Secondary Structures Comparison of Aquaporin-1 and Bacteriorhodopsin: A Fourier Transform Infrared Spectroscopy Study of Two-Dimensional Membrane Crystals

Véronique Cabiaux,* Keith A. Oberg,* Petr Pancoska,# Thomas Walz,¶ Peter Agre,§ and Andreas Engel¶
*Université Libre de Bruxelles, Laboratoire de Chimie Physique des Macromolécules aux Interfaces, CP 206/2, 1050 Brussels, Belgium; #Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60607-7061 USA: §Departments of Biological Chemistry and Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205-2185 USA; and ¶M. E. Mueller Institute for Microscopic Structural Biology at the Biozentrum, University of Basel, CH-4056 Basel, Switzerland

ABSTRACT Aquaporins are integral membrane proteins found in diverse animal and plant tissues that mediate the permeability of plasma membranes to water molecules. Projection maps of two-dimensional crystals of aquaporin-1 (AQP1) reconstituted in lipid membranes suggested the presence of six to eight transmembrane helices in the protein. However, data from other sequence and spectroscopic analyses indicate that this protein may adopt a porin-like β-barrel fold. In this paper, we use Fourier transform infrared spectroscopy to characterize the secondary structure of highly purified native and proteolyzed AQP1 reconstituted in membrane crystalline arrays and compare it to bacteriorhodopsin. For this analysis the fractional secondary structure contents have been determined by using several different algorithms. In addition, a neural network-based evaluation of the Fourier transform infrared spectra in terms of numbers of secondary structure segments and their interconnections [s_{ij}] has been performed. The following conclusions were reached: 1) AQP1 is a highly helical protein (42–48% α-helix) with little or no β-sheet content. 2) The α-helices have a transmembrane orientation, but are more tilted (21° or 27°, depending on the considered refractive index) than the bacteriorhodopsin helices. 3) The helices in AQP1 undergo limited hydrogen/deuterium exchange and thus are not readily accessible to solvent. Our data support the AQP1 structural model derived from sequence prediction and epitope insertion experiments: AQP1 is a protein with at least six closely associated α-helices that span the lipid membrane.

INTRODUCTION

Aquaporins are integral membrane proteins that mediate the permeability of plasma membranes to water molecules in diverse animal and plant tissues (reviewed by Chrispeels and Agre, 1994, and Knepper, 1994). Aquaporin-1 (AQP1; M_r = 28,000), the archetypal member of the aquaporins, is found in red blood cells (Denker et al., 1988; Smith and Agre, 1991). Although radiation inactivation studies of erythrocyte membranes indicate that the AQP1 water pore has a size of 30 kDa (van Hoeck et al., 1991), its function was unambiguously demonstrated after identification of the gene (Preston and Agre, 1991) by expression in *Xenopus laevis* oocytes (Preston et al., 1992). $HgCl_2$, which reversibly inhibits water channels in native membranes (Macey and Farmer, 1970), also inhibited AQP1-mediated water flow through oocyte membranes (Preston et al., 1993).

AQP1 exhibits an osmotic water permeability of $\sim 3 \times 10^9$ water molecules per monomer per second in native erythrocyte membranes, in reconstituted proteoliposomes (Zeidel et al., 1992, 1994), and in two-dimensional (2D) lipid-protein crystals (Walz et al., 1994b). The Arrhenius activation energy of 2–3 kcal/mol (Preston et al., 1992; Walz et al., 1994b) is similar to that for the diffusion of water in bulk solution. AQP1 is highly selective for water and does not allow passage of ions or other small solutes such as urea (Zeidel et al., 1994). In addition, the mercury-inhibitable site in AQP1 has been localized to a critical cysteine residue (C189; Preston et al., 1993). Therefore, a narrow and highly specific water pore structure is expected.

AQP1 is solubilized as a tetramer by various detergents (Smith and Agre, 1991) and readily forms two-dimensional tetragonal crystals ($a = b$ = 9.6 nm) during reconstitution with phospholipids when the detergent is removed (Walz et al., 1994b). These properties opened an avenue for structural analyses of AQP1 (Mitra et al., 1994; Walz et al., 1994a,b). A three-dimensional (3D) map of negatively stained crystals showed a unit cell that contains two tetramers with opposite orientations in the bilayer (Walz et al., 1994a). Higher resolution projection maps were obtained by electron crystallography of glucose-embedded (Jap and Li, 1995; Walz et al., 1995) or vitrified AQP1 2D crystals (Mitra et al., 1995). The projection maps derived from glucose-embedded crystals exhibited a clearly resolved mass density in the center of the trapezoid-shaped subunit (Jap and Li, 1995; Walz et al., 1995) and distinct density maxima, which can be interpreted as projections of α-helices along their axes. Moreover, the surface topography of AQP1 has been determined to 9 Å by electron and atomic force microscopy (Walz et al., 1996).

Received for publication 21 January 1997 and in final form 9 April 1997.

Address reprint requests to Dr. Véronique Cabiaux, Université Libre de Bruxelles, Laboratoire de Chimie Physique des Macromolécules aux Interfaces, CP 206/2, Boulevard du Triomphe, 1050 Brussells, Belgium. Tel.: 32-2-650-5365; Fax: 32-2-650-5113; E-mail: vcabiaux@ulb.ac.be.

Dr. Walz's present address is Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, P.O. Box 594, Sheffield S10 2UH, England.

0006-3495/97/07/406/12 $2.00

Despite the wealth of 2D data on the AQP1 structure, information on its 3D mass distribution is sparse. Sequence-based structure prediction has yielded a model comprising six α-helical spans and two long loops that include the most conserved features (Preston and Agre, 1991). However, other investigators have proposed a model with a porin-type β-barrel fold (Fishbarg et al., 1995). Although this model is supported by some spectroscopic analyses of reconstituted AQP1 (Haris et al., 1995; van Hoeck et al., 1993), it is not consistent with the projection data available. Thus a reevaluation of the secondary structure content as derived from optical spectroscopy is required. Because the secondary structure contents determined from the analysis of spectroscopic data can vary with the algorithm used, we present here a systematic comparison of different secondary structure assessment strategies. In addition, the orientation of AQP1-helices in the lipid membrane and their accessibility to solvent are compared with those of bacteriorhodopsin, a protein for which a 3D atomic model is available (Grigorieff et al., 1996).

EXPERIMENTAL PROCEDURES

Sample preparation

Purified AQP1 from human erythrocytes (Smith and Agre, 1991) was reconstituted into 2D crystals as described (Walz et al., 1994b). For proteolytic cleavage, crystalline vesicles were first deglycosylated with endoglycosidase F/N-glycosidase F (Boehringer, Mannheim, Germany) by incubation in 250 mM sodium acetate, pH 6, 20 mM EDTA, and 10 mM β-mercaptoethanol overnight at 37°C. Deglycosylated AQP1 2D crystals were washed twice with 25 mM sodium acetate, pH 6, 0.5 mM dithiothreitol, and subsequently incubated with carboxypeptidase Y (Sigma, St. Louis, MO) overnight at 37°C. The digested vesicles were finally washed twice with the above buffer. Four native AQP1 and two proteolyzed samples were prepared for the collection of infrared spectra.

Purple membranes were isolated from *Halobacterium salinarium* (ET 1001) as described by Oesterhelt and Stoeckenius (1974). Purple membranes were stored in double-distilled water (25 mg/ml). All samples were centrifuged at 10,000 rpm, and the pellets were resuspended in 20 μl of 2 mM HEPES, pH 7.2, and deposited on one side of a germanium crystal, as described below.

Infrared spectroscopy

Attenuated total reflection infrared spectra (4 cm^{-1} resolution) were obtained on a Perkin-Elmer 1720X Fourier transform infrared (FTIR) spectrophotometer as previously described (Goormaghtigh et al., 1994b). Measurements were carried out at room temperature. Thin films were obtained by applying a sample to one side of an attenuated total reflection (ATR) plate and slowly evaporating the solvent under a stream of nitrogen (Fringeli and Günthard, 1981; Goormaghtigh and Ruysschaert, 1990). The ATR plate was then sealed in a universal sample holder (Perkin-Elmer 186-0354). The sample was deuterated by flushing the sample compartment with D_2O-saturated N_2 at room temperature for 90 min.

Kinetics of deuteration

This analysis was carried out as previously described (Goormaghtigh et al., 1994b). The samples were spread on a germanium plate as described above. Before starting the deuteration, 10 spectra were recorded to verify the reproducibility of the measurements and the stability of the system. At time 0 a D_2O saturated N_2 flux was applied to the sample with a flow rate of 75 ml/min that was controlled with a Brooks flow meter. The Perkin-Elmer 1720X FTIR spectrophotometer was driven by a computer program. The spectra at each time point were the accumulation of 12 scans, with a resolution of 4 cm^{-1}. The signal from atmospheric water was subtracted as described by Goormaghtigh et al. (1994b). The amide I and II band areas were measured between 1702–1596 cm^{-1} and 1585–1502 cm^{-1}, respectively. The amide II area was divided by the amide I area for each spectrum to correct for any change in the total intensity of the spectra during the deuteration process. This ratio, expressed between 0 and 100%, was plotted versus deuteration time. The 100% value is defined by the amide II/amide I ratio obtained before deuteration, whereas the 0% value corresponds to a zero absorption in the amide II region. It has been shown previously (de Jongh et al., 1995; Raussens et al., 1996) on a series of proteins that can be fully denatured (and therefore fully deuterated) and then refolded to their native conformation, that complete H/D exchange results in 0 ± 5% absorption intensity in the amide II region. We are therefore confident that a zero absorbance in the amide II region corresponds to full deuteration of the protein.

