Data-driven and Simulation-assisted Synthesis of Hydrocarbon Polymer Electrolyte Membranes

Doctoral Thesis

by

Tushita Rohilla

(2016MSZ0002)

DEPARTMENT OF MECHANICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY ROPAR

December 2023

Data-driven and Simulation-assisted Synthesis of Hydrocarbon Polymer Electrolyte Membranes

A Thesis Submitted

in Partial Fulfilment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Tushita Rohilla

(2016MSZ0002)

DEPARTMENT OF MECHANICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY ROPAR
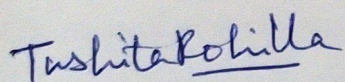
December 2023

ii

Dedicated to
Maa Saraswati
(The Bestower of All Knowledge and Wisdom)
and my beloved family

Declaration

I hereby declare that the work being presented in the thesis entitled **Data-driven and Simulation-assisted Synthesis of Hydrocarbon Polymer Electrolyte Membranes** has been solely authored by me. It presents the result of my own independent investigation/research conducted during the time period from July 2016 to December 2022 under the supervision of Dr.Dhiraj K. Mahajan, Associate Professor, Department of Mechanical Engineering, Indian Institute of Technology Ropar. To the best of my knowledge, it is an original work, both in terms of research content and narrative, and has not been submitted or accepted elsewhere, in part or in full, for the award of any degree, diploma, fellowship, associateship, or similar title of any university or institution. Further, due credit has been attributed to the relevant state-of-the-art collaborations (if any) with appropriate citations and acknowledgments, in line with established ethical norms and practices. I also declare that any idea/data/fact/source stated in my thesis has not been fabricated/ falsified/ misrepresented. All the principles of academic honesty and integrity have been followed. I fully understand that if the thesis is found to be unoriginal, fabricated, or plagiarized, the Institute reserves the right to withdraw the thesis from its archive and revoke the associated Degree conferred. Additionally, the Institute also reserves the right to appraise all concerned sections of society of the matter for their information and necessary action (if any). If accepted, I hereby consent for my thesis to be available online in the Institute's Open Access repository, inter-library loan, and the title & abstract to be made available to outside organizations.

Tushita Rohilla

Signature

Name: Tushita Rohilla
Entry Number: 2016MSZ0002
Program: Ph.D.
Department: Department of Mechanical Engineering
Indian Institute of Technology Ropar
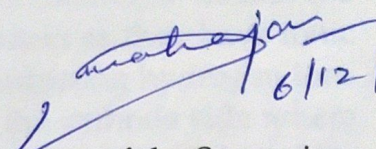Rupnagar, Punjab 140001
Date: 07/12/2023

Certificate

This is to certify that the thesis entitled **Data-driven and Simulation-assisted Synthesis of Hydrocarbon Polymer Electrolyte Membranes**, submitted by **Tushita Rohilla (2016MSZ0002)** for the award of the degree of **Doctor of Philosophy** of Indian Institute of Technology Ropar, is a record of bonafide research work carried out under my guidance and supervision. To the best of my knowledge and belief, the work presented in this thesis is original and has not been submitted, either in part or full, for the award of any other degree, diploma, fellowship, associate, or similar title of any university or institution.

In my opinion, the thesis has reached the standard of fulfilling the requirements of the regulations relating to the Degree.

Signature of the Supervisor
Dr. Dhiraj K. Mahajan
Department of Mechanical Engineering
Indian Institute of Technology Ropar
Rupnagar, Punjab 140001

Date: 6/12/2023

Lay Summary

The atmosphere is polluted and warming up rapidly due to the continuous use of fossil fuels in vehicles and to generate electricity. We need to replace the present sources of energy and adopt cleaner and more efficient sources to provide power to transport and energy-intensive infrastructure like data centers. Fuel cells convert electrochemical energy into usable electrical energy. Many cars, buses, trucks, trains, aircraft, drones, and forklifts run on fuel cells these days. These use hydrogen as fuel instead of conventional fossil fuels like petrol, diesel, and gasoline, and emit only water as the by-product rather than pollutants and greenhouse gases. In this way, these are greener, cleaner, and more efficient means of energy conversion that have a totally silent operation, unlike conventional IC engines. Fuel cells have several components which are made from different materials. The main component is the electrolyte membrane that is made from a special class of polymers known as ionomers as they have ionic groups along the chain. These polymers are specialized in conducting hydrogen ions or protons from the anode side where they are generated to the cathode side where they combine with oxygen to complete the reaction. The chemical design of electrolyte membranes decides how good or bad these membranes are at transporting protons. Generally, the commercially available fluorinated electrolyte membranes are good at conducting protons but only when conditions of humidity and temperature are right, not always. In contrast to these, hydrocarbon membranes can work efficiently over a wide range of humidity and temperatures and even are lesser in cost than fluorinated membranes. One such hydrocarbon ionomer class is Sulfonated Polyimides (SPIs) which have good thermal and mechanical stability. Usually, it is seen that experimentation to develop electrolyte membranes takes years and is very time-consuming and resource-intensive. Also, there often occurs a loss of effort as researchers are not able to get the desired levels of proton transport in these electrolyte membranes. However, in this age of artificial intelligence, we can use machine learning to train models based on past data to design newer materials for a targeted application. By doing that, we can also gather insights into what makes a good electrolyte membrane. This work does exactly that. Here tree-based simple machine learning algorithms have been trained using the past SPI data related to its chemical structure and proton conductivity to help in discovering newer chemical structures that can prove to be good electrolyte membranes for fuel cells. Moreover, the findings are evaluated by popular computational techniques of materials modeling and lab-scale experimental synthesis and testing. A good agreement was found between the results of all three techniques used in this work which point to the usefulness of this integrated methodology in predicting the electrolyte materials for fuel cells in other classes of hydrocarbon ionomers.

Abstract

Polymer Electrolyte Membrane Fuel Cells (PEMFCs) are versatile energy devices that provide useable electrical energy for a wide range of stationary and automotive applications. Polymer Electrolyte Membranes (PEMs) are solid ionomeric polymer membranes that conduct proton (H^+) from anode to cathode. These proton-conducting electrolyte membranes are the most important component of the fuel cell as the rate of proton transport taking place in the water channels governs the overall performance of the PEMFC. Currently, the commercially available perfluorinated membranes have limitations at higher temperatures and low humidity operations as usually faced in automobile applications. Sulfonated Polyimides (SPIs) are a class of versatile hydrocarbon ionomeric polymers that are being explored as a polymer electrolyte material for fuel cells due to their superior thermal and mechanical stability. Experimentation to discover alternative PEMs is extensive time-consuming and resource-intensive. There often occurs a loss of effort as the proton conductivity of the developed PEM is not able to be at par with the perfluorinated PEMs.

Also, there is a lack of investigations into the correlation of the nano-scale morphology of the PEM with the behavior of proton transport in SPI PEMs. Previously, researchers have gained some understanding of structure-property interplay through multi-scale computational models and extensive experimental synthesis and testing of SPI-based PEMs. However, combined efforts supported by the simulation-informed synthesis of hydrocarbon-based PEM can provide an understanding of the hydrocarbon structure-property relationship that is still elusive. Moreover, data-driven polymer discovery is a promising method of selecting polymers for target applications. Looking at the potential of data-driven polymer discovery of novel polymers as PEM, this thesis combines the effort put into ML-based identification of potential novel SPI PEMs

as an alternative to Nafion and their validation using MD simulation and extensive experimentation. The thesis work was divided into three different objectives.

In the first objective of the present work, a data set was prepared to comprise the physicochemical properties and proton conductivity data of SPI-based PEMs collected and organized from the reported literature. The data set also included the chemical structures of the repeat units of the SPIs in computer-parsable SMILES format. Semi-empirically calculated properties and Quantitative structure-property relationships(QSPR) properties were also included in the data set. Decision trees were trained to obtain certain rules for designing novel PEMs whose high proton conductivity could be ascertained with a high accuracy rate even before synthesizing them. Thus, following the rules, one SPI PEM, namely 1,4,5,8-naphthalene tetracarboxylic dianhydride/ 4,4'-diamino stilbene-2,2-disulfonic acid/4,4'-Diaminodiphenyl methane (NTDA/DSDSA/MDP) was designed for computational modeling using Molecular Dynamics(MD).

In the second objective, all-atom molecular dynamics simulations were used to model the nano-phase segregation, the morphology of the ionic domains, and the dynamics of proton transport in a novel hydrocarbon-based PEM identified through the work done in the first objective. The diffusion coefficients of hydronium ions and water molecules, radial distribution function (RDF) plots between sulfur atoms of sulfonate groups and solvent phase (hydronium ions and water molecules), as well as polymer solvent volume fractions and fractional free volumes, have been calculated at increasing levels of hydration (λ = 1, 5, 10 and 15) to understand the proton transport in the novel SPI PEM ionomers

In the final objective, the NTDA/DSDSA/MDP SPI membrane, identified through data-driven and computational techniques mentioned in the first and second objectives, was synthesized and proton conductivity was determined. Proton conductivity was found to be in the range of 0.1588 - 0.28636 S·cm^{-1} which is exceptionally well for a PEM while those obtained in the MD simulations were 0.03 - 0.18 S·cm^{-1}. Thus, a good agreement was observed between the proton conductivity values predicted using MD simulations and the values for stable stand-alone SPI PEMs.

Acknowledgements

I am grateful to the Almighty for giving me a supportive family and the opportunity to study at a premier institution. The constant encouragement and support of my family and friends helped me to complete this study with determination. I wish to express humble gratitude to my Ph.D. advisor, Dr. Dhiraj K. Mahajan for giving me this opportunity to design new polymer electrolytes for fuel cells and learn new skills for the same. I highly appreciate his patience and belief in my abilities while I was exploring new and advanced subjects of machine learning and materials modeling. His insights into my work have always been very helpful and valuable.

I wish to thank our collaborators from other departments who provided their expert advice for the completion of this interdisciplinary study. Firstly, Prof. Narinder Singh of the Department of Chemistry for his guidance related to polymer synthesis and characterization. Secondly, Dr. C. K. Narayanan for his constant monitoring and guidance which helped in completing the first objective of this study related to machine learning. Also, I wish to thank the members of my Doctoral Committee - Dr. Ravi Mohan Prasad, Dr. Prabhat K. Agnihotri, and the chairman, Prof. Navin Kumar.

I also want to thank my extended family in Ropar who made this journey incredible and memorable. I am indebted to Mr. Dinesh Singh and his family for their constant faith in me and their support during the tough times of the Covid-19 lockdown. I am highly grateful to our landlords for their love, kindness, and all the help they provided to make us feel at home in Ropar. I wish to acknowledge the support of my friends in the Chemistry department namely, Beant Kaur Billing, Monica Chaudhary, Richa Rani, and Mayank Sharma. I am extremely thankful to my friends - Mukesh Kumar and his wife, Dipika Kumari, Amanpreet Chander, Reshma Tudu, Binod Kumar, Isha Sharma, Shweta Dey, Swati Garg, and Ranjana Roy Chowdhury for their support and cheers, especially during the tough times.

I wish to extend deep gratitude to Dr. Ahmad Husain for his expertise in polymer synthesis which helped me in the development of robust PEM membranes and his genuine guidance. Also, I wish to extend thanks to the B.Tech. students who provided their enthusiastic support in the machine learning work (Mr. Ambuj Ghorai and Mr. Jigar Nilesh Mehta) and the experimental work (Mr. Rishabh Anchalia). I wish to thank Mr. Rupinder Singh for his support and expertise in developing setups for polymer synthesis and for his help with the fuel cell connections.

x

List of Publications

1. Rohilla, T., Singh, N., Krishnan, N.C., Mahajan, D.K. **Designing sulfonated polyimide-based fuel cell polymer electrolyte membranes using machine learning approaches**, Computational Materials Science, 2023 Feb 25;219:111974.

2. Rohilla, T., Husain, A., Singh, N., Mahajan, D.K. (2022), **Atomistic simulation and synthesis of novel sulfonated Polyimide polymer electrolyte membranes with facile proton transport**, Chemical Engineering Journal, 2023 Oct 15;474:145727.

Contents

List of Figures

List of Tables

List of Abbreviations

AFM	Atomic Force Microscopy
BMU	Best Matching Unit
DMFC	Direct methanol Fuel Cell
DSC	Differential Scanning Calorimetry
DSDSA	4,4′-Diamino Stilbene-2,2′-Disulfonic Acid
EIS	Electrochemical Impedance Spectroscopy
EW	Equivalent Weight
FFV	Fractional Free Volume
FTIR	Fourier Transform Infrared
GDL	Gas Diffusion Layer
HFBAPP	2-bis[4-(4-aminophenoxy) phenyl] Hexafluoropropane
HOR	Hydrogen Oxidation Reactions
IEC	Ion Exchange Capacity
MDP	4,4-Diamino-2,2-diphenyl Methane
MEA	Membrane Electrode Assembly
MSD	Mean Squared Displacement
NMR	Nuclear Magnetic Resonance
NTDA	1,4,5,8-Naphthalene Tetracarboxylic Dianhydride
OCV	Open Circuit Voltage
ORR	Oxygen Reduction Reaction
PEMFC	Polymer Electrolyte Membrane Fuel Cell
PFGSE	Pulsed Field Gradient Spin-Echo
QENS	Quasi-Elastic Neutron Scattering
RDF	Radial Distribution Function
RH	Relative Humidity
SAXS	Small-Angle X-ray Scattering
SEM	Scanning Electron Microscope
SOM	Self-Organization Maps
SPI	Sulfonated Polyimide
TGA	Thermogravimetric Analysis
TEM	Transmission Electron Microscopy
WAXD	Wide-Angle X-ray Diffraction

List of Symbols

$\alpha(t)$	Learning rate
$T_{j,I(x)}(t)(x_i - w_{ji})$	Gaussian neighborhood function
N_B	Number of B particles in the system
n_B	Number of atoms B surrounding atom A in a shell of certain thickness
e	Electron Charge
N	Number of charged ions in the system
$D_{H_3O^+}$	Diffusion coefficient of hydronium ions [cm^2·s^{-1}]
D_{H_2O}	Diffusion coefficient of water molecules [cm^2·s^{-1}]
V	Volume of the cell [Å^3]
$r_j(O)$	Initial coordinates of the particle
$r_j(t)$	Coordinates of the particle at time t
N	Total number of particles in the system
w_i	Weight vector
$\sigma_{||}$	Proton conductivity in the direction parallel to the plane [S·cm^{-1}]
t	Thickness of the membrane [cm]
T	Absolute temperature [K]
E_a	Activation energy for proton conduction [kJ.mol^{-1}]
σ_o	Pre-exponential factor
$\sigma_{||}$	Proton conductivity [S·cm^{-1}]
l	Distance between electrodes [cm]
R	Resistance observed from Nyquist plot [ohm]
A	Area of the sample [cm^2]
IEC_w	Ion exchange capacity by weight [meq·g^{-1}]
V_{NaOH}	Volume of NaOH consumed during titration [ml]
N_{NaOH}	Normality of NaOH
m_{wp}	Mass of polymer equilibrated with water [g]
m_{dp}	Mass of the dry polymer [g]
l_{dp}	Length of the dry polymer [mm]
l_{wp}	Length of the polymer equilibrated with water [mm]
Δl	Change in the length of the polymer sample due to absorption of water
t_{dp}	Thickness of the dry polymer [mm]
t_{wp}	Thickness of the polymer equilibrated with water [mm]
Δt	Change in the thickness of the polymer sample due to absorption of water
WU	Water uptake [%]

1 Introduction

1.1 Background

The world is confronting the constantly rising dangers of global warming and pollution caused by the continuous use of fossil fuels for over two centuries, leading to climate change and scarcity of energy resources (Manabe and Wetherald, 1967). To counter this, consistent research efforts are being directed to find more efficient and cleaner alternatives to replace the conventional sources of energy, and many renewable sources of energy are being explored and adopted (Jacobson, Colella, and Golden, 2005). Polymer electrolyte membrane fuel cells (PEMFCs) have emerged as clean energy devices that are non-polluting and non-greenhouse gas emitting and have high efficiency of energy conversion and silent operation (Larminie, Dicks, and McDonald, 2003). A schematic of a typical Hydrogen/Air PEM fuel cell is shown in Fig. 1.1. PEMFC consists of a pair of bipolar plates (BPs) and Gas diffusion electrodes (GDEs) coated with nanoparticles of Platinum/Carbon electrocatalyst. BPs have flow channels for the even distribution of gaseous reactants (fuel at the anode and oxidant at the cathode) over the micro-porous, hydrophobic, and electrically conductive gas diffusion electrodes.

PEMFC generates electrical energy from the continuous electro-chemical reactions taking place at the anode and cathode. Oxidation of hydrogen occurs at the anode (Eq. 1.1) and oxygen reduction reaction at the cathode (Eq. 1.2) proceeds with the protons obtained through the electrolyte membrane while electrons reach the cathode through an external circuit.

$$[\textbf{Anode}] : 2\,H_2 \longrightarrow 4\,H^+ + 4\,e^- \tag{1.1}$$

$$[\textbf{Cathode}] : O_2 + 4\,H^+ + 4\,e^- \longrightarrow 2\,H_2O \tag{1.2}$$
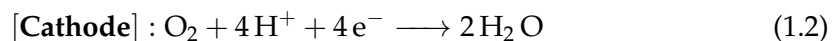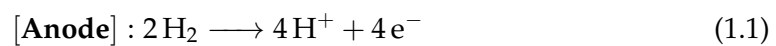
1.2   Polymer Electrolyte Membranes (PEMs)

The versatile design of PEMFCs and their potential for utilization in the wide-ranging automobile and stationary applications depend on the solid polymer electrolyte membrane (PEM), invented in the 1960s, which is the most important component of a PEMFC (Steele and Heinzel, 2011; Rikukawa and K, 2000). Polymer electrolyte membrane (PEM) is the most vital component of a PEMFC which is designed to selectively transport an ionic species generated during the electrochemical reaction and prevents fuel and oxidant from coming into direct contact. Specifically, a PEM is an ionomeric polymer that conducts hydrogen ions or protons generated due to the hydrogen oxidation reaction at the anode towards the cathode where it is utilized in the reduction reaction with oxygen to produce water. Ideally, PEM behaves as an insulator for electrons and also possesses high structural integrity and resistance to degradation arising owing to the harsh operating conditions. It should possess mechanical stability for withstanding stresses arising due to varying hydration conditions (Hickner, Ghassemi, et al., 2004).
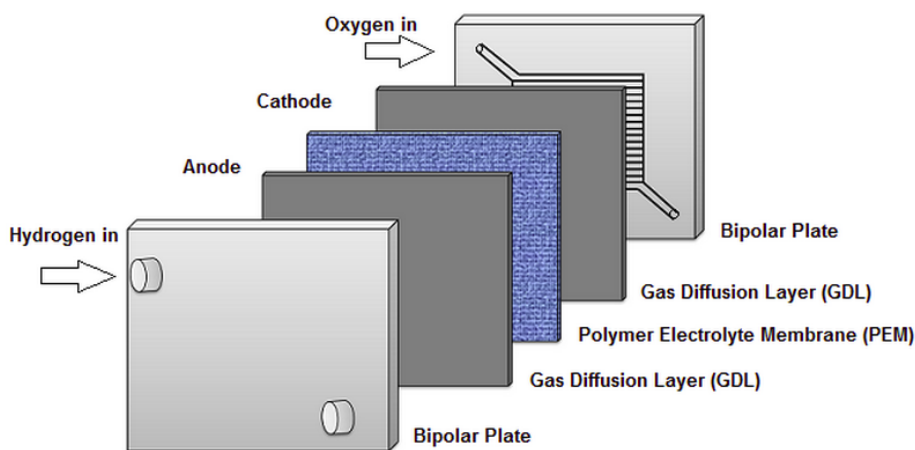
FIGURE 1.1: A schematic showing the components of a Polymer Electrolyte Membrane Fuel cell (PEMFC)

Polymer electrolyte membrane (PEM) is the most vital component of a PEM Fuel cell (PEMFC) which is designed to selectively transport an ionic species generated during the electrochemical reaction and prevents fuel and oxidant from coming into direct contact. Fluorinated backbone and side chain sequences are of different lengths in two

popular commercial Perfluorinated Sulfonic Acids (PFSAs)- Nafion® and Aquivion® ionomers (developed by DuPont and Solvay respectively) are shown in Fig. 1.2(Zhang and Shen, 2012). These perfluorinated ionomeric polymers possess a hydrophobic backbone, usually derived from Polytetrafluoroethylene (PTFE), decorated with side chains bearing hydrophilic acidic groups usually $-SO_3H$, $-PO_4$ (Rikukawa and K, 2000). Usually, the long or short side chains of PFSAs comprise multiple ethers (-C-O-C) and ($-CF_2$) groups terminating in acidic and hydrophilic sulfonic acid ($-SO_3H$) groups. This chemical composition leads to a characteristic microstructure that conducts proton on coming in contact with moisture (A., 1970). As shown in Fig. 1.2, these PFSAs are ionomeric polymers that possess hydrophobic backbone comprising of fluorine atoms and flexible side chains comprising ether groups and acidic sulfonate groups a the ends of the chains which are capable of binding water molecules around them weak and strong hydrogen bonds. This chemistry gives them unique microstructural segregation and characteristic morphology around inter-connected water channels in which large-scale proton transport takes place(Eisenberg, 1970). Thus, they have high proton conductivity in the order of 0.1 S.cm^1 in the operational temperature range of 50-80 °C and relative humidity range of 50% to 100%(Sone, Ekdunge, and Simonsson, 1996). The exorbitant cost and limited performance in low humidity and high-temperature operation of perfluorinated ionomers are the main reasons for the search for non-perfluorinated polyelectrolytes with better performance.

1.3 Interaction between PEM and water

The interaction of ionomers with water is not the same as other polymers. In non-ionomeric polymers, there is no chemical binding site for water molecules therefore the polymer chains interact according to Flory Huggins's theory of polymer solutions. However, in ionomers, the presence of acidic sites along the polymer chains alters the interaction behavior. Mechanism of water uptake in PEMs has been theorized using techniques that reveal their morphological details such as SAXS (Essafi, Gebel, and Mercier, 2004a). Fig. 1.3 describes the process of water uptake by PEMs. Water molecules tend to seep into empty pockets created by the entanglement of chains,
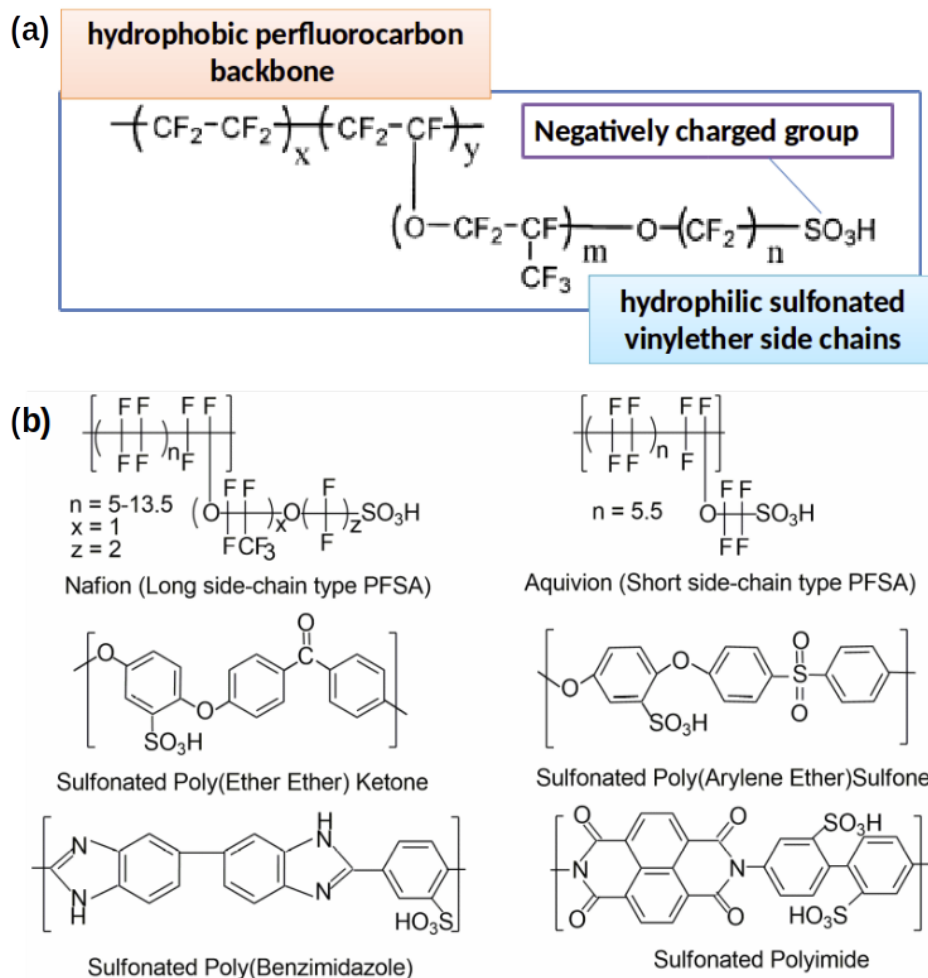
FIGURE 1.2: Hydrophobic and hydrophilic parts of ionomeric Perfluorinated Sulfonic Acid Polymer electrolyte membranes(PEMs) are shown; Chemical structures of the two PFSAs- Nafion®, Aquivion® and four Hydrocarbon PEMs Sulfonated Poly(Ether Ether) Ketone (SPEEK), Sulfonated Poly(Arylene Ether) Sulfone (SPAES), Sulfonated Poly(Benzimidazole) (SPBI) and Sulfonated Polyimide (SPI) have been shown

getting bound to ionic sites. As there is a rise in humidity, there is a tendency to create more water clusters in SPIs rather than enlargement of existing clusters as seen in Nafion®. At some threshold water level, the domains become interconnected with each other which leads to the initiation of proton transport. This is according to the percolation theory of water cluster formation and proton conduction in ionomers.

The polymer is brought in contact with a few molecules of water initially, phase segregation starts taking place at the nano-level. Hydrophobic and hydrophilic zones are demarcated and as hydration increases, interconnected ionic channels are formed. Initially, the few molecules of water are bound by the sulfonate groups but as hydration

increases the content of bulk water increases and becomes more than the content of bound water as shown in Fig. 1.3.

1.3.1 Mechanism of proton transport in PEMs

The characteristic chemical structure results in morphology that has overlapping water clusters or inter-connected ionic domains that are conducive for facile proton transport and thus demonstrates high proton conductivity with a sufficient level of hydration (Eikerling, Kornyshev, and Spohr, 2008). Transport properties or dynamic properties depend upon the structure of ionic domains or water domains formed around the ionic groups with simultaneous rearrangement of the polymer matrix. Chains rearrange to accommodate the increasing number of water molecules. Generally, aqueous ionic domains are connected to each other through narrow channels. In PFSAs, especially Nafion, different models have been proposed for explaining the nano-phase segregated morphology observed through x-ray and neutron scattering techniques. According to Small-angle X-ray Scattering (SAXS) studies, spherical ionic clusters or inverted micelles are interconnected by narrow water channels of diameter 1 nm spanning to the length of about 4-5 nm forming cluster-network morphology (Gierke, Munn, and Wilson, 1981). Similarly, the sandwiched core-shell model has been proposed based on Small-angle Neutron Scattering (SANS) studies (Haubold, Vad, et al., 2001). Understanding the interdependence between the structural aspect of the morphology and topological distribution of water clusters is crucial for new PEM design and development. Phase separation and resulting water channel morphology are closely related to the atomic-level details of the PEM architecture including the chemical structure of the repeat unit, nature of side-chains and main chains of the PEM, immediate chemical environment of the sulfonate groups and arrangement of the sulfonate groups with respect to each other. Generally, water molecules and hydronium ions are present in close vicinity of the sulfonate groups. Water molecules form hydrogen bonds of varying strength with the sulfonate groups as well as excess protons whereas hydronium ions interact with the sulfonate groups due to electrostatic forces. In a well-developed micro-phase segregated morphology, water clusters and channels are formed. In such water clusters and channels, water exists in two prominent states depending on the distance from the sulfonate group towards the center of the water channels and the overall

hydrogen bonding environment. Bound or non-freezable water is strongly bound to the sulfonate group through strong hydrogen bonds and bulk-like or freezable water is present in the central region of the water pore where both water molecules and hydronium ions tend to be highly mobile (Paddison, 2003).

On the basis of experimental and computational investigations into bulk water and ionomers, three prominent mechanisms of proton transport have been identified to be occurring- 1) Grotthus mechanism or proton hopping which involves traveling of the hydrogen cation or proton through continuous bond creating and breaking over a network of hydrogen-bonded water molecules [aka Structural diffusion] (Agmon, 1995), 2) Transport of hydrated proton in the form of hydronium ion or water molecule with one excess proton covalently bonded to the oxygen atom through conventional Fickian or sub-Fickian diffusion regime [Vehicular or Classical diffusion], and 3) Transport of protons along/in the interfacial region of polymer phase and solvent phase (water channel or cluster) wherein the proton travels along the series of adjacent sulfonate groups at low hydration conditions [Surface or en masse diffusion] (Eikerling, Kornyshev, and Spohr, 2008; Choi, Jalani, and Datta, 2005). Proton hopping takes place on a time scale of about 1.5 ps and is significantly faster than the time associated with vehicular diffusion(Eikerling, Kornyshev, et al., 2001). Besides these mechanisms of proton transports, fast and slow diffusive regimes have been identified to supplement the understanding based on the nano-scale confinement of protons in the water channels and clusters formed in hydrated ionomers(Berrod, Hanot, et al., 2017).

1.4   Search for an ideal PEM material

Despite their suitable chemical architecture, PFSAs have limitations in high-temperature and low-humidity operations. Also, there are concerns related to their high cost of production, a high crossover of fuel across them, and their detrimental effect on the environment (Roziere and Jones, 2003). These concerns have led researchers to develop alternatives for these PFSAs. In contrast to PFSAs, these are well-suited for high temperature and low humidity fuel cell operations and their production cost is also much lesser than the PFSAs (Kreuer, 2010). Hydrocarbon (HC) ionomeric polymers
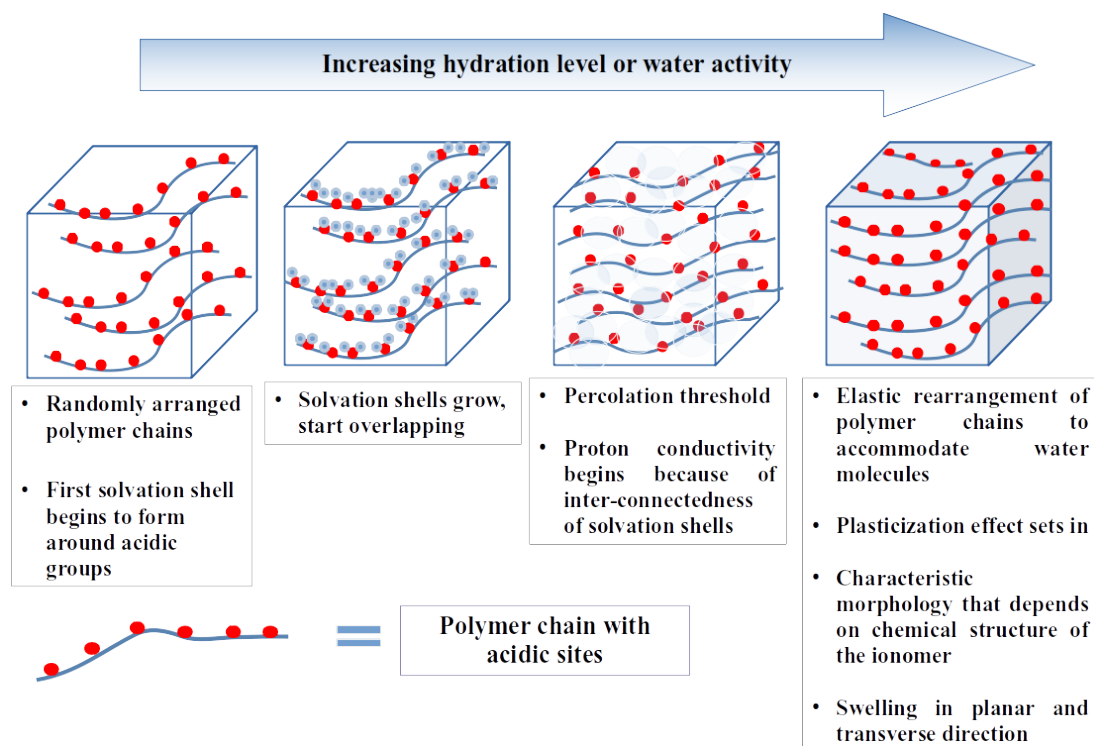
FIGURE 1.3: Mechanism of formation of ionic channels with increasing
water activity in ionomers

are specialized classes of non-fluorinated or partially fluorinated polymers for con-
ducting protons (hydrogen ions) in fuel cells. The central issue in designing HC PEMs
is the successful realization of the proton conductive morphology at the micro-scale
which supports facile proton transport on a large scale which is reflected by the high
proton conductivity of the ionomer and better overall performance of the fuel cell. Re-
searchers have reported many kinds of hydrocarbon PEMs such as Sulfonated Poly
(Ether Ether Ketone) (SPEEK), Sulfonated Poly (Aryl Ether Sulfone) (SPAES), and Sul-
fonated Polybenzimidazole(SPBI) have been reported(Sambandam and Ramani, 2007;
Park, Kim, et al., 2011; Li, He, et al., 2004; Hickner and Pivovar, 2005; Maier and Meier-
Haack, 2008; Zhang and Shen, 2012). However, it is important to identify promising
PEM candidates among the plethora of HC PEMs reported over the years. Fig. 2.13
shows the general chemical structure of a PFSA ionomer and the repeat units of two
PFSA ionomers- Nafion and Aquivion and four sulfonated hydrocarbons – SPEEK,
SPAES, SPBI, and SPI.

1.4.1 Hydrocarbon PEMs

The development of electrolyte membranes depends upon the targeted application. PEMs are expected to function optimally over a wide range of temperatures. Automobile application poses the real test on PEM performance as it is exposed to extremities of temperature- from sub-zero in cold start conditions to 120-160 °C in high current density conditions. Moreover, humidity conditions also vary during operation. Hydrocarbon ionomers offer comparable or better performance than PFSAs (Lee, Hwang, et al., 2006; Saito, Miyatake, and Watanabe, 2008). With some modifications, their performance can surpass the limits imposed by humidity and temperature conditions during the operation (Lee, Ogawa, et al., 2010). Sulfonated polymers are predominant for medium-temperature fuel cells (60-120 °C) while phosphoric acid (PA)-doped polymers have shown promise for High-Temperature PEMFC (HT-PEMFC) (Fang, Qiao, et al., 2015a). Hydrocarbon ionomers are typically non-fluorinated but can be semi-fluorinated optionally (Mistri, Mohanty, and Banerjee, 2012). Polyimide is a class of polymers that was first reported/synthesized by Dupont in the 1960s which was the same as the time of Nafion. This class of polymer found application in NASA's high-temperature mission due to extremely high thermal resistance. Sulfonated Polyimides were first reported around the 2000s.

1.4.2 Sulfonated Polyimide PEMs

One of these hydrocarbon PEMs is the class of Sulfonated Polyimide (SPI) which is a thermally-resistant hydrocarbon polymer possessing superior mechanical properties and chemical resistance This HC PEM was first reported in the early 2000s as a suitable candidate for PEM application(Guo, Fang, et al., 2002). SPIs also offer high mechanical strength, thermal resistance, and good phase segregation, proton conductivity, resistance to fuel crossover upon appropriate modification (Pu, 2014; Akbarian-Feizi, Mehdipour-Ataei, and Yeganeh, 2010a; Peighambardoust, Rowshanzamir, and Amjadi, 2010). Chemically, an SPI copolymer repeat unit comprises two parts namely, a dianhydride with sulfonated diamine and a dianhydride with a non-sulfonated diamine. These polymers imparted proton-conducting functionality by incorporating acidic functional groups along the backbone or in the side chains. Usually, the repeat

unit is designed by carefully selecting these chemical moieties in order to obtain favorable morphology with water uptake and thus high proton conductivity (Fang, Qiao, et al., 2015b). SPIionomers show better proton conductivity, especially at higher temperatures (Akbarian-Feizi, Mehdipour-Ataei, and Yeganeh, 2010b). Also, these membranes have shown competitive performance in H_2/Air fuel cells(Yin, Hayashi, et al., 2005a). This can be attributed to the fact that SPIs have imide groups formed from the condensation polymerization of dianhydride and diamine monomers. Such chemical structure favors the formation of rigid chains and the presence of Nitrogen atoms enhances hydrogen bonding stabilization as they form charge transfer complexes(Fang, Qiao, et al., 2019). SPIs can have a good range of proton conductivity with reduced Hydrogen, Oxygen, and Methanol crossover (Saito, Miyatake, and Watanabe, 2008; Yaguchi, Chen, et al., 2010; Woo, Oh, et al., 2003; Einsla, Kim, et al., 2005; Ganeshkumar, Bera, et al., 2014; Ito, Tanaka, and Kawakami, 2018). They are capable of delivering high Open Circuit Voltage (OCV) 0.9 V (Yin, Hayashi, et al., 2005b) and power density of 0.764 W.cm^{-2} (Lin, Ho, et al., 2013) in H_2/O_2 and Direct Methanol fuel cells respectively. The main challenges in SPI-based PEMs are related to hydrolytic stability and processability for commercial scale (Akbarian-Feizi, Mehdipour-Ataei, and Yeganeh, 2010a).

