General Model for Lipid-Mediated Two-Dimensional Array Formation of Membrane Proteins: Application to Bacteriorhodopsin

Mads C. Sabra, Joost C. M. Uitdehaag, and Anthony Watts
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England

ABSTRACT Based on experimental evidence for 2D array formation of bacteriorhodopsin, we propose a general model for lipid-mediated 2D array formation of membrane proteins in lipid bilayers. The model includes two different lipid species, "annular" lipids and "neutral" lipids, and one protein species. The central assumption of the model is that the annular lipids interact more strongly with the protein than with the neutral lipids. Monte Carlo simulations performed on this model show that 2D arrays of proteins only form when there are annular lipids present. In addition, no arrays form if all of the lipids present are annular lipids. The geometry of the observed arrays is for the most part hexagonal. However, for a certain range of low annular lipid/protein ratios, arrays form that have geometries other than hexagonal. Using the assumption that the hydrocarbon chains of the annular lipids are restricted in motion when close to a protein, we expand the model to include a ground state and an excited state of the annular lipids. The main result from the extended model is that within a certain temperature range, increasing the temperature will lead to larger and more regular protein arrays.

INTRODUCTION

The resolution of membrane protein structure is one of the major challenges of modern biophysics. The traditional method of protein structure determination, x-ray crystallography, fails in most cases to produce high-resolution structures of membrane proteins because of the difficulty of obtaining large well-ordered three-dimensional crystals. Only a few high-resolution structures of membrane proteins have been obtained with this method. Among these are the bacterial photosynthetic reaction center (Deisenhofer and Michel, 1989), the very first membrane protein structure to be determined to high resolution, bacterial porins (Weiss et al., 1990; Pebay-Peyroula et al., 1995; Cowan et al., 1995), light harvesting complexes (McDermott et al., 1995; Koepke et al., 1996), cytochrome *c* oxidases (Iwata et al., 1995; Tsukihara et al., 1996), and a photosystem I complex (Krauss et al., 1996; Schubert et al., 1997). Also, more recently, the structure of bacteriorhodopsin has been determined at 2.5 Å (Pebay-Peyroula et al., 1997).

An alternative approach to obtaining molecular structures of membrane proteins, albeit at lower ($\gtrsim$3 Å) resolution, is to use electron diffraction methods applied to two-dimensional (2D) membrane protein arrays. These methods were originally applied to obtain a three-dimensional (3D) model of the purple membrane (Henderson and Unwin, 1975) and, later, a detailed map of the transmembrane part of bacteriorhodopsin (Henderson et al., 1990). Recently the structure of the surface of bacteriorhodopsin has been determined to high resolution by electron diffraction methods (Kimura et al., 1997).

Electron diffraction applied to 2D protein arrays is also the basis for the progress made toward determining the structures of the bacterial porins OmpF (Sass et al., 1989) and PhoE (Jap, 1988; Jap et al., 1991), of plant light-harvesting complex II (LHC-II) (Kühlbrandt and Downing, 1989; Wang and Kühlbrandt, 1991; Kühlbrandt and Wang, 1991; Savage et al., 1996; Falchmann and Kühlbrandt, 1996), of erythrocyte band 3 protein (Wang et al., 1993), and of frog rhodopsin (Unger et al., 1997).

The applicability of electron diffraction methods to determine membrane protein structures relies heavily on the availability of large well-ordered 2D arrays of the membrane protein in question. Very few membrane proteins form 2D arrays in vivo. An example is bacteriorhodopsin (Blaurock and Stoeckenius, 1971), which also forms arrays in a predictable way in vitro (Sternberg et al., 1992). However, most membrane proteins do not form arrays in vivo, and it is therefore of great importance to develop methods to promote array formation in vitro (for reviews see (Kühlbrandt, 1992; Jap et al., 1992; Dolder et al., 1996; Rigaud et al., 1997)).

Two-dimensional arrays have been obtained of only a limited number of membrane proteins. Apart from the ones already mentioned, examples include Na^+,K^+-ATPase (Apell et al., 1992), maltoporin (Stauffer et al., 1992), aquaporin (Walz et al., 1997; Cheng et al., 1997), and a photosystem II core complex (Morris et al., 1997). Several other membrane proteins have also been reported to form arrays, but unfortunately, these arrays are often less reproducible or are not of a quality sufficient for electron diffraction (Kühlbrandt, 1992). These observations are very important, though, because they indicate that many different

Received for publication 1 December 1997 and in final form 5 June 1998.

Address reprint requests to Dr. Anthony Watts, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, England. Tel.: 44-1865-275268; Fax: 44-1865-275234; E-mail: awatts@bioch.ox.ac.uk.

Dr. Sabra's present address is Department of Chemistry, Technical University of Denmark, DK-2800 Lyngby, Denmark.

Dr. Uitdehaag's present address is Laboratory of Biophysical Chemistry, University of Groningen, 9747 AG Groningen, the Netherlands.

0006-3495/98/09/1180/09 $2.00

proteins may form arrays, and hence it is anticipated that arrays can be obtained more predictably, provided the necessary conditions can be controlled.

Reconstituting integral proteins into lipid bilayers may be difficult, and in the following we focus on the formation of 2D arrays after the proteins have already been reconstituted into the lipid bilayer. Hence it is assumed that the process of 2D array formation can be separated from the reconstitution process, although some evidence suggests that the micelle-to-vesicle transition may be important for the quality of the 2D arrays obtained (Kühlbrandt, 1992; Jap et al., 1992; Dolder et al., 1996; Rigaud et al., 1997).

One of the most studied membrane proteins with respect to 2D array formation is bacteriorhodopsin. The general model we propose in this paper to describe 2D array formation of membrane proteins is inspired by the experimental evidence of 2D array formation of bacteriorhodopsin (Sternberg et al., 1989, 1992, 1993; Watts, 1995). The basic assumption of the model is that some special lipids have a stronger attractive interaction with the membrane proteins than do other lipids present in the membrane. In the case of bacteriorhodopsin it has been shown that some specific polar lipids are essential for 2D array formation (Sternberg et al., 1992).

