MULTISCALE COMPUTATIONAL MODELING OF NANOSTRUCTURE AND TRANSPORT IN POLYMER ELECTROLYTE MEMBRANE FUEL CELLS

A Doctoral Dissertation

Presented to

The Academic Faculty

by

Robin M. Lawler

In Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy in the

School of Chemical and Biomolecular Engineering

College of Engineering

Georgia Institute of Technology

December 2021

THESIS APPROVAL PAGE

Seung Soon Jang (MSE, Chair, Thesis Reading Committee) – approved 7/19/2021

Paul Kohl (ChBE, Member, Thesis Reading Committee) – approved 7/19/2021

David Sholl (ChBE, Member, Thesis Reading Committee) – approved 7/19/2021

Christopher Jones (ChBE, Member, Thesis Reading Committee) – approved 7/19/2021

Thomas Fuller (ChBE, Member, Thesis Reading Committee) – approved 7/19/2021

Martha Grover (ChBE, School Chair/Graduate Coordinator) – approved 7/19/2021

Tatianna Richardson (OGS Staff) – approved 7/19/2021

To Helen, to my parents, to my Uncle John, and to my Grandma May

ACKNOWLEDGEMENTS

To my labmates who guided me through countless scientific challenges: Connor for helping me learn the ropes; Jinwon, Omar, and Junhe for our project collaborations; and especially Dr. Choi, who in addition to helping me solve problems in my research would regularly drop everything on a Friday or Saturday night to help me with a server issue.

To my advisor, Prof. Jang, for endless guidance and support of my scientific excursions. You were always available for a meeting, had invaluable insight when I was stuck on a problem, and encouraged me to work hard and push my boundaries. Thank you for renewing my passion for research and helping me to be so productive.

To Prof. Jones, Prof. Fuller, Prof. Sholl, and Prof. Kohl, for their support as my committee members. Thank you for your scientific insights and opportunities for collaboration. I am fortunate to have had your guidance.

To my collaborators, Prof. William Goddard, Prof. Seung Geol Lee, Prof. Seungjin Lee, Prof. Jin Young Kim, and Prof. Yu Huang. Thank you for your time and insightful input. It was an honor to work with you.

To my undergraduate researchers, David Liu, Inhee Baek, Jinhee Im, Kyungil Kim, Jihoon Kim, Bless Suh, Sahba Khorsandzadeh, Nessa Majaya, Laura Wu, Sami Al-Aidy, and Tyler Harrison, who pulled countless all-nighters to meet paper submission deadlines, worked diligently on research at all hours, and provided invaluable insight to our projects. Without your work, this thesis would not have been possible.

To my parents, for encouraging me to dream, for their effort in homeschooling me and for supporting my interest in science and engineering from an early age. Certainly, this thesis would not have been possible without your support.

To my Uncle John, for fostering my love of knowledge. Our study nights with Starbucks and Mike Oldfield are some of my favorite memories. Thank you for instilling in me the passion for studying and learning that carried me through college and my doctorate.

To Helen, for her endless emotional support, for making sure I always had a healthy meal, for helping with the house and dogs when my doctorate schedule didn't allow, for encouraging me to take risks, and for helping me to become the better person I am today. Without a doubt, none of this would have been possible without her.

Finally, I'd like to thank my Grandma May, for everything.

OUTLINE

LIST OF ABBREVIATIONS

PEMFC – polymer electrolyte membrane fuel cell

DFT – density functional theory

MD – molecular dynamics

MSD – mean squared displacement

KRR – kernel ridge regression

GPR – Gaussian process regression

ANN – artificial neural network

PFSA – perfluorosulfonic acid

DOS – density of states

VFE – vacancy formation energy

RCA – relative contribution analysis

RFE – recursive feature elimination

MAE – mean absolute error

CB – carbon black

ML – machine learning

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

Polymer electrolyte membrane fuel cells (PEMFCs) are predicted to revolutionize energy conversion for transportation due to a multitude of advantages over conventional methods. Their use of a renewable resource (hydrogen) and clean waste (water) makes them more likely to conform to increasingly stringent environmental protection laws compared to a traditional gasoline engine. Their high power density and lack of necessity for charging imparts a clear advantage over batteries.[1] Clearly, fuel cells show budding promise in the world of portable energies.

In a traditional PEMFC, hydrogen is initially injected into the fuel cell from an external tank, passing through a gas diffusion layer into the anode. Within the anode, this hydrogen is split over a platinum catalyst and broken into its component protons and electrons. The electrons pass through a circuit to provide electricity for the application, whereas protons are passed through a proton exchange membrane. Both electrons and protons reunite at the cathode. Meanwhile, air is injected into the opposite side of the cell, diffusing through another gas diffusion layer into the cathode. There, the electrons, protons, and O_2 from the input air are recombined, producing water as a waste product.

Indeed, PEMFCs are impressive feats of engineering with a promising multitude of advantages. Surprisingly, however, they are not as widespread as other portable energy technologies. This can be attributed to their lack of versatility in adverse conditions. Under optimal (e.g. >40% humidity, under 3000 hours of use) conditions, fuel cells perform superbly. However, in practice, these conditions cannot be maintained. In dry conditions or high temperatures, the polymer electrolyte membrane (PEM) dehydrates, compromising proton mobility and causing power loss.[2-3] Additionally, radical degradation could render the PEM unusable after a mere 5 years of automotive use.[4] It is also worth mentioning that to equip consumer cars with PEMFCs would require nearly four times the total estimated Pt reserves in the world.[5-6] In order to render PEMFCs suitable for extensive use, we must explore methods to enhance their performance, by improving conductivity in extreme conditions and lengthening their lifetime.

Previously, researchers have attempted to alleviate these issues using a variety of methods. Manipulation of the PEM's chemical properties (e.g. pK_a, equivalent weight) has been shown to enhance proton transport, but performance at low hydration remains insufficient.[7-8] To prevent degradation, researchers have incorporated cerium ion additives due to their versatility as a radical scavenger, and more recently ceria nanoparticles, but the efficiency of these scavengers could still be optimized.[9-11] Finally, low platinum loading in PEMFC electrodes offers an attractive low-cost alternative to traditional fuel cells, but these variants show diminished catalytic activity.[12-13] In order to perfect these promising technologies, we must guide development by 1) deepening our fundamental understanding of these systems, and 2) improving upon their chemistry.

Available experimental techniques limit fundamental understanding at scales relevant to these systems. For example, while small angle scattering experiments (SAXS and SANS) may offer some insight into the large-scale nanophase segregation of the PEM membrane, they provide little information regarding local structure surrounding functional groups.[14] Likewise, while microscopy techniques such as STM allows scientists to see defect structure on the ceria surface, they provide limited insight into finer geometric details and radical binding mechanisms.[15]

Multiscale computational modeling offers numerous tools including density functional theory (DFT) and molecular dynamics (MD) simulations which allow insight into the detailed structure and mechanisms of PEMFC components. DFT offers electronic-scale insight into system charges, geometries, and potential energies. These properties can be useful for elucidating properties such as atomic charges and binding energies. MD simulations allow researchers to observe molecular structure and dynamics on an atomic scale, which can provide insight into nanophase segregation and molecular diffusion in a system. To improve the accuracy of these models, results from all scales of computational modeling can be used as input for machine learning models like kernel ridge regression (KRR), Gaussian process regression (GPR), or artificial neural networks (ANN), thereby either improving their accuracy or broadening their scope.

As such, this thesis aims to address the issue of PEMFC durability by using multiscale computational simulations to provide fundamental understanding of PEM mechanisms and suggest superior chemistries for PEM components. Specifically, **Aim 1** aids in the design of PEMs resistant to hot or dry conditions by offering novel insight into how PEM nanostructure influences proton transport in low-humidity conditions. **Aim 2** involves the elucidation of the CeO_2 radical scavenging mechanisms in PEMs, as well as the suggestion of an improved CeO_2 surface chemistry. Finally, **Aim 3** expands upon our first aim by offering an algorithm which accurately predicts pK_a (and, consequently, approximates performance) of acids relevant to PEMs, streamlining the design of novel, durable chemistries.

II. THEORETICAL AND COMPUTATIONAL BACKGROUND

2.1. DENSITY FUNCTIONAL THEORY

Electronic structure theory enables the highly accurate calculation of electronic structure in systems with O(10^2) atoms. From this, researchers can extract properties such as optimized geometry, potential energy, charge and spin. In this research, we employ DFT, which is widely used due to its computational efficiency and high accuracy.[16-18]

In order to obtain the electronic structure, it is necessary to solve the time-independent Schrödinger equation:

$$\hat{H}\psi(x) = E\psi(x)$$

Wherein $\psi(x)$ represents the wave function such that $[\psi(x)]^2$ is the probability of finding an electron at point x, E is the total potential energy, and $\hat{H}$ is the Hamiltonian total energy operator:

$$\hat{H} = \left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]$$

Such that $\hbar$ is Planck's constant, m is the mass of an electron, and $V(x)$ is the potential energy operator.

Since an exact solution to the Schrödinger equation is unfeasible for systems larger than hydrogen, approximation techniques have been developed. Among these, DFT offers the best balance between computational speed and accuracy. DFT requires the solution of the Kohn-Sham equations as follows:

$$\left(-\frac{1}{2}\nabla^2 + V_{\text{eff}}\right)\varphi_i = \varepsilon_i\varphi_i$$

Where φ_i is a Kohn-Sham orbital (representing the system in terms of electron density rather than a wavefunction), ε_i is the orbital energy, and the effective potential is represented by:

$$V_{\text{eff}} = \int \frac{\rho(\mathbf{r}_2)}{\mathbf{r}_{12}} d\mathbf{r}_2 + \frac{\delta E_{XC}[\rho]}{\delta\rho} - \sum_A \frac{Z_A}{\mathbf{r}_{1A}}$$

Where $\sum_A \frac{Z_A}{\mathbf{r}_{1A}}$ represents the external potential (e.g. electron-nuclei interaction), $\int \frac{\rho(\mathbf{r}_2)}{\mathbf{r}_{12}} d\mathbf{r}_2$ represents the Coulomb energy, and E_{xc} is the exchange correlation functional. There are numerous methods to describe this exchange correlation energy, two of which used in this study include GGA/PBE and B3LYP.[19-25]

2.2. MOLECULAR DYNAMICS SIMULATIONS

MD simulations allow the extraction of structural and dynamic data for systems with $O(10^3)$ atoms. In any system, we can describe a potential energy surface (PES) based on the spatial relationships between atoms in this system. In a simple system (e.g. H_2 or H_2O), the PES has low dimensionality and can be easily calculated. In a system with thousands of atoms, the PES has high dimensionality and must be produced using a forcefield. A forcefield a collection of parameters and equations to describe interactions between atoms in a system. In our studies, we use a modified DREIDING forcefield[26] which incorporates both and non-bonded energy contributions as follows:

$$E_{total} = E_{vdW} + E_Q + E_{bond} + E_{angle} + E_{torsion} + E_{inversion}$$

Where E_{tot}, E_{vdW}, E_Q, E_{bond}, E_{angle}, $E_{torsion}$ and $E_{inversion}$ are the total, van der Waals, electrostatic, bond stretching, angle bending, torsion, and inversion energies, respectively. E_Q is calculated from atomic charges which are obtained from Mulliken population analysis. To prepare the system for submission, atoms are combined into a simulation cell with dimensions O(10^{-9}) meters. However, this small system size could potentially introduce boundary effects which impart unrealistic structure and behavior. As such, periodic boundary conditions are introduced. With periodic boundary conditions, if a particle passes through a boundary, it reappears on the opposite boundary while maintaining its previous velocity.

To run an MD simulation, it is necessary to compute the velocities and positions of the atoms at each timestep. In our applications, these parameters are computed using a velocity-Verlet algorithm as follows:[27]

$$\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2$$

$$\vec{v}(t+\Delta t) = \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t+\Delta t)}{2}\Delta t$$

Additionally, it is necessary to regulate the velocities within the system in order to maintain constant temperature. In our research, a Nose-Hoover thermostat is applied to maintain constant temperature and pressure.[28-30]

2.3. MACHINE LEARNING

The three types of machine learning models used in this study are kernel ridge regression (KRR), Gaussian process regression (GPR), and artificial neural network (ANN). KRR is a simple ML algorithm which utilizes a kernel trick with ridge regression.[31-32] Ridge regression involves the use of a complexity term in the objective function which penalizes large parameters and, thereby, prevents overfitting. Additionally, the kernel trick can accommodate nonlinear relationships among the data points by mapping them to a higher dimensional space with minimal computation. This technique, termed regularization, is especially known to be effective with small datasets. The target property is estimated as follows:

$$y = \sum_{i}^{N} \alpha_i k(X_i, x)$$

$$\alpha = (K + \lambda \mathrm{I})\mathrm{Y}(\mathrm{X})$$

$$\alpha = \underset{\alpha \in R^n}{\arg\min} \frac{1}{2} \left|\left|y - K\alpha\right|\right|_2^2 + \frac{\lambda}{2} \alpha^T K \alpha$$

where k is some positive definite kernel function, K is the matrix of kernel functions over all data, X and Y represent training data, x and y represent a new case, and λ is the regularization parameter. Kernels used in this work include linear, polynomial, radial basis, sigmoid, and Laplacian functions.[33-34] Please note that in this study, the linear kernel has a form of $k(x, x') = x \cdot x'$ and the polynomial has a form of $k(x, x') = (x \cdot x' + r)^d$ where d = 3 is the kernel degree and r = 1. The radial basis function (RBF) is a form of $k(x, x') = \exp\left(-\left|\left|x - x'\right|\right|_2^2\right)$ and the sigmoid function has a form of $k(x, x') = \tanh(x \cdot x' + r)$ where r = 1. Lastly, the Laplacian has a form of $k(x, x') = \exp\left(-\left|\left|x - x'\right|\right|_1\right)$.

On the other hand, GPR is similar to KRR in its use of a kernel trick, but differs in its methodology.[32, 35] Specifically, GPR uses Bayes' theorem to estimate the posterior mean function y given test point x and the training data X from a Gaussian prior mean function f_{prior} with covariance $k(X_i, x)$:

$$y = \sum_{i}^{N} w_i k(X_i, x) + f_{prior}$$

where *w* is obtained for a given *x* by solving the following equation:

$$\sum_{i}^{N} w_i k(X_i, x) = f - f_{prior}$$

where f is constructed from the training set such that f_m is the output corresponding to a set of descriptors X_m. In GPR, the kernel matrix $k(X_i, x)$ is modified with a value σ added to its diagonal to prevent numerical issues. Covariance functions screened in this study for GPR included RBF, Matérn, and rational quadratic.[32, 36-38] RBF in GPR has the following form:

$$k(X_i, x) = \exp\left(-\frac{d(X_i, x')^2}{2l^2}\right)$$

where $\mathrm{d}(X_i, x')$ is the Euclidean distance and *l* is the length-scale parameter, which is chosen during hyperparameter optimization. Matérn is as follows:

$$k(x, x') = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}d(x, x')\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}}{l}d(x, x')\right)$$

where $K_{\nu}(\,)$ is a modified Bessel function, $\Gamma(\,)$ is the gamma function, and $\nu = 1.5$ controls the smoothness of the resulting function. Rational quadratic has he following form:

$$k(x, x') = \left(1 + \frac{d(x, x')^2}{2\alpha l^2}\right)^{-\alpha}$$

where α is chosen during hyperparameter optimization.

ANNs are commonly used in materials science, in part due to their flexibility in capturing complex nonlinear correlations within data.[39-42] A neural network consists of an input layer (for descriptors), an output layer, and several hidden layers. Each node is connected to all nodes in the previous and subsequent layer by neurons. The value of the *k*th neuron in the *i*th hidden layer is defined as:

$$y_k^i = f_i\left(b_k^i + \sum_{j=1}^{N_{i-1}} w_{jk}^i y_j^{i-1}\right)$$

where f_i is the activation function (in this case, ReLu, hyperbolic tangent, or logistic sigmoid), N_{i-1} is the number of neurons, b_k^i is biases, and w_{jk}^i is weight between neuron j in layer i – 1 and neuron k in layer i.[43]

AIM 1. To investigate how PEM nanophase segregation can be manipulated to enhance proton transport in low-humidity conditions.

Introduction

As mentioned above, while boasting superior properties at standard conditions (e.g. 80 ºC and 100 % R.H.), PFSA membranes dehydrate at high temperatures (> 120 ºC) and low humidity (< 40 % R.H.), compromising their performance.[2, 44] In previous studies, researchers have attempted to alleviate this problem through use of inorganic additives to promote rehydration, or using non-aqueous solvents.[45-46] Although these alternatives yield respectable conductivities under low-hydration conditions compared to bare PFSA, their performance is still insufficient for various applications.[47] Alternatively, the functionalization of small molecules with a phosphate group, rather than a sulfonate group, has been reported to improve proton diffusion and conductivity under low-humidity conditions.[48-50] Previous literature has attributed this behavior to a reduced energy barrier for acid-to-acid proton transfer, which would facilitate non-aqueous charge transfer modes such as self-ionization coupled with proton hopping.[48-49, 51-52] Nevertheless, compared to sulfonate-functionalized molecules, their conductivity under standard humidity is substandard. To render these molecules suitable for widespread use, we must explore methods to enhance their performance.

One approach involves the manipulation of the nanophase segregation of the membrane. Nanophase segregation involves a nanoscopic separation of hydrophobic and hydrophilic phases in the membrane, establishing a percolated aqueous phase through which protons migrate via hopping and vehicular mechanisms. Various previous experimental and computational studies have demonstrated the importance of a well-established aqueous phase to PEMFC performance.[53-57] Modifications to PFSA chemistry have been extensively studied with the goal of improving nanophase segregation between hydrophobic and hydrophilic phases.[58-59] Researchers have controlled segregation through a variety of parameters, from membrane polymer chemistry (e.g., equivalent weight, backbone chemistry, molecular weight) to water content.[60-61] While there are many chemical modifications shown to affect nanophase segregation, side chain length is a well-studied option which involves minimal change to existing chemistry.

Previous studies have obtained mixed results regarding the effect of side chain length (given negligible change in equivalent weight) on nanophase segregation, diffusion, and conductivity.[7, 62-68] The prevailing hypothesis for PFSA membranes states that longer side chains, while enlarging water clusters (which enhances diffusion and conductivity), also interfere with their connectivity (which decreases diffusion and conductivity), ultimately posing an optimization problem.[60-62, 69-72] Despite these advances toward understanding this relationship in PFSA membranes, that of perfluorinated phosphoric acid (PFPA) membranes is far less explored. A comprehensive study is required to examine the complex relationships of side chain length with various variables such as functional group acidity, nanophase segregation, and water content in order to guide design efforts toward a durable PEM resistant to various environmental conditions.

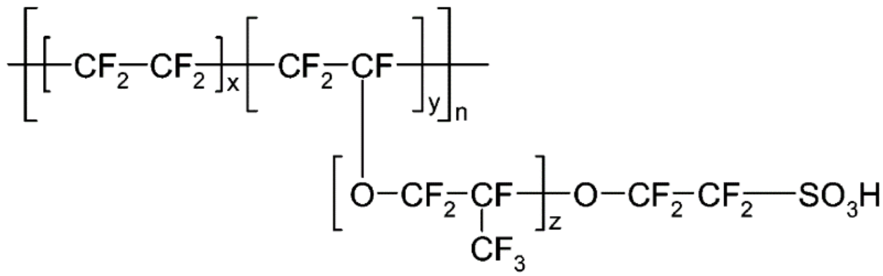
In this study, we investigate how membrane properties such as nanophase segregation, hydronium diffusion, and conductivity vary as a function of side chain length and acid strength PFSA and PFPA. We do this by employing molecular dynamics (MD) simulation methods to model hydrated PFPA-based and PFSA-based membranes in an equilibrium state at two different hydration levels.

Methods

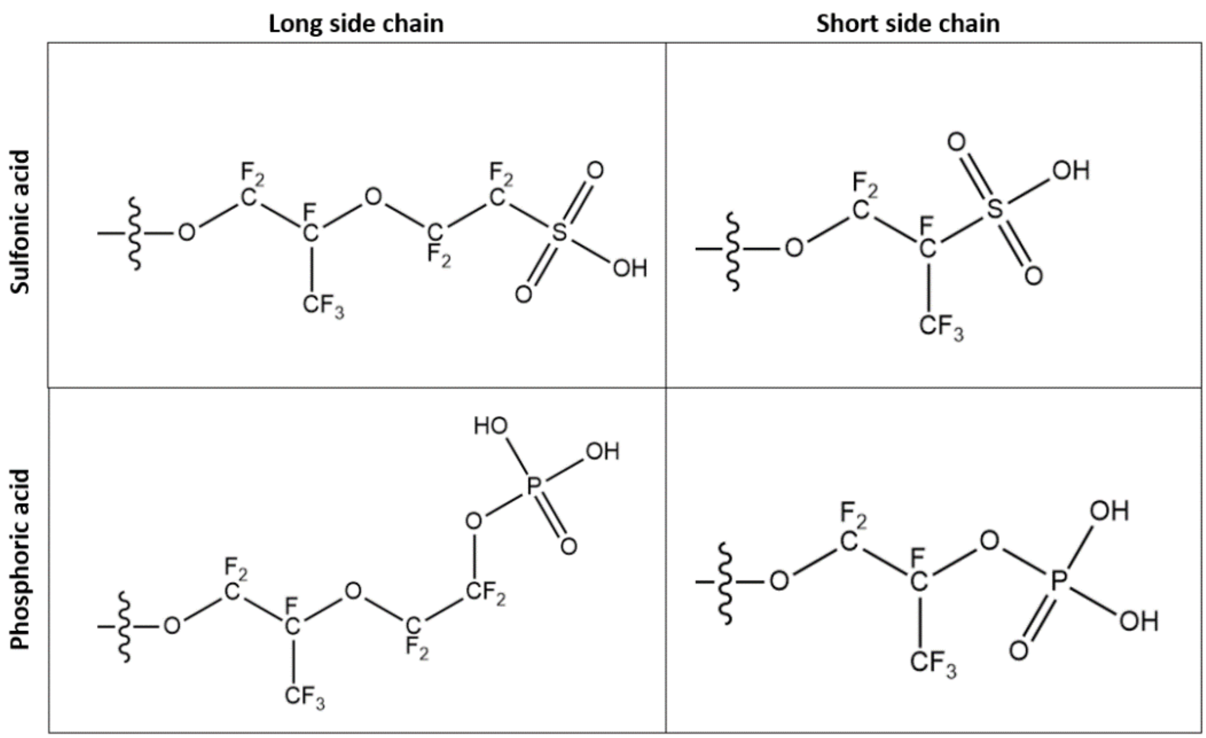
The fully atomistic simulation models were comprised of perfluorinated polymers, water, and hydronium ions. PFSA-based and PFPA-based polymer chains were prepared to have long and short side chains as presented in Figures 1.1a and 1.1b. The chemical structure of the polymer backbone is identical for all cases in this study while the chemical structures of the side chain and terminal acid group were varied. In this study, four perfluorinated polymers were simulated: S LSC (long side chain with sulfonic acid, original Nafion 117), S SSC (short side chain with sulfonic acid), P LSC (long side chain with phosphoric acid), and P SSC (short side chain with phosphoric acid).

First, the monomer structure for each polymer was optimized using DFT with B3LYP and 6-31G**. Each polymer chain consists of ten monomers, and each simulated membrane has four polymer chains. Then, Monte Carlo techniques were used to build amorphous structures with three-dimensional periodic boundary conditions. To investigate the effect of water content, water molecules were included to make 5 wt % and 20 wt % water contents in simulated membranes.[59] 40 hydronium ions were included due to the ionization of the acid groups. It was assumed that all sulfonic acid groups are ionized as supported by IR experimental observations[73-74] while all the phosphoric acid groups are also ionized based on their pK_a.[75]

To perform classical molecular dynamics (MD) simulation, we use the DREIDING force field as described above. Once the initial membrane structures were prepared, the annealing procedure was performed using LAMMPS software[76] which was modified to handle the force fields used in this study.[76] In this procedure, the volume and temperature were systematically varied to reach equilibrium. Experimental data shows that in a Nafion system at 300 K with a water content of 20 wt %, a density of 1.75 g/cm^3 is achieved.[77-78] To attain this density in simulation efficiently, the simulation cells were prepared to have an initial density of 1.9 g/cm^3 prior to the annealing procedure, so that, during the procedure, the simulation cells were allowed to expand and attain the desired density at equilibrium without direct input. The equations of motion were integrated via the velocity Verlet algorithm using a time step of 1.0 fs.[27] During the annealing procedure, the volume of the simulation cell is increased by 50% with increasing temperature from 373 K to 500 K over a period of 100 ps. Following this, a NVT MD simulation was conducted at 500 K for another 100 ps. The system was then compressed back to its initial size and simultaneously cooled to its original temperature over the course of 200 ps. This procedure was repeated five times. Thereafter, an NVT MD simulation and a subsequent NPT MD simulation were conducted using Nose-Hoover thermostat for 200 ps and 1 ns, respectively, to complete the annealing procedure.[28-30] Finally, NPT MD simulation was performed for 20 – 30 ns, whose equilibrated MD trajectory (Figures 1.2a and 1.2b) was used for various statistical analyses. The equilibrated structures at 20 wt % water content are presented in Figure 1.3. An identical procedure was repeated for all systems at 5 wt % water content.

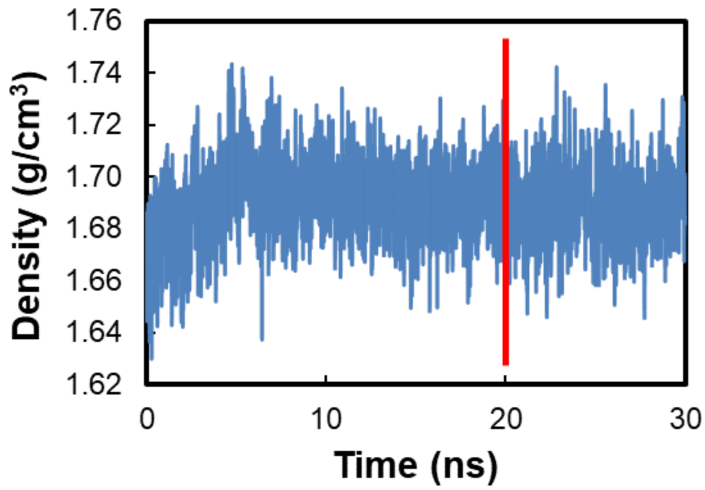


(a)

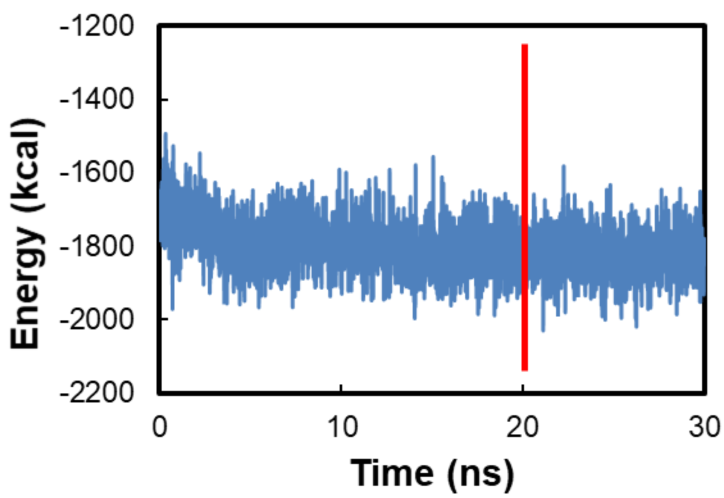


(b)

Figure 1.1. (a) Traditional Nafion polymer structure; (b) side chain variants used in this study.



(a)



(b)

Figure 1.2. Evolution of density (a) and potential energy (b). Red band indicates starting point of equilibrium state.

