Supplementary information for

Complex ligand adsorption on 3D atomic surfaces of synthesized nanoparticles investigated by machine-learning accelerated ab initio calculation

Dohun Kang^{1,2, #}, Sungin Kim^{1,3, #}, Junyoung Heo^{1,3}, Dongjun Kim¹, Hyeonhu Bae⁴, Sungsu Kang^{1,3}, Sangdeok Shim^{5,*}, Hoonkyung Lee^{4,*}, Jungwon Park^{1,3,6,7,*}

¹ School of Chemical and Biological Engineering, and Institute of Chemical Processes, Seoul National University, Seoul 08826, Republic of Korea

² Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA

³ Center for Nanoparticle Research, Institute for Basic Science (IBS), Seoul 08826, Republic of Korea

⁴ Department of Physics, Konkuk University, Seoul 05029, Republic of Korea

⁵ Department of Chemistry, Sunchon National University, Suncheon 57922, Republic of Korea

⁶ Institute of Engineering Research, College of Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea

⁷ Advanced Institutes of Convergence Technology, Seoul National University, 145, Gwanggyoro, Yeongtong-gu, Suwon-si, Gyeonggi-do, 16229, Republic of Korea

*Correspondence to: san90@scnu.ac.kr, hkiee3@konkuk.ac.kr, or jungwonpark@snu.ac.kr.

This PDF file includes:

Methods Supplementary references Supplementary Figures S1-S13 Supplementary Tables S1-S2

Methods

- Brownian one-particle reconstruction of synthesized Pt nanoparticles: Graphene liquid cell was fabricated by encapsulating the synthesized Pt nanoparticle solution in two graphene-supported transmission electron microscopy (TEM) grid.¹ In situ movies of Pt nanoparticles in liquid were obtained with TEAM I, an FEI Titan 80/300 TEM equipped with a post-specimen geometric- and chromatic-aberration corrector and Gatan K2 IS direct electron detector. Thousands of images with high temporal (400 frames/s) and spatial (0.358 Å pixel size) resolution were acquired at a dose rate of ~15 e⁻/Å² ·frame with an acceleration voltage of 300 kV. Pt nanoparticles showed tumbled dynamics in solution, which resulted in varying lattice information in a series of TEM images.^{1, 2} TEM movies of synthesized Pt nanoparticles were applied to 3D SINGLE (Structure Identification of Nanoparticles by Graphene Liquid cell Electron microscopy) based on SIMPLE/PRIME algorithm for ab-initio 3D reconstruction of a single particle protein structure.²⁻⁴ Then 3D density maps were obtained. 3D coordinates of constituent individual atoms were assigned to local maxima in density maps. The 3D density maps and atom position maps for the six nanoparticles studied here can be found in the previous publication.²

- **Computational details:** We utilized first-principles density functional theory (DFT) calculations as implemented in Vienna ab-initio simulation package $(VASP)^{5-8}$ with Perdew-Burke Ernzerhof (PBE) generalized gradient approximation (GGA)⁹ exchange-correlation functional and projector-augmented wave $(PAW)^{10,11}$ pseudopotentials. Basic plane waves were expanded with a cutoff energy of 400 eV. We included van der Waals (vdW) interaction with DFT-D3 method by Grimme.¹² Only Γ -point scheme was used for calculations. The vacuum

space was imposed to secure 10 Å within periodic images for each system. The adsorption energy (E_{ads}) was calculated with equation below:

$$E_{\rm ads} = E_{\rm *PVP} - (E_{\rm *} + E_{\rm PVP}) \tag{1}$$

The vdW interaction energy portion of adsorption energy, E_{vdw} is calculated with the same equation (1) using Evasp correction value in VASP output file. The short-range direct bonding energy E_{bind} is derived by subtracting E_{vdw} from E_{ads} .

- Calculation for CN and \overline{CN} : CN of an atom of interest in fcc nanoparticles is calculated as the number of neighboring atoms within specific distance *d* that is presented in equation (2):

$$d = \frac{d_1 + d_2}{2} \tag{2}$$

, where d_1 and d_2 are distance from an atom of interest to first and second nearest neighbor atoms, respectively. The d_1 and d_2 are defined by

$$d_1 = \frac{a_0}{\sqrt{2}} \tag{3}$$

$$d_2 = a_0 \tag{4}$$

, where a_0 is lattice constant of nanoparticle which is 3.92 Å for Pt. \overline{CN} of an atom i surrounded by the first nearest neighbor atoms of j is defined by equation below:

$$\overline{CN} = \sum_{k=1}^{j} \frac{CN(k)}{CN_{max}}$$
(5)

, where CN_{max} and CN(k) is the maximum CN of the first nearest neighbor atoms and CN of

each first nearest neighbor atom, respectively.

