Eur Biophys J (2004) 33: 506–512
DOI 10.1007/s00249-004-0391-6

ARTICLE

Jacki Goldman · Steven Andrews · Dennis Bray

Size and composition of membrane protein clusters predicted by Monte Carlo analysis

Received: 14 October 2003 / Revised: 20 January 2004/Accepted: 22 January 2004 / Published online: 2 March 2004

Abstract Biological membranes contain a high density of protein molecules, many of which associate into two-dimensional microdomains with important physiological functions. We have used Monte Carlo simulations to examine the self-association of idealized protein species in two dimensions. The proteins have defined bond strengths and bond angles, allowing us to estimate the size and composition of the aggregates they produce at equilibrium. With a single species of protein, the extent of cluster formation and the sizes of individual clusters both increase in non-linear fashion, showing a "phase change" with protein concentration and bond strength. With multiple co-aggregating proteins, we find that the extent of cluster formation also depends on the relative proportions of participating species. For some lattice geometries, a stoichiometric excess of particular species depresses cluster formation and moreover distorts the composition of clusters that do form. Our results suggest that the self-assembly of microdomains might require a critical level of subunits and that for optimal co-aggregation, proteins should be present in the membrane in the correct stoichiometric ratios.

Keywords Aggregation · Composition · Lattice gas · Microdomains · Phase change

J. Goldman · S. Andrews · D. Bray (✉)
Department of Zoology, University of Cambridge,
Downing Street, Cambridge, CB2 3EJ UK
E-mail: db10009@cam.ac.uk

J. Goldman
Computing Laboratory, University of Kent, Canterbury,
CT2 7NF UK

S. Andrews
Lawrence Berkeley National Laboratory, Berkeley, CA,
93720 USA

Introduction

The plasma membrane is a thin hydrophobic sheet containing about 50% lipid and 50% protein by weight. Proteins associated with the plasma membrane are highly heterogeneous in structure and function, and also differ widely in their location and mobility. Some proteins span the membrane or are associated only with one or other of its two faces. Some proteins diffuse readily in the plane of the membrane whereas others are associated, more or less firmly, with clusters of other proteins. Protein domains of different sizes and composition, such as synaptic endplates, cell junctions and receptor patches, can be held in place by attachments to the cytoskeleton beneath the membrane or to structures outside the cell, such as components of the extracellular matrix. Adding to this complexity, membranes undergo rapid changes as proteins are removed by endocytosis, added to the membrane by exocytosis, or modified as a result of signals received by the cell (Alberts et al. 2004).

Our interest in the question of membrane dynamics arose in the context of bacterial chemotaxis. Thousands of transmembrane receptors and associated signalling molecules, CheW and CheA, form a cluster, or plaque, usually at one end of an *Escherichia coli* cell (Bren and Eisenbach 2000). This cluster appears to form by the diffusion-limited association of its component proteins, since the deletion of individual components causes the other proteins to become delocalized (Maddock and Shapiro 1993). Furthermore, overexpression of either CheA or CheW to levels much higher than those found in a wild-type cell leads to a loss of chemotactic signalling, a result suggestive of the disruption of a lattice of defined stoichiometry (Bray and Bourret 1995; Liu and Parkinson 1989). Similar effects have been seen in other cell functions that depend on the formation of protein complexes of defined composition (Papp et al. 2003).

In order to examine this question further, we have now written computer programs that simulate the

diffusion and association of proteins in a two-dimensional membrane. The program is based on a simple random walk of protein molecules over a fixed lattice. A Metropolis algorithm is used to determine which states are of lower free energy and therefore favoured (specifically, states with large numbers of protein–protein bonds). We used this program to examine the formation of two-dimensional clusters in a defined area of membrane containing proteins of different densities, bond strengths and bond geometries. In particular, we measured the numbers and sizes of clusters formed at equilibrium and, in the case of mixed structures, the average composition of the clusters formed. Evidently, this model is only a crude approximation to a real membrane and can address only a limited number of questions. However, the results highlight features that we believe must be important for normal cellular processes.

Methods

Each simulation represents a small section of lipid membrane as a two-dimensional integer grid. This can have a trigonal, square, or hexagonal (honeycomb) geometry, depending on the bonding geometry of the protein monomers. Each position in a trigonal lattice has six neighbours spaced at 60° intervals; each position in a square lattice has four neighbours spaced at 90°; each position in a honeycomb lattice has three neighbours spaced at 120°. Lattice points are indexed with periodic (toroidal) coordinates so that a protein molecule leaving one side of the grid automatically re-emerges on the opposite side. The latter tactic allows us to avoid the issue of what happens when a protein collides with a boundary and keeps the protein density constant throughout the simulation.

