Photochemistry and Photobiology, 2006, 82: 1426–1435

Symposium-in-Print: Photoreceptors

Evolutionary Patterns of Retinal-Binding Pockets of Type I Rhodopsins and Their Functions†

Larisa Adamian, Zheng Ouyang, Yan Yuan Tseng and Jie Liang*

University of Illinois at Chicago, Bioengineering/Bioinformatics, Chicago, IL

Received 16 August 2006; accepted 20 August 2006; published online 21 August 2006 DOI: 10.1562/2006-02-14-RA-802

ABSTRACT

Genome sequencing projects resulted in the identification of a large number of new sequence homologs of archaeal rhodopsins in marine bacteria, fungi, and unicellular algae. It is an important task to unambiguously predict the functions of these new rhodopsins, as it is difficult to perform individual experiments on every newly discovered sequence. The transmembrane segments of rhodopsins have similar three-dimensional structures where the seven transmembrane helices form a tightly packed scaffold to accommodate a covalently bound retinal. We use geometric computations to accurately define the retinal-binding pockets in high-resolution structures of rhodopsins and to extract residues forming the wall of the retinal-binding pocket. We then obtain a tree defining the functional relationship of rhodopsins based on the short sequences of residues forming the wall of the retinal-binding pocket concatenated from the primary sequence, and show that these sequence fragments are often sufficient to discriminate phototactic *vs* transporting function of the bacterial and unicellular algal rhodopsins. We further study the evolutionary history of retinal-binding pockets by estimating the pocket residue substitution rates using a Bayesian Monte Carlo method. Our findings indicate that every functional class of rhodopsins has a specific allowed set of fast-rate amino acid substitutions in the retinal-binding pocket that may contribute to spectral tuning or photocycle modulation. The substitution rates of the amino acid residues in a putative retinal-binding pocket of marine proteorhodopsins together with the clustering of pocket sequences indicate that green-absorbing and blue-absorbing proteorhodopsins have similar function. Our results demonstrate that the evolutionary patterns of the retinal-binding pockets reflect the functional specificity of the rhodopsins. The approach we describe in this paper may be useful for large-scale functional prediction of rhodopsins.

INTRODUCTION

Rhodopsins are membrane receptors composed of seven transmembrane (TM) helices. There are two types of rhodopsins presently known: type I rhodopsins that were discovered in halophile archaea (including bacteriorhodopsin [BR], halorhodopsin [HR] and sensory rhodopsins I and II [SRI, SRII]) and type II rhodopsins, which are the animal visual photoreceptors (1). They are activated by light exposure, which results in the isomerization of the retinal that is covalently bound through Schiff base (2,3) to a lysine residue located on the seventh TM helix (G). The respective rhodopsin then changes its conformation and either transports the ions or transmits a signal to a membrane-bound or a cytosolic transducer. The two types of rhodopsins show no significant sequence similarity and may have different origins (4,5). We restrict this study to the type I rhodopsins.

Three-dimensional structures of several type I rhodopsins (BR, HR, SRII and anabaena sensory rhodopsin [ASR]) are currently available (see Table 1). Despite the low sequence identity of type I rhodopsins, there is a remarkable similarity in the architecture of the TM region, especially between helices C-G, which form a retinal-binding pocket. Structural comparison of TM helices C-G of archaeal type I rhodopsins shows that these helices have the highest structural similarity with the pairwise root mean square deviation (RMSD) of the main chain atoms of less than 0.9 Å (6).

In this study, we report the results of analysis of the rhodopsin retinal-binding pockets. We build a maximum likelihood relationship tree of short sequences obtained after concatenating all residues on the wall of the retinal-binding pocket and show that these pocket sequences are often sufficient to discriminate a phototactic or transporting function of the rhodopsin. We also estimate substitution rates of amino acid residues forming a retinal-binding pocket in BR, SRI, SRII and proteorhodopsins (PR) using a Bayesian Monte Carlo approach (7). Our findings indicate that every functional class of rhodopsins has its specific set of fast-rate amino acid substitutions allowed in the retinal-binding pocket. These fast-rate substitutions do not affect the function of the given rhodopsin, but may be important for the rhodopsin spectral tuning or for the photocycle modulation.

MATERIALS AND METHODS

Sequences and protein structures. We searched the protein sequence databases using BLAST and PSI BLAST (8) from the National Center for Biotechnology Information (NCBI) Web site (http://www.ncbi.nlm.nih.gov/). The outputs of BLAST searches were used to build multiple

*Corresponding author email: jliang@uic.edu (Jie Liang)

†This paper is dedicated to Professor Thomas Ebrey on the occasion of his retirement from the University of Washington.

Table 1. Retinal-binding pocket areas (Å^2) and volumes (Å^3) as obtained from CASTp database (http://sts.bioengr.uic.edu/castp) and retinal-binding pocket sequences from high-resolution structures of type I rhodopsins

PDB	Area	Volume	Resolution	POC*	Remarks	Pocket Sequence
Bacteriorhodopsin						
1M0L	430	506	1.5	21	Ground state	YWTLMIGGWWSTMWYPWFAK
1C3W	397	467	1.6	21	Ground state	YWTLMIGWSTMWYPWFAK
1KGB	442	526	1.7	24	Ground state	YWTLMIGGWWSTMWYPWFAK
1F50	384	450	1.7	20	E204Q mutant	YWTLMIGWSTMWYPWFAK
1M0K	436	516	1.4	23	K intermediate	YWTLMIGGWWSTMWYPWFAK
1M0M	440	523	1.4	21	M1 intermediate	YWTLMIGGWSTMWYPWFAK
1P8H	441	533	1.5	24	M1 intermediate	YWTLMIGGWSTMWYPWFAK
1O0A	430	501	1.6	18	I intermediate	YWTLMIGGWSTMWYPWFAK
1P8U	474	528	1.6	13	n′ intermediate	YWTLMIGGWSTMWYPWFAK
1QKO	430	513	2.1	25	Early intermediate	YWTLMIGGWWSTMWYPWFAK
Archaerhodopsin-1						
1UAZ	454	530	3.4	53		YWTTLMIGWSTMWYPWFDAK
Archaerhodopsin-2						
1VGO	438	530	2.5	63		YWTTLMIGWSTFWYPWFDAK
Halorhodopsin						
1E12	496	564	1.8	29		YWTIMCGAYSCFWYPWYAK
Sensory Rhodopsin II						
1H2S	609	630	1.9	35	With transducer	RYWTIVMAGFGAFLWYPIWPDIDTK
1H68	582	640	2.1	21	Ground state	RYWITIVMAGFGAFWYPWPDIDTK
1GU8	582	640	2.3	21	Ground state	RYWTIVMAGFGAFLWAYPFWPDIDTK
1GUE	619	702	2.3	23	K-state	RYWITIVMAGFGAFWYPIWPDIDTK
Anabaena Sensory Rhodopsin						
1XIO	476	549	2.0	19	*Cis*-retinal	YWTLQVIGAWYGVFWYPWFSK

*Pocket number as obtained from CASTp database.

sequence alignments (MSA) with ClustalW (9), which were manually adjusted using a multiple sequence editor Pfaat (10). Sequences used in this study are listed in Table 2.

PR sequences often come from the environmental samples or from uncultured bacteria, without any names assigned. We use bacterial strain names if those are available, *e.g.* PR_MedeBAC35C06 is used for a PR with accession number AAY82613. To track PR sequences obtained from the nameless strains, we name those by the residue type that occupies a position equivalent to Leu105 in green-absorbing proteorhodopsin (GPR) (11). If there are more than one PR with the same residue type at this position, a respective number is added to the residue type, *e.g.* PR_M7 indicates that this is the seventh PR with Met residue at the position equivalent to Leu105 in GPR. Supplemental Table 1 (See Supplemental Materials) contains a list of PR sequences obtained from the uncultured bacteria.

