Cell Motility as Persistent Random Motion: Theories from Experiments

David Selmeczi,*† Stephan Mosler,* Peter H. Hagedorn,‡ Niels B. Larsen,*‡ and Henrik Flyvbjerg*‡§

*Danish Polymer Centre, and ‡Biosystems Department, Risø National Laboratory, DK-4000 Roskilde, Denmark; †Department of Biological Physics, Eötvös Loránd University (ELTE), H-1117 Budapest, Hungary; and §Isaac Newton Institute for Mathematical Sciences, Cambridge, United Kingdom

ABSTRACT   Experimental time series for trajectories of motile cells may contain so much information that a systematic analysis will yield cell-type-specific motility models. Here we demonstrate how, using human keratinocytes and fibroblasts as examples. The two resulting models reflect the cells' different roles in the organism, it seems, and show that a cell has a memory of past velocities. They also suggest how to distinguish quantitatively between various surfaces' compatibility with the two cell types.

INTRODUCTION

Cell migration is essential in many physiological and pathological processes, e.g., embryogenesis, wound healing, inflammation, and metastasis. It is also essential to emerging medical technologies that rely on colonization of biomaterials by migrating cells (1–6). Many cellular signaling pathways that regulate migration have been described in recent years (7). The mechano-chemistry of migration is also studied and modeled in subtle detail (8–11). Less is known about the migratory pattern that results from a cell's processing of all external stimuli. Although motility models for bacteria have evolved to sophistication (12–15), phenomenological mathematical models for cells from higher organisms have, with few exceptions (16), remained simple. The Ornstein-Uhlenbeck (OU) process (17), an old model inspired by Brownian motion, remains the standard reference, and Fürth's formula (18)

$$\langle \vec{d}(t)^2 \rangle = 2nD(t - P(1 - e^{-t/P})), \quad (1)$$

for the mean square of a motile organism's displacement $\vec{d}(t) = \vec{r}(t) - \vec{r}(0)$ has remained a standard with which experimental data are analyzed; see, e.g., Gail and Boone and others (19–23). Here $\langle \ldots \rangle$ denotes expectation value, t is time, and $n = 1, 2,$ or 3 is the dimension of the space in which trajectories $\vec{r}(t)$ are studied. D is the diffusion coefficient of the OU process, and is referred to as the motility coefficient of the microorganism, and is sometimes written μ. P is the persistence time of the motion, and the notation $\beta = P^{-1}$ is often seen.

There is nothing as practical as a correct theory, so one should be conservative about well-proven formulas. However, we show below that human fibroblasts and keratinocytes move in a manner that rejects Fürth's formula and most other characteristic properties of the OU process. The data point clearly to new, cell-type-specific models, which we solve and

explain. The models are cell-type specific since the same model describes one cell type on different surfaces. Differences in the motility pattern of a given cell type on different surfaces are reflected in different values for model parameters. Consequently, these parameter values are quantitative cell-and-surface compatibility measures.

Our phenomenological approach to motility modeling is general, and may characterize other motility patterns as well. Models that result from this phenomenological approach, including ours, are natural targets for explanations in terms of biological processes.

MATERIALS AND METHODS

Cell culture substrata

Collagen I was acid-extracted from fetal bovine dermis and purified by a special sequence of differential salt precipitations (24). Coverslip glass was ultrasonicated in ethanol and water and dried in a clean bench. Molecular collagen-coated glasses resulted from adsorption from collagen I stem solution to the glass surface throughout 1 h at ambient conditions with 100% humidity, followed by thorough washing and drying. Fibrillar collagen-coated glass was fabricated by fibril reassembly of purified collagen I in a thin neutral buffer layer on the glass and its sedimentation at a density of 30 ng/mm^2, followed by thorough washing and drying (24). Tissue culture grade polystyrene, TCPS (Nunclon Delta, Nunc, Denmark), was used for cell line propagation and experiments.

Cell culture

HaCaT cells, a monolayer-forming transformed human keratinocyte line (German Cancer Research Center, DKFZ, Heidelberg, Germany), were cultured in DMEM/Nut Mix F-12 (Gibco, Carlsbad, CA) supplemented with 10% fetal calf serum (Gibco) and 4 mM glutamine. NHDF cells, normal human dermal fibroblasts (Clonetics, San Diego, CA), were cultured in fibroblast basal medium (Clonetics) supplemented with 2% fetal calf serum (Clonetics) and fibroblast growth factor (Clonetics). Both cell lines were incubated at 37°C and 7% CO$_2$. Before experiments with either cell type, the cells were dissociated by trypsin/EDTA (Gibco). The medium was changed to CO$_2$-independent medium (Gibco) with the same supplementation. Some 5000 cells/(cm^2 substratum) were seeded and incubated at 37°C. The motility data presented and analyzed below were obtained from HaCaT cells on substrata of molecular collagen and substrata of

fibrillar collagen, and from NHDF cells on the same two collagen substrata, as well as on glass and on TCPS, all at 37°C. HaCaT cells did not show measurable motility on glass and TCPS during the first 24 h of the experiment.

Monitoring of cell migration

A microscope stage was converted to a temperature-controlled incubation chamber with a computer-controlled step motor for parallel monitoring of several sites and samples under equal conditions. Cells were monitored in an inverted phase contrast microscope (Zeiss Axiovert 100 M, Zena, Germany) with autofocusing controlled by dedicated software. A digital camera (Olympus C-3040Zoom, Tokyo, Japan) recorded time-lapse movies with intervals Δt of 15 min for 24 h. Fig. 1 shows subsets of two frames from two movies, with cells marked.

Generation of trajectories

Only vital cells with full spreading on the substratum and long presence in the view field were selected from recorded image sequences. To ensure long presence, cells were excluded from the statistics if they started their trajectories within a 30-pixel-wide band around the edge of the 1024×768 pixels large view field. One pixel width corresponded to 0.85 μm in images of HaCaT cells, and 1.7 μm in images of NHDF cells.

One pixel in each cell nucleus was marked manually as the cell's coordinate. Because cells are located at random relatively to the grid of pixels with which they are viewed, the true position of a cell's center can be located anywhere inside the pixel used to mark it. Because we use the coordinates of the center of that square pixel instead of the unknown true coordinates of the cell's center, the coordinates that we use contain round-off errors. These errors come in all sizes between $\pm 1/2$ pixel width, with the same probability for all sizes. Hence the root-mean-square deviation (RMSD) of the round-off error distribution is $1/\sqrt{12}$ pixel width on the x-coordinate, and the same value on the y-coordinate.

Cell populations were dilute, so direct interactions between motile cells were rare. When they occurred, both cells' trajectories were excluded from the statistics. Cell trajectories crossing each other were less rare. No effect of this could be discerned. Data for cell coordinates were corrected for drift, and single cell trajectories were calculated from position and time data. A number of cells had to be removed from the statistics because they did not move. Visual inspection of trajectories would reveal nonmoving cells. From a number of such inspections, it was found that they could be excluded automatically, by excluding cells with a root-mean-square displacement that never exceeded 15 μm in the case of HaCaT cells and 20 μm in the case of NHDF cells. Fig. 2 shows examples of trajectories that were included in the statistics.

Data analysis

For each cell trajectory $\vec{r}(t)$, the positions $\vec{r}_j = \vec{r}(t_j)$, $t_j = j\Delta t, j = 1, 2, 3, \dots$ were recorded, and the velocities $\vec{v}_j = (\vec{r}_j - \vec{r}_{j-1})/\Delta t$ were calculated. From these velocities, the accelerations $\vec{a}_j = (\vec{v}_{j+1} - \vec{v}_j)/\Delta t$ were formed. Since surfaces were manufactured to be homogenous and isotropic to cells, and the cells' environment was kept constant in time, we tested and found velocity distributions consistent with spatial homogeneity and isotropy, as well as with temporal invariance; see Fig. 3. Consequently, when computing any ensemble average, we also averaged this quantity over time, space, and direction to improve statistics. Data were then plotted, analyzed, and modeled as described in the following sections and Appendix B.

Deriving motility models from experimental data
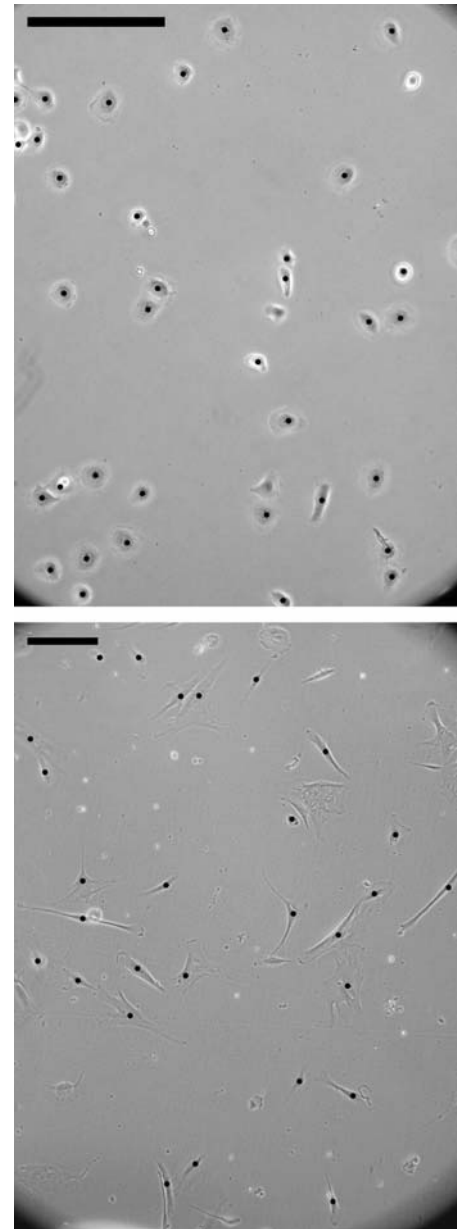
See Appendix A.

FIGURE 1 Snapshots of cells with markers. The markers shown are much larger than the one pixel in the center of a cell's nucleus that is used to denote the position of a cell and on which the marker is centered. Nonmoving cells and cells interfering with each other's trajectories were excluded from the statistics. Length bars measure 200 μm. (*Top*) HaCaT cells on substrate of molecular collagen. On glass and TCPS they are round in shape. Here they spread out and sometimes show an asymmetric migratory shape. (*Bottom*) NHDF cells on TCPS. On collagen surfaces they are elongated, whereas on TCPS, and especially on glass, they show a ''classical migratory shape'' with lamellipodia, riffling membrane, etc.

FÜRTH'S FORMULA AND THE ORNSTEIN-UHLENBECK PROCESS

Developed for protozoa in 1920, Fürth's formula has been used on motile cells since 1970 (19). Its agreement with data can be impressive (25) and is mostly satisfactory. This agreement is sometimes due to sizable experimental errors,
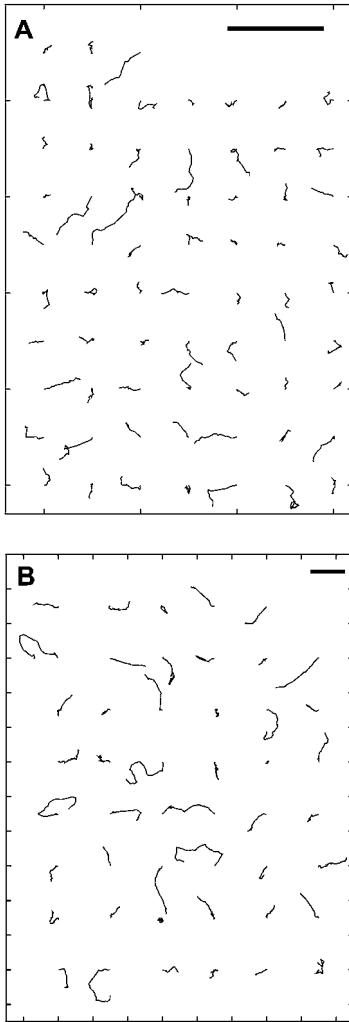
FIGURE 2   Collections of typical cell trajectories of the kind the ensuing analysis is based on. Each trajectory has been displaced so that the ensemble of starting points form a square lattice with a lattice spacing chosen large enough to prevent intersection of trajectories. The length of the two black bars shown is 200 μm. (A) HaCaT cells. (B) NHDF cells.

or to slow data sampling rate compared to P. Data sets with one or both of these properties cannot distinguish Eq. 1 from other functions of time that quickly approach a first-degree polynomial. The model-specific content of Eq. 1 is more obvious in the velocity autocorrelation function that Eq. 1 corresponds to,

$$\phi(t) \equiv \langle \vec{v}(t) \cdot \vec{v}(0) \rangle = \frac{nD}{P} e^{-|t|/P}, \qquad (2)$$

where $\vec{v} \equiv d\vec{r}/dt$. Equation 1 follows from Eq. 2, and vice versa, by integrating, respectively differentiating, twice. These equations do not define a specific motility model, however. A whole class of models have the property $\phi(t) \propto \exp(-|t|/P)$.

The Ornstein-Uhlenbeck process (17,26) is maybe the simplest such model for persistent random motion of motile cells, and a popular one. It is defined by the equation

$$P \frac{d\vec{v}}{dt} = -\vec{v} + \sqrt{2D}\,\vec{\eta}, \qquad (3)$$

where $\vec{\eta}$ is a normalized "white noise"; see Appendix A. The OU model's steady-state distribution of velocities, $p(\vec{v})$, is a simple Gaussian. This property conflicts with some experimental results (27–29). Some authors have observed that other distributions—an exponential (27) and Tsallis' distribution (28,29)—fit their velocity data better. Other authors have formulated new models with plausible details, but only in the form of computer algorithms. Our data are rich enough to allow a more radical approach. We let the data speak for themselves, as follows.

READING DATA

Experimental results for $\phi(t)$, for $p(\vec{v})$, for $d\vec{v}/dt$ as function of $\vec{v}$, and for the noise term $\propto \vec{\eta}$, all reveal the shortcomings of the OU model, we shall see. But the same experimental results yield its substitute: Eq. 3 states that a cell's acceleration $d\vec{v}/dt$ at any time t is a random vector with expectation value $-\vec{v}(t)/P$, and with equal RMSD in all directions. This statement can be compared with experimental data in a straightforward and model-independent manner: Fig. 4 *A* shows experimentally measured accelerations of HaCaT cells plotted against their instantaneous speed. If these data can be described by Eq. 3, Fig. 4 *A1* accelerations parallel to the velocity should average locally, as function of the speed v, to $-v/P$, with P a parameter to be fitted. The red data in Fig. 4 *B* show that they do.

Also, Fig. 4 *A2* accelerations orthogonal to the velocity should average to 0 locally, for all values of v. The green data in Fig. 4 *B* show that they do.

