Studies on a bacterial photosensor
Kort, R.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 2
The xanthopsin protein family: a new member in \textit{Rhodobacter sphaeroides}


The xanthopsins: a new family of eubacterial blue-light photoreceptors


Department of Microbiology, E.C.Slater Institute BioCentrum, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands and 1Department of Biochemistry, Physiology and Microbiology, Laboratory of Protein Structure and Function, State University of Ghent, Kedagneestraat 35, 9000 Ghent, Belgium

Corresponding author

Phototactic yellow protein (PYP) is a photoreceptor that has been isolated from three halophilic photosynthetic purple bacteria. The PYP from Ectothiorhodospira halophila BN9626 is the only member for which the sequence has been reported at the DNA level. Here we describe the cloning and sequencing of the genes encoding the PYPs from E.halophila SL-1 (type strain) and Rhodospirillum salxeigens. The latter protein contains, like the E.halophila PYP, the chromophore trans-p-coumaric acid, as we show here with high performance capillary zone electrophoresis. Additionally, we present evidence for the presence of a gene encoding a PYP homolog in Rhodobacter sphaeroides, the first genetically well-characterized bacterium in which this photoreceptor has been identified. An ORF downstream of the pyp gene from E.halophila encodes an enzyme, which is proposed to be involved in the biosynthesis of the chromophore of PYP. The pyp gene from E.halophila was used for heterologous overexpression in both Escherichia coli and R.sphaeroides, aimed at the development of a holoPYP overexpression system (an intact PYP, containing the p-coumaric acid chromophore and displaying the 446 nm absorbance band). In both organisms the protein could be detected immunologically, but its yellow color was not observed. Molecular genetic construction of a histidine-tagged version of PYP led to its 2500-fold overproduction in E.coli and simplified purification of the heterologously produced apoprotein. HoloPYP could be reconstituted by the addition of p-coumaric anhydride to the histidine-tagged apopPYP (PYP lacking its chromophore). We propose to call the family of phototactic yellow proteins the xanthopsins, in analogy with the rhodopsins.

Keywords: Ectothiorhodospira halophila/phototastic yellow protein/Rhodobacter sphaeroides/Rhodospirillum salxeigens/xanthopsins

Introduction

The phototactic yellow proteins (PYPs) constitute a new family of eubacterial photoreceptor proteins (Hoff et al., 1994b). Members have been isolated from the halophilic phototrophic purple eubacteria Ectothiorhodospira halophila (Meyer, 1985), Rhodospirillum salxeigens (Meyer et al., 1990) and Chromatium salxeigens (Koh et al., 1996). PYP is the first eubacterial photoreceptor to be characterized in detail and has recently been shown to contain a unique chromophoric group: thiol ester linked p-coumaric acid (Baca et al., 1994; Hoff et al., 1994a). This is the first demonstration of a co-factor role for p-coumaric acid in eubacteria, previously only known from higher plants (Goodwin and Mercer, 1983). The pathway of biosynthesis of p-coumaric acid has been extensively studied in higher plants (Hahlbrock and Scheel, 1989), but no information is available on the conservation of this pathway in E.halophila or other eubacteria. In higher plants, the two enzymes of central importance in the metabolic conversions relevant for p-coumaric acid are: phenylalanine ammonia lyase (PAL), which catalyses the reaction from either phenylalanine or tyrosine to p-coumaric acid, and p-coumaryl-CoA ligase (PCL), which activates p-coumaric acid through a covalent coupling to CoA, via a thiol ester bond (Hahlbrock and Scheel, 1989).

The PYP from E.halophila is by far the best-studied member of this photoreceptor family. Its crystal structure has recently been redetermined at 1.4 A resolution and shows that the protein has an e/ß fold, resembling (eukaryotic) proteins involved in signal transduction (Borgstahl et al., 1995). Evidence has been obtained indicating that PYP functions as the photoreceptor for a new type of negative phototaxis response (Sprenger et al., 1993). Absorption of a blue photon (λmax = 446 nm) induces PYP to enter a cyclic chain of reactions (Meyer et al., 1987). This photocycle involves two intermediates and strongly resembles the photochemistry of the archaeabacterial sensory rhodopsins (Meyer et al., 1987; Hoff et al., 1994c).

Recently, the ORF encoding PYP from E.halophila BN9626 was cloned and sequenced (Baca et al., 1995). Here we report the cloning and the complete sequence of the pyp genes from E.halophila SL-1 (the type strain) and R.salxeigens, which is the first gene cloned from this organism, through reverse genetics. Directly downstream of the pyp gene in E.halophila we located a gene encoding a CoA ligase homolog, suggesting a plant-like conversion of p-coumaric acid to its CoA derivative before linkage to PYP lacking its chromophore (apoPYP).

Previously, we have reported the presence of a single cross-reacting protein in a large number of eubacteria, with a highly specific polyclonal antibody against PYP (Hoff et al., 1994b). Here we report, using heterologous PCR techniques, the identification of a new PYP homolog in the genetically well-characterized Rhodobacter sphaeroides. This finding opens the way to molecular genetic
**Fig. 1.** The *pyp* gene from *Ehhalophila* SL-1 with flanking regions. (A) Physical map of the chromosomal region containing the *pyp* gene. The cloned 2.4 kb *PstI* fragment, which is located on the 5.2 kb *EcoR1*-Spel fragment, is shown in detail, indicating the position of the *dada*, *pyp* and *pel* genes. The open arrow indicates the direction of the genes. (B) DNA sequence of the 1.8 kb *PvuI*-PstI fragment containing a partial ORF1, the *Ehhalophila pyp* gene and a partial ORF3. The derived amino acid sequences are given at the first position of each codon by the one letter code. The stop codon is indicated by an asterisk. The putative AT-rich promoter region (41 mol% GC) is underlined. Putative ribosome binding sites are doubly underlined and an inverted repeat is overlined. Underlined amino acids are part of a highly conserved motif in AMP-binding proteins (Fulda et al. 1994). The bases indicated by a vertical arrow differ from the formerly published *Ehhalophila* BN9626 sequence (Baca et al., 1994).
The xanthopsin protein family

