

# Structure elucidation of the novel carotenoid gemmatoxanthin from the photosynthetic complex of *Gemmatimonas phototrophica* AP64

**Nupur**

Institute of Microbiology, Center Algatech, Třeboň

**Marek Kuzma**

Institute of Microbiology of the Czech Acad. Sci., Prague

**Jan Hájek**

Institute of Microbiology, Center Algatech, Třeboň

**Pavel Hrouzek**

Institute of Microbiology, Center Algatech, Třeboň

**Alastair T. Gardiner**

Institute of Microbiology, Center Algatech, Třeboň

**Martin Lukeš**

Institute of Microbiology, Center Algatech, Třeboň

**Martin Moos**

Biology Centre of the Czech Acad. Sci., České Budějovice

**Petr Šimek**

Biology Centre of the Czech Acad. Sci., České Budějovice

**Michal Koblížek (✉ [koblizek@alga.cz](mailto:koblizek@alga.cz))**

Institute of Microbiology, Center Algatech, Třeboň

---

## Research Article

**Keywords:** Gemmatimonadetes, *Gemmatimonas phototrophica* AP64, phototrophic, carboxyl group, methoxy group, aldehyde group, linear C40 carotenoid backbone

**Posted Date:** April 3rd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-376473/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Scientific Reports on August 5th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-95254-6>.

# Abstract

*Gemmatimonas phototrophica* AP64 is the first phototrophic representative of bacterial phylum Gemmatimonadetes. The cells contain photosynthetic complexes with bacteriochlorophyll *a* as the main light-harvesting pigment. In addition, the complexes contain a carotenoid with a single broad absorption band at 490 nm in methanol. A combination of nuclear magnetic resonance, high-resolution mass spectroscopy and Fourier-transformed infra-red spectroscopy was used to determine the chemical structure of *G. phototrophica* light-harvesting carotenoid that we have named gemmatoxanthin. It is a novel linear carotenoid containing 11 conjugated double bonds and further substituted by methoxy, carboxyl and aldehyde group. Its IUPAC-IUBMB semi-systematic name is 1'-Methoxy-19'-oxo-3',4'-dihydro-7,8,1',2'-tetrahydro- $\Psi$ ,  $\Psi$  carotene-16-oic acid. To our best knowledge, the presence of the carboxyl, methoxy and aldehyde groups on a linear C40 carotenoid backbone is reported here for the first time.

## Introduction

Photosynthesis is an ancient process that probably evolved more than 3 billion years ago <sup>1</sup>. It is believed that the earliest phototrophic organisms were anoxygenic (not producing oxygen) species <sup>2</sup>, which throughout evolution diverged into a number of bacterial phyla. The latest phylum from which anoxygenic phototrophic species have been isolated is *Gemmatimonadetes*. The phylum *Gemmatimonadetes* was formally established in 2003, with *Gemmatimonas (G.) aurantiaca* as the type species. Even though there exists only a handful of cultured strains, Gemmatimonadetes are known to be relatively common organisms in many natural habitats, such as soils, sediments and freshwaters <sup>3,4</sup>. To date, two phototrophic species have been described in the genus *Gemmatimonas*: *G. phototrophica* AP64, originating from a freshwater lake in the Gobi Desert <sup>5</sup>, and *G. groenlandica* TET16 isolated from a freshwater stream in Greenland <sup>6</sup>.

So far most of the research focussed on phototrophy in *Gemmatimonadetes* has been conducted with the facultatively photoheterotrophic strain, *G. phototrophica* AP64. It requires a supply of organic carbon substrates for metabolism and growth. However, light energy through photosynthesis can supplement the metabolic requirements of the cell and improve the organic carbon growth efficiency <sup>7</sup>. The photosynthesis genes in *G. phototrophica* are organized in a 42.3-kb photosynthesis gene cluster (PGC), the organization of which closely resembles that of Proteobacteria and indicates that phototrophic *Gemmatimonadetes* received their photosynthesis genes horizontally from Proteobacteria <sup>8</sup>.

*G. phototrophica* contains bacteriochlorophyll (BChl) *a* as the main light-harvesting pigment, however, the cells are strongly red-pink pigmented due to presence of more than 10 different, as yet mostly unknown, carotenoids <sup>8</sup>. Some polar carotenoids present in *phototrophica* have been previously identified in the heterotrophic species *G. aurantiaca* T-27. These carotenoids were identified as oscillol 2,2'-dirhamnosides <sup>8,9</sup> and are unique to this bacterial genus. These carotenoids presumably have a

photoprotective role to prevent the formation of reactive oxygen species (ROS) <sup>10</sup> and may also stabilize bacterial membranes.

Recently, the photosynthetic complex of *G. phototrophica* AP64 has been purified and characterized <sup>11</sup>. The complex consists of a bacterial type 2 reaction center (RC) surrounded by two concentric rings of light-harvesting antenna. This unique organization ensures that the complex is highly efficient at harvesting photons and directing them to the RC <sup>11</sup>. The purified complex has good structural stability and contains BChl molecules with phytol or geranyl-geranyl sidechains. The complex also contains carotenoid, however, not an oscillol-rhamnoside derivative, rather a purple colored, as yet unknown carotenoid.

This study aimed to purify and spectroscopically characterize this unknown carotenoid to determine its chemical structure and propose a tentative biosynthetic pathway.

## Results

### Identification of the unknown carotenoid from purified photosynthetic complex

*G. phototrophica* AP64 cells contain seven major and several minor carotenoids (Fig. 1). The main carotenoids are (peak 1-2) are oscillol 2,2'-dirhamnosides. These polar carotenoids were identified earlier in *G. aurantiaca*<sup>3,9</sup>. During purification of the PS complex, the majority of carotenoids were removed through sucrose density gradient centrifugation, ion-exchange chromatography and gel filtration (Fig. S1). The purified photosynthetic complex was subjected to solvent extraction and HPLC so that an unknown carotenoid eluted at 11.4 min with a single absorption maxima of 490 nm (Fig. 1A).

This unknown carotenoid was isolated by preparative HPLC. The HRMS analysis of the purified carotenoid provided molecular ion at  $m/z$  613.4268 [M+H]<sup>+</sup> and 611.4108 [M-H]<sup>-</sup> (calculated. for C<sub>41</sub>H<sub>56</sub>O<sub>4</sub> + H<sup>+</sup>, -2.7 ppm and C<sub>41</sub>H<sub>55</sub>O<sub>4</sub><sup>-</sup>, 0.305 ppm) in positive and negative modes, respectively. This led to an unambiguous determination of the summary molecular formula C<sub>41</sub>H<sub>55</sub>O<sub>4</sub>.

The absorption spectrum of the unknown carotenoid in methanol has a single broad band with  $\lambda_{\max}$  at 490 nm. The solubility of the pure carotenoid in hexane was very limited. In contrast to the methanol spectrum, the spectrum in hexane has the three characteristic carotenoid transition peaks at (455), 485 and 516 nm (Fig. 1B). Since such spectral behavior is typical for carotenoids containing a carbonyl group, we tested its presence through chemical reduction using NaBH<sub>4</sub>. As expected, the spectrum of the reduced compound in methanol revealed the three peaks with absorption maxima at 422, 447, 475 nm and its fine structure (% III/II ratio) is 37% (Fig. 1B). It is the ratio of the absorption peak height from the trough between the longest-wavelength peak (III) to the middle wavelength peak (II). The molecular mass of the reduced compound increased to 2.0165 Da. This indicated the presence of a single carbonyl group connected to the system of conjugated double bonds.

Since the initial spectroscopic studies and MS data of an unknown purified carotenoid indicated that it may represent a novel chemical structure, we named this carotenoid gemmatoxanthin and conducted its full structural characterization.

## Nuclear magnetic resonance

Due to the limited amount of gemmatoxanthin it was not possible to acquire a  $^{13}\text{C}$  NMR spectrum directly. Therefore, *G. phototrophica* cells were grown on the  $^{13}\text{C}$ -labeled substrate and the experiment was conducted with  $^{13}\text{C}$  enriched gemmatoxanthin. The information about carbon chemical shifts had to be obtained indirectly by  $^1\text{H}$ - $^{13}\text{C}$  HSQC and  $^1\text{H}$ - $^{13}\text{C}$  HMBC correlation experiments (Table 1, Fig. S6-S7). The  $^1\text{H}$  NMR spectra are provided in the supplementary Fig. S2-S4.