Each type of protein is defined by a class, or template, which contains the protein name and the number and arrangement of its binding sites. Multiple monomers are created using this template and distributed at random positions and with random orientations over the grid. Bonds form between binding sites on adjacent monomers and have a characteristic strength, depending on the bonding partners, defined in kilocalories per mole.

The simulation uses the Metropolis Monte Carlo algorithm (Metropolis et al. 1953) to explore different states of the system tending towards an ensemble of states of lowest free energy characteristic of the system at equilibrium. Briefly, an iteration consists of a trial move in which a randomly selected monomer takes a step to one of its (randomly chosen) adjacent positions. If the new position is already occupied, then the move is cancelled and the monomer remains in its initial location. As discussed elsewhere (Lamb 1994), this simple algorithm causes the protein molecule to perform a random walk over the surface. We confirmed that the movements of protein molecules in our simulation (under non-crowded, non-associating conditions) did in fact conform to two-dimensional diffusion. For a given set of conditions, the distance traversed from the starting point was found to follow the power law relationship $< x^2 > = 4Dt$, where D is the diffusion coefficient.

For a system containing proteins that can bind to each other, the decision to accept the move is based on the effect it has on the number and strength of intermolecular bonds. At each iteration, a monomer moves to a new position and undergoes a rotation in which its binding sites are shifted by one position (either clockwise or anticlockwise). If in the new position the total energy of its bonds is less than or equal to the energy in its previous position, the move is accepted. Otherwise, the move is accepted with a probability equal to $\exp(-E/kT)$, where E is the energy difference, k is the Boltzmann constant and T is the temperature in degrees kelvin. The number of iterations required to reach equilibrium varies, depending on the number of molecules in the simulation and the size of the lattice, but was usually within the range of 10^7–10^9 iterations for our simulations.

The program records the distribution of protein molecules in the lattice at every time step and allows it to be displayed graphically. Selected time frames can also be analysed within the program to determine the number, sizes and (where appropriate) the composition of each cluster of proteins.

Results

We ran a series of simulations to explore the ability of individual protein species, with sterically defined binding sites, to produce two-dimensional aggregates. Most of this work was performed with proteins on a trigonal lattice, each protein having six equally spaced homophilic binding sites, but similar results were also obtained with proteins on square and honeycomb lattices. Figure 1 shows the progress of a typical simulation. Proteins were initially distributed at random positions (Fig. 1a), but became associated as their independent random movements brought them into contact. After 10^6 steps (Fig. 1b), many small aggregates had formed. As the simulation continued, more clusters were formed which eventually, after 10^8 steps, condensed to a few large clusters (Fig. 1c). We assessed the extent of aggregation by counting the number of protein–protein bonds. Other measurements were also made, such as the mean cluster size, the size of the largest cluster, or the number of monomeric or unassociated protein molecules. However, the bond count was the most useful variable for most purposes, being closely associated with the free energy of the system.

Because of incipient fluctuations, it was impossible to tell by inspection when the system had reached a true equilibrium state. Several strategies were adopted to examine this question. One method was to run the simulation "in reverse", starting with a single large aggregate containing all protein molecules. The aggregate was then allowed to disperse through the action of Brownian motion and end points of the normal and "reverse" simulations were compared. A second strategy employed simulated annealing to accelerate the approach to equilibrium. In this method the simulation starts at a high temperature, which is then lowered stepwise. Simulated annealing helps break up recalcitrant configurations that are locked in unproductive aggregates (Kirkpatrick et al. 1983). Both of the above methods confirmed that the state reached after 10^8 iterations (Fig. 1c) was indeed close to the true (dynamic) equilibrium.

We then performed a series of simulations in which the equilibrium state was measured for different starting parameters, such as the starting density of protein molecules (number of molecules in the lattice) and the strength of the protein–protein bonds. Changing the temperature of the simulation was an equivalent procedure to changing bond strength since this value is relative to kT. A plot of equilibrium state against

