Chinese Science Bulletin 2006 Vol. 51 No. 8 889-896

DOI: 10.1007/s11434-008-0889-x

Proteorhodopsin—A new path for biological utilization of light energy in the sea

JIAO Nianzhi, FENG Fuying & WEI Bo

State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361005, China

Correspondence should be addressed to Jiao Nianzhi (email: Jiao@. xmu.edu.cn)

The breakthrough of environmental ge-Abstract nomics of marine microbes has revealed the existence of eubacterial rhodopsin in the sea, named proteorhodopsin (PR), which can take light to produce bio-energy for cell metabolism. Gene and protein sequence analysis and laser flash-induced photolysis experiments have validated the function of PR as light-driven proton-pump. During the pumping process, light energy is transformed into chemical gradient potential across plasma inner-membrane, the potential energy is then used to synthesize ATP. The finding of PR actually brings to light a novel pathway of sunlight utilization existing in heterotrophic eubacteria in contrast to the well-known chlorophyll-dependent photosynthesis in the sea. Since the group of PR-bearing bacteria is one of the numerically richest microorganisms on the Earth, accounting for 13% of the total in sea surface water, and with averaged cellular PR molecules of 2.5×10⁴, PRbearing bacteria are a key component not to be ignored in energy metabolism and carbon cycling in the sea. Based on the understanding of current literature and our own investigation on PR in the China seas which indicated a ubiquitous presence and high diversity of PR in all the marine environments, we propose a conceptual model of energy flow and carbon cycling driven by both pigment-dependent and -independent biological utilization of light in the ocean.

Keywords: proteorhodopsin, photosynthesis, marine microbial ecology, proton pump, model of carbon cycling.

The sea is the cradle of photosynthesis on the Earth. Bio-utilization of light in the sea is usually thought to be carried out by light-harvesting pigments, such as normal chlorophyll, divinyl chlorophyll and bacterial chlorophyll, which exist correspondingly in oxygenic

REVIEW

eukaryotic algae, oxygenic prokaryotic cyanobacteria and anoxygenic photosynthetic bacteria^[1]. However, recent studies of environmental genome library built on field samples from the Pacific have revealed, for the first time, that the carrier of a novel photosynthesis different from the common pigment-based ones proteorhodopsin (PR). PR can directly use light energy to drive proton transporting out of the plasma innermembrane, so that a proton gradient potential is built up. The ATP synthase embedded in membrane makes use of the potential to produce ATP for cell consummation^[2]; thus light energy is finally transformed into chemical energy. The photosynthetic function of PR is verified by studying their protein structure, spectroscopic characteristics and photochemical reactions^[3]. As PR-bearing bacteria account for up to 13% of total microorganisms in the euphotic zone^[4], this novel photosynthesis carried out by PR obviously plays a significant role in material and energy cycling in the marine ecosystem. The discovery of the new form of photosynthesis through which light energy is directly used makes it essential to renovate the structure of the marine foodweb and the flow and flux of the marine energy model (http://www.mbari.org/news/news releases/ 2002/feb07 delong.html). On the basis of analysis of recent PR-related literature and our filed investigations in the China Seas, this paper discusses the distribution, genetic diversity, and the role of PR in the sea, and proposes a light utilization based marine carbon cycling model.

1 Classification of rhodopsins and the discovery of proteorhodopsin

In nature, rhodopsin (retinylidene proteins) is found in organisms from Archaea to Eukarya^[5]. All known rhodopsins have very similar topological structures consisting of seven membrane-embedded α -helixes that form an internal pocket in which the chromophere retinal is bound. Some shifts in relative position of these helixes may occur in different rhodopsins. Sequence alignments split known rhodopsin into two distinct families, namely Type I and Type II. Type I (also called microbial type or archaeal type), is mainly found in lower trophics as archaea, while type II is well known as photosensitive receptor in animal eyes^[5].

Rhodopsin in microorganism was first found in archaea *Halobacterium salinarum* in the 1970s and recorded in some other archaea related to *H. salinarum* later on. Archaeal rhodopsins were the focus of

rhodopsin studies over thirty years since its discovery. Archaeal rhodopsins can be functionally divided into two classes ion-pump rhodopsin and sensory rhodopsin. The ion-pump rhodopsins can be further subclassified into proton-pump (bacteriorhodopsin, BR) and chloride pump (halorhodopsin, HR). Sensory rhodopsins (SR) can also be subclassified into SR I and SR II according to the differences in their gene sequences and spectroscopic characteristics. Both BR and HR's maximum absorption wavelength (λ_{max}) are at about 570 nm. Driven by light, BR transports H⁺ out of membrane and establishes an electronic gradient potential to be used by ATP synthase to synthesize ATP for cell metabolism, while HR can transport Cl⁻ into the inside of membrane under light so as to sustain the balance of pH in cell to acclimatize the cell to extreme environments. Of all rhodopsins, BR is to date the mostly studied protein and even one of the few membrane proteins with known crystal structure. Almost all of the functional residues in BR (full length of 248 amino acid residues) have been brought to light: the biding residues with retinal and Schiff base, protonation and deprotonation residues^[5,6]. SR I and II function as phototaxis to orange and green light respectively. SR I owns color-discriminating metabolism to ensue the cells be attracted to orange light and repelled from UV light, while SR II seems to have the repellent function only. In any case, its exertion of the particular function is closely related to photoisomerization of retinal and subsequent intermediates in the photochemical reaction^[6,7].

Recent studies have discovered rhodopsin-bearing bacteria other than halobiotic archaea. Since 1999, by genome sequencing of cultivated microorganisms, wide presence of archaeal rhodopsin homologies in both bacteria and eukarya domains was revealed^[8 10]. However, analysis of sequences and photochemical cycle have shown that these rhodopsins are of SR type and therefore are unable to use light energy $\frac{111}{2}$. Although proton-pumping rhodopsin beyond archaea is found in a fungi Leptosphaeria maculans most recently^[12], it was through the application of bacterial artificial chromosome library (BAC) of marine environmental samples that the proton-pump rhodopsin containing bacteria were brought to light^[2,13]. The gene recorded in BAC has been expressed in Escherichia coli (the gene-expressing E. coli cells appear red after the adding of additional retinal, indicating that the added retinal is bound to gene expressed product). Laser flash-induced photolysis experiments have shown that the time of photochemical cycle is as short as observed in BR (<20 ms)^[14], and the flux of proton from inside to outside of the membrane is detected. All the above evidence affirmed the function of proton-pump of the expressed product of the rhodopsin gene from BAC. The large fragment containing rhodopsin gene BAC31A08 bears a segment of 16S rDNA through which the large fragment clone is proved to be from proteobacteria SAR86. Thus the rhodopsin is named proteorhodopsin (PR)^[2]. In addition, the physical existence of PR has also been verified through laser flash-induced photolysis on the natural marine samples by Béjà and his collaborators^[15].

