LINKING REGIONS BETWEEN HELICES IN BACTERIORHODOPSIN REVEALED

DAVID A. AGARD AND ROBERT M. STROUD
Department of Biochemistry and Biophysics, University of California School of Medicine,
San Francisco, California 94143

ABSTRACT Three-dimensional electron-microscopic structural analysis requires the combination of many different tilted views of the same specimen. The relative difficulty of tilting the sample to high angles >60° without introducing severe distortion due to different focal distances across the specimen entails that the observable range of electron diffraction data is often limited to this range of angles. Thus, it is generally not possible to observe the diffraction maxima that lie within the conical region of reciprocal space around the direction perpendicular to the electron microscope grid. The absence of data in this region leads to a predictable distortion in the object, and for ±60° tilting makes the resolution essentially twice as bad in the direction perpendicular to the grid as it is for the in-plane image. Constrained density map modification and refinement methods can significantly reduce these effects. A method has been developed, tested on model cases, and applied to the electron-microscopic structure determination of bacteriorhodopsin in order to visualize the location of linking regions between helices. Electron-microscopic structural analysis of bacteriorhodopsin (Henderson and Unwin. 1975 Nature [Lond.] 257:28–32.) showed that the molecule consists of seven rods of density each nearly spanning the lipid bilayer. Owing to the distortion introduced by the missing conical region of reciprocal space data, no density was visible for the polypeptide segments linking the α-helices. Density in the refined maps indicates the location of at least five of the extrahelical segments of the polypeptide. The total number of possible ways of interconnecting the helices is reduced from 28! (5,040) to the five most consistent possibilities without recourse to other considerations. In addition, the density for the helical regions is more uniform and cylindrical throughout their length, and the length of the helices increases from 35 to 45 Å, close to the membrane thickness of 49 Å obtained for membranes dried in vacuo. Only three of the five structures consistent with the location of observed linkers place the seventh helix, onto which the chromophore can be attached by reduction in the light, at a position consistent with the main peak for deuterated retinal in the structure, as derived from neutron diffraction analysis. Two of these models are also consistent with the possible location of some of the reduced chromophore on helix B, at lys 40/41 after reduction in the dark, as well as lys 216 on helix G.

INTRODUCTION
The analysis of glucose-sustained two-dimensional crystals of catalase and of bacteriorhodopsin by Unwin and Henderson (1975), Henderson and Unwin (1975), and the similar analyses of frozen hydrated specimens by Glaeser and his colleagues (Hayward et al., 1978; Taylor, 1978; see also Hayward and Stroud, 1981) have secured the role of electron microscopy for moderate resolution structural studies of biological specimens, especially membranes where two-dimensional ordering is apparent.

Data collection for three-dimensional electron-microscope image reconstruction requires that many different tilted two-dimensional projections of the same material be combined. The effects of limited tilt angle on the final reconstructed image are most easily described with respect to the three-dimensional Fourier transform \( F(s) \), where \( s \) is the scattering vector of amplitude \( 2 \sin \theta / \lambda \) of the three-dimensional object \( \rho(r) \) (where \( r \) is the real-space vector) (Fig. 1). Any given projection of the object defines a plane of the Fourier transform \( F(s') \) perpendicular to the viewing direction \( s_0 \) (i.e., for \( s' \cdot s_0 = 0 \)), and passing through the origin \( |s| = 0 \). The final reconstructed image is the inverse Fourier transform of \( F(s) \). In practice, it is generally more difficult to collect electron microscope data when the specimen is tilted to high angles. Although it is possible to collect 80° tilted images, experimental limitations and a simpler analysis of images often are limited to ~60° from the horizontal, especially where phases are to be derived from electron images (Hayward and Stroud, 1981). As a result, no reciprocal space data would be obtained within a conical region of the reciprocal space transform \( F(s) \) about the axis normal to the specimen plane. The angular restriction arises from both instrumental limitations: curvature of the Ewald sphere and the large changes in focus occurring across a sample tilted to such steep angles (~8,700-Å focal change for a 1-μm diameter sample at 60° tilt angle).

The absence of data within this conical region intro-
The treatment presented is directed most specifically to the electron imaging problem, though the solution is applicable in identical form to the limited angle reconstruction problem in computer-aided tomography (CAT). In a more general sense, it can be usefully employed for resolution extension, or completion of any incomplete imaging data set.

