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

Kort, R.; Hoff, W.D.; van West, W.S.; Kroon, A.R.; Hoffer, S.M.; Vlieg, K.H.; Crielaard, W.; van Beeumen, J.J.; Hellingwerf, K.J.

Publication date
1996

Published in
EMBO Journal

The EMBO Journal vol.15 no.13 pp.3209–3218, 1996

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

R.Kort, W.D.Hoff[1], M.Van West, A.R.Kroon, S.M.Hoffer, K.H.Vlieg, W.Crielaard, J.J.Van Beeumen[2] and K.J.Hellingwerf[3]

Department of Microbiology, E.C.Slater Institute, BioCentrum, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands and [2]Department of Biochemistry, Physiology and Microbiology, Laboratory of Protein Structure and Function, State University of Ghent, Ledeganckstraat 35, 9000 Ghent, Belgium

[1]Present address: Department of Microbiology and Molecular Genetics, Health Science Center at Houston, University of Texas, 6431 Fannin, Houston, TX 77030, USA

[3]Corresponding author

Photoactive yellow protein (PYP) is a photoreceptor that has been isolated from three halophilic phototrophic 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 salexigens*. 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 apoPYP (PYP lacking its chromophore). We propose to call the family of photoactive yellow proteins the xanthopsins, in analogy with the rhodopsins.

Keywords: *Ectothiorhodospira halophila*/photoactive yellow protein/*Rhodobacter sphaeroides*/*Rhodospirillum salexigens*/xanthopsins

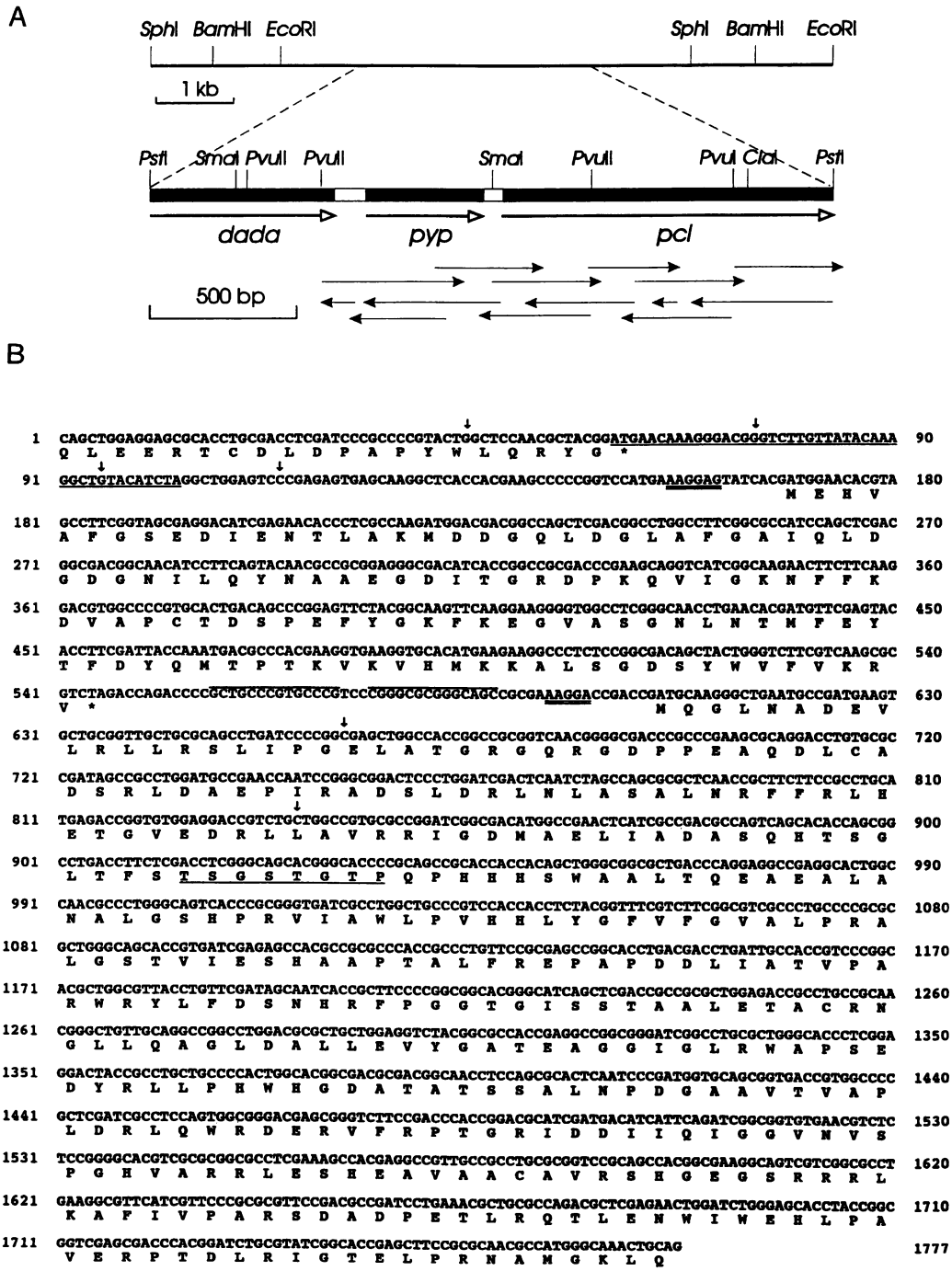
Introduction

The photoactive 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 salexigens* (Meyer *et al.*, 1990) and *Chromatium salexigens* (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 Å resolution and shows that the protein has an α/β 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 (λ_{max} = 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 archaebacterial 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 *Rs.salexigens*, 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

