



OPEN

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

Nupur¹, Marek Kuzma², Jan Hájek^{1,3}, Pavel Hrouzek^{1,3}, Alastair T. Gardiner¹, Martin Lukeš¹, Martin Moos⁴, Petr Šimek⁴ & Michal Koblížek^{1✉}

Gemmatimonas phototrophica AP64 is the first phototrophic representative of the bacterial phylum *Gemmatimonadetes*. The cells contain photosynthetic complexes with bacteriochlorophyll *a* as the main light-harvesting pigment and an unknown carotenoid with a single broad absorption band at 490 nm in methanol. The carotenoid was extracted from isolated photosynthetic complexes, and purified by liquid chromatography. A combination of nuclear magnetic resonance (¹H NMR, COSY, ¹H-¹³C HSQC, ¹H-¹³C HMBC, *J*-resolved, and ROESY), high-resolution mass spectroscopy, Fourier-transformed infra-red, and Raman spectroscopy was used to determine its chemical structure. The novel linear carotenoid, that we have named gemmatoxanthin, contains 11 conjugated double bonds and is further substituted by methoxy, carboxyl and aldehyde groups. Its IUPAC-IUBMB semi-systematic name is 1'-Methoxy-19'-oxo-3',4'-didehydro-7,8,1',2'-tetrahydro-Ψ, Ψ 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.

Photosynthesis is an ancient process that probably evolved more than 3 billion years ago¹. It is believed that the earliest phototrophic organisms were anoxygenic (not producing oxygen) species², 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^{3,4}. To date, two phototrophic species have been described in the genus *Gemmatimonas*: *G. phototrophica* AP64, originating from a freshwater lake in the Gobi Desert⁵, and *G. groenlandica* TET16 isolated from a freshwater stream in Greenland⁶.

So far most of the research focused 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 efficiency⁷. 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 Proteobacteria⁸.

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, carotenoids⁸. Some polar carotenoids present in *G. phototrophica* have been previously identified in the heterotrophic species *G. aurantiaca* T-27. These carotenoids were identified as oscilloxanthin 2,2'-dirhamnosides^{8,9} and are

¹Centre Algatech, Institute of Microbiology of the Czech Academy of Sciences, 379 81 Třeboň, Czech Republic. ²Laboratory of Molecular Structure Characterization, Institute of Microbiology of the Czech Academy of Sciences, Vídeňská 1083 Prague, Czech Republic. ³Faculty of Science, University of South Bohemia, Branišovská 1760 České Budějovice, Czech Republic. ⁴Biology Centre of the Czech Academy of Sciences, Branišovská 1160/31, 370 05 České Budějovice, Czech Republic. ✉email: koblizek@alga.cz

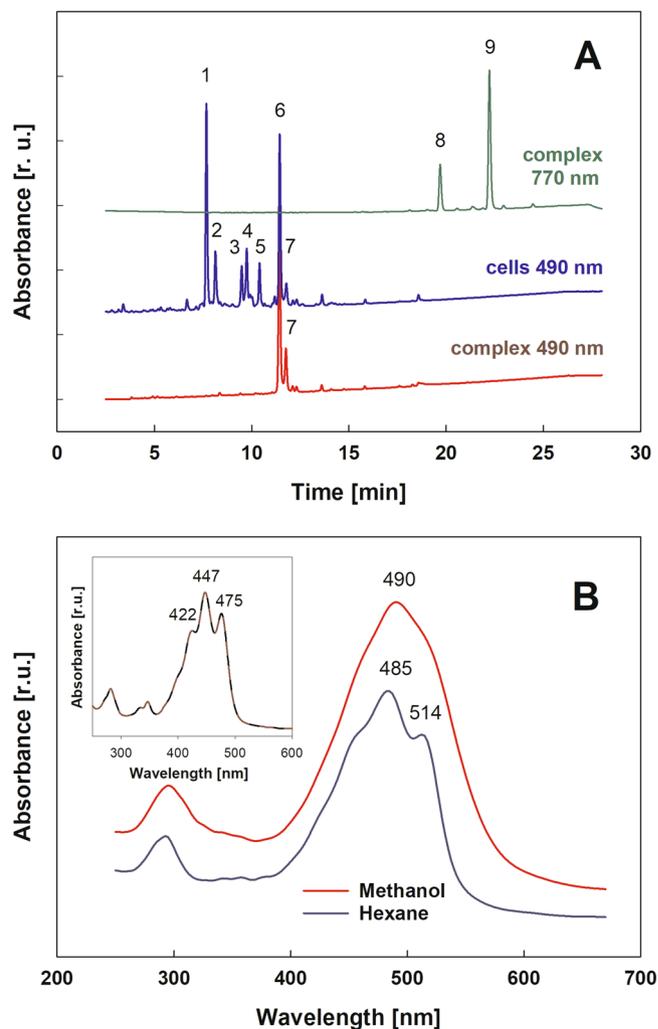


Figure 1. (A) HPLC elution profile of *Gemmatimonas phototrophica* AP64 whole-cell extract (green and blue trace) and of the isolated photosystems (red trace) at 490 nm for carotenoids and 770 nm for bacteriochlorophyll, respectively. Identified peaks: 1 and 2, putative (2*S*,2'*S*)-oscillol 2,2'-di-(α -L-rhamnoside); 3–5, unknown carotenoids; 6–7, gemmatoxanthin; 8, BChl *a*_{CG}; 9, BChl *a*_P. (B) Absorption spectrum of purified carotenoid corresponding to peak 6 in methanol (red) and hexane (blue). The insert depicts the absorption spectrum after NaBH₄ reduction in methanol (brown). r.u. – relative units.

unique to this bacterial genus. These carotenoids presumably have a photoprotective role to prevent the formation of reactive oxygen species (ROS)¹⁰ and may also stabilize bacterial membranes.

