Electronic Supplementary Information
For

Searching low-energy conformers of neutral and protonated di-, tri-, and tetra-glycine with first-principle accuracy assisted by the use of neural network potentials

Dong Cao Hieu,[a,b,c] Po-Jen Hsu,[a,] 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. International Graduate Program of Molecular Science and Technology (NTU-MST), National Taiwan University, Taipei 10617, Taiwan
d. Department of Chemistry, National Tsing Hua University, Hsinchu 30013, Taiwan.

***To whom correspondence should be addressed. E-mail:**

jlkuo@gate.sinica.edu.tw (Jer-Lai Kuo)

Contents

Figure S1. Analyze ΔE1, ΔF1, ΔE2, and RMSD between min structures by NNP and M06-2X.

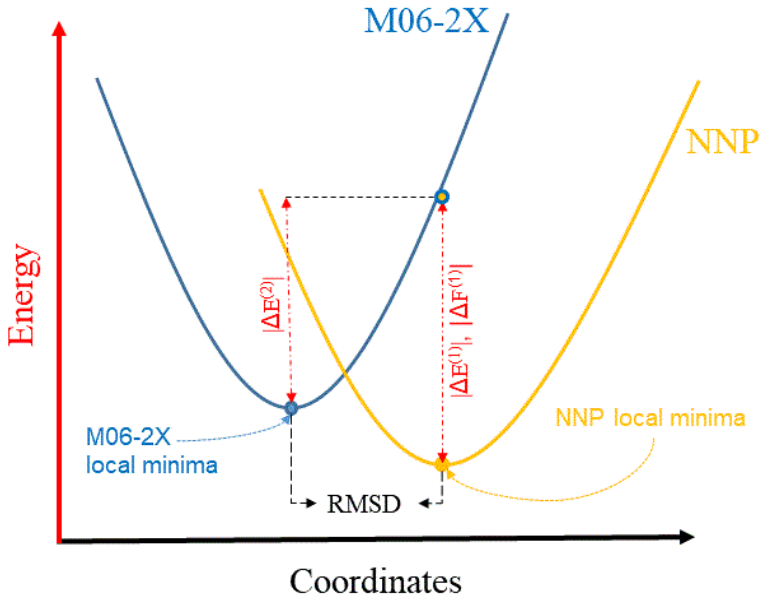


Figure S2. M06-2X minima of Gly_n_2-4 within 0-10 kJ/mol of ZPE-corrected energy.

