New Method for Modeling Connective-Tissue Cell Migration: Improved Accuracy on Motility Parameters

Matt J. Kipper,*† Hynda K. Kleinman,* and Francis W. Wang†
*National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland; and †National Institute of Standards and Technology, Gaithersburg, Maryland

ABSTRACT Mathematical models of cell migration based on persistent random walks have been successfully applied to describe the motility of several cell types. However, the migration of slowly moving connective-tissue cells, such as fibroblasts, is difficult to observe experimentally and difficult to describe theoretically. We identify two primary sources of this difficulty. First, cells such as fibroblasts tend to migrate slowly and change shape during migration. This makes accurate determination of cell position difficult. Second, the cell population is considerably heterogeneous with respect to cell speed. Here we develop a method for fitting connective-tissue cell migration data to persistent random walk models, which accounts for these two significant sources of error and enables accurate determination of the cell motility parameters. We demonstrate the usefulness of this method for modeling both isotropic cell motility and biased cell motility, where the migration of a population of cells is influenced by a gradient in a surface-bound adhesive peptide. This method can discern differences in the motility of populations of cells at different points along the peptide gradient and can therefore be used as a tool to quantify the effects of peptide concentration and gradient magnitude on cell migration.

INTRODUCTION

Cell migration is an important class of phenomena that play key roles in physiological processes such as wound healing, metastasis, and embryonic development. Cell migration is most commonly described as a persistent random walk,

$$\langle d^2 \rangle = n_d \langle S^2 \rangle P t \left[1 - \frac{P}{t} \left(1 - e^{-t/P} \right) \right], \quad (1)$$

where $\langle\ \rangle$ indicate the mean over a population of cells, n_d is the number of dimensions in which the cells are migrating, $\langle S^2 \rangle$ is a characteristic mean-squared speed of the population, P is a persistence time in direction, and t is time (1,2) (see Appendix for usage of the notation $\langle\ \rangle$). This model is arrived at by assuming that the orthogonal components of the velocity are described by two-parameter Langevin equations, subject to certain assumptions regarding velocity distributions (3,4) (see Appendix). Dunn and Brown have demonstrated that a discrete form of this two-parameter model is sufficient for describing fibroblast motility (5), and Stokes et al. have extended the analysis to describe continuous motion (6). Although Eq. 1 only accounts for isotropic motility, the model can be modified in a number of ways to account for anisotropies in the cell motility. The simplest of these modifications is to add an additional parameter that accounts for directional bias (6–9) (see Appendix). The reader is referred to previous works for a description of the development of the model (1–6,10,11).

This and other mathematical models of cell migration provide insights into the mechanisms by which cells migrate, and may predict what cellular responses will be manifested by a cell population in response to alternative sets of stimuli. Regardless of the model used, distinctions among mechanisms of cell migration cannot be unambiguously made without obtaining a significant amount of data describing the individual paths of a large population of cells. This is because the phenomena of interest (i.e., cell speed and changes in cell orientation) are always coincident with a multitude of cellular processes (e.g., the cell cycle) that are manifested in the cell population as statistical, rather than deterministic, parameters. Additionally, systematic and stochastic errors may be associated with experimental observation. These sources of uncertainty are amplified in the case of connective-tissue cells important for wound healing (e.g., fibroblasts, keratinocytes) because these cells migrate very slowly compared to cell types whose migration has been previously well characterized (e.g., bacteria, neutrophils, and macrophages).

In this work, we develop a method for analyzing cell migration data that improves the accuracy with which the persistent random walk model parameters can be determined. This is accomplished by accounting for two common sources of error: accurate determination of the cell position and

Submitted September 8, 2006, and accepted for publication April 18, 2007.

Address reprint requests to Francis W. Wang, National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD 20899. Tel.: 301-975-6726; Fax: 301-975-9143; E-mail: francis.wang@nist.gov.

Matt J. Kipper's present address is Dept. of Chemical and Biological Engineering, Colorado State University, Fort Collins, CO 80523-1370.

Editor: Elliot L. Elson.

0006-3495/07/09/1797/12 $2.00

doi: 10.1529/biophysj.106.096800

heterogeneity in the cell population with respect to speed. We demonstrate the usefulness of this method by comparing it to nonlinear regression on a persistent random walk model. We also briefly discuss applying this model to anisotropic cell migration on peptide gradient surfaces, and provide an example where this method can be used to differentiate among cell populations whose response is different at different points along a peptide gradient.

The method presented here is a general technique for obtaining persistent random walk model parameters. However, our work is motivated by our interest in developing assays for fibroblast migration that can be used to quantify the haptotactic and haptokinetic activity of peptides derived from the basement membrane protein, laminin-1. Laminin-1 contains at least 20 peptide sequences that have been associated with various biological activities, including promotion of cell adhesion, migration, and angiogenesis (12–14). We have recently developed a technique for fabricating covalent peptide gradients that is suitable for studies aimed at quantifying the biological activity of the peptides and studying cellular mechanisms of migration, and that can be applied to surfaces and hydrogels (15). The method presented here will allow more accurate quantitation of how these potentially haptotactic or haptokinetic peptides affect fibroblast migration.

MATERIALS AND METHODS

The details of the substrate preparation and the cell culture are given in Kipper et al. (15). Briefly, glass coverslips (18 mm × 18 mm, No. 1½, Corning, Corning, NY) were cleaned and oxidized by oxygen plasma etch and then coated with poly(L-lysine) hydrobromide (PLL, 70,000 g mol^{-1} to 150,000 g mol^{-1}, Sigma-Aldrich, Saint Louis, MO) by incubating in a PLL solution (600 μL, 0.1 mg mL^{-1} in Dulbecco's phosphate-buffered saline) for 30 min, followed by rinsing in deionized water. The PLL provides an amine-functionalized surface onto which we can couple peptides with reasonably accurate control over the surface peptide concentration. For the biased cell motility, surfaces were prepared with a gradient in covalently bound laminin B160 peptide, as reported previously (15). B160 comes from the β1 chain of laminin-1. B160 has the sequence ^{1607}VILLQQSAADAIR^{1618} (12). B160 has been shown to promote the integrin-mediated adhesion of human umbilical vein endothelial cells, angiogenesis, and the promotion of human umbilical vein endothelial cell migration when presented in solution (12). Its capacity to promote fibroblast migration has not been studied or quantified to our knowledge.

Primary human foreskin fibroblasts (HFF) were cultured in Dulbecco's modified Eagle's media (DMEM, Gibco, Carlsbad, CA), with 10% fetal bovine serum (10%), L-glutamine (2 mM), penicillin and streptomycin (100 U mL^{-1}), and 0.1 mg mL^{-1} gentamicin. Medium was replaced every three days and cells were passed 3:1 or 4:1 at confluence. Nonsynchronized cells were used between the eleventh and eighteenth passage. The PLL-coated coverslips were fixed to the bottom of wells of Nunclon 4-well plates (Nalge Nunc International, Rochester, NY) using a small amount of silicone grease. Cells were plated at low confluence (35–70 cells mm^{-2}), with 7 mL of serum-free DMEM in each well, and allowed to attach to the surfaces for 1 h. Medium was then replaced with 5 mL of serum-free DMEM containing 1×10^{-5} μmol L^{-1} 5-(and 6)-(((4-chloromethyl)benzoyl)amino)tetramethylrhodamine (CellTracker Orange CMTMR, Molecular Probes, Eugene, OR). Cells were incubated with the dye for 30 min, and then the medium was replaced with 7 mL of serum-free DMEM.

The stained cells were imaged using in situ time-lapse video microscopy on an Axiovert 100 inverted fluorescent microscope (Carl Zeiss International, Oberkochen, Germany) equipped with an incubation chamber kept at 37°C, humidified, with a 5% CO_2 atmosphere, and a LEP Bioprecision motorized stage (Ludl Electronic Products, Hawthorne, NY). A shutter was used to ensure cells were only exposed during image capture, and an excitation filter was used to minimize exposure to potentially toxic UV radiation. Images were captured every 10 min for 20 h, and analyzed using the Metamorph 5.0 software package. An automated image capture routine cycled the robotic stage through multiple fields of view at each time point so that multiple time-lapse videos could be captured in a single experiment. To obtain data on a sufficiently large number of individual cells, the cells were imaged at relatively low magnification. A 5× objective was used, resulting in images that were 1.8 mm across. Between 30 and 100 cells were typically present in each field of view. Cell migration on PLL-coated surfaces was studied in four replicate experiments. Cell migration on the B160 peptide gradient surfaces was studied in two replicates. Time-lapse videos were collected from nine adjacent fields of view across the surface and all of the cells within each field of view were pooled as a single population.

Cell tracks were reconstructed from the positions of individual cells at each time point using an automated image analysis algorithm. The algorithm attempts to match a fraction of the pixel intensities in each image to a template created from the previous image for each cell. Once this threshold is obtained, the template is updated for the next image, and the cell position is defined as the centroid of the cell shape. This algorithm is very efficient at not ''losing'' cells from one image to the next, and is able to account for changes in the cell shape and position with time.