Secondary structure analysis methods

Fourier self-deconvolution and curve fitting

This analysis was performed on the amide I region of deuterated samples as described previously (Goormaghtigh et al., 1990, 1994a,b; Cabiaux et al., 1989). Deuterated samples were analyzed because hydrogen/deuterium exchange allows differentiation of accessible α-helix and irregular structures, the absorption band of which shifts from ~1655 cm^{-1} to ~1642 cm^{-1} (Cortijo et al., 1982; Goormaghtigh et al., 1994a). For curve fitting, Lorentzian bands at 1689, 1683, 1673, 1665, 1660, 1654, 1647, 1638, 1631, and 1619 cm^{-1}, and 1686, 1681, 1673, 1664, 1657, 1649, 1641, 1634, and 1627 cm^{-1} were used as input parameters, respectively, for bacteriorhodopsin (BR) and AQP1. These frequencies were determined after Fourier self-deconvolution, but fitting was applied to the original spectra. The frequency limits used for assigning the resulting bands to different secondary structures were as follows: 1662–1645 cm^{-1}, α-helix; 1689–1682 cm^{-1} and 1637–1613 cm^{-1}, β-sheet; 1644.5–1637 cm^{-1}, random; 1682–1662.5 cm^{-1}, β-turns.

Resolution enhancement and curve fitting

This analysis used curve fitting with bands taken from "resolution-enhanced" spectra of undeuterated samples (Oberg et al., 1994; Oberg and Fink, 1995). The second derivative of the amide I region was first taken, then Fourier self-deconvolution (FSD) was performed, using the second-derivative spectrum as a reference for choosing optimal deconvolution parameters. The best match between FSD and second-derivative spectra was obtained using $\gamma = 5$ and a filtering factor, x, of 0.22 (Yang and Griffiths, 1984). In the next stage of the analysis, the FSD spectrum was curve-fitted, using a modified version of the Spectra Calc (Galactic Industries, Salem, NH) program CURVEFIT.AB with manually defined Gaussian/Lorentzian peaks. Only component peaks present in both the FSD and second-derivative spectra were used for the selection of initial parameters. During optimization, all parameters (peak heights, widths, position, and percentage Lorentzian) were allowed to vary. Because the component bands in the FSD spectrum are well resolved, the results of this fit were not strongly affected by variations in the initial bands. The final peaks from the FSD fit were then uniformly decreased in intensity and increased in width, so that their sum approximately matched the original (unenhanced) spectrum. These values were used as initial parameters for fitting the original spectrum. Because the contour of the amide I band is relatively smooth, peak positions were fixed during this stage to prevent drift. After optimization was complete, the quality of the fit was tested by allowing peak positions to vary for an additional 100 iterations. The maximum change in peak positions was typically less than 0.3 cm^{-1}.

Partial least-squares with a basis set of ATR-FTIR protein spectra (PLS-ATR)

This analysis used a commercially available partial least-squares package to determine fractional secondary structure composition (PLSPlus 2.1 for Spectra Calc, Galactic Industries). In contrast to the other methods presented here, the basis set was generated by collecting ATR-FTIR spectra of undeuterated thin films of globular proteins (nonmembrane proteins), with known x-ray structures, prepared in the same manner as the AQP1 samples. The basis proteins were chosen to represent a wide range of structural motifs. They were selected by several criteria, including their availability and cost, the quality of their crystal structure, and their classification in the CATH data base (Orengo et al., 1996). They are listed in Table 1 with their catalog numbers, CATH numbers, and the names of the PDB files used to tabulate their FC values (by the method of Kabsch and Sander, 1983). A bacteriorhodopsin (BR) spectrum was also included in the basis set.

Samples included in the basis set were prepared by gently evaporating the water from 20 μl of a 10 mg/ml protein solution containing 2 mM HEPES and 0.02% NaN_3, pH 7.2. Although the thin films are known to remain hydrated after the evaporation of solvent (de Jongh et al., 1996), no free water remains and thus it is not necessary to subtract its signal from that of the protein. Before analysis, all spectra were normalized to the same area in the amide I region, so that the PLS algorithm was forced to use band shapes, rather than absolute intensities, for secondary structure determination. The mean-centering option was selected for the analysis; this essentially is the same as the generation of difference (FD) spectra, as discussed below. The optimal number of factors used to determine the secondary structure fractions (FC values) of each structural type was found by using the diagnostic routines included in the PLSPlus software; these routines use a "one out" method of testing. The standard deviations of prediction errors for the basis set were 6.8% for α-helix, 6.0% for β-sheet, 4.2% for turn, and 6.1% for other structures (all other structure types were combined with irregular structure). It was not possible to calculate accurate FC values for 3_{10} helix with this method because of an insufficient distribution of FC values: the total range of 3_{10}-helix content in the PLS basis set was 0–14%, but most of the proteins contained 3–7%. Independent determinations were made of the FC values for α-helix, β-sheet, turn, and other structures; the details of the analysis will be presented elsewhere (Oberg et al., manuscript in preparation).

Restricted multiple regression determination of secondary structure fractions (FA/RMR)

The factor analysis used here was based on FTIR measurements of aqueous solutions of 23 globular proteins with known x-ray structures (Baumruk et al., 1996). To enhance the variable bandshape features of FTIR spectra, the experimental FTIR spectra of reference proteins, together with the AQP1 and BR spectra, were first transformed into FD spectra in the following way. The spectra of the reference set and of AQP1 and BR were normalized to the same amide I maximum intensity, and their average spectrum was calculated. This average spectrum was subtracted from all spectra. Before the quantitative analysis of the AQP1 and BR spectra, their compatibility with the spectra of the reference set was checked by using two cluster analysis schemes. First, the FD spectra were directly analyzed by six different cluster analysis algorithms using the Einsight software package (InfoMetrix Inc.). Second, the decomposition of the reference, AQP1, and BR spectra into common orthogonal component spectra (subspectra) was performed as described by Pancoska et al. (1994). The coefficients of the component spectra necessary to reconstruct each experimental FD spectrum were then used as the input for the cluster analysis, as before. The similarity and compatibility of the AQP1 and BR spectra with those of the

TABLE 1 Proteins included in the PLS-ATR basis set

Protein	Catalog no.*	PDB code	% α-helix	% β-sheet	CATH no.# (domain 1)	CATH no. (domain 2)
α-Lactalbumin	L6010	1alc	31.2	6.6	3.90.90.10	
Apolipoprotein E3 fragment	—§	1lpe	64.5	0.0	1.20.120.20	
Carbonic anhydrase	C3934	1hcb	8.5	29.0	3.10.200.10	
Citrate synthetase	C3260	1csh	59.6	2.3	1.10.580.10	1.10.230.10
α-Chymotrypsinogen	C4879	2cga	7.4	32.0	2.40.80.20	
Cytochrome *c*	C2506	1hrc	41.4	0.0	1.10.370.10	
Ferritin	F4503	1hrs	71.3	0.0	1.20.110.10	
Hemoglobin	H2500	1hda	68.9	0.0	1.10.490.10	
Hemocyanin (II)	H2133	1lla	30.3	17.0	N/A‖	
Hexokinase	H5250	2yhx	37.7	14.8	N/A‖	
Immunoglobulin γ	56834¶	8fab	2.2	50.0	2.60.40.210	2.60.40.120
Lectin, lentil	L9267	1len	1.7	48.1	2.60.140.80	4.10.120.10
Lactic (L) dehydrogenase	L2375	9ldt	39.6	17.0	3.40.330.130	3.90.110.10
Lysozyme	L6876	1hel	31.0	6.2	3.90.90.10	
Pepsin	P6887	4pep	11.0	43.6	2.170.10.10	2.170.20.10
Phosphoglyceric kinase	P7634	3pgk	34.5	11.1	3.40.110.10	
Ribonuclease A	R5500	6rat	17.7	33.1	3.10.130.10	
Superoxide dismutase (Fe)	S5389	1isc	47.9	10.9	N/A‖	
Superoxide dismutase (Cu,Zn)	86200¶	1sxc	2.7	39.1	2.60.40.200	
Subtilisin Carlsberg	P5380	1cse	30.7	17.9	3.40.320.190	
Trypsinogen	T1143	2tga	7.2	32.3	2.40.80.20	
Triose phosphate isomerase	T2507	7tim	38.9	16.2	3.20.40.70	
Ubiquitin	U6253	1ubi	15.8	31.6	3.10.20.40	

*All proteins purchased from Sigma Biochemicals unless otherwise noted.

#CATH numbers correspond to the protein fold (CATH is an acronym for Class Architecture Topology Homologous superfamily). In particular, the first digit represents the classification of a domain's overall structure: 1 denotes a primarily α-helical domain, 2 denotes primarily β-sheet, and 3 denotes a domain containing significant fractions of both α and β structures. For further details see the CATH web page (Orengo et al., 1996).

§This protein was a gift from Robert O. Ryan, University of Alberta, residues 1–183.

¶Purchased from Fluka Biochemicals.

‖CATH number not currently available for this protein.

reference proteins were analyzed by representing the clustering results as dendrograms.