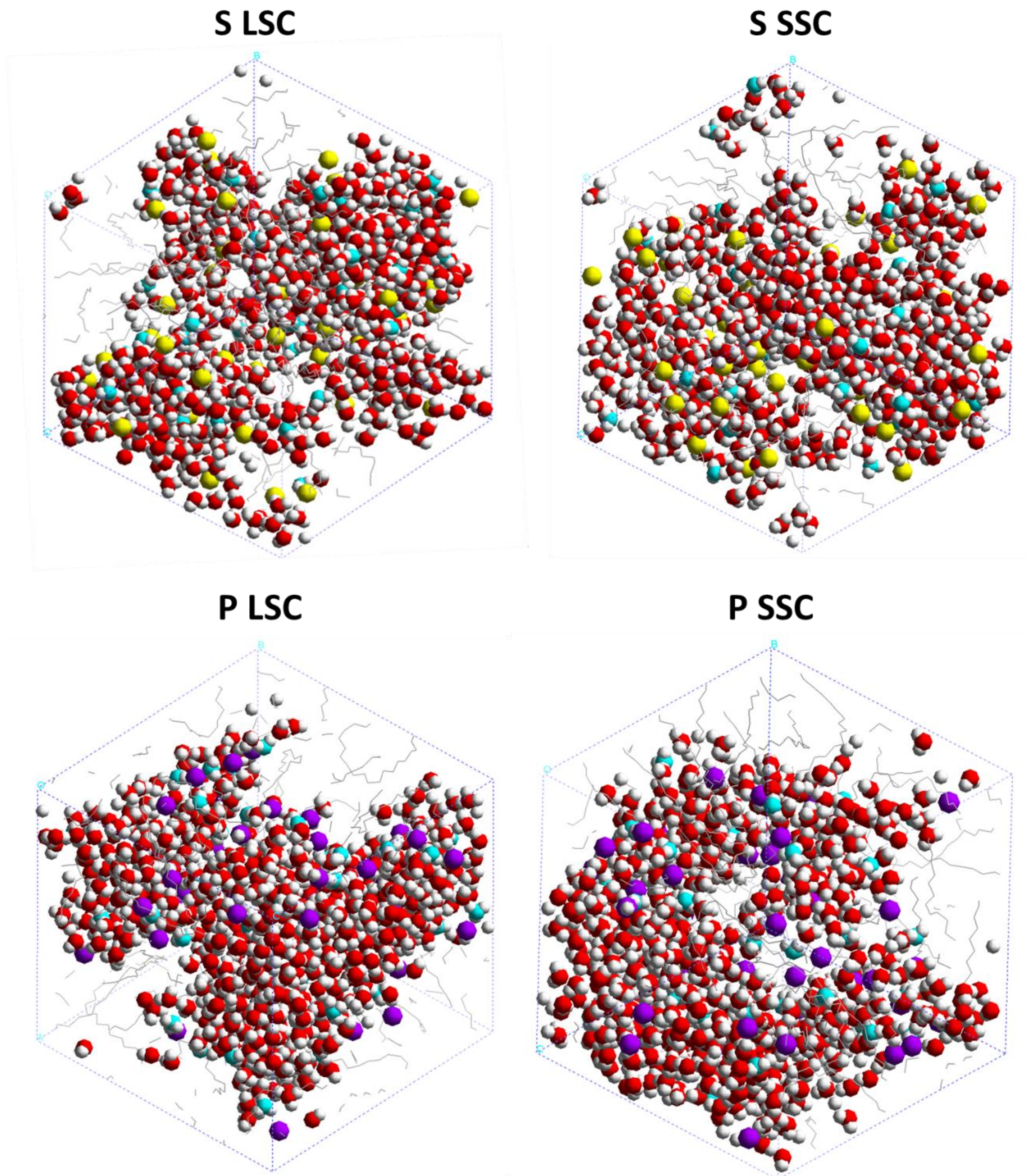


Figure 1.3. Nanophase segregated structures of PFSA and PFPA membranes at 20 wt % water content. The grey line denotes the carbon in polymer backbone. The blue, red and white denote oxygen in hydroxide, oxygen in water, and hydrogen, respectively. The yellow and magenta denote the sulfur and phosphorous, respectively.

To understand the behavior of PFSA and PFPA membranes, we analyze 1) static structural properties such as pair correlations and structure factors, and 2) dynamic transport properties such as molecular diffusion and proton hopping. These analyses serve to characterize nanophase segregation and transport. With these methods, we seek to compare the influence of side chain length on PFSA and PFPA membranes, thereby guiding future studies in membrane development.

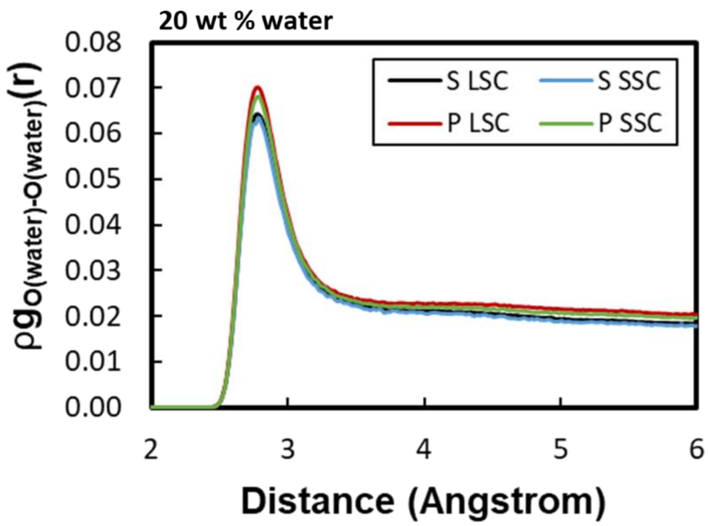
Pair Correlations and Coordination Numbers. Previous studies have led us to expect that in PFSA membranes, longer side chains may induce more matured water phase towards liquid water within the membrane as indicated by stronger pair correlations for adjacent water molecules.[62] Indeed, in Figures 1.4a and 1.4b, it is observed that $\rho \cdot g_{O(water)-O(water)}(r)$ for the first solvation shell is larger in S LSC than in S SSC, and likewise, P LSC membranes show greater water-water association than P SSC membranes. These are quantitatively confirmed by O(water)-O(water) coordination numbers (Table 1.1) which are calculated by integrating over the first solvation shell in the appropriate pair correlation function. In Figures 1.5a and 1.5b, water molecules are displayed as red spheres while polymer chains are hidden from view, providing a qualitative representation of these water phases and trends therein.

A noteworthy observation in Figures 1.4a and 1.4b is that the effect of side chain length on $\rho \cdot g_{O(water)-O(water)}(r)$ is stronger at 5 wt % water content than at 20 wt % water content. This is likely because at high hydration, the water phase is already well-established, and therefore the variation of side chain length has lesser effect. Another observation is that, at 20% water content, the water phase in PFPA membranes is more developed than that of the PFSA membranes regardless of side chain length (Figure 1.4a). We infer that the higher hydrophilicity of phosphate could be responsible for this observation.

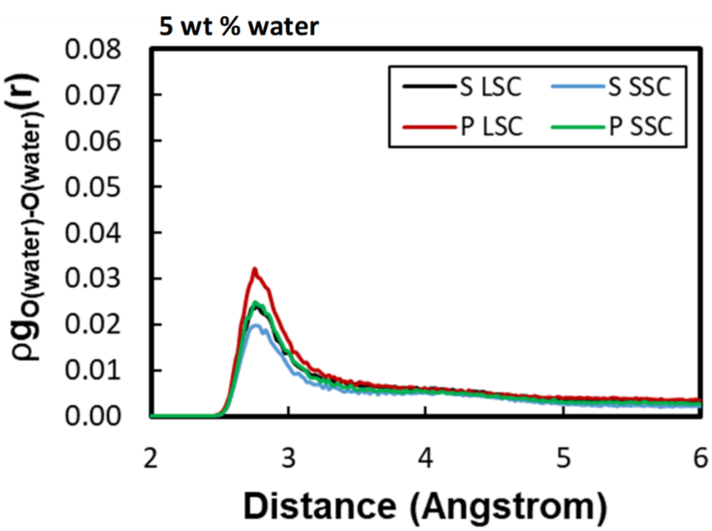
Table 1.1. Coordination numbers for all simulated systems

Water Content (wt %)	Polymer	FG[a] to $O(H_3O)$	FG[a] to $O(H_2O)$	$O(H_3O)$ to $O(H_2O)$	$O(H_2O)$ to $O(H_2O)$
20	S LSC	0.58	6.13	4.31	3.40
	S SSC	0.67	5.77	3.94	3.34
	P LSC	1.21	5.59	3.07	3.65
	P SSC	0.99	5.21	3.44	3.54
5	S LSC	2.26	3.23	1.75	1.28
	S SSC	2.31	2.53	1.59	1.12
	P LSC	2.03	2.73	1.40	1.57
	P SSC	1.97	2.24	1.41	1.36

[a] functional group

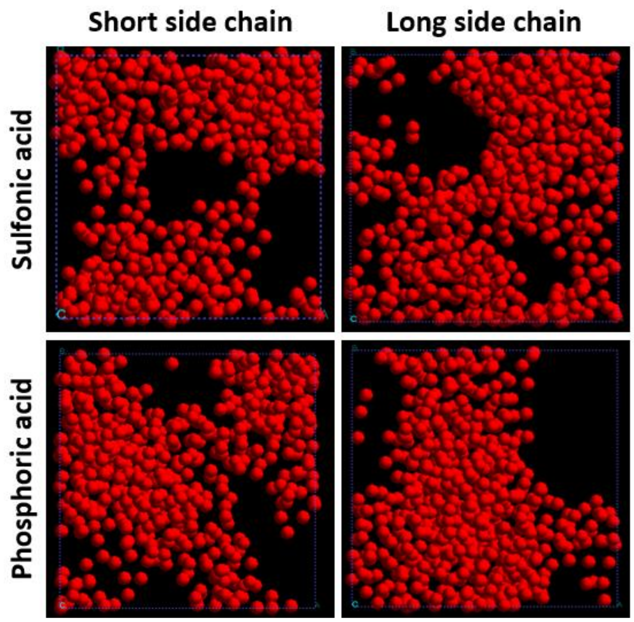


(a)

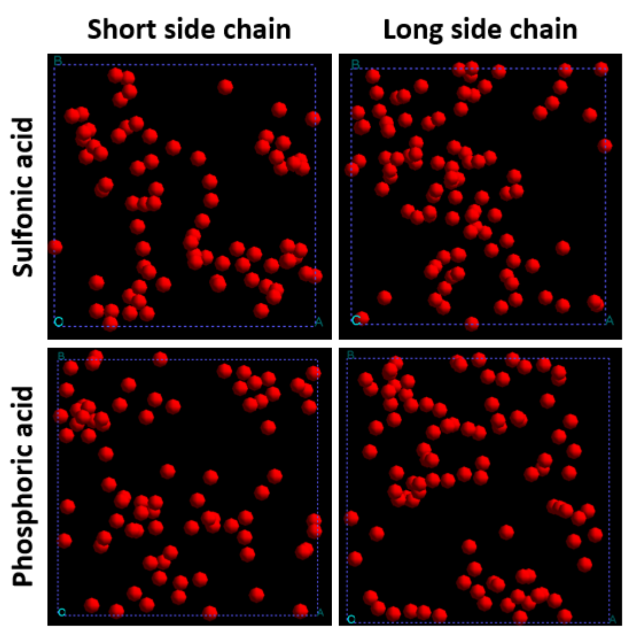


(b)

Figure 1.4. O(water)-O(water) pair correlations at 20 wt % water content (a) and 5 wt % water content (b).



(a) 20 wt % water content

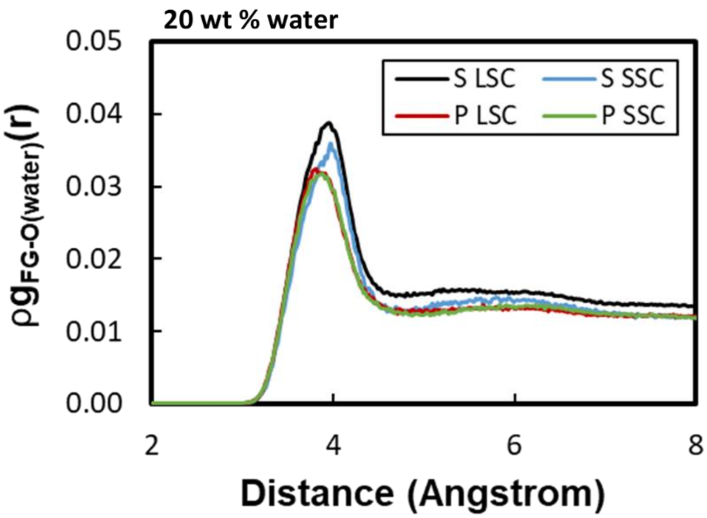


(b) 5 wt % water content

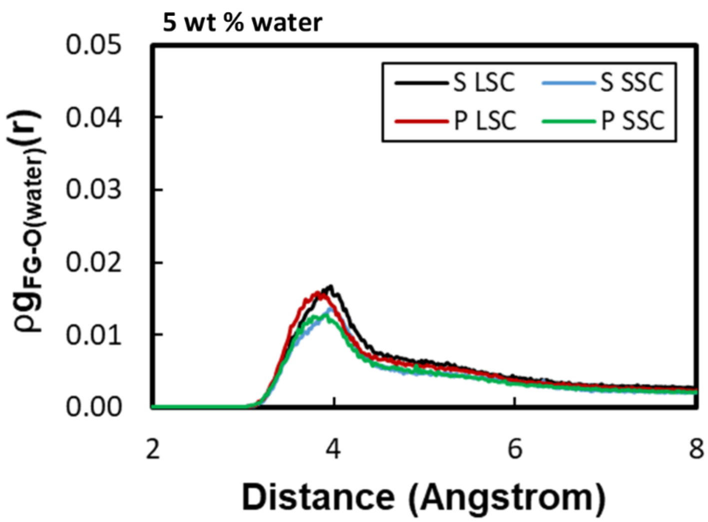
Figure 1.5. Nanophase segregation trends in PFSA and PFPA membranes at 20 wt % (a) and 5 wt % (b) water content. As chain length increases, water molecules (red spheres) tend to aggregate in larger clusters, especially in PFPA membranes.

Next, the solvation of phosphate and sulfonate groups by water molecules is characterized. For this, we calculate $\rho \cdot g_{FG-O(water)}(r)$ where FG denotes functional group as shown in Figures 1.6a and 1.6b. While $\rho \cdot g_{FG-O(water)}(r)$ is greater for LSC membranes at both water contents, which is quantitatively confirmed by the coordination number (Table 1.1), the change is more significant at 5 wt % water content. Hence, Figures 1.4 and 1.6 demonstrate that a well-established water phase leads to better solvation of phosphate and sulfonate groups as expected.

On the other hand, we would also expect that the hydroniums are dissolved in the water phase, and the extent of this dissolution should depend on the maturity of the water phase which is a measure of how the internal structure of water phase resembles that of bulk liquid water phase. Therefore, in Figures 1.7a and 1.7b, $\rho \cdot g_{FG-O(hydronium)}(r)$ for PFSA membranes is decreased as the water content is increased, indicating more dissociation of hydronium molecules from negatively charged functional groups at higher hydration condition. Likewise, in Figures 1.8a and 1.8b, $\rho \cdot g_{O(hydronium)-O(water)}(r)$ for PFSA membranes is increased as the water content is increased, indicating more solvation of hydronium molecules by water molecules. Together, these data indicate the greater dissociation of hydronium into the water phases of the S LSC membranes. The calculated coordination numbers support these observations (Table 1.1). However, for the PFPA membranes, we observe distinctly opposite trend around the phosphate groups in comparison to the sulfonate groups in PFSA membranes. As the side chain length increases, hydronium is less solvated and its association with functional groups increases, as indicated by the pair correlation functions (Figures 1.7 and 1.8) and as confirmed by their associated coordination numbers (Table 1.1).

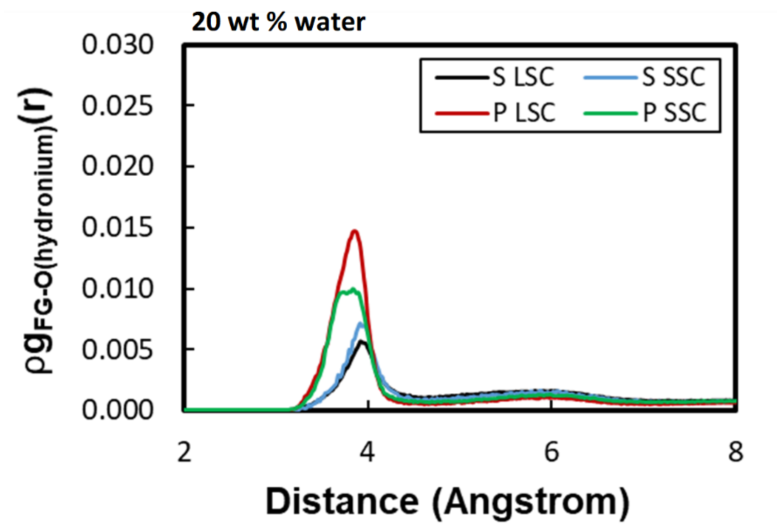


(a)

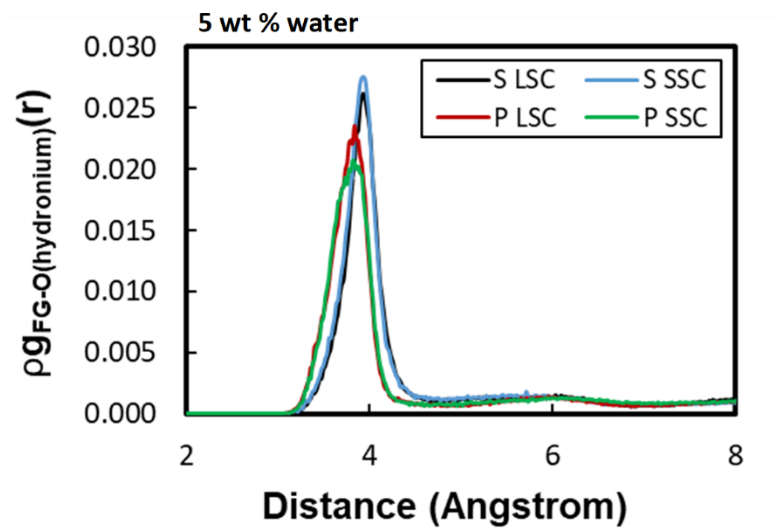


(b)

Figure 1.6. FG-O(water) pair correlations at 20 wt % water content (a) and 5 wt % water content (b). FG denotes S for sulfonate group or P for phosphate group.

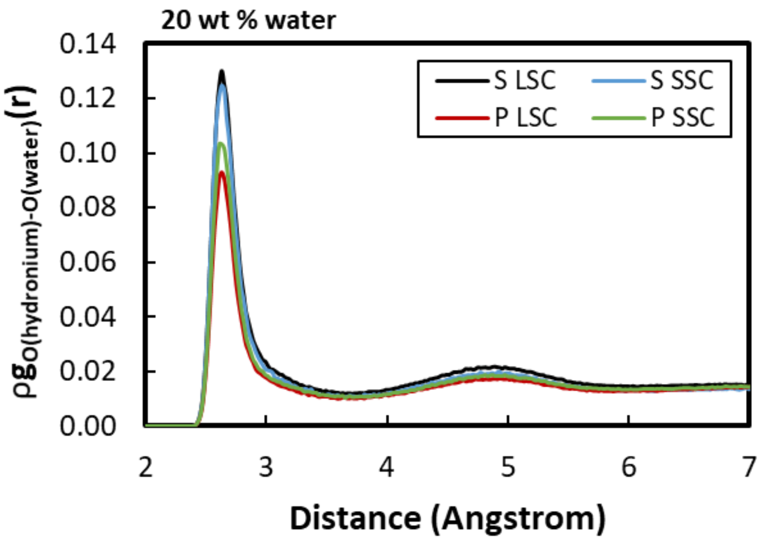


(a)

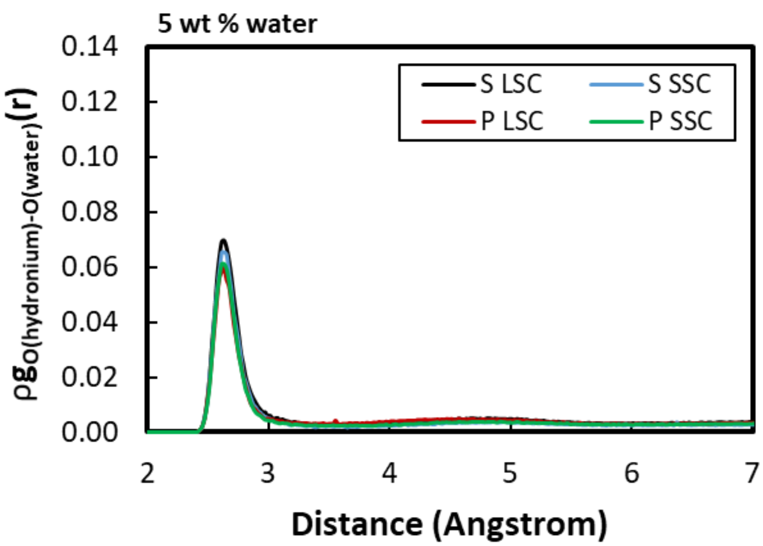


(b)

Figure 1.7. FG-O(hydronium) pair correlations at 20 wt % water content (a) and 5 wt % water content (b). FG denotes S for sulfonate group or P for phosphate group.



(a)



(b)

Figure 1.8. O(hydronium)-O(water) pair correlations 20 wt % water content (a) and 5 wt % water content (b).

An explanation for this puzzle is as follows: considering the pair correlations in Figures 1.6, 1.7, and 1.8, it appears that the hydronium molecules are dissociated from sulfonate groups, but associated with phosphate groups. This is corroborated by the $\rho \cdot g_{FG-FG}(r)$ results (Figures 1.9a and 1.9b), wherein the intensity of the first shell for P LSC membranes at 20 wt % water content is significantly higher than that of P SSC membranes (Figure 1.9a). This indicates that, given enhanced conformational diversity due to longer side chain, the phosphate groups prefer to more closely aggregate with themselves through the association with hydronium molecules (Figure 1.10). We call this a "hydronium-mediated bridge configuration." Although this bridge configuration has been observed in both PFSA and PFPA membranes, it previously had not been studied in conjunction with the effect of side chain length.[50] Notably, at 5 wt % water content, $\rho \cdot g_{FG-O(hydronium)}(r)$ (Figure 1.7b) and the first shell of $\rho \cdot g_{FG-FG}(r)$ in PFPA membranes (Figure 1.9b) are only slightly increased with the side chain length, which implies that, at low water content, the bridge configuration mediated by hydronium is not enhanced in PFPA membranes with side chain length.

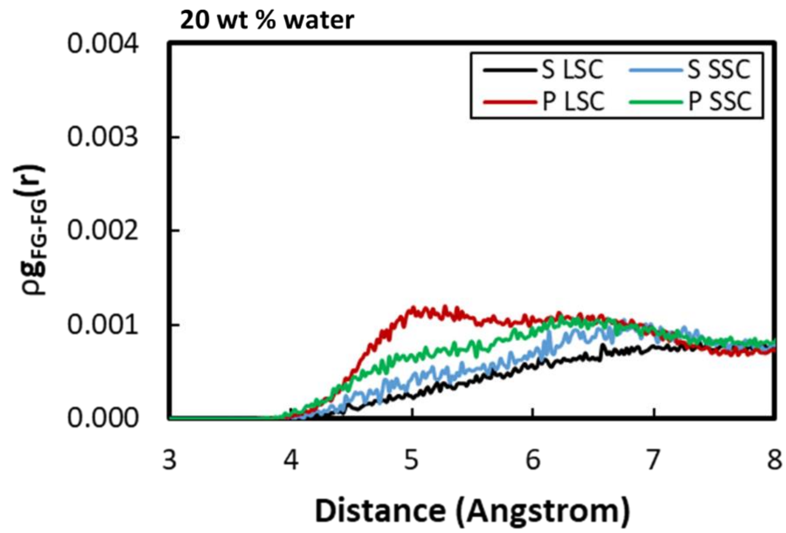
To confirm our hypothesis regarding hydronium localization from an energy standpoint, we compared the aqueous phase energy densities (EDs) between our systems at 20 wt % water content. Prior to completing this calculation, we obtained the H_3O^+ – H_2O and H_2O – H_2O pair potential energy curves in order to understand how these interactions could impact the ED of our systems. These potential curves are depicted in Figure 1.11. Since the H_3O^+ – H_2O potential minimum is comparatively negative, we surmise that this pair will bear a perceivable impact on the overall ED of the system despite the abundance of water molecules. To calculate this parameter, we divided the total energy of the aqueous phase (hydronium and water molecules) by its Connolly volume. The results from this study are tabulated in Table 1.2. These trends are expected considering that with increasing chain length, hydronium tends to associate with the aqueous phase in sulfonic acid membranes and tends to dissociate from this phase in phosphoric acid membranes (due to bridge formation).

Another intriguing point is that, at low hydration, the sulfonate groups in PFSA membranes tend to associate more tightly with other sulfonate groups as well as with hydronium in comparison to the phosphate groups in PFPA membranes (Figures 1.7b and 1.9b). This may be because, at low hydration, the less hydrophilic sulfonate groups are less solvated compared to the phosphate groups. Nonetheless, hydronium molecules in the PFSA membrane are still more solvated than those in the PFPA membrane, so that $\rho \cdot g_{O(hydronium)-O(water)}(r)$ in the PFPA membrane is not as repressed as in PFSA membranes despite sulfonate groups potentially becoming involved in hydronium bridging.

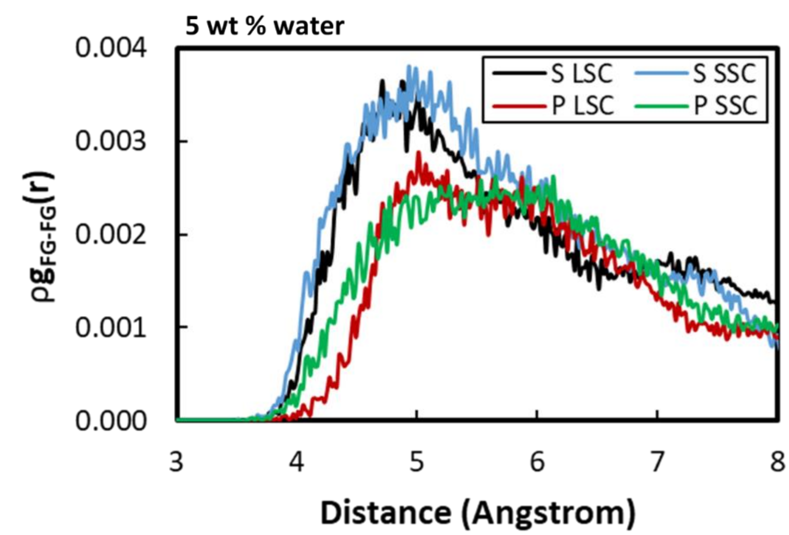
To solidify our findings, we quantified the amount of bridging in each system using the following criteria to denote a "bridge":

- 2 or more functional groups are within the vicinity of the hydronium
- Each functional group possesses at least one O atom located < 2.5 angstroms from an H atom in the hydronium

If two functional groups participate, it is counted as 1 bridge. If three functional groups participate, it is counted as 3 bridges. The results are detailed in Table 1.3. The trends in these bridge counts correspond to the trends in conductivity with side chain length, confirming our previous conclusion.



(a)



(b)

Figure 1.9. Pair correlation between functional groups at 20 wt % water content (a) and 5 wt % water content (b). FG denotes S for sulfonate group or P for phosphate group.

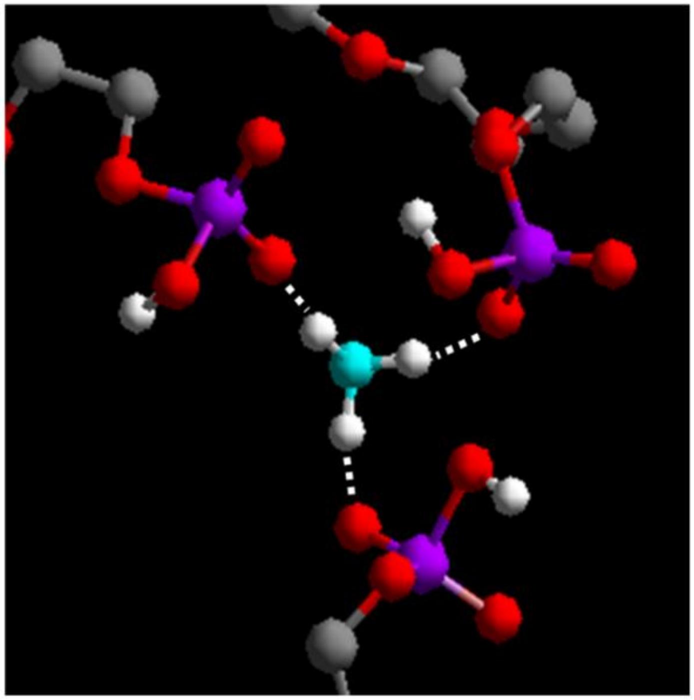


Figure 1.10. Hydronium-mediated bridge configuration consisting of one hydronium and three phosphate groups in P LSC membranes.

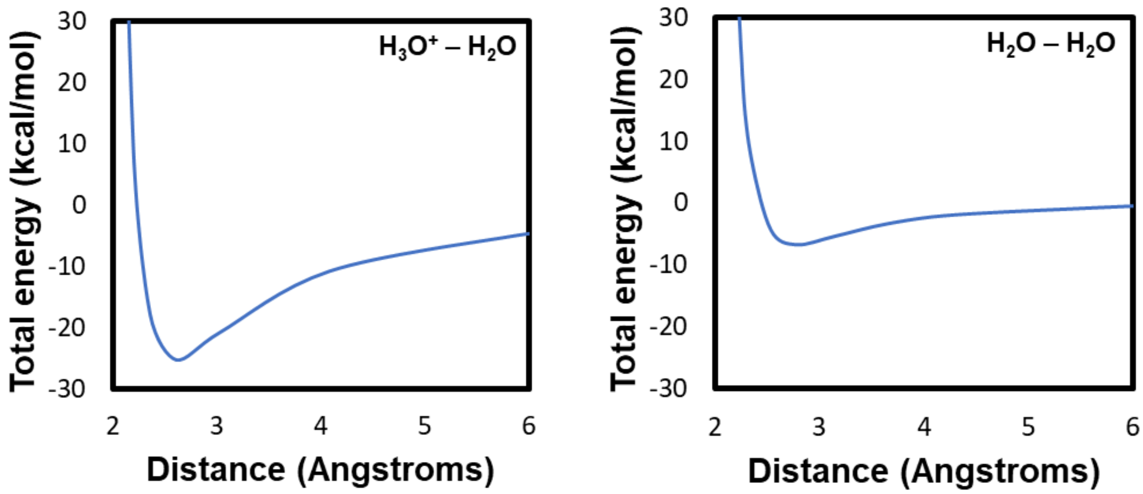


Figure 1.11. Pair potential energy curves for H_3O^+ – H_2O and H_2O – H_2O.

