Supplementary information for

"Neural Network Assisted Analysis of Bimetallic Nanoalloys using X-ray Absorption Near Edge Structure Spectroscopy"

Nicholas Marcella,[*a] Yang Liu, [b] Janis Timoshenko,[a] Erjia Guan,[a] Mathilde Luneau,[c] Tanya Shirman,[d,e] Anna M. Plonka,[a] Jessi E.S. van der Hoeven,[c,d] Joanna Aizenberg,[c,d,e] Cynthia M. Friend,[c,d] and Anatoly I. Frenkel[* a,f]

a. Department of Materials Science and Chemical Engineering, Stony Brook University, Stony Brook, New York 11794, United States.

b. Department of Chemistry, Stony Brook University, Stony Brook, New York 11794, United States.

c. Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States.

d. John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, United States.

e. Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts 02138, United States.

f. Chemistry Division, Brookhaven National Laboratory, Upton, New York 11973, United States.

Corresponding Author

*Emails: nicholas.marcella@stonybrook.edu, anatoly.frenkel@stonybrook.edu

Table S1. Experimental data used to test FEFF9 ability to simulate PdAu NPs.

Pd Concentration (at. %)	Average particle size (nm)	Support/Surfactant
83%	3.0 ± 0.6	Peptide R5
67%	3.3 ± 0.7	Peptide R5
50%	3.4 ± 0.7	Peptide R5
33%	3.8 ± 0.7	Peptide R5
25%	4.0 ± 0.7	Peptide R5
9%	6.0 ± 2.1	RCT-SiO2
2%	5.8 ± 1.4	RCT-SiO2

Note S1. Details of ab initio calculation of XANES

As in our previous work, [1-4] we used FEFF9[5] for XANES simulations. The non-structural parameters for XANES simulations were chosen to ensure the best agreement between the simulated spectrum for bulk Au and Pd and the corresponding experimental XANES data. FEFF version 9.6.4 was used for self-consistent calculations within full multiple scattering (FMS) and muffin-tin (MT) approximations. For Pd K-edge simulations, we use the complex Hedin-Lundqvist (HL) exchange correlation potential, cluster with radius 5.5 Å for self-consistent field (SCF) calculations, cluster of radius of 7 Å for FMS calculations, and S02 of 0.9, where the core-hole is treated with the random phase approximation (RPA). For Au L_3-edge simulations, we use the partially nonlocal exchange model: Dirac-Fock model for core + HL model for valence electrons + a constant imaginary part exchange correlation potential (which is corrected by a 0.2 eV shift of the Fermi level and a 0.5 shift to the optical potential), 3.1 Å cluster for SCF calculations, 7 Å cluster for FMS calculations, and S02 of 1.0, where the core-hole was treated with the final state rule (FSR).

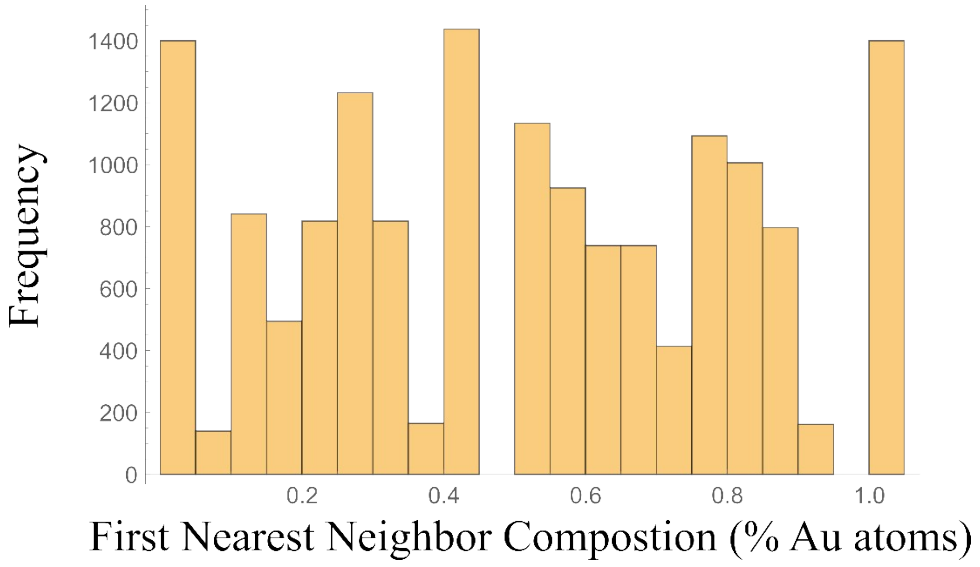

Figure S1. Au edge training data: the absorbing site local composition (defined as the number of Au neighbors divided by the total number of metal neighbors).

