Supplementary Information

Unveiling sodium storage mechanisms in hard carbon via machine learning-driven simulations with accurate site occupation identification

Zhaoming Wang,[ac] Guanghui Shi,[ac] Guanghui Wang,[ac] Man Wang,[ac] Feng Ding,[*ab] Xiao Wang[*ab]

a Institute of Technology for Carbon Neutrality, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

b Suzhou Laboratory, Suzhou 215000, China

c Nano Science and Technology Institute, University of Science and Technology of China, Suzhou 215123, China

Email: dingf@szlab.ac.cn; xiao.wang@siat.ac.cn

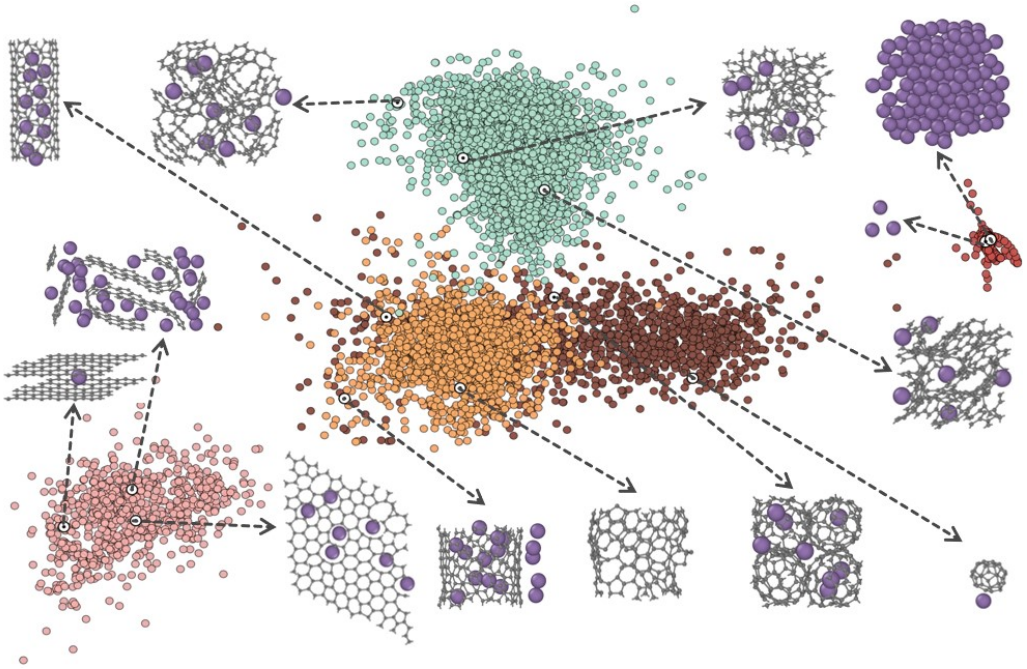

Fig. S1. Visualization of the dataset with component distribution. Each color represents a distinct structural type: green spheres denote amorphous carbon structures and their associated Na-C composite structures; pink spheres indicate layered carbon structures (e.g., graphite, graphene) and their Na-C composites; brown spheres correspond to spherical carbon structures (e.g., fullerenes) and their Na-C composites; yellow spheres denote tubular carbon structures (e.g., carbon nanotubes) and their Na-C composites; red spheres represent pure sodium systems, including sodium atoms, clusters, and bulk phase.

