Active learning of polarizable nanoparticle phase diagrams for the guided design of triggerable self-assembling superlattices

Journal:	*Molecular Systems Design & Engineering*
Manuscript ID	ME-ART-12-2021-000187.R1
Article Type:	Paper
Date Submitted by the Author:	20-Jan-2022

Complete List of Authors:	Dasetty, Siva; University of Chicago, Pritzker School of Molecular Engineering Coropceanu, Igor; University of Chicago Division of the Physical Sciences, Chemistry Portner, Josh; University of Chicago Division of the Physical Sciences, Chemistry Li, Jiyuan; University of Chicago, Pritzker School of Molecular Engineering de Pablo, Juan; Liew Family Professor of Molecular Theory and Simulation, Institute for Molecular Engineering Talapin, Dmitri; The University of Chicago, Department of Chemistry Ferguson, Andrew; University of Chicago, Institute for Molecular Engineering

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Active learning of polarizable nanoparticle phase diagrams for the guided design of triggerable self-assembling superlattices†

Siva Dasetty,[a] Igor Coropceanu,[b] Joshua Portner,[b] Jiyuan Li,[a] Juan J. de Pablo,[a] Dmitri Talapin,[a,b,c] and Andrew L. Ferguson*[,a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Polarizable nanoparticles are of interest in materials science because of their rich and complex phase behavior that can be used to engineer nanostructured materials with long-range crystalline order. To understand and rationally navigate the design space of polarizable nanoparticles for self-assembling highly ordered superlattices, we developed a coarse-grained computational model to describe the nanoparticle-nanoparticle interactions in implicit solvent and employ the computationally efficient image method to model many-body polarization interactions. We conducted high-throughput virtual screening over a five-dimensional particle design space spanned by temperature, particle size, particle charge, particle dielectric, and solvent dielectric using enhanced sampling molecular dynamics calculations within an active learning framework to efficiently map out the regions of thermodynamic stability of the self-assembled aggregates. We validate our predictions in comparisons against small angle x-ray scattering measurements of gold nanoparticles surface functionalized with metal chalcogenide ligands. Finally, we use our validated phase maps to computationally design switchable nanostructured materials capable of triggered assembly and disassembly as a function of temperature and solvent dielectric with potential applications as sensors, smart windows, optoelectronic devices, and in medical diagnostics.

Design, System, Application

Switchable materials that can change their properties in response to external stimuli are a desirable part of a "smart" materials toolbox. In this work, we combined coarse-grained models of polarizable nanoparticles explicitly accounting for many-body polarization interactions, enhanced sampling molecular dynamics calculations, and active learning to efficiently predict the phase behavior of self-assembling polarizable nanoparticles. We screen a five-dimensional particle design space spanned by temperature, particle size, particle charge, particle dielectric, and solvent dielectric to discern regions where assembly is thermodynamically favorable and unfavorable. We use this phase map to identify regions of design space where temperature and/or solvent dielectric can be used to drive the system back and forth over the phase boundary to achieve switchable assembly and disassembly of self-assembled nanostructured materials. Switchable materials that can change their appearance and properties as a function of external stimuli have diverse potential applications such as sensors, smart windows, optoelectronic devices, and in medical diagnostics. The design strategy is extensible to other self-assembling systems where accurate but efficient treatment of the many-body polarization interactions is critical to accurate prediction of phase behaviors.

1 Introduction

Self-assembly refers to the process of spontaneous organization of building blocks – typically molecules, nanoparticles, or colloids – for the bottom-up fabrication of materials with engineered properties that cannot easily be produced by top-down processing.[1–10] The structural and functional properties of a self-assembled material are directly linked to the individual building blocks, their supramolecular spatial arrangement, and their coupled interactions.[2,11] Building blocks including polymers, biomolecules, and nanoparticles (NPs) have been engineered to realize materials and devices with applications including light-emitting devices, self-healing materials, and sensors.[3,12–14] In many cases, the building block design process has been substantially accelerated and guided by computational studies that exposed design prin-

[a] Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 60637.; E-mail: andrewferguson@uchicago.edu
[b] Department of Chemistry, University of Chicago, Chicago, Illinois 60637.
[c] James Franck Institute, University of Chicago, Chicago, Illinois 60637.

† Electronic Supplementary Information (ESI) available: [Comparison between structure factors of metal chalcogenide complex functionalized gold nanoparticles in N-methylformamide from experiment and simulation, and computational design of temperature triggerable superlattices of particles with dielectric contrasts $\frac{\varepsilon_{in}}{\varepsilon_{out}} = 0.5$ and 1]. See DOI: 00.0000/00000000.

ciples by which to control the governing forces (e.g., hydrogen bonding, van der Waals forces, electrostatic interactions) between the building blocks and the environmental conditions (e.g., pH, temperature, salt concentration) in order to achieve a target assembled structure.[1,4,15,16]

Inorganic and polarizable NPs exhibit complex and rich phase behavior and have been used as building blocks for the self-assembly of materials with several promising applications such as solar cells, components in electronic circuitry, and imaging tools.[17–22] The colloidal stability of these NPs can be extremely high because of the strong electrostatic repulsions resulting from the surface charges. That the sample of gold NPs studied by Michael Faraday in 1850s remains stable to this day is a remarkable exemplar of the high colloidal stability of metallic NPs.[23] However, the situation can be changed dramatically by making even subtle changes to the environment in which the NPs are distributed. For example, the addition of a salt can lead to effectively instantaneous coalescence of the particles, typically into various classes of amorphous aggregates. As such, polarizable NPs can also serve as useful model systems in which to develop understanding and control of the many-body forces governing assembly and phase behavior.

A fascinating feature of inorganic and polarizable NPs that has important practical implications for assembly is the observation of anomalously high surface charge density for high dielectric constant particles stabilized by multivalent ionic ligands. This behavior is consistent with theoretical predictions that placing a high dielectric constant nanoparticle in a multivalent ionic medium results in significant polarization effects that shift the distribution of the cations or anions towards the surface of the sphere, thereby increasing the surface charge.[24] As a result of this additional surface charging, the interparticle potential remains repulsive up to very small separations before van der Waals forces become dominant and mediate short-ranged attraction. These large variations in the potential energy landscape switching from repulsive to attractive over nanometer length scales can have a dramatic impact on the assembly of the particles when the solution is destabilized. Unlike typical metal sols that generally form open disordered aggregates upon aggregation, in the case of metal particles coated with metal chalcogenide complex (MCC) ligands it is possible to form structures with long-range crystalline order.[25] The structural details of the self-assembled aggregates depend sensitively on a delicate interplay between the dispersion and the strength of the electrostatic interactions governed by the design parameters such as temperature, size, charge, polarizability, and the dielectric strength of the solvent. A mechanistic and predictive understanding of the role of the physicochemical features and properties of the NP building blocks and solvent upon the emergent self-assembled aggregates is a pre-requisite to the rational design and engineering of self-assembled NP aggregates with desired structure and function.

The primary objective of this work is to establish quantitative predictive models of the phase behavior of ligand functionalized NPs in order to rationally navigate the design space and engineer the triggerable assembly/disassembly of highly ordered superlattices by engineering the properties and conditions of the NPs and solvent. We employ high-throughput virtual screening (HTVS) using implicit-solvent coarse-grained computational models to predict the attractive and repulsive regimes within a five-dimensional particle design space spanned by temperature, particle size, particle charge, particle dielectric, and solvent dielectric, and use these data to construct self-assembly phase diagrams. Importantly, our coarse-grained model explicitly treats the many-body polarization interactions between NPs that is vital in accurate prediction of the interparticle interactions, but does so using a computationally efficient implementation of the image method[26,27] that enables our high-throughput screen. Additionally, we minimize the number of required calculations using an active learning protocol that builds an on-the-fly model of the phase diagram and simultaneously guides the next round of calculations towards those state points predicted to be most informative in accurately resolving the phase boundaries[28]. We validate our computational phase diagrams and predicted self-assembled structures against experimental observations and scattering measurements for gold NPs surface functionalized with MCC $K_4Sn_2S_6$ ligands. Finally, we use our validated phase maps to design switchable materials composed of polarizable NPs that are capable of triggered assembly and disassembly by modulating temperature or solvent. Example applications of such triggerable materials encompass diverse fields such as sensors[29,30], smart windows[31], optoelectronic devices[3] and in drug delivery[32]. The design strategy presented in this work is extensible to other self-assembling systems of inorganic building blocks where accurate but efficient treatment of the many-body polarization interactions are critical to accurate prediction of particle interactions and phase behaviors.

2 Methods

2.1 Computational methods

2.1.1 Coarse-grained model of nanoparticles

Our building blocks comprise homogeneous metallic NPs functionalized with $K_4Sn_2S_6$ ligands within a particular solvent.[18] To model the system, we represent the nanoparticle functionalized with the ligand as a spherical and polarizable particle with a total diameter $\sigma_{NP} = (\Delta + \sigma)$ – where σ is the length of the ligand and Δ is the diameter of the nanoparticle – a mass m_{NP}, a dispersion strength ε_{NP}, a charge q, and a dielectric constant ε_{in}. Solvent is modeled implicitly via its dielectric constant ε_{out}. The van der Waals interactions are described by the 12-6 Lennard-Jones potential[33]. The electrostatic interactions between a pair of identical spherical polarizable particles comprises a contribution from the bare Coulomb potential and that from induced surface charges. It can be shown that the total energy can be expressed as an infinite series $E = E_2 + E_3(g)\varepsilon + E_4(g)\varepsilon^2 + \ldots$, where E_n is a n-body interaction term, $g = (1 + \frac{\varepsilon_{in}}{\varepsilon_{out}})^{-1}$ and $\varepsilon = \frac{\varepsilon_{out} - \varepsilon_{in}}{\varepsilon_{out} + \varepsilon_{in}}$ [26,27]. E_2 is the familiar Coulomb interaction and is the only relevant term in the absence of any polarization effects where $\varepsilon_{in} = \varepsilon_{out}$ and $\varepsilon = 0$. The higher-order terms represent polarization contributions that depend on successively higher numbers of interacting particles and which decay geometrically in ε. The geometric combinatorics are such that the higher order terms become exponentially more expensive to compute. In the present work, we consider

only the leading order polarization contribution (i.e., the 3-body term) in order to make our calculations efficient and amenable to high-throughput virtual screening. This approximation becomes increasingly good as $|\varepsilon_{in}/\varepsilon_{out}| \to 1$: for $|\varepsilon_{in}/\varepsilon_{out}| > 0.8$ the 4-body term is approximately an order of magnitude smaller than the 3-body[27]. Furthermore, the polarization energy decreases by an order of $(\sigma_{NP}/2R)^3\varepsilon$ with each additional polarization of a surface, where R is the distance between the particles. This implies that higher order terms decay rapidly at small R and with increasing particle size.[26] We implement calculation of the 3-body term, described in more detail below, using the image method[26] as a fix command within LAMMPS named `colloid/image`[34]. This custom fix has been made freely available as part of the software development efforts of the Midwest Integrated Center for Computational Materials (MICCoM) center at `https://github.com/sivadasetty/image_method_MICCoM`. Mathematically, the total energy due to NP interactions is given by,[26,27]