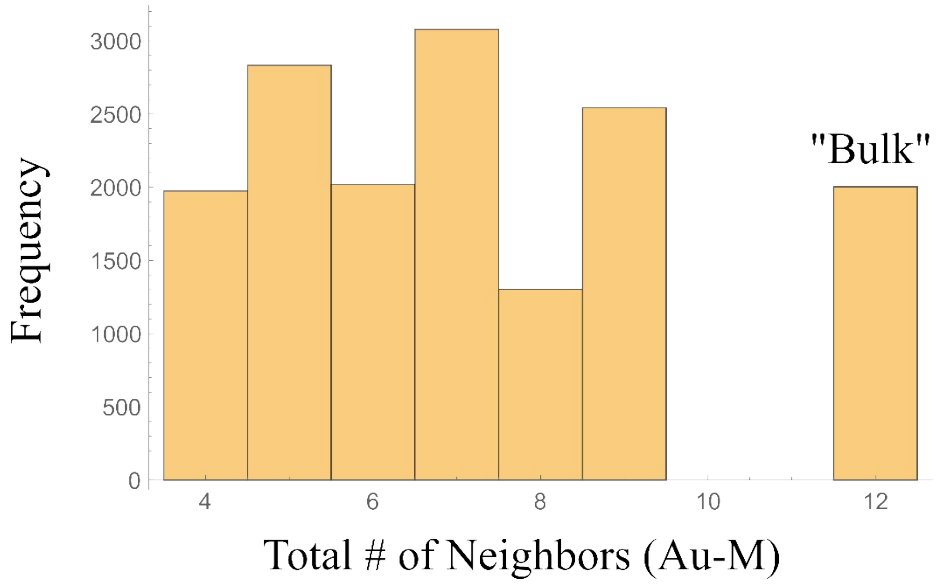

Figure S2. Au edge training data for PdAu nanoparticles: the location of the absorbing atoms is described by the total number of metal neighbors (Au-M). Label "Bulk" corresponds to absorbing atoms inside the NP (coordination number (CN) for Au-M: CN(Au-M) = 12), while the CN(Au-M) < 12 corresponds to undercoordinated sites at the surfaces, edges, and corners.

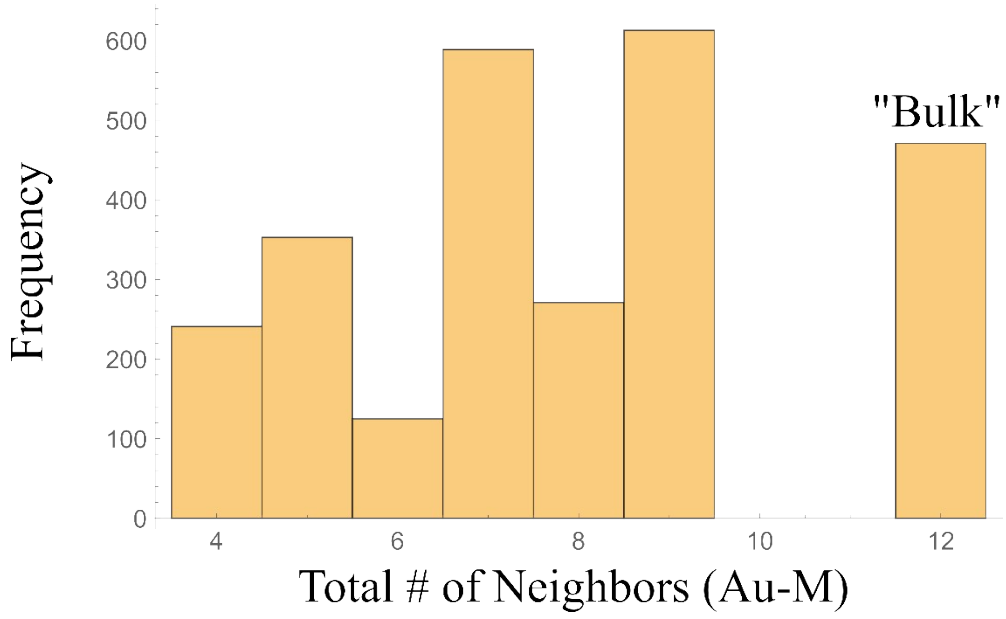

Figure S3. Au edge training data for Au nanoparticles: the location of the absorbing atoms are described by the total number of metal neighbors (Au-M). Label "Bulk" corresponds to absorbing atoms inside the NP (coordination number (CN) for Au-M: CN(Au-M) = 12), while the CN(Au-M) < 12 corresponds to undercoordinated sites at the surfaces, edges, and corners.