**METHOD**

**Theory**

The approach taken was to use the observed data and several global constraints that apply to the object, in order to extrapolate into the unobservable region. An iterative procedure is used to fill in the missing data so as to generate a structure that is consistent with the constraints and that correctly predicts the observed \(F_0(s)\). The iterative approach schematized in Fig. 2 was chosen over an analytic method (based on a Taylor series expansion, for example) because of the ease of incorporating constraints in either the Fourier-space or real-space domains.

All such extrapolation techniques are based in some way on the fact that each point in the Fourier transform \(F(s)\) of an object derives from each point in the entire object according to

\[
F(s) = \frac{1}{V} \int \rho(r) e^{i2\pi rv} dv,
\]

where \(V\) is the total volume of the unit cell, or repeated unit of the crystal. Thus, there is information in the observed data \(F(s)\) concerning the nature of the unobserved data, since each point in the unobserved portion of \(F(s)\) derives from the same \(\rho(r)\) (see, for example, Hoppe, 1969, or Crowther et al., 1970).

Any physically realistic object occupies a finite volume. In the case of a membrane-bound protein in a two-dimensional lattice, the thickness of the membrane is limited, as is also the volume \(V\) in the unit cell occupied by the ordered components, i.e., single protein molecules \((U < V)\). These constraints entail a powerful joint set of relationships among all the \(F(s)\), which provides the basis for the method. This single-boundary condition is very powerful and, in theory, allows a unique extension of the diffraction data into unobserved regions (Wolter, 1961; Frieden, 1975) since it implies that \(\{F(s)\}\) is a continuous analytic function. This same kind of constraint is also sufficient to allow de novo calculation of phases for continuous diffraction data \(F(s)\), as, for example, in our method for the determination of electron density profiles of biological membranes.

**FIGURE 1** The limited ability to tilt the object or sample to angles >60° results in a missing conical region of diffraction data (a, b). As a result, each point in the structure is convoluted with a function \(\rho^*\) which has the symmetry of a prolate ellipsoid of axial ratio ~1.8 (c). The product in diffraction space entails that each atom to be imaged within the structure is convoluted with the Fourier transform of \(G(s)\) schematized in Fig. 1 c, as \(\rho^*(r)\). Each atom is blurred in the density map to an approximately prolate ellipsoidal shape with a major axis ~1.8 × the minor axis. The resolution is consequently almost twice as bad in the z-direction perpendicular to the specimen. As a result, segments of chain running parallel to the z-axis tend to be tapered and smoothed by the large overlap of ellipsoidal functions \(\rho^*(r)\), whereas chains parallel to the plane of the specimen can, in some cases, be obliterated. Overlap in the plane or at the top and bottom of the sample is much less and density is consequently much lower. In the case of bacteriorhodopsin (Unwin and Henderson, 1975; Henderson and Unwin, 1975) it has been observed that there are seven rods of density (thought to correspond to α-helices, Henderson, 1975; Blaurock, 1975) running almost perpendicular to the membrane (approximately parallel to the z-direction). However, neither the connections between the rods nor the chromophore, whose major axis is known to lie ~23° away from the membrane plane (Bogomolni et al., 1977), could be observed. This is precisely the type of distortion expected to arise from the missing cone of information. As the mathematical nature of the distortion is known, its effects can be greatly reduced. We have developed a constrained iterative Fourier refinement procedure to remove this distortion. The method was applied to the bacteriorhodopsin data of Henderson and Unwin (1975), with the result that five regions of density between ends of the helices were revealed. The procedure has been tested on various model data sets, including one strictly analogous to the case with bacteriorhodopsin.

**FIGURE 2** Schematic diagram of the constrained refinement procedure. Here \(F_0(s)\) and \(\phi_0(s)\) are constrained to equal \(F_0(s)\) and \(\phi_0(s)\) in the observed region.
(Stroud and Agard, 1979; Roos et al., 1977) where the relevant \( F(s) \) are observable, but the phases are not. We have developed similar procedures for resolution enhancement of one- and two-dimensional electrophoretograms (Agard et al., 1981), and for phase refinement in protein crystallography (Agard and Stroud, 1981). In the present application both amplitudes and phases are known for a subset \( F_d(s) \).

The density map computed from the observed data alone,

\[
\rho_{o}(r) = \int_{v_{o}(s)} F_d(s) e^{-2\pi i r d s},
\]

is necessarily inconsistent with the known constraints of volume, thickness, etc., which apply to the real object, because \( F_d(s) \) is incomplete. The unobserved portion of \( F(s) \) must, therefore, account for the discrepancy between \( \rho(r) \) and the known constraints. At the same time, each point within the constrained volume \( U \) computed by transformation of \( F(s) \) is affected by each new value of \( F(s) \). There is thus a means for iterative refinement of the unknown \( F(s) \) that improves the density function both outside (to a constant level) and inside the constrained molecular volume. Some knowledge of the molecular volume \( U \), which includes the entire nonzero part of \( \rho(r) \), is required.

