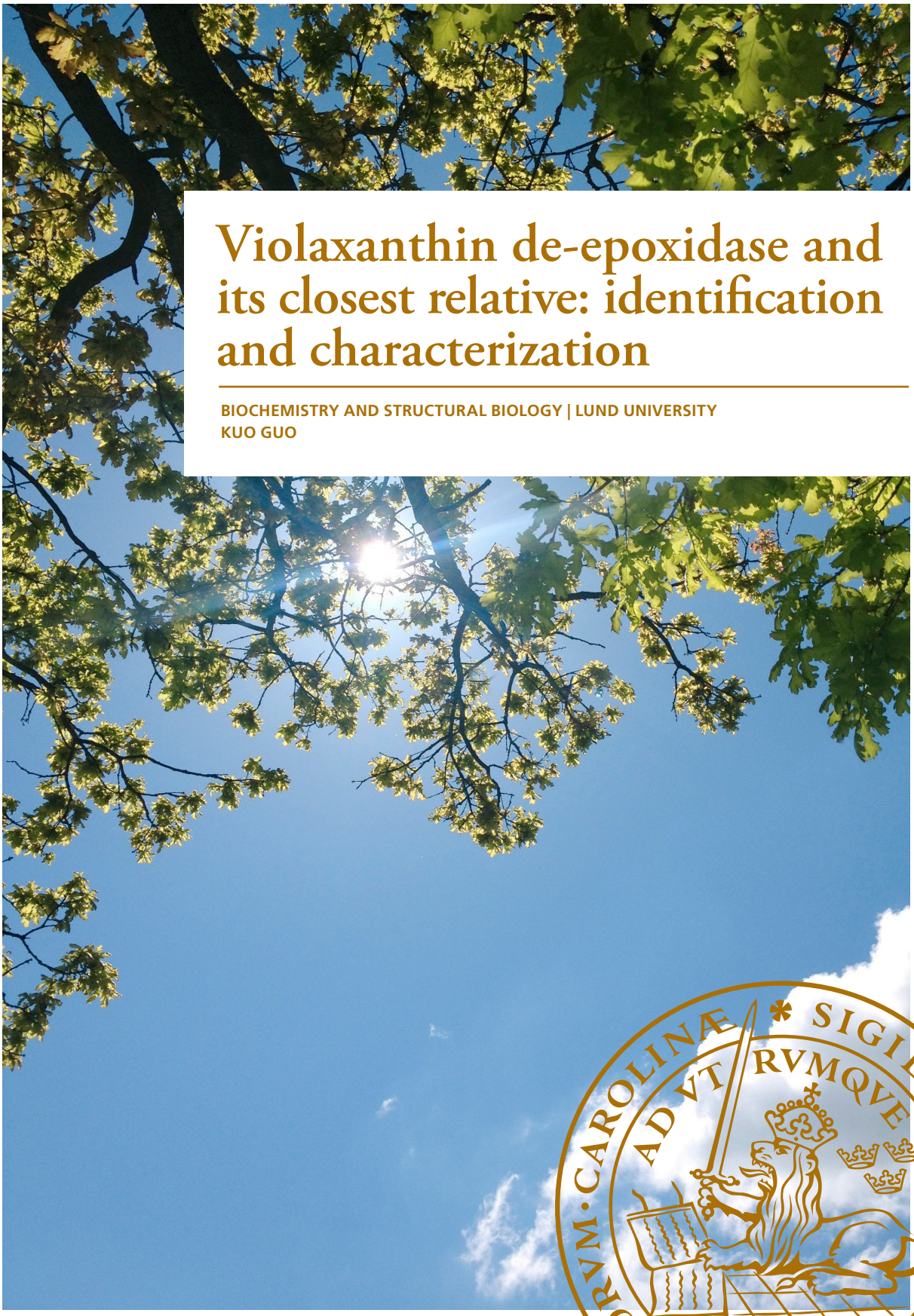
Violaxanthin de-epoxidase and its closest relative: identification and characterization

Guo, Kuo

2017

Document Version:
Publisher's PDF, also known as Version of record



Violaxanthin de-epoxidase and its closest relative: identification and characterization

BIOCHEMISTRY AND STRUCTURAL BIOLOGY | LUND UNIVERSITY
KUO GUO

Violaxanthin de-epoxidase and its closest relative: identification and characterization

Kuo Guo

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Science, Lund University, Sweden.
To be defended on Tuesday December 12th 2017 at 13.15 in Kemicentrum, Lecture Hall A

Faculty opponent
Dr. Cornelia Spetea Wiklund
Department of Biology and Environmental Sciences
University of Gothenburg

Organization
LUND UNIVERSITY
Department of Biochemistry and Structual Biology
Center for Molecular Protein Science

Author
Kuo Guo

Document name
Doctoral Dissertation

Date of issue
2017-11-16

Sponsoring organization

Title and subtitle
Violaxanthin de-epoxidase and its closest relative: Identification and Characterization

Abstract

Light is essential for plants and algae to process photosynthesis. However, excess of light will cause damage to the organism. A process called non-photochemical quenching (NPQ) is an important way for these organisms to protect themselves from photo oxidative damage. The NPQ process is depend on the xanthophyll cycle in thylakoids, which is controlled by the Violaxanthin de-epoxidase (VDE) and zeaxanthin epoxidase (ZE). VDE converts Violaxanthin to to Zeaxanthin on the lumen side of the thylakoid membrane under acid pH conditions caused by photosynthesis.

The mature VDE sequence can be divided into three domains, an N-terminal domain with conserved cysteine pattern, a lipocalin-like domain which expected to carry the substrate binding site, and a C-terminal domain rich in glutamic acids. In this work, we have constructed cysteine mutants that revealed that 12 of the 13 cysteines in VDE are essential for the activity; instead of directly contribute to catalytic function, these 12 cysteines formed disultphide bonds for VDE structural folding.

The catalytic center of VDE does not appear to be only located in the cysteine-rich N-terminal domain. The expressed N-terminal domain did not show activity and the N-terminal truncation of VDE also loss catalytic ability. When mixing these two separately expressed peptides together, the activity was regained. This shows that these two domains could fold independently to their active folding, and also indicates that the active site may be located in the interface of the two domains.

The closest relative of VDE has been found through bioinformatics analysis, the protein has a conserved cysteine pattern as VDE, and named as VDE related protein (VDR). From bioinformatics analysis, some characterization of VDR has been developed; VDR and VDE could be found in same organisms in tree of life, and they are suggested to have the same ancestor; cysteine-rich domain between the two protein are expected to have similar spatial structure, while the corresponding domain of lipocalin-like domain in VDR could have different structure as VDE. With 5' RACE method and western blot, mRNA and protein level evidence of VDR expression is confirmed, together with differential centrifugation, the localization of VDR in leaf would be suggested to follow chlorophyll.

Key words
Violaxanthin de-epoxidase; Violaxanthin de-epoxidase related protein; Cyesteine; function; sequence analysis; western blot; fractionation; 5' RACE

Classification system and/or index terms (if any)

Supplementary bibliographical information | Language: English

ISSN and key title | ISBN 978-91-7422-551-8 (print)
ISBN 978-91-7422-552-5 (pdf)

Recipient's notes | Number of pages 112 | Price

Security classification

Violaxanthin de-epoxidase and its closest relative: identification and characterization

Kuo Guo

LUND
UNIVERSITY

Supervisor: Hans-Erik Åkerlund

Co-supervisor: Cecilia Emanuelsson

Coverphoto by Yang LIU

Faculty of Science
Department of Biochemistry and Structural Biology

ISBN 978-91-7422-551-8 (print)
ISBN 978-91-7422-552-5 (pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2017

Media-Tryck is an environmentally certified and ISO 14001 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

Table of Contents

Abstract

Light is essential for plants and algae to process photosynthesis. However, excess of light will cause damage to the organism. A process called non-photochemical quenching (NPQ) is an important way for these organisms to protect themselves from photo oxidative damage. The NPQ process is depend on the xanthophyll cycle in thylakoids, which is controlled by the Violaxanthin de-epoxidase (VDE) and zeaxanthin epoxidase (ZE). VDE converts Violaxanthin to to Zeaxanthin on the lumen side of the thylakoid membrane under acid pH conditions caused by photosynthesis.

The mature VDE sequence can be divided into three domains, an N-terminal domain with conserved cysteine pattern, a lipocalin-like domain which expected to carry the substrate binding site, and a C-terminal domain rich in glutamic acids. In this work, we have constructed cysteine mutants that revealed that 12 of the 13 cysteines in VDE are essential for the activity; instead of directly contribute to catalytic function, these 12 cysteines formed disultphide bonds for VDE structural folding.

The catalytic center of VDE does not appear to be only located in the cysteine-rich N-terminal domain. The expressed N-terminal domain did not show activity and the N-terminal truncation of VDE also loss catalytic ability. When mixing these two separately expressed peptides together, the activity was regained. This shows that these two domains could fold independently to their active folding, and also indicates that the active site may be located in the interface of the two domains.

The closest relative of VDE has been found through bioinformatics analysis, the protein has a conserved cysteine pattern as VDE, and named as VDE related protein (VDR). From bioinformatics analysis, some characterization of VDR has been developed; VDR and VDE could be found in same organisms in tree of life, and they are suggested to have the same ancestor; cysteine-rich domain between the two protein are expected to have similar spatial structure, while the corresponding domain of lipocalin-like domain in VDR could have different structure as VDE. With 5’ RACE method and western blot, mRNA and protein level evidence of VDR expression is confirmed, together with differential

centrifugation, the localization of VDR in leaf would be suggested to follow chlorophyll.

Populärvetenskaplig sammanfattning

Växter behöver ljus för att växa. Ljuset omvandlas till kemisk energi genom fotosyntes i växtens blad. Den kemiska energin kan sedan användas av växten. Ljusabsorptionen i växter har en begränsning. Alltför mycket ljus kommer att orsaka bildning av högenergimolekyler som är skadliga för växterna. Därför måste växter justera mängden ljus som tas upp. Ljuset i miljön förändras ständigt och exponeringen för starkt ljus kan också variera med tiden. Växter har utvecklat olika lösningar på problemet med alltför starkt ljus i olika tidsskalor. Ett snabbt sätt för växter att omvandla överskott av ljusenergi till värme kräver ett pigment som kallas zeaxanthin. Zeaxanthin bildas från ett annat pigment som kallas violaxanthin med hjälp av ett enzym som kallas violaxanthin de-epoxidas (VDE). Denna reaktion aktiveras när bladen är under lätt ljus-stress.

En del av detta arbete handlar om att karakterisera VDE. Funktion av ett protein är alltid kopplad till dess sekvens och struktur. Byggstenarna för att göra ett protein kallas aminosyror där varje aminosyra har olika egenskaper. Olika aminosyror kopplas ihop i en specifik ordning och ger proteinet dess egenskaper. En typ av aminosyror, cystein, har en speciell förmåga att bindas till varandra under oxidativt tillstånd. Sådana cystein-cystein bryggor har tidigare visat sig bidra till proteiners struktur och katalytiska funktion. I detta arbete har vi bytt ut varje cystein till en annan aminosyra som har liknande egenskaper men som inte kan bilda bryggor. På så sätt identifierade vi cysteinernas betydelse i VDE. Sekvensen av VDE har tre delar. I detta arbete separerade vi VDE i delar och försökte ta reda på vikten av varje del.

En annan del av detta arbete handlar om ett protein som är nära släkt med VDE. Detta protein kallas VDE-relaterat protein (VDR). En viktig likhet mellan de två proteinerna är att de har mycket liknande arrangemang av cysteiner i sin sekvens. Vi gjorde många ytterligare analyser för att försöka kartlägga skillnader och likheter mellan dessa två proteiner. Vi gjorde också en del analyser på VDR för att försöka identifiera dess egenskaper, men dess funktion är fortfarande inte klargjord.