The conserved residues of type I rhodopsin in microorganisms, especially residues that form the "internal pocket", are distinctly different from those of the rhodopsins of higher organisms. So far, type I rhodopsin has been found in all the three life domains Archaea, Bacteria and Eukarya, possibly suggesting that the rhodopsin-induced photosynthesis emerged earlier than other photosynthesis on the Earth^[11].

2 Natural distribution and genetic diversity of PR in the sea

After the first record of PR in Monterey Bay, many other field investigations into environmental genome libraries and PCR techniques have shown that PR are rich in genetic diversity. At present, over 900 PR sequences have been registered in the GenBank, and some of them have been found to function as proton-pump. Although most of the obtained PR sequences have not yet been experimentally verified to be proton-pump, the same conserved amino acid residues as BAC31A08 at key positions in the sequences suggest that they are most likely to be the rhodopsins of light-driven proton-pump.

2.1 Natural distribution

So far, PR sequences have been detected in several marine sites, such as Monterey Bay of California, Hawaii Ocean-Time Station (HOT station), Red Sea, Mediterranean Sea, Sargasso Sea and Antarctic Peninsula^[4]. All the above areas except Monterey Bay are oligotrophic environments. However, PR in other marine environments has rarely been touched yet. In the present study, with published and home-designed primers ,we explored the presence and diversity of PR in the West Pacific marginal seas along the Chinese coasts

from 20°N 40°N, covering a variety of marine ecosystems from prawn aquatic culture pools, harbors, estuaries to open waters, and from sea surface to 500 m deep water (Fig. 1). PR positive results were observed in all the samples from the above areas. Theoretically, since PR-bearing bacteria can use light energy to supply their heterotrophic metabolism, and are more competitive than other bacteria under trophic stress, oligotrophic environments will therefore be more suitable for PR bacteria to thrive. However our data showed a ubiquitous distribution of PR in the marine environments, implying a high diversity of both PR proteins and PR-bearing bacteria in the sea. Particularly, we detected PR at a depth of 500 m in the South China Sea. This depth is much deeper than the previously reported 100 m, extending the depth boundary of PRbearing bacteria down to the aphotic zone. Although there was no direct reasoning for the recorded deep PR, this finding is supported by the illumination experiments on Pelagibacter ubique HTCC1062 (a-proteobacterium), one of the few lab cultivated SAR11 strains, which can normally grow and express PR proteins in the dark^[16]. Sequence alignment (Fig. 2) and phylogenetic analysis (Fig. 3) showed that our sequences were mainly from the groups of y-proteobacteria SAR86-I and α -proteobacteria.



Fig. 1. Locations of the PR investigation sites in the China Seas (\blacktriangle sampled depth at 0, 400 and 500 m; sampled of surface).

2.2 Genetic diversity

Although over 900 pieces of PR sequences have been acquired by environmental genomics and PCR analysis so far, we are still at the beginning of the understanding of the diversity of PR in the sea. Alignment of DNA sequences indicated that the similarities of PR from different geographical and ecological environments are divergent, and these PR sequences are allocated at different positions in the phylogenetic trees $\frac{[17]}{1}$. This suggests that the environmental control and natural selection significantly affect the distribution and evolution of $PR^{[18]}$. More studies on the diversity of PR would reveal their evolution history and diversion time in different life domains. Although the application of the first PR sequence revealed by environmental genome analysis in PR primer design is successful, further environmental genome sequencing analysis has demonstrated much divergent PR sequences that fall outside of the primary primers. e.g. in field samples from the Sargasso Sea, Venter et al.^[19] found 782 pieces of PR greatly different from the previously known ones. Therefore, it is essential to redesign primers in exploring PR diversity in different environments.

In the present study, we designed degenerate primers including specific ones for PR in α -proteobacteria. The alignment of protein sequences (Fig. 2) showed that the PR from China Seas also had seven conserved domains of α -helixes; among these helixes, Helix C and Helix G contained conserved motif of RXXD and DXXXK. It should be noticed that the corresponding residue in RXXD motif was not Asp85 but Gly85 in the two sequences China Seas (BHBH23-15 from and ECSXMS25-a) (residue position corresponding to BR). Since Asp85 in motif of RXXD is a must for a rhodopsin to be a functional proton-pump^[11,20], functions of these sequences at the position of 85 with Gly being not Asp need to be further confirmed, although BHBH23-15 and ECSXMS25-a group together with other marine PR in the phylogenetic tree (Fig. 3) which indicates that they are close relatives. On the other hand, as the two sequences were found in heavily contaminated areas, the questions as to whether they play a similar role in resistance or endurance with pollutant as other kind of bacterial photosynthesis (transformed light energy into ATP to supply the cell energy metabolism to live through some adversity), or whether the mutation at the site was induced by pollutant, and whether its function has been changed are still open.

