THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 5599−5609

pubs.acs.org/JPCB

Computational Signaling Protein Dynamics and Geometric Mass Relations in Biomolecular Diffusion

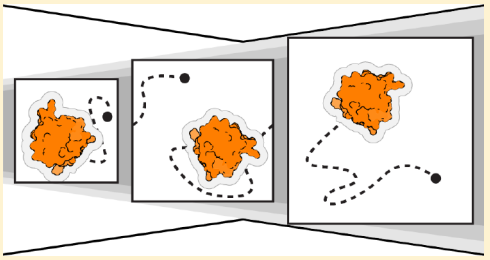
Christopher J. Fennell,*[,†] Neda Ghousifam,[‡] Jennifer M. Haseleu,[†,¶] and Heather Gappa-Fahlenkamp[‡]

[†]Department of Chemistry and [‡]School of Chemical Engineering, Oklahoma State University, Stillwater, Oklahoma 74078, United States

[¶]Department of Chemistry, Saint Vincent College, Latrobe, Pennsylvania 15650, United States

Supporting Information

ABSTRACT: We present an atomistic level computational investigation of the dynamics of a signaling protein, monocyte chemoattractant protein-1 (MCP-1), that explores how simulation geometry and solution ionic strength affect the calculated diffusion coefficient. Using a simple extension of noncubic finite size diffusion correction expressions, it is possible to calculate experimentally comparable diffusion coefficients that are fully consistent with those determined from cubic box simulations. Additionally, increasing the concentration of salt in the solvent environment leads to changes in protein dynamics that are not explainable through changes in solvent viscosity alone. This work in accurate computational determination of protein diffusion coefficients led us to investigate molecular-weight-based predictors for biomolecular diffusion. By introducing protein volume- and protein surface-area-based extensions of traditional statistical relations connecting particle molecular weight to diffusion, we find that protein solvent-excluded surface area rather than volume works as a better geometric property for estimating biomolecule Stokes radii. This work highlights the considerations necessary for accurate computational determination of biomolecule diffusivity and presents insight into molecular weight relations for diffusion that could lead to new routes for estimating protein diffusion beyond the traditional approaches.



INTRODUCTION

How proteins and other biomolecules move establishes a rate limit for function in living systems. No single molecule can perform all functions in and around cells, so diffusion of partner molecular species is necessary for given particles to play a role in a complex system. The primary role of signaling proteins is to do exactly this, translate through their environment and interact with other larger biomolecular structures in order to enact a more complex action.

Most studies on protein translation and function have been performed in experimental laboratories rather than through computational methods. Before the first structures of proteins were determined, diffusion analysis via ultracentrifugation sedimentation studies was one of the primary methods for assessing the shape and establishing the identity of these macromolecules.[1−4] Early such analyses in protein dynamics motivated the awarding of the 1926 Nobel Prize in Chemistry to Theodor Svedberg for his pioneering efforts in the practical development and implementation of experimental techniques in protein dynamics and separation. Protein diffusion studies tend to be challenging from a computational perspective, because the translational dynamics of multi-kDa mass particles that are greater than an order-of-magnitude larger than the particles in their surrounding solvent environment necessitate the use of reasonably large system sizes simulated over relatively long times, often multiple microseconds in length.

Complicating matters, molecular simulations are performed on finite-sized systems. While use of periodic boundary conditions and extended interaction correction techniques are mostly standardized for modeling condensed phase molecular systems, some calculated properties need to be corrected to account for the size of the simulation cell.[5] For example, while the shear viscosity is relatively insensitive to simulated system size,[6,7] the apparent three-dimensional diffusion coefficient will be slower in smaller systems and faster in larger ones, and corrections have been developed to account for these variations with system size.[6−13] We are interested in calculating diffusion coefficients that are quantitatively comparable with experiments, and some of the more complex models need to be simulated with noncubic simulation boxes. It would be beneficial to evaluate the accuracy of current strategies for converting apparent simulation calculated diffusion coefficients into viscosity corrected infinitely dilute diffusion coefficients. Converting a single simulation value would eliminate the need to perform multiple simulations in order to project out to the infinitely dilute value.

Special Issue: Ken A. Dill Festschrift

Received: November 30, 2017
Revised: February 28, 2018
Published: March 6, 2018

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

Here, we explore protein diffusion from atomistic level molecular dynamics simulations of a small signaling protein, monocyte chemoattractant protein-1 (MCP-1). The goals of this effort are severalfold and include (1) determining the steps and simulation scope necessary for calculation of experimentally comparable diffusion coefficients, (2) assessing how finite size corrections can be applied to single cubic and tetragonal simulations to determine similarly accurate diffusion coefficients, and (3) exploring how changes in the ionic strength of the surrounding environment affect translational diffusion of a protein. Performing this work led us to investigate protein-mass-based estimation of diffusion coefficients to evaluate how one might potentially make accurate estimations of protein diffusion without prior experimental knowledge of biomolecule dynamics.

METHODS

As a model system and for the purpose of comparison with experimental estimates, molecular dynamics simulations of pure water, water solvated monocyte chemoattractant protein-1 (MCP-1: PDB code 1DOL), and MCP-1 in increasingly concentrated aqueous NaCl solutions were performed at a target temperature and pressure of 310.15 K and 1 atm. In addition to these molecular dynamics simulations and companion analyses, we performed a surface area and volumetric investigation of a set of proteins of increasing mass in order to construct alternate mass-to-geometry relations for general estimation of biomolecule diffusivity.

In all molecular dynamics simulation systems, cubic simulation box geometries were used to form accurate finite size correction estimations for the diffusion coefficient calculations.[6−8] In the pure water and water solvated MCP-1 molecular simulation systems, additional tetragonal simulation box geometries were chosen to explore simple extensions of these finite size corrections beyond the typical cubic simulation box constraint. We have an interest in simulating the dynamics of proteins in the presence of collagen or other periodically replicated molecules, and these tetragonal simulations were used to test how accurate such extensions are in correcting diffusion calculations for large singular biomolecules.[8−13]

Figure 1 details the system geometry series used in calculating experimentally comparable diffusion coefficients. The pure water systems ranged from 3 to 6 nm short edge-length simulation box sizes, approximately 2700 to 21 500 atoms, respectively, while the water solvated protein systems ranged from 5 to 9 nm, corresponding to 12 400 to 72 500 atoms, respectively. The smaller tetragonal systems were up to twice the size of the corresponding cubic systems given the longer L_x edge-length. The pure water systems were both a verification test of finite size correction parameters detailed in previous studies[8,9] and necessary for determination of solvent viscosity η at the target temperature and pressure, needed for both finite size corrections and solvent model viscosity corrections in making experimental comparisons.[7,8] Specifically, for cubic systems with potentially charged molecules, Yeh and Hummer proposed the adapted cubic system finite size correction

$$D_{app}(L) = D_0 - \frac{k_B T \xi_{EW}}{6\pi\eta L}\alpha \quad (1)$$

where k_B is the Boltzmann constant, T is the simulation temperature, L is the simulation box edge-length, η is the solvent viscosity, $\xi_{EW} \approx 2.837298$ is a unitless simple cubic lattice self-term,[14] and α is an empirical free fitting parameter introduced to account for deviations from previous correction expressions, potentially introduced from charged solute interactions.[8] This expression gives a route to calculating the infinitely dilute diffusion coefficient, D_0, from the apparent diffusion coefficient, D_{app}, as calculated from a molecule's mean-square displacement over time. The resulting D_0 is not necessarily yet an experimentally comparable value, particularly if the solvent environment in the simulation does not reproduce the experimental environment and properties well. Multiplying the simulation D_0 by the ratio of simulation and experimental viscosities can correct for these differences.

