Empirical and theoretical analysis of particle diffusion in mucus

Antonio Cobarrubia,[1,2,*] Jarod Tall,[1,2,†] Austin Crispin-Smith,[1,2,‡] and Antoni Luque[1,3,4,§]

[1] *Viral Information Institute, San Diego State University, San Diego, CA 92182, USA*

[2] *Department of Physics, San Diego State University, San Diego, CA 92182, USA*

[3] *Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182, USA*

[4] *Computational Science Research Center, San Diego State University, San Diego, CA 92182, USA*

(Dated: November 9, 2021)

Abstract

Mucus is a complex fluid that coats multiple organs in animals. Various physicochemical properties can alter the diffusion of microscopic particles in mucus, impacting drug delivery, virus infection, and disease development. The simultaneous effect of these physicochemical properties in particle diffusion, however, remains elusive. Here, we analyzed 106 published experiments to identify the most dominant factors controlling particle diffusion in mucus. The effective diffusion—defined using a one-second sampling time window across experiments—spanned seven orders of magnitude, from 10^{-5} to 10^{2} $\mu m^2/s$. Univariate and multivariate statistical analyses identified the anomalous exponent (the logarithmic slope of the mean-squared displacement) as the strongest predictor of the effective diffusion, revealing an exponential relationship that explained 89% of the variance. A theoretical scaling analysis revealed that the stronger correlation of the anomalous exponent over the generalized diffusion constant occurs for sampling times two orders of magnitude larger than the characteristic molecular (or local) displacement time. This result predicts that, at these time scales, the molecular properties controlling the anomalous exponent, like particle-mucus unbinding times or particle-to-mesh size ratio—which depend on the underlying subdiffusion's mechanism—would be the most relevant physicochemical factors involved in the passive microrheology of particles in mucus. Our findings contrast with the fact that only one-third of the studies measured the anomalous exponent, and most experiments did not report the associated molecular properties predicted to dominate the motion of particles in mucus. The theoretical foundation of our work can be extrapolated to other systems, providing a guide to identify dominant molecular mechanisms regulating the mobility of particles in mucus and other polymeric fluids.

Keywords: mucus, microscopic particles, meta-analysis, random forest, subdiffusion, scaling analysis.

INTRODUCTION

Mucus is a complex fluid secreted by animals (Spagnolie 2015; Krajina *et al.* 2017). It protects organs against the invasion of pathogens and promotes the interaction with commensal microbes (Bäckhed *et al.* 2005; Bakshani *et al.* 2018; Silveira and Rohwer 2016). The diffusivity of particles in mucus is paramount for animal health. The infectivity of animal viruses, like HIV, decreases when their diffusion in mucus is impeded (Lai *et al.* 2009). On the flip side, enhancing the diffusivity of biomolecules in mucus facilitates the delivery of medical drugs in the body (Cone 2009), and reducing the diffusivity of commensal viruses that infect bacteria in the gut can increase their infectivity and protection against pathogens (Barr *et al.* 2013, 2015). Multiple factors modify the diffusivity of particles in mucus, including particle size (Cone 2009; Amsden and Turner 1999), particle charge and ionic strength (Abdulkarim *et al.* 2015; Arends *et al.* 2013; Hansing *et al.* 2016; Lieleg *et al.* 2010; Li *et al.* 2013), pH (Spagnolie 2015; Lai *et al.* 2009; Lieleg *et al.* 2010; Celli *et al.* 2009; Suk *et al.* 2011), and the concentration and polymeric organization of the characteristic glycoproteins in mucus called mucins (Barr *et al.* 2015). However, the combined effect of these physicochemical factors controlling particle diffusion in mucus remains puzzling.

Small biomolecules and biomolecular complexes tend to diffuse more readily through mucus, while larger particles are caught in the mucin network (Cone 2009; Amsden and Turner 1999). On the other hand, non-adhesive polystyrene particles with a diameter of 500 nm diffuse faster than smaller particles (200 nm) of the same type (Lai *et al.* 2007). These contrasting results highlight the major impact of parameters other than size in particle diffusion in mucus. In fact, neutrally charged particles display higher diffusivity in mucus than charged particles of the same size with a net negative charge (Abdulkarim *et al.* 2015; Arends *et al.* 2013; Hansing *et al.* 2016; Lieleg *et al.* 2010; Li *et al.* 2013). An increase in salt concentration shields charged particles, leading to diffusivities similar to neutrally charged particles (Arends *et al.* 2013; Hansing *et al.* 2016; Lieleg *et al.* 2010). Low pH increases the distribution of negative charges in mucins and alters mucus' viscoelasticity, reducing the diffusivity of most particles (Spagnolie 2015; Lai *et al.* 2009; Lieleg *et al.* 2010; Celli *et al.* 2009; Suk *et al.* 2011). Low pH also thickens mucus, reducing the diffusion and infection rate of viruses like HIV (Lai *et al.* 2009). Interaction with mucins also alters the diffusivity of particles in mucus. Commensal viruses that infect bacteria and reside in the gut display

immunoglobulin-like domains that are attracted to mucins. This interaction reduces their diffusivity and increases their infectivity against bacteria (Barr *et al.* 2013, 2015). Some of these observations may seem contradictory. However, the fact that mucus has been selected across animals suggests that there could be a more comprehensive emerging effect when these different physiochemical factors are combined (Lang *et al.* 2016).

To tackle this problem, we performed a meta-analysis of published experimental data measuring the passive diffusion of microscopic particles in mucus. The correlation between physicochemical properties and particle diffusion in mucus was investigated using univariate and multivariate correlation methods. A theoretical scaling analysis was applied to derive a theoretical framework justifying the empirical results. This framework provided a quantitative understanding of the regulation and control of particle diffusion in mucus and other hydrogels. Our findings predict an effective particle size (and diffusion threshold) where the anomalous exponent becomes dominant, anomalous exponent values for experiments that did not measure it, and molecular factors associated with the anomalous exponent that were not reported in most experiments but should be paramount to understand particle diffusion in mucus.

MATERIALS AND METHODS

Data extraction

We screened twenty-four published articles that reported diffusion of particles in mucus or mucus-like hydrogels (Source Data 1). Ten studies contained diffusion data for microscopic, spherical particles that could be compared at the same sampling time window (Lai *et al.* 2009; Barr *et al.* 2015; Abdulkarim *et al.* 2015; Lieleg *et al.* 2010; Suk *et al.* 2011; Lai *et al.* 2007; Olmsted *et al.* 2001; Newby *et al.* 2017; Schuster *et al.* 2013; Yildiz *et al.* 2015). WebPlotDigitizer (Rohatgi 2019) was used to extract 106 measurements of the effective diffusion, D_{eff}, measured at a time window of one second, $\Delta t_{eff} = 1$ s, that is,

$$D_{eff} = \frac{\langle MSD \rangle}{2k\,\Delta t_{eff}}\ , \tag{1}$$

where $\langle MSD \rangle$ was the ensemble mean squared displacement of a particle, and k was the dimensions of the particle diffusivity (Huang *et al.* 2013). The following variables were obtained in all the experiments analyzed: particle hydrodynamic diameter (d), particle type,

mucus source, dominant mucin expression, and temperature (T). If a study did not report the temperature explicitly, room temperature (298 K) was assumed. The following variables were obtained or derived when possible: anomalous diffusion exponent (α), particle effective surface charge (ζ), mucus pH, mucus salt concentration, and mucin concentration. The anomalous exponent—also known as the logarithmic slope of the mean-squared displacement in the microrheology community (McGlynn *et al.* 2020)—was obtained from the subdiffusion equation:

$$\langle MSD(\Delta t)\rangle = 2kD_{\alpha}\Delta t^{\alpha} \quad . \tag{2}$$

Here, D_{α} is the generalized diffusion, and Δt is the sampling time window (Barkai *et al.* 2012; Metzler *et al.* 2014; Hou *et al.* 2018). It is important noting that subdiffusion is predicted to be a transient regime in viscoelastic fluids (Grimm *et al.* 2011); at very short time scales, particles' motion is dominated by a ballistic motion and by non-anomalous diffusion at long time scales. Nonetheless, subdiffusion is a relevant phenomenon observed in mucus and other polymeric fluids at a range of time scales important in biological systems, from miliseconds to days (Cone 2009; Grebenkov *et al.* 2013; Chew *et al.* 2014). The experiments analyzed in this study fall within this time window where subdiffusion can be important.