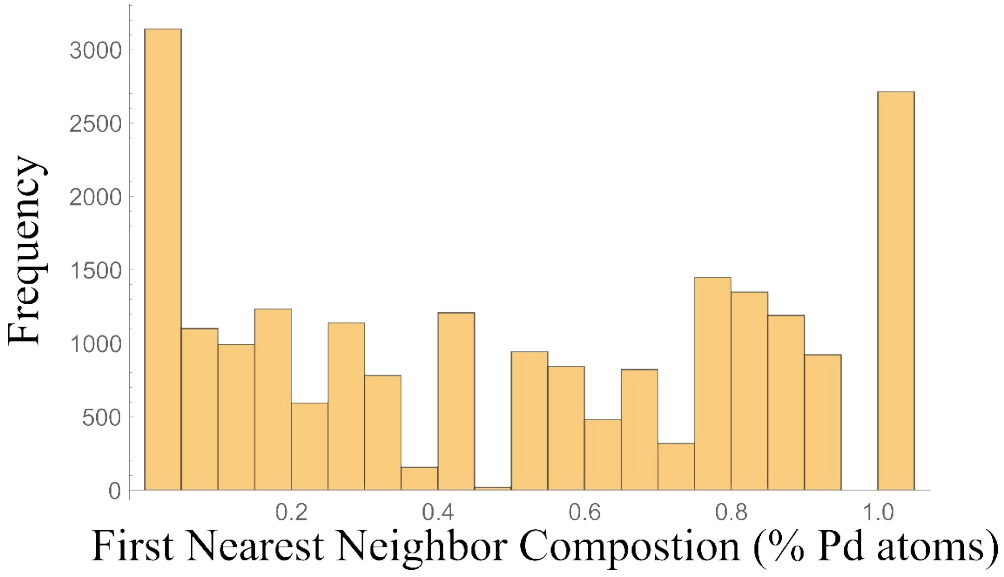

Figure S4. Pd edge training data: the absorbing site local composition (defined as the number of Pd neighbors divided by the total number of metal neighbors).

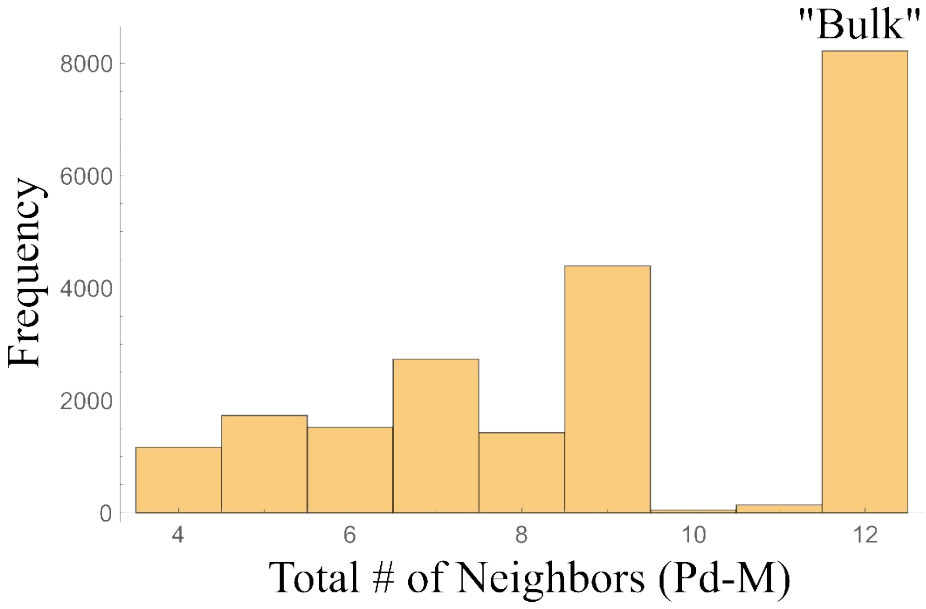

Figure S5. Pd edge training data for PdAu nanoparticles: the location of the absorbing atoms are described by the total number of metal neighbors (Pd-M). Label "Bulk" corresponds to absorbing atoms inside the NP (coordination number (CN) for Pd-M: CN(Pd-M) = 12), while the CN(Pd-M) < 12 corresponds to undercoordinated sites at the surfaces, edges, and corners.

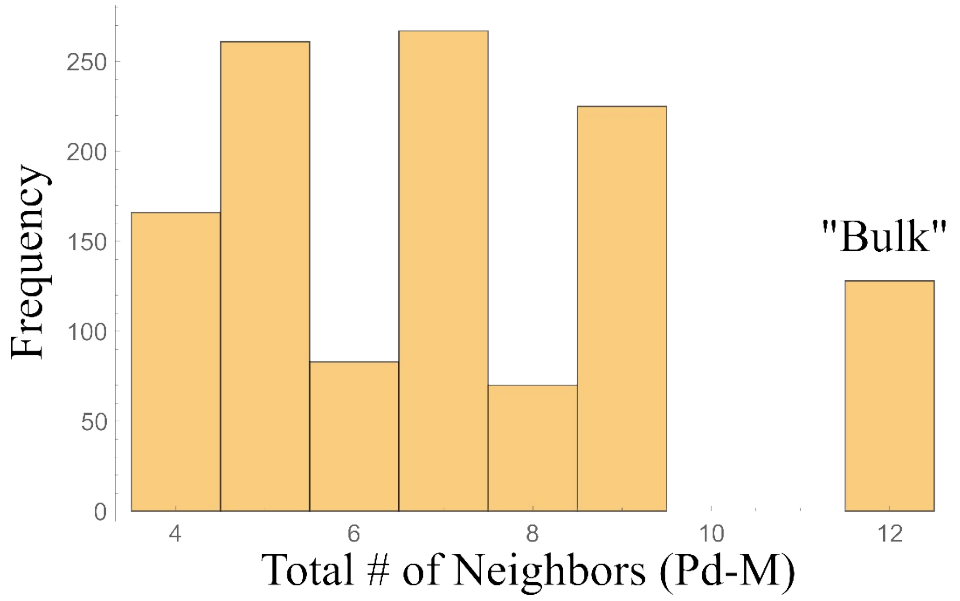

Figure S6. Pd edge training data for Pd nanoparticles: the location of the absorbing atoms are described by the total number of metal neighbors (Pd-M). Label "Bulk" corresponds to absorbing atoms inside the NP (coordination number (CN) for Pd-M: CN(Pd-M) = 12), while the CN(Pd-M) < 12 corresponds to undercoordinated sites at the surfaces, edges, and corners.

