Linear Discriminant Analysis Based Machine Learning and All-Atom Molecular Dynamics Simulations for Probing Electroosmotic Transport in Cationic-Polyelectrolyte-Brush-Grafted Nanochannels

Raashiq Ishraaq and Siddhartha Das[1]*

[1]Department of Mechanical Engineering, University of Maryland, College Park, MD 20742,

United States

*Email: sidd@umd.edu

1

Abstract

Deciphering the correct mechanism governing certain phenomenon in polyelectrolyte (PE) brush grafted systems, revealed through atomistic simulations, is an extremely challenging problem. In a recent study, our all-atom molecular dynamics (MD) simulations revealed a non-linearly large electroosmotic flow (in the presence of an applied electric field) in nanochannels grafted with PMETAC [Poly(2-(methacryloyloxy)ethyl trimethylammonium chloride] brushes. Given the lack of any formal mechanism that would have directed us to identify the correct factors responsible for such an occurrence, we needed to spend several months and devote significant analysis to unravel the involved mechanism. In this paper, we propose a Linear Discriminant Analysis (LDA) based ML approach to address this gap. At first, we obtain data on certain basic features from the all-atom MD data for a reference case (case with a smaller electric field) and the perturbed case in bins in which the nanochannel half height has been divided into. These datasets are high-dimensional dataset, to which the LDA is applied. This leads to the projection of the data (between the reference and the perturbed states) in a highly separated form on a 1D line. From such LDA calculations, we are able to identify the importance scores for the different features, which in turn tell us what to study and where to study. Such knowledge enables us to rapidly identify the key factors responsible for the non-linearly large EOS transport in PMETAC-brush-grafted nanochannels.

1. INTRODUCTION

Grafting solid surfaces with charged polyelectrolyte (PE) brushes[1-10] ensures effective functionalization of such surfaces enabling these surfaces to be used in a myriad of applications such as fabricating surfaces with desired wettability[11] and protein adsorbability,[12] making nanochannels capable of enhanced current rectification,[13] sensing ions and biomolecules,[14-16] and generating nanoscale electrokinetic energy,[17-20] and designing nanoparticles for drug delivery,[21] oil recovery,[22] and water harvesting.[23] All these applications rely on the ability of these functionalized surfaces to strongly respond to changes in environmental stimuli, such as the pH and the salt content of the solution in which the brushes are present, temperature and humidity of the surroundings, etc. Also, capturing the equilibrium behavior of these brushes (in response to these stimuli) *using* different theoretical,[3-5,7,8,24-28] computational (both coarse-grained simulations[29-44] and all-atom molecular dynamics simulations[45-59]), and experimental methods[60-81] has been essential to optimize these applications and unravel novel facets of equilibrated PE brush behaviors.

Previously, the current author and his team developed all-atom molecular dynamics (MD) simulation frameworks for capturing the equilibrium behavior of such hydrated PE brushes (both anionic and cationic brushes).[45-50,54-58] The use of the all-atom MD framework enabled the identification of the role of each atom of the PE brushes in determining their properties and at the same time provided a detailed description of the PE-brush-supported water molecules and counterions. Das and co-workers also employed this all-atom MD simulation framework for capturing the effect of functionalizing the nanochannel walls with such PE brushes in regulating the electroosmotic transport and pressure-driven electrokinetic transport.[51-53,59] Specifically, for the case where the nanochannels were grafted with Na$^+$-ion-screened anionic PAA brushes, there

3

occurred overscreening (OS) of the brush charges,[51] which in turn led to coion-driven EOS flow for smaller electric field strengths.[51] This OS disappeared at larger electric field strengths, enforcing a counterion-driven EOS flow at greater field strengths; hence there occurred a reversal in the EOS flow direction by a mere change in the strength of the electric field.[51] In another study, it was shown that this co-ion-driven EOS flow in nanochannels grafted with Na^+-ion-screened PAA brushes enforced simultaneous energy generation and flow enhancement in presence of a pressure-driven background flow.[52] Finally, in a third paper, the effect of changing the screening counterions in triggering the OS effect (and therefore the nature of the EOS transport) in PAA-brush-grafted (PAA: poly-acrylic acid) nanochannels was probed.[53] In a recent study, we have extended our analysis by employing all-atom MD simulations for capturing the EOS transport in nanochannels grafted with *cationic PE brushes* [chloride-ion-screened PMETA brushes or PMETAC brushes; PMETAC: Poly(2-(methacryloyloxy)ethyl trimethylammonium chloride].[59] Most remarkably, we observed a non-linear increase in the EOS flow strength: a 2-fold increase in the electric field strength led to a much greater than 2-fold increase in the EOS volume flow rate. In order to figure out the mechanism causing such a non-linear OS flow strength increase, we had to carry out an extensive set of analysis, as we had little clue about which quantity to study. It took us to several months and several rounds of supporting simulations to eventually identify that such a non-linear increase in the flow strength is due to three interrelated factors. First, there occurs a unique deformation of the PMETA brushes, where it deforms in such a manner that a part of it becomes parallel to the brush-bulk interface and this deformation characterizes an *anti-electrophoretic* response of the PMETA brushes since despite being positively charged, it bends in a direction opposite to that of the electric field. Second, in response to this unique brush deformation, the counterions, too, get expelled from the interior of the brush layer, but instead of

4

leaving the brush layer completely, they localize in this same brush-bulk interface. As a consequence, the overall brush layer remains fully screened, although a significant fraction of the counterions becomes "freer" to move as they are not located deep inside the brush layer. Third, these counterions pull with them hydrated water molecules; therefore, there is now a significant fraction of counterion-supported water molecules at the "freer" location of the brush-bulk interface. Such a scenario ensures that under the application of the electric field, a significant fraction of the counterions and the water molecules hydrating these counterions move much more freely and therefore, at a significant faster velocity, eventually causing this non-linearly large EOS transport.

As mentioned previously, figuring out this complex physics governing the EOS transport in the PMETA-brush-grafted nanochannels consumed a significant amount of energy, time, and computational resources. This prompted us to ponder this question: *Is there a logical way to* <u>*narrow down the analysis space*</u> *for this problem so that it becomes possible to explain the mechanism of non-linear EOS transport in PMETAC-brush-grafted nanochannels much more promptly?* A common way of extracting information from a complex system is to apply some sort of perturbation and observe changes in the variables or features of interest. However, finding appropriate features is a challenging task. Creating these features or variables that describe the state of a system requires a lot of subject or domain knowledge. For example, consider the tetrahedral order parameter or q parameter of water molecules. This parameter was developed[82] to discuss how close to a tetrahedron-like structure a water molecule and its surrounding other water molecules are. The functional form of the q parameter required a lot of domain knowledge such as the fact that the water is a polar molecule, water creates four hydrogen bonds with its four neighbors, the information on the structure of water and its electronic structure, the fact that it acts

5

as a donor to two of its neighboring water molecules and acceptor to the other twos, etc. Compared to water, the PE brush is a much more complex system. When analyzing such systems, it is not generally known *apriori* to the researcher what variables to use that have a good physical meaning or can help to extract the appropriate physics or give some sort of information about the underlying mechanism that governs such systems. Especially for a complex molecular system, finding a suitable and meaningful variable that navigates through the analysis space is a daunting task. It would be of great value if there exists a protocol or an analysis method that is grounded on solid physical arguments. We tried to tackle this class of problem, and our general philosophical argument is that many simple variables (that don't require high domain intelligence), when combined with an appropriate machine learning (ML) algorithm, can be used to extract meaningful information about a complex molecular system. Many variables generate a high dimensional data space, which naturally motivated us to look into the available machine learning (ML) algorithms capable of analyzing and handling such high-dimensional data. However, the journey from transforming the raw data to the data used into the ML algorithm, making an appropriate choice of the ML algorithm, and identifying the right procedures to interpret the data is a long one that must be supported by solid physics-based arguments. It should also have some form of generality to it, so that the same protocol or some slight modification of it could be used for other systems as well.

As a first attempt at this class of problems, we came up with a strategy that is based on Linear Discriminant Analysis (LDA). In this strategy, we employ a ML method (e.g., LDA) on some very basic features derived from the MD simulation data; the result is that we effectively replace *high-domain knowledge by employing ML on low-domain knowledge.* Specifically, for this problem (namely, understanding the mechanisms enforcing a non-linearly large EOS transport in

6

nanochannels grafted with chloride-ion-screened PMETA brushes), this basic feature is $n_{x\text{-}y}$ (number of "y" atoms around a chosen "x" atom). We choose several such basic features in different bins (in which the simulation domain has been divided) for the cases where the system (nanochannels grafted with chloride-ion-screened PMETA brushes) has been subjected to electric fields of different strengths: we identify one case as a reference case and other cases as perturbed cases. Subsequently, we employ the LDA on these features and obtain clusters (i.e., map the high-dimensional data onto a lower dimensional space in such a way that the difference between the datapoints in the reference and the perturbed states is maximum) corresponding to each feature in each bin. Such low dimensional space that enables maximum separability is characterized by an axis that is obtained as a weighted combination of these features. These "weights" are the "importance score" of these different features. Hence, by identifying the appropriate axis that ensures the maximum separation between the data from the reference and perturbed states, we can identify the change in the "importance score" corresponding to the different features across the bins (for a given perturbation or a given change in electric field) or in a given bin for different electric fields (or perturbations). From that analysis (which tells us which feature's importance score undergoes the maximum change), we can pinpoint the appropriate basic feature that should be investigated at a specific location (or bin) in greater detail to better explain the non-linear increase in the EOS flow strength. Thus, our approach can pinpoint (and thereby significantly reduce the analysis space and hence the time and cost of that analysis) "*what to study*" and "*where to study*". We explain that using this approach we are able to guide the analysis towards probing in greater details the distributions of the monomers, chloride ions, and water molecules in the brush-bulk interface; accordingly, it becomes possible to quickly identify the role of the

7

localization of the monomers, chloride ions, and water molecules in the brush-bulk interface in ensuring the non-linearly large EOS transport in PEMTAC brush grafted nanochannels.

2. METHODS

2.1. Molecular dynamics simulations

All-atom molecular dynamics (MD) simulations were performed to investigate the coupled dynamics of PMETAC polyelectrolyte (PE) brushes, brush-supported water molecules, counterions, and the resulting electroosmotic (EOS) flow (in the presence of an applied axial electric field) inside a nanochannel grafted with such PMETAC PE brushes. The nanochannel dimensions were 268.3 Å × 89.4 Å × 147.72 Å. We employed periodic boundary conditions in x and y directions, and fixed boundaries in z direction. The PE brushes were grafted on both the nanochannel walls in a 4 × 12 array; therefore, there were 96 PMETAC chains with a grafting density of 0.2 chains/nm^2. Each PE chain contains 24 monomers and methyl end cappings. This setup captures the brush-nanochannel interface while avoiding complications from long-chain effects like entanglement.

