Lawrence Berkeley National Laboratory
Recent Work

Title
ELECTRON DIFFRACTION OF PROTEIN CRYSTALS

Permalink
https://escholarship.org/uc/item/24f1k6n1

Author
Ho, M.

Publication Date
1984
ELECTRON DIFFRACTION OF PROTEIN CRYSTALS

M. Ho
(Ph.D. Thesis)

January 1984

TWO-WEEK LOAN COPY

This is a Library Circulating Copy which may be borrowed for two weeks. For a personal retention copy, call Tech. Info. Division, Ext. 6782.
DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.
ELECTRON DIFFRACTION OF PROTEIN CRYSTALS

Minghsiu Ho
(Ph.D. Thesis)

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

January 1984

This work was supported by the U.S. Department of Energy under Contract Number DE-AC03-76SF00098.
Abstract

Electron Diffraction of Protein Crystals

by

Minghsiu Ho

This research work concerns the use of electron diffraction in the structural study of biological macromolecule crystals. The first part is a theoretical investigation on the dynamical diffraction effects due to multiple-scattering of electrons in the specimen. Computer simulations are performed for both the dynamical Cowley-Moodie formula and the kinematic first Born approximation. The difference between the results from these two formulations are evaluated as a function of resolution, thickness and electron accelerating voltage. An "apparent temperature factor" associated with the dynamical diffraction effect is first reported for crystals of large unit cell and large number of low atomic weight atoms. The domain of validity for the use of the kinematic interpretation of electron diffraction data is discussed in terms of (i) crystallographic R values (ii) interpretability of difference Patterson function and (iii)
results of multiple isomorphio replacement phasing method.

The second part is experimental work in which an attempt is made to locate the C-terminal region of bacteriorhodopsin in the difference Fourier map calculated for native and papain treated purple membrane. A frozen-hydration specimen method is used in preparing both native and papain treated membranes.
Acknowledgements

I am grateful to Professor Robert M. Glaeser for his guidance of my graduate studies. His presentation of the development of electron crystallography and structural biology has provided an invaluable context for my studies. I deeply appreciate the support Professor Glaeser has kindly provided during this period. I would also like to thank the members of his group for their help. In particular, I wish to thank Dr. B. K. Jap for helping me to begin the theoretical calculation of dynamical diffraction and for his continued assistance and insightful suggestions on all aspects of this thesis research.
# Table of Contents

Acknowledgements .......................................................... i
Table of Contents .......................................................... ii

Chapter 1

Introduction and Born approximation .............................. 1

Chapter 2

Cowley-Moodie multislice formulation ....................... 6

Chapter 3

Dynamical calculation of electron diffraction ............ 11

Chapter 4

Electron diffraction of purple membrane .................... 46

Chapter 5

Conclusion and summary ................................................ 88
References ................................................................. 91
Electron diffraction was discovered in 1927, 15 years after the discovery of X-ray diffraction. Today the structures of more than 200 biological macromolecules including proteins, t-RNAs, oligonucleotides and even viruses have been determined to atomic resolution through X-ray crystallography (see Protein Data Bank, Brookhaven National Laboratory). On the other hand electron diffraction has not been fully employed in structural studies. There are several reasons for this. Keeping biological specimens hydrated in vacuum has been a problem, inaccessibility to the diffraction data in the vicinity of the $S_z$ axis leads to a hollow cone in Fourier space, and interpretation of diffraction data through the use of the first Born approximation (which is the foundation of X-ray crystallography) is not valid at greater thickness. The last problem is a fundamental one and is the main concern of this research. Specimen hydration has been achieved either through the method of frozen hydration (Taylor and Glaeser 1976) or the method of glucose embedment (Unwin and Henderson 1975). Several attempts have been tried to compensate for the absence of data in the missing cone, and their success remains to be seen.
In spite of the above potential difficulties, the temptation to use electron diffraction in structural studies is irresistible. The interaction between electrons and condensed material is about 100,000 stronger than that of X-rays. This makes structural analysis possible even when the available specimen is a tiny one. A practical specimen for X-ray diffraction would be at least of the order 0.1 µg while for electron a crystal of the dimension 1µ x 1µ x 100Å = (10^-14 g) is perfect, implying that sheet or lath-like two dimensional microcrystals can be studied by electron diffraction. Structures of some clay minerals such as celadonite and kaolinite which are considerably complicate for X-ray analysis have been successfully determined by electron diffraction (Vainshtein 1964 p. 348-350). Also the Z (atomic number)-dependence of scattering factor is greatly decreased for electrons compared to X-rays, this means that atoms of low atomic number such as hydrogen are more "visible" for electrons than X-rays. This facilitates the location of hydrogen atoms in earlier stage of structural analysis (Quon 1971). Or as a complement to X-ray crystallography which has been used to locate heavier atoms, electron diffraction can be employed to locate hydrogen positions (Vainshtein 1964 p. 356-370). Still another advantage is the high efficiency in data collection compared to that of X-ray. The wavelength of 100 keV electron is 0.037Å, implying a large Ewald sphere radius compared to a typical X-ray wavelength 1.5Å. In fact, one can assume a flat Ewald
sphere without introducing serious error at sufficient high resolution, if dynamical diffraction effects are not serious. A thorough investigation on the interaction between electrons and biological specimen is therefore necessary to justify the use of electron crystallography.

The theory governing elastic scattering of wave-particles by a potential field is briefly reviewed here. The equation of free motion of a particle is

\[ \Delta \psi^{(0)}(0) + k_0^2 \psi^{(0)} = 0 \]

where \( k_0 = \frac{2\pi}{\lambda} \) is the wave number for X-ray and \( k_0 = \frac{2\pi}{\hbar} \int 2mE = \frac{2W_0}{\hbar} \) for electrons in non-relativistic case. The equation has solutions \( \psi^{(0)}(r) = e^{-ik_0r} \), which represent incident plane waves and are usually the experimental cases for both X-rays and electrons. The scattered wave \( \psi(r) \) is expressed by the so called Born series.

\[
\psi(r) = \psi^{(0)}(r) + \psi^{(1)}(r) + \psi^{(2)}(r) + \cdots
\]

The equation for the first term in the Born series is

\[ \Delta \psi^{(1)}(r) + k^2 \psi^{(1)}(r) = \frac{8\pi^2mU(r)}{\hbar^2} \psi^{(0)}(r) \]

where \( U(r) \) is the scattering field potential. The solution for \( \psi^{(1)}(r) \) at large distance from the scattering field is

\[
\psi^{(1)}(r) = \frac{1}{r} (e^{ik_0r}) \left( -\frac{2\pi m}{\hbar^2} \int U(r)e^{-ixq}dv \right) \text{ where } q = k-k_0.
\]

In other words, the scattered wave measured at a large distance from the field is described by an outgoing spherical wave whose value at a particular scattered angle is proportional to the Fourier transformation of the scattering field evaluated at the angle.
The success of X-ray crystallography depends entirely on the validity of this approximation, namely the scattered waves are proportional to the Fourier components of the scattering field, which in this case, is the electron density distribution function \( \rho(r) \).

For 100 keV electrons, this approximation is not always valid. Glauber and Schomaker (1952, 1953) reported that the scattering of electrons by gaseous uranium hexafluoride can not be accounted for by the first Born approximation. Two distinct interatomic distances were shown in the analysis of the electron scattering data while data from infrared spectra and dipole moment measurements indicated a symmetrical structure giving a unique U-F bond length.

