Electronic Supplementary Information

For

A self-adapting first-principles exploration on the dissociation mechanism in sodiated aldohexose pyranoses assisted with neural network potentials

Pei-Kang Tsou,[a] Hai Thi Huynh,[a,b,c] Huu Trong Phan,[a,b,c] and Jer-Lai Kuo,[a,b,c,d]

a. Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei, 10617, Taiwan
b. Molecular Science and Technology Program, Taiwan International Graduate Program, Academia Sinica, Taipei, 11529, Taiwan
c. Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan.
d. International Graduate Program of Molecular Science and Technology (NTU-MST), National Taiwan University, Taipei 10617, Taiwan

Figure S1. Mean absolute error (MAE) of energy (in kJ/mol) and force (in kJ/mol/Å) obtained during training. The learning rate is controlled by the Reduce LR On Plateau hook and is reduced during training.

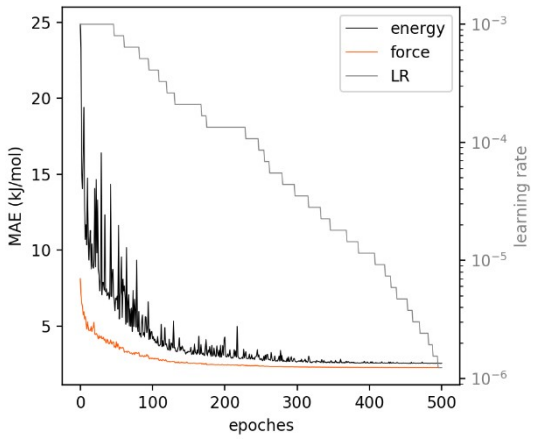
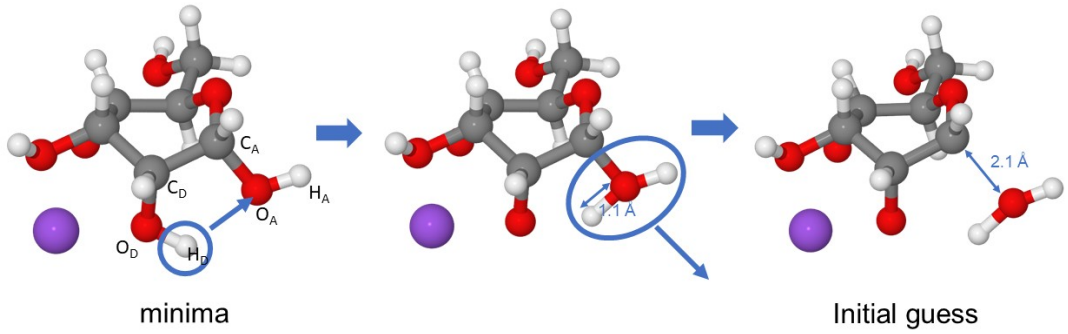


Figure S2. (a) Structure modifications for generating the initial guess in protocol-2. The minima conformers with the Na^+ binding to the proton donor were chosen as the candidates. The hydroxyl group of the proton donor was labeled as O_DH_D, while that of the proton accepter was O_AH_A. The C atom attached to O_A and O_D was labeled as C_A and C_D, respectively. For the dehydration, the initial guess is generated with the bond length of C_A–O_A elongated to 2.1Å, and the distance of O_A–H_D shortened to 1.1 Å. **(b)** The Structure modifications in protocol-2.2 (for QST3 calculation). The Structure-3 is the same as that used in protocol-2. The Structure-1 and Structure-2 are generated according to the direction of the normal mode of the C_A–O_A and O_A–H_D as shown in the figures.