The OPLS-all-atom forcefield[83] was used for modeling the PE brushes and validated in prior studies for capturing the solvation properties and the responses of the brushes to counterions. The system was solvated with SPC/E water, and counterions were added to neutralize the chains. Also, an additional 64 Na$^+$ and Cl$^-$ ions were included, representing a 0.2 M NaCl solution. Ion interactions were modeled using Joung-Chetham parameters.[84]

The system was first energy-minimized and equilibrated under NP$_z$T conditions (300 K, 1 atm) using a Nosé–Hoover thermostat and barostat, followed by further equilibration in the NVT ensemble.[45] Long-range Coulomb interactions between the charged entities present in the system were computed using the PPPM (particle-particle particle-mesh) method with slab geometry

9

adjustments.[45] A time step of 2 fs was used to integrate the velocity field using the Velocity Verlet algorithm.[45]

For flow simulations, axial electric fields (E=0.1,0.5,1.0 V/nm) were considered. The effect of the electric field on different atoms was represented by ascribing forces to these atoms that are proportional to the charges of these atoms (similar to that considered in our previous studies considering EOS transport in brush-grafted nanochannels[51,53,59]). Also, to remove the dissipated heat, a thermostat was applied to water and ions in one-quarter of the channel. A Langevin thermostat was applied to the brushes to mimic the thermal wall control. The system was considered to reach a steady state when the velocity profiles and the average brush heights stabilized. Much more details of this MD simulation procedure for capturing the EOS transport in nanochannels grafted with the PMETAC brushes have been discussed in our previous paper.[59]

2.2. Machine Learning Process

2.2.1. Philosophy of the Proposed Machine Learning Process

The standard approach of using all-atom MD simulations to reveal useful properties of a system (with or without the application of Machine Learning or ML) involves employing high-domain knowledge to the MD-simulation-derived trajectory information. For example, in several of our previous papers, we have utilized all-atom MD simulation generated trajectories for computing the water properties such as the tetrahedral order parameter (q) or dipole moment distribution (θ) inside the PMETAC brush layer.[55,57] In fact, in one of these studies,[57] we employed a ML protocol (a combination of Gaussian pre-processing and unsupervised clustering algorithm) on such all-atom MD generated information on q and θ of water molecules surrounding the $\{N(CH_3)_3\}^+$ group of the PMETAC brushes and demonstrated the occurrence of the multiple hydration states of the $\{N(CH_3)_3\}^+$ group.

Here, we propose an alternative: we show that we can employ a ML algorithm (namely Linear Discriminant Analysis or LDA) on very "basic" MD-derived features (or features representing "low-domain" knowledge) to generate information that replaces the need for high-domain knowledge for providing useful information on the system. In summary, therefore, we propose that ML-enabled low-domain knowledge can become equivalent to high-domain knowledge. Fig. 1 provides the summary of the ML protocol that will be employed here. The protocol consists of five basic steps: (1) selection of basic features; (2) generating appropriate data for feeding into the ML algorithm; (3) transforming the data so that the ML algorithm (when applied to this data) can provide more effective predictions; (4) actual application of the ML algorithm (LDA) to project the high-dimensional dataset into low-dimensional spaces in a manner such that the datasets have maximum separation; (5) utilizing the LDA-based separation to identify the importance score (and

11

their changes) associated with the different features and in the process identify what quantities should be analyzed and where for better understanding a given phenomenon. These five steps have been discussed in detail below.
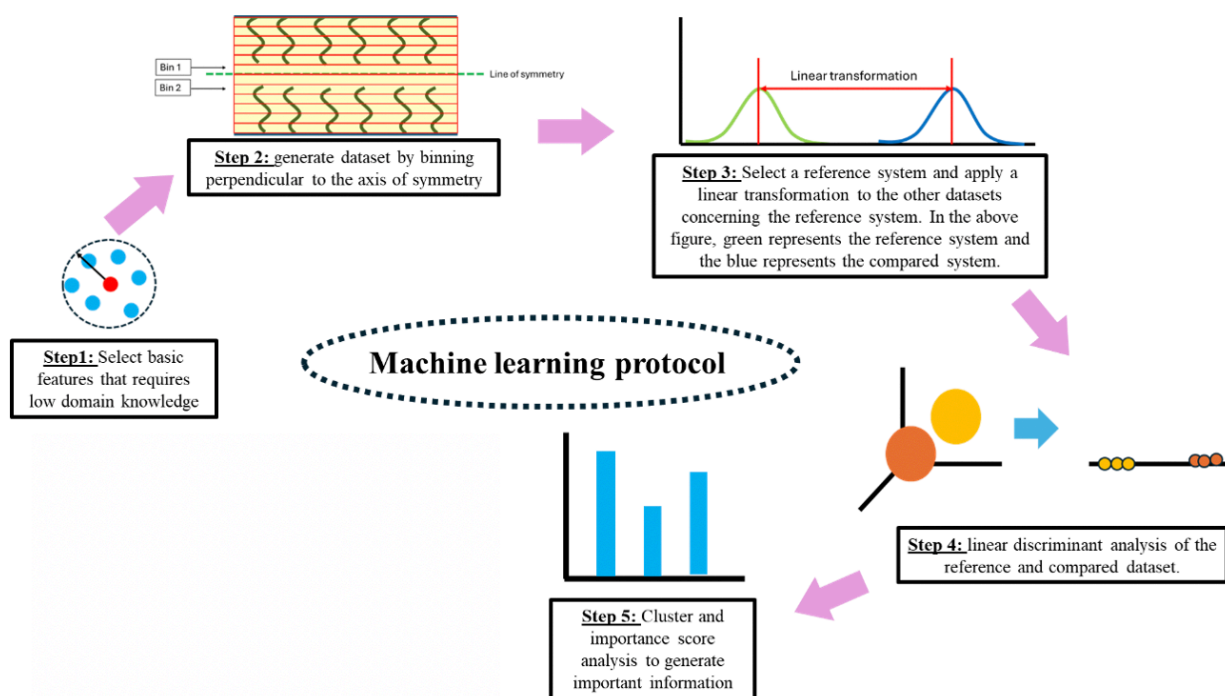
Figure 1. Linear Discriminant Analysis based Machine Learning Protocol used in the present study.

2.2.2. Selection of basic features

Molecular simulation algorithms generate trajectories of each particle in the system. Each particle represents either a single atom (for the case of all-atom MD simulations) or a collection of atoms (for the case of coarse-grained MD simulations). The relative positions of the different types of particles under different conditions carry the necessary information of the system. With this in mind, the simplest feature we could think of is the *number of particles of a certain type around a given center particle within a specified distance from this center particle*. We denote this number as n_{x-y}, where *x* is center particle and *y* is the particle surrounding this center particle, while the cut-off distance (considered as a hyper parameter) from the center particle used for this calculation is 7 Å (see Fig. 2). The biggest molecule in our system is the $\{N(CH_3)_3\}^+$ functional group of the PMETAC chain, which has a radius of 6 Å. The features that we chose are namely n_{N-Cl}, n_{N-Ow}, n_{Ow-Cl}, n_{Cl-Ow}, n_{N-N}, $n_{mon-mon}$ [where N, Cl, O_w, and "mon" refer to the nitrogen of the $\{N(CH_3)_3^+\}$, Chloride ion, water-oxygen-atoms, and PMETA monomers (or the carbon atoms of the PMETA monomer)].

A common way to probe the physics of a system is to apply some sort of perturbation to the system and observe how certain 'features' of the system change with the intensity of the perturbation. Identifying these 'features' requires a high domain knowledge. Our objective in this article is to create a ML protocol that uses features that require low human intelligence and reduce the analysis space.

Suppose there are '*m*' number of MD generated trajectories for an inhomogeneous, but asymmetric system. There are many things to analyze to come up with useful information or intuition. Such analyses mostly depend on the domain knowledge of the user. As mentioned above, we are using here a simple metric to gauge the physics of such a system: this metric is the number of particles

14

 ORCID: https://orcid.org/0000-0002-1705-721X

around a selected particle. Subsequently, we shall use a series of mathematical transformations that are based on physical intuition to extract meaningful information from the system.

The unperturbed system is denoted as a_0; suppose we apply a perturbation of 1, 2, and 3 to the system and the corresponding perturbed systems are denoted as a_1, a_2, a_3. At first, we create a high-dimensional dataset with low-dimensional features corresponding to each of these systems (a_0, a_1, a_2, and a_3). The comparison of the data (high-dimensional dataset with low-dimensional features) between the unperturbed and perturbed system holds the information needed to decipher the physics of the system. The difference between the features (associated with the unperturbed and perturbed states) will show us where the physics lies.
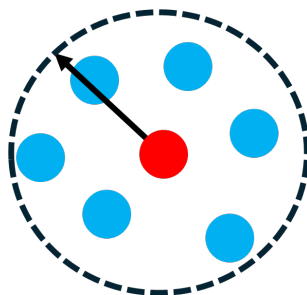
15

Figure 2. A schematic of the basic feature ($n_{x\text{-}y}$) used to probe the system. The feature is the number of particles (denoted as "y" and shown as blue circles) around a selected center particle (denoted as "x" and shown as the red circle) within a certain cutoff distance from the center particle (this cut-off distance is considered as 7 Å in the present case).

2.2.3. Generating data to feed into the ML algorithm

The features (or data describing the features) identified above contain information about the system; however, they need to be transformed in a manner that can be fed into the algorithm in such a way that the output has some physical meaning. The transformation needs to be in such a way that is inspired by the constitution of the problem or the system. There should have some form of generality, and the generality has to come from the nature of the problem. The system we are studying here (PE-brush grafted nanochannel) is an inhomogeneous, but a symmetric system. The system is inhomogeneous in the z-direction because the constitution of the system changes as one goes from the grafting surface to the middle of the nanochannel. At the same time, the system is symmetric (in z) about the of the centerline of the nanochannel. The system also has a symmetry in the x-y direction as well; in other words, for a constant z value, the composition along the x-y directions of the system can be thought of as homogeneous.

With such conditions under consideration, we applied two processes on the data. At first, we binned the features along the z direction; in other words, we divided the system (along z direction) into different bins and computed the features (corresponding to both unperturbed and perturbed systems) in these bins. Second, we averaged each of the features in each bin for a single snapshot. The purpose of averaging is that it will remove the noise.

Suppose there are 'm' number of MD generated trajectories for an inhomogeneous, but symmetric system. As mentioned above, we used a simple metric or feature (namely the quantity, n_{x-y}) to gauge the physics of such a system. Then we divided the system into 'd' number of bins above and below the line of symmetry. After that we obtained the features n_{x-y}s in each of these bins and averaged each of the features in each of the bins. After binning and averaging, we get therefore get $\mathbf{x} = \mathbf{m \times d}$ number of data points, with each data point having 'N' dimensions (the number of

17

features). The final dataset will look like as that shown in **Table 1** (with the information on the electric field and the bin number being identified as the "labels" of the data)

Table 1. Example of a representative dataset

Features							
E = 0.1 V/nm	*Bin #*	n_{N-Cl}	n_{N-Ow}	n_{O-Cl}	$n_{Cl-Ow,}$	n_{N-N}	$n_{mon-mon}$
	1	.	.	.	.	.	.
	2	.	.	.	.	.	.
	3	.	.	.	.	.	.
	4	.	.	.	.	.	.
	5	.	.	.	.	.	.
	6	.	.	.	.	.	.
	7	.	.	.	.	.	.
E = 0.5 V/nm	1	.	.	.	.	.	.
	2	.	.	.	.	.	.
	3	.	.	.	.	.	.
	4	.	.	.	.	.	.
	5	.	.	.	.	.	.
	6	.	.	.	.	.	.
	7	.	.	.	.	.	.
E = 1.0 V/nm	1	.	.	.	.	.	.
	2	.	.	.	.	.	.
	3	.	.	.	.	.	.
	4	.	.	.	.	.	.
	5	.	.	.	.	.	.
	6	.	.	.	.	.	.
	7	.	.	.	.	.	.

