Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting information–Capturing the potential energy landscape of large size molecular clusters from atomic interactions up to 4-body system using deep learning

Shweta Jindal,*,† Po-Jen Hsu,† Huu Trong Phan,†,‡,¶ Pei-Kang Tsou,† and Jer-Lai Kuo*,†,¶,§

†*Institute of Atomic and Molecular Sciences, Academia Sinica, No. 1 Roosevelt Road, Section 4, Daan District, Taipei City 10617, Taiwan*

‡*Molecular Science and Technology Program, Taiwan International Graduate Program, Academia Sinica, Taipei, 11529, Taiwan*

¶*Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan*

§*Molecular Science and Technology, National Taiwan University, Section 4, Daan District, Taipei City 10617, Taiwan*

E-mail: jindal.2292@gmail.com; jlkuo@pub.iams.sinica.edu.tw

Contents

List of Figures

List of Tables

Analogy to MBE

We chose 23 isomers of $(CH_3OH)_{14}$ and used QChem software to calculate the single point (SP) energy of the cluster without MBE approximation. We have also applied MBE and calculated the energy of the clusters by truncating at 3-body terms and 4-body terms, respectively. The level of theory used was B3LYP/6-31+G(d,p). In Fig. S1, we have plotted three differences between (a) Energy truncated at 3-body terms and SP energy, (b) Energy truncated at 4-body terms and SP energy and (c) Energy truncated at 3-body terms and energy truncated at 4-body terms. The difference between the SP energy and the energy from 3-body truncated clusters varies between 6 kJ/mol to 15 kJ/mol. The difference between the SP energy and the energy from 4-body truncated clusters varies between 0 kJ/mol to 3 kJ/mol. This shows that 4-body terms estimates the energy of cluster very near to the SP energy.

Inspired from the accuracy of MBE (truncated upto four-body terms), we create a database of molecular clusters containing one-body system, two-body cluster, three-body cluster and four-body cluster in the current work. A model trained on SP energy of these sub-clusters predicts the energy of large size clusters within an error of 3 kJ/mol (upto $(CH_3OH)_{14}$) (Table 3). Instead of partitioning the correction terms, SchNet can capture the atomic interactions with a good error cancellation by training a database upto four-body system. Similar to MBE, 4-body systems are essential for methanol clusters.

SchNet model trained upto three-body clusters

For training the SchNet model upto three-body clusters, the database consisted of 37645 points (1084 one-body system + 12721 two-body clusters + 23840 three-body clusters). The database was shuffled and randomly splitted into training set (31508), test set (3765) and validation set (2372). On obtaining the best model from training, we calculated the MAE for energy and force on the test-sets of large size clusters (Table S1). A small test set (2132

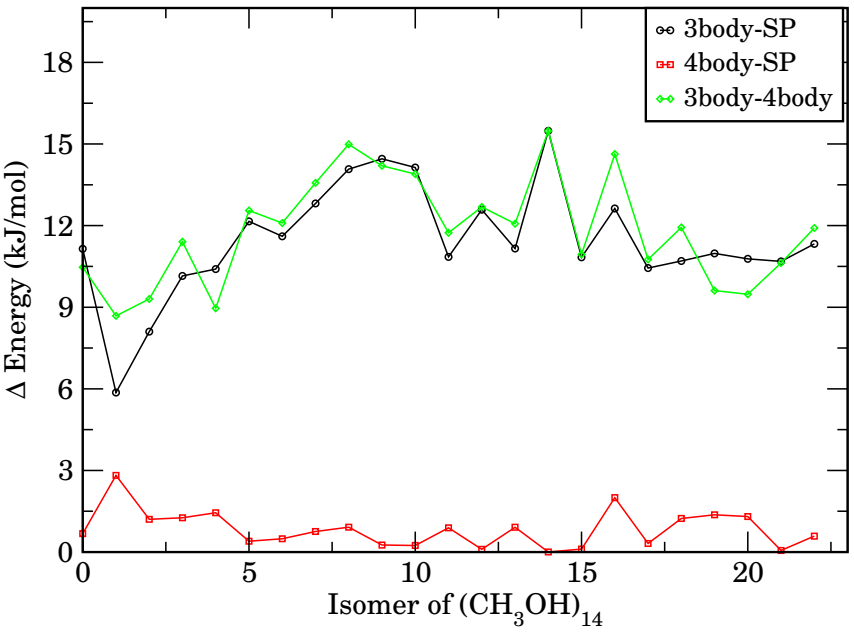


Figure S1: Plot representing the difference in energy between single point from B3LYP/6-31+G(d,p) and MBE-B3LYP/6-31+G(d,p) for 23 isomers of $(CH_3OH)_{14}$