Theoretical modeling of membranes is a compromise between, on the one hand, complexity to provide realism, and on the other hand, simplicity to allow for feasibility of the calculations and transparency of the results (Mouritsen et al., 1995). Accurate and elaborate molecular dynamics simulations have given substantial insight into protein-lipid interactions (Damodaran and Merz, 1994). For example, Edholm et al. (1995) made a model of bacteriorhodopsin in the membrane. However, because of the complexity of this kind of model, it is only possible to simulate rather small systems, typically only one protein imbedded in a matrix of a few hundred lipids. The aggregation behavior of proteins cannot be studied by this type of simulation. Therefore, we have chosen to use a simpler statistical mechanical model and to apply stochastic (Monte Carlo) computer simulation methods. This type of approach has been used before to study protein organization in membranes (Saxton, 1992; Dumas et al., 1997; Gil et al., 1997). A simple model is less applicable for a given specific experimental system than a detailed one, but a simple model will often be able to grasp the underlying physics of the phenomenon in question in a more transparent way than a complex model.

To use electron diffraction methods on membrane proteins, there are two requirements that must be fulfilled. First, the proteins must be arranged into 2D arrays, and second, the proteins must be rotationally ordered. In the model we propose, the proteins are modeled as disks and they have no internal structure. Hence we cannot study the rotational order of proteins in an array, and in the following we shall be concerned only with the formation of proteins into 2D arrays.

MICROSCOPIC MODEL AND CALCULATIONAL TECHNIQUE

The microscopic model we propose includes two different lipid species, "annular" lipids and "neutral" lipids, and one protein species. The basic assumption of the model is that the annular lipids bind more strongly to the proteins than do the neutral lipids. Furthermore, the proteins are taken to repel each other. This is done to make sure that observed array formation is lipid-mediated rather than a result of direct protein-protein interactions. It should be noted that in this paper the terms "bind" and "binding" do not refer to the formation of a chemical bond, but to a protein and an annular lipid being close to each other because of the attractive interaction.

To study the generic properties of the model, it is implemented in the simplest way possible. All three components are modeled as hard disks, which are taken to be of equal size. Apart from being simple, this choice of particle shape and of the relative particle size speeds up the calculations by several orders of magnitude. More insight could be gained if the shapes of lipids and proteins and their relative sizes were more realistic, but this is beyond the scope of this study and the general conclusions of the model are not changed in making this assumption. The hard disks interact via square well potentials. There are no assumptions about the origin of these potentials, although it is believed that electrostatic interactions are very important, and that hydrophobic mismatch interactions may also play a role. The generic properties of the model are not expected to depend on the relative sizes of the particles or on the shape of the potentials. However, details of some of the properties might very well be dependent on both. Likewise, the set of parameters is chosen to study the generic properties of the model, and not to fit a particular experimental system in every detail.

The Hamiltonian of the system can be written as

$$\mathcal{H} = \sum_{\langle i,j \rangle} V_{p_i p_j}(x_{ij}), \tag{1}$$

where $\langle i, j \rangle$ denotes each pair of particles in the system and x_{ij} is the distance between the particles. $p_i = \{p, a, n\}$ is the type of particle i. The neutral lipids are denoted by n, the annular lipids by a, and the proteins by p. $V_{p_i p_j} = V_{p_j p_i}$ are the potentials between particles and are given by

$$V_{nn} = V_{np} = V_{na} = V_{aa} = \begin{cases} \infty & \text{for } x \le 2r \\ 0 & \text{for } x > 2r, \end{cases} \tag{2}$$

$$V_{ap} = \begin{cases} \infty & \text{for } x \le 2r \\ -k & \text{for } 2r < x \le 2r + d/r \\ 0 & \text{for } x > 2r + d/r \end{cases} \tag{3}$$

and

$$V_{pp} = \begin{cases} \infty & \text{for } x \le 2r \\ k & \text{for } 2r < x \le 2r + d/r \\ 0 & \text{for } x > 2r + d/r. \end{cases} \tag{4}$$

d/r is the range of the interactions as measured from the surface of the particles, which all have equal radii, r.

Using the assumption that the hydrocarbon chains of the annular lipids are restricted in motion when they are close to a protein, we extend the model to include conformational energy and entropy of the hydrocarbon chains of the annular lipids. Excitations/deexcitations between conformationally ordered and disordered lipid chains are expected to occur in a lipid bilayer in the vicinity of the main gel-fluid phase transition (Mouritsen and Jørgensen, 1994), and hence we assume that the annular lipids can be in one of two states. Either a low-energy and low-entropy state (ground state), representing a state of relatively ordered hydrophobic chains, or an excited state in which the chains are disordered. The degeneracy of the excited state is taken to be 100 times that of the ground state. When in the excited state, interaction of the annular lipids with the proteins is less favorable than when in the ground state. For simplicity, when in the excited state the annular lipids have the same properties as the neutral lipids.

The Hamiltonian of the extended model is given by

$$\mathcal{H} = \sum_{\langle i,j \rangle} V_{p_i p_j}(x_{ij}) + N_x E_x, \tag{5}$$

where N_x is the number of excited annular lipids in the system and $E_x = 1.0$ is the internal energy associated with the excitation in units of $k_B T$, where k_B is the Boltzmann constant. The potentials between the particles remain the same, except that p_i can now take the value $p_i = x$, i.e., excited annular lipid. The potentials involving the excited annular lipids are given by

$$V_{xx} = V_{xn} = V_{xa} = V_{xp} = 0. \tag{6}$$

To determine the equilibrium properties of the model, we use the Metropolis Monte Carlo algorithm (Mouritsen, 1984) on a system consisting of 2000 particles. The system is confined to a square box with side length $L = 100r$. Particles are allowed to move in the plane in small steps, and long-range exchange of particles is applied.

The simulations are always initiated from a completely random configuration, and equilibrium is typically reached after $\sim 10^8$ Monte Carlo steps. During the subsequent 10^9 (approximately) Monte Carlo steps, the microconfigurations (i.e., the positions, and in the case of the extended model the internal state, of all particles) are recorded with certain intervals.

The structure factor, $S(\vec{q})$, of each microconfiguration characteristic of the equilibrium state is calculated as

$$S(\vec{q}) = \left(\sum_i e^{i\vec{R}_i \cdot \vec{q}} \right)^2, \tag{7}$$

where $\vec{R}_i$ is a two-dimensional vector giving the position of protein i, i.e., $S(\vec{q})$ is based on the positions of the proteins only.

The intensity of the most intense Bragg spot in $S(\vec{q})$, $\vec{q} \neq (0, 0)$ is extracted, and this value, relative to the intensity of $S(0, 0) = n_p^2$, is used as an approximate measure of the protein order and is denoted by S_{max}. In a fully ordered system, $S_{max} = 1$, and for a completely disordered system with an infinite number of proteins, $S_{max} = 0$. For a finite system of proteins $S_{max} > 0$, even when the proteins are completely randomly distributed. In this study, the number of proteins is fixed at $n_p = 50$, which gives $S_{max} = 0.2$ for a random distribution.