Results

The pyp genes from E. halophila and R. salexigens

The DNA sequence of a 1.8 kb PvuII-PstI fragment was determined (Figure 1A) and is shown in Figure 1B. The amino acid sequence of E. halophila PYP predicted on the basis of this sequence information is identical to the one determined by amino acid sequencing (Van Beeumen et al., 1993), except for position 56 which is a Gin instead of a Glu, as also observed in the DNA sequence of the pyp gene from E. halophila BN9626 (Baca et al., 1995). A potential AT-rich (41 mol% GC) promoter region can be identified upstream of the ORF encoding PYP (positions 60–103, Figure 1B), which may be essential for the formation of an open complex for initiation of transcription. Also, a potential ribosome binding site (RBS) is located directly upstream of the PYP ORF. Directly downstream of the PYP ORF an inverted repeat is located (positions 557–587, Figure 1B).

The pyp gene from R. salexigens is the first gene cloned from this bacterium. It was localized on a 1.4 kb PvuI-SaiI chromosomal fragment. Sequence analysis of this fragment (Figure 2B) showed that it contains the entire ORF encoding PYP; the predicted amino acid sequence contains 125 amino acids and completely matches the amino acid sequence of this protein (Koh et al., 1996). Upstream of the ORF, a potential AT-rich (35 mol% GC) promoter region (positions 638–680, Figure 2B) and a potential ribosome binding site can be recognized, while directly downstream of the ORF an inverted repeat is present (positions 1134–1164, Figure 2B).

Identification of a PYP homolog in R. sphaeroides

Chromosomal DNA from R. sphaeroides 2.4.1. was used as template in a PCR with two primers homologous to conserved pyp sequences to yield a 0.3 kb product. The validity of the PCR product was confirmed by Southern hybridization experiments with R. sphaeroides chromosomal DNA under stringent conditions, using the PCR fragment as a probe. This revealed strong and specific hybridization signals (data not shown). The DNA sequence of the product showed that the encoding protein sequence was homologous to PYP from E. halophila, R. salexigens and C. salexigens (Figure 3).

Comparison of PYP sequences

The complete amino acid sequences of the PYPs from E. halophila, R. salexigens and C. salexigens (Koh et al.,
Fig. 3. Sequence conservation in the family of photoactive yellow proteins: the xanthopsins. Model for the p-coumaric acid binding pocket based on crystallographic data (Borgstahl et al., 1995) and sequence conservation of the residues forming this pocket in the PYP sequences from E.halophila, Rs.sakxigens, Csalexigens, and Rs.sphaewides. Sequence conservation is indicated in gray, with the more and less essential residues for p-coumaric acid binding indicated in blue (asterisks) and orange respectively. The unique Cys69, which binds the chromophore, is indicated in green, the chromophore trans p-coumaric acid and the thiol ester linkage in yellow.

1996) are homologous, with 66% of the amino acids identical in all three sequences. This result enabled us to obtain the partial sequence of a PYP homolog from R.sphaewides (see above). A partial alignment of these four sequences is shown in Figure 3. All proteins contain the Cys residue that in the E.halophila protein has been shown to bind covalently to the chromophore (Van Beeumen et al., 1993). From the 1.4 Å crystal structure of PYP it can be concluded that Tyr42, Glu46, Arg52 and to a lesser degree Thr50 and Tyr98, in the E.halophila PYP, are important for the protein-chromophore interactions that lead to the deprotonation of the p-coumaric acid molecule and result in the tuning of the absorbance of this cofactor to 446 nm (Baca et al., 1995; Borgstahl et al., 1995; Kim et al., 1995). These residues are all conserved in the PYPs from E.halophila, R.sakxigens and C.salexigens (Figure 3), in line with the similarities between these proteins with respect to their absorbance spectrum and photochemical properties (Meyer, 1985; Meyer et al., 1990). In the sequence of the R.sphaewides PYP homolog these six residues, of central importance for the binding of the chromophore, are also conserved, with the exception of Thr 50 (Figure 3). Furthermore, a strong conservation is observed in the sequence VIGKNFF, which forms a type II tight turn between the a4-helix and the b3-strand of PYP (Borgstahl et al., 1995).

Analysis of pyp flanking regions
The 1.8 and 1.4 kb chromosomal fragments from E.halophila and R.salexigens respectively, were examined for the presence of ORFs. In addition to the PYP ORFs presented above, this analysis indicates the presence of a large partial ORF (391 residues) downstream of the pyp gene from E.halophila (Figure 1B). This ORF was not found in the chromosomal fragment from R.salexigens. In line with this, comparison of the 1.8 and 1.4 kb chromosomal fragments from E.halophila and R.salexigens showed that the sequence similarity in these fragments is confined to the ORFs encoding PYP.

Upstream of the pyp gene from E.halophila SL-1 an ORF is located that shows significant homology to the E.coli dada gene, encoding the small subunit of the membrane bound iron-sulfur flavoenzyme D-amino acid dehydrogenase (Olsiewski et al., 1980), as was found in E.halophila BN926 (Baca et al., 1994). The partial ORF downstream of the pyp gene from E.halophila was further analyzed by searching for sequence similarities with proteins in the SwissProt database. The most similar proteins were found to be a number of CoA ligases from various organisms with -24% sequence identity and 48% similarity over a stretch of 400 amino acids (Table II). Furthermore, this putative pcl gene (see Figure 1A) shows, like the pyp gene, a high GC-bias in the wobble position of its codons, which is indicative of its functionality. In R.salexigens the ORF encoding a CoA ligase homolog has not been found downstream from the pyp gene. This may suggest a larger intergenic region between pyp and the putative pcl in this latter organism. This is supported by a Southern blot, showing hybridization of R.salexigens chromosomal digests with the putative E.halophila pcl (M.K.Philips-Jones, unpublished observations).

