ELSEVIER

Biochimica et Biophysica Acta 1511 (2001) 206–223

BIOCHIMICA ET BIOPHYSICA ACTA

BBA

www.bba-direct.com

Review

Homologues of archaeal rhodopsins in plants, animals and fungi: structural and functional predications for a putative fungal chaperone protein

Yufeng Zhai [1], Wilbert H.M. Heijne [1], Douglas W. Smith, Milton H. Saier Jr. *

Department of Biology, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116, USA

Received 1 September 2000; received in revised form 12 December 2000; accepted 12 December 2000

Abstract

The microbial rhodopsins (MR) are homologous to putative chaperone and retinal-binding proteins of fungi. These proteins comprise a coherent family that we have termed the MR family. We have used modeling techniques to predict the structure of one of the putative yeast chaperone proteins, YRO2, based on homology with bacteriorhodopsins (BR). Availability of the structure allowed depiction of conserved residues that are likely to be of functional significance. The results lead us to predict an extracellular protein folding function and a transmembrane proton transport pathway. We suggest that protein folding is energized by a novel mechanism involving the proton motive force. We further show that MR family proteins are distantly related to a family of fungal, animal and plant proteins that include the human lysosomal cystine transporter (LCT) of man (cystinosin), mutations in which cause cystinosis. Sequence and phylogenetic analyses of both the MR family and the LCT family are reported. Proteins in both families are of the same approximate size, exhibit seven putative transmembrane α-helical spanners (TMSs) and show limited sequence similarity. We show that the LCT family arose by an internal gene duplication event and that TMSs 1–3 are homologous to TMSs 5–7. Although the same could not be demonstrated statistically for MR family members, homology with the LCT family suggests (but does not prove) a common evolutionary pathway. Thus, TMSs 1–3 and 5–7 in both LCT and MR family members may share a common origin, accounting for their shared structural features.

Keywords: Bacterio(archaeal)rhodopsin; Fungal chaperone; Nephropathic cystinosis; Cystinosin; Evolution; Molecular modeling

1. Introduction

In 1994 we published a phylogenetic analysis of a small family of archaeal rhodopsins, then referred to as the bacteriorhodopsin or BR family since the best characterized member is bacteriorhodopsin of *Halobacterium halobium* (now called *Halobacterium salinarum*) [1]. The family included five proton-translocating bacteriorhodopsins (BRs), four chloride-translocating halorhodopsins (HRs) and two photoreceptor proteins called sensory rhodopsins (SRI and SRII) [2–6]. Homology of all of these proteins was established, a sequence alignment revealed the conserved residues, an average hydropathy analysis revealed the relative degrees of hydrophobicity of the seven established transmembrane α-helical segments

* Corresponding author. Fax: 858-534-7108; E-mail: msaier@ucsd.edu

[1] These two authors contributed equally to the reported work.

0005-2736/01/$ – see front matter
PII: S0005-2736(00)00389-8

Table 1
Proteins of the microbial rhodopsin (MR) family included in this study[a]

Abbreviation	Name or description in database	Organism	Size (no. residues)	Database and accession No.[c]	gi No.[d]
Fungal chaperones (FC)					
HSP30	Heat shock protein 30 kDa	*Saccharomyces cerevisiae*	332	spP25619	140468
YRO2	YRO2 (YBR054w)	*Saccharomyces cerevisiae*	344	spP38079	586913
YDR033w	YDR033w	*Saccharomyces cerevisiae*	320	pirS61586	2132470
Spo	hypothetical protein	*Schizosaccharomyces pombe*	306	gbCAA21219	3702626
Cve	FDD123b	*Coriolus versicolor*	283	gbBAA76589	4587023
Bacteriorhodopsins; H^+ pumps (BR)					
Hsa	Bacteriorhodopsin precursor	*Halobacterium salinarum* (*halobium*)	262	spP02945	114811
Hva	Cruxrhodopsin-3 (COP-3) (CR-3)	*Haloarcula vallismortis*	250	spP94854	2829812
Har	Cruxrhodopsin-1 (COP-1) (CR-1)	*Haloarcula argentinensis*	250	spQ57101	2499386
Hmu	Cruxrhodopsin-2 (COP-2) (CR-2)	*Haloarcula mukohataei* sp. ARG-2	255	spQ53496	2499387
Hso	Archaearhodopsin-3 precursor (AR-3)	*Halorubrum sodomense*	258	spP96787	3023375
Hr2	Archaearhodopsin-2, retinal protein	*Halobacterium* sp. aus-2	259	gbAAB19870	235918
Ht4	Bacteriorhodopsin	*Haloterrigena* sp. ARG-4	250	gbBAA75200	4579714
HSG	Archaearhodopsin-1 precursor (AR-1)	*Halobacterium* sp. strain SG	260	spP19585	114807
Hme	Bacteriorhodopsin	*Halobacterium* strain mex	part: 220[b]	spP33969	461610
Hpo	Bacteriorhodopsin	*Halobacterium* strain port	part: 227[b]	spP33971	461611
Hsh*	Bacteriorhodopsin	*Halobacterium* strain shark	part: 209[b]	spP33972	461612
Halorhodopsins; Cl^- pumps (HR)					
Hsa	Halorhodopsin precursor	*Halobacterium salinarum* (*halobium*)	274	spP16102	114808
Hva	Cruxhalorhodopsin-3 precursor (CHR-3)	*Haloarcula vallismortis*	276	spP94853	2829811
Hso	Halorhodopsin	*Halorubrum sodomense*	282	gbBAA75202	4579718
Nph	Halorhodopsin	*Natromonas pharaonis*	291	spP15647	114809
Ht4	Halorhodopsin	*Haloterrigena* sp. ARG-4	297	gbBAA75201	4579716
HSG	Halorhodopsin	*Halobacterium* sp. strain SG	284	spP33742	461609
Hpo	Halorhodopsin precursor	*Halobacterium* strain port	276	spQ48315	2499383
Hsh	Halorhodopsin precursor	*Halobacterium* strain shark	276	spQ48314	2499384
Hme	Halorhodopsin	*Halobacterium* strain mex	part: 206[b]	spP33970	461608
Sensory rhodopsins I (SRI)					
Hsa	Sensory rhodopsins I	*Halobacterium salinarum* (*halobium*)	239	spP25964	114812
Hva	Bacterial rhodopsin CSR3	*Haloarcula vallismortis*	236	spQ48334	2499388
Hso	Sensory rhodopsin	*Halorubrum sodomense*	254	gbBAA75203	4579720
HSG	Sensory rhodopsin I	*Halobacterium* sp. strain SG	247	spP33743	461613
Sensory rhodopsins II; phoborhodopsins (SRII)					
Hsa	Sensory rhodopsin II	*Halobacterium salinarum* (*halobium*)	237	spP71411	2499389
Hva	Sensory rhodopsin II	*Haloarcula vallismortis*	236	spP42197	1168614
Nph	Sensory rhodopsin II	*Natromonas pharaonis*	239	spP42196	1168615

[a]Members missing a relevant part of the sequence (BR Hsh*), and the recently sequenced BR from *Haloarcula japonica* (three amino acid residues different from the *Haloarcula argentinensis* protein) were not included in this analysis. Additionally, the sequence published by Bieszke et al. [17,18] appeared after the completion of this study and consequently was not included.
[b]Part, partial sequence available.
[c]Databases included: sp, SwissProt; gb, GenBank; pir, Protein Information Resource.
[d]gi No., gene identification number of the National Center for Biotechnology Information (NCBI).

(TMSs), and an average similarity plot revealed that the TMSs were better conserved than the inter-TMS loops [1]. Moreover, a phylogenetic tree showed that the BRs clustered separately from the HRs, and that both of these two clusters were distant from the SRs [1]. Availability of high resolution three-dimensional structures of BR has allowed detailed mechanistic proposals that are supported by extensive experimentation [4,7–9]. Further, the high resolution structure of halorhodopsin has recently been reported [10].

Ihara et al. [11] confirmed the phylogenetic observations of Kuan and Saier [1] and extended the analyses to several newly sequenced proteins, all of which fell into the three groups (BR, HR and SR) observed previously. The family was expanded from 11 members to 25 members, all of which are archaeal retinal-containing chromophoric proteins. Ihara et al. [11], however, failed to note that the archaeal retinal-containing chromophoric proteins are, in fact, homologous to yeast and fungal proteins that apparently lack a chromophore and have been shown to be stress-induced [12–15]. These proteins appear to function in cellular responses to acid, organic solvent and heat stress signals [16]. Further, after completion of the analyses reported here, Bieszke et al. [17,18] sequenced a *Neurospora crassa* protein that encodes a retinal-binding homologue of the archaeal rhodopsins, and Béjà et al. [19] identified a retinal-containing, light-driven, proton-pumping homologue of the archaeal rhodopsins in a bacterium. It seems clear that these proteins are widespread in nature.

In this communication we present analyses of the fungal proteins that reveal their sequence and phylogenetic relationships to the three previously identified clusters of archaeal rhodopsins. Because the fungal proteins clearly share a common evolutionary origin with the archaeal and bacterial rhodopsins, we have renamed the family the microbial rhodopsin (MR) family (TC No. 3.E.1 [20,21]).

There is currently little information about the structures of the putative fungal (yeast) chaperone proteins (FCs) or about the functional relationships between FCs and the archaeal rhodopsins (ARs). In this paper, we predict the three-dimensional structure of one of the FC proteins based on the crystal structure of BR, depict the positions of potential functional residues in this FC protein and propose novel function/structure relationships between these two classes of proteins. We show further that the fungal proteins are homologues of, but distantly related to, a family of proteins in animals, plants, and fungi, one of which, the human disease protein, cystinosin, has been implicated in cystine transport across lysosomal membranes [22–27]. We have called this eukaryotic protein family the lysosomal cystine transporter (LCT) family (TC No. 2.A.43 [20,21]), named after its only functionally characterized member. We propose that all of the members of this eukaryotic-specific family are localized to intracellular organelles. Properties of LCT family members are described and compared with those of the MR family proteins. These studies have allowed us to suggest an evolutionary pathway by which the structurally well characterized archaeal rhodopsins arose.

2. Materials and methods

Pairwise sequence alignments were generated using the Genetics Computer Group (GCG) GAP program [28]. Standard deviation (S.D.) units were calculated with 500 random shuffles for statistical analysis. Multiple sequence alignments, with bootstrapping, were generated using the Clustal X program [29,60] and the TREE program [30], and confirmed with other programs [31,32]. Similarity determinations using other computational tools also revealed a common origin for bacteriorhodopsins and the yeast proteins [33]. Average hydropathy, average similarity, and average amphipathicity plots of the proteins in each of the four clusters of the MR family were based on multiple sequence alignments generated with the TREE program of Feng and Doolittle [30]. In all such plots, a sliding window of 21 residues was used. Hydropathy values were those of Kyte and Doolittle [34]. For the average amphipathicity plot, an angle of 100° was used as appropriate for an α-helix as described using the program of Le et al. [35].

The fungal chaperone protein selected for three-dimensional modeling studies is YRO2 from *Saccharomyces cerevisiae* (see Table 1). YRO2 is demonstrably homologous to nephropathic cystinosin and other members of the LCT family (comparison scores

using the GCG GAP program with 500 random shuffles in excess of 9 S.D.).