(a)



(b)

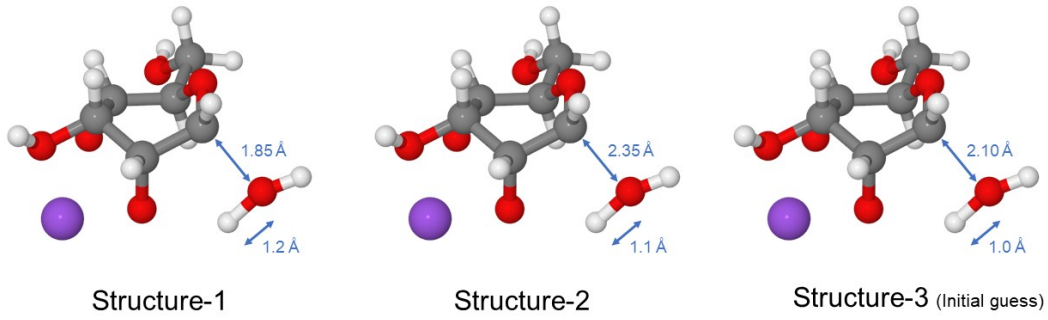
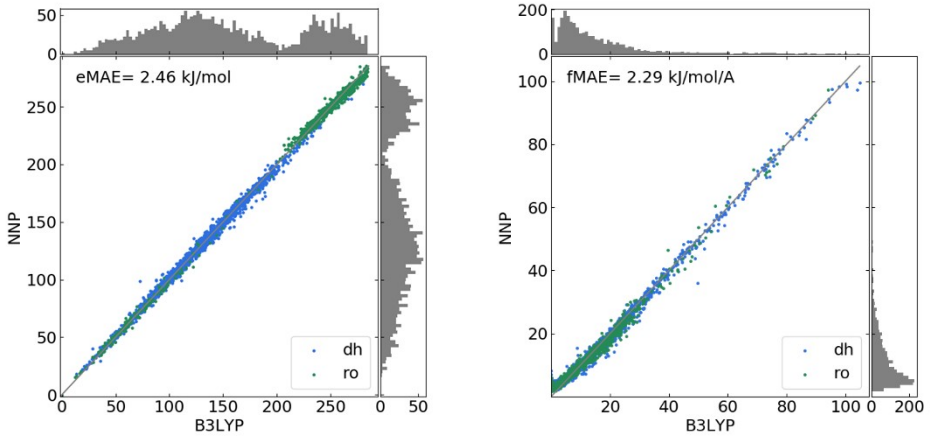


Figure S3. Performance of best NNP model trained on the DFT datapoints in AH-0. Relative energy predictions of NNP-0(y-axis) vs. B3LYP(x-axis) of the (a) Validation set and (b) Test set. The histogram of the actual value and the predicted value is shown in grey on the x-axis and y-axis. Data points along the reaction path leading to dehydration and ring-opening are shown in blue and green, respectively.

