*Article*

# The Functional Characteristics and Soluble Expression of Saffron *Cs*CCD2

Ying Wang [1,2,†], Siqi Li [1,2,†], Ze Zhou [1,2], Lifen Sun [1,2], Jing Sun [1,2], Chuanpu Shen [1,2], Ranran Gao [3], Jingyuan Song [4,*] and Xiangdong Pu [1,2,*]

1   Inflammation and Immune Mediated Diseases Laboratory of Anhui Province, Anhui Institute of Innovative Drugs, School of Pharmacy, Anhui Medical University, Hefei 230032, China
2   Center of Traditional Chinese Medicine Formula Granule, Anhui Medical University, Hefei 230032, China
3   Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China
4   Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China
*   Correspondence: jysong@implad.ac.cn (J.S.); xdpu@ahmu.edu.cn (X.P.)
†   These authors contributed equally to this work.

**Abstract:** Crocins are important natural products predominantly obtained from the stigma of saffron, and that can be utilized as a medicinal compound, spice, and colorant with significant promise in the pharmaceutical, food, and cosmetic industries. Carotenoid cleavage dioxygenase 2 (*Cs*CCD2) is a crucial limiting enzyme that has been reported to be responsible for the cleavage of zeaxanthin in the crocin biosynthetic pathway. However, the catalytic activity of *Cs*CCD2 on β-carotene/lycopene remains elusive, and the soluble expression of *Cs*CCD2 remains a big challenge. In this study, we reported the functional characteristics of *Cs*CCD2, that can catalyze not only zeaxanthin cleavage but also β-carotene and lycopene cleavage. The molecular basis of the divergent functionality of *Cs*CCD2 was elucidated using bioinformatic analysis and truncation studies. The protein expression optimization results demonstrated that the use of a maltose-binding protein (MBP) tag and the optimization of the induction conditions resulted in the production of more soluble protein. Correspondingly, the catalytic efficiency of soluble *Cs*CCD2 was higher than that of the insoluble one, and the results further validated its functional verification. This study not only broadened the substrate profile of *Cs*CCD2, but also achieved the soluble expression of *Cs*CCD2. It provides a firm platform for *Cs*CCD2 crystal structure resolution and facilitates the synthesis of crocetin and crocins.

**Keywords:** crocins; *Cs*CCD2; saffron; soluble expression; fusion tag

## 1. Introduction

Crocins, which are highly beneficial as medicines for human disorders and spices for flavoring and coloring, only accumulate in specific tissues, such as the stigmas of saffron (*Crocus sativus*), the fruits of *Gardenia jasminoides*, and the flowers of *Buddleja davidii*, of a few distantly related plants [1]. Numerous pharmacological studies have demonstrated that crocins exhibit several anticancer [2], anti-oxidative [3], anti-apoptotic [4], and anti-inflammatory effects [5]. Additionally, recent studies have reported the beneficial effects of crocins on Alzheimer's disease and depression [6,7]. Saffron is the primary natural source of crocins, which confer to its stigma a characteristic dark red color [8]. Owing to their complicated isolation from plants and chemical synthesis, crocins command high market prices [9]. Crocins' high price and remarkable properties in the treatment of central nervous system and cardiovascular diseases have led numerous scientists to investigate their biosynthetic pathway and heterologous production. With the rapid advancement in synthetic biology, the research on the synthesis of plant natural products using microbial fermentation has significantly progressed, overcoming the limitations of plant resources

and providing a new route for the green and efficient production of plant natural compounds [10], such as artemisinic acid [11,12], etoposide aglycone [13], and cannabinoids [14]. Consequently, the heterologous synthesis of crocins using microorganisms is a significant addition to the existing production methods. The biosynthetic pathway of crocins includes the cleavage of carotenoids by carotenoid cleavage dioxygenases (CCDs), oxidation of aldehydes by aldehyde dehydrogenases (ALDHs), and transfer of glycosyl groups by UDP-glucuronosyltransferases (UGTs) (Figure 1) [1,8,15–17]. Among them, CCDs are considered to be rate-limiting enzymes in the biosynthetic pathway of crocins. CCDs, which are essential enzymes that catalyze carotenoid cleavage, are classified into four subfamilies—CCD1, CCD4, CCD7, and CCD8 [18]. The CCD1 subfamily is responsible for the production of volatile terpenoids, which are crucial for the formation of plant aroma [19]; the CCD4 subfamily is crucial for plant color formation [20–22]. CCD7 and CCD8 participate in the synthesis of the plant hormone strigolactone and play an important role in the germination of lateral roots and lateral buds [23]. Thus, the screening and identification of CCDs responsible for the synthesis of crocin precursors are crucial areas of research.
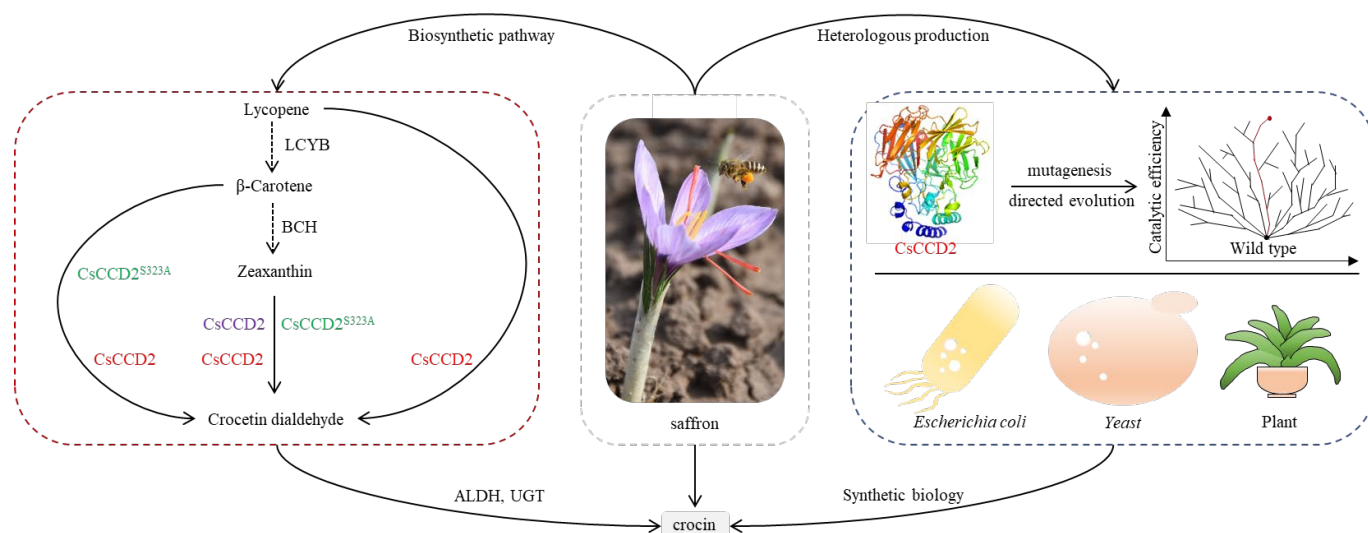


**Figure 1.** The function identification of *Cs*CCD2 in saffron and its application in synthetic biology. Note: The different color of *Cs*CCD2 represent different results reported in PNAS (purple), J Agric Food Chem (green) and this study (red).

