Electronic Supplementary Material (ESI) for Nanoscale.

Supporting Information for **Detection of Molecular Behavior that Characterizes Systems using a Deep Learning Approach**

Katsuhiro Endo,[1] Daisuke Yuhara,[1*] Katsufumi Tomobe,[1] Kenji Yasuoka[1*]

[1]Department of Mechanical Engineering, Keio University,
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, Japan 223-8522

* E-mail: daisuke@keio.jp (D.Y.); yasuoka@mech.keio.ac.jp (K.Y.)

Section S1

Structure of deep neural networks

For the calculation of Wasserstein distances between systems, deep neural networks were used. In the all experiments, the same structure of deep neural networks was applied. In total, the deep neural networks have 6 layers, which are as follows,
Input (64*3) - Affine (1024, LReLU) - Affine (1024, LReLU) - Affine (1024, LReLU) - Affine (1024, Non-bias) - Output (1).
Then, we defined "Output = f (Input)" and f was optimized by the Adam optimizer (learning rate = 1e-4, beta1 = 0.5, beta2 = 0.9). The dimension and activation function in each layer are written in (). For the initialization, the bias terms for Affine layers were set to 0 and the other parameters were randomly sampled from a normal distribution (mean = 0, standard deviation = 0.01). All activation functions were LReLU (a = 0.1).

Details of optimization process

To calculate the Wasserstein distances between each system, the following loss function[1] was optimized using deep neural networks.

$$L = \sum_{\boldsymbol{x}\sim\boldsymbol{P}} f(\boldsymbol{x}) - \sum_{\boldsymbol{x}\sim\boldsymbol{Q}} f(\boldsymbol{x}) - \sum_{\boldsymbol{x}\sim\boldsymbol{R}} (||\nabla f(\boldsymbol{x})|| - 1)^2 \tag{1}$$

When $\boldsymbol{x} \sim \boldsymbol{R} : \varepsilon = U[0,1]$, $\boldsymbol{x}_1 \sim \boldsymbol{P}$, $\boldsymbol{x}_2 \sim \boldsymbol{Q}$, $\boldsymbol{x} = \varepsilon \boldsymbol{x}_1 + (1-\varepsilon)\boldsymbol{x}_2$. When the loss function is minimized, the $\sum_{\boldsymbol{x}\sim\boldsymbol{P}} f(\boldsymbol{x}) - \sum_{\boldsymbol{x}\sim\boldsymbol{Q}} f(\boldsymbol{x})$ indicates the Wasserstein distance W[$\boldsymbol{P}, \boldsymbol{Q}$].

The data format of the input for deep neural networks is the 64 steps of molecular trajectories. For one molecule, the molecular trajectory is basically defined as the 64 steps of the 3-dimensional position data (x, y, z). According to the purpose, we can also use the 64 steps of certain vectors such as dipole moment. In the above loss function, $\boldsymbol{P}$ and $\boldsymbol{Q}$ denote the sets of molecular trajectories. In this study, we consider that the molecular trajectories are obtained by sampling from a certain probability distribution.

Thus, $\boldsymbol{P}$ (or $\boldsymbol{Q}$) is considered as a set of all molecular trajectories (= samples) obtained from one molecular system. In practice, if there are three different systems ($\boldsymbol{P}$, $\boldsymbol{Q}$, $\boldsymbol{R}$), three individual optimization processes (between $\boldsymbol{P}-\boldsymbol{Q}$, $\boldsymbol{Q}-\boldsymbol{R}$, $\boldsymbol{R}-\boldsymbol{P}$) are performed. When the loss function for all combinations is optimized, the calculation of Wasserstein distances is finished.