Table 1.2. Energy densities for all simulated systems at 20 wt % water content

Polymer	**Energy density (kcal/mol- Å^3)**
S LSC	–0.27
S SSC	–0.24
P LSC	–0.22
P SSC	–0.25

Table 1.3. Bridge counts for all simulated systems

Water Content (wt %)	Polymer	Bridge Count
20	S LSC	1
	S SSC	3
	P LSC	8
	P SSC	2
5	S LSC	23
	S SSC	31
	P LSC	16
	P SSC	13

Structure Factor Analysis. To quantitatively analyze the effect of side chain length on the large scale nanophase segregated morphology in PFPA and PFSA membranes, we calculate the structure factor, $\mathbf{S}(\mathbf{q})$, which is defined as:

$$\mathbf{S}(\mathbf{q}) = \left\langle \sum_{r_i} \sum_{r_j} \exp\left(i\mathbf{q} \cdot \mathbf{r}_{ij}\right)\left(\xi^i \xi^j - \langle \xi \rangle^2\right) \right\rangle / L^3 \qquad (2)$$

where the angular bracket denotes a thermal statistical average. ξ^i represents a local density contrast, $\mathbf{q}$ is the scattering vector, and $\mathbf{r}_{ij}$ is the vector between the sites i and j. Traditional SAXS and SANS experiments rely on the density contrast between hydrophilic and hydrophobic phases at a certain scattering length, to determine the structure factor. However, our structure factor is calculated from an artificial density contrast as follows: the local density variables are Φ_j^A and Φ_j^B: Φ_j^A is equal to 1 if the site j is occupied by a hydrophilic entity such as water or hydronium group and equal to 0 otherwise, and Φ_j^B is equal to 1 if the site is occupied by hydrophobic entities such as the polymer backbone or equal to zero otherwise. The quantity $\mathbf{S}(\mathbf{q})$ is spherically averaged as follows:

$$S(q) = \sum_{|\mathbf{q}|} \mathbf{S}(\mathbf{q}) \,/ \sum_{|\mathbf{q}|} 1 \qquad (3)$$

with $q = (2\pi/L)n$ where $n = 1,2,3,\cdots$ denotes that, for a given n, a spherical shell is taken as $n - 1/2 \leq qL/2\pi \leq n + 1/2$. This analysis is completed for five individual frames within the equilibrium state of each simulation, and these five $S(q)$ profiles are subsequently averaged.

The results from the structure factor analysis at 20 wt % water content are depicted in Figure 1.12. The location of the main peak is consistent with the experimental "ionomer peak" observed at q = ~0.13 A^{-1} indicated by the dotted line.[53] In accordance with our pair correlation studies and qualitative observations, P LSC membranes exhibit a higher $S(q)$ intensity than P SSC membranes, indicating a greater extent of nanophase segregation. The intensities of $S(q)$ for S

SSC and S LSC membranes are comparable to each other, in contrast to the improved local structure of the aqueous phase in S LSC membranes. This indicates that, at large length scales, chain length has little effect on phase segregation in PFSA membranes. Consistent with $\rho \cdot g_{O(water)-O(water)}(r)$, PFPA membranes show greater nanophase segregation than PFSA membranes, which may offset the detriment to conductivity due to the hydronium-mediated bridge configuration. The structure factor at 5 wt % water content was not analyzed due to insufficient water molecules to form a water phase for structure factor analysis.

Proton Transport. Table 1.4 summarizes the vehicular diffusion coefficients for hydronium in all simulated systems. For both PFSA and PFPA membranes, the vehicular diffusion coefficients for hydronium increase with the side chain length, which can be attributed to the more developed water phase providing a better path for molecular diffusion. However, it should be noted that, at 20 wt % water content, the increase in the hydronium diffusion coefficient in PFPA membranes are not as large as that of PFSA membranes, likely due to the hydronium-mediated bridge configuration discussed above, which restricts hydronium movement.

In order to estimate proton conductivity, it is first necessary to estimate proton hopping. For this purpose, we employ the method developed in Deng and his co-workers,[79-80] in which the hopping diffusion coefficient is calculated as follows:

$$D_{hopping} = \frac{1}{6Nt} \int_{0}^{t \to \infty} \sum_{i}^{N} \sum_{i}^{M} k_{ij} r_{ij}^{2} P_{ij} \mathrm{d}t$$

where N is the number of protons, P_{ij} is the probability with which a proton can jump from hydronium i to water j defined as $P_{ij} = k_{ij} / \sum_{j}^{M} k_{ij}$, r_{ij} is the distance between all donors and accepters measured from the equilibrium molecular dynamics trajectory, and k_{ij} is defined as follows:

$$k_{ij}(r) = \kappa(T, r)\frac{k_b T}{h}\exp\left(-\frac{E_{ij}(r) - 1/2h\omega(r)}{RT}\right)$$

where $\kappa(T, r)$ and $\omega(r)$ are the tunneling factor and frequency for zero-point energy correction (given in refs.[81-82]) and $E_{ij}(r)$ is the energy barrier for the proton to be transferred from donor to acceptor in water medium while they are at a distance (r). To assess this hopping energy barrier, we used the relative energy change as a function of the distance between the donor and acceptor oxygens and corrected by considering solvent effects using the Poisson-Boltzmann self-consistent reaction field model as described in Jang and his co-workers.[80]

In Table 1.4, we find hopping coefficients to be very similar in all membranes and in magnitude agreement with ref.[80], which predicted hopping coefficients to be on the order of 10^{-6} cm^2/s. This indicates that our proton hopping estimation is of a reasonable order of magnitude. After obtaining our hopping diffusion coefficients, we proceed to sum vehicular and hopping diffusion coefficients to obtain total diffusion coefficient values which are tabulated in Table 1.4. From these values, we calculate conductivity via the following Nernst-Einstein equation:

$$\sigma = \frac{Dcz^2F^2}{RT}$$

In this equation, σ is conductivity, D is diffusion coefficient, c is concentration of hydronium ions in mol/cm^3, F is the Faraday constant, R is the gas constant, T is temperature, and z is the charge on hydronium (+1 in this case). The estimated proton conductivities are tabulated in Table 1.5.

In PFSA membranes, the proton conductivity shows a small increase with increasing side chain length at both hydration levels, as expected based on the slightly improved local structures in the water phase. By contrast, in PFPA membranes, the proton conductivity decreases despite improvement of such local structures as a function of side chain length, which is inferred that the hydronium-mediated bridge configuration in PFPA membranes imposes a very effective restriction on proton transport.

This hypothesis is confirmed via an analysis of hydrogen bond lifetimes between hydronium and functional group in these systems. We used the following criteria to determine whether a bond is present:[83-88]

- The distance R_{OO} between the oxygens is shorter than 3.4 Å,
- The distance R_{OH} between the "acceptor" oxygen and the hydrogen of the "donor" molecule is shorter than 2.425 Å,
- The H – O – O angle is smaller than 30°.[89]

We first validated our protocol by monitoring the lifetime of hydrogen bonds between water molecule pairs in one of our systems. We obtained a value of 2.0 ps as an average lifetime, which falls within the reported range of 1 – 5 ps.[90] Then, we proceeded to monitor hydrogen bonding lifetime between functional group/hydronium pairs. Results are summarized in the Table 1.6. In both membrane chemistries lifetime increases with side chain length, potentially because conformational diversity provides more opportunities for favorable bond configurations. However, this increase in the phosphoric acid system is more drastic, likely due to bridging.

To further confirm, we analyzed the residence time of the entire hydronium molecule around the functional groups in each system. Residence time is defined as the amount of time that the oxygen in the hydronium ion spends at a distance < 4.3 angstroms (the cutoff distance for the first hydronium shell; refer to Figure 1.7) from the central atom in the functional group. Results are tabulated in Table 1.6. Like hydrogen bonding, residence time increases with chain length in both systems, albeit much more drastically in the phosphoric acid system.

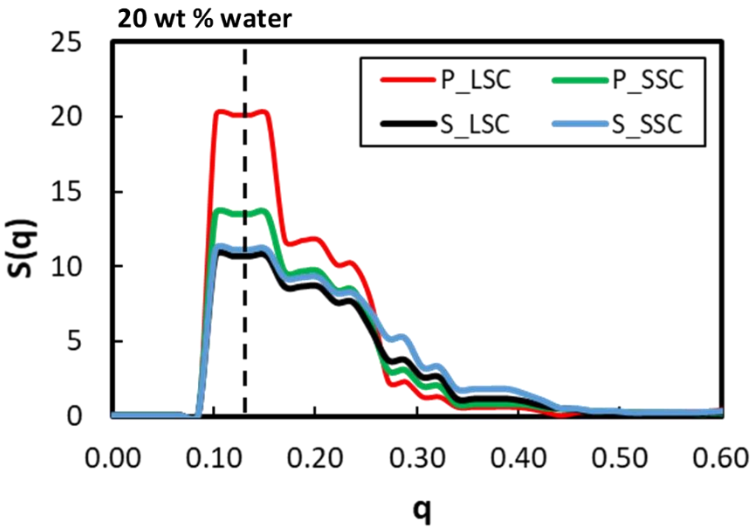


Figure 1.12. Structure factor profiles at 20 wt % water content.

Table 1.4. Proton diffusion coefficients

Water Content (wt %)	**Polymer**	**D(H_3O^+, vehicular)** (cm^2/s)	**D(hopping)** (cm^2/s)	**D(total)** (cm^2/s)
20	S LSC	5.41E-06	3.23E-06	8.64E-06
	S SSC	3.85E-06	2.94E-06	6.80E-06
	P LSC	2.70E-06	2.31E-06	5.01E-06
	P SSC	2.47E-06	2.58E-06	5.05E-06
5	S LSC	4.39E-08	1.32E-06	1.36E-06
	S SSC	2.09E-08	1.21E-06	1.23E-06
	P LSC	2.54E-08	1.08E-06	1.10E-06
	P SSC	8.35E-09	1.06E-06	1.07E-06

Table 1.5. Estimated proton conductivity

Water Content (wt %)	**Polymer**	**Conductivity (vehicular)** (S/cm)	**Conductivity (hopping)** (S/cm)	**Conductivity (Total)** (S/cm)
20	S LSC	0.0193	0.0125	0.0308
	S SSC	0.0156	0.0124	0.0275
	P LSC	0.0086	0.0095	0.0177
	P SSC	0.0087	0.0110	0.0198
5	S LSC	0.0002	0.0063	0.0065
	S SSC	0.0001	0.0063	0.0065
	P LSC	0.0001	0.0051	0.0052
	P SSC	0.0000	0.0055	0.0055

Table 1.6. Estimated hydrogen bond lifetimes and residence times between H_3O^+ and functional group at 20 wt % water content

Polymer	Hydrogen Bond Lifetime (ps)	Residence Time (ps)
P LSC	29.11	26.00
P SSC	10.35	10.51
S LSC	14.69	16.69
S SSC	7.47	9.30

Notably, the proton conductivities of all membranes are much more similar at 5 wt % water content than at 20 wt % water content. However, PFSA membranes still perform slightly better, likely due to the higher correlation between H_3O^+ and H_2O as evidenced in Figure 1.8b. This correlation facilitates the hopping mechanism, which is a major contributor to proton diffusion and membrane conductivity according to Tables 1.2 and 1.3.

Conclusions

We performed molecular dynamics simulations of PFSA and PFPA membranes to understand the effect of side chain length and acid strength on the nanophase segregation and transport. For PFPA and PFSA systems at 20 wt % and 5 wt % water contents, we find that the increase of the side chain length gives rise to more matured local structures in water phase. This is supported by the pair correlation functions and their associated coordination numbers. The structure factor analysis for PFPA also reflects greater segregation of hydrophobic and hydrophilic regions for long side chain membranes at 20 wt % water content. This phase segregation correlates with enhanced solvation of the sulfonate and phosphate groups as anticipated.

Consistent with the improved local structures in water phase, we note an increase in conductivity with side chain length for PFSA membranes at both hydration levels. In contrast, we do not observe a corresponding increase in conductivity for PFPA membranes. It turns out that this is because the longer side chain leads to a hydronium-mediated bridge configuration, reflected by the intensity increase in the $\rho \cdot g_{FG-O(hydronium)}(r)$ and $\rho \cdot g_{FG-FG}(r)$ pair correlation functions as well as a decrease in $\rho \cdot g_{O(hydronium)-O(water)}(r)$. This phenomenon is hypothesized to restrict hydronium mobility, which would be detrimental to conductivity. In addition, although still occurring, the bridging phenomenon appears less pronounced at low hydration conditions.

These results indicate a useful trend for PFPA membranes and PEM design in general. Hydronium transport is correlated with bridge structure formations, and at standard conditions can therefore be augmented simply by shortening the side chain. Upon further research into the scope of this phenomenon, it could be combined with other enhancement methods to design a polymer with increased dehydration resistance while boasting appropriate conductivity at standard conditions, thus allowing for adoption of fuel cell technologies into a wider range of conditions.

III. **AIM 2.** Radical scavenging capabilities of CeO_2 towards stability of polymer electrolyte membranes.

Introduction

Chemical stability of the PEMFC's polymer electrolyte membrane (PEM) is crucial to its prolonged performance in long-term applications such as solar cells, batteries, and fuel cells[91-92]. However, the electrochemical formation of radical species, such as •OH and •OOH, at the electrode degrades the PEM, causing a decrease in membrane conductivity and leading to eventual membrane failure.[93-96] Researchers initially attempted to combat this degradation using Mn(II) and Ce(III) ions in the fuel cell, later discovering that Ce(III) was superior from a kinetic standpoint[10, 97] due to their ability to rapidly cycle between a Ce^{3+} and a Ce^{4+} state. This cycling ability is important since fuel cell radicals such as •OH and •OOH bind to the Ce^{3+} ion and subsequently undergo conversion to more innocuous species[98-99], concurrent with the ion's transition to a Ce^{4+} state. Later, the ion can be reconverted to Ce^{3+} and effectively "recycled."

While promising, these ions exhibit migration within the membrane, diminishing their effectiveness over time.[100-101] A solution is found in the use of nanoparticles rather than ions. Due to a set of pioneering experimental studies by Trogadas et al., CeO_2 nanoparticles ("ceria") have recently garnered interest as a radical scavenger for fuel cell membranes.[102-106] In these particles, the generation of Ce^{3+} is concomitant with the desorption (i.e. vacancy formation) and adsorption of oxygen on the ceria surface[107]. Researchers have taken strides toward improving these nanoparticles by investigating the effects of particle size, facet, and notably, vacancy concentration upon scavenging performance.[108] However, the effect of defect geometry upon scavenging ability remains to be investigated.

According to previous studies, vacancies tend to form on the CeO_2 surface in either triangular or linear patterns.[15, 109-111] The relative incidence of triangular vs. linear vacancy clusters depends upon synthesis conditions and is therefore likely tunable.[109-110] Still, it is unclear which defect geometry is superior for radical scavenging. Thus, in this study, we attempt to illuminate the importance of defect geometry in the adsorption of •OH and •OOH radical species using DFT. Through these calculations, we will not only predict which defect geometry is favorable for radical scavenging, but also explore the electronic-scale reasoning behind this preference.

Computational Methods

The calculations reported herein were performed on the basis of spin-polarized DFT within the generalized gradient approximation (GGA) of the Perdew-Burke-Ernzerhof (PBE) functional, as implemented in the Vienna Ab-initio Simulation Package (VASP)[112-23]. The projector-augmented wave (PAW) method with a plane wave basis set was employed to describe the interaction between ion cores and valence electrons[113-114]. An energy cutoff of 430 eV was applied for the expansion of the electronic eigenfunctions. For the Brillouin zone integration, we initially used a (1×1×1) k-point grid to determine the optimal geometries and then increased to a (2×2×1) k-point grid to obtain total energies of the systems. In congruence with established techniques, we incorporate an empirical Hubbard U parameter (U = 5.0 eV) to overcome the poor description of GGA for the localized character of 4f electrons in ceria[115]. To account for van der Walls interactions, we used DFT-D3 corrections with Becke-Johnson damping[116-117].

For pristine and defective CeO_2, we constructed a slab of a cubic fluorite (Fm-3m) in a (4×4) supercell with 9 atomic layers, each of which contains 16 atoms. The (111) facet was chosen due to both high catalytic activity and a low surface energy compared to other low index surfaces, as summarized in literature.[108] As illustrated in Figure 2.1, three distinct defective surfaces were constructed: pristine, linear defective and triangular defective surfaces. The slab is separated from its periodic images in the z direction by a vacuum space spanning approximately 20 Å. The pristine, linear defective, and triangular defective CeO_2 slabs are geometrically optimized until residual forces on all the constituent atoms become smaller than 5×10^{-2} eV/Å, and the total free energy change in the self-consistent field procedure is less than 2×10^{-5} eV.

Next, we conducted a Mulliken population analysis from the optimized structures to determine the local charge and spin state across the defective ceria surfaces using the Cambridge Serial Total Energy Package (CASTEP).[118-119] The plane wave basis set and ultrasoft pseudopotential are utilized. The cutoff energy is chosen to be 370 eV and the k-point sampling is done using a (2×2×1) Monkhorst-Pack grid. Finally, a partial density of states (PDOS) analysis was conducted using VASP with k-points increased to a (3×3×1) Monkhorst-Pack mesh.

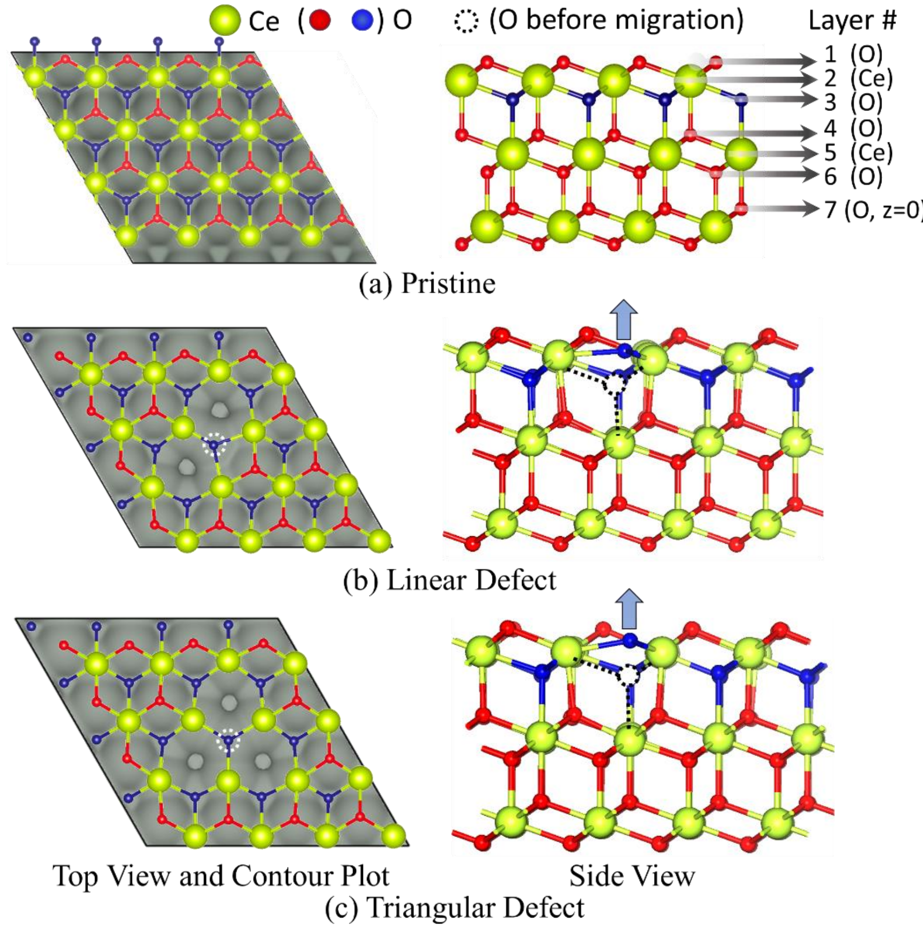


Figure 2.1. Charge density contour plots and optimized atomic structures of (a) pristine, (b) linear defect, (c) triangular defect on ceria (111) surface. For simplicity, the middle and bottom layers are hidden in the top views. In charge density contour plots, the defect sites are distinguished by the deficiency of charge density. Arrows in (b) and (c) indicate the shift direction of subsurface oxygen (dotted circle) during the geometry optimization.

Results and Discussion

Oxygen Vacancy Formation and Stability in CeO_2. Prior to analyzing the effect of vacancy cluster geometry on radical scavenging capability, we first confirmed the reasonableness of our structures. As shown in Figure 2.1, we have considered a pristine surface, a linear defective surface, and a triangular defective surface. These two defects were chosen because they are prevalent on ceria surfaces at standard vacancy density (> $10^{13}/cm^2$).[15, 109-110] For the linear vacancy, a divacancy was chosen since it falls within the range of commonly observed linear vacancy lengths while avoiding interactions between adjacent vacancies[109].

In order to validate our models, we compared the atomic z coordinates in the geometry-optimized pristine slab model with the experimental values from the literature as presented in Table 2.1. Since the z coordinates of atoms in our DFT calculations fall within 0.08 Å of the experimental values, our calculated structures are in good agreement with the structures observed in experiments, especially considering the uncertainty associated with low-energy electron diffraction.[120] In Table 2.2, we compared the bond lengths calculated via DFT with the experimental[120] and previous DFT[121] results. It is apparent that our bond lengths deviate from experimental values by approximately the same extent as those of the previous DFT studies (~1%)[121].

Table 2.1. Atomic z coordinates after geometric optimization compared with experimental results. Number in parenthesis denotes layer, with 1 being the topmost layer. O(7) is designated at z = 0. Experimental uncertainties in parenthesis

Atom	**Our DFT value (Å)**	**Experimental value[120] (Å)**
O(1)	6.35	6.28 (±0.06)
Ce(2)	5.55	5.42 (±0.02)
O(3)	4.74	4.66 (±0.06)
O(4)	3.19	3.12 (+0.04/-0.02)
Ce(5)	2.38	2.38 (±0.02)
O(6)	1.57	1.60 (+0.08/-0.04)

Table 2.2. Bond lengths in pristine slab after geometry optimization. Number in parenthesis denotes layer, with 1 being the topmost layer

Atom Pair	**Our DFT Value (Å)**	**Literature**	
		DFT[121] (Å)	**Experiment[120] (Å)**
O(1) – Ce(2)	2.350	2.365	2.37
Ce(2) – O(3)	2.353	2.370	2.34
Ce(2) – O(4)	2.353	2.322	2.30

After vacancy formation, the atoms adjacent to the defect are shifted due to charge redistribution. In our systems, we primarily observed lateral shifts. In the linear defect, cerium and oxygen atoms were shifted an average of 0.13 and 0.18 Å away from the vacancy in the x–y plane, respectively. Similarly, in the triangular defect, cerium and oxygen atoms were shifted an average of 0.10 and 0.17 Å away from the vacancy in the x–y plane, respectively. These lateral shifts are consistent with previous literature from an order of magnitude standpoint.[122-124] Intriguingly, the central, subsurface oxygen atoms in both linear and triangular defects were shifted significantly in the +z direction by 1.10 and 1.26 Å, respectively, as shown in Figure 2.1b and 2.1c. This is likely because the oxygen vacancy formation leaves two excess electrons that are localized at the vacancy sites reducing the neighboring Ce atoms.[124-125] Those surrounding undercoordinated Ce atoms would interact with the O atom associated with the vacancy site, increasing the Ce-O interaction. This strong interaction results in an atomic reconfiguration at the vacancy site that shifts the atomic position of the associated subsurface oxygen atom.

In order to evaluate the vacancy formation energy (VFE), we calculate *n*th vacancy formation energy (ΔE_{nth}) and average vacancy formation energy (ΔE_{avg}^{n})[124, 126] for given *n*-vacancy models using the following equations:

$$\Delta E_{nth} = E_{slab}^{n} - E_{slab}^{n-1} + \frac{1}{2} E_{O_2}$$

where *n* is the number of vacancies (*n* = 1, 2, 3), E_{slab}^{n} is the slab system energy with *n* vacancies, and E_{O_2} is the O_2 gas molecule energy, and

$$\Delta E_{avg}^{n} = (E_{slab}^{n} - E_{slab}^{0} + \frac{n}{2} E_{O_2})/n$$

where E_{slab}^{0} represents the energy of a pristine slab with zero vacancies. The results are summarized and compared in Table 2.3.

Table 2.3. Averaged formation energy per vacancy and formation energy of the n^{th} vacancy for point, linear, linear trimer, and triangular systems

System	**Avg. VFE [eV]**	**n^{th} VFE [eV]**
Point (n = 1)	2.44	2.44
Linear (n = 2)	2.42	2.40
Linear Trimer (n = 3)	2.52	2.73
Triangular (n = 3)	2.60	2.96

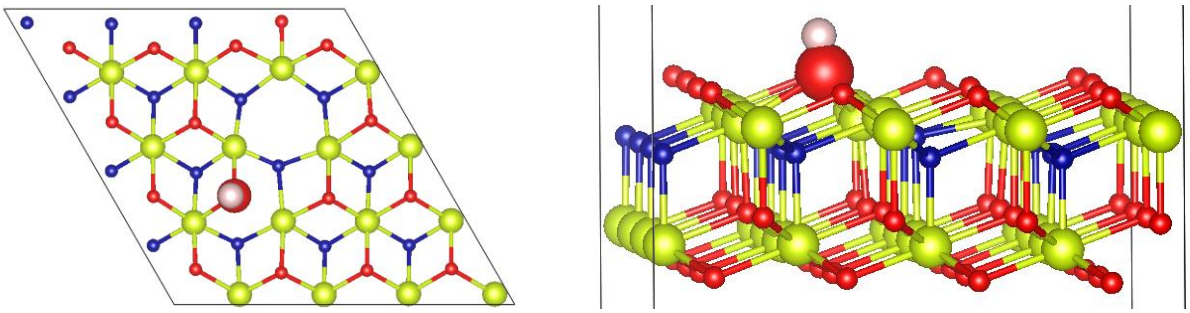
Both averaged and *n*th VFE increase with the the number of vacancies. The triangular defect system possesses the highest magnitudes on both energies, indicating less thermodynamic stability, which could predict a more favorable radical scavenging tendency. However, it should be noted that the magnitude of VFE is much higher than the energy difference between the VFE(n) and VFE(n+1), implying that even a single vacancy (n=1) could possess radical scavenging capabilities.

Adsorption of •OH and •OOH Radicals. We added radical species to the ceria surface and calculated binding energies to gauge the binding favorability of •OH or •OOH to a given defect site. Figures 2.2 and 2.3 depict the geometry-optimized structures after •OH and •OOH are bound to the linear and triangular defective surfaces. (Although this will be discussed in depth later, it should be noted that that the binding of •OH and •OOH on the defect-free CeO_2 surface is thermodynamically unfavorable).

First, we observe that the oxygen atom of •OH is positioned inside the vacancy in both triangular and linear defects. Since the defect contains no surface oxygen atom, the negatively polarized O atom of •OH is attracted to the defect by the adjacent, positively polarized cerium atoms. Additionally, since •OH is a radical molecule, its O atom prefers to interact with atoms of high spin state. As evident from our subsequent population analyses discussed below, the •OH is biased toward one cerium atom, which indicates the covalent bond characteristics through a change in spin state (Figures 2.4 and 2.5): for the linear defect, the spin of this Ce atom is decreased from 1.08 to 0.82, while for the triangular defect, it is decreased from 1.10 to 0.88.

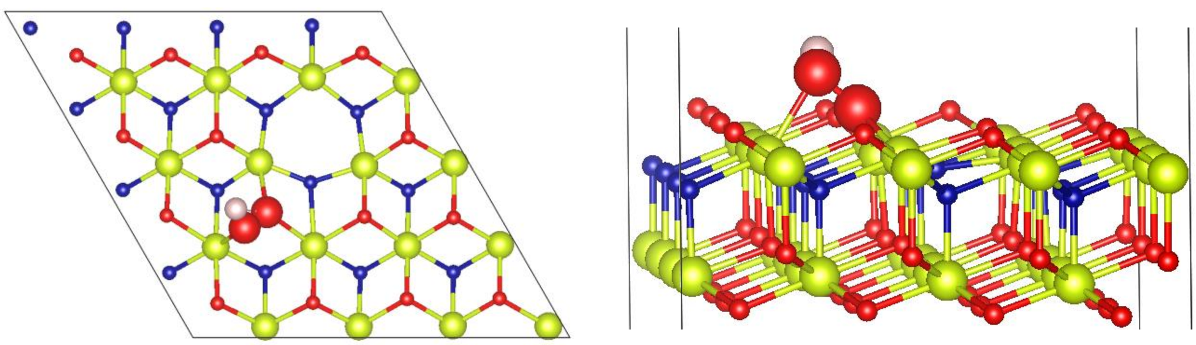
Similarly, we observed that the terminal oxygen atom of •OOH tends to centralize within the defect. This can be explained using the same reasoning as above: due to its negative polarization and high spin state, it prefers to interact with the positively polarized, high spin state cerium atoms surrounding the defect. The distance between this terminal oxygen atom and the surrounding cerium atoms ranges from 2.53 to 2.73 Å in the linear defect and from 2.40 to 2.68 Å in the triangular defect. Additionally, we observed that in both defects, the central oxygen atom of •OOH tends to settle near the one cerium atom (2.60 Å for linear defect and 2.56 Å for triangular defect), rendering the O-O bond at a tilted angle relative to the XY plane. This is because the negatively polarized central oxygen atom favors interaction with the positively polarized cerium. Finally, the H atom must be oriented upward for steric reasons, but intriguingly is tilted slightly downward to the surface in all systems. This is likely due to the interaction with the topmost, negatively polarized oxygen atom adjacent to the defect.

