THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Open Access
This article is licensed under CC-BY 4.0

pubs.acs.org/JPCA
Article

Machine-Learned Probability Landscapes Reveal Charge-Controlled Structural Motifs in Gold Nanoclusters

Gabriel Reynald Da Silva,* Mousa Salah, Milene M. Machado, Alexandre C. Dias, Carlos M. O. Bastos, Mauricio J. Piotrowski, Celso R. C. Rego,* and Diego Guedes-Sobrinho*

Cite This: https://doi.org/10.1021/acs.jpca.6c05562
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Configurational landscapes of metallic nanoclusters (NCs), particularly identifying thermodynamically relevant basins under the combined effects of charge state and temperature, remain a major challenge for catalytic applications. This work introduces an unsupervised machine-learning framework based on nonlinear dimensionality reduction, including sketch-map, t-SNE/FIt-SNE, and UMAP, to decode the high-dimensional configurational space and physicochemical properties of bare Au_{13}. By combining density functional theory-based replica-exchange Born−Oppenheimer molecular dynamics with Boltzmann reweighting via the multistate Bennett acceptance ratio, we constructed probability surfaces at 300 K from multiple-temperature canonical ensembles. From 2160000 sampled configurations, this framework projects complex atomistic trajectories onto low-dimensional manifolds that retain structural similarity, basin connectivity, and thermodynamic relevance. The resulting landscapes reveal that charge state and temperature cooperatively modulate the 2D/3D competition in Au_{13}; cationic and neutral clusters retain significant populations in fluxional 3D basins, whereas anionic clusters show pronounced migration toward 2D and *quasi*-2D regions. Local coordination, chirality, and the center of gravity of occupied *d*-states further connect configurational sampling to electronic reactivity and chiral-site formation. Implemented in the `DimRedPy` workflow, this strategy offers a transferable and reproducible route for decoding structure−property relationships in fluxional NCs under more realistic operating conditions.



1. INTRODUCTION

Metallic nanoclusters (NCs) exhibit unique properties because of their small size. Every atom in these clusters can participate in chemical reactions. When NCs interact with a surface, the support can significantly alter their electronic structure, charge distribution, and adsorption behavior.[1−4] This interaction creates new active sites at the interface that do not exist in the free cluster or on the bare surface. Among metallic NCs, Au NCs are especially attractive because they exhibit high activity and selectivity in a wide range of catalytic processes,[5−8] particularly as promising candidates for H_2 production[9,10] and photocatalysis across diverse chemical environments,[11−14] including ligand-protected and surface-supported systems.[15−20]

Due to their very small size, the evolution of possible morphologies is highly sensitive to the chemical environment and thermodynamic conditions, such as ligand-exchange dynamics,[21,22] temperature, and charge effects.[23,24] This pronounced structural sensitivity reflects the intrinsic fluxionality of subnanometer metal clusters wherein thermal excitation, environmental changes, or charge variations trigger dynamic structural interconversion across accessible potential energy basins rather than confining the system to a single, rigid ground-state geometry.[25−27] The total controlled design of new materials for specific purposes is highly sought for Au NCs (∼1.0 nm) and sub-NCs (<1.0 nm),[28] aiming for a better understanding of the morphology−property relationship. For catalytic selectivity and performance, a key challenge is identifying dominant configurations within the thermodynamic ensemble, particularly 2D and 3D configurations,[29] since higher catalytic activity has been reported in specific contexts for 2D configurations.[30] Moreover, chiral NCs are promising candidates for nanocatalysis, as even icosahedral Au_{13} NCs can exhibit chirality[31,32] arising from asymmetric ligand arrangements or intrinsic distortions of the metal kernel.[33] However, considering the impact of charge states in Au sub-NCs is crucial in *operando*, given their significant role during catalytic reactions,[34,35] in which, under temperature, multiple config-

Received: August 16, 2026
Revised: October 1, 2026
Accepted: October 5, 2026

ACS Publications
© XXXX The Authors. Published by American Chemical Society

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

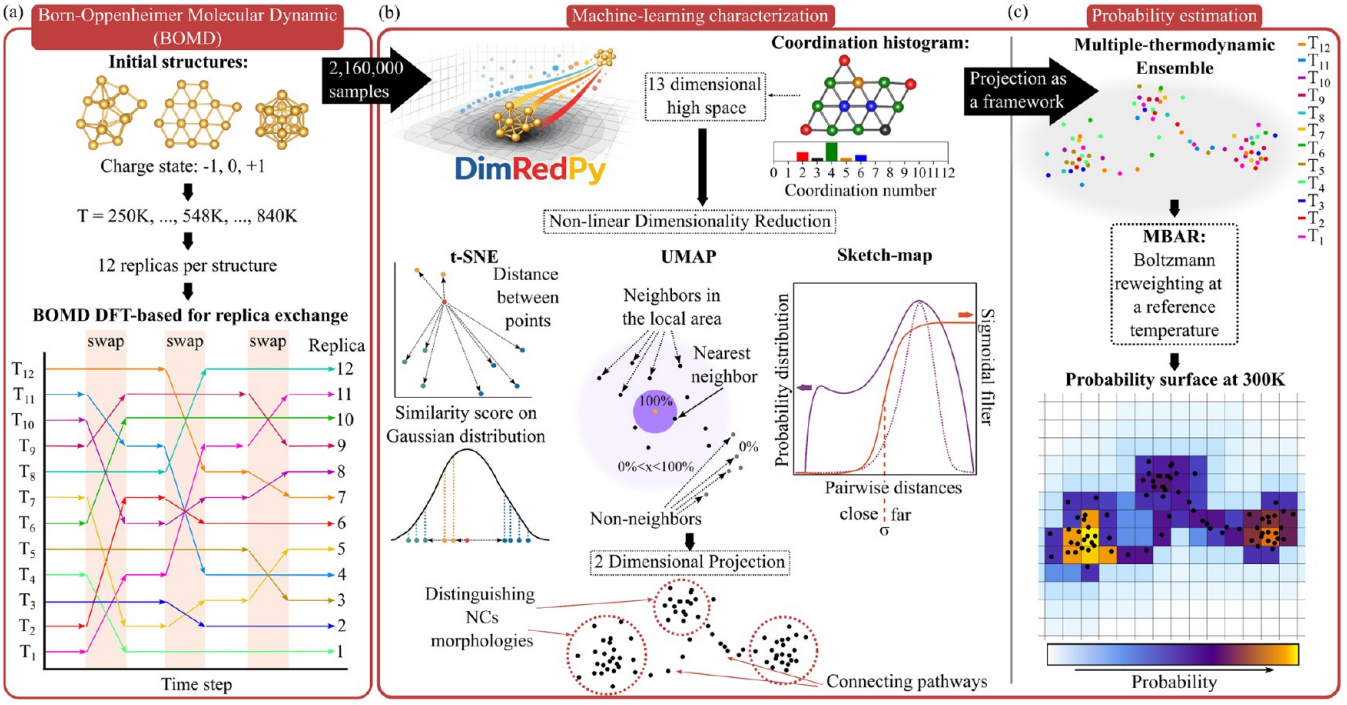


Figure 1. Schematic representation of the computational workflow used to map the configurational landscape of Au_{13} NCs. (a) Replica-exchange Born−Oppenheimer molecular dynamics (RE-BOMD), based on density functional theory (DFT), was carried out for planar, nonplanar, and high-symmetry initial structures in cationic (+), neutral (0), and anionic (−) charge states. We sampled 12 replicas per structure over the temperature ladder from 250.00 to 840.00 K, generating 2160000 configurations. (b) Coordination histograms extracted from the BOMD trajectories define a 13-dimensional configurational descriptor, which is embedded into a two-dimensional latent space using t-distributed stochastic neighbor embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP), and sketch-map. The resulting projections reveal morphological basins and low-density regions connecting structurally distinct configurational regions among planar, nonplanar, and intermediate configurations. (c) The NLDR projections are employed within a multiple-thermodynamic-ensemble canonical framework and reweighted using the multistate Bennett acceptance ratio (MBAR), yielding probability surfaces at 300 K derived from free-energy estimates.

urations may coexist at a given size.[36,37] Based on this, one question arises: how to rationalize the myriad possibilities based on the thermodynamic basin mapping of sub-NCs when different charge states (e.g., cationic, neutral, and anionic) are considered?

First-principles calculations (i.e., an *ab initio* approach based on density functional theory—DFT) have played a crucial role in understanding the structural and electronic properties of a wide range of systems, since the computed electronic structure can be qualitatively or semiquantitatively correlated with experimental measurements, e.g., ultraviolet photoelectron spectroscopy[38] and inverse photoemission spectroscopy.[39,40] In this context, bare and anionic NCs and sub-NCs $Au_n^{(-)}$ in the size range n = 4−25 are widely investigated by using photoelectron spectroscopy and trapped-ion electron diffraction in combination with DFT calculations,[41−47] highlighting the importance of explicitly considering charge states in the characterization of these systems. Despite the generally good agreement between the experimental and theoretical data,[48] a gap often remains between the physical systems probed experimentally and the models typically explored in simulations. Consequently, theoretical studies have frequently focused on describing charge-state or temperature effects separately. This leaves open the challenge of developing a unified framework to rationalize the structural and physicochemical properties of NCs under the combined influence of charge state and temperature.

Concerning temperature effects, not only is the structural evolution yielded from them, but they can also be understood as a strategy of configurational mapping for a particular system described by a connected thermodynamic basin,[49] kinetic restructuring barriers,[50] state-to-state interconversion pathways,[51] and thermalization time scales of fluxional clusters.[52] Through *ab initio* simulations, this configurational landscape is typically accessed through molecular dynamics simulations by considering Born−Oppenheimer molecular dynamics (BOMD), from which a new challenge emerges: the large volume of data generated to be treated and rationalized, making it difficult to obtain a straightforward interpretation based solely on chemical intuition.[53,54] In this context, unsupervised machine-learning-based nonlinear dimensionality reduction (NLDR) methods have increasingly been employed to unravel the configurational spaces sampled by atomistic MD simulations, including those of NCs and sub-NCs. Among NLDR methods, t-distributed stochastic neighbor embedding (t-SNE)[55] has long been one of the most powerful and widely used approaches across diverse applications.[56−58] More recently, Uniform Manifold Approximation and Projection (UMAP)[59] has gained significant popularity owing to its improved computational efficiency and scalability, together with its ability to capture both local and global structural relationships in the data, making it particularly suitable for large data sets.[60−62] In contrast to these more general-purpose manifold-learning approaches, the sketch-map algorithm was specifically designed for atomistic systems and has been demonstrated in studies of polymers and Lennard-Jones NCs,[53,54] as well as in investigations of the core−shell behavior of Au_{25}, Au_{38}, and Au_{40} NCs.[49] In the latter case, sketch-map

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

was employed as a robust framework for constructing probability surfaces based on multiple thermodynamic canonical ensembles, in combination with the multistate Bennett acceptance ratio (MBAR) method.[63]

Using NLDR algorithms to map the configurational landscape of sub-NCs—describing their thermodynamic basins for the 2D ↔ 3D competition under temperature effects—could open a new avenue for embedding methods, given the causal link between structural features and selectivity in nanocatalysed reactions.[64,65] From a broader representation-learning perspective, this idea is conceptually connected to World Models and Joint-Embedding Predictive Architectures (JEPAs), in which high-dimensional dynamical information is organized into a latent space that preserves the variables most relevant to the evolution and prediction of the system.[66−68] Here, rather than learning an autonomous predictive agent, we use this principle to construct chemically interpretable latent maps of atomistic trajectories, where proximity in the embedding reflects configurational similarity, thermodynamic recurrence, and the emergence of dominant structural motifs. Accordingly, we developed an innovative integration of t-SNE, UMAP, and sketch-map and constructed a multiple-thermodynamic canonical ensemble in combination with MBAR. Our framework, based on BOMD with a replica-exchange strategy, efficiently maps the configurational space of Au_{13} sub-NCs while accounting for charge-state effects. This feature has not been systematically addressed for this class of systems. Thus, by decoding the most probable configurations of bare Au_{13} clusters at +, 0, and − charge states, our approach allowed us to construct probability surfaces whose thermodynamic basins successfully distinguish the configurational occurrence of $Au_{13}^{(+,0,-)}$, particularly 2D, 3D, and intermediate structures at 300 K, calculated over a multiple-thermodynamic canonical ensemble. To ensure reproducibility,[69−71] meet the TRUE and FAIR principles,[72−74] and facilitate further applications, we also built a library that integrates t-SNE, UMAP, and sketch-map into a unified workflow, allowing users to perform their own NLDR calculations and reproduce the results presented here.

2. METHODOLOGY

2.1. Workflow and Data

Figure 1 summarizes the workflow employed in this work, outlining the protocols used for configurational landscape mapping and human-interpretable rationalization of the large volume of atomistic data. The following subsection discusses these protocols in detail. We started from three representative initial configurations of Au_{13}: a 2D structure, a 3D putative global minimum configuration (pGMC), and an icosahedral structure (ICO) with high symmetry, the latter being included to account for a core−shell-like structural motif. Each configuration was used as the initial structure for 12 replicas distributed along an increasing temperature ladder, namely, 250.00, 281.00, 312.00, 351.00, 390.00, 440.00, 487.00, 548.00, 608.00, 684.00, 760.00, and 840.00 K, considering cationic, neutral, and anionic charge states. These replicas were simulated using DFT-based replica-exchange Born−Oppenheimer molecular dynamics (RE-BOMD) under nonperiodic boundary conditions, as implemented in the FHI-aims package,[75] as shown in Figure 1a.

We performed the RE-BOMD simulations for a total of 40 ps, attempting temperature exchanges every 5 MD steps according to the Metropolis criterion.[76,77] All sampled configurations were then represented by coordination histograms, yielding a 13-dimensional configurational descriptor that was used as input for interpretable NLDR techniques, namely. sketch-map, UMAP, and t-SNE, to obtain two-dimensional projections, as shown in Figure 1b. The resulting projected spaces preserve the structural similarity encoded in the high-dimensional coordination histogram representation, allowing the configurations sampled along the RE-BOMD trajectories to be organized into low-dimensional maps.

