Electronic Supplementary Material (ESI) for Nanoscale.

Supplementary Information

The multivariate interaction between Au and TiO_2 colloids: the role of surface potential, concentration and defects

Kinran Lau,[a] Brian Giera,[b] Stephan Barcikowski*[a], Sven Reichenberger*[a]

[a]Technical Chemistry I, University of Duisburg-Essen and Center for Nanointegration Duisburg-Essen (CENIDE), Essen, Germany
[b]Center for Engineered Materials and Manufacturing, Lawrence Livermore National Laboratory, California, USA

Contents

---

1 Model configuration

Details of the model. To limit the simulation time, only $\frac{1}{8}$ of the square prism was simulated. The whole unit cell was generated by symmetry and the overall energy was obtained by multiplying by a factor of 8 (Figure S1). To account for the change in Debye length (κ^{-1}) due to concentration variations, the cell height was set to $10\kappa^{-1}$ instead of fixing it to an absolute length. Similarly, the cell width ($2w$) was also defined relative to κ^{-1}, where the exact relationship is described below. The "extra fine mesh" option was chosen for the mesh size in the COMSOL Multiphysics® software. The average simulation time for each set of parameters was 10.6 s on a computer with an i5-9500 CPU and 8 GB RAM, and a total of 90000 (6000×15 steps in κ^{-1}) conditions were simulated.

Definition of half width (w) (Figure S1c). The half width (w) of the model was limited to roughly ($R + \kappa^{-1}$) to shorten the simulation time. More specifically, w was defined by this relationship: $w = a[ceil(\frac{R+\kappa^{-1}}{a}) + 0.5]$, where R, κ^{-1} and a were the particle radius, Debye length, and defect separation respectively. For $Conc = 0.1$ mM, w was further limited to approximately ($R + 0.7\kappa^{-1}$) because κ^{-1} was much longer at this low concentration which greatly increased the simulation time. However, for all the cases, the potential at the edge of the unit cell was confirmed to decay to at least 20% ($<< \frac{1}{e} = 36.8\%$) of the original particle potential so the width of model should already be long enough to capture most of the potential changes in the solution. Also, the half width (w) was defined as multiples of $0.5a$, such that the full width ($2w$) would be multiples of the defect separation (a). This ensured all the defects were lying within the surface, and this square surface with defects was essentially a unit cell where an infinitely large version of itself could be constructed simply by translation.

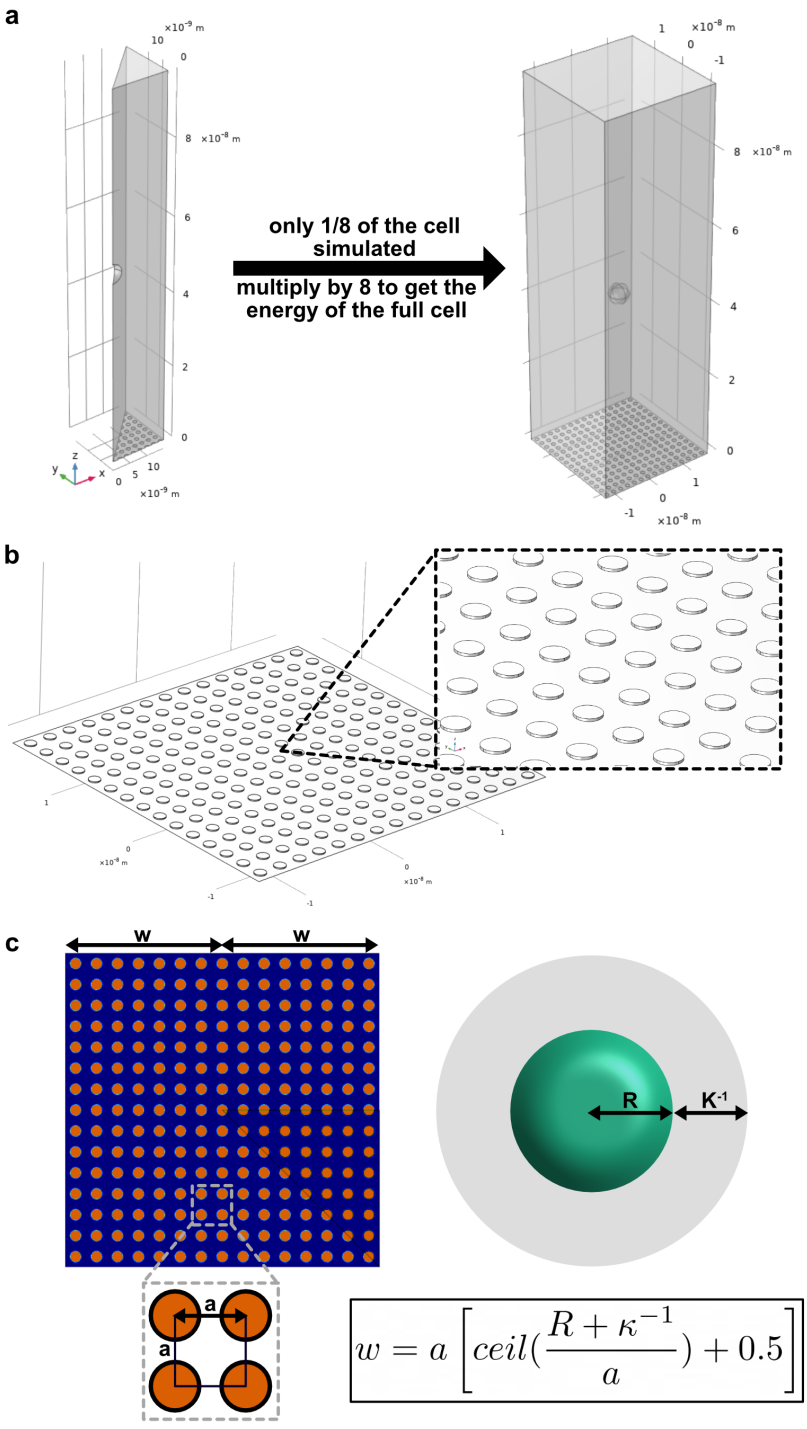

Figure S1: (a) Only $\frac{1}{8}$ of the square prism is simulated. The whole unit cell is generated by symmetry and the overall energy is obtained by multiplying by a factor of 8; (b) The base of the model which contains the defects; (c) The half width (w) of the square prism is not fixed, but defined relative to the particle radius (R), defect separation (a), and Debye length (κ^{-1}).

2 Electrostatic free energy (F_{el}) derivation

The electrostatic free energy (F_{el}) is calculated following the expressions derived by Theodoor and Overbeek,[1] and adapted by Krishnan.[2] The important equations in the main manuscript are repeated here followed by the derivations. Table S1 summarizes the equations in relation to the original papers.

Table S1: Summary of equation numbers in this work, Theodoor and Overbeek,[1] and Krishnan.[2]

This work	Theodoor and Overbeek[1]	Krishnan[2]
(S2)	(6)	/
(S3)	(6)	(1)
(S4) and (S8)	(15)	(5)
(S5) and (S11)	(5)	(2)
(S6) and (S12)	(7)	(3)
(S7)	(13)	in text before (5)

We denote φ as the dimensional potential with a unit of V, and ψ as the dimensionless potential. They can be interconverted by multiplying/dividing by the thermal voltage (Eq. S1). The additional subscript of φ_0 clarifies that this is the potential at the surface, but not that in the electrolyte.

$$\psi = \frac{e\varphi}{kT} \tag{S1}$$

Firstly, the electrostatic potential (dimensional φ or dimensionless ψ) is obtained by solving the Poisson-Boltzmann equation (Eq. S2).

$$\nabla^2\varphi = -\frac{\rho}{\epsilon_0\epsilon_r} \tag{S2}$$

where ρ is the charge density ($\mathrm{C\,m^{-3}}$), $\epsilon_0\epsilon_r$ is the permittivity of the medium.