2.2.4. Data transformation

As mentioned above, a common way to probe the physics of a system is to apply some sort of perturbation and observe how certain 'features' (associated with the system) changes with the change in the intensity of the perturbation. Ideally these changes need to be large enough so that they can be detected in experiments. Therefore, we compare the features of each perturbed state (i.e., the state generated in the presence of an electric field with a specified strength) with that of the reference state only after applying a nonlinear transformation to each of the features (of the perturbed state) in a manner such that the difference between the features of the perturbed and the reference state gets magnified. This transformation (or the change) is ascertained in the following manner:

New mean of a feature of the perturbed state = Old mean of the same feature of the same perturbed state + change, (1)

Change=K (mean of the perturbative state-mean of the reference state). (2)

Here K is the hyperparameter.

This transformation ensures that the datasets are highly separated and the differences between the features of the perturbed and the reference state are significantly magnified. After the transformation, the dataset has information about the change in its feature w.r.t the reference state and that information will help us figure out the physics.

20

2.2.5. LDA analysis

After the transformation, we generate a high-dimensional dataset describing the features of the perturbed and the reference states. These data are characterized by labels, namely the bin number and the perturbation intensity (or the strength of the electric field) (see Table 1). For many systems, it becomes extremely useful to project such high-dimensional dataset into a lower dimensional space in such a way that the difference between the dataset associated with the reference state and the perturbative state is maximum. This is precisely the reason for which we employ the Linear Discriminant Analysis (LDA) to the high-dimensional data points of the reference and perturbed states. This ensures that these data points are projected into a lower dimensional space, where the separation between the classes (datapoints of the reference and the perturbed states) is maximum; Figure 3 shows an explicit pictorial representation of the action of the LDA.

21
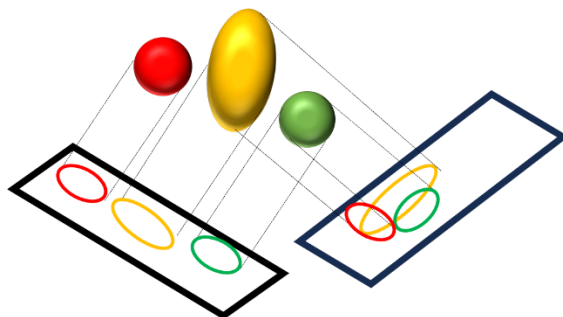
Figure 3. Pictorial representation showing the essence of the application of the LDA, enabling the projection of a high-dimensional (3D) data into a low-dimensional (2D) plane in two different ways. One type of projection (shown on the left) is more preferred than the other type of projection (shown on the right), since the former is more efficient in separating the high-dimensional data once they are projected into the 2D plane.

Mathematical formulation of the LDA process

Suppose our original high-dimensional dataset (consisting of data points representing the average of n_{x-y}s corresponding to the reference state and one perturbed state in a given bin of number k) is $\overline{\overline{x}}$. Thus, we can write $\overline{\overline{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, where x_1 and x_2 are the matrices representing the dataset (in a given bin of number k) corresponding to the unperturbed and the perturbed states, respectively. Thus, each of x_1 and x_2 consist of data points representing the averages of n_{x-y}s in the chosen bin. Therefore, if each of the class x_1 and x_2 contain N_d number of datapoints (here N_d is the same as the number of MD generated trajectories m, see section 2.2.3.), the $\overline{\overline{x}}$ matrix is a $6 \times 2N_d$ matrix (this "6" comes from the fact that each of these N_d datapoint in class x_1 or x_2 has a dimension of 6 or the number of features).

The use of the LDA will ensure that we employ a vector w to $\overline{\overline{x}}$ such that the result is a matrix $\overline{\overline{y}}$ (with $\overline{\overline{y}} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$), where the difference between the classes (i.e., between y_1 and y_2) will be maximum. Also, $\overline{\overline{y}}$ is a $1 \times 2N_d$ matrix since the high-dimensional dataset (of $\overline{\overline{x}}$ matrix) is projected (*via* LDA) on a 1D line. Under these circumstances, w should be a 1×6 matrix.

In other words, we start by considering the following operation:

$$\overline{\overline{y}} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = w\overline{\overline{x}} = w \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \qquad (3)$$

Given that we would like the classes (or y_1 and y_2) to be as far apart as possible (or the difference between the classes should to be maximized), we want to maximize the quantity $(\tilde{\mu}_1 - \tilde{\mu}_2)^2$, where $\tilde{\mu}_i$ ($i = 1,2$) denotes the matrix that represent the projected mean of class y_i ($i = 1,2$).

23

Using eq.(3), certain basic matrix identity [namely $(AB)^T = B^T A^T$, where A and B are matrices and X^T represents the transpose of matrix X], and considering μ_i to denote the mean of x_i (where $i = 1,2$), we can express the quantity $(\widetilde{\mu}_1 - \widetilde{\mu}_2)^2$ as:

$$(\widetilde{\mu}_1 - \widetilde{\mu}_2)^2 = (w\mu_1 - w\mu_2)^2 = [w(\mu_1 - \mu_2)][w(\mu_1 - \mu_2)]^T = w(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w^T =$$

$$w S_B w^T. \qquad (4)$$

In eq.(4),

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (5)$$

is the between-class scatter matrix.

In addition, we shall also like to minimize the scatter of the points among a given class (i.e., y_1 or y_2). Let us first consider the class y_1. Therefore, we would need to minimize the following quantity (with $j = 1,2,.....N_d$, where N_d is the total number of datapoints in the cluster represented by class y_1 and the dimension of each of these N_d datapoints is 1):

$$\widetilde{s}_1^2 = \sum_{j=1,2,..Nd}(y_1(j) - \widetilde{\mu}_1)^2 = \sum_{j=1,2,..Nd}(wx_1(j) - w\mu_1)^2 = \sum_{j=1,2,..Nd}[w(x_1(j) -$$

$$\mu_1)] [w(x_1(j) - \mu_1)]^T = \sum_{j=1,2,..Nd} w[(x_1(j) - \mu_1)] [(x_1(j) - \mu_1)]^T w^T =$$

$$w\{\sum_{j=1,2,..Nd}[(x_1(j) - \mu_1)] [(x_1(j) - \mu_1)]^T\}w^T = w S_1 w^T. \qquad (6)$$

In eq.(6), $y_1(j)$ and $x_1(j)$ represents the jth data point of the class y_1 (or the cluster represented by the class y_1) and the jth data point of the class x_1 (which is the class prior to the application of the LDA). Also in eq.(6),

$$S_1 = \sum_{j=1,2,..Nd}[(x_1(j) - \mu_1)] [(x_1(j) - \mu_1)]^T. \qquad (7)$$

Given that there are two classes (i.e., we have y_1 and y_2), we shall like to minimize

24

$$\tilde{s}_1^2 + \tilde{s}_2^2 = \boldsymbol{w}\boldsymbol{S}_1\boldsymbol{w}^T + \boldsymbol{w}\boldsymbol{S}_2\boldsymbol{w}^T = \boldsymbol{w}(\boldsymbol{S}_1\boldsymbol{w}^T + \boldsymbol{S}_2\boldsymbol{w}^T) = \boldsymbol{w}(\boldsymbol{S}_1 + \boldsymbol{S}_2)\boldsymbol{w}^T = \boldsymbol{w}\boldsymbol{S}_W\boldsymbol{w}^T, \quad (8)$$

where

$$\boldsymbol{S}_W = \boldsymbol{S}_1 + \boldsymbol{S}_2 = \sum_{j=1,2,..Nd}[(\boldsymbol{x}_1(j) - \boldsymbol{\mu}_1)]\,[(\boldsymbol{x}_1(j) - \boldsymbol{\mu}_1)]^T + \sum_{j=1,2,..Nd}[(\boldsymbol{x}_2(j) - \boldsymbol{\mu}_2)]\,[(\boldsymbol{x}_2(j) - \boldsymbol{\mu}_2)]^T. \qquad (9)$$

These two expressions that need to be minimized [see eqs.(4,5) and (8,9)] are captured together by Fisher's criterion, expressed as:

$$J(\boldsymbol{w}) = \frac{(\tilde{\mu}_1 - \tilde{\mu}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}. \quad (10)$$

Setting the derivative of the above quantity to zero we arrive at the following equation:

$$\boldsymbol{S}_W\boldsymbol{w} = \lambda\boldsymbol{S}_B\boldsymbol{w}. \quad (11)$$

Eq.(11) can be re-expressed as:

$$\boldsymbol{S}_B^{-1}\boldsymbol{S}_W\boldsymbol{w} = \lambda\boldsymbol{w}. \quad (12)$$

After solving the above eigenvalue equation [eq.(12)], we get the projection matrix $\boldsymbol{w}$, which will give us the required projected data. The top eigenvalue and the vectors are selected for projection.

2.2.6. Generating information from the LDA-generated data:

After being subjected to the LDA, corresponding to each bin, the datapoints appear as separated clusters (one cluster representing the datapoints corresponding to the reference state or the class $\boldsymbol{y}_1$ and the other cluster representing the datapoints corresponding to the perturbed state or the class $\boldsymbol{y}_2$) on a 1D line. Application of the LDA ensures that we now have projected the data onto the lower dimensional space in the **axis of maximum separability**: such dataset becomes easier to analyze. Of course, now that we have projected the data into the lower dimensional space (1 D line), the features lose their inherent physical meaning, given that the new axis is a linear combination of multiple features (or the averages of $n_{x\text{-}y}$s in the given bin for the reference and the perturbed states) and doesn't have any inherent physical meaning. The new axis provides us with the weights of the features. These weights are nothing but the components of the calculate $\boldsymbol{w}$ matrix (which is a 1×6 matrix). The six components of the 1×6 $\boldsymbol{w}$ matrix are the weight functions of six features. Thus, writing the $\boldsymbol{w}$ matrix as below [see eq.(13)]

$$\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix}, \qquad (13)$$

we can identify the weight functions of the six features as w_1, w_2, ….w_6. Therefore, the ***importance score*** associated with the six features are $|w_1|$, $|w_2|$, ….$|w_6|$. We calculate this importance score for a given bin for two different scenarios. In both the scenarios, we compare one reference state with one perturbed state. For both scenarios 1 and 2, the reference state is the case of the PMETAC brush grafted nanochannel subjected to an axial electric field of 0.1 V/nm.

26

On the other hand, the perturbed state for scenario 1 (2) is the case of the PMETAC brush grafted nanochannel subjected to an axial electric field of 0.5 V/nm (1 V/nm).

In addition to the importance score ($|w_i|$) associated with a given feature in a given bin for a given scenario (1 or 2), we also study the relative contribution of each importance score in a given bin for a given scenario by studying the quantity $\frac{|w_i|}{\Sigma_{\beta=1}^{6}|w_\beta|}$ (where the summation in the denominator is conducted over the number of features, which is 6).

There are three major forms of information that get revealed after applying LDA on the bins: (1) the absolute value of the feature; (2) the difference of the importance score (associated with a given feature) among the bins for different perturbation values; (3) difference of the importance score (associated with a given feature) within a specific bin for different perturbation values. Using such information, we will give some predictions about our system in a manner that is commensurate with the physics of the system. Since these are system-based quantifications, there is no-one-size-fit all procedure, but the reader will get an idea of how to modify it for a specific problem.