Sensitivity to data errors can be substantially reduced by incorporation of additional knowledge or constraints, such as non-negativity of scattering density (a further constraint incorporated in the refinement scheme we used). With the positivity constraint, additional error components in the extrapolated region produced by errors in the data are greatly minimized and are generally lower than the errors in the observed region, which makes the method remarkably insensitive to data errors. These two constraints are also sufficient to allow for phase refinement-phase extension (where amplitudes are available to higher resolution than are the phases). Neither constraint alone will suffice.

Initially, we performed an iterative test of the efficacy of the refinement procedure were carried out first for model structures in two and three dimensions with restricted sets of calculated data and data modified by inclusion of random errors in \( F_d(s) \), which were then mapped onto the experimental data for bacteriorhodopsin. Finally, a test was performed using data calculated from a molecular model built to match the density map derived for bacteriorhodopsin, except that every residue was replaced as alanine. The missing region was removed to parallel exactly the case with the real bacteriorhodopsin data set. In all cases, the distortions due to the angular limitations of the data set were dramatically reduced.

**Initial Test of the Method**

The first test model was a sequence of three connected \( \alpha \)-helices extracted from the myoglobin structure (residues 70-149, Brookhaven data bank) and slightly modified for clarity according to the protein interactive graphics system developed by R. Langridge. The model was used to compute an initial set of \( F(s) \), which was subsequently modified to resemble an observed set \( F_d(s) \) by removal of unobservable data, inclusion of error, resolution cut off, etc. For the two-dimensional studies only the main chain and \( \beta \)-carbon atoms were used to simplify the structure as viewed in projection. All atoms were used to simplify the structure as viewed in projection. All atoms were used in the three-dimensional studies. The models were placed in either a \( 64 \times 64 \times 64 \) two-dimensional unit cell, or in a \( 64 \times 64 \times 32 \) three-dimensional unit cell. Structure factors were calculated for all reflections to either 3.5-\( \AA \) (two-dimensional) or 7-\( \AA \) (three-dimensional) resolution with a temperature factor of 15 \( \AA^2 \) for each atom.

To model the angular limitations inherent in a \( \pm 60^\circ \) tilt three-dimensional electron-microscope data set, data were omitted from a \( \pm 30^\circ \) region of \( F(s) \) about what was designated the \( x \)-axis in Fig. 3 (i.e., perpendicular to the two longest helices). Under these conditions, fully \( 33\% \) of the two-dimensional data and 14% of the three-dimensional data were unobserved in the starting \( F_d(s) \) data set. Rough boundaries used to delineate the volume (or area) \( U \) occupied by protein were drawn so that 50% of each unit cell was considered solvent. This volume approximates values found in many protein crystals and in the purple membrane unit.

**Application to Bacteriorhodopsin**

The set of three-dimensional electron microscopic data recorded by Henderson and Unwin (1975) for bacteriorhodopsin was kindly provided by Dr. Richard Henderson. Although the data extend to a resolution of 7 \( \AA \) in the plane of the membrane, the resolution perpendicular to the plane is only 14 \( \AA \) (Henderson and Unwin, 1975). Thus, the effects of the unobserved data (53% of total 7-\( \AA \) data) are more serious than for the missing cone region alone. Refinement was used not only to extrapolate into the missing cone region but to extend the perpendicular resolution to 10 \( \AA \) (i.e., to predict the 47% missing data). The data were sampled at an interval corresponding to a 100-\( \AA \) \( c \)-axis cell length.

By comparison with the model cases, the task of properly defining real-space constraints was more complex. It was necessary to use smoothed boundary functions (Fig. 4); otherwise, serious ripples appeared in the density functions and the procedure failed to converge. The \( F(000) \) term corresponding to the absolute density level was chosen from the extrapolated value of a Wilson plot (Wilson, 1942) at \( s = 0 \). To minimize the deleterious effects of inaccuracies in the density levels outside the boundary function (outside of the membrane, and outside of the molecule within the bilayer) the very low angle data (>20-\( \AA \) resolution) were omitted from the starting data set. This serves to decouple the density assignments in different regions outside of the boundary. Thus, the constraints on density levels outside of a boundary become local instead of global in nature. Such an approach is a common feature of many two-dimensional image enhancement algorithms.