We then combined these NLDR projections with total energies from the RE-BOMD simulations to define multiple thermodynamic canonical ensembles. This information was then employed to construct probability surfaces through the MBAR method, as illustrated in Figure 1c. These probability surfaces enable the identification of the most probable configurations by forming thermodynamic basins at a given reference temperature. Additionally, we organized a structural-analysis protocol in which the configurational similarity sampled throughout the RE-BOMD simulations is correlated with structural parameters, described in the Supporting Information, such as the radius of gyration (R_g), effective coordination number (*ECN*), and radial distribution function (RDF or $g(r)$). We also introduced a procedure to calculate the Hausdorff chirality measure (*HCM*) and correlate it with the center of gravity of the occupied *d*-states (ε_d). We implemented this workflow in `DimRedPy`, a domain-agnostic computational library that integrates t-SNE, UMAP, and sketch-map within a unified Python protocol, along with MBAR-based thermodynamic reweighting, enabling users to perform their own NLDR calculations and reproduce the probability surfaces reported in this work. The implementation is publicly available at https://github.com/Celso0408/Decoding-Au-clusters. The repository provides all the information required to install and use the library, including the `DimRedPy` documentation, API usage, parameter settings, verification suite, and instructions for reproducing the calculations presented here.

2.2. Born−Oppenheimer Molecular Dynamics

To incorporate temperature effects and enhance the sampling of the configurational space, we performed *ab initio* replica-exchange Born−Oppenheimer molecular dynamics (RE-BOMD) simulations based on density functional theory (DFT).[41,42] Given the well-known competition between planar (2D) and three-dimensional (3D) morphologies in Au_{13} NCs,[78] we selected the lowest-energy 2D and 3D structures, together with the highly symmetric icosahedral (ICO) motif, as starting points for the simulations. For each cationic (+), neutral (0), and anionic (−) charge state, 12 replicas of each initial configuration were distributed across the temperature ladder of 250.00, 281.00, 312.00, 351.00, 390.00, 440.00, 487.00, 548.00, 608.00, 684.00, 760.00, and 840.00 K, as shown in Figure 1a.

This setup was motivated by the previously discussed fluxional behavior of the Au_{13} clusters. In particular, neutral Au_{13} has been reported as a highly fluxional cluster, exhibiting facile interconversion among 3D structures, even at low temperatures.[76] In contrast, anionic Au_{13} may experience restricted access to particular regions of the configurational space, so that high-temperature sampling becomes important to overcome the barriers required to reach the corresponding 2D morphological minima.[79] The RE-BOMD approach allows

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

simultaneous simulations at multiple temperatures, in contrast to sequential temperature protocols, thereby providing higher efficiency in escaping local minima and exploring high-energy regions of the probability surface as a function of the charge state.

Each trajectory was propagated at the temperature assigned to its replica, and exchanges between replica pairs were attempted after a predefined number of MD steps according to the Metropolis acceptance criterion. This criterion assesses whether swapping the temperature labels between two replicas preserves detailed balance given the difference in their potential energies and inverse temperatures. In the present simulations, we attempted exchanges every 5 MD steps, corresponding to 0.01 ps for a time step of 2 fs. The RE-BOMD simulations were propagated for a total of 40 ps, yielding 20000 sampled configurations per replica and 240000 configurations per initial structure. Because we used three initial geometries (2D, 3D, and ICO) and three charge states (cationic, neutral, and anionic), the full data set comprised 2160000 sampled configurations.

As a high level of accuracy is required for *ab initio* RE-BOMD, we performed DFT calculations using the all-electron full-potential formalism implemented in the FHI-aims package.[75] Numerically generated atomic orbitals (NAOs)[80] were used as basis sets to describe all electrons, employing `tight` numerical settings and the `tier-1` basis set to provide an accurate description. Since the $Au_{13}^{(+,0,-)}$ NCs were treated as isolated finite systems within the localized NAO formalism implemented in FHI-aims, no periodic boundary conditions or supercell representations were employed. Consequently, no compensating background charge or Makov−Payne-type electrostatic correction was required for the charged systems.

The exchange-correlation energy was described using the Perdew−Burke−Ernzerhof (PBE) functional,[81] including many-body dispersion (MBD) interactions[82] to account for nonlocal correlation effects. A total-energy convergence criterion of 1×10^{-6} eV was adopted for the self-consistent field cycle. Relativistic contributions were treated using the atomic zeroth-order regular approximation (ZORA),[83] derived from the Dirac equation. Canonical sampling in the NVT ensemble was performed using the stochastic Bussi−Donadio−Parrinello thermostat,[84] with a time step of 2 fs for the numerical integration of the MD equations. We used a Gaussian broadening of 5 meV for the electronic occupations to aid convergence.

To distinguish the morphologies of the configurations visited along the full trajectory, particularly between 2D and 3D structures, we employed coordination histograms as collective variables that describe each configuration in a high-dimensional space, as previously introduced by Ceriotti.[54] In this representation, the ith bin of the histogram corresponds to the fraction of atoms in the cluster having coordination number i, i.e., atoms that have i neighbors within a predefined bonding threshold, as shown in Figure 1b. The maximum coordination number was 12, yielding histograms of dimension 13, ranging from 0-fold-coordinated atoms, representing detached or evaporated atoms that were not observed in this study, to 12-fold-coordinated atoms. Although this descriptor reduces dimensionality from 3N Cartesian coordinates to a 13-dimensional representation, it retains the cluster's essential bonding topology and is invariant under rigid translations and rotations.

To explore the high-dimensional configurational space, we then analyzed these configurations using nonlinear dimensionality reduction techniques—sketch-map, UMAP, and FIt-SNE—yielding low-dimensional representations that preserve structural similarities relevant to the configurational organization. Our goal is to obtain a two-dimensional projection that separates structurally distinct regions associated with 2D, 3D, and *quasi*-2D morphologies while preserving the connectivity between configurations sampled along the molecular dynamics trajectories. By identifying the most-visited configurations within the stability basins, the projection provides a structural framework for evaluating the most probable configurations and the relative free-energy differences between basins, which we estimated using the multistate Bennett acceptance ratio (MBAR) method as a Boltzmann reweighting approach.

2.3. Nonlinear Dimensionality Reduction

To unravel the high-dimensional configurational space sampled by RE-BOMD, we employed nonlinear dimensionality reduction (NLDR) techniques to project the 13-dimensional coordination-histogram descriptors into two-dimensional latent representations. In this framework, each configuration is represented by a point in the low-dimensional space, and the relative proximity between points reflects their configurational similarity, as defined by each method's metric and optimization criterion. The charge-state and structural labels were not used during optimization; they were introduced only after projection for visualization and interpretation. Therefore, the resulting maps provide an unsupervised representation of the configurational landscape, allowing the identification of regions associated with 3D, 2D, and *quasi*-2D morphologies, with regions of sparsely sampled configurations connecting them.

In this work, sketch-map, FIt-SNE, and UMAP were used as complementary NLDR methods. We used sketch-map as the reference method because it was designed for atomistic configurational landscapes and provides direct control over the distance scale preserved in the projection. We used FIt-SNE as a fast interpolation implementation of t-SNE, enabling analysis of large data sets while preserving the ability to reveal clustered structures. UMAP was considered for its computational efficiency and graph-based representation of local neighborhood relationships. For each method, all charge states were embedded into the same configurational space, allowing a direct comparison of the regions most frequently sampled by $Au_{13}^{(+)}$, $Au_{13}^{(0)}$, and $Au_{13}^{(-)}$.

2.3.1. Sketch-Map. Sketch-map[53] is a nonlinear dimensionality-reduction algorithm designed to analyze molecular simulation data, preserving meaningful structural similarities by projecting high-dimensional configurations into a low-dimensional space. Starting from a set of pairwise dissimilarities D_{ij} between configurations in the high-dimensional coordination-histogram space and the corresponding distances d_{ij} in the reduced representation, sketch-map determines the optimal low-dimensional embedding by minimizing the stress function:

$$\chi^2 = \sum_{i<j} w_{ij} \left[F_{\sigma,A,B}(D_{ij}) - f_{\sigma,a,b}(d_{ij})\right]^2 \quad (1)$$

where w_{ij} are optional non-negative weights assigned to each pair of points and $F_{\sigma,A,B}$ and $f_{\sigma,a,b}$ are sigmoidal transformation functions applied to pairwise distances in the high- and low-dimensional spaces, respectively. Both functions share the same functional form:

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

$$F_{\sigma,A,B}(R) = 1 - \left[1 + (2^{A/B} - 1)\left(\frac{R}{\sigma}\right)^{A}\right]^{-B/A} \quad (2)$$

and

$$f_{\sigma,a,b}(r) = 1 - \left[1 + (2^{a/b} - 1)\left(\frac{r}{\sigma}\right)^{a}\right]^{-b/a} \quad (3)$$

where $R \equiv D_{ij}$ and $r \equiv d_{ij}$ denote pairwise distances in the high- and low-dimensional spaces, respectively. The parameter σ defines the critical length scale that the algorithm prioritizes during optimization. Thus, distances close to σ are emphasized, whereas very small intrabasin distances and very large interbasin distances are progressively down-weighted. The exponents (A,B) and (a,b) control the sharpness of the sigmoidal transition in the high- and low-dimensional spaces, respectively, with larger values yielding a sharper separation between distance regimes.

Because χ^2 is a nonconvex function and becomes increasingly stiff as the sigmoid sharpens, direct minimization is impractical. Instead, the optimization proceeds in two stages: an initial linear multidimensional scaling step provides a starting configuration, which is then refined by iteratively minimizing eq 1 with the full sigmoidal transformations. These nonlinear mappings allow the final projection to capture the connectivity between metastable basins and transition regions, while discarding irreproducible fine-grained structures within basins and the relative positions of distant, weakly related states. In the present work, the adopted sketch-map parameters were $\sigma = 1.2$, $A = 10.5$, and $B = b = a = 1.0$, selected after evaluating the pairwise-distance distribution of the coordination-histogram descriptors and the resulting separation among 3D, 2D, and *quasi*-2D regions. The original sketch-map code is available at https://github.com/lab-cosmo/sketchmap.

2.3.2. t-Distributed Stochastic Neighbor Embedding and FIt-SNE. t-distributed stochastic neighbor embedding (t-SNE)[55] is one of the most widely used unsupervised nonlinear dimensionality reduction techniques, particularly effective at revealing clustered structures in high-dimensional data sets. The method constructs a low-dimensional representation by first defining a conditional Gaussian probability for the similarity between points i and j in the high-dimensional space,

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2/2\sigma_i^2)} \quad (4)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the high-dimensional descriptors and σ_i is a locally adaptive bandwidth determined pointwise by enforcing a fixed perplexity on the conditional distribution P_i. Specifically,

$$\mathrm{Perp}(P_i) = 2^{H(P_i)} \text{ and } H(P_i) = -\sum_j p_{j|i}\log_2 p_{j|i} \quad (5)$$

where $H(P_i)$ is the Shannon entropy and `perplexity` is a user-defined hyperparameter controlling the effective number of neighbors considered around each point. Larger values of `perplexity` broaden σ_i, emphasizing more global structural relationships, whereas smaller values focus on finer local neighborhoods. The joint probability in the high-dimensional space is then obtained by symmetrization,

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (6)$$

where N is the number of data points.

In the low-dimensional embedding, pairwise similarities are modeled by a Student's t-distribution with one degree of freedom,

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k\neq l}(1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}} \quad (7)$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are the low-dimensional coordinates. The heavy tail of this distribution alleviates the crowding problem by allowing moderately dissimilar points to be placed farther apart in the embedded space. The optimal embedding is obtained by minimizing the Kullback–Leibler divergence between the high- and low-dimensional probability distributions,

$$\mathrm{KL}(P\|Q) = \sum_i \sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}} \quad (8)$$

This objective penalizes cases in which high-dimensional neighbors are mapped far apart in low-dimensional space, thereby emphasizing the preservation of local neighborhoods.

Despite its broad applicability, standard t-SNE scales approximately as $O(N^2)$ in both memory and computational cost, making it prohibitive for very large data sets.[56,58] To overcome this limitation, we employed Fast Interpolation-based t-SNE (FIt-SNE),[85] which accelerates the evaluation of repulsive forces through polynomial interpolation and the Fast Fourier Transform (FFT), reducing the computational cost to approximately $O(N\log N)$ while retaining the ability of t-SNE to identify clustered configurations. In the implementation used here, the main projection parameters were `perplexity`, `max_inter`, and `early_exag_coeff`. The `perplexity` parameter determines the effective neighborhood size in the high-dimensional space, `max_inter` controls the interpolation-mediated approximation of the repulsive interactions, and `early_exag_coeff` governs the initial exaggeration phase that promotes the formation of separated but connected clusters during early optimization. After testing the sensitivity of the projection to these parameters, the adopted values were `perplexity = 250`, `max_inter = 1000`, and `early_exag_coeff = 1000`. The original FIt-SNE code is available at https://github.com/KlugerLab/FIt-SNE.

2.3.3. Uniform Manifold Approximation and Projection. Uniform Manifold Approximation and Projection (UMAP)[59] is a nonlinear dimensionality reduction method grounded in fuzzy topological theory and manifold learning. Owing to its favorable balance between computational efficiency and preservation of neighborhood relationships, UMAP has become a widely used alternative to t-SNE for large data sets.[60−62] Similar to t-SNE, UMAP starts by constructing a weighted graph representing the local connectivity of the data in the high-dimensional space. For each point, edge weights are assigned through a locally adaptive kernel,

$$w_{j|i} = \exp\left[-\frac{\max(0, d(\mathbf{x}_i, \mathbf{x}_j) - \rho_i)}{\sigma_i}\right] \quad (9)$$

where $d(\mathbf{x}_i\mathbf{x}_j)$ is the pairwise distance computed according to the selected metric, ρ_i is the distance from point i to its nearest

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

neighbor, and σ_i is a locally adaptive bandwidth. The subtraction of ρ_i ensures that each point has at least one neighbor with strong connectivity, making the graph construction less sensitive to variations in local density. The bandwidth σ_i is determined by enforcing a fixed local connectivity condition,

$$\sum_{j=1}^{k} w_{j|i} = \log_2(k) \tag{10}$$

where k corresponds to the `n_neighbors` hyperparameter. Larger values of `n_neighbors` include more global information in the graph, whereas smaller values emphasize finer local relationships. The directed weights are then symmetrized through the fuzzy union operation,

$$w_{ij} = w_{j|i} + w_{i|j} - w_{j|i}w_{i|j} \tag{11}$$

The `metric` parameter defines the distance function used to compare the high-dimensional coordination-histogram descriptors. Three candidate metrics were evaluated in this work: the `euclidean` distance,

$$d_{\text{euc}} = \|\mathbf{x}_i - \mathbf{x}_j\|_2 \tag{12}$$

which treats all dimensions as independent and equally scaled; the `cosine` distance,

$$d_{\cos} = 1 - \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\|_2 \, \|\mathbf{x}_j\|_2} \tag{13}$$

which measures angular separation and is less sensitive to vector magnitude; and the `mahalanobis` distance,

$$d_{\text{mah}} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}} \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)} \tag{14}$$

where $\mathbf{S}$ is the sample covariance matrix of the descriptor space or its regularized inverse when required. The Mahalanobis metric accounts for interfeature correlations and heterogeneous scaling among histogram components, reducing to the Euclidean distance when $\mathbf{S} = \mathbf{I}$.