Identification of the chromophore of R.salexigens PYP
The chromophore of R.salexigens PYP was identified as p-coumaric acid in the purified protein with high-
The xanthopsin protein family

To overexpress PYP from *E. halophila*, a 0.45 kb *AvaII* fragment from pYAMA958 containing the *pyp* ORF, was inserted into the overexpression plasmid pT713 (Studier et al., 1990) to yield pTY13. After transformation of pTY13 to *E. coli* BL21, 50- to 100-fold overproduction of PYP was observed using Western blots and rocket immunoelectrophoresis (RIEP). However, absorbance spectra of the cytoplasmic fraction of these cells do not show an absorbance band at 446 nm, while this band was expected to be clearly visible on the basis of the concentration of PYP determined by RIEP (data not shown). This indicates that *E. coli* BL21/pTY13 mainly produces apoPYP, i.e. PYP without the chromophore.

In an attempt to obtain an overexpression system for holoPYP, the plasmid pART3 (see Table 1), containing the same 0.45 kb insert with the *pyp* gene from *E. halophila*, was conjugated to *R. sphaeroides* DD13. Since this organism is phototrophic, like *E. halophila*, and therefore produces a large array of pigments, it may also synthesize p-coumaric acid. The DD13 strain is mutared with respect to synthesis of the photosynthetic apoproteins (Jones et al., 1992), reducing the absorbance of the associating pigments, thereby facilitating the observation of the expected absorbance band at 446 nm, caused by holoPYP. RIEP experiments showed that the transconjugant *R. sphaeroides* DD13/pART3 also produces PYP at levels 100-fold higher than *E. halophila* (data not shown). Approximately 50% of the PYP produced was associated with the membrane fraction from these cells. However, also in this case the expected absorbance band at 446 nm for holoPYP was lacking (data not shown).

A chimeric version of the *pyp* gene from *E. halophila* was cloned in *E. coli*, which allows one to isolate PYP by the presence of a histidine affinity tag in the gene product and to confirm the lack of the chromophore in PYP produced in *E. coli*. Surprisingly, *E. coli* M15/pHisP (see Table 1) overproduces PYP at levels of 50 mg/l culture per OD660 unit, as determined by RIEP (Figure 5A), which is ~2500-fold higher than *E. halophila* and ~50-fold higher than in the case of the two overexpression systems described above. Cell-free extracts from *E. coli* M15/pHisP were used in Ni-affinity chromatography. This method yielded ~75% pure protein in a single step (Figure 5B). Incubation of the isolated histidine-tagged PYP with enterokinase yielded a product with a molecular weight visible region of the spectrum is completely lacking (Figure 5C). This indicates that the protein produced in this *E. coli* strain is histidine-tagged apoPYP (HAP).

To demonstrate the usefulness of HAP for further biophysical studies on PYP, we reconstituted HAP with p-coumaric anhydride into holoPYP. The following observations showed that reconstitution of holoprotein was achieved: (i) spectral analysis showed an absorption band at 446 nm, which increased (to saturation) with a stepwise addition of the p-coumaric anhydride; (ii) analysis of absorbance spectra in time showed an increase at 446 nm and a decrease at 350 nm, in line with an increase of holoPYP concentration and a decrease of the anhydride concentration; (iii) purified reconstituted holoPYP showed an absorbance spectrum like that of purified native PYP (Figure 5C); (iv) reconstituted holoPYP can be reversibly bleached after absorption of light (data not shown).
Table I. Strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Strains and plasmids</th>
<th>Description</th>
<th>Source or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.coli BL21</td>
<td>hsdS, gal, (ColE1 )5718d11, Sam7, rpsL5, lac UV5-T7 (gen I)</td>
<td>Studier and Moffat (1986)</td>
</tr>
<tr>
<td>E.coli M15[pREP4]</td>
<td>expression host with repressor plasmid. Km&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Qiagen</td>
</tr>
<tr>
<td>E.coli TG1</td>
<td>MLLac promoter, F&lt;sup&gt;+&lt;/sup&gt;, Tet&lt;sup&gt;R&lt;/sup&gt;, Amp&lt;sup&gt;R&lt;/sup&gt;, Sm&lt;sup&gt;R&lt;/sup&gt;, lacZ&lt;sup&gt;AM15&lt;/sup&gt;, recA, Ty&lt;sup&gt;K&lt;/sup&gt;, Sin&lt;sup&gt;R&lt;/sup&gt;</td>
<td>Gibson (1984)</td>
</tr>
<tr>
<td>E.coli S7-1</td>
<td>type strain</td>
<td>Gibson et al. (1983)</td>
</tr>
<tr>
<td>E.halophila SL1</td>
<td>type strain</td>
<td>Raymond and Sistrom (1969)</td>
</tr>
<tr>
<td>R.sphaeroides 2.4.1</td>
<td>type strain</td>
<td>Von Niel (1944)</td>
</tr>
<tr>
<td>R.sphaeroides DD13</td>
<td>type strain</td>
<td>Jones et al. (1992)</td>
</tr>
<tr>
<td>Rs.salzleitens WS 68</td>
<td>type strain</td>
<td>Drews (1981)</td>
</tr>
<tr>
<td>Plasmid</td>
<td></td>
<td>Benning and Sommerville (1992)</td>
</tr>
<tr>
<td>pCHB900</td>
<td>pRRK415 and pSH3 derivative, Te&lt;sup&gt;R&lt;/sup&gt;</td>
<td>this study</td>
</tr>
<tr>
<td>pAKT3</td>
<td>0.45 kb <em>E</em>.halophila Avrll fragment cloned into pCHB900</td>
<td>Qiagen</td>
</tr>
<tr>
<td>pQE30</td>
<td>RBSH, 6×His tag, ColE1 ori, amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>this study</td>
</tr>
<tr>
<td>pHsp</td>
<td>0.42 kb <em>E</em>.halophila PCR product cloned into pQE30</td>
<td>Gibco-BRL</td>
</tr>
<tr>
<td>pT713</td>
<td>expression vector, T7 promoter, Amp&lt;sup&gt;R&lt;/sup&gt;</td>
<td>this study</td>
</tr>
<tr>
<td>pTV15</td>
<td>0.45 kb <em>E</em>.halophila Avrll fragment cloned into pT713</td>
<td>Messing and Vieira (1982)</td>
</tr>
<tr>
<td>pM13mp18/19</td>
<td>M13mpl derived phages, lacZ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>this study</td>
</tr>
<tr>
<td>pYAMA18</td>
<td>2.4 kb <em>E</em>.halophila Pstl fragment cloned into M13mpl18</td>
<td>this study</td>
</tr>
<tr>
<td>pYAMA958</td>
<td>1.8 kb <em>E</em>.halophila Pstl fragment cloned into M13mpl18</td>
<td>this study</td>
</tr>
<tr>
<td>pS16</td>
<td>1.4 kb Rs.salzleitens Pstl-SalI fragment cloned into M13mpl18</td>
<td>this study</td>
</tr>
</tbody>
</table>