BR coordinates were obtained from Protein Data Bank (PDB) entry 1QHJ [36] which was submitted in 1999 at a resolution of 1.9 Å. Results obtained with this structure were fully in agreement with those based on the recent 1.55 Å resolution structure published by Luecke et al. [37]. The fungal chaperone structural determination was based on the structure of BR using the Molecular Simulation, Inc. (MSI) 3-D protein modeling program Insight II. The initial structure was refined by energy minimization (ME) and molecular dynamics (MD) using the Insight II program Discover (Version 98.0, MSI). Simulations were performed in vacuo with a dielectric constant of 4 using the consistent valence forcefield (CVFF). The protein was kept neutral during the structural refinement procedure. All of the ionizable amino acyl residues were maintained in their unionized forms. Structural refinement procedures were as follows: (1) poor contacts were eliminated by a 100 step steepest descents ME followed by a 1000 step conjugated gradients ME; (2) MD was performed at 600 K for 10 ps selecting a local minimum structure near the end of the trajectory to optimize by 1000 step conjugated gradients ME; (3) steps (1) and (2) were repeated at lower temperature intervals of 100 K to a final value of 300 K; and (4) the final structure was optimized with conjugated gradient ME and a convergence criterion of 0.05 kcal/mol.

Without constraints, the structure would be destined for denaturation because the ME and MD simulations were not conducted under native conditions in the presence of a lipid bilayer, an aqueous solution, and other potential factors. To prevent the helix bundle from disintegrating at higher temperatures, the polypeptide backbone atoms were subjected to a harmonic forcefield during the entire procedure of the refinement. The force constant was reduced gradually from the first to the last MD run. A hydrogen bond restraint was also used in the modeling calculations. Relatively strong distance restraints were imposed to maintain proper geometry for *i*, *i*+4 hydrogen bonds. The distances between the carbonyl oxygen of residues *i* and the backbone nitrogen of the amide protons of residues *i*+4 were constrained to a range of 2.7–3.2 Å and 1.8–2.3 Å, respectively, for the segments of an α-helix.

3. Results and discussion

3.1. The MR family (TC No. 3.E.1)

Table 1 presents the members of the MR family, subdivided into four groups: the BRs, the HRs, the SRs and the fungal/yeast putative FCs. One of the FC proteins, heat shock protein Hsp30, may function in energy conservation, downregulating stress-activation of the plasma membrane proton-ATPase so that this enzyme does not deplete the energy reserves of the cell [38]. Some of the fungal proteins are not currently believed to be retinal-binding proteins as several of the residues conserved in the archaeal proteins that are localized to the retinal-binding pocket are not conserved in the fungal proteins. In the analyses reported, all family members where full sequences were available at the time these studies were conducted are included with the exception of the *Haloarcula japonica* BR that differs from the *Haloarcula argentinensis* protein by only three amino acyl residues. Table 1 presents the protein abbreviations used in this report, the names or descriptions provided in the databases, the source organisms, and the sizes of the proteins in numbers of amino acyl residues. Database accession numbers and gi numbers allow easy web access to the sequences.

A partial multiple alignment of the protein members of the MR family is presented in Fig. 1. Four residues are fully conserved in all proteins including the fungal chaperones, and many additional residues are well conserved in all but one or a few proteins. It is important to note, however, that the lysyl residue (e.g., BR hso K226) in TMS 7 that provides the Schiff's base linkage to retinal is not retained in the fungal proteins. Moreover, some of the residues conserved in the archaeal proteins that are localized to the retinal-binding pocket are also not conserved in the fungal proteins [4,7,8,11,39].

Statistical analyses revealed that the FCs are homologous to the archaeal proteins (comparison scores of >10 S.D. using the GCG GAP program with 100 random shuffles) [28]. A considerable degree of sequence similarity is also evident from the alignment shown in Fig. 1. Fig. 2 shows average hydropathy (H, dark solid lines), average similarity (S, dotted lines) and average amphipathicity (A, light dashed lines) for the four clusters of proteins (A,

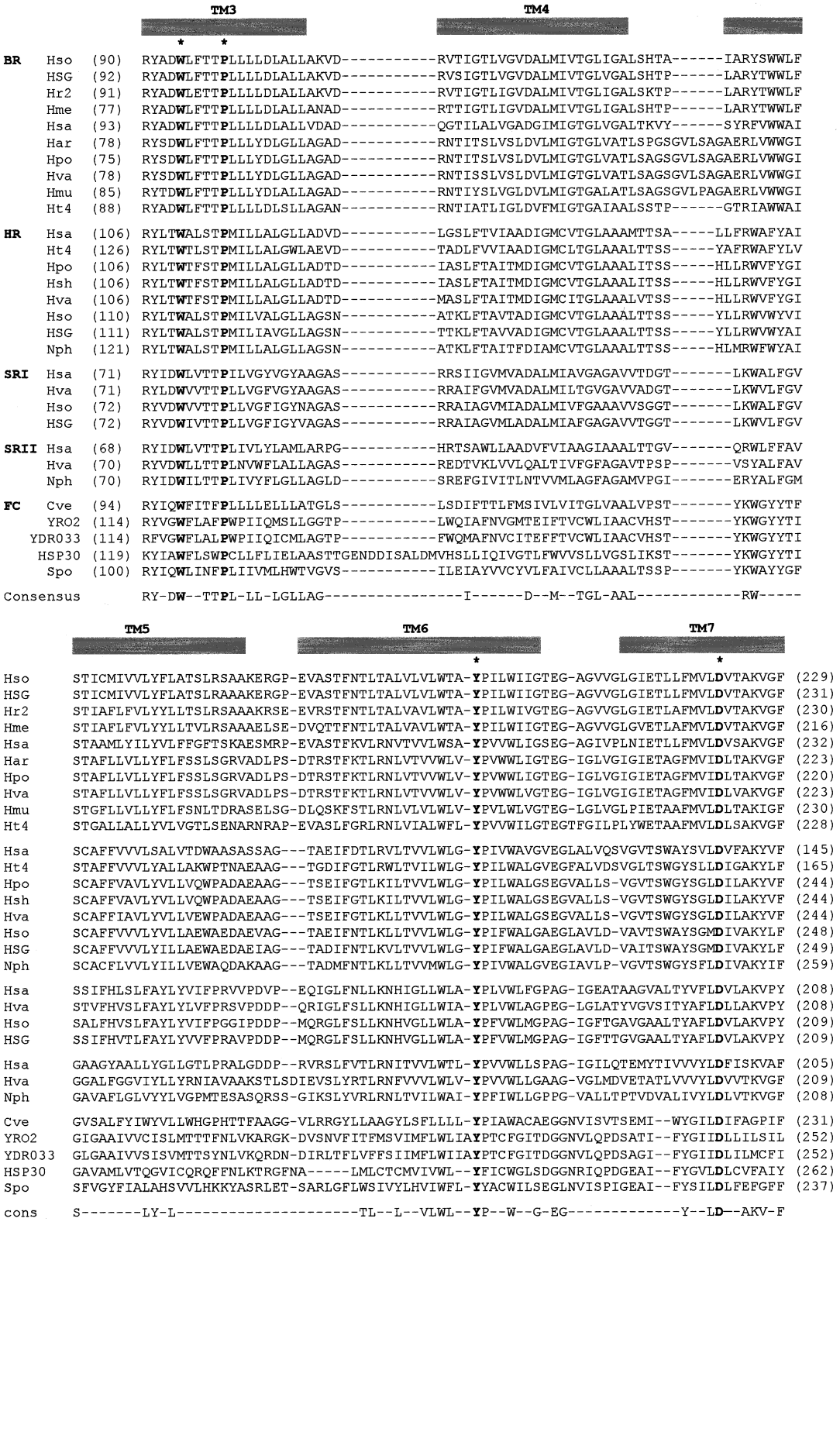
TM3
TM4
TM5
TM6
TM7
BR Hso (90) RYADWLFTTPLLLLDLALLAKVD----------RVTIGTLVGVDALMIVTGLIGALSHTA------IARYSWWLF
HSG (92) RYADWLFTTPLLLLDLALLAKVD----------RVSIGTLVGVDALMIVTGLVGALSHTP------LARYTWWLF
Hr2 (91) RYADWLETTPLLLLDLALLAKVD----------RVTIGTLIGVDALMIVTGLIGALSKTP------LARYTWWLF
Hme (77) RYADWLFTTPLLLLDLALLANAD----------RTTIGTLIGVDALMIVTGLIGALSHTP------LARYTWWLF
Hsa (93) RYADWLFTTPLLLLDLALLVDAD----------QGTILALVGADGIMIGTGLVGALTKVY------SYRFVWWAI
Har (78) RYSDWLFTTPLLLYDLGLLAGAD----------RNTITSLVSLDVLMIGTGLVATLSPGSGVLSAGAERLVWWGI
Hpo (75) RYSDWLFTTPLLLYDLGLLAGAD----------RNTITSLVSLDVLMIGTGLVATLSAGSGVLSAGAERLVWWGI
Hva (78) RYSDWLFTTPLLLYDLGLLAGAD----------RNTISSLVSLDVLMIGTGLVATLSAGSGVLSAGAERLVWWGI
Hmu (85) RYTDWLFTTPLLLYDLALLAGAD----------RNTIYSLVGLDVLMIGTGALATLSAGSGVLPAGAERLVWWGI
Ht4 (88) RYADWLFTTPLLLLDLSLLAGAN----------RNTIATLIGLDVFMIGTGAIAALSSTP------GTRIAWWAI
HR Hsa (106) RYLTWALSTPMILLALGLLADVD----------LGSLFTVIAADIGMCVTGLAAAMTTSA-----LLFRWAFYAI
Ht4 (126) RYLTWTLSTPMILLALGWLAEVD----------TADLFVVIAADIGMCLTGLAAALTTSS-----YAFRWAFYLV
Hpo (106) RYLTWTFSTPMILLALGLLADTD----------IASLFTAITMDIGMCVTGLAAALITSS-----HLLRWVFYGI
Hsh (106) RYLTWTFSTPMILLALGLLADTD----------IASLFTAITMDIGMCVTGLAAALITSS-----HLLRWVFYGI
Hva (106) RYLTWTFSTPMILLALGLLADTD----------MASLFTAITMDIGMCITGLAAALVTSS-----HLLRWVFYGI
Hso (110) RYLTWALSTPMILVALGLLAGSN----------ATKLFTAVTADIGMCVTGLAAALTTSS-----YLLRWVWYVI
HSG (111) RYLTWALSTPMILIAVGLLAGSN----------TTKLFTAVVADIGMCVTGLAAALTTSS-----YLLRWVWYAI
Nph (121) RYLTWALSTPMILLALGLLAGSN----------ATKLFTAITFDIAMCVTGLAAALTTSS-----HLMRWFWYAI
SRI Hsa (71) RYIDWLVTTPILVGYVGYAAGAS----------RRSIIGVMVADALMIAVGAGAVVTDGT-------LKWALFGV
Hva (71) RYLDWVVTTPLLVGFVGYAAGAS----------RRAIFGVMVADALMILTGVGAVVADGT-------LKWVLFGV
Hso (72) RYVDWVVTTPLLVGFIGYNAGAS----------RRAIAGVMIADALMIVFGAAAVVSGGT-------LKWALFGV
HSG (72) RYVDWIVTTPLLVGFIGYVAGAS----------RRAIAGVMLADALMIAFGAGAVVTGGT-------LKWVLFGV
SRII Hsa (68) RYIDWLVTTPLIVLYLAMLARPG----------HRTSAWLLAADVFVIAAGIAAALTTGV-------QRWLFFAV
Hva (70) RYVDWLLTTPLNVWFLALLAGAS----------REDTVKLVVLQALTIVFGFAGAVTPSP-------VSYALFAV
Nph (70) RYIDWILTTPLIVYFLGLLAGLD----------SREFGIVITLNTVVMLAGFAGAMVPGI-------ERYALFGM
FC Cve (94) RYIQWFITFPLLLLELLLATGLS----------LSDIFTTLFMSIVLVITGLVAALVPST-------YKWGYYTF
YRO2 (114) RYVGWFLAFPWPIIQMSLLGGTP----------LWQIAFNVGMTEIFTVCWLIAACVHST-------YKWGYYTI
YDR033 (114) RFVGWFLALPWPIIQICMLAGTP----------FWQMAFNVCITEFFTVCWLIAACVHST-------YKWGYYTI
HSP30 (119) KYIAWFLSWPCLLFLIELAASTTGENDDISALDMVHSLLIQIVGTLFWVVSLLVGSLIKST-------YKWGYYTI
Spo (100) RYIQWLINFPLIIVMLHWTVGVS----------ILEIAYVVCYVLFAIVCLLAAALTSSP-------YKWAYYGF
Consensus RY-DW--TTPL-LL-LGLLAG----------I------D--M--TGL-AAL----------RW-----
Hso STICMIVVLYFLATSLRSAAKERGP-EVASTFNTLTALVLVLWTA-YPILWIIGTEG-AGVVGLGIETLLFMVLDVTAKVGF (229)
HSG STICMIVVLYFLATSLRAAAKERGP-EVASTFNTLTALVLVLWTA-YPILWIIGTEG-AGVVGLGIETLLFMVLDVTAKVGF (231)
Hr2 STIAFLFVLYYLLTSLRSAAAKRSE-EVRSTFNTLTALVAVLWTA-YPILWIVGTEG-AGVVGLGIETLAFMVLDVTAKVGF (230)
Hme STIAFLFVLYYLLTVLRSAAAELSE-DVQTTFNTLTALVAVLWTA-YPILWIIGTEG-AGVVGLGVETLAFMVLDVTAKVGF (216)
Hsa STAAMLYILYVLFFGFTSKAESMRP-EVASTFKVLRNVTVVLWSA-YPVVWLIGSEG-AGIVPLNIETLLFMVLDVSAKVGF (232)
Har STAFLLVLLYFLFSSLSGRVADLPS-DTRSTFKTLRNLVTVVWLV-YPVWWLIGTEG-IGLVGIGIETAGFMVIDLTAKVGF (223)
Hpo STAFLLVLLYFLFSSLSGRVADLPS-DTRSTFKTLRNLVTVVWLV-YPVWWLIGTEG-IGLVGIGIETAGFMVIDLTAKVGF (220)
Hva STAFLLVLLYFLFSSLSGRVADLPS-DTRSTFKTLRNLVTVVWLV-YPVWWLVGTEG-IGLVGIGIETAGFMVIDLVAKVGF (223)
Hmu STGFLLVLLYFLFSNLTDRASELSG-DLQSKFSTLRNLVLVLWLV-YPVLWLVGTEG-LGLVGLPIETAAFMVLDLTAKIGF (230)
Ht4 STGALLALLYVLVGTLSENARNRAP-EVASLFGRLRNLVIALWFL-YPVVWILGTEGTFGILPLYWETAAFMVLDLSAKVGF (228)
Hsa SCAFFVVVLSALVTDWAASASSAG---TAEIFDTLRVLTVVLWLG-YPIVWAVGVEGLALVQSVGVTSWAYSVLDVFAKYVF (145)
Ht4 STAFFVVVLYALLAKWPTNAEAAG---TGDIFGTLRWLTVILWLG-YPILWALGVEGFALVDSVGLTSWGYSLLDIGAKYLF (165)
Hpo SCAFFVAVLYVLLVQWPADAEAAG---TSEIFGTLKILTVVLWLG-YPILWALGSEGVALLS-VGVTSWGYSGLDILAKYVF (244)
Hsh SCAFFVAVLYVLLVQWPADAEAAG---TSEIFGTLKILTVVLWLG-YPILWALGSEGVALLS-VGVTSWGYSGLDILAKYVF (244)
Hva SCAFFIAVLYVLLVEWPADAEAAG---TSEIFGTLKLLTVVLWLG-YPILWALGSEGVALLS-VGVTSWGYSGLDILAKYVF (244)
Hso SCAFFVVVLYVLLAEWAEDAEVAG---TAEIFNTLKLLTVVLWLG-YPIFWALGAEGLAVLD-VAVTSWAYSGMDIVAKYLF (248)
HSG SCAFFVVVLYILLAEWAEDAEIAG---TADIFNTLKVLTVVLWLG-YPIFWALGAEGLAVLD-VAITSWAYSGMDIVAKYLF (249)
Nph SCACFLVVLYILLVEWAQDAKAAG---TADMFNTLKLLTVVMWLG-YPIVWALGVEGIAVLP-VGVTSWGYSFLDIVAKYIF (259)
Hsa SSIFHLSLFAYLYVIFPRVVPDVP--EQIGLFNLLKNHIGLLWLA-YPLVWLFGPAG-IGEATAAGVALTYVFLDVLAKVPY (208)
Hva STVFHVSLFAYLYLVFPRSVPDDP--QRIGLFSLLKNHIGLLWIA-YPLVWLAGPEG-LGLATYVGVSITYAFLDLLAKVPY (208)
Hso SALFHVSLFAYLYVIFPGGIPDDP--MQRGLFSLLKNHVGLLWLA-YPFVWLMGPAG-IGFTGAVGAALTYAFLDVLAKVPY (209)
HSG SSIFHVTLFAYLYVVFPRAVPDDP--MQRGLFSLLKNHVGLLWLA-YPFVWLMGPAG-IGFTTGVGAALTYAFLDVLAKVPY (209)
Hsa GAAGYAALLYGLLGTLPRALGDDP--RVRSLFVTLRNITVVLWTL-YPVVWLLSPAG-IGILQTEMYTIVVVYLDFISKVAF (205)
Hva GGALFGGVIYLLYRNIAVAAKSTLSDIEVSLYRTLRNFVVVLWLV-YPVVWLLGAAG-VGLMDVETATLVVVYLDVVTKVGF (209)
Nph GAVAFLGLVYYLVGPMTESASQRSS-GIKSLYVRLRNLTVILWAI-YPFIWLLGPPG-VALLTPTVDVALIVYLDLVTKVGF (208)
Cve GVSALFYIWYVLLWHGPHTTFAAGG-VLRRGYLLAAGYLSFLLLL-YPIAWACAEGGNVISVTSEMI--WYGILDIFAGPIF (231)
YRO2 GIGAAIVVCISLMTTTFNLVKARGK-DVSNVFITFMSVIMFLWLIAYPTCFGITDGGNVLQPDSATI--FYGIIDLLILSIL (252)
YDR033 GLGAAIVVSISVMTTSYNLVKQRDN-DIRLTFLVFFSIIMFLWIIAYPTCFGITDGGNVLQPDSAGI--FYGIIDLILMCFI (252)
HSP30 GAVAMLVTQGVICQRQFFNLKTRGFNA-----LMLCTCMVIVWL--YFICWGLSDGGNRIQPDGEAI--FYGVLDLCVFAIY (262)
Spo SFVGYFIALAHSVVLHKKYASRLET-SARLGFLWSIVYLHVIWFL-YYACWILSEGLNVISPIGEAI--FYSILDLFEFGFF (237)
cons S-------LY-L----------TL--L--VLWL--YP--W--G-EG----------Y--LD--AKV-F

