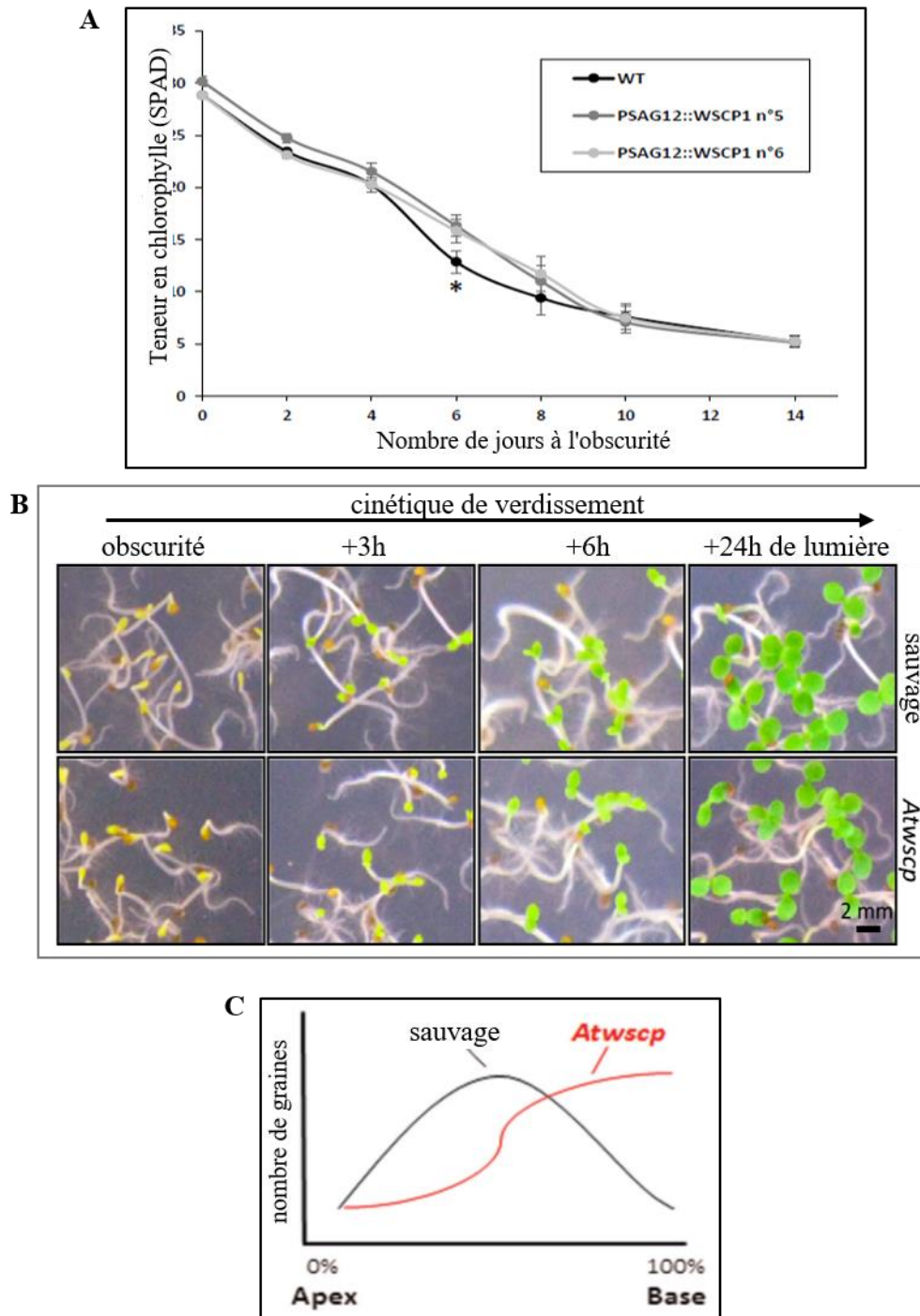
séquence d'adressage spécifique aux vacuoles ou aux chloroplastes, mais sont prédites *in silico* associées au système de sécrétion (Tab. 2; Satoh *et al.*, 2001). La localisation subcellulaire de certaines WSCPII a été précisée. Halls *et al.* (2006) ont montré dans des protoplastes de tabac une expression conjointe de BolWSCP et BP80 (Vacuolar-sorting receptor protein) marqueur des compartiments prévacuolaires, ce qui présageait d'une localisation vacuolaire à terme (Fig. 13A). Les surexpressions dans *A. thaliana* d'AtWSCP, LvWSCP, BolWSCP ou RshWSCP montrent leurs présences dans des structures particulières du réticulum endoplasmique, spécifiques aux brassicales, les ER-bodies et dans les vacuoles (Fig. 13B, Takahashi *et al.*, 2012, 2013b, 2015). Une autre étude a mis en évidence la présence d'AtWSCP dans les parois cellulaires et les espaces extracellulaires ce qui argumente favorablement l'association des WSCPII au système de sécrétion (Boex-Fontvieille *et al.*, 2015b). La localisation chloroplastique est plus controversée et seuls deux travaux vont dans ce sens. Satoh *et al.* (1998) ont réussi à purifier LvWSCP dans la fraction soluble des chloroplastes et BolWSCP a été localisée dans les chloroplastes après surexpression sous contrôle du promoteur *Sag12* dans *A. thaliana* (Fig. 13C; James, 2018).

#### **IV.2. Implication des WSCPII dans le métabolisme de la Chl**

Bien qu'à ce jour il n'a pas été possible de localiser les WSCPII dans les chloroplastes, leur capacité à se lier aux Chls, et à certains de ces précurseurs ou dérivés cataboliques est incontestable au moins *in vitro* (voir chapitre II.22.b). Pour rappel, le métabolisme des Chls est divisé en quatre parties (annexe 2) :

- La biosynthèse, commune à d'autres molécules tétrapyrroliques, dépendante de la lumière, se déroulant entièrement dans les chloroplastes et faisant intervenir quinze réactions enzymatiques. Parmi celles-ci, la photo-réduction de la protochlorophyllide, stockée préalablement dans les corps pro-lamellaires des étioplastes, en chlorophyllide est concomitante à la mise en place des structures thylakoïdiennes et à terme à la production de Chl
- La conversion enzymatique dans le chloroplaste de la Chlb à partir de la Chla
- Le recyclage de la Chl avec la conversion réversible de la Chla en Chlide (Lin *et al.*, 2016)
- Le catabolisme de la Chl qui débute dans les plastes puis finit dans la vacuole (Hörtensteiner et Kräutler, 2011; Kuai *et al.*, 2018).

Le contrôle des activités de biosynthèse et de dégradation des Chls est intimement lié à la biogénèse, à la maintenance et au démantèlement des photosystèmes et à la demande en Chls qui change dynamiquement au cours du développement des feuilles et en réponse aux modifications de l'environnement. Les WSCPII n'interviennent pas dans la photosynthèse alors que leur liaison avec les Chls entraîne une photoprotection de ces dernières (Horigome *et al.*, 2007; Bektas *et al.*, 2012; Palm *et al.*, 2019) ce qui soulève des questions. La propriété de liaison des WSCPII aux Chls et à



**Fig. 14 : Implication physiologique des WSCP II dans le métabolisme de la Chl (A et B) et le développement des graines (C).**

**A)** Cinétique de dégradation des chlorophylles au cours de la sénescence induite par l'obscurité dans les feuilles d'*A. thaliana* non transformée (WT) et de deux lignées transformées surexprimant BolWSCP (*pSAG12 :: WSCP1* n°5 et 6) (d'après James, 2018).

**B)** Phénotype des plantules de la souche sauvage et du mutant *Atwscp* au cours du verdissement des plantules étiolées (d'après Boex-Fontvieille *et al.*, 2015a).

**C)** Distribution du nombre de graines dans les fruits d'*A. thaliana* sauvage et du mutant *Atwscp* knockout (d'après Boex-Fontvieille *et al.*, 2015b). En abscisse, la longueur du fruit est divisée en portions de la partie la plus apicale (Apex) à la partie la plus basale (Base).

certaines de ses précurseurs ou dérivés cataboliques, a de fait suscité des spéculations quant à leur éventuelle implication dans la biosynthèse ou le catabolisme de la Chl. La liaison des WSCPII aux précurseurs de la Chl pourrait en effet supposer qu'elles fonctionnent comme transporteurs de ces précurseurs pendant la synthèse de la Chl et l'assemblage de l'appareil photosynthétique (Schmidt *et al.*, 2003; Reinbothe *et al.*, 2004). Ainsi, chez l'orge, Reinbothe *et al.* (2004) ont identifié une protéine FCBP (Free Chl(ide)-Binding Protein) reconnue par l'anticorps dirigé contre LvWSCP et qui présente une homologie avec les WSCPII en N-terminal. FCBP s'accumule au niveau de la membrane externe des étiooplastes et est importée dans les plastes à la lumière lors de la synthèse de Chlide à laquelle elle se lie. FCBP séquestrerait la Chlide pour prévenir d'éventuels dégâts occasionnés par des pigments photo-excitables libres. L'accumulation de protochlorophyllide (précurseur de la chlorophyllide) lors de la surexpression de BolWSCP dans le tabac corrobore le rôle des WSCPII dans un stockage temporaire des précurseurs immédiats de la Chl (Damaraju *et al.*, 2011). De plus, la surexpression de BolWSCP dans les feuilles d'*A. thaliana* est concomitante au ralentissement de la dégradation de la chlorophylle d'environ 20 % par rapport à la lignée sauvage six jours après induction de la sénescence foliaire par l'obscurité (James, 2018 ; Fig. 14A). En revanche, au stade plantule, *A. thaliana* ne présente aucune différence de verdissement entre le génotype sauvage et des mutants *Atwscp* knockout (Boex-Fontvieille *et al.*, 2016, Fig. 14B). Le profil d'expression d'*AtWSCP* dans les plantules n'est pas cohérent avec un rôle de transporteur transitoire de la chlorophyllide pendant l'établissement de l'appareil photosynthétique. *AtWSCP* est détecté uniquement dans les plantules étiolées, absent des cotylédons et des plastes qui sont les localisations attendues pour une protéine liant la Chl ou l'un de ses précurseurs.

Les WSCPII sont capables d'extraire *in vitro* les Chls des thylakoïdes, ce qui suggère un rôle possible lors de la dégradation des complexes LHCII suite à un dommage cellulaire ou lors de l'adaptation de l'appareil photosynthétique sous stress lumineux (Sato *et al.*, 1998). Les enzymes de dégradation des Chls sont situées dans les chloroplastes, au niveau de la membrane interne pour la chlorophyllase, alors que les Chls se trouvent dans les thylakoïdes (Hörtensteiner et Kräutler, 2011; Wang et Grimm, 2021). Pour Matile *et al.* (1997), cette séparation physique entre enzyme et substrat suppose l'existence d'une protéine de transfert. Ce rôle pourrait être tenu par les WSCPII, bien que leur présence chloroplastique ne soit pas démontrée. Par ailleurs, les résultats de Sakuraba *et al.* (2012) ont montré que le complexe protéique responsable de la dégradation de la chlorophylle ne comprend pas de WSCPII et ceux de Damaraju *et al.* (2011) que le catabolisme des Chls dans les plantes de tabac n'était pas perturbé par la surexpression de BolWSCP.



### **IV.3. Implication des WSCP II dans le contrôle de la protéolyse**

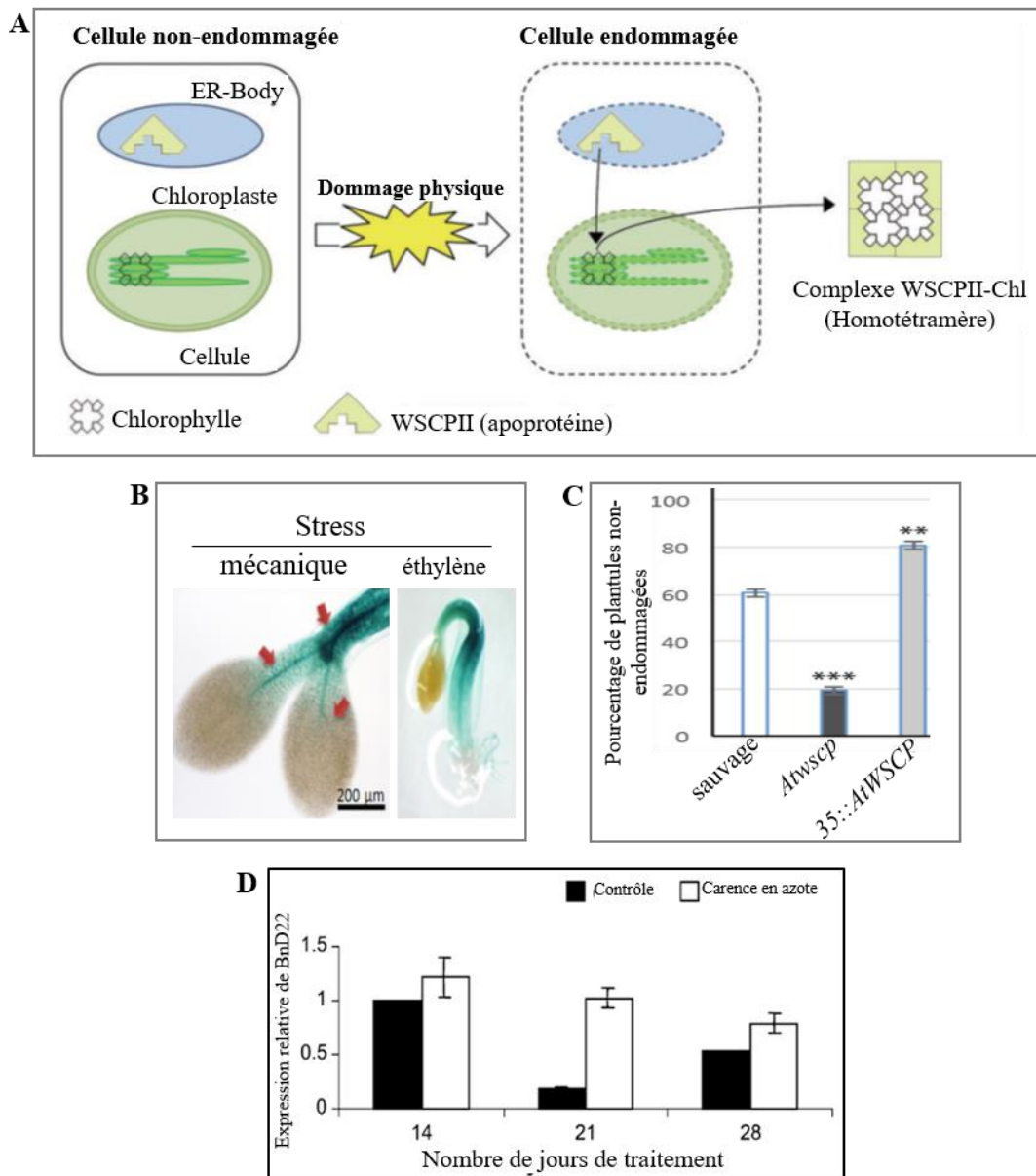
#### ***IV.3.1. Rôle dans les tissus chlorophylliens en croissance***

Il est avéré que les WSCP II sont préférentiellement exprimées dans les jeunes tissus chlorophylliens en croissance comme les siliques (AtWSCP, Bektas *et al.*, 2012) et les feuilles (Takahashi *et al.*, 2012, 2013b). Les deux fonctions des WSCP II sont en accord pour un rôle de protection des jeunes tissus : en tant qu'IP participant à un contrôle de la protéolyse à un niveau modéré et en tant que WSCP II, assurerait une protection des Chls (Avice et Etienne, 2004). Les études ont essentiellement porté sur BnD22 dans le cadre de la remobilisation foliaire des ressources des organes sources vers les organes puits au cours de la sénescence foliaire, le rôle de BnD22 en tant qu'inhibiteur de Ser protéases ayant été démontré dans les feuilles de colza (Desclos *et al.*, 2008) et suggéré *in vitro* (Ilami *et al.*, 1997). BnD22 présente un gradient d'accumulation basipède qui est opposé au gradient de la sénescence séquentielle des feuilles (Etienne *et al.*, 2007 ; Desclos *et al.*, 2008) et un gradient d'expression antagoniste à celui des protéases foliaires (Reviron *et al.*, 1992; Etienne *et al.*, 2007; Desclos *et al.*, 2008). Ainsi, elle est fortement accumulée dans les jeunes feuilles non sénescentes (Downing *et al.*, 1992; Reviron *et al.*, 1992). Takahashi *et al.* (2015) ont montré que les WSCP II (AtWSCP et RshWSCP) sont localisées dans les ER-bodies et les vacuoles où elles pourraient être impliquées dans l'inhibition des protéases comme RD21 dans ces mêmes compartiments cellulaires.

#### ***IV.3.2. Rôle dans les tissus non chlorophylliens : mort cellulaire programmée et skotomorphogénèse***

L'expression d'AtWSCP dans l'inflorescence est régulée de manière différentielle par les facteurs de transcription *NTT* (No Transmitting Tract) et *HEC* (HECATE), impliqués dans l'identité et l'organisation des différents tissus floraux (Bektas *et al.*, 2012; Boex-Fontvieille *et al.*, 2015a). Son implication dans le développement floral, plus précisément dans le contrôle de la mort cellulaire du tissu de transmission a été démontrée par Boex-Fontvieille *et al.* (2015a). AtWSCP pourrait interagir avec RD21 dans les tissus reproducteurs afin d'inhiber son activité protéolytique jusqu'aux stades tardifs du développement floral. Par la suite, le niveau de WSCP diminue dans le tissu de transmission induisant une activation consécutive de RD21 (Boex-Fontvieille *et al.*, 2015a). En absence d'AtWSCP, le contrôle de la mort cellulaire du tissu de transmission est altéré avec pour conséquence une perturbation de la progression des tubes polliniques qui influence le développement des graines (Fig. 14C). L'absence d'AtWSCP n'affecte pas ou peu le rendement en graines mais semble influencer leurs propriétés de germination (Boex-Fontvieille *et al.*, 2015a). De manière intéressante, les facteurs *NTT* et *HEC* sont également exprimés dans les plantules étiolées et régulent l'expression





**Fig. 15 : Implication des WSCPII dans la réponse aux contraintes environnementales**

**A)** Représentation schématique de l'hypothèse de formation des complexes BolWSCP-Chl dans les cellules végétales endommagées (d'après Takahashi *et al.*, 2012). Dans les cellules non endommagées, les apoprotéines WSCPII sont stockées dans des structures particulières du réticulum endoplasmique (ER-body). Lorsque les cellules sont endommagées, les apoprotéines BolWSCP sont libérées des ER-bodies et vont alors se lier à la Chl localisée dans les membranes de thylakoïdes endommagés.

**B)** Induction du gène *AtWSCP* lors d'un stress mécanique ou en présence d'éthylène dans la crosse apicale des plantules d'*A. thaliana*. Le promoteur du gène *AtWSCP* a été fusionné au gène rapporteur *GUS* ( $\beta$ -glucuronidase) (d'après Boex-Fontvieille *et al.*, 2016).

**C)** Pourcentage de plantules *A. thaliana* endommagées par les arthropodes chez le génotype sauvage, le mutant knockout (*Atwscp*) et le surexprimeur (*35::AtWSCP*) (d'après Boex-Fontvieille *et al.*, 2016).

**D)** Expression du gène *BnD22* par q-PCR dans les jeunes feuilles de colza après une carence azotée (d'après Desclos *et al.*, 2008).

du gène *AtWSCP* de façon différente de celle dans les tissus floraux. Le transcrit *AtWSCP*, régulé négativement par *NTT* et positivement par *HEC*, est exprimé dans les plantules étiolées au niveau de la crosse apicale, structure dont la fonction est de protéger les plantules au cours de l'émergence hors du sol. Cela laisse présager un rôle protecteur de la protéine *AtWSCP* au cours de la skotomorphogénèse (Boex-Fontvieille *et al.*, 2016a). Cette fonction serait liée à la capacité d'*AtWSCP* à s'associer à RD21 et à l'inhiber (Boex-Fontvieille *et al.*, 2015b, 2016a; Rustgi *et al.*, 2017).

#### **IV.4. Implication des WSCP II dans la réponse aux contraintes environnementales**

Sur la base de leurs expressions et localisations tissulaires, les WSCP II pourraient être impliquées dans les mécanismes de réponses aux contraintes environnementales. En l'absence de contrainte environnementale, les WSCP II seraient stockées dans les ER-bodies qui pourraient fusionner vers la vacuole lors de l'exposition aux stress (Takahashi *et al.*, 2012). Les ER-bodies peuvent être de type constitutif et leur nombre diminue avec développement foliaire ou de type inductible en réponse à des conditions de stress (Ogasawara *et al.*, 2009; Yamada *et al.*, 2011; Horibe *et al.*, 2013).

##### ***IV.4.1. Stress biotiques***

La nature des contraintes induisant l'apparition des ER bodies (blessure ou le traitement hormonal à l'acide jasmonique ou à l'éthylène) est concordante avec celles provoquant la surexpression de certaines WSCP II comme BnD22 ou *AtWSCP* (Desclos *et al.*, 2008; Ogasawara *et al.*, 2009; Gotté *et al.*, 2015; Boex-Fontvieille *et al.*, 2016a,b). Les ER-bodies sont associés à diverses réactions protectrices de la plante notamment à l'immunité innée et à la défense face à des stress biotiques. Ces structures accumulent la  $\beta$ -glucosidase, enzyme hydrolysant les glucosides comme la scopoline et l'esculine en produits toxiques pour les ravageurs et les pathogènes (Nagano *et al.*, 2005; Ahn *et al.*, 2010). Dans les tissus non endommagés, l'enzyme ne peut agir du fait de la compartimentation subcellulaire avec ses activateurs et substrats localisés dans la vacuole et le cytosol (Yamada *et al.*, 2011). La rupture des membranes internes des organites résultant des dommages cellulaires causés par un pathogène ou un ravageur, entraîne la libération des  $\beta$ -glucosidases des ER-bodies qui vont ainsi pouvoir hydrolyser les glucosides. Sur la base de ce modèle, Takahashi *et al.* (2012, 2013) ont émis l'hypothèse d'un mécanisme équivalent pour les WSCP II. En situation non contraignante, les WSCP sont isolées des Chls. A la suite d'un stress, les WSCP II seraient libérées des ER-bodies par rupture des membranes et se lieraient aux Chls libres (Fig. 15A). Ce mécanisme pourrait empêcher l'amplification des dommages photo-oxydants cellulaires en raison des propriétés photoprotectrices des WSCP II diminuant ainsi la production de ROS dérivées des Chls excitées (Schmidt *et al.*, 2003; Horigome *et al.*, 2007). Cette hypothèse est supportée par le phénotype du tabac surexprimant

BolWSCP qui montre un potentiel plus faible à générer des ROS et une faible activité de la peroxydase (Damaraju *et al.*, 2011). *A contrario*, il a été démontré *in vitro* une production non négligeable de ROS suite à l'illumination des complexes WSCP-II-Chls, qui seraient alors plutôt considérés comme une source de ROS de longue durée en raison de leur photo-stabilité (Agostini *et al.*, 2017; Palm *et al.*, 2019). Dans ce contexte, les WSCP-II seraient impliquées dans les réponses aux stress biotiques, abiotiques ainsi que de la mort cellulaire programmée *via* la voie de signalisation par les ROS (Laloi et Havaux, 2015; Agostini *et al.*, 2017; Girr *et al.*, 2020). La fonction potentielle des WSCP-II dans les processus de défense *via* la signalisation par les ROS ne remet pas en cause le rôle des ER-bodies lors de dommages cellulaires.

La fonction de protection des WSCP-II face à des stress biotiques *in planta* est à ce jour très peu documentée. Seule AtWSCP a fait l'objet d'études montrant sa participation à la protection des plantules au moment de leur émergence et de leur verdissement contre les herbivores arthropodes (Boex-Fontvieille *et al.*, 2015b, 2016; Fig. 15B-C). Cette protection passe par un mécanisme d'inhibition-activation de RD21, une cystéine protéase impliquée en outre dans les réponses de défense (Boex-Fontvieille *et al.*, 2015b, 2016a; Rustgi *et al.*, 2017). De plus, cette protéase inducible par le stress partage la même localisation subcellulaire que les WSCP-II (Matsushima *et al.*, 2002; Takahashi *et al.*, 2012, 2013b, 2015b).

#### **IV.4.1. Stress abiotiques**

Contrairement à AtWSCP, BnD22 est exprimée dans les jeunes feuilles où elle représente jusqu'à 9% des protéines solubles en fonction des conditions environnementales (Ilami *et al.*, 1997; Damaraju *et al.*, 2011). Ainsi, elle est fortement accumulée dans les jeunes feuilles lorsque les plantes sont soumises à des situations de stress tels que le déficit hydrique (Reviron *et al.*, 1992), le stress salin (Reviron *et al.*, 1992), la carence minérale en azote (Fig. 15D, Desclos *et al.*, 2008 ; Etienne *et al.*, 2007) ou en soufre (D'hooge *et al.*, 2013), ou suite à un traitement au méthyl jasmonate (Desclos *et al.*, 2008). Une relation inverse entre la sénescence et la présence de BnD22 a été établie dans les feuilles de tabac surexprimant BolWSCP (Damaraju *et al.*, 2011) ou celles de colza soumises ou non à une limitation en azote (Etienne *et al.* 2007). La réponse à la contrainte diffère selon le stade de développement des feuilles. Ainsi, la carence en N provoque une accélération de la sénescence des vieilles feuilles et un retard d'entrée en sénescence des feuilles plus jeunes comparées aux feuilles de plantes correctement alimentées en N. L'entrée prématurée en sénescence des vieilles feuilles est corrélée à une disparition rapide de BnD22 alors, qu'au contraire, le retard d'initiation de sénescence dans les jeunes feuilles est corrélé à une accumulation de BnD22. Ces résultats suggèrent que chez le colza soumis à une privation en N, la protéine BnD22 participerait à la fois à l'inhibition des protéases et à la protection des chlorophylles au sein de jeunes feuilles afin de maintenir le plus longtemps

possible leur intégrité physiologique dans l'attente de conditions de disponibilité en N plus favorables à la croissance (Etienne *et al.*, 2007; Desclos *et al.*, 2008). Ce constat est à rapprocher avec les travaux de James (2018) qui a démontré un rôle possible de BolWSCP dans le ralentissement de la progression de la sénescence induite par l'obscurité, et les travaux de Reviron *et al.* (1992) montrant, suite à un déficit hydrique, une moindre protéolyse dans les jeunes feuilles accumulant BnD22 en comparaison avec les feuilles plus âgées dans lesquelles BnD22 est absente.



## CHAPITRE I :

### CARACTERISATION FONCTIONNELLE DE BnD22 *IN VITRO*

---

#### Presentation

Ce premier chapitre repose sur la rédaction des deux premiers articles dont l'un est sous presse et l'autre est en cours de finalisation et sera soumis très prochainement :

**Article 1 :** Bouargalne Y., Raguénès-Nicol C., Guilbaud F., Cheron A., Clouet V., Deleu C. and Le Cahérec F. (2022). New insights into chlorophyll-WSCP (water-soluble chlorophyll binding proteins) interactions: the case study of BnD22 (*Brassica napus* Drought 22 kDa) (Plant Physiology and Biochemistry, sous presse)

**Article 2 :** Bouargalne Y., Guilbaud F., Macherel D., Delalande O., Deleu C. and Le Cahérec F. (2022). *Brassica napus* Drought 22 kD (BnD22), a water-soluble chlorophyll protein acting simultaneously as a cysteine protease inhibitor and chlorophyll-binding protein (en preparation).

**Contexte :** Les WSCP, spécifiques aux Brassicacées, sont des protéines non photosynthétiques ayant potentiellement la capacité de se lier aux chlorophylles et d'inhiber les protéases. Bien que leurs fonctions physiologiques soient à ce jour très peu documentées, certaines sont surexprimées en réponse à des contraintes abiotiques ce qui laisse supposer leur implication dans la réponse aux stress. Alors que les WSCP sont subdivisées en sous-classes IIA et IIB selon les ratios Chla/Chlb retrouvés dans les complexes natifs WSCP-Chls, leurs affinités à la Chla *versus* la Chlb ne sont pas encore précisément définies. Nous avons focalisé nos études sur la protéine BnD22, exprimée dans les jeunes feuilles de colza et accumulée sous stress hydrique. Cette protéine est répertoriée comme WSCPIIA alors que ses caractéristiques en tant qu'inhibiteur de protéases restent sujettes à controverse et sa capacité de liaison aux Chls n'a jamais été décrite.

**Objectif :** Les travaux présentés dans les deux articles ont pour objet d'étudier la capacité de BnD22 à se lier aux Chla et/ou Chlb (article 1) et/ou à inhiber les protéases (à sérine et/ou à cystéine) (article 2).

**Démarche expérimentale :** La stratégie a consisté à étudier *in vitro* les propriétés d'inhibition de protéases et de liaison aux Chls de BnD22-HisTag produite en système bactérien.

- L'aptitude de BnD22 à se lier à la Chla et/ou Chlb a été évaluée par égard à la formation des complexes oligomériques BnD22-Chls suivie par électrophorèse en condition non dénaturante. L'élution des complexes BnD22-Chls a permis l'obtention des complexes purifiés.

- L'état d'oligomérisation des complexes BnD22-Chls a été déterminé par chromatographie d'exclusion en collaboration avec Jean-François Hubert (IGDR, UMR 6290 CNRS-Université Rennes 1).
- Dans le but de valider l'interaction de BnD22 avec les Chls et de déterminer sa sélectivité-affinité envers la Chla ou la Chlb, nous avons développé une nouvelle méthode dérivée de la MicroScale Thermophoresis (MST), basée sur les taux de photoblanchiment des complexes BnD22-Chl en fonction des proportions en Chl/BnD22. Ce travail a été mené sur la plateforme de l'UMS Biosit (CNRS 3460-Université de Rennes 1) en collaboration avec Céline Raguénès-Nicol (IGDR, UMR 6290 CNRS-Université Rennes 1).
- Les activités anti-protéolytiques de BnD22 ont été menées contre deux protéases à serine, la trypsine et la chymotrypsine, et contre une protéase à cystéine, la papaine. Les activités anti-protéolytiques de BnD22 associée à la Chl ont été également testées.
- Des approches de modélisation structurale et de docking moléculaire ont été réalisées pour identifier plus précisément les zones d'interaction entre BnD22 et la papaine. Ce travail a été mené en collaboration avec Olivier Delalande (IGDR, UMR 6290 CNRS-Université Rennes 1).
- Pour tenter d'apporter des éléments de réponse quant au rôle physiologique de BnD22, sa localisation subcellulaire a été entreprise par expression transitoire dans des protoplastes de colza. Ce travail a été mené en collaboration avec David Macherel (IRHS, Institut Agro-INRAE, Angers).

**Résultats :** Malgré une localisation qui s'avère non chloroplastique, la BnD22 est capable de se lier à la Chla et/ou à la Chlb pour former des complexes tétramériques BnD22-Chls. Les affinités de BnD22 pour les Chls montrent une préférence de liaison à la Chla. *In vitro*, BnD22 ne présente pas de capacité d'inhibition envers les protéases à sérine mais une inhibition effective des protéases à cystéine. Cette fonction est tout à fait cohérente avec sa localisation vacuolaire mise en évidence par microscopie confocale. BnD22 peut assurer simultanément les deux fonctions de liaison à la Chl et à la papaine avec un pouvoir d'inhibition optimisé lorsqu'elle est complexée aux Chls résultant d'une interaction maximisée entre BnD22 et la protéase au sein des complexes.

**Conclusion :** Sur la base des résultats obtenus pour BnD22, nous pourrions imaginer que la double fonction de liaison à la Chl et d'inhibition de protéases, indépendamment ou simultanément, puisse être étendue à l'ensemble des WSCP. Cette hypothèse ne pourra être confirmée qu'à travers l'étude d'autres WSCP de cette famille. Pour l'heure, l'un des principaux défis sera de confirmer l'existence de cette double fonction *in planta*.

**Article 1:**

**New insights into chlorophyll-WSCP (water-soluble chlorophyll binding proteins) interactions: the case study of BnD22 (*Brassica napus* drought-induced 22 kDa)**





## New insights into chlorophyll-WSCP (water-soluble chlorophyll proteins) interactions : The case study of BnD22 (*Brassica napus* drought-induced 22 kDa)

Youssef Bouargalne<sup>a</sup>, Céline Raguénès-Nicol<sup>b</sup>, Florian Guilbaud<sup>a</sup>, Angélique Cheron<sup>c</sup>,  
Vanessa Clouet<sup>a</sup>, Carole Deleu<sup>a</sup>, Françoise Le Cahérec<sup>a,\*</sup>

<sup>a</sup> Univ Rennes 1, INRAE, Institut Agro, IGEPP - UMR 1349, 35653, Le Rheu, France

<sup>b</sup> Univ Rennes 1, INSERM, EHESP, Irset - UMR 1085, 35043, Rennes, France

<sup>c</sup> Univ Rennes1, CNRS, IGDR - UMR 6290, 35000, Rennes, France

### ARTICLE INFO

#### Keywords:

Chl binding constants

Photobleaching

Microscale thermophoresis

### ABSTRACT

The water-soluble chlorophyll-proteins (WSCP) of class II from *Brassicaceae* are non-photosynthetic proteins that bind chlorophylls (Chls) and chlorophyll derivatives. Their physiological roles, biochemical functions and mode of action are still unclear. It is assumed that the WSCPs have a protection function against Chl photodamage during stressful conditions. WSCPs are subdivided into class IIA and class IIB according to their apparent Chla/b binding ratio. Although their Chla/Chlb binding selectivity has been partly characterized, their Chl affinities are not yet precisely defined. For instance, WSCPs IIA do not show any Chl binding preference while WSCPs IIB have greater affinity to Chlb. In this study, we present a novel method for assessment of Chl binding to WSCPs based on the differences of Chl photobleaching rates in a large range of Chl/protein ratios. The protein we have chosen to study WSCP is BnD22, a WSCP IIA induced in the leaves of *Brassica napus* under water deficit. BnD22 formed oligomeric complexes upon binding to Chla and/or Chlb allowing a protective effect against photodamage. The binding constants indicate that BnD22 binds with high affinity the Chls and with a strong selectivity to Chla. Moreover, depending of Chl/protein ratio upon reconstitution, two distinct binding events were detected resulting from difference of Chl stoichiometry inside oligomeric complexes.

### 1. Introduction

Class II water-soluble chlorophyll proteins (WSCP), specific to *Brassicaceae*, strikingly differ from other Chl-binding proteins. They are water soluble, contain no carotenoid and are not involved in photosynthesis. Class II WSCPs can be categorized into two sub-classes: the class IIA with Chla/b ratio higher than 6 and the class IIB with Chla/b ratio lower than 3.5 (Sato et al., 2001; Prabakar et al., 2020). Although their biological function and mode of action remain unclear, it is assumed they have a protective function throughout environmental constraints. For example, some of them are induced under stressful abiotic conditions (Downing et al., 1992; Reviron et al., 1992; Annamalai and Yanagihara, 1999; Desclos et al., 2008). Furthermore, it was shown that AtWSCP from *Arabidopsis thaliana* was involved in defenses

against herbivores and played a role in programmed cell death during flower development (Boex-Fontvieille et al., 2015). As members of Kunitz-type protease inhibitors, WSCPs have potential function of inhibition against serine and/or cysteine proteases. Thus, AtWSCP and BnD22 (*Brassica napus* drought-induced 22 kDa) are able to inhibit cysteine and serine proteases respectively (Halls et al., 2006; Desclos et al., 2008; Boex-Fontvieille et al., 2015a; Rustgi et al., 2017). However, this ability has never been investigated for the widely studied WSCPs, LvWSCP from *Lepidium virginicum* and BoWSCP from *Brassica oleracea*. Because WSCPs are able to extract Chl from thylakoids, it was proposed that they could act as Chl carriers during reorganization process of photosynthetic apparatus under stress conditions and as Chl scavenger during cell damage (Takahashi et al., 2012, 2013b). Although the Chls bound to WSCP are partially shielded against photodamage, the benefit of Chl-binding to WSCP remains unclear (Damaraju et al., 2011).

**Abbreviations:** IR, infrared laser; CD, circular dichroism.

\* Corresponding author. UMR 1349 IGEPP (Institut de génétique et protection des plantes), INRAE, Institut Agro, Université de Rennes 1, BP 35327, 35650, Le Rheu, France.

E-mail address: [francoise.le-caherec@univ-rennes1.fr](mailto:francoise.le-caherec@univ-rennes1.fr) (F. Le Cahérec).

<https://doi.org/10.1016/j.plaphy.2022.03.023>

Received 15 December 2021; Received in revised form 1 March 2022; Accepted 18 March 2022



0.04% Nonidet-P40 (NP40) was mixed with Chla or Chlb (Sigma) solubilized in 100% ethanol to reach 5, 10, 15, 20, 40 or 60  $\mu\text{M}$  final concentrations (0, 0.25, 0.5, 1, 2 or 3 ratio Chls/protein) in a final volume of 100  $\mu\text{L}$ . NP40 was chosen for its suitability for the MST analyses. The mixtures were incubated in the dark at 4 °C for 30 min and subjected to EMSA using a detergent free 12% PAGE. Chl fluorescence was detected using a fluorescence imaging system (Fluorcam FC 800-O, Photon System Instruments) and the proteins were visualized by Coomassie blue staining.

For the absorption spectra, BnD22 was reconstituted in excess of Chls. 20  $\mu\text{M}$  recombinant protein were mixed with 60  $\mu\text{M}$  Chla or Chlb or mix Chla/Chlb at three different Chl/protein ratios (1/1, 3/1, 10/1). After binding step, the mixtures were subjected to 12% native PAGE and the greenish band corresponding to the reconstituted complexes BnD22-Chls was eluted from gel in 50 mM Tris-HCl (pH 7.5) buffer, 150 mM NaCl. The absorption spectra were measured between 300 and 750 nm using NanoDrop™ One (Thermo Scientific).

### 2.3. Détermination of Chl/protein stoichiometry

Based on results published by Palm et al. (2017), 5-fold Chl molar excess and 40-fold BnD22 molar excess were chosen for two binding conditions, to determine the molar Chl/protein ratio from absorption spectra of purified BnD22-Chla or BnD22-Chlb complexes. Chl concentrations were determined based on extinction coefficients of Chls bound to BnD22 and the absorption values at the Qy absorption maxima. Protein concentration was determined using absorption values at 280 nm taking into account the contributions of Chla or Chlb to the absorption at this wavelength and BnD22 extinction coefficients.

### 2.4. Development of a method based on use of microscale thermophoresis (MST)

To follow BnD22/Chl interactions, the intrinsic fluorescence of Chls was monitored to enable a label-free MST analysis. All MST equipment: Monolith NT.115 instrument, Monolith NT™ capillaries (standard or premium) and software (MO.control, NT.control, MO.affinity analysis) used are from NanoTemper technologies GmbH. Measurements were performed at the Spectroscopies-DCTP core facility (UMS Biosit, Université de Rennes 1- Campus de Villejean- 35043 RENNES Cedex, FRANCE).

#### 2.4.1. Buffer selection and optimal ligand concentration range

To determine the optimal buffer composition for Chl stability and

$$f(C) = \text{unbound} + \frac{(\text{bound} - \text{unbound}) \times (C + C_{\text{target}} + Kd - \sqrt{(C + C_{\text{target}} + Kd)^2 - 4 \times C \times C_{\text{target}}})}{2 \times C_{\text{target}}} \quad \text{Equation 2}$$

homogeneity during measurement, fluorescence capillary scan and MST signals were recorded in several buffers: Tris-buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl) or phosphate-buffer (50 mM (pH 7.5), 150 mM NaCl) supplemented with non-ionic detergents: 0.04% NP40, 0.3% n-dodecyl  $\beta$ -D-maltoside (DDM), 1% octyl  $\beta$ -D-glucopyranoside (OG) and 20% octylphenoxy poly(ethyleneoxy)ethanol (IGEPAL). Chla and Chlb solubilized in 100% ethanol were diluted in the different buffers at final concentrations of 50 and 80 nM respectively to yield detectable fluorescent signals. The samples were filled into standard capillaries for measurement.

#### 2.4.2. Preparation of reaction mixtures for MST experiments

For the binding assay, a constant concentration of Chla (50 nM) or

Chlb (80 nM) was mixed with a serial dilution of non-labeled BnD22 using Tris-buffer with 0.04% NP40 as binding buffer. The concentrations of BnD22 varied from 80  $\mu\text{M}$  to 0.124 nM (2.44-fold serial dilution) and from 80  $\mu\text{M}$  to 2.24 nM (2-fold serial dilution) with Chla and Chlb respectively. All binding mixtures were incubated for 30 min in the dark at 4 °C and filled into standard treated (Chla) or premium (Chlb) capillaries for MST measurements. All experiments were repeated three times with different purification batches of BnD22. As negative controls, trypsin and lysozyme (concentrations from 80  $\mu\text{M}$  to 2.24 nM), non-binding and unrelated Chl proteins, were used instead of the recombinant BnD22 protein.

#### 2.4.3. MST measurements, data acquisition and analysis

All MST measurements were conducted on a Monolith.NT115 instrument, at 100% LED (light emitting diode) power and 40% MST power. The data for Chla experiments were acquired with MO.control v2.3 software using nano-red fluorescence channel with laser off/on times of 5.5 and 6 s respectively. For Chlb MST experiments, NT.control v2.1.31 software with blue-red fluorescence channel with 20 s/5s laser off/on times was used (Fig. S1). The most appropriate wavelength for excitation/emission of Chls fluorescence in the Monolith.NT115 instrument was beforehand determined by fluorescence emission spectra after excitation at 470 nm or 625 nm (Fig. S1). The maximum emission wavelengths of Chla and Chlb ( $\lambda_{\text{max}}$  675 and 657 nm respectively) are compatible with Monolith NT.115 red detector.

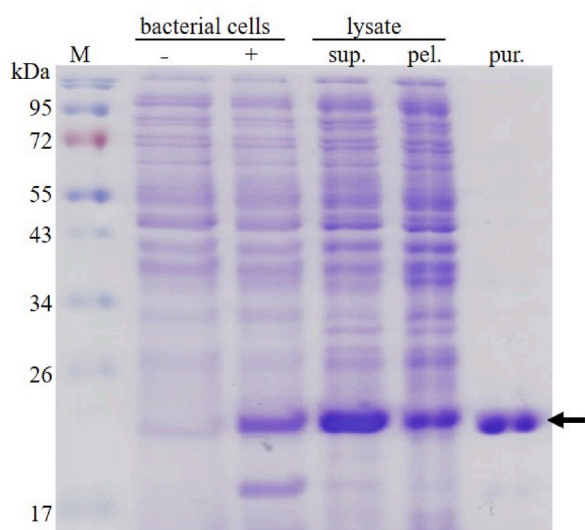
The photobleaching observed during excitation (100% LED) has been used for analyses. Ligand-induced photobleaching rates on the first 5 s of MST traces were calculated using the expert mode bleaching rate in the MO.affinity analysis software to derive binding affinities. The bleaching rates were normalized to response amplitude by equation (1), and plotted against the concentration of the BnD22. To determine the dissociation constant (Kd), the data were fitted to the Kd binding model (equation (2)) which describes a reversible molecular interaction with a 1:1 stoichiometry according to the law of mass action. The half-maximal effective concentration (EC50) and cooperativity Hill coefficient (nHill) were obtained from fitting the data with the Hill equation (equation (3)).

$$\text{response amplitude} = 1 - \left[ \frac{\text{value}(C) - \text{max value}}{\text{min value} - \text{max value}} \right] \quad \text{Equation 1}$$

where value (C) is the photobleaching value measured for the concentration C, max value is the photobleaching value for the unbound state (lowest BnD22 concentration) and min value is the photobleaching value for the fully bound state (highest BnD22 concentration).

$$f(C) = \text{unbound} + \left[ \frac{\text{bound} - \text{unbound}}{1 + \left[ \frac{\text{EC50}}{C} \right]^{n\text{Hill}}} \right] \quad \text{Equation 3}$$

where f(C) is the response amplitude at a given BnD22 concentration; unbound is the response amplitude of free Chl; bound is the response amplitude of the BnD22-Chl complex; Ctarget is the final concentration of Chl in the assay.



**Fig. 1.** Subcellular fractionation and purification of BnD22 expressed in *E. coli*. Proteins from cell extracts of uninduced *E. coli* (–) or after IPTG-induced BnD22 expression (+), from supernatant (sup.) or pellet (pel.) protein fractions of sonicated *E. coli*, and purified protein using Ni<sup>2+</sup> affinity chromatography (pur.) were separated by 15% SDS-PAGE and stained with Coomassie Blue. The protein markers are indicated as M. Black arrow indicates the position BnD22.

### 3. Results

#### 3.1. Purification of recombinant BnD22 expressed in *E. coli*

Recombinant mature form of BnD22, fused to a hexahistidine (His-tag) in C-terminal, was produced in *E. coli* under IPTG induction (Fig. 1). The produced 20 kDa protein was mainly soluble (about 60%). A high level of purified recombinant protein was obtained from the soluble fraction through nickel affinity chromatography, with more than 97% purity assessed by densitometric analysis (Fig. 1). A final large yield of 40 mg per liter of bacterial culture medium was obtained and purified BnD22 was concentrated to 2 mg mL<sup>-1</sup> (~0.1 mM).

#### 3.2. Qualitative assessment of BnD22 binding to Chls

Binding experiments using different concentrations of Chla or Chlb were performed in presence of a nonionic detergent NP40 instead of methanol (Takahashi et al., 2012), ethanol (Palm et al., 2017) or Triton-X114 (Palm et al., 2018) previously reported. In our study, the nonidet P40 was chosen during the binding conditions, for i) its

suitability for the MST analyses to provide an adequate level of fluorescence with low amounts of Chla or Chlb (Table 1), ii) its intrinsic properties: a low critical micelle concentration (0.06 mM against 0.2 mM for Triton-X114) that minimize the detergent concentration used for Chl solubility and a cloud point of 45–50 °C (23 °C for Triton-X114) that avoid heterogeneity of samples with the formation of two phases at room temperature.

Subsequently, the formation of BnD22-Chl complexes was visualized by EMSA (Fig. 2). The purified monomeric BnD22 was produced under an active form as confirmed by its ability to bind Chls. Dependently on the concentration of Chl, the binding induces a structural change with the formation of BnD22-Chl oligomeric complexes. When BnD22 was incubated at least with an equimolar level of Chla or Chlb, almost all the BnD22 formed oligomers, presumably on a tetrameric state. A minor additional supra-oligomeric complex was also detected. All oligomeric complexes emitted fluorescence refuting the possibility that oligomerization may result artificially from aggregates of BnD22 without Chl. Moreover, no difference in the behaviors of BnD22 reconstituted with Chla or Chlb was detected. This result might support the idea that BnD22 would not have Chl preference for either Chla or Chlb in our binding conditions. This was confirmed by absorption spectrum data (Fig. 3). The absorption spectra of BnD22 reconstituted with equal or 3 molar ratios of Chla/b (Fig. 3C and D) were a mix of the spectra of BnD22 reconstituted with Chla or Chlb only (Fig. 3A and B). This indicates that BnD22 would bind simultaneously both Chla and b. When the reconstitution was performed with a tenfold excess of Chla compared to Chlb, the absorption spectrum was almost identical to that of BnD22 reconstituted with Chla only (Fig. 3A, E). These spectra analyses suggest that there is no clear preferential binding of BnD22 to Chla or Chlb.

#### 3.3. Determination of the binding constant for BnD22-Chls

To assess the binding specificity of WSCPs to Chla or Chlb, interactions were investigated through a Monolith.NT115 instrument using the intrinsic fluorescence of Chls. The difficulty was to define buffer conditions to follow the binding between the hydrophobic molecule target (Chl) and the soluble BnD22 protein ligand.

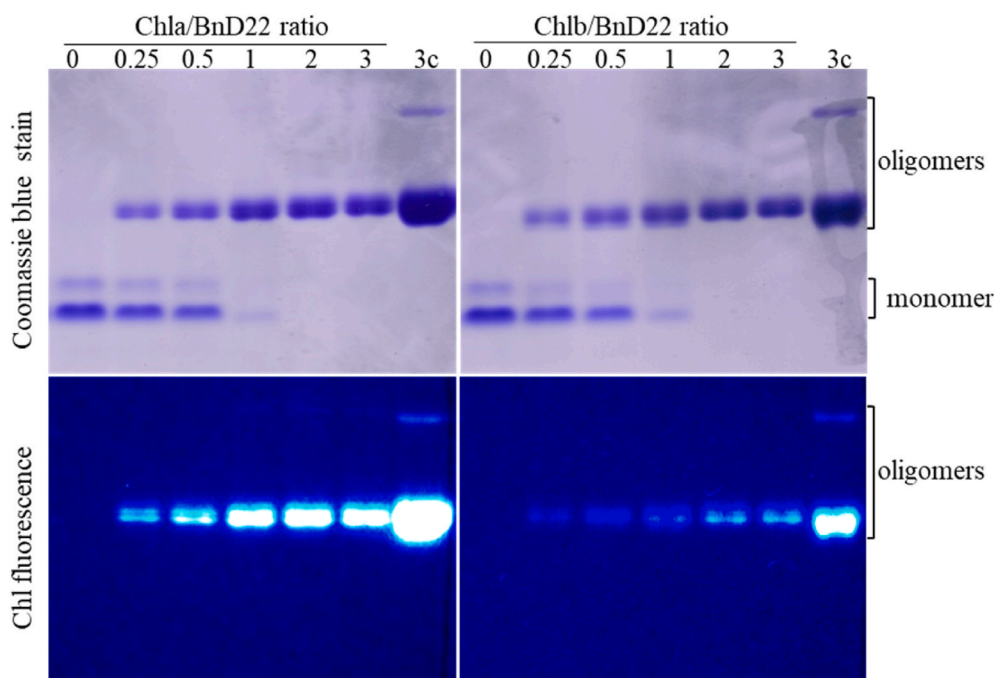
##### 3.3.1. Conditions for Chl fluorescence detection

The supplementation of buffers with detergent is essential to solubilize Chls and prevent their aggregations in aqueous solution. Two buffers Tris-HCl or sodium phosphate, supplemented with several non-ionic detergents (NP40, DDM, OG, IGEAL) were tested in presence of a constant concentration of Chla or Chlb (Table 1). None of these binding buffers led to Chl aggregation or Chl adsorption to the capillary walls (data not shown). Whatever buffer nature, the chosen concentrations of Chla at 50 nM or Chlb at 80 nM allow an adequate fluorescence

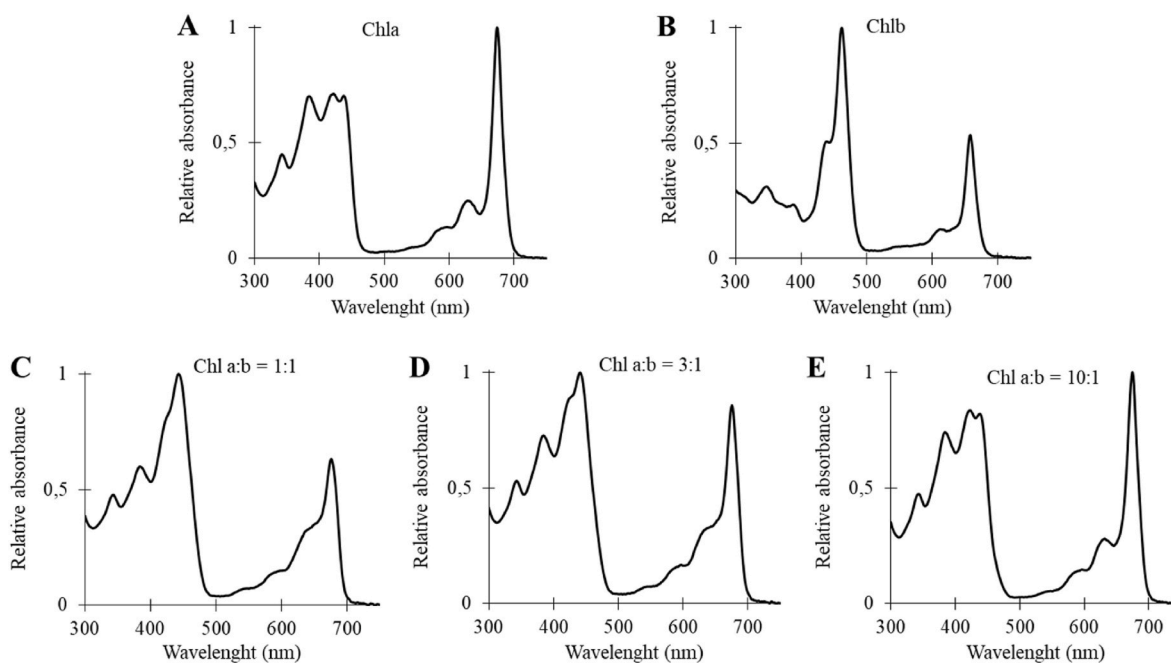
**Table 1**

Effect of binding buffers on initial fluorescence (IF) and bleaching rate (BR) of Chla and Chlb. 50 nM Chla or 80 nM Chlb were added to different buffers and filled into standard MST capillaries. Experiments were performed at 100% LED. IF was measured and the BR calculated in the first 5 s of excitation. Asterisks indicate significant effect of buffer or detergent for each variable in Kruskal-Wallis test (\*P ≤ 0.05; \*\*P ≤ 0.01; NS = not significant).

Buffer	Detergent	initial fluorescence (counts)		bleaching rate (%.s <sup>-1</sup> )		IF/BR ratio (%)	
		Chla	Chlb	Chla	Chlb	Chla	Chlb
Tris-HCl 50 mM, pH 7.5 150 mM NaCl	0.59 mM NP40	261 ± 2.8	219 ± 7.8	8.9 ± 0.2	28.0 ± 0.5	29.4 ± 0.3	7.8 ± 0.1
	5.87 mM DDM	241 ± 3.8	165 ± 16.0	8.4 ± 0.7	25.1 ± 0.3	29.0 ± 3.0	6.6 ± 0.6
	34.2 mM OG	293 ± 28.2	207 ± 12.1	11.8 ± 0.1	28.8 ± 0.5	24.9 ± 2.6	7.2 ± 0.3
	3.33 mM IGEAL	242 ± 6.3	184 ± 11.0	9.4 ± 0.1	28.4 ± 0.8	25.6 ± 0.5	6.5 ± 0.6
sodium phosphate 50 mM, pH 7.5 150 mM NaCl	0.59 mM NP40	262 ± 1.9	214 ± 26.9	9.4 ± 0.3	28.9 ± 4.1	28.0 ± 0.6	7.5 ± 2.0
	5.87 mM DDM	243 ± 5.5	176 ± 23.5	8.4 ± 0.1	29.3 ± 2.0	29.0 ± 0.7	6.0 ± 1.2
	34.2 mM OG	251 ± 4.0	240 ± 0.4	11.2 ± 0.3	29.4 ± 3.5	22.4 ± 0.2	8.2 ± 1.0
	3.33 mM IGEAL	234 ± 7.2	173 ± 16.6	9.1 ± 0.1	32.2 ± 2.5	25.6 ± 0.6	5.4 ± 0.9
	Buffer	NS	NS	NS	NS	NS	NS
	Detergent	**	**	**	NS	**	*



**Fig. 2.** Conformational changes monitoring of BnD22 upon binding to Chla (left) or Chlb (right). 20  $\mu$ M purified BnD22 were mixed with 0, 5, 10, 15, 20, 40, 60  $\mu$ M of Chla or Chlb (0, 0.25, 0.5, 1, 2 or 3 Chl/protein molar ratio) in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). After 30 min of binding, BnD22-Chls were visualized by electrophoretic mobility on 12% native PAGE stained with Coomassie blue (upper panel) or by detection of Chl fluorescence (lower panel). To visualize the high molecular mass oligomer, the complexes obtained at the Chl/BnD22 = 3 condition were concentrated (3c)



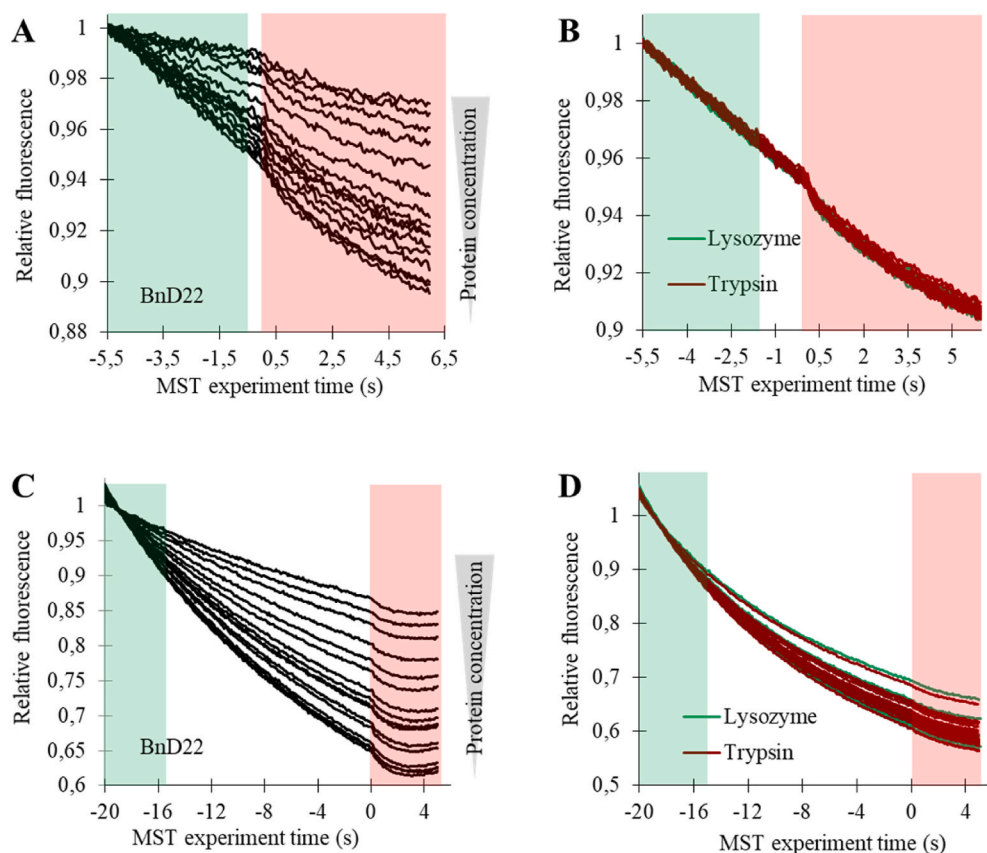
**Fig. 3.** UV-visible absorption spectra of reconstituted BnD22 with Chls. 20  $\mu$ M purified BnD22 were mixed with 60  $\mu$ M Chla (A), Chlb (B), or three different Chla/Chlb mixtures: 1/1 (C), 3/1 (D), 10/1 (E) ratio in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). After EMSA, the BnD22-Chl complexes were eluted from gels and their UV-visible absorption spectra were performed between 300 and 750 nm.

signal (IF) around 240 counts for Chla and 184 counts for Chlb consistent with the recommendations of Monolith NT.115 manufacturer (Table 1). IF was significantly influenced by nature of the detergent and thus NP40 and OG were the more appropriated detergents to allow highest IF. During excitation, a Chl bleaching was observed. The bleaching rates (BR) were more important for Chlb (25–32%.s<sup>-1</sup>) than for Chla (8–12%.s<sup>-1</sup>). As IF, bleaching was also significantly influenced by nature of the detergent and, could not be prevented. Our choice of binding buffer focused on the Tris buffer with NP40 for which the IF/BR ratio was the highest and because it requires a low concentration to

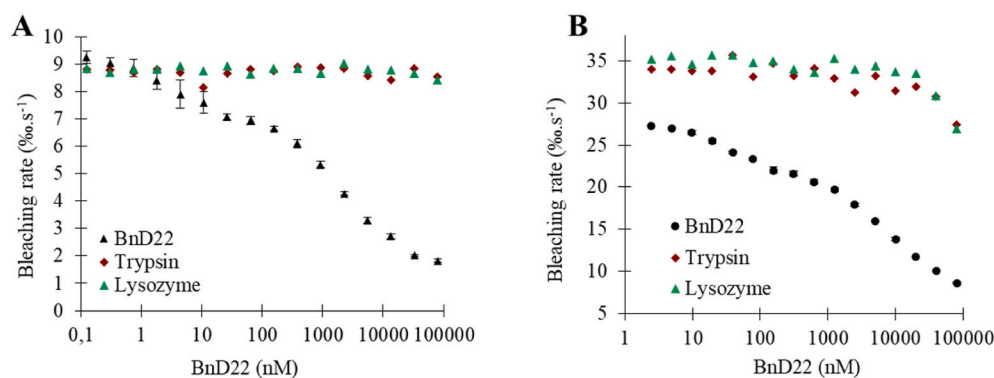
allow micelle formation.

