LOCATION OF AN EXTRINSIC LABEL IN THE PRIMARY AND TERTIARY STRUCTURE OF BACTERIORHODOPSIN

NANDINI V. KATRE, JANET FINER-MOORE, AND ROBERT M. STROUD
Department of Biochemistry and Biophysics, University of California, San Francisco, California 94143

STEVEN B. HAYWARD
State of California Department of Health Services, Berkeley, California 94704

ABSTRACT We located a heavy metal label, mercurilated phenylglyoxal, in both the primary sequence and in the tertiary structure of bacteriorhodopsin. This label modified arginines 225 and 227, which are on the COOH-terminal helix (G). In the projected electron potential difference map, the major site is close to the central inner helix. From this result we conclude that helix 1 could not be the COOH-terminal helix G. We tested the multiple isomorphous replacement method for obtaining phases for purple membrane by electron diffraction.

INTRODUCTION

Bacteriorhodopsin (bR), the major protein component of purple membranes from *Halobacterium halobium,* functions as a light-driven proton pump (1). The 27,000-dalton protein forms a two-dimensional trigonal lattice in the plane of the membrane and traverses the membrane seven times via predominantly α-helical rods (2). Susceptibility to chemical modification and to various proteases suggests that the known amino acid sequence of bR (3, 4) can be partitioned into seven transmembrane helical segments (Fig. 1). The assignment of the probable helical segments to particular electron density rods is not known, although several arrangements have been proposed (5–7). To better define the arrangement experimentally, we covalently labeled bR with a heavy-atom chemical label. We sought to locate the label attachment site in the sequence and in the structure by electron diffraction analysis. Success in this venture requires that the label have high specificity for a small number of sites, with a high stoichiometry of labeling, and that the label be ordered in the bR crystal lattice. We report on one successfully located label, mercurilated phenylglyoxal.

METHODS

Materials

Purple membranes were isolated from *Halobacterium halobium* (strain R_1) (8). Phenylglyoxal and mercuric acetate were obtained from Aldrich Chemical Co., Inc., Milwaukee, WI, and used without further purification. ^{203}Hg acetate was purchased from Amersham Corp., Arlington Heights, IL, at a specific activity of 0.95 mCi/mg. Formic acid was purchased from Aldrich Chemical Co. and was distilled before use.

Protein concentration was assayed by absorbance at 570 nm using the extinction coefficient for the light-adapted form of bR: $\epsilon_{570} = 63{,}000$ $cm^{-1}M^{-1}$ (9). Optical densities were measured on a Beckman DU spectrophotometer (Beckman Instruments, Inc., Palo Alto, CA) equipped with a digital detection system (Gilford Instrument Laboratories, Inc., Oberlin, OH). Amino acid analysis was carried out on a Beckman 121 M amino acid analyzer following acid hydrolysis in 6 N HCl at 110°C for 22 h. Radioactivity was counted in a Beckman model 300 Gamma Counter.

Synthesis of Mercurilated Phenylglyoxal

Mercurilated phenylglyoxal (HgPG) was synthesized as follows: 0.2 g phenylglyoxal in 4.2 ml 0.38 M potassium hydroxide, and mercuric acetate (0.42 g in 4 ml water and 0.013 ml glacial acetic acid) were each stirred separately at 60°C for 30 min and filtered. Mercuric acetate was added dropwise with constant stirring at 60°C to the filtered phenylglyoxal over a 3-min period. The reaction mixture was stirred for 10 min more and filtered hot through a sintered glass funnel. The precipitated compound was washed sequentially with 7% acetic acid, water, methanol, and ethylether and finally recrystallized from dioxane. Radiolabeled HgPG was synthesized using ^{203}Hg acetate added as a tracer to the filtered cold mercuric acetate solution just before addition to the phenylglyoxal solution.

Modification of Purple Membrane with HgPG

Purple membrane (PM) was labeled at 20°C in the dark using HgPG in dioxane (5.5 mg/ml) as follows: four 50-μl aliquots of this solution were added at intervals of 8 h to 20 mg of membranes suspended in 20 ml of 0.2 M sodium bicarbonate buffer (pH 8) with stirring. The reaction mixture was stirred for a further 24 h. Membranes were washed four times with the bicarbonate buffer (pH 8) by centrifugation in a Sorvall SS-34 rotor at 18,000 rpm (Ivan Sorvall Inc., Newtown CT). ^{203}HgPG-labeled membranes were used to locate the site of modification. An aliquot of these membranes was subsequently used for electron diffraction. ^{203}HgPG-modified PM was bleached with hydroxylamine and cut with α-chymotrypsin (CT) between residues 71 and 72 (10, 11). After CT cleavage, the membranes were dissolved in 2.5 ml formic acid, and 7 ml absolute ethanol was added to the dissolved membranes. This treatment delipidated the membrane. The chymotryptic fragments in the clear HCOOH/ETOH mixture were separated on Sephadex LH-60, cleaved with CNBr, and chromatographed as described earlier (10, 11). The CNBr peptide containing the ^{203}HgPG was subjected to amino acid analysis.

