J. Mol. Biol. (1996) **259**, 393–421

JMB

Electron-crystallographic Refinement of the Structure of Bacteriorhodopsin

N. Grigorieff[1], T. A. Ceska[1], K. H. Downing[2], J. M. Baldwin[1] and R. Henderson[1]

[1]*MRC Laboratory of Molecular Biology, Hills Road Cambridge CB2 2QH, UK*

[2]*Life Science Division Lawrence Berkeley National Laboratory, Berkeley CA 94720, USA*

Using electron diffraction data corrected for diffuse scattering together with additional phase information from 30 new images of tilted specimens, an improved experimental density map has been calculated for bacteriorhodopsin. The atomic model has then been rebuilt into this new map with particular attention to the surface loops. All the residues from 7 to 227 as well as ten lipid molecules are now included, although a few amino acid residues in three of the six surface loops, about half of the lipid hydrophobic chains and all of the lipid head groups are disordered. The model has then been refined against the experimental diffraction amplitudes to an *R*-factor of 28% at 3.5 Å resolution with strict geometry (0.005 Å bond length deviation) using the improvement of the ''free'' phase residual between calculated and experimental phases from images as an objective criterion of accuracy. For the refinement some new programs were developed to restrain the number of parameters, to be compatible with the limited resolution of our data.

In the final refined model of the protein (2BRD), compared with earlier co-ordinates (1BRD), helix D has been moved towards the cytoplasm by almost 4 Å, and the overall accuracy of the co-ordinates of residues in the other six helices has been improved. As a result the positions of nearly all the important residues in bacteriorhodopsin are now well determined. In particular, the buried, protonated Asp115 is 7 Å from, and so not in contact with, the retinal and Met118 forms a cap on the pocket occupied by the β-ionone ring. No clear density exists for the side-chain of Arg82, which forms a central part of the extracellular half-channel. The only arginine side-chain built into good density is that of Arg134 at the extracellular end of helix E, the others being disordered near one of the two surfaces. The interpretation of the end of helix F on the extracellular surface is now clearer; an extra loose helical turn has been built bringing the side-chain of Glu194 close to Arg134 to form a probable salt bridge. The model provides an improved framework for understanding the mechanism of the light-driven proton pumping. A number of cavities that could contain water molecules were found by searching the refined model, most of them above or below the Schiff base in the half-channels leading to the two surfaces. The ordered and disordered regions of the structure are described by the temperature factor distribution.

Keywords: bacteriorhodopsin; purple membrane; refinement; two-dimensional crystal; electron microscopy

Present address: T. A. Ceska, EMBL, Heidelberg, Germany.

Abbreviations used: bR, bacteriorhodopsin; PM, purple membrane; 1BRD, reference number for previous co-ordinates of bR in the Brookhaven Databank; 2BRD, reference number for refined co-ordinates in the Brookhaven Databank; IAM, independent atom model; RMS, root-mean-square; HK, Hendrickson–Konnert; CSF, cubic spline function; NMR, nuclear magnetic resonance; FTIR, Fourier transform infrared; FOM, figure of merit; PDB, format of co-ordinates for the Brookhaven Databank; CCP4, Collaborative Computational Project, number 4; H-bond, hydrogen bond.

0022–2836/96/230393–29 $18.00/0

Introduction

Bacteriorhodopsin (bR), the light-driven proton pump from Halobacteria, functions through a sequence of conformational changes involving the protein and the retinal chromophore (for reviews, see Mathies *et al*. (1991), Oesterhelt *et al*. (1992), Rothschild (1992), Ebrey (1993) and Lanyi (1993)). After absorbing a photon at 568 nm (light-adapted form) bR proceeds through a number of intermediate states originally characterised by their absorption in the visible (Lozier *et al*., 1975; Sharkov *et al*., 1985; Polland *et al*., 1986; Petrich *et al*., 1987):

$$bR_{568} + h\nu \rightarrow J_{610} \rightarrow K_{590} \leftrightarrow L_{550}$$

$$\leftrightarrow M_{412} \leftrightarrow N_{520} \leftrightarrow O_{640} \rightarrow bR_{568}$$

The driving force initiated by light absorption is isomerisation of all-*trans* retinal to 13-*cis* (Braiman & Mathies, 1982). The protonated Schiff base between the retinal and Lys216 releases its proton to Asp85 in the L to M transition (Braiman *et al*., 1988; Mogi *et al*., 1988; Gerwert *et al*., 1990) and is reprotonated in the M to N transition (Fodor *et al*., 1988; Gerwert *et al*., 1990). Reprotonation of the Schiff base from Asp96 (Butt *et al*., 1989; Gerwert *et al*., 1989; Otto *et al*., 1989) is thought to occur indirectly through a 12 Å long chain of intermediary groups (Henderson *et al*., 1990). Asp96 is protonated in bR_{568} and is believed to stay protonated up to and including M (Bousché *et al*., 1991). With the rise of M a proton is released on the extracellular side of the membrane (Alexiev *et al*., 1994) at a pH above 5.8, even though Asp85 is thought to remain protonated. The origin of the released proton is, therefore, assumed to be a surface group, XH, on the extracellular side (e.g. a side-chain with a pK_a of 5.8; Souvigner & Gerwert, 1992; Zimányi *et al*., 1992), which has recently been suggested to be, at least partly, Glu204 (Brown *et al*., 1995a). However, X may also be a distributed protonatable network in which a single proton hops from one position to another without leaving the extracellular channel until Asp85 is protonated in M. Asp96 is believed to reprotonate with the decay of N when a proton is taken up on the cytoplasmic side (Zimányi *et al*., 1993). Finally, the retinal is thought to re-isomerise during the formation of O (Smith *et al*., 1983) after which Asp85 loses its proton to X (Pfefferlé *et al*., 1991; Souvigner & Gerwert, 1992) and the protein returns to the bR ground state. Thus, the net result of each photocycle is the transfer of a single proton from the cytoplasmic to the extracellular side of the membrane.

The key features of any membrane pump are a change in the affinity of at least one binding site for the pumped species coupled with a change in accessibility to the two sides of the membrane (Jardetzky, 1966). To understand the mechanism of proton pumping in bR, it is these features that must be explained by any observed conformation changes. Through extensive characterisation of mutated bacteriorhodopsins, the Schiff base has been identified as the group whose pK_a changes during the photocycle. The pK_a is high in the ground state, drops after isomerisation and formation of the L intermediate to allow proton release on the extracellular side, and regains its high affinity later in the photocycle. The change in the accessibility of the Schiff base to protons from the two sides of the membrane, which can also be called the connectivity, the gating or the switching of the pump, is thought to occur immediately after proton release from the Schiff base on the extracellular side. An understanding of the causes of the Schiff base pK_a changes and the accessibility changes requires a knowledge of the structure and structural changes in sufficient detail to allow approximate estimates of the relative importance of the interactions between the protein and the retinal, and also between the protein and the large number of water molecules that fill the two half-channels leading from the Schiff base to the two surfaces. Clearly, this requires the determination of an atomic resolution structure for bR including any bound or ordered water molecules, with an accuracy comparable with that obtained for the best protein structures. Since the structural changes that parallel the gating are large (Dencher *et al*., 1989; Koch *et al*., 1991; Nakasako *et al*., 1991; Subramaniam *et al*., 1993), it is clear that similar high-resolution structures, including any bound or ordered water molecules, will be required for the intermediates.

Attempts to obtain such a high-resolution structure still fall short of what is needed, but have followed two paths. On the one hand, extensive efforts have been made to coax the protein to form well-ordered three-dimensional crystals suitable for high resolution X-ray crystallography (Henderson & Shotton, 1980; Michel & Oesterhelt, 1980; Michel, 1982; Schertler *et al*., 1993). However, three-dimensional crystals diffracting beyond 2 Å resolution and without other problems are still some way off. On the other hand, gradual improvement in the resolution obtained by electron crystallography of two-dimensional crystals (Henderson & Unwin, 1975; Unwin & Henderson, 1975; Leifer & Henderson, 1983; Henderson *et al*., 1986; Tsygannik & Baldwin, 1987) has produced a rough atomic model (Henderson *et al*., 1990), which has provided a useful framework to help us understand the wealth of other available data.

Earlier work (Henderson *et al*., 1990), which presented the first interpretable three-dimensional map of bR showing chemical detail, was based on the analysis of 70 electron microscopic images of tilted *p*3 crystals. In that map, the seven-helix structure for bR found earlier (Henderson & Unwin, 1975) was confirmed, and clearly resolved density peaks for most of the bulky, aromatic side-chains allowed the detailed interpretation of most of the map in terms of the amino acid sequence (Khorana *et al*., 1979; Ovchinnikov *et al*., 1979) of the protein. This allowed the building of the atomic model of

bR and the deposition of co-ordinates in the Brookhaven Protein Databank (1BRD). The model showed a hydrophobic binding pocket for the retinal lined with 21 residues including four tryptophan and two or three tyrosine residues, and provided a complete list of residues lining the two half-channels leading from the Schiff base to each membrane surface. The channel to the extracellular surface was wider, likely to be filled with water molecules, and contained many charged and hydrophilic side-chains. The cytoplasmic channel, in contrast, was narrow, hydrophobic and contained Asp96 as the only potentially charged side-chain. However, there were also many questions left open by the earlier work. Helix D contains no aromatic side-chains, and at the resolution then reached (3.5 Å in plane, 7.8 Å perpendicular), the interpretation of the density was less certain, so that the errors in the positions of, for example, Asp115 and Met118 on helix D were higher than for residues in the other six helices, especially in their vertical *z*-coordinate where the resolution, due to limited tilt data, was poor. Similarly, the interpretation of the structure of the hydrophilic loop sequences (three on each of the two surfaces), and the N and C termini was so uncertain that these regions were not included in the 1BRD co-ordinate list. Some of the surface peptides are genuinely disordered, for example residues 1 to 6 and 228 to 248 at the two termini, as well as several residues in the EF loop, since both the earlier map and that presented later in this paper show no density at all in these regions. Other loops contained weak density, which indicated partial disorder, but the density for both ordered and partly ordered surface loops needed to be improved.

During the photocycle, substantial structural changes in the M-intermediate occur as indicated by approximately 10% changes in the structure factor amplitudes observed by neutron diffraction (Dencher *et al*., 1989), X-ray diffraction (Koch *et al*., 1991; Nakasako *et al*., 1991) and electron diffraction (Subramaniam *et al*., 1993; Han *et al*., 1994). These changes appear to represent an increased ordering of the cytoplasmic C terminus of helix G and a tilt of the cytoplasmic end of helix F, outwards from the centre of the bR molecule (Subramaniam *et al*., 1993), which would be consistent with an opening of the cytoplasmic half-channel in the second half of the photocycle to allow reprotonation of the Schiff base from Asp96 followed by reprotonation of Asp96 from the cytoplasmic surface. Clearly, a full understanding of the mechanism will require a high-resolution three-dimensional structure for this and other intermediates as well as for the resting bR molecule.

Considerations Specific to Electron Crystallographic Refinement

The methods for refinement of structures solved by X-ray diffraction are well established and similar procedures can be used to refine against electron diffraction data. On the other hand, the phase information that is available from images taken with the electron microscope is not normally available in X-ray diffraction and is not normally used in refinement. The phase information can serve as an independent measure of the accuracy of the refined model (free phase residual) as explained in Methods. Owing to the different nature of both electron diffraction and two-dimensional crystals compared with X-ray diffraction and three-dimensional crystals, data obtained by electron microscopy or electron diffraction are more complex to interpret. In the early days, this led to some significant errors in structure determinations and gave electron diffraction a reputation for unreliability, leaving X-ray diffraction as the simpler and more direct method of choice for atomic resolution crystal structure determination. For this reason we thought it was very important to investigate or consider three sources of this complexity, which can also, in favourable cases, yield additional information on the structure.

Atomic electron scattering cross-section (electron form factor)

X-rays are scattered by electrons in the atomic shell and at low resolution (below 3 Å) scattering is fairly insensitive to details of the electronic distribution in the atom. At zero scattering angle the cross-section for X-rays depends solely on the number of electrons present. Only at very high resolution (beyond 2 Å or so) are there significant changes in the X-ray diffraction intensities due to the altered distribution of electrons in the outer shells of bonded atoms. The situation is different with electron diffraction, since an electron is scattered by the atomic Coulomb potential. The potential inside an atom arises from the negatively charged electron cloud as well as from the positively charged nucleus. In an overall neutral atom, the net result is very sensitive to the mean radius of the counteracting electric charges and at low resolution is proportional to the atomic volume or integrated potential. Therefore, the magnitude of the electron scattering cross-section (electron form factor) changes significantly at low resolution (up to about 2.5 Å) when the electron distribution inside the atom is modified, e.g. due to chemical bonding with another atom. Electron form factors are altered even more severely when an atom carries unshielded charge, e.g. the negative charge on the oxygen atoms in a carboxyl group. The changes become smaller at higher resolution (larger scattering angles) because scattering is then dominated by the positively charged unscreened nucleus. Changes in the radii of the valence electrons occur at a larger radius and, therefore, affect diffraction at lower resolution. Thus, electron diffraction data at a resolution of about 3 Å contain significant information on the nature of the chemical bonding of atoms, which can be neglected in X-ray diffraction

at this resolution. In the refinement process where differences between observed scattering intensities and calculated scattering intensities from the atomic model are minimised, a refinement of electron form factors gives additional detail on chemical bonding and charged residues. However, there are not enough accurate observations in this diffraction data set for bR to be certain about individual bonding effects, and only global scattering factors for similar atom types can be treated. On the other hand, ignoring significant deviations of form factors of atoms in the sample from the tabulated neutral atom form factors (independent atom model, IAM; Cowley, 1992) may result in inaccurate refinement of the positions of atoms in the structure. For this reason a small number of parameters describing the electron form factors were included in the refinement of the structure presented here. Further details on the refinement can be found in Methods.

Multiple (dynamical) scattering

Electron scattering cross-sections are about five orders of magnitude larger than those for X-rays. The strong interaction of electrons with matter is one of the key properties of electrons enabling the investigation of very thin specimens, e.g. two-dimensional protein crystals. With the larger cross-section the chance of an electron being scattered more than once while passing through the specimen is greatly increased. Multiple scattering alters the amplitude and phase of reflections from those expected in the single scattering (kinematical or Born) approximation (Fujiwara, 1959). In thick crystals, e.g. a silicon crystal a few thousand Å thick, multiple scattering causes the single scattering approximation to break down completely. The effect is not so extreme in thin protein crystals, due to their small thickness, their large unit cell size compared to that of a typical semiconductor, and their composition of light atoms. Measurement of the average difference in the intensity of Friedel-related reflections due to multiple scattering in the *p*3 form of bR shows a value of about 6%, for 120 keV electrons, in data to 7 Å resolution (Glaeser & Ceska, 1989). Calculation of the average Friedel difference shows a value of 11% for 100 keV electrons at 3.5 Å resolution (Glaeser & Downing, 1993). This means that the single scattering approximation is reasonably well obeyed and inaccuracies in the refined structure of bR should be small. Furthermore, averaging of Friedel-related reflections reduces the net perturbation introduced by multiple scattering. Glaeser & Downing (1993) estimated the residual dynamical scattering effect after Friedel averaging at 1%, and in the present study we have ignored the problem. However, a careful analysis of the small changes to the amplitudes of Friedel-related reflections affected by multiple scattering could provide additional phase constraints. Further work needs be done to investigate this possibility.

Diffuse scattering from partially disordered crystals

Diffuse background in electron diffraction patterns from thin protein crystals is caused by inelastic scattering, by electrons scattered by the carbon support film and the protein-embedding medium (e.g. vitrified ice or glucose), and by partial disorder in the protein crystal (Grigorieff & Henderson, 1995). After measurement of the integrated intensity of a reflection from the film, the diffuse background must be subtracted. This can be done very efficiently with most parts of the background for which the distribution is known. The diffuse scattering contribution due to disorder in the crystal is more difficult to deal with, since it has a modulation that depends on details of the disorder (Benoit & Doucet, 1995). In the *p*3 form of bR it could be shown that the bR trimer is displaced as a rigid unit with an RMS deviation from its average position by about 0.5 Å (Grigorieff & Henderson, 1995). Using this model for the disorder the remaining background in the diffraction patterns could be removed. The diffraction data thus obtained were used for refinement of the structure for bR presented here.

Refinement strategy at limited resolution

In the refinement of an X-ray structure solved at high resolution (e.g. 2.0 or 2.5 Å) the number of observations (structure factors) is usually significantly larger than the number of structural parameters to be refined. This allows accurate refinement of atomic positions as well as temperature factors for each atom, describing its degree of order, although, even then, atomic groups must be constrained and restricted independence of the temperature factors of adjacent atoms must be imposed. At a resolution of 3.5 Å in the plane of the membrane and 4.3 Å perpendicular to the membrane, as in our work, the refinement suffers from underdetermination. Thus, in the case of bR, there are about 2300 non-hydrogen atoms, including ten lipid molecules, with three degrees of freedom per atom, but only about 4700 independent structure factors to refine against. Refinement in such a situation leads to a structure with a very low *R*-factor. However, the accuracy of the structure usually decreases during such a refinement, as can be monitored with a free *R*-factor (Brünger, 1992, 1993), or with a free phase residual (see Methods) as has been done here. To prevent this from happening, very tight geometric restraints and global temperature factors with very few free parameters were maintained throughout the refinement. This reduced the number of degrees of freedom and led to a structure with a still relatively high *R*-factor of 28% but with a free phase residual steadily decreasing as the refinement progressed, indicating increasing accuracy of the structure.

To allow for varying disorder in the molecule it was essential to include some treatment of

temperature factors. For example, some interhelical loops and the lipid head groups show no or only weak density, suggesting that they are completely or partially disordered, whereas most of the helices around the chromophore have good density, indicating a well-ordered structure. To keep the added number of degrees of freedom small, each residue was given a single base temperature factor and these base temperature factors were constrained to be a smooth function of residue position in the sequence (see Methods). Within each residue, an increase in temperature factor with increasing distance from the backbone along a side-chain was then added to the residue base temperature factor with separate treatment for aromatic side-chains. This increased temperature factor with distance from C^{α} was applied as the average for all residues in the structure.

A strategy of tight geometry, highly constrained temperature factors and, consequently, a reduced number of degrees of freedom seems essential for reliable refinement of a structure at moderate resolution. As pointed out recently (Kleywegt & Jones, 1995), a more conservative management of degrees of freedom also serves well for the refinement of structures at higher resolution solved by X-ray diffraction.