For a symmetric monovalent electrolyte (e.g. NaCl), the expression simplifies to Eq. S3, where κ^{-1} is the characteristic Debye length of an electrical double layer. ($\frac{1}{\kappa} = \sqrt{\frac{\epsilon_0\epsilon_r kT}{2e^2 I}}$, where e is the elementary charge and I is the ionic strength (number $\mathrm{m^{-3}}$).)

$$\nabla^2\psi = \kappa^2 \sinh(\psi) \tag{S3}$$

This non-linear Poisson-Boltzmann equation is best solved numerically, which can be done by the software COMSOL Multiphysics® throughout the whole model to obtain a potential map in 3D. The computed spatial distribution of ψ is subsequently used to calculate the electrostatic free energy (F_{el}), which has 3 energy components: electrostatic potential energy (U_{el}), configurational

entropy involved in the distribution of ions (ΔS), and chemical free energy due to the adsorption and dissociation of ions (F_{chem}) (Eq. S4–S7),

$$F_{el} = U_{el} - T\Delta S + F_{chem} \tag{S4}$$

$$U_{el} = \frac{1}{2}\int_A \sigma\varphi_0 \, dA - c_0 \, k \, T \int_V \psi \, \sinh(\psi) \, dV \tag{S5}$$

$$\Delta S = 2 \, c_0 \, k \int_V \Big[- \psi \, \sinh(\psi) + \cosh(\psi) - 1 \Big] \, dV \tag{S6}$$

$$F_{chem} = - \int_A \sigma \, \varphi_0 \, dA \tag{S7}$$

where c_0, σ, and φ_0 are the bulk salt concentration (number m^{-3}), surface charge density ($\mathrm{C\,m}^{-2}$), and surface potential (V) respectively. σ can be computed from φ by $\sigma = -\epsilon \frac{\partial \varphi}{\partial n}$.

Substituting them back to F_{el} gives Eq. S8.

$$F_{el} = -\frac{1}{2}\int_A \sigma\varphi_0 \, dA + c_0 \, k \, T \int_V \psi \, \sinh(\psi) \, dV - 2 \, c_0 \, k \, T \int_V \Big[\cosh(\psi) - 1 \Big] \, dV \tag{S8}$$

Note that a boundary condition of constant potential is used in this work, which has a contribution from F_{chem} (Eq. S7). If a boundary condition of constant charge is used, the expression is simply $F_{el,\ cc} = U_{el} - T\Delta S$ without the F_{chem} term, which yields Eq. S9.

$$F_{el,\ cc} = +\frac{1}{2}\int_A \sigma\varphi_0 \, dA + c_0 \, k \, T \int_V \psi \, \sinh(\psi) \, dV - 2 \, c_0 \, k \, T \int_V \Big[\cosh(\psi) - 1 \Big] \, dV \tag{S9}$$

Nevertheless, both F_{el} and $F_{el,\ cc}$ consist of a surface integral and two volume integrals. The surface integral should be performed on all the surfaces. For our model, this would be the particle, the defects, and the normal surface. On the other hand, the volume integrals should be evaluated for the bulk electrolyte.

2.1 Derivation from Eq. S2 to S3

We begin with the standard Poisson-Boltzmann equation (Eq. S2).

$$\nabla^2 \varphi = -\frac{\rho}{\epsilon_0 \epsilon_r} \tag{S2}$$

The charge density ρ ($\mathrm{C\,m}^{-3}$) is the product of charges $z_i e$ (C) and the ion concentration n (number m^{-3}), where z_i is the charge number, e is the elementary charge, and c_0 is the bulk salt concentration (Eq. S10).

$$\begin{aligned}\rho &= \sum_i \underbrace{z_i e}_{\text{charges}} \underbrace{n}_{\text{conc.}} \\ &= \sum_i z_i e \left[c_0\, exp(-z_i\, \psi)\right]\end{aligned} \tag{S10}$$

Substitute Eq. S10 into Eq. S2 yields the following:

$$\nabla^2 \varphi = -\frac{e}{\epsilon_0 \epsilon_r} \sum_i z_i \left[c_0\, exp(-z_i\, \psi)\right]$$

Consider a symmetric electrolyte (e.g. NaCl) with an equal number of "+z" and"-z" ions, the expression becomes:

$$\nabla^2 \varphi = -\frac{e}{\epsilon_0 \epsilon_r} \left[\underbrace{(+z)\left[c_0\, exp(-z\, \psi)\right]}_{\text{+z ions}} + \underbrace{(-z)\left[c_0\, exp(+z\, \psi)\right]}_{\text{-z ions}} \right]$$

$$\nabla^2 \varphi = -\frac{zec_0}{\epsilon_0 \epsilon_r} \left[exp(-z\, \psi) - exp(+z\, \psi)\right]$$

Since $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$,

$$\nabla^2 \varphi = +\frac{2zec_0}{\epsilon_0 \epsilon_r} \sinh(z\psi)$$

Since $\frac{1}{\kappa} = \sqrt{\frac{\epsilon_0 \epsilon_r kT}{2e^2 I}}$ and $I = \frac{1}{2}\sum_i c_i z_i^2 = c_0$ for a symmetric monovalent electrolyte ($z = 1$),

$$\nabla^2 \varphi = \kappa^2 \left(\frac{kT}{e}\right) \sinh(\psi)$$

$$\left(\frac{e}{kT}\right) \nabla^2 \varphi = \kappa^2 \sinh(\psi)$$

The dimensional potential φ can be converted to the dimensionless potential ψ by multiplying a factor of $\frac{e}{kT}$ (Eq. S1), which gives Eq. S3.

$$\nabla^2 \psi = \kappa^2 \sinh(\psi) \tag{S3}$$

2.2 Derivation of Eq. S5

We begin with the following expression which is Eq. 5 in Theodoor and Overbeek[1] and Eq. 2 in Krishnan[2].

$$U_{el} = \frac{1}{2}\int_A \sigma \varphi_0\, dA + \frac{1}{2}\int_V \rho \varphi\, dV \tag{S11}$$

The first term already matches the final form in Eq. S5, but the second term can be further simplified. Substituting the expression for charge density (ρ) (Eq. S10) into second term of Eq. S11 gives:

$$\frac{1}{2}\int_V \rho\varphi\, dV = \frac{1}{2}\int_V \varphi \sum_i z_i e\,[c_0\, exp(-z_i\,\psi)]\; dV$$

For a symmetric monovalent electrolyte (e.g. NaCl), $z = \pm 1$:

$$\frac{1}{2}\int_V \rho\varphi\, dV = \frac{1}{2}\int_V \varphi\Big[\underbrace{(+1)e\,[c_0\, exp(-\psi)]}_{z=+1} + \underbrace{(-1)e\,[c_0\, exp(+\psi)]}_{z=-1}\Big]\, dV$$

$$= \frac{1}{2}\int_V \varphi\, c_0\, e[exp(-\psi) - exp(+\psi)]\, dV$$

Since $\sinh(x) = \frac{1}{2}(e^x - e^{-x})$,

$$\frac{1}{2}\int_V \rho\varphi\, dV = \frac{1}{2}\int_V -2\,\varphi\, c_0\, e \sinh(\psi)\, dV$$

$$= -c_0\, e \int_V \varphi\, \sinh(\psi)\, dV$$

Convert dimensional potential φ to dimensionless potential ψ with Eq. S1,

$$\frac{1}{2}\int_V \rho\varphi\, dV = -c_0\, e \int_V \psi \left(\frac{kT}{e}\right) \sinh(\psi)\, dV$$

$$= -c_0\, k\, T \int_V \psi\, \sinh(\psi)\, dV$$

Substitute this simplified $\frac{1}{2}\int_V \rho\varphi\, dV$ term back to Eq. S11 yields Eq. S5:

$$U_{el} = \frac{1}{2}\int_A \sigma\varphi_0\, dA - c_0\, k\, T \int_V \psi\, \sinh(\psi)\, dV \tag{S5}$$

2.3 Derivation of Eq. S6

We begin with Eq. 7 in Theodoor and Overbeek[1] and Eq. 3 in Krishnan[2].

$$\Delta S = k \int_V \left\{ \sum_i c_0[z_i\, \psi\, exp(-z_i\, \psi) + exp(-z_i\, \psi) - 1] \right\} dV \tag{S12}$$