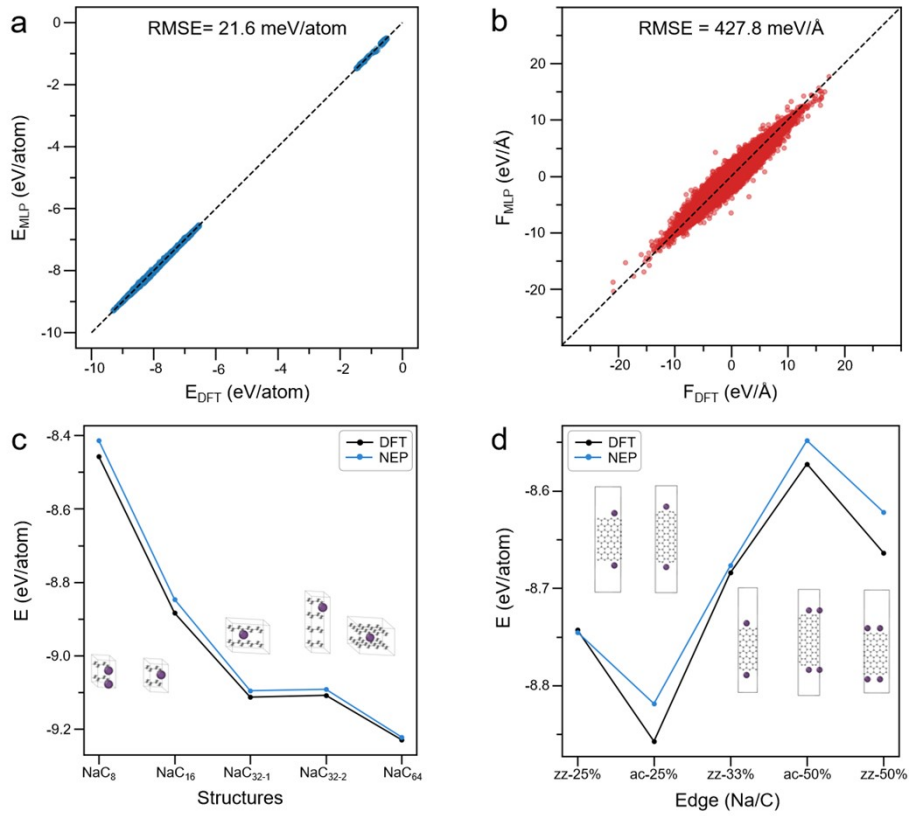

Fig. S2. MLP testing and validation. (a) Energy deviation; (b) Force deviation; (c) Energy profiles of various sodium-carbon structures calculated using density functional theory (DFT) and neuroevolution (NEP) potential methods. (d) Energy profiles of the adsorption of sodium at the graphene zigzag (zz) and armchair (ac) edges with varying Na/C ratios calculated by DFT and NEP methods. The percentage represents the atomic ratio between sodium and edge carbon atoms.

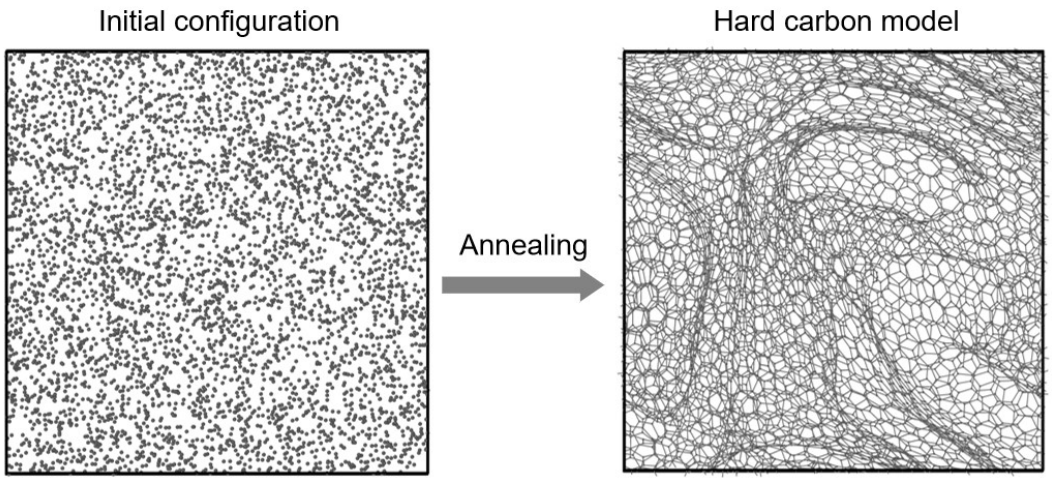

Fig. S3. The formation process of hard carbon (HC) model.

