Stokes-Einstein-Debye Relation: A Check of Validity for Proteins in Nanoconfinement

Navaneeth Haridasan[1], Sridhar Kumar Kannam[2], Santosh Mogurampelly[3], Sarith P Sathian[1]
[1]Indian Institute of Technology Madras
Chennai - 600036, India
navaneethharidasan@gmail.com; sarith@iitm.ac.in
[2]Swinburne University of Technology
Hawthorn, Victoria-3122, Australia
urssrisri@gmail.com
[3]Temple University
Pennsylvania -19122, USA
santoshcup6@gmail.com

Abstract - Combined kinetic theory-hydrodynamics treatment has been proven effective in the prediction of biomolecule dynamics, generally if a single biomolecule is present in the bulk solvent. But the validity of such a theory in many physiological conditions is controversial. In the present study, a sample protein surrounded by other large biomolecules is approximated as the protein in a cylindrical nanopore. The hydrodynamic radius of the protein is chosen as an indicator to check whether one of the widely used kinetic theory-hydrodynamics relation namely Stokes-Einstein-Debye relation, is genuine for confined conditions of the protein. It has been found that Stokes-Einstein-Debye relation cannot be satisfied by the protein if confinement dimensions are very close. The reason for the violation can be attributed to van der Waals interaction between pore and the protein.

Keywords: Molecular dynamics, Protein, Nanopore, Hydrodynamic radius, Stokes-Einstein-Debye equation.

1. Introduction

From the pioneering works of Einstein on kinetic theory [1], it has been well understood that a large solute in a homogeneous solvent, due to collisions with solvent molecules, is far from being stationary and diffuses randomly. This diffusional motion known as Brownian motion, quantified as diffusion coefficient D, is coupled to the microscopic form of friction coefficient given by:

$$D = \frac{k_b T}{\zeta}$$

(1)

where, k_b is the Boltzmann constant, T is the temperature and ζ is the microscopic friction coefficient. In the limit of low Reynolds number or creeping flow, the microscopic friction coefficient can be defined using classical hydrodynamics or specifically, Stokes law [2]. The friction coefficient related to translational motion in the limit of creeping flow is given by:

$$\zeta = f\pi\eta r$$

(2)

where, η is the solvent viscosity, r is the radius of solute and $f = 4$ or 6 based on slip or no-slip boundary condition respectively on the surface of the solute. Eqn. (2) is known as the Stokes-Einstein (SE) relationship. Eqn. (1) is also applicable for rotational diffusion of a large solute in the solvent. The Eqn. (1) can be rewritten to account for the rotational friction coefficient as:

$$D_r = \frac{k_b T}{8\pi\eta r^3} \tag{3}$$

Eqn. (3) is commonly known as Stokes-Einstein-Debye (SED) relation [3], left-hand side of which is the rotational diffusion coefficient. Even though SE and SED equations are derived from a classical hydrodynamic perspective, it is valid for many microscopic systems. One such instance is a biomolecule-solvent system, solvent typically being water.

An important aspect of SE and SED equation is that transport and structural properties of the solute can be related. With an *a priori* knowledge of viscosity of the solvent and temperature of the system, the size of the solute can be calculated from the diffusivity or vice versa. Often the solute, which may be a biomolecule such as protein or DNA, may not have a perfect spherical shape and thus *r* represents an approximate measure of biomolecule size, commonly known as Stokes radius or hydrodynamic radius (r_h). It has been reported in the literature that diffusivity is underestimated when the hydrodynamic radius is calculated based on the molecular volume of the biomolecule [4]. To alleviate the discrepancy, hydrodynamic radius based on the hydrated volume of the molecule can be used. The hydrated volume consists of the biomolecule and few layers of mobile solvent molecules in biomolecule's immediate vicinity.