$$U_{NP-NP} = U_{LJ} + U_{elec}, \tag{1}$$

where,

$$U_{LJ} = \frac{1}{2}\sum_{i,j} 4\varepsilon_{NP}\Big[\Big(\frac{\sigma}{R_{ij}-\Delta}\Big)^{12} - \Big(\frac{\sigma}{R_{ij}-\Delta}\Big)^{6}\Big], \tag{2}$$

$$U_{elec} = U_{Coul} + U_{pol-3}, \tag{3}$$

$$U_{Coul} = \frac{1}{2}\sum_{i,j}\frac{1}{4\pi\varepsilon_{out}R_{ij}}q_iq_j, \tag{4}$$

$$U_{pol-3} = \sum_{i,k,j}\frac{1}{8\pi\varepsilon_{out}}q_iq_j\mathscr{E}_{i,k,j}, \tag{5}$$

and,

$$\mathscr{E}_{i,k,j} = \varepsilon\frac{t_k}{\sqrt{R_{ik}R_{kj}}}\Big(\frac{1}{\mathscr{C}_{i,k,j}} - gt_k^{-2g}\int_0^{t_k^2}\frac{f^{g-1}}{\mathscr{D}_{i,k,j}}\mathrm{d}f\Big), \tag{6}$$

$$\varepsilon = \frac{\varepsilon_{out}-\varepsilon_{in}}{\varepsilon_{out}+\varepsilon_{in}},\ g = \frac{\varepsilon_{out}}{\varepsilon_{out}+\varepsilon_{in}},\ t_k = \frac{\sigma_{NP}}{2\sqrt{R_{ik}R_{kj}}}, \tag{7}$$

$$\mathscr{C}_{i,k,j} = \sqrt{1+t_k^4-2t_k^2\hat{\mathbf{R}}_{ik}.\hat{\mathbf{R}}_{kj}}, \tag{8}$$

$$\mathscr{D}_{i,k,j} = \sqrt{1+f^2-2f\hat{\mathbf{R}}_{ik}.\hat{\mathbf{R}}_{kj}}. \tag{9}$$

$R_{ij} = |\mathbf{R}_{ij}|$ is the scalar distance between i and j, $\mathbf{R}_{ik}$ and $\mathbf{R}_{kj}$ are the vectors pointing from scattering particle k to particle i and particle j to particle k, respectively, and $\hat{\mathbf{R}}_{ik}$ and $\hat{\mathbf{R}}_{kj}$ are the corresponding unit vectors. U_{LJ} refers to the portion of the total energy attributable to dispersion interactions. $U_{elec} = U_{Coul} + U_{pol-3}$ represents the total electrostatic energy decomposed into the Coulombic and 3-body polarization contributions.[26,27,35]

For numerical convenience, we conduct our simulations in reduced units. The NP-NP interactions described in Eq. 1-9 recast into reduced units are provided in Eq. 10-18. Since all particles are identical, we also make the additional specialization and simplification that q_i=q_j=q. We distinguish variables in reduced units from their real unit counterparts by an asterisk,

$$U^*_{NP-NP} = U^*_{LJ} + U^*_{elec}, \tag{10}$$

where,

$$U^*_{LJ} = \frac{1}{2}\sum_{i,j} 4\Big[\Big(\frac{1}{R^*_{ij}-\Delta^*}\Big)^{12} - \Big(\frac{1}{R^*_{ij}-\Delta^*}\Big)^{6}\Big], \tag{11}$$

$$U^*_{elec} = U^*_{Coul} + U^*_{pol-3}, \tag{12}$$

$$U^*_{Coul} = \frac{1}{2}\sum_{i,j}\frac{q^*q^*}{\varepsilon_{out}R^*_{ij}}, \tag{13}$$

$$U^*_{pol-3} = \sum_{i,k,j}\frac{1}{2\varepsilon_{out}}q^*q^*\mathscr{E}^*_{i,k,j}, \tag{14}$$

and,

$$\mathscr{E}^*_{i,k,j} = \varepsilon\frac{t^*_k}{\sqrt{R^*_{ik}R^*_{kj}}}\Big(\frac{1}{\mathscr{C}^*_{i,k,j}} - gt_k^{*-2g}\int_0^{t_k^{*2}}\frac{f^{*g-1}}{\mathscr{D}^*_{i,k,j}}df^*\Big), \tag{15}$$

$$\varepsilon = \frac{\varepsilon_{out}-\varepsilon_{in}}{\varepsilon_{out}+\varepsilon_{in}},\ g = \frac{\varepsilon_{out}}{\varepsilon_{out}+\varepsilon_{in}},\ t^*_k = \frac{\sigma^*_{NP}}{2\sqrt{R^*_{ik}R^*_{kj}}}, \tag{16}$$

$$\mathscr{C}^*_{i,k,j} = \sqrt{1+t_k^{*4}-2t_k^{*2}\hat{\mathbf{R}}^*_{ik}.\hat{\mathbf{R}}^*_{kj}}, \tag{17}$$

$$\mathscr{D}^*_{i,k,j} = \sqrt{1+f^{*2}-2f^*\hat{\mathbf{R}}^*_{ik}.\hat{\mathbf{R}}^*_{kj}}. \tag{18}$$

The mapping between real and reduced units is made by adopting a characteristic length scale of σ, a characteristic energy scale of ε_{NP}, and a characteristic mass of m_{NP},

$$U^*_{LJ} = \frac{U_{LJ}}{\varepsilon_{NP}},\ U^*_{elec} = \frac{U_{elec}}{\varepsilon_{NP}},$$

$$R^* = \frac{R}{\sigma},\ \sigma^*_{NP} = \frac{\sigma_{NP}}{\sigma},\ \Delta^* = \frac{\Delta}{\sigma},\ q^* = \frac{q}{\sqrt{4\pi\varepsilon_0\sigma\varepsilon_{NP}}},$$

$$T^* = \frac{k_BT}{\varepsilon_{NP}},\ \tau^* = \tau\sqrt{\frac{\varepsilon_{NP}}{m_{NP}\sigma^2}},\ M^* = \frac{M}{m_{NP}}, \tag{19}$$

where ε_0 is the vacuum permittivity, k_B is Boltzmann's constant, T is temperature, τ is time, and M is mass.

We define our five-dimensional nanoparticle design variables within the reduced unit mapping as temperature T^*, nanoparticle diameter σ^*_{NP}, nanoparticle charge q^*, and dielectric constant of the nanoparticle ε_{in} and solvent ε_{out}. We note that in reduced units, the relationship $\sigma_{NP} = (\Delta + \sigma)$ becomes $\sigma^*_{NP} = (\Delta^* + 1)$, so the reduced core nanoparticle diameter Δ^* is defined by the total diameter of the particle and the length of the ligand. Furthermore, we observe that ε_{in} and ε_{out} are already dimensionless quantities so are identical quantities in both the real and reduced unit sets. This defines the five-dimensional design space T^*-σ^*_{NP}-q^*-ε_{in}-ε_{out} within which we seek to quantify, predict, and ultimately engineer NP phase behavior.

2.1.2 Langevin dynamics simulations

Numerical simulations of NP behavior under the reduced interaction potential detailed in Eq. 10-18 were performed us-

ing Langevin dynamics implemented in LAMMPS – 3 March 2020[34,36] patched with the `colloid/image` fix[26]. Equations of motion were integrated using velocity Verlet integrator[37] with a time step of $0.005\tau^*$ and a Langevin thermostat with a damping time constant of $0.5\tau^*$. Simulations were conducted in an infinite simulation cell[38] with real-space cutoffs for dispersion and electrostatic interactions of $(2.5+\Delta^*)$ and $4.25\sigma^*_{NP}$, respectively. To control the concentration of nanoparticles and prevent the particles from drifting apart indefinitely, a spherical harmonic potential with force constant of $K^* = 10$ was used to restrain the particles when they approached within a distance $(2.5+\sigma^*_{NP}/2)$ of a fictitious spherical wall placed at a distance $2.125\sigma^*_{NP}$ from the origin of the simulation box. Dispersion and Coulombic interactions were respectively computed using the `lj/expand` and `coul/cut` LAMMPS pair types. The three-body polarization energy was computed using image method implemented as a custom LAMMPS `colloid/image` fix[26]. Execution times scale approximately cubically with system size[27], such that simulations of a particle pair execute at $\sim 1\times10^8\tau^*$/day and those of 10 particles at $\sim 5\times10^5\tau^*$/day on 20 × Intel® Xeon® Gold 6248 2.5GHz CPU cores using 20 OpenMP threads and 1 MPI task. Simulation snapshots were saved for analysis every 200 time steps, corresponding to a period of $1.0\tau^*$. Visualization of particle trajectories were conducted using OVITO[39] and VMD[40]. Numerical analyses of the trajectories were performed in Python using MDTraj[41].

2.1.3 Enhanced sampling of particle dimerization free energies

To determine the favorable regimes of NP aggregation within the five-dimensional T^*-σ^*_{NP}-q^*-ε_{in}-ε_{out} design space, we seek to identify the four-dimensional boundaries that separate regions where association of the particle pair is thermodynamically spontaneous from those where it is thermodynamically disfavored. We colloquially refer to this map defining the regimes of favorable and unfavorable interaction of the dimer as the 2-particle "phase diagram". Although our image method implementation of the 3-body polarization term is numerically efficient, the curse of dimensionality is such that simulations of multi-particle systems are too slow to support the thousands of calculations necessary to densely sample and accurately resolve the phase boundaries within the five-dimensional design space. For example, a very coarse grid of seven points per dimension would require 7^5=16,807 simulations, which, for simulations of N=100 particles, would entail a computational burden of more than 10,000 CPU-years. Accordingly, we adopt a two-step protocol wherein we first accurately determine the four-dimensional planes demarcating the boundaries for thermodynamically spontaneous formation of a dimer, and then use these results to guide us in performing a small number of expensive multi-particle simulations. Put simply, we take thermodynamically favorable pairwise interactions to be a prerequisite and reasonable proxy for spontaneous multi-particle self-assembly. As we will show, the 2-particle phase diagram does prove to be a good approximation for the many-body phase behavior as demonstrated through corroborating many-particle simulations and experimental measurements. Furthermore, accurate and comprehensive resolution of the 2-particle phase diagram is also very useful in exposing the interplay of the five design variables and understanding how these dictate the thermodynamically favorable and unfavorable regimes for spontaneous dimer formation.