Sammantaget låter dessa resultat oss förstå VDR bättre och samtidigt ökar vår kunskap om VDE, vilket i förlängningen kan användas för förståelse av växtens funktion och i växtförädling.

List of papers

This thesis is based on the following papers:

Paper I

Initial characterization of the closest relative to violaxanthin de-epoxidase
Kuo Guo, Erik Ingmar Hallin, Urban Johanson and Hans-Erik Åkerlund
Manuscript

Paper II

Violaxanthin de-epoxidase disulphides and their role in activity and thermal stability
Erik Ingmar Hallin, Kuo Guo and Hans-Erik Åkerlund
Photosynthesis Research; 2015 May; 124(2):191-8.
doi: 10.1007/s11120-015-0118-9

Paper III

Functional and structural characterization of domain truncated violaxanthin de-epoxidase
Erik Ingmar Hallin, Kuo Guo and Hans-Erik Åkerlund
Physiologia Plantarum; 2016 Aug; 157(4):414-21.
doi: 10.1111/ppl.12428.

Paper not included in the thesis:

Molecular studies on structural changes and oligomerisation of violaxanthin de-epoxidase associated with the pH-dependent activation.
Erik Ingmar Hallin, Hasan Mahmudul, Kuo Guo and Hans-Erik Åkerlund
Photosynthesis Research; 2016 Jul; 129(1):29-41.
doi: 10.1007/s11120-016-0261-y.

Author contributions

Paper I

Kuo Guo was the main contributor in the design of the study, took active part in the bioinformatics analysis, prepared the experiment materials, performed the fractionation experiment, the western blotting experiment, all of PCR works, and 5' RACE, and wrote majority of the manuscript.

Paper II

Kuo Guo contribute to the design of the study, performed the majority work of VDE mutants construction, took active part in VDE mutant expression, purification and activity measurement, and contribute to the manuscript writing.

Paper III

Kuo Guo contribute to the design of the study, performed the majority work of VDE truncations construction, took active part in VDE truncations expression, purification and activity measurement, and contribute to the manuscript writing.

Abbreviations

DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LHC	Light harvesting complex
MGDG	Monogalactosyldiacylglycerol
NADPH	Nicotinamide adenine dinucleotide phosphate
NPQ	Non-photochemical quenching
PCR	Polymerase chain reaction
PsbS	Photosystem II protein S
RACE	Rapid Amplification of cDNA Ends
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
VDE	Violaxanthin de-epoxidase
VDR	Violaxanthin de-epoxidase related protein

Introduction

Photosynthesis is one of the most important biochemical processes in nature; the energy generated from this process has supported the majority of consumption of the living creature on this planet. In photosynthesis process, light is converted into chemical energy by a well-developed bio-system. An antenna structure made by Light harvesting complex (LHC) together with chlorophylls bound to it. This antenna absorbs energy from light and passes the energy through adjacent antennas to the central reaction centre. In the reaction centre, the energy from light is used to split water molecule and pump the protons across the photosynthetic membrane. The proton gradient created by this process could be used by ATP synthase to generate ATP, a common "fuel" molecule which supporting the majority of bio reactions in plant cell.

In the plant cell, the organelle where photosynthesis is taken place in is named chloroplast. Inside the chloroplast, the aqueous fluid is called stroma; and the main components of photosynthetic system are located on an extensive membrane structure named thylakoid. Thylakoids frequently form stacks of disks, a structure which is called grana, and grana are connected by intergranal thylakoids. The continuous aqueous phase inside thylakoid is called lumen, the created proton gradient across the thylakoid membrane not only support the ATP synthesis, but also make the pH of lumen acidic.

Light is a very important environmental parameter, and numerous of biochemical and developmental response to light have evolved in plants to help photosynthesis and growth optimization (Muller, Li et al. 2001). One could easily assume that more light will form more harvestable energy, such assumption is true within a certain range of light intensities (Tamoi, Nagaoka et al. 2006), but above that level light could cause destructive effects for the organism (Barber and Andersson 1992). During the light energy transporting process, chlorophyll is excited to a short-lived singlet state by absorption of protons, and then the energy is transferred to adjacent chlorophyll molecules by resonance, until the energy reach to the reaction centre. However, when the absorbed energy exceed the limitation of what the whole antenna system could handle, the energy transporting from excited-state chlorophyll will be hindered; and therefore, the life time of singlet state chlorophyll is prolonged, which increase the risk of formation of triplet state

chlorophyll. The triplet state chlorophyll is long-lived, and in order to go back to its ground state, triplet state chlorophyll needs to react with other nearby molecules to pass the energy out. As an example, triplet state chlorophyll can react to oxygen and form a very reactive oxygen species, $^1O2$ (singlet oxygen), and cause further damage to the cell (figure 1).

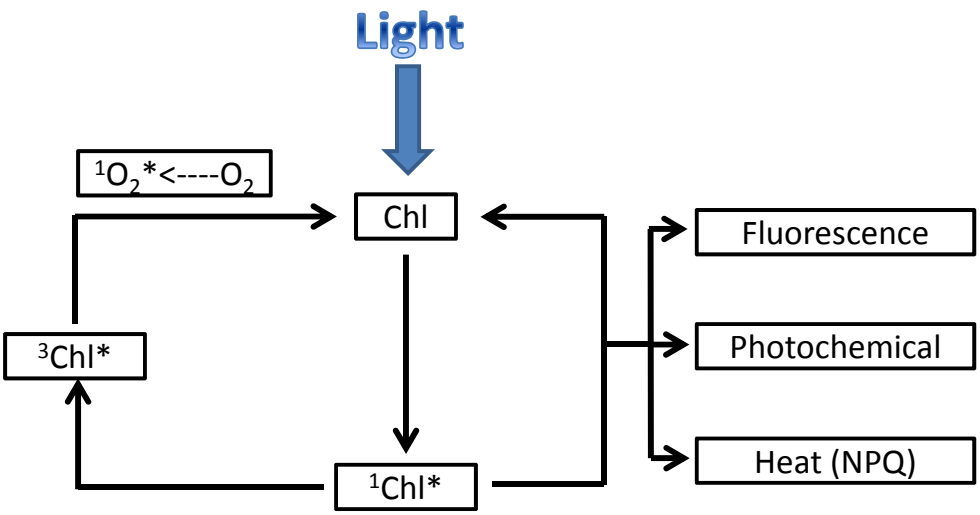


Figure 1 Various ways of de-excitation of excited chlorophyll (Chl). When ^{1}Chl* is formed by Chl absorbed light, in order to go back to ground state, it can either produce ^{3}Chl* and then produce ^{1}O2*, a very reactive oxygen species; or release the energy into light (seen as fluorescence), provide energy to photosynthetic reaction, or dissipate into heat (NPQ).

The light quantity in the natural environment never stops changing in different time scales. Several machineries have been evolved by plants to regulate how much of the light that will be absorbed and how to discard the surplus part. According to Pastenes et al, plant can move the entire leaf to adjust its angle towards the light source (Pastenes, Pimentel et al. 2005). And from Wada's discovery, chloroplasts also have the ability to adjust their position inside the plant cell (Wada 2013). The chloroplast are positioned along the cell edge and vertical to light source for optimized light absorption in low light intensity environment; and when exposed to high light intensity, chloroplasts tend to be positioned parallel to the light source and let them shade each other to reduce the light absorption. While as for the long-term regulation, plants can also control the light harvesting antennae size through gene expression level (Muller, Li et al. 2001).

However, when the adjustment speed of those described machinery is slower than the changing speed of environmental light intensity, plants need a fast alternative to get rid of the redundant light energy which have been captured by the antenna system. The alternative is named non-photochemical quenching (NPQ), which can quench singlet state chlorophylls and dissipate the excess excitation energy as heat, and by this way, converts the excess of the absorbed light into heat and makes the absorbed energy less harmful (Muller, Li et al. 2001) (figure1).

The details of how light energy is converted to heat in NPQ remain unknown, but previous research has discovered some key properties of this process. Gilmore and Yamamoto showed that the activation of NPQ requires a low lumen pH which can be generated by photosynthetic electron transport (Gilmore and Yamamoto 1992), removal of this pH gradient with uncoupling agent will strongly inhibit the NPQ. The pH gradient has been discovered to cause a structural changing of LHC which might beneficial for energy dissipates in heat form (Dekker and Boekema 2005). One important molecule require for NPQ is zeaxanthin, a xanthophyll located inside the thylakoid membrane and synthesized during light stress (Demmig-Adams 1990). The role of zeaxanthin in NPQ has also not been fully determined, but one could assume that zeaxanthin would have direct or indirect association to the LHC, where the excited-state chlorophyll is quenched.

The discoveries of two mutants provide new information of key components involved in NPQ. Both mutants are sensitive to swift light intensity changing. One mutant found by Li et al lacking a membrane bound protein named PsbS (Li, Bjorkman et al. 2000), this protein was suggested by Nield et al to interact with the LHC (Nield, Funk et al. 2000). The other mutant, *npq1*, is a natural mutant lacking an enzyme called violaxanthin de-epoxidase (VDE), and could not accumulate the zeaxanthin under light stress (Niyogi, Grossman et al. 1998). Due to this discovery, one could associate VDE and zeaxanthin with the formation of NPQ.

As shown in figure 2, the difference between violaxanthin and zeaxanthin is the two epoxides in violaxanthin changed to two double bonds. Both of them are xanthophylls and located inside the thylakoid membrane. The xanthophyll cycle consists of two parts, the epoxidation of zeaxanthin and the de-epoxidation of violaxanthin. The two reactions are catalyzed by two different enzymes located on the two different side of thylakoid membrane, and the intermediate product in both reactions is antheraxanthin. Zeaxanthin epoxidase (ZE) located on stroma side of thylakoid membrane, catalyze the epoxidation of zeaxanthin using molecular oxygen and NADPH (Eskling et al. 1997). On the lumen side, VDE catalyze de-epoxidation of violaxanthin using ascorbic acid, the optimal pH of VDE activity is 5.0 to 5.2 (Bratt, Arvidsson et al. 1995), such pH can be created by proton pumping from the photosynthesis process (Kramer et al. 1999). When the high

environmental light intensity triggered photosynthesis process speed up, the low stromal pH created by proton pumping will activate VDE converting violaxanthin to zeaxanthin, and therefore start the zeaxanthin dependent NPQ. When the NPQ slow down the formation of proton gradient, the stromal pH increased, and therefore inactivates the VDE and stops the zeaxanthin formation.