Some references shared the measured diffusion relative to the particle diffusion in water; in these cases, the particle diffusion in water was inferred applying the Stokes-Einstein equation, using the reported temperature and hydrodynamic particle diameter (Cruickshank Miller 1924). It is worth noting that the Stokes-Einstein relation was used only to infer particle diffusion in water. Particle diffusion in mucus is better described using the generalized diffusion equation due to mucus' viscoelastic properties (Spagnolie 2015). However, neither the Stokes-Einstein relation nor the generalized diffusion equation was applied to obtain the particle's effective diffusion in mucus. All the diffusion values in mucus were empirical and independent of theoretical assumptions regarding the generalized diffusion equation. The particle types were defined as a qualitative measure of particle-mucin bonds: COOH (carboxylated), PEG (pegylated), virus, amine, antibody/protein, or chitosan. The full data set containing the measurements collected in all experiments are available in Source Data 2.

The dominant mucin composition from each mucus source was obtained by evaluating the expression levels of mucin genes from the genome bioinformatics portal Ensembl (Zerbino *et al.* 2018). Mucins were identified assuming the tissue/organ associated with each mucus or closely associated tissues. Expression levels were collected by taking the average of the

reported median of transcript per million (TPM) RNA-sequence and the most explicitly stated low, medium, and high expression levels. Based on potential gene expression of mucins with reported levels below cutoff, TPM measured below the minimum (0.05 TPM) was distinguished from experiments with no data due to possible gene expression. Low, medium, and high expression levels were obtained over reports of below cutoff in the same tissue. The dominant mucin was determined by the highest expression level then, if necessary, by the highest average of median TPM. Identification of mucin expression based on tissues was associated with each mucus: human respiratory mucus and human cystic fibrosis mucus were associated with the human lung mucin genes; human cervical mucus and cervicovaginal mucus were associated with human cervix or uterus mucin genes; pig intestinal mucus was originally from jejunum part of the small intestine, however, due to a lack of reports for jejunum tissue, the associated mucin genes were taken as the average of the median of TPM of pig duodenum and pig ileum parts of the small intestine based on the close proximity to the jejunum; pig ileum intestinal mucus were associated with ileum tissue mucin genes; pig gastric mucus was collected from pig stomach mucin genes. Source Data 3 contains the full data set obtained from the bioinformatic analysis.

Statistical analysis

The multivariate analysis was performed using the non-parametric statistical method random forest, estimating permutation p-values in R, using the *rfpermute* package (Archer 2019). Random forest is a statistical learning method that relies on generating an average ensemble of random decision trees (James *et al.* 2013). Here, the effective diffusion was used as the supervised variable for the regression of the random forest algorithm using the rest of the variables as inputs. The Mean-square error (%MSE) was used to identify the importance of each variable as a predictor. The selection of the most relevant variables was obtained applying random forest in two consecutive rounds, discarding the variables that were not statistically significant in each round (p-value > 0.05). The average percentage increase of mean-square error (%MSE) was obtained by investigating permutations of three variables at a time. These permutations assessed if the p-value obtained was robust.

The single-variable correlation analysis was performed using the non-parametric Spearman's correlation coefficient and parametric linear regressions minimized by the least-squares

method. The effective diffusion was used as the predicted variable and compared with all the other variables as predictors. The linear regressions explored logarithmic and non-logarithmic scales for both the predictor and predicted variables. The values of best fit in the linear regression provided an average response and were compared with the mid-values from the measured physical factors for consistency.

Theoretical analysis

A scaling ansatz was applied to the subdiffusion equation, Eq. (2), to extract the explicit dependence on the anomalous diffusion α. This theoretical analysis assumed that the microscopic motion of a particle was associated with a characteristic molecular mobility time scale, t_D, and displacement scale, L_D. This first-order approximation aimed to identify the scaling relationship between the generalized diffusion and these microscopic observables. The rationale and explicit derivation of this theoretical analysis is included in the results section. Logarithmic derivatives of the effective diffusion were calculated to estimate the impact of each of these three physical parameters—α, t_D, L_D—in the rate of change of the effective diffusion. This determined the threshold condition where the anomalous diffusion, α, was predicted to dominate statistically over the other factors.. This theoretical prediction was compared with the empirical values obtained from the empirical statistical analysis.

RESULTS

Particles' effective diffusion in mucus spanned seven orders of magnitude.

The microscopic particles studied had diameters, d, covering three orders of magnitude, from 1 nm to 1300 nm, and they displayed an effective diffusion spanning seven orders of magnitude, from 10^{-2} μm^2/s to 10^5 μm^2/s (see Table I and Source Data 2). The anomalous exponent, α, ranged from strongly subdiffusive ($\alpha \approx 0.1$) to purely diffusive ($\alpha \approx 1$), but it was obtained only from a third of the dataset (n = 39; 37%). The zeta potential, ζ, measured the effective surface charge of particles in solution (Kumar and Dixit 2017). The values ranged from -70 mV to $+40$ mV and were obtained for half of the dataset (n = 57; 52%). The temperature range was narrow, 295 K to 310 K. The pH ranged from mildly acidic (pH = 3.0) to slightly alkaline (pH = 7.4); but most particle types were measured at a fixed

pH (Source Data 2). A third of the experiments had been conducted in artificial mucus-like hydrogels. The rest of the experiments had been conducted in mucus from four sources: human respiratory, human cervix, pig gastric, and pig intestines. The dominant mucins were MUC2 (n=59; pig intestines and pig stomach sources), MUC5B (n=30; human cervix source), and MUC5AC (n=14; human lung source). See Source Data 3 for the extended outputs of the dominant mucin analysis.

The anomalous exponent displayed the strongest correlation with the effective diffusion.

The random forest analysis selected five significant variables affecting the effective diffusion (window sampling time 1 sec), D_{eff} (Figure 1). The anomalous diffusion exponent, α, was the most relevant variable predicting effective diffusion in the random forest model, with an average percentage increase in mean square error (%MSE) of 22($\pm$3)% (std. dev.) (p-value = 0.0099). The second most relevant variable was particle type with 19($\pm$7)% in %MSE, followed by zeta potential with 15($\pm$5)%, mucus source with 13($\pm$7)%, and dominant mucin with 10($\pm$8)%.

When analyzing the selected variables individually, the anomalous diffusion exponent, α, displayed a significantly stronger correlation with the effective diffusion, D_{eff}, than the other variables. The non-parametric Spearman correlation was $\rho = 0.93$ (p-value $< 2.0 \cdot 10^{-16}$, n = 39) (Table S.1). The second strongest individual variable was the negative zeta potential, $\zeta < 0$ ($\rho = 0.6$, p-value = 0.0002, n = 36). See Table S.1 for the outputs of all correlations.

The effective diffusion, D_{eff}, increased exponentially with α (Figure 2a). The linear regression for the semi-logged data (log-linear axes) explained 89% of the variance ($\log_{10} D_{eff} = a\,\alpha + b$, $a = 5.3 \pm 0.3$, $b = -5.0 \pm 0.2$, $R^2 = 0.89$). The anomalous exponent was extracted for 37% (n=39) of the data, in particular, carboxylated, pegylated, and viral particles. The inverse statistical model was fitted to this dataset ($\alpha = a' \log_{10} D_{eff} + b'$, $a' = 0.17 \pm 0.01$, $b' = 0.92 \pm 0.02$, $R^2 = 0.88$, n=39) to estimate the mean value of α for the remaning 63% (n = 67) of the data, corresponding to amine, chitosan, antibodies, and proteins particles (Figure 2b). Particles with effective diffusion above $D_{eff} > 3.5$ $\mu m^2/s$ (n = 21) were predicted to display regular diffusion ($\alpha = 1$); none of the particles analyzed self-propelled, and, thus, superdiffusion ($\alpha > 1$) was discarded. The majority of particles

displaying regular diffusion ($\alpha \approx 1$) corresponded to human proteins (n = 18) and two viruses, Norwalk virus and human papilloma virus (HPV).