The transition from the kinematic (1st Born approximation) to dynamical range of single atom scattering for fast electrons can be estimated from the phase shift that results from the field. One of the criteria used is 

\[
\lambda |F_{h,k,l}| T = 1
\]

where \( \lambda \) is the wavelength of electrons, and \( F_{h,k,l} \) is the normalized Fourier transform of the field evaluated at \( h,k,l \) and \( T \) is the thickness of the scattering field. For mean value of \( h,k,l \) (e.g. \( \sin \frac{\theta}{\lambda} = 0.3\text{Å}^{-1} \)), the estimation of \( T \) for gold is \( 50\text{Å} \) and \( 100\text{Å} \) for silver (Vainshtein 1964 p. 151). Such estimates may, however, be too "optimistic" (too thick) by as much as a factor of 5 (i.e. realistic values would be 10 Å for gold and 20 Å for silver). A successful use of electron crystallography in
structural study involves the Fourier synthesis from multiple beams resulting from diffraction of the macromolecular crystal. This crude estimation for the domain of validity of kinematic theory is not enough.

The purpose of this research is to use some quantitative criteria to set a firm foundation for the application of kinematic theory to either electron diffraction pattern based or image based approach to structural study of biological macromolecule.

Besides the theoretical investigation, application of electron diffraction to the structural studies of bacteriorhodopsin is presented, in which an attempt is made to locate the C-terminal segment of this membrane protein in the two dimensional projection. The method of difference Fourier synthesis is used between native and papain treated purple membrane, prepared in the frozen hydrated state.
Chapter 2 Cowley-Moodie multislice formulation

The dynamical effect due to multiple scattering of high energy electrons (of the order 100 KeV) has been well investigated in the field of material science. However for biological specimens which are characterized by large unit cell and large number of low atomic weight atoms, the dynamical effect has not been studied thoroughly. In this chapter, a description is given of the approach used to evaluate the dynamical effect in protein crystals.

The calculation starts with the coordinates of the atoms in a protein crystal and its heavy atom derivatives, whose structure have been determined by X-ray crystallography. Computer simulations of electron diffraction of this model protein and its heavy atom derivatives are made through the "exact" dynamical formula (to be discussed in detail later) and also through the kinematic formula.

The residual difference between the results of calculation from the two formulas are evaluated as a function of resolution, thickness and electron accelerating voltages. Furthermore dynamical intensities are used to obtain the difference Patterson functions. The results are then compared with the "control" which is the self-correlation function of heavy atoms. Phase determination through multiple isomorphous replacement is carried out. The phases so determined are then compared with the original ones to see its potential applicability.
The protein crystal chosen in the calculation is Cytochrome b5 whose structure has been determined by Mathews et al (1972). The coordinates of the atoms and lattice constants are taken from the protein data bank of Brookhaven National Laboratory. This protein is chosen as a test object because of its relatively short c-axis repeat distance ~30 Å, and its convenient space group (P2₁2₁2₁) whose 2-D projection space group is Pgg; these factors facilitate the computation of diffracted waves, without, however, any sacrifice in generality regarding investigation of the dynamical effect. There are 4 asymmetric units in a unit cell, and a=64.54Å b=46.04Å and c=29.91Å. Each asymmetric unit has 469 carbon atoms, 140 oxygen atoms, 123 nitrogen atoms and 1 iron atom. Two mersaryl derivatives (each has 2 binding sites per asymmetric unit) and 1 uranyl derivative (with 6 binding sites per asymmetric unit) were used in the x-ray crystallographic study.

In order to simulate the diffraction of electrons by protein crystals, the Cowley-Moodie multislice approximation (Cowley and Moodie 1957) was used to calculate the diffracted waves. This approximation is expressed mathematically by

\[
\psi_n(x) = [\psi_{n-1}(x) \exp \left( \int_{z_{n-1}}^{z_n} eV(x, z) dz \right)] \ast \frac{i}{\hbar \Delta z_n} \exp \left( -\frac{i \pi x z}{\hbar \Delta z_n} \right)
\]

where \( \psi_n(x) \) is the wave function entering the nth slice and
\( \Delta z_n = z_n - z_{n-1} \). This formula can be interpreted in the following way: the transmission of electrons through a sample is approximated by transmission through successive slices of two dimensional planes, each separated by \( \Delta z \). Transmission through each slice is approximated by the "eikonal approximation", followed by Fresnel diffraction in vacuum. This new wave function then serves as the incident wave for the next slice, and so on. The eikonal approximation has been derived by Schiff (1956). This eikonal-type approximation is obtained by summing the infinite Born series, and each Born term is restricted to small scattering angle such that the stationary phase approximation is valid. In crystallographic terms, this is equivalent to a flat Ewald sphere. Therefore the eikonal or phase object approximation has a built-in assumption of zero wave length. By convoluting terms representing Fresnel diffraction, curvature of the Ewald sphere is taken into account in an approximate sense.

The Cowley-Moodie multislice approximation has been rederived by Jap and Glaeser (1978) through the use of the Feynman path-integral formulation where one sees a clear geometrical and physical picture regarding the "path" and "portion" of electrons passing through a specimen. In the impulse limit of this formula, that is \( \Delta z \to 0 \) and \( N \to \infty \) such that \( N \Delta z = T \), where \( T \) is the thickness of the specimen, this formulation becomes the Schrodinger equation for forward scattering (Goodman and Moodie 1974). The limitation to
only forward scattering is due to the fact that in using the Fresnel propagator, one uses a paraboloidal wavefront to approximate the spherical wavefront.

An immediately relevant question raised is what effect is there, in terms of evaluation of dynamical effects, by using a finite instead of infinitesimal value for $\Delta z$. Since the phase object approximation overestimates the dynamical effects, and Fresnel propagator reduces the dynamical effects by spreading the wavefront, therefore, for an identical total thickness of $T (=n \Delta z)$, if the chosen $\Delta z$ in the Cowley-Moodie formula is large, it tends to overestimate the dynamical effects. The calculation in this research takes c-axis length 29.91Å as $\Delta z$. The results are then an estimation of the "upperbound" of the dynamical effect. The Cowley-Moodie multislice approximation has been used extensively in material science, and the results of calculations match well with the experimental data (Lynch, Moodie and O'keefe 1975).

Atomic scattering factors used in the calculation of the projected potential for Cytochrome b$_5$ are taken from values of Doyle and Turner (1968). Atomic form factors for electrons, $f(s)$, are given in analytical form

$$f(s) = \sum_{i=1}^{4} a_i \exp(-b_i s^2)$$

where $s = \sin \theta / \lambda$ in Å$^{-1}$ and all the $a$'s and $b$'s are parameters calculated by the relativistic Hartree-Fock method and are tabulated in the paper. Only non-hydrogen atoms of cytochrome b$_5$ were included in the
calculation.