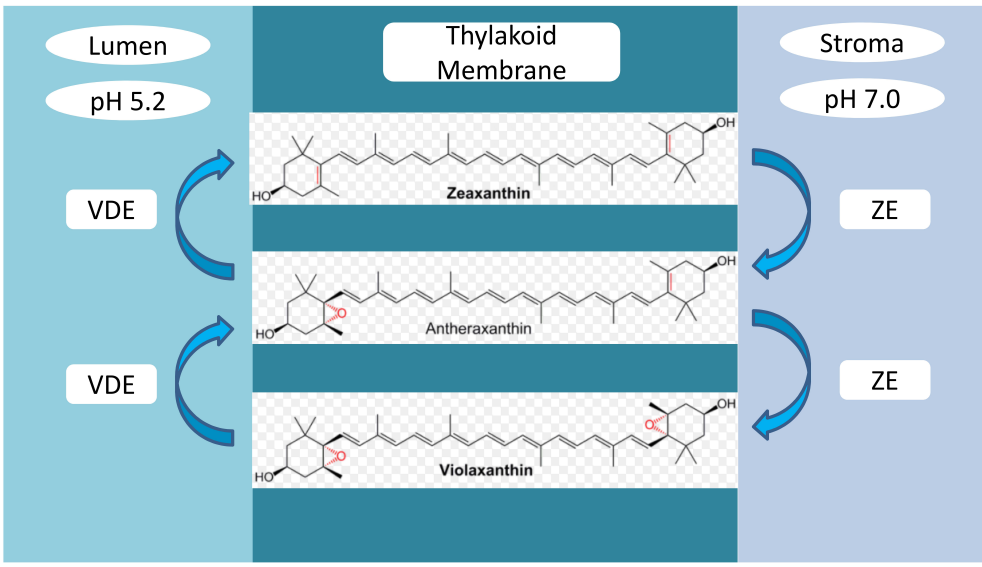


Figure 2 Overview of xanthophyll cycle. In lumen, at low pH (pH 5.2), VDE converts violaxanthin to Zeaxanthin, the reaction requires ascor acid as reductant; while in stroma, at pH 7.0, ZE converts zeaxanthin back to violaxanthin.

The mature sequence of *Spinacia oleracea* VDE is shown in figure 3, the entire sequence could be separated into three regions. The C-terminal region of VDE contains abundant of amino acids with charged side chain, the most significant among them is 25 glutamic acids. According to ProtParam tool calculation, theoretical isoelectric point of this C-terminal region is 4.6, which means this domain is negatively charged in neutral environment. When VDE is located in lumen, the dropping of pH created by proton pumping would make C-terminal domain partially neutralized, and help VDE anchor to thylakoid membrane. This theory is supported by Hieber et al, the C-terminal truncation they created has shown the loss of thylakoid membrane binding ability (Hieber, Bugos et al. 2002).

Figure 3 The sequence of mature spinach VDE. An extra methionine (in black) is added to N-terminal when expressed in *Escherichia coli* (*E.coli*). Cysteine (red)-rich N-terminal domain is marked in orange, lipocalin-like domain is marked in blue, glutamic acid (green)-rich C-terminal domain is marked in purple. The boundary separated N-terminal domain and lipocalin-like domain is determined according to results from Paper III, and will be discussed later.

The crystal structure of the lipocalin-like domain has been reported by Arnoux et al on 2009; the barrel structure consists of eight beta-strands corresponding to the classic lipocalin family (Arnoux, Morosinotto et al. 2009). As the typical function of lipocalin motif is to carry small hydrophobic molecules, the role of this hydrophobic cavity in VDE would be to hold violaxanthin during de-epoxidation process. However, the length of the violaxanthin molecule is too long to be hold by lipocalin-like domain alone, without the other side of the molecule protruding out of the domain. Which means violaxanthin molecule could be either hold by two lipocalin-like domains together, or by another domain in VDE. The dimer-form seen in the crystal structure of the lipocalin-like domain in pH5 from Arnoux et al would suggests that VDE works in dimer to bind one violaxanthin molecule during the reaction; even the lipocalin-like domain did not show activity, they still suggested that the catalytic center of de-epoxidation reaction locates in the lipocalin-like domain, and since VDE work in dimer, the conversion of the two epoxide groups in violaxanthin was suggested to be simultaneous. However, according to Hallin et al, the de-epoxidation of Violaxthin to zeaxanthin follows a non-symmetric conversion model (Hallin, Guo et al. 2016), which means the violaxanthin will be converted to antheraxanthin and then released from VDE during the epoxidation process. Another result that opposes the argument of catalytic centre locates in lipocalin-like domain is that removal of five amino acids from the N-terminal of VDE lead to a completely loss of activity (Hieber, Bugos et

al. 2002). This shows that the lipocalin-like domain alone could not possess the catalytic function.

On the other hand, the real function of N-terminal is still unknown. The unique cysteine pattern of VDE contains 11 cysteines. Hieber et al have shown that the entire N-terminal is essential to VDE activity. The typical function of cysteine is to form disulfide bonds which usually contribute to maintain protein spatial folding. According to Silverman et al (Silverman 1981), vitamin K epoxide reductase use a disulfide bond to convert the epoxide to a double bond. If the same occurs in the conversion from violaxanthin to zeaxanthin remains to be investigated.

From NCBI BLAST tool, the closest relative of VDE has been found; this protein is named as VDE related protein (VDR) and has not been studied before. In *Arabidopsis thaliana* (*A. thaliana*), VDR shared 20% sequence identity to VDE and contained a similar conserved cysteine pattern. This work is mainly focused on functional studies of VDE domains, and the initial study of VDR. By investigating the relationship between VDE and its closest relative, more knowledge of VDE in functional and structural fields could be acquired. With better knowledge of VDE, we could establish a better understanding of xanthophyll cycle, and furthermore, the light protection process in plants and green algae.

Methods and Materials

VDE gene cloning and expression in *E.coli*

The start of an analysis of a certain protein is to obtain sufficient amount of the protein. But it is often difficult to get enough quantity of many proteins by purifying them from its natural organism, when the natural content is relatively low. As for VDE, previous research has identified spinach VDE mature peptide sequence N-terminal, and subsequently obtained the whole sequence of mature spinach VDE (Arvidsson, Bratt et al. 1996). From previous research, with special-designed primers, mature VDE sequence was amplified from purified total spinach RNA, by using reverse-transcription polymerase chain reaction (RT-PCR) and PCR method; and cloned to carrier plasmid by using restriction endonuclease and DNA ligase.

With obtained mature VDE, we have been able to do various modifications on VDE sequence to analyze VDE properties of interest. I) In order to identify the importance of cysteines in VDE, a two-step PCR designed based on an overlap extension PCR method from Ho et al (Ho, Hunt et al. 1989) was used to replace the cysteine in mature VDE sequence to serine, one by one. II) To identify the functional domain of VDE, we created three variations of N-terminal truncations with different length, and their three corresponding N-terminal fragments, these six peptides were cloned and expressed separately. III) We also created three C-terminal truncations with different lengths to analyze the importance of VDE C-terminal. All the nine truncations peptide described were amplified by using PCR with correlated designed primer pairs.

To obtain sufficient amounts of VDE, VDE truncations, and VDE mutants for analysis, we chose to express the target peptides in a prokaryote model organism, *E.coli*. There are many advantages of using *E.coli* for protein expression: the culture conditions and techniques are well-developed; the yield of expression is higher than eukaryote organisms; various strains have been developed especially for protein expression; etc. The *E.coli* strain we used is BL21 (DE3), this strain is genetically-modified with deficiency in the Lon protease and the OmpT protease, and contains the λDE3 lysogen which carries the gene for T7 RNA polymerase.

The expression of T7 RNA polymerase is under control of the lacUV5 promoter, which can be induced by Isopropyl β-D-1-thiogalactopyranoside (IPTG).

The expression plasmid we chose is pET22b vector, this vector has the T7 promoter which can be recognized by T7 RNA polymerase, and has an ampicillin resistance gene as selection marker. The amplified peptide sequence was transferred to pET22b vector and transformed to *E.coli* top10 strain by using heat-shock method and transformed colonies were selected on ampicillin LB agar plate; the purified plasmid from *E.coli* top10 strain was sent to Eurofins Genomics for DNA sequencing with T7 promoter and T7 terminator specialized primer pair; the sequence-confirmed plasmid was transformed to *E.coli* BL21 (DE3) strain for expression. The detailed expression process is described in paper II.

VDE purification and activity assay

Once the target protein was expressed and host cells were harvested, the next step was to purify the target protein. The first step of purification was to release expressed protein from host cell into buffer at appropriate pH and protection reagent (like protease inhibitor to prevent target protein from degradation). Cell disruption methods usually can be divided into physical and chemical ways, but when the target for purification is macro biomolecule which is sensitive to solvent environment and components, for example, proteins; and the downstream processes are followed with series of purification steps, the physical cell disruption method, French press is our optimal choice. Such method can be performed at low temperature, and create much less heat during disruption process compared to sonication disruption method; the shearing force created by high pressure can assure the fragmentation of genomic DNA and make them dissolvable in solution and can be easily removed by using centrifugation.

Before designing detailed purification steps, the expression status of target protein need to be confirmed. The over-expressed protein sometimes shows to be soluble in the cytoplasm of host cell, and sometimes expressed as inclusion bodies. As for VDE, VDE truncations and VDE mutants, we applied the supernatant and solubilized pellet from centrifugation of cell homogenate to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and found inclusion body formation of all expressed target proteins (data not shown). Thus we focused on purification from inclusion bodies. The general idea of inclusion body purification is to keep inclusion bodies as pellet while removing other soluble and insoluble components in the cell homogenate, then to use chaotropic denaturants

such as urea to interfere with hydrogen bonds in inclusion bodies and make the protein unfold and dissolvable; as for VDE, VDE mutants and VDE truncations, we also use Dithiothreitol (DTT) to reduce the disulfide bonds in order to help protein dissolve in the denaturants. The unfolded protein need to be refolded into its native fold, before further functional or structural analysis. The removal of urea and DTT can be achieved either by using dialysis to slowly diffuse denaturants to dialysate, or by using desalting column method to remove denaturants step by step. Due to the presence of abundant cysteines in VDE, VDE mutants and VDE truncations, the correct cysteine-cysteine paring is obviously important for their native folding. We use a redox buffer containing reduced and oxidized glutathione to help cysteine bridge exchanges during refolding. The refolded protein from inclusion bodies was applied to two step size-exclusion chromatography then followed with concentrating using cut-off filters and a final ultracentrifugation to remove any aggregates. The detailed inclusion body purification and refolding process is described in paper II.

A classic method of VDE activity measurement was developed by Yamamoto and Higashi (1978) by using dual wavelength spectrophotometry, the theory of this method is based on the two xanthophyll, violaxanthin and zeaxanthin have different conjugated structure, and there for, different absorbance peak in spectrum. When de-epoxidation reaction starts, the difference in absorbance between two specific wavelengths can be measured to represent the relative value of de-epoxidated state of the two xanthophylls, and estimate the amount of active VDE. The drawback of this method is that it will not show the detailed quantity changing of violaxanthin, antheraxanthin and zeaxanthin during the de-epoxidation reaction; and some extremely low activity VDE mutants require a more sensitive method to detect the decrease of violaxanthin and formation of antheraxanthin and zeaxanthin after relatively long reaction time. A reverse phase-high-performance liquid chromatography (RP-HPLC)-based method has been developed to achieve such object. The components required for VDE activity assay among the two methods described above are same: violaxanthin, monogalactosyldiacylglycerol (MGDG), and ascorbic acid; but when we use RP-HPLC-based method measuring activity of VDE mutants and VDE truncations, we adjust the amount of each components according to specific protein and test different incubation time and other parameter until we get the optimized condition. The reaction was stopped by changing pH from optimal pH of VDE (pH 5.0 to 5.2) to neutral, and pigments extracted with organic solvents to remove all water-soluble components in the reaction solution. Then the extracted pigments were analyzed by RP-HPLC to quantify each component. The detailed activity measurement method of VDE, VDE mutants and VDE truncations are described in paper II and paper III.

Culturing of *Arabidopsis thaliana*

As a model plant, *A. thaliana* with its fully sequenced genome, make it a good start for researches on pant protein like VDR. However, one drawback of *A. thaliana* in protein purification is that the size of the plant is relatively small comparing to spinach (the leaf is even smaller). Unlike spinach, we cannot find *A. thaliana* in the supermarket, so a large batch of plantation is required for getting enough experiment materials.