A biomolecule seldom exists in the solvent as a single entity but is surrounded by other soluble biomolecules in a natural biological environment. The presence of other biomolecules in the neighborhood, generally known as macromolecular crowding, influences the structure and dynamics of the biomolecules and in turn their functions. When the crowding agents (surrounding biomolecules) are substantial compared to the biomolecule under consideration, the crowding can be represented by confinement. Confinement in a biological sense implies limiting a biomolecule within a fixed boundary. For a low concentration of the crowding agents, the macromolecular crowding can be approximated by cylindrical confinement [5]. It is of interest to us to examine whether the classical hydrodynamic treatment combined with the kinetic theory of Brownian motion is valid for physiological conditions of biomolecule existence. Since a stronger relationship exists between structure and transport properties through SED relation ($D \propto 1/r_h^3$), we choose the same equation as the basis for our investigation.

2. Methods

The first step of our investigation is to examine whether SED equation is valid for a sample biomolecule in water as the solvent, after which the critical question of whether SED equation is correct for a biomolecule in confinement can be addressed. As a sample molecule, we choose a globular protein, ubiquitin (Protein Data Bank id: 1UBQ) which is found in almost all biological entities and is responsible for marking proteins for degradation. To simulate the biological environment with ubiquitin in bulk solvent, we resort to the method of molecular dynamics (MD). Classical molecular dynamics simulations represent atoms of a molecule as point particles which interact with each other based on a forcefield. A typical forcefield defines the potential energy of the system, which consists of bonded (bonds and angles) and non-bonded interaction (van der Waals and electrostatic) between the atoms. Subsequently, Newton's second law of motion is used to estimate acceleration, velocity, and position of the point particle with respect to time. The entire system evolution can thus be tracked in terms of position and velocities of individual point particles.

The first and foremost quantity to be calculated for verifying SED relation is the rotational diffusion coefficient of the biomolecule. Once the rotational diffusion coefficient D_r of the biomolecule is calculated at a given temperature and solvent viscosity, hydrodynamic radius (r_h) of the biomolecule can be calculated, which will serve as the reference value indicating the validity of SED relation. With the assumption that the angular displacement of the biomolecule is minimal due to collisions with solvent molecules, rotational diffusion of a biomolecule in the solvent is governed by [3]:

$$\frac{\partial c(u,t)}{\partial t} = D_r \nabla^2 c(u,t) \tag{4}$$

where, *c(u,t)* is the probability of finding a unit vector *u* attached to the biomolecule at time *t*. We resort to the isotropic approximation (biomolecule considered as a perfect sphere) of rotational diffusion for the sake of simplicity and ease of calculation. The solution to Eqn. (4), diffusion equation in spherical coordinates, with isotropic approximation, is given by [6]:

$$C_l(t) \ = \langle P_l[u(0).u(t)]\rangle = \langle P_l[cos\ \theta(t)]\rangle = \ e^{\ -l(l+1)D_r\ t} \ = \ e^{-t/\tau_l} \qquad (5)$$

where, C_l represents the time correlation function of unit vector u, P_l is the Legendre polynomial of rank l, and τ_l is the rotational correlation time of the biomolecule. Legendre polynomial of the second rank ($l = 2$) is most widely used to estimate rotational diffusion coefficients in experimental methods such as Nuclear Magnetic Resonance (NMR) spectroscopy and fluorescence depolarization anisotropy measurements (cite). Hence the second rank Legendre polynomial is used for the calculation of the rotational diffusion coefficient in Eqn. (5). Our choices of the unit vector u are the three mutually perpendicular principal axes of moment of inertia. Evolution of each one of the principal axes of inertia with respect to time is used in Eqn. (5) and all of them are averaged to obtain a single time correlation function, C_2 [7].

2.1. Simulation Details

It has been reported in the literature that for an accurate prediction of the rotational diffusion coefficients, simulation time should be more than two orders of the rotational correlation time. From NMR experiments [8], τ_2 comes around 5 nanoseconds (ns) and hence more than 500 ns (0.5 microseconds) simulation time is required. With conventional all-atom molecular dynamics, significant computational resources are necessary to reach the time scales of microseconds with the particular system under consideration. One of the methods to reach microsecond timescales is by reducing the number of point particles to represent the system and hence reducing the number of interactions also. The method of assigning more than one atom to a point particle within the framework of MD is known as coarse-grained molecular dynamics (CGMD). CGMD using MARTINI forcefield [9] implemented in GROMACS molecular dynamics package [10] is selected as the method of simulation. In MARTINI forcefield, four solvent atoms are approximated as one point particle or bead and protein is coarse-grained based on its residues (amino acid groups). It is to be noted that water is modeled as a chargeless bead using this forcefield and hence electrostatic interactions of the solvent are not considered in the simulation.