Furthermore, the accelerations in Fig. 4 *A* should scatter about these averages with the same v-independent RMSD, $\sqrt{2D/\Delta t}/P$ according to the OU model. Here $\Delta t = 15$ min is the time lapse between successive position measurements $\vec{r}_j$ from which velocities and accelerations were calculated; see Appendix A. The blue and magenta data in Fig. 4 *B* show the RMSD of each component of the accelerations. Within their error bars, the two RMSDs are identical. But they depend on the speed v, essentially as a first-degree polynomial, in disagreement with the OU model.

Finally, the accelerations in Fig. 4 *A* should be Gaussian distributed about their averages and uncorrelated, according to the OU model. Fig. 4 *C* shows clearly that they are not Gaussian distributed. Both are more similar to two exponential distributions placed back-to-back. This is not just due to time-lapse recording, i.e., a discretization effect due to $\Delta t = 15$ min. One can prove that the OU process gives a purely Gaussian distribution when studied with any finite value for Δt.

Fig. 4 *C* was produced as follows: each component of $\vec{a}_j - \langle \vec{a}_j \rangle$ at given $\vec{v}_j$, parallel, respectively orthogonal to $\vec{v}_j$, was measured in units of its speed-dependent RMSD shown in Fig. 4 *B*. The distribution of each of these reduced,
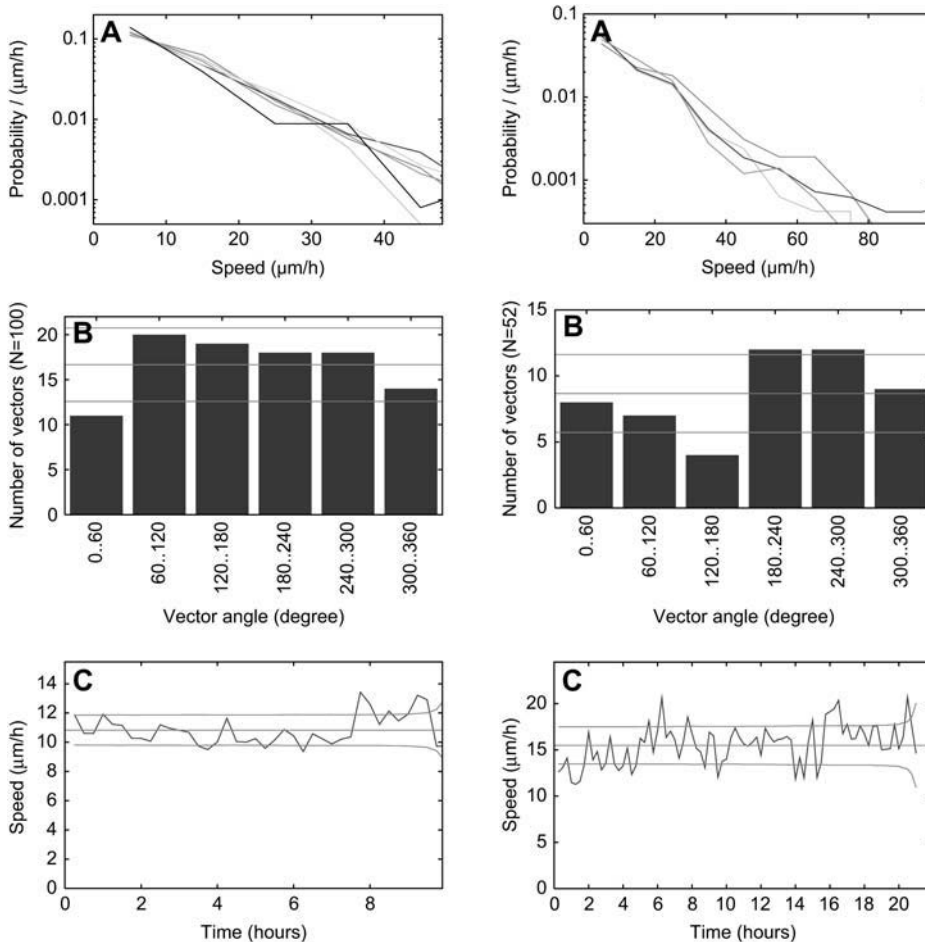
FIGURE 3 Experimental demonstration of spatial homogeneity, isotropy, and temporal invariance of cells' environment, expressed through the cells' pattern of motion. HaCaT (*left*) and NHDF (*right*) cells. (*A* panels) Distributions of observed speeds of HaCaT/NHDF cells on six/four different parts of the surface. The six/four distributions do not differ, apart from fluctuation due to finite statistics. (*B* panels) Six columns show the number of cells that moved in each of six different directions relative to their starting point. Directions are 60° intervals adding up to 360°. Total population of HaCaT/NHDF cells was 100/52. Three horizontal lines show theoretical expectation value and standard deviation for numbers shown as columns. The distribution on directions is consistent with isotropy. (*C* panels) Population-averaged speed as function of time. The population average is constant in time, up to stochastic fluctuations due to finite population size. Thick horizontal line is population- and time-averaged speed. Two thin horizontal lines are theoretical RMSD of population-averaged speed according to model presented below. The population inside the field of view is constant during the experiment, apart from a small decrease at the very end of the period of observation. This small decrease is seen in the theoretical RMSD, which increases correspondingly, as (population size)$^{-1/2}$.

dimensionless components was first measured as function of speed. Within fluctuations due to finite statistics, both distributions were found independent of speed, except at the lowest speeds. So we compounded data obtained at all speeds larger than half the mean speed. Fig. 4 *C* shows the resulting distributions.

Fig. 4 *D* shows our experimental results for the normalized autocorrelation function for the two components of $\vec{a}_j - \langle \vec{a}_j \rangle$. It is a Kronecker δ-function, i.e., there are no correlations, except for a small positive correlation in Fig. 4 *D1*. By neglecting the latter in a first attempt to find a model, we can achieve simplicity in that model-to-be.

Fig. 4 *C* looks discouraging since its non-Gaussian distributions seem to require a model with correlated, non-Gaussian noise. This turns out to be wrong: the speed-dependent noise amplitude combines with the 15-min discretization of data in a manner that allows a model with uncorrelated Gaussian noise to perfectly reproduce the non-Gaussian experimental noise in Fig. 4 *C*! Even the small positive correlation in Fig. 4 *D1* turns out to be explained by the model we arrive at below.

Fig. 4 *E* shows that the velocity autocorrelation function for HaCaT cells is not a simple exponential function, as it is

in the OU process. On the contrary, the data are fitted perfectly by a sum of two exponentials.

Fig. 4 *F* shows our experimental result for $p(\vec{v})$ as a histogram of observed velocities, binned on the v axis in panel *F1*, and binned on the v^2 axis in panel *F2*, both with logarithmic second axis. Isotropy of the surface on which the cells crawled, makes $p(\vec{v})$ depend only on the speed v, and not on the direction of the velocity. Consequently, if $p(\vec{v})$ is a Gaussian distribution on the $\vec{v}$-plane, as in the OU model, its graph is a straight line in panel *F2*. This is clearly not the case, according to the histogram of observed velocities.

If, on the other hand, speeds are exponentially distributed on the v axis, as suggested in Czirók et al. (27), the graph of that distribution is a straight line in panel *F1*. This is also not the case according to the histogram of observed velocities, though it could pass as an approximation away from zero velocity. If it were an exponential distribution also down to zero velocity, $p(\vec{v})$ would not be analytical in $\vec{v} = \vec{0}$, but diverge like $1/v$. Such singular behavior cannot be excluded à priori. But $p(\vec{v})$ must result from a dynamical theory of motility. And that results in a distribution that describes data much better, we shall find.
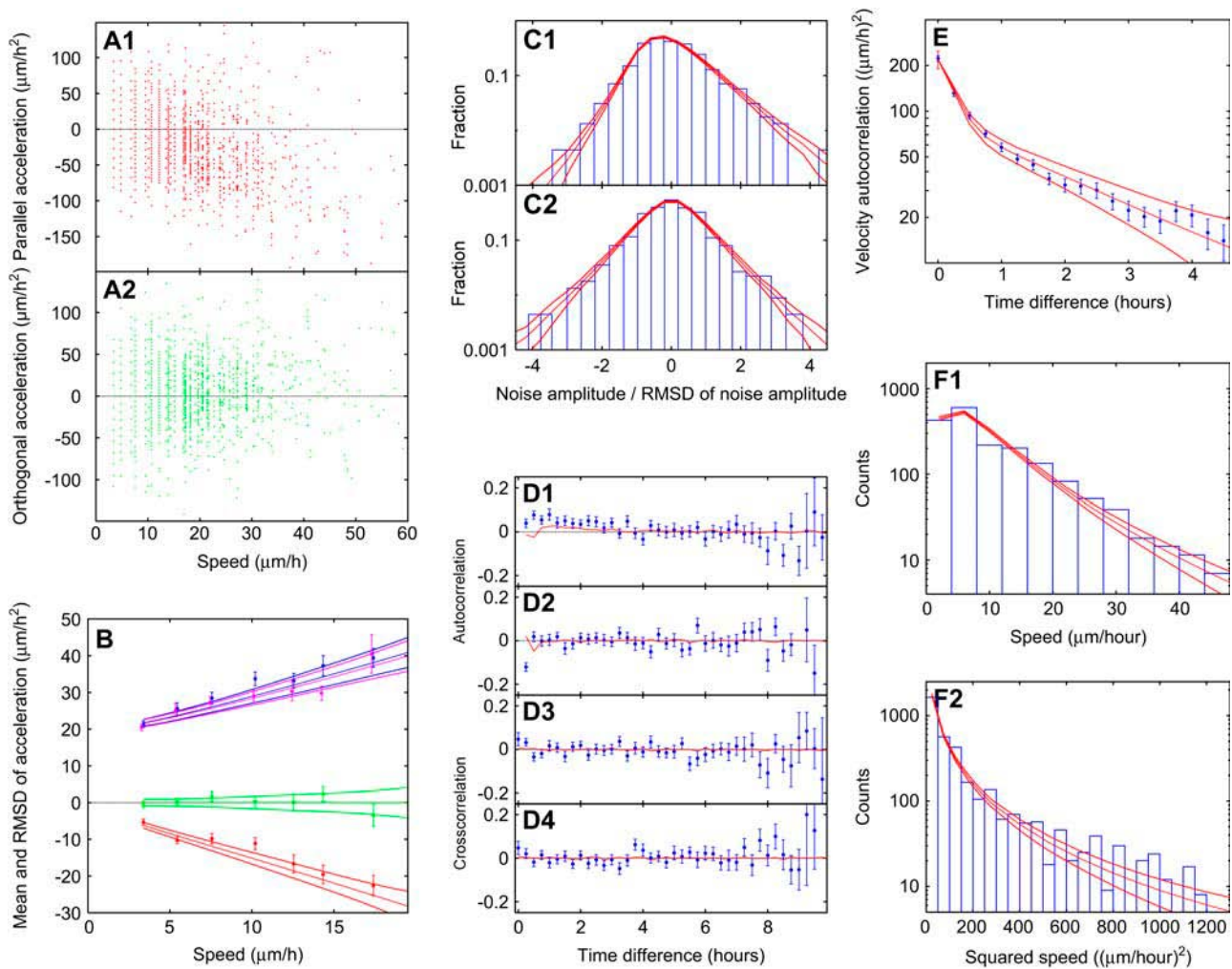
FIGURE 4   Motility of HaCaT cells. (A) Experimentally measured components of acceleration parallel (A1) and perpendicular (A2) to the cell's velocity, plotted against the cell's speed. The striped pattern in the scatter is due to finite resolution of positions, hence of speeds, caused by finite width of pixels. (B) Red and green data points are the mean values of the acceleration's two components shown in panels A1 and A2, respectively, as functions of speed. Blue/magenta data points show RMSD of same quantities, parallel/orthogonal component. Thick curves show theoretical expectation values of same quantities, according to the HaCaT model, Eq. 6, fitted to data in panels B, E, and F1 simultaneously. Pairs of thin lines next to thick curves show $\pm$ 1 SD from thick curve, according to HaCaT model's prediction of experimental data points' scatter. (C panels) Distribution of acceleration about its mean value, in units of distribution's RMSD, averaged over speeds v larger than $0.5\langle v \rangle$. Panels C1/C2 show distribution of component parallel/orthogonal to the velocity. Histograms show experimental results. Thick curves (averages) surrounded by two thin curves (average $\pm$ 1 SD) show predictions of the HaCaT model, not a fit to data shown here. (D) Correlation functions for acceleration's scatter about its mean. Panels D1/D2 are the autocorrelation function of parallel/orthogonal component, normalized to unity at time difference zero. Panels D3/D4 are the cross-correlation function between parallel and orthogonal component of acceleration's scatter about its mean, for positive/negative time difference, in units of (RMSD of parallel component) $\times$ (RMSD of orthogonal component). Curves are theoretical expectation values for the same quantities, according to the HaCaT model. (E) Data points are experimental velocity autocorrelation function $\phi(t)$. Error bars on data points underestimate true scatter as they were computed from experimental data that are correlated due to persistence of cell motion. The thick line is a guide to the eye connecting theoretical result for $\phi(t)$, computed from 15-min time-lapse position measurements, exactly as in experiment. Two thin lines are $\pm$ 1 SD on theoretical result for given value of $\phi(0)$. Red error bar on theoretical result for $\phi(0)$ is the theoretical RMSD of this quantity. Stochastic errors tend to make the whole data set shift up or down together. (F1) Histogram is distribution of observed speeds v, binned on v axis, and plotted with lin-log axis. Thick curve is HaCaT model's distribution of speeds, obtained as in experiment from 15-min time-lapse recordings of positions. Two thin curves are the theoretical RMSD of experimental result with regards to theoretical curve, obtained as RMSD of many speed distributions, each obtained from an independent simulated trajectory of duration equal to sum of durations of experimental trajectories. (F2) Histogram shows distribution of observed speeds v, binned on v^2 axis, and plotted with lin-log axis, so a Gaussian velocity distribution $p(\vec{v})$ would fall on a straight line. Curves are the same as in panel F1. Histogram values scatter more than the theoretical RMSD suggests they should, because they were computed from experimental data that are correlated due to persistence of cell motion, and unevenly redistributed between histogram bins by the effect of finite pixel width. The latter effect was not included in theoretical curves shown here and in panels C1 and C2, but was included in fitted curves. Both kinds of curves are shown in Fig. 10.

PROPERTIES OF AN UNKNOWN THEORY

Fig. 4 E shows that the velocity autocorrelation function for HaCaT cells is fitted perfectly by a sum of two exponentials. We consequently assume that

$$\phi(t) = \phi_1 e^{-|t|/P_1} + \phi_2 e^{-|t|/P_2}, \qquad (4)$$

and ask which generalization of the OU model might yield this function and the other properties of the data shown in Fig. 4. Such a generalization must, like the OU model, describe the rate of change of the cell's velocity. This acceleration cannot depend on a cell's position $\vec{r}(t)$, nor explicitly on spatial direction or time, because the cells crawl on surfaces that are homogenous and isotropic, and their environment is kept constant. So the acceleration must depend only on the cell's velocity, like in the OU model. We expect some memory in a cell, however, so the acceleration may depend not only on the current velocity, as in the OU model, but on past velocities as well. This dependence must be linear, like in the OU model, to ensure that $\langle d\vec{v}/dt \rangle_{\vec{v}} \propto \vec{v}$ as in the red data in Fig. 4 B.