| | | | | - - | Helix A | | - | Helix B | | • | F | He | olix C | | Heli. |
|--|--|--|---|---|--|--|---|---|--|--|--|---|----------------------------------|--|--|
| BAC31A08 | M KLLL I | LGSVIALPT | FAAGGGDLDA | SDYTGVSFWL | VT'AALLA | STVFFFVERD | RVSAKWKTSL | TVSGLVTGIA F | WHYMYMRGV | WIETG | DS | PTVF R YI D WL | LTVPLLICEF | YLILAAATNV | AGSLFI |
| SCSA10S13 | MGKVLL I | LGSVIALPT | FAAGGGDLDA | SDYTGVSFWL | VTAALLA S | STVFFFVERD | RDSAKWKTSL | TVSGLVTGIA F | WHYMYMRGV | WIETG | DS | PTVF R YI D WL | LTVPLLICES | YLILAAATNV | AGSLFI |
| SCSD6S5 | MGKKLV I | LGSVIALPT | FAAGGGDLDA | SDYTGVSFWL | VTTALLA S | STVFFFVERD | RVSAKWKTSL | TVSGLVTGIA F | WHYMYMRGV | WIETG | DS | PTVF R YI D WL | LSVPLLICES | YLILAAATNV | AGSLFI |
| ECSXCS14 | MGKILV I | LGSVIALPT | FAAGGGDLDA | SDYTGVSFWL | VTAALLA S | STVFFFVERD | RVSAKWKTSL | TVSGLVTGIA F | WHYMYMRGV | WIETG | DS | PTVF R YI D WL | LTVPLLICES | YLILAAATNV | AGSLFI |
| BHBHS23-15 | 5 | | GDLDA | GDPTGVSFWL | VTAALLA S | STVFFFVERD | RVSAKRKTSL | TVSGLVTGIA F | WHYMYMRGV | WIETG | DS | PTVF R YIGWL | LTVPLLICES | YLILAAATNV | AGSLFI |
| SCSA1500-1 | | | GDLDV | GDITGVSFWL | VTAASLA S | STVFFFVERD | RVSAKWKTSL | TVSGLVTGIA F | WHYMYMRGV | WIETG | DS | PTVF R YI D WL | LTVPLLICES | YLILAAATNV | AGSLFI |
| SCSA6S49 | | | GDLDV | GDPTGVSFWL | VTAALLA S | STVFSFIERD | RVAAKWKTSL | TVSGLVTGIA F | WHYLYMRGV | WVETG | ES | PTVF R YI D WL | LTVPLLICEF | YLILAAATNV | AGSLFI |
| HOT75 | MGKLLL I | LGSAIALPS | FAAAGGDLDI | SDTVGVSFWL | VTAGMLA | ATVFFFVERD | QVSAKWKTSL | TVSGLITGIA F | WHYLYMRGV | WIDTG | DT | PTVF R YI D WL | LTVPLQVVEF | YLILAACTSV | AASLFI |
| SCSA10S1-a | | | | | MVA | ATVFFFLERD | RVAPKWKTSL | TVAGLVTGIA A | WHYFYMRGV | WVATG | DS | PTVL R YI D WL | ITVPLQIVEF | YVILAAMTAV | ASCLF\ |
| ECSXMS25-a | | | | | MVA | ATVFFFLERD | RVAAKWKTSL | TVAGLVTGIA A | WHYFYMRGV | WVATG | DS | PTVL R YI G WL | ITVPLQIVEF | YVILAAMTAV | ASSLF\ |
| SAR11 | MKKLKLFALT A | VALMGVSGV | ANAETTLLAS | DDFVGISFWL | VSMALLA S | STAFFFIERA | SVPAGWRVSI | TVAGLVTGIA F | IHYMYMRDV | WVMTG | ES | PTVY R YI D WL | ITVPLLMLEF | YFVLAAVNKA | NSGIF\ |
| SRII | | | | -MALTTWFWV | GAVGMLA | GTVLPIRDCI | RHPSHRRYDL | VLAGITGLAA I | AYTTMGLGI | TATTVG | DRT | VYLARYIDWL | VTTPLIVL Y L | AMLARPGHRT | SAl |
| BR | | MLELLP | TAVEGVSQAQ | ITGRPEWIWL | ALGTALMGLG (| FLYFLVKGMG | VSDPDAKKFY | AITTLVPAIA F | TMYLSMLLG | YGLTMVPFGG | EQNP | IYWARYADWL | FTTPLLLLDL | ALLVDADQGT | ILAJ |
| SRI | | | | MDAVATAYL | GGAVALIVGV / | AFVWLLYRSL | DGSPHQSALA | PLAIIPVFAG I | SYVGMAYDI. | GTVIVN | GNQ | IVGL R YI D WL | VTTP ILVG Y V | GYAAGASRRS | IIG |
| HR | MTAVS T | TATTVLQAT | QSDVLQEIQS | NFLLNSSIWV | NIA-LAGVV | ILLFVAMGRD | LESPRAKLIW | VATMLVPLVS 1 | [SSYAGLASG | LTVGFLQMPP (| GHALAGQEVL | SPWGRYLTWT | FSTPMILLAL | GLLADTDIAS | LFT-A |
| | | | | | | | | | | | | | | | |
| | | → F | 1 | Helix E | → | | Helix | F | → | I | Helix G | → | | | |
| BAC31A08 | GSLVMLVFGY | MGEAGIMAA | W PAFIIGCLA | <i>Melix E</i> W VYMIYELW | AG E-GKSACNT | A SPAVQSAYN | <i>Helix</i> T MMYIIIFGW | <i>F</i> A IYPVGYFTGY | LMGDGGSAI | N -INLIYNLAI | <i>llelix G</i> D FVN K ILFGL | J IWNVAVKES | S NA | | 249 |
| BAC31A08 SCSA10S13 | GSLVMLVFGY GSLVMLVFGY | / MGEAGIMA/ | W PAFIIGCLA W PAFIIGCLA | <i>Helix E</i> AW VYMIYELWA AW VYMIYELWA | AG E-GKSACNTA | A SPAVQSAYN A SPAVRSAYN | <i>Helix</i> T MMYIIIFGW T MMYIIIFGW | • <u>F</u> A IYPVGYFTGY A IYPVGYFTGY | LMGDGGSAI | N -INLIYNLAI | Helix G FVNKILFGL FVNKILFGL | I IWNVAVKES | is na | | 249 250 |
| BAC31A08 SCSA10S13 SCSD6S5 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY | MGEAGIMAA MGEAGIMAA MGEAGIMAA | W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA | <i>Velix E</i> W VYMIYELW W VYMIYELW W VYMIYELW | AG E-GKSACNTA AG E-GKSACNTA AG E-GKFACNTA | A SPAVQSAYN A SPAVRSAYN A SPAVQSAHN | <i>Helix</i> T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW | - <u>F</u> A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY | IMGDGGSAI LVGDGGSAI LVGDGGSAI | N -INLIYNIAI N -INLIYNIAI N -INLIYNIAI | Helix G FVNKILFGL FVNKILFGL FVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY | / MGEAGIMAA / MGEAGIMAA / MGEAGIMAA / MGEAGIMAA | W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA | <i>Helix E</i> W VYMIYELW W VYMIYELW W VYMIYELW W VYMIYELW | AG E-GKSACNTA AG E-GKSACNTA AG E-GKFACNTA AG E-GKSACNTA | A SPAVQSAYN A SPAVRSAYN A SPAVQSAHN A SPAVQSAHN | <i>Helix</i> T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW | · <u>F</u> A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY | / LMGDGGSAI / LVGDGGSAI / LMGDGGSAI / LMGDGASAI | N -INLIYNLAI N -INLIYNLAI N -INLIYNLAI N -INLIYNLAI | Ilelix G FVNKILFGL FVNKILFGL FVNKILFGL FVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA S NA S NA | | 249 250 250 250 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY | / MGEAGIMAA / MGEAGIMAA / MGEAGIMAA / MGEAGIMAA / MGEAGIMAA | W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA | <i>Helix E</i> AW VYMIYELWA AW VYMIYELWA AW VYMIYELWA AW VYMIYELWA AW VYMIYELWA | AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKFACNT/ AG E-GKSACNT/ AG E-GKSACNT/ | A SPAVQSAYN A SPAVRSAYN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW | · <u>F</u> A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I | I.MGDGGSAI I.VGDGGSAI I.VGDGGSAI I.MGDGGSAI I.MGDGASAI | NLNLIYNLAI NLNLIYNLAI NLNLIYNLAI NLNLIYNLAI | Ilelix G D FVNKILFGL D FVNKILFGL D FVNKILFGL D FVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | 35 NA 35 NA 35 NA 35 NA | | 249 250 250 250 179 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 SCSA1500-1 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY | 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA | W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA | Welix E WW VYMIYELW WW VYMIYELW WW VYMIYELW WW VYMIYELW WW VYMIYELW WW VYMIYELW | AG E GKSACNTA AG E GKSACNTA AG E GKSACNTA AG E GKSACNTA AG E GKSACNTA TG E GKSACNTA | A SPAVQSAYN A SPAVRSAYN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIVFGW | · <u>F</u> A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I A | / LMGDGGSAI / LVGDGGSAI / LMGDGGSAI / LMGDGASAI | NLNLIYNLAI NLNLIYNLAI NLNLIYNLAI | Helix G FVNKILFGL FVNKILFGL FVNKILFGL FVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 250 179 178 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 SCSA1500-1 SCSA6S49 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY | 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA | W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL | <i>Helix E</i> W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W FYMIYELW, | AG E GKSACNT/ AG E GKSACNT/ AG E GKFACNT/ AG E GKSACNT/ AG E GKSACNT/ IG E GKSACNT/ AG E GKSACNT/ | A SPAVQSAYX A SPAVRSAYX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIFGW T MMYIIVFGW | F A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I A | 7 LMGDGGSAI 7 LVGDGGSAI 7 LMGDGGSAI 7 LMGDGASAI | N -INLIYNIAI N -INLIYNIAI N -INLIYNIAI | Ilelix G FVNKILFGL FVNKILFGL FVNKILFGL FVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 179 178 178 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 SCSA1500-1 SCSA6S49 H0T75 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY | MGEAGIMAA MGEAGIMAA MGEAGIMAA MGEAGIMAA MGEAGIMAA MGEAGIMAA MGEAGIMAA MGEAGIMAA MGEAGIMAA AGEAGIMAA | W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL L PAFIIGMAC | Welix E W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W FYMIYELW, W FYMIYELW, | AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKSACNT/ AG E-GKAAVST/ | A SPAVQSAYN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN A SPAVQSAHN A SPAVQSAN A SPAVQSAYN A SPAVNSAYN | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIFGW T MMYIIFGW T MMYIIFGW T MMYIIVFGW T MMYIIVFGW A MMMIVVGW | F A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I A A A IYPAGYAAGY | 7 LMGDGGSAI 7 LMGDGGSAI 7 LMGDGGSAI 7 LMGDGASAI 7 LMGGGGVY/ | N -INLIYNIA N -INLIYNIA N -INLIYNIA N -INLIYNIA N -INLIYNIA S NINLIYNIA | Ilelix G PVNKILFGL FVNKILFGL FVNKILFGL FVNKILFGL FVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 179 178 178 251 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 SCSA1500-1 SCSA6S49 H0T75 SCSA10S1-a | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLAGGF A STIMLLFCY | 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 AGEAGIAA 7 AGEAGIAA | W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL L PAFIIGMAC T LAFVIGMAC | Welix E W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W FYMIYELW, W FYMIYELW, W LYMIYELY, W LYMIYELY, | AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKAAVSTA | A SPAVQSAYX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAYX A SPAVQSAYX A SPAVXSAYX A XAAGPTAFX | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIFGW T MMYIIFGW T MMYIIVFGW A MMMIIVFGW A LRLIVTVGW | F A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I A A A IYPAGYAAGY A IYPAGYAAGY | I LMGDGGSAI I LVGDGGSAI I LMGDGGSAI I LMGDGGSAI I LMGDGASAI | N -INLIYNAU N -INLIYNAU N -INLIYNAU N -INLIYNAU N -INLIYNAU | Helix G FVNKILFGI. FVNKILFGI. FVNKILFGI. FVNKILFGI. | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 179 178 178 251 173 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 SCSA1500-1 SCSA6S49 H0T75 SCSA10S1-a ECSXMS25-a | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY A STIMLVFGY A STIMLVFGY | 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIMAA 7 MGEAGIAAN 7 MGETGAMN 7 MGETGAMN | W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL U PAFIIGCL T LAFVIGMAC T LAFVIGMAC | <i>Velix E</i> W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W FYMIYELW, W LYMIYELY W LYMIYELY W LYMIYEVF, | AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKSACNTA AG E-GKAAVSTA AR EPSKASAGS(| A SPAVQSAYX A SPAVQSAYX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAYX A SPAVQSAYX A SPAVXSAYX A XAAGPTAFX G XAAGQTAFX | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIVFGW T MMYIIVFGW A MMMIIVFGW A LRLIVTVGW A LRLIVTVGW | F A IYPVGYFTGY A IIPVGYFTGY A IIPVGYFTGY A IIPVGYAAGY A IIPVGYAAGY A IIPVGYAAGY A IIPVGYAAGY | / LMGDGGSAI / LVGDGGSAI / LMGDGGSAI / LMGDGGSAI / LMGDGGSAI / LMGGEGVY/ / LPG / LGGG | N -INLIYNAI N -INLIYNAI N -INLIYNAI N -INLIYNAI N -INLIYNAI | Helix G PVNKILFGL PVNKILFGL PVNKILFGL PVNKILFGL PVNKILFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 179 178 178 251 173 174 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBHS23-15 SCSA1500-1 SCSA6S49 H0T75 SCSA10S1-a ECSXMS25-a SAR11 | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLAGG A STIMLLFCY A STIMLVFGY GTLVMLIGGY | MGEAGIMAA MGETGAMNA MGETGAMNA MGETGAMNA LGEAGYINT | W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL W PAFIIGCL U PAFIIGCL T LAFVIGMAC T LAFVIGMAC T LGFVIGMAC | <i>Velix E</i> W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W FYMIYELW, W LYMIYELY W LYMIYELY W LYMIYELF, W LYMIYEVF, W LYMIYEVF, | AG E-GKSACNTA AG E-G | A SPAVQSAYX A SPAVQSAYX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAHX A SPAVQSAYX A SPAVQSAYX A XAAGPTAFX G XAAGQTAFX G XAAGQTAFX | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIVFGW T MMYIIVFGW A MMMIIVFGW A LRLIVTVGW A LRLIVTVGW A MRMIVTVGW | F A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I A I A I A I A I A I A I A I A I A I A I A IPIGYAAGY A IFPIGYAAGY A IYPIGYAAGY A IYPIGYAAGY | LMGDGGSAI LVGDGGSAI LVGDGGSAI LMGDGGSAI LMGDGASAI LMGGEGVY/ LPG | N -LNLIYNAR N -LNLIYNAR N -LNLIYNAR N -LNLIYNAR AS NLNLIYNAR | Helix G D FVNKILFGL D FVNKILFGL D FVNKILFGL D FVNKILFGL D FVNKILFGL D FLNKIAFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES | S NA | | 249 250 250 179 178 178 251 173 174 255 |