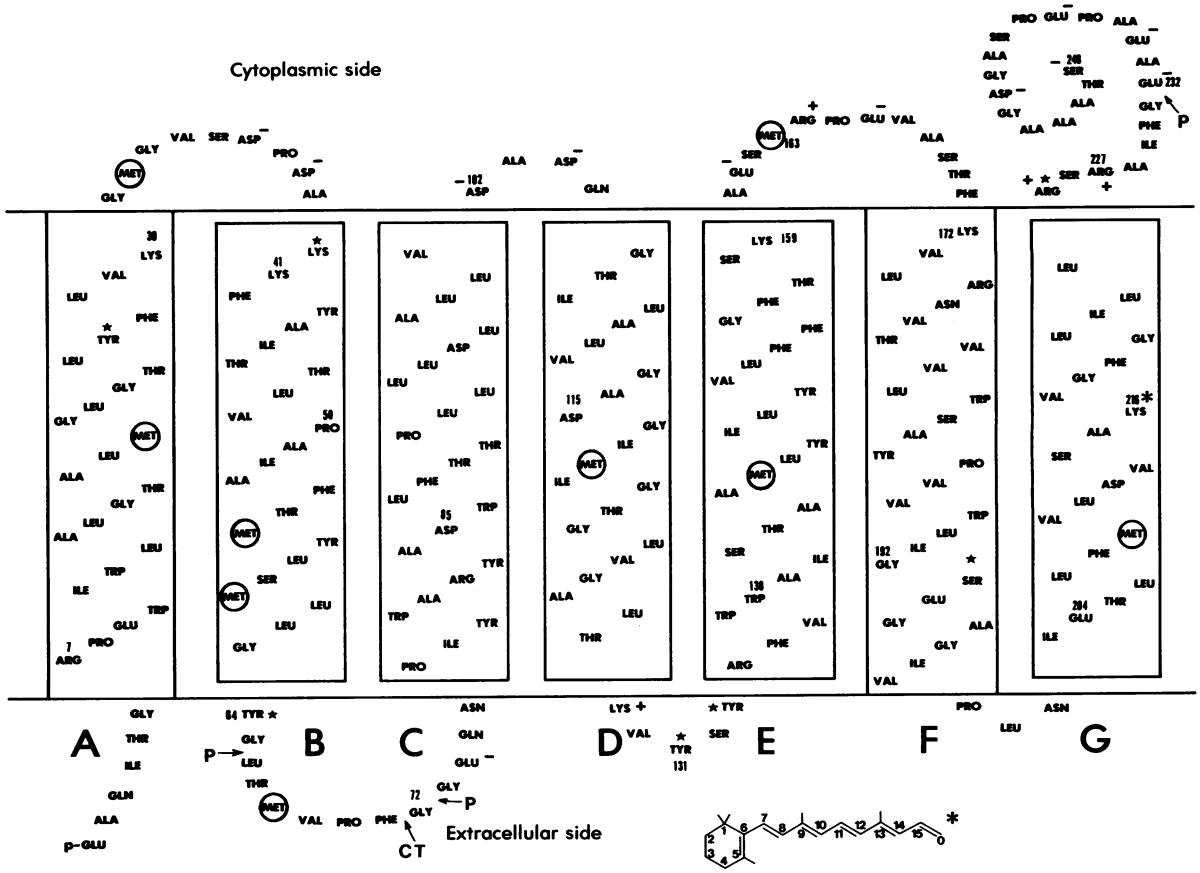

FIGURE 1 The primary sequence of bacteriorhodopsin shown arranged in seven transmembrane segments. This arrangement is consistent with chemical labeling (indicated by an asterisk), protease susceptibility (*P* and *CT* indicate papain and α-chymotrypsin cleavage sites), and hydrophobicity data. Methionine residues, sites of CNBr cleavage, are also marked. Residues are arranged as unfolded helices with 3.5 residues per turn and a helical pitch of 5.4 Å. Horizontal lines represent membrane surfaces separated by 45 Å. Charges are shown only on residues outside the membrane region. Retinal, shown in all-*trans* form, on the lower right, is attached as a Schiff's base to lysine 216.

Electron Diffraction

Electron diffraction was carried out as described (12) on a JEOL 100B electron microscope (JEOL USA, Peabody, MA), using the cold stage supplied by JEOL. Mercury-labeled and native membranes were applied to grids in a 1% glucose, 0.01% sodium azide solution and light-adapted for 5 min. Samples were maintained at −120°C, and the total electron dose for each electron diffraction pattern was 2 e/Å^2, as measured with a Si(Li) detector. Patterns were recorded on Kodak 4463 film (Eastman Kodak Co., Rochester, NY), which was developed in fresh Kodak D19 at 20°C for 12 min for maximum speed.

Data Processing

To avoid systematic errors in intensities, it was important to select diffraction patterns of membranes that were perpendicular to the electron beam. Membrane tilt (azimuth and tilt angle) for each pattern was calculated by least-squares minimization of the differences between observed F_{hk} values for all reflections to 7 Å resolution (including their five symmetry mates), and F's predicted from the gradients $(dF_{hk}/dz^*)_{z^*=0}$ known from the three-dimensional data (2). The patterns used were tilted <2.7°, with mean tilts of 1.26 ± 0.8°.

Electron diffraction patterns were digitized in 5 × 5 μM pixels along strips 49 pixels wide, centered on lattice rows parallel to one of the unit cell axes using a PDS microdensitometer (Perkin-Elmer Corp., Norwalk, CT). All reflection intensities from ∞ to 4.5 Å resolution were digitized. Optical densities (OD) were converted to electron exposure I_e according to:

$$\mathrm{OD} = A0 + A1\exp(-A2I_e), \quad (1)$$

where the parameters $A0$, $A1$, and $A2$ were determined by least-squares analysis of calibration films of known exposure created and scanned at the same time as the diffraction films. After conversion to electrons, data were smoothed by the following algorithm:

$$I_{x,y} = [5I_{x,y} + 2(I_{x-1,y} + I_{x+1,y} + I_{x,y-1} + I_{x,y+1}) - (I_{x-1,y-1} + I_{x-1,y+1} + I_{x+1,y-1} + I_{x+1,y+1})]/9.0, \quad (2)$$

where x and y are pixel coordinates for all pixels in the entire pattern scan. This smoothing was carried out to facilitate location of centers of diffraction spots.

The locations of diffraction spot centers were first estimated using approximate lattice parameters. A block of 49 × 49 pixels was defined about the estimated center of each spot in the scan direction. To locate the spot centers more precisely, the data around each spot were match-filtered using an algorithm similar to that applied to neutron diffraction intensities (13). Each block of data was searched for the group of intense pixels that best fit the expected spot size and shape (in this case, a circle of radius three pixels). Two-dimensional lattice parameters were then refined to the new spot centers and the centers were forced to lie on this lattice. Spot centers were twice recalculated by the matched-filter procedure during data processing, but in these subsequent determinations

the centers were not allowed to move outside the spot boundaries. After the first, but not the second, recentering, spot centers were constrained to lie on newly refined lattice parameters.

Average local backgrounds were calculated from pixels outside the spot boundaries and within the 49 × 49 blocks. Intensities were calculated by summing the background-corrected intensities of pixels within the spot boundaries. Spot size was refined by increasing the radius of the circular boundary in steps of one pixel until the increase in radius did not result in a significant increase (>4 SD) in background-corrected intensity.