3. RESULTS

3.1. Features of the equilibrated system

At the start, it is important to identify some basic features of the system (nanochannel grafted with chloride-ion-screened PMETA brushes) in equilibrium. There are four components in the system: PMETA chains, water, counterions (Cl⁻ ions) and co-ions (Na⁺ ions as part of the added NaCl salt). Fig. 4 provides the equilibrium distributions of these components. Near the walls, the PE chain density is high as evident from the monomer distribution [see Fig. 4(a)]. The counterions screen the charges of the PE chains; accordingly, their distribution is very similar to the monomer distribution [see Fig. 4(b)]. The water number density is small inside the PMETAC brush layer due to the lack of available space, but high in the middle of the nanochannel (i.e., outside the brush layer) [see Fig. 4(c)]. Finally, the coions are mostly remain in the middle of the nanochannel [see Fig. 4(d)]. These features of the system need to be kept in mind, which will further help to decipher the physics.
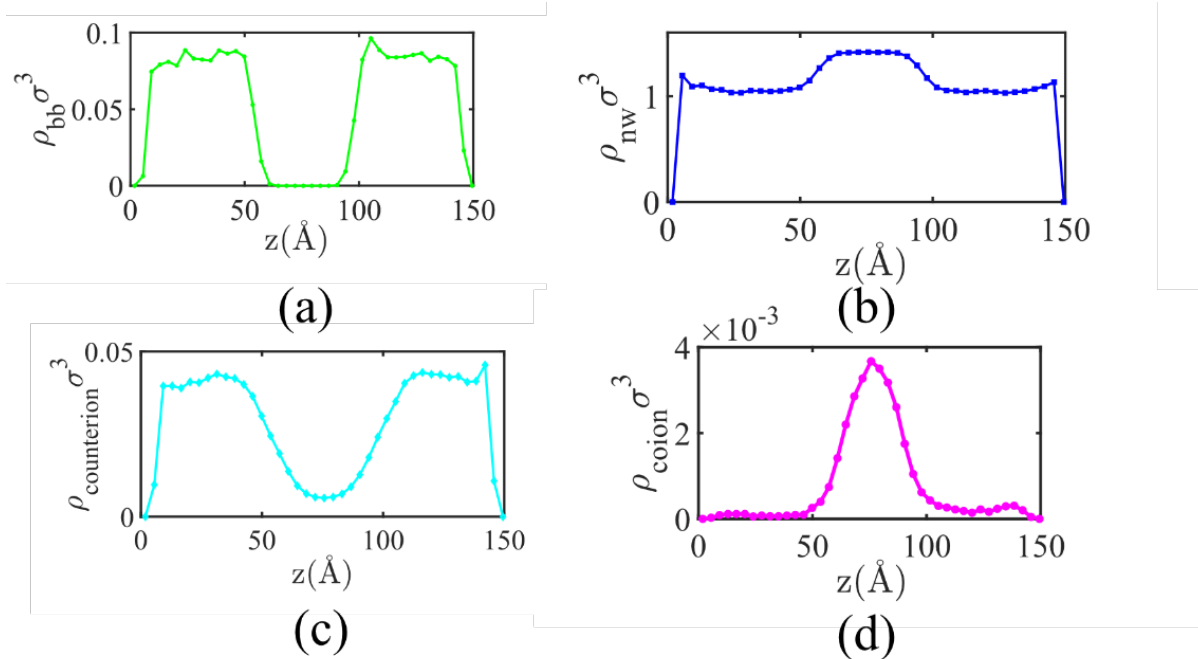
Figure 4. The dimensionless distribution of the various components of the PMETAC-brush-grafted grafted nanochannel system: (a) monomer distribution; (b) water number density distribution; (c) counterion distribution; and (d) coion distribution. In these figures, σ represents atomistic length scale.

https://doi.org/10.26434/chemrxiv-2024-7gs4p ORCID: https://orcid.org/0000-0002-1705-721X Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

3.2. Machine Learning Data

From the all-atom MD simulation generated data, we first obtain the average of each of the quantities n_{N-Cl}, n_{N-Ow}, n_{O-Cl}, n_{Cl-Ow}, n_{N-N}, $n_{mon-mon}$ in each of the seven bins in which the nanochannel half height has been divided into for each of the three conditions (E=0.1 V/nm, E=0.5 V/nm, and E=1.0 V/nm). The case corresponding to E=0.1 V/nm is referred to as the reference case, while those corresponding to E=0.5 V/nm and E=1.0 V/nm are referred to as the perturbed cases. We would like to compare each of these features n_{x-y} (or more appropriately, the average of each of these features) in a given bin between the reference case and the perturbed case. For example, we would like to know how n_{N-Cl} (or the average of n_{N-Cl}) in bin k (k = 1, 2,…,7) changes as one changes the applied electric field from 0.1 V/nm to either 0.5 V/nm or 1.0 V/nm. In order to ensure that this change is well magnified, we employ a non-linear transformation to n_{N-Cl} (or the average of n_{N-Cl}) in bin k corresponding to the perturbed cases (i.e., the cases corresponding to E=0.5 V/nm and E=1.0 V/nm). At this point, we now have high-dimensional and well-separated datasets (one corresponding to the reference case, while others corresponding to two perturbed cases).

We now apply the Linear Discriminant Analysis (see section 2.2.5. for the mathematical details) to this high-dimensional dataset and ensure that they are projected in a low-dimensional space (here a 1D line) in such a way that the clusters (representative of the datasets in a given bin) are maximally separated. Each cluster represent a given state: the reference state or any one of the two perturbed states, corresponding to a given bin. A given cluster in a given bin corresponding to a given state (reference or perturbed) consists of datapoints associated with the average value of n_{x-y} in that bin corresponding to the chosen state. Thus, we get clusters corresponding to bin k (k = 1, 2,…,7) for either the reference or the perturbed state.

As has been already elucidated in mathematical description (see section 2.2.5.), we start with a high dimensional dataset (represented by a $6 \times 2N_d$ matrix) and the application of the LDA projects this dataset into a $1 \times 2N_d$ space; in other words, application of the LDA leads to two highly separated data clusters (each having N_d number datapoints) projected on a 1D line.

Figs. 5 and 6 show these separated data clusters (projected on a 1D line) for the two separate situations for all the seven bins in which the nanochannel half height has been divided into. As already pointed out, situation 1 represents the case where we compare the reference case (case where we consider the transport in the PMETAC-brush-grafted nanochannels in the presence of an applied electric field of E=0.1 V/nm) with the perturbed case (case where we consider the transport in the PMETAC-brush-grafted nanochannels in the presence of an applied electric field of E=0.5 V/nm). On the other hand, situation 2 represents the case where we compare the reference case (same as situation 1) with the perturbed case (case where we consider the transport in the PMETAC-brush-grafted nanochannels in presence of an applied electric field of E=1 V/nm).
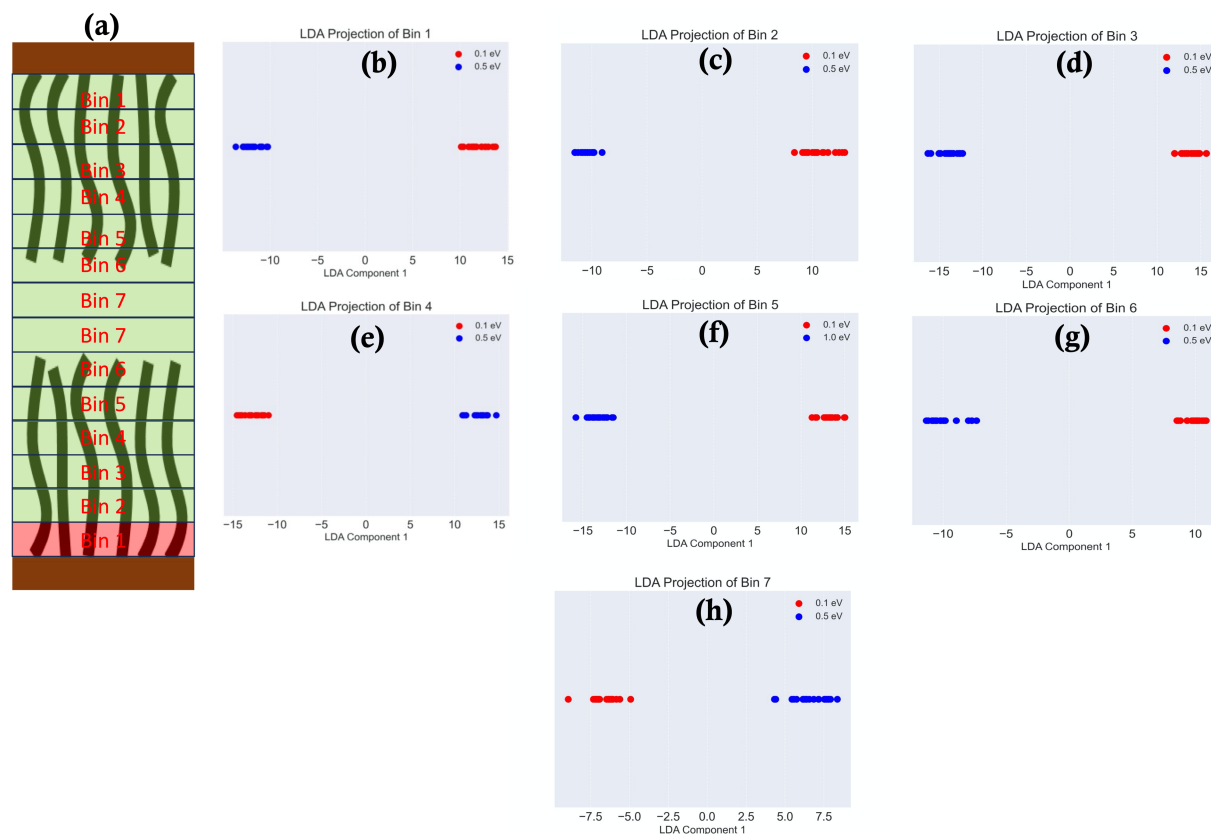
31

