

Microbial and Animal Rhodopsins: Structures, Functions and Molecular Mechanisms

Oliver P. Ernst¹, David T. Lodowski², Marcus Elstner³, Peter Hegemann⁴, Leonid S. Brown⁵, and Hideki Kandori⁶

¹Departments of Biochemistry and Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, Ontario M5S 1A8, Canada, ²Center for Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, 10900 Euclid Avenue, Cleveland, Ohio 44106, United States, ³Institute for Physical Chemistry, Karlsruhe Institute of Technology, Kaiserstrasse 12, 76131 Karlsruhe, Germany, ⁴Institute of Biology, Experimental Biophysics, Humboldt-Universität zu Berlin, Invalidenstrasse 42, 10115 Berlin, Germany, ⁵Department of Physics and Biophysics Interdepartmental Group, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada, ⁶Department of Frontier Materials, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan

Helix	A	B	B	B	C	C	C	C	C	C	C	D	D	D	E	E	F	F	F	F	-	G	G	G	G	G	G						
# in BR	20	49	53	56	57	82	83	85	86	89	90	91	93	96	115	118	122	141	145	150	182	185	186	189	194	204	208	212	215	216	219	220	
BR	M	V	A	M	Y	R	Y	D	W	T	T	P	L	D	D	M	G	S	M	Y	W	Y	P	W	E	E	F	D	A	K	F	G	
AR1	M	V	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	S	M	Y	W	Y	P	W	E	E	F	D	A	K	F	G	
AR2	M	V	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	S	F	Y	W	Y	P	W	E	E	F	D	A	K	F	G	
AR3	M	V	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	S	M	Y	W	Y	P	W	E	E	F	D	A	K	F	G	
MR	?	M	I	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	S	F	Y	W	Y	P	W	G	E	Y	D	T	K	F	G
GPR ¹	M	I	A	H	Y	R	Y	D	W	T	V	P	M	E	S	M	G	S	Y	Y	W	Y	P	Y	P	L	Y	D	N	K	F	G	
GPR ²	L	V	A	H	Y	R	Y	D	W	T	V	P	L	E	T	M	G	G	W	Y	W	Y	P	Y	M	L	Y	D	N	K	F	G	
GPR ³	L	V	A	H	Y	R	Y	D	W	T	V	P	L	E	T	M	G	G	W	Y	W	Y	P	Y	M	L	Y	D	N	K	F	G	
BPR ⁴	M	V	A	H	Y	R	Y	D	W	T	V	P	Q	E	S	M	G	G	W	Y	W	Y	P	Y	F	L	Y	D	N	K	F	G	
BPR ⁵	M	I	A	H	Y	R	Y	D	W	T	V	P	Q	E	S	M	G	G	W	Y	W	Y	P	Y	L	L	Y	D	N	K	F	G	
GPR ⁶	L	V	A	H	Y	R	Y	D	W	T	V	P	L	E	S	M	G	G	W	Y	W	Y	P	Y	L	L	Y	D	N	K	F	G	
GPR ⁷	L	V	A	H	Y	R	Y	D	W	T	V	P	L	E	S	M	G	G	W	Y	W	Y	P	Y	L	L	Y	D	N	K	F	G	
GR	M	V	A	H	Y	R	Y	D	W	T	V	P	L	E	S	M	G	S	F	Y	W	Y	P	Y	L	V	Y	D	A	K	F	G	
XR	M	V	A	H	Y	R	Y	D	W	T	V	P	L	E	A	M	G	S	F	Y	W	Y	P	Y	A	L	Y	D	A	K	Y	G	