Error estimation for the calculation of Wasserstein distance

Let us consider the error estimation for the calculation of Wasserstein distance. When the number of data is finite, the truth distributions $\boldsymbol{P}'$ and $\boldsymbol{Q}'$ given by infinite data can be approximated by the finite sum of δ function, $\boldsymbol{P}$ and $\boldsymbol{Q}$. According to the paper on optimization using the Wasserstein metric[2], the following equation is held under several assumptions,

$$\mathrm{P}[W(\boldsymbol{P},\boldsymbol{P'}) \geq \varepsilon] \sim \exp(-cN\varepsilon^a) \tag{2}$$

where c is a constant, a is a constant determined by ε, and N is the number of data. This equation means that the error by the finitude of data decreases exponentially with increasing N. Thus, if the N is sufficiently large, the error is negligible. In this study, N was set to a sufficiently large number (greater than 10^6). This theoretical estimation of error is one of the advantages of the using the Wasserstein distance.

Section S2

Preliminary experiment with the sequence of noise

To confirm the availability of our scheme, we conducted a preliminary experiment using the simple artificial data that had distinct differences. We assumed that there were nine systems, and sequences of non-correlated Gaussian noise were considered as the dynamics observable in each system. Thus, local dynamics ensembles $\boldsymbol{y}_i$ are defined as

$$\boldsymbol{y}_i = \mathcal{N}(\mu_i, \sigma_i^2) \tag{3}$$

where $\boldsymbol{y}_i$ denotes 64 (time-length) $\times$ 1 (channel) dimensional random variables. The 3 $\times$ 3 values were assigned as each (μ_i, σ_i) for i, $(-0.1, 1)$, $(-0.1, 2)$, $(-0.1, 3)$, $(0.0, 1)$, $(0.0, 2)$, $(0.0, 3)$, $(0.1, 1)$, $(0.1, 2)$, and $(0.1, 3)$. The Wasserstein distances between all the pairs of local dynamics ensembles were calculated (the distance matrix is shown in Fig. S1(a)). In Fig. S1(b), embeddings of the systems clearly demonstrate a two-dimensional structure. Each dimension of the embeddings definitely corresponds to the assigned mean and variance sets. This indicates that our scheme automatically identified the existing two types of differences.

To confirm the dynamics that characterize systems, the sequences of noise in S_0 and S_6 were compared using $g(\boldsymbol{x})$. The probability distribution of $g(\boldsymbol{x})$ for the sequence of noise in S_6 looked almost symmetric (Fig. S1(c)). This indicates that the dominant dynamics in each system (S_0 and S_6) appeared in S_6 with the same probability. From the local dynamics ensemble data of S_6, we exhibited the sequence of noise with high $g(\boldsymbol{x})$ and low $g(\boldsymbol{x})$ (Fig. S1(d)). The sequence of noise with high $g(\boldsymbol{x})$ demonstrated the higher average. This indicates that $g(\boldsymbol{x})$ appropriately represents dynamics that characterize systems because the average of the sequence of noise in S_6 was set to be higher than S_0.