The anomalous exponent's correlation is significant at time scales two orders of magnitude larger than the microscopic displacement time.

To elucidate the physical origin of the dominance of the anomalous exponent, α, its relationship with the effective diffusion, D_{eff}, was derived from Eqs. (1) and (2):

$$D_{eff} = \frac{D_\alpha}{\Delta t_{eff}} \Delta t_{eff}^\alpha \quad . \tag{3}$$

D_{eff} displays an explicit exponential dependency with α in the factor Δt_{eff}^α and an implicit dependency through the generalized diffusion coefficient, (D_α). The form of D_α depends on the specific underlying subdiffusion mechanism (Metzler *et al.* 2014; Joiner *et al.* 2019). Our meta-analysis contained a broad range of data (Table I), including particles with different chemistry, mucus of different types, different physicochemical conditions, and independent groups carrying different experimental implementations. Therefore, the functionality of D_α with α was not obvious, and Eq. (3) was not sufficient to justify the dependence and dominance of α in determining the effective diffusion of particles in mucus. To understand this phenomenon, D_α was further scrutinized.

The units of D_α depend on α, Eq. (2). In our study, these were $\mu m^2/s^\alpha$. Like any other physical quantity, α has an associated uncertainty (error or standard deviation) (Taylor 1997). Thus, the units of D_α are uncertain. In other words, the generalized diffusion coefficient is not a measurable physical quantity. The fact that D_α is not a physical quantity has been previously overlooked and mandates a revision of the classic subdiffusion equation, Eq. (2).

To reformulate the subdiffusion equation, the following ansatz was introduced. It was assumed that the diffusion emerges as the stochastic repetition of a particle's local physical motion with a characteristic displacement, L_D. This displacement is the consequence of a velocity v_D propelling the particle during a characteristic time, t_D:

$$L_D \sim v_D \, t_D \quad . \tag{4}$$

This is a general formulation independent on the underlying physical mechanism responsible for the particle's mobility. Other characteristic scales might play a role in the anomalous

exponent, α, as exemplified in the Discussion section. This led to the relationship

$$D_\alpha = \frac{L_D^2}{t_D^\alpha} \quad . \tag{5}$$

This ansatz was combined with the classic subdiffusion equation, Eq. (2), obtaining:

$$\langle MSD(\Delta t) \rangle = 2kL_D^2 \left(\frac{\Delta t}{t_D} \right)^\alpha \quad . \tag{6}$$

This reformulated subdiffusion equation is valid for time windows, Δt, larger than the characteristic mobility time scale, t_D, that is, $\Delta t \gg t_D$. For smaller time windows, the underlying mobility mechanism will dominate, requiring a different formulation for the displacement (Joiner *et al.* 2019).

The reformulated subdiffusion equation, Eq. (6), was combined with the definition of the effective diffusion, Eq. (1), obtaining

$$D_{eff}(\alpha) = \frac{L_D^2}{\Delta t_{eff}} \left(\frac{\Delta t_{eff}}{t_D} \right)^\alpha \quad . \tag{7}$$

The effective diffusion, thus, depends exponentially on the anomalous diffusion exponent, α, justifying the empirical relationship observed for the effective diffusion of particles in mucus (Figure 2**a**). The characteristic displacement, L_D, and time scale, t_D, depend on the specific physical mechanism responsible for the diffusion. Therefore, experiments using different particles and mucus properties are expected to introduced a variance in these two magnitudes, justifying the data dispersion in Figure 2**a**.

To determine the conditions that select α over L_D and t_D as the parameter with the strongest correlation with D_{eff} across multiple scales, the logarithm of Eq. (3) was investigated:

$$\log D_{eff}(\alpha) = \log \frac{L_D^2}{\Delta t_{eff}} + \alpha \log \frac{\Delta t_{eff}}{t_D} \quad . \tag{8}$$

For a fix time window, Δt_{eff}, the rate of change of D_{eff} with respect α is

$$\frac{\partial \log D_{eff}}{\partial \alpha} = \log \frac{\Delta t_{eff}}{t_D} \quad . \tag{9}$$

The impact of L_D and t_D were evaluated using the logarithms of L_D and t_D to obtain results valid across scales and independent of measuring units, respectively,

$$\frac{\partial \log D_{eff}}{\partial \log L_D} = 2 \tag{10}$$

and

$$\frac{\partial \log D_{eff}}{\partial \log t_D} = -\alpha \ . \tag{11}$$

The change with respect the length scale, L_D, was constant and equal to two, Eq. (10). The change with respect the time scale, t_D, was smaller than one in absolute value, Eq. (11), because the anomalous exponent had an upper limit of one, $\alpha \leq 1$ (in the experiments analyzed there were no self-propelled particles or active transport mechanisms that could display superdiffusion). Eqs. (9), (10), and (11) predict that the anomalous diffusion is the physical parameter with the strongest correlation in determining the rate of change in the effective diffusion,

$$\frac{\partial \log D_{eff}}{\partial \alpha} > \frac{\partial \log D_{eff}}{\partial \log L_D} > \left| \frac{\partial \log D_{eff}}{\partial \log t_D} \right| \ , \tag{12}$$

for sampling time windows two orders of magnitude larger than the characteristic mobility time scale,

$$\frac{\Delta t_{eff}}{t_D} > 10^2 \ . \tag{13}$$

This result applies to any physical system as far as the diffusion is the consequence of a local characteristic physical motion.

The theoretical ansatz is consistent with the statistical analysis.

The predictions from the reformulated subdiffusion equation were investigated for the particle diffusion in mucus data. The slope and intercept obtained from the linear regression in Figure 2**a** were interpreted with respect Eq. (8). The values of best fit represent an average response and were compared to the mid-values from the physicochemical factors to test the consistency of the ansatz, Eq. (5). The mean characteristic length and time scales obtained statistically were $L_D \approx 3$ nm and $t_D \approx 5$ μs, respectively. The sampling time window was $\Delta t_{eff} = 1$ s. Therefore, $\Delta t_{eff}/t_D \approx 10^6 \gg 10^2$, satisfying the inequality establishing the condition for the strong correlation of the anomalous diffusion exponent, Eq. (13). This implies that the experimental conditions investigating particle diffusion in mucus were on the regime where the anomalous exponent, α, was predicted theoretically to be the dominant factor determining the particle's effective diffusion, D_{eff}, Eq. (13).

To further confirm the consistency of the theoretical framework with the empirical data, it was necessary to justify that the mean values obtained from the linear regression of Eq. (8), that is, $L_D \approx 3$ nm and $t_D \approx 5$ μs, were physically sounding. Regardless of the physicochemical factors in mucus controlling α, one expects a local displacement caused by a tangible physical mechanism associated with a characteristic velocity v_D and a finite time scale t_D, Eq. (4). In all experiments analyzed, the particles were passive, and mucus was not forced externally to generate and active transport. It is reasonable to assume that most particles in the experiments acquired their transient velocity from absorbing kinetic energy from the water molecules in mucus, leading to the characteristic velocity $v_D^2 \sim k_B T/m$, where k_B is the Boltzmann constant, T is the temperature, and m is the mass of the particle. The particle's velocity v_D would dissipate due to the mucus' viscosity with a characteristic time $t_D \sim t_r \sim m/\gamma$, where γ is the friction coefficient (Zwanzig 2001). In the most general formulation, this friction coefficient contains the viscous and elastic effects of the fluid (Spagnolie 2015). This leads to the characteristic local displacement $L_D \sim \sqrt{k_B T m}/\gamma$. It was then assumed room temperature, a typical particle's mid-size in the data set, $d \sim 100$ nm, and a viscosity of mucus close to water, which was a reasonable assumption because most physiological conditions have a low weight per volume (Barr *et al.* 2015; Joiner *et al.* 2019). This led to a characteristic local displacement of $L_D \sim 1$ nm and a characteristic local displacement dissipation time of $t_D \sim 1$ μs. Therefore, the estimated characteristic scales were consistent with those obtained from the empirical and theoretical analysis, $L_D \approx 3$ nm and $t_D \approx 5$ μ. For the case discussed above, it is important to notice that in the limit of regular diffusion, $\alpha = 1$, the ansatz in Eqs (4) and (5) leads to $D \sim L_D^2/t_D$, recovering, as expected, the diffusion expression $D \sim kT/m$ associated with the fluctuation-dissipation theorem. However, the theoretical framework defined by the fundamental ansatz is general and does not require particles to be propelled by the adsorption of kinetic energy.