For assessing the validity of SED equation for the protein-solvent system, we simulated a single protein in solvent boxes of varying size. For bulk simulation, cubic simulation boxes with side lengths of 7.4 to 26 nm were chosen. The protein was first constrained to the center of the simulation box, after which simulation box was filled with solvent particles. The system is then simulated at constant pressure and temperature of 1 bar and 300 K with periodic boundary conditions in all three directions for 20 nanoseconds with a timestep of 10 femtoseconds (10^{-15}). After equilibration of the system, the constraint on the protein was removed and 1 microsecond production run with 20 femtoseconds timestep was performed at constant temperature (300 K) and volume of the simulation box. To add a nanoconfinement to the existing system of biomolecule-water, solvent particles outside the desired nanopore region were kept rigid. Thus the interaction parameters of the nanopore surface are the same as that of the solvent. In confined simulations, the nanopore length is kept fixed at 20 nm while the pore radius is varied between 2.5 to 10.0 nm. The equilibration and production run procedure is similar to the bulk system. The solvent viscosity in nanopores has been calculated in one of the previous works using MARTINI forcefield [11] and is taken as 0.00101 kgm^{-1}s^{-1} for both bulk as well as nanopore in this study.

2.2. HYDROPRO

A reference value of r_h is essential to check whether the SED relation is valid for the protein in bulk solvent. Since the rotational diffusion values obtained from different experimental techniques vary widely in the range, we rely on the predictions of a computer-program HYDROPRO [12]. HYDROPRO has been considered 'gold standard' in predicting hydrodynamic properties of rigid biomolecules based on atomic resolution specified by Protein Data Bank (PDB) structure file. The program output has been used as an 'experimental' value for the rotational diffusion coefficient in one of the previous studies [7]. As specified in the same literature, the hydrodynamic radius of the beads used for constructing the model was set to be 0.32 nanometer (nm). The reference value of the hydrodynamic radius of the protein, r_h was calculated based on Eqs. (1) - (3) with no-slip boundary condition as:

$$r_h = \sqrt{\frac{3}{4}\frac{D_t}{D_r}} \qquad (6)$$

where, D_t is the translational diffusion coefficient, and D_r is the rotational diffusion coefficient of the biomolecule predicted by HYDROPRO. A value of 1.67 nm was obtained from the above equation for the protein ubiquitin with PDB id 1UBQ.

3. Results and Discussions
3.1. Rigid Body Assumption
To compute rotational diffusion using Eqn. (5), it is necessary to obtain principal axes of inertia for every frame in the protein trajectory. It is assumed that the protein is rigid enough not to allow principal axes of inertia to change during the simulation. To justify the rigid body assumption, the structural conformation of the protein should be investigated. This check is done through the analysis of two quantities in the present study, namely root mean square deviation (RMSD) and radius of gyration (r_g). RMSD variation indicates the conformational change from a reference structure while the radius of gyration depicts the characteristic dimension corresponding to the globular approximation of the protein. RMSD and r_g are computed according to the following relation:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i(t) - x_i^{ref})^2} \qquad (7)$$

$$r_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i(t) - x_{CM}^{ref})^2} \qquad (8)$$

where, x_i corresponds to the position vector of ith bead of the protein, N indicates the number of beads constituting the protein, *ref* means the reference structure of the protein, which is the structure of the protein at the start of production run and x^{ref}_{CM} indicates the position vector of the protein center of mass in the reference structure of the protein. Figure 1 depicts the variation of RMSD and r_g of the protein in both bulk and pore cases for the entire simulation trajectory of 1 microsecond. For the bulk case, the largest simulation box size is considered while pore radius of 2.5 nm is considered from the pore simulations.