All rhodopsin structures were obtained from the Protein Data Bank. We chose high-resolution X-ray structures with resolution 2.5 Å or better if there was more than one structure for a given rhodopsin available.

Estimation of amino acid residue substitution rates in the retinal-binding pocket. We used a continuous-time Markov model to estimate substitution rates of amino acid residues forming a retinal-binding pocket of rhodopsins based on known three-dimensional structures. We aimed to estimate the values of the instantaneous substitution rate Q matrix. The continuous-time Markov model for residue substitutions has been implemented in several studies using maximum likelihood estimation, and has also been applied in a protein-folding study (12). Here, we adopt a Bayesian approach where we describe the instantaneous substitution rate $Q = \{q_{ij}\}$ by a posterior distribution $\pi(Q|S,T)$, which summarizes prior information available on the rates $Q = \{q_{ij}\}$ and the information contained in the observations sequences S and an optimal tree topology T. After integrating the prior information and the likelihood function, the posterior distribution $\pi(Q|S,T)$ can be estimated up to a constant and later, the posterior means $E_\pi = \int Q\, \pi(Q|S,T)dQ$ of rates in $\pi(Q)$ could be estimated in the next step.

We then ran a Markov chain to generate samples drawn from the target distribution $\pi(Q|S,T)$. Starting from a rate matrix Q_t at time t, we generate a new rate matrix Q_{t+1} using the proposal function $T(Q_t,Q_{t+1})$. The proposed new matrix Q_{t+1} will be either accepted or rejected, depending on the outcome of an acceptance rule $r(Q_t,Q_{t+1})$. This is achieved by using the Metropolis-Hastings acceptance ratio $r(Q_t,Q_{t+1})$ to either accept or reject Q_{t+1}, depending on whether the following inequality holds:

$$u \leq r(Q_t, Q_{t+1}) = \min\left\{1, \frac{\pi(Q_{t+1}|S,T) \cdot T(Q_{t+1}, Q_t)}{\pi(Q_t|S,T) \cdot T(Q_t, Q_{t+1})}\right\}$$

where u is a random number drawn from the uniform distribution $U[0,1]$. With the assumption that the underlying Markov process is ergodic, irreducible and aperiodic, a Markov chain generated following these rules will reach the stationary state. Details of our method can be found in Tseng and Liang (7). This method is used to obtain amino acid substitution matrices for residues in retinal-binding pockets for all type I rhodopsins, as well as for subsets of BRs, SRIs, SRIIs and PRs.

RESULTS AND DISCUSSION

Retinal-binding pockets in the structures of type I rhodopsins

High-resolution X-ray structures are available for several functional classes of type I rhodopsins. These are BR and HR from *Halobacterium salinarum*, archaerhodopsin-1 and archaerhodopsin-2 from *Halobacterium* sp., SRII from *Natronomonas pharaonis* and ASR from *Anabaena* sp. We have calculated the size and amino acid composition of the retinal-binding pockets in type I rhodopsin structures using the CASTp server http://sts.bioengr.uic.edu/castp/ (13). CASTp uses the weighted Delaunay triangulation and the alpha complex for shape measurements (14–16), which allow for the accurate determination of the atoms forming the wall of the pocket or void in the protein structure without a need for a cutoff (17). Figure 1(a) shows a structure of SRII (PDB: 1H2S) where atoms forming a retinal-binding pocket as determined by CASTp are shown in spacefill. Figure 1(b) shows the full sequence of SRII with the positions of the residues forming the retinal-

Table 2. Type I rhodopsin sequences used in this study

Species	Abbreviation	Accession number
Archaea		
Bacteriorhodopsins		
Haloarcula argentinensis	Harc_argent_BR	Q57101
Haloarcula japonica	Harc_jap_BR	BAA81816
Haloarcula marismortui	Harc_maris_BR	YP_137573
Haloarcula marismortui	Harc_maris_BR2	YP_136594
Haloarcula sp. AB19	Harc_AB19_BR	ABC72081
Haloarcula sp. ARG-2	Harc_BR	AAB32951
Haloarcula sp. YW016	Harc_YW016_BR	AAZ76910
Haloarcula vallismortis	Harc_val_CR3	P94854
Halobacterium salinarum	Hbac_sal_BR	AAG42454
Halobacterium salinarum (Mex)	Hbac_sal_Mex_BR	P33969
Halobacterium salinarum (Port)	Hbac_sal_Port_BR	P33971
Halobacterium salinarum (Shark)	Hbac_sal_Shark_BR	P33972
Halobacterium, SG1 AR1	Hbac_SG1_AR1	AAB26387
Halobacterium sp. arg-4	Hbac_sal_arg4_BR	O93740
Halobacterium sp. AUS-2	HspAUS2_BR	AAB19870
Halobiforma lacisalsi	Hbif_lac_BR	AAU04564
Halorubrum sodomense	Hrub_sod_AR3	P96787
Halorubrum sp. YW059	Hrub_YW059_BR	AAZ76911
Halorubrum xinjiangense	Hrub_xinj_BR	AAS15567
Natrinema ajinwuensis	Natr_ajin_BR	AAS87573
Natrinema altunense	Natr_altrun_BR	AAP36746
Natronococcus aibiensis	Natr_aib_BR	AAP36744
Halorhodopsins		
Haloarcula marismortui ATCC 43049	Harc_mar_HR	YP_136278
Haloarcula vallismortis	Harc_val_HR	P94853
Halobacteriaceae gen. sp.	Hbac_gen_HR	A56808
Halobacterium salinarum (Mex)	Hbac_sal_Mex_HR	P33970
Halobacterium salinarum (Shark)	Hbac_sal_Shark_HR	Q48314
Halobacterium SG1	Hbac_SG1_HR	AAB26388
Halobacterium sp. arg-4	Hbac_arg4_HR	O93741
Halobacterium sp. NRC-1	Hbac_NRC1_HR	AAG18795
Halorubrum sodomense	Hrub_sod_HR	O93742
Natronomonas pharaonis DSM 2160	Natr_phar_HR	YP_325981
Sensory rhodopsins I		
Haloarcula marismortui ATCC 43049	Harc_mar_SRI	YP_137680
Haloarcula vallismortis	Harc_val_SRI	Q48334
Halobacterium SG1	Hbac_SG1_SRI	AAB26389
Halobacterium sp. NRC-1	Hbac_NRC1_SRI	AAG19914
Halorubrum sodomense	Hrub_sod_SRI	BAA75203
Sensory rhodopsins II		
Haloarcula vallismortis	Harc_val_SRII	CAA84550
Halobacterium salinarum	Hbac_sal_SRII	AAC44370
Halobacterium sp. AUS-1	Hbac_AUS1_SRII	BAB86796
Halobacterium sp. NRC-1	Hbac_NRC1_SRII	AAG19988
Natronomonas pharaonis	SRII_1H2S	Z35086
Unknown function		
Haloarcula marismortui	Harc_mar_OUNK	YP_135281
Eukaryotes		
Fungi		
Botrytis cinerea	Bot_cin_R	AL115930
Chaetomium globosum	Chaet_glob_HP	EAQ83604
Cryptococcus neoformans	Cryp_neof_O1	AAW44466
Gibberella fujikuroi	Gibb_fuji_O	CAD97459
Gibberella zeae	Gibb_zeae_HP1	XP_383240
Gibberella zeae	Gibb_zeae_HP2	XP_387730
Leptosphaeria maculans	Lept_mac_O	AAG01180
Neurospora crassa	N_crassa_O1	XP_959421
Ustilago maydis 521	U_mayd_UM00371	XP_756518
Ustilago maydis 521	U_mayd_UM02629	XP_758776
Ustilago maydis 521	U_mayd_UM04125	XP_760272
Algae		
Acetabularia acetabulum	Acet_ace_R	CF259014
Chlamydomonas reinhardtii	CSRA	AAM44041
Chlamydomonas reinhardtii	CSRB	AAM15777
Chlamydomonas reinhardtii	Ch_rein_CHL-5	AAQ16277
Cryptomonas sp. S2	Cryptom_S2_O	ABA08439
Guillardia theta	Guill_theta_O1	ABA08437
Guillardia theta	Guill_theta_O2	ABA08438
Pyrocystis lunula	Plun_R	AF508258
Bacteria		
Cellulophaga sp. MED134	Cell_BR	ZP_01049273
Exiguobacterium sibiricum 255-15	Exi_sib_R	ZP_00537900
Kineococcus radiotolerans	Kin_rad_R	EAM73404
Polaribacter irgensii 23-P	Polar_irgen_BR	ZP_01117885
Psychroflexus torques	Psy_tor_BRa	ZP_01253360
Psychroflexus torques	Psy_tor_BRb	ZP_01255583
Rubrobacter xylanophilus DSM	Rubr_xyl_R	ZP_00600660
Salinibacter ruber DSM 13855	Sal_rub_SRIa	YP_446609
Salinibacter ruber DSM 13855	Sal_rub_SRIb	YP_446677
Salinibacter ruber DSM 13855	Sal_rub_XR	YP_445623
Salinibacter ruber DSM 13855	Sal_rub_HR	YP_446872
Tenacibaculum sp. MED152	Tenac_BR	ZP_01054176
Cyanobacteria		
Anabaena sp. PCC 7120	ASR_1XIO	BAB74864
Gloeobacter violaceus	Glo_vio_R	BAC88139
Eubacteria		
Candidatus pelagibacter HTCC1062(α)	Cand_pel_BR	AAZ21446
EBAC20E09 γ-Proteobacterium	PR_EBAC20E09_Q	AAS73014
eBACHOT4E07 γ-Proteobacterium	PR_eBACHOT4E07_Q	AAT38609
EBAC31A08 γ-Proteobacterium	EBAC31A08_GPR_L	AAG10475
HOT2C01 α-Proteobacterium	PR_HOT2C01_Q	AAR05342
HOT75M4 γ-Proteobacterium	BPR_Q	Q9AFF7
Photobacterium sp. SKA34 (γ)	Photo_SKA34_BR	ZP_01161099
Vibrio angustum S14 (γ)	Vib_ang_BR	ZP_01236264
Environmental samples		
BAC17H8 Uncultured bacterium	PR_BAC17H8_Q	AAY87315
MedeBAC35C06 Uncultured bacterium	PR_MedeBAC35C06_L	AAY82613
Uncultured bacterium	PR_A	AAO21449
Uncultured bacterium	PR_E2	AAO73911
Uncultured bacterium	PR_L24	AAO21450
Uncultured bacterium	PR_L34	AAO73919
Uncultured bacterium	PR_L45	AAO73935
Uncultured bacterium	PR_M7	AAT38637
Uncultured bacterium	PR_M8	AAT38639
Uncultured bacterium	PR_M9	AAT39128
Uncultured bacterium	PR_Q20	AAT09901
Uncultured bacterium	PR_Q22	AAY68058
Uncultured bacterium	PR_Q23	AAZ40531
Uncultured bacterium	PR_V	AAO21455