Figure 5. (a) Schematic showing the manner in which the brush grafted nanochannel is divided into two sets of 7 bins. One set of seven bins is from the nanochannel wall to the nanochannel centerline, with the bin number increasing as one goes from the wall to the centerline. LDA projection of the datapoints resulting in two separate clusters on a 1D line for situation 1 (reference state: case with E=0.1 V/nm and perturbed state: case with E=0.5 V/nm) for (b) Bin 1, (c) Bin 2, (d) Bin 3, (e) Bin 4, (f) Bin 5, (g) Bin 6, and (h) Bin 7.
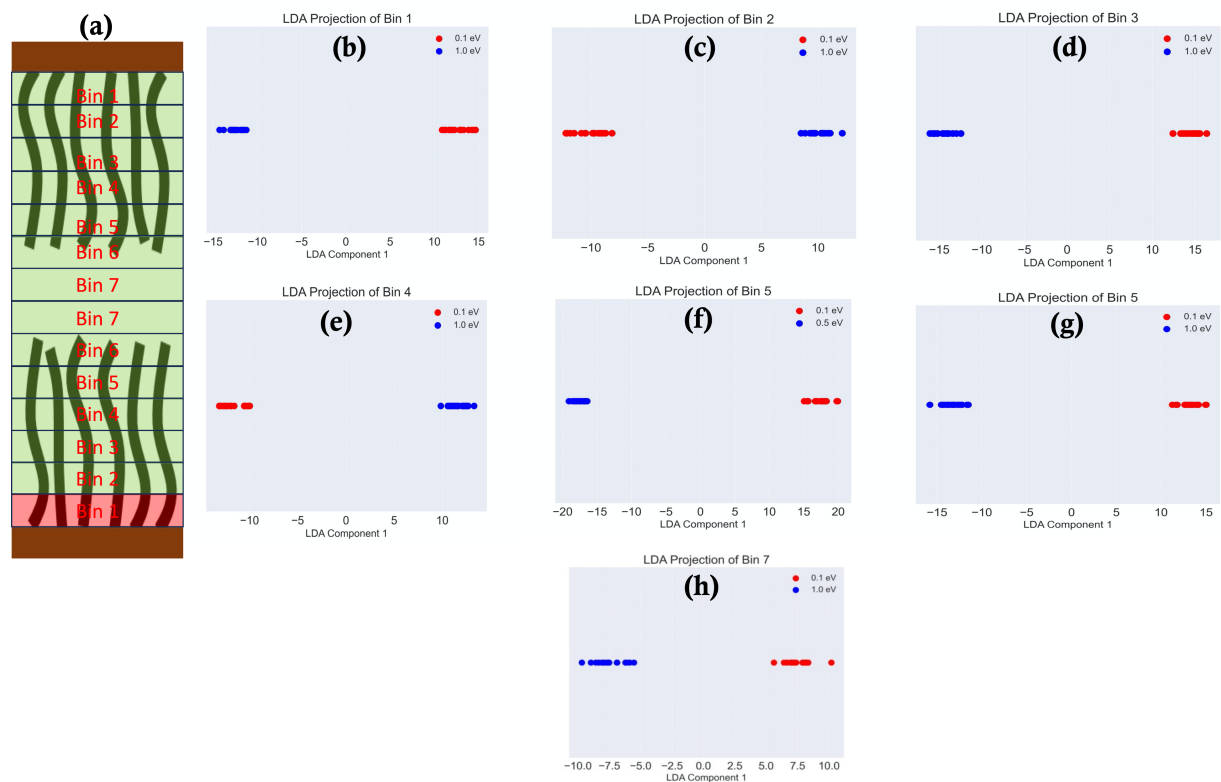
Figure 6. (a) Schematic showing the manner in which the brush grafted nanochannel is divided into two sets of 7 bins. One set of seven bins is from the nanochannel wall to the nanochannel centerline, with the bin number increasing as one goes from the wall to the centerline. LDA projection of the datapoints resulting in two separate clusters on a 1D line for situation 2 (reference state: case with E=0.1 V/nm and perturbed state: case with E=1 V/nm) for (b) Bin 1, (c) Bin 2, (d) Bin 3, (e) Bin 4, (f) Bin 5, (g) Bin 6, and (h) Bin 7.

3.3. Importance Score

Projecting the high-dimensional data into the clusters on a 1D line *via the application of the LDA* (see Figs. 5,6) is made possible by the calculation of the matrix w (a 1×6 matric). The absolute values of the components of w matric (namely, $|w_1|$, $|w_2|$, ….$|w_6|$) serve as the importance score associated with each feature in a given bin for a given situation. We use these importance score to also calculate the relative contributions of each feature in a given bin for a given situation. In Figs. 7 and 8 we plot these quantities for situations 1 and 2, respectively.
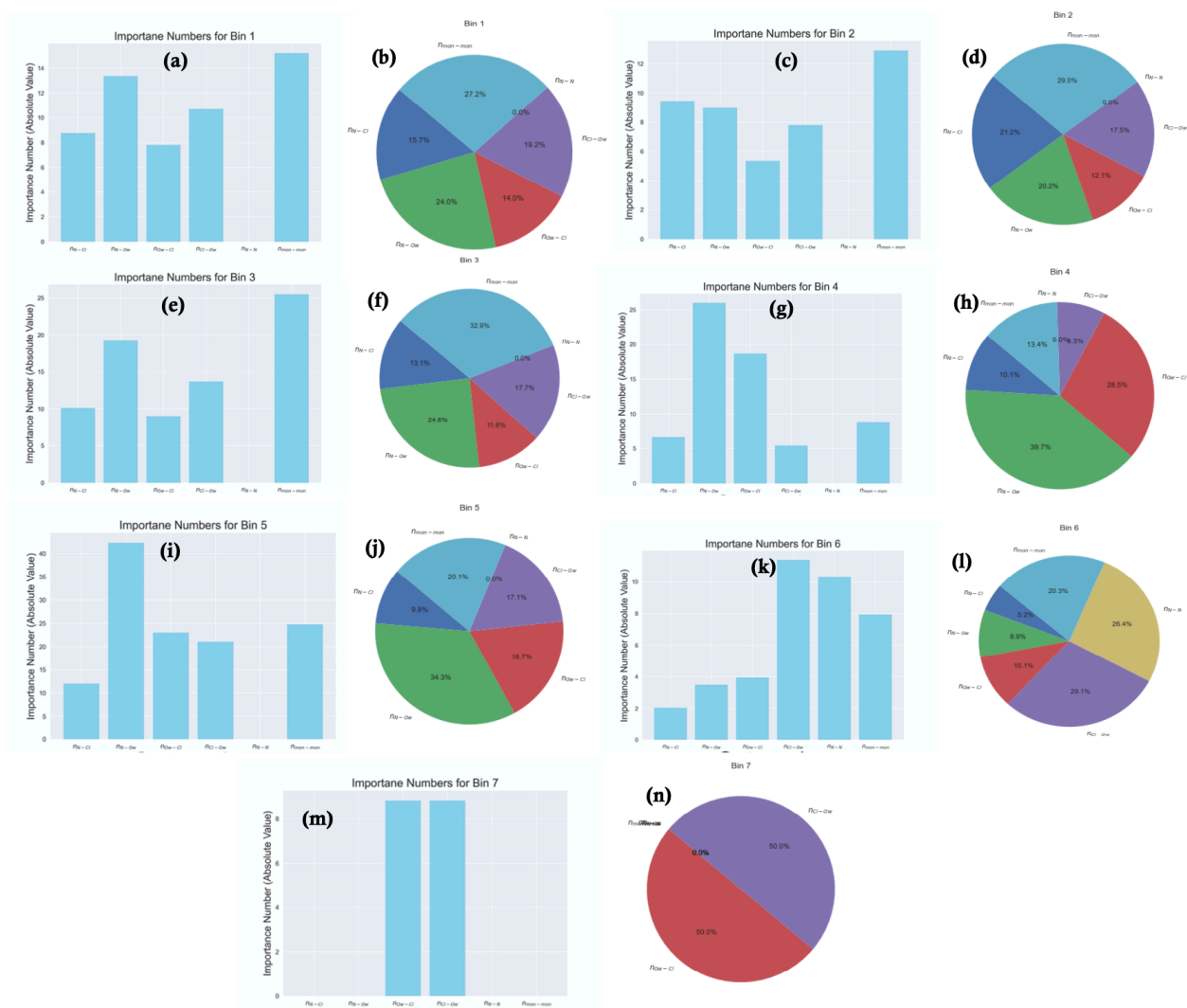
34

Figure 7. Importance score associated with different features for situation 1 (reference state: case with E=0.1 V/nm and perturbed state: case with E=0.5 V/nm) for (a) Bin 1, (c) Bin 2, (e) Bin 3, (g) Bin 4, (i) Bin 5, (k) Bin 6, and (m) Bin 7. Relative percentage contribution of the importance score associated with different features for situation 1 (reference state: case with E=0.1 V/nm and perturbed state: case with E=0.5 V/nm) (b) Bin 1, (d) Bin 2, (f) Bin 3, (h) Bin 4, (j) Bin 5, (l) Bin 6, and (n) Bin 7.
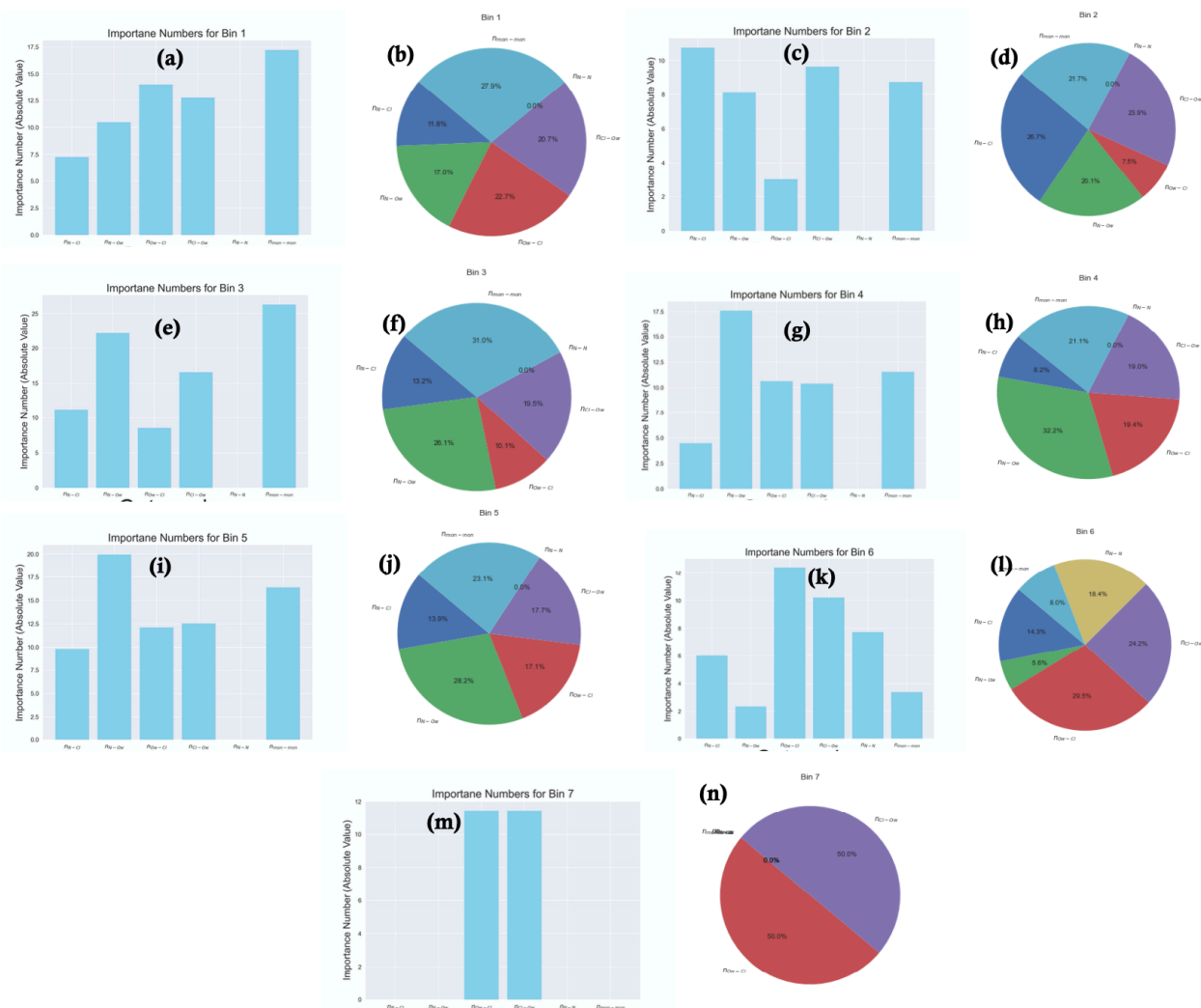
35

Figure 8. Importance score associated with different features for situation 2 (reference state: case with E=0.1 V/nm and perturbed state: case with E=1 V/nm) for (a) Bin 1, (c) Bin 2, (e) Bin 3, (g) Bin 4, (i) Bin 5, (k) Bin 6, and (m) Bin 7. Relative percentage contribution of the importance score associated with different features for situation 2 (reference state: case with E=0.1 V/nm and perturbed state: case with E=1 V/nm) (b) Bin 1, (d) Bin 2, (f) Bin 3, (h) Bin 4, (j) Bin 5, (l) Bin 6, and (n) Bin 7.