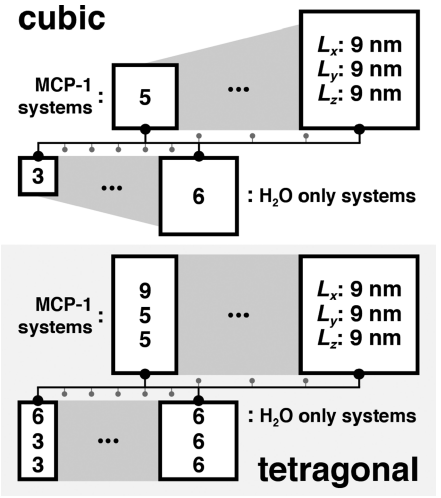


Figure 1. Illustrated series of simulation system geometries used in the calculation of experimentally comparable diffusion coefficients. The water-only simulations spanned cubic (top) and tetragonal (bottom) boxes with a short edge-length (L_y or L_z) ranging from approximately 3 to 6 nm in 0.5 nm increments, while the protein simulations needed to be larger with a short edge ranging from 5 to 9 nm in 1 nm increments.

$$D_\eta = \frac{\eta_{TIP3P}}{\eta_{expt}} \cdot D_0 \quad (2)$$

While insensitive to simulation system sizes, η_{TIP3P} will be strongly temperature and salt concentration dependent, and it needs to be estimated for each given simulation composition and state point.

Calculating a D_0 using eq 1 is potentially convenient assuming the necessary parameters, η and α, are easily determined or somewhat universal. In the case of α, this is not entirely clear given its empirical origins. For a short net $-2e$ RNA strand, Yeh and Hummer estimate this correction factor $\alpha \approx 0.76$, which was considered reasonable given a value somewhat close to unity.[8] Potentially more concerning is the requirement of eq 1 for perfectly cubic simulation geometries. When this is not possible, determining experimentally comparable diffusion coefficients becomes problematic. Recently, Kikugawa et al. developed fit functions that can be used to correct diffusion coefficient estimations in the case of tetragonal distortions of simulation box geometries. In the case of an elongated tetragonal cell, where $L_x \geq L_y = L_z$

$$D_{y,app} = D_{z,app} = D_0 + \frac{k_B T \xi_{EW}}{6\pi\eta L_z} \cdot \left(\frac{a_{x/z} - 1}{a_0 - 1} - 1 \right) \quad (3)$$

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

Here, the $D_{z,app}$ value represents the calculated one-dimensional apparent diffusion coefficient in the z-dimension, with similar $D_{x,app}$ and $D_{y,app}$ values for x and y, $a_{x/z} = L_x/L_z$, and $a_0 = 2.79336$, which is presented as a universal constant determined from the authors' extensive study of Lennard−Jones particle simulations.[9] In principle, the modification in parentheses is simply a fit function that approaches −1 when the $a_{x/z}$ aspect ratio approaches 1, restoring the unperturbed cubic correction term. In the case of $D_{x,app}$, the function in parentheses is taken to equal −1. This indicates that diffusion along the long x-axis is dependent upon the smaller $L_z = L_y$ according to the standard cubic correction, something observed in their, and other, Lennard−Jones particle simulations.[9,11−13]

It should be noted that the tetragonal corrections for Lennard−Jones particles proposed by Kikugawa et al. have recently been independently placed on more analytical footing by Vögele and Hummer,[12] versions of which are being further explored by Simonnin et al.[13] The extended expressions use box lengths rather than aspect ratios and slightly different numerical fitting constants, these with a focus on retracted tetragonal cells ($L_x < L_y = L_z$) that are more appropriate for membrane simulations. In this particular work, we focus on the original numerical forms proposed by Kikugawa et al.,[9] with any suggested extensions being fully modular and applicable to alternate expressions.

In the case of diffusion of charged proteins or similar macromolecular particles, it would seem a combination of eqs 1 and 3 would provide a route to calculating experimentally comparable diffusion coefficients. Such an expression array would take the form

$$D_{y,app} = D_{z,app} = D_0 + \frac{k_B T \xi_{EW}}{6\pi\eta L_z}\cdot\left(\frac{a_{x/z} - 1}{a_0 - 1} - 1\right)\alpha,$$
$$D_{x,app} = D_0 - \frac{k_B T \xi_{EW}}{6\pi\eta L_z}\cdot\alpha \quad (4)$$

where again $L_x \geq L_y = L_z$, and all the relevant variables are the same as described previously. This set of D_0 relations was tested first for pure water, where $\alpha = 1$, and then for MCP-1 using an α derived from the cubic systems.

The effect of salt concentration on the dynamics of the MCP-1 protein was also investigated by simulating it in the presence of additional ions beyond the five Cl^- ions needed to counter the net charge of MCP-1 at pH 7. Cubic simulation box systems with the same dimensions as the water solvated MCP-1 systems were prepared by random insertion of ions and water to target 0.125, 0.25, and 0.5 M NaCl concentrations. In addition to these buffered protein simulations, simulations of just NaCl in water at the same concentrations needed to be performed in order to calculate the solvent viscosity of the electrolyte solutions. Like the analogous pure water simulations, determination of solvent viscosity and the apparent diffusion coefficients of these salt solutions used 3 to 6 nm simulation box edge-lengths.

Simulation Methods. Molecular dynamics simulations were performed using GROMACS 4.5.5.[15−17] The Amber99SB-ILDN force field was used for modeling of protein,[18,19] while water and ions were modeled using the TIP3P water model with the associated Amber force field ion types.[20−22] Protein setup involved PDB2GMX processing of the 1DOL Protein Data Bank structure,[23] followed by charge neutralization with five Cl^- ions. All of the previously described systems were individually prepared using random insertion of protein, Na^+ ions, and Cl^- ions, followed by a culled-overlap insertion of equilibrated water boxes.

For each system size and simulation type, 10 unique starting configurations were prepared and seeded with random velocities from a Maxwell−Boltzmann distribution at a target temperature of 310.15 K following 1000 kJ mol^{-1} nm^{-1} force converged steepest descent minimization. All were equilibrated for 150 ps of isotropic constant-pressure (1 atm) and constant-temperature (310.15 K) equilibration with the Parrinello−Rahman barostat and v-rescale thermostat using time constants of 10 and 1 ps, respectively.[24,25] Integration of the equations of motion was performed using the leapfrog algorithm with a time step of 3 fs, LINCS to constrain protein bond vibrations, and SETTLE to keep TIP3P water molecules rigid.[26,27] Smooth particle-mesh Ewald[28] with a real-space cutoff of 1.0 nm, spline order of 4, and energy tolerance of 10^{-5} was used for long-ranged electrostatics corrections. Lennard−Jones interactions were cut off at 1.0 nm with applied long-range energy and pressure corrections.[29]

Molecular dynamics trajectories were recorded for each simulation immediately following the NPT equilibration. Cubic system protein containing simulations were each run for 300 ns, while those without protein were run for 30 ns. In the case of tetragonal protein containing simulations, simulations were extended to 450 ns to provide additional sampling, as the D_{app} needed to be decomposed into x-, y-, and z-dimension components. Given that there were 10 independent simulations for each system composition and size, each tetragonal protein, cubic protein, and nonprotein simulation data point involved 4.5 μs, 3 μs, and 300 ns of respective aggregate sampling.