### 3.3.2. Biphasic behavior of BnD22-Chl interaction

Serial dilutions of BnD22 were incubated with constant concentration of Chla or Chlb in Tris buffer supplemented with NP40. After loading in capillaries, cap scans showed little variations in IF (less than 10%), attesting the absence of ligand-dependent fluorescence enhancement or quenching effects (Figs. S2A and B). Furthermore, for both Chls, capillaries profiles showed gaussian distributed fluorescence peaks attesting absence of sticking or adsorption effects (Figs. S2C and D). MST



**Fig. 4.** Unprocessed MST traces of Chla (A, B) or Chlb (C, D) binding to BnD22 (A, C) or lysozyme or trypsin (B, D). Constant concentrations of Chla (50 nM) or Chlb (80 nM) were mixed with a titration series of BnD22 or non-relevant proteins (trypsin or lysozyme) used as negative controls in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). The protein concentrations used varied from 80 to 0.124  $\mu$ M. After 30 min of binding, MST experiments were performed at 100% LED and 40% MST power. Red area defines MST IR laser application. Data from the green highlighted region (the first 5 s of excitation) were used to calculate the bleaching rates.



**Fig. 5.** Bleaching rates of Chla (A) or Chlb (B) in presence of BnD22 or negative controls (lysozyme and trypsin). Constant concentrations of Chla (50 nM) or Chlb (80 nM) were mixed with a titration series of BnD22 or negative controls (trypsin or lysozyme) in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). After 30 min of binding, MST experiments were performed at 100% LED and 40% MST power. The bleaching rates, calculated in the first 5 s of excitation, were plotted against protein concentrations.

**Table 2**

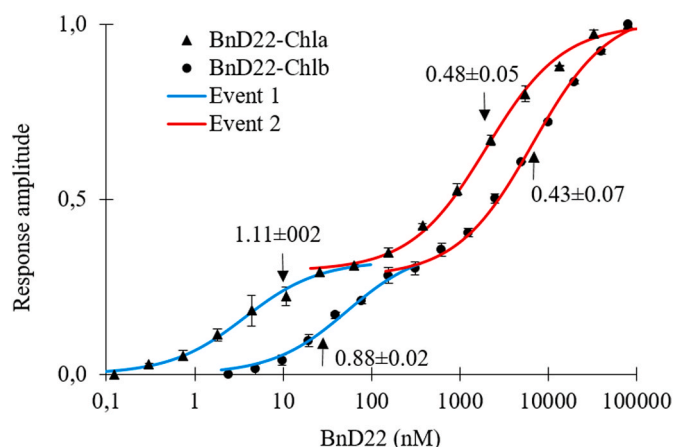
Effect of BnD22 on Chls photostability over excitation time. 50 nM Chla or 80 nM Chlb were mixed with titration series of BnD22 in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). After 30 min of binding, MST experiments were performed at 100% LED and 40% MST power. The bleaching rates (BR) were calculated in the first 5 s of excitation. Levels of photoprotection were determined by the difference between the highest and the lowest BR related to the highest BR. Data are expressed as the mean  $\pm$  SD (n = 3). Asterisks indicate significant differences between chlorophylls for each variable in the Student's *t*-test (\*\*\*P  $\leq$  0.001).

	photoprotection (%)		
	overall	event 1	event 2
BnD22-Chla	80.3 $\pm$ 0.3	33.2 $\pm$ 0.5	74.3 $\pm$ 0.6
BnD22-Chlb	68.4 $\pm$ 0.8	21.8 $\pm$ 2.3	61.4 $\pm$ 1.0
	***	***	***

traces were smooth without bumpiness indicating absence of aggregated molecules and good samples quality (Fig. 4).

As observed with Chls alone (Table 1), a Chl photobleaching was also detected in presence of BnD22 and was dependent on BnD22 concentrations (Fig. 4A and B). A photobleaching was also observed in presence of non-relevant proteins as trypsin or lysozyme, used as negative controls, but was not modified by their concentrations (Fig. 4B, D). High quantity of BnD22 reduced strongly the photobleaching from 9.2 to 1.8% $\cdot$ s $^{-1}$  for Chla, and from 27.2% $\cdot$ s $^{-1}$  to 8.6% $\cdot$ s $^{-1}$  for Chlb (34% $\cdot$ s $^{-1}$  for non-relevant proteins) (Fig. 5). This argues in favor of a protective effect of the BnD22-Chl binding against Chl photodamage. Moreover, it appeared that BnD22 protection was significantly higher for Chla (80.3  $\pm$  0.3%) than for Chlb (68.3  $\pm$  0.8%) (Table 2).

As the Chl photobleaching was BnD22-dependent, it prevented classical MST temperature-jump analysis. The binding parameters were so determined from the photobleaching rates, according to the recommendations of Monolith NT115 manufacturer. Response amplitude of



**Fig. 6.** Biphasic interactions between BnD22 and Chla or Chlb. Constant concentrations of Chla (50 nM) or Chlb (80 nM) were mixed with a titration series of BnD22 in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). After 30 min of binding, MST experiments were performed at 100% LED and 40% MST power. Response amplitude was calculated from the bleaching rates and plotted against BnD22 concentrations. Each binding event was fitted individually to define binding parameters (summarized in Table 3). Each value represents the mean  $\pm$  SD ( $n = 3$ ). For two binding conditions at 5-fold Chl molar excess (event 1) and 40-fold BnD22 molar excess (event 2), the stoichiometry of Chl/protein inside BnD22 complexes were determined from the absorption spectra of purified BnD22-Chla or BnD22-Chlb complexes (according to Palm et al., 2017) and indicated by arrows.

**Table 3**

Binding parameters of the interaction between BnD22 and Chls. 50 nM Chla or 80 nM Chlb were mixed with a titration series of BnD22 in binding buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40. After 30 min of binding, MST experiments were performed at 100% LED and 40% MST power. Response amplitude calculated from the bleaching rates were plotted against BnD22 concentrations (Fig. 6). Kd, EC50 and nHill values were calculated for each binding event. Each value represents the mean  $\pm$  SD ( $n = 3$ ). Asterisks indicate significant differences between chlorophylls for each variable in the Student's *t*-test (\*\* $P \leq 0.01$ ; \*\*\* $P \leq 0.001$ ; NS = not significant).

	Kd		EC50		nHill	
	event 1 (nM)	event 2 ( $\mu$ M)	event 1 (nM)	event 2 ( $\mu$ M)	event 1	event 2
BnD22-Chla	3.9 $\pm$ 2.2	2.0 $\pm$ 0.2	4.5 $\pm$ 3.2	2.1 $\pm$ 0.3	0.9 $\pm$ 0.2	0.8 $\pm$ 0.1
BnD22-Chlb	51.4 $\pm$ 8.6	6.7 $\pm$ 0.7	35.1 $\pm$ 9.8	8.1 $\pm$ 0.7	1.1 $\pm$ 0.2	0.8 $\pm$ 0.1
	***	***	**	***	NS	NS

Chls, calculated from photobleaching rates, were plotted against BnD22 concentrations (Fig. 6) and used to assess the binding constants of BnD22 to Chla or Chlb (Table 3). For both Chls, typical bi-phasic binding curves were obtained indicating that two binding events have occurred, each with a defined BnD22-Chls affinity. The first event (event 1) was occurred when the quantity of Chla or Chlb was in excess over the protein (low BnD22 concentrations), while the second binding event (event 2) occurred when Chl quantity was limited over protein (high BnD22 concentrations). For event 1, only estimated Kd values in the nM range could be determined. Anyway, it was possible to discriminate bindings between BnD22-Chla and BnD22-Chlb with Kds of  $3.9 \pm 2.2$  nM and  $51.4 \pm 8.6$  nM respectively, suggesting that BnD22 could preferentially bind the Chla (Table 3). This affinity of BnD22 to Chla, around 10 times higher than Chlb, was also supported by EC50 values. The bindings at the event 2 were characterized by affinities in order of  $\mu$ M, 1000-fold weaker than those of the event 1. The accurate Kd values of  $2.0 \pm 0.2$  and  $6.7 \pm 0.7$   $\mu$ M for Chla and Chlb respectively were in the same range than the EC50 values. These values revealed a higher affinity

(3-fold) of BnD22 for Chla than for Chlb. For both events, the Hill coefficients were around 1 indicating that the binding would be non-cooperative (Table 3).

According to Palm et al. (2017), the stoichiometry of Chl/protein inside complexes was determined (Fig. 6). When BnD22 was reconstituted with a 5-fold molar excess of Chla or Chlb (event 1), the Chl/protein ratio calculated was  $1.11 \pm 0.02$  and  $0.88 \pm 0.02$  respectively reflecting a binding stoichiometry of four Chl per tetrameric complex. When a 40-fold molar excess of BnD22 over Chls was used (event 2), the presence of only two Chls inside complex was outlined with Chl/protein ratio values of  $0.48 \pm 0.05$  for BnD22-Chla complexes and  $0.43 \pm 0.07$  for Chb complexes.

## 4. Discussion

In this study, WSCP-Chl interactions were monitored qualitatively by EMSA coupled with absorption spectra and quantified by decrease in photobleaching with a MST device. We focused on BnD22 protein which belongs to the class of WSCP IIA proteins. Recombinant mature BnD22 protein with histidine-tag in C-terminal was expressed in *E. coli*. As well established for numerous other WSCPs, the bacterial expression systems allowed high yields and high purity of recombinant target (Schmidt et al., 2003; Theiss et al., 2007; Takahashi et al., 2012, 2014).

### 4.1. BnD22 binds both Chla and Chlb

To our knowledge, this is the first report that demonstrates the ability of BnD22 to bind Chls. For this purpose, we used recombinant His-tagged BnD22 mature protein since it was shown by several authors that WSCP protein are appropriate for study their Chl interaction (Schmidt et al., 2003; Theiss et al., 2007; Takahashi et al., 2012, 2013a, 2013b, 2014). EMSA experiments and UV-visible absorption spectra revealed that BnD22 formed oligomer upon binding to Chla and/or Chlb. However, as shown for BoWSCPs, another WSCP IIA, no clear difference between Chla or Chlb in the binding behavior was detected by this method (Takahashi et al., 2012; Bednarczyk et al., 2015; Palm et al., 2018). To date, only LvWSCP, a WSCP of class IIB, showed a clear preference for Chlb compared to Chla (Takahashi et al., 2013a, 2013b; Palm et al., 2018).

### 4.2. MST derived method is a tool to characterize Chl-protein binding constants

To get more information about the Chla/Chlb selectivity, we have established a label free MST derived strategy to obtain direct binding constants between water-soluble BnD22 with highly hydrophobic Chls. MST approach has been successful in determining the binding constants for hydrophobic Chl-binding proteins as the light harvesting proteins (Mork-Jansson and Eichacker, 2018, 2019), but has never been used on WSCPs. Establishment of experimental conditions for introducing hydrophobic Chls into an aqueous protein environment compatible for MST measurement was challenging. In previous WSCP/Chl binding methods, Chls were provided either in organic solvents or detergents (Takahashi et al., 2012; Palm et al., 2017, 2018).

The presence of detergents is recommended for MST experiments since they may reduce the adsorption of the molecules to glass capillaries walls and the non-specific interactions driving to aggregation (Khavrutskii et al., 2013; Seidel et al., 2013). After comparison of results obtained for several detergents, we chose 0.04% NP40 providing an adequate level of fluorescence for as little as 50 nM Chla and 80 nM Chlb. These Chls concentrations were 20-fold lower than those previously used ( $\sim 1$   $\mu$ M) in a CD method where the detection limit of the method did not allow the determination of proper Kd values (Girr et al., 2020). The very low concentrations of Chl required for our proposed MST-derived analysis offer a real advantage by allowing accurate determination of dissociation constants for WSCPs-Chls interactions.



### 4.3. BnD22-Chls interactions depend on Chl/protein ratio

The dose-response curves of BnD22 binding Chls fit as a biphasic process, each phase representing a distinct binding event. Regardless Chl nature, when Chl was preponderant over proteins, corresponding to event 1, the affinity was very high, in the order of a few nM. These values were very close to the limit of detection 10 nM for the Monolith NT.115 device (Jerabek-Willemsen et al., 2014). To be more accurate, lower Chls concentrations are required (below the estimated Kd) and could only be analyzed with the Monolith NT Pico, an improved sensitivity instrument designed to detect low fluorophores concentrations (Entzian and Schubert, 2016). Under limiting Chl conditions, a second binding event (event 2) occurred with affinity in the order of  $\mu\text{M}$ , 1000-fold weaker than the first one.

This biphasic binding process could be explained by different affinities of the binding sites within the oligomeric complex, even if the Hill coefficients determined under our conditions indicate rather a lack of cooperative effect. The both event 1 and event 2 of binding between Chl and BnD22 could result also from a conformational change of the complexes. As reported for all studied WSCPs, the binding to Chls drives the formation of a tetramer complex, presumably the major complex observed in EMSA experiments. The additional supra-oligomeric complex detected could either be due to an hexamer conformational state as supported by Satoh et al. (1998) for a cauliflower WSCP, or could result from Chl or proteins aggregates (Schmidt et al., 2003; Bednarczyk et al., 2015). The conformation changes could also be due to a transition between dimer to tetramer oligomerizations, but such dimers have never been observed experimentally (Palm et al., 2017).

Another explanation of the biphasic binding event could be due to difference of Chl stoichiometry inside the major tetrameric complexes. In WSCP tetramer, each subunit binds a single Chl and the complex is organized into two open sandwich structures, each constituted by a WSCP-Chl dimer (Horigome et al., 2007; Bednarczyk et al., 2016; Agostini et al., 2019). Some studies suggested different Chl stoichiometries inside WSCP-Chl tetrameric complexes (Schmidt et al., 2003; Hughes et al., 2006). Recently, Palm et al. (2017) and Agostini et al. (2018) showed that at low Chl concentrations, a stable tetramer was formed by binding to only two Chl molecules. Such complexes are as

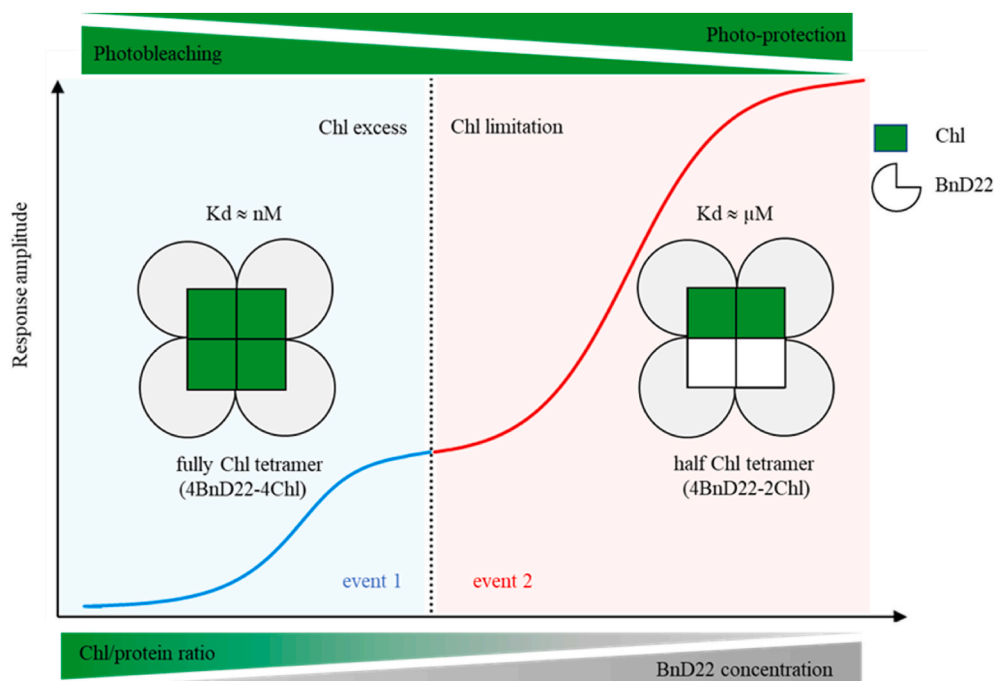
stable as those full of Chl. Moreover, the complexes containing two Chls can scavenge additional Chls to fill the vacant binding sites. We demonstrate herethat the biphasic binding event revealed the presence of two populations of BnD22-Chl complexes, a half Chl state (two Chls) and a fully Chl state (4 Chls). So, the event 1 corresponds to fully pigmented states (4BnD22-4Chl) where the BnD22 is reconstituted with excess of Chl, and the event 2 corresponds to the half-pigmented state (4BnD22-2Chl) where BnD22 was reconstituted with limited Chls (Fig. 7).

### 4.4. BnD22 shows a strong selectivity to Chla compared to Chlb

Although a preference of Chlb has been shown to LvWSCP (a WSCP IIB), there was no report describing the Chla preference for BoWSCP a WSCP IIA protein. These two WSCPs have respectively 36% and 93.5% similarity to BnD22 (Fig. S4). Girr et al. (2020) reported that the Chl selectivity was mainly controlled by the affinities. However, they failed to show any Chl preference to BoWSCP by the Kd values obtained for Chla and Chlb. In this study, the Kd values obtained by MST derived method for BnD22 to Chla were smaller than those of Chlb indicating that BnD22 displays a higher affinity for Chla and supporting the idea that Chla/Chlb preference is controlled by Kd values (Girr et al., 2020). The preference of BnD22 for Chla could partly be explained by the presence of PVCNEL motif in its polypeptide sequence (Palm et al., 2018) (Fig. S3). Moreover, it is in agreement with the Chla/Chlb content in native and recombinant WSCP IIA (Murata et al., 1971).

### 4.5. BnD22 has an effect on Chl photostability

The decrease of Chls photobleaching in presence of BnD22 corroborates the ability of BnD22 to protect the bound Chls against photo-damage as previously shown for several WSCPs (Schmidt et al., 2003; Agostini et al., 2017; Palm et al., 2019). This protective role would be due to the packaging of Chls within a well-protected hydrophobic core (Horigome et al., 2007; Agostini et al., 2017; Lemke and Götze, 2019; Palm et al., 2019). In our experimental conditions, highest photo-protection of Chla compared to Chlb in BnD22 complexes might due to a difference in the binding to BnD22 between Chls. Interestingly, in



**Fig. 7.** Model of biphasic BnD22-Chls interactions. The biphasic binding event resulted from two populations of BnD22-Chl tetrameric complexes, fully (4 Chl by tetramer, 4BnD22-4Chl) and a half (2 Chl by tetramer, 4BnD22-2Chl) Chl states, which depended on Chl/WSCP ratio in the binding condition. The Chls were positioned in the 2ChlaWSCP complex according to Palm et al. (2017), with one dimer binding a tightly coupled Chl pair and the other dimer with two empty binding sites. Binding affinities between Chl and BnD22 were more important during the formation of 4BnD22-2Chl complexes compare to fully Chl state (4BnD22-4Chl). Greater photo-protection of Chl inside the 4BnD22-2Chl complexes was observed. Dotted line indicates the equimolar Chl/protein ratio (=1). The blue and red highlighted regions indicate the Chl excess (Chl/protein ratio >1) and Chl limitation (Chl/protein ratio <1) respectively.

agreement with Palm et al. (2017), in our conditions, the Chla and Chlb are also protected against photooxidation in both types of BnD22-Chl complexes at half and fully pigmented states. But, in contrary to Palm et al. (2017) that observed a similar photobleaching rates whatever the stoichiometry of Chla inside the WSCP complexes, our results showed a photoprotection of Chla or Chlb better in 4BnD22-2Chls than in 4BnD22-4Chl complexes (Table 2). The difference in results obtained by Palm et al. (2017) with ours, could be explained by the time of illumination (1 h versus a few seconds respectively).

## 5. Conclusion

For the first time, we demonstrate that Chl photobleaching tracking by use of a Monolight equipment is an appropriate method to measure the binding affinities of WSCP to Chla or Chlb where circular dichroism method failed. By this novel method, we have demonstrated that BnD22, a WSCP IIA, is able to bind both Chla and/or Chlb with a Chla preference. Moreover, it appeared that depending on Chl/WSCP ratio, two distinct binding events were identified, each with specific binding affinities (range around nM and  $\mu$ M). This finding is consistent with previous results showing that WSCPs can bind 2 or 4 Chls molecules which could be protected differently from photodamage within WSCP-Chl complexes (Fig. 7). Our results demonstrate the technical feasibility of the method proposed in this study. It has the potential to become a valuable and fast approach for analyzing wild-type as well as mutated WSCPs difference in their affinity and their selectivity to Chls or its precursors or its derivatives. This derived-MST method as a screening tool is very promising and opens new possibilities to characterize assembly process between WSCPs and pigments. Future studies on other WSCP members than BnD22 could be performed to improve our knowledge on WSCP properties and their role in plants. This offers new opportunities to understand the presence of significant number of WSCPs genes in *Brassicaceae*.

## Data and materials availability statement

All materials and data analyzed in this study are available from the corresponding author.

## Funding

This research was supported by national collaborative project entitled RAPSODYN (ANR-11-BTBR-0004) funded by the program “Investments for the Future”. This work was also supported by PhD grant to Youssef Bouargalne from the French Research Ministry (MESR, Ministère de l’Enseignement Supérieur et de la Recherche) and by the Brittany regional authority.

## Author contributions

F. L.C. and C.D. acquired financial support, conceived, designed and supervised the research. Y.B. conceived, designed and performed experiments, analyzed data, prepared figures and wrote the original draft. V.C. and A.C. prepared expression plasmids and contributed to protein purification. F.G. performed the protein purification and EMSA experiment. C.R.N. conceived, designed, validated and performed technical guidance for the MST experiments. All authors have read and agreed to the published version of the manuscript.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank Jean-François Hubert for helpful discussions about recombinant protein production and purification. We thank Katarzyna Walkiewicz and Pierre Soule (NanoTemper) for their suggestions regarding the MST experiment design and for being our NanoTemper contact. We are grateful to the Spectroscopies-DCTP core facility (UMS Biosit, Université de Rennes 1- Campus de Villejean- 35043 RENNES Cedex, FRANCE) for access to their facilities.

## Appendix A. Supplementary data

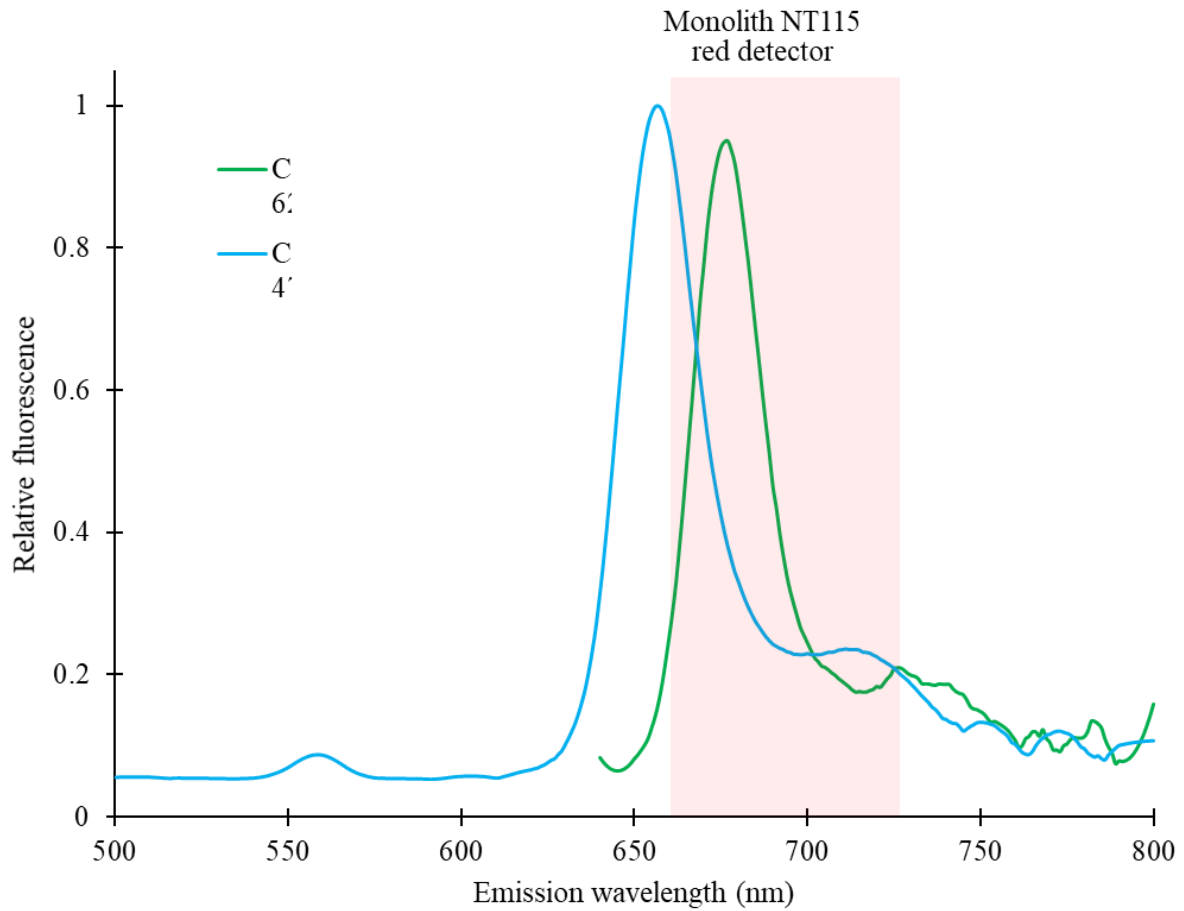
Supplementary data to this article can be found online at <https://doi.org/10.1016/j.plaphy.2022.03.023>.

## References

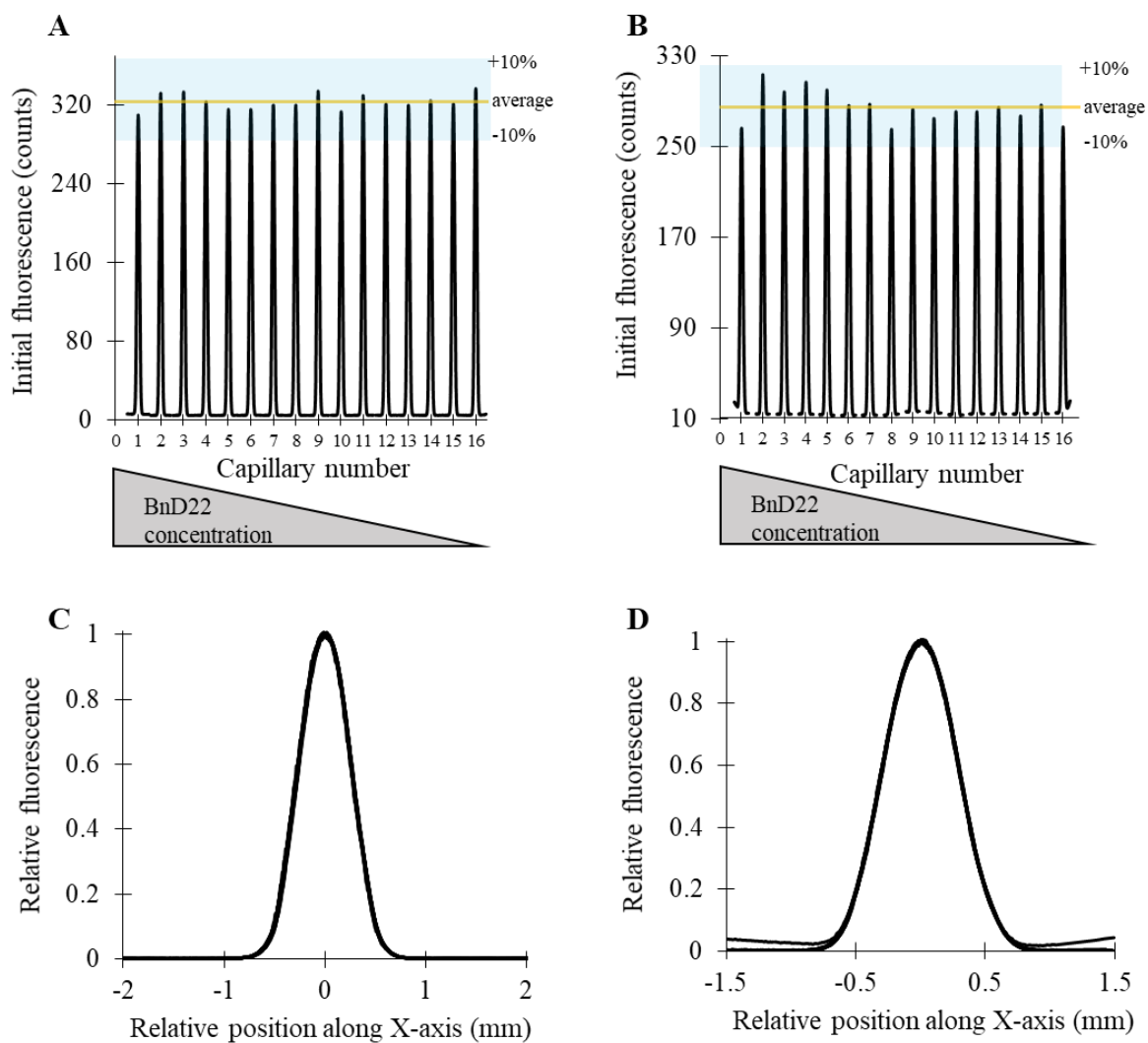
- Agostini, A., Palm, D.M., Schmitt, F.J., Albertini, M., Valentin, M Di, Paulsen, H., Carbonera, D., 2017. An unusual role for the phytol chains in the photoprotection of the chlorophylls bound to water soluble chlorophyll binding proteins. *Sci. Rep.* 7, 1–13. <https://doi.org/10.1038/s41598-017-07874-6>.
- Agostini, A., Palm, D.M., Paulsen, H., Carbonera, D., 2018. Optically detected magnetic resonance of chlorophyll triplet states in water soluble chlorophyll proteins from *Lepidium virginicum*: evidence for excitonic interaction among the four pigments. *J. Phys. Chem. B* 122, 6156–6163. <https://doi.org/10.1021/acs.jpbc.8b01906>.
- Agostini, A., Meneghin, E., Gewehr, L., Pedron, D., Palm, D.M., Carbonera, D., Paulsen, H., Jaenicke, E., Collini, E., 2019. How water-mediated hydrogen bonds affect chlorophyll a/b selectivity in water soluble chlorophyll protein. *Sci. Rep.* 9, 18255. <https://doi.org/10.1038/s41598-019-54520-4>.
- Annamalai, P., Yanagihara, S., 1999. Identification and characterization of a heat-stress induced gene in cabbage encodes a Kunitz type protease inhibitor. *J. Plant Physiol.* 155, 226–233. [https://doi.org/10.1016/S0176-1617\(99\)80011-X](https://doi.org/10.1016/S0176-1617(99)80011-X).
- Bednarczyk, D., Takahashi, S., Satoh, H., Noy, D., 2015. Assembly of water-soluble chlorophyll-binding proteins with native hydrophobic chlorophylls in water-in-oil emulsions. *Biochim. Biophys. Acta* 1847, 307–313. <https://doi.org/10.1016/j.bbabi.2014.12.003>.
- Bednarczyk, D., Dym, O., Prabakar, V., Peleg, Y., Pike, D.H., Noy, D., 2016. Fine tuning of chlorophyll spectra by protein-induced ring deformation. *Angew. Chem. Int. Ed.* 55, 6901–6905. <https://doi.org/10.1002/anie.201512001>.
- Boex-Fontvieille, E., Rustgi, S., Reinbothe, S., Reinbothe, C., 2015. A Kunitz-type protease inhibitor regulates programmed cell death during fower development in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 6119–6135. <https://doi.org/10.1093/jxb/erv327>.
- Boex-Fontvieille, E., Rustgi, S., von Wettstein, D., Reinbothe, S., Reinbothe, C., 2015a. Water-soluble chlorophyll protein is involved in herbivore resistance activation during greening of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* 112, 7303–7308. <https://doi.org/10.1073/pnas.1507714112>.
- Damaraju, S., Schleder, S., Eckhardt, U., Lokstein, H., Grimm, B., 2011. Functions of the water soluble chlorophyll-binding protein in plants. *J. Plant Physiol.* 168, 1444–1451. <https://doi.org/10.1016/j.jplph.2011.02.007>.
- Desclos, M., Dubouset, L., Etienne, P., Le Cahérec, F., Satoh, H., Bonnefoy, J., Ourry, A., Avicé, J.-C., 2008. A proteomic profiling approach to reveal a novel role of *Brassica napus* Drought 22 kD/water soluble chlorophyll binding protein in young leaves during nitrogen remobilization induced by stressful conditions. *Plant Physiol.* 147, 1830–1844. <https://doi.org/10.1104/pp.108.116905>.
- Downing, W.L., Mauxion, F., Fauvarque, M.-O., Reviron, M.-P., De Vienne, D., Vartanian, N., Giraudat, J., 1992. A *Brassica napus* transcript encoding a protein related to the Kunitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. *Plant J.* 2, 685–693. <https://doi.org/10.1046/j.1365-313X.1992.t01-11-00999.x>.
- Entzian, C., Schubert, T., 2016. Studying small molecule-aptamer interactions using MicroScale Thermophoresis (MST). *Methods* 97, 27–34. <https://doi.org/10.1016/j.ymeth.2015.08.023>.
- Halls, C.E., Rogers, S.W., Oufattole, M., Østergard, O., Svensson, B., Rogers, J.C., 2006. A Kunitz-type cysteine protease inhibitor from cauliflower and *Arabidopsis*. *Plant Sci.* 170, 1102–1110. <https://doi.org/10.1016/j.plantsci.2006.01.018>.
- Horigome, D., Satoh, H., Itoh, N., Mitsunaga, K., Oonishi, I., Nakagawa, A., Uchida, A., 2007. Structural mechanism and photoprotective function of water-soluble chlorophyll-binding protein. *J. Biol. Chem.* 282, 6525–6531. <https://doi.org/10.1074/jbc.M609458200>.
- Hughes, J.L., Razeghifard, R., Logue, M., Oakley, A., Wydrzynski, T., Krausz, E., 2006. Magneto-optic spectroscopy of a protein tetramer binding two exciton-coupled chlorophylls. *J. Am. Chem. Soc.* 128, 3649–3658. <https://doi.org/10.1021/ja056576b>.
- Ilami, G., Nespoulous, C., Huet, J.C., Vartanian, N., Pernollet, J.C., 1997. Characterization of BnD22, a drought-induced protein expressed in *Brassica napus* leaves. *Phytochem* 45, 1–8. [https://doi.org/10.1016/S0031-9422\(96\)00788-1](https://doi.org/10.1016/S0031-9422(96)00788-1).
- Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H.M., Duhr, S., Baaske, P., Breitsprecher, D., 2014. MicroScale thermophoresis: interaction analysis and beyond. *J. Mol. Struct.* 1077, 101–113. <https://doi.org/10.1016/j.molstruc.2014.03.009>.

- Khavrutskii, L., Yeh, J., Timofeeva, O., Tarasov, S., Pritt, S., Stefanisko, K., Tarasova, N., 2013. Protein purification-free method of binding affinity determination by microscale thermophoresis. *J. Vis. Exp.* 15, 50541. <https://doi.org/10.3791/50541>.
- Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685. <https://doi.org/10.1038/227680a0>.
- Lemke, O., Götze, J.P., 2019. On the stability of the water soluble chlorophyll binding protein (WSCP) studied by molecular dynamics simulations. *J. Phys. Chem. B* 123, 10594–10604. <https://doi.org/10.1021/acs.jpcc.9b07915>.
- Mork-Jansson, A.E., Eichacker, L.A., 2018. Characterization of chlorophyll binding to LIL3. *PLoS One* 13, 1–14. <https://doi.org/10.1371/journal.pone.0192228>.
- Mork-Jansson, A.E., Eichacker, L.A., 2019. A strategy to characterize chlorophyll protein interaction in LIL3. *Plant Methods* 15, 1–13. <https://doi.org/10.1186/s13007-018-0385-5>.
- Murata, T., Toda, F., Uchino, K., Yakushiji, E., 1971. Water soluble chlorophyll protein of *Brassica oleracea* var. botrys (cauliflower). *Biochim. Biophys. Acta* 245, 208–215. [https://doi.org/10.1016/0005-2728\(71\)90023-5](https://doi.org/10.1016/0005-2728(71)90023-5).
- Nakano, R.T., Yamada, K., Bednarek, P., Nishimura, M., Hara-Nishimura, I., 2014. ER bodies in plants of the Brassicales order: biogenesis and association with innate immunity. *Front. Plant Sci.* 5, 1–17. <https://doi.org/10.3389/fpls.2014.00073>.
- Palm, D.M., Agostini, A., Tenzer, S., Gloeckle, B.M., Werwie, M., Carbonera, D., Paulsen, H., 2017. Water soluble chlorophyll protein (WSCP) stably binds two or four chlorophylls. *Biochem* 56, 1726–1736. <https://doi.org/10.1021/acs.biochem.7b00075>.
- Palm, D.M., Agostini, A., Aversch, V., Girr, P., Werwie, M., Takahashi, S., Satoh, H., Jaenicke, E., Paulsen, H., 2018. Chlorophyll a/b binding-specificity in water soluble chlorophyll protein. *Nat. Plants* 4, 920–929. <https://doi.org/10.1038/s41477-018-0273-z>.
- Palm, D.M., Agostini, A., Pohland, A.C., Werwie, M., Jaenicke, E., Paulsen, H., 2019. Stability of water soluble chlorophyll protein (WSCP) depends on phytyl conformation. *ACS Omega* 4, 7971–7979. <https://doi.org/10.1021/acsomega.9b00054>.
- Prabhar, V., Afriat-Jurnou, L., Paluy, I., Peleg, Y., Noy, D., 2020. New homologues of Brassicaceae water-soluble chlorophyll proteins shed light on chlorophyll binding, spectral tuning, and molecular evolution. *FEBS J.* 287, 991–1004. <https://doi.org/10.1111/febs.15068>.
- Renger, G., Pieper, J., Theiss, C., Trostmann, I., Paulsen, H., Renger, T., Eichler, H.J., Schmitt, F.J., 2011. Water soluble chlorophyll binding protein of higher plants: a most suitable model system for basic analyses of pigment–pigment and pigment–protein interactions in chlorophyll protein complexes. *J. Plant Physiol.* 168, 1462–1472. <https://doi.org/10.1016/j.jplph.2010.12.005>.
- Reviron, M.-P., Vartanian, N., Sallantin, M., Huet, J.-C., Pernollet, J.-C., de Vienne, D., 1992. Characterization of a novel protein induced by progressive or rapid drought and salinity in *Brassica napus* leaves. *Plant Physiol.* 100, 1486–1493. <https://doi.org/10.1104/pp.100.3.1486>.
- Rustgi, S., Boex-Fontvieille, E., Reinbothe, C., von Wettstein, D., Reinbothe, S., 2017. Serpin1 and WSCP differentially regulate the activity of the cysteine protease RD21 during plant development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2212–2217. <https://doi.org/10.1073/pnas.1621496114>.
- Satoh, H., Nakayama, K., Okada, M., 1998. Molecular cloning and functional expression of a water soluble chlorophyll protein, a putative carrier of chlorophyll molecules in cauliflower. *J. Biol. Chem.* 273, 30568–30575. <https://doi.org/10.1074/jbc.273.46.30568>.
- Satoh, H., Uchida, A., Nakayama, K., Okada, M., 2001. Water soluble chlorophyll protein in Brassicaceae plants is a stress-induced chlorophyll-binding protein. *Plant Cell Physiol.* 42, 906–911. <https://doi.org/10.1093/pcp/pce117>.
- Schmidt, K., Fufezan, C., Krieger-Liszkay, A., Satoh, H., Paulsen, H., 2003. Recombinant water soluble chlorophyll protein from *Brassica oleracea* var. botrys binds various chlorophyll derivatives. *Biochem* 42, 7427–7433. <https://doi.org/10.1021/bi034207r>.
- Takahashi, S., Yanai, H., Nakamaru, Y., Uchida, A., Nakayama, K., Satoh, H., 2012. Molecular cloning, characterization and analysis of the intracellular localization of a water soluble chlorophyll binding protein from brussels sprouts (*Brassica oleracea* var. gemmifera). *Plant Cell Physiol.* 53, 879–891. <https://doi.org/10.1093/pcp/pc031>.
- Seidel, S., Dijkman, P., Lea, W., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F., Ladbury, J., Schreiber, G., Watts, A., Braun, D., Duhr, S., 2013. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods* 59, 301–315. <https://doi.org/10.1016/j.ymeth.2012.12.005>.
- Takahashi, S., Ono, M., Uchida, A., Nakayama, K., Satoh, H., 2013a. Molecular cloning and functional expression of a water soluble chlorophyll binding protein from Japanese wild radish. *J. Plant Physiol.* 170, 406–412. <https://doi.org/10.1016/j.jplph.2012.10.007>.
- Takahashi, S., Yanai, H., Oka-Takayama, Y., Zanma-Sohtome, A., Fujiyama, K., Uchida, A., Nakayama, K., Satoh, H., 2013b. Molecular cloning, characterization and analysis of the intracellular localization of a water soluble chlorophyll-binding protein (WSCP) from Virginia pepperweed (*Lepidium virginicum*), a unique WSCP that preferentially binds chlorophyll b *in vitro*. *Planta* 238, 1065–1080. [doi.org/10.1007/s00425-013-1952-7](https://doi.org/10.1007/s00425-013-1952-7).
- Takahashi, S., Uchida, A., Nakayama, K., Satoh, H., 2014. The C-terminal extension peptide of non-photoconvertible water-soluble chlorophyll binding proteins (Class II WSCPs) affects their solubility and stability: comparative analyses of the biochemical and chlorophyll-binding properties of recombinant *Brassica*, *Raphanus* and *Lepidium* WSCPs with or without their C-terminal extension peptides. *Protein J.* 33, 75–84. <https://doi.org/10.1007/s10930-013-9539-5>.
- Theiss, C., Trostmann, I., Andree, S., Schmitt, F.J., Renger, T., Eichler, H.J., Paulsen, H., Renger, G., 2007. Pigment–pigment and pigment–protein interactions in recombinant water soluble chlorophyll proteins (WSCP) from cauliflower. *J. Phys. Chem. B* 111, 13325–13335. <https://doi.org/10.1021/jp0723968>.
- Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D., Duhr, S., 2010. Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* 1. <https://doi.org/10.1038/ncomms1093>.

## Supplementary data

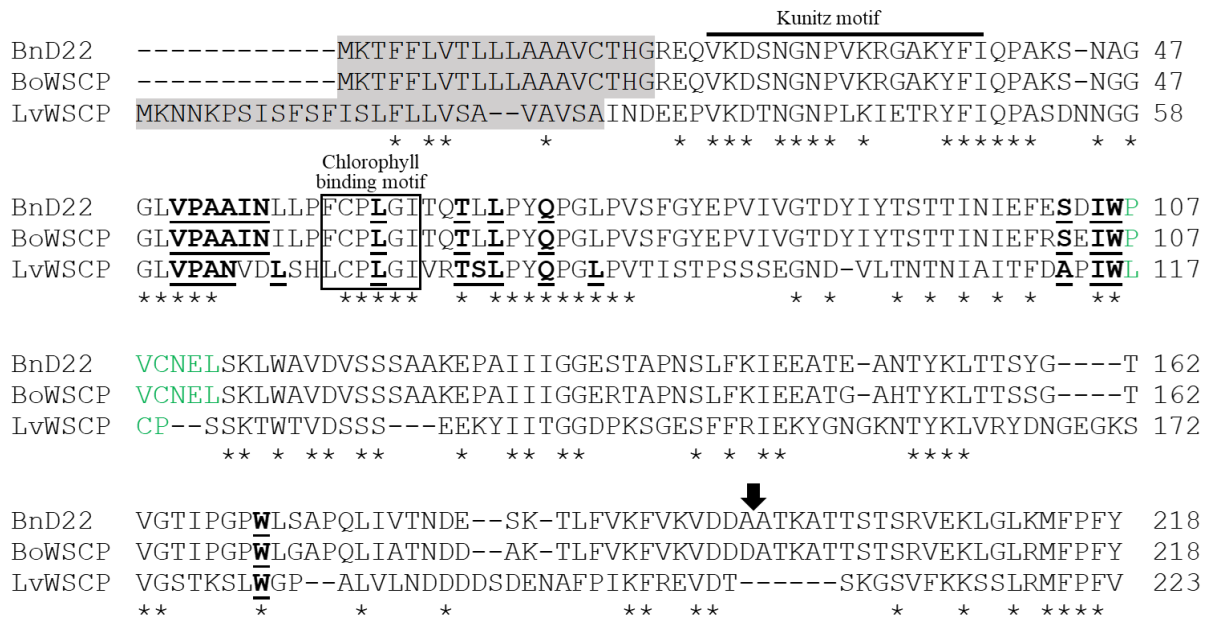


**Fig. S1.** Chla or Chlb fluorescence emission spectra after excitation at 470 nm and 625 nm. Chla or Chlb were solubilized in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04 % NP40) and fluorescence spectra were measured on a FluoroMax-2 instrument using two excitation wavelengths 470 nm (for Chlb) and 625 nm (for Chla). The emission wavelengths of the Monolith NT.115 red fluorescence detector are highlighted in red.



**Fig. S2.** Data inspection of MST binding experiments between BnD22 and Chla (left) or Chlb (right).

50 nM of Chla or 80 nM of Chlb were mixed with a titration series of BnD22 in binding buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% NP40). After 30 min of binding, fluorescence was measured at 100 % LED. Distribution of initial fluorescence was represented either individually for each capillary for Chla (A) or Chlb (B), or normalized and overlaid for Chla (C) or Chlb (D). The variation within  $\pm 10\%$  of the average is highlighted in blue.



**Fig. S3.** Sequence alignment of WSCPs from *Brassica napus* (BnD22, uniProt: Q43395), *Brassica oleracea* (BoWSCP, uniProt: Q7GDB3) and *Lepidium virginicum* (LvWSCP, uniProt: O04797).

The signal peptides are shown with a grey background. Black bar indicates the Kunitz motif. The box indicates the region homologous to the chlorophyll binding motif. Chla/b affinity motifs are in green. Amino acids close to the Chls chlorine macrocycle are underlined and bold. Black arrow indicates post-translational cleavage site. The asterisks indicate fully conserved residues.



## Article 2 :

### ***Brassica napus* Drought 22 kD (BnD22), a water-soluble chlorophyll protein acting simultaneously as a cysteine protease inhibitor and chlorophyll-binding protein.**

Bouargalne Youssef<sup>1</sup>, Guilbaud Florian<sup>1</sup>, Macherel David<sup>2</sup>, Delalande Olivier<sup>3</sup>, Deleu Carole<sup>1</sup> and Le Cahérec Françoise<sup>1\*</sup>

#### **Author affiliations**

1. Univ Rennes 1, INRAE, Institut Agro, IGEPP - UMR 1349, 35653, Le Rheu, France.
2. Univ Angers, Institut Agro, INRAE, IRHS, SFR 1207 QuaSaV, F-49000 Angers, France.
3. Univ Rennes 1, CNRS, IGDR – UMR 6290, F-35000 Rennes, France.

#### **Corresponding author**

\*Le Cahérec Françoise

UMR 1349 IGEPP (Institut de génétique et protection des plantes), INRAE, Institut Agro, Université de Rennes 1, BP 35327, 35650 Le Rheu, France

francoise.le-caherec@univ-rennes1.fr



**ABSTRACT**

Type II water-soluble chlorophyll-binding proteins from *Brassicaceae* (WSCP) are non-photosynthetic proteins that bind chlorophylls (Chls) and Chl derivatives. Although the physiological function of WSCPs is still unclear, it is assumed to be involved in some stress responses, which is likely related to their Chl-binding ability and protease inhibitors activity. Yet, the dual functions and their simultaneity are still poorly understood. Here, the biochemical functions of BnD22, a major WSCP expressed in *B. napus* leaves, were investigated using recombinant His-tag protein. We showed that BnD22 exhibits no inhibitory activity toward serine proteases but was a potent inhibitor of cysteine proteases such as papain. BnD22 can bind both Chla and Chlb to form a tetrameric complex. Unexpectedly, the tetrameric complex displays higher inhibition of cysteine proteases, indicating that BnD22 can bind Chl and inhibit protease simultaneously. Thus, physiological Chl-dependent activation of WSCP was suggested. Using 3D structural modeling and molecular docking, we showed that Chl-binding allows a better interaction between BnD22 and proteases. Despite its Chl-binding, BnD22 expressed as fluorescent-fused chimeric protein, was not found in chloroplasts, but was located in the endoplasmic reticulum and vacuole. In addition, the role of the C-terminal extension peptide, which cleaved off post-translationally *in vivo*, was investigated. It seems to be implicated neither in subcellular localization nor the functional properties of BnD22, but it drastically promotes the expression, solubility, and stability of the recombinant protein.

**Keywords:** Kunitz-type protease inhibitors, papain, inhibition constant, subcellular location.

**Abbreviations:**

At: *Arabidopsis thaliana*; BAEE: N $\alpha$ -benzoyl-L-arginine ethyl ester; BApNA: N-benzoyl-L-arginine-p-nitroanilide; BnD22: *Brassica napus* drought-induced 22 kDa protein; Bo: *Brassica oleracea*; BTEE: N-benzoyl-L-tyrosine ethyl ester; Chl: chlorophyll; EMSA: Electrophoretic mobility shift assay; ER: endoplasmic reticulum; GFP: green fluorescent protein; IC50: half-maximal inhibitory concentration; Kav: partition coefficient; Ki: inhibition constant; KTI: Kunitz type inhibitor; Lv: *Lepidium virginicum*; MES: 2-(N-morpholino) ethane sulfonic acid; MW: molecular weight; NP40: nonidet P40 substitute; PAGE: polyacrylamide gel electrophoresis; RFP: red fluorescent protein; SDS: Sodium dodecyl sulfate; Tm: melting temperature; WSCP: water-soluble chlorophyll proteins.

## INTRODUCTION

Hydrophilic chlorophyll (Chl)-binding non-photosynthetic proteins have been isolated from various *Brassicaceae* plants and are categorized into Class II water-soluble Chl-binding proteins (WSCPs) (Satoh *et al.*, 2001; Prabahar *et al.*, 2020). These Class II WSCPs bind chlorophylls (Chl) and their derivatives and can be further classified into two subclasses, class IIA and class IIB, based on their Chl a/b ratio. WSCP exhibits a typical  $\beta$ -trefoil fold with six antiparallel  $\beta$ -sheets and interconnecting unfolded loops (Horigome *et al.*, 2007). The binding of WSCP to Chls or its derivatives leads to the formation of a tetrameric complex (Nishio & Satoh, 1997; Horigome *et al.*, 2007; Palm *et al.*, 2017; Girr *et al.*, 2020; Bednarczyk *et al.*, 2021). The WSCP can extract Chl efficiently from isolated thylakoid membranes (Satoh *et al.*, 1998; Schmidt *et al.*, 2003; Girr & Paulsen, 2021).

The crystal structures of WSCPs from *L. virginicum* (LvWSCP, a WSCPIIB) and *Brassica oleracea* (BoWSCP, aWSCPIIA) with Chla or Chlb have been well documented (Horigome *et al.*, 2007; Palm *et al.*, 2018; Agostini *et al.*, 2019). The complexes WSCP-Chl are organized as an open sandwich dimer enclosing four tightly packed Chl molecules within a well-protected hydrophobic core (Horigome *et al.*, 2007; Bednarczyk *et al.*, 2016). This arrangement makes WSCPs remarkably stable towards denaturation and dissociation by harsh conditions and resistant to photo-oxidation upon illumination (Bektas *et al.*, 2012; Takahashi *et al.*, 2014; Palm *et al.*, 2017).

WSCPs can be overexpressed in *Escherichia coli* and used to study their structural and functional properties (Satoh *et al.*, 2001; Schmidt *et al.*, 2003; Prabahar *et al.*, 2020). Recombinant WSCP can be reconstituted *in vitro* with either 2 or 4 Chls to a tetrameric complex biochemically and spectroscopically equivalent to native ones (Satoh *et al.*, 2001; Palm *et al.*, 2017). WSCPs exhibit different relative affinities for Chla or Chlb. The class IIA showed (Bouargalne *et al.*, 2022, for BnD22) or not (Palm *et al.*, 2018, for BoWSCP) a clear preference for either Chla, and class IIB has a preference towards Chlb (Palm *et al.*, 2018; Girr *et al.*, 2020 for LvWSCP). A motif responsible for this Chl selectivity was identified (Palm *et al.*, 2018).