Computer programs are developed to calculate (1) the Fourier transform of the projected potential (2) Cowley-Moodie multislice approximation (3) crystallographic R values between Cowley-Moodie waves and kinematic waves (R values defined in next chapter). (4) Difference Patterson function and map based on Cowley-Moodie intensities. (5) Phasing through multiple isomorphous replacement in zonal plane of pgg 2-D space group, the projection of P2₁2₁2₁. (6) Least square fitting program used in Wilson plot is a package of Hewlett-Packard HP41c calculator. Programs for (1), (2), (3) were developed by Dr. B.K. Jap and software for contour plot map is a facility of BKY system at LBL. The computer used is CDC 7600.
Chapter 3 Dynamical calculation of electron diffraction

The first step in a series of calculations is to obtain the Fourier coefficients of the electric potential of crystalline cytochrome b5 in 2-D projection. All atoms are assumed to be stationary, that is the temperature factors are set to 0. Fourier coefficients are calculated to a resolution of 0.5Å for both native and heavy atom derivative structures. The array number used to register the coefficients is \((2\times64 + 1) \times (2\times46 + 1) = 11997\).

Out of this array only 9333 are nonvanishing as \(F(s(h,k)>2.0\frac{Å}{Å})\) are set to 0. The projected potential in the X-Y plane was calculated by taking the inverse Fourier transform of the \((H,K,0)\) structure factors. The 2-dimensional array of 129 x 93 elements are then floated on a larger array of dimension 245 x 175, which is used in the subsequent recursive Cowley-Moodie multislice formula. Since dynamical scattering may produce reflections that extend beyond the cutoff spatial frequency for the Fourier coefficients of electric potential function, a larger array is necessary to include such reflections thereby reducing the error in calculations and guaranteeing that unitarity condition (>99.7%) is satisfied up to a thickness of 600Å at all situations. In each recursive calculation of the Cowley-Moodie formula, beams with intensity less than \(10^{-8}\) are set to zero. The last measure is to eliminate residual computation error if any. The direction of the incident electron beam is assumed to be parallel to the crystal c-
axis. Three different accelerating voltages, namely 100, 500 and 1000 keV, are calculated.

The kinematic calculation was done by convoluting the structure factors with the crystal-thickness function and sampling the resulting function on the surface of the Ewald sphere. Figure 3.1 gives two examples of how diffraction intensities vary with the thickness of the crystal in the kinematic and the dynamical case, at low and high resolution. At a thickness of ~150-200 Å, the deviation is small, as it should be, but at a thickness of ~300 Å, there appears to be a significant difference. A "general comparison" between the kinematic and dynamical results can be inferred by evaluating the residual between $|F_C|$ , the Cowley-Moodie amplitude, and $|F_k|$ , the kinematic amplitude, according the following expression:

$$R(s_{\text{max}}) = \frac{s_{\text{max}} \sum_{h,k} |F_C(h,k) - \mu F_k(h,k)|^2}{s_{\text{max}} \sum_{h,k} |F_C(h,k)|^2}$$

where

$$\mu = \left( \frac{s_{\text{max}} \sum_{h,k} |F_C(h,k)|^2}{s_{\text{max}} \sum_{h,k} |F_k(h,k)|^2} \right)^{1/2}$$
Figure 3.1 The diffracted beam intensities for the (2,0,0) and (1,21,0) reflection are plotted as function of crystal thickness. The solid lines are the "exact" Cowley-Moodie dynamical intensities and dash lines are the kinematic intensities.
Cytochrome b$_5$
Voltage = 100 keV
- - - Kinematic approximation
--- Multislice dynamical

(2,0,0) Reflection

(1,21,0) Reflection

XBL772-3106
The data set in both cases are scaled such that the total diffracted intensities were the same at a given thickness. $s_{\text{max}}$ is the maximum spatial frequency. Figure 3.2 shows two representative examples of how $R(s)$ increases with increasing thickness for $s_{\text{max}} = \frac{1}{1.4}\text{Å}^{-1}$ and $s_{\text{max}} = \frac{1}{8.6}\text{Å}^{-1}$.

A "global" inference can be made from the above figure to define the "domain of validity" of the kinematic approximation for any numerical choice of tolerance of $R$ value. As shown in figure 3.3, once a value of $R$ is chosen, one can find a region of maximum resolution, at a particular thickness within which the kinematic approximation is valid. Thus for example, for the case of 100 keV electrons, the kinematic approximation is valid to a thickness of 250Å and a resolution of 2.5Å for a tolerance value of $R=0.05$, and a higher $R$ value permits the use of the kinematic approximation to a larger thickness. Cases for 500 keV and 1000 keV are given in figure 3.4 and 3.5. As expected, the domain of validity increases as higher accelerating voltage is used.

Simulation of phasing through multiple isomorphomorphic replacement (MIR):

The first step in MIR phasing is to determine the location of the heavy atom(s) from the difference Patterson map. The first work then is to see if an interpretable difference Patterson can be obtained even when dynamical intensities are used.
Comparison of Kinematic to Cowley-Moodie Multislice "Structure Factors". Cytochrome b$_5$ As The Test Object.

Figure 3.2 The reliability factors "R" (see text for definition) between kinematic and Cowley-Moodie Multislice structure factors are plotted for two resolution limits as a function of thickness.
Figure 3.3 The approximate boundary lines at various values of reliability factor "R" in terms of crystal thickness and resolution for the kinematic approximation. The electron accelerating voltage is indicated on the graph.
Cytochrome b$_5$
Kinematic approximation
Voltage = 100keV

Resolution (Å)

Spatial frequency (Å$^{-1}$)

Crystal thickness (Å)
Figure 3.4 The approximate boundary lines at various values of reliability factor "R" in terms of crystal thickness and resolution for the kinematic approximation. The electron accelerating voltage is indicated on the graph.
Voltage = 500 keV

Crystal thickness (Å)

Resolution (Å)

Spatial frequency (Å⁻¹)

R = 0.05

R = 0.10
Figure 3.5 The approximate boundary lines at various values of reliability factor "R" in terms of crystal thickness and resolution for the kinematic approximation. The electron accelerating voltage is indicated on the graph.
Voltage = 1 MeV

R = 0.05  R = 0.10

Resolution (Å)

Crystal thickness (Å)

Spatial frequency (Å⁻¹)
Using the same heavy atom derivative originally used by Matthew et al (1972) in their X-ray crystallography study, computer simulations are performed. Figure 3.6 shows the projected potential map of heavy atoms in the unit cell and their self-correlation function which serves as the control for the attempted difference Patterson map.

The difference Patterson is obtained by taking the inverse Fourier transform of \((|F_{p+h}| - |F_{p}|)^2\), that is \(F^{-1}(I_{p+h} - I_{p})^2\) where the I's are the dynamical intensities calculated through the Cowley-Moodie multislice approximation, and scaled as mentioned in detail below. The simulation of partial occupancy of heavy atoms in a protein crystal is extremely difficult to perform in dynamical calculations. We therefore approach this problem by simulating two limiting cases for 100 keV electrons. In the first case, the heavy atoms were assumed to have full occupancy at all binding sites. This should be the worst case in terms of concomitant dynamical effect. In the second case, the heavy atoms were replaced with heavy atoms of "effective" atomic number, thus for the mersalyl group, which contains a mercury atom with \(Z=80\), the effective heavy atoms have \(Z_i = 80f_i\), where \(f_i\) is the occupancy of the ith site. This approximation may underestimate the dynamical effect since it replaces a high atomic number element with one of lower atomic number. In this case the two Hg atoms are replaced by Molybdenum and Vanadium respectively.
Figure 3.6 An illustration showing the relationship between a function and its autocorrelation which is the functional relation between the heavy atom(s) binding sites and difference Patterson map.
a) The projected potential of the bound heavy atoms that were used to make an isomorphous derivative of Cyt b₅ crystal; the two-dimensional space group is pgg.

b) The autocorrelation of fig. a. This is the ideal difference Patterson map.
A total of 5 simulations have been done:

[1] voltage=100 keV, heavy atoms: 2 Hg, scaling of diffraction data: 

\[ \frac{I'(0)}{I(0)} = \frac{\Sigma f_2(0)}{\Sigma f_1(0)} \]

where ' denotes heavy atoms and I is the intensity, f is the atomic scattering factor for electron, and 0 denotes the undiffracted beam. Wilson plot is not used.