Ace1	M	I	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	G	L	W	W	Y	P	W	G	E	M	D	A	K	Y	A	
Ace2	M	I	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	G	Q	Y	W	Y	P	W	G	E	M	D	A	K	F	G	
LR	M	I	A	S	Y	R	Y	D	W	T	T	P	L	D	D	M	G	A	Y	W	W	Y	P	W	G	E	Y	D	A	K	F	G	
Ph1	A	V	A	A	Y	R	Y	D	W	T	T	P	L	D	D	M	G	G	L	Y	W	Y	P	W	G	E	Y	D	A	K	F	G	
Ph2	M	I	A	S	Y	R	Y	D	W	T	T	P	L	D	D	M	G	A	Y	W	W	Y	P	W	G	E	Y	D	A	K	F	G	
NR	?	M	I	A	S	Y	R	Y	D	W	T	T	P	L	E	D	M	G	G	Y	W	W	Y	P	W	G	E	Y	D	A	K	F	G
HsHR	A	I	S	S	Y	R	Y	T	W	S	T	P	I	A	D	M	G	S	F	S	W	Y	P	W	E	T	Y	D	A	K	F	A	
NpHR	A	V	S	S	Y	R	Y	T	W	S	T	P	I	A	D	M	G	S	F	Y	W	Y	P	W	E	T	Y	D	A	K	F	A	
SrHR	A	I	S	S	Y	R	Y	T	W	S	T	P	I	A	D	M	G	S	F	Y	W	Y	P	W	E	T	Y	D	A	K	F	T	
HsSRI	L	I	A	S	Y	R	Y	D	W	T	T	P	L	Y	D	M	G	S	H	A	W	Y	P	W	A	V	Y	D	A	K	Y	V	
SrSRI	M	I	A	M	Y	R	Y	D	W	T	T	P	L	Y	D	M	G	S	H	A	W	Y	P	W	-	I	I	D	A	K	F	V	
HvSRI	F	I	A	S	Y	R	Y	D	W	T	T	P	L	F	D	M	G	S	H	A	W	Y	P	W	E	V	Y	D	A	K	Y	V	
HsSRII	M	I	A	A	Y	R	Y	D	W	T	T	P	I	Y	D	V	G	G	Y	Y	W	Y	P	W	A	Y	D	S	K	F	V		
NpSRII	M	I	A	A	Y	R	Y	D	W	T	T	P	I	F	N	V	G	G	F	Y	W	Y	P	W	P	D	I	D	T	K	F	G	
HvSRII	E	I	A	A	Y	R	Y	D	W	T	T	P	N	F	Q	T	G	G	F	Y	W	Y	P	W	A	A	V	D	T	K	F	G	
ASR	M	I	S	A	Y	R	Y	D	W	T	T	P	L	S	Q	V	G	G	F	W	W	Y	P	W	S	D	F	P	S	K	F	S	
CrChR1	S	V	E	K	F	R	Y	E	W	T	C	P	I	H	D	T	G	G	G	F	W	F	P	F	E	S	H	D	S	K	W	S	
VcChR1	S	V	E	K	S	R	Y	E	W	T	C	P	L	H	D	C	G	S	G	F	W	F	P	F	E	S	H	D	A	K	W	G	
MvChR1	S	V	E	I	Y	R	Y	E	W	T	C	P	I	A	D	C	G	G	G	W	W	F	P	F	E	S	H	D	S	K	W	G	
CrChR2	S	V	E	K	V	R	Y	E	W	T	C	P	I	H	D	T	G	G	G	F	W	F	P	F	E	S	H	D	S	K	W	G	
VcChR2	S	V	E	K	V	R	Y	E	W	T	C	P	I	H	D	T	G	G	G	F	W	F	P	F	E	S	H	D	S	K	W	G	

1) MED134, 2) AND4, 3) SAR11, 4) LC1-200, 5) HOT 75m4, 6) SAR86, 7) HOT 0m1

Supporting Information Figure 1. Conservation of the key amino acids for microbial rhodopsins shown in Figure 9. Despite the varying global sequence identities among microbial rhodopsins (Supporting Information Figure 2), a subset of residues is conserved in sequence and in position in the TM helices, and plays a role in activity and selectivity. The numbering scheme corresponds to the position in BR. Negatively charged residues are shown in red, positively charged in blue, aromatic in yellow, polar in green, non-polar in grey.