The physiological role of WSCP is not yet clarified. Although their Chl-binding ability, WSCPs are non-photosynthesis proteins. They are not located in the chloroplast but rather in a special endoplasmic reticulum(ER)-derived compartment, so-called ER-body (Takahashi *et al.*, 2015). Some WSCPs are also located in non-photosynthetic tissues such as roots and flowers (Bektas *et al.*, 2012; Takahashi *et al.*, 2013b; Boex-Fontvieille *et al.*, 2015a), and some of them cannot bind Chl (Prabahar *et al.*, 2020). In several plant species, WSCP expression is induced by various stresses conditions (Downing *et al.*, 1992; Reviron *et al.*, 1992; Annamalai & Yanagihara, 1999; Desclos *et al.*, 2008). It seems likely that WSCP is involved in protective functions such as a Chl s(Takahashi *et al.*, 2012, 2013b) and ROS source in signaling pathways (Agostini *et al.*, 2017). Besides their Chl-binding

property, a function as Kunitz-type protease inhibitors (PI) was also assigned to them (Ilami *et al.*, 1997; Satoh *et al.*, 1998, 2001). For the two most-studied LvWSCP and BoWSCP, the PI activity has never been investigated, but other WSCPs were able to inhibit different proteases. It has been demonstrated that AtWSCP from *Arabidopsis thaliana* has no significant serine protease inhibitor properties (Boex-Fontvieille *et al.*, 2016), but it inhibits cysteine proteases like papain, proaleurain maturation protease (Halls *et al.*, 2006) and RD21 (responsive to desiccation 21 kDa protein), a papain-like cysteine protease implicated in programmed cell death, leaf senescence and response to biotic and abiotic stress (Gepstein *et al.*, 2003; Halls *et al.*, 2006; Boex-Fontvieille *et al.*, 2015a; Rustgi *et al.*, 2017). For BnD22 (*Brassica napus* drought-induced 22 kDa), it remains inconclusive whether or not it acts as a protease inhibitor. No trypsin inhibitor activity of BnD22 was never detected *in vitro* (Ilami *et al.*, 1997; Nishio & Satoh, 1997), whereas Desclos *et al.* (2008) reported that BnD22 inhibited trypsin activity in young leaves treated by methyl jasmonate. Boex-Fontvieille *et al.*, (2015a,b) showed that AtWSCP is unable to bind Chl and RD21 at the same time. The AtWSCP-RD21 complex, in which RD21 is inhibited, dissociates upon Chl-binding. So far, no data are available that show bifunctional propriety of other WSCPs as a Chl binding and protease inhibitor.

In the present study, using recombinant BnD22, we investigate its inhibition activities towards serine and cysteine proteases and its ability to keep them or not in Chl complexes. For the first time, we showed that BnD22 has a bifunctional activity with the ability to bind Chl and inhibit the cysteine proteases. Three-dimensional structural modeling and molecular docking methods were used to understand how BnD22 interacts with papain protease and how its interaction is kept inside Chls complexes causing a better inhibition against the protease.

## MATERIALS AND METHODS

### *E. coli* recombinant protein expression and purification

Mature BnD22s (BnaC06g33320D) with (BnD22C) or without (BnD22) C-terminal extension were PCR amplified from genomic DNA of *Brassica napus* (Darmor nain accession). Amplicons were cloned in-frame into a pET-21a plasmid (Novagen) using the Gibson assembly approach (New England Biolabs). The primers used are listed in Table S1. Expression of recombinant proteins fused with a C-terminal hexahistidine tag (His tag proteins) was performed in *E. coli* Rosetta-gami 2(DE3) (Novagen), and purification was carried out with Ni<sup>2+</sup>-chelating affinity chromatography as described previously (Bouargalne *et al.*, 2022). The bacterial expression profile and the purity level of His-tag proteins were analyzed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie staining (Laemmli, 1970). Western blotting was carried out using rabbit anti-His tag polyclonal antibody (Invitrogen) and detected by chemiluminescence (Thermo Scientific) to check the presence of the His-tag proteins.

### Protein thermal shift assay

The thermal shift assay experiments were carried out in 384 well PCR plates using LightCycler® 480 Real-Time PCR System (Applied Biosystems). In 20µl of 50 mM Tris-HCl (pH 7.5) buffer, 150 mM NaCl, 0.25g/µl of purified protein were mixed with 1:1000 protein Thermal Shift™ Dye (thermofisher). The dye's fluorescence was monitored using the excitation/ emission wavelength at 470/510 nm with a temperature gradient of 0.5°C/ min from 25 to 95°C. The melting temperature (T<sub>m</sub>) was calculated as the inflection of the resulting transition curve using a simple Boltzmann equation.

$$f(T) = f(\min) + \frac{f(\max) - f(\min)}{1 + e^{\frac{(T_m - T)}{c}}}$$

where f(T) is the fluorescence at particular temperature T; f(min) and f(max) are the fluorescence before and after transition, respectively; T<sub>m</sub> is the melting temperature; C is the enthalpy of the reaction.

### Reconstitution of protein- Chls complexes

To reconstitute proteins- Chl complexes, 20 µM of purified protein was mixed with 60 µM of Chla or Chlb (Sigma) in a binding buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.04% Nonidet-P40 (NP40). After 30 min in the dark, the BnD22s-Chls complexes were separated on a 10% native PAGE and eluted from gel in the binding buffer free of detergent. Protein concentrations were measured using the bicinchoninic acid protein assay kit (ThermoFisher). The eluted protein-Chl complexes were used for size-exclusion chromatography and protease inhibition assays.

### **Size-exclusion chromatography of protein-Chls complexes**

BnD22, BnD22C, and the recovered BnD22-Chl complexes were subjected to size-exclusion chromatography using a Superdex 200 10/300GL column on an AKTA system (GE Healthcare). Chromatography was performed at 10°C in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl. The chromatograms were monitored at 280 nm to check the presence of protein and at 420 or 452 nm for Chla or Chlb detections. The column was calibrated using a set of molecular weight (MW) markers (GE Healthcare) to determine the standard curve of the partition coefficient ( $K_{av}$ ) versus the logarithm of MW.  $K_{av}$  was calculated as  $(V_e - V_o)/(V_t - V_o)$ , where  $V_e$ ,  $V_o$ , and  $V_t$  are respectively the protein's elution, void and column volumes. The MW of the protein-Chl complexes and the subunits number were calculated based on the elution volumes and the MW of protein's monomer determined by the ExPasy tool (<https://web.expasy.org/protparam/>).

### **Trypsin and chymotrypsin inhibitory activity assays**

The anti-proteolytic properties of BnD22, BnD22C and BnD22-Chl against serine proteases using trypsin and chymotrypsin were assayed as described earlier with few modifications (Fang *et al.*, 2010). Briefly, 200  $\mu$ l of binding buffer containing 50 mM Tris-HCl (pH 7.5), 50 mM CaCl<sub>2</sub>, 7.5 pmoles of trypsin, or 15.75 pmoles of chymotrypsin were used pre-incubated with different amounts of BnD22, BnD22C, BnD22-Chla or BnD22-Chlb complexes (ratio protein/protease from 0 to 8). After 25 min at 26°C, 0.75 mM of the substrate, N $\alpha$ -benzoyl- L-arginine ethyl ester (BAEE) for trypsin and N-benzoyl-L-tyrosine ethyl ester (BTEE) for chymotrypsin was added. The hydrolysis of BAEE and BTEE was followed at 26°C for 5 min at 253 and 256 nm, respectively. The enzymatic activity was calculated from the absorbance vs. time curve slope. The anti-proteolytic properties against trypsin or chymotrypsin were expressed as a percentage of the residual activity. Experiments were performed in triplicate. All chemicals were from Sigma-Aldrich.

### **Papain inhibitory activity assay and kinetic parameters determination**

The anti-proteolytic properties of BnD22, BnD22C and BnD22-Chl against papain were analyzed as previously described (Esti *et al.*, 2013). Briefly, a fixed amount of protease (106.8 pmoles) was incubated for 15 min at 30°C with increasing amounts of apoproteins or protein-Chl complexes from 0 to 427.2 pmoles (ratio protein/protease from 0 to 4) in 200  $\mu$ l of 20 mM of 2-(N-morpholino) ethane sulfonic acid (MES) (pH 6.5) buffer, 5 mM cysteine. Papain activity was monitored by following the hydrolysis of 2 mM N-benzoyl-L-arginine-p-nitroanilide (BApNA) at 405 nm for 15 min. The enzymatic activity was calculated from the absorbance vs. time curve slope, expressed by  $\mu$ mol of hydrolyzed BApNA/min/mg papain. Inhibition of papain activity was determined. All experiments were carried out in triplicate.

The half maximal inhibitory concentration values (IC50) were calculated by fitting the data to a sigmoidal regression curve based on the following equation (Haupt *et al.*, 2015):

$$f(C) = \min + \frac{\max - \min}{1 + \left[\frac{C}{IC50}\right]^x}$$

Where  $f(C)$ , the inhibition effect at a concentration  $C$ ;  $\min$  and  $\max$ , minimal and maximal inhibitory effect respectively;  $x$ , the curve slope.

Inhibition constants ( $K_i$ ) were determined from Dixon plot (Dixon, 1953) using BApNA at four concentrations (1, 2, 4, and 8 mM) and different concentrations of apoproteins or reconstituted protein-Chl as described above. For each concentration of substrate and inhibitor, the specific enzymatic activities ( $v$ ) were calculated. The reciprocal velocity ( $1/v$ ) versus inhibitor concentrations was plotted for each substrate concentration.  $K_i$  was calculated from the intersection of the regressions line.

### Transient expression in protoplasts and confocal microscopy

The DNA encoding the full-length BnD22 with or without the C-terminal extension peptide was amplified by PCR (Table S2). The PCR amplicons were cloned first in pDONR207 (Invitrogen) and then recombined into the expression vectors p2GWF7,0 and p2GWR7,0 (Plant System Biology, VIB-Ghent University) using Gateway technology (Invitrogen). In the same way, an endoplasmic reticulum (ER) marker (Candat *et al.*, 2014) was inserted into p2GWF7,0.

*Brassica napus* protoplasts were prepared from cotyledons of four days old or leaves of two-week-old seedling as previously described by Candat *et al.* (2013). Seedlings were grown in a greenhouse with a thermo/photo period of 25°C/16 h - 18°C/8 h and a light of 200  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ . The protoplasts were co-transformed with BnD22 and BnD22C -p2GWF7,0 or -p2GWR7,0 and BnD22 (BnD22C)-p2GWR7,0 with ER- p2GWF7,0. Expressions of C-terminal GFP or RFP-fused protein were driven by 35S promoter. After overnight incubation, the subcellular localization of fluorescent proteins was examined using a Nikon A1 laser scanning confocal microscope and NIS-element software (Nikon). GFP, RFP, and chlorophyll autofluorescence were detected at excitation: emission wavelengths of 488:525-550 nm, 561:570-620 nm, and 638:662-737 nm, respectively.

### Molecular modeling

The 3D-modelling structure of BnD22 with or without C-terminal was performed using I-TASSER (Yang *et al.*, 2015). The crystal structures of BoWSCP complexed with Chla (PDB code: 5HPZ) (Bednarczyk *et al.*, 2016) or with Chlb (PDB code: 6S2Z) (Agostini *et al.*, 2019) were used to build the BnD22-Chl complexes. The energy minimization of the complexes was performed using YASARA (<http://www.yasara.org/minimizationserver.htm>). The crystal structure of papain available

in the RCSB protein data bank (ID code: 1PPN) was used. The protein-protein docking was carried out using ClusPro2.0 (Kozakov *et al.*, 2017). All figures displaying structural data were generated with the PyMOL molecular graphics system, Version 2.5.0 Schrödinger, LLC. The 2D plot of interaction was carried out using the DIMPLOT program of Ligplot<sup>+</sup> v.2.2.4 (Laskowski and Swindells, 2011).

### **Statistical analyses**

Statistical analyses for inhibition activities against papain of BnD22, BnD22C, BnD22-Chls complexes were performed by Kruskal- Wallis test.

## RESULTS

### Production, solubility, and stability of recombinant BnD22 proteins

BnD22C and BnD22 fused to a C-terminal His tag was expressed in *E. coli*. The two recombinant proteins were readily induced by IPTG and detected in both the supernatant and pellet fractions (Fig. 1). However, BnD22C was accumulated mainly in the insoluble fraction (Fig. 1B, about 60%) compared to BnD22 (Fig. 1A, almost 60% are soluble). After Ni-affinity chromatography, the recombinant proteins were purified to practically 97 % of purity. The yield of purified BnD22 (40 mg. L<sup>-1</sup>) was six times compared to BnD22C (7 mg. L<sup>-1</sup>).

The purified BnD22C and BnD22 were extremely thermostable with melting temperature (T<sub>m</sub>) above 70°C (Fig. 2A). However, in contrast to BnD22, BnD22C showed a very high initial fluorescent signal at low temperature, which gradually goes down with increasing temperature. This result indicates the presence of misfolded/aggregated BnD22C proteins. The size chromatography elution showed at least 20% of purified BnD22C (Fig. 2B). These findings suggest that the presence of C-terminal extension decreases the stability and solubility of recombinant proteins.

### Chl-binding ability

BnD22C and BnD22 were reconstituted with Chla or Chlb. Size- exclusion chromatography indicated that all the complexes were tetramers with a molecular weight of around 80 and 100 kDa for BnD22-Chl and BnD22C-Chl, respectively (Table 1, Fig. S1).

To pinpoint the interaction of BnD22 with Chl, 3D Molecular modeling was carried out. The structure of BnD22 closely resembles that of soybean Kunitz-type trypsin inhibitor or other WSCP proteins with 12 antiparallel  $\beta$ -strands that form a barrel-like structure (Song & Suh, 1998) (Fig. 3A-B). The C-terminal extension in BnD22C is very flexible (data not shown). The predicted structure of the BnD22-Chl complex was similar to that of other WSCP-Chl organized into two open-sandwich (dimers A-B and C-D) with four Chl compacted in the heart of the tetramer (Fig. 3A-B). There was no difference in the overall structure of BnD22-Chla and BnD22-Chlb. The complexes were due to interactions between subunit and Chls, particularly between their phytol tails and interaction with amino acids of three of the four chains of the complex (Fig. 3C-D). However, some slight differences were seen depending on the Chl natures. The Chl molecule mostly with the chain on which it is associated with the involvement of 15 amino acids for Chla and Chlb and six additional residues for Chlb. The other chain of the dimer is also engaged (3 and 5 amino acids for Chla and Chlb, respectively). Moreover, one of the two chains of the second dimer of the complex is implicated with four residues independently of Chl nature.



### Anti-proteolytic activity

As a member of the Kunitz family, the potential capacity to inhibit serine and/or cysteine protease of monomeric and tetrameric BnD22-Chls complexes was investigated. Both monomeric and tetrameric have no inhibitory activity against serine proteases (Fig. 4), but they display a concentration-dependent inhibition against a cysteine protease (Fig. 5A). The inhibition effect of BnD22 with or without the C-terminal was almost identical. However, its IC<sub>50</sub> and K<sub>i</sub> were affected by C-terminal. The IC<sub>50</sub> of BnD22 (1.60±0.08 μM) is 1.6-fold higher than that of BnD22C (0.98±0.05 μM), whereas the K<sub>i</sub> of BnD22, 1.29±0.05 μM, is slightly lower than that of BnD22C, 1.61±0.1 μM (Fig. 5B-C). The Dixon plot, IC<sub>50</sub> and K<sub>i</sub>, K<sub>m</sub>, and V<sub>max</sub> (data not shown) indicate that the BnD22 exhibited a mixed inhibition mode. Molecular docking studies supported this result. BnD22 binds papain in its L-domain by the amino acid from the β2 strand, loops II, III, V, and X (Fig. 6A). The hydrophobic contacts are mainly responsible for the interface between BnD22 and papain, and only 3 three hydrogen bonds are detected. The reactive site loop (RLS), loop V, seems to be the only one closest to the active pocket of papain (Fig. S2) and especially with binding sites S2 and S3 (Fig. 6B).

Surprisingly, even though BnD22 was complexed with Chl, its papain inhibition property activity was kept with a much stronger potency than that of the non-complexed monomeric form (Fig. 5A). For example, at an equimolar concentration of BnD22 and papain, the BnD22-Chl complexes inhibit 80% of papain activity, whereas only 20% inhibition at the same concentration of monomer (Fig. 5A). The BnD22-Chl complexes displays IC<sub>50</sub> values (0.13 μM for BnD22-Chlb and 0.25 μM for BnD22-Chla) and K<sub>i</sub> values (0.13 μM for BnD22-Chla and 0.10 μM for BnD22-Chlb) 5 to 10-fold higher than those of monomers. Moreover, at low concentrations (below 1 μM), the inhibition potency of BnD22-Chlb complexes was stronger than that of BnD22-Chla complexes. The BnD22-Chl complexes display a mixed inhibition mode with K<sub>i</sub> values of BnD22-Chlb slightly higher, suggesting that this complex exhibits a high affinity toward papain compared to BnD22-Chla complexes one (Fig. 5D-E).

The molecular docking prediction showed that the papain interacts with three of the four subunits of the tetrameric BnD22-Chl leading to leading to four interaction positions (Fig. S3). These four different orientations of the structures of BnD22 tetramer-papain complexes were defined as “frontside” when papain interacts with the A, B, and D subunits; “backside” when papain interacts with the B, A, and C subunits; “right-side” when papain interact with the C, D and B subunits; “left-side” when papain interact with the C, D and A subunits. the interaction of papain with BnD22-Chla and BnD22-Chlb exhibits a similar overall structure (data not shown). It should be noted that papain interacts with the three subunits similarly on all sides. For example, in the front-side model, where

the papain is mainly associated with the chain D, the loops II, VI, VII, and IX are involved with the papain L domain. The connections on the L domain of papain are achieved with loop X,  $\beta$ 10, and  $\beta$ 11 strands of chain A and the RLS loop V of chain B of the complex (Fig. 7B). Contrary to BnD22-papain, it is highlighted that in the BnD22-Chl-papain complexes, the VI and VII loops of BnD22 were especially interacted with the papain in close region of its active site pocket, whereas the RLS loop V interacted with the papain in distal region from its active site pocket (Table 2). The loop VI penetrates and binds tightly to the active site pocket, more precisely with the binding sites S2, S3, S1' and with the His159 or Cys25, two of the catalytic triad amino acids (Fig. S4A, Table 2). In addition, a few subunit A residues can also interact with the S3 pocket. The BnD22-Chl-papain complex was stabilized by sixteen H-bond, one salt bridge, and many hydrophobic contacts (Fig. S4A). These interactions allow an ideal interaction position of papain on BnD22-Chl complex to inhibit its proteolytic activity.

### **Subcellular localization of BnD22**

The subcellular distribution of BnD22 and BnD22C fused with GFP, or RFP at its C-terminal were determined in *B. napus* protoplasts (Fig. 8). Both BnD22 show similar localization (Fig. 8C-D). They were not located in plastids but appeared in ER and the vacuole of cotyledon protoplasts. A similar localization pattern was also observed in leaf protoplasts. Their presences in the ER were validated by co-expression with the ER-GFP marker (Fig. 8A\_B).

## DISCUSSION

In recent years, WSCPs have been used as model to identify and characterize protein-Chl interaction. Although members of Kunitz protease inhibitor, this aptitude is much less studied. BnD22, a WSCP in *B. napus* was the subject of this study. Here, we report the description of its inhibition potential against Ser and Cys proteases along with its chlorophyll-binding capacity.

### **BnD22 is an inhibitor against Cys proteases but not Ser proteases**

The WSCP proteins belong to the MEROPS sub-class I3A.028 of the Kunitz type inhibitor family (KTI). These KTI are classified as active against Ser protease, but they can inhibit other proteases such as Cys proteases (Habib & Fazili, 2007; Migliolo *et al.*, 2010; Rawlings, 2010). Desclos *et al.* (2008) reported that BnD22 induced by nitrogen starvation treatment inhibited trypsin activity in young leaves of *B. napus*. In contrast, it was shown that the purified native BnD22 and the recombinant AtWSCP did not show any protease inhibitor activity against trypsin or chymotrypsin (Ilami *et al.*, 1997; Boex-Fontvieille *et al.*, 2016). Our results showed that BnD22 was a KTI that cannot atypically inhibit serine proteases (trypsin and chymotrypsin) but was able to inhibit cysteine proteases such as papain. This is similar to earlier results, where it was shown that AtWSCP displayed an unusual target protease specificity by inhibiting cysteine proteases such as papain, and RD21A, instead of a serine protease (Halls *et al.*, 2006; Boex-Fontvieille *et al.*, 2015a; Rustgi *et al.*, 2017).

### **BnD22 binds Chls and inhibits cysteine protease at the same time.**

To investigate BnD22 inhibitory activity against papain after binding to Chls, we first purified the BnD22-Chl complexes to remove monomers that could inhibit papain. BnD22 was able to bind both Chla and b, yielding a tetramer BnD22-Chl complex, which is similar to that of other WSCPs (Prabakar *et al.*, 2020). In contrast to earlier results, where it was shown that the interaction between RD21 and AtWSCP is relieved upon Chl binding (Boex-Fontvieille *et al.*, 2015b), the BnD22-Chl can inhibit papain, indicating that BnD22 has a simultaneous bifunctional activity, an ability to bind Chl and inhibit protease. The Chl-dependent protease activation was previously discussed (Boex-Fontvieille *et al.*, 2015b). Those authors supposed that the Chl-binding of AtWSCP allows the activation of RD21. Our results revealed that BnD22 complexed with Chl exhibits a higher inhibitory activity against papain with 10-fold higher  $K_i$  than monomer, indicating that BnD22 binds Chl to become physiologically active. The activation of WSCP is therefore Chl-dependent. Thus, Chl-binding seems to be a pivotal step for the physiological functions of WSCP.

## Molecular basis of BnD22 binding to Chl and/or papain

The kinetic analyses were accompanied by docking studies to provide insights into the interactions responsible for the binding and inhibition of papain. To this end, a homology 3D model of BnD22 was built using the crystal structure of the known BoWSCP structures as a template (Bednarczyk *et al.*, 2016; Agostini *et al.*, 2019). Although protein sequence identity between BnD22 and BoWSCP is more than 93%, both possess the 2 conserved cysteine residues that form the disulfide bonds (Horigome *et al.*, 2007), suggesting similar folding. Obviously, the 3D structure of BnD22 and most KTIs exhibit a common fold, dominated by a b-trefoil formed by 12 antiparallel b-strands with long interconnecting loops (Oliva *et al.*, 2010). Further, the BnD22-Chl tetrameric complex containing one Chl per apoprotein was modeled (Horigome *et al.*, 2007; Bednarczyk *et al.*, 2016; Palm *et al.*, 2018; Agostini *et al.*, 2019). Some minor difference in the protein environment of the bound Chl molecules was observed between the 3D model of Chla and b but could not explain the Chl a/b specificities revealed by our previous titration experiments (Bouargalne *et al.*, 2022). It is important to note that the one Chl can interact with its principal chain and two other inter and intra-dimer chains within the tetramer. The second loop encompassing the LHCII signature and the fifth loop, which was designed to be the RSL, were involved in the Chl-binding.

The build model of BnD22 and BnD22Chl was docked to papain to underscore the differential interactions between BnD22-Chl-*protease*. The interaction model of papain with BnD22 is fundamentally different from that previously reported for AtWSCP-RD21 interaction (Boex-Fontvieille *et al.*, 2015b). The BnD22 interacts with only the papain's L-domain and exhibited tree H-bonds and many hydrophobic interactions. The interaction involved amino acids from three loops of BnD22, including the RLS, which is close to the S3 active site pocket of papain. Therefore, it is expected that BnD22 would slightly influence the papain activity as revealed by our kinetic experiment. The docking show that papain interacts with 3 subunits of the tetramer BnD22-Chl. The interacting residues are not the same in the implicated subunit and differ from those of the BnD22 monomer. The loopVI was predicted to penetrate into the papain active site pocket and forms crucial hydrophobic interactions with residues in the S1, S2, and S3 pocket and H-bound with His159 of the catalytic triad, thereby blocking access of small substrates (e.g., BAPNA) to the reactive site stopping the proteolytic activity of papain. It is not worthy that the BnD22 monomer RLS loop, which was supposed to intrude into the active site of papain, interacts with the papain in the distal region because it is mainly implicated in Chl-binding. In contrast to earlier results, where Boex-Fontvieille *et al.*, 2015b speculate pigment-dependent dissociation of the AtWSCP– RD21 complex. Our finding correlates well to the inhibitory activity observed, and validates that BnD22 can simultaneously bind Chl and inhibits *protease*. The molecular docking indicates that the interactions BnD22Chl-papain could take place in frontside, backside, right-side and left-side of the tetramer, and therefore, the

BnD22-Chl can inhibit 4 papain at the same time. The occurrence of many hydrogen and hydrophobic associations could stabilize the complex of BnD22-Chl-4papain. It has been reported that there are four pore-like vents in the Chl-binding cavity wall of WSCP-tetramer (Horigome *et al.*, 2007). However, in our model of BnD22-Chl-4papain interactions, the Chl-binding cavity was closed entirely. Thus, it could decrease the photoprotection of Chl within the tetramer.

### **Subcellular localization of BnD22**

The documented chlorophyll-binding properties and their ability to extract Chls from thylakoid membranes imply that WSCPs should be localized to the plastid compartment. All WSCPs, including BnD22, do not possess predictable chloroplast transit sequences but have a signal peptide targeting the endoplasmic reticulum (ER) and secretory pathway. Yet, it is still unknown how WSCP could enter the plastid since Chls are synthesized in chloroplasts, and all known WSCPs are never found in chloroplasts. Moreover, all most studied WSCPs, (AtWSCP, BoWSCP, LvWSCP, and RshWSCP) have been shown located in the ER-bodies and vacuoles (Takahashi *et al.*, 2012, 2013b, 2015). However, there have been no experimental data on the subcellular localization of BnD22. Here, we present data that BnD22 does not target chloroplasts but is localized in the ER and vacuole, either in cotyledon or leaf protoplasts. Such localization would be consistent with the inhibitory activity of BnD22, which could participate in modulating the cysteine protease activity predominantly detected in vacuoles (Otegui *et al.*, 2005; Martínez *et al.*, 2008). Most WSCPs such as LvWSCP, AtWSCP, and BoWSCP were previously detected in ER-bodies (Takahashi *et al.*, 2012, 2013a, 2015). The ER-bodies were found in different plant families belonging to the Brassicales order, like *Raphanus sativus* and *A. thaliana* (Ogasawara *et al.*, 2009; Gotté *et al.*, 2015). Since the presence of ER-bodies in *B. napus* has never been investigated, we cannot exclude that BnD22 might be located in ER bodies. It was also reported that AtWSCP-GFP fusion protein was detectable in the cytosol and in cell walls/extracellular spaces (Boex-Fontvieille *et al.*, 2015a). BnD22 could also be excreted in the cell wall, but it can't be assessed using transient expression in protoplasts.

### **Effect of C-terminal on BnD22 pro**

It has been observed previously that BnD22, a WSCP IIA of *B. napus*, possesses an N-terminal signal peptide for intracellular targeting and a C-terminal extension peptide (C-ter) cleaved off post-translationally *in vivo* (Ilami *et al.*, 1997). The C-terminal region is highly conserved in WSCPs, but its biochemical functions remain unclear. As shown for BoWSCP, RshWSCP, and LvWSCP (Takahashi *et al.*, 2014), we found that truncated recombinant BnD22 was soluble and remarkably stable. The presence of the C-terminal induced only a negligible effect on the 3D structure of BnD22, thus leading to a significant alteration of solubility, stability, and Chl binding but the protease

inhibitor was conserved. Although BnD22C exhibits similar proteases inhibition activity, its Chl-binding ability, especially for Chla, was slightly altered (Data not shown). This contradicts earlier results, where it was speculated that C-terminal does not affect the Chl-binding (Takahashi *et al.*, 2014). our results suggest that the cleavage of the C-terminal is needed for protein folding to produce the necessary conformation for optimal Chl binding ability. It was hypothesized that C-terminal might play a role in targeting a specific cellular compartment (Ilami *et al.*, 1997). However, as reported for LvWSCP (Takahashi *et al.*, 2013b), we found that the BnD22 with or without C-terminal has a very similar subcellular localization, indicating that this region is not a trailer peptide.

## CONCLUSION

This is the first report describing the relationship between the Chl-binding activity and the inhibitory activity of BnD22. We demonstrated that BnD22 can bind Chl and inhibit protease at the same time. The Chl-binding did cause an obvious increase of inhibitory activity. Our modeling studies shed new light on BnD22-Chl-protease interactions. Our data suggest an unprecedented mechanism of Chl-dependent activation of WSCP/BnD22. Furthermore, this study reveals for the first time the subcellular localization of a WSCP from *B. napus*. The present study provides some crucial insights on defining the physiological functions of WSCP/BnD22.

## REFERENCES

- Agostini A, Meneghin E, Gewehr L, Pedron D, Palm DM, Carbonera D, Paulsen H, Jaenicke E, Collini E. 2019.** How water-mediated hydrogen bonds affect chlorophyll a/b selectivity in Water-Soluble Chlorophyll Protein. *Scientific Reports* **accepted f**: 1–10.
- Agostini A, Palm DM, Schmitt FJ, Albertini M, Valentin M Di, Paulsen H, Carbonera D. 2017.** An unusual role for the phytyl chains in the photoprotection of the chlorophylls bound to Water-Soluble Chlorophyll-binding Proteins. *Scientific Reports* **7**: 1–13.
- Annamalai P, Yanagihara S. 1999.** Identification and characterization of a heat-stress induced gene in cabbage encodes a Kunitz type protease inhibitor. *Journal of Plant Physiology* **155**: 226–233.
- Bednarczyk D, Dym O, Prabaha V, Peleg Y, Pike DH, Noy D. 2016.** Fine Tuning of Chlorophyll Spectra by Protein-Induced Ring Deformation. *Angewandte Chemie - International Edition* **55**: 6901–6905.
- Bednarczyk D, Tor-Cohen C, Das PK, Noy D. 2021.** Direct Assembly in Aqueous Solutions of Stable Chlorophyllide Complexes with Type II Water-soluble Chlorophyll Proteins. *Photochemistry and Photobiology* **97**: 732–738.
- Bektas I, Fellenberg C, Paulsen H. 2012.** Water-soluble chlorophyll protein (WSCP) of Arabidopsis is expressed in the gynoecium and developing silique. *Planta* **236**: 251–259.
- Boex-Fontvieille E, Reinbothe S, Reinbothe C, Rustgi S, Von Wettstein D, Pollmann S. 2016.** An ethylene-protected achilles' heel of etiolated seedlings for arthropod deterrence. *Frontiers in Plant Science* **7**: 1–14.
- Boex-Fontvieille E, Rustgi S, Reinbothe S, Reinbothe C. 2015a.** A Kunitz-type protease inhibitor regulates programmed cell death during fower development in *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**: 6119–6135.
- Boex-Fontvieille E, Rustgi S, von Wettstein D, Reinbothe S, Reinbothe C. 2015b.** Water-soluble chlorophyll protein is involved in herbivore resistance activation during greening of *Arabidopsis thaliana* . *Proceedings of the National Academy of Sciences* **112**: 7303–7308.
- Desclos M, Dubouset L, Etienne P, Le Caherec F, Satoh H, Bonnefoy J, Ourry A, Avice J-C. 2008.** A Proteomic Profiling Approach to Reveal a Novel Role of *Brassica napus* Drought 22 kD/Water-Soluble Chlorophyll-Binding Protein in Young Leaves during Nitrogen Remobilization Induced by Stressful Conditions . *Plant Physiology* **147**: 1830–1844.
- Dixon M. 1953.** The determination of enzyme inhibitor constants. *Biochemical Journal* **55**: 177.
- Downing WL, Mauxion F, Fauvarque M -O, Reviron M -P, De Vienne D, Vartanian N, Giraudat J. 1992.** A *Brassica napus* transcript encoding a protein related to the Kunitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. *The Plant Journal* **2**: 685–693.
- Esti M, Benucci I, Lombardelli C, Liburdi K, Garzillo AMV. 2013.** Papain from papaya (*Carica papaya L.*) fruit and latex: Preliminary characterization in alcoholic–acidic buffer for wine application. *Food and Bioproducts Processing* **91**: 595–598.
- Fang EF, Wong JH, Ng TB. 2010.** Thermostable Kunitz trypsin inhibitor with cytokine inducing, antitumor and HIV-1 reverse transcriptase inhibitory activities from Korean large black soybeans. *Journal of Bioscience and Bioengineering* **109**: 211–217.
- Gepstein S, Sabehi G, Carp MJ, Hajouj T, Neshar MFO, Yariv I, Dor C, Bassani M. 2003.** Large-scale identification of leaf senescence-associated genes. *Plant Journal* **36**: 629–642.
- Girr P, Kilper J, Pohland AC, Paulsen H. 2020.** The pigment binding behaviour of water-soluble chlorophyll protein (WSCP). *Photochemical and Photobiological Sciences* **19**: 695–712.
- Girr P, Paulsen H. 2021.** How water-soluble chlorophyll protein extracts chlorophyll from membranes. *Biochimica et Biophysica Acta - Biomembranes* **1863**: 183479.
- Gotté M, Ghosh R, Bernard S, Nguema-Ona E, Vitré-Gibouin M, Hara-Nishimura I, Driouich A. 2015.** Methyl jasmonate affects morphology, number and activity of endoplasmic reticulum bodies in raphanus sativus root cells. *Plant and Cell Physiology* **56**: 61–72.

- Habib H, Fazili KM. 2007.** Plant protease inhibitors : a defense strategy in plants. *Biotechnology and Molecular Biology Review* **2**: 68–85.
- Halls CE, Rogers SW, Oufattole M, Østergard O, Svensson B, Rogers JC. 2006.** A Kunitz-type cysteine protease inhibitor from cauliflower and *Arabidopsis*. *Plant Science* **170**: 1102–1110.
- Haupt LJ, Kazmi F, Ogilvie BW, Buckley DB, Smith BD, Leatherman S, Paris B, Parkinson O, Parkinson A. 2015.** The reliability of estimating  $K_i$  values for direct, reversible inhibition of cytochrome P450 enzymes from corresponding  $IC_{50}$  values: A retrospective analysis of 343 experiments. *Drug Metabolism and Disposition* **43**: 1744–1750.
- Horigome D, Satoh H, Itoh N, Mitsunaga K, Oonishi I, Nakagawa A, Uchida A. 2007.** Structural mechanism and photoprotective function of water-soluble chlorophyll-binding protein. *Journal of Biological Chemistry* **282**: 6525–6531.
- Ilami G, Nespoulous C, Huet JC, Vartanian N, Pernollet JC. 1997.** Characterization of BnD22, a drought-induced protein expressed in *Brassica napus* leaves. *Phytochemistry* **45**: 1–8.
- Laskowski RA, Swindells MB. 2011.** LigPlot+: Multiple Ligand–Protein Interaction Diagrams for Drug Discovery. *Journal of Chemical Information and Modeling* **51**: 2778–2786.
- Martínez DE, Costa ML, Gomez FM, Otegui MS, Guamet JJ. 2008.** ‘Senescence-associated vacuoles’ are involved in the degradation of chloroplast proteins in tobacco leaves. *Plant Journal* **56**: 196–206.
- Migliolo L, de Oliveira AS, Santos EA, Franco OL, de Sales MP. 2010.** Structural and mechanistic insights into a novel non-competitive Kunitz trypsin inhibitor from *Adenanthera pavonina L.* seeds with double activity toward serine- and cysteine-proteinases. *Journal of Molecular Graphics and Modelling* **29**: 148–156.
- Nishio N, Satoh H. 1997.** A Water-Soluble Chlorophyll Protein in Cauliflower May Be Identical to BnD22, a Drought-Induced, 22-Kilodalton Protein in Rapeseed. *Plant Physiology* **115**: 841–846.
- Ogasawara K, Yamada K, Christeller JT, Kondo M, Hatsugai N, Hara-Nishimura I, Nishimura M. 2009.** Constitutive and inducible ER bodies of *Arabidopsis thaliana* accumulate distinct  $\beta$ -glucosidases. *Plant and Cell Physiology* **50**: 480–488.
- Oliva ML V., Silva MCC, Sallai RC, Brito M V., Sampaio MU. 2010.** A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. *Biochimie* **92**: 1667–1673.
- Otegui MS, Noh YS, Martínez DE, Vila Petroff MG, Staehelin LA, Amasino RM, Guamet JJ. 2005.** Senescence-associated vacuoles with intense proteolytic activity develop in leaves of *Arabidopsis* and soybean. *Plant Journal* **41**: 831–844.
- Palm DM, Agostini A, Aversch V, Girr P, Werwie M, Takahashi S, Satoh H, Jaenicke E, Paulsen H. 2018.** Chlorophyll a/b binding-specificity in water-soluble chlorophyll protein. *Nature Plants* **4**: 920–929.
- Palm DM, Agostini A, Tenzer S, Gloeckle BM, Werwie M, Carbonera D, Paulsen H. 2017.** Water-Soluble Chlorophyll Protein (WSCP) Stably Binds Two or Four Chlorophylls. *Biochemistry* **56**: 1726–1736.
- Prabahar V, Afriat-Jurnou L, Paluy I, Peleg Y, Noy D. 2020.** New homologues of Brassicaceae water-soluble chlorophyll proteins shed light on chlorophyll binding, spectral tuning, and molecular evolution. *FEBS Journal* **287**: 991–1004.
- Rawlings ND. 2010.** Peptidase inhibitors in the MEROPS database. *Biochimie* **92**: 1463–1483.
- Reviron M-P, Vartanian N, Sallantin M, Huet J-C, Pernollet J-C, de Vienne D. 1992.** Characterization of a Novel Protein Induced by Progressive or Rapid Drought and Salinity in *Brassica napus* Leaves . *Plant Physiology* **100**: 1486–1493.
- Rustgi S, Boex-Fontvieille E, Reinbothe C, von Wettstein D, Reinbothe S. 2017.** Serpin1 and WSCP differentially regulate the activity of the cysteine protease RD21 during plant development in *Arabidopsis thaliana* . *Proceedings of the National Academy of Sciences* **114**: 2212–2217.
- Satoh H, Nakayama K, Okada M. 1998.** Molecular cloning and functional expression of a water-soluble chlorophyll protein, a putative carrier of chlorophyll molecules in cauliflower. *Journal of Biological Chemistry* **273**: 30568–30575.
- Satoh H, Uchida A, Nakayama K, Okada M. 2001.** Water-soluble chlorophyll protein in Brassicaceae plants



is a stress-induced chlorophyll-binding protein. *Plant and Cell Physiology* **42**: 906–911.

**Schmidt K, Fufezan C, Krieger-Liszkay A, Satoh H, Paulsen H. 2003.** Recombinant water-soluble chlorophyll protein from *Brassica oleracea* var. *Botrys* binds various chlorophyll derivatives. *Biochemistry* **42**: 7427–7433.

**Song HK, Suh SW. 1998.** Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *Journal of Molecular Biology* **275**: 347–363.

**Takahashi S, Aizawa K, Nakayama K, Satoh H. 2015.** Water-soluble chlorophyll-binding proteins from *Arabidopsis thaliana* and *Raphanus sativus* target the endoplasmic reticulum body. *BMC Research Notes* **8**: 1–5.

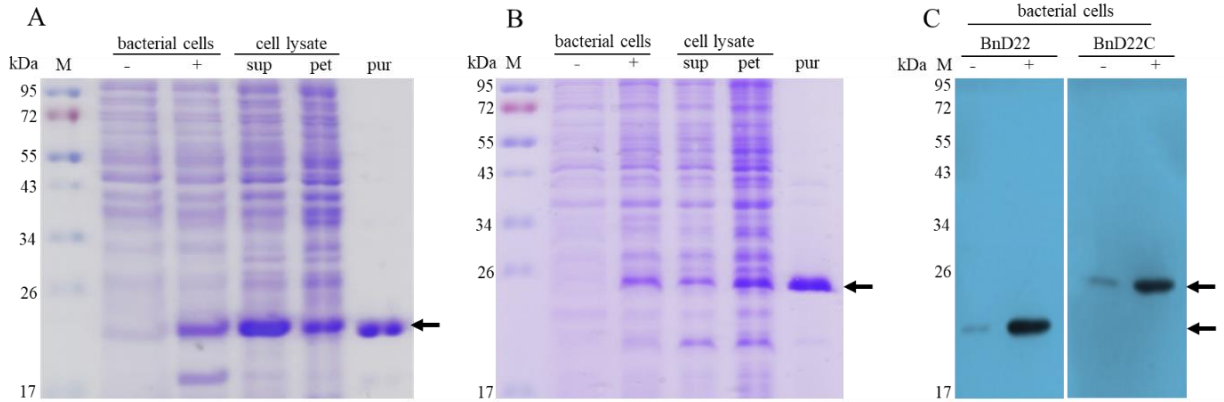
**Takahashi S, Ono M, Uchida A, Nakayama K, Satoh H. 2013a.** Molecular cloning and functional expression of a water-soluble chlorophyll-binding protein from Japanese wild radish. *Journal of Plant Physiology* **170**: 406–412.

**Takahashi S, Uchida A, Nakayama K, Satoh H. 2014.** The C-terminal extension peptide of non-photoconvertible water-soluble chlorophyll-binding proteins (Class II WSCPs) affects their solubility and stability: Comparative analyses of the biochemical and chlorophyll-binding properties of recombinant brassica. *Protein Journal* **33**: 75–84.

**Takahashi S, Yanai H, Nakamaru Y, Uchida A, Nakayama K, Satoh H. 2012.** Molecular cloning, characterization and analysis of the intracellular localization of a water-soluble chl-binding protein from brussels sprouts (*brassica oleracea* var. *gemmifera*). *Plant and Cell Physiology* **53**: 879–891.

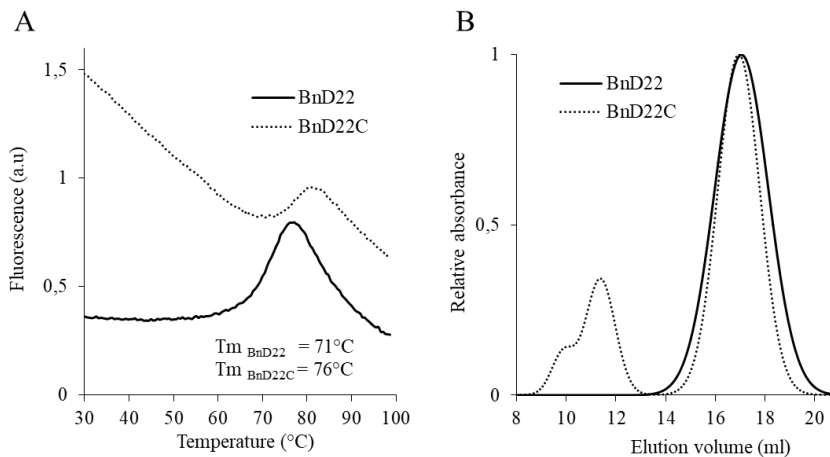
**Takahashi S, Yanai H, Oka-Takayama Y, Zanma-Sohtome A, Fujiyama K, Uchida A, Nakayama K, Satoh H. 2013b.** Molecular cloning, characterization and analysis of the intracellular localization of a water-soluble chlorophyll-binding protein (WSCP) from Virginia pepperweed (*Lepidium virginicum*), a unique WSCP that preferentially binds chlorophyll b in vitro. *Planta* **238**: 1065–1080.

## FIGURES



**Fig. 1: SDS-PAGE profile (A et B) and western-blotting (C) of recombinant protein.**

A-B) Expression, subcellular fractionation in *E. coli* and purification of BnD22 (A) and BnD22C (B). Proteins from cell extracts of uninduced *E. coli* (-) or after IPTG inducing (+), from supernatant (sup) or pellet (pet) fractions of sonicated *E. coli* and purified protein using Ni<sup>2+</sup> affinity chromatography (pur), were separated by 15% SDS-PAGE and stained with coomassie Blue. Detection of recombinant proteins from uninduced (-) or IPTG-induced (+) *E. coli* by western-blotting using His tag antibodies (C). Black arrows indicate the recombinant proteins. M indicates the protein markers.



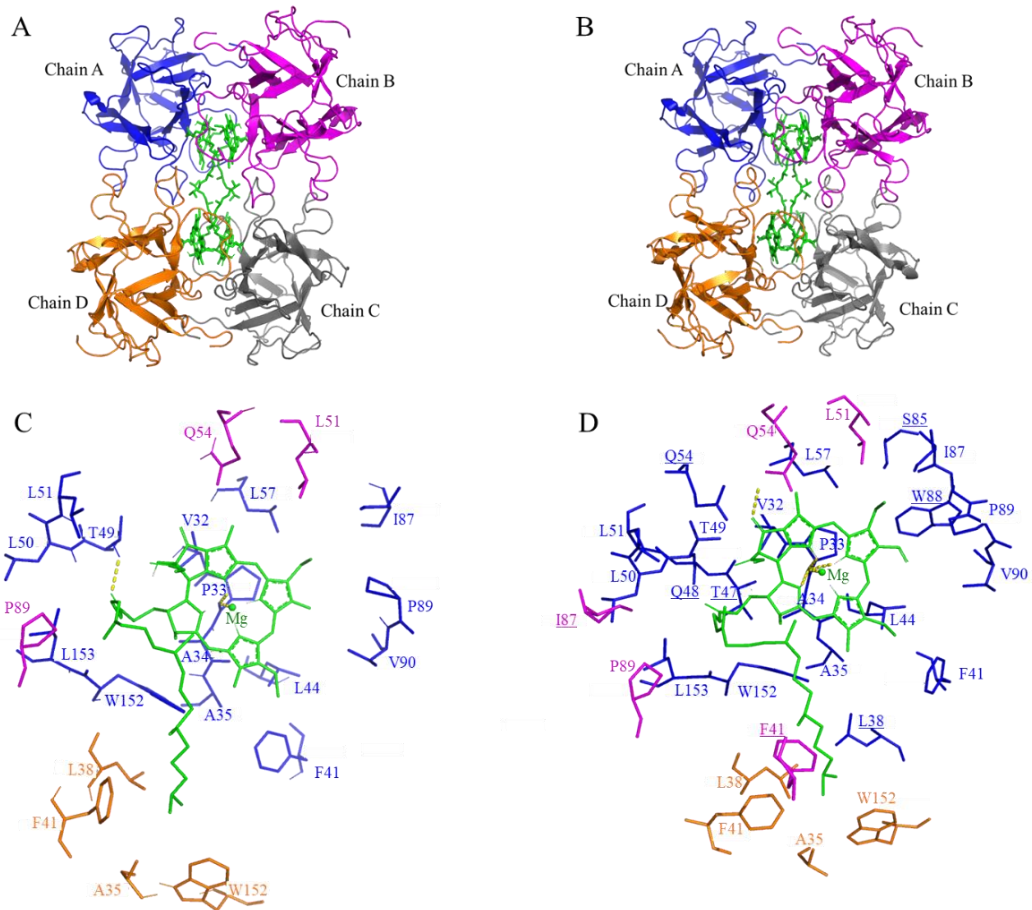
**Fig. 2: Thermostability and solubility of recombinant proteins with with (BnD22C) or without (BnD22) C-terminal extension peptide.**

### A) protein thermal shift assay

0.25 µg/µl of purified proteins were mixed with a fluorescent dye in 50 mM Tris-HCl (pH 7.5) buffer, 150 mM NaCl. The fluorescence intensity was monitored according a gradient of temperature. The data were fitted to the Boltzmann equation to determine the melting temperature (T<sub>m</sub>).

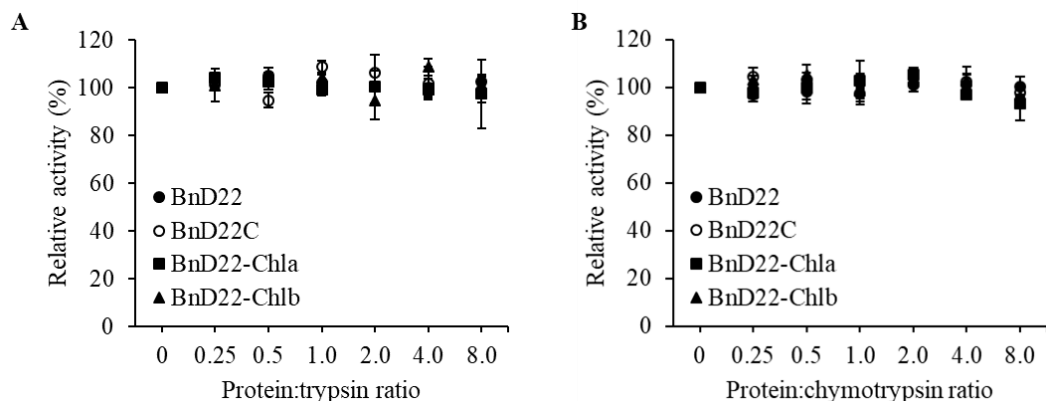
### B) size-exclusion chromatography elution profile

0.5 g/l of purified proteins were analyzed by analytical size exclusion chromatography in 50 mM Tris-HCl (pH 7.5) buffer, 150 mM NaCl, using a Superdex 200 column. The absorbance of proteins was monitored at 280 nm and normalized.



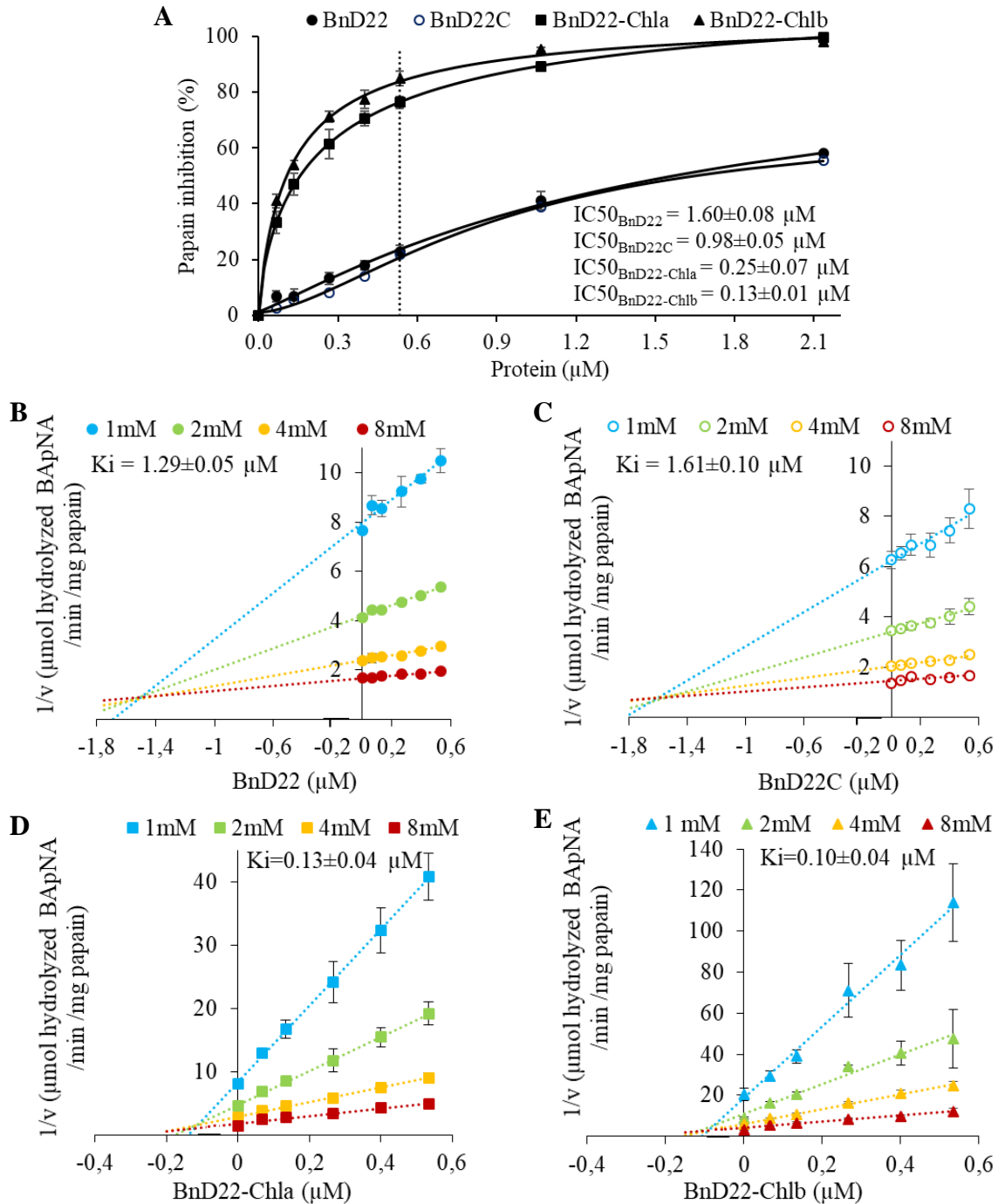
**Fig. 3: Structural model for the BnD22-Chls complexes**

Chla or Chlb are represented in green sticks. The chains A, B, C, D are shown in blue, magenta, grey and orange respectively. **A-B**) Ribbon diagrams of the BnD22-Chla (A) or BnD22-Chlb (B) tetrameric complexes. **B-C**) Details of the interaction between Chla (C) or Chlb (D) fixed by the chain A, with the whole tetrameric complex. The amino residues within 4 Å from the Chls are shown as stick with the same color of the polypeptide chains to which they belong. Amino acids specific to Chlb interaction are underlined. Hydrogen bond are indicated by yellow dashed lines.



**Fig. 4: Activities of trypsin (A) and chymotrypsin (B) proteases in presence of BnD22, BnD22C, BnD22-Chla or BnD22-Chlb complexes.**

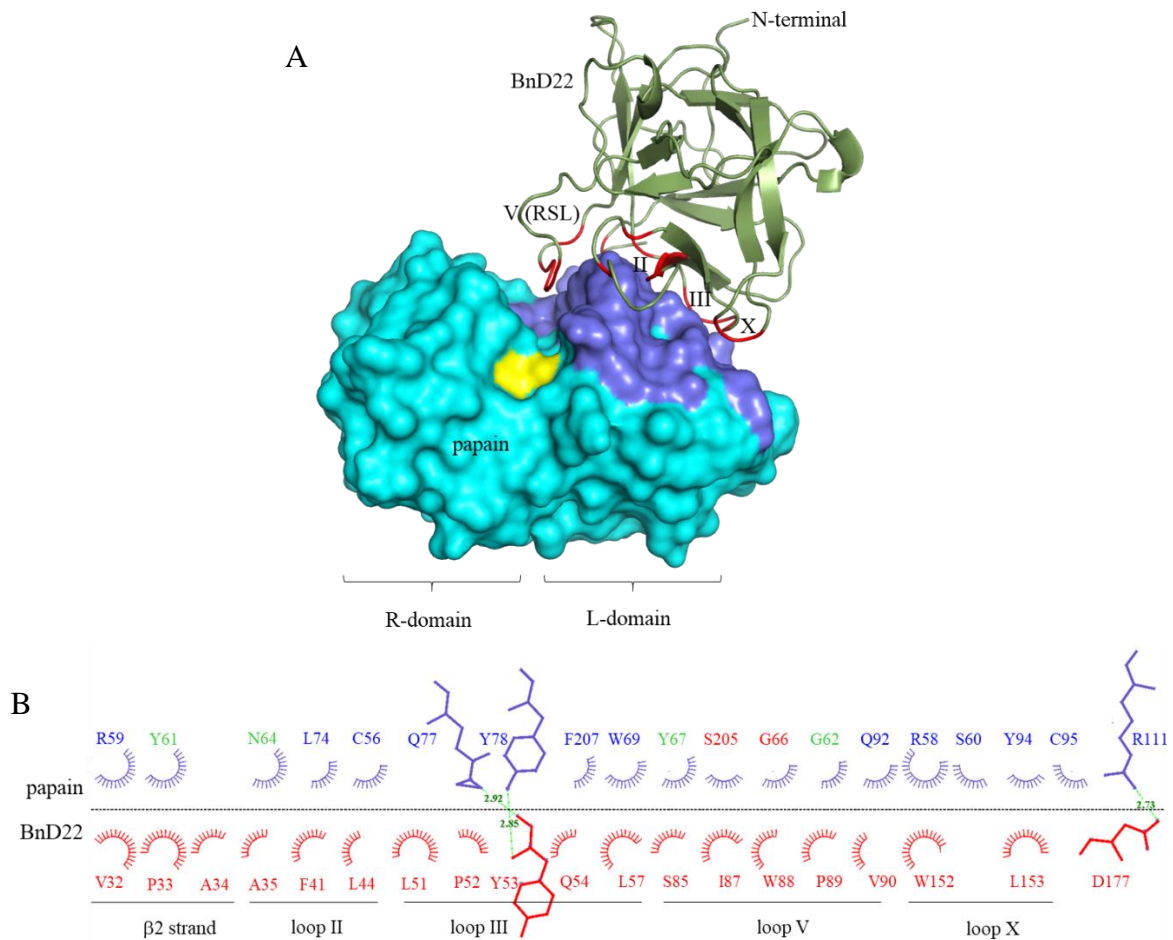
After incubation of trypsin or chymotrypsin in the presence of increasing concentrations of BnD22, BnD22-Chla or BnD22-Chlb complexes, enzyme assays were carried out by following hydrolysis of BAEE and BTEE respectively. Results are expressed as percentage of the residual enzyme activity compared to that without BnD22.



**Fig.5 Inhibition of papain protease in presence of BnD22C, BnD22, BnD22-Chla or BnD22-Chlb complexes.**

**A)** Titration curve of papain inhibition in presence of BnD22 with or without C-terminal extension peptide, or BnD22 associated with Chla or Chlb. Enzyme assays were carried out by following hydrolysis of 2 mM of BApNA. Papain inhibitions were determined for each concentration of WSCP and fitted to estimate  $IC_{50}$ . The dotted line indicated the WSCP:papain ratio =1.

**B-C-D-E)** Dixon plots. Papain activities with the presence of increasing concentrations of BnD22, BnD22C (**B-C**), BnD22-Chla or BnD22-Chlb complexes (**D-E**) at different concentrations of BApNA. The reciprocals of papain activity were plotted against the WSCP concentrations and the dissociation constants ( $K_i$ ) values were obtained from the intercepts of all lines at concentrations of substrate.