The factor analysis/restricted multiple regression (FA/RMR) analysis of undeuterated AQP1 and BR FD spectra was performed after their decomposition into the orthogonal subspectra common to them and the reference spectra. The relationship between the calculated contributions of these orthogonal component spectra and the secondary structure fractions (FC values, determined by the DSSP program of Kabsch and Sander, 1983) was optimized by calibration with the known FC values of the reference proteins, using the "one out" method of testing. This involved the least-squares optimization of a multiple regression function relating the components describing a given FD spectrum to the FC values of that protein. This was done successively 23 times for all possible numbers and combinations of these components using the spectra of the reference set, from which one protein was systematically removed. The spectral FC values of the reference protein left out were then determined, and the difference from the x-ray FC values calculated. These differences, determined for all reference proteins, yielded the standard deviation (σ) of the FC prediction. The multiple regression function providing the minimum σ was then used to determine the secondary structure fractions for AQP1 and BR from the principal components of their FD spectra. Selection of the optimal combination of spectral components for the determination of FC values was performed independently for five secondary structures considered: helices, sheets, bends, turns, and other categories.

Neural network determination of the matrix descriptor of interconnectivity of secondary structure segments

This analysis of the AQP1 and BR FD spectra involved a neural network approach to determining a matrix descriptor of secondary structure segment interconnectivities, as described recently (Pancoska et al., 1996; Pancoska et al., manuscript submitted for publication). The network was trained to determine the correct values of 3×3 matrix descriptors of protein structure, $[s_{ij}]$, from the 23 reference protein spectra. The resulting output matrix is a description of the protein in terms of how many continuous segments of each secondary element are present and how many of them are adjacent along the sequence in the native protein fold, rather than the fractional contribution (FC) of each structure type to the spectrum. Thus this method can be used to determine possible topological arrangements of secondary structure segments in a protein. Each element in the output matrix represents the predicted number of occurrences of a particular structure in the protein. The matrix element definitions are as follows: number of helical segments (S_{11}), number of sheet segments (S_{22}), number of "other" (C) segments (S_{33}), number of helix-sheet connections (S_{12}), number of helix-C connections (S_{13}), number of sheet-helix connections (S_{21}), number of sheet-C connections (S_{23}), number of C-helix connections (S_{31}), and number of C-sheet connections (S_{32}). The input for the network (Pancoska et al., 1996) was 200 equidistant FD intensities in the 1450–1750 cm^{-1} interval taken from all of the spectra in the FA/RMR set. For each protein, a reference descriptor was generated from the secondary structure assignments of Kabsch and Sander (1983) with a Turbo Pascal program (Janota, 1993). To ensure the optimal performance of the network (Pancoska et al., 1996), the numerical values of the $[s_{ij}]$ descriptors for all proteins were individually scaled into the interval $\langle -0.8, 0.8 \rangle$, based on the maximum element for each protein $[s_{ij}]$ matrix. The neural network was constructed with NeuralWorks Professional II+ software (v. 5.2; Neural Ware, Pittsburgh, PA). The network topology (number of neurons in respective hidden layers and the details of weight-updating algorithms) was optimized as described earlier (Pancoska et al., 1996). To train the network scheme, all of the FD spectra (input) and x-ray matrix descriptors of the reference proteins (output) were utilized, and the network state that produced the closest match to the actual BR matrix descriptor, as found by the "save-best" algorithm (Neural Ware documentation; Pancoska et al., 1996), was then used to process the FD spectra of AQP1. Renormalization of the network output into $[s_{ij}]$ of AQP1 was finally performed, using the norm for the BR matrix descriptor.

The final step of this analysis was reliability checking (and correction) of the scaled matrix descriptors, which is based on the fact that proteins are unbranched polymers. This requires that (with the exception of the number of "other" segments, which should be higher by one) the row and column sums of the off-diagonal elements of $[s_{ij}]$ should yield the corresponding diagonal entry (Pancoska et al., 1996). In structural terms, this condition means that the number of occurrences of a particular secondary structure in the protein, helix for example (given by the diagonal elements of $[s_{ij}]$), must be the same as the number of helix "heads" (lower left off-diagonal elements of $[s_{ij}]$) as well as the number of helix "tails" (upper-right off-diagonal elements of $[s_{ij}]$). For example, the analysis of the AQP1 spectra (which is, as in any such procedure, subjected to experimental and numerical error) resulted in a violation of this structural rule in the matrix element (S_{23}, Table 2), which was easily corrected by elimination of the single offending sheet-other (ec) connection segment. Such correction is justified because the value, 1, at this position is not supported by any of the other independently determined matrix elements. Similarly, for the BR $[s_{ij}]$ matrix, the number of helix-sheet (he) and sheet-other (ec) connections in the output matrix were not supported by the rest of the matrix elements, and so were set to zero, values consistent with the rest of the matrix.

RESULTS

Secondary structure determination

Fig. 1 shows the ATR-FTIR spectra of AQP1 (*solid line*) reconstituted in lipid vesicles as described by Walz et al. (1994b) and bacteriorhodopsin (BR) in purple membranes (*dashed line*), spread from an aqueous sample (Fig. 1 *A*) or from an aqueous sample flushed with N_2-saturated D_2O to allow H/D exchange (Fig. 1 *B*). Fourier self-deconvolution of the deuterated spectra (Fig. 1 *C*) shows that the main components of AQP1 (*a*) and BR (*b*) are located at 1658 and 1660 cm^{-1}, respectively. These maxima indicate the presence of a high percentage of α-helical structure in both proteins. Their locations are not modified by deuteration. The 1636 cm^{-1} component observed in BR has been interpreted as proof of some β-sheet in the BR structure (as a comparison for β-sheet absorption, see porin, Fig. 1 *C*, *upper trace*). However, several studies (Miick et al., 1992; Prestrelski et al., 1990, 1991; Zhang et al., 1995) have suggested that this component can also be attributed to the presence of non-α-helical structure, such as 3_{10}-helix. According to a recent refinement of the structure of BR, a few residues may indeed adopt 3_{10}-helical structure (Grigorieff et al., 1996).

Before presenting the quantitative analysis of AQP1 and BR spectra, we must address the question of the compatibility of their spectral features with those of globular proteins, which are used, either directly or indirectly, as a structural or calibration basis for all of the "empirical" algorithms used in this work to determine protein secondary structure from FTIR spectra. As an objective assessment, a cluster analysis was performed on the FD spectra and the parameters derived from the FA/RMR method. This only assumes that the bandshapes of FTIR spectra, as a whole, are sensitive to secondary structure, and that similarity between the reference and analyzed data ensures the required correspondence in our analyses. The results of both clustering calculations justified the choice of BR as the reference molecule for our study. BR FD spectra were classified as most similar to those of AQP1 (Fig. 2 *a*); the

TABLE 2 Interconnectivity matrix descriptors $[S_{ij}]$ of secondary structure segments determined from their FD spectra by the neural network analysis

Bacteriorhodopsin

Electron crystallography matrix*		Renormalized NN output#		Integer form§		
$\begin{bmatrix} 7 & 0 & 7 \\ 0 & 0 & 0 \\ 7 & 0 & 8 \end{bmatrix}$	$\Rightarrow$	$\begin{bmatrix} 6.9 & 1.1 & 7.1 \\ 0.3 & 0.2 & 0.6 \\ 6.8 & 0.3 & 8.4 \end{bmatrix}$	$\rightarrow$	$\begin{bmatrix} 7 & 1 & 7 \\ 0 & 0 & 1 \\ 7 & 0 & 8 \end{bmatrix}$	$\rightarrow$ *correction* $\rightarrow$	$\begin{bmatrix} 7 & 0 & 7 \\ 0 & 0 & 0 \\ 7 & 0 & 8 \end{bmatrix}$

AQP1

Renormalized NN output¶		Integer form§		
$\begin{bmatrix} 6.1 & 0.0 & 5.9 \\ 0.0 & 0.5 & 1.1 \\ 6.5 & 0.5 & 7.0 \end{bmatrix}$	$\Leftrightarrow$	$\begin{bmatrix} 6 & 0 & 6 \\ 0 & 0 & 1 \\ 6 & 0 & 7 \end{bmatrix}$	$\rightarrow$ *correction* $\rightarrow$	$\begin{bmatrix} 6 & 0 & 6 \\ 0 & 0 & 0 \\ 6 & 0 & 7 \end{bmatrix}$

Layout of secondary structure segments in AQP1 based on $[s_{ij}]$ results‖

C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C

*$[s_{ij}]$: Interconnectivity descriptor of secondary structure segments. See Scheme I (below) for the structural meaning of matrix elements, secondary structure definition, and assignment.

#Result of the "save best" optimization of the neural network analysis of FD spectra of reference protein spectra in terms of $[s_{ij}]$. The matrix is the raw output of the network in a state, determining best the BR $[s_{ij}]$ descriptor from its FD spectra (1400 iteration cycles, testing and saving of the network with BR data was performed every 200 iterations).

§Integer form of the descriptor is generated by the algebraic rounding of the real network output and as such is subjected to numerical error. Corrected form of the matrix descriptor is generated using sum rules for $[s_{ij}]$ elements as explained in the text.

¶Renormalized raw output of the neural network determination of the $[s_{ij}]$ descriptor from AQP1 FD spectra using the network state, optimized to determine the BR descriptor.

‖Symbols as in Scheme 1.