In 2014, Frusciantea et al. reported a novel key enzyme *Cs*CCD2 from saffron, which is considered to be the first key enzyme in the biosynthetic pathway of crocins [8]. It catalyzed the cleavage of zeaxanthin to produce crocetin dialdehyde but did not use β-carotene and lycopene as substrates. To further increase catalytic efficiency, researchers have engineered *Cs*CCD2 variants with broader substrate profiles for crocin biosynthesis using a "hybrid-tunnel" strategy. Based on the results of a directed evolution study, $Cs\text{CCD2}^{S323A}$ was reported to have additional catalytic activity on β-carotene and improved catalytic efficiency [24]. Despite having better catalytic efficiency than the wild-type, the yield of the engineered strain remains in the milligram range and needs further optimization. In contrast, heterologous protein expression of CCDs is a significant challenge that impedes the discovery of molecular and physiological functions to some extent [25]. The low catalytic efficiency of *Cs*CCD2 may be attributed to its misfolding during heterologous expression in *E. coli*. However, a protein expression optimization study of *Cs*CCD2 remains missing.

The biosynthetic pathway of carotenoids in *Erwinia* species has been precisely identified [26]. crtE (GGPP synthase) can catalyze farnesyl pyrophosphate to form geranylgeranyl pyrophosphate (GGPP). crtB (PPPP synthase) is responsible for phytoene formation from GGPP, and phytoene is catalyzed by crtI (phytoene desaturase) to form lycopene. crtY (lycopene cyclase) is involved in β-carotene production from lycopene, and crtZ (β-carotene

hydroxylase) can transfer β-carotene to zeaxanthin. Three plasmids were constructed based on these genes and used in carotenoid biosynthesis in *E. coli*. *E. coli* with pACCAR25ΔcrtX, which contained *crtE*, *crtI*, *crtB*, *crtY*, and *crtZ* from *E. uredovora*, was used in zeaxanthin production. *E. coli* with pACCAR16Δcrt harboring *crtE*, *crtI*, *crtB*, and *crtY* from *E. uredovora* can produce β-carotene. *E. coli* with pACCRT-EIB, which comprises *crtE*, *crtI*, and *crtB* from *E. uredovora*, is responsible for lycopene accumulation [26]. It establishes a solid foundation for the in vivo functional identification of the *CCD* gene.

In this study, we constructed the *Cs*CCD2 gene in a pET32a prokaryotic expression vector, and three engineered *E. coli* harboring pACCAR25ΔcrtX, pACCAR16Δcrt, and pACCRT-EIB were used for functional identification via in vivo assays [26]. Using ultra-performance liquid chromatography (UPLC) and UPLC with tandem mass spectrometry (UPLC-MS/MS), we demonstrated that the wild *Cs*CCD2 not only exhibited a catalytic impact on zeaxanthin but also possessed cleavage activity for β-carotene and lycopene. The molecular basis of the varied functionality of *Cs*CCD2 was also elucidated via a truncation study. Furthermore, a protein expression optimization study of *Cs*CCD2 revealed that maltose-binding protein (MBP)-*Cs*CCD2 was soluble using *E. coli* as the host organism at an induction temperature of 16 °C with 0.8 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). Additionally, the catalytic efficiency of *Cs*CCD2 was enhanced as protein solubility increased. This will establish the groundwork for further research into the catalytic mechanism of *Cs*CCD2 and provide important genetic tools for the synthesis of crocetin and crocins.

## 2. Results

### 2.1. Functional Characteristics of CsCCD2

The prokaryotic expression vector TRX-*Cs*CCD2 (Table 1, Figure 2A) was constructed and co-transformed with pACCAR25ΔcrtX, pACCRT-EIB, or pACCAR16Δcrt into *E. coli* BL21(DE3) to create the TRX-CsCCD2-Z/L/B engineered strains. The pET32a vector (Table 1, Figure 2A) was also co-transformed with these three engineered plasmids to form pET32a-Z/L/B as a control. Following IPTG induction, the products were extracted and identified using UPLC and UPLC-MS/MS. The characteristic spectrums of crocetin dialdehyde (with mainly two maximum absorptions at 443.90 nm and 467.01 nm), zeaxanthin (with mainly two maximum absorptions at 450.30 nm and 473.90 nm), lycopene (with mainly three maximum absorptions at 446.71 nm, 468.47 nm, and 498.65 nm), and β-carotene (with mainly two maximum absorptions at 449.57 nm and 474.39 nm) are demonstrated in Figure S1A–D, respectively. The UPLC results depicted that the extracts of TRX-*Cs*CCD2-Z/L/B produced a new peak at 14.55 min in accordance with the retention time of the crocetin dialdehyde standard. The characteristic spectrum of the new peak was parallel to the crocetin dialdehyde standard. Based on the UPLC-MS/MS analysis, the m/z of the new peak was 297.1855, which also corresponded to the crocetin dialdehyde standard (Figure 3C). Moreover, the fragmentation pattern of this new peak was consistent with the standard. However, no crocetin dialdehyde product was detected in pET32a-Z/L/B. These results revealed that TRX-*Cs*CCD2 cleaved zeaxanthin, β-carotene, and lycopene.

The function of DAN1-*Cs*CCD2, was identified again in this study using the same methodology as previously reported [8]. The prokaryotic expression vector DAN1-*Cs*CCD2 was co-transformed with pACCAR25ΔcrtX, pACCRT-EIB, or pACCAR16Δcrt into *E. coli* BL21(DE3) to construct the DAN1-CsCCD2-Z/L/B engineered strains. The pTHIO-DAN1 vector was co-transformed with these three engineered plasmids to form pTHIO-DAN1-Z/L/B as a control (Table 1). After being induced by arabinose, the products were extracted and tested using UPLC and UPLC-MS/MS. The UPLC results indicated that only DAN1-CsCCD2-Z exhibited a chromatographic peak with the same retention time as the crocetin dialdehyde standard. Mass spectrometry further confirmed that the product was the substance crocetin dialdehyde. These results demonstrated that DAN1-CsCCD2 only exhibited catalytic activity against zeaxanthin and had no activity on β-carotene or lycopene, which was consistent with the results reported in the literature (Figure 3B).

**Table 1.** Plasmids used in this study.

| Plasmids | Relevant Properties or Genetic Marker [a] | Source or Reference |
|---|---|---|
| pACCAR25ΔcrtX | pACYC184 plus crtE, crtI, crtB, crtY, and crtZ from *E. uredovora*, Cm$^R$ | [26] |
| pACCAR16Δcrt | pACYC184 plus crtE, crtI, crtB, and crtY from *E. uredovora*, Cm$^R$ | [26] |
| pACCRT-EIB | pACYC184 plus crtE, crtI, and crtB from *E. uredovora*, Cm$^R$ | [26] |
| pTHIO-DAN1 | *pBR322* ori and *pUC* ori, Amp$^R$ | [27] |
| DAN1-*Cs*CCD2 | pTHIO-DAN1 plus *Cs*CCD2 from saffron | [8] |
| DAN1-*Cs*CCD2-T | pTHIO-DAN1 plus truncated *Cs*CCD2 | This study |
| pET32a | *pBR322* ori and *f1* ori, Amp$^R$ | Novagen |
| TRX-*Cs*CCD2 | pET32a plus *Cs*CCD2 from saffron | This study |
| pET41a | *pBR322* ori and *f1* ori, Kan$^R$ | Novagen |
| GST-*Cs*CCD2 | pET41a plus *Cs*CCD2 from saffron | This study |
| pET28a | *pBR322* ori and *f1* ori, Kan$^R$ | Novagen |
| pET28a-SUMO | pET28a plus SUMO•Tag | This study |
| pET28a-MBP | pET28a plus MBP•Tag | This study |
| SUMO-*Cs*CCD2 | pET28a-SUMO•Tag plus *Cs*CCD2 from saffron | This study |
| MBP-*Cs*CCD2 | pET28a-MBP•Tag plus *Cs*CCD2 from saffron | This study |