Note S2. Neural network training and validation

The Pd and Au absorber NNs were implemented and trained using Mathematica 12 using the NetGraph and NetTrain functions. The complete architectures are presented in Figs. S9 and S10. For the Pd absorber-specific NN, the site-specific XANES calculations were aligned to a bulk Pd reference standard, and then interpolated to a 95 point non-uniform energy mesh that spanned energies from E_{min} = 24339.8 eV to E_{max} = 24416.6 eV with a step size of 0.6 eV for data points near the adsorption edge, which gradually increased up to 1.7 eV for points approaching E_{max}. Furthermore, a simulated bulk Pd spectra, also calculated by FEFF9, aligned to the bulk reference standard, and interpolated to the same mesh, was subtracted from each interpolated site-specific XANES spectrum. Finally, the energies and absorption coefficients were normalized between 0 and 1 using the commonly used min-max normalization procedure: $Z = (X - \min(X))/(\max(X) - \min(X))$. X is the training data input, min(X) is the smallest input, max(X) is the largest input, and Z is the normalized input. The target length 2 vector was composed of the first partial coordination numbers C_{Pd-Pd} and C_{Pd-Au} as calculated from the atomistic models. Before training, the partial coordination numbers were normalized between 0 and 1 using the min-max normalization. For

the Au absorber-specific NN, the site-specific XANES calculations were aligned to a bulk Au reference standard and then interpolated to an 83 point non-uniform energy mesh before use in training. The energy mesh spanned energies from E_{min} = 11917.8 eV to E_{max} = 11994.6 eV with a step size of 0.6 eV for data points near the adsorption edge, which gradually increased up to 1.7 eV for points approaching E_{max}. For the Au edge, bulk subtraction did not help minimize the validation loss, so it was not utilized. This was likely due to the Au L_3-edge being less sensitive to the energy resolution effects as compared to the higher energy Pd K-edge. The target length 2 vector was composed of the first coordination numbers C_{Au-Au} and C_{Au-Pd} as calculated from the atomistic models. Before being combined linearly to form synthetic data, the absorption coefficients, energy values, and coordination numbers were normalized between 0 and 1, again using min-max normalization. For training, we initialized the weights randomly, using the Xavier method.[6] For the Pd absorber-specific NN, the validation cost function was minimized after training on 405,248 synthetic examples with a batch size of 64 per training round. Training was stopped at the lowest validation loss. The validation cost function was minimized using the ADAM optimizer (with Epsilon = 0.001, Beta1= 0.9, Beta2 = 0.99 and initial learning rate of 0.001) and a L2 Regularization of 0.001 was employed. For the Au absorber specific NN, training was stopped at the lowest validation loss, which occurred after training on 1,617,920 synthetic examples with a batch size of 64 per training round. Minimization was completed using the ADAM optimizer (with Epsilon = 0.001, Beta1= 0.9, Beta2 = 0.99 and initial learning rate of 0.001) and a L2 Regularization of 0.001. The additional regularization methods (Dropout with a probability of 0.2 and Batch Normalization) helped to minimize the validation cost in the case of the Au absorber-specific NN, and so were included in the final architecture.