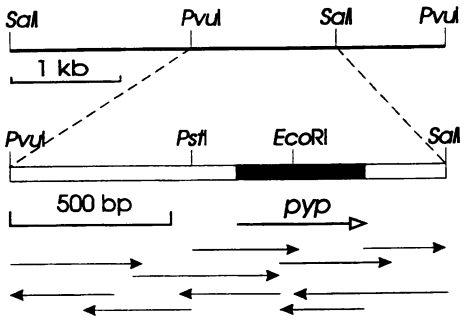



```
1    CAGCTGGAGGAGCGCACCTGCGACCTCGATCCCGCCCCGTACTGGCTCCAACGCTACGGATGAACAAAGGGACGGGTCTTGTTATACAAA   90
     Q  L  E  E  R  T  C  D  L  D  P  A  P  Y  W  L  Q  R  Y  G  *

91   GGCTGTACATCTAGGCTGGAGTCCCGAGAGTGAGCAAGGCTCACCACGAAGCCCCCGGTCCATGAAAGGAGTATCACGATGGAACACGTA   180
                                                                  M  E  H  V

181  GCCTTCGGTAGCGAGGACATCGAGAACACCCTCGCCAAGATGGACGACGGCCAGCTCGACGGCCTGGCCTTCGGCGCCATCCAGCTCGAC   270
     A  F  G  S  E  D  I  E  N  T  L  A  K  M  D  D  G  Q  L  D  G  L  A  F  G  A  I  Q  L  D

271  GGCGACGGCAACATCCTTCAGTACAACGCCGCGGAGGGCGACATCACCGGCCGCGACCCGAAGCAGGTCATCGGCAAGAACTTCTTCAAG   360
     G  D  G  N  I  L  Q  Y  N  A  A  E  G  D  I  T  G  R  D  P  K  Q  V  I  G  K  N  F  F  K

361  GACGTGGCCCCGTGCACTGACAGCCCGGAGTTCTACGGCAAGTTCAAGGAAGGGGTGGCCTCGGGCAACCTGAACACGATGTTCGAGTAC   450
     D  V  A  P  C  T  D  S  P  E  F  Y  G  K  F  K  E  G  V  A  S  G  N  L  N  T  M  F  E  Y

451  ACCTTCGATTACCAAATGACGCCCACGAAGGTGAAGGTGCACATGAAGAAGGCCCTCTCCGGCGACAGCTACTGGGTCTTCGTCAAGCGC   540
     T  F  D  Y  Q  M  T  P  T  K  V  K  V  H  M  K  K  A  L  S  G  D  S  Y  W  V  F  V  K  R

541  GTCTAGACCAGACCCCGCTGCCCGTGCCCGTCCCGGGCGCGGGCAGCCGCGAAAGGACCGACCGATGCAAGGGCTGAATGCCGATGAAGT   630
     V  *                                                                  M  Q  G  L  N  A  D  E  V

631  GCTGCGGTTGCTGCGCAGCCTGATCCCCGGCGAGCTGGCCACCGGCCGCGGTCAACGGGGCGACCCGCCCGAAGCGCAGGACCTGTGCGC   720
     L  R  L  L  R  S  L  I  P  G  E  L  A  T  G  R  G  Q  R  G  D  P  P  E  A  Q  D  L  C  A

721  CGATAGCCGCCTGGATGCCGAACCAATCCGGGCGGACTCCCTGGATCGACTCAATCTAGCCAGCGCGCTCAACCGCTTCTTCCGCCTGCA   810
     D  S  R  L  D  A  E  P  I  R  A  D  S  L  D  R  L  N  L  A  S  A  L  N  R  F  F  R  L  H

811  TGAGACCGGTGTGGAGGACCGTCTGCTGGCCGTGCGCCGGATCGGCGACATGGCCGAACTCATCGCCGACGCCAGTCAGCACACCAGCGG   900
     E  T  G  V  E  D  R  L  L  A  V  R  R  I  G  D  M  A  E  L  I  A  D  A  S  Q  H  T  S  G

901  CCTGACCTTCTCGACCTCGGGCAGCACGGGCACCCCGCAGCCGCACCACCACAGCTGGGCGGCGCTGACCCAGGAGGCCGAGGCACTGGC   990
     L  T  F  S  T  S  G  S  T  G  T  P  Q  P  H  H  H  S  W  A  A  L  T  Q  E  A  E  A  L  A

991  CAACGCCCTGGGCAGTCACCCGCGGGTGATCGCCTGGCTGCCCGTCCACCACCTCTACGGTTTCGTCTTCGGCGTCGCCCTGCCCCGCGC  1080
     N  A  L  G  S  H  P  R  V  I  A  W  L  P  V  H  H  L  Y  G  F  V  F  G  V  A  L  P  R  A

1081 GCTGGGCAGCACCGTGATCGAGAGCCACGCCGCGCCCACCGCCCTGTTCCGCGAGCCGGCACCTGACGACCTGATTGCCACCGTCCCGGC  1170
     L  G  S  T  V  I  E  S  H  A  A  P  T  A  L  F  R  E  P  A  P  D  D  L  I  A  T  V  P  A

1171 ACGCTGGCGTTACCTGTTCGATAGCAATCACCGCTTCCCCGGCGGCACGGGCATCAGCTCGACCGCCGCGCTGGAGACCGCCTGCCGCAA  1260
     R  W  R  Y  L  F  D  S  N  H  R  F  P  G  G  T  G  I  S  S  T  A  A  L  E  T  A  C  R  N

1261 CGGGCTGTTGCAGGCCGGCCTGGACGCGCTGCTGGAGGTCTACGGCGCCACCGAGGCCGGCGGGATCGGCCTGCGCTGGGCACCCTCGGA  1350
     G  L  L  Q  A  G  L  D  A  L  L  E  V  Y  G  A  T  E  A  G  G  I  G  L  R  W  A  P  S  E

1351 GGACTACCGCCTGCTGCCCCACTGGCACGGCGACGCGACGGCAACCTCCAGCGCACTCAATCCCGATGGTGCAGCGGTGACCGTGGCCCC  1440
     D  Y  R  L  L  P  H  W  H  G  D  A  T  A  T  S  S  A  L  N  P  D  G  A  A  V  T  V  A  P

1441 GCTCGATCGCCTCCAGTGGCGGGACGAGCGGGTCTTCCGACCCACCGGACGCATCGATGACATCATTCAGATCGGCGGTGTGAACGTCTC  1530
     L  D  R  L  Q  W  R  D  E  R  V  F  R  P  T  G  R  I  D  D  I  I  Q  I  G  G  V  N  V  S

1531 TCCGGGGCACGTCGCGCGGCGCCTCGAAAGCCACGAGGCCGTTGCCGCCTGCGCGGTCCGCAGCCACGGCGAAGGCAGTCGTCGGCGCCT  1620
     P  G  H  V  A  R  R  L  E  S  H  E  A  V  A  A  C  A  V  R  S  H  G  E  G  S  R  R  R  L

1621 GAAGGCGTTCATCGTTCCCGCGCGTTCCGACGCCGATCCTGAAACGCTGCGCCAGACGCTCGAGAACTGGATCTGGGAGCACCTACCGGC  1710
     K  A  F  I  V  P  A  R  S  D  A  D  P  E  T  L  R  Q  T  L  E  N  W  I  W  E  H  L  P  A

1711 GGTCGAGCGACCCACGGATCTGCGTATCGGCACCGAGCTTCCGCGCAACGCCATGGGCAAACTGCAG                        1777
     V  E  R  P  T  D  L  R  I  G  T  E  L  P  R  N  A  M  G  K  L  Q
```