The COSY spectrum (Fig. S5) in combination with  $^1\text{H}$ - $^{13}\text{C}$  HSQC (Fig. S6) allowed us to identify partial spin systems (Fig. 2A), namely one  $-\text{CH}_2-\text{CH}=\text{CH}-$ , three  $=\text{CH}-\text{CH}=\text{CH}-$ , one  $=\text{CH}-\text{CH}=\text{CH}-\text{CH}=\text{}$ , and two  $-\text{CH}_2-\text{CH}_2-\text{CH}=\text{}$ . Moreover, the  $^1\text{H}$  NMR also contained an aldehyde signal ( $\delta_{\text{C}}$  194.2,  $\delta_{\text{H}}$  9.539), signals of six methyl on a double bond, methoxyl ( $\delta_{\text{C}}$  48.6,  $\delta_{\text{H}}$  3.114) and one methyl of a double intensity ( $\delta_{\text{C}}$  24.7,  $\delta_{\text{H}}$  1.091). The  $^1\text{H}$ - $^{13}\text{C}$  HMBC (Fig. S7) further pointed to a carbon atom (C16) with a chemical shift 169.5, which indicates the presence of a carboxyl functional group, finding further confirmed by FTIR and MS measurements (see below). The identified structure fragments were partially interconnected by  $^1\text{H}$ - $^{13}\text{C}$  HMBC depicted in Fig. 2B. As the quaternary carbon C2 and carbon of methine C1 were not observable, we have performed an additional  $^1\text{H}$ - $^{13}\text{C}$  HMBC experiment using highly  $^{13}\text{C}$  labeled sample obtained by cultivation in SILEX media. As a result,  $^{13}\text{C}$  signals were obtained proving the connection of C1 and C2 at 130.9 and 137.0 ppm, respectively (Suppl. Fig. S8). This connection was further confirmed by the MS/MS data presented below.

The ROESY experiment was used to establish stereochemistry on the double bonds. The detected through space interactions are depicted in Fig. 2C. Unfortunately, the arrangement on the double bonds H-8' to H-12' was not approved.

## Mass spectrometry fragmentation analysis

The fragmentation experiment performed in the negative ion mode provided a loss of methoxy- group positioned at C1' (fragment X1<sup>-</sup>, Fig. 3). It was followed by a consecutive CO<sub>2</sub> loss that formed a fragment at  $m/z$  535.39. The decarboxylation was also observed directly from the molecular ion ( $m/z$  567.4213 - fragment X2<sup>-</sup>, Fig. 3) that established the presence of a carboxyl group on the NMR unassigned C16 carbon ( $^{13}\text{C}$ -shift 171.1 ppm). Further fragments X<sub>3</sub><sup>-</sup> to X<sub>7</sub><sup>-</sup> represented the cleavage of the aliphatic chain of the gemmatoxanthin molecule. Although the negative ion MS spectra did not provide a direct neutral loss of the formyl substituent, a consecutive loss of methoxy- (X<sub>1</sub><sup>+</sup>, Fig. 3) and formyl groups forming a fragment at  $m/z$  553.41 (X<sub>1</sub><sup>+</sup> → X<sub>3</sub><sup>+</sup>, Fig. 3) was observed in the positive ion MS spectrum. Both the gemmatoxanthin positive and negative ion spectra provided fragments corresponding

to the loss of a water molecule (at  $m/z$  563.3910 and 561.37438, respectively). However, the mechanism of water cleavage remains unexplained. Detailed information on the recorded ion fragments is provided in the Supplementary information (Tables S1 and S2).

### Fourier-transform infrared spectroscopy and Raman spectroscopy

The FTIR spectra of gemmatoxanthin in methanol show several absorbance peaks ranging from 4000  $\text{cm}^{-1}$  to 500  $\text{cm}^{-1}$  (Fig. 4). Peaks were identified using literature library data <sup>12-15</sup>. A broad absorbance band at a spectrum frequency of 3378  $\text{cm}^{-1}$  indicates the presence of a hydroxyl group <sup>12</sup>. An absorbance band at 1741  $\text{cm}^{-1}$  corresponds to a C=O stretch of carboxylic acid <sup>16</sup>. A band at 2954  $\text{cm}^{-1}$  corresponds to an asymmetric vibration of methyl groups together with a rather weak symmetric vibration at 2870  $\text{cm}^{-1}$ . Prominent bands at 2921  $\text{cm}^{-1}$  and 2851  $\text{cm}^{-1}$  correspond to asymmetric and symmetric vibrations of methylene groups <sup>12,13,17</sup>. The vibration band of C-H from the methoxy group (-O-CH<sub>3</sub>) is not visible as it could be overlapped by stronger signals originating from the aforementioned CH<sub>3</sub> and CH<sub>2</sub> vibrations <sup>12,14</sup>. The vibration visible as a shoulder at 1685  $\text{cm}^{-1}$  indicates the presence of conjugated aldehyde <sup>13</sup>. The symmetric 1577  $\text{cm}^{-1}$  and asymmetric 1540  $\text{cm}^{-1}$  stretchings of C=C confirms the conjugated system of double bonds in the carbon backbone of the carotenoid <sup>13</sup>. The absence of a vibration band at ~3006  $\text{cm}^{-1}$  proves that none of the double bonds in the system originate from a *cis* conformation (Fig. 4). The strong band of methylene/ methyl at 1477  $\text{cm}^{-1}$  together with weak methyl band at 1378  $\text{cm}^{-1}$  (umbrella deformation vibration of C-H of CH<sub>3</sub>) and band at 721  $\text{cm}^{-1}$  (methylene rocking vibration) is indicative for long-chain linear aliphatic structures <sup>17</sup>. A weak, but visible, band at 1259  $\text{cm}^{-1}$  originates from C-H rocking. A weak band at 1038  $\text{cm}^{-1}$  belongs to C-O bending <sup>18</sup>. A weak band at 1011  $\text{cm}^{-1}$  is designated as out-of-plane bending and C $\alpha$ =C $\alpha$  torsion of the backbone chain <sup>19</sup>.

The recorded Raman spectra contained three characteristic bands typical for carotenoids (Fig. S9). These three bands originate from (1) a rocking vibration of the molecule's methyl component (1015 and 1014  $\text{cm}^{-1}$ ) <sup>20</sup>, (2) from C-C single-bond stretch vibrations (1158  $\text{cm}^{-1}$ ) <sup>21,22</sup> and (3) C=C stretching vibrations (1525 and 1518  $\text{cm}^{-1}$ ) of the molecule's backbone <sup>20,22,23</sup>. Other Raman bands elicited by the lasers were at 910 and 908  $\text{cm}^{-1}$  and considered as C-C symmetric stretch <sup>24</sup>. A vibrational band appearing as a non-resolved band at 1176 and 1175  $\text{cm}^{-1}$  belonged to CH<sub>3</sub> rocking and a weak band at 1594 and 1589  $\text{cm}^{-1}$  was assigned to C=C stretching vibrations <sup>20</sup>.

## Discussion

In the presented study we have identified the structure of the main carotenoid present in the photosynthetic complex of *G. phototrophica* AP64. A combination of NMR and MS experiments has led to the unambiguous structure assignments presented in Figures 2 and 3, and having a IUPAC-IUBMB semi-systematic name 1'-Methoxy-19'-oxo-3',4'-didehydro-7,8,1',2'-tetrahydro- $\Psi,\Psi$  carotene-16-oic acid.

C40 carotenoids containing a carboxyl group or its variants are more common in the case of algae, fungi and yeast <sup>25</sup>. Only a limited number of photosynthetic prokaryotes are known to have carboxylic carotenoids, for instance, thiothece-484 from *Thiocystis gelatinosa* <sup>26</sup> or synechoxanthin from cyanobacterium *Synechococcus* sp. PCC 7002 <sup>27</sup>. The presence of a carbonyl group or hydroxyl group in the C40 linear carotenoids is commonly known and have been isolated from photosynthetic bacteria. However, the presence of a carboxylic group in the C40 linear carotenoids is rare <sup>25,28,29</sup>.

Gemmatoxanthin's unique structure is corroborated by a unique MS/MS APCI fragmentation pattern. The majority of carotenoids provide neutral losses, 92 and 106 Da, corresponding to the electrocyclic elimination of toluene and xylene, respectively <sup>30</sup>. This cleavage was not found in gemmatoxanthin and is very likely due to the presence of a formyl group at C9' of the conjugated backbone, resulting in a completely different fragmentation pathway. Secondly, although the molecule does not contain any free hydroxyl substituents, it loses a water molecule in both ionization modes and this neutral loss is likely associated with the carboxyl group fragmentation. The consecutive loss of water and CO was reported in the MS/MS spectrum of retinoic acid also lacking hydroxyl and featuring the plausible connection of the carboxyl group to the conjugated linear chain <sup>31</sup>. However, subsequent CO loss in the gemmatoxanthin MS/MS spectrum might be also interpreted as a loss of the formyl functional group. Finally, the CO<sub>2</sub> loss confirming the presence of the carboxylic group was detected only in the negative ion mode, probably due to the fact the carboxyl loss is manifested as formic acid HCOOH (H<sub>2</sub>O + CO) loss in the positive APCI spectrum as mentioned above.