RESULTS

In the following we present the results obtained from Monte Carlo simulations performed on the basic model (Eq. 1), followed by the results from the extended model (Eq. 5). Typical microconfigurations from the simulations will be shown to give a qualitative description of the data. The corresponding experimental data would be the micrographs obtained from electron microscopy. As a more quantitative measure, we use the structure factor, which corresponds to the diffraction pattern, which can be obtained in an electron diffraction experiment. The intensity of the most intense spot in the structure factor is used as an approximate scalar measure of the order in the system.

The basic model

In the following, the number of proteins in the system is fixed at $n_p = 50$, while the number of annular lipids in the system is varied. This is effectively the same as varying the annular lipid/protein ratio. One might have also chosen to vary n_p. However, it is easier to compare the protein order in different systems when n_p is fixed.

The upper panel of Fig. 1 shows the effect of having both annular and neutral lipids present in the membrane. When all of the lipids present are neutral lipids (Fig. 1 *A*), the proteins are seen to be randomly distributed throughout the system. This is supported by the calculated structure factor (*inset*), which shows no spots, except for the 0th order spot. The situation is the same in Fig. 1 *C*, where all the lipids present are annular lipids. The proteins are randomly distributed, and the structure factor (*inset*) shows no spots, except for the 0th order spot.

When both annular and neutral lipids are present, a well-ordered hexagonal array of proteins is formed. This is seen in Fig. 1 *B*, where the concentration of annular lipids is 20 mol%. The proteins are embedded in a matrix of annular lipids such that each protein has six annular lipid neighbors and each annular lipid has three protein neighbors and three other annular lipids as neighbors. The corresponding structure factor (*inset*) also reveals a very high degree of order in the system, showing several orders of well-defined Bragg peaks arranged in a hexagonal pattern.

For low concentrations of the annular lipids, an interesting effect is observed, as shown in the lower panel of Fig. 1. For 5 mol% annular lipids (Fig. 1 *D*), the arrangement resembles that of 20 mol% annular lipids (Fig. 1 *B*), show-

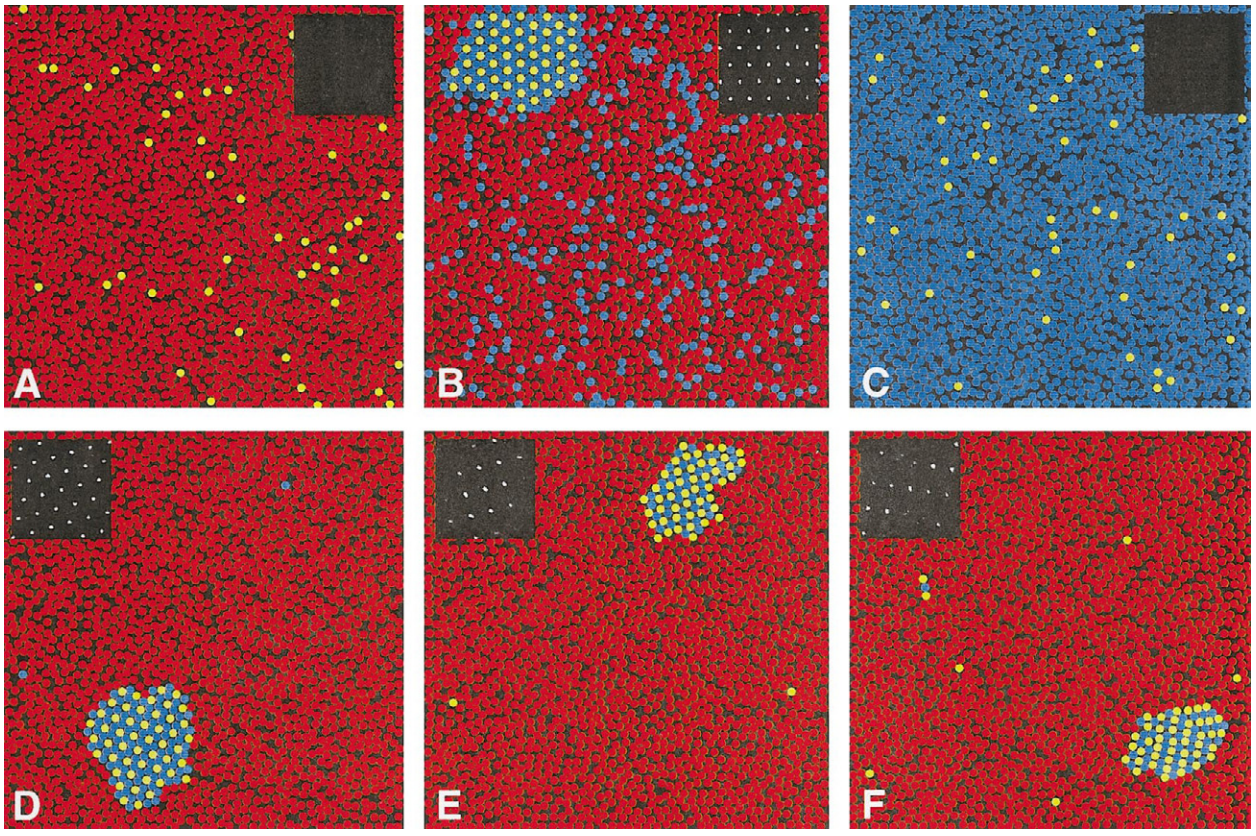

FIGURE 1 Typical microconfigurations obtained from simulations of the model (Eq. 1). The top row shows the effect of having two different kinds of lipids present in the membrane. (*A*) No annular lipids are present. (*B*) 20 mol% annular lipids are present in the membrane. (*C*) All lipids present are annular lipids. The bottom row shows different geometries of 2D arrays formed at low annular lipid concentrations. (*D*) 5 mol% annular lipids. (*E*) 2 mol% annular lipids. In *F* there are also 2 mol% annular lipids, but the range, d/r, of the interactions is changed from $d/r = 0.4$ to $d/r = 0.2$. The proteins are denoted by yellow circles, the annular lipids by blue circles, and the neutral lipids are represented by red circles. The strength of the interactions is $k/(k_BT) = 5$, the number of proteins in the system is $n_p = 50$, and the total number of particles is $N = 2000$. The range of the interactions is $d/r = 0.4$, except in *F*. Contour plots of the structure factor $S(\vec{q})$ in the (q_x, q_y) plane are shown in the insets for $-5 < q_x < 5$ and $-5 < q_y < 5$. $S(\vec{q})$ is calculated from each microconfiguration by using Eq. 7.

ing a well-ordered hexagonal array of proteins in the microconfiguration and clear spots in the structure factor. Lowering the concentration of annular lipids to 2 mol% (Fig. 1 *E*) results in a degradation of the hexagonal array. The structure factor (*inset*) shows only a few Bragg spots arranged in a seemingly quadratic pattern corresponding to a quadratic array. It is not clear, however, whether the observed structure is slightly rectangular or perhaps even orthorhombic. An answer to this question might be found by decreasing the range of the interactions, d/r, but keeping the concentration of annular lipids at 2 mol% (Fig. 1 *F*). Now the proteins are clearly arranged in rows, with rows of annular lipids between them, and the array is not quadratic, but more rectangular or slightly orthorhombic, as also indicated by the structure factor.