| BAC31A08 SCSA10S13 SCSD655 ECSXCS14 BHBHS23-15 SCSA1500-1 SCSA6549 HOT75 SCSA10S1-a ECSXMS25-a SAR11 SRII | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLGGF A STIMLFGY GSLVMLGGY DVFVIAAGIA | WGEAGIMAN WGEAGIMAN WGEAGIMAN WGEAGIMAN WGEAGIMAN WGEAGIMAN WGEAGIMAN GGEAGIMAN WGEGAIMAN WGEGAMNN MGETGAMNN LGEAGYINI AALTTGVQF | W PAFIIGCL ^J W PAFIIGCL ^J W PAFIIGCL ^J W PAFIIGCL ^J W PAFIIGCL ^J W PAFIIGCL ^J L PAFIIGMAG T LAFVIGMAG T LAFVIGMAG T LGFVIGMAG W LFFAVGAAG | Welix E W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW W FYMIYELW W LYMIYELY W LYMIYEVF, W LYMIYEVF, W FYMLYEVF, Y AALLYGLL | AG E-GKSACNTA AG E-G | A SPAVQSAYX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAYX A SPAVQSAYX A SPAVQSAYX G NAAGQTAFX G NAAGQTAFX G NAAGQTAFX G NKALVTAFG D DPRVRSLFV | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW A MMYIIVFGW A MMMIIVFGW A LRLIVTVGW A LRLIVTVGW A MRMIVTVGW T LRNITVVLW | F A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A I A I A I A IYPAGYAAGY A IYPIGYAAGY A IYPIGYAAGY A IYPIGYAAGY A IYPIGYAAGY A IYPIGYAAGY | Y LMGDGGSAU Y LVGDGGSAU Y LMGDGGSAU Y LMGDGASAU Y LMGGEGVY/ Y LPG | N -LNLIYNAU N -LNLIYNAU N -LNLIYNAU N -LNLIYNAU N -LNLIYNAU SLNVIYNAU M TTIVWYL-T | Helix G D FVNKILFGL D FVNKILFGL D FVNKILFGL D FVNKILFGL D FVNKILFGL D FLNKIAFGL D FLNKIAFGL D FLNKIAFGL | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWAAAMSQI F AVLGADAVS | IS NA | PA TAEPTPDG | 249 250 250 179 178 178 251 173 174 255 D 237 |
| BAC31A08 SCSA10513 SCSD655 ECSXCS14 BHBHS23-17 SCSA1050-1 SCSA6549 H0775 SCSA1051-r ECSXM525-r SAR11 SR11 BR | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLAGF A STIMLFGY GTLVMLIGG DVFVIAAGIA GIMIGTGLVG | Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGETGAMON Y MGETGAMON Y LGEAGYINT A AALTGVQF G ALTKVYSYF | W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA W PAFIIGCLA U PAFIIGCLA T LAFVIGMAC T LAFVIGMAC T LAFVIGMAC G LFFAVGAAC F VWWAISTAN | Welix E W VYMIYEL#, W VYMIYEL#, W VYMIYEL#, W VYMIYEL#, W VYMIYEL#, W VYMIYEL#, W FYMIYEL#, W FYMIYEL#, W LYMIYEL#, W LYMIYEL#, W FYLLYEVF, W FYLLYEVF, Y AALLYGLJ M LYLLYVFF | AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG E-GKSACNT, AG EASKASAGSG GG EAGKNAKSG GGTTLPRALG FG FTSKAESJ | A SPAVQSAYX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAYX A SAAQTAYX G XAAQTAYX G XAAQTAYX G XAAQTAYX G XAAQTAYX G XAAQTAYX G XPAVSAYX A XAAQTAYX G XPAVQSAYX A XAAQTAYX G XPAVQSAYX A XAAQTAYX G XPAVQSAYX A SPAVQSAYX A SPAVX A SPAVQSAYX A SPAVX A SPAVQSAYX A SPAVX A SP | Helix T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIIFGW T MMYIIVFGW A MMMIIVGGW A LRLIVTVGW A LRLIVTVGW A KMMIVYGW T LRNITVVLW V LRNITVVLW | F A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPVGYFTGY A IYPAGYTGY A IYPAGYTGY A IYPAGYAGY A IYPAGYAAGY A IYPIGYAAGY A IYPIGYAAGY A IYPIGYAAGY A IYPIGYAAGY T IYPWULLSY S AYPVWLLGSS | Y LMGDGGSAI Y LVGDGGSAI Y LVGDGGSAI Y LMGDGSAI Y LMGGEGVY/ Y LPG Y LGGG Y MTGGMDASS AGIGILQTI S EGAGIVPL2 | N -LNLIYNAI N -LNLIYNAI N -LNLIYNAI N -LNLIYNAI S -LNLIYNAI S -LNUIYNAI S -LNVIYNAI M TTIWWYL-I S - FLLFWU-I | Helix G) FVNKH.FGL) FVNKH.FGL) FVNKH.FGL) FVNKH.FGL) FVNKH.FGL) FLNKHAFGL) FLNKHAFGL) FLNKHAFGL) FLNKHAFGL) FLNKHAFGL | I IWNYAVKES I IWNYAVKES I IWNYAVKES I IWNYAVKES I IWNYAVKES I IWNYAVKES I IWNAAMSQF F AVLGADAYS I LLRSRAIFC | IS NA | PA TAEPTPDG 3D GAAATSD | 249 250 250 179 178 178 251 173 174 255 0 237 262 |
| BAC31A08 SCSA10S13 SCSD6S5 ECSXCS14 BHBIS23-11 SCSA6S49 H0175 SCSA10S1-* ECSXB25-* SAR11 BR SRI SRI | GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GSLVMLVFGY GTLVML1GGY GTLVML1GGY DVFV1AAGIA GIMIGTGLVG | Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGEAGIMAN Y MGETGANCY Y MGETGANCY Y MGETGANCY Y AALTTGVQF G ALTKVYSYF G AVVTDGTLF | // W PAFIIGCL/ W PAFIIGCL/ W PAFIIGCL/ W PAFIIGCL/ W PAFIIGCL/ W PAFIIGCL/ W PAFIIGCL/ T LAFVIGMAC T LAFVIGMAC T LAFVIGMAC T LAFVIGMAC W LFFAVGAAC W ALFAVSSIH W ALFGVSSIH | Welix E W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW, W VYMIYELW W FYMIYELW, W LYMIYELY W LYMIYELY W LYMIYELY W LYMIYEVF, W LYMIYEVF, Y AALLYGLL M LYMYYEF, H LSLFAYLY | AG E-GKSACNTI AG E-G | A SPAVQSAYX A SPAVQSAYX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAIX A SPAVQSAYX A SPAVQSAYX A SAAGPTAPX A SAAGPTAPX A SAAGPTAPX A SAAGPTAPX A SAAGPTAPX A SAAGPTAPX A SAAGYAPX A SAAGYAPX A SPAVSAYX A SPAVSAYX A SPAVSAYX A SPAVQSAYX A SPAVX A S | Helix T MUYIIIFGW T MUYIIIFGW T MUYIIFGW T MUYIIFGW T MUYIIFGW T MUYIIVFGW A MUYIVFGW A LRLIVTVGW A LRLIVTVGW A MRMIVTGW T LRXIVVLW V LRXVVVLW L LKXHIGLLW | F A IYPVGYFTGI A IYPVGYFTGI A IYPVGYFTGI A IYPVGYFTGI A IYPVGYFTGI A IYPVGYFTGI A IYPIGYAGI A IYPIGYAGI A IYPIGYAGI A IYPIGYAGI A IYPIGYAGI I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII | Y LMGDGGSAH Y LVGDGGSAH Y LVGDGGSAH Y LMGDGSAH Y LMGDGSAH Y LMGGEGVY/ Y LPG Y LDGG | N -LNLIYNAI N -LNLIYNAI N -LNLIYNAI N -LNLIYNAI S -LNLIYNAI | Helix G) FVNKILFGL) FVNKILFGL) FVNKILFGL) FVNKILFGL) FVNKILFGL) FLNKIAFGL) FLNKIAFGL) FLNKIAFGL) FLNKIAFGL) FLNKIAFGL) VLAKVPVVY | I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES I IWNVAVKES F AVLGADAVS I LLRSRAIFC F FYARRRYFA | IS NA | PA TAEPTPDG 5D GAAATSD 2A TVEATAAD | 249 250 250 179 178 251 173 174 255 D 237 - 262 - 239 |