Results

An improved experimental density map

With the increased number of images contributing phase information to the experimental map, the features in regions where previously there was difficulty in interpretation have become much clearer. However, the orientation of most side-chains is still not very accurately defined by the density. Figure 1 and Figure 2 show two regions of the map to illustrate the improved quality. Figure 1 shows a cross-section though the helices where previously the densities for side-chains such as leucine, isoleucine, valine and methionine were poor. This was especially important in helix D, which contains none of the bulky aromatic side-chains that were necessary in the 1990 study (Henderson *et al*., 1990) to provide unambiguous guide points. In each Figure: (a) shows part of the previous map with 2750 measured phases (Henderson *et al*., 1990); (b) shows the new map (3766 measured phases); and (c) shows a $2F_o - F_c$ map (4743 calculated phases) after refinement as described below. Figure 1(a) shows only a hint of resolved side-chain density for residues smaller than tyrosine, tryptophan and phenylalanine, but this was sufficient to allow the previous, correct overall interpretation. Figure 1(b) shows clearly resolved density for side-chains of leucine, isoleucine, valine, methionine and threonine, which is sufficient now to be certain of the interpretation of helix D; its position is some 4 Å closer to the cytoplasmic surface (larger *z*-coordinate) than in the previous 1BRD model. This correction was not unexpected in view of the limited (7.8 Å) vertical resolution of the previous data. Finally, Figure 1(c) shows, with the $2F_o - F_c$ map, that the increase to a 60°-tilt limit in the electron diffraction amplitudes together with the improved phases from the model is adequate to produce a good map.

The vertical resolution, using the method of evaluating the point spread function described by Henderson *et al*. (1990), is 7.8 Å, 6.0 Å and 4.3 Å in each of the three maps. The improvement in the vertical resolution is due to the increase in the number of phases at higher tilt angles in each case. The effect of this is best seen in a view parallel to the membrane plane of density, whose interpretation has always been clear. The side-chains of residues Tyr185 and Phe208 from helices F and G are in van der Waals contact and vertically above one another. The densities for the two side-chains, as pointed out previously, were unresolved (Figure 2(a)). There is now a clear minimum between the two side-chains (Figure 2(b)) indicating that our vertical resolution is just sufficient to resolve atoms in van der Waals contact with one another. The building of the atomic model can then be done with more confidence, and refinement of the co-ordinates against the higher-resolution diffraction data can then improve the map further (Figure 2(c)). The point spread functions are shown as insets in each case. Finally, Figure 2(d) also gives a view of the retinal and its environment parallel to the membrane plane.

These two Figures show the steps involved in the work as illustrated by two parts of the structure. The entire map was carefully examined and a new model built, initially using FRODO (Jones, 1978, 1985) and subsequently using O (Jones & Kjeldgaard, 1993). Apart from the change in the vertical *z*-coordinate of helix D, the new model is substantially the same as that obtained previously. This is illustrated in Figure 3 where the difference between Φ and Ψ angles in the backbone, and also the RMS difference between co-ordinates of each residue in the previous and the refined model are plotted (excluding residues not included in the previous model). The overall RMS deviation between the two models is 2.2 Å. However, excluding the larger deviation of helix D the RMS deviation is only 1.9 Å. In the previous model, uncertainties about the position of residues in non-helical regions at the two surfaces meant that only 170 residues out of the 248 present in the protein could be identified. The new map has allowed us to build residues 7 to 227, though a number of surface residues included in this model appear to be disordered in addition to those missing at the N and C termini. Of the six surface loops linking adjacent helices, there is clear density for the CD, DE and FG loops, but poorly ordered density for AB, BC and EF (cf. Figure 6(a)). Ten lipid molecules have also been built into the remaining density features that cannot be part of the protein. Some of these lipid features are very clear, but others are very weak; these are differentiated by

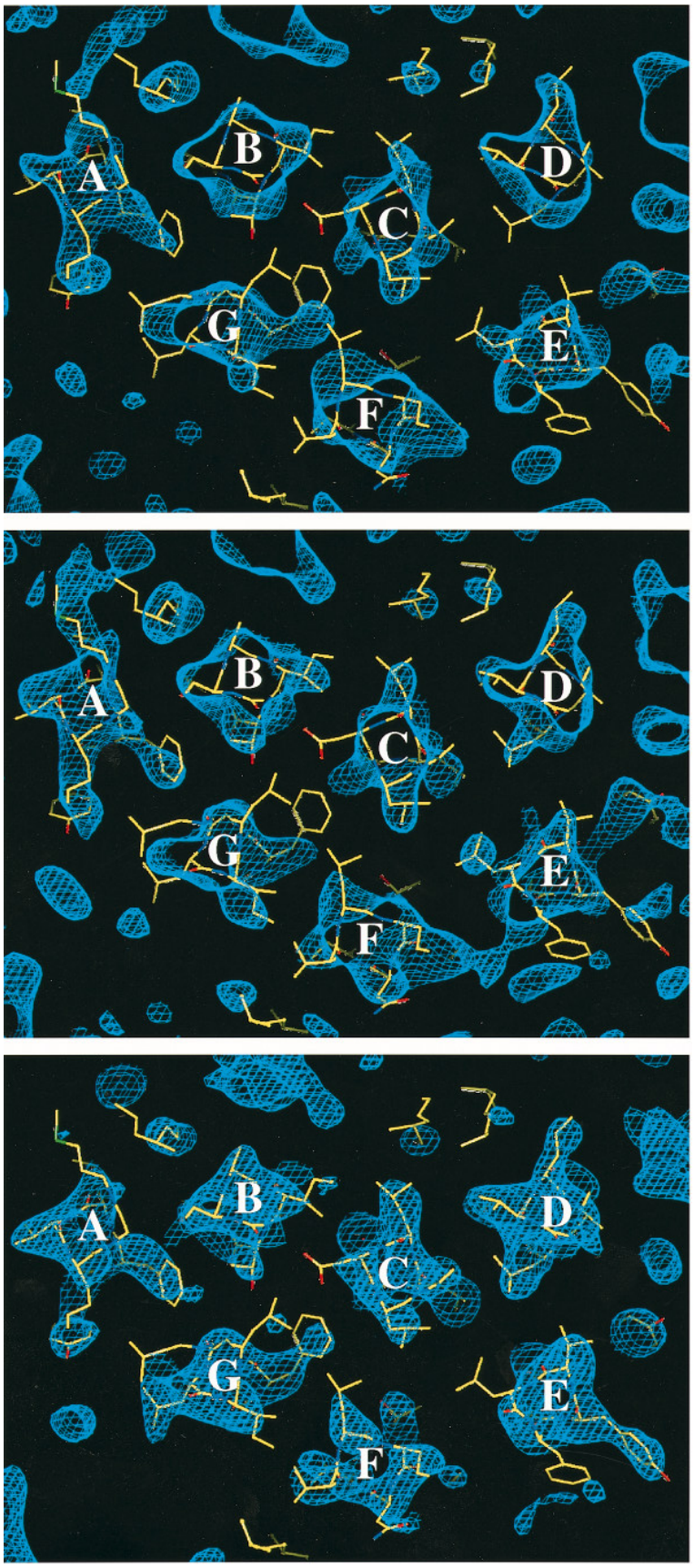


Figure 1. Cross-sections through the density map viewed perpendicular to the membrane plane, showing improved density for the helices. (a) Previous map with 2750 measured phases (Henderson *et al*., 1990); (b) improved map with 3766 measured phases; (c) $2F_o - F_c$ map with 4743 calculated phases from the refined model. The successive improvement of density can be seen, for example, at Leu109 and Leu111 (above and below letter D in the Figure) in helix D.

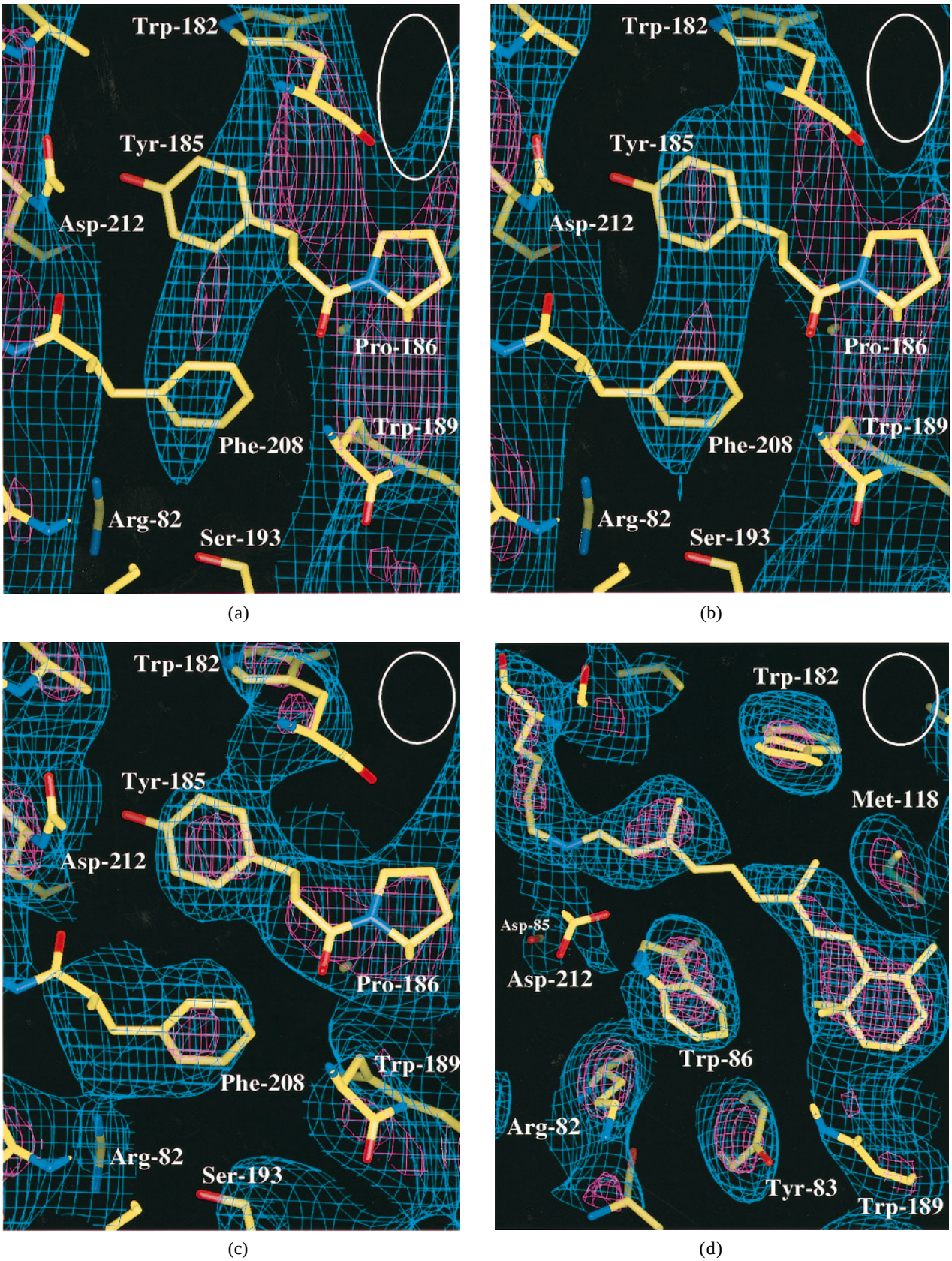


Figure 2. Views parallel to the membrane plane. Three views of the density for Tyr185 and Phe208, which are in van der Waals contact are shown. (a) Previous map with 2750 measured phases (Henderson *et al*., 1990); (b) improved map with 3766 measured phases showing the residues separately resolved; (c) $2F_o - F_c$ map with 4743 calculated phases from the refined model. The improved vertical resolution in (b) and (c) is illustrated by the density profile between the side-chains of Tyr185 and Phe208 which are not resolved in (a). (d) $2F_o - F_c$ map of the retinal and its environment. The vertical point spread functions are included as insets in the upper right corners.

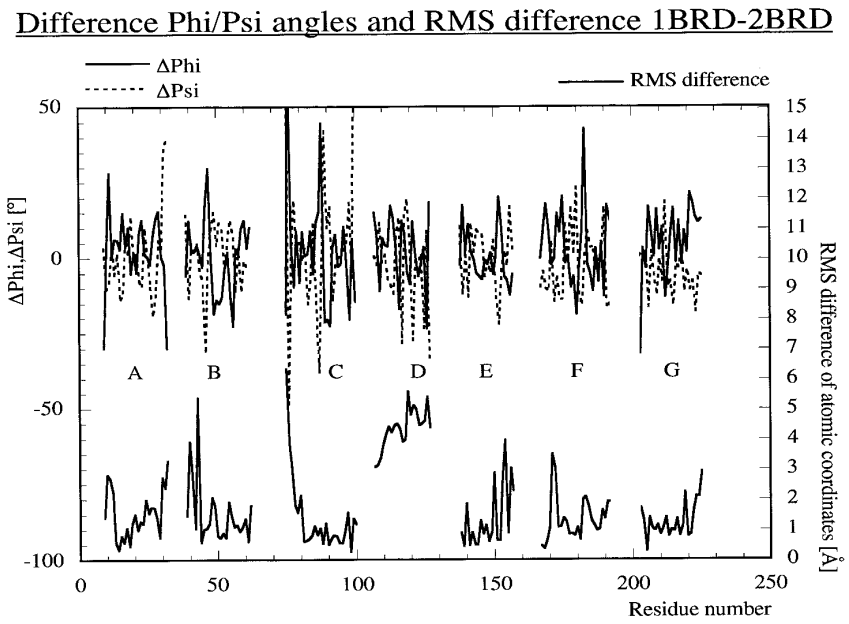


Figure 3. Difference between Φ and Ψ angles in the backbone, as well as the RMS difference between co-ordinates of every residue in the previous and the refined model (surface loops were not included in the earlier model). Larger deviations from zero in either plot occur in regions where structure has been reinterpreted on the basis of the new improved map. For example, the RMS difference in helix D co-ordinates is due to a 4 Å vertical shift of the helix.

their temperature factors determined during refinement.

Refinement

Progress of refinement

The initial model built into the three-dimensional density map had an *R*-factor of 47% with an average phase residual of about 62°. Temperature factors of atoms were assigned to residues by inspection of the density map by eye and were at first kept fixed. Refinement of the structure using Hendrickson–Konnert (HK) least-squares refinement (Hendrickson & Konnert, 1981) was done with tight geometric restraints (see Methods). Initially, the *R*-factor dropped to about 29% (see Figure 4). However, the free phase residual increased to 64°, suggesting that

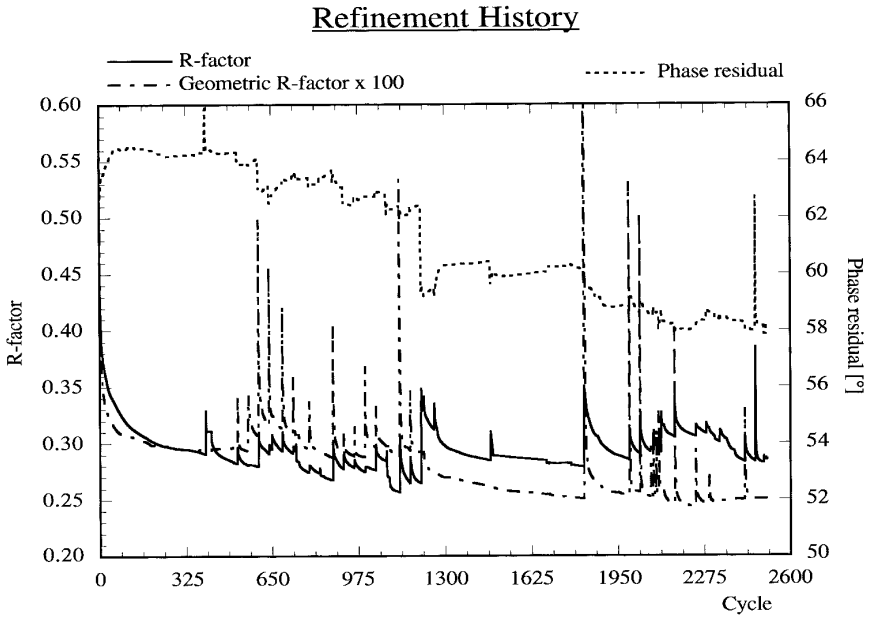


Figure 4. *R*-factor, free phase residual and geometric *R*-factor during the refinement. The final *R*-factor is 28% with a free phase residual of 58°. The irregular jumps indicate points at which changes in co-ordinates or restraints were introduced manually as described in the text.