Particles larger than 100 nm are more sensitive to anomalous diffusion.

Particle size, d, was not selected as a significant predictor in the random forest analysis (Figure 1**a**). However, the anomalous exponent analysis predicted that a certain group of particles would display regular diffusion (Figure 2**b**). This suggested that particle size could have an important indirect role in the effective diffusion. In fact, The analysis of D_{eff} as a

function of d displayed a clear threshold around $d^{\star} \sim 100$ nm (Figure 3**a**). Larger particles, $d > 100$ nm, displayed a lower effective diffusion, D_{eff}, although with no apparent statistical correlation with size (Spearman correlation $\rho = -0.24$, p-value $= 0.19$). Smaller particles, $d < 100$ nm, displayed an effective diffusion with a significant statistical correlation (Figure 3**a**). The slope for the log-log data was $m = -2.2 \pm 0.3$ ($R^2 - 0.67$), that is, the effective diffusion displayed apparently a power law of order two with particle size, $D_{eff} \sim 1/d^2$. Thus, diffusion overall decreased with particle size much rapidly than for regular diffusion, which is consistent with viscoelastic effects. However, a subset of particles (n=21) diffused normally ($\alpha = 1$) in Figure 2**b**, displaying an effective diffusion inversely proportional to particle size with a power function exponent $m = -1.0 \pm 0.1$ (p-value $= 1.4 \cdot 10^{-7}$, $R^2 = 0.77$) (Figure 3**a**). This empirical scaling, $D_{eff} \sim 1/d$, is expected for particles displaying regular diffusion, in agreement with the anomalous exponent prediction in Figure 2**b**. A similar analysis was performed comparing the rescaled effective diffusion by particle size, $D_{eff}\, d$, as a function of particle size, d. As expected, the particles predicted to display regular diffusion had slope zero, and the conclusions of the analysis were analogous (Figure S.1). Statistically, the analysis of D_{eff} was preferred over $D_{eff}\, d$ because the use of d as input and output in Figure S.1 can introduce biases and increase uncertainty.

Most parameters reported display weak correlations with the effective diffusion

The other four variables selected in the random forest analysis (Figure 1), that is, particle type, particle charge, mucin source, and mucin type, displayed weak correlations or no apparent correlations with D_{eff} as single predictors (Table S.1).

Particle type. Particle type was selected as the second most relevant variable to predict the effective diffusion based on the random forest analysis (Figure 1). Comparing the effective diffusion for the different particles confirmed this prediction (Figure S.2**a**). Antibodies and proteins displayed the fastest effective diffusion with a median of 48.9 $\mu m^2/s$. Viruses were the second fastest group with a mean effective diffusion an order of magnitude smaller, 3.5 $\mu m^2/s$. Pegylated and amine particles formed the third group. They displayed statistically similar effective diffusion with medians 0.99 $\mu m^2/s$ and $2 \cdot 10^{-2}$ $\mu m^2/s$. This was followed by carboxylated particles, median $3 \cdot 10^{-2}$ $\mu m^2/s$, and finally chitosan $4 \cdot 10^{-3}$ $\mu m^2/s$. Differences in particle size could explain the reduction in effective diffusion for an-

tibodies/proteins, viruses, and PEG (pegylated) particles (Figure S.2**b**). They had median sizes of $\sim$ 10 nm, $\sim$ 100 nm, and $\sim$ 1000 nm, respectively. It is unclear what were the physicochemical factors behind the slower diffusion of amine, COOH, and chitosan particles (Figure S.2).

Particle charge. The third predictor for effective diffusion was particle charge, expressed as the zeta-potential ζ (Figure 1). Particles with negative zeta potential displayed a positive correlation with the effective diffusion with a Spearman correlation of $\rho = 0.6$ (p = 0.002, n = 36) (Figure 4**a**). The relationship was approximated by an exponential function, $D_{eff} \sim 10^{m\zeta}$. The potential rate was $m = (0.024 \pm 0.006)$ mV^{-1} (p = 0.0002) obtained from a least-square linear regression using the log-linear data. This exponential model accounted for 30% of the variance ($R^2 = 0.30$). The largest effective diffusions were achieved at neutral zeta potentials. Positive zeta potentials (n=21) had lower values but did not display a statistically significant correlation for the effective diffusion. Particle size or other properties did not seem to explain the trend observed for negatively charged zeta potentials (Figure S.3). These particles, however, displayed a linear positive correlation with the anomalous diffusion (Figure 4**b**).

Mucus source. The mucus source and dominant mucin were the last two significant predictors of effective diffusion (Figure 1). The effective diffusion was faster in human cervix samples with a median $\sim$ 10 $\mu\text{m}^2/\text{s}$, although the values spanned six orders of magnitude, from $\sim 10^{-4}$ to $\sim 10^{2}$ $\mu\text{m}^2/\text{s}$ (Figure S.4). The effective diffusion was the slowest in mucus from human lung (median $\sim 10^{-2}$ $\mu\text{m}^2/\text{s}$) and pig intestine (median $\sim 10^{-2}$ $\mu\text{m}^2/\text{s}$). The median particle size in empirical data from human cervix mucus was more than an order of magnitude smaller, $\sim$ 10 nm, than for the empirical data from the other sources. The median pH for the empirical data from human cervix mucus was significantly lower pHs (median 5.5) compared to the other sources (median 7). Lower pH tends to thicken mucus (Hwang *et al.* 1969), thus expecting a slower effective diffusion. But the particle size may have offset this trend. The transcription analysis identified MUC5B, which is dominant in human cervix, displaying the largest effective diffusion (median $\sim$ 10 $\mu\text{m}^2/\text{s}$) compared to the other dominant mucins, MUC2 common in intestinal mucus (median diffusion $\sim 10^{-1}$ $\mu\text{m}^2/\text{s}$), and MUC5AC common in respiratory mucus (median diffusion $\sim 10^{-2}$ $\mu\text{m}^2/\text{s}$) (Figure S.4).

DISCUSSION

The meta-analysis of particle diffusion in mucus revealed that diffusion of microscopic particles spanned seven orders of magnitude in passive conditions (no self-propulsion or active mucus transport) (Table I). The anomalous diffusion exponent, α, was the factor displaying the strongest correlation with the effective diffusion, D_{eff} (Figure 1). Statistically, the effective diffusion displayed an exponential dependence with respect to the anomalous diffusion, explaining 89% of the variance in the data (Figure 2a). This result was based on 39 out of 106 experiments (about a third), which had measured the anomalous exponent. Among the remaining 67 experiments, our statistical model predicts that the anomalous exponent was dominant in determining the effective diffusion in 46 of those experiments, that is, 69% of them (Figure 2b). In the other 21 experiments, the model predicts that particles followed regular diffusion, that is, the anomalous exponent would have no power predicting the change in the effective diffusion (Figure 2b). Therefore, the anomalous exponent was a strong predictor of the effective diffusion in 80% of all experiments analyzed. The anomalous exponent is an emerging property, and this result offers the opportunity to compare the diffusion of particles subjected to different molecular mechanisms. It is puzzling, however, that only a third of the experiments measured the anomalous exponent. One possible explanation is the fact that the anomalous exponent is a well-known emerging property, but the relation between this exponent and underlying molecular factors determining its value are not that established in the field yet (McGlynn *et al.* 2020). Below we argue that investigating the molecular basis of the anomalous exponent is the key to characterizing and controlling particle diffusion in mucus and other polymeric fluids at relevant biological time scales.