The noise term seems to have an isotropic amplitude σ, as in the OU model, because blue and magenta data in Fig. 4 B coincide. But unlike the OU model's amplitude, this amplitude must be speed dependent, $\sigma = \sigma(v)$, according to the same data. The noise itself was uncorrelated to a good approximation, when measured with our 15-min time resolution; see Fig. 4 D. We consequently model the noise as uncorrelated on all timescales, to keep the model as simple as our data allow.

This list of properties of the yet unknown model, narrows it down to the integro-differential equation

$$\frac{d\vec{v}}{dt} = -K * \vec{v} + \sigma(v)\vec{\eta}, \qquad (5)$$

where K is a memory kernel yet to be determined, except causality demands that $K(t) = 0$ for $t < 0$: the future must not affect the present. The asterisk denotes convolution, i.e., $(K * \vec{v})(t) = \int_{-\infty}^{t} K(t - t')\vec{v}(t')dt'$. K is a scalar function of t, as opposed to a tensor of rank two, because the surface on which the cells move, is isotropic to them. Isotropy means that all directions are equivalent. When this equivalence principle is enforced on the theory we wish to find, mathematics gives that K must transform under spatial rotations as an invariant tensor, i.e., be a scalar. If the isotropy of space is broken, as it is in chemotactic and galvanotactic experiments, K is a tensor of rank two.

HaCaT THEORY

Equations 4 and 5 have only one solution for K; see Appendix A. Inserted in Eq. 5, it results in the equation of motion

$$\frac{d\vec{v}}{dt}(t) = -\beta\vec{v}(t) + \alpha^2 \int_{-\infty}^{t} dt'\, e^{-\gamma(t-t')}\vec{v}(t') + \sigma(v(t))\vec{\eta}(t).$$
$$(6)$$

Here α, β, and γ are known functions of P_1, P_2, and ϕ_1/ϕ_2, and satisfy $\beta\gamma > \alpha^2$, as they must for velocities to remain finite under the dynamics of Eq. 6; see Appendix A. The term in Eq. 6 containing β represents "loss of memory" of velocity at average rate β^{-1}. The term containing α^2, on the other hand, represents "memory" with characteristic time γ^{-1} and strength α^2/γ. This combination of memory was anticipated in Dunn and Brown (20). If γ is given a sufficiently large imaginary component, this model has an oscillatory velocity autocorrelation function similar to the one suggested in Shenderov and Sheetz (16) for *Dictyostelium discoideum*.

The experimental data shown in red in Fig. 4 B show proportionality between velocity and the mean acceleration at that velocity,

$$\langle d\vec{v}/dt \rangle_{\vec{v}} = -\beta_{\text{eff}}\,\vec{v}, \qquad (7)$$

like in the OU model. One can prove that Eq. 6 results in the very same proportionality and find $\beta_{\text{eff}}(\alpha, \beta, \gamma)$ (see Appendix A). So, apart from the experimental value for β_{eff}, the red data contain no information that is not already built into the model Eq. 6. But they do, of course, thereby confirm the choice of Eq. 6 as model.

Although $K(t)$ was determined from $\phi(t)$, the only information about $\sigma(v)$ that is contained in $\phi(t)$, is the theory's "fluctuation-dissipation theorem"; see Appendix A, Eq. 24. Thus, we are free to choose

$$\sigma(v) = \sigma_0 + \sigma_1 v, \qquad (8)$$

inspired by the data in Fig. 4 B, as long as the "fluctuation-dissipation theorem" is respected. This defines our model for HaCaT cells.

With this choice, one can derive the steady-state distribution of velocities (see Appendix A),

$$p_{\text{HaCaT}}(\vec{v}) = \frac{a e^a}{2\pi v_\sigma^2}\, \frac{\exp\left(-\dfrac{a}{1 + v/v_\sigma}\right)}{(1 + v/v_\sigma)^{2+a}}, \qquad (9)$$

where $v_\sigma \equiv \sigma_0/\sigma_1$ and $a \equiv 2\beta_{\text{eff}}/\sigma_1^2$. The moments $\langle v^k \rangle$ are also known analytically and are finite only for $k < a$. Because data show $\langle \sigma^2(v) \rangle$ is finite, $a > 2$ is necessary for the model to be consistent with the choice in Eq. 8.

Clearly, $p_{\text{HaCaT}}(\vec{v})$ is not a Gaussian distribution for $\sigma_1 \neq 0$. But it is the same function for all models having the same values for β_{eff}, σ_0, and σ_1, including, e.g., the case of no memory kernel, $\alpha = 0$. This demonstrates the value of plotting data as in Fig. 4 B. Experimental results for $p_{\text{HaCaT}}(\vec{v})$ and $\phi(t)$ contain precise, important, but incomplete information about the dynamics causing these functions. Fig. 4 B presents the experimental information in a manner that suggests this dynamics more directly.

COMPARING THEORY AND EXPERIMENT: HaCaT CASE

The theory in Eq. 6 was compared with experimental results as follows: Eq. 6 was simulated numerically to obtain

theoretical trajectories of persistent random motions. These trajectories were sampled with the same time lapse as was used to record experimental trajectories, to have theoretical results that can be compared with experimental data. Values for coordinates obtained by this sampling were rounded to the nearest-integer multiple of pixel widths, as described in Appendix B, before results were computed from them with the purpose of fitting the theory to data. After the theory has been fitted in this manner, it is more convenient to compare data with the smoother curves obtained from nonrounded coordinates, if rounding only makes a theoretical curve jagged. Direct inspection (Fig. 10 in Appendix B) shows when this is legitimate (panels *B–D*), and when not (panel *A*).

Thus, Fig. 4 *F* shows $p_{\mathrm{HaCaT}}(\vec{v})$—not the distribution given in Eq. 9—but the distribution of velocities obtained by time-lapse sampling the theory after it had been fitted to experimental data. For fitting, the theory was both time-lapse sampled and rounded to the nearest pixel. The velocity distribution resulting from the latter procedure was fitted to the experimental distribution of velocities of HaCaT cells, while simultaneously the theory's mean acceleration and the RMSD of its acceleration were fitted to the experimental data in Fig. 4 *B*, and its velocity autocorrelation function was fitted to the data in Fig. 4 *E*. As fitting parameters we used α, β, γ, σ_0, σ_1—five fitting parameters in a simultaneous fit of five functions to the experimental data shown in Fig. 4, *B*, *E*, and *F*. Fig. 4, *B–F*, shows the resulting fit.

The HaCaT model in Eq. 6 clearly captures the nature of the data, including the data in Fig. 4 *C*. The latter aspect of the data was not used when fitting the theory. So the theoretical distributions that are shown as curves in Fig. 4 *C* are predictions made by the theory for this aspect of the experimental data, after the theory was fitted to other aspects of the same data.

Experimental data and theory both differ from zero in the same manner for the first few hours in Fig. 4 *D1*. The same behavior is seen in Fig. 5 *D1*. This agreement between theory and experimental data is caused by a property of the theory that it picked up from other aspects of the data: $\vec{a}_\mathrm{j} - \langle \vec{a}_\mathrm{j} \rangle$ is an uncorrelated Gaussian noise in the OU model, but in the HaCaT model it also contains a small term linear in $\vec{v}$ because of the memory kernel. On the differential timescale it reads $(d\vec{v}/dt - \langle d\vec{v}/dt \rangle)|_{\vec{v}} = \beta_{\mathrm{eff}} \vec{v} - K * \vec{v} + \sigma(v)\vec{\eta}$. In the OU model the first two terms on the right-hand side of this identity cancel, because $\beta_{\mathrm{eff}} = \beta$ and $K(t) = \beta\delta(t)$. In the HaCaT model they don't cancel, and cause the observed small correlation.

NHDF THEORY

Fig. 5 shows results for NHDF cells. They move approximately twice as fast, on average, and are also in other ways more dynamic than keratinocytes. Their mean acceleration parallel to their velocity, e.g., decreases somewhat faster than proportional to v (red data points in Fig. 5 *B*. Fig. 12 in Appendix B shows this with maximal clarity, while fully

accounting for the time-lapse nature of the data.) Also, the RMSD of the acceleration differs for the two directions: its component orthogonal to the direction of motion is almost constant as in the OU model. But its component parallel to the direction of motion increases with v and doubles its value in the window shown. This behavior is captured by two modifications of Eq. 6: firstly, we replace the constant β with an increasing function $\beta(v)$. A first-degree polynomial, $\beta(v) = \beta_0 + \beta_1 v$, turns out to do the job, whereas a second degree polynomial does not further improve the agreement between theory and data. Secondly, we replace the scalar function $\sigma(v)$ with the tensor $\bar{\bar{\sigma}}(\vec{v}) \equiv \sigma_\parallel(v)\hat{v}\otimes\hat{v} + \sigma_\perp(v)\check{v}\otimes\check{v}$. Here $\hat{v} = \vec{v}/v$, $\check{v}$ is a unit vector orthogonal to $\vec{v}$, and $\sigma_\parallel(v)$ and $\sigma_\perp(v)$ are the RMSDs of the random components of the acceleration, parallel and orthogonal to the velocity, respectively. The data in Fig. 5 *B* suggest that $\sigma_\parallel(v)$ and $\sigma_\perp(v)$ are different first-degree polynomials in v, but with $\sigma_\parallel(0) = \sigma_\perp(0)$, so $\bar{\bar{\sigma}}(\vec{v})$ contains a total of three free parameters to be fitted, where $\sigma(v)$ contained two. This defines our motility model for NHDF cells.

Like the HaCaT model, this NHDF model has a non-Gaussian velocity distribution, shown in Fig. 5 *F*, but it cannot be found analytically. Neither can the correlation function $\phi(t)$, shown in Fig. 5 *E*, though its simple exponential behavior for $t > 1$ h indicates that good analytical approximations might be found there. The result of a simultaneous fit of this NHDF model to all experimental data shown in Fig. 5, *B*, *E*, and *F*, is shown in all panels in Fig. 5.

COMPARING THEORY AND EXPERIMENT: NHDF CASE

Equation 6 was simulated numerically in its NHDF variant and fitted to the NHDF data shown in Fig. 5, *B*, *E*, and *F*, using α, β_0, β_1, γ, σ_0, $\sigma_{1\parallel}$, $\sigma_{1\perp}$ as fitting parameters. The curves shown in Fig. 5, *C* and *D*, are the theory's predictions for the aspect of the experimental data that are also shown there.

We tested the possibility that ambiguity in the choice of pixel that marks a cell's position might play a role. This was tested by introducing this ambiguity in the theory, and then fitting the theory to the data with the ambiguity as a parameter. That resulted in a slightly better fit, but there is little ambiguity: only 1.5% of the times that a pixel is chosen, should one of its four nearest neighbors be chosen at random instead. So the data confirm that the round-off error associated with using a pixel as coordinate is larger than the ambiguity of this coordinate.

The NHDF model's agreement with data could be better in Fig. 5 *B*. The data could also be richer, however, and a theory should not be built on a single data set. The NHDF theory is not. Below, where we compare the model to data taken on different surfaces, we find better agreement between fitted theory and data for the average acceleration as function of speed. There is room for improvement of the NHDF model. We just don't know a systematic way of going about it,
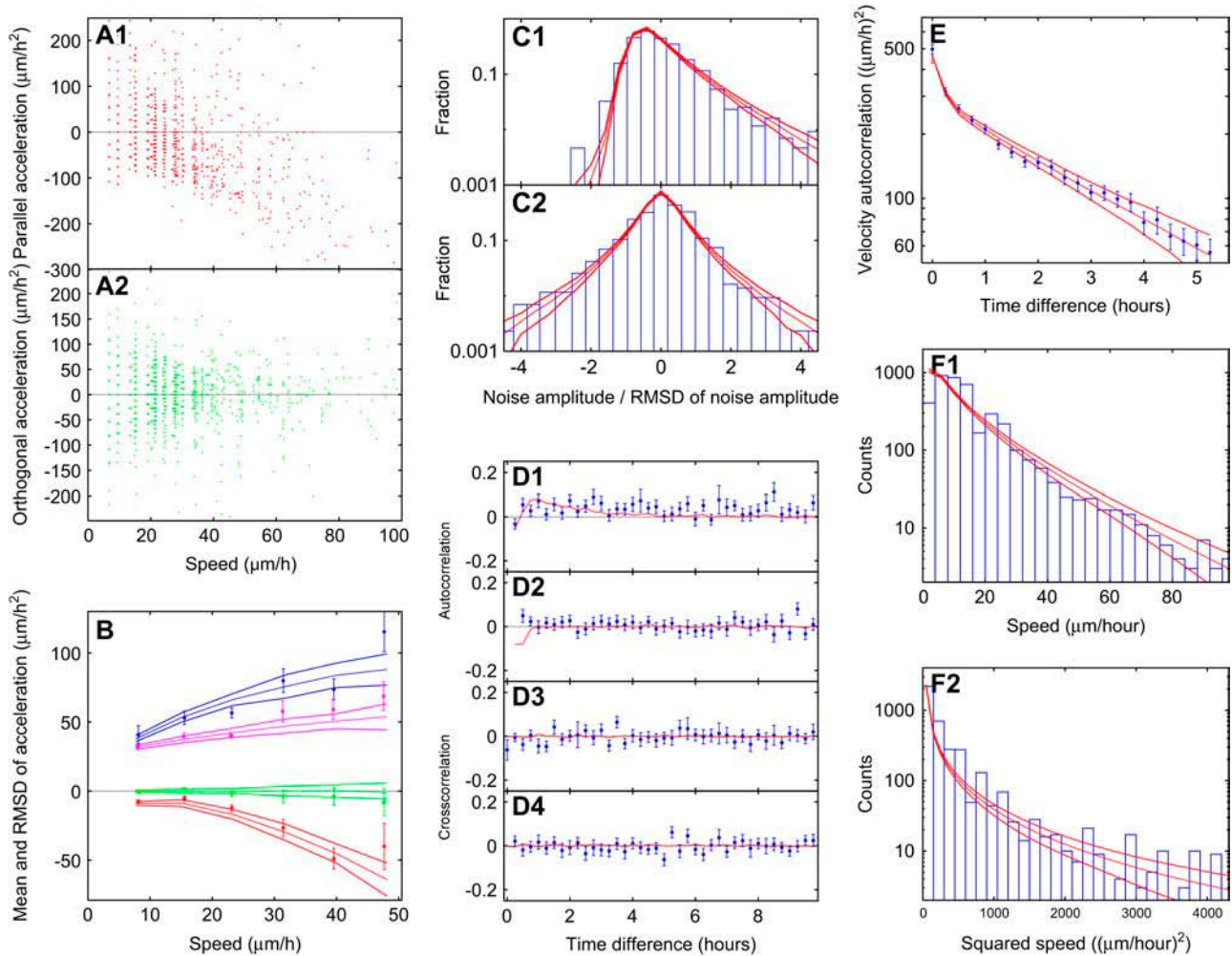
FIGURE 5   Same as Fig. 4, but for NHDF cells and NHDF model. (*C1*) The asymmetry of the distribution for NHDF cells is more pronounced than is the case for HaCaT cells. The NHDF model's prediction for this distribution and that in panel *C2* is shown as red lines. It is not a fit. The model was fitted only to data in panels *B*, *E*, and *F1*. (*D1*) Red curve is the NHDF model's prediction for the autocorrelation function for the acceleration's scatter about its mean in direction parallel to the velocity. Like the experimental data, it is small, but nonvanishing, though the model's noise term is uncorrelated Gaussian. (*E*) Velocity autocorrelation function. See caption to Fig. 4 *E*. (*F* panels) Distribution of speeds; see caption to Fig. 4 *F*.

except to suggest that more and better data are studied, since it is a nonlinear theory that is called for.