To improve background estimates, local backgrounds along each scanned strip were fit to an empirically determined curve:

$$B = A1 + A2X + A3\exp(-X^2/A4) + A5\exp(-|X|^3/A6), \quad (3)$$

where X is the (vector) distance of a block center relative to the center of the strip and $A1$–$A6$ are refined for each strip. This procedure eliminated effects of any abnormally large noise spikes. Fitted backgrounds were used in the final calculations of integrated intensities.

Diffraction amplitudes from native samples were six-fold symmetry-averaged, weighted according to their variance, and then scaled and averaged between patterns to yield $|F_N(h,k)|$. Amplitudes from labeled patterns were similarly treated and scaled to the native set using a two-parameter exponential scale factor (14, 15) to yield $|F_D(h,k)|$. All data between 26.0 and 4.5 Å were used in scaling and in Fourier calculations.

Fourier Analysis

Phases and figures of merit used in the difference Fourier calculations were those determined for the 3.7-Å projected structure of PM at low temperature[1] from low-dose electron images (12). Difference diffraction amplitudes were multipled by the figures of merit m to give the most error-free difference map (16). The average figure of merit for all reflections used was $m = 0.88$. The positions, temperature factors, and occupancies of HgPG sites selected from the difference map were refined by least-squares minimization (17) of:

$$\sum_{h,k} \Delta\Delta F^2 = \sum_{h,k} (|F_D| - |F_N + F_H|)^2, \quad (4)$$

where F_H is the calculated structure factor contribution from the HgPG label, F_N and F_D are as defined previously, and the sum is over all independent reflections. Scattering amplitudes for electrons were from the International Tables for X-Ray Crystallography (18).

RESULTS

Location of the Label in the Primary Sequence

HgPG, synthesized as described, gave the elemental analysis listed in Table I. The percent by weight of each element in the sample depends on the percent of the compound that was mercurilated. The Hg and C percentages determine the percent mercurilation as 56% ± 1% and 65% ± 1%, respectively, and the Hg/C ratio determines it to be 60% ± 0.4%. The degree of mercurilation is much less precisely determined by the percentages of oxygen and hydrogen in the sample since these vary much less between 0 and 100% mercurilation than do percentages of carbon and mercury. The effective formula weight of the label was calculated from the molecular weight of PG (134) and HgPG (393) to be 289.4 (134 + [0.6 × (393 − 134)]).

TABLE I
ELEMENTAL ANALYSIS OF MERCURILATED PHENYLGLYOXAL

Calculated percentage of the composition (by weight) for X% mercurilated PG				Y = Found percentage of composition (by weight)	Y/atomic weight	Calculated percentage of mercurilation
X =	0%	50%	100%			
C	71.7	41.0	30.5	38.7 ± 0.2	3.2	64.6 ± 0.9
H	4.5	2.7	2.1	2.8 ± 0.2	2.7	45.0 ± 10
O	23.9	18.3	16.3	14.2 ± 0.4	0.9	57.0 ± 8
Hg	0.0	38.1	51.1	40.4 ± 0.4	0.2	56.0 ± 1

The specific activity of the label (371.1 cpm/μm) was used to calculate the molar ratio of total label (and of Hg) to protein. After extensive washing, the membrane fraction retained 237,500 cpm/μm protein, which is equivalent to 2.21 mol of total label and 1.33 mol of Hg per mole of protein. This stoichiometry applies to the sample used for diffraction.

After bleaching and chymotryptic cleavage, both at 37°C and pH 7 for a total of 6 hr, but before column separation, the specific activity fell to 1.34 mol of total label and 0.80 mol of Hg per mole protein. After separation of CTI/CTII, radioactivity was detected only on the CTI peptide. After CNBr cleavage of CTI (24 h, 37°C), the specific activity corresponded to 0.60 mol of label and 0.36 mol Hg per mole of protein. Radioactivity appeared in a single major peak after CNBr treatment, which suggests that essentially all of the Hg-containing label was attached to a residue(s) of a single CNBr peptide. By the criterion of A_{280} absorbance, there was 90–95% recovery of all peptides after passage through the Sephadex columns. Hence the diminishing specific activity of the radiolabeled peptide reflects the reversibility of the phenylglyoxal modification (see Discussion).

The separation profiles of the CT and CNBr peptides were similar to those obtained earlier (10) and suffered from some overlap of bands. The radioactive CNBr fragment, which contained the HgPG label, ran at the same position as the COOH-terminal peptide (residues 210–248), to which retinal was bound after reduction in light with $NaBH_4$ (11). Amino acid analysis of the HgPG-labeled fragment (Table II) gave a composition most similar to the COOH-terminal peptide (fragment 5). Since papain cleavage of 17 amino acid residues of the COOH-terminus in the HgPG-modified PM did not alter the radioactivity of the sample, the modified residues must be in the sequence 210–232.

Phenylglyoxal (PG) reacts covalently primarily with

[1]Imaging was carried out at −90°C using a high stability cold stage; diffraction was carried out at −120°C on the JEOL cold stage.

TABLE II
AMINO ACID ANALYSIS OF THE CNBr FRAGMENT CONTAINING THE RESIDUE MODIFIED BY HgPG

	Found composition*	Expected composition‡				
		I	II	III	IV	V
	nmol					
Lys	1.67	0	1	1	1	1
Arg§	1.30	1	1	0	2	2
Asp	2.28	5	0	0	2	2
Thr	1.75	3	3	1	3	1
Ser	3.73	0	2	2	3	4
Glu	4.44	3	0	1	3	3
Pro	2.02	2	0	0	3	2
Gly	8.07	5	3	1	3	5
Ala	3.23	6	4	1	3	8
Val	3.03	2	3	1	8	3
Ile	1.15	3	2	1	3	2
Leu	2.56	10	2	3	7	4
Tyr	0.76	2	2	2	1	0
Phe	1.29	1	1	0	2	2
RMS deviation, $\sqrt{\chi^2/12}$‖		± 2.5	± 2.6	± 2.5	± 2.8	±1.6

Weighted least-squares fit of a linear combination of the five fragments to the observed data indicated a composition for the analyzed sample of the following: 67% CNBr V, 24% CNBr III, and <5% each of fragments I, II, and IV. CNBr V contains residues 210–248.

*Found composition is measured in nanomoles and scaled to ~37 total amino acids.

‡The expected composition of the CNBr fragments for CTI are given. Fragments are numbered from the CTI NH_2-terminus.