(a) Energy: **Gradient:**



(b) energy gradient

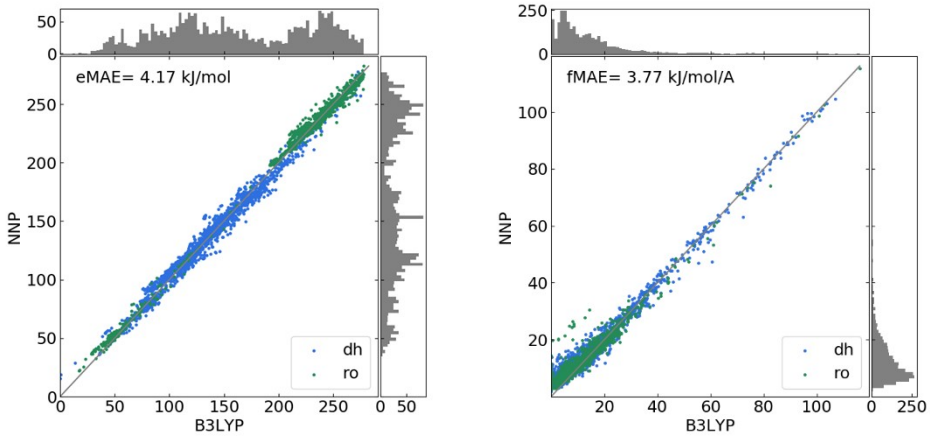