points) for $(CH_3OH)_4$ was generated from the four-body clusters data to validate the model trained upto three-body clusters. The MAE in energy and force increases for large size clusters by 2 kJ/mol and 2 kJ/mol/Å respectively as compared to the model trained upto four-body sub-clusters (Table 3 in manuscript). This shows that we need four-body clusters to improve the SchNet model and make it more predictable for large size clusters. Power spectrum is not that sensitive to energy and SchNet can capture the essential vibrational modes by using a model trained upto three-body clusters only (Fig. S11). The peak positions are similar to the powerspectrum generated using the SchNet model trained upto tetramers.

SchNet model trained upto five-body clusters

To investigate the impact of five-body clusters on the quality of PES, we added a fraction of them in the training database. The five-body sub-clusters can be extracted from $(CH_3OH)_{6-8}$ in the same way as smaller fragments are extracted. On splitting $(CH_3OH)_{6-8}$ into 5-body

Table S1: The MAE in energy (kJ/mol) and forces (kJ/mol/Å) on the test sets of sub-clusters ($(CH_3OH)_N$) N = 1 to 3, and the test sets of different size clusters N = 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14, using the best model fitted from SchNet till three-body cluster.

Property	Energy	Force
N = 1 to 3	0.84	0.85
N = 4	1.01	1.02
N = 5	1.93	1.67
N = 6	3.89	2.03
N = 7	4.41	2.59
N = 8	3.71	3.37
N = 9	3.70	3.85
N = 10	4.05	4.01
N = 11	4.57	4.18
N = 12	4.98	4.31
N = 13	5.53	4.57
N = 14	5.28	4.44

sub-clusters using the cut-off of 8Å we get 329 sub-clusters from $(CH_3OH)_6$(86 clusters), 1991 sub-clusters from $(CH_3OH)_7$(178 clusters) and 9671 sub-clusters from $(CH_3OH)_8$(382 clusters). On screening the duplicates, we get a total of 8812 5-body sub-clusters from $(CH_3OH)_{6-8}$. To include a fraction of these clusters, we randomly picked 2196 clusters and calculated their single point energy and forces from B3LYP/6-31+G(d,p). We added the 2196 clusters to 61832 datapoints (one-body to four-body sub-clusters database) and split the data into train set (54743 points), test set (6403 points) and validation set (2881 points). On obtaining the best model, we calculated the MAE for the test sets of large size clusters (Table S2). On adding the five-body sub-clusters to the training set have not put much impact on the accuracy of SchNet model. The MAE in energy has increased by 1-2 kJ/mol for $(CH_3OH)_{8-14}$ as compared to the model trained upto four-body sub-clusters (Table 3 in manuscript). The MAE for gradients decreases on addition of five-body sub-clusters. Thus, the inclusion of higher-body clusters does not significantly improve the quality of the SchNet model.

Table S2: The MAE in energy (kJ/mol) and forces (kJ/mol/Å) on the test sets of sub-clusters ($(CH_3OH)_N$) N = 1 to 5, and the test sets of different size clusters N = 5, 6, 7, 8, 9, 10, 11, 12, 13 and 14, using the best model fitted from SchNet till five-body clusters.

Property	Energy	Force
N = 1 to 5	0.61	0.41
N = 5	0.78	0.83
N = 6	1.32	1.04
N = 7	1.78	1.13
N = 8	3.75	1.32
N = 9	3.91	1.77
N = 10	5.17	1.78
N = 11	5.57	1.77
N = 12	5.40	1.83
N = 13	5.18	2.18
N = 14	4.78	2.05

Comparing SchNet model with MBE results

To compare the predictions of SchNet model (trained upto four-body) with MBE (truncated at four-body terms), we picked 50 clusters of $(CH_3OH)_7$ and calculated their energy using QChem as shown in Fig. S2. The difference between SchNet model (trained upto four-body) and MBE energy is 0.07-6 kJ/mol with average at 2.2 kJ/mol.