Effects of heterogeneity within the population on cell motility parameter estimation

Equation 1 implies that the mean-squared displacement should be a smooth, monotonic function of time. Measuring the $\langle d^2 \rangle$ of a population of cells over time should therefore provide us with data on which we can perform a nonlinear regression to obtain the parameters $\langle S^2 \rangle$ and P. In fact, the experimentally observed $\langle d^2 \rangle$ data are dominated by features that are not described by Eq. 1, even for a relatively large number of cells (i.e., $n > 50$). This is illustrated in Fig. 1. In Fig. 1 *a*, $\langle d^2 \rangle$ (*solid line*) ± SE (*dashed lines*) is plotted over 20 h for a population of 73 HFF on a PLL-coated surface (see Appendix for the definition of standard error). The deviations from the expected smooth, monotonic behavior may be due to some error in the cell position measurements or some true behavior of the cell population that the model represented by Eq. 1 fails to describe, or they may be due to both effects. The latter contribution represents a true component of the cell population behavior and should therefore be included in our description of the cell migration. First, we briefly discuss the heterogeneity within the population. The effects of cell position uncertainty are discussed in the following section.

Heterogeneity of cell migration parameters within a population is demonstrated in Fig. 1 *c*, which shows four of the individual cell paths plotted as the discrete positions that are observed in the experiment. In Fig. 1 *c*, we can see that at some times cells move with relatively high speed (indicated by relative distance between successive points) and long persistence time (indicated by consistency of direction of travel of the cell) for a considerable distance, and at other times they change direction more frequently or move more slowly.

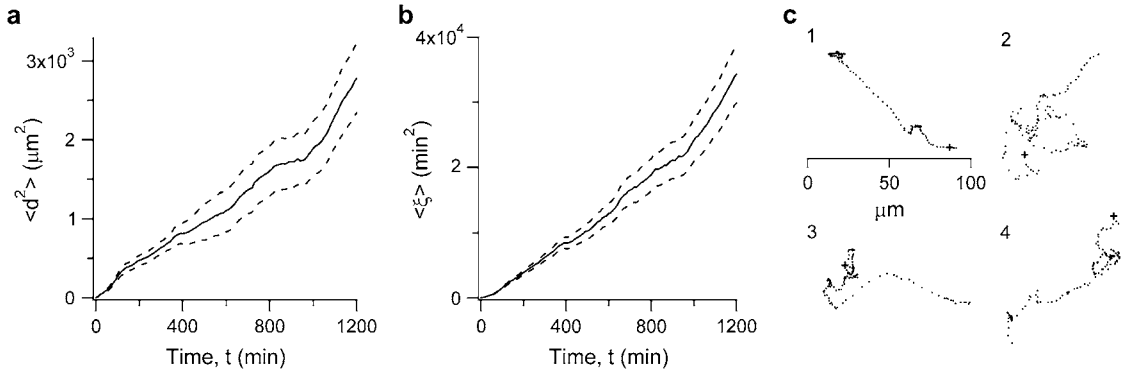

FIGURE 1 Mean-squared displacement (*a*) and $\langle\xi\rangle$ (*b*) as a function of time for a population of 73 HFF on a PLL-coated glass surface. Broken lines represent the standard error of the mean. Normalizing the mean-squared displacement by the cumulative mean-squared speed significantly reduces the standard error. (*c*) Cell tracks for four of the cells shown as the discrete cell positions measured in the experiment. The + indicates the initial position of each cell, and the scale bar applies to all four cell tracks.

Dunn (1) proposed a method for dealing with this heterogeneity. In this method, the mean-squared cell speed is computed directly over all of the m nonoverlapping (or overlapping) time intervals of equal length, Δt, over the entire time of the experiment, $m\Delta t$ (1). This method averages changes in cell behavior over the course of an experiment and does not leave the mean-squared speed as a fit parameter in Eq. 1. This method also provides a larger set of speed data from which to approximate $\langle S^2\rangle$ at the expense of the information about changes within the population as a function of time. Dunn's approach is valuable when data are only available on a limited number of cells over a relatively short time. However, we now have the capability of collecting and storing many more images than were feasible when Dunn's experiments were conducted. Furthermore, Bergman and Zygourakis have recognized that accounting for changes in cell motility parameters throughout the course of an experiment may improve the predictive ability of persistent random walk models as well as providing information about the temporal evolution of a cell population (16).

We propose an alternative approach to dealing with the heterogeneity in cell speed. According to Eq. 1, the expected value of d^2 for times $t \gg P$ increases at a rate proportional to $\langle S^2\rangle$. Therefore, normalizing d^2 for each cell by that cell's cumulative mean-squared speed should eliminate some undesired noise arising from speed fluctuations. The cumulative mean-squared speed of cell i, which we denote $\overline{S}_i^2(t)$ to distinguish it from the fit parameter in Eq. 1, is computed from the displacements of each cell at each previous time point. That is, for each cell, i, we define a time-dependent, normalized migration parameter $\xi_i(t)$

$$\xi_i(t) = \frac{d_i^2(t)}{n_d \overline{S}_i^2(t)} = \frac{d_i^2(t)}{n_d\left(\frac{1}{\tau}\sum_{j=1}^{\tau}\frac{\underline{d}_{i,j}^2}{\Delta t^2}\right)}. \quad (2)$$

(We use the notation $f(t)$ and $f(\Delta t)$ to indicate that f is a function of time or time interval, not that time is a multiplicative factor.) The displacement of the ith cell over the jth time interval is denoted $\underline{d}_{i,j}$, where the underscore distinguishes this displacement from the total displacement of the i^{th} cell from its starting position at $t = 0$, $d_i(t)$. τ is the number of time intervals at time t, given by $\tau = t/(\Delta t)$. $\overline{S}_i^2(t)$ is approximated by the expression in parentheses in the denominator of the righthand side of Eq. 2. Equation 2 assumes that we can accurately estimate the cell speed from the positions of a cell at successive time points. In doing so, we make the approximation

$$\frac{\underline{\delta}_{i,j}^2}{\Delta t^2} \approx \frac{\underline{d}_{i,j}^2}{\Delta t^2}, \quad (3)$$

where $\underline{\delta}_{i,j}$ is the path length of the i^{th} cell over the j^{th} time interval. Assuming that the mean of the ratio is a valid estimate for the ratio of the means, we can estimate the population mean, $\langle\xi\rangle(t)$:

$$\langle\xi\rangle(t) = \left\langle\frac{d_i^2}{n_d\overline{S}_i^2}\right\rangle(t) \approx \frac{\langle d^2\rangle(t)}{n_d\langle S^2\rangle} = Pt\left[1 - \frac{P}{t}\left(1 - e^{-t/P}\right)\right]. \quad (4)$$

The righthand side of Eq. 4 comes from Eq. 1. The mean normalized migration parameter, $\langle\xi\rangle(t)$, for the data in Fig. 1 *a* is plotted in Fig. 1 *b*. This normalization reduces the standard error by ~50% over the entire range of t. That is, the standard error on $\langle d^2\rangle(t)$ is about twice the standard error on $\langle\xi\rangle(t)$ as a percentage of the respective means. This reduction in the standard error occurs because cells that have relatively large $d_i^2(t)$ typically also have a relatively large $\overline{S}_i^2(t)$, while cells with relatively small $d_i^2(t)$ typically also have a relatively small $\overline{S}_i^2(t)$.

Effects of cell position uncertainty on cell motility parameter determination

To more fully investigate the effects of cell position uncertainty, suppose the position of cell i is measured at two

successive time points, $r_{i,j-1} = (x_{i,j-1}, y_{i,j-1})$ and $r_{i,j} = (x_{i,j}, y_{i,j})$. Its corresponding squared displacement over the j^{th} time interval is

$$\underline{d}^2_{i,j} = (x_{i,j} - x_{i,j-1})^2 + (y_{i,j} - y_{i,j-1})^2. \quad (5)$$

Further, assume that the error on each of the measured coordinates is an independent random variable with a normal distribution, and standard deviation, σ_x, and that the errors on the two coordinates are independent and identically distributed. The uncertainty (type B standard uncertainty) on the measurement of $\underline{d}^2_{i,j}$ arising from the uncertainty on the cell position is

$$\sigma_{\underline{d}^2_{i,j}} = \sqrt{8}\sigma_x\left[(x_{i,j} - x_{i,j-1})^2 + (y_{i,j} - y_{i,j-1})^2\right]^{1/2} = \sqrt{8}\sigma_x\underline{d}_{i,j}. \quad (6)$$

(See Appendix for notes about uncertainties.) To consider how this uncertainty influences the squared speed, recognize that the true squared speed of a cell would be given by the lefthand side of Eq. 3 if it were possible to measure the path length, $\underline{\delta}_{i,j}$. If the persistent random walk model (Eq. 1) is accurate, then $\langle d^2\rangle(t)$ is related to the mean-squared path length, $\langle\delta^2\rangle(t)$ by

$$\frac{\langle d^2\rangle(t)}{\langle\delta^2\rangle(t)} = n_d\frac{P}{t}\left[1 - \frac{P}{t}\left(1 - e^{-t/P}\right)\right]. \quad (7)$$

For computing the squared speed over a short time interval, Δt, we can replace $\langle d^2\rangle(t)$ with $\langle\underline{d}^2\rangle(\Delta t)$. This ratio approaches unity as Δt goes to zero (for $n_d = 2$, $\lim_{\Delta t\to 0}\langle\underline{d}^2\rangle(\Delta t) = \lim_{\Delta t\to 0}\langle\underline{\delta}^2\rangle(\Delta t)$). For nonzero Δt, the path length approximation, $\underline{d}^2_{i,j} \approx \underline{\delta}^2_{i,j}$, introduces an approximation on the cumulative mean-squared speed of cell i.