Table 1. H-bond restraints in the final model for bR

Atom I	Atom II	Atom I	Atom II	Atom I	Atom II	Atom I	Atom II
H-bonds in helices		Ala81 O	Asp85 N	Ala139 O	Ala143 N	Leu206 O	Val210 N
Pro8 O	Ile11 N	Arg82 O	Trp86 N	Ile140 O	Ala144 N	Leu207 O	Leu211 N
Glu9 O	Leu13 N	Tyr83 O	Leu87 N	Ser141 O	Met145 N	Phe208 O	Asp212 N
Trp10 O	Ala14 N	Ala84 O	Phe88 N	Thr142 O	Leu146 N	Met209 O	Val213 N
Ile11 O	Leu15 N	Asp85 O	Thr89 N	Ala143 O	Tyr147 N	Val210 O	Ser214 N
Trp12 O	Gly16 N	Trp86 O	Thr90 N	Ala144 O	Ile148 N	Leu211 O	Ala215 N
Leu13 O	Thr17 N	Phe88 O	Leu92 N	Met145 O	Leu149 N	Asp212 O	Lys216 N
Ala14 O	Ala18 N	Thr89 O	Leu93 N	Leu146 O	Tyr150 N	Val213 O	Val217 N
Leu15 O	Leu19 N	Thr90 O	Leu94 N	Tyr147 O	Val151 N	Ser214 O	Gly218 N
Gly16 O	Met20 N	Pro91 O	Leu95 N	Ile148 O	Leu152 N	Ala215 O	Phe219 N
Thr17 O	Gly21 N	Leu92 O	Asp96 N	Leu149 O	Phe153 N	Lys216 O	Gly220 N
Ala18 O	Leu22 N	Leu93 O	Leu97 N	Tyr150 O	Phe154 N	Val217 O	Leu221 N
Leu19 O	Gly23 N	Leu94 O	Ala98 N	Val151 O	Gly155 N	Gly218 O	Ile222 N
Met20 O	Thr24 N	Leu95 O	Leu99 N	Leu152 O	Phe156 N	Phe219 O	Leu223 N
Gly21 O	Leu25 N	Asp96 O	Leu100 N	Phe153 O	Thr157 N	Gly220 O	Leu224 N
Leu22 O	Tyr26 N	Leu97 O	Leu 100 N	Pro165 O	Ala168 N	Leu221 O	Arg225 N
Gly23 O	Phe27 N	Ala98 O	Val101 N	Glu166 O	Thr170 N	Ile222 O	Ser226 N
Thr24 O	Leu28 N	Ala98 O	Asp102 N	Val167 O	Phe171 N	Leu223 O	Arg227 N
Leu25 O	Val29 N	Asp104 O	Ile108 N	Ala168 O	Lys172 N		
Tyr26 O	Lys30 N	Gln105 O	Leu109 N	Set169 O	Val173 N		
Pro37 O	Lys40 N	Gly106 O	Ala110 N	Thr170 O	Leu174 N	*H-bonds for Thr and Ser*	
Asp38 O	Lys41 N	Thr107 O	Leu111 N	Phe171 O	Arg175 N	Leu13 O	Thr17 OG1
Ala39 O	Tyr43 N	Ile108 O	Val112 N	Val173 O	Val177 N	Met20 O	Thr24 OG1
Lys40 O	Ala44 N	Leu109 O	Gly113 N	Leu174 O	Thr178 N	Tyr43 O	Thr47 OG1
Lys41 O	Ile45 N	Ala110 O	Ala114 N	Arg175 O	Val179 N	Ala51 O	Thr55 OG1
Phe42 O	Thr46 N	Leu111 O	Asp115 N	Asn176 O	Val180 N	Thr55 O	Ser59 OG
Tyr43 O	Thr47 N	Val112 O	Gly116 N	Val177 O	Leu181 N	Asp85 O	Thr89 OG1
Ala44 O	Leu48 N	Gly113 O	Ile117 N	Thr178 O	Trp182 N	Trp86 O	Thr90 OG1
Thr47 O	Ala51 N	Ala114 O	Met118 N	Val179 O	Ser183 N	Ile117 O	Thr121 OG1
Leu48 O	Ile52 N	Asp115 O	Ile119 N	Val180 O	Ala184 N	Trp137 O	Ser141 OG
Val49 O	Ala53 N	Gly116 O	Gly120 N	Leu181 O	Tyr185 N	Trp138 O	Thr142 OG1
Pro50 O	Phe54 N	Ile117 O	Thr121 N	Ser183 O	Val187 N	Phe153 O	Thr157 OG1
Ala51 O	Thr55 N	Met118 O	Gly122 N	Ala184 O	Val188 N	Leu174 O	Thr178 OG1
Ile52 O	Met56 N	Ile119 O	Leu123 N	Tyr185 O	Trp189 N	Val179 O	Ser183 OG
Ala53 O	Tyr57 N	Gly120 O	Val124 N	Pro186 O	Leu190 N	Val210 O	Ser214 OG
Phe54 O	Leu58 N	Thr121 O	Gly125 N	Val187 O	Ile191 N		
Thr55 O	Ser59 N	Gly122 O	Ala126 N	Val188 O	Gly192 N	*Other H-bond restraints*	
Met56 O	Met60 N	Leu123 O	Leu127 N	Leu190 O	Ala196 N	Tyr57 OH	Asp212 OD2
Tyr57 O	Leu61 N	Val124 O	Thr128 N	Ile191 O	Gly197 N	Tyr83 OH	Trp189 NE1
Leu58 O	Leu62 N	Tyr133 O	Val136 N	Leu201 O	Glu204 N	Ala126 O	Arg134 NH2
Ser59 O	Gly63 N	Arg134 O	Trp138 N	Asn202 O	Thr205 N	Tyr185 OH	Asp212 OD1
Pro77 O	Ala81 N	Phe135 O	Ala139 N	Asn202 O	Leu206 N	Gly192 O	Gly195 N
Ile78 O	Arg82 N	Val136 O	Ile140 N	Ile203 O	Leu207 N	Glu194 OE1	Arg134 NH1
Tyr79 O	Tyr83 N	Trp137 O	Ser141 N	Glu204 O	Phe208 N	Glu194 OE2	Arg134 NH2
Trp80 O	Ala84 N	Trp138 O	Thr142 N	Thr205 O	Met209 N	Asp212 OD1	Trp86 NE1

The shaded regions indicate the location of 3_{10} H-bonds or the helical H-bonding in the extra loose turn at the end of helix F.

the accuracy of the structure was reduced. This interpretation may be misleading for the following reason: features in a density map are much more sensitive to phases than they are to amplitudes (Ramachandran & Srinivasan, 1961). Building a structure into a map is therefore likely to lead to a biased phase residual. For this reason the phase residual at the beginning of the refinement should not be regarded as ''free'' and, therefore, it is not a measure of the accuracy of the model. With progressive refinement of the model and manual rebuilding of parts of it, the phase residual was steadily improved to a final value of 58°. This gave us confidence at each level that the accuracy of the model increased. On the other hand, the *R*-factor dropped further to about 26% and then went up again to a final value of 28%. This increase was caused by the addition of further geometrical restraints during the refinement in the form of explicitly defined hydrogen bond and helical Φ, ψ restraints. Hydrogen bonds were only added between atoms which were already within H-bonding distance (about 2.9 Å) of each other. In the case of serine and threonine residues an H-bond was assumed between the side-chain oxygen and the backbone oxygen of the residue in the previous turn of the helix, only if there was no other possible interaction with the side-chain. A complete list of H-bonds is given in Table 1. The geometric *R*-factor also plotted in Figure 4 is calculated for each refinement cycle (program PROLSQ in the CCP4 suite, Collaborative Computational Project, 1994), and is an overall measure of the agreement of bond lengths, angles and planarity of planar groups with tabulated ideal values. Every manual rebuilding of parts of the model produced a temporarily bad geometric *R*-factor (spikes in Figure 4) with a quick recovery of the initial value in subsequent refinement. The final geometric *R*-factor indicates significant improvement of the overall geometry of

Table 2. Refined atomic form factors

Chemical group	Relative IAM scattering strength	Relative scattering strength after refinement	Change (%)	No. atoms in bR
Backbone carbon	1.00	0.67	–33	240
Carbon without H	1.00	0.96	–4	77
Aliphatic CH	1.00	1.05	+5	307
Aliphatic CH_2	1.00	0.98	–2	215
Aliphatic CH_3	1.00	1.09	+9	191
Aromatic CH	1.00	0.91	–9	144
Backbone nitrogen	0.89	0.97	+9	230
NH_2	0.89	0.76	–15	5
NH_2^+	0.89	1.03	+16	21
NH_3^+	0.89	1.60	+80	6
O^-	0.80	0.89	+11	28
Backbone oxygen	0.80	1.09	+36	226
OH	0.80	0.80	0	38
S	2.06	1.69	–18	9

Form factors are relative to the neutral form factor for carbon.

the model. The RMS bond length deviation at the end of the refinement was 0.005 Å.

Refinement of backbone temperature factors using a cubic spline function (CSF) along the protein sequence (see Methods) first started with 30 knots. The number of knots was later increased to 70, which gave the lowest phase residual (see Methods and Figure 15). Similarly, refinement of the lipids was first done using a CSF with five knots for the average temperature factor of all lipids along the *z*-coordinate. Later, the lowest phase residual was found when independent CSFs with two knots for each lipid were used. Form factor refinement and side-chain temperature factor refinement were carried out iteratively due to the interdependence of form factors and temperature factors (see Discussion). We distinguished 14 chemically different groups (see Table 2) and allowed a separate parameter for each of their scattering factors.

Manual rebuilding of the model focused on the loop regions and the lipids that were not included in the previous model. Rebuilding of the loops was helped by work on spin-labelled bR mutants (Altenbach *et al*., 1990), polarity measurements using pH indicators (Alexiev *et al*., 1994) and low-resolution X-ray diffraction of a mercury-labelled Cys103 mutant (Krebs *et al*., 1993; see discussion of structural details). It was found that the convergence of the refinement of the positions of newly built structural elements could be speeded up by resetting the temperature factors for the corresponding residues by hand to low

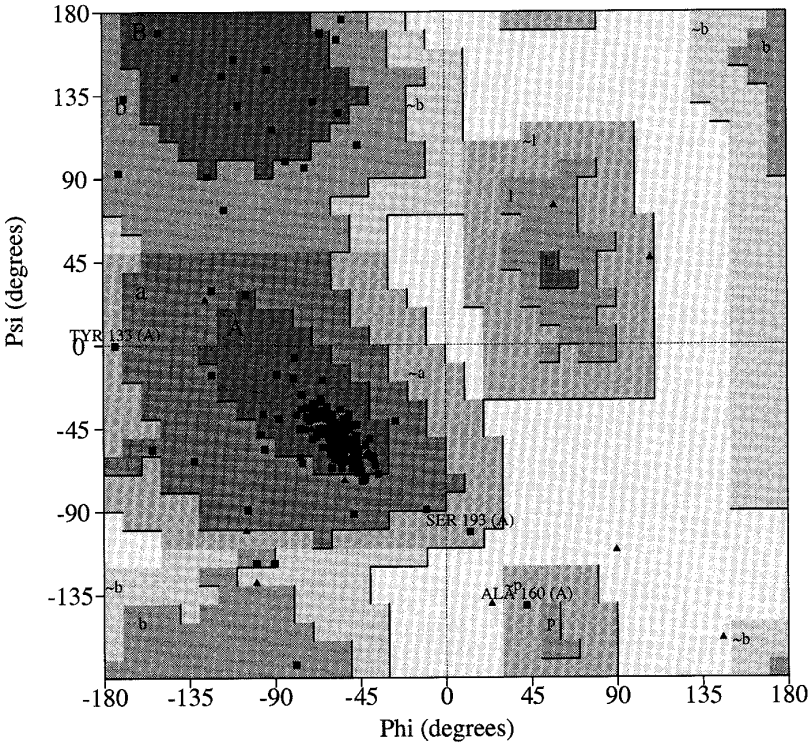


Figure 5. Ramachandran plot for the refined protein produced using the program PROCHECK (Laskowski *et al*., 1993). Domains A, B and L are most favoured regions; domains a, b, l and p are additional allowed regions; and domains ~a, ~b, ~l and ~p are generously allowed regions. Most residues have an α-helical conformation (domain A) with a scatter of about 10° in the Φ and Ψ angles. The triangles are glycine, and the three labelled residues are in barely allowed regions. Tyr133 and Ser193 are associated with the putative salt bridge between Arg 134 and Glu194.

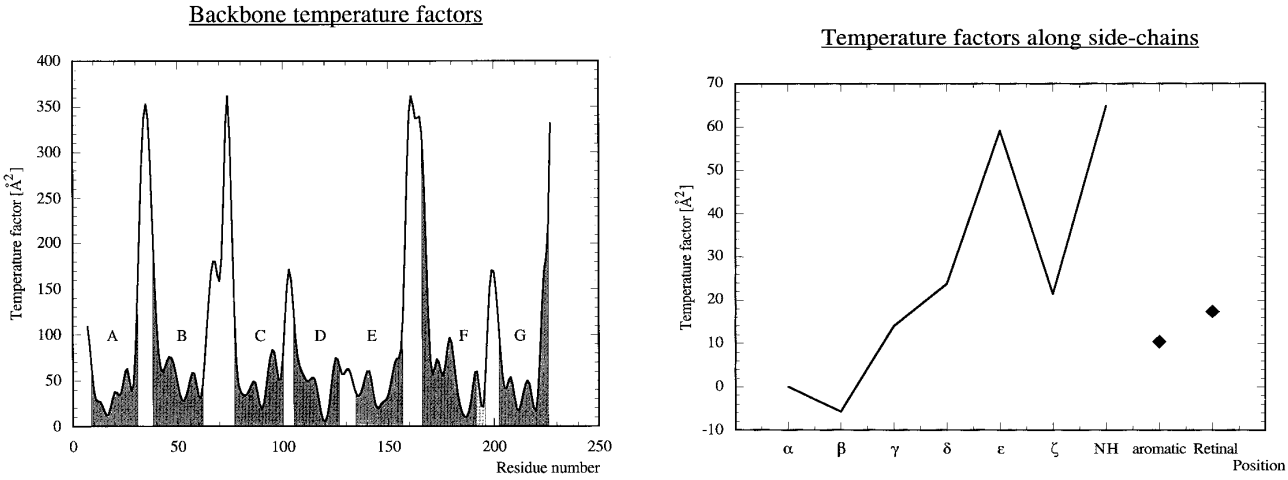


Figure 6. (a) Refined backbone temperature factors using cubic spline smoothing with 70 knots. Regions of low temperature factors are labelled with the corresponding helix number, and the extent of each helix is indicated by a shaded region. The lighter shaded region next to helix F shows the location of the extra loose turn. (b) Refined temperature factors grouped according to distance from the backbone. The retinal atoms are in a separate group. All protein atoms are put into one of the nine groups and plotted *versus* position along the side-chain. ζ contains six lysine nitrogen atoms and seven arginine carbon atoms; NH contains 14 arginine nitrogen atoms.

values (20 to 60), followed by a few cycles of co-ordinate refinement and subsequent temperature factor refinement. In this way the surrounding structure could readjust to the changed domain while critical residues were held in the density. After completion of the refinement a few omit maps were calculated to try to get an independent view of one or two of the ambiguous regions of the model, such as the position of Arg82, but this did not alter our interpretation. In the final model there were no residues in disallowed regions of the Ramachandran plot (Figure 5).

Refinement results

Temperature factor refinement. The temperature factors along the backbone are plotted in Figure 6(a) and clearly identify well-ordered domains in the helices (temperature factors from 10 Å^2 to 50 Å^2) as well as partially or completely disordered domains in the loop regions (temperature factors from 50 Å^2 to 350 Å^2). Thus, domains corresponding to helices A to G can be identified (see Figure 6(a)). Inspection of the density map shows significant density for the CD, DE and FG loops, indicating a notable degree of order in those loops. This can also be seen in the plot in Figure 6(a) where the temperature factors for these loops assume intermediate values between 50 Å^2 and 150 Å^2. The remaining loops have very high temperature factors and the density map shows very weak or no density in these regions, suggesting that these loops are disordered. The sections of helices D and F in contact with the β-ionone ring of the retinal have the highest degree of ordering and, therefore, the lowest temperature factors.

Figure 6(b) shows average side-chain temperature factors relative to the main-chain (α position). Generally, the temperature factors increase with the distance of the side-chain atoms from the backbone, in good agreement with the expected increase in mobility of atoms more distant to the rigid backbone. Deviations from this general increase in temperature factor at position β and ζ are likely to be artefacts due to the limited resolution and noise in the data. The plot in Figure 6(b) shows similar temperature factors for atoms in the α and β positions as well as in retinal and the aromatic ring structures (phenylalanine, tyrosine and tryptophan) implying comparable density and, therefore, mobility for these atoms on average.

Since the helices in bR are arranged almost perpendicular to the plane of the membrane and there are no further layers of protein molecules helping to maintain long-range order like in a three-dimensional crystal, the temperature factor is highly correlated with the depth in the membrane (*z*-coordinate). Figure 7 shows plots of temperature factors along *z* for each helix. The plots confirm that the protein is most rigid in the middle of the membrane near the retinal-binding pocket ($z \approx +5$ Å) and becomes more flexible at the surfaces. It is interesting to note that proline residues near the middle of the membrane (Pro50 at $z \approx +8$ Å in helix B and Pro186 at $z \approx +1$ Å in helix F) seem to be located where an increase in the temperature factor towards the cytoplasmic side occurs. This is most visible in helix F where the temperature factor goes up by about 50 Å^2. An increase of the temperature factor in helical regions indicates increased flexibility of the protein in this region and may help to identify locations of conformational changes in the protein during the

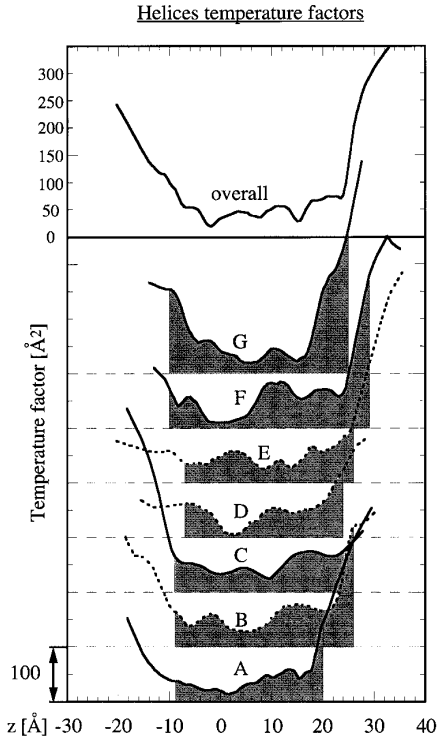


Figure 7. Temperature factors for atoms in each helix (A to G) and the average over the whole structure plotted as a function of the *z*-coordinate. The temperature factors for some helices have been plotted with a broken line for clarity. For each plot a shaded region has been drawn to indicate the location of the helical conformation. Note that the centre of gravity of the protein has turned out to be at about $z = +5$ Å; there is no absolute origin in space group $p3$. Roughly half of each loop has been arbitrarily assigned to each of the seven plots.

photocycle (see Discussion). Figure 7 also shows a plot of the average temperature factor for all atoms in the structure as a function of *z*-coordinate. The plot suggests that the cytoplasmic side of the protein (*z*-coordinate > +5 Å) is more disordered (higher temperature factor) than the extracellular side (*z*-coordinate < +5 Å). The variation of temperature factors within the protein is illustrated in Figure 8 using a colour scheme where blue regions are ordered domains, and red and yellow regions represent progressively more disordered domains. The blue sections of helices D and F next to the β-ionone ring of the retinal are the most ordered parts of the protein.

The temperature factors for each lipid molecule along *z* are plotted in Figure 9. All lipids except lipid 267 have high temperature factors at the head groups and lower values halfway along the hydrocarbon chain, increasing again at the end of the chain. This corresponds well with the observed disorder of lipids in a bilayer (Seelig & Seelig, 1980; Moser *et al*., 1989). On average, the temperature factors are much higher than those found in the protein due to the fluid-like behaviour of the lipids (Seelig & Seelig, 1980). Some lipids have below average temperature factors halfway along their polyene chains and are partially resolved (e.g. lipids 261, 263 and 267; see Table 6, below). Other lipids seem to have higher temperature factors, although there is good density for them in the map (e.g. lipid 262). This may require further attention.

Form factor refinement. Refined electron form factors are listed in Table 2, in units relative to neutral carbon. The starting values were taken from Cowley (1992) and show decreasing scattering strength from neutral carbon to oxygen. This order is reversed for the backbone atoms after refinement. For the carbon groups there is a clear trend in the form factors to higher values as hydrogen atoms are added. Values for carbon atoms in aromatic groups remain relatively small. In the case of nitrogen and oxygen, addition of hydrogen is paralleled by form factors smaller than the backbone values. This is most likely to be due to the limited resolution of the diffraction data (see Discussion). However, the positively charged nitrogen groups (arginine and lysine) assumed significantly higher values after refinement than the nitrogen backbone group, and the negatively charged oxygen groups (aspartate and glutamate) show a refined value notably smaller than the corresponding backbone group. The form factor for the nine sulphur atoms in the nine methionine residues dropped after refinement by about 20%. The refinement of the form factors led only to a marginal improvement of the *R*-factor (0.3%) and no change in the phase residual.