1.5 Integration of computational and experimental effort

Extensive experimental studies have been performed and reported but at the expense of a lot of effort and resources (Cornet, Diat, et al., 2000; Ye, Bai, and Ho, 2006b). Only a few studies have utilized scattering techniques to investigate the structure-property interplay (Essafi, Gebel, and Mercier, 2004b). The proton conductivity of SPIs has been experimentally and theoretically examined in a number of studies however, the shape and size of ion transport channels and the nature of nanophase segregation as well as resulting morphology are still less established compared to the well-studied conventional PFSAs especially Nafion (Shin, Guiver, and Lee, 2017). In general, the development and testing of a novel PEM is a critical polymer discovery problem that requires enormous effort. Designing, synthesis, characterization, and testing of an ionomer is a time-consuming process and there is little certainty with regard to the final fuel

cell performance outcome of the synthesized PEM. Computational screening of the prospective hydrocarbon PEM candidates using Molecular Dynamics is an extensive technique that often requires coarse-graining of the molecules due to the large number of atoms involved in the chain(Hu, Lu, and Guo, 2018).

The work presented in this thesis attempts to integrate the predictive abilities of data analysis and machine learning techniques along with the understanding gained from computational atomistic modeling to discover high-throughput SPI ionomers for PEMFCs. Such an approach not only minimizes experimentation but also provides an understanding of structure-property interdependence in SPI ionomers.

1.6   Thesis outliine

Chapter 1: Introduction: The first chapter of the thesis provides a background to the problem addressed in this work.

Chapter 2: Polymer discovery: Chapter 2 presents details on data collection, pre-processing of data, featurization, model training, and testing for discovering novel SPI PEM.

Chapter 3: Atomistic study: In Chapter 3, molecular dynamics simulations have been performed to obtain hydronium ion and water molecule diffusion behavior in NTDA/DSDSA/HFBAPP and NTDA/DSDSA/MDP SPI membranes.

Chapter 4: Experimental study: In Chapter 4, synthesis, physicochemical character-ization, and in-situ testing of NTDA/DSDSA/MDP SPI membranes have been done and compared with commercial Nafion PFSA membranes.

Chapter 5: Conclusion and future outlook: In the concluding chapter, the key find-ings and results of the present study are discussed. Also, key areas which require further computational and experimental efforts have been elaborated.

2 Polymer Discovery

2.1 Chapter summary

Due to the high cost and limited range of operating parameters, alternatives of per-fluorinated ionomers-based commercial Polymer Electrolyte Membranes (PEMs) are urgently required. Sulfonated polyimides (SPIs) based hydrocarbon PEMs have exhibited better proton conductivity (an important property determining the performance of fuel cells) even at low hydration levels and high temperatures, making them possible candidates for replacing commercial PEMs. However, finding alternative SPI PEMs is a critical polymer discovery problem that requires enormous experimental efforts where Machine learning (ML) approaches can help to reduce such efforts. To this end, both supervised and unsupervised ML approaches are developed to predict the proton conductivity of SPIs. A hybrid dataset of 81 unique SPIs is generated that consists of collected chemical structure-properties data from reported literature and calculated quantitative structure-property and semi-empirical quantum chemical descriptors. Using decision trees, rules that lead to a low or high class of proton conductivity labels with high accuracy are identified. The random forest regression model, on the other hand, provided a set of features that can predict proton conductivity with reasonable error. These findings are key to designing novel SPI PEMs while correlating proton transport at the ionomer level with factors such as the morphology of the microstructure and inter-chain interactions.

2.2 Introduction

Among all computational approaches, Machine Learning (ML) offers efficient techniques for discovering newer materials by training models or performing statistical

analyses on past data. Supervised ML involves the use of labeled data wherein an algorithm is trained to predict the value of a target variable from data provided in the form of a set of attributes. Broadly, all of the supervised ML problems fall in the purview of either classification or regression. While regression is performed when the target variable to be predicted is numerical, classification involves the prediction of categorical target variables formulated in the form of two or more classes. Unsupervised learning relies on the formation of association rules to group unlabelled but related instances into clusters(Mueller, Kusne, and Ramprasad, 2016). Machine learning involves learning predictive and descriptive algorithms that can learn complex inter-relationships among the multiple attributes from the patterns in the training data and then make accurate predictions about the outcome of unseen data points(Agrawal and Choudhary, 2016). ML expedites the process of discovery of materials by taking into account the structure-property relationships that might not seem apparent. With respect to HC PEMs, ML can help in identifying polymers that are cost-effective with the same or even better performance parameters than the state-of-the-art Nafion®.

2.2.1   Polymer discovery using Data-driven techniques

Several studies have been reported in the literature on the implementation of machine learning for polymer discovery applications and for the identification of relevant features for a target property. Ward et al. have emphasized why machine learning techniques based on atomistic parameters derived from first principles calculation give accurate predictions and help in studying cause-and-effect relationships between different attributes(Ward and Wolverton, 2017). Afzal et al. took the SVR approach to make predictions for refractive index using a data set of only 112 optical polymers(Afzal, Cheng, and Hachmann, 2018). Xu et al. trained a Support Vector Classifier on the data of 284 polymers comprising four blocks repeating units for predicting the band gap of the polymers. Their feature space comprised 5270 descriptors derived from DFT and Quantitative Structure-Prediction Relationship (QSPR) calculations. They identified that the band gap is related to a subset of the most relevant features selected using the Maximum Relevance Minimum Redundancy (mRMR) techniques. These features included the compositional formation and topological and geometrical structural information (Xu, Lu, et al., 2021).

There are several reported studies on the successful implementation of machine learning for polymer discovery applications and for the identification of relevant features for a target property. Xu et. al. trained a Support Vector Classifier on the data of 284 block polymer repeating units for predicting the bandgap of the polymers. Their feature set was derived from DFT and Quantitative Structure-Prediction Relationship (QSPR) calculations. They selected the most relevant features using Maximum Relevance Minimum Redundancy (mRMR) techniques(Xu, Lu, et al., 2021). Huang et. al. utilized an unsupervised learning algorithm to predict the Power Conversion Efficiency (PCE) of organic photovoltaic polymers from a feature set of 8 properties and 11 molecular descriptors. They used an Organic Photovoltaic (OPV) donor molecule database and also used dimension reduction methods for feature selection. They found that features like photon energy loss and the number of fluorine atoms are important features that help in predicting the values of PCEs. They also observed that subtle causal relationships among attributes can be found using statistical analysis techniques. Also, they made observations related to certain design aspects of the OPV donor molecules. For example, they found that polymers demonstrating high PCE usually have 6-9 aromatic rings in their chemical structure(Huang, Zhang, et al., 2020). Yan et al. applied a Dual Convolutional Neural Network model approach to the problem of finding efficient Thermoset Shape Memory Polymers (TSMPs). They represented polymers using BIGSMILES and validated the obtained results using MD simulations and synthesis(Yan, Feng, et al., 2021). Table 2.1 summarizes a few more ML studies performed for predicting polymer properties.

2.2.1.1 Data-driven techniques for polymer electrolyte study

Recently, the ion transport problem in solid polymer electrolytes has also been explored through the ML route backed by MD simulations. Wheatle et al. derived parameters that represent ionic transport and mechanical properties from coarse-grained molecular dynamics simulations of polymeric blend electrolytes (PBEs) for Lithium-ion batteries (Wheatle, Fuentes, et al., 2020). They trained a Gaussian Process Regression model with an 80/20 train-test split for predicting anionic and cationic diffusivities in the PBEs. Bayesian Optimization was used to obtain optimal parameter vectors

TABLE 2.1: Machine learning approaches demonstrated for the discovery of novel polymers and identifying relevant features

Reference	Predicted polymer property/ Data points / Descriptor-generation methods	Feature selection splitting/ Machine Learning algorithm	Selected features
Palomba, Vazquez, and Díaz, 2012	Glass transition temperature/ 88 High Molecular Weight polymers/ Novel descriptors generated using molecular modeling	A combination of Genetic algorithm, Regression techniques and Artificial Neural Networks (ANNs)	Main chain surface area, Number of rotatable bonds, Sidechain mass
Jabeen, Chen, et al., 2017	Refractive Index/ 133 polymers/ QSPR	Genetic algorithm(GA)/ Random test-train splitting (75-25)/ Multiple linear regression analysis	Ncsp2, ATSC2p, GATS1p, F01[C-F]
Mercader and Duchowicz, 2016	Polyacrylates/ Glass transition temperature/126 polyacrylates	Enhanced Replacement Method (ERM)/ Leave-One-Out (loo) and the Leave-More-Out Cross-Validation	Eight molecular descriptors: IDDM, piPC01, MWC06, Se, BELv8, nRCONHR, BEHm3, Neoplastic
Bhowmik, Sihn, et al., 2021	Specific heat at constant Pressure (Cp) of polymers/68 polymers/Molecular dynamics	Decision Tree and Principal Component Analysis/ 5-fold cross-validation	Bonding descriptors related to C1–C1 and C2–O1 bonds

that can facilitate an understanding of the ion diffusivity dependence on bulk properties like viscosity. Liu et al. applied ML approaches to the solid polymer electrolyte application to assess device-level performance. They fitted six different regression models including DTs to data sets derived from 200 literature reports of ionic conductivity data of PEO-LiTSFI polymer electrolytes. Their random forest model was able to capture the inverse relationship of the activation energy variation with the composition (weight%) of LiTSFI and ionic conductivity appropriately(Liu, Clement, et al., 2021).

2.2.2 Challenges in polymer discovery

The key to the successful implementation of any ML algorithm in any field is the quality of the data set used. Generally, in materials design problems, there is a lack of quality data that can be directly used to train ML algorithms (Zhang and Ling, 2018). The problems with the data sets include sparse data, missing values, inadequate number of data points, multicollinear feature space, noise in the attributes, and dimensionality. Generally, non-linear regression models have been used to describe the material properties and the underlying dependencies of the properties on the structure of the materials. Recently, several authors have reported the successful implementation of ML algorithms to problems related to Materials Sciences. Wang et al. performed Support Vector Machine based regression (SVR) to predict Elastic constants for metals and metallic binary alloys(Wang, Yang, et al., 2017). Jorgensen et al. employed a neural network model to predict the HOMO-LUMO gap in molecules(Jørgensen, Mesta, et al., 2018).

ML approaches offer efficient techniques for discovering newer materials by training models or performing statistical analyses on past data(ref). Polymer discovery has been equally expedited with the ML intervention. Recently, several data-driven initiatives have been reported to accelerate polymer discovery for various applications. For this purpose, a number of polymer data sets have been created. There are a few polymer data sets available such as PolyInfo (NIMS, Japan)(Otsuka, Kuwajima, et al., 2011) and Polymer Genome(Kim, Chandrasekaran, et al., 2018). In spite of this, polymer informatics and polymer discovery still have some challenges. Firstly, the collection of data related to polymers such as the chemical structure of the repeat unit and the associated properties from reported literature is a tedious task that also results

in sparse data. Moreover, the inclusion of conditions of measurement is also diffi-
cult to include. Secondly, the representation of information related to polymers is a
difficult task. Usually, simple monomers are represented as Simplified Molecular In-
put Line Entry System (SMILES), and IUPAC International Chemical Identifier(Inchl)
while more advanced formats for polymer data representation have been developed,
such as; Curly SMILES(Drefahl, 2011) and BIGSMILES(Lin, Coley, et al., 2019). The
encoded representation at higher length scales to describe oligomers, chains, local-
order (amorphous, semi-crystalline, and crystalline domains) and microstructure is
still a problematic issue in polymer structure representation(Wu, Yamada, et al., 2020).
Thirdly, the featurization or generation of information-rich features for training mean-
ingful predictive and descriptive algorithms is a challenge. Usually, there are fea-
ture vectors such as encoded fingerprints, Coulomb matrices(Hansen, Biegler, et al.,
2015), and Bag of Bonds(Rupp, Tkatchenko, et al., 2012). There are also Quantitative
Structure-Property Relationship (QSPR) features that quantify the chemical structure
of the monomers/oligomer/polymers such as the number of rotatable bonds, number
of aromatic rings, Topological Polar Surface Area, molecular surface area, and so on.
In addition to this, computational resources are used to derive certain computational
features that could be based on Density Functional Theory, Molecular Dynamics, and
other multiscale computations.

Feature selection is a critical step before training a model as high dimensionality (or
too many features) is not desirable for obtaining a good predictive model. There are
a few techniques to reduce dimensionality. Pearson Correlation Coefficient gives the
correlations among the features, Principal Component Analysis (PCA) uses orthogonal
linear combinations of features that show the highest variance and Recursive Feature
Elimination (RFE) assigns weights to features iteratively and gives ranking according
to their importance and eliminates the least important features.

In the present chapter, we collated the data set and characterized the data statisti-
cally by plotting histograms of collected features and their distributions and calculat-
ing statistical parameters namely mean, standard deviation, and range. We also deter-
mined correlations among the collected properties but due to sparse data, these corre-
lations can not be generalized. This way, we develop an understanding of the prop-
erties and their distribution in the collated data set. We also developed SQL queries

to average out the values reported for different experimental and testing conditions. Data preparation is an essential step in polymer discovery for any application, more so when the polymer class is relatively new for the target application. The polymer data sets that are available are not appropriate for problem formulation as the target property as well as other features vary along with experimental conditions. Moreover, chemical structures are collected in the form of SMILES codes which are processable by the algorithms. Also, we reduce the dimensionality of the collected data by identifying redundant features. Further, we use WEKA machine learning software to train Decision Tree Classifiers and iteratively eliminate lesser significant features using 10-fold cross-validation. In particular, we use two algorithms - Relief F and Gain ratio for iterative feature elimination of low-ranking features. Further, three DTCs with the highest classification accuracy were identified to predict low proton conductivity (LPC) and high proton conductivity (HPC) class labels. Along with the prediction of proton conductivity class labels, we also discuss the physical and chemical significance of the identified features. We also take the help of Self-organizing maps (SOMs) to visualize the distributions of the identified features along with studying the correlation among these features.

2.2.3 Machine Learning techniques

Among the various ML algorithms, classification provides predictive models wherein the target variable to be predicted is a categorical variable i.e. has class labels while other attributes could be continuous or numerical. Decision tree (DT) algorithms are tools that convert the inter-dependence of the attributes of a data set into a set of simple rules or 'If-Else' statements. Generally, DTs are white-box, non-linear classifiers provided the depth of the tree remains small. DT classifiers perform several tasks namely classification (a data mining task), predictive analysis (a machine learning task), feature selection (a dimensionality reduction task), and rule extraction (a task for establishing descriptive inter-dependence among attributes).

On the other hand, the other branch of supervised ML includes regression where the target variable is also continuous in nature similar to the input feature variables. Usually, regressor algorithms are accurate but complex and uninterpretable which makes them a black box and complicates their application to polymer design. Random

forest regressor(RFR) is a tree-based regression algorithm that involves numerous decision trees trained with random training subsets derived from the original training set. RFRs are less prone to over-fitting, enable the ranking of features based on their importance in training the trees, and offer generalized predictions as they involve a combination of DTs.

Self-organizing maps (SOMs) are a class of unsupervised machine learning algorithms and a kind of artificial neural network (ANNs) developed by T. Kohonen in 1990. SOM training involves transforming input data onto a two-dimensional lattice or grid of cells or neurons through a distance-based iterative algorithm. SOMs offer the ability to visualize the input feature space and obtain information related to the importance of features in predicting the target variable and the similarities in the data points(Kohonen, 1990).

Huang et al. utilized an unsupervised learning algorithm Self-organizing maps (SOMs) to predict the Power Conversion Efficiency (PCE) of organic photovoltaic polymers from a feature set of 8 properties and 11 molecular descriptors. They used an Organic Photovoltaic (OPV) donor molecule database and dimension reduction methods for feature selection. They found that features like photon energy loss and the number of fluorine atoms are important features that help in predicting the values of PCEs. They also observed that subtle causal relationships among attributes can be found using statistical analysis techniques. Also, they made observations related to certain design aspects of the OPV donor molecules. For example, they found that polymers demonstrating high PCE usually have 6-9 aromatic rings in their chemical structure(Huang, Zhang, et al., 2020). Yan, Feng, et al., 2021 applied a Dual Convolutional Neural Network model approach to the problem of finding efficient Thermoset Shape Memory Polymers (TSMPs) (Yan, Feng, et al., 2021). They represented polymers using BIGSMILES and validated the obtained results using MD simulations and synthesis. Moreover, Dennis et al. trained supervised and semi-supervised self-organizing maps through the Python package SUSI to generate 15×15 grid size lattices for obtaining information about glass transition temperature from the chemical structures of high glass transition temperature polyimides. They used an extremely meager data set of 23 polyimides(Dennis and Zubarev, 2021).

2.2.4 Proton Conductivity (Target property)

Proton conductivity is the most important property of the PEM. It depends on PEM properties such as ion exchange capacity (IEC) of the membrane, water uptake, direction of measurement (through-plane or in-plane direction), level of degradation as well as temperature and humidity(Kim and Lee, 2015a). It is usually measured using a four-probe method from Electrochemical Impedance Spectroscopy(EIS) which is an ex-situ electrochemical technique(Mikhailenko, Guiver, and Kaliaguine, 2008).

2.3 Methodology

The method followed for identifying DT classifiers for SPI PEMs involved three steps/stages: 1. Collection of chemical structures and data from the literature, 2. Generation of QSPR features from the collected chemical structures and calculation of quantum chemical properties, and 3. Finally, the generation of DTs is based on the obtained data sets. This workflow has been schematically shown in Fig. 2.1.
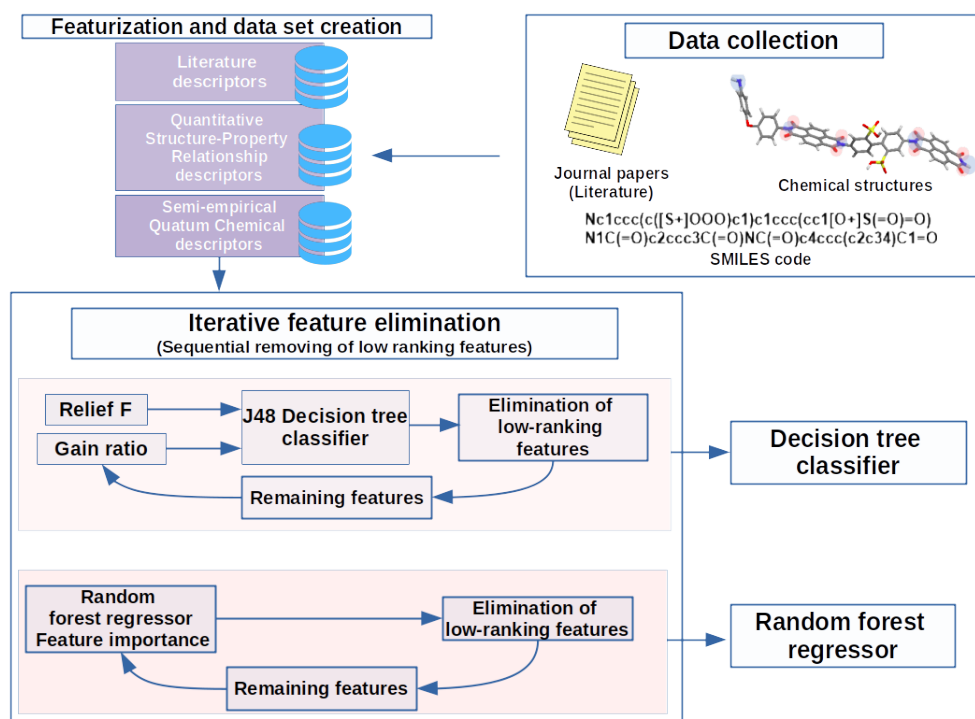
FIGURE 2.1: Workflow followed in the present study for predicting low proton conductivity (LPC) and high proton conductivity (HPC) class labels of SPI PEMs using DT classifiers

2.3.1 Data collection of SPI PEMs repeat units and their properties from Literature

Firstly, an extensive literature review covering 200+ journal articles that reported the synthesis and performance of SPI PEMs for Hydrogen Fuel Cells (HFC) or Direct Methanol Fuel cells (DMFC) from the year 2000 to 2019 was performed. The reported data of properties (molar content of sulfonate group and proton conductivity) were collected along with the chemical structures. All units were standardized for uniformity. To eliminate the effects of measurement condition parameters such as temperature and humidity, the values of the molar content and proton conductivity were averaged to obtain one tuple for each unique repeat unit instead of the multiple values reported.

2.3.1.1 Representation of repeat units

The chemical structures of the reported monomers or repeat units were collected in the form of conventional SMILES code using the IUPAC names given in the experimental section of the reported literature. ACD Chemsketch was used to generate the chemical structures from IUPAC names which were corroborated with the reported structure in the literature(Hunter, 1997). The chemical compositions of the SPI repeat units were named according to the following naming convention usually followed for any condensation polymer i.e. Dianhydride/Sulfonated Diamine/Non-sulfonated Diamine. Based on this, for example, an SPI named NTDA/BDSA/ODA means that it includes one unit of the BDSA (a sulfonated diamine) with one unit of NTDA (a dianhydride) and also one unit of ODA (a non-sulfonated diamine) with one unit of NTDA, as shown in Fig. 2.2.

2.3.1.2 Dataset preparation

A total of 81 unique SPI monomers/repeat units and their corresponding properties were calculated. Other properties such as IEC, water uptake, and swelling ratio were ignored as that could have resulted in sparse data, and also since they are not truly designed attributes they are measured once the polymer is synthesized. The target variable i.e. proton conductivity (which was extracted as a numeric and continuous variable) was formulated into binary class labels i.e. converted into a categorical variable.
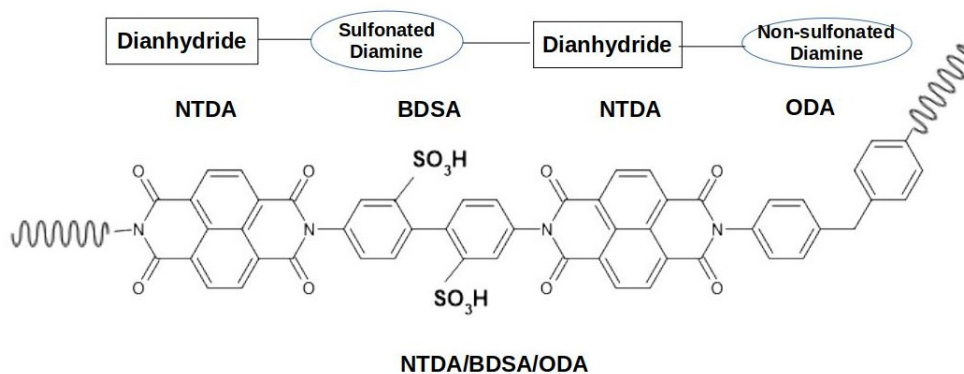
FIGURE 2.2: Naming convention for repeat units of Sulfonated polyimides

Data points were categorized into low proton conductivity (LPC) and high proton conductivity (HPC) classes with associated proton conductivity range between 0.05-0.08 S.cm^{-1} and 0.11-0.26 S.cm^{-1}, respectively. The class-wise distribution of the collected data from the literature is shown in Fig. 2.3.

2.3.1.3 Calculation of semi-empirical descriptors

In the next step, the calculation of quantum chemical descriptors of all SPI PEM monomers collected from literature was performed using MOPAC2016 (Molecular Orbital PACkage) which is a software that performs semi-empirical quantum chemical calculations on molecules(James, 2016). It is a semi-empirical quantum chemistry program based on Dewar and Thiel's Neglect of Diatomic Differential Overlap (NDDO) approximation. NDDO-type semi-empirical methods use molecular orbitals to solve the Hartree-Fock equations in which PM7 Hamiltonian was modified by Throssel and Frisch for continuous potential energy surfaces (Stewart, 2007). MOPAC works best for any general molecule since it makes the fewest approximations while performing calculations. It is especially suitable for large chemical structures as the calculations are fast and the results are of comparable accuracy as DFT calculations(Stewart, 2012; James, 2016). Semi-empirical calculations on the obtained chemical structures were performed after geometry optimization using the PM7 method and simple energy calculations were performed. Semi-empirical calculations are usually carried out in a minimal basis set and incorporate results from spectroscopy. This data set is referred as MopacD in this paper.
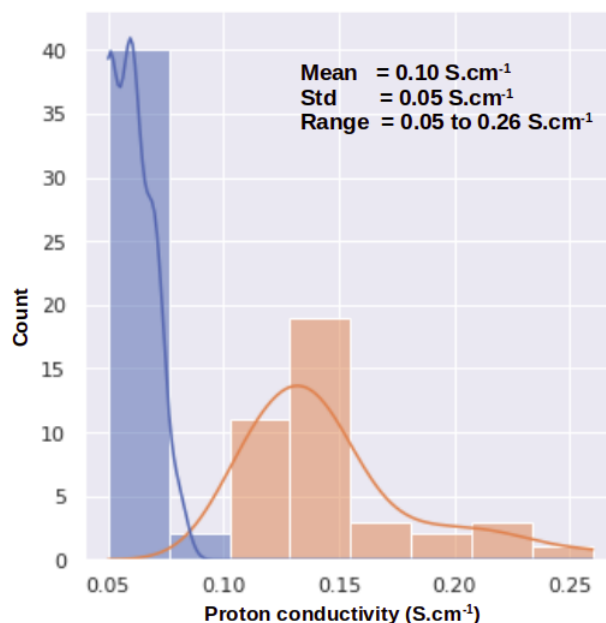
FIGURE 2.3: Class-wise distribution of proton conductivity data into Low Proton conductivity (LPC) and High Proton Conductivity (HPC) class with an associated range of 0.05 to 0.08 S.cm^{-1} and 0.11 to 0.26 S.cm^{-1}, respectively.

Secondly, another data set (MopacW) was collected by performing the same geometry optimization and energy calculations after adding three water molecules in the vicinity of each acidic site in the repeat units using the Avogadro chemical editor tool(Hanwell, Curtis, et al., 2012). To be precise, the water molecules were added near the H-atom attached to the O-atom of the sulfonate (-SO$_3$H) group. H-bond energy and dispersion energy were estimated after geometry optimization with the PM7 method. All H-bonds with the strength of 1.0 kcal·mol^{-1} were then monitored. In MOPAC, hydrogen bond energy is calculated using the local charge around the bond depending on the presence of Oxygen, Nitrogen, and H-ions in their vicinity. This approach was taken to include the hydrogen bonding-related parameters like the number of hydrogen bonds and energy of hydrogen bonds around sulfonate groups in the presence of water molecules and how the chemical structure supports the hydrogen bonds.

2.3.1.4   Calculation of QSPR descriptors

Thirdly, the QSPR data sets were generated using two open-source molecular QSPR calculation tools namely, Mordred(Moriwaki, Tian, et al., 2018) and Babel(O'Boyle, Banck, et al., 2011) named as *Babel* and *Mordred* data sets, respectively. A molecular

descriptor is a way of representing different characteristics of a chemical structure, quantitatively. These are derived from calculations that are performed on the digital representation of chemical structures of compounds (SPI monomers in this case) so that quantitative comparison among different monomers for any particular characteristic, say the number of rotatable bonds, becomes easier. This way, creating a data set becomes easier and meaningful statistics of scientific information can be obtained through data mining. Mordred package can calculate more than 1800 2-D and 3-D descriptors while OpenBabel can calculate 16 numerical 2D descriptors. In this work, only 2D descriptors were calculated that do not require geometry optimization but are based on the chemical composition of the molecule alone.

2.3.1.5   Final dataset compilation

Only one attribute namely the molar content of the sulfonate group (Molar) was collected from the reported literature along with the target variable (PC). IEC as an attribute had to be removed as its information content surpassed all other attributes and prevented the formation of detailed trees with higher classification accuracy. Removal of IEC led to the formation of this tree which demonstrated the highest classification accuracy. In total 13 QSPR descriptors were generated using Mordred (*Mordred* data set), and 10 QSPR descriptors were generated using Open Babel (*Babel* data set). Using MOPAC, 7 semi-empirical descriptors were calculated for as-such repeat units (*MopacD* data set) while 9 were calculated by adding three water molecules around each acidic group of the repeat unit (*MopacW* data set). Combinations of these descriptors are named as *Babel + Molar*, *Mordred + Molar*, *Babel + Mordred + Molar*, *MopacD + Molar*, *MopacW + Molar* and *MopacD + MopacW + Molar* which represent a combination of *Babel* and *Molar* descriptors, *Mordred* and *Molar* descriptors, *Babel*, *Mordred* and *Molar* descriptors, *MopacD* and *Molar*, *MopacW + Molar* and *Molar* descriptors, respectively.

2.3.2   Feature selection and random forest regression model training

Similar to DT classifiers, RFR algorithms also act as a means of feature selection as the basis for the training of DTs in RFRs is the variance among the input features. Thus, high-variance features tend to appear as higher-ranking features in the feature importance histogram. Similarly, training of RFR was done with all the (40 features)

initially and the features lying below a threshold in the feature importance histogram were removed. RFRs were obtained using the SK learn RFR function in Python. This iterative method was repeated till the test error was minimal. Hyperparameters were optimized each time while training RFRs.

2.3.3   Self Organizing Map train and feature visualization

SOMs transform the input vectors into a two-dimensional map wherein each neuron in the grid is a representation of the input data and is associated with a weight vector. The process of SOM training involves initialization of the weight vector, w_i, and selection of a random data vector (x_i) from the input data. As shown in Eq. 2.1, w_i is an n-dimensional randomly initialized weight vector where n is the dimensionality of the input. On the grid, the neuron whose weight vector is closest to the input vector in terms of the Euclidean distance is termed as the winning neuron or the best matching unit (BMU) as shown in Eq. 2.2. The 'self-organizing' part of the SOM algorithm is called so as there is iterative updating of the weight vector of the BMU and its immediate neighboring neurons to which it is connected. The updating of the weight vectors takes place according to Eq. 2.3 till the weight vector of the BMU is closest to the input vector.

$$w_i = [w_{i1}, w_{i2}, ..., w_{in}] \tag{2.1}$$

$$d_i(x) = \sum_{i=1}^{D} [x_i - w_i]^2 \tag{2.2}$$

$$w_i(t+1) = w_{i1}(t) + \alpha(t).T_{j,I(x)}(t)(x_i - w_{ji}) \tag{2.3}$$

where, $T_{j,I(x)}(t)(x_i - w_{ji})$ is the Gaussian neighborhood function and $\alpha(t)$ is the learning rate.

SOMs were generated using MINISOM python package (Vettigli, 2018)for visualizing individual feature planes for each of the features selected by the tree-based supervised ML algorithms (DT and RFR) and also plotting a combined map wherein the features are clustered according to their maximum weights.

2.4 Results and Discussion

The Results and Discussion section is organized into four sub-sections. The first and second sub-sections discuss the results of feature selection and training and testing of DT classifiers. Further, certain decision rules have been identified that can be key in designing newer SPI PEMs while correlating proton transport at the ionomer level with factors including the morphology of the microstructure and inter-chain interactions. The third sub-section discusses the relevance of the features identified from these trees, the correlation among these features and their distribution, and the evaluation metrics for classification. The fourth sub-section discusses the feature planes obtained from the training of SOMs with the selected features. Features identified at each stage of the training process are mentioned in the brackets.

2.4.1 Decision Trees using different data sets

Different combinations of feature subsets were used for feature selection and DT training to find the combination of features from different data sets with maximum classification accuracy. For example, features derived from *Babel* data set were used as such for predicting class labels. Then the molar content feature (*Molar*) data was used along with *Babel* data set features to predict the classification accuracy. Similarly, feature selection and DT training procedure were repeated using feature subsets derived from Mordred (*Mordred*), Semi-empirical calculations on repeat units as such (*MopacD*), and after adding water molecules near acidic groups (*MopacW*). Classification accuracy was recorded along with the number of features that gave that accuracy. Figure 2.4 shows the classification accuracy obtained with different data sets individually and after hybridization with other data sets. Figure 2.4 also shows the maximum classification accuracy obtained for a particular data set along with the number of features that led to the construction of the DT.

The *Babel* features provided a classification accuracy of only 67.90% with four features (abonds, dbonds, F, TPSA) according to the Gain ratio algorithm and 71.60% with three features (dbonds, MW, F) according to the ReliefF algorithm. Further, the addition of the literature feature (Molar content of the repeat unit) increased the classification accuracy as among *Babel+Molar* features, six features (Molar, abonds, F, TPSA,
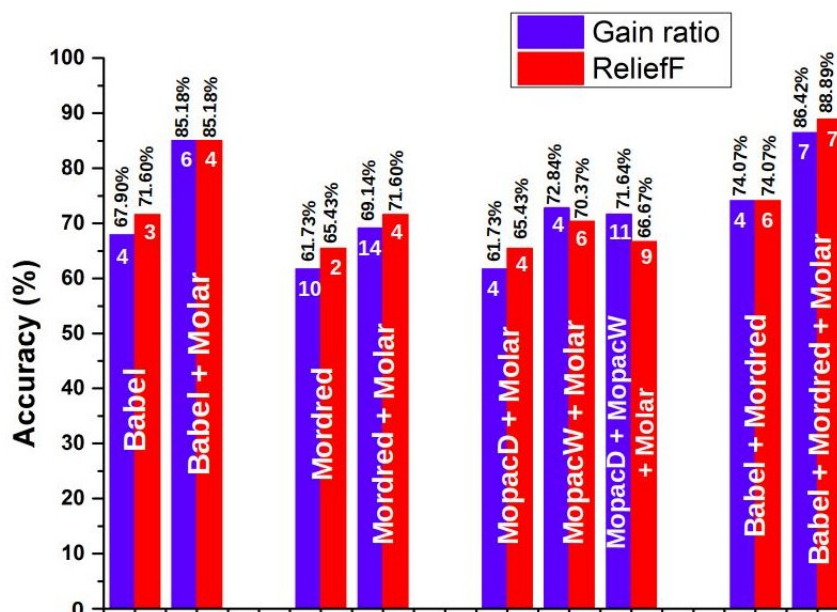
FIGURE 2.4: Maximum classification accuracies (%) for different combinations of *Molar*, *Mordred*, *Babel*, *MopacD*, *MopacW* data sets is obtained from iterative feature selection and DT training process using (a) Gain ratio algorithm (b) ReliefF algorithm, with sequential removal of the lowest ranking feature.

sbonds, dbonds) were identified by the Gain ratio feature selection algorithm after sequential removal of low ranking features and training of DT after each removal. The DT trained by these features provided a classification accuracy of 85.18%. Another feature subset identified by the ReliefF algorithm also shows the same classification accuracy but with a different set of features (Molar, dbonds, MW, F). Clearly, Molar content adds important information for the prediction of the class labels.

The DTs trained using feature subset obtained from *Mordred* data set showed classification accuracy of 61.73% with ten features(nAromAtom, nAromBond, nHeavyAtom, nHetero, nN, nO, nBondsD, nBondsA, nBondsKS, nBondsKD) as identified by Gain ratio algorithm whereas inclusion of Molar content feature *Mordred+Molar* could improve it to 69.13% with fourteen features (Molar, ABC, ABCGG, nAromAtom, nAromBond, nHeavyAtom, nHetero, nN, nO, nBondsS, nBondsD, nBondsA, nBondsKS, nBondsKD). Clearly, these feature subsets do not provide parsimonious and accurate DTs. The ReliefF algorithm provided a classification accuracy of 65.43% with two features (nN and nBondsD) and 71.60% with four features (Molar, nAromBond, nN, nBondsD) among the *Mordred+Molar* features.

DTs trained with only semi-empirical (SE) feature subsets (*MopacD* and *MopacW*) showed poor accuracy and only one node (not shown in Fig. 2.4). DTs trained with feature subsets including Molar feature along with the semi-empirical features (*MopacD+Molar*, *MopacW+Molar* and *MopacD+MopacW+Molar*) identified with the feature selection algorithms showed moderate accuracy. Among these, *MopacD+Molar* showed a maximum classification accuracy of 61.72% with four features (Molar, Dispersion energy, Core-core repulsion, Ionization potential) identified by Gain ratio algorithm and 65.43% with four features (Molar, Dispersion energy, Core-core repulsion, COSMO area, COSMO volume) identified by the ReliefF algorithm. On the other hand, *MopacW+Molar* performed slightly better. The gain ratio algorithm identified four features (molar, Dispersion energy (with water molecules), COSMO volume (with water molecules), Ionization potential (with water molecules)) with which DT gave the accuracy of 72.84% and ReliefF algorithm identified six features (Molar, Dispersion energy (with water molecules), H-bond energy (with water molecules), COSMO area (with water molecules), COSMO volume (with water molecules), Ionization potential (with water molecules)) with which the DT showed an accuracy of 70.37%.

DTs trained with a subset of features identified by feature selection algorithms from a combination of semi-empirical features of repeat unit as such and repeat units with water molecules near acidic groups, showed 71.64% accuracy with eleven features (Molar, Dispersion energy, Core-core repulsion, COSMO area, COSMO volume, H-bond energy, number of hydrogen bonds, Core-core repulsion (with water molecules), COSMO area (with water molecules), COSMO volume (with water molecules), ionization potential (with water molecules)) in case of Gain ratio and merely 66.67% with nine features (Molar, dispersion energy, COSMO area, COSMO volume, dispersion energy (with water molecules), Core-core repulsion (with water molecules), COSMO area (with water molecules), COSMO volume (with water molecules), Ionization potential (with water molecules)) in case of features identified by ReliefF algorithm.