Recently, the photosynthetic complex of *G. phototrophica* AP64 has been purified and characterized¹¹. 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 RC¹¹. 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 (Fig. 1A). These polar carotenoids were identified earlier in *G. aurantiaca*⁹. 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. 1B).

This unknown carotenoid was isolated by preparative HPLC. The HRMS analysis of the purified carotenoid provided molecular ions at m/z 613.4268 [M+H]⁺ and 611.4108 [M-H]⁻ (calculated for C₄₁H₅₇O₄, –2.7 ppm and

$C_{41}H_{55}O_4^-$, 0.305 ppm) in positive and negative modes, respectively. This led to an unambiguous determination of the summary molecular neutral formula $C_{41}H_{56}O_4$.

The absorption spectrum of the unknown carotenoid in methanol has a single broad band with λ_{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 514 nm (Fig. 1B). Since such spectral behavior is typical for carotenoids containing a carbonyl group¹², we tested its presence through chemical reduction using $NaBH_4$. 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 peaks (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}C NMR spectrum directly. Therefore, *G. phototrophica* cells were grown on the ^{13}C -labeled substrate and the experiment was conducted with ^{13}C enriched gemmatoxanthin. The information about carbon chemical shifts had to be obtained indirectly by 1H - ^{13}C HSQC and 1H - ^{13}C HMBC correlation experiments (Table 1, Fig. S2-S10). The 1H NMR spectra are provided in the supplementary Fig. S11-S14.

The COSY spectrum (Fig. S15-S19) in combination with 1H - ^{13}C HSQC (Fig. S2-S5) allowed us to identify partial spin systems (Fig. 2A), namely one $-CH_2-CH=CH-$, three $=CH-CH=CH-$, one $=CH-CH=CH-CH=$, and two $-CH_2-CH_2-CH=$. Moreover, the 1H NMR also contained an aldehyde signal (δ_C 194.2, δ_H 9.539), signals of six methyl on a double bond, methoxyl (δ_C 48.6, δ_H 3.114) and one methyl of a double intensity (δ_C 24.7, δ_H 1.091). The 1H - ^{13}C HMBC (Fig. S10) further pointed to a carbon atom (C16) with a chemical shift 169.5 ppm indicating the presence of a carboxyl functional group, which was further confirmed by FTIR and MS measurements (see below). The identified structure fragments were partially interconnected by 1H - ^{13}C HMBC depicted in Fig. 2B. As the quaternary carbon C2 and carbon of methine C1 were not observable, we have performed an additional 1H - ^{13}C HMBC experiment using highly ^{13}C labeled sample obtained by cultivation in SILEX media. As a result, ^{13}C signals were obtained proving the connection of C1 and C2 at 130.9 and 137.0 ppm, respectively (Fig. S20). This connection was further confirmed by the MS/MS data presented below.

The ROESY experiment was used to establish stereochemistry on the double bonds (Fig. S21-S23). The crucial through space interactions detected in the ROESY spectrum 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' (product ion X_1^- , Fig. 3). It was followed by a consecutive CO_2 loss that formed a product ion at m/z 535.39. The decarboxylation was also observed directly from the molecular ion (m/z 567.4213—product ion X_2^- , Fig. 3) that established the presence of a carboxyl group on the NMR unassigned C16 carbon (^{13}C -shift 169.5 ppm). Further product ions X_3^- to X_7^- 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_1^+ , Fig. 3) and formyl groups forming a product ion at m/z 553.41 ($X_1^+ \rightarrow X_3^+$, Fig. 3) was observed in the positive ion MS spectrum. Both the gemmatoxanthin positive and negative ion spectra provided product ions 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 product ions is provided in the Supplementary information (Tables S1, S2 and Fig. S24).

Fourier-transform infrared spectroscopy and Raman spectroscopy. The FTIR spectrum of gemmatoxanthin in methanol shows several absorbance peaks ranging from 4000 to 500 cm^{-1} (Fig. 4). Peaks were identified using literature library data¹³⁻¹⁶. A broad absorbance band at a spectrum frequency of 3378 cm^{-1} indicates the presence of a hydroxyl group¹⁶. An absorbance band at 1741 cm^{-1} corresponds to a C=O stretch of carboxylic acid¹⁷. A band at 2954 cm^{-1} corresponds to an asymmetric vibration of methyl groups together with a rather weak symmetric vibration at 2870 cm^{-1} . Prominent bands at 2921 cm^{-1} and 2851 cm^{-1} correspond to asymmetric and symmetric vibrations of methylene groups^{13,16,18}. The vibration band of C-H from the methoxy group ($-O-CH_3$) is not visible as it could be overlapped by stronger signals originating from the aforementioned CH_3 and CH_2 vibrations^{14,16}. The vibration visible as a shoulder at 1687 cm^{-1} indicates the presence of conjugated aldehyde¹³. The symmetric 1577 cm^{-1} and asymmetric 1540 cm^{-1} stretchings of C=C confirms the conjugated system of double bonds in the carbon backbone of the carotenoid¹³. The absence of a vibration band at in region of 3015–3007 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 1466 cm^{-1} together with weak methyl band at 1378 cm^{-1} (umbrella deformation vibration of C-H of CH_3) and band at 721 cm^{-1} (methylene rocking vibration) is indicative for long-chain linear aliphatic structures¹⁸. A weak, but visible, band at 1259 cm^{-1} originates from C-H rocking. A weak band at 1038 cm^{-1} belongs to C-O bending¹⁹. A weak band at 1011 cm^{-1} is designated as out-of-plane bending and $Ca=Ca'$ torsion of the backbone chain²⁰.

The recorded Raman spectra²¹ contained three characteristic bands typical for carotenoids (Fig. S25). These three bands originate from (1) a rocking vibration of the molecule's methyl component (1015 and 1014 cm^{-1})²², (2) from C-C single-bond stretch vibrations (1158 cm^{-1})^{23,24} and (3) C=C stretching vibrations (1525 and

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

Table 1. The ¹H and ¹³C NMR data of gemmatoxanthin (**1**) (700.13 MHz for ¹H, 176.05 for ¹³C, DMSO-*d*₆, 303.1 K). s: singlet, d: doublet, t: triplet, q: quartet, m: multiplet.