In the low-dimensional space, UMAP models the edge weights as

$$\hat{w}_{ij} = (1 + a\|\mathbf{y}_i - \mathbf{y}_j\|_2^{2b})^{-1} \tag{15}$$

where $\mathbf{y}_i$ and $\mathbf{y}_j$ are the embedded coordinates. The parameters a and b are determined by fitting a smooth curve controlled by `min_dist`, which defines the minimum distance allowed between nearby points in the low-dimensional representation. Smaller values of `min_dist` allow tighter packing and sharper cluster boundaries, whereas larger values lead to more diffuse embeddings. The optimal embedding is obtained by minimizing the binary cross-entropy between the high- and low-dimensional fuzzy graph representations,

$$\mathcal{L} = \sum_i \sum_j \left[w_{ij} \log \frac{w_{ij}}{\hat{w}_{ij}} + (1 - w_{ij}) \log \frac{1 - w_{ij}}{1 - \hat{w}_{ij}} \right] \tag{16}$$

This objective penalizes both false connections and false disconnections between the high- and low-dimensional graphs, unlike the asymmetric KL divergence used in t-SNE.

In the present work, the UMAP projections used for the MBAR analysis were generated with `metric = "mahalanobis"`, `n_neighbors = 50`, and `min_dist = 0.001`, which provided the best compromise between structural separation and the preservation of connectivity among 3D, 2D, and *quasi*-2D configurations. UMAP projections were performed using the Python implementation available at https://github.com/lmcinnes/umap.

Finally, we integrated all NLDR calculations into the `DimRedPy` workflow developed in this work, which provides a unified protocol for applying sketch-map, FIt-SNE, and UMAP to coordination-histogram descriptors and for using the resulting two-dimensional projections as collective variable spaces for subsequent MBAR reweighting. The implementation is publicly available at https://github.com/Celso0408/Decoding-Au-clusters and includes the information needed to install and use the library, including documentation, API usage, parameter settings, verification scripts, and instructions for reproducing the calculations presented here.

2.4. Free-Energy Surface Calculation Method

We used the multistate Bennett acceptance ratio (MBAR)[63] as a Boltzmann reweighting framework to reconstruct probability and free energy surfaces from multiple thermodynamic canonical ensembles generated by the RE-BOMD simulations. MBAR is a statistically optimal approach for combining samples collected from different thermodynamic states to estimate free energy differences and thermodynamic observables with minimum variance. It generalizes the Bennett acceptance ratio (BAR),[86] which corresponds to the two-state limit, and is formulated within the extended bridge sampling formalism, in which all configurations sampled across all states are treated collectively via importance sampling.[87] Because of its maximum-likelihood foundation, MBAR provides a rigorous route to reconstruct thermodynamic observables from simulations carried out at multiple temperatures.[88,89]

In the present work, MBAR was applied independently for each charge state, since $Au_{13}^{(+)}$, $Au_{13}^{(0)}$, and $Au_{13}^{(-)}$ correspond to distinct electronic Hamiltonians and different electron numbers. Therefore, we did not use MBAR to estimate free-energy differences between charge states. Instead, for each charge state, we combined the configurations sampled across the temperature ladder to reconstruct charge-specific probability surfaces at a reference temperature, chosen here as 300 K. We used three initial structures, namely, 2D, 3D, and ICO, to improve configurational sampling, while the replica temperatures defined the thermodynamic states in the MBAR calculation.

For a given charge state, the reduced potential of configuration x_n evaluated at thermodynamic state k is defined as

$$u_k(x_n) = \beta_k[U(x_n) + V_k(x_n)] \tag{17}$$

where $\beta_k = (k_B T_k)^{-1}$, $U(x_n)$ is the DFT potential energy of configuration x_n, and $V_k(x_n)$ represents an eventual biasing potential. In the present unbiased RE-BOMD simulations, $V_k(x_n) = 0$, so that the reduced potential is determined only by the DFT energy and the temperature of the corresponding thermodynamic state. Since the probability surfaces are configurational, kinetic-energy contributions are integrated out and do not enter the relative configurational weights. For numerical stability, the DFT energies can be shifted by a constant reference value within each charge state, since constant energy shifts do not alter the MBAR-reweighted probabilities.

Given K sampled thermodynamic states, with N_k configurations collected at temperature T_k and $N = \sum_k N_k$ total configurations, MBAR estimates the reduced free energies $\hat{f}_k$ by solving the self-consistent equations

$$\hat{f}_i = -\ln \sum_{n=1}^{N} \frac{\exp[-u_i(x_n)]}{\sum_{k=1}^{K} N_k \exp[\hat{f}_k - u_k(x_n)]} \tag{18}$$

Once the reduced free energies are obtained, the normalized statistical weight of configuration x_n in a target thermodynamic state i is given by

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

$$W_{n,i} = \frac{\exp[\hat{f}_i - u_i(x_n)]}{\sum_{k=1}^{K} N_k \exp[\hat{f}_k - u_k(x_n)]} \tag{19}$$

These weights satisfy $\sum_n W_{n,i} = 1$ and allow the reconstruction of equilibrium averages at any sampled or reweighted thermodynamic state. In particular, the 300 K reference state was treated as an evaluation state by computing its reduced potential for all configurations sampled along the RE-BOMD trajectories.

To construct the probability surfaces, the two-dimensional coordinates obtained from the NLDR projections were used as collective variables,

$$\mathbf{s}(x_n) = [s_1(x_n), s_2(x_n)] \tag{20}$$

where $\mathbf{s}(x_n)$ corresponds to the sketch-map, FIt-SNE, or UMAP coordinates of configuration x_n. The projected space was discretized into two-dimensional bins, and the MBAR-reweighted probability associated with each bin B_l at the target state i was calculated as

$$P_i(B_l) = \sum_{n=1}^{N} W_{n,i} \mathbb{I}[\mathbf{s}(x_n) \in B_l] \tag{21}$$

where $\mathbb{I}[\mathbf{s}(x_n) \in B_l]$ is an indicator function equal to one when the projected configuration belongs to bin B_l and zero otherwise. The corresponding free-energy surface, or potential of mean force, is then obtained as

$$F_i(B_l) = -k_B T_i \ln P_i(B_l) + C \tag{22}$$

where the constant C is chosen such that the global minimum satisfies $F_{min} = 0$. Equivalently, the probability surfaces shown in this work are reported as $\log(P)$, normalized with respect to the most probable bin,

$$\log P_i(B_l) = \ln P_i(B_l) - \max_l [\ln P_i(B_l)] \tag{23}$$

so that $\log(P) = 0$ corresponds to the highest-probability region of the configurational landscape.

In practice, the workflow consists of i) assigning each RE-BOMD configuration to the thermodynamic state corresponding to the temperature at which it was propagated; ii) computing the reduced potentials of all configurations over the temperature ladder and at the 300 K reference state; iii) solving the MBAR equations to obtain the reduced free energies and statistical weights; iv) projecting the configurations onto the two-dimensional NLDR coordinates; and v) accumulating the MBAR weights in the projected space to construct the probability and free-energy surfaces. Thus, the NLDR projections provide the structural framework, whereas MBAR provides the thermodynamic weighting required to identify the most probable configurational basins and their connections.

The bin resolution used to construct the probability surface is a user-controlled parameter: increasing the number of bins yields a more detailed landscape but also requires higher sampling density to avoid noisy or sparsely populated regions. Importantly, MBAR itself does not require discretization of the energy space; binning is applied only after statistical reweighting for visualization and interpretation of the free-energy surface on the two-dimensional configurational map. The MBAR calculations were performed using the Python implementation available in `pymbar`: https://github.com/choderalab/pymbar.

2.5. Energetic Properties and Catalytic-Relevant Electronic Descriptors

To quantify the energetic stability associated with the 2D vs 3D competition, we computed the relative energy (ΔE_{rel}) of each optimized isomer with respect to the structure obtained from the ICO starting motif. At $T = 0$ K, ΔE_{rel} was defined for each charge state $q \in \{+,0,-\}$ as

$$E_{rel,i}^{(q)} = E_i^{(q)} - E_{ICO}^{(q)} \tag{24}$$

where $E_i^{(q)}$ and $E_{ICO}^{(q)}$ are the total energies of configuration i and of the ICO-reference structure, respectively, both evaluated at the same charge state and computational level. Within this convention, $\Delta E_{rel} = 0$ corresponds to the ICO reference, whereas negative values indicate configurations energetically more stable than the ICO-derived structure. We applied the same definition to calculations with and without many-body dispersion (MBD) corrections, allowing us to directly assess the role of vdW interactions in the 2D/3D energetic balance. These $T = 0$ K relative energies were used only for the optimized-isomer stability analysis. In contrast, the finite-temperature MBAR probability surfaces were constructed from the DFT potential energies of the configurations sampled along the RE-BOMD trajectories.

Moreover, we assessed the average Au−Au bonding strength through the binding energy per atom (E_b). To ensure charge consistency between the cluster and the isolated-atom reference, E_b was defined for each charge state $q \in \{+,-\}$ as

$$E_b^{(q)} = \frac{E_{Au_{13}}^{(q)} - 12E_{Au}^{(0)} - E_{Au}^{(q)}}{13} \tag{25}$$

where $E_{Au_{13}}^{(q)}$ is the total energy of the Au_{13} cluster in charge state q, $E_{Au}^{(0)}$ is the total energy of a neutral isolated Au atom, and $E_{Au}^{(q)}$ is the total energy of an isolated Au atom carrying the same charge as the cluster. For the neutral case, eq 25 reduces to

$$E_b^{(0)} = \frac{E_{Au_{13}}^{(0)} - 13E_{Au}^{(0)}}{13} \tag{26}$$

With this sign convention, more negative E_b values indicate stronger average binding within the NC, while less-negative values indicate weaker cohesive stabilization.

To characterize the electronic structure in a manner commonly used in heterogeneous catalysis, we evaluated the center of gravity of the occupied d-states (ε_d), which corresponds to the average energy ε weighted by the occupied d-projected density of states, $n_d(\varepsilon)$. This descriptor originates from the d-band center model,[90] widely used for transition-metal surfaces, as it provides insight into catalytic activity by revealing the relative positions of metal d-states with respect to the frontier electronic levels.[91] For extended metallic systems, the d-band center is commonly referenced to the Fermi level (E_f), which is well-defined as the electronic chemical potential. However, for finite systems such as NCs, where the electronic spectrum is discrete, the definition of E_f becomes less straightforward. Therefore, to enable consistent comparisons across configurations and charge states, we adopted the HOMO energy as the internal reference level. Moreover, ε_d should be regarded as a comparative electronic descriptor rather than as a direct measure of catalytic activity.

In this way, the occupied d-state center was calculated as

$$\varepsilon_d^{(q)} = \frac{\int_{-\infty}^{\varepsilon_{HOMO}^{(q)}} (\varepsilon - \varepsilon_{HOMO}^{(q)}) n_d^{(q)}(\varepsilon) d\varepsilon}{\int_{-\infty}^{\varepsilon_{HOMO}^{(q)}} n_d^{(q)}(\varepsilon) d\varepsilon} \tag{27}$$

where $\varepsilon_{HOMO}^{(q)}$ is the HOMO energy of the cluster in charge state q and $n_d^{(q)}(\varepsilon)$ is the density of occupied d-states projected onto the Au atomic orbitals. For spin-polarized calculations,

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

the spin components of the projected density of states were summed before evaluating eq 27. Since the HOMO energy is used as the reference, ε_d values are generally negative, and larger, or less negative, values indicate that the center of the occupied *d*-states lies closer to the frontier electronic level.

Within this convention, ε_d is used as a comparative descriptor of electronic reactivity among structurally related configurations. Configurations with ε_d closer to the HOMO level are expected to exhibit stronger interaction with adsorbate molecular orbitals. In contrast, configurations with more negative ε_d values are associated with deeper occupied *d*-states and, consequently, weaker electronic reactivity. Therefore, the combined analysis of ΔE_{rel}, E_b, and ε_d provides a direct connection between thermodynamic stability, average bonding strength, and catalytic descriptors for the $Au_{13}^{(+,0,-)}$ configurational landscape. Additionally, Hirshfeld charge[92] analysis was performed to assess the charge distribution across different sites within the 2D and 3D $Au_{13}^{(+,0,-)}$ configurations.

3. RESULTS AND DISCUSSION

3.1. Systems at T = 0 K

Before addressing finite-temperature charge-state effects in $Au_{13}^{(+,0,-)}$, we first analyzed the energetic and structural properties of several $Au_{13}^{(+,0,-)}$ models at T = 0 K, previously investigated for the neutral state.[78,93] These structures were named according to their morphology, namely, Biplanar (BBP), Cuboctahedral (CUBO), Double Simple Cubic (DSC), Hexagonal Bilayer (HBL), One-Folded Parallelogram (OFP), Four-Folded Parallelogram (FFP), Three-Folded Hexagon (TFH), and Icosahedral (ICO), as shown in Figure 2a. We computed their relative energies (ΔE_{rel}) using eq 24, with the ICO structure obtained via the Mackay scheme as the reference.[94] Previously, we highlighted the energetic contributions governing the 2D/3D competition in neutral Cu_{13}, Ag_{13}, and Au_{13} clusters, demonstrating that vdW contributions significantly modulate their relative stability.[78] As depicted in Figure 2b, the same trend is observed for $Au_{13}^{(0)}$: i) in the absence of vdW corrections, the 2D-FFP motif is favored by −0.47 eV; ii) the inclusion of MBD corrections destabilizes this motif by +0.58 eV; and iii) consequently, a close energetic competition emerges between the 3D-TFH and 2D-FFP motifs, with only a 0.05 eV difference in ΔE_{rel}.