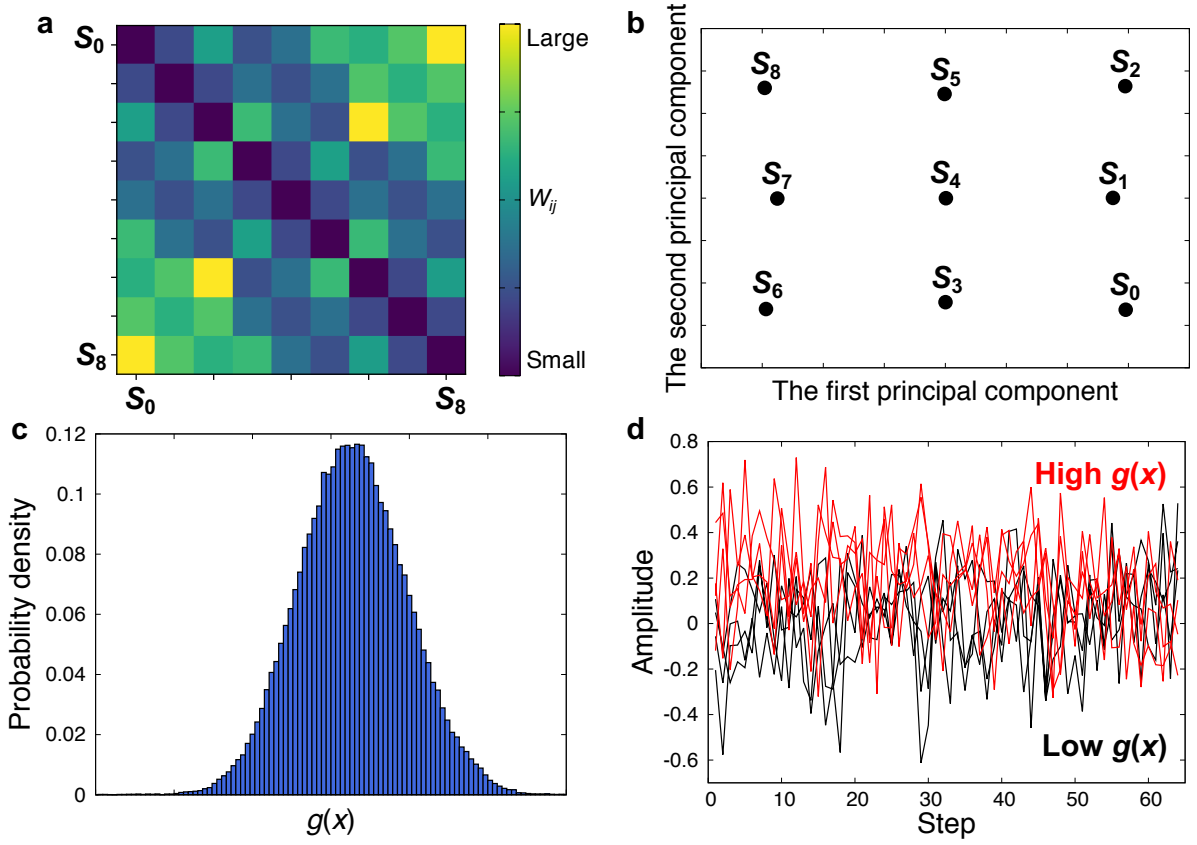

Fig. S 1. Experimental analysis (sequence of noise). (a) The distances matrix. The elements of the matrix are colored by the gradation between yellow (high) and blue (low). (b) Embeddings of systems based on the Wasserstein distances between the local dynamics ensembles. (c) Probability distribution of $g(\boldsymbol{x})$ for the trajectories of S_6. (d) Comparison of the sequence of noise, which indicates high $g(\boldsymbol{x})$ and low $g(\boldsymbol{x})$.

Section S3

Supplementary data for the numerical experiments

Distance matrix

The distance matrix obtained in each experiment using MD data is shown in Fig. S2. In the experiment for the Lennard-Jones particle system, the temperature of each system was raised with the increase of the system's index. This relationship clearly appears in Fig. S2(a). This indicates that the calculated Wasserstein distances accurately represented the temperature differences. In the experiment for the ethanol and water system, the mole fraction of ethanol in each system was reduced with the increase of the system's index. Although this relationship appears in Fig. S2(b), it looks noisier than Supplementary Fig. S2(a). This noise was caused by the lack of samples for the local dynamics ensemble of each system (the details are discussed in the Results section of the main text). In the experiment for the amino acid solution system, the differences in water dynamics had no relation to the system's index. This is also confirmed in Fig. S2(c). To visualize the differences more clearly, the embeddings were obtained using this distance matrix (Fig. 4(a)).