Discussion

We report here the DNA sequence of two genes encoding proteins known to be yellow and photoactive. The sequence of *pyp* from *E*.halophila SL1 (type strain) is identical to the sequence reported for the *pyp* gene from *E*.halophila BN9626 (Bacal et al., 1994). In the flanking regions six differences between the two sequences were found, which in five cases did not lead to changes in amino acid residues (see Figure 1B); this indicates the close similarity but distinctness of these two strains. Interestingly, all silent mutations are from T in the *pyp* gene. This gene encodes a protein of 124 residues, which cross-reacts with a polyclonal antiserum raised against *E*.halophila PYP (data not shown). The amino acid sequence of the *R*.sphaeroides PYP homolog is ~46% identical to the sequence of the *pyp* genes from *E*.halophila, *R*.salexigens and *C*.salinum, indicating that this PYP belongs to a different sub-group of the yellow proteins (R.Kort and S.M.Hoffer, unpublished observations). Since *R*.sphaeroides is genetically accessible, this opens up possibilities for genetic studies concerning the function of PYP. The identification of this PYP homolog raises the question whether the *R*.sphaeroides protein also binds a p-coumaric acid chromophore. The conservation of Cys69, Tyr42, Glu46, Arg52 and Tyr98 in the *R*.sphaeroides sequence suggests that this may indeed be so. This leads to the prediction that *R*.sphaeroides, in addition to its well-studied positive phototactic and chemotactic responses (for a review see Armitage, 1992), displays additional phototactic and chemotactic responses (see Sprunger et al., 1993). This prediction is currently being tested.

Directly downstream of the *pyp* gene from *E*.halophila an ORF is located that shows the highest sequence similarity to a range of CoA ligases (Table II), including *p*-coumaryl-CoA ligases. The putative *E*.halophila CoA ligase contains the motif TSGSTGTP (Figure 1B), which is conserved in all members of the AMP-binding protein family, of which the coumaryl-CoA ligases form a distinct subfamily (Fulda et al., 1994). This motif resembles the known loop-forming adenosine-binding motif (Saraste et al., 1999). In plants, coumaryl-CoA ligase is of central importance in the metabolism of *p*-coumaric acid (Hahlbrock and Scheel, 1989). This suggests that in *E*.halophila, *p*-coumaric acid is likewise activated by the formation of a thiol ester bond with CoA.
The xanthopsin protein family

Esterification was demonstrated by the fact that in vitro reconstitution of holoPYP was observed with the thio-phenyl ester of p-coumaric acid and not with p-coumaric acid (Imamoto et al., 1995). A further indication for a functional coupling of the _pyp_ and _pel_ gene products is the presence of an inverted repeat between these two coding regions and the absence of a recognizable promoter sequence, directly upstream of the _pel_ gene (see Figure 1B). This indicates that transcription of the _pel_ gene occurs by readthrough of this inverted repeat from the promoter directly upstream of the _pyp_ gene.

The biosynthesis of p-coumaric acid, which in plants can be performed in one step by phenylalanine ammonia lyase (Hahlbrock and Scheel, 1989), may consist of three consecutive steps in prokaryotes (compare the amino acid fermentation scheme of the anaerobic bacterium _Clostridium sporogenes_; Bader et al., 1982). If so, an aromatic aminotransferase, a 2-keto-acid reductase and a dehydratase respectively, would be involved. In the first reaction, pyruvate may be the amino acceptor, as shown for many aminotransferases. The reformation of pyruvate would then be carried out by alanine dehydrogenase. Interestingly, the _dada_ gene upstream of the _pyp_ gene (Figure 1A), encodes an alanine dehydrogenase.

Based on the observations described above, one can conclude that the organization of the genes encoding the PYP sensory system is completely different from that of the only other well-studied class of bacterial photoreceptors: the archaeabacterial sensory rhodopsins. For sensory rhodopsin I (SR-I) it has recently been shown that transcription of the _sopi_ gene (encoding the SR-I apoprotein) is transcriptionally coupled to an ORF immediately upstream of the _sopi_ gene; this upstream ORF (the _htrl_ gene) encodes the signal transducer interacting with SR-I (Yao and Spudich, 1992; Ferrando-May et al., 1993; Spudich, 1994).

In the soluble protein fraction of _Rs.salexigens_ cells, we could detect the PYP chromophore p-coumaric acid (8 pmol/mg soluble protein). This finding made a protocol available for straightforward screening of intact cells for the presence of this chromophore. This may be of great importance, since the nature of the chromophore in receptors for a large number of blue-light responses, observed in microorganisms as well as in plants, has not yet been elucidated (Senger, 1987). The amount of chromophore identified in _Rs.salexigens_ is equivalent to 0.1 µg PYP per mg soluble cell protein, similar to the cellular content of PYP in _E.halophila_ (Meyer et al., 1985).

We propose to designate the family of PYPs 'xanthopsins', which is derived from the Greek words ξανθός - xanthos.
Materials and methods

Bacterial strains and plasmids

The strains and plasmids used in this study are listed in Table I. E.halophilia SL-1, the type strain, was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSM), Braunschweig, strain number 244.