$$\frac{\overline{S}^2_{i\,\mathrm{Observed}}}{\overline{S}^2_{i\,\mathrm{Expected}}} = \frac{\frac{1}{\tau}\sum_{j=1}^{\tau}\frac{\underline{d}^2_{i,j}}{\Delta t^2}}{\frac{1}{\tau}\sum_{j=1}^{\tau}\frac{\underline{\delta}^2_{i,j}}{\Delta t^2}} = \frac{\langle\underline{d}^2_i\rangle_t}{\langle\underline{\delta}^2_i\rangle_t} \approx \left(\frac{\langle d^2\rangle(\Delta t)}{\langle\delta^2\rangle(\Delta t)}\right) = n_d\frac{P}{\Delta t}\left[1 - \frac{P}{\Delta t}\left(1 - e^{-\Delta t/P}\right)\right]. \quad (8)$$

The expected value is the value we would expect if Eq. 1 were an accurate model for the motility of the population. We use the notation $\langle\ \rangle_t$ to indicate the mean over time. In Eq. 8, we approximate that the mean-squared displacement and mean-squared path length of a single cell over many time intervals are equal to the mean-squared displacement and the mean-squared path length of many cells over a single time interval. The righthand side of Eq. 8 is from Eq. 7, where t is replaced with Δt, because the squared speed is computed over the interval Δt. The cell position uncertainty can then be propagated to Eq. 7 to determine the uncertainty on Eq. 8:

$$\sigma^2_{\frac{\langle\underline{d}^2_i\rangle_t}{\langle\underline{\delta}^2_i\rangle_t}} = \frac{1}{\langle\underline{\delta}^2_i\rangle_t^2}\frac{1}{\tau^2}\sum_{j=1}^{\tau}\left(\sigma_{\underline{d}^2_{i,j}}\right)^2 = \frac{8\sigma_x^2}{\langle\underline{\delta}^2_i\rangle_t}\frac{1}{\tau}\left(\frac{\langle\underline{d}^2_i\rangle_t}{\langle\underline{\delta}^2_i\rangle_t}\right) \approx \frac{8\sigma_x^2}{\langle\delta^2\rangle(\Delta t)}\frac{1}{\tau}\left(\frac{\langle d^2\rangle(\Delta t)}{\langle\delta^2\rangle(\Delta t)}\right). \quad (9)$$

We continue to use (Δt) to remind ourselves that the last factor in Eq. 9 is a function of Δt. To understand the sources of this uncertainty, we replace $\langle\delta^2\rangle(\Delta t)$ on the righthand side of Eq. 9 with $\langle S^2\rangle\Delta t^2$, where $\langle S^2\rangle$ is the fit parameter from Eq. 1.

$$\sigma^2_{\frac{\langle\underline{d}^2_i\rangle_t}{\langle\underline{\delta}^2_i\rangle_t}} = \frac{8\sigma_x^2}{\langle S^2\rangle\Delta t}\frac{1}{\Delta t}\frac{1}{\tau}\left(\frac{\langle d^2\rangle(\Delta t)}{\langle\delta^2\rangle(\Delta t)}\right) = \frac{4\sigma_x^2}{DP}\left(\frac{P}{\Delta t}\right)^2\frac{1}{\tau}\left(\frac{\langle d^2\rangle(\Delta t)}{\langle\delta^2\rangle(\Delta t)}\right). \quad (10)$$

Here, we introduce the diffusion coefficient, which is also referred to as a random motility coefficient (17,18), for the cell population, $D = 0.5\langle S^2\rangle P$. Equation 10 represents the squared uncertainty on the mean of the path length approximation, not the uncertainty on a single measurement. For a single measurement of the squared cell speed, we consider the case of $\tau = 1$. Taking the square root of Eq. 10 and the identity in Eq. 8 with $\tau = 1$ gives the uncertainty of a squared speed measurement arising from the uncertainty in the cell position, and accounting for the path length approximation (Eq. 3):

$$\sigma_{\frac{S^2_{i,j\,\mathrm{Observed}}}{S^2_{i,j\,\mathrm{Expected}}}} = \sigma_{S^*} = 2\left(\frac{\sigma_x}{(DP)^{1/2}}\right)\left(\frac{\Delta t}{P}\right)^{-1}\left(\frac{\langle d^2\rangle(\Delta t)}{\langle\delta^2\rangle(\Delta t)}\right)^{1/2}. \quad (11)$$

(See Appendix for notes about the uncertainties in Eqs. 9–11.) Because Eq. 11 represents the special case of the squared speed of a single cell over a single time interval, we write $\overline{S}^2_{i\,\mathrm{Observed}}/\overline{S}^2_{i\,\mathrm{Expected}}$ as $S^2_{i,j\,\mathrm{Observed}}/S^2_{i,j\,\mathrm{Expected}}$. For simplicity, we call this ratio S^*. The righthand side of Eq. 11 is written as a product of three important dimensionless groups. Each of these dimensionless groups has a physical significance. The left most group on the righthand side is the ratio of the uncertainty on the cell position to the expected distance the cells travel in one persistence time (equivalent to a mean free path). We will call this the normalized position uncertainty. The center group on the righthand side is a normalized observation time. The rightmost group on the righthand side represents the path length approximation. Note that the error arising from the path length approximation can be estimated by Eq. 8, and is also determined by the normalized observation time, $\Delta t/P$. As the observation time is reduced, the uncertainty on the squared speed measurement increases, but increasing the observation time too much reduces the accuracy of the path length approximation.

Combining Eqs. 8 and 11, with $\tau = 1$, we find that the ratio $S^* = S^2_{i,j\,\mathrm{Observed}}/S^2_{i,j\,\mathrm{Expected}}$ and its uncertainty can be estimated by

$$S^* \approx \mu_{S^*} \pm \sigma_{S^*} = \left(\frac{\langle d^2\rangle(\Delta t)}{\langle \delta^2\rangle(\Delta t)}\right) \pm 2\left(\frac{\sigma_x}{(DP)^{1/2}}\right)\left(\frac{\Delta t}{P}\right)^{-1} \times \left(\frac{\langle d^2\rangle(\Delta t)}{\langle \delta^2\rangle(\Delta t)}\right)^{1/2}, \quad (12)$$

where μ_{S^*} is the value of S^* predicted by Eq. 8 (see Appendix, Eq. A10). Equation 12 is only valid for $S^* > 0$. Equation 12 is plotted in Fig. 2 *a*, with the solid line representing the first term on the righthand side, and the broken lines representing different values of $\sigma_x(DP)^{-1/2}$. Fig. 2 *a* shows that for $\Delta t > P$, the squared speed is underestimated, due to the path length approximation, but the error diverges as Δt approaches zero, due to the normalized position uncertainty.

S^* does not have a normal distribution because it cannot take on a negative value. Thus, the probability density function (PDF), $\Pi(S^*)$, can be modeled as

$$\Pi(S^*) = \frac{1}{\sigma_{S^*}\sqrt{2\pi}} \times \left\{\exp\left[-\frac{(S^* - \mu_{S^*})^2}{2\sigma_{S^*}^2}\right] + \exp\left[-\frac{(S^* + \mu_{S^*})^2}{2\sigma_{S^*}^2}\right]\right\} \quad (0 \le S^* < \infty). \quad (13)$$

By integrating $S^*\Pi(S^*)$ over the range $0 \le S^* < \infty$, we can obtain the error that we would expect on the mean of many measurements of the squared speed

$$\varepsilon_{S^2} = \left\langle \frac{S^2_{\text{Observed}}}{S^2_{\text{Expected}}} \right\rangle - 1 = \frac{\sqrt{2}\sigma_{S^*}}{\sqrt{\pi}}\left[e^{-\frac{\mu_{S^*}^2}{2\sigma_{S^*}^2}} + \frac{\mu_{S^*}\sqrt{\pi}}{\sqrt{2}\sigma_{S^*}}\,\text{erf}\left(\frac{\mu_{S^*}}{\sqrt{2}\sigma_{S^*}}\right)\right] - 1 \quad (14)$$

(see Eq. A11). Here, *erf* is the error function. The mean of the squared speed normalized by its expected value, $\varepsilon_{S^2} + 1$, is plotted in Fig. 2 *b* for different values of the normalized position uncertainty. Notice that the average of the observed squared speed is likely to significantly overestimate the squared speed for short observation times, due to the nonnormality of the probability distribution function (Eq. 13). Thus, accurate determination of the squared cell speed is difficult from individual measurements of cell position, particularly when the persistence time is not known. This divergence at short observation times has not been dealt with analytically by others who have analyzed the error on determination of cell speed (1,18). Dunn recognized that at short observation times the uncertainty in the cell displacement becomes large, but in his formulation, shorter observation times provide more nonoverlapping time intervals over which to average the cell speed. However he noted that accurate measurements of cell speed would be limited by the resolution of the images at short observation times.