The theoretical scaling analysis of subdiffusion identified that the anomalous diffusion exponent, α, displays a stronger correlation over other physical factors when the diffusion is characterized at sampling times two orders of magnitude larger than the microscopic timescale fueling the diffusive motion, Eqs. (13) and (12). In our analysis, the experiments focused on passive diffusion conditions, but the principles behind the theoretical scaling can be expanded to situations with motile particles as well as energetically active mucus transport. The theoretical derivation just assumes that there is a characteristic time scale and length scale governing the particle's local motion. For the experiments analyzed here, the

kinetic energy and viscosity of the fluid were assumed to be associated with the particle's local motion and were used to investigate the consistency with the theory. But in other contexts, the same analysis can be applied to replace the dependency of the characteristic scales with other mechanisms. For example, if particles run and tumble, like the bacterium *E. coli*, the transient velocity of the particle depends on the viscosity and concentration of the polymeric network and food sources instead of the kinetic energy (Martinez *et al.* 2014; Patteson *et al.* 2015). The scaling analysis is also consistent with the generalized diffusion equation in complex fluids, which extends the Stokes-Einstein relation to viscoelastic fluids (Spagnolie 2015). In this case, the characteristic length and time scales would incorporate the elastic effects of the network. The role of the general characteristic scales in the revised diffusion equation, Eqs. (6), aimed to accommodate a diverse set of scenarios. Additionally, it solved the issue of relying on the generalized diffusion constant, which has undefined physical units and is not strictly a well-defined physical magnitude, Eq. (5). The theoretical and empirical analyses presented here highlighted the dominance of the anomalous diffusion exponent in determining the range of effective diffusions.

Thus, the problem now translates into identifying the factors that determine the anomalous diffusion exponent. These physical factors depend on the underlying mechanism responsible for the subdiffusion (Metzler *et al.* 2014). This is particularly relevant to understand the emergence of the critical particle size $d^\star \sim 100$ nm in Figure 3. This critical value may represent the particle size's onset when the effects of the mucus mesh become relevant. Most experiments analyzed did not report the mucus mesh size, but the critical size observed was consistent with mesh sizes measured in mucus samples, which range from 100 to 400 nm (Spagnolie 2015; Suk *et al.* 2009; Lai *et al.* 2010, 2011; Ensign *et al.* 2012). For particles with sizes similar or larger than this mesh size, the theoretical description of the molecular displacement, L_D, should include the effects of the mesh size. If the mesh hinders mobility, this will lead to a reduction of the average molecular displacement. If the mesh streamlines the mobility, then the molecular displacement could increase; for example, changes in the chemical coating of relatively large particles (200-500 nm) compared to the critical size, $d^\star \sim 100$ nm, can display larger diffusivities in mucus than in water (Lai *et al.* 2007). The key point is that for particles larger than 100 nm, the impact of the mesh size in particle diffusivity will be very diverse. In each case, it is necessary to assess the underlying molecular mechanism responsible for the anomalous diffusion to identify the key physical

factors governing the diffusivity. We clarified this below for two mechanisms that may play an important role in mucus. First, microscopic particles can bind to mucin fibers leading to subdiffusion (Barr *et al.* 2015). Second, mucin fibers form a polymeric mesh that can trap particles as observed in other hydrogels (Wong *et al.* 2004). These two scenarios are particularly relevant in passive conditions. Scenarios involving the activation of the mucus network via cilia or peristalsis are also of interest but fall beyond the scope of this work.

Binding to mucins does not necessarily lead to subdiffusion. If a particle has a single binding site, the characteristic binding time t_b would dilate the characteristic time to estimate the diffusivity of the particle, $t_D \sim t_r + t_b$, where t_r is the relaxation time. The microscopic diffusion would be $D \sim v_D^2 t_r^2 / t_D \sim f_r k_B T/m$. The diffusion would be reduced by the factor $f_r < 1$, which is the fraction of time spent dissipating the particle's speed, $f_r = t_r/(t_r + t_b)$. This would not impact α unless more than one region of the particle can bind stochastically to mucins, increasing the binding time beyond the sampling time, $t_b \gg \Delta t_{eff}$. This would lead to an effective power-law distribution of binding times with no apparent characteristic binding time (Xu *et al.* 2011). The emergence of long-tailed attachment time distributions leads to subdiffusion. The anomalous exponent, α, would be equal to the exponent, ν, of the asymptotic approximated power-law distribution of attachment times (Barkai *et al.* 2012; Metzler *et al.* 2014; Joiner *et al.* 2019). In this case, the continuous-time random walk approximation leads to the generalized subdiffusion expression $D_\alpha = D\,\tau_D/\tau_D^\alpha$ (Joiner *et al.* 2019). Here, D is the diffusion of the particle in the absence of interactions with mucins, τ_D is the average diffusion time of a particle before attaching again to a mucin fiber. This result is consistent with the ansatz introduced in Eq. (5). In this particle-mucin affinity mechanism, the distribution of binding times would control α becoming the most relevant factor impacting the effective diffusion, D_{eff}. Unfortunately, the experiments analyzed did not explore the particle affinities to mucus explicitly.

The microenvironment trapping mechanism was observed in F-actin networks, where microscopic tracers were shown to follow anomalous diffusion (Wong *et al.* 2004). The anomalous exponent was a linear function of the ratio between the particle size (d) and the network's mesh size (ξ). The empirical dependency obtained was $\alpha \approx 1$ for $d/\xi < 0.1$, $\alpha \approx -1.25\,d/\xi + 1.38$ for $0.1 < d/\xi < 1.1$, and $\alpha \approx 0.1$ for $d/\xi > 1.1$. Thus, particles with sizes a 10% of the mesh size or smaller diffused normally, while particles with a size similar or larger to the mesh displayed a reduced diffusivity with a low anomalous exponent.

The specific parameters of the relationship were not derived, but one would expect similar behavior in mucus. The average mucus in humans has a typical mesh size between 100 to 1000 nm (Cone 2009). In this mechanism, d/ξ controls α becoming the most important factor determining D_{eff}. This could explain the threshold observed on the effective diffusion as a function of particle size (Figure 3**a**). Larger particles, $d > 100$ nm, displayed lower effective diffusions, D_{eff}, although with no apparent statistical correlation. The variation of α, from 0.15 to 1, could be due to a change in the mesh size (ξ). Unfortunately, the mesh size (or a proxy, like the concentration of mucins) was not measured or reported in most experiments analyzed here.

The two mechanisms discussed above could also help interpret the statistically significant correlations obtained between the effective diffusion and the surface charge of particles (Figure 4). Given the negative charge of mucin fibers, a particle with a larger negative charge would display a larger effective radius within the mucin network. This would increase the particle size to network mesh ratio, thus, reducing the anomalous exponent, and, consequently, the particle's diffusivity. This scenario would explain the statistical trends observed for the effective diffusion and anomalous exponent for negative zeta potentials (Fig. 4). However, one cannot discard other scenarios. For example, negatively charged carboxylated particles competing for cations at high densities can expose hydrophobic regions in mucus, leading mucus fibers to form bundles (Lai *et al.* 2009, 2007). This might be a less likely but still plausible scenario. For positively charged particles, the particle-mucin binding mechanism could be responsible for the relatively low anomalous exponents observed. The framework explored here suggests that measuring the particle-mucin binding times and mucin mesh size would help disentangle the variance in the data. This framework should also apply to other polymeric fluids. It has been observed, for example, that particles of 1 μm (1000 nm) display subdiffusion and trapping in biofilms in the time scale of days, and 0.5 μm (500 nm) particles display lower mean-squared displacement in regions of higher effective cross-linking (Chew *et al.* 2014). These observations were done on the time scale of days. These results resonate with our finding that particles larger than 100 nm in mucus are very sensitive to subdiffusion behavior and trapping. It would be necessary to characterize systematically particle-binding to polymer fibers, polymer mesh sizes, and time scales to compare, extrapolate, and unify results across different polymeric fluid systems.