Further, a combination of *Babel* and *Mordred* showed an accuracy of 74.07% with four features (dbonds, F, nAromAtom, nBondsA) identified by Gain ratio algorithm and six features (dbonds, F, ABCGG, nAromBond, nN, nBondsD) identified by ReliefF algorithm. Maximum classification accuracy was shown by the feature subset

from the combination of *Babel*, *Mordred* and *Molar* data sets (referred in Fig. 2.4 as *Babel+Mordred+Molar*). According to the Gain ratio feature selection algorithm, seven features (molar, abonds, dbonds, F, nAromAtom, nAromBond (or nBondsA)) were identified. They showed a classification accuracy of 86.42%. ReliefF feature selection algorithm also identified seven features but included the number of aromatic bonds (nAromBond) instead of the number of aromatic atoms (nAromAtom). This set of features showed a classification accuracy of 88.89%. The inclusion of the molar content feature from the *Molar* data set caused a dramatic increase in the classification accuracy.

Following the principle of parsimony, the DTs that provide maximum accuracy with the minimum number of features were selected for further evaluation as the features involved in the construction of these DTs offer maximum information content for the prediction of the class labels. Three high-accuracy DTs were selected for further evaluation: 1. *Babel+Molar* (ReliefF) (Fig. 2.5), 2. *Babel+Mordred+Molar*(Gain ratio)(Fig. 2.6), and *Babel+Mordred+Molar*(ReliefF)(Fig. 2.7). All the DTs show a similar structure in terms of the root node and internal nodes. The difference is in one of the leaf nodes. All three DTs show that the molar content of the sulfonate group is the most important attribute as it is the root node. The splitting value of 55% is reasonable as there is always a critical value for the concentration (or content) of sulfonate group sites (or acidic sites) along the chain. Insufficient acidic sites do not provide adequate water binding sites to reach the percolation threshold and the proton concentration remains inadequate for efficient proton conduction. Whereas too many acidic sites provide a lower percolation threshold but decrease the overall mechanical properties(Knauth and Di Vona, 2014). Thus, a value of 55% molar content of the sulfonate group per monomer is an agreeable value. This is also indicated by the fact that there are 17 training instances (rule coverage = 20.98%) that have molar content less than 55% and fall in the LPC class. Out of these 17 instances, 16 training instances (rule precision= 94.11%) are classified correctly as Low while one is classified incorrectly as HPC. Whether the polymer is fluorinated or not is the next important attribute. Two branches grow out based on whether the monomer is fluorinated or not. If the monomer is not fluoridated then the number of double bonds present in the monomer plays an important role in predicting the proton conductivity class. The splitting value of 12 indicates that if the number of double bonds in the monomer is higher than 12 then a total of 19 training

instances (rule coverage = 23.45%) took that decision path and out of them only one got misclassified whereas 18 (rule precision = 94.73%) of them got correctly classified as having HPC class label. This demonstrates a strong design principle for designing a non-fluoridated SPI monomer.
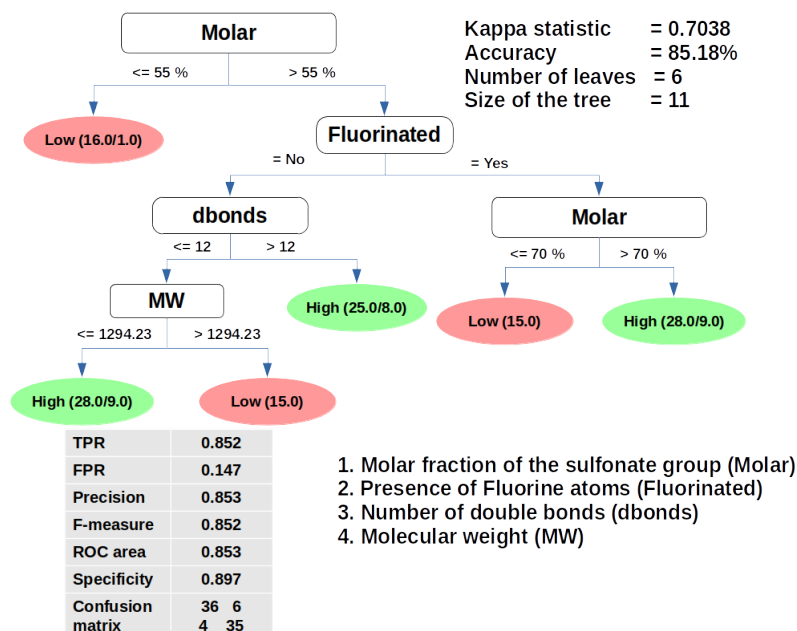
FIGURE 2.5: Pruned DT trained using four features identified by the ReliefF algorithm from *Babel+Molar* features.

If the monomer is fluorinated, then again the molar content becomes a key factor in deciding the proton conductivity behavior. Molar content being a numerical attribute got utilized again with another splitting value due to the high information gain it brings to the tree classifier. This is proved by the fact that it is the highest-ranking feature identified by both attribute selection algorithms. In the case of fluorinated monomers, a molar content of more than 70% is required for the SPI monomer to belong to the class of highly proton-conductive monomers. This can be due to the fact that fluorinated monomers tend to be more hydrophobic in comparison with their non-fluorinated counterparts. Thus probably additional acidic sites are required to attain the characteristic nano- and micro-phase segregation and thus the optimum percolation behavior which facilitates the facile movement of protons. There are only 11 training instances that are fluorinated and out of them, 5 have a molar content of more than 70% while 6 (Rule coverage = 7.40%) have a molar content of less than or equal to 70%. All 6 monomers having lesser than 70% molar content are correctly classified
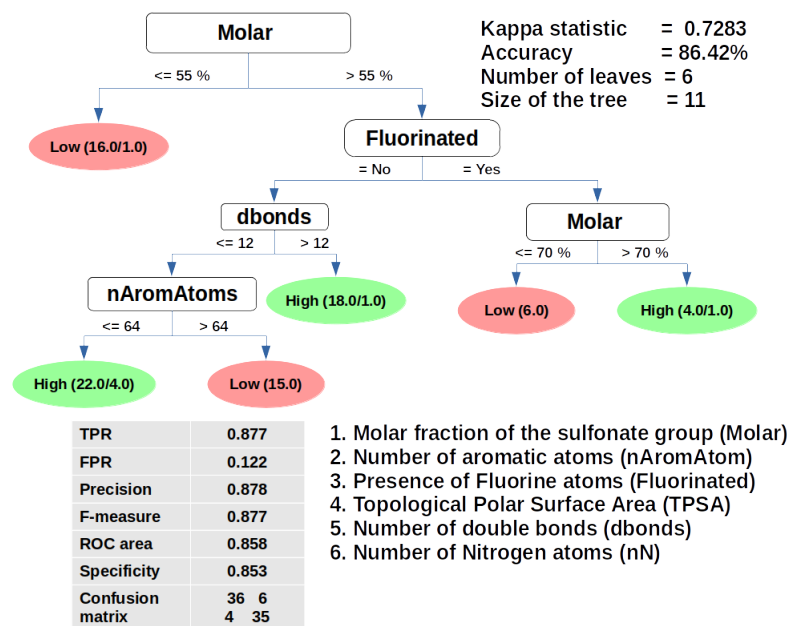
FIGURE 2.6: Pruned DT trained using four features identified by the Gain ratio algorithm from *Babel+Mordred+Molar* features.

(Rule precision= 100%) as having LPC by the classifier while out of 5 training instances (Rule coverage = 6.17%) having more than 70% molar content, 4 are correctly classified (Rule precision = 80.00%) as having HPC and one is misclassified.

On the other hand, if the number of double bonds is lesser than 12 in a monomer, then according to DT trained with the four features identified by the ReliefF algorithm, MW is the next important feature as shown in Fig. 2.5. A non-fluorinated test instance will be classified as LPC if it has less than or equal to 12 double bonds and if its molecular weight is more than 1294.23 as 15 instances (rule coverage = 18.52%) follow this rule and all of them are classified correctly as LPC (rule precision = 100.00%). On the other hand, if the test instance has a molecular weight lesser than or equal to 1294.23 then it is more likely to be classified as HPC as 37 instances (rule coverage= 45.68%) follow this decision path. Out of those 37 instances, only 25 instances (rule precision = 75.67%) were identified as HPC while 9 were misclassified as LPC. The DT trained with the featured sub-set identified by the Gain ratio algorithm from the pool of QSPR features along with the molar content of the sulfonate group is shown in Fig. 2.6. According to this DT, there should be lesser than or equal to 64 Aromatic atoms in the non-fluorinated repeat unit. This can lead to the designing of a high proton conductive SPI PEM. Here 26 training instances (Rule coverage = 32.09%) out of 81 were observed
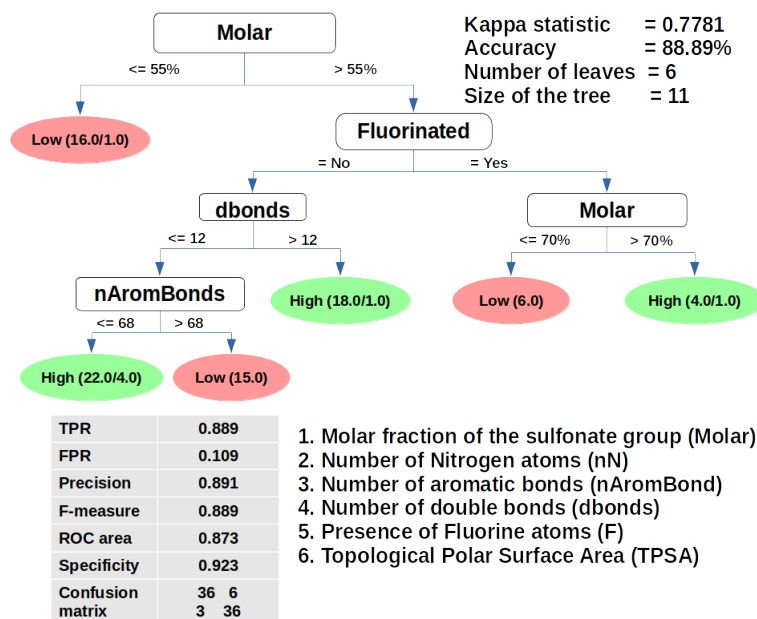
FIGURE 2.7: Pruned DT trained using four features identified by the ReliefF algorithm from *Babel+Mordred+Molar* features.

to follow this rule and 22 (Rule precision = 84.61%) out of them were correctly classified by the tree as belonging to the HPC class. Another branch rules that if the number of aromatic atoms increased to more than 64 in a monomer then the SPI PEM is more likely to have a lower proton conductivity. This is shown by the terminal node in Fig. 2.6. Fifteen training instances (Rule coverage = 18.51%) followed this rule all of which were classified as belonging to class LPC (rule precision = 100%). Also, in the case of DT trained with the feature subset identified by the ReliefF algorithm from QSPR features along with the molar content of the sulfonate group as shown in Fig. 2.7, it can be seen that the terminal node is replaced by the number of aromatic bonds. This has increased the accuracy further. Table 2.2 summarizes the rules obtained from the decision tree shown in Fig. 2.7 i.e. the decision tree with the maximum classification accuracy.

2.4.2 Feature selection using Random Forest Regressor

Figure 2.8 shows the residual plots generated for predicted and actual proton conductivity on the train and test sets for different feature subsets which were generated as a result of iterative feature elimination. Firstly, RFR was trained using the full feature set i.e. 40 features. For this subset, R^2 of 0.546 on the train set with RMSE, MSE, and MAE

TABLE 2.2: Rules derived from the DTs trained with feature subsets identified by the feature selection algorithms show maximum classification accuracy.

Rule	Class labels
If Molar <= 55%	LPC
If Molar > 55%, F=No and dbonds >12	HPC
If Molar > 55,% F=No, dbonds <= 12 and nAromBonds > 68	LPC
If Molar > 55%, F=No, dbonds <= 12 and nAromBonds <= 68	HPC
If Molar > 55%, F=Yes and molar > 70%	HPC
If Molar > 55%, F=Yes and molar <= 70%	LPC

of 0.0322, 0.0010, and 0.0283 respectively while R^2 of 0.2486, RMSE of 0.0402, MSE of 0.0016, MAE of 0.0325 for the test set. After the removal of features based on low variance, 16 features remained with which another RFR was trained. It showed a R^2 of 0.7497, RMSE of 0.024, MSE of 0.0005, MAE of 0.0281 while for the test set R^2 of 0.2532, RMSE of 0.0401, MSE of 0.0016, and MAE of 0.0321 was obtained. At last, the top 7 features were selected which showed R^2 of 0.50067, RMSE of 0.0338, MSE of 0.00114, and MAE of 0.0296 for the train set, and R^2 of 0.352, RMSE of 0.0373, MSE of 0.0014, and MAE of 0.0301 for the test set. Clearly, the RFR is a moderate model for this data set as shown by the moderate train and test R^2 value for these data sets. Nevertheless, since the range of the target variable is very less (0.08 to 0.22 S.cm^{-1}), the other error metrics can be considered important.

2.4.3 Feature visualization

2.4.3.1 Feature planes

Figure 2.9 shows the feature planes or component planes derived after training SOMs with the feature subset with a quantization error of 0.2567 on output grid size 20 x 20. These feature planes are based on the weights assigned to the neurons for each of the selected features and provide information about the distribution followed by the feature. In the feature maps, red regions represent higher proton conductivity and blue

		40 features				**16 features**				**7 features**	
		Train set	**Test set**			**Train set**	**Test set**			**Train set**	**Test set**
R^2		0.546	0.2486	R^2		0.7497	0.2532	R^2		0.50067	0.352
RMSE		0.0322	0.0402	RMSE		0.024	0.0401	RMSE		0.0338	0.0373
MSE		0.0010	0.0016	MSE		0.0005	0.0016	MSE		0.00114	0.0014
MAE		0.0283	0.0325	MAE		0.0281	0.0321	MAE		0.0296	0.0301
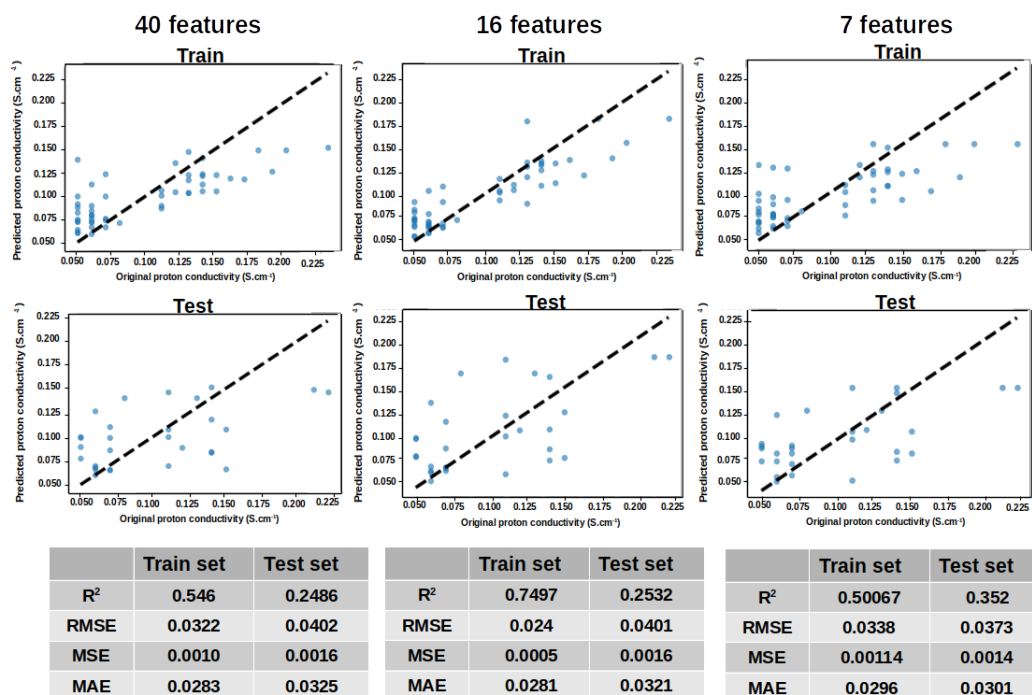
FIGURE 2.8: Residual plots showing the predicted and actual proton conductivity obtained using RFR for training and test sets for 40, 16, and 7 features. Also shown are the R^2, RMSE, MSE, and MAE values obtained randomly for models.

regions have lower proton conductivity. These feature maps are also helpful in high-lighting the correlations that exist among the features. It can be seen that feature planes of the QSPR features namely dbonds, MW, TPSA, nAromAtom, nAromBond, and nN are quite identical yet cannot be considered entirely correlated except for nAromAtom and nAromBond. The feature plane of the categorical feature showing whether the monomer is fluorinated or not has a major blue region representing LPC. Molar repre-senting molar content of the sulfonate group in the monomer exhibits a slightly similar pattern as the semi-empirical feature, Dispersion energy. The other semi-empirical fea-ture shows similarity with the QSPR features.

2.4.3.2 Cluster map

Figure 2.10 shows a cluster map wherein the neurons are clustered and color-coded according to the maximum weight associated with the features. Thus a map of regions where individual features have higher values is obtained which enables us to visualize the contribution of the individual features to the prediction of the target variable.
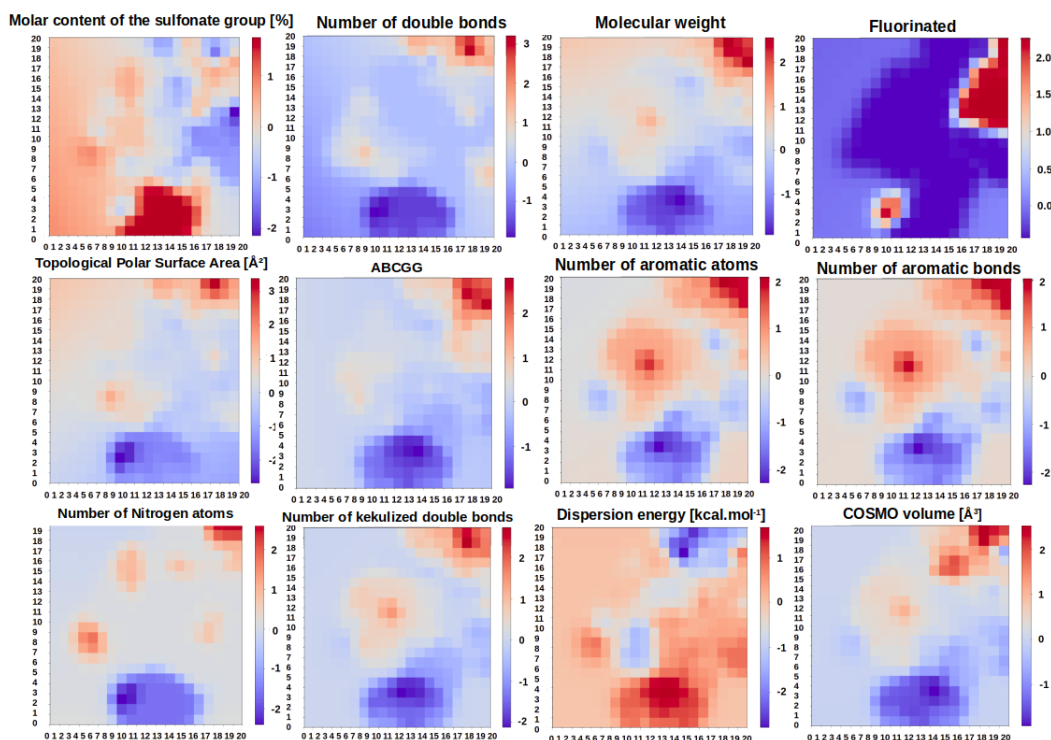
FIGURE 2.9: Feature planes generated using MINISOM python package with a quantization error of 0.2567 and grid size of 20 x 20 for the features identified through DT classifier

2.4.3.3 Correlation map

A Pearson correlation analysis was done to observe the kind of interaction that exists among these features as identified by the three DTs. The Pearson correlation was observed between these attributes and the numerical proton conductivity output variable as the Proton conductivity class is a categorical variable and not appropriate for correlation analysis. Figure 2.11 shows a color-coded matrix plot with values of the Pearson correlation shown in the boxes and the features outlined on the sides. These correlations have been obtained using the data set of 81 points and averaged proton conductivity values neglecting the conditions of measurement. Therefore, it represents the correlation obtained from the available data set. These correlations may change upon the addition of more data points. The molar content of sulfonate groups exhibits a high positive correlation (0.53) with proton conductivity whereas the number of aromatic atoms shows the highest negative correlation (-0.47). Apart from the number of aromatic atoms, proton conductivity is negatively correlated with the molecular weight(-0.41), number of Nitrogen atoms(-0.37), number of double bonds(-0.3), and
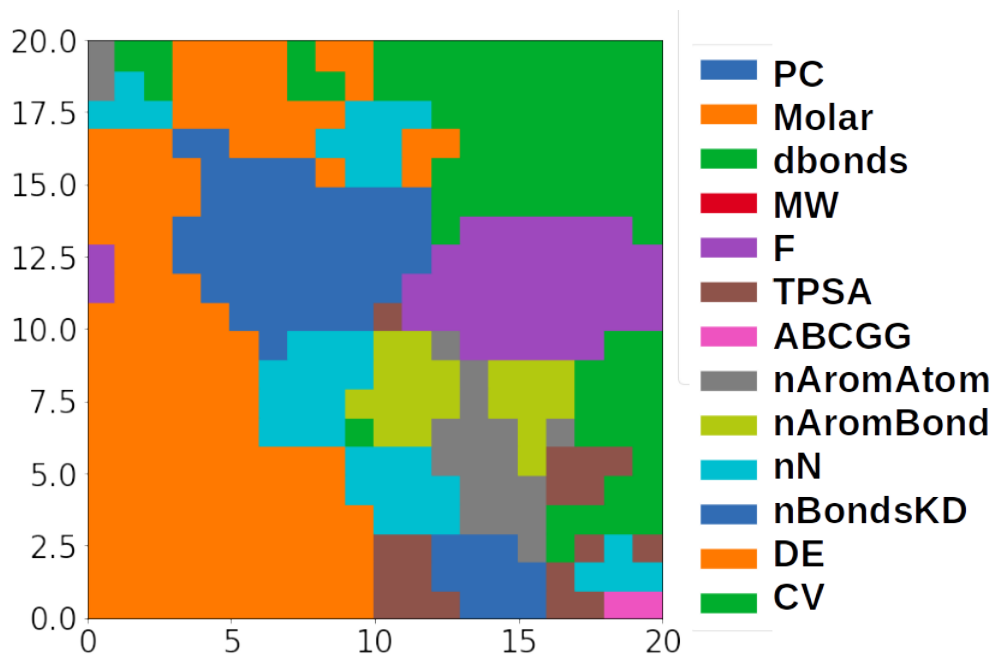
FIGURE 2.10: Feature cluster map obtained using MINISOM python package demonstrates the contribution of each feature in color-coded form.

TPSA(-0.26). Features discovered by the RFR namely Dispersion energy and COSMO volume also have only distant correlations with the target variable. Dispersion energy has a mild positive (0.36) correlation with PC while COSMO volume has a negative correlation (-0.44) with PC. However, it should be kept in mind that correlation does not indicate causation and is merely a tool for supplementing the understanding of the dependence of proton conductivity on the features identified by the feature selection algorithms and DTs.

2.4.3.4 Kernel disctribution

Figure 2.12 shows the class-wise distributions of the features identified by the feature selection algorithms through both classification (DTs) and regression (RFR) routes for their maximum relevance in predictions of the target variable i.e. class labels (LPC and HPC) and proton conductivity as such. It can be seen from the histograms that these features do not follow any conventional distribution and the overall distribution is data-driven and unknown for each of the identified features. It is important to observe that there are limitations related to the availability of sufficient data points in Polymer informatics for training ML algorithms. The true distribution of a variable

1	0.51	-0.26	-0.38	-0.25	-0.22	-0.37	-0.44	-0.44	-0.33	-0.39	0.36	-0.44	PC
0.51	1	-0.6	-0.5	-0.16	-0.49	-0.51	-0.51	-0.51	-0.6	-0.5	0.41	-0.53	Molar
-0.26	-0.6	1	0.86	0.15	0.94	0.87	0.73	0.71	0.81	0.84	-0.65	0.78	Dbonds
-0.38	-0.5	0.86	1	0.34	0.88	0.97	0.92	0.91	0.82	0.97	-0.71	0.86	MW
-0.25	-0.16	0.15	0.34	1	0.066	0.43	0.2	0.2	0.21	0.19	-0.03	0.16	F
-0.22	-0.49	0.94	0.88	0.066	1	0.85	0.79	0.78	0.83	0.87	-0.71	0.81	TPSA
-0.37	-0.51	0.87	0.97	0.43	0.85	1	0.86	0.84	0.81	0.93	-0.67	0.82	ABCGG
-0.44	-0.51	0.73	0.92	0.2	0.79	0.86	1	1	0.74	0.96	-0.73	0.83	nAromAtom
-0.44	-0.51	0.71	0.91	0.2	0.78	0.84	1	1	0.74	0.94	-0.72	0.82	naromBond
-0.33	-0.6	0.81	0.82	0.21	0.83	0.81	0.74	0.74	1	0.78	-0.54	0.79	nN
-0.39	-0.5	0.84	0.97	0.19	0.87	0.93	0.96	0.94	0.78	1	-0.76	0.87	nbondsKD
0.36	0.41	-0.65	-0.71	-0.03	-0.71	-0.67	-0.73	-0.72	-0.54	-0.76	1	-0.8	DE
-0.44	-0.53	0.78	0.86	0.16	0.81	0.82	0.83	0.82	0.79	0.87	-0.8	1	CV

PC · Molar · Dbonds · MW · F · TPSA · ABCGG · nAromAtom · naromBond · nN · nbondsKD · DE · CV

FIGURE 2.11: Matrix plot showing correlations among the features that were found to be important for the construction of the DTs that show high classification accuracy.

can be obtained only when a large number of data points are available. Therefore, the non-parametric method of Kernel Density Estimation (KDE) was selected to obtain the distributions for these features. KDE method is based on the calculation of probability density function in which a linear combination of suitable kernel functions is taken which provides a good approximation of the calculated probability density. Thus a continuous probability density curve is obtained. These KDEs were generated using the Gaussian kernel function in the Seaborn package of Python. Gaussian kernel function tends to provide smoother estimation than other kernel functions. These features have been selected by the tree-based algorithms because of their information richness, ordered structure, and variance which are relevant for the prediction of the proton conductivity. The histograms and KDE plots have been color-coded according to the class labels for ease of understanding.

In the case of the molar content of the sulfonate group (Molar) in the monomer, the distribution tends to be skewed normal for the LPC class while for the HPC class, it

is a bimodal distribution with an edge peak present towards the value 100%. Also, it was observed that in the case of repeat units associated with LPC class labels, there are two prominent peak counts of Molar content. The first highest peak count is between 60-70% molar content and the second highest is between 40-50%. In the case of HPC, the prominent peak count of molar content is between 90-100%. This represents homopolymers as there is no non-sulfonated part of the repeat unit. The second peak is between 70-80%. The histogram for the distribution of the number of double bonds (dbonds) is a scattered one but the KDE plot shows the near-normal distribution for the LPC class with a prominent peak at 12.5. For the HPC class, a bimodal KDE distribution is obtained with a prominent peak around 8. Then again another peak area is obtained between 12.5 and 15. Molecular weight(MW) for LPC is near normally distributed around 1250 - 1500 g.mol^{-1} with a slight edge peak towards 2000 g.mol^{-1}. For HPC, MW shows bimodal distribution with a prominent peak at 1500 g.mol^{-1}. TPSA shows a sharp/centered normal distribution for LPC class labels between 300-400Å^3 while its distribution for HPC class labels is slightly merged bimodal and spread over 200-300Å^3 and around 400Å^3. Though threshold value could not be achieved from pruned DT, LPC distribution tends to give a threshold value for TPSA at around 350Å^3. The number of Nitrogen atoms shows bimodal distributions in the case of both LPC and HPC classes. HPC class labels are majorly associated with three and 5-6 Nitrogen atoms while LPC class labels are associated with 5-6 class labels. This indicates that a lesser number of N-atoms is somewhat related to higher proton conductivity while more number of N-atoms may or may not help in increasing proton conductivity. Graovac-Ghorbani atom-bond connectivity index (ABCGG) shows near-normal distribution with a slight edge peak towards the right for LPC and a slight bimodal distribution with a slightly higher peak between 40 and 50. In the case of the LPC class, the number of aromatic atoms shows a slightly normal distribution with the major peak centering around the values 60-80%. The number of aromatic atoms in the HPC shows a bimodal distribution with merged densities means the two peaks are not well-separated. The peaks are centered around 40 and 70-80 aromatic atoms. For the number of aromatic bonds in the repeat units, the distributions for both LPC and HPC follow distribution trends of the number of aromatic atoms. The peak for LPC appears at around 80 and the peak for HPC appears at around 40 and 80-90. The

number of kekulized double bonds (nBondsKD) show a normal distribution for LPC centered around 40 and 50 while a merged bimodal for HPC with high peak counts between 25-30 and 45-50. Dispersion energy has an edge peak towards the lower value of -100 kcal.mol^{-1}. The rest of the distribution is slightly skewed towards the higher end for both LPC and HPC. COSMO volume shows a skewed normal distribution centered around 15 Å^3 for LPC while a clear bimodal distribution with peaks around 7.5 Å^3 and 12.5 Å^3 for HPC. Detailed discussion on the features identified and the possible scientific basis for their relevance in the prediction of proton conductivity has been discussed in the 'Discussion on identified features' subsection.
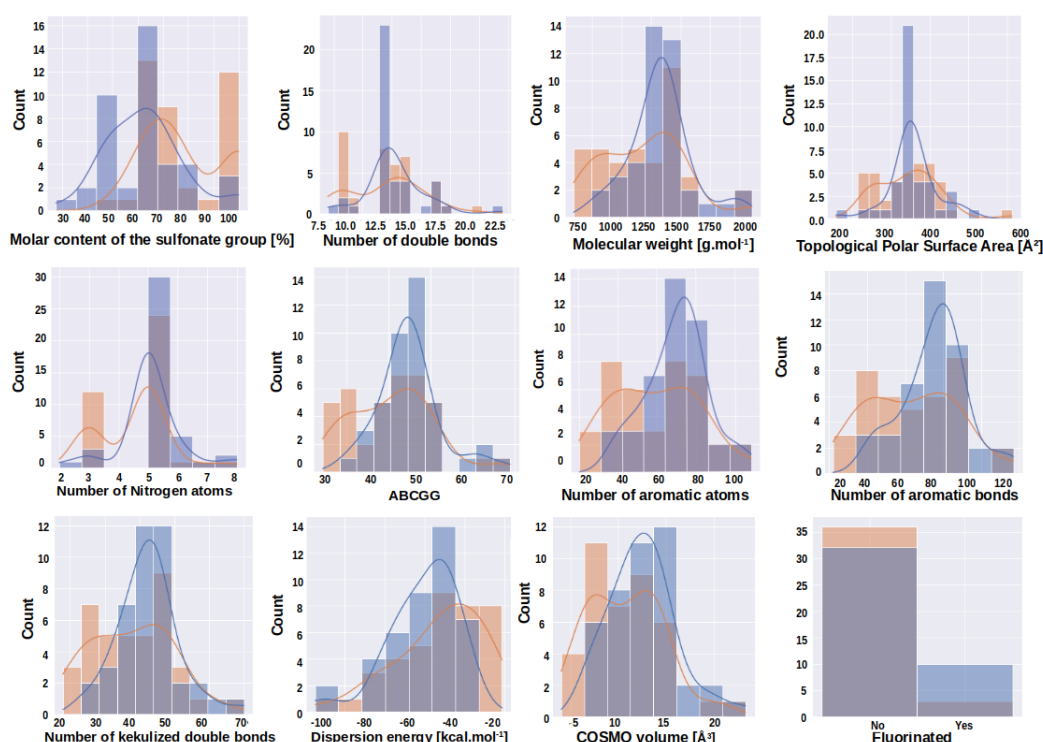
FIGURE 2.12: Class-wise univariate distribution and Kernel Density Estimation (KDE) plots of molar content of the sulfonate group(%), dbonds, MW, TPSA, nN, ABCGG, nAromAtom, nAromBond, Presence of Fluorine atoms; Blue represents LPC, Orange represents HPC.

2.4.4   Discussion on identified features

A total of 12 features namely Molar, dbonds, MW, TPSA, nN, ABCGG, nAromAtoms, nAromBonds, nBondsKD, Dispersion energy, COSMO volume, and presence of fluorine atoms have been identified to be important for the prediction of proton conductivity in SPI PEMs by the supervised learning in the form of tree-based ML algorithms

TABLE 2.3: Mean, standard deviation, and range of the selected features

Feature	Mean	Std deviation	Range
Molar fraction of the sulfonate group (%)	70.26 %	17.54 %	27.5% to 100.00%
Number of Nitrogen atoms	4.83	1.16	2.00 to 8.00
Topological Polar Surface Area [Å^2]	350.95	70.46	157.96 to 612.93
Molecular weight	1268.72	317.21	593.54 to 2105.57
Number of double bonds	12.41	2.82	7.00 to 22.00
Number of aromatic atoms	64.74	19.59	20 to 108
Number of aromatic bonds	69.85	21.20	22 to 117
ABCGG	43.143	8.84	24.37 to 68.74
nBondsKD	40.34	10.387	18 to 70
Dispersion energy [kcal.mol^{-1}]	-57.21	19.607	-114.62 to -21.34
COSMO volume [Å^3]	1311.22	338.027	577.64 to 2170

DT classifier and RFR.

These features have been identified from the number of points available in the present data set and thus may not provide a complete picture of how proton conductivity is related to these empirically. However, even if more data points are made available, ML algorithms can only identify the non-parametric and statistically significant features related to the target variable rather than the parametric and empirically related features. Therefore, features identified by ML algorithms for any chemical physics variable can only point in the direction of the more reliable empirical yet unexplored dimensions. Moreover, only a few attributes appear in the pruned DT classifier, nonetheless, all the identified features are important for the construction of this tree classifier. Removal of any of these led to decreased classification accuracy. Since proton conductivity is a compound effect of many factors, the scope of this work is limited to discussing what differences can be made at the monomer designing level so that there are good chances of obtaining a higher proton conductivity in the final SPI polymer regardless of other factors.

On exploring the relevance of these features, we find that these features can be divided into two broad categories. Firstly, the features that are directly related to the chemical structure of the repeat unit i.e. MW, dbonds, nN, nAromAtoms, nAromBonds, nBondsKD, ABCGG, and the presence of fluorine atoms play a role in deciding the shape of the microstructure or morphology of the polymer electrolyte membrane which in turn is responsible for determining the nature of proton transport. Morphology can be controlled by carefully designing the monomer molecule and incorporating

design principles obtained by the DT. Functional groups, terminal groups, and bridging linkages determine the kind of microstructure or morphology that will be formed by the polymer chains by interacting with each other and also with water molecules. QSPR features provide quantitative values for these functional groups and chemical moieties.

Apart from the feature related to the chemical structures, there are features that are related to the prominent interfacial region that exists between the polymer chain, water molecules, and protons. This interfacial region has many interesting properties which determine the extent of proton transport in the short as well as long range. Dispersion energy, COSMO volume, Molar, and TPSA are the features that can have a significant effect on the interfacial region. The interface brings about some structural changes in the water and its hydrogen bond network and also affects the proton transport properties because of the confinement of water in a nanoscopic volume or pore which could have any shape- cylindrical, slab-like, spherical, or lamellar. The properties of the interfacial region are also determined by the molar content of sulfonate groups in PFSAs as well as SPIs. The distribution of partial charges over the acidic sulfonate group determines the pKa or acidic potential of the sulfonate group. Dissociation of the proton takes place in the interfacial region thus readiness of the acidic group to donate the proton to the surrounding water molecules.

2.4.4.1   Molar content of the sulfonate group

The molar content of the sulfonate group (Molar) is perhaps the most critical factor in designing a PEM. On a broad level, insufficient sulfonate groups result in lesser bound water. More water molecules will be required to make the solvation shells grow and formation of bulk water domains. The polymer matrix may or may not have sufficient free volume to accommodate that much water and thus overlapping solvation shells and proper bulk water domains may never exist in such SPIs which have molar content lesser than 55%. The interconnectedness of the water domains and the presence of adequate proton concentration in those domains depends on the interplay of the Molar content of acidic groups in the polymer, IEC, and water uptake. The rest of the chemical structure may not be too useful if an adequate number of water molecules is not bound by the ionic groups. Moreover, even if there is sufficient free volume to

accommodate excess water around sulfonate groups such that bulk water channels are formed, there may not be a sufficient concentration of protons that travel in the bulk water. As Berrod et al. have pointed out in the study of PFSAs, there are three slow diffusing protons in the form of interfacial hydronium ions bound to each sulfonate group in Nafion(Berrod, Hanot, et al., 2017). Similar slow diffusing bound hydronium ions are not enough to provide facile proton transport as happens in the bulk water through the Grotthus mechanism. Also, the surface diffusion of the proton which happens when solvation shells of adjacent sulfonate groups overlap becomes difficult when the sulfonate groups are located far away from each other.

