Subdiffusion of proteins and oligomers on membranes

David Lepzelter and Muhammad Zaman

Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA

(Received 13 September 2012; accepted 5 October 2012; published online 6 November 2012)

Diffusion of proteins on lipid membranes plays a central role in cell signaling processes. From a mathematical perspective, most membrane diffusion processes are explained by the Saffman-Delbrück theory. However, recent studies have suggested a major limitation in the theoretical framework, the lack of complexity in the modeled lipid membrane. Lipid domains (sometimes termed membrane rafts) are known to slow protein diffusion, but there have been no quantitative theoretical examinations of how much diffusion is slowed in a general case. We provide an overall theoretical framework for confined-domain (“corralled”) diffusion. Further, there have been multiple apparent contradictions of the basic conclusions of Saffman and Delbrück, each involving cases in which a single protein or an oligomer has multiple transmembrane regions passing through a lipid phase barrier. We present a set of corrections to the Saffman-Delbrück theory to account for these experimental observations. Our corrections are able to provide a quantitative explanation of numerous cellular signaling processes that have been considered beyond the scope of the Saffman-Delbrück theory, and may be extendable to other forms of subdiffusion. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4764305>]

INTRODUCTION

The diffusion of proteins and other particles on membranes has been of interest to scientists for several decades. It can have obvious effects on clustering of signaling molecules and reaction rates of any molecules that sit on the membrane or are associated with membrane molecules, which is a large and extraordinarily important subsection of cellular proteins and other molecules. In 1975, Saffman and Delbrück published a theory regarding membrane diffusion that has defined the field ever since.¹ Though there was a lack of direct experimental evidence, and what evidence eventually existed has not always initially seemed to bear out those predictions, the overall framework associated with the theory tends to be generally accepted. Even the most comprehensive work on membrane diffusion tends to be formulated as extensions of the overall theory,² and others serve as specific-case corrections.³ One of the most important specific corrections, that of confined (also called “corralled”) diffusion, has been examined only in extremely limited ways via analytical theory^{4,5} or in ways that ignore important edge effects;⁶ we present a much more complete theory here. However, we suggest that we can only do so properly when we consider another seemingly unrelated phenomenon.

The validity of the Saffman-Delbrück model was called into question in 2006, when Gambin and colleagues published direct diffusion measurements of proteins with reasonably well-measured size on a membrane of controllable width, finding a diffusion constant with much more significant dependence on protein radius than expected, seemingly $D \propto 1/R$ instead of $D \propto \ln 1/R$.⁷ Such results gave rise to an interesting hypothesis involving local lipid rafts in the area of the protein changing the momentum dissipation of the membrane;³ however, theoretical results have since shown that such an effect is not large enough in appropriate-sized proteins for the large-

scale change in diffusivity seen in the experiments.⁸ More, another direct measurement of multiple well-characterized proteins of different radii have found the $D \propto \ln 1/R$ expected from the Saffman-Delbrück theory.⁹ To date, the discrepancies are still not properly accounted for; even if Gambin and colleagues did not measure D versus R behavior as they believed, it is clear that they observed something of interest. This observation cannot be explained without an additional correction to the Saffman-Delbrück model.

In this article, we propose a completely different reason for the apparent $D \propto 1/R$ behavior seen in the work by Gambin and colleagues. Further, we suggest that the relevant correction to the Saffman-Delbrück theory could be important to a significant subclass of membrane proteins. We therefore add to our theory of confined diffusion a correction factor which properly accounts for the apparent behavior.

The proteins used by Gambin and colleagues had either a single hydrophobic region or seven hydrophobic regions connected by hydrophilic regions. These hydrophobic regions are referred to as transmembrane regions, resting in the lipid membrane. In the second case, the seven transmembrane regions are tethered to each other by the connecting hydrophilic regions. From the viewpoint of the standard Saffman-Delbrück model, this detail is mostly irrelevant. However, we argue that the distinctness of the transmembrane regions can be central to diffusion on a membrane that contains multiple lipid phases. (Both types of membranes used in Gambin's work have been seen to have multiple effective phases. The C₁₂E₅ surface used has been shown to have a very small island of stability for its sponge phase,¹⁰ and even when stable in its sponge phase it seems to have uneven lipid heights. The other, SOPC (1,3-stearoyl-2,3-oleoyl-phosphatidylcholine), by itself has a gel-to-liquid-crystal transition at approximately 6–10 °C, depending on amount of acyl migration,¹¹ and

FIG. 1. Stylized illustration of how a single protein can face multiple separate energy barriers in passing a lipid phase transition barrier.

the transition can broaden or shift with the inclusion of any imperfections, e.g., proteins or cholesterol. For that matter, multiple phases are common in currently used unilamellar vesicles in general,¹² and of course in actual cells. Even in membranes considered to be in one phase, small domains in other phases are not uncommon.¹³⁾

The idea that separate areas on a membrane in different lipid phases (also known as lipid rafts or membrane rafts) could partially confine membrane proteins and thus slow their diffusion is not a new one. Lipid rafts have been identified as one of the ways in which membrane proteins are corralled, giving rise to slowed diffusion on large scales (subdiffusion) in realistic membranes,¹⁴ though there has been a lack of general modeling surrounding the phenomenon. In more explicit terms, a membrane protein may be slowed by either a necessary conformation change to adjust to the size of surrounding lipids of different heights or a small semi-permanent membrane raft of lipids of the size appropriate to the protein's transmembrane region. A necessary conformation change could reasonably be represented by an energy barrier between lipid phases. A raft, in contrast, would treat other rafts of the same lipid phase as an energetically preferable environment. In either case, a protein that might otherwise cross a lipid phase boundary could face a large energy barrier and turn back.

The conformation change concept, supported by more recent work from Gambin and colleagues,¹⁵ is central to our proposal regarding multiple transmembrane regions. We suggest that each transmembrane region in a protein can undergo a separate conformation change as pictured in Fig. 1, and that therefore a protein with multiple transmembrane regions faces multiple energy barriers when it crosses a lipid phase boundary. The mathematics of this situation is identical to those giving rise to the analytical expression derived in our previous work regarding clusters of integrins and cytoskeletal barriers, which in previous publication lacked consideration for cluster geometry. Here we present our derivation of the single-transmembrane-region diffusion constant, in addition to our previously published equation without geometry and updated equations that take geometry into account. We then show that we can easily account for the apparent $D \propto 1/R$ behavior seen in experiment.