Apparent diffusion coefficients were calculated using the Einstein mean-square displacement (MSD) as a function of time relation

$$D_{app} = \lim_{t\to\infty}\frac{1}{2dt}\langle|r(t) - r(0)|^2\rangle \quad (5)$$

where d is the desired dimension (1, 2, or 3), and the displacement, $r(t) - r(0)$, is taken only over the dimensions of interest. The D_{app} is simply calculated from the slope of the linear region of the MSD as a function of the time interval. For the diffusion of water, the sheer number of water molecules results in excellent averaging, leading to a linear trend over most time intervals greater than the subpicosecond mean first collision time. The regression interval for water D_{app} values was taken over the 1-to-15 ns, and standard errors of these values were determined over the 10 independent simulation trajectories. For single molecules, like MCP-1 in this study, the averaging is much poorer. The regressions to determine protein D_{app} values were taken over the 2.5-to-25 ns, the observed most linear region. This choice was validated by subsequent iterative expansion of the regression range to the point where changes in the calculated D_{app} were not distinguishable from the accumulated error, and the resulting values overlapped with those from this selected time interval. In cubic systems, $d = 3$ in eq 5, while for tetragonal systems, $d = 1$ and separate $D_{x,app}$, $D_{y,app}$, and $D_{z,app}$ values were determined from particle displacement solely in the respective dimensions.

Mass Relation Analysis. In order to derive general insight into protein molecular weight (MW) to diffusion coefficient expressions, surface area and volume calculations were performed on a set of 40 protein structures of increasing mass, from 3.7 to 48 kDa.[30] All nonstandard components of the

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

structures, such as ligands, ions, hemes, and water particles, were removed from the selected structures in the interest of placing a protein sequence knowledge-based restriction on the explored series. Hydrogen atoms were cleared and added in idealized positions to Protein Data Bank structures for all proteins (see Table S1 in the Supporting Information for the list of proteins), and in the case of NMR structures with multiple deposited conformers, only the first conformer was used. MSMS 2.6.1 was used to calculate the solvent-accessible surface area (SASA),[31] solvent-excluded surface area (SESA),[32] and solvent-excluded surface area volume, all using a probe radius of 0.14 nm. While the same probe is used for both the SASA and SESA calculations, the surfaces differ in that the SESA represents the mostly smooth protein contact surface, while the SASA is the surface traced out by the center of the probe sphere and is not smooth. As the SASA originates from the solvent probe center, it is inflated relative to the SESA and typically reports larger values (see the table in the Supporting Information). One could physically describe the SESA as the protein surface and the SASA as the protein surface modulated by bound solvent. Regression of all these quantities versus protein MW provided a slope, and potentially nonzero intercept if desired, to form a function for estimation of a hydrodynamic radius, R_H, given the approximation that the SASA, SESA, and/or volume be applied to a sphere. These regression results were applied as coefficient inputs in the volume or surface area forms of the following related extensions of the standard Stokes−Einstein relation

$$D_0 = \frac{C_0 T}{\eta (\mathrm{MW} + C_1)^{1/3}} \tag{6}$$

for the MW to R_H approximation via general protein volumes, and

$$D_0 = \frac{C_0 T}{\eta \sqrt{\mathrm{MW} + C_1}} \tag{7}$$

for the MW to R_H approximation via general protein surface areas, either SASAs or SESAs. In both of these equations, C_0 and C_1 come from the slope and intercept of the protein geometric quantity versus MW regression, respectively. If the regression intercept passes through 0 nm^2 or nm^3 in the surface area or volume regressions, respectively, the C_1 parameter is eliminated, leaving only C_0 as a protein geometry to MW connection parameter. The performance of these geometry-only tethered diffusion relations was assessed through comparisons to experimentally measured values.

RESULTS AND DISCUSSION

Model viscosity and experimentally comparable diffusion coefficients for TIP3P water were initially determined from a series of increasingly sized cubic system simulations. The viscosity of the model at 310.15 K is needed for correction of noncubic system diffusion, water solvated protein diffusion, and *all* conversions of D_0 to D_η values for experimental comparisons. The TIP3P water solvent was used in this study due to its general acceptance in biomolecular simulations more so than its ability to accurately reproduce the experimental properties of real water. The large downside to this is that TIP3P is known to diffuse two to three times faster than real water, indicating that it has a significantly lower viscosity than experiment. In agreement with previous findings,[7] we observe this reduced viscosity with respect to experiment, where η_{expt} = 6.88 × 10^{-4} kg m^{-1} s^{-1} at this temperature.[33] The top panel of Figure 2 shows the system size dependence of η_{TIP3P}, and as

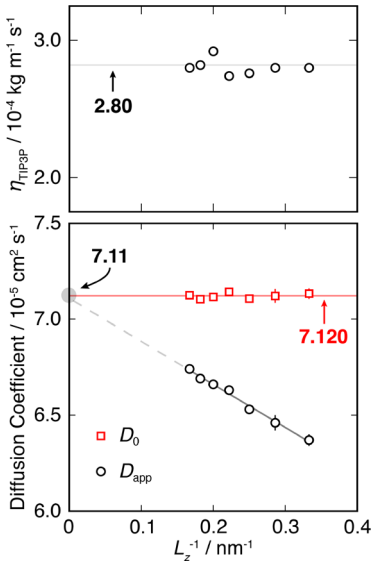


Figure 2. Viscosity (upper) and calculated D_{app} and corrected D_0 values (lower) for TIP3P water at 310.15 K as a function of inverse box size. The η_{TIP3P} is insensitive to changes in system size, while the D_{app} is quite strongly dependent on the choice of simulation size. The linear projection of D_{app} to infinite system size and average corrected value using eq 1 are in agreement at 7.11 ± 0.09 and 7.120 ± 0.005 × 10^{-5} cm^2 s^{-1}.

seen in other studies, it is insensitive to changes in simulation system size. The average value of 2.80 ± 0.02 × 10^{-4} kg m^{-1} s^{-1} is in agreement with the projected temperature trend observed by Venable et al.,[34] and it is expectedly much lower than η_{expt}, resulting in a viscosity correction ratio in eq 2 of 0.408.

From the lower panel of Figure 2, it is clear to see why it is important to consider correction of the D_{app} from individual molecular simulations. The D_{app} values increase linearly with inverse box edge-length up to a D_0 value of 7.11 ± 0.09 × 10^{-5} cm^2 s^{-1} at infinite box size. When applying eq 1 with a correction factor α = 1, there is strong agreement with this infinitely dilute value given that the average 7.120 ± 0.005 × 10^{-5} cm^2 s^{-1} overlaps the D_{app} regression value within error. Both of these values are significantly larger than the D_{expt}(310.15 K) value of 3.04 × 10^{-5} cm^2 s^{-1}.[33] Correcting the calculated D_0 value with the viscosity correction eq 2 results in a similar D_η = 2.904 ± 0.003 × 10^{-5} cm^2 s^{-1}, which, while slightly low, is in significantly better agreement with the experimental value.

Corrected Diffusion Coefficients for Water from Noncubic Simulation Boxes Are in Good Agreement with Cubic Box and Experimental Values. Moving beyond perfectly cubic simulation cells, the ability to accurately predict the diffusion coefficient for a molecule from an alternatively shaped simulation box would be advantageous for modeling studies of systems where a constraint is applied along one of the cell dimensions. To evaluate the accuracy of eq 3, we performed TIP3P water simulations in elongated tetragonal simulation boxes, where the *x*-dimension box length L_x was kept fixed at the maximum cubic box length of 6 nm, and L_y = L_z was reduced following the same series of smaller box size dimensions explored in the cubic box correction test above. To determine an overall D_η value from such noncubic simulations,

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

each dimension was corrected independently, and the resulting independent values were averaged. Table 1 shows a comparison of resulting D_η values from these simulations alongside the cubic system results. They agree well as the calculated error bars just overlap.