Let us now consider the ratio of the observed squared displacement to the expected squared displacement for the i^{th} cell at time t. This ratio, which we abbreviate, d^*, can be arrived at by extension of Eq. 5 to compute the uncertainty

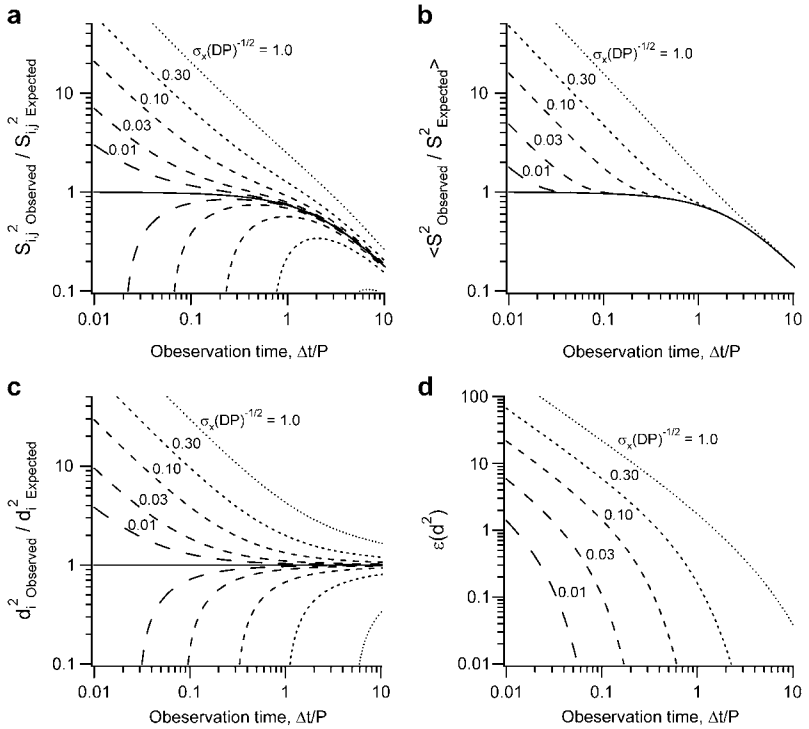

FIGURE 2 Ratio of the observed squared speed to the expected squared speed (*solid line*) ± standard uncertainty (*dashed lines*) according to Eq. 12 (*a*), and the mean of the ratio of observed squared speed to expected squared speed (see Eq. 14) (*b*) for different values of the normalized position uncertainty. Ratio of the observed squared displacement to the expected squared displacement (*solid line*) ± standard uncertainty (*dashed lines*) according to Eq. 15 (*c*), and the error of the of the observed squared displacement to the expected squared displacement according to Eq. 16 (*d*) for different values of the normalized position uncertainty.

on a squared displacement measurement from the origin at $t = 0$ and Eq. 1 for the expected value of the squared displacement:

$$\frac{d^2_{\mathrm{iObserved}}}{d^2_{\mathrm{iExpected}}} = d^* = 1 \pm \sigma_{\mathrm{d}^*} = 1 \pm 2 \times \left(\frac{\sigma_{\mathrm{x}}}{(DP)^{1/2}}\right)\left(\frac{t}{P}\right)^{-1/2}\left[1 - \left(\frac{t}{P}\right)^{-1}\left(1 - e^{-\mathrm{t/P}}\right)\right]^{-1/2} \tag{15}$$

(see Eq. A10). Here, for simplicity, we drop the notation (t), which indicates a function of time. As with Eq. 12, Eq. 15 is only valid for $d^2_{\mathrm{iObserved}}/d^2_{\mathrm{iExpected}} > 0$. Equation 15 is expressed in terms of the same important dimensionless groups that appear in Eq. 12, the normalized position uncertainty and the normalized observation time. The PDF of d^* is similar to the PDF given by Eq. 13 for S^*, with μ_{S^*} replaced by $\mu_{d^*} = 1$, and σ_{S^*} replaced by σ_{d^*}, equal to the uncertainty indicated in Eq. 15 by the terms after the symbol $\pm$. Thus, the mean error on the squared displacement has a similar form as in Eq. 14:

$$\varepsilon_{\mathrm{d}^2} = \left\langle \frac{d^2_{\mathrm{Observed}}}{d^2_{\mathrm{Expected}}} \right\rangle - 1 = \frac{\sqrt{2}\sigma_{\mathrm{d}^*}}{\sqrt{\pi}} \left[e^{-\frac{1}{2\sigma^2_{\mathrm{d}^*}}} + \frac{\sqrt{\pi}}{\sqrt{2}\sigma_{\mathrm{d}^*}} \mathrm{erf}\left(\frac{1}{\sqrt{2}\sigma_{\mathrm{d}^*}}\right)\right] - 1 \tag{16}$$

(see Eq. A11). Equations 15 and 16 are plotted in Fig. 2, *c* and *d*, respectively, for different values of the normalized position uncertainty. Just as for S^* (Fig. 2 *a*), d^* and its error diverge for small observation times due to the normalized position uncertainty. However, d^2_{Observed} converges to d^2_{Expected} for long observation times. Dickinson and Tranquillo predicted a different result for the effect of the position error on the mean-squared displacement (18). In their analysis, the position error results in a constant offset of 2γ, where γ is the variance of the cell position uncertainty. In developing Eqs. 15 and 16, we have assumed that uncertainty on the cell position is uncorrelated to the direction of the cell displacement.

Based on Fig. 2, *c* and *d*, one might assume that it would be possible to obtain $\langle S^2 \rangle$ and P from nonlinear regression of Eq. 1, provided we use sufficiently long observation times to reduce the error on $\langle d^2 \rangle$. However, as we have already noted from Fig. 1, these data may be too noisy to perform a reliable regression. Let us then consider the alternative mean normalized migration parameter, $\langle \xi \rangle$, which we proposed in Eq. 2. The uncertainty on a measurement of the normalized migration parameter for a single cell, ξ_{i}

$$\sigma_{\xi_{\mathrm{i}}} = \frac{2\sigma_{\mathrm{x}}}{(\overline{S}^2_{\mathrm{i}})^{1/2}} \xi_{\mathrm{i}}^{1/2}\left(1 + \frac{n_{\mathrm{d}}}{t\Delta t}\xi_{\mathrm{i}}\right)^{1/2} \tag{17}$$

(see Eq. A9). If we fit the cell position data to Eq. 4 rather than to Eq. 1, then the ratio of the observed value of ξ_{i} to the expected value of ξ_{i}, is given by

$$\frac{\xi_{\mathrm{i,\,Observed}}}{\xi_{\mathrm{i,\,Expected}}} = \xi^* = \frac{d^2_{\mathrm{iObserved}}}{d^2_{\mathrm{iExpected}}} \frac{\overline{S}^2_{\mathrm{iExpected}}}{\overline{S}^2_{\mathrm{iObserved}}} \pm \sqrt{2}\left(\frac{\sigma_{\mathrm{x}}}{(DP)^{1/2}}\right)\left(\frac{t}{P}\right)^{-1/2} \times \left\{\left[1 - \left(\frac{t}{P}\right)^{-1}\left(1 - e^{-\mathrm{t/P}}\right)\right]^{-1} + n_{\mathrm{d}}\left(\frac{\Delta t}{P}\right)^{-1}\right\}^{1/2} = \frac{1}{\mu_{S^*}} \pm \left(\frac{\sigma^2_{\mathrm{d}^*}}{2} + \frac{\sigma^2_{S^*}}{\tau\mu_{S^*}}\right)^{1/2} \tag{18}$$

(see Eq. A10). Again, we wish to emphasize that ξ_{i} and $\langle \xi \rangle$ are functions of time, though we drop the notation (t) for simplicity. Equation 18 is arrived at by using the definition in Eq. 2 for the observed value of ξ_{i}. $\xi_{\mathrm{i,Expected}}$ is the population mean, $\langle \xi \rangle$, and is computed according to the approximation in Eq. 4. Equation 18 has two normalized timescales, $\tau\Delta t/P = t/P$ (the time over which ξ_i is measured) and $\Delta t/P$ (the time step over which the squared speed is measured), which appear in σ_{d^*} and σ_{S^*}, respectively. The probability density function for ξ^* is also similar to Eq. 13, and results in a mean error on the observed value of the normalized migration parameter:

$$\varepsilon_\xi = \left\langle \frac{\xi_{\mathrm{Observed}}}{\xi_{\mathrm{Expected}}} \right\rangle - 1 = \frac{\sqrt{2}\sigma_{\xi^*}}{\sqrt{\pi}}\left[e^{-\frac{\mu^2_{\xi^*}}{2\sigma^2_{\xi^*}}} + \frac{\mu_{\xi^*}\sqrt{\pi}}{\sqrt{2}\sigma_{\xi^*}}\mathrm{erf}\left(\frac{\mu_{\xi^*}}{\sqrt{2}\sigma_{\xi^*}}\right)\right] - 1$$

$$\mu_{\xi^*} = \frac{1}{\mu_{S^*}}; \quad \sigma_{\xi^*} = \left(\frac{\sigma^2_{\mathrm{d}^*}}{2} + \frac{\sigma^2_{S^*}}{\tau\mu_{S^*}}\right)^{1/2} \tag{19}$$

(see Eq. A11). ξ^* and $(\varepsilon_\xi + 1)$ are plotted in Fig. 3. In Fig. 3, *a* and *b*, the effect of varying the observation time is illustrated, with the total number of time steps held constant at $\tau = 25$. The effect of varying τ is illustrated in Fig. 3, *c* and *d*, where $\Delta t/P$ is held constant at 0.3. Comparing Fig. 2, *c* and *d*, to Fig. 3, *a* and *b*, we can see that normalizing d^2_{i} by $\overline{S}^2_{\mathrm{i}}$ reduces the error on the measurement for small observation times. However, at longer observation times, ξ^* diverges due to divergence in $\overline{S}^2_{\mathrm{i}}$, whereas the error on $d^2_{\mathrm{iObserved}}/d^2_{\mathrm{iExpected}}$ converges. Thus, while ξ_{i} may have significantly less error than d^2_{i}, ε_ξ does not converge to 0.