Cell culturing

E.halophilia SL-1 (Raymond and Cistrom, 1969) and Rs.salexigens WS88 (Drews, 1981) were cultured photoautotrophically as described (Meyer, 1985 and 1990 respectively), unless specified otherwise. R.sphaeroides strain 2.4.1 (van Niel, 1914) was grown aerobically in a 500 ml screw-cap bottle under illumination with 60 W (yellow) and 0.01 μg (eyesight). The bacterial xanthopins resemble the archaeabacterial sensory rhodopsins at the level of photochemistry (Hoff et al., 1994c), as well as of function, which is proposed to be that of a photoreceptor in negative phototaxis (Sprenger et al., 1993). Further evidence for the xanthopins, as a cubatypical protein family, has been obtained by studies with a highly specific polyclonal antiserum against E.halophilia PYP, which showed the presence of a single, cross-reacting protein, with a size of ~15 kDa, in a large number of prokaryotic microorganisms (Hoff et al., 1994b).

The results reported here define the xanthopins as a protein family of photoreceptors with strong sequence conservation and a highly conserved chromophore binding site. In addition, we have identified a gene that most likely encodes an enzyme involved in p-coumaric acid activation and that therefore is essential for in vivo holopPYP synthesis. The heterologously produced apoPYP was used as substrate for in vitro holopPYP reconstitution, which is essential for further biophysical studies on intact and directionally mutagenized PYP and for hybrid forms of PYP, containing chromophore analogs (A.R.Kroon and H.P.M.Fennema, unpublished observations). In addition, the discovery of a PYP homolog in R.sphaeroides renders this new photoreceptor family genetically accessible.

Identification of the chromophore of Rs.salexigens PYP

A colorless Rs.salexigens culture, grown aerobically in the dark in Huter modified medium as described (Hoff et al., 1994b), was diluted twice in the same medium and incubated anaerobically at 42°C in a completely filled 500 ml screw-cap bottle under illumination with 60 W
tungsten light bulbs, yielding a red culture after 96 h. The soluble cell fraction of 300 ml of aerobically and anaerobically grown cultures was prepared as described elsewhere. Proteins were precipitated with 10% (v/v) trichloro-acetic acid and washed once with deionized water. Pellets were resuspended in 5 ml deionized water and incubated overnight at pH 12 (leading to a complete solubilization of the proteins) to hydrolyze thiol ester bonds, followed by acidification to pH 4 with hydrochloric acid and acetic acid to neutralize the chromophore for optimal extraction. Before extraction, protein concentrations were determined with the Bio-Rad protein assay, as described by the manufacturer. Chromophore extractions were performed with mixing thoroughly with 15 ml ethyl acetate, followed by 5 min of centrifugation at 120 g. The organic phase was washed twice with 5 ml deionized water and extracted by 5 min with 20 ml of 10% (v/v) trichloro-acetic acid. The supernatant was adjusted to pH 7 and subsequently grown in liquid medium under semi-anaerobic conditions.

Subsequent experiments were carried out using the purified Rs.salexigens PYP (Meyer et al., 1980). Air-dried samples were dissolved in distilled water and injected on a 50 µm fritted silica capillary TSP05375 (Composite Metal Services LTD) with an injection time of 0.2 min and injection pressure of 40 mbar. The sample was analyzed in 60 mM Tris, 50 mM sodium chloride, pH 8.2, through a capillary with an effective length of 55 cm, at 25 kV and -12 U.A. on-column detection was performed at 284 nm (determined as the wavelength at which trans-p-coumaric acid maximally absorbs in the Tris/sodium chloride buffer), with a UV/vis detector (Linear, Fremont). As a reference, 11.0 ml of trans-p-coumaric acid (Sigma, St Louis, MO) was used. To confirm this identification, the eluted protein was subjected to electrophoresis in 25 mM borax buffer, pH 9.0, at 25 kV and -35 U.A. The amount of the detected trans-p-coumaric acid was calculated from the peak area using the software Cesar for Windows (version 4.0.2, Perkin Elmer Technologies). As a reference, 11.0 ml of trans-p-coumaric acid (Sigma) was injected in the concentration range from 2.5 to 75 µM, showing a linear relation to the detected peak areas.

**Construction of overexpression plasmids and overproduction strains**

A 0.5 kb Avall fragment from pYAMA58, containing the pyp ORF from Ethalophilus, was ligated into the Smal-linearized overexpression plasmid pT713 (Studier et al., 1990) to yield pT713, which was transformed to E.coli BL21. Overexpression in pT713 is based on the strong viral T7 promoter 601. The gene coding for the viral RNA polymerase is located on the chromosome of E.coli BL21, downstream of an inducible tac promoter (Studier et al., 1990). A conjugative broad host range overexpression system was constructed by ligating the 0.45 kb Avall fragment, described above, into the PstI polylinker site of pCH500/pCH160, which is a broad host range vector, containing two promoters directly upstream of the polylinker site: the E.coli Pτ promoter and the Pλ promoter that supports anabolic expression of the yca gene from R.sphaeroides (Benning and Sommerville, 1992). The resulting plasmid pART3 was transformed into the conjugal strain E.coli S17 and then transferred to R.sphaeroides D1011 (Jones et al., 1992) by conjugation on LB agar plates for 4.5 h. Transconjugants were selected on minimal agar containing xylan (10 µg/ml), streptomycin (5 µg/ml) and kanamycin (20 µg/ml). The transconjugants were subsequently grown in liquid medium under semi-anaerobic conditions, allowing pigmentation synthesis.