Population Analyses. The results from the Mulliken population analyses for charge and spin states on bare surfaces are depicted in Figures 2.4, 2.5, and 2.6. When an oxygen atom is removed from the ceria surface, its valence electrons are redistributed to the neighboring cerium atoms, affecting their partial charges as shown in Figure 2.4. This is reflected in the population analysis: cerium atoms adjacent to defects bear a less positive charge. Intriguingly, the central, subsurface oxygen atom in both systems is more negatively charged compared to surrounding oxygen atoms, suggesting that this subsurface oxygen atom shares the burden of electron redistribution with the cerium atoms. The additional electrons also affect the local spin state of the neighboring cerium atoms, as depicted in Figure 2.5. The four and six cerium atoms adjacent to the linear and triangular defects, respectively, possess a higher spin state compared to non-adjacent cerium.



Top view Side view

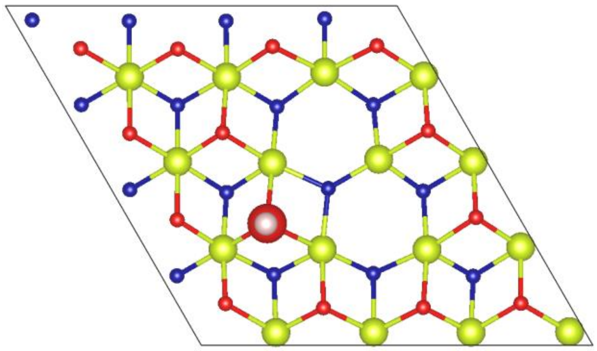
(a) •OH to Linear Defect

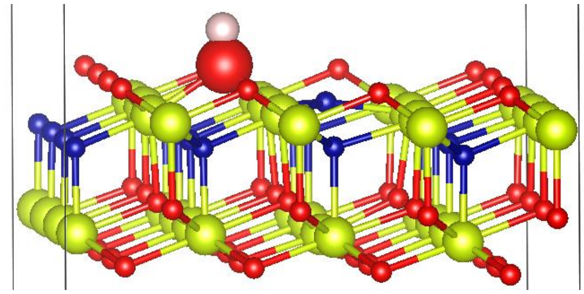


Top view Side view

(b) •OOH to Linear Defect

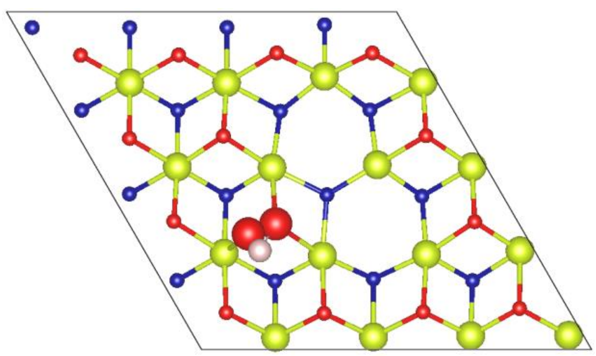
Figure 2.2 Geometry-optimized structures for (a) •OH bound on linear defect; (b) •OOH bound on linear defect. Yellow green, red, and white spheres indicate ceria, oxygen, and hydrogen atoms, respectively. The subsurface oxygen atoms are represented by blue spheres and the •OH and •OOH radicals are enlarged for better graphical view. For simplicity, the middle and bottom layers are hidden in the top views.

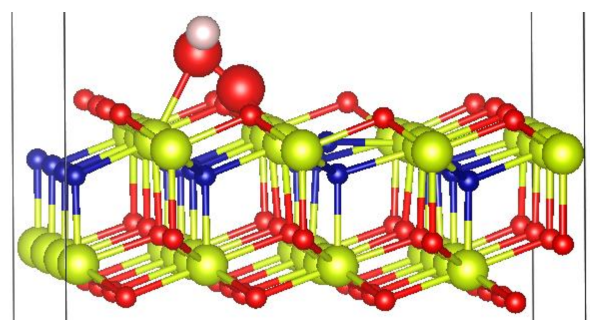




Top view Side view

(a) •OH to Triangular Defect

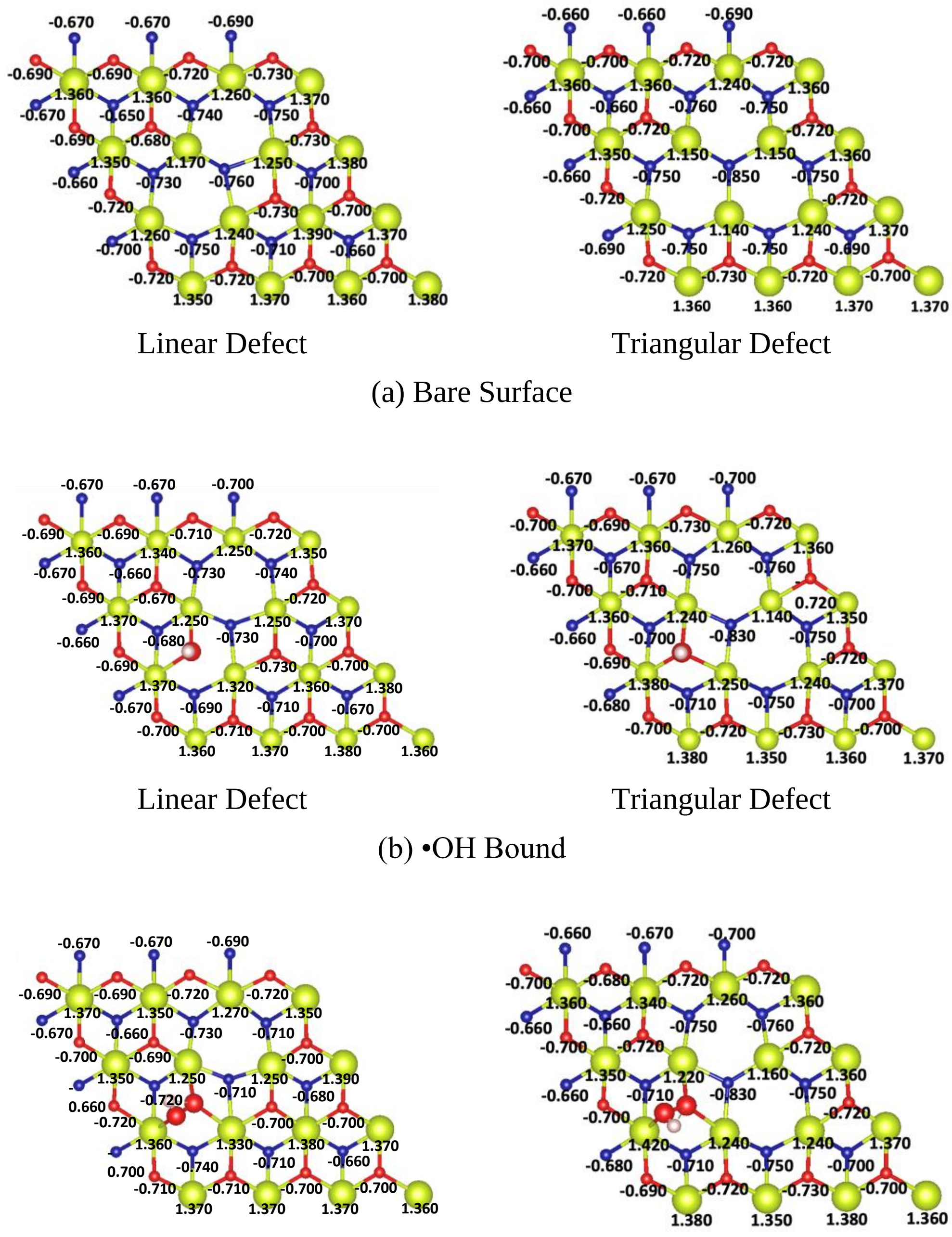




Top view Side view

(b) •OOH to Triangular Defect

Figure 2.3. Geometry-optimized structures for (a) •OH bound on triangular defect; (b) •OOH bound on triangular defect. Yellow green, red, and white spheres indicate ceria, oxygen, and hydrogen atoms, respectively. The subsurface oxygen atoms are represented in blue spheres and the •OH and •OOH radicals are enlarged for better graphical view. For simplicity, the middle and bottom layers are hidden in the top views.



Linear Defect | Triangular Defect

(a) Bare Surface

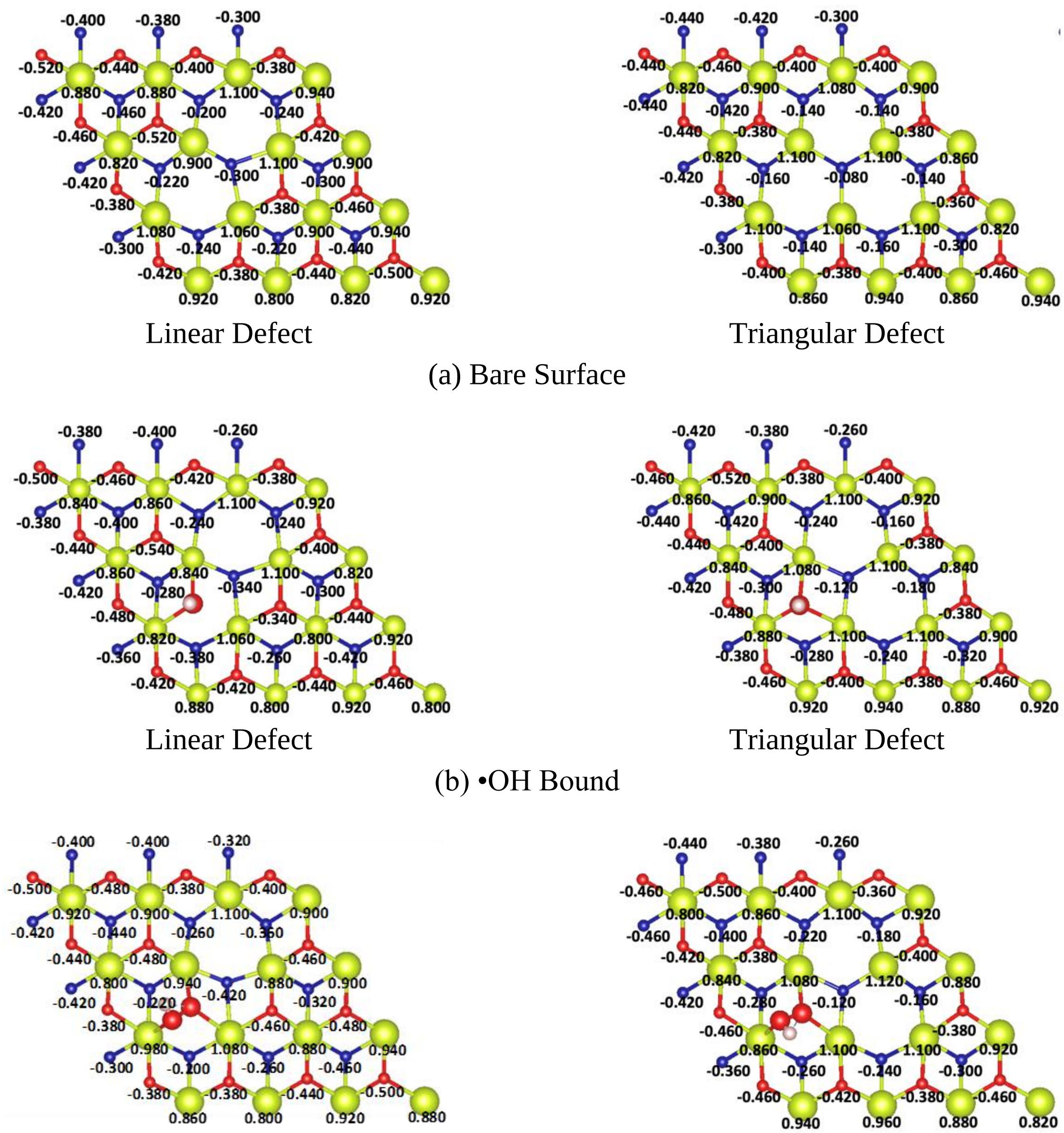
Linear Defect | Triangular Defect

(b) •OH Bound

Linear Defect | Triangular Defect

(c) •OOH Bound

Figure 2.4. Mulliken charges analysis for (a) radical-free bare surface; (b) •OH bound surface; (c) •OOH bound surfaces calculated using CASTEP. Yellow, white, red, and blue spheres indicate ceria, hydrogen, surface oxygen, and subsurface oxygen atoms, respectively.



Linear Defect Triangular Defect

(a) Bare Surface

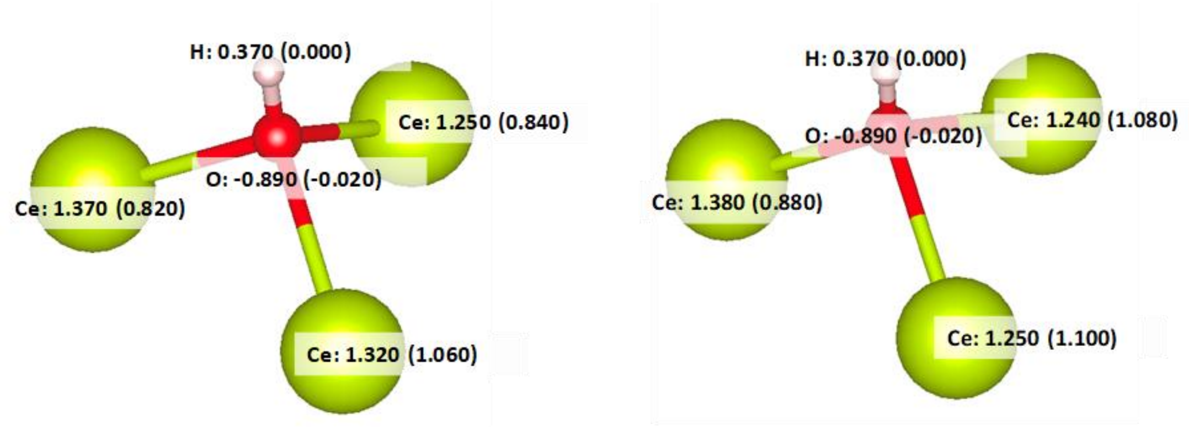
Linear Defect Triangular Defect

(b) •OH Bound

Linear Defect Triangular Defect

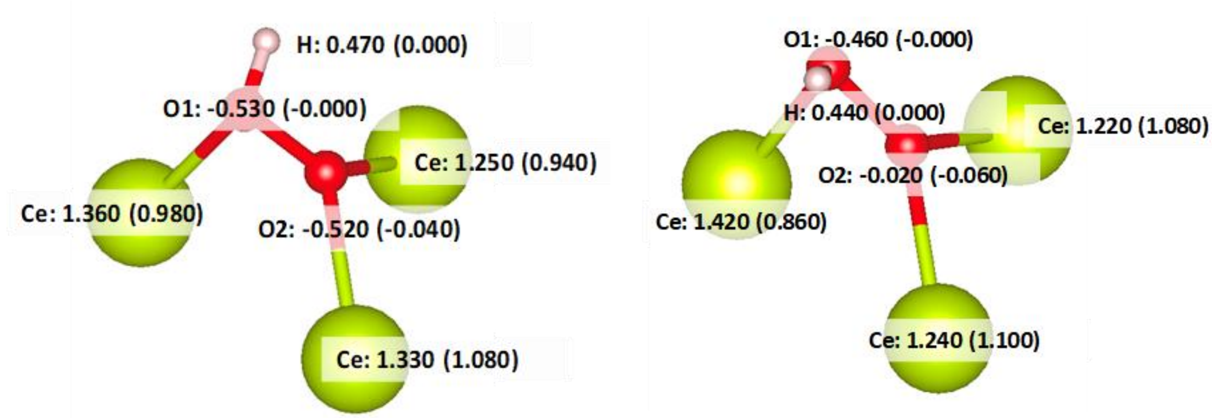
(c) •OOH Bound

Figure 2.5. Mulliken spin analysis for (a) radical-free bare surface; (b) •OH bound surface; (c) •OOH bound surfaces calculated using CASTEP. Yellow, white, red, and blue spheres indicate ceria, hydrogen, surface oxygen, and subsurface oxygen atoms, respectively.



Linear Defect Triangular Defect

(a) •OH



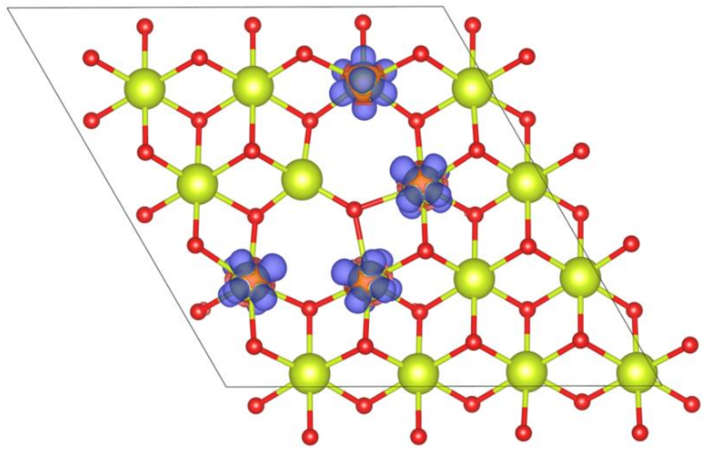
Linear Defect Triangular Defect

(b) •OOH

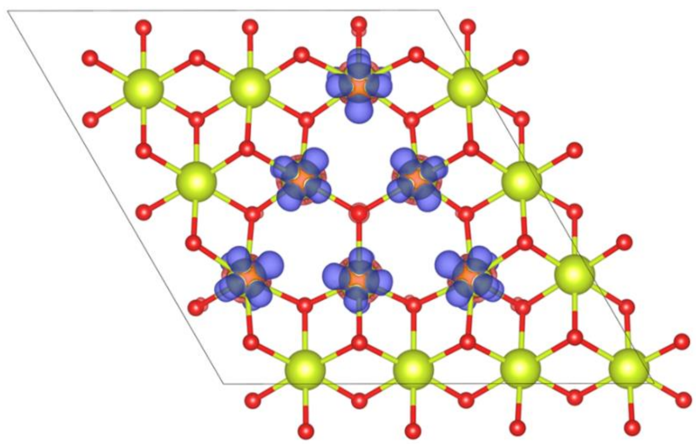
Figure 2.6. Structures of •OH and •OOH bound at the defect sites with charge values and parenthesized spin values for individual atoms. Yellow, red, and white spheres indicate ceria, oxygen, and hydrogen atoms, respectively.

Mulliken population analyses were also performed for the •OH and •OOH-bound surfaces. When a radical species is bound, the charge values in the neighboring cerium atoms become more positively polarized relative to the bare surface (Figure 2.4), while their spin values are reduced (Figure 2.5). In the triangular defect, we observed a decrease of the spin for the target cerium from 1.10 to 0.86, and neutralization of the net spin, indicating that a radical coupling is made. In the linear defect, the effect of •OOH binding on spin is more complicated. When the •OOH is bound on the target cerium, the spin of this cerium is reduced. Additionally, the spin on the cerium diagonal to the target cerium is decreased from 1.10 to 0.88. Such a complicated set of events is likely necessary due to the asymmetry of the linear defect. For further understanding of such changes in charge and spin states presented in Figures 2.4 and 2.5, we investigated the states of •OH and •OOH. Compared to a net zero charge and net 1.0 spin in the unbound state, •OH and •OOH exhibited more polarized charge values and lesser spin values in the bound state, as depicted in Figure 2.6. In other words, when the covalent bond is formed, the electronegativity difference causes more charge polarization while the electron pairing lowers the atomic spin values.

In addition, we conducted an excess spin density analysis to determine the localization of the electrons, which is presented in Figure 2.7. Consistent with the population analyses in Figures 2.4, 2.5, and 2.6, the cerium atoms adjacent to each defect exhibit a higher net spin than the other cerium atoms, meaning that they are the most probable participants in radical scavenging.



(a) Linear Defect



(b) Triangular Defect

Figure 2.7. Net spin analysis for (a) linear defect and (b) triangular defect. Yellow, red, and blue spheres indicate ceria, surface oxygen, and subsurface oxygen atoms, respectively.

Binding Energy. Next, we perform the binding energy calculations to further investigate the effect of the defect site structure on the adsorption of •OH and •OOH as follows:

$$\Delta E_{bind} = E(CeO_{1.x} + \ radical) - \ E(CeO_{1.x}) - E_{radical}$$

where ΔE_{bind} is the binding energy of the •OH or •OOH to the ceria surface, $E(CeO_{1.x} + \ radical)$ is the energy of the •OH-bound or •OOH-bound surface, $E(CeO_{1.x})$ is the energy of the non-stoichiometric defective surface, and $E_{radical}$ represents the energy of the •OH or •OOH. Atomic configurations for the binding geometries are presented in Figure 2.6, and the radical binding energies are plotted as a function of number of vacancies in Figure 2.8. The •OH binding energy is found to be consistently stronger than that of •OOH for the pristine and defective surface. It appears that, for •OH and •OOH, the binding energies to triangular defects are –4.54 eV and –3.20eV, respectively, indicating that this process is energetically more favorable than the binding to linear defects, whose energies are –4.12eV and –2.80eV, respectively. Such binding behavior can be explained through the population analyses: a linear defect presents an additional oxygen atom adjacent to the bound radical with a charge of -0.730 (Figure 2.4a), which repels the negatively polarized radical oxygen and decreases the magnitude of the binding energy (i.e. weaker binding). Additionally, the averaged spin value of the defect-adjacent cerium atoms (Figure 2.5a) is lower in the linear defect (1.05) than in the triangular defect (1.09). Since the radical oxygen possesses a high spin state and prefers to associate with other high spin state atoms, it would tend to be more attracted to the cerium atoms surrounding the triangular defect, increasing the magnitude of the binding energy for stronger adsorption. Of note, we also observed that •OH binds more strongly than •OOH. This is likely due to the Coulombic interactions: •OOH features two negatively charged oxygen atoms whereas •OH possesses only one. Therefore, •OOH has more electrostatic repulsion with the surface oxygen atoms nearby the defect, compared to •OH.

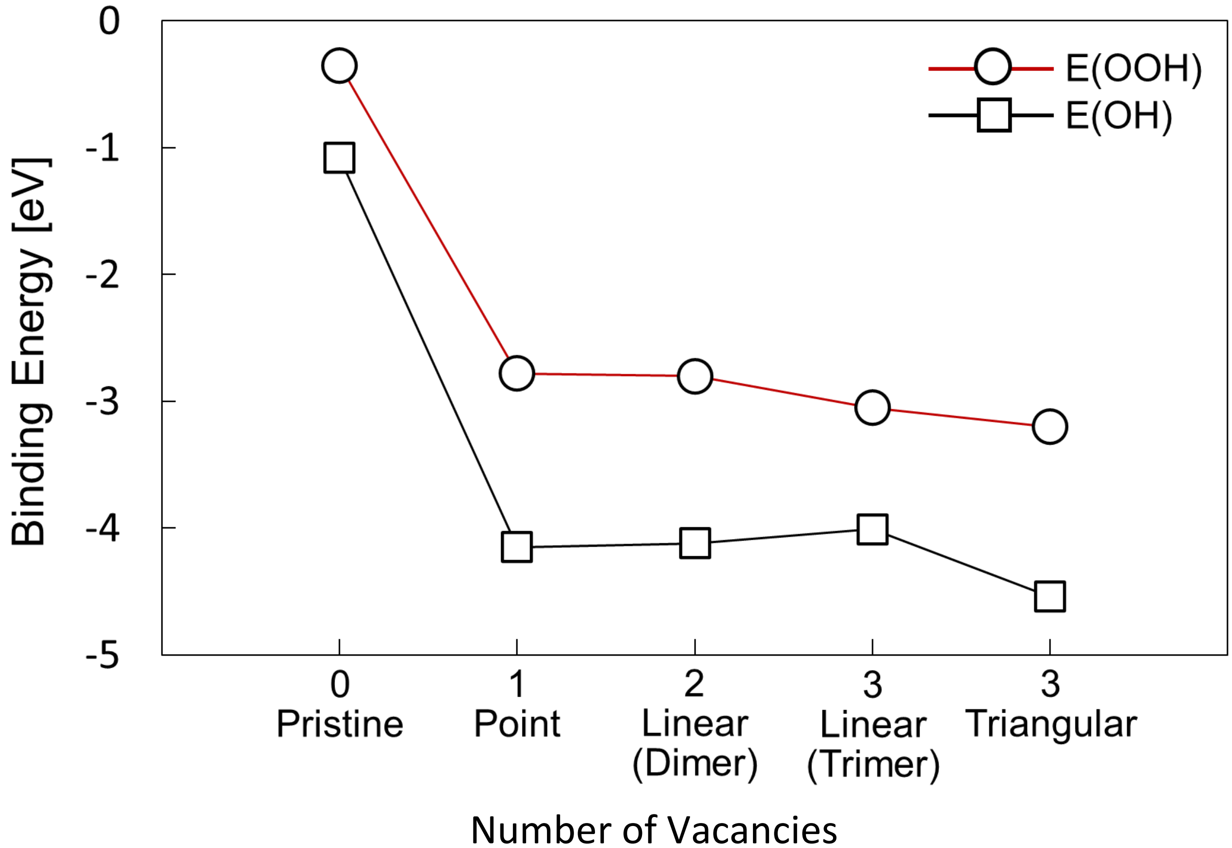
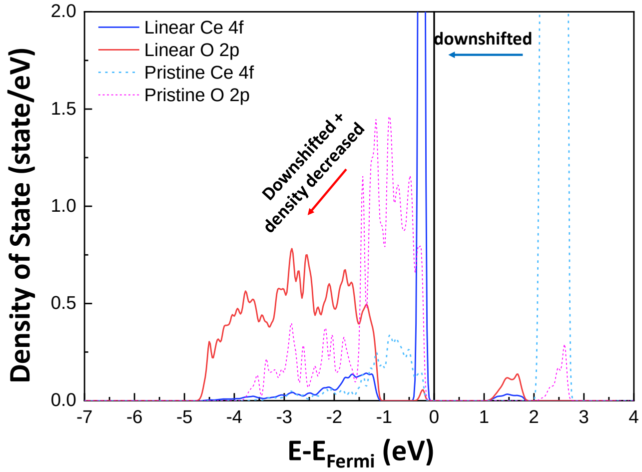


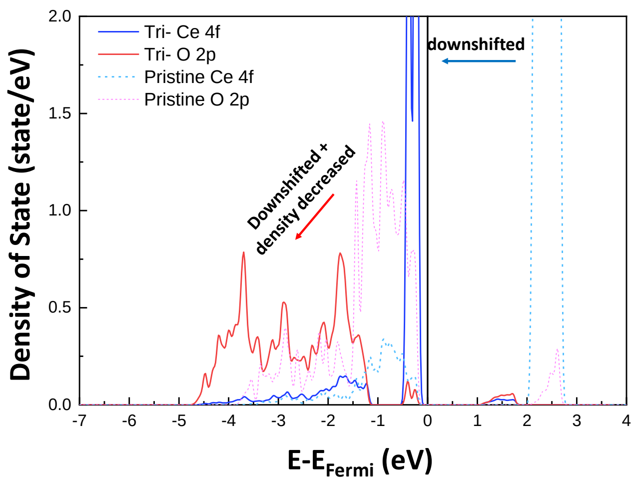
Figure 2.8. Adsorption energy plot of the •OH and •OOH on pristine ceria (111) surface, and defective surfaces including point, linear dimer, linear trimer, and triangular defects. The •OH binding energy, E(OH), is found to be stronger than that of •OOH binding, E(OOH). The strong binding tendency is consistently found through all evaluated vacancy models, even on the pristine CeO_2 surface.

Density of States Analysis. To further analyze the difference between linear and triangular defects in terms of binding energy and population (charge and spin state), we calculated the density of states (DOS) for the *f*- and *p*-bands of surface Ce and O atoms, respectively, in the defective ceria surface.

Figure 2.9 demonstrates that the energy levels of Ce 4*f*- and O 2*p*-bands of the ceria surface undergo a significant change during generation of the oxygen vacancy defects. For both linear and triangular defects, the sharp peak for the 4*f*-band of the pristine surface (indicated by the dotted sky-blue line) at 2 – 3eV is shifted down below the Fermi level. In previous studies, it is revealed that the valence band near the Fermi level, especially the high (low) peak over the energy range of –0.5 to 0 eV, is responsible for the strong (weak) kinetic interaction between the surface and adsorbates.[127-131] From this result, it is expected that, unlike the 4*f*-band of the Ce atoms on the defect-free pristine ceria surface, the 4*f*-band of Ce on the defective surface would play a crucial role in forming a bond with the radical species. The binding of radical species •OH and •OOH, in fact, is stronger in the defective ceria than in the pristine CeO_2, which is in good agreement with the abovementioned studies. We also noticed a slight downshift and decrease in the DOS of the O 2*p*-band within the same bonding region. These downshifted O 2*p*-bands are hybridized with Ce 4*f*-bands to form covalent bonds between oxygen and cerium atoms. Simultaneously, the HOMO level is decreased by 1 eV, and is therefore positioned around -1.1eV in systems with oxygen defects (around -0.1eV for pristine CeO_2).



