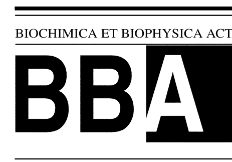
Biochimica et Biophysica Acta 1385 (1998) 1–6



Short sequence-paper

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 salexigens* and *Chromatium salexigens* [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

* Corresponding author.
[1] The nucleotide sequences reported in this paper have been submitted to the EMBL nucleotide sequence database under accession numbers AJ002398 and X98889.

0167-4838/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0167-4838(98)00050-8

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 *Ase*I 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 *Pst*I fragment, which was cloned into the *Pst*I-digested cloning vector pBS SK$^+$ (Stratagene, La Jolla, CA), resulting in pATC3. To obtain single-stranded DNA for sequencing, a 0.5-kb *Pst*I *Bam*HI 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*, *pcl* 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 S_1 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-

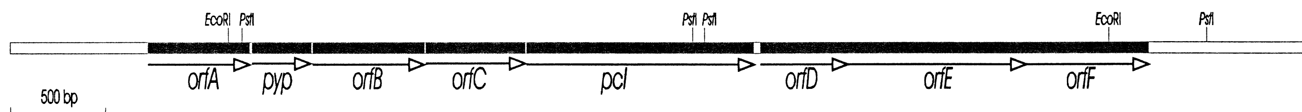


Fig. 1. *R. sphaeroides* NCIB8253 6.5-kb DNA fragment containing *Orfs A–F*, *pyp* and *pcl* genes. Positions of the restriction sites of *Eco*RI and *Pst*I are indicated in italics.

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

ORF	AA	Position	rc	FASTA	AL/gaps	sim/id	Ref.
orfA	185	717–1274	5				
pyp	124	1299–1673	0	PYP	125/1	64/49	[5]
orfB	188	1688–2254	5				
orfC	166	2251–2751	2				
pcl	411	2763–3998	0	PCL	413/12	55/36	[5]
orfD	176	4031–4561	5				
orfE	336	4561–5571	14				
orfF	227	5571–6254	0	SRI	232/5	47/22	[19]

The columns indicate the number of encoded amino acids (AA), the position on the DNA fragment (position), the number of rare codons (rc), 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.*

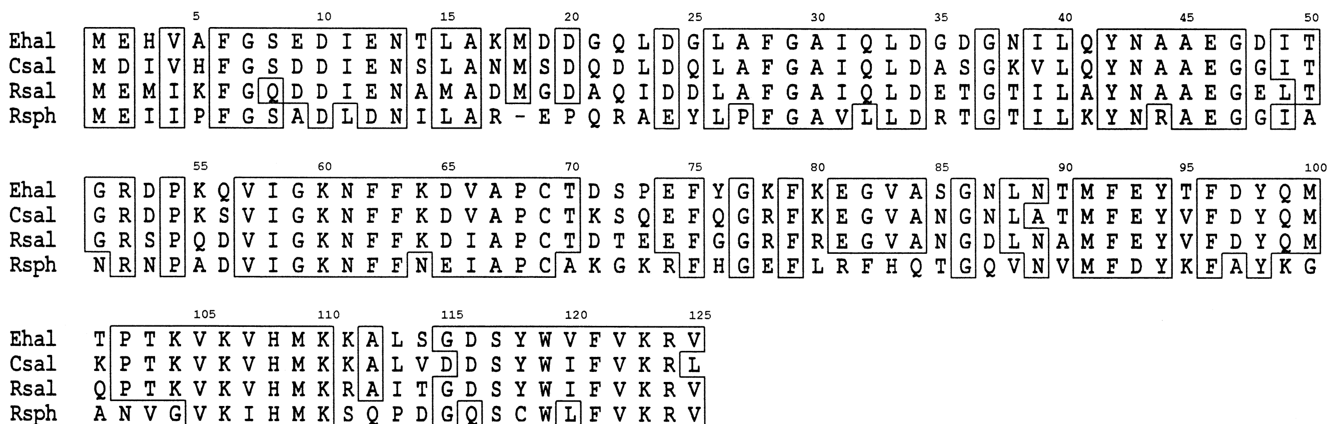


Fig. 2. Multiple sequence alignment of the 4 known PYP amino acid sequences. Abbreviations: Ehal = *E. halophila*, Csal = *C. salexigens*, Rsal = *Rh. salexigens*, Rsph = *R. sphaeroides*. Sequences have been aligned with the programs PILEUP and PRETTYPLOT, using default settings.