A third overexpression system involved the heterologous overproduction of an affinity-tagged version of PYP from Ethalophilus in E.coli. The expression vector was constructed by directional insertion of a PCR product into the expression plasmid pQE30 (Qiagen, Hilden). The PCR product was obtained using pYAMA18 as template in a reaction with the primers GCCTTCGG and NTPYP2h containing the restriction sites of GCCTTCGG and NTPYP2h. The PCR product was ligated into the SmaI-linearized plasmid pQE30 (Qiagen) to yield pHl1. The PCR product was digested with BamHI and HindIII ligated into pHl1 (Qiagen) to yield pH2, and finally transformed to E.coli M15. The colonies, resistant against ampicillin (100 µg/ml) and kanamycin (25 µg/ml), were shown to contain the construct by colony PCR, using the two primers described above.

**SDS-PAGE, Western blotting and RIEP**

SDS-PAGE was performed in a Bio-Rad mini slab gel apparatus (Bio-Rad, Hercules, CA) according to Laemmli (1970) as modified by Schagger and Jagow (1987) for improvement of resolution in the 5-20 kDa range. Gels were stained with Coomassie brilliant blue G250. Western blotting and immunodetection were performed as described previously (Tosbir et al., 1979; Hoff et al., 1994b). RIEP was carried out as described (Hoff et al., 1994b).

**Heterologous expression of PYP**

E.coli BL21(pT713) and E.coli M15(pHl1) were induced to express the heterologous gene by the addition of 1 mM IPTG to well-aerated cultures of exponentially growing cells at an OD600 of 1. Cells were grown at 37°C in well-shaken Erlenmeyer, or in a well-aerated 10 l fermenter (New Brunswick Scientific, New Brunswick). Production of PYP in R.sphaeroides was induced by growing the organism semi-aerobically in two-fifths filled, slowly shaking Erlenmeyer, using Luria Bertani broth with appropriate antibiotics. The resulting E.coli and R.sphaeroides cultures were sorbed three times for 1 h while cooled on ice, and centrifuged at 200 000 g for 3 h at 4°C to obtain a clear supernatant containing the overexpressed product. Absorbance spectra of these fractions were measured with an Amicon DW2000 spectrophotometer (SEM Instruments). In addition, these fractions were used for SDS-PAGE, Western blotting and RIEP analysis, as described above.

**Isolation and cleavage by enterokinase of histidine-tagged PYP**

Ultracentrifugation supernatants from E.coli M15(pHl1), induced with IPTG, were incubated with Ni-NTA resin for 1 h at 4°C, as described by the manufacturer (Qiagen). The resin was packed in a column and related, either by adjusting the metal concentration or by a pH gradient, as described by the manufacturer. The protein elution pattern was analyzed by measuring the absorption of the eluting fractions at 280 nm. Cleavage of histidine-tagged apoPYP was performed at 37°C for 5-24 h using an enterokinase PYP ratio of 1:50 (w/w).

**Reconstitution of holoPYP**

Reconstitution of the heterologously produced apoPYP was achieved by addition of the p-coumaric anhydride, dissolved in dimethyl formamide (DMF), as described for the reconstitution of the apoPYP obtained from Ethalophilus (Inamoto et al., 1995). The p-coumaric anhydride was synthesized as described (Inamoto et al., 1995).

**Mass spectrometry**

The integrity of histidine-tagged apoPYP and reconstituted histidine-tagged holoPYP was verified by electrospray mass spectrometry (ESMS). Typically, 20 nmol of protein was dissolved in 10 ml CH3CN/water formic acid (1:9,0.1%; v/v) and injected into the electrospray source of a VG Bio-Q mass spectrometer (VG Organic, Altrenhim, UK) at a flow rate of 6 μl/min, delivered by a Harvard Syringe Pump 11 (Harvard, South Natick, MA). Nine-second scans, covering the 650-1550 amu range, were accumulated during 2.5 min. The spectra were collected and processed using the maslynx software provided with the instrument.

**Acknowledgements**

The authors are very grateful to R.Kok and Ivan Thr for their advice on the use of molecular genetic techniques and to Ivan DJik and W.Spikes for cloning the Rs.salexigens pyp gene. We thank K.Xu and H.Vonk for performing the capillary electrophoresis experiments and H.P.M.Fermana for the synthesis of the p-coumaric anhydride. We would like to thank B.Poelman for his help with the initial PCR experiments. We are very thankful to T.E.Meyer for supplying information concerning the RIEP protocol and to their advice for improving the resolution in the mass chromatograms. We thank J. Rauer for the generation of the expression vector pART3. The xanthopsin protein family has been reviewed by I. Hoff and we are very grateful to R. Dijkstra for the distribution of the information.

**References**


The xanthopsin protein family

Sequence, chromophore extraction and 3-D model of the photoactive yellow protein from *Rhodobacter sphaeroides* {1}

Remco Kort {a}, Mary K. Phillips-Jones {b}, Daan M.F. van Aalten {c}, Andrea Haker {a}, Sally M. Hoffer {a}, Klaas J. Hellingwerf {a}, Wim Crielaard {a, *}

{a} Laboratory for Microbiology, EC Slater Institute, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, Netherlands

{b} Department of Microbiology, University of Leeds, Leeds LS2 9JT, UK

{c} Keck Structural Biology, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA

Received 2 March 1998; accepted 19 March 1998

Abstract

The photoactive yellow protein (*pyp*) gene has been isolated from *Rhodobacter sphaeroides* by probing with a homologous PCR-product. A sequence analysis shows that this *pyp* gene encodes a 124 AA protein with 48% identity to the three known PYPs. Downstream from *pyp*, a number of adjacent open reading frames were identified, including a gene encoding a CoA-ligase homologue (*pCL*). This latter protein is proposed to be involved in PYP chromophore activation, required for attachment to the apoprotein. We have demonstrated the presence of the chromophoric group, previously identified in PYP from *Ectothiorhodospira halophila* as trans 4-hydroxy cinnamic acid, in phototrophically cultured *R. sphaeroides* cells by capillary zone electrophoresis. The basic structure of the chromophore binding pocket in PYP has been conserved, as shown by a 3D model of *R. sphaeroides* PYP, constructed by homology-based molecular modelling. In addition, this model shows that *R. sphaeroides* PYP contains a characteristic, positively charged patch. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Photoactive yellow protein; Chromophore extraction; (*Rhodobacter sphaeroides*)