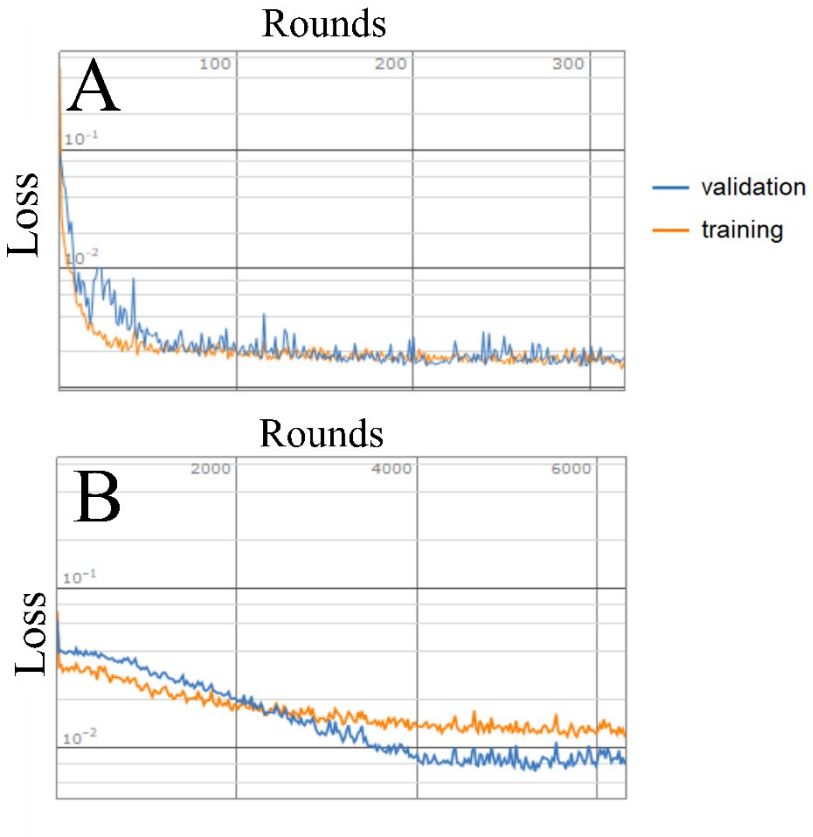

Figure S7. The MSE loss vs. training round for A) The Au absorber specific NN and B) Pd absorber specific NN. Training is stopped at the validation minimum.

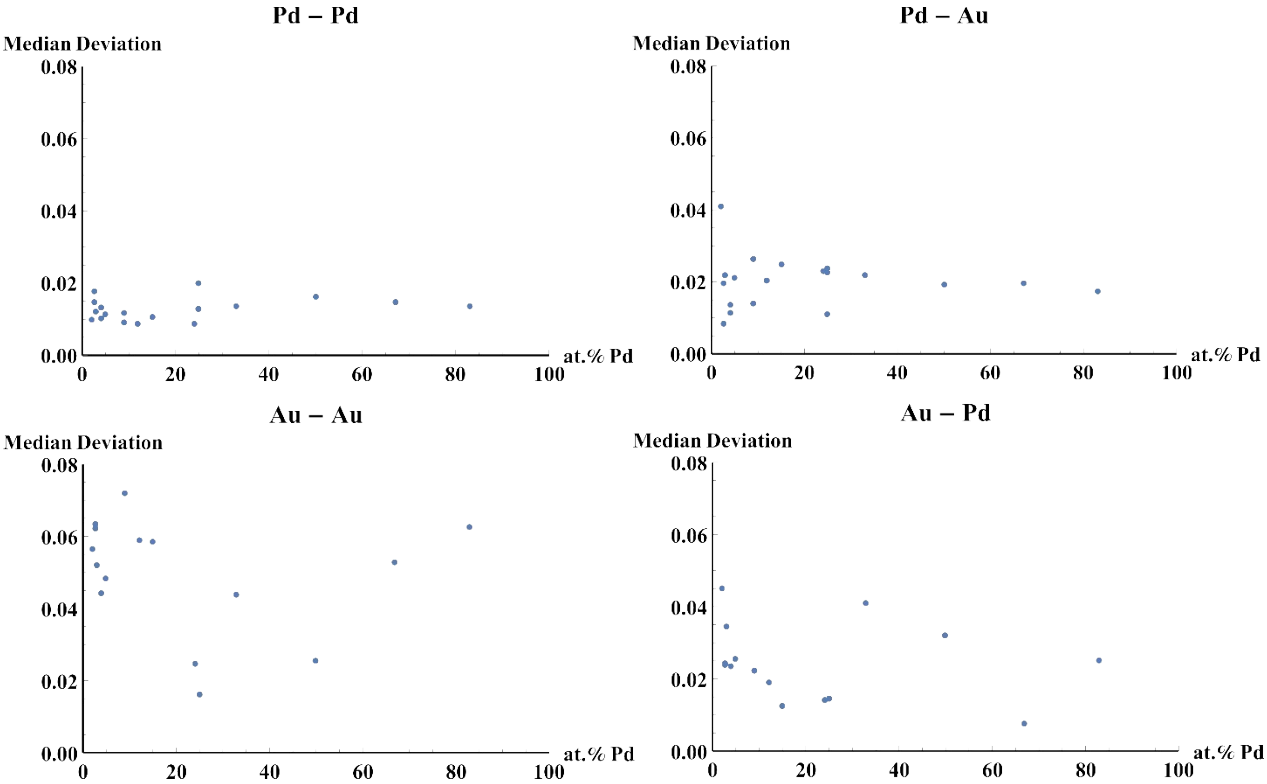

Figure S8. The median absolute deviation in NN-XANES predictions of the coordination numbers Pd-Pd, Pd-Au, Au-Au, and Au-Pd. The median was calculated for predictions on labelled experimental data by 10 separate Pd and Au absorber-specific NNs.