Fig. 1. The *pyp* gene from *E.halophila* SL-1 with flanking regions. (**A**) Physical map of the chromosomal region containing the *pyp* gene. The cloned 2.4 kb *Pst*I fragment, which is located on the 5.2 kb *Eco*RI–*Sph*I fragment, is shown in detail, indicating the position of the *dada*, *pyp* and *pcl* genes. The open arrow indicates the direction of the genes. (**B**) DNA sequence of the 1.8 kb *Pvu*II–*Pst*I fragment containing a partial *ORF1*, the *E.halophila 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 *E.halophila* BN9626 sequence (Baca *et al.*, 1994).

studies of the function of PYP. The *E.halophila pyp* gene was heterologously overexpressed in *Escherichia coli* and *R.sphaeroides*, yielding (mainly) apoPYP. The purification of a histidine affinity-tagged derivative of PYP from *E.halophila*, overproduced in *E.coli*, yielded a 2500-fold overproduction of apoPYP. Intact PYP, containing the *p*-coumaric acid chromophore and displaying the 446 nm absorbance band (HoloPYP) could be reconstituted by the addition of *p*-coumaric anhydride to the recombinant apoPYP as described for apoPYP (Imamoto *et al.*, 1995). These results will facilitate detailed biophysical studies on a protein with a unique set of characteristics: it is water soluble, photoactive and its structure is known at 1.4 Å resolution.

A



B

```
   1  CGATCGCCTCCTCGGCTCCAGAACCAGGCGATGGCGCAGATAGGGGCTCAGGTGCGAGACGCTGTCGTCGGCATCGGGACCGTAATCGGT  90
  91  GTTGCGGTCGGCCGCATAGATCCGCCCGGCCCGGGGCGCGAAGGCGTCGAGCCGGGCCAGGCCAGCGGCGCGGGTCGTCCGCTGGCGTCT 180
 181  GGGGTGAATCGGCGTGCGCTGGCGCGACCGCTTCGACGCGGGAGGCTTCAATCATGATGGCGGCGGTTCCTTGGCGGCGGTGCGAGCCTG 270
 271  CCGGGCGCTGCGCCGGCAGCGTTCCGAGCAATAGCGCACGCGCTCCCAATCCCGGGCCCACTTCTTGCGCCAGGTGAAGGGGCGGTTGCA 360
 361  TTCGGCGCATTGTTTTTGGGGCAGGTCGGCCTTTTTGCGCATGACGGGGTTTCGATGTGACCACATGGTATCCGGTTCGTGCCAACCGGG 450
 451  GCCCCGGATCTGCCCCAGTTGCACTCGCTGGCGCGATGCGACAATGTGGCCGACCGGCAGGCGAGGGAGGCCGACAGCGCTGCAGCACAG 540
 541  GGCGGCGGCGGACGTTGACGGCCCACCGCAGCGACACGACGAACCGTTGACCTCTTGCGTCCGGTTCCCCCAAACGAGGCGCCGTGTGCG 630
 631  CACCGGGTTCATTTATCTGTATTGTATAAAGGGCGCCGATATCTATGTCTGGGCGTGGCGATGGAATCGTCCGCCTATCATAACCGTCAC 720
 721  GGGAGACTTGCTATGGAAATGATCAAATTCGGCCAGGACGACATCGAGAACGCCATGGCGGATATGGGCGACGCGCAGATCGACGACCTG 810
                M  E  M  I  K  F  G  Q  D  D  I  E  N  A  M  A  D  M  G  D  A  Q  I  D  D  L
 811  GCTTTTGGCGCCATTCAACTGGACGAGACCGGCACGATCCTGGCCTATAACGCGGCCGAGGGCGAACTAACCGGCCGCAGTCCCCAAGAC 900
       A  F  G  A  I  Q  L  D  E  T  G  T  I  L  A  Y  N  A  A  E  G  E  L  T  G  R  S  P  Q  D
 901  GTGATCGGCAAGAACTTCTTCAAGGACATAGCGCCGTGCACCGACACCGAGGAATTCGGCGGCCGGTTCCGCGAAGGGGTGGCCAATGGC 990
       V  I  G  K  N  F  F  K  D  I  A  P  C  T  D  T  E  E  F  G  G  R  F  R  E  G  V  A  N  G
 991  GACCTGAACGCGATGTTCGAATATGTCTTCGACTATCAGATGCAGCCGACCAAGGTGAAGGTGCACATGAAGCGCGCTATCACCGGCGAC 1080
       D  L  N  A  M  F  E  Y  V  F  D  Y  Q  M  Q  P  T  K  V  K  V  H  M  K  R  A  I  T  G  D
1081  AGCTACTGGATCTTCGTCAAGCGTGTCTGATCCTCTCGACGCCCCAGCACCCTCGGTCCCGGCGCGCGTCGCGTGGTGGCGCCGGCGCGA 1170
       S  Y  W  I  F  V  K  R  V  *
1171  TTTGCTGGTCCGGGTGGTCGCTGACTGTGTGGCGGATGCCGCCGCGCGCACGGTGCTGCCGGCCTTCTGGTCCGACGACACGCCGCTGAC 1260
1261  CGATGCGGCGGGGTCGGGTGACGGACCTGGGCTCGACTCGCTCGGCCTCGTCAATGCGGCCAGCGACCTGACCCGGCGCCTGCACCTGGA 1350
1351  CGAGAGTGGCCAGGAGGAACGCTTGCTGCGCCAGCGCACGTTCCCACACTGGGTCGAC                                 1408
```

Fig. 2. The *pyp* gene from *Rs.salexigens* with flanking regions. (**A**) Detailed physical map of the cloned 1.4 kb *Pvu*I–*Sal*I fragment from *Rs.salexigens*, indicating the position of the *pyp* gene. The open arrow indicates the direction of the gene. (**B**) DNA sequence of the 1.4 kb *Pvu*I–*Sal*I chromosomal fragment from *Rs.salexigens* containing the *pyp* gene with flanking regions. The derived amino acid sequence is given at the first position of each codon by the one letter code. The putative AT-rich promoter region (35 mol% GC) is underlined. The putative ribosome binding site is doubly underlined.

Results

The pyp genes from E.halophila and Rs.salexigens

The DNA sequence of a 1.8 kb *Pvu*II–*Pst*I 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 Gln 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 *Rs.salexigens* is the first gene cloned from this bacterium. It was localized on a 1.4 kb *Pvu*I–*Sal*I 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 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*, *Rs.salexigens* and *Chromatium salexigens* (Figure 3).

