Structural Analysis of Benzene-Methanol Azeotrope: A Molecular Dynamics Approach

Thesis submission towards the partial fullfillment of

Integrated PhD Programme - Master's

By,

Anant Onkar Bhasin

Under the guidance of

Dr. Arnab Mukherjee

Department of Chemistry

IISER Pune, India.

Department of Chemistry

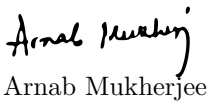
Indian Institute of Science Education and Research

Pune - 411 008

April,2021

Certificate

This is to certify that this project entitled "Structural Analysis Of Benzene-Methanol Azeotrope A Molecular Dynamics Approach" in partial fulfillment of the Integrated PhD program at the Indian Institute of Science Education and Research, Pune represents original research work carried out by Anant Onkar Bhasin at IISER Pune under the supervision of Dr. Arnab Mukherjee, Associate Professor, IISER Pune during the academic year 2020-2021.



Date: 30-04-2021
Pune(MH), India

Arnab Mukherjee
(Thesis Supervisor)

Declaration

I hereby declare that the matter embodied in this report entitled "Structural Analysis of Benzene-Methanol Azeotrope A Molecular Dynamics Approach" are the results of the investigations carried out by me at the Department of Chemistry, IISER Pune, under the supervision of Dr. Arnab Mukherjee, and the same has not been submitted elsewhere for any other purpose. I declare that I have adequately cited and referenced the original sources and I have adhered to all principles of academic honesty

Date: 30-04-2021
Pune(MH), India

Anant Onkar Bhasin
(20182013)

Acknowledgement

I would like to express my deep sense of gratitude to my mentor, **Dr. Arnab Mukherjee** for providing me with an opportunity to work in his lab. He always provided me with his scholarly advice, meticulous scrutiny and scientific approach that helped me to great extent throughout my project. He overlooked multiple mistakes I made, also helped me with providing constant motivation during the project. I want to thank him for his constructive criticism and for aiding me to develop professional behaviour. He helped me to develop not professionally but personally as well.

I would like to express my deepest gratitude towards my mother **Mrs. Sanyogita Bhasin** and to my father **Mr. Jitender Kumar Bhasin** for their unbreakable faith, unconditional love the much-needed criticism throughout the good, bad and worst patches of this journey. I am deeply indebted to my sister and brother **Ms. Yagyesmita Bhasin** and **Mr. Maheshwar Bhasin** for raising me into what I am today. I would like to express my special thanks to my special friend **Ms. Gunjan Singh** for always providing me with the shoulder to lean on to. Had it not been for the support from all of you, I probably would not have achieved this milestone so easily.

My sincere gratitude to my senior **Mr. Amal Vijay** and **Mr. Abhijeet Gupta** for their continuous help and support. Their immense knowledge coupled with timely suggestions has guided me throughout the project. They helped me to understand the basics of the methodologies followed. I would like to thank **Mr. Manjul Yadav** for refreshing tea time discussions covering from science to diets. I want to thank **Dr. Vanitha M.** for answering my questions and helping me wherever needed. Without these people, I might not be able to finish even half of my thesis work. Outside the lab, I would like to thank **Mr. Rahul Rajput** for helping me with this project.

I am grateful to **IISER Pune** for allowing me to pursue my masters in such a vibrant environment and providing every platform to help me grow as a person. I thank **Almighty God** for giving me the required perseverance and strength and surrounding me with all the beautiful people in my life.

Dedicated To....

My beloved family and their constant support

Table of Contents

List of Figures

List of Tables

Abstract

Azeotropes are the most intriguing form of mixtures that shows the unique boiling point. These are laborious systems to deal with when we compare them with the separation of zeotropic mixtures. The structure gained by the azeotrope is still a mystery as experiments can only infer results from data. The simulations explicitly look at the interactions between molecules and can show the structure of such systems. Multiple research groups have studied the behaviour of methanol in benzene and tried predicting the unit structure using spectroscopy techniques. The NMR, IR, Raman and other spectroscopies have been highly beneficial to determine the behaviour of pure compounds and in an azeotropic mixture. In this study, we try to understand the structure attained by this minimum boiling azeotrope of methanol and benzene and the role of temperature in it. For analysis of the system, we performed an agglomerative clustering algorithm and metadynamics simulations. This study shows the temperature role in azeotrope formation. Also, we explored the forces between methanol and benzene that help the system attain a stable structure.

Chapter 1

Introduction

Boiling is a phenomenal process in which the vapour pressure of the liquid reaches the external pressure as the temperature is increased, resulting in the change of phase from liquid to vapour. This is a common process observed in our day-to-day lives. The mixtures follow the same pattern, the component of the mixture boils and exits the system once its boiling temperature is reached, this helps in the process of distillation to separate two liquids but this fails in the case of azeotropic mixtures.[1]

1.1. Azeotropes

Azeotrope has a constant boiling temperature at a fixed mole ratio i.e. azeotropic composition. It can be easily distinguished from the pure component. As the composition changes, the boiling point changes for an azeotrope which is not the case for a pure liquid. The azeotropes are categorized based on the number of constituents as binary and ternary. There are different types of azeotropes known as minimum-boiling azeotropes and maximum boiling azeotropes. Also, azeotropes are differentiated based on phases of components as a homogeneous and non-homogeneous azeotrope. The existence of azeotrope is based upon their deviation from Raoults law [2](deviation of vapour pressure w.r.t. mole fraction).

For an ideal liquid mixture, Raoult's law states that the pressure applied by a component in a mixture which is the partial pressure of the component is equal to the product of mole fraction and pressure of that component, in the pure state. Raoults law equation (1) inherently assumes the ideality of mixture and takes $\gamma = 1$ (activity coefficient) i.e. the interactions of like and unlike molecules as similar. As the interaction between like and unlike molecules becomes different, there is a deviation of activity coefficient from unity causing the formation of non-ideal mixtures.

Ideal Raoult's law can be expressed by equation 1, for simplification we will consider only binary mixture;

$$p_1 = x_1 P_1^0 \quad , \qquad y_1 = \frac{x_1 P_1^0 T}{P_{tot}} \qquad \& \qquad y_2 = \frac{(1-x_1) P_2^0 T}{P_{tot}} \tag{1}$$

where y is the composition of vapour phase; x is the composition of liquid phase; P_o and P_{tot} are the pure component and total pressure. A mixture is an azeotrope if x becomes equal to y for all components.
The entire pressure of the mixture is given as

$$P_{tot} = p_1 + p_2 \tag{2}$$

A positively deviated mixture or if total pressure is higher than ideal makes a minimum boiling azeotrope is formed as less amount of energy is required to convert the system between liquid and vapour state and a negatively deviated makes a maximum boiling azeotrope due to the potential barrier is heightened for the boiling process. Mixtures that dont make up an azeotrope are called zeotropic mixtures (means zeo-boil; tropos-way) in which the components boils individually irrespective of composition at their boiling points. The observation regarding azeotrope is summarized in Ghemling et al[3].

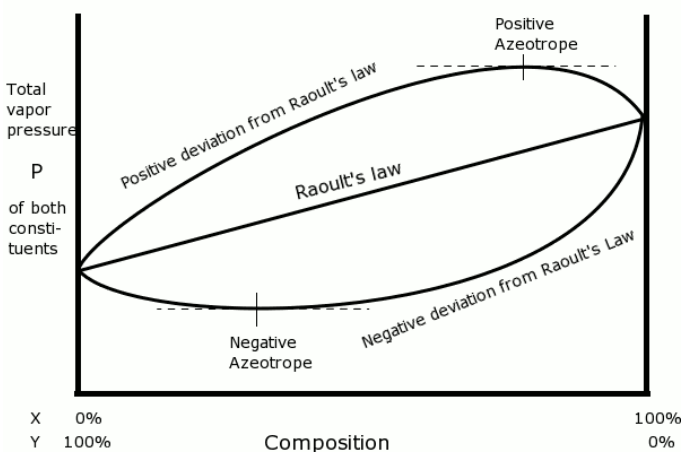


Figure 1.1: Vapour pressure variation with composition. (https://en.wikipedia.org/wiki/Azeotrope)

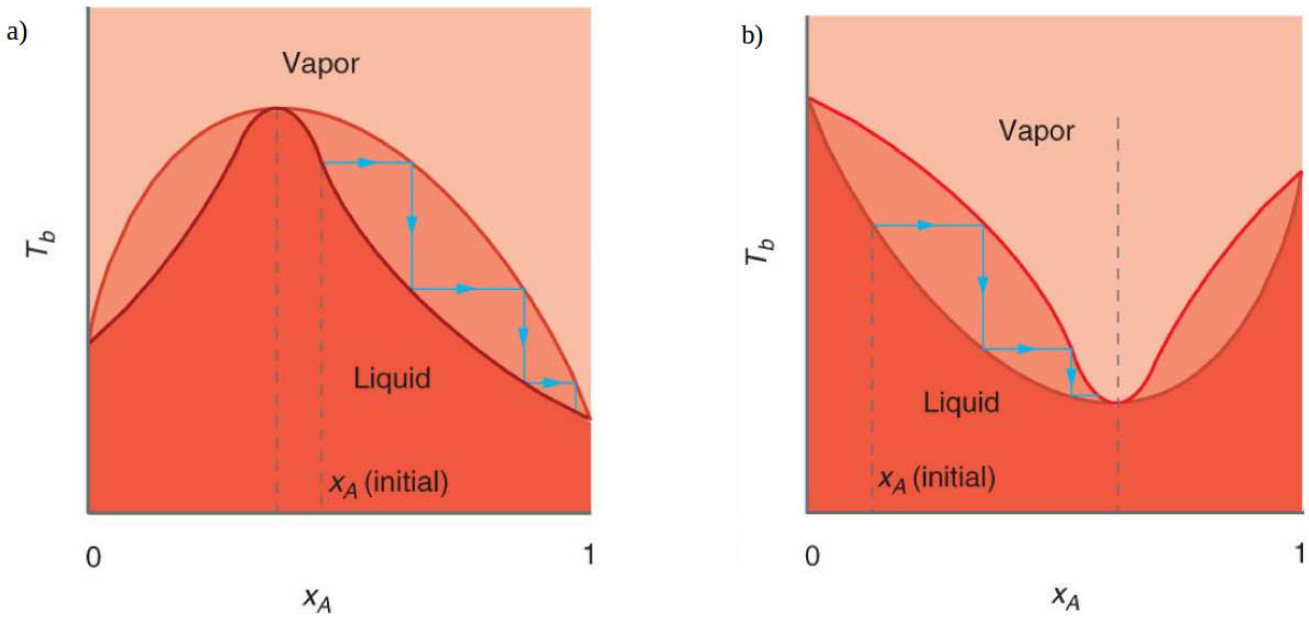
The relative volatility is important to understand the azeotrope formation. The relative volatility, α, is used for comparison of volatilities of components and aids in the distillation process.[4]

$$\alpha = \frac{y_i/x_i}{y_j/x_j} \tag{3}$$

Azeotrope is formed when the compositions are equivalent, or, when relative volatility becomes unity. Any deviation from this value implies that the vapour phase can be more enriched by one component specifically. This value signifies that ordinary distillation techniques used for separation are ineffective in this scenario.

The azeotrope can be divided into three different categories depending upon the interactions between components as minimum boiling azeotropes (positive deviation from Raoults law), maximum boiling azeotrope (negative deviation from Raoults law) and ideal mixture (follows Raoult's law).

The boiling point of liquids and the amount of non-ideality are the two key factors to facilitate the formation of



(a) Maximum boiling azeotrope

(b) Minimum boiling azeotrope

Figure 1.2: Variation of boiling point of mixture with composition. Minimum Boiling azeotrope and Maximum boiling azeotrope. Copyright ©pchemandyou.blogspot.

azeotropes [5, 6]. As a thumb rule for homogeneous mixtures, it is said that if the difference between boiling points of two components $> 30^oC$, the likelihood of formation of the azeotrope is higher [7]. Swietoslawki [8] et al. explored the azeotropes formed by organic molecules and found that the majority of azeotropes formed by these are minimum boiling non-ideal mixtures and contains a polar moiety. The ideal behaviour is seen among the molecules with almost similar structure. Although it is generally the scenario if there is existence of Bancrofts point the mixture still shows an azeotropic behaviour [9].

1.2. Vapour Liquid Equilibrium

The phase diagram of a mixture depends upon the temperature, pressure and composition that are conventionally known as the phase variables. In separation processes, these variables are tailored such that one component of the mixture dominates in a phase. In Vapour- Liquid equilibrium, the component's composition doesn't vary over time. For mixture, Raoult's law is depicted as follows:

$$y_i P = x_i \gamma_i P_i^{sat} T \qquad i = 1, 2, ..., n \tag{4}$$

where T, P and n are pressure, temperature and the number of components in a mixture. If the value of $\gamma > 1$ the mixture is said to exhibit a positive deviation and if $\gamma < 1$ represents the negative deviation from Raoult's law. The amount of deviation from the unity of γ can even explain the type of azeotrope formed homo or hetero. The stronger the attraction between unlike molecules than like ones, the greater is the negative deviation from Raoult's law but, if the converse is true, it leads to positive deviation.

Essentially, the VLE exists when the system tends to oppose any change in free energy i.e.

$$dG = 0 \tag{5}$$

for any change in composition, the chemical potential follows the same equation.

The Wilson[10], Antoine[11], NRTL[12], UNIFAC[13] and UNIQUAC[14] equations are usually employed while calculating the activity coefficient that corresponds well with the relation between γ and compositions. Generally, the use of Henry's law[15] $y_i p = x_i p_i^{sat}$ is done for dilute solutions. The knowledge of VLE and phase diagram is essential for multiple industrial applications and the azeotrope's formation.

As explained, the azeotropes are formed due to the interactions that behave non-ideally, therefore, it is vital to figure out the unit structure of azeotropes formed at the azeotropic composition and to study the role of temperature in that structure. In this study, we will consider a methanol benzene system that exhibits the azeotropic behaviour at 3:2 (methanol: benzene) ratio and boils at 58.3^oC i.e. lower than boiling point of both the components, hence forming a minimum boiling azeotrope.

Previous literature indicates that the methanol can form clusters with the help of hydrogen bonding, with the formula $[CH_3OH]_n$, where n ranges from 1-7 [16–22]. To gain insight on the frequency of hydroxyl group ion-dip infrared spectroscopy is also done on clusters of methanol where n=1-6[23]. The hydrogen bonding is ultimately responsible for the formation of the cluster is also explained by these works of literature.[24] Dimer study done by Del Bene was further studied by Curtius using the minimal basis set[25, 26]. A theoretical study by Matisz et al[27] was based upon ab-initio calculations and DFT to understand the formation of clusters of the methanol present in Benzene.

The milestone towards understanding the methanol cluster size in Benzene was done by Pribble which showed the presence of$[CH_3OH]_4$ is the most prominent[23]. Experimentally, the densities of methanol and benzene at different compositions were calculated by Irving and Simpson[28]. Jalilian et al used implicit techniques such as NMR, IR and FT Raman to illustrate the structure gained by benzene and methanol at azeotropic composition[29]. The techniques used by Jalilian et al doesnt give the explicit structure of the benzene, hence the structured derived was based upon interpretation of peaks and comparison of them with pure benzene and methanol. The structure predicted by Jalilian et al was to satisfy the spectral data only and might not depict the true unit structure. Nonetheless, they observed that methanol is not separated much from each other at the azeotropic composition, since the azeotropic boiling point is much closer to methanols boiling point than benzene.

In this study, we try to find the structure obtained by the azeotrope explicitly using Molecular Dynamics and we study the effect of temperature and composition on the structure. Finally, we will explore the phase space and plot the free energy surfaces for the methanol-benzene azeotropic system and try to find out the unit structure formed and compare it with the structure predicted by Jalilian et al.

1.3. Outline

The thesis follows the folowing section for the analysis of the binary azeotropic system:

- **Methodology**:

  This section explains the various softwares and programs employed in the analysis of azeotrope study. Moreover, this part touches upon the background of various techniques used and explains the choices made for the analysis.

- **Azeotrope Structure Determination**:

  This part shown the results gained from the simulation and the interpretation made from those results. The entire section is divided on the basis of the analysis. The relative terms are explained in between the section whereever necessary.

- **Discussion**:

  This part discussed the important results gained from the simulation and the future direction of the work. This section compares the results gained from simulation with previously reported results.

Bibliography

[1] Atkins P, De Paula J. Atkins physical chemistry. 9th ed. London, England: Oxford University Press; 2010.

[2] M RF. General law of methyl alcohol at room temperature,Comptes rendus. Vol. 104. 1886. p. 14301433.

[3] Gmehling J, Menke J, Krafczyk J, Fischer K. A data bank for azeotropic data status and applications. Fluid Phase Equilib. 1995;103(1):5176.

[4] Kister HZ. Distillation Design. New York, NY: McGraw-Hill Professional; 1992.

[5] Horsley LH. Table of Azeotropes and Nonazeotropes. In: Advances in Chemistry. 1155 Sixteenth Street, N.W. Washington, D. C.: AMERICAN CHEMICAL SOCIETY; 1973. p. 1314.

[6] Tamir E, Tamir A, King MB. Explaining characteristic azeotropic behaviour. Chem Eng Sci. 1981;36(4):76571.

[7] Perry RH, Chilton CH, Green. Perrys chemical engineers handbook. 6th ed. New York, NY: McGraw-Hill; 1984.