36

Finally, in Figure 9, we summarize the variation of the relative percentage of the importance score associated with the six features for each bin and for each situation. This figure has all the message that we need. It tells us (1) how a given feature varies across different bins for a given situation (or a given perturbation) and (2) how a given feature in a given bin varies with the perturbation (i.e., between the two different situations).

The most prominent and the abrupt change occurs for the feature $n_{N\text{-}N}$ [see Fig. 10(a)]. For both the situations, it is zero for all the bins except for the bin at the brush-bulk interface (bin 6). In fact, in this bin, it becomes significantly large ~20% for both the cases. This feature $n_{N\text{-}N}$ represents the presence of one $\{N(CH_3)_3\}^+$ group of the PMETA chain in close vicinity of another $\{N(CH_3)_3\}^+$ group. Given that $n_{N\text{-}N}=0$ for bins deep inside the brushes implies that inside the brush layer, such an occurrence (where one $\{N(CH_3)_3\}^+$ group of the PMETA chain comes in close vicinity of another $\{N(CH_3)_3\}^+$ group) is not possible. This implies that something with regards to the structure of the PMETAC brushes is changing at the brush-bulk interface (or bin 6) that is making $n_{N\text{-}N}$ suddenly change to a finite value (for both the situations) at the brush-bulk interface (bin 6). This points to the fact that in order to better understand the response of the PMETAC brush-grafted nanochannel system to the applied electric field (and the resulting non-linear EOS transport), we must investigate the structure of the PMETAC brushes at the brush-bulk interface.

The variation of the feature $n_{mon\text{-}mon}$ across different bins and for different situations is unremarkable [see Fig. 10(b)]. It is significantly high at near wall locations (bin 1), and progressively decreases as one moves away from the wall. This is an expected occurrence characterizing the decreases of the monomer density as one moves away from the grafting surface.

The features $n_{N\text{-}Cl}$ and $n_{Cl\text{-}Ow}$ also show interesting result at the bin corresponding to the brush-bulk interface [see Figs. 10(c,f)]. At such brush-bulk interface, where we expect some alteration in the

37

PMETAC brush structure, we find a significant percentage contribution of n_{N-Cl} for either of the situations. This implies that the chloride ion is still having a significantly large interaction with the part of the brushes that are in the brush-bulk interface. Therefore, in addition to probing the structure of the brushes in greater detail at the brush-bulk interface, one should also probe the distribution of the chloride ions (and their association with the brushes) at the brush-bulk interface. Finally, the significantly high n_{Cl-Ow} percentage contribution at the brush-bulk interface points to a close association of the water oxygen atoms with the chloride ions at the brush-bulk interface. Therefore, the water number density distribution at the brush-bulk interface should be the third quantity that should be probed.

In summary, this entire effort of treating the all-atom MD simulation data to LDA-based ML protocol boils down to answering the question *what should be probed and where should be probed?* The answer to this question for the specific problem of the electric field response of the PMETAC brush-grafted nanochannel and the resulting non-linear EOS flow is as follows: *we should probe the behavior of the brushes, chloride counterions, and counterion-solvating water molecules at the location of the brush-bulk interface.*
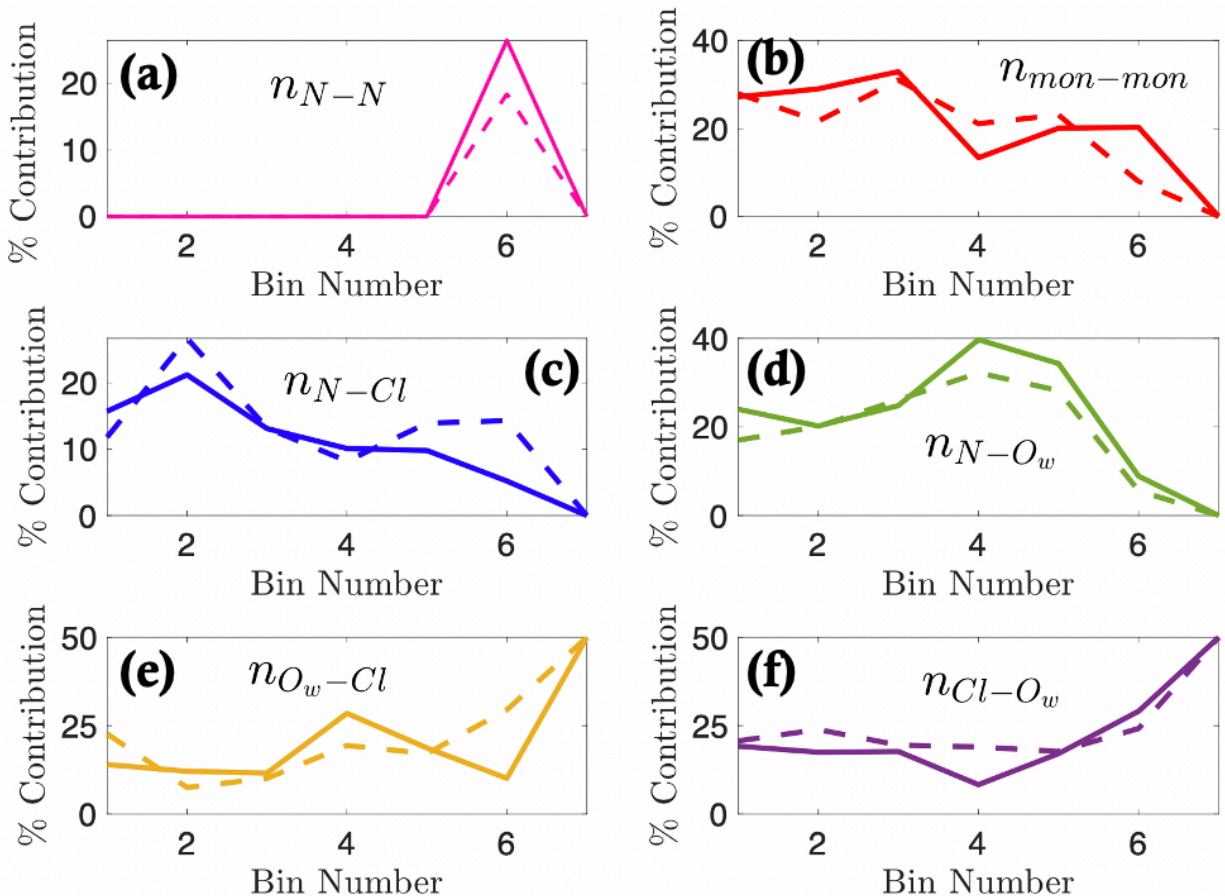
Figure 9. Variation of the percentage contribution of the importance score associated with the different features as a function of the bean number for (a) feature n_{N-N}; (b) feature $n_{mon-mon}$; (c) feature n_{N-Cl}; (d) feature n_{N-Ow}; (e) feature n_{Ow-Cl}; and (f) feature n_{Cl-Ow}. In each sub-figure, the results are shown by bold line for situation 1 (reference state: case with E=0.1 V/nm and perturbed state: case with E=0.5 V/nm) and by dashed line for situation 2 (reference state: case with E=0.1 V/nm and perturbed state: case with E=1 V/nm).

3.4 Utilization of the findings from the LDA based ML approach in explaining the mechanism of non-linear EOS transport in PMETAC brush grafted nanochannels

Fig. 10(a) shows the result confirming the non-linear EOS transport in PMETAC brush grafted nanochannels: one can see that the volume flow rate increases non-linearly with the electric field strength. Application of the LDA-based ML protocol to the all-atom MD data revealing such non-linear EOS transport directs us (see the previous subsection) to first study the PMETAC brush structure at the brush-bulk interface. This is done by calculating the monomer number distribution [see Fig. 10(b)]: most remarkably, there is an increase in the monomer number density at the brush-bulk interface with the increase in the electric field strength. A deeper analysis reveals that at larger electric fields, the brushes tilt/bend and occupy an unprecedented parallel (to the grafting substrate) configuration at the brush-bulk interface [see Figs. 10(c,d)].

The LDA-based ML protocol also directs us to study the chloride ion distribution at the brush-bulk interface. We do that and find that chloride ions, too, show a similar increase in the concentration with the electric field at the brush-bulk interface [see Fig. 10(e)]. This suggests that while the brushes are undergoing a significant deformation and getting parallel to the brush-bulk interface, the counterions *are not leaving the brushes*: this is also confirmed from the corresponding screening efficiency (SE) result (for all electric field values, the SE is ~100%, indicating that the chloride ions never leave the brush layer) [see Fig. 10(f)]. Such accumulated counterions at the location of the brush-bulk interface, in the presence of an applied electric field, experience a significantly large velocity [see Fig. 10(g)] as they are "much freer" and are only subjected to much weaker resistance from the monomers (at such brush-bulk interface).

Finally, the LDA-based ML protocol has also indicated to study the water molecules that are closely associated (hydrating) the chloride ions at the brush-bulk interface. Therefore, we study

40

the water velocity, and it closely resembles the velocity of the chloride counterions [see Fig. 10(h)]: the large velocity of the chloride counterions and their strong association with the water molecules ensure such large velocity of the water molecules in the presence of the applied electric field, justifying the non-linearly large EOS transport in PMETAC-brush-grafted nanochannels.

All these results on the EOS transport in the PMETAC brush grafted nanochannels were obtained without any guidance from the LDA-based ML protocol (see our paper in Ref. 59); hence it took us several months and a large number of analyses to converge on these results (and explanations for the non-linearly large EOS transport). On the other hand, with the guidance from the results of the LDA-based ML protocol, we are able to capture the appropriate physics governing the EOS transport in PMETAC brush grafted nanochannels much more quickly.
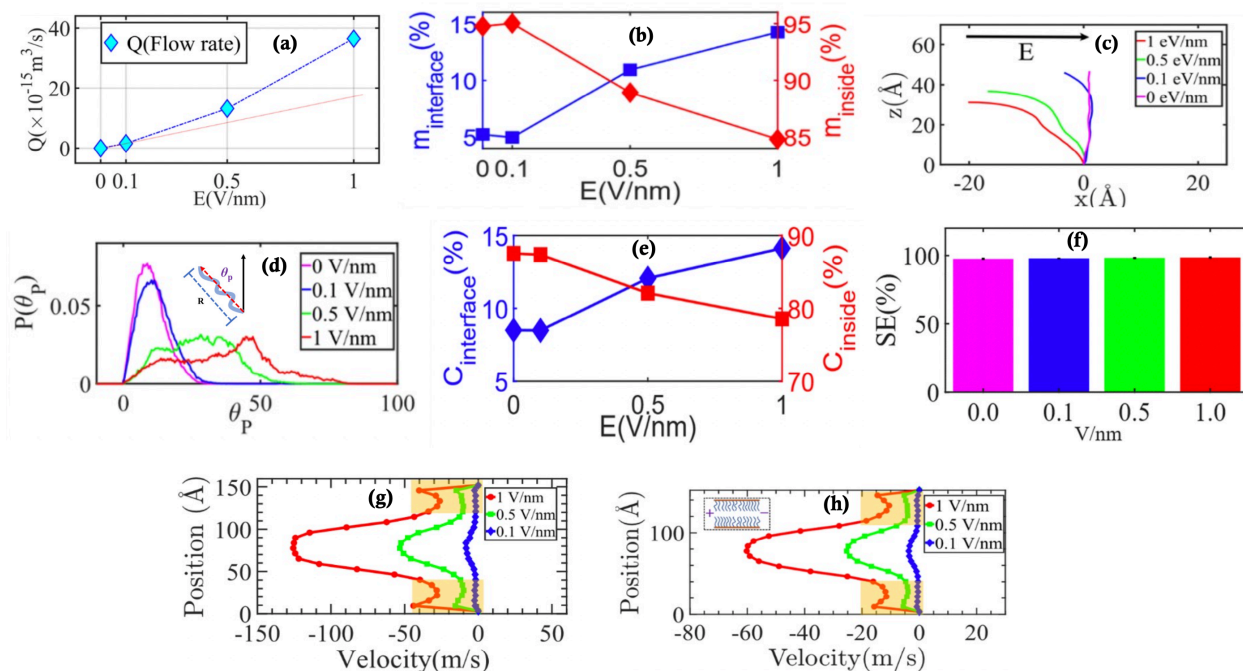
Figure 11. (a) Variation of the EOS volume flow rate as a function of the applied electric field strength in PMETAC-brush-grafted nanochannels. (b) Variation in the percentage of the PMETAC monomers in the "inside the brush" region (m_{inside}) and in the "brush-bulk interface" region ($m_{interface}$) with the strength of the applied electric field. (c) Chain profile of the PMETAC chains for different electric field strengths. (d) Probability distribution of the tilt angle (θ_p) of the brushes [the tilt angle is shown in the caption of (d)] for as a function of the applied electric field. (e) Variation in the percentage of counterions inside the brush" region (C_{inside}) and in the "brush-bulk interface" region ($C_{interfaces}$) with the electric field. (f) Screening Efficiency (SE) of the PMETAC PE brushes (by the counterions) for the cases of the different electric fields. (g,h) Variation of the velocity of the (g) counterions and (h) water molecules across the nanochannel. All the results have been taken from Ref. 59.