Comparison of PYP sequences

The complete amino acid sequences of the PYPs from *E.halophila*, *Rs.salexigens* and *C. salexigens* (Koh *et al.*,

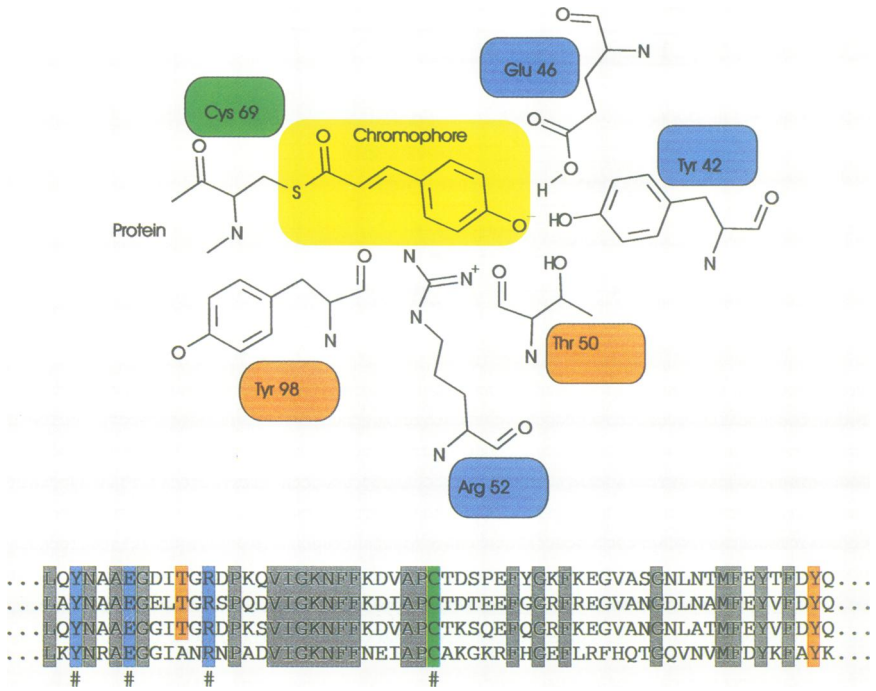


Fig. 3. Sequence conservation in the family of photoactive yellow proteins: the xanthopsins. Model for the *p*-coumaric acid binding pocket based on crystallographic data (Borgstahl *et al.*, 1995) and sequence conservation of the residues forming this pocket in the PYP sequences from *E.halophila*, *Rs.salexigens*, *C.salexigens*, and *R.sphaeroides*. 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.sphaeroides* (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*, *Rs.salexigens* 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.sphaeroides* 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 α4-helix and the β3-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 *Rs.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 *Rs.salexigens*. In line with this, comparison of the 1.8 and 1.4 kb chromosomal fragments from *E.halophila* and *Rs.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* BN9626 (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 *Rs.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 *Rs.salexigens* chromosomal digests with the putative *E.halophila pcl* (M.K.Phillips-Jones, unpublished observations).

Identification of the chromophore of Rs.salexigens PYP

The chromophore of *Rs.salexigens* PYP was identified as *p*-coumaric acid in the purified protein with high-

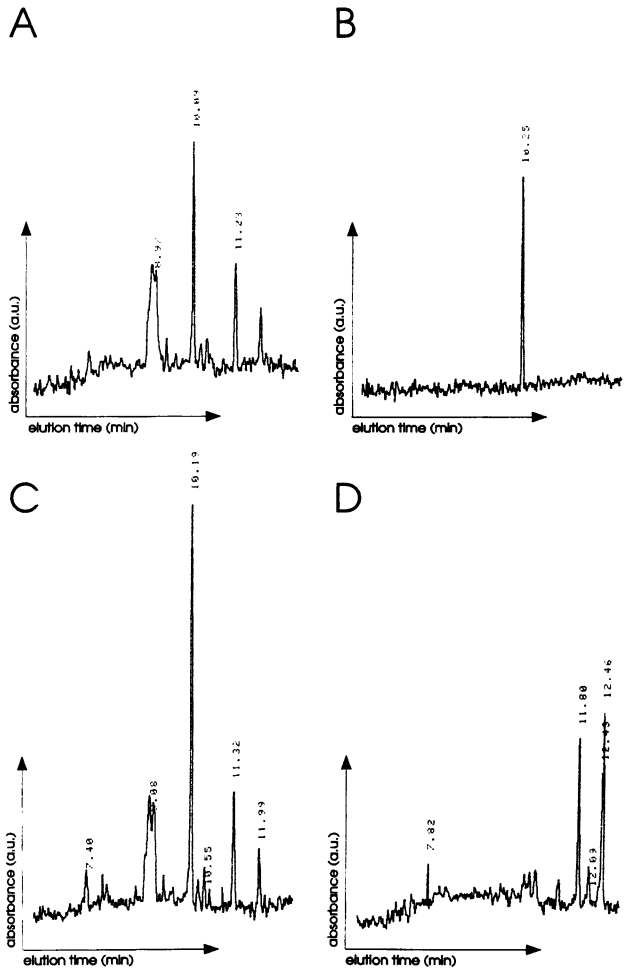


Fig. 4. Identification of the *Rs.salexigens* chromophore with capillary electrophoresis. (**A**) Electropherogram of ethyl acetate extract from soluble protein fraction of anaerobically grown *Rs.salexigens*; *p*-coumaric acid elutes at 10 min. (**B**) Electropherogram of *p*-coumaric acid, predominantly the *trans* isomer (Sigma). (**C**) Electropherogram of extracted chromophore from anaerobically grown *Rs.salexigens* co-injected with *p*-coumaric acid, showing an increase of the *p*-coumaric acid peak at 10 min. (**D**) Electropherogram of extracted chromophore from aerobically grown *Rs.salexigens*.

performance capillary zone electrophoresis (data not shown), which uses the electrophoretic mobility of ions as separation principle (for a review see Karger *et al.*, 1989). After injection and electrophoresis of an ethyl acetate extract from the soluble protein fraction of *Rs.salexigens*, the electropherogram shows a major component at 10 min (Figure 4A), in an amount of 0.1 pmol of *p*-coumaric acid (see Materials and methods), which corresponds with 8 pmol of detected *p*-coumaric acid per mg soluble protein. As a control, co-elution of *p*-coumaric acid (Figure 4B) with the chromophore in the extraction mixture was demonstrated by the increase in size of the peak at 10 min (Figure 4C). Furthermore, our analysis shows that no *p*-coumaric acid is bound to soluble proteins in aerobically grown *Rs.salexigens* cells (Figure 4D), which is independent proof of regulation of PYP expression in this organism (compare Hoff *et al.*, 1994b).