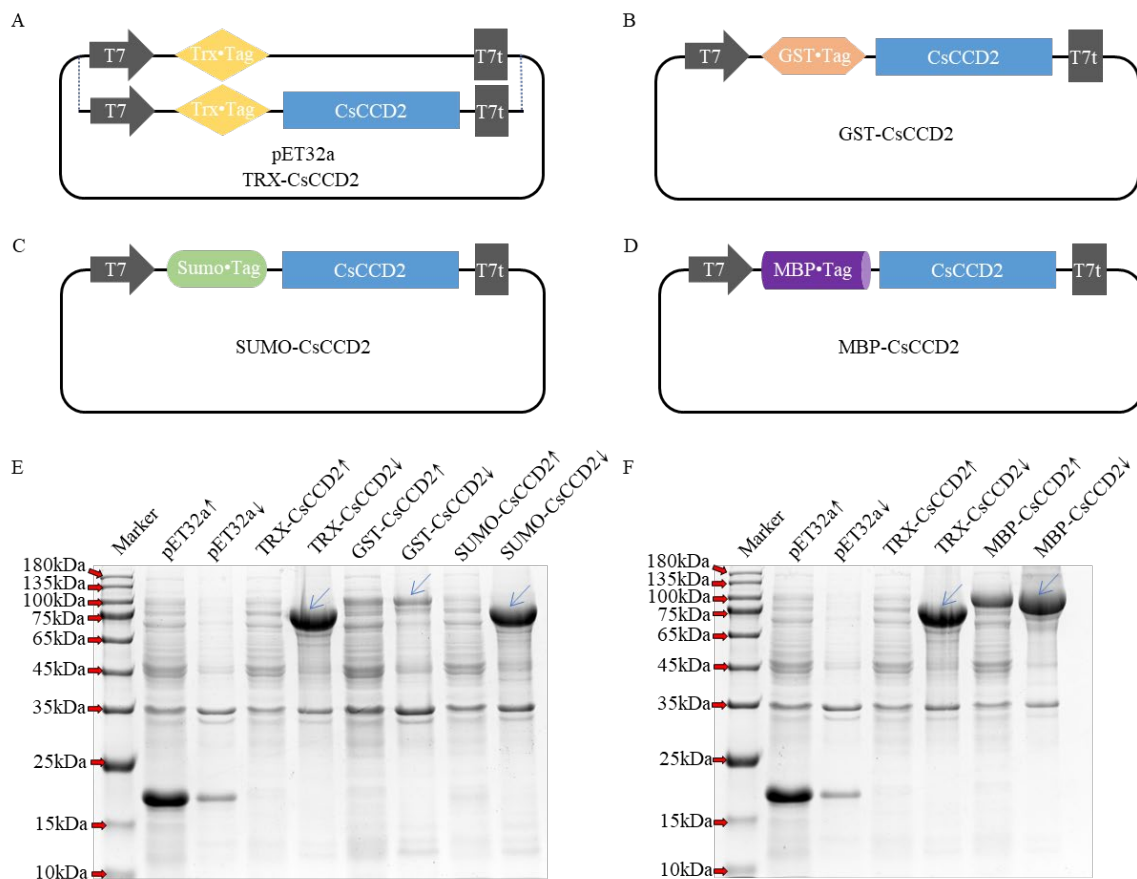[a] Amp$^R$, Kan$^R$, and Cm$^R$ represent ampicillin, kanamycin, and chloramphenicol, respectively.



**Figure 2.** Different expression patterns plasmid maps and its protein expression analysis. (**A–D**), five plasmid maps of pET32a, TRX-*Cs*CCD2, GST-*Cs*CCD2, Sumo-*Cs*CCD2 and MBP-*Cs*CCD2. (**E,F**), SDS-PAGE analysis of the *Cs*CCD2 with different fusion tags at 16 °C. ↑, indicates the protein supernatant. ↓, indicates the protein precipitate. The band of *Cs*CCD2 fusion protein was labeled with blue arrows in (**E,F**).
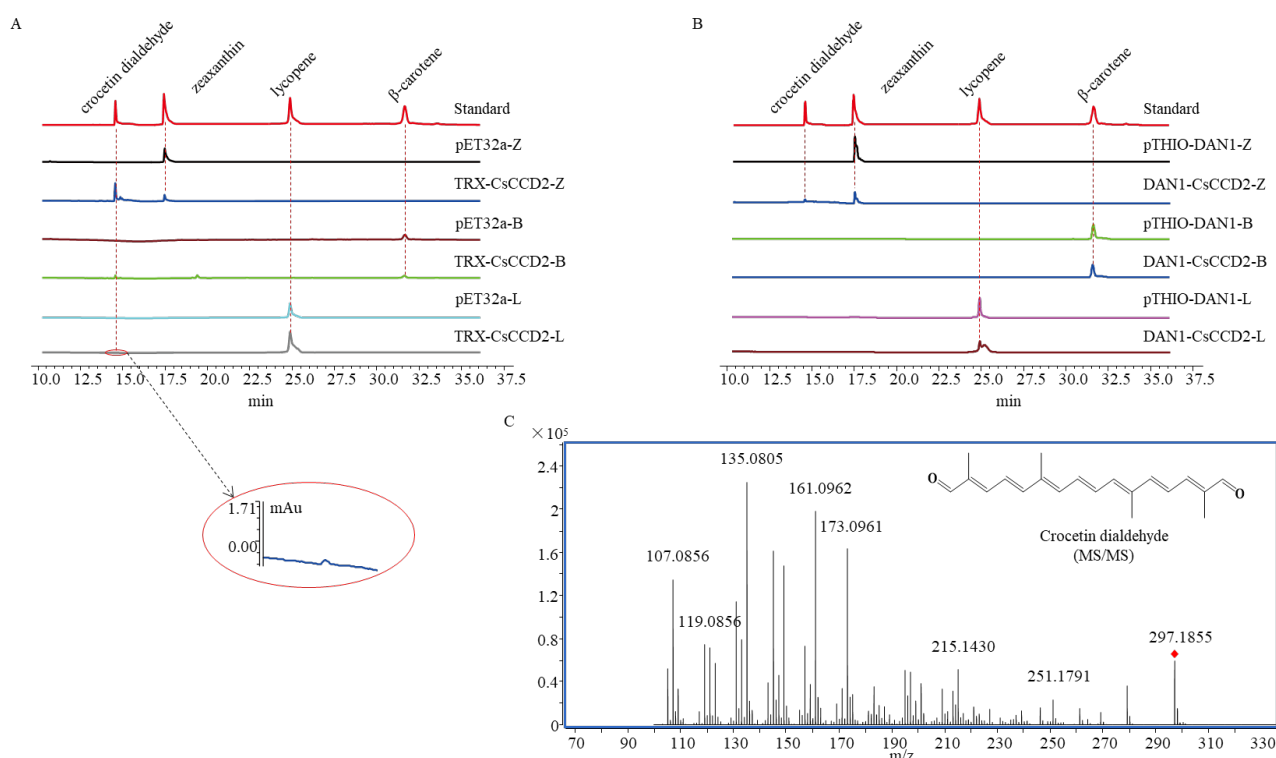
**Figure 3.** Functional identification of *Cs*CCD2. (**A**), UPLC-DAD chromatograms (abs at 440 nm) of pET32a-Z/B/L and TRX-*Cs*CCD2-Z/B/L extracts. (**B**), UPLC-DAD chromatograms (abs at 440 nm) of pTHIO-DAN1-Z/L/B and DAN1-*Cs*CCD2-Z/L/B extracts. (**C**), the MS/MS fragmentation pattern of crocetin dialdehyde. The red diamond represents the $[M+H]^+$ of crocetin dialdehyde.