The effect of the concentration of annular lipids on the order of the proteins is summarized in Fig. 2, which shows the protein order, S_{max}, defined in above as a function of the concentration of annular lipids in the system. When there are no annular lipids present, S_{max} is low, implying that no 2D protein arrays are present. Adding even small amounts of annular lipid results in the formation of 2D arrays, as revealed by the significantly higher S_{max}. From 0 mol% to 5 mol% annular lipids, there is a steep increase in S_{max}.

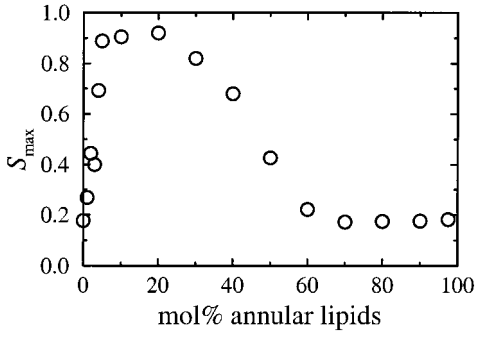

FIGURE 2 The protein order S_{max} as a function of the concentration of annular lipids in the system. S_{max} is defined in the text. The ratio between the interaction constant and the temperature is fixed at $k/(k_BT) = 5$, and the range of the potential is $d/r = 0.4$. The protein concentration is 2.5 mol%, and the system consists of 2000 particles in total.

Increasing the concentration further to ~20 mol% causes the protein order to improve only slightly. Increase of the annular lipid concentration beyond this level leads to a decrease in the protein order up to ~70 mol%, where S_{max} becomes low again, indicating that no 2D protein arrays are formed at high concentrations of annular lipids.

Two important parameters in the model are the strength, k (the interaction constant), and the range, d/r, of the potentials. For simplicity, the repulsive interaction between proteins is taken to be equal in strength and range to the attractive interaction between annular lipids and proteins. It should be noted that in this very simple model, varying k is equivalent to varying the temperature. The controlling parameter is the ratio between the interaction constant and the temperature, $k/(k_BT)$, that is, large values of k are equivalent to low temperatures and vice versa.

The effect on protein order of varying $k/(k_BT)$ is shown in Fig. 3. At low values ($k/(k_BT) < 2.5$), the protein order, S_{max}, is below 0.2, indicating that the proteins are completely disordered. Increasing $k/(k_BT)$ to above 2.5 results in a rise in the protein order, revealing the presence of 2D protein arrays. The slope of the curve is relatively steep to begin with, but levels off, reaching a maximum at $k/(k_BT) = 5.0$; stronger interactions have not been studied.

Fig. 4 shows the protein order, S_{max}, as a function of the range of the interactions, d/r. At very short-range interactions ($d/r < 0.2$), no 2D arrays of proteins are formed, as seen from the low protein order. Increasing the range to $d/r = 0.2$ results in a steep increase of S_{max} to ~0.8, which indicates the presence of highly ordered 2D protein arrays. Further increases of the range cause S_{max} to decrease monotonically until it reaches a minimum at about $d/r = 0.6$, where no 2D arrays are formed. At longer range potentials ($d/r > 0.6$), no arrays are formed.

The extended model

We turn now to a description of the results obtained from the extended model (Eq. 5), which includes two different states of the annular lipids, a ground state and an excited state. Because the relative stability of the two states depends strongly on the temperature, but not on k, the reciprocal relation between the interaction constant and the temperature does not hold for the extended model. By changing the temperature, the fraction of annular lipids in the excited state is also changed, and hence the effective number of annular lipids varies with the temperature. It is therefore not surprising that 2D array formation in the extended model depends strongly on the temperature.

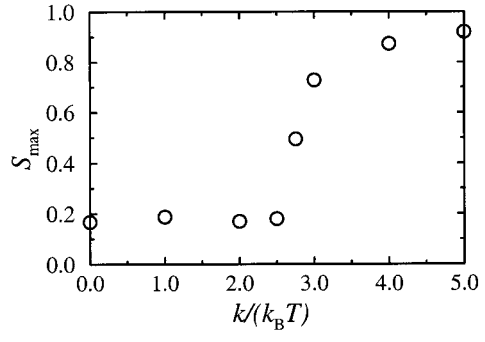

FIGURE 3 The protein order, S_{max}, as a function of the interaction constant, $k/(k_BT)$. The range of the potential is $d/r = 0.4$, and the concentration of annular lipids is 20 mol%. The protein concentration is 2.5 mol%, and the system consists of 2000 particles in total.

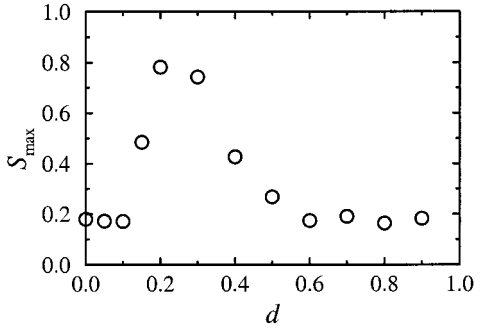

FIGURE 4 The protein order, S_{max}, as a function of the range of the potential, d/r. The concentration of annular lipids is 50 mol%, and the ratio between the interaction constant and the temperature is $k/(k_BT) = 5$. The protein concentration is 2.5 mol%, and the system consists of 2000 particles in total.

At fixed temperatures, the extended model is very much like the basic model. The only difference is that at a given temperature, a fraction of the annular lipids will be in the excited state and therefore have no attractive interaction with the proteins. It should be noted, although it is not important for the results, that this fraction is not constant, but varies in time around a mean value because of the thermal fluctuations of the system.