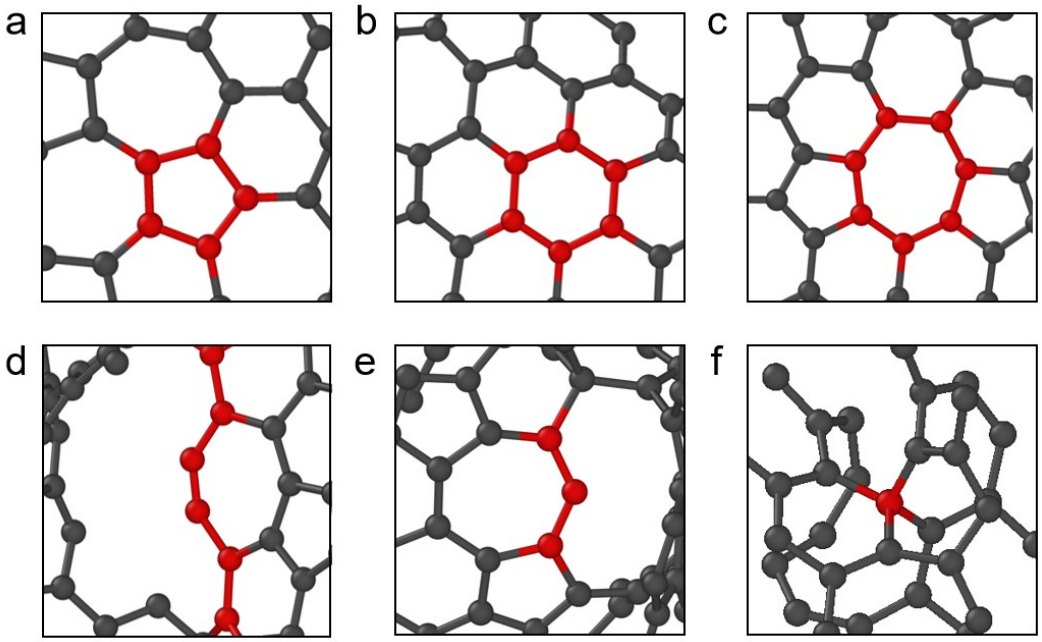

Fig. S4. The local feature of the HC model. (a) pentagon, (b) hexagon, (c) heptagon, (d) armchair edge, (e) zigzag edge, (f) amorphous carbon.