[2] voltage=100 keV, heavy atoms: 1 Mo, 1 V, scaling of diffraction data: same as [1]

[3] voltage=100 keV, heavy atoms: 1 Mo, 1 V. Wilson plot is used to scale diffraction data, "apparent temperature factors" corrected for all diffraction data.

[4] voltage=500 keV, the rest is same as [3].

[5] voltage=1000 keV, the rest is same as [3].

The results of case [1] and [2] are presented here. Figure 3.7 shows the difference Patterson maps for specimen thicknesses of 30, 150, and 300 Å. Full occupancy of heavy atoms is assumed. Figure 3.8 shows the difference Patterson maps for specimen thicknesses of 30, 150 and 300 Å. Partial occupancy of heavy atoms is accounted for by replacing 2 Hg with Mo and V. In all maps, the whole numerical range of the function is represented in the contour plots. In both cases, difference Patterson maps obtained with dynamical intensities can clearly be correctly interpreted for the
Figure 3.7 The difference Patterson maps for the full-occupied Hg derivative of cytochrome b 5. (a) control, (b) one, (c) five and (d) ten unit cells thick. The intensities used are calculated by the Cowley-Moodie multislice formulation.
Difference Patterson for Hg derivative of Cyt b_{5} control, one, five and ten unit cells thick.

Cowley-Moodie Dynamical
Full occupancy of heavy atoms is assumed
Figure 3.8 The difference Patterson maps for the partially-occupied Hg derivative of cytochrome b5. (a) control, (b) one, (c) five and (d) ten unit cells thick. The intensities used are calculated by the Cowley-Moodie multislice formulation.
Difference Patterson for Hg derivative of Cyt. b\textsubscript{15}
Control. one, five and ten unit cells thick
Cowley-Moodie Dynamical
thinnest possible specimen, which is just one unit cell thick. For specimen thickness of 150 Å, the difference Patterson would probably still be correctly interpreted. But by a thickness of 300 Å, the background level in the difference Patterson map has gotten to be unacceptably high, and the relative values of the Patterson peaks are greatly in error.

The fact that full occupancy and partial occupancy give similar results can probably be explained in this way: while full occupancy is favorable in effectively perturbing the diffraction data to give a high signal to noise ratio, the concomittant dynamical effect causes the data to deviate from the kinematic assumption. At these thickness of 30, 150 and 300 Å, we can not tell to what degree each factor contributes.

Dynamical effect introduces an "apparent temperature factor"

As mentioned earlier, all the atoms are assumed to be stationary, that is to say, no thermal disorder is introduced in the calculation. The intensities resulting from the Cowley-Moodie multislice formulation do, however, show an apparent temperature factor. The introduction of a temperature factor can be seen from the Wilson plots of figure 3.9 for accelerating voltage of 100, 500 and 1000 keV. The Wilson plot is commonly used by X-ray crystallographers for scaling together different sets of diffraction data. In the plot,
Figure 3.9 The Wilson plots for the dynamical electron diffraction intensities calculated by the Cowley-Moodie multislice formulation for cytochrome b 5 of various thickness and electron accelerating voltage, as indicated in each graph.
one takes the logarithm of the ratio of the average value of
the diffraction intensities within a circular zone to the
corresponding total atomic scattering factors squared, and
plot this logarithm of the ratio against the square of the
spatial frequency. For an ideal kinematic situation, in the
absence of thermal disorder, the curve should be horizontal,
as both the denominator and the nominator are, by defini­
tion, the differential cross section at a particular angle
of scattering. This is indeed the case for a specimen thick­
ness of 30 Å. However for a specimen thickness of 300 Å, the
slope of the Wilson plot is quite steep, corresponing to an
apparent temperature factor of 2.2 Å² for 100 keV electron,
and 2.5 Å² for 500 and 1000 keV electrons. These values of
temperature factor correspond to an r.m.s. atomic displace­
ment of 0.13 Å to 0.17 Å. The random fluctuation in the
average structure factor for different circular zones in
scattering angle are also more pronounced for greater speci­
men thickness. This kind of "temperature factor" has been
calculated up to 600 Å, and it seems to level off at 2.5 Å²
to 2.7 Å² for all three accelerating voltages. These
numbers are tabulated in table 3.1.

The intercept together with the slope in the Wilson
plot are then used to scale the dynamical diffraction data
for the native and heavy atom derivative. Difference
Patterson maps are then generated from these scaled, tem­
perature factor corrected diffraction data. Figure 3.10
shows the case of 100 keV. The left half of the figure is a
Table 3.1 The "apparent" temperature factors (-2B) determined by the Wilson plots based on the diffraction intensities calculated by the Cowley-Moodie multislice method for the crystal thickness and electron accelerating voltage indicated. The relation between B and r.m.s. atomic displacement u is \( B = 8 \times u^2 \). For \( 2B = 4 \) u=0.159, \( 2B = 5 \) u=0.178
<table>
<thead>
<tr>
<th>Thickness (x 29.91 Å)</th>
<th>100 kev</th>
<th>500 kev</th>
<th>1000 kev</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>2</td>
<td>-4.98</td>
<td>-0.09</td>
<td>0.22</td>
</tr>
<tr>
<td>3</td>
<td>-3.30</td>
<td>-0.97</td>
<td>-0.08</td>
</tr>
<tr>
<td>4</td>
<td>-4.52</td>
<td>-2.49</td>
<td>-0.52</td>
</tr>
<tr>
<td>5</td>
<td>-3.77</td>
<td>-5.04</td>
<td>-1.13</td>
</tr>
<tr>
<td>6</td>
<td>-4.33</td>
<td>-5.42</td>
<td>-1.96</td>
</tr>
<tr>
<td>7</td>
<td>-3.97</td>
<td>-4.61</td>
<td>-3.09</td>
</tr>
<tr>
<td>8</td>
<td>-4.29</td>
<td>-4.06</td>
<td>-4.44</td>
</tr>
<tr>
<td>9</td>
<td>-4.11</td>
<td>-4.24</td>
<td>-5.10</td>
</tr>
<tr>
<td>10</td>
<td>-4.33</td>
<td>-5.04</td>
<td>-4.99</td>
</tr>
<tr>
<td>11</td>
<td>-4.40</td>
<td>-5.10</td>
<td>-4.59</td>
</tr>
<tr>
<td>12</td>
<td>-4.26</td>
<td>-4.84</td>
<td>-4.21</td>
</tr>
<tr>
<td>13</td>
<td>-4.56</td>
<td>-4.67</td>
<td>-4.00</td>
</tr>
<tr>
<td>14</td>
<td>-4.35</td>
<td>-4.84</td>
<td>-4.02</td>
</tr>
<tr>
<td>15</td>
<td>-4.57</td>
<td>-5.19</td>
<td>-4.28</td>
</tr>
<tr>
<td>16</td>
<td>-4.49</td>
<td>-5.11</td>
<td>-4.66</td>
</tr>
<tr>
<td>17</td>
<td>-4.60</td>
<td>-4.90</td>
<td>-4.85</td>
</tr>
<tr>
<td>18</td>
<td>-4.55</td>
<td>-4.85</td>
<td>-4.86</td>
</tr>
<tr>
<td>19</td>
<td>-4.57</td>
<td>-5.03</td>
<td>-4.78</td>
</tr>
<tr>
<td>20</td>
<td>-4.62</td>
<td>-5.30</td>
<td>-4.69</td>
</tr>
<tr>
<td>21</td>
<td>-4.65</td>
<td>-5.22</td>
<td>-4.65</td>
</tr>
</tbody>
</table>
Figure 3.10 The difference Patterson maps for Hg derivative of cytochrome b 5 based on the intensities calculated by the Cowley-Moodie multislice formulation with (right) and without (left) correction for the "apparent" temperature factors. (a) control, (b) one, (c) five and (d) ten unit cell thickness.
Difference Patterson for Hg derivative of Cyt. b\textsubscript{5}.
Control, one, five and ten unit cells thick.
Cowley-Moodie Dynamical