2.4.4.2   Double bonds

The number of double bonds and kekulized double bonds (kBondsKD) was found to be important for the prediction of proton conductivity by the DT classifier and RFR respectively. The share of double bonds comes from imide groups and some from aromatic rings and the remaining from any ketone or sulfone groups present in dianhydrides and diamines. The Sulfonate group (-SO$_3$H) itself has double bonds. These possibly play important roles in molecular ordering by governing the interaction of one chain with the other as they are rigid in nature. Kekulization of double bonds is related to the localization of the electron density in aromatic rings and gives an idea about the distribution of charge density only at double bond regions. The QSPR feature nBondsKD calculated by Mordred provides the number of the kekulized double bonds as opposed to the delocalized double bonds in aromatic rings. Another factor identified by the DT classifier algorithm is the aromaticity quantified in terms of nAromAtoms (number of Aromatic atoms) and nAromBonds (number of Aromatic bonds). Aromatic atoms are present in the form of Phenyl or Phenylene components of dianhydrides and diamines. The phenyl group is known to be hydrophobic in nature. The presence of biphenylene rings contributes to the rigidity of the SPI chains. Generally, due to the rigid nature of SPI chains, a rigid polymer matrix is formed around ionic domains which resist drastic rearrangement upon filling and defiling of water molecules. Observations made by the Essafi group were for NTDA/BDSA/ODA in which naphthalene rings tend to make chains rigid and increase the interchain spacings. These results cannot be generalized for all the SPIs(Essafi, Gebel, and Mercier, 2004b).

2.4.4.3 Nitrogen atoms

Thirdly, the presence of Nitrogen atoms in the repeat units in the form of imide groups and heterocyclic diamines provides electronegative centers. Nitrogen atoms have a higher electron density than the adjoining carbonyl groups thus intra-chain and inter-chain charge transfer complexes can form. Also, Nitrogen atoms are basic in nature and hence they tend to act as proton acceptors near sulfonic acid sites and hence leading to the stabilization of the Hydrogen bond network. These two factors play a major role in the formation of the morphology of the polymer matrix around the ionic domain. A negative correlation indicates that too many nitrogen atoms along the chain may lead to a decrease in the overall proton conductivity of the SPI. This could be due to the fact that too many Nitrogen atoms tend to divert the protons away from the water domains and decrease the concentration of protons in the channels. It is pertinent to add here that fluorination is an important factor in the design of a PEM as the presence of fluorinated groups results in a dramatic change in the overall properties of SPIs. This is due to the combination of electronic and steric effects which reduce the ability for interchain interactions and hinder the formation of charge-transfer complexes. These are major factors in the packing of the polymer chains arising from the molecular design. These introduce intractability in aromatic SPIs. C-F bond is a high energy bond and SPIs containing fluorine in general result in low moisture absorption and a well-segregated morphology.

2.4.4.4 Topological Polar Surface Area (TPSA

TPSA is an important attribute as it is calculated by the presence of polar atoms namely Oxygen and Nitrogen atoms along with their attached Hydrogen atoms. Such polar atoms have high electronegativity and are known to form Hydrogen bonds with Hydrogen atoms of water molecules. In SPIs, Nitrogen, and Oxygen atoms are present in the imide group and contribute to the polarity of the dianhydride part of the repeat unit. The sulfonate group also has oxygen atoms that have enough polarity to stabilize the water molecule network around them which helps in the facile dissociation of protons. The presence of polar atoms in the monomer other than the sulfonate group can have effects that may or may not promote proton transport through hydrogen

bonding. If the polar atoms are placed such that they tend to stabilize the Hydrogen bonding network away from the sulfonate group, then they are in a way taking up the water molecules that could have formed a solvation shell around the ionic group. Such an effect will decrease the proton conductivity. More polar atoms can be incorporated into the repeat unit through careful selection of diamines. Due to its role in forming hydrogen bonding, TPSA also contributes to the formation of interfacial regions between hydrophilic and hydrophobic regions and thus is an important parameter. The critical (or threshold) value of TPSA could not be known from the DTs as these are pruned trees and it is possible that TPSA appeared as a terminal node and got removed in the post-pruning process but its removal caused a decrease in accuracy.

2.4.4.5 Dispersion energy and COSMO volume

Dispersion energy originates from the intermolecular dispersion forces that arise from the correlated fluctuations in the electron density of the molecules. Both proton dissociation and proton transport phenomena can be considered as a many-body problem involving acidic groups, water molecules, hydrogen cation, and hydronium ions. Thus polarizabilities of the interacting molecules i.e. repeat units and water (solvent) play an important role in determining the nature of the interaction between repeat units of adjacent chains and repeat units and water molecules. Polarization arises due to the coupling of charge environments of the interacting molecules and the formation of instantaneous dipoles. This charge density on the repeat unit and hence the polarizability that arises due to the accumulation of water molecules around the acidic groups plays a role in proton dissociation. As a rule, a more polarizable repeat unit tends to bind the solvent more strongly. Repeat units with more aromatic rings are usually highly polarizable. COSMO volume is calculated by MOPAC by constructing a conducting polygonal surface around the ion or molecule at van der Waals distance and is related to the surface polarity distribution. Together dispersion energy and COSMO volume can be considered to influence the interfacial region and hydrogen bonding significantly.
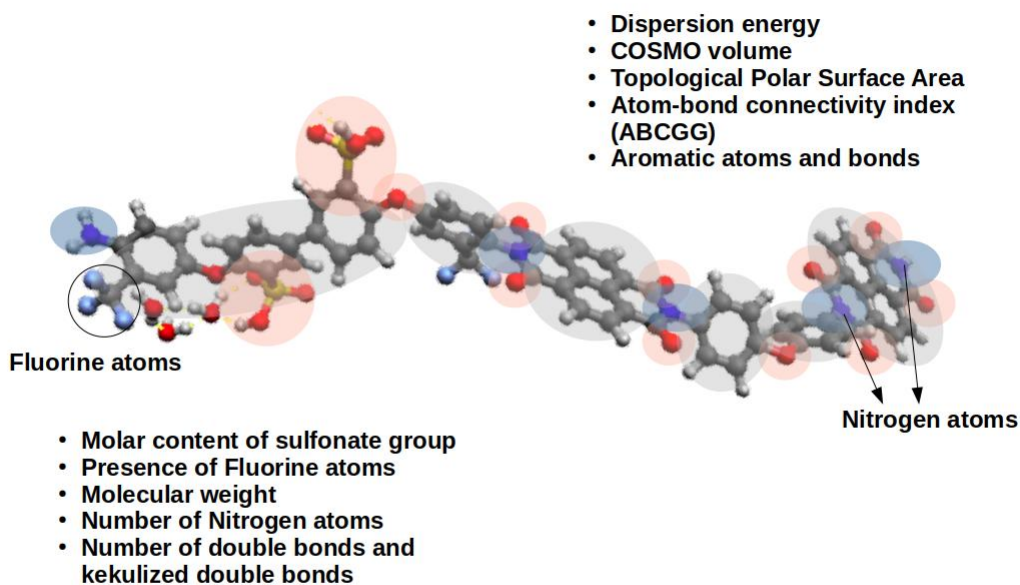
FIGURE 2.13: A sample fluorinated SPI repeat unit showing the identi-
fied features at the monomer level

2.5   Conclusion

In the present study, a combination of supervised and unsupervised machine learning approaches was utilized to predict the proton conductivity of sulfonated polyimide polymer electrolyte membranes using data sets generated from properties collected from reported literature, QSPR descriptors, and semi-empirical quantum chemical descriptors. Using DT classifiers, simplistic and interpretable models were obtained in the form of If-else statements from a data set of just 81 unique SPI monomers. Random forest regression models were also trained which made low-error predictions. To the best of our knowledge, this is the first demonstration of the prediction of class labels for estimating proton conductivity of novel monomers belonging to a specific class of hydrocarbon polymer electrolyte membrane i.e. sulfonated polyimide in this case.

Also, a subset of the most relevant features was selected by manually including the features that increased the prediction accuracy of the DT classifier. Best prediction accuracy was obtained from the data set which included QSPR features - Presence

of Fluorine atoms (F), Number of double bonds (dbonds), Topological Polar Surface Area (TPSA), Molecular weight (MW), number of Nitrogen atoms (nN), number of aromatic atoms (nAromAtoms), number of aromatic bonds (nAromBonds) along with the literature feature related to the monomer design feature molar content of the sulfonate group(Molar). 10-fold cross-validation ensures that despite the less number of data points reliable predictions, feature importance, and monomer design insights can be obtained. RFR identified some additional QSPR features namely a number of kekulized double bonds (nBondsKD) and Graovac-Ghorbani atom-bond connectivity index (ABCGG) as well as dispersion energy and COSMO volume which are the features obtained from semi-empirical quantum chemical calculations.

The rules obtained encompass different descriptors or monomer designing criteria like the critical number of double bonds, number of Nitrogen atoms, and number of aromatic atoms and bonds which make a difference at the monomer level to improve inter-chain interactions, polymer chain stacking and micro-phase segregation so that high proton conductivity is obtained at the material level as the cumulative effect. The critical value of the molar content of the sulfonate group in the monomer/repeat unit determines the water-binding behavior of the polymer and also is indicative of how much proton concentration can be expected to be present in the polymer with increasing levels of hydration. Thus, using open-source descriptor calculators and data mining/machine learning approaches, rich insights have been gained into designing newer Sulfonated Polyimide Polymer Electrolyte Membranes.

3 Atomistic study

3.1 Chapter summary

Two novel hydrocarbon-based PEM belonging to the Sulfonated Polyimide (SPI) class of PEM ionomers- a partially fluorinated SPI and a non-fluorinated SPI have been designed in-silico and their properties compared with commercially available Nafion PFSA ionomer. In the present work, detailed all-atomistic Molecular Dynamics (MD) simulations have been used to investigate the state of nanophase segregation, the morphology of the ionic domains, and dynamics of proton transport with increasing hydration level ($\lambda = 1, 5, 10$ and 15). The diffusion coefficients of hydronium ions and water molecules and corresponding proton conductivity owing to the transport of hydronium ions in water channels were calculated. Proton conductivity values were found to be highest for non-fluorinated SPI lying within the range 0.03-0.18 S·cm^{-1} for ($\lambda = 1$ to 15). Also, structural analyses have been done using radial distribution function (RDF) plots between sulfur atoms of sulfonate groups and solvent phase (hydronium ions and water molecules) along with polymer and solvent volume fractions have been calculated.

3.2 Introduction

Multi-scale Molecular Dynamics (MD) simulations have revealed rich molecular-level insights into the mechanisms of nano-phase segregation and proton transport in PFSAs. In PFSAs, MD studies have led to the well-established understanding that the nano-segregated morphology is significantly influenced by the equivalent weight, length of the side chain, and water content of the membrane. Devanathan et. al. have shown using mean residence time studies that there is an active movement of the hydronium ions and water molecules in water channels formed and their dynamics change

with increasing hydration levels. Usually, at lower hydration levels, hydronium ions are surrounded by more sulfonate groups acting as bound bridges between sulfonate groups. With increasing water content, more water molecules surround hydroniums instead of sulfonate groups. Moreover, hydronium tends to move from the first solvation shell around the sulfonate groups to the center of the pore or water cluster (Devanathan, Venkatnathan, and Dupuis, 2007a). Karo et. al. compared the structural and transport properties of Nafion PFSA ionomer with shorter side chain PFSA ionomer namely, Hyflon using all-atomistic studies. They found that there is slightly greater sulfonate-group clustering in Hyflon compared to Nafion. Moreover, diffusion coefficients of water and hydronium were more in Hyflon (6.5×10^{-6} cm$^2 \cdot$s^{-1} and 25.2×10^{-6} cm$^2 \cdot$s^{-1} respectively) compared to that in Nafion (6.1×10^{-6} cm$^2 \cdot$s^{-1} and 21.3×10^{-6} cm$^2 \cdot$s^{-1} respectively) because of the presence of more hydronium ions in the center of the water channel rather than bound to the sulfonate group as in case of longer side chain PFSA (Karo, Aabloo, et al., 2010). Cui et. al. studied Nafion PFSA ionomer with an equivalent weight (EW) of 1143 with a degree of hydration varying from 5 to 20 wt% of water. They found that the distribution of water clusters is more dispersed (Cui, Liu, et al., 2007).

It is pertinent to mention the limitation of MD simulations in describing the proton conductivity of PEMs. In conventional MD simulations, diffusivities of hydronium ions and water molecules are studied which contribute only a fraction to the total diffusivity of proton in bulk water and also in PEM ionomers. Structural diffusion or the Grotthus mechanism occurs through the hopping of the proton from one water molecule to the adjacent water molecule. Several researchers have incorporated bond-creating and bond-breaking provisions through Multi-state- Empirical Valence Bond (MS-EVB) in conventional MD simulations to characterize the extent of structural diffusion taking place in PFSA ionomers (Tse, Herring, et al., 2013).

3.2.1   Molecular dynamics studies of SPI PEMs

Nevertheless, insights gathered from MD simulations could help in designing more efficient hydrocarbon PEMs with predictable morphology that offers a conducive environment for facile vehicular and structural diffusion of the protons. These computational tools have also been utilized to study the structural and dynamic properties

of SPI-based PEM. For instance, Park et. al. utilized atomistic simulations to model NTDA/ODADS/ODA SPI PEM model with the COMPASS II force field. With these simulations, they observed the effect of the level of hydration on the morphology of the hydrated SPI PEM. They observed that too many water clusters and channels observed were not formed in hydrated SPI, unlike the Nafion PFSA ionomer. The critical threshold level of hydration in the studied SPI PEM was identified between 6 to 8.5 (Park, Kim, et al., 2019). In a previous study, block copolymers of NTDA/ODADS/ODA were studied and it was concluded that SPI ionomers have generally distributed water molecules and hydronium ions in the free volumes rather than an ordered lamellar cluster formation like that observed in Nafion majorly due to the rigid nature of SPI polymer chains. Hu et. al. conducted a mesoscale study wherein different sequence arrangements of block copolyimide namely NTDA/BSPA/DMDA comprising blocks of different lengths of hydrophilic and hydrophobic repeat units were modeled (1,4,5,8-Naphthalenetetracarboxylic dianhydride (NTDA) and 3,3'-Bis(sulfopropoxy)-4,4'-diamino diphenyl (BSPA) and hydrophobic repeat unit synthesized by NTDA and 1,10-decamethylenediamine (DMDA)). They found that a uniform arrangement of hydrophobic and hydrophilic segments/blocks along the SPI chain led to the formation of a homogeneous microphase structure. This ensured rapid water transport in the channels, high dimensional stability, and better mechanical properties of the SPI PEM (Hu, Lu, and Guo, 2018).

In some elaborate efforts, experimental synthesis backed by computational efforts has also been reported. Garrido et. al. studied the diffusion coefficients of protons using PFG-NMR in NTDA/ODADS/ODA membranes at different hydration levels ranging from minimal hydration to high levels of hydration and also performed MD simulations on corresponding all-atomistic models. They found that these naphthalenic imide membranes were able to hold water molecules and diffusible protons even at minimal hydration conditions (Garrido, Pozuelo, et al., 2009). In sum, experimentation to discover alternative PEMs is extensive and time-consuming which often results in wasted efforts as the proton conductivity of the developed PEM is not able to be at par with the perfluorinated PEMs nor is there any understanding of the exact nano-scale morphology of the PEM and behavior of proton transport in them. To

this end, a combined effort toward the simulation-informed synthesis of SPI PEM pro-

viding an understanding of the structure-property relationship is still elusive. Still,

an elaborate discussion on the formation of ionic domains and the dependence of the

proton transport dynamics on the morphology in SPI PEMs is still required.
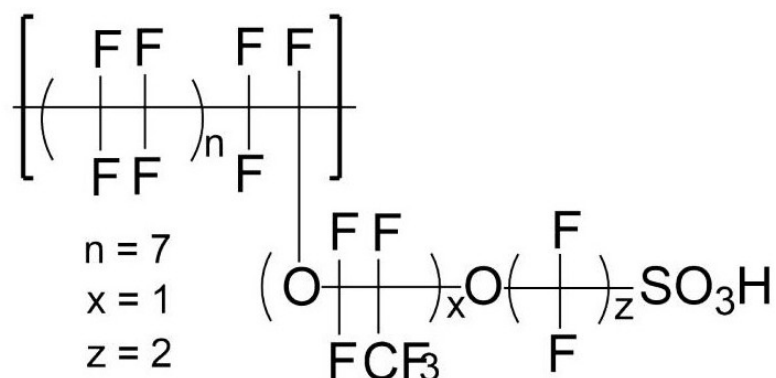
FIGURE 3.1: Chemical structures of the repeat units of Nafion perfluori-
nated sulfonic acid ionomer modeled in the present study

(a)

(b)
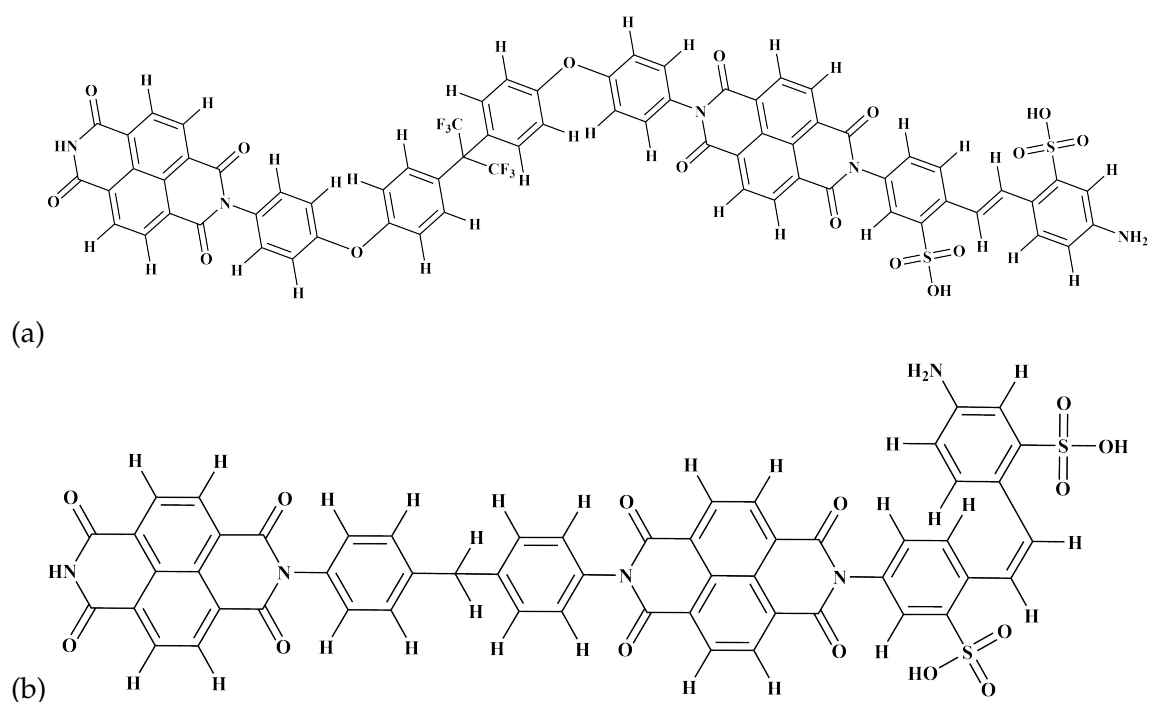
FIGURE 3.2: Chemical structures of the repeat units of (a)
NTDA/DSDSA/HFBAPP (fluorinated SPI) ionomer, and (b)
NTDA/DSDSA/MDP (non-fluorinated SPI) ionomer modeled in
the present study; NTDA= 1,4,5,8-naphthalene tetracarboxylic dian-
hydride, DSDSA = 4,4-diamino stilbene-2,2-disulfonic acid, HFBAPP
= 2-bis[4-(4-amino phenoxy) phenyl] hexafluoro propane, and MDP =
4,4'- Diaminodiphenyl methane

3.3 Computational details

The selection of these particular SPIs namely the semi-fluorinated SPI i.e. NTDA/ DS-DSA/HFBAPP and the non-fluorinated SPI i.e. NTDA/DSDSA/MDP was based on a previous study conducted by the authors which are reported elsewhere. The study involved the utilization of decision trees trained on datasets derived from previously reported proton conductivity data and Quantitative structure-property relationship (QSPR) and semi-empirical descriptors calculated from the chemical structures of SPI repeat units. The data-driven approach provided certain rules for designing SPI repeat units that provide PEMs with high proton conductivity. These rules are provided in Table 3.1 in the Supporting Information. The repeat units modeled in the present work conform to these rules.

TABLE 3.1: Rules for designing high proton conductivity SPI PEMs obtained by training Decision trees with previous data on SPI PEMs

	Number of double bonds (dbonds)	Molecular weight (MW) [gm·mol^{-1}]	Number of aromatic atoms (nAromAtoms)	Number of aromatic bonds (nAromBonds)
FSPI	13	1370.18	68	74
NFSPI	13	1034.98	56	62

All-atom molecular dynamics simulations were performed to investigate the proton transport behavior in FSPI, NFSPI, and Nafion. Related analyses were performed to characterize the local morphology of the ionic domains. The local structure was derived using RDF analysis while structural diffusion was characterized using diffusion coefficients obtained from MSD of hydronium ions and water molecules. Connolly surface analysis was performed to characterize the overall morphology and interconnectedness of the ionic domains.

3.3.1 Simulation framework

3.3.1.1 Initial geometry/Model construction

Repeat units of the novel FSPI, NFSPI, and Nafion were created in the Material Studio visualizer according to the chemical structures shown in Fig. 4.1. Three chains of

the polymer each comprising four repeat units were packed into simulation cells with an increasing number of water molecules. The sulfonate groups were considered to be completely deprotonated and the excess protons were considered hydronium ions. Simulation cells of FSPI were prepared in the Amorphous cell module of the Accelrys Material Studio. PCFF91, a class II force field for polymer modeling, was applied to the polymer chains, water molecules, and hydronium ions. PCFF has been previously used for modeling polyelectrolytes - Nafion, SPEEK, SPI, water, and hydronium. Table 3.2 provides the details related to the Molecular weight of the chain [gm $\cdot$ mol^{-1}], Number of sulfonate groups or degree of sulfonate, Equivalent weight [gm$\cdot$mol^{-1}] and Ion Exchange capacity [milliequiv.$\cdot$gm^{-1} of dry polymer] for Nafion PFSA ionomer, FSPI, and NFSPi modeled in the present work. An amorphous cell module was used for generating the FSPI, NFSPI, and Nafion PFSA ionomer models at different hydration numbers represented by the values of λ = 1,5,10, and 15. Partial charges were assigned according to the Gasteiger method.

TABLE 3.2: Physico-chemical details of the modeled PEMs- Nafion PFSA, NTDA/ DSDSA/HFBAPP (FSPI) and NTDA/DSDSA/MDP (NFSPI)

	Molecular weight of the chain [gm$\cdot$mol^{-1}]	Number of sulfonate groups (degree of sulfonation[%])	Equivalent weight (EW) [gm$\cdot$mol^{-1} of SO$_3^-$]	Ion Exchange Capacity (IEC) [milliequiv. $\cdot$gm^{-1} of dry polymer]
Nafion	4610.8	4 (100%)	1152.7	0.867
FSPI	5414.6	8 (75%)	676.83	1.477
NFSPI	4125.8	8 (75%)	515.73	1.939

3.3.1.2 Minimization and Equilibration

Molecular dynamics protocol as followed for simulating NTDA/ODADS/ODA(3/1) by Garrido et. al. (Garrido, Pozuelo, et al., 2009). All MD simulations and corresponding analyses related to RDF and MSD were done in the Forcite module of Accelrys Material Studio. Energy minimization was performed using the Conjugate Gradient algorithm with maximum iterations keeping the maximum derivative smaller than 0.1

kcal·(Å^{-1}·mol^{-1}) and a limit of 5,000 steps. The criteria for stopping the minimization run was set at 0.1 kcal·(Å^{-1}·mol^{-1}). In order to obtain realistic geometries, the energy-minimized geometries were subjected to a 300ps run of Constant temperature, constant pressure (NPT) (McDonald, 1972) ensemble at 300K and 1bar maintained by the Berendsen barostat. The simulation cells attained densities in the realistic density range for Nafion. Constant temperature, constant volume (NVT) (Panagiotopoulos, 1987) simulation was performed at 300K maintained by Andersen thermostat. The simulation cells were run for 300ps for NVT equilibration such that minimized total energy remains approximately constant with respect to the simulation time. The final data collection run was done for 3ns in the NVT ensemble. A timestep of 1fs was taken for the integration of the equations of motion. The coulombic interactions were calculated using the Ewald summation method (Veld, Ismail, and Grest, 2007) with Verlet method(Grubmüller, Heller, et al., 1991). NVT and NPT thermostat relaxation parameters were kept at 1ps and 0.1ps respectively while the NPT barostat relaxation time parameter was kept at 0.3ps. van der Waals and Coulombic non-bonding interactions were calculated by the Cell Multipole method (CMM).

3.3.2 Analysis studies

3.3.2.1 Mean Square Displacement (MSD) and Diffusion Coefficient (DC)

Diffusion coefficients of water molecules provide information about the self-diffusivity of water molecules in the ionic domains formed in the ionomer membrane. Mean-square displacement (MSD) data collected over adequate simulation time provides information regarding the nature of the movement of particles diffusing through a medium. MSD data of the oxygen atoms of water molecules (O_{H_2O}) and the oxygen atoms of the hydronium ions ($O_{H_3O^+}$) were collected at the interval of every 0.5 ps i.e. at every 500 steps with each step of 1 fs. A total simulation time of 3 ns was used for each sample for calculating the diffusion coefficients for increasing water content represented by the values of λ = 1, 5, 10, and 15. MSD data was used in the Einstein equation [Eq. 3.1].

$$MSD = \frac{1}{N} \sum_{j=1}^{N} [(r_j(t) - r_j(O))^2] \tag{3.1}$$

$$D = \frac{1}{6N} \lim_{t \to \infty} \frac{d}{dt} \sum_{j=1}^{N} [|r_j(t) - r_j(O)|^2] \tag{3.2}$$

where $r_j(O)$ = initial coordinates of the particle; $r_j(t)$ = coordinates of the particle at time t; N = total number of particles in the system

3.3.2.2   Proton conductivity

Proton conductivity was calculated from the diffusion coefficient of the hydronium ions according to the Einstein equation Eq. 3.3.

$$\sigma = \frac{N.z^2.e^2.D_{H_3O^+}}{V.k.T} \tag{3.3}$$

where e =Electron Charge, N = Number of charged ions in the system, Boltzmann constant (K) = 1.38×10^{-23} J·K^{-1}, T = 298K, $D_{H_3O^+}$ = Diffusion coefficient of hydronium ions [cm^2·s^{-1}], V = Volume of the cell [Å^3].

3.3.2.3   Radial Distribution Function (RDF)

Radial Distribution Function (RDF) denoted by $g_{A-B}(r)$ is proportional to the probability of finding an atom B at a distance r from a reference atom A inside a shell of thickness dr as given in Eq. 3.4.

$$g_{A-B}(r) = \frac{V}{N_B} . \frac{n_B}{4\pi r^2 dr} \tag{3.4}$$

where N_B = Number of B particles in the system; n_B = Number of atoms B located at the distance r from atom A in a shell of thickness dr; V = total volume of the system.

Coordination Numbers (CN) for the polymer and solvent phases were calculated by integrating the peaks of RDF plots.

3.3.2.4   Volume analysis

FFV is the volume of the simulation cell not occupied by the polymer. Water-channel structure influences proton transport properties. Analyzing the volume of water gives information on the nano-phase structure. Connolly surfaces were calculated with a probe radius of 0 Å.

$$FFV = \frac{V_{total} - 1.3V_{vdW}}{V_{total}} \tag{3.5}$$

3.3.2.5 Visual analysis

Snapshots of the simulations were captured and represented using OVITO (Stukowski, 2009). As the dimensions of features (ionic domains and channels) are roughly in Å and nm scale, the structure of water channels formed due to sequential rearrangement of polymer matrix with increasing water content could be estimated.

3.4 Results and discussion

3.4.1 Equilibrium validation

Table 3.3 lists the densities obtained after equilibrating the samples in unprotonated form and in protonated form with hydronium ions and an increasing number of water molecules to represent increasing water content. In the dry state or unprotonated state, the density of Nafion was found to be 2.227 g.cm^{-3}. The experimental density of Nafion 117 (EW =1100) at 300 K has been reported to be 2 g.cm^{-3} (Eikerling, Kornyshev, et al., 2001). The densities for FSPI and NFSPI in dry/unprotonated state were 1.334 g.cm$-3$and 1.35 g.cm-3. The equilibration steps are thus providing the equilibrated samples for Nafion and novel SPIs. Moreover, for increasing water content, density was found to decrease for Nafion, FSPI, and NFSPI. This is due to an increase in the volume with an increasing number of water molecules. The density of Nafion is higher than both of the novel SPIs due to the presence of bulky fluorine atoms in the backbone and side chains. Fig. 3.3 shows the variation in density of the equilibrated models for dry and increasingly hydrated Nafion PFSA, FSPI, and NFSPI ionomers.

TABLE 3.3: Density [g·cm^{-3}] obtained after equilibration

Level of hydration (λ)	Nafion PFSA	FSPI	NFSPI
0	2.227	1.334	1.35
1	2.107	1.391	1.328
5	2.019	1.333	1.263
10	1.914	1.244	1.17
15	1.79	1.12	1.080
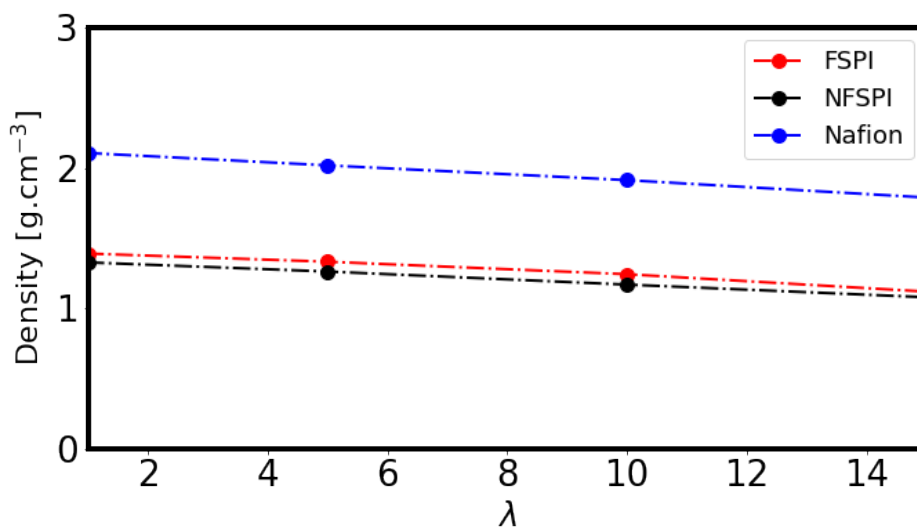
FIGURE 3.3: Densities of the equilibrated systems

3.4.2 Proton transport

Proton transport in hydrated ionomer membranes is a complex/cumulative/compound phenomenon and is subject to many factors including the formation of water channels and clusters with increasing water content, the presence of ionic species formed from different combinations of water molecules, and excess protons which aid in the transport of proton namely hydronium, Zundel and Eigen ions. Table. 3.4 in the Supporting Information presents the values of diffusivities of hydronium ions and water molecules in different ionomers at 300 K and different levels of hydration and also in bulk water as reported in the literature. It is pertinent to mention the limitations of atomistic simulations which usually span only a few nanoseconds. Despite best efforts to model the systems as closely and reasonably as possible, there are limitations related to length and time scale and there may be over- or under-estimation of dynamic properties such as diffusion behavior.

3.4.2.1 Diffusion coefficients of water molecules

Fig. 3.4 shows the variation of self-diffusion coefficient of water molecules in the water channels formed in the polymer matrices of Nafion PFSA (Blue), FSPI (Red) and NFSPI (Black) ionomers with increasing levels of hydration. In Nafion, there is a negligible

TABLE 3.4: Values reported in the literature for diffusion coefficients of hydronium ions ($D_{H_3O^+}$) [cm^2·s^{-1}]] and diffusion coefficients of water molecules (D_{H_2O}) [cm^2·s^{-1}] for increasing levels of hydration (λ)

Ionomer	Level of hydration (λ)	$D_{H_3O^+}$ [cm^2·s^{-1}]	D_{H_2O} [cm^2·s^{-1}]	Reference
SPI-8 (NTDA/ODADS/ ODA)	2 6 16.4	3.87×10^{-7} 6.3×10^{-7} 9.93×10^{-7}		Park, Kim, et al., 2019
SPI-8 (NTDA/ODADS/ ODA)	23	2.6×10^{-5}		Garrido, Pozuelo, et al., 2009
Nafion PFSA (QENS)	5 11 13.5	0.14×10^{-6} 0.63×10^{-6} 1.13×10^{-6}	0.96×10^{-6} 3.46×10^{-6} 5.27×10^{-6}	Devanathan, Venkat- nathan, and Dupuis, 2007b
Bulk water		5.98×10^{-5}		Agmon, 1995
Nafion PFSA (PFG-NMR)	3 6		0.73×10^{-6} 3.6×10^{-6}	Zawodzinski Jr, Neeman, et al., 1991
Nafion PFSA 1143 EW (Simulation)	4.4 9.6	0.2×10^{-6} 1.9×10^{-6}	1.02×10^{-6} 6.39×10^{-6}	Cui, Liu, et al., 2007
Nafion 117 PFSA (Experiment)	4.4 9.6	0.8×10^{-6} 5×10^{-6}	0.5×10^{-6} 4.4×10^{-6}	Cui, Liu, et al., 2007
Nafion 117 PFSA (Experiment)	4.4 9.6	0.8×10^{-6} 5×10^{-6}	0.5×10^{-6} 4.4×10^{-6}	Cui, Liu, et al., 2007
Nafion PFSA	10 14 16		3.38×10^{-6} 5.49×10^{-6} 5.88×10^{-6}	Ozmaian and Naghd- abadi, 2014
Nafion PFSA	3.5 7 10 13 16 25	1.0×10^{-6} 5.0×10^{-6} 1.0×10^{-5} 1.3×10^{-5} 1.6×10^{-5} 2.3×10^{-5}	3.5×10^{-7} 2.21×10^{-6} 6.32×10^{-6} 9.41×10^{-6} 1.65×10^{-5}	Zhang, Yang, et al., 2021

diffusion coefficient (4.18×10^{-8} cm^2·s^{-1}) at low hydration level as hydrophobic polymer backbone is densely packed, and thus the movement of water molecules is hindered in the absence of unobstructed pathways. Also, since hydrophilic sulfonic acid sites are inaccessible to water molecules there is no accumulation of water molecules and they are scattered in the simulation cell. At a moderate hydration level ($\lambda = 5$),

an increase in the level of diffusion can be observed (3.84×10^{-6} cm$^2 \cdot$s^{-1}). The percolation threshold for Nafion has been reported in the literature to be $\lambda = 5$ (Laporta, Pegoraro, and Zanderighi, 1999; Devanathan, Venkatnathan, et al., 2010; Liu, Cavaliere, et al., 2018). At optimal hydration level ($\lambda = 10$), the value obtained (1×10^{-5} cm$^2 \cdot$s^{-1}) is in good agreement with the value reported in the literature (0.909×10^{-5} cm$^2 \cdot$s^{-1}) (Zhang, Yang, et al., 2021) [refer Fig. 3.4 in the Supplementary Information]. This represents a facile proton transport regime where the water molecules are mobile in the water channels developed due to nano-phase segregation. Increasing water content facilitates the rearrangement of the polymer matrix and widens the water channels/pathways. At a higher level of water content ($\lambda = 15$), the diffusion behavior of water molecules starts to resemble bulk-water-like diffusion behavior. The diffusion coefficient of 1.508×10^{-5} cm$^2 \cdot$s^{-1} is observed which is about 1.8 times more than the experimental value for almost the same hydration level(Cui, Liu, et al., 2007). Hence the equilibrated model for Nafion can be considered reliable and the choice of force field can be considered valid. Thus these MD models for the novel SPIs can be relied upon to provide a close estimation of information regarding diffusion behavior and water channel morphology.

We observed that the diffusion of water molecules in both novel SPIs tends to be slightly different on the following accounts. As seen from Table **??**, the value of DC_{H_2O} for FSPI at ($\lambda = 1$) is 6.28×10^{-7} cm$^2 \cdot$s^{-1} which is about 15 times higher than that observed in Nafion at the same hydration level. Similarly, for NFSPI, the diffusivity of water molecules is 5.81×10^{-6} cm$^2 \cdot$s^{-1} at the lowest hydration level which is about 63 times higher than Nafion. Similar to Nafion, there seems to exist a percolation threshold level of hydration above which a drastic increase in the diffusivity of water molecules is observed. High levels of water molecule diffusion coefficients can be ascribed to various molecular details of the SPIs. Firstly, it is a general observation that naphthalenic SPIs have rigid chains packed in a planar arrangement which leads to the presence of large empty pockets or free volumes (Essafi, Gebel, and Mercier, 2004b; Park, Kim, et al., 2019). FSPI and NFSPI comprise higher free volumes compared to Nafion [discussed later in later sections]. The higher free volume provides more space for water molecules to occupy with increasing hydration levels and the large water-filled pores provide broad pathways for the diffusion of water molecules.
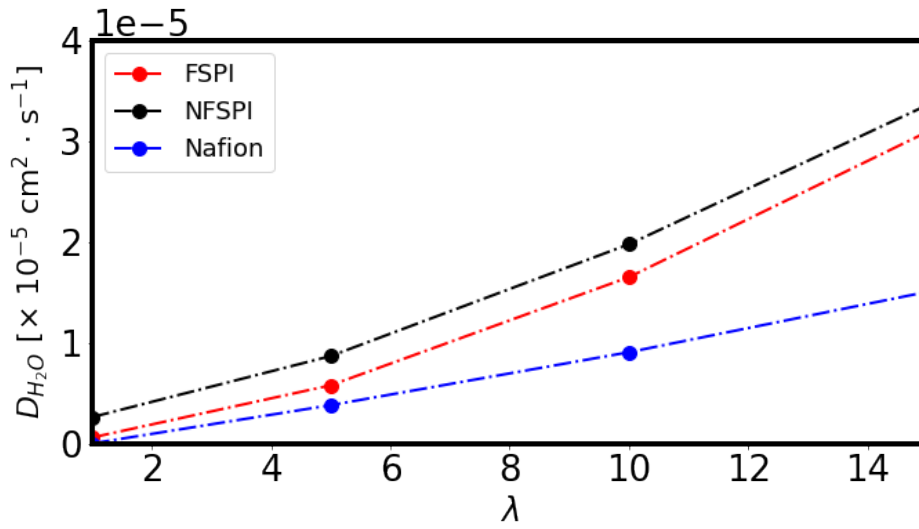
FIGURE 3.4: Variation observed in the diffusion coefficients of the water molecules with increasing hydration level

Usually, it is observed that the mobility of water molecules is more in the central region of the pore than near the wall. The diffusion coefficients tend to be higher if there is a sizeable central region in a water cluster.