RESULTS FOR MODEL PARAMETERS

Table 1 shows values for parameters in the HaCaT and NHDF models, obtained by fitting these models to data as just described. The table also gives values for some quantities that characterize experimental data more directly. The values given are those that resulted from the fits, i.e., those corresponding to the listed model-parameter values. In the HaCaT case, the model parameters α, β, γ, σ_0, σ_1 were determined by fitting. Then P_0, P_1, P_2, g_1, g_2, β_{eff}, ϕ_1, ϕ_2, $\langle v \rangle$, $\langle v^2 \rangle$, $\langle \sigma^2(v) \rangle$ were calculated from Eqs. 34–39. In the NHDF case, α, β_0, β_1, γ, σ_0, $\sigma_{1\parallel}$, $\sigma_{1\perp}$ were determined by fitting. Then the values of $\langle v \rangle$, $\langle v^2 \rangle$ were measured in a simulation of the fitted model, and $\langle \sigma_\parallel^2(v) \rangle$, $\langle \sigma_\perp^2(v) \rangle$ were computed from these values.

As a test of the computer algorithm used in the fitting procedure, $\langle v \rangle$ and $\langle v^2 \rangle$ were also measured in a computer simulation of the HaCaT model. The validity of Eq. 39 was verified for the measured values.

SAME CELLS ON DIFFERENT SURFACES

So far, we have studied HaCaT cells crawling on substrates of collagen and NHDF cells crawling on TCPS. The qualitative differences between the two models we have found, reflect qualitative differences between the two data sets they were found from. These differences may be due to differences between the cell types, or between the surfaces, or both. We now demonstrate that the differences are associated with the cells, and not with the surfaces, by varying the surfaces.

Fig. 6 shows motility data for NHDF cells on four different surfaces: glass, TCPS, fibrillar collagen, and molecular

TABLE 1   Values for parameters in HaCaT and NHDF models fitted to data

	HaCaT	NHDF
α (1/h)	1.5	2.9
β, respectively β_0 (1/h)	4.1	2.1
β_1 (1/μm)	–	0.06
γ (1/h)	1.0	2.1
σ_0 (μm/h$^{3/2}$)	11	7.3
$\sigma_{1\parallel}$ (1/h$^{1/2}$)	1.3	2.8
$\sigma_{1\perp}$ (1/h$^{1/2}$)	1.3	0.9
β_{eff} (1/h)	3.4	–
P_0 (h)	0.51	–
P_1 (h)	0.21	–
P_2 (h)	2.4	–
g_1 (dimensionless)	0.86	–
g_2 (dimensionless)	0.14	–
ϕ_1 ((μm/h)^2)	209	–
ϕ_2 ((μm/h)^2)	94	–
$\langle v \rangle$ (μm/h)	11	15
$\langle v^2 \rangle$ ((μm/h)^2)	260	614
$\langle \sigma_\parallel^2(v) \rangle^{1/2}$ (μm/h$^{3/2}$)	31	74
$\langle \sigma_\perp^2(v) \rangle^{1/2}$ (μm/h$^{3/2}$)	31	28

Values above the line are for parameters in the HaCaT and NHDF models fitted to data in Figs. 4 and 5. Below the line are corresponding values for some quantities that characterize experimental data, but cannot be compared directly with data in figures above because the latter are finite-difference results.

collagen. Fig. 8 shows motility data for HaCaT cells on the two collagen substrates. HaCaT cells did not show measurable motility on glass and TCPS during the first 24 h of the experiment.

Comparing the *A* panels in Figs. 6 and 8 we note a qualitative difference: for HaCaT cells, the RMSD of the acceleration's two components are virtually identical functions of speed. For NHDF cells, on the other hand, the RMSD of the acceleration's component orthogonal to the velocity is almost independent of speed, whereas the RMSD of the component parallel to the velocity is strongly speed dependent. No theory is needed to tell which cell type produced the data in any one of the six *A* panels in the two figures. We have identified a cell-type-specific pattern in the data that persists across all the different surfaces we have data for. This pattern is reflected in the two models we found above, and Figs. 6 and 8 show that the two models describe the data in all six cases. It follows that the two models are cell-type specific. Only their parameter values vary between surfaces. These parameter values can then be used to characterize the surfaces.

SCALING ANALYSIS, DATA COLLAPSE, AND PARAMETERIZATION OF SURFACE DIFFERENCES

In principle, the two models put five, respectively seven, parameters at our disposal for the characterization of surfaces. In practice, these parameters may not all parameterize independent properties of surfaces. Also, even if they do, we need at least eight different surfaces to demonstrate this for

the seven parameters of the NHDF model, and six different surfaces for the five parameters of the HaCaT model. Finally, even if we had data for a sufficient number of surfaces, finite statistics might blur subtle differences. For these reasons, a phenomenological analysis of the actual information content of our data is a healthier approach. It has the additional advantage of being model independent. Thus, it is a practical tool that can describe how motile cells and microorganisms experience different surfaces without requiring a model for their motility.

NHDF data collapse

Fig. 6 shows motility data for NHDF cells on glass, on TCPS, on substrate of molecular collagen, and on substrate of fibrillar collagen. Compared to the other three surfaces, glass results in elevated scatter in the acceleration's components (panel *A*), in shorter correlations (panel *C*), and in higher velocities (panel *D*). However, the two distributions describing the acceleration's scatter about its average (panel *B*) do not differ between surfaces when measured as here, in units of the scatter's RMSD. The speed distribution looks like it may have the same property: invariant if rescaled to give the distribution of speeds in units of the average speed. Fig. 7 *B* shows quite convincingly that this is the case. So the mean speed is the only aspect of the velocity distribution of NHDF cells that will distinguish surfaces. Other aspects of the distribution is invariant between surfaces, at least with our experimental resolution and for the four surfaces studied here.

Maybe the mean speed exhausts differences between other quantities of dimension speed and acceleration. To find out, we rescaled all velocities so they are measured in units of their corresponding mean speed. Fig. 7 *A* shows how the accelerations of these dimensionless velocities average and scatter as functions of dimensionless speed. The figure shows data collapse: the mean speed captures the whole difference between surfaces as far as this plot is concerned.

Fig. 7 *C* shows the velocity autocorrelation functions in Fig. 6 *C*, but in units of mean-squared velocity, $\langle \vec{v}^2 \rangle$. This normalization, rather than $\langle v \rangle^2$, is conventional because it is more practical, as $\phi(t)/\phi(0) = 1$ for $t = 0$. These functions discriminate between surfaces in a manner that is not captured by the mean speed. This was to be expected, because another physical dimension, time, is involved.

Apart from the initial rapid decrease, the velocity autocorrelation functions in Fig. 7 *C* show a slower, exponential decrease over most of the times where they can be measured. This exponential decrease has a different characteristic time, t_0, for glass, TCPS, and collagen, whereas the two kinds of collagen cannot be distinguished. So t_0 discriminates between surfaces in a manner that might be independent of the mean speed $\langle v \rangle$.

To test for this independence, we form their product, a length. The theories presented here do not contain any natural length scales, but cells do, e.g., their size. A theory
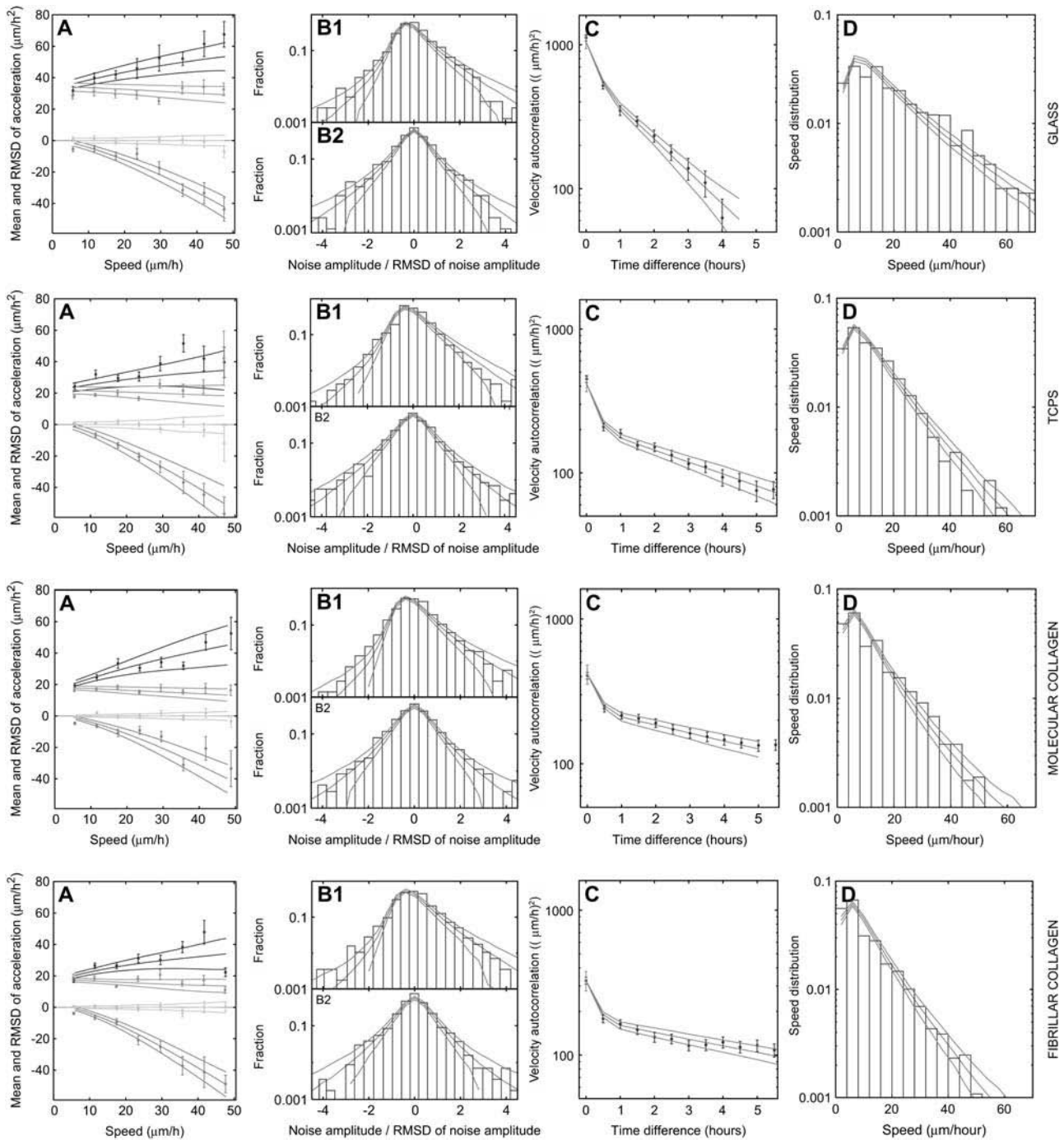
FIGURE 6   Motility data for NHDF cells on four different surfaces, from 30-min time-lapse records. (*Rows from top*) Glass, TCPS, molecular collagen, and fibrillar collagen. (*A* panels) Average acceleration and RMSD of acceleration, both as functions of speed, for directions parallel and orthogonal to velocity. Compare Fig. 5 *B*. (*B* panels) Distribution of acceleration minus mean acceleration. Panels *B1/B2* show distribution of component parallel/orthogonal to the velocity, in units of its RMSD. Compare Fig. 5 *C*. (*C* panels) Velocity autocorrelation function $\phi(t) = \langle \vec{v}(t) \cdot \vec{v}(0) \rangle$. Compare Fig. 5 *E*. (*D* panels) Speed distribution $2\pi v p(\vec{v})$ against speed *v*. Compare Fig. 5 *F1*.

might resolve such a length scale in motility data, and if it is the same on different surfaces, t_0 is just the inverse of $\langle v \rangle$ up to a constant, hence no independent discriminator. Table 2 lists t_0, $\langle v \rangle$, and their product. The product is not constant for a given cell type, so t_0 is an independent characteristic of

compatible cell-and-surface combinations. We note that the product has a value similar to the cells' size. Any radically different value would have been surprising.