## 2.2. Optimization of CsCCD2 Protein Expression

The low soluble expression of *Cs*CCD2 may be responsible for its poor activity or inactivation. To enhance the soluble expression of *Cs*CCD2, three different fusion expression plasmids were constructed, including GST-*Cs*CCD2, SUMO-*Cs*CCD2, and MBP-*Cs*CCD2 (Figure 2B–D). Additionally, different induction temperatures and IPTG concentrations were tested. Following 24 h of induction with 0.5 mM IPTG, the *Cs*CCD2 proteins were successfully expressed among the three vectors and displayed as 96 kDa, 75 kDa, and 104 kDa protein bands on sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE), respectively. At a 37 °C induction temperature, *Cs*CCD2 protein was only detected in the inclusion body section of all three protein expression patterns (Figure S2A,B). At a 28 °C induction temperature, *Cs*CCD2 protein was only detected in the inclusion body section of GST-*Cs*CCD2 and SUMO-*Cs*CCD2 expression patterns; however, it exhibited slightly soluble expression in MBP-*Cs*CCD2 (Figure S2C,D). When the induced temperature decreased to 16 °C, *Cs*CCD2 was only detected in the inclusion body section of SUMO-*Cs*CCD2. Notably, in GST-*Cs*CCD2 and MBP-*Cs*CCD2, *Cs*CCD2 was expressed with more soluble protein. The SDS-PAGE results demonstrated that MBP-*Cs*CCD2 had better solubility than GST-*Cs*CCD2 (Figure 2E,F). It can be inferred that a lower induction temperature is more suitable for CsCCD2 expression. Furthermore, the results of induction with different IPTG concentrations on MBP-*Cs*CCD2 at 16 °C indicated that 0.8 mM was the best concentration for *Cs*CCD2 expression (Figure S3). In conclusion, the best soluble expression pattern of *Cs*CCD2 was obtained from MBP-*Cs*CCD2 at an induction temperature of 16 °C with a 0.8 mM IPTG concentration for 24 h.

## 2.3. The Catalytic Activity Study of CsCCD2-Fusion Protein

To examine the catalytic activity of *Cs*CCD2-fusion protein, GST-*Cs*CCD2, SUMO-*Cs*CCD2, and MBP-*Cs*CCD2 were constructed and transformed into *E. coli* BL21 (DE3) harboring pACCAR25ΔcrtX, pACCAR16Δcrt, or pACCRT-EIB to form GST/SUMO/MBP-

*Cs*CCD2-Z/B/L engineered strains, respectively. After induction at 16 °C for 24 h, the UPLC results revealed that all engineered strains produced crocetin dialdehyde but with varying crocetin dialdehyde yields. MBP-*Cs*CCD2 exhibited higher conversion efficiency than GST-*Cs*CCD2 and SUMO-*Cs*CCD2 in vivo, which was consistent with the results of protein expression (Figure 4A–C). Thus, the use of the MBP tag can greatly improve the soluble expression of *Cs*CCD2 and result in higher CsCCD2 enzyme activity in the engineered *E. coli*.
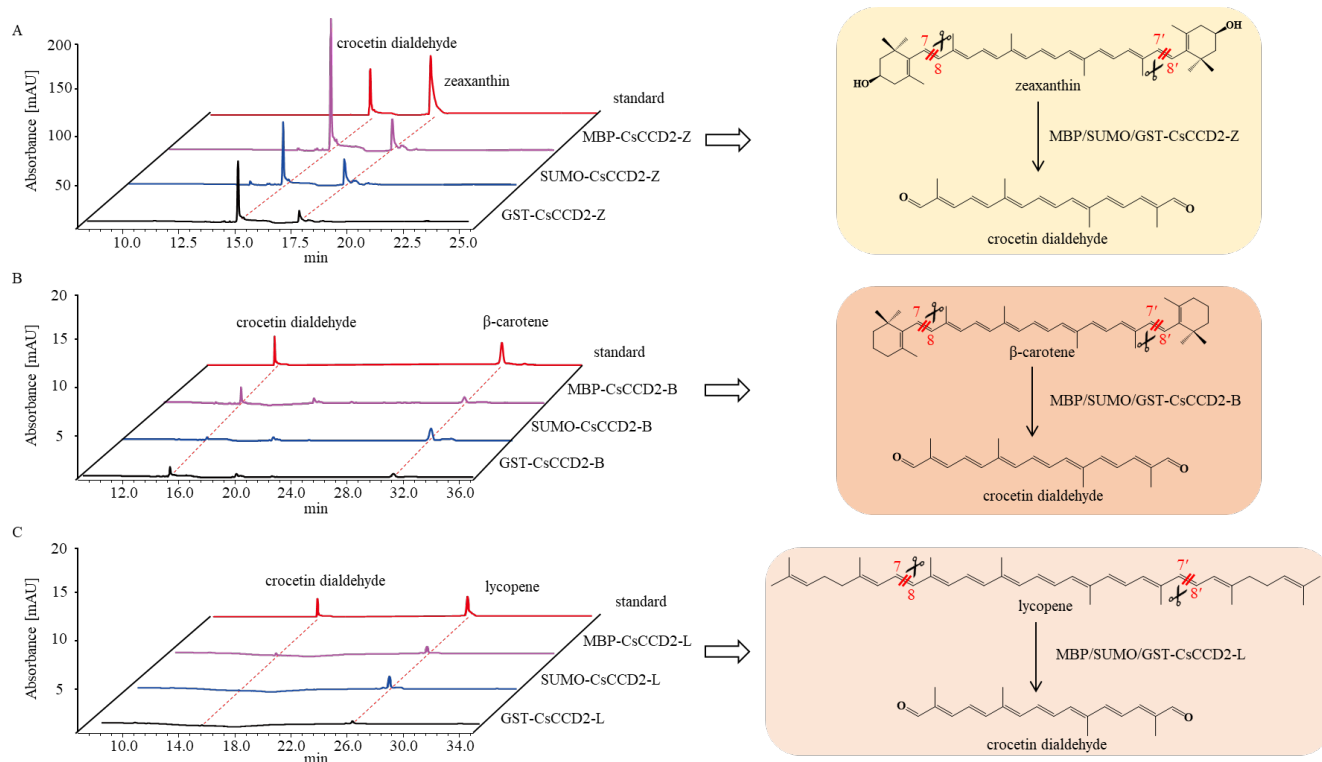


**Figure 4.** Functional characteristic of *Cs*CCD2-fusion protein. (**A**), UPLC detection of MBP/SUMO/GST-*Cs*CCD2-Z extracts at 440 nm. (**B**), UPLC detection of MBP/SUMO/GST-*Cs*CCD2-B extracts at 440 nm. (**C**), UPLC detection of MBP/SUMO/GST-*Cs*CCD2-L extracts at 440 nm.