42

4. CONCLUSIONS

In this paper, we have proposed a new ML approach, based on Linear Discriminant Analysis, enabling the use of basic features (low domain knowledge data) from all-atom MD simulations to provide rapid dissemination of the mechanism dictating certain perplexing phenomena revealed by the MD simulations. Specifically, we consider the all-atom MD simulation-revealed highly non-linear EOS transport in the PMETAC-brush-grafted nanochannels. We show that the use of the LDA approach enables us to quickly identify what factors to study and in what locations of the nanochannel: such knowledge helps to identify the mechanism of such non-linear EOS transport very quickly. On the contrary, when such LDA-based information is unavailable, it takes a significantly larger effort and analyses to identify the mechanism of the non-linear EOS transport. This approach will be further utilized in future studies for quickly explaining the mechanisms associated with a variety of different electrokinetic and pressure-driven liquid flows in nanochannels grafted with different types of PE brushes and screened by a variety of different counterions.

References

1. Das, S., Banik, M., Chen, G., Sinha, S., Mukherjee, R., "Polyelectrolyte brushes: theory, modelling, synthesis and applications", *Soft Matter* **2015**, *11*, 8550-8583.

2. Milner, S. T., Witten, T. A., and Cates, M. E., "Theory of the grafted polymer brush", *Macromolecules*, **1988**, *21*, 2610–2619.

3. Misra, S., Varanasi, S., and Varanasi, P. P., "A polyelectrolyte brush theory", *Macromolecules*, **1989**, *22*, 4173–4179.

4. Zhulina, E. B., Boulakh, A. B., and Borisov, O. V., "Repulsive forces between spherical polyelectrolyte brushes in salt-free solution", *Z. Phys. Chem.*, **2012**, *226*, 625–643.

5. Zhulina, E. B., and Borisov, O. V., "Poisson–Boltzmann theory of pH-sensitive (annealing) polyelectrolyte brush", *Langmuir*, **2011**, *27*, 10615–10633.

6. Ishraaq R., Das, S. "All-atom molecular dynamics simulations of polymer and polyelectrolyte brushes". *Chem. Comm.* **2024**, *60*, 6093-6129.

7. Zhulina, E. B., and Borisov, O. V., "Structure and interaction of weakly charged polyelectrolyte brushes: Self-consistent field theory", *J. Chem. Phys.*, **1997**, *107*, 5952–5967.

8. Zhulina, E. B., Klein Wolterink, J., and Borisov, O. V., "Screening effects in a polyelectrolyte brush: Self-consistent-field theory", *Macromolecules*, **2000**, *33*, 4945–4953.

9. Chen, L., Merlitz, H., He, S. Z., Wu, C. X., and Sommer, J. U., "Polyelectrolyte brushes: Debye approximation and mean-field theory", *Macromolecules*, **2011**, *44*, 3109–3116.

10. Zhulina, E. B., Boulakh, A. B., and Borisov, O. V., "Repulsive forces between spherical polyelectrolyte brushes in salt-free solution", *Z. Phys. Chem.*, **2012**, *226*, 625–643.

44

11. Azzaroni, O., Brown, A. A., and Huck, W. T. S., "Tunable Wettability by Clicking Counterions Into Polyelectrolyte Brushes", *Adv. Mater.* **2007**, *19*, 151−154.

12. Han, L. *et al.*, "Tuning protein adsorption on charged polyelectrolyte brushes via salinity adjustment", *Colloid. Surf. A* **2018**, *539*, 37−45.

13. Vilozny, B., Wollenberg, A. L., Actis, P., Hwang, D., Singaram, B., and Pourmand, N., "Carbohydrate-actuated nanofluidic diode: Switchable current rectification in a nanopipette", *Nanoscale* **2013**, *5*, 9214−9221.

14. Ali, M., Yameen, B., Neumann, R., Ensinger, W., Knoll, W., and Azzaroni, O., "Biosensing and supramolecular bioconjugation in single conical polymer nanochannels. Facile incorporation of biorecognition elements into nanoconfined geometries", *J. Am. Chem. Soc.* **2008**, *130*, 16351−16357.

15. Umehara, S., Karhanek, M., Davis, R. W., and Pourmand, N., "Label-free biosensing with functionalized nanopipette probes", *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4611−4616.

16. Cai, S-L., Cao, S-H., Zheng, Y-B., Zhao, S., Yang, J-L., and Li, Y-Q., "Surface charge modulated aptasensor in a single glass conical nanopore", *Biosens. Bioelectron.* **2015**, *71*, 37−43.

17. Chanda, S., Sinha, S., and Das, S., "Streaming potential and electroviscous effects in soft nanochannels: towards designing more efficient nanofluidic electrochemomechanical energy converters", *Soft Matter*, **2014**, 10, 7558–7568.

18. Sachar, H. S., Sivasankar, V. S., and Das, S., "Electrokinetic energy conversion in nanochannels grafted with pH-responsive polyelectrolyte brushes modelled using augmented strong stretching theory", *Soft Matter*, **2019**, *15*, 5973–5986.

19. Patwary, J., Chen, G., and Das, S., "Efficient electrochemomechanical energy conversion in nanochannels grafted with polyelectrolyte layers with pH-dependent charge density",

Microfluid. Nanofluid., **2016**, *20*, 37.

20. Chen, G., and Das, S., "Streaming potential and electroviscous effects in soft nanochannels beyond Debye–Hückel linearization", *J. Colloid Interface Sci.*, **2015**, *445*, 357–363.

21. Yang, Q., Li, L., Zhao, F., Han, H., Wang, W., Tian, Y., Wang, Y., Ye, Z., and Guo, X. "Hollow silica-polyelectrolyte composite nanoparticles for controlled drug delivery". *J. Mater. Sci.* **2019**. 54, 2552–2565.

22. ShamsiJazeyi, H., Miller, C.A., Wong, M.S., Tour, J.M., and Verduzco, R. "Polymercoated nanoparticles for enhanced oil recovery." *J. Appl. Polym. Sci.* **2014** 131, 40576.

23. Liu, G., Cai, M., Wang, X., Zhou, F., and Liu, W. "Core-shell-corona-structured polyelectrolyte brushes-grafting magnetic nanoparticles for water harvesting." *ACS Appl. Mater. Interfaces*, **2014,** 6, 11625–11632.

24. Sachar, H. S., Sivasankar, V. S., and Das, S., "Revisiting the strong stretching theory for pH-responsive polyelectrolyte brushes: effects of consideration of excluded volume interactions and an expanded form of the mass action law", *Soft Matter*, **2019**, *15*, 559–574.

25. Sachar, H. S., Sivasankar, V. S., Etha, S. A., Chen, G., and Das, S., "Ionic current in nanochannels grafted with pH-responsive polyelectrolyte brushes modeled using augmented strong stretching theory", *Electrophoresis*, **2020**, *41*, 554-561.

26. Etha, S. A., Sivasankar, V. S., Sachar, H. S., and Das, S., "Strong stretching theory for pH-responsive polyelectrolyte brushes in large salt concentrations", *Phys. Chem. Chem. Phys.*, **2020**, *22*, 13536-13553.

27. Borisov, O. V., and Zhulina, E. B., "Conformations of polyelectrolyte molecular brushes: A mean-field theory", *J. Chem. Phys.*, **2018**, *149*, 184904.

28. Chen, L., Merlitz, H., He, S. Z., Wu, C. X., and Sommer, J. U., "Polyelectrolyte brushes: Debye approximation and mean-field theory", *Macromolecules*, **2011**, *44*, 3109–3116.

29. Csajka, F. S., and Seidel, C., "Strongly charged polyelectrolyte brushes: A molecular dynamics study", *Macromolecules*, **2000**, *33*, 2728–2739.

30. He, G. L., Merlitz, H., and Sommer, J. U., "Molecular dynamics aimulations of polyelectrolyte brushes under poor solvent conditions: Origins of bundle formation", *J. Chem. Phys.*, **2014**, *140*, 104911.

31. Sandberg, D. J., Carrillo, J. M. Y., and Dobrynin, A. V., "Molecular dynamics simulations of polyelectrolyte brushes: From single chains to bundles of chains", *Langmuir*, **2007**, *23*, 12716–12728.

32. Desai, P. R., Sinha, S., and Das, S., "Polyelectrolyte brush bilayers in weak interpenetration regime: Scaling theory and molecular dynamics simulations", *Phys. Rev. E*, **2018**, *97*, 032503.

33. Mei, Y., Hoffmann, M., Ballauff, M., and Jusufi, A., "Spherical polyelectrolyte brushes in the presence of multivalent counterions: The effect of fluctuations and correlations as determined by molecular dynamics simulations", *Phys. Rev. E*, **2008**, *77*, 031805.

34. Carrillo, J. M. Y., and Dobrynin, A. V., "Morphologies of planar polyelectrolyte brushes in a poor solvent: Molecular dynamics simulations and scaling analysis", *Langmuir*, **2009**, *25*, 13158–13168.

35. Hehmeyer, O. J., and Stevens, M. J., "Molecular dynamics simulations of grafted polyelectrolytes on two apposing walls", *J. Chem. Phys.*, **2005**, *122*, 134909.

36. Jackson, N. E., Brettmann, B. K., Vishwanath, V., Tirrell, M., and de Pablo, J. J., "Comparing solvophobic and multivalent induced collapse in polyelectrolyte

brushes", *ACS Macro Lett.*, **2017**, *6*, 155–160.

37. Merlitz, H., Li, C., Wu, C., and Sommer, J. U., "Polyelectrolyte brushes in external fields: Molecular dynamics simulations and mean-field theory", *Soft Matter*, **2015**, *11*, 5688–5696.