Fig. 1. Partial multiple sequence alignment of the fungal chaperones (FCs) with established members of the archaeal rhodopsin family. The proteins (abbreviations as indicated in Table 1) are grouped from top to bottom as follows: BR, HR, SRI, SRII and FC. The numbers of the first and last residues in each protein are provided in parentheses. Fully conserved residues are presented in bold print while residues conserved in a majority of proteins appear in the consensus sequence at the bottom of the alignment. Horizontal numbered shaded bars indicate the positions of the TMSs in BR. The alignment shown and subsequent alignments presented in this paper were generated with the Clustal X program [29].
←

BRs; B, HRs; C, SRs; and D, FCs). The archaeal proteins all exhibit similar patterns of hydropathy with peaks 3, 4 and 5, and peaks 6 and 7 close to each other while peaks 1 and 2, 2 and 3, and 5 and 6 are more distant from each other. The observation that the N-terminus of BR is outside the cell while the C-terminus is inside [37] reveals that inter-TMS cytoplasmic loops are shorter than external loops. Moreover, peaks 3, 6 and 7 are generally less hydrophobic than peaks 1, 2, 4 or 5.

These features can be contrasted with those for the FCs. Although the relative spacing of peaks is similar to that found for the archaeal proteins, with the troughs between peaks 2 and 3, and peaks 5 and 6 being the deepest, definition of the peaks for the fungal proteins is clearly greater than that for the archaeal proteins. This fact reflects the greater loop sizes observed for the FCs, particularly for the external loops (see below).

The average similarity plots generally reveal comparable degrees of similarity throughout most of the archaeal proteins with greatest conservation in their TMSs. By contrast, the putative fungal chaperone proteins show distinct peaks of average similarity that do not correspond to the seven peaks of hydrophobicity. The first peak of average similarity precedes TMS 1; the second peak follows TMS 2; the third peak precedes TMS 3; the fourth peak is midway between TMSs 4 and 5; and the last major peak of average similarity is found between TMSs 6 and 7. Thus, the highest degree of conservation precedes odd numbered TMSs (1, 3, 5 and 7) while the lowest degree of conservation precedes even numbered TMSs (2, 4 and 6). According to the established topological model for BR, and assuming a similar topology for the FCs, external loops in the FC proteins are well conserved while internal loops in the FCs are poorly conserved. This observation leads to the suggestion that if these proteins function as chaperones, their activities involve the extracytoplasmic loops that together comprise the chaperone active sites. Regardless of the biochemical function(s) of these proteins, we predict that the conserved external inter-TMS loops, possibly together with the TMSs, provide the primary sites of action of these proteins.

Average amphipathicity plots [35] revealed another interesting feature that appears to be a characteristic of both the fungal and the archaeal proteins. The first (putative) TMS is always preceded by a large peak of amphipathicity when an angle of 100° is selected as is appropriate for an α-helix. This qualitative observation has been noted previously for BR [40] and may have biogenic significance [41,42]. Other peaks of amphipathicity are not as well conserved in the four groups of proteins represented in Fig. 2.

A phylogenetic tree for the MR family is shown in Fig. 3 (obtained using the Clustal X program which generates a neighbor joining tree [29]). A comparable tree was generated with the Phylip Protpars program that produces a parsimony tree ([31]; data not shown). Bootstrapping with 1000 replications was applied to the Clustal X tree. The closeness of the positions from which the four major branches emanate suggests that the gene duplication events that gave rise to the four functional types of MR family proteins (BR, HR, SR and FC) occurred at about the same time in evolutionary history [32]. Although previously derived phylogenetic trees for the MR family do not include the fungal proteins (see [17,18,33]), the portions of the trees that include archaeal proteins are in good agreement with those published previously ([1,11]; reviewed in [43]). The recent phylogenetic analyses reported by Bieszke et al. [17,18], demonstrating the presence of a retinal-binding protein member of the MR family in *N. crassa*, are in agreement with the results reported in Fig. 3.