The refinement of bacteriorhodopsin data was carried out in three stages starting with both in-plane and perpendicular volume constraints as well as non-negativity. After 10 cycles, the non-negativity constraint
and bacteriorhodopsin, and our refinement scheme was applied exactly as for bacteriorhodopsin.

RESULTS

Two-dimensional Model Studies

The distortions introduced into the structure as a result of placing $\pm 60^\circ$ angular restriction on the observed 3.5-Å resolution two-dimensional data set are shown in Fig. 5a, b. Comparison with the undistorted structure in Fig. 5a shows how regions of chain or helix running perpendicular to the $x$-axis (horizontal, assumed to be the electron beam direction) are almost totally obliterated. After 30 cycles of constrained refinement starting from the observed data shown in Fig. 5b and with the boundary area shown in Fig. 3, much of the original structure is restored as shown in Fig. 5c. Although the restoration is certainly not perfect, the improvement is dramatic and sufficiently accurate for the polypeptide chain to be unambiguously traced.

The restoration of the missing sector of Fourier data is also shown in Fig. 5 by the half-tone representation of the diffraction amplitudes $|F(s)|$. The residual,

$$R = \sum_s \frac{|F_T(s)| - |F_c(s)|}{\sum_s |F_T(s)|},$$

provides a quantitative assay of the mean discrepancy between the original or true value of $F(s)$, $F_T(s)$ (Fig. 5a), and the calculated values $F_c(s)$ (Fig. 5c). The residual for the true-vs.-calculated data in the missing sector is 45.3% while the rms phase error,

$$\phi_{rms} = \left(\sum_s \left[\phi_T(s) - \phi_c(s)\right]^2\right)^{1/2},$$

is 70°. This corresponds to a mean figure of merit $(\cos \left[\phi_T(s) - \phi_c(s)\right])$ of 0.52 (Blow and Crick, 1959). The residual calculated within the observed region is 4.7%. Although the errors for the extended region $F_c(s)$ may seem large, it is necessary to keep in mind that fully one-third of the total data is predicted by the refinement. Even so, these errors are no worse than occur in the early stages of any macromolecular x-ray structure determination. The nonzero residual in the observed region results from errors introduced by the constraints: at any finite resolution there are series termination effects that imply non-positivity of density values in the resolution limited map, when correct $F_T(s)$ values are used for $s$ less than some $s_{\text{max}}$. The chosen constraints do not allow for non-negative values in the refined structure. In this sense, the positivity constraint, although physically correct, is not procedurally correct. The errors introduced are negligible, however, and the power of the constraint is considerable. Too much is gained, especially in the early stages of refinement, to relax this constraint, and little is to be gained even in the end by altering the lower boundary.

**Figure 4** The mask functions used for the bacteriorhodopsin experiments. The in-plane mask is shown contoured in (a) and the transmembrane mask is shown in (b).

**Table 1** The mask functions used for the bacteriorhodopsin experiments. The in-plane mask is shown contoured in (a) and the transmembrane mask is shown in (b).

**Table 2** The mask functions used for the bacteriorhodopsin experiments. The in-plane mask is shown contoured in (a) and the transmembrane mask is shown in (b).

**Table 3** The mask functions used for the bacteriorhodopsin experiments. The in-plane mask is shown contoured in (a) and the transmembrane mask is shown in (b).

**Table 4** The mask functions used for the bacteriorhodopsin experiments. The in-plane mask is shown contoured in (a) and the transmembrane mask is shown in (b).
FIGURE 5 The results of the two-dimensional refinement. Shown are the structures and their Fourier transforms (halftone images) for (a) the true structure, (b) before refinement, indicating the severity of the distortion produced by a 30° semiangle missing cone parallel to the horizontal axis, and (c) after 30 cycles of constrained Fourier refinement. Note that the missing region lies along the horizontal axis.

Three-dimensional Experiments
Severe distortions are also introduced into a three-dimensional structure when a conical region of half-angle 30° is removed from the total (true) 7-Å data set (Fig. 6a, b). Proper chain connectivity is lost for those segments running perpendicular to the axis of the missing conical region (horizontal). The resultant distortion is less severe than in the two-dimensional case, owing to the reduced percentage of missing data (14 vs. 33%). 25 cycles of refinement lead to nearly perfect restoration (Fig. 6c); chain connectivity is restored and the elongated regions return to their undistorted sizes and shapes. At convergence the residual for the predicted amplitudes is 15.3%; the rms phase error is 28.2° (a mean figure of merit of 0.92). As in the two-dimensional case, the nonzero residual for observed data (R = 5.5%) reflects inaccuracies in the assumptions and gives an empirical estimate of the error introduced by these assumptions. These errors are slightly more severe at 7 than at 3.5 Å because of the larger series termination ripples. The use of a lower cutoff, which accommodates series termination ripples in early stages of refinement, slowed convergence; the low density cutoff was therefore modified only in the final stages.