Some of the weak and highly-dispersed correlations analyzed above might be obscured

due to the combined effect of multiple variables, for example, particle charge and pH (Leal *et al.* 2017). Unfortunately, at a given particle chemistry and zeta potential (charge), most particle diffusion measurements were reported at a fixed pH (Source Data 2). Thus, the data analyzed posed an intrinsic limitation to disentangle the convoluted effects of charge and pH on particle diffusion. The impact of mucin-type, mucin-mucin interactions and mucin concentration can also depend on pH. Low pH alters the molecular structure of mucins from a random coil to extended conformations, facilitating cross-linking and transitioning to a mucus solid-gel phase (Cao *et al.* 1999). However, these key physicochemical mucus' properties were difficult to extract from most experiments analyzed.

In any case, our results indicate a common approach to investigate these co-dependent properties on particle diffusion: measuring the anomalous exponent and establishing the underlying mechanism responsible for it. The physical factors controlling the anomalous exponent will be the dominant factors in particle diffusion. The theory introduced here is based on a generic ansatz that represents a first approximation. More refined theoretical approaches will be necessary to identify correction factors associated with specific underlying molecular mechanisms. One possible direction would be adapting the continuum models that characterize polymer fluids as viscoelastic Maxwell fluids (Spagnolie 2015; Grimm *et al.* 2011; Grebenkov *et al.* 2013; Levine and Lubensky 2000). In the presence of viscous delays incorporated with the Basset force term, these models predict an emerging subdiffusive transient region at time scales of miliseconds (Grebenkov *et al.* 2013). That time scale is much shorter than the one investigated here (above seconds), and that subdiffusion does not emerge from the molecular mechanisms (particle-mucus interaction mechanism and the caging effect) identified here as relevant in mucus. Incorporating the molecular characteristics of these two mechanisms in Maxwell fluids would offer a more sophisticated framework to predict particle diffusion in mucus.

In conclusion, our meta-analysis revealed that the anomalous exponent displays the strongest correlation with the effective diffusion of particles in mucus compared to other commonly measured factors.. It explained almost 90% of the variance of diffusions across seven orders of magnitude. Our theoretical scaling analysis justified this observation assuming the characteristic displacement length and time of the local physical motion. This led to a reformulated subdiffusion equation in terms of these characteristic scales of the underlying mobility mechanism, and it demonstrated that the widely accepted generalized

diffusion constant is not a measurable physical quantity. The theoretical analysis predicted that the anomalous exponent determines the order of magnitude of the effective diffusion for sampling time windows two orders of magnitude larger than the microscopic mobility time scale. This prediction applied to any physical system and was consistent with the data from particles diffusing in mucus. Our theoretical analysis indicates that the factors regulating the anomalous exponent are essential to characterize the diffusion of particles. At least two of these factors can control the anomalous exponent in mucus: the distribution of particle-mucin binding times and the particle size-to-mucin mesh ratio. These factors regulate the anomalous exponent and, subsequently, the effective diffusion of microscopic particles. However, these key properties were not reported in most experiments analyzed. Therefore, our study provides a guide on how to characterize, study, and modify the diffusion of particles in mucus and other hydrogels.

ACKNOWLEDGMENTS

This manuscript has been released as a pre-print at bioRxiv (Cobarrubia *et al.* 2020). The original contributions presented in the study are included in the article and supplementary material. Further inquiries can be directed to the corresponding author. We thank the Biomath Working Group at San Diego State University for their feedback during the development of the project, and we would like to make a special mention of the insightful comments from Professor Parag Katira and Arlette Baljon. A.L. acknowledges the support received by the New Investigator Award received from the California State University (CSU) Program For Education and Research in Biotechnology (CSUPERB), the CSU Faculty Innovation and Leadership Award, and the National Science Foundation Award 1951678 from the Mathematical Biology Program.

*tony.cobarrubia38@gmail.com

†jarod.tall@gmail.com

‡augustuscrispin@gmail.com

§aluque@sdsu.edu

S. Spagnolie, *Complex fluids in biological systems experiment, theory and computation* (Springer, New York, 2015).

B. Krajina, C. Tropini, A. Zhu, P. DiGiacomo, J. L. Sonnenburg, S. C. Heilshorn, and A. J. Spakowitz, ACS Cent. Sci. **3**, 1294 (2017).

F. Bäckhed, R. E. Ley, J. L. Sonnenburg, D. A. Peterson, and J. I. Gordon, Science **307**, 1915 (2005).

C. R. Bakshani, A. L. Morales-Garcia, M. Althanus, M. D. Wilcox, J. P. Pearson, J. C. Bythell, and J. G. Burgess, NPJ Biofilms Microbiomes **4** (2018).

C. B. Silveira and F. L. Rohwer, NPJ Biofilms Microbiomes **2** (2016).

S. K. Lai, K. Hida, S. Shukair, Y. Wang, A. Figueiredo, R. Cone, T. J. Hope, and J. Hanes, J. Virol. **83**, 11196 (2009).

R. Cone, Adv. Drug Deliv. Rev. **61**, 75 (2009).

J. J. Barr, R. Auro, M. Furlan, K. L. Whiteson, M. L. Erb, J. Pogliano, A. Stotland, R. Wolkowicz, A. S. Cutting, K. S. Doran, P. Salamon, M. Youle, and F. Rohwer, Proc. Natl. Acad. Sci. U.S.A. **110**, 10771 (2013).

J. J. Barr, R. Auro, N. Sam-Soon, S. Kassegne, G. Peters, N. Bonilla, M. Hatay, S. Mourtada, B. Bailey, M. Youle, B. Felts, A. Baljon, J. Nulton, P. Salamon, and F. Rohwer, Proc. Natl. Acad. Sci. U.S.A. **112**, 13675 (2015).

B. Amsden and N. Turner, Biotechnol. Bioeng. **65**, 605 (1999).

M. Abdulkarim, N. Agullo, B. Cattoz, P. Griffiths, A. Bernkop-Schnurch, S. G. Borros, and M. Gumbleton, Eur. J. Pharm. Biopharm. **97**, 230 (2015).

F. Arends, B. R., and L. O., Langmuir **29**, 15965 (2013).

J. Hansing, C. Ciemer, W. K. Kim, X. Zhang, J. E. DeRouchey, and R. R. Netz, Eur. Phys. J. E **39** (2016).

O. Lieleg, I. Vladescu, and K. Ribbeck, Biophys. J. **98**, 1782 (2010).

L. D. Li, T. Crouzier, A. Sarkar, L. Dunphy, J. Han, and K. Ribbeck, Biophys. J. **105**, 1357 (2013).

J. P. Celli, B. S. Turner, N. H. Afdhal, S. Keates, I. Ghiran, C. P. Kelly, R. H. Ewoldt, G. H. McKinley, P. So, S. Erramilli, and R. Bansila, Proc. Natl. Acad. Sci. U.S.A. **106**, 14321 (2009).

J. S. Suk, S. K. Lai, N. J. Boylan, M. R. Dawson, M. P. Boyle, and J. Hanes, Nanomedicine(Lond) **6**, 365 (2011).

S. K. Lai, D. E. O'Hanlon, S. Harrold, S. T. Man, Y. Wang, R. Cone, and J. Hanes, Proc. Natl. Acad. Sci. U.S.A. **104**, 1482 (2007).