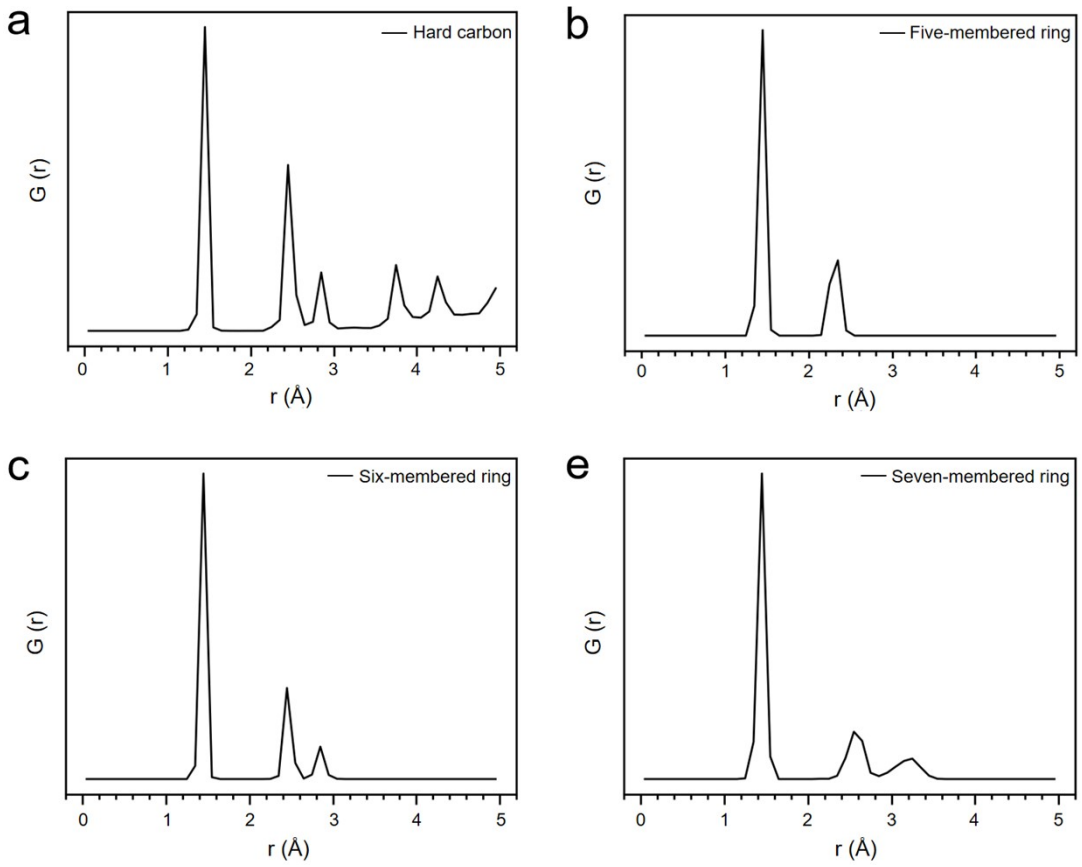

Fig. S5. The radial distribution function (RDF) of the (a) HC model and the extracted (b) five-membered, (c) six-membered and (d) seven-membered rings within the HC model.

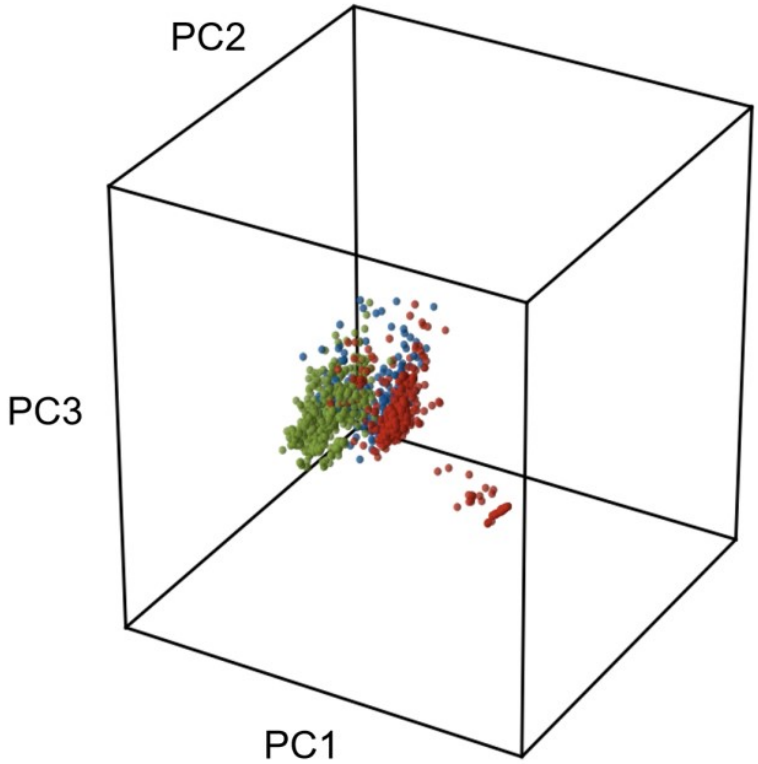

Fig. S6. The principal component analysis (PCA) on the chosen features, where blue, green, and red spheres denote adsorbed, intercalated, and filled sodium atoms, respectively.