Fig. 2. Alignment of primary sequences from the China Seas and other marine areas. *Halobacterium salinarum* BR; *H. salinarum* SRI; *H. salinarum* SRI; *H. salinarum* SRI; proteorhodopsin: BAC31A08 from Monterey Bay, blue-light absorbing; HOT75 from Pacific HOT Station at the depth of 75 m, green-light absorbing (refer to sec. 2); SAR11from *P. ubique* HTCC1062; the rest from the present study. Conserved residues are marked in bold, different conserved residues are marked with grey background. The alignment was done on the software of ClUSTALX (1.81) and checked manually.

In addition, the discrepancy of L (Leu93) and Q (Gln93) between BAC31A08 and HOT75, leads to their apparent different maximum light absorption wavelength; after all the ~22% difference makes them grouped into different evolutionary branches^[15].

2.3 Biodiversity of PR-bearing bacteria

The diversity of rhodopsin-bearing bacteria is obviously richer than initially thought. In addition to the known photo-sensory rhodopsin in cyanobacteria^[21] and PR in proteobacteria^[2,16], there are rhodopsin homologies found in Cytophaga-Flavobacterium-Bacteriodes (CFB), one richest group of marine bacteria in nature^[19]. However the CFB rhodopsin sequences have not been identified in terms of their functions and it is unknown whether or not they can be expressed in nature. In contrast, there is plenty of evidence for PR of SAR86 to support its proton-pump function^[2]. Further studies by Sabehi *et al.*^[22] detailed the bio-distribution of PR in subgroups of SAR86-I, II, III a and IIIb and revealed that SAR86's PR did not evolve in a parallel pattern with their 16S rRNA gene. The first reported PR of BAC31A08 from SAR86-II subgroup is the mostly studied and known as PR. Analyses of PR-bearing genomic fragment (HOT2C01) from the North Pacific subtropical gyre revealed a segment of 16S rRNA which is confirmed to be from the most abundant marine bacterial group α -proteobacteria. Unfortunately the carrier bacterium is not yet further classified due to the fact that the obtained 16S rRNA segment is not long enough. Likewise many divergent rhodopsin homologies observed in the Sargasso Sea BAC are probably originating from variant microorganisms, suggesting that there exist more divergent PR bearing bacteria in the sea. Our results showed similar evidence on the ubiquitous distribution of PR bearing bacteria in diverse marine environments, but the dominant PR carrier were SAR11 bacteria as seen from 16S rDNA analysis rather than the mostly reported SAR86 bacteria. Both the groups of SAR86 and SAR11 are very hard to cultivate in lab. To date, only P. ubique HTCC1062 was successfully grown in lab^[23], which is thought to be the smallest cultured microbe with a cell volume of 0.01 μ m³. By sequencing the whole genome, it is believed



Fig. 3. Comparative phylogenetic analysis of PR sequences from the China Seas and other marine areas (Neighbor-joining phylogenetic tree). The alignment was done on CLUSTALX (1.81). Software MEGA3 was used to construct the neighbor-joining tree (Kumar *et al.* 2004). The scale bar represents an estimated number of amino acid replacements per position. Percentages on the branch refer to bootstrap values of 100 calculated trees. *Halobacterium salinarum* BR; *H. salinarum* SR I; *H. salinarum* SR II; proteorhodopsin: BAC31A08 are from Monterey Bay, blue-light absorbing; HOT75 from Pacific HOT Station at a depth of 75 m, green-light absorbing (refer to sec. 3); SAR11from *P. ubique HTCC1062*; All others are from the China Seas.

that *P. ubique* HTCC1062 also has the most contracted genome structure. It is also shown by proteomics analysis to have the similarity of 49.4% to BAC31A08^[16,24]. Comparative alignment of protein sequences showed that the protein from SAR11 bacterium had the sequence characteristics of proton-pump PR.

3 Spectroscopic characteristics of PR

The sequences of PR from surface waters are much different from those in deeper waters, although they all expressed in *E. coli* and verified to be able to bind added retinals to form proton-transporting active proteins. These differences in sequence are related to the differences in PR protein structure which is adapted to the different wavelengths of light at different water depths^[15]. The spectroscopic characteristics of PR split the known active proteins of PR into three groups, namely green-light PR (GPR), blue-light PR (BPR) and blue-green-light PR (B-G-PR). The shift of absorption maxima wavelength of PR mainly result from the mu-

tations of residues at the sites of 65, 70 and 105 which bring on red or blue shifts.

GPR is the first reported PR and the known GPRs accounts for a majority of known PRs. GPR is found in the illumination surface waters of Pacific Oceans, Mediterranean and Red Seas, etc. Spectroscopic studies on active GPR by isolation and purification have showed that it has maximum absorption on green-light with the absorption peak (λ_{max}) of 525 nm. In contrast to GPR, BPR and B-G-PR have λ_{max} at 490 and 515 nm respectively $\frac{[25 \ 27]}{2}$. The residues at the sites of 65, 70 and 105 of GPR BAC31A08 found in surface waters of Monterey Bay are Ser, Gly and Leu, respectively, while corresponding residues of BPR from 75 m at the HOT Station are Ser, Gly and Gln, respectively. Obviously the spectroscopic difference between GPR and BPR is the reflections of the different residues (Leu105 and Gln105, respectively). Since B-G-PR contains Gln instead of Leu at position 105, its spectrum is expected to shift to blue wavelengths, i.e. to 490 500 nm. However, the residues of B-G-PR at the positions of 65 and

70 are respectively Pro and Asp not corresponding to Ser and Gly in GPR and BPR. Therefore, it appears that the algebraic sum of red shifts caused by the two additional mutations at positions 65 and 70 and the blue shift caused by the change in position 105 causes the intermediate maximum of 515 nm^[27]. The effects of these specific residues on absorption spectroscopy were supported by the point mutation experiments.