Microconfigurations from the simulations of the extended model are shown in Fig. 5. It is seen that increasing the temperature in the chosen interval has the effect of enhancing the ordering of the proteins. At $k_BT = 0.18$ (Fig. 5 *A*), there is some ordering in the form of a cluster containing ~10 proteins. There is also a tendency of the remaining proteins to be close to each other, surrounded by annular lipids in the ground state. The corresponding structure factor shows only a few weak spots, and hence supports the observation of a lack of protein order in the system. When the temperature is increased to $k_BT = 0.19$ (Fig. 5 *B*), at least three protein clusters are formed. Two of these have a considerable degree of order, as supported by the structure factor (*inset*), which reveals several orders of weak Bragg spots. At $k_BT = 0.20$ (Fig. 5 *C*), there is one large and relatively well-ordered 2D array of proteins. The order, however, is not perfect. The array has defects, and not all proteins in the system are incorporated into the array. The structure factor (*inset*) shows several orders of bright spots.

Fig. 6 shows quantitatively the effect of the temperature on 2D array formation as calculated from the extended model. At low temperatures ($T < 0.18$) the protein order is around $S_{max} = 0.2$, revealing that the proteins are not

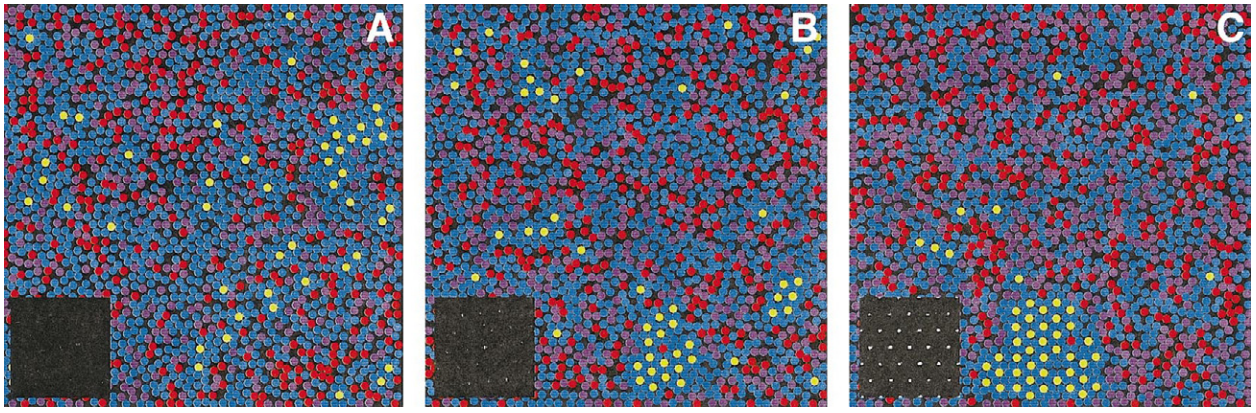

FIGURE 5 Microconfigurations obtained from simulations of the extended model (Eq. 5), showing the effect of increasing the temperature in a certain interval. (*A*) $k_BT = 0.18$; (*B*) $k_BT = 0.19$; (*C*) $k_BT = 0.20$. The interaction constant is $k = 0.2$, and the range of the interactions is $d/r = 0.4$. The proteins are denoted by yellow circles, and the annular lipids in the ground and excited states by blue and purple circles, respectively. The neutral lipids are represented by red circles. The number of proteins in the system is $n_p = 50$, and the total number of particles is $N = 2000$. Contour plots of the structure factor $S(\vec{q})$ in the (q_x, q_y) plane are shown in the insets for $-5 < q_x < 5$ and $-5 < q_y < 5$. $S(\vec{q})$ is calculated from each microconfiguration by using Eq. 7.

ordered. At $k_BT = 0.19$, S_{max} is significantly above 0.2, indicating that the proteins are ordered to some degree, and at temperatures between $k_BT = 0.25$ and $k_BT = 0.35$, S_{max} is large as revealed in large, well-ordered arrays. Increasing the temperature to $k_BT = 0.40$ results in a decrease in S_{max} to ~0.2, which means that the proteins are disordered at temperatures above $k_BT = 0.4$.

DISCUSSION

We have presented the results from a computer simulation study of a general model for 2D array formation of membrane proteins. In addition to proteins, the model involves two different kinds of lipids. One of these, the annular lipid, has a more attractive interaction with the proteins than the other, neutral lipid.

Monte Carlo simulations performed on this model show that large regular hexagonal 2D arrays may form if both kinds of lipids are present. If all of the lipids present are annular lipids, no arrays form (Fig. 1, *top row*). The reason for this is that all protein configurations have approximately the same interaction enthalpy because the number of interacting lipid-protein pairs is almost constant. Whether ordered or disordered, each protein interacts with six annular lipids. To minimize the free energy, the system has to maximize the entropy by distributing the proteins randomly.

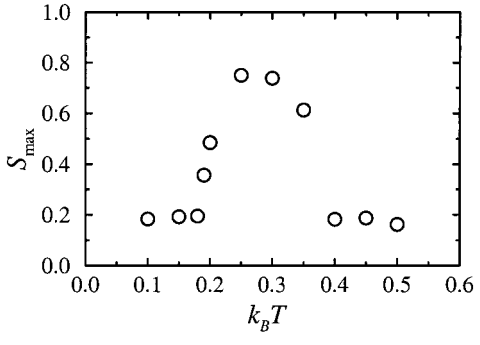

FIGURE 6 The protein order, S_{max}, as a function of the temperature as calculated from the extended model (Eq. 5). The interaction constant is $k = 0.2$, the range of the interactions is $d/r = 0.4$, and the concentration of annular lipids is 80 mol%. The protein concentration is 2.5 mol%, and the system consists of 2000 particles in total.

Decreasing the concentration of annular lipids below 60 mol% leads to an increase in the protein order, indicating that 2D arrays begin to form. The most ordered arrays are formed at 20 mol% annular lipids, but the order remains high down to 5 mol%.

When the total concentration of annular lipids is lowered, the concentration is still relatively high around the proteins because of the attractive interaction. A given protein therefore finds the highest concentration of annular lipids in the vicinity of other proteins. This favors the intermolecular association of proteins leading to 2D array formation.

A more appropriate way of explaining the formation of 2D protein arrays in a mixture of annular and neutral lipids is in terms of entropy maximization. Because of the attractive interaction between proteins and annular lipids, the enthalpy of the system is at a minimum when each protein has six annular lipids surrounding it. The enthalpy of the system can be minimized, for example, by arranging the proteins in hexagonal 2D arrays like the one seen in Fig. 1 *B*. In the system shown in this microconfiguration, there are 2.5 mol% proteins and 20 mol% annular lipids, that is, eight annular lipids per protein. Hence, an alternative way of minimizing the enthalpy would be to distribute the proteins randomly throughout the system each with an annulus of six annular lipids. This "random" configuration maximizes the entropy of the proteins, and at first sight this seems to minimize the total free energy of the system, suggesting this "random" configuration to be the most favorable configuration.