3.4.2.2 Diffusion coefficients of hydronium ions

Fig. 3.5 represent the diffusion coefficients of hydronium ions ($DC_{H_3O^+}$) for Nafion (Blue), FSPI (Red) and NFSPI (Black) for various hydration levels (or increasing water contents). The diffusion coefficient of hydronium describes the rate of vehicular diffusion of the hydrated proton. The range of diffusion coefficient of hydronium ions ($D_{H_3O^+}$) obtained for Nafion in our simulated model is in good agreement with those reported in the literature (Zhang, Yang, et al., 2021) [Table 3.4]. Also, there is an agreement with the experimentally obtained diffusion coefficients for Nafion at varying water contents (Devanathan, Venkatnathan, and Dupuis, 2007b; Zawodzinski Jr, Neeman, et al., 1991). Evidently, at $\lambda = 1$ hydration level which denotes extremely low hydration conditions, no large-scale proton transport through vehicular mechanism is taking place. Only surface (en masse) diffusion can be considered to exist in this low hydration region but the movement or diffusion of the hydronium ions does not describe that. It can be seen that in extremely low hydration conditions ($1 \leq \lambda \leq 5$) FSPI and Nafion demonstrate almost similar diffusion behavior. In the moderate water content
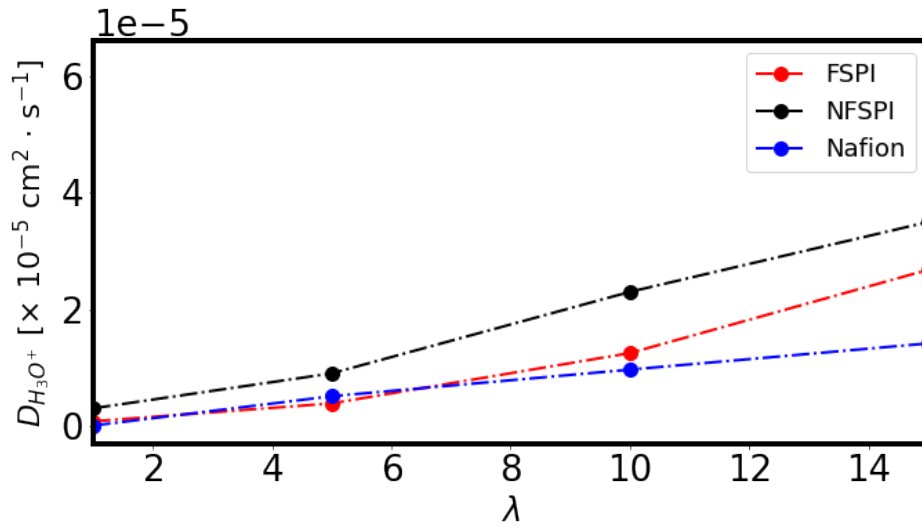
FIGURE 3.5: Variation observed in the diffusion coefficients of the hydronium ions with increasing hydration level

range ($5 \leq \lambda \leq 20$), $D_{H_3O^+}$ increases dramatically for both FSPI and NFSPI while the increase is more for the non-fluorinated SPI. This is ascribable to the fact that fluorination causes hydrogen bond formation between hydronium ions and fluorine atoms which hinders the movement of the former. At low and moderate hydration levels, values of $D_{H_3O^+}$ are highest for NFSPI. The values are much higher not only among SPIs but also more than that observed in Nafion PFSA.

Notably, the confinement effect is absent in SPIs i.e. large water clusters do not involve the confinement of water molecules and protons in nanoscopic volumes. Thus confinement effect can be considered negligible and there is a negligible impediment to the flow of hydrated protons due to this phenomenon which is prominent in PFSA (Berrod, Hanot, et al., 2017). The most notable observation here is that there is non-zero diffusivity of $D_{H_3O^+}$ in the NFSPI even at minimal hydration conditions. With increasing levels of hydration, the dramatic rise in the diffusivities is ascribed to the presence of larger water clusters and lesser bonding interaction between the Sulfonate groups and the hydronium ion as revealed by RDF plots in Fig. **??**. Moreover, the diffusion of hydronium ions is impeded by the presence of a strong hydrogen bonding network around sulfonate groups in PFSAs whereas there seems a lack of stable hydrogen bond network in the first solvation shell in NFSPI thus the movement of hydronium ions is unimpeded.

3.4.2.3 Proton conductivity

Figure 3.6 shows the proton conductivity derived from the diffusivities of the hydronium ions ($D_{H_3O^+}$) at different hydration levels in the three ionomers FSPI, NFSPI, and Nafion according to the Eq.3.3. At the lowest hydration level, Nafion is out of sufficient water molecules and hence does not conduct protons at all. FSPI also does not provide many levels of proton transport despite having double the number of sulfonate groups and water molecules than Nafion. On the other hand, the non-fluorinated NFSPI has non-zero proton conductivity at the lowest hydration level. This supports the argument that hydrocarbon-based PEMs are able to hold sufficient water molecules at lower humidity conditions thus preventing total dry condition which develops in fluorinated PEMs at higher temperatures. This also makes these PEMs well-suited for automobile applications. After the threshold value of percolation level is achieved at about λ = 5 for Nafion, there is a drastic increase in the proton conductivity of NFSPI whereas there is only a moderate increase in the conductivity of Nafion. FSPI also shows a drastic increase but proton transport in NFSPI surpasses the values of FSPI and Nafion. This is due to the ability of NFSPI to hold more bound water along the chain and not only around the anionic groups. This has been supported by the volume analysis studies also. At moderate level, hydration (λ = 10), the value proton conductivity of NFSPI lies in the range of 0.1-0.15 S·cm^{-1} which is an acceptable operating range for a PEM. It is to be noted that these values of proton conductivity only rely on the contribution of hydronium ions and not the movement of the proton through the Grotthus mechanism. Corresponding values for FSPI and Nafion are around 0.07 and 0.05 S·cm^{-1} respectively which are moderate values considering the contribution of hydronium ions only. Table 3.5 provides the values of proton conductivity for the modeled systems at different hydration levels.

Such a high rate of proton transfer in NFSPI is attributable to the high water volume fraction in the polymer. The movement of hydronium is most favored in the freezable bulk water. SPIs generally provide broad water pathways as close packing of polymer chains is prevented by the rigid nature of the binaphthalenic moieties. At a high level of hydration (λ = 15), the proton conductivity of NFSPI reaches the value of 0.18 S.cm^{-1} which is quite high. The vehicular mechanism of proton transport is favored

Level of hydration (λ)	Nafion PFSA	FSPI	NFSPI
1	0.00 (4.68%)	0.01 (7.99%)	0.03 (10.48%)
5	0.03 (23.44%)	0.02 (39.93%)	0.06 (104.80%)
10	0.05 (46.88%)	0.07 (79.85%)	0.14 (104.80 %)
15	0.07 (70.33%)	0.12 (119.78%)	0.18 (157.19%)

Values in parentheses represent Water uptake (%)
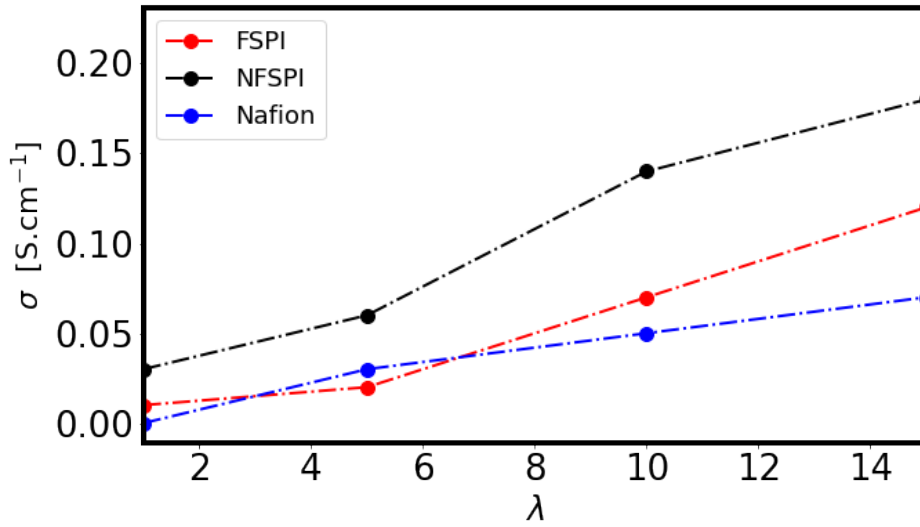
FIGURE 3.6: Proton conductivity calculated for Nafion PFSA, FSPI, and NFSPI from diffusion coefficient of hydronium ions with increasing hydration level

by increasing hydration levels.

3.4.3 Local structure of water molecules and hydronium ions

3.4.3.1 RDF plot of $O_{H_3O^+}$-O_{H_2O}

Fig. 3.7 (a) to (c) shows the RDF plots between the oxygen atoms of hydronium ions ($O_{H_3O^+}$) and the oxygen atoms of water molecules (O_{H_2O}) for Nafion, FSPI, and NFSPI with increasing levels of hydration. As observed from Fig.3.7(a) for Nafion, at a lower hydration level ($\lambda = 1$), the first solvation shell appears at a distance of about 5 Å as a broad peak of suppressed intensity. Notably, hydrogen bonds do not exist at such

large distances so these water molecules cannot be considered strongly bound to the hydronium ions. At higher hydration levels ($\lambda = 5$, 10, and 15), there is an intense peak at a distance of 3 Å and a slight second peak at about 6-7 Å. This information provides an idea about the immediate environment of hydronium ions in the water clusters or pores. With increasing water content, the number of water molecules around hydronium ions seems to be almost constant. In the case of novel SPIs, almost similar RDF plots were obtained at all levels of hydration. In both FSPI and NFSPI, peaks of varying intensities appear at a distance of 3 Å suggesting that hydronium ions are surrounded by a solvation shell of water molecules at this distance with a number of water molecules varying with water content (Fig. (3.7 (b) and (c)). Notably, in both HC PEMs, FSPI, and NFSPI, even at extremely low hydration ($\lambda = 1$) conditions, there seem to be existing hydrogen-bonded overlapping shells of water molecules around hydronium ions.

3.4.4 Local structure of sulfonate group

Generally, the immediate environment of the hydrophilic sulfonate groups in an ionomer comprises shells of water molecules and hydronium ions arranged in the form of overlapping shells bound by hydrogen bonds and electrostatic forces. RDF plots obtained from MD simulations provide information regarding these shells, particularly the first and the second solvation shells. Following RDF plots were obtained in the present study to obtain information on the local structure of the sulfonate groups:

- RDF plots between sulfur atoms of the sulfonate groups (S) and sulfur atoms of other sulfonate groups (S) $[S - S]$

- RDF plots between sulfur atoms of the sulfonate groups (S) and oxygen atoms of water molecules (O_{H_2O}) [S- O_{H_2O}]

- RDF plots between sulfur atoms of the sulfonate groups (S) and oxygen atoms of hydronium ions ($O_{H_3O^+}$) $[S - O_{H_3O^+}]$

(a)

(b)

(c)

FIGURE 3.7: Radial distribution plots of sulfur atom of the oxygen atom of the hydronium ions (O_h) and the oxygen atom of the water molecules (O_w) for (a) Nafion, (b) FSPI, and (c) NFSPI

3.4.4.1   RDF plot of S-O$_{H_3O^+}$

Fig. 3.8 (a) to (c) presents the RDF plots between sulfur atoms of the sulfonate groups and oxygen atoms of hydronium ions (S- $O_{H_3O^+}$) for Nafion, FSPI and NFSPI with increasing levels of hydration. In Fig. 3.8(a), it can be seen that the nearest hydronium ions are located at a distance of about 5 Å from the sulfur atom of the sulfonate group

at the lowest hydration level ($\lambda = 1$). This suggests that only weak or negligible electrostatic interaction with the sulfonate group can exist given the distance. As the level of hydration increases, hydronium ions move to the closer vicinity of the sulfonate groups and become strongly bound to sulfonate groups through electrostatic interaction as suggested by the broad peak at the distance of about 4 Å for hydration levels λ =5, 10 and 15. This is consistent with the cutoff distance of 4.3 Å for the first solvation shell around the sulfonate group for hydronium ions (Devanathan, Venkatnathan, and Dupuis, 2007b; Cui, Liu, et al., 2007; Kwon, Kang, et al., 2019; Zhang, Yang, et al., 2021). In the case of FSPI, a broad first peak of suppressed intensity appears at a distance of about 4 Å along with a flattened second peak extending to about 7 Å for the low level of hydration ($\lambda = 1$) (Fig. 3.8(b)). At higher levels of hydration ($\lambda = 5, 10$, and 15), the first coordination shell of hydronium ions strongly bound to sulfonate groups appears at a distance of about 4 Å, and a second peak indicating weakly bound hydronium ions appears at a distance of 7-8 Å. Similarly, in NFSPI, the first and second coordination shells appear at distances similar to FSPI, the only difference is observed in the intensity of the second peak being more pronounced in the case of NFSPI (Fig. 3.8(c)).

3.4.4.2   RDF plot pf S-O$_{H_2O}$

Fig. 3.9 (a) to (c) presents the RDF plots between sulfur atoms of the sulfonate groups and oxygen atoms of water molecules (S- O$_{H_2O}$) for Nafion, FSPI, and NFSPI with increasing levels of hydration. As seen in Fig.3.7(a), the first solvation shell of water molecules appears around sulfonate groups at a distance of about 5 Å at a low level of hydration ($\lambda = 1$). Clearly, these are weakly bound water molecules and no hydrogen bonding can be considered to exist, and no possibility of clustering of water molecules. As the water content in the ionomer increases, more water molecules move nearer to the sulfonate groups, and the intensity of the peak increases as more water molecules accumulate closer to the sulfonate group at a distance of about 4 Å. As in the case of hydronium ions, this finding is also consistent with the cutoff distance of 4.3 Å for the first solvation shell around the sulfonate group for water molecules (Devanathan, Venkatnathan, and Dupuis, 2007b; Cui, Liu, et al., 2007; Zhang, Yang, et al., 2021). The appearance of a slight second peak indicates the presence of a second solvation shell of weakly bound freezable water. Noticeably, in the novel SPIs, even at a lower level

(a)

(b)

(c)
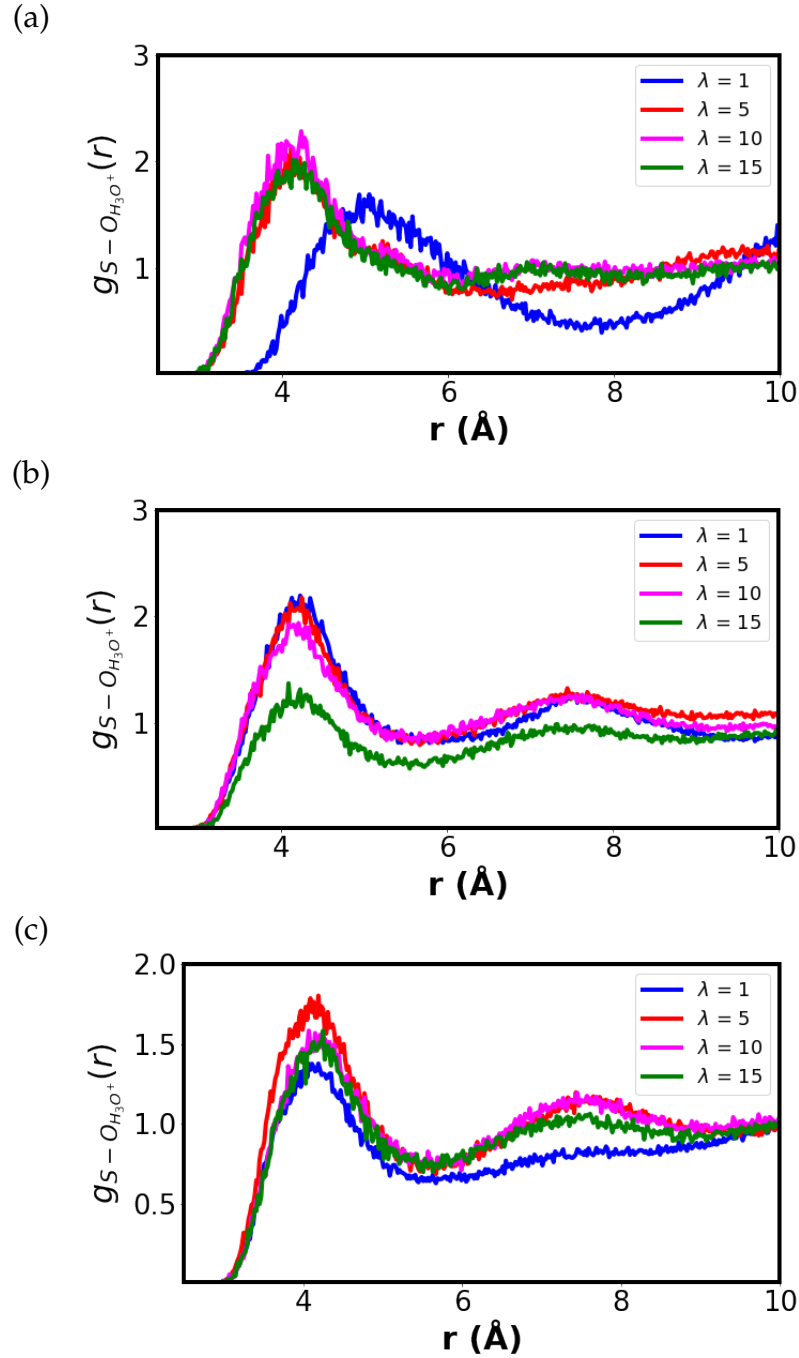
FIGURE 3.8: Radial distribution plots of sulfur atom of the sulfonate group (S) and the oxygen atom of the hydronium ions ($O_{H_3O^+}$) for (a) Nafion, (b) FSPI, and (c) NFSPI

of hydration ($\lambda = 1$), there is a first solvation shell comprising strongly bound water molecules at a distance of about 4 Å from the sulfur atoms of the sulfonate groups (Fig. 3.8(b) and (c)). There is also a second solvation shell which is of lesser intensity in FSPI compared to NFSPI. This prominent second peak in NFSPI apparently indicates a weakly bound shell of water molecules which acts as a buffer zone between strongly

bound water molecules around the sulfonate groups and the bulk water molecules in the center of the water cluster or pore.

(a)

(b)

(c)
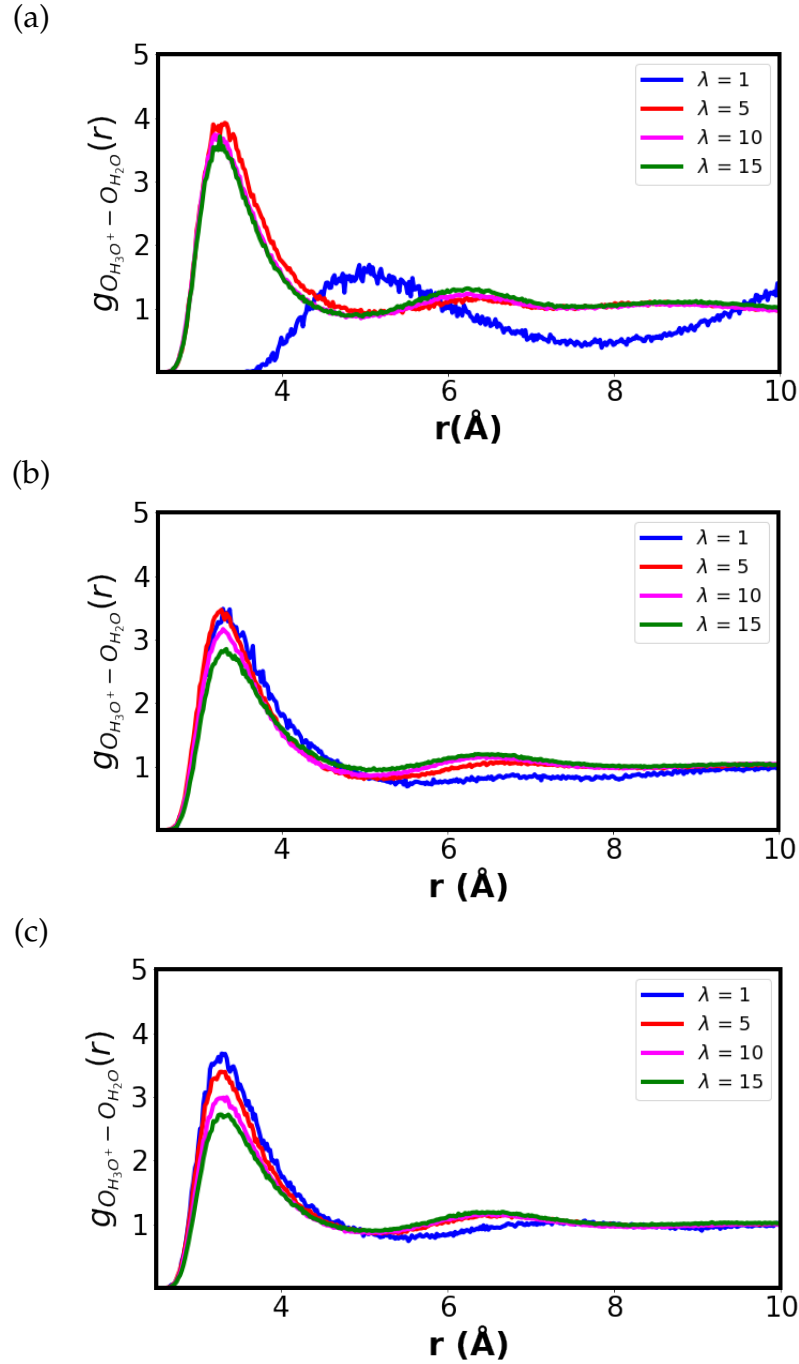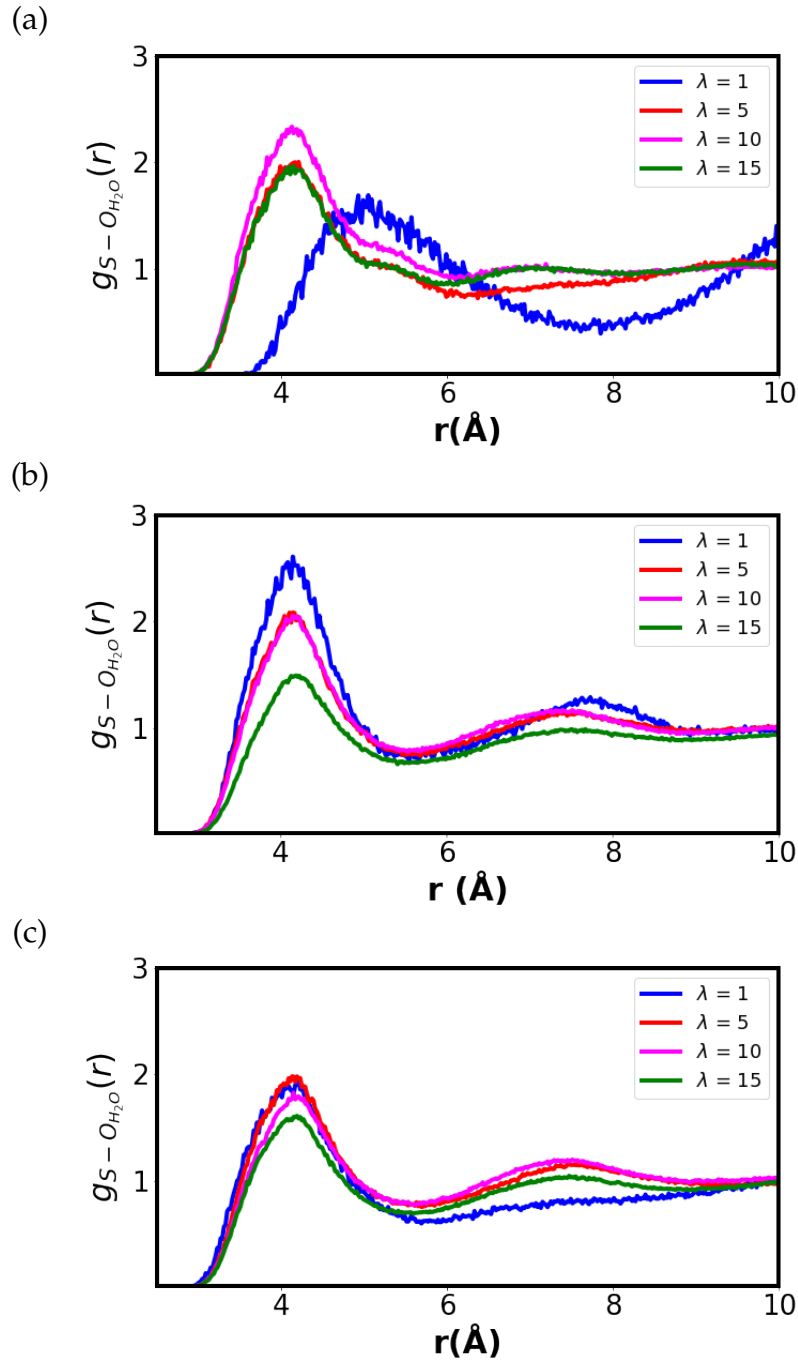

FIGURE 3.9: Radial distribution plots of sulfur atom of the sulfonate group (S) and the oxygen atom of the water molecules (O_{H_2O}) for (a) Nafion, (b) FSPI, and (c) NFSPI

Here it is relevant to discuss the significant role of second peaks in the RDF plots between sulfur atoms of sulfonate groups and oxygen atoms of hydronium, shown in Fig. 3.9 and Fig. 3.8. The second peak represents an interfacial region wherein water

molecules are bound to the sulfonate groups through hydrogen bonds and hydronium ions are bound to sulfonate groups through electrostatic interaction. This interfacial zone facilitates the transition from bound water molecules and hydronium ions near the sulfonate groups to freely moveable water molecules and hydronium ions in the center of the water cluster pore filled with bulk water.

3.4.4.3   RDF plot pf S-S

Fig. 3.10(a) to (c) presents the RDF plots between inter-molecular and intra-molecular sulfur atoms of the sulfonate groups (S-S) for Nafion, FSPI, and NFSPI with increasing levels of hydration. This RDF plot provides significant information related to the arrangement of the hydrophilic sulfonate groups along the walls of the water clusters and channels. As shown in Fig. 3.10 (a), there appears a single extended peak at about 5 Å suggesting that sulfonate groups are located close to each other. This is in agreement with the pair correlation for sulfonate groups in Nafion ionomer at 300 K reported in literature (Cui, Liu, et al., 2007). It is an expected trend considering the sulfonate groups are located at the end of the flexible side chain and nanophase segregation ensures close placement of sulfonate groups. While such close proximity of hydrophilic groups is significant for the relay transport of hydronium ions, it can also be considered to be an impediment due to the electrostatic interaction between hydronium ions and sulfonate groups. The second peak is completely missing in Nafion PFSa ionomer while it is quite prominent in SPI ionomers (Fig. 3.10 (b) and (c)).

It should be noted that the number of sulfonate groups in FSPI and NFSPI is double that of Nafion. Fig. 3.10 (b) and (c) shows the RDF plots between intermolecular and intra-molecular sulfur atoms of the sulfonate groups ($S - S$) for FSPI and NFSPI. At lower hydration levels, ($\lambda = 1$ and 5), almost three coordination shells are clearly visible with the first peak appearing at about 5 Å. A well-ordered arrangement of sulfonate groups suggested by a clearly distinct second peak at about 7 Å appears only at higher hydration levels ($\lambda = 10$ and 15). In NFSPI, the first peak occurs at about 5 Å while the second peak appears at about 7 Å. Apparently, it turns out that a symmetrical long-range order of sulfonate groups is conducive for proton transport considering the fact that NFSPI displays the highest proton conductivity and water and hydronium diffusivities among the studied ionomer systems.

(a)

(b)

(c)
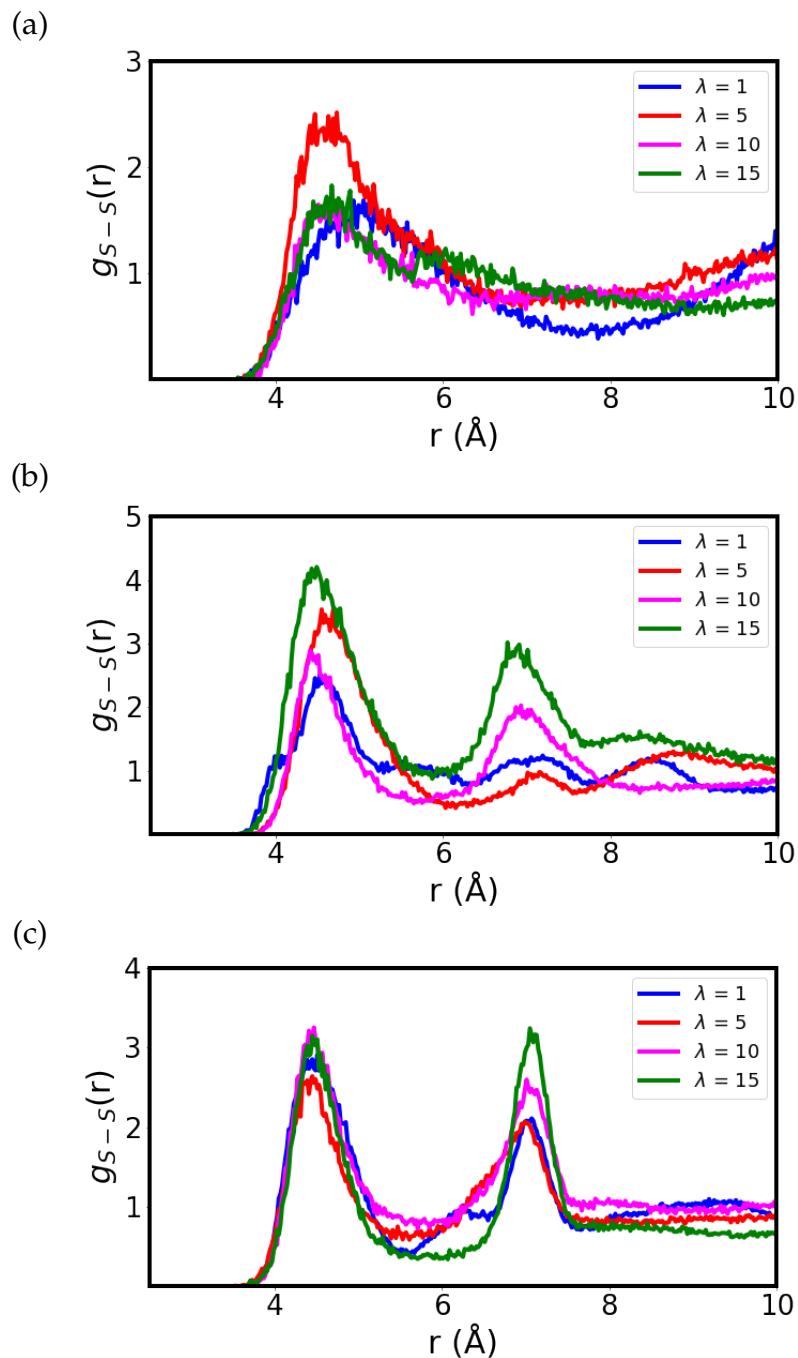
FIGURE 3.10: Radial distribution plots of sulfur atoms of the sulfonate group (S) for (a) Nafion, (b) FSPI, and (c) NFSPI

3.4.4.4 RDF plot pf S-F

In the case of Nafion and FSPI, RDF plots of S-F were also obtained from the trajectories to investigate the distribution of the fluorinated (hydrophobic) group with respect to the hydrophilic sulfonate group (refer Fig. 3.11 and Fig. 3.12).
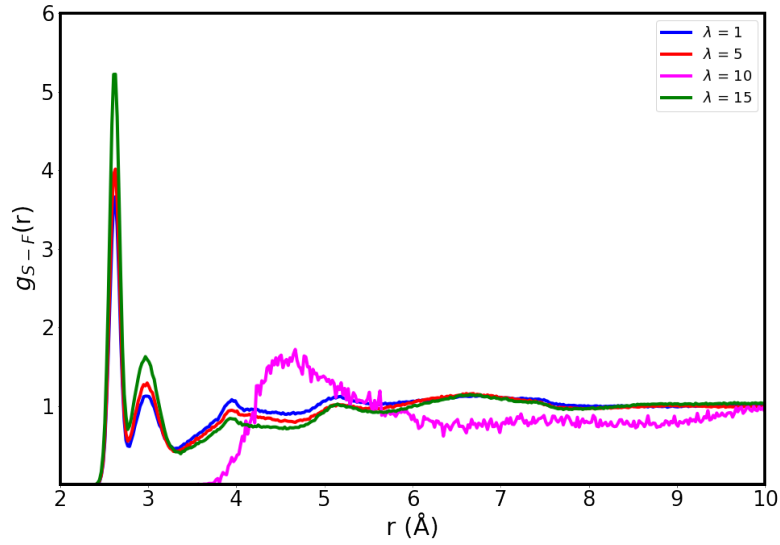
FIGURE 3.11: Radial Distribution Function (RDF) plot between the Sulfur atom of the sulfonate group (S) and the fluorine atoms for Nafion PFSA ionomer
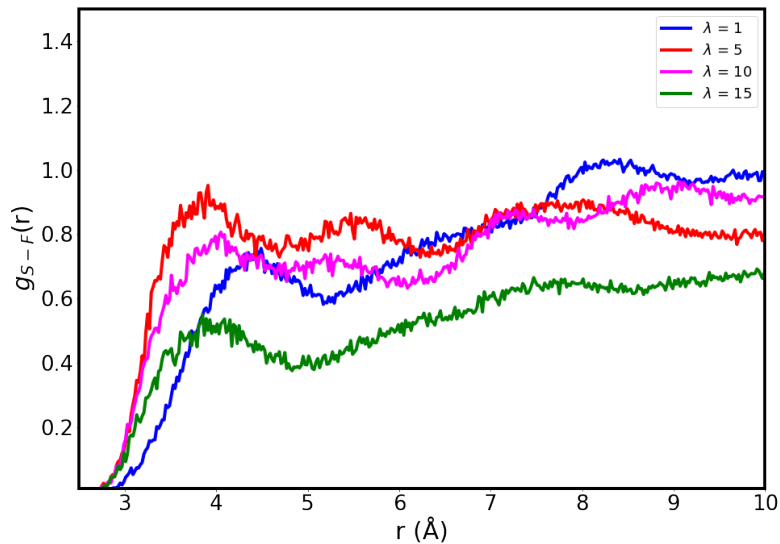
FIGURE 3.12: Radial Distribution Function (RDF) plot between the Sulfur atom of the sulfonate group (S) and the fluorine atoms for FSPI ionomer

3.4.5 Coordination Number analysis

Table 3.6 presents the coordination numbers and respective cutoff distances (given in parentheses) for Oxygen atoms of Hydronium ions around Sulfur atoms of Sulfonate groups. There are optimal numbers of hydronium ions and water molecules in the vicinity of sulfonate groups even at minimal hydration level represented by $\lambda = 1$. The hydronium ions are located at a distance of 4 Å distance while water molecules are

placed at 3.5 Å. There is an absence of such arrangement in Nafion at minimal hydrations whereas there are adequate numbers at higher levels of hydration, unlike the SPI ionomers. There is a minimal level of water molecules around the sulfonate groups of the SPI membranes at a distance of 3.5 Åas shown in Table 3.7. In the case of the solvation shell of water molecules around the hydronium ions, it can be seen in Table 3.8, that there are sufficient water molecules at even minimal hydration conditions.