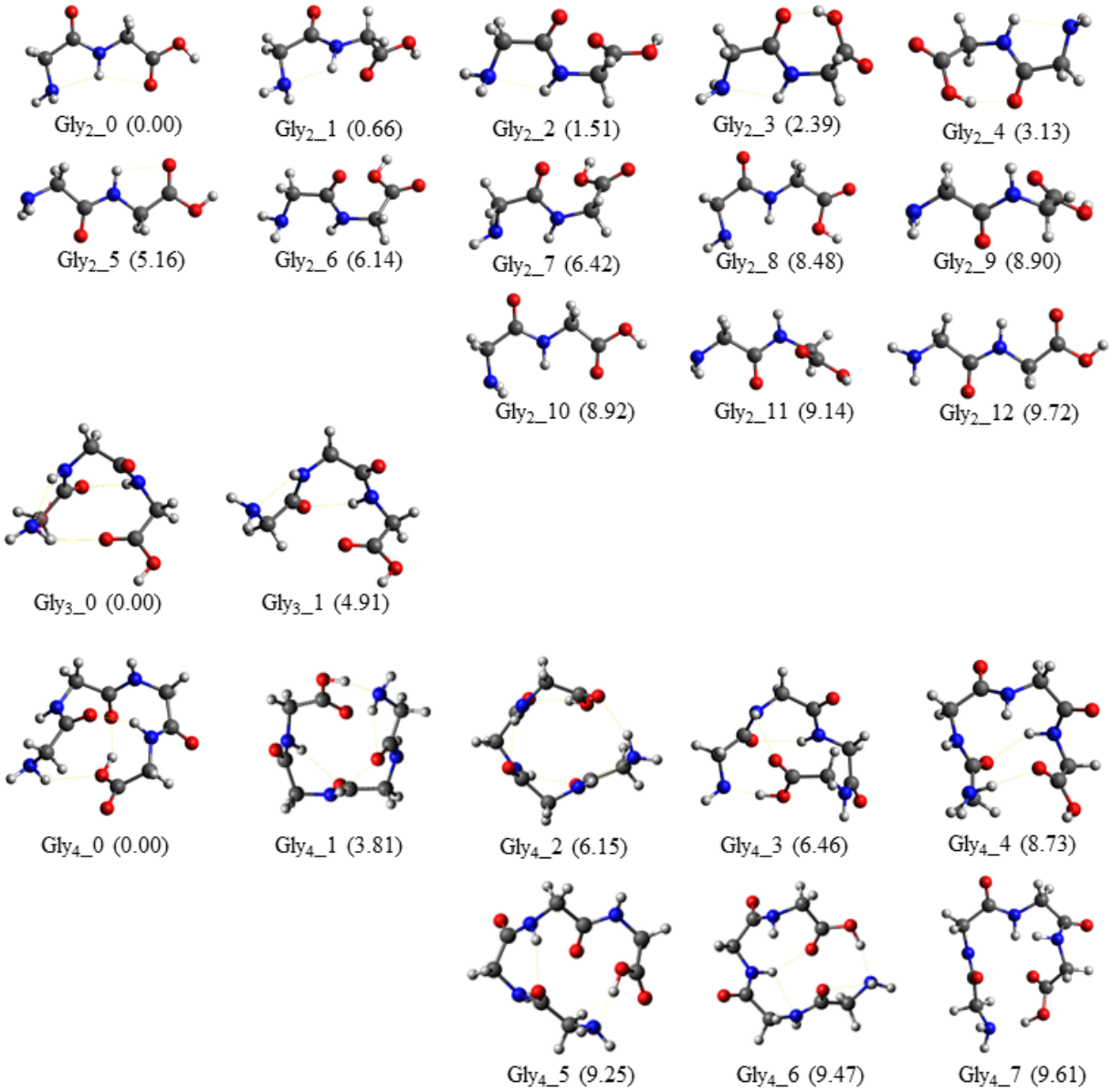


Figure S3. M06-2X minima of H^+Gly_n_2-4 within 0-10 kJ/mol of ZPE-corrected energy.