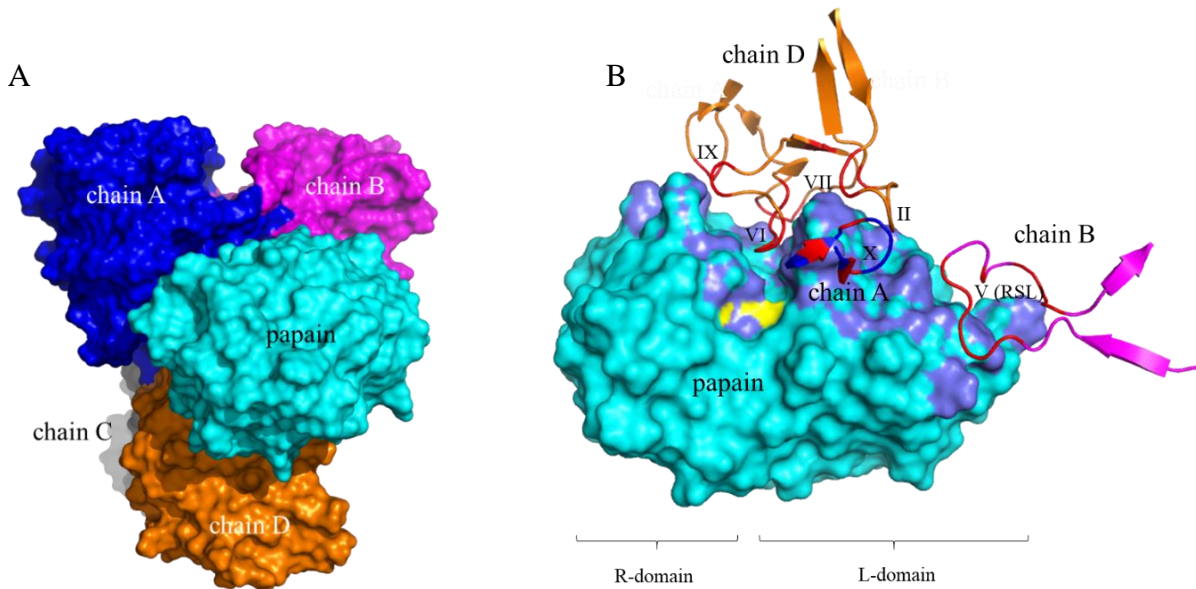
**Fig. 6: Interaction model of papain and BnD22.**

**A.** overall 3D model of BnD22-papain interaction.

The papain with their right (R) and left (L) domains is shown as cyan surface and its catalytic triad site (Cys25, His159, Asn175) in yellow. The BnD22 is shown in green cartoon. The interaction regions, with the amino acids within 4 Å from the two proteins, are indicated in blue for papain and in red for the loops II, III, V, X and  $\beta$ 2 strand for BnD22. RSL = reactive site loop.

**B.** 2D plot of interacting residues between papain and BnD22 .

The interface between BnD22 and papain is represented by the horizontal dashed line. The residues making hydrogen bonds are shown in sticks with hydrogen bonds as green dotted lines. The spoked arcs represent residues involved in hydrophobic contacts. In papain, the amino acid residue identified in the binding sites S2 and S3 are indicated in green and red respectively. Strand and loops are referred for BnD22. The 2D plot was generated by LigPlot+ software.



**Fig. 7: Structural model of the interaction between papain and BnD22-Chl complex**

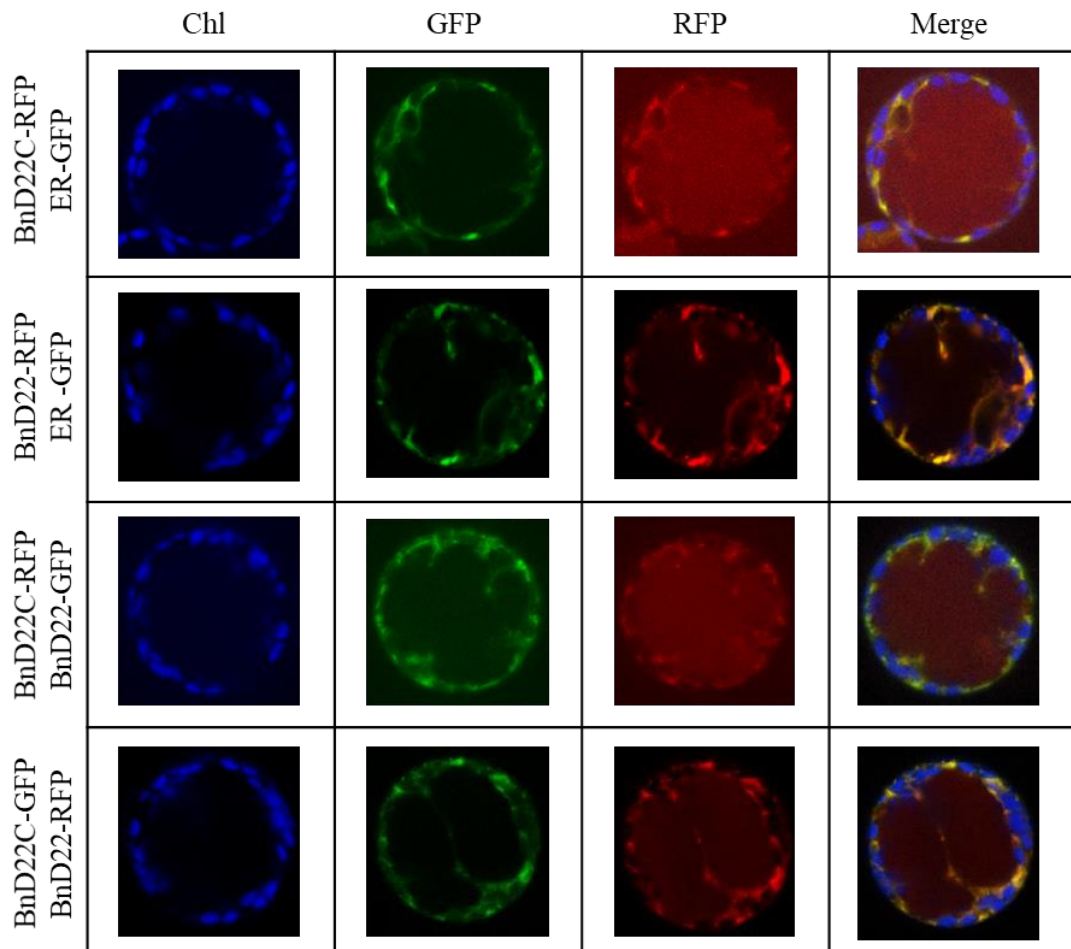
**A.** overall structural model of the interaction

The papain is shown as cyan surface. Its catalytic triad site (C25, H159, N175) is not visible. The four BnD22 (chains A to D) in the Chl complexes are indicated. Chls are omitted for clarity.

Papain can be associated with each sub-unit of the Chl complex. The papain linked mainly to chain D is also slightly associated to chains B and A. This binding into the BnD22-Chla complex represents the “front-side” interaction. On the other side, for the back-side interaction, papain is linked mainly to chain C. Papain associated with BnD22 of chain A or chain B have been called respectively left and right sides (see Fig. S3).

**B.** Focus on the interface between BnD22s and papain in the front-side model

The papain with their right (R) and left (L) lob domains is shown as cyan surface and its catalytic triad site (C25, H159, N175) in yellow. The interaction regions, with the amino acids within 4 Å from protease and the Chl complex, are indicated in blue for papain and in red for the BnD22 loops of chains A, B and D. RSL = reactive site loop.



**Fig. 8: Subcellular localization of BnD22 and BnD22C proteins in rapeseed protoplasts.**

Protoplasts from *B. napus* cotyledons were co-transformed with BnD22-GFP/RFP, BnD22C-GFP/RFP or with an endoplasmic reticulum protein marker (ER-GFP). Laser-scanning confocal fluorescence microscopy was used to visualize the proteins and the Chl.

## TABLES

### Table 1: Oligomeric state of the protein-Chl complexes

Recombinant proteins with (BnD22C) or without (BnD22) C-terminal extension peptide were reconstituted with chlorophyll (Chla or Chlb). The molecular weights (MW) of protein-Chl complexes were determined by size exclusion chromatography and subunits number in the complexes were calculated. MW of protein's monomer were determined by ExPasy tool (<https://web.expasy.org/protparam/>).

	Monomer MW (kDa)	Protein-Chl complexes				
			Elution volume (ml)	MW (kDa)	sub-units number	Oligomeric state
BnD22	20.08	Chla	13.88±0.02	79.03±0.63	3.9±0.03	tetramer
		Chlb	13.91±0.01	78.07±0.21	3.9±0.01	tetramer
BnD22C	22.71	Chla	13.32±0.01	102.27±0.27	4.5±0.01	tetramer
		Chlb	13.43±0.02	96.96±0.68	4.3±0.03	tetramer



**A**

		loop II			loop III			loop V			
		β2 strand	loop II		loop III			loop V			
		V32 P33 A34 A35 I36 N37 L38 L39 P40 F41 P43 L44 T47 Q48 T49 L50 L51 P52 Y53 Q54 L57	S85 D86 I87 W88 P89 Y90 N92 E93 L94 S95 K96								
in interaction with Chl	Chla	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
	Chlb	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue	Blue
in interaction with papain	monomer	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red
	Chla complexes	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red
	Chlb complexes	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red

		loop VI			loop VII			loop IX			loop X			loop XI		
		β7	loop VI		loop VII			loop IX			loop X			loop XI		
		A99 V100 D101 V102 S103 S104 S105 A106 K108 E109 P110 I111 I113 I114 G115 G116 E117 T119 A120 P121 N122 S141 Y142 P149 P151 W152 L153 S154 V160 T161														
in interaction with Chl	Chla	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	
	Chlb	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	
in interaction with papain	monomer	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	
	Chla complexes	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	
	Chlb complexes	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	Yellow	

**B**

		S3			S2			S1			S2			
		Y61 G62 C63 N64	G65	G66 Y67	W69 L74 Q77 Y78 V91 Q92 R93 Y94 C95 R98 E99 R111 Q112 Y116 V133 K156 V157	D158	H159	T204	S205	F207				
BnD22 monomer	monomer	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
	BnD22-Chla	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
	BnD22-Chlb	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green

**Table 2: BnD22-interacting residues with Chls and papain**

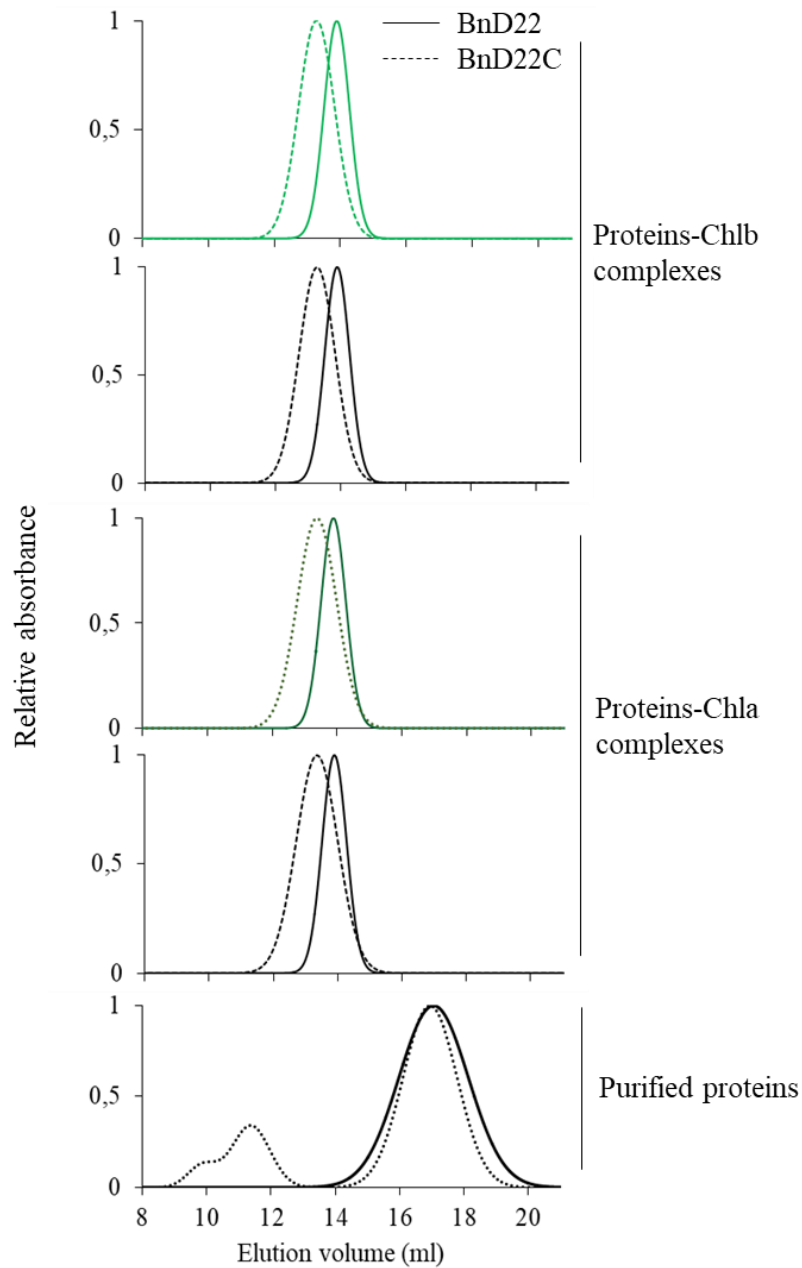
The amino acids within 4 Å from the Chls, the Cys protease and BnD22 were identified.

The papain linked to chain D is called “front-side” interaction. The colors of residues represent their membership to BnD22 chain A (blue), chain B (magenta) or chain D (orange) of the Chl complex. Strands and loops are referred for each BnD22. In papain, the amino acid residue identified in the binding sites S2, S3 or S1’ are indicated in green, red or orange respectively. In yellow, Cys 25 or H159 are residues of the triad which form the catalytic site of the enzyme.

**A.** Amino residues of BnD22 in interaction with Chls or with papain.

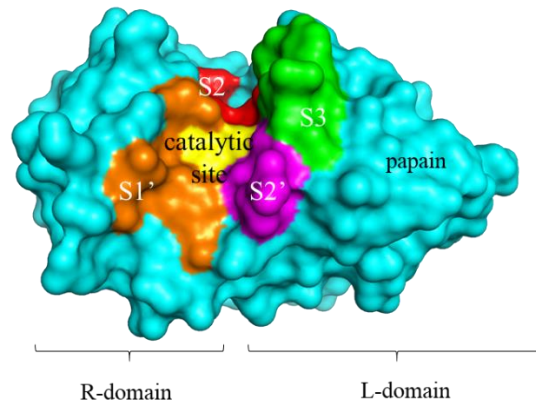
**B.** Amino acids of papain in interaction with BnD22 or BnD22-Chl.





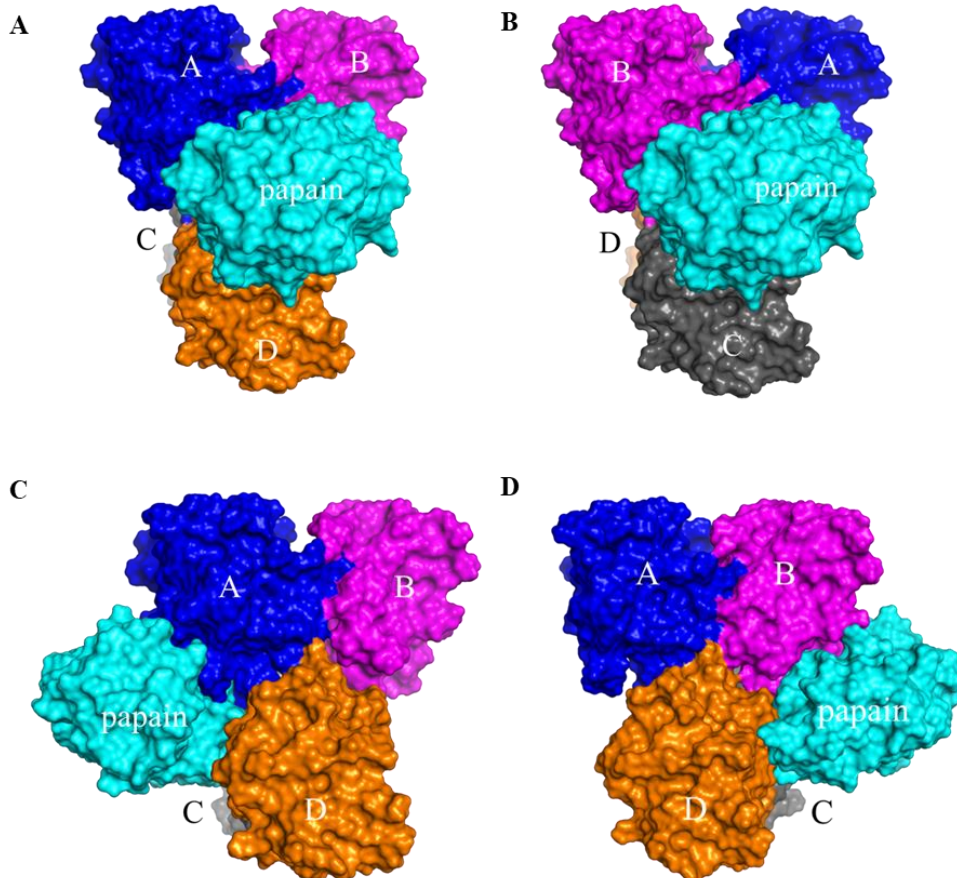
**Fig. S1: Size-exclusion chromatography elution profiles of purified protein and protein-Chl complexes**

0.5 g/l of purified proteins or proteins-Chl complexes were submitted to an analytical size exclusion chromatography in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, using a Superdex 200 column. The presence of proteins, Chla and Chlb were monitored at 280 (black), 420 and 452 nm (Green) respectively. BnD22C and BnD22 indicate protein with or without C-terminal extension peptide, respectively



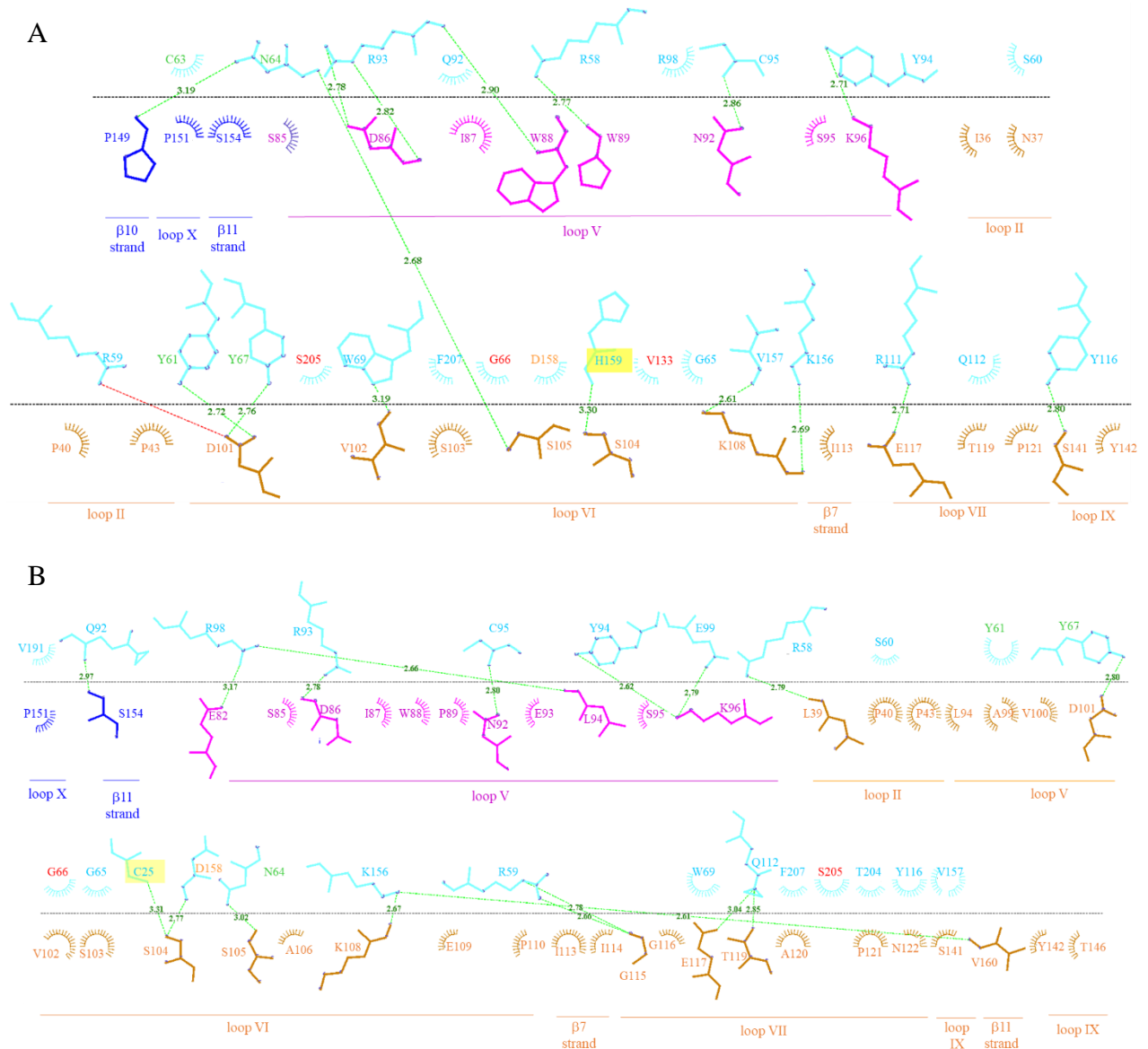
**Fig. S2: Active site of papain.**

The enzyme essentially consists of two domains separated on the top in an active site cleft in which the binding sites S3- S2'- S1'-S2' are represented. The different binding sites (S) are constituted by W27 G66 P68 V133 A160 S205 for S2 (in red), Y61 G62 C63 N64 Y67 for S3 (in green), A136 D158 Q142 W177 for S1' (orange), Q19 G20 S21 C22 G23 S24 for S2'(magenta). In the middle of active site cleft, in yellow, are the residues C25, H159, N175 triad which form the catalytic site of the enzyme.



**Fig. S3: structural model of different interaction positions between papain and BnD22-Chl complexes.**

By docking prediction, four papain could be associated in the BnD22-Chla or BnD22-Chlb complexes. Papain is represented in cyan and A, B, C, D designed the BnD22 chains inside the Chl complexes. Chls are omitted for clarity. The papain mainly linked to chain D and also to chains B and A is named front side position (A). The papain mainly linked to chain C and also to chains A and B is named back side position (B). The papain mainly linked to chain A and also to chains C and D is named left side position (C). The papain mainly linked to chain B and also to chains D and C is named right side position.



**Fig. S4: 2D plot of interacting residues between papain and BnD22-Chl: example of the front side model**

The papain linked mainly to chain D is also slightly associated to chains B and A. This binding into the BnD22-Chla complex represents the “front-side” interaction.

The amino acids within 4 Å from the Cys protease and BnD22 were identified for papain located at the front side linked to the chain D of the complexes constituted with Chla (**A**) or Chlb (**B**). The interface between BnD22 and papain is represented by the horizontal dashed line. The colors of residues represent their membership to BnD22 chain A (blue), chain B (magenta) or chain D (orange) of the Chl complex. Strand and loops are referred for each BnD22. In papain, the amino acid residue identified in the binding sites S2, S3 or S1’ are indicated in green, red or orange respectively. In yellow, H159 is one residue of the triad which form the catalytic site of the enzyme. The residues making hydrogen bonds are shown in sticks with hydrogen bonds as green dotted lines. The spoked arcs represent residues involved in hydrophobic contacts. The salt bridge is shown as red dotted lines. The 2D plots were generated by LigPlot+ software.

## CHAPITRE II :

### IDENTIFICATION DE LA FAMILLE DES WSCP II DU COLZA

---

#### Presentation

Le deuxième chapitre est consacré à l'identification des membres de la famille WSCP II du colza sur la base de leur caractérisation fonctionnelle et de leur expression tissulaire. Il est présenté sous la forme d'un troisième article en préparation :

**Article 3 :** Clouet V., Bouargalne Y., Rousseau-Gueutin M., Bouchereau A., Deleu C. and Le Cahérec F. (2022). Genome-wide identification and expression analysis of *WSCP II* genes in *Brassica napus* (en preparation).

**Contexte :** Plusieurs études suggérant la présence d'isoformes WSCP II chez colza ou le chou présageaient la présence de familles multigéniques (Ilami *et al.*, 1997; Desclos *et al.*, 2008; Prabahar *et al.*, 2020). La famille des *WSCP II* n'avait pas encore été étudiée de manière exhaustive chez le colza, espèce allotétraploïde, qui en fait un organisme modèle pour l'étude de l'évolution des polyploïdes.

**Objectif :** Les travaux présentés ont pour objet i- d'identifier de manière la plus aboutie possible la famille *WSCP II* du colza, ii- de retracer son histoire évolutive au sein des Brassicaceae et iii- de définir la sous-fonctionnalisation de ses membres par la caractérisation de leurs propriétés définies *in silico* et de leur expression tissulaire.

**Démarche expérimentale :** Après avoir répertorié et classé l'ensemble des gènes identifiés comme *WSCP II* à partir des données disponibles chez le colza et les espèces parentales, la stratégie a consisté à expliquer la présence des multiples copies alors qu'une seule WSCP est identifiée chez *A. thaliana*. La démarche repose sur :

- Une étude phylogénique menée en prenant en compte les *WSCP II* identifiés chez plusieurs Brassica et enrichie avec les copies de propriétés de liaison à la Chl définies.
- Une évaluation des propriétés fonctionnelles par une approche *in silico* en prenant en compte la présence ou non des acides aminés démontrés comme importants dans les interactions avec les Chl ou avec les Cys protéases, ou la présence du motif de liaison pour les fonctions de Ser IP.
- Un suivi des niveaux d'expression des *WSCP II* par RT-PCR quantitative dans différents tissus de colza cultivé en condition témoin en fonction du stade de développement (stade plantule, stades végétatifs et reproducteurs).

**Résultats :** 68 gènes de la famille WSCP II ont été identifiés pour quatre espèces Brassica dont 28 pour le colza. La famille des WSCP II se divise en 4 groupes phylogénétiques associés à leur préférence de liaison à la Chla ou à la Chlb ou à la perte de cette fonction. La présence des multiples copies chez les Brassica résulterait de duplications ancestrales avant l'évènement de triplication. Au sein de chaque groupe phylogénétique, des propriétés supplémentaires d'IP de protéases de type sérine ou cystéine ont été définies. Les études des niveaux d'expression des WSCP II ont révélé que *B. napus* pouvait présenter des profils d'expression des gènes WSCP II spécifiques aux organes photosynthétiques ou non, dépendamment des stades de développement des plantes.

**Conclusion :** Chez les Brassica, les WSCP II sont regroupés au sein d'une famille multigénique présentant des sous-groupes dépendamment de leur liaison aux chlorophylles. La présence de ces nombreuses copies de gènes WSCP pourrait s'expliquer par une sous-fonctionnalisation avec d'une part une divergence des propriétés fonctionnelles quant à leurs capacités de liaison aux Chls et d'inhibition envers les Ser et Cys protéases, et d'autre part une diversité d'expression tissulaire.

**Article 3:****Genome-wide identification and expression analysis of WSCPII genes in  
*Brassica napus***

Vanessa Clouet<sup>1</sup>, Youssef Bouargalne<sup>1</sup>, Mathieu Rousseau-Gueutin<sup>1</sup>, Alain Bouchereau<sup>1</sup>, Carole Deleu<sup>1</sup> and Françoise Le Cahérec<sup>1\*</sup>

**Author affiliations**

1. Univ Rennes 1, INRAE, Institut Agro, IGEPP - UMR 1349, 35653, Le Rheu, France.

**Corresponding author**

\*Le Cahérec Françoise

UMR 1349 IGEPP (Institut de génétique et protection des plantes), INRAE, Institut Agro, Université de Rennes 1, BP 35327, 35650 Le Rheu, France

francoise.le-caherec@univ-rennes1.fr

**ABSTRACT**

Class II's Brassicaceae water-soluble chlorophyll-binding proteins (WSCP) are non-photosynthetic proteins that bind chlorophylls (Chls) and Chl derivatives. They also belong to a subclass of the Kunitz protease inhibitor and could assume two different functions. Although their biological function and mode of action remain unclear, it is assumed they have a protective role, especially during stressful conditions. In the present study, to gain insight into the molecular diversity of WSCPs, a genome-wide analysis was performed in *B. napus*, an allotetraploid and model organism for the study of polyploid evolution, and in the two parentals *B. rapa* and *B. oleracea*. 55 WSCP genes were identified and characterized, including phylogeny analyses, gene structure, chromosomal localization, and expression pattern. The WSCP genes were divided into 4 phylogenetic groups, which shared highly conserved structural features within the same subfamily. Expression profiling revealed diverse patterns of different family members over various tissues. Certain genes were highly expressed in the non-photosynthetic organs as root or stem, and others are expressed in leaves or developing seeds-siliques. Our results provide a valuable reference to characterize WSCP genes, investigate the family evolution in Brassicaceae species beyond, and provide useful information for further functional analysis of the WSCP family in *B. napus* and related Brassica species.

**Keywords:** Kunitz protease inhibitor, gene duplication, pseudogene, chlorophyll-binding,

**Abbreviations**

At: *Arabidopsis thaliana*; BnD22: *Brassica napus* drought 22 kDa protein; Bol: *Brassica oleracea*; Bni: *B. nigra*; Bra: *Brassica rapa*; Chl: chlorophyll; LCH: light-harvesting chlorophyll a/b-binding protein of photosystem II; Lv: *Lepidium virginicum*; PI: protease inhibitor; UPGMA: unweighted pair group method with arithmetic mean; WSCP: water-soluble chlorophyll proteins.



## INTRODUCTION

The class II water-soluble chlorophyll proteins (WSCP II) constitute a small family of proteins that have been found in only a limited number of plant species belonging to Brassicaceae. Proteins of this family, discovered in 1971 (Murata *et al.*, 1971), have been further classified into two subclasses, class IIA and class IIB, based on their Chl *a/b* ratio. Although they share a short sequence motif for chlorophyll *a/b* (Chl) binding in light-harvesting binding proteins of photosystem II (LCH), they are not involved in photosynthesis. WSCPs differ from the classical LCH in several aspects: (i) they are not membrane-bound, (ii) they bind Chl*a* and Chl*b* at different ratio to form a tetrameric complex (Horigome *et al.*, 2007; Bednarczyk *et al.*, 2016; Palm *et al.*, 2018; Agostini *et al.*, 2019), (iii) they do not interact with carotenoids (Satoh *et al.*, 1998; Schmidt *et al.*, 2003; Hugues *et al.*, 2006), and (iv) they belong to a subclass of the Kunitz protease inhibitor (PI) (family I3 clan IC of the MEROPS classification) (for reviews, see Satoh *et al.*, 2001; Maleeva *et al.*, 2019).

Definite answers on the actual role of WSCPs *in planta* are still missing. Although their Chl-binding ability and Chl protection against photodamage (Schmidt *et al.*, 2003; Agostini *et al.*, 2017; Palm *et al.*, 2019), until now, no WSCP has been found located in the chloroplast but rather in a special endoplasmic reticulum(ER)-derived compartment so-called ER-body and vacuole (Takahashi *et al.*, 2012). Some WSCPs are also located in non-photosynthetic tissues such as roots flowers (Takahashi *et al.*, 2013ab; Boex-Fontvieille *et al.*, 2015) and cannot bind Chl (Prabakar *et al.*, 2020). Their presence in different tissues also changes according to the plant development. For example, AtWSCP is expressed in seedlings, absent in leaves, but present in inflorescence and siliques (Boex-Fontvieille *et al.*, 2016a). To date, no WSCP II is detected in the seeds. The induction of several WSCPs under various stresses (Downing *et al.*, 1992; Reviron *et al.*, 1992; Annamalai & Yanagihara, 1999; Desclos *et al.*, 2008) suggests protective functions in young tissue in growth, as a Chl scavenger (Takahashi *et al.*, 2012, 2013b) or ROS source in signaling pathway (Agostini *et al.*, 2017).

In contrast to their Chl binding capacity, their protease inhibition activities (PI) have never been deeply investigated. It has been demonstrated that AtWSCP from *Arabidopsis thaliana* has no significant serine protease inhibitor properties (Halls *et al.*, 2006) but inhibits cysteine proteases (Halls *et al.*, 2006; Boex-Fontvieille *et al.*, 2015a,b; Rustgi *et al.*, 2017). For BnD22 (*Brassica napus* drought 22 kDa), a WSCP II identified in rapeseed, it remains inconclusive whether or not it acts as a protease inhibitor. No trypsin inhibitory activity was detected *in vitro* (Ilami *et al.*, 1997; Nishio and Satoh, 1997). In contrast, Desclos *et al.* (2008) reported that BnD22 inhibited trypsin activity in young leaves treated with methyl jasmonate. Boex-Fontvieille *et al.* (2015b) showed that AtWSCP is not able to bind Chl and protease simultaneously. No supplemental data showing bifunctional propriety of WSCPs as Chl binding and protease inhibitor is available.

Some studies suggested the presence of the *WSCP II* multigenic family in Brassica. Thus, in rapeseed leaves, the purification of BnD22 is concomitant with other protein isoforms presenting very similar molecular masses (Ilami *et al.*, 1997), and *WSCP II* protein isoforms have been detected in rapeseed leaves (Desclos *et al.*, 2008). In addition, Prabahar *et al.* (2020) identified several *WSCP II*s in *Brassica oleracea* and *Brassica rapa*. In contrast, *A. thaliana* possess only a single copy of *WSCP II*. Concerning, only one phylogenetic study was performed on *WSCP* Chl-binding property, but the correlation, if any, between Chl-binding and proteases inhibition property was not highlighted (Prabahar *et al.*, 2020). Through numerous rounds of gene duplication, a common ancestor of *WSCP* genes drove the structural and functional diversification of the *WSCP* gene family (Prabahar *et al.*, 2020). This evolutionary diversification may be related to acquiring new specialized functions needed to adjust and adapt to diverse and complex plant growth conditions. *Brassica napus*, an amphidiploid (AACC,  $2n = 4x = 38$ ), originated from a natural crossing between two ancestral diploid parents, *B. oleracea* (CC,  $2n = 2x = 18$ ) and *B. rapa* (AA,  $2n = 2x = 20$ ), approximately 7500 years ago (Chalhoub *et al.*, 2014). Brassica species belonging to the Brassica genus and Brassicaceae family offer a valuable model for studying polyploid genome evolution, mechanisms associated with gene duplications, loss of duplicated genes and gene neo- and sub- functionalization (Chen *et al.*, 2018; Liu *et al.*, 2018). The *WSCP II* of *B. napus*, *B. rapa*, and *B. oleracea* genome sequences has provided an opportunity to identify and characterize key genes from a genome-wide perspective (Belser *et al.*, 2018; Rousseau-Gueutin *et al.*, 2020; Istace *et al.*, 2021).

This study provides a complete description of the *WSCP II* gene family in *B. napus* by performing a genome-wide *in silico* identification, characterization, and evolutionary and functional analysis of the *WSCP* gene family. We report identification of 28 genes from *B. napus*, members of *WSCP* family, clustered into four groups. The physical properties of proteins, chromosomal location, gene structure, motif analysis and phylogenetic relationships with gene duplication were investigated. Orthology and synteny analysis were also carried out to explore the evolutionary history and divergence of the *WSCP* family in *B. napus*, *B. rapa*, *B. oleracea*. *WSCP II* expression profiles in different tissues at seedling, vegetative and reproductive stages highlighted their potential roles in regulating distinct and diverse developmental processes and stress responses.

## MATERIALS AND METHODS

### 1. Genome-wide identification of WSCP in *Brassica napus* and its progenitor species

*B. napus* BnD22 (X65637, Downing *et al.*, 1992), *Arabidopsis thaliana* (At1g72290, Halls *et al.*, 2006) and *Lepidium virginicum* LvWSCP (AB002589, Takahashi *et al.*, 2013b) members of class WSCPIIB were used as queries to a BlastN search of the whole genome sequences of *B. napus* (cv. Darmor-bzh, v10, Rousseau-Gueutin *et al.*, 2020), *B. rapa* (cv. Z1, v2, Istace *et al.*, 2021) and *B. oleracea* (cv HDEM, v1, Belser *et al.*, 2018). The revealing of WSCP II pseudogenes in *B. napus* was checked by genomic DNA sequencing. All WSCP genes identified from *B. napus* were used as queries to complete the screen of the *B. napus* and the parental genomes. Amino acid sequences were also blasted in MEROPS database (<http://merops.sanger.ac.uk>) (Rawlings *et al.*, 2013), and only genes coding proteins belonging to the I3A.028 subclass were retained. The exon-intron structures of the WSCP genes were monitored from information given by the automatic annotations.

### 2. Chromosomal distribution and gene duplication

All identified WSCP genes were mapped to their respective chromosomes based on the physical location from genomic information using package R ‘chromPlot’ v 1.22.0. To examine the syntenic relationships of the orthologous WSCP family genes obtained from *B. napus* and parental species, synteny maps were constructed using the Circos software (Krzywinski *et al.*, 2009).

### 3. Protein characteristics

Signal peptides were determined with SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP/>). Cleavage site signal, the physicochemical parameters of Bna- Bol- Bra-WSCP proteins, including the molecular weights and isoelectric points (pIs), were determined by CLC Main Workbench Version 22.0 (QIAGEN). The aliphatic index of protein, regarded as a positive factor for increasing the thermostability of globular proteins, is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It is calculated as  $X(\text{Ala}) + a * X(\text{Val}) + b * (X(\text{Ile}) + X(\text{Leu}))$  where  $X(\text{Ala})$ ,  $X(\text{Val})$ ,  $X(\text{Ile})$ , and  $X(\text{Leu})$  are mole percent (100 X mole fraction) of alanine, valine, isoleucine, and leucine. The coefficients  $a$  and  $b$  are the relative volume of the valine side chain ( $a = 2.9$ ) and Leu/Ile side chains ( $b = 3.9$ ) to the side chain of alanine. Subcellular localization was predicted from the translated nucleotidic sequences with PSORT software (Horton *et al.*, 2007).

Kunitz and WSCP chlorophyll-binding motifs were defined by homology to consensus sequences [LIVM]-x-D-[EK]-[EDNTY]-[DG]-[RKH DENQ]-x-[LIVM]-x(5)-Y-x-[LIVM] (PROSITE ID: PS00283) and “PFCPLGI” defined by Satoh *et al.* (1998, 2001) respectively. Chl interacting and Chl affinity residues were identified by comparison with those described by Horigome

*et al.* (2007) for WSCPIIB, Agostini *et al.* (2019), and Bednarczyk *et al.* (2016) for WSCPIIA BoWSCP, and by Palm *et al.* (2018) for the Chl affinity motif. The Ser protease reactive-site loop location, activity and location of the highest potential sites for trypsin or chymotrypsin cleavages were obtained through Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) and PeptideCutter (<http://www.expasy.org/tools/peptidecutter/>). The Cys protease inhibition was estimated by the presence or absence of amino acids defined in the AtWSCP and papain-like Cys proteases (RD21, PMP, or XBCP3) interactions (Boex-Fontvieille *et al.*, 2015b; Rustgi *et al.*, 2017).

#### 4. Phylogenetic analysis of the WSCP family

To illustrate the evolutionary relationships of the WSCP, nucleic sequences from various Brassicaceae species whose genomes are sequenced and representative of the Brassicaceae family evolutionary (*A. thaliana*, *B. napus*, *B. oleracea*, *B. nigra*, *B. rapa*, *L. virginicum*) and the proteins identified by Prabahar *et al.* (2020) were used to generate the phylogenetic tree. WSCP of *B. nigra* cv Ni100 and *Aethionema arabicum* were identified by genome-wide analysis (Nguyen *et al.*, 2019; Sampath *et al.*, 2020). WSCP from *Aethionema arabicum* was used as an outgroup. Sequence alignments, models tests, and phylogenetic analysis were performed with MEGA 11.0.11 software (Tamura *et al.*, 2013). The best substitution models were found with the maximum likelihood statistical method. The evolutionary phylogenetic gene tree was generated with the maximum likelihood statistical method (1000 bootstraps replications) and the best nucleotide substitution model: Tamura 3-parameter (T92) with a gamma-distributed rate among sites (G).

#### 5. Quantitative analysis of transcript levels

##### - Plant material and growth conditions

Seeds of oilseed rape *B. napus* genotype Aviso were obtained from our 'Bracysol' biological resource center. Seedlings were grown on Murashige and Skoog (MS) agarose medium. After 2-3 days at 4°C in the dark to synchronize germination, plates were moved to a growth room at 22°C having a 12 h-light period (light intensity of 100  $\mu\text{mol.m}^{-2}.\text{s}^{-1}$ ) and 70% relative humidity. Plates with seedlings were kept vertically, and their tops were not wrapped to allow transpiration. After 4 days of development, roots, cotyledons, and hypocotyls were collected 7 h after light onset to consider the circadian rhythms of plants. For the vegetative stage, plants, individually sown in two-liters pots filled with perlite, were grown in a greenhouse with a thermo-photoperiod of 25°C/16 h day and 18°C/8 h night under natural light supplemented to ensure a minimum photosynthetically active radiation of 200  $\mu\text{mol photons m}^{-2} \text{s}^{-1}$  at the top of the leaf canopy. They were regularly watered to full capacity with a solution containing 8 mM nitrate (Albert *et al.*, 2012). Sixty days after sowing (DAS), stems,

leaves and roots of three plants were collected individually. The laminae of leaves were separated from midribs and the larger veins and petioles and organized in three sets according to their stage of development. For the reproductive stage, after 40 DAS of growth, plants were placed for two months in a 4 °C growth chamber for vernalization and put back into the greenhouse. At 130 DAS, three plants were harvested and sampled. From each plant, the four oldest and the four youngest leaves were collected individually, separated into veins (including midribs and the larger veins), petioles, and laminae. Flowers, fine roots, taproots, stems and young and mature pods were also collected (Faës *et al.*, 2014).

*- RNA extraction and reverse transcription*

Total RNAs extraction, reverse transcription and quantitative PCR were performed according to Delloero *et al.* (2020). Briefly, total RNAs were extracted from frozen tissue with the NucleoSpin RNA Plus Kit (Macherey Nagel) following the manufacturer's instructions. After checking for the absence of genomic DNA and RNA integrity, RNA treated with the DNase was reverse-transcribed using oligo(dT)15 primer and Goscript reverse transcriptase (Promega) according to the manufacturer's instructions. For quantitative PCR (qPCR) experiments, diluted cDNA was mixed with the LightCycler 480 SYBR Green I Master mix (Roche) and primers pairs (Table S1). qPCRs were achieved on a Light Cycler LC480 (Roche) under the following conditions: an initial step at 95°C for 5 min, then 40 cycles of 95°C for 15 s and 60°C (or T<sub>m</sub>) for 40 s. Using the GeNorm method (Vandesompele *et al.*, 2002), different couples of reference genes were selected for their expression stability in the samples harvested from different tissues. For the samples harvested at the seedling stage, *RibS3* and *UBQ11* were selected and for those from the vegetative and reproductive stages, *Bna.PTB* and *Bna.PP2A* were used. Gene expression was normalized using the geometrical mean expression of the reference genes and the efficiency of qPCR primers.

## RESULTS

### 1. Identification and characterization of *WSCP* gene family in *Brassica napus* and its diploid progenitors

The identification of the *WSCP* family genes in *B. napus* was performed by similarity with BnD22 (X65637) the only *WSCP* known in *B. napus*, *AtWSCP* (At1g72290) from *A. thaliana* and *LvWSCP* (AB002589) from *L. virginicum*, the WSCPIB reference. 28 *WSCP* genes were identified, all predicted by the genome annotation. 9 of them were shown to encode pseudogenes and were not further analyzed (Fig. 1, Fig. S1, Table S2). All the pseudogenes, except *C02p29790.1\_BnaDAR* and *A07p38390.1\_BnaDAR*, are found on chromosome C06 (Fig. 2). The *BnaWSCP* genes were distributed only on a few chromosomes with 19 *BnaCWSCP* positioned on sub-genome C (4, 1, 13 and 1 on chromosomes C02, C05, C06, C07 respectively) and 9 *BnaAWSCP* located on sub-genome A (3, 5 and 1 on chromosomes A02, A07, A09 respectively) (Fig. 2, Table S2). Similarly, we also identified 9 *BraWSCP* and 19 *BolWSCP* in *B. rapa* and *B. oleracea*, respectively (Fig. 2, Table S3). 8 pseudogenes were detected in *B. oleracea* with 7 located in chromosome C06 and one in chromosome C02. Chromosomal locations of *WSCP* in *B. rapa* and *B. oleracea* are identical to that of *B. napus* with *BraWSCP*s mainly positioned in chromosomes A02 and A07 with one gene in chromosome A09 and *BolWSCP*s clustered in chromosomes C06 with three in chromosome C02 and one in chromosomes C05 and C07.

*WSCP* genes were clustered in four groups (Fig. 1). Belong the group, group 1 is the largest with 25 genes including 11 pseudogenes (Table 1). *AtWSCP*, a WSCPIIA, is found in this group. In *B.napus*, the copies are located on chromosomes C06-A07 and C02-A02. These locations are not specific to group 1. The Group 2 is constituted by 8 genes including 1 pseudogene. The 13 genes of the group 3, including 3 pseudogenes, are also found in chromosomes C02-A02 and C06-A07 respectively. The genes of group 4 are the only ones located in the chromosomes C05-A09 and C07. The group 4, where *LvWSCP* colocalized, would correspond to the class *IIB*.

### 2. Synteny and duplicated gene analysis of the *WSCP* family genes in *B. rapa*, *B. oleracea* and *B. napus*

To investigate the patterns of retention or loss in the orthologous *WSCP* family genes, we compared the relationships of the *WSCP* genes between *B. napus*, *B. oleracea* and *B. rapa* (Fig. 3). Conserved synteny between these three species can be observed. The pair of homoeologous genes *BnaA.WSCP*II and *BnaC.WSCP*II is located on homeologous chromosomes (for example, pairs A02-C02, A07-C06). *BnaWSCP*II genes have experimented with independent tandem duplications located in chromosomes A02, C02, C06 and A07 and produce clusters with respectively 3, 4, 13 and 5

paralogs (Fig. 2). Furthermore, the positions of *WSCP II* inside the groups located on chromosomes C02-A02, C06-A07 are respected between *B. napus* and the parental chromosomes.

All *BraWSCP* genes and *BolWSCP* genes showed a syntenic relationship with the *BnaA*- and *BnaC*- *WSCP* genes (Table 1). Furthermore, 9 and 19 orthologous pairs were identified in *BraWSCP*-*BnaAWSCP*, and *BolWSCP*-*BnaCWSCP* respectively. *WSCP IIA* genes could be associated with each other in complete syntenic groups with homologous genes including two *B. napus* of each sub-genome, one *B. rapa* and one *B. oleracea*. This is not the case for most of the *BnaCWSCP IIA*s located on Chr C06 which have no homeologs on the A subgenome and no *B. rapa* paralogs found (Table 1). It should also be noted that only one *BraWSCP* gene, *A07p40690.2\_BraZ1*, is related to a pseudogene in *B. napus* (*A07p38390.1\_BnaDAR*). *BolC2t09022H* and *BolC2t09025H* are duplicated in *B. napus* genome. The pseudogenes in *B. oleracea* remain nonfunctional in *B. napus* genome or are lost as *BolC2t09020H* and *BolC6t39704H*. These results show that the syntenic *WSCP II* gene pairs are widely distributed on the genomes of the allotetraploid *B. napus* and its diploid progenitors.

### 3. Phylogenetic analysis

To investigate the evolutionary relationships among the *WSCP* family, the nucleotidic sequences of *WSCP*s from various Brassicaceae species whose genomes were sequenced (*B. napus*, *B. nigra*, *B. oleracea* and *B. rapa*). 12 *WSCP*s were thus identified in *B. nigra* (sub-genome B). Additional 5 *WSCP*s identified by Prabahar *et al.* (2020) were selected to generate the phylogenetic tree. The phylogenetic tree was rooted on the group composed of *A. arabicum* *WSCP*s. 67 *WSCP*s were thus identified (Fig. S1-S2). The phylogenetic tree was rooted on the group composed of *A. arabicum* *WSCP*s. The others *WSCP*s were splitted into two major clades (Fig. 4). One major clade includes all *WSCP IIB*, as *LvWSCP*, that are phylogenetically distinct from the other *WSCP*s clustered into 3 different phylogenetic groups corresponding to *WSCP IIA*, *WSCP IIX* as described by Prabahar *et al.* (2020) and an additional one named *WSCP IIC*. The phylogenetic tree illustrates the expansion and divergence of the *WSCP* genes from *Arabidopsis* to *B. napus* with an apparent disassociation of *AtWSCP* from all the Brassica *WSCP*s genes. Moreover, the IIB and IIX clades are sub-divided into two sub-groups named IIB (1-2) and IIX (1-2). In chromosomes, sub-group IIA is mixed with IIX1 or IIB and the subgroup IIC with IIX1. Moreover, *WSCP* in the allotetraploid *B. napus* and the diploids *B. rapa*, *B. oleracea* and *B. nigra* are related in the same phylogenetic branch, suggesting ancestral conserved duplication (Fig. 5). Subgroup IIB2, without *BniWSCP*, seems to result from more recent duplications and some genes located in chromosome B03 of *B. nigra* have no orthologous in the genome of *B. rapa*. We assigned new identifiers to the *WSCP*s by using the prefix “Bna, Bni, Bol or Bra” for *B. napus*, *B. nigra*, *B. oleracea*, or *B. rapa*, taking into account i) A, B or C genome location, ii) their phylogenetic membership in the major and sub-groups clades IIA, IIB (1-2), IIC or

IIX(1-2) and iii) their homology to BnD22 (*C06p46460.1\_BnaDAR* in v10 renamed *BnaC.WSCP IIA.a1*, Table 1).

#### 4. WSCP II proteins properties

The physical and chemical properties of WSCP proteins are outlined in Table S2. The molecular weight of WSCP proteins was around 20 kDa, except for WSCP IIBs which have sizes of 24 kDa. WSCPs are acid, neutral or basic proteins with a predicted isoelectric point (pI) ranging from 4.95 to 9.12. They show an important aliphatic index reflecting a potential high thermostability (Fig. 6, Table S2). Globally, even if the orthologous and homeologous proteins display same properties (molecular weight, isoelectric point and aliphatic index), acid and basic proteins are present in a same phylogenetic group. For example, proteins in sub-groups B1 and IB2 differ in their pI with acid WSCP IIB1 proteins and basic WSCP IIB2 proteins with a size larger than the average WSCP II.

The WSCP II proteins from *B. napus* and parents show a high diversity with identities at the lowest at 26%, with a greater homogeneity within the same phylogenetic group (Table S4). Some of WSCP II are identical (Table S4E). Despite this diversity, all the identified WSCP II proteins contain regions closely matching the Kunitz consensus motif [LIVM]-x-D-{EK}-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-(x)5-Y-x-[LIVM] (PROSITE PS00283) (Fig. S3). These proteins were classed as WSCP, because of the presence of a conserved region “chlorophyll-binding motif” “PFCPLGI” as noted by Satoh *et al.* (1998, 2001) designed from the signature motif called the “(F/Y)DPLGL motif” of light-harvesting Chl a/b-binding protein (Green and Kühlbrandt, 1995). To confirm their identity as WSCP, sequences were tested in the MEROPS database (Rawlings *et al.*, 2013). All the sequences have been classed as protease inhibitors type I3, clan IC. 11 amino acids, EKLG(LK/R)MFPFY, are highly conserved in the carboxy-terminal terminus, which is removed during the post-translational maturation.

#### 5. *In silico* properties of WSCP protein as Chl-binding and protease inhibition abilities

The presence of signature motifs prevents any reliable prediction of the putative active function as PI or WSCP. PI activity is due to a reactive site, which reacts with the binding site of the protease. This reactive site would be located in loop IV for Ser PI activities. Although the sequences of the reactive sites are highly variable, depending on the nature of the amino acids, the potential of PI against Ser protease (trypsin and chymotrypsin) can be predicted. The Cys PI activity was estimated according the presence or not of amino acids defined for interaction between AtWSCP and papain-like Cys proteases (RD21, PMP or XBCP3) (Boex-Fontvielle *et al.*, 2015; Rutsgi *et al.*, 2017). The Chl binding capacity prediction was based on the presence of amino acids in contact with chlorophyll and Chl binding domains (CBD) defined by Horigome *et al.*, (2007); Palm *et al.*, (2018);



and Prabahar *et al.*, (2020). Chl binding specificity was determined by the presence of Chl-affinity motif defined by Palm *et al.* (2018) (Fig. 7).

Globally, each pair of orthologous (BnaA-BraA, BnaC-BolC) shows a similar pattern of possible inhibition against Ser- Cys- protease and Chl binding capacity. Different properties can differentiate homeologues as for example the WSCP IIB2 A07p11180.1\_BnaDAR with *in silico* a strong detected PI missing for the homolog C07p16850.1\_BnaDAR. Only A07p38420.1\_BnaDAR also shows a chymotrypsin PI potential. In addition, WSCP IIA proteins A07p38410.1\_BnaDAR, C06p46430.1\_BnaDAR and WSCP IIB proteins A02p17890.1\_BnaDAR, C06p46550.1\_BnaDAR have a strongly possible P1 residue for a typical trypsin cleavage. It is also possible to class those proteins regarding their degrees of Cys PI activity. C06p46460.1\_BnaDAR and C06p46430.1\_BnaDAR (WSCP IIA) and some WSCP IIX1 as C02p29730.1\_BnaDAR, C02p29750.1\_BnaDAR, A02p17860.1\_BnaDAR (Fig. 7A) are the only rapeseed WSCPs that have most of the amino acids shown to be important in the AtWSCP-RD21 interaction and would therefore be able to be Cys PI. It should be noted that PI properties of WSCP II are thus no specific to one phylogenetic group and no WSCP II presents the potential to inhibit Ser and Cys proteases. WSCP IC proteins seem to exhibit any properties as PI or Chl binding molecules.

For Chl binding capacity, WSCP IIX clearly showing no amino acids compatible with Chl interaction are not able to bind chlorophyll as expected. Only WSCP IIA C06p46520.1\_BnaDAR and C06p46460.1\_BnaDAR have all the amino acids predicted essential and the others proteins present substitutions which could cause either gradual affinities or no connection at all with Chl. Moreover WSCP IIB is divided according to the nature of the selectivity binding motif defined by Palm *et al.* (2018) with sub-group IIB2 that should be more affine to Chla compared to IIB1 (Fig. 7B).

#### 4. Expression profiles of the *BnaWSCP II* family genes in various tissues and organs

The presence of multiple copies of *WSCP II* genes in *B. napus* genome raised the question about the specific localization of each gene expression. *BnaWSCP* transcripts were quantified in oilseed rape, grown under control conditions, at seedling, vegetative and reproductive stages.

The WSCP IIA C06p46430.1\_BnaDAR is the major WSCP II expressed specifically in seedling cotyledons. Three other WSCP IIA are detected in cotyledons: C06p46460.1\_BnaDAR (*BnD22*) and two WSCP IIB (C05p27460.1\_BnaDAR and C07p16850.1\_BnaDAR). The WSCP IIB was also detected in hypocotyl (Fig. 8, table 2). At the vegetative and reproductive stages, the WSCP IIB was very weakly expressed or not detected in plants grown in control conditions and their expression pattern does not show any tissue specificity. WSCP IC genes also show a similar pattern without real expression and tissue preference (Table 2). The WSCP IIX is strongly expressed in stems and roots and its expression seems to be specific to the vegetative stage, whereas WSCP IIX2 RNAs persist in

the reproductive stage and extend to the reproductive organs (flowers and pods) (Fig. 9). *WSCP IIA* genes are the major *WSCP II* detected in leaves with strong expressions for *C06p46460.1* and *C06p46430.1* (Fig. 10). The highest expression site of some of them becomes the reproductive organs (flowers and pods). Moreover, the homeologous *C06p43520.1-A07p38410.1* shows different expression patterns with a very weak expression for copy C compared to copy A, which is detected in leaves and represents the major form *WSCP* in pods and flowers (Fig. 10, table 2).

## DISCUSSION

The phylogenetic tree illustrates the expansion and divergence of the *WSCP II* genes from *Arabidopsis* to *B. napus* with a clear disassociation of *AtWSCP* from all the Brassica *WSCP* genes. The *WSCP II* family genes had undergone a whole-genome triplication since their divergence from *A. thaliana*, results in a significant increase in the duplicated genes. Although the allotetraploid *B. napus* was formed by the natural hybridization and polyploidization of *B. rapa* and *B. oleracea*, the number of *WSCP II* and their existence in *B. nigra* genome indicates that the expansion of the *WSCP II* genes was mainly due to a result of earlier whole-genome and segmental duplications. However, different numbers of *WSCP*s were identified in the diploid species *B. rapa*, *B. nigra* and *B. oleracea* suggesting that gene loss or duplication events might have occurred in the *WSCP* family genes. Sub-genome C appears subject to many more tandem duplications with numerous pseudogenes. This suggests that the evolutionary history of *WSCP II* is not finalized.