Despite of the differences in spectroscopy, both GPR and BPR function as proton-pump as revealed by laser flash-induced photolysis experiments. However, whether B-G-PR is rhodopsin of proton-pump is still unclear. Similar to the photochemical reaction of BR, laser flash induces the photochemical reaction, which is a repeat cycle. The measurement of time-resolution absorption spectroscopy of intermediate in the cycle may characterize the intermediates and tell us what the intermediate is and when it appears. One photochemical cycle of GPR and BPR only takes 20 ms or less. The short-time cycle is the typical profile of the protonpump rhodopsin in comparison with the 300 ms half time of the cycle of sensory rhodopsin^[7]. Although the expressed B-G-PR in E. coli can transport proton out of cell, the data of laser flash-induced measurements have shown that its time of photochemical cycle (halftime>600 ms) is the longest among the known PR. Previous work indicated that SRII could transport proton in the absence of the co-transmitting signal molecule^[28,29]. It is no doubt that there is less signal molecule assorted to B-G-PR, so proton-pump function of B-G-PR with so long a cycle time is questioned. Nevertheless we cannot disapprove the proton-transporting function of B-G-PR under the explanation that the long cycle time is an adaptation to the light intensity and spectra in their natural niches^[11]. In the present study,</sup> the PR we detected at the depth of 500 m has the same essential key residues as BAC31A08, and presumably has a λ_{max} of 525 nm. Although it is hard to interpret the observations at this stage, the following possibilities might be useful clews for further exploration: The PR functions as a light sensor, probably related to the light-repellent behavior of the organisms, and there should be some previously unknown residues in the PR sequences controlling tightly the spectroscopic shifts; As an assistant way to supply energy, PR may remain silent and do not express protein under certain conditions just like the rhodopsin in the animal brain; Marine PR may have unknown function other than light-sensor or ion-transporter. At any rates, to answer these questions, further studies and successful isolation of cultured strain turns to be necessary.

4 Ecological significance of PR

As indicated by recent literature, PR is involved in α -proteobacteria, γ -proteobacteria and CFB which are the most abundant and widely distributed bacteria in marine environments. According to BAC analysis, PR-bearing bacteria may account for up to 13% of the total microorganisms in surface water^[4]. Having in mind the wide distribution, high diversity and great abundance of PR bacteria, one could estimate a rough photosynthesis flux through the cellular PR content of 2.4×10^4 molecule as with the case of SAR86^[15] and photoreceptor quantum conversion efficiency^[30,31], which would be a great number not to be ignored in any model of energy flow or carbon cycling. Giovannoni et al.^[16] have even proposed a higher estimation of 25000 PR molecules per cell, and indicated that PR molecules occupy 20% of inner superficial area of a cell, adding further accounts to the role of PR in marine photosynthesis. Such an estimation would become more accurate only if more strains are isolated and cultured in the lab and further specialized techniques become available to the field. In any case, light energy utilized through the PR path would play a significant role in the marine ecosystem.

Covering about 71% of the surface of the Earth, the ocean is the biggest carbon pool of this planet. Carbon fixation by photosynthesis is the key link of ocean carbon cycling. Previous understanding of photosynthesis in the sea is mainly established on the basis of photosynthetic pigments. The discovery of PR has brought to light a novel pathway of marine photosynthesis which is independent of pigments. This is a breakthrough in the conventional concept of biological utilization of light energy in the sea and is challenging the assessment of primary production based on photosynthetic pigments either through *in situ* measurements or remote sensing observation.

We hereby propose a frame model of bio-utilization of light and carbon cycling in the ocean based on the understanding of literature knowledge and our own data (Fig. 4). In this model, light energy utilization pathways are sorted into three categories: Path I is chlorophyll dependent oxygenic photosynthesis with photosystem I and photosystem II, through which light is captured and transformed into chemical energy and carbon dioxide is fixed to organic carbon. This photosynthesis is carried



Fig. 4. A model of light bio-utilization and carbon cycling in the sea. Chl, Chlorophyll; BChl, bacteriochlorophyll; PR, proteorhodopsin; PS I and II, photosystem I and II; PP, proton pump.

out by marine algae/phytoplankton (with normal chlorophyll) and cyanobacteria (normal chlorophyll or divinyl chlorophyll). Path II is bacteriochlorophyll-dependent anoxygenic photosynthesis, and the owners of this pathway are aerobic and anaerobic anoxygenic photosynthetic bacteria, which can transform light into chemical energy but can not fix carbon due to lack of photosystem II. Path III is PR-depending light utilization. Lacking photosystem I and II, the carrier, PR-bearing bacteria directly trap light by PR and use proton-pump to produce ATP. These three pathways constitute the major mechanisms of light bio-utilization in the sea and drive the energy flow and carbon cycling in different ways: Pathway I directly leads to the fixation of carbon dioxide, whereas path II and III produce ATP so as to reduce the consummation of organic carbon contributing indirectly to carbon fixation by the ocean (Fig. 4).

5 Prospects

So far, only about 1% of total marine microorganisms are cultured. The understanding of the role of bacteria in marine ecosystem is still very limited. Application of molecular techniques in the studies on marine microorganism has broken the limitation of culture-dependent methods. It is no doubt that molecular techniques are a powerful tool in exploration of the role of functional bacteria. Especially, the recent developed environmental genomic techniques that have great potentials in expedition of unknown microbes as demonstrated by the environmental genome shotgun sequencing of the Sargasso Sea^[19] is honored as one of the breakthroughs of the year 2004 of Science magazine^[32]. Meanwhile other techniques should also be developed

www.scichina.com www.springerlink.com

so as to compensate for BAC in exact information of specific function genes, in-situ physiological, ecological functions, etc. Our preliminary results indicated that in-situ PCR can mend partially the shortcoming of BAC. Meanwhile, quantification of the PR-bearing bacteria is essential in understanding the significance of the role of PR in an ecosystem. As in the case of another functional group of aerobic anoxygenic phototrophic bacteria (AAPB)^[33], accurate enumeration of AAPB had been a limiting factor in correct understanding of their distribution pattern in marine environments until the technique of time-series-observation-based Infrared epifluorescence microscopy (TIREM) was established $\frac{[34]}{}$. Fig. 4 shows the theoretical frame of the role of PR in carbon cycling in the marine ecosystem, while a better understanding of the degree of the significance of such a role is awaiting accurate PR quantification techniques.

On the other hand, culture of PR-bearing bacteria in the lab is in urgent need as well. Only when enough typical strains of PR-bearing bacteria become available, could one take a detailed look at their physiological and ecological aspects including photochemical reaction, in-site function as proton-pump or photo-sensor; photosynthetic quantum yield efficiency; photosynthesis switch on or off, etc. In addition, attentions should also be paid to future applications of PR proteins, not only proton-pump type but also SR ones. e.g., the signal transmitting mechanism of SR is quite promising in information technology, such as optical information storage, processing, light-controlling switch, lightelectronic detection, light modulator, light filtration, and light transistor^[35].

In a word, studies on PR has brought us new under-

standings of energy flow and materials cycling in the sea. Future in-depth study would not only promote the development of the discipline, but also contribute to the exploitation and utilization of marine microbial resources.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 40232021, 40576063 & 40521003) and the Ministry of Science and Technology Projects (Grant Nos. G2000078500, 2001CB409700, 2003DF000040 & 2005AA635240).