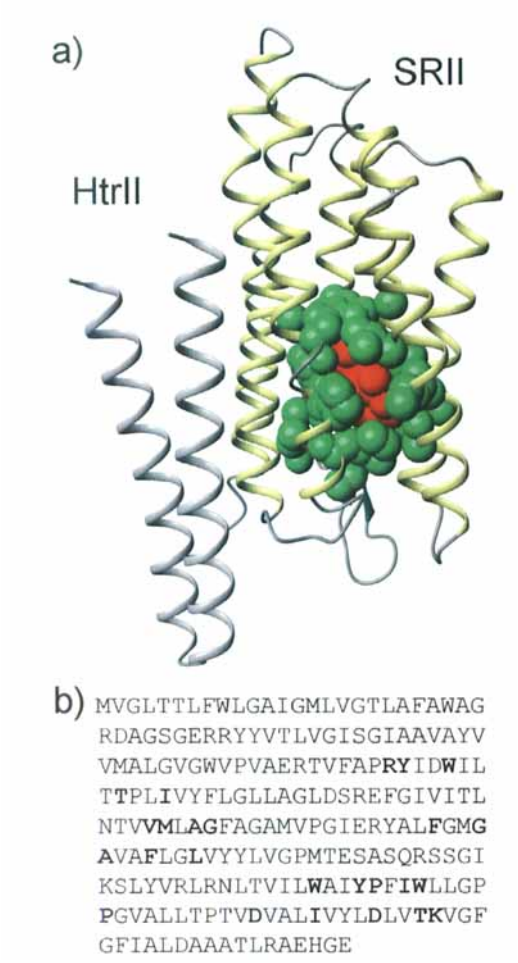

Figure 1. (a) Structure of sensory rhodopsin II (SRII) with transducer HtrII (PDB: 1H2S). Atoms forming a retinal-binding pocket are shown in green. Retinal atoms are in red. (b) The full sequence of SRII with residues contributing atoms to the retinal-binding pocket wall shown in bold face. (c) Sequence of retinal-binding pocket.

Figure 2. Mapping of residues forming a retinal-binding pocket to the type I rhodopsin sequences of known structures. Multiple sequence alignment contains sequences of bacteriorhodopsin (1M0L, 1M0M, 1MOK, 1P8H, 1O0A, 1QKO, 1KGB and 1F50), archaerhodopsin-1 (1UAZ), archaerhodopsin-2 (1VGO), halorhodopsin (1E12), sensory rhodopsin II (1GU8, 1GUE, 1H68, 1H2S), and anabaena sensory rhodopsin (1XIO). Residues forming the retinal-binding pocket in individual structure are shown in bold. The residues are numbered following the convention of bacteriorhodopsin.

binding pocket wall mapped onto the full sequence (shown in bold). We map residues forming the walls of the calculated retinal-binding pockets in all of the rhodopsin structures to the respective aligned sequences in the multiple sequence alignment of the type I rhodopsin sequences listed in Table 1 (Fig. 2). There are 18 positions (83, 86, 90, 93, 118, 119, 122, 138, 141, 142, 145, 182, 185, 186, 189, 208, 215 and 216 in BR residue numbering) in the MSA where residues always contribute to the retinal-binding pocket independently of the function, structure or the physical state of the particular rhodopsin. Nine of these positions (83, 86, 90, 122, 182, 185, 186, 189 and 216 in BR residue numbering) are invariant in the given set of the structures.

We build pocket sequences by concatenating the residues forming the wall of the calculated pocket as described by Binkowski *et al.* (18). Residues in the pocket sequences are often nonconsecutive. The pocket sequence of SRII as calculated for structure 1H2S is shown on Fig. 1(c). The pocket sequences for other structures are shown in Table 1. BR, HR, archaerhodopsin-1 and archaerhodopsin–2 have retinal-binding pockets of about 18–20 residues long with volumes between 450 Å^3 and 533 Å^3. ASR has a 21-residues-long retinal-binding pocket, whereas SRII has the largest pocket, which is 24 to 26 residues long with the volume ranging between 630 Å^3 and 702 Å^3 depending on the specific Protein Data Bank structure. A slight variation in the pocket size and, as a consequence, of the amino acid composition of the pockets from different structures of the same protein is likely due to the uncertainty in the assignment of residues to the electronic density that is detected by highly sensitive geometry calculations used in this study.