This vdW-induced modulation is also observed when charge effects are introduced. For the cationic clusters, in addition to the TFH → Cut Prism Box (CPB) transition, $Au_{13}^{(+)}$ preferentially stabilizes in a 3D morphology after including MBD corrections, whereas the anionic $Au_{13}^{(-)}$ state always favors the 2D arrangement. It is worth noting the ICO → CUBO transition during geometry optimization for the cationic state, as well as the TFH → Inverse Arrow Head (IAH) transition for the anionic state. These transitions observed for $Au_{13}^{(+)}$ and $Au_{13}^{(-)}$, as detailed in Figure S1, could be associated with a tendency toward more symmetric configurations, which we suggest may be driven by their even electron count, in contrast to the odd-electron character of $Au_{13}^{(0)}$. Furthermore, the incorporation of MBD corrections changes the 2D/3D energetic balance in the cationic state, shifting the 2D → 3D stabilization from +0.14 eV without vdW corrections to −0.25 eV with MBD corrections. In contrast, for

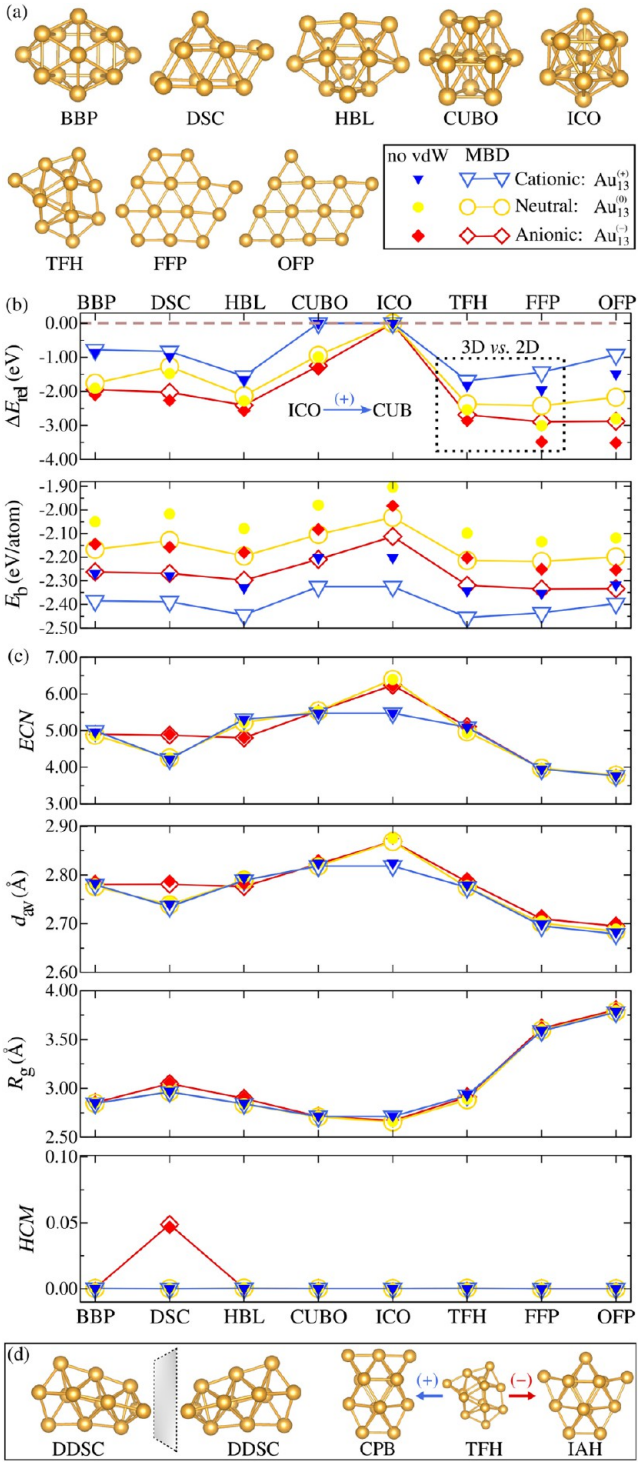


Figure 2. (a) Representation of Au_{13} structures in biplanar (BBP), cuboctahedral (CUBO), double simple cubic (DSC), hexagonal bilayer (HBL), one-folded parallelogram (OFP), four-folded parallelogram (FFP), three-folded hexagon (TFH), and icosahedral (ICO) configurations. (b) Relative energy (ΔE_{rel}), calculated with respect to the ICO configuration, and average binding energy (E_b) obtained with and without many-body dispersion (MBD) van der Waals corrections. (c) Structural-parameter evaluation through the effective coordination number (*ECN*), average bond distance (d_{av}), radius of gyration (R_g), and Hausdorff chirality measure (*HCM*). (d) Specular image of distorted DSC (DDSC) and representative structural transitions of $Au_{13}^{(+,-)}$.

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

the anionic state, MBD corrections reduce the 3D → 2D stabilization from −0.62 eV without vdW corrections to −0.21 eV.

Through the calculation of the average binding energy (E_b), according to eq 25, one observes an increase in the magnitude of E_b across the sequence neutral → anionic → cationic. Including MBD corrections consistently enhances bonding interactions compared with the no-vdW protocol. For all configurations, E_b varies nearly linearly across the charge states, with an average stabilization of approximately −0.11 eV/atom from neutral to anionic and from anionic to cationic species.

Moving to the characterization of structural changes in $Au_{13}^{(+,0,-)}$ clusters, we analyzed the effective coordination number (*ECN*), average bond distance (d_{av}), radius of gyration (R_g), and Hausdorff chirality measure (*HCM*). As shown in Figure 2c, no significant charge-induced variations are observed for a given morphology, except for the anionic DSC and HBL structures. Therefore, this observation reinforces the importance of temperature effects in configurational assessment under *operando* conditions. As previously mentioned, upon the ICO → CUBO transition in the cationic state, the anionic DSC structure distorts into a distorted DSC (DDSC) configuration, as reflected by a slight increase in the *HCM* value to approximately 0.05. Figure 2d shows the corresponding mirror image. Despite this distortion, and although the anionic HBL configuration also transitions to CPB, the transition structures remain essentially achiral, except for the weak chirality induced in the DDSC motif.

To quantitatively distinguish between 2D and 3D structures, we defined the following thresholds: $ECN \leq 3.75$ and $R_g \geq 3.75$ Å as indicative of 2D structures and $ECN \geq 4.25$ and $R_g \leq 3.00$ Å as characteristic of 3D morphologies. For d_{av}, values ≤ 2.70 Å correspond to 2D structures, whereas values ≥ 2.75 Å correspond to 3D configurations.

3.2. Finite-Temperature Effects: RE-BOMD Calculations and Structural Classification

We employed *ab initio* calculations within a RE-BOMD regime to explore the configurational space of $Au_{13}^{(+,0,-)}$ clusters across multiple thermodynamic states, enabling a broad temperature interval to sample the thermodynamic basins associated with different structural patterns, as well as the connections among them. Our RE-BOMD protocol attempts replica exchanges according to the Metropolis acceptance criterion, whereby temperature labels associated with 250.00, 281.00, 312.00, 351.00, 390.00, 440.00, 487.00, 548.00, 608.00, 684.00, 760.00, and 840.00 K are swapped, or not, every 5 MD steps.[77] This procedure is suitable for efficient configurational mapping through *ab initio* molecular dynamics.[76] The effectiveness of the sampling was confirmed by an average swapping rate higher than 60% for all exchange attempts throughout the RE-BOMD simulations (see Table S1 and Figure S2), indicating efficient communication among replicas and supporting the construction of multiple thermodynamic canonical ensembles for each charge state.

Thus, as detailed in the Methodology section and represented in Figure 1a, three initial configurations were considered for each $Au_{13}^{(+,0,-)}$ charge state: the lowest-energy 2D configuration, the lowest-energy 3D configuration, and the ICO structure as a high-symmetry reference. We propagated each initial configuration using 12 replicas distributed along the aforementioned temperature ladder. Therefore, with the setup detailed in the Methodology section, the RE-BOMD simulations yielded 2160000 points in the configurational space of $Au_{13}^{(+,0,-)}$ clusters across the three charge states. Each point corresponds to a DFT total energy, temperature, and structural configuration. From these configurations, we used coordination histograms to describe local atomic environments, yielding 13-dimensional vectors that define the high-dimensional configurational space used as input for the NLDR algorithms for structural classification.

3.2.1. Dimensionality Reduction. The sampled high-dimensional configurational space obtained from the RE-BOMD calculations—as schematized in Figure 1b—was systematically analyzed through sketch-map, FIt-SNE, and UMAP as NLDR techniques. The two-dimensional projections preserve configurational similarities from the high-dimensional space, thereby rationalizing the sampled structures and the global differences in connectivity among densely sampled regions, which can be interpreted as the competition between 2D and 3D structural patterns under temperature effects. Thus, sparsely sampled regions correspond to transition configurations that connect densely sampled regions associated with well-defined configurations because these regions have reduced sampling probability due to higher energies and smaller Boltzmann weight contributions.

Figure 3 depicts the two-dimensional projections obtained for all configurations visited by the cationic, neutral, and anionic charge states. These charge states are highlighted, respectively, by blue, yellow, and red points, selected as 1 point every 4000 MD steps from the full sampling shown in gray. We embedded all charge states in the same configurational space to identify which regions of the configurational landscape are most frequently sampled during the RE-BOMD simulations for each charge state.

We extensively tested the sketch-map, FIt-SNE, and UMAP algorithms across their main metrics and hyperparameters to identify the embeddings that best distinguish the 2D and 3D configurational regions. Across all embeddings, cationic configurations occupy a densely populated central region of the projections. In contrast, the anionic configurations are more prevalent in the outer regions of the configurational map, as highlighted by the central and outer regions in Figure 3a−c. Here, we observed that the neutral charge state occupies an intermediate regime, bridging the preferred spatial distributions of the cationic and anionic NCs. This systematic migration from central → outer regions from cationic → anionic charge states is consistently captured by all three dimensionality reduction methods. As shown later, the central and outer regions correspond to 3D- and 2D-like configurations, respectively.

The resulting low-dimensional representations provide a structural framework for identifying highly sampled configurational regions that are separated by regions with sparsely sampled configurations. We then quantified the corresponding thermodynamic stability and free-energy differences between these regions using the MBAR formalism. The influence of the hyperparameters associated with each method, as well as the values adopted in this work, is discussed in Section 3.4.1.

3.3. Probability Surface Interpretation

Aiming to identify the high-stability regions obtained from the unsupervised NLDR approaches, in which the statistical contributions of all configurations sampled within the 250 K−840 K temperature interval were considered, we employed

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

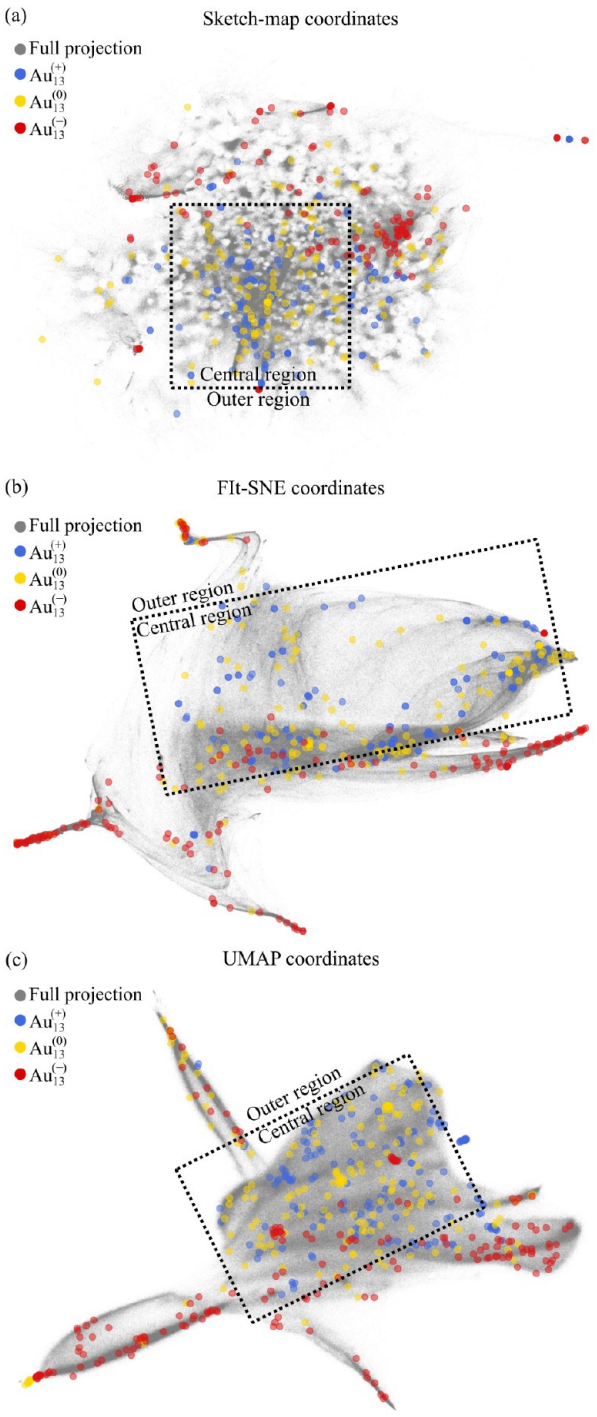


Figure 3. (a) Sketch-map, (b) FIt-SNE, and (c) UMAP projections obtained from the full trajectory, represented by gray points, and separated by charge state. Blue, yellow, and red points correspond to $Au_{13}^{(+)}$, $Au_{13}^{(0)}$, and $Au_{13}^{(-)}$, respectively. We subsampled the charge-state configurations by selecting 1 point every 4000 MD steps from the full sampling for better visualization. Colored points are 600 times larger than gray ones.

the MBAR method to calculate probability surfaces for each Au_{13} charge state through the protocol depicted in Figure 1c. Although the RE-BOMD simulations were performed over a set of temperatures, MBAR estimates probability surfaces from a multiple-thermodynamic-canonical-ensemble perspective by reweighting the sampled configurations to a reference temperature, chosen here as 300 K. This protocol allows us to construct thermodynamic basins and their configurational connections using the logarithm of the resulting probabilities. Since the free energy is related to the probability as $F = -k_B T \ln(P) + C$, regions with larger $\log(P)$ correspond to lower free-energy regions at the reference temperature for a given charge state.[49]

Because standard MD may suffer from kinetic trapping at 300 K, our RE-BOMD framework improves configurational sampling by facilitating barrier crossing through exchanges between low- and high-temperature replicas, reducing the risk of trapping in local minima. Since replica exchange disrupts the direct temporal continuity of the trajectories, the apparent connections in the low-dimensional maps presented here should not be interpreted as kinetic pathways but regions of structural similarity connecting the high-probability configurations. We then feed the sampled configurations from the resulting canonical thermodynamic ensembles into MBAR, which we use as a statistical reweighting framework to reconstruct the thermodynamic observables and probability surfaces.