Table S2. Pd K-edge Neural Network implemented in Mathematica 12 [7]

Layer	Neurons	Activation	Kernel	Stride
Convolution	32	ReLU	1	1
Max Pooling	N/A	N/A	2	N/A
Convolution	32	ReLU	1	N/A
Fully Connected	516	ReLU	N/A	N/A
Fully Connected	2	N/A	N/A	N/A

Table S3. Au L_3-edge Neural Network implemented in Mathematica 12 [7]

Layer	Neurons	Activation	Kernel	Stride
Convolution	40	ReLU	2	2
Batch Norm	N/A	N/A	N/A	N/A
Fully Connected	300	ReLU	N/A	N/A
Fully Connected	300	ReLU	N/A	N/A
Dropout	N/A	N/A	N/A	N/A
Fully Connected	300	ReLU	N/A	N/A
Fully Connected	2	N/A	N/A	N/A

Figure S9. Au absorber specific NN architecture as implemented in Mathematica 12. The layers are connected sequentially from 1 to 11.

Figure S10. Pd absorber specific NN architecture as implemented in Mathematica 12. The layers are connected sequentially from 1 to 8.

Note S3. Details of error estimation

Here we describe two major sources of uncertainty associated with the NN-XANES method, 1) the systematic differences that exist between the theoretical training data and the experimental data, and 2) the stochastic training process. The uncertainties from source 1 are best estimated with the help of our labeled experimental data. We estimate the uncertainties associated with the absorber-specific NN predictions by assuming that there exists mean absolute error associated with the predictors. Therefore, based on the absolute errors calculated for our validation and test predictions, we can estimate the minimum and maximum values of the true mean absolute error within a 95% confidence interval, where the maximum of the confidence interval is a conservative estimate of the maximum error bar for the pair-specific predictions. This is accomplished by calculating the mean errors of the validation and test predictions (11 examples for the Pd absorber-specific NN and 13 examples for the Au absorber-specific NN). Equation 2 is then used to calculate the max of the confidence interval (the maximum error bar for our purposes), where $\bar{x}$ is the mean absolute error, σ is the standard deviation of the mean absolute error, n is the number of examples, and 1.96 is the Z value for a 95% confidence interval.

$$\max error = \bar{x} + \frac{1.96\sigma}{\sqrt{n}} . \quad (2)$$

After obtaining the mean and standard deviations of the absolute errors, we calculated the maximum value of the 95% confidence interval and take this to be a very conservative estimate of the error bars associated with the NN predictions. For the partial coordination numbers $C_{Pd\text{-}Pd}$, $C_{Pd\text{-}Au}$, $C_{Au\text{-}Au}$, and $C_{Au\text{-}Pd}$, we estimate the error bars as ± 1.0, ± 1.3, ± 0.8, and ± 0.9 respectively. The uncertainties from source two are determined by comparing the predictions made on the validation and testing data for 10 independently trained NNs, for both Pd and Au. The median absolute deviations of the predictions are shown in Fig. S8 as a function of Pd concentration. For both Pd and Au absorber specific NNs, the errors due to stochastic variation are relatively stable, and are an order of magnitude lower than those due to systematic theoretical-experimental differences, and so we report the latter error bars in Fig. 4 and Table S4. However, the error due to source two is important when one is interested in relative changes in, rather than absolute predictions of, coordination numbers.

Note S4. Details of RCT-2 dataset

The synthesis methods for all RCT-SiO_2 supported PdAu catalysts were reported previously (Pd_2Au_{98},[8, 9] Pd_4Au_{96},[8] Pd_9Au_{91},[8, 9] and $Pd_{25}Au_{75}$.[10]) XAFS measurements were performed at beamline 8-ID (ISS) of the National Synchrotron Light Source II (NSLS-II) using a focused 0.5 mm by 0.5 mm beam, Si 111 monochromator, and Pt collimating mirror for higher harmonic rejection. The RCT-SiO_2 supported Pd_4Au_{96}, Pd_9Au_{91}, and $Pd_{25}Au_{75}$ powders were packed into 2.0 mm ID 2.4 mm OD quartz capillaries for insertion into the flow cell. The XAFS measurements were taken at the Pd K edge (24350 KeV) in fluorescence mode using a PIPS detector. The cell was positioned 45 degrees to the beam direction to reduce elastic scattering, and the PIPS was positioned perpendicular to the sample. A Pd foil reference standard was positioned between two ionization chambers directly after the flow cell. All XAFS was taken under He, and $Pd_{25}Au_{75}$ was measured under He and H_2. The data was aligned and edge-step normalized using Athena,[11] after which is was preprocessed for use with NN-EXAFS according to the procedure reported in Ref. [9]