We determine the spontaneity of dimer formation at a particular T^*-σ^*_{NP}-q^*-ε_{in}-ε_{out} point in the design space by estimating the 2-particle potential of mean force (PMF) along the distance R between the NPs and quantifying the change in Helmholtz free energy ΔA^* upon bringing the particles into contact. We efficiently estimate the PMF using the umbrella sampling[42] approach implemented in SSAGES[43,44] using LAMMPS[34] as the simulation engine. A variety of enhanced sampling techniques could be employed to estimate the PMF, including free energy perturbation (FEP)[45], metadynamics[46], temperature accelerated molecular dynamics (TAMD)[47], and adaptive biasing force (ABF)[48]. In this work we choose to employ umbrella sampling for its simplicity and embarrassing parallelism that is well-suited to our parallel compute resources and our active learning framework to be described below. We achieved converged PMF profiles by sampling 80 umbrella windows over the range $R^* = \sigma^*_{NP}-(\sigma^*_{NP}+100)$ applying harmonic spring constants in the range $K^* = 0.5$–100 and running up to 1×10^9 simulation time steps. In all cases we ensured that we reached the effectively non-interacting regime at sufficiently large R^* that defines the arbitrary zero of free energy, and deposited additional umbrella windows as necessary to achieve good overlaps between neighboring umbrella histograms. The unbiased PMF was estimated from the biased umbrella sampling data using the weighted histogram analysis method (WHAM)[49] implemented by Grossfield[50] and applying the entropic correction[51–53]. Uncertainties in the PMF profiles were estimated using block averaging employing blocks of 1×10^8 simulation time steps[54]. We extracted $\Delta A^* = (A^*_\infty - A^*_{\text{contact}})$ from the converged PMF as the free energy difference between the A^*_∞-large R^* plateau corresponding to the non-interacting state and the free energy A^*_{contact} in the contact state defined as $R^* \leq (\Delta^*+2.5)$. The free energy $A^*_{\text{contact}} = -\ln\int_0^{(\Delta^*+2.5)} e^{-A^*}\,\mathrm{d}R^*$ was determined by integrating the probability distribution over the contact region using Simpson's rule implemented in scipy[55]. Similarly, the free energy A^*_∞ was determined by integrating the probability distribution over the non-interacting plateau at a sufficiently large R^*_∞ that the free energy profile is flat by considering the same volume as that in contact region, $A^*_\infty = -\ln\int_{R^*_\infty}^{(R^*_\infty+\Delta^*+2.5)} e^{-A^*}\,\mathrm{d}R^*$.

2.1.4 Active learning of the 2-particle phase diagram

To minimize the number of calculations required to accurately estimate the 2-particle phase diagram, we employed the active learning framework proposed by Dai and Glotzer[28] that we illustrate schematically in Fig. 1. This approach employs an iterative framework comprising three components: (i) enhanced sampling calculations to estimate ΔA^* at a number of state points $x = \{T^*,\ \sigma^*_{NP},\ q^*,\ \varepsilon_{in},\ \varepsilon_{out}\}$, (ii) construction of a Gaussian process regression (GPR) surrogate model over all simulation data collected to date to predict ΔA^* over the full design space, and (iii) interrogation of the GPR model using Bayesian optimization (BO) to identify the most informative state points x to sample in the next round of simulations to accurately define the $\Delta A^* = 0$

phase boundary.

We train the GPR surrogate model $\Delta\hat{A}^* = f(x)$ to predict ΔA^* as a function of location x within the five-dimensional design space[56]. The GPR is fully specified by the choice of covariance function, for which we choose the twice differentiable Matérn-5/2 kernel[28]. The Matérn-5/2 length scale parameter l is treated as hyperparameter that is optimized during training by maximizing the log marginal likelihood.[28,56]. We train GPR models using scikit-learn[57] and GPytorch[58]. The trained model effectively extrapolates and interpolates from the ΔA^* values computed from our enhanced sampling simulations to predict the locus of points x s.t. $\Delta A^* = 0$ defining the phase boundary. Importantly, the surrogate model associates both a prediction of $\Delta\hat{A}^*$ and an associated uncertainty $var(\Delta\hat{A}^*)$ to each point x in the design space.

The GPR surrogate model is passed to a BO protocol employing the acquisition function $\alpha(x) = \frac{\sqrt{var(\Delta\hat{A}^*)}}{|\Delta\hat{A}^*|+e}$ proposed by Dai and Glotzer[28], where $e = 0.05$ is a small positive constant added for numerical stability. The BO calculates $\alpha(x)$ for each point in the design space to provide a measure of the relative desirability of sampling each point x in the next round of simulations. This acquisition function is custom designed to highly rank points that have small $|\Delta\hat{A}^*|$ and are therefore expected to lie close to the phase boundary and large uncertainty $var(\Delta\hat{A}^*)$ and are therefore benefit from sampling to reduce the uncertainty of the location of the phase boundary. The most desirable next point to sample x' is defined as $x' = \underset{x}{\operatorname{argmax}}\,\alpha(x)$, which we compute using optimization routines in scipy[55]. We employ Kriging believers[59] to perform batched sampling of the $N = 20$ top ranked points for the next batch of enhanced sampling calculations.

We initialize the iterative loop by performing simulations at 480 randomly selected state points $x = \{T^*,\ \sigma^*_{NP},\ q^*,\ \varepsilon_{in},\ \varepsilon_{out}\}$ within the design space and limit exploration over the range $\{T^* = 0.05-4,\ \sigma^*_{NP} = 10-80,\ q^* = -150-0,\ \varepsilon_{in} = 0.8-800000,\ \varepsilon_{out} = 80-8000\}$ that was designed to span the range of physically realizable systems. We then proceeded to perform 115 rounds of GPR model building, BO selection, and enhanced sampling Langevin dynamics simulation each comprising 20 new state points for a total of 2780 simulations requiring approximately 148,500 CPU-h and 18 GPU-h of compute time. We terminated the process once the $\Delta A^* = 0$ isotherms had been resolved to an accuracy of $T^*\pm 0.09$ for different values of σ^*_{NP}, q^*, ε_{in} and ε_{out} in the design space domain.

A Jupyter notebook containing the process used to perform active learning sampling of the 2-particle phase diagram using umbrella sampling and WHAM is available at `https://github.com/sivadasetty/activeLearningPB2NP`.

2.1.5 Multi-particle simulated annealing simulations and structural characterization

We conducted simulated annealing multi-particle simulations to either draw comparisons with experimental measurements or, guided by the 2-particle phase diagrams, conduct simulations in regions of design space where we expected to be able to trigger assembly and disassembly by manipulating temperature and/or solvent dielectric constant. We performed Langevin dynamics simulations of 10 NPs initially randomly distributed on the surface of a sphere of diameter $4.25\sigma^*_{NP}$. We then applied a simulated annealing protocol[60] to predict the stable self-assembled structure of the system at $T^* = 0.05$. We first applied 50,000 steps of conjugate gradient energy minimization to eliminate any large forces in the system, followed by Langevin dynamics simulations within the following annealing protocol: (i) a 4,000,000 linear cooling ramp from $T^* = 1.0$ to 0.05, (ii) a 500,000 step hold at $T^* = 0.05$, (iii) a 4,000,000 linear heating ramp from $T^* = 0.05$ to 1.0, (iv) a second 4,000,000 linear cooling ramp from $T^* = 1.0$ to 0.05, (iv) a 500,000 step hold at $T^* = 0.05$, and (v) a final 200,000 step hold at $T^* = 0.01$. When comparing multi-particle assembly with experimental observations at 300 K, the temperature in the last step (v) was replaced with a temperature $T^* = \frac{k_B \times 300\ \mathrm{K}}{\varepsilon_{NP}}$ using the estimated ε_{NP} for each experimentally studied gold NP reported in Table 1. We performed three independent annealing calculations commencing from different random initializations for each state point x.

We performed structural characterizations for the terminal self-assembled aggregates produced at the end of the annealing protocol using in-house Python scripts to estimate computational structure factor $S(\rho)$ profiles for the purposes of comparison with experimental measurements. (To avoid confusion with our symbol for charge, we denote the scattering vector by ρ.) These were computed by taking a Fourier transform of the radial distribution function[61,62] and applying volume corrections according to Gereben and Petkov.[63]

2.2 Experimental methods

2.2.1 Synthesis

A series of 4-6 nm Au NPs were prepared according to a modified literature procedure.[64] Briefly, $HAuCl_4 \cdot 3H_2O$ (500 mg, 1.25 mmol) was dissolved in a solution of 50 mL oleylamine and 20 mL hexane. The solids were dissolved by stirring under a flow of N_2 for one hour at room temperature. Varying concentrations of a tert-butylamine-borane complex solution (0.10-0.35 M) in 5 mL oleylamine and 2 mL hexane were injected after full dissolution by sonication and filtration through a 0.22 μm PTFE filter. The NPs were isolated by precipitation with ethanol and redispersing in hexane. Larger Au NPs up to 12 nm in diameter were grown from smaller 4-5 nm seeds using previously described methods.[64]

2.2.2 Ligand exchange

Native oleylamine capped Au NPs were transferred into N-methylformamide (NMF) and hydrazine via a two-phase ligand exchange. For NPs smaller than 7 nm, the native oleylamine ligands were first replaced with dodecanethiol ligands by vortexing NPs in a 10% dodecanethiol solution in hexane for one minute. The NPs were washed three times with ethanol and redispersed in hexane. NPs larger than 7 nm were directly phase transferred into polar solvents with oleylamine ligands. In a nitrogen filled glovebox, ~100 mg of NPs in 6 mL of hexane were stirred with 4 mL of 25 mM $K_4Sn_2S_6$ ligand solution until complete phase transfer into either NMF or hydrazine. Nanocrystals were washed three times by precipitation with acetonitrile and stored as ~100