To compare the residues forming a retinal-binding pocket across different type I rhodopsins, we created "superpocket" sequences from the full-length sequences of each structure by taking residues at positions that appear in the retinal-binding pocket in SRII, plus an additional position, Gly125 in BR, which doesn't occur in SRII but occurs in BR, HR and ASR retinal-binding pockets. The total length of a superpocket is 26 residues. We chose SRII pocket because SRII has the largest retinal-binding pocket and would provide more complete coverage of the retinal-interacting residues across the different classes of type I rhodopsins. Of the four SRII structures listed, we chose 1H2S structure because it represents a native functional form of SRII with a bound transducer.

We found that the sequence identity between superpocket sequences in the structural data set (Table 1) ranges between 39% and 61% for proteins with different functions. The respective full-

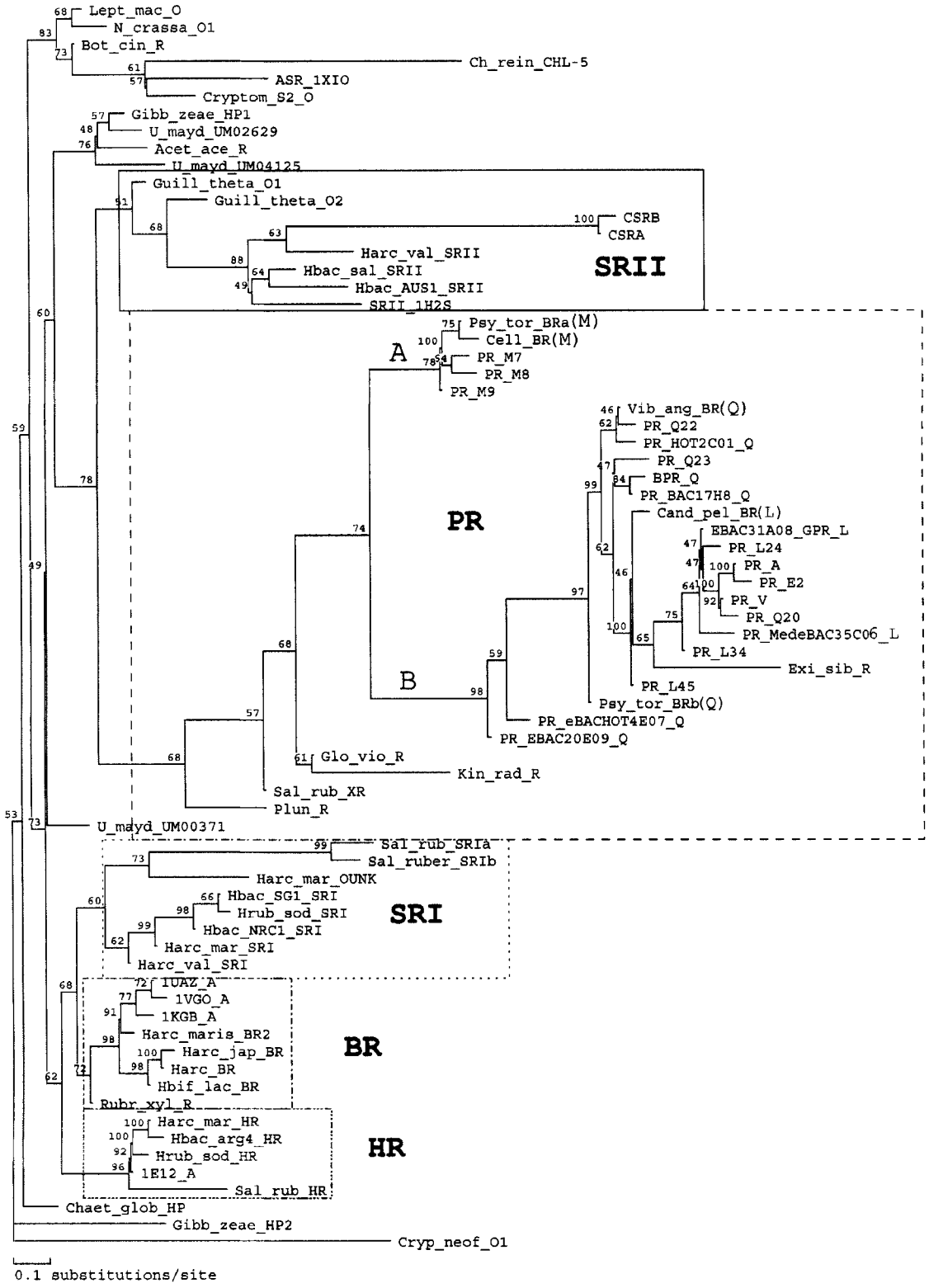

Figure 3. Maximum likelihood relationship tree of the retinal-binding superpockets. Rhodopsins of different biological functions are automatically grouped into different branches: sensory rhodopsin II branch (SRII), proteorhodopsin branch (PR, groups A and B), sensory rhodopsin I branch (SRI), bacteriorhodopsin branch (BR), halorhodopsin branch (HR), and a few smaller branches containing fungal rhodopsins. Abbreviations are from Table 2.

length sequences are usually only 23–35% identical. Pockets obtained from the structures belonging to the same functional class of proton pumps, namely BR, archaeorhodopsin-1 (PDB: 1UAZ) and archaeorhodopsin-2 (PDB: 1VGO), are much more similar, with typical sequence identity of 85–94% and 55–86% for the superpocket and the full-length sequences, respectively. This demonstrates that residues interacting with retinal chromophore are more conserved than the full sequence between different rhodopsins (5).

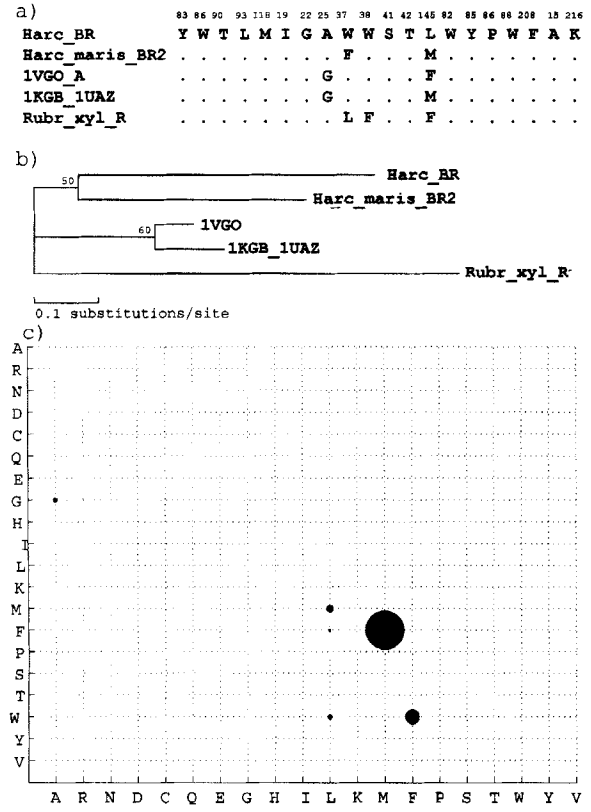

Figure 4. Amino acid substitution rates in the retinal-binding pockets of bacteriorhodopsins. (a) Alignment of unique pocket sequences. The 20 pocket residue positions are from retinal-binding pocket in structure 1KGB. Residues that are identical to the residues in the first sequence are substituted with " • ". (b) Maximum likelihood phylogeny tree of the full-length BR sequences. (c) A plot of amino acid substitution rates. The area of the circles is proportional to the substitution rate. There are three high-rate amino acid substitutions in bacteriorhodopsins: F↔M, F↔W, and L↔M.