XBL817-4038
reproduction of figure 3.8 presented here for comparison, the right half is the map generated from the same diffraction data with correction made for the effective temperature factor and scaled according to the Wilson plot. The correction for incurred temperature factor has sharpened the Patterson peaks, and the relative values at the peaks are improved. Figure 3.11 shows the difference Patterson maps for 500 and 1000 keV electron. The numerical range of the difference Patterson function and contour plots representing contour levels are tabulated in table 3.2.

The occurrence of a negative valued background in the difference Patterson maps in certain cases is unexpected, the background level in the contour plot is adjusted to avoid negative numbers, as the gray scale representing negative background smears the plot. The minimum value in the contour plot is \( \frac{1}{4} \) of the total numerical range from the Patterson function. The peaks shown in the plots are indeed significantly higher than the background.

The dynamical diffraction intensities for the native and heavy atom derivative protein crystal were then used to phase the \((h,k,0)\) reflections. Since this zonal projection is centro-symmetric, the phase choice is limited to 0 or \( \pi \). The phases of the native structure were assumed to be the same as the phases of the heavy atoms alone, whenever the intensity of the derivative is larger than that of the native. When the intensity of the derivative was less than
Figure 3.11 The difference Patterson maps for Hg derivative of cytochrome b 5 based on the intensities calculated by the Cowley-Moodie multislice formulation with correction for the "apparent" temperature factors. (a) control, (b) one, (c) five and (d) ten unit cell thickness. Electron accelerating voltage is indicated.
Difference Patterson for Hg derivative of Cyt. b₅.
Control, one, five and ten unit cells thick.
Cowley-Moodie Dynamical (500 keV)

Difference Patterson for Hg derivative of Cyt. b₅.
Control, one, five and ten unit cells thick.
Cowley-Moodie Dynamical (1000 keV)
<table>
<thead>
<tr>
<th>Voltage (x 29.91 A)</th>
<th>Thickness</th>
<th>Minimum</th>
<th>Increment</th>
<th>Maximum</th>
<th>Range Plotted</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 kev</td>
<td>1</td>
<td>0.0000</td>
<td>0.0004</td>
<td>0.0068</td>
<td>100%</td>
</tr>
<tr>
<td>100 kev</td>
<td>5</td>
<td>-0.0007</td>
<td>0.0007</td>
<td>0.0105</td>
<td>100%</td>
</tr>
<tr>
<td>100 kev</td>
<td>10</td>
<td>-0.002</td>
<td>0.001</td>
<td>0.017</td>
<td>100%</td>
</tr>
<tr>
<td>500 kev</td>
<td>1</td>
<td>0.0000</td>
<td>0.0004</td>
<td>0.0064</td>
<td>100%</td>
</tr>
<tr>
<td>500 kev</td>
<td>5</td>
<td>-0.009</td>
<td></td>
<td>0.042</td>
<td></td>
</tr>
<tr>
<td>500 kev</td>
<td>5</td>
<td>0.0014</td>
<td>0.0027</td>
<td>0.0419</td>
<td>80%*</td>
</tr>
<tr>
<td>500 kev</td>
<td>10</td>
<td>-0.015</td>
<td></td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>500 kev</td>
<td>10</td>
<td>0.006</td>
<td>0.0038</td>
<td>0.063</td>
<td>74%*</td>
</tr>
<tr>
<td>1000 kev</td>
<td>1</td>
<td>0.0000</td>
<td>0.0004</td>
<td>0.0064</td>
<td>100%</td>
</tr>
<tr>
<td>1000 kev</td>
<td>5</td>
<td>0.0000</td>
<td>0.0005</td>
<td>0.007</td>
<td>100%</td>
</tr>
<tr>
<td>1000 kev</td>
<td>10</td>
<td>-0.018</td>
<td></td>
<td>0.078</td>
<td></td>
</tr>
<tr>
<td>1000 kev</td>
<td>10</td>
<td>0.007</td>
<td>0.0046</td>
<td>0.076</td>
<td>74%*</td>
</tr>
</tbody>
</table>

*From maximum.

Table 3.2 The numerical range of values that occur in the calculated difference Patterson functions, and the range of values that were actually used along with the increment between contour levels in the plots of Figure 3.10 and Figure 3.11.
that of the native, then the phase of the native structure was taken to be \( \pi \) plus the phase of the heavy atom.

Structure factors in the \((h, k, 0)\) plane, for the heavy atoms, were assumed to be the same for all cases, and were calculated by the same program used to calculate the fourier coefficients of the protein crystal. Heavy atoms were assumed to be Mo and V as previously taken. Their positions were the same as those used in the X-ray work by Matthew et al. This is so assumed in order to separate the errors due solely to dynamical diffraction effects from those due to experimental inaccuracy in the interpretation of the difference Patterson function.

Strickly speaking, the vectorial relation \( F_{ph} = F_p + F_h \) where \( F_{ph} \), \( F_p \), \( F_h \) are the structural factors of heavy atom derivative, native structure and heavy atoms respectively, is valid in the kinematic case only. While using this kinematic relation, dynamical intensities for native and heavy atom derivative and a common kinematic structure factor for heavy atoms to obtain phases, we believe this phasing scheme suffices as a pilot study for the evaluation of the dynamical effect in electron protein crystallography.

Figure 3.12 shows the results. The upper 3 curves show the percentage of beams assigned wrong phases as a function of thickness, and the lower 2 curves represent the percentages of their associated intensities. A significant improvement is seen from 100 keV to 500 keV, but not much
Figure 3.12 The number of diffraction spots assigned the wrong phase (top 3 curves) and their associated total intensities (lower 2 curves) as a percentage of all diffracted beams for different crystal thickness. Phases in the centric projection of cytochrome b 5 were deduced from calculated dynamical intensities for the native structure and one heavy atom isomorphous derivative, and corrected for the "apparent" temperature factors.
improvement is seen from 500 to 1000 keV.