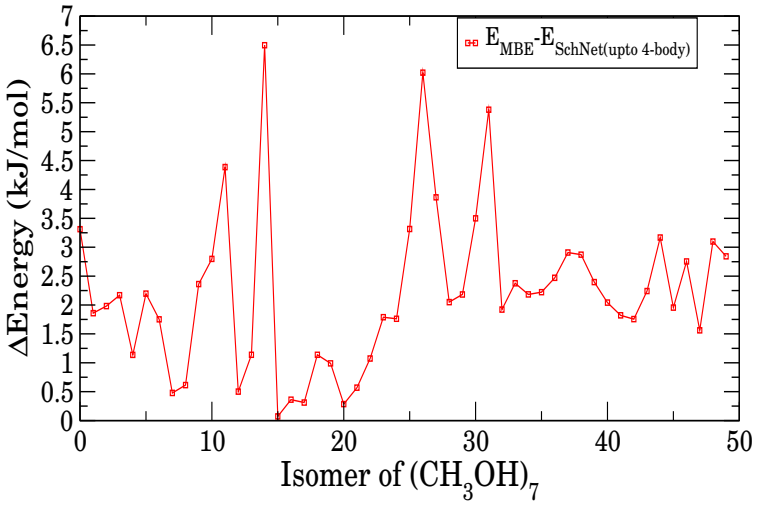


Figure S2: Plot representing the prediction error between SchNet (models trained upto four-body clusters) and MBE truncated at 4-body terms using MBE-B3LYP/6-31+G(d,p) for 50 isomers of $(CH_3OH)_7$.

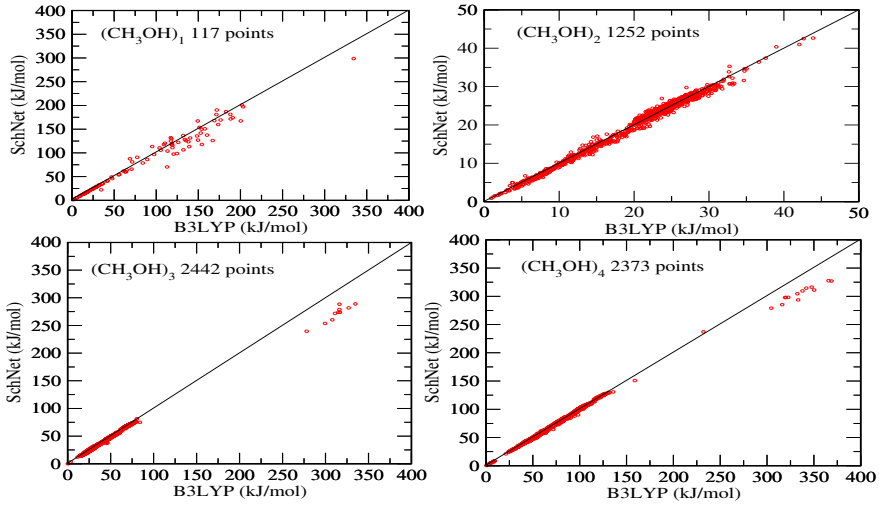


Figure S3: The correlation plots of energy predicted from SchNet model and energy obtained from B3LYP/6-31+G(d,p) for the monomer (117 points), dimer (1252 points), trimer (2442 points) and tetramer (2373 points) in the test set (6184 points)