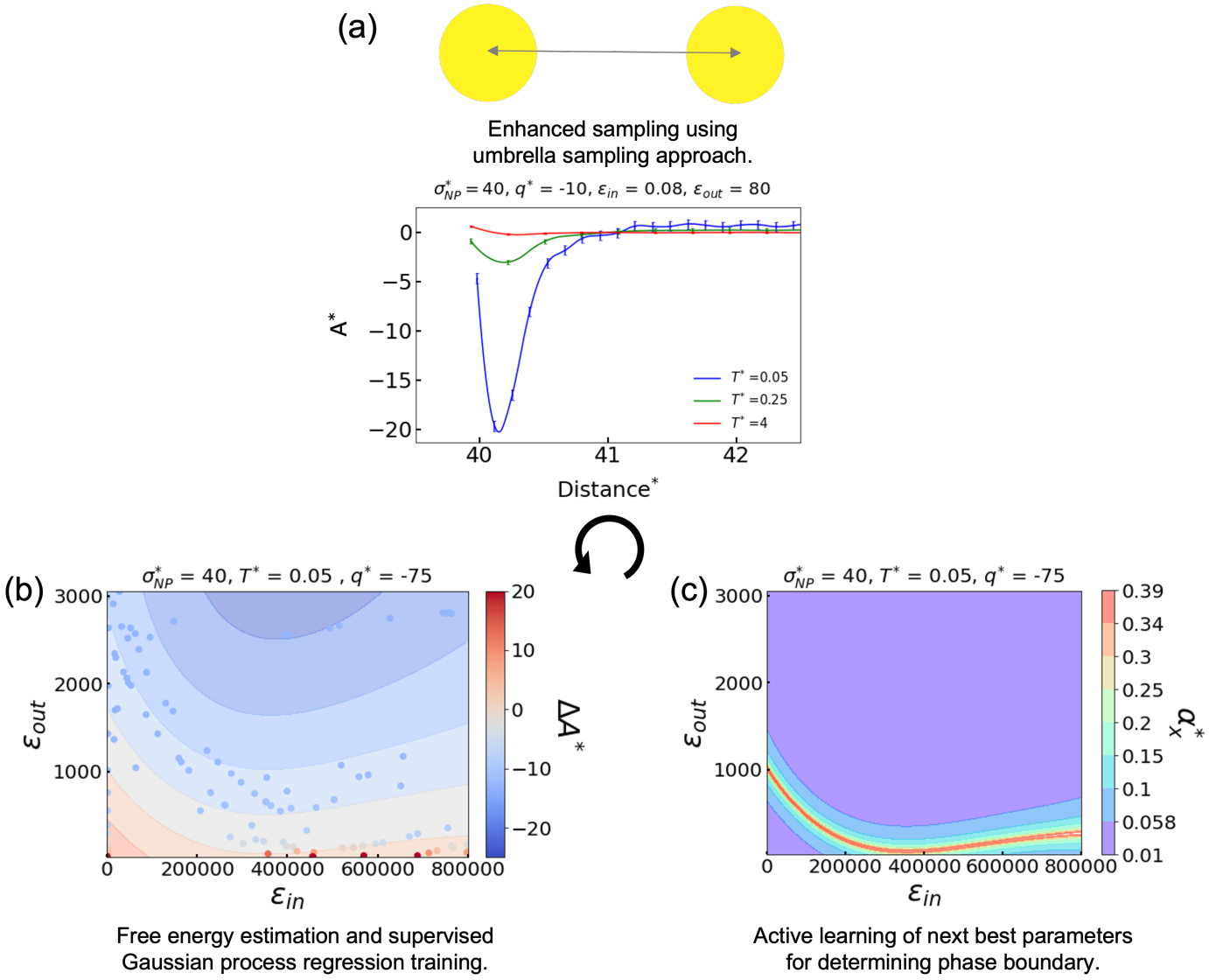

Fig. 1 Illustration of the enhanced sampling active learning framework used to efficiently map the 2-particle phase diagram within the five-dimensional design space. (a) Enhanced sampling Langevin dynamics simulations are used to estimate the Helmholtz free energy ΔA^* for 2-particle dimerization at a particular state point, here $x = \{T^* = 0.05, 0.25, 4, \sigma^*_{NP} = 40, q^* = -10, \varepsilon_{in} = 0.08, \varepsilon_{out} = 80\}$ within the five-dimensional design space. (b) A Gaussian process regression (GPR) model is used to construct a surrogate model for ΔA^* (shaded areas) over the full design space by training over calculated ΔA^* values (filled circles). Here we show a two-dimensional ε_{in}-ε_{out} slice through the five-dimensional design space at $T^* = 0.05$, $\sigma^*_{NP} = 40$, $q^* = -75$. (c) We interrogate the GPR model using Bayesian optimization (BO) to identify the most informative state points to sample next state point in our numerical simulations to accurately resolve the phase boundary at $\Delta A^* = 0$. Here we show the same two-dimensional ε_{in}-ε_{out} projection as in panel (b), but now colored according to the Bayesian optimization acquisition function α^*_x. State points x with a high value of the acquisition function are more likely to be selected for the next round of simulations.

mg/mL stock solutions in the glovebox.

2.2.3 Self-assembly and structural characterization

Au NPs in hydrazine were assembled by flocculation with excess 0.5 M $K_4Sn_2S_6$ ligand solution in hydrazine. Above a critical ligand concentration (~50-100 mM), the colloid fully precipitated to form a solid powder. In NMF solutions, the ligand concentration was first increased to 100-250 mM with a 0.2 M $K_4Sn_2S_6$ ligand solution in NMF, then a small amount of acetonitrile was added to fully flocculate the solution. Solid powders were isolated by centrifugation and washed with acetonitrile before being deposited on Kapton tape for analysis by small-angle x-ray scattering.

3 Results and Discussion

3.1 Analysis and design rules are exposed by the 2-particle phase diagrams

We present in Fig. 2 the converged 2-particle phase diagrams calculated using our active learning-guided enhanced sampling calculations and which demarcate the thermodynamically favorable and unfavorable regimes for NP dimerization. We present the phase boundaries in the five-dimensional T^*-σ^*_{NP}-q^*-ε_{in}-ε_{out} design space as a series of panels containing two-dimensional T^*-q^*

slices at particular choices of σ^*_{NP}. Within each panel, we show different colored curves for each choice of ε_{out} and different line styles for different choices of $\varepsilon_{in}/\varepsilon_{out}$. The lines correspond to the locus of points for which $\Delta A^* = (0\pm0.005)$ separating the attractive regime where dimerization is thermodynamically spontaneous below the lines from the repulsive regime where it is thermodynamically disfavored above. The phase diagrams calculated using our active learning-guided enhanced sampling calculations reveal a number of trends and design rules for the NPs in order to engineer the pairwise interactions to favor attractive self-assembly.

First, polarizability plays a small but important role on the phase boundaries. For example, the attractive regime expands towards lower q^* and higher T^* with increasing dielectric contrast ($\varepsilon_{in}/\varepsilon_{out}$) for fixed σ^*_{NP} and ε_{out}. This can be seen as the trend in the different line styles of a particular color in each of the figure panels. This effect arises from polarization interactions and correlates with trends in electrostatic potential between particles reported in Ref.[27]. The effect is more pronounced for large values of ε_{out} and larger particles, where neglecting polarization effects can result in large errors. For instance, for $K_4Sn_2S_6$ functionalized gold nanopoarticles with diameter 6 nm in water (ε_{out}=80) carrying a charge of ~(-6) e, ligands of length 0.25 nm, and approximating the effectively infinite dielectric constant of gold[65] as $\varepsilon_{in}=100\times\varepsilon_{out}$, neglecting the polarization contributions to the interaction potential could result in errors in the phase boundary predictions in excess of 85 K.

Second, increase in the dielectric constant of solvent ε_{out} results in a wider attractive regime in the T*-q* space for a fixed σ^*_{NP} and $\varepsilon_{in}/\varepsilon_{out}$. This can be seen as the trend in lines of different colors at a fixed line style in each of the figure panels. This trend is expected because an increase in ε_{out} increases the strength of screening, which results in a decrease in the net electrostatic repulsions.

Third, the attractive regime widens with increase in particle size. This can be seen as the expansion of the attractive area below the curves in progressing from panel a-f. This effect can be understood as primarily due to a decrease in the surface charge density with increasing in particle size, which results in a net reduction in the unfavorable electrostatic interactions at non-zero charge.

Fourth, when the particles are uncharged, the total electrostatic energy including the polarization energy is zero and the particles interact only by dispersion forces. Consequently, the temperature at the phase boundary for any ε_{out} and $\varepsilon_{in}/\varepsilon_{out}$ at q^*=0 should be identical. We confirm that within error bars we do observe convergence of T^* as $q^* \to 0$ for each value of σ^*_{NP}.

Fifth, the dimerization of the particles can be tuned using multiple design parameters simultaneously. One example is that the transition into the attractive regime can be triggered by simultaneously decreasing T and increasing ε_{out} for a given σ^*_{NP}, q^* and $\varepsilon_{in}/\varepsilon_{out}$. As another example, both q^* and ε_{out} can be simultaneously increased to move the particles from the repulsive into the attractive regime for a given σ^*_{NP}, T^*, and $\varepsilon_{in}/\varepsilon_{out}$.

3.2 Validation of the coarse-grained NP model against experimental measurements

The main goal of this study is to develop a predictive model of the phase behavior of gold NPs surface functionalized with $K_4Sn_2S_6$ ligands to rationally engineer triggerable assembly/disassembly of ordered superlattices. Before proceeding to this design step, we first validate the predictive capacity of our coarse-grained computational model and the calculated 2-particle phase diagrams to accurately predict experimental phase transitions and the structure of the self-assembled aggregates. To do so, we compare our predictions of the self-assembled structures at selected locations within the five-dimensional design space against experimental scattering measurements. Experiments are performed on $K_4Sn_2S_6$ functionalized gold NPs in hydrazine and n-methylformamide (NMF) solvent. Three different diameters of gold NPs are considered in each solvent: σ_{NP} = 4, 5 and 6 nm in hydrazine and σ_{NP} = 5.5, 8 and 12 nm in n-methylformamide. The dielectric constant of hydrazine and n-methylformamide are ε_{out} = 53 and 171, respectively. We mimic experimental conditions in our simulations by employing an estimated length of the $K_4Sn_2S_6$ ligands of σ = 0.25 nm, corresponding to approximately half of their ~0.58 nm fully extended length. We then define $\Delta = (\sigma_{NP} - \sigma)$ accordingly. We approximate the effectively infinite dielectric constant[65] of gold as $\varepsilon_{in} = 100\times\varepsilon_{out}$. Assuming that the dispersion interactions of the $K_4Sn_2S_6$ functionalized gold NPs is dominated by the gold core, we define ε_{NP} by matching the minimum of Lennard-Jones potential (Eq. 2) with the minimum of Hamaker potential[66,67] and adopting a Hamaker constant of 30×10^{-20} J consistent with experimental measurements for gold NPs.[68]. Table 1 shows the estimated dispersion strengths ε_{NP} for each σ_{NP}. It is challenging to estimate experimental surface charge because it is difficult to determine surface potential for opaque systems, but using the 2-particle phase boundaries as a guide we select a fixed value of the reduced charge q^*=(-10) that results in particle dimerization expected in the experimentally studied systems at 300 K and translate this into real units for each particle within Table 1.

Table 1 Estimated σ_{NP}, ε_{NP}, and q for each of the experimentally studied gold NPs.