---

Photoactive yellow protein is a small (125 amino acids), water-soluble protein found in the three halophilic purple bacteria *Ectothiorhodospira halophila*, *Rhodospirillum saleni*gens and *Chromatium saleni*gens [1–3]. The encoding gene has been cloned from two of these species [4,5]. The protein is proposed to play a role as a photoreceptor for negative phototaxis [6]. Upon blue light absorption, it enters a rhodopsin-like photocycle, starting with the fast formation of a red-shifted intermediate, followed by the formation of a blue-shifted intermediate and a relatively slow recovery of the ground state [7,8]. The crystal structure of PYP has been elucidated to 1.4 Å resolution [9]. Recently, structural information about the long-lived photocycle intermediate has also become available, showing conformational changes, including the ejection of the chromophore from the binding pocket [10]. In addition to these studies, it was shown that the chromophore of
PYP is trans 4-hydroxy-cinnamic acid [4,11], which is present as a deprotonated phenolate anion in the ground state [9,12]. This chromophore photo-isomerizes to the cis isomer after light absorption [13], in a two-bond isomerization reaction, from 7-trans 9-S-cis to 7 cis 9-S-trans [14], and becomes protonated in the long-lived photocycle intermediate [15].

Recently, we proposed the name Xanthopsins for the PYP protein family and reported the identification of a new PYP homologue in R. sphaeroides by its partial amino acid sequence obtained from the DNA sequence of a cloned PCR product [5]. We propose to designate this R. sphaeroides strain to RK1 (previously assigned to the type strain 2.4.1), since its genomic DNA shows an Asel digestion pattern that differs from the 2.4.1 strain, as obtained by transverse alternating field electrophoresis (data not shown). Here, we report the cloning and sequencing of the pyp genes from R. sphaeroides strains RK1 and NCIB8253.

The R. sphaeroides RK1 pyp gene was identified on 2.3-kb PstI fragment, which was cloned into the PstI-digested cloning vector pBS SK+ (Stratagene, La Jolla, CA), resulting in pATC3. To obtain single-stranded DNA for sequencing, a 0.5-kb PstI BamHI fragment, containing the entire pyp gene, was subcloned into the phages M13mp18/19. In addition, the pyp gene was cloned from R. sphaeroides strain NCIB8253 with the use of a pSUP202 plasmid library, constructed by Hunter and Coomber [16]. Part of this isolated plasmid (6.5 kb; see Fig. 1), which is designated to pSUP202.79, was sequenced according to described methods [17,18]. Processing of sequencing data was carried out with the program Sequencher 2.1 software (Gene Codes, Ann Arbor, MI). DNA and protein sequence analysis was performed with the Genetics Computer Group software package from the University of Wisconsin.

The 0.5-kb DNA sequence from the R. sphaeroides RK1 pyp gene and its flanking regions shows 99% identity to the DNA sequence of the same region from the NCIB8253 strain; the amino acid sequences of both PYPs are identical. Putative coding regions were identified using R. sphaeroides codon preference tables and the GC-bias at the third position of each codon. A total of 8 sequential open reading frames were identified, including 3 that show significant similarity to proteins in the SwissProt database: pyp, pel and orfF, encoding photoactive yellow protein, a CoA ligase homologue and a protein most homologous to sensory rhodopsin I, respectively (Fig. 1 and Table 1). The sequence alignment of the four known PYP sequences (Fig. 2) shows conserved amino acids, which play a crucial role in PYP function: Cys69, to which the chromophore is covalently linked by a thiol ester bond; Tyr42 and Glu46 (present in the protonated state), which hydrogen-bond to the phenolic oxygen of the chromophore; Arg52, which stabilizes the negative charge on the chromophore and Tyr98, which hydrogen-bonds to Arg52, keeping the chromophore shielded from the solvent [9]. In addition, the alignment shows conserved amino acids, which are part of a new structural motif: Asp34, Gly37, Asn43, Ala45 and Gly59, found in many proteins with a regulatory function [20]. This structural motif is known as the S1 box of the PAS domain [21].

Immunoscreening experiments have indicated that PYP-like proteins are widely distributed among bacteria, including R. sphaeroides [22]. A more recent report, using the same technique, claimed that PYP-like proteins are only present in three halophilic purple bacteria and suggested the involvement of an artefact due to incomplete purification of antiserum [23]. The current report shows that a PYP-like protein is also present in a non-halophilic species of the purple bacteria; whether the Xanthopsin protein family also extends beyond this group of anoxic photosynthetic proteobacteria, remains to be solved.

Interestingly, we identified about 1 kb downstream of pyp a gene encoding a p-coumaryl-CoA-ligase homologue (PCL), as indicated in Fig. 1. In a previ-
The xanthopsin protein family

ORF  AA  Position  re  FASTA  AL/gaps  sim/id  Ref.
orfA  185  717-1274  5  0  124  1299-1673  0  0  [5]
orfB  188  1688-2254  5  0  166  2251-2751  2  0  [5]
orfC  188  2251-2751  2  0  188  2763-3998  0  0  [6]
orfD  176  4031-4561  5  0  413  55/36  [5]
orfE  336  4561-5571  14  0  411  2763-3998  0  0  [5]

Table 1
Putative coding regions of the R. sphaeroides NCIB8253 6.5-kb DNA fragment

The columns indicate the number of encoded amino acids (AA), the position on the DNA fragment (position), the number of rare codons (re), using a threshold of 0.02, significant similarity to a protein found in the SwissProt protein database (FASTA), the length of the alignment to the subsequent proteins (AL), the number of gaps in the alignment (gaps), the percentage of similarity (sim) and identity (id) and the reference to the homologous protein (Ref.).