1518 cm⁻¹) of the molecule's backbone^{22,24,25}. Other Raman bands elicited by the lasers were at 910 and 908 cm⁻¹ and considered as C—C symmetric stretch²⁶. A vibrational band appearing as a non-resolved band at 1176 and 1175 cm⁻¹ belonged to CH₃ rocking and a weak band at 1591 cm⁻¹ (532 nm) and 1585 cm⁻¹ (785 nm) was assigned to C=C stretching vibrations²².

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 Figs. 2 and 3, and having a IUPAC-IUBMB semi-systematic name 1'-Methoxy-19'-oxo-3',4'-didehydro-7,8,1',2'-tetrahydro- Ψ,Ψ -caroten-16-oic acid.

C40 carotenoids containing a carboxyl group or its variants are more common in the case of algae, fungi and yeast²⁷. Only a limited number of photosynthetic prokaryotes are known to have carboxylic carotenoids, for instance, thiothece-484 from *Thiocystis gelatinosa*²⁸ or synechoxanthin from cyanobacterium *Synechococcus*

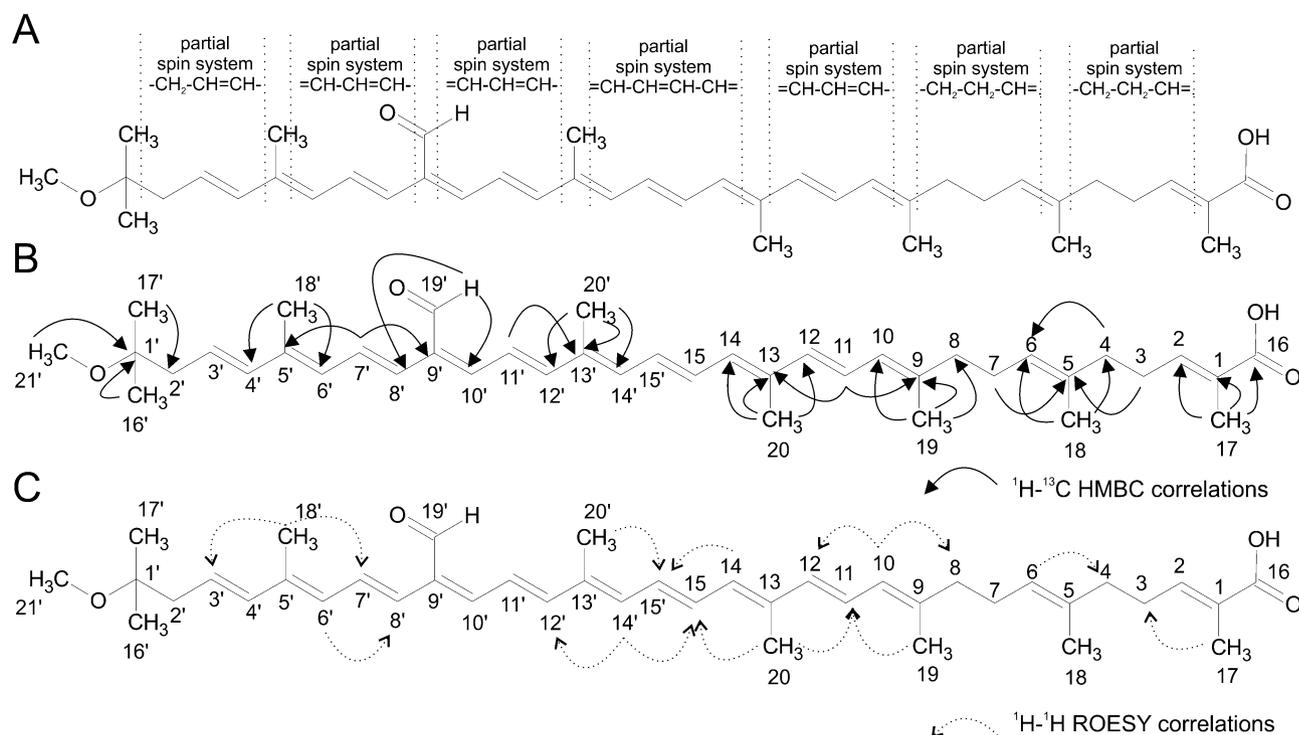


Figure 2. NMR assignment of gemmatoxanthin structure. (A) Spin systems based on ^1H - ^1H COSY spectra. (B) Crucial ^1H - ^{13}C HMBC and (C) crucial ROESY contacts.

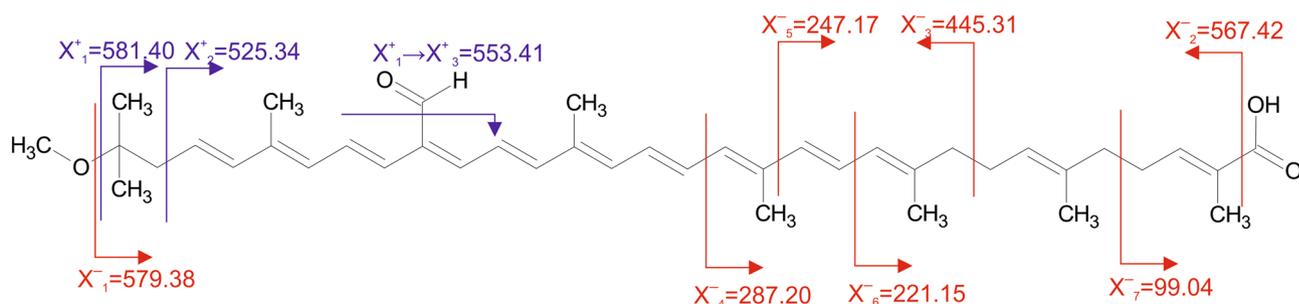


Figure 3. The most prominent MS/MS product ions provided by gemmatoxanthin molecular and demethoxylated ions in the positive (blue) and negative ionization mode (red). For detailed information, refer to the Supplementary Tables S1, S2 and Fig. S24.

sp. PCC 7002²⁹. 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^{27,30,31}.