Heterologous overproduction of the E.halophila PYP

To overexpress PYP from *E.halophila*, a 0.45 kb *Ava*II 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 oberved 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 I), 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 mutated 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 I) overproduces PYP at levels of 50 mg/l culture per OD_{660} 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 indistinguishable from native *E.halophila* apoPYP (Figure 5B). The absorbance spectrum of the isolated histidine-tagged PYP shows that the typical absorbance band in the 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). The

Table I. Strains and plasmids used in this study

Strains and plasmids	Description	Source or reference
Strain		
E.coli BL21	*hsd*S, *gal*, (λcI*ts* 857*ind*1, *S*am7, *nin*5, *lac* UV5-T7 gen 1)	Studier and Moffat (1986)
E.coli M15[pREP4]	expression host with repressor plasmid, Km^R	Qiagen
E.coli TG1	*supE*, Δ(*lac-proAB*), *hsd*Δ5, F'[*traD*36, *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15]	Gibson (1984)
E.coli S17-1	RP4-2(Tc::Mu)(Km::Tn*7*), *thi*, *pro*, *hsd*R, *hsd*M+, *rec*A, Tp^R, Sm^R	Simon *et al.* (1983)
E.halophila SL1	type strain	Raymond and Sistrom (1969)
R.sphaeroides 2.4.1	type strain	Van Niel (1944)
R.sphaeroides DD13	RC-, LH1-, LH2-, Km^R, Sm^R	Jones *et al.* (1992)
Rs.salexigens WS 68	type strain	Drews (1981)
Plasmid		
pCHB500	pRK415 and pSH3 derivative, Tc^R	Benning and Sommerville (1992)
pART3	0.45 kb *E.halophila* *Ava*II fragment cloned into pCHB500	this study
pQE30	RBSII, 6×His tag, ColE1 *ori*, amp^R	Qiagen
pHisp	0.42 kb *E.halophila* PCR product cloned into pQE30	this study
pT713	expression vector, T7 promoter, Amp^R	Gibco BRL
pTY13	0.45 kb *E.halophila* *Ava*II fragment cloned into pT713	this study
M13mp18/19	M13mp1 derived phages, *lacZ'*	Messing and Vieira (1982)
pYAMA18	2.4 kb *E.halophila* *Pst*I fragment cloned into M13mp18	this study
pYAMA958	1.8 kb *E.halophila* *Pvu*II fragment cloned into M13mp18	this study
pS16	1.4 kb *Rs.salexigens* *Pvu*I–*Sal*I fragment cloned into M13mp19	this study

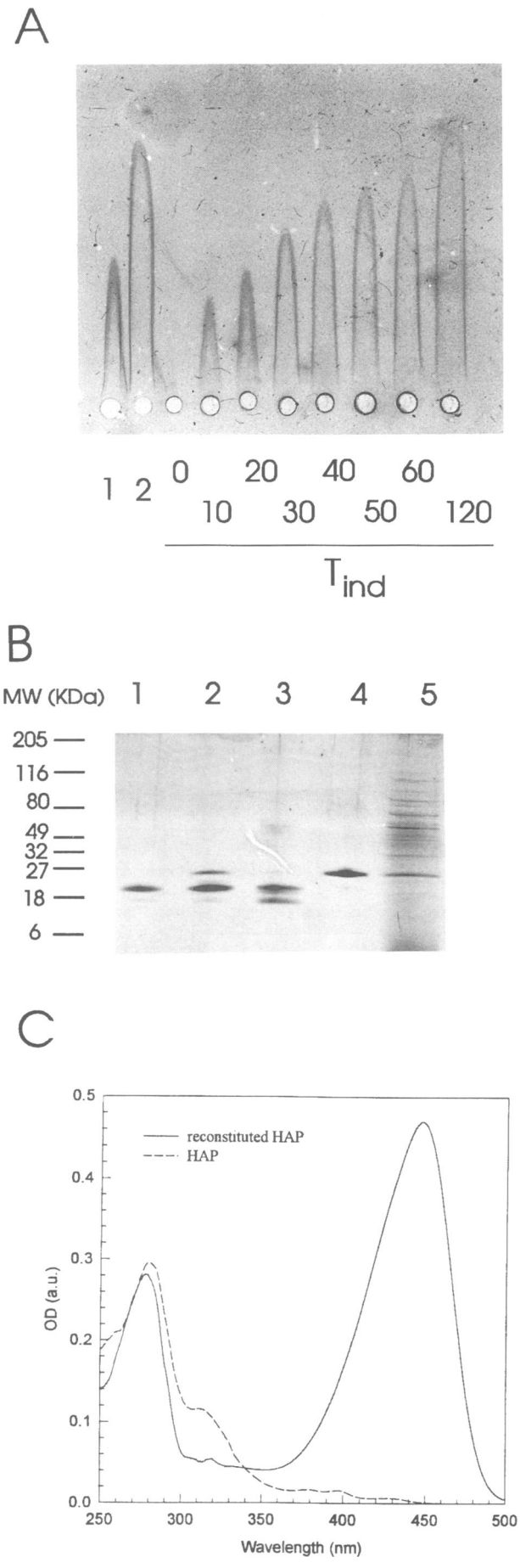
masses of the histidine-tagged holo- and apoPYP were determined by ESMS to be respectively, 16.0081 and 15.8625 kDa. These values correspond well to the calculated molecular weights of 16.0081 and 15.8611.

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 (Baca *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 *E.halophila* BN9626 strain to G or C in the *E.halophila* SL-1 strain. This may be explained by a slight difference in the overall GC-content between the two strains, which have been isolated from different environments; the BN9626 strain was isolated from the Wadri Natrun, Lake Abu Gabara near Bir Hooker, Egypt (Imhoff *et al.*, 1978) and the type strain SL-1 from Summer Lake, OR, USA (Raymond and Sistrom, 1969). The GC-content of the cloned DNA fragments from *E.halophila* SL1 and *Rs.salexigens* was calculated to be 67.3 and 65.8% respectively, which matches well with the overall GC-content from these organisms, being 68.4% (Raymond and Sistrom, 1969) and 64 ± 2% (Drews, 1981) respectively. The lack of a signal peptide sequence upstream from the two *pyp* genes is in line with the intracellular localization of PYP in *E.halophila*, as determined with immuno-gold labeling experiments (Hoff *et al.*, 1994b). Furthermore, the isoelectric points of the PYPs from *E.halophila* and *Rs.salexigens* are predicted to be 4.63 and 4.23 respectively. For *E.halophila* PYP, this parameter was experimentally determined to be 4.3 (McRee *et al.*, 1986).