For a symmetric monovalent electrolyte (e.g. NaCl), $z = \pm 1$:

$$\Delta S = k \int_V \left\{ \underbrace{c_0[\psi\, exp(-\psi) + exp(-\psi) - 1]}_{\text{z = +1}} + \underbrace{c_0[-\psi\, exp(+\psi) + exp(+\psi) - 1]}_{\text{z = -1}} \right\} dV$$

$$= c_0\, k \int_V \left\{ \psi \Big[\underbrace{exp(-\psi) - exp(+\psi)}_{\sinh(x)=\frac{1}{2}(e^x - e^{-x})} \Big] + \Big[\underbrace{exp(-\psi) + exp(+\psi)}_{\cosh(x)=\frac{1}{2}(e^x + e^{-x})} \Big] - 2 \right\} dV$$

$$= c_0\, k \int_V \left\{ \psi \Big[-2\sinh(\psi) \Big] + \Big[2\cosh(\psi) \Big] - 2 \right\} dV$$

Upon further simplification, it becomes Eq. S6:

$$\Delta S = 2\, c_0\, k \int_V \Big[-\psi\, \sinh(\psi) + \cosh(\psi) - 1 \Big]\, dV \tag{S6}$$

3 Comparing perfect and defective surfaces

By comparing perfect (uniformly charged) and defective (non-uniformly charged) surfaces, the objective is to identify conditions where defects are expected to play a significant role. This can be done by comparing perfect and defective surfaces which have the same effective surface potential (V_{Eff}). For perfect surfaces, V_{Eff} is simply the surface potential (V_{Surf}) due to the lack of defects. For defective surfaces, V_{Eff} is the area-averaged surface potential of the normal surface and the defects, resulting from various combinations of the surface potential (V_{Surf}), defect potential (V_{Def}) and defect density (DD).

Table S2: Parameter space of perfect (uniformly charged) surfaces without defects ($DD = 0$ and $V_{Def} = 0\,\mathrm{mV}$), which covers 1260 unique combinations of these 4 parameters (V_{Part}, V_{Surf}, $Conc$, R). For each condition, the particle is moved towards the surface in 15 steps (D).

	Parameter	Values	Unit
V_{Part}	Particle potential	-51.4, -38.5, -25.7, -12.8	mV
V_{Surf}	Surface potential	-51.4 to 0.0 (21 steps)	mV
V_{Def}	Defect potential	0.0	mV
DD	Defect density	0.0	/
$Conc$	Salt concentration	0.1, 0.5, 1.0	mM
R	Particle radius	0.5, 1.5, 2.5, 3.5, 4.5	nm
D	Particle-Surface separation	0.1 to 4.1 (15 steps)	κ^{-1} (Debye length)

In our defect-containing model, we considered 6 parameters: V_{Part}, V_{Surf}, V_{Def}, DD, $Conc$, and R. For perfect surfaces, since both DD and V_{Def} were set to 0, only 4 variables (V_{Part}, V_{Surf}, $Conc$, R) remained. To explore the effectiveness of V_{Eff} as a descriptor, we conducted additional simulations spanning 1260 unique conditions of these 4 parameters for perfect surfaces (Table S2). For V_{Part}, $Conc$, and R, the same values as those in the defective parameter space (Table 1) were used to perform parametric sweeps. However, for V_{Surf}, we explored 21 steps from -51.4 mV to 0 mV to encompass the entire range of V_{Eff}. This extensive range in V_{Surf} was selected to account for the various combinations of V_{Def} and DD that could yield a broad spectrum of V_{Eff}. Throughout all simulations, the Hamaker constant was fixed at 100 zJ to calculate the van der Waals contributions.

To evaluate how well V_{Eff} could represent defective surfaces, we examined whether the interaction energy barriers aligned when both perfect and defective surfaces exhibited the same V_{Eff}. Specifically, when the interaction energy barriers were plotted against V_{Eff}, the energy barriers of both the perfect and defective cases should fall on the same curve in the ideal case when V_{Eff} could fully describe defective surfaces. However, such perfect alignment was not anticipated. To quantify the degree of alignment, we computed the coefficient of determination (R^2) for each case (Eq. S13). This involved treating the defective cases as the "data points" (y_i) and the perfect surfaces as the "predicted values" ($\hat{y}_i$).

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{S13}$$

where R^2 is the coefficient of determination, y_i is the observed value at point i, $\hat{y}_i$ is the predicted value at point i, and $\bar{y}$ is the mean of the observed values. Larger R^2 values indicate better agreement between the perfect and defective surfaces, and hence lower importance of non-uniformities in the surface charge distribution.

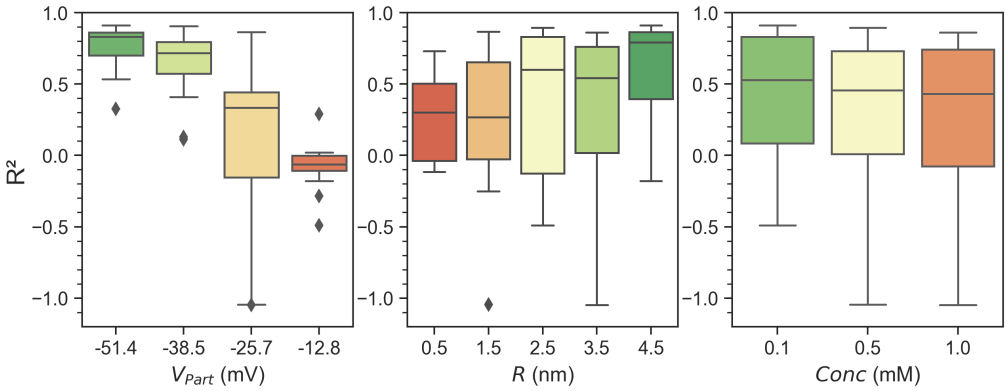

Figure S2: Box plot showing the distribution of R^2 values by parameters (V_{Part}, R, and $Conc$).

Our analysis yielded a total of 60 curves depicting the interaction energy barriers against V_{Eff}, considering all the possible combinations of V_{Part}, R, and $Conc$ (Figure S3-S6). To provide a more meaningful overview, we examined the distribution of the 60 R^2 values by parameters (Figure S2). Notably, there were some instances when the defective energy barriers mostly located near the perfect surface curves, as indicated by the high R^2 values > 0.75. However, in other cases, the dispersion of energy barriers in the defective scenarios deviated significantly from the perfect surface curves, resulting in small and even negative R^2 values. From the box plot in Figure S2, it appears that the largest R^2 values could be achieved when V_{Part} was the most negative (-51.4 mV), R was the largest (4.5 nm), and $Conc$ was the lowest (0.1 mM), which suggested that under these conditions, perfect surfaces provided a reasonable description of defective surfaces with the same V_{Eff}. Among these factors, V_{Part} had the most pronounced influence on the resultant R^2 values, while R and $Conc$ had comparatively less effect on the outcome. Specifically, when particles possessed highly negative charges (V_{Part} = -51.4 mV), the deviations in the energy barriers between the two scenarios were relatively small (large R^2 values). However, as V_{Part} became less negative, there was a significant decrease in the R^2 values, highlighting the growing importance of including defects in the model. Consequently, defects were explicitly considered in our study to comprehensively describe surface interactions and capture the nuanced impacts of defects on these interactions. Furthermore, the defect-containing model also enabled the possibility to explore surface heterogeneity and examine the effect of local conditions (V_{Def} and DD), which would not be captured by V_{Eff} alone.