*WSCP II* family can be categorized into two sub-classes: the class IIA with Chla/b ratio higher than 6 and the class IIB with Chla/b ratio lower than 3.5 (Sato *et al.*, 2001). Phylogenetic analysis revealed that *WSCP*s were clustered according to their Chl-binding ability to bind Chl. They could be divided into two major clades with a clear separate evolution of the *WSCP IIB*. Prabahar *et al.* (2020) suggested that during *WSCP II* gene family evolution, the ability to bind Chlb (*WSCP IIB*) was gained later than the ability to bind Chla (*WSCP IIA*), and that a second evolutionary trend toward the loss of Chl-binding function (*WSCP IIX*). Our phylogenetic analysis of an extended set of Brassica *WSCP*s in *B. napus*, *B. nigra*, *B. oleracea* and *B. rapa* confirmed a Chl binding-dependent evolution event of *WSCP II*.

Girr *et al.* (2020) reported that the affinities mainly controlled the Chl selectivity. The preference of *WSCP IIB* for Chlb has been shown for several proteins, while proteins of class IIA have no clear preference (Takahashi *et al.*, 2013a, 2013b; Bednarczyk *et al.*, 2015; Prabahar *et al.*, 2020). BoWSCP (BoIC6t39693H), a *WSCP IIA*, shows any Chl preference (Girr *et al.*, 2020). In contrast, in the same sub-group, BnD22 (C06p464601\_BnaDAR) displays a higher affinity for Chla (Bouargalne *et al.*, 2022, in press). It is in agreement with the Chla/Chlb contents in native and recombinant *WSCP IIA* (Murata *et al.*, 1971). The Chla/Chlb specificity involves a loop in close contact to the C<sub>7</sub> side chain of the chlorin macrocycle, where Chla differs structurally from Chlb (Agostini *et al.*, 2019; Palm *et al.*, 2019). The preference of BnD22 for Chla could partly be explained by the presence of the PVCNEL motif in this loop, causing an unfavorable steric environment for interaction with the Chlb (Palm *et al.*, 2018). However, the same motif is present in BoWSCP that failed to show any Chl preference. In the class IIA, amino acid substitutions in the motif « PVCNEL » defined by Palm *et al.* (2018) could reveal different levels of affinity to Chla (Fig. 7). The sub-division inside the *WSCP IIB* can also be explained by a difference in Chl selectivity. Thus, sub-group IIB1

collects the most Chlb affine proteins characterized by the specific motif “LCPS” defined by Palm *et al.* (2018). The WSCP IIB2 would have a higher affinity to Chla compared to WSCP IIB1 because of a single amino acid exchange in the motif (PCPS instead of LCPS) as demonstrated by Palm *et al.* (2018). The new method developed by Bouargalne *et al.* (2022) should offer new opportunities further explore the WSCP II affinity and selectivity to Chls.

Besides chlorophyll selectivity, pigment binding capacity inside each phylogenetic group is also questionable. Thus, in Bna- Bol- and Bna- WSCP IIB, unlike the ancestral proteins defined by Prabahar *et al.* (2020), one of the tryptophan residue defined as important (Horigome *et al.*, 2007) is replaced by a tyrosine (IIB2) or a phenylalanine (IIB1). This amino acid absence is also noted for the rapeseed WSCP IIA C06p46430.1 and the parental proteins. Similarly, other substitutions are pointed out for the other amino acids targeted by Palm *et al.* (2018). Thus, WSCP II of the same phylogenetic group does not seem to share similar Chl functions. For the WSCP IIX and WSCP IIC identified in *B. napus*, *B. oleracea* and *B. rapa*, it is clear that they have lost their ability to bind to Chl which is highlighted by the absence of the primordial amino acids for the Chl interaction. This loss of capacity and additional functions as PI could be consistent with the specific expression of WSCP IIX in nongreen plant tissues such as roots and stems. The distinction between the two subgroups could be explained by the maintenance of WSCP IIX2 expression at the reproductive stage. WSCP IIC, which show no apparent PI or Chl functions, are poorly expressed in oilseed rape grown without environmental constraints. A very low expression of WSCP IIB is also observed even if these proteins could bind to Chlb, or have inhibitory functions towards Ser proteases like A07p11180.1. It should be noted that the ancestral proteins AnIIXA and AnIIB defined by Prabahar *et al.* (2020) would be Cys PI, a property lost in the brassica WSCP II targeted in this study.

Acquisition of PI function is not limited to WSCP IIX or WSCP IIB or to WSCP II without Chl binding potential. It has been demonstrated that AtWSCP the WSCP IIA in *A. thaliana*, inhibits Cys proteases as papain, proaleurain maturation protease and RD21 (responsive to desiccation 21 kDa protein), a papain-like cysteine protease (Halls *et al.*, 2006; Boex-Fontvieille *et al.*, 2015a; Rustgi *et al.*, 2017). BnD22 (C06p46460.1\_BnaDAR), the closest rapeseed protein to AtWSCP, and others WSCP IIA are predicted *in silico* as Cys PI, with potentially additional inhibitory activity against Ser proteases as trypsin or chymotrypsin types. However, whether or not WSCP IIA acts as Ser protease inhibitor remains inconclusive. No significant Ser protease inhibitory properties were detected *in vitro* for AtWSCP and BnD22 (Ilami *et al.*, 1997; Nishio & Satoh, 1997; Boex-Fontvieille *et al.*, 2016). In contrast, Desclos *et al.* (2008) reported that BnD22 inhibited trypsin activity in young leaves treated with methyl jasmonate. This Ser PI function could be provided not by BnD22, but by another protein oilseed rape WSCP IIA expressed in leaves. Indeed, the main WSCP II expressed in leaves are WSCP IIA and more specifically C06p46460.1 (BnD22), C06p46430.1 and A07p38410.1. Moreover,

C06p46460.1 and A07p38410.1 are mainly expressed in flowers, seeds or siliques at the reproductive stage.

WSCP<sub>II</sub> expressions followed on plants grown under control conditions highlighted a plasticity of expression. These profiles could be altered entirely by environmental constraints since it is assumed that WSCP<sub>II</sub> are involved in some stress response, likely related to their Chl-binding ability and protease inhibitors activity(ies). The induction of several WSCP<sub>II</sub> under various stresses suggests protective functions in young tissue in growth, as a Chl scavenger (Takahashi *et al.*, 2012, 2013) or ROS source in signaling pathway (Agostini *et al.*, 2017). WSCP<sub>II</sub> may also serve as Cys- and/ or Ser protease inhibitors to protect against biotic (Boex-Fontvieille *et al.*, 2015a) and abiotic stress (Downing *et al.*, 1992; Reviron *et al.*, 1992; Annamalai & Yanagihara, 1999; Desclos *et al.*, 2008) or to regulate endogenous protease activities (Avice and Etienne, 2007).

## CONCLUSION

In this study, a systematic analysis of four Brassica species identified a total of 68 WSCP<sub>II</sub>. Based on the phylogenetic analysis, the WSCP<sub>II</sub> family is classified into four groups IIA, IIB, IIC and IIX according to their ability to bind to Chl. A WSCP<sub>II</sub> sub-functional process is highlighted with some tissue-specific expression suggesting the role of WSCP<sub>IIA</sub> in organ development, the loss of the Chl binding ability for WSCP<sub>IIIX</sub> and potential additional PI properties. These results bring new elements in understanding of the evolution and function of WSCP<sub>II</sub> family genes in *B. napus* and its progenitors, providing valuable clues for further research.

## REFERENCES

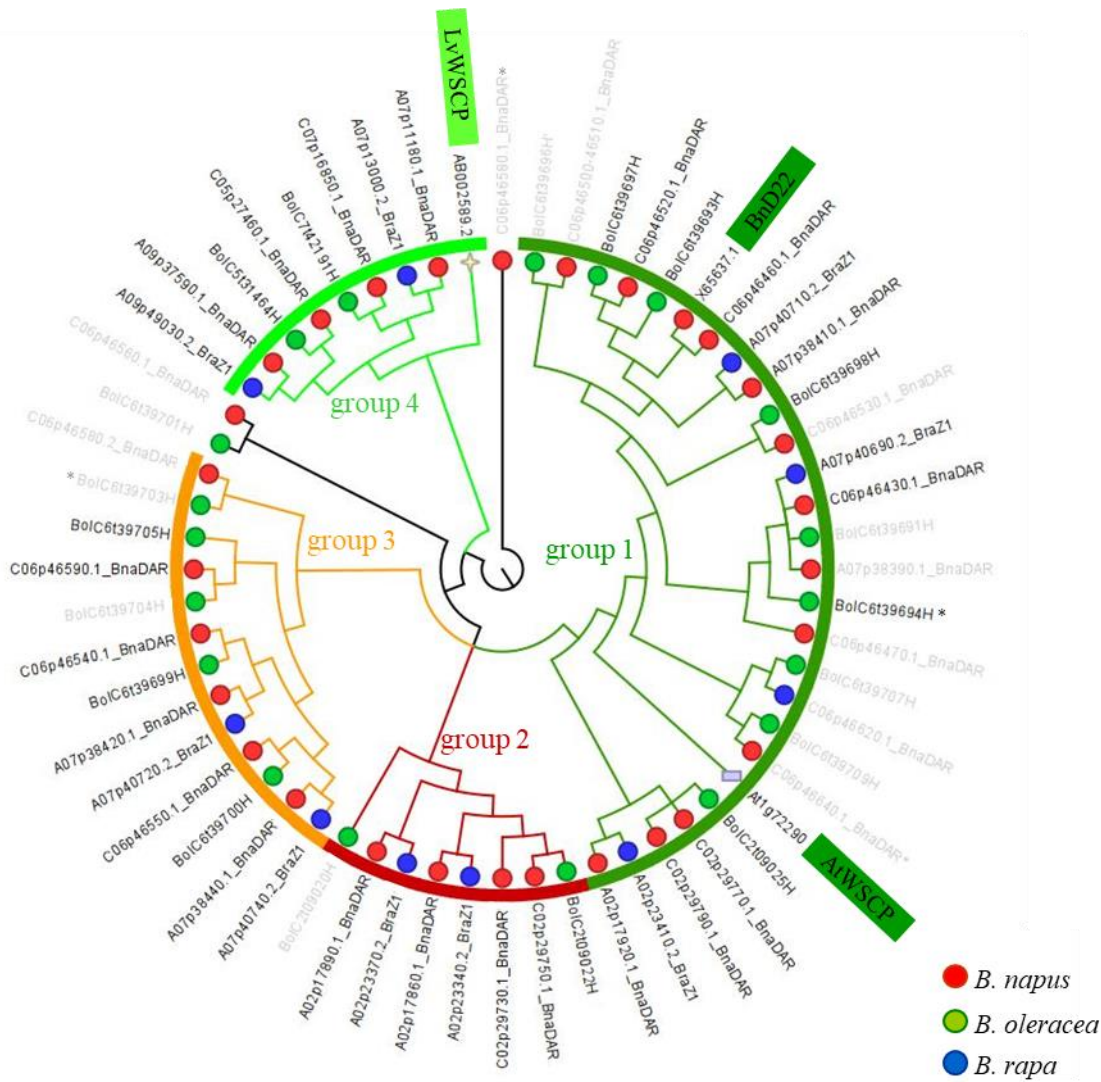
- Agostini A. et al.**, 2017. An unusual role for the phytyl chains in the photoprotection of the chlorophylls bound to water soluble chlorophyll binding proteins. *Sci. Rep.*, **7**: 1-13. <https://doi.org/10.1038/s41598-017-07874-6>.
- Agostini A. et al.**, 2019. How water-mediated hydrogen bonds affect chlorophyll a/b selectivity in water soluble chlorophyll protein. *Sci. Rep.*, **9**:18255. <https://doi.org/10.1038/s41598-019-54520-4>
- Annamalai and Yanagihara**, 1999. Identification and characterization of a heat-stress induced gene in cabbage encodes a Kunitz type protease inhibitor. *J. Plant Physiol.*, **155**: 226-233. [https://doi.org/10.1016/S0176-1617\(99\)80011-X](https://doi.org/10.1016/S0176-1617(99)80011-X).
- Bailey et al.**, 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**: W202–W8. <https://doi.org/10.1093/nar/gkp335>
- Bednarczyk D. et al.**, 2016. Fine tuning of chlorophyll spectra by protein-induced ring deformation. *Angewandte Chemie-International Edition*, **55**: 6901-6905. <https://doi.org/10.1002/anie.201512001>.
- Boex-Fontvieille E. et al.**, 2015a. A Kunitz-type protease inhibitor regulates programmed cell death during flower development in *Arabidopsis thaliana*. *J. Exp. Bot.*, **66**: 6119-6135. <https://doi.org/10.1093/jxb/erv327>.
- Boex-Fontvieille E. et al.**, 2015b. Water-soluble chlorophyll protein is involved in herbivore resistance activation during greening of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, **112**: 7303–7308. <https://doi.org/10.1073/pnas.1507714112>.
- Boex-Fontvieille E et al.**, 2016. An ethylene-protected achilles' heel of etiolated seedlings for arthropod deterrence. *Frontiers in Plant Science*, **7**: 1-14. <https://doi.org/10.3389/fpls.2016.01246>
- Belser C. et al.**, 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat Plants*, **4** : 879-887. <https://doi.org/10.1038/s41477-018-0289-4>.
- Chalhoub B. et al.**, 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**: 950–953. <https://doi.org/10.1126/science.1253435>.
- Cheng F. et al.**, 2018. Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants*, **4** : 258–268. <https://doi.org/10.1038/s41477-018-0136-7>.
- Dellero Y. et al.**, 2020. Leaf status and environmental signals jointly regulate proline metabolism in winter oilseed rape. *J Exp Bot.*, **71**: 2098–2111. <https://doi.org/10.1093/jxb/erz538>
- Desclos M. et al.**, 2008. A proteomic profiling approach to reveal a novel role of *Brassica napus* Drought 22 kD/ water soluble chlorophyll binding protein in young leaves during nitrogen remobilization induced by stressful conditions. *Plant Physiol.*, **147**: 1830-1844. <https://doi.org/10.1104/pp.108.116905>.
- Downing WL. et al.**, 1992. A *Brassica napus* transcript encoding a protein related to the Kunitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. *Plant J.*, **2**: 685-693. <https://doi.org/10.1046/j.1365-313X.1992.t01-11-00999.x>.
- Faës P. et al.**, 2015. Molecular evolution and transcriptional regulation of the oilseed rape proline dehydrogenase genes suggest distinct roles of proline catabolism during development. *Planta*, **241**:403–419. <https://DOI.10.1007/s00425-014-2189-9>.
- Green RR. and Kühlbrandt W.**, 1995. Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosynth Res.*, **44**: 139-48. <https://doi.org/10.1007/BF00018304>.
- Halls CE. et al.**, 2006. A Kunitz-type cysteine protease inhibitor from cauliflower and *Arabidopsis*. *Plant Sci.*, **170**: 1102-1110. <https://doi.org/10.1016/j.plantsci.2006.01.018>.
- Horigome D. et al.**, 2007. Structural mechanism and photoprotective function of water-soluble chlorophyll-binding protein. *J. Biol. Chem.*, **282**: 6525-6531. <https://doi.org/10.1074/jbc.M609458200>.

- Horton P. et al.**, 2007. WoLF PSORT: protein localization predictor. *Nucleic Acids Res.*, **35**(Web Server issue): W585–W587. <https://doi.org/10.1093/nar/gkm259>
- Hugues et al.**, 2006. Magneto-optic spectroscopy of a protein tetramer binding two exciton-coupled chlorophylls. *J. Am. Chem. Soc.*, **128**: 3649-3658. <https://doi.org/10.1021/ja056576b>.
- Ilami et al.**, 1997. Characterization of BnD22, a drought-induced protein expressed in *Brassica napus* leaves. *Phytochem.*, **45**: 1-8. [https://doi.org/10.1016/S0031-9422\(96\)00788-1](https://doi.org/10.1016/S0031-9422(96)00788-1).
- Istace B. et al.**, 2021. Sequencing and Chromosome-Scale Assembly of Plant Genomes, *Brassica rapa* as a Use Case. *Biology*, **10**, 732. <https://doi.org/10.3390/biology10080732>
- Krzywinski M. et al.**, 2009. Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**: 1639-45. <https://doi.org/10.1101/gr.092759.109>.
- Maleeva et al.**, 2019. Water soluble Chlorophyll-binding Proteins of plants: structure, properties and functions. *Mol. Biol.*, **53**, 876-888. <https://doi.org/10.1134/S0026893319060128>
- Murata T. et al.**, 1971. Water soluble chlorophyll protein of *Brassica oleracea* var. botrys (cauliflower). *Biochim. Biophys. Acta*, **245**: 208-215. [https://doi.org/10.1016/0005-2728\(71\)90023-5](https://doi.org/10.1016/0005-2728(71)90023-5).
- Nguyen TP. et al.**, 2019. Genome improvement and genetic map construction for *Aethionema arabicum*, the first divergent branch in the *Brassicaceae* family. *G3 (Bethesda)*, **9**: 3521-3530. <https://doi.org/10.1534/g3.119.400657>
- Nishio and Satoh**, 1997. A water-soluble chlorophyll protein in cauliflower may be identical to BnD22, a drought-induced, 22-kilodalton protein in rapeseed. *Plant Physiol.*, **115**: 841-846. <https://doi.org/10.1104/pp.115.2.841>.
- Palm DM. et al.**, 2018. Chlorophyll a/b binding-specificity in water soluble chlorophyll protein. *Nat. Plants*, **4**: 920-929. <https://doi.org/10.1038/s41477-018-0273-z>.
- Palm DM. et al.**, 2019. Stability of water soluble chlorophyll protein (WSCP) depends on phytol conformation. *ACS Omega*, **4**: 7971-7979. <https://doi.org/10.1021/acsomega.9b00054>.
- Prabakar V. et al.**, 2020. New homologues of Brassicaceae water-soluble chlorophyll proteins shed light on chlorophyll binding, spectral tuning, and molecular evolution. *FEBS J.*, **287**: 991-1004. <https://doi.org/10.1111/febs.15068>.
- Rawlings N. et al.**, 2013. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, **42**: D503–D509. <https://doi.org/10.1093/nar/gkt953>.
- Reviron MP. et al.**, 1992. Characterization of a novel protein induced by progressive or rapid drought and salinity in *Brassica napus* leaves. *Plant Physiol.*, **100**: 1486-1493. <https://doi.org/10.1104/pp.100.3.1486>.
- Rousseau-Gueutin M. et al.**, 2020. Long-reads assembly of the *Brassica napus* reference genome, Darmor-bzh. *Giga Science*, **9**: 1-16. <https://doi.org/10.1093/gigascience/giaa137>
- Rustgi S. et al.**, 2017. Serpin1 and WSCP differentially regulate the activity of the cysteine protease RD21 during plant development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA*, **114**: 2212-2217. <https://doi.org/10.1073/pnas.1621496114>
- Sampath P. et al.**, 2020. High contiguity long read assembly of *Brassica nigra* allows localization of active centromeres and provides insights into the ancestral Brassica genome. *BioRxiv*, in press. <https://doi.org/10.1101/2020.02.03.932665>
- Satoh H. et al.**, 1998. Molecular cloning and functional expression of a water soluble chlorophyll protein, a putative carrier of chlorophyll molecules in cauliflower. *J. Biol. Chem.*, **273**: 30568-30575. <https://doi.org/10.1074/jbc.273.46.30568>.
- Satoh H. et al.**, 2001. Water soluble chlorophyll protein in Brassicaceae plants is a stress-induced chlorophyll-binding protein. *Plant Cell Physiol.*, **42**: 906-911. <https://doi.org/10.1093/pcp/pce117>

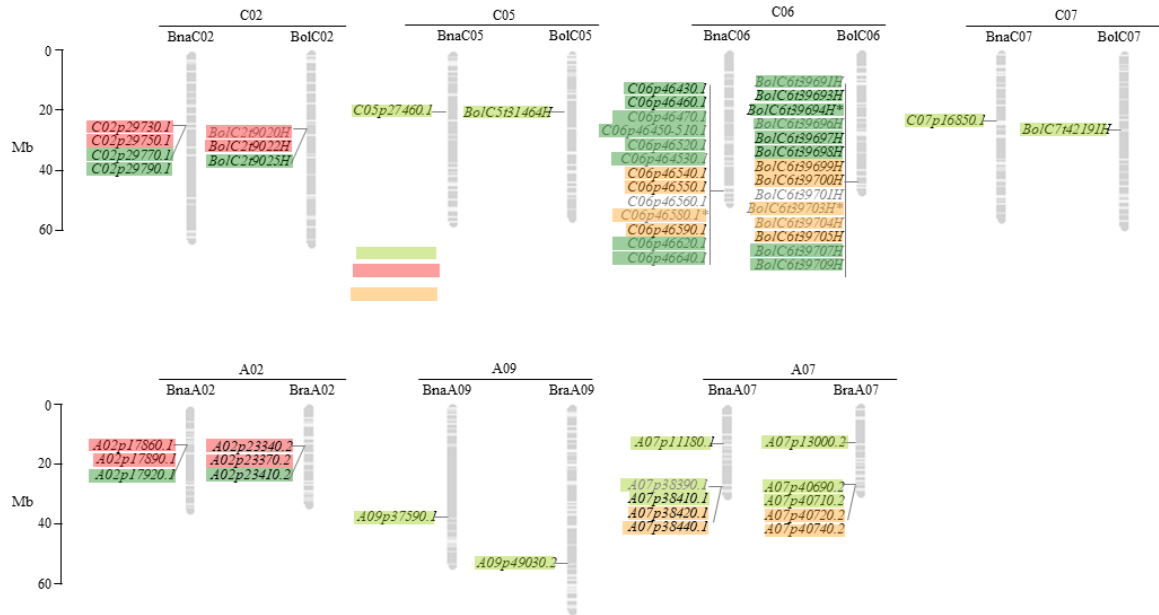
- Schmidt K. et al.**, 2003. Recombinant water soluble chlorophyll protein from *Brassica oleracea* var. botrys binds various chlorophyll derivatives. *Biochem.*, **42**: 7427-7433. <https://doi.org/10.1021/bi034207r>.
- Takahashi S. et al.**, 2012. Molecular cloning, characterization and analysis of the intracellular localization of a water soluble chlorophyll binding protein from brussels sprouts (*Brassica oleracea* var. gemmifera). *Plant Cell Physiol.*, **53**: 879-891. <https://doi.org/10.1093/pcp/pcs031>.
- Takahashi S. et al.**, 2013a. Molecular cloning and functional expression of a water soluble chlorophyll binding protein from Japanese wild radish. *J. Plant Physiol.*, **170**: 406-412. <https://doi.org/10.1016/j.jplph.2012.10.007>.
- Takahashi S. et al.**, 2013b. Molecular cloning, characterization and analysis of the intracellular localization of a water soluble chlorophyll- binding protein (WSCP) from Virginia pepperweed (*Lepidium virginicum*), a unique WSCP that preferentially binds chlorophyll b *in vitro*. *Planta*, **238**: 1065-1080. <https://doi.org/1065-1080.10.1007/s00425-013-1952-7>.
- Tamura K. et al.**, 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.*, **30**: 2725-2729. <https://doi.org/10.1093/molbev/mst197>.
- Vandesompele J. et al.**, 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**: RESEARCH0034. <https://doi.org/10.1186/gb-2002-3-7-research0034>.



FIGURES

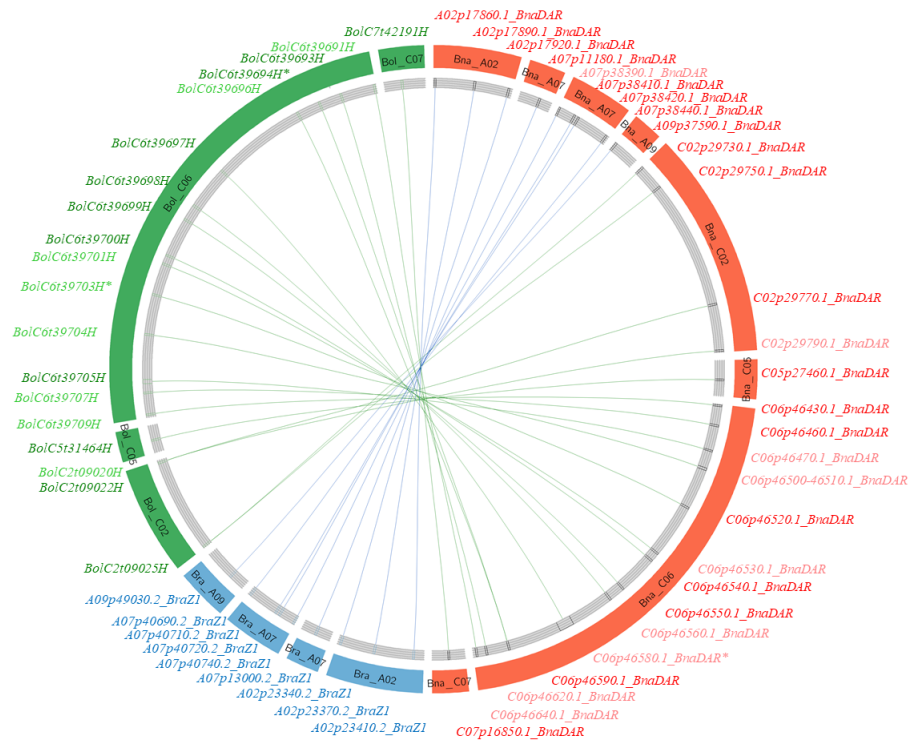


**Fig. 1: UPGMA tree of the WSCP II genes in *Brassica napus*, *B. oleracea* and *B. rapa*.** Sequence alignment and construction of the tree were performed using CLC software. Gene names are designed according automatic annotation. The pseudogenes are indicated in grey. Genes with intron were indicated as “\*”. AtWSCP, a WSCP IIA and LvWSCP, a WSCP IIB were positioned. C06p46460.1\_BnaDAR corresponds to BnD22, the only know WSCP from *B. napus* is indicated in green.



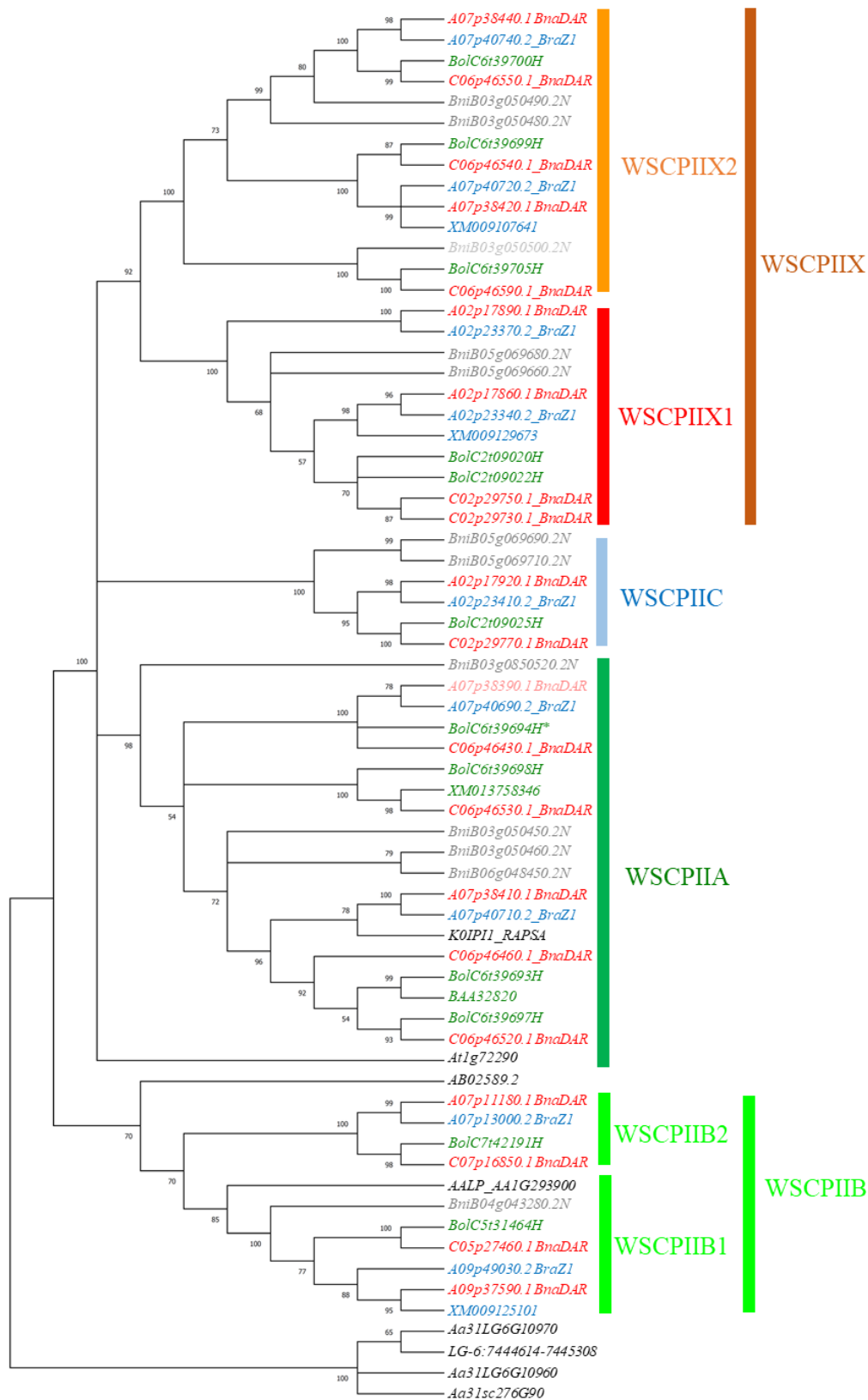
**Fig. 2. Chromosomal distribution of WSCP family genes among the *B. napus*, *B. oleracea* and *B. rapa* species.**

Genes from the same clade are represented by the same color (hard green for group 1-WSCPIIA, light green for group 4-WSCPIIB, red and orange for groups 2-3-WSCPIIX). The labels on the corresponding chromosomes indicate the name of the source organism and sub-genome. The scales indicate the sizes of the various Brassica chromosomes (Mb). The pseudogenes are indicated in grey. Genes with introns were designed as “\*”. Bna, *B. napus*; Bol, *B. oleracea*; Bra, *B. rapa*; Chr, chromosome.



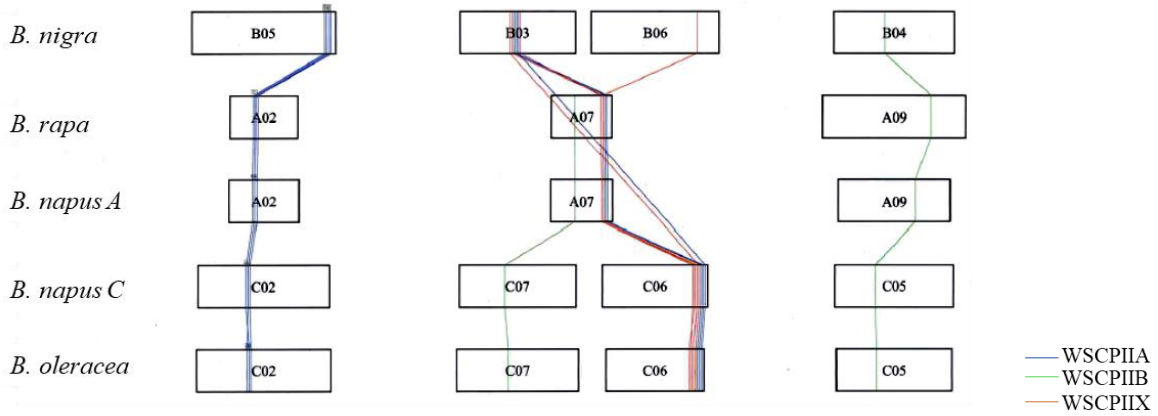
**Fig. 3: Genome-wide synteny analysis of all identified WSCP family genes among *B. napus*, *B. rapa*, and *B. oleracea*.**

The syntenic genes are linked with blue (A sub-genome) and green lines (C sub-genome) respectively. Genes with intron as “\*”. Pseudogenes are indicated in lighter colour. BnaDAR: *B. napus* cv Darmor nain, BraZ1: *B. rapa* cv Z1, Bol: *B. oleracea* cv HDEM.



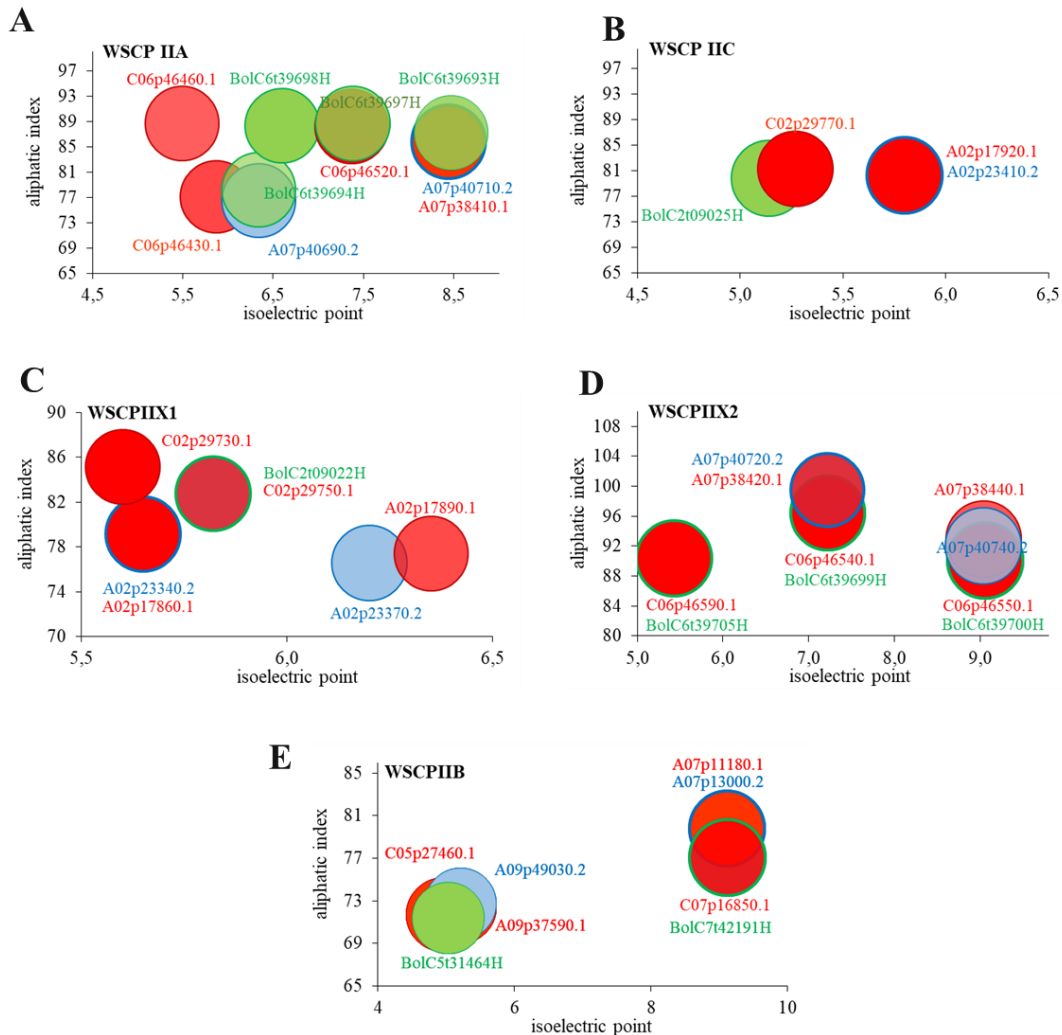
**Fig. 4: Phylogenetic tree of WSCPII.**

A set of 67 sequences of WSCPII from *A. arabicum* (*Aa*), *A. thaliana* (*At*), *Arabis alpina* (*Aa*), *B. napus* cv Darmor nain (*BnaDAR*), *B. nigra* cv Ni100 (*Bni*), *B. oleracea* cv HDEM (*Bol*), *B. rapa* cv Z1 (*BraZ1*), *L. virginicum* (*Lv*), *Raphanus sativus* (*RAPSA*) were used with WSCPII from *A. arabicum* as an outgroup. Sequence alignments, models tests and phylogenetic analysis were performed with MEGA 11.0.11 software (Tamura *et al.*, 2013). Pseudogenes are indicated in lighter colour. \* indicates genes with intron.



**Fig. 5 : Orthologous relationships between WSCP II from *B. napus*, *B. nigra*, *B. oleracea* and *B. rapa*.**

Colours of lines represent phylogenetic groups. In chromosomes B05, C02 and A02, WSCP IIC are includes in blue lines.



**Fig. 6: Characteristics of WSCP II proteins from *B. napus*, *B. oleracea* and *B. rapa***

According to amino acid sequence of mature proteins without peptide signal, molecular weight, isoelectric point and aliphatic index are determined. The sizes of circles are proportional to protein molecular weights. **A.** WSCP IIA, **B.** WSCP IIC, **C.** WSCP IIX1, **D.** WSCP IIX2, **E.** WSCP IIB

**A.**

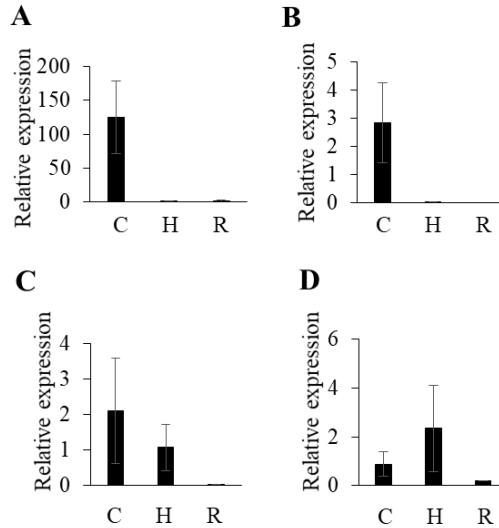
protein name	PI activities							Chl binding					
	Ser protease				Cys protease			Horigone <i>et al.</i> (2007); Palm <i>et al.</i> (2018); Prabahar <i>et al.</i> (2020)					
	reactive-site	% of cleavage probability (P1 residue and position in reactive site)		Amino acids defined in interaction AtWSCP- papain like proteases (Boex-Fontvieille <i>et al.</i> , 2015; Rutsigi <i>et al.</i> , 2017)					AA in interaction with chlorophyll				chl a binding specificity
		trypsin	chymotrypsin	RD21	PMP	XBCP3	CBD1	CBD1-2	CBD3	CBD4	CBD3		
<b>AnIXA</b>	YPPSIMGPN	0	36 (M6)	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAAIN	L T L Q	S I W W	PVCNEF			
<b>AnIIA</b>	YPPSIVGTN	0	5	<u>L P W P N</u>	<u>A A L F N</u>	<u>A V</u>	VPAAIN	L T L Q	S I W W	PVCNEF			
WSCPIIA	<b>Atlg72290 (AtWSCP)</b>	0	0	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAAIT	L T L Q	S I W W	PFCNEF			
	C06p46460.1_BnaDAR BoIC63963H BoIC63969H	PVIVGTD	2 (P1)	0	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAAIN	L T L Q	S I W W	PVCNEL		
	C06p46520.1_BnaDAR A07p38410.1_BnaDAR A07p40710.2_BraZ1	PPVVGTD	73 (E3)	0	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAAIK	L T L Q	S I W W	PVCNEF		
	C06p46430.1_BnaDAR BoIC63694H	YTETIDRD	99 (Y1)	64 (R7)	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAATN VPAATK VPAATN	L T L Q	S I W W	PVCSEF		
	A07p40690.2_BraZ1 BoIC63698H	YHPNILGR	70 (Y1)	0	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAATN	L T L Q	S I W W	PVCNEF		
WSCPIIC	C02p29770.1_BnaDAR BoIC210925H	SDSVTEP	4 (T5)	0	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAAIR	L T L Q	S I W W	PVCKEF		
	A02p17920.1_BnaDAR A02p23410.2_BraZ1	SDSVTEP	4 (T5)	0	<u>L P W P N</u>	<u>A A L F N</u>	<u>A F</u>	VPAAIR	L T L Q	S I W W	PVCKEY		
WSCPIIX1	<b>AnIX</b>	YPPSLIGPN	0	51 (L)	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	VPVAVK	L T L Q	S I W W	PVCKEF		
	C02p29730.1_BnaDAR C02p29750.1_BnaDAR BoIC2109022H	YYPYLMDT	53 (L5)	0	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	LPVPVK	L T L Q	S I W W	PGCEEFF		
	A02p17860.1_BnaDAR A02p23340.2_BraZ1 A02p17890.1_BnaDAR A02p23370.2_BraZ1	PYALMDT	92 (Y2)	0	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	LPVPVK	L T L Q	S I W W	PGCEEFF		
WSCPIIX2	C06p46540.1_BnaDAR BoIC63969H	PPLLPRP	31 (L3)	100 (R6)	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	IPVAAI	L T L Q	S I W W	XXCNGI		
	A07p38420.1_BnaDAR A07p40720.2_BraZ1	SLIPPIP	51 (L2)	0	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	VPLAAK	L T L Q	S I W W	XVCKDI		
	A07p38440.1_BnaDAR C06p46550.1_BnaDAR BoIC639700H	PPLPESP	72 (F5)	0	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	VPLAAK	L T L Q	S I W W	XVCKDI		
	C06p46590.1_BnaDAR BoIC639705H	YPPGFIT	4 (Y1)	0	<u>L P W P K</u>	<u>V A V F N</u>	<u>V F</u>	APLAAI	L T L Q	S I W W	XXCKEL		

**B**

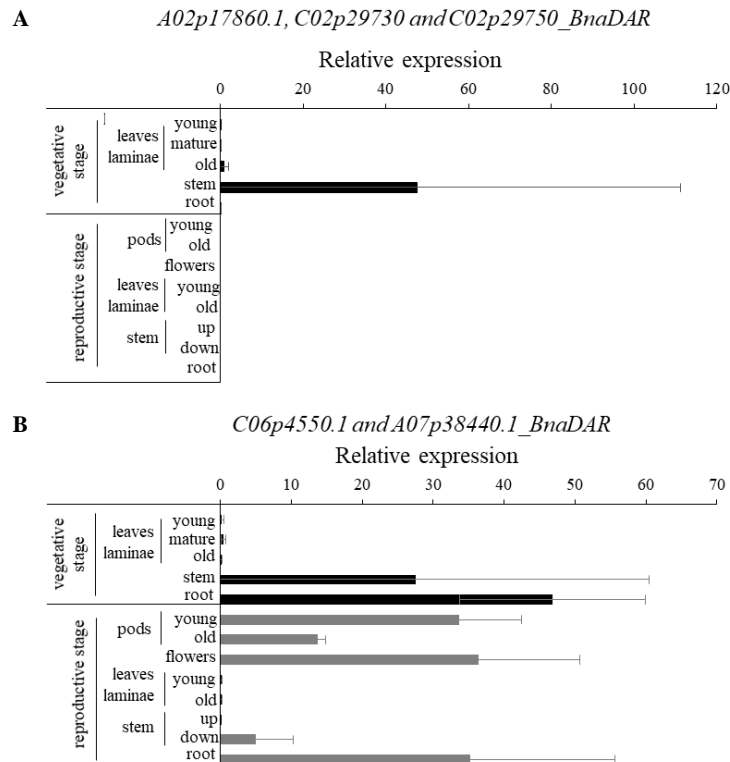
protein name	PI activities							Chl binding					
	Ser protease				Cys protease			Horigone <i>et al.</i> (2007); Palm <i>et al.</i> (2018); Prabahar <i>et al.</i> (2020)					
	reactive-site	% of cleavage probability (P1 residue and position in reactive site)		Amino acids defined in interaction AtWSCP- papain like proteases (Boex-Fontvieille <i>et al.</i> , 2015; Rutsigi <i>et al.</i> , 2017)					AA in interaction with chlorophyll				binding specificity
		trypsin	chymotrypsin	RD21	PMP	XBCP3	CBD1	CBD1-2	CBD3	CBD4	CBD3	chla	chlb
<b>AnIXA</b>	SPSSSINGN	0	5	<u>L P W L P A N S L P</u>	<u>S S L P</u>	<u>P S</u>	VPAS	L L T L L Q L S	I W W			LCPS	
<b>AnIIB</b>	TPSSSNG	5 (S3)	0	<u>L P W L P A S S L P</u>	<u>S S L P</u>	<u>S S</u>	VPAS	L L T S L Q L S	I W W			LCPS	
WSCPIIB	<b>LvWSCP (AB002589.2)</b>	PSSSEGN	5 (S2)	0	<u>S H L P N S L P</u>	<u>N S L P</u>	VPAN	L L T S L Q L S	I W W				
	C07p16850.1_BnaDAR BoIC0742191H	TPFFFDIG	60 (F3)	0	<u>S H L P N S L P</u>	<u>N S L P</u>	VPAN	L L T S L Q L S	I W W				
	A07p11180.1_BnaDAR A07p13000.2_BraZ1	TPFFDRGN	74 (F4)	100 (R6)	<u>S H L P N S L P</u>	<u>N S L P</u>	VPAN	L L T S L Q L S	I W W			CPS	
	C05p27460.1_BnaDAR BoIC05131464H A09p37590.1_BnaDAR A09p49030.2_BraZ1	NPSSQNDD	6 (Q5)	0	<u>S H L P N S L P</u>	<u>N S L P</u>	VPAN	L L T S L Q L S	I W W				LCPS

**Fig. 7. In silico properties of WSCP II**

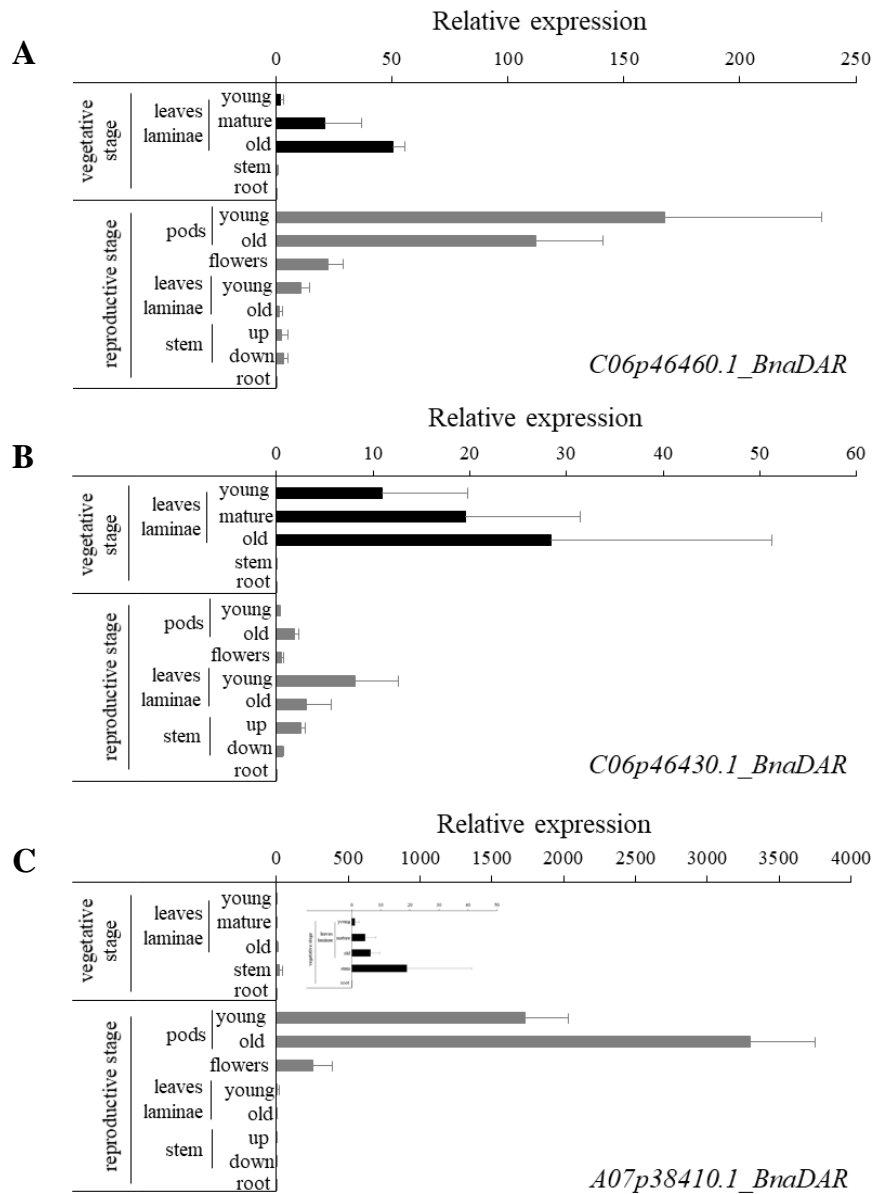
The reactive-site for Ser anti proteolytic activity, located in the IV loop, is determined with the Conserved Domain Database software (<https://www.ncbi.nlm.nih.gov/cdd>). Activity and location of the highest potential P1 sites for trypsin and chymotrypsin were obtained through PeptideCutter (<http://www.expasy.org/tools/peptidecutter/>). The Cys PI activity was estimated according amino acids at the same positions as those defined by for interaction between AtWSCP and papaine like Cys proteases (RD21, PMP or XBCP3) are underlined in red in loop II (Ala<sub>59</sub> Ala<sub>60</sub> Leu<sub>64</sub> Pro<sub>65</sub> Phe<sub>66</sub>) and loop V (Asn<sub>109</sub> Trp<sub>111</sub> Pro<sub>112</sub> Phe<sub>113</sub> Lys<sub>115</sub>) (Boex-Fontvielle *et al.*, 2015; Rutsigi *et al.*, 2017). The Chl binding was estimated by the presence of amino acids defined to be in contact with chlorophyll in the homo-tetrameric WSCP-Chl complex are in green background (Horigome *et al.*, 2007; Palm *et al.*, 2018; Prabahar *et al.*, 2020) and located in the Chl domain binding domains (CBD) defined by Prabahar *et al.* (2020). Chl binding specificity was determined by comparison of the amino acids present in the motifs defined by Palm *et al.* (2018). In addition, referent proteins AtWSCP, LvWSVP and ancestral proteins defined by Prabahar *et al.* (2020) are included.



**Fig. 8: Expression patterns of *B. napus* WSCP II in 4-DAS (4-days after sowing) seedling.** Quantitative RT-PCR analyses of *C06p46460.1\_BnaDAR* (A), *C06p46430.1\_BnaDAR* (B), *C05p27460.1\_BnaDAR* (C) and *C07p16850.1\_BnaDAR* (D) expressions in cotyledon (C), hypocotyl (H) and root (R) of oilseed rape seedlings at 4 days after sowing.



**Fig. 9: Expression patterns of *B. napus* WSCP II X.** Quantitative RT-PCR analyses performed at vegetative and reproductive stages in different tissues and organs. Several copies were amplified using qPCR primers designed. Examples of expression patterns of three WSCP II X1 (A) and for the WSCP II X2 genes, the homeologous *C06p44540.1-A07p38420.1\_BnaDAR* (B).



**Fig. 10: Expression patterns of *B. napus* WSCP II A.**

Quantitative RT-PCR analyses were performed at vegetative and reproductive stages in different tissues and organs. Examples of expression patterns *C06p44660.1* (A), *C06p44430.1* (B), *A07p38410.1* (C).







## Supplementary data

### Fig. S1: WSCP II genomic sequences of *B. napus*, *B. rapa* and *B. oleracea*

The WSCP II genomic sequences from *B. napus* cv. Darmor-bzh (Rousseau-Gueutin *et al.*, 2020), *B. rapa* cv. Z1 (Istace *et al.*, 2021) and *B. oleracea* cv HDEM (Belser *et al.*, 2018) were listed. The names of genes were assigned according to the automatic annotation and to the Ostergaard and King (2008) nomenclature, taking into account A or C genome location, the closest *BnaC.WSCP IIA.a* (*C06p46460.1\_BnaDAR* identical to *BnD22*) sequence homology and their membership in phylogenetic sub-groups.

Nucleotides in bold and underlined correspond to start and stop codons. Nucleotides in italic correspond to intron region (gene designed with \*). The pseudogenes are indicated in grey. Genes with modification according the automatic annotation were designed “\*\*\*”. Chr = chromosome.

<https://filesender.renater.fr/?s=download&token=8c695998-289a-4196-a920-da6720d735bf>

### Fig. S2. Additional WSCPS used for the phylogenic analysis

Additional WSCP II genomic sequences from *B. nigra* cv. Ni100 (Sampath *et al.*, 2020), *Aethionema arabicum* (Nguyen *et al.*, 2019) and some from Prabahar *et al.* (2020) are listed. Nucleotides in bold and underlined correspond to start and stop codons. The pseudogenes are indicated in grey. Genes with modification (sequence length, ATG or stop codon positions, taking into account or not intron related region) according the automatic annotation were designed as “\*\*\*”. Chr = chromosome. Protein sequences are given without peptide sequence.

<https://filesender.renater.fr/?s=download&token=8c695998-289a-4196-a920-da6720d735bf>

### Fig. S3: WSCP II protein sequences from *B. napus*, *B. oleracea* and *B. rapa*.

Amino acids (AA) in italic correspond to the peptide signal defined with SignalP 4.0 software (<http://www.cbs.dtu.dk/services/SignalP/>).

- For PI activities:

AA in boxed red constituted the Kunitz consensus sequence “[LIVM]-x-D-[EK]-[EDNTY]-[DG]-[RKH DENQ]-x-[LIVM]-x(5)-Y-x-[LIVM]” (PROSITE ID: PS00283). The reactive-site for Ser anti proteolytic activity, located in the IV loop, is determined with the Conserved Domain Database software (<https://www.ncbi.nlm.nih.gov/cdd>) and shown in red. Amino acids at the same positions as those defined by for interaction between AtWSCP and papaine like Cys proteases (RD21, PMP or XBCP3) are underlined in red in loop II (Ala<sub>59</sub> Ala<sub>60</sub> Leu<sub>64</sub> Pro<sub>65</sub> Phe<sub>66</sub>) and loop V (Asn<sub>109</sub> Trp<sub>111</sub> Pro<sub>112</sub> Phe<sub>113</sub> Lys<sub>115</sub>) (Boex-Fontvielle *et al.*, 2015; Rutsgi *et al.*, 2017).

- For chlorophyll (Chl) binding:

AA in boxed green correspond to the WSCP signature chlorophyll binding motif “PFCPLGI” as defined by Satoh *et al.* (1998, 2001). AA defined to be in contact with chlorophyll in the homo-tetrameric WSCP-Chl complex are in green background (Horigome *et al.*, 2007; Palm *et al.*, 2018; Prabahar *et al.*, 2020). AA important for Chl selectivity are in green italic (Palm *et al.*, 2018). Locations of the Chl domain binding domains (CBD) defined by Prabahar *et al.* (2020) are underlined in green. Inside those CBDs, AA in green correspond to the consensus sequence.

In addition, referent proteins AtWSCP, LvWSVP, BolWSCPs and CaWSP are included.