Scheme 1: Example construction of the interconnectivity descriptor $[s_{ij}]$ of secondary structure segments for bacteriorhodopsin

a) *Linear layout of secondary structure segments.*

C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C-ch-H-hc-C

Symbols: H, C—helical and "other" secondary structure segments, respectively (continuous sequences of amino acid residues assigned to the respective conformations by DSSP processing of the protein data bank file 2brd); ch, hc—ordered contacts between C $\rightarrow$ H and H $\rightarrow$ C segment ends.

b) *Definitions of the $[s_{ij}]$ matrix descriptor for bacteriorhodopsin based on the layout in a):*

$$n(\mathrm{H}) = S_{11} = 7, \quad n(\mathrm{he}) = S_{12} = 0; \quad n(\mathrm{hc}) = S_{13} = 7,$$
$$n(\mathrm{eh}) = S_{21} = 0; \quad n(\mathrm{E}) = S_{22} = 0; \quad n(\mathrm{ec}) = S_{23} = 0;$$
$$n(\mathrm{ch}) = S_{31} = 7; \quad n(\mathrm{ce}) = S_{32} = 0; \quad n(\mathrm{C}) = S_{33} = 8;$$

Symbols: n(X), number of segments and segment connections in the linear segment layout; S_{11}, number of helical segments; S_{22}, number of sheet segments; S_{33}, number of "other" segments (C); S_{12}, number of connections between helical and sheet segments in the protein fold; S_{13}, number of helix-C connections; S_{21}, number of sheet-helix connections; S_{23}, number of sheet-C connections; S_{31}, number of C-helix connections; S_{32}, number of C-sheet connections.

c) *Final form of the matrix descriptor $[s_{ij}]$ for bacteriorhodopsin*

$$[S_{ij}] \equiv \begin{bmatrix} 7 & 0 & 7 \\ 0 & 0 & 0 \\ 7 & 0 & 8 \end{bmatrix}$$

same was observed for the principal component spectra (Fig. 2 *b*). More importantly, in both clustering calculations, the primary cluster of AQP1/BR was mixed into the larger clusters of reference globular protein spectra of high α-character (hemoglobin, myoglobin, etc.). According to our experience, such a result can be found only if the AQP1 and BR spectra do not contain component bands and/or relative spectral intensities that are absent from the reference protein spectra. In this sense, the spectral set used to calibrate the spectra-structure relation most probably covers all structural features contributing to AQP1 and BR spectra. Therefore the use of this reference set of globular protein spectra and structures for the quantitative analysis of AQP1 is also justified.

As previously emphasized, the determination of secondary structure content from spectroscopic data is strongly dependent upon the algorithm and input parameters used for the analysis. Table 3 gives the percentage of AQP1 α-helical, β-sheet, and turns obtained from different analyses of the data presented in Fig. 1. To facilitate the discussion, we have also included FC values of AQP1 published in other studies, as well as the structure of BR obtained by electron crystallography (Grigorieff et al., 1996).

All secondary structure evaluation techniques underestimate the proportion of α-helical structure in BR. For the frequency-based algorithms, this underestimation most probably results from the observed shoulder at 1636 cm^{-1}, which is usually attributed to β-structure in FTIR analysis.

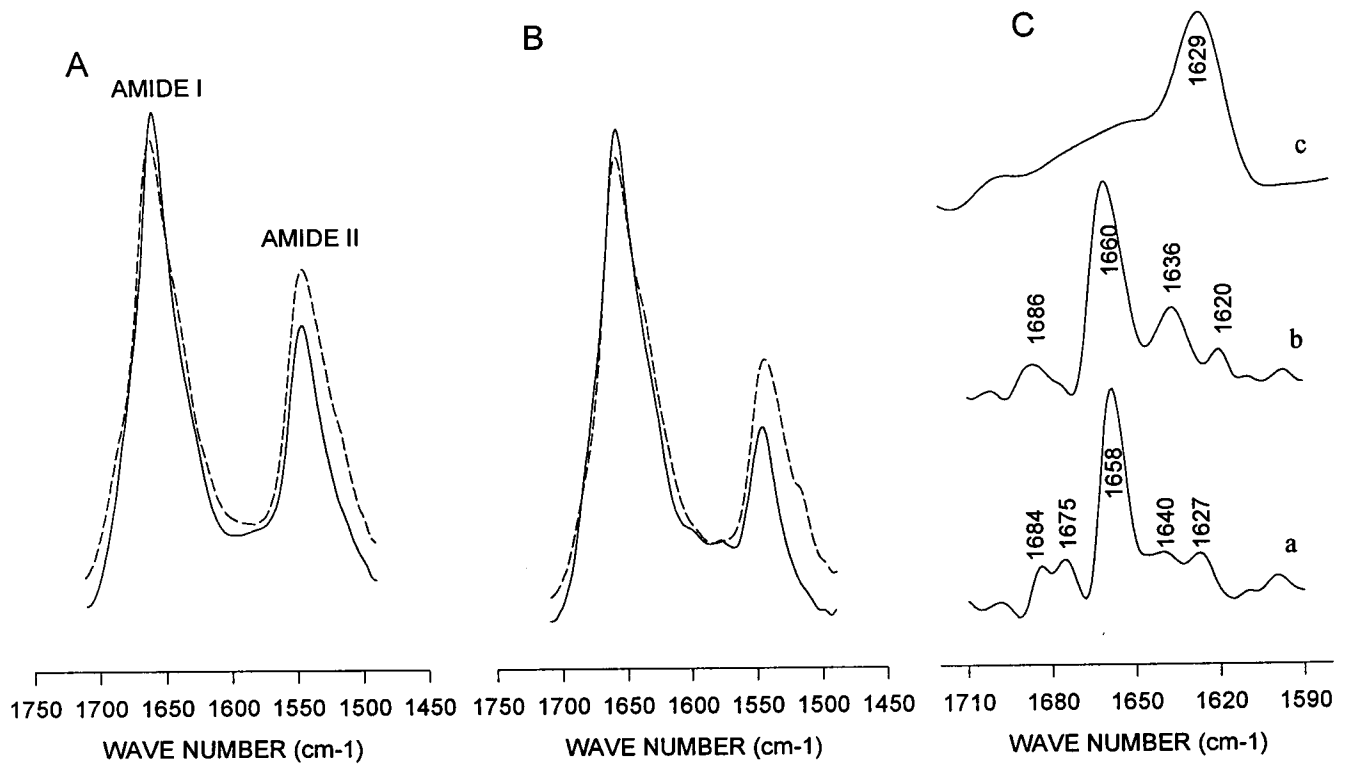

FIGURE 1 FTIR-ATR spectra. (*A*) Thin film (H_2O) spectra of AQP1 reconstituted in proteoliposomes (——) and of BR in purple membranes (– – –). (*B*) Thin film spectra of AQP1 reconstituted in proteoliposomes (——) and of BR in purple membranes (– – –) recorded after an overnight deuteration. (*C*) Fourier self-deconvolution of the deuterated spectra of (*a*) AQP1 reconstituted in proteoliposomes (resolution factor 2), (*b*) BR in the purple membranes (resolution factor 2), and (*c*) porin reconstituted in dimyristoylphosphatidylcholine vesicles, as described by Goormaghtigh et al. (1990) (resolution factor 1.8).

When the area of the 1636 cm^{-1} band identified in the resolution enhancement and curve-fitting procedure is combined with the α-helix bands, the total calculated helix content for BR rises to 68%. However, because such an assignment depends on some prior knowledge of the structure of the protein being analyzed, it is not of practical use in the analysis of proteins of unknown structure. The FA/RMR and partial least-square-attenuated total reflection (PLS-ATR) methods, which do not depend on frequency assignments, seem to compensate for this problematic spectral feature because they rely on the overall similarity of the spectral bandshapes of studied and reference proteins. Thus these methods yield the highest unbiased estimates of the helical content in BR. The results from all analyses of AQP1 spectra indicate that it is a protein with a highly helical character and little or no β-sheet content. The proportion of α-helical structure in AQP1 reconstituted in proteoliposomes determined by all of the methods presented here (including previously published data) ranges from 38% to 48%. Based on the number of residues in this protein, these values correspond to 102–129 residues in a helical conformation.

Analysis of BR and AQP1 FD spectra by the neural network approach described above yielded the secondary structure segment interconnectivity matrix descriptors summarized in Table 2. The main result obtained from this approach is that AQP1 is an all-α-protein with six helical domains connected by loop ("other") segments. The final descriptors are marked as the correction results in Table 2. Because AQP1 was found to be an all-α-helical protein, there is only one possible layout of the predicted number of helical and "other" segments consistent with the predicted $[s_{ij}]$ descriptor for the protein, which is shown in Table 2. More importantly, the results of this analysis are consistent with the FC analyses of the spectra. First, by two different statistical analysis methods (PLS-ATR and FA/RMR), we have found that there are no spectral features attributable to β-sheet segments in the AQP1 spectra. This is in accordance with the close bandshape similarity of the AQP1 and BR FD spectra as determined by cluster analysis. Second, the added structural information in the $[s_{ij}]$ matrix descriptor allows us to check the consistency of the complete set of spectral analysis methods applied here by combining the six helical domains determined by $[s_{ij}]$ with the number of helical residues found by the other methods. We can thus estimate the average length of each of the six helical domains in AQP1 as 17–21.5 (i.e., 102/6 to 129/6) amino acid residues. This purely computational result is in excellent agreement with the physical constraint of the helical segment length needed to span a membrane, which is 18–20 residues.