T. Lang, E. Larsson, M. E. V. Johnansson, G. C. Hansson, and T. Samuelsson, Mol. Biol. Evol. **33**, 1921 (2016).

S. S. Olmsted, J. L. Padgett, A. I. Yudin, K. J. Whaley, T. R. Moench, and R. Cone, Biophys. J. **81**, 1930 (2001).

J. Newby, J. L. Schiller, T. Wessler, J. Edelstein, M. G. Forest, and S. K. Lai, Nat. Commun. **8** (2017).

B. S. Schuster, J. S. Suk, G. F. Woodworth, and J. Hanes, Biomaterials **34**, 3439 (2013).

H. M. Yildiz, C. A. McKelvey, P. J. Marsac, and R. L. Carrier, J. Drug Target **23** (2015).

A. Rohatgi, "Webplotdigitizer 4.2," (2019).

F. Huang, E. Watson, C. Dempsey, and J. Suh, Methods Mol. Biol. **991**, 211 (2013).

J. McGlynn, N. Wu, and K. Schultz, J. of App. Phys. **127**, 201101 (2020).

E. Barkai, Y. Garini, and R. Metzler, Phys. Today **65**, 29 (2012).

R. Metzler, J. Jeon, A. G. Cherstvy, and E. Barkai, Phys. Chem. **16**, 24128 (2014).

R. Hou, A. G. Cherstvy, R. Metzler, and T. Akimoto, Phys. Chem. Chem. Phys. **20**, 20827 (2018).

M. Grimm, S. Jeney, and T. Franosch, Soft Matter **7**, 2076 (2011).

D. Grebenkov, M. Vahabi, E. Bertseva, L. Forró, and S. Jeney, Physical Review E **88**, 040701 (2013).

S. C. Chew, B. Kundukad, T. Seviour, J. Van der Maarel, L. Yang, S. Rice, P. Doyle, and S. Kjelleberg, MBio **5**, e01536 (2014).

C. Cruickshank Miller, Proc. Royal Soc. B **106**, 724 (1924).

D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Giron, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, T. Lavidas, Z. Liu, J. E. Loveland, T. Mau-

rel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuh, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, and P. Flicek, Nucleic Acids Res. **46**, D754–D761 (2018).

E. Archer, *rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics* (2019), r package version 2.1.7.

G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, Vol. 112 (Springer, 2013).

A. Kumar and C. K. Dixit, in *Advances in nanomedicine for the delivery of therapeutic nucleic acids* (Woodhead Publishing, 2017) pp. 43–58.

K. L. Joiner, A. Baljon, J. Barr, F. Rohwer, and A. Luque, Sci. Rep. **9** (2019).

J. Taylor, *Introduction to error analysis, the study of uncertainties in physical measurements* (University Science Books, New York, 1997).

R. Zwanzig, *Nonequilibrium statistical mechanics* (Oxford University Press, 2001).

S. Hwang, M. Litt, and W. C. Forsman, Rheol. Acta **8**, 438 (1969).

V. Martinez, J. Schwarz-Linek, M. Reufer, L. Wilson, A. Morozov, and W. Poon, Proc. Natl. Acad. Sci. U.S.A. **111**, 17771 (2014).

A. Patteson, A. Gopinath, M. Goulian, and P. Arratia, Sci. Rep. **5**, 1 (2015).

J. Suk, S. Lai, Y.-Y. Wang, L. Ensign, P. Zeitlin, M. Boyle, and J. Hanes, Biomaterials **30**, 2591 (2009).

S. Lai, Y.-Y. Wang, K. Hida, R. Cone, and J. Hanes, Proc. Natl. Acad. Sci. U.S.A. **107**, 598 (2010).

S. Lai, J. Suk, A. Pace, Y.-Y. Wang, M. Yang, O. Mert, J. Chen, J. Kim, and J. Hanes, Biomaterials **32**, 6285 (2011).

L. Ensign, B. Tang, Y.-Y. Wang, A. Terence, T. Hoen, R. Cone, and J. Hanes, Sci. Transl. Med. **4**, 138ra79 (2012).

I. Y. Wong, M. L. Gardel, D. R. Reichman, E. R. Weeks, M. T. Valentine, A. R. Bausch, and D. A. Weitz, Phys. Rev. Lett. **92**, 178101 (2004).

Q. Xu, L. Feng, R. Sha, N. C. Seeman, and P. M. Chaikin, Phys. Rev. Lett. **106**, 228102 (2011).

J. Leal, H. Smyth, and D. Ghosh, Int. J. Pharm. **532**, 555 (2017).

X. Cao, R. Bansil, K. Bhaskar, B. Turner, J. LaMont, N. Niu, and N. Afdhal, Bioph. J. **76**, 1250 (1999).

A. J. Levine and T. Lubensky, Physical review letters **85**, 1774 (2000).

A. Cobarrubia, J. Tall, A. Crispin-Smith, and A. Luque, bioRxiv (2020), 10.1101/2020.07.25.221416.

FIGURES AND TABLES

Property	Symbol	Range	Data points
Effective diffusion	D_{eff}	$3.1 \cdot 10^{-5}$ to $1.3 \cdot 10^{2}$ $\mu m^2/s$	106
Anomalous exponent	α	0.16 to 1.02	39
Diameter	d	3.5 to 1280.0 nm	106
Zeta potential	ζ	-73.0 to $+33.3$ mV	57
Temperature	T	295 to 310 K	106
pH	pH	3.0 to 7.4	63
Mucus source		Hydrogel, human lung, human cervix, pig stomach, pig intestines	106
Mucin type		MUC2, MUC5AC, MUC5B	103

TABLE I. Summary of empirical data. The effective diffusion, D_{eff}, was obtained for a common time window of 1 second.

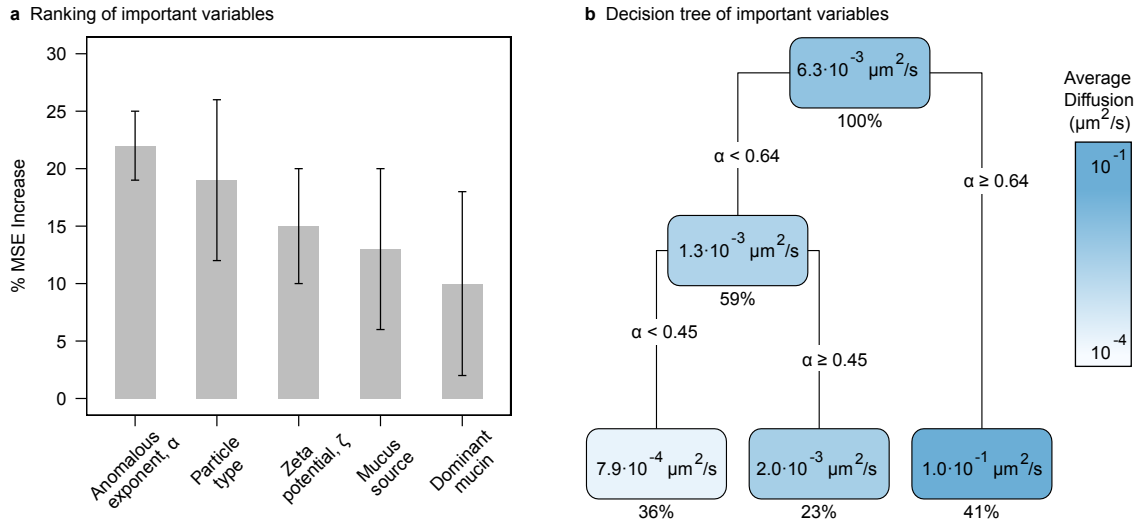

FIG. 1. Selected variables impacting effective diffusion. **a**, Average percentage increase of mean-square error (% MSE) for the selected variables. The error bars correspond to the standard deviation. **b**, Decision tree for the most important variables. Each node contains the predicted average D_{eff} and percentage of data predicted. The gradient display diffusion values from $\sim 10^{-4}$ μm^2/s (white) to $\sim 10^{-1}$ μm^2/s (blue).