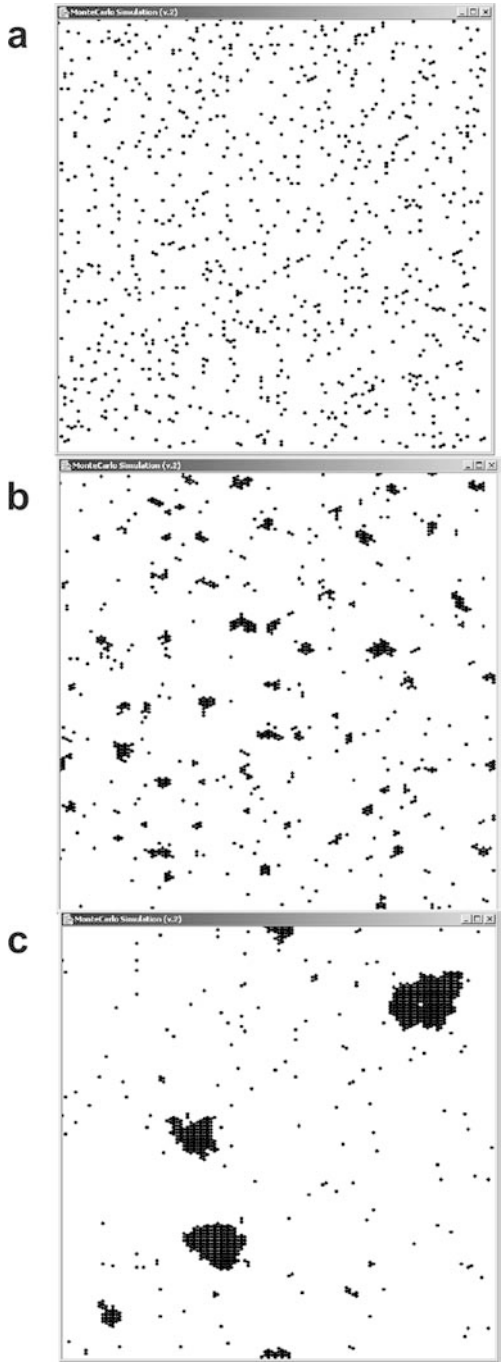

Fig. 1a–c Typical aggregation sequence for a single protein species. Each protein molecule had six symmetrically positioned homophilic (that is, self-binding) sites. Diffusion in two dimensions was simulated by the algorithm described in the text. (**a**) 820 copies of the protein monomer were randomly distributed on a trigonal lattice with 128×128 points. Assuming a unit spacing of 5 nm, the total area represents approximately 1000 nm^2 of lipid membrane. (**b**) After 10^6 iterations, small aggregates of protein molecules had formed, with a mean cluster size of 6.78. The total number of protein–protein bonds formed at this stage was 985. (**c**) After 10^8 iterations the number of bonds formed was 1752 and the average cluster size was 29.42 monomers

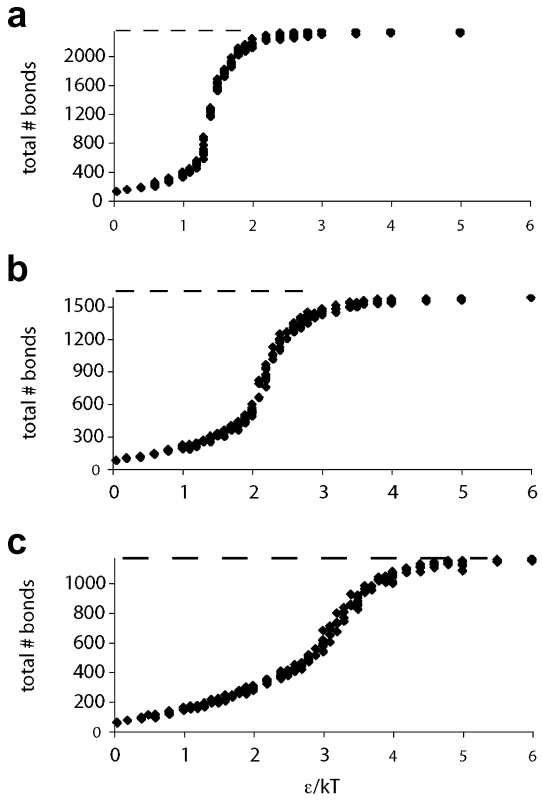

Fig. 2a–c Effect of bond energy. Numbers of protein–protein bonds reached at equilibrium are plotted as a function of bond energy. Simulations were performed on a lattice of 128×128 points populated by 820 protein monomers (5% occupancy). (**a**) Trigonal lattice (six binding sites/molecule). (**b**) Square lattice (four binding sites per molecule). (**c**) Honeycomb lattice (three biding sites per molecules). Simulations were run for 10^8 iterations at each value of the bond strength

progressively increasing bond strength is shown in Fig. 2 for three kinds of lattice. It may be seen that very few protein bonds form at low bond strengths (or high temperatures), but their number rises abruptly in a non-linear fashion. Equivalent results (albeit with a greater degree of noise) were obtained by measuring the size of the largest aggregate formed. Simulations in which the starting concentration of protein monomers was increased incrementally (Fig. 3) also displayed a similar, although less dramatic, threshold. The existence of this ''phase change'', in which the system changes abruptly from a diffuse to a more highly condensed state, is consistent with the behaviour of our simulation as a lattice gas, as discussed below.