The plantation of *A. thaliana* is under the guidelines from *Arabidopsis* Biological Resource Center (ABRC). The detailed growth condition is described in paper I. One main problem in *A. thaliana* planting is the larva from fruit flies, those insects are small and their eggs can be easily carried by potting soil bought from supermarket. The problem of fly larva is not only because they eat leaves, but also they created stress to the plant and make the plant response to the predation, and in this way, the plant will not grow under a "normal" condition any more. To prevent any larva disturbing the growth of plant, we pre-treat potting soil in 70 °C Celsius degree for 2 hours before sowing, and the following planting showed the elimination of possible fly eggs in growth environment. But the drawback of heating soil is that the high temperature could kill some beneficial bacteria in the soil as well, and the unbalance of microorganism may lead to non-limitation growth of some fungi in the soil. To reduce the growth speed of fungi, water supplement need to be controlled: flooding the soil should be avoided, and it is highly recommended to use sprinkling can spray water on leaf surface instead of directly pour water into soil.

VDR identification

To identify the localization of VDR, we use fractionation methods together with the western blotting method. One of classic fractionation method is to use density gradient centrifugation. This method requires preparation of a density gradient with Percoll or different concentrations of sucrose in a centrifuge tube, and load cell homogenate on top of the gradient. After centrifugation, particles with different mass/density in the homogenate will be separated, and end up in different layers of the density gradient. This method can be used for further fractionation of organelles, for example, separation of lumen, thylakoid and envelope membrane of chloroplast, but it requires more advanced centrifuge, like swinging ultracentrifuge can operate centrifugation at 200000g (Flores-Perez and Jarvis 2017).

In our experiment, we followed the condition of the differential centrifugation method developed by Aryal. et al (Aryal, Xiong et al. 2014). In their method, *A. thaliana* leaf homogenate was separated into four major fractions: crude nuclear and chloroplast fraction, mitochondria fraction, microsomal fraction, and cytosolic fraction. To prove each fraction has been correctly separated, they use western blot method to detect a series of protein marker which is enriching in different fractions, and the result showed they had done a decent fractionation of the *A. thaliana* leaf cell. Since the chloroplast is a big organelle and can be broken during the homogenization, small components from broken chloroplast may not sediment as fast as entire chloroplast during centrifugation. In our experiment, after the fractionation process, we measured chlorophyll concentration of each fraction to determine the arrangement of chloroplast and its inclusions more precisely. The measurement of chlorophyll concentration was according to the method developed by Bruinsma J on 1961 (Bruinsma 1961).

The western blot method is a widely used analytical technique to detect specific protein from tissue homogenate or extract (Mahmood and Yang 2012). In our experiments, we used this method to trace VDR location in wild type *A. thaliana* leaf fractionations and in the T-DNA insertion *A. thaliana* mutant. The sensitivity of VDR detection mainly depends on the primary antibody specificity; there-for, the well-designed antigen peptide constructs the foundation of a successful western blot reaction. The design of antigen peptide should consider both the availability and specificity; the antigen peptide should be enough hydrophilic, and the special position of the peptide should be expose to the surface of protein, which means it can be recognized by immune cell and stimulate the formation of specific antibody which targeting the antigen peptide; on the other hand, the chosen peptide sequence should be unique for the target protein among all the proteins which can be found in the tissue homogenate. In VDR sequence, we found the last 20 amino acids in C-terminal fulfill both requirements, and considering that the mature sequence of VDR remains unknown, an antigen sequence in the C-terminal could assure the immunogenicity of the mature protein.

When it comes to the detailed step of the western blot reaction, many parameters and condition need to be optimized to get a decent signal under low background. The choice of transfer membrane, the concentration of NaCl in TBS buffer, the choice of blocking reagent and its concentration, the concentration of primary and secondary antibody; the components in primary and secondary antibody solution, the incubation time and environmental conditions, etc.; all these parameter need to optimized by series of experiments. And when using plant leaf tissue as experiment material for western blotting, one important issue need to be considered is the present of endogenous peroxidase in leaf material. Some peroxidases are still active after SDS-PAGE and membrane transferring, which

can react to ECL western blotting substrate and give a false positive signal. Such phenomenon is caused by one of the common conjugated enzyme to secondary antibody is horseradish peroxidase (HRP), which use the same substrate as these peroxidases in leaf tissue. One solution to reduce the influence of endogenous peroxidase is to treat the membrane after transferring with 3% Hydrogen peroxide buffer for 0.5 to 1 hour, therefor to inhibit the endogenous peroxidase and then start the following blocking steps.

VDR T-DNA insertion mutant

The transfer DNA (T-DNA) is a transferred DNA in tumor-inducing (Ti) plasmid. Ti plasmid exists in some bacteria species like *Agrobacterium tumefaciens*, and T-DNA can transfer from these bacteria into the nuclear genome of host plant (Barker, Idler et al. 1983). In 2010, The Salk Institute Genomic Analysis Laboratory started a project to analyze around 25000 protein-coding genes in *A. thaliana* by creating homozygous T-DNA insertion mutants. From bioinformatics database, we found that *A. thaliana* line SALK_200643C showed a T-DNA insertion upstream of the starting codon of the VDR gene, which means in such mutant, expression of VDR could be stopped (Insertion mutant information was obtained from the SIGnAL website at http://signal.salk.edu) (Alonso, Stepanova et al. 2003).

The seed of SALK_200643C is available in ABRC. We planted the same number of SALK_200643C and *A. thaliana* Col-0 seeds and cultured them under the same conditions. We harvested rosette stage of both plants and compared them by using physical and chemical analysis techniques (pigment analysis by using HPLC, pulse amplitude modulated chlorophyll fluorometry, proteomics analysis, western blotting of VDR, etc.).

To confirm the insertion of T-DNA in SALK_200643C and get the detailed information of the insert position in the genome, we purified the genomic DNA from *A. thaliana* Col-0 plant and SALK_200643C plant and designed a series of PCR reaction followed by a sequencing reaction (performed by Eurofins Genomics). Due to the T-DNA left border sequence information obtained from SIGnAL website, two primer pairs were designed, the anti-sense primer was located in T-DNA sequence, 416 base pairs (bp) from the left border; two sense primers located in the VDR gene, one before the intron sequence and the other after the intron sequence. As shown in figure 4, the direction of VDR gene and T-DNA gene are opposite, which make the 5’ of the VDR gene is closed to the left

border of T-DNA. After PCR reaction, the amplification products were purified and send to sequencing to acquire sequence information between the two DNAs.

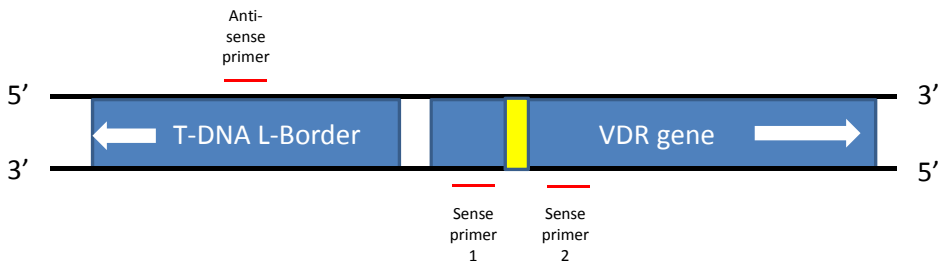


Figure 4 schematic of position T-DNA and VDR gene on chromosome DNA, and the relative position of designed primer pairs. Double strand DNA are shown as thick black lines, T-DNA and VDR gene are shown in blue blocks, the intron of VDR gene is shown in yellow block, the positions of primers are marked in red line. (Paper I)

5' RACE

Mature VDE peptide begins with the cysteine-rich N-terminal, while according to alignment result between VDE and VDR sequences, the cysteine-rich domain in VDE aligned to the middle part of VDR, and the 238 amino acids-length region before this part in VDR has no significant identity with the corresponding region in VDE. If the mature VDR sequence also starts with cysteine-rich region, this 238 amino acids-length would be signal peptides region, and its length is much longer than normal signal peptides need. According to discovery of Wang X et al (Wang, Hou et al. 2016), gene transcription in mammals can be process by alternative transcription sites. We suspected that there might be alternative transcription sites in VDR gene and the normal VDR mRNA is shorter than previous literature reported. To find out other possible length of VDR mRNA, we need to find out the information of 5' end of mRNA which contains VDR gene.

A method called Rapid Amplification of cDNA Ends (RACE) is developed to amplify a nucleic acid sequence from a mRNA template, the amplification is between a defined specific position on mRNA and an unknown sequence at either the 3'end or the 5'end of target mRNA (Frohman, Dush et al. 1988). Unlike normal PCR reaction which requires two specific primers flanking the target sequence, the RACE method only require one primer that recognize the specific

site on mRNA template, such “one-sided” PCR can amplify the sequence from the chosen position to the beginning or the end on target mRNA (Ohara, Dorit et al. 1989).

With 5’ RACE method, we could amplify the VDR mRNA sequence from the binding site of the specific designed primer to the 5’ end of the mRNA into DNA product; and get the 5’ end of VDR mRNA information after purification and sequencing of DNA product. The entire 5’ RACE experiment is processed carefully by following 5’ RACE protocol from Invitrogen, still some focal points are worth to be noticed. Since the starting material of RACE reaction is total RNA from leaf tissue, handling RNA sample requires more careful treatment with lab instruments and reagents to avoid any possible contact between RNA sample and RNase. The success of 5’ RACE reaction largely depends on the specificity of the primer used to bind the target sequence on target mRNA. Such primer is often called the gene specific primer (GSP); the 5’ RACE experiment requires two nested GSPs, one for the cDNA synthesis (GSP1) and another for the following PCR reaction (GSP2). As a common sense, protein peptide is usually constituted by twenty different amino acids, while DNA sequence only use four different nucleotides; which means peptide sequence has much more complex than DNA sequence. As for a specific chosen GSP1 sequence, one can easily find a repeat part in other gene in the same organism by using BLAST tools, the coverage percentage of repeat sequence in GSP1 might be relative low (less than 50%), but the identity between repeat sequence and non-target gene sequence is usually close to 100%. Under such condition, if the length of GSP1 is not sufficient, unspecific binding between GSP1 and non-targeting mRNA could have higher chance to be prolonged by reverse transcriptase. When we design the GSP1 for 5’ RACE, we carefully pick a DNA sequence close to 3’end of VDR mRNA, this 25bp DNA showed well discrimination between VDR and other *Arabidopsis* genes, it only share 7bp with the second best hit besides VDR in BLAST result, and give a good 5’ RACE result.

Bioinformatics analysis

The first time we found VDR sequence was when we use NCBI BLAST (Boratyn, Camacho et al. 2013) tool searching VDE sequence, as the first non-VDE hit in BLAST result, the similarity of cysteine pattern between the two proteins attracts us. To study an unknown protein, bioinformatics tools can usually provide sequence-based analysis and predictions, these results are always good starting point of thinking or experiment designing. When this unknown protein has a known “relative” protein, by comparing the two proteins with different

bioinformatics tools, some properties of the known protein can be associated to the unknown protein, and provide more information of the unknown protein. The detailed bioinformatics tools used are listed in paper I.

Results and discussion

Sequence analysis of VDR

From BLAST results of searching *A. thaliana* VDR peptide sequence among all available organisms, VDR was particularly found in plantae and chromalveolata in eukaryota phyla and could not be found in bacteria, archaea, animalia or fungi (figure 5). At least one organism in green algae, non-vascular land plants, clubmosses, gymnosperms and flowering plants can be found to have VDR. The only subgroup where VDR gene could not be found is ferns; such phenomenon can be explained by the fact that there is no fern organism that has been completely sequenced by now. On the other hand, VDE is presented in all the subgroups in Plantae except ferns, such consistent behavior of VDE compare to VDR could be caused by same reason.