Orientation

Information about the orientation of a given secondary structure can be obtained by recording ATR-FTIR protein

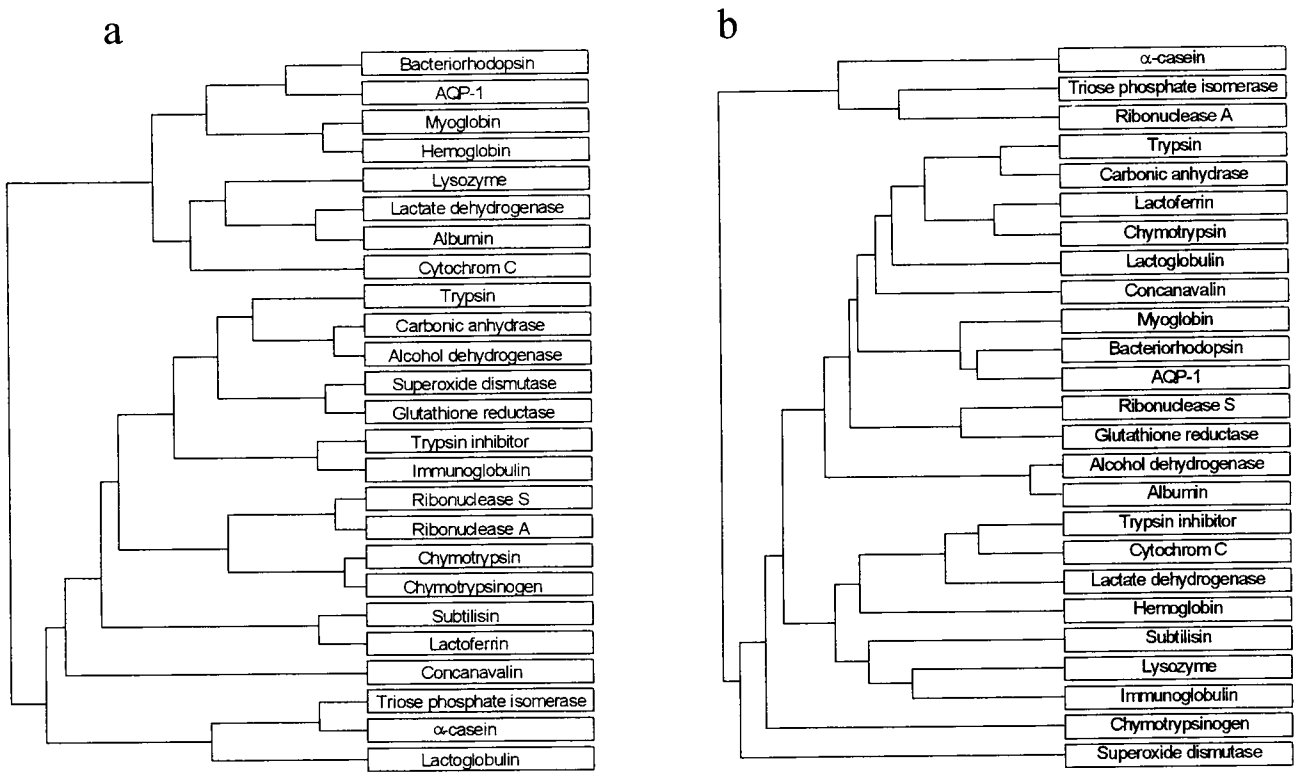

FIGURE 2 (*a*) Cluster analysis of FD spectra of the reference set of globular proteins used for FA/RMR and neural network analyses of AQP-1 and bacteriorhodopsin. Clustering was calculated using the group-averaged algorithm as described, e.g., by Massart and Kaufman (1983); the dendrogram is presented in the relative similarity distance scale. (*b*) Cluster analysis of the reference set of globular proteins with AQP-1 and bacteriorhodopsin, using principal component coefficients needed for the reconstruction of their spectra in PC/FA algorithm. Clustering was calculated using the group-averaged algorithm; the dendrogram is presented in the relative scale of similarity distances.

spectra with polarized light, provided that the proteins are oriented with respect to the internal reflection element (the germanium plate). This prerequisite is met with films formed from membrane proteins reconstituted in lipid vesicles (Fringeli and Günthard, 1981). We have confirmed this by acquiring images of purple membrane films on germanium plates with an atomic force microscope (data not shown). Fig. 3 displays the 90° and 0° polarization spectra

TABLE 3 Determination of secondary structure of AQP1 and BR

Method	Reference	% α (±SD %)	% β	% turns
FSD and curve fitting (Cabiaux et al., 1989)	This paper	AQP1: 48 ± 5	18	21
		BR: 51	18	21
Second derivative and curve fitting (Oberg and Fink, 1995)	This paper	AQP1: 45 ± 5	20	24
		BR: 52	21	19
PLS-ATR	This paper	AQP1: 48 ± 7*	0	11
		BR: 67 ± 7*	0	10
Pattern recognition (FA/RMR algorithm) (Baumruk et al., 1996)	This paper	AQP1: 42 ± 8*	0	19
		BR: 58 ± 8*	1	18
Factor analysis (Lee et al., 1990)	Haris et al. (1995) (FTIR)	AQP1: 38 ± 4	45	19
Singular value decomposition (Sarver and Krueger, 1990)	van Hoeck et al. (1993) (ATR-FTIR)	AQP1: 38 ± 5	18	22
Electron diffraction	Grigorieff et al. (1996)	BR: 68–70		

*Standard deviation σ, based on optimization of a basis set of reference protein spectra, as described under Materials and Methods. The other deviations are calculated on the basis of four independent experiments.

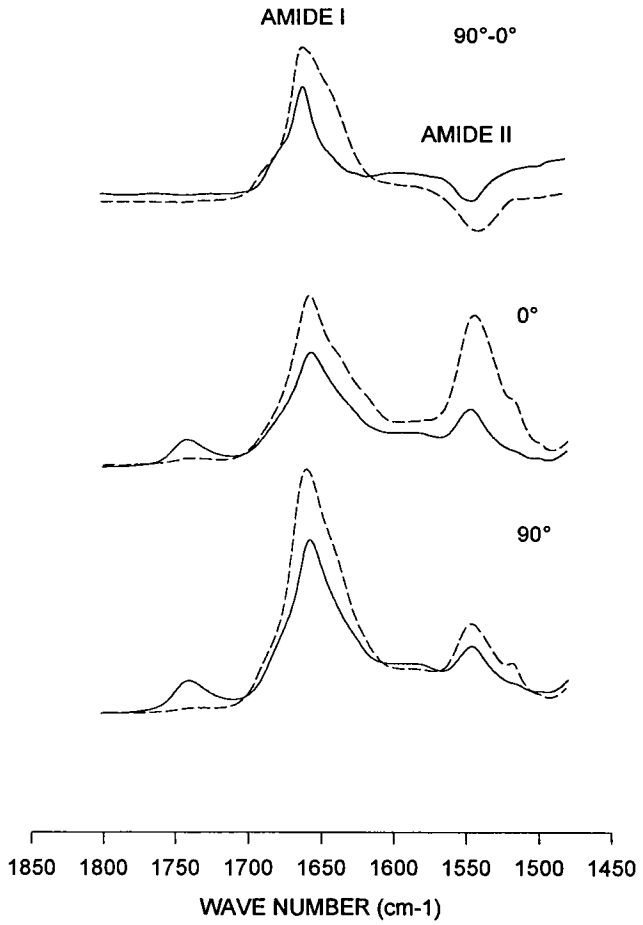

FIGURE 3 Polarized deuterated spectra of AQP1 reconstituted in proteoliposomes (——) and of BR in purple membranes (– – –). The subtraction coefficient was chosen to cancel the $\upsilon(CH_2\text{-}CH_3)$ between 3000 and 2824 cm^{-1}. The difference spectra were enlarged 1.7 and 4 times as compared to the 90° spectrum for BR and AQP1, respectively.

of AQP1 (*solid line*) and BR in purple membranes (*dashed line*) as well as the 90°-0° difference spectrum. Because of differences in the relative power of the 90° and 0° polarized evanescent fields, the subtraction coefficient (R_{iso}, see below) differs from unity in ATR spectroscopy and is best assessed experimentally (Goormaghtigh and Ruysschaert, 1990). In the present case, the subtraction coefficient (R_{iso}, see below) was chosen to zero the $\upsilon(CH_2\text{-}CH_3)$ bands between 3000 and 2824 cm^{-1} instead of the lipid $\upsilon(C{=}O)$ band (1763–1711 cm^{-1}), as described previously (Cabiaux et al., 1989). This was necessary because very little absorption was observed in the lipid $\upsilon(C{=}O)$ frequency range (1750–1700 cm^{-1}) of BR because of the high protein/lipid ratio in the purple membranes. For the AQP1 samples, we found very similar coefficients using the lipid $\upsilon(C{=}O)$ and the $\upsilon(CH_2\text{-}CH_3)$ bands (data not shown).

A positive deviation in a polarization difference spectrum indicates an orientation of the considered dipole parallel to the lipid acyl chains. For both proteins, a positive deviation was observed in the amide I (C=0) region of the difference spectrum, whereas a negative deviation was observed in the amide II (N-H) region. Both of these facts indicate that the α-helical component of AQP1 has a strong transmembrane orientation. Quantitative evaluation of the helix orientation can be obtained from the integrated dichroic ratio by using the thin film hypothesis and either 1.7 or 1.55 as the refractive index of the film (see below) (Goormaghtigh and Ruysschaert, 1990). The dichroic ratio of the α-helical component was calculated with the following equation (Raussens et al., 1997), assuming that α-helices are the only strongly oriented components:

$$R_\alpha = \frac{R - [(R + 2)/(2R_{iso} + 1)](1 - x)}{1 - (1/R_{iso})\,[(R + 2)/(2R_{iso} + 1)](1 - x)}$$

where R_α is the dichroic ratio of the α-helical component, R is the measured dichroic ratio (area amide I 90°/area amide I 0°), R_{iso} is the ratio of the intensities at 90° and 0° of a dipole with an isotropic orientation (the $\upsilon(CH_2\text{-}CH_3)$ between 3000 and 2824 cm^{-1}), and x is the fraction of α-helical structure. A value of x = 70% was used for the percentage of α-helix in bacteriorhodopsin (Grigorieff et al., 1996). Because there is some variation in the determined percentages of α-helical structure of AQP1, the dichroic ratio was calculated using values of 40% and 50%. The dichroic ratio determined for the BR spectrum was 3.5, whereas four independent experiments with reconstituted AQP1 gave an average value of 2.0 ± 0.25 for both x = 40% and 50%. As a control, we also calculated the dichroic ratio of AQP1 amide I using the lipid $\upsilon(C{=}O)$ area to determine R_{iso}, and obtained the same value within the error limits.