Having examined the behaviour of a single protein species, we turned to systems containing two or more types of protein able to form heterophilic protein–protein bonds. In the T_2A_3 system, one protein, designated T, had three binding sites, spaced at 120°, each with affinity for the second protein A. Protein A had two equivalent binding sites with affinity for protein T, oriented at 180° to each other. This configuration is loosely based on the lattice of gephyrin molecules at inhibitory synapses (Choquet and Triller 2003). Spontaneous aggregation of T and A should then generate a lattice of

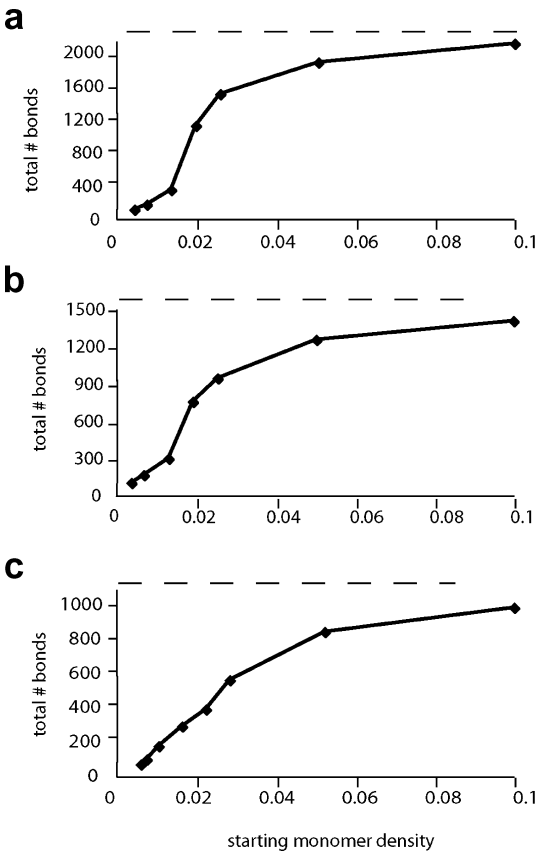

Fig. 3 Effect of monomer concentration. Numbers of protein–protein bonds reached at equilibrium are plotted as a function of the starting concentration of monomers. Simulations were performed as described for Fig. 2

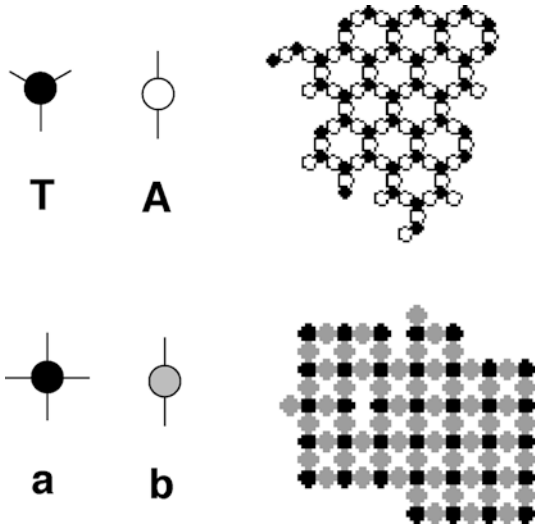

Fig. 4 Large clusters formed by the co-aggregation of two proteins. Proteins T and A form a T_2A_3 cluster on a trigonal lattice. Proteins a and b form an ab_2 cluster on a square lattice

alternating T and A with a characteristic composition (for an infinitely large lattice) of T_2A_3.

The starting proteins were placed at random positions on a 128×128 trigonal lattice with 351 copies of T and 469 copies of A. Diffusion was simulated as before, by allowing individual molecules to step onto vacant adjacent positions. A and T molecules were able to form bonds if they occupied adjacent lattice points and if their binding sites were correctly oriented. In this manner, clusters of molecules containing T and A were established. The time-course of aggregation had similar features to that already shown for a single protein species, with a steep sigmoidal rise occurring at some characteristic bond strength or density of proteins.