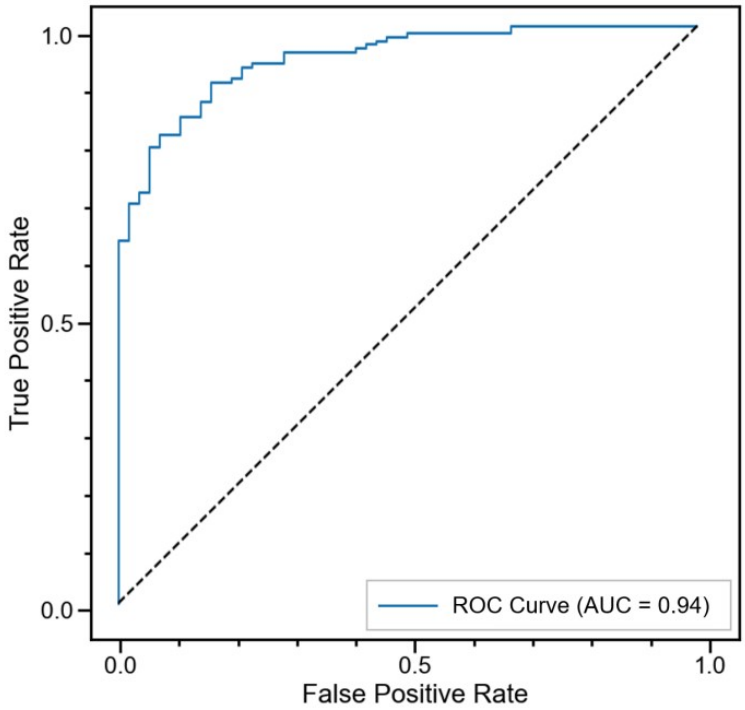

Fig. S7. The receiver operating characteristic (ROC) curve based on our trained random forest (RF) algorithm.

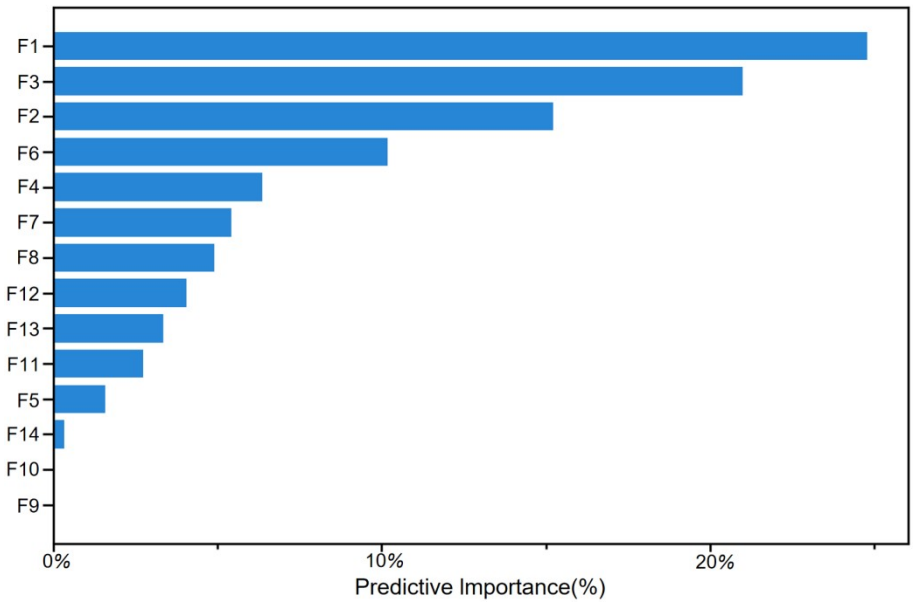

Fig. S8. The feature importance evaluation of the RF model. F1/F2: coordination number of sodium ions / carbon atoms, F3/F4: distance to the nearest sodium ion/carbon atom, F5: number of fitted carbon planes, F6: maximum flatness of fitted carbon planes, F7: distance to the structure formed by projecting carbon network onto fitted planes, F8: dihedral angle between fitted carbon planes, F9-F14: numbers of 3-8-membered carbon rings.