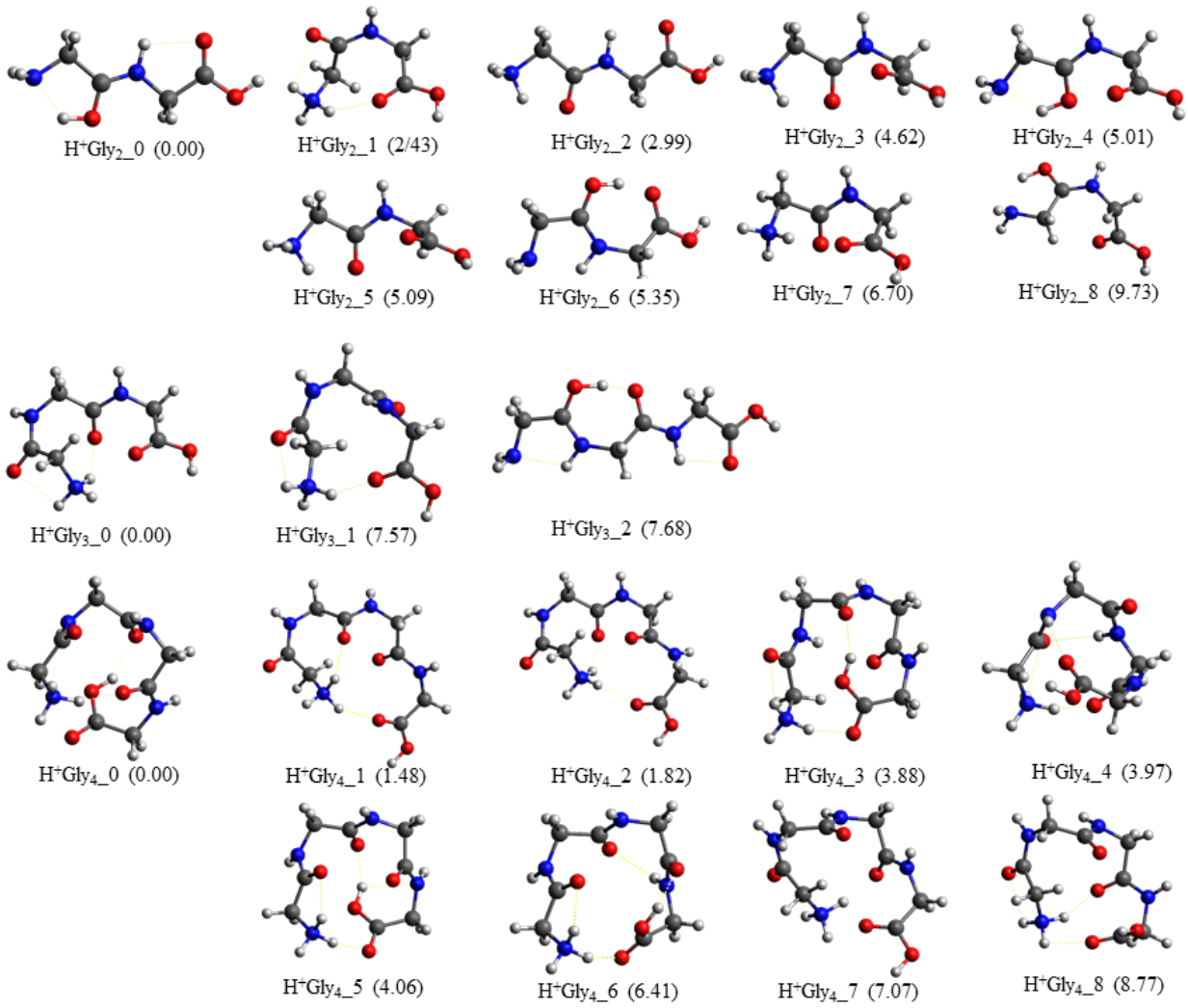


Figure S4. Root Mean Square Deviation (RMSD) between NNP-2 and M06-2X minima (Å).

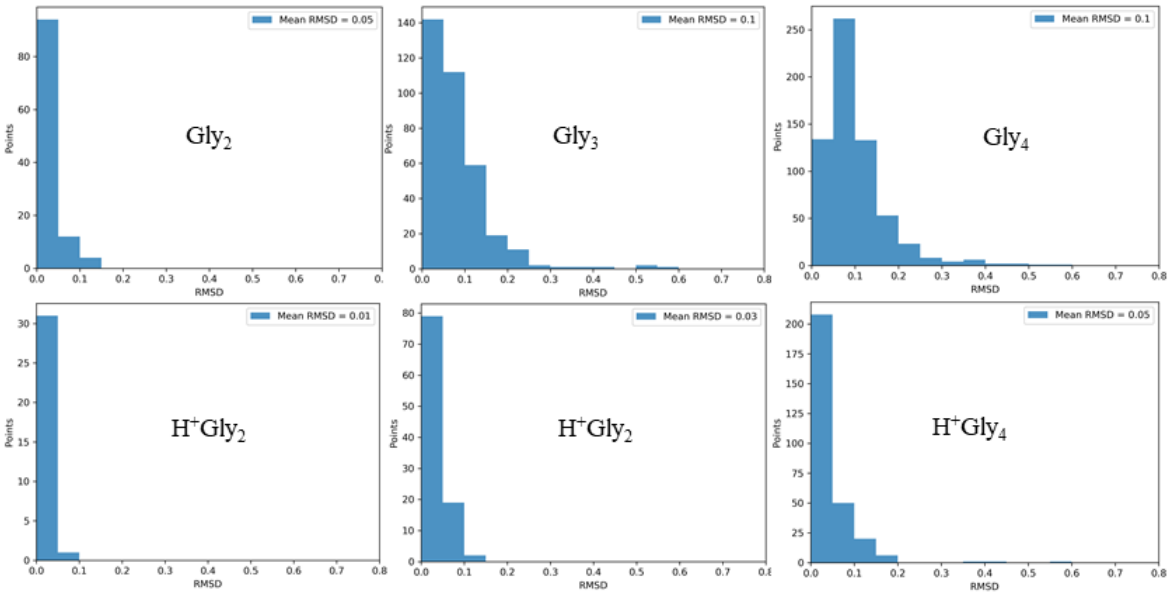


Figure S5. The correlation between NNP-2 energy (y-axis) and M06-2X energy (x-axis) of all NNP-2 stable configurations within 0-50 kJ/mol. The zero of the relative energies is calculated with respect to the lowest M06-2X energy (ΔE1 in Table 2).