### 2.4. Transforming the Catalytic Activity of CsCCD2 Using Tailored Truncation

In order to better understand the catalytic difference between TRX-*Cs*CCD2 and DAN1-*Cs*CCD2, a comparative analysis of the open reading frames (ORFs) of *Cs*CCD2, TRX-*Cs*CCD2, and DAN1-*Cs*CCD2 was conducted. There were two distinct variances between them, namely sections I and II (Figure 5). To the best of our knowledge, the redundant or missing amino acid sequence can have a significant impact on the three-dimensional structure of proteins, which in turn affects protein function. Thus, we truncated section II of DAN1-*Cs*CCD2 and obtained DAN1-*Cs*CCD2-T mutant (Figure 6A). Following functional characteristic analysis, DAN1-*Cs*CCD2-T indicated cleavable activity toward β-carotene and lycopene in addition to zeaxanthin. Furthermore, DAN1-*Cs*CCD2-T exhibited greater conversion efficiency than that observed before truncation (Figure 6B,C). This result indicated that the redundant amino acid sequences at the carbon-terminal end of CsCCD2 had a remarkable influence on its catalytic activity.
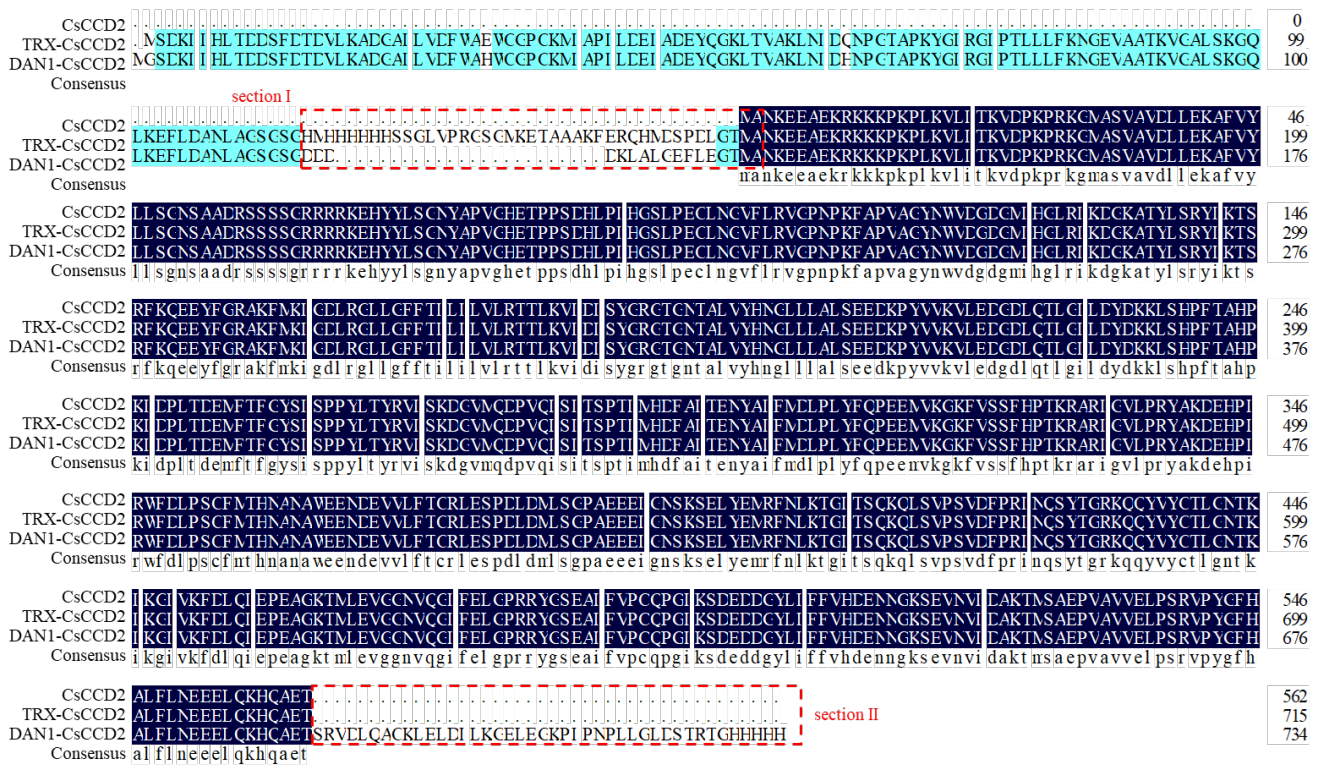
**Figure 5.** Amino acid sequence comparative analysis of *Cs*CCD2, TRX-*Cs*CCD2 and DAN1-*Cs*CCD2. Dark blue represents the same amino acid sequence among *Cs*CCD2, TRX-*Cs*CCD2 and DAN1-*Cs*CCD2. Wathet blue indicates the same amino acid sequence between TRX-*Cs*CCD2 and DAN1-*Cs*CCD2.
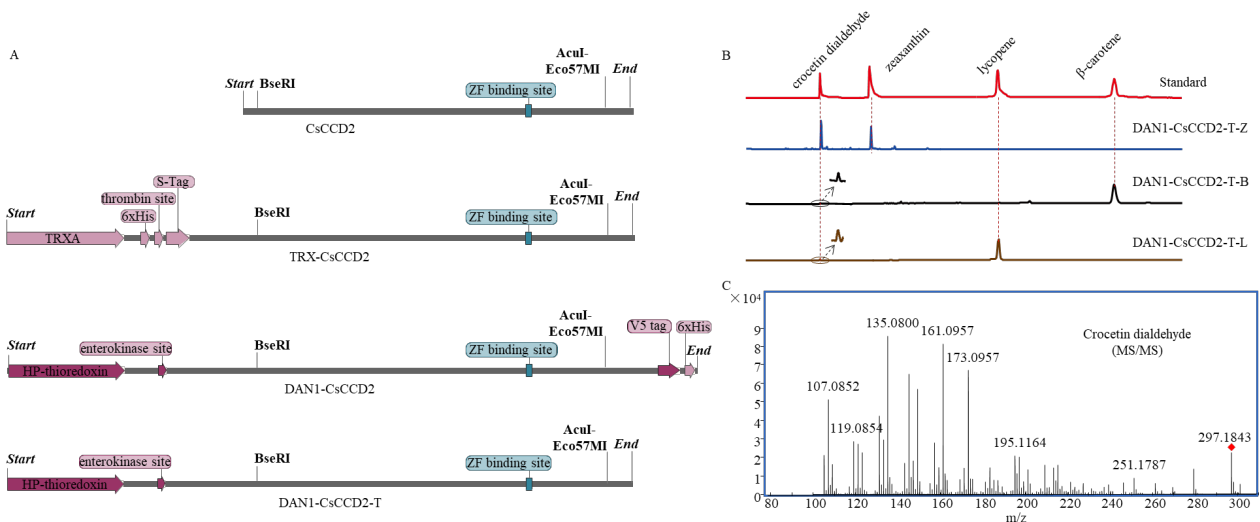


**Figure 6.** The truncation study of *Cs*CCD2. (**A**), the common features of *Cs*CCD2 ORF. (**B**), UPLC-DAD chromatograms (abs at 440 nm) of DAN1-*Cs*CCD2-T-Z/B/L extracts. (**C**), the MS/MS fragmentation pattern of crocetin dialdehyde. The red diamond represents the [M+H]$^+$ of crocetin dialdehyde.

## 3. Discussion

Crocins, a class of highly valuable apocarotenoids derived primarily from saffron, have significant pharmacological activity for treating human disorders [28]. *Cs*CCD2 is the rate-limiting enzyme involved in the biosynthetic pathway of crocins in saffron. It has been reported that *Cs*CCD2 can only cleave the 7,8 and 7′,8′ double bonds of zeaxanthin sequentially to produce crocetin dialdehyde, but not without the cleavage of β-carotene

and lycopene [8]. In this study, we replicated the functional assay of *Cs*CCD2 and obtained results consistent with those of previous investigations. Further research revealed that MBP-*Cs*CCD2 can effectively express its soluble form using *E. coli* as the host organism. Intriguingly, beta-carotene and lycopene, previously believed to be non-recognizable as substrates, were indeed accepted as substrates by *Cs*CCD2 in this study. What caused this difference?

First, in this study, the *Cs*CCD2 was subcloned into a pET32a prokaryotic expression vector to form TRX-*Cs*CCD2. The ORF in TRX-*Cs*CCD2 was TRXA•Tag-6×His-thrombin-S•Tag, which exhibited cleavage activity of zeaxanthin, β-carotene, and lycopene. However, in a previous report, researchers used the pTHIO-DAN1 prokaryotic expression vector to examine the functional characteristics of *Cs*CCD2. pTHIO-DAN1 is a derivative of pBAD/Thio (Invitrogen, Paisley, UK) carrying the pUC18 polylinker [27]. The ORF in DAN1-*Cs*CCD2 was HP-thioredoxin-enterokinase-*Cs*CCD2-V5•tag-6×His. The distinct open expression frame may account for the divergent functionality of *Cs*CCD2.