<https://filesender.renater.fr/?s=download&token=8c695998-289a-4196-a920-da6720d735bf>

**Table S1: Real-time qRT-PCR primers, melting temperature and amplicon sizes.**

Gene name	Forward primer	Reverse primer	T <sub>m</sub> (°C)	Fragment size (pb)
<i>Bna.PTB</i>	AGCGAAAGAAGCATTGGAAG	CACGAGTAGGCTCAGATCAGG	60	148
<i>Bna.PP2A</i>	GCTTCAAATCCCTCATTCCAA	ATCCTCGCTTAGCTCCACAA	60	82
<i>RibS3</i>	CCTGTGCGCCAGAAGAAG	AGAGCCGCTGATAAGGTTTG	60	
<i>UBQ11</i>	GTTGATCTTCGCTGGAAAAC	CCATTAAAGACGGCTCGATG	60	
<i>C06p46460.1_BnaDAR</i>	ATAGAAGAAGCTACAGAAGCAA	CTTAGTAGCAGCATCATCAACT	60	165
<i>A07p38410.1_BnaDAR</i>	TTCGGATATTATCCACCTTTC	ATGTCGGACTCGAACTTGAC	61	80
<i>C06p46520.1_BnaDAR</i>	TCCATAAACATCGAGTTTAG	GTAGTCTCTCATCTTAAACAG	56	161
<i>A02p17920.1_BnaDAR</i>	ACTCCGTCACAGAAGCCAT	GAGGATAATTGCTGGCTCAT	62	146
<i>C02p29770.1_BnaDAR</i>	CGTCACGGGGATTTTCT	TCTCAACACGAGAAGCCGAC	66	116
<i>C06p46530.1_BnaDAR</i>	CGCATATCATCCAAACATTC	AGGCCAGATGTTGGACAC	61	86
<i>C06p46430.1_BnaDAR</i>	GGGAAAAATACTTCATGCTG	ACAAGACCACCTCCATCG	60	58
<i>C02p29750.1_BnaDAR/C02p29730.1_BnaDAR</i>	TTTAAAAATTGAGAGAAAAGAACC	GTGGTGCATTTTCCCAAG	61	106
<i>A02p17890.1_BnaDAR</i>	GAAAGTCGATGAATCTCATCT	AACTAGTTGTGGGGCTTTG	60	192
<i>A02p17860.1_BnaDAR</i>	CCAGGGACACGAACCTGC	GGTGTCCATGAGCGGGTAC	66	220
<i>C02p29730.1_BnaDAR</i>	GGAGCCTGCTATTCTGACC	ACTTCTCTGACGTGAAGTAGAT	62	248
<i>A07p38440.1_BnaDAR</i>	GGAGGTGGTCTTATCCCAG	CCCAGAACTTGGAGATGC	61	211
<i>C06p46550.1_BnaDAR</i>				
<i>C06p46590.1_BnaDAR</i>	TCCAACGACTGTAAGGAGC	TTGTGGTGTAGTCCCAATAG	60	213
<i>A07p38440.1_BnaDAR</i>				
<i>C06p46550.1_BnaDAR</i>	GAGAAAAGTGGGAGAAGAAGG	AGTAACATCCGTAGTAACAGCA	60	174
<i>A07p38440.1_BnaDAR</i>				

**Table S2: Basic information concerning WSCP II in *B. napus*, *B. rapa* and *B. oleracea* genomes and protein characteristics.**

Gene locations and presence of introns were determined according automatic annotation. \*gene with intron(s). The pseudo genes are indicated in grey. The peptide signal was predicted with SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP/>). The characteristics of proteins were determined for proteins without peptide signal by CLC Main Workbench Version 22.0 (QIAGEN). Subcellular locations were predicted with PSORT software (Horton *et al.*, 2007). aa: amino acids; bp: pair base; extr: extracellular; nucl: nuclear; plat: chloroplast; vacu: vacuolar.

phylogenetic group	gene										protein					
	automatic annotation	gene name	Chr	start	end	exons	length (bp)	peptide signal cleavage site	protein (aa)	MW (kDa)	pI	aliphatic index	subcellular location (kNN score/14)			
sub-genome A	A02p17860.1_BnaDAR	<i>BnaA.WSCPPIX1.a1</i>	A02	12 359 254	12 359 934	1	681	24 <sup>∧</sup> 25: TQG-HE	202	22.14	5.65	79.16	extr: 8			
	A02p17890.1_BnaDAR	<i>BnaA.WSCPPIX1.a2</i>	A02	12 373 799	12 374 470	1	672	24 <sup>∧</sup> 25: GE-NR	199	22.16	6.35	77.34	nucl: 7			
	A02p17920.1_BnaDAR	<i>BnaA.WSCPPII.a</i>		12 386 406	12 387 059	1	654	19 <sup>∧</sup> 20: THG-QE	198	21.07	5.8	80.25	extr: 6			
	A07p11180.1_BnaDAR	<i>BnaA.WSCPPII.b</i>		11 557 089	11 557 808	1	720	23 <sup>∧</sup> 24: AAA-VT	216	24.26	9.12	79.81	vacu: 5			
	A07p38390.1_BnaDAR	<i>BnaA.WSCPPII.c1</i>		26 223 010	26 223 636	1	627	-	-	-	-	-	-			
	A07p38410.1_BnaDAR	<i>BnaA.WSCPPII.a</i>	A07	26 229 083	26 229 754	1	672	23 <sup>∧</sup> 24: THG-KE	200	21.51	8.44	85.8	extr: 6			
	A07p38420.1_BnaDAR	<i>BnaA.WSCPPIX2.b</i>		26 230 686	26 231 327	1	642	23 <sup>∧</sup> 24: THG-QV	190	20.51	7.22	99.47	extr: 8			
	A07p38440.1_BnaDAR	<i>BnaA.WSCPPIX2.a</i>		26 242 660	26 243 325	1	666	23 <sup>∧</sup> 24: THG-QE	198	21.63	9.04	92.93	plas: 5			
	A09p37590.1_BnaDAR	<i>BnaA.WSCPPII.b</i>	A09	36 930 217	36 930 885	1	669	23 <sup>∧</sup> 24: VSA-GD	199	21.83	5.21	72.36	extr: 4			
	C02p29730.1_BnaDAR	<i>BnaC.WSCPPIX1.a1</i>		23 593 059	23 593 729	1	671	11 <sup>∧</sup> 12: GH-EH	198	21.95	5.6	85.15	cyto: 6			
	C02p29750.1_BnaDAR	<i>BnaC.WSCPPIX1.a2</i>	C02	23 602 457	23 603 128	1	672	24 <sup>∧</sup> 25: TQG-HE	199	22.04	5.82	82.76	extr: 9			
	C02p29770.1_BnaDAR	<i>BnaC.WSCPPII.a1</i>		23 650 032	23 650 685	1	654	19 <sup>∧</sup> 20: THG-QE	198	21.15	5.27	81.26	extr: 7			
	C02p29790.1_BnaDAR	<i>BnaC.WSCPPII.a2</i>		23 667 139	23 667 791	1	653	-	-	-	-	-	-			
	C05p27460.1_BnaDAR	<i>BnaC.WSCPPII.b</i>	C05	19 204 465	19 203 767	1	699	23 <sup>∧</sup> 24: VSA-ED	209	22.91	4.95	71.72	extr: 5			
C06p46430.1_BnaDAR	<i>BnaC.WSCPPII.a1</i>		45 974 190	45 974 816	1	627	21 <sup>∧</sup> 22: VCA-HG	187	20.57	5.87	77.06	extr: 5				
C06p46460.1_BnaDAR	<i>BnaC.WSCPPII.a1</i>		45 981 557	45 982 213	1	657	19 <sup>∧</sup> 20: THG-RE	199	21.51	5.49	88.69	extr: 6				
C06p46470.1_BnaDAR	<i>BnaC.WSCPPII.a2</i>		45 990 950	45 991 557	1	608	-	-	-	-	-	-				
C06p46500.1_BnaDAR	<i>BnaC.WSCPPII.b1</i>		45 997 765	45 998 672	1	908	-	-	-	-	-	-				
C06p46520.1_BnaDAR	<i>BnaC.WSCPPII.a2</i>		46 013 390	46 014 058	1	669	23 <sup>∧</sup> 24: THG-KE	199	21.34	7.37	88.19	extr: 5				
C06p46530.1_BnaDAR	<i>BnaC.WSCPPII.b2</i>		46 033 377	46 034 046	1	670	-	-	-	-	-	-				
C06p46540.1_BnaDAR	<i>BnaC.WSCPPIX2.b</i>	C06	46 036 743	46 037 411	1	669	23 <sup>∧</sup> 24: THG-QV	190	20.56	7.22	96.42	extr: 9				
C06p46550.1_BnaDAR	<i>BnaC.WSCPPIX2.a</i>		46 051 286	46 051 954	1	669	23 <sup>∧</sup> 24: THG-QE	199	21.68	9.06	90.05	plas: 4				
C06p46560.1_BnaDAR	<i>BnaC.WSCPPII.a</i>		46 057 219	46 057 964	1	746	-	-	-	-	-	-				
C06p46580.1_BnaDAR*	<i>BnaC.WSCPPIX2.d</i>		46 073 630	46 079 045	2	5 416	-	-	-	-	-	-				
C06p46590.1_BnaDAR	<i>BnaC.WSCPPIX2.c</i>		46 098 439	46 099 095	1	657	24 <sup>∧</sup> 25: THG-EE	194	21.44	5.43	90.36	extr: 7				
C06p46620.1_BnaDAR	<i>BnaC.WSCPPII.d1</i>		46 106 725	46 107 390	1	666	-	-	-	-	-	-				
C06p46640.1_BnaDAR	<i>BnaC.WSCPPII.d2</i>		46 110 344	46 111 032	1	689	-	-	-	-	-	-				
C07p16850.1_BnaDAR	<i>BnaC.WSCPPII.a</i>	C07	21 735 562	21 736 281	1	720	23 <sup>∧</sup> 24: AAA-VT	216	24.23	9.12	77.08	extr: 7				

*B. napus*

phylogenetic group	gene					protein							
	automatic annotation	gene name	Chr	start	end	exons	length (bp)	peptide signal cleavage site	protein length (aa)	MW (kDa)	pI	aliphatic index	subcellular location (kNN score/14)
<i>B. rapa</i>	A02p23340.2_BraZI	<i>Bra.WSCP1IX1.a1</i>		12 273 429	12 274 109	1	681	24*25: TQG-HE	202	22.14	5.65	79.16	extr: 8
	A02p23370.2_BraZI	<i>Bra.WSCP1IX1.a2</i>	A02	12 287 984	12 288 658	1	675	19*20: ASA-VC	205	22.79	6.20	76.54	extr: 6
	A02p23410.2_BraZI	<i>Bra.WSCP1IA.e</i>		12 300 589	12 301 242	1	654	19*20: THG-QE	198	21.07	5.80	80.25	extr: 6
	A07p13000.2_BraZI	<i>Bra.WSCP1IB.a</i>		11 065 153	11 065 872	1	720	23*24: AAA-VT	216	24.23	9.11	79.81	vacu: 5
	A07p40690.2_BraZI	<i>Bra.WSCP1IA.c1</i>		25 436 076	25 436 714	1	639	21*22: VCA-HG	191	20.96	6.34	76.49	extr: 5
	A07p40710.2_BraZI	<i>Bra.WSCP1IA.a</i>	A07	25 442 043	25 442 714	1	672	23*24: THG-KE	200	21.51	8.44	85.80	extr: 6
	A07p40720.2_BraZI	<i>BraA.WSCP1IX2.b</i>		25 443 638	25 444 279	1	642	23*24: THG-QV	190	20.51	7.22	99.47	extr: 8
	A07p40740.2_BraZI	<i>Bra.WSCP1IX2.a</i>		25 454 861	25 455 526	1	666	23*24: THG-QE	198	21.75	9.04	92.03	plas: 4
	A09p49030.2_BraZI	<i>Bra.WSCP1IB.b</i>	A09	51 763 642	51 764 310	1	669	24*25: SAG-DE	198	21.81	5.21	72.73	vacu: 4
		<i>BolC2i09020H</i>	<i>Bol.WSCP1IX1.a2</i>		24 843 202	24 843 805	1	604	-	-	-	-	-
<i>B. oleracea</i>	<i>BolC2i09022H</i>	<i>Bol.WSCP1IX1.a1</i>	C2	24 850 153	24 850 821	1	669	23*24: TQG-HE	199	22.04	5.82	82.76	extr: 9
	<i>BolC2i09025H</i>	<i>Bol.WSCP1IA.e</i>		24 884 722	24 885 375	1	654	19*20: THG-QE	198	21.18	5.14	79.79	extr: 7
	<i>BolC5i31464H</i>	<i>Bol.WSCP1IB.b</i>	C5	19 094 879	19 094 211	1	669	23*24: VSA-ED	199	21.85	5.03	71.41	extr: 5
	<i>BolC6i39691H</i>	<i>Bol.WSCP1IA.c1</i>		42 354 690	42 355 270	1	581	-	-	-	-	-	-
	<i>BolC6i39693H</i>	<i>Bol.WSCP1IA.a1</i>		42 362 605	42 363 261	1	657	19*20: THG-RE	199	21.47	8.47	87.23	extr: 6
	<i>BolC6i39694H*</i>	<i>Bol.WSCP1IA.c2</i>		42 366 294	42 367 863	2	1 570	25*26: GTA-EI	193	21.41	6.34	78.19	extr: 7
	<i>BolC6i39696H</i>	<i>Bol.WSCP1IA.b1</i>		42 383 994	42 384 642	1	649	-	-	-	-	-	-
	<i>BolC6i39697H</i>	<i>Bol.WSCP1IA.a2</i>		42 398 714	42 399 382	1	669	23*24: THG-KE	199	21.39	7.38	88.69	extr: 5
	<i>BolC6i39698H</i>	<i>Bol.WSCP1IA.b2</i>		42 410 459	42 411 127	1	669	20*21: THG-LE	202	21.83	6.6	88.32	extr: 9
	<i>BolC6i39699H</i>	<i>Bol.WSCP1IX2.b</i>	C6	42 413 833	42 414 474	1	642	23*24: THG-QV	190	20.56	7.22	96.42	extr: 9
	<i>BolC6i39700H</i>	<i>Bol.WSCP1IX2.a</i>		42 428 392	42 429 060	1	669	23*24: THG-QE	199	21.68	9.06	90.05	plas: 4
	<i>BolC6i39701H</i>	<i>Bol.WSCP1I.a</i>		42 434 322	42 435 067	1	746	-	-	-	-	-	-
	<i>BolC6i39703H*</i>	<i>BolC.WSCP1IX2.d</i>		42 450 680	42 451 282	2	603	-	-	-	-	-	-
	<i>BolC6i39704H</i>	<i>Bol.WSCP1IX2.c1</i>		42 485 428	42 486 085	1	658	-	-	-	-	-	-
	<i>BolC6i39705H</i>	<i>Bol.WSCP1IX2.c2</i>		42 494 124	42 494 780	1	657	24*25: THG-EE	194	21.44	5.43	90.36	extr: 7
		<i>BolC6i39707H</i>	<i>Bol.WSCP1IA.d1</i>		42 502 536	42 503 181	1	646	-	-	-	-	-
		<i>BolC6i39709H</i>	<i>Bol.WSCP1IA.d2</i>		42 515 095	42 515 743	1	649	-	-	-	-	-
		<i>BolC7i42191H</i>	<i>Bol.WSCP1IB.a</i>	C7	24 711 202	24 711 921	1	720	23*24: AAA-VT	216	24.29	9.12	77.08

**Table S3: WSCP II genes number according to their chromosomic location and group membership.**

Pseudogenes numbers are indicated in grey. Chr, chromosome

**A.**

		Chr	<i>B. napus</i> sub-genome C	<i>B. oleracea</i>	Chr	<i>B. napus</i> sub-genome A	<i>B. rapa</i>
WSCPIIA		C06	3 (5)	4 (4)	A07	1 (1)	2
WSCPIIC		C02	1 (1)	1	A02	1	1
WSCPIIX	X1	C02	2	1 (1)	A02	2	2
	X2	C06	3 (1)	3 (2)	A07	2	2
WSCPIIB	B1	C05	1	1	A09	1	1
	B2	C07	1	1	A07	1	1
no group		C06	(1)	(1)		-	-

**B.**

		Chr	<i>B. nigra</i>
WSCPIIA		B03	3
		B06	1
WSCPIIC		B05	2
WSCPIIX	X1	B03	3
	X2	B05	2
WSCPIIB	B1	B04	1
	B2	-	-

**Table S4. Sequence identities of *B. napus*, *B. oleracea* and *B. rapa* WSCP II proteins**

Identities >80% are in red. Identities <40% are in grey. Proteins identical are underlined. WSCP IIA (A), WSCP IIC (B), WSCP IIX1 (C); WSC IIX2 (D), WSCP IIB (E) and identical WSCP II proteins (E).

**A.**

WSCP IIA	<u>C06p46520.1_BnaDAR</u>	<u>BoIC6i39697H</u>	<u>BoIC6i39693H</u>	<u>C06p46460.1_BnaDAR</u>	<u>A07p40710.2_BraZl</u>	<u>A07p38410.1_BnaDAR</u>	<u>BoIC6i39698H</u>	<u>C06p46430.1_BnaDAR</u>	<u>A07p40690.2_BraZl</u>	<u>BoIC6i39694H.1_BnaDAR</u>
<u>C06p46520.1_BnaDAR</u>		98,5	93,5	89,5	88,0	88,0	75,9	64,0	63,7	60,2
<u>BoIC6i39697H</u>	98,5		95,0	90,0	89,5	89,5	76,9	65,0	64,7	60,7
<u>BoIC6i39693H</u>	93,5	95,0		93,0	89,5	89,5	77,3	64,5	64,2	60,2
<u>C06p46460.1_BnaDAR</u>	89,5	90,0	93,0		88,0	88,0	75,4	65,0	64,2	60,7
<u>A07p40710.2_BraZl</u>	88,0	89,5	89,5	88,0		100,0	78,8	64,2	64,4	61,3
<u>A07p38410.1_BnaDAR</u>	88,0	89,5	89,5	88,0	100,0		78,8	64,2	64,4	61,3
<u>BoIC6i39698H</u>	75,9	76,9	77,3	75,4	78,8	78,8		67,5	68,1	63,5
<u>C06p46430.1_BnaDAR</u>	64,0	65,0	64,5	65,0	64,2	64,2	67,5		95,3	88,9
<u>A07p40690.2_BraZl</u>	63,7	64,7	64,2	64,2	64,4	64,4	68,1	95,3		88,4
<u>BoIC6i39694H.1_BnaDAR</u>	60,2	60,7	60,2	60,7	61,3	61,3	63,5	88,9	88,4	

**B.**

WSCP IIC	<u>C02p29770.1_BnaDAR</u>	<u>BoIC2i09025H</u>	<u>A02p23410.2_BraZl</u>	<u>A02p17920.1_BnaDAR</u>	<u>C02p29790.1_BnaDAR</u>
<u>C02p29770.1_BnaDAR</u>		99,5	93,9	93,9	61,6
<u>BoIC2i09025H</u>	99,5		94,4	94,4	62,1
<u>A02p23410.2_BraZl</u>	93,9	94,4		100,0	58,6
<u>A02p17920.1_BnaDAR</u>	93,9	94,4	100,0		58,6
<u>C02p29790.1_BnaDAR</u>	61,6	62,1	58,6	58,6	

**C.**

WSCP IIX1	<u>C02p29750.1_BnaDAR</u>	<u>BoIC2i09022H</u>	<u>C02p29730.1_BnaDAR</u>	<u>A02p23340.2_BraZl</u>	<u>A02p17860.1_BnaDAR</u>	<u>A02p23370.2_BraZl</u>	<u>A02p17890.1_BnaDAR</u>
<u>C02p29750.1_BnaDAR</u>		100,0	94,3	93,1	93,1	80,1	78,2
<u>BoIC2i09022H</u>	100,0		94,3	93,1	93,1	80,1	78,2
<u>C02p29730.1_BnaDAR</u>	94,3	94,3		87,7	87,7	79,2	81,0
<u>A02p23340.2_BraZl</u>	93,1	93,1	87,7		100,0	80,2	78,3
<u>A02p17860.1_BnaDAR</u>	93,1	93,1	87,7	100,0		80,2	78,3
<u>A02p23370.2_BraZl</u>	80,1	80,1	79,2	80,2	80,2		97,6
<u>A02p17890.1_BnaDAR</u>	78,2	78,2	81,0	78,3	78,3	97,6	

**D.**

group IIX2	A07p40740.2_BraZ1	A07p38440.1_BnaDAR	C06p46550.1_BnaDAR	BoIC6t39700H	BoIC6t39705H	C06p46590.1_BnaDAR	C06p46540.1_BnaDAR	BoIC6t39699H	A07p40720.2_BraZ1	A07p38420.1_BnaDAR
A07p40740.2_BraZ1		99,0	94,0	94,0	56,9	56,9	60,5	60,5	59,0	59,0
A07p38440.1_BnaDAR	99,0		94,0	94,0	56,9	56,9	60,5	60,5	58,5	58,5
C06p46550.1_BnaDAR	94,0	94,0		100,0	59,3	59,3	62,7	62,7	61,2	61,2
BoIC6t39700H	94,0	94,0	100,0		59,3	59,3	62,7	62,7	61,2	61,2
BoIC6t39705H	56,9	56,9	59,3	59,3		100,0	51,7	51,7	50,7	50,7
C06p46590.1_BnaDAR	56,9	56,9	59,3	59,3	100,0		51,7	51,7	50,7	50,7
C06p46540.1_BnaDAR	60,5	60,5	62,7	62,7	51,7	51,7		100,0	96,8	96,8
BoIC6t39699H	60,5	60,5	62,7	62,7	51,7	51,7	100,0		96,8	96,8
A07p40720.2_BraZ1	59,0	58,5	61,2	61,2	50,7	50,7	96,8	96,8		100,0
A07p38420.1_BnaDAR	59,0	58,5	61,2	61,2	50,7	50,7	96,8	96,8	100,0	

**E.**

group IIB		A09p49030.2_BraZ1	A09p37590.1_BnaDAR	C05p27460.1_BnaDAR	BoIC5t31464H	C07p16850.1_BnaDAR	BoIC7t42191H	A07p13000.2_BraZ1	A07p11180.1_BnaDAR
IIB1	A09p49030.2_BraZ1		98,0	90,0	95,0	60,2	59,7	60,2	60,2
	A09p37590.1_BnaDAR	98,0		90,4	95,5	59,7	59,3	59,7	59,7
	C05p27460.1_BnaDAR	90,0	90,4		93,8	55,3	54,9	54,9	54,9
	BoIC5t31464H	95,0	95,5	93,8		58,8	58,3	58,3	58,3
IIB2	C07p16850.1_BnaDAR	60,2	59,7	55,3	58,8		99,5	95,8	96,3
	BoIC7t42191H	59,7	59,3	54,9	58,3	99,5		95,4	95,8
	A07p13000.2_BraZ1	60,2	59,7	54,9	58,3	95,8	95,4		99,5
	A07p11180.1_BnaDAR	60,2	59,7	54,9	58,3	96,3	95,8	99,5	

**F.**

	<i>B.napus</i>	<i>B.rapa</i>	<i>B.oleracea</i>
group IIA	A07p38410.1_BnaDAR	A07p40710.2_BraZ1	
group IIC	A02p17920.1_BnaDAR	A02p23410.2_BraZ1	
group IIX1	C02p29750.1_BnaDAR		BoIC2t09022H*
	A02p17860.1_BnaDAR	A02p23340.2_BraZ1	
group IIX 2	C06p46550.1_BnaDAR		BoIC6t39700H*
	C06p46590.1_BnaDAR		BoIC6t39705H
	C06p46540.1_BnaDAR		BoIC6t39699H
	A07p38420.1_BnaDAR	A07p40720.2_BraZ1	





## CHAPITRE III :

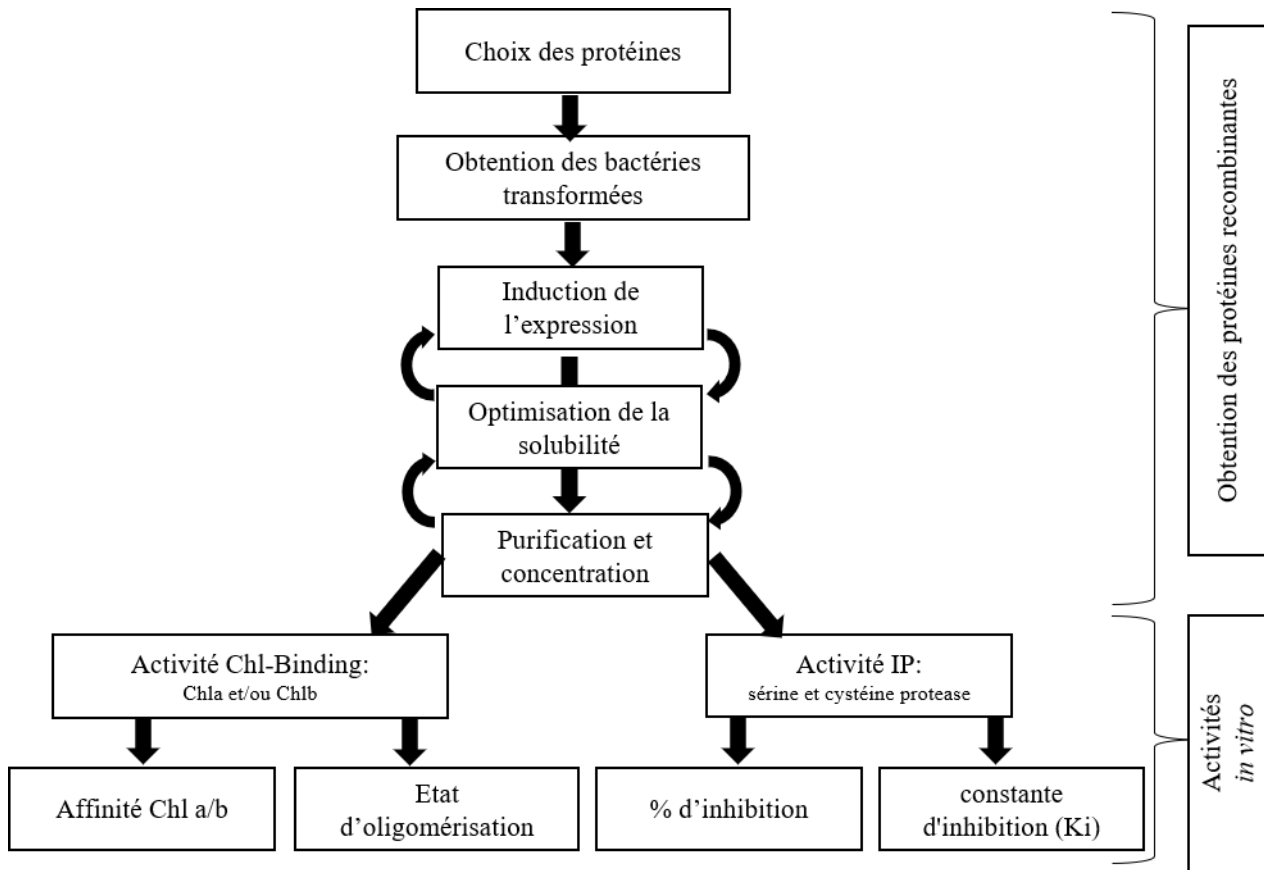
# DETERMINATION DES PROPRIETES FONCTIONNELLES DES WSCP II DE COLZA A PARTIR DE PROTEINES RECOMBINANTES

---

### Introduction

Chez le colza, les WSCP II constituent une famille multigénique de 28 gènes chez la variété Darmor nain dont 9 pseudogènes (chapitre II). Cette redondance d'homologues peut s'expliquer d'une part par la présence des deux sous génomes parentaux A et C et d'autre part par une spécificité tissulaire des WSCP II se répartissant entre des tissus chlorophylliens (feuilles, tiges, siliques) et non chlorophylliens (fleurs, racines, graines) (Bektas *et al.*, 2012; Takahashi *et al.*, 2013; Clouet *et al.*, en préparation). Le niveau d'expression de ces gènes diffère également selon les conditions environnementales, certains étant induits en réponse à des contraintes abiotiques ou biotiques (Ilami *et al.*, 1997; Desclos *et al.*, 2008; Boex-Fontvieille *et al.*, 2015b). La présence de multiples isoformes protéiques peut aussi s'expliquer par une redondance ou une diversité fonctionnelle en termes d'inhibition de certaines classes de protéases et d'affinité différente aux chlorophylles (Chls). De plus, des modifications post-traductionnelles possibles comme le clivage en C-terminal éliminant une séquence de 2 à 3 kDa (Ilami *et al.*, 1997; Takahashi *et al.*, 2012, 2013b), pourraient augmenter la diversité de WSCP II et influencer leurs propriétés.

Ce volet de mon travail vise à spécifier l'aptitude des WSCP II du colza à se lier à la chlorophylle et/ou à agir en tant qu'inhibiteurs de protéases (IP). Pour se faire, un nombre restreint de six protéines WSCP II, représentatives de la diversité retrouvée chez le colza, ont été sélectionnées sur la base d'une approche *in silico* reposant sur la présence de domaines spécifiques : i) les CBD qui comportent des séquences d'acides aminés indicatrices d'une aptitude à se lier aux Chls et le motif de sélectivité Chla/Chlb (Horigome *et al.*, 2007; Palm *et al.*, 2018; Prabakar *et al.*, 2020) ii) le domaine IP qui comporte le site réactif de liaison entre protéases et inhibiteurs siège d'activité d'inhibition de serine protéases (Ser-IP) (De Vonis Bidlingmeyer *et al.*, 1972) et iii) la présence des acides aminés impliqués dans l'interaction et l'inhibition de RD21, cystéine protéase, par AtWSCP (Boex-Fontvieille *et al.*, 2015b). Takahashi *et al.* (2014) ont montré *in vitro*, que l'absence de la séquence C-terminale, clivée lors de la maturation post-traductionnelle, jouerait un rôle dans la stabilité/solubilité de la protéine sans toutefois affecter l'activité de liaison aux Chls. Cependant, l'influence de cette séquence dans l'activité anti-protéolytique, l'affinité aux Chl ou la stabilité/solubilité des WSCP II de colza n'a jamais été caractérisée.



**Fig. 1 : Principales étapes pour l'obtention des protéines recombinantes WSCP II de Colza et leur caractérisation**

Après amplification par PCR du gène cible, le clonage est réalisé dans le vecteur d'expression pET21a(+). La sélection des bactéries transformées se fait en présence d'antibiotique. L'expression des protéines recombinantes est induite en présence d'IPTG dans le système bactérien et nécessite des étapes d'optimisation de production pour un rendement maximal de protéines recombinantes solubles. Après purification de la fraction soluble protéique par chromatographie d'affinité Ni<sup>+</sup>, les protéines recombinantes concentrées sont testées pour leurs propriétés de liaison à la Chl et leur inhibition contre les protéases à sérine ou à cystéine.

L'étude des interactions protéine-protéine ou protéine-chlorophylle nécessite la production au préalable de quantités importantes de WSCP II purifiées qui ont gardé leurs propriétés fonctionnelles. Nous avons fait le choix de tester les activités d'inhibition de protéases (Ser-IP et/ou Cys-IP) et de liaison aux Chls (a et/ou b) *in vitro* à partir de protéines recombinantes portant une séquence de 6 histidines (His-tag) en C-terminal produites en système bactérien. Cette démarche a déjà été validée avec succès pour l'étude de nombreuses WSCP II d'espèces végétales diverses (Schmidt *et al.*, 2003; Takahashi *et al.*, 2012, 2013b, 2014b) et pour BnD22 (chapitre I). Les principales étapes de la démarche nécessaires pour produire des WSCP II recombinantes et entreprendre leur caractérisation sont résumées dans la figure 1.

## Matériels et méthodes

### I. Clonage et obtention des bactéries transformées

Les séquences des WSCP II sélectionnées, dépourvues de leur peptide signal, ont été amplifiées par PCR à partir de l'ADN génomique du génotype «Darmor nain» avec l'ADN polymérase *PfuI* selon le protocole du fournisseur (Promega). Les séquences des amorces (annexe 4) ont été obtenues par le logiciel CLC Main Workbench (Version 8.1) suivant la méthode de clonage de Gibson Assembly (Gibson *et al.*, 2009). Les produits d'amplification ont été insérés dans le plasmide d'expression pET21a(+) par insertion dans les sites de restrictions *NdeI* et *XhoI* à l'aide du kit Gibson Assembly® Master Mix (NEB). La transformation des bactéries *E. coli* DH5 $\alpha$  (NEB) permet l'obtention des plasmides en quantité suffisante pour la transformation ultérieure de souches bactériennes spécialisées dans la production de protéines recombinantes (*E. coli* Rosetta-gami 2 (DE3), *E. coli* C41(DE3) et *E. coli* Lemo21(DE3), annexe 6).

### II. Production et purification des WSCP II recombinantes dans *E. coli*

#### II.1. Expression des protéines recombinantes

Les bactéries transformées *E. coli* Rosetta-gami 2 (DE3), *E. coli* C41(DE3) et *E. coli* Lemo21(DE3) sont mises en culture pendant 12 à 16 h dans 6 mL de milieu LB-ampicilline supplémenté par 50  $\mu$ g/mL de tétracycline (résistance portée par Rosetta-gami 2 et C41) ou 50  $\mu$ g/mL de chloramphénicol (résistance portée par Lemo21), à 37°C sous agitation à 200 rpm. Une aliquote de 100  $\mu$ l est utilisée pour inoculer une culture en petit volume de 10 ml qui permet de définir les paramètres optimaux d'induction de chaque WSCP II. La totalité des 6 ml de pré-culture est utilisée pour une culture en grand volume de 500 ml nécessaire à la production en quantité importante de protéines. La croissance bactérienne s'effectue sous agitation à 37°C jusqu'à atteindre une DO<sub>600nm</sub> comprise entre 0,6 et 1. L'expression de la protéine recombinante est induite par ajout de 1 mM

d'isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG) dans le milieu et les bactéries sont incubées pendant 4-16 h à 25 ou 37°C sous agitation douce. L'induction de la protéine est vérifiée par PAGE-SDS. Les bactéries sont collectées par centrifugation (15 min à 5000 $\times$ g) puis les culots sont stockés à -20°C.

## **II.2. Extraction des protéines recombinantes**

Les culots bactériens sont repris dans le tampon de lyse (50 mM tampon phosphate pH 8, 300 mM NaCl, 5 mM imidazole) contenant 1 mg/ml de lysozyme et 3,3  $\mu$ g/ml de DNase, des inhibiteurs de protéases (complete™, mini EDTA-free, Roche) et 0,5% (v/v) de Triton X100, puis placés sur un agitateur rotatif (40 rpm) pendant 10 min. Après une sonication de 3 pulses de 30 s entrecoupés de 1 min de refroidissement à 4°C, le lysat est centrifugé à 14.000 $\times$ g pendant 15 minutes à 4°C, afin de séparer les fractions soluble (surnageant) et insoluble (culot). La présence de la protéine recombinante dans les fractions soluble et/ou insoluble est vérifiée par PAGE-SDS.

## **II.3. Purification des protéines recombinantes**

La purification se fait sur AktaStart par chromatographie d'affinité Ni<sup>2+</sup> (résine de nickel Protino Ni-NTA 5 ml column, Macherey-Nagel). La protéine His-tag est purifiée par compétition avec l'imidazole, analogue structural de l'histidine. La fraction protéique soluble, filtrée sur membrane 0,22  $\mu$ m, est injectée sur la colonne de chromatographie d'affinité préalablement équilibrée avec 25 mL du tampon de lyse. Une étape de lavage de 50 ml de tampon de lyse contenant des concentrations croissantes d'imidazole de 10 à 70 mM permet d'éliminer le plus efficacement possible les protéines bactériennes contaminantes sans provoquer de décrochage massif de la protéine d'intérêt. L'élution finale de la protéine portant le motif His-tag est réalisée par 10 mL de tampon de lyse contenant 250 mM imidazole. Toutes les étapes (chargement, lavages et élution) sont suivies à 280 nm. Les fractions contenant les protéines d'intérêt sont ensuite dessalées (élimination du NaCl et de l'imidazole) et concentrées dans du tampon 50 mM Tris-HCl pH 7,5 sur concentrateurs Amicon® Ultra-4 10K (Merck Millipore). La concentration en protéines purifiées est déterminée à 280 nm en utilisant le coefficient d'extinction molaire estimé à partir de la séquence protéique (<http://web.expasy.org/protparam/>). Au cours des étapes de purification, les fractions de lavage et d'élution sont systématiquement analysées par PAGE-SDS pour contrôler les étapes de purification et estimer au final la pureté de la protéine recombinante.

## **II.4. Analyse par PAGE-SDS et Western-Blot**

Le fractionnement des protéines des échantillons issus de différentes étapes d'induction, de lyse et de purification des protéines recombinantes est suivi par électrophorèse en condition dénaturante selon la technique de PAGE-SDS (Laemmli, 1970). Des aliquotes contenant de 10 à 20  $\mu$ g de

protéines solubilisées dans le tampon de charge et dénaturées à 100°C pendant 5 min sont déposés sur le gel constitué d'un gel de concentration (4 % d'acrylamide) et d'un gel de séparation (15 % d'acrylamide). L'électrophorèse est conduite à 80 V pour le gel de concentration et à 150 V pour le gel de séparation en présence de tampon de migration. Après migration, les polypeptides sont révélés par coloration au bleu de Coomassie (annexe 7).

Pour le Western-Blot, après électrophorèse, les protéines sont transférées sur une membrane de type PVDF (polyfluorure de vinylidène) par électro-transfert semi-sec pendant 2 h en présence de la solution de transfert (annexe 7). Après transfert, la détection des protéines recombinante est réalisée par marquage immunologique en utilisant des anticorps primaires anti-poly Histidine révélés par des anticorps secondaires anti IgG de lapin couplés à la peroxydase. La révélation des protéines immuno-marquées est réalisée par photo-imager par suivi de la chemo-luminescence émise par le luminol oxydé en présence de H<sub>2</sub>O<sub>2</sub> par la peroxydase.

### **III. Mesure des activités anti- protéolytiques**

La trypsine et la chymotrypsine, deux Ser protéases, sont mises ou non en présence de différentes quantités de WSCP II (ratio molaires de 0 à 8) dans le tampon réactionnel composé de 50 mM Tris-HCl pH 7,5 pour la trypsine, supplémenté avec 50 mM CaCl<sub>2</sub> pour la chymotrypsine. L'activité enzymatique est obtenue par suivi cinétique de l'hydrolyse des substrats spécifiques des enzymes : 0,25 mM BAEE pour la trypsine et 0,25 mM BTEE pour la chymotrypsine. Les absorbances sont suivies à 253 nm (BAEE) et à 256 nm (BTEE) toutes les 15 secondes pendant 20 min. L'activité enzymatique correspond à la vitesse initiale (V<sub>i</sub>) équivalente au coefficient directeur de la tangente à l'origine de la cinétique. Le même principe est utilisé pour la papaïne, protéase à cystéine en présence de 2 mM BApNA avec suivi des absorbances à 405 nm dans le tampon réactionnel 20 mM MES pH 6,5, 5 mM cystéine. L'effet inhibiteur des WSCP II est visualisé par détermination de l'activité résiduelle des protéases en référence à son activité maximale en absence de WSCP II.

### **IV. Capacités de liaison à la chlorophylle**

#### **IV.1. Retard sur gel ou Electrophoretic Mobility Shift Assay (EMSA)**

Les protéines WSCP II recombinantes purifiées (5 µM) sont mises en contact de Chla ou de Chlb (15 µM ; ratio Chl/protéine = 3) pendant 30 minutes à 4°C dans 250 µl de tampon de liaison (50 mM Tris-HCl pH 7,5, 150 mM NaCl, 0,04% Nonidet-P40 (NP40)). Après électrophorèse sur gel de polyacrylamide 12% en conditions non-dénaturantes (conditions identiques à celles définies pour le PAGE-SDS mais sans détergent), les protéines sont colorées au bleu de Coomassie. La présence d'oligomères WSCP II est indicatrice de liaison avec les pigments.



## IV.2. Détermination de l'affinité des WSCPII aux chlorophylles

La détermination de l'affinité des WSCPII aux chlorophylles a été réalisée en utilisant l'appareillage Monolith NT. 115 prévu pour le suivi par Micro-Scale Thermophoresis (MST). Cette technique permet de mesurer les interactions entre deux molécules en temps réel avec de faibles quantités d'échantillon et de déterminer les constantes d'affinité pour une gamme se situant entre le nM et le mM (Jerabek-Willemsen *et al.*, 2014). Le principe est basé sur la détection du changement de fluorescence par la température d'une molécule cible en fonction de la concentration de son ligand potentiel non fluorescent. Ce changement de fluorescence est indicateur d'un changement dans le mouvement des molécules conséquent d'une liaison entre les deux molécules. Les chlorophylles étant naturellement fluorescentes, la méthode a été développée en utilisant leur fluorescence intrinsèque. Les chlorophylles (50 nM Chla ou 200 nM Chlb) ont été mises en présence de 16 concentrations de WSCPII recombinantes purifiées (0,15 nM à 10 µM) et leur fluorescence a été suivie dans le tampon de liaison composé de 50 mM Tris-HCl pH 7,8, 1 % octyl β-D-glucopyranoside (OG). La présence de photoblanchiment des Chls variant avec les concentrations de WSCPII avant application du gradient de température a empêché une analyse des constantes d'affinités (Kd) telle que menée habituellement par cette technique. Une solution alternative pour la détermination des Kd a été proposée sur la base des taux de photoblanchiment, calculés sur les 4-5 premières secondes d'illumination, et normalisés en fraction liée de Chl par l'équation 1. Les Kd sont ensuite calculés par régression non-linéaire (équation 2) à partir de courbes de dose-réponse de la fraction liée en fonction des concentrations en WSCPII.

Equation 1 :

$$Fraction\ liée = 1 - \left[ \frac{valeur\ (C) - libre}{liée - libre} \right]$$

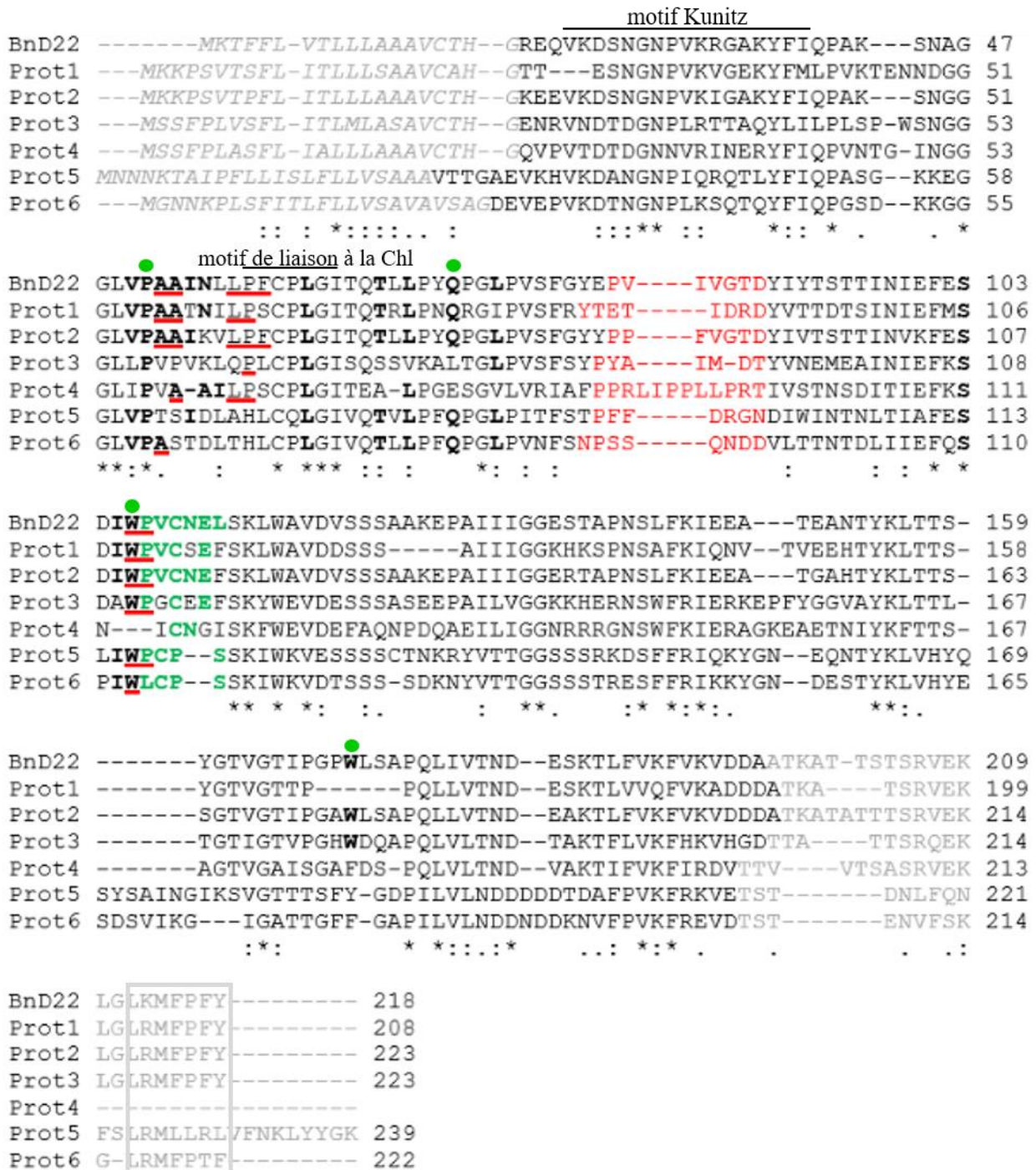
Avec valeur (C) correspondant au taux de photoblanchiment de la Chl à une concentration C de WSCPII; libre et liée correspondant respectivement aux taux de photoblanchiment de la Chl à la concentration minimale et maximale de WSCPII.

Equation 2 :

$$f(C) = Max + \frac{(Max - Min) \times (C + C_{target} + Kd - \sqrt{(C + C_{target} + Kd)^2 - 4 \times C \times C_{target}})}{2 \times C_{target}}$$

avec Max et Min correspondant à la fraction liée à la concentration (C) minimale et maximale de WSCPII respectivement ; C<sub>target</sub>, la concentration de Chl dans le milieu réactionnel.





**Fig. 2 : Alignement multiple des protéines WSCP II de colza sélectionnées et BnD22**

L'alignement est réalisé à l'aide du logiciel disponible sur Uniprot (<https://www.uniprot.org/align/>). Les astérisques indiquent les résidus conservés: les acides aminés de même nature biochimique. La séquence signal est en gris italique, et l'extrémité C-terminale subissant un clivage post-traductionnel est en gris avec la zone hautement conservée encadrée. Les motifs Kunitz et de liaison à la Chl caractéristiques des WSCP II sont indiqués.

Activité IP: En rouge, le site réactif siège de l'activité Ser-IP défini selon Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>); soulignés en rouge, les résidus impliqués dans l'activité Cys-IP définis par Boex-Fontvieille *et al.* (2015a).

Activité de liaison à la Chl : en gras, les résidus impliqués dans la liaison à la chlorophylle; rond vert, la position des 4 résidus interagissant avec le noyau tétrapyrrolique de la Chl selon Horigome *et al.* (2007); en vert, les motifs d'affinité à la Chl définis par Palm *et al.* (2018).

**Tabl. 1 : Caractéristiques géniques, physico-chimiques et fonctionnelles des 6 WSCP II sélectionnées**

Les caractéristiques physico-chimiques sont prédites par EXPASY (<https://web.expasy.org/protparam/>).

Pour l'activité IP, selon la nature des acides aminés constitutifs du site réactif (Conserved Domain Database, <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>), l'activité Ser-IP (type trypsine ou chymotrypsine) a été prédite (PeptideCutter, <http://www.expasy.org/tools/peptidecutter/>). L'activité Cys-IP est basée sur la présence ou non des acides aminés impliqués dans l'interaction entre AtWSCP et RD21 (Boex-Fontvielle *et al.*, 2015a). La même stratégie a été utilisée pour définir la capacité de liaison à la chlorophylle, en référence aux acides aminés impliqués dans l'interaction LvWSCP et BolWSCP avec la Chla ou Chlb (Horigome *et al.*, 2007; Palm *et al.*, 2018). De plus, l'affinité à la Chla ou Chlb a été définie selon l'étude menée par Palm *et al.* (2018). -, absence d'activité; +, présence d'activité avec le nombre de + proportionnel à la probabilité d'activité; nd, non déterminée.

Caractéristiques géniques			Caractéristiques physico-chimiques						Prédiction <i>in silico</i> des fonctions				
Prot	Groupes WSCP II	nom	avec C-terminal			sans C-terminal			Inhibition de protéase		liaison Chl		
			Nb AA	MW (kDa)	pI	Nbr AA	MW (kDa)	pI	Ser protéase		Cys protéase	Chl-liaison	Chl spécificité
									Tryp	Chym			
Prot1	A	C06p46430.1_BnaDAR	187	20,57	5,87	176	18,44	4,98	+++	++	+	++	nd
Prot2		A07p38410.1_BnaDAR	200	21,51	8,44	178	19,01	5,14	++	-	+	++	nd
Prot3	X	A02p17890.1_BnaDAR	199	22,16	6,35	180	19,93	5,56	+++	-	-	-	nd
Prot4		A07p38420.1_BnaDAR	190	20,51	7,22	178	19,26	5,54	-	+++	-	-	nd
Prot5	B	A07p11180.1_BnaDAR	216	24,26	9,12	189	20,96	7,84	++	+++	-	+++	Chla
Prot6		A09p37590.1_BnaDAR	199	21,83	5,21	182	19,87	4,74	-	-	-	+++	Chlb

**Tabl. 2 : Identité entre les séquences protéiques des 6 WSCP II sélectionnées entières et BnD22**

Les pourcentages d'identité sont représentés dans les cases non grisées et les distances entre paires de séquences dans les cases grisées. Les scores sont issus des alignements en Fig.2.

	BnD22	Prot1	Prot2	Prot3	Prot4	Prot5	Prot6
BnD22		0,32	0,12	0,60	0,70	0,90	0,84
Prot1	64,6		0,30	0,68	0,70	1,06	0,86
Prot2	86,5	66,5		0,59	0,70	0,97	0,84
Prot3	51,7	47,1	53,3		0,73	1,10	0,90
Prot4	43,1	40,9	44,4	42,9		1,19	1,07
Prot5	34,1	28,4	32,4	28,9	24,6		0,48
Prot6	37,4	34,9	37,34	35,8	28,4	58,1	



## Résultats

### I. Sélection et choix des protéines représentatives des WSCP II du colza

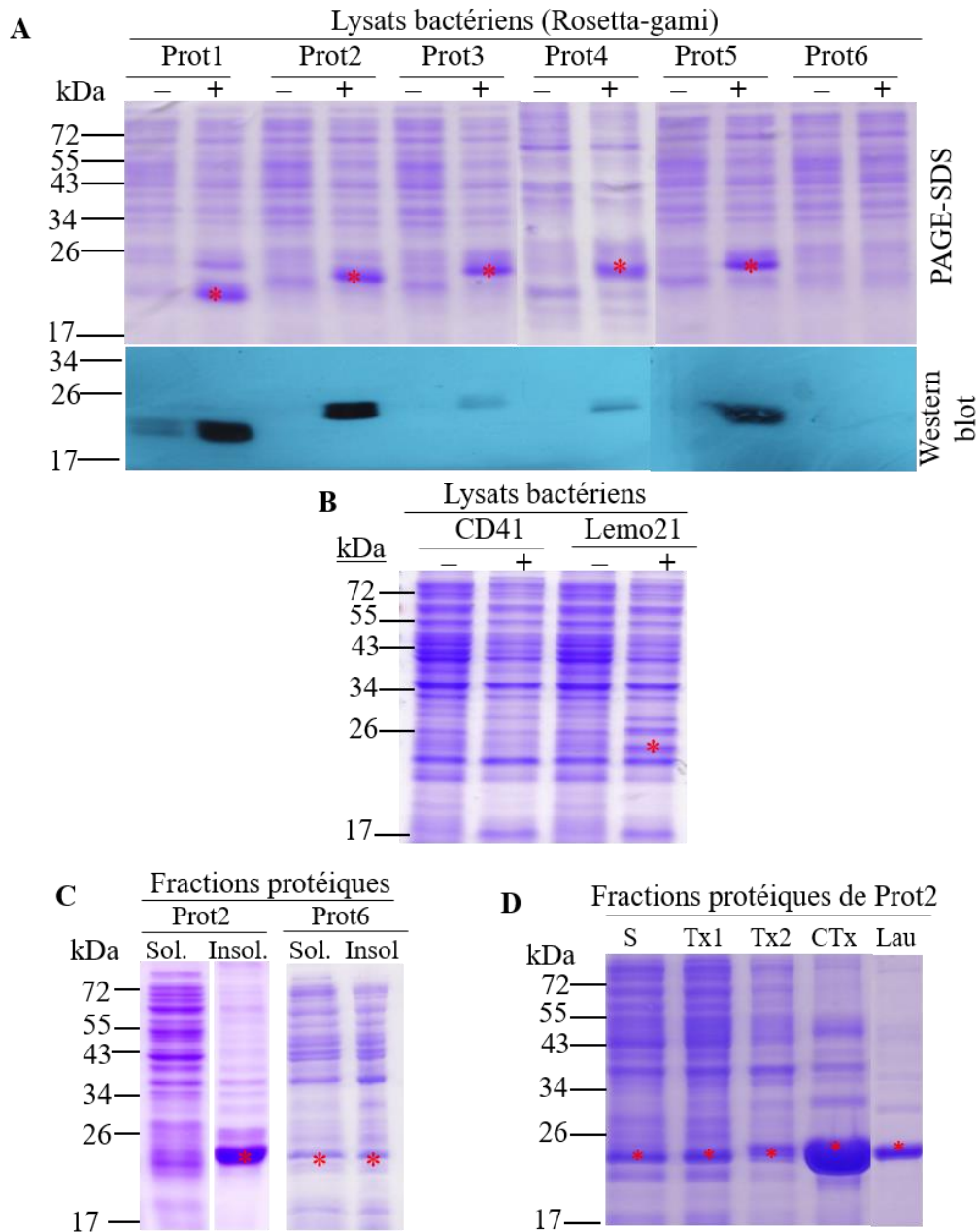
Sur les 19 WSCP II fonctionnelles identifiées chez le colza, notre choix s'est porté sur six protéines représentatives de la variabilité fonctionnelle *in silico* en fonction de leurs capacités de liaison à la Chl, d'inhibition aux protéases et en tenant compte de l'appartenance aux groupes phylogéniques IIA, IIB et IIX (Tab.1, Fig. 2). Ces protéines sont nommées Prot1 à Prot6. Les Prot1 et Prot2 appartiennent au groupe WSCP IIA et seraient, de ce fait, à même de se lier aux Chls et susceptibles d'inhiber les Ser protéases de type trypsine. En raison des différences de séquence sur le motif de spécificité de liaison à la chlorophylle tel que défini par Palm *et al.* (2018), leur affinité vis-à-vis des chlorophylles ne peut être prédite. Les Prot3 et Prot4 appartiennent au groupe WSCP IIX, protéines ayant perdu leur potentiel de liaison aux Chls (décrit par Prabahar *et al.*, 2020). La Prot3 serait une Ser-IP de type trypsine alors que la Prot4 une Ser-IP de type chymotrypsine. Dans le groupe WSCP IIB, les Prot5 et Prot6 se distinguent quant à leur spécificité de liaison soit à la Chla, soit à la Chlb. La Prot5 aurait de plus un pouvoir d'inhibition anti Ser protéase. En ce qui concerne la capacité d'inhibition vis-à-vis des papaine-like-Cys protéases, aucune de ces protéines ne possède l'ensemble des acides aminés définis comme importants lors de l'interaction entre AtWSCP et RD21 (Boex-Fontvieille *et al.*, 2015b). Les protéines retenues sont des protéines acides ou basiques avec des masses moléculaires allant de 20 à 24 kDa. A noter que ces 6 WSCP II montrent une diversité importante avec des identités comprises entre 24 et 66% entre elles (Tabl. 2). Les Prot1 et Prot2 (WSCP IIA) sont les plus similaires à BnD22 qui appartient au même groupe. Les différences les plus extrêmes sont trouvées entre les WSCP IIA (Prot1 et Prot2) et les WSCP IIB (Prot5 et Prot6).

Afin de clarifier le rôle de la séquence C-terminale dans l'activité ou/et la stabilité/solubilité des WSCP II, nous avons fait le choix de produire les protéines recombinantes avec ou sans cette séquence C-terminale (clivage de 2-3 kDa). L'absence de l'extrémité C-terminale abaisse le pI de toutes les protéines ce qui suggère que cette extrémité aurait un effet sur la charge globale de la protéine.

### II. Production et purification des WSCP II recombinantes

#### II.1. Expression et degré de solubilité des protéines recombinantes

Les vecteurs d'expression pour les 6 protéines entières ou tronquées (sans C-terminal) ont été utilisés pour la transformation de trois souches bactériennes disponibles. L'efficacité de production bactérienne des WSCP II entières a été testée, en premier lieu, pour la souche Rosetta-gami 2 (DE3) après induction par l'IPTG (Fig. 3A). A l'exception de la Prot6, la présence de protéines de la taille



**Fig. 3 : Induction et solubilité des WSCPII recombinantes entières (avec l'extrémité C-terminale)**

**A)** Expression des WSCPII dans les lysats bactériens (Rosetta-gami) après (+) ou non (-) induction par 1 mM d'IPTG, révélés par PAGE-SDS et par Western-blot

**B)** Expression de Prot6 dans les lysats bactériens des souches CD41 et Lemo21

**C)** Présence des Prot2 et Prot6 dans les fractions solubles (Sol) et insolubles (Insol)

**D)** Effet des détergents sur la solubilisation de la fraction insoluble de Prot2. S : fraction soluble du lysat bactérien; Tx1, Tx2: fraction solubilisée en présence de 0,5% Triton X100 ; CTx : fraction insoluble restant après extraction par le Triton X100 ; Lau : fraction solubilisée en présence de N-lauroyl sarcosine

Les lysats bactériens et les fractions (Sol.) et (Insol.) obtenues après centrifugation des lysats bactériens sont analysés par PAGE-SDS 15%. Les protéines sont révélées après coloration du gel au bleu de Coomassie ou immunodétectées par western blot avec l'anticorps anti-HisTag. Les protéines recombinantes sont indiquées par un astérisque.

attendue autour de 20 kDa a été révélée dans les lysats bactériens (extraits protéiques bruts) par PAGE-SDS, et confirmée par Western blot avec un anticorps anti-histidine. Deux autres souches bactériennes (CD41 et Lemo 21) ont été testées pour la production de la Prot6 en raison de sa réticence à s'exprimer dans la souche Rosetta-gami 2 (DE3) et seule la souche Lemo21 a permis l'expression de la Prot6 (Fig. 3B). A l'issue de cette première série d'expériences nous avons réussi à obtenir la production des 6 protéines sélectionnées entières.