§Arg was not used in the least-squares fit or the RMS calculations.

‖RMS deviation derived from the observed data after linear scaling to the values in the first column (under Found composition heading).

arginine residues, but also in some cases with histidines, cysteines, and lysines (19). However, reactivity of PG with histidine, cysteine, and lysine is very low (~10% of the reactivity with arginine) and easily reversible. There are no histidines or cysteines in bR. We conclude that the bR residues modified by HgPG were arginines 225 and/or 227. The postulated modification of arginine attaches two molecules of HgPG to the side chain (Fig. 2). Modification

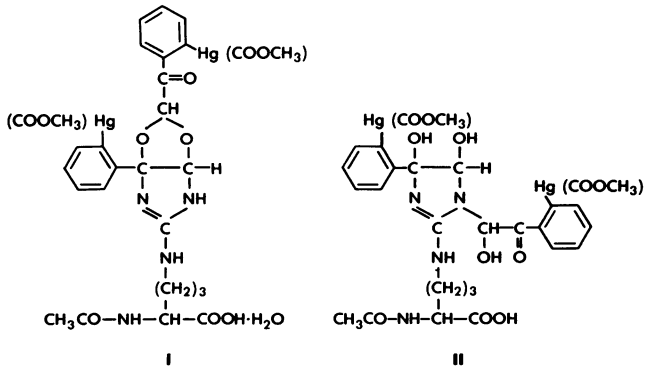

FIGURE 2 Two possible structures for arginine modified by mercurilated phenylglyoxal, derived from those reported by Takahashi (19).

by HgPG did not result in a detectable shift of the 570-nm absorption maximum for the light-adapted PM, or in a change in the PM lattice, as measured by x-ray diffraction, which indicates that labeling did not induce major structural changes in the protein.

Electron Diffraction

Accurate amplitudes were particularly crucial for the calculation of electron scattering potential difference maps since the electron scattering power for mercury is only ~6 times that for nitrogen (the ratio is ~12 for x-rays). Assuming a random distribution of atoms, the expected average change in the amplitudes of reflections from PM on the addition of 2.2 mol $Hg_{0.6}PG$ is:

$$\langle|\Delta F|\rangle/\langle|F_N|\rangle = 2/\pi^* \left| \sum_{i-1}^{ne} f_i^2 \Big/ \sum_{j-1}^{np} f_j^2 \right|^{1/2} = 11\%, \quad (5)$$

for *ne* label atoms of scattering power f_i added to *np* protein atoms of scattering power f_j. For the best determined amplitudes, from 20 to 7 Å resolution, the native amplitudes are about three times higher than estimated in this calculation because of the nonrandom distribution of atoms in helices and the expected relative change in amplitude is therefore even less than 11%; after refinement the calculated values of $\langle|\Delta F|\rangle/\langle|F_N|\rangle$ ranged from 6% at low resolution to 12% at high resolution. Therefore, it was especially important to minimize all sources of error. Errors in the amplitudes were substantially reduced by the matched filter data processing approach, by averaging symmetry equivalent reflections, and by averaging reflections from several patterns. Uncertainties in the data from individual patterns were evaluated before and after averaging of symmetry equivalent reflections using residuals:

$$R = \sum_{h,k} \sum_{i-1}^{n} ||F_i| - |\bar{F}|_{h,k}| / \sum n|\bar{F}|_{h,k}, \quad (6)$$

where $|F_i|$ are the equivalent amplitudes and n is the number of equivalent (hk) reflections compared. The residuals quoted below are for amplitudes and (in brackets) intensities between 26 and 4.5 Å resolution.

Patterns from four unlabeled and five labeled samples were processed. Residuals for Friedel pairs $|F(h,k)|$ and $|F(-h,-k)|$ ranged from R = 8(12)% to 10(16)% for the native patterns and 10(14)% to 13(21)% for labeled patterns. Because the intensities of Friedel mates should be the same regardless of the tilt of the specimen with respect to the incident beam, these residuals reflect random errors in the data. Residuals for threefold symmetric reflections $|F(h,k)|$, $|F(-h-k,h)|$, and $|F(k,-h-k)|$, where each reflection is averaged with its Friedel mate, are sensitive to specimen tilt. These residuals ranged from 7(10)% to 10(15)% for native patterns and from 9(14)% to 15(22)% for labeled patterns. After sixfold averaging of the symmetry equivalent amplitudes of each pattern, the mean residuals between the amplitudes of equivalent patterns were

7.1% (native) and 9.5% (labeled). Amplitudes from equivalent patterns were averaged (Table III), thereby reducing random errors by $\sim 1/\sqrt{N}$, where N is the number of patterns averaged ($N = 4$ for native data and $N = 5$ for labeled data). If all of the errors contributing to the residuals between patterns were random, the average estimated error for the 83 independent native amplitudes $|F_N|$ would be 3.5% and the average estimated error for the 83 independent labeled bR amplitudes $|F_D|$ would be 4.25%.

Fourier Maps

The map calculated with coefficients ΔF and experimentally determined phases (12) is shown in Fig. 3 *a*. Fig. 3 *b* is the 3.7-Å projected structure of the PM reported previously (12). An absolute scale factor for the native amplitudes was worked out by the method of Engelman et al. (5). A difference map was calculated with ΔF's multiplied by the absolute scale and with an $F(0,0,0)$ term for 2.2 molecules $Hg_{0.6}PG$ per molecule bR. The integrated intensity of the main peak on this map corresponds to ~0.83 $Hg_{0.6}PG$ molecules. The phase approximation used in a difference Fourier synthesis results in real features with densities approximately half as large as their theoretical values (20); thus, the peak corresponds to 1.66 ordered $Hg_{0.6}PG$ molecules. Least-squares refinement of the occupancy of the major peak using the scaled amplitudes gave a value of 2.14 ($Hg_{0.6}PG$).