Second, to explain the divergent functionality of *Cs*CCD2 in the aforementioned two plasmids, an amino acid sequence comparative analysis was conducted, demonstrating two different sections between them, namely section I and section II (Figure 5). Thus, we hypothesized whether the amino acid sequence redundancy and deletion caused functional divergence. To verify this hypothesis, a DAN1-*Cs*CCD2 without section II was constructed, namely DAN1-*Cs*CCD2-T. Functional identification demonstrated that DAN1-*Cs*CCD2-T exhibited splitting activity on zeaxanthin, β-carotene, and lycopene. To some extent, the truncation study shed light on the molecular mechanism of functional divergence. This further suggests that using wild-type genes and thus avoiding additional modification sequences is better for conducting functional gene identification studies.

Third, as an important enzyme involved in the carotenoid cleavage process, *Cs*CCD2 exhibits relatively high specificity in substrate recognition and catalytic activity. The low catalytic efficiency of *Cs*CCD2 may also be attributed to its insoluble expression in *E. coli*. Thus, it is critical to investigate the soluble protein expression of *Cs*CCD2. Furthermore, a crystallographic structure analysis is the most direct and efficient way to gain a better understanding of the molecular catalytic mechanism of *Cs*CCD2. To the best of our knowledge, the crystal structures of only three CCD family members have been determined, including SynACO (4OU9) in *Synechocystis* sp. PCC 6803 [29], viviparous-14 (3NPE) in maize (*Zea mays*) [30], and NdCCD (6VCF) in *Nitrosotalea devanaterra* [25]. However, the crystal structure analysis of the CCD subfamily in the plant kingdom remains a significant challenge, and the insoluble expression of plant CCD limits its research progress.

Fusion expression is one of the most effective strategies, which can either increase the recombinant protein expression or participate in the protein-folding process, to enhance the soluble expression of recombinant protein [31,32]. Many studies have demonstrated that some highly soluble proteins promoted the soluble expression of fusion proteins after fusion [33]. GST and MBP are the most commonly used fusion tags that can improve fusion-protein solubility and enable one-step purification via affinity chromatography [34,35]. Furthermore, SUMO can improve protein solubility by facilitating proper protein folding and enhancing binding stability [36]. *Cs*CCD2 fusion proteins with three different fusion tags (GST, MBP, and SUMO) were constructed using the fusion expression technique, including GST-*Cs*CCD2, SUMO-*Cs*CCD2, and MBP-*Cs*CCD2. The optimization of protein expression demonstrated that MBP-*Cs*CCD2 achieved soluble expression at an induction temperature of 16 °C with 0.8 mM IPTG. Concurrently, the catalytic efficiency was enhanced as the protein solubility increased. This breakthrough in the soluble expression of *Cs*CCD2 will greatly promote the progress of CCD crystallographic structure studies in the plant kingdom.

## 4. Materials and Methods

### 4.1. Chemicals and Strains

Standards of lycopene (CAS, 502-65-8), β-carotene (CAS, 7235-40-7), zeaxanthin (144-68-3), and crocetin dialdehyde (CAS, 502-70-5) were purchased from Sigma-Aldrich (Sigma-Aldrich Corp., St. Louis, MO, USA). Phanta Max Super-Fidelity DNA Polymerase was purchased from Vazyme (Nanjing Vazyme Biotech Co., Ltd., Nanjing, China). A ClonExpress II One Step Cloning Kit C112 was purchased from Vazyme. IPTG was purchased from Solarbio (Beijing Solarbio Science & Technology Co., Ltd., Beijing, China). All restriction endonucleases were purchased from Takara (Takara Biomedical Technology (Beijing) Co., Ltd., Beijing, China). T4 DNA Ligase was purchased from Takara. A DNA plasmid isolation kit and DNA gel extraction kit were purchased from Tiangen (TIANGEN Biotech (Beijing) Co., Ltd., Beijing, China). Tryptone and yeast extract were purchased from Thermo (Thermo Fisher Scientific Inc., Waltham, MA, USA). *E. coli* DH5α and *E. coli* BL21 (DE3) were purchased from Tiangen. Z/B/L represent the BL21 (DE3) harboring pACCAR25ΔcrtX, pACCAR16Δcrt, or pACCRT-EIB, respectively. All the chemical reagents utilized in this experiment were of analytical grade.

### 4.2. Plasmid Construction and Culture Conditions

The *Cs*CCD2 (Genbank Accession Number: KJ541749.1) gene with a single-base mutation G1179T from *Crocus sativus* was synthesized by GenScript (GenScript Biotech Corp., Nanjing, China) and subcloned into a pET32a vector using KpnI and EcoRI restriction endonucleases (the G1179T single-base mutation is a synonymous mutation that aims to remove an EcoRI site from the coding sequence). In order to optimize *Cs*CCD2 protein soluble expression, pET41a-GST-F/R, pET28a-SUMO-F/R, and pET28a-MBP-F/R primers were used for the cloning of *Cs*CCD2, which carried corresponding restriction sites (Table 2). Next, the *Cs*CCD2 was subcloned into the pET41a (with GST tag) and the modified pET28a (with SUMO tag or MBP tag) prokaryotic expression vectors using digestion and ligation. To further delve into the reasons behind the non-conversion of DAN1-*Cs*CCD2, a truncated mutation study was carried out. CCD2-T-F and CCD2-T-R were used for the truncation study of CsCCD2 using one-step cloning. All the recombinant plasmids were confirmed using colony polymerase chain reaction and Sanger sequencing. *E. coli* DH5α was used for plasmid propagation and *E. coli* BL21 (DE3) was used for the expression of recombinant protein. Lysogeny broth medium (per liter: 10 g tryptone, 5 g yeast extract, and 10 g NaCl) was used for the propagation of *E. coli*.

**Table 2.** Primers used in this study.

| Primers | Sequence (5′–3′) | Restriction Endonuclease (Underlined) |
|---|---|---|
| pET41a-GST-F | CATG<u>CCATGG</u>GCGAAAACCTGTACTTTCAAGGCATGGCAAATAAGGAGGAG | NcoI |
| pET41a-GST-R | CCC<u>AAGCTT</u>TCATGTCTCTGCTTGGTGCTTCTG | HindIII |
| pET28a-SUMO-F | CCC<u>AAGCTT</u>CCGAAAACCTGTACTTTCAAGGCATGGCAAATAAGGAGGAG | HindIII |
| pET28a-SUMO-R | AAGGAAAAAA<u>GCGGCCGC</u>TCATGTCTCTGCTTGGTGCTTCTGAAGTTC | NotI |
| pET28a-MBP-F | CTA<u>GCTAGC</u>GAAAACCTGTACTTTCAAGGCCATATGATGGCAAATAAGGAGGAGGCAG | NheI |
| pET28a-MBP-R | CCC<u>AAGCTT</u>TCATGTCTCTGCTTGGTGCTTCTG | HindIII |
| CCD2-T-F | CCAAGCAGAGACATGAGTTTAAACGGTCTCCAGCTT | - |
| CCD2-T-R | ACTCATGTCTCTGCTTGGTGCTTCTGAAGTTCT | - |

### 4.3. The Function Assay of CsCCD2 in Bacteria

The TRX-*Cs*CCD2 vector was co-transformed with pACCAR25ΔcrtX, pACCRT-EIB, or pACCAR16Δcrt into *E. coli* BL21 (DE3). The engineered strains were precultured overnight, and 2 mL cultures were inoculated into 40 mL LB with 50 μg/mL ampicillin and 34 μg/mL chloramphenicol. After being grown at 37 °C and 200 rpm for approximately 3 h (OD$_{600}$ about 0.6), 1.0 M IPTG was added for a final concentration of 0.3 mM, and the culture was induced at 16 °C and 160 rpm for 24 h. The cultures were centrifuged at 4500× *g* for 10 min, and the pellets were ultrasonically extracted with 800 μL acetone to detect the catalyst.