Structural details

Helices. The improved density map allowed better interpretation of some of the loop regions at the surfaces of the protein (see above). As a result, the exact beginning and end points of the helices listed in Table 3 deviate slightly, by one residue on average, from their previous values (Henderson *et al*., 1990). The new residue numbers are still uncertain by at least one residue, although the interpretation of the loop regions make them more reliable than before. The largest changes were made in helices C, D, E and F. Most of these are due to residues that became part of one of the better-resolved loops. The single extra loose turn at the end of helix F is a new feature of the model, which is well supported by the experimental density map. With the movement of helix D towards the cytoplasm by 4 Å, Asp115 is now somewhat more distant from the β-ionone ring (9 Å from the C^5 methyl and 7.3 Å from the C^9 methyl) than in the previous model. Met118 forms a cap on the pocket for the ring and is in contact with the 5–6 double bond region. Many helices begin with one or two residues in the conformation of a 3_{10} helix at the N terminus (Table 1). This conformation was first suggested by the positions assumed during the refinement and later added as geometrical constraints for these residues. α-Helices beginning and ending with half a turn of 3_{10} helix are also a common feature in other proteins (Barlow & Thornton, 1988).

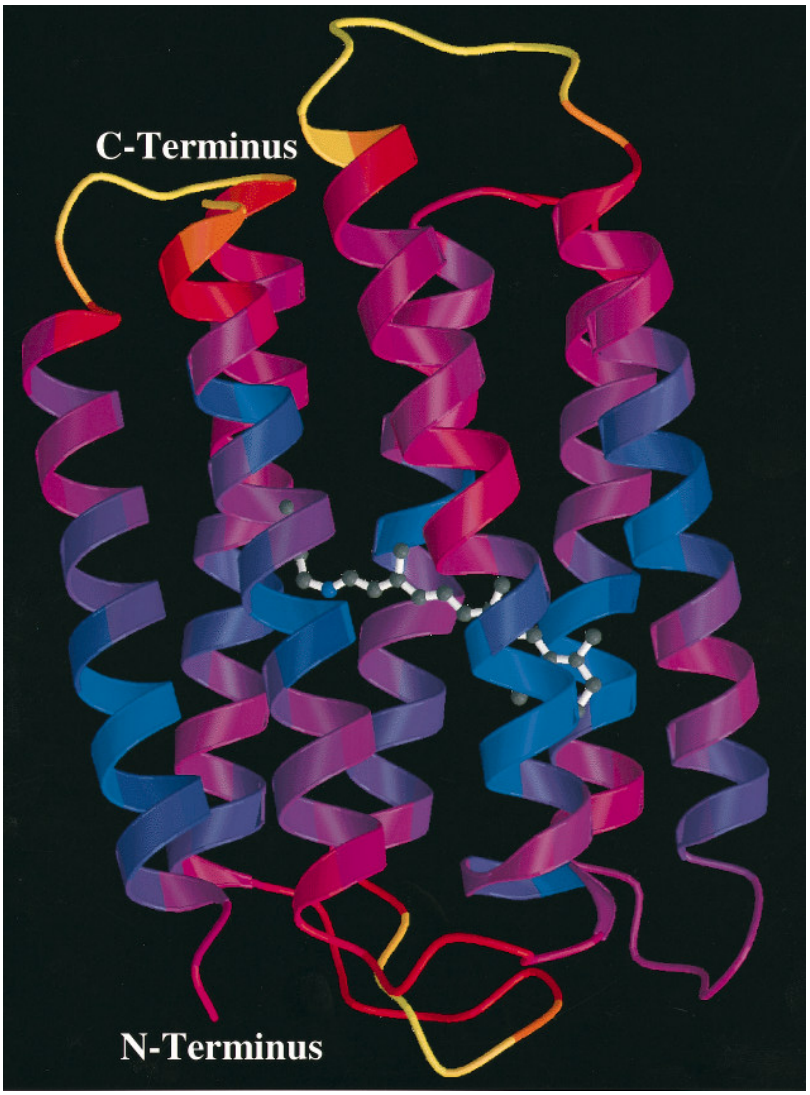


Figure 8. Ribbon representation of bR with colours representing different degrees of order in the protein. Blue changes to red and yellow as the degree of order decreases. Loops AB, BC and EF (yellow) are the most disordered parts of the protein, whereas the blue sections of helices D and F next to the β-ionone ring of the retinal are the most ordered parts.

Loops. Weak density for the BC loop indicates that it is partially disordered. However, some degree of order was suggested by restricted mobility of a spin label at residue 72 (Altenbach *et al*., 1989), by the destabilising effect the deletion of the loop had on the protein (Gilles-Gonzalez *et al*., 1991) and by the strong conservation of the amino acid sequence at some positions (e.g. Gly63, Gly65 and Thr67). By assuming maximum H-bonding between residues 69 and 73 in the loop, a β-sheet structure with a turn at residue 71 could be built. After rebuilding the loop, the residues assumed slightly different positions in later refinement cycles. The structure of the BC loop, therefore, remains essentially a guess. Consistent with the disorder in the cytoplasmic half of the structure, the AB and EF loops refine with temperature factors of over 100. Density for residues 164 to 166 in the EF loop is completely absent in all maps. The interpretation of residues 102 to 105 in the CD loop has been made taking into account the *xy*-position found in the low-resolution projection difference map of Hg-Cys103-labelled bR (Krebs *et al*., 1993). The structure of the loop as well as the beginning of helix D are in excellent agreement with results from spin-label studies (Greenhalgh *et al*., 1991). On the extracellular surface, the interpretation of the very short DE loop between Lys129 and Arg134 has been helped by extensive spin-labelled cysteine mutant analyses (Altenbach *et al*., 1990) and is consistent with polarity measurements using pH indicators (Alexiev *et al*., 1994). The density at the end of helix F loops out to a much greater diameter than allowed in an α-helix and we interpret this density as an extra helical turn into which two additional residues are inserted, namely Ser193 and Glu194. The positions of Arg134 and Glu194 then come into close juxtaposition. There is good density for our interpretation of the side-chain of Arg134, which is now buried in the protein near the extracellular surface, but there is nothing clear for Glu194. Nevertheless, we believe it is likely there is a salt bridge between the two residues and this has been built into the model. There is good density for both the DE and FG loops. Features along the FG loop could be interpreted as the side-chains of Ile198 and Pro200. However, although we have made a definite interpretation for the DE and FG

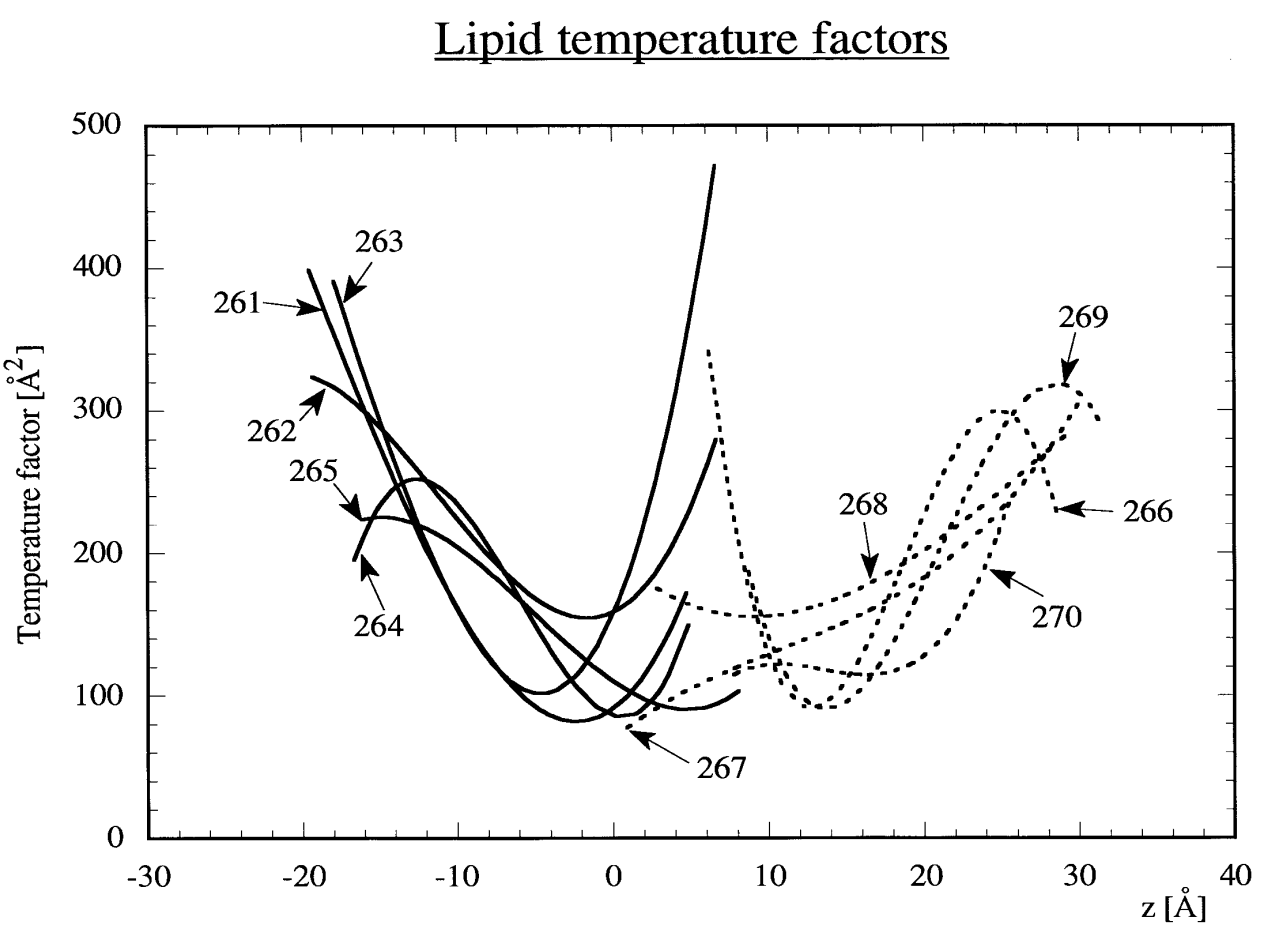


Figure 9. Refined temperature factors for the ten lipid molecules using cubic spline smoothing with two knots per lipid. Lipids on the extracellular side are plotted as a continuous line and those on the cytoplasmic side are plotted as a broken line.

loops, we still cannot be completely sure it is correct.

Critical side-chains. The orientation of the side-chain of Arg82 is still unclear. Since there is no clear density that could correspond to Arg82 in the improved density map, it could either point up towards Asp85, downwards away from Asp85 or, most likely, because of the absence of density, move between the two positions. Similarly, because of the poor scattering power of oxygen compared with, for example, the methyl groups of leucine and valine, there are no well-resolved densities for the side-chains of Asp85 and Asp96, only bulges. However, since aspartic acid residues in helices can only take up one conformation, the refinement against the more highly tilted electron diffraction data could proceed without ambiguity, giving rise to clear $2F_o - F_c$ maps. The orientation of the side-chain of Asp212 is determined by H-bonds with neighbouring residues (Table 1) which form part of the complex counterion of the Schiff base proton. We believe that Asp85 and Asp212 appear with particularly weak density because they are negatively charged, and that Asp96 is weak because, being near the cytoplasmic end of helix C, it has poorer order. In keeping with this, Asp115 is clearer and its order is likely to be increased further by an H-bond to Thr90 not included in the refinement restraints.

Table 3. Residues in helical conformation

Helix	Residues (previous)	Number
A	9–31 (10–32)	23
B	38–62 (38–62)	25
C	77–100 (80–101)	24
D	105–127 (108–127)	23
E	134–157 (136–157)	24
F	166–191 (167–193)	26
Loose turn terminating helix F	192–196 (–)	5
G	202–226 (203–227)	25

Residues at the start and end of some helices have the conformation of a 3_{10} helix (see Table 1). The listed terminating residues have their Φ and ψ angles both in a helical conformation.

Chromophore. Model angles and distances for the retinal were provided by the structure of retinylidene determined by X-ray diffraction (Santarsiero *et al.*, 1990). In the refined protein structure the relative orientations of the methyl groups of the retinal deviate on average by 4° (maximum 8°) from those in the retinylidene model. The deviation is

Table 4. Orientation of some bonds and position in projection of some atoms in the retinal

	Angle between vector and membrane normal (deg)	
Vector in retinal	Refined model	Other measurements
C^1–C^{16}	100.7	94[a]
C^1–C^{17}	67.4	75[a]
C^5–C^{15}	70.7	70.3[b]
C^5–C^{18}	38.6	37[c]
C^9–C^{19}	29.3	40[c]
C^{13}–C^{20}	16.1	32[c]
Atom in retinal	Position in projection (Å)	
C^1	x = 15.3, y = 9.9	x = 14.5, y = 11.0[d]
C^2	x = 14.5, y = 10.7	x = 15.3, y = 9.2[e]
C^{19}	x = 16.6, y = 8.6	x = 16.9, y = 8.7[e]
C^{20}	x = 17.9, y = 3.6	x = 17.6, y = 3.2[d]
		x = 17.5, y = 4.3[e]

[a] NMR of deuterated retinal (Ulrich *et al*., 1992).
[b] Linear dichroism measurements (Lin & Mathies, 1989). The transition dipole of the long axis of retinal may not exactly coincide with the C^5–C^{15} vector, and the actual angle of the C^5–C^{15} vector may, therefore, be slightly larger than measured.
[c] NMR of deuterated retinal (Ulrich *et al*., 1994).
[d] Neutron diffraction of deuterated retinal (Hauss *et al*., 1994). C^1 and C^{20} do not coincide exactly with the centres of deuteration.
[e] X-ray diffraction of bromine and mercury-labelled retinal derivatives (Büldt *et al*., 1991).

within the range found in small molecule crystal structures of various retinals and retinal derivatives (Hamanaka *et al*., 1972; Stam, 1972; Simmons *et al*., 1981, 1986). The angle between the polyene chain of the retinal in the refined model and the membrane normal (70.7°), the orientation of methyl groups C^{16}, C^{17}; C^{18}, C^{19} and C^{20} and the position of some atoms of the retinal within the unit cell of purple membrane (PM) have been estimated also by other methods, and the results are summarised and compared with the refined model in Table 4. Methyl groups C^{19} and C^{20} in the refined model make angles with the membrane normal that are somewhat smaller (about 15°) than found by NMR of deuterated retinal (Ulrich *et al*., 1994). There are only bulges in the density for those methyl groups in the map (see Figure 2(d)) and, hence, the angles assumed after refinement will be close to those provided by the model structure retinylidene. Differences between the refined model and other measurements, such as for the methyl groups C^{19} and C^{20}, may therefore be due to differences between retinylidene and the actual retinal structure in the protein. However, the NMR results might also be less accurate than claimed. The chromophore and its environment are shown in stereo in Figure 10, including all residues that are part of the retinal-binding pocket.

Cavities. The refined model for bR was searched for cavities large enough to accommodate one or more water molecules. The search was carried out using the program VOIDOO (Kleywegt & Jones, 1994), and 12 enclosed cavities with space for at least one water molecule each were found (Table 5 and Figure 11). VOIDOO searches the protein using a specified probe size. A cavity is detected if there is space for the probe and the space is inaccessible from the outer surface of the protein (enclosed cavity). The calculation of free space takes into account the van der Waals radii of atoms in the protein. To make sure all cavities were detected the protein was searched with different probe sizes. A small probe detects more cavities buried inside the protein but misses others near the surface that may appear to be accessible from the outside by a probe with reduced size. A larger probe misses small cavities inside the protein but may pick up new cavities near the surface. Probe sizes with radii between 0.5 Å and 1.4 Å were used and every cavity found in any of the searches was looked at to check its plausibility. In a final search a probe size with radius 1.2 Å was used to select all cavities large enough to accommodate one or more water molecules. The effective radius of a water molecule is somewhat larger than that (radius about 1.4 Å). The smaller probe should allow for inaccuracies in the orientation of side-chains inside the protein and eliminate cavities near the surface that are accidental, owing to the assumed conformation of surface loops that are disordered. The co-ordinates of the centres of gravity and lining residues for all cavities are listed in Table 5. In Figure 11 the cavities are shown as large blue spheres to indicate their positions within the protein. The size of each sphere matches approximately that of a water molecule although some cavities are larger than that (see Table 5). Figure 11 also shows the side-chains of all arginine, lysine, aspartic and glutamic acid residues that could take part in proton conduction.

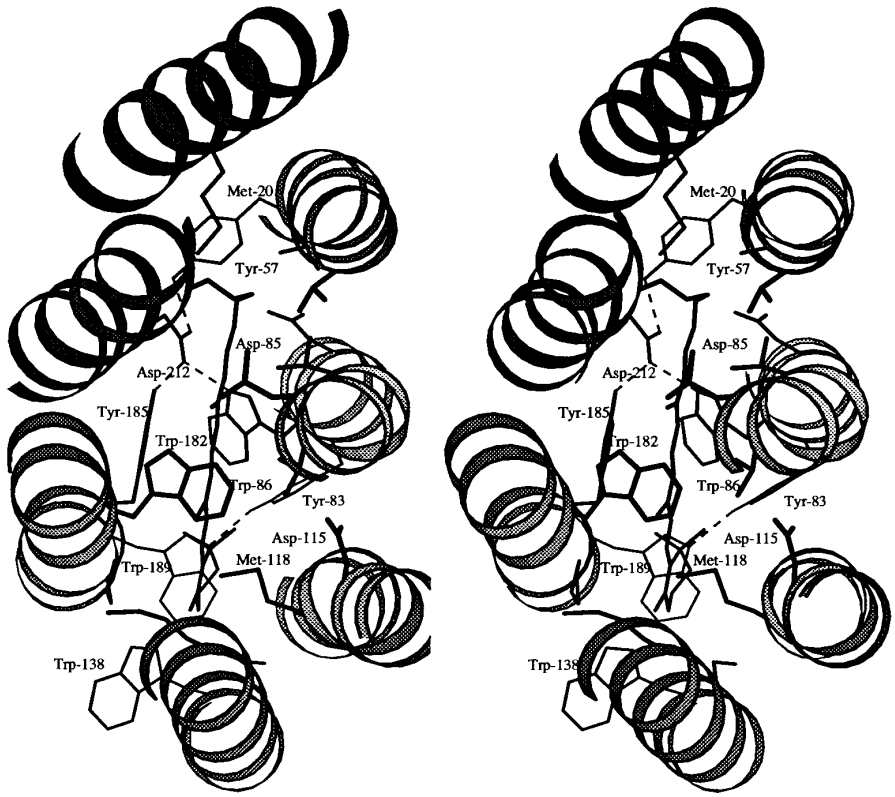


Figure 10. Stereo view of the retinal-binding pocket showing the retinal and all residues that are part of the binding pocket. The H-bond restraints for Asp212 and Trp189 are also indicated (see Table 1). The Figure was rotated by 90° from the standard orientation to allow the use of a larger scale.