Effects of Data Errors
The effects of errors in the data were investigated by introducing random errors into the observed data set. Errors introduced into the data set are described by a residual between modified $|F_0(s)|$ and unmodified $|F(s)| = |F_0(s)| = |F_0(s)|$ amplitudes (Eq. 2), and by an rms phase difference $\Delta \phi_{\text{rms}}$ between the two sets (Eq. 3). In the tests described in Table I, random errors were introduced either into the amplitudes alone ($R = 19.7\%$, column 2) or into

FIGURE 6 Several sections through the three-dimensional density maps for noise-free data: (a) the true structure, (b) before refinement, and (c) after 25 cycles. As in Fig. 5, the axis of the missing cone of data is taken along the horizontal axis.
TABLE I
ANALYSIS OF THE SENSITIVITY TO DATA ERRORS

<table>
<thead>
<tr>
<th></th>
<th>$F_0$, no error</th>
<th>$F_0$, errors in $F_0$ only</th>
<th>$F_0$, errors in $F_0$, $\phi$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R$</td>
<td>$\Delta \phi$</td>
<td>$R$</td>
</tr>
<tr>
<td>Errors introduced in observed data</td>
<td>0</td>
<td>0</td>
<td>19.7</td>
</tr>
<tr>
<td>Unobserved region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolated vs. true data</td>
<td>17.9</td>
<td>25.5</td>
<td>18.4</td>
</tr>
<tr>
<td>Observed region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Refined vs. true data</td>
<td>7.5</td>
<td>9.0</td>
<td>10.9</td>
</tr>
<tr>
<td>Refined vs. observed data</td>
<td>7.5</td>
<td>9.0</td>
<td>15.9</td>
</tr>
</tbody>
</table>

Shown are the residuals and rms phase errors for restricted data sets with no errors, errors in the amplitudes alone, and errors in both the amplitudes and phases. The values reported are after 10 cycles of constrained refinement. There is almost no change in the accuracy of the extrapolated data when errors are incorporated into the starting data set. Since the errors themselves are inconsistent with the constraints, data in the observed region refines towards error-free values. The remarkable insensitivity to data errors is readily apparent.

both amplitudes and phases ($R = 17.4\%$, $\Delta \phi_{\text{rms}} = 29°$, column 3) and a conical region of semiangle 30° was removed to generate the observed data sets $F_0(s)$, $F_0''(s)$.

10 cycles of refinement gave the results summarized in Table I and Fig. 7, which demonstrate a remarkable degree of immunity to error. When perfect error-free data were used in the observed region, the solution after 10 cycles gave residuals for extrapolated-vs.-true data of 17.9% ± 25.5°. Introduction of errors into the observed region resulted in residuals only very slightly worse (18.4% ± 28.0° and 21.7% ± 25.1° for the $F_0$ and $F_0''$ data sets, respectively). Thus, the effective figures of merit are 0.92–0.94 for the derived data, even when errors of the order of 20% ± 30° are present in the observed data set.

There is an obvious tendency for refined data in the observed region to converge toward the observed data set, since at each cycle observed amplitudes and phases are reassociated with extrapolated values in the unobserved region. Nevertheless, the errors between the values computed in the final cycle and the true data in the observed region were only slightly worse (10.9% ± 12.0°, 11.2% ± 18.5°) when error-flawed data were used than when refinement was based on error-free data (7.5% ± 9.0°). Thus, there is a powerful tendency to correct for data errors that are themselves inconsistent with the constraints. This potential for improvement of data errors is not fully used in the method as presented here since observed values are always given full weight at each cycle. A weighting scheme would allow for some refinement of observed amplitudes and phases. This approach is also of use in refinement of early phases in protein crystallography (Agard and Stroud, 1981).

For completeness, residuals calculated between refined and observed (i.e., error-flawed rather than true) data in the observed region are listed in Table I. These residuals rise more rapidly than for the comparison of refined and
true data as errors are introduced, again symbolic of the inconsistency between flawed data and a self-consistent result, and of the tendency to minimize effects of random error.