TABLE III
STRUCTURE FACTOR AMPLITUDES FOR NATIVE (FN) AND HgPG-LABELED (FD) PURPLE MEMBRANE

H	K	FN	FD	H	K	FN	FD	H	K	FN	FD
1	2	8.75	7.84	3	7	4.41	4.59	6	4	3.59	3.75
1	3	4.57	3.74	3	8	2.49	2.48	6	5	2.66	2.38
1	4	9.78	10.07	3	9	1.10	1.15	6	6	0.54	0.57
1	5	15.57	14.48	3	10	1.76	2.16	6	7	0.60	0.65
1	6	2.89	2.91	4	0	8.98	9.04	7	0	6.05	5.66
1	7	10.22	11.34	4	1	12.95	14.05	7	1	7.88	6.07
1	8	2.93	2.78	4	2	5.19	4.56	7	2	2.21	2.87
1	9	3.64	3.08	4	3	24.46	28.69	7	3	2.32	2.95
1	10	1.04	1.28	4	4	3.61	3.70	7	4	2.36	2.55
1	11	2.14	2.43	4	5	4.46	4.09	7	5	0.78	0.86
2	1	5.28	4.64	4	6	2.20	2.70	7	6	2.38	3.29
2	2	11.01	11.37	4	7	3.03	2.83	8	0	2.47	2.89
2	3	5.96	4.72	4	8	3.52	3.18	8	1	2.22	2.42
2	4	19.84	20.92	4	9	2.17	2.38	8	2	3.21	3.24
2	5	4.74	3.69	5	0	15.77	16.16	8	3	2.90	3.16
2	6	4.60	4.35	5	1	3.47	3.28	8	4	1.59	1.97
2	7	5.93	5.36	5	2	13.78	15.52	8	5	3.76	3.86
2	8	7.44	4.96	5	3	3.61	3.77	9	0	2.35	2.56
2	9	3.53	4.14	5	4	2.57	2.84	9	1	4.39	3.68
2	10	0.91	1.05	5	5	2.49	2.74	9	2	2.12	2.53
3	0	4.58	4.04	5	6	2.50	2.56	9	3	2.89	2.80
3	1	14.36	13.81	5	7	2.48	2.52	9	4	2.61	2.80
3	2	10.45	9.24	5	8	1.27	1.56	10	0	1.45	1.87
3	3	3.28	3.03	6	0	8.03	7.62	10	1	1.71	1.72
3	4	13.37	13.43	6	1	10.79	13.07	10	2	2.98	3.08
3	5	12.08	13.41	6	2	4.23	3.18	10	3	3.83	2.74
3	6	2.46	2.28	6	3	2.05	2.38	11	0	2.84	2.88

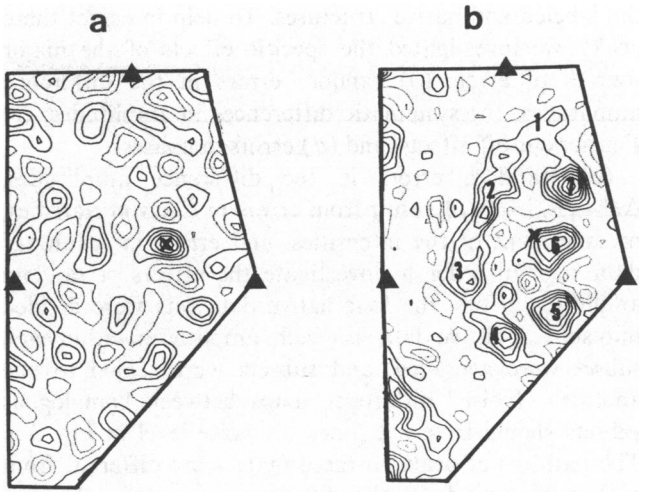

FIGURE 3 (*a*) 120° sector of the 4.5-Å projected difference map calculated with amplitudes $\Delta F = |F_D| - |F_N|$ and plotted with a contour interval of 2.0×10^{-3} Å^{-2}. No $\Delta F(0,0)$ term was used in this or subsequent maps. (*b*) 120° sector of the 3.7-Å projected electron scattering density of purple membrane showing one monomer of bR. The contour level is 1.17×10^{-2} Å^{-2}. The mercury site found in *a* is marked with an X.

Error Analysis

The mean squared noise level for the difference map $\langle \Delta\rho^2 \rangle$ can be estimated using an equation derived by Henderson and Moffat (20):

$$\langle \Delta\rho^2 \rangle = 1/V^2 \sum_{hk} \Delta F(h,k)^2 \cdot [2 - m(h,k)^2] + \delta(h,k)^2, \quad (7)$$

where m is the phase figure of merit, δ is the standard deviation of the difference amplitude ΔF, and the sum is the over all reflections used in the map calculation. Using the figures of merit determined for the native phases (12) and estimating the standard deviation of a reflection by its variation over equivalent patterns, we estimate the error in the map $\langle \Delta\rho^2 \rangle^{1/2}$ to be 0.00318 Å^{-2}, or ~1.6 contours of the difference map in Fig. 3 *a*. This estimated value is 3.4 times smaller than the major peak. Blundell and Johnson (21) derive a different formula for estimation of errors in a difference map calculated with coefficients $2^*\Delta F$:

$$\langle \Delta\rho^2 \rangle = 1/V^2 \sum_{hk} 2^*\Delta F(h,k)^2 + 4\delta(h,k)^2. \quad (8)$$

This formula ignores errors in the protein phases. The major peak in Fig. 3 *a* is ~4.3 times the standard deviation estimated with this formula. By either estimate the major peak, while small, has a <0.1% probability of arising from random errors (22).

The remaining peaks on the map may be the result of errors and approximations or of real differences between

the labeled and native structures. To help interpret these peaks, we investigated the specific effects of the major sources of error: (*a*) random errors in the difference amplitudes; (*b*) systematic differences in amplitudes; (*c*) Fourier cut-off effects; and (*d*) errors in phases.

(*a*) Random errors in the difference amplitudes, $\Delta F = |F_D| - |F_N|$, result from errors in electron statistics, measurement of the intensities, and errors in scaling of data sets. In order to investigate the effects of random amplitude errors, the four native data sets were divided into subsets of two data sets each, amplitudes within each subset were averaged, and subsets were scaled to one another. "Noise" difference maps between homologous subsets should have $\sqrt{2}$ times the noise level in Fig. 3 *a*. The patterns of peaks in these maps were different from each other and from the difference map (Fig. 3 *a*), as expected. One such map is shown in Fig. 4. None of the peaks in these maps were as high as 2σ above the estimated noise level or as large as the largest trough.