σ_{NP} [nm]	$\varepsilon_{NP}\times10^{-19}$ [J]	q [e]
4	1.54	~(-4)
5	2.09	~(-5)
5.5	2.37	~(-5)
6	2.66	~(-5)
8	3.84	~(-6)
12	6.26	~(-8)

3.2.1 Self-assembly of $K_4Sn_2S_6$ functionalized gold nanoparticles in hydrazine

We present in Fig. 3a the calculated phase boundaries of the gold NPs with diameters 4, 5 and 6 nm in hydrazine solvent. With increase in the NP diameter, the attractive regime in the 2-particle phase boundaries in T-q space widens as observed in the previous section. We note that the extremely high temperatures on the 2-

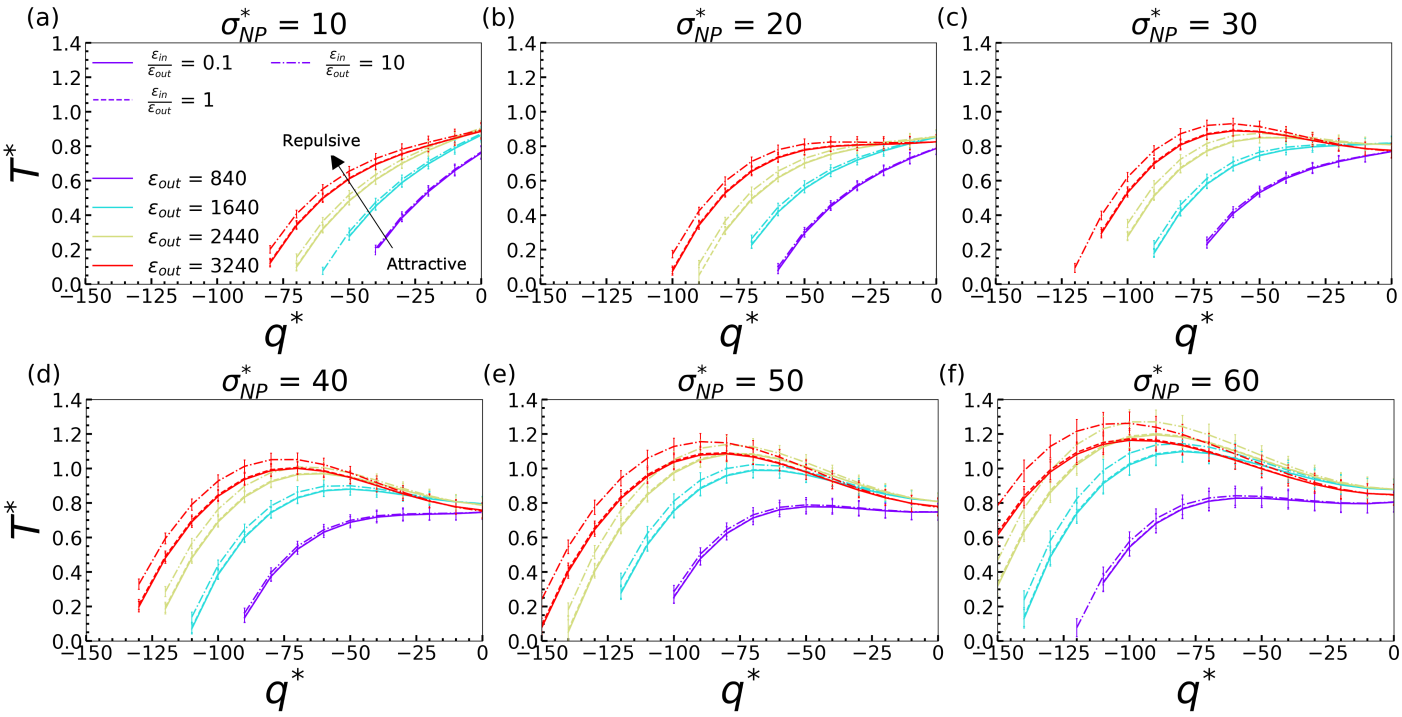

Fig. 2 Phase boundaries of 2-particles system along T*-q* for a σ^*_{NP} = (a) 10, (b) 20, (c) 30, (d) 40, (e) 50, and (f) 60. Within each panel different values of ε_{out} are represented by different line colors and different values of $\frac{\varepsilon_{in}}{\varepsilon_{out}}$ by different line styles. The lines correspond to the locus of points for which $\Delta A^* = 0$ separating the attractive and repulsive regimes that lie below and above each curve, respectively. Error bars on each phase boundary represent the standard deviation in ΔA^* estimated by the terminal fitted GPR model.

particle phase boundaries are clearly unphysical and represent only a hypothetical extrapolation under our model that does not account for NP melting, ligand decomposition, or solvent boiling.

We predict the equilibrium self-assembled structures at 300 K by conducting simulated annealing calculations[60] in a 10-particle system. As detailed in Section 2.1.5, we perform multiple heating to cooling cycles across the predicted 2-particle phase boundaries and then gradually anneal the system to the target experimental temperature of T = 300 K. The temperature range of the annealing calculations are shown as vertical arrows in Fig. 3a at the respective particle charges, and the target temperature 300 K defined by a horizontal dotted line. The changes in the 10-particle potential energy with temperature during the simulated annealing are illustrated in Fig. 3b. The transition in the 10-particle potential energy from disassembled configuration at the highest annealing temperature to attractive (i.e., negative potential energy) is correlated with the trends in the 2-particle phase boundaries. This can be observed from the temperature at which the potential energy drops during the cooling cycle of annealing simulation. For example, the potential energy becomes negative at $T \approx$ 9000 K for the larger 6 nm particle that is consistent with its wider attraction regime observed in the 2-particle phase diagram. Similarly, the smaller 4 nm system containing 10-particles transitions into assembled state at a lower temperature of $T \approx$ 5448 K, again consistent with the predictions of the 2-particle phase diagram. This indicates that our 2-particle phase boundaries serve as good predictors for the multi-particle phase behavior.

We present in Fig. 3c the computational structure factors $S(\rho)$ calculated from the self-assembled aggregates produced at the end of the simulated annealing calculations and compare these against the experimental structure factors measured by small-angle x-ray scattering (SAXS). A quantitative comparison is challenging due to the finite system size used in simulations compared to the much larger system size studied in experiments and the intrinsically coarse-grained nature of the model, but the qualitative trends are similar between computational predictions and experimental observations for all three NP diameters. First, the location of the primary (i.e., low-ρ) peak moves to lower ρ values with increasing particle diameter as expected by changing the size of the NP building block. Second, the higher-ρ peaks become less pronounced as NP diameter increases, which is consistent with the poorer packing and reduced ordering of the self-assembled aggregates observed in our simulations. Snapshots of the aggregates produced by the simulated annealing protocol are presented in Fig. 3d. All three particle diameters produce similar self-assembled aggregates occupying a icosahedral crystal lattice with bond orientational order parameter Q_6=~0.62-0.64[69,70]. Consistent with the better packing of the smaller NPs observed in the SAXS spectra, the calculated potential energy per unit area reduces (i.e., becomes more favorable) as the NP diameter decreases (Fig. 3b).

3.2.2 Self-assembly of $K_4Sn_2S_6$ functionalized gold nanoparticles in n-methylformamide (NMF)

In Fig. 4a we present the calculated phase boundaries of the gold NPs with diameters 5.5, 8 and 12 nm in NMF solvent. Again, we observe that the attractive regime in the T-q plane widens with

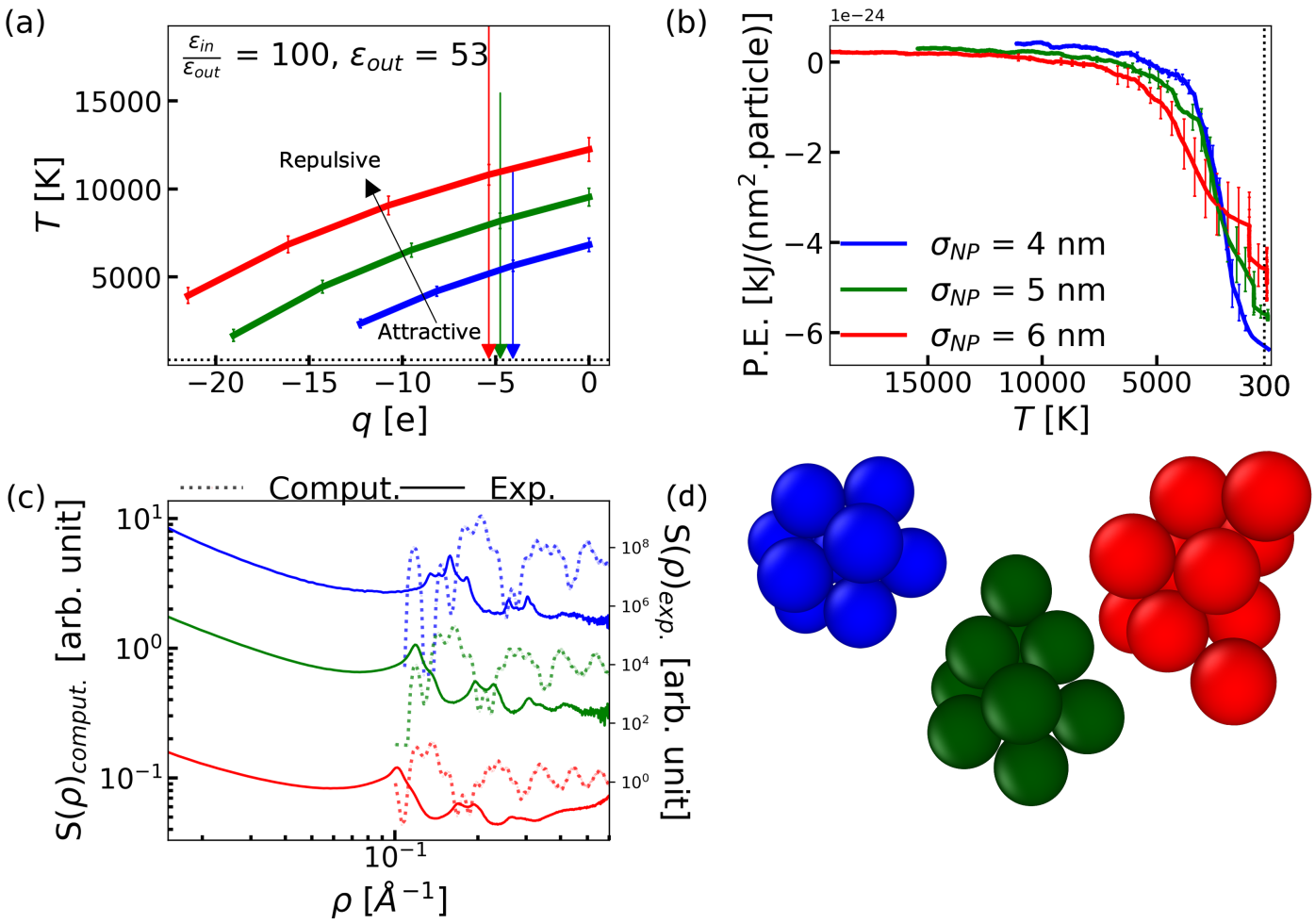

Fig. 3 Computational and experimental comparison of the self-assembly of $K_4Sn_2S_6$ functionalized gold nanoparticles in hydrazine. (a) Calculated phase diagram for 2-particle assembly projected into the T-q plane for NPs of diameter σ_{NP} 4 nm (blue), 5 nm (green), and 6 nm (red). The lines correspond to the locus of points for which $\Delta A^* = 0$ separating the attractive and repulsive regimes are below and above each curve, respectively. Error bars on each phase boundary represent the standard deviation in ΔA^* estimated by the terminal fitted GPR model. The vertical arrows indicate the simulated annealing temperature quenches used to determine the stable self-assembled structures of our 10-particle systems from an initial high temperature down to the 300 K target temperature indicated by the horizontal dashed line. The location of each vertical arrow on the x-axis corresponds to the q value assigned to each NP in our simulations. (b) Potential energies (P.E.) (U_{NP-NP} in Eq. 1 normalized by particle surface area $4\pi(\sigma_{NP}/2)^2$ and number of particles) observed during the terminal temperature quench of the 10-particle simulated annealing calculations. (c) Comparison of computational (dotted) and experimental (solid) structure factors $S(\rho)$ for the terminal self-assembled aggregates. For clarity of viewing, the 4 nm (blue) experimental data are vertically shifted in log-scale by +16.52 arb. unit and the 5 nm (green) data by +10.12 arb. unit. The computational data are then aligned with the experimental results. To avoid confusion with our symbol for charge, we denote the scattering vector by ρ. (d) Snapshots of the self-assembled structures produced at the end of the simulated annealing calculations.