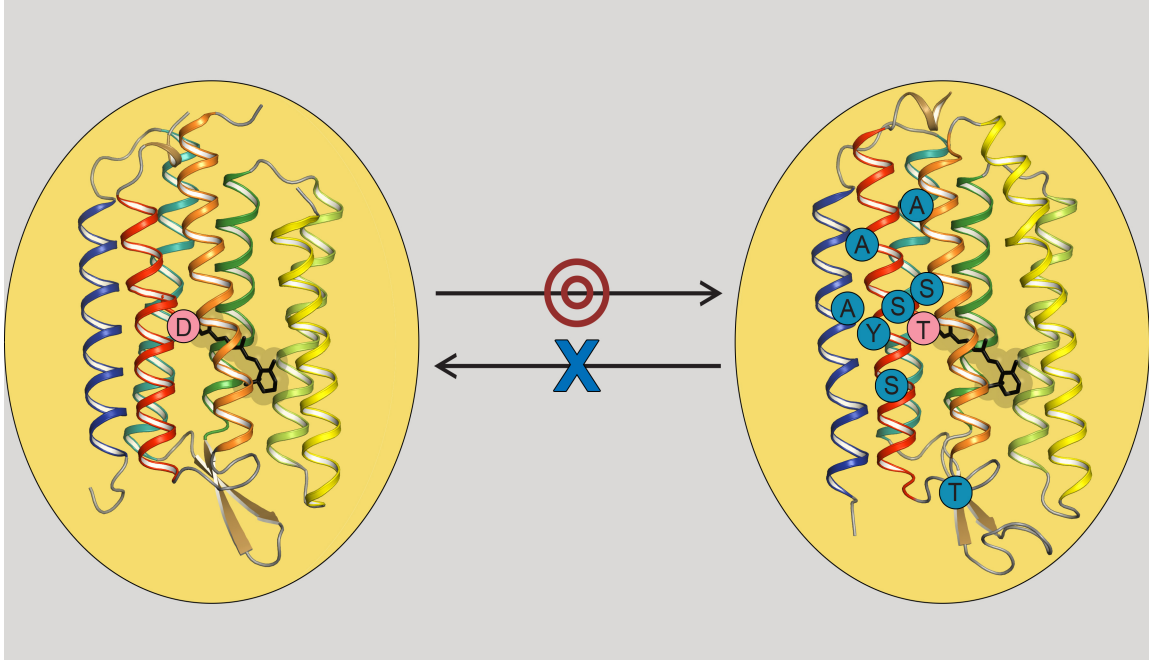
BR; Bacteriorhodopsin from Archaea (*Halobacterium salinarum*), AR1; Archaeorhodopsin-1 from Archaea (*Halorubrum chaoviator*), AR2; Archaeorhodopsin-2 from Archaea (*Halobacterium sp.* AUS-2), AR-3; Archaeorhodopsin-3 from Archaea (*Halorubrum sadomense*), MR; Midrhodopsin from Eubacteria (*Haloquadratum walsbyi* DSM 16790), GPR(1); Green-absorbing Proteorhodopsin from Eubacteria (*Dokdonia donghaensis* MED134; EAQ40507), GPR(2); Green-absorbing Proteorhodopsin from