We were interested to explore the consequences of changing the ratio of the two proteins, T and A. The intrinsic bonding ratio of these two proteins, specified by their binding sites, is 2T:3A, and we observed that the large aggregates formed from this input ratio indeed had the correct composition (Fig. 4). However, what would happen if the starting mixture contained a different ratio, with either T or A present in excess over that required for stoichiometric binding? We examined this question by running a series of simulations of the T_2A_3 system in which the total number of protein monomers was kept constant and the proportion of T and A molecules was varied. We found that in this case the extent of aggregation at equilibrium, measured either by the numbers of bonds formed or the size of the largest cluster, was sensitive to the input ratio. A maximum extent of aggregation was reached at a specific stoichiometric ratio but decreased on either side of this optimum (Fig. 5a). Similar results were obtained by keeping T constant and varying A; in other words, it appears that an excess of either T or A above the amount required for stoichiometry decreases aggregation. Similar results were obtained for other two-dimensional lattices, such as the ab_2 lattice (Fig. 5b). The basis of the suppression of aggregation became apparent when we inspected aggregates of different sizes in the T_2A_3 system. Small aggregates produced if T was in excess had mostly T on their perimeter, whereas the opposite was true if A was in excess (Fig. 6). Evidently, a bias towards either T or A on the perimeter of a cluster will have the largest effect on overall composition when clusters are small. In a series of tests in which the starting ratio of A and T was changed and the composition of all aggregates examined, we found that the composition of small aggregates was indeed very sensitive to the input ratio (Fig. 7). As the size of the aggregates increased, however, their composition came closer to the stoichiometric endpoint of 0.66 T to 1 A.

Discussion

The program we developed in this work resembles the "lattice gas" models of statistical physics in which a large number of particles representing the molecules of a gas move around on the vertices of a lattice (Newman and Barkema 1999). Lattice gas models are simpler to

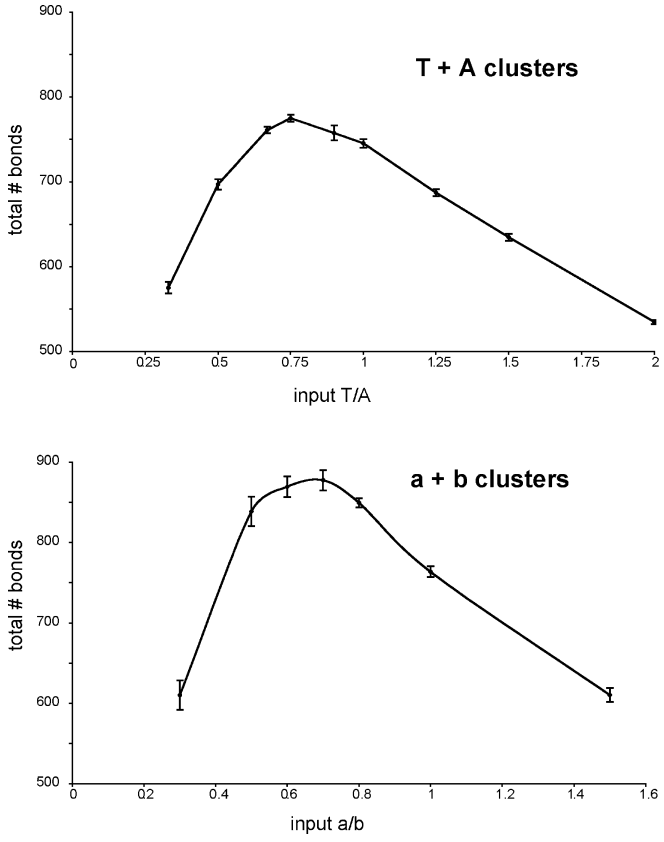

Fig. 5 How the extent of co-aggregation of two proteins is affected by the relative concentrations. Numbers of protein–protein bonds reached at equilibrium are plotted against the input ratio of monomers for the T_2A_3 and ab_2 systems (see Fig. 4). In both cases the total number of monomers in the simulation was fixed