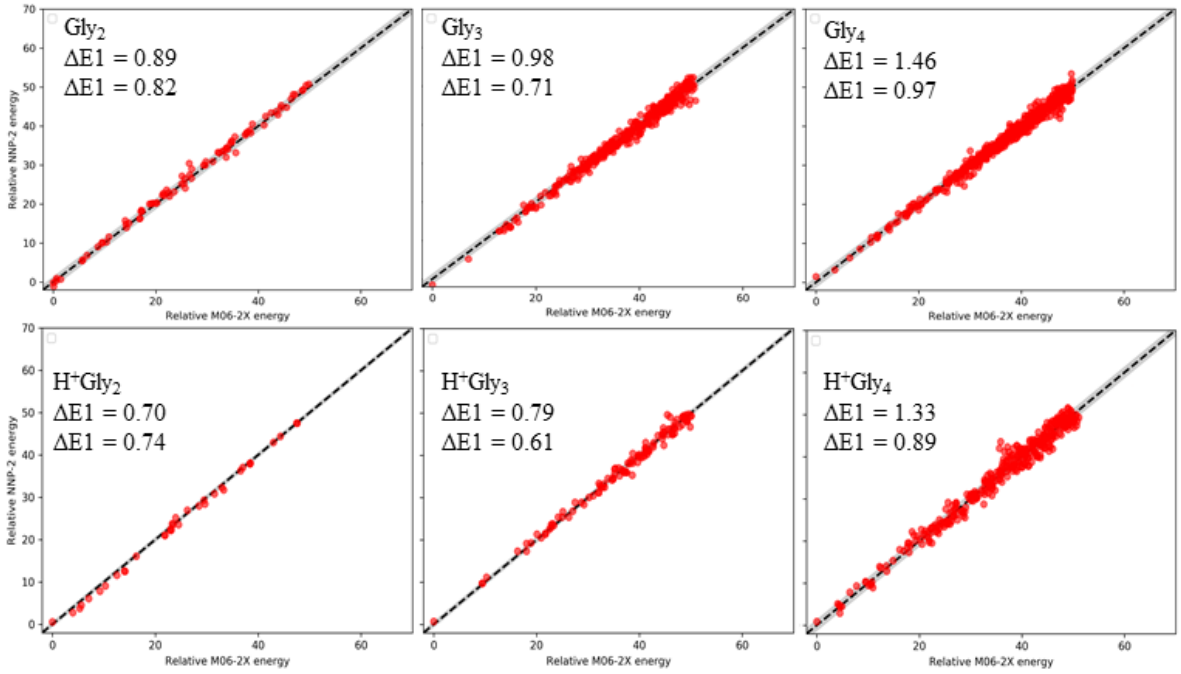


Figure S6. The energy histogram of Gly_n (left) and H+Gly_n (right) at M06-2X level.