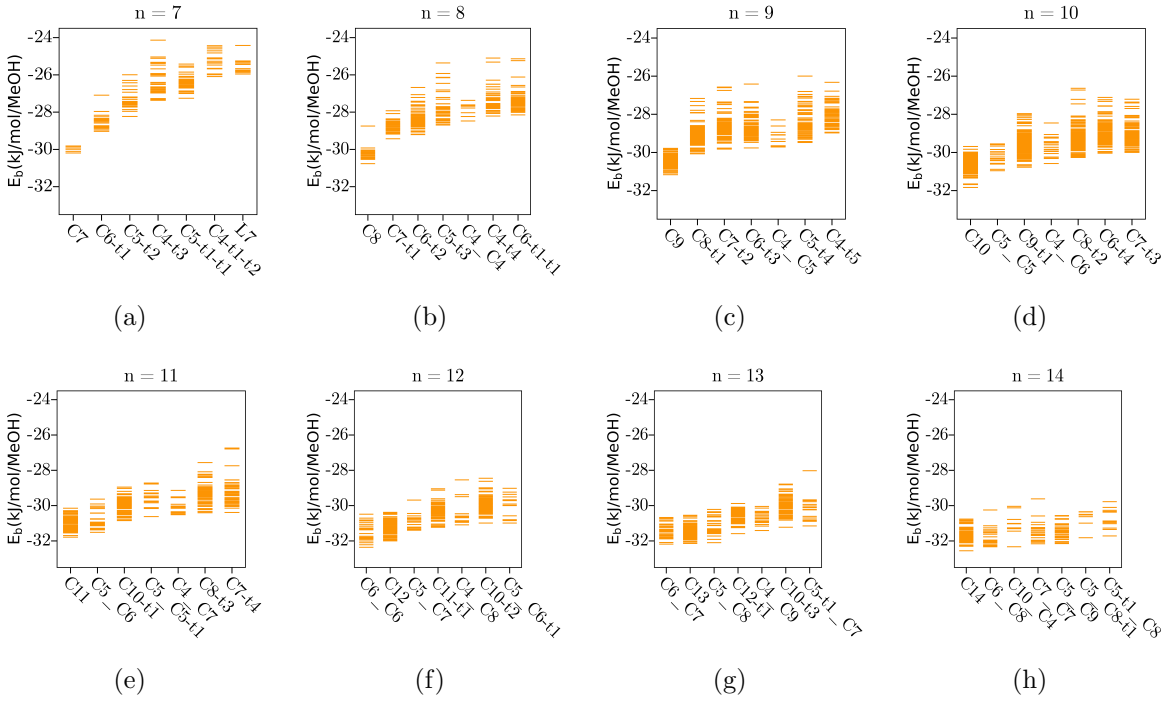


Figure S4: The binding energy per molecule for different isomers of $(CH_3OH)_N$, $N = 7$ to 14