It is important to point out that, while we believe our interpretation to be correct in this instance, we offer no opinion on how many (or which) membrane proteins cross lipid phase barriers via conformation change as opposed to with an attached small membrane raft. Our later equations (after the

section labeled “Theory: Constrained Diffusion”) are clearly not valid for the second situation without significant alteration. Additionally, this entire work depends on the domains involved being semi-stable, in the sense that it is reasonable for the protein to diffuse from one side of the domain to the other in significantly less time than the domain takes to stop existing. This is not an unreasonable assumption; membrane raft lifetimes have been measured to be on the scale of milliseconds to seconds, and their sizes are on the nanometer scale.¹⁶

THEORY: CONSTRAINED DIFFUSION

Here we build upon our previous, geometry-free arguments for tethered membrane proteins or transmembrane regions passing through an energy barrier.¹⁷ The general mathematics can be used for multiple types of confined domains, not just those in our previous work and those primarily considered here, and likely could even be generalized for other forms of subdiffusion (e.g., on a protein folding energy landscape^{18,19}). However, our language in this article will focus on lipid phase regions. Consider a protein with a single transmembrane region diffusing within a confined area (in this case, a portion of the membrane in a certain phase). It will diffuse within the area with a high diffusion constant D . When it attempts to leave this area there is a high probability of hitting the energy barrier and being turned back. This will continue for some time, but eventually it will manage to get past the energy barrier into the next confined lipid area, in which it will continue. As seen in measurements with long enough time scale that the protein is not confined to a single confined area in the entire time, there will be an apparent diffusion constant $D_1 \ll D$.

We now estimate D_1 in as general a manner as possible. While D_1 will depend in part on the membrane geometry and the properties of the diffusing protein, we approximate a lipid phase region here as a square in order to make the problem solvable, and we introduce a single parameter to account for protein behavior beyond basic diffusion. First, we explore the size of the square. If its physical size is $L \times L$ and the protein's radius is R , the effective size of the square (the amount of space in which the protein can move freely without interaction with the walls) is $L' \times L' = (L - 2R) \times (L - 2R)$.

There are three relevant time scales to consider in confined-domain diffusion. Each of these scales has one component that corresponds to larger-scale diffusion assumptions and one that does not. First is the time the protein takes to diffuse to an edge of the confined domain; this is usually called “first passage time,” though the name in this case is slightly misleading. This first time scale is not the time the protein takes to pass into another domain because the protein is confined: there is some probability $(1 - \mathcal{P})$ that the protein will be blocked from passing. The time scale on which the protein actually passes through to another confined domain is the second of the three.

Both of these time scales contain components that, because of the system's geometry, favor one side of the confined domain over the other. (Specifically, they favor the edge with which the protein has most recently been in contact—the backwards direction, as it were.) The third time scale is

split into the amount of time the protein takes to fail at diffusing out of a confined domain because of these effects and the amount of time it takes to succeed. Combined properly, these components completely describe the time taken for the protein traveling from domain to domain without favoring any particular direction, corresponding to traditional diffusion assumptions. Using this amount of time as a time step Δt and the space traveled L as a spatial step, we can treat the problem as diffusion on a larger scale.

First time scale

First, we calculate the time scale on which a protein comes in contact with a lipid domain edge. Let us assume the protein has just entered the square of interest and is no longer in contact with the lipid phase boundary. (By this we mean that the protein is *fully* not in contact with the boundary, so none of it is adjacent to any lipid in the previous square; it is entirely engulfed by lipids from its current square. In a system with finite lipid size, this implies the edge of the protein is one lipid away from the phase boundary, with this distance being called x_0 .) We allow the protein to be at any point in y .

It would be tempting to use the amount of time taken to leave a lipid phase region in either the x or the y direction, as the two would seem to be intertwined in the versions of the below equations taken for motion in both directions. However, this is not necessary; passage across a barrier in the x direction has no effect on the y position of the protein. For verification purposes, our approach gives the same answer as a fully 2D simulation (data shown at the end of this section).

Using Fourier series analysis on the position distribution $\phi(x, t = 0) = \delta(x - x_0)$ with absorbing (Dirichlet) boundaries, we find

$$\phi(x, t = 0) = \frac{2}{L'} \sum_{m=1}^{\infty} \sin \frac{\pi m x_0}{L'} \sin \frac{\pi m x}{L'},$$

which evolves in time via the diffusion equation $\frac{\partial \phi}{\partial t} = D \nabla^2 \phi$ to yield

$$\phi(x, t) = \frac{2}{L'} \sum_{m=1}^{\infty} \sin \frac{\pi m x_0}{L'} \sin \frac{\pi m x}{L'} e^{-\frac{D \pi^2 m^2}{L'^2} t}. \quad (1)$$

The mean first passage time (average amount of time the protein takes to diffuse to an edge) is therefore

$$\begin{aligned} t_1^{\text{tot}} &= - \int_0^{\infty} dt \int_0^{L'} dx t \frac{\partial \phi}{\partial t} \\ &= \int_0^{\infty} dt \frac{4L'}{\pi^3 D} \sum_{m \text{ odd}} \frac{1}{m^3} \sin \left(\frac{\pi m x_0}{L'} \right) \\ &= t_1 = \int_0^{\infty} dt \frac{4L'}{\pi^3 D} \left[\sum_{m=1}^{\infty} \frac{1}{m^3} \sin \left(\frac{\pi m x_0}{L'} \right) \right. \\ &\quad \left. - \sum_{m \text{ even}} \frac{1}{m^3} \sin \left(\frac{\pi m x_0}{L'} \right) \right]. \end{aligned}$$

Using the Fourier series for the third Bernoulli polynomial,

$$t_1^{\text{tot}} = \frac{L'^2 x_0}{2D L'} \left(1 - \frac{x_0}{L'} \right). \quad (2)$$

Again, however, this is not the same as the mean first passage time in most diffusion problems for two reasons. First, one side is favored over another. Second, the protein experiences confinement. Confinement, which we will discuss further later, simply means there is some probability that the protein will actually pass through the edge of the modeled area, and some probability it will not. Weighting (favoring of one side over another) is a more complicated issue.