- Symmetry Function and Neural Network: The radial part and the angular part of the symmetry function were derived from G^2 and G^5 functions in Behler 2011^{4413} with the following equations.

For each surface atom or characteristic point i,

$$G_i^2 = \sum_j e^{-\eta (R_{ij} - R_s)^2} \cdot f_c(R_{ij})$$
(6)

$$G_{i}^{5} = \sum_{j,k\neq i}^{all} (1 + \cos(\theta_{ijk} - \theta_{s}))^{\zeta} \cdot e^{-\eta(\frac{R_{ij} + R_{ik}}{2} - R_{s})^{2}} \cdot f_{c}(R_{ij}) \cdot f_{c}(R_{ik})$$
(7)

where,
$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot \left[\cos\left(\frac{R_{ij}}{R_c}\right) + 1\right] & \text{for } R_{ij} \le R_c \\ 0 & \text{for } R_{ij} > R_c \end{cases}$$

 R_{ij} is the distance between i and neighboring atom j, and θ_{ijk} is the angle between R_{ij} and R_{ik} . R_c is a cutoff radius for cutoff cosine function, f_c . Twelve different radial shifts (R_s , [1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7]) for G² and seven R_s ([0.84, 1.40, 1.96, 2.52, 3.09, 3.65, 4.21]) with four angular shifts (θ_s , [0.9, 1.55, 2.2, 2.85]) are set for G⁵ functions to calculate 40 symmetry functions. R_c for G² and G⁵ are set to 7 and 5 respectively and η for G² and G⁵ are set to 16 and 8 respectively. ζ is set to 32.

Two-layered fully connected neural network for the prediction of PVP adsorption energy was implemented using PyTorch.¹⁴ The model was trained with AdamW optimizer¹⁵ and mean square error (MSE) was set as a loss function. The learning rate of the optimizer was modulated by reduce on a plateau learning rate scheduler.

- Calculation for bimodality index: bimodal index is calculated by the method from BimodalIndex package included in R software.¹⁶ Following bimodal expression is estimated using this algorithm.

$$y = \pi N(\mu_1, \sigma) + (1 - \pi) N(\mu_2, \sigma)$$
 (8)

, where y is adsorption energy, π is the proportion of two distribution, μ_1 and μ_2 are the means of the two distributions and σ is the common standard deviation. From these evaluated parameters, bimodality index is defined by equation below:

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma} \tag{9}$$

Bimodality Index =
$$[\pi(1-\pi)]^{\frac{1}{2}}\delta$$
 (10)

, where $\boldsymbol{\delta}$ is the standardized distance between two distributions.

Supplementary References

1. J. M. Yuk, J. Park, P. Ercius, K. Kim, D. J. Hellebusch, M. F. Crommie, J. Y. Lee, A. Zettl, and A. P. Alivisatos, *Science* **2012**, *336*, 61-64.

2. B. H. Kim, J. Heo, S. Kim, C. F. Reboul, H. Chun, D. Kang, H. Bae, H. Hyun, J. Lim, H. Lee, B. Han, T. Hyeon, A. P. Alivisatos, P. Ercius, H. Elmlund, and J. Park, *Science* **2020**, *368*, 60-67.

- 3. C. F. Reboul, M. Eager, D. Elmlund, and H. Elmlund, Protein Sci. 2018, 27, 51-61.
- 4. C. F. Reboul, J. Heo, C. Machello, S. Kiesewetter, B. H. Kim, S. Kim, D. Elmlund, P. Ercius,
- J. Park, and H. Elmlund, Sci. Adv. 2021, 7 (5), eabe6679.
- 5. G. Kresse and J. Hafner, Phys. Rev. B 1994, 49 (20), 14251.
- 6. F. Shimojo, K. Hoshino and Y. Zempo, Comput. Phys. Commun. 2001, 142 (1-3), 364-367.
- 7. G. Kresse and J. Furthmuller, Comput. Mater. Sci. 1996, 6 (1), 15-50.
- 8. G. Kresse and J. Furthmuller, Phys. Rev. B 1996, 54 (16), 11169.
- 9. J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 1996, 77 (18), 3865.
- 10. P. E. Blöchl, Phys. Rev. B 1994, 50 (24), 17953.
- 11. G. Kresse and D. Joubert, Phys. Rev. B 1999, 59 (3), 1758.
- 12. S. Grimme, J. Comput. Chem. 2006, 27 (15), 1787-1799.
- 13. J. Behler, J. Chem. Phys. 2011, 134 (7), 074106.
- 14. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. De Vito, M. Raison, A. Tejani, S.

Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *Advances in neural information* processing systems **2019**, *32*.