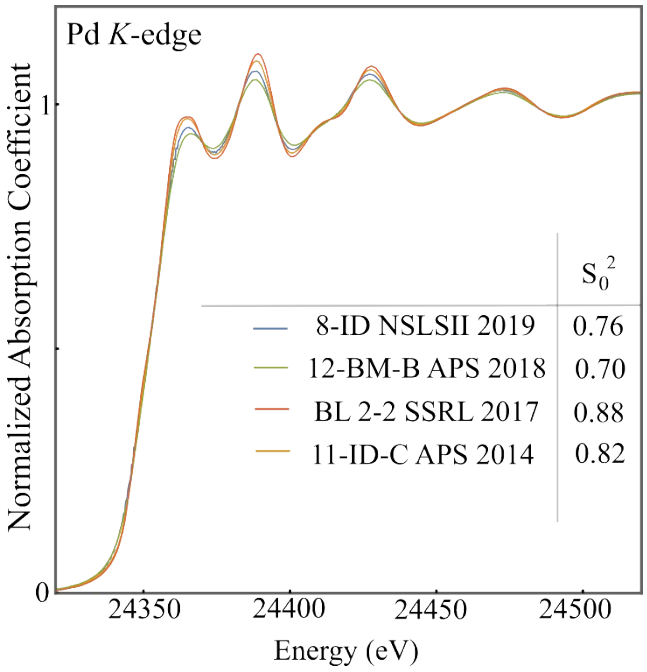

Figure S11. Pd K-edge for Pd foil references collected at beamlines 8-ID (RCT-2 data set collected at NSLSII in 2019), 12-BM-B (TiO_2 supported data set collected at APS in 2018), BL 2-2 (RCT-1 data set collected at SSRL in 2017), and 11-ID-C (peptide stabilized data set collected at APS in 2014).

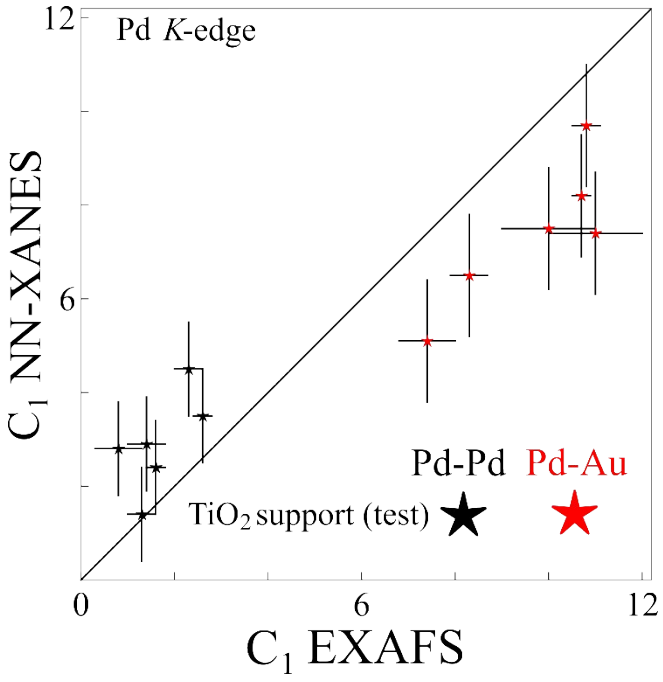

Figure S12. The XANES-derived vs. EXAFS-derived first partial coordination numbers with respect to the Pd K-edge. Pd-Pd and Pd-Au coordination numbers predicted for the TiO_2 supported data set are in black and red, respectively.

Table S4. EXAFS/NN-EXAFS and NN-XANES-derived CNs for experimental datasets

		EXAFS (or NN-EXAFS)				NN-XANES			
Experiment	Samples	$C_{Pd\text{-}Pd}$	$C_{Pd\text{-}Au}$	$C_{Au\text{-}Au}$	$C_{Au\text{-}Pd}$	$C_{Pd\text{-}Pd}$ a	$C_{Pd\text{-}Au}$ b	$C_{Au\text{-}Au}$ c	$C_{Au\text{-}Pd}$ d
NSLSII RCT support (RCT-2)	Pd_9Au_{91} (He)	1.2(1) Δ	10.6(3) Δ	-	-	1.0	10.3	-	-
	Pd_4Au_{96} (He)	1.2(2) Δ	10.3(5) Δ	-	-	1.1	10.1	-	-
	$Pd_{25}Au_{75}$ (He)	0.6(5) Δ	9.7(12) Δ	-	-	2.1	8.7	-	-
	$Pd_{25}Au_{75}$ (H_2)	1.1(6) Δ	9.8(12) Δ	-	-	2.2	8.5	-	-
SSRL RCT support (RCT-1)	Pd_2Au_{98} (He)	0.9(4) Δ	10.4(6) Δ	11.2(2) Δ	0.5(2) Δ	0.0	12.6	10.3	0.1
	Pd_9Au_{91} (He)	1.1(6) Δ	10.3(9) Δ	10.6(3) Δ	0.8(2) Δ	0.8	10.7	10.4	0.1
APS Peptide surfactant (Peptide)	83%PdAu	6.9(3) Δ	2.4(5) Δ	3.6(5) Δ	7.1(5) Δ	5.4	3.6	3.1	7.3
	67%PdAu	4.3(2) Δ	3.8(4) Δ	5.2(3) Δ	6.1(4) Δ	4.7	4.7	5.2	5.5
	50%PdAu	3.9(2) Δ	6.5(6) Δ	6.7(2) Δ	4.0(3) Δ	3.8	6.0	6.6	4.2
	33%PdAu	2.8(4) Δ	6.5(6) Δ	8.1(3) Δ	1.9(3) Δ	3.5	6.5	8.5	2.2
	25%PdAu	1.7(5) Δ	9.5(4) Δ	8.4(3) Δ	1.6(2) Δ	2.6	7.9	9.1	1.4
APS TiO_2 support (TiO_2)	24%PdAu	2.6(2) †	8.3(4) †	7.8(5) †	3.1(2) †	3.5	6.5	9.3	1.3
	15%PdAu	1.0 (4) †	11(1) †	9.1(9) †	1.9(4) †	2.9	7.4	9.8	0.6
	12%PdAu	2.1(3) †	7.4(6) †	10.2(8) †	1.4(4) †	4.5	5.1	10.1	0.4
	5%PdAu	0.8(5) †	10(1) †	10(1) †	0.7(4) †	2.8	7.5	10.2	0.2
	4%PdAu	1.3(3) Δ	10.8(3) Δ	10.3(8) †	0.5(4) †	1.4	9.7	10.1	0.2
	3%PdAu	1.6(2) Δ	10.7(2) Δ	12(1) †	0.0(4) †	2.4	8.2	10.3	0.1