The labeled data sets were also divided into subsets and data within each subset was averaged and scaled to native data. Maps between native and labeled subsets (Fig. 5) showed the same pattern of peaks as in difference map Fig. 3 *a*. In each of these maps, the major peak was the same, although its height above the background features varied and was always less than in Fig. 3 *a*. This shows that the similarly disposed peaks in Figs. 3 and 5 are not due to random errors.

(*b*) The tilt of a specimen can introduce a systematic difference into the observed amplitudes that depends on the resolution and on the rate of change of the amplitudes with respect to the Z^* coordinate at $Z^* = 0$. Tilt error is small in reflections where $[dF(z^*)/dz^*]_{z^*=0} = 0$. In cases where $d^2F(z^*)/dz^{*2} = 0$, errors are also small since the six observations within one pattern will be equally distributed in pairs about the correct zero tilt value. For the remaining reflections, any tilt in a pattern will result in a symmetry-averaged $|F|$ that is systematically too large or too small, depending on the shape of $F(z^*)$ vs. z^* curve. This systematic error will partially cancel when calculating ΔF's since the errors are nearly always in the same direction for both native and derivative $F(z^*)$.

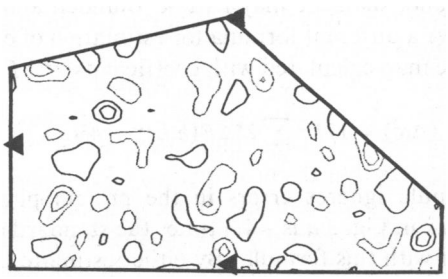

FIGURE 4 120° sector of a test map calculated with amplitudes $|F_{N1}| - |F_{N2}|$, where F_{N1} and F_{N2} are derived from two different subsets of the native bR electron diffraction patterns. The contour interval is the same as for Fig. 3 *a*.

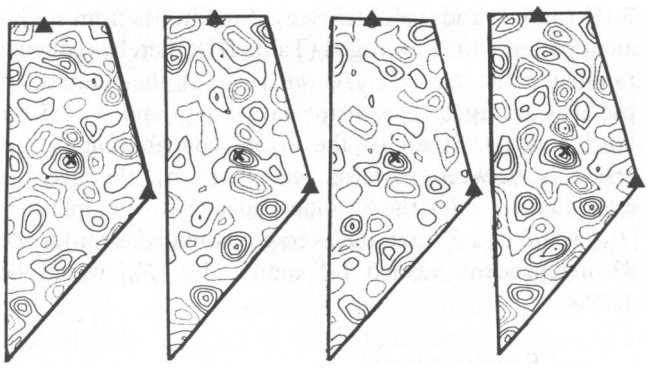

FIGURE 5 120° sectors of test maps calculated with coefficients $\Delta F = |F_D| - |F_N|$, where F_D and F_N are subsets of the electron diffraction patterns. Contour intervals are the same as for Fig. 3 *a*.

The tilt angles for the individual native and labeled PM patterns could be determined to about ±1.0° in azimuth by the method described. Refinement of tilt angles was very sensitive to the subset of reflections and weighting schemes used. Tilt errors in our patterns, where the mean tilt angle was 1.26 ± 0.8° are apparently far less significant than random errors at resolutions of <7 Å. Fig. 6 shows "observed" random errors in ΔF [$\sigma(\Delta F) = (\delta_N^2 + \delta_D^2)^{1/2}$, where δ_N and δ_D are calculated for 24 and 30 observations of each reflection, respectively], predicted systematic differences in ΔF that arise from not correcting for tilt, and observed and calculated $\langle|\Delta F|\rangle$, all vs. resolution. The tilt errors are small for all but a few reflections. Removing the

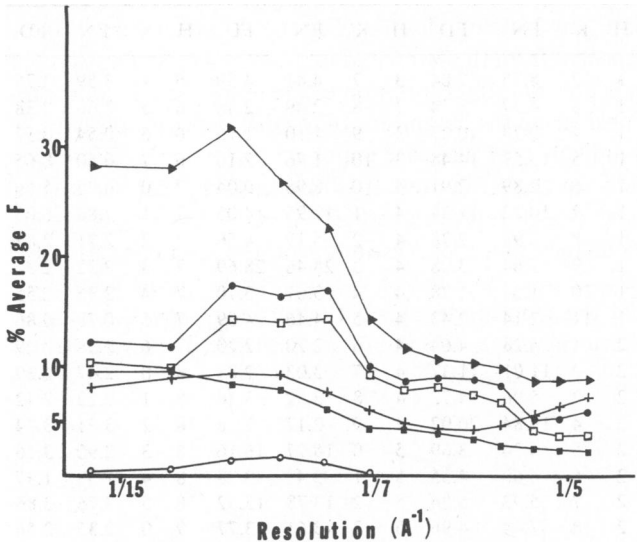

FIGURE 6 A graphic representation of the scattering and errors for the mercurial derivative, after refinement of occupancy and position, versus resolution. (▼) Mean protein amplitude $\langle|F_P|\rangle/6$. (●) Mean difference in amplitude between derivative and native $\langle|\Delta F|\rangle$. (■) Calculated value of $\langle|\Delta F|\rangle$ based on the refined derivative contribution. (○) The errors expected due to tilt for reflections to 7 Å resolution. (+) Observed random errors calculated from multiple observations. (□) Lack of closure errors ($\Delta\Delta F$). Each point on the graph is for 18 reflections centered at that resolution.

tilt-sensitive reflections had little effect on the difference map. Beyond 7 Å, tilt errors are expected to be larger and significant for 10% of the reflections. These errors may contribute to noise in the difference map. Fig. 5 shows effects of tilt (plus random errors) in that the two central difference maps arise from conjugation of the more tilted pairs of native and derivative data sets.

(*c*) The effects of Fourier cut-off were examined by calculating maps with the computed structure factors for a Gaussian distribution of electron potential at the major site, using a refined occupancy and temperature factor. The maps were calculated at 2.0 and 5.5 Å resolution, and in each case the background densities were <10% of the peak. Eliminating high resolution data decreased the intensity of the peak and increased the background slightly. The background features on these maps did not in general correspond to the features of the true difference map shown in Fig. 3 *a*. Therefore, Fourier cut-off is a minor source of the observed peaks.

(*d*) A map was calculated at 4.5 Å resolution using difference amplitudes calculated from the refined potential at the site and observed native phases (12). The background peaks in this map (Fig. 7) were up to 30% of the label peak and overlapped many of the positive features in Fig. 3 *a*. This result suggests that many of the secondary peaks in Fig. 3 *a* result from approximating the ΔF phases with phases from the native structure and from errors in the native phases.