**Bacteriorhodopsin**

As was found in the model examples, the iterative Fourier refinement technique was able to minimize much of the distortion caused by the absence of observed data. The final residual for predicted data in the observed region was 11.2%. Views of a model built to the reconstructed density map showing connectivity at both the cytoplasmic and extracellular surfaces of the membrane are shown in Fig. 8. Before processing, the helices rapidly tapered to points near the end and were only ~35 Å in length. After refinement, the molecule was up to 45-Å thick and the helices no longer tapered and extended to both membrane surfaces. The transmembrane electron density profile, derived by Blaurock and King (1977) and by Stroud and Agard (1979), with data to 14-Å resolution, shows peak densities for the lipid head groups separated by 41.7 Å. The total thickness of the 14-Å resolution profile is 66-Å full width, or 55-Å full width at half-height. Perhaps the best estimate of membrane thickness is the minimum multilayer spacing of membranes dried in vacuo, found to be 49 Å (Blaurock and Stoeckenius, 1971).

Four linking densities are seen on the extracellular surface, which we interpret to be the N-terminus and the three intrahelical linking sequences. Only one linker is seen on the cytoplasmic surface, two being left undefined. This could be because the linking sequences on the cytoplasmic side (Fig. 9) are more highly charged (13–16 charged residues) and so possibly more disordered than those on the extracellular surface (2-4 charged residues).

Since the N-terminus lies on the extracellular side (Fig. 8, bottom), the visible linkers between helices 2-3 (extra-
**Figure 9** Sequence of bacteriorhodopsin obtained by Gerber et al. (1979) and Ovchinnikov et al. (1979) arranged in seven transmembrane α-helical segments. Where the sequences differ, the sequence of Gerber et al. (1979) is above that of Ovchinnikov et al. (1979). This arrangement is consistent with chemical labeling and protease susceptibility experiments. Residues are arranged as unfolded helices with 3.5 residues/turn and a helical pitch of 5.4 Å. The horizontal lines represent the membrane surfaces separated by 45 Å. Side chains that would be charged at neutral pH are indicated. Proton donors (●) and proton acceptors (○) in helical segments are shown, and the shaded regions may represent intramolecular contact surfaces. Though independently derived this arrangement is similar to that proposed by Engelman et al. (1980).

Cellular), 3-4 (cytoplasmic), 4-5 (extracellular) imply that none of these includes the N-terminus helix, and that either 2 or 5 could be the C-terminal helix. The extracellular density connecting helices 1, 7, and 6 can only be explained if either helix 1 or 6 is the N-terminal helix, whereupon helices 2, 5, or 7 must be the second, or B helix. Since no cytoplasmic linker is seen for any of these helices, the order of the first three helices can be 1-7-6 (Fig. 10a, b), 6-7-1 (Fig. 10 c), 6-5-4 (Fig. 10 d), or 1-2-3 (Fig. 10 c). In these most probable models there is density for the N-terminus, and for four of the six linkers. There is some residual density on the cytoplasmic side, attached to helix 5, which could accomodate the first few residues of the C-terminal chain of ~24 residues in Fig. 10 b, c.

There are two additional arms of density, one on the cytoplasmic side of helix 4 stretching ~10 Å to reach the cytoplasmic surface, and another attached to the center of helix 1 extending 10 Å toward the extracellular surface. These could be due to tightly bound lipid chains: the 3.7 Å electron-microscope projection structure derived in our laboratory did also show peaks in the lipid region (Hayward and Stroud, 1981).

The connectivity information provided by refinement alone reduces the possible number of ways of assigning the amino acid sequence (Ovchinnikov et al., 1979; Gerber et al., 1979) to the structure from 5,040 (7!) to the five most probable models (Fig. 10). In our laboratory (Katre et al., 1981) and elsewhere (Bayley et al., 1981; Lemke and Oesterhelt, 1981) it was shown that the retinyl moiety is primarily linked to lysine 216, on helix G, by reduction of photocycling purple membranes: We found evidence in favor of possible attachment of some retinyl also to lysine 40/41 (helix B) in the dark. Three of the models (Fig. 10a, d, e) place helix G (lysine 216) close enough to the...
projected location for retinal found by King et al. (1980) when they compared deuterated and hydrogenated retinal in their neutron diffraction studies. Two of these (Fig. 10 a, e) leave helix B suitably close to this site also. The helices in the model are labeled as they were by Engelman et al. (1980) in their discussion of more favored models. Their criteria were based on the density of regions in the density maps, the proximity of charged residues, and the length of supposed linking regions between helices. Their two most favored models are consistent with the arrangement of visible linking regions and the chromophore location (Fig. 10 a, e). Another model (Fig. 10 b) is number 30 (out of their most probable 35), whereas the third and fourth models (Fig. 10 c, d) are not among the most favored by their criteria, having an arrangement of chains that did not assign the least dense of the chosen helical sequences to helices 1 and 5, and because of a somewhat less favorable arrangement of charged residues. Since the exact termination of each helix is not known, and the apparent densities in the map may be affected by other factors, the exclusion of these as possible models is not yet fully justified.