Superpocket sequences of type I rhodopsins allow a better functional discrimination than the whole sequences

There is a significant divergence among the sequences of type I rhodopsins. Traditional phylogenetic tree is obtained from full-length sequences and incorporates both selection pressure due to protein folding and selection pressure due to protein function. It reflects the overall evolutionary ancestral and sibling relationship between different rhodopsins, which does not always permit assignment of particular sequences as encoding transport or sensory proteins (19,20). Here, we use a different approach and build a relationship tree for the short sequences of the putative retinal-binding superpockets from the type I rhodopsins. This tree is different from the phylogenetic tree because it describes the relatedness of rhodopsins in terms of retinal-binding pockets, and hence is more informative for understanding how they are related in their functions.

Table 2 lists type I rhodopsin sequences used in this study. We collected 84 sequences from archaea, eukaryotes (fungi and algae) and different types of bacteria. Our data set also contains type I rhodopsin sequences obtained from environmental samples. Table 2 lists 14 such sequences, whereas the full list is available from Table 1 in the supplementary material. We have aligned all 98 full-length sequences listed in Table 2 using ClustalW multiple-sequence alignment algorithm with BLOSUM series matrices (21) and manually adjusted the obtained MSA using Pfaat program (10). We then built a superpocket alignment by removing non-pocket positions from the full-length alignment. Superpockets from closely related proteins often have the same sequences. We have eliminated the duplicates and obtained a final set of 71 unique superpocket sequences. These 71 superpocket sequences were re-aligned using a customized matrix for retinal-binding pockets of type I rhodopsins obtained with a Bayesian Monte Carlo approach (7) as described in Methods section. The use of this matrix improves the characterization of the functional relationship among different proteins. We then built a relationship tree of the superpocket sequences (Fig. 3) using a maximum likelihood algorithm PAML (22). Several distinct branches of rhodopsins of different functions are observed on this tree: SRII branch, PR branch, SRI branch, BR branch, HR branch and a few smaller branches containing fungal rhodopsins.

SRII branch groups superpockets of archaeal SRII as well as rhodopsins from cryptomonad alga with experimentally demonstrated sensory function. These are sensory rhodopsins A and B (CSRA and CSRB, respectively) from *Chlamydomonas reinhardtii* (23), and opsins 1 and 2 from *Guillardia theta*, which were recently shown to possess a sensory function and likely form a dual rhodopsin-based photosensory system as described by Sineschekov *et al.* (24).

PR branch contains superpockets from PRs as well as rhodopsins from a few other organisms, which often exist in the extreme environments. These include xanthorhodopsin from *Salinibacter ruber*, a bacterium surviving at extremely high concentration of NaCl (25). A recent study by Balashov *et al.* (26) showed that xanthorhodopsin not only includes the chromophore retinal and pumps protons but also utilizes a carotenoid antenna, which enables the rhodopsin to absorb light across a greater absorbance spectrum than is possible with retinal alone. There is also a rhodopsin-like protein from *Gloeobacter violaceus*, a unique cyanobacterium that lacks thylakoid membranes (accession number BAC88139) (27). *Gloeobacter* rhodopsin expressed in *Escherichia coli* is an efficient light-driven proton transporter with a fast photocycle similar to that of the high-pH form of GPR (28). Finally, there is a pocket sequence with a putative proton pump function from *Kineococcus radiotolerans*, an aerobic bacterium thriving in high-level radioactive waste. *Pyrocystis lunula* pocket is also on this branch, indicating a possibility for a proton-pumping activity.

PR retinal-binding superpockets are clearly separated into two groups: A and B. The smaller group A contains superpockets with the Met residue at the position equivalent to 105 of GPR, whereas the larger group B contains PRs with a wide array of residue types at this position, such as Ala, Glu, Leu, Gln and Val. GPR as well as deepwater BPR, which contain Leu and Gln residues at this position, respectively, are found in the same group B of the retinal-binding pockets. Close to this group is a retinal-binding pocket of rhodopsin from *Exiguobacterium* (*Exi_sib_R* on Fig. 3) isolates from ancient permafrost sediments (ZP_00537900) (29), thriving at extremely low temperatures.

SRI branch contains classical archaeal SRIs including a *Haloarcula marismortui* rhodopsin (accession number YP_135281) with unknown function, as well as two SRIs from *Salinibacter ruber* (25). BR branch contains archaeal BRs and a recently discovered BR from thermophilic bacterium *Rubrobacter xylanophilus* (30). Retinal-binding pockets of archaeal HRs are also grouped into a separate branch.

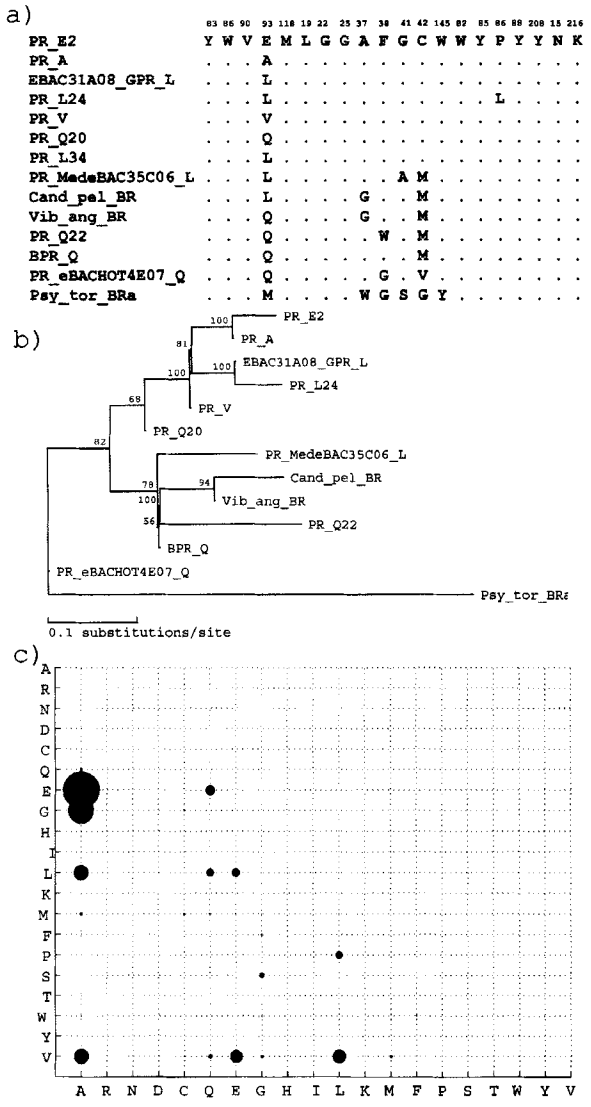

Figure 5. Amino acid substitution rates in the putative retinal-binding pockets of proteorhodopsins. (a) Alignment of putative pocket sequences. The 20 pocket residue positions are from retinal-binding pocket in BR structure 1KGB. Residues that are identical to the residues in the first sequence are substituted with " • ". (b) Maximum likelihood phylogeny tree of the full-length PR sequences. (c) A plot of amino acid substitution rates. The area of the circles is proportional to the substitution rate. The exchange pairs with the fastest rates are mainly found at positions 93 and 137 in PR (following BR numbering). These are: A↔L, A↔V, A↔E, E↔Q, E↔L, L↔Q, L↔V and M↔T.