It is clear from figure 1 that the structural variations from the reference structure are minimal when protein is in bulk as well as in nanopore since the values of RMSD and r_g remains almost constant. Hence rigid body assumption is used in the present study for assessing rotational diffusion coefficient using the evolution of inertial principal axes. The radius of gyration of the protein is found to be ~1.2 nm and the same is used as the characteristic property for representing protein dimension throughout this study.
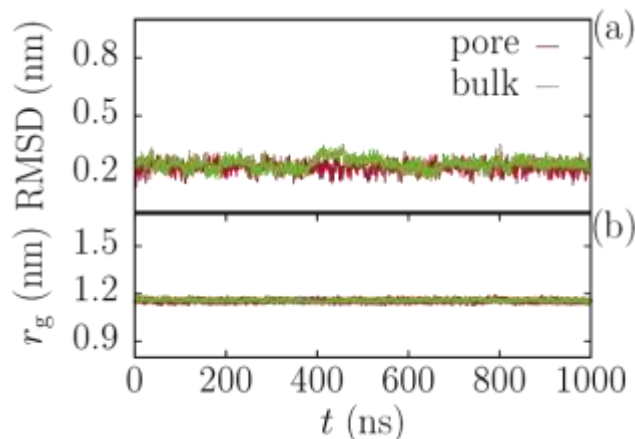
Fig. 1: Root Mean Square Deviation (RMSD) and radius of gyration (r_g)of the protein for the entire simulation trajectory. Both pore (r_p=2.5 nm) and bulk (L = 26 nm) cases are analyzed for the variation in RMSD and r_g.

3.2. Validity of SED Equation in Bulk

We computed the hydrodynamic radius of the protein, r_h in solvent for different box sizes i.e., by varying the number of solvent molecules contained in the box. Figure 2 shows the values of r_h for different box sizes denoted by the ratio L_0/L, L_0 being the least box length of 7.4 nm. The average value of r_h obtained from the bulk simulations is 1.65 ± 0.024 nm, which agrees well with the reference value of 1.67 nm obtained from HYDROPRO. This result indeed proves that SED relation is valid for the sample protein in water as a solvent.
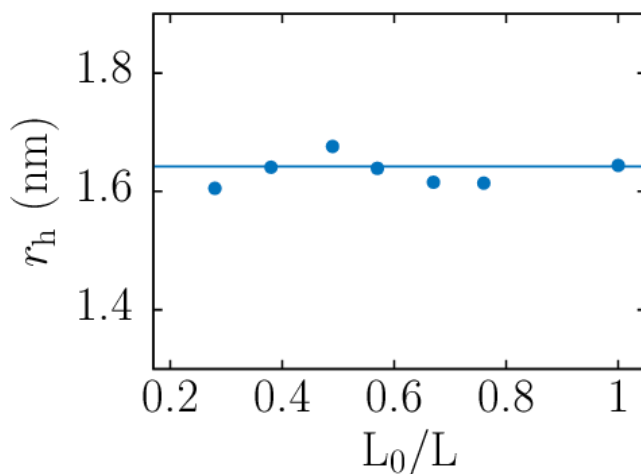

Fig. 2: Hydrodynamic radius of protein, r_h for different box sizes. L is the box length while L_0 is the minimum box length of 7.4 nm.

3.3. Hydrodynamic Radius Variation of the Protein in Nanopores

Hydrodynamic radius of the protein in nanopores was calculated subsequently to verify whether SED relation is valid for the protein in nanopores. The values of the hydrodynamic radius when pore radii r_p is varied from 2.5 to 10.0 nm are shown in figure 3. The figure also shows the average value of r_h for the protein in bulk solvent. For most of the pore radii considered in the simulation, r_h value for the protein in the pore is close to the corresponding bulk value. The least three values of pore radii, 2.5, 3 and 3.5 nm shows noticeable deviation from the bulk value. This shows a clear violation of SED behavior when protein is present in tighter confinement or when large crowding agents are closer to the biomolecule.
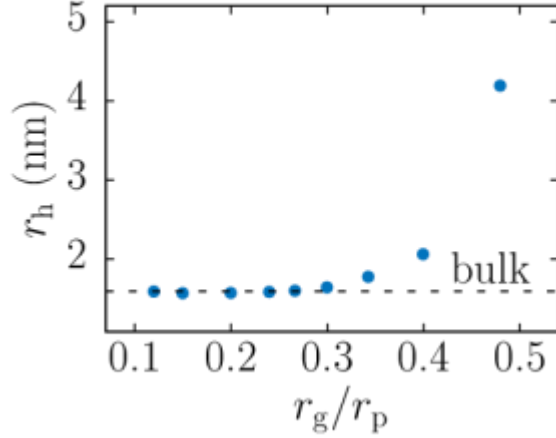
Fig. 3: Hydrodynamic radius of protein, r_h for varying pore radii r_p. Pore radii are normalized by protein radius of gyration r_g in *x axis*.