Determining the uncertainty on the cell position

Unfortunately, there is no straightforward way to regress either the $\langle d^2 \rangle$ data or the $\langle \xi \rangle$ data due to the unknown uncertainty associated with determination of the cell position. However, if σ_{x} could be estimated, then it would be possible to correct the ξ_{i} data. We can see from Fig. 2 *b* that the shape of ε_{S^2} is sensitive to the value of $\sigma_{\mathrm{x}}(DP)^{-1/2}$. Thus, we can estimate the position uncertainty by computing $\langle S^2_{\mathrm{Observed}} \rangle$ for

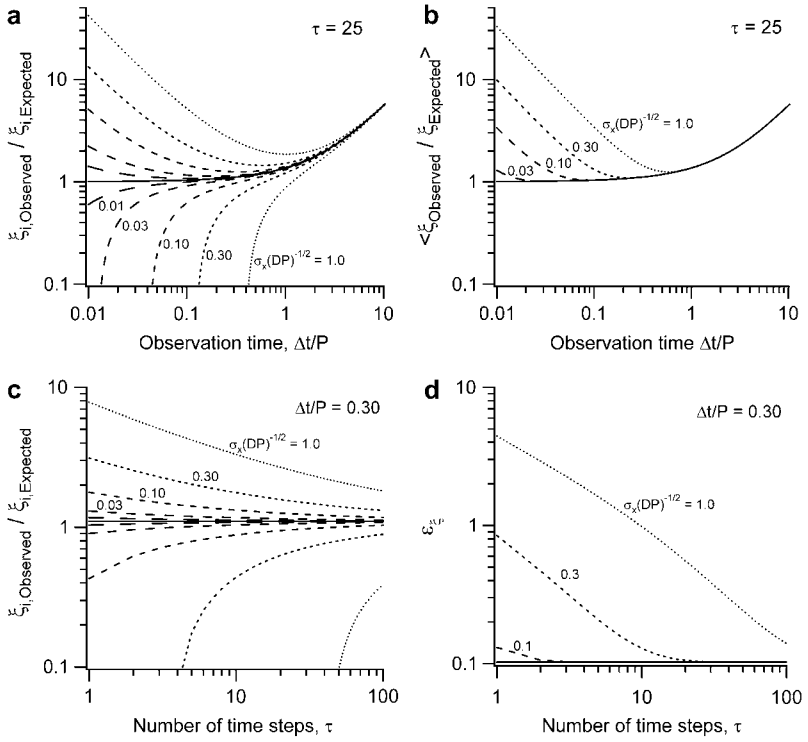

FIGURE 3 The ratio of the observed value of ξ_i to the expected value of ξ_i (*solid line*) ± standard uncertainty (*dashed lines*) for different values of the normalized position uncertainty according to Eq. 18 (*a* and *c*), and the error on $\langle\xi\rangle$ computed from Eq. 19 for different values of the normalized position uncertainty (*b* and *d*). In *a* and *b*, the number of time steps, τ, is held constant at 25. In *c* and *d*, the normalized observation time is held constant at 0.30.

multiple values of Δt and fitting to Eq. 14. In determining σ_x, we assume that this parameter is time invariant, thus, the mean-squared speed used is the average over all n cells and all m time intervals of length Δt (i.e., the definition used by Dunn (1), $\langle S^2_{\text{Observed}}\rangle = (1)/(\tau)\sum_{j=1}^{\tau}\left((1)/(n)\sum_{i=1}^{n}(\underline{d}^2_{i,j})/(\Delta t^2)\right)$). This regression requires three fit parameters. We choose the set: P, $\langle S^2_{\text{Expected}}\rangle$, and $\sigma_x(DP)^{-1/2}$. Fig. 4 *a* shows a fit of Eq. 14 to the data from the cell paths from Fig. 1, along with the computed values of the fit parameters. The effects of both the cell position uncertainty (at $\Delta t/P < 1$) and the path length approximation (at $\Delta t/P > 1$) are apparent. (Compare Fig. 4 *a* to the family of curves in Fig. 2 *b*.) The error on the cell position, σ_x ± type A combined standard uncertainty, is 1.2 μm ± 0.3 μm. This corresponds to the size of one pixel (1.3 μm × 1.3 μm) in the microscopy images. Although the uncertainty on cell position is likely determined by a variety of factors (including our definition of the cell position), we find it instructive that our uncertainty appears to be near the limit of our camera resolution.

The above analysis assumes a constant mean-squared speed for the population; however, there is a systematic time-dependent variation in the mean-squared speed. The root mean-squared (RMS) cell speed (mean ± SE) computed from the 10-min time intervals as a function of time is plotted in Fig. 4 *b* for the first 900 min of the experiment (the range used to compute the data shown in Fig. 4 *a*). Excluding the data from the first 100 min improves the fit somewhat, and provides an even smaller value for σ_x, approaching the theoretical limit of half of a pixel width (~0.7 μm). We will demonstrate how the prediction of P and $\langle S^2\rangle$ can be improved in the following section.

Improved method for fitting cell migration data to the persistent random walk model

Now that we have an estimate for σ_x, we can develop a practical method for fitting the experimental data to the model. We use Eq. 19, and the approximation first introduced in Eq. 4 to model the mean normalized migration parameter $\langle\xi\rangle$ as

$$\langle\xi\rangle = \frac{\langle\xi_{\text{Observed}}\rangle}{[f(t)]} = \langle\xi_{\text{Observed}}\rangle\left\langle\frac{\xi_{\text{Expected}}}{\xi_{\text{Observed}}}\right\rangle = \langle\xi_{\text{Observed}}\rangle \times\frac{\sqrt{\pi}}{\sqrt{2}\sigma_{\xi^*}}\left[e^{-\frac{\mu^2_{\xi^*}}{2\sigma^2_{\xi^*}}} + \frac{\mu_{\xi^*}\sqrt{\pi}}{\sqrt{2}\sigma_{\xi^*}}\,\text{erf}\left(\frac{\mu_{\xi^*}}{\sqrt{2}\sigma_{\xi^*}}\right)\right]^{-1}. \quad (20)$$

Here, $\langle\xi\rangle$ is an estimate for the true value and is also the expected value from the model, and we define the factor $f(t)$ as the ratio of the observed value to the true value of $\langle\xi\rangle$. We drop the subscript ''Expected'' from $\langle\xi\rangle$ to reflect the assumption that the model is accurate. The ratio $f(t)$ can be computed from μ_{ξ^*} and σ_{ξ^*}, which can be computed from P and $\sigma_x(DP)^{-1/2}$ (using Eqs. 12, 15, and 19) . We use the values of P and $\sigma_x(DP)^{-1/2}$ that were obtained from the regression in Fig. 4 *a* as initial guesses to compute $f(t)$. Once $f(t)$ is known, $\langle\xi_{\text{Observed}}\rangle$ can be transformed to $\langle\xi\rangle$ using Eq. 20. An improved prediction for the persistence time can be obtained from a fit to Eq. 4. Finally, we use this improved value

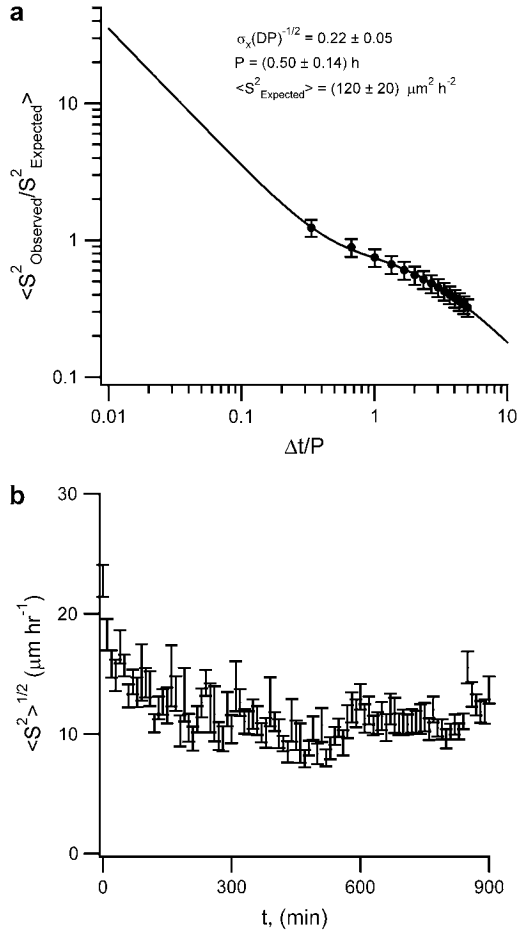

FIGURE 4 (*a*) Fit of the observed mean-squared speed as a function of observation time to Eq. 14 to obtain estimates for the normalized position uncertainty, persistence time, and expected mean-squared speed. Model parameters are reported as mean ± the standard deviation. The speed data were computed from the displacements of the 73 HFF for the first 900 min of migration using the overlapping time intervals with increasing observation time from 10 min to 150 min. Error bars represent the standard deviation of the speed. RMS speed as a function of time for the population of 73 HFF from Fig. 1 *b*. Error bars represent the standard error of the mean.

of P to compute the error on the observed mean-squared speed (Eq. 14) to find the expected value of the mean-squared speed. This procedure can be iterated using the improved prediction of P to reevaluate $f(t)$ if suitable agreement between the values of P from the two fits is not obtained after the first iteration. Using this method, the correct values of $\langle S^2 \rangle$ and P will be obtained, regardless of the time interval, Δt, at which images are collected, provided that there are data at normalized time intervals below the convergence of the curves in Fig. 2 *b*.