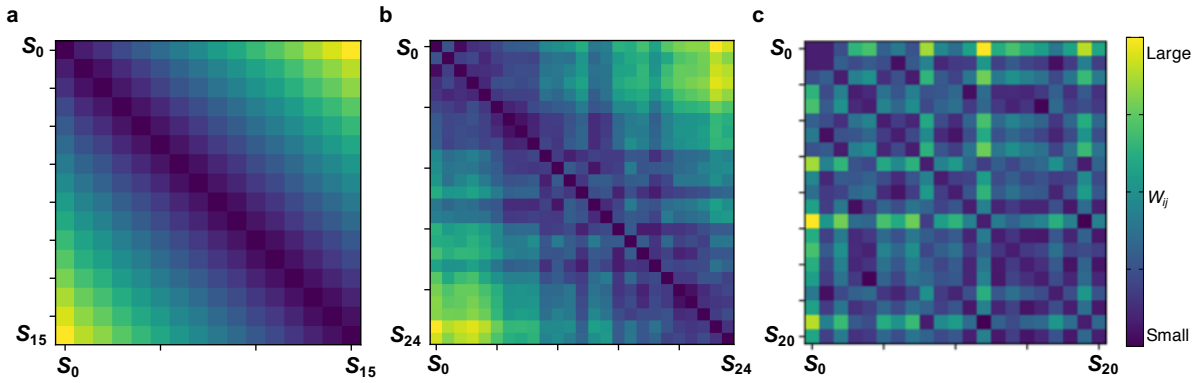

Fig. S 2. Distance matrix obtained in the experiments using MD data. (a) Lennard-Jones particle system. (b) Ethanol and water mixture system. (c) Amino acid solution system.

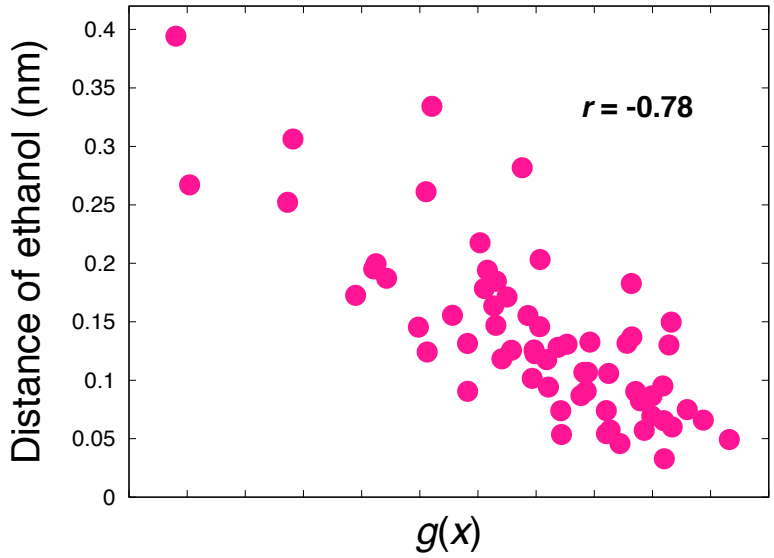

Fig. S 3. Correlation between the distance of ethanol molecules and $g(\boldsymbol{x})$.

Ethanol and water mixture system

In the experiment for the ethanol and water mixture system, the distance from the initial position for several ethanol molecules in S_{23} is shown in Fig. 3(d). This figure implies that the distance seems to be large with the decrease of $g(\boldsymbol{x})$. To quantitatively confirm this tendency, we obtained the correlation between the distance of ethanol molecules after 128 ps elapsed and $g(\boldsymbol{x})$. In Fig. S3, we see a strong correlation between them, in which the correlation coefficient is -0.78.