We cannot rule out the possibility that some of the secondary peaks in Fig. 3 *a* result from the differences between the labeled and native PM structures. The fact that the lack of closure (Fig. 6) is much larger than expected from the estimated random and systematic errors indicates that interpretation of the map is still incomplete. Considering the magnitude of the difference amplitudes, it is unlikely that there are major conformational differences between the structures. It is also unlikely that the secondary peaks represent other HgPG sites of significant occupancy since our estimate of the occupancy of the major site is close to the stoichiometry of protein labeling and the magnitudes of all the secondary peaks are $<3\sigma$ above the noise. We therefore limit our discussion to the single most significant site.

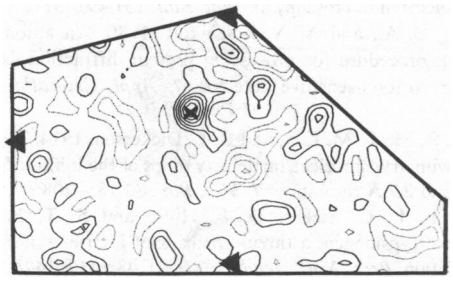

FIGURE 7 120° sector of an "ideal" Fourier synthesis using calculated difference amplitudes for the refined site shown in Fig. 3 *a*, and experimentally determined (12) phases.

Phase Determination by Isomorphous Replacement

The methods so far used for determining high resolution electron diffraction phases include averaging phase information from several electron images (12) and constrained crystallographic refinement of an atomic model to three dimensional electron diffraction data (Agard, D. A., and R. Henderson, personal communication). These two methods have given nearly identical phases to 7 Å resolution but beyond 6 Å resolution, the phases are only very weakly correlated (Agard, D. A., and R. Henderson, personal communication). Therefore, the multiple isomorphous replacement method (MIR) (23, 24) commonly used for protein crystallography may provide a means for phase determination to the highest resolution. We tested an MIR approach using the platinum derivative of Dumont et al. (25) at 5.5 Å resolution, and our $Hg_{0.6}$PG site. The MIR phases out to 7 Å were on average 63° from the other phase sets; from 7 to 5.5 Å, the MIR phases are as different from either set as the other two sets are from each other (~80°). The reasons for the poor results were: (*a*) the platinum and mercury sites are very close to one another in the projected structure; therefore the ambiguities that arise when a single heavy-atom derivative is used to determine phases (24) were not completely resolved; and (*b*) the large lack of closure errors for both sets of data (see Fig. 6). Under more ideal circumstances, perhaps with metal clusters, the MIR procedure provides a second experimental means of phase determination.

DISCUSSION

The specific labeling of arginines 225 and/or 227 by the mercurilated phenylglyoxal implies that these two arginines are accessible from the aqueous phase. This conclusion is consistent with the proposed arrangement (Fig. 1) of the bR sequence in the membrane. In this arrangement, arginines 225 and 227 are on the cytoplasmic surface of the membrane, approximately above helix G. Under certain experimental conditions, additional arginines have been modified with phenylglyoxal or 2,3-butanedione (26). This result suggests that some of the remaining five arginines in bR may be accessible from the aqueous phase but normally in conformations that restrict their interaction with the aqueous phase reagents.

A decrease in the stoichiometry of labeling from 2.2 mol label per mole of purple membrane to 0.6 mol label per mole of bleached, CT cut, delipidated, and CNBr-treated protein suggests that the modification of arginine partially reverts under the conditions for location of the modified protein site. The reversion of PG-modified arginine, which is more likely to occur at neutral or high pH than at low pH, has been discussed by Takahashi (19). Samples for electron diffraction experiments were stored at pH 5.5 in the cold after preparation and examined as soon as possible. Very little reversion of the reaction could have

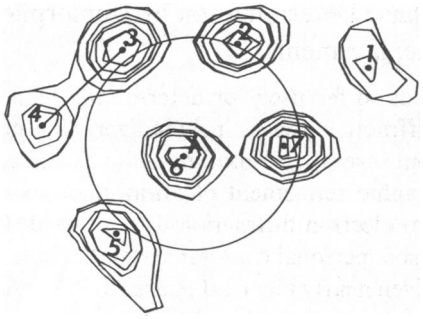

FIGURE 8 Sections from the cytoplasmic side of the three-dimensional electron scattering density of bR (7). The position of the center of label density is marked and a circle of 11 Å radius delineates the area in which the Cα of the modified residue is expected to lie. 11 Å is the distance between the Cα of a fully extended arginine and the center of electron scattering density in the extrinsic label.

occurred before the diffraction experiments since the found occupancy of the label (1.66) was only slightly less than that expected from the stoichiometry (2.2).

The combined length of the phenylglyoxal label and the arginine side chain, ~11 Å, prevents unique location of the C-terminal helix in the three-dimensional structure of bR using the label site identified in the projection map in Fig. 3 *a*. However our result eliminates density labeled 1, and possibly density 4 from being helix G (Fig. 8). This result is fully consistent with all of the five most probable models of Agard and Stroud (7) and with the locations for retinal determined by neutron diffraction (27; Jubb, J. S., D. L. Worcester, H. L. Crespi, and G. Zaccai, manuscript submitted for publication) and by fluorescence energy transfer (28). It is less consistent with the claim of Trewhella et al. (6) that density 4 is helix G.

Mercury is known to associate with sulfur (29), which could order the HgPG enough to be seen in the difference map. Such an association might be detected by an extended x-ray absorbtion fine structure spectroscopy (EXAFS) experiment and would further restrict the possible linking schemes by placing either the *E-F* linker Met 163 or the *A-B* linker Met 32 within 7 Å of the HgPG peak. Three of the five models presented by Agard and Stroud (7) meet this criterion.

CONCLUSION

This is the first reported study to locate a heavy metal label in both the sequence and structure of bR. A major problem with using heavy metal labels to study structure is that the resolution of the structural information that can be obtained from the position of a heavy atom is limited by the size and flexibility of the label, and by the flexibility of the labeled region in the protein. The mercurilated phenylglyoxal we used to label bR is bulky, and the arginine side chain is also long and flexible. As a result, we were not able to uniquely identify the modified helix (G) in the projected structure of the purple membrane, although the results limit the possible models and the locations of arginines 225/227.