It has been reported previously that if conventional SE relation cannot be satisfied with a single hydrodynamic radius, one should look for the connection of intermolecular interactions with effective hydrodynamic radius [13]. Since both SE and SED equations are complementary to each other, we next probe into intermolecular interactions that occur within the system. Intermolecular interaction that affects protein dynamics in this particular study consists of van der Waals (vdW) interaction between protein-solvent and protein-nanopore. The variation of both theses intermolecular interactions are shown in figure 4 for different pore radii.
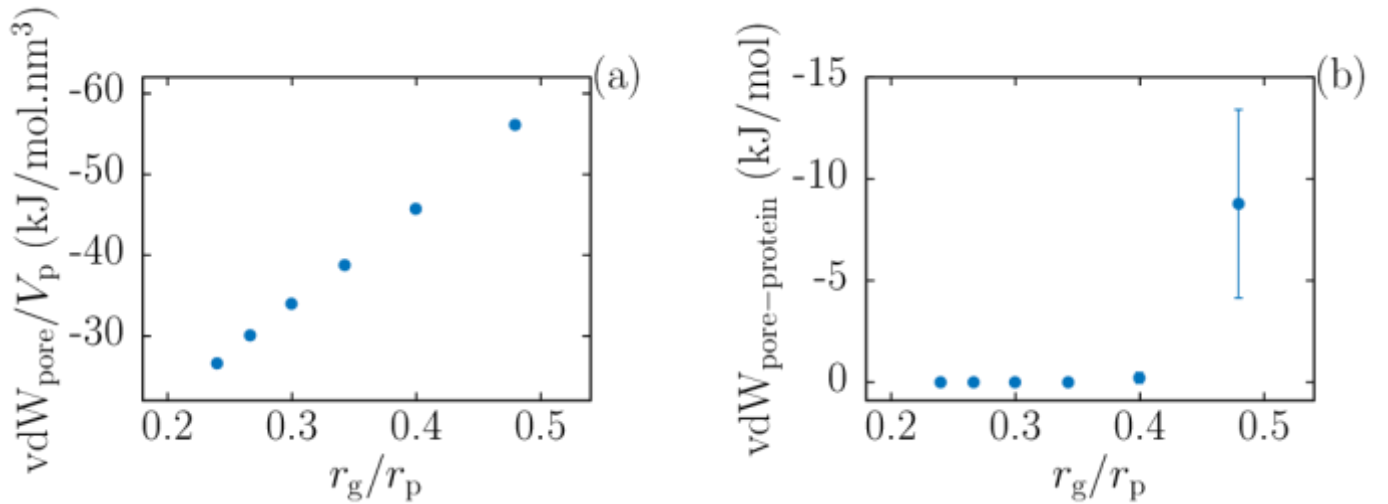

Fig. 4: Van der Waals (vdW) interaction that affects protein dynamics as a function of pore radii. On the *y axis*: (a) The interaction between solvent and protein occurring inside the pore normalized by pore volume V_p (b) Pore-protein interaction. For brevity, r_p 6, 8 and 10 nm cases are not shown in the figures.

Figure 4(a) shows the van der Waals interaction between protein and solvent occurring inside the pore normalized by the pore volume V_p. It can be seen that the interaction strength increase linearly with a decrease in pore radii. Since the solvent is interceding protein and pore, the interaction between protein and solvent is influenced by pore-solvent interactions. The increasing trend thus indicates a scale up in the interaction energy between pore and protein mediated through the solvent as pore radii decrease. No clear correlation between solvent-mediated interactions and hydrodynamic radius can be drawn from figures (3) and (4a). On the other hand, figure (4b) shows the direct interaction between pore and protein which is non-zero for the last three pore radii considered in the study. An excerpt of figures (3) and (4b) is given in table 1. A variation in the hydrodynamic radius from bulk value is observed when the pore-protein vdW interaction is non-zero. Thus, it can be

concluded from table 1 and figures (3) and (4) that direct protein-pore van der Waals interaction is responsible for the violation of SED relation in nanopore rather than solvent-mediated interactions between pore and the protein.