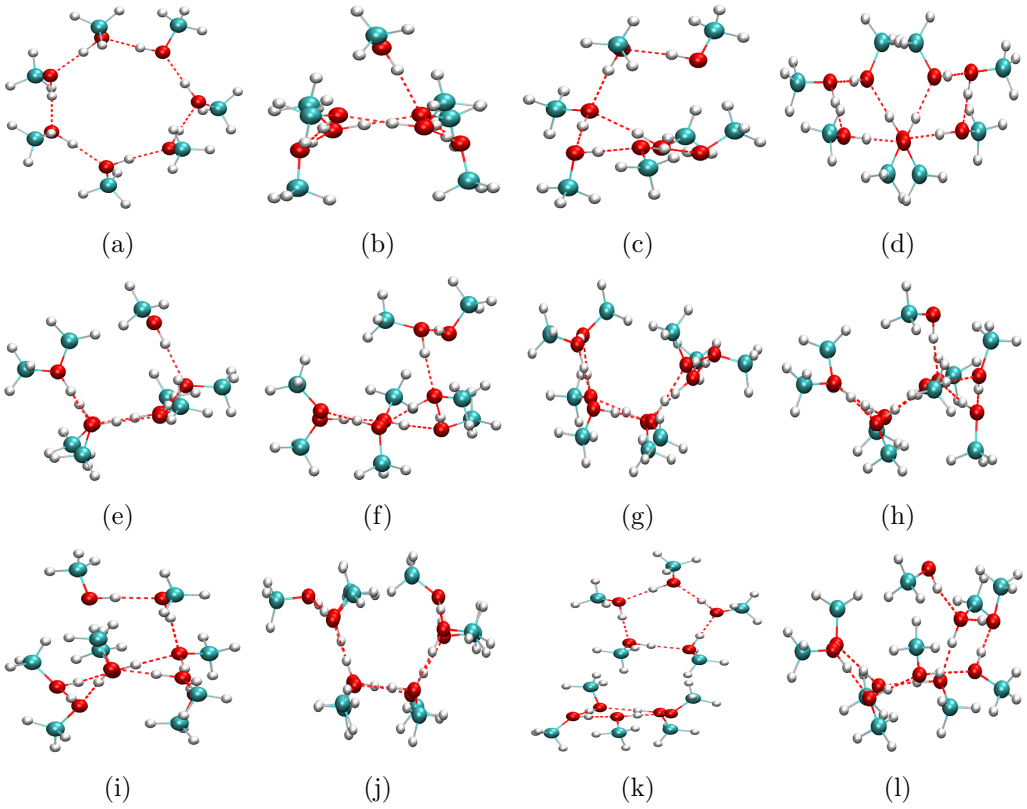


Figure S5: The lowest energy isomer of top 3 topologies of $(CH_3OH)_N$, $N = 7$ to 10, (a-c) $(CH_3OH)_7$ - C7, C6-t1, C5-t2; (d-f) $(CH_3OH)_8$ - C8, C7-t1, C6-t2; (g-i) $(CH_3OH)_9$ - C9, C8-t1, C7-t2; (j-l) $(CH_3OH)_{10}$ - C10, C5-C5, C9-t1.