We thank Dr. Walter Stoeckenius for purple membranes, Dr. John Kane of the Cardio-Vascular Research Institute for the amino acid analyses, and Dr. Robert Glaeser for use of the JEOL 100B. We also thank Drs. Robert H. Fairclough and David Agard for helpful suggestions, and Ric Bornstein and Mel O. Jones for technical assistance.

This work was carried out under support from the National Institutes of Health (NIH) (GM27057). J. F.-M. was supported by NIH Training Grant 5-T32-CA09270.

Received for publication 20 July 1983 and in final form 9 April 1984.

REFERENCES

1. Stoeckenius, W., and R. A. Bogomolni. 1982. Bacteriorhodopsin and related pigments of Halobacteria. *Annu. Rev. Biochem.* 52:587–616.
2. Henderson, R., and P. N. T. Unwin. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature (Lond.).* 257:28–32.
3. Ovchinnikov, Yu. A, N. G. Abdulaev, M. Yu. Feigina, A. V. Kiselev, and N. A. Lobanov. 1979. The structural basis of the functioning of bacteriorhodopsin: an overview. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 100:219–224.
4. Khorana, H. G., G. E. Gerber, W. C. Herlihy, C. P. Gray, R. J. Anderegg, K. Nihei, and K. Biemann. 1979. Amino acid sequence of bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 76:5046–5050.
5. Engelman, D. M., R. Henderson, A. D. McLachlan, and B. A. Wallace. 1980. Path of the polypeptide in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 77:2023–2027.
6. Trewhella, J., S. Anderson, R. Fox, E. Gogol, S. Khan, D. Engleman, and G. Zaccai. 1983. Assignment of segments of the bacteriorhodopsin sequence to positions in the structural map. *Biophys. J.* 42:233–241.
7. Agard, D. A., and R. M. Stroud. 1982. Linking regions between helices in bacteriorhodopsin revealed. *Biophys. J.* 37:589–602.
8. Oesterhelt, D., and W. Stoeckenius. 1974. Isolation of the cell membrane of *Halobacterium halobium* and its fractionation into red and purple membrane. *Methods Enzymol.* 31:667–678.
9. Rehorek, M., and M. P. Heyn. 1979. Binding of *all-trans*-retinal to the purple membrane. Evidence for cooperativity and determination of the extinction coefficient. *Biochemistry.* 18:4977–4983.
10. Katre, N. V., and R. M. Stroud. 1981. A probable linking sequence between two transmembrane components of bacteriorhodopsin. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 136:170–174.
11. Katre, N. V., P. K. Wolber, W. Stoeckenius, and R. M. Stroud. 1981. Attachment site of retinal in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 78:4068–4072.
12. Hayward, S. B., and R. M. Stroud. 1981. Projected structure of purple membrane determined to 3.7 Å resolution by low temperature electron microscopy. *J. Mol. Biol.* 151:491–517.
13. Spencer, S. A., and A. A. Kossiakoff. 1980. An automated peak fitting procedure for processing protein diffraction data from a linear position-sensitive detector. *J. Appl. Crystallogr.* 13:563–571.
14. Stroud, R. M., L. M. Kay, and R. E. Dickerson. 1974. The structure of bovine trypsin: electron density maps of the inhibited enzyme at 5 Å and 2.7 Å resolution. *J. Mol. Biol.* 83:185–208.
15. Kraut, J., L. C. Sieker, D. F. High, and S. T. Freer. 1962. Chymotrypsinogen: a three-dimensional Fourier synthesis at 5 Å resolution. *Proc. Natl. Acad. Sci. USA.* 48:1417–1424.
16. Blow, D. M., and F. H. C. Crick. 1959. The treatment of errors in the isomorphous replacement method. *Acta Crystallogr.* 12:794–802.
17. Dickerson, R. E., J. E. Weinzierl, and R. A. Palmer. 1968. A least

squares refinement method for isomorphous replacement. *Acta. Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* B24:997–1003.

18. Lonsdale K., editor. 1968. International Tables for X-Ray Crystallography. Kynoch, Birmingham, England. 3:216–227.
19. Takahashi, K. 1968. The reaction of phenylglyoxal with arginine residues in proteins. *J. Biol. Chem.* 243:6171–6179.
20. Henderson, R., and J. K. Moffat. 1971. The difference Fourier technique in protein crystallography: errors and their treatment. *Acta. Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* B27:1414–1420.
21. Blundell, T. L., and L. N. Johnson. 1976. Protein Crystallography. Academic Press, Inc., London. 409–411.
22. Cruickshank, D. W. J. 1949. The accuracy of electron-density maps in X-ray analysis with special reference to dibenzyl. *Acta. Crystallogr.* 2:65–79.
23. Bokhoven, C., J. C. Schoone, and J. M. Bijvoet. 1951. The Fourier synthesis of the crystal structure of strychnine sulphate pentahydrate. *Acta. Crystallogr.* 4:275–280.
24. Phillips, D. C. 1966. Advances in protein crystallography. *In* Advances in Structure Research by Diffraction Methods. R. Brill and R. Mason, editors. Wiley-Interscience, New York. 2:75–140.
25. Dumont, M. E., J. W. Wiggins, and S. B. Hayward. 1981. Location of platinum binding sites on bacteriorhodopsin by electron diffraction. *Proc. Natl. Acad. Sci. USA.* 78:2947–2951.
26. Packer, L., S. Tristram, J. M. Herz, C. Russell, and C. L. Borders. 1979. Chemical modification of purple membranes: role of arginine and carboxylic acid residues in bacteriorhodopsin. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 108:243–248.
27. King, G. I., P. C. Mowery, W. Stoeckenius, H. L. Crespi, and B. P. Schoenborn. 1980. Location of the chromophore in bacteriorhodopsin. *Proc. Natl. Acad. Sci. USA.* 77:4726–4730.
28. Kouyama, T., Y. Kimura, K. Kinosita, and A. Ikegami. 1981. Location and orientation of the chromophore in bacteriorhodopsin. *J. Mol. Biol.* 153:337–359.
29. Karush, F., N. R. Klinman, and R. Marks. 1964. An assay method for disulfide groups by fluorescence quenching. *Anal. Biochem.* 9:100–114.