Fig. 7 *D* shows the same velocity autocorrelation functions as in Fig. 7 *C*, but as functions of dimensionless relative
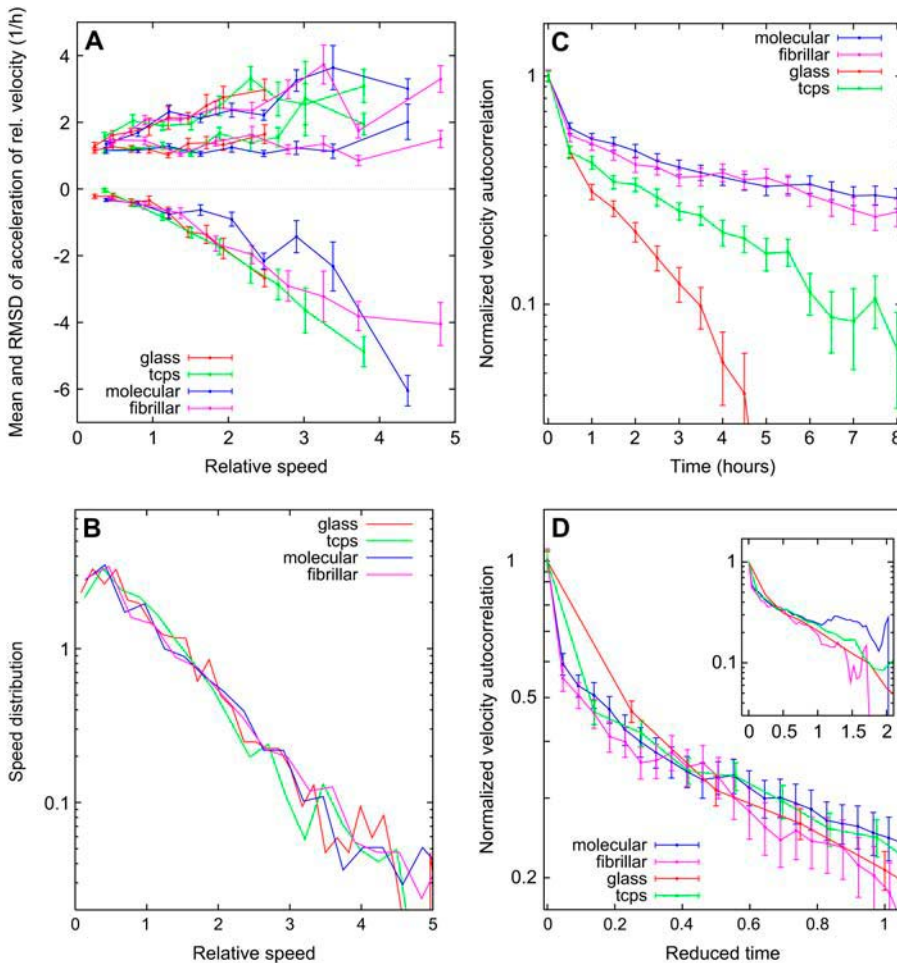
FIGURE 7 Same data as in Fig. 6, but with all velocities and accelerations given in units of the mean velocity on a given surface. (*A*) Average and RMSD of acceleration's components, as function of speed/(mean speed). (*B*) Velocity distribution as function of speed/(mean speed). (*C*) Dimensionless velocity autocorrelation functions $\phi(t)/\phi(0)$. They clearly distinguish three kinds of surfaces, the discriminator being the characteristic time of their long, exponentially decreasing tails, and clearly cannot distinguish substrates of molecular collagen from fibrillar ditto. (*D*) Same as in panel *C*, but as function of reduced time, i.e., time divided by characteristic time of tail. Plotted this way, the tails all have slope 1. The data collapse is an additional result. The normalized correlation function of reduced time cannot distinguish surfaces.

time: time in units of the characteristic time for the exponential decrease. Plotted this way, the correlation functions are guaranteed to have the same slope, except during their rapid initial decrease. We see a data collapse within underestimated error bars. Our resolution of time could be better at earlier times, especially for glass and TCPS. But for this 30-min time-lapse data, we must conclude that the data

collapse leaves no other discriminator of dimension time to be found in the velocity autocorrelation functions.

HaCaT data collapse

In the HaCaT theory, t_0 is identical to P_2. Table 1 lists a 12-times-smaller value for P_1, the shortest characteristic time

TABLE 2 Experimental values and parameter values

	HaCaT		NHDF			
	Molecular	Fibrillar	Glass	TCPS	Molecular	Fibrillar
$\langle v \rangle$ (μm/h)	11	11	25	16	14	13
t_0 (h)	3.1	1.8	2	3.6	1.1	1.1
$\langle v \rangle \, t_0$ (μm)	34	20	49	56	15	14
α (1/h)	1.7	1.6	2.2	1.9	2.3	1.8
β, respectively β_0 (1/h)	4.7	4.8	1.5	2.1	6.8	3.5
β_1 (1/μm)	–	–	0.043	0.073	0.015	0.057
γ (1/h)	1.0	1.3	1.6	0.8	0.8	0.7
σ_0 (μm/h$^{3/2}$)	11	14	33	25	24	21
$\sigma_{1\parallel}$ (1/h$^{1/2}$)	1.4	1.3	3.0	3.0	2.6	2.8
$\sigma_{1\perp}$ (1/h$^{1/2}$)	1.4	1.3	−0.02	0.006	0.29	−0.12

Values above the line are experimental values. Below the line are parameter values from fit of models to experimental data. HaCaT data obtained with 15-min time lapse; NHDF data with 30-min time lapse.

in $\phi(t)$. Its product with $\langle v \rangle$ listed in the same place, is 2.3 μm, much shorter than HaCaT cells' size, and consistent with the idea that when we resolve finer timescales, we resolve smaller spatial properties of the cells, in this case possibly their pseudopodia. Cells do move in an uneven manner, as they haul themselves forward on their pseudopodia (10).

Fig. 8 shows motility data for HaCaT cells on substrate of fibrillar collagen, and on substrate of molecular collagen. They look indistinguishable, except for the velocity auto-correlation functions. The mean speed is the same on the two surfaces, to within 4%, an insignificant difference. So no significant rescaling was done to check for data collapse. Fig. 9 A seems to show a collapse. The average acceleration as function of speed seems smaller on molecular collagen, but this may be a coincidence due to limited statistics and the fact that the data are correlated. Fig. 9 B shows a very convincing identity of p_{HaCaT} on the two surfaces.

Fig. 9 C shows the normalized velocity autocorrelation functions. They are the only aspect of data from the two surfaces that seem to distinguish the surfaces. But the statistics are poor, the data correlated, and we have less than three correlation-times worth of data, so, as indicated by the theoretical standard deviation, the two functions may not be significantly different. For this reason we merged these data to obtain better statistics in the HaCaT analysis presented in the first part of this article. If we nevertheless attempt to extract a characteristic time from each of the tails in Fig. 9 C,

and replot the correlation functions as functions of reduced time, we get the data collapse shown in Fig. 9 D.

Different or not, the HaCaT model captures the nature of these data in Figs. 8 and 4. We have also reanalyzed the 15-min time-lapse trajectories on which Fig. 4 is based, but using only every second data point, as if we did 30-min time-lapse recording. The velocity autocorrelation function is then less well resolved, but the qualitative features that distinguish the HaCaT model from the NHDF model become more manifest (not shown here): the average acceleration is distinctly linear in its speed dependence, and the RMSD of the acceleration's two components is more clearly the same function $\sigma(v)$ of the speed.

DISCUSSION

All those parameters...

With enough parameters one can fit anything. Have we sufficiently justified the five-parameter HaCaT model and the seven-parameter NHDF model? Or will other models with less (as many) parameters do as well (better)? How many parameters are really needed?

Any model that is a mathematical relationship between velocity and time, must contain at least one dimensionful parameter, because velocity and time have different dimensions. The previous section's phenomenological analysis
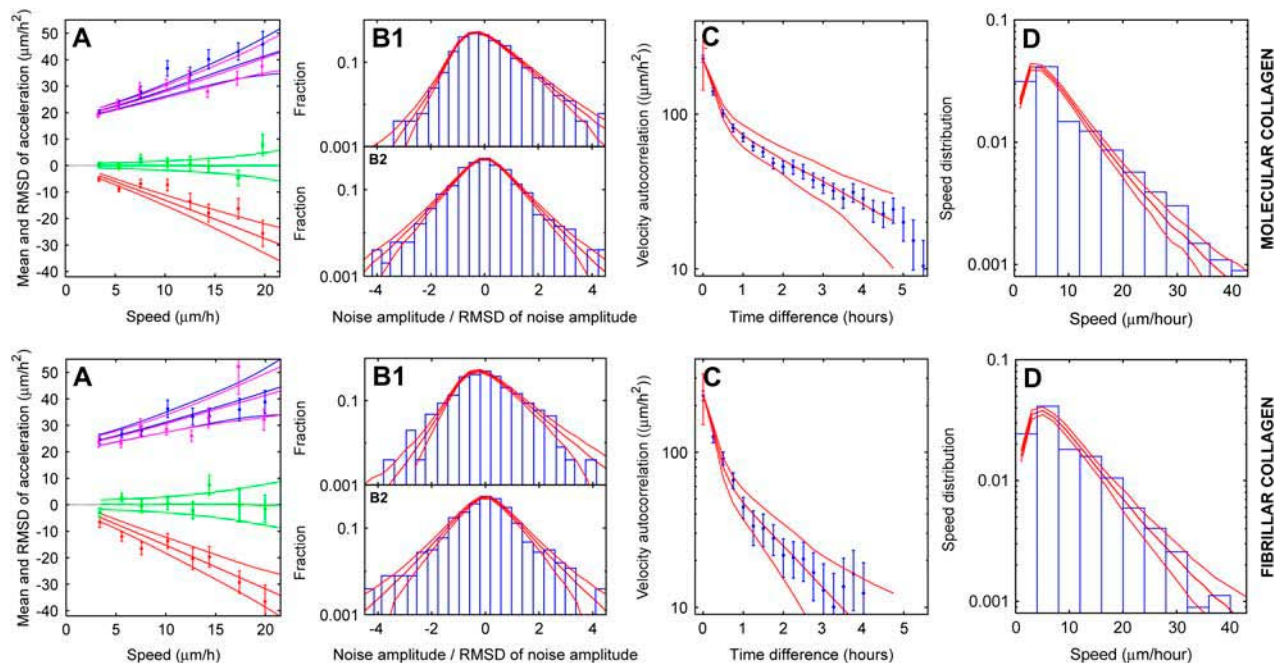

FIGURE 8 Motility data for HaCaT cells on two different surfaces, from 15-min time-lapse records. (*Rows from top*) Molecular collagen, fibrillar collagen. (*A panels*) Average acceleration and RMSD of acceleration, both as functions of speed, for directions parallel and orthogonal to velocity. Compare Fig. 4 *B*. (*B panels*) Distribution of acceleration minus mean acceleration. Panels *B1/B2* show distribution of component parallel/orthogonal to the velocity, in units of its RMSD. Compare Fig. 4 *C*. (*C panels*) Velocity autocorrelation function $\phi(t)$. Compare Fig. 4 *E*. (*D panels*) Speed distribution $2\pi v p(\vec{v})$ against speed v. Compare Fig. 4 *F1*.
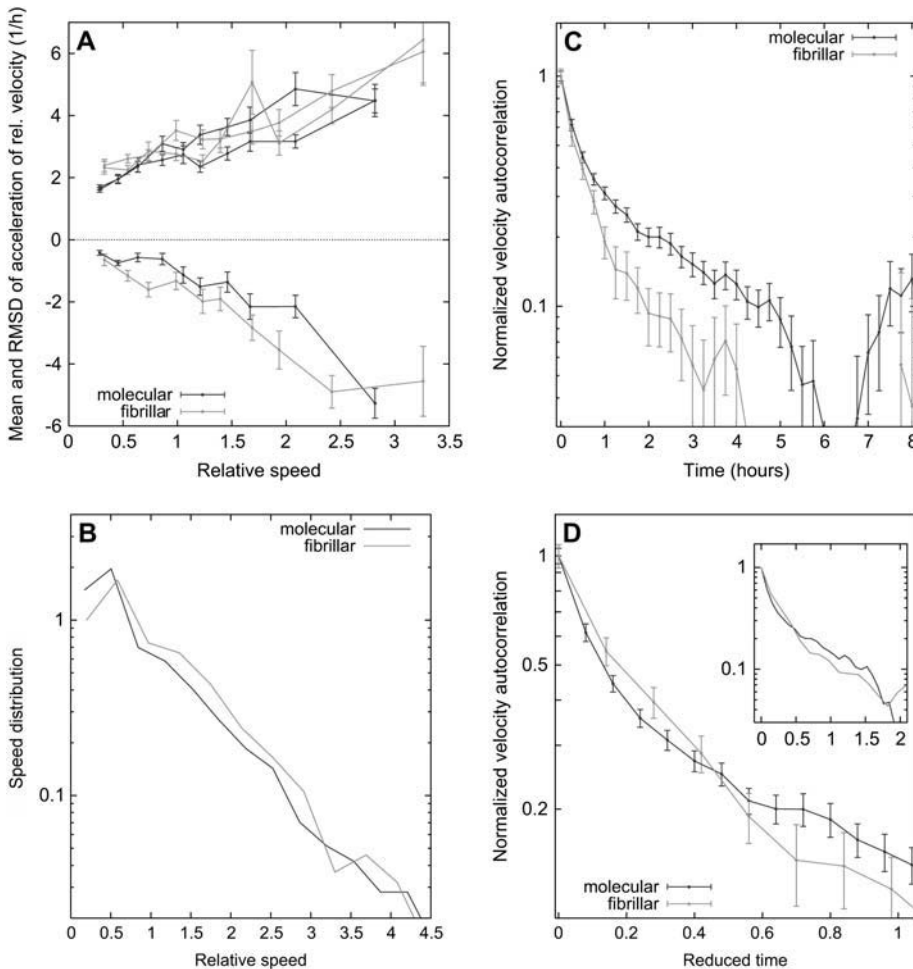
FIGURE 9 Same data as in Fig. 8, but with all velocities and accelerations given in units of the mean velocity on a given surface. (*A*) Average and RMSD of acceleration's components, as function of speed/(mean speed). (*B*) Velocity distribution as function of speed/(mean speed). (*C*) Dimensionless velocity autocorrelation functions $\phi(t)/\phi(0)$. They maybe distinguish two kinds of surfaces, the discriminator being the characteristic time of the exponentially decreasing tails. (*D*) Same as in panel *C*, but as function of reduced time, i.e., time divided by characteristic time of tail. Plotted this way, the tails all have slope 1. The data collapse is an additional result: the normalized correlation function of reduced time cannot distinguish surfaces.

revealed that NHDF data require at least two independent dimensionful parameters from a model that is to describe the data on different surfaces. The data collapse achieved in the previous section indicates that no more than two parameters are needed, because in mathematical terms the data collapse means that any relation between the variables can be written $F(\vec{v}/\langle v \rangle, t/t_0) = 0$, where F is a dimensionless functional of dimensionless variables. (The relationship F may involve differentiation and integration w.r.t. time, as it does in Eq. 6 and the NHDF version of it, and it may depend only on $v/\langle v \rangle$, as in $p(v)$, or only on t/t_0, as in $\phi(t)$.) The collapse of the HaCaT data leads to the same conclusion, except there is no experimental proof that the two dimensionful parameters are independent of each other. So a single parameter might in principle do.

If there were no data collapse, F would depend on the surface under consideration, and that would introduce additional, dimensionless parameters in F. But our 30-min time-lapse data do collapse, so we do not need such extra dimensionless parameters to model them. Data taken with better time resolution can resolve the initial rapid decrease in the velocity autocorrelation that we found in our 15-min time-lapse data. This initial decrease contains another time-

scale, and another amplitude of dimension (velocity). Divided by t_0, respectively $\langle v \rangle^2$, this additional timescale and amplitude make up two dimensionless parameters that probably will vary with surfaces. We have yet to address this question experimentally.

So when we here refer to the HaCaT model as being a five-parameter model, most of the parameters we refer to, were not forced upon us by experiments that prove the need for them by proving that they change with some external condition, such as choice of surface. They are parameters of a different, less absolute status. They parameterize the function space in which we find F, and their number may depend on our choice of basis functions in that abstract space. Suppose, e.g., that a feature is described by an exponential function, but we have not realized this, and use a polynomial description in the interval in which we observe this feature. An exponential function has an amplitude and a characteristic range, i.e., two parameters. A polynomial of degree higher than one contains more than two parameters—unnecessary extra parameters in this example, as they can be expressed as functions of the first two parameters if one is aware of the exponential nature of the feature. So when we parameterize features in the data that we model, this potential ambiguity

must be borne in mind when we count the number of parameters introduced. We do, and find an unambiguous result.