In which direction does the protein diffuse out of the confined domain? It would be easy to assume that either direction is equally likely; after all, normal diffusion assumes that motion in the $\hat{x}$ direction is the same as motion in the $-\hat{x}$ direction. However, while the local diffusion of the protein itself is unbiased, the time taken to leave the domain is dependent on x_0 because there is a greater chance of the protein diffusing in the $-\hat{x}$ direction if x_0 is small. We can calculate how much more likely this is, using $P(\text{contact}) \propto \hat{n} \cdot \nabla \phi$, where $\hat{n}$ is the unit vector pointing inward from the boundary in question.

$$P(\text{contact}) \propto \int dt (\hat{n} \cdot \nabla) \frac{2}{L'} \sum_{m=1}^{\infty} \sin \frac{\pi m x_0}{L'} \sin \frac{\pi m x}{L'} e^{-\frac{D \pi^2 m^2}{L'^2} t}. \quad (3)$$

We are interested in the value of this quantity at $x = 0$ and $x = L'$, simplifying the equation greatly. The $\hat{n} \cdot \nabla$ term is, in this case, either $\frac{d}{dx}$ at $x = 0$ or $-\frac{d}{dx}$ at $x = L'$.

$$P(\text{contact}) \propto \begin{cases} \frac{2}{\pi D} \sum_{m=1}^{\infty} \frac{1}{m} \sin \frac{\pi m x_0}{L'}, & x = 0, \\ -\frac{2}{\pi D} \sum_{m=1}^{\infty} \frac{(-1)^m}{m} \sin \frac{\pi m x_0}{L'}, & x = L'. \end{cases}$$

With some tinkering and the Fourier series for the first Bernoulli polynomial this yields

$$P(\text{contact}) \propto \begin{cases} \frac{1}{D} - \frac{x_0}{DL'}, & x = 0, \\ \frac{x_0}{DL'}, & x = L'. \end{cases}$$

As the two probabilities must total to 1, the appropriate proportionality constant is simply D .

Clearly, the system's geometry has caused the protein to favor the edge of the domain with which it has had recent contact. However, we can separate this weighted motion from the more traditional unweighted motion, as illustrated in Figs. 2 and 3: if we add to the probability of forward passage an exact image of itself, taken from the probability of passage at $x = 0$, we find a probability of unweighted motion. The remaining probability is then that of weighted motion. Thus

$$P_1^{\text{unweighted}} = 2 \frac{x_0}{L'}, \text{ and} \quad (4)$$

$$P_1^{\text{weighted}} = 1 - 2 \frac{x_0}{L'}. \quad (5)$$

Alternatively, we could picture the probability of unweighted motion to be the probability that the protein diffuses to the center of the lipid phase region before coming into contact with an edge. The two are equivalent at this point; indeed,

FIG. 2. Illustration of weighted (center) and unweighted (right) contact with an edge immediately following a barrier crossing (left).

the way in which the protein stops favoring one edge over another is by diffusing to the center of the domain. However, future calculations will be simpler if we view the system as per Fig. 3.

Again clearly from the example in that figure, the time the system takes to go through weighted motion is not the same as that needed for unweighted motion. This can be calculated:

$$t_1^{\text{unweighted}} = t_{\text{pass}}^{x=L'} = -\frac{D}{P_{\text{pass}}^{x=L'}} \int_0^\infty dt t \frac{2}{L'} \sum_{m=1}^\infty \frac{d}{dx} \sin \frac{\pi m x}{L'} \times \sin \frac{\pi m x_0}{L'} e^{-\frac{D\pi^2 m^2}{L'^2} t} \Big|_{x=L'}, \quad (6)$$

$$t_1^{\text{weighted}} = \frac{D}{P_1^{\text{weighted}}} \left(\int_0^\infty dt t \frac{2}{L'} \sum_{m=1}^\infty \frac{d}{dx} \sin \frac{\pi m x}{L'} \times \sin \frac{\pi m x_0}{L'} e^{-\frac{D\pi^2 m^2}{L'^2} t} \Big|_{x=0} - P_{\text{pass}}^{x=L'} t_{\text{pass}}^{x=L'} \right). \quad (7)$$

With some substitutions these become

$$t_1^{\text{unweighted}} = \frac{L'^2}{6D} \left(1 - \left(\frac{x_0}{L'} \right)^2 \right), \quad (8)$$

$$t_1^{\text{weighted}} = \frac{L'^2}{6D} \frac{x_0}{L'} \left(1 - \frac{x_0}{L'} \right). \quad (9)$$

Of course, these must total to the appropriate amount, from (2):

$$t_1^{\text{tot}} = P_1^{\text{weighted}} t_1^{\text{weighted}} + P_1^{\text{unweighted}} t_1^{\text{unweighted}} = \frac{L'^2}{2D} \frac{x_0}{L'} \left(1 - \frac{x_0}{L'} \right).$$

Second time scale

We now calculate the time necessary to pass into another lipid domain by approaching the second complication mentioned in the previous calculation. For historical reasons, the calculated times in (8) and (9) are called “first passage times,” but they are not truly times of passage out of the confined area. Instead, they are times to contact with the boundary of the confined area, the energy barrier. Once a protein comes into contact with a barrier, it may be reflected or may manage to get through.

Mathematically, we can call the probability of successful passage across that boundary $\mathcal{P}$. (The precise value of $\mathcal{P}$ should depend on, among other things, lipid phase composition and protein structure. In cases where lipid domains are not responsible for the trapping, $\mathcal{P}$ should of course depend on whatever relevant physical parameters are involved.) We can combine this number with (4) to find the probability that the protein will diffuse to an edge of the domain, regardless of favor from domain geometry, and then cross the boundary: $2\mathcal{P} \frac{x_0}{L'}$.