Cavities I, II, IV, V and VI are located in the cytoplasmic half of the proton channel, whereas cavities IX, X and XII are part of the extracellular half of the proton channel. Cavities VIII and XI were found at opposite ends of the retinal binding pocket. The remaining cavities (III and VII) appear to be further away from the proton channel and the binding pocket. Water located in those cavities is unlikely to be involved in the translocation mechanism of protons. Cavity III under the CD loop becomes accessible to the solvent if the side-chain of Asp102 is rotated.

There are spaces for many more water molecules in open cavities at the surfaces of the protein, particularly on the extracellular side where the proton channel is more hydrophilic and opens up to the water phase (Henderson *et al*., 1990). These cavities are not well defined, have a complex structure with many charged side-chains, and will not be discussed here. The function of water in these cavities is mainly to deliver and export protons to and from the surfaces.

Lipids. The positions of ten lipid molecules included in the model are shown in Figure 12(a) and (b) for lipids on the cytoplasmic and the extracellular side, respectively. Table 6 summarises lipid positions and indicates which of the lipids are resolved in the density map. All lipids in the model are built as phosphatidyl glycerophosphate with dihydrophytol chains, although in PM as purified from Halobacteria (Kates *et al*., 1982) 30% are actually glycolipids (on the extracellular side; Henderson *et al*., 1978) with a small proportion of other lipids. Since all head groups are disordered and, therefore, not resolved in the density map, and the hydrocarbon chains all have the same structure, a single type of lipid was used for simplicity. Recent work has shown, by mass spectroscopy (Tsujimoto *et al*., 1989; Kates *et al*., 1993), that the principal lipid in Halobacteria is not phosphatidyl glycerophosphate but its terminal O-methyl ester, so the lipid in our model does not actually occur naturally to any extent. There are four out of five lipid molecules on the extracellular side of the membrane with one or both polyene chains resolved as density, but only two out of five lipids with resolved chains on the intracellular side. The position of a lipid within the membrane was adjusted to match the density if the lipid was resolved, and other lipid molecules were added to fill space with similar *z*-coordinates if there was no density for the lipid. In some cases a lipid could be positioned so that its head group was near a hydrophilic side-chain on the surface of the protein such as arginine or lysine. The degree of order in the lipids parallels that found in the protein, being greater on the extracellular side.

There was clear density for one lipid hydrocarbon chain in a crevice between adjacent bR monomers within the trimer. A squalene molecule was

Table 5. Cavities in bR, their location, number of water molecules they can accommodate, and lining residues

Cavity	Location	Centre of gravity (Å)	No. of H_2O	Lining residues
I	CP, close to the cytoplasmic surface	x = 19.3 y = –1.6 z = 24.2	1	**Phe42 Asp96 Leu100 Leu223 Ser226 Arg227**
II	CP, close to the cytoplasmic surface	x = 19.4 y = –5.4 z = 21.1	1	**Phe27 Phe42 Tyr43 Thr46 Leu223** Leu224 **Arg227**
III	Under CD loop, close to cytoplasmic surface	x = 14.0 y = 10.1 z = 21.6	7	**Leu97 Ala98 Val101 Asp102 Thr107 Leu111 Val151 Leu152 Phe154** Gly155 **Phe171**
IV	CP, close Asp96	x = 15.3 y = –1.2 z = 15.6	1	**Ile45** Thr46 **Val49** Leu92 **Leu93 Asp96 Phe219**
V	CP, between helices A and B, close to backbone oxygen of Thr46	x = 22.8 y = –7.0 z = 13.1	3–4	Gly23 **Tyr26 Phe27 Thr46 Pro50 Lys216 Val217** Gly220 **Leu221**
VI	CP, close to retinal-binding pocket	x = 21.9 y = 4.2 z = 13.0	1	**Leu93 Leu174 Val177 Thr178 Leu181 Trp182 Ala215 Phe219 Ile222**
VII	Between helices D and E	x = 10.5 y = 13.0 z = 15.1	1	**Leu94 Leu111 Ala114 Tyr147 Ile148 Val151**
VIII	BP, between helices A and G, close to Lys216	x = 21.0 y = –5.3 z = 8.2	1	**Met20** Gly23 **Pro50 Ala53 Val213 Lys216 Val217**
IX	EC, part of complex counterion of the Schiff base	x = 17.2 y = –0.3 z = 0.3	1	**Tyr57 Met60 Arg82 Asp85 Trp86 Asp212**
X	EC, part of complex counterion of the Schiff base	x = 20.5 y = 0.4 z = –2.1	1–2	**Leu13 Tyr57 Met60 Arg82 Trp86 Thr205 Phe208 Met209 Asp212**
XI	BP, close to β-ionone ring of the retinal	x = 11.8 y = 13.3 z – 2.2	1	Thr121 Gly122 Gly125 Arg134 **Trp137 Trp138 Ser141 Trp189 Retinal**
XII	EC, close to extracellular surface	x = 16.5 y = 5.8 z = –4.8	2	**Tyr79 Arg82 Tyr83 Trp86 Trp189 Ser193 Glu194 Phe208**

In bold: lining residues including atoms other than backbone carbon, oxygen, nitrogen or C^{α}. CP, cytoplasmic channel; EC, extracellular channel; BP, retinal binding pocket.

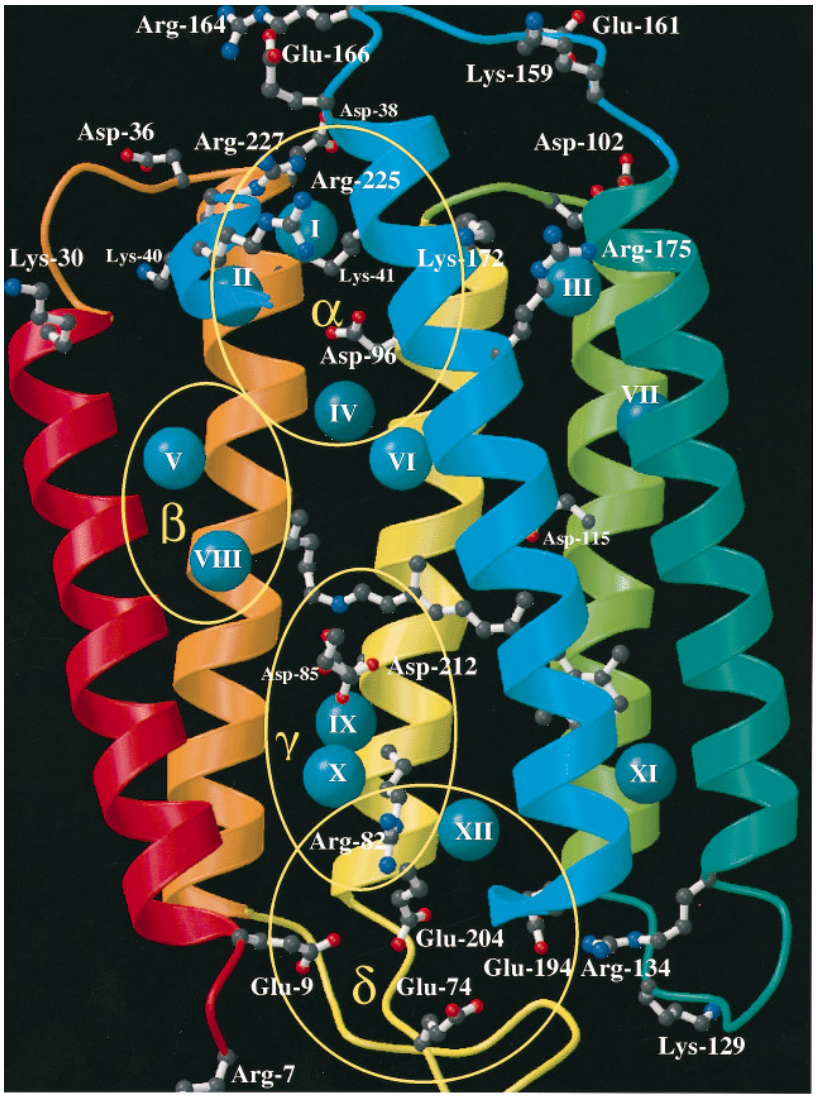
considered to be a possible interpretation of this density. Squalene has a single, methylated carbon chain with a length of 24 carbon atoms (excluding branching methyl groups), which is 1.5 times longer than the dihydrophytol chains of all the other lipids. About two molecules of squalene per bR trimer are thought to be present in PM (Kates *et al*., 1982). However, it was found that squalene was too long to fit the space in the crevice, and it was concluded that a more likely occupant of the resolved density is a dihydrophytol chain of one of the phospholipids (lipid 269) with a length of only 16 carbon atoms (excluding branching methyl groups). One or more of the lipid chains in the current model may be replaced by squalene in reality. Alternatively, the squalene may share positions statistically with ordered or disordered lipids.

Discussion

Helix conformation and packing

The parallel seven-helix bundle remains an excellent description of the overall structural arrangement. The inter-helical angles are summarised in Table 7 together with other interesting information. Three of the seven helices have a conserved proline residue about half the way along, each one associated with a kink. All three kink angles are about 20°, and the kinking appears to have three different functional roles. Pro50 forms a kink in helix B in which the resultant unpaired carbonyl oxygen of Thr46 forms part of the channel (see below). Pro91 bends helix C around the retinal to ensure closer packing above and below the

(a)



(b)

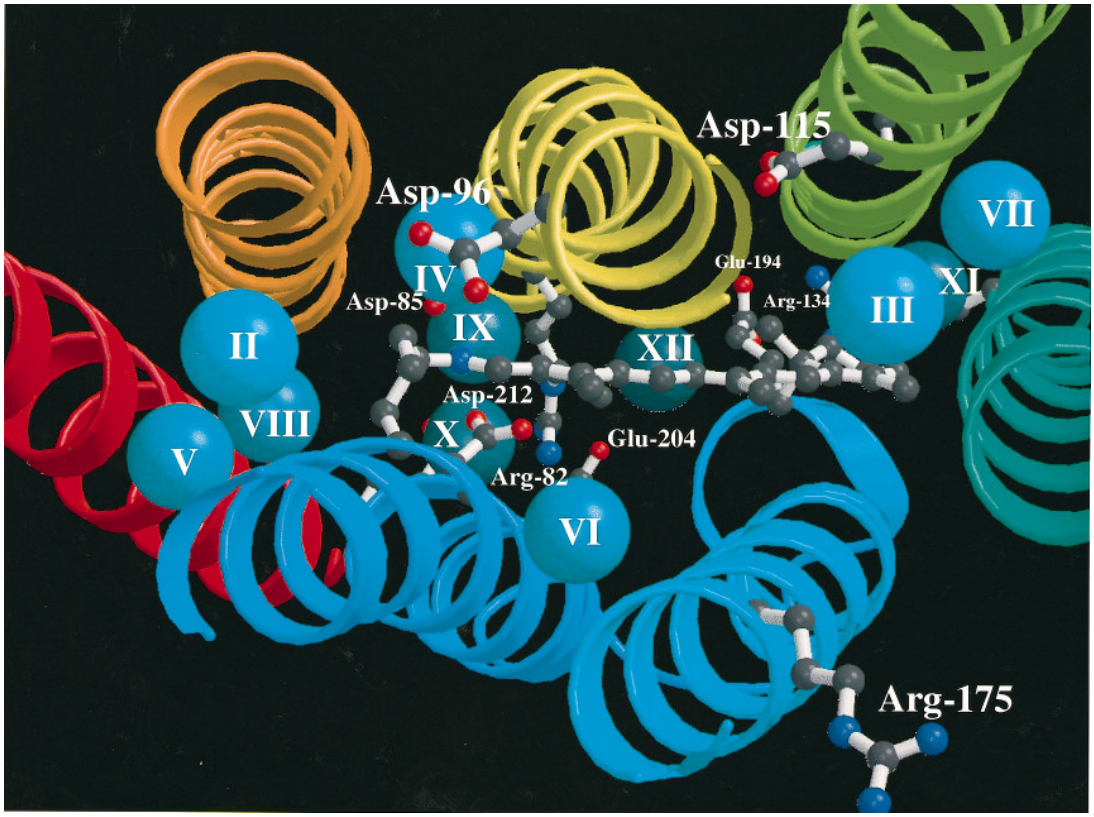


Figure 11. Cavities and H-bonded networks in bR. The side-chains of all arginine, lysine, aspartic acid and glutamic acid residues are also shown. Cavities are represented as blue spheres with a size similar to one water molecule. In (a) most of helix G was cut away to give a clearer presentation of the proton channel. Possible H-bonded networks, termed α, β, γ and δ, are outlined in yellow. They outline the proton channel starting with α on the cytoplasmic side and finishing with δ on the extracellular side. In (b) the proton channel is shown viewed from the cytoplasmic side. Cavity I near the cytoplasmic surface was omitted to make others below visible.

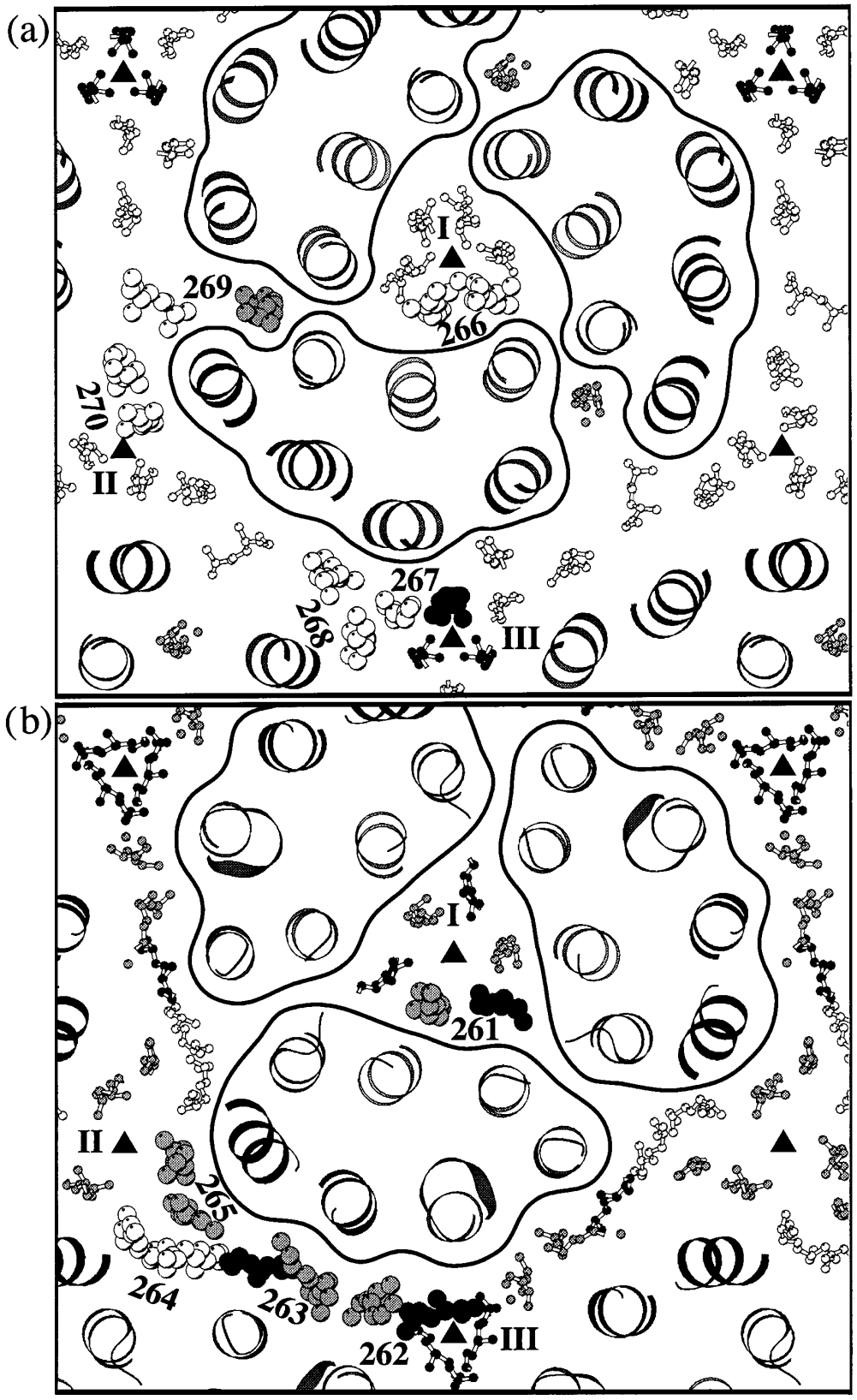


Figure 12. Schematic display of 10 Å slices of the cytoplasmic and extracellular halves of the structure to show the positions of the ten lipid molecules in the hexagonal unit cell of native purple membrane. (a) Cytoplasmic side; (b) extracellular side. The 3-fold axes are indicated with a triangle and the three different 3-fold axes are numbered with Roman numerals. Each monomer within one trimer is highlighted by a surrounding line. The unique lipids in one asymmetric unit associated in the co-ordinate list with the lower monomer are plotted using larger atoms; the symmetry-related lipids are plotted with smaller atoms. The shading of the lipids indicates their density in the map (see also Table 6): black, good density; grey, some density; white, no density.

Table 6. Positions of lipids in PM and density for the polyene chains

Lipid	Position	Density for C^{11}–C^{30}	Density for C^{41}–C^{60}
261	I, extracellular	+ +	+
262	III, extracellular	+ +	+
263	Between II and III, extracellular	+ +	+
264	II, extracellular	0	0
265	II, extracellular	+	+
266	I, intracellular	0	0
267	III, intracellular	+ +	0
268	III, intracellular	0	0
269	Crevice between bR monomers, intracellular	+	0
270	II, intracellular	0	0

+ +, Good density; +, some density; 0, no density. Lipids 262, 263 and 267 are likely to be responsible for the integrity of the two-dimensional crystal lattice, which is held together by protein–lipid–protein interactions.

retinal. Pro186 forms a very close contact with the β-ionone ring of the retinal, and may act as a hinge to allow movement on the top half of helix F during the photocycle (see below).

The mean rise per residue in the helices is 1.48 Å but there is a good deal of scatter about this value, due to the relatively low resolution of the data. Similarly, the helical screw rotation is scattered about a mean value of 98°. The mean values of Φ and Ψ within the helices are –58° and –49° and agree within 1° with the constraints used in the refinement. Examination of the Ramachandran plot (Figure 5) shows a spread along the diagonal that reflects variations in the angles between the peptide planes and the helix axes that are not explicitly constrained in our refinement. There is no evidence for the systematic outward tilt of the carbonyl groups (Krimm & Dwivedi, 1982), so the abnormal amide I carbonyl stretch infrared absorption peak at 1665 cm^{-1} must have another explanation (perhaps the low dielectric constant in purple membrane). The most deviant helical backbone conformation is that of Asp212 in helix G. Asp212 is also the residue with the largest number of definite side-chain H-bonds (to Tyr57, Tyr185 and Trp86; see Figure 10) that have been used as constraints in the refinement (see Table 1). The influence of these H-bond constraints together with the relatively low resolution of the data used in the refinement may cause the helical conformation to appear spuriously distorted. However, since we are sure of the hydrogen bonds, we prefer to present the results in this way until higher-resolution data are available.