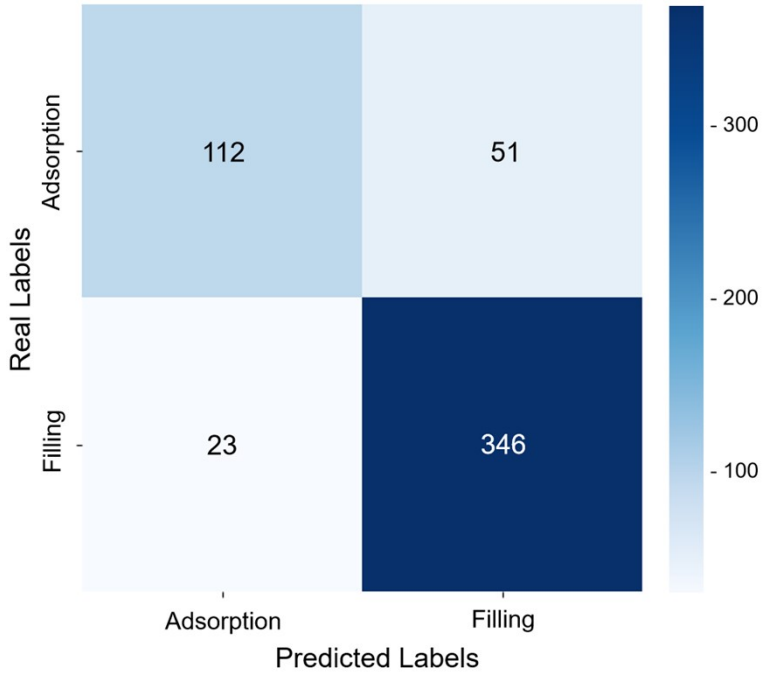

Fig. S9. The confusion matrix based on the RF model.

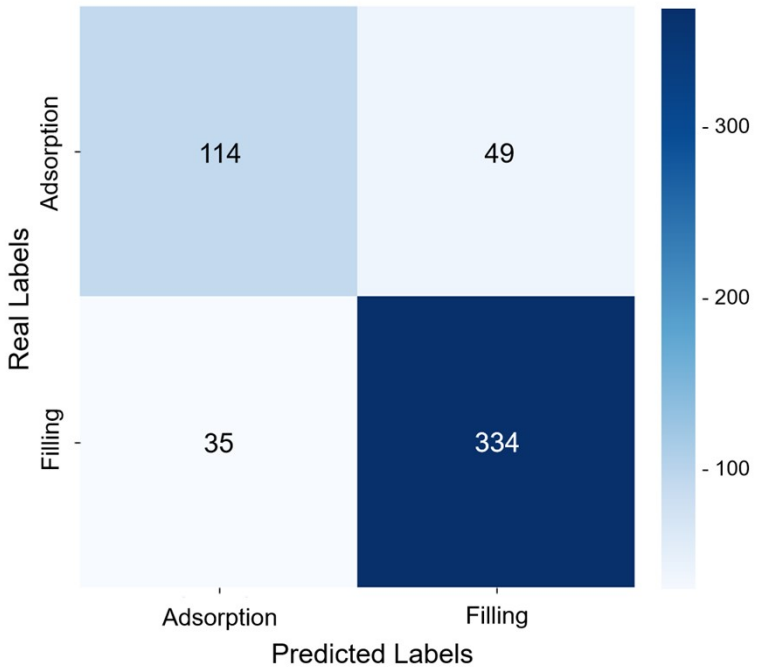

Fig. S10. The confusion matrix based on supporting vector machine (SVM) model.

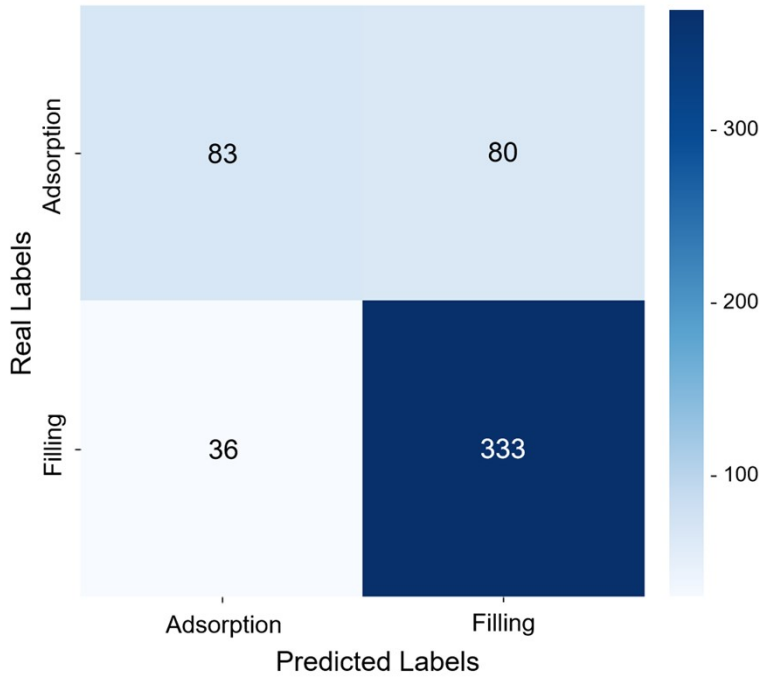

Fig. S11. The confusion matrix based on the clustering model.