However, in both cases, each protein and the area around it are not available for the neutral lipids because of the accumulation of annular lipids, that is, there is an "excluded

area" around each protein. In the random configuration this excluded area is roughly given by the area of the six annular lipids surrounding each protein. In the 2D array in Fig. 1 *B*, the excluded areas of each pair of neighboring proteins overlap, and therefore the excluded area corresponding to the 2D array is only about one-third of the excluded area corresponding to the random configuration. Consequently, each time a protein is incorporated into an array, each neutral lipid gains translational entropy due to a decrease in the excluded area. The total entropy gain of the system therefore decreases with the number of neutral lipids, and the tendency to form arrays decreases with an increasing concentration of annular lipids. This is the reason for the decrease in S_{max} with increasing concentration of annular lipids seen in Fig. 2.

There is also a contribution to the entropy from the "setting free" of about three annular lipids each time a protein is incorporated into a 2D array. Because of the attractive interaction, each protein imposes order on the annular lipids in its annulus. In a 2D array, only about three annular lipids per protein are ordered, compared to six if the proteins were randomly distributed. Hence entropy is gained by the annular lipids when a 2D protein array is formed.

At low concentrations of annular lipids, the 2D array formation cannot be explained by entropy maximization. In the microconfigurations shown in the bottom row of Fig. 1, all annular lipids present are incorporated into the arrays, no annular lipids have been set free, and hence no entropy has been gained by the neutral lipids. On the contrary, there is a loss of entropy (compared to a random configuration) due to the proteins and annular lipids being restricted in motion when in the arrays. The reason that 2D arrays still form is that the minimization of the enthalpy contributes more to the total free energy than to the loss in entropy.

The different geometries of arrays seen in the bottom row of Fig. 1 can also be explained in terms of enthalpy minimization. When the number of proteins exceeds the number of annular lipids, it is not possible for each protein to have six annular lipids in its annulus. Because of the repulsive interaction between proteins, the opposite configuration with six proteins surrounding each annular lipid is unfavorable. However, it is possible to have four proteins interacting with each annular lipid, giving rise to the nonhexagonal geometries seen in Fig. 1, *E* and *F*. The difference between these two microconfigurations arises from the range of the repulsive interaction, which determines how close the proteins can be without interacting. It is clearly seen that when the range of the interaction is decreased from $d/r = 0.4$ to $d/r = 0.2$, the proteins come closer and form long rows, with rows of annular lipids in between. It is possible that additional geometries of protein arrays could be observed, if there were no repulsive interactions between the proteins at all.

When the concentration of annular lipids is decreased to below 2 mol% as shown in Fig. 1, *E* and *F*, the 2D arrays start to dissolve, and when no annular lipids are present, the only interaction in the system (except for the hard disk interactions) is the repulsive protein-protein interaction, and then the proteins do not order.

It has been shown experimentally that when reconstituted into lipid bilayers consisting of only DMPC, bacteriorhodopsin trimers do not form 2D arrays (Sternberg et al., 1989). Array formation of bacteriorhodopsin trimers in DMPC bilayers only occurs when certain specific lipids are also present in the membrane (Sternberg et al., 1989, 1992). In these lipid mixtures, trimers of bacteriorhodopsin form at least two different kinds of arrays, specifically hexagonal arrays and orthorhombic arrays (Sternberg et al., 1993). These orthorhombic arrays of bacteriorhodopsin trimers should not be confused with the orthorhombic arrays of the monomers, which have also been observed (Michel et al., 1980). The experimentally observed hexagonal arrays of the trimers have a lattice constant of 9.2 nm, and the orthorhombic unit cell dimensions are 9.9×5.9 nm (Sternberg et al., 1993). The results from the model calculations show a similar difference in the dimensions of the two observed geometries; one of the orthorhombic unit cell dimensions is equal to the hexagonal lattice constant, and the other is shorter.

Fig. 2 shows that the optimal concentration for 2D array formation is 20 mol%. However, it is obvious from the microconfigurations shown in Fig. 1 that this optimal concentration would be higher if the protein concentration were higher, and vice versa. Furthermore, from geometric considerations, and from the suggested mechanism for 2D array formation presented above, the size of the proteins as well as the relative size of protein and lipid, is also important for estimating the optimal concentration of annular lipids. A protein such as bacteriorhodopsin, which is much larger than the lipids, would attract more lipids to its annulus, thereby increasing the optimal concentration of annular lipids. Furthermore, as the larger protein would be able to share more annular lipids, the direct enthalpic forces keeping the array together would also be larger.

Arrays only form if the attractive interaction between the annular lipids and the proteins is strong enough when compared to the temperature, as shown in Fig. 3. If the attraction is too low, or if the temperature is too high, 2D arrays are not observed. In an experiment it is the temperature that is most easily controlled, and it has been shown that 2D bacteriorhodopsin arrays undergo a melting transition when the temperature is increased (Koltover et al., 1997). No assumptions about the origin of the attractive interaction between the proteins and the annular lipids in the model have been made. In the case of bacteriorhodopsin, it seems clear that electrostatic interactions play a major role, because the lipids needed for 2D array formation are highly charged (Sternberg et al., 1992; Watts, 1995). However, hydrophobic mismatch interactions may also play an important role. It has been shown theoretically as well as experimentally that lipids with a hydrophobic length which matches that of bacteriorhodopsin have a tendency to be closer to the protein than lipids with another hydrophobic

length (Dumas et al., 1997). It is not clear whether this tendency is strong enough to drive 2D array formation.

The range of the interactions, d/r, is also very important for array formation. As seen in Fig. 4, the range of the interactions has to be in a certain interval for 2D arrays to form. If d/r is either too large or too small, no 2D arrays form. This can be explained in terms of the entropy gain from the "setting free" of annular lipids. When d/r is increased, the annular lipids and the proteins are free to move within a greater distance from each other without increasing the enthalpy. Hence the order that the proteins impose on the annular lipids decreases, that is, the entropy gap between "bound" and "free" annular lipids becomes smaller, as d/r increases. Consequently, increasing d/r decreases the drive for 2D array formation, and therefore 2D arrays are not formed when the range of the interactions is long. When an annular lipid is bound to a protein, entropy is lost. How much entropy is lost depends on the range of the interactions. A bound annular lipid is more restricted when the interactions are short than when they are long, i.e., more entropy is lost for the shorter interactions. Hence binding of an annular lipid is less favorable for the shorter interactions, and decreasing the range of the interactions will lead (on average) to fewer annular lipids around each protein, and thereby to a lower drive for 2D array formation. Consequently, increasing d/r when it is small gives rise to more ordered 2D arrays.