[8] Reshe M. Azeotropy and Polyazeotropy. VonW. Swietoslawski. (Aus dem Polnischen). Herausgeg. v.K. Ridgway. Pergamon Press, Oxford-London-New York-Paris 1963. 1. Aufl., 226 S., zahlr. Abb. und Tab., Gzln. 3.10.0. Angew Chem Weinheim Bergstr Ger. 1965;77(8):392392.

[9] Malesinkski,W. . "Azeotropy and Other Theoretical Problems of. Vapor-Liquid Equilibrium". Interscience, New York: 1965.

[10] Ghosh SK, Chopra SJ. Activity coefficients from the Wilson equation. Ind eng chem process des dev. 1975;14(3):3048.

[11] Thomson GW. The Antoine equation for vapor-pressure data. Chem Rev. 1946;38(1):139.

[12] Ball FX, Frst W, Renon H. An NRTL model for representation and prediction of deviation from ideality in electrolyte solutions compared to the models of Chen (1982) and Pitzer (1973). AIChE J. 1985;31(3):3929.

[13] Kikic I, Alessi P, Rasmussen P, Fredenslund A. On the combinatorial part of the UNIFAC and UNIQUAC models. Can J Chem Eng. 1980;58(2):2538.

[14] Demirel Y. Calculation of infinite dilution activity coefficients by the NRTL and UNIQUAC models. Can J Chem Eng. 1990;68(4):697701.

[15] Gerrard W. Henrys Law and Raoults Law. In: Solubility of Gases and Liquids. Boston, MA: Springer US; 1976. p. 2954.

[16] Magini M., Paschina G., Piccaluga G. On the structure of methyl alcohol at room temperature. J. Chem. Phys.77 (1982) 2051

[17] HaughneyM., Ferraio M., McDonald T.R. Molecular-Dynamics simulation of aqueous mixtures: (methanol, acetone and ammoniaMols. Phys. 58 (1986) 849.

[18] Pettitt B.M., Rossky P.Contribution of hydrogen bonding to the structure of liquid methanol.J. Chem. Phys.78 (1983) 7296.

[19] Mammone J.F., Sharma S.K., Nikol M.Raman spectra of methanol and ethanol at pressures up to 100kBarJ. Phys. Chem., 84 (1980) 3130.

[20] Narten A.H., Habenschuss A.,XRD study of some liquid alkanesJ.Chem.Phys. 80 (1984) 1984.

[21] Monatgue D.G., Gibson I.P., Dore J.C.,neutron diffraction study of liq. Iso-propanolMol. Phys. 44 (1981) 1355.

[22] Jogersen W.L., Ibrahim M.I.,Pressure dependence of hydrogenbonding in liquid methanol.J. Am. Chem. Soc.104 (1982) 373.

[23] Pribble R. N., Hagemeister F. C., Zevier T.S., Resonant ion-dip infrared spectroscopy of benzene-(methanol)nclusters n=1-6 J. Chem. Phys.,106 (1997) 2145.

[24] Jonas, J., Akai, J. A. (1977). "Transport processes in compressed liquid methanol". J. Chem.Phy., 66(11), 49464950.

[25] Bene, J. E. Del. (1971). Theoretical Study of Open Chain Dimers and Trimers ContainingCH 3 OH and H 2 O. J.Chem.Phy., 55(9), 46334636.

[26] Curtiss, L. A. (1977). Molecular orbital studies of methanol polymers using a minimal basis set.J. Chem.Phy., 67(3), 11441149.

[27] Matisz G., Kelterer A.M., FabinaW. M. F., Mate S. K., Coordination of methanol clusters to benzene: A computational study. J. Phys. Chem. A., 115 (2011) 10556-10564.

[28] Irving, H. M. N. H., Simpson, R. B. (1972). "The viscosity of mixtures of benzene and methanol". J.Inorg.Nucl.Chem., 34(7), 22412247.

[29] Jalilian, M. R., Tayyari, S. F. (2009). "Spectra and structure of binary azeotropes VI-benzene-methanol." Spectrochimica Acta -Part A: Molecular and Biomolecular Spectroscopy, 73(5), 828832.

Chapter 2

Methodology

2.1. Introduction

Azeotrope formation is a spontaneous process given the conditions for its formation are satisfied such as composition, temperature and pressure. The azeotrope is said to be formed when the composition is the same in both liquid and vapour state and the mixture boils at one temperature. These mixtures are commonly not separable using the common distillation process as they tend to boil at the same temperature irrespective of the component's individual boiling points which is the main differences between zeotropic and azeotropic mixtures. We will focus in this study on binary azeotropic mixture having a lower boiling point than its individual components i.e. an example of the positive azeotrope.

To study the formation of the cluster we employed the agglomerative clustering techniques with the distance as the measure to clustering molecules together. Thereafter different linkage was taken into account to choose the most optimum one which is illustrated in section 2.3 and Appendix A. The primary difference between these linkages is the parameter that these employ to cluster the molecules together.

To study the stable state of the azeotrope we plotted the free energy surfaces for the structure of azeotrope as a function of the coordination number. This chapter is primarily focused on the methodology background and gives an explanation about the choices made to study the system.

2.2. Molecular Dynamics

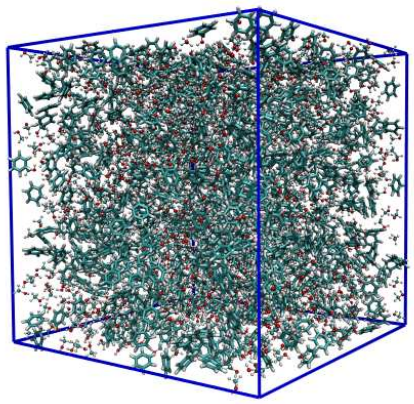
2.2.1. System Preparation

The benzene and methanol was prepared using the AVOGADRO software[1, 2] and were converted into the format i.e. compatible for simulation using Gromacs in-built functions. Different topology generators were used to generate

the parameter files for the OPLS-AA force field[3] and CHARMM force field[4, 5] namely LigParGen[6–8] from the Jorgensen group and CGenFF. The Gaussian software[9] with 6-31G* basis set and ambertools[10, 11] were used to generate the topology for the GAFF force field[12] via antechamber. The topology thus created was further converted to the format which Gromacs understand using Perl script amb2gmx.pl[13]. The system was created using the packing software PACKMOL[14] in various compositions. Gromacs[15–17] was used for the simulation of the systems.

2.2.2. Simulation Details

All the simulations were performed using the Gromacs software. Initially, the steepest descent method[18] was used to minimize the energy of the system such that the potential of the system becomes negative. The simulations are then subjected to 4 ns equilibration having the modified Brendensen thermostat[19] to heat the system at different temperature. The annealing was done in all systems for effective mixing of the system while using 0.001 ps as the time step. Compressibility of the system was taken from previously reported results[20, 21]. The Parrinello- Rahman barostat[22, 23] was used while equilibrating the system with LINCS[24] algorithms to constraint all bonds. The particle- mesh Ewald or PME[25] method was employed for the calculation of electrostatics within the distance cutoff of 1nm. Afterwards, the simulation was submitted for a production run for 100ns and the analysis was done on this 100 ns long trajectory. Almost half the simulations of the methanol- benzene azeotrope was done at the azeotropic composition (3:2 or 1.5:1) with varying the temperature from the room temperature till T=450K. The pure methanol and pure benzene were also simulated using the same parameter but only in the OPLS-AA force field. The visualization software we used is VMD[26] for our system. The analysis was done using the Gromacs in-built functions and Plumed software.



Figrue 2.1: Screenshot of methanol-benzene azeotropic system at 3:2 (methanol: benzene ratio).

2.3. Clustering Analysis

The formation of clusters from the individual yet similar objects present in large data is known as clustering. Clustering is a robust method effectively used in generating information from a large number of datasets. This method uses certain functions based upon the algorithm presented to cluster the objects together, from large data,

that have some amount of resemblance either in pattern or structure[27]. Cluster analysis is a more common practice used in data mining, finding a pattern in climate [28], the grouping of similar genes together, computer vision etc.

There are multiple clustering analysis algorithm present namely K means clustering[29], Self organizing maps[30] and Hierachical clustering[31] , all of these are unsupervised learning algorithms. In the K-means clustering it is a prerequisite to know about the number of clusters in advance. This algorithm employs equivalent centroid in the data set and then performs iteration to cluster together the data points having resemblance in some manner and the centroid changes till the change in succesive steps becomes negligiable or centroid becomes static. The disadvatage that this clustering suffers is that it cant be applied to smaller data sets and it requires a prior knowledge of number of clusters gained from other algorithms commonly. We have opted for hierarichal clustering for our system study due to its highly reliable clustering algorithm and even though it time consuming for larger datasets for our study it can cluster in a reasonable amount of time. The other advatage is it doesnt require any prior knowledge of number of clusters formed ergo it can perform optimally with least human intervention.

2.3.1. Hierarchical Clustering

The Hierarchical clustering is based upon the creation of a binary tree of similar data and it returns the clusters having the potential to provide essential information about cluster ergo showing its different approach than other flat clustering algorithms[32]. Unlike K-means clustering, it doesnt require any prior set number of clusters for clustering analysis. This algorithm combines the nearby clusters until we hit the root of the tree created i.e. all the data is clustered into cluster 0. As depicted in Figure 2.1, show how a sample set can be clustered together using hierarchical clustering. The full data is at the top (A, B, C, D, W, X, Y, Z) as shown in figure 2.1 and the individual clusters are at the bottom as shown at the leaves of the tree. Hierarchical clustering is based upon the idea that similar objects are present nearby and objects with different properties are far apart which is used for our advantage in this case as the molecules which are interacting with each other will be found nearby. Ergo, Hierarchical Clustering is used for the clustering analysis in the methanol-benzene azeotrope system over the entire trajectory.

As depicted in the adjacent figure the data points were all different at the bottom and in the successive steps these are combined based upon the criteria specified in a binary fashion i.e. one-to-one cluster combination. This approach continues until all the data points are present in any one or more clusters. This approach can be halted at any point wherever it seems best.

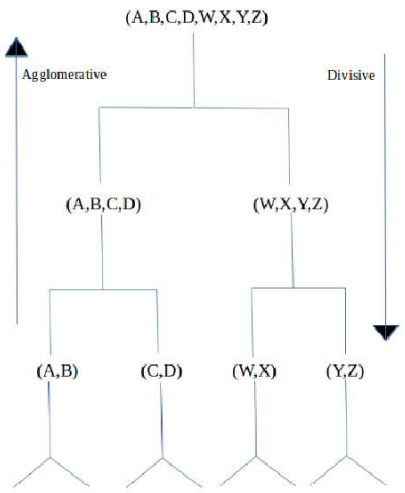


Figure 2.2: Hierarchical Clustering.

After carefully analysing the dendrogram (figure 2.1) it can be seen that since the clustering is done in a binary fashion ergo we can create the following two approaches either going from root to leaves (top-down approach) or from leaves to roots (bottom-up approach). Indeed these approaches divide the

hierarchical approach into two different methods:
(i) Agglomerative Bottom-up approach i.e. starting from all the data points and gradually making the clusters larger.
(ii) Divisive Top-Down approach i.e. starting from the largest cluster possible and dividing it into smaller clusters till the number of clusters becomes equal to the number of data points.

Even though the hierarchical approach suffers the disadvantage of being slow for a huge data set we can still use it for analysis of our system. We will utilise the agglomerative hierarchical clustering (HAC) approach for our analysis. Divisive clustering is more complex than agglomerative clustering.

2.3.2. Agglomerative Clustering

Agglomerative hierarchical clustering or HAC approach utilises the bottom-up approach that means it starts with all the data points as separate entities and then clusters together till a specific condition is met or all the data points are clustered together. This approach takes into account the local orientation without considering the global picture of the system which although a flaw is an advantage for our system. This approach is clustering together the data points based upon distances between them i.e. calculated using Euclidean distance or Manhattan distance. The euclidean distance in 3-D is calculated using the equation

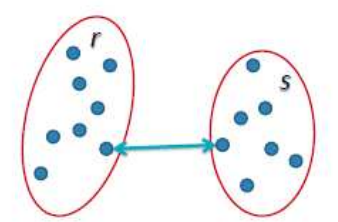
$$d(p,q) = \sqrt{(p_1+q_1)^2 + (p_2+q_2)^2 + (p_3+q_3)^2} \tag{1}$$

where, p1, p2 and p3 and q1, q2 and q3 are Cartesian coordinates of points p and q present on the euclidean plane. The criteria used to specify clusters is distance and that can be beneficial since at large distances it is rather unlikely that the first molecule will interact with the last one. Apart from this we also tend to use different linkage which tells the algorithm which distance to choose for clustering, such as Single linkage, complete or maximum linkage, average linkage or Ward linkage. All these types of linkages differ in the mathematical function that they choose to optimize The following are the types of linkages that are present for agglomerative hierarchical clustering[33]:

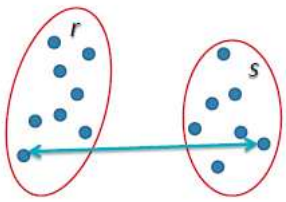
i. Single Linkage: In this linkage the clusters are clubbed together depending upon the shortest distance between the data points in different clusters which are in proximity. Mathematically, it can be express as linkage L between two clusters r and s where the distance is the parameter i.e. minimized and if the distance is less than the threshold distance these clusters are merged together. The disadvantage of this linkage is that it tends to form larger clusters which can be sub divided and provide much more detailed information from data.

ii. Complete or Maximum Linkage: In this linkage the maximum distance between all data points is taken into account and based upon whether it is satisfying the distance threshold it clubs together the two clusters. Mathematically, it is expressed as linkage L between two clusters r and s where the distance is the parameter between the two farthest data points that must be below the distance threshold.

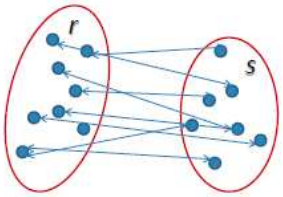
iii. Average Linkage: In this linkage the distance is taken as the average distance between all the data points. Mathematically, it is expressed as linkage L between two clusters r and s where the distance between all points are



$$L(r,s) = \min(D(x_{ri}, x_{sj}))$$



$$L(r,s) = \max(D(x_{ri}, x_{sj}))$$



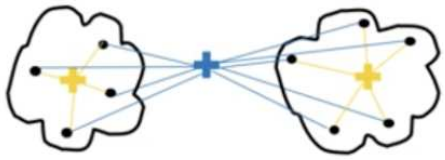
$$L(r,s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} D(x_{ri}, x_{sj})$$

calculated using euclidean distance and then averaged to check does it satisfy the distance threshold. This type of linkage is beneficial but it tends to create different shape of clusters which is not an ideal scenario for our analysis.

iv. Ward Linkage[34]: In this linkage the objective function is variance which we tend to minimize within the cluster. Ergo, the clusters that will finally pair in binary fashion will be based upon the change in variance i.e. lower the change higher the probability of such clusters to pair. The distance d in this linkage is calculated using the following

$$d_{i,j} = d(X_i, X_j) = (||X_i - X_j||)^2 \tag{2}$$

This type of linkage is beneficial for our system analysis since it creates almost similar shape of clusters which is vital for clustering analysis as the system is present in liquid state and it is not possible to see a definite boundary between clusters.



The following methods have been employed and can be seen in Appendix A where complete and average linkages were used at different temperatures and at different cutoff distance or distance threshold. It is interesting to note that the size of cluster varies extremely when the linkages are different at same distance.

2.4. Metadynamics

Metadynamics is a robust technique used to calculate the free energy or other state functions for a simulation such that the systems entire landscape can be gained by increasing the probability of the system attaining a rare event rather quickly. The metadynamics is explained as filling the energy wells with computational sand[35]. It works based on a beforementioned collective variable with the help of which the energy of the entire system can be varied. Ergo,

collective variable in a sense forces the system to navigate through the entire energy landscape hopping from one minimum to another. The various values of the collective variable are attained by the addition of gaussian potential to the true systems potential such that it can visit multiple states without failure and this addition also decreases the probability of the system reaching the same state again[36]. This addition of gaussian hills to the systems potential creates the entire energy landscape until all the state that are plausible are visited at least once. Thereafter the system tends to float in the space and as the entire well is filled with computational sand the system now can vary the collective variable as needed resulting in high fluctuation in the collective variable. The potential added by depositing gaussian potential hills over the entire simulation is:

$$V(\vec{s},t) = \sum_{k\tau} W(k\tau) exp(-\sum_{i=1}^{d} \frac{(s_i - s_i(q(kt)))^2}{2\sigma_i^2}) \tag{3}$$

where τ is the frequency of addition of gaussian, σ is the width of the gaussian hill i and W(k) is the height of the gaussian hill i.e. deposited.