15. I. Loshchilov, F. Hutter, arXiv preprint arXiv:1711.05101 2017.

16. J. Wang, S. Wen, W. Symmans, L. Pusztai, and K. R. Coombes, *Cancer Inform.* 2009, 7, CIN. S2846.

Supplementary Figures and Table



Figure S1. Distribution of each datapoint generated by particle 1 which is used for the training of ANN (red empty dot) and ideal truncated octahedron with the range of 2 to 6 nm truncated from octahedron to cuboctahedron (blue dot). G⁵ symmetry functions with $R_s = 2.52$, $\theta_s = 1.55$ are extracted from the adsorption center (x axis) and 36 characteristic points with 10-degree intervals for each adsorption center (y axis).



Figure S2. Surface vector criteria for 10 Å Pt cluster model. The center Pt atom is defined as a surface atom when all angles between the center Pt to neighbor Pt vector and surface vector are larger than 30 degrees.



Figure S3. Energy convergence benchmark calculation for the binding energy of PVP on Pt (100) surface (a) and amorphous Pt surface (b). All Pt atoms in the cluster are fixed during geometry optimization. The energy difference between bulk and cluster was 0.025 eV for (100) surface and 0.035 eV for the amorphous surface.

Distance (Å)	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
RMSE (eV)	0.1242	0.1239	0.1223	0.1211	0.1220	0.1212	0.1336	0.1426	0.1504	0.1521
Distance (Å)	5.5	6.0	6.5	7.0	7.5	8.0	8.5	9.0	9.5	10.0
RMSE (eV)	0.1578	0.1592	0.1570	0.1519	0.1612	0.1555	0.1548	0.1561	0.1594	0.1512

Table S1. The average of test errors in 5-fold cross-validation for each distance of the characteristic point of neural network (the minimum RMSE is 0.1211 eV where the characteristic point is located at 2.0 Å)



Figure S4. Training and validation error profiles for the neural network. The training was earlystopped at 716 epochs when the validation error was 0.2250 eV and the training error was 0.2169 eV



Figure S5. 3D atomic structures of six synthesized Pt nanoparticles colored by averaged PVP adsorption energy. Each Pt nanoparticle has a diameter of 2.25, 2.41, 2.42, 2.52, 2.66, and 2.92 nm, respectively. Scale bar, 1 nm.



Figure S6. Relative frequency histograms for the averaged E_{bind} in surface atoms of Pt nanoparticles. Each energy distribution is close to bimodal shape.



Figure S7. Relative frequency histograms for the averaged E_{vdW} in surface atoms of Pt nanoparticles. Each energy distribution is close to unimodal shape.



Figure S8. Bimodality indices (BI) of each energy distribution of Pt nanoparticles. All E_{bind} are higher than E_{vdW} , which means the energy distributions of E_{bind} are close to bimodal shape, while those of the E_{vdW} are close to unimodal distribution.



Figure S9. Adsorption energy profiles of E_{bind} and E_{vdW} along the different angular configuration of PVP. The outermost surface layers and the lower layers in each facet are colored with yellow and blue, respectively. Among the colored surface atoms, the orange atoms indicate the adsorption centers.



Figure S10. The most stable configuration of PVP ligands binding on different terrace atoms of islands. The characteristic vectors are consistently toward to near terraces.



Figure S11. The most stable configuration of PVP ligands binding on different (a) edge and (b) corner atoms. The characteristic vectors are consistently toward the steps nearby.



Figure S12. The most stable configuration of PVP ligands binding under the islands. The characteristic vectors are consistently toward the steps nearby.

Adsorption site	Configurational angle (degree)	DFT calculated E _{ads} (eV)	ANN predicted E _{ads} (eV)
Row 1 blue	115(103)	-1.560	-1.659
Row 1 orange	115(110)	-1.900	-2.060
Row 2 blue	130(130)	-1.522	-1.673
Row 2 orange	130(144)	-1.992	-1.987
Row 3 blue	20(37)	-1.912	-1.912
Row 3 orange	20(21)	-2.346	-2.236

Table S2. The DFT calculation results from the exemplary adsorption sites in Figure 5. The configurational angles in column 2 are used for the DFT relaxation where ANN predicted the largest adsorption energy difference and angles in the parenthesis are the adsorption angles after the DFT relaxation. ANN predicted adsorption energies at DFT relaxed angles are listed in column 4.



Figure S13. Relative frequency of the cosine similarity of adsorption direction of each PVP ligand with most stable E_{vdW} and E_{bind} compared with adsorption direction with the most stable E_{ads} . The adsorption directions of each PVP ligand with the most stable E_{ads} , E_{vdW} and E_{bind} are defined as θ_{ads} , θ_{vdW} , and θ_{bind} , respectively. The cosine similarity of θ_{ads} to θ_{vdW} and to θ_{bind} is displayed as orange and green bar, respectively. The mean value of the former is 0.760 and that of the latter is 0.399.