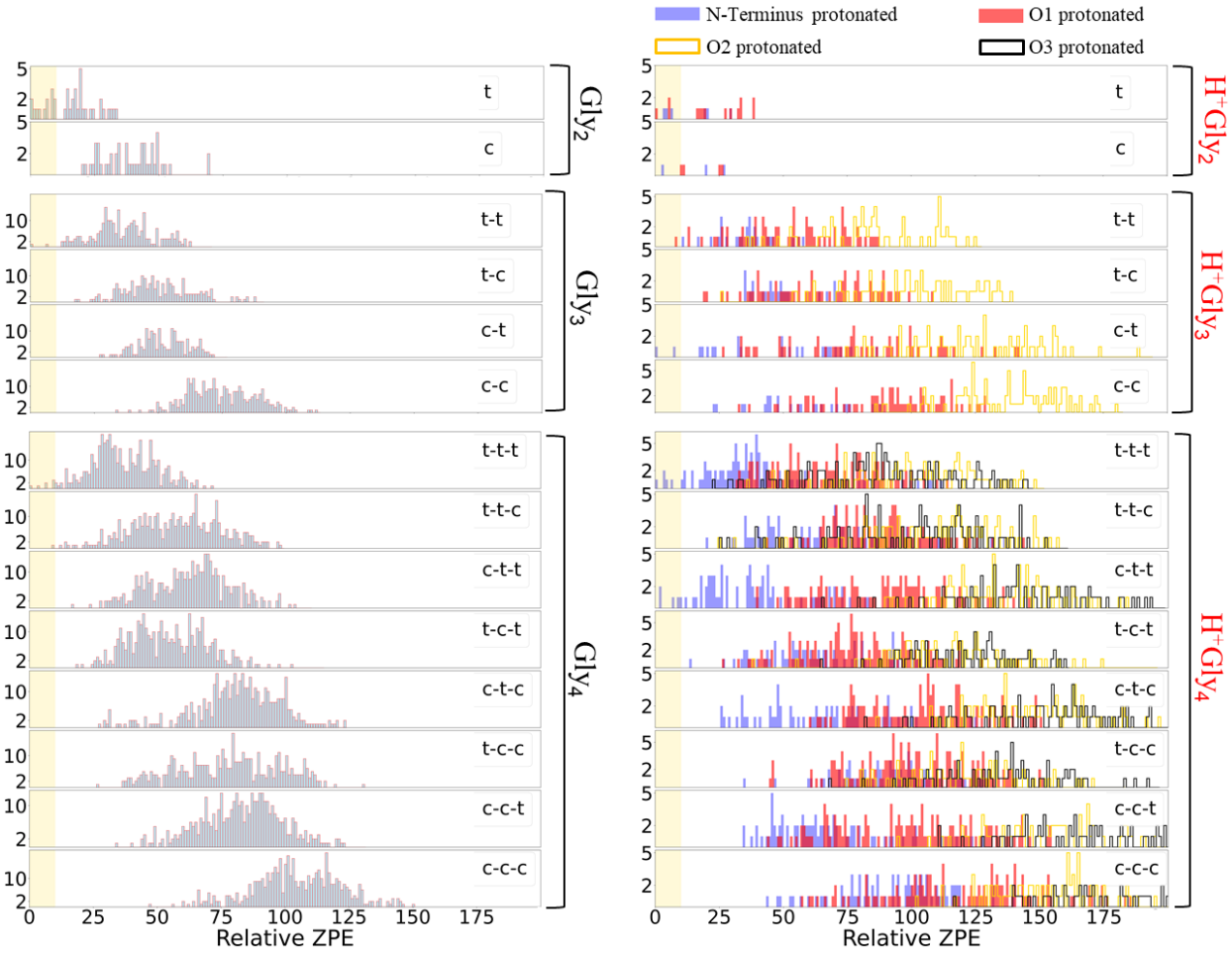
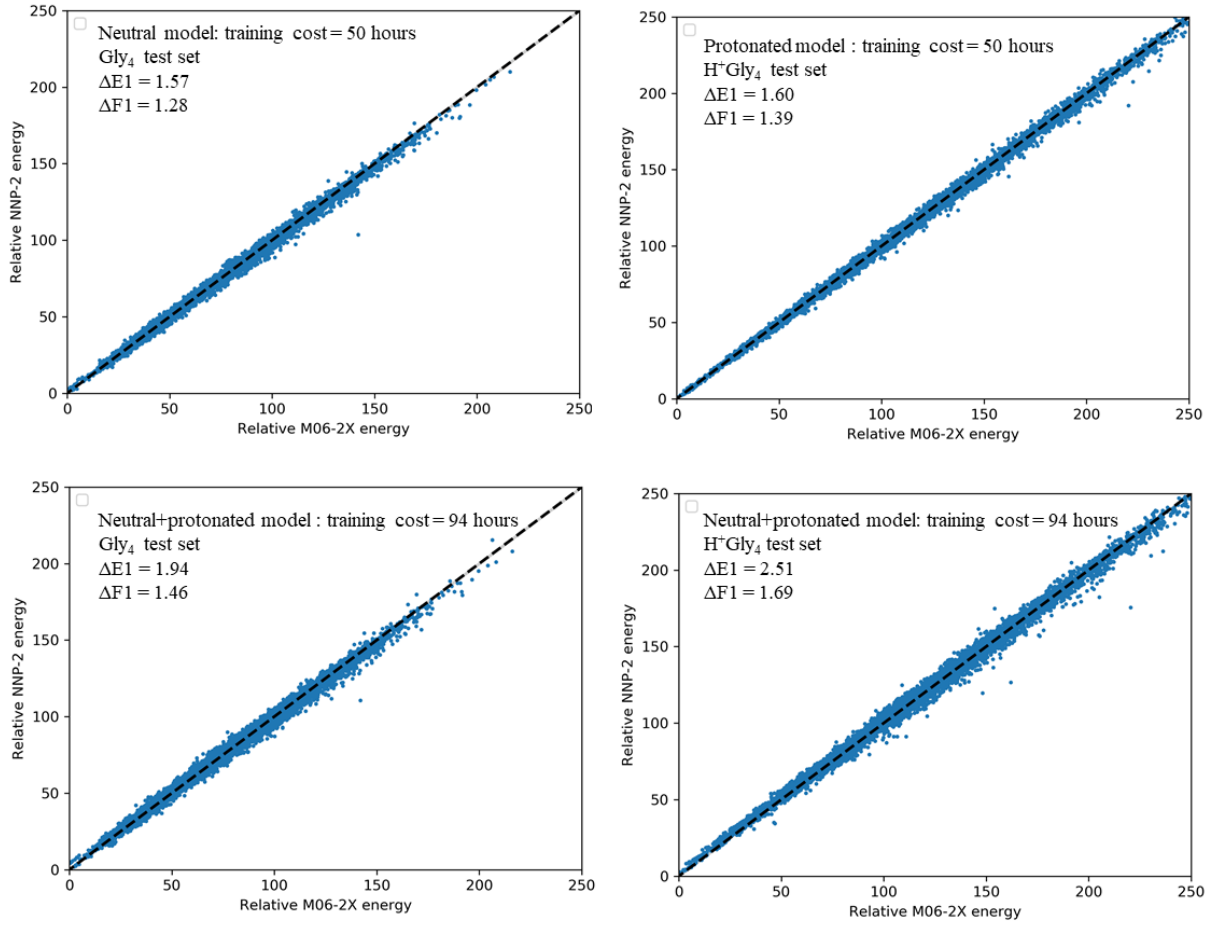