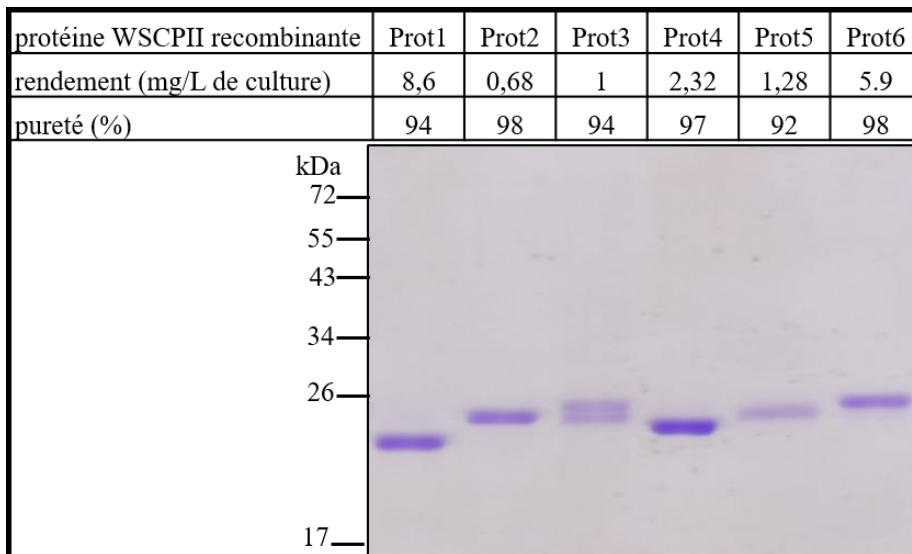
Comme l'induction de protéines hétérologues dans un système bactérien peut entraîner la formation de corps d'inclusion, ce qui constitue un obstacle majeur pour leur purification, la présence des protéines recombinantes a été recherchée dans les fractions soluble et insoluble obtenues après centrifugation du lysat bactérien. Il s'est avéré que, quelle que soit la nature de la protéine, la majorité des protéines recombinantes était retrouvée dans la fraction insoluble (Fig. 3C, exemple de Prot2). La Prot6 fait exception avec une proportion égale entre fractions soluble et insoluble (Fig. 3C). Suite à cette observation, de nombreuses conditions de culture ont été expérimentées afin d'obtenir le maximum de protéines recombinantes dans la fraction soluble, en utilisant cette fois-ci aussi bien les protéines entières que les protéines ne possédant plus la partie C-terminale clivée lors de la maturation post-traductionnelle. Les résultats de la production optimale de protéines solubles obtenue pour chaque protéine ont été rassemblés dans le tableau 3. Ils montrent que les conditions de culture (température, durée d'induction par IPTG) mais aussi la souche bactérienne utilisée pour obtenir une solubilité maximale sont très différentes selon les protéines. Par exemple, pour Prot1, les meilleurs résultats étaient obtenus avec la souche Lemo21 cultivée à 26°C et induite pendant 16 h alors que pour Prot5 c'est la souche CD41 cultivée à 37°C et induite pendant 4 à 6 h. Par ailleurs, les niveaux de solubilité diffèrent en fonction du type de protéines recombinantes. Il faut noter que pour une même protéine les conditions sont différentes selon qu'elle est entière ou sans son C-terminal. Par exemple Prot1 entière est produite dans Lemo21 alors que Prot1 sans C-terminal est produite dans Rosetta. A l'exception de Prot4 qui montre une solubilité améliorée lorsque le C-terminal est absent, les autres protéines ne montrent pas de différence de niveau de solubilité entre les 2 formes (Tabl. 3).

Lors de ces étapes d'optimisation, comme une proportion non négligeable des protéines recombinantes était encore systématiquement retrouvée dans la fraction insoluble, nous avons eu recours à l'utilisation de détergents non dénaturants comme le Triton X100 ou la N-lauroyl sarcosine qui ont permis d'améliorer la solubilisation (Fig. 3D, exemple Prot2). Toutefois comme la présence de N-lauroyl sarcosine s'est avérée par la suite incompatible avec l'étape de purification par chromatographie d'affinité, nous avons opté pour la solubilisation des protéines lors de la lyse bactérienne en présence de Triton X100 avec deux extractions successives. Un rendement de 30%

**Tabl. 3 : Conditions de culture optimales produisant le meilleur taux de solubilité pour chacune des 6 WSCP II recombinantes entières ou tronquées**

Rosetta : *E. coli* Rosetta-gami 2 (DE3); CD41 : *E. coli* C41(DE3); Lemo21 : *E. coli* Lemo21(DE3); durée incubation, temps de culture après addition d'IPTG; solubilité, pourcentage de protéines recombinantes dans la fraction soluble par rapport à la totalité de protéines produites.

protéine	avec C-terminal				sans C-terminal			
	souche bactérienne	température (°C)	durée incubation (h)	solubilité (%)	Souche bactérienne	température (°C)	durée incubation (h)	solubilité (%)
Prot1	Lemo21	26	4-6	50	Rosetta	37	4-6	50
Prot2	Lemo21	37	16	30	Rosetta	37	4-6	30
Prot3	CD41	26	4-6	30	Lemo21	37	16	30
Prot4	Rosetta	26	16	50	Rosetta	26	16	75
Prot5	CD41	37	4-6	40	Rosetta	26	4-6	30
Prot6	Lemo21	37	16	50	---	---	---	---



**Fig. 4 : Analyse par PAGE-SDS des protéines WSCP II recombinantes entières (avec C-terminal) après purification**

Les protéines recombinantes sont purifiées par chromatographie d'affinité sur colonne Ni<sup>+</sup> à partir de la fraction soluble du lysat bactérien. Après purification, le NaCl et l'imidazole sont éliminés et les protéines sont concentrées. Le niveau de pureté des protéines recombinantes est estimé par « GelAnalyzer » après coloration au bleu de Coomassie du gel.

à 50% des protéines recombinantes entières dans la fraction soluble a été ainsi rendu possible. Pour les protéines recombinantes tronquées, à l'exception de Prot6, leur solubilité après optimisation des conditions de culture varie entre 30 à 75% (Tabl. 3).

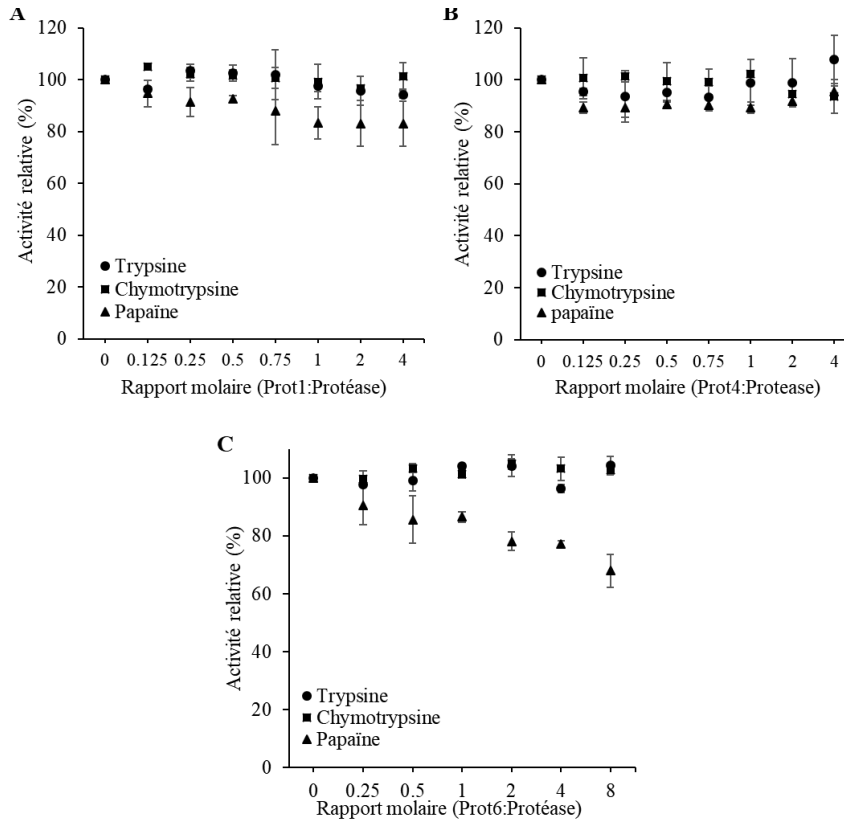
Lorsque les conditions d'induction et de solubilité ont été jugées satisfaisantes, la production en grand volume a été effectuée pour obtenir des quantités compatibles avec les études fonctionnelles, lesquelles nécessitent des protéines pures et fonctionnelles en quantités importantes. Selon la nature des protéines, le passage d'un petit à un grand volume de culture a amené parfois des problèmes d'induction et de solubilité nécessitant de réadapter à nouveau les conditions de culture. L'étape de purification des protéines par chromatographie d'affinité n'a été effectuée que pour les protéines entières.

## **II.2. Purification des protéines recombinantes solubles**

La purification des protéines recombinantes repose sur l'affinité de liaison entre le nickel et le His-tag des protéines et leur élution par compétition avec l'imidazole, un analogue structural de l'histidine.

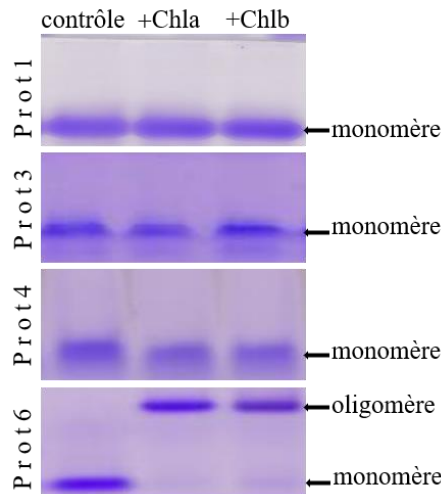
Des tests de purification des protéines solubles ont été réalisés en premier lieu avec des billes de nickel sur des volumes réduits de cultures bactériennes afin de vérifier l'affinité des protéines recombinantes pour le nickel et de définir les conditions d'élution pour leur purification. Lors du passage de la production en grand volume, il a été nécessaire de modifier les conditions d'élution pour l'adapter à la purification par chromatographie basse pression : i) faible débit lors de l'étape de chargement et de fixation des protéines sur les sites Ni<sup>+</sup> de la chromatographie, ii) modification du nombre, du volume, du débit et de la concentration en imidazole des étapes de lavages afin de permettre l'élimination d'un maximum de protéines contaminantes tout en évitant l'élution de la protéine d'intérêt. Après purification, l'élimination du NaCl et de l'imidazole, qui à fortes concentrations peuvent nuire aux activités, a été réalisée ainsi que la concentration des protéines purifiées (Fig. 4). Les protéines recombinantes obtenues sont très pures (degrés de pureté entre 92 et 98%) mais avec des rendements très différents. L'ensemble des tests d'activité IP et liaison à la Chl en prenant en compte les triplicatas (% d'inhibition IP, Ki, affinité à la Chl par MST) nécessite environ 1 mL de protéine recombinante purifiée à la concentration de 1,8mg/mL. Les rendements des Prot1 et Prot6 sont compatibles pour assurer l'ensemble des études fonctionnelles envisagées. En revanche, les rendements trop faibles obtenus pour les Prot2, Prot3, Prot4 et Prot5 s'avèrent insuffisants pour mener l'ensemble des études fonctionnelles. Pour ces protéines, plusieurs purifications à partir de volumes plus conséquents de culture bactérienne et/ou l'extraction et la purification des protéines recombinantes à partir de la fraction insoluble pourraient être envisageables.





**Fig. 5 : Evaluation de l'inhibition des activités protéolytiques de la trypsine, chymotrypsine et papaine par des quantités croissantes de WSCPII recombinantes entières purifiées.**

Après pré-incubation de la protéase (trypsine, chymotrypsine ou papaine) avec différentes quantités de WSCPII purifiées **A)** Prot1 **B)** Prot4 **C)** Prot6 (rapports molaires WSCP II/protéase de 0 à 8), les activités enzymatiques sont estimées par suivi cinétique de l'hydrolyse des substrats spécifiques de la trypsine (BAEE), de la chymotrypsine (BTEE) et de la papaine (BApNA). Les absorbances des produits de clivage sont mesurées respectivement à 253, 256 et 405 nm. Les activités sont exprimées en activité relative (%) par rapport à la protéase en absence de WSCPII.



**Fig. 6 : Tests d'oligomérisation des WSCPII recombinantes purifiées Prot1, Prot3, Prot4 et Prot6 entières en absence de Chl (contrôle) ou en présence de la Chla (+Chla) ou de la Chlb (+Chlb)**

Après incubation de la protéine recombinante (5  $\mu$ M) avec 15  $\mu$ M de Chla ou Chlb, les mélanges réactionnels sont séparés par PAGE en condition native (EMSA) puis les complexes sont révélés par coloration du gel au bleu de Coomassie.

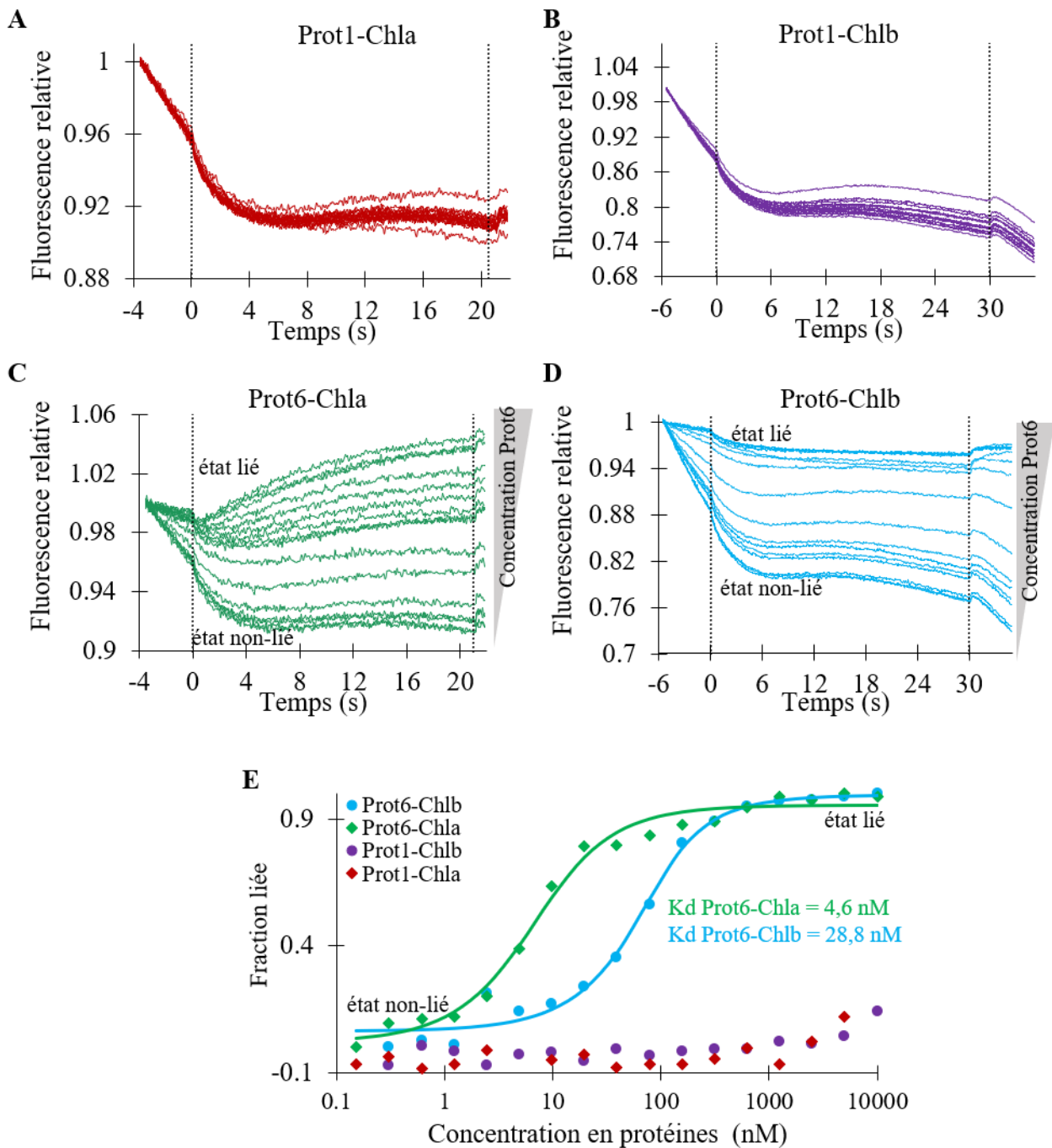
### III. Evaluation des propriétés fonctionnelles des WSCP II recombinantes

Les études préliminaires pour le suivi des activités IP et de liaison aux Chls ont été réalisées pour les Prot1, Prot4 et Prot6 (activités IP et de liaison à la Chl). Pour la Prot3, seule sa capacité à se lier à la Chl a été testée.

#### III.1. Activités anti-protéolytiques

Afin de confirmer ou infirmer les propriétés prédites d'inhibition vis-à-vis des protéases à sérine et/ou à cystéine des WSCP II sélectionnées (Tabl. 1), les activités protéolytiques de deux protéases à sérine (trypsine et chymotrypsine) et d'une protéase à cystéine (papaine) ont été suivies en présence des quantités croissantes de protéines WSCP II (Fig. 5). Les conditions de suivi de ces activités protéolytiques ont été mises au point en fonction de la nature du tampon, du pH du milieu réactionnel, des concentrations en substrats ou en enzymes. Ces conditions étant fixées, les propriétés d'inhibition des WSCP II sont suivies par pré-incubation de la protéase cible en présence des concentrations croissantes de WSCP II et mesure de l'activité résiduelle de la protéase après addition du substrat. Comme le montre la Fig. 5, dans nos conditions expérimentales, et quels que soient les ratios molaires protéines recombinantes/protéase, aucune inhibition des activités protéolytiques n'a été observée pour les Prot1 et Prot4 (Fig. 5 A, B) alors que celles-ci étaient susceptibles *in silico* d'inhiber respectivement la trypsine et la chymotrypsine. La Prot6 ne présente pas non plus de capacité IP contre les Ser protéases (résultats non montrés). En revanche, elle inhibe les Cys protéases de type papaine avec un maximum de 30% d'inhibition pour la concentration la plus forte en WSCP II (Fig. 5C).

Concernant les trois autres WSCP II sélectionnées (Prot2, Prot3 et Prot5), leur capacité d'IP pourra être réalisée dès lors que ces protéines pures et concentrées auront été obtenues. Au vu des rendements obtenus, plusieurs purifications (>5) seront réalisées pour obtenir les quantités suffisantes pour l'ensemble des activités IP. L'ensemble des activités Ser IP (% d'inhibition et  $K_i$  pour les ratios molaires protéine/protéase de 0 à 8 en triplicatas) nécessite environ 1 ml de protéine purifiée à 0,5 mg/ml et 300  $\mu$ l de protéine purifiée à 1,4 mg/ml sont nécessaires pour l'ensemble des activités Cys IP (% d'inhibition et  $K_i$  pour les ratios molaires protéine/protéase de 0 à 4 en triplicatas). De telles concentrations (1,4 mg/ml) pourraient être difficiles à atteindre à cause de la formation d'agrégats. Cette limitation nous conduit à envisager une détermination de l'absence/présence de l'activité IP à une seule condition qui serait le ratio protéine/protéase équimolaire.



**Fig. 7 : Caractérisation de la liaison des WSCP II recombinantes entières à la Chl**

**A-D)** Tracés bruts thermophorétiques des interactions Prot1-Chla (**A**), Prot1-Chlb (**B**) Prot6-Chla (**C**) et Prot6-Chlb (**D**).

**E)** Courbes de dose réponse de l'interaction entre la Chl et Prot1 ou Prot6 à partir des taux de photoblanchiments représentés par la fraction liée de Chl en fonction des doses croissantes de protéines.

Après incubation des Chls (50 nM Chla ou 200 nM Chlb) avec des quantités croissantes de WSCP II purifiées (0,15 nM à 10 µM), la fluorescence intrinsèque des Chls a été suivie. Les pointillés en noir définissent la zone d'application du laser. Les taux de photoblanchiment pour chaque concentration en WSCP II pendant la phase d'illumination avant application du laser sont calculés puis normalisés en fraction de Chl liée, ce qui permet la détermination des  $K_d$ .

### III.2. Capacité de liaison à la chlorophylle

La liaison de WSCPII avec les Chls a été réalisée dans le tampon de liaison en présence de Nonidet P40 (0,04%) nécessaire à la solubilisation de la Chl. La liaison des WSCPII avec les Chls entraîne la formation de complexes tétramériques visualisables par électrophorèse en conditions non dénaturantes (Fig. 6). Parmi les quatre protéines Prot1 (WSCPIIA), Prot3 et Prot4 (WSCPIIX) et Prot6 (WSCPIIB), testées dans les conditions expérimentales avec un ratio Chl/prot =3, seule la Prot6 montre une capacité de liaison aussi bien à la Chla qu'à la Chlb se traduisant par la formation d'oligomères Prot6-Chl. Les Prot2 et Prot5 sont des protéines basiques qui, de ce fait, ne sont pas compatibles à une visualisation de leur état éventuel d'oligomérisation en présence de Chl par EMSA. L'approche par chromatographie d'exclusion de taille pour déterminer la formation ou non de complexe avec la Chl peut-être une alternative possible. L'ajout de Chl directement dans les lysats bactériens suivi d'une purification par chromatographie d'affinité Ni<sup>2+</sup> pourrait être également envisagé. Cette méthode présenterait le double avantage i-de surmonter les obstacles relatifs aux problèmes de solubilité lors de l'extraction et ii-d'obtenir directement les complexes protéines-Chl purifiés (Takahashi *et al.*, 2013a; Bednarczyk *et al.*, 2015).

Afin de déterminer l'affinité des protéines pour les Chls et de confirmer ou non leur capacité de liaison en jouant sur les proportions entre WSCPII et Chl, des expériences ont été menées par une méthode dérivée de la MST basée sur les taux de photoblanchiment observé avant application du gradient de température par le laser. Seules les Prot1 et Prot6 ont été testées. Aucune variation du signal MST n'est observée pour la Prot1 indépendamment des concentrations mises en jeu, ce qui confirme l'incapacité de cette protéine à fixer la Chla ou la Chlb (Fig. 7 A, B). La variation graduelle du signal MST lors des interactions entre Prot6- Chla ou Prot6-Chlb valide la capacité de cette WSCPIIB à se lier aux chlorophylles (Fig. 7 C, D). De plus, lors de la phase d'illumination, un phénomène de photoblanchiment est observé lors de l'interaction des Prot6 et Prot1 avec la Chla ou la Chlb mais seules les interactions Prot6-Chla et Prot6-Chlb présentent un photoblanchiment des chlorophylles qui est dépendant de la concentration en protéines. La diminution du photoblanchiment concomitante à l'augmentation des concentrations en Prot6 suggère très fortement une photoprotection des Chl liées à la Prot6. Les constantes de dissociation (Kd) à 4,6 nM pour la Chla et 28,8 nM pour la Chlb vont en faveur d'une préférence de la WSCPIIB pour la Chla par rapport à la Chlb. Ce résultat est surprenant car contraire aux propriétés définies pour les WSCPIIB telle que LvWSCP et en référence avec le motif d'affinité de la Chlb présent dans le Prot6. Les valeurs de Kd sont surestimées en raison des concentrations importantes de Chl et de détergent utilisées et devront être affinées en prenant en compte les conditions mises au point pour la BnD22 (chapitre 1). Par ailleurs, contrairement à BnD22 (chapitre 1), l'interaction biphasique entre Prot6 et la Chl n'est pas

observée. Les concentrations utilisées pour Prot6 seront à revoir afin de pouvoir comparer de façon plus précise les affinités de ces deux WSCPIIA pour les Chls.

Pour les autres WSCPII sélectionnées, leur capacité et leur affinité de liaison à la Chl par la méthode dérivée de thermophorèse pourront être évaluées dès lors que ces protéines pures et concentrées auront été obtenues. Plusieurs purifications (>4) seront à réaliser pour obtenir les quantités suffisantes pour le suivi de l'ensemble des activités de liaison à la Chl.

## Discussion, conclusion et perspectives

La détermination des propriétés fonctionnelles des WSCPII en tant qu'IP et de leur aptitude à se lier aux Chl dépend de l'obtention des protéines WSCPII recombinantes de qualité et en quantité satisfaisante. Nous avons montré que dans la fraction soluble, la qualité et la quantité de protéines WSCPII produites dépendaient de la nature des protéines WSCPII de colza sélectionnées pour cette étude, lesquelles présentent entre 24 et 66% d'identité entre elles. Ces résultats corroborent ceux obtenus par Takahashi *et al.* (2014) qui ont montré que l'efficacité de l'induction, de la solubilité et de la stabilité variait selon la nature des WSCPII. Les WSCPII His tag entières produites ont de très bons niveaux de pureté. Cependant, malgré une optimisation des conditions de production, la proportion en protéines solubles reste faible avec pour conséquence la nécessité de volumes importants de culture bactérienne pour obtenir des quantités suffisantes en protéines pour le suivi de leur propriété fonctionnelle.

En tant que membres de la famille des IP de type Kunitz, les protéines WSCPII présentent un site réactif potentiel d'inhibition aux Ser protéases dans la boucle IV (Satoh *et al.*, 2001). A ce jour, aucune WSCPII recombinante testée *in vitro* n'a présenté d'activité IP significative contre cette classe de protéases en prenant comme cible la trypsine ou la chymotrypsine (Nishio & Satoh, 1997; Satoh *et al.*, 1998; Boex-Fontvieille *et al.*, 2016a). Il en est de même pour les trois protéines du colza, Prot1, Prot4 et Prot6 testées alors qu'*in silico* la probabilité d'inhibition était proche de 100%. Une autre Ser protéase cible comme la subtilisine d'origine végétale sera également à étudier. AtWSCP, la WSCPIIA d'*Arabidopsis thaliana*, est capable d'inhiber les protéases à cystéine comme la papaine et RD21 (Halls *et al.*, 2006; Rustgi *et al.*, 2017). Parmi les 3 WSCPII de colza testées, seule la Prot6 (WSCPIIB) se comporte comme un Cys-IP alors qu'elle ne possède pas les acides aminés définis dans l'interaction entre AtWSCP et RD21 et pourrait donc inhiber une autre protéase à Cys. Pour compléter les potentialités des WSCPII comme IP, les protéases à aspartate considérées également comme des cibles des IP Kunitz (Mosolov & Valueva, 2005) devront faire l'objet de futurs travaux. De plus, des modifications post-traductionnelles tel que le clivage en C-terminal peuvent modifier les propriétés anti-protéolytiques des IP à domaine Kunitz (Momonoki *et al.*, 2002). Dans ce contexte, les tests d'inhibition de protéases avec les protéines recombinantes tronquées seront à envisager. L'absence ou une faible activité IP des WSCPII peut également s'expliquer par des problèmes de conformation des protéines suite à leur production dans le système bactérien. Une possibilité de confirmer ou infirmer l'absence d'activité IP serait d'effectuer une détection des activités IP sur des protéines natives purifiées à partir des tissus végétaux. Toutefois, la présence de multiples isoformes protéiques WSCPIIA dans les feuilles complique la purification spécifique de chacune d'entre elles (Ilami *et al.*, 1997; Etienne *et al.*, 2007; Desclos *et al.*, 2008). Une autre approche est l'identification

des protéases partenaires (RD21 ou autre protéase) soit par co-immunoprécipitation, méthode utilisée avec succès par Boex-Fontvieille *et al.* (2015 a-b) pour AtWSCP ou par pull-down.

La capacité de liaison à la Chla ou la Chlb été évaluée *in vitro* par formation d'oligomères pour quatre des 6 protéines sélectionnées (Prot1, Prot3, Prot4 et Prot6) et validée par la technique dérivée de la thermophorèse pour deux d'entre elles (Prot1 et Prot6). Malgré une forte probabilité de liaison à la Chl de la Prot1 prédite *in silico*, aucune liaison à la Chla ou à la Chlb n'a été détectée expérimentalement alors qu'elle est membre des WSCP IIA décrites comme efficaces. L'absence de formation d'oligomères pour Prot3 et Prot4 (WSCP IIX) en présence de Chls confirme la prédiction *in silico*. Cette perte de la propriété de liaison à la Chl s'explique par l'absence ou la substitution des acides aminés (Pro, Gln, Trp, Trp) primordiaux pour la liaison à la chlorophylle (Horigome *et al.*, 2007). L'absence de liaison à la Chl des Prot3 et Prot4 confirme l'existence du groupe WSCP IIX chez le colza (Prabahar *et al.*, 2020) concordante avec leur expression dans les tissus non chlorophylliens que sont les tiges et les racines (chapitre 2). On pourrait s'attendre cependant à ce que ces protéines gardent une propriété d'IP pour expliquer un rôle physiologique *in planta*.

La capacité de la Prot6 à se lier aux Chls et former des complexes oligomériques est démontrée *in vitro* avec des affinités fortes de l'ordre du nM. En contradiction avec la prédiction *in silico* selon laquelle la Prot6 serait plus affine à la Chlb car membre du groupe WSCP IIB, nos données expérimentales montrent une préférence pour la Chla six fois supérieure à la Chlb. Cette particularité n'est pas expliquée par les acides aminés constitutifs du motif de sélectivité défini par Palm *et al.* (2018). La Prot6 possède les 4 acides aminés LCPS du motif retrouvé également chez LvWSCP de *L. virginicum*, WSCP IIB avec une préférence 16 fois plus importante pour la Chlb par rapport à la Chla (Palm *et al.*, 2018). Les valeurs de Kd pour la Prot6 sont vraisemblablement surestimées en raison des concentrations importantes de Chl et de détergent utilisées dans les conditions expérimentales. Les conditions décrites pour BnD22 (chapitre I, Bouargalne *et al.*, 2022) avec des concentrations faibles en Chl (50 nM Chla et 80 nM Chlb) et en détergent (0,04% Nonidet-P40) pourront être appliquées lors de futurs travaux sur la Prot6 et les autres WSCP II de colza. De plus, Takahashi *et al.* (2014) ont montré que l'extrémité C-terminale ne perturbe pas la liaison à la Chl pour BoIWSCP, RshWSCP et LvWSCP. Des expériences de type thermophorèse devront également être menées pour étudier l'interaction de la Prot6 tronquée avec les Chls afin de préciser si cette séquence a un rôle ou non dans l'affinité aux Chls.

Les études phylogénétiques de Prabahar *et al.* (2020) ont démontré que les membres de la famille des WSCPII ont évolué en fonction de leur aptitude à se lier préférentiellement aux Chla ou Chlb avec un sous-groupe parmi les WSCPIIA qui a perdu cette aptitude. Cette évolution n'est pas incompatible avec au sein des groupes des spécificités en terme d'IP. A ce jour, seule AtWSCP (Boex-Fontvieille *et al.*, 2015b, 2016a), BnD22 (chapitre 1) et Prot6 (WSCPIIB) portent une double fonction. Parmi les protéines testées, seules les protéines ayant une capacité de liaison aux Chl possèdent également une activité Cys-IP ce qui pourrait suggérer une interdépendance de ces deux fonctions. Comme pour BnD22, il serait intéressant de vérifier si la Prot6 complexée avec la Chl conserve son pouvoir IP. L'absence de fonction des deux WSCPIIX (Prot3 et Prot4) n'est pas compatible avec le maintien de leur expression dans les plantes et sous-entend la possibilité d'autre(s) propriété(s) qui restent à déterminer. Ces fonctions devront être abordées dans les futures études biochimiques et physiologiques en s'appuyant sur la classification des WSCPII de colza présentée dans le chapitre 2.



**Tabl.1 : Propriétés fonctionnelles de WSCP II caractéristiques de la diversité naturelle chez le colza et sites principaux d'expression.**

ND= non détecté ; ? = non déterminé

groupe	gène	protéine	Activité ou fonction biochimique										Expression transcriptionnelle		
			chlorophylle				Ser IP				Cys IP		stade plantule	stade végétatif	stade reproducteur
			liaison		affinité		trypsine		chymotrypsine		papaïne				
			<i>in silico</i>	<i>in vitro</i>	<i>in silico</i>	<i>in vitro</i>	<i>in silico</i>	<i>in vitro</i>	<i>in silico</i>	<i>in vitro</i>	<i>in silico</i>	<i>in vitro</i>			
WSCP IIA	C06p46460.1	BnD22	+++	+++	Chla	Chla	-	-	-	-	+++	+++	cotylédon	feuilles	feuilles, fleurs, siliques+graines
	C06p46430.1	Prot1	+	-	-	-	+++	-	-	-	++	-	cotylédon	feuilles	feuilles
	A07p38410.1	Prot2	-	-	-	-	+++	-	-	-	+	-	ND	feuilles	fleurs, siliques+graines
WSCP IIX	A02p17890.1	Prot3	-	-	-	-	+	-	-	-	-	-	?	?	?
	A07p38420.1	Prot4	-	-	-	-	-	-	+++	-	-	-	ND	tige	tige + fleurs
WSCP IIB	A07p1180.1	Prot5	+	-	Chla	-	++	-	++	-	-	-	ND	ND	ND
	A09p37590.1	Prot6	++	+++	Chlb	Chla	-	-	-	-	-	+++	ND	ND	ND

## DISCUSSION GENERALE ET PERSPECTIVES

---

Les WSCP<sub>II</sub>, membres des inhibiteurs de protéases de type Kunitz, sont des protéines qui, bien que n'étant pas impliquées dans le processus de photosynthèse, sont capables de se complexer avec les chlorophylles et de les protéger contre la photo-dégradation. Les WSCP<sub>II</sub> sont devenues des protéines modèles pour étudier les interactions protéines-Chl. Alors que les WSCP<sub>II</sub> sont subdivisées en sous-classes IIA et IIB selon les ratios Chl<sub>a</sub>/Chl<sub>b</sub> retrouvés dans les complexes natifs WSCP<sub>II</sub>-Chl, leurs affinités à la Chl<sub>a</sub> *versus* la Chl<sub>b</sub> ne sont pas encore précisément définies. De plus, certaines WSCP<sub>II</sub>, appelées WSCP<sub>IIX</sub>, ont perdu leur aptitude de liaison aux Chls (Prabahar *et al.*, 2020), et la fonction des WSCP<sub>II</sub> en tant qu'inhibiteur de protéases (IP) est très peu étudiée. A l'exception de la WSCP<sub>II</sub> unique d'*A. thaliana*, aucune étude n'indique clairement qu'une même protéine puisse assurer les deux fonctions.

Le cœur du travail présenté dans ce manuscrit a consisté à réaliser une étude fonctionnelle i) de BnD22, une WSCP<sub>II</sub> de colza dont les caractéristiques en tant qu'inhibiteur de protéases restent sujettes à controverse et dont la capacité de liaison aux Chls n'a jamais été décrite et ii) d'autres WSCP<sub>II</sub> de colza nouvellement identifiées, représentatives de la diversité naturelle, et sélectionnées selon leur appartenance aux groupes phylogéniques (IIA, IIB et IIX).

La présence de multiples WSCP<sub>II</sub> (19 protéines chez le colza) peut s'expliquer par un processus de sous-fonctionnalisation qui conduit à l'acquisition de spécificités fonctionnelles ou d'expression spécifiques soit tissulaires soit en réponse aux contraintes abiotiques et biotiques. Les expressions transcriptionnelles des WSCP<sub>II</sub> de colza, ont été suivies par qRT-PCR dans différents organes du colza cultivé en conditions témoins au stade plantule, végétatif et reproducteur (Tab.1). Les propriétés fonctionnelles des WSCP<sub>II</sub> de colza ciblées ont été suivies *in vitro* sur des protéines recombinantes portant une étiquette His-tag et produites en système bactérien (Tab1). Cette stratégie, déjà validée pour de nombreuses WSCP<sub>II</sub> de Brassicacées telles que AtWSCP, LvWSCP, BoIWSCP et RshWSCP (Nishio & Satoh, 1997; Schmidt *et al.*, 2003; Halls *et al.*, 2006; Takahashi *et al.*, 2012, 2013a,b, 2014b; Boex-Fontvieille *et al.*, 2016a), s'est avérée efficace pour les 7 WSCP<sub>II</sub> ciblées (BnD22 et Prot1 à 6) et a permis d'obtenir des protéines recombinantes avec un degré de pureté mais avec des rendements de production très variables.

Pour le suivi du potentiel de liaison aux chlorophylles, une méthode précise de détermination de l'affinité a été mise au point en prenant profit de la propriété de photoprotection des Chls au sein des complexes WSCP-Chl. Les taux variables de photoblanchiment des Chls associées aux WSCP<sub>II</sub>, dépendamment des quantités en protéines, ont ainsi permis de définir des constantes d'affinités. Les aptitudes de BnD22 et de la Prot6 (WSCP<sub>IIB</sub>) à se lier aux Chls ont été validées, alors que les

WSCPIIX ciblées (Prot 3 et Prot4) ne se lient pas à la Chl ce qui conforte la perte de cette fonction dans ce groupe phylogénique. Par contre, une autre WSCPIIA, la Prot1 s'est avérée incapable de se lier à la Chl dans les conditions expérimentales testées, peut-être du fait de l'absence d'un des acides aminés définis comme primordiaux pour l'interaction entre les WSCP et la Chl (Horigome *et al.*, 2007). Il semble donc que toutes les WSCPIIA ne soient pas aptes à se lier aux chlorophylles. Ces protéines seraient, d'un point de vue évolutif, des protéines WSCPIIX en devenir, tout comme les protéines classées dans le groupe IIC (chapitre 2). Nos résultats laissent à penser que les WSCPII capables de se lier aux chlorophylles pourraient également présenter des affinités différentes aux Chls. Ainsi, nous avons montré que :

- BnD22 est la première WSCPIIA caractérisée plus affine à la Chla qu'à la Chlb.
- les affinités et la stœchiométrie en Chl au sein des complexes BnD22-Chl diffèrent selon la disponibilité en Chls.

La Prot6 a également un profil de sélectivité similaire à celui de BnD22 alors qu'elle est classée comme WSCPIIB et possède le motif de sélectivité défini par Palm *et al.*, (2018). Dans ce contexte, il s'avère indispensable de poursuivre une caractérisation fine des WSCPII pour permettre de comparer leur potentiel de liaison à la Chla et la Chlb qui pourrait refléter une diversité fonctionnelle au sein des groupes phylogéniques IIA et IIB.

Hormis pour les WSCPIIX, la propriété de liaison aux Chls des WSCPII n'est pas remise en cause. Cependant, aucune étude n'a démontré *in vivo* l'adressage de ces protéines aux chloroplastes. Les WSCPII sont présentes dans des structures spécifiques du réticulum endoplasmique, appelées ER bodies, et dans la vacuole (Takahashi *et al.*, 2012, 2013b, 2015b; Bouargalne *et al.*, 2022). Sur la base des rôles des ER bodies dans la protection cellulaire suite à des contraintes environnementales, Takahashi *et al.*, (2012, 2013b) ont émis l'hypothèse d'un adressage chloroplastique des WSCPII pour assurer une protection des Chls suite à un dommage cellulaire. En effet, il est démontré que les WSCPII ont un effet photoprotecteur envers les Chl (Agostini *et al.*, 2017) et que les complexes WSCPII-Chl sont résistants à la dénaturation par la chaleur, aux pH extrêmes et aux protéases (Schmidt *et al.*, 2003; Bektas *et al.*, 2012; Takahashi *et al.*, 2014b; Palm *et al.*, 2019). L'hypothèse d'un adressage aux chloroplastes inductible par une contrainte pourrait être testée dans des protoplastes sur-exprimant BnD22 soumis à une contrainte comme un stress lumineux ou thermique. Plusieurs auteurs suggèrent aussi que les WSCPII pourraient récupérer les molécules de Chl libérées lors de la dégradation des complexes pigment-protéines qui se produit soit pendant la sénescence foliaire, soit lors de l'adaptation de l'appareil photosynthétique en réponse à différentes contraintes environnementales (Sato *et al.*, 2001; Bektas *et al.*, 2012; Takahashi *et al.*, 2012). Cette hypothèse repose sur l'aptitude des WSCPII à extraire les Chls des membranes des thylakoïdes (Sato *et al.*, 1998; Takahashi *et al.*, 2012; Girr & Paulsen, 2021). Cette propriété est remise en question par Girr

et Paulsen (2021) qui démontrent un pouvoir d'extraction lent et inefficace. Les WSCPII ne seraient donc pas capables d'extraire directement la Chl des complexes protéines-pigments intégrées dans les membranes des thylakoïdes mais récupérerait la Chl non liée.

De récentes études ont mis en avant que les complexes WSCPII-Chl, du fait de leur stabilité, seraient à l'origine d'une source continue de ROS qui ont, en outre, un rôle signalétique dans la réponse aux stress biotiques, abiotiques ainsi que dans le processus de mort cellulaire programmée (Agostini *et al.*, 2017). De plus, les WSCPII peuvent se lier à différentes molécules tétrapyroliques comme l'hème, certains précurseurs et produits de dégradation des Chls (Schmidt *et al.*, 2003; Bektas *et al.*, 2012; Girr *et al.*, 2020). Ces molécules tétrapyroliques sont impliquées dans la signalisation plastide-noyau pour réguler leur métabolisme et induire des mécanismes de réponse au stress (Terry et Bampton, 2019). Bien que le mécanisme de signalisation par les tétrapyroles est à ce jour mal compris, il apparaît que les tétrapyroles produisent des ROS qui seraient des inhibiteurs de l'expression des gènes liés à la photosynthèse et activateur des gènes liés au stress et à la mort cellulaire (Terry & Bampton, 2019). Étant donné que les WSCPII se lient à différents tétrapyroles qui produisent des ROS lors de l'illumination, il est possible que ces protéines soient impliquées dans la signalisation plastide-noyau lors de la réponse aux stress. Cette hypothèse est renforcée par l'induction de certaines WSCPIIA, telles que BnD22, BolWSCP ou AtWSCP, par diverses contraintes environnementales (Downing *et al.*, 1992; Ilami *et al.*, 1997; Nishio & Satoh, 1997; Desclos *et al.*, 2008; Boex-Fontvieille *et al.*, 2016a,b). Dans ce cadre, l'aptitude des WSCPII de colza à se lier à d'autres tétrapyroles que les Chls et induire la production de ROS sera à étudier. Cette hypothèse permettrait également d'expliquer la présence spécifique des WSCPIIX dans les tissus non chlorophylliens (tige, racine) qui aurait une capacité à lier d'autres tétrapyroles que les Chls (Tanaka & Tanaka, 2007; Girr *et al.*, 2020) sans écarter des propriétés additionnelles en tant qu'IP.

Membres des IP de type Kunitz, les WSCPIIA peuvent présenter des propriétés additionnelles d'inhibition aux protéases. Etienne *et al.*, (2007) et Desclos *et al.*, (2008) ont mis en évidence une activité d'inhibition des activités protéolytiques de type Ser pour BnD22. Par l'approche *in silico*, les 7 protéines de colza sélectionnées, hormis la Prot6, auraient une fonction Ser-IP et seules les 3 WSCPIIA seraient des Cys-IP. Notre approche *in vitro* n'a pas permis de confirmer la fonction d'inhibition contre les Ser protéases de type trypsine ou chymotrypsine. Aucune propriété Ser-IP n'a également été démontrée *in vitro* pour AtWSCP, la seule autre WSCPII testée pour son potentiel IP (Boex-Fontvieille *et al.*, 2016a). L'absence de détection *in vitro* pour BnD22 ne remet pas en cause la possibilité d'une inhibition effective *in vivo* contre une ou plusieurs Ser protéase(s) foliaire(s). Pour les fonctions de Cys-IP, nos résultats démontrent une inhibition effective pour BnD22 et la Prot6 (WSCPIIB) mais pas pour la Prot4 appartenant aux WSCPIIX. Chez *A. thaliana*, l'unique WSCPII

est également un Cys-IP qui peut s'associer *in vivo* avec RD21 dans la plantule étiolée où elle est impliquée dans la résistance contre les herbivores au moment de l'émergence (Halls *et al.*, 2006; Boex-Fontvieille *et al.*, 2015b, 2016a). L'implication des WSCPII face à l'attaque biotique concorde avec leur lieu de stockage que sont les ER-bodies, système de défense *via* l'activité de  $\beta$ -glucosidase (Yamada *et al.*, 2011; Takahashi *et al.*, 2012). Il serait judicieux d'élargir cette étude à d'autres stress biotiques tels que la hernie (*Plasmodiophora brassicae*) et le phoma (*Leptosphaeria* et *Leptosphaeria maculans*) afin de vérifier si l'expression des copies présentes dans les racines ou les tiges diffèrent en présence des pathogènes.

RD21, cible d'AtWSCP, intervient dans de nombreux processus physiologiques, en réponse aux contraintes environnementales (déficit hydrique, stress salin), lors de la sénescence foliaire ou encore pendant l'imbibition des graines (Koizumi *et al.*, 1993; Yamada *et al.*, 2001; Kikuchi *et al.*, 2008). L'inhibition de RD21 par AtWSCP est Chl-dépendante. Ainsi lors du verdissement des plantules, AtWSCP se dissocie de RD21 pour se lier aux Chls ou à ses précurseurs (Boex-Fontvieille *et al.*, 2015a). Nos études *in vitro* ont mis en évidence une fonction simultanée de BnD22 avec une augmentation de son pouvoir d'inhibition contre la papaïne par les complexes BnD22-Chl par rapport à BnD22 non complexée aux Chls. Cette activation pourrait être considérée comme un niveau supplémentaire de la régulation et le contrôle de l'activité des protéases dans les tissus chlorophylliens ou dans l'attente de conditions plus favorables à la croissance. La possibilité d'activité IP des complexes WSCP-Chl devra être élargie à d'autres WSCPII présentant une double fonction et validée *in vivo* par des approches d'immunoprécipitation. Par ailleurs des approches par mutagenèse dirigée dans le site de fixation de protéase et/ou la Chl devraient permettre de définir précisément les acides aminés importants pour ces deux activités. De plus, pour avoir une vision plus fine, les approches complémentaires de suivi *in vitro* de la liaison WSCPII-protéase peuvent être envisagées par approche de thermophorèse. Pour se rapprocher au plus près des fonctions physiologiques, l'identification des protéases cibles *in planta* pourrait se faire par des approches complémentaires de co-immunoprécipitation et de pull-down dans les tissus d'intérêt à choisir selon le site d'expression de la WSCPII étudiée (par exemple les jeunes feuilles pour BnD22, les tiges et racines pour les WSCPIIX, les graines pour A07p38410.1...). Afin de spécifier au mieux quelle WSCPII cibler, des analyses complémentaires par approche transcriptomique (RNAseq) devront valider les données d'expressions transcriptionnelles acquises par qRT-PCR en conditions témoins pour permettre de confirmer l'identité des copies les plus exprimées et leur réactivité en réponse à une contrainte. En effet, des copies faiblement exprimées en condition témoin peuvent devenir les formes majeures en réponse à une contrainte. La présence tissulaire effective des protéines sera également à vérifier.

Une différence majeure entre *A. thaliana* et le colza est la présence importante de WSCPII dans les feuilles en développement. Sous contrainte hydrique, BnD22 représente jusqu'à 9% des protéines solubles foliaires et elle est également surexprimée suite à des carences minérales (Reviron *et al.*, 1992; Etienne *et al.*, 2007; Desclos *et al.*, 2008; D'Hooghe *et al.*, 2013). BnD22 présente un gradient d'expression antagoniste à celui des protéases foliaires suggérant son implication dans le contrôle du processus de protéolyse associé à la sénescence naturelle ou induite (Downing *et al.*, 1992; Reviron *et al.*, 1992; Desclos *et al.*, 2008). Nous pourrions alors imaginer qu'au sein des jeunes feuilles, les WSCPIIA participent à la force d'appel des organes puits par inhibition des protéases et à la protection des chlorophylles afin de maintenir l'intégrité physiologique des jeunes feuilles pour leur croissance et qu'à l'opposé, leur absence dans les feuilles sénescentes permet une protéolyse accrue et une dégradation de la chlorophylle, est en accord avec ce qui est observé dans les tissus foliaires en sénescence. Nos résultats ont clairement démontré que BnD22 est bifonctionnelle ce qui présente un réel intérêt dans la modulation de la progression de la sénescence foliaire.

Les Cys protéases comme Sag12 et RD21 et les Ser protéases sont les protéases majeures impliquées dans la sénescence foliaire chez *A. thaliana* ou le colza (Yamada *et al.*, 2001; Otegui *et al.*, 2005; Martínez *et al.*, 2008b; Girondé *et al.*, 2015b; Diaz-Mendoza *et al.*, 2016; Poret *et al.*, 2016; Havé *et al.*, 2017; James *et al.*, 2018). La fonction IP des WSCPII est en concordance avec leur localisation dans la vacuole qui est le siège principal des protéases (Otegui *et al.*, 2005; Martínez *et al.*, 2008a). Les WSCPII dans les ER-bodies (Takahashi *et al.*, 2012, 2013b, 2015b) pourraient aussi exercer une inhibition envers deux Cys protéases RD21 et VPE (Vacuolar Processing Enzymes) également présentes dans ce compartiment dérivé du RE (Hayashi *et al.*, 2001). De plus, en réponse à un stress abiotique, les ER bodies fusionnent à la vacuole permettant ainsi la libération des protéases et des IP (Hayashi *et al.*, 2001).

De nombreuses Cys et Ser protéases ont également été identifiées au sein des chloroplastes (Nishimura *et al.*, 2017). Leur inhibition par les WSCPII est moins probable car aucune étude n'indique clairement la localisation chloroplastique des WSCPII. Cependant, la question de savoir comment les WSCPII sont exposées *in vivo* à la Chl et aux protéases reste encore inexplicée. La localisation non chloroplastique des WSCPII conduit à émettre des hypothèses d'une interaction entre WSCP et Chl/protéases en dehors des chloroplastes. En effet, il a été démontré récemment que les Senescence-Associated Vacuoles (SAVs), petites vacuoles spécifiquement induites lors de la sénescence, contiennent en plus des Cys protéases comme SAG12, de la Chl avec un rapport Chl a/b proche de 5 (Martínez *et al.*, 2008a; Gomez *et al.*, 2019). D'autres vésicules nommées CCVs (Chloroplast-Containing Vesicles) contenant des protéines chloroplastiques (stroma + membranes thylakoïdiennes) et des produits de dégradation des Chls ont été également mis en évidence chez *A*



*thaliana* (Wang & Blumwald, 2014; Buet *et al.*, 2019; Gomez *et al.*, 2019). La composition de ces vésicules est encore relativement peu étudiée et leur rôle physiologique encore mal élucidé. Bien que les WSCPII n'aient pas été identifiées dans ces vésicules, leur interaction avec les Chl et/ou les protéases pourrait être envisagée. Une autre piste de réflexion concerne le rôle des WSCPII pendant la biogénèse des chloroplastes. En effet, les WSCPIIA assurent un rôle protecteur envers les précurseurs de la Chl que sont la protochlorophyllide et la chlorophyllide (Schmidt *et al.*, 2003). Ces précurseurs de la Chl sont accumulés dans les étiooplastes des plantules étiolées et la présence des WSCPIIA pourrait être bénéfique afin d'optimiser leur conversion en Chl pendant la biogénèse des chloroplastes après passage à la lumière. Cette fonction de protection pourrait être retrouvée pendant la morphogénèse des feuilles avec les chloroplastes formés à partir de proplastides indifférenciés des tissus du méristème apical (Solymosi & Schoefs, 2010; Putarjunan *et al.*, 2013). La dégradation des protéines chloroplastiques représente la source majoritaire de l'azote. Au stade reproducteur, les siliques, par leur activité photosynthétique et la sénescence des péricardes, vont également jouer un rôle essentiel dans l'approvisionnement en éléments chez des graines en développement (Liu *et al.*, 2015). De nombreuses protéases ont été identifiées dans les siliques de colza et certaines WSCPII A07p38410.1 ou BnD22 sont exprimées majoritairement dans les siliques ou les graines. Le suivi des expressions des WSCPIIA dans les siliques et les graines au cours de leur développement demande à être approfondi. S'il est avéré une abondance de certaines WSCPII dans les siliques et les graines, leurs rôles physiologiques devront être étudiés.

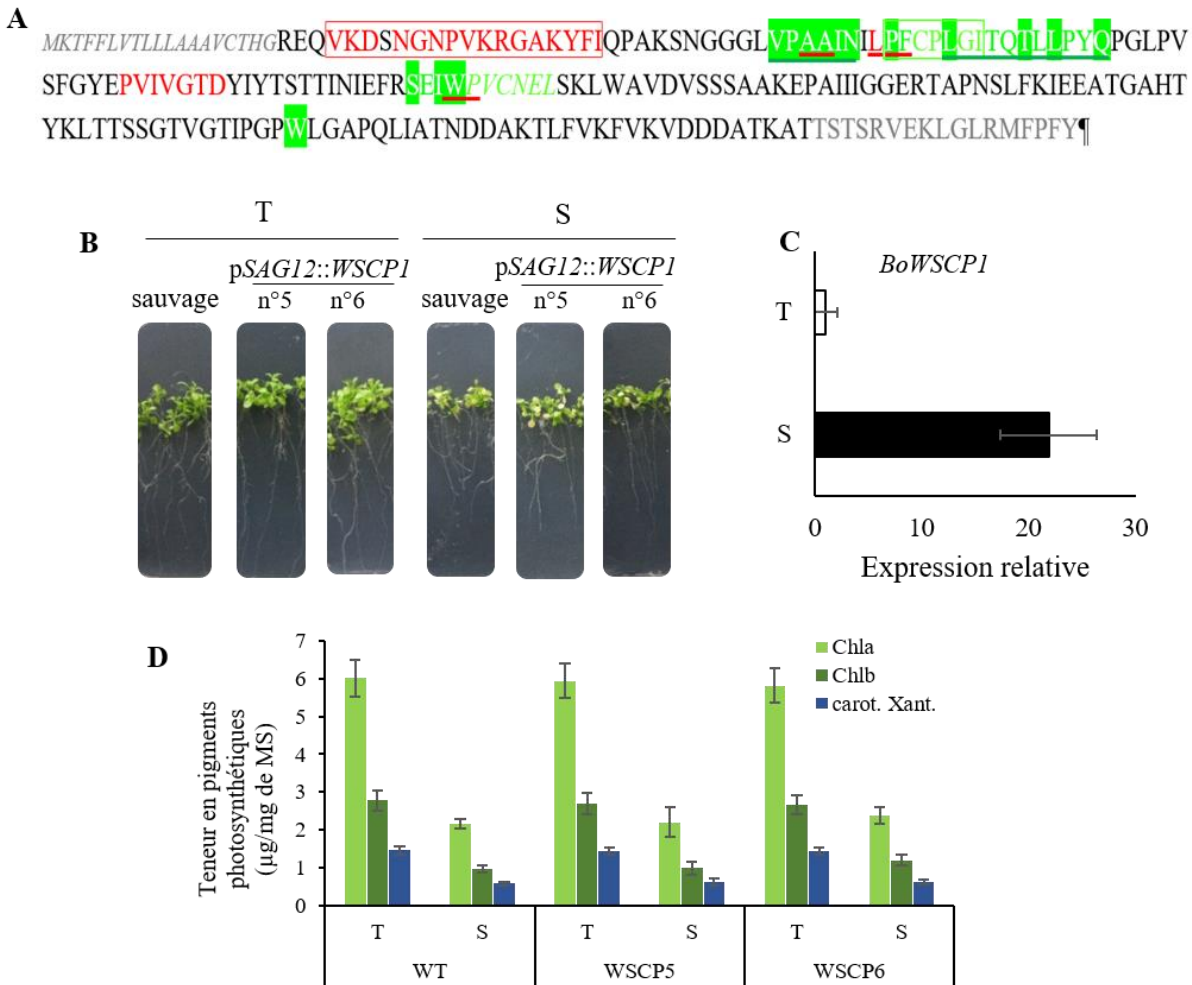
Peu d'études par génomique fonctionnelle ont été entreprises pour comprendre le(s) rôle(s) physiologique(s) des WSCPII. Plusieurs possibilités s'offrent :

- Les lignées de tabac sur-exprimant BoWSCP (Damaraju *et al.*, 2011) pourraient être exploitées pour suivre la sénescence foliaire naturelle ou induite en réponse à des stress environnementaux.

- La complémentation des mutants *Atwscp* knockout par des gènes de colza permettrait de déterminer si les WSCPII de colza exprimés au stade plantule (C06p46430.1 et BnD22) ou dans les fleurs (WSCPIIA : BnD22 et A07p38410.1=Prot2 ; WSCPIIX : A07p38420.1=prot4 ou son homologue, A07p38440.1 ou son homologue), peuvent assurer des fonctions similaires à celles décrites pour AtWSCP.

- La surexpression de WSCPIIA sous contrôle de promoteur spécifique à l'expression foliaire pourrait être également envisagée dans les lignées d'*A. thaliana* déficientes en AtWSCP pour assurer une protection maximale des feuilles en croissance. A l'inverse nous pourrions nous attendre à ce que la surexpression de WSCPIIA dans les feuilles sénescentes induise un retard de sénescence. Cette stratégie a été choisie par James et collaborateurs (James, 2018) par surexpression de BoWSCP1 sous contrôle du promoteur Sag12 dans *A. thaliana*. Les résultats ont montré qu'au stade végétatif, les lignées *pSAG12::WSCP1* présentent un ralentissement de la dégradation de la chlorophylle après





**Fig. 1 : Phénotypage des plantules *pSAG12::WSCP1* en présence de mannitol.**

**A.** Séquence protéique de *BoWSCP1* (BAA32820).

Le motifs Kunitz et de liaison à la Chl sont encadrés en rouge ou vert (Sato *et al.*, 2001). Les acides aminés soulignés en rouge sont ceux identifiés dans l'interaction *AtWSCP-RD21* (Boex-Fontvieille *et al.*, 2015a). Les acides aminés en fond vert sont ceux identifiés importants dans l'interaction avec la Chl avec en italique le motif d'affinité à la Chla (Horigone *et al.*, 2008; Palm *et al.*, 2018)

**B.** Photographies des plantules sauvages ou *pSAG12::WSCP1* (lignées 5 et 6) cultivées en condition témoin (T) ou de stress (S) par addition de mannitol 225 mM.