Figure 4 shows the impact of cationic and anionic charge states on the probability surface of the Au_{13} cluster as a result of the MBAR and sketch-map projection combination; the neutral $Au_{13}^{(0)}$ case is shown in Figure S3. In Figure 4a, most of the probability associated with $Au_{13}^{(+)}$ lies in the central region, leading to a landscape characterized by multiple metastable basins associated with distinct 3D configurations. This behavior reflects the highly fluxional nature of these structures, which interconvert through thermally accessible configurational connections. One realizes that the presence of a positive charge amplifies this behavior when compared to the neutral $Au_{13}^{(0)}$ state, since the transition $Au_{13}^{(0)} \rightarrow Au_{13}^{(+)}$ is accompanied by a probability shift toward the central basins and by a decrease in the outer regions associated with dominant 2D configurational basins. This effect may be rationalized by the increased electrophilic character of Au atoms in $Au_{13}^{(+)}$ upon electron removal, which favors higher coordination and, consequently, the preferential formation of compact 3D configurations.

Thus, competition among Au atoms for charge compensation further enhances the fluxionality of the 3D $Au_{13}^{(+)}$ structures, with charge dynamically delocalizing across the NC. Figure S5a shows the Hirshfeld partial charges for the cationic 2D and 3D configurations. Interestingly, both the higher- and lower-coordinated atoms display the greatest charge depopulation and the most pronounced cationic character. The central atoms of the 3D-ICO/CUB and the low-coordination atoms of the 2D-OFP/FFP configurations exhibit partial charge populations higher than +0.10 e, whereas atoms with intermediate coordination show lower values. The high electrophilicity of Au, together with the tendency to avoid highly cationic states, may contribute to the structural preference for 3D configurations observed for the cationic NCs at $T = 0$ K. Nevertheless, the combined electronic and entropic contributions still favor the 2D over 3D configurations.

We identified basins related to TFH, IAH, CPB, and Prism Box (PB) configurations, as represented with their coordination histograms in Figure 4a, which have been previously reported as stable or metastable Au_{13} motifs in earlier studies

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

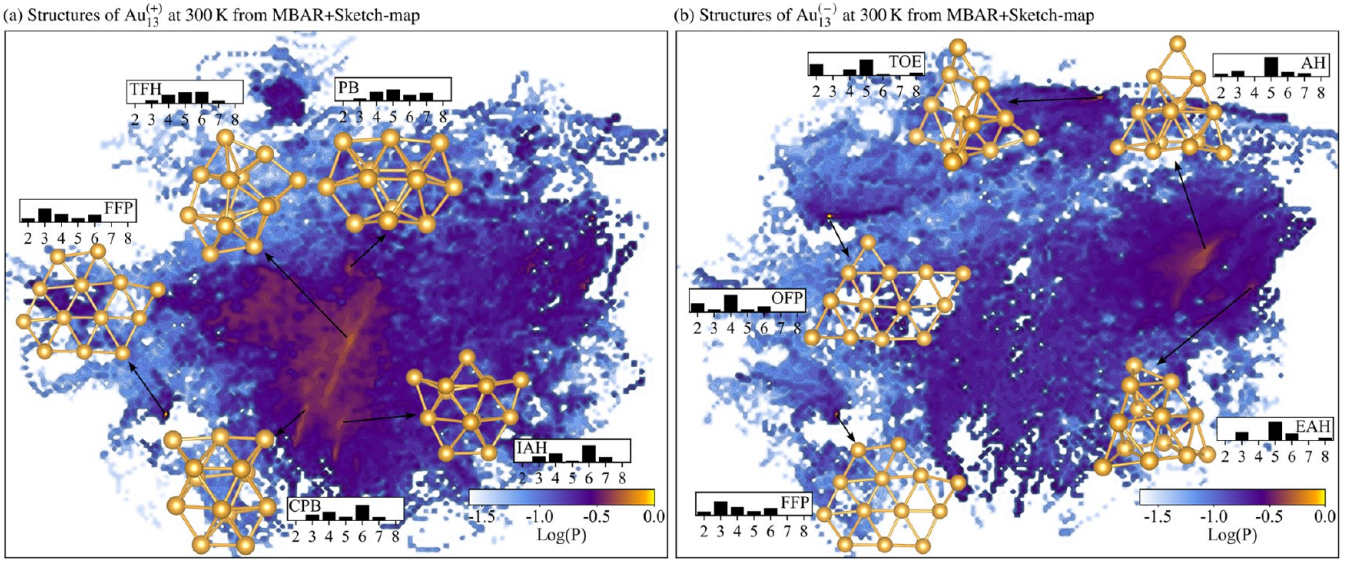


Figure 4. Probability surface for (a) cationic $Au_{13}^{(+)}$ and (b) anionic $Au_{13}^{(-)}$ nanocluster configurations at 300 K, calculated via MBAR as a function of the sketch-map coordinates. For each selected structure, we also show the corresponding 13-dimensional coordination histogram and its acronym.

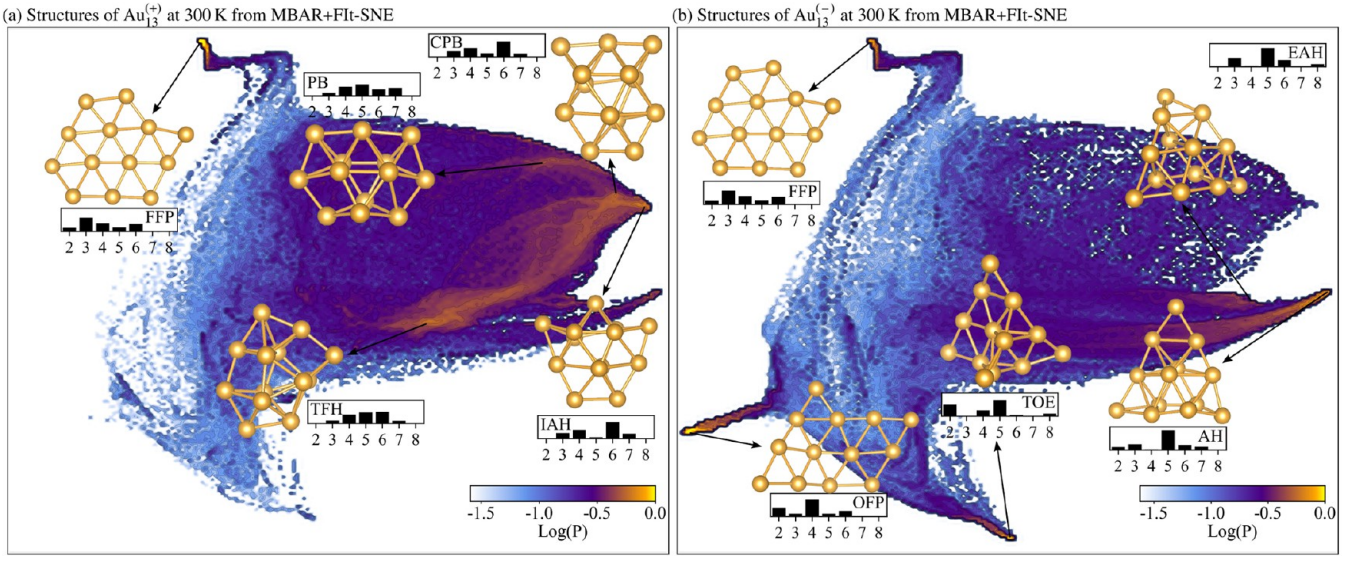


Figure 5. Probability surface for (a) cationic $Au_{13}^{(+)}$ and (b) anionic $Au_{13}^{(-)}$ nanocluster configurations at 300 K, calculated via MBAR as a function of the FIt-SNE coordinates. For each selected structure, we also show the corresponding 13-dimensional coordination histogram and its acronym.

for different charge states.[45,95−97] Interestingly, the TFH isomer presents a notable probability of occurrence among the 3D basins, despite the observation of a TFH → CPB transition in the cationic state at T = 0 K, which reinforces the importance of temperature effects. Moreover, the 2D-FFP configuration also contributes to a high-probability basin in the outer region for $Au_{13}^{(+)}$. Together with the increased probability of the 3D-TFH isomer, this finding highlights the crucial role of temperature in increasing the population of configurations that are less favored at T = 0 K.

Several trends observed in the cationic state also appear in $Au_{13}^{(0)}$, albeit with a distinct balance between the central and outer regions. For the neutral cluster, the most probable basin corresponds to the 2D-FFP configuration. However, a fraction of the probability migrates from the central region toward the outer region, where it becomes associated with the 2D-OFP configuration and emerges as the second most probable basin. This behavior shows that the charge state reduces the basin depth associated with central 3D configurations, indicating partial suppression of the fluxional behavior observed for the cationic cluster.

For the anionic state, Figure 4b shows the opposite behavior compared to the cationic system. Taking the transition of the charge state as $Au_{13}^{(0)} \rightarrow Au_{13}^{(-)}$, a suppression of the fluxionality associated with the 3D configurations in the central region is observed, owing to an almost complete migration of probability toward the outer regions of the surface. In this case, the anionic cluster exhibits an additional distinctive feature: the formation of basins with configurations intermediate between 2D and 3D structures. Here, we refer to these motifs as *quasi*-2D, since they exhibit structural signatures that can be locally classified as either 2D or 3D patterns, as shown

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

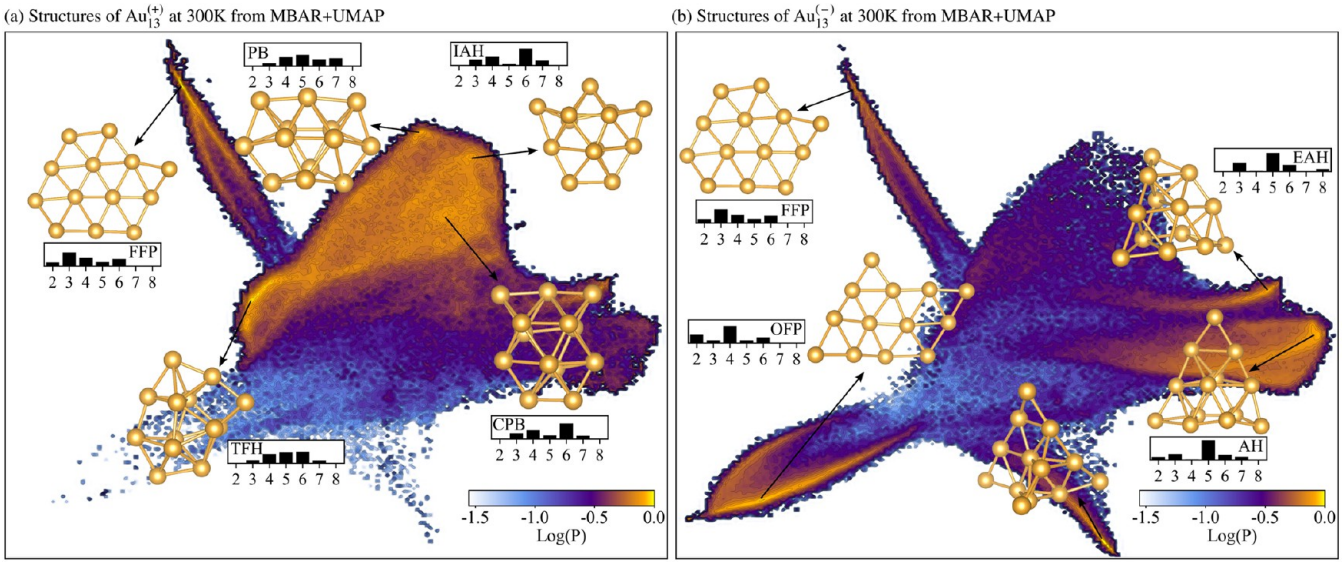


Figure 6. Probability surface for (a) cationic $Au_{13}^{(+)}$ and (b) anionic $Au_{13}^{(-)}$ nanocluster configurations at 300 K, calculated via MBAR as a function of the UMAP coordinates. For each selected structure, we also show the corresponding 13-dimensional coordination histogram and its acronym.

in Section 3.5. Thus, we label these configurations as *Triangle Opened Edge* (TOE), *Arrow Head* (AH), and *Edgeless Arrow Head* (EAH).

The Hirshfeld partial charges for the anionic configurations shown in Figure S5c reveal a distinct charge distribution depending on the configuration. The 3D configurations exhibit excess charge distributed throughout the cluster, giving rise to a repulsive electrostatic contribution. In contrast, the 2D and quasi-2D configurations display slightly cationic (≈+0.02e) or neutral central atoms, while the largest excess charge is concentrated on the less-coordinated atoms (≈−0.15e), resulting in an attractive electrostatic character.

Thereby, the enhanced probability of 2D and *quasi*-2D configurations for $Au_{13}^{(-)}$ is consistent with the earlier discussion, suggesting that the excess electronic charge renders the NC structurally softer. Under these conditions, Au atoms can preferentially occupy less-coordinated sites, thereby favoring the formation of more open structural motifs. This structural flexibility increases the propensity for 2D and *quasi*-2D configurations, in contrast to the cationic state. For instance, the OFP configuration has the highest probability of occurrence, again highlighting the temperature-induced shift in stability relative to that observed for AH at T = 0 K in Figure S4. This basin is followed by slightly shallower stability basins for the FFP and TOE configurations, whereas the AH and EAH structures have markedly lower probabilities.

Thus, although the dominant tendency of $Au_{13}^{(+)}$ and $Au_{13}^{(-)}$ clusters to adopt compact 3D and *quasi*-2D configurations, respectively, inferred from T = 0 K calculations is largely preserved, thermal effects in the $Au_{13}^{(+,0,-)}$ systems promote enhanced stabilization, and consequently a higher probability, of 2D configurations. Nevertheless, the influence of the charge state remains clearly discernible: for $Au_{13}^{(+)}$ and $Au_{13}^{(0)}$, central regions associated with 3D configurations still host significant probability basins, whereas for $Au_{13}^{(-)}$, the probability is predominantly displaced toward outer regions associated with 2D and *quasi*-2D motifs. Therefore, these findings highlight the strong interplay between charge and temperature in determining the configurational landscape of Au_{13} clusters.

3.4. Probability Surface Reproducibility

Figure 5 presents the probability surfaces obtained from the MBAR results projected onto the FIt-SNE coordinates, for which the main observations remain consistent with those obtained from the MBAR+sketch-map protocol. One observes that MBAR+FIt-SNE consistently identifies the high-probability configurations associated with the outer 2D subregion, while also distinguishing the basins related to the fluxionality of the central 3D configurations for $Au_{13}^{(+)}$ and $Au_{13}^{(0)}$, and the probability migration toward the outer *quasi*-2D subregion in the case of $Au_{13}^{(-)}$.