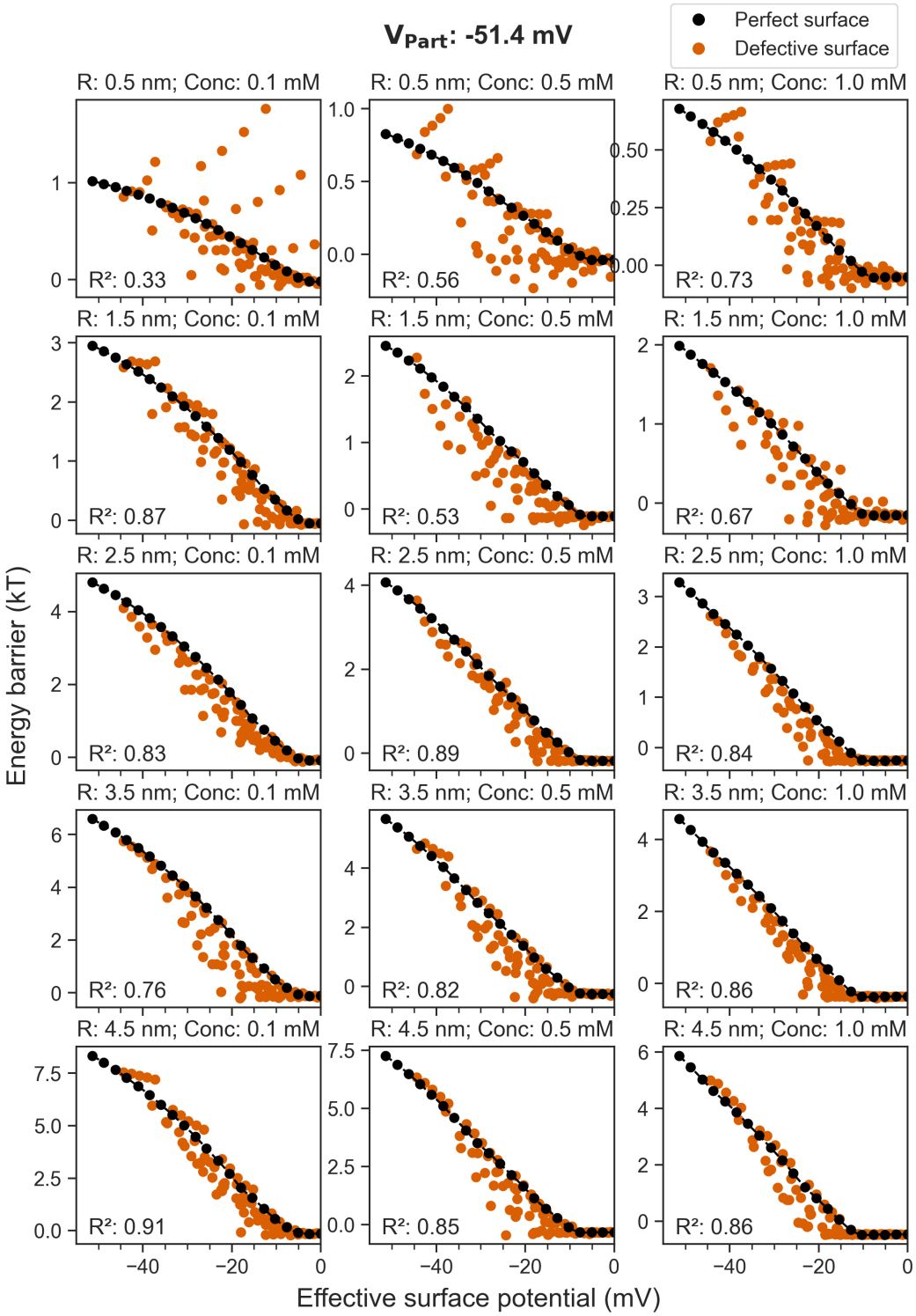

Figure S3: 15 plots of interaction energy barriers against the effective surface potential (V_{Eff}) comparing perfect and defective surfaces, with V_{Part} fixed at -51.4 mV while varying R and $Conc$.

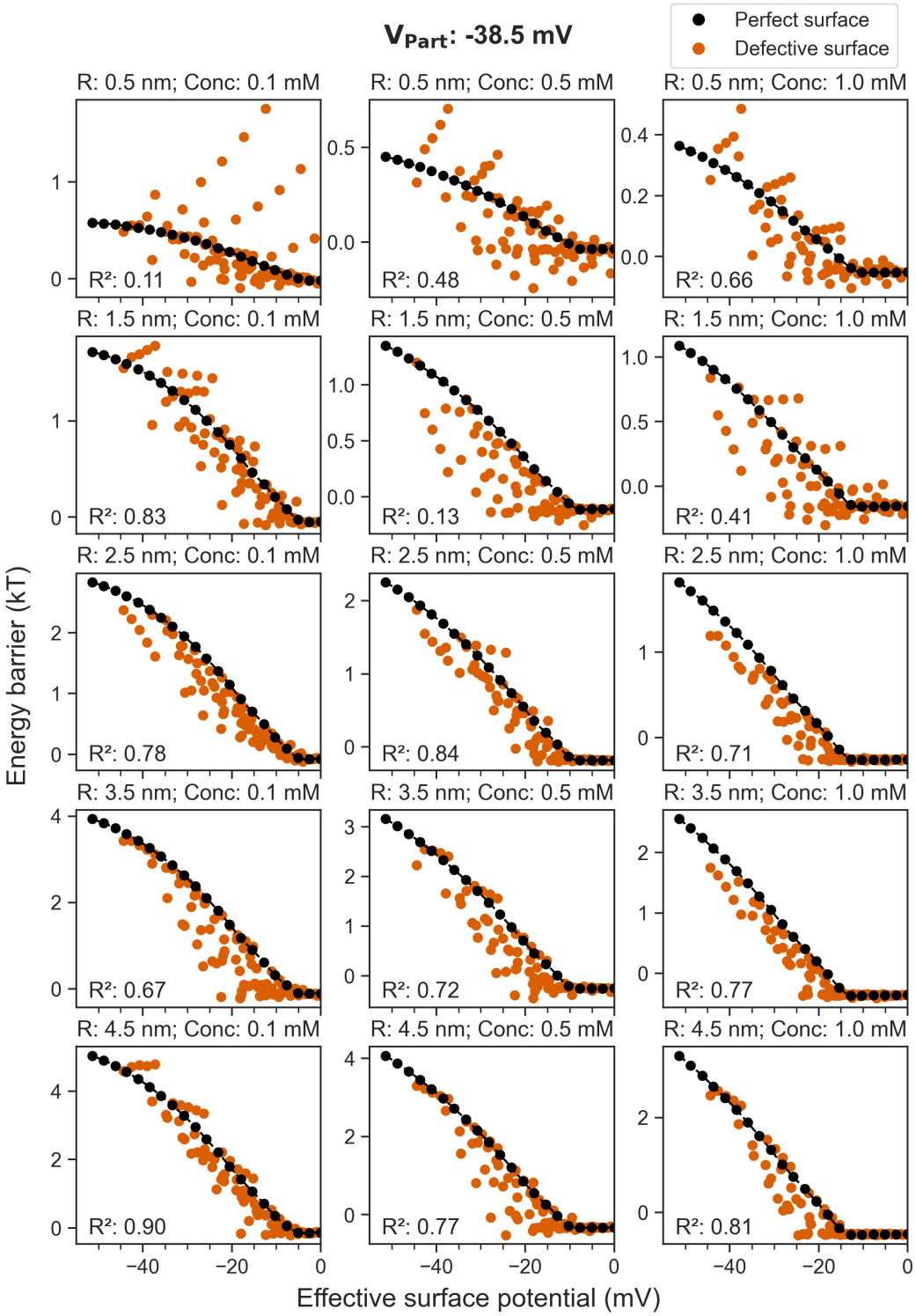

Figure S4: 15 plots of interaction energy barriers against the effective surface potential (V_{Eff}) comparing perfect and defective surfaces, with V_{Part} fixed at -38.5 mV while varying R and $Conc$.