The sequence data for these two PYPs were used to design primers for the amplification of a fragment from chromosomal DNA by heterologous PCR, leading to the identification of a PYP homolog in *R.sphaeroides*. The PCR product obtained was used as a probe to clone the *R.sphaeroides 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 PYPs from *E.halophila*, *Rs.salexigens* and *C. salexigens*, indicating that this PYP belongs to a different subgroup 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 phototaxis response(s), based on PYP (see Sprenger *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 adenine-binding motif (Saraste *et al.*, 1990). 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 importance of this



esterification was demonstrated by the fact that *in vitro* reconstitution of holoPYP was observed with the thiophenyl 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 *pcl* gene products is the presence of an inverted repeat between these two coding regions and the absence of a recognizable promotor sequence, directly upstream of the *pcl* gene (see Figure 1B). This indicates that transcription of the *pcl* 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 archaebacterial 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 *htrI* 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 ξανθος

Fig. 5. Overproduction, purification and *in vitro* reconstitution of histidine-tagged PYP. (**A**) RIEP analysis of PYP production in *E.coli* M15/pHisp after induction with IPTG. Wells 1 and 2 contain solutions of purified PYP from *E.halophila* with known concentrations; the following wells contain cell material from *E.coli* M15/pHisp taken at the indicated induction times (T_{ind} in min) after the addition of IPTG. (**B**) SDS–PAGE of cell-free extracts from *E.coli* M15/pHisp (lane 5), histidine-tagged PYP isolated from this extract by Ni affinity chromatography (lane 4), the same preparation after 5 h (lane 2) and 24 h (lane 3) of incubation with enterokinase, and PYP purified from *E.halophila* (lane 1). (**C**) Absorbance spectrum of the histidine-tagged PYP (HAP) isolated from *E.coli* M15/pHisp and the spectrum of HAP after reconstitution with the *p*-coumaric anhydride and subsequent purification.

Table II. Homology of the putative coumaryl-CoA ligase from *E.halophila* with CoA ligases from other organisms

Enzyme (number of amino acids)	Organism	Identity (%)	Similarity (%)	Reference
CoA ligase homolog (391)	*E.halophila*	100	100	this paper
Acetate-CoA ligase (660)	*A.eutrophus*	25.1	49.5	Priefert and Steinbuechel (1992)
Acetate-CoA ligase (672)	*M.soehngenii*	20.6	47.5	Eggen *et al.* (1991)
Long-chain-fatty-acid-CoA ligase (558)	*E.coli*	26.3	51.5	Black *et al.* (1992)
Long-chain-fatty-acid-CoA ligase (700)	yeast	22.8	47.5	Duronio *et al.* (1992)
Coumaryl-CoA ligase (545)	potato	22.4	45.5	Becker-Andre *et al.* (1991)
Coumaryl-CoA ligase (563)	rice	25.3	49.2	Zhao *et al.* (1990)

Identity and similarity values are based on full length alignments made with the Genetics Computer Group package program BESTFIT using a gap weight of 3.0 and a length weight of 0.1.

(yellow) and οψις (eyesight). The bacterial xanthopsins resemble the archaebacterial sensory rhodopsins at the level of photochemistry (Hoff *et al.*, 1994c), as well as of function, which is proposed to be that of a photosensor in negative phototaxis (Sprenger *et al.*, 1993). Further evidence for the xanthopsins, as a eubacterial protein family, has been obtained by studies with a highly specific polyclonal antiserum against *E.halophila* 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 xanthopsins as a protein family of photosensors 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* holoPYP synthesis. The heterologously produced apoPYP was used as substrate for *in vitro* holoPYP 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.

Materials and methods

Bacterial strains and plasmids

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

Cell culturing

E.halophila SL-1 (Raymond and Cistrom, 1969) and *Rs.salexigens* WS68 (Drews, 1981) were cultured phototrophically as described (Meyer, 1985 and 1990 respectively), unless specified otherwise. *R.sphaeroides* strain 2.4.1 (van Niel, 1944) was grown aerobically in Luria Bertani broth.

DNA manipulation

Chromosomal DNA was isolated according to standard procedures (Sambrook *et al.*, 1989) from *E.halophila*, *Rs.salexigens* and *R.sphaeroides*. All additional molecular genetic techniques were performed as described in Sambrook *et al.* (1989).

Southern hybridization

Southern blots of chromosomal DNA from both *E.halophila* and *Rs.salexigens* were probed using a 94 bp PCR product consisting of an internal fragment from the *E.halophila pyp* gene (see below). The probe was labeled with the Klenow enzyme by random priming using the DIG DNA labeling kit and detected with Nitroblue tetrazolium salt, as described by the manufacturer (Boehringer, Mannheim). Southern blots of chromosomal DNA from *E.halophila* and *R.sphaeroides* were hybridized at 65°C and washed at 65°C with 0.1× SSC buffer containing 0.1% SDS. The blots containing chromosomal DNA from *Rs.salexigens* were hybridized at 50°C and washed at 50°C with 0.5× SSC buffer containing 0.1% SDS.

Cloning of the E.halophila pyp gene

*Pst*I-digested *E.halophila* chromosomal DNA was used as template in a PCR-reaction with degenerated oligonucleotides YS-1 and YS-2 with the sequences AARAAYTTYTTYAARGA and GTCATYTGMTARTCRAA respectively, as based on the PYP amino acid sequence (Van Beeumen *et al.*, 1993). PCR was performed with the enzyme *Taq* polymerase (HT Biotechnology, Cambridge, UK) for 30 cycles with 1 min denaturation at 94°C, 1 min annealing at 20°C and 1 min elongation at 70°C. Based on the sequence of the PCR product a new probe was constructed, completely homologous to the *pyp* gene in *E.halophila*. This probe was used to isolate a positive clone (pYAMA18) by screening a mini library of 2.4 kb *Pst*I chromosomal fragments from *E.halophila* in phage M13mp18. A 950 bp *Pvu*II fragment from pYAMA18, containing the *pyp* ORF, was subcloned in M13mp18 to give pYAMA958.

Cloning of the Rs.salexigens pyp gene

The probe used to clone the *pyp* gene from *E.halophila* was used in heterologous Southern hybridization experiments with *Rs.salexigens* chromosomal digests. A mini library, containing sized *Pvu*I–*Sal*I fragments in phage M13 was screened by hybridization with the same probe, leading to the identification of two positive clones. A 1.4 kb fragment containing the *pyp* gene was made blunt by Klenow treatment and reinserted into the *Sma*I linearized phage M13mp19, yielding pS16.

Sequencing

Both strands of the 1.8 kb *E.halophila Pvu*II–*Pst*I fragment and the *Rs.salexigens* 1.4 kb *Pvu*I–*Sal*I fragment were sequenced using universal and gene-specific oligonucleotides; the sequence strategies are indicated in Figures 1A and 2A. Sequence information was obtained by the dideoxy chain termination method (Sanger *et al.*, 1977), using [^{35}S]dATP and a modified T7 DNA polymerase sequencing kit (Sequenase; US Biochemical Corporation, Cleveland, OH), as well as through the use of fluorescently labeled dideoxy nucleotides and a thermostable *Taq* polymerase with the Dyedeoxy terminator cycle sequencing kit (Applied Biosystems, Foster City).