In contrast, the MBAR+UMAP results, shown in Figure 6, do not reproduce the probability surfaces obtained by the MBAR+sketch-map and MBAR+FIt-SNE protocols. Although UMAP efficiently separates regions associated with 3D, *quasi*-2D, and 2D configurations qualitatively, its embeddings are more diffuse and less suitable for constructing the compact low-dimensional regions required for stability basin identification. More specifically, 2D configurations, although structurally identified as a distinct class, appear to be much more dispersed in the projected space. This behavior does not indicate that UMAP fails to distinguish the structural families; rather, it indicates that the UMAP coordinates impose a different geometric organization on the sampled points, which affects the construction of probability basins when these coordinates are used directly as collective variables for the MBAR reweighting.

Insufficiently compact clustering, which is essential for defining stability basins within a multiple-thermodynamic-canonical-ensemble framework, biases estimates of the most probable configurations. As a consequence, although MBAR+UMAP reproduces the main features of the $Au_{13}^{(+)}$ probability surface captured by MBAR+sketch-map and MBAR+FIt-SNE, it produces more diffuse embeddings that hinder the recovery of the 2D-FFP basin for $Au_{13}^{(0)}$ (Figure S3) and the 2D-OFP

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

basin for $Au_{13}^{(-)}$ (Figure 6b). Instead, it assigns the highest probability for $Au_{13}^{(-)}$ to the *quasi*-2D TOE configuration. Notably, these poorly resolved basins correspond to the second-most-stable motifs according to the MBAR+sketch-map and MBAR+FIt-SNE analyses. Thus, even when 2D configurations are lower in free energy than their 3D counterparts, increased sample dispersion within the 2D regions of the UMAP embedding prevents these motifs from being resolved as high-probability basins.

Overall, constructing reliable probability surfaces depends not only on the structural separation achieved by the dimensionality reduction method but also on its ability to generate sufficiently compact regions that enable effective clustering of the sampled configurational space. This observation underscores the need to carefully tune hyperparameters to determine the optimal settings for each NLDR algorithm, as discussed in the following.

3.4.1. Hyperparameter Evaluation. Despite the similar global behavior of the projections in Figure 3, each method separates configurational regions in markedly different ways. The sketch-map projection reproduces the configurational landscape remarkably well, revealing compact, densely populated regions corresponding to specific configurations, connected by sparse transition configurations that surround and link them. Our earlier mention of sketch-map as the most suitable approach stems from the fact that its hyperparameters are directly tied to the pairwise-distance distribution in the high-dimensional space, which the sketch-map code itself can provide. It is important to emphasize that the embeddings were evaluated qualitatively and visually, based on which representation yielded the best physicochemical interpretation, and the hyperparameters were optimized accordingly, given that the analysis relies on the relative positioning of the projected structures.

The parameter σ, which is the most critical hyperparameter, controls the length scale used to assess structural similarity and should be chosen within the range between the Gaussian fluctuations and the peak of the pairwise-distance distribution in the high-dimensional space. Specifically, points separated by distances smaller than σ are considered similar, while those separated by distances larger than σ are considered distinct. As a result, too small values of σ tend to overseparate the NC structures into scattered regions, while too large values fail to distinguish between them, as shown in Figure S5. In addition, the parameter b should be selected such that the sigmoidal function approaches 1 before the distance histogram vanishes, while also satisfying the relation $a/A = d/D$. Figure S6 shows the pairwise-distance distribution for the sampled embedding considered here. Based on these guidelines, we adopted the recommended values as an initial reference. We tuned hyperparameters to improve separation among regions associated with 3D, 2D, and *quasi*-2D configurations. Specifically, we set $\sigma = 1.2$, $A = 10.5$, and $B = b = a = 1.0$.

FIt-SNE and UMAP, in turn, tend to impose defined boundaries in the low-dimensional space, generating delimited regions whose shapes may reflect the structures of the underlying high-dimensional manifold. They differ mainly in how dispersed or clustered the points are within the bounded areas.

FIt-SNE exhibited the widest variability, ranging from highly dispersed embeddings to overclustered representations depending on the chosen parametrization. As in the original t-SNE formulation, the key parameter is `perplexity`, which controls the effective number of nearest neighbors considered in the high-dimensional space and is used to determine the point-specific Gaussian bandwidth σ_i. Larger `perplexity` values lead to broader σ_i, emphasizing more global structures, whereas smaller values focus on local relationships. In our analysis, we started from the FIt-SNE default value of 30 and increased it up to 250. As shown in Figure S7a, the embedding evolves from a highly dispersed configuration to more structured, clustered, and interconnected regions; however, above 200, we observe no significant qualitative changes.

As highlighted in previous studies,[56,58,85] parameters related to the approximation of pairwise interactions, such as `max_inter`, as well as the early exaggeration factor, `early_exag_coeff`, play important roles in the embedding projection. The `max_inter` parameter limits the number of interpolation-mediated interactions used to approximate repulsive forces, thereby controlling the number of pairwise interactions that are explicitly resolved. Too few interactions lead to excessive, artificial clustering, whereas too many interactions increase the computational cost. As shown in Figure S7b, we observe an optimal regime in which connectivity is maximized without losing structural coherence; beyond this point, the embedding begins to dissolve into poorly defined clusters.

In contrast, `early_exag_coeff` governs the balance between global dispersion and cluster separation during the early phase of optimization without substantially altering the overall projection shape. Low values yield diffuse embeddings with poorly defined cluster boundaries, whereas high values compress clusters excessively, reducing intercluster connectivity. As illustrated in Figure S7c, the optimal regime corresponds to well-separated yet interconnected structures. On the basis of these observations, the best compromise was achieved with `perplexity = 250`, `max_inter = 1000`, and `early_exag_coeff = 1000`, which were subsequently adopted for the MBAR calculation.

UMAP, unlike FIt-SNE, demonstrated greater robustness to parameter variations, consistently producing embeddings with a well-defined separation between 2D and 3D regions. First, the `metric` parameter defines the pairwise-distance function in the high-dimensional space. Common choices include the `euclidean` metric, which treats all dimensions as independent and equally scaled, and the `cosine` metric, which measures the angular similarity between the feature vectors. However, these metrics proved inadequate for capturing structural differences among NCs, resulting in scattered, poorly clustered regions, as shown in Figure S8a. To address this limitation, we evaluated the `mahalanobis` distance, which incorporates the full covariance structure of the data and effectively normalizes the correlated and heterogeneously scaled dimensions. This metric proved to be particularly well-suited for the coordination-histogram descriptor, yielding projections with improved separation between 2D and 3D regions.

After computing the distance, UMAP constructs a weighted, fuzzy graph in the high-dimensional space, where the hyperparameters `n_neighbors` and `min_dist` control the connectivity between points. The `n_neighbors` parameter determines the number of nearest neighbors considered for each point, thereby setting the scale of the local connectivity. As shown in Figure S8b, smaller values produce scattered projections, whereas larger values incorpo-

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

rate more global context, better defining 2D and 3D regions, albeit at increased computational cost. Beyond a certain threshold, however, further increases yield no additional improvement and instead displace the 2D regions away from the central manifold.

The `min_dist` parameter controls the minimum distance between points in the low-dimensional embedding, thereby balancing the local compactness and global spread. Larger values produce more diffuse representations that emphasize global structure, whereas smaller values allow tighter clustering and sharper regional definition. Although Figure S8c shows only minimal improvements in cluster compactness as `min_dist` decreases, it shows no significant impact on the global separation among configurational regions. On the basis of these analyses, the UMAP embedding used in the MBAR calculations was generated using `metric = "mahalanobis"`, `n_neighbors = 50`, and `min_dist = 0.001`.

3.5. Structural Classification from NLDR Maps

We have shown that NLDR algorithms can rationalize correlated features from large data sets, such as those generated from RE-BOMD simulations. Thus, sketch-map, and in principle other NLDR approaches, can be employed to connect structural parameters extracted from a large number of Au_{13} configurations under temperature effects, such as *ECN*, d_{av}, and R_g, allowing several descriptors to be correlated simultaneously toward a detailed structural-pattern characterization. Therefore, sketch-map projections can characterize the 2D and 3D behavior within the projected regions, as shown in Figures S9–S11. A more concise representation is shown in Figure 7, in which nearby symbols in the sketch-map projection correspond to structurally similar configurations, while being simultaneously characterized by *ECN* through the color map and by R_g through the symbol size for T = 250 and 840 K for each charge state. This representation can also be linked to electronic descriptors, providing a direct route to connect configurational sampling to catalytic performance indicators.

Based on Figure 7a for $Au_{13}^{(+)}$, one observes that the central region of the projection is predominantly associated with 3D-like structures. In contrast, the outer regions are enriched in 2D-like configurations, as highlighted across the charge-state panels. The central region is significantly populated by the $Au_{13}^{(+)}$ and $Au_{13}^{(0)}$ states up to 684 K. In contrast, beyond this temperature, the population of these charge states becomes increasingly distributed toward the outer regions. However, the 3D regions have characteristic structural-parameter values, which, for all temperatures, are approximately $ECN \gtrsim 4.50$, $d_{av} \gtrsim 2.77$ Å, and $R_g \lesssim 3.00$ Å.

In the outer regions, $Au_{13}^{(-)}$ predominantly populates the region across the entire temperature range, with increasing contributions from the cationic and neutral states above 760 K. In contrast, these outer regions show a broader distribution of structural parameters, allowing subdivision into two distinct subregions. The first subregion, denoted as 2D-like configurations, is characterized by two main basins located in the leftmost outer region of the sketch-map projection, on the right-hand side and bottom of the UMAP projection, and in the upper and bottom regions of the FIt-SNE projection. These basins are associated with characteristic values of *ECN* $\lesssim 3.75$, $d_{av} \lesssim 2.73$ Å, and $R_g \gtrsim 3.40$ Å.

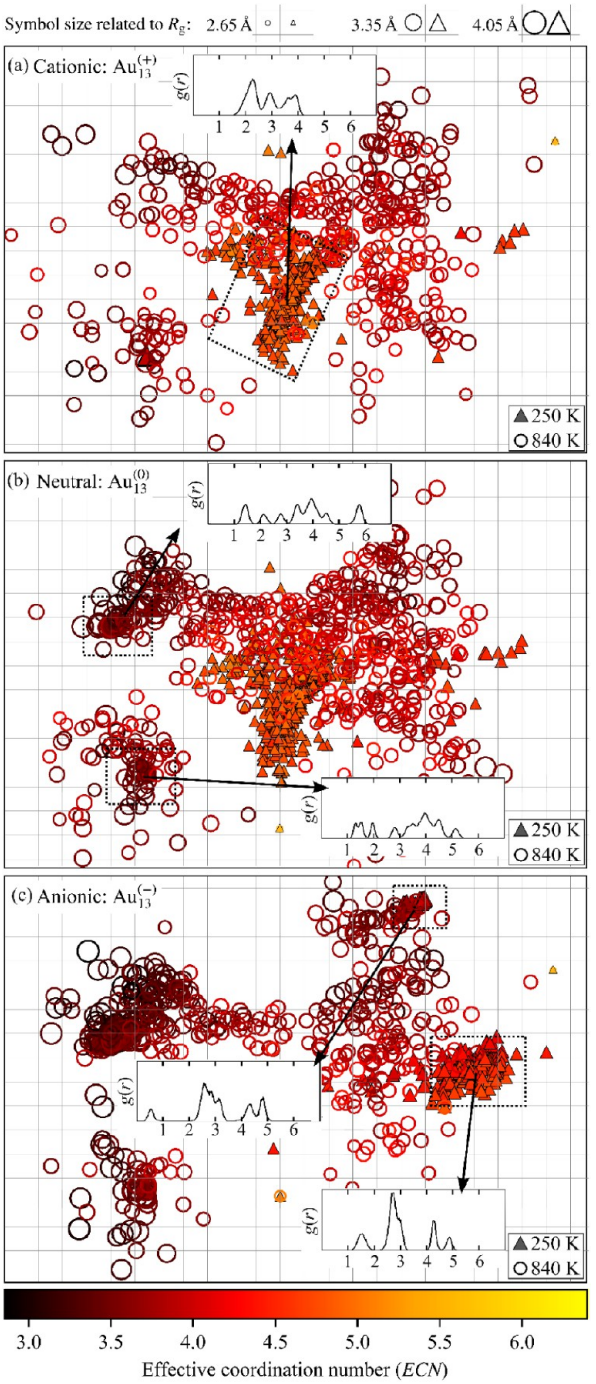


Figure 7. Structural-parameter mapping for (a) $Au_{13}^{(+)}$, (b) $Au_{13}^{(0)}$, and (c) $Au_{13}^{(-)}$ as a function of the sketch-map coordinates at 250 and 840 K. The color map and symbol size represent the effective coordination number (*ECN*) and radius of gyration (R_g), respectively. Triangles and circles correspond to configurations sampled at 250 and 840 K, respectively. Radial distribution profiles, $g(r)$, were also calculated for selected 3D, 2D, and *quasi*-2D regions.

Moving to the second outer subregion, corresponding to structures not classified as purely 2D or 3D, the intermediate character is described as *quasi*-2D configurations by particular structural-parameter values. These structures are characterized by the coexistence of nearly planar motifs and three-dimensional structural features, so that they are clearly distinguishable from both 2D and 3D configurations by

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

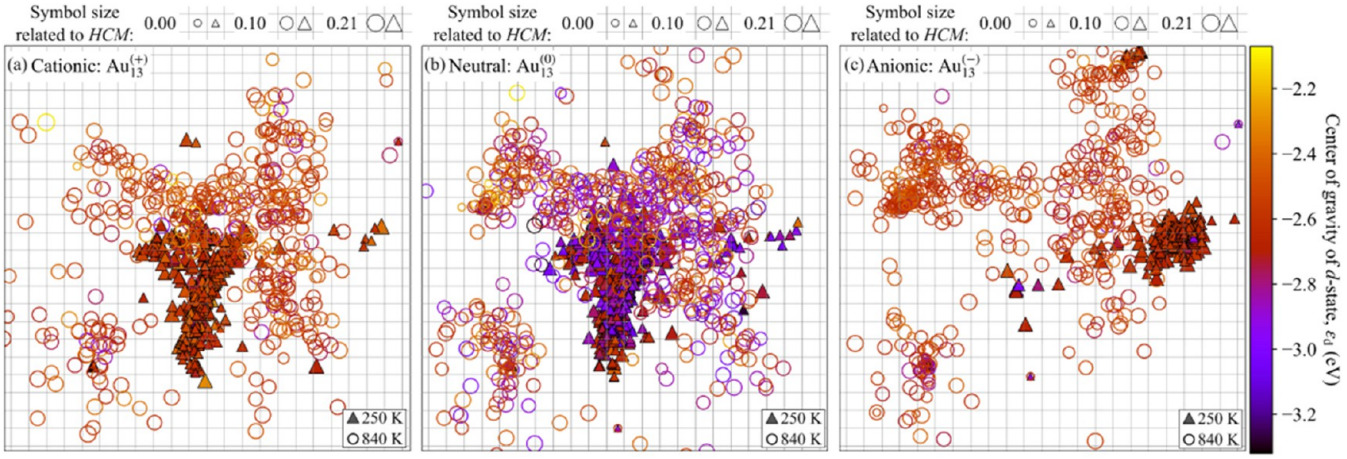


Figure 8. Center of gravity of the occupied *d*-states (ε_d) and Hausdorff chirality measure (*HCM*) mapped onto the sketch-map coordinates for (a) $Au_{13}^{(+)}$, (b) $Au_{13}^{(0)}$, and (c) $Au_{13}^{(-)}$ at 250 and 840 K. The color map and symbol-size range represent ε_d and *HCM*, respectively.

exhibiting intermediate values of *ECN* between 3.75 and 4.50, d_{av} between 2.73 and 2.77 Å, and R_g between 3.10 and 3.40 Å. These thresholds, used to distinguish 2D, *quasi*-2D, and 3D configurations, should be interpreted as operational classification criteria derived from the configurational distributions observed for the present Au_{13} data set, rather than as universal or sharply defined physical boundaries. We introduced these ranges to facilitate qualitative identification of distinct morphological regimes across the configurational landscape.