TABLE 3.6: Coordination number and respective cutoff distances (given in parentheses) for Oxygen atoms of Hydronium ions around Sulfur atoms of Sulfonate groups

Level of hydration (λ)	Nafion PFSA	FSPI	NFSPI
1	0.85 (4.0 Å)	0.71 (3.5 Å)	0.60 (3.5 Å)
5	1.49 (3.5 Å)	0.77 (3.5 Å)	0.81 (3.5 Å)
10	2.06 (3.5 Å)	0.76 (3.5 Å)	0.61 (3.5 Å)
15	1.83 (3.5 Å)	0.44 (3.5 Å)	0.59 (3.5 Å)

TABLE 3.7: Coordination number and respective cutoff distances (given in parentheses) for Oxygen atoms of Water molecules around Sulfur atoms of Sulfonate groups

Level of hydration (λ)	Nafion PFSA	FSPI	NFSPI
1	0.85 (4.0 Å)	0.95 (3.5 Å)	1.12 (3.5 Å)
5	1.97 (3.0 Å)	0.88 (3.5 Å)	0.83 (3.5 Å)
10	1.99 (3.0 Å)	0.82 (3.5 Å)	0.74 (3.5 Å)
15	1.90 (3.0 Å)	0.50 (3.5 Å)	0.69 (3.5 Å)

TABLE 3.8: Coordination number and respective cutoff distances (given in parentheses) Oxygen atoms of Water molecules around Oxygen atoms of Hydronium ions

Level of hydration (λ)	Nafion PFSA	FSPI	NFSPI
1	0.85 (4.0 Å)	0.71 (3.0 Å)	0.84 (3.5 Å)
5	1.97 (3.0 Å)	0.85 (3.0 Å)	0.80 (3.5 Å)
10	1.99 (3.0 Å)	0.74 (3.0 Å)	0.68 (3.5 Å)
15	1.90 (3.0 Å)	0.60 (3.0 Å)	0.59 (3.5 Å)

3.4.6 Volume analysis

Generally, there is an increase in the volume of the simulation cell with an increasing content of water or an increasing number of water molecules. Absolute volumes do not provide complete information about the distribution of water molecules and

polymer matrix as different simulation cells have different volumes. Information on free volumes provides an idea of how the structural features of the chains influence the packing of the chains and also the extent of cavities and voids for the collection of water molecules. Volume analysis is a useful technique for understanding the basis of differences in the proton transport properties of ionomers which are direct results of nano-phase segregation (Kwon, Kang, et al., 2019). Polymer volume fractions, solvent (water molecules and hydronium ions) volume fractions, and fractional free volumes have been obtained by calculating van Der Waals volumes of the atoms using a probe of 0 Å radius. To obtain solvent volume fractions (and thus polymer volume fractions), volumes occupied by individual water molecules and hydronium ions were calculated and subtracted from the total volume according to the number of hydronium and water molecules present in the simulation cells. FFVs for the membranes have been calculated from total volumes and van der Waals volume using Eq.3.5.

Fig. 3.13(a) and (b) show the effect of increasing levels of hydration on the polymer and solvent (water molecules and hydronium ions) volume fraction in Nafion PFSA (Blue), FSPI (Red) and NFSPI (Black) ionomers. There are several features to observe in these systems. Firstly, Nafion and FSPI have similar volume fractions of the polymer and solvent phase although their chemical structures and absolute number of water molecules are different in this system (the level of hydration is the same but the number of water molecules is double in FSPI compared to Nafion due to a double number of sulfonate groups present). This similarity in polymer and solvent volume fraction is ascribable to the presence of fluorinated groups (-CF$_2$-) in Nafion and (-CF$_3$-) in FSPI as well as ether linkages (C-O-C) in side chain Nafion and in the main chain in FSPI. Secondly, the solvent volume fraction of NFSPI is the highest among the three modeled polymers even though the water content is similar to FSPI. Apparently, oxygen and nitrogen atoms in the imide groups (O=C—N—C=O) have electronegative nature and thus are capable of forming hydrogen bonds with water molecules. So, water molecules have more sites to form hydrogen bonding (however weaker) other than the hydrophilic sulfonate groups. Moreover, in Nafion and FSPI ionomers, the presence of fluorine (hydrophobic) atoms prevents the bonding of water molecules at sites other than sulfonate groups.

Fig. 3.13(c) shows the effect of an increased level of hydration on the fractional

free volume (FFV) of Nafion PFSA(Blue), FSPI(Red), and NFSPI(Black) ionomers. FFV is the volume that is not occupied by the polymer matrix and solvent phase (water molecules and hydronium ions). It can be seen that Nafion PFSA ionomer has consistent and low FFV whereas FFVs of FSPI and NFSPI steadily increase with the increasing water content values even though these are already higher than Nafion. Moreover, FFV seems to be inversely related to the equivalent weight (EW) of the ionomer (Table 3.2). The EW of Nafion is higher and thus it is observed to be occupying a higher volume in the simulation cell. On the other hand, FSPI and NFSPI have lesser EWs hence they are seen to be occupying lesser space in their respective cells. Moreover, NFSPI shows the highest FFV and solvent volume fraction which is probably due to the presence of Nitrogen and oxygen atoms and sulfonate groups which are able to bind water molecules and thus leave empty spaces. Also, the MDP non-sulfonated dianhydride has an angular ethylene component which breaks the linearity of chains and prevents close packing of the adjacent chains. Thirdly, the rigid main chain of SPIs allows parallel packing of chains and there are large empty pockets along the chains which allow the formation of elongated water channels. Thus, it can be considered that a higher solvent volume fraction of NFSPI facilitates better proton dynamics and proton conductivity. As seen in earlier sections, it was observed that proton transport is taking place in FSPI and NFSPI even at the lowest hydration levels, it can be considered that even at lower water fractions, the water molecules are rather freely dispersed than confined in a closed space as in Nafion. Free dispersion of the water molecules allows the existence of loosely overlapping networks through which hydronium can diffuse. At moderate hydration level (λ =5), the polymer volume fraction of Nafion ionomer is about 1.25% lesser than FSPI but 5.3% more than NFSPI whereas the water volume fraction of FSPi is the same as Nafion ionomer and that of NFSPI is 16% more than Nafion. Compared with FSPI, NFSPI has the presence of more water phases despite the same number of water molecules.

3.4.7 Visual analysis

Though visual analysis of a simulated model sample of the order of 2nm is insufficient to gather information on the distribution of water domains, overall water channel topology, and transport of protons within the channels. Nevertheless, a general

(a)

(b)

(c)
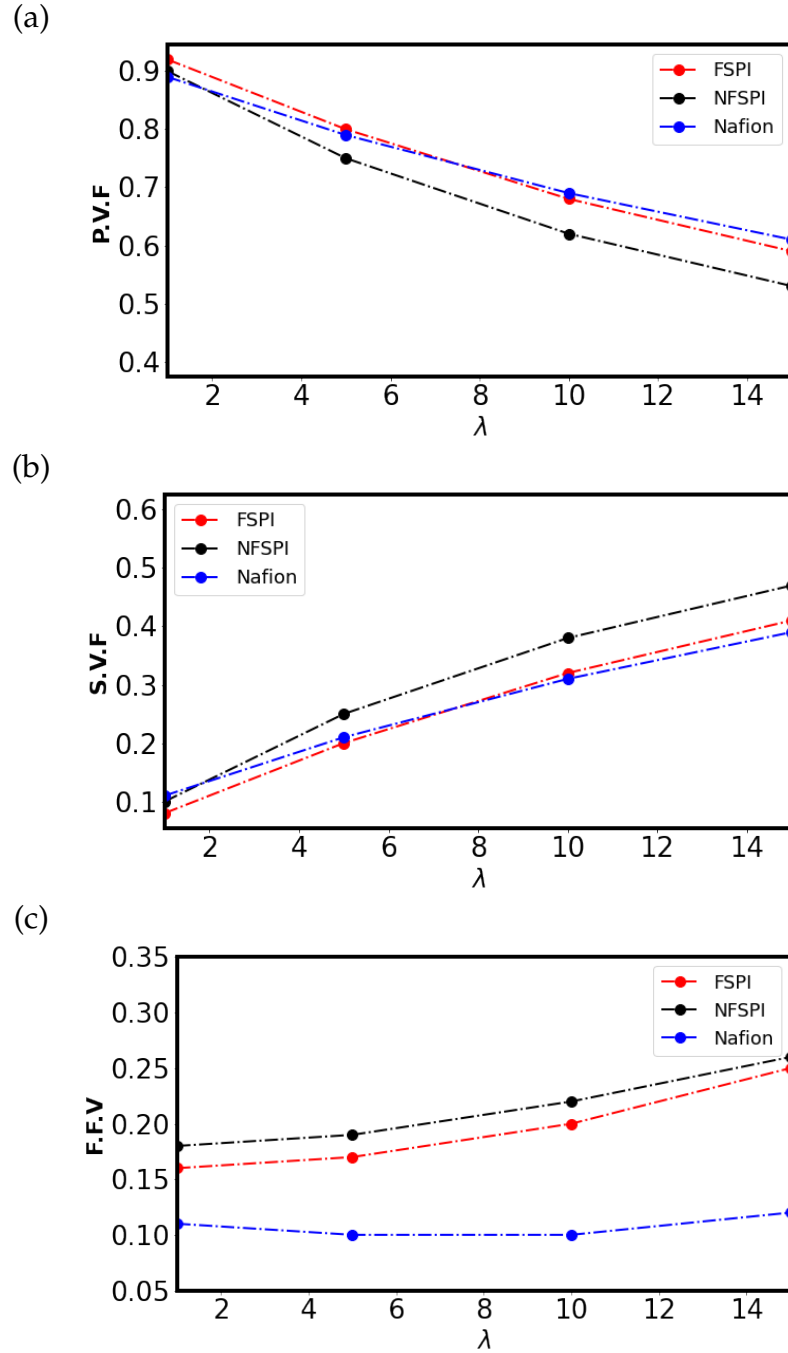
FIGURE 3.13: Variations in (a) Polymer volume fractions (b) Water volume fractions (c) Fractional free volumes observed with increasing levels of hydration for Nafion(Blue), FSPI(Red), and NFSPI(Black)

estimate of the form, structure, and distribution of ionic domains can be obtained from such oligomer-level all-atomistic MD studies.

Fig. 3.14(a) to (d) show the change in morphology of the water channels and rearrangement of polymer matrix upon increasing hydration level values from 1, 5, 10, and 15 in Nafion PFSA ionomer. Here Fluorine atoms are represented by sky blue,

Oxygen atoms are red, Hydrogen atoms are white, Sulfur atoms are yellow and Carbon atoms are gray. Clearly, there are significant changes in the morphology with increasing levels of hydration as the number of water molecules increase and gets accommodated initially in the empty pockets or free volume and gradually there is an elastic rearrangement of polymer chains for accommodating excess solvent phase water molecules.

Fig. 3.14(a) shows the state of the polymer at the lowest hydration level ($\lambda = 1$). At such low hydration levels, the sulfonate groups are inaccessible to water and hydronium due to hydrophobic Fluorine atoms in the polymer backbone. Hydrophobic perfluorinated backbone shields the hydrophilic sulfonate groups from whatever water molecules are present. Water domains and connecting channels are not well-formed due to an inadequate number of water molecules. At moderate water level ($\lambda = 5$), narrow channels can be observed to be beginning to form and an interconnected network is evident/visible [Fig. 3.14(b)]. Clearly, the percolation threshold is reached and transport of the proton begins to occur through vehicular and surface diffusion mechanisms. Fig. 3.14(c) shows the hydration level which represents the normal operating conditions ($\lambda = 10$), small water clusters begin to be apparent and the proton transport tends to be more facile. Fig. 3.14(d) shows that at high hydration level ($\lambda = 15$), sizeable water domains exist which are interconnected with other water domains through narrow channels. Such morphology is consistent with the ionomeric cluster and channel model proposed by Gierke et. al. for hydrated PFSA ionomers (Gierke, Munn, and Wilson, 1981).

Fig. 3.14(e) to (h) shows the snapshots collected from the simulation of the hydrated FSPI with increasing levels of hydration values $\lambda = 1, 5, 10$, and 15. Here Fluorine atoms are represented by sky blue, Oxygen atoms are red, Hydrogen atoms are white, Sulfur atoms are yellow, and Carbon atoms are gray. At a low hydration level ($\lambda = 1$), most of the volume of the simulation cell is seen to be occupied by the polymer matrix and there are only merely visible water molecules and hydronium ions. As the water content increases to $\lambda = 5$ [Fig. 3.14(f))], small, irregular clusters of the solvent phase (water) begin to appear in flattened free-volume spaces. Well-formed, narrow, and connected water channels appear at $\lambda = 10$ which connect a few larger water clusters
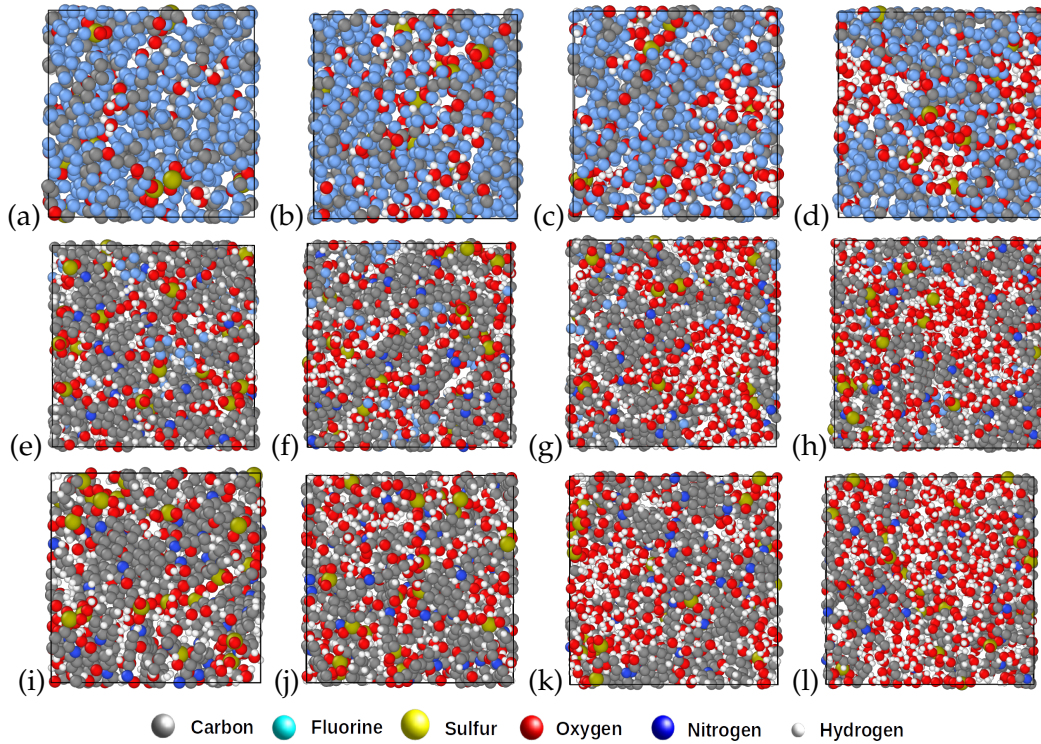
FIGURE 3.14: Snapshots of Nafion PFSA after equilibration (a) $\lambda = 1, (b)\lambda = 5, (c)\lambda = 10, (d)\lambda = 15; FSPI ionomer (e)\lambda = 1, (f)\lambda = 5, (g)\lambda = 10, (h)\lambda = 15; NFSPI ionomer (i)\lambda = 1, (j)\lambda = 5, (k)\lambda = 10, (l)\lambda = 15;$

(Fig. 3.14(g)). These water clusters appear to be comparatively larger in size than those formed in the Nafion PFSA ionomer but are not exactly spherical as in Nafion [Fig. 3.14(c)]. As shown in Fig. 3.14(h), at a high hydration level, the solvent phase increases and takes a flattened shape.

Fig. 3.14(i) to (l) show the snapshots from the simulations of hydrated NFSPI as the water content increases λ = 1, 5, 10, and 15. The polymer matrix is composed of NTDA/DSDSA/MDP backbone and sulfonate groups are present on the backbone itself. At a minimal hydration level (λ = 1), some pathways of water molecules are visible [Fig. 3.14(i)]. The molecular design prevents close packing of the polymer chains hence several empty pockets remain present which can be filled with water molecules. At λ = 5, numerous scattered flat water domains exist that provide a continuous pathway[Fig. 3.14(j)]. At a working hydration level of λ = 10, large nearly spherical pockets filled with water and hydronium appear lined by sulfonate groups [Fig. 3.14(k)]. At high hydration levels (λ = 5), irregular water domains or clusters are visible [Fig. 3.14(l)]. These are interconnected through narrow water channels. The irregular shape

of hydrophilic domains is probably due to the presence of N-atoms along the main chain which is also responsible for holding water molecules through hydrogen bonding.

3.5 Conclusion

To sum up, MD is an informative tool for exploring the structure-property interplay in ionomers and this study provides encouraging results for the novel SPI ionomers. Our atomistic simulations have revealed interesting insights into several aspects through which molecular design affects the structural and dynamic properties of an ionomer. The study involved molecular dynamics simulations of Nafion PFSA ionomer and two novel SPI membrane materials namely - NTDA/DSDSA/HFBAPP (partially fluorinated SPI) and NTDA/DSDSA/MDP (non-fluorinated SPI)) at different levels of hydration (λ = 1, 5, 10 and 15). Diffusion coefficients of water molecules and hydronium ions were found to be drastically affected by the level of hydration in all three ionomers. Maximum differences among the dynamic properties of these ionomers appeared at higher levels of hydration when the water clusters and channels became well-developed. Partially fluorinated SPIs were found to have better proton conductivity than Nafion PFSA but fully hydrocarbon (non-fluorinated SPI) was found to be still better than the partially-fluorinated SPI.

RDF plots, analyses of polymer, solvent phase, and free volume, and visual analyses through snapshots of equilibrated models were also performed to gather information related to the structure of water clusters and channels. Particularly, RDF plots of Sulfur atoms revealed that a symmetrical arrangement can be correlated with higher proton conductivity as in NFPSI. Unlike Nafion, irregular-shaped, flattened, and more dispersed water clusters were found to be present in the novel SPIs. Spherical water clusters connected through narrow channels were clearly observed in Nafion.

4 Experimental study

4.1 Chapter summary

Non-fluorinated SPI PEM (NTDA/DSDSA/MDP) with 50% degree of sulfonation was developed through condensation polymerization and consequent solvent casting method. Polymerization was confirmed by confirming the chemical structure of the synthesized polymer using Fourier Transform infrared (FTIR) and Nuclear Magnetic Resonance (NMR) spectroscopies. Physico-chemical properties namely Ion Exchange Capacities (IEC), water uptake, and swelling ratio along length and thickness were measured and it was found that there was good agreement with the water uptake values modeled in the MD simulations according to the hydration numbers. In-plane proton conductivity values were found using an in-house developed four-probe setup and Electrochemical Impedance Spectroscopy (EIS). The proton conductivity values found for the synthesized SPI membranes were in the range of 0.1588-0.28636 S·cm^{-1} which is exceptionally well for a PEM while those obtained in the MD simulations were 0.03-0.18 S·cm^{-1}. A good agreement was observed between the proton conductivity values predicted using MD simulations in which only the proton conductivity was determined using the hydronium ion diffusivities representing only one of the three operational mechanisms of proton diffusion. The SPI membranes were found to be stable in the harsh oxidative environment represented by the 30% Ferrous Sulphate solution in the Fenton reagent during oxidative testing as well as in the hydrolytic stability testing.

4.2   Introduction

4.2.1   Synthesis of SPI PEMs

One of the first reports of naphthalaneic co-SPI was presented by Genies et al (Genies, Mercier, et al., 2001). Though the rigid structure does not allow much room for flexibility and polymer chain rearrangement, they are found to have better hydrolytic stability (Essafi, Gebel, and Mercier, 2004a; Jang, Lee, et al., 2005). 4,4'-diamino stilbene-2,2'-sulfonic acid (DSDSA), a Type 1 sulfonated diamine was used in the synthesis of SPI along with bulky fluorenylidene group containing diamine. The bulky group created enough free volume to retain water even at high temperatures and offer good proton conductivities even at temperatures as high as 140 ºC (Ye, Bai, and Ho, 2006a). Several groups have reported the experimental synthesis of SPI-based ionomers Einsla, Kim, et al., 2005; Ganeshkumar, Bera, et al., 2014; Chen, Chen, et al., 2009; Chen, Yin, et al., 2006a; Chen, Chen, et al., 2010; Einsla, Hong, et al., 2004; Feng, Kondo, et al., 2018; Pandey and Shahi, 2015. The general route of Polyimide synthesis involves two-step- Polycondensation of dianhydrides ($-(CO)_2$-O) and diamines ($(-NH_2)_2$) followed by chemical or thermal imidization (Mittal, 2005). PI copolymers with controlled IEC are fabricated by careful consideration of comparative molar amounts of sulfonated diamines (bearing $-SO_3H$ group) and non-sulfonated diamines.

Polyimides are known for their rigid backbone which imparts thermal and mechanical stability to them but at the same time makes their processability difficult by increasing their glass transition temperature. The backbones of the polyimides also govern the symmetry, crystalline/amorphous behavior and the glass transition temperature (Akbarian-Feizi, Mehdipour-Ataei, and Yeganeh, 2012). Dianhydrides (DAs) make up a major fraction of the PI main chain. Usually, their rigid structure and interaction between basic imide N-atoms and acidic sulfonated group are responsible for limited solubility but flexible DAs make these interactions weaker and lead to improved solubility (Chen, Chen, et al., 2010). Benzoic acid (BA) is used as a catalyst in the polycondensation reaction, especially in the reactions of naphthalenic dianhydrides to enhance their reactivity towards diamines Sk, Pijet, and Wanic, 1992. BA aids in formation of trans-isoimide (Sek, Wanic, and Schab-Balcerzak, 1995). In addition, an organic base Triethylamine (TEA) is also added to these reactions. This base

deprotonates the sulfonated groups in the polymer chain thus an additional step of re-protonation of these acidic sites is performed. Acid treatment by immersing the films in an acid solution usually leads to good proton conductivity but can cause a decrease in the mechanical properties while protonation using ion exchange resins led to good proton conductivity (14.4×10^{-2} S.cm^{-1} at 80 °C) membrane with an IEC of 1.98 mEq g^{-1}) (Blázquez, Iruin, et al., 2005).

4.2.2 Properties of SPI PEMs

The properties of the PEMs are evaluated using various spectroscopic, gravimetric, and imaging methods (Kim and Lee, 2015b). Physical properties originating from the chemical composition are categorized as physicochemical properties. In ionomers, properties such as IEC, water uptake, swelling ratio, and ion conductivity are primary physicochemical properties that are directly related to the acidic content of the membranes.

4.2.2.1 Ion exchange capacity (IEC)

IEC, expressed in mEq.g^{-1}, is the ability of an ionomer to bind the counter-ion. It depends directly upon the content of the acidic group in the polymer and inversely on the non-ionomeric hydrophobic content in the polymer (i.e. the relative content of the rest of the polymer with respect to the acidic component). Proton conductivity is governed by IEC but is not solely dependent on it. It depends on polymer composition, morphology, the direction of measurement, temperature, and humidity. Effective IEC observed in the fully processed PEM is often different from the value calculated from the molar ratio of the ionic and neutral content of the polymer. This is due to fabrication processes and also due to the final resulting structure where the ionic groups could be less accessible. Effective IEC is calculated using the back titration method and NMR-H[1]. The back titration method is based on neutralizing the acidic content of the known weight of the PEM specimen with a base of known molarity and then determining the concentration of the residual base. This method includes two steps after the conversion of membranes from TEA form (SO_3-N$^+$H-$(CH_2CH_3)_3$) to proton form (-SO_3H).

4.2.2.2   Proton Conductivity

Proton conductivity is the most important property of the PEM. It depends on IEC, water uptake, proton mobility, the direction of measurement, level of degradation, and temperature and humidity. It is usually derived from an ex-situ electrochemical technique of Electrochemical Impedance Spectroscopy using the four-probe method. EIS or AC impedance Spectroscopy is a technique in which the sample is perturbed with a small AC signal of known frequency and its response is captured and plotted in the form of Nyquist plots. An equivalent circuit is derived to evaluate the behavior of the different components contributing to the impedance faced by H^+ in the membrane. These measurements are carried out in a specially designed setup using the four-probe method. This method ensures accuracy, especially in the lower frequency region. Usually, a rectangular strip of a specimen of known dimensions is immobilized in a cell such that it is in contact with two current-carrying electrodes and two voltage-measuring electrodes. Since proton conductivity is sensitive to temperature and humidity, provisions for accurate temperature and humidity control are made in this cell and proton conductivity is determined with respect to these parameters.

4.2.2.3   Water uptake

Uptake of water is a measure of the maximum capacity of the PEM to hold water. It depends on the content of acidic groups in the polymer sample and is evaluated using the gravimetric method after the conversion of the membrane sample to proton form. Membrane specimens are kept immersed in deionized water for a certain time (usually 6-24 hours). After careful wiping and vacuum drying, a percentage rise in their weight is calculated. Apart from acidic groups, the water uptake behavior of the electrolyte membrane is also dependent on temperature. At high temperatures ($>100^{\circ}C$), the mobility of polymer chains and water molecules gives rise to a different water uptake behavior of the membrane than at lower temperatures. Usually, the water uptake is seen to rise with the increase in temperature. However, this rise may be linear or nonlinear. Even for lower IEC values, WU increases with temperature, and a sharp increase is observed at $100^{\circ}C$.

4.2.2.4 Swelling

Swelling is the dimensional change that the PEM undergoes upon absorption of water. Usually, it is isotropic in the case of PFSA PEMs that are similar in planar as well as transverse direction. It is tested by measuring the dimensions before and after soaking PEM samples in water for a certain time period. The interaction of SPI with water is closely related to the swelling ratio of the membrane which is another important design consideration. SPIs have rigid polymer chains aligned in the in-plane direction. Upon absorbing water, swelling is observed more in the thickness direction than plane direction. Such anisotropic swelling is a characteristic feature of SPI-based PEMs but unfortunately, this behavior tends to have detrimental effects on the membrane's mechanical integrity in the fuel cell.

4.2.2.5 Oxidative stability

Lee et. al. observed that flexible crosslinkers improved OS but only at a high degree of crosslinking (Lee, Hwang, et al., 2006). Chemically, oxidative stability is governed by the basic tertiary nitrogen groups having high electron density. Similar products of degradation were observed in real FC effluents and the Fenton test. Usually, the durability of the PEM material is inversely proportional to the time when degradation is observed in this test but still other factors also control the durability of PEMs.

4.2.2.6 Hydrolytic stability

SPIs comprise imido rings in their main chains which make them susceptible to hydrolysis. The hydrolytic stability of the membranes is determined by electron density over imido rings. Lesser electron density amounts to more positive charge over imido carbonyl groups and thus lesser hydrolytic stability. Imido rings are formed from the dianhydride monomers. Thus, essentially dianhydride monomers decide the hydrolytic stability. NTDA has four carbonyl groups and thus distributed positive charge thus decreasing electron-withdrawing characteristics of the carbonyl group (Chen, Yin, et al., 2006b; Zhang, Li, et al., 2008; Zhang, Li, et al., 2010). The electron density over imido rings should be lesser and the carbonyl group should be more to prevent an attack on

the imido rings. Hydrolysis takes place in two stages initiating with the breaking of imide rings and eventually complete dissolution.

4.3 Methodology

4.3.1 Materials

1,4,5,8-Naphthalene tetracarboxylic anhydride (NTDA), 4,4-diamino stilbene-2,2-sulfonic acid (DSDSA), and 4,4-diamino diphenyl methane (MDP) were supplied by Tokyo Kasei (TCI), Japan. Triethyl amine (TEA) (>98%) was received from Spectrochem Chemicals, India, and purified using molecular sieves (4 Å). m-Cresol (LOBA Chemie, India), acetone (LOBA Chemie, India), methanol (FINAR Chemicals, India), and benzoic acid (SDFCL India) were used as received.
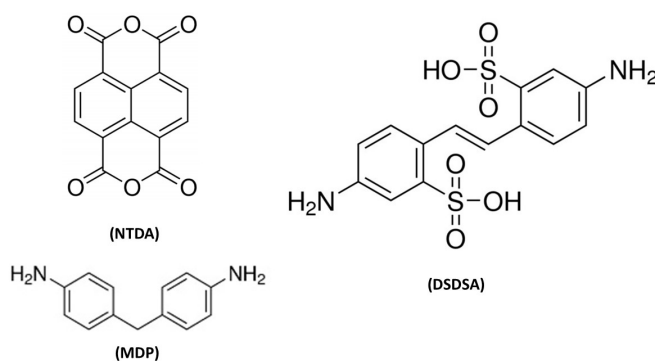
FIGURE 4.1: Chemical structures of the dianhydride, sulfonated and non-sulfonated diamines

4.3.2 Synthesis of NTDA/DSDSA/MDP SPI

First of all, a 250 mL round bottom flask, kept in a vessel containing silicon oil was fitted with a condenser over a magnetic stirrer connected with a thermocouple for accurate temperature measuring. Then, 0.345 g (0.5 mmol) of DSDSA, 20 mL of m-cresol, and 0.5 mL of triethylamine (TEA) were added sequentially to the thoroughly dried round bottom flask. After DSDSA was completely dissolved, and then 0.185 g (0.5 mmol) MDP and 0.5 g (1.0 mmol) NTDA were added. After 10 minutes, 0.16 g benzoic acid was added as a catalyst to start the polymerization. The mixture was stirred at room temperature for a couple of minutes and then heated at 80 °C for 4 h to obtain polyamic acid (PAA). The resulting PAA was further heated at 180 °C for 24 h. The

obtained highly viscous and thick solution was diluted with 10 mL m-cresol before cooling at room temperature. Then SPI solution was transferred into a beaker filled with acetone to obtain SPI precipitate. Next, the SPI precipitate was washed with acetone several times and dried at 80 °C overnight in a vacuum oven to produce pure and dry SPI (Chhabra and Choudhary, 2009; Dhra, Balasubramanian, and Kannan, 2021).

4.3.3 Membrane preparation and proton exchange reaction

An adequate amount of as-prepared SPI was dissolved in m-cresol to produce a homogeneous viscous solution. The SPI solution was spread on a flat glass container and spread uniformly using a film applicator of known thickness. Then the glass container was kept in a vacuum oven at 60 °C for 24 h to remove the solvent. The dried membrane was soaked in methanol for 2 h to remove the remaining m-cresol. The dried membrane was immersed in 1.0 M hydrochloric acid (HCl) at ambient temperature in order to implement a proton reaction exchange reaction for 48 h. The resulting proton-exchanged membrane was then washed thoroughly with double distilled water. Finally, the membrane was dried at 100 °C for 12 h in a vacuum oven (Chhabra and Choudhary, 2009; Dhra, Balasubramanian, and Kannan, 2021).

FIGURE 4.2: Picture of the synthesized SPI membrane

4.3.4   Structure characterization

Nuclear Magnetic Resonance (NMR) spectra (^{1}H) were recorded with JEOL JNMECS 400 MHz NMR Spectrometer using Deuterated DMSO D$_6$ as solvent. FT-IR (Fourier Transform Infrared) spectra were recorded on the IRAffinity-1S instrument. The morphology of the synthesized membranes was observed using Scanning Probe Microscopy (Multimode 8, Bruker, USA) in tapping mode.

4.3.5   Physicochemical properties

4.3.5.1   Ion-exchange capacity (IEC) (meq·g^{-1})

Theoretical IEC was calculated from molar fractions of the reactant feeds. IEC was also calculated using the back titration method.

$$IEC_w = \frac{V_{NaOH} \cdot N_{NaOH}}{m_{wp}} \tag{4.1}$$

where IEC$_w$ is the Ion exchange capacity by weight (meq·g^{-1}), V$_{NaOH}$ is the volume of NaOH consumed during titration (ml), N$_{NaOH}$ is the Normality of NaOH, and m$_{dp}$ is the mass of dry (g).

4.3.5.2   Water uptake (%)

Water uptake, Swelling ratio, and Dimensional stability were evaluated by immersing membrane samples of known weight and dimensions in deionized water for 12 hours at room temperature and using these

$$WU\% = \frac{m_{wp} - m_{dp}}{m_{dp}} \cdot 100 \tag{4.2}$$

where m$_{wp}$ is the mass of polymer equilibrated with water (g), m$_{dp}$ is the mass of polymer equilibrated with water (g).

4.3.5.3   Dimensional stability

Dimensional stability was determined along the thickness and longitudinal direction according to Eq.4.3 and Eq.4.4.

$$\Delta t = \frac{t_{wp} - t_{dp}}{t_{dp}} \tag{4.3}$$

where t_{wp} is the thickness of polymer equilibrated with water (mm) and t_{dp} is the thickness of dry polymer (mm).

$$\Delta l = \frac{l_{wp} - l_{dp}}{l_{dp}} \tag{4.4}$$

where l_{wp} is the length of polymer equilibrated with water (mm) and l_{dp} is the length of dry polymer (mm).

4.3.6 Hydrolytic and oxidative stability

Hydrolytic stability was assessed by immersing the membranes in deionized water at 80 °C. Oxidative stability was assessed by keeping the membranes immersed in Fenton's reagent (30% H_2O_2 solution with 30 ppm $FeSO_4$) at 90 °C. The time taken for the samples to completely dissolve or break down is monitored.

4.3.7 Proton conductivity

Proton conductivity was determined in the in-plane direction using an in-house developed four-probe apparatus. Using Electrochemical Impedance Spectroscopy (EIS) according to Eq. 4.5 as

$$\sigma_{||} = \frac{l}{R.A} \tag{4.5}$$

where $\sigma_{||}$ is the proton conductivity (S·cm^{-1}), l = distance between electrodes, R = is the Resistance observed in the direction parallel to the plane (Ohm), derived from Nyquist plot (), A = Area of the sample (cm²). Temperature dependence of proton conductivity follows an Arrhenius equation given by Eq.4.6 and the activation energy for proton conduction is also obtained from this equation:

$$\sigma = \sigma_o.exp\frac{E_a}{R.T} \tag{4.6}$$

where σ_o is the pre-exponential factor, Ea is the activation energy for proton conduction (kJ.mol^{-1}), R is the Universal Gas constant (= 8.314 J.mol^{-1}.K^{-1}), and T is the absolute temperature (K).