3.2. Sequence conservation between BR and YRO2

Fig. 4 shows the sequence alignment for BR and YRO2 that was used in the homology modeling.

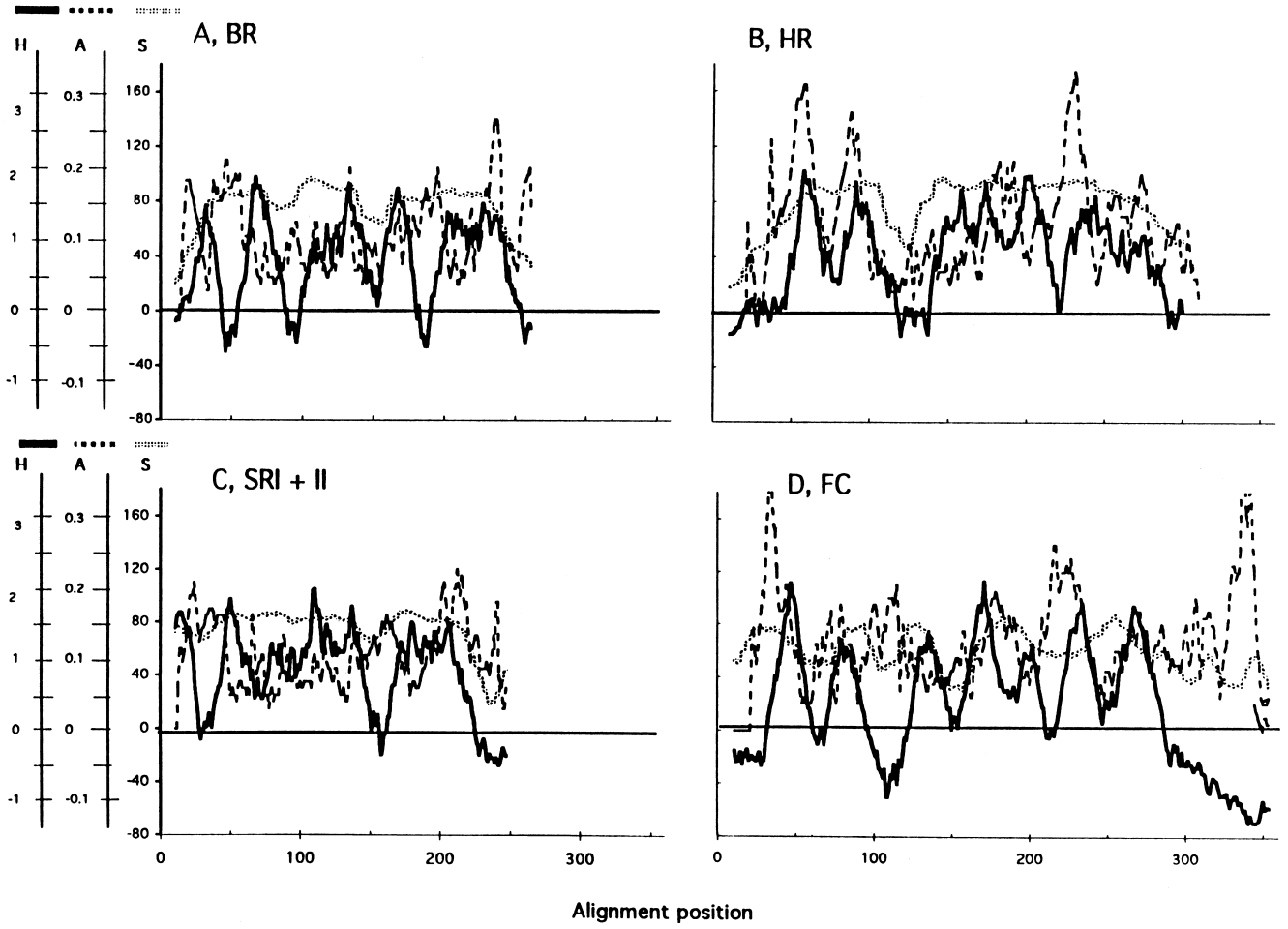

Fig. 2. Average hydropathy (———), similarity (• • •) and amphipathicity (– – –) for the four clusters of proteins in the MR family: A, BR; B, HR; C, SRI and II; D, FC. In all plots a sliding window of 21 residues was used. The programs for average hydropathy and average amphipathicity (angle of 100° for α-helix) have been described [34,35,53].

YRO2 was selected because of its sequence similarity with members of both the MR and LCT families. The sequence identity and similarity values of YRO2 with BR are 18% and 55%, respectively. Transmembrane helices 3–7 and the loops between them are well conserved with 63% similarity. Helices 1 and 2 and loops 1-2 and 2–3 are relatively poorly conserved. In order to check the reliability of the sequence alignment, we also aligned the sequences of YRO2-NOP-1 [17,18] and NOP1-BR, which have sequence identities of 27% and 33%, and sequence similarities of 66% and 68%, respectively. Thus, both BR and YRO2 exhibit far greater sequence similarity to NOP-1 than they do to each other. The combination of these two alignments gives the alignment shown in Fig. 4. We found from the alignments that almost all of the gaps are located in loop regions. One exception is an insertion of Ile in helix 6. This insertion, preceding the fully conserved Y, exists only in YRO2 and YDR033 and not in NOP-1 or the other fungal proteins (see Fig. 1).

3.3. General shape of the TM bundle

The YRO2 model maintains a seven-helix bundle closely resembling that of BR. In this and the subsequent sections of the text, the three letter abbreviations of the amino acids will be used for the fungal chaperone protein, YRO2, while the one letter abbreviations will be used for bacteriorhodopsin. The four fully conserved residues, Trp90, Pro95, Tyr189 and Asp215, are all located in the middle of the

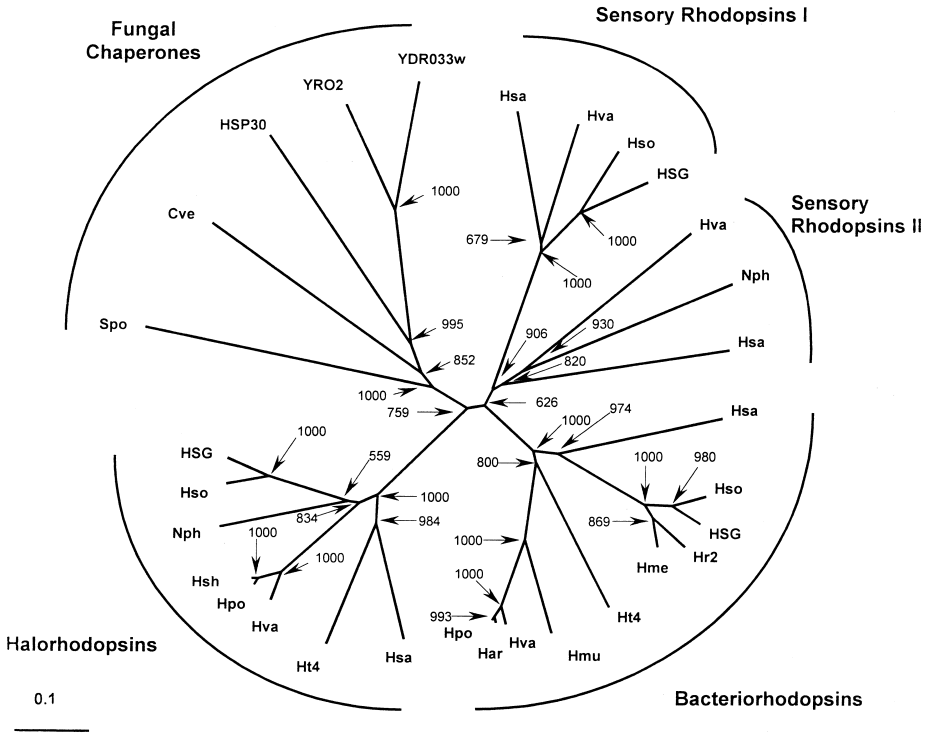

Fig. 3. Phylogenetic tree for the MR family based on the Clustal X program [29]. The results of bootstrapping, with 1000 replications, applied to the tree are provided at the nodes of the branches [31].

molecule as shown in Fig. 5 (residues in green color). Pro95, located in the middle of helix 3, causes this helix to bend. A similar structure exists in the predicted model for SRI [44]. The other three fully conserved amino acyl residues are in the retinal-binding pocket, are retinal-binding residues in BR, HR and SR, and are important for the functions of these proteins. At present, it is not known whether other FCs (except NOP-1) can bind a chromophore since the lysine which covalently binds retinal is not present in these proteins [18]. However, experiments have shown that the lysine is not a prerequisite for retinal binding [45–47]. In YRO2, the similarity of residues that compose the retinal-binding pocket is 57%. Considering the sequence conservation and the presence of the chromophore in the retinal-binding protein, NOP-1, we consider it possible that FCs bind retinal or a retinal analogue, bypassing the need for a Schiff's base linkage.

The other four residues that are fully conserved in all archaeal proteins and most of the FCs including YRO2 are Tyr49, Arg86, Tyr87 and Trp185. Among them, Tyr49 and Trp185 are residues that comprise the retinal-binding pocket in BR, HR and SR. R86 is important for the pathway of proton transfer (Fig. 5, residues in blue color). Its counterpart in BR is Arg82 that could connect the proton pathway between D85 and E204/E196 [48].

There are 12 residues that are conserved in the FCs but not in the archaeal rhodopsins. These residues are Phe10, Ala48, Phe50, Ala53, Gly57, Arg80, Tyr136, Lys137, Trp138, Tyr140, Tyr141, Gly198 and Tyr211. These residues are displayed in Fig. 6. Among them, the counterpart of Tyr141 in BR, W138, may contribute to the retinal-binding pocket. However, we found that all of these residues are located at or near the extracytoplasmic surface of the protein (Fig. 6) although the highly hydrophilic residues Arg80 and Lys137 face inwards. The location of these conserved residues and the directions of the hydrophilic side chains suggest that they have important biochemical functions.

3.4. A possible proton transfer pathway in YRO2

Although experimental evidence is currently lacking, residues showing a high degree of similarity with those in the archaeal rhodopsins are positioned

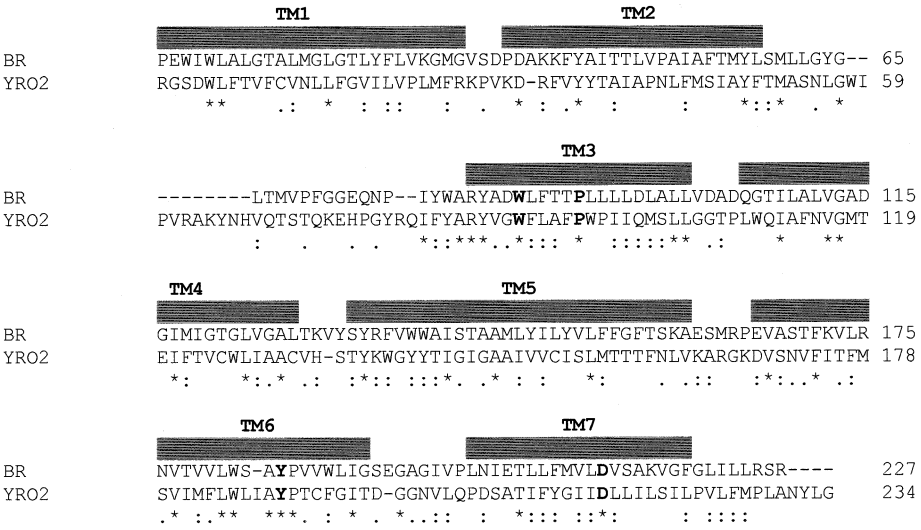

Fig. 4. A partial sequence alignment of BR and YRO2 used for homology modeling. TMSs 1–7 are indicated by numbered, shaded bars. The fully conserved residues shown in bold print in Fig. 1 are also presented in bold print here. Asterisks below the alignment indicate conserved residues while single or double dots indicate distant and close similarities, respectively.

throughout the transmembrane regions of YRO2 in a fashion that theoretically could allow proton transport (Fig. 7). The conserved proton acceptor, D85 in BR, is not present in YRO2. The counterpart of this residue in members of the FC family is not conserved and seems to have no biological significance. However, there is another acidic residue, D212, near D85 in BR. Its counterpart in the FC, YRO2, is Asp215. It is fully conserved in all of the proteins of the MR family. This residue is located near the middle of helix 7, towards the cytoplasmic side of the protein, just opposite D85 of helix 3 in BR. D212 hydrogen bonds to Y57 and Y185 that gives it an unusually low pK_a and renders it difficult to accept a proton.