Using infrared techniques such as FTIR, the key functional groups were identified. A methoxy group exhibits a weak band in the 2860 - 2800 cm<sup>-1</sup> region of the infrared spectrum, at a lower frequency than that of the main C-H absorption <sup>14</sup>. The strong absorption of CH<sub>3</sub> and CH<sub>2</sub> groups may overlap the absorption region for a methoxy group, which may hinder precise detection <sup>14</sup>. The key functional group vibration at 3378 cm<sup>-1</sup> corresponds to the hydroxyl group and a band at 1741 cm<sup>-1</sup> to the C=O stretch. Hence, the carboxyl group characteristic vibration has been proved by FTIR <sup>12</sup>. The absence of the *cis* C=C absorbance band in the region >3000 cm<sup>-1</sup> which usually absorbs at ~3006 cm<sup>-1</sup> <sup>32</sup> confirmed an all-trans conformation of the carotenoid. Therefore, based on the spectroscopic data analysis, we can conclude the gemmatoxanthin is a linear, all-trans, conjugated C<sub>40</sub> keto carotenoid containing formyl and a carboxylic group.

The presence of the carbonyl-containing carotenoids is randomly distributed in different families of anoxygenic phototrophic bacteria where keto- carotenoids are much more common than carotenoids containing an aldehyde group. Spheroidenone, one of the most studied ketocarotenoids is the main light harvesting carotenoid in phototrophic species belonging to the 'so called' Roseobacter clade, which represents a very common group of marine bacteria <sup>33</sup>. It was speculated that shift from spheroidene to spheroidenone, which occurs in *Rhodobacter sphaeroides* under aerobic conditions, may help to quench potentially dangerous singlet oxygen <sup>34</sup>. Similar characteristics may have also predicted in the

carotenoid bacteriorubixanthinal present in the photosynthetic complexes of *Erythrobacter* species<sup>35</sup>. Numerous studies have been published on the energy transfer efficiency of carbonyl carotenoids to the BChl *a* in photosynthetic complexes<sup>36,37</sup>. The energy transfer efficiency of rhodopinal to BChl *a* found to be ~100%<sup>38</sup> and similarly, for okenone the energy transfer efficiency is ~ 95 %<sup>39</sup>.

The biosynthetic pathway of the novel carotenoid is another important issue. As demonstrated previously, *G. aurantiaca* contains a biosynthetic pathway of (2S,2'S)-Oscillol 2,2'-di-( $\alpha$ -L-rhamnoside) with lycopene as its intermediate<sup>9</sup>. It is, therefore, plausible to assume that gemmatoxanthin synthesis begins from phytoene and lycopene (Fig. 5). Phytoene, the first C<sub>40</sub>-carotene, is converted to lycopene by incorporating double bonds. This 4-step reaction is carried out by phytoene dehydrogenase (phytoene desaturase) encoded by *crtI* (GEMMAAP\_12150) responsible for extending the conjugated double bonding. Lycopene undergoes hydroxylation and then oxidation at the C-19 position forming thus lycopene-19-al probably by the hydroxylase/oxidase activity of enzymes. This hypothesis is based on the fact that lycopene-19-al was first isolated from *Lamprocystis roseopersicina* belonging to the *Chromatiaceae* family and this pathway was previously tentatively predicted<sup>39,40</sup>. A similar pathway has also been predicted for lycopene-20-al, but as yet, no enzyme has been identified for carrying out the hydroxylation/oxidation reaction either at C-19 or the C-20 carbon position<sup>41</sup>. Furthermore, the addition of the terminal methoxy group is probably initiated by a carotenoid 1,2-hydratase encoded by *cruF* gene. Subsequent to this, the methyl is transferred by a carotenoid O-methyltransferase, which is likely encoded by *crtF* gene present in the PGC<sup>8</sup>. We can hypothesize that the methyl group of 1' carbon undergoes a hydroxylation and oxidation reaction yielding carboxylic acid, which is already known for synechoxanthin<sup>27</sup> and torulahadrin carotenoids<sup>42</sup> isolated from the unicellular Cyanobacterium *Synechococcus* sp. strain PCC 7002 and from the yeast *Rhodotorula*, respectively. Synechoxanthin is an aromatic carotenoid, whose pathway is derived from  $\beta$ -carotene. One of the intermediates of the synechoxanthin pathway consists of  $\beta,\chi$ -carotene ring at one end and renierapurpurin ( $\chi,\chi$ -carotene) at the other end of the carotenoid backbone. The methyl group of renierapurpurin is eventually oxidized to carboxylic acid by a hydroxylase/oxidase encoded by *cruH* gene<sup>27,43</sup>. In order to confirm that AP64 has similar enzymes carrying out a similar hydroxylation/oxidation reaction, we searched the CruH protein sequence against the genome of *G. phototrophica* AP64 (CP011454.1) using tblastn. We found two unidentified proteins showing 36.61% similarity (GEMMAAP\_15370) and 23.03% identity (GEMMAAP\_15125) that may be responsible for methyl conversion to carboxyl in the gemmatoxanthin pathway. Similarly, torularhodin is a C40 carotenoid having a linear carboxylic acid at one end and the other end consists of a 6-membered ring. Torulene, an initial precursor of the pathway, then undergoes hydroxylation and oxidation to yield carboxylic acid in a similar manner as described for synechoxanthin. The enzymes involved in the torularhodin pathway are, unfortunately, not known<sup>42,44</sup>.

In summary, gemmatoxanthin represents a novel natural compound. To our knowledge, this is the first linear C40 carotenoid that contains a carboxyl group at one end of the molecule, a methoxy substituent at the other end and a central formyl group connected to the system of conjugated double bonds, all of which strongly affects the spectral characteristics of the molecule.

# Methods

## Cultivation of *G. phototrophica* AP64

The strain *G. phototrophica* AP64<sup>5</sup> was grown on solid agar media containing per litre 0.5 g yeast extract, 0.5 g peptone, 0.2 g pyruvate, 0.2 g glucose, 1 g K<sub>2</sub>HPO<sub>4</sub>, pH 7.7, 15 g of Bacto™ Agar, and 1 mL of modified SL8 trace metal solution (190 µg CoCl<sub>2</sub>·6H<sub>2</sub>O, 5.2 mg Na<sub>2</sub>-EDTA, 24 µg NiCl<sub>2</sub>·6H<sub>2</sub>O, 17 µg CuCl<sub>2</sub>·2H<sub>2</sub>O, 70 µg ZnCl<sub>2</sub>, 1.8 mg SrCl<sub>2</sub>·6H<sub>2</sub>O, 20.3 mg MgCl<sub>2</sub>, 62 µg H<sub>3</sub>BO<sub>3</sub> per mL) and 1 mL vitamins (200 µg B1, 20 µg B3, 10 µg B7, 10 µg B12 per mL). Cells were streaked on agar media plates and incubated for approx. 2 weeks in microaerophilic conditions (10% O<sub>2</sub> +90% N<sub>2</sub>) at 28 ± 1°C, and pH-7.7 in the dark. The colonies were scraped from the agar media plates, using a plastic scraper, into a suitable tube, pelleted, washed and stored at -20°C until needed.

## Extraction and purification of photosynthetic membranes

The harvested cells (9-10 g wet) were re-suspended in 20 mM, Tris. Cl, 50 mM NaCl pH 8.0 and homogenized thoroughly with a few grains of DNase and a few mg of MgCl<sub>2</sub>. The cells were broken by passage three times through an Emulsiflex-CS cell disrupter and unbroken cells removed by low-speed centrifugation (10 min, 12,000 g, 4°C). The membranes were pelleted by ultra-centrifugation (180,000×g, 2 h at 4°C) and gently resuspended in 20 mM, Tris. Cl pH 8.0 and adjusted to an optical density (OD) = ~10 cm<sup>-1</sup> at the Qx band (~580). The re-suspended membranes were then solubilized with 2 % n-dodecyl β-D-maltoside (DDM), 0.2% Triton X-100 for 60 min, stirred at room temperature in the dark and then centrifuged to remove any un-solubilized material. The photosynthetic complex was then purified by layering on top of a stepwise sucrose gradient and run overnight (208,000×g, 16 h at 4°C). The gradient comprised sucrose dissolved in TD buffer (20 mM Tris.Cl pH 8.0, 0.02 % DDM) layered in 0.2 M steps (1.6 to 0.4 M). The following day the band containing photosynthetic complex was carefully removed from the gradient and loaded on to a gravity Q-Sepharose ion-exchange column. The column was pre-equilibrated with 20 mM, Tris. Cl pH 8.0 and after loading the sample was washed with copious amounts of TD buffer to remove sucrose and any non-specifically bound proteins. The photosynthetic complex was eluted by progressively increasing the NaCl concentration in the TD buffer. The resulting fractions were pooled and concentrated prior to gel filtration. Gel filtration was performed using an XK16/Superdex S300 column (GE Healthcare) at a flow rate of 0.5 mL min<sup>-1</sup> at 22°C. The collected fractions with the best optical ratio were pooled and concentrated as required. The final photosynthetic complex pool typically had an A817/A260 = ~ 1.4 – 1.3.