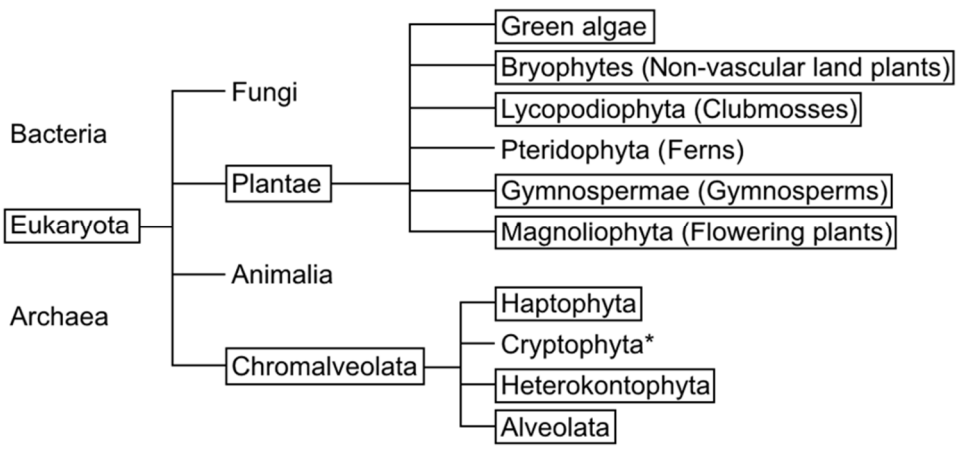


Figure 5 The occurrence of VDE and VDR among different phylogenetic groups of organisms is illustrated by boxed phyla in a schematic phylogenetic tree. It should be noticed that there is no species in fern phylum that has been completely sequenced. cryptophyta phylum is marked by astray because that VDR4 has been found instead of VDE, VDR4 sequence showed an intermediate sequence of VDE and VDR1. The tree is

assembled based on phylogenetic studies of (Woese and Fox 1977), (Kenrick and Crane 1997) and (Adl, Simpson et al. 2005). (Paper I)

The existence of both VDR and VDE in chromalveolata shows that the origin of these proteins is older than Plantae phylum. However, when sequence analysis comes to more primitive organisms, sequences of VDRs among the limited numbers of species from chromalveolata showed more variations than comparison of VDRs from organisms in plant kingdom. In figure 6, alignment result of VDR sequences from the four organisms in chromalveolata shows variation that has influenced some part of cysteine pattern which is considered as the most conserved feature in VDR. In order to classify the different variants of VDR, we defined VDR into four types. VDRs found in aveaolata and heterokontophyta which contain the classic conserved cysteine pattern between VDE and VDR were renamed as VDR1. Some organisms in heterokontophyta also contain a type of VDR which lack some conserved cysteines in VDR1, and showed a slightly different cysteine pattern, this type of VDR is renamed as VDR2. The only VDR we found in haptophyta is renamed as VDR3; this type of VDR does not contain many of the conserved cysteines, but on the other hand, sharing a high identity sequence with other VDRs.

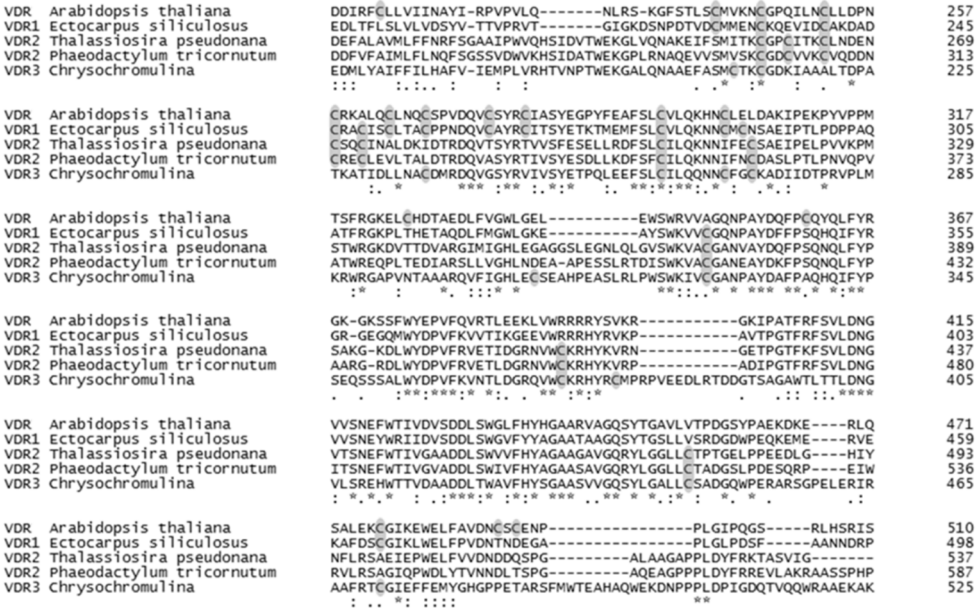


Figure 6 Multiple alignment of 3 different VDR types found in chromalveolata phylum together with VDR in *Arabidopsis thaliana*. (Paper I)

In cryptophyta phylum, we found only one hit by searching VDE and VDR1 sequence from *A. thaliana*, this protein found in the organism *Guillardia theta*, shows sequence similarity with both VDE and VDR (figure 7). Phylogenetic analysis result shows VDR4 has slightly higher sequence identity with VDE than VDR1, but the alignment result shows VDR4 does not have C-terminal domain as in VDE; which means, VDR4 is close VDE on sequence level, but more similar to VDR1 on structural level.

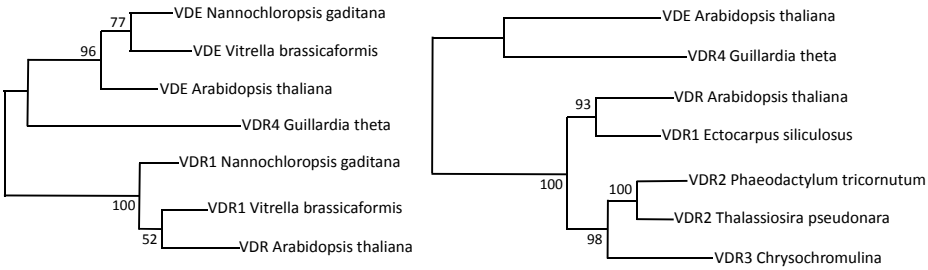


Figure 7 Phylogenetic analyses of VDEs and VDRs. The left phylogram is among VDEs, VDR1s and VDR4, VDEs and VDR1s are clearly separate into two branches, and VDR4 locates close to the basal of VDEs branch. The right phylogram is among 4 VDRs and VDE, shows VDR3 is steadily associated with VDR2 branch. Both trees are midpoint rooted and generated by MEGA7 with Maximum Likelihood option. Bootstrap values in % are showed on the nodes and generated by using 1000 repeats. (Paper I)

Among the four types of VDR variants, VDR1, VDR2 and VDR3 are more close to each other than VDR4 (figure 7). One can assume VDR4 is more close to a common ancestor of VDE and VDR1. However, VDR4 is found in only one organism, to draw such conclusion, further more sequencing information of cognate species of *Guillardia theta* is required. On the other hand, the variation of cysteine pattern in VDR1, VDR2 and VDR3 shows that some cysteines may not contribute to the function of VDR, or, these three VDR variants may have a different function. Since we found VDR1 and VDR2 in two organisms from heterokontophyta, VDR2 can be predicted to have closer relationship with VDR1 than VDR3 does. The phylogenetic analysis result also supports this prediction.

Cysteine pattern in VDE and VDR

According to proteomic search result from NCBI BLAST tool, VDR shows the highest sequence similarity to VDE in *A. thaliana*. VDR has 20% sequence identity to VDE with an expectancy value of 5×10^{-9}, while the second most similar protein in result list gives an expectancy value at 8.3. As shown in figure 8, the most significant resemblance between VDE and VDR is the highly conserved cysteine pattern. In the alignment between the N-terminal domain of VDE and the corresponding region in VDR, only the first cysteine in VDE is not conserved (C120 in VDE aligned to L238 in VDR from *A. thaliana*).

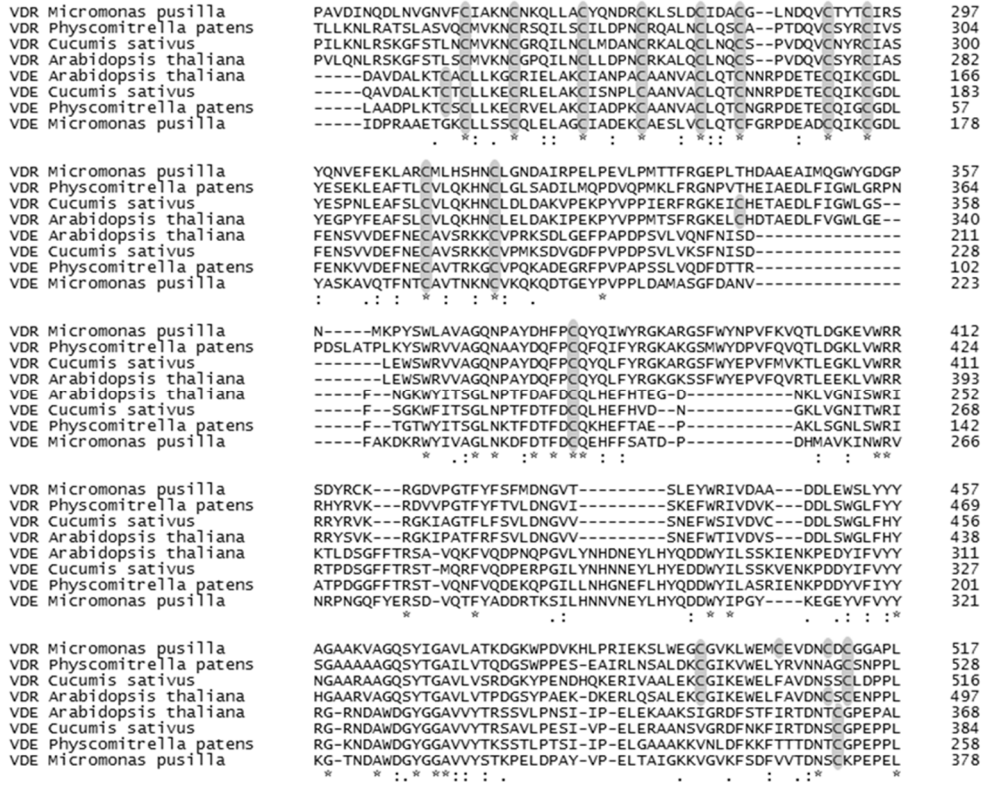


Figure 8 Multiple alignment of VDEs and VDRs from organisms representing Flowering plants, Mosses and Green algae, showing the region corresponding to the N-terminal domain and the lipocalin domain of VDE. (Paper I)

As it was found for VDE in the activity measurement of the 13 cysteine to serine mutants (figure 9), all of the mutants have very low activity except one of them. The mutation to serine in 12 of 13 cysteine leads to a loss of over 99% of VDE activity, while the mutation of first cysteine showed higher activity than native VDE. In accordance with the disulfide connectivity analysis result from Hallin et al (Hallin, Guo et al. 2015), the first cysteine does not form disulfide bond with any other cysteine in VDE, this free cysteine could react with other VDE molecule and influence VDE activity in negative way; thus, the mutation of this cysteine could lead to an increase of VDE stability, and therefor this mutant showed an increased activity. It should be noticed that, this first cysteine is also missing in

green algae VDE, and not found in the VDR sequence, which suggests that the first cysteine might have another function than catalysis and only required by higher plants, while VDR might be more primitive than VDE.