(a)

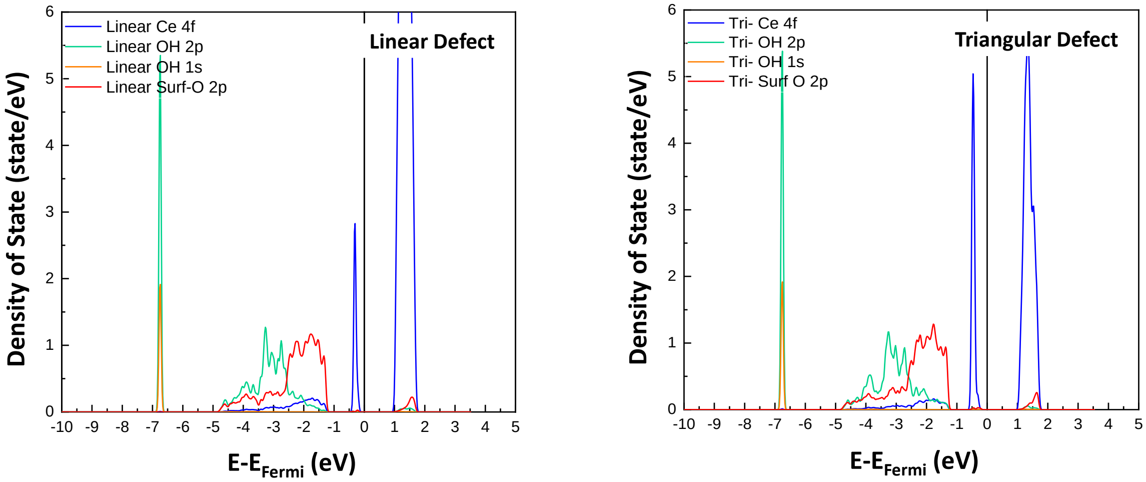


(b)

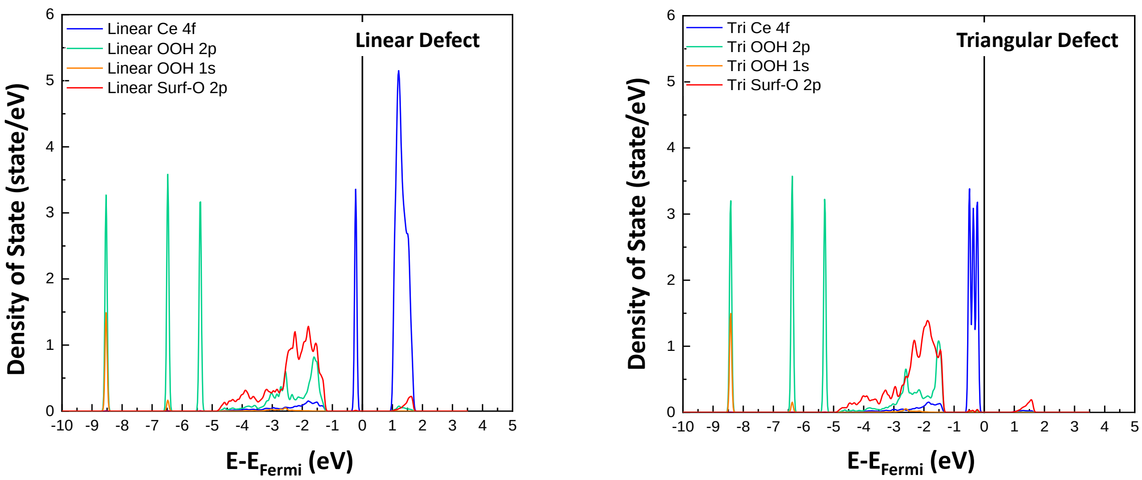
Figure 2.9. Projected density of states (PDOS) analysis of the *f*- and *p*-bands of (a) linear defective and (b) triangular defective surfaces. The vertical line at 0 eV denotes Fermi level position. The PDOS of pristine CeO_2 are shown in the dotted lines.

Moreover, we performed a PDOS analysis on the •OH- and •OOH-bound linear and triangular defect systems (Figure 2.10) to further investigate the modifications to Ce's 4*f*-bands as a function of radical scavenging. First, it is found that the 2*p*-bands of surface oxygen atoms (solid red line) are very similar regardless of the defect type, which is reasonable since those oxygen atoms are covalently connected with cerium atoms and hence do not directly participate in radical scavenging. While the 2*p*-bands of oxygen in •OH and •OOH radicals in the linear and triangular defective systems are also indistinguishable, we observed a significant difference in the density of Ce 4*f*-bands between the two defective systems. With respect to the linear defects, the density of the sharp 4*f* peak below the Fermi level in the triangular defective system is higher than that of the linear defective system, and the opposite is true above the Fermi level region. Consequentially, the binding energy of •OH is stronger in triangular defects than in linear defects as demonstrated in Figure 2.8. Additionally, in Figure 2.10b we observe a more dramatic change in the 4*f*-band of the •OOH-bound triangular defective system compared to the linear defective system. Specifically, the Ce 4*f* peak above the Fermi level is shifted down to below the Fermi level (-0.5 ~ -0eV) and multiple dense peaks are formed. These 4*f* peaks around the -0.5 ~ 0eV region in the triangular defective system are responsible for the more enhanced •OOH radical scavenging ability compared to the linear defective system.

Therefore, our PDOS calculations suggest that the oxygen vacancy causes the electronic spin localization on Ce atoms in the defect site, which induces a downshift of Ce 4*f*-band to below the Fermi level. Such a transition of the Ce 4*f*-band is key to ceria's radical scavenging capability.



(a) •OH Bound



(b) •OOH Bound

Figure 2.10. Projected density of states (PDOS) analysis for (a) •OH bound surfaces; (b) •OOH bound surfaces. The vertical line at 0 eV denotes Fermi level position.

Conclusion

CeO_2 has been established as an effective scavenger for destructive oxygen radicals in fuel cell membranes. In this study, we aimed to enhance fundamental understanding of radical binding mechanisms and discern the ideal CeO_2 surface condition for radical scavenging using DFT. In particular, we compared the structures, defect formation energies, radical binding energies, and electronic states between linear and triangular defect sites. After the addition of •OH and •OOH to the ceria surface, cerium atoms adjacent to the defect displayed increased charge and reduced spin state compared to other cerium atoms, indicating radical coupling. Additionally, by analyzing the PDOS, it is found that the defective surface with reduced cerium (Ce^{+3}) atoms exhibits a downshift of the Ce 4*f*-band located below the Fermi level and hybridizes with oxygen's 2*p*-band in systems containing bound •OH and •OOH.

Here, the binding strength of •OH and •OOH largely changes in the presence of a vacancy, whereas the pristine surface showed much weaker binding strength, especially for the •OOH of ~0.35 eV. This implies the surface migration of the radical toward preferable vacancy sites. Radical binding on vacancy sites was more exothermic in the triangular defect (-4.54 eV and -3.20 eV, respectively) than in the linear defect (-4.12 eV and -2.80eV, respectively), meaning that the triangular defect is a more effective scavenging site than the linear defect, and the •OH is adsorbed more strongly compared to •OOH. It is found that the average spin state of the defect-adjacent cerium atoms is higher in the triangular defect (1.09) than in the linear defect (1.05); likewise, the linear defect presents an additional negatively charged oxygen atom (-0.730 in Figure 2.8a) adjacent to the defect which is not present in the triangular defect. Overall, these properties cause the radical species to bind more strongly to the triangular defect compared to the linear defect. As such, considering their stability and radical scavenging capability, we suggest the use of ceria with primarily triangular defects when designing radical scavengers against •OH and •OOH to enhance the durability of polymer electrolyte membranes in fuel cells.

IV. **AIM 3.** To aid design of high performance PEMFCs by creating a tool which predicts pK_a of acids relevant to PEMs.

Introduction

Oxoacids such as sulfonic, phosphonic, and phosphoric acids are highly relevant to the performance of polymer electrolyte membranes (PEMs) in fuel cells.[49, 132-133] Studies have demonstrated that the pK_a of these acids is a crucial parameter in predicting the equilibrium structure and transport properties of these PEMs, and therefore their performance.[49, 132, 134-136] Specifically, studies have found correlations between pK_a and membrane conductivity, though the trend and magnitude of these correlations differ depending on water content.[49, 132, 135, 137] Such pK_a-conductivity correlation has been attributed to the effect of pK_a on hydrogen bond network formation as well as proton dissociation.[132, 134, 138] As such, knowing the pK_a of candidate acid groups in PEM could serve as a convenient estimation of its performance. In addition to the tedium of experimental measurement of pK_a, the high strength of some PEM–relevant oxoacids renders experimental pK_a estimation unreliable, since results are highly sensitive to small fluctuations in environmental conditions.[139-140] In this context, computational predictions could provide researchers with faster, more robust, and more accurate pK_a estimations.

Computational pK_a prediction methods can be sorted into two broad categories: (1) data–driven methods[141-150] which use regression algorithms by utilizing experimentally determined pK_a values, and (2) quantum mechanics (QM) based methods[151-152] which calculate pK_a from the free energy change during the deprotonation process in solution via a thermodynamic cycle.

Data-driven methods may include linear free energy relationships (LFERs)[153] or quantitative structure activity/property relationships (QSARs/QSPRs). LFERs employ the Hammet-Taft equation to calculate pK_a from that of a reference molecule using empirically determined constants to describe the effect of each substituent.[141, 153] This method is very popular and has been employed in several commercial software packages.[142-144] QSARs/QSPRs are one of the most common pK_a prediction techniques, and involve linear regression or least squares models fitted to experimental data.[143, 154-156] These algorithms have also been employed in a variety of commercial pK_a prediction platforms.[145-147]

While these data-driven methods can achieve impressive accuracy, their usefulness is limited to the availability and reliability of experimental data.[148-150] Due to the time-consuming nature of experiments, lack of available data for nontraditional acids (e.g. phosphorous-based acids), and experimental inaccuracy at low pK_a values (e.g. in sulfur-based acids), data-driven methods are often limited to small, well-characterized chemical spaces and are not ideal for the prediction of pK_a in acids relevant to PEMs.[148, 157] Furthermore, obtaining sufficient experimental data is usually time- and resource-consuming.

On the other hand, QM-based methods require no experimental data, and are capable of achieving desirable (<0.5 pK_a) mean absolute errors (MAEs) for multiple classes of acids.[158] However, to achieve this level of accuracy, complete basis set (CBS) methods[159-160] and inclusion of explicit solvent molecules[152, 161-162] must be used, and these methods involve laborious computational procedures.[161] Many researchers have attempted to predict pK_a using simpler basis sets and implicit solvation methods while maintaining accuracy by employing methods such as solvent cavity scaling.[163-164] However, it should be stressed that these remedies function well when optimized for a narrow range of acids, meaning that generalization to other chemistries frequently results in large error.

Recently, researchers have attempted to further enhance QSAR/QSPR models by upgrading from regression to ML algorithms such as deep neural networks, extreme gradient boosting, support vector machines, or tree-based methods,[148, 154-155] which have been utilized in commercial packages.[165-166] While ML reduces the prediction error significantly, these algorithms are still limited to the availability and accuracy of experimental data. However, the use of QM-calculated properties as ML descriptors could boost the performance of these methods over a more diverse range of acids.

In this study, we overcome the shortcomings of QM and ML by introducing a hybrid of experimentally-trained ML and mid-level DFT methods. Specifically, we combine the DFT–calculated pK_a with several other descriptors in a supervised learning algorithm trained against experimental pK_a data extracted from literature. The ML algorithm allows us to harness the precision of modern data-driven prediction techniques, while the computationally inexpensive DFT method allows us to maintain accuracy in less-characterized chemical spaces. As such, we create a pK_a prediction tool that is efficient (low demand on computational resources and time), accurate, versatile (able to predict pK_a for the large range in acid strength and structure relevant to oxoacids for fuel cells), and robust to experimental error. This tool will be invaluable in the development of novel PEMs.

Computational Methods

Calculation of pK_a using DFT. First, we used DFT to predict pK_a values for six test acid molecules including $PhCH_2PO_3H_2$, $CH_3OPO_3H_2$, CH_3COOH, CF_3COOH, CH_3SO_3H, and CF_3SO_3H whose pK_a values have been well-documented and whose values span over our range of interest (~ $-5.9 < pK_a < 4.8$).[167-172] pK_a was calculated using Jaguar.[173] The DFT computation conditions used for prediction of pK_a are tabulated with the corresponding mean absolute errors (MAEs) in Table 3.1.

In the computational procedure, the pK_a value was calculated from the free energy change ($\Delta G^o_{deprot,aq}$) over the deprotonation process using the following equation:[159, 174-177]

$$pK_a = \frac{\Delta G^o_{deprot,aq}}{2.303RT}$$

where R is the universal gas constant and T is absolute temperature. $\Delta G^o_{deprot,aq}$ is calculated as follows:

$$\Delta G^o_{deprot,aq} = \Delta G^o_{aq}(A^-) + \Delta G^o_{aq}(H^+) - \Delta G^o_{aq}(HA)$$

where $\Delta G^o_{aq}(A^-)$ and $\Delta G^o_{aq}(H^+)$ denote free energies of the deprotonated species and proton in aqueous solution, respectively, and $\Delta G^o_{aq}(HA)$ denotes free energy of the protonated species. Since $\Delta G^o_{aq}(X) = \Delta G^o_g(X) + \Delta G^o_{solv}(X)$ where X $= A^-, H^+, or\ HA$, and $\Delta G^o_g(X)$ and $\Delta G^o_{solv}(X)$ denote free energies of the gas phase species at 298.15 K and for solvation, respectively, we used a thermodynamic cycle to calculate $\Delta G^o_{deprot,aq}$ as shown in Figure 3.1.

Table 3.1. Performance of various DFT calculation conditions to predict pK_a of $PhCH_2PO_3H_2$, $CH_3OPO_3H_2$, CH_3COOH, CF_3COOH, CH_3SO_3H, and CF_3SO_3H

Functional/Basis Set for Geometry Optimization		**Solvation model**	**Mean Absolute Error (MAE)**
in Gas phase	**in Solution phase**		
B3LYP/6-31+G**	None	PBF	1.17
B3LYP/6-31++G**	None	PBF	1.19
B3LYP/6-31+G**	B3LYP/6-31+G**	PBF	1.29
B3LYP/6-31++G**	B3LYP/6-31++G**	PBF	1.18
B3LYP/6-311++G**	None	PBF	1.25
PBE0/6-31++G**	PBE0/6-31++G**	PBF	1.37
B3LYP/6-31+G**	None	SM8	1.38
B3LYP/6-31++G**	None	SM8	0.82

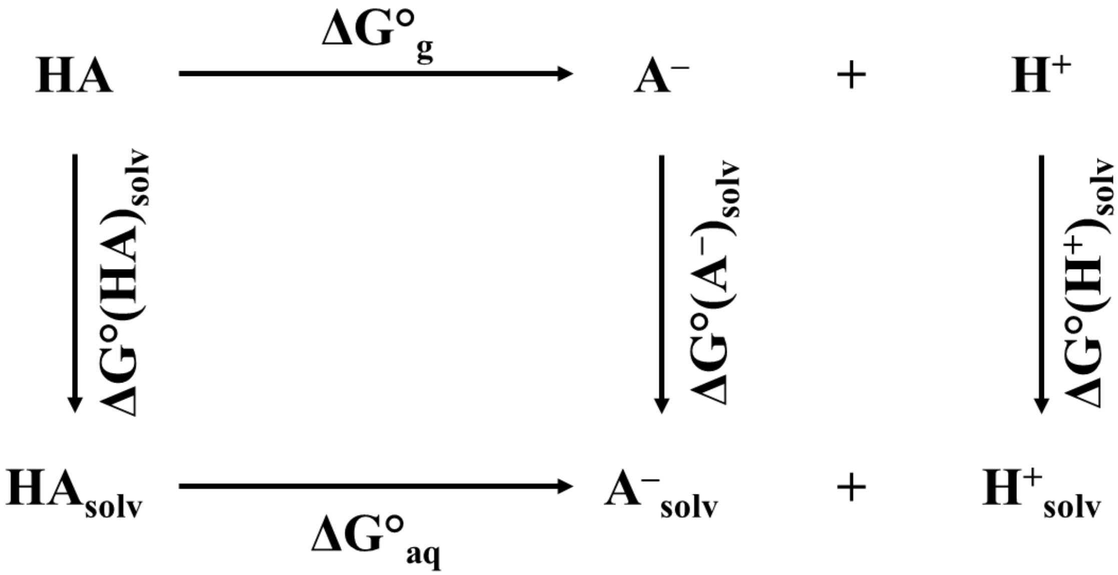


Figure 3.1. Thermodynamic cycle to calculate $\Delta G^o_{deprot,aq}$.

According to this procedure, geometry optimization of the molecular structure was performed in the gas phase prior to calculation of ΔG_g^o. Since $\Delta G_{solv}^o(H^+)$ has varied significantly in experimental reports (from –254 to –264 kcal/mol),[175, 178-179] its value was adjusted to minimize prediction error,[180] and its final value fell between –253.12 and –259.82 kcal/mol depending on the level of theory. After calculating pK_a values for all six test acids, the level of theory yielding the lowest MAE was chosen for future DFT predictions, where MAE is defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - f(x_i)|$$

where $f(x_i)$ is the predicted pK_a value, y_i is the literature pK_a value, and n = 6 for our test set. If multiple pK_a values were reported for a given acid, the highest and lowest values were set as bounds and error was calculated as deviation from these bounds.

To thoroughly represent the pK_a of a variety of acids relevant to fuel cell membranes, a set of 80 oxoacids including sulfur (12), phosphorous (28), and carboxylic (41) acids was prepared to train our model. The full list of acids, their descriptors, and literature references for experimental values are tabulated in Table 3.2. Once we chose a level of theory to predict pK_a values of test acid molecules with high accuracy, we calculated the pK_a values for all the acids in Table S1 using the same level of theory. $\Delta G_{solv}^o(H^+)$ was reoptimized to minimize DFT prediction error for this expanded set of acids, ultimately remaining within the reported literature range at –257.29 kcal/mol. Finally, these DFT-predicted pK_a values were used as descriptors for the ML models.

Development of Machine Learning Models. In order to develop machine learning models for predicting pK_a with high accuracy, the following 10 properties were selected as descriptors (Table 3.3): electronegativity of the central atom (i.e. of P, S, or C) in the acid molecule, the magnitude of the molecule's dipole moment, degree of oxidation of the central atom, number of hydrogens (in the entire completely deprotonated molecule), number of fluorine atoms, number of carbons, molecular weight of the neutral molecule, Connolly volume of the neutral molecule, solvation free energy (SFE), and DFT-calculated pK_a. SFE and DFT-calculated pK_a were both chosen due to their presumed correlation with the true pK_a values for acid molecules (refer to Figure 3.1).

Electronegativity of the central atom, dipole magnitude, and number of fluorine atoms were chosen due to their contribution toward the electron withdrawing capacity of the acid, suggesting stability of the deprotonated state. Size-related descriptors including Connolly volume, molecular weight, and number of carbon atoms were selected due to their hypothesized correlation with positive induction of the substituent group, a suggested predictor of pK_a.[181]

To confirm that the selected descriptors were not highly correlated with one another, we performed a Pearson correlation analysis on descriptors from all acids (novel and training sets) to scale the level of correlation for each pair of features. Results for this analysis are depicted in Figure 3.2.

Table 3.2. List of all acids used in this study, their descriptors, and references (where relevant) for literature pK_a validation

Formula	1	2	3	4	5	6	7	8	9	10	Ref.
CF3COOH	2.55	2	0	3	73.6	114.0	2.9	1	-3.2	-1.92	182-185
CH3COOH	2.55	2	3	0	56.3	60.1	4.6	1	-11.0	3.25	167, 185-186
HCO2H	2.55	2	1	0	39.9	46.0	4.9	0	-8.8	1.66	185, 187
(CH3)2CHCO2H	2.55	2	7	0	91.5	88.1	1.6	3	-6.4	4.21	185
(CH3)3CCO2H	2.55	2	9	0	110.6	102.1	1.6	4	-6.5	2.93	185
CF3CH2CH2CO2H	2.55	2	4	3	110.5	142.1	3.5	3	-12.0	7.55	182, 184-185
CF3CF2CF2CO2H	2.55	2	0	7	129.9	214.0	2.8	3	-0.8	0.61	185
CH3CH2CH2CO2H	2.55	2	7	0	93.9	88.1	2.6	3	-6.1	5.45	185
CF3CH2CH2CH2CO2H	2.55	2	6	3	130.5	156.1	4.5	4	-12.6	7.96	188
CH2=CHCO2H	2.55	2	3	0	66.9	72.1	3.4	2	-7.3	3.01	185
CH2=CFCO2H	2.55	2	2	1	73.3	90.1	4.6	2	-8.8	3.86	185
CF2=CHCO2H	2.55	2	1	2	79.7	108.0	1.2	2	-5.6	7.70	185
CF2=CFCO2H	2.55	2	0	3	83.0	126.0	3.0	2	-4.4	5.56	185
CH2=CHCH2CO2H	2.55	2	5	0	87.8	86.1	2.2	3	-6.9	5.14	185
CH2CH=CHCO2H (trans)	2.55	2	4	0	84.8	85.1	3.2	3	-8.0	4.45	185
(CH3)2C=CHCO2H	2.55	2	7	0	101.8	100.1	2.1	4	-8.7	4.57	185
cyclopropyl-carboxylic acid	2.55	2	5	0	81.6	86.1	1.7	3	-7.9	5.79	185
cyclobutyl-carboxylic acid	2.55	2	7	0	99.4	100.1	2.3	4	-6.9	5.84	185
cyclopentyl-carboxylic acid	2.55	2	9	0	115.7	114.1	5.6	5	-9.3	4.57	185
cyclohexyl-carboxylic acid	2.55	2	11	0	134.7	128.2	6.3	6	-8.2	3.18	185
1-methyl-cyclohexyl-carboxylic acid	2.55	2	13	0	151.0	142.2	2.6	7	-6.0	4.39	185
2-methyl-cyclohexyl-carboxylic acid (trans)	2.55	2	13	0	150.6	142.2	6.2	7	-7.4	1.75	185
2-methyl-cyclohexyl-carboxylic acid (cis)	2.55	2	13	0	153.5	142.2	2.6	7	-5.1	5.33	185
3-methyl-cyclohexyl-carboxylic acid (trans)	2.55	2	13	0	152.4	142.2	6.3	7	-7.5	3.38	185
3-methyl-cyclohexyl-carboxylic acid (cis)	2.55	2	13	0	153.8	142.2	5.7	7	-7.7	2.00	185
4-methyl-cyclohexyl-carboxylic acid (trans)	2.55	2	13	0	155.0	142.2	2.9	7	-5.0	5.09	185
4-methyl-cyclohexyl-carboxylic acid (cis)	2.55	2	13	0	157.0	142.2	2.4	7	-5.2	6.30	185
Cyclohexyl-ethanoic acid	2.55	2	13	0	152.0	142.2	5.9	7	-8.4	0.28	185
Phenyl-CO2H	2.55	2	5	0	105.8	122.1	5.1	6	-12.2	0.87	185
o-Methylbenzoic acid	2.55	2	7	0	124.2	136.2	5.0	7	-12.5	1.78	185
m-Methylbenzoic acid	2.55	2	7	0	129.0	136.2	5.5	7	-12.6	0.86	185
o-Fluorophenyl-CO2H	2.55	2	4	1	112.3	140.1	3.5	6	-12.0	3.24	185
m-Fluorophenyl-CO2H	2.55	2	4	1	115.7	140.1	3.0	6	-9.8	5.84	185
p-Fluorophenyl-CO2H	2.55	2	4	1	119.6	140.1	1.4	6	-9.6	6.56	185

CF2H-CO2H	2.55	2	1	2	67.9	96.0	3.0	1	-7.3	1.15	183-184
CH2F-CO2H	2.55	2	2	1	60.8	78.0	3.5	1	-9.7	2.98	184-185, 188
CH2CH3-CO2H	2.55	2	5	0	74.4	74.1	5.5	2	-10.2	3.59	184-185
CH2CF3-CO2H	2.55	2	2	3	92.3	128.1	3.0	2	-9.3	4.41	182, 184-185
p-Methylbenzoic acid	2.55	2	7	0	127.6	136.2	2.6	7	-10.6	3.74	185, 189
CF3(CF2)2(CH2)2CO2H	2.55	2	4	7	168.7	242.1	2.4	5	-5.6	9.75	185
(CH3CH2)2HCO2H	2.55	2	11	0	131.4	104.2	4.9	4	-7.9	0.18	185
benzenesulfinic acid	2.58	2	5	0	115.6	142.2	4.1	6	-20.9	2.70	190
p-methyl-benzenesulfinic acid	2.58	2	7	0	133.7	156.2	3.5	7	-24.7	4.28	191
p-chloro-benzenesulfinic acid	2.58	2	4	0	134.6	176.6	2.5	6	-20.4	1.13	190-191
p-bromo-benzenesulfinic acid	2.58	2	4	0	137.4	221.1	2.5	6	-36.8	-6.23	190-191
p-nitro-benzenesulfinic acid	2.58	2	4	0	143.3	187.2	1.9	6	-23.5	3.95	190-191
p-methoxybenzenesulfinic acid	2.58	2	7	0	145.2	172.2	5.6	7	-25.8	3.44	191
m-chloro-benzenesulfinic acid	2.58	2	4	0	134.1	176.6	2.8	6	-20.0	0.55	191
m-nitro-benzenesulfinic acid	2.58	2	4	0	145.8	187.2	3.0	6	-22.8	2.84	190-191
CF3SO3H	2.58	3	0	3	90.2	150.1	2.2	1	-13.0	-12.01	172, 192
CH3SO3H	2.58	3	3	0	72.1	96.1	3.0	1	-22.1	-4.55	193
p-Toluenesulfonic acid	2.58	3	7	0	144.3	172.2	5.7	7	-21.4	-5.43	193
benzenesulfonic acid	2.58	3	5	0	124.7	158.2	3.9	6	-20.7	-5.73	185
sulfuric acid	2.58	4	0	0	63.3	98.1	3.4	0	-21.0	-7.16	192
Methylphosphinic acid	2.19	2	3	0	68.7	79.0	3.9	1	-14.8	5.05	185
Ethylphosphinic acid	2.19	2	5	0	87.6	93.0	4.1	2	-14.0	6.01	185
N-Propylphosphinic acid	2.19	2	7	0	106.5	107.1	3.8	3	-13.6	6.07	185
Isopropylphosphinic acid	2.19	2	7	0	105.6	107.1	4.0	3	-12.8	5.60	185
N-Butylphosphinic acid	2.19	2	9	0	122.3	121.1	4.1	4	-13.5	6.38	185
tert-Butylphosphinic acid	2.19	2	9	0	123.0	121.1	3.9	4	-11.6	3.92	185
Phenylphosphinic acid	2.19	2	5	0	120.4	141.1	3.2	6	-17.6	5.65	185
CF3PO3H2	2.19	3	0	3	95.2	150.0	1.6	1	-9.4	0.52	187
CH3PO3H2	2.19	3	3	0	75.3	96.0	1.4	1	-11.9	5.26	185, 187
CH2CH3-PO3H2	2.19	3	5	0	94.2	110.1	2.4	2	-11.1	5.08	185, 194
CH3CH2CH2PO3H2	2.19	3	7	0	112.4	124.1	4.7	3	-11.8	3.43	185, 194
CH3CH2CH2CH2PO3H2	2.19	3	9	0	128.9	138.1	2.3	4	-10.6	5.06	185, 194
Isopropyl-PO3H2	2.19	3	7	0	111.0	124.1	2.2	3	-10.0	3.98	185, 194
N-Butyl-2-Phosphonate	2.19	3	9	0	134.9	138.1	2.2	4	-9.5	3.64	185
Isobutylphosphonate	2.19	3	9	0	131.0	138.1	2.4	4	-10.3	4.31	185
Tert-butylphosphonate	2.19	3	9	0	128.6	138.1	1.4	4	-9.4	3.40	185

2,2'-dimethyl-n-propyl-phosphonate	2.19	3	11	0	145.2	152.1	2.4	5	-9.9	4.23	185
Tert-amyl-phosphonate	2.19	3	11	0	147.9	152.1	2.2	5	-8.7	3.29	185
PhPO3H2	2.19	3	5	0	127.1	158.1	2.4	6	-13.8	2.89	185, 187-188
o-Methyl-PhPO3H2	2.19	3	7	0	147.1	172.1	1.5	7	-14.1	2.19	185, 188
m-Methyl-PhPO3H2	2.19	3	7	0	145.7	172.1	2.4	7	-14.3	2.28	185, 188
p-Methyl-PhPO3H2	2.19	3	7	0	147.0	172.1	2.9	7	-14.4	2.08	185, 188
PhCH2PO3H2	2.19	3	7	0	147.2	172.1	1.5	7	-15.6	3.51	187
CH3OPO3H2	2.19	4	3	0	81.7	112.0	0.8	1	-12.3	1.21	195
CH2CH3-OPO3H2	2.19	4	5	0	102.0	126.1	3.4	2	-12.8	1.39	195
CH3CH2CH2OPO3H2	2.19	4	7	0	123.2	140.1	3.8	3	-10.5	-0.93	195
CH3CH2CH2CH2OPO3H2	2.19	4	9	0	138.6	154.1	4.1	4	-10.5	-0.46	195
o-fluorophenyl-phosphonate	2.19	3	4	1	135.2	176.1	3.0	6	-14.3	2.17	185

Table 3.3. Descriptors used in initial machine learning models

Descriptors	Feature ID (for reference)
Functional group electronegativity	1
Degree of oxidation	2
Number of hydrogen atoms in the deprotonated state	3
Number of fluorine atoms	4
Connolly volume	5
Molecular weight	6
Dipole magnitude	7
Number of carbon atoms	8
Solvation free energy (DFT)	9
pK_a (DFT)	10

	1	2	3	4	5	6	7	8	9	10
1	1.0									
2	-0.4	1.0								
3	-0.1	-0.3	1.0							
4	0.1	0.2	-0.6	1.0						
5	-0.1	-0.1	0.6	0.0	1.0					
6	-0.1	0.3	-0.1	0.5	0.7	1.0				
7	0.2	-0.2	0.2	-0.2	0.0	-0.2	1.0			
8	0.1	-0.4	0.7	-0.4	0.8	0.3	0.1	1.0		
9	-0.1	-0.3	0.1	0.2	-0.1	-0.2	-0.2	-0.1	1.0	
10	-0.3	-0.5	0.4	-0.3	0.2	-0.2	-0.1	0.3	0.3	1.0

Figure 3.2. Pearson correlation analysis for primary features used in this study. Redness indicates higher (positive or negative) correlation.