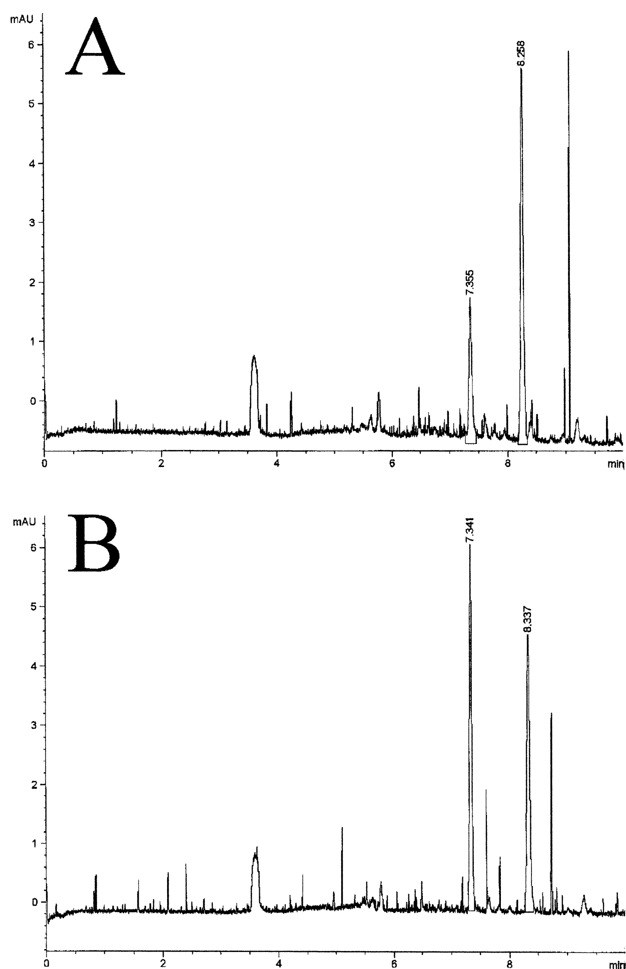


Fig. 3. Electropherograms of ethyl acetate extracts from *R. sphaeroides* RK1 cells. The eluate was analysed at 284 nm. Trace (A) shows 4-hydroxy cinnamic acid at 7.3 min and an unidentified compound at 8.3 min. Trace (B) shows the result of co-injection analysis of the extract with 4-hydroxy cinnamic acid (Sigma), showing enhancement of the peak at 7.3 min.

sphaeroides are strongly reminiscent of the information available for the purple bacterium *Rs. salexigens*, 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. halophila* 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. halophila* 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