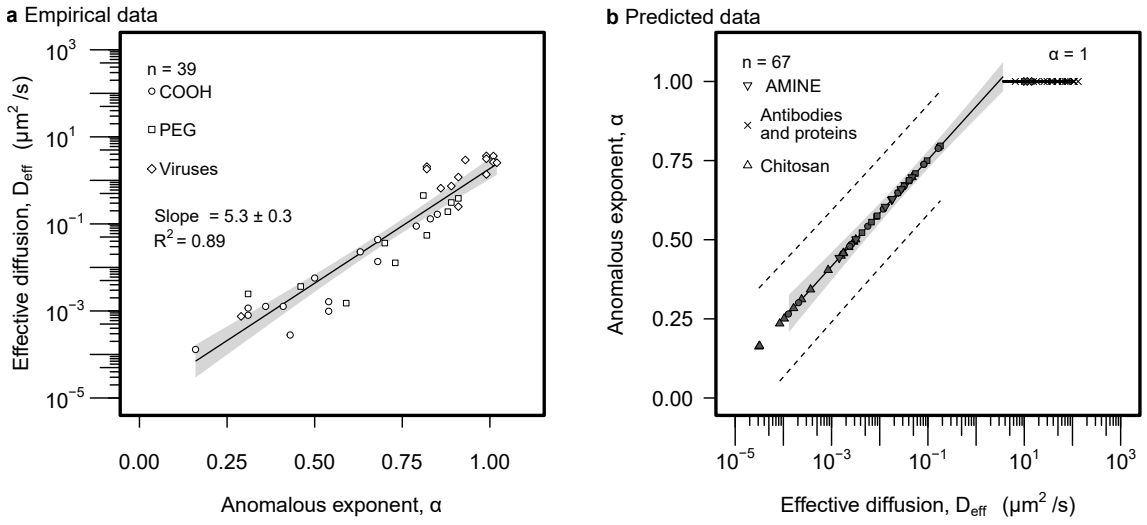

FIG. 2. Effective diffusion and anomalous exponent analysis. **a**, effective diffusion was plotted as a function of the anomalous exponent. The solid line represents the regression model. The grey area represents the 95% confidence interval. Statistically significant slope and R^2 of linear regression are displayed. **b**, the anomalous exponent was predicted based on the model found empirically in **a**. The solid line designates the predicted linear model. The grey area represents the 95% confidence interval of the predicted linear model. The dashed line represents a 95% prediction interval **a-b**, distinguished particle types are represented in the legend of both panels.

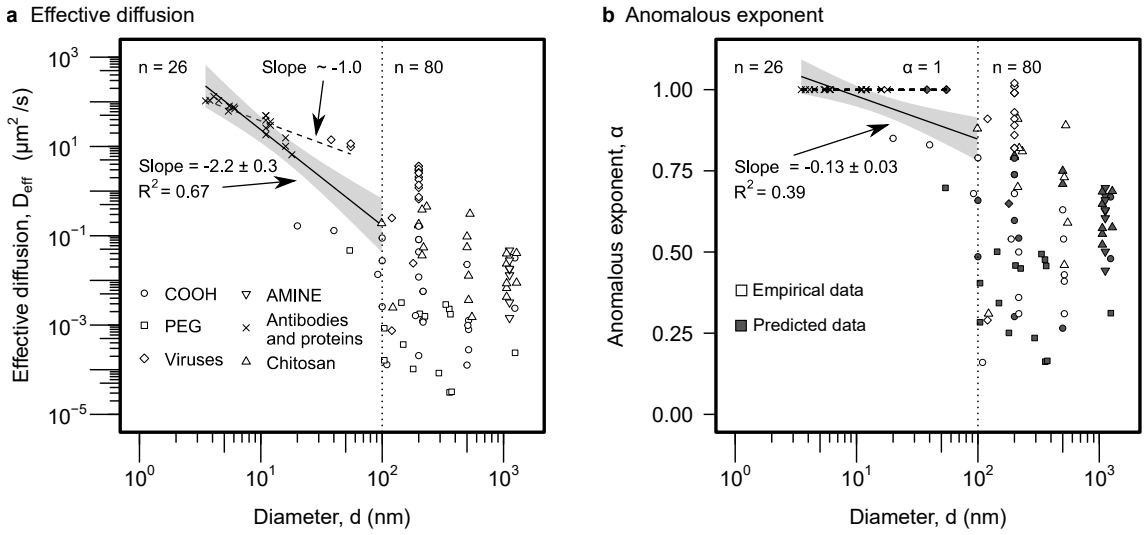

FIG. 3. Particle size analysis. **a**, The effective diffusion was plotted against particle size, d. The different symbols correspond to different particles types as indicated in the legend. The solid line indicates the linear regression for $d < 100$ nm particles using log-log data (n = 26), and it displays the slope and coefficient of determination, R^2. The grey area represents the 95% confidence interval. The group of particles with $d > 100$ nm (n=80) did not display a statistically significant relationship and no solid line is included. The dashed line corresponds to the linear regression of the subset of particles (n = 21) displaying regular diffusion, $\alpha = 1$, in Fig. 2**b**, using the log-log data. The slope was appoximately one as expected ($slope = -1.0 \pm 0.1$, p-value $= 1.4 \cdot 10^{-7}$, $R^2 = 0.77$). **b**, The anomalous exponent was plotted as a function of particle size. The symbols and lines are analogous to panel **a**. As in panel **a**, the solid line is the regression for the particles with $d < 100$ (n = 26), while the dashed line represents the subset (n = 21) predicted to display regular diffusion, $\alpha = 1$. Empty symbols anomalous exponents obtained empirically. The solid symbols correspond to the predicted anomalous exponents for the subset of data that did not include empirical values. The predictions were obtained using the model derived from Figure 2**b**.

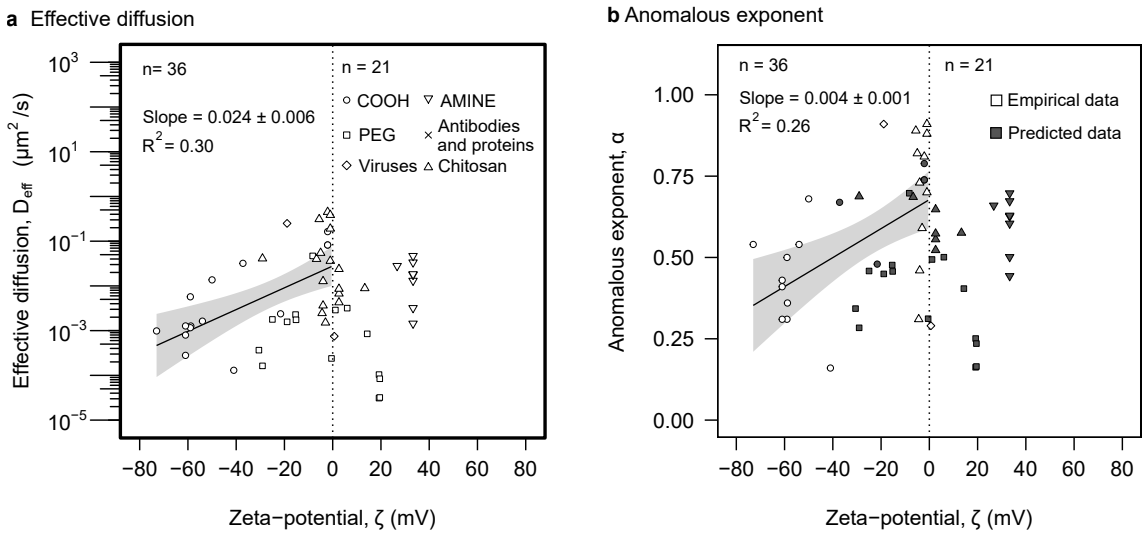

FIG. 4. Electrostatic analysis. **a** The effective diffusion was plotted against zeta potential. **b** The anomalous exponent was plotted as a function of zeta potential. **a-b** The distinction between empirical and predicted data as well as particle types are represented in the legend. The dotted line indicates $\zeta = 0$. The solid lines correspond to statistically significant linear regressions. The grey areas represent 95% confidence intervals of the linear regression. The slopes and R^2 of each linear regression are also displayed in the panels.