Eubacteria (*Vibrio sp.* AND4; ZP_02194911), GPR(3); Green-absorbing Proteorhodopsin from Eubacteria (*Candidatus Pelagibacter ubique* HTCC1062, “SAR11” group; Q4FMZ3), BPR(4); Blue-absorbing Proteorhodopsin from Eubacteria (*Photobacterium sp.* LC1-200; BAL68143), BPR(5); Blue-absorbing Proteorhodopsin from uncultured bacterium (HOT 75m4; AAK30179), GPR(6); Green-absorbing Proteorhodopsin from uncultured marine gamma proteobacterium (“SAR86” group; AAG10475), GPR(7); Green-absorbing Proteorhodopsin from uncultured bacterium (HOT 0m1; AAK30176), GR; Gloeobacter Rhodopsin from Eubacteria (*Gloeobacter violaceus*), XR; Xanthorhodopsin from Eubacteria (*Salinibacter ruber* DSM 13855), Ace1; Acetabularia Rhodopsin 1 from Eukaryote (*Acetabularia acetabulum*), Ace2; Acetabularia Rhodopsin 2 from Eukaryote (*Acetabularia acetabulum*), LR; Leptosphaeria Rhodopsin from Eukaryote (*Leptosphaeria maculans*), Ph1; Phaeosphaeria Rhodopsin 1 from Eucaryote (*Phaeosphaeria nodorum* SN15), Ph2; Phaeosphaeria Rhodopsin 2 from Eukaryote (*Phaeosphaeria nodorum* SN15), NR; Neurospora Rhodopsin from Eukaryote (*Neurospora crassa* OR74A), HsHR; Halorhodopsin from Archaea (*Halobacterium salinarum*), NpHR; Halorhodopsin from Archaea (*Natronomonas pharaonis*), SrHR; Halorhodopsin from Eubacteria (*Salinibacter ruber* DSM 13855), HsSRI; Sensory Rhodopsin I from Archaea (*Halobacterium salinarum*), SrSRI; Sensory Rhodopsin I from Eubacteria (*Salinibacter ruber* M8), HvSRI; Sensory Rhodopsin I from Archaea (*Haloarcula vallismortis* ATCC 29715), HsSRII; Sensory Rhodopsin II from Archaea (*Halobacterium salinarum*), NpSRII; Sensory Rhodopsin II from Archaea (*Natronomonas pharaonis*), HvSRII; Sensory Rhodopsin II from Archaea (*Haloarcula vallismortis* ATCC 29715), ASR; Anabaena Sensory Rhodopsin from Eubacteria (*Nostoc sp.* PCC 7120), CrChR1; Channelrhodopsin 1 from Eukaryote (*Chlamydomonas reinhardtii*), VcChR1; Channelrhodopsin 1 from Eukaryote (*Volvox carteri f. nagariensis*), MvChR1; Channelrhodopsin 1 from Eukaryote (*Mesostigma viride*), CrChR2; Channelrhodopsin 2 from Eukaryote (*Chlamydomonas reinhardtii*), VcChR2; Channelrhodopsin 2 from Eukaryote (*Volvox carteri f. nagariensis*).