Amino acid solution system

In the experiment for the amino acid solution system, the differences between the rotation of water molecules were clarified by $g(\boldsymbol{x})$. To quantitatively confirm the relationship between the rotation of water molecules and $g(\boldsymbol{x})$, we obtained the correlation between them. To represent the rotation of water, we calculated an average of inner product of the dipole moments from a certain time to 128 ps later. In Fig. S4, we see the

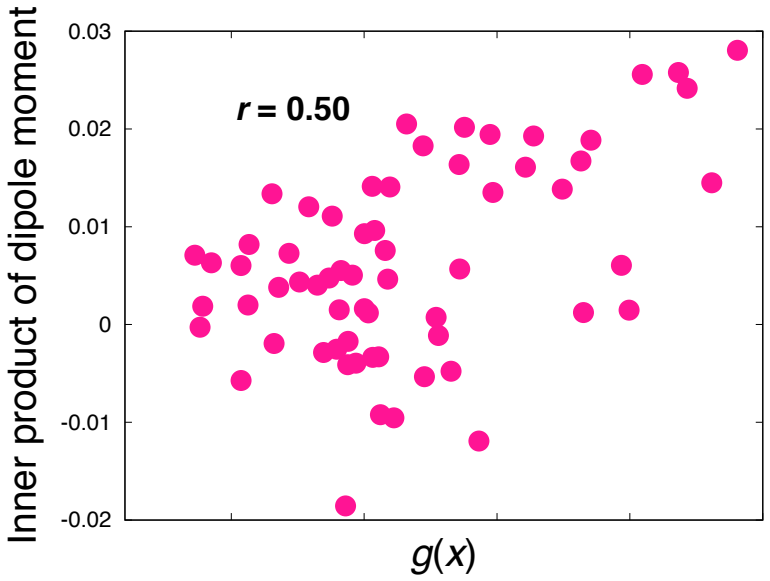

Fig. S 4. Correlation between the inner product of the dipole moments and $g(\boldsymbol{x})$.

correlation between the inner product and $g(\boldsymbol{x})$, in which the correlation coefficient is 0.50. The correlation coefficient is not so high, because most of the behavior of the water molecules was similar (see the histogram of $g(\boldsymbol{x})$ in Fig. 4(c)). To confirm the relationship between the rotation of water molecules and $g(\boldsymbol{x})$, we also calculated the rotational time-correlation function of water molecules in each system. The relaxation time given by the correlation function has a very strong correlation with the first principal component of the embeddings (Fig. S5), in which the correlation coefficient is 0.99.

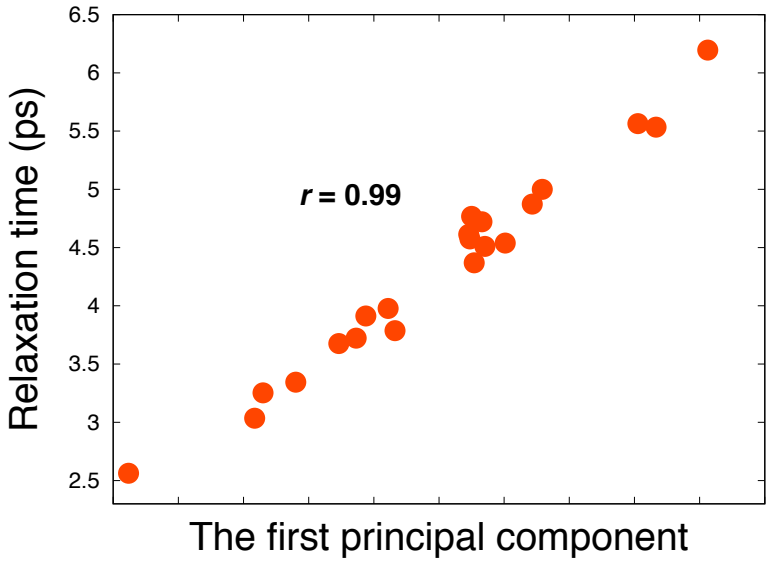

Fig. S 5. Correlation between the rotational relaxation time of water molecules and the first principal component in the embedding space.