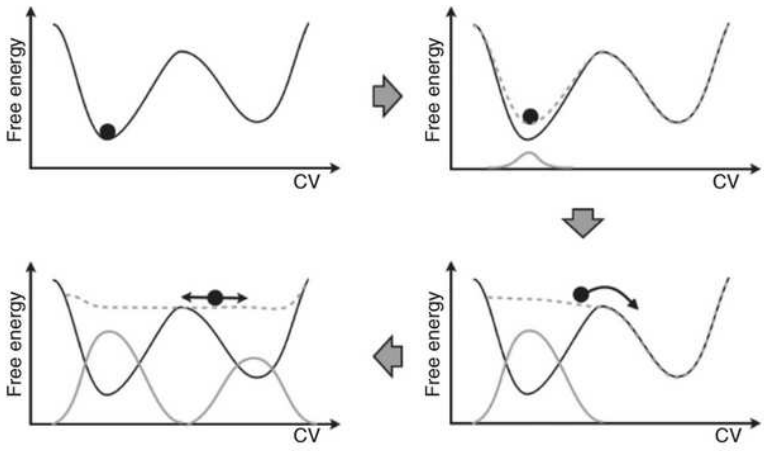


Figure 2.3: Moving from normal MD trajectory to metadynamcis by addition of gaussian plots (grey) that in turn increases the potential of system and alters the free energy landscape (dashed grey line). After filling up of first potential well the system hops to another minima where the again the gaussian hills starts to deposit. In the end the system start to floats over the entire space.

Metadynamics simulation is essential at times where the MD simulations are not enough for instance processes that occur rarely or shifting from one metastable state to another takes a long MD run which might not be feasible. Obtaining the required state after employing the all-atom force field is even more expensive from MD simulation alone in terms of computational. Therefore, metadynamics simulation helps to hop from one metastable state to another based upon the choice of collective variable. The advantages of metadynamics also include that it doesnt require any prior knowledge of energy landscape and it is easier to plot the entire energy-landscape once the system reached the floating state or when the collective variable converges by simply subtracting the amount of gaussian hills added in the potential well[37]. Along with these advantages, the metadynamics approach of accelerating rare events has serious demerits such as the system tends to float around the space once it fills the potential wells and visits all the states. This inherently adds up error while constructing the free energy landscape from the metadynamics simulation which is commonly known as overfilling of the well. As mentioned, it is important to look for the convergence of

collective variable but on multiple occasions i.e. not an easy task. The other task to take into consideration is opting for a decent collective variable for the free energy calculation.

Metadynamics has been employed even before in different manners[38–40]. . As discussed above the metadynamics drawbacks are quite significant and to tackle these the metadynamics is distributed into two different types based upon the implementation of gaussian hills namely ordinary and well-tempered metadynamics. The well-tempered metadynamics is introduced to reduce the overfilling of the well and as a result, reducing the overall error in the calculation of the free energy landscape. In the ordinary metadynamics, the gaussian hill i.e. deposited is of constant height i.e. not the case in well-tempered metadynamics simulation.

2.4.1. Well-Tempered Metadynamics

In well-tempered metadynamics, the deposition of the bias factor gradually drops to decline the error caused by overfilling of the well. This is done by declining the height of the gaussian hills added to the total potential over time[39]. The hill height W(k) is decreased according to the following equation:

$$W(k\tau) = W_0 \; exp(-\frac{V(\vec{s}(q(k\tau)), k\tau)}{k_B \Delta T}) \tag{4}$$

where , W_0 is the first hill height of the gaussian added, ΔT is the parmeter that the user specifies with the unit of temperature and k_B is the boltzmann constant. The convergence of bias is smooth but the compensation of free energy is not complete, as depicted by

$$V_G(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \tag{5}$$

where T is the systems temperature and ΔT is the parameter which if tends to zero becomes MD simulation and when tends to infinity becomes ordinary metadynamics simulation. The bias factor γ is often described in the metadynamics simulation which corresponds to the ratio

$$\gamma = \frac{T + \Delta T}{\Delta T} \tag{6}$$

This should be chosen after extreme consideration as it can influence the efficiency of the metadynamics simulation to overcome a certain barrier.

The well-tempered metadynamics is based upon a previous approach known as the self-healing sampling method[41]. This approach helps the system to spend more time in the phase space of interest rather than simulation all the rare events. The collective variable is chosen and the width of the gaussian hill are the parameters at hand which we can play around with to plot the free energy landscape by mandating the exploration done by the metadynamics simulation. Tuning the T parameter can help the system to overcome the barrier in free energy. The metadynamics simulation is also discouraging the system to visit the same state twice. The system tends to move from the initial state to the final state with help of added potential as gaussian hills which tends to compensate for the free energy

required for this transverse process. When the entire system reaches the floating state or the state where the collective variable varies drastically the free energy can be calculated employing the equation 2.5.

The most essential part is to understand when to halt the metadynamics simulation or how to know the system has reached the convergence. This is non-trivial to see the oscillations done by the collective variable such that we can restrict the overfilling of potential well and reduce the error while plotting the free energy landscape.

2.5. Collective Variable

The collective variable is an indirect way of monitoring the entire process that we are interested in since it is nearly unrealistic to take note of every motion to understand the entire motion of the particle. The choice of collective variable is one of the most important tasks while setting up metadynamics simulation and it can influence the entire analysis. It is the reduction of the dimensionality of the phase space. Ergo, it is paramount to understand the characteristics that the collective variable should exhibit. Firstly, the collective variable will be the one that can significantly distinguish between the different states that the system visits and second, it should be able to accommodate the entire modes exhibited by the system. Lastly, it is essential for the number of collective variables to be less as the dimensionality of the metadynamics simulation increases as soon as we increase the collective variable.

For our system, the optimum collective variable that satisfies all these criteria is coordination between all possible pairs namely Benzene- Benzene, Benzene- Methanol, Methanol- Benzene and Methanol- Methanol. These collective variable are used to plot a 2-D metadynamics landscape. The collective variable is explained below.

2.5.1. Coordination

In structural determination, the metadynamics studies using the coordination number as a collective variable which is not an uncommon pratice[43**?** , 44]. The coordination number is almost self-explanatory i.e. the coordination or the neighbouring molecules that a reference molecule have in the proximity. The cut-off distance is set up using the highest RDF value i.e. the distance between benzene rings which is calculated to be 0.78 nm by setting up the centre of mass of benzene as a reference point and neighbouring group. The following equation is used for the calculation of the coordination number between molecules

$$CN_{ij} = \sum_{i \in A} \sum_{j \in B} s_{ij} \tag{7}$$

Where A and B are the two different groups that are defined while calculation of coordination number i.e. this will calculate the coordination number till the cut-off between molecule i of group A with all the molecules of group B. The function must be smoothly decaying at the cut-off by which we mean that it should not decline to 0 exactly at the next step from 0.78 as this might result in derivatives being non-continuous. In this case, the s_{ij} is called the switching function which has many variants but we are particularly interested in rational switching function i.e. given by the equation:

$$s_{ij} = \frac{1 - (\frac{r_{ij} - d_0}{r_0})^n}{1 - (\frac{r_{ij} - d_0}{r_0})^m} \tag{8}$$

where if the $r_{ij} \leq d_0$ the sij will be 1 whereas for $r_{ij} > d_0$ the value of the switching function is gradually decreasing till

0. In the scenario where the two groups are the same then the coordination between the same molecule is considered as 0 in other words self-interactions are not considered.

The entire analysis is done by considering the centre of masses of all the molecules in both groups. The four different groups will be now between the COM of Benzene- Benzene, COM of Benzene- Methanol, COM of Methanol-Benzene and COM of Methanol- Methanol. The first group here specifies the reference molecules and the second group specifies the other molecule it should take into consideration for the calculation of coordination number.

Bibliography

[1] Avogadro: an open-source molecular builder and visualization tool. Version 1.2.0.

[2] Hanwell M.D., Curtis D. E., Lonie D. C., Vandermeersch T., Zurek E., Hutchison G. R., (2012) Avogadro: An advanced semantic chemical editor, visualization and analysis platform. J. Cheminf. 4, 17.

[3] Severance, D. L., Jorgensen, W. L. (1990). Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform, and Liquid Benzene.J.Am.Chem.Soc., 112(12), 47684774.

[4] Vanommeslaeghe K, MacKerell A. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. Journal of Chemical Information and Modeling. 2012;52(12):3144-3154.

[5] Vanommeslaeghe K, Raman E, MacKerell A. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. Journal of Chemical Information and Modeling. 2012;52(12):3155-3168.

[6] Jorgensen, W. L.; Tirado-Rives,(2005)Potential energy functions for atomic-level simulations of water and organic and biomolecular systems J. Proc. Nat. Acad. Sci. USA, 102, 6665-6670.

[7] Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L.(2017) 1.14*CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase SimulationsJ. Phys. Chem. B, 121 (15), 3864-3870.

[8] Dodda, L. S.;Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. (2017)LigParGen web server: An automatic OPLS-AA parameter generator for organic ligands. Nucleic Acids Research, 45, W1, W331-W336.

[9] Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., Cheeseman J. R., Scalmani G., Barone V., Petersson G. A., Nakatsuji H., Li X., Caricato M., Marenich A. V., Bloino J., Janesko B. G., Gomperts R., Mennucci B., Hratchian H. P., Ortiz J. V., Izmaylov A. F., Sonnenberg J. L., Williams-Young D., Ding F., Lipparini F., Egidi F., Goings J., Peng B., Petrone A., Henderson T., Ranasinghe D., Zakrzewski V. G., Gao J., Rega N., Zheng G., Liang W., Hada M., Ehara M., Toyota K., Fukuda R., Hasegawa J., Ishida M., Nakajima T., Honda Y., Kitao O., Nakai H., Vreven T., Throssell K., Montgomery J. A. Jr., Peralta J. E., Ogliaro F., Bearpark M. J., Heyd J. J., Brothers E. N., Kudin K. N., Staroverov V. N., Keith T. A., Kobayashi R., Normand J., Raghavachari K., Rendell A. P., Burant J. C., Iyengar S. S., Tomasi J., Cossi M., Millam J. M., Klene M., Adamo C., Cammi R., Ochterski J. W., Martin R. L., Morokuma K., Farkas O., Foresman J. B. and Fox D. J., Gaussian 03, 2016.

[10] Cornell W, Cieplak P, Bayly C, Kollman P. Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. Journal of the American Chemical Society. 1993;115(21):9620-9631.

[11] Case D, Cheatham T, Darden T, Gohlke H, Luo R, Merz K et al. The Amber biomolecular simulation programs. Journal of Computational Chemistry. 2005;26(16):1668-1688.

[12] Wang J, Wolf R, Caldwell J, Kollman P, Case D. Development and testing of a general amber force field. Journal of Computational Chemistry. 2004;25(9):1157-1174.

[13] Sorin E, Pande V. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. Biophysical Journal. 2005;88(4):2472-2493.

[14] L. Martnez, R. Andrade, E. G. Birgin, J. M. Martnez.(2009) Packmol: A package for building initial configurations for molecular dynamics simulations.J. Comp. Chem., 30(13):2157-2164

[15] Berendsen H, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Computer Physics Communications. 1995;91(1-3):43-56.

[16] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark A, Berendsen H. GROMACS: Fast, flexible, and free. Journal of Computational Chemistry. 2005;26(16):1701-1718.

[17] van Gunsteren W, Berendsen H. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. Angewandte Chemie International Edition in English. 1990;29(9):992-1023.

[18] Greenwell y., E.(1994)Numerical receptors in FORTRAN-The art of scientific computing 2nd ed.SIAM Rev.,36,149.

[19] BussiG., DonadioD., and Parrinello M. (2007), Canonical sampling through velocity rescaling,J. Chem. Phys.,126014101.

[20] Marcus Y. ,(1998) The properties of solvent, Wiley

[21] CalemanC.MaarenP.J.V. and HongM.and Jochen S. Hub and CostaL.T.and SpoelD.V.D.(2012). Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant, J. Chem. Theory Comput.8, 61-7.

[22] Nos S. and Klein M. L., (1983) Constant pressure molecular dynamics for molecular systems,Mol. Phys.,5010551076.

[23] ParrinelloM. andRahmanA.,(1981) Polymorphic transitions in single crystals: A new molecular dynamics method,J. Appl. Phys.,5271827190.

[24] Hess B., Bekker H., Berendsen H. J. C., Fraaije J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. J. Comp. Chem., 18(12), 14631472

[25] Darden T., York D., PedersenL. (1993). "Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems."J.Chem.Phy., 98(12), 1008910092.

[26] Humphrey W., Dalke A., Schulten, K., (1996) VMD-Visual molecular dynamics, J. Molec. Graphics 14.1 33-38.

[27] Sasirekha K. and Baby P. "Agglomerative Hierarichal Clustering Algorithm", IJSRP 2013, 3, 3

[28] Attributed datasets. 2021 [cited 19 April 2021]. Available from: https://www.ecse.rpi.edu/homepages/cvrl/database/AttributeDataset.h

[29] Na S, Xumin L, Yong G. Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm. 2010 Third International Symposium on Intelligent Information Technology and Security Informatics. 2010.

[30] ACADEMIC PRESS, INC. Analytical Chemistry. 1973;45(12):1010A-1010A.

[31] Roux M. A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms. Journal of Classification. 2018;35(2):345-366.

[32] elsen F. Introduction to HPC with MPI for Data Science.

[33] Hierarchical Clustering [Internet]. Saedsayad.com. 2021 [cited 19 April 2021]. Available from: http://www.saedsayad.com/clustering hierarchical.htm

[34] Ward JH. Hierarchical Grouping to Optimize an Objective Function". Journal of the American Statistical Association. 1963;58:236244.

[35] Grs-sim.de. [cited 2021 Apr 20]. Available from: http://www.grs-sim.de/cms/upload/Carloni/Presentations/Marinelli.ppt

[36] Bussi G, Laio A. Using metadynamics to explore complex free-energy landscapes. Nat Rev Phys. 2020;2(4):20012.

[37] Barducci A, Bonomi M, Parrinello M. Metadynamics: Metadynamics. Wiley Interdiscip Rev Comput Mol Sci. 2011;1(5):82643.

[38] Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci U S A. 2002;99(20):125626.

[39] Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett. 2008;100(2):020603.

[40] Laio A, Rodriguez-Fortea A, Gervasio FL, Ceccarelli M, Parrinello M. Assessing the accuracy of metadynamics. J Phys Chem B. 2005;109(14):671421.

[41] Marsili S, Barducci A, Chelli R, Procacci P, Schettino V. Self-healing umbrella sampling: a non-equilibrium approach for quantitative free energy calculations. J Phys Chem B. 2006;110(29):140113.

[42] Fiorin G, Pastore A, Carloni P, Parrinello M. Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. Biophys J. 2006;91(8):276877.

[43] Ikeda T, Hirata M, Kimura T. Hydration structure of Y3+ and La3+ compared: an application of metadynamics. J Chem Phys. 2005;122(24):244507.

[44] Fiorin G, Klein ML, Hnin J. Using collective variables to drive molecular dynamics simulations. Mol Phys. 2013;111(2223):334562.

Chapter 3

Azeotropic Structure Determination

3.1. Introduction

Methanol and benzene shows a single azeotropic point on the phase diagram at 3:2 (methanol: Benzene). This binary mixture signifies the attraction between a non-polar and polar molecule where the van der Waals attraction plays an important role. The temperature plays an important role in the azeotrope formation which can aid the system to overcome some specific barriers over the entire simulation. This study incorporates the effect of temperature over the azeotropic composition qualitatively. This structure exploration study is done using the OPLS-AA force field, amber and CHARMM general force field to comprehend and hand-pick the optimum force field.

3.2. Structure of Study

The simluations were first validated using the density as a parameter and analysis was done primarily using RDF and corrdination number. The findings were compared with previously reported results on the benzene and methanol azeotrope done using Monte-Carlo simulations. The analysis of hydrogen bonding between methanol was explored and the unconventional hydrogen bonds such as between two benzene rings were found using the angle and distance analysis. Along with this we tried to gain the exactboiling point of the azetrope with OPLS-AA atom which was higher than expected for which we simulated using Amber and CgenFF force fields. As a result of no significant change between the RDF, coordination number and Boiling ponits we used OPLS-AA force field which is frequently used for liquid simulation (except protein systems). To study the uniquesness of this azeotrope we simulated different systems as well as pure methanol and benzene. This clearly showed qualitatively that the methanol and benzene is an exceptional case.

Later we incorporated the machine learning technique known as agglomerative clustering with ward linkage for finding the clusters based upon distance. Moreover, we used metadynamics simulation using coordination number as a collective variable to get an insight into the structure.

3.3. Simulation Validation

Every simulation's density was analyzed and compared with experimental values reported by Irving et al [1]. These values are found to be in the acceptable range with the experimental data.

Temperature Mole-fraction	300K	310K	315K	320K
0.75:1 (0.43:0.57)	847.7	835.2	828.2	821.8
1:1 (0.5:0.5)	841.4	829.3	822.7	815.8
1.5:1 (0.6:0.4)	832.2	819.0	812.8	805.9

At Room Temperature					
Composition (Me:Bz)	1:4	0.33:1	1:2	2:1	4:1
Density (Kg m-3)	818.91	822.15	835.069	859.517	866.245

Table 3.1: Density (Kg m-3) values at different temperature and composition.

The radial distribution function (RDF) was plotted using the gromacs inbuilt analysis function for the systems at room temperature and different compositions. The Centre of mass of the molecules were taken to construct the RDF plot.