increasing NP diameter and that the extremely high temperatures for the phase boundaries represent an unphysical extrapolation of our model. As in the case of hydrazine solvent, we conduct 10-particle simulated annealing calculations through the predicted phase boundaries and represent as vertical arrows the terminal cooling stage to the 300 K experimental target temperature. As was the case for hydrazine solvent, the temperatures at which we observe a downturn in the 10-particle potential energy curves displayed in Fig. 4b occur at the approximate temperatures predicted by the 2-particle phase diagrams (i.e., $T \approx$ 10,000 K, 14,000 K, 16,000 K for the 5.5 nm, 8 nm, and 12 nm NPs, respectively).

We present in Fig. 4c a comparison between the computational and experimental structure factors $S(\rho)$ for the calculated and observed self-assembled aggregates. Again, a quantitative comparison is challenging due to the finite system size used in simulations compared to the much larger system size studied in experiments and the intrinsically coarse-grained nature of the model. Nevertheless, a semi-quantitative comparison is possible, and, similar to the S(ρ) trends in hydrazine, we see good correspondence between our computational predictions and experimental observations in terms of the shift in the location of the primary (i.e., low-ρ) peak and attenuation in the magnitude of the higher-ρ peaks with increasing particle size. In NMF we see larger differences in trends in $S(\rho)$ as a function of NP diameter compared to hydrazine solvent seemingly due to the more pronounced reduction in packing quality with increased NP diameter (Fig. 4d).

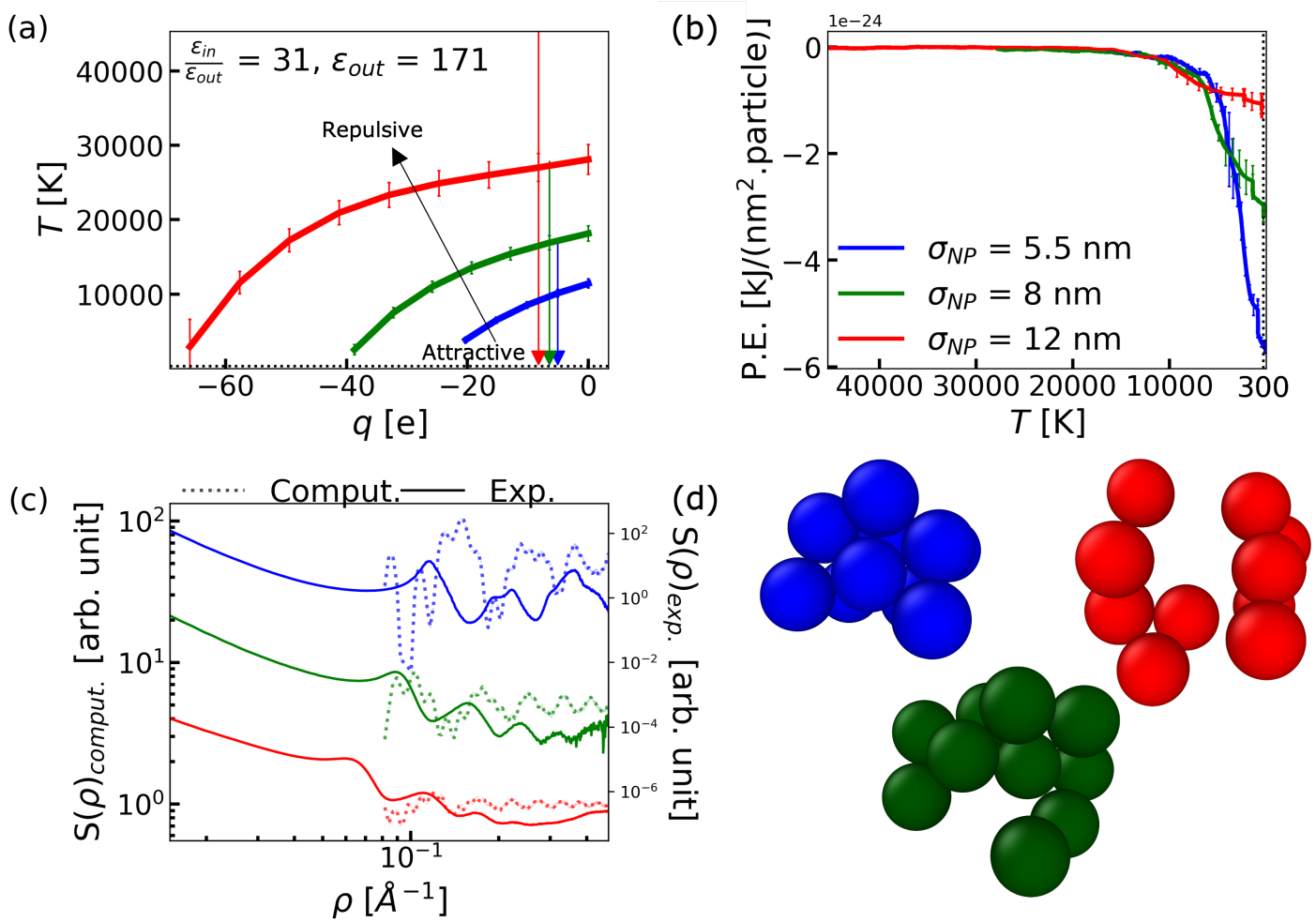

Fig. 4 Computational and experimental comparison of the self-assembly of $K_4Sn_2S_6$ functionalized gold nanoparticles in NMF. (a) Calculated phase diagram for 2-particle assembly projected into the T-q plane for NPs of diameter σ_{NP} 5.5 nm (blue), 8 nm (green), and 12 nm (red). The lines correspond to the locus of points for which $\Delta A^* = 0$ separating the attractive and repulsive regimes are below and above each curve, respectively. Error bars on each phase boundary represent the standard deviation in ΔA^* estimated by the terminal fitted GPR model. The vertical arrows indicate the simulated annealing temperature quenches used to determine the stable self-assembled structures of our 10-particle systems from an initial high temperature down to the 300 K target temperature indicated by the horizontal dashed line. The location of each vertical arrow on the x-axis corresponds to the q value assigned to each NP in our simulations. (b) Potential energies (P.E.) (U_{NP-NP} in Eq. 1 normalized by particle surface area $4\pi(\sigma_{NP}/2)^2$ and number of particles) observed during the terminal temperature quench of the 10-particle simulated annealing calculations. (c) Comparison of computational (dotted) and experimental (solid) structure factors $S(\rho)$ for the terminal self-assembled aggregates. For clarity of viewing, the 5.5 nm (blue) experimental data are vertically shifted in log-scale by (−18.42) arb. unit and the 8 nm (green) data by (−6.50) arb. unit. The computational data are then aligned with the experimental results. To avoid confusion with our symbol for charge, we denote the scattering vector by ρ. (d) Snapshots of the self-assembled structures produced at the end of the simulated annealing calculations.

The smallest 5.5 nm NP (blue) exhibits the most structural ordering, forming a compact, close packed icosahedral crystal lattice with bond orientational order parameter Q_6=∼0.63, and which exhibits the lowest potential energy per unit area of the three systems (Fig. 4b). The next largest 8 nm particle (green) forms a more loosely associated aggregate with intermediate potential energy per unit area, and the largest particle remains the most loosely cluster with the highest potential energy per unit area. The differences in the packing is also captured by the average number of nearest neighbors per particle, which are ∼5, ∼4, and ∼2 for 5.5 nm NP, 8 nm NP, and 12 nm NP, respectively. We confirm that the trends in the potential energy, $S(\rho)$, and structure of the aggregate morphologies are maintained upon doubling the size of the system to 20 NPs as illustrated in Fig. S1 in the ESI†.

3.3 Computational design of triggerable assembly/disassembly of self-assembled ordered superlattices

Given that the 2-particle phase diagrams have been validated as good predictors of multi-particle behaviors, we can use them to design triggerable systems that can be pushed across the phase boundaries by manipulating one or more of the five design parameters. In this manner we can engineer superlattices that can be self-assembled and disassembled in response to an external stimulus. Physically, by manipulating some combination of temperature (T), particle size (σ_{NP}), particle charge (q), particle dielectric (ε_{in}), and solvent dielectric (ε_{out}), we change the thermo-

dynamic state point of the system such that the equilibrium state of the system changes from an assembled state to a disassembled state, or *vice versa*. Example applications of such switchable self-assembled material components include sensors, smart materials and optoelectronic devices.[3,29–31]

3.3.1 Temperature triggerable superlattices

We first design ligand-functionalized NP systems that can be triggered to assemble/disassemble by modulating temperature. We present in Fig. 5a the T^*-q^* 2-particle phase diagram for σ^*_{NP}=10 and ε_{out}=80 for $\frac{\varepsilon_{in}}{\varepsilon_{out}}$ = 0.5, 1.0, and 100. Adopting a particle charge of q^*=(-10), our phase diagram predicts that we should be able to toggle between the attractive and repulsive regimes by modulating temperature to push the system back and forth across the phase boundary. This locus of points is indicated by the vertical dotted line. We consider three points on this line corresponding to temperatures of T^*=0.05 (blue star), 0.3 (green star), and 0.6 (red star). We performed simulated annealing simulations of 10 particle systems from T*=1 to each of these temperatures and computed the structure factors $S(\rho)$ of the self-assembled aggregates formed in each of these runs. We present the calculated $S(\rho)$ profiles in Fig. 5b and representative snapshots of the terminal aggregates in Fig. 5c for a dielectric contrast of $\frac{\varepsilon_{in}}{\varepsilon_{out}}$=100. As anticipated by the proximity of the calculated phase boundaries, similar results are observed for $\frac{\varepsilon_{in}}{\varepsilon_{out}}$ = 0.5 and 1.0, which are presented in Fig. S2 and S3 in the ESI†. As is apparent from the degree of structure in the $S(\rho)$ curves and in the degree of crystallinity visible in the representative snapshots, elevating temperature from T^*=0.05 (blue) to 0.3 (green) to 0.6 (red) corresponds to a marked loss in structure of the aggregate due to temperature-induced mitigation of the attractive interactions and entropic stabilization of the disaggregated state.