Table 1. Calculated D_η Values[a] for TIP3P Water at 310.15 K from Tetragonal and Cubic Simulations of Increasing System Size

tetragonal			cubic		
$1/\langle L_z\rangle$ (nm^{-1})	N_{wat}	D_η (10^{-5} cm^2 s^{-1})	$1/\langle L_z\rangle$ (nm^{-1})	N_{wat}	D_η (10^{-5} cm^2 s^{-1})
0.332	1780	2.92(2)	0.333	895	2.91(2)
0.284	2447	2.91(1)	0.286	1410	2.90(1)
0.247	3257	2.90(1)	0.250	2180	2.90(1)
0.221	4017	2.916(8)	0.222	3009	2.912(9)
0.199	4993	2.912(9)	0.200	4142	2.902(8)
0.182	5935	2.904(9)	0.182	5439	2.897(5)
0.167	7161	2.905(4)	0.167	7161	2.905(5)
avg:		2.911(4)			2.904(3)

[a]With standard error of the last digits in parentheses. .

These tetragonal system diffusion results used the a_0 aspect ratio constant numerically computed from Lennard–Jones simulations.[9] We found that the resulting D_η values were not strongly perturbed by moderate changes to this parameter. In fact, rounding the original a_0 = 2.79336 value to a_0 = 3 gives an average D_η result that overlaps within error. While we used the recommended constant in all tetragonal simulation correction calculations, it is possible that this empirical correction could be further refined or simplified.

Corrected MCP-1 Diffusion Coefficients Are Equivalent to Those Determined from Infinitely Dilute Projections. While the corrective techniques worked well in pure water simulations, charged biomolecule simulations potentially pose a greater challenge given the need to introduce and determine a potentially system specific α correction factor. Such a correction factor needs to be extracted from a linear regression of cubic simulation box D_{app} values as a function of increasing inverse box edge-length. A plot of this trend for the water solvated MCP-1 signaling protein is shown in Figure 3, with the regression intercept of 0.447 ± 0.027 × 10^{-5} cm^2 s^{-1}.

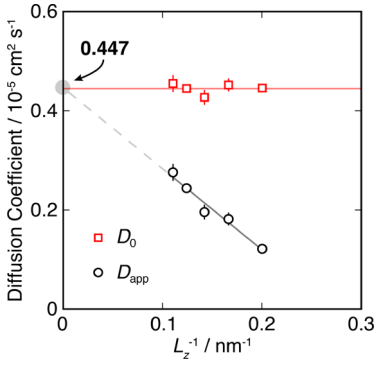


Figure 3. D_{app} and corrected D_0 values for TIP3P solvated MCP-1 at 310.15 K as a function of inverse simulation box size. The α correction factor in eq 1 was determined to be 0.710 from the intercept of the D_{app} regression, resulting in an average value of D_0 = 0.447 ± 0.005 × 10^{-5} cm^2 s^{-1}.

This intercept was used to set the α correction factor in eq 1 to α = 0.710, resulting in an identical average D_0 value of 0.447 ± 0.005 × 10^{-5} cm^2 s^{-1}. Correcting for the viscosity of the model solvent results in a D_η of 0.182 ± 0.002 × 10^{-5} cm^2 s^{-1}.

With this α value, we used eq 4 to evaluate the tetragonal simulation cell diffusion coefficients. Each of the dimension $D_{i,app}$ values were separately corrected, and the averaged D_η results over all dimensions are displayed in Table 2. The

Table 2. Calculated D_η Values[a] for MCP-1 in TIP3P Water at 310.15 K from Tetragonal and Cubic Simulations of Increasing System Size

tetragonal		cubic	
$1/\langle L_z\rangle$ (nm^{-1})	D_η (10^{-5} cm^2 s^{-1})	$1/\langle L_z\rangle$ (nm^{-1})	D_η (10^{-5} cm^2 s^{-1})
0.200	0.191(4)	0.200	0.183(3)
0.166	0.179(3)	0.166	0.185(5)
0.142	0.170(4)	0.142	0.175(6)
0.124	0.173(4)	0.124	0.182(3)
0.111	0.182(2)	0.111	0.186(7)
avg:	0.179(2)		0.182(2)

[a]With standard error of the last digits in parentheses.

averaged noncubic and cubic results overlap within the accumulated error, indicating that using this extended correction equation is a reasonable strategy for correcting charged biomolecule diffusion coefficients from tetragonal simulation cells. One other general thing to note comes from comparing the error values in Table 1 with those in Table 2. In the case of water diffusion, the statistical error gets progressively smaller with increasing system size, while for MCP-1 diffusion, there is no such trend. The lack of trend in the MCP-1 results is expected, because as the system size increases, the number of water molecules increases, but only a single protein is present. Since there is no increase in MSD data, unlike in the case of the pure water trajectories, the trend in statistical error should be flat with random fluctuations.

Effect of Salt Concentration on Protein Diffusion Is Only Partly Modeled by Solvent Viscosity Corrections. Salt or other cosolvents change the viscosity of solution environments. Incorporation of salt effects on protein dynamics is thus handled in an implicit fashion through determination or estimation of a system specific solvent η value. Given that we model salt ions explicitly in our molecular simulations, we can test this implicit salt effect by calculating a protein's diffusion coefficient as a function of salt concentration. To do this, TIP3P water simulations with increasing concentrations of NaCl were performed in order to determine a trend in η for these electrolyte solutions. While this trend is important for application in finite size simulation corrections described above, it is critical for viscosity correction of D_0 if experimental comparisons are expected. Table 3 shows the calculated viscosities of electrolyte solutions as a function of increasing salt concentration. In general, a linear trend is observed, as in experiments,[33] though the reported statistical error is generally somewhat large as these values are averaged over only four system sizes, excepting the 0 M concentration result which was calculated from the pure water simulations discussed previously.

These viscosity values were used in the construction of viscosity correction ratios for computing D_η values from simulations of an MCP-1 protein monomer in aqueous

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

Table 3. Calculated η Values, Experimental η Values,[33] α Correction Parameter, and Calculated D_η for NaCl in TIP3P Water at 310.15 K with Increasing NaCl Concentration

[NaCl] (M)	η_{sim} (10^{-4} kg m^{-1} s^{-1})	η_{expt} (10^{-4} kg m^{-1} s^{-1})	α	D_η (10^{-5} cm^2 s^{-1})
0.0	2.80(2)	6.88	0.710	0.182(2)
0.125	2.84(11)	6.95	0.662	0.166(1)
0.25	2.99(6)	7.04	0.642	0.159(2)
0.5	3.25(10)	7.20	0.606	0.157(1)

environments of increasing ionic strength. To account for possible changing of the α correction factor in eq 1 with changes in ionic strength, both α and D_η values were determined via D_{app} regression as a function of system size, and a new α value was determined for each concentration. The resulting α values in Table 3 follow a generally decreasing trend, indicating that this correction factor may not be purely dependent upon the formal charge of the solute.[8] There appears to be some degree of coupling between the solute protein and solvent charges at the given cosolvent concentrations.

In hindsight, this change in α should not be that unexpected given how solvent permittivity changes in response to increasing ionic strength.[35] With the introduction of salt, water is increasingly electrostricted in solvation shells about the ions, decreasing the static dielectric constant of the environment. For example, with a +5 net charge protein like MCP-1 here, the experimental Bjerrum length λ_B, the distance when the Coulombic potential energy between the solute and an external formal charge is equal to k_BT, is 3.5 nm at 298.15 K. If the static dielectric constant of the environment decreases from 80 to 70, λ_B would increase to 4.0 nm. Even before considering nonuniform perturbations of the electrostatic environment about the protein, there is an increasing interaction range with the decreasing solution permittivity. This increased range can further emphasize the need for the overly strong interaction correction that the α parameter provides.