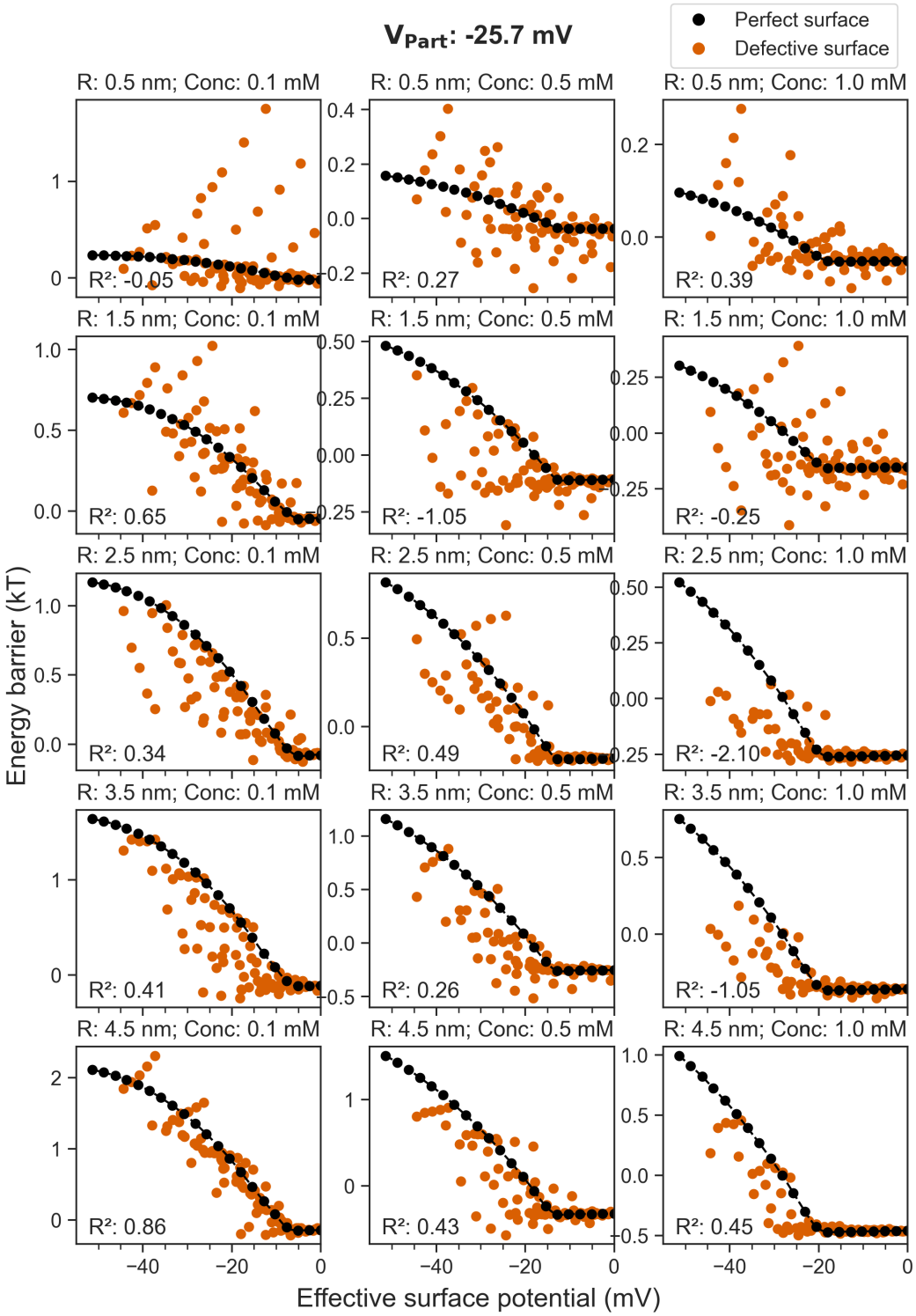

Figure S5: 15 plots of interaction energy barriers against the effective surface potential (V_{Eff}) comparing perfect and defective surfaces, with V_{Part} fixed at -25.7 mV while varying R and $Conc$.

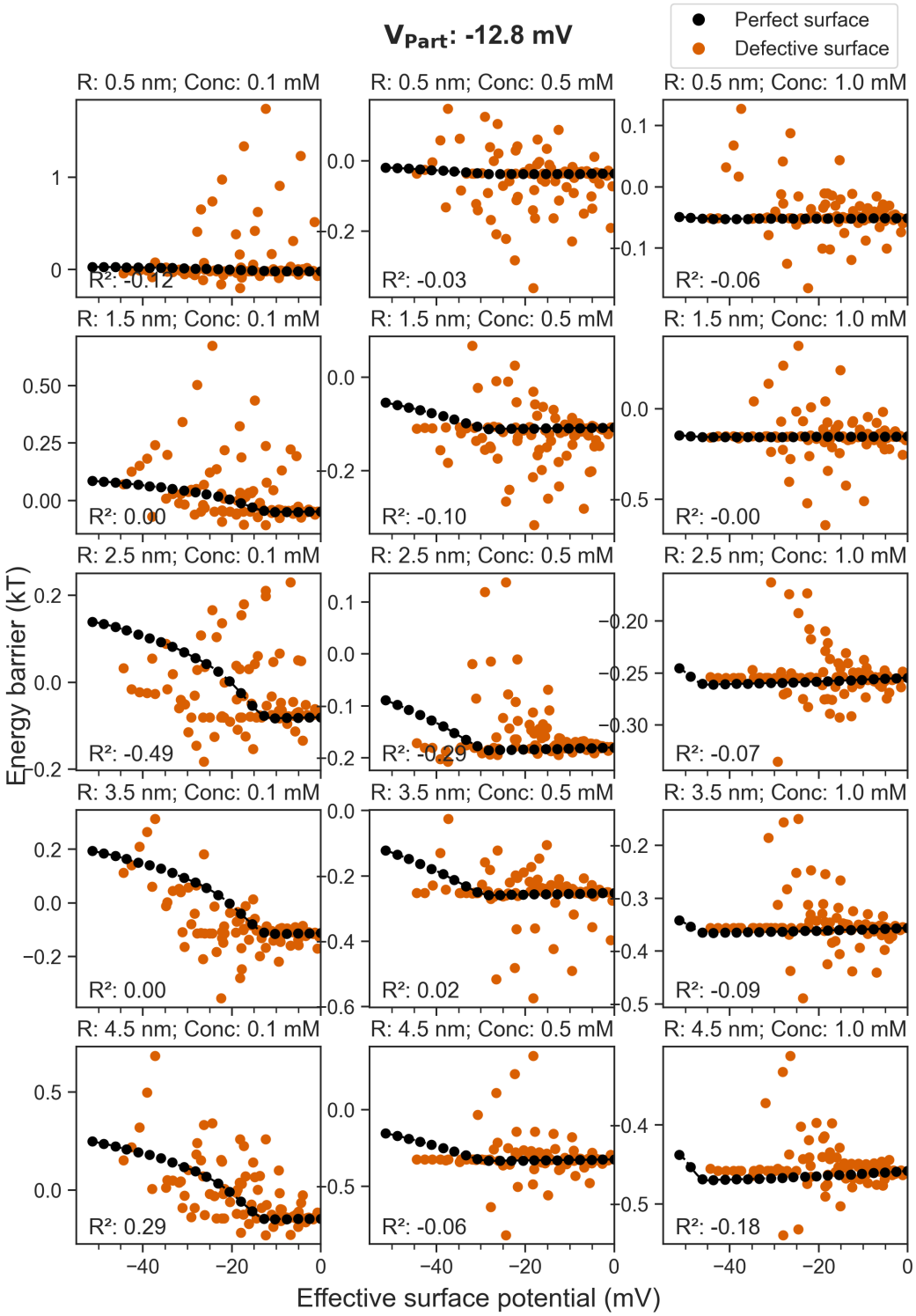

Figure S6: 15 plots of interaction energy barriers against the effective surface potential (V_{Eff}) comparing perfect and defective surfaces, with V_{Part} fixed at -12.8 mV while varying R and $Conc$.

4 Energy distributions with varying Hamaker constants

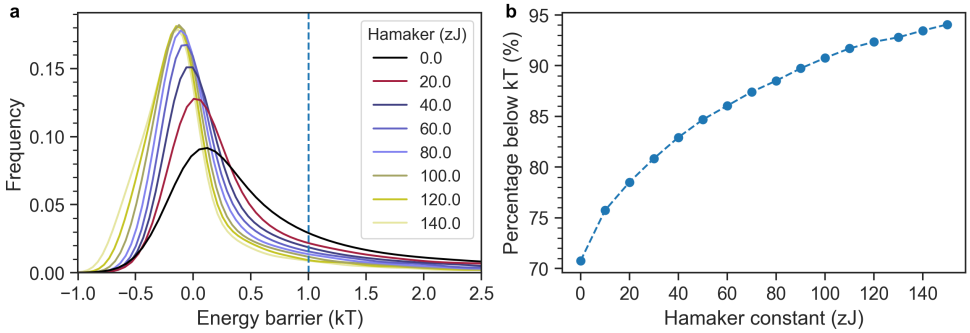

Figure S7: (a) Distribution of energy barriers and (b) percentage of data points below kT with varying Hamaker constants (A_H).