Identification of the R.sphaeroides pyp gene

Chromosomal DNA from *R.sphaeroides* 2.4.1 was used as template in a PCR using 10 cycles of annealing for 1 min at 25°C and 25 cycles at 35°C. Denaturation and elongation were performed in all 35 cycles for 1 min at 95°C and 72°C respectively. Primers were based on known *pyp* sequences and restriction sites *Bam*HI and *Hin*dIII (underlined) were introduced to enable directional cloning: GCGGATCCGCCTTCGGCGCCATCCAGCTCGAC (NTPYP1) and GCGCAAGCTTCTAGACGCGCTTGACGAAGACCC (CTPYP1). The PCR product obtained was isolated from agarose gel and inserted into phages M13mp18/19. Both strands of the PCR product were sequenced. Hybridization of the PCR product with *R.sphaeroides* chromosomal DNA was performed as described (Engler-Blum *et al.*, 1993).

Identification of the chromophore of Rs.salexigens PYP

A colorless *Rs.salexigens* culture, grown aerobically in the dark in Hutner 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 500 ml of aerobically and anaerobically grown cultures was prepared as described (Hoff *et al.*, 1994b). Proteins were precipitated with 10% (v/v) trichloro-acetic acid and washed once with demineralized water. Pellets were resuspended in 5 ml demineralized water and incubated overnight at pH 12 (leading to a complete solublization 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 by 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 demineralized water and dried by air. To substantiate the result of our analysis, the same chromophore extraction procedure was carried out using the purified *Rs.salexigens* PYP (Meyer *et al.*, 1990). Air-dried samples were dissolved in distilled water and injected in a 50 μm fused silica capillary TSP050375 (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/30 mM valeric acid pH 8.2, through a capillary with an effective length of 55 cm, at 25 kV and ~12 μA. On-column detection was performed at 284 nm (determined as the wavelength at which *trans p*-coumaric acid maximally absorbs in the Tris/valeric acid buffer), with a UVIS 200 detector (Linear, Fremont). As a reference *trans p*-coumaric acid (Sigma, St Louis, MO) was used. To confirm this identification, *p*-coumaric acid was also subjected to electrophoresis in 25 mM borax buffer, pH 9.0 at 25 kV and ~35 μA. The amount of detected *trans p*-coumaric acid was calculated from the peak area using the software Caesar for Windows (version 4.02, 1990, Prince Technologies). As a reference, 11.0 nl 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 0.45 kb *Ava*II fragment from pYAMA958, containing the *pyp* ORF from *E.halophila*, was ligated into the *Sma*I-linearized overexpression plasmid pT713 (Studier *et al.*, 1990) to yield pTY13, which was transformed to *E.coli* BL21. Overexpression in pT713 is based on the strong viral T7 promoter ϕ10. The gene coding for the viral RNA polymerase is located on the chromosome of *E.coli* BL21, downstream of an inducible *lac* promoter (Studier *et al.*, 1990).

A conjugative broad host range overexpression system was constructed by ligating the 0.45 kb *Ava*II fragment, described above, into the *Pst*I polylinker site of pCHB500. pCHB500 is a broad host range vector, containing two promoters directly upstream of the polylinker site: the *E.coli* P_{lac} promoter and the P_{cyc} promoter that supports anaerobic expression of the *cycA* gene from *R.capsulatus* (Bennig and Sommerville, 1992). The resulting plasmid pART3 was transformed into the conjugative strain *E.coli* S17 and then transferred to *R.sphaeroides* DD13 (Jones *et al.*, 1992) by conjugation on LB agar plates for 4.5 h. Transconjugants were selected on LB plates containing tetracyclin (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 pigment synthesis.

A third overexpression system involved the heterologous overproduction of an affinity-tagged version of PYP from *E.halophila* 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 GCGGATCCGATGACGATGACAAAATGGAACACGTA-GCCTTCGG (NTPYP2), containing the *Bam*HI site (underlined) and CTPYP1 (see above). Use of NTPYP2 results in the presence of an enterokinase site in the recombinant protein, allowing proteolytic removal of the affinity tag. This tag is formed by six His residues, encoded by pQE30 (Qiagen). The PCR was performed using an annealing temperature of 60°C for 30 s and extension at 70°C for 30 s in 30 cycles. The resulting PCR product was digested with *Bam*HI and *Hin*dIII, ligated into pQE30 (Qiagen) to yield pHisp and 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 Schägger and Jagow (1987) for improvement of resolution in the 5–20 kDa range. Gels were stained with Coomassie brilliant blue G250. Western blotting and immunodecoration were performed as described previously (Towbin *et al.*, 1979; Hoff *et al.*, 1994b). RIEP was carried out as described (Hoff *et al.*, 1994b).

Heterologous expression of PYP

E.coli BL21/pTY13 and *E.coli* M15/pHisp were induced to express the heterologous gene by the addition of 1 mM IPTG to well-aerated cultures of exponentially growing cells at an OD_{660} of 1. Cells were grown at 37°C in well-shaken Erlenmeyers, or in a well-aerated 10 l fermentor (New Brunswick Scientific, New Brunswick). Production of PYP in *R.sphaeroides* was induced by growing the organism semi-anaerobically in two-thirds filled, slowly shaking Erlenmeyers, using Luria Bertani broth with appropriate antibiotics. The resulting *E.coli* and *R.sphaeroides* cells were sonified three times for 1 min 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 Aminco DW2000 spectrophotometer (SLM 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/pHisp, 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 eluted, either by an imidazole gradient or by a pH gradient, as described by the manufacturer. The protein elution pattern was analyzed by measuring the absorbance 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 *E.halophila* (Imamoto *et al.*, 1995). The *p*-coumaric anhydride was synthesized as described (Imamoto *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 pmol of protein was dissolved in 10 ml CH_3CN:water:formic acid (1:0.9:0.1; v/v) and injected into the electrospray source of a VG Bio-Q mass spectrometer (VG Organic, Altrincham, UK) at a flow rate of 6 ml/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 masslynx software provided with the instrument.

Acknowledgements

The authors are very grateful to R.Kok and J.van Thor for their advice on the use of molecular genetic techniques and to J.van Dijk and W.Spijker for cloning the *Rs.salexigens pyp* gene. We thank X.Xu and H.Vonk for performing the capillary electrophoresis experiments and H.P.M.Fennema for synthesis of the *p*-coumaric anhydride. We would like to thank B.Poolman for his help with the initial PCR experiments. We are very thankful to T.E.Meyer for supplying information concerning oligonucleotides YS-1 and YS-2 and the purified *Rs.salexigens* PYP. W.D.Hoff was supported by the Netherlands Organization for Scientific Research (NWO) via the Foundation for Biological Research (BION). J.Van Beeumen is indebted to the National Fund for Joint Basic Research for financial support (Contract 32001891).