Section S4

Details of MD simulations

In the three numerical experiments, we performed isothermal-isobaric MD simulations using the GROMACS packages[3]. The leapfrog algorithm was used for time integration. The SHAKE algorithm was used to constrain the bond lengths and angles of water molecules. A time step of 2 fs was used, with short-range interactions truncated at 1 nm. Long-range electrostatic interactions were calculated using the particle mesh Ewald method[4,5]. In the calculation of interactions between different molecular species, the Lorentz-Berthelot combining rules[6] were used. The temperature was controlled by the Nosé-Hoover thermostat[7,8], except for the Lennard-Jones particle system (see below). The pressure was controlled by the Parrinello-Rahman barostat[9]. Periodic boundary conditions were applied in all directions of systems. The specific simulation conditions are as follows.

Lennard-Jones particles

We prepared 16 systems, each of which consisted of 216 argon molecules. The OPLS-UA force field[10] was used for the Lennard-Jones parameters of argon molecules. The temperature was controlled using the velocity scaling method. The MD data of each system consisted of positions (x, y, z coordinates) of all molecules every 1 ps. The local dynamics ensembles were extracted from the 5 ns of data.

Ethanol and water mixture

We prepared 25 systems, each of which consisted of 1023 molecules. The S_0 contained only ethanol molecules, and the 33 ethanol molecules were exchanged for water molecules, with the increase of the system's index. OPLS-UA and TIP4P/2005[11] force fields were applied to the ethanol and water molecules, respectively[12]. The MD data of each system consisted of positions (x, y, z coordinates) of all molecules every 2 ps. The

local dynamics ensembles were extracted from the 1 ns of data.

Amino acid solution

We prepared 20 amino acid solutions (800 water and 40 single-species amino acid molecules in each system) plus one pure water system (800 water molecules). The modified AMBER FF99SB-ILDN[13,14] was applied to model zwitterionic forms of amino acids, and TIP4P-Ew[15] force fields were applied to the water molecules. The partial charges for amino acids reported by Horn[16] were used. The MD data of each system consisted of positions (x, y, z coordinates) of all molecules every 2 ps. The local dynamics ensembles were extracted from the 20 ns of data.

Analysis method for the rotation of water molecules

The rotation of water molecules is typically described by the rotational time-correlation function $\langle P_2[\boldsymbol{u}(0)\cdot\boldsymbol{u}(t)]\rangle$, where P_2 is the second-order Legendre polynomial, and $\boldsymbol{u}(t)$ is the rotation of the dipole moment at time t[17]. The rotational relaxation of water molecules near amino acids demonstrates a slower decay than that in a bulk water system. For each time-correlation function, the rotational relaxation time of water molecules is given by an exponential fit on the 2-10 ps interval. In this interval, the contributions of fast librational relaxation and long-time migration to other environments are avoided.

Notes and references

[1] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin and A. C. Courville, *Advances in Neural Information Processing Systems*, 2017, 5767–5777.

[2] P. M. Esfahani and D. Kuhn, *Math. Program.*, 2018, **171**, 115–166.

[3] B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

[4] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

[5] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

[6] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, 1989.

[7] S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.

[8] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695.

[9] M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

[10] W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.*, 1984, **106**, 6638–6646.

[11] J. L. Abascal and C. Vega, *J. Chem. Phys.*, 2005, **123**, 234505.

[12] G. Guevara-Carrion, J. Vrabec and H. Hasse, *J. Chem. Phys.*, 2011, **134**, 074508.

[13] P. S. Nerenberg, B. Jo, C. So, A. Tripathy and T. Head-Gordon, *J. Phys. Chem. B*, 2012, **116**, 4524–4534.

[14] D. E. Chapman, J. K. Steck and P. S. Nerenberg, *J. Chem. Theory Comput.*, 2013, **10**, 273–281.

[15] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.

[16] A. H. Horn, *J. Mol. Model.*, 2014, **20**, 2478.

[17] D. Laage, G. Stirnemann, F. Sterpone, R. Rey and J. T. Hynes, *Annu. Rev. Phys. Chem.*, 2011, **62**, 395–416.