The Hamaker constant describing the interaction between TiO_2 and Au can be estimated by taking the geometric mean of the individual A_H, i.e. $\sqrt{(A_{H,TiO_2})(A_{H,Au})}$. The A_H of TiO_2 in water is fairly consistent in the literature ($\sim$50 zJ),[3,4] but that of Au can vary considerably depending on the study ($\sim$100–300 zJ).[5,6] Therefore, a reasonable estimate of the A_H between TiO_2 and Au ranges from $\sqrt{(50)(100)} \approx 70$ zJ to $\sqrt{(50)(300)} \approx 120$ zJ, where $A_{H,\mathrm{TiO_2}-Au} = 100$ zJ is used in this work for the sake of discussion. Within this range of A_H, although the exact distribution of energy barriers varies slightly, the qualitative conclusions in the paper still hold. Moreover, the whole range of A_H (0–150 zJ, steps of 10 zJ) is considered when fitting the data with a support vector machine for predicting whether the interaction is favorable.

5 Representative cases of energy barriers

In our model, we investigated the colloidal interaction between a negatively charged particle and a negatively charged surface with defects. Specifically, our focus was on understanding whether colloidal particles possess sufficient thermal energy to overcome the energy barrier and potentially adsorb onto the surface. This energy barrier arises from the balance between the repulsive electrostatic forces and the attractive van der Waals forces, similar to the treatment in the classical DLVO theory. The total interaction energy (E_{tot}) is the sum of the repulsive electrostatic free energy (F_{el}) and the attractive van der Waals energy (E_{vdw}), and the energy barrier is determined by the maximum of E_{tot}.

To illustrate the different types of energy barriers resulting from F_{el} and E_{vdw}, we have included three representative cases in Figure S8: a) Energy barrier $> kT$, b) Energy barrier $< kT$, c) No apparent energy barrier.

The conditions in Figure S8a are the same as those in Figure 2 of the manuscript ($V_{Part} = -51.4\,\text{mV}$, $V_{Surf} = -51.4\,\text{mV}$, $V_{Def} = +12.8\,\text{mV}$, $DD = 0.2$, $Conc = 1\,\text{mM}$, $R = 2.5\,\text{nm}$). In these 3 examples, the Hamaker constant (A_H) and particle radius (R) are kept constant, such that the attractive van der Waals components remain identical. However, by varying the surface potential (V_{Surf}) and defect density (DD), we can adjust the degree of repulsion between the particle and the surface, and examine its effect on the interaction energy barrier.

In Figure S8a, F_{el} is sufficiently repulsive that the energy barrier exceeds kT, even when considering the attractive E_{vdw}. This implies that adsorption is unlikely under these conditions. However, when we reduce V_{Surf} from -51.4 mV to -38.5 mV (Figure S8b), F_{el} remains repulsive but the resulting energy barrier now falls below kT, which suggests that adsorption could be possible with thermal fluctuations. In Figure S8c, we increase DD from 0.2 to 0.3. This results in an even weaker repulsive F_{el}, and when combined with the attractive E_{vdw}, there is no apparent energy barrier for the particle to move towards the surface.

To examine whether additional data points are required between the first and second points (0.1 and 0.4 κ^{-1}) along the x-axis (particle-surface separation), we have also simulated four additional data points between them for the 3 representative cases, as indicated by red diamonds in the energy curves in Figure S8. Notably, the inclusion of these points does not affect the interaction energy barriers, which are extracted as the maximum energy of E_{tot}. This observation agrees with our analysis of the distribution of energy barriers depicted in Figure S9, which indicates that the majority of energy barriers happen at particle-surface separations of 0.5 – 1.0 κ^{-1}. Therefore, additional simulations between the first and second points (0.1 and 0.4 κ^{-1}) are unlikely to yield substantial changes in the interaction energy barriers.

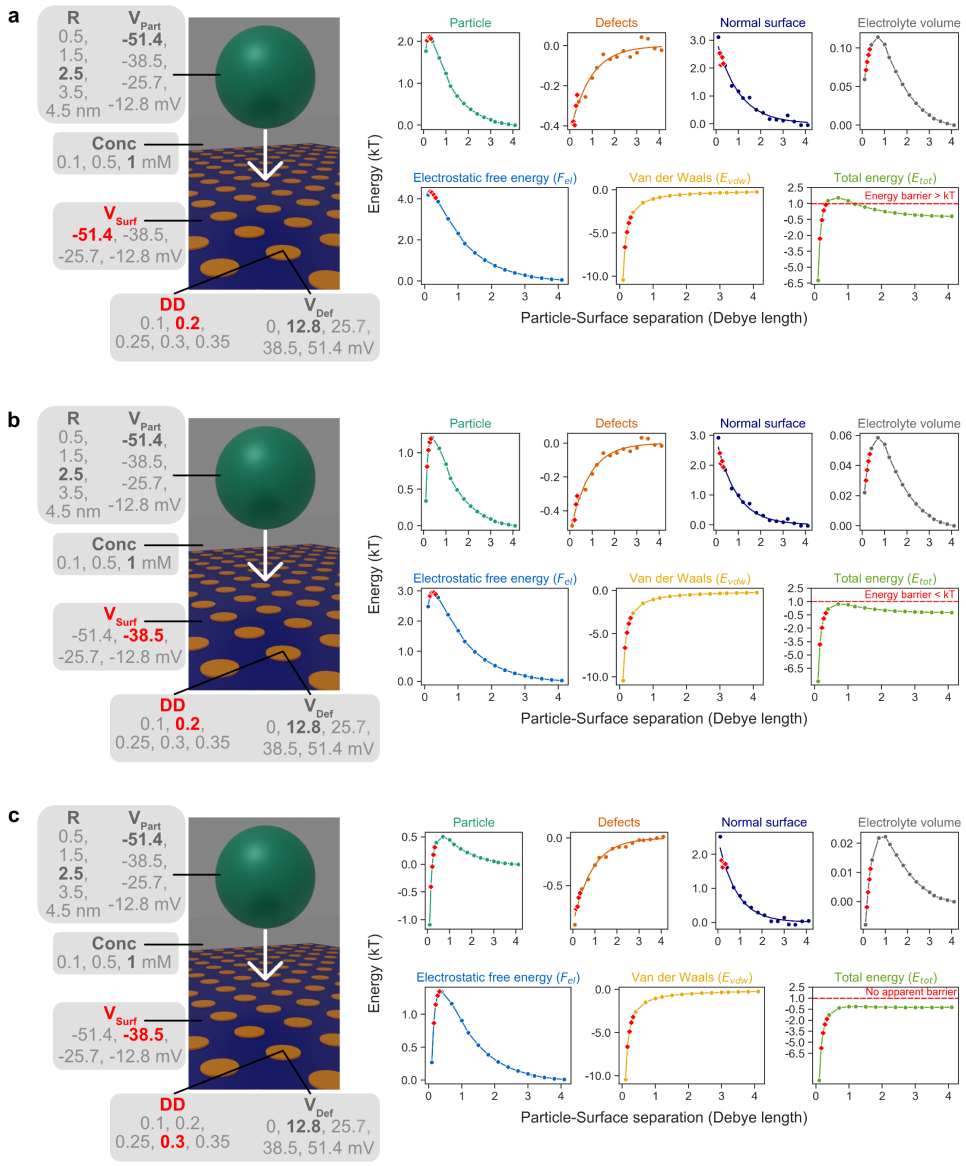

Figure S8: Three representative cases of energy barriers illustrated by varying V_{Surf} and DD, while keeping other factors the same ($V_{Part} = -51.4\,\text{mV}$, $V_{Def} = +12.8\,\text{mV}$, $Conc = 1\,\text{mM}$, $R = 2.5\,\text{nm}$). (a) Energy barrier $> kT$ ($V_{Surf} = -51.4\,\text{mV}$, $DD = 0.2$, identical conditions as Figure 2); (b) Energy barrier $< kT$ ($V_{Surf} = -38.5\,\text{mV}$, $DD = 0.2$); (c) No apparent barrier ($V_{Surf} = -38.5\,\text{mV}$, $DD = 0.3$). The red diamonds are additional data points simulated between 0.1 and 0.4 κ^{-1} to demonstrate that they do not affect the energy barriers.