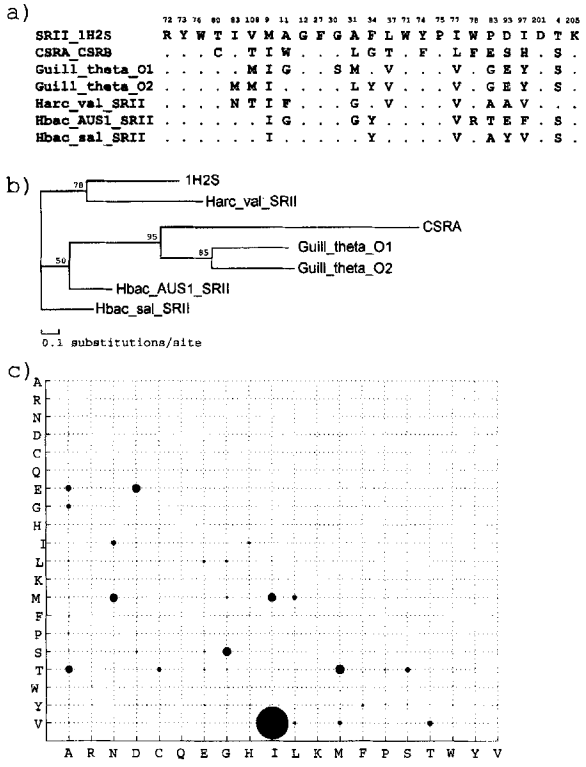

Figure 6. Amino acid substitution rates in the retinal-binding pockets of sensory rhodopsins II. (a) Alignment of pocket sequences. The 25 pocket residue positions are from retinal-binding pocket in structure 1H2S. Residues that are identical to the residues in the first sequence are substituted with " • ". (b) Maximum likelihood phylogeny tree of the full-length SRII sequences. (c) A plot of amino acid substitution rates. The area of the circles is proportional to the substitution rate. The fastest-exchanging amino acid pairs are I↔V, M↔T and G↔S.

ASR (PDB: 1XIO) (31) is different from archaeal sensory rhodopsins in two aspects. First, it interacts with a cytosolic transducer, whereas archaeal SRI and SRII interact with other integral membrane proteins. Second, it has an inverse mode of retinal isomerization (as compared with BR) with a primarily all-trans configuration of retinal in the dark-adapted state and a primarily 13-cis configuration in the light-adapted state (32). ASR superpocket has the highest sequence identity (65%) and clusters together with a retinal-binding superpocket from algae *Cryptomonas sp* S2, chlamiopsin-5 from *Chlamydomonas reinhardtii*, and an opsin from fungi *Botritis cinerea*, which tentatively suggests a similarity in signal transduction mechanism in these proteins, although more sequence information would be required to fully establish this.

Superockets from fungal rhodopsins used in this study (excluding Bot_cin rhodopsin) did not cluster with the pockets from other species with known function. The putative retinal-binding superpockets from *Neurospora crassa* NOP-1 and *Leptosphaeria maculans* rhodopsins grouped together because they differ only by one residue. However, it has been experimentally demonstrated that these two rhodopsins have different functions such as proton pump and sensory rhodopsin, respectively (33–35). It seems that pocket residue positions obtained from archaeal rhodopsins may not be directly transferred to fungal rhodopsins, which may have a different arrangement of the residues around the bound retinal molecule.

Substitution rates of amino acid residues in the retinal-binding pockets

We calculated substitution rates for the amino acid residues of the retinal-binding pockets to examine their function-specific evolutionary patterns. The substitution rates are calculated using a Bayesian Markov Chain Monte Carlo method developed by Tseng & Liang (7). This method was developed with the assumption that different regions of the protein experience different evolutionary pressure and have different substitution rates for amino acid residues. For example, Tseng & Liang showed that the residue substitution rates are significantly different for the buried and exposed residues, as well as for residues forming specific binding surfaces on protein structures (7).

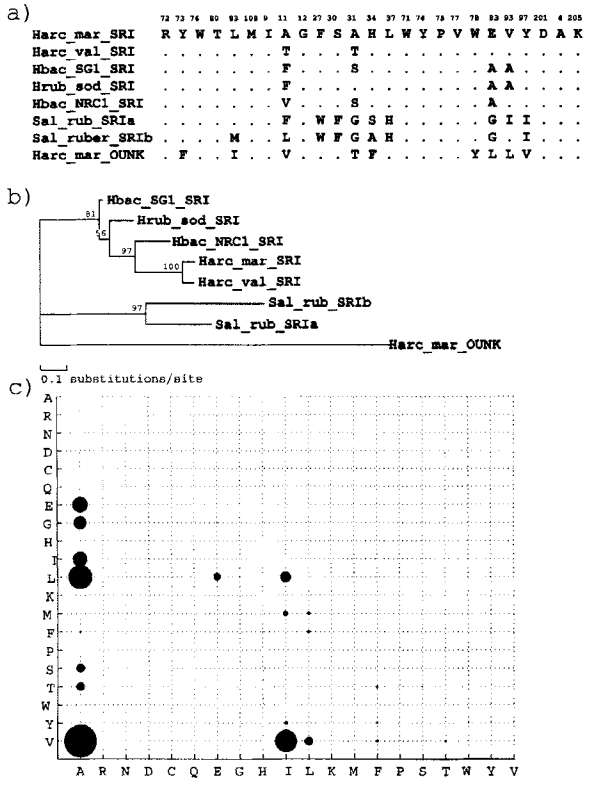

Figure 7. Amino acid substitution rates in the putative retinal-binding pockets of sensory rhodopsins I. (a) Alignment of pocket sequences. The 25 pocket residue positions are from retinal-binding pocket in structure 1H2S. Residues that are identical to the residues in the first sequence are substituted with " • ". (b) Maximum likelihood phylogeny tree of the full-length SRI sequences. (c) A plot of amino acid substitution rates. The area of the circles is proportional to the substitution rate. The high-rate exchange pairs include A↔V, I↔V, A↔L, A↔E, A↔G and A↔I.

In this study, we use function-specific sets of rhodopsin sequences based on the clustering of the superpockets as presented on Fig. 3. We use residue positions from the structure-defined retinal-binding pockets for calculation of substitution rates in BRs and SRIIs, and putative retinal-binding pockets for calculation of substitution rates in SRIs and PRs. The analysis of HR retinal-binding pockets is not feasible because of the extremely high conservation of the retinal-binding pocket: there are only three unique pocket sequences available when amino acid residue positions from 1E12 retinal-binding pocket are used.

Figures 4–7 present multiple sequence alignments of the pocket sequences where only pocket residue positions specific for a given rhodopsin function, as obtained by the geometric calculations on the respective PDB structure, were used. This approach produces shorter pocket sequences in comparison with the superpocket sequences that were used to generate a relationship tree on Fig. 3. In addition, shorter pocket sequences are often identical. This is why the number of sequences on Figs. 4–7 is smaller than can be found on Fig. 3 for a specific function. For example, BR branch contains eight superpockets, whereas only five unique pocket sequences are available when the number of included positions decreased from 26 to 20. Each figure also contains the plots where the area of the filled circles is proportional to the substitution rate. Only a handful of high-rate substitutions marked by prominently large filled circles is observed in all four plots. These are M↔F substitution in BR (Fig. 4), A↔E/G substitutions in PR (Fig. 5), I↔V substitution in SRII (Fig. 6) and A↔V substitutions in SRI (Fig. 7).

In BR, substitutions occur in 4 of 20 pocket amino acid residue positions. Figure 4 shows that three high-rate amino acid substitutions in BR are F↔M, F↔W and L↔M. Substitution pairs F↔M and L↔M occur at the same position 145, pair F↔W is found in two different positions (137 and 138). F↔W and L↔M are widely accepted in membrane proteins, whereas substitution F↔M is rarely observed (36). A recent crystallographic study by Enami *et al.* (37) showed how F↔M exchange affects the thermodynamic properties of retinal in BR. Namely, the 13-cis isomer is better stabilized in methionine-containing pockets of BR and AR-1 because of the protruding side-chain of Met145 that contacts the C5 methyl of retinal and pushes the ionone ring downwards. On the other hand, the all-trans isomer is better stabilized in AR-2 because the benzene ring of Phe145 is far from the C5 methyl of retinal (37).