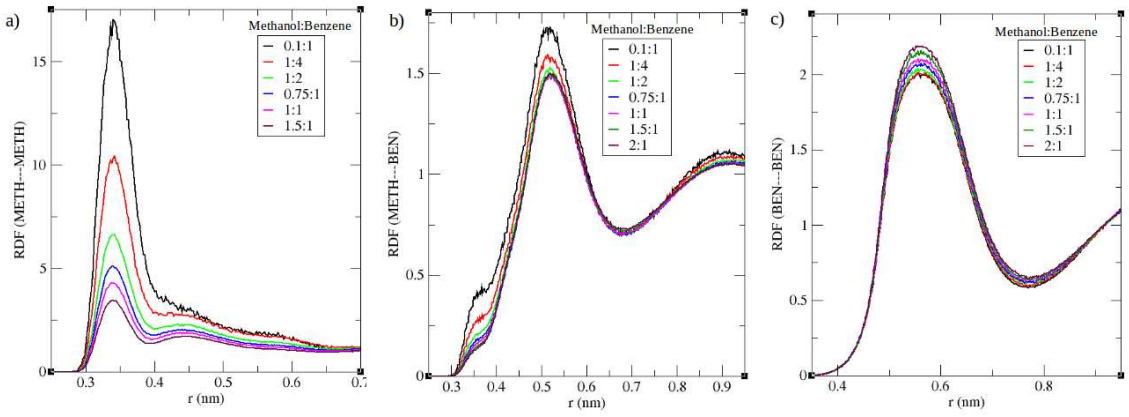


Figure 3.1: Radial distribution function between COMs of (a) Methanol-Methanol (b) Methanol-Benzene (c) Benzene-Benzene. T=300K. Composition (Methanol:Benzene)

The RDF plot between Methanol-Methanol shows a peak around 0.34 nm, depicting the first coordination sphere was found of methanol relatively more closely compared to bulk i.e. the case for methanol-benzene at 0.56nm and Benzene-Benzene at 0.598nm. The RDF goes to unity in all the plots as the bulk density becomes equivalent to local density. In figure 3.1 a, the peak decreases as the number of methanol increases. The plot for Benzene as reference and methanol is deliberately omitted as it was equivalent to figure 3.1 b.

We further looked at the coordination number at three different compositions namely 0.75:1, 1:1 and 1.5:1 (Azeotropic composition). The number of methanol in the proximity of reference methanol is found to be increasing as the methanol concentration increases, which was expected. The observed can be explained as the number of methanol increase but the density increases too ergo, there is a decrease in RDF function even though the number of methanol increases in the system.

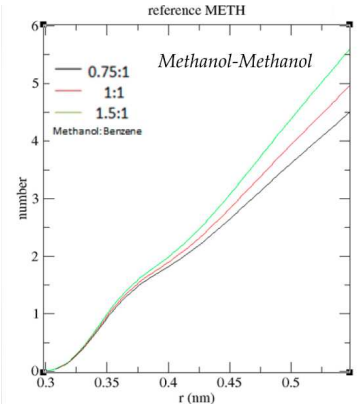


Figure 3.2: Coordination number of Methanol- Benzene system.

Moreover, the simulation carried by MD here was compared with a similar system but using monte Carlo simulation which was previously reported [2]. The RDF of the hydroxyl group was plotted and found to be similar to the reported result. Similarly, the RDF of plotted between oxygen from one group and another hydroxyl proton also gave a similar result. In figure 3.3 b, the nearby hydroxyl group was found in approximately hydrogen bond length. This encourages us to look for the hydrogen bond between methanols.

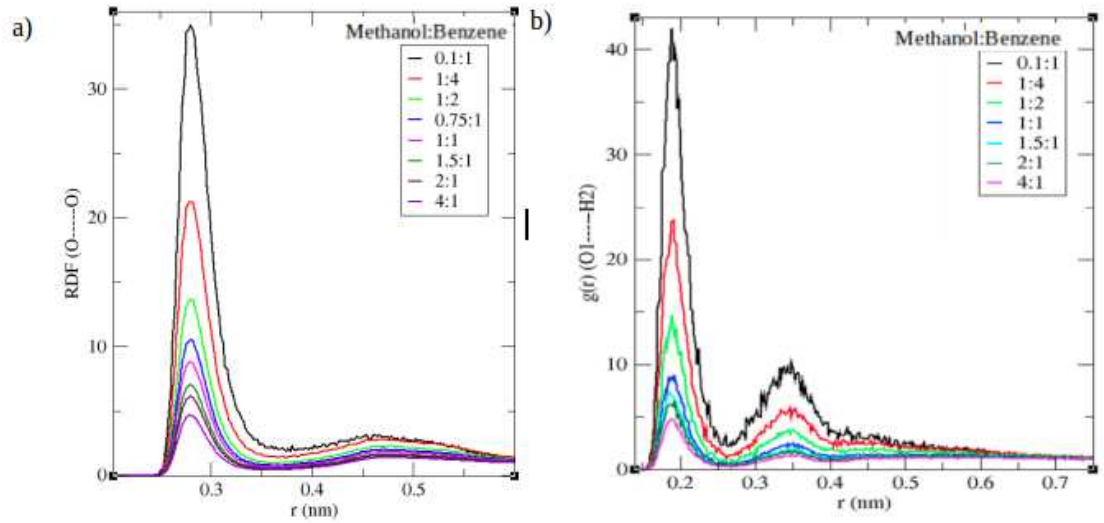


Figure 3.3: (a)RDF between oxygen of hydroxyl group. (b) RDF between oxygen and hydrogen of different methanols.

Coordination number analysis was done between the two hydroxyl groups which were found to be similar to the reported result. As depicted in figure 3.4, the coordination number between two methanol group in the first coordination number was found to be 1 at the distance of the first peak in figure 3.3 b.

The coordination number between all four possible combinations was plotted, see figure 3.5, at all different compositions and found that it was following the same trend as reported by previous studies. The minor difference in the

plot for methanol-methanol in figure 3.5 can be attributed to the fact that methanol is smaller than ethanol molecule.

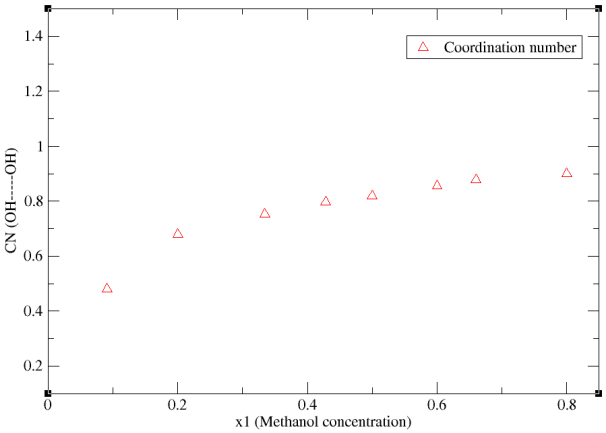


Figure 3.4: Variation of coordination number of hydroxyl groups between methanols.

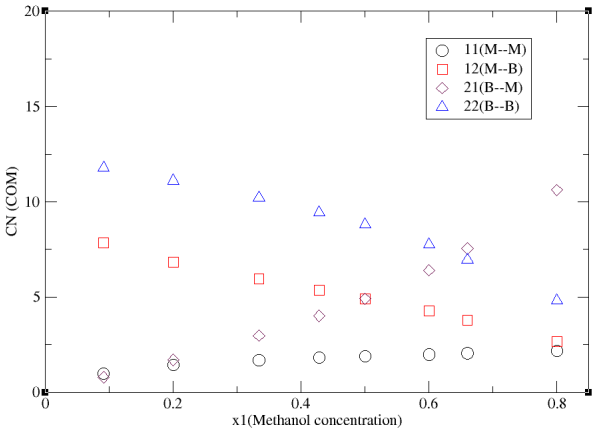


Figure 3.5: Variation of coordination number of all groups in the system using Centre of masses of molecules.

From figure 3.5, the methanol in proximity is found to be 2, which is consistent with the prediction made by Pribble et al[3] that methanol tetramer cyclic structure is most stable in benzene. Previously, Provencal et al[4] showed that methanol tends to form a cyclic structure rather than linear.

3.4. Hydrogen Bonding

The analysis of the number of hydrogen bonds depends upon two factors distance between donor and acceptor as well as the angle between these groups. The number of hydrogen bond between methanols can be found using the conventional definition of hydrogen bonding but we have to use a different approach for finding the unconventional hydrogen bond to verify the presence of this bonding between benzene and methanol.

Table 3.2 shows the number of hydrogen bond between methanols at azeotropic composition at different temperatures. As expected the number of hydrogen bond decreases with an increase in temperature where the total number of methanol is 1200.

Temperature	300K	310K	315K	320K	325K	330K	335K	340K	345K	355K	365K	375K
Number of Hydrogen bonds	981.8	950.9	934.8	918.1	901.2	884.4	867.0	849.8	825.1	795.5	757.7	720.6

Table 3.2: Variation of number of hydrogen bonds of azeotropic composition at different temperature.

The total number of hydrogen bond can be greater than the value shown in the table due to the highly constrained algorithm used by gromacs, especially in the case of angle, for an interaction to be considered as hydrogen bonding. The trend seen in this table is the key feature that satisfies the expected result. This doesnt essentially mean that all methanol is hydrogen-bonded as even in pure methanol some lone methanol molecules are not hydrogen-bonded to any other methanol.

In section 3.3, we tested the coordination number of methanol molecules with other methanols is 2 and hence it becomes essential to see whether they form hydrogen bonds. As shown in figure 3.6, the average number of hydrogen bond methanol makes is 2 which further satisfy the reported result by Pribble et al.

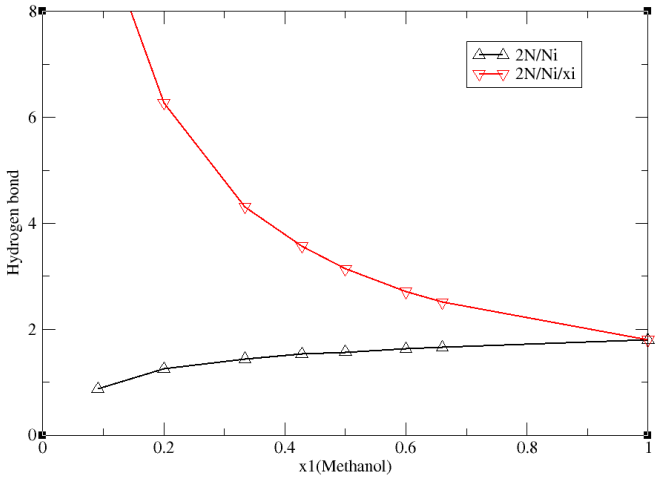


Figure 3.6: Average number of hydrogen bonds between methanols at different compositions per methanol molecules and scaled by methanol mole fraction.
N= Number of hydrogen bonds
Ni= Number of methanol molecules
xi= Mole fraction of methanol

3.4.1. Determination of unconventional hydrogen bonding

Hydrogen bonding is one of the most important factors binding the unit structure together. The methanol hydrogen bonding has a conventional form of acceptor and donor i.e. not the case if we look at benzene between -cloud and other protons in the system. The unconventional hydrogen bonding presence can be judges on two factors, first the distance between protons and Benzene and second, the angle between the acceptor and donor.

In figure 3.7, it can be seen that the shortest distance between the Benzene ring with other molecules is approximately similar over the three chosen compositions and is almost the distance of the hydrogen bond. Although this gives a hint of the presence of some van-der Waals bonding between two molecules, it is vital to check for the angle between these molecules.

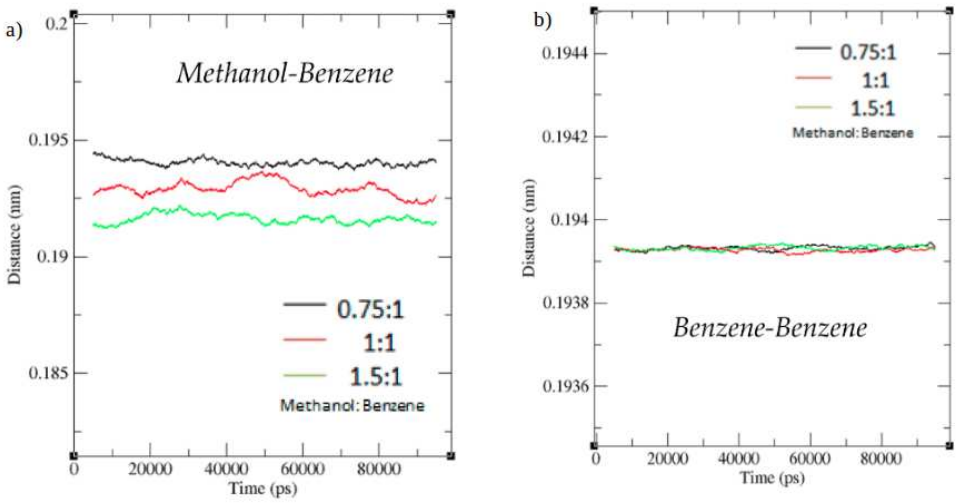


Figure 3.7: (a) Variation of shortest distance between Methanol and Benzene with composition (b) Variation of shortest distance between Benzene and Benzene with composition.

The angle between benzene rings is studied at different compositions. In figure 3.8, the angle between benzene at different compositions is analyzed over the entire 100ns trajectory. In the pure Benzene system, the stacked conformation of benzene rings and T-shape conformations are both present and almost similar amount. In the presence of methanol, the Benzene tends to stay in parallel conformation than T-shaped which can be seen by the increase in peaks as the methanol concentration increases. This result is consistent with the experimentally proposed structure by Jalilian et al [5], which shows the presence of stacked conformation. This angle analysis also confirms the presence of van der Waals interaction between benzene molecules.

Along with these, the angle variation was analysed at different compositions which are depicted in figure 3.9. The analysis was done on the 100ns trajectory at the azeotropic composition 3:2 (1.5:1) (Methanol: Benzene). It can be seen that even though the temperature is increased from room temperature there is no significant change in the plot.

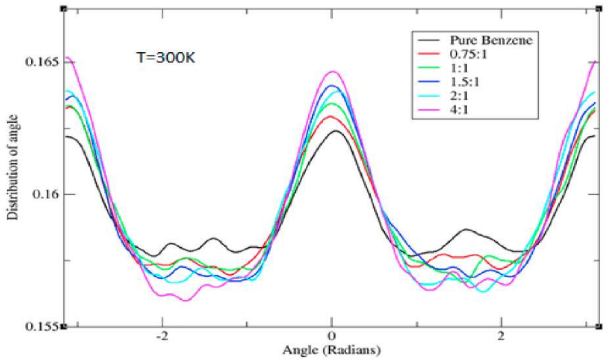


Figure 3.8: Angle analysis at different compositions of Methnaol-Benzene mixture. Comparision of angles with pure Benzene.

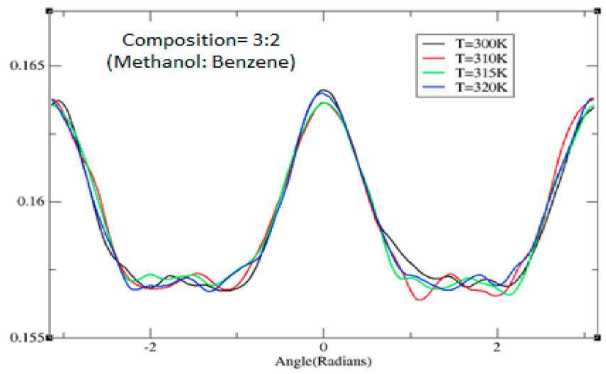


Figure 3.9: Angle analysis between benzene rings at different temperature of azeotropic composition.

Alternatively, this unconventional bonding may be present in between the methanol proton and benzenes -cloud, which is suggested by the distance analysis. Similarly, we plotted the angle analysis for the OH vector and normal vector to benzene. The distance between hydroxyl proton and benzene is shown in figure 3.10 which is in the length

of 1.5-2.5 [6]. The angle analysis as shown in figure 3.11 and 3.12, gives an explicit reason to believe that the bonding is present between hydroxyl proton and benzene. This angle analysis gives a contradictory result to the predicted unit structure by Jalilian et al[5].

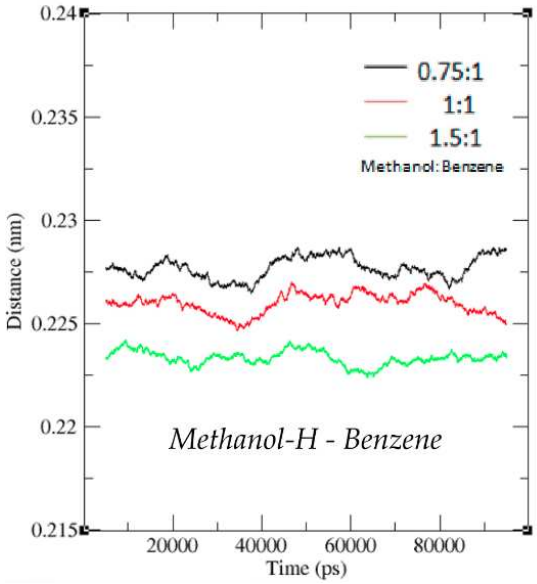


Figure 3.10: Shortest distance between hydroxyl proton and Benzene.

The angle and distance are the two most important factors to determine the hydrogen bonding presence. From figures 3.7-3.12 it is clear that there is van-der Waals attraction is present between methanol and benzene in the form of hydrogen bonding between benzene molecules and benzene-methanol as well.