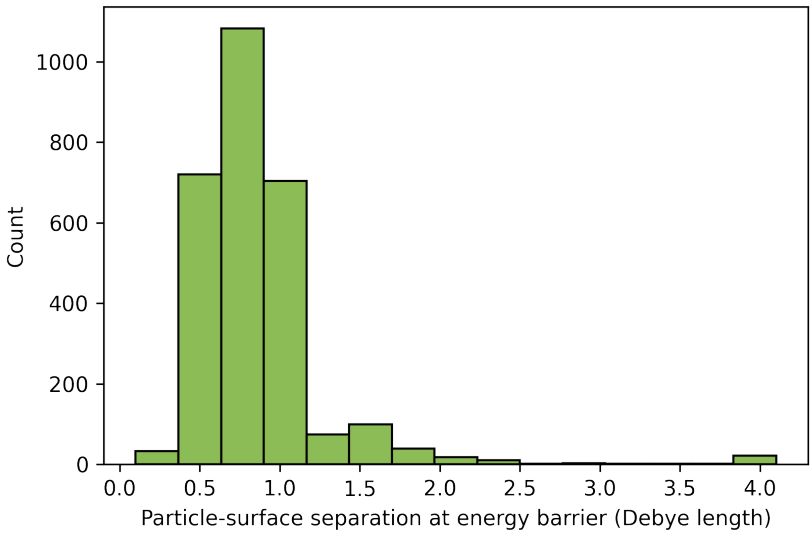

Figure S9: Distribution of particle-surface separations at the energy barriers.

6 Pairwise comparison of parameters by the Tukey's HSD (honestly significant difference) test

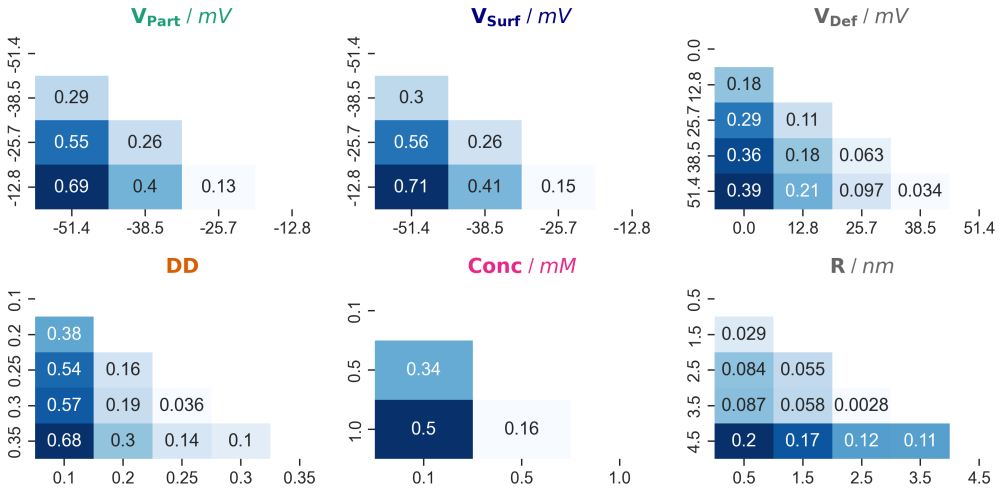

Figure S10: The q-statistics obtained by performing pairwise comparison with Tukey's HSD (honestly significant difference) test.

7 Logistic regression for classifying the "below kT" and "above kT" cases

The colloidal interaction was simulated under 6000 unique conditions, where each condition is a specific combination of these 6 parameters: V_{Part} (particle potential), V_{Surf} (surface potential), V_{Def} (defect potential), DD (defect density), $Conc$ (concentration of salt), and R (particle radius). For each set of parameters, the interaction energy barrier is extracted, and is classified into one of these two classes based on the energy barrier: "below kT" (favorable) or "above kT" (unfavorable). Accordingly, the 6000 cases have 6 parameters (V_{Part}, V_{Surf}, V_{Def}, DD, $Conc$, R) as the input and "below kT" or "above kT" as the label, which becomes a binary classification problem and can be modelled by logistic regression. The logistic regression equations with the fitted coefficients are shown in Eq. S14 and S15.

$$\begin{aligned} y = {} & 6.572 + 2.516||V_{Surf}|| + 2.449||V_{Part}|| + 1.696||DD|| \\ & + 1.417||Conc|| - 1.382||R|| + 1.096||V_{Def}|| \end{aligned} \tag{S14}$$

$$z = \frac{1}{1 + \exp\left(-y\right)} \tag{S15}$$

Given a particular set of parameters, this fitted expression (Eq. S14 and S15) can determine whether it is "below kT" or "above kT" with an accuracy of 95% evaluated using a standard train/validation split of 80%/20%. z takes a value between 0 and 1 since it is a sigmoid function, and it is essentially the probability of "below kT" because the "below kT" and "above kT" cases are arbitrarily assigned to a label of "1" and "0" respectively. Note the 6 parameters (V_{Part}, V_{Surf}, V_{Def}, DD, $Conc$, R) are scaled by z-score normalization before feeding into logistic regression, where the additional $||\ ||$ symbol is used to denote a normalized quantity. Z-score normalization can be performed by subtracting the mean (μ_i) and dividing by the standard deviation (σ_i) of the distribution (Eq. S16), where the Table S3 summarizes the μ_i and σ_i for each factor in this study. As an example, V_{Surf} has a mean and standard deviation of -32.1 mV and 14.36 mV respectively in the distribution, which means that a V_{Surf} of -25 mV converts to a dimensionless $||V_{Surf}||$ of 0.49 after normalization ($||V_{Surf}|| = \frac{V_{Surf} - \mu_i}{\sigma_i} = \frac{-25-(-32.1)}{14.36} = 0.49$). These normalized quantities are used to visualize the decision boundary shown in Figure 4, where the x-axis is the "Electrostatics" metric: $2.516||V_{Surf}|| + 2.449||V_{Part}||$, and y-axis is the "Environment" descriptor: $1.696||DD|| + 1.417||Conc||$.

$$||x_i|| = \frac{x_i - \mu_i}{\sigma_i} \tag{S16}$$

Table S3: Mean (μ_i) and standard deviation (σ_i) of each variable for z-score normalization (Eq. S16).

Parameter		Unit	μ_i	σ_i
V_{Part}	Particle potential	mV	-32.10	14.36
V_{Surf}	Surface potential	mV	-32.10	14.36
V_{Def}	Defect potential	mV	+25.68	18.16
DD	Defect density	/	0.24	0.086
$Conc$	Salt concentration	mM	0.533	0.368
R	Particle radius	nm	2.5	1.414

8 Attempts to cluster data points with handcrafted dimensionless numbers

In our model, a total of 6000 conditions were simulated by varying 6 parameters (V_{Part}, V_{Surf}, V_{Def}, DD, $Conc$, R). Each data point carries a label of either "below kT" (1) or "above kT" (0) derived from its computed energy barrier. To unambiguously specify a data point in the dataset, a particular combination of the 6 parameters is needed. However, when fewer than 6 parameters are considered, there will be overlapping points existing in this low-dimensional space. By evaluating the average of the labels (0 or 1) at the overlapping points, the probability of "below kT" can be estimated. This likelihood of "below kT" is used to visualize the clustering of data points in Figure 4 and S11: green for "below kT", red for "above kT", and yellow for the uncertain cases.