We used a BR pocket as a template for constructing putative retinal-binding pockets in PRs because it has been experimentally shown that some PRs function as light-driven proton pumps (4,38). The substitution patterns in PR retinal-binding pockets are different from those observed in BR pockets (Fig. 5). For example, position 93 (following BR numbering) is conserved in BRs, but is highly variable in PRs; position 125 is conserved in PR, whereas the same position is variable in BR. The exchange pairs with the fastest rates are mainly found at positions 93 and 137 in PR (following BR numbering). These are: A↔L, A↔V, A↔E, E↔Q, E↔L, L↔Q, L↔V and M↔T. Another fast-exchanging pair A↔G occurs at positions 138, 142 and 145.

In the alignment of PR pocket sequences (Fig. 5), position 93 (following BR numbering) corresponds to position 105 in GPR and 106 in BPR where Leu and Gln residues are found, respectively. Wang *et al.* (11) demonstrated that the mutation L105Q in GPR *vs* Q106L in BPR nearly completely interconverts the absorption spectra of the two pigments and affects the photocycle of the mutated proteins. It has also been suggested that the BPR may be a light sensor rather than a proton pump because of its longer photocycle as was experimentally determined by Wang *et al.* (11) and the fact that the sites evolving under a positive Darwinian selection are far from the retinal-binding pocket (the opposite effect was found for GPR) and dispersed over the modeled structure of BPR, implying that the exterior surface of BPR was the main subject of the positive Darwinian selection (39). On the basis of our data, we propose that similarly to the fast-exchanging position 145 in BR, the fast rate of amino acid exchange at position 105/106 in PRs may be an adaptation of the same function to modulate a photocycle in different environments.

PR pocket sequences that contain residues Ala, Glu, Gln, Leu and Val at position 93 (Fig. 5) cluster separately from the PR pocket sequences that contain Met at this position (Fig. 3). It should be noted that five Met-contaning superpockets (26 amino acid residues) from the relationship tree on Fig. 3 converged to one sequence when only BR pocket positions (20 residues) were selected for the calculation of the substitution rates. Interestingly, all (excluding M↔T) exchange pairs with Met residue such as A↔M, C↔M, M↔Q, E↔M and L↔M have slow substitution rates. This may indicate that Met93-containing sequences have a different evolutionary history and may represent PRs that have a different function. An additional argument comes from the positioning of pocket sequences from two new rhodopsins from *Psychroflexus torques*, a sea-ice-derived psychrofilic bacteria (40)

(Psy_tor_Bra and Psy_tor_BRb on Fig. 3 with accession numbers ZP_01253360 and ZP_01255583, respectively). Retinal-binding pockets of *Psychroflexus torques* belong to two different branches: Psy_tor_BRa contains Met at position 93 and groups exclusively with other Met-containing PRs in group A. Psy_tor_BRb contains Gln at position 93 and groups with the remaining PRs in group B (Fig. 3). There are two possible scenarios for *Psychroflexus torques* rhodopsins. First, two rhodopsins represent two different functions, *e.g.* ion pumping and signal transduction. Second, this is a two-rhodopsin system similar to those found in *Chlamydomonas reinhardtii* (CSRA and CSRB) (23), *Guillardia theta* (24) or in *Salinibacter ruber* (there are two SRI rhodopsins in this species) (25). However, the pocket sequences from a two-rhodopsin system are usually closely related paralogues that cluster together. Here, the clustering of *Psychroflexus torques* pocket sequences on two rather distant branches supports a hypothesis about a possible different function for the PRs with Met at position 93.

We used a set of pocket sequences that cluster together with SRII pocket from structure 1H2S to estimate substitution rates in retinal-binding pockets of SRIIs. Figure 6 shows that SRII sequences have a unique set of substitution patterns as well. The fastest-exchanging amino acid pairs are I↔V, M↔T and G↔S. Again, the positions, namely the locations where these pairs occur, may indicate special adaptations of the retinal-binding pocket. For example, G↔S pair occurs only at position 130 (following SRII-1H2S numbering) and is found only between two rhodopsins from *Guillardia theta*. This could be a particular position that modulates the light response in this two-rhodopsin system. Another fast-exchanging pair in pockets of *Guillardia theta* is I↔M. Unfortunately, position 109 also contributes to the exchange rate of these amino acid residues and it is difficult to separate the contribution from these positions.

There are no structures of SRIs, so we used the pocket positions from SRII to analyze the evolutionary patterns in SRI putative pockets. The set of available SRI sequences shows fewer substitution positions in the pocket when compared to SRII and has a different set of high-rate exchange pairs, which includes A↔V, I↔V, A↔L, A↔E, A↔G and A↔I. The largest number of the fast-exchange pairs comes from the positions 193, 183 and 111 following SRII (PDB: 1H2S) residue numbering.

The results from substitution rate calculations demonstrate that although the retinal binding pockets are well conserved within every functional class of type I rhodopsins, there are specific sets of high-rate amino acid substitutions that do not change the function of the particular class of proteins, but rather provide an adaptation to specific environmental conditions or modulate the photocycle of the given rhodopsin.

SUMMARY AND CONCLUSIONS

Our findings indicate that every functional class of type I rhodopsins has a specific allowed set of nonfunctional substitutions of amino acid residues in the retinal-binding pocket. These mutations occur in the course of evolution but they do not fundamentally change the protein function. Rather, they provide a specific adaptation of the particular class of rhodopsins to different environments. We also demonstrate that the short concatenated sequences of amino acid residues forming a wall of the retinal-binding pocket are often sufficient to discriminate a phototactic *vs* transporting function of the bacterial or unicellular algal rhodopsins.

Our study also generates hypotheses that should be further verified by experimental works and additional computational analysis when more sequence and structure data are available. For example, the Met93-containing PRs may have a distinct evolutionary history and different function from PRs that contain other amino acid residues at this position. By this hypothesis, the two PRs from *Psychroflexus torques* (ZP_01253360 and ZP_01255583) may play different biochemical roles. The patterns of close grouping of ASR, algae *Cryptomonas sp. S2* and chlamiopsin-5 from *Chlamydomonas reinhardtii* may suggest that these rhodopsins bear similarity in their mechanisms of signal transduction.

Previous studies showed that experimentally characterized bacterial and eukaryotic rhodopsins may not follow the same grouping as that of the phylogenetic tree (19,20). Our results demonstrate that the evolutionary patterns of the retinal-binding pockets reflect the functional specificity of the rhodopsins better. A traditional phylogenetic tree is obtained from full-length sequences and incorporates both selection pressure due to protein folding and selection pressure due to protein function. Our approach takes advantage of structural information about the retinal-binding pocket as obtained from geometric computation, a detailed analysis of residue substitution rates within the binding pocket and is the first that effectively isolates the effect of selection pressure due to the biological function from the selection pressure due to maintaining protein stability and folding accessibility.

SUPPLEMENTAL MATERIALS

Table 1 contains a list of PR sequences obtained from the uncultured bacteria. It can be found at DOI: 10.1562/2006-02-14-RA-802.s1.

Acknowledgements—This work has been supported in part by grants from the National Science Foundation (CAREER_DBI0133856), National Institutes of Health (GM68958), and Office of Naval Research (N00014-06-0100). We thank the anonymous reviewer for valuable suggestions that allowed us to significantly improve the manuscript.