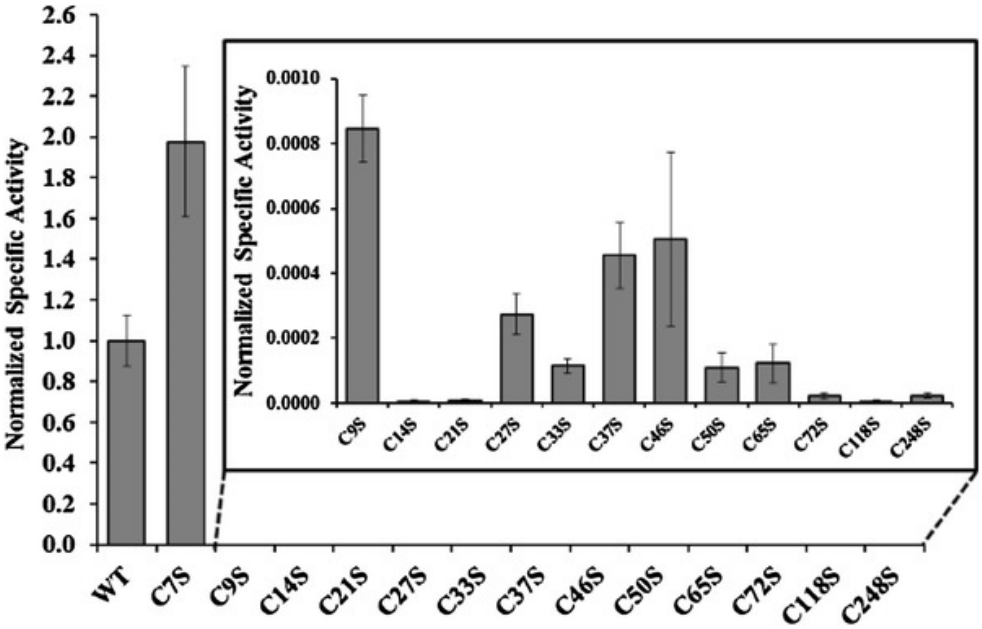


Figure 9 The specific activity of the cysteine to serine mutants normalized to the specific activity of native VDE (WT). A magnification of the mutants with low activity is shown in a window to show the difference in remaining activity. The number after cysteine marks the amino acid ordinal from mature spinach VDE sequence. (Paper II)

As we know, cysteine and serine have similar physical and chemical properties; the main difference between them is the ability of forming disulfide bond. After the great loss of activity, these mutants still retain the ability to catalyze violaxanthin to antheraxanthin or zeaxanthin, this suggests that the spatial structure created and maintained by disulfide bonds might be very important for VDE to perform its catalytic function, but still none of cysteines is directly involved in the de-epoxidation reaction, unless an nearby residue or the serine itself can partly perform the features of removed cysteine and replace it during the catalytic process. The residues directly involved in catalytic process in VDE still need further investigation.

On the other hand, the sequence identity between VDE and VDR in *A. thaliana* is about 20%, while cysteine pattern (except first cysteine) is highly conserved between these two proteins. It could be suggested that cysteines in VDR are required to maintain a spatial structure essential to VDR function, and this maintained spatial structure from VDR might share decent similarity with the structure in VDE (will be discussed later). However, besides the highly conserved cysteine pattern, the low identity in the rest of the sequence between the two proteins would suggest that VDR has a different function as VDE does.

Domains in VDE and VDR

As shown in figure 10, secondary structure prediction results of VDE and VDR are well-matched following the sequence alignment result of the two proteins. N-terminal domain of VDE has been predicted as alpha-helix-rich region, the predicted arrangement of alpha-helixes in VDE is also seen in the prediction of VDR, when the two protein sequences are aligned according to the cysteine pattern. Thus, the corresponding domain in VDR is likely to fold into a spatial structure with high similarity to the cysteine-rich domain in VDE.

Although the secondary structure prediction coincidence could be found between lipocalin-like domain of VDE and the relevant domain in VDR, it is still hard to make the prediction that VDR has a lipocalin-like domain, one reason could be the position of eight beta-strands in the prediction result is not as clear as VDE, while the major evidence turn out to be the lack of structural conserved region (SCR) in VDR sequence. According to Flower (Flower 1996), even the sequence identity of lipocalin family is relatively low (below 20%), still three major SCRs could be found in their sequence. In VDE lopocalin-like domain, two SCR could be found: the GXWY sequence in the first beta-strand, and the R in the end of last beta strand, although one SCR is missing, the lipocalin-like domain crystallization result from Arnoux et al has shown the classic beta-barrel structure (Arnoux, Morosinotto et al. 2009). While in the corresponding region of VDR sequence, all the three SCRs are missing (figure 10). According to this, the beta-strand region of VDR may have a different structure to that of VDE.

The major difference between VDR and VDE sequence is the lack of C-terminal domain. From our findings, the C-terminal truncation VDE with entire C-terminal lacking (lacking the sequence begin with the proline between the grey and black box in figure 10) showed extremely low activity (0.005 % of native VDE activity at pH 5.1), but the formation of anthelaxanthin and zealaxthan was still detected

by RT-HPLC-based method. This illustrate that C-terminal of VDE may not directly be involved in catalytic function but contribute to the process of VDE docking to thylakoid membrane, such opinion is supported by Hieber et al, the C-terminal truncation they created turned out to loss the function of binding to thylakoid membrane (Hieber, Bugos et al. 2002). In VDR, the C-terminal is much shorter than VDE, but still contain many hydrophilic residues, which illustrate this small part of sequence could have certain function while interacting with aqua-environment.

Figure 10 Sequence alignment between VDE and VDR from *A. thaliana* together with their predicted secondary structure (H - α-helix, E - β-strand). The three domains in mature VDE are marked with: N-terminal domain – white box; lipocalin-like domain – grey box; C-terminal domain – black box. The two lipocalin protein SCRs are marked with red box in VDE sequence. (Paper I)

As shown in figure 10, a part between the Cysteine-rich domain and lipocalin-like domain can be found, where neither alpha helix nor beta-strand can be predicted. We suggest that this region is a flexible linker region, and to test our assumption, an N-terminal peptide and its corresponding N-terminal truncation peptide have been expressed and purified separately. As shown in figure 11, the activity measurement based on RP-HPLC showed that the two peptides could catalyze the de-epoxidation reaction without being directly connected by the linker region. Such result also supported the classification of three functional domains in VDE.

In VDR, the region between the cysteine-rich region and beta-strand region is longer than linker in VDE, due to the uncertainty of structure of this beta-strand region, the relationship between this "linker" region and beta-strand region still need further investigation. Although the cysteine pattern behaves as a remarkable evidence of the correlation between VDE and VDR, the uncertainty of beta-strand region and the short C-terminal domain would suggest that VDE and VDR could have different function from each other.

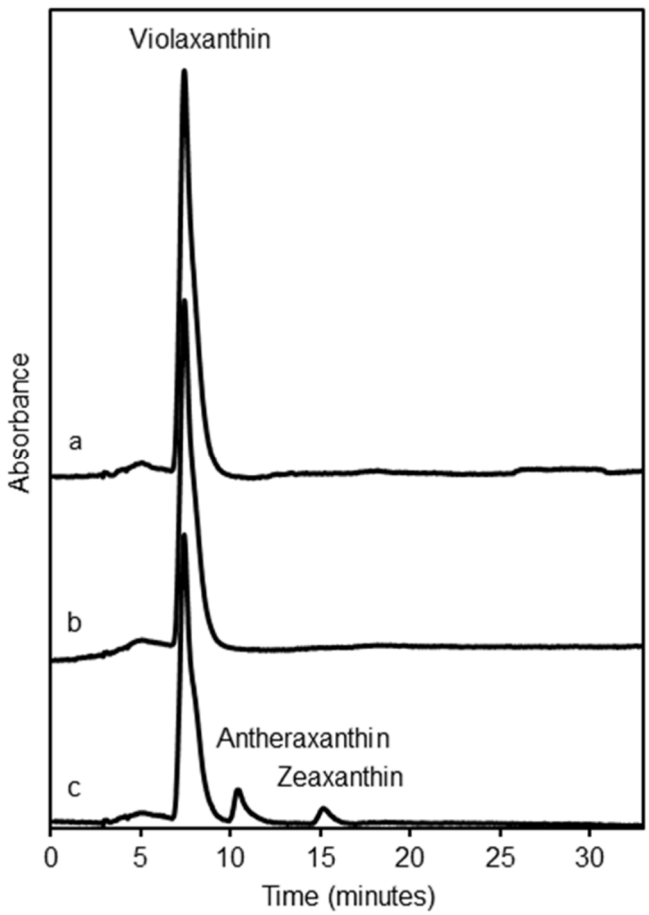


Figure 11 RP-HPLC chromatogram (445 nm) of the extracted xanthophylls from a reaction mixture of violaxanthin, monogalactosyldiacylglycerol and ascorbic acid at pH 5.2 with (a) VDE N-terminal, (b) VDE N-terminal truncation, (c) VDE N-terminal and VDE N-terminal truncation incubated for 24 h at 22°C. (Paper III)

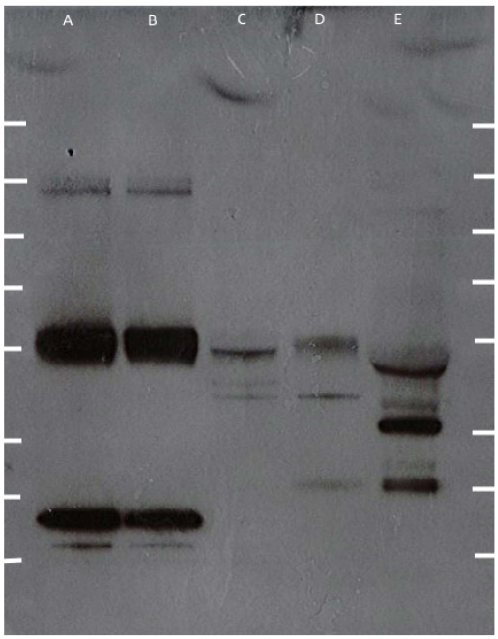
Identification of VDR by fractionation

As shown in figure 12, *A. thaliana* leaf homogenate is separated into 5 fractions by differential centrifugation. According to experiment from Aryal et al (Aryal, Xiong et al. 2014), the unbroken chloroplast are gathered in fraction A, and fraction B should contain mitochondria. While according to the chlorophyll concentration measurements, chlorophylls are mainly arranged in fraction A and B. This can be explained by those thylakoid released from broken chloroplast contribute to the chlorophylls in fraction B. The size of thylakoids are much smaller than entire chloroplast, which means it takes longer time for them to sediment down, and appeared in fraction B. In western blotting result, the same pattern of immunoreaction was found in both fraction A and B, one band at around 55 kDa and one band at 26 kDa followed chlorophyll closely. Thus VDR appears to be located in the thylakoids or the thylakoid membrane. The entire VDR sequence size calculated by ProtParam tool is about 59 kDa, if the 55 kDa is the un-mature VDR, then it is hard to explain how the entire VDR can be transferred to the chloroplast with such a short signal peptide. On the other hand, using the program TargetP, VDR is predicted to have a signal peptide directing the protein to the chloroplast. The TargetP analysis result predicted that the chloroplast transit peptide might be from M1 to V56, leaving a 52.9 kDa product, which is smaller than 55 kDa, this illustrate that either the prediction from TargetP is not accurate, or the 55 kDa band is the dimerization of 26 kDa peptide or some post-translation modified VDR peptide.