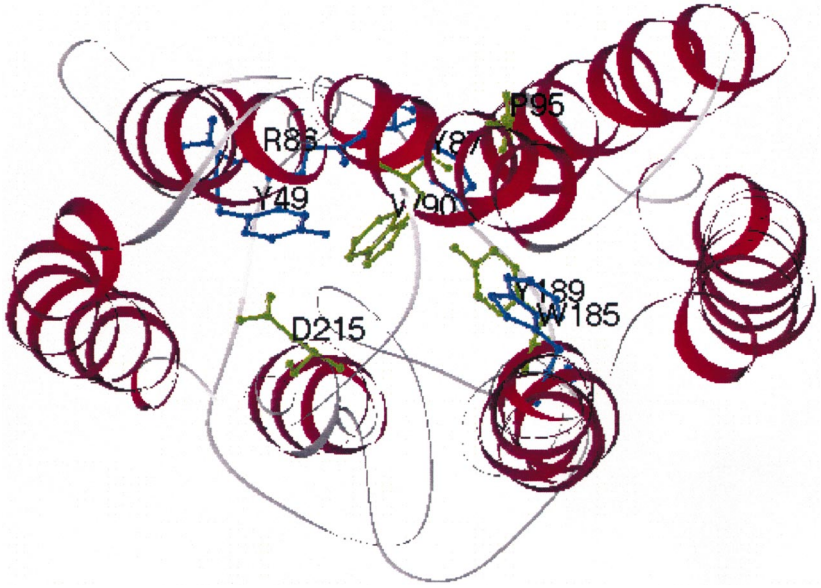

Fig. 5. Three-dimensional structural model for YRO2 showing well conserved residues in all of the proteins that comprise the four clusters of the MR family. Residues in green are those that are fully conserved while residues in blue are well conserved in the BRs, HRs and SRs, as well as in the FCs.

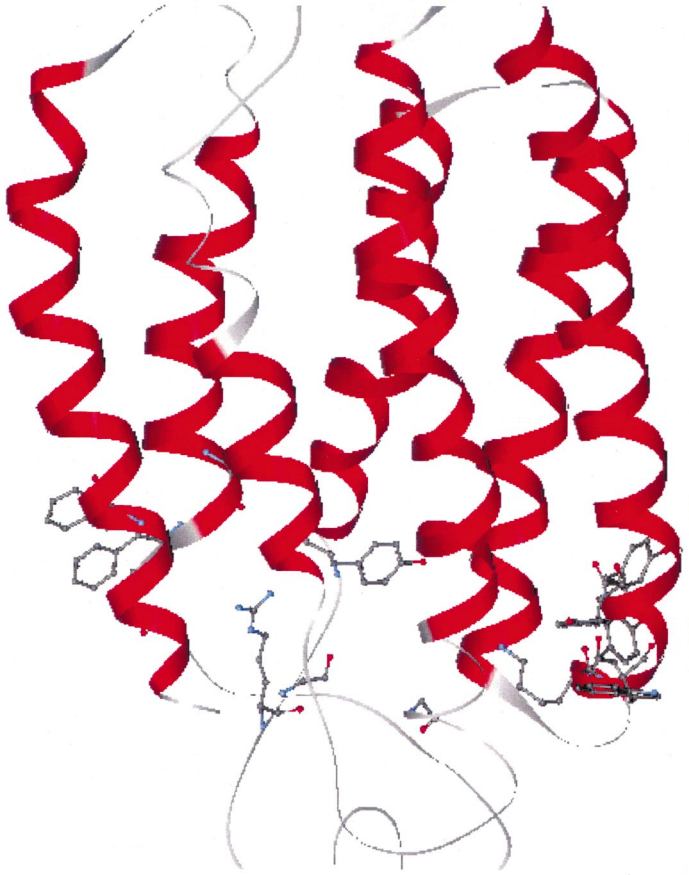

Fig. 6. Three-dimensional structural model for YRO2 showing residues that are conserved only in the FCs. All of these residues are localized to the external membrane surface of the protein.

Y185 in BR is fully conserved in the FCs, and Y57 is replaced by a Phe only in Hsp30. The side chain of Tyr189 in YRO2 tilts slightly towards helix 4, and Asp215 tilts slightly towards helix 1. The conformation of these side chains obliterates a potential hydrogen bond between them, and the hydrogen bond interaction between Asp215 and Tyr49, if any, is also likely to be weak. In Hsp30, there can be no hydrogen bonding interaction since Tyr is replaced by a Phe. This substitution would cause Asp215 to have a larger pK_a value so that it can act as a proton acceptor.

Another important residue, Arg86, is well conserved in all AR families as well as in the FCs. Only in Hsp30 is it replaced by the similar residue, Lys. Different models have different side chain conformations for this residue in BR. In the Hartmat Luecke model, this residue is hydrogen bonded to D212 [7], but in Grigorieff's model, it points toward the extracellular surface [49]. It is the bridge that contacts the proton transfer pathway between D85 and E194/E204 in BR.

E194 is conserved in the BRs and HRs. It is located in the loop region on the extracytoplasmic side of the membrane where it serves as the terminal residue in the proton transfer pathway where the proton is released. In our sequence alignment, it has no equivalent in YRO2, but Asp197, the first residue in the loop between TMSs 6 and 7 (or the Glu in Cve and Spo, see Fig. 1), perhaps acts in the same capacity. Since it is located in the loop region, the position and conformation should be variable, but there is a reasonable possibility that this residue plays an important role. Another hydrophilic residue that is fully conserved in the FCs, Arg80, is located in the loop region between helix 2 and helix 3. Its side chain faces inwards and may be important for proton transfer. Finally, in BR, D96 is located on the

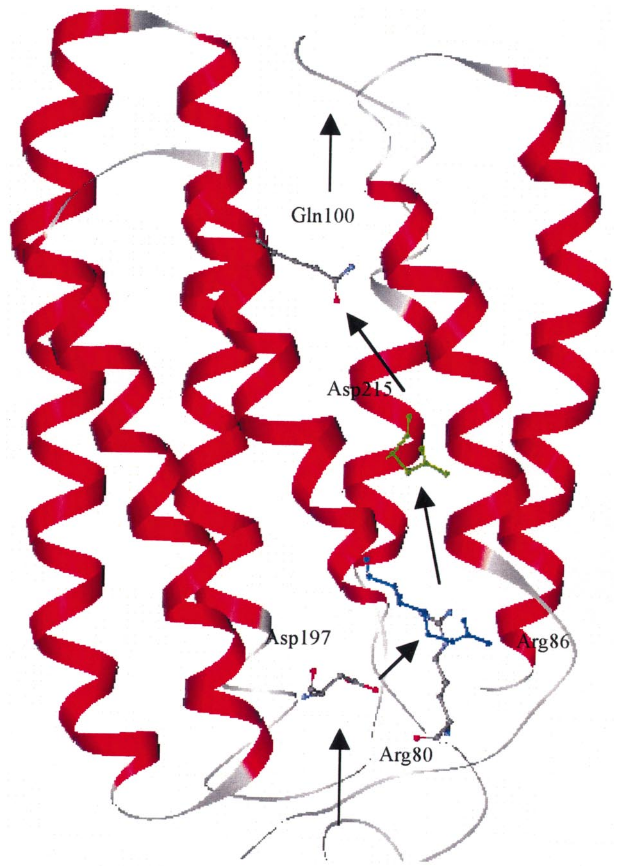

Fig. 7. Three-dimensional structural model for YRO2 showing a possible proton transfer pathway through the protein. The rationale for choosing these residues is described in the text.

cytoplasmic side of the membrane where it plays an important role in reprotonating the Schiff's base. In YRO2, its counterpart is Gln100, located in helix 3, and it may transfer a proton into the aqueous solution on the cytoplasmic side of the membrane.

In summary, we have postulated a proton transfer pathway in the FC protein, YRO2. Each of the residues that may play an important role has its counterpart in BR. The postulated proton transfer pathway through YRO2 is: external milieu → Arg80 or Asp197 → Arg86 → Asp215 → Gln100 → cytoplasm.

The proton thus flows down its electrochemical gradient so that it can provide the energy for protein folding, just as ATP provides the energy for chaperone-mediated protein folding in the cytoplasm of all living cells. In BR, the proton flows electrogenically in the opposite direction in response to light absorption instead of electrophoretically as a source of energy for protein folding.

3.5. The LCT family

Table 2 lists the proteins of the LCT family, a eukaryotic family with representation in plants, animals and yeast. The abbreviations, database descriptions, organismal sources, protein sizes and predicted numbers of TMSs are provided. Although these proteins were originally identified on the basis of homology, they are also similar in size and predicted topology. Thus, all MR and LCT family members are of 250–350 residues in size and exhibit seven putative TMSs. While the spacing of the TMSs is relatively constant for the members of the MR family, this is not true for members of the LCT family (Table 1).

Molecular functional data are available only for human cystinosin, the human lysosomal cystine transporter [22]. This protein is believed to actively transport cystine out of lysosomes into the cell cytoplasm by a proton motive force (pmf)-dependent mechanism [22]. The pmf across the lysosomal membrane is generated by a vacuolar-type (V-type) ATPase that pumps protons into the lysosomal lumen upon hydrolysis of cytoplasmic ATP [22]. The gene encoding cystinosin, the human *CTNS* gene, has been identified [23,25,50], and the deletion mutation giving rise to cystinosis has been characterized [26]. Clinical and molecular aspects of the disease have recently been reviewed [24,27].

Distant homologues of cystinosin include the Lec15/Lec35 suppressor, SL15, of Chinese hamster ovary (CHO) cells [51] and ERS1, the ERD suppres-

```
                          *     *  *                                                  *              *
YRO2 Sce    (159) ACVHSTYKWGYYTIGIGAAIVVCISLMTTTFNLVKARGKDVSNVFITFMSVIMFLWLIAYPTCFGITDG-GNVLQPDSATIFYGI (243)
YDR033 Sce  (159) ACVHSTYKWGYYTIGLGAAIVVSISVMTTSYNLVKQRDNDIRLTFLVFFSIIMFLWIIAYPTCFGITDG-GNVLQPDSAGIFYGI (243)
YDR090 Sce  (119) PLYEKGVKWPDLIFGISASVLLAVGLLPPYFELAKRKGRVIGINFAFLFIDSLGAWLSIISVILGNMDIMGIILYSIVAGMELGI (204)
YDO3 Spo    (131) LGIRRHVEWPVVFMGVLATVLVNIGFLPQYISIFRAR-AVTGISYLFLAIDSSGSLFSFLSLPFDRWDVLAAVDYGLLFIIEMGV (215)
```

Fig. 8. Partial alignment of two yeast members of the MR family with two yeast members of the LCT family (YDR090 and YDO3). Residues presented in bold print are those conserved in at least three of the four proteins, and residues highlighted with asterisks are those that are fully conserved. Residue numbers in each of the proteins are indicated in parentheses before and after the aligned sequence.