Gemmatoxanthin's unique structure is corroborated by a specific MS/MS APCI product ion pattern. The majority of carotenoids provide neutral losses, 92 and 106 Da, corresponding to the electrocyclic elimination of toluene and xylene, respectively³². 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 chain³³. 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_2 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 ($\text{H}_2\text{O} + \text{CO}$) 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 $2860\text{--}2800\text{ cm}^{-1}$ region of the infrared spectrum, at a lower frequency than that of the main C-H absorption¹⁴. The strong absorption of CH_3 and CH_2 groups may overlap the absorption region for a

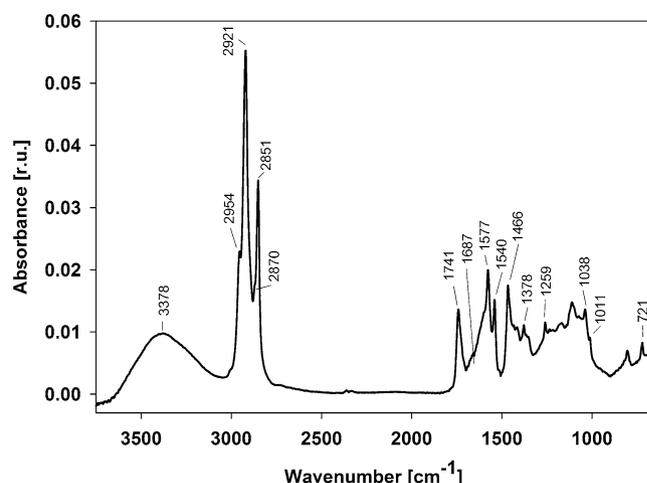


Figure 4. ATR-FTIR spectrum in the (4000–500 cm^{-1} , mid-infrared region) of gemmatoxanthin. r.u.- relative units.

methoxy group, which may hinder precise detection¹⁴. The key functional group vibration at 3378 cm^{-1} corresponds to the hydroxyl group and a band at 1741 cm^{-1} to the C=O stretch. Hence, the carboxyl group characteristic vibration has been proved by FTIR¹⁶. The absence of the *cis* C=C absorbance band in the region $>3000 \text{ cm}^{-1}$ which usually absorbs at $\sim 3006 \text{ cm}^{-1}$ ¹³⁴ 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_{40} keto carotenoid containing formyl and a carboxylic group.

Resonance Raman spectroscopy showed typical simple spectra that were exclusively dominated by bands at 1525 cm^{-1} assigned to $\nu(\text{C}=\text{C})$ in phase stretching, 1158 cm^{-1} assigned to $\nu(\text{C}-\text{C})$ stretching, and 1015 cm^{-1} assigned to a combination of $\delta(\text{C}=\text{CH})$ methyl in-plane rocking and $\delta(\text{C}-\text{H})$ out-of-plane bending modes when excited with 785 nm laser and have shown the similar bands at 1518 cm^{-1} , 1158 cm^{-1} , and 1014 cm^{-1} when excited by 532 nm laser. The wavenumber position of $\nu(\text{C}=\text{C})$ stretch is usually influenced by the length of the conjugated carbon chain together with changes in the substitution pattern of the chain and by interactions with other constituents in the matrix^{35–37}. The linear relationship between the conjugation length N and the $\nu(\text{C}=\text{C})$ Raman band is not always readily followed by β -rings, ketones, conjugated end-cycles containing carotenoids^{21,38}. For instance, β -carotene^{39,40} and gemmatoxanthin both possess 11 conjugated double bonds exhibiting the different $\nu(\text{C}=\text{C})$ Raman band at wavenumber 1517 cm^{-1} and 1525 cm^{-1} respectively. The reason being, β -carotene is a cyclic carotenoid where two conjugated double bonds are in cyclohexane rings resulting lower wavenumber 1517 cm^{-1} while gemmatoxanthin is a linear molecule with a central formyl group C=O connected to the system of conjugated bonds resulting in a shift of higher wavenumber 1525 cm^{-1} . No strong bands appearing in the region of 1290–1200 cm^{-1} are indicative for all-*trans* configuration⁴¹. The $\nu(\text{C}-\text{C})$ stretching wavenumber position is similar to carotenoids such as lutein, lycopene, and β -carotene and their mixtures⁴². The $\delta(\text{C}=\text{CH})$ methyl in-plane rocking and $\delta(\text{C}-\text{H})$ out-of-plane bending is slightly shifted by $\sim 10 \text{ cm}^{-1}$ to higher wavenumber $\sim 1015 \text{ cm}^{-1}$ in gemmatoxanthin than in other known carotenoids with exception of crocetin and *cis*-Bixin³⁵, together with appearance of a band at 1175–76 cm^{-1} may be indicative for contamination with some other carotenoid.

Carotenoids, owing to their electron-rich polyene chain, are known to photodegrade or isomerize upon exposure to light, heat, oxygen, acids or alkaline base^{43,44}. Therefore, to avoid the photooxidation, carotenoids were measured in dark, are kept in low temperature and in the inert environment (under nitrogen). However, during the NMR measurement, the sample was measured in $\text{DMSO}-d_6$ at 30 $^{\circ}\text{C}$ without excess of light. Under these conditions the sample was stable for three days and between NMR measurements, the sample was stored at $-80 \text{ }^{\circ}\text{C}$. We observed substantial degradation of the sample after the third round of freezing/melting.