REFERENCES

1. Ebrey, T. and Y. Koutalos (2001) Vertebrate photoreceptors. *Prog. Retin. Eye Res.* **20**, 49–94.
2. Collins, F. D. (1953) Rhodopsin and indicator yellow. *Nature* **171**, 469–472.
3. Morton, R. A. and G. A. Pitt (1955) Studies on rhodopsin. 9. pH and the hydrolysis of indicator yellow. *Biochem. J.* **59**, 128–134.
4. Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich and E. F. DeLong (2000) Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906.
5. Spudich, J. L., C. S. Yang, K.-H. Jung and E. N. Spudich (2000) Retinylidene proteins: Structures and functions from archaea to humans. *Annu. Rev. Cell Dev. Biol.* **16**, 365–392.
6. Royant, A., P. Nollert, K. Edman, R. Neutze, E. M. Landau, E. Pebay-Peyroula and J. Navarro (2001) X-ray structure of sensory rhodopsin II at 2.0 A resolution. *Proc. Natl. Acad. Sci. USA* **98**, 10131–10136.
7. Tseng, Y. Y. and J. Liang (2006) Estimation of amino acid residue substitution rates at local spatial regions and application in protein function inference: A Bayesian Monte Carlo approach. *Mol. Biol. Evol.* **23**, 421–436.
8. Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. L. Lipman (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
9. Higgins, D. G., A. J. Bleasby and R. Fuchs (1991) Clustal V: Improved software for multiple sequence alignment. *CABIOS* **5**, 151–153.

10. Johnson, J. M., K. Mason, C. Moallemi, H. Xi, S. Somarro and E. S. Huang (2003) Protein family annotation in a multiple alignment viewer. *Bioinformatics* **19**, 544–545.
11. Wang, W.-W., O. A. Sineshchekov, E. N. Spudich and J. L. Spudich (2003) Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J. Biol. Chem.* **278**, 33985–33991.
12. Tseng, Y. Y. and J. Liang (2004) Are residues in a protein folding nucleus evolutionarily conserved? *J. Mol. Biol.* **335**, 869–880.
13. Binkowski, T. A., S. Naghibzadeh and J. Liang (2003) CASTp: Computed atlas of surface topography of proteins. *Nucleic Acids Res.* **31**, 3352–3355.
14. Edelsbrunner, H. and E. P. Mücke (1994) Tree-dimensional alpha-shapes. *ACM Trans. Graphics* **13**, 43–72.
15. Edelsbrunner, H. and N. R. Shah (1996) Incremental topological flipping works for regular triangulations. *Algorithmica* **15**, 223–241.
16. Facello, M. A. (1995) Implementation of a randomized algorithm for Delaunay and regular triangulation in three dimensions. *Comput. Aided Geom. D.* **12**, 349–370.
17. Adamian, L. and J. Liang, (2001) Helix-helix packing and interfacial pairwise interactions of residues in membrane proteins. *J. Mol. Biol.* **311**, 891–907.
18. Binkowski, T., L. Adamian and J. Liang (2003) Interring functional relationships of proteins from local sequence and spatial surface patterns. *J. Mol. Biol.* **332**, 505–526.
19. Spudich, J. L. and K.-H. Jung (2005) *Handbook of Photosensory Receptors*. Wiley-VCH Verlag, Weinheim, UK.
20. Ruiz-Gonzalez, M. X. and I. Marin (2004) New insights into the evolutionary history of type 1 rhodopsins. *J. Mol. Evol.* **58**, 348–358.
21. Henikoff, S. and J. G. Henikoff (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
22. Yang, Z. (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.
23. Sineshchekov, O. A., K.-H. Jung and J. L. Spudich (2002) Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc. Natl. Acad. Sci. USA* **99**, 8689–8694.
24. Sineshchekov, O. A., E. G. Govorunova, K.-H. Jung, S. Zauner, U.-G. Maier and J. L. Spudich (2005) Rhodopsin-mediated photoreception in cryptophyte flagellates. *Biophys. J.* **89**, 4310–4319.
25. Mongodin, E. F., K. E. Nelson, S. Daugherty, R. T. DeBoy, J. Wister, H. Khouri, J. Weidman, D. A. Walsh, R. T. Papke, G. S. Perez, A. K. Sharma, C. L. Nesbo, D. MacLeod, E. Bapteste, W. F. Doolittle, R. L. Charlebois, B. Legault and F. Rodriguez-Valera (2005) The genome of *Salinibacter ruber*: Convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. USA* **102**, 18147–18152.
26. Balashov, S. P., E. S. Imasheva, V. A. Boichenko, J. Anton, J. M. Wang and J. K. Lanyi (2005) Xanthorhodopsin: A proton pump with a light-harvesting carotenoid antenna. *Science* **309**, 2061–2064.
27. Nakamura, Y., T. Kaneko, S. Sato, M. Mimuro, H. Miyashita, T. Tsuchiya, S. Sasamoto, A. Watanabe, K. Kawashima, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, N. Nakazaki, S. Shimpo, C. Takeuchi, M. Yamada and S. Tabata (2003) Complete genome structure of *Gloeobacter violaceus PCC 7421*, a cyanobacterium that lacks thylakoids (supplement). *DNA Res.* **10**, 181–201.
28. Brown, L. S. and K.-H. Jung (2006) Bacteriorhodopsin-like proteins of eubacteria and fungi: The extent of conservation of the haloarchaeal proton-pumping mechanism. *Photochem. Photobiol. Sci.* **5**, 538–546.
29. T. Vishnivetskaya, S. Kathariou, J. McGrath, D. Gilichinsky and J. M. Tiedje (2000) Low-temperature recovery strategies for the isolation of bacteria from ancient permafrost sediments. *Extremophiles* **4**, 165–173.
30. Ferreira, A. C., M. F. Nobre, E. Moore, F. A. Rainey, J. R. Battista and M. S. da Costa (1999) Characterization and radiation resistance of new isolates of *Rubrobacter radiotolerans* and *Rubrobacter xylanophilus*. *Extremophiles* **3**, 235–238.
31. Vogeley, L., O. A. Sineshchekov, V. D. Trivedi, J. Sasaki, J. L. Spudich and H. Luecke (2004) *Anabaena* sensory rhodopsin: A photochromic color sensor at 2.0 Å. *Science* **306**, 1390–1393.
32. Sineshchekov, O. A., V. D. Trivedi, J. Sasaki and J. L. Spudich (2005) Photochromicity of *Anabaena* sensory rhodopsin, an atypical microbial receptor with a *cis*-retinal light-adapted form. *J. Biol. Chem.* **280**, 14663–14668.
33. Waschuk, S. A., A. G. Bezerra, L. Shi and L. S. Brown (2005) *Leptosphaeria rhodopsin*: Bacteriorhodopsin-like proton pump from a eukaryote. *Proc. Natl. Acad. Sci. USA* **102**, 6879–6883.
34. Sumii, M., Y. Furutani, S. A. Waschuk, L. S. Brown and H. Kandori (2005) Strongly hydrogen-bonded water molecule present near the retinal chromophore of *Leptosphaeria* rhodopsin, the bacteriorhodopsin-like proton pump from a eukaryote. *Biochemistry* **44**, 15159–15166.
35. Bieszke, J. A., E. N. Spudich, K. L. Scott, K. A. Borkovich and J. L. Spudich (1999) A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment. *Biochemistry* **38**, 14138–14145.
36. Jones, D. T., W. R. Taylor and J. M. Thornton (1994) A mutation data matrix for transmembrane proteins. *FEBS Lett.* **339**, 269–275.
37. Enami, N., K. Yoshimura, M. Murakami, H. Okumura, K. Ihara and T. Kouyama (2006) Crystal structures of archaerhodopsin-1 and -2: Common structural motif in archaeal light-driven proton pumps. *J. Mol. Biol.* **358**, 675–685.
38. Beja, O., E. N. Spudich, J. L. Spudich, M. Leclerc and E. F. DeLong (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411**, 786–789.
39. Bielawski, J. P., K. A. Dunn, G. Sabehi and O. Beja (2004) Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl. Acad. Sci. USA* **101**, 14824–14829.
40. Bowmann, J. P., S. A. McCammon, T. Lewis, J. H. Skerratt, J. L. Brown, D. S. Nichols and T. A. McMeekin (1998) *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al, 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* **144**, 1601–1609.