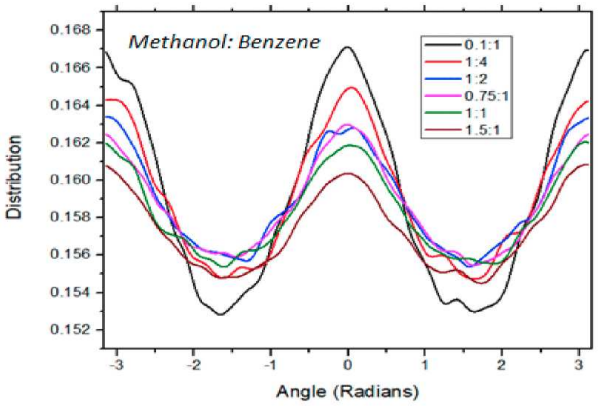


Figure 3.11: Variation of angle at different compositions between O-H vector and benzene normal.

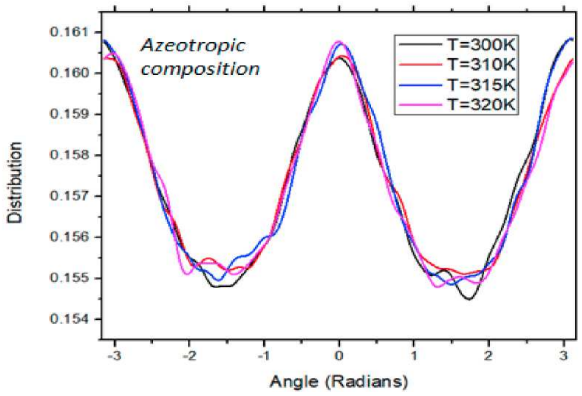


Figure 3.12: Variation of angle at different temperature of azeotropic composition between O-H vector and benzene normal.

3.5. Determination of Boiling points

As previously stated, it is vital to understand the role of temperature as it assists the system to form the azeotropic unit structure. As shown in figure 3.13, the boiling point of azeotrope is 443K, which is higher than the experimental result. The system simulation was again done using amber and Cgenff force field and we observed the systems didnt

boil till 440K. OPLS-AA force field is used for qualitative analysis of the system as it is the most commonly used force field for liquid systems except protein.

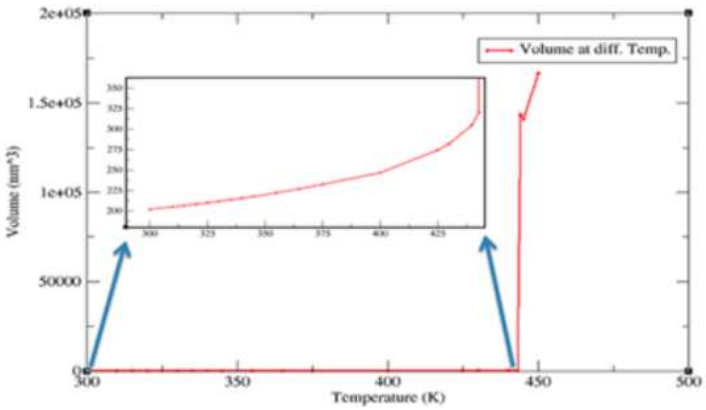


Figure 3.13: Boiling point of azeotropic system using the OPLS-AA force field

The boiling point determination of methanol and benzene is also done to look at whether the force field can distinguish between the pure system and azeotropic ratio. Figure 3.14 a and b shows the boiling point of methanol to be 380K and for benzene, it is 494K, showing that the force field is overestimating the interaction between methanol and benzene molecules but we assume this wont affect the qualitative study of unit-structure attained by the azeotrope at the boiling point.

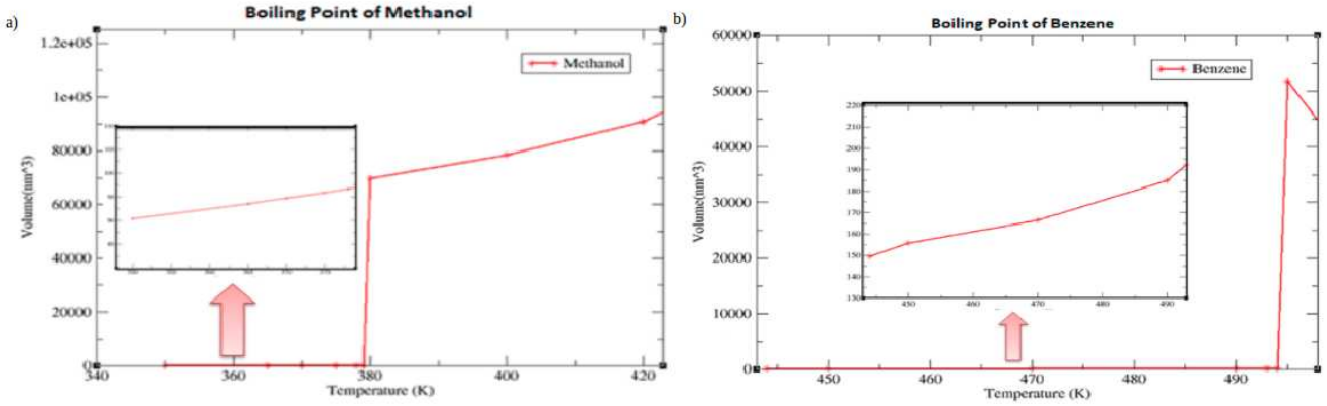


Figure 3.14: Boiling point of pure systems (a) Methanol and (b) Benzene

3.6.Agglomerative Clustering Analysis

Agglomerative clustering analysis was done over the entire trajectory of 100ns. The Centre of masses was chosen as the points to look for clusters taking the linkage as ward as it will create the clusters of almost similar structures. The distance parameter was taken to look for the clusters in the trajectory.

The initial distance was chosen as 0.68nm i.e. the value of the first peak between methanol and benzene as shown in figure 3.1 b. The distance was explored approximately till the half of the box i.e. 2.28nm. Figure 3.15 shows the number of clusters formed having the clusters of ratio 3:2 at five different temperatures.

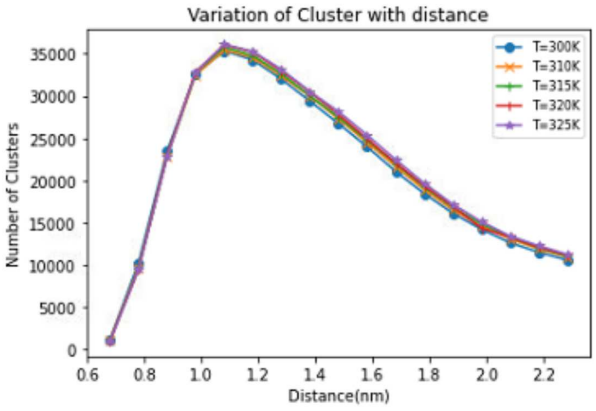


Figure 3.15: Number of desired clusters variation with temperature.

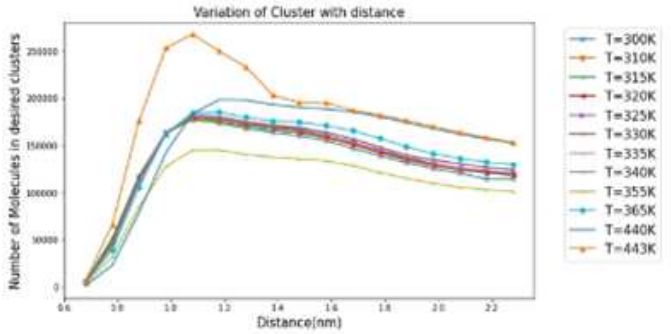


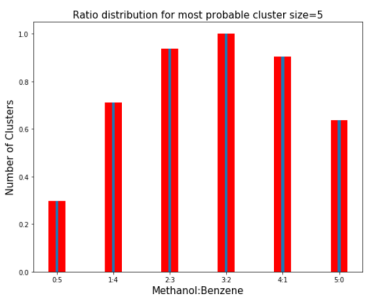
Figure 3.16: Clustering analysis for choosing the most optimum distance for agglomerative clustering.

Figure 3.16 illustrates the distance at which the clusters formed have the maximum of 3:2 ratio and is found to be 1.08 nm. Figure 3.16 gives an unexpected shift peak near the boiling point at 443K, which further supports the fact that temperature helps to overcome the barrier required for the formation of azeotropic structure at the azeotropic composition. To further look at this, the analysis was done at 1.08nm to verify this trend.

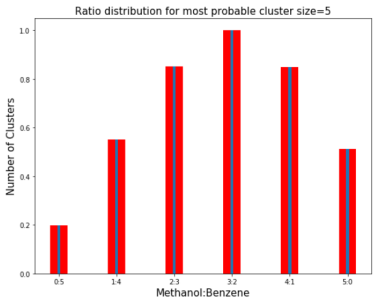
Moreover, it became essential to understand the statistical distribution of ratio present in the most probable cluster at 1.08nm such that we can confirm the most probable ratio is 3:2. Table 3.3 shows the distribution of the ratio of molecules with temperature. The graphs show that the 3:2 is the most dominant peak in all the temperatures and the role of temperature is depicted in all the plots.

There is some significant trend that can be observed from table 3.3. Firstly, the 3:2 ratio is the most significant peak over all the temperature range which becomes progressively exclusive as the temperature is increased since the nearby ratios diminish gradually. An increase in percentage difference between nearby peaks shows explicitly this trend and explains the role of temperature in the formation of the azeotrope. Secondly, it can be seen that the clusters which contain pure benzene and pure methanol decreases in number with temperature which shows that benzene and methanol tend to move far from each other. The decrease in the number of clusters of pure benzene is more than pure methanol. This distance created among the benzene results in the decrease of the boiling point of benzene i.e. higher than methanol. This result is similar to the previously reported study of the azeotrope system by Jalilian et al.

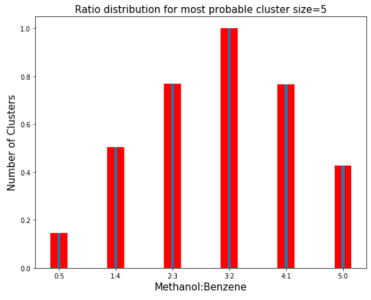
Table 3.3 shows that the temperature is exhibiting a vital role to overcome the barrier for the formation of azeotropic unit structure. Other clusters are becoming unlikely to form as the temperature reaches closer to the boiling point of azeotrope for the simulation.



(a) T=300K, 6.2%, Highest Peak= 32320 clusters.



(b) T=340K, 14.7%, Highest Peak= 34410 clusters.



(c) T=440K, 23.1%, Highest Peak= 44059 clusters.

Table 3.3: Variation of ratio of molecules at most probable cluster with cutoff 1.08nm. T= temp. Percentage explains the difference between azeotropic composition and nearby peaks (2:3 and 3:2).

Figure 3.17 illustrates the variation of the number of molecules in the desired cluster with the increase of temperature till boiling point i.e. T=443K. Figure 3.17 shows an 60% increase in azeotropic clusters near boiling point. This explicitly shows the role of temperature in the formation of azeotropic composition.

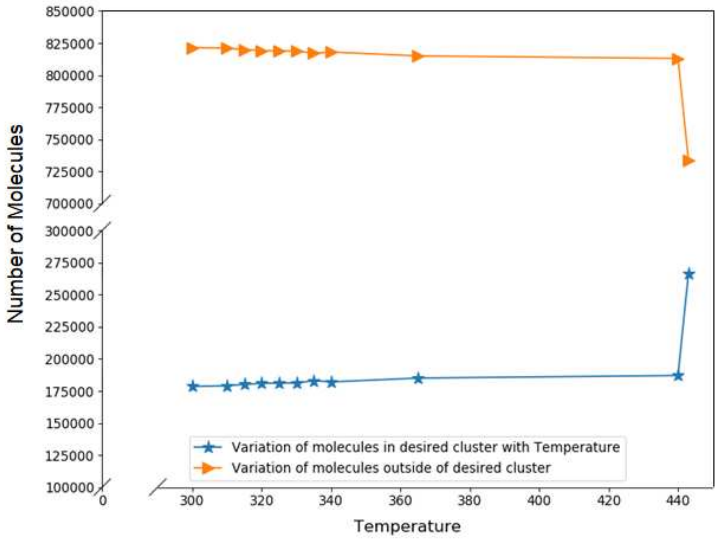


Figure 3.17: Clusters of 3:2 ratio at different temperature at 1.08nm.

Clustering analysis was compared with the other ratios as well. As depicted in figure 3.18, this was done at T=443K where the clusters at 3:2 ratio are found to be maximum. Further explaining the role of temperature in the formation of azeotropic structure. The difference between the nearby ratios 2:3 and 4:1 is seen to be increasing as the temperature is increased. This explicitly shows the key role of temperature in overcoming the thermodynamic barrier to gain an azeotropic structure. The difference between the two peaks was found to be 23.7% and this allows us to make a reasonable assumption that during the boiling the heat supplied which is latent heat is also consumed to overcome the barrier present in the formation of unit-structure of methanol-benzene azeotrope such that it maximizes the number of cluster of 3:2 ratio.

This entire cluster analysis was done using the ward linkage as it is based upon decreasing the variance of the

different clusters that in turn results in similar size clusters formation. This is necessary as using other linkages can results in some unrealistic clusters as shown in chapter 2.

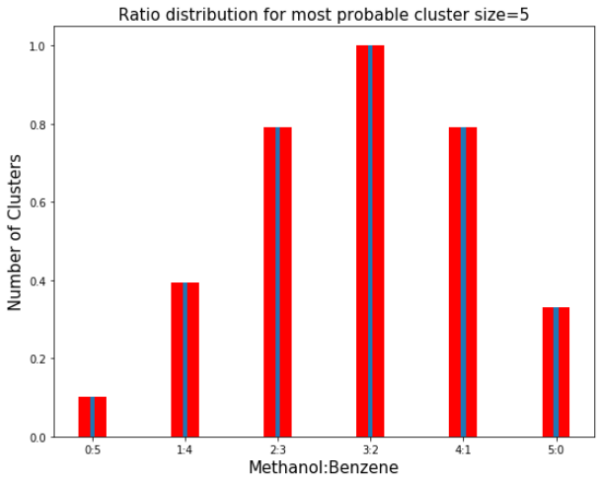
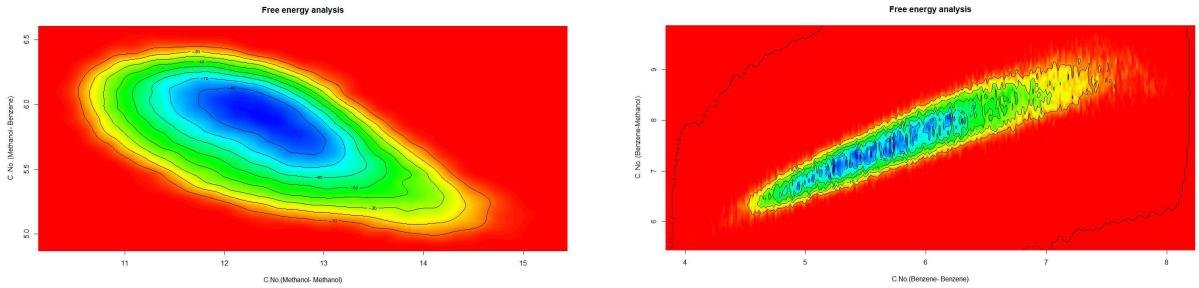


Figure 3.18: Comparision of different ratio clusters over the trajectory of T=443K. Highest Peak= 44059

3.7. Metadynamics

The well-tempered metadynamics simulation was done using the coordination number as a collective variable such that we can simulate the rare events and increase the probability of the system to reach such an event by the addition of potential to the system. The simulation is done at various temperatures such as T=330K, 350K, 440K and 443K.

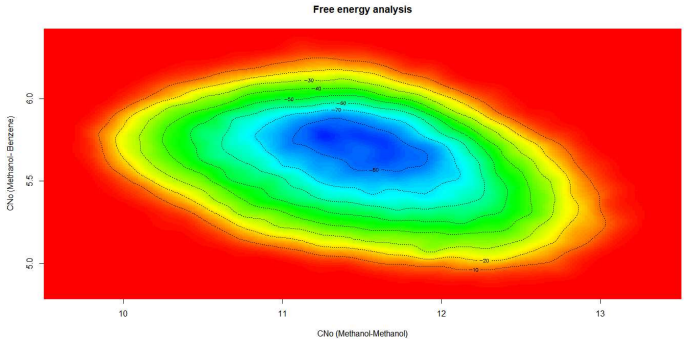
It is observed at T=330K the simulation is showing minima at ratio 1:2 (Benzene: Methanol) for taking methanols COM as a reference, while if we take benzene as COM the ratio is 4:5 (Benzene: Ratio). As soon as the temperature is increased the free energy minima is shown to be around 3:2 (Methanol: Benzene) for both the references Benzene or Methanols COM i.e. observed around T=440K. A similar trend is seen around T=443K but the barrier for the vapour phase is too low that the system goes in the gaseous state in some nanosecond simulation.