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 bacteria⁴⁵. It was speculated that shift from spheroidene to spheroidenone, which occurs in *Rhodospira sphaeroides* under aerobic conditions, may help to quench potentially dangerous singlet oxygen⁴⁶. Similar characteristics may have also predicted in the carotenoid bacteriorubixanthin present in the photosynthetic complexes of *Erythrobacter* species⁴⁷. Numerous studies have been published on the energy transfer efficiency of carbonyl carotenoids to the BChl *a* in photosynthetic complexes^{47,48}. The energy transfer efficiency of rhodopinal to BChl *a* found to be $\sim 100\%$ ⁴⁹ and similarly, for okenone the energy transfer efficiency is $\sim 95\%$ ⁵⁰.

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-(α -L-rhamnoside) with lycopene as its intermediate⁹. It is, therefore, plausible to assume that gemmatoxanthin synthesis begins from phytoene and lycopene (Fig. 5). Phytoene, the first C_{40} -carotene, is converted to lycopene by incorporating double bonds.

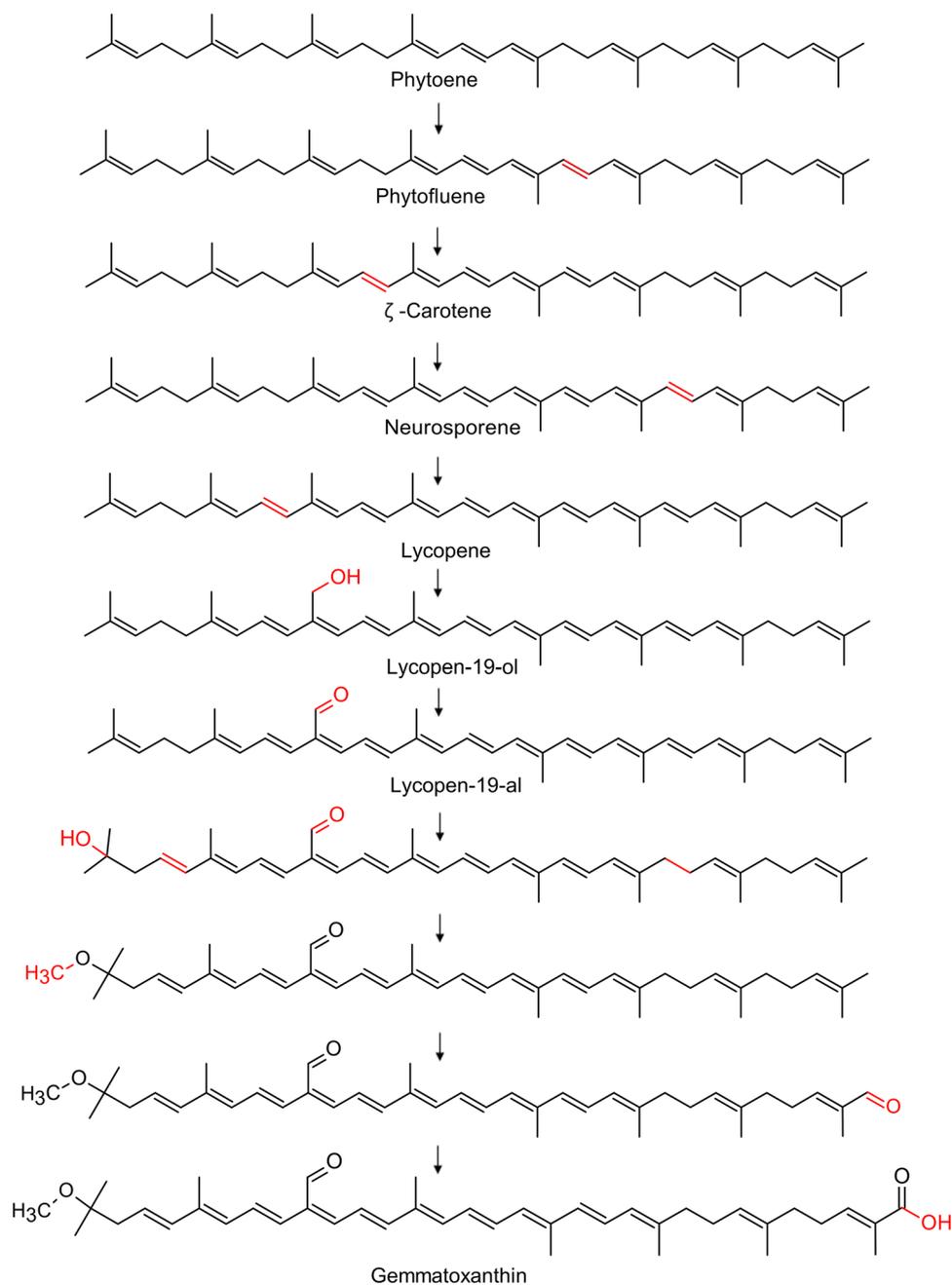


Figure 5. Tentative biosynthetic pathway of gemmatoxanthin. The biosynthesis of the gemmatoxanthin pathway begins from phytoene and lycopene⁹. Lycopene undergoes modification by the hydroxylase/oxidase activity of enzymes (genes are not yet known) to form lycopene-19-al which was tentatively predicted in *Lamprocystis roseopersicina* belonging to the *Chromatiaceae* family^{50,51}. The *cruF* gene and *crtF* gene are probably responsible for the addition of terminal methoxy group in the pathway⁸. 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^{29,54}. 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).

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^{50,51}. 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 position⁵². 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 PGC⁸. We can hypothesize that the methyl group of 1' carbon undergoes a hydroxylation and oxidation reaction yielding carboxylic acid, which is already known for synechoxanthin²⁹ and torulahadrin carotenoids⁵³ isolated from the unicellular Cyanobacterium *Synechococcus* sp. strain PCC 7002 and from the yeast *Rhodotorula*, respectively. Synechoxanthin is an aromatic carotenoid, the pathway of which is derived from β -carotene. One of the intermediates of the synechoxanthin pathway consists of β,χ -carotene ring at one end and renierapurpurin (χ,χ -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^{29,54}. 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^{53,55}.