In an attempt to split the data points into regions of "below kT" and "above kT", the following dimensionless numbers are constructed out of the 6 parameters: $\frac{V_{Surf}}{V_{Part}}$, $\frac{V_{Surf}}{V_{Def}}$, $\frac{V_{Part}}{V_{Def}}$, $\frac{(V_{Surf}+V_{Part})}{V_{Def}}$, DD, $\frac{\kappa^{-1}}{R}$. It is noted that even though R and V_{Def} are not among the most influential factors as suggested by ANOVA and logistic regression, they are involved in formulating the dimensionless numbers such that the units can be cancelled out. $Conc$ (mM) does not have any complementary parameters with the same unit, and is therefore converted to the Debye length (κ^{-1}, nm), and divided by R (nm) to yield a dimensionless number $\frac{\kappa^{-1}}{R}$.

However, these handcrafted dimensionless numbers do not separate the "below kT" (green) and "above kT" (red) cases well, as indicated by the lack of red dots on the graphs, i.e. the "below kT" and "above kT" cases are overlapping, resulting in mostly green and some yellow/orange points. The plots of DD vs. $\frac{(V_{Surf}+V_{Part})}{V_{Def}}$ and $\frac{\kappa^{-1}}{R}$ vs. $\frac{(V_{Surf}+V_{Part})}{V_{Def}}$ are capable of separating some red dots out of the dataset. Yet, it is also apparent that some intermediate (yellow/orange) cases can also be found in the green region of "below kT". By contrast, the linear combination of z-score normalized parameters ($2.516||V_{Surf}||+ 2.449||V_{Part}||$ and $1.696||DD|| + 1.417||Conc||$) is much more effective to cluster the data points, as manifested by the diagonal transition from red to yellow and eventually to green from the bottom left to the top right of Figure 4.

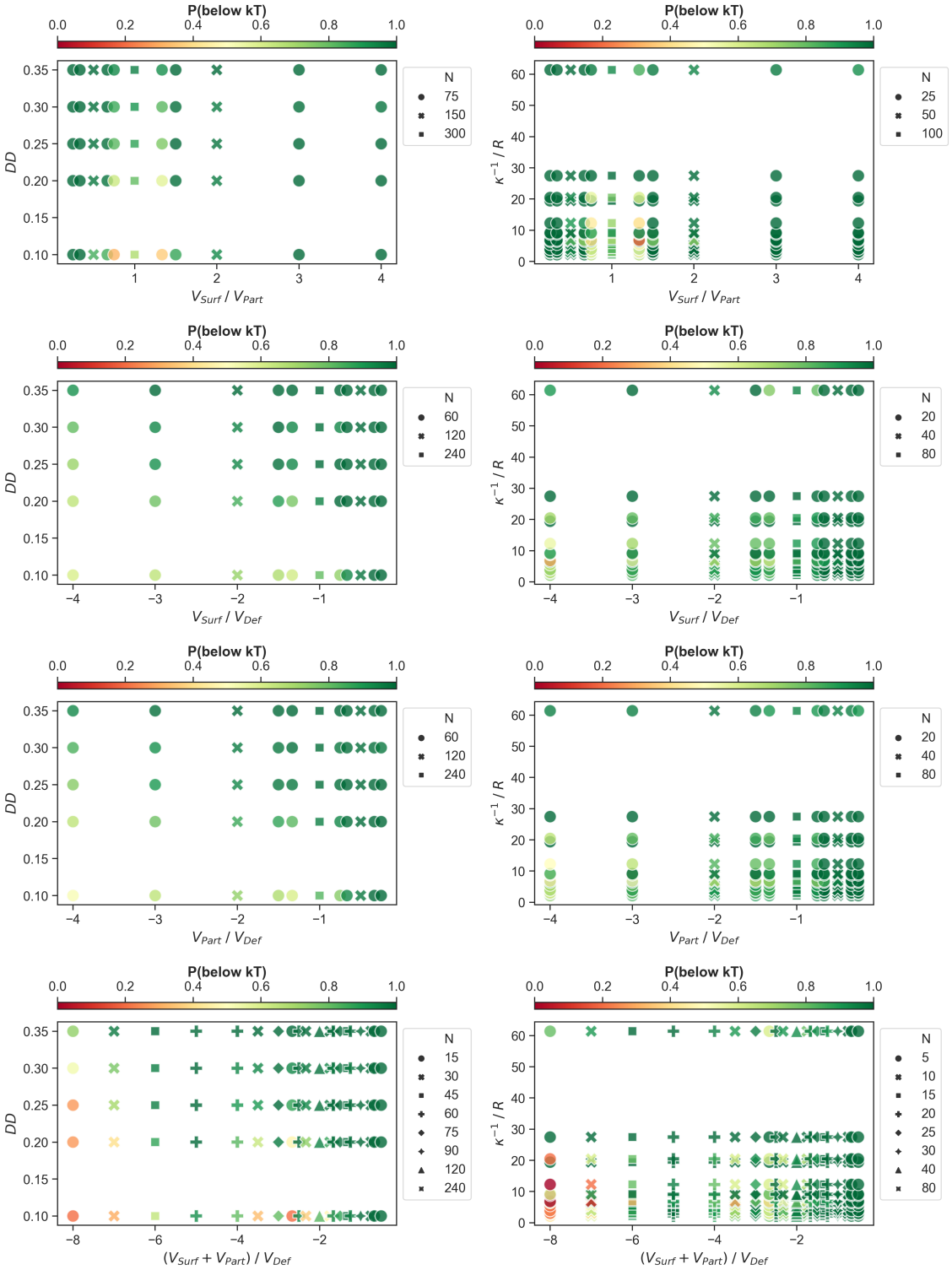

Figure S11: Attempts to cluster data points into regions of "below kT" (green) and "above kT" (red) with different combinations of dimensionless numbers.

9 Support vector machine (SVM) for predicting the interaction outcome

For defective surfaces, we conducted simulations covering 6000 combinations of these 6 parameters (V_{Part}, V_{Surf}, V_{Def}, DD, $Conc$, R) (Table 1). Additionally, to examine perfect surfaces without defects, we simulated an additional 1260 conditions (Table S2). This yielded a total of $6000 + 1260 = 7260$ electrostatic free energy (F_{el}) curves computed with the software COMSOL. However, it is important to note that the total interaction energy (E_{tot}) comprises both electrostatic free energy (F_{el}) and van der Waals energy (E_{vdw}). Previously, the Hamaker constant (A_H) was fixed at 100 zJ for the sake of discussion. To consider the whole range of van der Waals interactions in the SVM, A_H was varied from 0 to 150 zJ with a step size of 10 zJ. This resulted in an extended dataset with $7260 \times 16 = 116160$ entries. To effectively work with this comprehensive dataset, we employed a Support Vector Machine (SVM) for fitting. Similar to the logistic regression analysis described in Section 7, the parameters were first scaled by z-score normalization before being fed into the SVM. The 116160 energy barriers were classified into one of these two groups: "below kT" or "above kT". The SVM was then trained to determine the decision boundary between these two cases given a specific set of the 7 parameters (V_{Part}, V_{Surf}, V_{Def}, DD, $Conc$, R, A_H). An accuracy of 97% was achieved, which was assessed by a standard train/validation split of 80%/20%. The radial basis function (RBF) was used as the kernel of the SVM, and default C and gamma values of the scikit-learn python package were used when training the SVM. Alongside the prediction of "below kT" or "above kT", a decision function indicating the confidence of the prediction would also be outputted (a larger absolute value means higher confidence) to alert the users for cases relatively close to the decision boundary.

References

[1] J. Theodoor and G. Overbeek, *Colloids and Surfaces*, 1990, **51**, 61–75.

[2] M. Krishnan, *The Journal of Chemical Physics*, 2013, **138**, 114906.

[3] L. Bergström, *Advances in Colloid and Interface Science*, 1997, **70**, 125–169.

[4] J. M. Fernández-Varea and R. Garcia-Molina, *Journal of Colloid and Interface Science*, 2000, **231**, 394–397.

[5] V. Parsegian and G. H. Weiss, *Journal of Colloid and Interface Science*, 1981, **81**, 285–289.

[6] K. Jiang and P. Pinchuk, *Nanotechnology*, 2016, **27**, 345710.