References

- 1 Fenchel T. Marine bugs and carbon flux. Science, 2001, 292: 2444 2445[DOI]
- 2 Béjà O, Koonin E V, Aravind L, et al. Bacterial rhodopsin: evidence for a new type of phototrophy. Science, 2000, 289: 1902 1906[DOI]
- 3 Friedrich T, Geibel S, Kalmbach R, et al. Proteorhodopsin is a light-driven proton pump with variable vectoriality. J Mol Biol, 2002, 321: 821 838[DOI]
- 4 Sabehi G, Loy A, Jung K-H, et al. New insights into metabolic properties of marine bacteria encoding proteorhodopsins. PLoS Boil, 2005, 3(8): e273
- 5 Spudich J L, Yang C-S, Jung K-H, et al. Retinylidene proteins: Structures and functions from archaea to humans. Annu Rev Cell Dev Biol, 2000, 16: 365 392[DOI]
- 6 Haupts U, Tittor J, Oesterhelt D. Closing in on bacteriorhodopsin: Progress in understanding the molecule. Annu Rev Biophys Biomol Struct, 1999, 28: 367 399[DOI]
- 7 Bogomolni R A, Stoeckenius W, Szundi I, et al. Removal of transducer HtrI allows electrogenic proton translocation by sensory rhodopsin I. Proc Natl Acad Sci USA, 1994, 91: 10188 10192
- 8 Bieszke J A, Braun E L, Bean L E, et al. The nop-1 gene of Neurospora crassa encodes a seven transmembrane helix retinal-binding protein homologous to archaeal rhodopsins. Proc Natl Acad Sci USA, 1999, 96: 8034 8039[DOI]
- 9 Sineshchekov O A, Jung K H, Spudich J L. Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. Proc Natl Acad Sci USA, 2002, 99: 8689 8694
- 10 Jung K H, Trivedi V D, Spudich J L. Demonstration of a sensory rhodopsin in eubacteria. Mol Microbiol, 2003, 47: 1513 1522[DOI]
- 11 Spudich J L, Jung K-H. Microbial rhodopsins: Phylogenetic and functional diversity. In: Briggs W R, Spudich J L, eds. Handbook of Photosensory Receptors. Weinheim: Wiley-VCH, 2005. 1 23
- 12 Waschuk S A, Bezerra A G Jr, Shi L, et al. Leptosphaeria rhodopsin: Bacteriorhodopsin-like proton pump from a eukaryote. Proc Natl Acad Sci USA, 2005, 102: 6879 6883[DOI]
- 13 Béjà O, Koonin E V, Aravind L, et al. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ Microbiol, 2000, 2: 516 529[DOI]
- Váró G, Brown L S, Lakatos M, et al. Characterization of the photochemical reaction cycle of proteorhodopsin. Biophysi, 2003, 84: 1202 1207
- 15 Béjà O, Aravind L, Eugene V, et al. Proteorhodopsin phototrophy in the ocean. Nature, 2001, 411: 786 789[DOI]
- 16 Giovannoni S J, Bibbs L, Cho J C, et al. Proteorhodopsin in the ubiquitous marine bacterium SAR11. Nature, 2005, 438: 82 85[DOI]

- 17 Sabehi G, Massana R, Bielawski J P, et al. Novel proteorhodopsin variants from the Mediterranean and Red Seas. Environ Microbiol, 2003, 5: 812 849
- 18 Bielawski J P, Dunn K A, Sabehi G. Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. Proc Natl Acad Sci USA, 2004, 101: 14824 14829[DOI]
- Venter J C, Remington K, Heidelberg J F, et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science, 2004, 306: 66 74[DOI]
- 20 Kyndt J, Meyer T E, Cusanovich M A. Photoactive yellow protein bacteriophytochrome, and sensory rhodopsin in purple phototrophic bacteria. Photochem Photobiol Sci, 2004, 3: 519 530[DOI]
- 21 Torre J R, Christianson L M, Béjà O, et al. Proteorhodopsin genes are distributed in divergent marine bacterial taxa. Proc Natl Acad Sci USA, 2003, 100: 12830 12835[DOI]
- 22 Sabehi G, Béjà O, Suzuki M T, et al. Different SAR86 subgroups harbour divergent proteorhodopsins. Environ Microbiol, 2004, 6: 903 910[DOI]
- Rappé M S, Connon S A, Vergin K L, et al. Cultivation of the ubiquitous SAR11 marine bacterioplanton clade. Nature ,2002, 418:
 630 633[DOI]
- 24 Giovannoni S J, Tripp H J, Givan S, et al. Genome streamlining in a cosmopolitan oceanic bacterium. Science, 2005, 309(5738): 1242 1245[DOI]
- 25 Man D, Wang W W, Sabehi G, et al. Diversification and spectral tuning in marine proteorhodopsins. EMBO J, 2003, 22: 1725 1731[DOI]
- 26 Wang W W, Sineshchekov O A, Spudich E N, et al. Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. J Biol Chem, 2003, 278: 33985 33991[DOI]
- 27 Man-Aharonovich D, Sabehi G, Sineshchekov O A, et al. Characterization of RS29, a blue-green proteorhodopsin variant from the Red sea. Photochem Photobiol Sci, 2004, 3: 459 462[DOI]
- 28 Hoff W D, Jung K H, Spudich J L. Molecular mechanism of photosignaling by archaeal sensory rhodopsins. Annu Rev Biophys Biomol Struct, 1997, 26: 223 258[DOI]
- 29 Haupts U, Haupts C, Oesterhelt D. The photoreceptor sensory rhodopsin I as a two-photon-driven proton pump. Proc Natl Acad Sci USA, 1995, 92: 3834 3838
- 30 Michel H, Oesterhelt D. Electrochemical proton gradient across the cell membrane of Halobacterium halobium: Effect of N,N9-dicyclohexylcarbodiimide, relation to intracellular adenosine triphosphate, adenosine diphosphate, and phosphate concentration, and influenceof the potassium gradient. Biochemistry, 1980, 19: 4607 4614[DOI]
- 31 Tsutomu M, Koichi K. Quantum conversion and image detection by a bacteriorhodopsin-based artificial photoreceptor. Science, 1992, 255: 342 344
- 32 Kennedy D. Breakthrough of the year. Science, 2004, 306: 2001[DOI]
- 33 Jiao N Z, Sieracki M E, Zhang Y, et al. Aerobic anoxygenic phototrophic bacteria and their roles in marine ecosystems. Chin Sci Bull, 48(11): 1064 1068
- 34 Jiao N Z, Zhang Y, Chen Y. Time series observation based Infrared epifluorescence microscopic approach for accurate enumeration of bacteriochlorophyll containing microbes in Marine Environments. J Microbiol Methods, 2006, 65(3): 428 439
- 35 Hampp N, Oesterhelt D. Bacteriorhodopsin and its potential in technical applications. In: Niemeyer C, Mirkin C, eds. Nano-biotechnology. Weinheim: Wiley-VCH-Verlag, 2004. 146 167

(Received November 8, 2005; accepted March 1, 2006)