Most descriptor pairs have low correlation, with the exception of Connolly volume with number of carbons (0.81). However, since we employ feature engineering techniques to reduce the presence of correlated features in most ***Pipelines***, we therefore used all proposed descriptors in our initial ML models. Data for model training and model validation were extracted from literature. Although most of data were primarily from experimental studies, the pK_a values of strong acids such as CF_3SO_3H were estimated from reports employing several distinct, rigorous modeling methods. If multiple pK_a values were reported, the highest and lowest reported values were averaged to obtain the validation value.[141, 167-171, 185, 196-206] We used an 85% training/15% testing ratio and at least 50 shuffles for each condition to ensure that the mean absolute error (MAE) was representative.

Several methods are available to optimize ML algorithms, including 1) ML algorithm selection, 2) hyperparameter tuning, and 3) feature engineering. Due to the multitude of available ML algorithms with varying prediction capabilities, it is often beneficial to test the performance of multiple algorithms. Hyperparameter tuning is a necessary step which involves the optimization of any model parameters apart from the set of weights and coefficients which depend directly on data. Feature engineering involves the creation of a set of features which minimizes prediction error, using methods such as feature elimination, transformation, or compositing. These techniques are typically integrated into a sequence known as a ***Pipeline.*** Notably, a more complex ***Pipeline*** does not necessarily guarantee a more accurate prediction, so it is appropriate to test several ***Pipelines*** to ensure that an optimal sequence is chosen. In this study, we tested three ***Pipelines*** of increasing complexity (Figure 3.3). ***Pipeline 1*** involves only hyperparameter optimization and model selection, whereas ***Pipelines 2 – 3*** add feature selection and transformation, respectively.

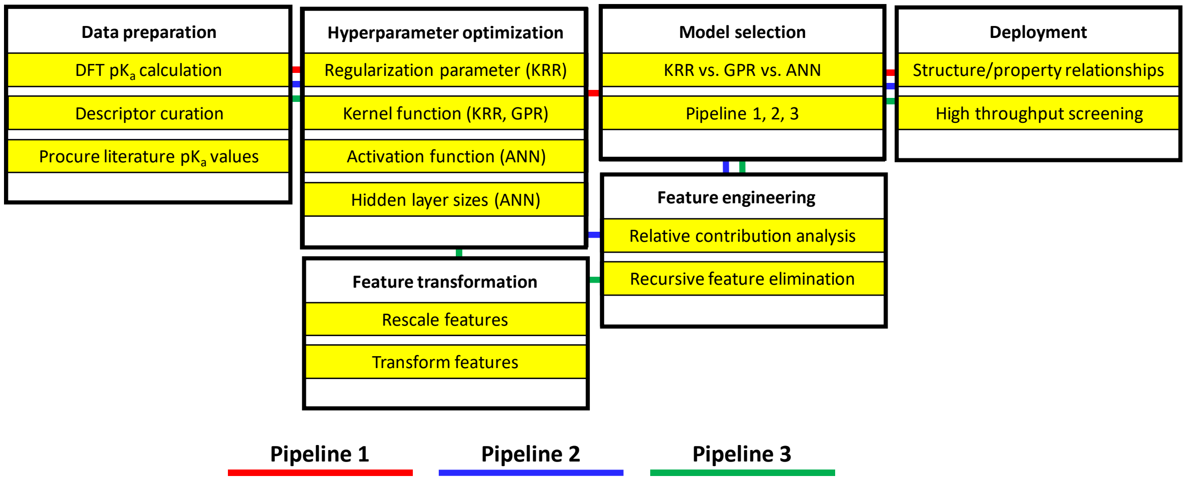


Figure 3.3. Conceptualization of machine learning ***Pipelines*** tested in this study.

In ***Pipeline 1***, we selected the best of three candidate models: KRR, GPR, and ANN. Within each model, hyperparameters are chosen to yield the lowest MAE. For KRR, the regularization parameter (λ) and the kernel were simultaneously optimized. We screened 50 values for λ scaled logarithmically from 1×10^{-5} to 1×10^{0}. Kernels screened for KRR included linear, polynomial, radial basis, sigmoid, and Laplacian functions.[33-34] Covariance functions screened in this study for GPR included RBF, Matérn, and rational quadratic. Additionally, constant scaling of each covariance function, the sum of each covariance function with a white kernel, and sum combinations of each pair of covariance functions were screened. Other relevant hyperparameters are optimized prior to each fitting by maximizing the log-marginal likelihood function as described in previous literature.[32]

Our ANN consists of an input layer (for descriptors), an output layer represented by a single neuron (for pK_a), and 2 hidden layers. Hyperparameters optimized in this study included the number of nodes per hidden layer (1 through 10) and the activation function. In ***Pipeline 2***, this tuning is followed by a relative contribution analysis (RCA) and recursive feature elimination (RFE). RCA and RFE enhance ML models by 1) reducing the number of input features to streamline curation/training, and 2) potentially improving prediction accuracy by eliminating highly correlated features. In RCA, the initial features are removed individually and the resultant MAE is recorded. The features are then ranked in an "order of importance" depending on their level of impact on the MAE. Finally, an RFE is performed wherein features are removed cumulatively in their order of importance. A final, dimensionally reduced model is chosen based on which set of features yields the best performance during RFE.

Finally, in ***Pipeline 3***, the 10 primary features are supplemented with 40 transformed features prior to RCA/RFE. Feature transformations can improve accuracy of ML models by reducing skewness, equalizing spread, and increasing linearity of the input data.[207-208] Prior to transformation, features were normalized (to mean 0 and variance 1) in order to avoid log errors and high-magnitude input values. Then, features were transformed as follows: features 11 – 20 are X^2, features 21 – 30 are $|X|^{1/2}$, features 31 – 40 are $\log(|X| + 1)$, and features 41 – 50 are $\exp(X)$, where X represents the normalization of features 1 – 10. For example, feature 22 is (absolute value of normalized feature 2)$^{1/2}$. RCA/RFE was then performed to reduce the dimensionality of these models.

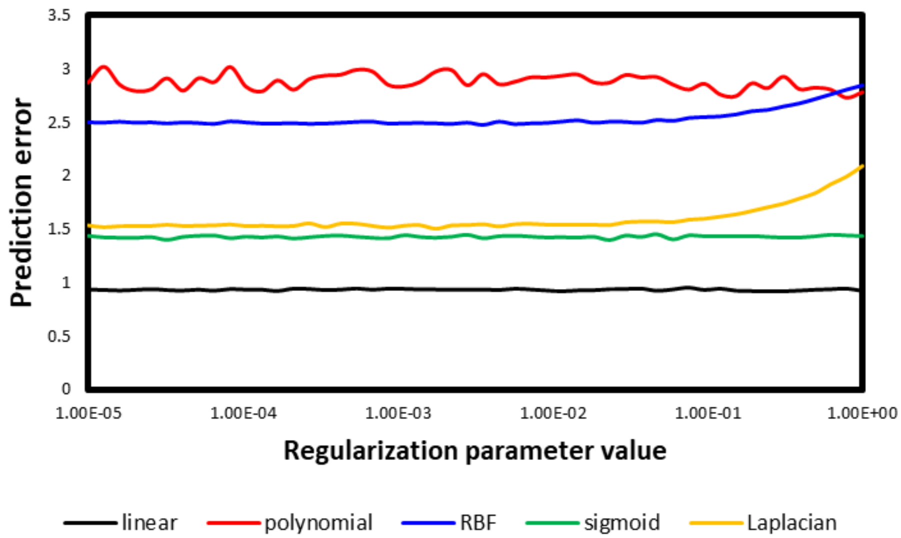
Results and Discussion

DFT Calculation for pK_a. As summarized in Table 3.1, B3LYP[209] functional performed well compared with PBE0,[210-211] and that 6-31++G** provided comparable accuracy to 6-311++G**. On the other hand, it seems that the inclusion of a solution phase geometry optimization did not improve the accuracy significantly. Regarding implicit solvation models, the Solvation Model 8 (SM8)[212] offered superior performance compared to the Poisson-Boltzmann continuum solvation model (PBF).[213-214] As such, we used the condition consisting of B3LYP, 6-31++G**, and SM8 with no solution phase geometry optimization to generate pK_a and SFE descriptors.

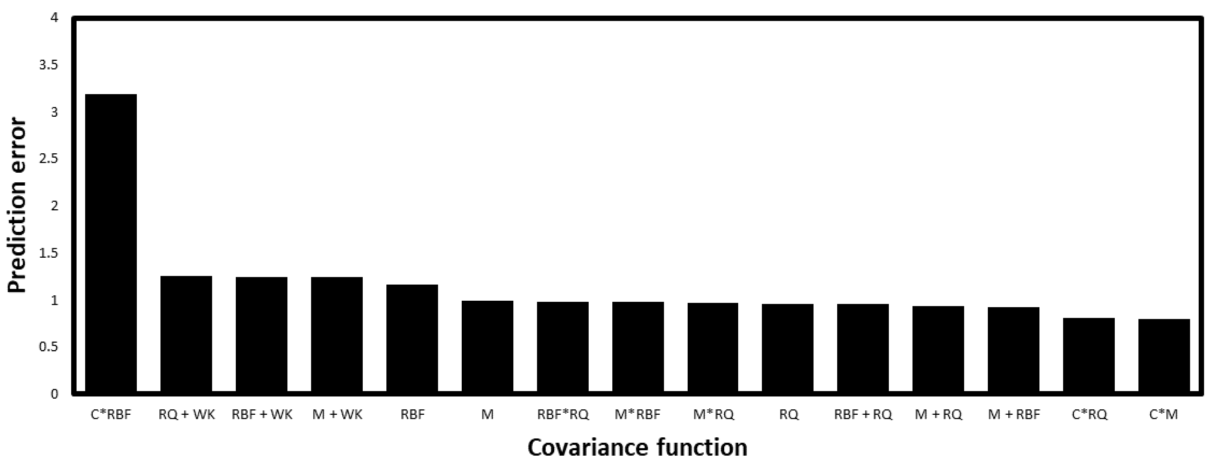
Hyperparameter Optimizations. It is found that for KRR, the linear kernel offers the best performance (Figure 3.4). Additionally, the value of the regularization parameter (λ) does not seem to significantly affect MAE. As such, $\lambda = 1.0$ was chosen arbitrarily. For GPR, the product between a constant kernel and the Matérn covariance function yielded the lowest MAE (Figure 3.4). Of the three ANN activation functions, logistic sigmoid yielded the lowest average error (0.94 over all hidden layer configurations) followed by hyperbolic tangent (1.08) and ReLu (1.43). The performance of ANN as a function of the hidden layer configuration with logistic sigmoid is summarized in Figure 3.5, wherein rows correspond to the number of neurons in the 1st hidden layer, columns correspond to the number of neurons in the 2nd hidden layer, greenness indicates a lower MAE, and redness indicates a higher MAE. The optimal configuration consisted of two hidden layers wherein 10 nodes and 4 nodes are equipped in the first and second hidden layers, respectively as presented in Figure 3.5.

Relative Contribution Analysis (RCA) and Recursive Feature Elimination (RFE). Results for the RCA and RFE analyses are summarized in Figures 3.5 and 3.6, respectively. Final descriptors were chosen based on which set yielded the lowest MAE during RFE. In other words, during RFE, features were removed in the order of importance (as determined during RCA) until we obtained the model with the lowest MAE. Figure 3.6 shows that, in ***Pipeline 2***, the optimal KRR and GPR models use 8 features: (3, 8, 4, 9, 7, 2, 1, 10) and (3, 2, 4, 8, 7, 9, 5, 10), respectively, and ANN uses 9 features (3, 9, 2, 7, 8, 6, 1, 4, 10).

On the other hand, Figure 3.7 shows that, in ***Pipeline 3***, KRR uses the final 28 features (15 through 20 in order of significance), GPR uses the final 37 features (1 through 10), and ANN uses the final 24 features (38 through 5).



(a)

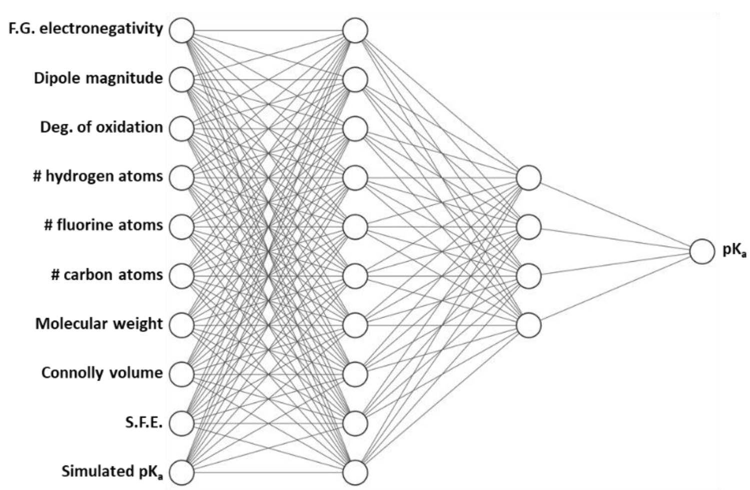


(b)

Figure 3.4. Variation of mean absolute error as a function of (a) KRR regularization parameter and kernel, and (b) GPR covariance function, where "C" signifies a constant, "RQ" signifies the rational quadratic covariance function, "WK" signifies a white kernel, and "M" signifies the Matérn covariance function.

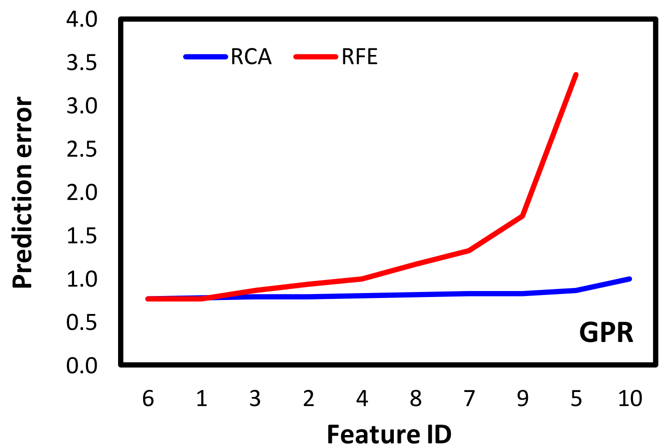
# of nodes in 1st hidden layer \ # of nodes in 2nd hidden layer	1	2	3	4	5	6	7	8	9	10
1	1.29	1.25	1.27	1.23	1.22	1.22	1.22	1.26	1.24	1.26
2	1.11	1.12	1.10	1.10	1.04	1.12	1.11	1.05	1.08	1.10
3	1.03	1.00	0.97	0.98	0.96	0.96	1.00	0.95	0.99	0.97
4	1.00	0.94	0.94	0.91	0.93	0.93	0.92	0.87	0.90	0.90
5	0.97	0.91	0.89	0.83	0.85	0.87	0.88	0.87	0.88	0.85
6	1.02	0.88	0.89	0.87	0.81	0.85	0.86	0.86	0.84	0.83
7	0.99	0.87	0.83	0.82	0.86	0.83	0.81	0.80	0.86	0.86
8	1.03	0.85	0.82	0.85	0.81	0.83	0.81	0.82	0.85	0.80
9	0.93	0.85	0.83	0.81	0.81	0.82	0.81	0.83	0.80	0.83
10	0.97	0.89	0.83	0.79	0.79	0.82	0.85	0.80	0.81	0.79

(a)

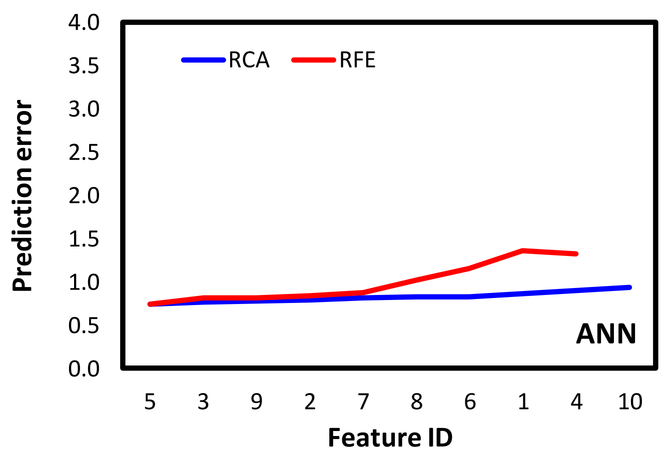


(b)

Figure 3.5. a) MAEs of several ANN hidden layer configurations (with logistic sigmoid activation function); b) Illustration of optimal neural network configuration.

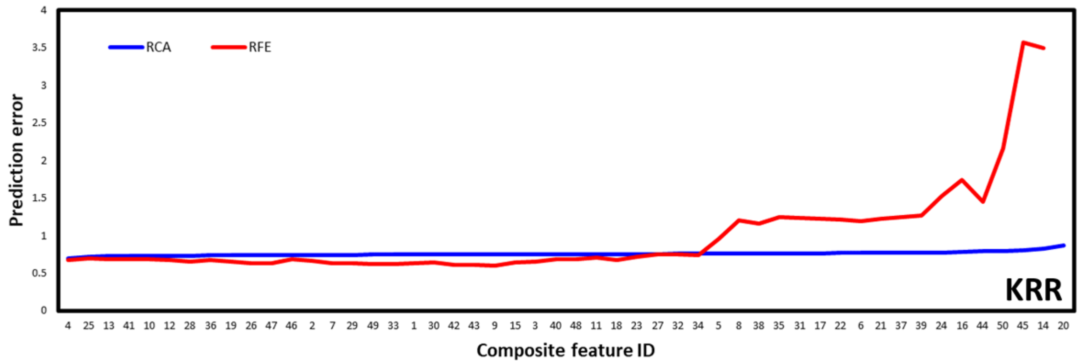


(b)

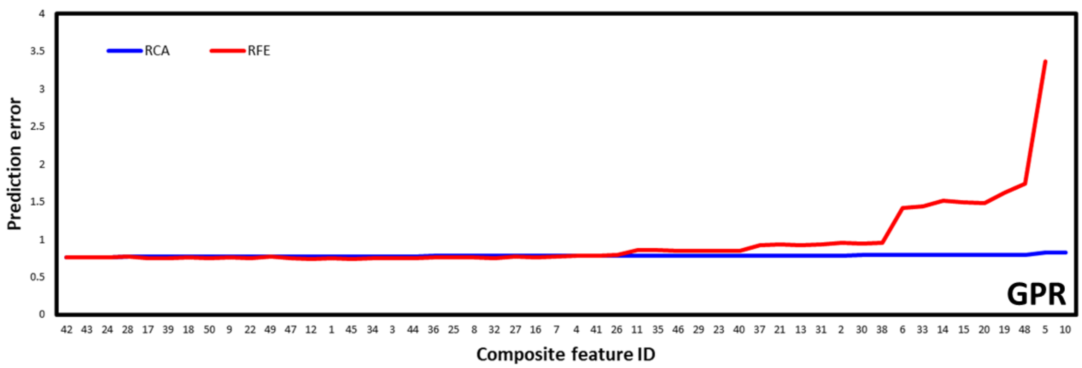


(c)

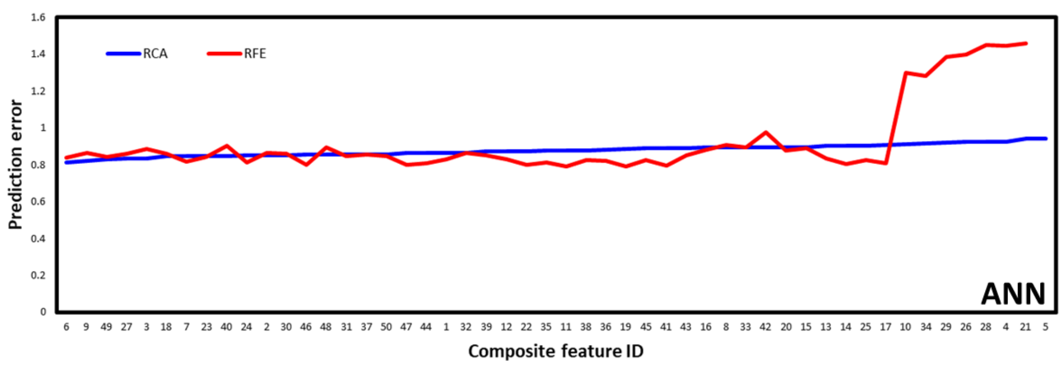
Figure 3.6. Relative contribution analysis (RCA) and recursive feature elimination analysis (RFE) in ***Pipeline 2***: a) KRR, b) GPR, and c) ANN. Prediction error denotes MAE by removing the feature from training.



(a)



(b)



(c)

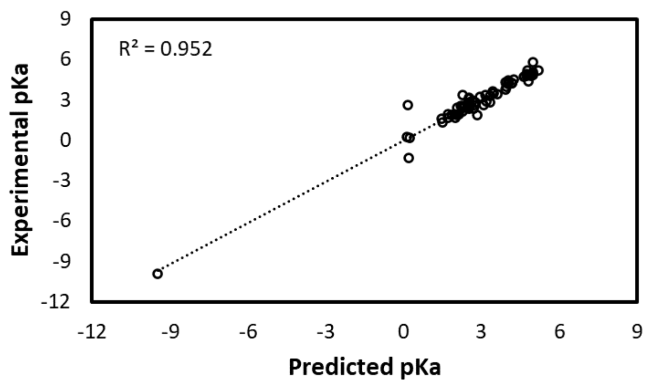
Figure 3.7. Relative contribution analysis (RCA) and recursive feature elimination analysis (RFE) in ***Pipeline 3***: a) KRR, b) GPR, and c) ANN. Prediction error denotes MAE by removing the feature from training.

Final Model for pK_a Prediction. Table 3.4 tabulates the performance of each model after ***Pipeline*** optimization. It is well-demonstrated that the MAE values obtained from the ML models in this study are smaller than that from the DFT approach (1.85). Additionally, we assessed MAE as well as maximum absolute error and standard deviation of each model's best pipeline using leave-one-out (LOO) cross validation. It should be noted that LOO cross validation increases the size of the training set compared to the shuffle method used in the paper, thereby decreasing the error. Additionally, CF_3SO_3H is problematic when assessing maximum absolute error and standard deviation due to the large reported range of pK_as in literature (-14 to -5.9). If we remove this data point, the MAE/maximum absolute error/standard deviation for each model are as follows: 0.53/5.30/0.81 for KRR/***Pipeline 3***, 0.63/4.07/0.72 for GPR/***Pipeline 3***, and 0.51/5.37/0.77 for ANN/***Pipeline 2***. While the maximum absolute error is large, it should be noted that the highest maximum absolute errors usually exist among the sulfur based acids. Certainly, the experimental data for these acids is sparser and less accurate due to the strength of these acids. Among the non-sulfur acids, the highest maximum absolute error for KRR/***Pipeline 3*** is 1.14 for $(CH_3CH_2)_2HCO_2H$.

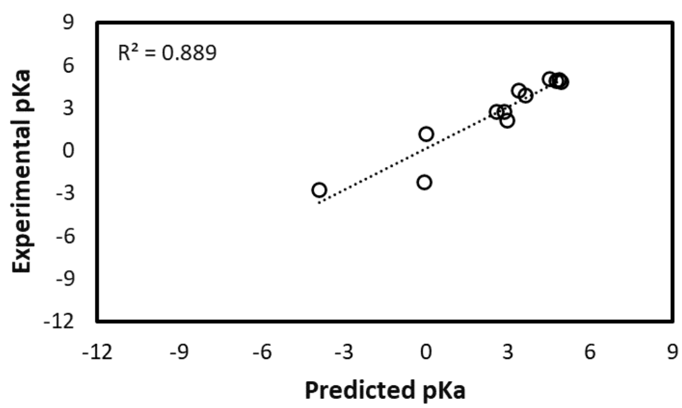
The performance of KRR via ***Pipeline 3*** for an arbitrary and representative (MAE for selected test data set is +/– 0.02 of the overall MAE in Table 3.4) train/test shuffle is depicted in Figure 3.8. The R^2 value of 0.889 for the test data is respectable, and qualitatively each individual prediction appears reasonable. To further investigate the role of machine learning in correcting DFT predictions, we decomposed the data by acid class (sulfur-based, phosphorous-based, and carboxylic). Due to the small nature of the dataset, LOO cross validation was used to predict individual pK_as for this analysis. Additionally, CF_3SO_3H is removed as mentioned above due to the large reported range of pK_as in literature. Results are tabulated in Table 3.5.

Table 3.4. Mean absolute errors in pK_a predictions by KRR, GPR, and ANN models optimized via three ***Pipelines.*** Please note that the DFT result refers to the averaged DFT/SM8 error on the entire dataset

Pipeline	**Mean Absolute Errors (MAE)**			
	DFT	**KRR**	**GPR**	**ANN**
1		0.93	0.80	0.79
2	1.85	0.87	0.77	0.74
3		0.60	0.74	0.79



(a)



(b)

Figure 3.8. Prediction performance of KRR via ***Pipeline 3*** on (a) training and (b) test data.

Table 3.5. Comparison of MAE in prediction between DFT and KRR/***Pipeline 3***

Acid	**DFT MAE (B3LYP/6-31++G**/SM8)**	**ML MAE (KRR/Pipeline 3)**
Carboxylic	1.88	0.30
Sulfur	3.07	1.61
Phosphorous	1.27	0.33

LOO cross validation confirms that in every case, ML corrects DFT significantly. The stellar performance for both phosphorous and carboxylic acid also reiterates our algorithm's flexibility. Sulfur-based acids have comparatively poor performance. This is likely due to the small dataset and difficulty of determining pK_a experimentally. However, our algorithm still offers a significant improvement over DFT methods.

Based on its superior MAE, KRR in ***Pipeline 3*** is chosen to predict the pK_a of novel acids after refitting the model to 100% of the validation set (i.e. 100/0 train/test split). The novel acids were oxoacids with minor structural variations compared to the acids in the training data (e.g. a slightly different chain length or number of fluorine atoms in structure). We chose structures similar to those in the training data for two reasons: 1) documenting their pK_a will aid in PEMFC design since these types of oxoacids are viable candidates for PEMs, and 2) they can be directly compared to acids in the training data to better understand the effects of minor structural variations upon pK_a. Additionally, prediction of pK_a for novel acids demonstrates the model's capacity for high throughput predictions, since determining these pK_a values experimentally is a comparatively tedious process. To prepare the transformed features while ensuring proper normalization, novel acid descriptors were appended to those of the training data, and subsequently all features were simultaneously normalized, scaled, and transformed. Since the predicted pK_a values of these novel acids may be useful in future studies, results of all predictions are documented in Table 3.6.

To provide a heuristic in estimating pK_a for a given acid, the Pearson correlation coefficients between ML-predicted pK_a and other features for the novel and training acids were calculated and plotted in Figure 3.9.