Table 2
Established members and distant homologues of the lysosomal cystine transporter (LCT) family[a]

Abbreviation	Name or description in database	Organism	Size (no. residues)	Database and accession No.	gi No.	No. TMS[b]; topology	Pos. PQ-1	Pos. PQ-2	Distance[c]
Ath-1	putative protein	*Arabidopsis thaliana*	288	gbCAA16618	2827664	7; 3-1-3	23	196	173
Ath-2	putative protein	*Arabidopsis thaliana*	374	gbCAB16817	4006887	7; 3-1-3	50	263	213
Y43H11 Cel[d]	Y43H11.120 orf	*Caenorhabditis elegans*	409	no accession No.	no gi No.	7; 3-1-2–1	58	217	159
YBZ7 Sce	YBZ7 (YBR147w) probable membrane protein 33.5 kDa	*Saccharomyces cerevisiae*	296	spP38279	586551	7; 3–1-1-2	30	212	182
YOL092w Sce	probable membrane protein	*Saccharomyces cerevisiae*	308	pirS57377	1362441	7; 1-2–1-1-2	30	216	186
YDR352w Sce	probable membrane protein	*Saccharomyces cerevisiae*	317	pirS61149	2132508	7; 1-2-1-3	28	189	161
YDFA Spo	YDFA hypothetical 30.6 kDa protein	*Schizosaccharomyces pombe*	271	spQ10482	1723569	7; 1-2–1-1-2	28	182	154
ERS1 Sce	ERD-1 suppressor (suppresses retention of endogenous ER proteins)	*Saccharomyces cerevisiae*	260	spP17261	119562	7; 2-1-2-2	21	159	138
pCTNS Hsa	cystinosis nephropathic lysosomal cystine transporter	*Homo sapiens*	367	gbNP004928	4826682	7; 2-1-2-2	143	270	127
Cel-3	similarity to TM domains of Sce ERS1 (C41C4.7)	*Caenorhabditis elegans*	487	gbCAA88102	3874889	7; 1-1-2-2–1	145	272	127
YDO3 Spo	YDO3 hypothetical 31.3 kDa protein	*Schizosaccharomyces pombe*	283	spQ10227	1723416	7; 1-1-2-2-1	35	146	111
YDR090c Sce	probable membrane protein	*Saccharomyces cerevisiae*	310	pirS58090	2132473	7; 1-3-1-2	23	135	112
YMPO Sce	YMPO (YMR010w)	*Saccharomyces cerevisiae*	405	spQ03687	1078557	7; 3-1-3	–	269	–
YDX3 Spo	YDX3	*Schizosaccharomyces pombe*	302	gbCAB11174	2330801	7; 2-1-2-2	–	201	–
Cel-2	F38E1.9	*Caenorhabditis elegans*	238	gbAAA83473	1123071	7; 1-1-3-1-1	51	154	103
supl15h Mmu	supl15h	*Mus musculus*	247	dbjBAA78781	5103142	7; 1-2-2-1-1	60	163	103
SL15 Cgr	SL15: Lec15 suppressor	*Cricetulus griseus*	247	gbAAD30976	4838367	7; 1-2-2-1-1	60	163	103
SL15h Hsa	SL15	*Homo sapiens*	247	gbNP004861	4759110	7; 1-2-2-1-1	60	163	103
SL15 Ath	Homologue of SL15	*Arabidopsis thaliana*	143	gbAAD48939	5732040	3; 1–2	47	–	–

[a]Format of presentation is essentially as for Table 1. Proteins below the first hairline exhibit only the second repeat sequence. Proteins below the second hairline exhibit both repeat sequences but are divergent in sequence, relative to the proteins listed above the first hairline.
[b]All proteins listed exhibit seven putative TMSs except SL15 Ath that is half size and exhibits three TMSs. Relative spacing of TMSs is indicated after the semicolon. Numbers indicate numbers of closely allied TMSs. Short and long dashes indicate increased degrees of spacing between TMSs, respectively.
[c]Distance between the two PQ motifs in numbers of residues.
[d]The DNA accession No. for this protein is RC006765, and the gi No. is 426173.

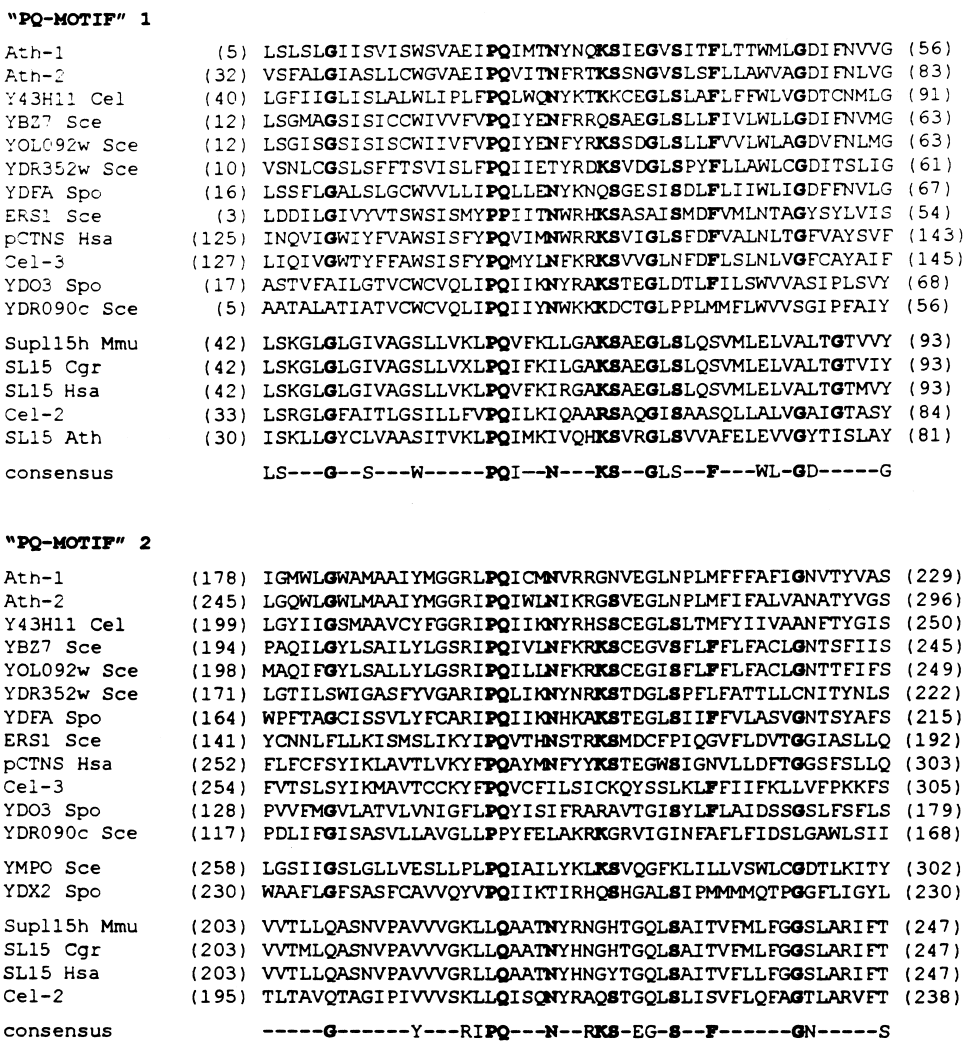

Fig. 9. Partial multiple sequence alignments of LCT family proteins showing the repeat sequences preceding and overlapping TMSs 1 and 2 (PQ-motif 1) and TMSs 5 and 6 (PQ-motif 2). The convention of presentation is the same as in Fig. 8 except that residues presented in bold are those that are well conserved and illustrate the repeat sequence motif. A consensus sequence (a majority of the residues conserved) is presented at the bottom of each alignment. Protein abbreviations are as presented in Table 2. In the top figure (PQ-motif 1), SL15 Ath is of half-length, exhibits only the first motif sequence, and is therefore lacking in the bottom figure. Proteins YMPO Sce and YDX2 Spo exhibit only the second motif sequence and are therefore lacking in the top figure.

sor in *S. cerevisiae* [52]. Both of these suppressors, when overexpressed, have been reported to influence retention of luminal endoplasmic reticular proteins and glycosylation in the Golgi apparatus. The Lec15 and Lec35 mutations are characterized by inefficient synthesis and utilization, respectively, of mannose-phosphate-dolichol for glycolipid (and glycoprotein) biosynthesis [51].

Fig. 8 shows a partial sequence alignment of two members of the MR family with two members of the LCT family. Residues conserved in at least three of the four proteins are presented in bold print. As can be seen, the numbers of shared residues are considerable, and comparison scores using the GAP program with 500 random shuffles were in excess of 9 S.D. (probability that the sequence similarity exhibited arose by chance is less than 10^{-19}) when a MR family protein was compared with an LCT family member. This result establishes that at least major portions of the proteins of these two families are related, and that they therefore probably arose from a single protein precursor. While we cannot

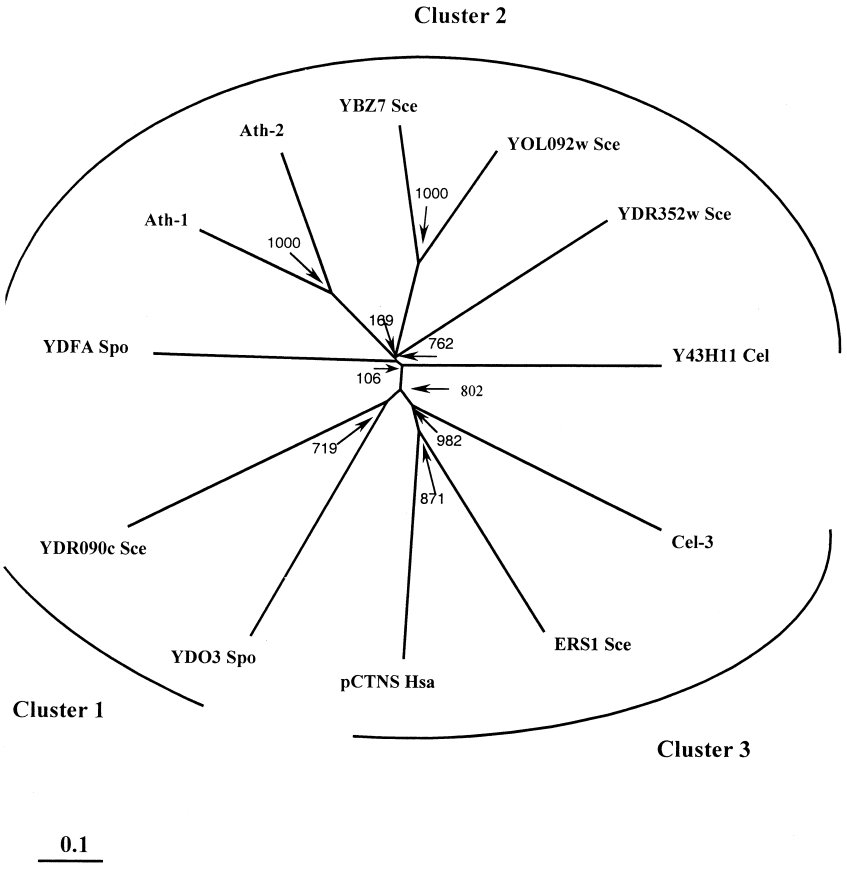

Fig. 10. Phylogenetic tree of the LCT family proteins. The proteins represented are those that are above the first double line in Table 2. These proteins are well conserved and exhibit both repeat sequences as illustrated in Fig. 4. The complete protein sequences were used to derive the tree. Bootstrap values are presented at the nodes of each branch. Values vary from 1000 (100% confidence) to 106 (10.6% confidence). Branch length is approximately proportional to phylogenetic distance.

establish that other portions of these proteins are homologous, this seems a very likely possibility (see [53]).