Although the conformation in the main body of the helices is essentially that of a standard α-helix, some of the ends have a tighter 3_{10} helix conformation or a looser extra turn. Table 1 (shaded regions) shows the regions in five of the seven helices that we believe have one or two residues with a 3_{10} H-bond pattern at the N terminus, and the region of helix F that terminates with a looser turn. We have also looked carefully for particular features at the N and C termini that have been pointed out in other structures and, in some cases, have found evidence of either Ncap or Ccap features (Harper & Rose, 1993; Aurora *et al*., 1994). Four of the seven helices have proline residues at or near the N terminus, but none of these is either strongly conserved or has an obvious structural role. The presence of an Ncap (Harper & Rose, 1993) hydrogen-bonding pattern at the N termini of helices D, F and G, although possible when examining the sequence, does not occur in our model. The termination of helices by a specific conformation only allowed for a glycine residue (Aurora *et al*., 1994) is observed at the C termini

Table 7. The kinks in the helices and the inter-helical angles

Helix	Residues	Height range (Å)	Angle of kink (deg)	height of kink (Å)	Θ (deg)	Ω (deg)	Inter-helix angle (deg)
A	9 to 31	–9 to 20	—	—	21	–119	A/B1 18
							A/B2 22
B1	39 to 45	10 to 26	16	10	12	–62	B1/C2 20
B2	Pro50 to 62	–9 to 10			5	150	B2/C1 6
C1	77 to 86	–9 to 11	19	11	11	129	C1/D –9
C2	Pro91 to 100	11 to 25			15	34	C2/D –23
D	105 to 127	–8 to 24	—	—	9	–178	D/E 23
E	134 to 157	–7 to 26	—	—	15	–29	E/F1 15
							E/F2 10
F1	166 to 181	2 to 29	24	2	18	–82	F1/G 6
F2	Pro186 to 196	–10 to 2			16	12	F2/G 25
G	202 to 226	–10 to 25	—	—	14	–100	G/A 9

The two parts of the kinked helices are labelled 1 and 2. Θ is angle between helix axis and *z* axis; Ω is the azimuthal angle measured to *x* axis. All seven segments are helical within the height range –7 Å to 20 Å; the middle of the transmembrane region is at *z* = 6.5 Å. The longest helix (F) extends from –10 Å to 29 Å.

of all three helices for which a conserved glycine occurs. These are Gly31, Gly63 and Gly195 in helices A, B and F.

Ordered and disordered parts of the structure

The chromophore, its environment and the main parts of the proton channel remain essentially unchanged in the refined model. Therefore, the interpretation of the model given by Henderson *et al*. (1990) to explain the mechanism of the proton pump remains valid. Nevertheless, the refinement has produced a more accurate model and increased our understanding of its functioning. Thus, the order in the structure indicated by temperature factors could give important clues to conformational changes in the protein during the photocycle. For example, in helix F, Pro186 is part of the retinal pocket (Henderson *et al*., 1990) and was suggested to act as a hinge (Subramaniam *et al*., 1993) to allow movement of the cytoplasmic half of helix F in the formation of M as suggested by time-resolved electron diffraction (Subramaniam *et al*., 1993; Han *et al*., 1994), X-ray diffraction (Koch *et al*., 1991), neutron diffraction (Dencher *et al*., 1989) and FTIR difference spectroscopy (Ludlam *et al*., 1995). The higher temperature factor for the cytoplasmic half of helix F found in Figure 7 points to an increased flexibility in this part of the protein, which fits well with the observed movement of helix F on the cytoplasmic side. The sudden change in the temperature factor at Pro186 suggests that, indeed, this residue acts as a hinge between the more rigid extracellular half of helix F and the more mobile half on the cytoplasmic side to allow the formation of the M intermediate.

Helix D exhibits an interesting temperature factor profile (Figure 7). Close to the middle of the membrane (–2 Å to +6 Å) the temperature factors are significantly lower than in the rest of the helix. These co-ordinates coincide with the segment of helix D (Asp115 to Gly122) lining the retinal-binding pocket (Henderson *et al*., 1990). This suggests that the low temperature factors in this region of the protein are due to stabilising contacts between the helix and the retinal, particularly the relatively rigid β-ionone ring.

In general, the structure appears to be well ordered in the lower, extracellular half, so that apart from residues 66 to 80 in the BC loop, we can build into well-determined density. Higher in the structure, closer to the cytoplasmic surface, there seems to be progressively more disorder. The spline-fitted temperature factors when plotted as a function of *z*-coordinate (Figure 6(a) and Figure 7) show this most clearly. Thus, all the helices have a weaker and lower resolution appearance at their cytoplasmic surface. We cannot say whether this is due to internal flexibility and possible motion of the helices with respect to one another, or whether it is an overall rigid body rotational or librational freedom of each bR molecule with no relative helix–helix motion, or partly both. It is also possible that some of the disorder is introduced as a consequence of the specimen preparation as an air-dried, glucose-embedded thin film. The interpretation (Subramaniam *et al*., 1993) of structural changes in the M intermediate as an increased ordering at the cytoplasmic end of helix G suggests that individual helices can undergo internal movements relative to one another, as would be required to accommodate opening and closing of the proton channel leading to the cytoplasmic surface.

Cavities, water molecules and H-bonding networks

It is thought that water molecules form a major part of the complex counterion for the protonated Schiff base in the extracellular half-channel (de Groot *et al*., 1989; Deng *et al*., 1994) and that they also participate in the proton pathway in the cytoplasmic half-channel (Marti *et al*., 1991; Maeda *et al*., 1992). The entire proton channel was shown to contain at least four water molecules (Papadopoulos *et al*., 1990). Ten of the 12 cavities found in the refined structure are located either in the proton channel or in the retinal-binding pocket (Table 5 and Figure 11). These cavities suggest a pathway for protons involving both water molecules, which might occupy the cavities and ionisable side-chains (Figure 11).

Starting from the cytoplasmic side where a proton is taken up later in the photocycle, water in cavities I, II and IV (space for one water molecule each) together with Thr46, Asp96, Ser226 and Arg227 may form an H-bonded network (termed α in Figure 11), which could include water from the bulk. The volume of cavity I could be contiguous with the solvent if the side-chain of Arg227 was slightly rotated. The idea of an H-bonded network near Asp96 was also put forward on the basis of mutagenisis of these residues (Marti *et al*., 1991; Brown *et al*., 1994) and delayed reprotonation of the Schiff base from Asp96 later in the photocycle in osmotically dehydrated bR (Cao *et al*., 1991). More evidence for such a network was recently obtained by FTIR spectroscopy (le Coutre *et al*., 1995). The network could explain the interaction of Arg227 with Asp96 inferred from the slowed photocycle observed in the Arg227 to Gln mutant (Stern & Khorana, 1989; Lin *et al*., 1991) and the decreased pK_a of Asp96 in M (Zimányi *et al*., 1993). The suggested salt bridge between Arg227 and Asp96 (Stern & Khorana, 1989) seems unlikely in view of the distance (about 9 Å) between the side-chains of these two residues in the refined model. Further along the proton channel a second hydrogen-bonded network (β in Figure 11) may exist, including cavities V (space for three to four water molecules) and VIII (space for one water molecule), and the side-chain and backbone oxygen atoms of Thr46. The backbone oxygen of Thr46 cannot form an H-bond with the backbone nitrogen one turn ahead in the helix, due to it being Pro50 and,

therefore, it is capable of forming other H-bonds (Marti *et al*., 1991). Cavities V and VIII are very close and form an almost continuous volume. In agreement with the locations for cavities IV and V, FTIR studies suggest there are at least two water molecules, one in contact with Thr46 and Asp96, the other interacting only with Thr46 (Yamazaki *et al*., 1995a). The first water molecule could be located in cavity IV and the second in cavity V.

Cavity VI is close to the retinal binding pocket and between Thr178 and Trp182. It has space for a water molecule that could make H-bonds to both Thr178 and Trp182. H-bonding of Trp182 was also suggested by FTIR results (Yamazaki *et al*., 1995b). The distance between cavity VI and the other H-bonded networks suggests that a potential water molecule cannot take part in the proton pathway.

Cavities IX and X give space for one water molecule each to form a complex counterion to the Schiff base together with Tyr57, Arg82, Asp85, Trp86, Thr89, Tyr185 and Asp212 (network γ in Figure 11). The two cavities are adjacent to Arg82, which could partly occupy cavity IX after rotation of its side-chain. The participation of water in the active site of bR was suggested by NMR (de Groot *et al*., 1989), mutagenesis (Sonar *et al*., 1994), FTIR (Fischer *et al*., 1994; Maeda *et al*., 1994; Kandori *et al*., 1995) and resonance Raman spectroscopy (Deng *et al*., 1994). Cavity XI in the retinal-binding pocket appears much too distant from the proton channel to participate in proton translocation. However, a water molecule in cavity XI could H-bond to Ser141.

Near the extracellular surface, cavity XII is close to Glu9, Tyr79, Arg82, Tyr83, Arg134, Trp189, Ser193, Glu194, Glu204 and Thr205, which may all be part of a complex H-bonded network (termed δ in Figure 11) including a water molecule in cavity XII and water from the bulk phase.

Networks α to δ line the proton channel and suggest a pathway for protons. Network γ, which includes the Schiff base, would be expected to be perturbed by the isomerisation of the retinal in the K intermediate. The perturbation leads to the release of a proton on the extracellular side in the M intermediate *via* network δ near the extracellular surface. Network δ overlaps with network γ and proton transfer is likely to occur through a sequence of transfers between adjacent proton polarisable groups (proton wire; Merz & Zundel, 1981). The Schiff base is then reprotonated *via* network β in the cytoplasmic half of the channel from Asp96 in network α. Again, proton conduction may happen *via* proton polarisable groups in networks α and β with Thr46 linking the two domains. With a gap of about 6 Å between networks β and γ there is no clear pathway for protons from network β to the Schiff base. It is possible that structural changes during the photocycle bring networks β and γ closer together as part of the switch action to enable proton transfer. Alternatively, since there is some uncertainty in the orientation of side-chains leading to inaccuracies in the location of cavities, a water molecule could be closer to the Schiff base than is actually suggested by the location of cavity VIII.

The complexity of the identified networks makes it difficult to assign specific functions within the photocycle to single residues. As was also pointed out by Brown *et al*. (1995b), it is more likely that domains with many interacting groups including water are the functional elements of the proton pump. The pK_a values of individual groups then depend on other groups within the same domain. The net charge of a domain is a critical parameter for the behaviour of the domain. Networks α to γ identify four such domains in the proton channel, and may provide the basis for further work in characterising the function of each domain.

Electron form factors

Electron form factors for atoms in amino acids and proteins have not been measured before. However, gas phase electron diffraction shows that form factors of atoms in small molecules deviate significantly from calculated values assuming neutral atoms (see Methods). The refinement of electron form factors gave only a marginal improvement in the *R*-factor (about 0.3%). This may be surprising in view of the relatively large changes in the form factors after refinement. One should keep in mind, however, that the number of parameters (14 + 1 for atoms not included in any chemical group, i.e. lipids) used in the refinement is very small compared to the number of co-ordinates and temperature factors. Thus, the change in *R*-factor is still significant. It is also important to remember that, due to the limited amount of data at 3.5 Å resolution, the refinement problem is underdetermined. Addition or subtraction of a small number of parameters is unlikely to change the *R*-factor much, simply because more parameters, especially temperature factors, are needed for a correct description of the real structure. In any case, the refined value for a global parameter will still reflect a true physical property of the protein, but with limited accuracy in proportion to the number of observations available.

The refined form factors suggest that the scattering strength of atoms deviates significantly from quantum mechanical calculations assuming neutral, spherical and unbonded atoms (the independent atom model, IAM). According to Table 2 the scattering of neutral carbon, nitrogen and oxygen in the backbone of the protein increases in that order and, therefore, becomes more similar to the order found in X-ray diffraction (the refined value found for nitrogen may be somewhat too large, since it includes scattering from a hydrogen atom). The new values are in agreement with what may be expected from chemical considerations. The electron orbitals of single, neutral atoms in group IV, V and VI, such as carbon, nitrogen and oxygen, become progressively filled, thereby increasing the binding energy of electrons and making the electron distribution more compact around the nucleus. This

has the effect of more efficient screening of opposite charges in the shell and the nucleus, and a weakening of the scattering potential. Hence, the scattering amplitude for the IAM goes down from carbon to nitrogen and oxygen, but this is not true in real molecules. Atoms in a molecule or protein enter covalent bonds and atomic orbitals are filled with electrons that increase their average binding energy. Again, screening of opposite charges in the shell and the nucleus becomes more efficient so that the IAM trend is reversed. The change is largest for carbon, which has the highest valency. The progressive decrease in scattering strength between carbon, nitrogen and oxygen in the IAM, therefore, becomes smaller or may be reversed in real, bonded molecules, as suggested by the refinement results in Table 2.

Since hydrogen atoms are not included in the model their scattering contribution appears as part of the atoms they are bound to. This explains the increasing scattering strength of CH, CH_2 and CH_3 carbon groups in Table 2, which becomes larger with the number of hydrogen atoms included. The same would be expected for nitrogen and oxygen groups. The contrary result of the refinement shown in Table 2 may have two causes. First, the number of atoms found in bR for these groups is relatively small (five for NH_2 and 38 for OH), giving poor accuracy for the refined values. Second, there is an interdependence between atomic occupancies (the quantity refined in our form factor treatment) and temperature factors. An atom with a large temperature factor and a high occupancy may have an average scattering strength similar to an atom with a small temperature factor and a low occupancy. This interdependence is partly removed in the refinement of side-chain temperature factors described in Methods, where atoms from different chemical groups share the same temperature factors relative to the backbone. However, a small ambiguity may still remain. The values we find in Table 2 need to be improved by future work on a better ordered crystal, so that more accurate values can be obtained and then used without refinement in future protein structure determinations by electron crystallography.

The effect of an unscreened charge is larger than that due to covalent bonding, especially at low resolution. This is confirmed by the refined form factors in Table 2 where the scattering strengths of the positively charged nitrogen groups takes on larger values and the scattering of the negatively charged oxygen group is reduced, relative to the values for the neutral backbone groups. The reduced scattering for negatively charged side-chains in aspartic and glutamic acids may explain why these side-chains are not well resolved in the density map (e.g. Asp85 and Asp212). The only exception is Asp115, suggesting that it is protonated. Asp96 is also thought to be protonated in bR_{568} but being near the cytoplasmic surface it forms part of the structure with a higher temperature factor.

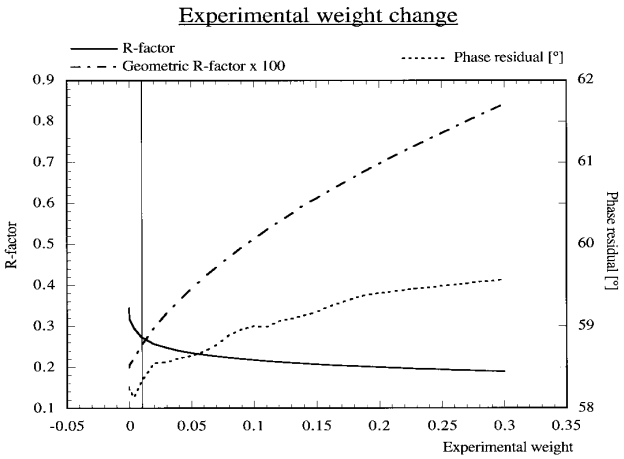


Figure 13. Variation of the relative weight of the experimental data in the refinement and change of *R*-factor, free phase residual and geometric *R*-factor. The weight is relative to the weight of geometrical restraints. The weight used in the final refinement described in the paper is indicated by a vertical line that is 1/30 of the weight usually used in X-ray work (about 0.3), and which still enables successful refinement against the data. The RMS bond length deviations corresponding to the extreme geometric *R*-factors were 0.004 and 0.022 Å.

Lipids

For the lipids, there is more resolved density on the extracellular side than there is on the cytoplasmic side (see Table 6), suggesting more disorder of the lipids on the cytoplasmic side, paralleling the behaviour of the protein. This could be due to either (1) a difference in space accessible to the lipids allowing them to move, or (2) a generally higher flexibility of the whole structure on the cytoplasmic side. In Figure 12(a) and (b) there is more space for lipids on the extracellular side, especially near 3-fold axis II, making the first possibility unlikely. Figure 7 (overall) shows that, on average, the protein structure is more disordered on the cytoplasmic side, supporting the second possible cause. Despite the stronger density on the extracellular side (Table 6) only a small decrease in the temperature factors on average for lipids on the extracellular side is visible in Figure 9.

Refinement at relatively low resolution

Apart from providing a more accurate model, the work presented here also suggests ways and strategies for the refinement of a model against data with limited resolution (3.5 Å in our case), whether from electron or X-ray diffraction. The use of the free phase residual as an independent measure of the accuracy of the model is vital for reliable refinement. Both tight geometry (Figure 13) and the use of additional geometrical restraints during the refinement (Figure 4) increase the *R*-factor, with the decreased phase residual indicating that the model had been improved.

Methods

Electron diffraction

The data used in this paper were those collected by Ceska & Henderson (1990). In that paper, three sets of high-resolution electron diffraction data were obtained, one for native purple membranes containing wild-type bacteriorhodopsin made from *Halobacterium halobium* strain R_1 and two for purple membrane treated with heavy-atom compounds in an attempt to make heavy-atom derivatives. All three sets of data were collected by electron diffraction of specimens tilted at angles up to 60° on a Philips 400 or 420 electron microscope operated at 120 kV and with the specimens at −120°C. Figure 14 shows the degree of completeness of the data. The Ceska & Henderson (1990) data were obtained using the procedure of Baldwin & Henderson (1984). Later work by Grigorieff & Henderson (1995) showed that diffuse electron scattering in the regions of the diffraction patterns between the diffraction spots was a significant source of systematic error because of the effect of the subtraction of the local background on the estimated intensities of the peaks. The electron diffraction amplitudes used to calculate both the experimental density maps and to carry out the reciprocal space refinement in the present work were therefore those corrected for the diffuse scattering (Grigorieff & Henderson, 1995). In this work, 4743 independent electron diffraction amplitudes to 3.5 Å resolution, out of 6750 collected by Ceska & Henderson (1990) to somewhat higher resolution, were used. For data beyond 3.5 Å it was found that the diffuse scattering contribution was too large to be reliably corrected. The cell dimension was taken as $a = b = 62.45$ Å in space group $p3$ with z^* sampling on a c-axis of 100.9 Å. In the model, the orthogonal co-ordinate axes have been altered from the use of (x^*, y) to (x, y^*) to fit in with the Brookhaven Databank convention that has now become standard. This means that the 1990 co-ordinates (1BRD) must be rotated by 30° before they can be compared with those described in this paper (2BRD).