## Gemmatoxanthin purification and HPLC-PDA analysis

Pigments were extracted twice in 100% methanol (HPLC grade, VWR Czech Republic) from cells or purified complexes. The extracted pigments were centrifuged at top speed for 3 min in the Eppendorf centrifuge and the supernatant evaporated to dryness under a stream of nitrogen. The dried sample was dissolved in 50 µl of HPLC grade methanol mixed with 5 µl of 25% of ammonium acetate buffer and 20 µl

was injected into a SCL-40 HPLC (Nexera series, Shimadzu, Japan) equipped with SPD-M40 PDA detector. The pigments were separated using a reverse-phase analytical C8 column (Kinetex, 2.6  $\mu\text{m}$ , 100 $\text{\AA}$ , 150 x 4.6 mm column, Phenomenex, USA) heated at 40°C and the following mobile phase was employed: 25% 28 mM ammonium acetate in water with 75% methanol (solvent A) and 100% methanol (solvent B) at a flow rate of 0.8 ml/min. Gradient is as follows: A/B 0/100 (23 min), 0/100 (25 min), 100/0 (27 min), and 100/0 (28 min). Eluting pigments peaks were monitored at spectral range of 250-800 nm and was collected manually. Samples were handled in ice under the dim light condition to minimize its oxidation and isomerization.

## Spectroscopic analysis

Dry and purified carotenoid was dissolved in hexane or methanol, and its absorbance spectrum was measured and recorded in UV 2600 spectrophotometer (Shimadzu, Japan). The reduction was carried out by adding a few crystals of  $\text{NaBH}_4$  in purified carotenoid dissolved in 1 ml of methanol<sup>45</sup>. Absorbance spectra were measured from 250 to 700 nm, with a resolution of 0.5 nm in 1 cm quartz cuvettes against the pure solvent as a blank.

## HPLC-APCI-HRMS of carotenoids

Carotenoid samples were analysed on a Dionex UltiMate 3000 UHPLC+ (Thermo Scientific, Sunnyvale, CA, USA) equipped with a diode-array detector. Purity and determination of  $m/z$  of carotenoids was performed on a reversed phase C18 column (Kinetex, 150 x 4.6 mm, 2.6  $\mu\text{m}$ , Phenomenex, Torrance, CA, USA) using  $\text{H}_2\text{O}$  (A)/methanol (B) both containing 0.1%  $\text{HCOOH}$  as a mobile phase with the flow rate of 0.6 ml  $\text{min}^{-1}$ . The gradient was as follows: A/B 85/15 (0 min), 85/15 (in 1 min), 0/100 (in 20 min), 0/100 (in 25 min) and 85/15 (in 30 min). The chromatographic instrument and conditions for measuring positive and negative mass spectra were identical. The HPLC was connected to an Impact HD high-resolution mass spectrometer (Bruker, Billerica, Massachusetts, USA) atmospheric pressure chemical ionization operating in positive mode. The following APCI probe settings were used: drying temperature, 250°C; drying gas flow, 12 l  $\text{min}^{-1}$ ; nebulizer gas pressure, 3 bar; capillary voltage, 4.0 kV; endplate offset, 500 V. The APCI MS spectra were collected in the range  $m/z$  20–2000 with the spectral rate of 3 Hz. Collision energy was set from 28 to 35 eV. The calibration of the instrument was performed using sodium formate clusters at the beginning of each analysis. The formulas of obtained molecular peaks and fragments were calculated using Smart Formula in Bruker Compass Data Analysis software (version 5.1). Negative ion mass spectra were acquired by a Q Exactive Plus Orbitrap HRMS (Thermo Fisher Scientific, San Jose, CA, USA) equipped with an APCI ion source operated at -5 kV spray voltage; capillary temperature, 250°C, sheath gas at 30 au, aux gas at 10 au, spare gas at 0 au; probe temperature, 400°C and S-Lens level at 60 au. Full scan HRMS spectra were recorded in a mass range of  $m/z$  50-750 at 70,000 resolution (R,  $m/z$  200); scan rate,  $\pm$  3 Hz; automatic gain control (AGC) target,  $3 \times 10^6$  and maximum ion injection time (IT), 100 ms. The MS/MS experiments (611.4106 Da; 579.3844 Da) were conducted at R =17,500 ( $m/z$  200), automatic gain control target,  $2 \times 10^5$  and maximum ion injection time (IT), 100 ms; isolation window, 3  $m/z$ ; stepped normalized collision energy, 20, 35, 45. For the

accurate mass measurements, the lock mass at  $m/z$  301.9981 (2,4,6-Tris(trifluoromethyl)-1,3,5 triazene, 25  $\mu\text{mol L}^{-1}$ , 2  $\mu\text{l min}^{-1}$ ) was used. The data were processed using a 4.0 Xcalibur™ software (Thermo Fisher Scientific, San Jose, CA, USA).

### **Fourier Transform Infrared Spectroscopy (FTIR) and Raman Spectroscopy**

Purified carotenoid was dissolved in 10  $\mu\text{L}$  of methanol and 1  $\mu\text{L}$  was deposited onto ATR crystal via Hamilton syringe. After evaporation of the solvent, the Fourier transform spectra were obtained with a Nicolet IS10 (Thermo Nicolet) spectrometer equipped with Smart iTR accessory with the installed ZnSe ATR crystal. Absorbance spectra were collected in the spectral range from 600  $\text{cm}^{-1}$  to 4000  $\text{cm}^{-1}$ , at a spectral resolution of 4  $\text{cm}^{-1}$ ; 32 scans were co-added. A Norton-Beer apodization function was used, with a zero-filling factor of 2. The background was measured using pure methanol after evaporation. Omnic software (Nicolet) was used for measurement and data processing.

DXR2 Raman Microscope with 532 and 785 nm excitation (Thermo Fisher Scientific, USA) was employed where the confocal measurement was through the glass while direct measurement was performed on a thin film on a metal slide. The purified carotenoid sample was dissolved in methanol. We used 10 mW on the sample for 532 nm excitation laser, with laser spot approx. 1  $\mu\text{m}$ . Spectra were collected with one second exposure time and 300 exposures for one Raman spectrum. For 785 nm excitation laser, we used 30 mW on the sample with laser spot approx. 1.5  $\mu\text{m}$ . Spectra were collected with one second exposure time and 30 exposures for one Raman spectrum. Spectra were recorded over a wavenumber range of 50-3500  $\text{cm}^{-1}$ . Omnic software (Nicolet) was used for measurement and data processing.

### **Gemmatoxanthin purification for NMR analysis**

For NMR studies, the *G. phototrophica* AP64 cells were grown on modified agar media containing per liter: 0.5 g of  $^{13}\text{C}$  labelled glucose, 0.5 g yeast extract, 1 g  $\text{K}_2\text{HPO}_4$ , 15 g of Bacto™ Agar, and 1 mL of modified SL8 trace element solution as mentioned previously. The incorporation of  $^{13}\text{C}$  glucose displayed low labelling in the gemmatoxanthin. Therefore, in order to obtain highly substituted  $^{13}\text{C}$  gemmatoxanthin, the second batch of *G. phototrophica* AP64 cells was grown with  $^{13}\text{C}$  labelled SILEX *E. coli* media (0.5 g  $\text{L}^{-1}$ ) (Silantes GmbH, Munich, Germany) instead of  $^{13}\text{C}$  labelled glucose. Approx. 15 g (wet weight) of cells scraped from agar plates were used as the starting material for the extraction and purification of the complexes as mentioned earlier. Pigments were extracted in 100% methanol from the purified complexes until it becomes colorless and was centrifuged. The supernatant was concentrated under a nitrogen stream. Carotenoids were purified using a Shimadzu Prominence-*i* high-performance liquid chromatography system equipped with PDA detector. The purification of carotenoid was carried out into two steps: reverse phase semi-preparatory C8 column (250 × 10 mm, Luna 5  $\mu\text{m}$ , Phenomenex Inc., USA) heated at 40°C was eluted with the following mobile phases: 25 % HPLC water with 75% methanol (solvent A) and 100% methanol (solvent B) at a flow rate of 3.0  $\text{ml min}^{-1}$  using a gradient as follows: A/B 75/25 (0.01 min), 50/50 (in 16 min), 50/50 (in 23 min), 0/100 (in 40 min), 0/100 (in 42 min), 75/25 (in 43 min), 75/25 (in 56 min). Eluting pigments were monitored at 490 nm wavelength using PDA. The

gemmatoxanthin peak was collected and evaporated completely under the stream of nitrogen. In the second and final step of purification, partially purified and dried carotenoid sample was dissolved in methanol and was injected on HPLC semi-preparative phenyl column (8 × 250 mm, Reprosil 100 phenyl 5 μm, Watrex) eluted with HPLC grade water (A) and 100% methanol (B) at a flow rate of 2.0 ml min<sup>-1</sup> using following gradient: A/B 20/80 (0.01 min), 10/90 (10 min), 5/95 (20 min), 0/100 (21 min), 0/100 (26 min) and 20/80 (28 min).