FIG. 3. Example of how one can find unweighted diffusion to an edge and separate it from the weighted diffusion to an edge. The probability of unweighted diffusion is the probability of diffusing “forward,” plus an equivalent piece of the probability of diffusing “backward.” The probability of weighted diffusion is what remains of the “backward” probability.

Immediately after this crossing, though, the protein's position is again precisely x_0 away from the domain edge. This means that the probability of the protein immediately passing back across the just-crossed barrier is $\mathcal{P}(1 - \frac{x_0}{L'})$, and the chances of it doing so in a weighted way are $\mathcal{P}(1 - 2\frac{x_0}{L'})$. Such a calculation is incomplete, however. What of the remaining $(1 - \mathcal{P})(1 - 2\frac{x_0}{L'})$ probability, in which the protein has diffused back to the previously-crossed edge in a weighted way, but failed to cross the boundary again? In that case, the protein is *again* x_0 away from the same domain edge, and once again faces a $\mathcal{P}(1 - 2\frac{x_0}{L'})$ probability of weighted passage across the just-crossed barrier.

If we follow the probabilities accordingly and continue down this line of reasoning, the probability of weighted passage due to favoring the backwards direction is

$$P_2^{\text{weighted}} = \mathcal{P}P_1^{\text{weighted}} + (1 - \mathcal{P})P_1^{\text{weighted}}\mathcal{P}P_1^{\text{weighted}} + (1 - \mathcal{P})^2P_1^{\text{weighted}^2}\mathcal{P}P_1^{\text{weighted}} + \dots,$$

$$P_2^{\text{weighted}} = \mathcal{P}P_1^{\text{weighted}} \sum_{n=0}^{\infty} (1 - \mathcal{P})^n P_1^{\text{weighted}^n},$$

which is simply a geometric series. Thus

$$P_2^{\text{weighted}} = \frac{\mathcal{P}P_1^{\text{weighted}}}{1 - (1 - \mathcal{P})P_1^{\text{weighted}}} = \frac{\mathcal{P}(1 - 2\frac{x_0}{L'})}{1 - (1 - \mathcal{P})(1 - 2\frac{x_0}{L'})}. \quad (10)$$

Notice that the probability of unweighted contact with an edge, even followed by an unsuccessful passage, is not included in the probability of weighted passage across the edge. The reason for this is simple. Consider a protein which has diffused to an edge without favoring either side and then failed to pass through the barrier (probability $2\frac{x_0}{L'}(1 - \mathcal{P})$). At this point, there is an equal probability of the protein being adjacent to the forward or the backward edge. Future weighted passage would equally favor either edge (see Fig. 4). As equal favor in this case is equivalent to no favor at all, a single unweighted diffusion to an edge is sufficient to remove directional favoring from the system.

The probability of unweighted passage, therefore, is

$$P_2^{\text{unweighted}} = P_1^{\text{unweighted}} + (1 - \mathcal{P})P_1^{\text{weighted}}P_1^{\text{unweighted}} + (1 - \mathcal{P})^2P_1^{\text{weighted}^2}P_1^{\text{unweighted}},$$

$$P_2^{\text{unweighted}} = P_2^{\text{unweighted}} = P_1^{\text{unweighted}} \sum_{n=0}^{\infty} (1 - \mathcal{P})^n P_1^{\text{weighted}^n},$$

yielding

$$P_2^{\text{unweighted}} = \frac{2\frac{x_0}{L'}}{1 - (1 - \mathcal{P})(1 - 2\frac{x_0}{L'})}. \quad (11)$$

As should be expected, $P_2^{\text{unweighted}} + P_2^{\text{weighted}} = 1$.

The time taken for either situation is only slightly more complicated. We begin with weighted passage. Keeping track of the probabilities involved, the time taken for n weighted

FIG. 4. Immediately after unweighted motion to an edge (left), further weighted motion favors *both* directions (right). Therefore, at this point favored directions are removed.

contacts is nt_1^{weighted} , so

$$P_2^{\text{weighted}} t_2^{\text{weighted}} = \mathcal{P}P_1^{\text{weighted}} t_1^{\text{weighted}} + (1 - \mathcal{P})P_1^{\text{weighted}}\mathcal{P}P_1^{\text{weighted}} \cdot 2t_1^{\text{weighted}} + (1 - \mathcal{P})^2P_1^{\text{weighted}^2}\mathcal{P}P_1^{\text{weighted}} \cdot 3t_1^{\text{weighted}} + \dots, \\ = \frac{\mathcal{P}t_1^{\text{weighted}}}{1 - \mathcal{P}} \sum_{n=1}^{\infty} n(1 - \mathcal{P})^n P_1^{\text{weighted}^n}.$$

This series yields a time per improper passage of

$$t_2^{\text{weighted}} = \frac{t_1^{\text{weighted}}}{1 - (1 - \mathcal{P})P_1^{\text{weighted}}} \\ = \frac{L'^2 x_0}{6D L'} \frac{1 - \frac{x_0}{L'}}{1 - (1 - \mathcal{P})(1 - 2\frac{x_0}{L'})}. \quad (12)$$

Time taken for unweighted passage is most easily done in two parts: the time taken to have diffusion to an edge independent of favor (thus guaranteeing eventual unbiased passage as per our above arguments) and the time taken after an unweighted diffusion to an edge to pass through a boundary. The first of these is given by

$$P_2^{\text{unweighted}} t_2^{\text{unweighted}-1} = P_1^{\text{unweighted}} t_1^{\text{unweighted}} + (1 - \mathcal{P})P_1^{\text{weighted}}P_1^{\text{unweighted}} \\ \times (t_1^{\text{weighted}} + t_1^{\text{unweighted}}) + (1 - \mathcal{P})^2P_1^{\text{weighted}^2}P_1^{\text{unweighted}} \\ \times (2t_1^{\text{weighted}} + t_1^{\text{unweighted}}) + \dots, \\ t_2^{\text{unweighted}-1} = t_1^{\text{unweighted}} + t_1^{\text{weighted}} \frac{(1 - \mathcal{P})(1 - 2\frac{x_0}{L'})}{1 - (1 - \mathcal{P})(1 - 2\frac{x_0}{L'})}.$$