Plotting the explicitly calculated D_η values alongside projected D_η values due solely to changes in viscosity as a function of increasing NaCl concentration in Figure 4 highlights a somewhat unexpected trend. Projecting the trend in diffusion simply through changes in experimental viscosity with salt concentration shows a modestly sloped decrease in D_η. Explicit calculation of the diffusion coefficient shows a significantly more dramatic sudden drop with the addition of salt ions and a somewhat similar smooth progression beyond this point. Why is this D_η response so different? Figure 5 shows the normalized ion probability density function with respect to distance from the protein surface for both Cl^- and Na^+ ions over these salt concentrations. Without the introduction of NaCl, Cl^- counterions distribute preferentially near the surface of the protein, forming a negatively charged ion cloud. With the introduction of 0.125 M NaCl, both Cl^- and Na^+ ions partition into this cloud in a drive to neutralize the net protein charge (see illustrations in Figure 5). This increases the effective concentration of ions in the immediate protein environment, localizing ions at or near the protein surface to increase both the protein hydrodynamic radius and the viscosity of the local solvent environment. This effective cloud of ions moves with the protein, acting to slow the diffusion of MCP-1 more than would be expected in a uniform salt environment.

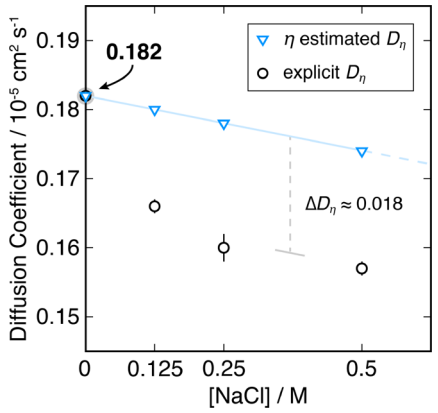


Figure 4. Simulation D_η for MCP-1 at 310.15 K as a function of NaCl concentration (black circles) alongside a simple experimental viscosity-based prediction of the diffusion coefficient (blue triangles). Of interest is the deviation of the actual diffusivity from the predicted trend based solely on change in solvent viscosity. The simulation D_η values show a stronger dependence upon salt concentration, primarily through a sudden drop with the initial introduction of NaCl.

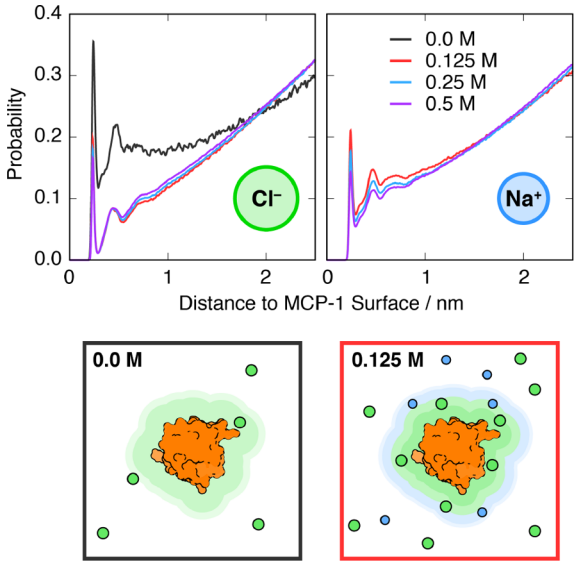


Figure 5. Normalized probability density function of ion occupancy as a function of distance from the MCP-1 protein surface over the different simulation salt concentrations. In the case of Cl^- ions (left), the counterion cloud at 0 M sees an enhanced occupancy probability near the protein surface relative to further out in solution as illustrated below. As the salt concentration increases, there is an influx of both Cl^- and Na^+ ions into the inner protein ion cloud in a drive to neutralize the net protein charge.

To better illustrate quantitatively how the ions partition about the MCP-1 signaling protein, we radially integrated populations of both the Cl^- and Na^+ ions about the geometric center of the flexible/dynamic protein. Figure 6 shows the to-scale results of this ion population analysis as a function of increasing salt concentration. As the protein is flexible in the molecular dynamics simulations, it is difficult to tightly resolve the preferred absolute location of ions relative to protein features, so we integrated the protein atom density out to the 90% occupancy level to use as an average protein contact surface. This is shown in the panels of Figure 6 as an orange background circle with radius r_{p90}. The r_{s1} and r_{s2} come from integrating the Cl^- counterion density out to the first and

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

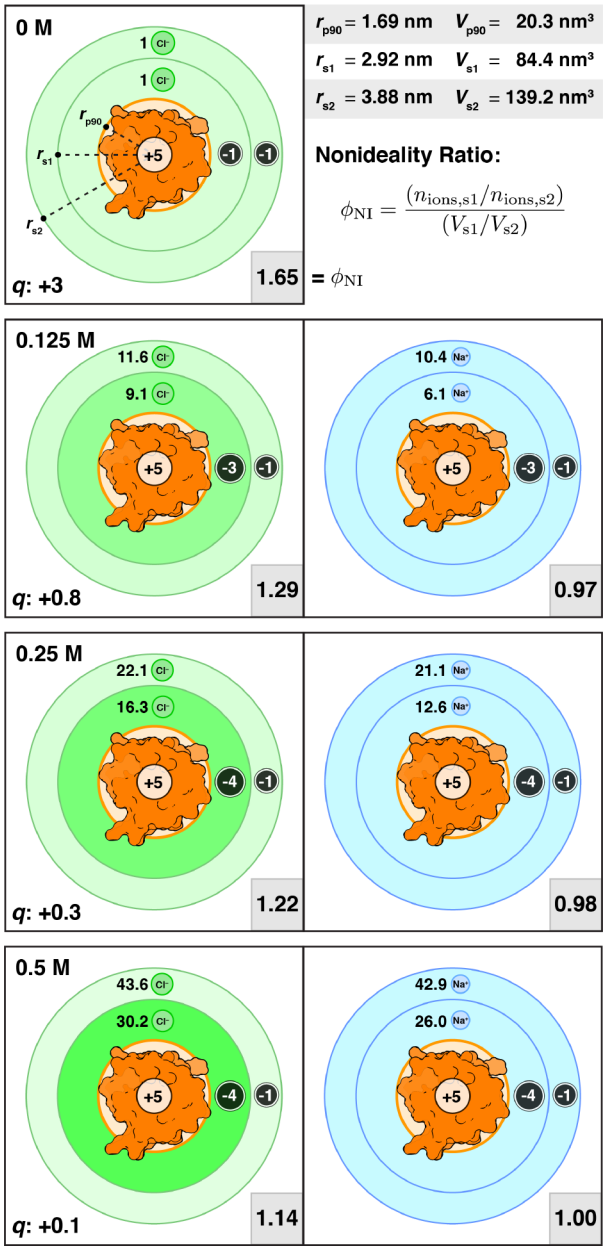


Figure 6. Integrated Cl^- (left panels) and Na^+ (right panels) ion populations about the 90% protein particle occupancy surface as a function of salt concentration. The bottom right of each panel lists the ϕ_{NI} nonideality ratio for ion populations in the ion cloud shells, while the bottom left of the Cl^- panels lists the net charge within the outer ion cloud sphere. The circled numbers indicate the net charge within each cloud shell, rounded to the nearest formal charge, and the color of the Cl^- shells indicates the asymmetry in absolute charge between them. In all Cl^- panels, the ϕ_{NI} is greater than unity, indicating a neutralization driven pressure for populating anions at the protein surface. This is further seen by the sudden drop of the net charge within the r_{s2} sphere upon addition of salt going from 0 to 0.125 M NaCl, with the protein near fully neutralized by 0.5 M NaCl.

second ion occupancy values, respectively, and these are taken to represent the first and second ion cloud shells about the protein. As a metric of the nonuniformity or nonideality of ion populations in the environment around the protein, we calculate a nonideality ratio ϕ_{NI}

$$\phi_{NI} = \frac{(n_{ions,s1}/n_{ions,s2})}{(V_{s1}/V_{s2})} = \frac{\rho_{ions,s1}}{\rho_{ions,s2}} \quad (8)$$

where $n_{ions,s\#}$ is the integrated number of Cl^- or Na^+ ions out to either the first or second ion cloud shells as indicated, and $V_{s\#}$ is the volume of the indicated ion cloud shells. A ϕ_{NI} value is the ratio of ion number densities in the inner ($\rho_{ions,s1}$) to the outer ($\rho_{ions,s2}$) cloud shell, so a value of 1 indicates a balanced ion density between the two ion cloud shells. Values greater or less than 1 indicate increasingly biased ion populations either near to or away from the protein's surface, respectively.