Protein	BR	AR1	AR2	AR3	GPR				BPR		GR	XR	LR	Ph1	Ph2	NR	Ace1	Ace2	HsHR	NpHR	SrHR	HsSRI	SrSRI	HvSRI	HsSRII	NpSRII	HvSRII	MR	ASR	CrChR1	VcChR1	MvChR1	CrChR2	VcChR2		
					SAR86	HOT_0m1	SAR11	MED134	Vibrio sp. AND4	HOT_75m4	LC1-200																									
BR		60	55	58	26	26	26	24	27	26	28	24	24	29	30	35	30	27	31	34	33	31	26	31	25	30	28	33	43	30	20	23	20	18	18	
AR1	60		85	92	25	26	24	25	25	26	22	25	26	32	32	31	29	28	32	31	34	30	26	32	28	31	30	35	42	32	24	26	21	21	20	
AR2	55	85		86	24	25	23	31	24	24	25	25	28	36	33	31	30	29	32	33	30	28	25	32	28	32	33	35	45	31	23	24	24	23	21	
AR3	58	92	86		26	26	24	25	25	25	22	24	29	32	31	30	29	30	32	32	30	28	27	31	29	33	31	34	42	33	24	23	21	23	21	
GPR	SAR86	26	25	24	26		97	51	40	54	79	57	29	28	28	25	30	26	27	23	26	20	20	17	18	21	18	26	27	26	23	19	20	18	17	17
	HOT_0m1	26	26	25	26	97		52	40	55	78	57	29	28	24	31	26	27	24	25	19	20	17	18	21	18	26	27	26	23	19	20	18	20	17	17
	SAR11	26	24	23	24	51	52		43	65	51	55	28	26	27	25	24	24	22	22	15	19	20	19	18	21	27	27	24	23	20	17	17	18	18	19
	MED134	24	25	31	25	40	40	43		43	39	41	32	29	24	23	22	27	21	24	22	22	18	21	19	25	25	28	28	26	21	16	18	16	17	18
	Vibrio sp. AND4	27	25	24	25	54	55	65	43		55	57	28	29	26	25	26	24	22	21	17	22	17	22	25	22	26	28	29	26	18	20	20	16	21	21
BPR	HOT_75m4	26	26	24	25	79	78	51	39	55		55	27	27	24	23	24	25	18	23	21	20	17	22	20	24	29	31	30	24	18	19	17	19	17	13
	LC1-200	28	22	25	22	57	57	55	41	57	55		28	27	22	27	25	30	19	23	21	20	17	21	25	20	27	28	30	24	20	20	18	19	17	17
GR	24	25	25	24	29	29	28	32	28	27	28		50	20	25	28	33	21	24	19	24	19	25	22	27	26	26	32	20	20	26	19	25	28	23	
XR	24	26	28	29	28	28	26	29	29	27	27	50		19	23	20	20	52	26	19	20	16	25	26	27	28	25	27	24	26	17	24	13	19	28	
LR	29	32	36	32	25	24	27	24	26	24	22	20	19		34	78	53	33	30	33	29	29	27	28	33	30	28	32	27	20	20	21	22	22		
Ph1	30	32	33	31	30	31	25	23	25	23	27	25	23	34		32	30	30	33	28	27	26	24	30	25	29	27	25	29	25	22	25	*	20	21	
Ph2	35	31	31	30	26	26	24	22	26	24	25	28	20	78	32		55	34	34	31	28	29	30	28	35	32	28	29	33	24	21	22	22	23	21	
NR	30	29	30	29	27	27	24	27	24	25	30	33	20	53	30	55		29	32	31	31	29	25	23	30	31	27	27	29	25	20	23	21	25	22	
Ace1	27	28	29	30	23	24	22	21	22	18	19	21	52	33	30	34	29		54	24	29	23	31	26	29	31	25	25	28	29	22	21	20	25	22	
Ace2	31	32	32	32	26	25	22	24	21	23	23	24	26	30	33	34	32	54		25	27	25	28	30	28	27	29	27	28	27	22	21	22	20	22	
HsHR	34	31	33	32	20	19	15	22	17	21	21	19	19	33	28	31	31	24	25		55	56	28	29	27	31	28	30	30	22	17	23	21	20	18	
NpHR	33	34	30	30	20	20	19	22	22	20	20	24	20	29	27	28	31	29	27	55		58	23	24	27	29	29	32	29	22	23	22	24	22	22	
SrHR	31	30	28	28	16	17	20	18	17	17	17	19	16	29	26	29	29	23	25	56	58		22	23	30	31	27	30	32	22	24	22	20	24	23	
HsSRI	26	26	25	27	18	18	19	21	22	22	21	25	25	27	24	30	25	31	28	28	23	22		41	69	33	30	30	25	25	25	22	22	20	23	
SrSRI	31	32	32	31	21	21	18	19	25	20	25	22	26	28	30	28	23	26	30	29	24	23	41		40	33	30	33	32	26	25	26	19	23	21	
HvSRI	25	28	28	29	18	18	21	25	22	24	20	27	27	33	25	35	30	29	28	27	27	30	69	40		35	32	30	27	24	25	25	21	23	22	
HsSRII	30	31	32	33	26	26	27	25	26	29	27	26	28	30	29	32	31	31	27	31	29	31	33	33	35		38	36	36	27	29	27	26	26	30	
NpSRII	28	30	33	31	27	27	28	28	31	28	26	25	30	27	28	27	25	29	28	29	27	30	30	30	32	38		43	38	28	21	23	24	20	21	
HvSRII	33	35	35	34	26	26	24	28	29	30	30	26	27	28	25	29	27	25	27	30	29	30	30	33	30	36	43		34	25	23	25	25	25	24	
MR	43	42	45	42	23	23	23	26	26	24	24	32	24	32	29	33	29	28	28	30	32	32	25	32	27	36	38	34		29	21	19	20	21	21	
ASR	30	32	31	33	19	19	20	21	18	18	20	20	26	27	25	24	25	29	27	22	29	22	25	26	24	27	28	25	29		24	19	23	21	23	
CrChR1	20	24	23	24	20	20	17	16	20	19	20	20	17	25	22	31	20	22	22	17	22	24	25	25	25	29	21	23	21	24		58	32	54	55	
VcChR1	23	26	24	23	18	18	17	18	20	17	20	26	24	20	25	22	23	21	21	23	23	22	22	26	25	27	23	25	21	19		58		29	54	52
MvChR1	20	21	24	21	20	20	18	16	16	19	18	19	14	20	*	22	20	20	22	21	22	20	22	19	21	26	24	25	19	23		32	29	30	45	
CrChR2	18	21	23	23	18	17	18	17	21	17	19	25	19	21	20	23	25	25	20	20	24	24	20	23	23	26	20	25	20	21		54	54	30	54	
VcChR2	18	20	21	21	17	17	19	18	21	13	17	28	23	22	21	22	22	22	21	18	22	23	23	21	22	30	21	24	21	23		55	52	45	54	