Figure S7. The prediction of the NNP-2 neutral model (top-left) and NNP-2 protonated model (top-right) model on the test set of Gly_4 and H^+Gly_4. The prediction of one neutral+protonated model (bottom) on the test set of Gly_4 and H^+Gly_4. The performance of the NNP-2 neutral or NNP-2 protonated model surpasses that of the neutral_protonated model trained on neutral+protonated data. The training duration for the NNP-2 neutral model or NNP-2 protonated model, which is 50 hours, is shorter than that of the neutral_protonated model, which takes 94 hours.



Amplitudes		0.1π	0.2π	0.3π	0.4π	0.5π
Gly_3	Trial structures	20k	20k	20k	20k	20k
	DFTB-3 min	3618	4117	4134	4082	3942
	DFTB-3 min (0-50)	3276	3682	3883	3484	3687

Table S1. Number of distinct DFTB-3 minima with varying amplitudes from 0.1π to 0.5π for Gly_3

Glycine	Gly_2	Gly_3	H^+Gly_2	H^+Gly_3
camB3LYP/def2TZVPP	55	119	48	107
M06-2X/6311+G(d,p)	87	146	55	110
MP2/6-311+G(d,p)	85	139	50	95

Table S2. Number of distinct minima at cam-B3LYP/def2TZVPP, M06-2X/6-311+G(d,p), and MP2/6-311+G(d,p) level (from 400 DFTB-3 minima)

	Name of peptide	NNP-0	NNP-1	NNP-2	Test set
The training set for Gly_n	Gly-2	1450	6543	6543	1613
	Gly-3	5622	15768	25263	5724
	Gly-4	8190	23433	27908	7607
The training set for H^+Gly_n	H^+Gly-2	1700	6752	6752	1629
	H^+Gly-3	4418	18593	25759	4262
	H^+Gly-4	6736	25863	27385	6915

Table S3. Number of training sets of the model's generations