Table 3.6. Results for pK_a calculation for novel acids using KRR with ***Pipeline 3***. Columns 1 – 10 are descriptor IDs 1 through 10

Formula	1	2	3	4	5	6	7	8	9	10	pKa
CF2CF3-CO2H	2.55	2	0	5	101.9	164.0	3.2	2	-1.7	-0.71	0.80
CFHCH3-CO2H	2.55	2	4	1	77.9	92.1	3.7	2	-8.5	2.23	2.72
CFHCF3-CO2H	2.55	2	1	4	97.1	146.0	3.1	2	-6.1	2.58	2.27
CF2CH3-CO2H	2.55	2	3	2	86.2	110.1	1.2	2	-8.1	1.44	2.57
CF3OCOOH	2.55	3	0	3	82.6	130.0	3.2	1	-3.2	-14.10	-12.10
CH3OCOOH	2.55	3	3	0	65.1	76.1	3.7	1	-9.2	0.35	2.66
CH2F-SO3H	2.58	3	2	1	78.9	114.1	5.2	1	-24.5	-4.96	-2.46
CF2CF3-SO3H	2.58	3	0	5	118.9	200.1	3.2	2	-10.7	-8.76	-2.14
CF2CH3-SO3H	2.58	3	3	2	101.9	146.1	3.4	2	-18.5	-7.71	-3.80
CF2H-SO3H	2.58	3	1	2	86.9	132.1	3.9	1	-19.6	-6.62	-3.60
CH2CF3-SO3H	2.58	3	2	3	108.5	164.1	3.1	2	-20.9	-3.09	0.00
CFHCF3-SO3H	2.58	3	1	4	114.9	182.1	2.3	2	-16.8	-6.63	-1.29
CH2CH3-SO3H	2.58	3	5	0	91.7	110.1	4.9	2	-20.4	-4.05	0.01
CFHCH3-SO3H	2.58	3	4	1	97.6	128.1	3.1	2	-20.8	-5.14	-1.98
CH2CH3-OSO3H	2.58	4	5	0	97.5	126.1	4.3	2	-19.0	-9.26	-4.18
CF3CF2-OSO3H	2.58	4	0	5	125.0	216.1	2.8	2	-11.6	-14.95	-8.29
CF3OSO3H	2.58	4	0	3	97.0	166.1	2.6	1	-12.6	-14.25	-11.00
CF2CH3-OSO3H	2.58	4	3	2	110.4	162.1	4.1	2	-19.1	15.52	4.16
CF2H-OSO3H	2.58	4	1	2	92.1	148.1	3.1	1	-17.1	-12.48	-9.49
CFHCF3-OSO3H	2.58	4	1	4	122.1	198.1	3.2	2	-16.5	-10.42	-3.63
CFHCH3-OSO3H	2.58	4	4	1	104.7	144.1	6.2	2	-20.9	-10.22	-6.17
CH2CF3-OSO3H	2.58	4	2	3	114.6	180.1	1.6	2	-19.7	-8.57	-2.74
CH2F-OSO3H	2.58	4	2	1	87.1	130.1	4.1	1	-20.4	-6.92	-3.58
CH3OSO3H	2.58	4	3	0	80.8	112.1	3.9	1	-19.8	-8.62	-4.26
CF2CF3-PO3H2	2.19	3	0	5	121.9	200.0	4.3	2	-7.9	2.41	3.25
CF2CH3-PO3H2	2.19	3	3	2	109.8	146.0	4.3	2	-13.6	1.56	1.10
CFHCH3-PO3H2	2.19	3	4	1	98.9	128.0	3.6	2	-14.1	3.23	0.96
CH2CF3-PO3H2	2.19	3	2	3	113.1	164.0	3.7	2	-13.6	5.32	2.10
CF3OPO3H2	2.19	4	0	3	101.3	166.0	0.0	1	-7.6	-1.06	1.90
CF2CF3-OPO3H2	2.19	4	0	5	129.5	216.0	2.2	2	-6.2	-0.98	4.45
CF2CH3-OPO3H2	2.19	4	3	2	116.0	162.0	5.2	2	-13.1	1.07	1.82
CF2H-OPO3H2	2.19	4	1	2	96.2	148.0	2.7	1	-11.3	-0.44	0.37
CFHCF3-OPO3H2	2.19	4	1	4	125.4	198.0	2.6	2	-9.4	1.85	3.77
CFHCH3-OPO3H2	2.19	4	4	1	106.3	144.0	2.7	2	-11.7	-0.18	0.86
CH2CF3-OPO3H2	2.19	4	2	3	120.9	180.0	2.0	2	-12.9	2.48	2.97
CH2F-OPO3H2	2.19	4	2	1	89.9	130.0	3.2	1	-12.3	-0.51	0.02
CF2H-PO3H2	2.19	3	1	2	90.4	132.0	2.4	1	-13.3	1.50	0.12

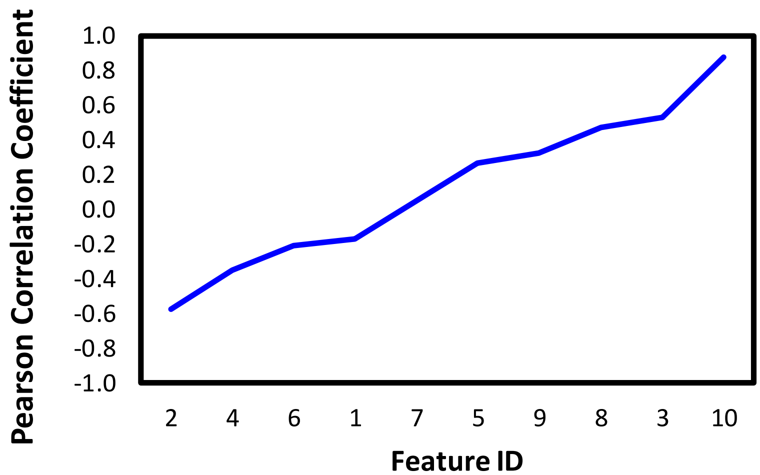


Figure 3.9. Pearson correlation coefficients between the predicted pK_a and features 1 – 10.

As expected, DFT-predicted pK_a shows a strong correlation with pK_a. Surprisingly, SFE shows little correlation with pK_a, though it is possible that our ML model ignores this feature due to its redundancy with DFT-predicted pK_a. Due to their contribution to the electron withdrawing capacity of the molecule, the number of fluorine atoms and degree of oxidation exhibit a negative correlation with $pK_{a,}$. Conversely, the number of hydrogen atoms shows a positive correlation with pK_a. This can be rationalized by acknowledging that due to their positive polarization, the effect of hydrogen atoms on electron withdrawal is opposite to that of fluorine, wherein additional hydrogen atoms signify weaker electron withdrawal capacity of the molecule (and therefore lesser stability of its deprotonated moiety).

Finally, consistent with the "positive induction hypothesis" mentioned above, the number of carbons shows a positive correlation with pK_a. It should be noted that most correlation coefficients are relatively low (magnitude < 0.6), indicating that no single structural or chemical property is a particularly strong predictor of pK_a.

Conclusion

Although the DFT method alone can be used to predict the pK_a of acids without experimental data, its mean absolute error in our study is ~1.85 pK_a units despite choosing the best condition consisting of B3LYP/6-31++G** and SM8. Therefore, these DFT-based pK_a values were employed as one of the descriptors for ML algorithms to develop a highly accurate DFT-ML method for pK_a prediction.

For this purpose, we employed three ML algorithms including KRR, GPR, and ANN, which were optimized using three ***Pipelines***. ***Pipeline 1*** is the most basic strategy involving only hyperparameter optimization (HPO), ***Pipeline 2*** performs a relative contribution analysis (RCA) .and recursive feature elimination (RFE) in addition to ***Pipeline 1***, and ***Pipeline 3*** utilizes an expanded set of transformed features in ***Pipeline 2***. It was found that KRR with ***Pipeline 3*** demonstrated the lowest MAE of 0.60 pK_a units. Using this model, the most effective structural contributors to pK_a were identified as degree of oxidation and number of hydrogen atoms in the molecular structure. This model offers respectable accuracy and flexibility for the prediction of pK_a in PEM-relevant acids, and will be an invaluable tool in the design of novel, durable PEM chemistries.

V. CONCLUSION

In this thesis, I have conducted a computational study which addresses pertinent barriers to the universal adoption of portable PEMFC devices in various applications, by using multiscale computational simulations to enhance fundamental understanding and suggest superior chemistries for PEMs. In **Aim 1**, I addressed the issue of PEM dehydration by identifying a crucial factor in PEM conductivity (bridge structure formation), thereby informing design of dehydration-resistant membranes. In **Aim 2**, I addressed the issue of short PEMFC lifetimes by providing novel insight into the OH/OOH binding mechanism to ceria radical scavengers, and by proposing a superior CeO_2 surface defect geometry to enhance membrane preservation. Finally, in **Aim 3**, I created a tool which expands upon Aim 1 by utilizing DFT and ML to quickly and accurately predict pK_a of PEM-relevant acids, to advise the design of novel PEMs with superior conductivity and performance in various conditions. I am confident this study constitutes a significant contribution toward the design of PEMFCs for widespread use in many applications.

VI. FUTURE WORK

In addition to the previous studies, the author is currently pursuing a project to investigate structure and transport characteristics of the PEMFC's 3-phase interfacial system at the cathode. The motivation is as follows: while PEMFCs have been integrated into numerous industrial processes, their high cost prohibits widespread use in consumer vehicles.[215] Addressing this cost barrier is a crucial step in rendering PEMFC technology accessible. The most significant contribution to PEMFC cost is the use of platinum catalyst in its electrodes.

Platinum is a powerful catalyst capable of rapid and efficient conversion of reactant gases to provide an uncontested power density. Unfortunately, platinum is an exceedingly rare metal: there are an estimated 28,000 tons of platinum reserves in existence, whereas 150,000 tons would be necessary to equip all consumer vehicles with current PEMFC technology.[216-217] To address this issue, scientists have proposed several solutions, including alternative catalysts such as iridium, as well as platinum recycling.[217-218] However, the former does not offer acceptable performance, and the latter doesn't sufficiently offset the cost of raw platinum. As such, an alternative solution must be considered.

Over the past decade, low-Pt PEMFCs have gained popularity as a viable alternative to traditional PEMFCs.[219] Lowering platinum loading significantly reduces the cost of PEMFCs while retaining the efficiency of a platinum catalyst and slowing the rate of recycling. However, low-Pt PEMFCs generate notoriously low power densities compared to traditional PEMFCs.[216] This has been attributed to increased transport resistance through the cathode's ionomer membrane.

One potential solution lies in the manipulation of the ionomer's phase segregation. Theoretical studies have demonstrated that ionomer phase segregation and transport changes depending on surface chemistry (i.e. CB support vs. Pt catalyst).[220] Currently, we are examining the effect of hydrophilic CB functionalization upon ionomer properties. Specifically, we employ multiphase atomistic simulations to model the three-phase interfacial system including amorphous carbon black, graphitic carbon, ionomer membrane (both Nafion and Aquivion are examined), and gaseous phase including O_2 molecules. The graphitic carbon surface is functionalized with 0%, 3%, 6%, 9%, and 12% oxygen (hydroxyl or epoxide) and the resultant structure and transport within the system is analyzed through pair correlation functions, structure factor analyses, and MSD calculations. Through such analysis, we strive to deepen our understanding of structure-property relationships in the PEMFC's three phase interfacial system.

VII. SOURCES

1. Dyer, C. K., Fuel Cells for Portable Applications. *Journal of Power Sources* **2002**, *106*, 31-34.
2. Mauritz, K. A.; Moore, R. B., State of Understanding of Nafion. *Chem. Rev.* **2004**, *104*, 4535-4585.
3. Devanathan, R., Recent Developments in Proton Exchange Membranes for Fuel Cells. *Energ. Environ. Sci.* **2008**, *1*, 101-119.
4. Pei, P.; Chen, H., Main Factors Affecting the Lifetime of Proton Exchange Membrane Fuel Cells in Vehicle Applications: A Review. *Applied Energy* **2014**, *125*, 60-75.
5. Brouzgou, A.; Song, S. Q.; Tsiakaras, P., Low and Non-Platinum Electrocatalysts for Pemfcs: Current Status, Challenges and Prospects. *Applied Catalysis B: Environmental* **2012**, *127*, 371-388.
6. Ganesan, A.; Narayanasamy, M., Ultra-Low Loading of Platinum in Proton Exchange Membrane-Based Fuel Cells: A Brief Review. *Materials for Renewable and Sustainable Energy* **2019**, *8*.
7. Edmondson, C. A.; Stallworth, P. E.; Chapman, M. E.; Fontanella, J. J.; Wintersgill, M. C.; Chung, S. H.; Greenbaum, S. G., Complex Impedance Studies of Proton-Conducting Membranes. *Solid State Ionics* **2000**, *135*, 419-423.
8. Idupulapati, N.; Devanathan, R.; Dupuis, M., Atomistic Simulations of Perfluoro Phosphonic and Phosphinic Acid Membranes and Comparisons to Nafion. *J. Phys. Chem. B* **2011**, *115*, 2959-2969.
9. Donnadio, A.; D'Amato, R.; Marmottini, F.; Panzetta, G.; Pica, M.; Battocchio, C.; Capitani, D.; Ziarelli, F.; Casciola, M., On the Evolution of Proton Conductivity of Aquivion Membranes Loaded with Ceo2 Based Nanofillers: Effect of Temperature and Relative Humidity. *Journal of Membrane Science* **2019**, *574*, 17-23.
10. Endoh, E., Development of Highly Durable Pfsa Membrane and Mea for Pemfc under High Temperature and Low Humidity Conditions. *ECS Transactions* **2008**, *16*, 1229-1240.
11. Coms, F. D.; Liu, H.; Owejan, J. E., Mitigation of Perfluorosulfonic Acid Membrane Chemical Degradation Using Cerium and Manganese Ions. In *Proton Exchange Membrane Fuel Cells 8, Pts 1 and 2*, Fuller, T.; Shinohara, K.; Ramani, V.; Shirvanian, P.; Uchida, H.; Cleghorn, S.; Inaba, M.; Mitsushima, S.; Strasser, P.; Nakagawa, H.; Gasteiger, H. A.; Zawodzinski, T.; Lamy, C., Eds. 2008; Vol. 16, pp 1735-+.
12. Nesselberger, M., et al., The Effect of Particle Proximity on the Oxygen Reduction Rate of Size-Selected Platinum Clusters. *Nat Mater* **2013**, *12*, 919-24.
13. Sievers, G.; Vidakovic-Koch, T.; Walter, C.; Steffen, F.; Jakubith, S.; Kruth, A.; Hermsdorf, D.; Sundmacher, K.; Brüser, V., Ultra-Low Loading Pt-Sputtered Gas Diffusion Electrodes for Oxygen Reduction Reaction. *Journal of Applied Electrochemistry* **2018**, *48*, 221-232.
14. Jang, S.; Molinero, V.; Cagin, T.; Goddard, W., Nanophase-Segregation and Transport in Nafion 117 from Molecular Dynamics Simulations: Effect of Monomeric Sequence. *J. Phys. Chem. B* **2004**.
15. Norenberg, H.; Briggs, G. A. D., Defect Formation on Ceo2(111) Surfaces after Annealing Studied by Stm. *Surface Science* **1999**, *424*, L352-L355.
16. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Phys Rev* **1965**, *140*, A1133-A1138.

17. Dreizler, R. M.; Gross, E. K. U., *Density Functional Theory*; Springer-Verlag: Berlin, 1990.
18. Parr, R. G.; Yang, W., *Density Functional Theory of Atoms and Molecules*; Oxford: New York, 1989.
19. Perdew, J. P.; Yue, W., Accurate and Simple Density Functional for the Electronic Exchange Energy: Generalized Gradient Approximation. *Phys Rev B Condens Matter* **1986**, *33*, 8800-8802.
20. Langreth, D. C.; Mehl, M. J., Beyond the Local-Density Approximation in Calculations of Ground-State Electronic Properties. *Physical Review B* **1983**, *28*, 1809-1834.
21. Becke, A. D., Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys Rev A Gen Phys* **1988**, *38*, 3098-3100.
22. Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C., Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation. *Phys Rev B Condens Matter* **1992**, *46*, 6671-6687.
23. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, *77*, 3865-3868.
24. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **1994**, *98*, 11623-11627.
25. Becke, A. D., Density-Functional Thermochemistry .3. The Role of Exact Exchange. *Journal of Chemical Physics* **1993**, *98*, 5648-5652.
26. Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - a Generic Force-Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897-8909.
27. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R., A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637-649.
28. Nose, S.; Klein, M. L., A Study of Solid and Liquid Carbon Tetrafluoride Using the Constant Pressure Molecular-Dynamics Technique. *J. Chem. Phys.* **1983**, *78*, 6928-6939.
29. Nose, S., A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511-519.
30. Nose, S., A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255-268.
31. Murphy, K. P., *Machine Learning: A Probabilistic Perspective*; The MIT Press, 2012.
32. Rasmussen, C. E.; Williams, C. K., *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, 2006.
33. Shawe-Taylor, J.; Cristianini, N., *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*; Cambridge University Press, 2000.
34. Shawe-Taylor, J.; Cristianini, N., *Kernel Methods for Pattern Analysis*; Cambridge University Press, 2004.
35. Kamath, A.; Vargas-Hernandez, R. A.; Krems, R. V.; Carrington, T., Jr.; Manzhos, S., Neural Networks Vs Gaussian Process Regression for Representing Potential Energy Surfaces: A Comparative Study of Fit Quality and Vibrational Spectrum Accuracy. *J. Chem. Phys.* **2018**, *148*, 241702.
36. Genton, M. G., Classes of Kernels for Machine Learning: A Statistics Perspective. *J Mach Learn Res* **2002**, *2*, 299-312.

37. Stoyan, D.; Matern, B., *Spatial Variation*, 2 ed.; Springer: Berlin, 1988.
38. Vert, J. P.; Tsuda, K.; Scholkopf, B., *Kernel Methods in Computational Biology*. MIT Press: Cambridge, MA, 2004.
39. Bhadeshia, H. K. D. H., Neural Networks in Materials Science. *ISIJ Int.* **1999**, *39*, 966-979.
40. Bhadeshia, H. K. D. H.; Dimitriu, R. C.; Forsik, S.; Pak, J. H.; Ryu, J. H., Performance of Neural Networks in Materials Science. *Materials Science and Technology* **2013**, *25*, 504-510.
41. Rumelhart, D. E.; McClelland, J. L., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*; MIT Press, 1986.
42. Mcculloch, W. S.; Pitts, W., A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biol.* **1990**, *52*, 99-115.
43. Zuo, J.; Chen, Q.; Hu, X.; Guo, H.; Xie, D., Theoretical Investigations of Rate Coefficients for H + O_3 and Ho_2 + O Reactions on a Full-Dimensional Potential Energy Surface. *J. Phys. Chem. A* **2020**, *124*, 6427-6437.
44. Devanathan, R., Recent Developments in Proton Exchange Membranes for Fuel Cells. *Energy & Environmental Science* **2008**, *1*, 101-119.
45. Hung, J.; Stone, J.; Suriano, J.; Zhou, H. In *Proceedings of Hydrogen Fuel Cells and Infrastructure Technologies Program Review*, Crystal City, VA, Crystal City, VA, 2006.
46. Herring, A. In *Proceedings of Hydrogen Fuel Cells and Infrastructure Technologies Program Review*, Crystal City, VA, Crystal City, VA, 2006.
47. Garland, N. L.; Kopasz, J. P., The United States Department of Energy's High Temperature, Low Relative Humidity Membrane Program. *J. Power Sources* **2007**, *172*, 94-99.
48. Herath, M. B.; Creager, S. E.; Kitaygorodskiy, A.; DesMarteau, D. D., Perfluoroalkyl Phosphonic and Phosphinic Acids as Proton Conductors for Anhydrous Proton-Exchange Membranes. *ChemPhysChem* **2010**, *11*, 2871-2878.
49. Schuster, M.; Rager, T.; Noda, A.; Kreuer, K. D.; Maier, J., About the Choice of the Protogenic Group in Pem Separator Materials for Intermediate Temperature, Low Humidity Operation: A Critical Comparison of Sulfonic Acid, Phosphonic Acid and Imidazole Functionalized Model Compounds. *Fuel Cells* **2005**, *5*, 355-365.
50. Idupulapati, N.; Devanathan, R.; Dupuis, M., Atomistic Simulations of Perfluoro Phosphonic and Phosphinic Acid Membranes and Comparisons to Nafion. *J. Phys. Chem. B* **2011**, *115*, 2959-2969.
51. Dippel, T.; Kreuer, K. D., Proton Conductivity in Fused Phosphoric Acid; a 1h/31p Pfg-Nmr and Qns Study. *Solid State Ion.* **1993**, *61*.
52. Paddison, S. J.; Kreuer, K.-D.; Maier, J., About the Choice of the Protogenic Group in Polymer Electrolyte Membranes: Ab Initio Modelling of Sulfonic Acid, Phosphonic Acid, and Imidazole Functionalized Alkanes. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4530-4542.
53. Rollet, A.-L.; Diat, O.; Gebel, G., A New Insight into Nafion Structure. *J. Phys. Chem. B* **2002**, *106*, 3033-3036.
54. Rollet, A.-L.; Gebel, G.; Simonin, J.-P.; Turq, P., A Sans Determination of the Influence of External Conditions on the Nanostructure of Nafion Membrane. *J. Polym. Sci.* **2001**, *39*, 548-558.
55. Sharaf, O. Z.; Orhan, M. F., An Overview of Fuel Cell Technology: Fundamentals and Applications. *Renew. Sustain. Energy Rev.* **2014**, *32,* 810-853.
56. Vishnyakov, A.; Neimark, A. V., Molecular Dynamics Simulation of Microstructure and Molecular Mobilities in Swollen Nafion Membranes. *J. Phys. Chem. B* **2001**, *105*, 9586-9594.

57. Yeager, H. L.; Steck, A., Cation and Water Diffusion in Nafion Ion-Exchange Membranes - Influence of Polymer Structure. *J. Electrochem. Soc.* **1981**, *128*, 1880-1884.
58. Elliott, J. A.; Paddison, S. J., Modelling of Morphology and Proton Transport in Pfsa Membranes. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2602-2618.
59. Jang, S. S.; Molinero, V.; Cagin, T.; Goddard, W. A., Nanophase-Segregation and Transport in Nafion 117 from Molecular Dynamics Simulations: Effect of Monomeric Sequence. *J. Phys. Chem. B* **2004**, *108*, 3149-3157.
60. Liu, J.; Suraweera, N.; Keffer, D. J.; Cui, S.; Paddison, S. J., On the Relationship between Polymer Electrolyte Structure and Hydrated Morphology of Perfluorosulfonic Acid Membranes. *J. Phys. Chem. C* **2010**, *114*, 11279-11292.
61. Cui, S.; Liu, J.; Selvan, M. E.; Paddison, S. J.; Keffer, D. J.; Edwards, B. J., Comparison of the Hydration and Diffusion of Protons in Perfluorosulfonic Acid Membranes with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2008**, *112*, 13273-13284.
62. Jinnouchi, R.; Okazaki, K., Molecular Dynamics Study of Transport Phenomena in Perfluorosulfonate Ionomer Membranes for Polymer Electrolyte Fuel Cells. *J. Electrochem. Soc.* **2003**, *150*, E66-E73.
63. Luo, X.; Holdcroft, S.; Mani, A.; Zhang, Y.; Shi, Z., Water, Proton, and Oxygen Transport in High Iec, Short Side Chain Pfsa Ionomer Membranes: Consequences of a Frustrated Network. *Physical Chemistry Chemical Physics* **2011**, *13*, 18055-18062.
64. Stassi, A.; Gatto, I.; Passalacqua, E.; Antonucci, V.; Arico, A. S.; Merlo, L.; Oldani, C.; Pagano, E., Performance Comparison of Long and Short-Side Chain Perfluorosulfonic Membranes for High Temperature Polymer Electrolyte Membrane Fuel Cell Operation. *Journal of Power Sources* **2011**, *196*, 8925-8930.
65. Devanathan, R.; Dupuis, M., Insight from Molecular Modelling: Does the Polymer Side Chain Length Matter for Transport Properties of Perfluorosulfonic Acid Membranes? *Phys. Chem. Chem. Phys.* **2012**, *14*, 11281-95.
66. Karo, J.; Aabloo, A.; Thomas, J. O.; Brandell, D., Molecular Dynamics Modeling of Proton Transport in Nation and Hylton Nanostructures. *Journal of Physical Chemistry B* **2010**, *114*, 6056-6064.
67. Wu, D.; Paddison, S. J.; Elliott, J. A., A Comparative Study of the Hydrated Morphologies of Perfluorosulfonic Acid Fuel Cell Membranes with Mesoscopic Simulations. *Energy & Environmental Science* **2008**, *1*, 284-293.
68. Kreuer, K.-D.; Schuster, M.; Obliers, B.; Diat, O.; Traub, U.; Fuchs, A.; Klock, U.; Paddison, S. J.; Maier, J., Short-Side-Chain Proton Conducting Perfluorosulfonic Acid Ionomers: Why They Perform Better in Pem Fuel Cells. *J. Power Sources* **2008**, *178*, 499-509.
69. Wang, Y.; Chen, K. S.; Mishler, J.; Cho, S. C.; Adroher, X. C., A Review of Polymer Electrolyte Membrane Fuel Cells: Technology, Applications, and Needs on Fundamental Research. *Appl. Energy* **2011**, *88*, 981-1007.
70. Paddison, S. J.; Paul, R.; Zawodzinski, T. A., Proton Friction and Diffusion Coefficients in Hydrated Polymer Electrolyte Membranes: Computations with a Non-Equilibrium Statistical Mechanical Model. *J. Chem. Phys.* **2001**, *115*, 7753-7761.
71. Brandell, D.; Karo, J.; Liivat, A.; Thomas, J. O., Molecular Dynamics Studies of the Nafion (R), Dow (R) and Aciplex (R) Fuel-Cell Polymer Membrane Systems. *J. Mol. Model.* **2007**, *13*, 1039-1046.
72. Kuo, A.-T.; Takeuchi, K.; Tanaka, A.; Urata, S.; Okazaki, S.; Shinoda, W., Exploring the Effect of Pendent Side Chain Length on the Structural and Mechanical Properties of Hydrated

Perfluorosulfonic Acid Polymer Membranes by Molecular Dynamics Simulation. *Polymer* **2018**, *146*, 53-62.
73. Laporta, M.; Pegoraro, M.; Zanderighi, L., Perfluorosulfonated Membrane (Nafion): Ft-Ir Study of the State of Water with Increasing Humidity. *Phys. Chem. Chem. Phys.* **1999**, *1*, 4619-4628.
74. Gruger, A.; Regis, A.; Schmatko, T.; Colomban, P., Nanostructure of Nafion (R) Membranes at Different States of Hydration - an Ir and Raman Study. *Vib. Spectrosc* **2001**, *26*, 215-225.
75. Thaplyal, P.; Bevilacqua, P. C., Experimental Approaches for Measuring Pka's in Rna and DNA. *Methods in enzymology* **2014**, *549*, 189-219.
76. Plimpton, S., Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.* **1995**, *117*, 1-19.
77. Takamatsu, T.; Eisenberg, A., Densities and Expansion Coefficients of Nafion Polymers. *Journal of Applied Polymer Science* **1979**, *24*, 2221-2235.
78. Morris, D. R.; Sun, X., Water-Sorption and Transport-Properties of Nafion-117-H. *J. Appl. Polym. Sci.* **1993**, *50*, 1445-1452.
79. Deng, W.-Q.; Molinero, V.; Goddard, W. A., Fluorinated Imidazoles as Proton Carriers for Water-Free Fuel Cell Membranes. *J. Am. Chem. Soc.* **2004**, *126*, 15644-15645.
80. Jang, S. S.; Lin, S.-T.; Cagin, T.; Molinero, V.; Goddard, W. A., Nanophase Segregation and Water Dynamics in the Dendrion Diblock Copolymer Formed from the Frechet Polyaryl Ethereal Dendrimer and Linear Ptfe. *J. Phys. Chem. B* **2005**, *109*, 10154-10167.
81. Lill, M. A.; Helms, V., Reaction Rates for Proton Transfer over Small Barriers and Connection to Transition State Theory. *J. Chem. Phys.* **2001**, *115*, 7985-7992.
82. Lill, M. A.; Helms, V., Compact Parameter Set for Fast Estimation of Proton Transfer Rates. *J. Chem. Phys.* **2001**, *114*, 1125-1132.
83. Tamai, Y.; Tanaka, H.; Nakanishi, K., Molecular Dynamics Study of Polymer-Water Interaction in Hydrogels. 2. Hydrogen-Bond Dynamics. *Macromolecules* **1996**, *29*.
84. Guàrdia, E.; Martí, J.; García-Tarrés, L.; Laria, D., A Molecular Dynamics Simulation Study of Hydrogen Bonding in Aqueous Ionic Solutions. *Journal of Molecular Liquids* **2005**, *117*, 63-67.
85. Ai, L. Q.; Zhou, Y. S.; Chen, M., A Reactive Force Field Molecular Dynamics Simulation of the Dynamic Properties of Hydrogen Bonding in Supercritical Water. *Journal of Molecular Liquids* **2019**, *276*, 83-92.
86. Liu, J.; He, X.; Zhang, J. Z. H.; Qi, L. W., Hydrogen-Bond Structure Dynamics in Bulk Water: Insights from Ab Initio Simulations with Coupled Cluster Theory. *Chemical Science* **2018**, *9*, 2065-2073.
87. Naberukhin, Y. I.; Voloshin, V. P., Distributions of Hydrogen Bond Lifetimes in Instantaneous and Inherent Structures of Water. *Zeitschrift Fur Physikalische Chemie-International Journal of Research in Physical Chemistry & Chemical Physics* **2009**, *223*, 1119-1131.
88. Voloshin, V. P.; Naberukhin, Y. I., Hydrogen Bond Lifetime Distributions in Computer-Simulated Water. *Journal of Structural Chemistry* **2009**, *50*, 78-89.
89. Markovitch, O.; Agmon, N., Structure and Energetics of the Hydronium Hydration Shells. *Journal of Physical Chemistry A* **2007**, *111*, 2253-2256.
90. Antipova, M. L.; Petrenko, V. E., Hydrogen Bond Lifetime for Water in Classic and Quantum Molecular Dynamics. *Russian Journal of Physical Chemistry A* **2013**, *87*, 1170-1174.