Another unexpected finding turns out to be the different arrangement of peptide in the fraction D and E. These two fractions are from the same supernatant, but separated by careful pipetting. In the total 150 µl supernatant, the upper 100 µl (fraction D) behaves significant different as the lower 50 µl (fraction E) closed to the tiny green pellet. Two clear immunoreaction signals was found at around 38 kDa and 29 kDa in fraction E instead of both fraction D and E, such illustrated that these two peptides are sedimenting during the ultracentrifugation instead of maintained soluble in cytosolic fraction means that they are bound to larger structures. The reason of these two bands absence in the pellet is not clear yet.

According to ProtParam tool calculation, the size of peptide from the last cysteine in cysteine-rich domain to c-terminal of VDR is around 25.4 kDa, this size is close to the 26 kDa band in fraction A and B. If this is right then, the mature peptide of VDR has lost the cysteine-rich domain, and the two bands seen in fraction E could be under-transporting state of mature VDR. This can explain why they are missing in the chlorophyll-rich fractions. As we have discussed above, the beta-strand region in VDR did not show any typical lipocalin family conserved sequences, this

region might have a different function as hydrophobic molecules carrier. If the cysteine-rich domain is not in the mature VDR sequence, the mature VDR function could be totally different as VDE and require further investigation.



	Fraction description
Fraction A	Resuspend pellet from 1000g 15mins centrifugation
Fraction B	Resuspend pellet from 10000g 15mins centrifugation
Fraction C	Resuspend pellet from 100000g 2 hours centrifugation
Fraction D	Upper part (100µl) of supernatant from 100000g 2 hours centrifugation
Fraction E	lower part (50µl) of supernatant from 100000g 2 hours centrifugation

	Chlorophyll concentration (mg/L)
Fraction A	11.47
Fraction B	5.65
Fraction C	0.31
Fraction D	0.08
Fraction E	0.07

Figure 12 Summary of fractionation result of *A. thaliana* leaf homogenate. (Paper I)

5’ RACE

Besides the first start codon (ATG) in VDR DNA sequence, there are 7 Met in VDR sequence, and theoretically any of them has the possibility to become the alternative transcription starting position. In order to fish out any mRNA that carries VDR gene, the position of two designed GSP1 close to 3’ end of VDR gene were selected. The calculated full length VDR gene is 1566bp, and the calculated distance between positions of two gene specific primers (GSP1s) to the starting codon was GSP1A - 1090bp; GSP1B - 1540bp. The real product will have an additional 36bp provided by Abridged Anchor Primer (AAP) in 5’ RACE kit.

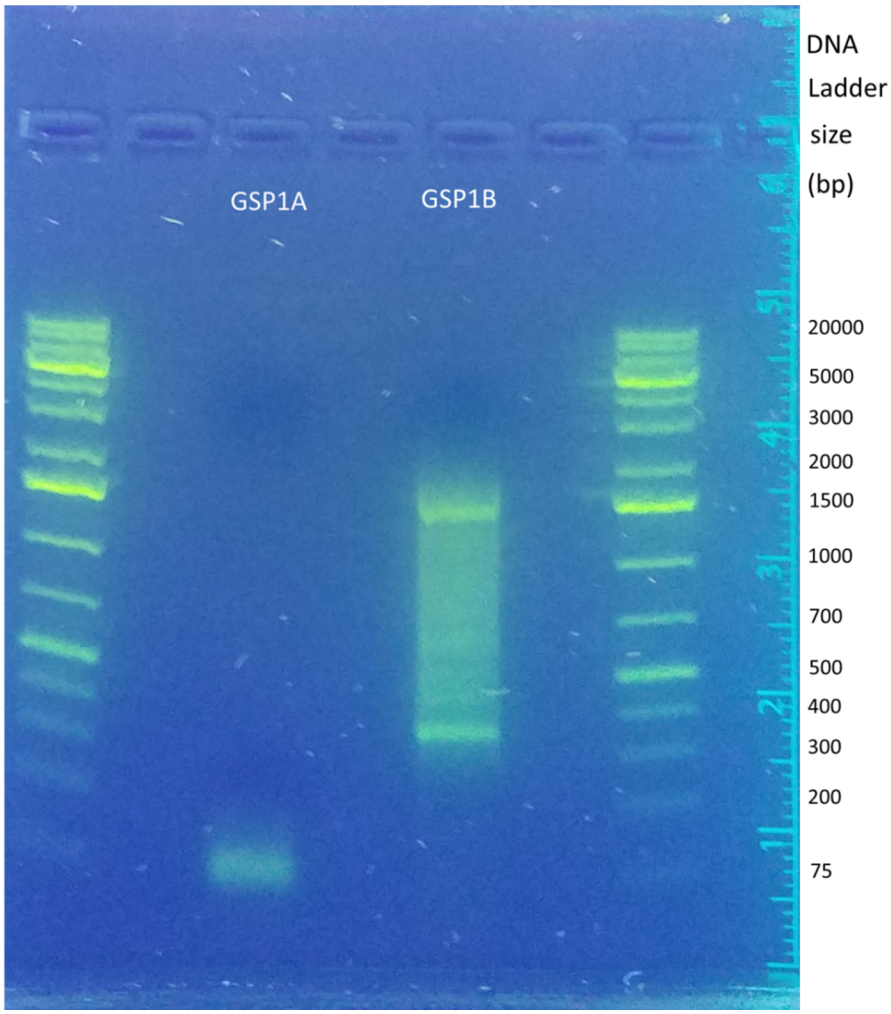


Figure 13 Agarose gel electrophoresis result of PCR products from the two 5' RACE reactions using two designed GSP1s with purified total RNA from *A. thaliana*. (Paper I)

As shown in figure 13, no significant band was found in GSP1A column, except a potential primer-dimer with length less than 75bp; in the GSP1B column, two clear bands (1500bp and 350bp) with a smear are visible, according to manuscript of 5' RACE kit, the smear might be caused by unspecific binding between GSP1 and RNA template. The sequencing results from purified two bands in column GSP1B showed that the 1500bp band has been sequenced 200bp and gave 100% identity to the part of VDR gene. Thus, we could draw the conclusion that the 1500bp band should be the full-length VDR gene, due to the 100% identity of the

sequenced part and the band size is at 1500bp. The sequencing of the 1500bp band only give 200bp length result, this might because of impurity of provided DNA sample and low amount of target DNA (according to trouble shoot of Eurofins Genomics). So far, the only available mRNA of VDR gene that we could detect is the full length mRNA, there might be other alternative mRNA of VDR gene exists, but the extreme low amount made them beyond the detection limit of the 5' RACE method. On the other hand, the sequencing result showed that 350bp band belongs to 5'end of 18s ribosome RNA (rRNA), while the designed GSP1B sequence has no significant correlation to the 18s rRNA sequence. So far, the reason of GSP1B amplified 18s rRNA remain unknown.

VDR T-DNA insertion mutant

To compare the wild type *A. thaliana* and SALK_200643C mutant, as described in paper I, we did pigment analysis, pulse amplitude modulated chlorophyll fluorometry with leaves at same growth stage from both wild type *A. thaliana* and SALK_200643C mutant and the results showed that no difference could be detected between the two plants. We did the fractionation experiment and 5' RACE experiment followed same protocol we have described above with SALK_200643C plant leaves homogenate. The results were the same as in wild type *A. thaliana* (data not shown). Such results would lead to the consequence that the T-DNA insertion process is unsuccessful to silence the VDR expression in SALK_200643C *A. thaliana*. Therefor we designed the PCR test to confirm the T-DNA insertion in SALK_200643C mutant.

5’ of VDR gene:
5’-**TGTCTGAGAGACGTTAACAACTGTTTTCTTCGCC**ATGGCGGCGATTCCTCTC
AAAGCTCCTTCTCTTCCTGGAACGAGAGAGTGTTCTCCCTTCCTTAACCGGTTC
GTCACTCATCAACCGGTTAATTTCATCCGGAAGTTTCATCCTTCCTTAGATGAAG
CATCGTTCTCCAGACGGGGACGAACCACTGTTGTTCGCGTTTTGGA-3’

L border sequence of T-DNA in SALK project:
5’-CTGATGGGCTGCCTGTATCGAGTGGTGATTTTGTGCCGAGCTGCCGGTCGG
GGAGCTGTTGGCTGGCTGGTGGCAGGATATATTGTGGTGTAAACAAATTGACG
CTTAGACAACTTAATAACACATTGCGGACGTTTTTAATGTACTGGGGTGGTTTTT
CTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTG
AGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCT
GTTTGATGGTGGTTCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCC
GAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAAC
GTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACT
ACGTGAAC-3’

Sense sequencing result of PCR amplification1 product from SALK_200643C genomic DNA:
5’-GTTAATGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCA
CGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGG
CTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGAACCACCATCAAACAGGATTT
TCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAG
GCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCC
AGTACATTAAAAACGTCCGCAATGTGTTATTAAGTTGTCTAAGCGTCAATTTGTTTACA
CCACAATATATCGAATACGTTATACAACTGTTTTCTTCGCCATGGCGGCGATTCCTCT
CAAAGCTCCTTCTCTTCCTGGAACGAGAGAGTGTCTCCCTCATAC-3’

Figure 14 The sequencing result of the PCR product by using sense primer 1 and antisense primer in SALK_200643C genomic DNA (bottom). Useful sequence information of the VDR gene (top; 5’ end of VDR gene with 34bp-length 5’ UTR) and the expected inserted T-DNA sequence (middle; reverse and complement relative the bottom sequence) are shown for reference. The yellow bold sequence is the upstream sequence of VDR gene, while incorrect and extra nucleotides are marked with red color. The red boxes show suspected alternative TATA sequences forming a putative promoter in the SALK_200643C genomic DNA upstream of the VDR gene (initiation codon marked in green). The black sequences mark regions that have not been confirmed in wild type and SALK_200643C genomic DNA; sense primer 1 sequence is underlined; detected sequences from the VDR gene in wild type *A. thaliana* (marked as yellow) and from the T-

DNA insertion sequence (marked with blue) are shown. The purple sequences highlight incorrect parts of sequencing results due to limitations of the sequencing method. (Paper I)

The sequencing result of PCR amplification 1 in SALK_200643C genomic DNA shows the left border of T-DNA has inserted to upper stream of VDR gene in chromosome DNA successfully. While the reason of the base pairs changing in the VDR gene upper stream sequence remain unknown, and also the reason of the missing of the 5' end sequence in Left border of T-DNA (figure 14). Theoretically, a T-DNA insertion would lead to interruption of target gene expression, for the inserted sequence cut off the connection between transcription start site and TATA sequence. In SALK_200643C genomic sequence result, one TTATA sequence is found at -20 position to the starting codon, and one TATAT sequence is found at -35 position to the starting codon (red block in figure 14). A typical TATA box usually located at 25-35 base pairs upstream from transcription starting site (TSS); while in the native VDR gene upstream sequence, a TATAA sequence appeared at -6 position to the TSS. The length of 5' untranslated region (UTR) in native VDR gene is only 34bp, which is much shorter than normal 5'UTR. The insertion of T-DNA in the SALK_200643C genome removed some base pairs in original 5' UTR of VDR gene, and at the same time, apart the original TATA sequence from VDR gene, but the loss of 5' end of T-DNA lead to the result that a TATA sequence at -40 to the starting codon replaced by a TATA sequence appeared at -35 to the starting codon. This could explain the expression of VDR gene in the SALK_200643C plant, and indicates that SALK_200643C mutant has a successful T-DNA insertion but the T-DNA insert did not block the translation. Thus, SALK_200643C cannot be used for VDR function analysis.