The power of symmetry

If an unknown theory must possess certain symmetry properties, the enforcement of these properties is an extremely powerful tool for narrowing down the search for this theory and for defining remaining properties that must be determined experimentally.

The heuristic, data-driven derivation of the motility models presented above may not have done justice to the symmetry properties that guided us through their derivation. So we briefly repeat them here: we demanded that theories must be covariant under spatial rotations and invariant under translations in space and time. We first checked consistency of these assumptions with data. That done, these symmetries permeated our phenomenological approach. Data were averaged and plotted in manners based on the assumption of these symmetries. Only then did we identify parameters to be fitted. In this process, it turned out that HaCaT data were consistent with isotropic noise amplitude, so that further reduction in the number of parameters was achieved with the assumption that $\bar{\bar{\sigma}}(\vec{v}) = \sigma(v)\bar{\bar{1}}$. Our NHDF data do not demonstrate this isotropy of noise amplitudes, so in this case isotropy of space only simplifies $\bar{\bar{\sigma}}(\vec{v})$ to $\sigma_{\parallel}(v)\hat{v} \otimes \hat{v} + \sigma_{\perp}(v)\check{v} \otimes \check{v}$.

Parameter counting

In the HaCaT case, space-time symmetries and the choice of simple, uncorrelated noise narrowed down potential models to those of the form given by Eq. 5. This form also assumes that the cell population studied is made up of a single kind of cells, as opposed to several subpopulations of different kinds of cells. The latter possibility is discussed below and eliminated with Occam's razor after some rigorous math.

Parameterization of the experimental result for $\phi(t)$ required four parameters, two for each of the two exponential functions with which we modeled $\phi(t)$. Exponential functions are the natural choice of basis functions for $\phi(t)$, and leave no room for further simplification: Four parameters are required.

From this assumption and Eq. 5 followed with mathematical rigor explicit expressions for K and for $\langle \sigma^2(v) \rangle$. The latter expression is the "fluctuation-dissipation theorem" for the theory, and a constraint on the choice of the otherwise undetermined function $\sigma(v)$. That function was chosen to be a first-degree polynomial inspired by the RMSD data in Fig. 4 B. A first-degree polynomial contains two parameters. So does any other nonconstant function that one might choose to describe the RMSD data in Fig. 4 B with. This follows from dimensional arguments alone: the two axes in Fig. 4 B have different dimensions, and both slope and intercept of the RMSD data are nonzero, hence must be parameterized. So no less than two parameters will do. We have already one constraint between them. Consequently, one additional

parameter must be introduced here. No more parameters are introduced in the HaCaT model, so it is a bona fide five-parameter model, and the only such model that describes the data, given the assumptions we have made: These were assumptions of maximum simplicity: symmetry, uncorrelated noise, and only one type of cell. Our HaCaT model cannot be replaced by a simpler model, nor with a model with fewer parameters—not with our HaCaT data. The only possibility that we have not considered is whether a mixed population of cells with a simpler dynamics will describe the data with better intellectual economy. This is the subject of the next subsection.

As for the NHDF model, its deterministic part is nonlinear, so no similar mathematical rigor can be achieved for it. The nonlinearity itself, β's dependence on v, is modeled with a first-degree polynomial, which introduces one additional parameter, β_1. Any other function basis will also require two parameters to describe $\beta(v)$, for dimensional reasons and because both slope and intercept of $\beta(v)$ are nonvanishing.

But the choice to make β v-dependent, while leaving α and γ constant, was ad hoc, and not based on any mathematical rigor, only mathematical simplicity and the fact that it works. The data themselves only tell us that β_{eff} is not a constant. We did try to give α, β, and γ the same dependence on v, that which β_{eff} has according to the data in Fig. 5 B. According to the "equivalent theory" in Appendix A, this is a mathematically symmetric manner to introduce v-dependence in these coefficients, and would, to the extent the definition of β_{eff} in Eq. 37 remained true also for this nonlinear theory, result in a theoretical $\beta_{\mathrm{eff}}(v)$ that by construction agrees with the data. Unfortunately, Eq. 37 does not remain true, we found, and choosing only β v-dependent results in a model that agrees with experimental data, whereas this more elegant attempt does not. Ad hoc or not, another parameter has been introduced in the model, which now contains a total of four parameters in its deterministic part.

The two noise amplitudes $\sigma_{\parallel}(v)$ and $\sigma_{\perp}(v)$ were modeled with two first-degree polynomials having the same value at $v = 0$, i.e., a total of three parameters describe the model's stochastic part. Fewer parameters will not do, also not in a different function basis. We conclude that despite space-time symmetries, the richer structure of the NHDF data forces us to use seven parameters in the NHDF model, and, worse, one of them was introduced in an ad hoc manner. Other models may describe the same data as well or better, but will hardly need fewer parameters to do that.

Why "two simpler subpopulations" is not a simpler explanation

Obviously, a cell population consisting of two subpopulations, 1 and 2, having simple exponential velocity autocorrelation functions $\phi_i(t) = \phi_i(0) \exp(-|t|/P_i)$, $i = 1, 2$, would also result in Eq. 4: let w_i denote the fraction of cells in subpopulation i. Then $w_1 + w_2 = 1$ and

$$\phi(t) = w_1\phi_1(0)e^{-|t|/P_1} + w_2\phi_2(0)e^{-|t|/P_2}. \qquad (10)$$

So how can we tell whether we have two subpopulations or not? That could easily be done if we had enough data in individual cell trajectories to calculate velocity autocorrelation functions for individual trajectories. If trajectory-specific autocorrelation functions all are simple exponentials with one of two persistence times, then we have two subpopulation. Cells could then be classified by the value of their individual persistence time, as computed from each cell's trajectory. On the other hand, if we have only one population, velocity autocorrelation functions computed from individual trajectories should all be identical, and each be a sum of the same two exponentials. Unfortunately, our data are far from rich enough to support this approach. We need to average over the cell population to get sufficient statistics.

Doing that, we ask whether two simpler motility models, one for each subpopulation, will describe the HaCaT data as well as the HaCaT model does. To answer this question, we must consider all motility models that result in a velocity autocorrelation function that is just a simple exponential. Then we limit the search to the subset of such models which have uncorrelated noise term, because the HaCaT model has that, and we are looking for simpler models. For this subset of models, we can reuse the derivation given in Appendix A. It proves that motility models with these properties are differential equations of the form

$$\frac{d\vec{v}}{dt} = -\vec{v}/P_i + \sigma_i(v)\vec{\eta}_i. \qquad (11)$$

So we use one such model for each subpopulation, each model having its own persistence time P_i and speed-dependent noise $\sigma_i(v)\vec{\eta}_i$.

The two steady-state velocity distributions $p_i(\vec{v})$ that result from these two equations, must add up to $p_{\text{HaCaT}}(\vec{v}) = w_1 p_1(\vec{v}) + w_2 p_2(\vec{v})$ that agrees with the experimental distribution in Fig. 4 D. But from two subpopulations it is very difficult to get the property $\langle d\vec{v}/dt\rangle_{\vec{v}} \propto \vec{v}$ demonstrated by the experimental data in Fig. 4 C. Equation 11 yields

$$\left\langle \frac{d\vec{v}}{dt} \right\rangle_{\vec{v}} = -\vec{v}/P_i, \qquad (12)$$

for expectation values within a single subpopulation, hence for the full population

$$\left\langle \frac{d\vec{v}}{dt} \right\rangle_{\vec{v}} = -\beta_{\text{eff}}(v)\vec{v}, \qquad (13)$$

with

$$\beta_{\text{eff}}(v) \equiv \frac{\sum_{i=1,2} w_i p_i(\vec{v})/P_i}{\sum_{i=1,2} w_i p_i(\vec{v})}. \qquad (14)$$

This expression for $\beta_{\text{eff}}(v)$ is only constant, as in the HaCaT data, if $p_1(\vec{v}) = p_2(\vec{v})$ for all values of $\vec{v}$. This is

a very special condition with precise consequences: Eq. 11, like the HaCaT model, leads to Eqs. 42 and 43 with $\beta_i = 1/P_i$ in place of β_{eff}. So we know that $p_i(\vec{v}) = p(v)$ given in Eq. 45 for $i = 1, 2$. From these two identities follows that $\sqrt{P_1}\sigma_1(v) = \sqrt{P_2}\sigma_2(v)$ for all values of v. Using this in Eq. 11, we see that the two subpopulations obey identical equations of motion, when time t is measured in units of characteristic time P_i: the two subpopulations move in identical manners, except to clocks that move with different speeds, one being a factor P_1/P_2 slower than the other.

If we assume that this very special property of the cell population is satisfied and assume the simplest possible form for $\sigma(v)$ that is allowed by the data, Eq. 8, then we are still left with a five-parameter model: P_1, P_2, σ_0, σ_1, and the ratio between the two subpopulations' sizes, are to be determined by fitting.

We could not fit this model to the data. We could fit it perfectly well to some of the data sets plotted in Fig. 4. We knew that before trying from the built-in properties of this model. But, surprisingly, we could not fit it to all of the data that we fitted the HaCaT model so well to.

Thus, we appear to have excluded those two-population models that à priori seemed able to describe the HaCaT data with five parameters. This is not a rigorous proof, but it does exclude the obvious. Because we also have no reason to believe that we have two subpopulations, Occam's razor favors the explanation that the HaCaT cell population is a single population described by the HaCaT model with memory in Eq. 6, as this model is simple, generic, and embodies only assumptions that were read off the data.

CONCLUSIONS

We recommend that motility data are plotted like we have done repeatedly above. Even if one has no intention of modeling the data, these plots are useful. As we have seen, they may expose qualitative features in the data that can be associated with the motile organism, here, cell type.

They may also point to quantitative phenomenological features in the data that may be extracted and used, both to characterize the data quantitatively, and to bring out the qualitative features that are invariant, we saw, in a data collapse. This is much sounder than blindly fitting Eq. 1 to the mean-squared displacement.

Finally, plotting as we did revealed the models that do capture the nature of the data. The manner in which these models were found is readily applied to other cells and other motile organisms. As illustrated here, it may capture qualitative aspects of motility specific to a given organism. It may also capture quantitative differences in a given organism's motility in different environments. More generally, the cell-specific nature of models found with this approach may help to distinguish intrinsic differences in cell behavior from effects caused by surroundings.

Biological origins of motility patterns are not revealed by our top-down approach. But because our approach does capture motility patterns, its models are natural targets for bottom-up explanations in terms of biological processes in the cell (30,31). These models also invite reflection over possible origins and purposes of the features they describe. For example the two different timescales of the velocity autocorrelation function beg an explanation. Does the shorter of the two timescales, e.g., describe a short-lived component in the cell's motion associated with actions of individual pseudopodia? Or the relative motion of the nucleus within the cell? Or is there a third explanation?

Also, the HaCaT model's random changes in velocity occur with almost the same amplitude in all directions. This matches the facts that keratinocytes move to find—in any direction— other, similar cells with which they connect to form sheets. In contrast, the NHDF model has stronger accelerations at low speeds and stronger decelerations at high speeds than the HaCaT model has. Thus, speeds near the mean speed are more favored in the NHDF model. Motion is also more persistent in the NHDF model, because random changes of direction occur with much lower amplitude than random changes of speed. This may relate to the fact that fibroblasts in tissue synthesize and arrange collagen fibers into oriented bundles. This activity requires the cells to move with orientational persistence, and may be optimized at a specific speed.

APPENDIX A: SOME MATHEMATICAL DETAILS

Deriving a motility model from the experimental data

Figs. 4 D and 5 D show the correlation functions for fluctuations in experimental accelerations $\vec{a}_j$ about their speed-dependent averages in Figs. 4 B and 5 B. One sees that all correlations are zero, except a few that are fairly negligible, at least in a first attempt to model the data. We consequently model these fluctuations with an uncorrelated noise $\vec{\eta}(t) \equiv d\vec{W}(t)/dt$, where the components of $\vec{W}(t)$ are uncorrelated Wiener processes (32), so

$$\langle \vec{\eta}(t) \rangle = \vec{0}; \quad \langle \eta_j(t) \eta_k(t') \rangle = \delta_{j,k} \delta(t-t'), \quad (15)$$

for all times t and t', with $\delta_{j,k}$ and $\delta(t-t')$ Kronecker's and Dirac's δ-functions, respectively.

Equation 5 is formally solved by

$$\vec{v} = g * (\sigma(v)\vec{\eta}), \quad (16)$$

where the propagator g satisfies

$$\frac{dg}{dt} = -K * g + \delta. \quad (17)$$

Here $\delta(t)$ is Dirac's δ-function, and the retarded solution must be chosen to ensure causality, i.e., $g(t) = 0$ for $t < 0$ so that

$$\vec{v}(t) = \int_{-\infty}^{t} dt' \, g(t-t') \sigma(v(t')) \vec{\eta}(t'). \quad (18)$$

The future must not affect the present, so $\vec{v}(t)$ depends only on the noise $\sigma(v(t'))\vec{\eta}(t')$ at times $t' < t$.

After Fourier transformation, Eq. 17 gives

$$\tilde{g}(\omega) = \frac{1}{\tilde{K}(\omega) - i\omega}, \quad (19)$$

where

$$\tilde{g}(\omega) = \int_{-\infty}^{\infty} dt \, e^{i\omega t} g(t) = \int_{0}^{\infty} dt \, e^{i\omega t} g(t), \quad (20)$$

and a similar expression for K. By Cauchy's theorem, causality of $g(t)$ and $K(t)$ is synonymous with $\tilde{g}(\omega)$ and $\tilde{K}(\omega)$ being analytical in the upper complex ω-half-plane where $Im \, \omega > 0$.

K is found by determining $\tilde{g}(\omega)$ from the Fourier transformed version of Eq. 4,

$$\tilde{\phi}(\omega) = \frac{2\phi_1/P_1}{P_1^{-2} + \omega^2} + \frac{2\phi_2/P_2}{P_2^{-2} + \omega^2}, \quad (21)$$

and Wiener-Khintchine theorem, which states what we find if we insert Eq. 16 in the definition of $\phi(t)$:

$$\tilde{\phi}(\omega) = n\langle \sigma^2(v) \rangle \tilde{g}(\omega) \tilde{g}^*(\omega). \quad (22)$$

Equation 19 shows that

$$\tilde{g}(\omega) \sim \frac{1}{-i\omega} \text{ for } \omega \to \infty \quad (23)$$

because we expect and demand that $\tilde{K}(\omega)$ is bounded for $\omega \to \infty$. This we do because $\tilde{K}(\omega) \propto \omega^k$ asymptotically will make the kth time derivative of $\vec{v}(t)$ appear on the right-hand side of Eq. 5, whereas K should be an integral kernel, at most as singular as a Dirac δ-function. Inserting Eq. 23 in Eq. 22 and comparing with Eq. 21, one finds the only information about $\sigma(v)$ contained in $\phi(t)$:

$$n\langle \sigma^2(v) \rangle = 2(\phi_1/P_1 + \phi_2/P_2), \quad (24)$$

which is the theory's ''fluctuation-dissipation theorem''. See also Eq. 36 below.