## Nuclear Magnetic Resonance (NMR) analysis

Bruker Avance III 700 MHz spectrometer equipped with TCI CryoProbe (700.13 MHz for <sup>1</sup>H, 176.05 MHz for <sup>13</sup>C, Bruker Biospin GmbH, Rheinstetten, Germany) was used to acquire NMR spectra in DMSO-*d*<sub>6</sub> at 303 K. The solvent signals were used as an internal standard (*d*<sub>H</sub> 2.499 ppm and *d*<sub>C</sub> 39.46 ppm). The <sup>1</sup>H NMR, COSY, <sup>1</sup>H-<sup>13</sup>C HSQC, <sup>1</sup>H-<sup>13</sup>C HMBC, *J*-resolved, and ROESY spectra were measured using the standard manufacturer's software. Pre-saturation sequence was used to eliminate strong signals of residual signal in DMSO-*d*<sub>6</sub> and water signal in <sup>1</sup>H NMR, COSY, and <sup>1</sup>H-<sup>13</sup>C HSQC due to a low concentration of the sample. The <sup>1</sup>H NMR spectrum was zero-filled to 4-fold data points, and line broadening was applied (0.3 Hz) before the Fourier transformation. Protons were assigned by COSY, and the assignment was transferred to carbons by HSQC. The chemical shifts are given in the *d* scale (ppm) and coupling constants are given in Hz. The digital resolution (0.3 Hz for <sup>1</sup>H) allowed us to present the proton chemical shifts to three and coupling constant to one decimal places, respectively. The carbon chemical shifts are readouts from <sup>1</sup>H-<sup>13</sup>C HSQC and <sup>1</sup>H-<sup>13</sup>C HMBC and are reported to one decimal place.

## Declarations

### Acknowledgements

This research was supported by the *PhotoGem+* project No. 19-28778X financed by the Czech Science Foundation (Nupur, ATG, JH, MK). The post-doctoral fellowship of Nupur was supported under the PPLZ program by the Czech Academy of Sciences. Authors thank Dr. Jan Pásztor and Jan Šec from Nicolet CZ s.r.o. Prague, for providing Raman spectra of gemmatoxanthin.

### Author contributions

Nupur, M.Kuzma, K.H., J.H., A.T.G., M.L., M.M., P.Š., performed research. Nupur wrote the paper with the help of P.H. M.Kuzma and M.Koblížek. P.H and M.Koblížek conceptualized the research.

### Additional information

The authors declare no competing interests.

## References

- 1 Blankenship, R. E. Early evolution of photosynthesis. *Plant Physiol***154**, 434-438, doi:10.1104/pp.110.161687 (2010).
- 2 Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of photosynthesis. *Annu Rev Plant Biol***62**, 515-548, doi:10.1146/annurev-arplant-042110-103811 (2011).
- 3 Zeng, Y. *et al.* Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environmental Microbiology Reports***8**, 139-149, doi:10.1111/1758-2229.12363 (2016).
- 4 Mujakić, I. *et al.* Common Presence of Phototrophic *Gemmatimonadota* in Temperate Freshwater Lakes. *mSystems***6**, e01241-01220, doi:10.1128/mSystems.01241-20 (2021).
- 5 Zeng, Y. *et al.* Characterization of the microaerophilic, bacteriochlorophyll a-containing bacterium *Gemmatimonas phototrophica* sp. nov., and emended descriptions of the genus *Gemmatimonas* and *Gemmatimonas aurantiaca*. *Int J Syst Evol Microbiol***65**, 2410-2419, doi:10.1099/ij.s.0.000272 (2015).
- 6 Zeng, Y. *et al.* *Gemmatimonas groenlandica* sp. nov. Is an Aerobic Anoxygenic Phototroph in the Phylum Gemmatimonadetes. *Frontiers in Microbiology***11**, doi:10.3389/fmicb.2020.606612 (2021).
- 7 Koblížek, M. *et al.* Utilization of light energy in phototrophic Gemmatimonadetes. *J Photochem Photobiol B***213**, 112085, doi:10.1016/j.jphotobiol.2020.112085 (2020).
- 8 Zeng, Y., Feng, F. Y., Medová, H., Dean, J. & Koblížek, M. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *P Natl Acad Sci USA***111**, 7795-7800 (2014).
- 9 Takaichi, S., Maoka, T., Takasaki, K. & Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2'-dirhamnoside. *Microbiology***156**, 757-763, doi:10.1099/mic.0.034249-0 (2010).
- 10 Yurkov, V. & Csotonyi, J. T. in *The Purple Phototrophic Bacteria* (eds C. Neil Hunter, Fevzi Daldal, Marion C. Thurnauer, & J. Thomas Beatty) 31-55 (Springer Netherlands, 2009).
- 11 Dachev, M. *et al.* Unique double concentric ring organization of light harvesting complexes in *Gemmatimonas phototrophica*. *Plos Biol***15**, e2003943, doi:10.1371/journal.pbio.2003943 (2017).
- 12 Coates, J. in *Encyclopedia of Analytical Chemistry* (2006).
- 13 Lorand, T., Deli, J., Molnar, P. & Toth, G. FT-IR study of some carotenoids. *Helv Chim Acta***85**, 1691-1697, doi:10.1002/1522-2675(200206)85:6<1691::Aid-Hlca1691>3.0.Co;2-G (2002).

- 14 Degen, I. A. Detection of Methoxyl Group by Infrared Spectroscopy. *Appl Spectrosc***22**, 164-166, doi:Doi 10.1366/000370268774383444 (1968).
- 15 Colthup, N. B., Daly, L. H. & Wiberley, S. E. in *Introduction to Infrared and Raman Spectroscopy (Third Edition)* (eds N. B. Colthup, L. H. Daly, & S. E. Wiberley) 327-337 (Academic Press, 1990).
- 16 Jagannadham, M. V., Rao, V. J. & Shivaji, S. The Major Carotenoid Pigment of a Psychrotrophic *Micrococcus-Roseus* Strain - Purification, Structure, and Interaction with Synthetic Membranes. *J Bacteriol***173**, 7911-7917, doi:DOI 10.1128/jb.173.24.7911-7917.1991 (1991).
- 17 Norman B. Colthup, L. H. D., Stephen E. Wiberley,. in *Introduction to Infrared and Raman Spectroscopy* 327-337 (Academic Press, 1990).
- 18 Fujioka, N., Morimoto, Y., Arai, T. & Kikuchi, M. Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy. *Cancer Detect Prev***28**, 32-36, doi:10.1016/j.cdp.2003.11.004 (2004).
- 19 Schulz, H. & Baranska, M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib Spectrosc***43**, 13-25, doi:10.1016/j.vibspec.2006.06.001 (2007).
- 20 Tschirner, N. *et al.* Resonance Raman spectra of beta-carotene in solution and in photosystems revisited: an experimental and theoretical study. *Phys Chem Chem Phys***11**, 11471-11478, doi:10.1039/b917341b (2009).
- 21 Edwards, H. G. M., Moody, C. D., Villar, S. E. J. & Wynn-Williams, D. D. Raman spectroscopic detection of key biomarkers of cyanobacteria and lichen symbiosis in extreme Antarctic habitats: Evaluation for Mars Lander missions. *Icarus***174**, 560-571, doi:10.1016/j.icarus.2004.07.029 (2005).
- 22 Jehlička, J. *et al.* Potential and limits of Raman spectroscopy for carotenoid detection in microorganisms: implications for astrobiology. *Philos T R Soc A***372**, doi:ARTN 20140199 10.1098/rsta.2014.0199 (2014).
- 23 Shao, Y. N., Gu, W. M., Jiang, L. J. & Zhu, Y. M. Study on the Visualization of Pigment in *Haematococcus pluvialis* by Raman Spectroscopy Technique. *Sci Rep-Uk***9**, doi:ARTN 12097 10.1038/s41598-019-47208-2 (2019).
- 24 Ferreira, G. B., Comerlato, N. M., Wardell, J. L. & Hollauer, E. Vibrational spectra of bis(dmit) complexes of main group metals: IR, Raman and *ab initio* calculations. *J Brazil Chem Soc***15**, 951-963, doi:Doi 10.1590/S0103-50532004000600025 (2004).
- 25 Yabuzaki, J. Carotenoids Database: structures, chemical fingerprints and distribution among organisms. *Database***2017**, doi:10.1093/database/bax004 (2017).