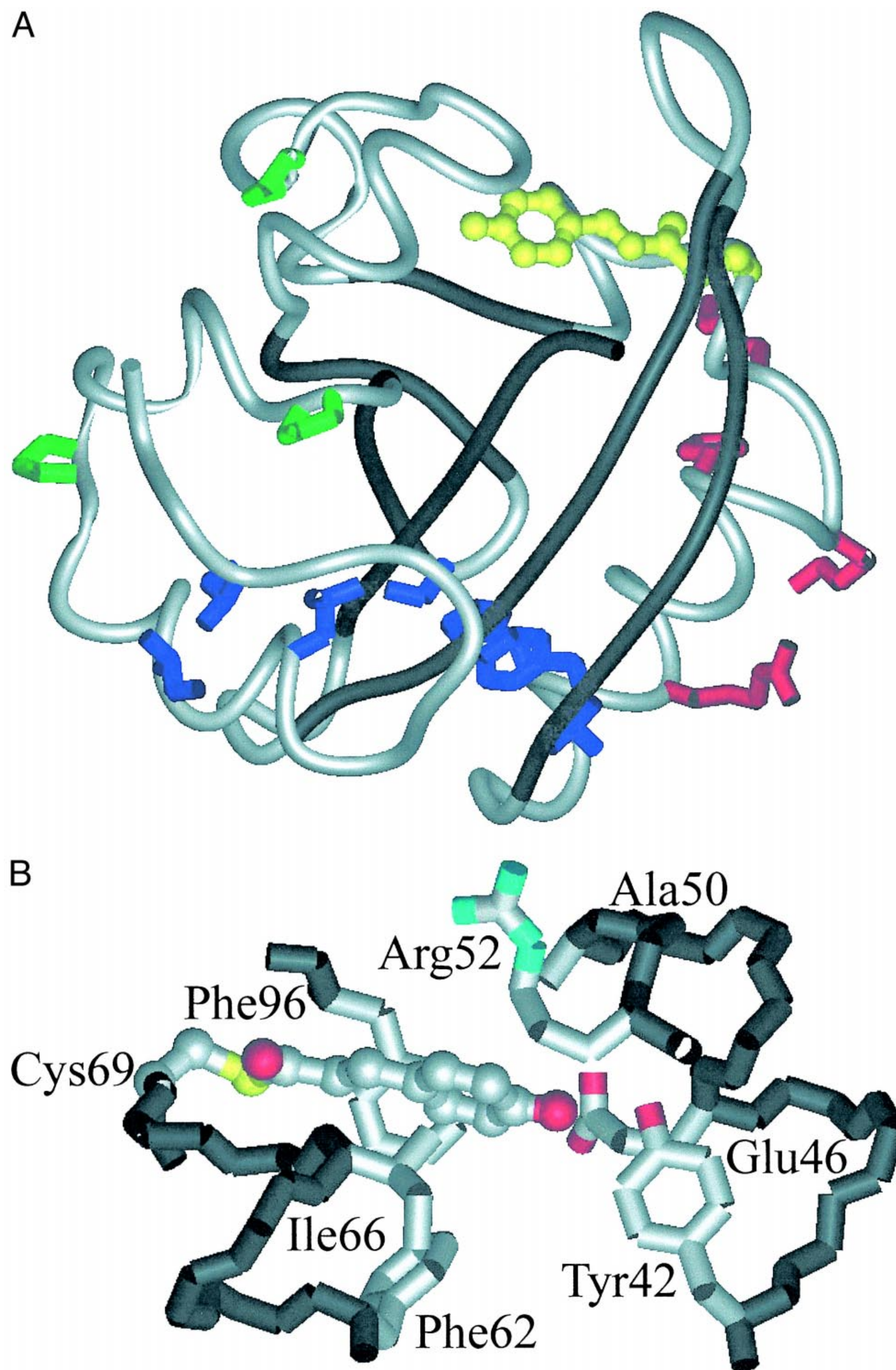


Fig. 4. Spatial model for *R. sphaeroides* PYP. (A) Overall view of the model. The C-α trace is coloured grey, using a darker shade for the central β sheet. The four positively charged amino acids, which are part of the new patch around residue 75 (see text) are coloured red. Three prolines leading to 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.