The relationship between the dichroic ratio and the orientation of a given structure (α-helix in this case) is dependent upon the order parameter, which is in turn strongly dependent upon the angle between the transition dipole being considered and the primary axis of the structure (Goormaghtigh et al., 1994a). A dichroic ratio of 3.5, as found for BR, can only be obtained with an order parameter of 1 and a value of 17° for the angle between the C=O dipole and the helical axis (for a discussion of this angle, see Goormaghtigh et al., 1994a). Another parameter influencing the angle determination is the refractive index (n) of the film. Values of 1.7 and 1.55 are currently used for protein and lipid films, respectively (Goormaghtigh and Ruysschaert, 1990). A dichroic ratio of 3.5 corresponds to a maximum average tilt of the α-helices, with respect to the membrane surface normal, of 4° (n = 1.7) or 13° (n = 1.55). This result is in excellent agreement with the average angle of 13° found in the BR structure (Grigorieff et al., 1996), suggesting that the thin film hypothesis is valid for our analysis. When the same C=O dipole-helical axis angle and order parameter values are used to calculate the average orientation of the helices in AQP1, the maximum average angle to the membrane normal was found to be 21 ± 4° (n = 1.7) or 27 ± 4° (n = 1.55) (four independent experiments). Thus our data suggest that the helices in AQP1 are

significantly more tilted than those in BR. We note that 21° or 27° is the maximum possible angle, because values for the order parameter smaller than 1, or a C=O dipole-helical axis angle higher than 17° would decrease the calculated helix axis-membrane normal angle.

H/D exchange

At constant experimental conditions (pH, temperature, etc.), hydrogen/deuterium exchange is related to the solvent accessibility of the protein and/or to the stability of its secondary structure. Membrane proteins are expected to be resistant to exchange because much of their sequence is inserted in the bilayer and therefore is inaccessible to solvent. Such a low level of exchange has previously been reported for BR, in which more than 70% of the amide protons remain unexchanged after 48 h of exposure to D_2O (Earnest et al., 1990). Fig. 4 illustrates a similar behavior under our experimental conditions: ~71% of the BR residues remained unexchanged after 200 min of exposure to D_2O. The extent of exchange for AQP1 was also quite low: 60% of the residues remained unexchanged after 200 min of deuteration.

Characterization of an AQP1 proteolyzed sample

Treatment of an AQP1 sample by carboxypeptidase Y leads to the removal of a C-terminal fragment of ~5 kDa (Smith and Agre, 1991). As shown in Fig. 5 *A*, the spectrum of proteolyzed AQP1 has the same features as the spectrum of AQP1. Using the resolution enhancement and curve fitting procedure, we found that after digestion the α-helix content increased to 49% with a concomitant decrease in the "random" component. This indicates that the C-terminal peptide adopts a mainly disordered conformation. The location of the C-terminus in the protruding cytosolic domain (Walz et al., 1996) that is likely to become disordered during sample preparation corroborates this observation. In two independent experiments, proteolyzed AQP1 was found to undergo deuterium exchange to a lesser extent than the intact protein (Fig. 5 *C*), ~67% of residues being protected from exchange. The dichroic ratios (Fig. 5 *B*) calculated from two independent experiments with the proteolyzed AQP1 were 2.2 and 1.87 (x = 50%), and 2.3 and 1.9 (x = 40%) (x is the percentage of α-helical structure; see above). These dichroic ratios are in the same range as those calculated for AQP1, which suggests that the helices in AQP1 retain their transmembrane orientation after digestion.

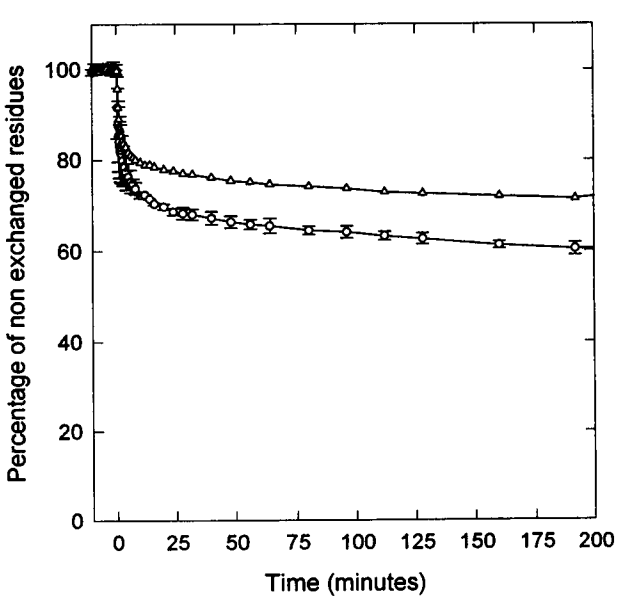

FIGURE 4 Evolution of nonexchanged residues of AQP1 reconstituted in proteoliposomes (○) and BR in purple membranes (△) as a function of deuteration time. The percentage of nonexchanged residues is evaluated from the ratio of (amide II area)/(amide I area) as described under Materials and Methods. The AQP1 curve is the mean of four independent experiments, and the error bars represent the standard deviation.

DISCUSSION

Various determinations of AQP1 secondary structure have led to controversies. Projection maps of 2D crystals reconstituted in lipid membranes suggest the presence of six to eight helices in the protein (Jap and Li, 1995; Walz et al., 1995). A previous study using ATR-FTIR found that partially purified AQP1 contains ~40% α-helical structure, but also a significant amount (18%) of β-sheet (van Hoeck et al., 1993). A high proportion of β-sheet structure (45%) was found in the FTIR study by Haris et al. (1995). However, this last study should be viewed with caution, because the absorption of the amide I′ (D_2O) band is not at a minimum at 1600 cm^{-1}; this is a strong indication that the signal in this region may be the result of something other than the protein C=O vibration. Finally, on the basis of a predictive algorithm, Fishbarg et al. (1995) have proposed that AQP1 should fold in the membrane as a porin-like 16-stranded antiparallel β-barrel.

In an effort to resolve the contradictions in the literature, a new ATR-FTIR study was initiated using highly purified AQP1 protein in membrane crystalline arrays. In this study we attempted to identify any artifact that could arise from sources such as the use of partially dehydrated samples or the nature of the algorithm used for secondary structure determination, by systematically treating BR membranes the same way as AQP1 membranes. The results we have obtained for BR from infrared spectroscopy agree with the BR structure determined by electron crystallography (Grigorieff et al., 1996), suggesting that the ATR-FTIR technique we have developed can provide reliable information about the structure and orientation of a membrane protein.

Several algorithms were used in this study for secondary structure determination. The two first algorithms (Table 3) rely on a frequency-based structural assignment of the components bands by Fourier self-deconvolution or differentiation. However, these algorithms do not always allow unam-