- 26 Andrewes, A. G. & Liaaen-Jensen, S. Bacterial carotenoids. XXXVII. Carotenoids of thiorhodaceae 9. Structural elucidation of five minor carotenoids from *thiothece gelatinosa*. *Acta Chem Scand***26**, 2194-2204, doi:10.3891/acta.chem.scand.26-2194 (1972).
- 27 Graham, J. E., Lecomte, J. T. J. & Bryant, D. A. Synechoxanthin, an Aromatic C40 Xanthophyll that Is a Major Carotenoid in the *Cyanobacterium Synechococcus* sp. PCC 7002. *Journal of Natural Products***71**, 1647-1650, doi:10.1021/np800310b (2008).
- 28 Nupur *et al.* ProCarDB: a database of bacterial carotenoids. *BMC Microbiology***16**, 96, doi:10.1186/s12866-016-0715-6 (2016).
- 29 Britton, G., Liaaen-Jensen, S. & Pfander, H. in *Carotenoids* 271-280.272 (Birkhäuser Basel, 2004).
- 30 Neto, F. C. *et al.* Re-investigation of the fragmentation of protonated carotenoids by electrospray ionization and nanospray tandem mass spectrometry. *Rapid Commun Mass Spectrom***30**, 1540-1548, doi:10.1002/rcm.7589 (2016).
- 31 Kane, M. A. & Napoli, J. L. Quantification of endogenous retinoids. *Methods Mol Biol***652**, 1-54, doi:10.1007/978-1-60327-325-1\_1 (2010).
- 32 Vlachos, N. *et al.* Applications of Fourier transform-infrared spectroscopy to edible oils. *Anal Chim Acta***573-574**, 459-465, doi:10.1016/j.aca.2006.05.034 (2006).
- 33 Koblížek, M., Zeng, Y., Horák, A. & Oborník, M. Regressive evolution of photosynthesis in the Roseobacter clade. *Advances in Botanical Research* **66**, 385-405, doi:10.1016/B978-0-12-397923-0.00013-8. (2013).
- 34 Šlouf, V. *et al.* Photoprotection in a purple phototrophic bacterium mediated by oxygen-dependent alteration of carotenoid excited-state properties. *P Natl Acad Sci USA***109**, 8570-8575, doi:10.1073/pnas.1201413109 (2012).
- 35 Šlouf, V. *et al.* Carotenoid Charge Transfer States and Their Role in Energy Transfer Processes in LH1-RC Complexes from Aerobic Anoxygenic Phototrophs. *The Journal of Physical Chemistry B***117**, 10987-10999, doi:10.1021/jp309278y (2013).
- 36 Khan, T., Litvín, R., Šebelík, V. & Polívka, T. Excited State Evolution of Keto-carotenoids after Excess Energy Excitation in the UV Region. *Chemphyschem*, doi:10.1002/cphc.202000982 (2020).
- 37 Šlouf, V. *et al.* Carotenoid to bacteriochlorophyll energy transfer in the RC-LH1-PufX complex from *Rhodobacter sphaeroides* containing the extended conjugation keto-carotenoid diketospirilloxanthin. *Photosynthesis research***135**, 33-43, doi:10.1007/s11120-017-0397-4 (2018).
- 38 Magdaong, N. M. *et al.* High efficiency light harvesting by carotenoids in the LH2 complex from photosynthetic bacteria: unique adaptation to growth under low-light conditions. *J Phys Chem B***118**,

11172-11189, doi:10.1021/jp5070984 (2014).

39 Ola Andersson, P., Cogdell, R. J. & Gillbro, T. Femtosecond dynamics of carotenoid-to-bacteriochlorophyll a energy transfer in the light-harvesting antenna complexes from the purple bacterium *Chromatium purpuratum*. *Chemical Physics***210**, 195-217, doi:[https://doi.org/10.1016/0301-0104\(96\)00172-3](https://doi.org/10.1016/0301-0104(96)00172-3) (1996).

40 Pfennig, N., Markham, M. C. & Liaaen-Jensen, S. Carotenoids of Thiorhodaceae. *Archiv für Mikrobiologie***62**, 178-191, doi:10.1007/BF00410404 (1968).

41 Takaichi, S. in *The Photochemistry of Carotenoids. Advances in Photosynthesis and Respiration* Vol. 8 (ed Young A.J. Frank H.A., Britton G., Cogdell R.J. ) (Springer, Dordrecht, 1999).

42 Kot, A. M., Błażejczak, S., Kurcz, A., Gientka, I. & Kieliszek, M. *Rhodotorula glutinis*—potential source of lipids, carotenoids, and enzymes for use in industries. *Applied Microbiology and Biotechnology***100**, 6103-6117, doi:10.1007/s00253-016-7611-8 (2016).

43 Graham, J. E. & Bryant, D. A. The Biosynthetic Pathway for Synechoxanthin, an Aromatic Carotenoid Synthesized by the Euryhaline, Unicellular *Cyanobacterium Synechococcus* sp. Strain PCC 7002. *J Bacterio***190**, 7966-7974, doi:10.1128/jb.00985-08 (2008).

44 Moliné, M., Libkind, D. & van Broock, M. in *Microbial Carotenoids From Fungi: Methods and Protocols* (ed José-Luis Barredo) 275-283 (Humana Press, 2012).

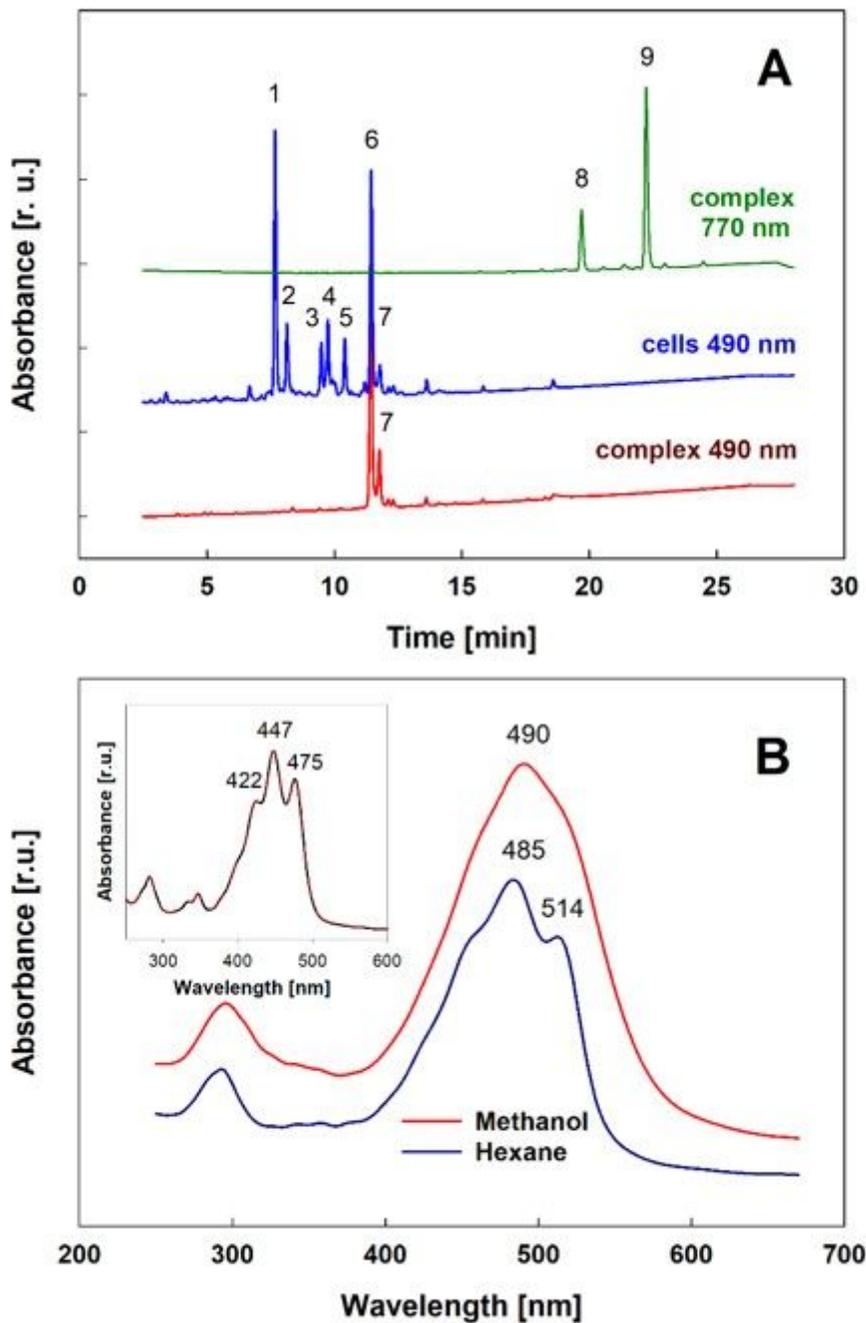
45 Takaichi, S. & Shimada, K. in *Methods in Enzymology* Vol. 213 374-385 (1992).