Improved experimental phase information from 30 additional images

Seventy electron micrographs of two-dimensional crystals tilted up to 45° were used to produce the density map obtained previously (Henderson *et al*., 1990). Careful analysis of the accuracy of the phases produced by the merging of that data suggested that, although higher tilted data would have been desirable, for the same amount of effort more information in poorly determined regions of reciprocal space would be obtained by collecting and processing further information at lower tilt angles. Accordingly, 30 more images at tilt angles of 20° and 45° were collected (at Berkeley using Spotscan and a 100 keV field emission electron source) from air-dried, glucose-embedded specimens, and processed using exactly the same protocol as described before (Henderson *et al*., 1990). Because the new images were obtained with the benefit of experience gained during the earlier work, they were of a more consistently high quality, so that their contribution to the combined information was greater than their number might suggest.

The merged, TTF and beam tilt corrected phases were then combined with the corrected electron diffraction amplitudes using the program LATLINE (Agard, 1983) to produce fitted curves that described 3766 independent phases of the three-dimensional transform of the crystals. Individual weighting for the film quality and strength of each spot was calculated prior to and used in LATLINE. The resulting amplitudes and phases, with figure-of-merit (FOM) weighting was used to calculate the experimental map into which a completely new starting model was built. The phases were also used subsequently to calculate a free phase residual, following on from the idea of the free R-factor (Brünger, 1992, 1993).

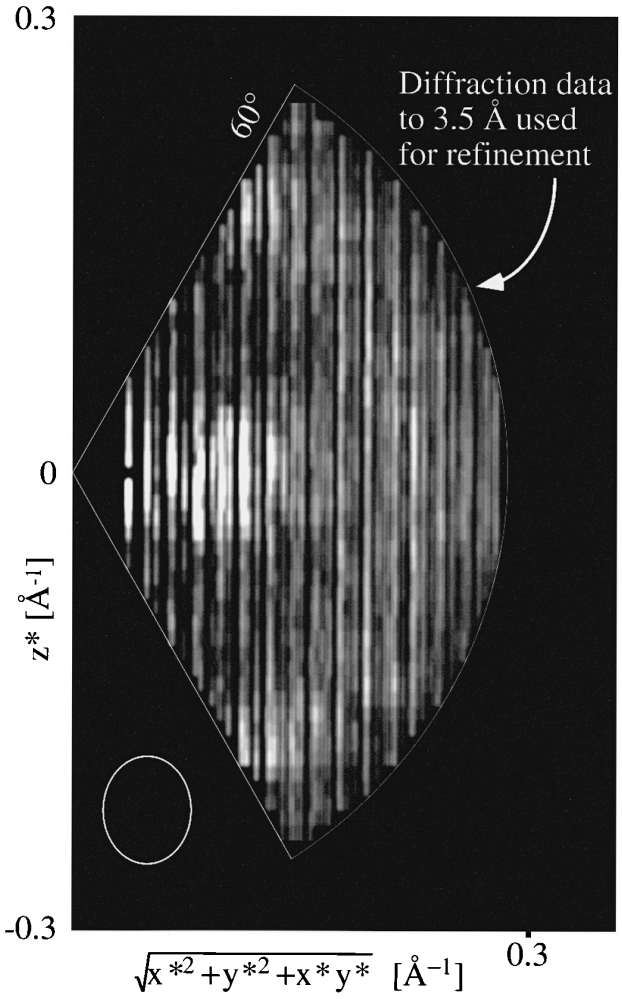


Figure 14. The diffraction data used for refinement. Data up to 3.5 Å resolution and 60° tilts were included. The vertical lines in the plot represent individual lattice lines. Their shading is modulated according to their diffraction intensity. The inset in the lower left corner shows the point spread function for a $2F_o - F_c$ map calculated using the data in the plot. The effective Fourier cut-off resolutions are 3.5 Å in the plane of the membrane and 4.3 Å perpendicular.

The free phase residual

As was pointed out by Brünger (1992, 1993) the R-factor alone is an unreliable indicator of the accuracy of a structure. The accuracy can also be independently monitored by the free R-factor, which measures objectively the agreement of the model with a small part of the experimental diffraction data not included in any refinement. This works well in X-ray diffraction work

where the amount of data is sufficient to allow exclusion of a small part without significantly affecting the refinement. In the electron diffraction from bacteriorhodopsin, the present diffraction data are more limited. On the other hand, image phase information need not be used in the refinement and, thus, could be used as a ''free phase residual'' to monitor the accuracy of the model. Since the free *R*-factor is a good measure of the phase accuracy of the model (Brünger, 1993), the free phase residual may be used to assess the model in the same way. In Figure 4, the phase residual includes 3457 comparisons between model phases and those experimentally determined with figures of merit above 0.50. The final phase residual, 58.1°, is smaller if the comparison is restricted to measurements with higher experimental FOM. For example, for FOM > 0.80 (2640 measurements) the residual is 51.4, and for FOM > 0.95 (1450 measurements) the residual is 40.7°.

Electron form factor refinement

As explained above, changes in the electron form factors of atoms due to chemical bonding are expected to be more significant than in X-ray diffraction experiments. Electron form factors for neutral atoms are listed by Cowley (1992) and are the result of quantum mechanical calculation. As was shown by gas phase electron diffraction (Bartell, 1988; Konaka, 1988) theoretical models for form factors give good results at scattering angles corresponding to resolution higher than about 1 Å. However, at resolutions below 2 to 3 Å, calculations for small molecules (e.g. H_2O or CO_2) assuming spherical atoms (independent atom model, IAM) deviate by 25% or more from experimental observations (Hirsch *et al*., 1965; Bartell, 1988; Konaka, 1988). The approach we have taken here to arrive at more accurate form factors for both neutral and charged atoms in the protein is empirical and based on the experimental data themselves.

Numerically, form factors are represented as linear combinations of Gaussians. Refinement of both exponential coefficients and exponents for each atom would have added more degrees of freedom to the model than can be justified at a resolution of 3.5 Å. Instead, one parameter describing the occupancy of each atom type was introduced. To keep the number of additional parameters small, atoms in similar chemical bonding states were grouped to share a common occupancy. Hydrogen atoms have not been explicitly included, which is appropriate for our limited resolution. However, they are implicitly included in the groupings used for the refined form factors. Fourteen chemical groups were defined for bR and are listed in Table 2. The number of additional parameters (14 + 1 for atoms other than those included in the listed chemical groups) is insignificant in comparison with the number of observations. For the refinement a new program, FREF, was written. It allows definition of each chemical group and makes alterations to both the co-ordinate file (PDB) and the form factor library. In the co-ordinate file the atomic numbers (e.g. 6 for carbon) normally serve as a reference to an entry in the form factor library during the calculation of structure factors and X-ray gradients by the program SFALL (part of the CCP4 suite, Collaborative Computational Project, 1994). FREF replaces the atomic numbers in the co-ordinate file with new identification numbers of atom types describing the new groups (e.g. 50) and amends the form factor library file. At a later stage when FREF refines form factors the entries in the library file for the newly defined atomic numbers are updated.

For the refinement of the occupancies (FREF), partial structure factors for all the atoms in each chemical group are calculated. In a linear combination of these partial structure factors the coefficients correspond to the occupancies. They are refined by minimising the *R*-factor between the linear combination and the observed diffraction amplitudes. This method has the disadvantage that refinement of occupancies, co-ordinates and temperature factors in a single refinement cycle is not possible and, therefore, several cycles of separate refinement of each of these sets of parameters in a sequence is necessary. The benefit lies in the simplicity of the implementation on a computer, since established procedures for the refinement of co-ordinates and temperature factors can be used. It was found that convergence of the refinement was satisfactory after about three cycles of sequential refinement of each type of parameter.

Temperature factor refinement

The resolution obtained in electron diffraction from two-dimensional protein crystals so far is insufficient to allow refinement of temperature factors for each atom in an atomic model (see Introduction). The density map of bR suggests that the order of the protein varies smoothly along the main-chain, with residues near the retinal binding site being well ordered and residues near the surface, especially on the cytoplasmic side, being partially or completely disordered. For the lipids, it was found that density is highest near the middle of the membrane and weak or absent for the head groups near the surface. The density for side-chains of most residues decreases with increasing distance from the backbone, due to increasing mobility. Thus, the number of parameters for the temperature factors was limited by restraining the average temperature factors for the backbone atoms in each residue to a smooth function along the main-chain. The increase in temperature factor within a residue along the side-chain was restrained to be the same for all residues. Temperature factors for the lipids were restrained to a smooth function along the axis perpendicular to the membrane surface. In all cases a cubic spline function (CSF) was chosen for the smoothing function. This allowed variations in the number of knots to be explored easily and, thus, the number of degrees of freedom for each function. In the refinement of temperature factors, the application of the restraints described was performed by a new program, BFIT.

The optimal number of knots for the cubic spline functions for the main-chain and the lipids was found by observing the free phase residual during the refinement using different numbers of knots. The result for the main-chain refinement is shown in Figure 15 where the lowest phase residual is found for 70 knots. For the lipids two knots per lipid molecule gave the lowest phase residual. For simplicity, the temperature factor refinement was carried out alternately with co-ordinate refinement.

Relative X-ray weights, the free phase residual and the *R*-factor

The *R*-factor is quoted in all crystallographic analysis as an index of agreement between observed and calculated structure factors. With the introduction of the free *R*-factor (Brünger, 1992, 1993), a more objective measure for the accuracy of a model was introduced. In the present work the free phase residual (see above) was used in a similar way. In the Hendrickson–Konnert procedure used,

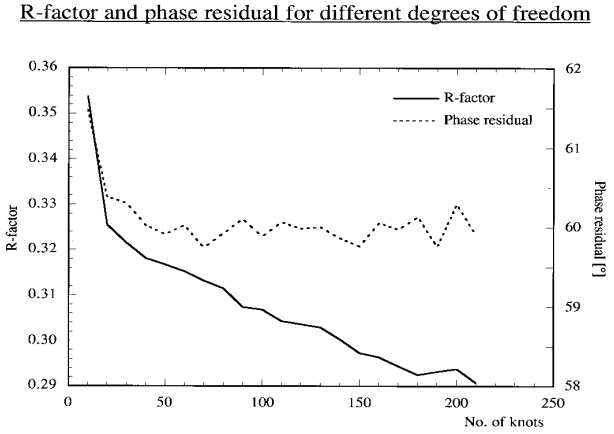


Figure 15. Variation of *R*-factor and phase residual with the number of knots for the cubic spline function describing the temperature factors along the backbone of the protein. There is a minimum at 70 knots, which was therefore the number used in subsequent refinement, and there was no further improvement in the phase residual with an increasing number of knots. The value of the *R*-factor at this early stage in the refinement is higher than that obtained later. The plot shows that the lower *R*-factor values beyond 70 knots are the result of overfitting the data with too many degrees of freedom and no improvement in the accuracy of the model.

the relative weight of the gradient or shift from the experimental data and geometrical restraints is set as an input parameter in the program PROLSQ (CCP4 program suite, Collaborative Computational Project, 1994). The weight determines the equilibrium between low *R*-factor and good geometry measured by a geometrical *R*-factor. The relative weight of the two terms normally used in X-ray work is about 0.3. To reduce the number of degrees of freedom in the refinement the weight used here was reduced 30-fold. However, to compare the refinement carried out for bR with X-ray results we investigated how the *R*-factor, free phase residual and geometrical *R*-factor change as this relative weight is changed. Forty further HK cycles of refinement of the final structure were performed using each of many different weights and the results are plotted in Figure 13. As expected, when the experimental weight is increased from 0.01 of the geometrical weight to 0.3 the *R*-factor is reduced (from 28% to 18%) and the phase residual as well as the geometrical *R*-factor rises. The RMS bond length deviation (not shown) increases from 0.005 Å to 0.022 Å.

On the other hand, when the experimental weight is reduced to 0.00 the *R*-factor increases substantially (to 35%), the phase residual first decreases slightly by 0.2° and then increases again by 0.1°, and the geometrical *R*-factor is reduced as expected.

Acknowledgements

The work described here on the refinement of the structure is a continuation of earlier work on bacteriorhodopsin in which key contributions have been made by many others. It is a pleasure to thank Erich Beckmann and Fritz Zemlin, who have recorded many of the electron micrographs on which the earlier work was based; Yoshinori Fujiyoshi, who has sent occasional micrographs for evaluation; and Bob Glaeser and Elmar Zeitler, who have constantly supported the work by encouraging us to collaborate with their laboratories. We are also very grateful to Christian Altenbach, Maarten Heyn, Janos Lanyi, Dieter Oesterhelt, Sriram Subramaniam and their colleagues for examining the co-ordinates before deposition in the databank. Finally, we thank our colleagues for their comments on the manuscript. This work was supported by EC grant ERBCHRX-CT94-0587.

References

Agard, D. A. (1983). A Least-squares method for determinating structure factors in three-dimensional tilted-view reconstructions. *J. Mol. Biol.* **167**, 849–852.

Alexiev, U., Marti, T., Heyn, M. P., Khorana, H. G. & Scherrer, P. (1994). Covalently bound pH-indicator dyes at selected extracellular or cytoplasmic sites in bacteriorhodopsin. 2. Rotational orientation of helices D and E and kinetic correlation between M formation and proton release in bacteriorhodopsin micelles. *Biochemistry*, **33**, 13693–13699.

Altenbach, C., Flitsch, S., Khorana, H. G. & Hubbell, W. L. (1989). Structural studies on transmembrane proteins. 2. Spin labeling of bacteriorhodopsin mutants at unique cysteines. *Biochemistry*, **28**, 7806–7812.

Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. (1990). Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science*, **248**, 1088–1092.

Aurora, R., Srinivasan, R. & Rose, G. D. (1994). Rules for α-helix termination by glycine. *Science*, **264**, 1126–1130.

Baldwin, J. M. & Henderson, R. (1984). Measurement and evaluation of electron diffraction patterns from two-dimensional crystals. *Ultramicroscopy*, **14**, 319–336.

Barlow, D. J. & Thornton, J. M. (1988). Helix geometry in proteins. *J. Mol. Biol.* **201**, 601–619.

Bartell, L. S. (1988). Status of electron scattering theory with respect to accuracy in structure analyses. In *Stereochemical Applications of Gas-phase Electron Diffraction* (Hargittai, I. & Hargittai, M., eds), pp. 55–83, VCH Verlagsgesellschaft mbH, Weinheim.

Benoit, J.-P. & Doucet, J. (1995). Diffuse scattering in protein crystallography. *Quart. Rev. Biophys.* **28**, 131–169.

Bousché, O., Braiman, M., He, Y.-W., Marti, T., Khorana, H. G. & Rothschild, K. J. (1991). Vibrational spectroscopy of bacteriorhodopsin mutants. *J. Biol. Chem.* **266**, 11063–11067.

Braiman, M. & Mathies, R. (1982). Resonance Raman spectroscopy of bacteriorhodopsin's primary photoproduct: Evidence for a distorted 13-*cis* retinal chromophore. *Proc. Natl Acad. Sci. USA*, **79**, 403–407.

Braiman, M. S., Mogi, T., Marti, T., Stern, L. J., Khorana, H. G. & Rothschild, K. J. (1988). Vibrational spectroscopy of bacteriorhodopsin mutants: light-driven proton transport involves protonation changes of aspartic acid residues 85, 96 and 212. *Biochemistry*, **27**, 8516–8520.

Brown, L. S., Yamazaki, Y., Maeda, A., Sun, L., Needleman, R. & Lanyi, J. K. (1994). The proton transfers in the cytoplasmic domain of bacteriorhodopsin are facilitated by a cluster of interacting residues. *J. Mol. Biol.* **239**, 401–414.

Brown, L. S., Sasaki, J., Kandori, H., Maeda, A., Needleman, R. & Lanyi, J. K. (1995a). Glutamic acid 204 is the terminal proton realease group at the

extracellular surface of bacteriorhodopsin. *J. Biol. Chem*, **270**, 27122–27126.

Brown, L. S., Váró, G., Hatanaka, M., Sasaki, J., Kandori, H., Maeda, A., Friedman, N., Sheves, M., Needleman, R. & Lanyi, J. K. (1995b). The complex extracellular domain regulates the deprotonation and reprotonation of the retinal Schiff base during the bacteriorhodopsin photocycle. *Biochemistry*, **34**, 12903–12911.

Brünger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of the structure. *Nature*, **355**, 472–475.

Brünger, A. T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and application. *Acta Crystallog. sect. B*, **49**, 24–36.

Büldt, G., Konno, K., Nakanishi, K., Plöhn, H.-J., Rao, B. N. & Dencher, N. A. (1991). Heavy-atom labelled retinal analogues located in bacteriorhodopsin by X-ray diffraction. *Photochem. Photobiol.* **54**, 873–879.

Butt, H. J., Fendler, K., Bamberg, E., Tittor, J. & Oesterhelt, D. (1989). Aspartic acids 96 and 85 play a central role in the function of bacteriorhodopsin as a proton pump. *EMBO J.* **8**, 1657–1663.

Cao, Y., Váró, G., Chang, M., Ni, B., Needleman, R. & Lanyi, J. K. (1991). Water is required for proton transfer from aspartate-96 to the bacteriorhodopsin Schiff base. *Biochemistry*, **30**, 10972–10979.

Ceska, T. A. & Henderson, R. (1990). Analysis of high-resolution electron diffraction patterns from purple membrane labelled with heavy-atoms. *J. Mol. Biol.* **213**, 539–560.

Collaborative Computational Project, Number 4. (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallog.* **50**, 760–763.

Cowley, J. M. (1992). Scattering factors for the diffraction of electrons by crystalline solids. In *International Tables for Crystallography* (Wilson, A. J. C., ed.), pp. 223–338, Kluwer Academic Publishers, Dordrecht, Holland.

de Groot, H. J. M., Harbison, G. S., Herzfeld, J. & Griffin, R. G. (1989). Nuclear magnetic resonance study of the Schiff base in bacteriorhodopsin: counterion effects on the ^{15}N shift anisotropy. *Biochemistry*, **28**, 3346–3353.

Dencher, N. A., Dresselhaus, D., Zaccai, G. & Büldt, G. (1989). Structural changes in bacteriorhodopsin during proton translocation revealed by neutron diffraction. *Proc. Natl Acad. Sci. USA*, **86**, 7876–7879.

Deng, H., Huang, L., Callender, R. & Ebrey, T. (1994). Evidence for a bound water molecule next to the retinal Schiff base in bacteriorhodopsin and rhodopsin: a resonance Raman study of the Schiff base hydrogen/deuterium exchange. *Biophys. J.* **66**, 1129–1136.