The results of this calculation are encouraging in terms of (1) crystallographic R values (2) interpretability of difference Patterson map and (3) the result of MIR phasing. It must be pointed out that Cytochrome b5 is a particularly favorable crystal for these theoretical simulations, as the low molecular weight (~10000) and two binding sites for Hg per asymmetric unit make it an ideal case for structural study. In this work only centric-zone diffraction data are calculated, non-centric data whose phase choice is not limited to 0 or π, have not been investigated.
Chapter 4 Electron diffraction of purple membrane

In this chapter, experimental work of electron diffraction of a membrane protein is presented. An attempt is made to see if one can use the method of difference Fourier synthesis to determine the structural difference between two membrane protein crystals, in which one crystal is chemically modified from the other. The specimen is purple membrane of bacteriorhodopsin of *Halobacterium halobium* and the papain treated purple membrane, where the C-terminal region is removed (Ochinickov 1979).

Bacteriorhodopsin is the only integral protein in the specialized purple membrane patches of *Halobacterium halobium*. It absorbs light and uses the photon energy to pump protons across the cell membrane, thereby generating an electro-chemical gradient from which energy is derived to power metabolism under anaerobic conditions (Stoeckienius 1979).

The purple membrane occurs in nature as a two-dimensional crystalline array suitable for structural study through electron diffraction and microscopy (Unwin and Henderson 1975, Henderson and Unwin 1975, Leifer and Henderson 1983, Hayward and Stroud 1981). In previous studies, a
three-dimensional map of resolution 7 Å parallel to the plane of the membrane and 14 Å perpendicular to the plane have been obtained. The two-dimensional projected map has been determined to a resolution of 3.7 Å. The three-dimensional map shows a structure of 7 α helices that span across the membrane. Based on this three-dimensional map of Unwin and Henderson and the amino acid sequence information of Khorana et al. (1979), various models have been proposed (Ovchinikov et al. 1979, Engelman et al. 1980, Engelman and Zaccai 1980, Katre et al. 1981). The two-dimensional projected map of bacteriorhodopsin at 3.7 Å resolution (Hayward and Stroud 1981), however shows some features that may suggest possible alternatives, notably the outer helices designated by Engelman et al. (1980) as helices 1, 2, 3 pinched off and gave a narrower structure which can not be matched by α helix. Moreover, the circular dichroism and infrared spectra of bacteriorhodopsin measured by Jap et al. (1983) suggests that there is substantial β sheet (~20 %) conformation and that α-helix content (~50 %) is sufficient only for 5 Helices. A 7-helix structure would require α content to be greater 70 %. The location of the C-terminal region in a two-dimensional in-plane projection would be a useful constraint to possible models.

Govindjee et al. (1982), using the dye, p-nitrophenol, as reporting molecule for measuring the proton concentration, and absorbance at 412 nm as measurement of bR M412 concentration, have determined the ratios [H+] : [M412] for
both native bR and for bR with the C-terminal removed. The result showed that removal of the C-terminal region of bR by trypsin or papain reduces the yield of light-induced proton by 50-70%. Based on these results, the authors concluded that the C-terminal region of bR plays an important role in light-driven proton release by the purple membrane.

Wallace and Henderson (1982), using the method of difference Fourier synthesis, attempted to locate the C-terminal region in the in-plane projection. In this work, intensity difference between glucose-embedded native and papain cleaved purple membrane were used together with the phases of glucose-embedded native purple membrane for the Fourier synthesis. The map generated, according to the authors, is not interpretable. The question left is then whether the C-terminal region is crystalline-ordered or not in glucose.

Renthal et al (1983), have presented experimental evidence that the C-terminal region is rigidly held at the membrane surface. The authors measured the depolarization of fluorescence (for theory of polarization of fluorescence see e.g. Cantor and Schimmel (1980)) of the labelling dye, dansylhydrazine, attached to the C-terminus. They found that the steady state polarization of this dansyl fluorescence on the tail was 0.24 at 25°C before cleavage, and after papain proteolysis, the polarization decreased to 0.1. The result convincingly implies that the C-terminal region
is rigidly linked to the remaining bulk of bR.

It must be pointed out that even if the C-terminal region is ordered, the density difference, (hence the contrast), is very small between glucose and protein, and the signal could be non-detectable. In view of the fact that the density of peptide is $\approx 1.3$ of glucose is $\geq 1.3$, and the density of ice is 0.917, it is worthwhile to tackle this problem again, by the frozen hydrated specimen method.

The method of "double-carbon film" is employed to prepare the frozen-hydrated specimen. This procedure is described here briefly. A thin layer of carbon film is coated onto pieces of mica $\approx 1$cm $\times$ 3cm. The mica is then dipped into distilled water, and the carbon film is floated onto the water surface. One then passes an electron microscope grid from inside the water through the carbon film. This first layer of carbon film serves as the supporting film. Sample is added on top of this film when this supporting film is dried. The grid with the sample on it is then dipped again into water and passed through the same floated carbon film, thereby the sample is sandwiched between double carbon films. One then waits until the grid gives a characteristic reflection color, then puts the grid into the specimen holder, already sitting in liquid nitrogen, and then transfers the frozen hydrated specimen to the cold stage of the electron microscope (Hayward and Glaeser 1980).
Processing of Diffraction Data

The bR used in this work is from strain JW-3 (kindly provided by Dr. H.J. Weber) which is a bR overproducing bacterium that forms a large spatially coherent patches of purple membrane especially suitable for electron diffraction.

Figure 4.1 shows the polyacrylimide gel electrophoresis pattern of native and papain-treated bR. The electron microscope used is the JEOL 100B with liquid nitrogen cooled cold stage mentioned earlier. The film used to record diffraction patterns is Kodak electron image film. Diffraction patterns of two native and two papain-treated purple membrane were processed.

The diffraction pattern on electron image film is digitized by scanning the films on the Perkin-Elmer densitometer installed at the Department of Astronomy. All scanning work was done in the same session with the same physical and electronic setups. In order to assure that the scanned regions of the plates are the desired region, that no misalignment or off-tracking has taken place, a colored Dicomed display of the scanned data is made as a control. An example is shown in figure 4.2. Such Dicomed displays were done for all diffraction data. The characteristic blackening curve of electron image film, that is the optical density response to a given dose of electron density, is fitted to a parabolic equation (Figure 4.3) \( y = 64.73 + 1027.08x - 348.04x^2 \)
Figure 4. PAGE (polyacrylimide gel electrophoresis) pattern of native and papain treated bacteriorhodopsin of purple membrane used in the electron diffraction experiment.
Figure 4.2 A color Dicom display of scan data file. Such display gives a clear picture of the area scanned. One can thereby tell if the scan has been properly done. (BBC 830-10836)
Figure 4.3 The curve fitting for the blackening characteristic of the electron image film used in recording the diffraction patterns. Data with O.D. greater than 700 are discarded.
where \( y \) is the numerical output of the Perkin-Elmer densitometer, a measurement of optical density, and \( x \) is the electron dose in number of electrons per micron\(^2\). Each data pixel is converted to its corresponding electron dose according to the above equation. The cut-off value is set at 700, that is diffraction spots containing pixel value(s) greater than 700 are discarded. Only a very few spots, near the origin, had to be rejected.