Work in progress on the localization of several linker-directed heavy metal labels in our laboratory (Katre et al. 1981) and elsewhere (R. Henderson, private communica-

tion; and Dumont et al., 1981) should decide between the five most probable connection schemes.

Evaluation Test

Seven-helical models of 227 alanine residues were computer generated to match the number of residues in helical and linker regions shown in Fig. 9 and the density shown in Fig. 8 derived for bacteriorhodopsin; 21 residues were deleted from the actual C-terminal sequence of bacteriorhodopsin, since the effect of their removal on the diffraction pattern is small, probably owing to disorder (B. A. Wallace, R. Henderson, private communications). The structure built to model A is shown in Figs. 11 a and 12 a. The true density map computed with all data to 7-Å resolution in-plane, 10-Å perpendicular resolution, is shown in Figs. 11 b, and 12 b. Data were removed to produce an observed data set $F''(s)$ including only the terms in the bacteriorhodopsin data set. The density map computed from this set is shown in Figs. 11 c, and 12 c. Finally, refinement was carried out exactly as for bacteriorhodopsin to give the map represented in Figs. 11d and 12 d. The character of the flawed density map (Figs. 11 c and 12 c) closely resembles the density map for bacteriorhodopsin of Henderson and Unwin (1975). The helices are shorter and tapered, and the linker regions are no longer visible. As seen in Figs. 11 d, and 12 d the restored density map shows the three linkers on the extracellular surface, and two of the three on the cytoplasmic side. The density in the linking regions was always somewhat lower than in the true density maps (Fig. 11 b and 12 b) and no density was introduced by the procedure where linkers were not put into the starting model, both factors being consistent with the stability of the retrieval algorithm.

The one linker not retrieved (5 to 6 on the cytoplasmic side) is one of the two not retrieved in the case with bacteriorhodopsin. Thus, even when a (short) 5-6 linking sequence was present in the starting model, this linker density was for some reason more resistant to retrieval, and in general the linker densities on the extracellular side were slightly more faithfully restored, perhaps because they have higher density contours in the true density map, simply owing to disposition in the unit cell, in the limited resolution map. Density for six residues of the C-terminal sequence, also on the cytoplasmic side were also poorly retrieved, presumably for similar reasons. Nevertheless, these tests showed remarkably similar behavior to the bacteriorhodopsin analysis, showing weaker, or absent density at precisely the same locations, even when linking, and part of the C-terminal sequences were present in the starting model at these locations.

To demonstrate the amount and amplitude of information predicted by our algorithm, the values of $F_{h}(Z)$ for the (4, 2) and for the (3, 0) reflections after refinement are graphed in Fig. 13. Both of these reflections are graphed by Henderson and Unwin (1975) along with their observed values. This figure shows good restoration, and

AGARD AND STROUD  Linking Regions Between Helices in Bacteriorhodopsin Revealed 597
FIGURE 11  (a) α-Carbon positions for the trial model, which was constructed from 227 alanine residues. The model is shown left-handed, and sheets in b, c, and d are stacked in reverse order, for direct comparison with Figs. 10 and 12. This view represents the cytoplasmic side (b) The "true" density map computed at 7 Å in-plane, 10 Å perpendicular to plane resolution. A 10-Å thick section of sheets is shown. (c) Density map computed from error-flawed data with the missing region removed exactly as with bacteriorhodopsin. (d) Restored density map after refinement to fill in the missing data region.
FIGURE 12  (a–d) The model and map shown in Fig. 11 is viewed from the extracellular side, in right-handed axes, and shows a 10-Å thick density region on this side.
FIGURE 13 Predicted (refined) values for $|F(4,2)|$ and $|F(3,0)|$. These curves can be compared with original data points and the smooth curve plots presented by Henderson and Unwin (1975). Arrows mark the limits of observed data.

also indicates the rather small contributions obtained in the extrapolated region, i.e., the amounts added to the map by the extrapolated data are a relatively small modulation of the starting structure and probably correspond to the genuinely weaker intensities in that region (Henderson and Unwin, 1975).