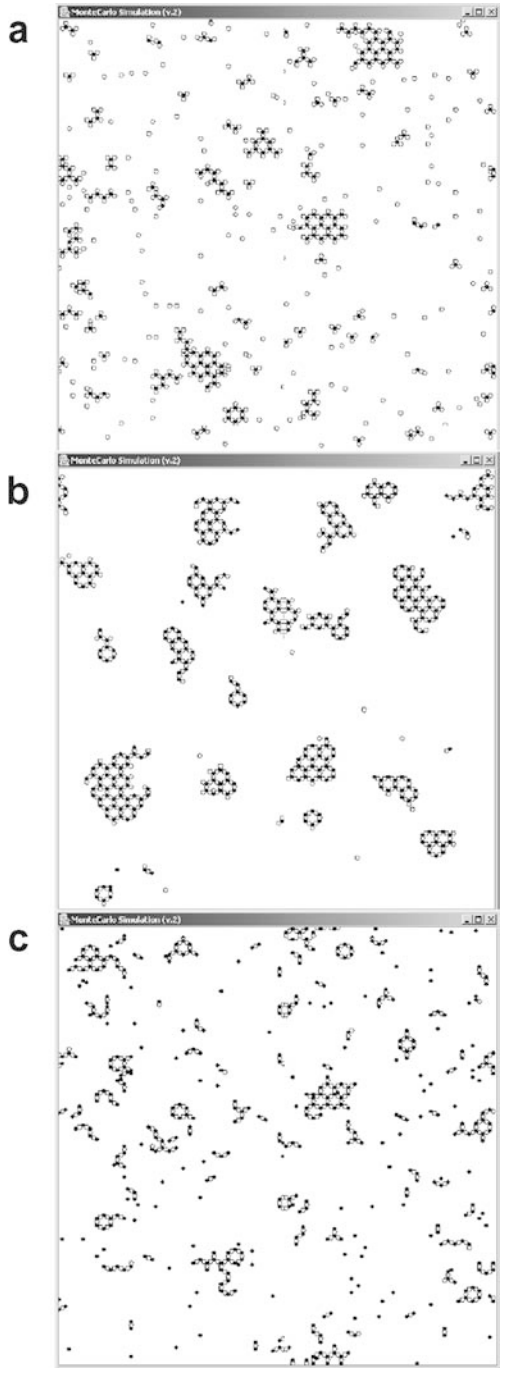

Fig. 6a–c Equilibrium states for different input ratios. The T_2A_3 system was sampled after 10^8 iterations. (**a**) Starting ratio T/A = 0.33, bonds formed = 594, mean cluster size = 9.3 monomers. (**b**) Starting ratio T/A = 0.75, bonds formed = 851, mean cluster size = 35 monomers. (**c**) Starting ratio T/A = 1.5, bonds formed = 646, mean cluster size = 8.6 monomers. Note that each simulation used the same number of protein monomers (T + A = 820)

deal with mathematically than models in which particles can take any position in space, and are less realistic for this reason. However, they have been found to give valuable insight into the general behaviour of real gases. In particular, the equilibrium states reached by a lattice gas should be an accurate representation of the states of the real system, since they depend only on the interaction energies between the molecules.

Similarly, in our simulations we have not attempted to take accurate account of the physical sizes of the protein molecules or of their true rates of diffusion. As a rough guide, lattice points will be separated by 5 nm, since this is a typical packing distance of proteins in an aggregate. A trigonal lattice with 128×128 points (the value in most of our simulations) therefore represents an area of roughly 3.5×10^5 nm^2 (0.35 μm^2) of membrane. Time intervals between iterations, which depend on the number of molecules in a simulation, were chosen to give diffusion rates around 5×10^{-9} cm^2 s^{-1}, which corresponds to values measured for rhodopsin (Poo and Cone 1974).

In broad terms, the simulations met our expectations for a system of proteins diffusing in a lipid bilayer. If individual monomers were tracked in a simulation with zero binding energy, they moved in random walks that carried them a distance proportional to the square root of time, as expected for a diffusing particle. If interactions between monomers were allowed, the extent of cluster formation (measured by the number and size of aggregates or the numbers of bonds formed) increased with time to a plateau value. The plateau values had the

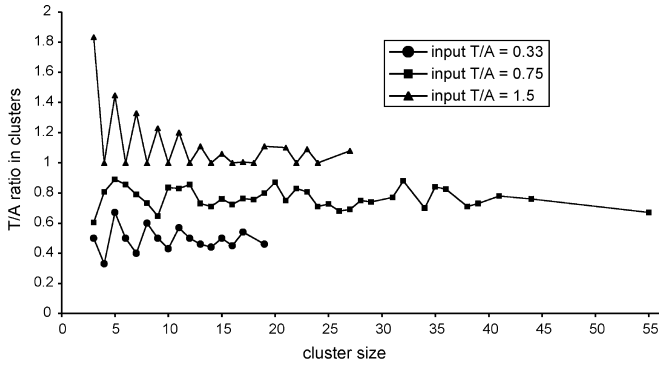

Fig. 7 Plot of composition versus aggregate size for different input ratios. Simulations of the T_2A_3 system were run with the same total number of protein molecules but with different ratios of T and A. Each simulation was allowed to run to equilibrium and then the composition of different size aggregates was calculated

properties expected of steady-state equilibria, displaying fluctuations around a mean value but without a consistent increase or decrease over time. Tests such as running the simulation in reverse and simulated annealing confirmed that plateau values were indeed close to the thermodynamic equilibrium.

