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, 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

`

Evaluations of the NNP's prediction in each iteration

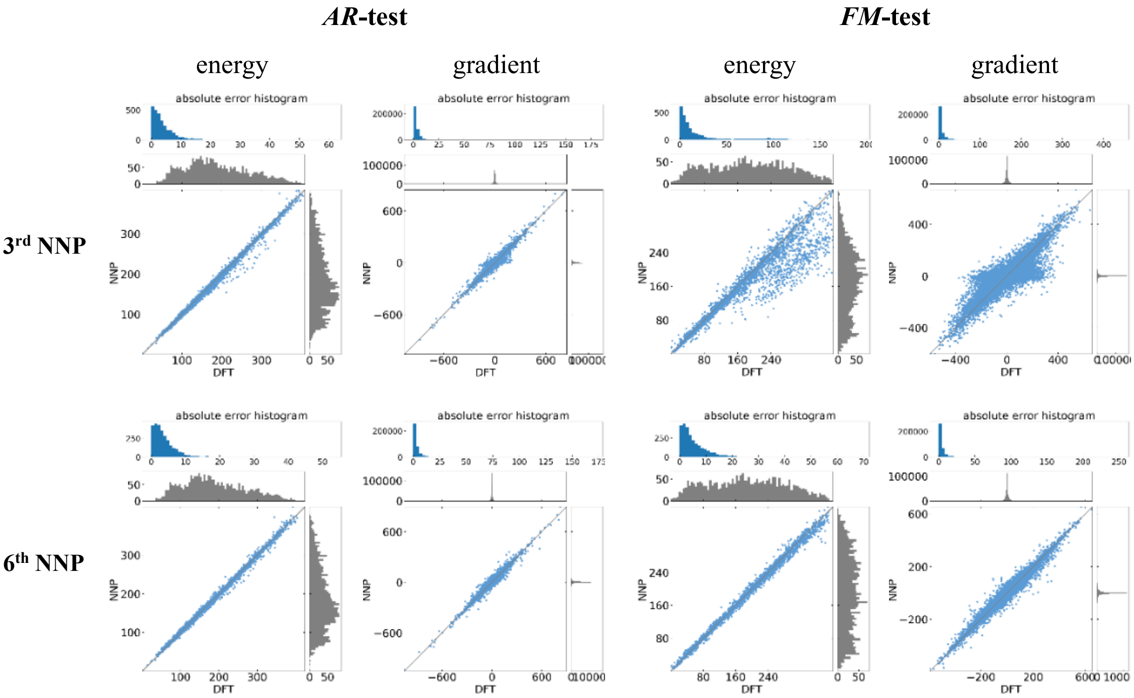
Table S1. The evaluations for the energy prediction. The units of energy are given in kJ mol^{-1}.

	AR-test					*FM*-test				
Iter.	DH	RO	C1Z1	B1Y1	Avg.	DH	RO	C1Z1	B1Y1	Avg.
1	9.3	4.7	5.9	6.8	6.7	70.5	10.6	24.6	15.8	36.9
2	6.0	3.2	3.7	3.9	4.1	47.9	8.6	13.3	8.3	24.4
3	6.1	3.6	3.0	4.2	4.2	47.9	8.0	10.6	9.0	23.8
4	5.0	3.4	2.9	4.5	4.1	9.1	6.1	5.1	6.2	7.1
5	4.8	2.8	2.9	4.7	3.9	7.6	4.4	4.6	5.3	5.8
6	4.1	3.6	2.9	4.2	3.8	7.5	4.0	3.8	4.5	5.4

Table S2. The evaluations for the gradient prediction. The units of gradient are given in kJ mol^{-1} Å^{-1}

	AR-test					*FM*-test				
Iter.	DH	RO	C1Z1	B1Y1	Avg.	DH	RO	C1Z1	B1Y1	Avg.
1	5.6	4.2	4.8	4.5	4.7	15.0	7.4	10.3	6.8	10.7
2	4.4	3.1	3.4	3.2	3.5	14.4	6.1	8.1	5.6	9.5
3	3.8	2.7	3.0	3.2	3.2	13.0	5.5	7.1	5.3	8.6
4	3.7	2.8	3.0	3.2	3.1	6.2	4.0	4.4	4.0	4.9
5	3.5	2.7	2.9	3.1	3.0	5.4	3.3	3.9	3.6	4.3
6	3.3	2.9	2.8	3.0	3.0	5.4	2.9	3.6	3.4	4.1

Figure S1. Scatter and histogram plots for the evaluations of the 3rd and 6th NNP iterations. The units are given in kJ mol^{-1}, kJ mol^{-1} Å^{-1} for energy and gradient, respectively.