DISCUSSION

The refinement method was designed specifically to remove the distortion produced in electron-microscopically determined structures as a result of the limited data set obtainable. The mathematical basis is in a sense derivative from our constrained-volume Fourier model refinement approach to phasing of continuous x-ray scattering from membranes (Stroud and Agard, 1979; Ross et al., 1977). In a more general sense, it is clear that this method is applicable to resolution extension (Agard et al., 1981), refinement of phases where intensities are known (Agard and Stroud, 1981), or determination of some subset of the phases in the Fourier transform of any structure or image, with significant gain in each case. The use of constraints which reflect a priori information about density in atomic structure, positivity, and atomicity, was fundamental to the development of direct methods in crystallography (Karle and Hauptman, 1950; Sayre, 1952). Analogous procedures (Gerchberg and Saxton, 1972) and use of density modification schemes (e.g., Hoppe and Gassman, 1968) have also been proposed for electron microscopy (Hoppe, 1970). (For an information-theoretic approach to structure analysis, see Gassmann, 1977.) The use of volume and density constraints form a particularly powerful combination.

Theoretically, only the volume/thickness restriction is required to extrapolate into the missing region from the observed $F_0(s)$. Methods, such as those proposed by Wolter (1961) or Harris (1964), that rely strictly on spatial limitations are extremely sensitive to data errors. The incorporation of the positivity constraint dramatically increases the stability of the algorithm as demonstrated by the remarkable noise immunity.

An additional piece of information that could be incorporated would be the meridional scattering data obtained from small-angle solution x-ray experiments [$F(s)$ for $s$ parallel to $z$]. When corrected for the absence of a constant solvent density, absent in electron microscopy, these scattering data can provide a set of experimental data along the $z$-axis. This direction is otherwise the most difficult to predict, since no values of $F(s)$ along the $z$-axis are obtained from electron microscopy. However, the meridional x-ray data reflect the one-dimensionally well ordered lipid bilayer profile, whereas the electron diffraction peaks sample the truly repetitive in-plane order or protein. Disordered lipids contribute incoherent and small angle scattering; thus, it is incorrect to incorporate the meridional x-ray data in refinement of the phases for electron diffraction amplitudes. The phased meridional x-ray $F(s)$ could be added later to increase density at both sides of the membrane though this was not done: The evaluation test shows that density in the retrieved region is somewhat weaker than in the true resolution limited density map, but addition of density to the outer regions of the bilayer everywhere from terms $F(s)$ where $s$ is perpendicular to the bilayer would lend false confidence since it arises from the unordered in-plane, as well as the ordered components.

The application to bacteriorhodopsin produces a map...
that is significantly different from the previous starting model of Henderson and Unwin (1975). It is also of note that the sequence alignment, as presented in Fig. 9, identifies several patches of amino acid side chains, or main chain carbonyls one turn behind proline residues, which include all the protein donors and acceptors. Particularly striking is the fact that the fifth and sixth helices contain patches that extend the entire length of the helix, and are tilted toward the left, consistent with that face always being on the inside of the molecule and in an arrangement of left-handed, tilted helices. In the density model, helices number 1, 2, 3, and 4 in the map are the most tilted (Henderson and Unwin, 1975), and three of the linker-derived models (Fig. 10 a, b, c) place the fifth and sixth helices at sites 3 and 4, which makes them among the most tilted.

The 6-7 linker is the longest and the most dense observed in the reconstructed model (Fig. 8). In both Figs. 10 a and b this corresponds to the linker between second and third helices in the sequence (Fig. 9). On the basis of protease susceptibility, this linker is most probably the longest linker sequence.

There are three regions where density connects helices inside the molecule: (a) 16 Å from the cytoplasmic surface between helices 1 and 7, (b) 18 Å from the cytoplasmic surface between helices 2 and 3, and (c) 16 Å from the extracellular surface between helices 4 and 5. It is conceivable that these could represent close packing of aromatic side chains, such as might be possible between the extracellular ends of the first and second, or second and third helices, or the close clustering of ion pairs possible between the sixth and seventh helices.

Overall, models 10 a, and e are consistent with three completely independent sets of criteria; the linker results presented here, the retinyl location in the sequence and structure, and the criteria of Engelman et al. (1980). We therefore consider them to be the most likely structures.

We are grateful to Richard Henderson and Nigel Unwin who kindly sent us their 1975 data set for bacteriorhodopsin.

This work was supported by the National Institutes of Health (grants GM27057-01 and GM24485-04) and by the National Science Foundation (grant PCM80-21433).

Received for publication 11 August 1980 and in revised form 8 August 1981.

REFERENCES


AGARD AND STRoud Linking Regions Between Helices in Bacteriorhodopsin Revealed 601