The second is

$$t_2^{\text{unweighted}-2} = \mathcal{P} \cdot 0 + (1 - \mathcal{P})\mathcal{P} \cdot t_1 + (1 - \mathcal{P})^2\mathcal{P} \cdot 2t_1 + \dots \\ = \mathcal{P} \frac{1 - \mathcal{P}}{(1 - (1 - \mathcal{P}))^2} t_1^{\text{tot}} \\ = \frac{1 - \mathcal{P}}{\mathcal{P}} t_1.$$

So, in total, the time taken for proper passage, per proper passage, is

$$\begin{aligned} t_2^{\text{unweighted}} &= t_2^{\text{unweighted}-1} + t_2^{\text{unweighted}-2} \\ &= t_1^{\text{unweighted}} + t_1^{\text{weighted}} \frac{(1 - \mathcal{P})(1 - 2\frac{x_0}{L'})}{1 - (1 - \mathcal{P})(1 - 2\frac{x_0}{L'})} \\ &\quad + \frac{1 - \mathcal{P}}{\mathcal{P}} t_1^{\text{tot}}. \end{aligned} \quad (13)$$

Third time scale

Now we are prepared to discuss the total amount of time the system takes to travel properly from one domain to another. The central assumption one makes in the larger-scale diffusion system is that the protein passes from one confined domain to another without favoring any direction, and then does so again from its new domain, repeating the process for as long as it can. We must therefore find the time scale on which this happens: the time it takes for a protein to go from a single passage from one domain, without favoring any direction, to another passage from a neighboring domain without favoring any direction.

This requires consideration, illustrated by Fig. 5. Weighted passage is more common than unweighted passage, and if a protein successfully passes a barrier favoring the backwards direction after an unweighted passage it returns to the previous domain, invalidating the unweighted passage. If, though, it immediately passes favoring the backwards direction *again*, it again crosses over the barrier with which the protein has had most recent contact, and so returns to the new domain. Thus, an odd number of weighted passages in a row essentially cancel out an unweighted passage, but zero or an even number of weighted passages do not. If unweighted passage is cancelled, the protein must start over completely. Therefore, total time taken is given by

$$\begin{aligned} t_{\text{pass}}^{\text{tot}} &= t_2^{\text{unweighted}} + P_2^{\text{weighted}} [P_2^{\text{unweighted}} (t_2^{\text{weighted}} + t_{\text{pass}}^{\text{tot}}) \\ &\quad + P_2^{\text{weighted}} [P_2^{\text{unweighted}} (2t_2^{\text{weighted}}) \\ &\quad + P_2^{\text{weighted}} [P_2^{\text{unweighted}} (3t_2^{\text{weighted}} + t_{\text{pass}}^{\text{tot}}) \\ &\quad + P_2^{\text{weighted}} [P_2^{\text{unweighted}} (4t_2^{\text{weighted}}) + P_2^{\text{weighted}} [\dots \end{aligned}$$

(Note that $t_{\text{pass}}^{\text{tot}}$ is on *both* sides of this equation.),

$$\begin{aligned} t_{\text{pass}}^{\text{tot}} &= t_2^{\text{unweighted}} + P_2^{\text{weighted}} P_2^{\text{unweighted}} (t_2^{\text{weighted}} + t_{\text{pass}}^{\text{tot}}) \\ &\quad + P_2^{\text{weighted}^2} P_2^{\text{unweighted}} (2t_2^{\text{weighted}}) \\ &\quad + P_2^{\text{weighted}^3} P_2^{\text{unweighted}} (3t_2^{\text{weighted}} + t_{\text{pass}}^{\text{tot}}) \\ &\quad + P_2^{\text{weighted}^4} P_2^{\text{unweighted}} (4t_2^{\text{weighted}}) + \dots \\ &= t_2^{\text{unweighted}} + \sum_{n=1}^{\infty} P_2^{\text{weighted}^n} P_2^{\text{unweighted}} \cdot n t_2^{\text{weighted}} \\ &\quad + P_2^{\text{weighted}^{2n-1}} P_2^{\text{unweighted}} \cdot t_{\text{pass}}^{\text{tot}}, \end{aligned}$$

FIG. 5. An even number of improper passages after a proper passage only adds those passage times to the total amount of time taken; an odd number puts the protein back in its original domain, effectively forcing the protein to start over.

$$\begin{aligned} t_{\text{pass}}^{\text{tot}} &= t_2^{\text{unweighted}} + P_2^{\text{unweighted}} \frac{P_2^{\text{weighted}}}{(1 - P_2^{\text{weighted}})^2} t_2^{\text{weighted}} \\ &\quad + P_2^{\text{unweighted}} \frac{P_2^{\text{weighted}}}{1 - P_2^{\text{weighted}^2}} t_{\text{pass}}^{\text{tot}} \\ t_{\text{pass}}^{\text{tot}} &= (1 + P_2^{\text{weighted}}) \left(t_2^{\text{unweighted}} + t_2^{\text{weighted}} \frac{P_2^{\text{weighted}}}{P_2^{\text{unweighted}}} \right). \end{aligned} \quad (14)$$

Combining Eqs. (12)–(14), and simplifying the expressions, we find

$$t_{\text{pass}}^{\text{tot}} = \frac{L'^2}{2D} \left(\frac{(\mathcal{P}(1 - 2\frac{x_0}{L'}) + \frac{x_0}{L'})(1 - \frac{x_0}{L'})}{\mathcal{P}} \right). \quad (15)$$

It is also possible to add a term accounting for the time the protein actually takes to cross the barrier; in most cases, this should be insignificant, but a simple modification to the above mathematics should account for it.