References

Armitage,J.P. (1992) *Annu. Rev. Physiol.*, **54**, 683–714.
Armitage,J.P. and Mcnab,R.M. (1987) *J. Bacteriol.*, **169**, 514–518.
Baca,M., Borgstahl,G.E.O., Boissinot,M., Burke,P.M., Williams,W.R., Slater,K.A. and Getzoff,E.D. (1994) *Biochemistry*, **33**, 14369–14377.
Bader,J., Rauschenbach,P. and Simon,H. (1982) *FEBS Lett.*, **140**, 67–72.
Becker-Andre,M., Schulze-Lefert,P. and Hahlbrock,K. (1991) *J. Biol. Chem.*, **266**, 8551–8559.
Bennig,C. and Sommerville,C. (1992) *J. Bacteriol.*, **174**, 2352–2360.

Black,P.N., Dirusso,L.L., Metzyer,A.K. and Heimert,T.L. (1992) *J. Biol. Chem.*, **267**, 255113–25520.
Borgstahl,G.E.O., Williams,D.R. and Getzoff,E.D. (1995) *Biochemistry*, **34**, 6278–6287.
Drews,G. (1981) *Arch. Microbiol.*, **130**, 325–327.
Duronio,R.J., Knoll,L.J. and Gordon,J.I. (1992) *J. Cell. Biol.*, **117**, 515–529.
Eggen,R.I.L., Geerling,A.C.M., Boshoven,A.B.P. and de Vos,W.M. (1991) *J. Bacteriol.*, **173**, 6383–6389.
Engler-Blum,G., Meier,M., Frank,J. and Müller,G.A. (1993) *Anal. Biochem.*, **210**, 235–244.
Ferrando-May,E., Krahl,M., Marwan,W. and Oesterhelt,D. (1993) *EMBO J.*, **12**, 2999–3005.
Fulda,M., Heinz,E. and Wolter,F.P. (1994) *Mol. Gen. Genet.*, **242**, 241–249.
Gibson,T.J. (1984) In *Studies on the Epstein-Barr Virus Genome*. Ph.D. thesis, Cambridge University, UK.
Goodwin,T.W. and Mercer,E.I. (1983) *Introduction to Plant Biochemistry*. 2nd edn. Pergamon Press, Oxford, UK.
Hahlbrock,K. and Scheel,D. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **40**, 347–369.
Hoff,W.D. *et al.* (1994a) *Biochemistry*, **33**, 13959–13962.
Hoff,W.D., Sprenger,W.W., Postma,P.W., Meyer,T.E., Veenhuis,M., Leguijt,T. and Hellingwerf,K.J. (1994b) *J. Bacteriol.*, **176**, 3920–3927.
Hoff,W.D., Van Stokkum,I.H.M., Van Ramesdonk,H.J., Van Brederode, M.E., Brouwer,A.M., Fitch,J.C., Meyer,T.E., Van Grondelle,R. and Hellingwerf,K.J. (1994c) *Biophys. J.*, **67**, 1691–1705.
Imamoto,Y., Ito,T., Kataoka,M., and Tokunaga,F.(1995) *FEBS Lett.*, **374**, 157–160.
Imhoff,J.F., Hashwa,F., and Trüper,H.G. (1978) *Arch. Hydrobiol.*, **84**, 381–388.
Jones,M.R., Fowler,G.J.S., Gibson,L.C.D., Grief,G.G., Olsen,J.D., Crielaard,W. and Hunter,C.N. (1992) *Mol. Microbiol.*, **6**, 1173–1184.
Karger,B.L., Cohen,A.S. and Guttman,A. (1989) *J. Chromatogr.*, **492**, 585–614.
Kim,M., Mathies,R.A., Hoff,W.D. and Hellingwerf,K.J. (1995) *Biochemistry*, **34**, 12669–12672.
Koh,M., Van Driessche,G., Samyn,B., Hoff,W.D., Meyer,T.E., Cusanovich,M.A. and Van Beeumen,J.J. (1996) *Biochemistry*, **35**, 2526–2534.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
McRee,D.E., Meyer,T.E., Cusanovich,M.A., Parge,H.E. and Getzoff,E.D. (1986) *J. Biol. Chem.*, **261**, 13850–13851.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Meyer,T.E. (1985) *Biochim. Biophys. Acta*, **806**, 175–183.
Meyer,T.E., Yakali,E., Cusanovich,M.A. and Tollin,G. (1987) *Biochemistry*, **26**, 418–423.
Meyer,T.E., Fitch,J.C., Bartsch,R.G., Tollin,G. and Cusanovich,M.A. (1990) *Biochim. Biophys. Acta*, **1016**, 364–370.
Olsiewski,P.J., Kaczorowski,G.J. and Walsh,C. (1980) *J. Biol. Chem.*, **255**, 4487–4494.
Priefert,H. and Steinbuechel,A. (1992) *J. Bacteriol.*, **174**, 6590–6599.
Raymond,J.C. and Sistrom,W.R. (1969) *Arch. Microbiol.*, **69**, 121–126.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
Saraste,M., Sibbald,P.R. and Wittighofer,A. (1990) *Trends. Biochem. Sci.*, **15**, 430–434.
Schägger,H. and von Jagow,G. (1987) *Ann. Biochem.*, **166**, 368–379.
Senger,H. (1987) *Blue Light Responses: Phenomena and Occurrence in Plants and Microorganisms*. CRC Press, Boca Raton, FL.
Simon,R., Priefer,U. and Puhler,A. (1983) *Biotechnology*, **1**, 845–856.
Sprenger,W.W., Hoff,W.D., Armitage,J.P. and Hellingwerf,K.J. (1993) *J. Bacteriol.*, **175**, 3096–3104.
Spudich,J.L. (1994) *Cell*, **79**, 747–750.
Studier,F.W. and Moffat,B.A. (1986) *J. Mol. Biol.*, **89**, 113–130.
Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 4350–4354.
Van Beeumen,J., Devreese,B., Van Bun,S., Hoff,W.D., Hellingwerf,K.J., Meyer,T.E., McRee,D.E. and Cusanovich,M.A. (1993) *Protein Sci.*, **2**, 1114–1125.
Van Niel,C.B. (1944) *Bacteriol. Rev.*, **8**, 1–118.
Woese,C.R. (1987) *Microbiol. Rev.*, **51**, 221–271.
Yao,V.J. and Spudich,J.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11915–11919.
Zhao,Y., Kung,S.D. and Dube,S.K. (1990) *Nucleic Acids Res.*, **18**, 6144.

Received on December 14, 1995; revised on February 26, 1996

Note added in proof

Recent results cast doubt on our strain assignment in *E.halophila*.