## Table

**Table 1:** The  $^1\text{H}$  and  $^{13}\text{C}$  NMR data of gemmatoxanthin (1) (700.13 MHz for  $^1\text{H}$ , 176.05 for  $^{13}\text{C}$ , DMSO- $d_6$ , 30 °C). \* The data was obtained from highly substituted  $^{13}\text{C}$  gemmatoxanthin.

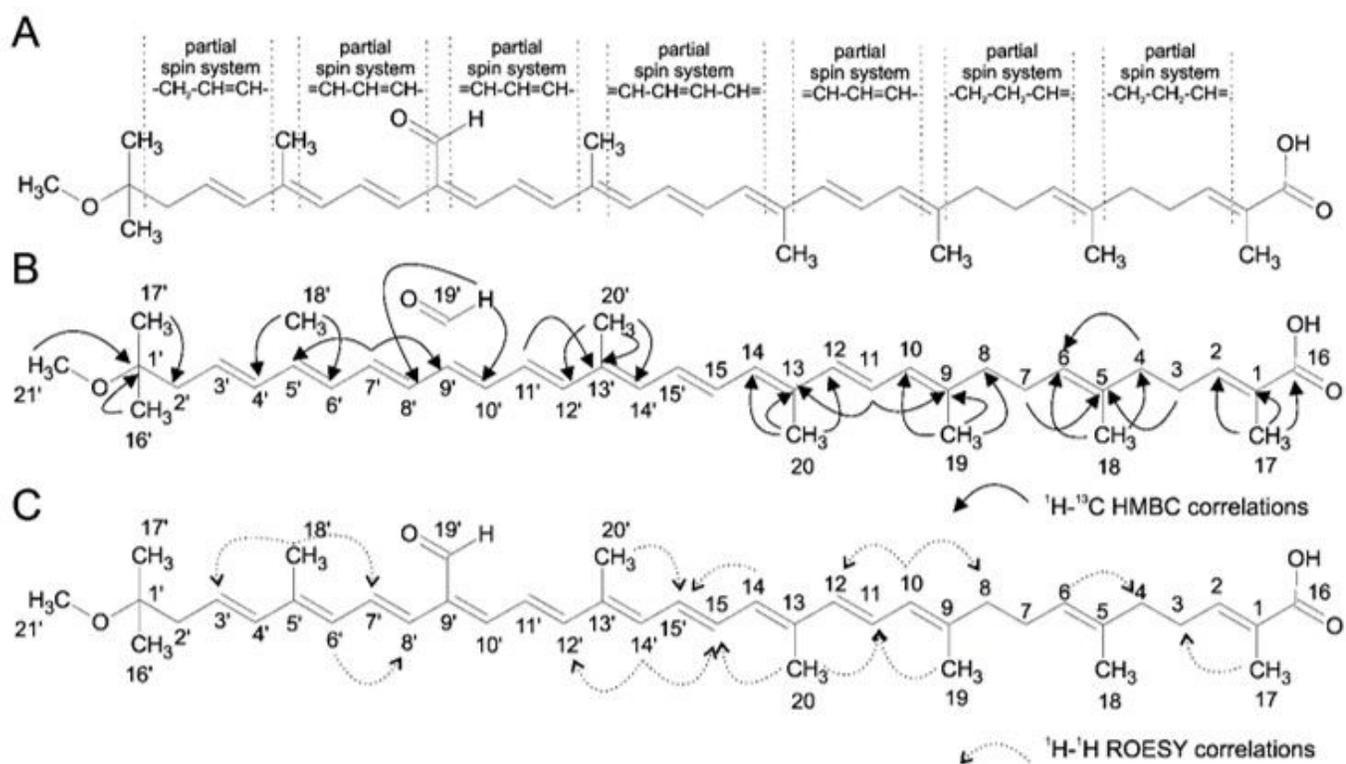
Atom #	$\delta_C$	m.	$\delta_H$	m	$J_{HH}$ [Hz]
1	n.d.				
2	n.d.		6.499	br s	
3	26.7	t	2.182	m	-
4	38.0	t	2.033	m	-
5	134.2	s	-		
6	123.9	d	5.126	m	-
7	26.0	t	2.107	m	-
8	39.6	t	2.099	m	-
9	140.0	s	-		
10	125.7	d	5.970	d	10.9
11	126.0	d	6.570	dd	10.9, 15.1
12	134.9	d	6.281	d	15.1
13	138.4	s	-		
14	131.2	d	6.318	d	11.8
15	133.6	d	6.898	m	-
16	169.5*	s*	-		
17	12.6	q	1.697	s	-
18	15.7	q	1.591	s	-
19	16.7	q	1.802	s	-
20	12.6	q	1.952	s	-
1'	74.3	s	-		
2'	43.1	t	2.292	d	7.4
3'	126.9	d	5.771	dt	7.4, 15.5
4'	136.6	d	6.186	d	15.5
5'	136.7	s	-		
6'	131.0	d	6.219	d	11.4
7'	130.1	d	7.675	dd	11.4, 15.2
8'	123.0	d	6.752	d	15.2
9'	133.0	s	-		
10'	148.9	d	7.055	d	11.5
11'	122.8	d	7.013	dd	11.5, 13.9
12'	147.1	d	6.943	d	13.9
13'	135.4	s	-		
14'	138.3	d	6.627	d	11.6
15'	129.3	d	6.743	m	-
16', 17'	24.7	q	1.091	s	-
18'	12.8	q	1.884	s	-
19'	194.2	d	9.539	d	2.1
20'	12.4	q	2.014	s	-
21'-OMe	48.6	q	3.114	s	-

## Figures



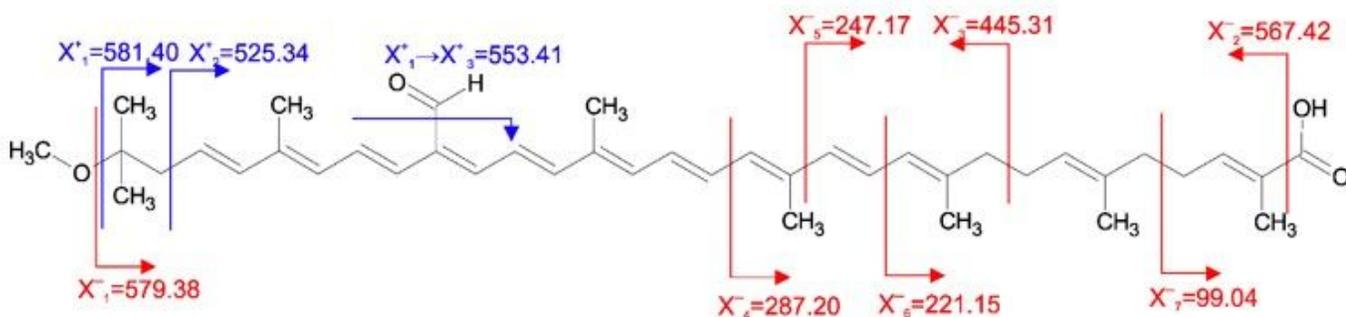
**Figure 1**

A: HPLC elution profile of *Gemmatimonas phototrophica* AP64 whole-cell extract (green and blue trace) and of the isolated photosystems (red trace) at 490 nm for carotenoids and 770 nm for bacteriochlorophyll, respectively. Identified peaks: 1 and 2, putative (2S,2'S)-oscillool 2,2'-di-( $\alpha$ -L-rhamnoside); 3–5, unknown carotenoids; 6–7, gemmatoxanthin; 8, BChl aGG; 9, BChl aP. B: Absorption spectrum of purified carotenoid corresponding to peak 6 in methanol (red) and hexane (blue). The insert depicts the absorption spectrum after NaBH<sub>4</sub> reduction in methanol (dashed).