The extract was filtered into vials for UPLC and UPLC-MS/MS analysis. Prof. Giovanni Giuliano generously contributed the plasmids pTHIO-DAN1 and DAN1-*Cs*CCD2. Their function assays were repeated as described [8]. Following co-transformation of pTHIO-DAN1 and DAN1-*Cs*CCD2 with pACCAR25ΔcrtX, pACCRT-EIB, or pACCAR16Δcrt into *E. coli* BL21 (DE3), the positive colony was cultured overnight. Two milliliters of cultures were inoculated into 50 mL of LB (containing 50 μg/mL ampicillin and 34 μg/mL chloramphenicol) and grown at 37 °C for 3 h to an $OD_{600}$ of 0.7. The cells were induced using 0.2% (wt/vol) arabinose for 16 h at 20 °C. Moreover, the fusion proteins (GST-*Cs*CCD2, SUMO-*Cs*CCD2, and MBP-*Cs*CCD2) were induced following the protocol of TRX-*Cs*CCD2 induction. Truncated protein (DAN1-*Cs*CCD2-T) was examined based on the methods of DAN1-*Cs*CCD2 induction.

### 4.4. UPLC and LC-MS/MS Analysis of the Product

The samples were analyzed using a Thermo Ultimate 3000 system (Thermo Fisher Scientific Inc., Waltham, MA, USA) equipped with a Waters Acquity UPLC® BEH C18 column (1.7 μm, 100 × 2.1 mm, Waters Co., Milford, MA, USA) at 35 °C with a wavelength detector set to 440 nm. The mobile phases of acetonitrile comprising 0.1% formic acid (A) and water comprising 0.1% formic acid (B) were used for UPLC. At a flow rate of 0.2 mL/min, the following gradient elution program was used: 0–8 min, linearly increasing from 10% A to 50% A; 8–12 min, linearly increasing from 50% A to 90% A; 12–13 min, linearly increasing from 90% A to 100% A; sustained 30 min.

The qualitative analysis of each product was performed using Agilent Technologies 1290 Infinity II and 6545 Q-TOF, together with Dual Agilent Jet Stream Electrospray Ionization sources (Agilent Technologies, Santa Clara, CA, USA). The drying gas was set at 350 °C and 8 L/min; the sheath gas was set at 350 °C, with a gas flow rate of 11 L/min. The nebulizer was set at 35 PSIG; the VCap was set at 3500 V. The data were analyzed using MassHunter (version B.07.00).

### 4.5. Optimization of CsCCD2 Protein Expression Pattern

The fusion expression strategy was adopted to enhance the soluble expression of *Cs*CCD2. Three expression vectors, including GST-*Cs*CCD2, SUMO-*Cs*CCD2, and MBP-*Cs*CCD2, were constructed as aforementioned. Furthermore, various inducing temperatures (37 °C, 28 °C, or 16 °C) and inducer concentrations (0.1 mM, 0.3 mM, 0.5 mM, 0.8 mM, 1.0 mM, 1.5 mM, and 2.0 mM) were also examined. When the $OD_{600}$ of engineered strains reached 0.6, 0.3 mM IPTG was added to induce protein expression at 160 rpm for 24 h. SDS-PAGE was used to evaluate the protein expression level. Subsequently, the function of fusion *Cs*CCD2 proteins was confirmed using an in vivo assay, and the product was detected as previously stated.

### 4.6. The Bioinformatic Analysis of CsCCD2, TRX-CsCCD2, and DAN1-CsCCD2

The ORFs of *Cs*CCD2, TRX-*Cs*CCD2, and DAN1-*Cs*CCD2 were precisely identified and translated into amino acids using MEGA (Version 6). Their amino acid sequences were further subjected to multiple sequence alignment via DNAMAN (Version 6) to predict differences in sequence. The common characteristics of the ORFs were predicted using SnapGene (Version 4.1.9). The three-dimensional structure of *Cs*CCD2 was predicted using the online Robetta service (https://robetta.bakerlab.org/, accessed on 22 May 2023).

## 5. Conclusions

In summary, this study demonstrated that *Cs*CCD2, a first key enzyme in saffron crocin biosynthesis, can catalyze not only the cleavage of zeaxanthin but also the cleavage of β-carotene and lycopene. The substrate profile of *Cs*CCD2 has been expanded to allow for the use of less-expensive raw materials (especially β-carotene) in crocin synthesis via substrate-feeding strategies. Additionally, the soluble expression of *Cs*CCD2 was determined

using *E. coli* as the host organism, paving the path for *Cs*CCD2 crystal structure resolution and facilitating the synthesis of crocetin and crocins using microbial fermentation.

## References

1.  Xu, Z.; Pu, X.; Gao, R.; Demurtas, O.C.; Fleck, S.J.; Richter, M.; He, C.; Ji, A.; Sun, W.; Kong, J.; et al. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. *BMC Biol.* **2020**, *18*, 63. [CrossRef] [PubMed]
2.  Khorasanchi, Z.; Shafiee, M.; Kermanshahi, F.; Khazaei, M.; Ryzhikov, M.; Parizadeh, M.R.; Kermanshahi, B.; Ferns, G.A.; Avan, A.; Hassanian, S.M. *Crocus sativus* a natural food coloring and flavoring has potent anti-tumor properties. *Phytomedicine* **2018**, *43*, 21–27. [CrossRef] [PubMed]
3.  Rahaiee, S.; Hashemi, M.; Shojaosadati, S.A.; Moini, S.; Razavi, S.H. Nanoparticles based on crocin loaded chitosan-alginate biopolymers: Antioxidant activities, bioavailability and anticancer properties. *Int. J. Biol. Macromol.* **2017**, *99*, 401–408. [CrossRef] [PubMed]
4.  Albalawi, G.A.; Albalawi, M.Z.; Alsubaie, K.T.; Albalawi, A.Z.; Elewa, M.A.F.; Hashem, K.S.; Al-Gayyar, M.M.H. Curative effects of crocin in ulcerative colitis via modulating apoptosis and inflammation. *Int. Immunopharmacol.* **2023**, *118*, 110138. [CrossRef] [PubMed]
5.  Fernandez-Albarral, J.A.; Ramirez, A.I.; de Hoz, R.; Lopez-Villarin, N.; Salobrar-Garcia, E.; Lopez-Cuenca, I.; Licastro, E.; Inarejos-Garcia, A.M.; Almodovar, P.; Pinazo-Duran, M.D.; et al. Neuroprotective and anti-inflammatory effects of a hydrophilic saffron extract in a model of glaucoma. *Int. J. Mol. Sci.* **2019**, *20*, 4110. [CrossRef] [PubMed]
6.  D'Onofrio, G.; Nabavi, S.M.; Sancarlo, D.; Greco, A.; Pieretti, S. *Crocus sativus* L. (Saffron) in Alzheimer's disease treatment: Bioactive effects on cognitive impairment. *Curr. Neuropharmacol.* **2021**, *19*, 1606–1616. [CrossRef] [PubMed]
7.  El Midaoui, A.; Ghzaiel, I.; Vervandier-Fasseur, D.; Ksila, M.; Zarrouk, A.; Nury, T.; Khallouki, F.; El Hessni, A.; Ibrahimi, S.O.; Latruffe, N.; et al. Saffron (*Crocus sativus* L.): A source of nutrients for health and for the treatment of neuropsychiatric and age-related diseases. *Nutrients* **2022**, *14*, 597. [CrossRef] [PubMed]
8.  Frusciante, S.; Diretto, G.; Bruno, M.; Ferrante, P.; Pietrella, M.; Prado-Cabrero, A.; Rubio-Moraga, A.; Beyer, P.; Gomez-Gomez, L.; Al-Babili, S.; et al. Novel carotenoid cleavage dioxygenase catalyzes the first dedicated step in saffron crocin biosynthesis. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 12246–12251. [CrossRef]
9.  Ahrazem, O.; Rubio-Moraga, A.; Nebauer, S.G.; Molina, R.V.; Gomez-Gomez, L. Saffron: Its phytochemistry, developmental processes, and biotechnological prospects. *J. Agric. Food Chem.* **2015**, *63*, 8751–8764. [CrossRef]
10. Luo, Y.; Li, B.Z.; Liu, D.; Zhang, L.; Chen, Y.; Jia, B.; Zeng, B.X.; Zhao, H.; Yuan, Y.J. Engineered biosynthesis of natural products in heterologous hosts. *Chem. Soc. Rev.* **2015**, *44*, 5265–5290. [CrossRef]
11. Paddon, C.J.; Westfall, P.J.; Pitera, D.J.; Benjamin, K.; Fisher, K.; McPhee, D.; Leavell, M.D.; Tai, A.; Main, A.; Eng, D.; et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **2013**, *496*, 528–532. [CrossRef] [PubMed]
12. Ro, D.K.; Paradise, E.M.; Ouellet, M.; Fisher, K.J.; Newman, K.L.; Ndungu, J.M.; Ho, K.A.; Eachus, R.A.; Ham, T.S.; Kirby, J.; et al. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006**, *440*, 940–943. [CrossRef] [PubMed]
13. Lau, W.; Sattely, E.S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **2015**, *349*, 1224–1228. [CrossRef] [PubMed]
14. Luo, X.; Reiter, M.A.; d'Espaux, L.; Wong, J.; Denby, C.M.; Lechner, A.; Zhang, Y.; Grzybowski, A.T.; Harth, S.; Lin, W.; et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **2019**, *567*, 123–126. [CrossRef] [PubMed]