Table 1 compares the results obtained from the procedure outlined above to results obtained from fitting the experimentally observed displacement data to Eq. 1. Data are fit to the models using the Levenberg-Markuardt nonlinear regression algorithm. The model fits to the experimental $\langle d^2 \rangle$ and $\langle \xi \rangle$ data are shown in Fig. 5, *a* and *b*, respectively. For each technique, only the first 900 min of data were used; there appears to be a change in the persistence time after ~900 min, which the persistent random walk model does not account for. We attribute this to some phenotypic change in the cell population that may arise from extended culture in serum-free media. Because the regressions are performed over the mean values, $\langle d^2 \rangle$ and $\langle \xi \rangle$, the standard uncertainties indicated for the model parameters (P and $\langle S^2 \rangle$ for Eq. 1 and P for Eq. 20) represent the accuracy with which the model is able to predict the parameters, and do not reflect the variance in the cell population. Therefore the uncertainties on these quantities shown in Table 1 provide us with a measure of the performance of each of the methods.

Clearly, Eq. 1 is not at all suitable for fitting the data to the persistent random walk. However, eliminating the variation that arises from heterogeneous cell speed, and then accounting for the uncertainty in the cell position, greatly improves the ability of the model to describe the experimental data. Also, we note that the uncertainty on the persistence time is much smaller than the uncertainty on the RMS speed, because in the second method, the model is used to predict a single value for the persistence time that describes the behavior of the population, whereas the value for the RMS speed contains the variation within the population.

Model development for anisotropic cell motility

For the case where cell motility is affected by some anisotropic property of the surface, such as a gradient in adhesion ligand concentration in one direction, the cell motility is biased in one direction. The anisotropy can be accounted for by adding a term to the Langevin equations for the velocity

TABLE 1 Comparison of two different techniques for fitting cell migration data to the persistent random walk model

	Parameter					
Technique	P (h)	$\langle S^2 \rangle^{1/2}$ (μm h^{-1}) (measured)	$\Delta t/P$	$\langle d^2(\Delta t) \rangle / \langle \delta(\Delta t)^2 \rangle$	$\langle S^2 \rangle^{1/2}$ (μm h^{-1})	D (μm^2 h^{-1})
Eq. 1 for d^2	*0.01 ± 0.32*	—	—	—	*70 ± 1500*	35 ± 1700
Eq. 20 for ξ	*0.34 ± 0.01*	11.9* ± 2.0	0.50 ± 0.02	0.9 ± 0.8	12.1† ± 2.0	24.4 ± 6.2

Italicized values are fit parameters to the respective models; other values are computed from the data. Values are reported as mean ± SD (type A combined standard uncertainty).

*Computed from the cell paths as $(1)/(\tau)\sum_{j=1}^{\tau}\left((1)/(n)\sum_{i=1}^{n}(d_{i,j})/(\Delta t)\right)$.

†Computed from Eq. 12.

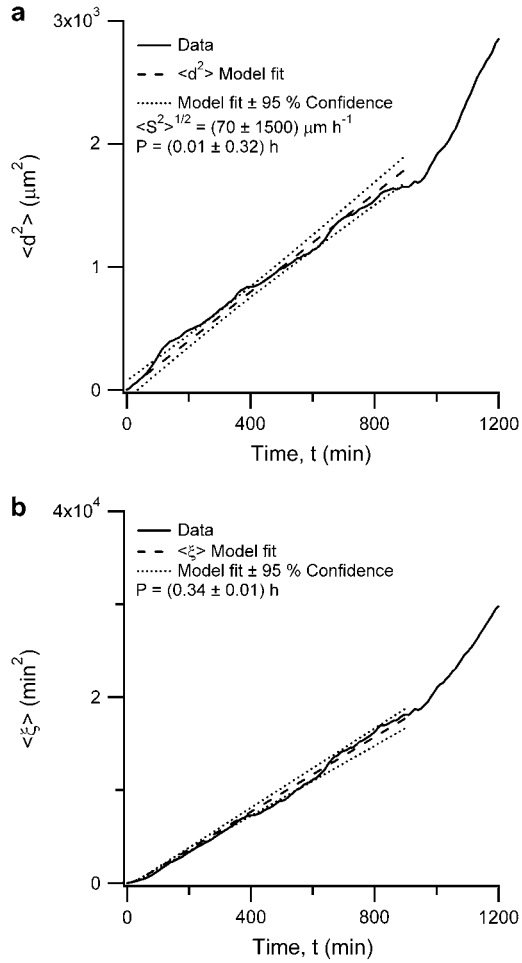

FIGURE 5 Model fits of the data from (*a*) Fig. 1 *a* to Eq. 1, and (*b*) Fig. 1 *b* to Eq. 4. Model parameters are reported as mean ± standard deviation.

distributions resulting in a linear drift in the population, parallel to the gradient, which is characterized by a bias speed, S_{Bias} (6–9) (see Appendix). The movement of the cells in the direction orthogonal to the gradient is unaffected. That is, for a gradient in the x direction, $\langle y(t)\rangle \approx 0$, whereas $\langle x(t)\rangle \approx S_{\text{Bias}}t$. S_{Bias} can therefore be estimated by fitting the average x position of the population as a function of time to a line. To correct for the anisotropic contribution to the total displacement of each cell, we compute the isotropic component of each cell's squared displacement as

$$d^2_{\text{iRandom}}(t) = [x(t) - S_{\text{Bias}}t - x(0)]^2 + [y(t) - y(0)]^2. \quad (21)$$

The isotropic component of the cell motility, d^2_{iRandom}, can then be further analyzed as above. It is important to note that there will be some variance within the population with respect to the degree to which each cell responds to the gradient. Thus, S_{Bias} cannot be computed for each cell. Rather, S_{Bias} is a statistical quantity associated with the cell population.

On the B160 peptide gradient surface, we observe an RMS bias speed of $\sim$1.6 μm h^{-1} in the direction of increasing peptide concentration. More interestingly, there is a significant change in the isotropic component of the cell migration with changing peptide concentration. The isotropic cell motility parameters are shown for populations of cells at nine different adjacent fields of view across a B160 peptide gradient in Fig. 6. Statistical significance among populations was determined using Welch's t-test ($p < 0.05$). Because the speed data are log-normally distributed, statistics were computed on log-transformed speed and diffusivity data. We do not necessarily expect that the populations in adjacent fields of view should be statistically different since each population represents cells on a distribution of surface peptide concentrations. Fig. 6 *c* illustrates that the diffusivity of the cell population can be more than doubled by selecting an appropriate peptide concentration. However, if the surface is too adhesive, having too much peptide, the cell motility is reduced to about half that on unmodified PLL. The observation of a maximum in motility at an intermediate adhesion ligand concentration is consistent with our previous observations of cell motility on other laminin peptide gradients (15), as well as the observations of others who have studied the migration of anchorage-dependent cells (19,20) and the responses of neutrophils to soluble chemoattractants (21,22).

DISCUSSION

Fig. 5 and Table 1 provide a dramatic example of the improvement that our technique offers over simple nonlinear regression on Eq. 1. In this particular case, nonlinear regression on Eq. 1 is a very poor method, in part because the RMS speed of the population drops significantly over the first few minutes of the migration (c.f. Fig. 4 *b*). The increased cell speed at early times may be because the cells are still forming adhesive contacts with the surface. The cells may also be changing their speed early because of some toxicity of the fluorescent dye or in response to the change to serum-free medium. This change in the cell speed may alter the specific responses to substrate bound peptides. Furthermore, this change in the RMS speed tends to inflate the $\langle d^2\rangle$ at early times compared to the later times, making it nearly linear with time with an intercept very near zero. Equation 1 can be rewritten as

$$\langle d^2\rangle = n_{\text{d}}\langle S^2\rangle[Pt - P^2] + n_{\text{d}}\langle S^2\rangle P^2 e^{-t/P}. \quad (22)$$

Equation 22 is the sum of a linear component, with slope $n_{\text{d}}\langle S^2\rangle P$ and intercept -$n_{\text{d}}\langle S^2\rangle P^2$, and an exponential component with a time constant of P. However, because our $\langle d^2\rangle$ data in this case are nearly linear with zero intercept, the fit eliminates the exponential and the intercept by setting P close to zero, and essentially fits the slope to the product of two parameters. This contributes to the large uncertainties on $\langle S^2\rangle^{1/2}$ and P indicated in Table 1. Therefore, we might expect to improve the fit by rejecting the first 100 min of migration data, setting the initial position of each cell to its