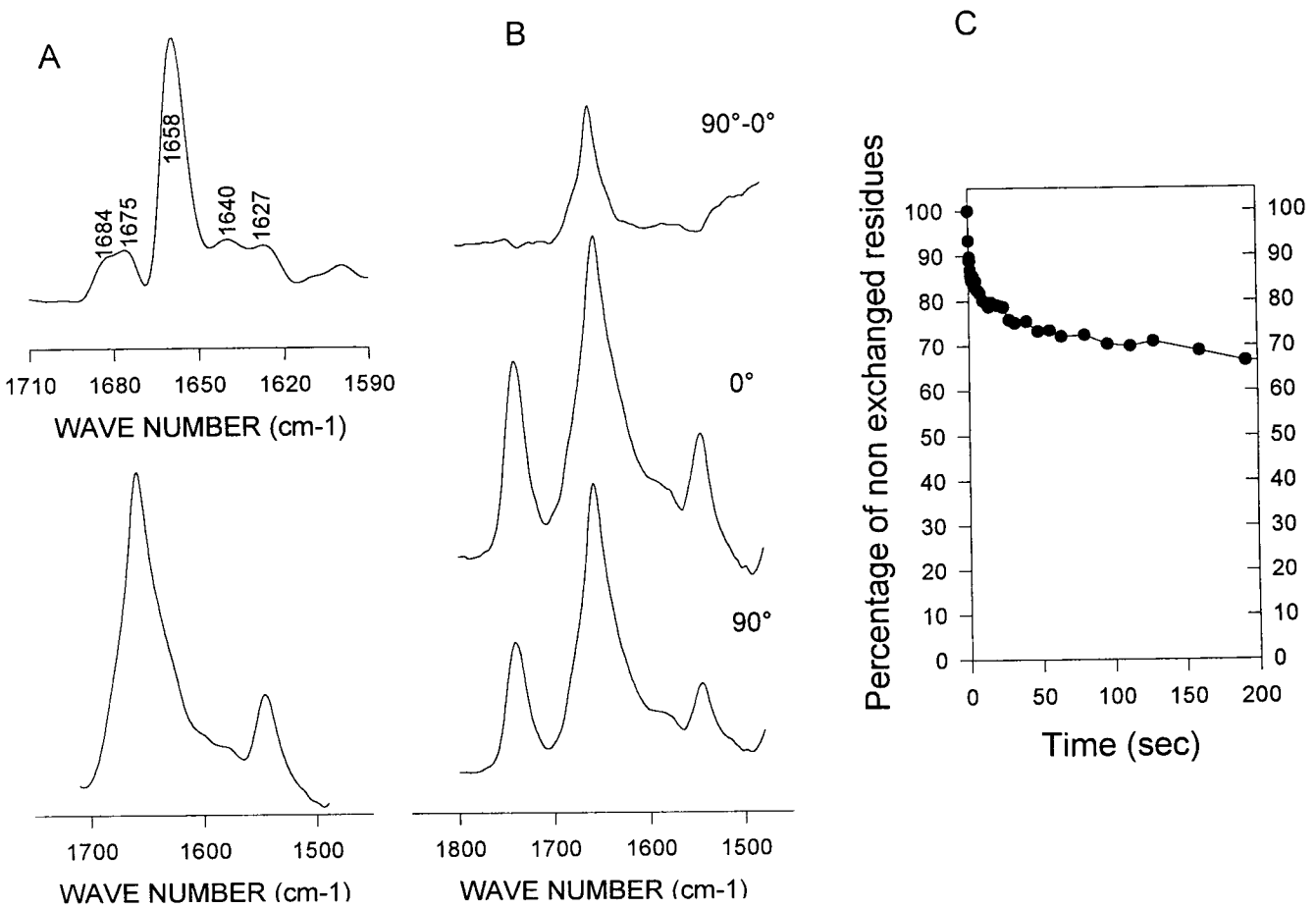

FIGURE 5 Characterization of a proteolyzed AQP1 sample. (*A*) Deuterated FTIR spectra. *Bottom curve*: Original spectrum; *top curve*: deconvoluted spectrum (resolution factor 2). (*B*) Polarization spectra. The difference spectrum has been enlarged three times with respect to the 90° spectrum. (*C*) Evolution of nonexchanged residues as a function of deuteration time.

biguous assignments of a given frequency to a particular secondary structure (Pancoska et al., 1993). The second two algorithms, FA/RMR and PLS-ATR, are methods of bandshape analysis that avoid the need for the assignment of component bands based on frequency; instead they evaluate the overall bandshape of the FTIR spectrum and use its similarity to the reference spectra for the determination of secondary structure fractions.

The AQP1 spectrum is characterized by an absorption maximum at 1658 cm^{-1}, very close to the absorption maximum of BR, both of which are indicative of a high α-helix content. The helix content found in this study for AQP1 ranged from 42% to 48%, depending on the algorithm used for secondary structure determination. As far as β-structure is concerned, no β-sheet was detected by the PLS-ATR method of Oberg et al. (manuscript in preparation) or the FA/RMR algorithm of Pancoska et al. (1994). Values of up to 20% β-sheet were found by using the other algorithms (Table 3), but it must be kept in mind that the set of proteins used as a basis for the frequency assignments does not allow unambiguous discrimination between β-sheets and 3_{10}-helices. Thus 20% should be considered as a maximum value. Therefore we cannot completely eliminate the possibility that a small amount of β-structure may play a role in channel activity. Overall, however, the bulk of our data supports the notion that AQP1 is an all-α-helical protein, its spectral characteristics being close to those of BR. In addition, the $[s_{ij}]$ matrix analysis of the AQP1 spectra suggests that its helical residues are distributed within six helices of sufficient length to span a lipid bilayer. Moreover, the higher percentage of α-helical structure found in the proteolyzed samples shows that the helices are protected against proteolytic cleavage, which in turn suggests that they have a transmembrane orientation. Additional strong evidence for transmembrane orientation is the polarization data. The helices of both BR and AQP1 showed a preferential absorption of 90° polarized light, which demonstrates that the helical axes are oriented parallel to the membrane normal. The data also indicate that the helices in AQP1 have, on average, more tilt than the helices in BR. Finally, the dichroic ratio of the proteolyzed AQP1 samples is very similar to the dichroic ratio of the nonproteolyzed protein, which further reinforces our conclusion that all of the helices in AQP1 adopt a transmembrane orientation.

We have compared the hydrogen/deuterium exchange characteristics of BR and AQP1. As previously reported, BR undergoes a limited hydrogen/deuterium exchange (Earnest et al., 1990). In a recent report, Haris et al. (1995) found a high exchange rate for AQP1, similar to that of the human erythrocyte glucose transporter in which 81% of the hydrogen bonds exchange within 60 min (Alvarez et al., 1987). However, as we show here, the exchange of AQP1 is actually similar to that of BR. The discrepancy between these two studies may arise from the fact that the significant absorbance below 1600 cm^{-1} obtained by Haris et al. (1995) may have interfered with their determination of the amide II area. The low extent of exchange found in this study supports the view that AQP1 has a rigid structure that does not undergo a folding/unfolding reaction during the deuteration time, and that it has a narrow channel with few water-accessible sites, as expected from a pore that allows the rapid passage of water molecules, but not ions or small solutes (Zeidel et al., 1994). Moreover, the hydrogen/deuterium exchange behavior we observed is in agreement with the surface topography of AQP1, as determined by electron microscopy and atomic force microscopy (Walz et al., 1996). These data show that the AQP1 tetramer protrudes by 1.1 nm from the membrane on the cytosolic side and surrounds a narrow central cavity on the other side. Removal of the 5-kDa C-terminal peptide reduces the major protrusion, hence reducing the number of residues exposed to the water. This change is in agreement with our observed decrease in the extent of hydrogen/deuterium exchange of the proteolyzed samples.

In conclusion, our data strongly support the model of an α-helical AQP1 as derived by structure prediction from the sequence (Preston and Agre, 1991), and subsequently tested by epitope insertion experiments (Preston et al., 1994). The main arguments are the high α-helical content of the protein, and the number, length, and orientation of these helices found with the methods applied in this study. None of our data are compatible with a β-barrel topological organization, as seen with the porins, which are characterized by an infrared absorption maximum at 1629 cm^{-1} that preferentially absorbs 90° polarized light (Goormaghtigh et al., 1990).

The authors thank Barbara L. Smith for purification of AQP1 protein and E. Goormaghtigh for helpful discussions.

VC is a research associate of the National Fund for Scientific Research (Belgium). The work at UIC was supported by a grant from the National Institutes of Health (GM30147) and at Johns Hopkins University by HL 33991 and HL 48268. Part of the work was supported by a grant "Télévie" from the National Fund for Scientific Research (Belgium). The work in Basel was funded by the M. E. Müller Foundation, Switzerland, and grant 31-42435.94 to AE from the Swiss National Foundation for Research.

REFERENCES

Alvarez, J., D. C. Lee, S. A. Baldwin, and D. Chapman. 1987. Fourier-tranform infrared spectroscopy study of the structure and conformational changes of the human erythrocyte glucose transporter. *J. Biol. Chem.* 262:3502–3509.

Baumruk, V., P. Pancoska, and T. A. Keiderling. 1996. Predictions of secondary structure using statistical analyses of electronic and vibrational circular dichroism and Fourier transform infrared spectra of proteins in H_2O. *J. Mol. Biol.* 259:774–791.

Cabiaux, V., R. Brasseur, R. Wattiez, P. Falmagne, J. M. Ruysschaert, and E. Goormaghtigh. 1989. Secondary structure of diphtheria toxin and its fragments interacting with acidic liposomes studied by polarized infrared spectroscopy. *J. Biol. Chem.* 264:4928–4938.

Chrispeels, M. J., and P. Agre. 1994. Aquaporins: water channels of plant and animal cells. *Trends Biochem Sci.* 19:421–425.

Cortijo, M., A. Alonso, J. C. Gomez-Fernandez, and D. Chapman. 1982. Instrinsic protein-lipid interactions: infrared spectroscopic studies of gramicidin A, bacteriorhodopsin and Ca^{++} ATPase in biomembranes and reconstituted systems. *J. Mol. Biol.* 157:597–618.

de Jongh, H. H. J., E. Goormaghtigh, and J. M. Ruysschaert. 1995. Tertiary stability of native and methionine-80 modified cytochrome c detected by proton-deuterium exchange using on-line Fourier transform infrared spectroscopy. *Biochemistry.* 35:172–179.

de Jongh, H. H. J., E. Goormaghtigh, and J. M. Ruysschaert. 1996. The different molar absorptivities of the secondary structure types in the amide I region: an attenuated total reflection infrared study on globular proteins. *Anal. Biochem.* 242:95–103.

Denker, B. M., B. L. Smith, F. P. Kuhajda, and P. Agre. 1988. Identification, purification, and partial characterization of a novel M_r 28,000 integral membrane protein from erythrocytes and renal tubules. *J. Biol. Chem.* 263:15634–15642.

Earnest, T. N., J. Herzfeld, and K. J. Rotschild. 1990. Polarized FTIR of bacteriorhodopsin: transmembrane α-helices are resistant to hydrogen deuterium exchange. *Biophys. J.* 58:1539–1546.

Fishbarg, J., J. Li, M. Cheung, F. Czegledy, P. Iserovitch, and K. Kuang. 1995. Predictive evidence for a porin-type β-barrel fold in CHIP28 and other membres of the MIP family. A restricted-pore model common to water channels and facilitators. *J. Membr. Biol.* 143:177–188.

Fringeli, U. P., and H. H. Günthard. 1981. Infrared membrane spectroscopy. *In* Membrane Spectroscopy. E. Grell, editor. Springer Verlag, Berlin. 270–332.

Goormaghtigh, E., V. Cabiaux, and J. M. Ruysschaert. 1990. Secondary structure and dosage of soluble and membrane proteins by attenuated total reflection Fourier-transform infrared spectroscopy on hydrated films. *Eur. J. Biochem.* 193:409–420.