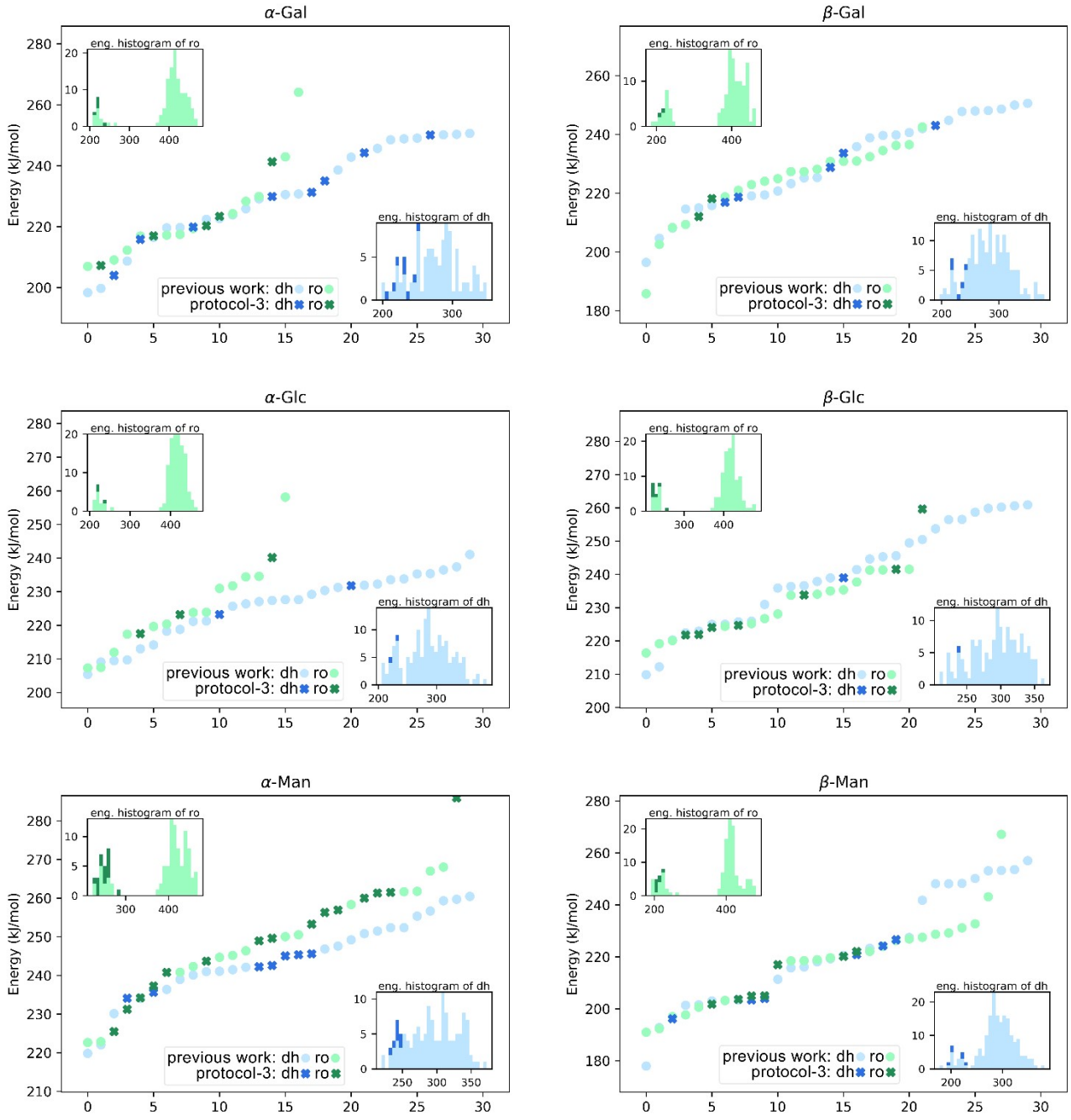
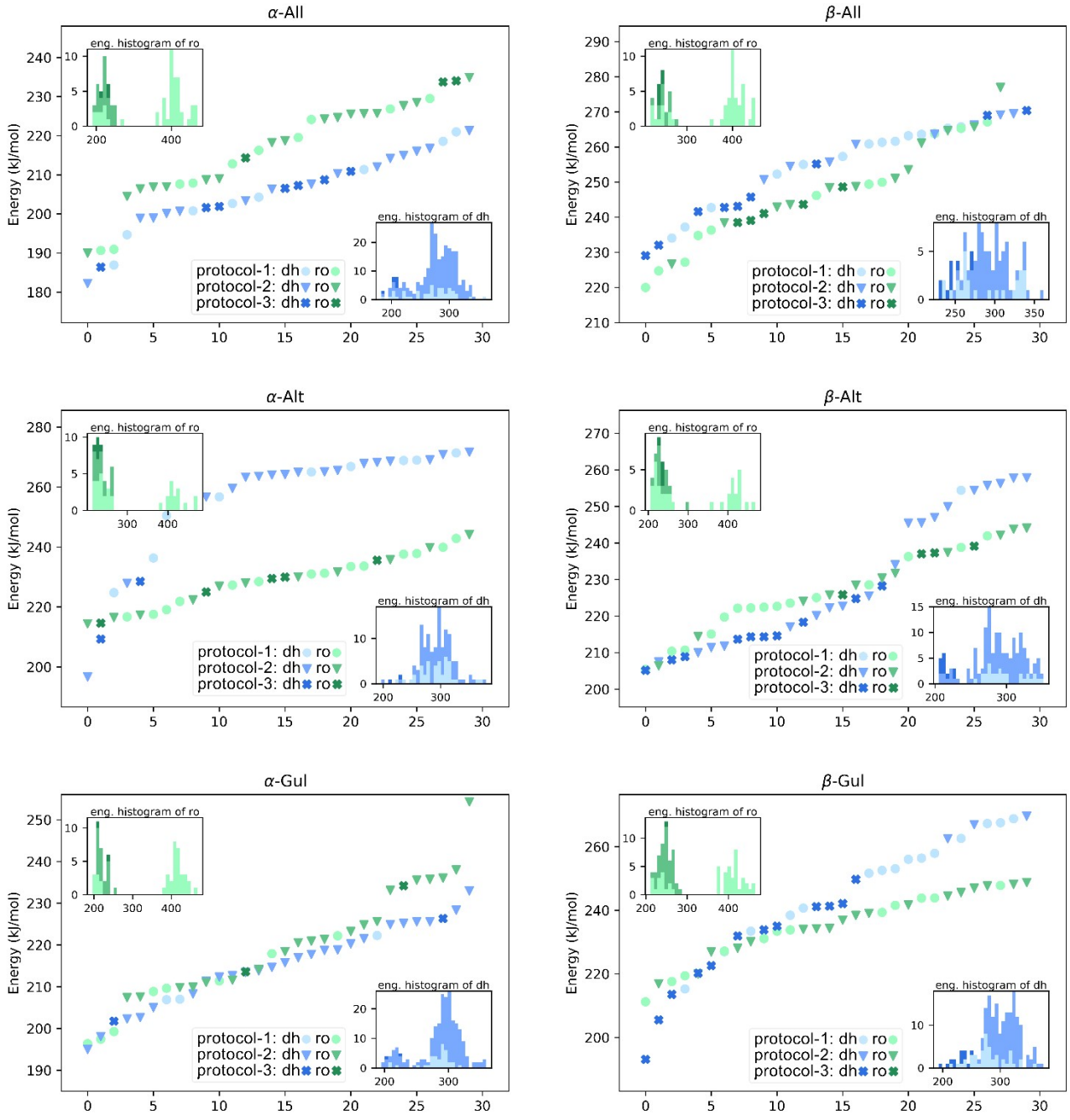