The large effect of the range of the interactions observed in this study is due partly to the shape of the potential used, but it is anticipated that qualitatively similar effects, although smaller, can be obtained with a potential of another shape.

Using the assumption that the hydrocarbon chains of the annular lipids are restricted in motion when bound to a protein, the model is extended to include conformational entropy of the hydrocarbon chains of the annular lipids. The annular lipids can be in one of two states. Either a low-energy and low-entropy state (ground state), representing a state of relatively ordered hydrophobic chains, or in an excited state where the chains are disordered. When they are in the excited state, binding of the annular lipids to the proteins is less favorable than when they are in the ground state. The main result from the extended model is that within a certain temperature range, increasing the temperature will lead to larger and more regular arrays. The ordering of proteins by increasing temperature is due to the effect of temperature on the number of annular lipids with an attractive interaction with the proteins. As the temperature increases, the number of ground-state annular lipids decreases, and this leads to larger arrays via the same mechanism as when the total number of annular lipids is decreased as described above. Consequently, within the extended model, arrays may form even when all lipids present are annular lipids, provided that the temperature has a value so that the right fraction of the annular lipids is in the excited state. It has been observed experimentally that bacteriorhodopsin forms larger and more regular arrays when quenched from higher temperatures (Watts, 1995). Furthermore, most of the 2D arrays obtained for other membrane proteins have been produced at room temperature or higher (Kühlbrandt, 1992). However, if the temperature is too high, no arrays will form. This can easily be understood as a melting transition. The arrays go from an ordered, low-temperature state to a disordered, high-temperature state. In the case of bacteriorhodopsin, the melting of the arrays has been studied by x-rays (Koltover et al., 1997).

In conclusion, from a practical viewpoint for producing 2D arrays for structural studies of membrane proteins, the model presented predicts that 2D array formation of membrane proteins may be promoted by having two different lipid species present in the membrane. One of these species has to interact more strongly with the protein than the other. The relative amount of the two lipid species giving the most optimal conditions for 2D array formation depends on the size of the protein as well as the protein concentration. Furthermore, the model offers an explanation for why increasing the temperature might promote 2D array formation, which is somewhat counterintuitive.

Ling Miao and Tamir Gil are thanked for enlightening discussions. Ole G. Mouritsen is thanked for valuable comments on the manuscript. MCS is affiliated with the Danish Centre for Drug Design and Transport, which is supported by the Danish Medical Research Council.

REFERENCES

Apell, H. J., J. Colchero, A. Linder, O. Marti, and J. Mlynek. 1992. Na,K-ATPase in crystalline form investigated by scanning force microscopy. *Ultramicroscopy.* 42–44:1133–1140.

Blaurock, A. E., and W. Stoeckenius. 1971. Structure of the purple membrane. *Nature New Biol.* 233:152–155.

Cheng, A., A. N. van Hoek, M. Yeager, A. S. Verkman, and A. K. Mitra. 1997. Three-dimensional organization of a human water channel. *Nature.* 387:627–630.

Cowan, S. W., R. M. Garavito, J. N. Jansonius, J. A. Jenkins, R. Karlsson, N. König, E. F. Pai, R. A. Pauptit, P. J. Rizkallah, J. P. Rosenbusch, G. Rummel, and T. Schirmer. 1995. The structure of OmpF porin in a tetragonal crystal form. *Structure.* 3:1041–1050.

Damodaran, K. V., and K. M. Merz. 1994. Computer simulation of lipid systems. *In* Reviews of Computational Chemistry, Vol. V. K. B. Lipkowitz and D. B. Boyd, editors. VCH Publishers, New York. 269–298.

Deisenhofer, J., and H. Michel. 1989. The photosynthetic reaction center from the purple bacterium *Rhodopseudomonas viridis. Science.* 245: 1463–1473.

Dolder, M., A. Engel, and M. Zulauf. 1996. The micelle to vesicle transition of lipids and detergents in the presence of a membrane proteins: towards a rationale for 2D crystallization. *FEBS Lett.* 382:203–208.

Dumas, F., M. M. Sperotto, C. Lebrun, J.-F. Tocanne, and O. G. Mouritsen. 1997. Molecular sorting of lipids by bacteriorhodopsin in DLPC/DSPC lipid bilayers. *Biophys. J.* 73:1940–1953.

Edholm, O., O. Berger, and F. Jähnig. 1995. Structure and fluctuations of bacteriorhodopsin in the purple membrane: a molecular dynamics study. *J. Mol. Biol.* 250:94–111.

Falchmann, R., and W. Kühlbrandt. 1996. Crystallisation and identification of an assembly defect of recombinant antenna complexes produced in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA.* 93:14966–14971.

Gil, T., M. C. Sabra, J. H. Ipsen, and O. G. Mouritsen. 1997. Wetting and capillary condensation as means of protein organization in membranes. *Biophys. J.* 73:1728–1741.

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

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

Iwata, S., C. Ostermeier, B. Ludwig, and H. Michel. 1995. Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans. Nature.* 376:660–669.

Jap, B. K. 1988. High-resolution electron diffraction of reconstituted PhoE porin. *J. Mol. Biol.* 199:229–231.

Jap, B. K., K. H. Downing, and P. J. Walian. 1991. Structural architecture of an outer membrane channel as determined by electron crystallography. *Nature.* 350:167–170.

Jap, B. K., M. Zulauf, T. Scheybani, A. Hefti, W. Burmeister, U. Aebi, and A. Engel. 1992. 2D crystallization: from art to science. *Ultramicroscopy.* 46:45–84.

Kimura, Y., D. G. Vassylyev, A. Miyazawa, A. Kidera, M. Matsushima, K. Mitsouko, K. Murata, T. Hirai, and Y. Fujiyoshi. 1997. Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature.* 389:206–211.

Koepke, J., X. Hu, C. Muenke, K. Schulten, and H. Michel. 1996. The crystal structure of the light-harvesting complex II (B800–850) from *Rhodospirillum molischianum. Structure.* 4:581–597.

Koltover, I., T. Salditt, J. Raedler, C. Safinya, and K. J. Rothschild. 1997. Melting of two-dimensional crystals of the membrane-protein bacteriorhodopsin. *Biophys. J.* 72:A311.