38. Kumar, N. A., and Seidel, C., "Polyelectrolyte brushes with added salt", *Macromolecules*, **2005**, *38*, 9341–9350.

39. Miliou, K., Gergidis, L. N., and Vlahos, C., "Molecular dynamics simulation of brushes formed by star polyelectrolytes under theta solvent conditions", *J. Polym. Sci., Part B: Pol. Phys.*, **2017**, *55*, 1110–1117.

40. Washizu, H., Kinjo, T., and Yoshida, H., "Structure of polyelectrolyte brushes studied by coarse grain simulations", *Friction*, **2014**, *2*, 73–81.

41. Zhang, F., Ding, H. D., Duan, C., Zhao, S. L., and Tong, C. H., "Molecular dynamics simulation of the response of bi-disperse polyelectrolyte brushes to external electric fields", *Chinese Phys. B*, **2017**, *26*, 088204.

42. Carrillo, J. M. Y., and Dobrynin, A. V., "Molecular dynamics simulations of grafted layers of bottle-brush polyelectrolytes", *Langmuir*, **2010**, *26*, 18374–18381.

43. Li, L., Cao, Q., and Zuo, C., "Effect of counterion valence on conformational behavior of spherical polyelectrolyte brushes confined between two parallel walls", *Polymers*, **2018**, *10*, 363.

44. He, G.L., Merlituz, H., and Sommer, J. U., "Molecular dynamics simulations of polyelectrolyte brushes under poor solvent conditions: origins of bundle formation", *J. Chem. Phys.,* **2014,** 140, 104911.

45. Sachar, H. S. Pial, T. H., Desai. P.R., Etha, S. A., Wang, Y., and Das. S., "Densely Grafted

Polyelectrolyte Brushes Trigger "Water-in-Salt"-like Scenarios and Ultraconfinement Effect", *Matter*, **2020**, *2*, 1509–1521.

46. Sachar, H. S., Pial, T. H., Chava, B. S., and Das, S., "All-atom molecular dynamics simulations of weak polyionic brushes: influence of charge density on the properties of polyelectrolyte chains, brush-supported counterions, and water molecules", *Soft Matter*, **2020**, *16*, 7808–7822

47. Sachar, H. S., Chava, B. S., Pial, T. H., and Das, S., "Hydrogen Bonding and Its Effect on the Orientational Dynamics of Water Molecules inside Polyelectrolyte Brush-Induced Soft and Active Nanoconfinement", *Macromolecules*, **2021**, *54*, 2011–2021.

48. Sachar, H. S., Pial, T. H., Chava, B. S., and Das, S., "All-atom Molecular Dynamics simulations of the temperature response of densely grafted polyelectrolyte brushes", *Macromolecules*, **2021**, *54*, 6342–6354.

49. Pial, T. H., Sachar, H. S., and Das, S., "Quantification of Mono- and Multivalent Counterion-Mediated Bridging in Polyelectrolyte Brushes", *Macromolecules*, **2021**, *54*, 4154–4163.

50. Pial, T. H., Prajapati, M.; Chava, B.; Sachar, H. S., and Das, S., "Charge-density-specific response of grafted polyelectrolytes to electric fields: Bending or tilting?" *Macromolecules*, **2022**, *55*, 2413–2423.

51. Pial, T. H., Sachar, H. S., Desai, P. R., and Das, S., "Overscreening, Co-Ion-Dominated Electroosmosis, and Electric Field Strength Mediated Flow Reversal in Polyelectrolyte Brush Functionalized Nanochannels", *ACS Nano*, **2021**, *15*, 6507–6516.

52. Sachar, H. S., Pial, T. H., Sivasankar, V. S.; and Das, S., "Simultaneous energy generation and flow enhancement (Electroslippage Effect) in polyelectrolyte brush functionalized

nanochannels", *ACS Nano*, **2021**, *15*, 17337–17347.

53. Pial, T. H.; Das, S. "Specific ion and electric field controlled diverse ion distribution and electroosmotic transport in a polyelectrolyte brush grafted nanochannel", *J. Phys. Chem. B*, **2022**, *49*, 10543–10553.

54. Pial, T. H.; Das, S. "Machine learning enabled quantification of the hydrogen bonds inside the polyelectrolyte brush layer probed using all-atom molecular dynamics simulations", *Soft Matter*, **2022**, *18*, 8945–8951.

55. Ishraaq, R.; Akash, T. S.; Pial, T. H.; and Das, S. "Hydrophilic and apolar hydration in densely grafted cationic brushes and counterions with large mobilities", *J. Phys. Chem. B*, **2024**, *128*, 381–392.

56. Ishraaq, R.; Das, S. "All-atom molecular dynamics simulations of cationic polyelectrolyte brushes in the presence of halide counterions", *Macromolecules* **2024**, *57*, 3037–3046.

57. Ishraaq, R.; Akash, T. S.; Das, S. "Combined Machine Learning and Molecular Dynamics reveal two states of hydration of a single functional group of cationic polymeric brushes", *Macromolecules* **2024**, *57*, 5300–5312.

58. Bera, A.; Akash, T. S.; Ishraaq, R.; Pial, T. H.; Das, S. "Hydrogen bonding inside anionic polymeric brush layer: Machine Learning-driven exploration of the relative roles of the polymer steric effect, charging, and type of screening counterions", *Macromolecules* **2024**, *57*, 1581–1592.

59. Ishraaq, R.; Das, S. "Anti-electrophoretic response driven bending-tilting deformation of cationic polyelectrolyte brushes drives massively non-linear electroosmotic transport in brush-grafted nanochannels" DOI: 10.26434/chemrxiv-2024-p9sw8

60. Wittemann, A., Drechsler, M., Talmon, Y., and Ballauff, M., "High elongation of

polyelectrolyte chains in the osmotic limit of spherical polyelectrolyte brushes: A study by cryogenic transmission electron microscopy", *J. Am. Chem. Soc.*, **2005**, *127*, 9688–9689.

61. Biesalski, M., Johannsmann, D., and Rühe, J., "Electrolyte-induced collapse of a polyelectrolyte brush", *J. Chem. Phys.*, **2004**, *120*, 8807–8814.

62. Balastre, M., Li, F., Schorr, P., Yang, J., Mays, J. W., and Tirrell, M. V., "A study of polyelectrolyte brushes formed from adsorption of amphiphilic diblock copolymers using the surface forces apparatus", *Macromolecules*, **2002**, *35*, 9480–9486.

63. Yu, J., Jackson, N. E., Xu, X., Morgenstern, Y., Kaufman, Y., Ruths, M., de Pablo, J. J., and Tirrell, M., "Multivalent counterions diminish the lubricity of polyelectrolyte brushes", *Science*, **2018**, *360*, 1434–1438.

64. Yu, J., Jackson, N. E., Xu, X., Brettmann, B. K., Ruths, M., de Pablo, J. J., and Tirrell, M., "Multivalent ions induce lateral structural inhomogeneities in polyelectrolyte brushes", *Sci. Adv.*, **2017**, *3*, eaao1497.

65. Yu, J., Mao, J., Yuan, G., Satija, S., Chen, W., and Tirrell, M., "The effect of multivalent counterions to the structure of highly dense polystyrene sulfonate brushes", *Polymer*, **2016**, *98*, 448–453.

66. Yu, J., Mao, J., Yuan, G., Satija, S., Jiang, Z., Chen, W., and Tirrell, M., "Structure of polyelectrolyte brushes in the presence of multivalent counterions", *Macromolecules*, **2016**, *49*, 5609–5617.

67. Farina, R., Laugel, N., Pincus, P., and Tirrell, M., "Brushes of strong polyelectrolytes in mixed mono-and tri-valent ionic media at fixed total ionic strengths", *Soft Matter*, **2013**, *9*, 10458–10472.

68. Schneider, C., Jusufi, A., Farina, R., Pincus, P., Tirrell, M., and Ballauff, M., "Stability

behavior of anionic spherical polyelectrolyte brushes in the presence of La (III) counterions", *Phys. Rev. E*, **2010**, *82*, 011401.

69. Brettmann, B., Pincus, P., and Tirrell, M., "Lateral Structure Formation in Polyelectrolyte Brushes Induced by Multivalent Ions", *Macromolecules*, **2017**, 50, 1225−1235

70. Xu. X., Mastropietro. D., Ruths, M., Tirrell, M., and Yu. J., "Ion-Specific Effects of Divalent Ions on the Structure of Polyelectrolyte Brushes", *Langmuir*, 2019, 35, 15564−15572

71. Mahalik, J. P., Yang, Y., Deodhar, C., Ankner, J. F., Lokitz, B. S., Kilbey, S. M., Sumpter, B. G., and Kumar, R., "Monomer volume fraction profiles in pH responsive planar polyelectrolyte brushes", *J. Polym. Sci., Part B: Polym. Phys.*, **2016**, *54*, 956–964.

72. Willott, J. D., Humphreys, B. A., Webber, G. B., Wanless, E. J., and De Vos, W. M., "Combined experimental and theoretical study of weak polyelectrolyte brushes in salt mixtures", *Langmuir*, **2019**, *35*, 2709–2718.

73. Azzaroni, O., Moya, S., Farhan, T., Brown, A. A., and Huck, W. T., "Switching the properties of polyelectrolyte brushes via "hydrophobic collapse"". *Macromolecules*, 2005, *38*, 10192–10199.

74. Guo, X., and Ballauff, M., "Spherical polyelectrolyte brushes: Comparison between annealed and quenched brushes", *Phys. Rev. E*, **2001**, *64*, 051406.

75. Ahrens, H., Förster, S., Helm, C. A., Kumar, N. A., Naji, A., Netz, R. R., and Seidel, C., "Nonlinear osmotic brush regime: Experiments, simulations and scaling theory", *J. Phys. Chem. B*, **2004**, *108*, 16870–16876.

76. Wittemann, A., and Ballauff, M., "Secondary structure analysis of proteins embedded in spherical polyelectrolyte brushes by FT-IR spectroscopy", *Anal. Chem.*, **2004**, *76*, 2813–

2819.

77. Hur, J., Witte, K. N., Sun, W., and Won, Y. Y., "On the origins of the salt-concentration-dependent instability and lateral nanoscale heterogeneities of weak polyelectrolyte brushes: Gradient brush experiment and Flory-type theoretical analysis", *Langmuir*, **2009**, *26*, 2021–2034.

78. Tran, Y., Auroy, P., and Lee, L. T., "Determination of the structure of polyelectrolyte brushes", *Macromolecules*, **1999**, *32*, 8952–8964.

79. Kaewsaiha, P., Matsumoto, K., and Matsuoka, H., "Synthesis and nanostructure of strong polyelectrolyte brushes in amphiphilic diblock copolymer monolayers on a water surface", *Langmuir*, **2004**, *20*, 6754–6761.

80. Stephan, T., Muth, S., and Schmidt, M., "Shape changes of statistical copolymacromonomers: From wormlike cylinders to horseshoe-and meanderlike structures", *Macromolecules*, **2002**, *35*, 9857–9860.

81. Förster, S., Hermsdorf, N., Böttcher, C., and Lindner, P., "Structure of polyelectrolyte block copolymer micelles", *Macromolecules*, **2002**, *35*, 4096–4105.

82. Duboué-Dijon, E., Laage, D., "Characterization of the Local Structure in Liquid Water by Various Order Parameters", *J. Phys. Chem. B* **2015**, *119*, 8227-8474.

83. Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J. "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

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