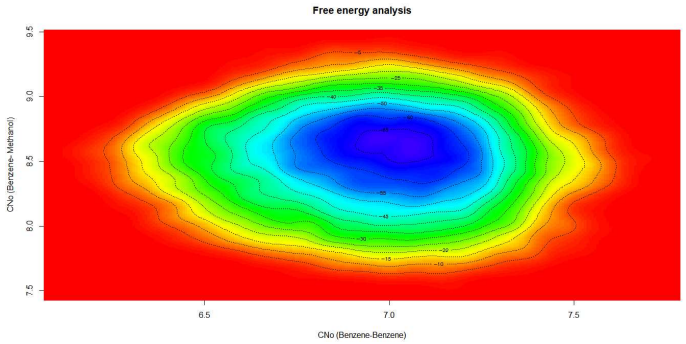
(a) Methanol as reference

(b) Benzene as reference

Figure 3.19: Metadynamics at T=330K

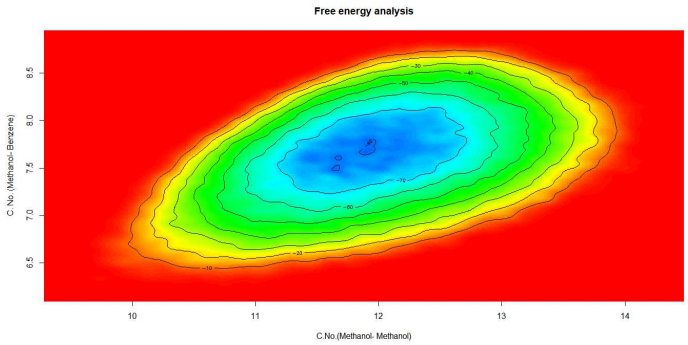


(a) Methanol as reference

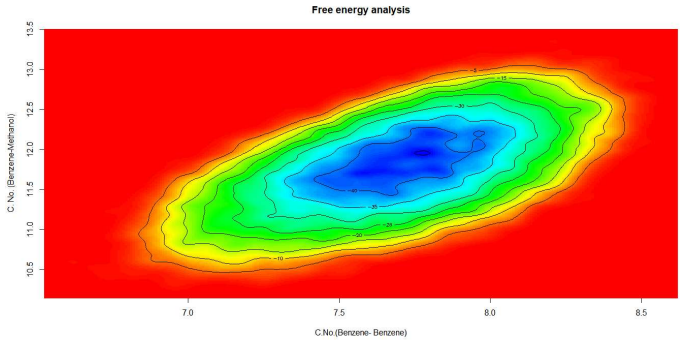


(b) Benzene as reference

Figure 3.20: Metadynamics at T=345K

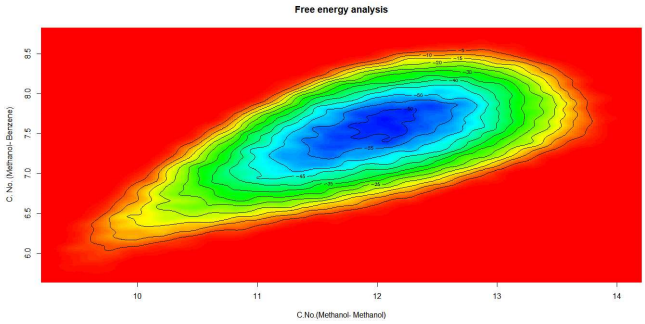


(a) Methanol as reference

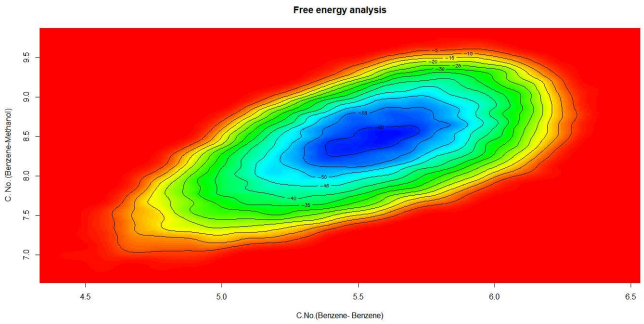


(b) Benzene as reference

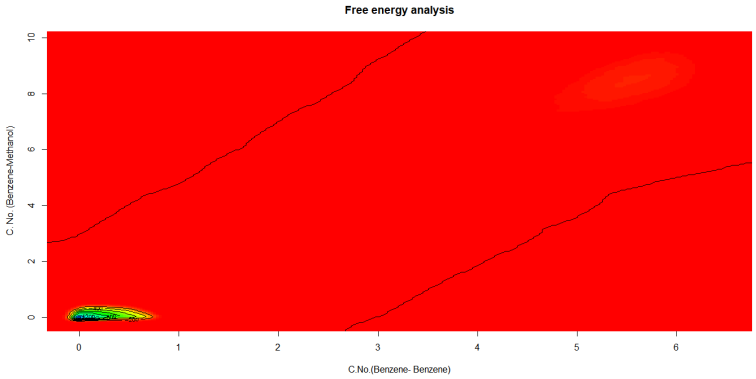
Figure 3.21: Metadynamics at T=440K



(a) Methanol as reference



(b) Benzene as reference



(c) Benzene as reference

Figure 3.22: Metadynamics at T=443K

Bibliography

[1] Irving HMNH, Simpson RB. The viscosity of mixtures of benzene and methanol. J inorg nucl chem. 1972;34(7):22417.

[2] Li D, Gao Z, Vasudevan NK, Li H, Gao X, Li X, et al. Molecular mechanism for azeotrope formation in ethanol/benzene binary mixtures through Gibbs ensemble Monte Carlo simulation. J Phys Chem B. 2020;124(16):337186.

[3] Pribble R. N., Hagemeister F. C., Zevier T.S., Resonant ion-dip infrared spectroscopy of benzene-(methanol)nclusters n=1-6 J. Chem. Phys.,106 (1997) 2145.

[4] Provencal RA, Casaes RN, Roth K, Paul JB, Chapo CN, Saykally RJ, et al. Hydrogen bonding in alcohol clusters: A comparative study by infrared cavity ringdown laser absorption spectroscopy. J Phys Chem A. 2000;104(7):14239.

[5] Jalilian, M. R., Tayyari, S. F. (2009). "Spectra and structure of binary azeotropes VI-benzene-methanol." Spectrochimica Acta -Part A: Molecular and Biomolecular Spectroscopy, 73(5), 828832.

[6] Jeffrey GA. An introduction to hydrogen bonding. New York, NY: Oxford University Press; 1997.

Chapter 4

Conclusion

This study aims to understand the unit structure of methanol-benzene azeotrope and the role of temperature using the molecular dynamics approach due to its advantage over the experimental techniques. The boiling point analysis shows that the OPLS-AA force field underestimates the self-interaction between methanol and benzene molecules in the pure state and also in the mixture as all the temperatures are higher than experimental results. The methanol and benzene azeotrope mixture boils at T=443K.

The hydrogen bonding analysis shows the presence of unconventional hydrogen bond between methanol and benzene as depicted by angle and distance analysis. It is not surprising to note that all methanol is not hydrogen bonded as this can be seen in the pure methanol system as well. The Van-Der Walls forces seem to be the most prominent between this azeotrope which is expected.

The clustering analysis explains the role temperature played during the formation of azeotrope formation. The temperature gradually helps the system to overcome the barrier for the formation of azeotropic structure. The molecules tend to come closer to form more 3:2 (methanol: benzene) ratio azeotrope as the temperature is increased and we observe the peak near boiling point.

The metadynamics study supports this analysis as we increase the temperature the coordination number minima is seen to be getting closer to the azeotropic ratio near boiling point. At T=443K the system is seen to be more inclined towards the formation of a gaseous state.

To study the unit structure we still have to look for the most probable structure present in the clusters from the entire trajectory near boiling point. The structure identification can be more efficiently seen using the Monte Carlo simulations. Small alterations in the force field can be also effective to gain quantitative agreement with experiments but those alterations will be specific to this system only.

Appendix A

The following agglomerative clustering analysis was performed using the different linkage (other than the ward) as explained in chapter 2. The following cluster analysis was done at different temperature but due to their inherent disadvantages, they were unlikely to use in clustering analysis for our system. We deliberately omit out the single linkage.

First, we will look into the average linkage for our agglomerative clustering. In figure A.1 it can be seen that at a distance of 0.68 nm the most probable cluster size is 4 at T=300K with azeotropic composition. This plot is made using the values at different frames in the 100 ns trajectory. Distance 0.68 nm has opted as it is the RDF value for the benzene-methanol system. The ratio clustering analysis was done as shown in Fig A.2. This doesnt give any information about the system and now we omit the clustering analysis for the systems that cant satisfy the 3:2 (Methanol: Benzene) composition i.e. we are interested in clusters with size 5,10,15 and so on.

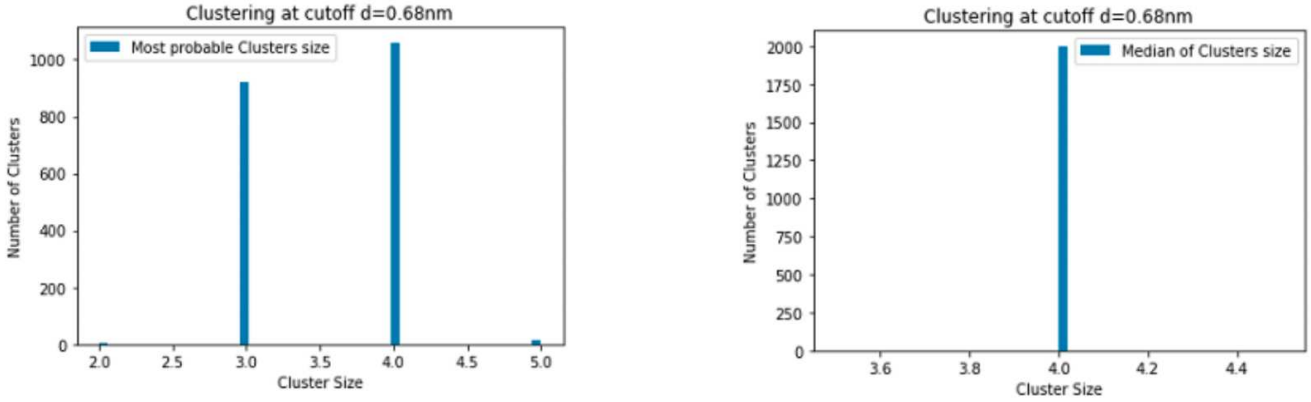


Figure A.1: Mode and median at the 0.68 nm distance at T=300K.

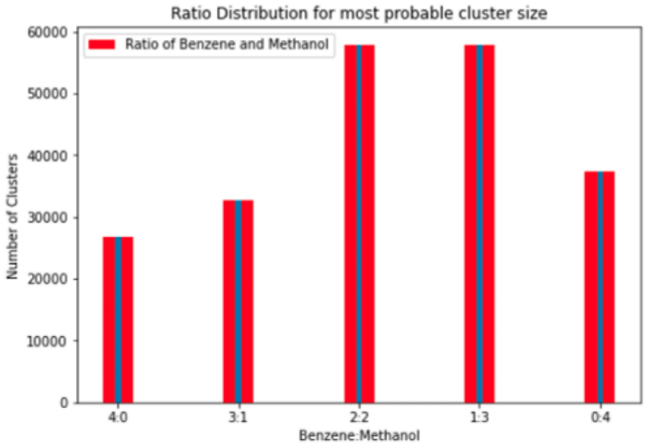


Figure A.2: Ratio distribution for the cluster size of 4 at T=300K.

Afterwards, the clustering analysis was done at different distances using the same linkage as shown in Figure A.3. The distance is varied at a distance of 0.1nm. In this plot, it is evident that a cluster of 5 contains the azeotropic ratio as the highest peak.

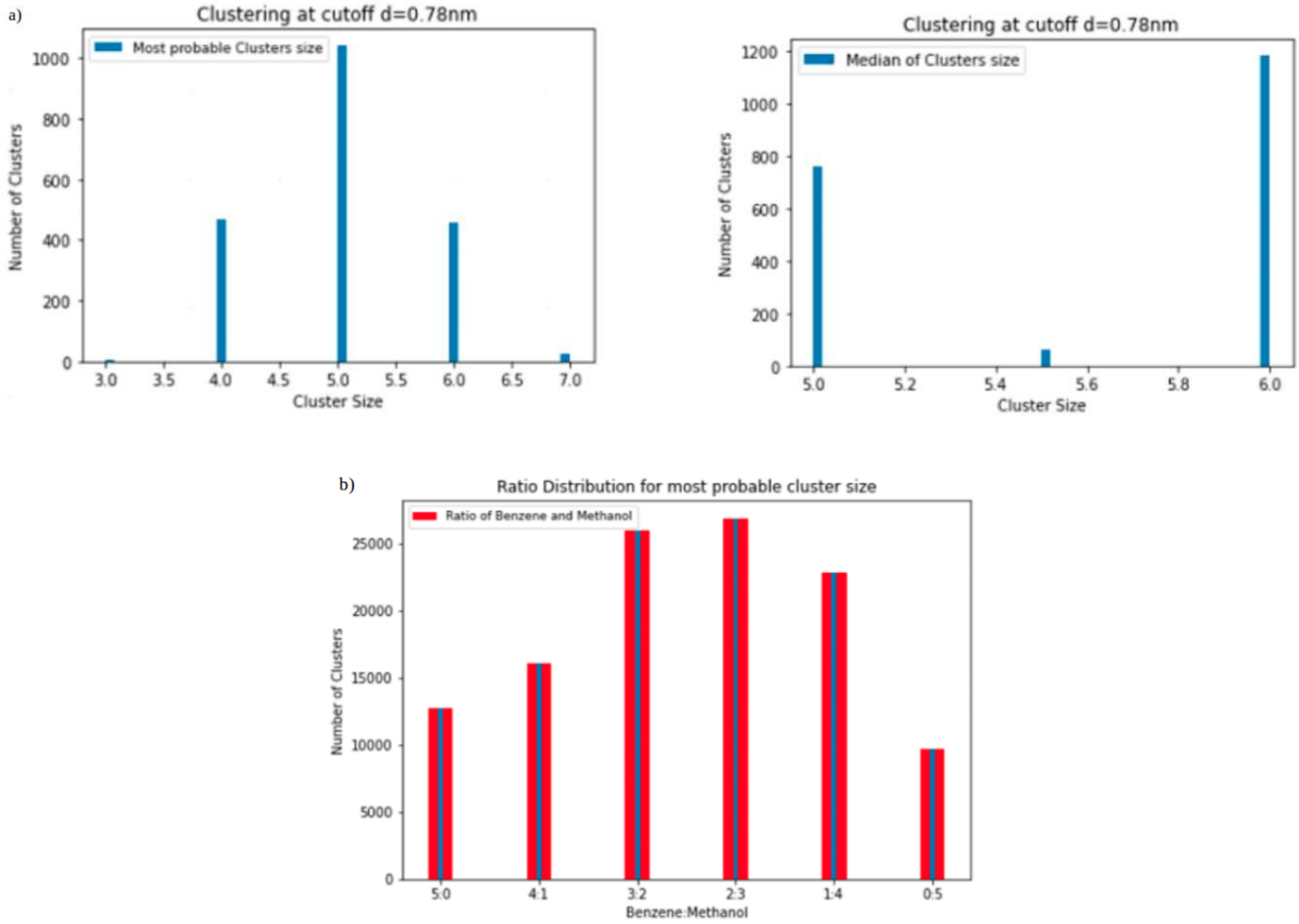
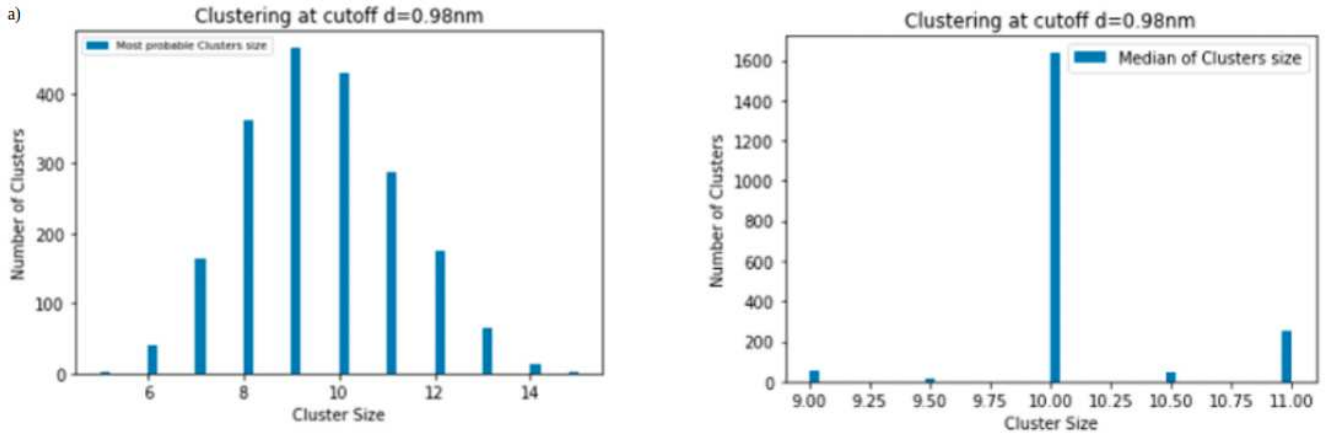
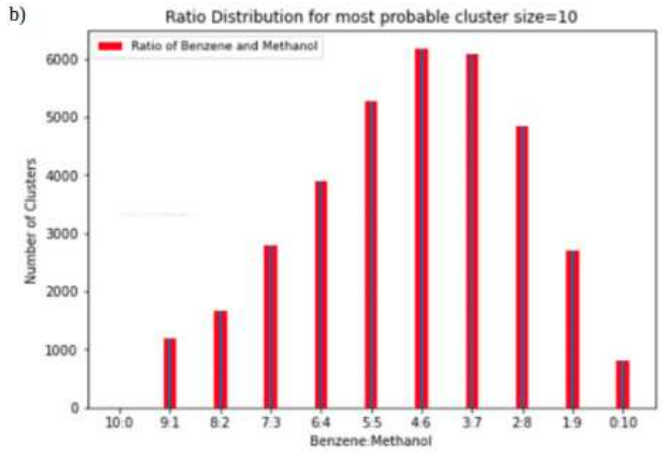
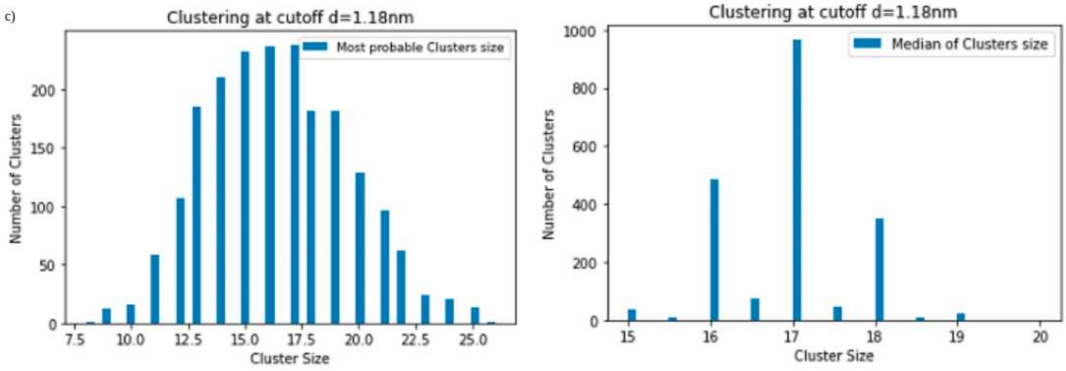
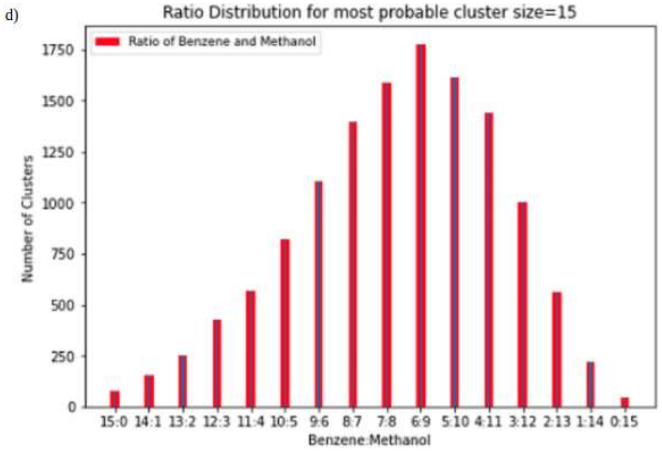