LCT family proteins are characterized by two well conserved repeat sequences including a characteristic PQ dipeptide (Fig. 9). These repeat sequences encompass TMSs 1 and 2 and the intervening loop as well as TMSs 5 and 6 and their intervening loop. Thus, in all of the LCT family proteins tabulated in Table 2 above the double line, the highly conserved PQ motif is present at the ends of TMS 1 and TMS 5 or in the loop region between these and the subsequent TMSs. Comparison scores obtained for the first and second repeat sequences of several of these proteins were of 9 S.D. or greater, establishing that these repeat segments are homologous.

Proteins below the second line in Table 2 are more distant homologues that exhibit at least one and often both such repeat sequences (Fig. 9). One of the latter proteins, SL15 Ath, is only 143 residues long and exhibits only three TMSs. It nevertheless is a clear homologue of LCT family members. It may either be a true representative of a half-sized member of the LCT family, or this sequence may have resulted from artifactual truncation of a full-length protein. Examples of authentic half sized proteins homologous to recognized internally duplicated transporters have been described [54,55].

The relative approximate spacing between putative TMSs in LCT family proteins is summarized in Table 2. Considerable variability was noted. Thus, in the proteins from *Arabidopsis thaliana*, the first and last three TMSs are closely spaced while greater spacing is observed between the third, fourth and fifth TMSs. However, substantial variation in loop length is observed for these proteins (see Table 2).

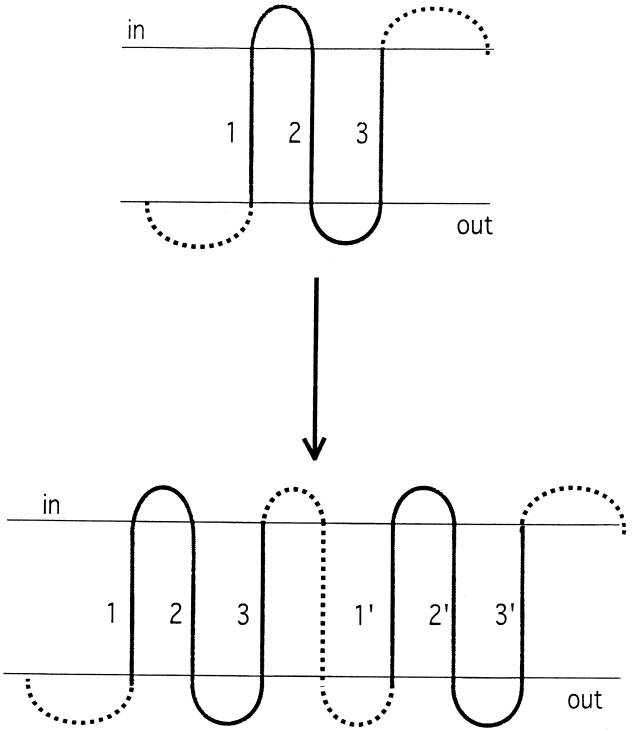

Fig. 11. Schematic depictions of the proposed internal gene duplication event that gave rise to members of the LCT family (including human cystinosin) and that, by analogy, is proposed to have given rise to members of the MR family (including bacteriorhodopsin). In this latter case, there is, however, insufficient sequence similarity between the two halves to establish common origin. Creation of a new, central TMS (TMS 4) allowed both repeat segments (TMSs 1–3 and TMSs 5–7) to retain the same orientation in the membrane.

As expected, the loop sequences in general proved to be less well conserved than the putative transmembrane sequences (data not shown). Average amphipathicity plots revealed that like the MR family proteins, LCT family proteins usually exhibit an amphipathic structure (when displayed as an α-helix) preceding TMS 1 (data not shown; see Fig. 2 for corresponding data for the MR family).

The phylogenetic tree for the LCT family is shown in Fig. 10. As shown, the two *A. thaliana* proteins, and two of the three *S. cerevisiae* proteins cluster convincingly together at the top of the tree. Moreover, several proteins in the lower portion of the tree exhibit loose clustering. Bootstrap values provide some degree of confidence at each node but do not evaluate the assumptions upon which the program is based [32].

4. Conclusions

The analyses reported in this communication provide detailed information about conserved residues in putative fungal chaperones and archaeal photoreceptor proteins which, although divergent in function, appear to be conserved in structure. This postulate is based on: (1) common ancestry as revealed by high degrees of sequence similarity (see Fig. 1); (2) common apparent topology as revealed by the hydropathy plots shown in Fig. 2A–D; and (3) the results of average amphipathicity analyses suggesting that these proteins may share biogenic properties as discussed for other classes of integral membrane proteins [41,42]. Of equal significance are the evolutionary implications resulting from our phylogenetic analyses. Thus, homologues of the MR family are found in bacterial, archaeal, and eukaryotic kingdoms. It can be argued that the primordial gene giving rise to all MR family members probably arose before divergence of these three kingdoms from each other.

Our phylogenetic results extend the results presented in earlier publications [1,11,17,33]. We show that the MR family, which includes archaeal rhodopsins, fungal chaperones, one sequenced fungal retinal-containing rhodopsin [17] and one bacterial homologue [19], not included in our analyses, is distantly related to a family of animal, plant and yeast proteins of similar size and topology. This second family, which we have called the LCT family, exhibits a clear internal repeat not easily recognized in the proteins of the MR family. Although we cannot prove it, we propose that MR family proteins may similarly have arisen from a smaller precursor polypeptide chain following an internal gene duplication event. Such an event is depicted schematically in Fig. 11. Either this type of duplication event occurred twice, with that in the MR family occurring before that in the LCT family, thus accounting for the differences in degrees of sequence conservation, or else, the two halves of the proteins in the two families have undergone sequence divergence at different rates. Current techniques are inadequate to distinguish these two possibilities experimentally.

The results presented here strongly suggest that archaeal rhodopsins, with well defined three-dimensional structures [9,56], are homologous to plant,

animal, yeast and fungal proteins, most of which lack rhodopsin. Functional data currently available suggest that these proteins display dissimilar functions. Nevertheless, if the LCT and MR family proteins both derived from a three TMS precursor, as proposed, one would predict that TMSs 1–3 would share structural features with TMSs 5–7 in all of these proteins. For bacteriorhodopsin, this prediction appears to be verified [9,57]. It is interesting to note that Taylor and Agarwal [58] predicted that such a tandem intragenic duplication event occurred in the early evolution of bacteriorhodopsins although statistical evidence justifying this postulate was not available [1].

Several conclusions can be drawn from the three-dimensional structural modeling of the fungal chaperone protein, YRO2. First, the degree of sequence similarity between BR and YRO2 was sufficient to allow reliable alignment of the two sequences. Second, the alignment was confirmed by aligning both of these sequences with NOP1, a protein that showed greater sequence similarity to both BR and YRO2 than either of these two proteins did to the other. Third, the modeled structure of YRO2 proved to be very similar to that of BR. Fourth, based on the BR-modeled structure, residues well conserved between all members of the MR family proved to be localized to the central retinal-binding pocket of the transmembrane regions of these proteins. Fifth, residues conserved *only* in the fungal chaperone proteins were localized exclusively to the extracytoplasmic face of the proteins. Finally, residues showing a high degree of similarity with those in the archaeal rhodopsins were positioned through the transmembrane regions of YRO2 in a fashion that should allow the rapid passage of protons through the protein. Thus, like the archaeal rhodopsins [56,59], the fungal chaperones may exhibit a transmembrane proton conducting pathway.

The findings reported here allow us to propose a model for the biochemical mode of action of the fungal chaperones. These proteins are proposed to facilitate the folding of denatured proteins in the extracytoplasmic space. Thus, following heat shock, acid shock or organic solvent shock, denatured, or partially denatured periplasmic or cell wall associated proteins may be refolded in an energy-dependent mechanism that utilizes the pmf. Passage of the proton through the proton conducting pathway in an FC protein might induce a series of conformational changes that promote substrate protein folding, driven by proton flux through the FC, in a fashion that is functionally analogous to ATP-driven protein folding by cytoplasmic chaperones. Our proposal therefore introduces the concept of an entirely novel mode of energy coupling for protein folding. Further experimentation will be required to substantiate or refute the proposal presented here.

Acknowledgements

We are grateful to Jerry A. Schneider for useful discussions, Kevin T. Harley for his assistance with some of the in silico analyses reported, and Mary Beth Hiller and Milda Simonaitis for their assistance in the preparation of this manuscript. Work in our laboratory was supported by USPHS grants 5RO1 AI21702 from the National Institutes of Allergy and Infectious Diseases and 9RO1 GM55434 from the National Institute of General Medical Sciences, as well as by the M.H. Saier, Sr. Memorial Research Fund.

References

[1] G. Kuan, M.H. Saier Jr., Phylogenetic relationships among bacteriorhodopsins, Res. Microbiol. 145 (1994) 273–285.
[2] J. Sasaki, L.S. Brown, Y.-S. Chon, H. Kandori, A. Maeda, R. Needleman, J.K. Lanyi, Conversion of bacteriorhodopsin into a chloride ion pump, Science 269 (1995) 73–75.
[3] W. Zhang, A. Brooun, M.M. Mueller, M. Alam, The primary structures of the Archaeon *Halobacterium salinarum* blue light receptor sensory rhodopsin II and its transducer, a methyl-accepting protein, Proc. Natl. Acad. Sci. USA 93 (1996) 8230–8235.
[4] J.K. Lanyi, Mechanism of ion transport across membranes. Bacteriorhodopsin as a prototype for proton pumps, J. Biol. Chem. 272 (1997) 31209–31212.
[5] J.l. Spudich, Variations on a molecular switch: transport and sensory signalling by archaeal rhodopsins, Mol. Microbiol. 28 (1998) 1051–1058.
[6] X.-N. Zhang, J. Zhu, J.L. Spudich, The specificity of interaction of archaeal transducers with their cognate sensory rhodopsins is determined by their transmembrane helices, Proc. Natl. Acad. Sci. USA 96 (1999) 857–862.
[7] H. Luecke, H.-T. Richter, J.K. Lanyi, Proton transfer path-