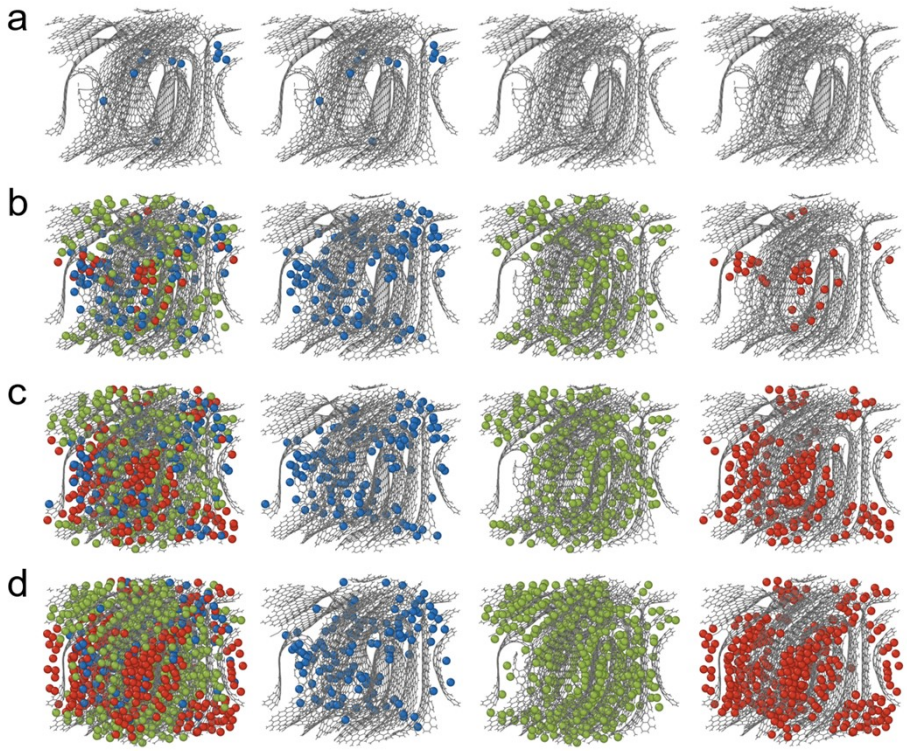

Fig. S12. The visualization of sodium ion occupation site distribution when (a) 11 (adsorption only stage), (b) 327 (intercalation and adsorption stage) and (c) 585 (pore-filling dominated stage) and (d) 1000 sodium ions (overcharging stage) were inserted into the HC model.

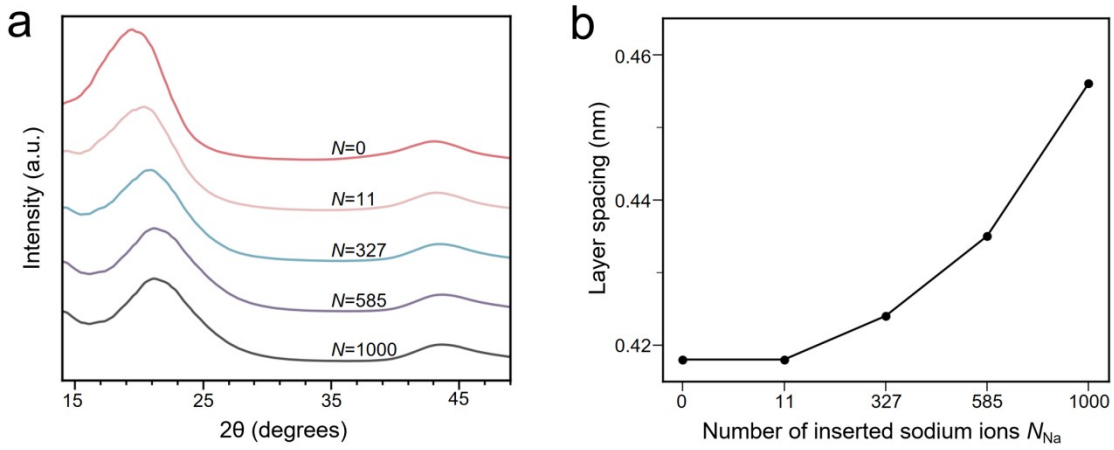

Fig. S13. (a) X-ray diffraction (XRD) patterns of the HC model and the (b) corresponding interlayer spacing with the number of inserted sodium atoms.