**C.** Expression de *BoWSCP1* dans la partie aérienne des plantules *pSAG12::WSCP1*

**D.** Teneurs en Chla, Chlb et caroténoïdes (carot.) et xanthophylle (xant.) des plantules sauvages ou *pSAG12::WSCP1* (lignées 5 et 6)

Les graines d'*A. thaliana RIL232* sont semées sur milieu agarose (6g/L agar) Murashige et Skoog (MS). Après 2-3 jours à 4°C dans l'obscurité pour synchroniser la germination, les boîtes de cultures sont placées dans une salle de culture à 22°C avec une période de lumière de 12 h (intensité lumineuse de 100 µmol.m<sup>-2</sup>.s<sup>-1</sup>) et une humidité relative de 70%. Après 10 jours de développement, les plantules sont transférées ou non sur milieu enrichi en mannitol 225 mM pour 4 jours. La partie aérienne est prélevée 7 h après le début de la lumière afin de prendre en compte les rythmes circadiens des plantules.

induction de la sénescence foliaire par l'obscurité. Afin d'approfondir cette étude, des expériences préliminaires ont été menées dans l'équipe de recherche sur des plantules de la lignée *pSAG12::WSCPI* soumises à un stress osmotique par du mannitol afin d'induire le promoteur *Sag12* (Fig. 1), mais aucun phénotype majeur n'a été observé en terme de développement et de teneurs en pigments des plantules. Le profil protéolytique de ces lignées sera entrepris ainsi que l'étude des effets de la sénescence induite par l'obscurité au stade plantule et du stress hydrique au stade végétatif.

Pour conclure, ce travail a permis d'identifier l'existence d'une famille multigénique WSCPII chez le colza. Un processus de sous-fonctionnalisation a été mis en évidence conduisant, pour certaines copies, à une expression tissulaire spécifique dans des tissus chlorophylliens ou non. De plus, ces protéines sont caractérisées par une diversité fonctionnelle avec des protéines aptes ou non à se lier à la chlorophylle, capables ou non d'agir en tant qu'inhibiteurs de protéases. Nos travaux ont ainsi permis de mieux cerner les interactions spécifiques entre les WSCPII, Chl et protéase. L'étude approfondie des propriétés fonctionnelles de BnD22 a démontré une double fonction pour cette WSCPIIA, ce qui suggère sa contribution à réguler la protéolyse cellulaire et/ou maintenir l'activité photosynthétique et pourrait permettre d'améliorer l'efficacité de remobilisation de l'azote pendant la sénescence foliaire. Les analyses fonctionnelles sur les autres WSCPII du colza doivent être poursuivies ainsi que l'analyse de la plasticité d'expression des WSCPII face à diverses contraintes biotiques et abiotiques. L'identification *in planta* des protéases partenaires des WSCPII pourraient permettre de conclure sur leur rôle dans régulation de protéolyse. Pour finir, vue les avancées technologiques chez le colza de l'approche CRISPR-cas9 pour éteindre spécifiquement plusieurs gènes cibles, cette approche pourrait être envisagée pour la famille des WSCPII d'autant plus que des duplications en tandem sont concentrées sur des portions restreintes du génome.

Finalement, l'ensemble de ce travail de thèse a permis d'apporter de nouvelles connaissances sur ces protéines particulières que sont les WSCP et plus spécifiquement les WSCPII du colza. Ils contribueront à l'avenir à faire des choix quant aux protéines qui seront jugées les plus pertinentes en tant que candidates potentielles pour améliorer la résistance du colza en réponse à des contraintes environnementales, qui est l'un des enjeux majeurs des prochaines décennies. Il est plus que probable que les WSCPII ne pourront pas résoudre à elles seules les problèmes liés au stress biotiques ou abiotiques mais tout au plus elles apporteront leur contribution à l'obtention de plantes génétiquement plus performantes dans leur aptitude à supporter les stress environnementaux.



## REFERENCES BIBLIOGRAPHIQUES

---

- Agostini A, Palm DM, Paulsen H, Carbonera D. 2018.** Optically Detected Magnetic Resonance of Chlorophyll Triplet States in Water-Soluble Chlorophyll Proteins from *Lepidium virginicum*: Evidence for Excitonic Interaction among the Four Pigments. *Journal of Physical Chemistry B* **122**: 6156–6163.
- Agostini A, Palm DM, Schmitt FJ, Albertini M, Valentin M Di, Paulsen H, Carbonera D. 2017.** An unusual role for the phytyl chains in the photoprotection of the chlorophylls bound to Water-Soluble Chlorophyll-binding Proteins. *Scientific Reports* **7**: 1–13.
- Ahn YO, Shimizu BI, Sakata K, Gantulga D, Zhou Z, Bevan DR, Esen A. 2010.** Scopolin-hydrolyzing  $\beta$ -glucosidases in roots of Arabidopsis. *Plant and Cell Physiology* **51**: 132–143.
- Avice JC, Etienne P. 2014.** Leaf senescence and nitrogen remobilization efficiency in oilseed rape (*Brassica napus* L.). *Journal of Experimental Botany* **65**: 3813–3824.
- Bednarczyk D, Dym O, Prabakar V, Peleg Y, Pike DH, Noy D. 2016.** Fine Tuning of Chlorophyll Spectra by Protein-Induced Ring Deformation. *Angewandte Chemie - International Edition* **55**: 6901–6905.
- Bednarczyk D, Takahashi S, Satoh H, Noy D. 2015.** Assembly of water-soluble chlorophyll-binding proteins with native hydrophobic chlorophylls in water-in-oil emulsions. *Biochimica et Biophysica Acta - Bioenergetics* **1847**: 307–313.
- Bektas I, Fellenberg C, Paulsen H. 2012.** Water-soluble chlorophyll protein (WSCP) of Arabidopsis is expressed in the gynoceium and developing silique. *Planta* **236**: 251–259.
- Boex-Fontvieille E, Reinbothe S, Reinbothe C, Rustgi S, Von Wettstein D, Pollmann S. 2016a.** An ethylene-protected achilles' heel of etiolated seedlings for arthropod deterrence. *Frontiers in Plant Science* **7**: 1–14.
- Boex-Fontvieille E, Rustgi S, Reinbothe S, Reinbothe C. 2015a.** A Kunitz-type protease inhibitor regulates programmed cell death during fower development in *Arabidopsis thaliana*. *Journal of Experimental Botany* **66**: 6119–6135.
- Boex-Fontvieille E, Rustgi S, Von Wettstein D, Pollmann S, Reinbothe S, Reinbothe C. 2016b.** Jasmonic acid protects etiolated seedlings of *Arabidopsis thaliana* against herbivorous arthropods. *Plant Signaling and Behavior* **11**: 1–7.
- Boex-Fontvieille E, Rustgi S, von Wettstein D, Reinbothe S, Reinbothe C. 2015b.** Water-soluble chlorophyll protein is involved in herbivore resistance activation during greening of *Arabidopsis thaliana* . *Proceedings of the National Academy of Sciences* **112**: 7303–7308.
- Bouargalne Y, Raguénès-Nicol C, Guilbaud F, Cheron A, Clouet V, Deleu C, Le Cahérec F. 2022.** New insights into chlorophyll-WSCP (water-soluble chlorophyll binding proteins) interactions the case study of BnD22 (*Brassica napus* drought-induced 22 kDa). *Plant Physiology and Biochemistry*.
- Buet A, Costa ML, Martínez DE, Guiamet JJ. 2019.** Chloroplast protein degradation in senescing leaves: Proteases and lytic compartments. *Frontiers in Plant Science* **10**.
- D'Hooghe P, Escamez S, Trouverie J, Avice JC. 2013.** Sulphur limitation provokes physiological and leaf proteome changes in oilseed rape that lead to perturbation of sulphur, carbon and oxidative metabolisms. *BMC Plant Biology* **13**: 1.
- Damaraju S, Schlede S, Eckhardt U, Lokstein H, Grimm B. 2011.** Functions of the water soluble chlorophyll-binding protein in plants. *Journal of Plant Physiology* **168**: 1444–1451.
- Desclos M, Dubouset L, Etienne P, Le Caherec F, Satoh H, Bonnefoy J, Ourry A, Avice J-C. 2008.** A Proteomic Profiling Approach to Reveal a Novel Role of *Brassica napus* Drought 22 kD/Water-Soluble Chlorophyll-Binding Protein in Young Leaves during Nitrogen Remobilization Induced by Stressful Conditions . *Plant Physiology* **147**: 1830–1844.
- Diaz-Mendoza M, Velasco-Arroyo B, Santamaria ME, González-Melendi P, Martinez M, Diaz I. 2016.** Plant senescence and proteolysis: Two processes with one destiny. *Genetics and Molecular Biology* **39**: 329–338.

- Downing WL, Mauxion F, Fauvarque M -O, Reviron M -P, De Vienne D, Vartanian N, Giraudat J. 1992.** A *Brassica napus* transcript encoding a protein related to the Kunitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. *The Plant Journal* **2**: 685–693.
- Etienne P, Desclos M, Le Gou L, Gombert J, Bonnefoy J, Maurel K, Le Dily F, Ourry A, Avice JC. 2007.** N-protein mobilisation associated with the leaf senescence process in oilseed rape is concomitant with the disappearance of trypsin inhibitor activity. *Functional Plant Biology* **34**: 895–906.
- Gholizadeh A. 2018.** Chlorophyll Binding Ability of Non-chloroplastic DUF538 Protein Superfamily in Plants. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences* **88**: 967–976.
- Girondé A, Etienne P, Trouverie J, Bouchereau A, Le Cahérec F, Leport L, Orsel M, Niogret MF, Nesi N, Carole D, et al. 2015a.** The contrasting N management of two oilseed rape genotypes reveals the mechanisms of proteolysis associated with leaf N remobilization and the respective contributions of leaves and stems to N storage and remobilization during seed filling. *BMC Plant Biology* **15**.
- Girondé A, Poret M, Etienne P, Trouverie J, Bouchereau A, Le Cahérec F, Leport L, Niogret M-F, Avice J-C. 2015b.** A Comparative Study of Proteolytic Mechanisms during Leaf Senescence of Four Genotypes of Winter Oilseed Rape Highlighted Relevant Physiological and Molecular Traits for NRE Improvement. *Plants*.
- Girr P, Kilper J, Pohland AC, Paulsen H. 2020.** The pigment binding behaviour of water-soluble chlorophyll protein (WSCP). *Photochemical and Photobiological Sciences* **19**: 695–712.
- Girr P, Paulsen H. 2021.** How water-soluble chlorophyll protein extracts chlorophyll from membranes. *Biochimica et Biophysica Acta - Biomembranes* **1863**: 183479.
- Goedheer JC, Vos M. 1977.** On the origin of the photoconvertible chlorophyll-protein complex Cp668→Cp743 in chenopodium and amaranthus species\*. *Acta Botanica Neerlandica* **26**: 289–298.
- Gombert J, Etienne P, Ourry A, Le Dily F. 2006.** The expression patterns of SAG12/Cab genes reveal the spatial and temporal progression of leaf senescence in *Brassica napus* L. with sensitivity to the environment. *Journal of Experimental Botany* **57**: 1949–1956.
- Gomez FM, Carrión CA, Costa ML, Desel C, Kieselbach T, Funk C, Krupinska K, Guamet J. 2019.** Extra-plastidial degradation of chlorophyll and photosystem I in tobacco leaves involving ‘senescence-associated vacuoles’. *Plant Journal*: 1–13.
- Gotté M, Ghosh R, Bernard S, Nguema-Ona E, Vicré-Gibouin M, Hara-Nishimura I, Driouich A. 2015.** Methyl jasmonate affects morphology, number and activity of endoplasmic reticulum bodies in raphanus sativus root cells. *Plant and Cell Physiology* **56**: 61–72.
- Green BR, Kühlbrandt W. 1995.** Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosynthesis Research* **44**: 139–148.
- Halls CE, Rogers SW, Oufattole M, Østergard O, Svensson B, Rogers JC. 2006.** A Kunitz-type cysteine protease inhibitor from cauliflower and *Arabidopsis*. *Plant Science* **170**: 1102–1110.
- Havé M, Marmagne A, Chardon F, Masclaux-Daubresse C. 2017.** Nitrogen remobilization during leaf senescence: Lessons from *Arabidopsis* to crops. *Journal of Experimental Botany* **68**: 2513–2529.
- Hayashi Y, Nishimura M, Yamada K, Shimada T, Matsushima R, Hara-Nishimura I, Nishizawa NK. 2001.** A proteinase-storing body that prepares for cell death or stresses in the epidermal cells of *Arabidopsis*. *Plant and Cell Physiology* **42**: 894–899.
- Hirabayashi H, Amakawa M, Kamimura Y, Shino Y, Satoh H, Itoh S, Tamiaki H. 2006.** Analysis of photooxidized pigments in water-soluble chlorophyll protein complex isolated from *Chenopodium album*. *Journal of Photochemistry and Photobiology A: Chemistry*.
- Horigome D, Satoh H, Itoh N, Mitsunaga K, Oonishi I, Nakagawa A, Uchida A. 2007.** Structural mechanism and photoprotective function of water-soluble chlorophyll-binding protein. *Journal of Biological Chemistry* **282**: 6525–6531.
- Horigome D, Satoh H, Uchida A. 2003.** Purification, crystallization and preliminary X-ray analysis of a water-soluble chlorophyll protein from *Brassica oleracea* L. var. *acephala* (kale). *Acta Crystallographica - Section D Biological Crystallography* **59**: 2283–2285.

- Hörtensteiner S, Kräutler B. 2011.** Chlorophyll breakdown in higher plants. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1807**: 977–988.
- Hughes JL, Razeghifard R, Logue M, Oakley A, Wydrzynski T, Krausz E. 2006.** Magneto-optic spectroscopy of a protein tetramer binding two exciton-coupled chlorophylls. *Journal of the American Chemical Society* **128**: 3649–3658.
- Ilami G, Nespoulous C, Huet JC, Vartanian N, Pernollet JC. 1997.** Characterization of BnD22, a drought-induced protein expressed in *Brassica napus* leaves. *Phytochemistry* **45**: 1–8.
- James M. 2018.** Etude de systèmes protéolytiques et anti-protéolytiques impliqués dans la remobilisation de l'azote au cours de la sénescence chez les Brassicacées (*Arabidopsis thaliana* L., *Brassica napus* L). Thèse de l'Université de Caen Normandie au sein de l'École Doctorale Normande de Biologie Intégrative, Santé, Environnement (Spécialité : Sciences agronomiques, Biotechnologies agro-alimentaires).
- James M, Masclaux-Daubresse C, Marmagne A, Azzopardi M, Laîné P, Goux D, Etienne P, Trouverie J. 2019.** A New Role for SAG12 Cysteine Protease in Roots of *Arabidopsis thaliana*. *Frontiers in Plant Science* **9**: 1–11.
- James M, Poret M, Masclaux-Daubresse C, Marmagne A, Coquet L, Jouenne T, Chan P, Trouverie J, Etienne P. 2018.** SAG12, a major cysteine protease involved in nitrogen allocation during senescence for seed production in *Arabidopsis thaliana*. *Plant and Cell Physiology* **59**: 2052–2063.
- Jerabek-Willemsen M, André T, Wanner R, Roth HM, Duhr S, Baaske P, Breitsprecher D. 2014.** MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure* **1077**: 101–113.
- Jmel MA, Aounallah H, Bensaoud C, Mekki I, Chmelař J, Faria F, M'ghirbi Y, Kotsyfakis M. 2021.** Insights into the role of tick salivary protease inhibitors during ectoparasite-host crosstalk. *International Journal of Molecular Sciences* **22**: 1–23.
- Kamimura Y, Mori T, Yamasaki T, Katoh S. 1997.** Isolation, properties and a possible function of a water-soluble chlorophyll a/b-protein from Brussels sprouts. *Plant and Cell Physiology* **38**: 133–138.
- Khamrui S, Dasgupta J, Dattagupta JK, Sen U. 2005.** Single mutation at P1 of a chymotrypsin inhibitor changes it to a trypsin inhibitor: X-ray structural (2.15 Å) and biochemical basis. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1752**: 65–72.
- Kikuchi Y, Saika H, Yuasa K, Nagahama M, Tsuji A. 2008.** Isolation and biochemical characterization of two forms of RD21 from cotyledons of daikon radish (*Raphanus sativus*). *Journal of Biochemistry* **144**: 789–798.
- Koeslin-Findeklee F, Rizi VS, Becker MA, Parra-Londono S, Arif M, Balazadeh S, Mueller-Roeber B, Kunze R, Horst WJ. 2015.** Transcriptomic analysis of nitrogen starvation- and cultivar-specific leaf senescence in winter oilseed rape (*Brassica napus* L.). *Plant Science* **233**: 174–185.
- Koizumi M, Yamaguchi-Shinozaki K, Tsuji H, Shinozaki K. 1993.** Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*. *Gene* **129**: 175–182.
- Krowarsch D, Cierpicki T, Jelen F, Otlewski J. 2003.** Canonical protein inhibitors of serine proteases. *Cellular and Molecular Life Sciences* **60**: 2427–2444.
- Kuai B, Chen J, Hörtensteiner S. 2018.** The biochemistry and molecular biology of chlorophyll breakdown. *Journal of Experimental Botany* **69**: 751–767.
- Laemmli UK. 1970.** Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680–685.
- Laloi C, Havaux M. 2015.** Key players of singlet oxygen-induced cell death in plants. *Frontiers in Plant Science* **6**.
- Laskowski Jr. M, Kato I. 1969.** Protein Inhibitors. *Nature* **223**: 426–426.
- Laskowski JM, Kato I. 1980.** Protein inhibitors of proteinases. *Annual Review of Biochemistry* **49**: 593–626.
- Laskowski M, Qasim MA. 2000.** What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology* **1477**: 324–337.

- De Leo FD. 2002.** PLANT-PIs: a database for plant protease inhibitors and their genes. *Nucleic Acids Research* **30**: 347–348.
- Lin YP, Wu MC, Charng YY. 2016.** Identification of a Chlorophyll Dephytylase Involved in Chlorophyll Turnover in Arabidopsis. *The Plant Cell* **28**: 2974–2990.
- Liu H, Yang Q, Fan C, Zhao X, Wang X, Zhou Y. 2015.** Transcriptomic basis of functional difference and coordination between seeds and the silique wall of *Brassica napus* during the seed-filling stage. *Plant Science* **233**: 186–199.
- Malagoli P, Laine P, Rossato L, Ourry A. 2005.** Dynamics of nitrogen uptake and mobilization in field-grown winter oilseed rape (*Brassica napus*) from stem extension to harvest: I. Global N flows between vegetative and reproductive tissues in relation to leaf fall and their residual N. *Annals of Botany* **95**: 853–861.
- Maleeva Y V., Neverov K V., Obukhov YN, Kritsky MS. 2019.** Water Soluble Chlorophyll-Binding Proteins of Plants: Structure, Properties and Functions. *Molecular Biology* **53**: 876–888.
- Martínez DE, Costa ML, Gomez FM, Otegui MS, Guamet JJ. 2008a.** ‘Senescence-associated vacuoles’ are involved in the degradation of chloroplast proteins in tobacco leaves. *Plant Journal* **56**: 196–206.
- Martínez DE, Costa ML, Guamet JJ. 2008b.** Senescence-associated degradation of chloroplast proteins inside and outside the organelle. *Plant Biology* **10**: 15–22.
- Matsushima R, Hayashi Y, Kondo M, Shimada T, Nishimura M, Hara-Nishimura I. 2002.** An Endoplasmic Reticulum-Derived Structure That Is Induced under Stress Conditions in Arabidopsis. *Plant Physiology* **130**: 1807–1814.
- De Meester P, Brick P, Lloyd LF, Blow DM, Onesti S. 1998.** Structure of the Kunitz-type soybean trypsin inhibitor (STI): Implication for the interactions between members of the STI family and tissue-plasminogen activator. *Acta Crystallographica Section D: Biological Crystallography* **54**: 589–597.
- Momonoki YS, Sugawara M, Watanabe T. 2002.** Change in activity of soybean trypsin inhibitor by removal of C-terminal amino acid residues during seed germination. *Plant Production Science* **5**: 51–57.
- Mosolov V V., Valueva TA. 2005.** Proteinase inhibitors and their function in plants: A review. *Applied Biochemistry and Microbiology* **41**: 227–246.
- Murata T, Ishikawa C. 1981.** Chemical, physicochemical and spectrophotometric properties of crystalline chlorophyll-protein complexes from *Lepidium virginicum* L. *Biochimica et Biophysica Acta* **635**: 341–347.
- Murata T, Toda F, Uchino K, Yakushiji E. 1971.** water-soluble chlorophyll protein of *brassica oleracea* var. *botrys* (cauliflower). *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **245**: 208–215.
- Nagano AJ, Matsushima R, Hara-Nishimura I. 2005.** Activation of an ER-body-localized  $\beta$ -glucosidase via a cytosolic binding partner in damaged tissues of *Arabidopsis thaliana*. *Plant and Cell Physiology* **46**: 1140–1148.
- Nishimura K, Kato Y, Sakamoto W. 2017.** Essentials of Proteolytic Machineries in Chloroplasts. *Molecular Plant* **10**: 4–19.
- Nishio N, Satoh H. 1997.** A Water-Soluble Chlorophyll Protein in Cauliflower May Be Identical to BnD22, a Drought-Induced, 22-Kilodalton Protein in Rapeseed. *Plant Physiology* **115**: 841–846.
- Ogasawara K, Yamada K, Christeller JT, Kondo M, Hatsugai N, Hara-Nishimura I, Nishimura M. 2009.** Constitutive and inducible ER bodies of *Arabidopsis thaliana* accumulate distinct  $\beta$ -glucosidases. *Plant and Cell Physiology* **50**: 480–488.
- Ohtsuki T, Ohshima S, Uchida A. 2007.** Purification, crystallization and preliminary X-ray diffraction analysis of water-soluble chlorophyll-binding protein from *Chenopodium album*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*.
- Otegui MS, Noh YS, Martínez DE, Vila Petroff MG, Staehelin LA, Amasino RM, Guamet JJ. 2005.** Senescence-associated vacuoles with intense proteolytic activity develop in leaves of Arabidopsis and soybean. *Plant Journal* **41**: 831–844.
- Otlewski J, Jaskólski M, Buczek O, Cierpicki T. 2001.** Structure-function relationship of serine protease – pro tein in -. *Acta Biochimica Polonica* **48**: 419–428.

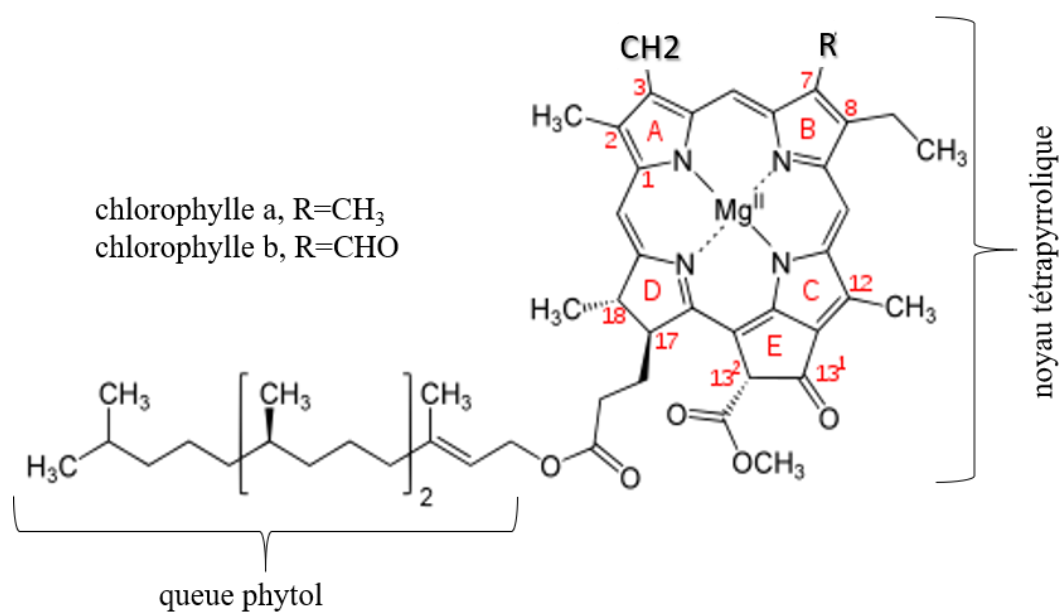
- Palm DM, Agostini A, Aversch V, Girr P, Werwie M, Takahashi S, Satoh H, Jaenicke E, Paulsen H. 2018.** Chlorophyll a/b binding-specificity in water-soluble chlorophyll protein. *Nature Plants* **4**: 920–929.
- Palm DM, Agostini A, Pohland AC, Werwie M, Jaenicke E, Paulsen H. 2019.** Stability of Water-Soluble Chlorophyll Protein (WSCP) Depends on Phytol Conformation. *ACS Omega* **4**: 7971–7979.
- Palm DM, Agostini A, Tenzer S, Gloeckle BM, Werwie M, Carbonera D, Paulsen H. 2017.** Water-Soluble Chlorophyll Protein (WSCP) Stably Binds Two or Four Chlorophylls. *Biochemistry* **56**: 1726–1736.
- Poret M, Chandrasekar B, van der Hoorn RAL, Avice JC. 2016.** Characterization of senescence-associated protease activities involved in the efficient protein remobilization during leaf senescence of winter oilseed rape. *Plant Science*.
- Prabahar V, Afriat-Jurnou L, Paluy I, Peleg Y, Noy D. 2020.** New homologues of Brassicaceae water-soluble chlorophyll proteins shed light on chlorophyll binding, spectral tuning, and molecular evolution. *FEBS Journal* **287**: 991–1004.
- Putarjunan A, Liu X, Nolan T, Yu F, Rodermeil S. 2013.** Understanding chloroplast biogenesis using second-site suppressors of *immutans* and *var2*. *Photosynthesis Research* **116**: 437–453.
- Rawlings ND, Tolle DP, Barrett AJ. 2004.** Evolutionary families of peptidase inhibitors. *Biochemical Journal* **378**: 705–716.
- Reinbothe C, Satoh H, Alcaraz JP, Reinbothe S. 2004.** A Novel Role of Water-Soluble Chlorophyll Proteins in the Transitory Storage of Chlorophyllide. *Plant Physiology* **134**: 1355–1365.
- Reviron M-P, Vartanian N, Sallantin M, Huet J-C, Pernollet J-C, de Vienne D. 1992.** Characterization of a Novel Protein Induced by Progressive or Rapid Drought and Salinity in *Brassica napus* Leaves. *Plant Physiology* **100**: 1486–1493.
- Roberts IN, Caputo C, Criado MV, Funk C. 2012.** Senescence-associated proteases in plants. *Physiologia Plantarum* **145**: 130–139.
- Rustgi S, Boex-Fontvieille E, Reinbothe C, von Wettstein D, Reinbothe S. 2017.** Serpin1 and WSCP differentially regulate the activity of the cysteine protease RD21 during plant development in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **114**: 2212–2217.
- Rustgi S, Boex-Fontvieille E, Reinbothe C, von Wettstein D, Reinbothe S. 2018.** The complex world of plant protease inhibitors: Insights into a Kunitz-type cysteine protease inhibitor of *Arabidopsis thaliana*. *Communicative and Integrative Biology* **11**: 1–8.
- Sakuraba Y, Schelbert S, Park SY, Han SH, Lee BD, Andrés CB, Kessler F, Hörtensteiner S, Paek NC. 2012.** STAY-GREEN and Chlorophyll Catabolic Enzymes Interact at Light-Harvesting Complex II for Chlorophyll Detoxification during Leaf Senescence in *Arabidopsis*. *The Plant Cell* **24**: 507–518.
- Satoh H, Nakayama K, Okada M. 1998.** Molecular cloning and functional expression of a water-soluble chlorophyll protein, a putative carrier of chlorophyll molecules in cauliflower. *Journal of Biological Chemistry* **273**: 30568–30575.
- Satoh H, Uchida A, Nakayama K, Okada M. 2001.** Water-soluble chlorophyll protein in Brassicaceae plants is a stress-induced chlorophyll-binding protein. *Plant and Cell Physiology* **42**: 906–911.
- Schmidt K, Fufezan C, Krieger-Liszkay A, Satoh H, Paulsen H. 2003.** Recombinant water-soluble chlorophyll protein from *Brassica oleracea* var. *Botrys* binds various chlorophyll derivatives. *Biochemistry* **42**: 7427–7433.
- Shinashi K, Satoh H, Uchida A, Nakayama K, Okada M, Oonishi I. 2000.** Molecular characterization of a water-soluble chlorophyll protein from main veins of Japanese radish. *Journal of Plant Physiology* **157**: 255–262.
- Solymsi K, Schoefs B. 2010.** Etioplast and etio-chloroplast formation under natural conditions: The dark side of chlorophyll biosynthesis in angiosperms. *Photosynthesis Research* **105**: 143–166.
- Song HK, Suh SW. 1998.** Kunitz-type soybean trypsin inhibitor revisited: Refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator. *Journal of Molecular Biology* **275**: 347–363.



- Takahashi S, Abe E, Nakayama K, Satoh H. 2015a.** Identification of genes encoding photoconvertible (Class I) water-soluble chlorophyll-binding proteins from *Chenopodium ficifolium*. *Bioscience, Biotechnology and Biochemistry*.
- Takahashi S, Aizawa K, Nakayama K, Satoh H. 2015b.** Water-soluble chlorophyll-binding proteins from *Arabidopsis thaliana* and *Raphanus sativus* target the endoplasmic reticulum body. *BMC Research Notes* **8**: 1–5.
- Takahashi S, Ono M, Uchida A, Nakayama K, Satoh H. 2013a.** Molecular cloning and functional expression of a water-soluble chlorophyll-binding protein from Japanese wild radish. *Journal of Plant Physiology* **170**: 406–412.
- Takahashi S, Seki Y, Uchida A, Nakayama K, Satoh H. 2014a.** Cysteine-2 and Cys30 are essential for chlorophyll-binding activity of the water-soluble chlorophyll-binding protein (WSCP) of *Chenopodium album*. *Bioscience, Biotechnology and Biochemistry* **78**: 1825–1832.
- Takahashi S, Uchida A, Nakayama K, Satoh H. 2014b.** The C-terminal extension peptide of non-photoconvertible water-soluble chlorophyll-binding proteins (Class II WSCPs) affects their solubility and stability: Comparative analyses of the biochemical and chlorophyll-binding properties of recombinant brassica. *Protein Journal* **33**: 75–84.
- Takahashi S, Yanai H, Nakamaru Y, Uchida A, Nakayama K, Satoh H. 2012.** Molecular cloning, characterization and analysis of the intracellular localization of a water-soluble chl-binding protein from brussels sprouts (*brassica oleracea* var. *gemmifera*). *Plant and Cell Physiology* **53**: 879–891.
- Takahashi S, Yanai H, Oka-Takayama Y, Zanma-Sohtome A, Fujiyama K, Uchida A, Nakayama K, Satoh H. 2013b.** Molecular cloning, characterization and analysis of the intracellular localization of a water-soluble chlorophyll-binding protein (WSCP) from Virginia pepperweed (*Lepidium virginicum*), a unique WSCP that preferentially binds chlorophyll b in vitro. *Planta* **238**: 1065–1080.
- Takahashi S, Yoshikawa M, Kamada A, Ohtsuki T, Uchida A, Nakayama K, Satoh H. 2013c.** The photoconvertible water-soluble chlorophyll-binding protein of *Chenopodium album* is a member of DUF538, a superfamily that distributes in Embryophyta. *Journal of Plant Physiology* **170**: 1549–1552.
- Takamiya A. 1973.** Distribution of photoconvertible, water-soluble chlorophyll protein complex CP668 in plants related to *Chenopodium album*. *Carnegie Inst. Wash. Year Book* **72**: 330–336.
- Tanaka R, Tanaka A. 2007.** Tetrapyrrole Biosynthesis in Higher Plants. *The Annual Review of Plant Biology* **58**: 321–346.
- Terry MJ, Bampton J. 2019.** The role of tetrapyrroles in chloroplast-to-nucleus retrograde signaling. In *Advances in Botanical Research* **91**: 225–246. Academic Press. Elsevier Ltd.
- Theiss C, Trostmann I, Andree S, Schmitt FJ, Renger T, Eichler HJ, Paulsen H, Renger G. 2007.** Pigment - Pigment and pigment - Protein interactions in recombinant Water-Soluble Chlorophyll Proteins (WSCP) from cauliflower. *Journal of Physical Chemistry B* **111**: 13325–13335.
- De Vonis Bidlingmeyer U, Leary TR, Laskowski M. 1972.** Identity of the Tryptic and  $\alpha$ -Chymotryptic Reactive Sites on Soybean Trypsin Inhibitor (Kunitz). *Biochemistry* **11**: 3303–3310.
- Wang S, Blumwald E. 2014.** Stress-Induced Chloroplast Degradation in Arabidopsis Is Regulated via a Process Independent of Autophagy and Senescence-Associated Vacuoles. *The Plant Cell* **26**: 4875–4888.
- Wang P, Grimm B. 2021.** Connecting Chlorophyll Metabolism with Accumulation of the Photosynthetic Apparatus Plant Science. *Trends in Plant Science* **26**: 484–495.
- Yakushiji E, Uchino K, Sugimura Y, Shiratori I, Takamiya F. 1963.** Isolation of water-soluble chlorophyll protein from the leaves of *Chenopodium album*. *BBA - Biochimica et Biophysica Acta* **75**: 293–298.
- Yamada K, Hara-Nishimura I, Nishimura M. 2011.** Unique defense strategy by the endoplasmic reticulum body in plants. *Plant and Cell Physiology* **52**: 2039–2049.
- Yamada K, Matsushima R, Nishimura M, Hara-Nishimura I. 2001.** A Slow Maturation of a Cysteine Protease with a Granulin Domain in the Vacuoles of Senescing Arabidopsis Leaves. *Plant Physiology* **127**: 1626–1634.

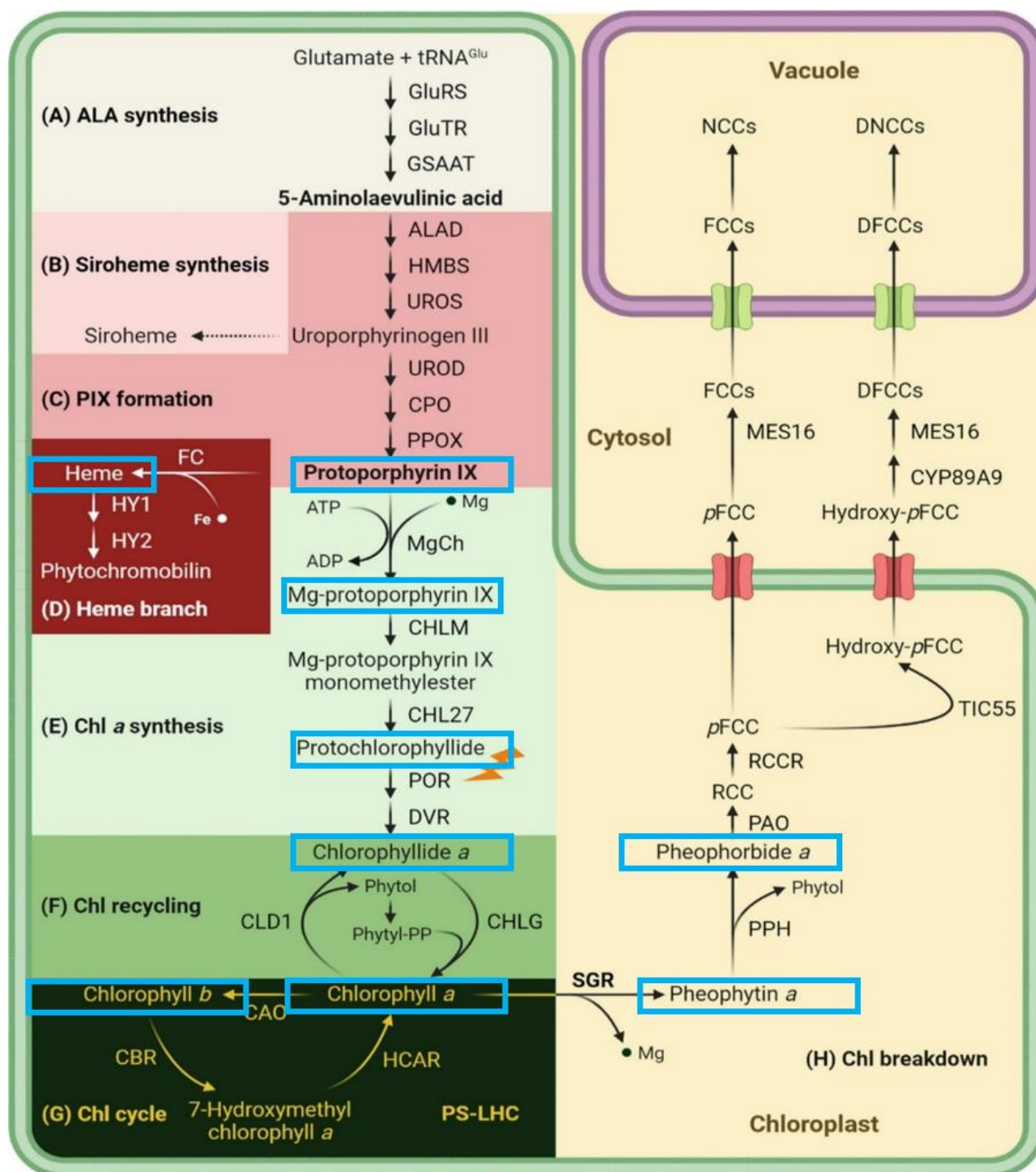
## ANNEXES

## Annexe 1 : Structures de la chlorophylle a et de la chlorophylle b



## Annexe 2 : Représentation schématique du métabolisme des chlorophylles chez les plantes (d'après Wang & Grimm, 2021)

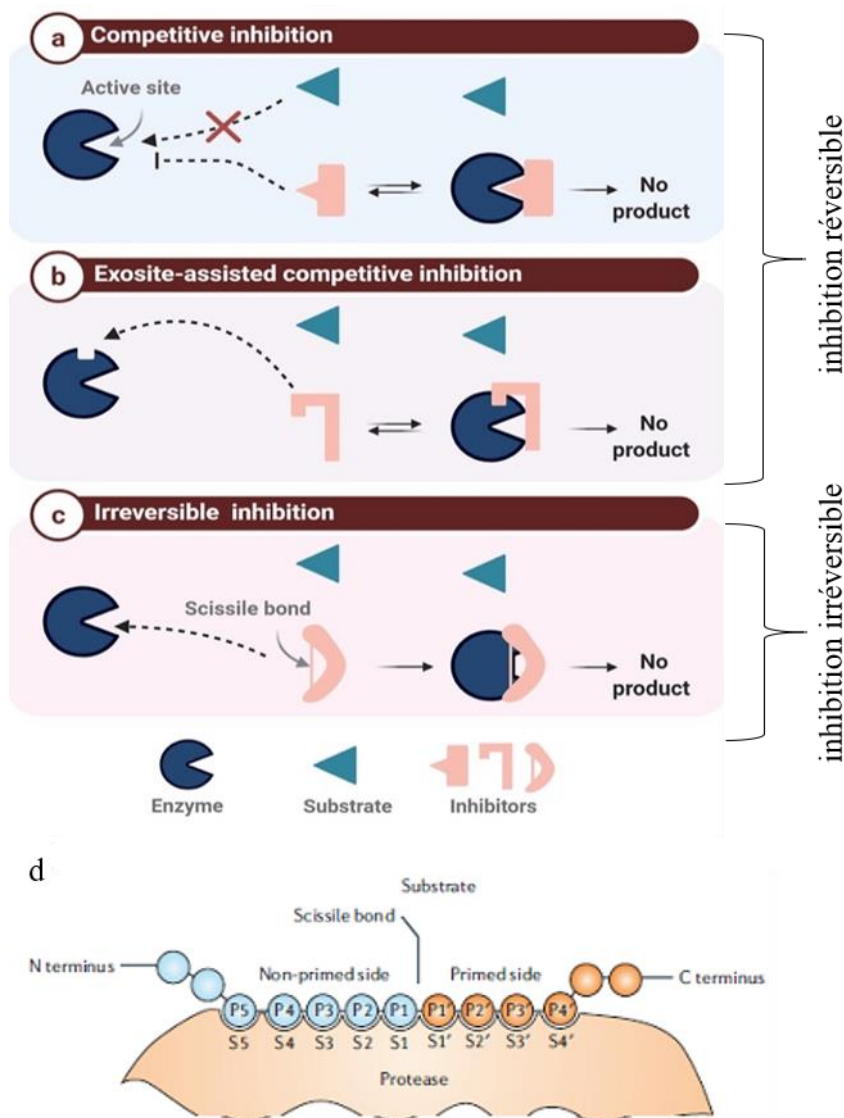
La localisation subcellulaire des enzymes de synthèse et de dégradation des Chls est indiquée. Les molécules pouvant lier les WSCPPI sont encadrées en bleu.



### Annexe 3 : Les différents types d'inhibition des protéases (d'après Jmel *et al.*, 2021).

Les inhibiteurs de protéases assurent une inhibition dite “compétitive réversible” en occupant le site de liaison au substrat naturel (a) ou en se positionnant sur un site proche du site de liaison (exosite) créant un encombrement stérique empêchant la liaison du substrat (b). Le troisième modèle d'inhibition appelé “inhibition irréversible” est consécutif à un encombrement stérique entraînant la réticulation irréversible de la protéase à son inhibiteur (c).

L'interaction se fait par le domaine inhibiteur. La nature de l'acide aminé en position P1 du domaine inhibiteur est essentiel pour la spécificité d'interaction avec la protéase cible (d).

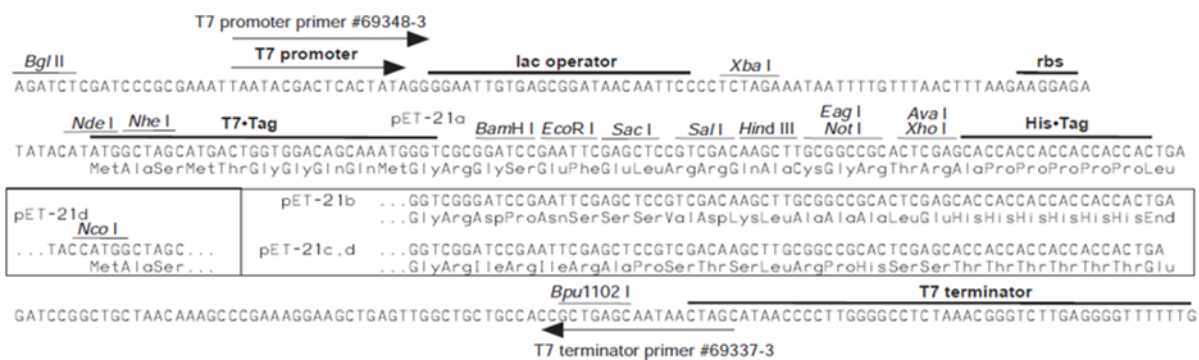
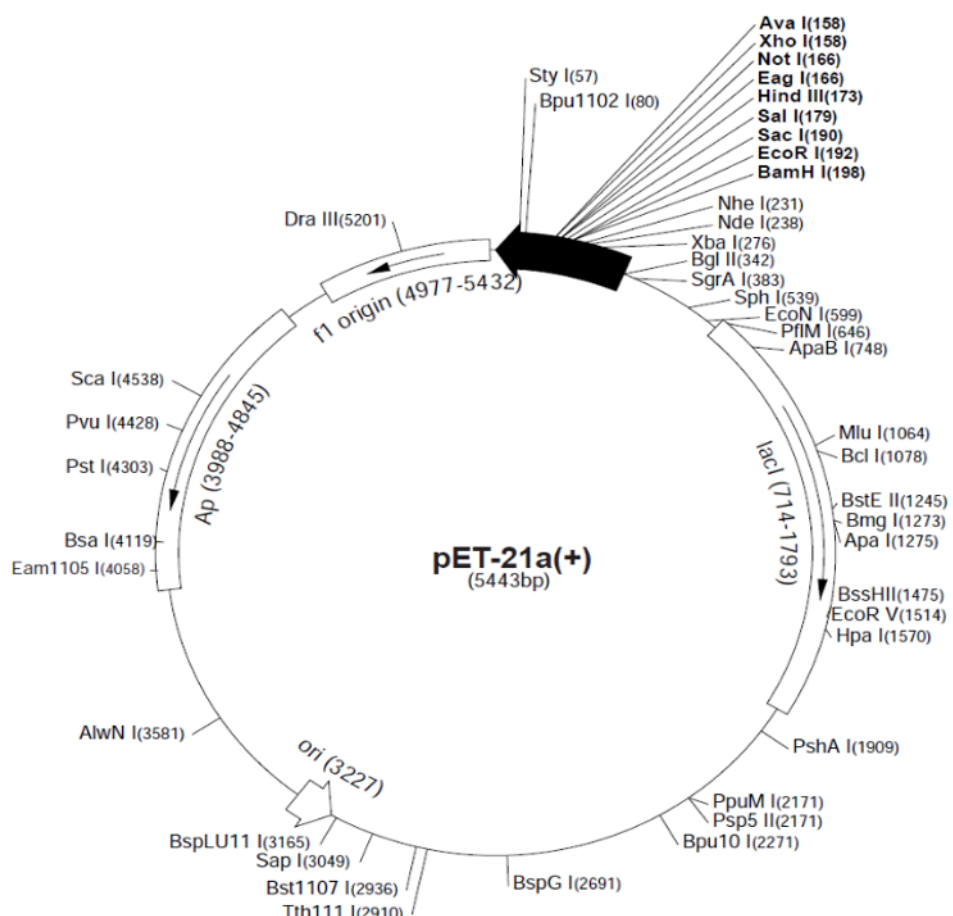


#### Annexe 4 : Liste des amorces utilisées pour l'amplification des gènes *WSCPII* sélectionnés.

Les amorces sont constituées d'une séquence spécifique de l'insert (soulignée) et d'une séquence du vecteur. F, forward; R, reverse

Protéine	Amorce	
avec C-terminal	Prot1	F: <u>ACTTTAAGAAGGAGATATACATGACTACAGAATCCAACGGAAAT</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGGTAGAATGGGAACATCCTTAGAC</u>
	Prot2	F: <u>ACTTTAAGAAGGAGATATACATGAAAGAAGAGGTGAAGGACTCCA</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGGTAGAATGGGAACATCCTTAGAC</u>
	Prot3	F: <u>ACTTTAAGAAGGAGATATACATGGAAAATCGCGTGAATGAC</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGGTAGAAGGGGAACATCCTTAG</u>
	Prot4	F: <u>ACTTTAAGAAGGAGATATACATGCAAGTACCGGTGACGGACA</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGCTTTTCAACACGAGAAGCAGAA</u>
	Prot5	F: <u>ACTTTAAGAAGGAGATATACATGGTAACAACCGGGGCAGAA</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGTTCATAATACTTATTAATACC</u>
	Prot6	F: <u>ACTTTAAGAAGGAGATATACATGGGAGATGAAGTGGAACCCGT</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGGAAAGTAGGGAACATCCTAAGACC</u>
sans C-terminal	Prot1	F: <u>ACTTTAAGAAGGAGATATACATGACTACAGAATCCAACGGAAAT</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGAGCATCATCATCAGCCTTCAC</u>
	Prot2	F: <u>ACTTTAAGAAGGAGATATACATGAAAGAAGAGGTGAAGGACTCCA</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGAGCATCATCATCAACCTTCAC</u>
	Prot3	F: <u>ACTTTAAGAAGGAGATATACATGGAAAATCGCGTGAATGACAC</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGATCACCATGAACCTTGTGGAA</u>
	Prot4	F: <u>ACTTTAAGAAGGAGATATACATGCAAGTACCGGTGACGGACA</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGAACATCTCTGATGAATTTGACG</u>
	Prot5	F: <u>ACTTTAAGAAGGAGATATACATGGTAACAACCGGGGCAGAA</u> R: <u>CTCAGTGGTGGTGGTGGTGGTGCTCAACCTTACGGAACCTAAC</u>
	Prot6	F: <u>ACTTTAAGAAGGAGATATACATGGGAGATGAAGTGGAACCCGT</u> R: <u>CTCAGTGGTGGTGGTGGTGGTCAACCTCACGGAACCTTAAGTGA</u>

**Annexe 5 : Carte du vecteur pET-21a(+) (Invitrogen) utilisé pour l'expression des protéines WSCPII-His-tag recombinantes dans *E.coli*.**



**pET-21a-d(+) cloning/expression region**

**Annexe 6 : tableau regroupant les caractéristiques des souches bactériennes utilisées pour la production des protéines recombinantes (d'après le site web du fournisseur)**

Strain	Features
<i>E. coli</i> Rosetta-gami 2 (DE3) (Novagen #71351)	<p>-Rosetta-Gami 2 host strains allows for enhanced disulfide bond formation and enhanced expression of eukaryotic proteins that contain codons rarely used in <i>E. coli</i>.</p> <p>-Rosetta-gami 2 host strains combine the advantages of Rosetta 2 and Origami 2 strains to alleviate codon bias and enhance disulfide bond formation in the cytoplasm when heterologous proteins are expressed in <i>E. coli</i>. These <i>trxB/gor</i> mutants are compatible with kanamycin-resistant vectors, and carry the chloramphenicol-resistant pRARE2 plasmid, which supplies seven rare tRNAs.</p> <p>-DE3 indicates that the host is a lysogen of <math>\lambda</math>DE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the lacUV5 promoter. T7 lysozyme expression suppresses basal T7 expression. Such strains are suitable for production of protein from target genes cloned in pET vectors by induction with IPTG.</p> <p>-Genotype: <math>\Delta(\text{ara-leu})7697 \Delta\text{lacX74} \Delta\text{phoA PvuII phoR araD139 ahpC galE galK rpsL (DE3) F'[\text{lac}^+ \text{lacIq pro}] \text{gor522::Tn10 } \text{trxB pRARE2 (CamR, StrR, TetR)}</math></p>
<i>E. coli</i> C41(DE3) (Sigma-Aldrich #CMC0017)	<p>-The C41(DE3) is a permissive strain derived from BL21(DE3)</p> <p>-This strain has a mutation that reduces the level of T7 RNAP activity, thereby preventing cell death associated with overexpression of many recombinant toxic proteins.</p> <p>-Express genes cloned into any T7 vector with these BL21(DE3) derivatives</p> <p>-Effective in expressing toxic &amp; membrane proteins</p> <p>-cells express at least 2x the level of other engineered BL21s with my minimally toxic proteins</p>
<i>E. coli</i> Lemo21(DE3) (NEB#C2528J)	<p>-BL21(DE3) containing the Lemo System™</p> <p>-Tunable T7 Expression Strain for difficult targets: membrane proteins, toxic -Lemo21(DE3) offers the host features of BL21(DE3) while also allowing for tunable expression of difficult clones. Tunable expression is achieved by varying the level of lysozyme (<i>lysY</i>), the natural inhibitor of T7 RNA polymerase. The level of lysozyme is modulated by adding L-rhamnose to the expression culture at levels from zero to 2000 <math>\mu\text{M}</math>.</p> <p>-proteins and proteins prone to insoluble expression</p> <p>-Deficient in proteases Lon and OmpT</p> <p>-Resistant to phage T1 (<i>fhuA2</i>)</p> <p>-Lemo System™ maintained by chloramphenicol</p> <p>-Genotype: <math>\text{fhuA2 [lon] ompT gal } (\lambda \text{ DE3) [dcm] } \Delta\text{hsdS/ pLemo(CamR) } \lambda \text{ DE3} = \lambda \text{ sBamHI} \Delta\text{EcoRI-B int::}(\text{lacI::PlacUV5::T7 gene1}) \text{ i21 } \Delta\text{nin5 pLemo} = \text{pACYC184-PrhaBAD-lysY}</math></p>

**Annexe 7 : Liste et composition des solutions pour SDS-PAGE et Western-blot****PAGE-SDS :**

- Gel de concentration : acrylamide 17 % ; bis-acrylamide 0,4 % ; Tris-HCl 0,375 M pH 8,8 ; SDS 0,1 %
- Gel de séparation : acrylamide 12 % ; bis-acrylamide 0,4 % ; Tris-HCl 0,125 M pH 6,8 ; SDS 0,1 %
- Tampon de charge : Tris-HCl 1,5 M pH 6,8 ; SDS 0,35 M ; glycérol 50 % ; bleu de bromophérol 0,001 % ; 5%  $\beta$ -mercaptoéthanol.
- Tampon de migration : Tris-base 25 mM, glycine 190 mM, SDS 7 mM
- Solution de coloration Bleu de Coomassie : Bleu R250 0,25% (m/v); méthanol 40 % (v/v) ; acide acétique 10% (v/v)
- Solution de décoloration : éthanol 40 % ; acide acétique 10 %

**Transfert semi-sec**

Tampon de transfert : 48 mM Tris-Base 1,5 M pH 8,6 ; 39 mM de glycine ; SDS 0,04% ; éthanol 10 %

**Immunodétection**

- Tampon TBS 1X : Tris-Base 20 mM ; NaCl 0,5 M
- Tampon TBS-T : Tween 20 0.05% ; TBS
- Solution de saturation : TBS 1X, 5 % lait écrémé en poudre (m/v), 0,01 % de Tween 20 (v/v)
- Solution de dilution anticorps : TBS 1X, lait écrémé en poudre 0,2%



**Titre :** Caractérisation fonctionnelle des protéines WSCP « Water-Soluble Chlorophyll Proteins » du colza

**Mots clés :** BnD22, protéines recombinantes His-tag, liaison à la chlorophylle, inhibiteur de protéase Kunitz, expression transitoire, microscale thermophorèse.

**Résumé :** Les WSCP II, spécifiques des Brassicaceae, sont des protéines hydrosolubles non-photosynthétiques, qui peuvent se lier aux chlorophylles (Chls). Ces protéines font partie des inhibiteurs de protéases (IP) de type Kunitz et certaines sont surexprimées en réponse à des contraintes abiotiques. L'existence de telles protéines avec une double fonctionnalité de liaison aux Chls et/ou IP reste un sujet à débat et leurs rôles dans les processus physiologiques sont à ce jour très peu documentés. L'objectif de cette thèse est d'apporter des connaissances nouvelles sur ces protéines particulières par leur caractérisation fonctionnelle. Nous avons montré que BnD22, WSCP II majeure foliaire de colza, est une protéine bifonctionnelle capable d'inhiber les Cys protéases et de se lier aux Chls avec une affinité pour la Chla. Ces deux fonctions peuvent être simultanées avec un pouvoir d'inhibition accru lorsque les protéines sont complexées aux Chls. De plus, BnD22 n'est pas localisée dans le chloroplaste mais dans le réticulum endoplasmique et la vacuole. Par ailleurs, nos études phylogénétiques ont révélé chez le colza l'existence d'une famille multigénique divisée en groupes distincts en fonction de leur capacité ou non à se lier aux Chls. La présence des multicopies s'expliqueraient par des profils d'expression différents selon les organes et les stades de développement, et des propriétés fonctionnelles diverses ou redondantes. L'ensemble de ces travaux a permis de valider la diversité fonctionnelle naturelle (IP et/ou liaison aux Chls) des WSCP II du colza ce qui permettra à l'avenir de mieux appréhender les propriétés fonctionnelles des WSCP II et de comprendre leurs rôles physiologiques.

**Title:** Functional characterization of WSCP "Water-Soluble Chlorophyll Proteins" from rapeseed

**Keywords:** His-tag recombinant protein, BnD22, chlorophyll-binding, Kunitz protease inhibitors, transient expression, microscale thermophoresis.

**Abstract:** Class II WSCP (WSCP II) are non-photosynthetic soluble proteins that bind chlorophyll, and have been found so far only in Brassicaceae. These proteins belong to the Kunitz-type protease inhibitor (PI) family and some of them are overexpressed under abiotic stress conditions. Little is known about the dual function of WSCP II and their physiological role is much more limited. The aim of this study was to characterize the function of WSCP II from rapeseed. We first showed that BnD22, the major rapeseed leaf WSCP II, is a bifunctional protein that inhibit Cys proteases and bind Chls with preferential affinity to Chla. These two functions can be simultaneous with an enhanced inhibitory activity when proteins are complexed with Chls. Moreover, we demonstrated that BnD22 is not localized in the chloroplast but in the endoplasmic reticulum and the vacuole. Furthermore, we have demonstrated that rapeseed possesses a multigene family of WSCP II, split into distinct groups associated to preferential binding to Chla or Chlb or the loss of Chl-binding capacity. Rapeseed WSCP II exhibit organ- and developmental stage- specific expression patterns and they possess diverse or redundant functional properties. Altogether, these results allow to understand the functional properties of WSCP II that may open new gate to understand their physiological roles.