91. Sun, Y.; Ogden, J.; Delucchi, M., Societal Lifetime Cost of Hydrogen Fuel Cell Vehicles. *International Journal of Hydrogen Energy* **2010**, *35*, 11932-11946.
92. Advancing Clean Electric Power Technologies. In *QUADRENNIAL TECHNOLOGY REVIEW*, U.S. Department of Energy: 2015.
93. Gubler, L.; Dockheer, S. M.; Koppenol, W. H., Radical (Ho Center Dot, H-Center Dot and Hoo Center Dot) Formation and Ionomer Degradation in Polymer Electrolyte Fuel Cells. *Journal of the Electrochemical Society* **2011**, *158*, B755-B769.
94. de Bruijn, F. A.; Dam, V. A. T.; Janssen, G. J. M., Review: Durability and Degradation Issues of Pem Fuel Cell Components. *Fuel Cells* **2008**, *8*, 3-22.
95. Uegaki, R.; Akiyama, Y.; Tojo, S.; Honda, Y.; Nishijima, S., Radical-Induced Degradation Mechanism of Perfluorinated Polymer Electrolyte Membrane. *Journal of Power Sources* **2011**, *196*, 9856-9861.
96. Sugawara, T.; Kawashima, N.; Murakami, T. N., Kinetic Study of Nafion Degradation by Fenton Reaction. *Journal of Power Sources* **2011**, *196*, 2615-2620.
97. Coms, F. D.; Liu, H.; Owejan, J. E., Mitigation of Perfluorosulfonic Acid Membrane Chemical Degradation Using Cerium and Manganese Ions. *ECS Transactions* **2008**, *16*, 1735-+.
98. Kim, J.; Chung, K.; Lee, H.; Bae, B.; Cho, E.-B., Mesoporous Ceria-Silica/Poly(Arylene Ether Sulfone) Composite Membranes for Durability of Fuel Cell Electrolyte Membrane. *Microporous and Mesoporous Materials* **2016**, *236*, 292-300.
99. Endoh, E.; Onoda, N.; Kaneko, Y.; Hasegawa, Y.; Uchiike, S.; Takagi, Y.; Take, T., Membrane Degradation Mitigation of Pefc During Cold-Start Application of the Radical Quencher Ce3+. *ECS Electrochemistry Letters* **2013**, *2*, F73-F75.
100. Zaton, M.; Prelot, B.; Donzel, N.; Roziere, J.; Jones, D. J., Migration of Ce and Mn Ions in Pemfc and Its Impact on Pfsa Membrane Degradation. *Journal of the Electrochemical Society* **2018**, *165*, F3281-F3289.
101. Baker, A. M.; Dumont, J. H.; Mukundan, R.; Advani, S. G.; Prasad, A. K.; Spernjak, D.; Borup, R. L., The Effects of Cerium Migration on Pem Fuel Cell Performance. *ECS Transactions* **2017**, *80*, 643-650.
102. Lim, C.; Alavijeh, A. S.; Lauritzen, M.; Kolodziej, J.; Knights, S.; Kjeang, E., Fuel Cell Durability Enhancement with Cerium Oxide under Combined Chemical and Mechanical Membrane Degradation. *Ecs Electrochemistry Letters* **2015**, *4*, F29-F31.
103. Pearman, B. P.; Mohajeri, N.; Brooker, R. P.; Rodgers, M. P.; Slattery, D. K.; Hampton, M. D.; Cullen, D. A.; Seal, S., The Degradation Mitigation Effect of Cerium Oxide in Polymer Electrolyte Membranes in Extended Fuel Cell Durability Tests. *Journal of Power Sources* **2013**, *225*, 75-83.
104. Trogadas, P.; Parrondo, J.; Ramani, V., Degradation Mitigation in Polymer Electrolyte Membranes Using Cerium Oxide as a Regenerative Free-Radical Scavenger. *Electrochemical and Solid State Letters* **2008**, *11*, B113-B116.
105. Wang, L.; Advani, S. G.; Prasad, A. K., Degradation Reduction of Polymer Electrolyte Membranes Using Ceo2 as a Free-Radical Scavenger in Catalyst Layer. *Electrochimica Acta* **2013**, *109*, 775-780.
106. Trogadas, P.; Parrondo, J.; Ramani, V., Ceo2 Surface Oxygen Vacancy Concentration Governs in Situ Free Radical Scavenging Efficacy in Polymer Electrolytes. *Acs Applied Materials & Interfaces* **2012**, *4*, 5098-5102.

107. Xue, Y.; Luan, Q.; Yang, D.; Yao, X.; Zhou, K., Direct Evidence for Hydroxyl Radical Scavenging Activity of Cerium Oxide Nanoparticles. *The Journal of Physical Chemistry C* **2011**, *115*, 4433-4438.
108. Trovarelli, A.; Llorca, J., Ceria Catalysts at Nanoscale: How Do Crystal Shapes Shape Catalysis? *Acs Catalysis* **2017**, *7*, 4716-4735.
109. Namai, Y.; Fukui, K.; Iwasawa, Y., Atom-Resolved Noncontact Atomic Force Microscopic Observations of Ceo2(111) Surfaces with Different Oxidation States: Surface Structure and Behavior of Surface Oxygen Atoms. *Journal of Physical Chemistry B* **2003**, *107*, 11666-11673.
110. Norenberg, H.; Briggs, G. A. D., Defect Structure of Nonstoichiometric Ceo2(111) Surfaces Studied by Scanning Tunneling Microscopy. *Physical Review Letters* **1997**, *79*, 4222-4225.
111. Esch, F.; Fabris, S.; Zhou, L.; Montini, T.; Africh, C.; Fornasiero, P.; Comelli, G.; Rosei, R., Electron Localization Determines Defect Formation on Ceria Substrates. *Science* **2005**, *309*, 752-755.
112. Kresse, G.; Hafner, J., Ab Initio Molecular-Dynamics Simulation of the Liquid-Metal-Amorphous-Semiconductor Transition in Germanium. *Physical Review B: Condensed Matter and Materials* **1994**, *49*, 14251-14269.
113. Kresse, G.; Joubert, D., From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Physical Review B: Condensed Matter and Materials* **1999**, *59*, 1758-1775.
114. Blochl, P. E., Projector Augmented-Wave Method. *Physical Review B: Condensed Matter and Materials* **1994**, *50*, 17953-17979.
115. Grieshammer, S., Influence of the Lattice Constant on Defects in Cerium Oxide. *Physical Chemistry Chemical Physics* **2018**, *20*, 19792-19799.
116. Johnson, E. R.; Becke, A. D., A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J Chem Phys* **2006**, *124*, 174104.
117. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (Dft-D) for the 94 Elements H-Pu. *J Chem Phys* **2010**, *132*, 154104.
118. Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C., First Principles Methods Using Castep. *Zeitschrift fur Kristallographie* **2005**, *220*, 567-570.
119. *Materials Studio*, Accelrys Software Inc., 2009.
120. Siegel, D. A.; Chueh, W. C.; El Gabaly, F.; McCarty, K. F.; de la Figuera, J.; Blanco-Rey, M., Determination of the Surface Structure of Ceo2(111) by Low-Energy Electron Diffraction. *Journal of Chemical Physics* **2013**, *139*, 114703.
121. Fronzi, M.; Soon, A.; Delley, B.; Traversa, E.; Stampfi, C., Stability and Morphology of Cerium Oxide Surfaces in an Oxidizing Environment: A First-Principles Investigation. *Journal of Chemical Physics* **2009**.
122. Fabris, S.; Vicario, G.; Balducci, G.; de Gironcoli, S.; Baroni, S., Electronic and Atomistic Structures of Clean and Reduced Ceria Surfaces. *Journal of Physical Chemistry B* **2005**, *109*, 22860-22867.
123. Yang, Z.; Yu, X.; Lu, Z.; Li, S.; Hermansson, K., Oxygen Vacancy Pairs on Ceo2(110): A Dft + U Study. *Physics Letters A* **2009**, *373*, 2786-2792.

124. Zhang, C.; Michaelides, A.; King, D. A.; Jenkins, S. J., Oxygen Vacancy Clusters on Ceria: Decisive Role of Cerium F Electrons. *Physical Review B: Condensed Matter and Materials* **2009**, *79*.
125. Skorodumova, N. V.; Simak, S. I.; Lundqvist, B. I.; Abrikosov, I. A.; Johansson, B., Quantum Origin of the Oxygen Storage Capability of Ceria. *Physical Review Letters* **2002**, *89*, 166601.
126. Han, Z. K.; Zhang, L.; Liu, M.; Ganduglia-Pirovano, M. V.; Gao, Y., The Structure of Oxygen Vacancies in the near-Surface of Reduced Ceo2 (111) under Strain. *Frontiers in Chemistry* **2019**, *7*, 436.
127. Cho, J.; Jang, I.; Park, H. S.; Choi, S. H.; Jang, J. H.; Kim, H. J.; Yoon, S. P.; Yoo, S. J.; Ham, H. C., Computational and Experimental Design of Active and Durable Ir-Based Nanoalloy for Electrochemical Oxygen Reduction Reaction. *Applied Catalysis B: Environmental* **2018**, *235*, 177-185.
128. Mavrikakis, M.; Hammer, B.; Nørskov, J. K., Effect of Strain on the Reactivity of Metal Surfaces. *Physical Review Letters* **1998**, *81*, 2819-2822.
129. Kitchin, J. R.; Nørskov, J. K.; Barteau, M. A.; Chen, J. G., Role of Strain and Ligand Effects in the Modification of the Electronic and Chemical Properties of Bimetallic Surfaces. *Physical Review Letters* **2004**, *93*, 156801.
130. Cho, J.; Lee, S.; Han, J.; Yoon, S. P.; Nam, S. W.; Choi, S. H.; Lee, K.-Y.; Ham, H. C., Importance of Ligand Effect in Selective Hydrogen Formation Via Formic Acid Decomposition on the Bimetallic Pd/Ag Catalyst from First-Principles. *The Journal of Physical Chemistry C* **2014**, *118*, 22553-22560.
131. Cho, J.; Lee, S.; Yoon, S. P.; Han, J.; Nam, S. W.; Lee, K.-Y.; Ham, H. C., Role of Heteronuclear Interactions in Selective H2 Formation from Hcooh Decomposition on Bimetallic Pd/M (M = Late Transition Fcc Metal) Catalysts. *ACS Catalysis* **2017**, *7*, 2553-2562.
132. Paddison, S. J.; Kreuer, K.-D.; Maier, J., About the Choice of the Protogenic Group in Polymer Electrolyte Membranes: Ab Initio Modelling of Sulfonic Acid, Phosphonic Acid, and Imidazole Functionalized Alkanes. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4530-4542.
133. Vilčiauskas, L.; de Araujo, C. C.; Kreuer, K.-D., Proton Conductivity and Diffusion in Molten Phosphinic Acid (H3po2): The Last Member of the Phosphorus Oxoacid Proton Conductor Family. *Solid State Ionics* **2012**, *212*, 6-9.
134. Clark, J. K.; Paddison, S. J., Proton Dissociation and Transfer in Proton Exchange Membrane Ionomers with Multiple and Distinct Pendant Acid Groups: An Ab Initio Study. *Electrochimica Acta* **2013**, *101*, 279-292.
135. Idupulapati, N.; Devanathan, R.; Dupuis, M., Atomistic Simulations of Perfluoro Phosphonic and Phosphonic Acid Membranes and Comparisons to Nafion. *J. Phys. Chem. B* **2011**.
136. Lawler, R.; Caliendo, C.; Ju, H.; Kim, J. Y.; Lee, S. W.; Jang, S. S., Effect of the Side-Chain Length in Perfluorinated Sulfonic and Phosphoric Acid-Based Membranes on Nanophase Segregation and Transport: A Molecular Dynamics Simulation Approach. *J. Phys. Chem. B* **2020**, *124*, 1571-1580.
137. Sumner, J. J.; Creager, S. E.; Ma, J. J.; DesMarteau, D., Proton Conductivity in Nafion® 117 and in a Novel Bis[(Perfluoroalkyl)Sulfonyl]Imide Ionomer Membrane. *Journal of the Electrochemical Society* **1998**, *145*, 107-110.

138. Tayouo, R.; David, G.; Ameduri, B.; Roziere, J.; Roualdes, S., New Fluorinated Polymers Bearing Pendant Phosphonic Acid Groups. Proton Conducting Membranes for Fuel Cell. *Macromolecules* **2010**, *43*, 5269-5276.
139. Zhang, S. M., A Reliable and Efficient First Principles-Based Method for Predicting Pka Values. 4. Organic Bases. *Journal of Computational Chemistry* **2012**, *33*, 2469-2482.
140. Eckert, F.; Diedenhofen, M.; Klamt, A., Towards a First Principles Prediction of Pk(a): Cosmo-Rs and the Cluster-Continuum Approach. *Molecular Physics* **2010**, *108*, 229-241.
141. Perrin, D. D.; Dempsey, B.; Serjeant, E. P., *Pka Prediction for Organic Acids and Bases*; Chapman and Hall: New York, 1981.
142. Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M., Epik: A Software Program for Pka Prediction and Protonation State Generation for Drug-Like Molecules. *J. Comput. Aided Mol. Des.* **2007**, *21*, 681-91.
143. Milletti, F.; Storchi, L.; Sforna, G.; Cruciani, G., New and Original Pka Prediction Method Using Grid Molecular Interaction Fields. *Journal of Chemical Information and Modeling* **2007**, 2172-2181.
144. Ribeiro, A. R.; Schmidt, T. C., Determination of Acid Dissociation Constants (Pka) Of. Cephalosporin Antibiotics: Computational and Experimental Approaches. *Chemosphere* **2017**, *169*, 524-533.
145. *Biobyte*, BioByte Corporation: Claremont, CA, 2019.
146. *Acd/Pka*, Advanced Chemistry Development ACDLabs: Toronto, 2021.
147. *Chemicalize*, ChemAxon Ltd.: Boston, MA, 2021.
148. Mansouri, K.; Cariello, N. F.; Korotcov, A.; Tkachenko, V.; Grulke, C. M.; Sprankle, C. S.; Allen, D.; Casey, W. M.; Kleinstreuer, N. C.; Williams, A. J., Open-Source Qsar Models for Pka Prediction Using Multiple Machine Learning Approaches. *J. Cheminformatics* **2019**, *11*, 60.
149. Fourches, D.; Muratov, E.; Tropsha, A., Trust, but Verify Ii: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model* **2016**, *56*, 1243-52.
150. Mansouri, K.; Grulke, C. M.; Richard, A. M.; Judson, R. S.; Williams, A. J., An Automated Curation Procedure for Addressing Chemical Errors and Inconsistencies in Public Datasets Used in Qsar Modelling. *SAR QSAR Environ Res* **2016**, *27*, 939-965.
151. Shields, G. C.; Seybold, P. G., *Computational Approaches for the Prediction of Pka Values* CRC Press: Boca Raton, FL, 2014.
152. Ho, J. M., Predicting Pka in Implicit Solvents: Current Status and Future Directions. *Aust. J. Chem.* **2014**, *67*, 1441-1460.
153. Alkorta, I.; Popelier, P. L. A., Linear Free-Energy Relationships between a Single Gas-Phase Ab Initio Equilibrium Bond Length and Experimental Pka Values in Aqueous Solution. *ChemPhysChem* **2015**, *16*, 465-469.
154. Li, M.; Zhang, H.; Chen, B.; Wu, Y.; Guan, L., Prediction of Pka Values for Neutral and Basic Drugs Based on Hybrid Artificial Intelligence Methods. *Sci Rep* **2018**, *8*, 3991.
155. Lee, A. C.; Crippen, G. M., Predicting Pka. *Journal of Chemical Information and Modeling* **2009**, *49*, 2013-2033.
156. Hansch, C.; Maloney, P. P.; Fujita, T., Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178-180.
157. Rupp, M.; Korner, R.; Tetko, I. V., Predicting the Pka of Small Molecules. *Comb. Chem. High Throughput Screen.* **2011**, *14*, 307-327.

158. Thapa, B.; Raghavachari, K., Accurate Pka Evaluations for Complex Bio-Organic Molecules in Aqueous Media. *J. Chem. Theory Comput.* **2019**, *15*, 6025-6035.
159. Liptak, M. D.; Shields, G. C., Accurate Pka Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-N Models Combined with Cpcm Continuum Solvation Methods. *Journal of the American Chemical Society* **2001**, *123*, 7314-7319.
160. Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, M. J., Calibration and Comparison of the Gaussian-2, Complete Basis Set, and Density Functional Methods for Computational Thermochemistry. *J. Chem. Phys.* **1998**, *109*, 10570-10579.
161. Thapa, B.; Schlegel, H. B., Improved Pka Prediction of Substituted Alcohols, Phenols, and Hydroperoxides in Aqueous Medium Using Density Functional Theory and a Cluster-Continuum Solvation Model. *J. Phys. Chem. A* **2017**, *121*, 4698-4706.
162. Alongi, K. S.; Shields, G. C., Theoretical Calculations of Acid Dissociation Constants: A Review Article. *Annu. Rep. Comput. Chem.* **2010**, *6*, 113-138.
163. Xu, L. K.; Coote, M. L., Methods to Improve the Calculations of Solvation Model Density Solvation Free Energies and Associated Aqueous Pka Values: Comparison between Choosing an Optimal Theoretical Level, Solute Cavity Scaling, and Using Explicit Solvent Molecules. *J. Phys. Chem. A* **2019**, *123*, 7430-7438.
164. Lian, P.; Johnston, R. C.; Parks, J. M.; Smith, J. C., Quantum Chemical Calculation of Pk(a)S of Environmentally Relevant Functional Groups: Carboxylic Acids, Amines, and Thiols in Aqueous Solution. *Journal of Physical Chemistry A* **2018**, *122*, 4366-4374.
165. Fraczkiewicz, R.; Lobell, M.; Goller, A. H.; Krenz, U.; Schoenneis, R.; Clark, R. D.; Hillisch, A., Best of Both Worlds: Combining Pharma Data and State of the Art Modeling Technology to Improve in Silico Pka Prediction. *J. Chem. Inf. Model* **2015**, *55*, 389-97.
166. *Admet Predictor*, 7.0; Simulations Plus, Inc.: Lancaster, CA, 2014.
167. Vanysek, P.; Weast, R. C.; Astle, M. J.; Beyer, W. H., *Crc Handbook of Chemistry and Physics*; CRC Press, Inc.: Boca Raton, FL, 1986.
168. Henne, A. L.; Fox, C. J., Ionization Constants of Fluorinated Acids. *Journal of the American Chemical Society* **1951**, *73*, 2323-2325.
169. Kurz, J. L.; Farrar, J. M., The Entropies of Dissociation of Some Moderately Strong Acids. *Journal of the American Chemical Society* **1969**, *91*, 6057-6062.
170. Covington, A. K.; Freeman, J. G.; Lilley, T. H., Ionization of Moderately Strong Acids in Aqueous Solution. I. Trifluoro- and Trichloroacetic Acids. *Journal of Physical Chemistry* **1970**, *74*, 3773-3780.
171. Guthrie, J. P., Hydrolysis of Esters of Oxy Acids: Pka Values for Strong Acids; Brønsted Relationship for Attack of Water at Methyl; Free Energies of Hydrolysis of Esters of Oxy Acids; and a Linear Relationship between Free Energy of Hydrolysis and Pka Holding over a Range of 20 Pk Units. *Canadian Journal of Chemistry* **2017**, *56*, 2342-2354.
172. Ripin, D. H.; Evans, D. A. Evans Pka Table. http://ccc.chem.pitt.edu/wipf/MechOMs/evans_pKa_table.pdf (accessed 14 Jan.).
173. *Jaguar*, Schrödinger: New York, NY, 2008.
174. Kallies, B.; Mitzner, R., Pka Values of Amines in Water from Quantum Mechanical Calculations Using a Polarized Dielectric Continuum Representation of the Solvent. *Journal of Physical Chemistry B* **1997**, *101*, 2959-2967.
175. Lim, C.; Bashford, D.; Karplus, M., Absolute Pka Calculations with Continuum Dielectric Methods. *J. Phys. Chem.* **1991**, *95*, 5611-5620.

176. Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A., Calculation of Absolute and Relative Acidities of Substituted Imidazoles in Aqueous Solvent. *Journal of Physical Chemistry A* **1997**, *101*, 10075-10081.
177. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G., Aqueous Solvation Free Energies of Ions and Ion-Water Clusters Based on an Accurate Value for the Absolute Aqueous Solvation Free Energy of the Proton. *Journal of Physical Chemistry B* **2006**, *110*, 16066-16081.
178. Reiss, H., The Absolute Potential of the Standard Hydrogen Electrode: A New Estimate. *Journal of Physical Chemistry* **1985**, *89*, 4207-4213.
179. Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V., The Proton's Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *Journal of Physical Chemistry A* **1998**, *102*, 7787-7794.
180. Rogstad, K. N.; Jang, Y. H.; Sowers, L. C.; Goddard, W., First Principles Calculations of the Pka Values and Tautomers of Isoguanine and Xanthine. *Chem. Res. Toxicol.* **2003**, *16*, 1455–1462.
181. Ege, S. N., *Organic Chemistry: Structure and Reactivity*, Fifth ed.; CENGAGE Learning: Plymouth, MI, 2018.
182. Henne, A. L.; Fox, C. J., Ionization Constants of Fluorinated Acids. *J. Am. Chem. Soc.* **1951**, *73*, 2323-2325.
183. Kurz, J. L.; Farrar, J. M., The Entropies of Dissociation of Some Moderately Strong Acids. *J. Am. Chem. Soc.* **1969**, *91*, 6057-6062.
184. Goss, K.-U., The Pka Values of Pfoa and Other Highly Fluorinated Carboxylic Acids. *Environ. Sci. Technol.* **2008**, *42*, 456-458.
185. Kortum, G.; Vogel, W.; Andrussow, K., *Dissociation Constants of Organic Acids in Aqueous Solution*; Butterworth: Oxford, 1961.
186. Schuurmann, G.; Cossi, M., Prediction of the Pka of Carboxylic Acids Using the Ab Initio Continuum-Solvation Model Pcm-Uahf. *J. Phys. Chem. A* **1998**, *102*, 6706-6712.
187. Goura, J.; Chandrasekhar, V., Molecular Metal Phosphonates. *Chem. Rev.* **2015**, *115*, 6854-965.
188. Speight, J., *Lange's Handbook of Chemistry*; McGraw Hill: New York, NY, 2005.
189. Sengupta, M.; Pal, K.; Chakravarti, A.; Mahapatra, P., Dissociation Constants of Toluic Acids in Aqueous Solution at Different Temperatures and Study of Related Thermodynamic Parameters. *J. Chem. Eng. Data* **1978**, *23*, 103-107.
190. Burkhard, R. K.; Sellers, D. E.; Decou, F.; Lambert, J. L., The Pka's of Aromatic Sulfinic Acids. *J. Org. Chem.* **1959**, *24*, 767-769.
191. Defilippo, D.; Momicchioli, F., A Study of Benzenesulfinic and Seleninic Acids: Determination and Theoretical Interpretation of Pkappa. *Tetrahedron* **1969**, *25*, 5733-5744.
192. Guthrie, J. P., Hydrolysis of Esters of Oxy Acids: Pka Values for Strong Acids; Brønsted Relationship for Attack of Water at Methyl; Free Energies of Hydrolysis of Esters of Oxy Acids; and a Linear Relationship between Free Energy of Hydrolysis and Pka Holding over a Range of 20 Pk Units. *Can. J. Chem.* **2017**, *56*, 2342-2354.
193. Serjeant, E. P.; Dempsey, B., *Iupac Chemical Data Series No. 23*; Pergamon Press: New York, NY, 1979.
194. Freedman, L. D.; Doak, G. O., The Preparation and Properties of Phosphonic Acids. *Chem. Rev.* **1957**, *57*, 479-523.

195. Kumler, W. D.; Eiler, J. J., The Acid Strength of Mono and Diesters of Phosphoric Acid. The N-Alkyl Esters from Methyl to Butyl, the Esters of Biological Importance, and the Natural Guanidine Phosphoric Acids. *J. Am. Chem. Soc.* **1943**, *65*, 2355-2361.
196. Bonner, O. D.; Prichard, P. R., The Ionization of Trichloroacetic Acid and Evidence for an Unusual Type of Ion Pairing. *J. Solution Chem.* **1979**, *8*, 113.
197. Dean, J. A., *Handbook of Organic Chemistry*; McGraw-Hill Book Co.: New York, NY, 1987.
198. Freedman, L. D.; Doak, G. O., The Preparation and Properties of Phosphonic Acids. *Chemical Reviews* **1957**, *57*, 479-523.
199. Goss, K.-U., The Pka Values of Pfoa and Other Highly Fluorinated Carboxylic Acids. *Environmental Science & Technology* **2008**, *42*, 456-458.
200. Goura, J.; Chandrasekhar, V., Molecular Metal Phosphonates. *Chem. Rev.* **2015**, *115*, 6854-965.
201. Kumler, W. D.; Eiler, J. J., The Acid Strength of Mono and Diesters of Phosphoric Acid. The N-Alkyl Esters from Methyl to Butyl, the Esters of Biological Importance, and the Natural Guanidine Phosphoric Acids. *Journal of the American Chemical Society* **1943**, *65*, 2355-2361.
202. Schuurmann, G.; Cossi, M., Prediction of the Pka of Carboxylic Acids Using the Ab Initio Continuum-Solvation Model Pcm-Uahf. *Journal of Physical Chemistry A* **1998**, *102*, 6706-6712.
203. Serjeant, E. P.; Dempsey, B., *Ionisation Constants of Organic Acids in Aqueous Solution*; Pergamon Press: New York, 1979.
204. Defilippo, D.; Momicchioli, F., A Study of Benzenesulfinic and Seleninic Acids Determination and Theoretical Interpretation of Pk. *Tetrahedron* **1969**, *25*, 5733-5744.
205. Burkhard, R. K.; Sellers, D. E.; Decou, F.; Lambert, J. L., The Pka's of Aromatic Sulfinic Acids. *J. Org. Chem.* **1959**, *24*, 767-769.
206. Muckerman, J. T.; Skone, J. H.; Ning, M.; Wasada-Tsutsui, Y., Toward the Accurate Calculation of Pka Values in Water and Acetonitrile. *Biochim Biophys Acta* **2013**, *1827*, 882-91.
207. Kuhn, M.; Johnson, K., *Applied Predictive Modeling*; Springer: New York, 2013.
208. Duboue, P., *The Art of Feature Engineering : Essentials for Machine Learning*; Cambridge University Press, 2020.
209. Becke, A. D., A New Mixing of Hartree–Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
210. Adamo, C.; Barone, V., Toward Reliable Density Functional Methods without Adjustable Parameters: The Pbe0 Model. *Journal of Chemical Physics* **1999**, *110*, 6158-6170.
211. Ernzerhof, M.; Scuseria, G. E., Assessment of the Perdew-Burke-Ernzerhof Exchange-Correlation Functional. *Journal of Chemical Physics* **1999**, *110*, 5029-5036.
212. Marenich, A. V.; Olson, R. M.; Kelly, C. P.; Cramer, C. J.; Truhlar, D. G., Self-Consistent Reaction Field Model for Aqueous and Nonaqueous Solutions Based on Accurate Polarized Partial Charges. *Journal of Chemical Theory and Computation* **2007**, *3*, 2011-2033.
213. Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A.; Honig, B., Accurate First Principles Calculation of Molecular Charge-Distributions and Solvation Energies from Ab-Initio Quantum-Mechanics and Continuum Dielectric Theory. *Journal of the American Chemical Society* **1994**, *116*, 11875-11882.
214. Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B., New Model for Calculation of Solvation Free Energies: Correction of

Self-Consistent Reaction Field Continuum Dielectric Theory for Short-Range Hydrogen-Bonding Effects. *Journal of Physical Chemistry* **1996**, *100*, 11775-11788.
215. Wang, Y.; Ruiz Diaz, D. F.; Chen, K. S.; Wang, Z.; Adroher, X. C., Materials, Technological Status, and Fundamentals of Pem Fuel Cells – a Review. *Materials Today* **2020**, *32*, 178-203.
216. Yu, X.; Ye, S., Recent Advances in Activity and Durability Enhancement of Pt/C Catalytic Cathode in Pemfc. *Journal of Power Sources* **2007**, *172*, 133-144.
217. Sealy, C., The Problem with Platinum. *Materials Today* **2008**, *11*, 65-68.
218. Xu, F.; Mu, S.; Pan, M., Recycling of Membrane Electrode Assembly of Pemfc by Acid Processing. *International Journal of Hydrogen Energy* **2010**, *35*, 2976-2979.
219. Ohma, A.; Mashio, T.; Sato, K.; Iden, H.; Ono, Y.; Sakai, K.; Akizuki, K.; Takaichi, S.; Shinohara, K., Analysis of Proton Exchange Membrane Fuel Cell Catalyst Layers for Reduction of Platinum Loading at Nissan. *Electrochimica Acta* **2011**, *56*, 10832-10841.
220. Kang, H.; Kwon, S. H.; Lawler, R.; Lee, J. H.; Doo, G.; Kim, H.-T.; Yim, S.-D.; Jang, S. S.; Lee, S. G., Nanostructures of Nafion Film at Platinum/Carbon Surface in Catalyst Layer of Pemfc: Molecular Dynamics Simulation Approach. *The Journal of Physical Chemistry C* **2020**, *124*, 21386-21395.