Fluctuation-dissipation theorems occur in statistical mechanical models quite similar in form to the one derived here. Quotation marks are nevertheless necessary when using that name here, because of differences of content: Eq. 6 is not Newton's second law, its right-hand side does not describe forces, and there is no kinetic energy to dissipate. The mathematics is the same, however, except there is no Boltzmann equilibrium statistical mechanics with a temperature that must be matched in equilibrium, so it is somewhat natural to use the name in quotation marks.

Equation 21 shows that $\tilde{\phi}$ is a rational function of ω, and the only rational function $\tilde{g}$ that reproduces $\tilde{\phi}$'s singularities and asymptotic behavior at $\omega \to \infty$ while being analytical for $Im \, \omega > 0$, is

$$\tilde{g}(\omega) = \frac{g_1}{P_1^{-1} - i\omega} + \frac{g_2}{P_2^{-1} - i\omega}, \quad (25)$$

where g_1 and g_2 are constants satisfying $g_1 + g_2 = 1$ because of Eq. 23. Inserting Eq. 25 in Eq. 22 and comparing with Eq. 21 gives

$$g_1 = \frac{P_2 - P_0}{P_2 - P_1}; \quad g_2 = \frac{P_0 - P_1}{P_2 - P_1}, \quad (26)$$

where we have introduced the notation

$$P_0 = \sqrt{\frac{\phi_1 P_1 + \phi_2 P_2}{\phi_1/P_1 + \phi_2/P_2}}. \qquad (27)$$

With $\tilde{g}$ thus known, we solve Eq. 19 for $\tilde{K}$, finding

$$\tilde{K}(\omega) = i\omega + \frac{1}{\tilde{g}(\omega)} = \beta - \frac{\alpha^2}{\gamma - i\omega}, \qquad (28)$$

where we have introduced the notation

$$\alpha = \frac{\sqrt{(P_2 - P_0)(P_0 - P_1)}}{P_1 P_2}$$
$$\beta = \frac{P_1 + P_2 - P_0}{P_1 P_2}$$
$$\gamma = \frac{P_0}{P_1 P_2}. \qquad (29)$$

Fourier transforming this result, we have

$$K(t) = \beta\delta(t) - \alpha^2\theta(t)e^{-\gamma t}, \qquad (30)$$

which, inserted in Eq. 5 gives Eq. 6.

Obviously, α, β, and γ are independent functions of P_1, P_2, and, through P_0, of ϕ_1/ϕ_2. They satisfy the inequality $\beta\gamma > \alpha^2$. This ensures that the deterministic part of Eq. 6 on the average diminishes velocities. This is necessary in order for velocities to remain finite under the influence of the noise term. When this inequality is satisfied, one can invert the relationship, and find P_1, P_2, and ϕ_1/ϕ_2 as functions of α, β, and γ. Thus, α, β, γ, σ_0, and σ_1 can be chosen as a more natural set of independent parameters for the motility model.

Equivalent theory

A small trick simplifies numerical integration of Eq. 6 as well as some analytical considerations. Rather than doing the convolution integral in Eq. 6 at each time step of its integration, we introduce the auxiliary velocity

$$\vec{V}(t) = \alpha \int_{-\infty}^{t} dt'\, e^{-\gamma(t-t')}\vec{v}(t'), \qquad (31)$$

and solve a mathematically equivalent problem of two coupled, but ordinary, differential equations,

$$\frac{d\vec{v}}{dt}(t) = -\beta\vec{v}(t) + \alpha\vec{V}(t) + \sigma(v(t))\vec{\eta}(t)$$
$$\frac{d\vec{V}}{dt}(t) = \alpha\vec{v}(t) - \gamma\vec{V}(t). \qquad (32)$$

Analytical solution of HaCaT model

The speed-dependent noise amplitude $\sigma(v)$ makes the HaCaT model nonlinear. Some of its properties can nevertheless be determined analytically. Its velocity autocorrelation function, $\phi(t)$ in Eq. 4, is one example. Because the theory was constructed over $\phi(t)$'s known form, Eq. 4, this result is not surprising. First one finds the propagator as function of α, β, γ,

$$g(t) = \theta(t)(g_1 e^{-t/P_1} + g_2 e^{-t/P_2}), \qquad (33)$$

where

$$\frac{1}{P_0} = \beta - \alpha^2/\gamma$$
$$\frac{1}{P_1} = \frac{1}{2}(\beta + \gamma + \sqrt{(\beta - \gamma)^2 + 4\alpha^2})$$
$$\frac{1}{P_2} = \frac{1}{2}(\beta + \gamma - \sqrt{(\beta - \gamma)^2 + 4\alpha^2})$$
$$g_1 = \frac{1}{2}\left(1 + \frac{\beta - \gamma}{\sqrt{(\beta - \gamma)^2 + 4\alpha^2}}\right)$$
$$g_2 = \frac{1}{2}\left(1 - \frac{\beta - \gamma}{\sqrt{(\beta - \gamma)^2 + 4\alpha^2}}\right). \qquad (34)$$

Then one has ϕ_1 and ϕ_2 as function of α, β, γ through

$$\phi_j = n\langle\sigma(v)^2\rangle\left(g_j^2 t_j/2 + g_1 g_2 P_1 P_2/(P_1 + P_2)\right), \qquad (35)$$

for $j = 1, 2$. Here, in general

$$n\langle\sigma(v)^2\rangle = 2\beta_{\text{eff}}\langle\vec{v}^2\rangle, \qquad (36)$$

where we have introduced the notation

$$\beta_{\text{eff}} = \frac{(\beta + \gamma)(\beta\gamma - \alpha^2)}{\beta\gamma - \alpha^2 + \gamma^2}, \qquad (37)$$

and thereby obtained a more recognizable form for the theory's "fluctuation-dissipation theorem". Equation 36 follows from Eq. 24 combined with $\phi(0) = \langle\vec{v}^2\rangle$ and the definitions of α, β, and γ.

Both sides of Eq. 36 depend on our choice for $\sigma(v)$, however, as well as on α, β, γ. With our choice, Eq. 8,

$$\langle\sigma(v)^2\rangle = \sigma_0^2 + 2\sigma_0\sigma_1\langle v\rangle + \sigma_1^2\langle v^2\rangle, \qquad (38)$$

where

$$\langle v\rangle = \frac{2M(3, a + 2, a)}{(a + 1)(a - 1)} v_\sigma$$
$$\langle v^2\rangle = \frac{6M(4, a + 2, a)}{(a + 1)(a - 1)(a - 2)} v_\sigma^2, \qquad (39)$$

where M is Kummer's function. These two moments are special cases of

$$\langle v^k\rangle = a B(k + 2, a - k)M(k + 2, a + 2, a)v_\sigma^k, \qquad (40)$$

for k any real number satisfying $-2 < k < a$, which follows from Eq. 9. Here B is Euler's integral of the first kind.

Isotropy and linearity of Eq. 32 ensures

$$\langle\vec{V}(t)\rangle_{\vec{v}(t)} = \xi\vec{v}(t), \qquad (41)$$

where ξ is a constant of proportionality and the subscripted brackets denote "expectation value for given value of subscript", here $\vec{v}(t)$. By taking the vector product with $\vec{v}(t)$ on both sides of this equation, and then averaging over $\vec{v}(t)$, one obtains $\langle\vec{v}\cdot\vec{V}\rangle = \xi\langle\vec{v}^2\rangle$, from which follows that $\xi = \alpha\gamma/(\beta\gamma - \alpha^2 + \gamma^2)$, hence Eq. 7. This and other useful relations are obtained by using the fact that $\langle\vec{v}(t)^2\rangle$, $\langle\vec{v}(t)\cdot\vec{V}(t)\rangle$, and $\langle\vec{V}(t)^2\rangle$ are independent of time. Taking their time derivatives and using Eq. 32, one obtains a closed, linear set of equations for these second moments with $\langle\sigma^2(v)\rangle$ as inhomogeneous term.

Equation 41 also has as consequence that the Fokker-Planck equation (33) for $p(\vec{v}, \vec{V}, t)$, the probability distribution at time t on the space of velocities $(\vec{v}, \vec{V})$, can be integrated over all velocities $\vec{V}$ to obtain a closed
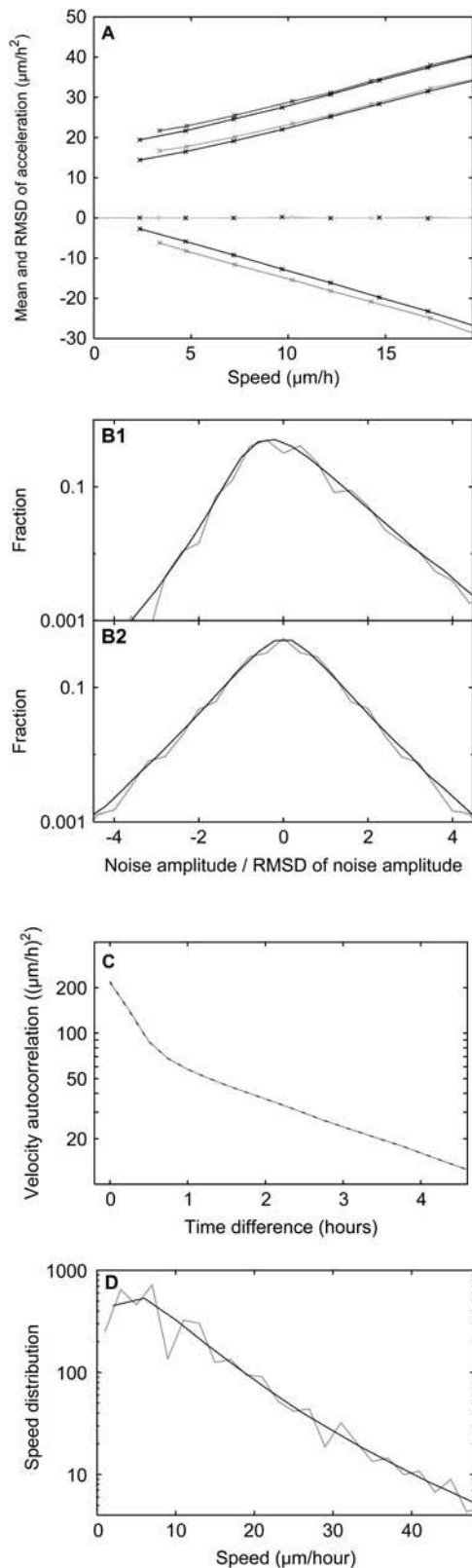
equation for $p(\vec{v}, t) \equiv \int d^n V \, p(\vec{v}, \vec{V}, t)$. Closure depends decisively on the fact that the first moment $\int d^n V \, p(\vec{v}, \vec{V}, t)\vec{V} / \int d^n V \, p(\vec{v}, \vec{V}, t) = \langle \vec{V}(t) \rangle_{\vec{v}(t)}$ is proportional to $\vec{v}$. The resulting equation reads

$$\partial p(\vec{v}, t)/\partial t = -\partial \vec{j}(\vec{v}, t)/\partial \vec{v}, \qquad (42)$$

where

$$\vec{j}(\vec{v}, t) = -\beta_{\text{eff}} \vec{v} p(\vec{v}, t) - \partial\left(\frac{1}{2} \sigma(v)^2 p(\vec{v}, t)\right)/\partial \vec{v}. \qquad (43)$$

Its stationary solution is rotation invariant, $p(\vec{v}) = p(v)$, so the corresponding current $\vec{j}(\vec{v})$ is radial, hence must vanish because of the reflecting boundary condition at $v = 0$. Thus,

$$0 = -\beta_{\text{eff}} \, v p(v) - d\left(\frac{1}{2} \sigma(v)^2 p(v)\right)/dv, \qquad (44)$$

which is solved by

$$p(v) = \frac{\beta_{\text{eff}}}{\pi \sigma^2(v)} \exp\left(-\int_0^v dv' \frac{2\beta_{\text{eff}} \, v'}{\sigma^2(v')}\right), \qquad (45)$$

where the normalization chosen corresponds to our case of interest, $n = 2$,

$$\int_0^\infty dv \, 2\pi \, v p(v) = 1. \qquad (46)$$

Insertion of $\sigma(v) = \sigma_0 + \sigma_1 v$ yields $p_{\text{HaCaT}}(\vec{v})$ in Eq. 9, whereas only $\sigma(v)$ constant, independent of v, yields a Gaussian distribution of velocities.

Note that Eq. 44 can be rearranged as

$$\sigma(v)^2/\beta_{\text{eff}} = p(v)^{-1} \int_{v^2}^\infty d(v'^2)p(v')). \qquad (47)$$

Here, the right-hand side is known experimentally from the histogram in Fig. 4 *D*. This equation shows how far the red data (''parallel mean'') take us towards finding the appropriate model. The simple proportionality between speed and mean acceleration that these data demonstrate, leads to the assumption of a linear equation of the form Eq. 5. The latter immediately gives the relationship in Eq. 7, which both confirms that the form Eq. 5 is appropriate and tells us how to determine β_{eff} experimentally. That done, $\sigma(v)$ can then be determined experimentally from the histogram in Fig. 4 *D*. The noise term assumed in Eq. 5 can be as general in form as allowed by symmetries, i.e., it can be of the form used in the NHDF model. In that case it is $\sigma_\parallel(v)$ that is determined in Eq. 47, whereas $\sigma_\perp(v)$ still must be chosen from a plot like Fig. 5 *C*.

APPENDIX B: SOME TECHNICAL DETAILS

The importance of being discrete

The effect of rounding experimental cell coordinates to the nearest pixel coordinates was accounted for by doing the same to the models' cell coordinates: cell coordinates were modeled with real-valued numbers that evolved continuously in real-valued time, when trajectories were simulated. But simulated trajectories were time-lapse sampled exactly as in experiments. The sampled coordinate values were rounded to the nearest-integer multiple of the pixel width used in the experiment with which we wanted to compare.