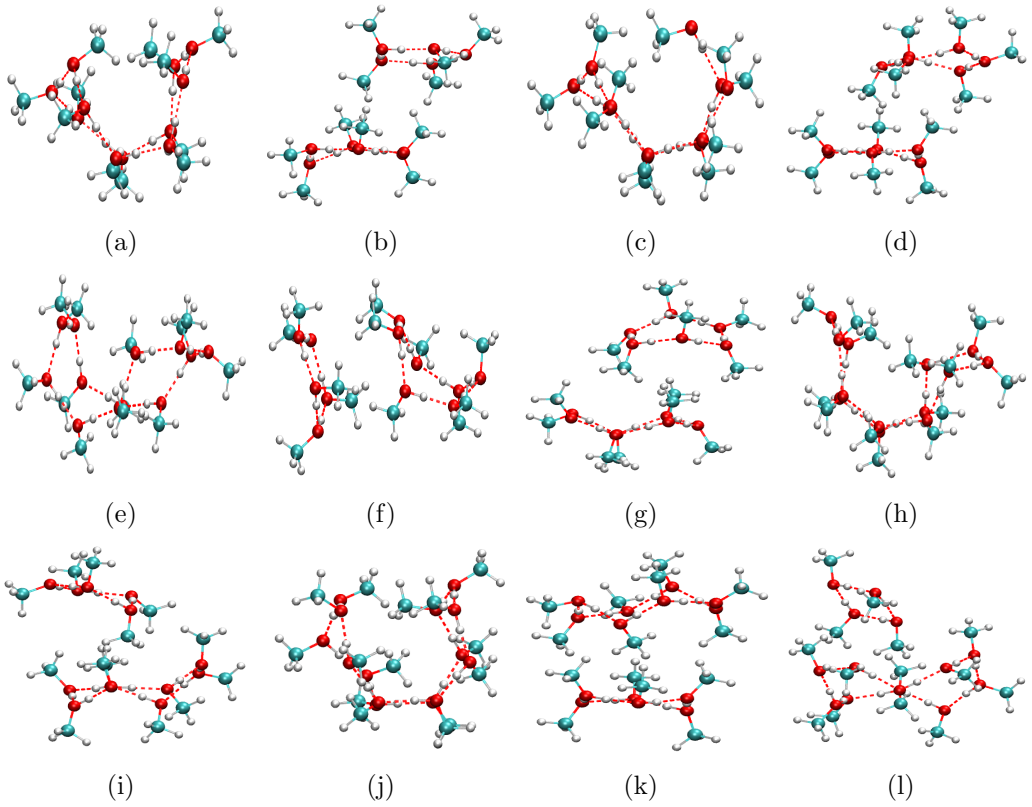
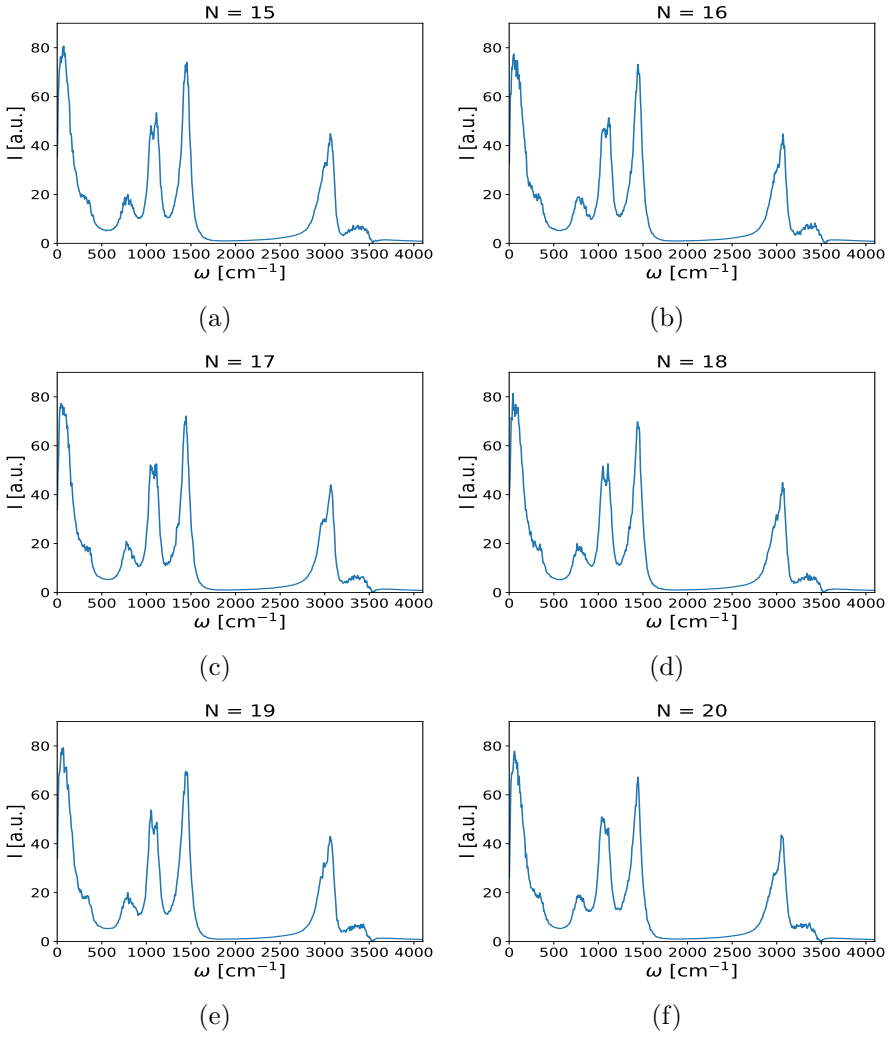


Figure S6: The lowest energy isomer of top 3 topologies of $(CH_3OH)_N$, $N = 11$ to 14, (a-c) $(CH_3OH)_{11}$ - C11, C5-C6, C10-t1; (d-f) $(CH_3OH)_{12}$ - C6-C6, C12, C5-C7; (g-i) $(CH_3OH)_{13}$ - C6-C7, C13, C5-C8; (j-l) $(CH_3OH)_{14}$ - C14, C6-C8, C10-C4.

Table S3: The total number of one-body, two-body, three-body and four-body sub-clusters from the low energy isomers of $(CH_3OH)_N$ (N = 5 - 8) without using any cut-off to extract

N	Isomers	one-body system	two-body clusters	three-body clusters	four-body clusters	Total combinations
5	38	5×38	10×38	10×38	5×38	1140
6	86	6×86	15×86	20×86	15×86	4816
7	178	7×178	21×178	35×178	35×178	17444
8	382	8×382	28×382	56×382	70×382	61884
Total	684	5008	16104	29722	34450	85284



(a) (b) (c) (d) (e) (f)

Figure S7: The vibrational power spectrum of $(CH_3OH)_N$, N = 15 to 20 obtained from running the MD simulation using schnet model at 100K.