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

[1] T.E. Meyer, Biochim. Biophys. Acta 806 (1985) 175–183.
[2] T.E. Meyer, J.C. Fitch, R.G. Bartsch, G. Tollin, M.A. Cusanovich, Biochim. Biophys. Acta 1016 (1990) 364–370.
[3] M. Koh, G. Van Driessche, B. Samyn, W.D. Hoff, T.E. Meyer, M.A. Cusanovich, J.J. Van Beeumen, Biochemistry 35 (1996) 2526–2534.
[4] M. Baca, G.E.O. Borgstahl, M. Boissinot, P.M. Burke, R.W. DeWight, K.A. Slater, E.D. Getzoff, Biochemistry 33 (1994) 14369–14377.
[5] R. Kort, W.D. Hoff, M. Van West, A.R. Kroon, S.M. Hoffer, K.H. Vlieg, W. Crielaard, K.J. Hellingwerf, EMBO J. 15 (1996) 3209–3218.
[6] W.W. Sprenger, W.D. Hoff, J.P. Armitage, K.J. Hellingwerf, J. Bacteriol. 175 (1993) 3096–3104.
[7] T.E. Meyer, E. Yakali, M.A. Cusanovich, G. Tollin, Biochemistry 6 (1987) 418–423.
[8] W.D. Hoff, I.H.M. Van Stokkum, H.J. Van Ramesdonk, M.E. Van Brederode, A.M. Brouwer, J.C. Fitch, T.E. Meyer, R. Van Grondelle, K.J. Hellingwerf, Biophys. J. 67 (1994) 1691–1705.
[9] G.E.O. Borgstahl, R.W. DeWight, E.D. Getzoff, Biochemistry 34 (1995) 6278–6287.
[10] U. Genick, G. Borgstahl, K. Ng, Z. Ren, C. Pradervand, P.M. Burke, V. Srajer, T.-Y. Teng, W. Schildkamp, D.E. McRee, K. Moffat, E.D. Getzoff, Science 275 (1997) 1471–1475.
[11] W.D. Hoff, P. Düx, K. Hard, B. Devreese, I.M. Nugteren-Roodzant, W. Crielaard, R. Boelens, R. Kaptein, J.J. Van Beeumen, K.J. Hellingwerf, Biochemistry 33 (1994) 13959–13962.
[12] M. Kim, R.A. Mathies, W.D. Hoff, K.J. Hellingwerf, Biochemistry 34 (1995) 12669–12672.
[13] R. Kort, H. Vonk, X. Xu, W.D. Hoff, W. Crielaard, K.J. Hellingwerf, FEBS Lett. 382 (1996) 73–78.
[14] A. Xie, W.D. Hoff, A.R. Kroon, K.J. Hellingwerf, Biochemistry 35 (1996) 14671–14678.
[15] Y. Imamoto, K. Mihara, O. Hisatomi, M. Kataoka, F. Tokunaga, N. Bojkova, K. Yoshihara, J. Biol. Chem. 272 (1997) 12905–12908.
[16] C.N. Hunter, S.A. Coomber, J. Gen. Microbiol. 134 (1988) 1491–1497.
[17] A.P.N. De Boer, W.N.M. Reijnders, J.G. Kuenen, A. Stouthamer, R.J.M. Spanning, Antonie van Leeuwenhoek 66 (1994) 111–127.
[18] M. Choudhary, C. Mackenzie, K. Nereng, E. Sodergren, G.M. Weinstock, S. Kaplan, Microbiology 143 (1997) 3085–3099.
[19] A. Blanck, D. Oesterhelt, E. Ferrando, E.S. Schegk, F. Lottspeich, EMBO J. 8 (1989) 3963–3971.
[20] D.M. Lagarias, S.-H. Wu, J.C. Lagarias, Plant Mol. Biol. 29 (1995) 1127–1142.
[21] I.B. Zhulin, B.L. Taylor, R. Dixon, TIBS 22 (1997) 331–333.
[22] W.D. Hoff, W.W. Sprenger, P.W. Postma, T.E. Meyer, M. Veenhuis, T. Leguijt, K.J. Hellingwerf, J. Bacteriol. 176 (1994) 3920–3927.
[23] B. Thiemann, J.F. Imhoff, Biochim. Biophys. Acta 1253 (1995) 181–188.
[24] M. Fulda, E. Heinz, F.P. Wolter, Mol. Gen. Genet. 242 (1994) 241–249.
[25] G. Vriend, J. Mol. Graph. 8 (1990) 52–56.
[26] G. Vriend, C. Sander, J. Appl. Crystallogr. 26 (1993) 47–60.
[27] W.F. Van Gunsteren, H.J.C. Berendsen, Gromos Manual. BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry, Univ. of Groningen, The Netherlands, 1987.
[28] D.M.F. Van Aalten, W.D. Hoff, J.B.C. Findlay, K.J. Hellingwerf, Protein Eng. (1997) submitted.