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* AP64⁵ 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 K₂HPO₄, pH 7.7, 15 g of Bacto™ Agar, and 1 ml of modified SL8 trace metal solution (190 μ g CoCl₂·6H₂O, 5.2 mg Na₂-EDTA, 24 μ g NiCl₂·6H₂O, 17 μ g CuCl₂·2H₂O, 70 μ g ZnCl₂, 1.8 mg SrCl₂·6H₂O, 20.3 mg MgCl₂, 62 μ g H₃BO₃ per ml) and 1 ml vitamins (200 μ g B1, 20 μ g B3, 10 μ g B7, 10 μ g B12 per ml). Cells were streaked on agar media plates and incubated for approx. 2 weeks in microaerophilic conditions (10% O₂ + 90% N₂) 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 MgCl₂. 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-centrifuge Beckman L8-55 M equipped with the fixed angle rotor Ti-55.2 (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⁻¹ 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⁻¹ 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 to 1.3.

Gemmatoxanthin purification and HPLC–PDA analysis. Pigments were extracted twice in 100% methanol (HPLC grade, VWR Czech Republic) from wet 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 μ l of HPLC grade methanol mixed with 5 μ l of 25% of ammonium acetate buffer and 20 μ l 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 μ m, 100 Å, 150×4.6 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 NaBH₄ in purified carotenoid dissolved in 1 ml of methanol⁵⁶.

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, extracted and purified from 1 gm of wet cells, 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, 150×4.6 mm, $2.6 \mu\text{m}$, Phenomenex, Torrance, CA, USA) using H_2O (A)/methanol (B) both containing 0.1% HCOOH as a mobile phase with the flow rate of 0.6 ml 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. To acquire the fragmentation spectra in positive mode the HPLC was connected to an Impact HD high-resolution mass spectrometer (Bruker, Billerica, Massachusetts, USA) equipped with atmospheric pressure chemical ionization (APCI) ionization source. The following APCI probe settings were used: drying temperature, $250 \text{ }^\circ\text{C}$; drying gas flow, 12 l 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. The MS/MS experiments were performed using molecular ion at m/z 613.4106 Da $[\text{M} + \text{H}]^+$ and demethoxylated ion at m/z 581.3844 Da as precursor ions with collision energies 35 and 28 eV, respectively. The calibration of the instrument was performed using sodium formate clusters at the beginning of each analysis. The molecular formulas of obtained molecular peaks and fragments were calculated using Smart Formula function 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, $250 \text{ }^\circ\text{C}$, sheath gas at 30 au, aux gas at 10 au, spare gas at 0 au; probe temperature, $400 \text{ }^\circ\text{C}$ 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, $\pm 3 \text{ Hz}$; automatic gain control (AGC) target, 3×10^6 and maximum ion injection time (IT), 100 ms. The negative ion MS/MS experiments using molecular ion at m/z 611.4106 Da $[\text{M} - \text{H}]^-$ and demethoxylated ion at m/z 579.3844 Da as precursor ions were conducted at R = 17,500 (m/z 200), automatic gain control target, 2×10^5 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 triazine, $25 \mu\text{mol L}^{-1}$, $2 \mu\text{ l min}^{-1}$) 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. $10 \mu\text{g}$ of pure 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 to 4000 cm^{-1} , at a spectral resolution of 4 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. $10 \mu\text{g}$ of pure carotenoid 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 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}C labelled glucose, 0.5 g yeast extract, 1 g K_2HPO_4 , 15 g of Bacto™ Agar, and 1 ml of modified SL8 trace element solution as mentioned previously. The incorporation of ^{13}C glucose displayed low labelling in the gemmatoxanthin. Therefore, in order to obtain highly ^{13}C labeled substituted gemmatoxanthin, the second batch of *G. phototrophica* AP64 cells was grown with ^{13}C labelled SILEX *E. coli* media (0.5 g L^{-1}) (Silantes GmbH, Munich, Germany) instead of ^{13}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 $40 \text{ }^\circ\text{C}$ 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 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 $5 \mu\text{m}$, Watrex) eluted with HPLC grade water (A) and 100% methanol (B) at

a flow rate of 2.0 ml min⁻¹ 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 ¹H, 176.05 MHz for ¹³C, Bruker Biospin GmbH, Rheinstetten, Germany) was used to acquire NMR spectra in DMSO-*d*₆ at 303.1 K. The solvent signals were used as an internal standard (δ_{H} 2.499 ppm and δ_{C} 39.5 ppm). The ¹H NMR, COSY, ¹H-¹³C HSQC, ¹H-¹³C 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*₆ and water signal in ¹H NMR, COSY, and ¹H-¹³C HSQC due to a low concentration of the sample. The ¹H NMR spectrum was zero-filled to fourfold 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 scale (ppm) and coupling constants are given in Hz. The digital resolution (0.3 Hz for ¹H) 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 ¹H-¹³C HSQC and ¹H-¹³C HMBC and are reported to one decimal place.