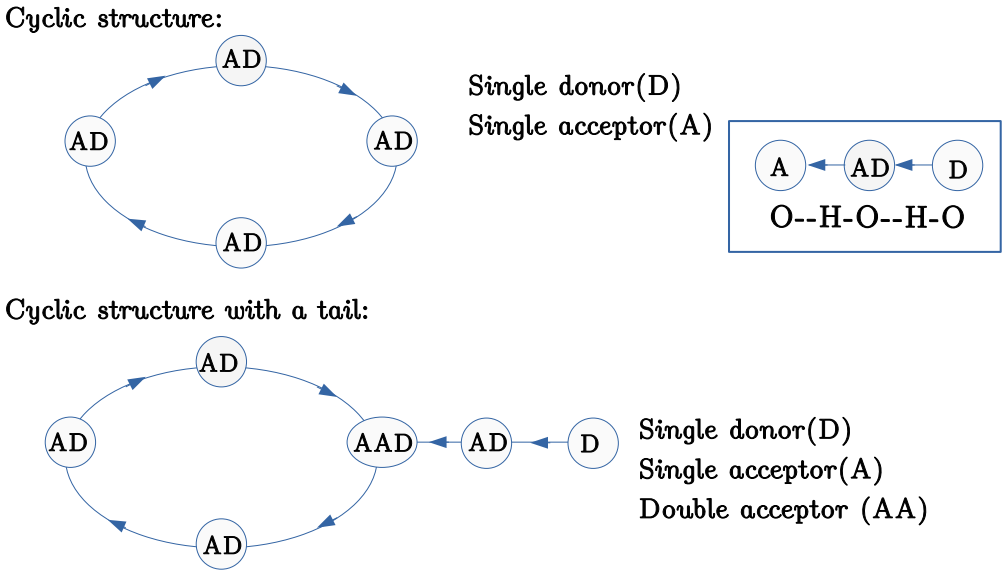


Figure S8: Symbolic representation of H-bond network in $(CH_3OH)_n$

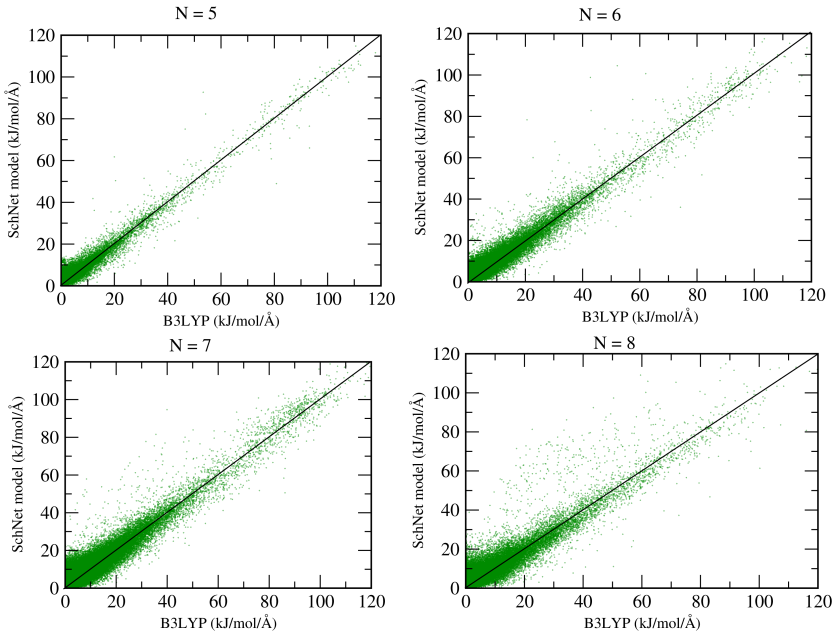


Figure S9: Correlation plot for norm of atomic gradients for the test set of $(CH_3OH)_{5-8}$ between SchNet model and B3LYP