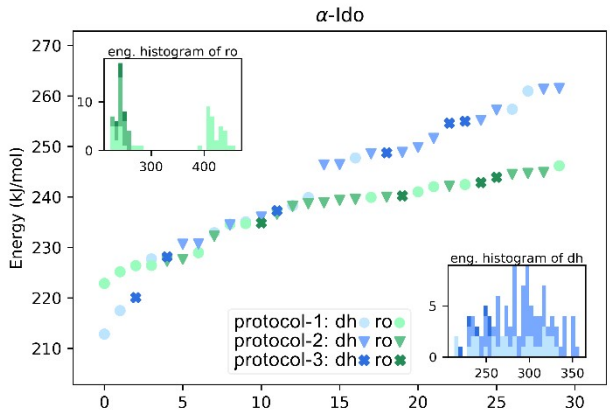
Figure S4. The plot for the 30 lowest energy TS-dh and TS-ro of AH-0 and AH-1. The TSs located by protocol-1, protocol-2, and protocol-3 are labeled with different notations. The energy histograms of all the TS-dh and TS-ro are shown in the subfigures.

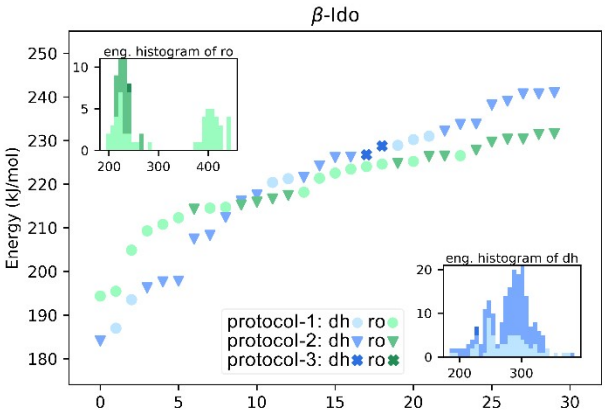
AH-0:

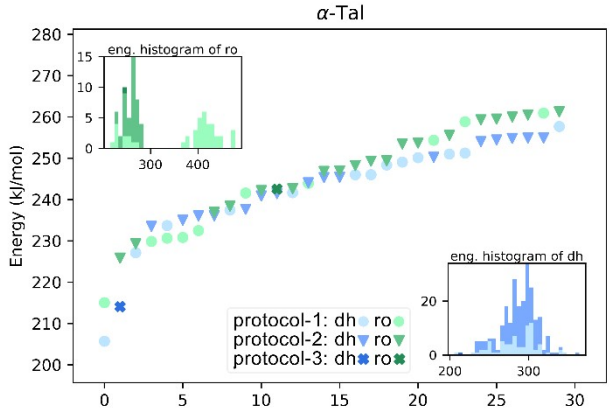


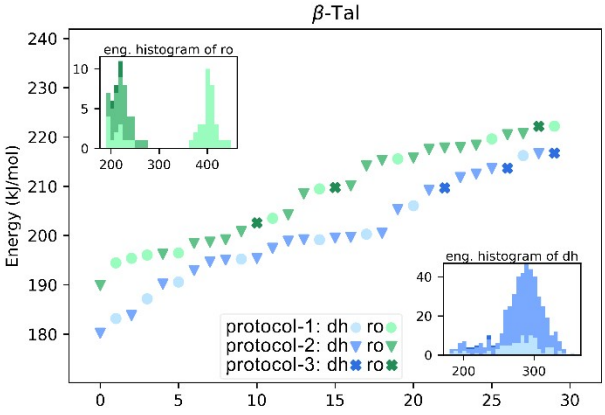
AH-1:


α-All
β-All
α-Alt
β-Alt
α-Gul
β-Gul
Energy (kJ/mol)
eng. histogram of ro
eng. histogram of dh
protocol-1: dh ro
protocol-2: dh ro
protocol-3: dh ro


α-Ido
Energy (kJ/mol)
eng. histogram of ro
eng. histogram of dh
protocol-1: dh ro
protocol-2: dh ro
protocol-3: dh ro


β-Ido
Energy (kJ/mol)
eng. histogram of ro
eng. histogram of dh
protocol-1: dh ro
protocol-2: dh ro
protocol-3: dh ro


α-Tal
Energy (kJ/mol)
eng. histogram of ro
eng. histogram of dh
protocol-1: dh ro
protocol-2: dh ro
protocol-3: dh ro


β-Tal
Energy (kJ/mol)
eng. histogram of ro
eng. histogram of dh
protocol-1: dh ro
protocol-2: dh ro
protocol-3: dh ro