†Conventional EXAFS ΔNN-EXAFS Max absolute error in predictions: [a] ± 1.0, [b] ± 1.3, [c] ± 0.8, [d] ± 0.9

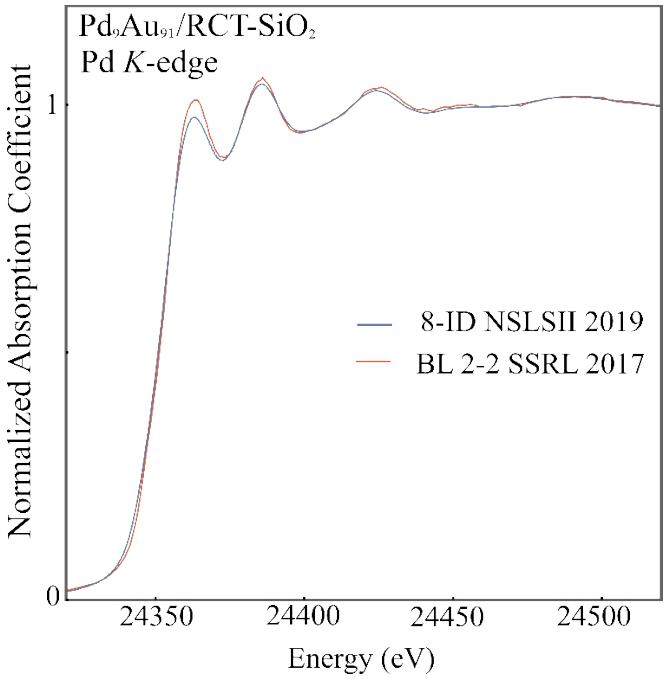

Figure S13. Pd K-edge for Pd_9Au_{91} on RCT-SiO_2 that was measured at beamlines 8-ID (RCT-2 data set collected at NSLSII in 2019), and BL 2-2 (RCT-1 data set collected at SSRL in 2017). The qualitative differences in the XANES are due to the different monochromators used (Si 111 at 8-ID and Si 220 at BL 2-2). Despite the qualitative differences, NN-XANES predicts the same CNs for both spectra, which is expected as both spectra were measured for the same sample.

References

1 Y. Liu, N. Marcella, J. Timoshenko, A. Halder, B. Yang, L. Kolipaka, M. J. Pellin, S. Seifert, S. Vajda, P. Liu and A. I. Frenkel, *J. Chem. Phys.*, 2019, **151**, 164201.

2 J. Timoshenko, D. Lu, Y. Lin and A. I. Frenkel, *J. Phys. Chem. Lett.*, 2017, **8**, 5091-5098.

3 J. Timoshenko, S. Roese, H. Hövel and A. I. Frenkel, *Radiat. Phys. Chem.*, 2018.

4 J. Timoshenko and A. I. Frenkel, *ACS Catal.*, 2019, **9**, 10192-10211.

5 J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange and K. Jorissen, *Phys Chem Chem Phys*, 2010, **12**, 5503-5513.

6 X. Glorot and Y. Bengio, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Chia Laguna Resort, Sardinia, Italy, 2010.

7 Mathematica, version 12.0, Wolfram Research, Inc.: Champaign, Illinois,2019.

8 M. Luneau, T. Shirman, A. C. Foucher, K. Duanmu, D. M. A. Verbart, P. Sautet, E. A. Stach, J. Aizenberg, R. J. Madix and C. M. Friend, *ACS Catal.*, 2020, **10**, 441-450.

9 J. Timoshenko, C. J. Wrasman, M. Luneau, T. Shirman, M. Cargnello, S. R. Bare, J. Aizenberg, C. M. Friend and A. I. Frenkel, *Nano Lett.*, 2019, **19**, 520-529.

10 E. Guan, A. C. Foucher, N. Marcella, T. Shirman, M. Luneau, A. R. Head, D. M. A. Verbart, J. Aizenberg, C. M. Friend, D. Stacchiola, E. A. Stach and A. I. Frenkel, *ChemCatChem*, 2020, **12**, 717-721.

11 B. Ravel and M. Newville, *J. Synchrotron Radiat.*, 2005, **12**, 537-541.