Received: 30 March 2021; Accepted: 20 July 2021

Published online: 05 August 2021

References

- Blankenship, R. E. Early evolution of photosynthesis. *Plant Physiol.* **154**, 434–438. <https://doi.org/10.1104/pp.110.161687> (2010).
- Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of photosynthesis. *Annu. Rev. Plant Biol.* **62**, 515–548. <https://doi.org/10.1146/annurev-arplant-042110-103811> (2011).
- Zeng, Y. *et al.* Metagenomic evidence for the presence of phototrophic Gemmatimonadetes bacteria in diverse environments. *Environ. Microbiol. Rep.* **8**, 139–149. <https://doi.org/10.1111/1758-2229.12363> (2016).
- Mujakić, I. *et al.* Common presence of phototrophic Gemmatimonadota in temperate freshwater lakes. *mSystems* **6**, e01241–e11220. <https://doi.org/10.1128/mSystems.01241-20> (2021).
- 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. <https://doi.org/10.1099/ijs.0.000272> (2015).
- Zeng, Y. *et al.* Gemmatimonas groenlandica sp. nov. Is an aerobic anoxygenic phototroph in the phylum Gemmatimonadetes. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2020.606612> (2021).
- Koblížek, M. *et al.* Utilization of light energy in phototrophic Gemmatimonadetes. *J. Photochem. Photobiol. B* **213**, 112085. <https://doi.org/10.1016/j.jphotobiol.2020.112085> (2020).
- Zeng, Y., Feng, F. Y., Medová, H., Dean, J. & Koblížek, M. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. *Proc. Natl. Acad. Sci. USA* **111**, 7795–7800 (2014).
- Takaichi, S., Maoka, T., Takasaki, K. & Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyosicollol 2-rhamnoside, and proposed biosynthetic pathway of osicollol 2,2'-dirhamnoside. *Microbiology* **156**, 757–763. <https://doi.org/10.1099/mic.0.034249-0> (2010).
- Yurkov, V. & Csotonyi, J. T. in *The Purple Phototrophic Bacteria* (eds C. Neil Hunter, Fevzi Daldal, Marion C. Thurnauer, & J. Thomas Beatty) pp 31–55 (Springer, Berlin, 2009).
- Dachev, M. *et al.* Unique double concentric ring organization of light harvesting complexes in *Gemmatimonas phototrophica*. *PLoS Biol.* **15**, e2003943. <https://doi.org/10.1371/journal.pbio.2003943> (2017).
- Zigmantas, D. *et al.* Effect of a conjugated carbonyl group on the photophysical properties of carotenoids. *Phys. Chem. Chem. Phys.* **6**, 3009–3016. <https://doi.org/10.1039/B315786E> (2004).
- Lóránd, T., Deli, J., Molnár, P. & Tóth, G. FT-IR study of some carotenoids. *Helv. Chim. Acta* **85**, 1691–1697. [https://doi.org/10.1002/1522-2675\(200206\)85:6%3c1691::aid-hlca1691%3e3.0.co;2-g](https://doi.org/10.1002/1522-2675(200206)85:6%3c1691::aid-hlca1691%3e3.0.co;2-g) (2002).
- Degen, I. A. Detection of methoxyl group by infrared spectroscopy. *Appl. Spectrosc.* **22**, 164–166. <https://doi.org/10.1366/000370268774383444> (1968).
- Colthup, N. B., Daly, L. H. & Wiberley, S. E. in *Introduction to Infrared and Raman Spectroscopy*. 3rd edn (eds N. B. Colthup, L. H. Daly, & S. E. Wiberley) pp. 327–337 (Academic Press, Cambridge, 1990).
- Coates, J. in *Encyclopedia of Analytical Chemistry* (eds Meyers, R. A. & McKelvey, M. L.). <https://doi.org/10.1002/9780470027318.a5606> (2006).
- 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. <https://doi.org/10.1128/jb.173.24.7911-7917.1991> (1991).
- Colthup, N. B., Daly, L. H. & Wiberley, S. E. *Introduction to Infrared and Raman Spectroscopy* 327–337 (Academic Press, Cambridge, 1990).
- 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. <https://doi.org/10.1016/j.cdp.2003.11.004> (2004).
- Schulz, H. & Baranska, M. Identification and quantification of valuable plant substances by IR and Raman spectroscopy. *Vib. Spectrosc.* **43**, 13–25. <https://doi.org/10.1016/j.vibspec.2006.06.001> (2007).
- Mendes-Pinto, M. M. *et al.* Electronic absorption and ground state structure of carotenoid molecules. *J. Phys. Chem. B* **117**, 11015–11021. <https://doi.org/10.1021/jp309908r> (2013).
- 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. <https://doi.org/10.1039/b917341b> (2009).
- 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. <https://doi.org/10.1016/j.icarus.2004.07.029> (2005).
- Jehlička, J. *et al.* Potential and limits of Raman spectroscopy for carotenoid detection in microorganisms: implications for astrobology. *Philos. Trans. R. Soc. A.* <https://doi.org/10.1098/rsta.2014.0199> (2014).
- 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* <https://doi.org/10.1038/s41598-019-47208-2> (2019).