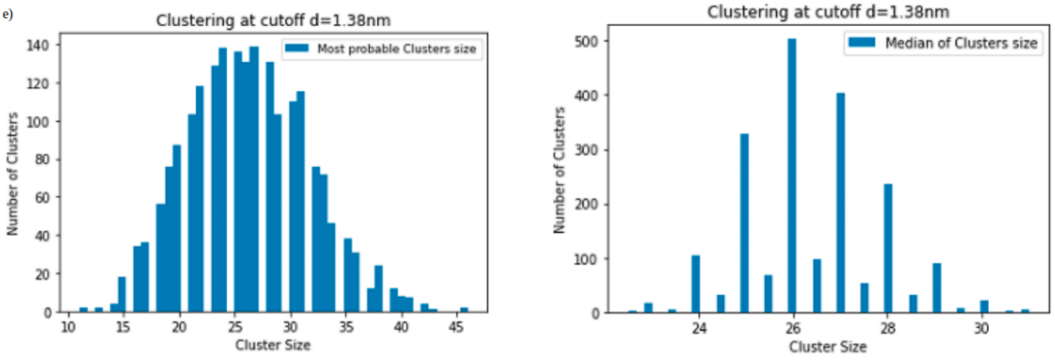
Figure A.3: Clustering analysis at distance 0.78 nm at T=300K. (a) Mode and median of clustering at 0.78nm (b) Ratio analysis for most probable cluster.




b)
Ratio Distribution for most probable cluster size=10
Ratio of Benzene and Methanol
Number of Clusters
0
1000
2000
3000
4000
5000
6000
10:0
9:1
8:2
7:3
6:4
5:5
4:6
3:7
2:8
1:9
0:10
Benzene:Methanol


c)
Clustering at cutoff d=1.18nm
Most probable Clusters size
Number of Clusters
0
50
100
150
200
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Cluster Size
Clustering at cutoff d=1.18nm
Median of Clusters size
Number of Clusters
0
200
400
600
800
1000
15
16
17
18
19
20
Cluster Size


d)
Ratio Distribution for most probable cluster size=15
Ratio of Benzene and Methanol
Number of Clusters
0
250
500
750
1000
1250
1500
1750
15:0
14:1
13:2
12:3
11:4
10:5
9:6
8:7
7:8
6:9
5:10
4:11
3:12
2:13
1:14
0:15
Benzene:Methanol


e)
Clustering at cutoff d=1.38nm
Most probable Clusters size
Number of Clusters
0
20
40
60
80
100
120
140
10
15
20
25
30
35
40
45
Cluster Size
Clustering at cutoff d=1.38nm
Median of Clusters size
Number of Clusters
0
100
200
300
400
500
24
26
28
30
Cluster Size

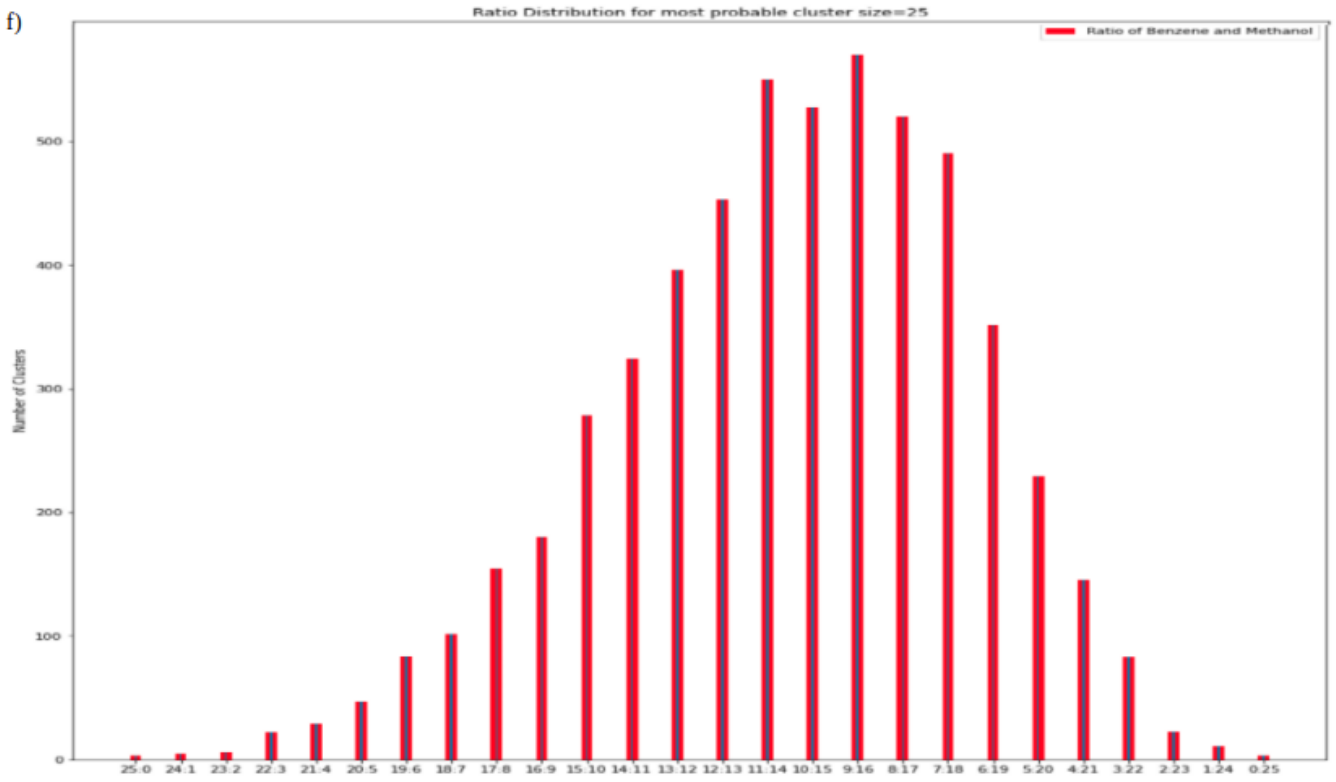
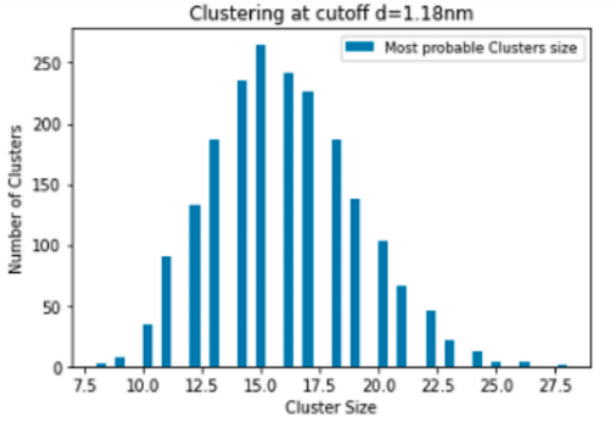
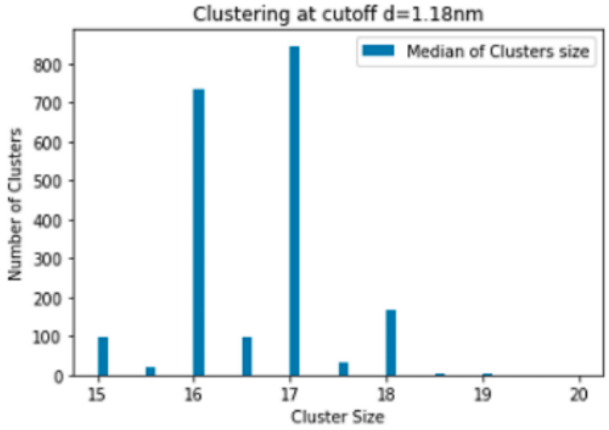


Figure A.4:Mode and median values at different distance and the ratio analysis at T=300K. (a) Mode and Median at 0.98nm. (b) Ratio analysis at 0.98nm for cluster size=10. (c) Mode and median at 1.18nm. (d) Ratio analysis at 1.18nm for cluster size=15. (e) Mode and median at 1.38nm. (f) Ratio analysis at 1.38nm for cluster size=25.

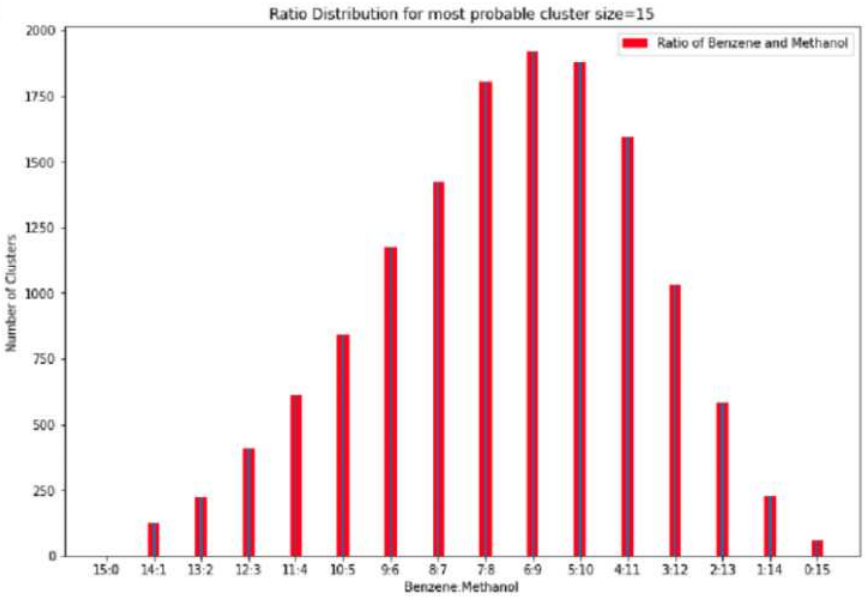
In figure A.5 the clustering was done at T=325K utilizing the same average linkage for clustering analysis. This analysis will only contain the distances 1.18 and 1.28nm.

a)

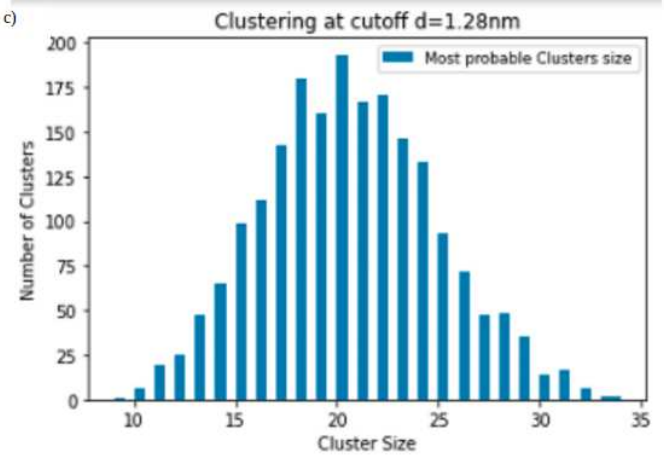

Clustering at cutoff d=1.18nm
Most probable Clusters size
Number of Clusters
Cluster Size

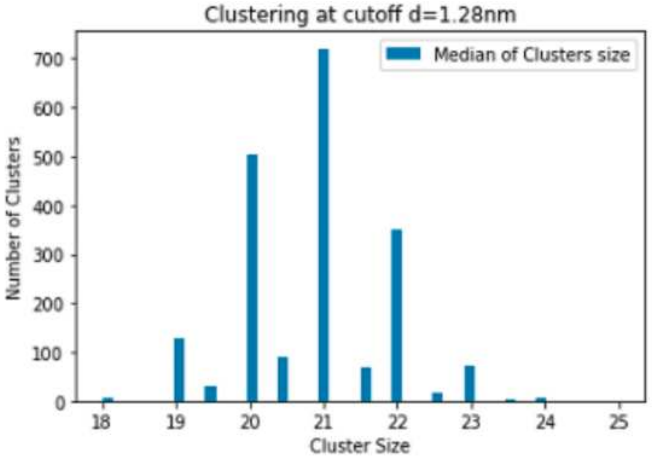

Clustering at cutoff d=1.18nm
Median of Clusters size
Number of Clusters
Cluster Size

b)


Ratio Distribution for most probable cluster size=15
Ratio of Benzene and Methanol
Number of Clusters
15:0
14:1
13:2
12:3
11:4
10:5
9:6
8:7
7:8
6:9
5:10
4:11
3:12
2:13
1:14
0:15
Benzene:Methanol

c)


Clustering at cutoff d=1.28nm
Most probable Clusters size
Number of Clusters
Cluster Size


Clustering at cutoff d=1.28nm
Median of Clusters size
Number of Clusters
Cluster Size