4.3.8   Single cell PEMFC testing

Membrane Electrode Assemblies (MEAs) were prepared with Nafion (NR212) and NF-SPI SPI ionomer PEMs using Toray Carbon paper as Gas Diffusion Layers(GDLs). Pt/C electrocatalyst loading of 0.25 mg.cm^{-2} on the Anode and 0.50 mg.cm^{-2} on the Cathode. The I-V characteristics or polarization behavior of the synthesized membrane was measured using a single cell of H$_2$/Air PEMFC having an effective area of 7.29 cm^2 (2.7 cm X 2.7 cm). The temperature of reactant gases maintained at 80°C was used. Humidified Hydrogen and Air were supplied to the cell with a 50% RH level at flow rates of 80 ml.min^{-1} H$_2$ and 100 ml.min^{-1} respectively. The pressure was maintained at 2.5 bar for both reactants.Fig. 4.3 shows the schematic of the single-cell PEMFC testing apparatus.
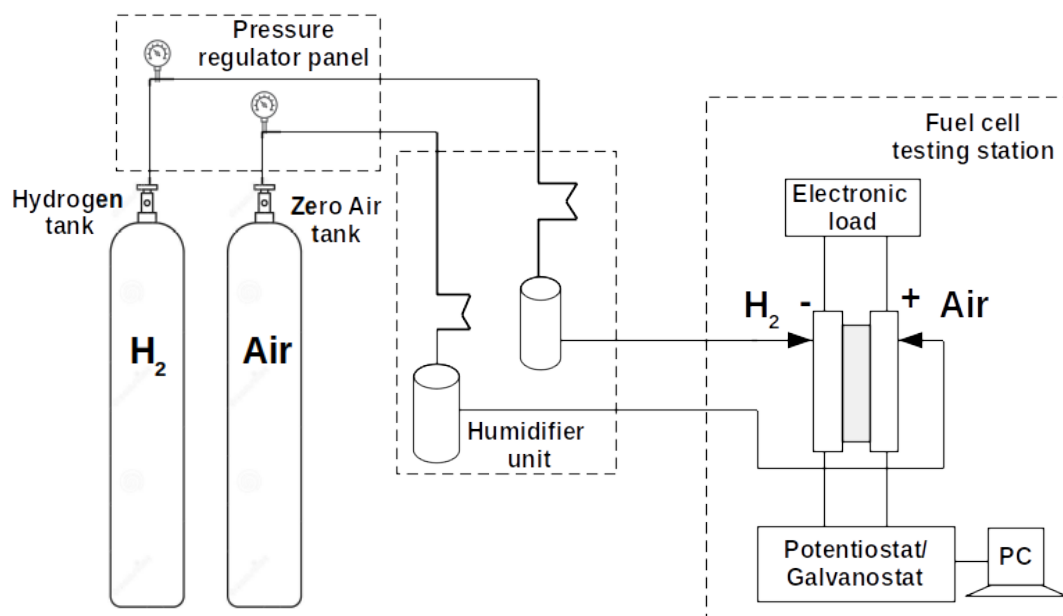
FIGURE 4.3: Schematic of the single cell PEMFC test apparatus

4.4 Results and discussion

4.4.1 Chemical structure characterization

4.4.1.1 FTIR

Fig. 4.4 shows two strong absorption bands corresponding to asymmetric and symmetric stretching vibrations are observed between 1178 cm^{-1} and 1085-1095 cm^{-1} respectively. The absorbed water shows a band around 3500 cm^{-1}. -C-N asymmetric stretching occurs at 1347 cm^{-1}, -C=O bond symmetric stretching appears as peak around 1664 cm^{-1}, and -C=O bond asymmetric stretching occurs at 1709 cm^{-1}.
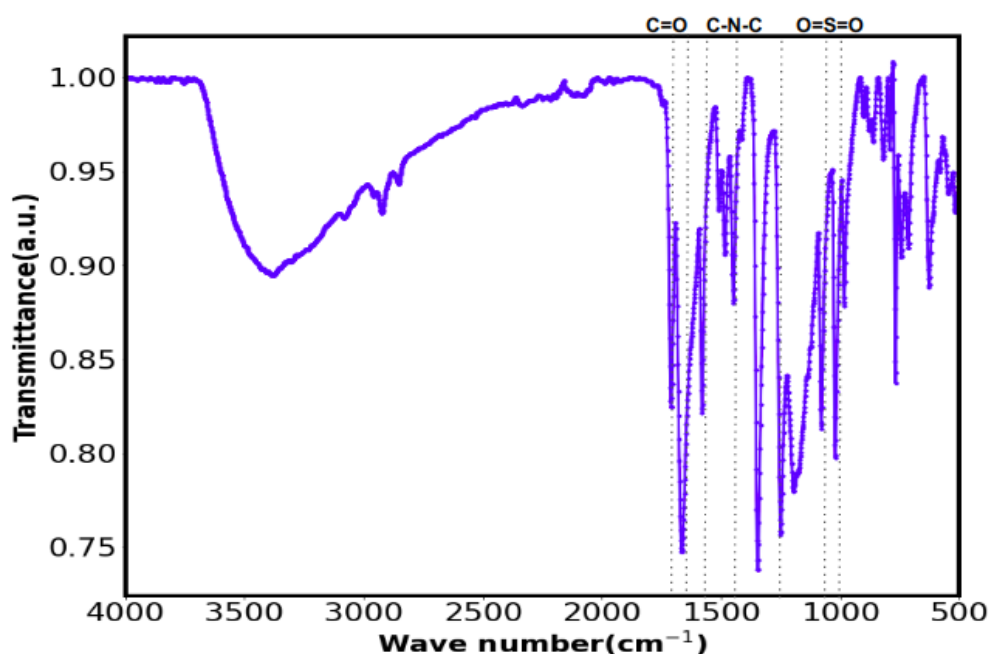
FIGURE 4.4: FTIR spectrum of the synthesized SPI ionomer

4.4.1.2 NMR

Fig. 4.5 shows the NMR shifts observed in the SPI samples. Naphthalene (Aromatic H-atoms (8.736)), Benzene (Aromatic H-atoms (6.963, 7.091), Sulfonate group [-S(=O)(=O)OH] (8.306)), Stilbene (Aromatic H-atoms (7.444, 7.501, 7.845, 7.913), Ethylene H-atoms (7.218)). No unreacted amine groups were found to be present as there is no shift around 5 ppm.
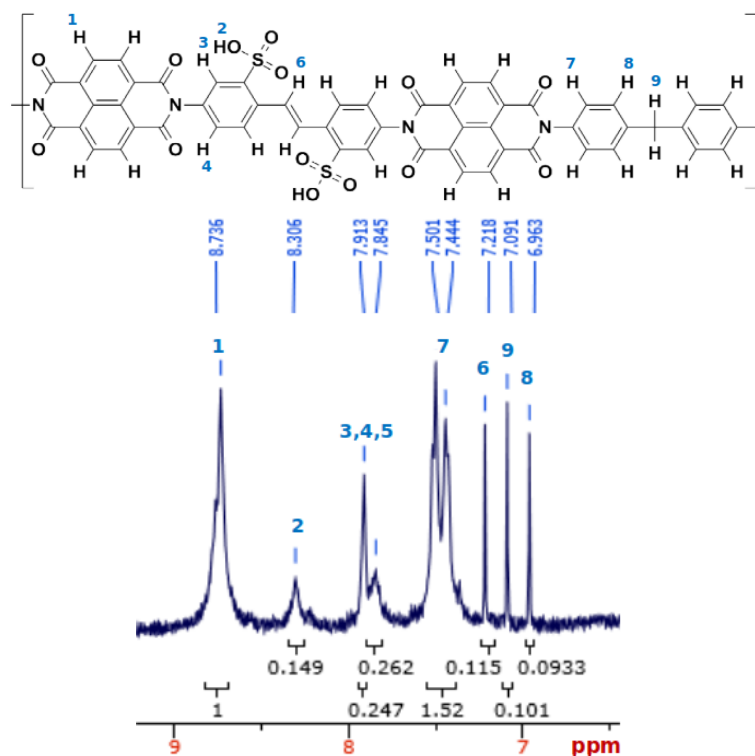
FIGURE 4.5: NMR spectrum of the synthesized SPI ionomer
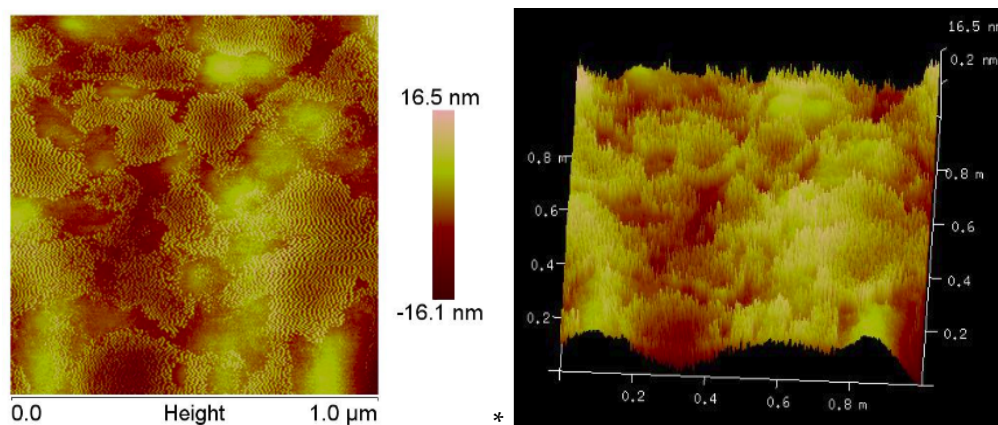
FIGURE 4.6: Scanning Probe Microscope image (Above) of the synthesized SPI ionomer, height images: z = 16 nm

4.4.2 Morphology

Fig. 4.6 shows the topology of the synthesized ionomer observed through Atomic Force Microscopy (AFM) or Scanning Probe Microscopy (SPM) on a 1-micron length scale. Seemingly well-defined morphology of parallel arrangement of polymer chains is visible in the dry ionomer samples. Fig. 4.7(a,c) shows notable clusters of water distributed within the solid phase polymeric matrix while the overall topography is visible in the 3D height images as seen in Fig 4.7(b,d).
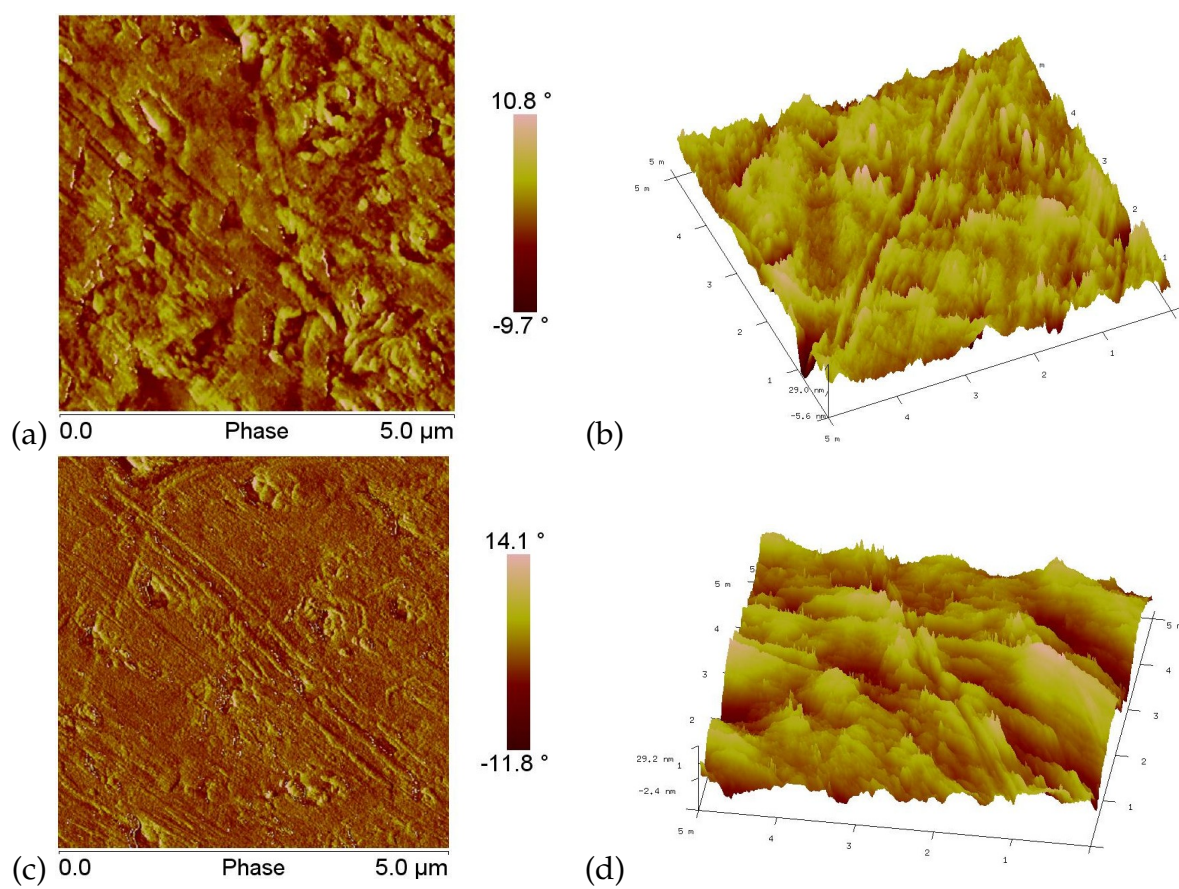
FIGURE 4.7: Scanning Probe Microscope image showing phase contrast
between water clusters and NFSPI ionomer segments

4.4.3 Physicochemical properties and Dimensional stability

Table 4.1 provides a comparison of the physicochemical properties namely IEC, water uptake, and hydration number along with proton conductivity for different SPI membranes reported in the literature as well as the SPI membrane synthesized in this study. Also, the dimensional stability in terms of swelling ratio along the length and thickness direction has been shown. There is only marginal change in the swelling ratio in both directions which suggests high dimensional stability of the synthesized SPI membrane. Also, even at comparatively lower IEC and moderate water uptake, high proton conductivity could be observed.

TABLE 4.1: Comparison of physicochemical properties of the synthesized SPI membrane with other PEMs

Ionomer	IEC[a] (meq.g^{-1})	WU[b] (%)	λ[c]	SRT[d] (%)	SRL[e] (%)	PC[f] (S·cm^{-1}) (In-plane)	Reference
NTDA/ ODADS/ MPDA	2.40	78	18	0.19	0.15	0.105 (20 °C) 0.184 (80 °C)	Li, Cui, et al., 2007
BTDA/ 2,2′-BSBB/ BAHF	1.80	18.8	37.5	0.13	0.01	0.175 (80 °C)	Li, Cui, et al., 2007
NTDA/ BAPBDS	2.43	107	22	0.26	0.22	0.2	Watari, Fang, et al., 2004
NTDA/ 3,3′-BSPB/ TAPB	2.49	114	NR	0.68	0.02	0.13 (50 °C)	Yin, Hayashi, et al., 2005a
[NTDA/ DSDSA] -DABA-QA] - DQDN	2.29	23.66	5.7	0.110	0.01	0.0649 (30 °C) 0.1364 (90 °C)	Mistri, Mohanty, and Banerjee, 2012
NTDA/ DSDSA/ BATh	1.98	27	7.2	15	2	0.047 (30 °C) 0.077 (90 °C)	Mistri, Banerjee, et al., 2015
NTDA/ DSDSA/ MDP	1.54	21.98	8	0.06	0.0048	0.1588 (30 °C)	This work

[a] IEC = Ion Exchange Capacity (meq·g^{-1}), [b] WU = Water Uptake (%), [c] λ = Hydration Number, [d] SRT = Swelling ratio (thickness direction) (%), [e] SRL = Swelling ratio (length direction) (%), [f] PC = Proton Conductivity (S·cm^{-1})

4.4.4 Hydrolytic and oxidative stability

Table 4.2 gives a comparison of hydrolytic and oxidative stability values for Nafion PFSA membranes and SPI membranes as reported in the literature and obtained for the synthesized SPI membrane. The synthesized SPI membranes show moderate stability in hydrolytic and oxidative environments.

TABLE 4.2: Comparison of Hydrolytic and Oxidative Stability of the synthesized SPI membrane with other PEMs

	Hydrolytic Stability	Oxidative Stability	Reference
Nafion 117	240 h		Sundar, Jang, et al., 2005
NTDA/ODADS/ODA	25 h	24 h	Fang, Guo, et al., 2002
BTDA/2,2'-BSBB/BAHF	160 h	52 h	Li, Cui, et al., 2007
NTDA/S-DHPZDA/ODA	38 h	51 h	Zhu, Pan, et al., 2008
NTDA/DSDSA/MDP SPI	51 h	42 h	This work

4.4.5 Proton conductivity

Proton conductivity is the most important property of the PEM. It depends on IEC, water uptake, proton mobility, the direction of measurement, temperature, humidity, and level of degradation. Proton conductivity was measured in the in-plane or longitudinal direction measured at for a temperature range of 30 °C - 80 °C using an in-house developed four-probe apparatus. As shown in Fig. 4.8, values in the range of 0.15 - 0.28 S·cm^{-1} were observed. The phenomenon of proton conduction incorporates two dominant mechanisms of proton diffusion namely proton hopping or structural diffusion and vehicular diffusion of hydronium ions. Fig. 4.8 shows the variation of the activation energy of proton transport with increasing temperature. At 30 °C, an activation energy of 9.9144 kJ.mol^{-1} was observed.

Fig. 4.9 shows the Nyquist plots for SPI ionomer at increasing temperatures. With an increase in temperature, there is a decrease in the resistance due to the lowering of the activation energy of proton transport. At higher temperatures(333K, 343K, and 353K), the polymer chains undergo substantial rearrangement, and better connectivity among the hydrophilic domains is achieved hence lesser resistance is observed. It can
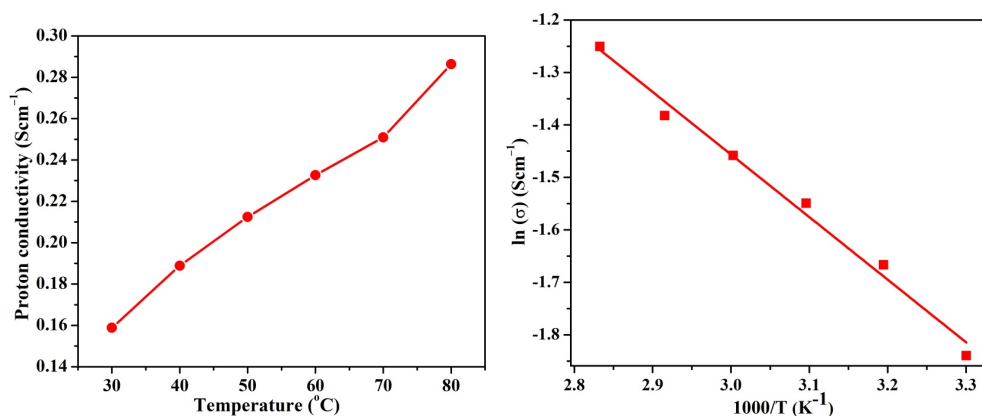
FIGURE 4.8: Proton conductivity (Left) and Activation energy (Right)
values for SPI PEM observed at temperatures from 30 °C - 80 °C synthe-
sized SPI ionomer

also be observed that at lower temperatures (303K, 313K, 323K), the high-frequency re-
gion is dominated by the H^+ diffusion-limited processes in the membrane whereas
at high temperatures (333K, 343K, 353K), there is improvement in kinetics though
diffusion-limited processes still exist but decrease in intensity.
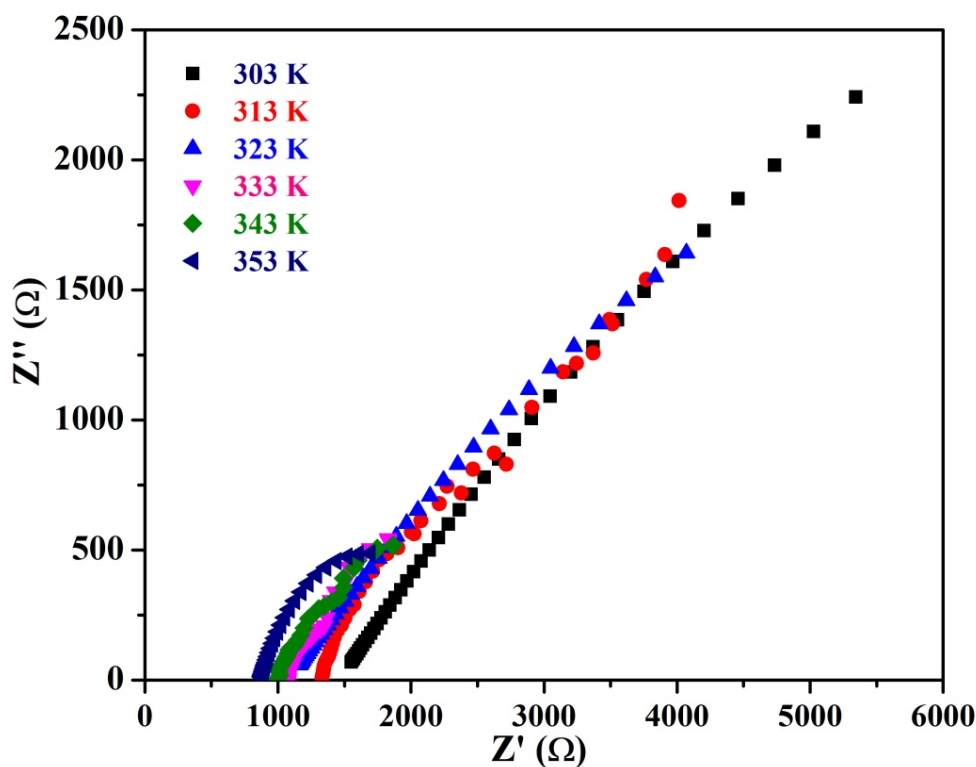
FIGURE 4.9: Nyquist plots observed for temperatures from 30 °C-80 °C
synthesized SPI ionomer

Table 4.3 gives a comparison of the activation energy values for Nafion PFSA mem-
branes and SPI membranes as reported in the literature and obtained in the present

TABLE 4.3: Comparison of Activation energy of the synthesized SPI membrane with other PEMs

	Activation energy (E_a)	Reference
Nafion 212	4.9	Lin, Chang, and Wang, 2013
Nafion 117	7.02	Yao, Shi, et al., 2016
NDTA/BDSA/ODA	12.3	Lin, Chang, and Wang, 2013
BTDA/BAPBDS/DAB	4.54	Pandey and Shahi, 2013
NTDA/ODADS/TFVBPA	9.52	Yao, Shi, et al., 2016
NTDA/DSDSA/MDP	9.91	This work

study. The activation energy of proton conduction is on the higher side compared to Nafion and some other SPIs.

4.4.6 Single cell PEMFC performance

Fig. 4.10 shows the current density-voltage (j-V) characteristic plot for Nafion MEA(Green) and 50% DS NTDA/DSDSA/MDP SPI MEA(Blue) for the single cell PEMFC operated at 80 °C. The activation characteristics of the PEMFC for Nafion and SPI PEM are comparable with the O.C.P. values of 0.85 V for Nafion and 0.80 V for the SPI PEM. As the current density increases, the kinetics losses increase drastically in both cases as shown by the decrease in the cell potential. In the ohmic region of the j-V plot, the SPI membrane shows higher resistance than Nafion despite having higher proton conductivity. At higher current density, the mass transport losses take over in both cases, however, the power density is marginally high in the SPI MEA. This is attributed to the bound water in minimal hydration conditions in NFSPI. The j-V characteristic plot is consistent with a similar SPI reported in the literature (Pan, Chen, et al., 2015).
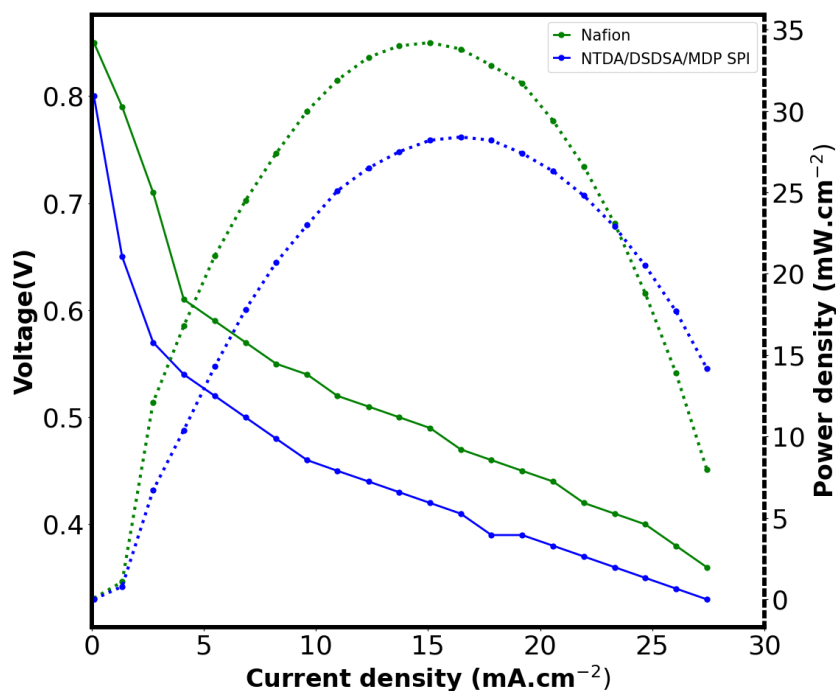
FIGURE 4.10:   Current density-Voltage (j-V) characteristic plot for
Nafion(Green) and 50% DS NTDA/DSDSA/MDP SPI(Blue) observed
at 80°C

4.5   Conclusion

Non-fluorinated SPI PEM (NTDA/DSDSA/MDP) with 50% degree of sulfonation was
developed through condensation polymerization and consequent solvent casting method.
Polymerization was confirmed by confirming the chemical structure of the synthe-
sized polymer using Fourier Transform infrared (FTIR) and Nuclear Magnetic Reso-
nance (NMR) spectroscopies. Physico-chemical properties namely Ion Exchange Ca-
pacities (IEC), water uptake, and swelling ratio along length and thickness were mea-
sured. There was good agreement with the water uptake values modeled in the MD
simulations according to the hydration numbers. In-plane proton conductivity val-
ues were found using an in-house developed four-probe setup and Electrochemical
Impedance Spectroscopy (EIS). The proton conductivity values found for the synthe-
sized SPI membranes were in the range of 0.1588-0.28636 S·cm^{-1} which is exceptionally
well for a PEM while those obtained in the MD simulations were 0.03-0.18 S·cm^{-1}. A
good agreement was observed between the proton conductivity values predicted us-
ing MD simulations in which only the proton conductivity was determined using the

hydronium ion diffusivities representing only one of the three operational mechanisms of proton diffusion. The SPI membranes were found to be stable in the harsh oxidative environment represented by the 30% Ferrous Sulphate solution in the Fenton reagent during oxidative testing as well as in the hydrolytic stability testing.

5 Conclusion and future outlook

5.1 Concluding remarks

Broadly, this work integrates the three realms of polymer discovery using data-driven techniques, polymer modeling using molecular dynamics, and polymer synthesis and testing for the targeted application of electrolyte membrane development for PEMFC. Herein, this integrated methodology has been successfully implemented for the Sulfonated Polyimides class of hydrocarbon PEMs. In fact, this is the first report of a data-based approach taken for designing novel polymer electrolyte membranes for fuel cells and it can be extended to other classes of prospective hydrocarbon PEM materials as well. Also, it was found that SPIs hold promise as electrolyte membrane materials showing a high range of proton conductivity.

Specifically, through this work, a data set of SPI-based ionomer membranes has been created to identify suitable SPI candidates for computational modeling and experimental synthesis based on insights obtained from data-driven approaches. Also, a combination of QSPR, machine learning, and data mining approaches was utilized to predict the proton conductivity of novel sulfonated polyimide polymer electrolyte membranes. Using decision tree classifiers, simplistic and interpretable Machine Learning models were obtained in the form of If-else statements using 81 unique SPI monomers. Also, a subset of the most relevant features was selected by manually including the features that increased the prediction accuracy of the decision tree classifier.

Further, our atomistic simulations have revealed interesting insights into several aspects through which molecular design affects the structural and dynamic properties of an ionomer. The study involved molecular dynamics simulations of Nafion PFSA ionomer and two novel SPI membrane materials namely - NTDA/DSDSA/HFBAPP (partially fluorinated SPI) and NTDA/DSDSA/MDP (non-fluorinated SPI) at different levels of hydration (λ = 1, 5, 10 and 15). Diffusion coefficients of water molecules and

hydronium ions were found to be drastically affected by the level of hydration in all three ionomers.

Finally, NTDA/DSDSA/MDP membrane was developed through condensation polymerization and solvent casting methods. The proton conductivity values found for the synthesized SPI membranes were in the range of 0.1588-0.28636 S·cm^{-1} which is exceptionally well for a PEM while those obtained in the MD simulations were 0.03-0.18 S·cm^{-1}. A good agreement was observed between the proton conductivity values predicted using MD simulations in which only the proton conductivity was determined using the hydronium ion diffusivities representing only one of the three operational mechanisms of proton diffusion.

5.2   Future outlook

Through this work, certain areas have been identified where future research and development work can be targeted:

- The data set can be expanded by a collection of data spanning across other classes of hydrocarbon PEMs.

- Advanced machine learning algorithms can be explored for polymer discovery with the increased number of datapoints in the data set.

- Molecular Dynamics studies can be designed to include the proton hopping mechanism of diffusion by making suitable modifications in the code or through use of additional packages suited for bond breaking and bond creation.

- There is a need for more in-depth studies into the structure-properties relationship especially the effect of morphology on hydrolytic stability and fuel cell performance.

- Moreover, it is important to study the role of acidic sites, nano-sized ionic domains, presence of bound and bulk water in the ionomer with regard to the diffusion of fuel (Hydrogen) and oxidant.

- Long-term durability studies are indispensable in creating robust SPI PEMs.

A Descriptor information/details

A1   QSPR and Semi-empirically calculated attributes:

The following attributes were considered for feature selection and model training in the next step.

- Literature attributes (Molar): Molar content of the sulfonate group

- Mordred descriptors (Mordred): ABC (Atom-bond connectivity index descriptor, a degree–base descriptor), ABCGG (Graovac-Ghorbani atom-bond connectivity index descriptor, a distance-based topological descriptor), nAromAtom (Aromatic atoms count), nAromBond (Aromatic bonds count), nheavyAtom (Number of heavy atoms, a heavy atom is any atom other than Hydrogen), nhetero (Number of hetero atoms, any atom that is not a Carbon atom or a Hydrogen atom)), nN (Number of Nitrogen atoms), nO (Number of Oxygen atoms), nbondsS (Number of single bonds in non-kekulized structure), nbondsD (Number of double bonds in non-kekulized structure), nbondsA (Number of aromatic bonds in non-kekulized structure), nbondsKS (Number of single bonds in kekulized structure Kekulization, refers to localization of the bond over an aromatic structure), nbondsKD (Number of double bonds in kekulized structure)

- Babel descriptors (Babel): nAromBond, dbonds (Number of double bonds), HBA1 (Number of Hydrogen Bond Acceptors), HBD (Number of Hydrogen Bond Donors), LogP (Partition coefficient, generally refers to the concentration ratio of nonionized species of a compound in two solvents e.g. when one of the solvents is water and the other is a non-polar solvent, then the logP value is a measure of

lipophilicity or hydrophobicity), molar refractivity (a measure of the total polar-
izability of a mole of a substance which is dependent on the temperature, index of
refraction, and pressure), Molecular Weight, nF or Fluorinated (Number of Fluo-
rine Atoms, initially a numeric descriptor but converted to categorical attribute),
sbonds (Number of single bonds), TPSA(Topological polar surface area)

- MOPAC semi-empirical descriptors (MopacD): Dispersion Energy [kcal.mol^{-1}]
  (van der Waals or VDW interaction arising from the instantaneous correlation
  of electrons), Core-core repulsion [eV] (Correction term added to compensate for
  approximations in calculating energies as only valence electrons are considered
  and the core electrons are treated together with the nuclei as one effective core
  potential), COSMO area [Å^2] and COSMO Volume [Å^3] (Conductor-like Screen-
  ing Model, which is a continuum approach to generates a conducting polygonal
  surface around the ion or molecule at the van der Waals' distance to determine
  the electrostatic interaction of a molecule with a solvent), Ionization Potential
  [eV] (energy needed to remove an electron and create a positive ion), HOMO En-
  ergies [kcal.mol^{-1}], LUMO Energies [kcal.mol^{-1}]

- MOPAC semi-empirical descriptors after adding three water molecules around
  each sulfonate (MopacW): Dispersion Energy (with water molecules) [kcal.mol^{-1}],
  H-bond Energy [kcal.mol^{-1}], Number of Hydrogen bonds, COSMO area (with
  water molecules) [Å^2] and COSMO Volume (with water molecules) [Å^3], Ion-
  ization Potential (with water molecules) [eV], Core-core repulsion (with water
  molecules) [eV], HOMO energies (with water molecules) [kcal.mol^{-1}], LUMO en-
  ergies (with water molecules) [kcal.mol^{-1}]

A2   Feature selection

A2.1   Gain ratio

DT induction in WEKA machine learning platform is based on C4.5 algorithm of In-
formation theoryQuinlan, 1996; Holmes, Donkin, and Witten, 1994. According to this,

information content ($Info(D)$) is calculated by the potential information produced by splitting the training data D into n partitions according to the outcomes or class labels as shown in Eq.A1. Net gain of information is obtained by subtracting the information content of the individual attribute ($Info_A(D)$) i.e. potential information obtained if the data points are partitioned according to the m outcomes or class labels of a test on attribute A from the total information content of the data set as shown in Eq.A2.

$$Info(D) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{A1}$$

$$Gain(A) = Info(D) - Info_A(D) \tag{A2}$$

where p_i is the probability of a data instance belonging to a class label (total instances of a class label divided by all the instances); D = total number of instances; A = Attribute.

Split information takes into account all possible splits of an attribute ($Split\ Info_A(D)$) and the occurrence of the data points lying in these split intervals as shown in Eq.A3.

$$Split\ Info_A(D) = \sum_{i=1}^{m} \frac{D_j}{D} \log_2(\frac{|D_j|}{|D|}) \tag{A3}$$

where D_j is the number of instances belonging to a particular class label in an attribute.

Gain Ratio is obtained by normalizing the gain by split information value so that the bias for more heterogeneous attributes or the attributes with multiple outcomes is removed as shown in Eq.A4.

$$Gain\ Ratio(A) = \frac{Gain(A)}{Split\ Info_A(D)} \tag{A4}$$

A2.2 ReliefF algorithm

The feature weight vector is predicted based on the Manhattan nearest neighbor distance method from a randomly selected initial instance Kononenko, Šimec, and Robnik-Šikonja, 1997. Using this, the distance between the nearest five same class instances

and the different class labels instances is calculated. The vector is updated with every ith iteration as mentioned below in as shown in Eq.A5:

$$W_i = W_i - (x_i - y_{same}^2) + (x_i - y_{diff}^2) \tag{A5}$$

where W_i is the weight vector whose elements represent the weights associated with each feature; x_i is the randomly selected data point; y_{same} is the nearest data point belonging to the same class label as x_i; y_{same} is the nearest data point belonging to the class label different from x_i.

A3   Classification metric/evaluation scores

The performance of a classifier is evaluated on the basis of how accurately it classifies the data instances according to the class labels to which they belong. There are many evaluation criteria based on which different characteristics of the classifier are drawn. These evaluators are derived from the confusion matrix that is a representation of true class labels and predicted class labels. The number of rightly and wrongly predicted class labels is given in the columns and actual numbers of the class labels are given in rows. It represents how many misclassification errors were performed by the classifier. If an HPC class label is predicted as 'HPC' then it is termed as True positive(TP) and if an LPC class label is predicted as 'LPC' then it is termed as True Negative(TN). On the contrary, if the HPC class label is predicted as 'LPC' then it is considered as False Negative(FN) and if an LPC class label is predicted as 'HPC' then it is a False Positive(FP) count. Based on this,the following classification metric/evaluation scores are evaluated:

A3.1   Accuracy

Accuracy for the classification algorithm is defined by the number of correctly classified instances whether they are HPC or LPC labels divided by the total number of predictions. In terms of SPI PEM repeat units, accuracy determines the percentage of all the repeat units corresponding to HPC and LPC which are classified as such among

all the classifications. The relationship of accuracy with TP, TN, FN and FP is given as

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{A6}$$

A3.2 True Positive Rate or Recall or Sensitivity

It represents the proportion of the labels which are actually HPC among the total number of HPC predictions. It indicates the completeness of the classifier and how exactly the classifier predicts the HPC cases as HPC. The low value of recall indicates that there are many HPC class label cases that were predicted as LPC and the high value of this parameter indicates that most of the HPC repeat units were identified and predicted correctly. The relationship of True Positive Rate (TPR) with TP, and FN is given as

$$TPR = \frac{TP}{TP + FN} \tag{A7}$$

A3.3 Precision

Precision is a measure of retrieved relevant instances e.g it represents actual HPC class labels among the total number of predicted HPC class labels. This value provides a measure of how much relevant the classifier is in predicting HPC as HPC. High precision indicates that only a small number of LPC repeat units are getting falsely predicted as HPC class labels whereas high precision indicates that a large number of repeat units that are associated with LPC class labels are getting classified as HPC which is not an ideal case. The relationship of TPR with TP and FP is given as

$$Precision = \frac{TP}{TP + FP} \tag{A8}$$

A3.4 Cohen's Kappa Statistics

It is the measure of the level of agreement between the ground reality and the predicted class labels. The higher the Cohen's Kappa statistic (K) the better it is as $K = 1$ means perfect agreement and $K = 0$ means chance agreement. The value between $0.80 - 0.90$ represents a strong classification model. The relation of K is given as follows

$$K = \frac{p_o - p_e}{1 - p_e} \tag{A9}$$

where p_o is the percentage agreement between classifier and ground truth, and p_e is the percentage of chance agreement.

A3.5 F_1 measure

F_1 measure combines precision and recall by taking the harmonic mean of these parameters. We want the repeat unit in the test case to be rightly predicted whether it is HPC or LPC (precision) such that the right predictions should be done for as many repeat units in the test case as possible (recall). The F_1 measure balances this trade-off.

$$F_1 = 2 * \frac{precision.recall}{precision + recall} \tag{A10}$$

A3.6 False Positive Rate

False Positive Rate (FPR) measure represents the proportion of the wrong predictions i.e. actual LPC predicted as HPC among all actual negatives or LPC class labels. The high value of this measure indicates that a large number of actual LPC repeat units are classified as HPC which is an undesirable scenario as LPC repeat units might get screened for further processing.

$$FPR = \frac{FP}{FP + TN} \tag{A11}$$

A3.7 Specificity or TNR

Specificity relates to the classifier's ability to identify LPC or negative class labels. A higher value of this measure indicates that a large number of actual LPC repeat units are classified as LPC and misclassification of LPC as HPC (i.e. FP) is low. The relationship of TNR with FP and TN is given as

$$TNR = \frac{TN}{FP + TN} \tag{A12}$$

A3.8   Receiver Operating Characteristics

The area under the Receiver Operating Characteristics (ROC) graph is the area under the graph of the FPR on the x-axis and TPR on the y-axis. This value indicates the quality of the classifier for both class labels individually. The perfect classifier lies on or nearer to the y-axis and the value is closer to 1. Its value determines the frequency with which a class label is predicted.

A4   Regression metrics/evaluation scores

Similar to classification evaluation scores, there are several criteria based on which RFR algorithms are evaluated, as mentioned below:

A4.1   Mean squared error (MSE)

Mean Squared Error (MSE) represents the average absolute value of the squared difference between the original and predicted values.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} [y_i - \hat{y}]^2 \tag{A13}$$

A4.2   Root mean squared error (RMSE)

Root Mean Squared Error (RMSE) is the square root of the Mean Squared error. It is always a positive value and is sensitive to outliers in the data set.

$$RMSE = \frac{1}{N} \sqrt{\sum_{i=1}^{N} [y_i - \hat{y}]^2} \tag{A14}$$

A4.3   R-squared (R^2)

The coefficient of determination or R-squared (R^2) represents the goodness of fit for the predicted value.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} [y_i - \hat{y}]^2}{\sum_{i=1}^{N} [y_i - \bar{y}]^2} \tag{A15}$$

A4.4   Mean absolute error (MAE)

Mean absolute error (MAE) represents the averaged absolute difference between the original and predicted values in the dataset.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} [y_i - \hat{y}] \tag{A16}$$

B Force field details

A1 Polymer Consistent Force Field

Bonded and non-bonded parameters were assigned according to the PCFF (Polymer Consistent Force Field) force field.

The total energy of all the bonded and non-bonded interactions is given by Eq. A1.

$$E = E_b + E_\theta + E_\phi + E_\chi + E_{bb'} + E_{b\theta} + E_{b\phi} + E_{\theta\phi} + E_{\theta\theta'} + E_{\theta\theta'\phi} + E_q + E_{vdW} \qquad \text{(A1)}$$

- The energy contribution of quartic bond-stretching is determined by Eq. A2 with b_o representing the equilibrium bond length:

$$E_b = K_{b2}(b - b_0)^2 + K_{b3}(b - b_0)^3 + K_{b4}(b - b_0)^4 \qquad \text{(A2)}$$

where coefficients K_{b2}, K_{b3} and K_{b4} are constants.