From this equation, we can find an effective diffusion constant: $D_1 = \frac{L^2}{2t_{\text{pass}}^{\text{tot}}}$, using the actual size of the domain L instead of the reduced size L' . (This uses the equation for one-dimensional diffusion, as all the motion we have calculated is in one dimension. The protein is, of course, diffusing in two dimensions, for which we could account by dividing both $t_{\text{pass}}^{\text{tot}}$ and the right-hand side of the equation for D_1 by 2. The result is mathematically identical, so we use the slightly simpler 1D version.)

$$D_1 = \frac{DL^2}{L'^2} \frac{\mathcal{P}}{(\mathcal{P}(1 - 2\frac{x_0}{L'}) + \frac{x_0}{L'})(1 - \frac{x_0}{L'})}. \quad (16)$$

We can simplify Eqs. (15) and (16) for some of the problem's extrema. For free diffusion, we can assume L is large compared to both R and x_0 , and $\mathcal{P} = 1$. (Alternatively, we could assume larger x_0 and R , adjusting $\mathcal{P}$ and adding an extra time term to account for passage across the “barrier” that is an empty membrane between arbitrarily defined squares. This may be more realistic in some sense, but it is also needlessly

FIG. 6. (Left) Simulated diffusion on a two-dimensional membrane with square lipid domains (corrals) of size L with $D = 0.5 \text{ nm}^2$ per time unit, and with probability of passage across a barrier $\mathcal{P}$, graphed in blue against our analytical result in red. (Right) Percent difference between simulation results and analytical prediction, including sign of difference.

complicated.) We then find

$$t_{\text{pass}}^{\text{tot}} = \frac{L^2}{2D}, \quad D_1 = D, \quad (17)$$

recovering proper diffusion.

For highly confined diffusion, we have $\mathcal{P} \ll 1$, which yields

$$t_{\text{pass}}^{\text{tot}} \approx \frac{L^2 \frac{x_0}{L} (1 - \frac{x_0}{L})}{2D \mathcal{P}}, \quad D_1 \approx \frac{L^2}{L^2 \frac{x_0}{L} (1 - \frac{x_0}{L})} \frac{D\mathcal{P}}{D} \propto D\mathcal{P}. \quad (18)$$

The strong dependence on $\frac{x_0}{L}$ may be considered a surprising result, but it does follow logically from the system; after unweighted diffusion to an edge, further weighted motion is enough to retry a low- $\mathcal{P}$ passage as many times as necessary, and a small $\frac{x_0}{L}$ makes this occur quickly. The effective diffusion constant is proportional to D and $\mathcal{P}$, as we should expect.

To further verify our mathematics, we have programmed a small Markov Chain Monte Carlo simulation in C, in which a protein diffuses on a grid (grid size 1 nm) in x and y directions. Without a barrier in the way, it travels one grid point in every time step in a random direction in both x and y ($D = 0.5 \text{ nm}^2/\text{time unit}$). We placed a barrier at every L grid points in either direction, with probability of passage $\mathcal{P}$. The results averaged over 1000 simulations are shown graphed against our analytical results in Fig. 6, including percent differences. The small and unpatterned disagreement between the two implies that the differences are caused by the finite number of simulations run, meaning that our analytical results accurately represent our proposed system.

THEORY: LINEAR PROTEIN

We summarize and then expand upon some of our previous work here, which relied on the intuitive Eq. (18) with simple modifications and estimated $\mathcal{P}$ for one protein or oligomer given that of a simpler protein. Consider, now, a protein with two transmembrane regions connected by a hydrophilic region. Instead of hitting a barrier and either being reflected or entirely passing through it, this protein will eventually have one of its transmembrane regions undergo a conformation

change and get through the barrier. At this point, it will be stuck on the boundary between lipid phases, until either its other transmembrane region undergoes the same conformation change and joins the first in the new lipid phase region or the first transmembrane region reverts back to its original conformation. In the first case, of course, the protein will have successfully crossed the phase boundary; in the second, it will have been reflected. What are the chances of either situation, once one transmembrane region is on one side of the boundary and the other on the other side? Assuming there is no significant energy difference in the transmembrane region configurations, the chance of either is 1/2.

We can expand this approach to account for n separate, sequential transmembrane regions. In this case, the phase change (energy barrier) will be between one transmembrane region and another until the entire protein is on one side of it or another. The chance of going forward is 1/2, as is the chance of going backward. An easy representation for the evolution of the system is then an $n + 1$ by $n + 1$ matrix H , for which we give an example at $n = 4$:

$$\begin{pmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \end{pmatrix}.$$

We can name these rows: row 1, all transmembrane regions on the original side of the barrier. Row 2, one transmembrane region on the opposite side. Row 3, two transmembrane regions on the opposite side, etc., to row 5, where all transmembrane regions have crossed. The probability that the entire protein ends up crossing the energy barrier is represented by an element of the matrix $P = H + H^2 + H^3 + H^4 + \dots$, specifically $P(n + 1, 2)$. This probability ends up simply being equal to $1/n$, meaning that $\mathcal{P}$ has decreased by a factor of n compared to that of a single-transmembrane-region protein. If $\mathcal{P}$ for both is small and the time the protein spends trapped straddling the transition is negligible compared to the diffusion time, this yields an effective diffusion constant

$D_{\text{eff}} = D_1/n$, with D_1 being the derived D_{eff} with a single transmembrane region from (15).

Slightly more complex is the calculation of how much time the protein spends trapped on the edge between the two lipid phases. We can reasonably expect it to be fairly small. Again following our previous work, with t_s the time taken for a successful pass through and t_f for a failed attempt, the total time taken is $t_s + (n-1)t_f$. Defining the amount of time it takes a single transmembrane region to cross a barrier as τ , $t_s = \frac{\tau}{P(n+1,2)} \sum_{i=1}^{\infty} i \vec{e}(n+1) \cdot H^i \vec{e}(2)$ where $\vec{e}(j)$ is the j th unit vector in the space of the matrix H . In this case, $\vec{e}(n+1) \cdot H^i \vec{e}(2)$ is simply the $(n+1, 2)$ element of the matrix H^i . Likewise, $t_f = \frac{\tau}{P(n+1,2)} \sum_{i=1}^{\infty} i \vec{e}(1) \cdot H^i \vec{e}(2)$. When calculated, $t_s + (n-1)t_f = \tau(n^2 - n)$. Using τ and an area characteristic to the system, we can define a second “diffusion constant” D_2 which makes the overall effective diffusion constant

$$D_{\text{eff}}^n = \frac{D_1}{n} \frac{1}{\left(\frac{D_1}{D_2}\right)(n-1)+1}.$$

Depending on how much larger D_2 is than D_1 , this expression can range from a $1/n$ behavior to a $1/n^2$ behavior. However, as mentioned previously, this expression lacks proper accounting for geometry. More explicitly, it assumes the protein’s transmembrane regions are perfectly and rigidly linear as the protein passes through (or fails to pass through) the lipid phase barrier. This is clearly unrealistic. We now present two more reasonable protein configurations, both of which have not been published previously.