26. 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. Braz. Chem. Soc.* **15**, 951–963. <https://doi.org/10.1590/S0103-50532004000600025> (2004).
27. Yabuzaki, J. Carotenoids database: structures, chemical fingerprints and distribution among organisms. *Database* <https://doi.org/10.1093/database/bax004> (2017).
28. 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. <https://doi.org/10.3891/acta.chem.scand.26-2194> (1972).
29. Graham, J. E., Lecomte, J. T. J. & Bryant, D. A. Synechocanthin, an Aromatic C40 Xanthophyll that is a major carotenoid in the Cyanobacterium *Synechococcus* sp. PCC 7002. *J. Nat. Prod.* **71**, 1647–1650. <https://doi.org/10.1021/np800310b> (2008).
30. Nupur, *et al.* ProCarDB: a database of bacterial carotenoids. *BMC Microbiol.* **16**, 96. <https://doi.org/10.1186/s12866-016-0715-6> (2016).
31. Britton, G., Liaaen-Jensen, S. & Pfander, H. *Carotenoids* 271–280.272 (Birkhäuser, Basel, 2004).
32. 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. <https://doi.org/10.1002/rcm.7589> (2016).
33. Kane, M. A. & Napoli, J. L. Quantification of endogenous retinoids. *Methods Mol. Biol.* **652**, 1–54. https://doi.org/10.1007/978-1-60327-325-1_1 (2010).
34. Vlachos, N. *et al.* Applications of Fourier transform-infrared spectroscopy to edible oils. *Anal. Chim. Acta* **573–574**, 459–465. <https://doi.org/10.1016/j.aca.2006.05.034> (2006).
35. Schulz, H., Baranska, M. & Baranski, R. Potential of NIR-FT-Raman spectroscopy in natural carotenoid analysis. *Biopolymers* **77**, 212–221. <https://doi.org/10.1002/bip.20215> (2005).
36. de Oliveira, V. E., Castro, H. V., Edwards, H. G. M. & de Oliveira, L. F. C. Carotenes and carotenoids in natural biological samples: a Raman spectroscopic analysis. *J. Raman Spectrosc.* **41**, 642–650. <https://doi.org/10.1002/jrs.2493> (2010).
37. de Oliveira, V. E. *et al.* Carotenoids and β -cyclodextrin inclusion complexes: Raman spectroscopy and theoretical investigation. *J. Phys. Chem. A* **115**, 8511–8519. <https://doi.org/10.1021/jp2028142> (2011).
38. Fuciman, M., Keşan, G., LaFountain, A. M., Frank, H. A. & Polivka, T. Tuning the spectroscopic properties of aryl carotenoids by slight changes in structure. *J. Phys. Chem. B* **119**, 1457–1467. <https://doi.org/10.1021/jp512354r> (2015).
39. Parker, S. F. *et al.* Raman spectrum of beta-carotene using laser lines from green (514.5 nm) to near-infrared (1064 nm): implications for the characterization of conjugated polyenes. *Appl. Spectrosc.* **53**, 86–91 (1999).
40. Leopold, L. F. *et al.* Cellular internalization of beta-carotene loaded polyelectrolyte multilayer capsules by Raman mapping. *Molecules* **25**, 1477 (2020).
41. Tarantilis, P. A., Beljebbar, A., Manfait, M. & Polissiou, M. FT-IR, FT-Raman spectroscopic study of carotenoids from saffron (*Crocus sativus* L.) and some derivatives. *Spectrochimica Acta Part A Mol. Biomol. Spectrosc.* **54**, 651–657. [https://doi.org/10.1016/S1386-1425\(98\)00024-9](https://doi.org/10.1016/S1386-1425(98)00024-9) (1998).
42. Amaya, D. B. *A Guide to Carotenoid Analysis in Food* (OMNI Research. ILSI Press, 2001).
43. Xiao, Y.-D. *et al.* Thermal degradation kinetics of all-trans and cis-carotenoids in a light-induced model system. *Food Chem.* **239**, 360–368. <https://doi.org/10.1016/j.foodchem.2017.06.107> (2018).
44. Borsarelli, C. & Mercadante, A. Z. in *Carotenoids: Physical, Chemical, and Biological Functions and Properties* (ed Landrum, J. T.) 229–254 (CRC Press, Boca Raton, 2009).
45. Koblížek, M., Zeng, Y., Horák, A. & Oborník, M. Regressive evolution of photosynthesis in the Roseobacter clade. *Adv. Bot. Res.* **66**, 385–405. <https://doi.org/10.1016/B978-0-12-397923-0.00013-8> (2013).
46. Šlouf, V. *et al.* Photoprotection in a purple phototrophic bacterium mediated by oxygen-dependent alteration of carotenoid excited-state properties. *Proc. Natl. Acad. Sci. USA* **109**, 8570–8575. <https://doi.org/10.1073/pnas.1201413109> (2012).
47. Šlouf, V. *et al.* Carotenoid charge transfer states and their role in energy transfer processes in LH1-RC complexes from aerobic anoxygenic phototrophs. *J. Phys. Chem. B* **117**, 10987–10999. <https://doi.org/10.1021/jp309278y> (2013).
48. Šlouf, V. *et al.* Carotenoid to bacteriochlorophyll energy transfer in the RC-LH1-PufX complex from *Rhodobacter sphaeroides* containing the extended conjugation keto-carotenoid diketopirilloxanthin. *Photosynthesis Res.* **135**, 33–43. <https://doi.org/10.1007/s11120-017-0397-4> (2018).
49. 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. <https://doi.org/10.1021/jp5070984> (2014).
50. Ola Andersson, P., Cogdell, R. J. & Gillbro, T. Femtosecond dynamics of carotenoid-to-bacteriochlorophyll energy transfer in the light-harvesting antenna complexes from the purple bacterium *Chromatium purpuratum*. *Chem. Phys.* **210**, 195–217. [https://doi.org/10.1016/0301-0104\(96\)00172-3](https://doi.org/10.1016/0301-0104(96)00172-3) (1996).
51. Pfennig, N., Markham, M. C. & Liaaen-Jensen, S. Carotenoids of thiorhodaceae. *Arch. Mikrobiol.* **62**, 178–191. <https://doi.org/10.1007/BF00410404> (1968).
52. 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).
53. Kot, A. M., Błażejczak, S., Kurcz, A., Gientka, I. & Kieliszek, M. *Rhodotorula glutinis*—potential source of lipids, carotenoids, and enzymes for use in industries. *Appl. Microbiol. Biotechnol.* **100**, 6103–6117. <https://doi.org/10.1007/s00253-016-7611-8> (2016).
54. Graham, J. E. & Bryant, D. A. The biosynthetic pathway for synechocanthin, an aromatic carotenoid synthesized by the euryhaline, unicellular *Cyanobacterium Synechococcus* sp. Strain PCC 7002. *J. Bacteriol.* **190**, 7966–7974. <https://doi.org/10.1128/jb.00985-08> (2008).
55. Moliné, M., Libkind, D. & van Broock, M. in *Microbial Carotenoids From Fungi: Methods and Protocols* (ed José-Luis Barredo) 275–283 (Humana Press, Totowa, 2012).
56. Takaichi, S. & Shimada, K. Characterization of carotenoids in photosynthetic bacteria. *Methods Enzymol.* **213**, 374–385 (1992).

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, J.H., P.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. All authors reviewed the MS.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-95254-6>.

Correspondence and requests for materials should be addressed to M.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021