Particular regions on the projections depicted in Figure 7 were selected to calculate the radial distribution profiles, $g(r)$, as an additional descriptor for the 3D, 2D, and *quasi*-2D characterization. For instance, 3D structures are typically characterized by higher atomic densities at distances around 2.25, 3.00, and 4.00 Å. In contrast, 2D configurations exhibit a broader spread of atomic density toward both shorter distances, below 2 Å, and longer distances, above 4 Å. Among the 2D structures, although both selected regions display a peak close to 1.50 Å, they can be distinguished by a shift of the last peak from 5.80 Å to 5.10 Å when moving from the top to the bottom of the 2D region in the projection.

Furthermore, this approach yields a *quasi*-2D profile with an intermediate $g(r)$ behavior between 3D and 2D configurations. These structures show atomic densities extending beyond typical 3D limits while maintaining a considerably narrower peak distribution than that observed for 2D structures. Overall, the $g(r)$ profiles of the *quasi*-2D configurations are characterized by four main peaks, with those located at 3.00, 4.25, and 4.90 Å being comparable in position. The most pronounced difference arises from a shift of the shortest-distance peak from 1.50 Å to 0.50 Å when moving from the right to the top *quasi*-2D region of the projection.

We have demonstrated the efficiency of this analysis protocol for population characterization based on structural parameters associated with NLDR, which is generally independent of single-temperature sampling. All $g(r)$ analyses performed for the full trajectories of each charge state and temperature are shown in Figure S12. At 250 K, a clear distinction in the $g(r)$ profiles is observed among the different charge states. For $Au_{13}^{(+)}$, three main peaks located at approximately 2.25, 3.00, and 3.75 Å are identified, which are characteristic of 3D configurations. $Au_{13}^{(0)}$ exhibits the same dominant peaks, together with additional minor features at 1.25, 1.50, 4.50, and 5.25 Å, indicating a largely 3D structural sampling with emerging contributions from 2D-like motifs. In contrast, $Au_{13}^{(-)}$ displays a markedly different $g(r)$ profile, with main peaks at 1.50, 2.75, and 4.25 Å, as well as secondary peaks at 2.25, 3.40, 3.90, 4.90, and 5.75 Å. This distribution is characteristic of 2D and *quasi*-2D configurations and reflects the loss of compact 3D structural signatures. As the temperature increases, these distinctive features in $g(r)$ are progressively smoothed out, such that at 840 K, no well-defined peaks remain, indicating a broad structural sampling spanning both central and outer regions of the configurational space.

3.6. Catalytic-Relevant Electronic and Chirality Descriptors from NLDR Maps

Advancing toward catalytic and chirality-related perspectives based on NLDR protocols, we evaluated the center of gravity of the occupied *d*-states (ε_d) as a descriptor of the catalytic potential of NCs, as introduced in Section 2.5. Figure 8 summarizes these results for $Au_{13}^{(+,0,-)}$ at 250 and 840 K, in which ε_d and the Hausdorff chirality measure (*HCM*) are correlated on the sketch-map projections. Figures S13−S14 provide more detailed representations for each charge state and temperature.

Two main effects emerge from the ε_d analysis, both directly associated with the influence of charge state and NC configuration. First, ε_d undergoes a systematic downshift along the sequence $Au_{13}^{(+)} \rightarrow Au_{13}^{(0)} \rightarrow Au_{13}^{(-)}$ for structurally similar configurations. The cationic NCs exhibit ε_d values typically in the range of −2.6 to −2.8 eV, the neutral NCs around −2.8 to −2.9 eV, and the anionic NCs generally at lower energies. However, low-coordinated 2D motifs can partially shift ε_d back toward the Fermi level. This trend reflects a charge-induced redistribution of the *d*-state electron population: in the cationic state, electron removal increases the number of unoccupied states. It shifts ε_d toward the Fermi level, whereas electron addition in the anionic state produces the opposite effect. Consequently, for comparable structural configurations, the NCs become progressively less reactive as the charge state varies from cationic to anionic.

Because the simulations were performed within the Born−Oppenheimer approximation, temperature is not assigned to the electronic degrees of freedom. In this context, temperature

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

serves to enhance the exploration of the nuclear configurational space, while changes in ε_d arise indirectly from the different geometries sampled along the trajectories.

In this context, we show that for the $Au_{13}^{(+)}$, $Au_{13}^{(0)}$, and $Au_{13}^{(-)}$ charge states, ε_d shifts along the structural sequence 3D → *quasi*-2D → 2D, which can be interpreted as a progressive increase in reactivity associated with lower-coordinated motifs. For $Au_{13}^{(+)}$, the central region associated with 3D configurations exhibits ε_d values around −2.6 eV, whereas configurations sampled in the outer regions reach values close to −2.4 eV. In the neutral case, $Au_{13}^{(0)}$, the 3D configurations in the central region display ε_d values of approximately −2.8 eV, while the 2D configurations exhibit values above −2.6 eV. For $Au_{13}^{(-)}$, the outer regions dominated by *quasi*-2D structures tend toward ε_d values near −3.0 eV, whereas the more open 2D configurations show values shifted upward relative to the *quasi*-2D region. These findings further support the notion that lower-coordinated sites are intrinsically more reactive, since the fraction of such sites increases along the 3D → *quasi*-2D → 2D structural sequence.

Another relevant catalytic perspective is to assess the chirality of the $Au_{13}^{(+,0,-)}$ clusters, quantified here using *HCM*, to probe the emergence of chiral sites under temperature effects. Figure 8 reports *HCM* mapped onto the sketch-map coordinates. Although no significant chirality is observed at T = 0 K, except for weak distortions in specific optimized structures, the configurations associated with the 3D stability basins begin to exhibit *HCM* values around 0.100 at 250 K, indicating that temperature promotes distortions that can generate chiral sites. At this temperature, *quasi*-2D configurations show similar behavior, with the regions associated with the EAH, TOE, and AH motifs displaying *HCM* values of approximately 0.200, 0.125, and 0.075, respectively.

Increasing temperature enhances structural distortions and gradually increases chirality for 3D configurations up to 548 K, although above this temperature, sampling within the 3D region becomes less frequent. In contrast, the stability basins associated with strictly 2D configurations present negligible chirality, with *HCM* ≈0.000. Even at elevated temperatures, *HCM* remains low, up to approximately 0.050, within the OFP and FFP basins. However, configurations surrounding these basins can reach *HCM* values as high as 0.200, consistent with strongly distorted structures.

Overall, these results suggest that 3D and *quasi*-2D configurations are more likely to provide chiral active sites. In contrast, strictly 2D configurations are comparatively less prone to developing chirality under the conditions investigated. Therefore, combining NLDR mapping with structural, electronic, and chirality descriptors provides a compact framework for connecting configurational populations to catalytic descriptors in fluxional Au_{13} nanoclusters.

4. CONCLUSION

We have established a general framework for decoding the configurational landscapes of metallic $Au_{13}^{(+,0,-)}$ nanoclusters (NCs) under the combined effects of charge state and temperature. By integrating density functional theory-based replica-exchange Born−Oppenheimer molecular dynamics (RE-BOMD) with nonlinear dimensionality reduction techniques, such as sketch-map, t-SNE/FIt-SNE, and UMAP, and Boltzmann reweighting through the multistate Bennett acceptance ratio (MBAR), we demonstrate that the complex high-dimensional configurational space generated by different charge and temperature conditions can be unraveled by projections onto low-dimensional manifolds that retain thermodynamically meaningful information.

Within this framework, MBAR+sketch-map emerges as the most suitable approach for constructing probability surfaces, owing to its ability to preserve basin connectivity and enable effective clustering of configurations in the context of a multiple-thermodynamic canonical ensemble. While FIt-SNE, as a t-SNE-based algorithm, reproduces the main thermodynamic features with good consistency, UMAP, despite its efficiency in separating structurally distinct regions, produces more diffuse embeddings that are unsatisfactory for the construction of compact low-dimensional representations used in the subsequent free-energy analysis. In this sense, we note that, when the purpose of applying NLDR extends beyond separating the configurations sampled during the dynamics (a task that all methods employed here accomplish effectively) to the construction of a probability surface, and a given NLDR method yields an excessively diffuse projection that hinders the clustering required to resolve well-defined basins even with hyperparameter tuning, we recommend comparing the resulting embedding against that of an NLDR method with more clustering features, such as sketch-map. These results highlight that the choice of dimensionality reduction method, even when it achieves similar structural separation among 3D, 2D, and *quasi*-2D regions, can affect the thermodynamic interpretation of configurational landscapes.

Our results reveal that temperature and charge state act cooperatively to modulate the probability landscape of the Au_{13} NCs. Although compact 3D configurations are strongly favored in the cationic state and remain highly competitive in the neutral state at T = 0 K, thermal effects increase the population of 2D and *quasi*-2D configurations across charge states. The charge state controls the redistribution of probability among the configurational basins: the cationic state preferentially stabilizes central 3D basins, reflecting enhanced structural compactness and fluxionality; the neutral state exhibits a more balanced competition between 3D and 2D basins; and the anionic state promotes more open configurations, characterized by dominant 2D and *quasi*-2D basins. Structural descriptors, radial distribution functions, and MBAR-derived probability analyses consistently supported these findings.

Beyond structural stability, the proposed framework directly connects configurational sampling to physicochemical properties. In particular, correlations among local coordination, chirality, and the center of gravity of the occupied *d*-states reveal that lower-coordinated motifs, especially 2D configurations, exhibit enhanced electronic reactivity, whereas temperature-induced distortions in 3D and *quasi*-2D configurations promote the emergence of chiral features. For structurally comparable configurations, the occupied *d*-state center exhibits a systematic charge-state dependence, with the cationic, neutral, and anionic clusters showing progressively ε_d positions relative to the HOMO. Within the *d*-band-center framework, these shifts may indicate different tendencies for cluster-adsorbate interactions, but they should not be interpreted as a direct prediction of catalytic activity. These effects highlight the critical role of charge and temperature in

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX

defining configurational diversity, electronic response, and catalytic functionality.

This work provides a unified strategy for integrating atomistic simulations, machine learning, and statistical thermodynamics and, in this case, applies it to study nanoscale systems. In this sense, our approach is complementary to recent autonomous-simulation frameworks such as GENIUS,[98] which uses agentic AI, knowledge-graph guidance, and automated error recovery to generate and validate electronic-structure simulation protocols. While GENIUS addresses the autonomous design and execution of reliable DFT workflows, the present framework focuses on decoding, classifying, and thermodynamically interpreting the high-dimensional atomistic data generated by such simulations. Together, these developments point toward an integrated materials-discovery paradigm in which simulations can be generated, validated, executed, analyzed, and interpreted within increasingly autonomous and reproducible workflows. The methodology proposed here is readily transferable to other classes of clusters and NCs, offering a powerful route to rationalizing structure−property relationships and guiding the design of nanocatalysts under realistic operating conditions.

■ ASSOCIATED CONTENT

Data Availability Statement

The authors confirm that the code and data supporting the study's findings are available in the article's GitHub repository: https://github.com/Celso0408/Decoding-Au-clusters.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.6c05562.

Additional information on nanocluster transitions, atomic charge distributions, probability surface of neutral Au_{13}, hyperparameter tests for nonlinear dimensionality reduction methods, and temperature- and charge-state-resolved sketch-map projections, together with structural parameters and the center of gravity of the *d*-states (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Gabriel Reynald Da Silva − *Quantum Chemistry and Thermodynamic Materials Group, Department of Chemistry, Federal University of Paraná, Curitiba 81531-980, Brazil*; Email: gabriel.reynald@ufpr.br

Diego Guedes-Sobrinho − *Quantum Chemistry and Thermodynamic Materials Group, Department of Chemistry, Federal University of Paraná, Curitiba 81531-980, Brazil*; orcid.org/0000-0002-3313-2822; Email: guedes.sobrinho.d@gmail.com

Celso R. C. Rego − *Institute of Nanotechnology Hermann-von-Helmholtz-Platz, Karlsruhe Institute of Technology, Karlsruhe 76021, Germany*; orcid.org/0000-0003-1861-2438; Email: celso.rego@kit.edu

Authors

Mousa Salah − *Kiel University, Kiel 24118, Germany*

Milene M. Machado − *Quantum Chemistry and Thermodynamic Materials Group, Department of Chemistry, Federal University of Paraná, Curitiba 81531-980, Brazil*

Alexandre C. Dias − *Institute of Physics and International Center of Physics, University of Brasilia, Brasilia 70910-900, Brazil*

Carlos M. O. Bastos − *Institute of Physics and International Center of Physics, University of Brasilia, Brasilia 70910-900, Brazil*

Mauricio J. Piotrowski − *Department of Physics, Federal University of Pelotas, Pelotas 96010-900, Brazil*; orcid.org/0000-0003-3477-4437

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.6c05562

Author Contributions

D.G.-S., C.R.C.R., and G.R.S. conceptualized this research and guided the data analysis/interpretation and the manuscript writing. D.G.-S., M.M.M., and G.R.S. performed the calculations and the preliminary formal analysis, as well as the manuscript writing. M.M., M.J.P., A.C.D., and C.M.O.B. performed formal analysis and developed the DymRedPy platform and its documentation. D.G.-S. and G.R.S. reviewed earlier drafts to produce the final version of the paper. All authors contributed to additional calculations, writing, data analysis, and manuscript revision.