ways in bacteriorhodopsin at 2.3 angstrom resolution, Science 280 (1998) 1934–1937.
[8] D. Oesterhelt, The structure and mechanism of the family of retinal proteins from halophilic archaea, Curr. Opin. Struct. Biol. 8 (1998) 489–500.
[9] K. Mitsuoka, T. Hirai, K. Murata, A. Miyazawa, A. Kidera, Y. Kimura, Y. Fujiyoshi, The structure of the bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution, J. Mol. Biol. 286 (1999) 861–882.
[10] M. Kolbe, H. Besir, L.-O. Essen, D. Oesterhelt, Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution, Science 288 (2000) 1390–1396.
[11] K. Ihara, T. Umemura, I. Katagiri, T. Kitajima-Ihara, Y. Sugiyama, Y. Kimura, Y. Mukohata, Evolution of the archaeal rhodopsins: evolution rate changes by gene duplication and functional differentiation, J. Mol. Biol. 285 (1999) 163–174.
[12] B. Panaretou, P.W. Piper, The plasma membrane of yeast acquires a novel heat-shock protein (hsp30) and displays a decline in proton-pumping ATPase levels in response to both heat shock and the entry to stationary phase, Eur. J. Biochem. 206 (1992) 635–640.
[13] M. Regnacq, H. Boucherie, Isolation and sequence of HSP30, a yeast heat-shock gene coding for a hydrophobic membrane protein, Curr. Genet. 23 (1993) 435–442.
[14] Y. Iimura, K. Tatsumi, Isolation of mRNAs induced by a hazardous chemical in white-rot fungus, *Coriolus versicolor*, by differential display, FEBS Lett. 412 (1997) 370–374.
[15] I.J. Seymour, P.W. Piper, Stress induction of HSP30, the plasma membrane heat shock protein gene of *Saccharomyces cerevisiae*, appears not to use known stress-regulated transcription factors, Microbiology 145 (1999) 231–239.
[16] P.W. Piper, C. Ortiz-Calderon, C. Holyoak, P. Coote, M. Cole, hsp30, the integral plasma membrane heat shock protein of *Saccharomyces cerevisiae*, is a stress-inducible regulator of plasma membrane H(+)-ATPase, Cell Stress Chaperones 2 (1997) 12–24.
[17] J.A. Bieszke, E.L. Braun, L.E. Bean, S. Kang, D.O. Natvig, K.A. Borkovich, The *nop-1* gene of *Neurospora crassa* encodes a seven transmembrane helix retinal-binding protein homologous to archaeal rhodopsins, Proc. Natl. Acad. Sci. USA 96 (1999) 8034–8039.
[18] J.A. Bieszke, E.N. Spudich, K.L. Scott, K.A. Borkovich, J.L. Spudich, A eukaryotic protein, NOP-1, binds retinal to form an archaeal rhodopsin-like photochemically reactive pigment, Biochemistry 38 (1999) 14138–14145.
[19] O. Béjà, 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, E.F. DeLong, Bacterial rhodopsin: evidence for a new type of phototrophy in the sea, Science 289 (2000) 1902–1906.
[20] M.H. Saier, Jr., Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya, in: R.K. Poole (Ed.), Advances in Microbial Physiology, Academic Press, San Diego, CA, 1998, pp. 81–136.
[21] M.H. Saier Jr., A functional/phylogenetic classification system for transmembrane solute transporters, Microbiol. Mol. Biol. Rev. 64 (2000) 354–411.
[22] M.L. Smith, A.A. Greene, R. Potashnik, S.A. Mendoza, J.A. Schneider, Lysosomal cystine transport. Effect of intralysosomal pH and membrane potential, J. Biol. Chem. 262 (1987) 1244–1253.
[23] G. Jean, A. Fuchshuber, M.M. Town, O. Gribouval, J.A. Schneider, M. Broyer, W. van't Hoff, P. Niaudet, C. Antignac, High-resolution mapping of the gene for cystinosis, using combined biochemical and linkage analysis, Am. J. Hum. Genet. 58 (1996) 535–543.
[24] G.A. McDowell, M.M. Town, W. van't Hoff, W.A. Gahl, Clinical and molecular aspects of nephropathic cystinosis, J. Mol. Med. 76 (1998) 295–302.
[25] M. Town, G. Jean, S. Cherqui, M. Attard, L. Forestier, S.A. Whitmore, D.F. Callen, O. Gribouval, M. Broyer, G.P. Bates, W. van't Hoff, C. Antignac, A novel gene encoding an integral membrane protein is mutated in nephropathic cystinosis, Nat. Genet. 18 (1998) 319–324.
[26] L. Forestier, G. Jean, M. Attard, S. Cherqui, C. Lewis, W. van't Hoff, M. Broyer, M. Town, C. Antignac, Molecular characterization of *CTNS* deletions in nephropathic cystinosis: development of a PCR-based detection assay, Am. J. Hum. Genet. 65 (1999) 353–359.
[27] Y. Anikster, C. Lucero, J. Guo, M. Huizing, V. Shotelersuk, I. Bernardini, G. McDowell, F. Iwata, M.I. Kaiser-Kupfer, R. Jaffe, J. Thoene, J.A. Schneider, W.A. Gahl, Ocular nonnephropathic cystinosis: clinical, biochemical and molecular correlations, Pediatr. Res. 47 (2000) 17–23.
[28] J. Devereux, P. Hasberli, O. Smithies, A comprehensive set of sequence analysis programmes for the VAX, Nucleic Acids Res. 12 (1984) 387–395.
[29] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[30] D.-F. Feng, R.F. Doolittle, Progressive alignment and phylogenetic tree construction of protein sequences, Methods Enzymol. 183 (1990) 375–387.
[31] J. Felsenstein, PHYLIP – phylogeny inference package (version 3.2), Cladistics 5 (1989) 164–166.
[32] G.B. Young, D.L. Jack, D.W. Smith, M.H. Saier Jr., The amino acid/auxin:proton symport permease family, Biochim. Biophys. Acta 1415 (1999) 306–322.
[33] R.C. Graul, W. Sadée, Evolutionary relationships among proteins probed by an iterative neighborhood cluster analysis (INCA). Alignment of bacteriorhodopsins with the yeast sequence YRO2, Pharm. Res. 14 (1997) 1533–1541.
[34] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1982) 105–132.
[35] T. Le, T.-T. Tseng, M.H. Saier Jr., Flexible programs for the prediction of average amphipathicity of multiply aligned homologous proteins: application to integral membrane transport proteins, Mol. Membr. Biol. 16 (1999) 173–179.

[36] H. Belrhali, P. Nollert, A. Royant, C. Menzel, J.P. Rosenbusch, E.M. Landau, E. Pebay-Peyroula, Protein, lipid and water organization in bacteriorhodopsin crystals: a molecular view of the purple membrane at 1.9 Å resolution, Struct. Fold Des. 7 (1999) 909–917.
[37] H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, J.K. Lanyi, Structure of bacteriorhodopsin at 1.55 Å resolution, J. Mol. Biol. 291 (1999) 899–911.
[38] P.W. Piper, The heat shock and ethanol stress responses of yeast exhibit extensive similarity and functional overlap, FEMS Microbiol. Lett. 134 (1995) 121–127.
[39] P.J.F. Henderson, M.C.J. Maiden, Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes, Philos. Trans. R. Soc. London Ser. B Biol. Sci. 326 (1990) 391–410.
[40] G. von Heijne, J.P. Segrest, The leader peptides from bacteriorhodopsin and halorhodopsin are potential membrane-spanning amphipathic helices, FEBS Lett. 213 (1987) 238–240.
[41] M.H. Saier Jr., P.K. Werner, M. Müller, Insertion of proteins into bacterial membranes: mechanism, characteristics, and comparisons with the eucaryotic process, Microbiol. Rev. 53 (1989) 333–366.
[42] Y. Yamada, Y.-Y. Chang, G.A. Daniels, L.-F. Wu, J.M. Tomich, M. Yamada, M.H. Saier Jr., Insertion of the mannitol permease into the membrane of *Escherichia coli*. Possible involvement of an N-terminal amphiphilic sequence, J. Biol. Chem. 266 (1991) 17863–17871.
[43] Y. Mukohata, K. Ihara, T. Tamura, Y. Sugiyama, Halobacterial rhodopsins, J. Biochem. 125 (1999) 649–657.
[44] S.L. Lin, B. Yan, Three-dimensional model of sensory rhodopsin I reveals important restraints between the protein and the chromophore, Protein Eng. 10 (1997) 197–206.
[45] N. Friedman, S. Druckmann, J. Lanyi, R. Needleman, A. Lewis, M. Ottolenghi, M. Sheves, A covalent link between the chromophore and the protein backbone of bacteriorhodopsin is not required for forming a photochemically active pigment analogous to the wild type, Biochemistry 33 (1994) 1971–1976.
[46] U. Schweiger, J. Tittor, D. Oesterhelt, Bacteriorhodopsin can function without a covalent linkage between retinal and protein, Biochemistry 33 (1994) 535–541.
[47] E.A. Zhukovsky, P.R. Robinson, D.D. Oprian, Transducin activation by rhodopsin without a covalent bond to the 11-cis-retinal chromophore, Science 251 (1991) 558–560.
[48] U. Haupts, J. Tittor, D. Oesterhelt, Closing in on bacteriorhodopsin: progress in understanding the molecule, Annu. Rev. Biophys. Biomol. Struct. 28 (1999) 367–399.
[49] N. Grigorieff, T.A. Ceska, K.H. Downing, J.M. Baldwin, R. Henderson, Electron-crystallographic refinement of the structure of bacteriorhodopsin, Science 259 (1996) 393–421.
[50] J. Thoene, R. Lemons, Y. Anikster, J. Mullet, K. Paelicke, C. Lucero, W. Gahl, J. Schneider, S.G. Shu, H.T. Campbell, Mutations of *CTNS* causing intermediate cystinosis, Mol. Genet. Metab. 67 (1999) 283–293.
[51] F.E. Ware, M.A. Lehrman, Expression cloning of a novel suppressor of the Lec15 and Lec35 glycosylation mutations of Chinese hamster ovary cells, J. Biol. Chem. 271 (1996) 13935–13938.
[52] K.G. Hardwick, H.R.B. Pelham, ERS1: a seven transmembrane domain protein from *Saccharomyces cerevisiae*, Nucleic Acids Res. 18 (1990) 2177.
[53] M.H. Saier Jr., Computer-aided analyses of transport protein sequences: gleaning evidence concerning function, structure, biogenesis, and evolution, Microbiol. Rev. 58 (1994) 71–93.
[54] D.H. Nies, S. Koch, S. Wachi, N. Peitzsch, M.H. Saier Jr., CHR, a novel family of prokaryotic proton motive force-driven transporters probably containing chromate/sulfate antiporters, J. Bacteriol. 180 (1998) 5799–5802.
[55] T.-T. Tseng, K.S. Gratwick, J. Kollman, D. Park, D.H. Nies, A. Goffeau, M.H. Saier Jr., The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins, J. Mol. Microbiol. Biotechnol. 1 (1999) 107–125.
[56] H. Luecke, B. Schobert, H.-T. Richter, J.-P. Cartailler, J.K. Lanyi, Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution, Science 286 (1999) 255–261.
[57] P. Tuffery, C. Etchebest, J.-L. Popot, R. Lavery, Prediction of the positioning of the seven transmembrane α-helices of bacteriorhodopsin. A molecular simulation study, J. Mol. Biol. 236 (1994) 1105–1122.
[58] E.W. Taylor, A. Agarwal, Sequence homology between bacteriorhodopsin and G-protein coupled receptors: exon shuffling or evolution by duplication?, FEBS Lett. 325 (1993) 161–166.
[59] R.B. Gennis, T.G. Ebrey, Proton pump caught in the act, Science 286 (1999) 252–253.
[60] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.