The overall scheme to obtain the integrated intensities for each diffraction spot is described here. A typical scan data file for the entire pattern of electron diffraction spots on one electron image film consists of 3720 (linear samples) \( \times \) 75 (lines per row of spots) \( \times \) 23 (rows of spots) pixels. Each pixel corresponds to an area of 5micron \( \times \) 5micron on the film. These parabolically-corrected pixel values are then summed in blocks of 4 \( \times \) 75, to give 23 stripes, each stripe corresponds to an area of 375 micron \( \times \) 1.86cm that covers a row of diffraction spots (see the Dicomed display of figure 4.2). Each data file now is a one-dimensional array of 930 elements and each element represents the number of electrons in a specified area of 20micron \( \times \) 375micron. The next step is to find a background curve for each integrated line. By subtracting these background values and integrating this difference, over a window wide enough to cover the diffraction spots at each lattice point, we finally obtain the required intensities for each diffraction spot.
Three different methods are used to determine the background level. In all three cases, the first step is to eliminate the diffraction spots from the scan data. This is done by taking a window with sufficient width to cover a diffraction spot and replacing the values within this window by linear-interpolated values. A smooth background is then evaluated for this "spot-free" curve.

[1] low-pass filtered background:

This is done by taking the Fourier transform of each line, setting the coefficients of higher harmonics to zero and taking an inverse Fourier transform to obtain a smooth background level. An example is given in figure 4.4. Such measure is meant to correct any systematic local or global variation due to electronic problems of the densitometer or other unknown factors.

[2] Lorentzian-Gaussian fitting:

In this approach, a function which is a sum of Lorentzian and Gaussian functions is attempted to fit the "spot-free" zig-zag curve. Mathematically, this is

\[ y = \frac{a_1 w}{2\pi \left[ \left( x - x_0 \right)^2 + \left( \frac{w}{2} \right)^2 \right]} + a_2 \exp \left( -a_3 (x - x_0)^2 \right) + a_4 \]

where \( y \) is the optical density, \( x \) is the one-dimensional array pixel number representing position. Six adjustable parameters \( a_1, w, x_0, a_2, a_3, a_4 \) are determined to obtain a best fit, and the resonance point, \( x_0 \) for both the Lorentzian and the Gaussian is set to be identical. \( a_4 \) is
the y offset. The physics of inelastic scattering which
gives the background is here mathematically described as a
damped oscillation with resonance point \( x_0 \), half-width \( w \),thereby accounted for by the Lorentzian function, the
Gaussian term is added to take care any residual. An example
is given in figure 4.5.

[3] Lorentzian-Gaussian with "free" segment:

In this approach, the fitted curve has the same functional form as in [2], except during the determination of
the parameters, a certain portion of the fitted curve is
allowed to take any numerical value. This can be done as
follows. In the fitting process, the residual between fitted
numbers, \( Y_i \), and original numbers, \( y_i \), that is \( (Y_i - y_i)^2 \) is
not included in the overall residual which is to be minim-
ized. An easy way to do this is to set these differences
\( Y_i - y_i \) to be 0 in the "free" segment. The "free" segment is
chosen so as to cover symmetrically the saturation region of
the plate. This will become clear by looking at the example
in figure 4.6.

Once the background levels are determined, the
integrated intensities are obtained by integrating the
difference between the signal and background at each lattice
point with a window sufficiently wide to cover the diffrac-
tion spot. Indexing of the spots is done manually, and care
is taken to ensure that all diffraction patterns are of the
same parity as conventionally assigned.
Figure 4.4 The line plot of a summed scan line and the associated background determined by the low-pass filter method.
Figure 4.5 The line plot of a summed scan line and the associated background determined by the curve fitting to a Lorentzian-Gaussian function.
Figure 4.6 The line plot of a summed scan line and the associated background determined by the curve fitting to a Lorentzian-Gaussian function with a "free" segment.
To evaluate the quality of the data, $R_{\text{sym}}$ values are calculated and plotted against the intensity for the four sets of data used, as shown in figure 4.7. $R_{\text{sym}}$ is defined as:

$$R_{\text{sym}} = \frac{\sum |I_i - \bar{I}_i|}{\sum I_i}$$

where $I_i$ indicates the intensities of a set of symmetry related reflections and $\bar{I}_i$ is the average of $I_i$.

Up to now, there are four sets of independent measurements, and they must be scaled to a common base before any algebraic operation can be applied. To do this, the difference Wilson plot is used, in which the value of

$$\ln \frac{\ln I_1(s^2)}{\ln I_2(s^2)} = \ln I_1(s^2) - \ln I_2(s^2)$$

is plotted against $s^2$, the square of the spatial frequency. The four sets of data are designated b1, b2, c1 and c2 for the two native bR and the two papain-cleaved bR. It is found that b2 has the lowest temperature factor and hence is chosen as a standard. Figure 4.8 shows the difference Wilson plots of b1 vs b2, c1 vs b2 and c2 vs b2 respectively. Once the slope and the Y-intercept are determined, a corresponding scaling factor is applied for each diffraction intensity of each data set. It should be noted that the line in the Wilson plot is a least square fitting weighted by

$$\frac{1}{\sqrt{R_1^2 + R_2^2}}$$

$R_{\text{sym}}$ values of the mentioned diffraction intensity in the two sets of data to be scaled together. That is given $Y_i(x_i)$, a linear equation $y = ax + b$ is to be fitted with
Figure 4.7 The $R_{sym}$ (see text for definition) plotted as a function of intensity in arbitrary scale. (a) b1, (b) b2, (c) c1 and (d) c2.
Figure 4.8 The difference Wilson plots for data set (a) $b_1$ vs $b_2$, (b) $c_1$ vs $b_2$ and (c) $c_2$ vs $b_2$. 
\[ \sigma = \sum (y_i - \bar{y_i})^2 \frac{1}{\sqrt{R_1^2 + R_2^2}} \]
to be minimized. Once a and b are determined, each set of data is scaled to a common base and relative temperature factors between different sets of data are corrected. The Kraut factor (Kraut et al 1962) which is an objective and independent "figure of merit" for comparison of scaling (the limit is 1) for bl vs b2 is 0.984, for cl vs b2 is 0.960 and for c2 vs b2 is 0.981, indicating a reasonable scaling. Fourier synthesis is done by taking the difference \( \sqrt{I_1} - \sqrt{I_2} \) as amplitude, and phases available in this laboratory originally worked out by Dr. Hayward.

**Method and theory of difference Fourier**

It is well known that in the process of structural determination through crystallographic methods, it takes great effort to obtain phase information. It is therefore desirable to extract as much information as possible from the phase information available. One such attempt is the method of difference Fourier synthesis, in which structural difference between two similar objects can be synthesized by using the amplitude difference and a common set of phases. Mathematically, this is illustrated in figure 4.9. The two vectors F and F' are the structure factors of two objects with a small structure difference. \( \Delta F = F' - F \) is the true structure factor representing their difference. Suppose only vector F, \( |F'| \) and \( \sigma \) are available, this is usually the experimental situation where one has obtained the structure
Figure 4.9 The Argand diagram showing the relationship between $F$, the parent structure factor, and $F'$, the derivative structure factor. In the method of difference Fourier synthesis, vector $\Delta F$ is approximated by vector $A$. 
of an object with amplitude $F$ and phase $Q$, and one has also measured the diffraction intensity of a similar object with amplitude $F'$. Can we extract more information regarding the difference between the two similar objects? In the method of difference Fourier synthesis, the amplitude difference $|F'| - |F|$ and phase of $|F|$, i.e., $Q$, are used to approximate the true difference structure factor $\Delta F$, that is, vector $\Delta F$ is approximated by vector $A$ in figure 4.9. In other words, only the component along $F$ of $\Delta F$ is used to synthesize the structural difference. Such an approach undoubtedly will introduce error. It is therefore important to estimate the associated error.