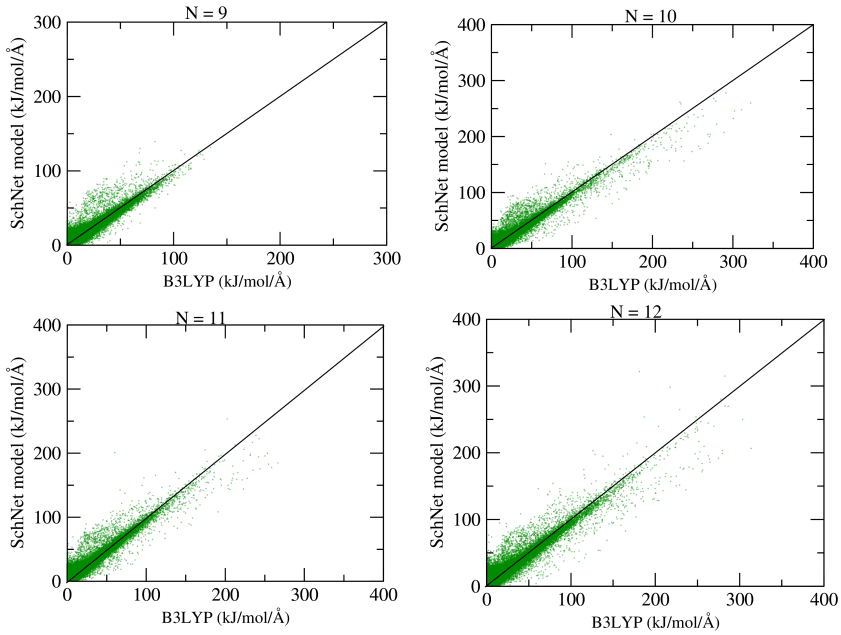


Figure S10: Correlation plot for norm of atomic gradients for the test set of $(CH_3OH)_{9-12}$ between SchNet model and B3LYP

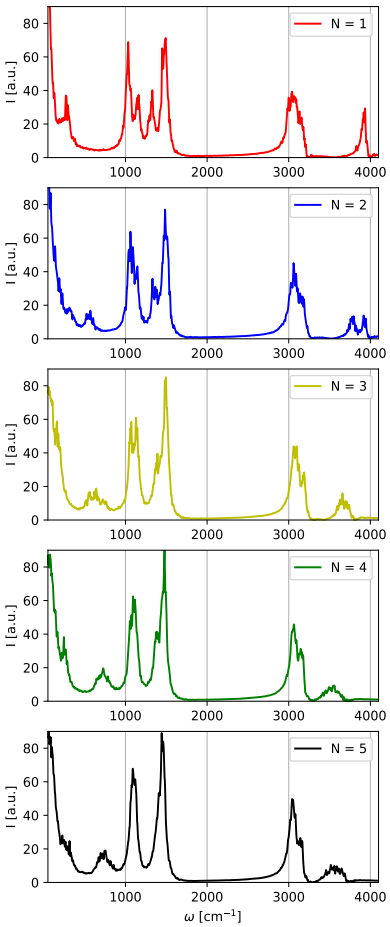


Figure S11: The vibrational power spectrum of $(CH_3OH)_1$ to $(CH_3OH)_5$ obtained from running the MD simulation using SchNet model at 100K trained upto trimer

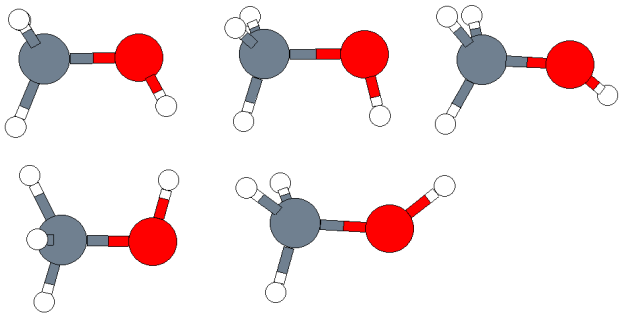


Figure S12: A few distinct clusters of $(CH_3OH)_1$