FIGURE 10   Relative importance of finite pixel width in HaCaT theory, demonstrated by comparing various functions measured with pixel width 0.85 μm (*gray curves*) with the same functions measured with vanishing pixel width (*black curves*) in the exact same HaCaT theory. (*A*) Same quantities as in Fig. 4 *B*. The values shown for the RMSD of the scatter orthogonal to the velocity were shifted down to prevent collapse with RMSD of scatter orthogonal to velocity. (*B* panels) Same quantities as in Fig. 4 *C*. (*C*) Same quantity $\phi(t)$ as in Fig. 4 *E*. (*D*) Same quantity $p(v)$ as in Fig. 4 *F*.
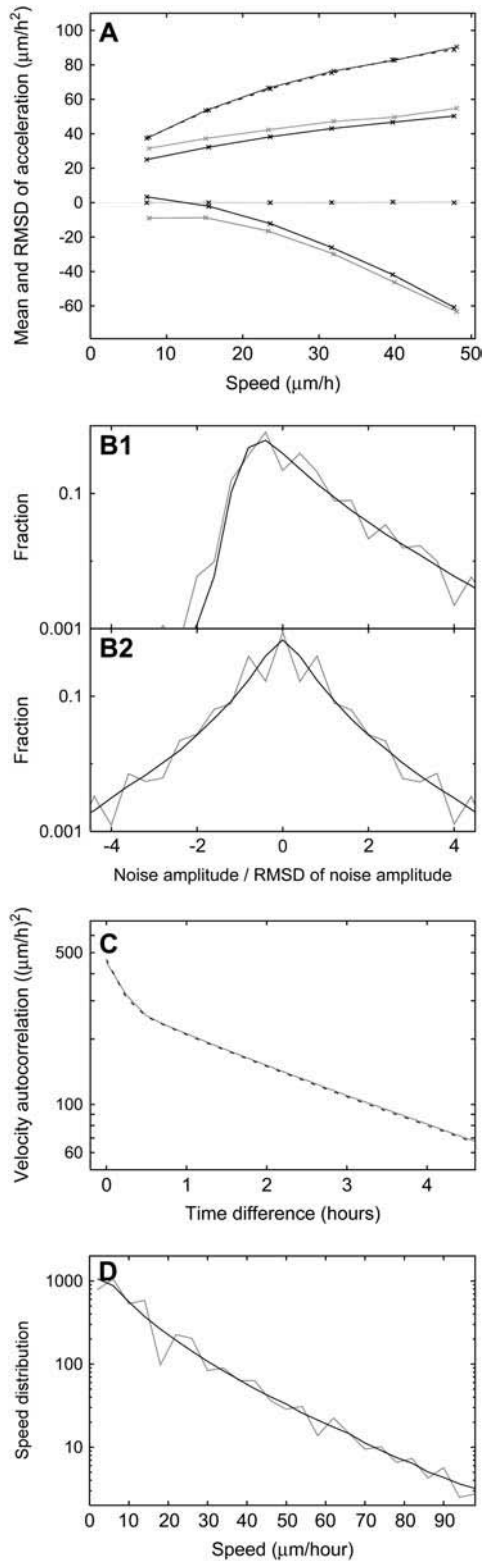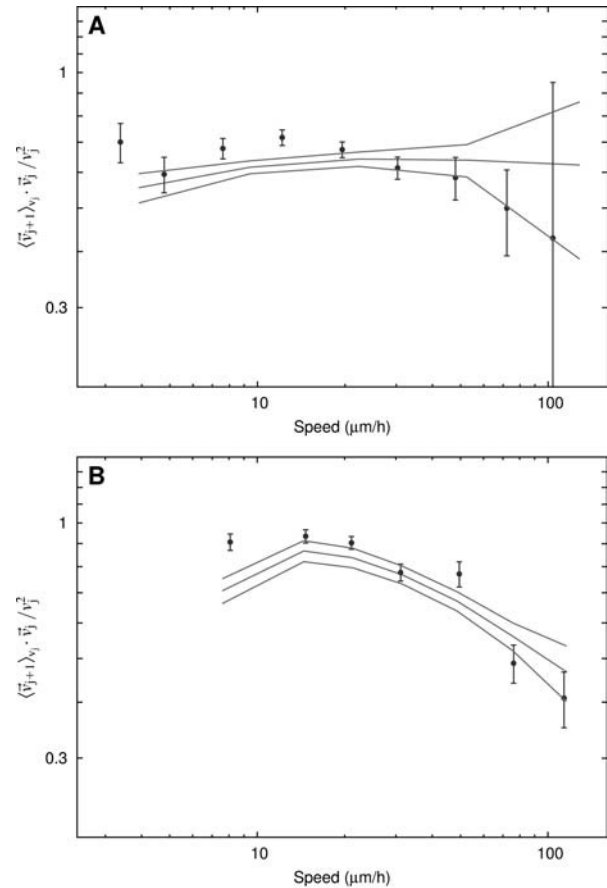
FIGURE 12 Expectation value of $\vec{v}_{j+1}$ for given $\vec{v}_j$, projected onto the latter, and measured in units of the latter's length, i.e., $\langle\vec{v}_{j+1}\rangle_{\vec{v}_j} \cdot \vec{v}_j/\vec{v}_j^2$ as function of $|\vec{v}_j|$. Here $\vec{v}_j$ and $\vec{v}_{j+1}$ are consecutive velocities computed from positions measured with 15-min time lapses; see Appendix A. In the OU model, the quantity plotted here is constant, independent of speed. (A) Data points are experimental values from HaCaT cell trajectories. Their consistency with a constant value suggests a linear velocity dependence of the noise-averaged acceleration in a HaCaT model. Thick curve surrounded by two thin curves is the theoretical result for same quantity, ± 1 SD, computed from HaCaT model in Eq. 6 fitted to data shown in Fig. 4, *B*, *E*, and *F*, and here. It's small, but nonvanishing dependence on speed, is due to a combination of time-lapse sampling and speed dependence of the model's noise term. (B) Data points are experimental values from NHDF cell trajectories. Their distinct speed dependence suggests a nonlinear velocity dependence of the acceleration in an NHDF model. Thick curve surrounded by two thin curves is the theoretical result for same quantity, ± 1 SD, computed from the NHDF model fitted to data shown in Fig. 5, *B*, *E*, and *F*, and here.

The relative importance of round-off errors varies, depending on which quantity one considers. We determined this importance for all quantities considered by recomputing all results for the fitted theories, using coordinates that had not been rounded off. Both types of results were plotted on top of each other for easy comparison; see Figs. 10 and 11.

Fig. 10 demonstrates the relative importance of round-off errors for the HaCaT data. Results for the HaCaT theory ''observed'' with finite pixel width are replotted on top of the same results for the very same theory, observed with vanishing pixel width. Finite pixel width is seen to matter most at low speeds, where round-off by a fixed amount causes the largest relative change. It does not affect the velocity autocorrelation function discernibly. It affects the average acceleration as function of speed, as well

FIGURE 11 Relative importance of finite pixel width in NHDF data (*gray curves*), demonstrated by comparison with exact same NHDF theory with vanishing pixel width (*black curves*). (A) Same quantities as in Fig. 5 *B*. No curves were shifted here. They are well separated by nature. (*B* panels) Same quantities as in Fig. 5 *C*. (*C*) Same quantity $\phi(t)$ as in Fig. 5 *E*. (*D*) Same quantity $p(v)$ as in Fig. 5 *F*.

as the two RMSDs of the acceleration's components, as functions of speed. The pattern seen in $p(v)$ in panel D is the same as the pattern of vertical stripes seen in Fig. 4 A.

Fig. 11 demonstrates the relative importance of finite pixel width in the NHDF data. Compare with Fig. 10. We see the same dependence in the importance here on the function considered and at which speed. But we note that the mean acceleration of the theory observed unhampered by finite pixel width actually is positive for speeds below 12 μm/h: NHDF cells accelerate in a deterministic manner if moving slower than that, if we are to believe this result. An indication that we are, is that we found the same acceleration when the same trajectories were analyzed with 30-min time lapse (not shown). By doing that, we reduced pixel round-off effects to a negligible level.

A better window on β

The speed dependence of the mean acceleration $\langle d\vec{v}/dt \rangle_{\bar{v}}$ can be exhibited better than done in Figs. 4 B and 5 B. The acceleration plotted there is $(\vec{v}_{j+1} - \vec{v}_j)/\Delta t$ as function of $v_j = |\vec{v}_j|$. Thus, the component parallel to $\vec{v}_j$ is $(\vec{v}_{j+1} \cdot \vec{v}_j / v_j - v_j)/\Delta t$, and its average value for given value of v_j is $(\langle \vec{v}_{j+1} \cdot \vec{v}_j \rangle_{v_j}/v_j - v_j)/\Delta t$. Because the second term in this expression always is proportional to v_j, only the first term is of real interest. Its speed dependence is brought out clearly in a plot of $\langle \vec{v}_{j+1} \cdot \vec{v}_j \rangle_{v_j}/v_j^2$ against v_j. If this quantity is constant, as the case is for the HaCaT data shown in Fig. 12 A, we limit our search for a model to models with acceleration proportional to the velocity, a great simplification. If it is not, as the case is for the NHDF data shown in Fig. 12 B, we cannot make this simplifying assumption. Instead, we must assume a more complex speed dependence of the acceleration in an NHDF model. The less-than-perfect agreement between theory and data in Fig. 12 B shows that there is room for improvement in the NHDF model we landed on when we assumed that $\beta(v)$ is not a constant, but a low-degree polynomial in v, and demonstrated that a first-degree polynomial describes data as well as a second-degree polynomial does.

REFERENCES

1. Langer, R. S., and J. P. Vacanti. 1999. Tissue engineering: the challenges ahead. *Sci. Am.* 280:86–89.

2. Hench, L. L., and J. M. Polak. 2002. Third-generation biomedical materials. *Science.* 295:1014–1017.

3. Davis, M. E. 2002. Ordered porous materials for emerging applications. *Nature.* 417:813–821.

4. Magnani, A., A. Priamo, D. Pasqui, and R. Barbucci. 2003. Cell behaviour on chemically microstructured surfaces. *Mater. Sci. Eng. C.* 23:315–328.

5. Sarikaya, M., C. Tamerler, A. K. Jen, K. Schulten, and F. Baneyx. 2003. Molecular biomimetics: nanotechnology through biology. *Nat. Mater.* 2:577–585.

6. Verrier, S., J. J. Blaker, V. Maquet, L. L. Hench, and A. R. Boccaccini. 2004. PDLLA/Bioglass composites for soft-tissue and hard-tissue engineering: an in vitro cell biology assessment. *Biomaterials.* 25: 3013–3021.

7. Ridley, A. J., M. A. Schwartz, K. Burridge, R. A. Firtel, M. H. Ginsberg, G. Borisy, J. T. Parsons, and A. R. Horwitz. 2003. Cell migration: integrating signals from front to back. *Science.* 302: 1704–1709.

8. Gaudet, C., W. A. Marganski, S. Kim, C. T. Brown, V. Gunderia, M. Dembo, and J. Y. Wong. 2003. Influence of type I collagen surface density on fibroblast spreading, motility, and contractility. *Biophys. J.* 85:3329–3335.

9. Ponti, A., M. Machacek, S. L. Gupton, C. M. Waterman-Storer, and G. Danuser. 2004. Two distinct actin networks drive the protrusion of migrating cells. *Science.* 305:1782–1786.

10. Lauffenburger, D. A., and A. F. Horwitz. 1996. Cell migration: a physically integrated molecular process. *Cell.* 84:359–369.

11. Joanny, J. F., F. Jülicher, and J. Prost. 2003. Motion of an adhesive gel in a swelling gradient: a mechanism for cell locomotion.. *Phys. Rev. Lett.* 90:168102.

12. Berg, H. C., and D. A. Brown. 1972. Chemotaxis in Escherichia coli analysed by three-dimensional tracking. *Nature.* 239:500–504.

13. Berg, H. C. 2000. Motile behavior of bacteria. *Phys. Today.* 53:24–29.

14. Berg, H. C. 2003. *E. coli* in Motion. Springer-Verlag, New York.

15. Gerbal, F., P. Chaikin, Y. Rabin, and J. Prost. 2000. An elastic analysis of *Listeria monocytogenes* propulsion.. *Biophys. J.* 79:2259–2275.

16. Shenderov, A. D., and M. P. Sheetz. 1997. Inversely correlated cycles in speed and turning in an ameba: an oscillatory model of cell locomotion. *Biophys. J.* 72:2382–2389.

17. Ornstein, L. S. 1919. On the Brownian motion. *Proc. Amst.* 21:96–108.

18. Fürth, R. 1920. Die Brownsche Bewegung bei Berücksichtigung einer Persistenz der Bewegungsrichtung. Mit Anwendungen auf die Bewegung lebender Infusorien. [in German]. *Z. Physik.* 2:244–256.

19. Gail, M. H., and C. W. Boone. 1970. The locomotion of mouse fibroblasts in tissue culture. *Biophys. J.* 10:980–993.

20. Dunn, G. A., and A. F. Brown. 1987. A unified approach to analysing cell motility. *J. Cell Sci. Suppl.* 8:81–108.

21. Maheshwari, G., and D. A. Lauffenburger. 1998. Deconstructing (and reconstructing) cell migration. *Microsc. Res. Tech.* 43:358–368.

22. Bergman, A. J., and K. Zygourakis. 1999. Migration of lymphocytes on fibronectin-coated surfaces: temporal evolution of migratory parameters. *Biomaterials.* 20:2235–2244.

23. Ionides, E. L., K. S. Fang, R. R. Isseroff, and G. F. Oster. 2004. Stochastic models for cell motion and taxis. *J. Math. Biol.* 48:23–37.

24. Gadegaard, N., S. Mosler, and N. B. Larsen. 2003. Biomimetic polymer nanostructures by injection molding. *Macromol. Mater. Eng.* 288:76–83.

25. Hartman, R. S., K. Lau, W. Chou, and T. D. Coates. 1994. The fundamental motor of the human neutrophil is not random—evidence for local non-Markov movement in neutrophils. *Biophys. J.* 67:2535–2545.

26. Uhlenbeck, G. E., and L. S. Ornstein. 1930. On the theory of Brownian motion. *Phys. Rev.* 36:823–841.

27. Czirók, A., K. Schlett, E. Madarász, and T. Vicsek. 1998. Exponential distribution of locomotion activity in cell cultures. *Phys. Rev. Lett.* 81:3038–3041.

28. Upadhyaya, A., J. P. Rieu, J. A. Glazier, and Y. Sawada. 2001. Anomalous diffusion and non-Gaussian velocity distribution of Hydra cells in cellular aggregates. *Physica A.* 293:549–558.

29. Thurner, S., N. Wick, R. Hanel, R. Sedivy, and L. Huber. 2003. Anomalous diffusion on dynamical networks: a model for interacting epithelial cell migration. *Physica A.* 320:475–484.

30. Chicurel, M. 2002. Cell migration research is on the move. *Science.* 295:606–609.

31. Pollard, T. D. 2003. The cytoskeleton, cellular motility and the reductionist agenda. *Nature.* 422:741–745.

32. Karatzas, I., and S. E. Shreve. 1991. Brownian motion and stochastic calculus, 2nd Ed. Springer-Verlag, Berlin, Germany.

33. Van Kampen, N. G. 2001. Stochastic Processes in Physics and Chemistry. Elsevier Science, Amsterdam, The Netherlands.