The most distinct aspects of the panels in Figure 6 are the large ϕ_{NI} values of the Cl^- ion cloud populations versus the near uniform spatial distribution of Na^+ ions and the change in net charge within the ion cloud environments. The MCP-1 protein has a +5 net charge, so counteranions populate preferentially nearby in order to neutralize this large formal charge. With no added salt, the large ϕ_{NI} value indicates that this neutralization pressure is a strong drive, though it is opposed by counterion distribution entropy. With the addition of 0.125 M NaCl, there is a sudden drop of the net charge within the ion cloud environment, indicating that the introduction of salt makes it easier to locally neutralize the protein charge, and the ϕ_{NI} greater than unity means that this neutralization is preferentially localized at the immediate protein surface. Increasing the salt concentration to 0.5 M leads to near complete neutralization of the protein within this ion cloud environment.

This ion cloud population behavior has implications for the protein diffusion coefficient in that the protein does not diffuse independent of its environment. The ion cloud has an outsized gain of Cl^- ions with even a small addition of salt in the protein solution environment. This leads to an increased drag, as the denser ion cloud needs to be pulled with the protein as it translates. This condition is consistent with the observed sudden drop in the diffusion coefficient seen in Figure 4 with addition of 0.125 M NaCl.

Protein Surface Area Connects to Hydrodynamic Radius More Strongly than Protein Volume. The calculated diffusion coefficients in the previous work are internally consistent, with cubic and noncubic simulation results giving nearly identical values. In the case of pure water, the D_η values compare well with experimental values. How do the MCP-1 diffusion coefficients compare with experimental results?

In the absence of direct experimental diffusion data, estimations of a given protein's diffusion coefficient can be crafted from molecular-weight-based relations. One of the most common is[3]

$$D = \frac{A}{MW^{1/3}} \quad (9)$$

where A is a connection constant that can be determined from known MW and diffusion data from single or sets of proteins. For example, using both the MW and $D_{20,w}$ of hemoglobin,[4,36] this constant has been estimated to be 2.82×10^{-5} cm^2 s^{-1} g$^{1/3}$ mol$^{-1/3}$.[37] Using this coefficient, the diffusion coefficient of MCP-1 (mass: 8.144 kg mol^{-1}) in pure water at 20 °C would be D_{expt} = 0. 140 × 10^{-5} cm^2 s^{-1}. This is ~30% off the calculated value, and this difference mostly seems to come from inflexibility in altering system temperature and viscosity conditions in eq 9. If one considers the temperature change

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

and resulting change in viscosity moving to 37 °C in the standard Stokes−Einstein relation

$$D = \frac{k_B T}{6\pi\eta R_H} \quad (10)$$

the estimated D_{expt} would increase to 0.216 × 10^{-5} cm^2 s^{-1}. Instead of the calculated MCP-1 value being 30% faster than this experimental estimate, the calculated value now appears to be ∼16% slower. A closer look at the data behind the recommended A coefficient uncovered that the MW of hemoglobin reported in the cited source is about 5.5% too large.[4] The mass comes from an early study on horse hemoglobin reported by Svedberg and Pedersen,[2] while the $D_{20,w}$ comes from a different study on human hemoglobin by Lamm and Polson.[38] Correcting this to use the known weight of human hemoglobin lowers the experimental estimate to 0.210 × 10^{-5} cm^2 s^{-1}, making the calculated MCP-1 D_η only 13% too low. Polson actually recommended an A value of 2.74 × 10^{-5} cm^2 s^{-1} $g^{1/3}$ $mol^{-1/3}$, built from a wider statistical fit to accepted diffusion coefficients and MWs at that time.[3] Using this A and applying corrections for differences in temperature and viscosity gives a D_{expt} = 0. 194 × 10^{-5} cm^2 s^{-1}, which is much closer to the calculated value of 0.182 × 10^{-5} cm^2 s^{-1}. Further corrections for molecular asymmetry and bound water could potentially lead to even greater agreement,[4,34,39−41] though the statistical nature of the fitting process behind the A value likely incorporates these considerations to some degree.

While it is encouraging that eq 9 can be readily adjusted to alter results and potentially improve agreement with expected values, this variability and statistical foundation also hides some of the reasoning for how and why it works. This particular mass-to-diffusion relation, like many other similar relations,[42−44] is based on two implicit assumptions: (1) proteins and other such molecules have uniform densities, and (2) the hydrodynamic radius, R_H, is well approximated by the radius of a sphere with a volume similar to that from the protein mass divided by this uniform density. These approximating assumptions came from before the first crystal structures of proteins were determined[45] and were actually adopted to estimate molecular weights from measured diffusion coefficients.[3]

Unlike when many of these relations where first devised, we now have many known protein structures whose volumes and surface areas we can directly calculate, and this body of knowledge provides significantly more detailed information from which to construct MW to diffusion relations. Is it possible to form an estimate of the diffusion coefficient from only a protein's sequence without statistically building in prior experimental knowledge of diffusion coefficients? Also, there have been many studies exploring volume and surface area relations for estimating the hydrodynamic radii of small molecules up through protein-sized structures.[39,40,44,46−49] Is the volume or the surface area a better geometric property to work through in estimating a hydrodynamic radius? Answering these questions could potentially lead to some additional insight into the driving forces behind diffusion in biomolecular systems.

To test this assumption, we calculated the SASA, SESA, and SESA volume for a set of 40 proteins with MW ranging from 3.7 to 48 kDa.[30] In order to base this effort purely upon sequence information, all nonstandard residue components, such as ligands, ions, hemes, water, etc., were removed from this set of structures, as mentioned in the Methods section. Figure 7 shows the data trends for these calculations as a function of MW. In all three cases, the trends are linear, though there is slightly more scatter in the surface area trends than the volume trend. Inserting the MW into linear fit functions from these trends would be a quick way to directly estimate a volume or surface area of a protein in the case that we only know the protein's sequence and thus MW. These values could then be converted directly into the R_H needed in the Stokes−Einstein relation. This is the goal of the proposed volume and surface area mass-to-diffusion relations in eqs 6 and 7.

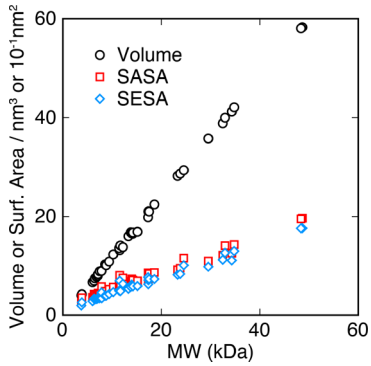


Figure 7. Calculated SESA volume, SASA, and SESA for a series of 40 protein structures, plotted as a function of molecular weight. All three are linear trends, though with differing slopes and differing intercepts if regression of the data is not set to pass through the origin.

The slopes and intercepts of the linear functions in Figure 7 were converted into a form that could be directly inserted into the two mass-to-diffusion relations. Two sets of parameters were determined for each of the SASA, SESA, and SESA volume trends. In one set, the regression is fixed to pass through the origin, so only the slope, and hence C_0 parameter, is used. In the other set, both the slope and intercept are used in the form of C_0 and C_1 parameters. In practice, the C_1 parameter acts as a general mass correction, providing some flexibility to the fitting trend if needed. Table 4 shows the resulting parameters for each of the fits. In these cases, the Vol0 and Vol1 sets are intended for the volume-based relation, eq 6. The remaining surface area sets are intended for the surface area-based relation, eq 7.