Krauss, N., W.-D. Schubert, O. Klukas, P. Fromme, H. T. Witt, and W. Saenger. 1996. Photosystem I at 4 Å resolution represents the first structural model of a joint photosynthetic reaction centre and core antenna system. *Nature Struct. Biol.* 3:965–973.

Kühlbrandt, W. 1992. Two-dimensional crystallization of membrane proteins. *Q. Rev. Biophys.* 25:1–49.

Kühlbrandt, W., and K. H. Downing. 1989. Two-dimensional structure of plant light harvesting complex at 3.7 Å resolution by electron crystallography. *J. Mol. Biol.* 207:823–828.

Kühlbrandt, W., and D. N. Wang. 1991. Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature.* 350:130–134.

McDermott, G., S. M. Prince, A. A. Freer, A. M. Hawthornthwaite-Lawless, M. Z. Papiz, R. J. Cogdell, and N. W. Isaacs. 1995. Crystal structure of an integral membrane light-harvesting complex from photosynthetic bacteria. *Nature.* 374:517–521.

Michel, H., D. Oesterhelt, and R. Henderson. 1980. Orthorhombic two-dimensional crystal form of purple membrane. *Proc. Natl. Acad. Sci. USA.* 77:338–342.

Morris, E. P., B. Hankamer, D. Zheleva, G. Friso, and J. Barber. 1997. The three-dimensional structure of a photosystem II core complex determined by electron crystallography. *Structure.* 5:837–849.

Mouritsen, O. G. 1984. Computer Studies of Phase Transitions and Critical Phenomena. Springer-Verlag, New York.

Mouritsen, O. G., B. Dammann, H. C. Fogedby, J. H. Ipsen, C. Jeppesen, K. Jørgensen, J. Risbo, M. C. Sabra, M. M. Sperotto, and M. J. Zuckermann. 1995. The computer as a laboratory for the physical chemistry of membranes. *Biophys. Chem.* 55:55–68.

Mouritsen, O. G., and K. Jørgensen. 1994. Dynamical order and disorder in lipid bilayers. *Chem. Phys. Lipids.* 73:3–25.

Pebay-Peyroula, E., R. M. Garavito, J. P. Rosenbusch, M. Zulauf, and P. A. Timmins. 1995. Detergent structure in tetragonal crystals of OmpF porin. *Structure.* 3:1051–1059.

Pebay-Peyroula, E., G. Rummel, J. P. Rosenbusch, and E. M. Landau. 1997. X-ray structure of bacteriorhodopsin at 2.5 Angstroms from microcrystals grown in lipidic cubic phases. *Science.* 277:1676–1681.

Rigaud, J.-L., G. Mosser, J.-J. Lacapere, A. Olofsson, D. Levy, and J.-L. Ranck. 1997. Biobeads: an efficient strategy for two-dimensional crystallization of membrane proteins. *J. Struct. Biol.* 118:226–235.

Sass, H. J., E. Beckmann, F. Zemlin, M. van Heel, E. Zeitler, J. P. Rosenbusch, D. L. Dorset, and A. Massalski. 1989. Densely packed β-structure at the protein-lipid interface of porin is revealed by high-resolution cryo-electron microscopy. *J. Mol. Biol.* 209:171–175.

Savage, H., M. Cyrklaff, G. Montoya, W. Kühlbrandt, and I. Sinning. 1996. Two-dimensional structure of light harvesting complex II (LHII) from the purple bacterium *Rhodovulum sulfidophilum* and comparison with LHII from *Rhodopseudomonas acidophilia. Structure.* 4:243–252.

Saxton, M. J. 1992. Lateral diffusion and aggregation. A Monte Carlo study. *Biophys. J.* 61:119–128.

Schubert, W.-D., O. Klukas, N. Krauss, W. Saenger, P. Fromme, and H. T. Witt. 1997. Photosystem I of *Synechococcus elongatus* at 4 Å resolution: comprehensive structure analysis. *J. Mol. Biol.* 272:741–769.

Stauffer, K. A., A. Hoenger, and A. Engel. 1992. Two-dimensional crystals of *Escherichia coli* maltoporin and their interaction with the maltose-binding protein. *J. Mol. Biol.* 223:1155–1165.

Sternberg, B., P. Gale, and A. Watts. 1989. The effect of temperature and protein content on the dispersive properties of bacteriorhodopsin from *H. halobium* in reconstituted DMPC complexes free of endogenous purple membrane lipids: a freeze fracture electron microscopy study. *Biochim. Biophys. Acta.* 980:117–126.

Sternberg, B., C. L'Hostis, C. A. Whiteway, and A. Watts. 1992. The essential role of specific *Halobacterium halobium* polar lipids in 2D-array formation of bacteriorhodopsin. *Biochim. Biophys. Acta.* 1108: 21–30.

Sternberg, B., A. Watts, and Z. Cejka. 1993. Lipid-induced modulation of the protein packing in two-dimensional crystals of bacteriorhodopsin. *J. Struct. Biol.* 110:196–204.

Tsukihara, T., H. Aoyama, E. Yamashita, T. Tomizaki, H. Yamaguchi, K. Shinzawa-Itoh, R. Nakashima, R. Yaono, and S. Yoshikawa. 1996. The whole structure of the 13-subunit oxidized cytochrome *c* oxidase at 2.8 Å. *Science.* 272:1136–1144.

Unger, V. M., P. A. Hargrave, J. M. Baldwin, and G. F. X. Schertler. 1997. Arrangement of rhodopsin transmembrane α-helices. *Nature.* 389: 203–206.

Walz, T., T. Hirai, K. Murata, J. B. Heymann, K. Mitsuoka, Y. Fujiyosi, B. L. Smith, P. Agre, and A. Engel. 1997. The three-dimensional structure of aquaporin-1. *Nature.* 387:624–627.

Wang, D. N., and W. Kühlbrandt. 1991. High-resolution electron crystallography of light-harvesting chlorophyll a/b-protein complex in three different media. *J. Mol. Biol.* 217:691–699.

Wang, D. N., W. Kühlbrandt, V. E. Sarabia, and R. A. Reithmeier. 1993. Two-dimensional structure of the membrane domain of human band 3, the anion transport protein of the erythrocyte membrane. *EMBO J.* 12:2233–2239.

Watts, A. 1995. Bacteriorhodopsin: the mechanism of 2D-array formation and the structure of retinal in the protein. *Biophys. Chem.* 55:137–151.

Weiss, M. S., T. Wacker, J. Weckesser, W. Welte, and G. E. Schulz. 1990. The three-dimensional structure of porin from *Rhodobacter captulus* at 3 Å resolution. *FEBS Lett.* 267:268–272.