Table 1: Van der Waals (vdW) interactions between pore and protein and hydrodynamic radius (r_h) of the protein for different pore radii. The first three vdW interaction values are non-zero which can be related to the increase in r_h in smaller pores.

Pore radius, r_p (nm)	vdW$_{pore-protein}$ (kJ/mol)	Hydrodynamic radius, r_h (nm)
2.5	-8.79	4.19
3	-0.22	2.06
3.5	0.000437	1.77
4	0.0	1.64
4.5	0.0	1.61
5	0.0	1.63

4. Conclusion

Biomolecules seldom exist as a single entity in its natural environment but are surrounded by other soluble molecules. The validity of Stokes-Einstein-Debye (SED) relation in the limit of very large biomolecules surrounding a specific protein is investigated through protein in cylindrical confinements. The proximity effects of the surrounding biomolecules on the protein are recreated through varying the pore radii. It was found that when the confinement is tight or large biomolecular crowders are very close to the protein, SED relation is violated. The main reason behind the deviation from the SED relation can be related to the direct protein-pore or protein-crowder intermolecular interactions. While solvent-mediated interactions between crowders and protein (pore and protein) do increase with a decrease in proximity, the influence of solvent-mediated interactions on rotational diffusion coefficient is marginal and in turn, does not contribute to the violation of SED equation.

Acknowledgments

References

[1] A. Einstein, *Investigations on the Theory of the Brownian Movement*. New York: Dover Publications, 1956.
[2] G. G. Stokes, "On the effect of internal friction of fluids on the motion of pendulum," *Trans. Camb. Phil. Soc.*, vol. 9, no. 8, p.106, 1850.
[3] P. Debye, *Polar Molecules*. New York: Dover, 1929.
[4] H. Bertil and M. Davidovic, "Biomolecular hydration: from water dynamics to hydrodynamics," *Proc. Nat. Acad. Sci.,* vol. 10, no. 21, pp. 12135-12140, 2003.
[5] J. Mittal and R. B. Best, "Thermodynamics and kinetics of protein folding under confinement," *Proc. Nat. Acad. Sci.,* vol. 105, no. 51, pp.20233-20238, 2008.
[6] V. Wong and D.A. Case, "Evaluating rotational diffusion from protein MD simulations," *J. Phys. Chem B*, vol. 112, no. 19, pp.6013-6024, 2008.
[7] T. Frembgen-Kesner and A. H. Elcock, "Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins," *J.Chem. Theory. Comput.,* vol. 5, no. 2, pp. 242-256, 2009.

[8] A. L. Lee and A. J. Wand, "Assessing potential bias in the determination of rotational correlation times of proteins by NMR relaxation," *J. Biomol. NMR*, vol. *13*, no. 2, pp.101-112, 1999.

[9] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, "The MARTINI force field: coarse grained model for biomolecular simulations," *J. Phys. Chem. B*, vol. 111*, no. 27, pp.7812-7824, 2007.

[10] H. J. Berendsen, D. van der Spoel and R. van Drunen, "GROMACS: a message-passing parallel molecular dynamics implementation," *Comp. Phy. Comm,* vol. 91, pp. 43–56, 1995.

[11] S. K. Kannam and M.T. Downton, "Translational diffusion of proteins in nanochannels," *J. Chem. Phys.*, vol. *146,* no. 5, pp. 054108, 2017.

[12] A. Ortega, D. Amoros and J. García de la Torre, "Prediction of hydrodynamic and other solution properties of rigid proteins from atomic and residue-level models," *Biophys. J.*, vol. 101, pp. 892-898, 2011.

[13] R. Zwanzig and A. K. Harrison, "Modifications of the Stokes–Einstein formula," *J.Chem. Phys.*, vol.83, no. 11, pp. 5861-5862, 1985.