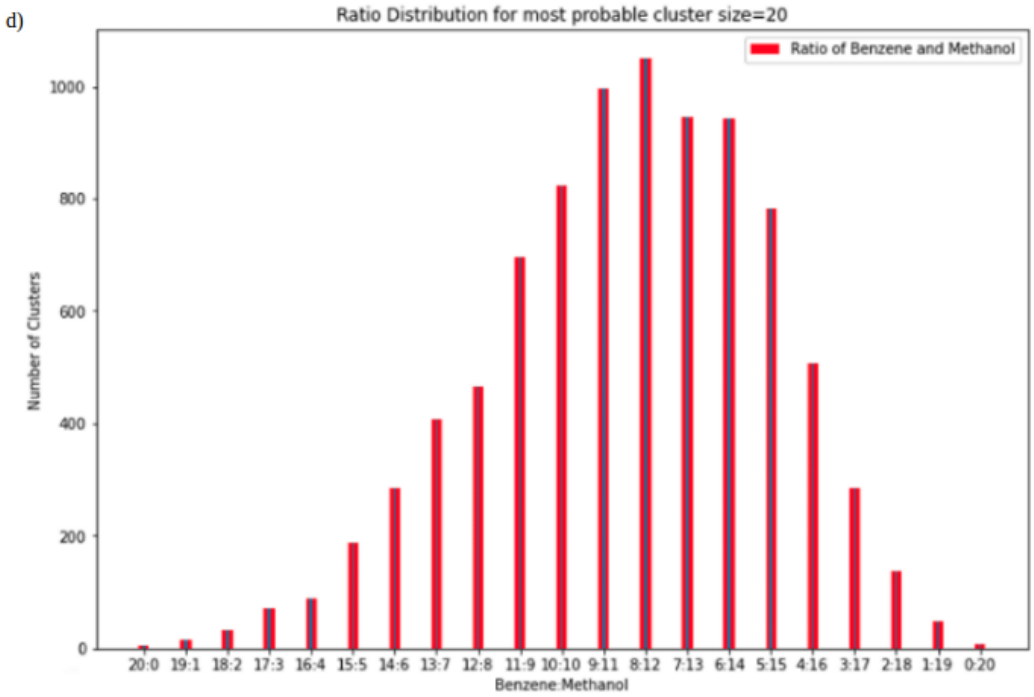
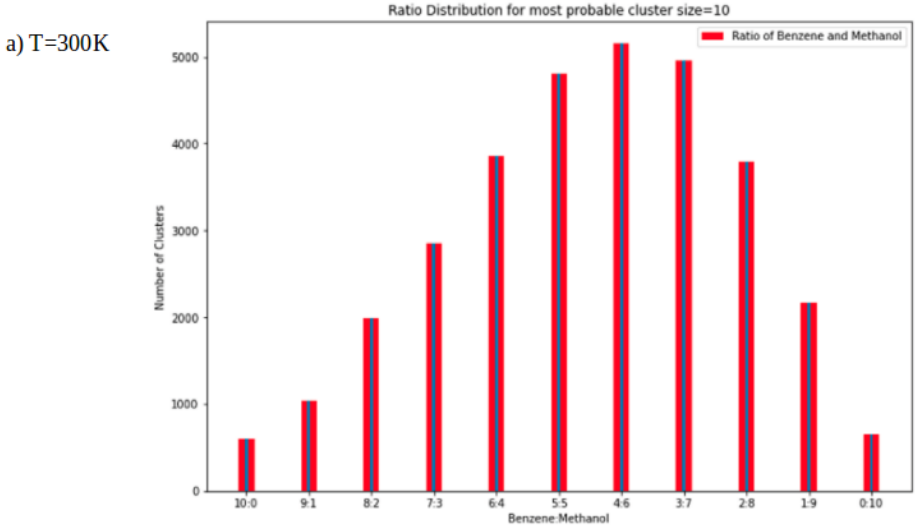
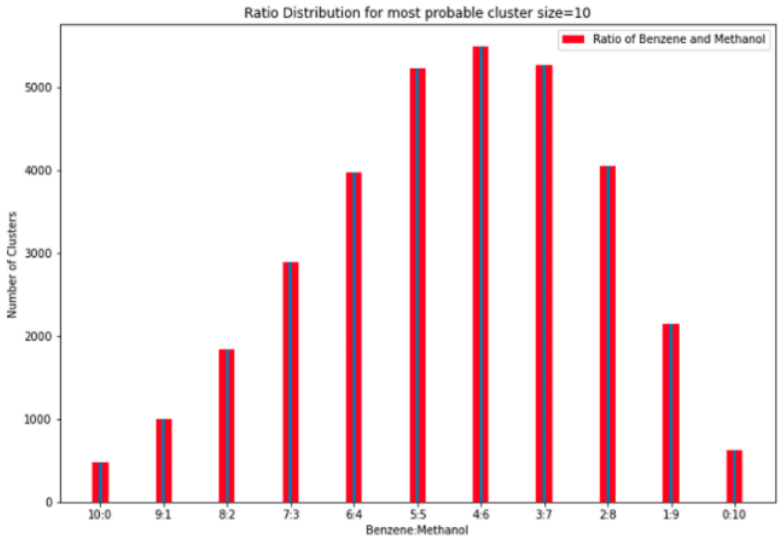


Figure A.4:Mode and median values at different distance and the ratio analysis at T=300K. (a) Mode and Median at 0.98nm. (b) Ratio analysis at 0.98nm for cluster size=10. (c) Mode and median at 1.18nm. (d) Ratio analysis at 1.18nm for cluster size=15. (e) Mode and median at 1.38nm. (f) Ratio analysis at 1.38nm for cluster size=25.

Moreover, the analysis was done using the Complete or Maximum linkage for agglomerative clustering analysis i.e. already discussed in chapter 2 and as shown in figure A.6. This was done at different Temperature T=300K, 315K and 330K but at a distance of 0.98nm. This result shows a similar trend as ward analysis discussed in chapter 3 i.e. as the temperature increases the azeotropic composition ratio becomes exceedingly significant than the previous one. This provides a hint towards the role of temperature in the formation of azeotrope unit- structure.



b) T=315K



c) T=330K

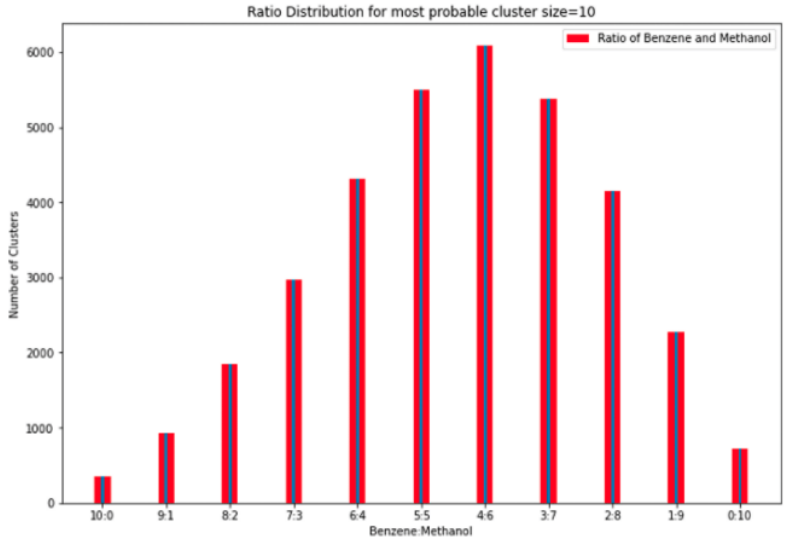
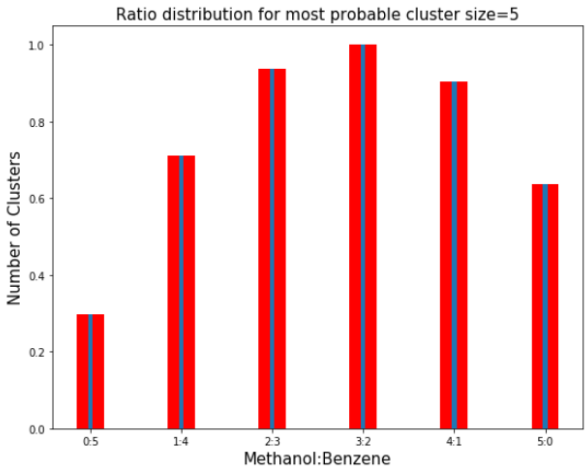


Figure A.6: Ratio analysis at distance threshold 0.98nm with complete linakge in HAC. (a)T=300K (b) 315K and (c) 330K

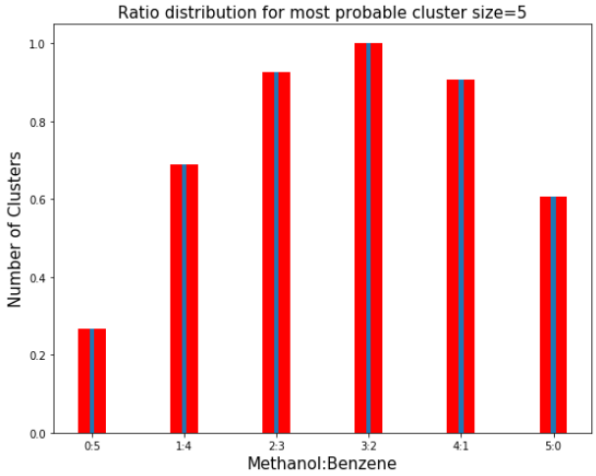
These two linkage were not utilized in the main analysis due to the cluster sizes that were formed were quite different when visualized using VMD. On the contrary, ward linkage was found to form almost similar type of clusters while analyzing the 100ns long trajectories. All the linkages shows the similar trend when the temperature is increased making the azeotropic ratio cluster as the dominant cluster.

Appendix B

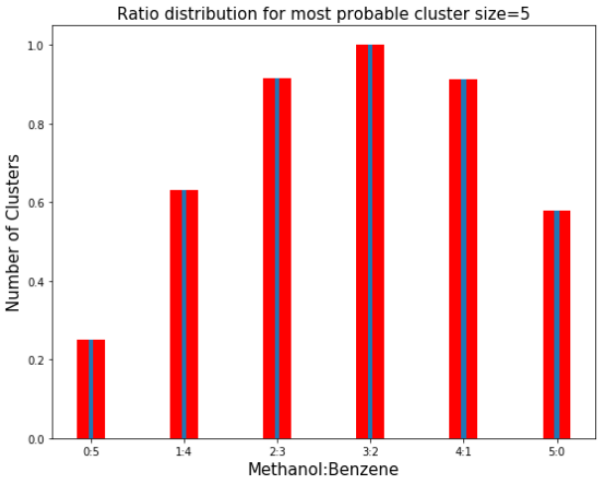
Clustering analysis at different temperature.



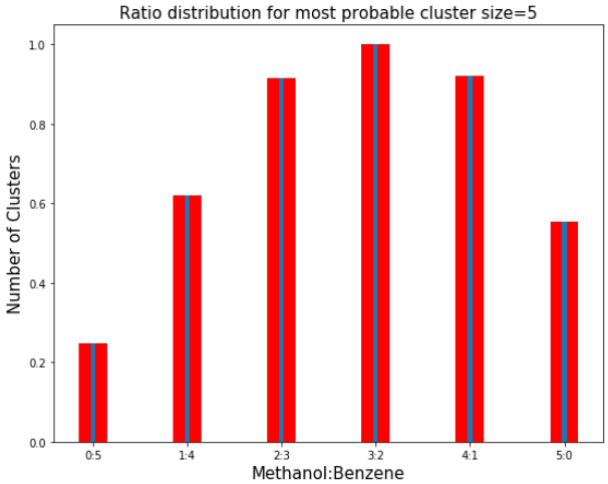
(a) T=300K, 6.2%, Maximum peak=32320



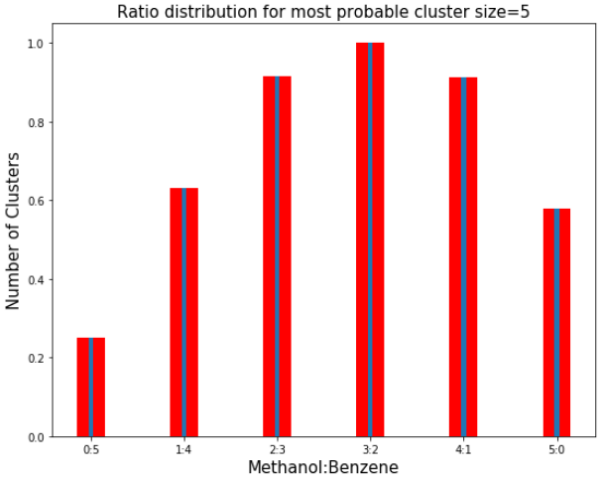
(b) T=310K, 7.3%, , Maximum peak=33410



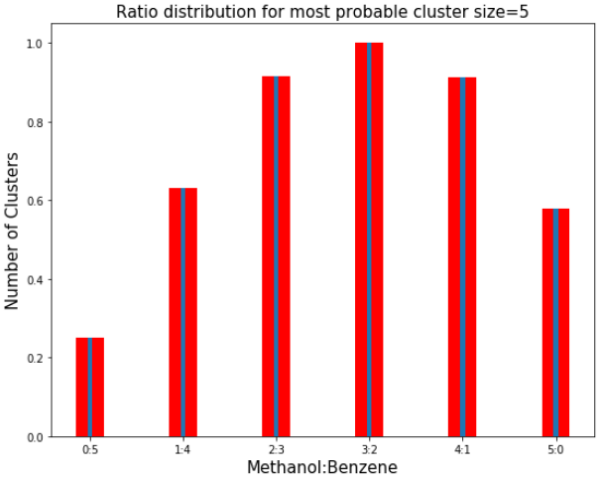
(c) T=315K, 8.5%, , Maximum peak=33890



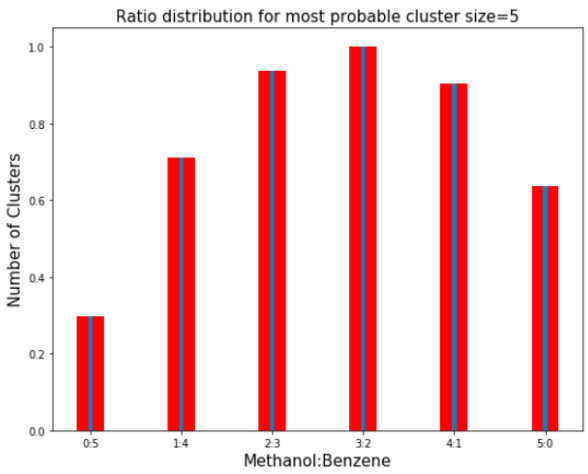
(d) T=320K, 8.5%, , Maximum peak=34112



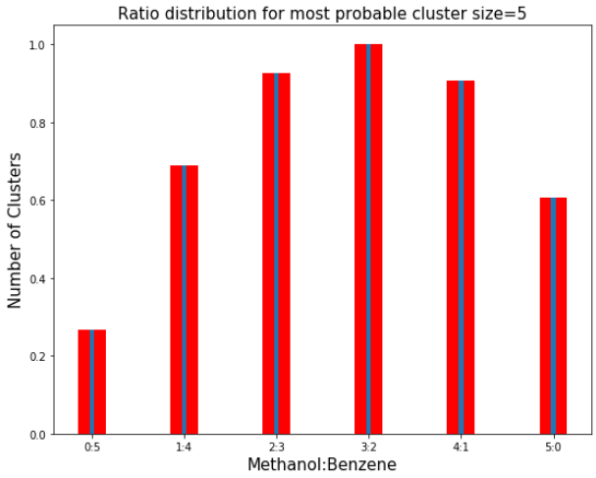
(e) T=325K, 9.4%, , Maximum peak=34210



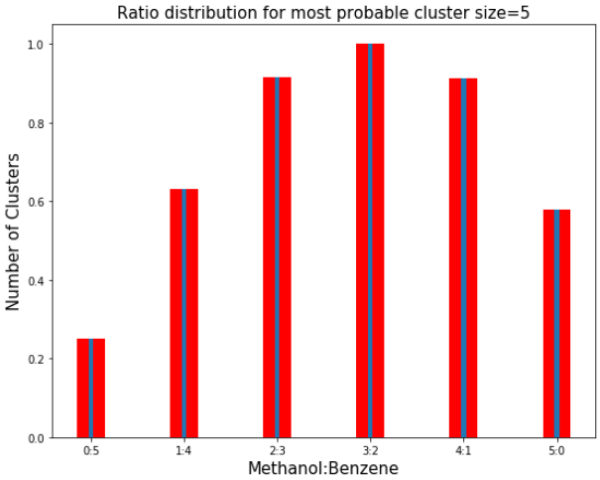
(f) T=330K, 11.4%, , Maximum peak=34316



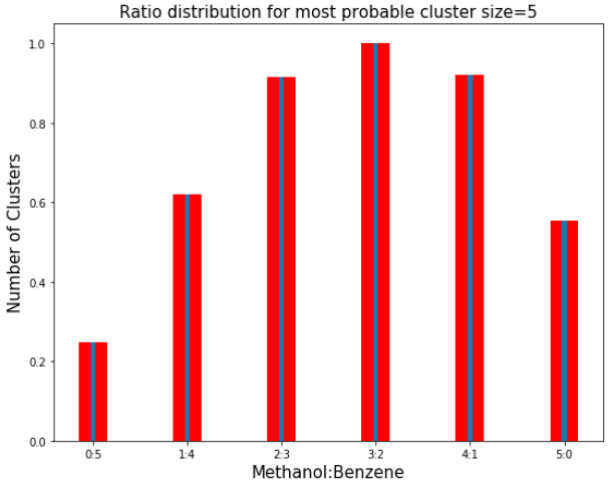
(g) T=335K, 13.3%, Maximum peak=34615



(h) T=340K, 14.7%, , Maximum peak=34410



(i) T=365K, 16.8%, , Maximum peak=35004



(j) T=440K, 23.1%, , Maximum peak=34975

Figure B.1: Clustering analysis at different temperature

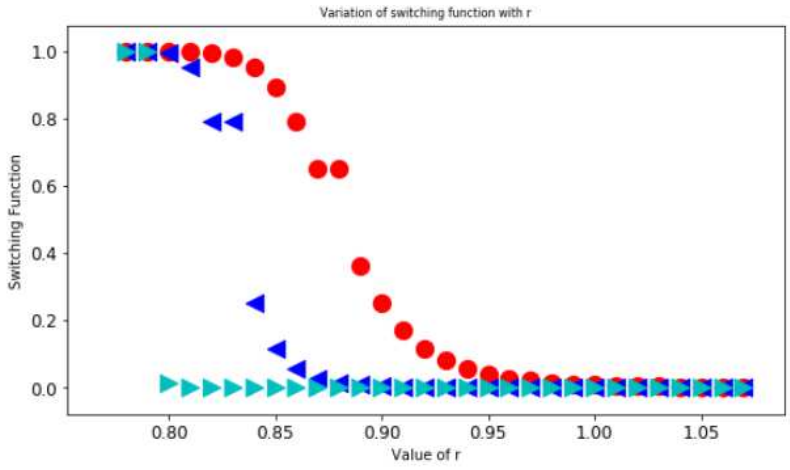


Figure B.2: Decline of switching function with r using different r_0 values 0.01, 0.5 and 0.1