There are three major sources of error: (1) error in the measurement of $F$, (2) error in the measurement of $F'$, and (3) phase $Q$ is used instead of the true phase. Henderson and Moffat (1971) and later Jaffe (1982) have shown the error in the density map is

$$\langle \Delta \rho^2 \rangle = \frac{1}{V^2} \sum ((2 - m^2) \Delta \rho^2 + \sigma^2)$$

where $m$ is the figure of merit for the phase and $\sigma^2$ is the variance for $\Delta F$. In the case where $m=1$ and $\sigma^2 = 0$, this equation becomes Parseval's equation, also known as the energy equation. The total energy associated with the signal in Fourier space is $\sum \Delta \rho^2$ and total energy in real space is $\langle \Delta \rho^2 \rangle V^2$. $\langle \Delta \rho^2 \rangle$ is the average energy per pixel if the total energy is evenly distributed, and since this is not the case, the signal will emerge.

The final difference Fourier map generated is shown in
figure 4.10. The amplitudes used in the synthesis are the differences between the average of the two native bR and the average of the two papain cleaved bR. (symbolically $\frac{b_1 + b_2 - c_1 + c_2}{2}$) As a control, the difference maps between the two native bR ($b_1 - b_2$) and that between the two papain cleaved bR ($c_1 - c_2$) are also generated as shown in figure 4.11. The maps of $b_1$, $b_2$, $c_1$, and $c_2$ are shown in figure 4.12. The contour levels in figure 4.10 are chosen such that there are a total of six contours, and $<\Delta \rho^2>^{1/2}$ is 0.95 of this contour increment (which is 0.000735). The contour level in maps of $b_1$, $b_2$, $c_1$ and $c_2$ is such that the $\alpha$-helix is represented by eight contours from zero. If we assume there are 20 - 24 amino acid residues in each $\alpha$-helix, then this contour increment (which is ~ 0.0018) represents ~ 3 amino acid residues. The $<\Delta \rho^2>^{1/2}$ of the difference map then corresponds to ~ 1.2 amino acid. The largest peak in the difference map corresponds to 4 amino acid residues.

The location of the C-terminus region thus obtained extends from the rim of the inner central $\alpha$-helix to across the outer slur region which Jap and Glaeser (1983) suggested could be $\beta$ sheet. If we look at the amino acid sequence near the C-terminus, such speculation is not without merit. It should be noted that there are a few amino acids near the N-terminus also cleaved by papain, the candidate for the N-terminus can not be inferred from this work.
Figure 4.10 The contour plot for the average difference map between the native and papain treated bacteriorhodopsin of purple membrane. The contour level is 1.05 of the standard deviation and corresponds to \( \sim 1.3 \) amino acids.
MAP OF B-C

ROWS 1 TO 7

2 OF 72 COLS 1 TO 125 OF 125

10-JAN-84 00:39:

LEVELS: -0.202E-02 TO 0.239E-02 BY 0.735E-03

XBL 841-126
Figure 4.11 The contour plots for the difference between (a) two native and (b) two papain treated purple membranes.
MAP OF B1-B2
ROWS 1 TO 7
2 OF 72 COLS 1 TO 125 OF 125
10-JAN-84 00:47:
C LEVELS: -0.122E-02 TO 0.107E-02 BY 0.459E-03

(a) XBL 841-114
MAP OF C1-C2

ROWS 1 TO 7

2 OF 72 COLS 1 TO 125 OF 125

C LEVELS: -0.252E-02 TO 0.259E-02 BY 0.852E-03

(b)

XBL 841-129
Figure 4.12 The contour plots for the maps of two native (a) b1 and (b) b2 and two papain treated (c) c1 and (d) c2 bacteriorhodopsin of purple membrane. The contour levels in all cases are chosen so that the $\alpha$ helix is represented by eight contours from zero.
MAP OF B1
ROWS 1 TO 7
2 OF 72 COLS 1 TO 125 OF 125
10-JAN-84 00:25:
C LEVELS: -0.132E-01 TO 0.129E-01 BY 0.174E-02
(a) XBL 841-113
MAP OF B2

ROWS 1 TO 7
2 OF 72 COLS 1 TO 125 OF 125
10-JAN-84 00:30:

C LEVELS: -0.133E-01 TO 0.136E-01 BY 0.180E-02

XBL 841-122
MAP OF C1
ROWS 1 TO 7
2 OF 72 COLS 1 TO 125 OF 12510-JAN-84 00:32:
C LEVELS: -0.133E-01 TO 0.123E-01 BY 0.170E-02
(c) XBL 841-125
MAP OF C2

ROWS 1 TO 7

2 OF 72 COLS 1 TO 125 OF 125

10-JAN-84 00:34:

LEVELS: -0.133E-01 TO 0.121E-01 BY 0.169E-02

(d) XBL 841-128
Chapter 5 Conclusion and summary

Ever since the first crystallographic approach to the interpretation of electron micrographs of periodic biological specimens in 1968 (De Rosier and Klug), there has been great progress in electron crystallography. Specimen preservation through the use of glucose embedment or frozen hydration, together with the method of image enhancement have made low dose high-resolution electron microscopy of unstained biological specimen possible. However, the possible failure of kinematic interpretation for either image-based or diffraction-pattern-based data has been a worry. The pilot study presented here serves a direct and systematic investigation to this problem, with the hope of setting a firm theoretical foundation for the use of electron structural study of biological macromolecules.

The use of the Cowley-Moodie multislice formulation to represent the "exact" case is justified, as it has been known to match well with the experimental results in material science, and in the theoretical aspect, it is also the solution of the Schroedinger equation.

The deviation of this "exact" data from the kinematic case is explored and expressed in terms of (i)
crystallographic R values, (ii) interpretability of the
difference Patterson function obtained from dynamical dif-
fraction intensities and (iii) the results from the Multiple
Isomorphic Replacement phasing scheme for the centric zone.
The results presented in this study have an immediate
relevance. In order to obtain three-dimensional structural
data, the specimen must be tilted, if the specimen thickness
is \( T \) and the angle of tilt is \( \Theta \), then the effective electron
path or thickness is \( \frac{T}{\cos \Theta} \). From the results of this work,
once the tolerance allowed is determined, the maximum angle
of tilt can then be estimated.

In the second part, an experimental work, a comparison
is made between the projected structures of native bac-
teriorrhodopsin and bacteriorhodopsin in which the C-terminus
removed by papain treatment. The method of data collection,
processing and results are presented. A new method to treat
the background is shown. The background in the diffraction
pattern is considered as a systematic-global phenomenon, and
is treated in that sense. Three types are used to determined
the background level, namely (i) low-pass filter, (ii)
Lorentzian-Gaussian function and (iii) Lorentzian-Gaussian
function with a "free segment".

Just as X-ray crystallography has led us to the under-
standing of structure-function relation at atomic resolution
for water-soluble proteins, electron crystallography will
provide us the structure information at higher resolution
for a large class of integral membrane proteins which are known to play a wide variety of functions.
REFERENCES


This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.