The details of the *FM* sampling:

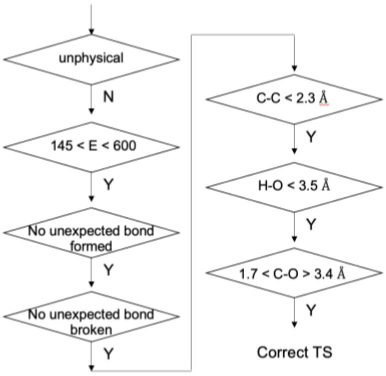
In *FM* sampling, the minima conformers explored in the previous studies were utilized to make the initial guess of TS [1]. On the assumption that the reactions are initiated from the proton transfer between the oxygen atoms, each hydrogen bond can be considered a reaction channel. Among the channels, the conformers with the Na^+ binding to the proton donor were chosen as the candidates. The proton donor's hydroxyl group was labeled as O_DH_D, while that of the proton accepter was O_AH_A. The C atoms bonding to O_A and O_D were labeled as C_A and C_D, respectively. For the DH, 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. Only the pathway of a proton from O1 to Oring was considered for the RO because it is the only favored mechanism. The initial guesses were generated with the O_{ring}—H_D shortened to 1.25 Å. For the B1Y1 and C1Z1 channels, the initial guess is generated with the bond length of C_A—O_A elongated to 2.1Å with all the atoms in the monomer residue moved. The distance of O_A—H_D is shortened to 1.1 Å.

The details of the pre-screening for TSs at the NNP level

Given the large number of TSs located at the NNP level, performing frequency and intrinsic reaction coordinate (IRC) calculations would be costly. Therefore, the extensive TS confirmation procedure is not performed; instead, the pre-screening is performed by checking the key bond length to remove the unreasonable or unrelated TS. The key bond length is defined as the length of the broken or the formed bond in the TS structures in each type of transition state.

The figure demonstrates the flowchart of the pre-screening procedure. The unphysical structures are first removed by checking any bond distance between the heavy atoms shortened than 1.2 Å and the distance between hydrogen atoms and heavy atoms shortened than 0.8 Å. The TSs are filtered with an energy range of 145-600 kJ/mol. Then, the structures with any unexpected bond formed or broken are removed. The reaction mechanism may involve the C–C, H–O, and C–O bond cleavage. The structures are examined by these bond lengths.

Scheme S1. The flowchart of the pre-screening.



The comparison of the TSs in this and the previous study [2]

The TSs structure taken from the study of Ni et al. was re-optimized using M06-2X/6-311+G(d,p), and the relative energies are calculated with the energy of the global minima of alpha-Maltose.

Table S3. The comparision between the barriers of the TS structures searched in previous and this study. The units of energy are given in kJ mol^{-1}.

Barrier	Dr. Ni's study	This work
DH	223.0	220.3
RO	207.0	204.8
B1Y1	228.0	212.1
C1Z1	299.0	269.1

The benchmark of the accuracy on the structures optimized using NNP and M06-2X

Table S4. Summary of the lowest-energy NNP-optimized TS (within ~25 kJ/mol per dissociation channel), re-optimized using M06-2X/6-311+G(d,p). Correct DFT-optimized TS was validated based on geometry and the imaginary frequency's normal mode. RMSD values between NNP-TS and DFT-TS are included.

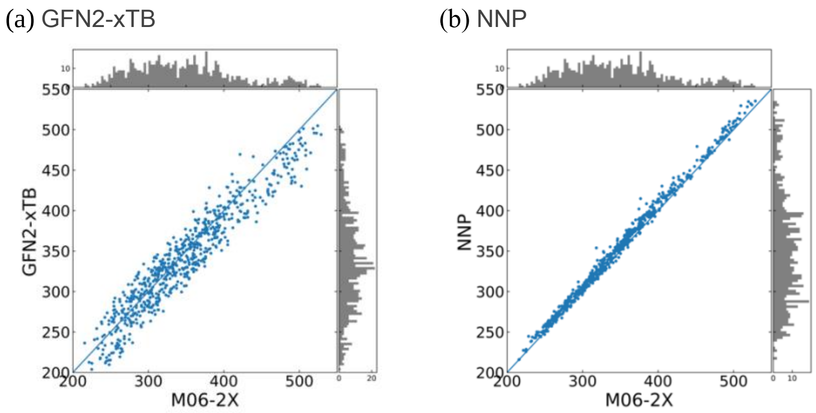
	# NNP-TS	# DFT-TS	# distinct	RMSD
DH	108	103 (95.4 %)	93 (86.1%)	0.131
RO	41	41 (100 %)	36 (87.8 %)	0.125
B1Y1	119	78 (65.5 %)	73 (61.3 %)	0.201
C1Z1	79	73 (92.4 %)	60 (75.9 %)	0.133

Table S5. Summary of 200 NNP-TS per dissociation channel, re-optimized using M06-2X/6-311+G(d,p). Correct DFT-optimized TS was validated based on geometry and the imaginary frequency's normal mode. RMSD values between NNP-TS and DFT-TS are included.

	# NNP-TS	# DFT-TS	# distinct	RMSD
DH	200	181 (90.5%)	181 (90.5%)	0.268
RO	200	193 (96.5%)	193 (96.5%)	0.247
B1Y1	200	118 (59.0%)	118 (59.0%)	0.488
C1Z1	200	168 (84.0 %)	167 (83.5%)	0.363

Energetic Benchmarking of GFN2-xTB and NNP Against M06-2X

Figure S2. Benchmarking GFN2-xTB and NNP against M06-2X. Single-point calculations were performed on TSs optimized at the M06-2X/6-311+G(d,p) level. The units of energy are given in kJ mol^{-1}.



The structural features of theTSs@NNP

Figure S3. The energy diagram of the Puckering configurations of reducing (upper) and non-reducing (lower) end. The units of energy are given in kJ mol^{-1}.