Supporting Information Figure 2. Pairwise sequence identities for the selected microbial rhodopsins from Figure 9. Despite their similarity of topology and conservation of the key intramembrane residues, especially in the retinal binding pocket, global sequence homology amongst the microbial rhodopsins can be as low as ~15%.



Supporting Information Figure 3. Asymmetric functional conversion between light-driven proton and chloride pumps. Light-driven proton pump BR can be converted into a chloride pump BR by a single amino acid replacement (Asp-to-Thr mutation). In contrast, substituting the BR-like ten amino acids into HR creates a NpHR mutant which is not capable of pumping protons. FTIR spectroscopy revealed that the failure of functional conversion of HR is due to the lack of strongly hydrogen-bonded water(s), the functional determinant of proton pumping.^{160,301}

the A_{2A}-adenosine and PAR1 receptor structures (PDB ID: 4EIY, 3VW7) in a similar position occupied by water in Rho. TM helices are connected by cytoplasmic (CL1-CL3) and extracellular (EL1-EL3) loops. Asn2 and Asn15 are glycosylated (not shown). Cys322 and Cys323 at end of H8 are palmitoylated (not shown). Cys110 on TM3 and Cys187 in EL2 are linked by a disulfide bridge. Residues Asn55^{1.50}, Asp83^{2.50}, Arg135^{3.50}, Trp161^{4.50}, Pro215^{5.50}, Pro267^{6.50} and Pro303^{7.50} are most conserved residues in a TM helix. The superscript relates to Ballesteros–Weinstein numbering (cf. ref 521). Conserved motifs are the (D/E)RY motif in TM3 (Glu134-Arg135-Tyr136) and the NPxxY(x)_{5,6}F motif in TM7 and H8 (Asn302-Pro303-x-x-Tyr306-x_{5,6}-Phe313). The inverse agonist 11-*cis* retinal is attached to Lys296^{7.43}. Residues Tyr223 (Y^{5.58}) and Lys231 (K/R^{5.66}) of the Y^{5.58}(x)₇K(R)^{5.66} motif in TM5 are microswitches for stabilizing the active receptor state.⁵²¹