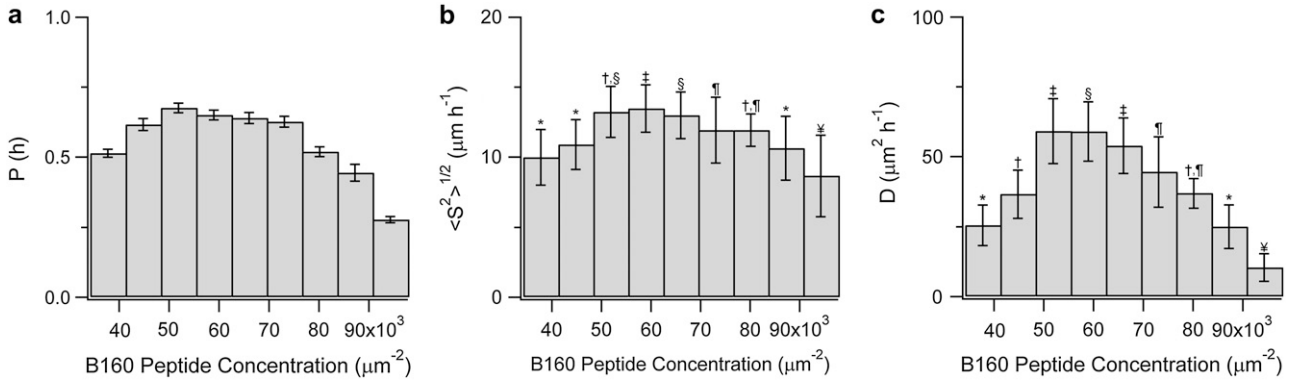

FIGURE 6 Persistence time (*a*), root mean-squared speed (*b*) and diffusivity (*c*) for nine populations of HFF on a laminin B160 peptide gradient. Error bars represent the standard deviations over the mean values obtained at each time point. In *b* and *c*, the populations marked with symbols are statistically different from populations not marked with the same symbol, as determined using Welch's *t*-test ($p < 0.05$) on the log-transformed speed and diffusivity data. When applying the *t*-test, a reduced number of degrees of freedom was assumed to account for the fact that the cell speed is correlated over the persistence time. This correlation means that not all speed observations are necessarily independent measurements, requiring that the number of degrees of freedom be normalized by the factor $P/\Delta t$.

position at $t = 100$ min. However, this results in $P = (0.02 \pm 0.31)$ h and $\langle S^2 \rangle^{1/2} = (60 \pm 500)\ \mu\text{m h}^{-1}$, which is little improvement over the original fit shown in Table 1.

Alternative methods of analyzing cell motility that eliminate immotile cells from the analysis (e.g., (23,24)) may intrinsically correct for the large error in cell speed at short observation times by only considering cells whose distance traveled is greater than the uncertainty on the cell position. Demou and McIntire suggest that a description of the cell migration is not complete without information on how the measured cell speeds are distributed (24). Fig. 7 is a probability distribution of the cell speed data shown in Fig. 4 *b*, taken only from 100 to 900 min (eliminating the relatively high speeds observed during the first 100 min). The curve in Fig. 7 is a log-normal fit to the data.

$$\Pi(S) = \frac{1}{\sigma S\sqrt{2\pi}} e^{-\frac{[\ln(S)-\mu]^2}{2\sigma^2}}. \tag{23}$$

The log-transformed speed data have a mean of 1.95 and standard deviation of 0.75. From this distribution, the RMS speed is computed as $\langle S^2 \rangle^{1/2} = (e^{2\mu+2\sigma\sigma})^{1/2} = 12.3\ \mu\text{m h}^{-1}$, and the standard deviation of the speed is computed as $\sigma_S = (e^{2\mu+2\sigma\sigma} - e^{2\mu+\sigma\sigma})^{1/2} = 8.1\ \mu\text{m h}^{-1}$. This observed RMS speed of $(12.3 \pm 8.1)\ \mu\text{m h}^{-1}$ accounts for the log-normal distribution. This is probably the best estimate for the RMS speed of the population if we want to describe the population dynamics.

We have developed an improved method for fitting data to a simple persistent random walk model. However, other models of cell motility are also parameterized by experimental data on cell positions and displacements. Therefore, this method could be generalized to more complex models of bacterial cell and leukocyte motility for instance, that account for other cell movements (e.g., turning frequency and alignment with external fields) (10,25–27) and subcellular phenomena (e.g., receptor-ligand binding and internal signaling) (10,25). Related models (11,22,23) that are based upon cell population density rather than individual cell position may not be subject to the same sources of uncertainty in determining the model parameters. However, if these models are parameterized from any measurements of individual cell motions, they are subject to the same concerns regarding position uncertainty and heterogeneity raised here.

CONCLUSIONS

We have demonstrated an improved method for fitting cell motility data to a persistent random walk model, which is suitable for slow moving connective tissue cells, such as fibroblasts. This method permits the accurate determination of cell motility parameters, by reducing the sensitivity to two

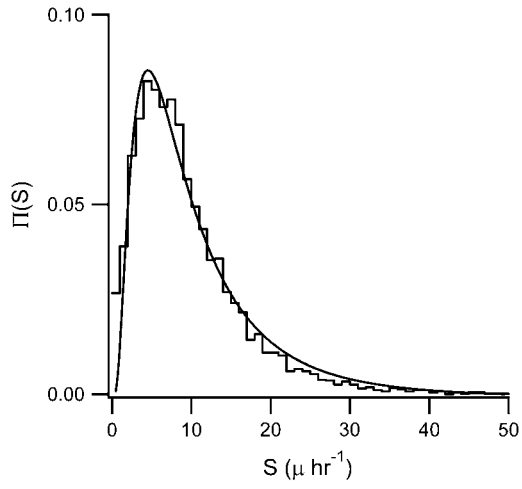

FIGURE 7 Probability distribution function for the speed data from Fig. 4 *b* over the range 100 min $\leq t \leq$ 900 min. The curve represents a log-normal fit to the data (Eq. 23).

common sources of error: uncertainty on determination of the cell position, and large variance in cell speed. There are three benefits obtained by using the procedure outlined above: First, the noise in the mean-squared displacement data associated with heterogeneity in the population is reduced by ~50% by normalizing each cell's squared displacement by its cumulative mean-squared speed. Second, as illustrated in Fig. 2, the accuracy of the observed cell speed and displacements are extremely sensitive to cell position uncertainty for short observation times when the cell position uncertainty is not close to zero. The method outlined above corrects for this. Third, the cell speed measurements are sensitive to the path length approximation at long observation times. Our method accounts for the path length approximation by adding the term $f(t)$. The method proposed here will be used to analyze the migration of fibroblasts on laminin peptide gradients and will enable accurate determination of cell migration parameters.

APPENDIX

Isotropic and anisotropic persistent random walks from Langevin equations

Equation 1 is derived from the Ornstein-Uhlenbeck process, which uses Langevin equations for the orthogonal components of the velocity (1,3,4,6):

$$\frac{dv_x}{dt} = -\beta v_x(t) + \sqrt{\alpha}\frac{dW_x}{dt}; \quad \frac{dv_y}{dt} = -\beta v_y(t) + \sqrt{\alpha}\frac{dW_y}{dt}, \tag{A1}$$

where W_x and W_y are continuous random ''white noise'' functions such that $W(t) - W(t')$ for $t \neq t'$ has a normal distribution with mean zero and a spectrum of magnitude α. The cell position as a function of time is then found by integrating these equations, for the x-component, for instance:

$$x(t) = \int_0^t v_x(t')dt'. \tag{A2}$$

Doob (3) (after Uhlenbeck and Ornstein (4)) showed that by integrating the stochastic differential equations for each of the velocity components, the expected value of the squared displacement components, $E\{[x(t) - x(0)]^2\}$ and $E\{[y(t) - y(0)]^2\}$ become

$$\mathrm{E}\left\{[x(t)-x(0)]^2\right\} = \mathrm{E}\left\{[y(t)-y(0)]^2\right\} = \frac{\alpha}{\beta^3}(\beta t - 1 - e^{-\beta t}). \tag{A3}$$

The orthogonal components of the squared displacements are added by the Pythagorean theorem to obtain the expected squared displacement in n dimensions. Making the substitutions $\langle S^2\rangle = \alpha/\beta$ and $P = 1/\beta$, one recovers for the persistent random walk (Eq. 1). Thus, the cell motility can be described by either pair of parameters ($\langle S^2\rangle$ and P or α and β).

To describe effects of external fields, or, in our case, the anisotropic response to the gradient, the original Langevin equations can be modified to account for external forces by adding acceleration terms, $A(t)$:

$$\frac{dv_x}{dt} = -\beta v_x(t) + \sqrt{\alpha}\frac{dW_x}{dt} + A_x(t);$$
$$\frac{dv_y}{dt} = -\beta v_y(t) + \sqrt{\alpha}\frac{dW_y}{dt} + A_y(t). \tag{A4}$$

Uhlenbeck and Ornstein considered the case where $A(t)$ represents a harmonic oscillator (4). Stokes et al. allowed $A(t)$ to be a function of the cell orientation, exerting a stronger influence over cells aligned with the gradient (6). Distasi et al. discussed the case where $A(t)$ is a constant (7). Setting $A_y = 0$ and A_x as constant, the simplest solution is obtained:

$$E[x(t) - x(0)] = \frac{v_x(0)}{\beta}(1 - e^{-\beta t}) + \frac{A_x}{\beta}t;$$
$$E[y(t) - y(0)] = \frac{v_y(0)}{\beta}(1 - e^{-\beta t}). \tag{A5}$$

Assuming that the initial x and y components of the velocities are distributed with a mean of 0 within the population, we obtain

$$E[x(t) - x(0)] = \frac{A_x}{\beta}t; \quad E[y(t) - y(0)] = 0. \tag{A6}$$

Thus, $\langle x\rangle$ increases linearly with time, at the rate A_x/β, whereas $\langle y\rangle$ remains near 0. In our analysis, we do not require that $v_x(0) = v_y(0) = 0$, we merely assume that $\langle y(t)\rangle \approx 0$, whereas $\langle x(t)\rangle \approx S_{\mathrm{Bias}}t$, where $A_x/\beta = S_{\mathrm{Bias}}$. If we subtract the cumulative bias, $S_{\mathrm{Bias}}t$, from the x position of each cell, we should obtain for the mean x position a value near 0. Furthermore the expected value for the corrected x displacements for each of the cells behaves as

$$E[x(t) - x(0) - S_{\mathrm{Bias}}t] = \frac{v_x(0)}{\beta}(1 - e^{-\beta t}) + \frac{A_x}{\beta}t - \frac{A_x}{\beta}t$$
$$= \frac{v_x(0)}{\beta}(1 - e^{-\beta t}), \tag{A7}$$

which is identical to the expected x displacement for the case $A_x = 0$. Therefore, correcting the x position of each cell with the mean bias yields an isotropic population that behaves identically to the original isotropic persistent random walk (Eq. 1).