THEORY: NONLINEAR PROTEINS

The first of these configurations is fairly rigid but not linear. We assume an essentially random configuration (transmembrane regions not lined up as grid points), and point out that in this case geometric constraints make it difficult though not impossible for more or fewer than two transmembrane regions to be adjacent to a straight-line lipid phase boundary. In this case, we simplify the case slightly by saying there are always two transmembrane regions adjacent to the boundary on either side unless only one transmembrane region exists on a side of the boundary. The chances of going through the boundary from, e.g., the L_o phase to the L_α phase are then $2/3$ if there is only one transmembrane region on the L_o side of the boundary, $2/4$ or $1/2$ if at least two transmembrane regions are on both sides, and $1/3$ if there is only one transmembrane region on the L_o side of the lipid phase transition. In this case, the evolution of the system is represented by, again for $n = 4$,

$$\begin{pmatrix} 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 2/3 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 0 \end{pmatrix}.$$

Using this version of H , we do precisely the same calculation as before, getting this time

$$D_{\text{eff}}^n = \frac{D_1}{\frac{n}{2} + 1} \frac{1}{\left(\frac{D_1}{D_2}\right)(2n-3)+1}.$$

Next, we loosen the constraints further, and assume the protein is somewhat flexible though not so flexible that it resembles a Gaussian chain or similar configuration. In this case, it is not unreasonable to expect consistently three transmembrane regions to be approximately adjacent to the boundary. Similarly to before, the chances of going through the boundary from the L_o phase to the L_α phase are $3/4$ if there is only one transmembrane region on the L_o side of the boundary, $2/5$ or if there are two transmembrane regions on the L_o side, $3/6$ or $1/2$ if at least three transmembrane regions are on both sides, and so forth.

In this case, the geometry is essentially invalid until at least $n = 5$, with a restricted version valid at $n = 4$. Here we present the case of $n = 6$.

$$\begin{pmatrix} 0 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2/5 & 0 & 0 & 0 & 0 \\ 0 & 3/4 & 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 3/4 & 0 \\ 0 & 0 & 0 & 0 & 2/5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/4 & 0 \end{pmatrix}$$

and

$$D_{\text{eff}}^n = \frac{D_1}{\frac{2n+16}{9}} \frac{1}{\left(\frac{D_1}{D_2}\right)(9n/2-11)+1}.$$

Clearly, all three of these equations are approximations; a real protein geometry might be expected to come somewhere in between them, presumably between the second and third in most cases. Regardless, the overall picture is fairly clear: if the time a protein spends trapped on the phase boundary is insignificant ($D_1/D_2 \ll 1$), its diffusion constant should be decreased from its Saffman-Delbrück value by a factor linear in the number of transmembrane regions n (Fig. 7), and if not the constant should be reduced by a factor quadratic in n (Fig. 8).

Additionally, all of these equations are valid for oligomers as well as single proteins of the appropriate geometries.

COMPARISON TO EXPERIMENTS

As our derivation of D_1 for a single-transmembrane-region protein depends on parameters which are difficult to measure, we know of no reasonable test for it between the obvious extremes of high and low $\mathcal{P}$. In either of these cases, however, our derived D_1 corresponds well to observed behavior (true diffusion with the correct diffusion constant for high $\mathcal{P}$ and the obvious $D\mathcal{P}$ subdiffusion for low $\mathcal{P}$). We do predict that, in cases where $\mathcal{P}$ is small and the size of the lipid domain region can be altered without changing $\mathcal{P}$ or D , the

FIG. 7. Ratio of diffusion constant with n transmembrane regions to diffusion constant with only one transmembrane region, assuming $(D_1/D_2 \ll 1)$. Note that, since the radius of a protein might be expected to be proportional to some power of n roughly between 0.5 and 1, this would seem to indicate diffusion that is not in agreement with the Saffman-Delbrück model unless the relevant trapping (subdiffusion) behavior is seen.

size of the domain should have an effect on the total diffusion constant very roughly proportional to L .

Additionally, there have been several tests of diffusion with multiple transmembrane regions as explored in our later sections. The case of $n = 7$ has relevance to the work by Gambin and colleagues. As previously mentioned, their proteins had one and seven transmembrane regions, and an approximate diffusion constant ratio of 3.8. Our derived relationships between the single-transmembrane-region diffusion constant D_1 and the seven-transmembrane-region diffusion constant D_{eff}^7 give 4.5 for a rigid nonlinear geometry and 3.3 for a flexible one, assuming insignificant time spent on the boundary. (If we do not assume $D_1/D_2 \ll 1$, we can find a precise value of D_1/D_2 to give a value of 3.8 for a flexible protein. However, it is unclear that this is necessary, since Gambin and colleagues gave no uncertainty in their measurements and uncertainties in ratios of diffusion constants tend to be high enough that either theoretical value is potentially reasonable.)