- The energy contribution of quartic angles is determined by Eq. A3 with θ_o representing the equilibrium angle:

$$E_\theta = K_{\theta2}(\theta - \theta_0)^2 + K_{\theta3}(\theta - \theta_0)^3 + K_{\theta4}(\theta - \theta_0)^4 \qquad \text{(A3)}$$

where coefficients $K_{\theta2}$, $K_{\theta3}$ and $K_{\theta4}$ are constants.

- The term for torsion is given by Eq. A4:

$$E_\phi = \sum_\phi [K_{\phi1}(1 - cos\phi) + K_{\phi2}(1 - cos2\phi) + K_{\phi3}(1 - cos3\phi)] \qquad \text{(A4)}$$

where coefficients $K_{\phi1}$, $K_{\phi2}$ and $K_{\phi3}$ are constants.

- The term for Wilson out-of-plane interactions is given by Eq. A5

$$E_\chi = \sum_\chi K_\chi \chi^2 \tag{A5}$$

where K_χ is constant.

The cross-term parameters are included as:

- The bond-bond parameters are given by Eq. A6:

$$E_{bb'} = K_{bb'}(b - b_0)(b' - b'_0) \tag{A6}$$

where $K_{bb'}$ is constant

- The bond-angle parameters are given by Eq. A7:

$$E_{b\theta} = K_{b\theta}(b - b_0)(\theta - \theta_0) \tag{A7}$$

where $K_{b\theta}$ are constants.

- The middle bond-torsion parameters are given by Eq. A8:

$$E_{b\phi} = (b - b_0)[K_{b\phi 1}\cos\phi + K_{b\phi 2}\cos 2\phi + K_{b\phi 3}\cos 3\phi] \tag{A8}$$

where $K_{b\phi 1}$, $K_{b\phi 2}$ and $K_{b\phi 3}$ are constants.

- The end bond-torsion and middle bond-torsion parameters are given by Eq. A9:

$$\begin{aligned} E_{b'\phi} = (b' - b'_0)\left[K_{b'\phi 1}\cos\phi + K_{b'\phi 2}\cos 2\phi + K_{b'\phi 3}\cos 3\phi\right] + \\ (b'' - b''_0)\left[K_{b''\phi 1}\cos\phi + K_{b''\phi 2}\cos 2\phi + K_{b''\phi 3}\cos 3\phi\right] \end{aligned} \tag{A9}$$

where $K_{b'\phi 1}$, $K_{b'\phi 2}$, $K_{b'\phi 3}$, $K_{b''\phi 1}$, $K_{b''\phi 2}$ and $K_{b''\phi 3}$ are constants.

- The angle-torsion parameters are given by Eq. A10:

$$E_{\theta\phi} = \sum_{\theta,\phi}(\theta - \theta_o)[K_{\theta\phi 1}(1 - cos\phi) + K_{\theta\phi 2}(1 - cos2\phi) + K_{\theta\phi 3}(1 - cos3\phi)] \tag{A10}$$

where $K_{\theta\phi 1}$, $K_{\theta\phi 2}$ and $K_{\theta\phi 3}$ are constants.

- The angle-angle parameters are given by Eq. A11:

$$E_{\theta\phi} = (\theta - \theta_0) \left[K_{\theta\phi 1} \cos \phi + K_{\theta\phi 2} \cos 2\phi + K_{\theta\phi 3} \cos 3\phi \right] +$$
$$(\theta' - \theta_0') \left[K_{\theta'\phi 1} \cos \phi + K_{\theta'\phi 2} \cos 2\phi + K_{\theta'\phi 3} \cos 3\phi \right] \tag{A11}$$

  where $K_{\theta\phi 1}$, $K_{\theta\phi 2}$, $K_{\theta\phi 3}$, $K_{\theta'\phi 1}$, $K_{\theta'\phi 2}$ and $K_{\theta'\phi 3}$ are constants.

- The angle-angle-torsion parameters are given by Eq. A12:

$$E_{\theta\theta'\phi} = K_{\theta\theta'\phi}(\theta - \theta_0)(\theta' - \theta_0') \cos \phi \tag{A12}$$

  where $K_{\theta\theta'\phi}$ are constants.

- The electrostatic interactions are defined by the Coulombic term according to Eq. A13

$$E_q = \sum_{ij} \frac{q_i q_j}{r_{ij}} \tag{A13}$$

- The non-bonded or van der Waals interactions are governed by Eq. A14 where an inverse 9th-power term is used for the repulsive part as:

$$E_{vdW} = \sum_{ij} \epsilon_{ij} [2(\frac{r_{ij}^o}{r_{ij}})^9 - 3(\frac{r_{ij}^o}{r_{ij}})^6] \tag{A14}$$

- The Waldman-Hagler combining rules are used calculating r_{ij}^o and ϵ_{ij} as given by Eq. A15 and Eq. A16:

$$r_{ij}^o = (\frac{(r_i^o)^6 + (r_j^o)^6}{2})^{\frac{1}{6}} \tag{A15}$$

$$\epsilon_{ij} = 2\sqrt{\epsilon_i \cdot \epsilon_j} [\frac{(r_i^o)^3 \cdot (r_j^o)^3}{(r_i^o)^6 + (r_j^o)^6}] \tag{A16}$$

Bibliography

A., E. (1970). "Clustering of ions in organic polymers. A theoretical approach". In: *Macromolecules* 3.2, pp. 147–54.

Afzal, M. A. F., C. Cheng, and J. Hachmann (2018). "Combining first-principles and data modeling for the accurate prediction of the refractive index of organic polymers". In: *The Journal of chemical physics* 148.24, p. 241712.

Agmon, N. (1995). "The grotthuss mechanism". In: *Chemical Physics Letters* 244.5-6, pp. 456–462.

Agrawal, A. and A. Choudhary (2016). "Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science". In: *Apl Materials* 4.5, p. 053208.

Akbarian-Feizi, L., S. Mehdipour-Ataei, and H. Yeganeh (2010a). "Survey of sulfonated polyimide membrane as a good candidate for nafion substitution in fuel cell". In: *International Journal of Hydrogen Energy* 35.17, pp. 9385–9397.

Akbarian-Feizi, L., S. Mehdipour-Ataei, and H. Yeganeh (2010b). "Survey of sulfonated polyimide membrane as a good candidate for nafion substitution in fuel cell". In: *international journal of hydrogen energy* 35.17, pp. 9385–9397.

Akbarian-Feizi, L., S. Mehdipour-Ataei, and H. Yeganeh (2012). "Synthesis of new sulfonated copolyimides in organic and ionic liquid media for fuel cell application". In: *Journal of Applied Polymer Science* 124.3, pp. 1981–1992.

Berrod, Q., S. Hanot, A. Guillermo, S. Mossa, and S. Lyonnard (2017). "Water subdiffusion in membranes for fuel cells". In: *Scientific reports* 7.1, pp. 1–14.

Bhowmik, R., S. Sihn, R. Pachter, and J. P. Vernon (2021). "Prediction of the specific heat of polymers from experimental data and machine learning methods". In: *Polymer* 220, p. 123558.

Blázquez, J. A., J. J. Iruin, S. Eceolaza, C. Marestin, R. Mercier, D. Mecerreyes, O. Miguel, A. Vela, and R. Marcilla (2005). "Solvent and acidification method effects in the

performance of new sulfonated copolyimides membranes in PEM-fuel cells". In: *Journal of Power Sources* 151, pp. 63–68.

Chen, K, X Chen, K Yaguchi, N Endo, M Higa, and K. Okamoto (2009). "Synthesis and properties of novel sulfonated polyimides bearing sulfophenyl pendant groups for fuel cell application". In: *Polymer* 50.2, pp. 510–8.

Chen, S., Y. Yin, K. Tanaka, H. Kita, and K.-i. Okamoto (2006a). "Synthesis and properties of novel side-chain-sulfonated polyimides from bis[4-(4-aminophenoxy)-2-(3-sulfobenzoyl)]phenyl sulfone". In: *Polymer* 47.8, pp. 2660–2669. DOI: 10.1016/j.polymer.2006.02.045.

Chen, X, K Chen, P Chen, M Higa, K. Okamoto, and T Hirano (2010). "Effects of tetracarboxylic dianhydrides on the properties of sulfonated polyimides". In: *Journal of Polymer Science Part A: Polymer Chemistry* 48.4, pp. 905–15.

Chen, X, Y Yin, K Tanaka, H Kita, and K. Okamoto (2006b). "Synthesis and characterization of novel sulfonated polyimides derived from naphthalenic dianhydride". In: *High Performance Polymers* 18.5, pp. 637–54.

Chhabra, P and V Choudhary (2009). "Synthesis and characterization of sulfonated naphthalenic polyimides based on 4, 4-diamino diphenyl ether-2, 2-disulfonic acid, and bis [4-(4-aminophenoxy) phenylhexafluoropropane] for fuel cell applications." In: *European Polymer Journal* 45.5, pp. 1467–75.

Choi, P., N. H. Jalani, and R. Datta (2005). "Thermodynamics and proton transport in nafion: I. membrane swelling, sorption, and ion-exchange equilibrium". In: *Journal of The Electrochemical Society* 152.3, E84.

Cornet, N, O. Diat, G Gebel, F Jousse, D Marsacq, R. Mercier, and M. Pineri (2000). "Sulfonated polyimide membranes: a new type of ion-conducting membrane for electrochemical applications". In: *Journal of new materials for electrochemical systems* 3.1, pp. 33–42.

Cui, S., J. Liu, M. E. Selvan, D. J. Keffer, B. J. Edwards, and W. V. Steele (2007). "A molecular dynamics study of a nafion polyelectrolyte membrane and the aqueous phase structure for proton transport". In: *The Journal of Physical Chemistry B* 111.9, pp. 2208–2218.

Dennis, J. M. and D. Y. Zubarev (2021). "Hebbian Learning on Small Data Enables Experimental Discovery of High T g Polyimides". In: *The Journal of Physical Chemistry A* 125.31, pp. 6829–6835.

Devanathan, R., A. Venkatnathan, and M. Dupuis (2007a). "Atomistic simulation of Nafion membrane. 2. Dynamics of water molecules and hydronium ions". In: *The Journal of Physical Chemistry B* 111.45, pp. 13006–13013.

Devanathan, R., A. Venkatnathan, and M. Dupuis (2007b). "Atomistic simulation of nafion membrane: I. Effect of hydration on membrane nanostructure". In: *The Journal of Physical Chemistry B* 111.28, pp. 8069–8079.

Devanathan, R., A. Venkatnathan, R. Rousseau, M. Dupuis, T. Frigato, W. Gu, and V. Helms (2010). "Atomistic simulation of water percolation and proton hopping in Nafion fuel cell membrane". In: *The Journal of Physical Chemistry B* 114.43, pp. 13681–13690.

Dhra, G., A. Balasubramanian, and T. Kannan (2021). "Polyelectrolyte proton exchange membranes: synthesis and characterization of sulfonated polyimide membranes using novel stilbene-containing diamine". In: *Polymer Bulletin*, pp. 1–19.

Drefahl, A. (2011). "CurlySMILES: a chemical language to customize and annotate encodings of molecular and nanodevice structures". In: *Journal of cheminformatics* 3.1, pp. 1–7.

Eikerling, M, A. Kornyshev, A. Kuznetsov, J. Ulstrup, and S Walbran (2001). "Mechanisms of proton conductance in polymer electrolyte membranes". In: *The Journal of Physical Chemistry B* 105.17, pp. 3646–3662.

Eikerling, M., A. A. Kornyshev, and E. Spohr (2008). "Proton-conducting polymer electrolyte membranes: water and structure in charge". In: *Fuel cells I*, pp. 15–54.

Einsla, B., Y. Hong, Y Seung Kim, F Wang, N Gunduz, and J. McGrath (2004). "Sulfonated naphthalene dianhydride based polyimide copolymers for proton-exchange-membrane fuel cells. I. Monomer and copolymer synthesis." In: *Journal of Polymer Science Part A: Polymer Chemistry* 42.4, pp. 862–74.

Einsla, B. R., Y. S. Kim, M. A. Hickner, Y.-T. Hong, M. L. Hill, B. S. Pivovar, and J. E. McGrath (2005). "Sulfonated naphthalene dianhydride based polyimide copolymers for proton-exchange-membrane fuel cells: II. Membrane properties and fuel cell performance". In: *Journal of Membrane Science* 255.1-2, pp. 141–8.

Eisenberg, A. (1970). "Clustering of ions in organic polymers. A theoretical approach". In: *Macromolecules* 3.2, pp. 147–154.

Essafi, W, G Gebel, and R Mercier (2004a). "Sulfonated polyimide ionomers: a structural study". In: *Macromolecules* 37.4, pp. 1431–40.

Essafi, W., G. Gebel, and R. Mercier (2004b). "Sulfonated polyimide ionomers: a structural study". In: *Macromolecules* 37.4, pp. 1431–1440.

Fang, J., X. Guo, S. Harada, T. Watari, K. Tanaka, H. Kita, and K.-i. Okamoto (2002). "Novel sulfonated polyimides as polyelectrolytes for fuel cell application. 1. Synthesis, proton conductivity, and water stability of polyimides from 4, 4 '-diaminodiphenyl ether-2, 2 '-disulfonic acid". In: *Macromolecules* 35.24, pp. 9022–9028.

Fang, J., J. Qiao, D. P. Wilkinson, and J. Zhang (2015a). *Electrochemical polymer electrolyte membranes*. Vol. 7. CRC Press New York.

Fang, J., J. Qiao, D. P. Wilkinson, and J. Zhang (2015b). *Electrochemical polymer electrolyte membranes*. Vol. 7. CRC Press New York.

Fang, J., J. Qiao, D. P. Wilkinson, and J. Zhang (2019). *Electrochemical polymer electrolyte membranes*. CRC Press.

Feng, S., S. Kondo, T. Kaseyama, T. Nakazawa, T. Kikuchi, R. Selyanchyn, S. Fujikawa, L. Christiani, K. Sasaki, and M. Nishihara (2018). "Development of polymer-polymer type charge-transfer blend membranes for fuel cell application". In: *Journal of Membrane Science* 548, pp. 223–231. DOI: 10.1016/j.memsci.2017.11.025.

Ganeshkumar, A., D. Bera, E. A. Mistri, and S. Banerjee (2014). "Triphenyl amine containing sulfonated aromatic polyimide proton exchange membranes". In: *European Polymer Journal* 60, pp. 235–246. DOI: 10.1016/j.eurpolymj.2014.09.009.

Garrido, L., J. Pozuelo, M. Lopez-Gonzalez, J. Fang, and E. Riande (2009). "Simulation and experimental studies on proton diffusion in polyelectrolytes based on sulfonated naphthalenic copolyimides". In: *Macromolecules* 42.17, pp. 6572–6580.

Genies, C, R. Mercier, B. Sillion, R. Petiaud, N Cornet, G Gebel, and M Pineri (2001). "Stability study of sulfonated phthalic and naphthalenic polyimide structures in aqueous medium". In: *Polymer* 42.12, pp. 5097–5105.

Gierke, T. D., G. Munn, and F. Wilson (1981). "The morphology in nafion perfluorinated membrane products, as determined by wide-and small-angle x-ray studies". In: *Journal of Polymer Science: Polymer Physics Edition* 19.11, pp. 1687–1704.

Grubmüller, H., H. Heller, A. Windemuth, and K. Schulten (1991). "Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions". In: *Molecular Simulation* 6.1-3, pp. 121–142.

Guo, X., J. Fang, T. Watari, K. Tanaka, H. Kita, and K.-i. Okamoto (2002). "Novel sulfonated polyimides as polyelectrolytes for fuel cell application. 2. Synthesis and proton conductivity of polyimides from 9, 9-bis (4-aminophenyl) fluorene-2, 7-disulfonic acid". In: *Macromolecules* 35.17, pp. 6707–6713.

Hansen, K., F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. Von Lilienfeld, K.-R. Muller, and A. Tkatchenko (2015). "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space". In: *The journal of physical chemistry letters* 6.12, pp. 2326–2331.

Hanwell, M. D., D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison (2012). "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform". In: *Journal of Cheminformatics* 4.1, pp. 1–17.

Haubold, H.-G., T. Vad, H Jungbluth, and P Hiller (2001). "Nano structure of NAFION: a SAXS study". In: *Electrochimica Acta* 46.10-11, pp. 1559–1563.

Hickner, M. and B. Pivovar (2005). "The chemical and structural nature of proton exchange membrane fuel cell properties". In: *Fuel cells* 5.2, pp. 213–229.

Hickner, M. A., H. Ghassemi, Y. S. Kim, B. R. Einsla, and J. E. McGrath (2004). "Alternative polymer systems for proton exchange membranes (PEMs)". In: *Chemical reviews* 104.10, pp. 4587–4612.

Holmes, G., A. Donkin, and I. H. Witten (1994). "Weka: A machine learning workbench". In: *Proceedings of ANZIIS'94-Australian New Zealnd Intelligent Information Systems Conference*. IEEE, pp. 357–361.

Hu, C., T. Lu, and H. Guo (2018). "Mesoscale modeling of sulfonated polyimides copolymer membranes: Effect of sequence distributions". In: *Journal of Membrane Science* 564, pp. 146–158.

Huang, Y., J. Zhang, E. S. Jiang, Y. Oya, A. Saeki, G. Kikugawa, T. Okabe, and F. S. Ohuchi (2020). "Structure–property correlation study for organic photovoltaic polymer materials using data science approach". In: *The Journal of Physical Chemistry C* 124.24, pp. 12871–12882.

Hunter, A. D. (1997). *ACD/ChemSketch 1.0 (freeware); ACD/ChemSketch 2.0 and its tau-tomers, dictionary, and 3D plug-ins; ACD/HNMR 2.0; ACD/CNMR 2.0.*

Ito, G., M. Tanaka, and H. Kawakami (2018). "Sulfonated polyimide nanofiber framework: Evaluation of intrinsic proton conductivity and application to composite membranes for fuel cells". In: *Solid State Ionics* 317, pp. 244–255.

Jabeen, F., M. Chen, B. Rasulev, M. Ossowski, and P. Boudjouk (2017). "Refractive indices of diverse data set of polymers: A computational QSPR based study". In: *Computational Materials Science* 137, pp. 215–224.

Jacobson, M. Z., W. Colella, and D. Golden (2005). "Cleaning the air and improving health with hydrogen fuel-cell vehicles". In: *Science* 308.5730, pp. 1901–1905.

James, J. (2016). *Stewart, Stewart Computational Chemistry MOPAC2016.*

Jang, W., C. Lee, S. Sundar, Y. G. Shul, and H. Han (2005). "Thermal and hydrolytic stability of sulfonated polyimide membranes with varying chemical structure". In: *Polymer Degradation and Stability* 90.3, pp. 431–440. DOI: 10.1016/j.polymdegradstab.2005.04.012.

Jørgensen, P. B., M. Mesta, S. Shil, J. M. García Lastra, K. W. Jacobsen, K. S. Thygesen, and M. N. Schmidt (2018). "Machine learning-based screening of complex molecules for polymer solar cells". In: *The Journal of chemical physics* 148.24, p. 241735.

Karo, J., A. Aabloo, J. O. Thomas, and D. Brandell (2010). "Molecular dynamics modeling of proton transport in nafion and hyflon nanostructures". In: *The Journal of Physical Chemistry B* 114.18, pp. 6056–6064.

Kim, C., A. Chandrasekaran, T. D. Huan, D. Das, and R. Ramprasad (2018). "Polymer genome: a data-powered polymer informatics platform for property predictions". In: *The Journal of Physical Chemistry C* 122.31, pp. 17575–17585.

Kim, Y. S. and K.-S. Lee (2015a). "Fuel cell membrane characterizations". In: *Polymer Reviews* 55.2, pp. 330–370.

Kim, Y. S. and K.-S. Lee (2015b). "Fuel cell membrane characterizations". In: *Polymer Reviews* 55.2, pp. 330–370.

Knauth, P. and M. L. Di Vona (2014). "Hydration and proton conductivity of ionomers: the model case of Sulfonated Aromatic Polymers". In: *Frontiers in Energy Research* 2, p. 50.

Kohonen, T. (1990). "The self-organizing map". In: *Proceedings of the IEEE* 78.9, pp. 1464–1480.

Kononenko, I., E. Šimec, and M. Robnik-Šikonja (1997). "Overcoming the myopia of inductive learning algorithms with RELIEFF". In: *Applied Intelligence* 7.1, pp. 39–55.

Kreuer, K. (2010). "Hydrocarbon membranes". In: *Handbook of fuel cells*.

Kwon, S. H., H. Kang, J. H. Lee, S. Shim, J. Lee, D. S. Lee, C. M. Kim, and S. G. Lee (2019). "Investigating the influence of the side-chain pendants of perfluorosulfonic acid membranes in a PEMFC by molecular dynamics simulations". In: *Materials Today Communications* 21, p. 100625.

Laporta, M, M Pegoraro, and L Zanderighi (1999). "Perfluorosulfonated membrane (Nafion): FT-IR study of the state of water with increasing humidity". In: *Physical Chemistry Chemical Physics* 1.19, pp. 4619–4628.

Larminie, J., A. Dicks, and M. S. McDonald (2003). *Fuel cell systems explained*. Vol. 2. J. Wiley Chichester, UK.

Lee, C. H., S. Y. Hwang, J. Y. Sohn, H. B. Park, J. Y. Kim, and Y. M. Lee (2006). "Water-stable crosslinked sulfonated polyimide–silica nanocomposite containing interpenetrating polymer network". In: *Journal of Power Sources* 163.1, pp. 339–348. DOI: 10.1016/j.jpowsour.2006.09.023.

Lee, S.-Y., A. Ogawa, M. Kanno, H. Nakamoto, T. Yasuda, and M. Watanabe (2010). "Nonhumidified intermediate temperature fuel cells using protic ionic liquids". In: *Journal of the American Chemical Society* 132.28, pp. 9764–73.

Li, N., Z. Cui, S. Zhang, and W. Xing (2007). "Sulfonated polyimides bearing benzimidazole groups for proton exchange membranes". In: *Polymer* 48.25, pp. 7255–7263.

Li, Q, R He, J. Jensen, and N. Bjerrum (2004). "PBI-based polymer membranes for high temperature fuel cells–preparation, characterization and fuel cell demonstration". In: *Fuel cells* 4.3, pp. 147–159.

Lin, C.-C., C.-B. Chang, and Y.-Z. Wang (2013). "Preparation and properties of crosslinked sulfonated poly (imide-siloxane) for polymer electrolyte fuel cell application". In: *Journal of power sources* 223, pp. 277–283.

Lin, C.-C., K.-S. Ho, C.-F. Chou, W.-F. Lien, and Y.-Z. Wang (2013). "A facile approach toward preparing low methanol permeable sulfonated polyimide composites: The

sulfonated polyimide blended with poly (vinyl alcohol-co-vinyl acetate)". In: *Solid State Ionics* 244, pp. 40–7.

Lin, T.-S., C. W. Coley, H. Mochigase, H. K. Beech, W. Wang, Z. Wang, E. Woods, S. L. Craig, J. A. Johnson, J. A. Kalow, et al. (2019). "BigSMILES: a structurally-based line notation for describing macromolecules". In: *ACS central science* 5.9, pp. 1523–1531.

Liu, H., S. Cavaliere, D. J. Jones, J. Roziere, and S. J. Paddison (2018). "Morphology of hydrated nafion through a quantitative cluster analysis: a case study based on dissipative particle dynamics simulations". In: *The Journal of Physical Chemistry C* 122.24, pp. 13130–13139.

Liu, M., C. Clement, K. Liu, X. Wang, and T. D. Sparks (2021). "A data science approach for advanced solid polymer electrolyte design". In: *Computational Materials Science* 187, p. 110108.

Maier, G. and J. Meier-Haack (2008). "Sulfonated aromatic polymers for fuel cell membranes". In: *Fuel cells II*, pp. 1–62.

Manabe, S. and R. T. Wetherald (1967). "Thermal equilibrium of the atmosphere with a given distribution of relative humidity". In.

McDonald, I. (1972). "NpT-ensemble Monte Carlo calculations for binary liquid mixtures". In: *Molecular Physics* 23.1, pp. 41–58.

Mercader, A. G. and P. R. Duchowicz (2016). "Encoding alternatives for the prediction of polyacrylates glass transition temperature by quantitative structure-property relationships". In: *Materials Chemistry and Physics* 172, pp. 158–164.

Mikhailenko, S., M. Guiver, and S Kaliaguine (2008). "Measurements of PEM conductivity by impedance spectroscopy". In: *Solid State Ionics* 179.17-18, pp. 619–624.

Mistri, E. A., S. Banerjee, H. Komber, and B. Voit (2015). "Structure–property correlation of semifluorinated 6-membered co-SPIs for proton exchange membrane". In: *European Polymer Journal* 73, pp. 466–479.

Mistri, E. A., A. K. Mohanty, and S. Banerjee (2012). "Synthesis and characterization of new fluorinated poly (ether imide) copolymers with controlled degree of sulfonation for proton exchange membranes". In: *Journal of Membrane Science* 411, pp. 117–129.

Mittal, K. (2005). *Polyimides and other high-temperature polymers: synthesis, characterization, and applications*. CRC Press.

Moriwaki, H., Y.-S. Tian, N. Kawashita, and T. Takagi (2018). "Mordred: a molecular descriptor calculator". In: *Journal of cheminformatics* 10.1, pp. 1–14.

Mueller, T., A. G. Kusne, and R. Ramprasad (2016). "Machine learning in materials science: Recent progress and emerging applications". In: *Reviews in Computational Chemistry* 29, pp. 186–273.

O'Boyle, N. M., M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison (2011). "Open Babel: An open chemical toolbox". In: *Journal of cheminformatics* 3.1, pp. 1–14.

Otsuka, S., I. Kuwajima, J. Hosoya, Y. Xu, and M. Yamazaki (2011). "PoLyInfo: Polymer database for polymeric materials design". In: *2011 International Conference on Emerging Intelligent Data and Web Technologies*. IEEE, pp. 22–29.

Ozmaian, M. and R. Naghdabadi (2014). "Modeling and simulation of the water gradient within a Nafion membrane". In: *Physical Chemistry Chemical Physics* 16.7, pp. 3173–3186.

Paddison, S. (2003). "Proton conduction mechanisms at low degrees of hydration in sulfonic acid-based polymer electrolyte membranes". In: *Annual Review of Materials Research* 33, p. 289.

Palomba, D., G. E. Vazquez, and M. F. Díaz (2012). "Novel descriptors from main and side chains of high-molecular-weight polymers applied to prediction of glass transition temperatures". In: *Journal of Molecular Graphics and Modelling* 38, pp. 137–147.

Pan, H., S. Chen, Y. Zhang, M. Jin, Z. Chang, and H. Pu (2015). "Preparation and properties of the cross-linked sulfonated polyimide containing benzimidazole as electrolyte membranes in fuel cells". In: *Journal of Membrane Science* 476, pp. 87–94.

Panagiotopoulos, A. Z. (1987). "Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble". In: *Molecular Physics* 61.4, pp. 813–826.

Pandey, R. P. and V. K. Shahi (2013). "Aliphatic-aromatic sulphonated polyimide and acid functionalized polysilsesquioxane composite membranes for fuel cell applications". In: *Journal of Materials Chemistry A* 1.45, pp. 14375–14383.

Pandey, R. P. and V. K. Shahi (2015). "Phosphonic acid grafted poly (ethyleneimine)-silica composite polymer electrolyte membranes by epoxide ring opening: improved

conductivity and water retention at high temperature". In: *International Journal of Hydrogen Energy* 40.41, pp. 14235–14245.

Park, C. H., T.-H. Kim, S. Y. Nam, and Y. T. Hong (2019). "Water channel morphology of non-perfluorinated hydrocarbon proton exchange membrane under a low humidifying condition". In: *International Journal of Hydrogen Energy* 44.4, pp. 2340–2348.

Park, K. T., S. G. Kim, J. H. Chun, D. H. Jo, B.-H. Chun, W. I. Jang, G. B. Kang, S. H. Kim, and K. B. Lee (2011). "Composite membranes based on a sulfonated poly (arylene ether sulfone) and proton-conducting hybrid silica particles for high temperature PEMFCs". In: *International Journal of Hydrogen Energy* 36.17, pp. 10891–10900.

Peighambardoust, S., S Rowshanzamir, and M Amjadi (2010). "Review of the proton exchange membranes for fuel cell applications". In: *International Journal of Hydrogen Energy* 35.17, pp. 9349–9384.

Pu, H. (2014). *Polymers for PEM fuel cells*. John Wiley Sons.

Quinlan, J. R. (1996). "Improved use of continuous attributes in C4. 5". In: *Journal of artificial intelligence research* 4, pp. 77–90.

Rikukawa, M. and S. K (2000). "Proton-conducting polymer electrolyte membranes based on hydrocarbon polymers". In: *Progress in Polymer Science* 25.10, pp. 1463–502.

Roziere, J. and D. J. Jones (2003). "Non-fluorinated polymer materials for proton exchange membrane fuel cells". In: *Annual Review of Materials Research* 33.1, pp. 503–555.

Rupp, M., A. Tkatchenko, K.-R. Muller, and O. A. Von Lilienfeld (2012). "Fast and accurate modeling of molecular atomization energies with machine learning". In: *Physical review letters* 108.5, p. 058301.

Saito, J., K. Miyatake, and M. Watanabe (2008). "Synthesis and Properties of Polyimide Ionomers Containing 1 H-1, 2, 4-Triazole Groups". In: *Macromolecules* 41.7, pp. 2415–20.

Sambandam, S. and V. Ramani (2007). "SPEEK/functionalized silica composite membranes for polymer electrolyte fuel cells". In: *Journal of power sources* 170.2, pp. 259–267.

Sek, D., A. Wanic, and E. Schab-Balcerzak (1995). "Investigation of polyimides containing naphthalene units. II. Model compounds synthesis". In: *Journal of Polymer Science Part A: Polymer Chemistry* 33.3, pp. 547–554.

Shin, D. W., M. D. Guiver, and Y. M. Lee (2017). "Hydrocarbon-based polymer electrolyte membranes: importance of morphology on ion transport and membrane stability". In: *Chemical reviews* 117.6, pp. 4759–4805.

Sone, Y., P. Ekdunge, and D. Simonsson (1996). "Proton conductivity of Nafion 117 as measured by a four-electrode AC impedance method". In: *Journal of the Electrochemical Society* 143.4, p. 1254.

Steele, B. C. and A. Heinzel (2011). "Materials for fuel-cell technologies". In: *Materials for Sustainable Energy: A Collection of Peer-Reviewed Research and Review Articles from Nature Publishing Group*. World Scientific, pp. 224–231.

Stewart, J. J. (2007). "Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements". In: *Journal of Molecular modeling* 13.12, pp. 1173–1213.

Stewart, J. J. (2012). "MOPAC2012. Stewart Computational Chemistry, Colorado Springs". In: *CO, USA*.

Stukowski, A. (2009). "Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool". In: *Modelling and simulation in materials science and engineering* 18.1, p. 015012.

Sundar, S., W. Jang, C. Lee, Y. Shul, and H. Han (2005). "Crosslinked sulfonated polyimide networks as polymer electrolyte membranes in fuel cells". In: *Journal of Polymer Science Part B: Polymer Physics* 43.17, pp. 2370–2379.

Sk, D., P. Pijet, and A. Wanic (1992). "Investigation of polyimides containing naphthalene units: 1. Monomer structure and reaction conditions". In: *Polymer* 33.1, pp. 190–193.

Tse, Y.-L. S., A. M. Herring, K. Kim, and G. A. Voth (2013). "Molecular dynamics simulations of proton transport in 3M and Nafion perfluorosulfonic acid membranes". In: *The Journal of Physical Chemistry C* 117.16, pp. 8079–8091.

Veld, P. J. in't, A. E. Ismail, and G. S. Grest (2007). "Application of Ewald summations to long-range dispersion forces". In: *The Journal of chemical physics* 127.14, p. 144711.

Vettigli, G. (2018). *MiniSom: minimalistic and NumPy-based implementation of the Self Organizing Map*.

Wang, J., X. Yang, Z. Zeng, X. Zhang, X. Zhao, and Z. Wang (2017). "New methods for prediction of elastic constants based on density functional theory combined with machine learning". In: *Computational Materials Science* 138, pp. 135–148.

Ward, L. and C. Wolverton (2017). "Atomistic calculations and materials informatics: A review". In: *Current Opinion in Solid State and Materials Science* 21.3, pp. 167–176.

Watari, T., J. Fang, K. Tanaka, H. Kita, K.-i. Okamoto, and T. Hirano (2004). "Synthesis, water stability and proton conductivity of novel sulfonated polyimides from 4, 4-bis (4-aminophenoxy) biphenyl-3, 3-disulfonic acid". In: *Journal of Membrane Science* 230.1-2, pp. 111–120.

Wheatle, B. K., E. F. Fuentes, N. A. Lynd, and V. Ganesan (2020). "Design of Polymer Blend Electrolytes through a Machine Learning Approach". In: *Macromolecules* 53.21, pp. 9449–9459.

Woo, Y., S. Y. Oh, Y. S. Kang, and B. Jung (2003). "Synthesis and characterization of sulfonated polyimide membranes for direct methanol fuel cell". In: *Journal of Membrane Science* 220.1-2, pp. 31–45.

Wu, S., H. Yamada, Y. Hayashi, M. Zamengo, and R. Yoshida (2020). "Potentials and challenges of polymer informatics: exploiting machine learning for polymer design". In: *arXiv preprint arXiv:2010.07683*.

Xu, P., T. Lu, L. Ju, L. Tian, M. Li, and W. Lu (2021). "Machine Learning Aided Design of Polymer with Targeted Band Gap Based on DFT Computation". In: *The Journal of Physical Chemistry B* 125.2, pp. 601–611.

Yaguchi, K., K. Chen, N. Endo, M. Higa, and K.-i. Okamoto (2010). "Crosslinked membranes of sulfonated polyimides for polymer electrolyte fuel cell applications". In: *Journal of Power Sources* 195.15, pp. 4676–84.

Yan, C., X. Feng, C. Wick, A. Peters, and G. Li (2021). "Machine learning assisted discovery of new thermoset shape memory polymers based on a small training dataset". In: *Polymer* 214, p. 123351.

Yao, H., K. Shi, N. Song, N. Zhang, P. Huo, S. Zhu, Y. Zhang, and S. Guan (2016). "Polymer electrolyte membranes based on cross-linked highly sulfonated co-polyimides". In: *Polymer* 103, pp. 171–179.

Ye, X, H Bai, and W. Ho (2006a). "Synthesis and characterization of new sulfonated polyimides as proton-exchange membranes for fuel cells". In: *Journal of Membrane Science* 279.1-2, pp. 570–7.

Ye, X., H. Bai, and W. W. Ho (2006b). "Synthesis and characterization of new sulfonated polyimides as proton-exchange membranes for fuel cells". In: *Journal of Membrane Science* 279.1-2, pp. 570–577.

Yin, Y., S. Hayashi, O. Yamada, H. Kita, and K.-I. Okamoto (2005a). "Branched/crosslinked sulfonated polyimide membranes for polymer electrolyte fuel cells". In: *Macromolecular rapid communications* 26.9, pp. 696–700.

Yin, Y., S. Hayashi, O. Yamada, H. Kita, and K.-I. Okamoto (2005b). "Branched/Crosslinked Sulfonated Polyimide Membranes for Polymer Electrolyte Fuel Cells". In: *Macromolecular Rapid Communications* 26.9, pp. 696–700. DOI: 10.1002/marc.200500014.

Zawodzinski Jr, T. A., M. Neeman, L. O. Sillerud, and S. Gottesfeld (1991). "Determination of water diffusion coefficients in perfluorosulfonate ionomeric membranes". In: *The Journal of Physical Chemistry* 95.15, pp. 6040–6044.

Zhang, F., N. Li, Z. Cui, and S. Zhang (2008). "Novel acid-base polyimides synthesized from binaphthalene dianhydrie and triphenylamine-containing diamine as proton exchange membranes". In: *Journal of Membrane Science* 314.1-2, pp. 24–32.

Zhang, F., N. Li, S. Zhang, and S. Li (2010). "Ionomers based on multisulfonated perylene dianhydride: Synthesis and properties of water-resistant sulfonated polyimides". In: *Journal of Power Sources* 195.8, pp. 2159–2165.

Zhang, G., G. Yang, S. Li, Q. Shen, H. Wang, Z. Li, Y. Zhou, and W. Ye (2021). "Effects of Hydration and Temperature on the Microstructure and Transport Properties of Nafion Polyelectrolyte Membrane: A Molecular Dynamics Simulation". In: *Membranes* 11.9, p. 695.

Zhang, H. and P. K. Shen (2012). "Recent development of polymer electrolyte membranes for fuel cells". In: *Chemical reviews* 112.5, pp. 2780–2832.

Zhang, Y. and C. Ling (2018). "A strategy to apply machine learning to small datasets in materials science". In: *Npj Computational Materials* 4.1, pp. 1–8.

Zhu, X., H. Pan, Y. Liang, and X. Jian (2008). "Synthesis and properties of novel sulfonated polyimides containing phthalazinone moieties for PEMFC". In: *European polymer journal* 44.11, pp. 3782–3789.