Notation for the population mean

Throughout the text we use the notation $\langle\ \rangle$ to indicate the result of the mean operation over a population of size n (rather than the operation itself). Thus, we write the mean of f_i over the population as $\langle f\rangle$ (rather than $\langle f_i\rangle$):

$$\langle f\rangle = \frac{1}{n}\sum_{i=1}^{n} f_i. \tag{A8}$$

Uncertainties

The standard deviation refers to the combined standard uncertainty, and the standard error of the mean refers to the combined standard uncertainty of the mean. Eq. 6, 9–11, and 17 are for the theoretical uncertainty on an observation of the cell displacement, displacement to path length ratio, squared speed, and normalized migration parameter. These uncertainties represent type B standard uncertainties, because they are estimated assuming that they can be related to the standard uncertainty on the cell position coordinates, σ_x. For each of these quantities, we use the general formula for estimating the uncertainty of $g(x_1, x_2, \ldots x_n)$, σ_g, from the normally-distributed uncertainties on the independent variables x_i, $\sigma_{x,i}$

$$\sigma_g^2 = \sum_{i=1}^{n} \sigma_{x,i}^2 \left(\frac{\partial g}{\partial x_i}\right)^2. \tag{A9}$$

The uncertainty on the x and y components of the cell position are assumed to be independent and identically distributed with normal distributions, and are therefore indicated by a single variable, σ_x.

Errors

In Eqs. 12, 15, and 18, we represent the observed value of a function, f, normalized by its expected value as

$$\frac{f_{\text{Observed}}}{f_{\text{Expected}}} = f^* = \mu_{f^*} \pm \sigma_{f^*}. \tag{A10}$$

The expected value is the value obtained assuming that the model given by Eq. 1 is an accurate description of the cell migration. μ_{f^*} is the predicted value of the ratio, f^*, accounting for the path length approximation, and σ_{f^*} is the type B standard uncertainty of f^*, computed from Eq. A9. For the squared speed (Eq. 12) and the squared displacement (Eq. 15), we use S^* and d^* (rather than S^{2^*} and d^{2^*}).In Eqs. 14, 16, and 19 the error, ε_f, is computed from

$$\varepsilon_f = \left\langle \frac{f_{\text{Observed}} - f_{\text{Expected}}}{f_{\text{Expected}}} \right\rangle = \left\langle \frac{f_{\text{Observed}}}{f_{\text{Expected}}} \right\rangle - 1. \tag{A11}$$

Funding for this work was provided by the National Research Council through a Research Associateship to M.J.K.

REFERENCES

1. Dunn, G. A. 1983. Characterising a kinesis response: time averaged measures of cell speed and directional persistence. *Agents Actions Suppl.* 12:14–33.
2. Gail, M. H., and C. W. Boone. 1970. Locomotion of mouse fibroblasts in tissue culture. *Biophys. J.* 10:980–993.
3. Doob, J. L. 1942. The Brownian movement and stochastic equations. *Ann. Math.* 43:351–369.
4. Uhlenbeck, G. E., and L. S. Ornstein. 1930. On the theory of the Brownian motion. *Phys. Rev.* 36:0823–0841.
5. Dunn, G. A., and A. F. Brown. 1987. A unified approach to analysing cell motility. *J. Cell Sci. Suppl.* 8:81–102.
6. Stokes, C. L., D. A. Lauffenburger, and S. K. Williams. 1991. Migration of individual microvessel endothelial cells: stochastic model and parameter measurement. *J. Cell Sci.* 99:419–430.
7. Distasi, C., P. Ariano, P. Zamburlin, and M. Ferraro. 2002. In vitro analysis of neuron-glial cell interactions during cellular migration. *Eur. Biophys. J.* 31:81–88.
8. Maheshwari, G., and D. A. Lauffenburger. 1998. Deconstructing (and reconstructing) cell migration. *Microsc. Res. Tech.* 43:358–368.
9. Smith, J. T., J. K. Tomfohr, M. C. Wells, T. P. Beebe, T. B. Kepler, and W. M. Reichert. 2004. Measurement of cell migration on surface-bound fibronectin gradients. *Langmuir.* 20:8279–8286.
10. Alt, W. 1980. Biased random walk models for chemotaxis and related diffusion approximations. *J. Math. Biol.* 9:147–177.
11. Lauffenburger, D. 1983. Measurement of phenomenological parameters for leukocyte motility and chemotaxis. *Agents Actions Suppl.* 12:34–53.
12. Malinda, K. M., M. Nomizu, M. Chung, M. Delgado, Y. Kuratomi, Y. Yamada, H. K. Kleinman, and M. L. Ponce. 1999. Identification of laminin α 1 and β 1 chain peptides active for endothelial cell adhesion, tube formation, and aortic sprouting. *FASEB J.* 13:53–62.
13. Nomizu, M., W. H. Kim, K. Yamamura, A. Utani, S. Y. Song, A. Otaka, P. P. Roller, H. K. Kleinman, and Y. Yamada. 1995. Identification of cell-binding sites in the laminin α-1 chain carboxyl-terminal globular domain by systematic screening of synthetic peptides. *J. Biol. Chem.* 270:20583–20590.
14. Nomizu, M., Y. Kuratomi, K. M. Malinda, S. Y. Song, K. Miyoshi, A. Otaka, S. K. Powell, M. P. Hoffman, H. K. Kleinman, and Y. Yamada. 1998. Cell binding sequences in mouse laminin α 1 chain. *J. Biol. Chem.* 273:32491–32499.
15. Kipper, M. J., F. W. Wang, and H. K. Klienmann. 2007. Covalent surface chemistry gradients for presenting bioactive peptides. *Anal. Biochem.* 363:175–184.
16. Bergman, A. J., and K. Zygourakis. 1999. Migration of lymphocytes on fibronectin-coated surfaces: temporal evolution of migratory parameters. *Biomaterials.* 20:2235–2244.
17. Dickinson, R. B. 2000. A generalized transport model for biased cell migration in an anisotropic environment. *J. Math. Biol.* 40:97–135.
18. Dickinson, R. B., and R. T. Tranquillo. 1993. Optimal estimation of cell-movement indexes from the statistical-analysis of cell tracking data. *AIChE J.* 39:1995–2010.
19. Dimilla, P. A., J. A. Stone, J. A. Quinn, S. M. Albelda, and D. A. Lauffenburger. 1993. Maximal Migration of Human Smooth-Muscle Cells on Fibronectin and Type-Iv Collagen Occurs at an Intermediate Attachment Strength. *J. Cell Biol.* 122:729–737.
20. Wu, P., J. B. Hoying, S. K. Williams, B. A. Kozikowski, and D. A. Lauffenburger. 1994. Integrin-binding beptide in solution inhibits or enhances endothelial-cell migration, predictably from cell-adhesion. *Ann. Biomed. Eng.* 22:144–152.
21. Moghe, P. V., R. D. Nelson, and R. T. Tranquillo. 1995. Cytokine-stimulated chemotaxis of human neutrophils in a 3-D conjoined fibrin gel assay. *J. Immunol. Methods.* 180:193–211.
22. Tranquillo, R. T., S. H. Zigmond, and D. A. Lauffenburger. 1988. Measurement of the chemotaxis coefficient for human-neutrophils in the under-agarose migration assay. *Cell Motil. Cytoskeleton.* 11: 1–15.
23. Farrell, B. E., R. P. Daniele, and D. A. Lauffenburger. 1990. Quantitative relationships between single-cell and cell-population model parameters for chemosensory migration responses of alveolar macrophages to C5a. *Cell Motil. Cytoskel.* 16:279–293.
24. Demou, Z. N., and L. V. McIntire. 2002. Fully automated three-dimensional tracking of cancer cells in collagen gels: determination of motility phenotypes at the cellular level. *Cancer Res.* 62:5301–5307.
25. Tranquillo, R. T., D. A. Lauffenburger, and S. H. Zigmond. 1988. A stochastic model for leukocyte random motility and chemotaxis nased on receptor binding fluctuations. *J. Cell Biol.* 106:303–309.
26. Othmer, H. G., S. R. Dunbar, and W. Alt. 1988. Models of dispersal in biological systems. *J. Math. Biol.* 26:263–298.
27. Rivero, M. A., R. T. Tranquillo, H. M. Buettner, and D. A. Lauffenburger. 1989. Transport models for chemotactic cell populations based on individual cell behavior. *Chem. Eng. Sci.* 44:2881–2897.