Ebrey, T. G. (1993). Light energy transduction in bacteriorhodopsin. In *Thermodynamics of Membrane Receptors and Channels* (Jackson, M. B., ed.), pp. 353–387, CRC Press, Boca Raton, FL.

Fischer, W. B., Sonar, S., Marti, T., Khorana, H. G. & Rothschild, K. J. (1994). Detection of a water molecule in the active-site of bacteriorhodopsin: hydrogen bonding changes during the primary photoreaction. *Biochemistry*, **33**, 12757–12762.

Fodor, S. P. A., Ames, J. B., Gebhard, R., van den Berg, E. M. M., Stoeckenius, W., Lugtenburg, J. & Mathies, R. A. (1988). Chromophore structure in bacteriorhodopsin's N intermediate: implications for the proton-pumping mechanism. *Biochemistry*, **27**, 7097–7101.

Fujiwara, K. (1959). Application of higher order Born approximation to multiple elastic scattering of electrons by crystals. *J. Phys. Soc. Japan*, **14**, 1513–1524.

Gerwert, K., Hess, B., Soppa, J. & Oesterhelt, D. (1989). Role of aspartate-96 in proton translocation by bacteriorhodopsin. *Proc. Natl Acad. Sci. USA*, **86**, 4943–4947.

Gerwert, K., Souvignier, G. & Hess, B. (1990). Simultaneous monitoring of light-induced changes in protein side-group protonation, chromophore isomerization, and backbone motion of bacteriorhodopsin by time-resolved Fourier-transform infrared spectroscopy. *Proc. Natl Acad. Sci. USA*, **87**, 9774–9778.

Gilles-Gonzalez, M. A., Engelman, D. M. & Khorana, H. G. (1991). Structure-function studies of bacteriorhodopsin XV. Effects of deletions in loops B-C and E-F on bacteriorhodopsin chromophore and structure. *J. Biol. Chem.* **266**, 8545–8550.

Glaeser, R. M. & Ceska, T. A. (1989). High-voltage electron diffraction from bacteriorhodopsin (purple membrane) is measurably dynamical. *Acta Crystallog. sect. A*, **45**, 620–628.

Glaeser, R. M. & Downing, K. H. (1993). High-resolution electron crystallography of protein molecules. *Ultramicroscopy*, **52**, 478–486.

Greenhalgh, D. A., Altenbach, C., Hubbell, W. L. & Khorana, H. G. (1991). Locations of Arg-82, Asp-85, and Asp-96 in helix C of bacteriorhodopsin relative to the aqueous boundaries. *Proc. Natl Acad. Sci. USA*, **88**, 8626–8630.

Grigorieff, N. & Henderson, R. (1995). Diffuse scattering in electron diffraction data from protein crystals. *Ultramicroscopy*, **60**, 295–309.

Hamanaka, T., Mitsui, T., Ashida, T. & Kakudo, M. (1972). The crystal structure of all-*trans* retinal$_1$. *Acta Crystallog. sect. B*, **28**, 214–222.

Han, B.-G., Vonck, J. & Glaeser, R. M. (1994). The bacteriorhodopsin photocycle: direct structural study of two substates of the M-intermediate. *Biophys. J.* **67**, 1179–1186.

Harper, E. T. & Rose, G. D. (1993). Helix stop signals in proteins and peptides: the capping box. *Biochemistry*, **32**, 7605–7609.

Hauss, T., Büldt, G., Heyn, M. P. & Dencher, N. A. (1994). Light-induced isomerization causes an increase in the chromophore tilt in the M intermediate of bacteriorhodopsin: a neutron diffraction study. *Proc. Natl Acad. Sci. USA*, **91**, 11854–11858.

Henderson, R. & Shotton, D. (1980). Crystallization of purple membrane in three dimensions. *J. Mol. Biol.* **139**, 99–109.

Henderson, R. & Unwin, P. N. T. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*, **257**, 28–32.

Henderson, R., Jubb, J. S. & Whytock, S. (1978). Specific labelling of the protein and lipid on the extracellular surface of purple membrane. *J. Mol. Biol.* **123**, 259–274.

Henderson, R., Baldwin, J. M., Downing, K. H., Lepault, J. & Zemlin, F. (1986). Structure of purple membrane from *Halobacterium halobium*: recording, measurement and evaluation of electron micrographs at 3.5 Å resolution. *Ultramicroscopy*, **19**, 147–178.

Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy. *J. Mol. Biol.* **213**, 899–929.

Hendrickson, W. A. & Konnert, J. H. (1981). Stereochemically restrained crystallographic least-squares refinement of macromolecule structures. In *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., ed.), pp. 43–57, Pergamon, Oxford.

Hirsch, P. B., Howie, A., Nicholson, R. B., Pachley, D. W. & Whelan, M. J. (1965). *Electron Microscopy of Thin Crystals*, p. 91, Butterworths, London, UK.

Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature*, **211**, 969–970.

Jones, T. A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallog.* **11**, 268–272.

Jones, T. A. (1985). Interactive computer graphics: FRODO. *Methods Enzymol.* **115**, 157–171.

Jones, T. A. & Kjeldgaard, M. (1993). *O–The Manual*, Uppsala University, Uppsala, Sweden.

Kandori, H., Yamazaki, Y., Sasaki, J., Needleman, R., Lanyi, J. K. & Maeda, A. (1995). Water-mediated proton transfer in proteins: an FTIR study of bacteriorhodopsin. *J. Am. Chem. Soc.* **117**, 2118–2119.

Kates, M., Kushwaha, S. C. & Sprott, G. D. (1982). Lipids of purple membrane from extreme halophiles and of methanogenic bacteria. *Methods Enzymol.* **88**, 98–111.

Kates, M., Moldoveanu, N. & Stewart, L. C. (1993). On the revised structure of the major phospholipid of *Halobacterium salinarium*. *Biochim. Biophys. Acta*, **1169**, 46–53.

Khorana, H. G., Gerber, G. E., Herlihy, W. C., Gray, C. P., Anderegg, R. J., Nihei, K. & Biemann, K. (1979). Amino acid sequence of bacteriorhodopsin. *Proc. Natl Acad. Sci. USA*, **76**, 5046–5050.

Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallog. sect. D*, **50**, 178–185.

Kleywegt, G. J. & Jones, T. A. (1995). Where freedom is given, liberties are taken. *Structure*, **3**, 535–540.

Koch, M. H. J., Dencher, N. A., Oesterhelt, D., Plöhn, H.-J., Rapp, G. & Büldt, G. (1991). Time-resolved X-ray diffraction study of structural changes associated with the photocycle of bacteriorhodopsin. *EMBO J.* **10**, 521–526.

Konaka, S. (1988). Small-angle electron scattering by gas molecules: experimental and theoretical aspects. In *Stereochemical Applications of Gas-phase Electron Diffraction* (Hargittai, I. & Hargittai, M., eds), pp. 85–105, VCH Verlagsgesellschaft mbH, Weinheim.

Krebs, M. P., Behrens, W., Mollaaghababa, R., Khorana, H. G. & Heyn, M. P. (1993). X-ray diffraction of a cysteine-containing bacteriorhodopsin mutant and its mercury derivative. Localization of an amino acid residue in the loop of an integral membrane protein. *Biochemistry*, **32**, 12830–12834.

Krimm, S. & Dwivedi, A. M. (1982). Infrared spectrum of the purple membrane: clue to a proton conduction mechanism. *Science*, **216**, 407–408.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Lanyi, J. K. (1993). Proton translocation mechanism and energetics in the light-driven pump bacteriorhodopsin. *Biochim. Biophys. Acta*, **1183**, 241–261.

le Coutre, J., Tittor, J., Oesterhelt, D. & Gerwert, K. (1995). Experimental evidence for hydrogen-bonded network proton transfer in bacteriorhodopsin shown by Fourier-transform infrared spectroscopy using azide as catalyst. *Proc. Natl Acad. Sci. USA*, **92**, 4962–4966.

Leifer, D. & Henderson, R. (1983). Three-dimensional structure of orthorhombic purple membrane at 6.5 Å resolution. *J. Mol. Biol.* **163**, 451–466.

Lin, G. C., El-Sayed, M. A., Marti, T., Stern, L. J., Mogi, T. & Khorana, H. G. (1991). Effects of individual genetic substitutions of arginine residues on the deprotonation and reprotonation kinetics of the Schiff base during the bacteriorhodopsin photocycle. *Biophys. J.* **60**, 172–178.

Lin, S. W. & Mathies, R. A. (1989). Orientation of the protonated retinal Schiff base group in bacteriorhodopsin from absorption linear dichroism. *Biophys. J.* **56**, 653–660.

Lozier, R. H., Bogomolni, R. A. & Stoeckenius, W. (1975). Bacteriorhodopsin: a light-driven proton pump in *Halobacterium halobium*. *Biophys. J.* **15**, 955–962.

Ludlam, C. F. C., Sonar, S., Lee, C.-P., Coleman, M., Herzfeld, J., RajBhandary, U. L. & Rothschild, K. J. (1995). Site-directed isotope labeling and ATR-FTIR difference spectroscopy of bacteriorhodopsin: the peptide carbonyl group of TYR185 is structurally active during the bR → N transition. *Biochemistry*, **34**, 2–6.

Maeda, A., Sasaki, J., Shichida, Y. & Yoshizawa, T. (1992). Water structural changes in the bacteriorhodopsin photocycle: analysis by Fourier transform infrared spectroscopy. *Biochemistry*, **31**, 462–467.

Maeda, A., Sasaki, J., Yamazaki, Y., Needleman, R. & Lanyi, J. K. (1994). Interaction of aspartate-85 with a water molecule and the protonated Schiff base in the L intermediate of bacteriorhodopsin: a Fourier-transform infrared spectroscopic study. *Biochemistry*, **33**, 1713–1717.

Marti, T., Otto, H., Mogi, T., Rösselet, S. J., Heyn, M. P. & Khorana, H. G. (1991). Bacteriorhodopsin mutants containing single substitutions of serine or threonine residues are all active in proton translocation. *J. Biol. Chem.* **266**, 6919–6927.

Mathies, R. A., Lin, S. W., Ames, J. B. & Pollard, W. T. (1991). From femtoseconds to biology: mechanism of bacteriorhodopsin's light-driven proton pump. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 491–518.

Merz, H. & Zundel, G. (1981). Proton conduction in bacteriorhodopsin via a hydrogen-bonded chain with large proton polarizability. *Biochem. Biophys. Res. Commun.* **101**, 540–546.

Michel, H. (1982). Characterization and crystal packing of three-dimensional bacteriorhodopsin crystals. *EMBO J.* **1**, 1267–1271.

Michel, H. & Oesterhelt, D. (1980). Three-dimensional crystals of membrane proteins: bacteriorhodopsin. *Proc. Natl Acad. Sci. USA*, **77**, 1283–1285.

Mogi, T., Stern, L. J., Marti, T., Chao, B. H. & Khorana, H. G. (1988). Aspartic acid substitution affect proton translocation by bacteriorhodopsin. *Proc. Natl Acad. Sci. USA*, **85**, 4148–4152.

Moser, M., Marsh, D., Meier, P., Wassmer, K.-H. & Kothe, G. (1989). Chain configuration and flexibility gradient in phospholipid membranes. *Biophys. J.* **55**, 111–123.

Nakasako, M., Kataoka, M., Amemiya, Y. & Tokunaga, F. (1991). Crystallographic characterization by X-ray diffraction of the M-intermediate from the photocycle of bacteriorhodopsin at room temperature. *FEBS Letters*, **292**, 73–75.

Oesterhelt, D., Tittor, J. & Bamberg, E. (1992). A unifying concept for ion translocation by retinal proteins. *J. Bioenerg. Biomem.* **24**, 181–191.

Otto, H., Marti, T., Holz, M., Mogi, T., Lindau, M.,

Khorana, H. G. & Heyn, M. P. (1989). Aspartic acid-96 is the internal proton donor in the reprotonation of the Schiff base of bacteriorhodopsin. *Proc. Natl Acad. Sci. USA*, **86**, 9228–9232.

Ovchinnikov, Y. A., Abdulaev, N. G., Feigina, M. Y., Kiselev, A. V. & Lobanov, N. A. (1979). The structural basis of the functioning of bacteriorhodopsin: an overview. *FEBS Letters*, **100**, 219–224.

Papadopoulos, G., Dencher, N. A., Zaccai, G. & Büldt, G. (1990). Water molecules and exchangeable hydrogen ions at the active centre of bacteriorhodopsin localized by neutron diffraction. *J. Mol. Biol.* **214**, 15–19.

Petrich, J. W., Breton, J., Martin, J. L. & Antonetti, A. (1987). Femtosecond absorption spectroscopy of light-adapted and dark-adapted bacteriorhodopsin. *Chem. Phys. Letters*, **137**, 369–375.

Pfefferlé, J.-M., Maeda, A., Sasaki, J. & Yoshizawa, T. (1991). Fourier transform infrared study of the N intermediate of bacteriorhodopsin. *Biochemistry* **30**, 6548–6556.

Polland, H.-J., Franz, M. A., Zinth, W., Kaiser, W., Kölling, E. & Oesterhelt, D. (1986). Early picosecond events in the photocycle of bacteriorhodopsin. *Biophys. J.* **49**, 651–662.

Ramachandran, G. N. & Srinivasan, R. (1961). An apparent paradox in crystal structure analysis. *Nature*, **190**, 159–161.

Rothschild, K. J. (1992). FTIR difference spectroscopy of bacteriorhodopsin: toward a molecular model. *J. Bioenerg. Biomembr.* **24**, 147–167.

Santarsiero, B. D., James, M. N. G., Mahendran, M. & Childs, R. F. (1990). Crystal structure of *N*-methyl-*N*-phenylretinal iminium perchlorate: a structural model for the bacteriorhodopsin chromophore. *J. Am. Chem. Soc.* **112**, 9416–9418.

Schertler, G. F. X., Bartunik, H. D., Michel, H. & Oesterhelt, D. (1993). Orthorhomic crystal form of bacteriorhodopsin nucleated on benzamidine diffracting to 3.6 Å resolution. *J. Mol. Biol.* **234**, 156–164.

Seelig, J. & Seelig, A. (1980). Lipid conformation in model membranes and biological membranes. *Quart. Rev. Biophys.* **13**, 19–61.

Sharkov, A. V., Pakulev, A. V., Chekalin, S. V. & Matveetz, Y. A. (1985). Primary events in bacteriorhodopsin probed by subpicosecond spectroscopy. *Biochim. Biophys. Acta*, **808**, 94–102.

Simmons, C. J., Liu, R. S. H., Denny, M. & Seff, K. (1981). The crystal structure of 13-*cis*-retinal. The molecular structures of its 6-*s*-*cis* and 6-*s*-*trans* conformers. *Acta Crystallog. sect. B*, **37**, 2197–2205.

Simmons, C. J., Asato, A. E. & Liu, R. S. H. (1986). Structure of all-*trans*-3,4-didehydroretinal (Retinal$_2$). *Acta Crystallog. sect. B*, **42**, 711–715.

Smith, S. O., Pardoen, J. A., Mulder, P. P. J., Curry, B., Lugtenburg, J. & Mathies, R. (1983). Chromophore structure in bacteriorhodopsin's O_{640} photointermediate. *Biochemistry*, **22**, 6141–6148.

Sonar, S., Marti, T., Rath, P., Fischer, W., Coleman, M., Nilsson, A., Khorana, H. G. & Rothschild, K. J. (1994). A redirected proton pathway in the bacteriorhodopsin mutant Tyr-57 → Asp. *J. Biol. Chem.* **269**, 28851–28858.

Souvigner, G. & Gerwert, K. (1992). Proton uptake mechanism of bacteriorhodopsin as determined by time-resolved stroboscopic-FTIR-spectroscopy. *Biophys. J.* **63**, 1393–1405.

Stam, C. H. (1972). The crystal structure of a monoclinic modification and the refinement of triclinic modification of vitamin A acid (retinoic acid), $C_{20}H_{28}O_2$. *Acta Crysallog. sect. B*, **28**, 2936–2945.

Stern, L. J. & Khorana, H. G. (1989). Structure-function studies on bacteriorhodopsin. X. Individual substitutions of arginine residues by glutamine affect chromophore formation, photocycle, and proton translocation. *J. Biol. Chem.* **264**, 14202–14208.

Subramaniam, S., Gerstein, M., Oesterhelt, D. & Henderson, R. (1993). Electron diffraction analysis of structural changes in the photocycle of bacteriorhodopsin. *EMBO J.* **12**, 1–8.

Tsujimoto, K., Yorimitsu, S., Takahashi, T. & Ohashi, M. (1989). Revised structure of a phospholipid obtained from *Halobacterium halobium*. *J. Chem. Soc. Chem. Commun*. 668–670.

Tsygannik, I. N. & Baldwin, J. M. (1987). Three-dimensional structure of deoxycholate-treated purple membrane at 6 Å resolution and molecular averaging of three crystal forms of bacteriorhodopsin. *Eur. Biophys. J.* **14**, 263–272.

Ulrich, A. S., Heyn, M. P. & Watts, A. (1992). Structure determination of the cyclohexene ring of retinal in bacteriorhodopsin by solid-state deuterium NMR. *Biochemistry*, **31**, 10390–10399.

Ulrich, A. S. & Watts, A., Wallat, I. & Heyn, M. P. (1994). Distorted structure of the retinal chromophore in bacteriorhodopsin resolved by ^{2}H-NMR. *Biochemistry*, **33**, 5370–5375.

Unwin, P. N. T. & Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* **94**, 425–440.

Yamazaki, Y., Hatanaka, M., Kandori, H., Sasaki, J., Karstens, W. F. J., Raap, J., Lugtenburg, J., Bizounok, M., Herzfeld, J., Needleman, R., Lanyi, J. K. & Maeda, A. (1995a). Water structural changes at the proton uptake site (the Thr46-Asp96 domain) in the L intermediate of bacteriorhodopsin. *Biochemistry*, **34**, 7088–7093.

Yamazaki, Y., Sasaki, J., Hatanaka, M., Kandori, H., Maeda, A., Needleman, R., Shinada, T., Yoshihara, K., Brown, L. S. & Lanyi, J. K. (1995b). Interaction of tryptophan-182 with the retinal 9-methyl group in the L intermediate of bacteriorhodopsin. *Biochemistry*, **34**, 577–582.

Zimányi, L., Váró, G., Chang, M., Ni, B., Needleman, R. & Lanyi, J. K. (1992). Pathways of proton release in the bacteriorhodopsin photocycle. *Biochemistry*, **31**, 8535–8543.

Zimányi, L., Cao, Y., Needleman, R., Ottolenghi, M. & Lanyi, J. K. (1993). Pathway of proton uptake in the bacteriorhodopsin photocycle. *Biochemistry*, **32**, 7669–7678.

Edited by R. Huber

(Received 11 December 1995; received in revised form 22 March 1996; accepted 28 March 1996)