By inverting the reduced unit mapping defined by Eq. 19, we can identify a physically realizable system exhibiting this temperature-triggered assembly/disassembly behavior. Specifically, if we consider iron NPs with a Hamaker constant of 5.38×10^{-20} J[68] and functionalized with ligands with a length of σ=0.25 nm, this defines a length scale of σ=0.25 nm, energy scale of ε_{NP} = 1.36×10^{-20} J, sets the size of the particles to be σ_{NP} = 2.5 nm, and (neglecting the mass of the ligands relative to the metal NP core) defines a characteristic mass of m_{NP} = 6.44×10^{-23} kg. This implies a real units charge of $q\approx$(-6) e. We model the high dielectric constant of iron[65] as ε_{in} = $100\times\varepsilon_{out}$ and assume a water solvent ε_{out}=80, although the insensitivity of the phase boundary to the dielectric contrast $\frac{\varepsilon_{in}}{\varepsilon_{out}}$ means that our predictions are not sensitive to the specific values of these quantities. Finally, the temperatures T^*=0.05, 0.3, and 0.6 correspond to ∼49 K, ∼296 K, and ∼593 K, respectively. Our calculations imply that we may trigger assembly and disassembly of the ligand-functionalized iron NPs by modulating temperature between approximately the freezing (273 K) and boiling (373 K) points of water at atmospheric pressure.

3.3.2 Solvent triggerable superlattices

We now design ligand-functionalized NP systems that can be triggered to assemble/disassemble by modulating solvent dielectric. We present in Fig. 5d the ε_{out}-q^* 2-particle phase diagram for σ^*_{NP}=10, ε_{in}=8000 and T^*=0.05. Adopting a particle charge of q^*=(-50), our phase diagram predicts that we should be able to toggle between the attractive and repulsive regimes by modulating solvent dielectric to push the system back and forth across the phase boundary. This locus of points is indicated by the vertical dotted line. We consider three points on this line corresponding to solvent dielectric of ε_{out}=80 (blue star), 1000 (green star), and 2000 (red star). We performed simulated annealing simulations of 10 particle systems at each of these solvent dielectric constants using a high temperature of T^*=1 to a target temperature T^*=0.05. We then computed the structure factors $S(\rho)$ of the self-assembled aggregates formed at T^*=0.05 for each solvent dielectric. We present the calculated $S(\rho)$ profiles in Fig. 5e and representative snapshots of the terminal aggregates in Fig. 5f. As is apparent from the degree of structure in the $S(\rho)$ curves and in the degree of crystallinity visible in the representative snapshots, elevating solvent dielectric from ε_{out}=80 (blue) to 1000 (green) to 2000 (red) results in improved packing of the aggregate due to solvent-induced mitigation of the repulsive electrostatic interactions.

By inverting the reduced unit mapping defined by Eq. 19, we can identify a physically realizable system exhibiting this solvent dielectric-triggered assembly/disassembly behavior. Specifically, if we consider silver NPs with a Hamaker constant of 28×10^{-20}J[68] and functionalized with ligands with a length of σ=0.25 nm, this defines a length scale of σ=0.25 nm, energy scale of ε_{NP} = 7.10×10^{-19} J, sets the size of the particles to be σ_{NP} = 2.5 nm, and (neglecting the mass of the ligands relative to the metal NP core) defines a characteristic mass of m_{NP} = 8.58×10^{-23} kg. This implies a real units charge of $q\approx$(-14) e. We model the high dielectric constant of silver[65] as ε_{in} = 8000, and a temperature of T^*=0.05 corresponds to ∼257.3 K. Our calculations imply that we may trigger assembly and disassembly of the ligand-functionalized silver NPs by modulating solvent dielectric. We predict, for example, that in N,N-dimethylformamide (ε_{out}∼47[71]) the system should be disassembled, and by changing to a high dielectric solvent such as N-methylformamide (ε_{out}∼217[71]) or N-methylpropionamide (ε_{out}∼267[71,72]) we may approach or cross the multi-particle phase boundary and trigger assembly.

4 Conclusions

We have developed a coarse-grained computational model for polarizable NP-NP interactions to understand and rationally navigate the particle design space. Many-body polarization interactions between NPs are efficiently treated using the image method[26,27]. To understand and explore N-particle systems, we mapped the phase-diagram within the high-dimensional particle design space of a 2-particle system efficiently using an active learning framework. The trends in the phase boundaries quantitatively predict assembly behavior and reveal that the attraction regime widens with increase in particle size, dielectric constant of solvent, and dielectric contrast between NP and solvent. The structure and assembly behaviors predicted by our model and 2-particle phase diagrams are in good agreement with multi-

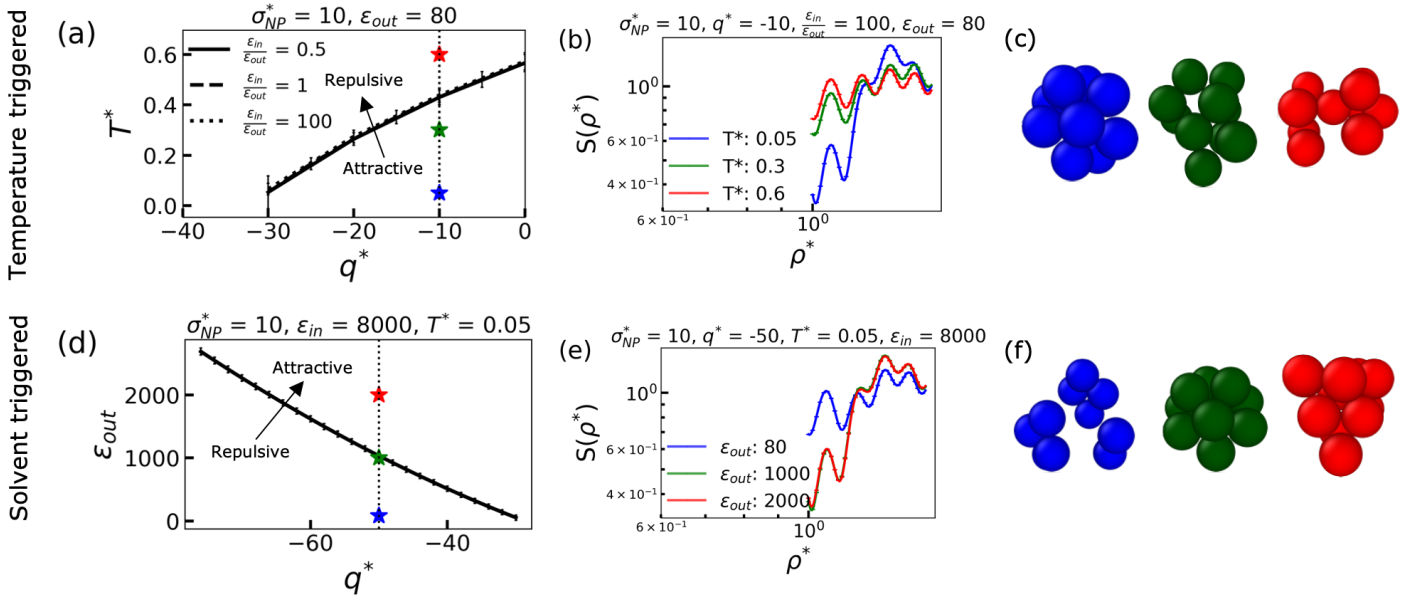

Fig. 5 Computational design of (a-c) temperature and (d-f) solvent-triggerable assembly/disassembly in self-assembled NP superlattices. (a) Calculated phase diagram for 2-particle assembly projected into the T^*-q^* plane for NPs of diameter σ^*_{NP}=10 in a solvent with a dielectric constant ε_{out} = 80 and dielectric contrasts $\frac{\varepsilon_{in}}{\varepsilon_{out}}$ = 0.5, 1, and 100. The lines correspond to the locus of points for which ΔA^* = 0 separating the attractive and repulsive regimes are below and above each curve, respectively. Different line styles correspond to different $\frac{\varepsilon_{in}}{\varepsilon_{out}}$. Error bars on each phase boundary represent the standard deviation in ΔA^* estimated by the terminal fitted GPR model. The location of the vertical dotted line on the x-axis corresponds to the q value assigned to the NP in our simulations. The stars on the vertical dotted line represent the temperature at which structure analysis was performed – 0.05 (blue), 0.3 (green), and 0.6 (red). (b) Computed structure factors $S(\rho)$ for the terminal self-assembled aggregates at each temperature highlighted in (a) spanning across the phase boundary for $\frac{\varepsilon_{in}}{\varepsilon_{out}}$ 100. (c) Snapshots of the self-assembled structures corresponding to computed $S(\rho)$ shown in (b) at each temperature. (d) Calculated phase diagram for 2-particle assembly projected into the ε_{out}-q^* plane for NPs of diameter σ^*_{NP} 10 and dielectric constant of NPs ε_{in} = 8000 at temperature T^* = 0.05 and solvents with dielectric constants ε_{out} = 8, 1000, and 2000. The location of the vertical dotted line on the x-axis corresponds to the q value assigned to the NP in our simulations. The stars on the vertical dotted line represent the different solvent dielectric at which structure analysis was performed – 8 (blue), 1000 (green), and 2000 (red). (e) Computed structure factors $S(\rho)$ for the terminal self-assembled aggregates at each solvent dielectric highlighted by stars in (d) spanning across the phase boundary. (f) Snapshots of the self-assembled structures corresponding to computed structure factors shown in (e) at each solvent dielectric.

particle calculations and selected experimental observations. We used the calculated phase diagrams to rationally design ligand-functionalized NPs that can be triggered to assemble/disassemble by modulating temperature or changing the solvent environment. These triggerable materials have potential applications in sensing[29,30], smart windows[31], optoelectronics[3] and drug delivery[32]. In future work, we aim to computationally design additional physically realizable triggerable systems, subject these designs to experimental testing, and extend our design protocol to other self-assembling systems of inorganic building blocks where many-body polarization interactions are important in governing particle interactions and phase behaviors.

Author Contributions

S.D.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. I.C.: Methodology, Investigation, Writing - Review & Editing. J.P.: Investigation, Writing - Review & Editing. J.L.: Software. J.J.dP.: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition. D.T.: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision, Funding acquisition. A.L.F.: Conceptualization, Methodology, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Conflicts of interest

A.L.F. is a co-founder and consultant of Evozyne, Inc. and a co-author of US Provisional Patents 62/853,919 and 62/900,420 and International Patent Applications PCT/US2020/035206 and PCT/US20/50466.