Table 4. Protein MW to Geometry C_0 and C_1 Parameters Extracted from Fits to Data in Figure 7 for the Proposed Volume[a] or Surface Area[b] Mass-to-Diffusion Relations, Both alongside the Predicted D_η for MCP-1 at 310.15 K Using the Listed Parameters in Their Associated Equation

method	C_0 (10^{-4} cm K^{-1} $kg^{4/3}$ $mol^{-1/3}$)[c] (10^{-4} cm K^{-1} $kg^{3/2}$ $mol^{-1/2}$)[d]	C_1 (kg mol^{-1})	MCP-1 D_η (10^{-5} cm^2 s^{-1})
Vol0	1.112	0.000	0.2466(7)
Vol1	1.107	−0.382	0.249(1)
SASA0	1.251	0.000	0.196(4)
SASA1	1.389	5.550	0.167(3)
SESA0	1.334	0.000	0.209(4)
SESA1	1.432	3.648	0.186(4)

[a]Eq 6. [b]Eq 7. [c]Units for Vol0 and Vol1. [d]Units for SASA0, SASA1, SESA0, and SESA1.

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609

Table 5. Volume-[a] and Surface-Area-Based[b] D_0 Predictions Using Only MW Knowledge of a Protein Sequence Compared with Experimental Protein D_0 Values[4]

protein	MW (kDa)	expt.	Vol0	Vol1	SASA0 (10^{-5} cm^2 s^{-1})	SASA1	SESA0	SESA1
ribonuclease	13.7	0.1190	0.1360(4)	0.1366(5)	0.099(2)	0.093(2)	0.105(2)	0.101(2)
lysozyme	14.3	0.1040	0.1340(4)	0.1345(5)	0.097(2)	0.091(2)	0.103(2)	0.099(2)
chymotrypsinogen	23.4	0.0950	0.1137(3)	0.1137(5)	0.076(2)	0.075(2)	0.081(2)	0.080(2)
β-lactoglobulin	35.2	0.0782	0.0993(3)	0.0991(4)	0.062(1)	0.064(1)	0.066(1)	0.067(1)
ovalbumin	41.9	0.0776	0.0937(3)	0.0935(4)	0.057(1)	0.059(1)	0.060(1)	0.062(1)
hemoglobin	64.5	0.0690	0.0812(2)	0.0809(3)	0.0456(9)	0.049(1)	0.049(1)	0.051(1)
serum albumin	66.3	0.0594	0.0822(2)	0.0820(3)	0.0465(9)	0.049(1)	0.050(1)	0.052(1)
catalase	227.1	0.0410	0.0533(2)	0.0531(2)	0.0243(5)	0.0266(5)	0.0259(5)	0.0276(6)
MUE			0.25	0.24	0.24	0.22	0.19	0.18
RMSD			0.26	0.26	0.26	0.23	0.21	0.19

[a]Eq 6. [b]Eq 7.

Initial testing of these relations was performed on the MCP-1 system to see which of the approaches was able to best reproduce the explicit simulation results. In them, T = 310.15 K, and the η was set to the experimental viscosity of water at this temperature, 6.95×10^{-4} kg m^{-1} s^{-1}. The primary trend in these results is that the surface-area-based relations outperform the volume-based relation when it comes to agreeing with the molecular simulation value of 0.182×10^{-5} cm^2 s^{-1}. Given that diffusion in molecular systems is primarily affected by collisions between particles, it is understandable that surface area would be a more well connected geometric property to the R_H. Also, in these cases, the inclusion of the C_1 parameter seemed to make little if any difference. SESA1 seems to benefit from this mass correction, but Vol1 and SASA1 show only minor to no improvement over the Vol0 and SASA0 variants.

It is important to note that the only structurally dependent input is the MW of MCP-1. The resulting predictions rely upon the assumption that the MW is connected to volume or surface area in the same manner as other proteins of similar MW, specifically the broad set used in developing the MW-to-geometry relation as depicted in Figure 7. Without specific considerations of protein flexibility,[50] cosolvent characteristics, and dimerization propensity,[51] an MW-based prediction will be intrinsically limited in accuracy. Of these specific considerations, multimeric propensity would be the easiest to treat. The predicted diffusion coefficient would simply come from an ensemble average of the properly weighted predictions from the multimer state MWs. In the case of MCP-1 here, the comparisons are to calculated values that by design do not have any possibility for dimerization.

The above result is only a single protein molecular weight data point compared with our corrected computational diffusion coefficients, making it difficult to fully evaluate the relations. Table 5 shows the results for prediction of D_0 using only sequence MW for a series of proteins provided by Tanford.[4] Again, only the molecular weight of each of these proteins was used as an input for estimating the diffusion coefficient. Across this series, it appears that approximating the R_H, and thus the overall diffusion coefficient at a given T and solvent η, via a MW to volume relation tends to predict slower diffusion coefficients than experiment, just as in the diffusion predictions for MCP-1 in Table 4. In the case of the MW to SASA trend, the predicted diffusion coefficients are faster than experiment, which also appears consistent with previous results. SESA sits between these two extremes and better reproduces the experimental D_0 values. It should be noted that while a 19–21% root-mean-squared deviation is an improvement over volume and SASA, pinning the trend to a known diffusion coefficient, as done using the traditional eq 9 relation, will be about twice as accurate. This is particularly true if the molecules used for determining the A coefficient are part of the evaluation set. The interest here is that no previously known diffusion information was provided. The trend in predicted diffusion coefficients with increasing MW seems to hold regardless of using volume or surface area. This indicates that a better accounting of the solvent-excluded protein contact surface would potentially lead to more accurate predictions of protein diffusion than currently possible using knowledge-based diffusion statistics.

CONCLUSIONS

Accurate computational determination of the diffusion coefficient of biomolecules is a challenging problem. The finite sizes necessary for performing molecular simulations of analogous systems pose a significant computational burden, and even accounting for this, a series of increasingly large simulations often needs to be performed in order to correct the dynamical quantities to make them experimentally comparable. Here, we show that it is possible to determine refined diffusion coefficients for a small signaling protein, MCP-1, even in noncubic simulation cells. The approaches used in this work were validated against experimental values for the diffusion coefficient and viscosity of water as well as experimentally based diffusion coefficient estimation from protein molecular weight.

In addition to pure water solvent simulations of counterion neutralized MCP-1, we explored the effect of salt concentration by performing similar simulations in the presence of 0.125, 0.25, and 0.5 M NaCl. We find that the ions populate the local environment around the protein to better counter its net charge and ion induced solvent charge inhomogeneities. This action, even at very low ionic strength, works to alter the local solvent environment around the protein and cause it to slow down more than expected from simple uniform viscosity changes due to the presence of salt.

Finally, this computational study of biomolecular diffusion motivated us to present a geometric investigation of molecular-weight-based relations for prediction of diffusion coefficients. Using only the connection between a protein's mass and the volume, SASA, or SESA, we show that it is possible to predict the diffusion coefficients of proteins without previous input knowledge of their diffusivity. While the predictions using the

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599–5609

proposed relations are not as accurate as those that use statistical fitting to known protein diffusion data, this effort provides an initial foundation for exploring alternate routes to general independent predictions of biomolecule dynamics.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.7b11846.

Plotted data for calculated signaling protein and water diffusion coefficients, as well as protein mass, volume, and surface areas used in this study (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: christopher.fennell@okstate.edu.

ORCID

Christopher J. Fennell: 0000-0001-8963-4103

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.J.F. would like to acknowledge partial financial support of the National Institutes of Health grant GM063592. This work was also supported in part by a grant from the National Institute of Biomedical Imaging and Bioengineering (1R15EB009527-01). The computing for this project was performed at the OSU High Performance Computing Center at Oklahoma State University supported in part through the National Science Foundation grant OCI-1126330. This material is in part based upon work supported by the National Science Foundation REU program under Grant No. CHE-1559874.