Though it is often difficult to extract sufficient detail from a scientific article to test a hypothesis mostly unrelated to the

FIG. 8. Ratio of diffusion constant with n transmembrane regions to diffusion constant with only one transmembrane region, with $D_1/D_2 = 0.1$. For this very high value of D_1/D_2 , the protein clearly makes up for a higher transition probability for more complicated geometries by taking more time in the transition.

original article's purpose, a brief review of the literature has shown two other experiments which seem quite suggestive of our hypothesized behavior. Single molecule tracking of amyloid oligomers and non-amyloid oligomers has shown that amyloid oligomers tend to diffuse more slowly on cell surfaces, and that both show clear signs of subdiffusion.²⁰ As an amyloid oligomer would likely be a stiffer, perhaps even linear, collection of transmembrane regions, this agrees with our model. Additionally, in another study dimerization of a protein on the endoplasmic reticulum (which does have lipid phase changes²¹) seemed to be related to a drop in diffusion constant by (within error bars) a factor of 2 in multiple cases,²² which would agree with our predicted result for $n = 2$. Though this was confused by a protein variant which was faster as a monomer and did not slow upon dimerization, our model would still seem to be relevant to the experiment; the non-responsive variant may have needed a particularly simple conformation shift from region to region, and thus encountered an insignificant energy barrier ($\mathcal{P} \sim 1$).

CONCLUSIONS

In summary, we have developed a general theoretical framework to describe confined (corrallied) diffusion, spanning the extremes from fully free to highly confined, in the equation

$$D_1 = \frac{DL^2}{L'^2} \frac{\mathcal{P}}{(\mathcal{P}(1 - 2\frac{x_0}{L}) + \frac{x_0}{L})(1 - \frac{x_0}{L})}.$$

Though we used a specific geometry, we do expect that other realistic geometries should give similar results. As a number of proteins should be expected to fall into the middle range between free and highly confined, an equation describing that middle in addition to the extremes should be useful. We have also described how, if a protein or oligomer passing through a lipid phase barrier undergoes a conformation change each time one of its transmembrane regions passes through that barrier, it is slowed via a reduction in $\mathcal{P}$ by a factor that is effectively linear in n (possibly quadratic with high n or very low D_2), the number of distinct transmembrane regions that it contains. Our derived equations provide solutions to the apparent discrepancies with the Saffman-Delbrück model presented by Gambin and colleagues in 2006, who saw what seemed to be $D \propto 1/R$ diffusion behavior. We note that this portion of the effect does not depend strictly on the size of a region of lipid in a given phase, though it assumes this size is significantly larger than the diameter of the protein or oligomer and smaller than the distance measured for the sake of the diffusion measurement (as the effect relies on the slowed apparent diffusion behavior inherent in subdiffusion due to lipid phase barriers). Importantly, we predict stronger subdiffusion behavior in proteins with multiple distinct transmembrane regions than in those with only a single transmembrane region. Our derived equations for this slowing include the two most reasonable protein conformations, and show agreement with other experiments as well.

ACKNOWLEDGMENTS

D.L. would like to thank Barry Hughes for helpful suggestions. The authors acknowledge the support of Grants 1R01CA132633 and 1RC2CA147925 to M.Z. D.L. is supported, in part, by NIH NCI R25-CA153955 and Boston University Cross-Disciplinary Training in Nanotechnology for Cancer.

- ¹P. G. Saffman and M. Delbrück, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3111 (1975).
- ²B. D. Hughes, B. A. Pailthorpe, and L. R. White, *J. Fluid Mech.* **110**, 349 (1981).
- ³A. Naji, A. J. Levine, and P. Pincus, *Biophys. J.* **93**, L49 (2007).
- ⁴J. C. Owicki and H. M. McConnell, *Biophys. J.* **30**, 383 (1980).
- ⁵E. D. Siggia, J. Lippincott-Schwartz, and S. Bekiranov, *Biophys. J.* **79**, 1761 (2000).
- ⁶H. A. Soula, A. Coulon, and G. Beslon, *BMC Biophys.* **5**, 6 (2012).
- ⁷Y. Gambin, R. Lopez-Esparza, M. Reffay, E. Sieracki, N. S. Gov, M. Genest, R. S. Hodges, and W. Urbach, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2098 (2006).
- ⁸G. Guigas and M. Weiss, *Biophys. J.* **95**, L25 (2008).
- ⁹S. Ramadurai, A. Holt, V. Krasnikov, G. van den Bogaart, J. A. Killian, and B. Poolman, *J. Am. Chem. Soc.* **131**, 12650 (2009).
- ¹⁰T. D. Le, U. Olsson, H. Wennerström, and P. Schurtenberger, *Phys. Rev. E* **60**, 4300 (1999).
- ¹¹P. J. Davis, B. D. Fleming, K. P. Coolbear, and K. M. W. Keough, *Biochemistry* **20**, 3633 (1981).
- ¹²N. Kahya, D. Scherfeld, K. Bacia, B. Poolman, and P. Schwille, *J. Biol. Chem.* **278**, 28109 (2003).
- ¹³S. Tristram-Nagle and J. F. Nagle, *Chem. Phys. Lipids* **127**, 3 (2004).
- ¹⁴G. J. Schlitz, G. Kada, V. P. Pastushenko, and H. Schindler, *EMBO J.* **19**, 892 (2000).
- ¹⁵Y. Gambin, M. Reffay, E. Sieracki, F. Homble, R. S. Hodges, N. S. Gov, N. Taulier, and W. Urbach, *J. Phys. Chem. B* **114**, 3559 (2010).
- ¹⁶K. Simons and J. L. Sampaio, *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
- ¹⁷D. Lepzelter and M. H. Zaman, *Biophys. J.* **99**, L106 (2010).
- ¹⁸J. Chahine, R. J. Oliveira, V. B. P. Leite, and J. Wang, *Proc. Natl. Acad. Sci.* **104**, 14646 (2007).
- ¹⁹R. J. Oliveira, P. C. Whitford, J. Chahine, V. B. Leite, and J. Wang, *Methods* **52**, 91 (2010).
- ²⁰M. Calamai and F. S. Pavone, *J. Am. Chem. Soc.* **133**, 12001 (2011).
- ²¹G. v. Meer, D. R. Voelker, and G. W. Feigenson, *Nat. Rev. Mol. Cell Biol.* **9**, 112 (2008).
- ²²E. D. Covington, M. M. Wu, and R. S. Lewis, *Mol. Biol. Cell* **21**, 1897 (2010).