Summary

One of the key findings in this work is some characterization of VDE. Cysteine plays very important role in VDE activity, 12 out of 13 cysteine mutations would lead to great loss of VDE catalytic ability, while the remained product formation also illustrated that none of cysteines is directly involved in the catalytic process. On the structural level, the VDE truncation research showed that both the cysteine-rich domain and the lipocalin-like domain are important for VDE activity, and when the two truncated peptides were put together, the catalytic ability was recovered, which indicated that the two domains could fold independently to its functional form.

On the other hand, research of VDR started from sequence-based analysis. From bioinformatics analysis, VDE and VDR were found in almost the same organisms, the intermediate peptide sequence found in one primitive organism illustrated VDE and VDR would have a common ancestor; the comparison between VDE and VDR on sequence similarity and secondary structure prediction showed the cysteine-rich domain in both proteins would have similar folding, but the corresponding domain in VDR to the lipocalin-like domain in VDE could have totally different folding and function.

In order to prove and identify VDR expression, we use differential centrifugation to fractionate *A. thaliana* leaves homogenate together with western blot method, the result showed VDR expression following chlorophyll present and the mature protein is a 26 kDa peptide. The 5' RACE method also detect the full length mRNA of VDR, which provide the mRNA level evidence of VDR expression.

To analyze a VDR T-DNA insertion from SALK project, we used sequence specific primer together with molecular biology methods; the result confirmed the unsuccessfully T-DNA insertion and VDR expression in this mutant is confirmed by immuno-blotting method and with 5' RACE.

Future work

One of future research on characterization of VDE would be to obtaining the structure. Several attempts on crystallization conditions have been done by our groups but still no crystal has been obtained, one solution of such problem could be trying to crystalize different domain separately. Since cysteine-rich domain and N-terminal truncation of VDE could fold into active form independently, it is worth trying to obtain crystal structure of these parts separately.

Another future research would focus on catalytic site and substrate binding of VDE. By using site-mutation method together with our high sensitivity RP-HPLC-based activity assay, we could mutate amino acids which suspected to be directly involved in the catalytic process, and select amino acids would be involved in substrate (like ascorbic acid) binding sites.

On the other hand, future work on VDR will mainly focus on obtaining mature sequence and functional analysis. To obtain mature peptide sequence, we could use large scale preparation of *A. thaliana* leaves homogenate and chloroplast preparation; once VDR band is visible on coomassie stained SDS-PAGE gel, the band could be cut off and sequenced by Edman degradation method. According to the band size visualized on western blot, we could also estimate the mature peptide size, and use molecular biology method to construct expression vectors containing designed VDR truncated peptide, and express the peptide in the host cell for further analysis; since the beginning of the mature peptide is unknown, we could express various length of VDR truncation to screen out correct peptide size.

For functional analysis of VDR, we could continue search for VDR knock-out mutant plants and compare the ecotype between mutant and wild type. We could also incubate purified expressed VDR peptide with isolated thylakoid or chloroplast, and analysis component changes after incubation by using HPLC and other analytical methods. In such way, possible biochemical process which VDR might be involved in could be found.

Acknowledgement

Many people I would like to thank during my Ph.D study and my research life.

First of all, I want to thank to my supervisor, Hans-Erik Åkerlund, for he provided me this opportunity to become his group member and join this attractive project, and shared his knowledge with me. He is the best supervisor I have ever met, thank you for everything.

I would like to thank to my co-supervisor, Cecilia Emanuelsson, when I need help, I could always get help from her. I want to thank Urban Johanson, for he shared so much experience with *A.thaliana* culture and bioinformatics analysis. My colleague Erik Hallin, we shared so much interesting idea and work load, thank you for everything. I also want to thank everyone in CMPS, for this excellent research and study environment.

I deeply appreciate to my parents, I know that four-years study far away from home is a hard feeling to my parents, but they still respect my choice, and encourage me to study abroad and let me choose the work I am interested in. I also thank my girlfriend, for her love and supports during my study life.

Additionally, I would also give thanks to my scholarship provider, for their supporting my study life.

References

Adl, S. M., A. G. Simpson, M. A. Farmer, R. A. Andersen, O. R. Anderson, J. R. Barta, S. S. Bowser, G. Brugerolle, R. A. Fensome, S. Fredericq, T. Y. James, S. Karpov, P. Kugrens, J. Krug, C. E. Lane, L. A. Lewis, J. Lodge, D. H. Lynn, D. G. Mann, R. M. McCourt, L. Mendoza, O. Moestrup, S. E. Mozley-Standridge, T. A. Nerad, C. A. Shearer, A. V. Smirnov, F. W. Spiegel and M. F. Taylor (2005). "The new higher level classification of eukaryotes with emphasis on the taxonomy of protists." J Eukaryot Microbiol 52(5): 399-451.

Alonso, J. M., A. N. Stepanova, T. J. Leisse, C. J. Kim, H. Chen, P. Shinn, D. K. Stevenson, J. Zimmerman, P. Barajas, R. Cheuk, C. Gadrinab, C. Heller, A. Jeske, E. Koesema, C. C. Meyers, H. Parker, L. Prednis, Y. Ansari, N. Choy, H. Deen, M. Geralt, N. Hazari, E. Hom, M. Karnes, C. Mulholland, R. Ndubaku, I. Schmidt, P. Guzman, L. Aguilar-Henonin, M. Schmid, D. Weigel, D. E. Carter, T. Marchand, E. Risseeuw, D. Brogden, A. Zeko, W. L. Crosby, C. C. Berry and J. R. Ecker (2003). "Genome-wide insertional mutagenesis of Arabidopsis thaliana." Science 301(5633): 653-657.

Arnoux, P., T. Morosinotto, G. Saga, R. Bassi and D. Pignol (2009). "A structural basis for the pH-dependent xanthophyll cycle in Arabidopsis thaliana." Plant Cell 21(7): 2036-2044.

Arvidsson, P. O., C. E. Bratt, M. Carlsson and H. E. Akerlund (1996). "Purification and identification of the violaxanthin deepoxidase as a 43 kDa protein." Photosynth Res 49(2): 119-129.

Aryal, U. K., Y. Xiong, Z. McBride, D. Kihara, J. Xie, M. C. Hall and D. B. Szymanski (2014). "A proteomic strategy for global analysis of plant protein complexes." Plant Cell 26(10): 3867-3882.

Barber, J. and B. Andersson (1992). "Too much of a good thing: light can be bad for photosynthesis." Trends in Biochemical Sciences 17(2): 61-66.

Barker, R. F., K. B. Idler, D. V. Thompson and J. D. Kemp (1983). "Nucleotide sequence of the T-DNA region from theA grobacterium tumefaciens octopine Ti plasmid pTi15955." Plant Mol Biol 2(6): 335-350.

Boratyn, G. M., C. Camacho, P. S. Cooper, G. Coulouris, A. Fong, N. Ma, T. L. Madden, W. T. Matten, S. D. McGinnis, Y. Merezhuk, Y. Raytselis, E. W. Sayers, T. Tao, J. Ye and I. Zaretskaya (2013). "BLAST: a more efficient report with usability improvements." Nucleic Acids Res 41(Web Server issue): W29-33.

Bratt, C. E., P. O. Arvidsson, M. Carlsson and H. E. Akerlund (1995). "Regulation of violaxanthin de-epoxidase activity by pH and ascorbate concentration." Photosynth Res 45(2): 169-175.

Bruinsma, J. (1961). "A comment on the spectrophotometric determination of chlorophyll." Biochimica et Biophysica Acta 52(3): 576-578.

Dekker, J. P. and E. J. Boekema (2005). "Supramolecular organization of thylakoid membrane proteins in green plants." Biochim Biophys Acta 1706(1-2): 12-39.

Flores-Perez, U. and P. Jarvis (2017). "Isolation and Suborganellar Fractionation of Arabidopsis Chloroplasts." Methods Mol Biol 1511: 45-60.

Flower, D. R. (1996). "The lipocalin protein family: structure and function." Biochemical Journal 318(Pt 1): 1-14.

Frohman, M. A., M. K. Dush and G. R. Martin (1988). "Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer." Proceedings of the National Academy of Sciences of the United States of America 85(23): 8998-9002.

Gilmore, A. M. and H. Y. Yamamoto (1992). "Dark induction of zeaxanthin-dependent nonphotochemical fluorescence quenching mediated by ATP." Proceedings of the National Academy of Sciences of the United States of America 89(5): 1899-1903.

Hallin, E. I., K. Guo and H. E. Akerlund (2015). "Violaxanthin de-epoxidase disulphides and their role in activity and thermal stability." Photosynth Res 124(2): 191-198.

Hallin, E. I., K. Guo and H. E. Akerlund (2016). "Functional and structural characterization of domain truncated violaxanthin de-epoxidase." Physiol Plant 157(4): 414-421.

Hieber, A. D., R. C. Bugos, A. S. Verhoeven and H. Y. Yamamoto (2002). "Overexpression of violaxanthin de-epoxidase: properties of C-terminal deletions on activity and pH-dependent lipid binding." Planta 214(3): 476-483.

Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen and L. R. Pease (1989). "Site-directed mutagenesis by overlap extension using the polymerase chain reaction." Gene 77(1): 51-59.

Kenrick, P. and P. R. Crane (1997). "The origin and early evolution of plants on land." Nature 389(6646): 33-39.

Li, X. P., O. Bjorkman, C. Shih, A. R. Grossman, M. Rosenquist, S. Jansson and K. K. Niyogi (2000). "A pigment-binding protein essential for regulation of photosynthetic light harvesting." Nature 403(6768): 391-395.

Mahmood, T. and P. C. Yang (2012). "Western blot: technique, theory, and trouble shooting." N Am J Med Sci 4(9): 429-434.

Muller, P., X. P. Li and K. K. Niyogi (2001). "Non-photochemical quenching. A response to excess light energy." Plant Physiol 125(4): 1558-1566.

Nield, J., C. Funk and J. Barber (2000). "Supermolecular structure of photosystem II and location of the PsbS protein." Philos Trans R Soc Lond B Biol Sci 355(1402): 1337-1344.

Niyogi, K. K., A. R. Grossman and O. Bjorkman (1998). "Arabidopsis mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion." Plant Cell 10(7): 1121-1134.

Ohara, O., R. L. Dorit and W. Gilbert (1989). "One-sided polymerase chain reaction: the amplification of cDNA." Proceedings of the National Academy of Sciences of the United States of America 86(15): 5673-5677.

Pastenes, C., P. Pimentel and J. Lillo (2005). "Leaf movements and photoinhibition in relation to water stress in field-grown beans." J Exp Bot 56(411): 425-433.

Silverman, R. B. (1981). "Chemical model studies for the mechanism of vitamin K epoxide reductase." Journal of the American Chemical Society 103(19): 5939-5941.

Tamoi, M., M. Nagaoka, Y. Miyagawa and S. Shigeoka (2006). "Contribution of fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase to the photosynthetic rate and carbon flow in the Calvin cycle in transgenic plants." Plant Cell Physiol 47(3): 380-390.

Wada, M. (2013). "Chloroplast movement." Plant Sci 210: 177-182.

Wang, X., J. Hou, C. Quedenau and W. Chen (2016). "Pervasive isoform‐specific translational regulation via alternative transcription start sites in mammals." Molecular Systems Biology 12(7): 875.

Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: The primary kingdoms." Proceedings of the National Academy of Sciences 74(11): 5088-5090.

Printed by Media-Tryck, Lund 2017 NORDIC ECOLABEL, 3041 0903

ISBN 978-91-7422-551-8 (print)
ISBN 978-91-7422-552-5 (pdf)

Biochemistry and Structural Biology
Faculty of Science
Lund University