**Figure 2**

NMR assignment of gemmatoxanthin structure. (A) spin systems based on <sup>1</sup>H-<sup>1</sup>H COSY spectra. (B) Crucial <sup>1</sup>H-<sup>13</sup>C HMBC and (C) crucial ROESY contacts.



**Figure 3**

The most prominent MS/MS product ions provided by gemmatoxanthin in the positive (blue) and negative ionization mode (red). For detailed information, refer to the Supplementary Tables S1 and S2.

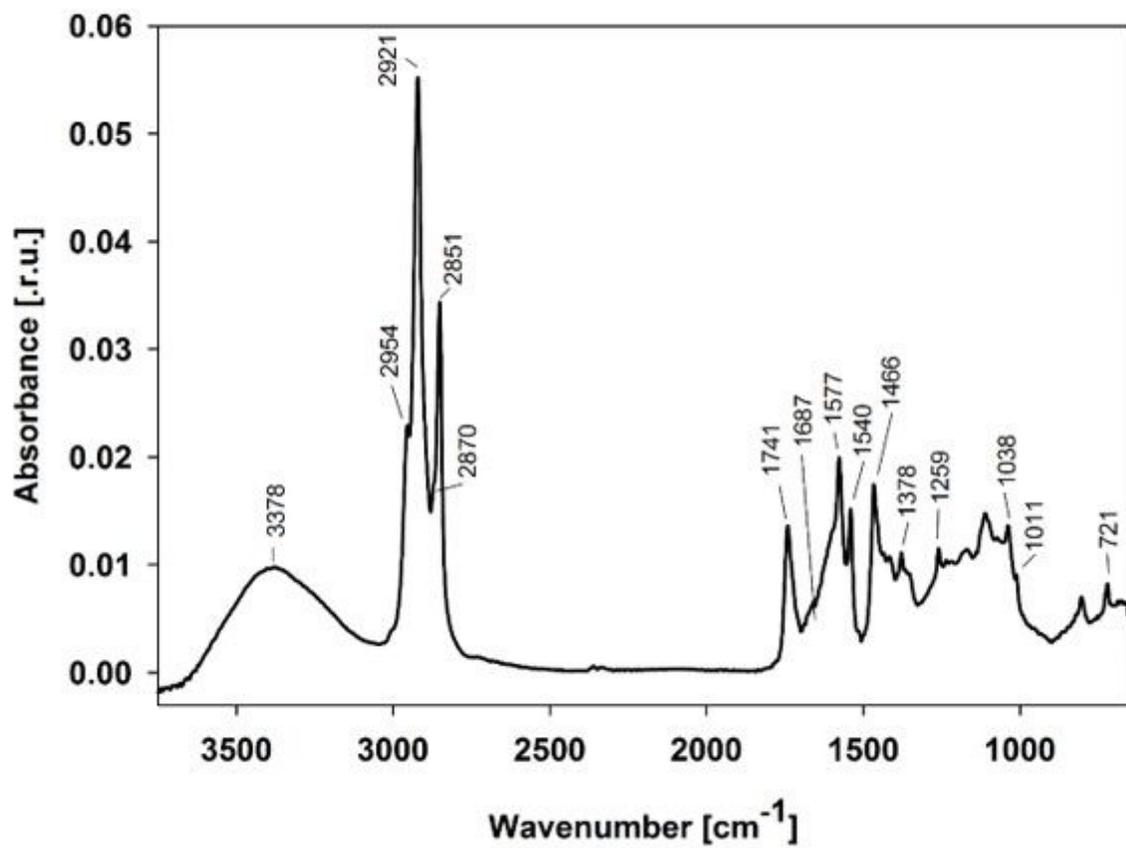
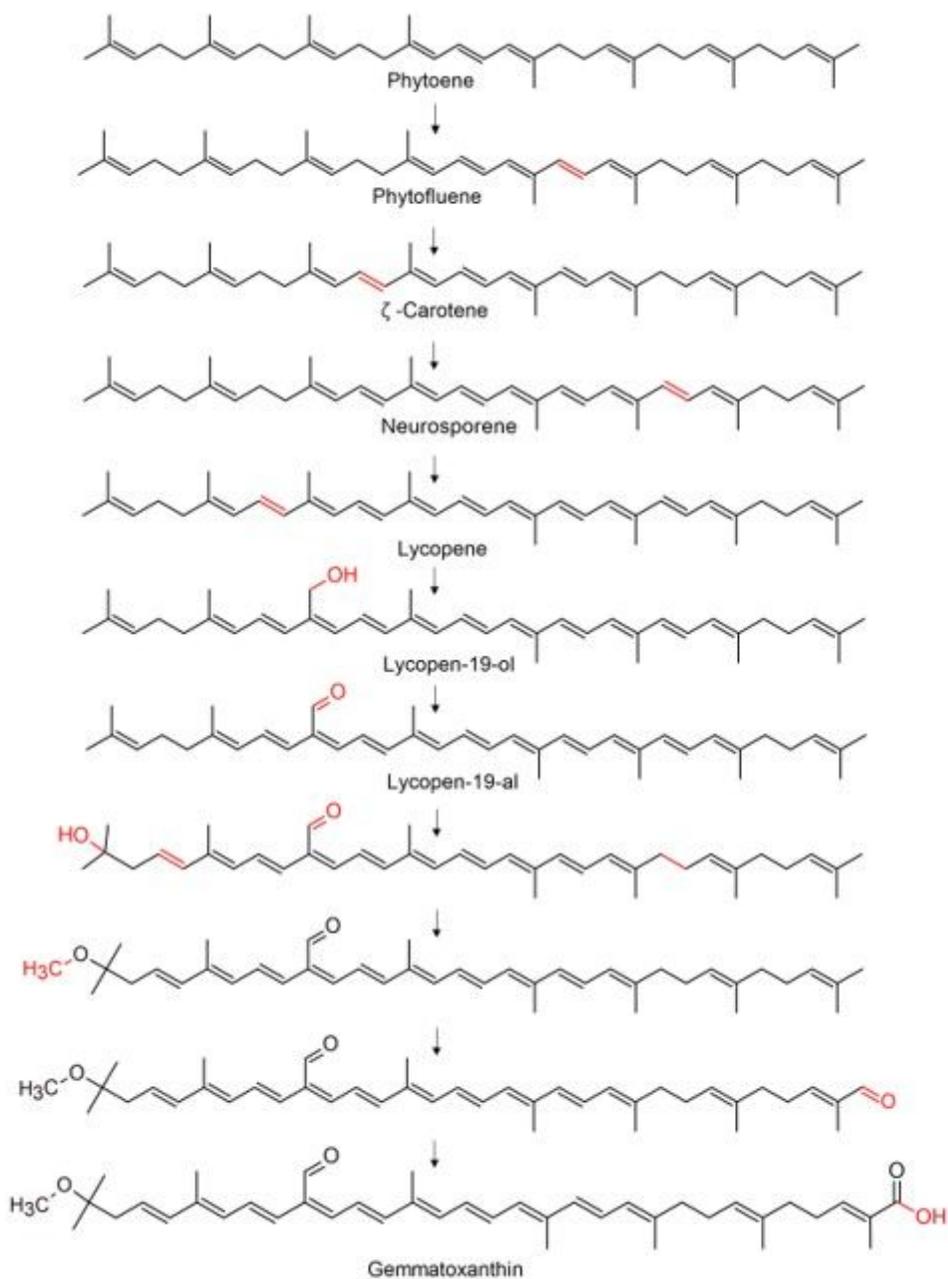


Figure 4

ATR-FTIR spectrum in the (4000-500 cm<sup>-1</sup>, mid-infrared region) of gemmatoxanthin



**Figure 5**

Tentative biosynthetic pathway of gemmatoxanthin. The biosynthesis of the gemmatoxanthin pathway begins from phytoene and lycopene<sup>9</sup>. Lycopene undergoes modification by the hydroxylase/oxidase activity of enzymes (genes are not yet known) to form lycopene-19-al which was tentatively predicted in *Lamprocystis roseopersicina* belonging to the Chromatiaceae family<sup>39,40</sup>. The *cruF* gene and *crtF* gene are probably responsible for the addition of terminal methoxy group in the pathway<sup>8</sup>. We hypothesize that the 1' carbon of methyl group undergoes modification by the hydroxylase/oxidase enzyme encoded by *cruH* gene yielding carboxylic acid, which is already known in synechococcal pathway<sup>27,43</sup>. Two unidentified genes similar to *cruH* gene was found in the genome of the *G. phototrophica* AP64 using *tblastn* showing 36.607% similarity (GEMMAAP\_15370) and 23.034% identity (GEMMAAP\_15125).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [2021CarotenoidpaperSuppl26032021.pdf](#)