■ REFERENCES

(1) Svedberg, T. Zentrifugierung, Diffusion und Sedimentations-gleichgewicht von Kolloiden und Hochmolekularen Stoffen. *Colloid Polym. Sci.* **1925**, *36*, 53−64.

(2) Svedberg, T.; Pedersen, K. O. *The Ultracentrifuge*; The Clarendon Press: Oxford, U.K., 1940.

(3) Polson, A. Some Aspects of Diffusion in Solution and a Definition of a Colloidal Particle. *J. Phys. Colloid Chem.* **1950**, *54*, 649−652.

(4) Tanford, C. *Physical Chemistry of Macromolecules*; Wiley: New York, 1961.

(5) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: New York, 1987.

(6) Dünweg, B.; Kremer, K. Molecular Dynamics Simulation of a Polymer Chain in Solution. *J. Chem. Phys.* **1993**, *99*, 6983−6997.

(7) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873−15879.

(8) Yeh, I.-C.; Hummer, G. Diffusion and Electrophoretic Mobility of Single-Stranded RNA from Molecular Dynamics Simulations. *Biophys. J.* **2004**, *86*, 681−689.

(9) Kikugawa, G.; Nakano, T.; Ohara, T. Hydrodynamic Consideration of the Finite Size Effect on the Self-Diffusion Coefficient in a Periodic Rectangular Parallelepiped System. *J. Chem. Phys.* **2015**, *143*, 024507.

(10) Kikugawa, G.; Ando, S.; Suzuki, J.; Naruke, Y.; Nakano, T.; Ohara, T. Effect of the Computational Domain Size and Shape on the Self-Diffusion Coefficient in a Lennard-Jones Liquid. *J. Chem. Phys.* **2015**, *142*, 024503.

(11) Botan, A.; Marry, V.; Rotenberg, B. Diffusion in Bulk Liquids: Finite-Size Effects in Anisotropic Systems. *Mol. Phys.* **2015**, *113*, 2674−2679.

(12) Vögele, M.; Hummer, G. Divergent Diffusion Coefficients in Simulations of Fluids and Lipid Membranes. *J. Phys. Chem. B* **2016**, *120*, 8722−8732.

(13) Simonnin, P.; Noetinger, B.; Nieto-Draghi, C.; Marry, V.; Rotenberg, B. Diffusion Under Confinement: Hydrodynamic Finite-Size Effects in Simulation. *J. Chem. Theory Comput.* **2017**, *13*, 2881−2889.

(14) Placzek, G.; Nijboer, B. R. A.; Hove, L. V. Effect of Short Wavelength Interference by Dense Systems of Heavy Nuclei. *Phys. Rev.* **1951**, *82*, 392−403.

(15) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.

(16) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.

(17) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(18) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.

(19) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1950−1958.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(21) Case, D.; Cheatham, T.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. J.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(22) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.

(23) Lubkowski, J.; Bujacz, G.; Boque, L.; Domaille, P. J.; Handel, T. M.; Wlodawer, A. The Structure of MCP-1 in Two Crystal Forms Provides a Rare Example of Variable Quaternary Interactions. *Nat. Struct. Mol. Biol.* **1997**, *4*, 64−69.

(24) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(25) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(26) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.

(27) Miyamoto, S.; Kollman, P. A. SETTLE: An Analytical Version of the SHAKE and RATTLE Algorithms for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(29) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740−5761.

(30) Drechsel, N. J. D.; Fennell, C. J.; Dill, K. A.; Villà-Freixa, J. TRIFORCE: Tessellated Semi-analytical Solvent Exposed Surface Areas and Derivatives. *J. Chem. Theory Comput.* **2014**, *10*, 4121−4132.

(31) Lee, B.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−490.

(32) Connolly, M. L. Analytical Molecular Surface Calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548−558.
(33) Lide, D. R., Ed. *CRC Handbook of Chemistry and Physics*, 84th ed.; CRC Press, Inc.: Boca Raton, FL, 2003.
(34) Venable, R. M.; Hatcher, E.; Guvench, O.; MacKerell, A. D., Jr.; Pastor, R. W. Comparing Simulated and Experimental Translation and Rotation Constants: Range of Validity for Viscosity Scaling. *J. Phys. Chem. B* **2010**, *114*, 12501−12507.
(35) Hasted, J. B.; Ritson, D. M.; Collie, C. H. Dielectric Properties of Aqueous Ionic Solutions. Parts I and II. *J. Chem. Phys.* **1948**, *16*, 1−21.
(36) Lauffenburger, D. A.; Zigmond, S. H. Chemotactic Factor Concentration Gradients in Chemotaxis Assay Systems. *J. Immunol. Methods* **1981**, *40*, 45−60.
(37) Zhao, X.; Jain, S.; Larman, H. B.; Gonzalez, S.; Irvine, D. J. Directed Cell Migration via Chemoattractants Released from Degradable Microspheres. *Biomaterials* **2005**, *26*, 5048−5063.
(38) Lamm, O.; Polson, A. LXXVIII. The Determination of Diffusion Constants of Proteins by a Refractometric Method. *Biochem. J.* **1936**, *30*, 528−541.
(39) Edward, J. T. Molecular Volumes and the Stokes-Einstein Equation. *J. Chem. Educ.* **1970**, *47*, 261−270.
(40) Pastor, R. W.; Karplus, M. Parametrization of the Friction Constant for Stochastic Simulations of Polymers. *J. Phys. Chem.* **1988**, *92*, 2636−2641.
(41) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin from ^{15}N NMR Relaxation. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566.
(42) Scheraga, H. A.; Mandelkern, L. Consideration of the Hydrodynamic Properties of Proteins. *J. Am. Chem. Soc.* **1953**, *75*, 179−184.
(43) Young, M. E.; Carroad, P. A.; Bell, R. L. Estimation of Diffusion Coefficients of Proteins. *Biotechnol. Bioeng.* **1980**, *22*, 947−955.
(44) Tyn, M. T.; Gusek, T. W. Predition of Diffusion Coefficients of Proteins. *Biotechnol. Bioeng.* **1990**, *35*, 327−338.
(45) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042−1046.
(46) Bloomfield, V. A.; Dalton, W. O.; Van Holde, K. E. Frictional Coefficients of Multisubunit Structures. I. Theory. *Biopolymers* **1967**, *5*, 135−148.
(47) Venable, R. M.; Pastor, R. W. Frictional Models for Stochastic Simulations of Proteins. *Biopolymers* **1988**, *27*, 1001−1014.
(48) Hubbard, J. B.; Douglas, J. F. Hydrodynamic Friction of Arbitrarily Shaped Brownian Particles. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1993**, *47*, R2983−R2986.
(49) Carrasco, B.; García de la Torre, J. Hydrodynamic Properties of Rigid Particles: Comparison of Different Modeling and Computational Procedures. *Biophys. J.* **1999**, *76*, 3044−3057.
(50) Handel, T. M.; Domaille, P. J. Heteronuclear (^{1}H, ^{13}C, ^{15}N) NMR Assignments and Solution Structure of the Monocyte Chemoattractant Protein-1 (MCP-1) Dimer. *Biochemistry* **1996**, *35*, 6569−6584.
(51) Paolini, J. F.; Willard, D.; Consler, T.; Luther, M.; Krangel, M. S. The Chemokines IL-8, Monocyte Chemoattractant Protein-1, and I-309 are Monomers at Physiologically Relevant Concentrations. *J. Immunol.* **1994**, *153*, 2704−2717.

DOI: 10.1021/acs.jpcb.7b11846
J. Phys. Chem. B 2018, 122, 5599−5609