It is important to recognize that our lattice gas treatment of membrane proteins ignores many important factors. Interactions with lipid molecules, for example, can lead to dramatic changes in the location of proteins. The thickness of the lipid bilayer and its curvature, as well as its local composition and the dynamic formation of lipid domains, can all influence protein aggregation in a real cell (Gil et al. 1998). Membrane proteins can be tethered to the cell cortex inside the cell, to extracellular matrix molecules outside the cell, or to proteins on the surface of another cell (Alberts et al. 2004). In some cells, such as epithelial cells, diffusion barriers restrict selected proteins to a particular region. Other membrane proteins are in a dynamic rather than an equilibrium state, being rapidly inserted into membranes and removed in response to external signals, as seen for example during T cell recognition (Davis et al. 2003). Each of these issues is a complex subject in its own right and none is addressed in our simple analysis. Nevertheless, just as is it important to measure diffusion coefficients of membrane proteins under defined conditions, we believe it is essential to understand their aggregation behaviour in an idealized environment free of extraneous factors. The phenomena revealed in our study will always be present, even if they are masked by other factors in specific situations.

The extent of aggregation was increased by high concentrations of monomers and large bonding energies but decreased by elevated temperatures. In all of the systems we examined there was an abrupt transition in the extent of clustering with changes in bond strength, concentration or temperature (see Figs. 2 and 3). Transitions of this kind are commonly seen in lattice gases and correspond to the ''phase change'' occurring when molecules of a gas condense to a liquid or solid. Moreover, fundamental thermodynamic considerations of the self-assembly of amphiphilic molecules, such as lipids in two dimensions, predict a critical monomer concentration (CMC) below which no aggregation occurs, but above which monomers will assemble into large aggregates (Israelachvili 1992).

Our results suggest that an analogous transition will occur in the assembly of membrane proteins into aggregates. One possible example of such a phenomenon is seen in the bacteriophage-encoded proteins called holins. These are small membrane proteins that accumulate in the bacterial membrane until, at a specific time characteristic of each holin gene, the membrane undergoes lysis and the cell bursts (Wang et al. 2000). It has recently been suggested that the holins accumulate until they reach a critical level and then assemble into a complex that triggers lysis (Gründling et al. 2001).

The inclusion of two protein species able to form two-dimensional aggregates of mixed composition revealed other phenomena of interest. If the starting ratio of the two proteins was fixed, then the extent of aggregation increased with concentration and bond strength and showed an abrupt phase transition, as for a single protein. However, if the total protein concentration was fixed and the ratio of the two species allowed to vary, then this also affected the final extent of aggregation. Maximum aggregation, measured by numbers of bonds or size of clusters, was reached when the two proteins were in a specific ratio. If one or other protein was present in excess of this value, then the extent of aggregation was diminished.

It is of interest to note that the input ratio of proteins giving the greatest clustering is not identical to the final ratio of these proteins in an extended lattice. For example, the maximum bond number for the T + A system was achieved at an input T/A ratio of about 0.75. However, from simple geometry we know that the stoichiometry of an infinitely extended lattice of the same aggregate is 0.67. This difference arises because (1) the proteins form hexagons wherever they can, since these are more stable, and (2) the perimeter of a cluster formed entirely of hexagons contains equal umbers of T and A and therefore has a T/A ratio of 1.0. Hence, the stoichiometric ratio of subunits in the interior of a cluster, which is indeed 0.67, will be increased by the higher proportion of T proteins on the cluster perimeter. The magnitude of this bias will depend on the size of the cluster and the irregularity of its contour.