Acknowledgements

This work was supported by MICCoM (Midwest Center for Computational Materials), as part of the Computational Materials Science Program funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division. This work was completed in part with resources provided by the University of Chicago Research Computing Center. We gratefully acknowledge computing time on the University of Chicago high-performance GPU-based cyberinfrastructure supported by the National Science Foundation under Grant No. DMR-1828629.

Notes and references

1 M. Dijkstra and E. Luijten, *Nat. Mater.*, 2021, **20**, 762–773.

2 D. V. Talapin, M. Engel and P. V. Braun, *MRS Bull.*, 2020, **45**, 799–806.

3 M. A. Boles, M. Engel and D. V. Talapin, *Chem. Rev.*, 2016, **116**, 11220–11289.

4 Z. M. Sherman, M. P. Howard, B. A. Lindquist, R. B. Jadrich and T. M. Truskett, *J. Chem. Phys.*, 2020, **152**, 140902.

5 S. C. Glotzer and M. J. Solomon, *Nat. Mater.*, 2007, **6**, 557–562.

6 G. M. Whitesides and M. Boncheva, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 4769–4774.

7 S. C. Glotzer, *Science*, 2004, **306**, 419–420.

8 G. M. Whitesides and B. Grzybowski, *Science*, 2002, **295**, 2418–2421.

9 Z. Zhang and S. C. Glotzer, *Nano Lett.*, 2004, **4**, 1407–1413.

10 K. Ariga, J. P. Hill, M. V. Lee, A. Vinu, R. Charvet and S. Acharya, *Sci. Technol. Adv. Mater.*, 2008, **9**, 014109.

11 X. Li, X. Liu and X. Liu, *Chem. Soc. Rev.*, 2021, **50**, 2074–2101.

12 N. C. Seeman and H. F. Sleiman, *Nat. Rev. Mater.*, 2017, **3**, 1–23.

13 A. Levin, T. A. Hakala, L. Schnaider, G. J. Bernardes, E. Gazit and T. P. Knowles, *Nat. Rev. Chem.*, 2020, **4**, 615–634.

14 J.-F. Lutz, J.-M. Lehn, E. W. Meijer and K. Matyjaszewski, *Nat. Rev. Mater.*, 2016, **1**, 1–14.

15 Y. Ma and A. L. Ferguson, *Soft Matter*, 2019, **15**, 8808–8826.

16 K. Shmilovich, R. A. Mansbach, H. Sidky, O. E. Dunne, S. S. Panda, J. D. Tovar and A. L. Ferguson, *J. Phys. Chem. B*, 2020, **124**, 3873–3891.

17 H. Zhang and D. Wang, *Angew. Chem.*, 2008, **120**, 4048–4051.

18 M. V. Kovalenko, M. Scheele and D. V. Talapin, *Science*, 2009, **324**, 1417–1420.

19 A. Nag, M. V. Kovalenko, J.-S. Lee, W. Liu, B. Spokoyny and D. V. Talapin, *J. Am. Chem. Soc.*, 2011, **133**, 10612–10620.

20 M. V. Kovalenko, L. Manna, A. Cabot, Z. Hens, D. V. Talapin, C. R. Kagan, V. I. Klimov, A. L. Rogach, P. Reiss, D. J. Milliron, P. Guyot-Sionnnest, G. Konstantatos, W. J. Parak, T. Hyeon, B. A. Korgel, C. B. Murray and W. Heiss, *ACS Nano*, 2015, **9**, 1012–1057.

21 K. Barros and E. Luijten, *Phys. Rev. Lett.*, 2014, **113**, 017801.

22 A. Azari, J. J. Crassous, A. M. Mihut, E. Bialik, P. Schurtenberger, J. Stenhammar and P. Linse, *Langmuir*, 2017, **33**, 13834–13840.

23 M. Faraday, *Philos. Trans. R. Soc. London*, 1857, **147**, 145–181.

24 G. I. Guerrero García and M. Olvera de la Cruz, *J. Phys. Chem. B*, 2014, **118**, 8854–8862.

25 I. Coropceanu, E. M. Janke, J. Portner, D. Haubold, T. D. Nguyen, A. Das, C. P. Tanner, J. K. Utterback, S. W. Teitelbaum, M. Hudson, N. Sarma, A. M. Hinkle, C. J. Tassone, A. Eychmüller, D. T. Limmer, M. O. de la Cruz, N. S. Ginsberg and D. V. Talapin, *unpublished work*.

26 J. Qin, J. J. de Pablo and K. F. Freed, *J. Chem. Phys.*, 2016, **145**, 124903.

27 J. Qin, J. Li, V. Lee, H. Jaeger, J. J. de Pablo and K. F. Freed, *J. Colloid Interface Sci.*, 2016, **469**, 237–241.

28 C. Dai and S. C. Glotzer, *J. Phys. Chem. B*, 2020, **124**, 1275–1284.

29 L. La Spada and L. Vegni, *Materials*, 2018, **11**, 603.

30 V. Pardo-Yissar, R. Gabai, A. N. Shipway, T. Bourenko and I. Willner, *Adv. Mater.*, 2001, **13**, 1320–1323.

31 S. Lee, K. Sim, S. Y. Moon, J. Choi, Y. Jeon, J.-M. Nam and S.-J. Park, *Adv. Mater.*, 2021, 2007668.

32 S. Mura, J. Nicolas and P. Couvreur, *Nat. Mater.*, 2013, **12**, 991–1003.

33 S. Jungblut and A. Eychmüller, in *Chemical Modelling*, ed. M. Springborg and J.-O. Joswig, The Royal Society of Chemistry, UK, 2020, vol. 15 of SPR - Chemical Modelling, pp. 1–27.

34 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

35 J. Qin, *Soft Matter*, 2019, **15**, 2125–2134.

36 S. Plimpton, A. Kohlmeyer, A. Thompson, S. Moore and R. Berger, *LAMMPS Stable Release 3 March 2020*, 2020, `https://doi.org/10.5281/zenodo.3726417`, (accessed March 2020).

37 L. Verlet, *Phys. Rev.*, 1967, **159**, 98.

38 X. Jiang, J. Li, V. Lee, H. M. Jaeger, O. G. Heinonen and J. J. de Pablo, *J. Chem. Phys.*, 2018, **148**, 234302.

39 A. Stukowski, *Modell. Simul. Mater. Sci. Eng.*, 2009, **18**, 015012.

40 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

41 R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane and V. S. Pande, *Biophys. J.*, 2015, **109**, 1528 – 1532.

42 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

43 H. Sidky, Y. J. Colón, J. Helfferich, B. J. Sikora, C. Bezik, W. Chu, F. Giberti, A. Z. Guo, X. Jiang, J. Lequieu, J. Li, J. Moller, M. J. Quevillon, M. Rahimi, H. Ramezani-Dakhel, V. S. Rathee, D. R. Reid, E. Sevgen, V. Thapar, M. A. Webb, J. K. Whitmer and J. J. de Pablo, *J. Chem. Phys.*, 2018, **148**, 044104.

44 H. Sidky, Y. J. Colón, J. Helfferich, B. J. Sikora, C. Bezik, W. Chu, F. Giberti, A. Z. Guo, X. Jiang, J. Lequieu, J. Li, J. Moller, M. J. Quevillon, M. Rahimi, H. Ramezani-Dakhel, V. S. Rathee, D. R. Reid, E. Sevgen, V. Thapar, M. A. Webb, J. K. Whitmer and J. J. de Pablo, *SSAGESproject/SSAGES*, `https://github.com/SSAGESproject/SSAGES`, 2018, (accessed March 2020).

45 R. W. Zwanzig, *J. Chem. Phys*, 1954, **22**, 1420–1426.

46 A. Barducci, M. Bonomi and M. Parrinello, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 826–843.

47 L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.*, 2006, **426**, 168–175.

48 E. Darve, D. Rodríguez-Gómez and A. Pohorille, *J. Chem. Phys*, 2008, **128**, 144120.

49 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and R. J. M., *J. Comput. Chem.*, 1992, **13**, 1011–1021.

50 A. M. Grossfield, *WHAM: the weighted histogram analysis method version 2.0.10*, `http://membrane.urmc.rochester.edu/wordpress/?page_id=126`, (accessed March 2020).

51 S. Páll, M. J. Abraham, C. Kutzner, B. Hess and E. Lindahl, in *Solving Software Challenges for Exascale*, ed. S. Markidis and E. Laure, Springer International Publishing, Cham, 2014, vol. 8759, pp. 3–27.

52 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19–25.

53 R. M. Neumann, *Am. J. Phys.*, 1980, **48**, 354–357.

54 A. Grossfield, P. N. Patrone, D. R. Roe, A. J. Schultz, D. W. Siderius and D. M. Zuckerman, *Living J. Comp. Mol. Sci.*, 2018, **1**, 5067.

55 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, *Nat. Methods*, 2020, **17**, 261–272.

56 C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2005.

57 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.

58 J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger and A. G. Wilson, Advances in Neural Information Processing Systems, 2018, p. 7587âĂŞ7597.

59 B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. De Freitas, *Proc. IEEE*, 2015, **104**, 148–175.

60 P. J. van Laarhoven and E. H. Aarts, in *Simulated annealing*, Springer Netherlands, Dordrecht, 1st edn., 1987, pp. 7–15.

61 K. Zhang, *arXiv*, 2016, preprint, arXiv:1606.03610, `https://arxiv.org/abs/1606.03610`.

62 M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, UK, 2nd edn., 2017.

63 O. Gereben and V. Petkov, *J. Phys.: Condens. Matter*, 2013, **25**, 454211.

64 B.-H. Wu, H.-Y. Yang, H.-Q. Huang, G.-X. Chen and N.-F. Zheng, *Chin. Chem. Lett.*, 2013, **24**, 457–462.

65 G. Kofod, S. Risse, H. Stoyanov, D. N. McCarthy, S. Sokolov and R. Kraehnert, *ACS Nano*, 2011, **5**, 1623–1629.

66 H. C. Hamaker, *Physica*, 1937, **4**, 1058–1072.

67 R. Everaers and M. R. Ejtehadi, *Phys. Rev. E*, 2003, **67**, 041710.

68 A. R. Petosa, D. P. Jaisi, I. R. Quevedo, M. Elimelech and N. Tufenkji, *Environ. Sci. Technol.*, 2010, **44**, 6532–6549.

69 P. J. Steinhardt, D. R. Nelson and M. Ronchetti, *Phys. Rev. B*, 1983, **28**, 784–805.

70 J. S. van Duijneveldt and D. Frenkel, *J. Chem. Phys.*, 1992, **96**, 4655–4668.

71 S. J. Bass, W. I. Nathan, R. M. Meighan and R. H. Cole, *J. Phys. Chem.*, 1964, **68**, 509–515.

72 S. Takagi, *Jpn. J. Appl. Phys.*, 1986, **25**, 1612–1613.