Goormaghtigh, E., V. Cabiaux, and J. M. Ruysschaert. 1994a. Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. *In* Subcellular Biochemistry, Vol. 23, Physicochemical Methods in the Study of Biomembranes. H. J. Hilderson and G. B. Ralston, editors. Plenum Press, New York. 329–450.

Goormaghtigh, E., and J. M. Ruysschaert. 1990. Polarized attenuated total reflection spectroscopy as a tool to investigate the conformation and orientation of membrane components. *In* Molecular Description of Biological Membrane Components by Computer Aided Conformational Analysis. R. Brasseur, editor. CRC Press, Boca Raton, FL. 285–329.

Goormaghtigh, E., L. Vigneron, G. Scarborough, and J. M. Ruysschaert. 1994b. Tertiary conformational changes of the *Neurospora crassa* plasma membrane H^+-ATPase monitored by hydrogen/deuterium exchange kinetics. *J. Biol. Chem.* 269:27409–27413.

Grigorieff, N., T. A. Ceska, K. H. Downing, J. M. Baldwin, and R. Henderson. 1996. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* 259:393–421.

Haris, P. I., D. Chapman, and G. Benga. 1995. A Fourier-transform infrared spectroscopic investigation of the hydrogen-deuterium exchange and secondary structure of the 28 kDa channel-forming integral membrane protein (CHIP28). *Eur. J. Biochem.* 233:659–654.

Janota, V. 1993. Determination of protein structure by mathematical analyses of optical spectra. Thesis. Charles University, Prague.

Jap, B. K., and H. Li. 1995. Structure of the osmo-regulated, AQP-CHIP, in projection at 3.5 Å resolution. *J. Mol. Biol.* 251:413–420.

Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical feature. *Biopolymers.* 22:2577–2637.

Knepper, M. A. 1994. The aquaporin family of molecular water channels. *Proc. Natl. Acad. Sci. USA.* 91:6355–6258.

Lee, D. C., P. I. Haris, D. Chapman, and R. C. Mitchell. 1990. Determination of protein secondary structure using factor analysis of infrared spectra. *Biochemistry.* 29:9185–9193.

Macey, R. I., and R. E. I. Farmer. 1970. Inhibition of water and solute permeability in human red cells. *Biochim. Biophys. Acta.* 211:104–106.

Massart, D. L., and L. Kaufman. 1983. The interpretation of analytical chemical data by the use of cluster analysis. *In* Chemical Analysis, Vol. 65. Wiley, New York.

Miick, S. M., G. V. Martinez, W. R. Fiori, A. P. Todd, and G. L. Millhauser. 1992. Short alanine-based peptides may form 3_{10}-helices and not α- helices in aqueous solution. *Nature.* 359:653–655.

Mitra, A. K., A. N. van Hoek, M. C. Wiener, A. S. Verkman, and M. Yeager. 1995. The CHIP28 water channel visualized in ice by electron crystallography. *Nature Struct. Biol.* 2:726–729.

Mitra, A. K., M. Yeager, A. N. van Hoek, M. C. Wiener, and A. S. Verkman. 1994. Projection structure of the CHIP28 water channel in lipid bilayer membranes at 12 Å resolution. *Biochemistry.* 33: 12735–12740.

Oberg, K. A., B. A. Chrunyk, R. Wetzel, and A. L. Fink. 1994. Nativelike secondary structure in interleukin-1β inclusion bodies by attenuated total reflectance FTIR. *Biochemistry.* 33:2628–2634

Oberg, K. A., and A. L. Fink. 1995. Methods for collecting and analyzing attenuated total reflectance FTIR spectra of proteins in solution. *In* Techniques in Protein Chemistry VI. John W. Crabb, editor. Academic Press, San Diego. 475–484.

Oesterhelt, D., and W. Stoeckenius. 1974. Isolation of the cell membrane of *Halobacterium halobium* and its fractionation into red and purple membrane. *Methods Enzymol.* 31:667–678.

Orengo, C. A., A. D. Michie, S. Jones, M. B. Swindells, D. T. Jones, and J. M. Thornton. 1996. CATH Protein Structure Classification Database, Version 1.0. University College London. http://www.biochem.ucl.ac.uk/bsm/cath/.

Pancoska, P., E. Bitto, V. Janota, and T. A. Keiderling. 1994. Quantitative analysis of vibrational circular dichroism spectra of proteins. Problems and perspectives. *Faraday Discuss.* 99:287–310.

Pancoska, P., H. Fabian, G. Yoder, V. Baumruk, and T. A. Keiderling. 1996. Protein structural segments and their interconnections derived from optical spectra. Thermal unfolding of ribonuclease T1 as an example. *Biochemistry.* 35:13094–13106.

Pancoska, P., L. Wang, and T. A. Keiderling. 1993. Comparison of protein FTIR absorption and vibrational circular dichroism. VCD frequency analyses in terms of secondary structure. *Protein Sci.* 2:411–419.

Preston, G. M., and P. Agre. 1991. Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: member of an ancient channel family. *Proc. Natl. Acad. Sci. USA.* 88:11110–11114.

Preston, G. M., T. P. Carroll, W. B. Guggino, and P. Agre. 1992. Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science.* 256:385–387.

Preston, G. M., J. S. Jung, W. B. Guggino, and P. Agre. 1993. The mercury-sensitive residue at cysteine 189 in the CHIP28 water channel. *J. Biol. Chem.* 268:17–20.

Preston, G. M., J. S. Jung, W. B. Guggino, and P. Agre. 1994. Membrane topology of aquaporin CHIP: analysis of functional epitope-scanning mutants by vectorial proteolysis. *J. Biol. Chem.* 269:1668–1673.

Prestrelski, S. J., D. M. Byle, and M. N. Liebman. 1991. Comparison of various molecular forms of bovine trypsin: correlation of infrared spectra with x-ray crystal structures. *Biochemistry.* 30:133–143.

Prestrelski, S. J., D. M. Byler, and M. P. Thompson. 1990. Infrared spectroscopy discrimination between α- and 3_{10}-helices in globular proteins: reexamination of amide I infrared bands of α-lactalbumin and their assignment to secondary structures. *Int. J. Pept. Protein Res.* 29:508–512.

Raussens, V., V. Narayanaswami, E. Goormaghtigh, R. O. Ryan, and J. M. Ruysschaert. 1996. Hydrogen/deuterium exchange kinetics of apolipophorin-III in lipid-free and phospholipid-bound states. *J. Biol. Chem.* 271:23089–23095.

Raussens, V., J. M. Ruysschaert, and E. Goormaghtigh. 1997. Fourier transform infrared spectroscopy study of the secondary structure of the gastric H^+,K^+-ATPase and of its membrane-associated proteolytic peptides. *J. Biol. Chem.* 276:262–270.

Sarver, R. W., and W. Krueger. 1990. Protein secondary structure from Fourier-transform infrared spectroscopy: a database analysis. *Anal. Biochem.* 194:89–100.

Smith, B. L., and P. Agre. 1991. Erythrocyte M_r 28,000 transmembrane protein exists as a multisubunit oligomer similar to channel proteins. *J. Biol. Chem.* 266:6407–6415.

van Hoeck, A. N., M. L. Hom, L. H. Luthjens, M. D. deJong, J. A. Dempster, and C. H. van Os. 1991. Functional unit of 30 kDa for proximal tubule water channel as revealed by radiation inactivation. *J. Biol. Chem.* 266:16633–16635.

van Hoeck, A. N., M. Wiener, S. Bicknese, L. Miercke, J. Biwersi, and A. S. Verkman. 1993. Secondary structure analysis of purified functional CHIP28 water channels by CD and FTIR spectroscopy. *Biochemistry.* 32:11847–11856.

Walz, T., B. L. Smith, P. Agre, and A. Engel. 1994a. The three-dimensional structure of human erythrocyte aquaporin CHIP. *EMBO J.* 13:2985–2993.

Walz, T., B. L. Smith, M. L. Zeidel, A. Engel, and P. Agre. 1994b. Biologically active two-dimensional crystals of aquaporin CHIP. *J. Biol. Chem.* 269:1583–1586.

Walz, T., P. Tittman, K. H. Fuchs, D. Müller, B. L. Smith, P. Agre, H. Gross, and A. Engel. 1996. Surface topographies at subnanometer-resolution reveal asymmetry and sidedness of aquaporin-1. *J. Mol. Biol.* 264:907–918.

Walz, T., D. Typke, B. L. Smith, P. Agre, and A. Engel. 1995. Projection map of aquaporin-1 determined by electron crystallography. *Nature Struct. Biol.* 2:730–732.

Yang, W. J., and P. R. Griffiths. 1984. Optimization of parameters for Fourier self-deconvolution. 2. Band multiplets. *Computer-Enhanced Spectrosc.* 2:69–74.

Zeidel, M. L., S. V. Ambudkar, B. L. Smith, and P. Agre. 1992. Reconstitution of functional water channels in liposomes containing purified red cell CHIP28 protein. *Biochemistry.* 31:7436–7440.

Zeidel, M. L., S. Nielsen, B. L. Smith, S. V. Ambudkar, A. B. Maunsbach, and P. Agre. 1994. Ultrastructure, pharmacologic inhibition, and transport selectivity of aquaporin CHIP in proteoliposomes. *Biochemistry.* 33:1606–1615.

Zhang, Y. P., R. N. A. H. Lewis, G. D. Henry, B. D. Sykes, R. S. Hodges, and R. N. McElhaney. 1995. Peptide models of helical hydrophobic transmembrane segments of membrane proteins. 1. Studies of the conformation, intrabilayer orientation and amide hydrogen exchangeability of Ac-K_2-$(LA)_{12}$-K_2-amide. *Biochemistry.* 34:2348–2361.