The reduction of aggregation noted in our study resembles the suppression of complex formation by unbalanced monomers sometimes called the ''prozone'' effect (Bray and Lay 1997). Examples include the suppression of dynactin formation by an excess of the 50 kD subunit (Escheverri et al. 1996); suppression of the formation of a yeast myosin V by an excess of either its heavy or light chains (Stevens and Davis 1998); and the inhibition of microtubule formation by an excess of

β-tubulin (Abruzzi et al. 2002). We are not aware that similar phenomena have been seen in two-dimensional membrane complexes, but the results of our simulations suggest that it should occur. In order to produce large protein clusters of defined composition, the interacting components should be present at the correct stoichiometric ratios.

Perhaps the most intriguing result arising from our simulations is that the composition of membrane clusters is also malleable. If one component is present in excess at the start of the simulation, then the clusters that form have a bias in composition toward that component. The basis for this bias is readily understood if we examine specific clusters. Clusters made with an excess of protein A tend to have molecules of A on their perimeter. This can distort the overall composition, especially if the cluster is relatively small. Since, as already noted, an excess of A also has the effect of reducing the size and number of clusters, the effect of input ratios on composition can be significant. One possible situation in which this could arise is if a membrane complex is reconstituted in vitro. A common experimental strategy is to incubate membrane fragments enriched in a specific protein with other purified proteins in order to form a functional membrane complex, such as that associated with bacterial chemotaxis (Levit et al. 2002). It is obvious from our simulations that complexes obtained by this procedure could differ significantly in composition from the actual complexes in the living cell.

Acknowledgements We are grateful to Matthew Levin for useful suggestions. Supported by NIGMS grant GM64713.

References

Abruzzi KC, Smith A, Chen W, Solomon F (2002) Protection from free β-tubulin by the β-tubulin binding protein Rb12p. Mol Cell Biol 22:138–147

Alberts B, Bray D, Hopkin K, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2004) Essential cell biology, 2nd edn. Garland, New York

Bray D, Bourret RB (1995) Computer analysis of the binding reactions leading to a transmembrane receptor-linked multiprotein complex involved in bacterial chemotaxis. Mol Biol Cell 6:1367–1380

Bray D, Lay SW (1997) Computer-based analysis of the binding steps in protein complex formation. Proc Natl Acad Sci USA 94:13493–13498

Bren A, Eisenbach M (2000) How signals are heard during bacterial chemotaxis: protein-protein interactions in sensory signal propagation. J Bacteriol 182:6865–6873

Choquet D, Triller A (2003) The role of receptor diffusion in the organization of the postsynaptic membrane. Nat Rev Neurosci 4:251–265

Davis MM, Krogsgaard M, Huppa JB, Sumen C, Purbhoo MA, Irvine DJ, Wu LC, Ehrlich L (2003) Dynamics of cell surface molecules during T cell recognition. Annu Rev Biochem 72:717–742

Escheverri CJ, Paschal BM, Vaughan KT, Vallee RB (1996) Molecular characterisation of the 50-kD subunit of dynactin reveals function for the complex in chromosome alignment and spindle organization during mitosis. J Cell Biol 132:617–633

Gil T, Ipsen JH, Mouritsen OG, Sabra MC, Sperotto MM, Zuckermann MJ (1998) Theoretical analysis of protein organization in lipid membranes. Biochim Biophys Acta 1376:245–266

Gründling A, Manson MD, Young R (2001) Holins kill without warning. Proc Natl Acad Sci USA 98:9348–9352

Israelachvili JN (1992) Intermolecular and surface forces, 2nd edn. Academic Press, London

Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. Science 220:671–680

Lamb TD (1994) Stochastic simulation of activation in the G-protein cascade of phototransduction. Biophys J 67:1439–1453

Levit MN, Grebe TW, Stock JB (2002) Organization of the receptor-kinase signaling array that regulates *Escherichia coli* chemotaxis. J Biol Chem 277:36748–36754

Liu JD, Parkinson JS (1989) Role of CheW protein in coupling membrane receptors to the intracellular signaling system of bacterial chemotaxis. Proc Natl Acad Sci USA 86:8703–8707

Maddock JR, Shapiro L (1993) Polar location of the chemoreceptor complex in the *Escherichia coli* cell. Science 259:1717–1723

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:1087–1092

Newman MEJ, Barkema GT (1999) Monte Carlo methods in statistical physics. Oxford University Press, Oxford

Papp B, Pál C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. Nature 424:194–197

Poo M-M, Cone RA (1974) Lateral diffusion of rhodopsin in the photoreceptor membrane. Nature 247:438–441

Stevens RC, Davis TN (1998) M1c1p is a light chain for the unconventional myosin Myo2p in *Saccharomyces cerevisiae*. J Cell Biol 142:711–722

Wang IN, Smith DL, Young R (2000) Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol 54:799–825