15. Bouvier, F.; Suire, C.; Mutterer, J.; Camara, B. Oxidative remodeling of chromoplast carotenoids: Identification of the carotenoid dioxygenase *Cs*CCD and *Cs*ZCD genes involved in crocus secondary metabolite biogenesis. *Plant Cell* **2003**, *15*, 47–62. [CrossRef] [PubMed]

16. Pu, X.; He, C.; Yang, Y.; Wang, W.; Hu, K.; Xu, Z.; Song, J. In Vivo production of five crocins in the engineered *Escherichia coli*. *ACS Synth. Biol.* **2020**, *9*, 1160–1168. [CrossRef] [PubMed]

17. Ji, A.; Jia, J.; Xu, Z.; Li, Y.; Bi, W.; Ren, F.; He, C.; Liu, J.; Hu, K.; Song, J. Transcriptome-guided mining of genes involved in crocin biosynthesis. *Front. Plant Sci.* **2017**, *8*, 518. [CrossRef]

18. Walter, M.H.; Strack, D. Carotenoids and their cleavage products: Biosynthesis and functions. *Nat. Prod. Rep.* **2011**, *28*, 663–692. [CrossRef]

19. Pu, X.; Li, Z.; Tian, Y.; Gao, R.; Hao, L.; Hu, Y.; He, C.; Sun, W.; Xu, M.; Peters, R.J.; et al. The honeysuckle genome provides insight into the molecular mechanism of carotenoid metabolism underlying dynamic flower coloration. *New Phytol.* **2020**, *227*, 930–943. [CrossRef]

20. Ohmiya, A.; Kishimoto, S.; Aida, R.; Yoshioka, S.; Sumitomo, K. Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. *Plant Physiol.* **2006**, *142*, 1193–1201. [CrossRef]

21. Brandi, F.; Bar, E.; Mourgues, F.; Horváth, G.; Turcsi, E.; Giuliano, G.; Liverani, A.; Tartarini, S.; Lewinsohn, E.; Rosati, C. Study of 'Redhaven' peach and its white-fleshed mutant suggests a key role of CCD4 carotenoid dioxygenase in carotenoid and norisoprenoid volatile metabolism. *BMC Plant Biol.* **2011**, *11*, 24. [CrossRef] [PubMed]

22. Buah, S.; Mlalazi, B.; Khanna, H.; Dale, J.L.; Mortimer, C.L. The Quest for golden bananas: Investigating carotenoid regulation in a Fe'i group musa cultivar. *J. Agric. Food Chem.* **2016**, *64*, 3176–3185. [CrossRef] [PubMed]

23. Alder, A.; Jamil, M.; Marzorati, M.; Bruno, M.; Vermathen, M.; Bigler, P.; Ghisla, S.; Bouwmeester, H.; Beyer, P.; Al-Babili, S. The path from beta-carotene to carlactone, a strigolactone-like plant hormone. *Science* **2012**, *335*, 1348–1351. [CrossRef] [PubMed]

24. Liang, N.; Yao, M.D.; Wang, Y.; Liu, J.; Feng, L.; Wang, Z.M.; Li, X.Y.; Xiao, W.H.; Yuan, Y.J. *Cs*CCD2 access tunnel design for a broader substrate profile in crocetin production. *J. Agric. Food Chem.* **2021**, *69*, 11626–11636. [CrossRef] [PubMed]

25. Daruwalla, A.; Zhang, J.; Lee, H.J.; Khadka, N.; Farquhar, E.R.; Shi, W.; von Lintig, J.; Kiser, P.D. Structural basis for carotenoid cleavage by an archaeal carotenoid dioxygenase. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 19914–19925. [CrossRef]

26. Misawa, N.; Satomi, Y.; Kondo, K.; Yokoyama, A.; Kajiwara, S.; Saito, T.; Ohtani, T.; Miki, W. Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J. Bacteriol.* **1995**, *177*, 6575–6584. [CrossRef] [PubMed]

27. Trautmann, D.; Beyer, P.; Al-Babili, S. The ORF slr0091 of Synechocystis sp. PCC6803 encodes a high-light induced aldehyde dehydrogenase converting apocarotenals and alkanals. *FEBS J.* **2013**, *280*, 3685–3696. [CrossRef]

28. Liu, T.; Yu, S.; Xu, Z.; Tan, J.; Wang, B.; Liu, Y.G.; Zhu, Q. Prospects and progress on crocin biosynthetic pathway and metabolic engineering. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3278–3286. [CrossRef]

29. Kloer, D.P.; Ruch, S.; Al-Babili, S.; Beyer, P.; Schulz, G.E. The structure of a retinal-forming carotenoid oxygenase. *Science* **2005**, *308*, 267–269. [CrossRef]

30. Messing, S.A.; Gabelli, S.B.; Echeverria, I.; Vogel, J.T.; Guan, J.C.; Tan, B.C.; Klee, H.J.; McCarty, D.R.; Amzel, L.M. Structural insights into maize viviparous14, a key enzyme in the biosynthesis of the phytohormone abscisic acid. *Plant Cell* **2010**, *22*, 2970–2980. [CrossRef]

31. Esposito, D.; Chatterjee, D.K. Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol.* **2006**, *17*, 353–358. [CrossRef] [PubMed]

32. Waugh, D.S. Making the most of affinity tags. *Trends Biotechnol.* **2005**, *23*, 316–320. [CrossRef] [PubMed]

33. Eicholt, L.A.; Aubel, M.; Berk, K.; Bornberg-Bauer, E.; Lange, A. Heterologous expression of naturally evolved putative de novo proteins with chaperones. *Protein Sci.* **2022**, *31*, e4371. [CrossRef]

34. Smith, D.B.; Johnson, K.S. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* **1988**, *67*, 31–40. [CrossRef]

35. di Guan, C.; Li, P.; Riggs, P.D.; Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **1988**, *67*, 21–30. [CrossRef] [PubMed]

36. Butt, T.R.; Edavettal, S.C.; Hall, J.P.; Mattern, M.R. SUMO fusion technology for difficult-to-express proteins. *Protein Expres. Purif.* **2005**, *43*, 1–9. [CrossRef]