ous report, we showed that a pcl gene was also present in E. halophila, but directly downstream of the pyp gene [5]. There, we proposed that its gene product is involved in the conversion of the chromophore to its CoA derivative, before the latter is linked to apoPYP. Now, with the conservation of this downstream gene in R. sphaeroides, a functional involvement of its product with PYP appears even more likely. The R. sphaeroides PCL is most homologous to the E. halophila PCL (Table 1) and contains the highly conserved motif present among AMP-binding proteins. This stretch of amino acids is present in a large number of enzymes, forming an acyl-adenylate from a fatty acid and ATP, followed by the transfer of the acyl group to the sulfhydryl group of CoA and subsequent release of AMP [24]. This could very well be the mechanism of p-coumaric acid activation in R. sphaeroides. Furthermore, 4 kb downstream from pyp, a gene (orfF) was identified, encoding a product that shows a striking similarity to sensory rhodopsin I, a membrane spanning photoreceptor from the archaeon Halobacterium salinarum, mediating negative phototaxis [19]. The significance of this finding is not yet clear, since the bacterial rhodopsin signature and the retinal-binding site signature sequences are not conserved. However, the observed homology is very likely to indicate the presence of membrane spanning regions in Orff, as is also supported by its hydrophobicity plot (data not shown).

We identified the 4-hydroxy-cinnamic acid chromophore in phototrophically grown Rhodobacter cells by capillary zone electrophoresis (Fig. 3), performed according to methods described in [5]. Experiments aimed at the identification of the chromophore from R. sphaeroides cells grown semi-aerobically in the dark showed that this compound was not present in these cells (data not shown). This is compatible with the proposed photoreceptor role for PYP, needed under phototrophic conditions, where the protein mediates a response resulting in migration from too high (blue) light intensities [6]. These findings for R.
**Fig. 3.** Electropherograms of ethyl acetate extracts from *R. sphaeroides* RK1 cells. The eluate was analysed at 284 nm. Trace (A) shows 4-hydroxy cinnamic acid at 7.3 min and an unidentified compound at 8.3 min. Trace (B) shows the result of co-injection analysis of the extract with 4-hydroxy cinnamic acid (Sigma), showing enhancement of the peak at 7.3 min.

*sphaeroides* are strongly reminiscent of the information available for the purple bacterium *R. salsii-gens*, in which protein-attached chromophore, as well as PYP, could only be identified in cells grown anaerobically in the light, and not in aerobically grown cells in the dark [5,22].

A structural model for *R. sphaeroides* PYP, based upon the new sequence reported here, was constructed using the homology modelling procedure in the program WHATIF [25]. Rotamers of conserved residues were left unchanged, and all other residues were initially mutated to alanines. Rotamers were then modelled using the WHATIF backbone-dependent rotamer libraries. At each position, rotamer quality was checked by hydrogen bonding, van der Waals bumps and packing quality [26]. The resulting model was subjected to energy minimisations prior and subsequent to a 2-ps molecular dynamics run, using the GROMOS87 suite of programs [27]. Calculations were performed in vacuo with crystallographic waters, using the GROMOS reduced charges forcefield. The chromophore *p*-coumaric acid was included in the calculations, using a topology described elsewhere [28]. There are two clusters of mutations that are buried in the protein (positions 4, 11, 14 and 82, 83, 88, 118; Fig. 4). In both cases, cavities created by mutations to smaller residues are compensated by mutations to larger residues at complementary positions in the cluster. This mutational complementarity emphasises the quality of the model. Due to several mutations of (acidic) residues to neutral and basic residues, a positively charged patch has emerged in the region 71–81 (Fig. 4A). So far, this group of solvent-accessible positive amino acids has only been found in *R. sphaeroides* PYP, contributing to a striking shift upwards in the calculated iso-electric point in comparison to the other three known PYPs (10.10 vs. 5.00 ± 0.77). The basic structure of the chromophore pocket has been conserved (Fig. 4B). Two residues close to the chromophore, however, that have been changed compared to the *E. halopila* sequence, lead to small changes in the chromophore binding pocket. At position 50, there is an Ala in the *R. sphaeroides* sequence, while there is a Thr in the *E. halopila* PYP sequence, which hydrogen-bonds to Tyr98. By changing it to an Ala, there are two effects: (i) this hydrogen bond is lost, and Tyr98 may become more mobile and (ii) a small cavity next to the chromophore is created, giving it more flexibility. Similarly, there is a residue change Thr to Ala at position 70. A side chain-backbone hydrogen bond is thereby lost and a small cavity is created, possibly leading to similar effects.

We thank Prof. C.N. Hunter from the University of Sheffield, UK, for kindly providing the *R. sphaeroides* gene bank. We thank M. Gomelski, R. Ng and D. Needleman from the Department of Microbiology and Molecular Genetics, University of Texas Medical School Houston, USA for carrying out
Fig. 4. Spatial model for *R. sphaeroides* PYP. (A) Overall view of the model. The C-α trace is coloured grey, using a darker shade for the central β sheet. The four positively charged amino acids, which are part of the new patch around residue 75 (see text) are coloured red. Three prolines leading to a more rigid N-terminal domain are coloured green. The two clusters of buried mutations are coloured blue and the 4-hydroxy cinnamic acid chromophore is shown in yellow. (B) Chromophore binding pocket. The C-α trace is shown in dark grey and side chains in grey. Oxygen atoms are red, the sulphur atom is yellow and nitrogen atoms are coloured blue. Residues contacting the chromophore pocket are labelled. The 4-hydroxy cinnamic acid chromophore, covalently attached to Cys69, is shown in stick and ball representation.
Chapter 2

TAFE and part of the DNA sequencing. The latter was performed with a grant to Prof. S. Kaplan (GM 15590, The National Institute of Health). We thank W.N.M. Reijnders from the Free University of Amsterdam, The Netherlands, for expert assistance in DNA sequencing. We thank J. Hendriks from the University of Amsterdam, The Netherlands for work on the 3-D figures. M.P.-J., K.J.H. and W.C. thank the Royal Society, London and the British council-NWO for financial support. R.K. thanks NWO for the SIR-travel grant 14-1779 to Houston, TX, USA.

References