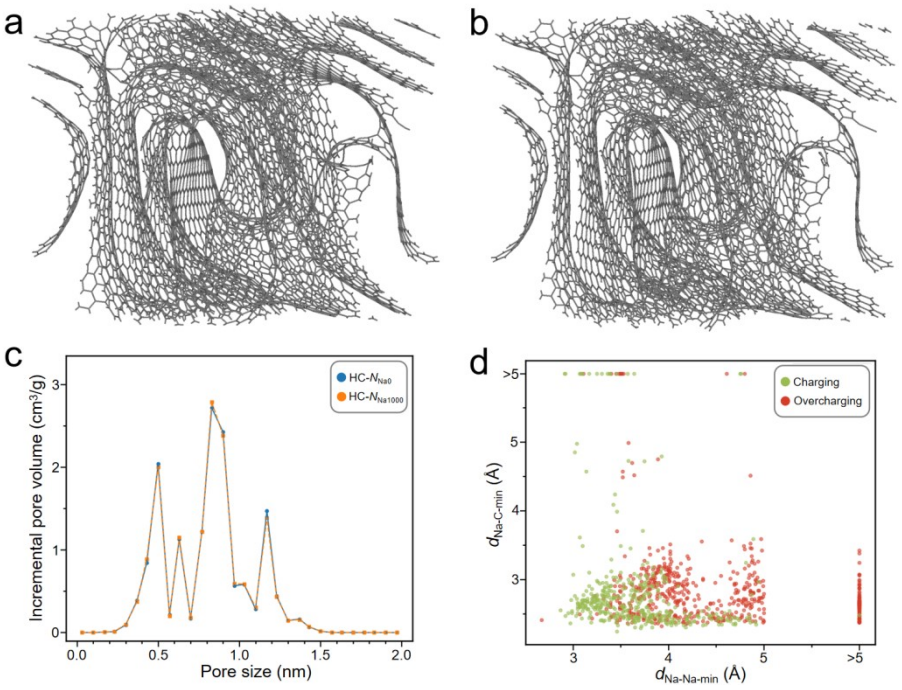

Fig. S14. The structural evolution of HC at the overcharging stage. (a) The view of HC before (a) and after (b) the overcharging stage. (c) Variation of pore distribution after overcharging stage. (d) The distribution of the shortest Na-Na and Na-C distance at the charging and overcharging stage.

Table S1. The composition and distribution of the MLP dataset.

Structures		Training set	Test set
Na	Sodium atoms	83	9
	Sodium cluster	90	10
	Sodium bulk	48	5
C	Amorphous carbon	440	49
	Layered carbon	50	6
	Spherical carbon	197	22
	Tubular carbon	173	19
Na_xC	Sodium-amorphous carbon	1853	205
	Sodium-layered carbon	636	71
	Sodium-spherical carbon	812	90
	Sodium-tubular carbon	1183	132
Total		5565	618

Table S2. The contribution of the 14 structural features for the three principal components (PC1, PC2, and PC3).

	PC1	PC2	PC3
F1	0.3271	-0.1748	0.2792
F2	-0.4791	-0.0595	-0.0940
F3	-0.2807	0.1295	-0.3967
F4	0.3904	0.0014	-0.2579
F5	-0.3806	-0.1967	0.3113
F6	-0.0699	0.4828	-0.0747
F7	0.0964	0.0276	-0.6411
F8	0.2654	-0.2752	0.0572
F9	-0.0103	0.2556	0.0996
F10	-0.0246	-0.0157	-0.0219
F11	-0.1587	0.4152	0.2572
F12	-0.3832	-0.2711	-0.2617
F13	-0.1630	0.4193	0.1328
F14	-0.0587	0.3369	-0.0996

Table S3. Distribution of each type of site across different stages.

Sites	Adsorption only stage	Intercalation and adsorption stage	Pore-filling dominated stage
Adsorption	8.5%	79.2%	12.3%
Intercalation	0	67.4%	32.6%
Pore-filling	0	16.8%	83.2%