Funding

The Article Processing Charge for the publication of this research was funded by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), Brazil (ROR identifier: 00x0ma614).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are thankful for financial support from the National Council for Scientific and Technological Development (CNPq, grant numbers 409792/2024-1, 303206/2025-0, 444431/2024-1, 408144/2022-0, 305174/2023-1, and 444069/2024-0), the Rio Grande do Sul Research Foundation (FAPERGS, grant number 24/2551-0001551-5), the Federal District Research Support Foundation (FAPDF, grants 00193-00001817/2023-43 and 00193-00002073/2023-84), PDPG-FAPDF-CAPES Centro-Oeste (grant number 00193-00000867/2024-94), and the Coordination for Improvement of Higher Level Education (CAPES). In addition, the authors thank the "Centro Nacional de Processamento de Alto Desempenho em São Paulo" (CENAPAD-SP, UNICAMP/FINEP—MCTI project) for resources into the 897 and 570 projects, Lobo Carneiro HPC (NACAD) at the Federal University of Rio de Janeiro (UFRJ) for resources (133 project), CIMATEC SENAI at Salvador, BA, Brazil, for the partnership, support through the Ogun Supercomputer, and "Laboratório Central de Processamento de Alto Desempenho" (LCPAD) financed by FINEP through CT-INFRA/UFPR projects and Santos Dumont/LNCC. The project (HGF ZT-I-PF-5-261 GENIUS) underlying this publication was funded by the Initiative and Networking Fund of the Helmholtz Association in the framework of the Helmholtz AI project call.

■ REFERENCES

(1) Palheta, J. M. T.; Filho, O. R.; Soleymanibrojeni, M.; Dias, A. C.; Guedes-Sobrinho, D.; Wenzel, W.; Aydin, R.; Rêgo, C. R.; Piotrowski,

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

M. J. Teaching a transformer to think like a chemist: predicting nanocluster stability. *Npj Comput. Mater.* **2026**, *12*, 264.
(2) Morais, W. O.; Felix, J. P. C.; Silva, G. R. D.; Bastos, C. M. D. O.; Dias, A. C.; Flores, E. M.; Rêgo, C. R. C.; Sousa, V. D. S. R. D.; Guedes-Sobrinho, D.; Piotrowski, M. J. Understanding stability and reactivity of transition metal single-atoms on graphene. *Sci. Rep.* **2025**, *15*, 15496.
(3) Rêgo, C. R. C.; Tereshchuk, P.; Oliveira, L. N.; Da Silva, J. L. F. Graphene-supported small transition-metal clusters: A density functional theory investigation within van der Waals corrections. *Phys. Rev. B* **2017**, *95*, 235422.
(4) Monteiro, R. F.; Palheta, J. M. T.; Grison, T. G.; Filho, O. R.; Parreira, R. L. T.; Guedes-Sobrinho, D.; Rêgo, C. R. C.; Dias, A. C.; Batista, K. E. D. A.; Piotrowski, M. J. Interpretable, physics-informed learning reveals sulfur adsorption and poisoning mechanisms in 13-atom icosahedra nanoclusters. *Sci. Rep.* **2026**, *16*, 14174.
(5) Li, C.; Chai, O. J. H.; Yao, Q.; Liu, Z.; Wang, L.; Wang, H.; Xie, J. Electrocatalysis of gold-based nanoparticles and nanoclusters. *Mater. Horiz.* **2021**, *8*, 1657−1682.
(6) Guan, Z.-J.; Li, J.-J.; Hu, F.; Wang, Q.-M. Structural engineering toward gold nanocluster catalysis. *Angew. Chem.* **2022**, *134*, No. e202209725.
(7) Li, S.; Du, X.; Liu, Z.; Li, Y.; Shao, Y.; Jin, R. Size effects of atomically precise gold nanoclusters in catalysis. *Precis. Chem.* **2023**, *1*, 14−28.
(8) Dong, J.; Li, Y.; Ding, Y.; Su, H.-F.; Cui, X.; Wang, Y.-X.; Li, H. Electrosynthesis of atomically precise Au nanoclusters. *Adv. Sci.* **2025**, *12*, 2414057.
(9) Zhao, S.; Jin, R.; Abroshan, H.; Zeng, C.; Zhang, H.; House, S. D.; Gottlieb, E.; Kim, H. J.; Yang, J. C.; Jin, R. Gold nanoclusters promote electrocatalytic water oxidation at the nanocluster/$CoSe_2$ interface. *J. Am. Chem. Soc.* **2017**, *139*, 1077−1080.
(10) Zhao, S.; Jin, R.; Song, Y.; Zhang, H.; House, S. D.; Yang, J. C.; Jin, R. Atomically precise gold nanoclusters accelerate hydrogen evolution over MoS_2 nanosheets: the dual interfacial effect. *Small* **2017**, *13*, 1701519.
(11) Negishi, Y.; Mizuno, M.; Hirayama, M.; Omatoi, M.; Takayama, T.; Iwase, A.; Kudo, A. Enhanced photocatalytic water splitting by $BaLa_4Ti_4O_{15}$ loaded with 1nm gold nanoclusters using glutathione-protected Au_{25} clusters. *Nanoscale* **2013**, *5*, 7188−7192.
(12) Yan, L.; Wang, F.; Meng, S. Quantum mode selectivity of plasmon-induced water splitting on gold nanoparticles. *ACS Nano* **2016**, *10*, 5452−5458.
(13) Feng, Q.; Zhang, Y.; Feng, H.; Liu, D.; Li, Q. Photocatalytic water splitting in the gap between plasmonic gold nanoclusters. *Phys. Fluids* **2023**, *35*, 073334.
(14) Bera, D.; Mahata, S.; Biswas, M.; Kumari, K.; Rakshit, S.; Nonappa Ghosh, N.; Ghosh, S.; Goswami, N. Efficient Photocatalytic Hydrogen Production Using *In-Situ* Polymerized Gold Nanocluster Assemblies. *Small* **2025**, *21*, 2406551.
(15) Zhu, Y.; Qiu, X.; Zhao, S.; Guo, J.; Zhang, X.; Zhao, W.; Shi, Y.; Tang, Z. Structure regulated catalytic performance of gold nanocluster-MOF nanocomposites. *Nano Res.* **2020**, *13*, 1928−1932.
(16) Moon, C. W.; Choi, M.-J.; Hyun, J. K.; Jang, H. W. Enhancing photoelectrochemical water splitting with plasmonic Au nanoparticles. *Nanoscale Adv.* **2021**, *3*, 5981−6006.
(17) Kawawaki, T.; Kataoka, Y.; Hirata, M.; Akinaga, Y.; Takahata, R.; Wakamatsu, K.; Fujiki, Y.; Kataoka, M.; Kikkawa, S.; Alotabi, A. S.; et al. Creation of High-Performance Heterogeneous Photocatalysts by Controlling Ligand Desorption and Particle Size of Gold Nanocluster. *Angew. Chem., Int. Ed.* **2021**, *60*, 21340−21350.
(18) Mousavi, H.; Yin, Y.; Sharma, S. K.; Gibson, C. T.; Golovko, V.; Andersson, G. G.; Shearer, C. J.; Metha, G. F. Factors influencing catalytic activity of size-specific triphenylphosphine-ligated gold nanoclusters in the electrocatalytic hydrogen evolution reaction. *J. Phys. Chem. C* **2022**, *126*, 246−260.
(19) Gao, Z.-H.; Wei, K.; Wu, T.; Dong, J.; Jiang, D.-E.; Sun, S.; Wang, L.-S. A heteroleptic gold hydride nanocluster for efficient and selective electrocatalytic reduction of CO_2 to CO. *J. Am. Chem. Soc.* **2022**, *144*, 5258−5262.
(20) Wan, X.-K.; Liu, T.-T.; Li, N.-L.; Dai, Q.; Wei, J.; Wang, Q.-M. Tailoring atomically precise gold nanoclusters for boosting selective hydrogenation of Nitrostyrene with H_2. *ACS Nano* **2025**, *19*, 11371−11380.
(21) Gupta, R. K.; Wang, Z.; Mohan, B.; Tung, C.-H.; Sun, D. Advancements in Atomically Precise Nanocluster Protected by Thiacalix[4]arene. *Adv. Mater.* **2024**, *36*, 2410054.
(22) Bralick, A. K.; Walker, M. M.; Anderson, I. D.; Tofanelli, M. A.; Ackerson, C. J. Toward Regiochemical Control of Ligands on Thiol-Protected Gold Clusters. *J. Phys. Chem. C* **2025**, *129*, 18380−18390.
(23) de Jong, J. A.; Workentin, M. S.; Hesari, M. Charge State Tuning and Photochemical Stability of $Au_{25}(SR)_{18}$ Nanoclusters. *ChemPhotoChem* **2025**, *9*, No. e202400236.
(24) Kong, J.; Huo, D.; Jie, J.; Wu, Y.; Wan, Y.; Song, Y.; Zhou, M. Effect of single electrons on the excited state dynamics of rod-shaped Au_{25} nanoclusters. *Nanoscale* **2021**, *13*, 19438−19445.
(25) Zhai, H.; Alexandrova, A. N. Fluxionality of catalytic clusters: When it matters and how to address it. *ACS Catal.* **2017**, *7*, 1905−1911.
(26) Zhai, H.; Alexandrova, A. N. Local fluxionality of surface-deposited cluster catalysts: the case of Pt_7 on Al_2O_3. *J. Phys. Chem. Lett.* **2018**, *9*, 1696−1702.
(27) McCandler, C. A.; Sanwal, D.; Rogal, J. Dynamic Ensembles of Phosphine-Stabilized Gold Nanoclusters. *arXiv*, 2026, .
(28) Zhai, X.-J.; Luo, M.-Y.; Luo, X.-M.; Cheng, D.; Guo, M.; Sun, K.-X.; Si, Y.; Han, Z.; Huang, J.-H.; Zang, S.-Q.; et al. Visualization of Total Synthesis from Au_{11} to Au_{13} Nanoclusters. *J. Am. Chem. Soc.* **2025**, *147*, 32560−32570.
(29) Lethbridge, S.; Pavloudis, T.; McCormack, J.; Slater, T.; Kioseoglou, J.; Palmer, R. E. Stabilization of 2D raft structures of Au nanoclusters with up to 60 atoms by a carbon support. *Small Sci.* **2024**, *4*, 2400093.
(30) Zhang, C.; Yoon, B.; Landman, U. Predicted oxidation of CO catalyzed by Au nanoclusters on a thin defect-free MgO film supported on a Mo(100) surface. *J. Am. Chem. Soc.* **2007**, *129*, 2228−2229.
(31) Albright, E. L.; Malola, S.; Jacob, S. I.; Yi, H.; Takano, S.; Mimura, K.; Tsukuda, T.; Hakkinen, H.; Nambo, M.; Crudden, C. M. Enantiopure chiral Au_{13} nanoclusters stabilized by ditopic *N*-heterocyclic carbenes: synthesis, characterization, and electrocatalytic reduction of CO_2. *Chem. Mater.* **2024**, *36*, 1279−1289.
(32) Yang, Y.; Zhang, Q.; Guan, Z.-J.; Nan, Z.-A.; Wang, J.-Q.; Jia, T.; Zhan, W.-W. Enantioselective synthesis of homochiral Au_{13} nanoclusters and their chiroptical activities. *Inorg. Chem.* **2019**, *58*, 3670−3675.
(33) Deng, G.; Malola, S.; Yan, J.; Han, Y.; Yuan, P.; Zhao, C.; Yuan, X.; Lin, S.; Tang, Z.; Teo, B. K.; et al. From symmetry breaking to unraveling the origin of the chirality of ligated $Au_{13}Cu_2$ nanoclusters. *Angew. Chem., Int. Ed.* **2018**, *57*, 3421−3425.
(34) Wang, Y.-G.; Yoon, Y.; Glezakou, V.-A.; Li, J.; Rousseau, R. The role of reducible oxide−metal cluster charge transfer in catalytic processes: new insights on the catalytic mechanism of CO oxidation on Au/TiO2 from *ab initio* molecular dynamics. *J. Am. Chem. Soc.* **2013**, *135*, 10673−10683.
(35) Weiher, N.; Beesley, A. M.; Tsapatsaris, N.; Delannoy, L.; Louis, C.; van Bokhoven, J. A.; Schroeder, S. L. Activation of oxygen by metallic gold in Au/TiO2 catalysts. *J. Am. Chem. Soc.* **2007**, *129*, 2240−2241.
(36) Schebarchov, D.; Baletto, F.; Wales, D. Structure, thermodynamics, and rearrangement mechanisms in gold clusters−insights from the energy landscapes framework. *Nanoscale* **2018**, *10*, 2004−2016.
(37) Settem, M.; Ferrando, R.; Giacomello, A. Tempering of Au nanoclusters: capturing the temperature-dependent competition among structural motifs. *Nanoscale* **2022**, *14*, 939−952.
(38) Nakayama, Y.; Kera, S.; Ueno, N. Photoelectron spectroscopy on single crystals of organic semiconductors: Experimental electronic

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

band structure for optoelectronic properties. *J. Mater. Chem. C* **2020**, *8*, 9090−9132.
(39) Yubero, F.; Tőkési, K. Identification of hydrogen and deuterium at the surface of water ice by reflection electron energy loss spectroscopy. *Appl. Phys. Lett.* **2009**, *95*, 084101.
(40) Arantes, C.; Scholz, M.; Schmidt, R.; Dehm, V.; Rocco, M.; Schöll, A.; Reinert, F.; Würthner, F. Comparative analysis of the energy levels of planar and core-twisted perylene bisimides in solution and solid state by UV/VIS, CV, and UPS/IPES. *Appl. Phys. A: Mater. Sci. Process.* **2012**, *108*, 629−637.
(41) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.
(42) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133.
(43) Häkkinen, H.; Yoon, B.; Landman, U.; Li, X.; Zhai, H.-J.; Wang, L.-S. On the electronic and atomic structures of small Au_N^-(N = 4−14) clusters: a photoelectron spectroscopy and density-functional study. *J. Phys. Chem. A* **2003**, *107*, 6168−6175.
(44) Xing, X.; Yoon, B.; Landman, U.; Parks, J. H. Structural evolution of Au nanoclusters: From planar to cage to tubular motifs. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2006**, *74*, 165423.
(45) Lechtken, A.; Schooss, D.; Kappes, M. M.; Furche, F.; Johansson, M. P. 2D-3D transition of gold cluster anions resolved. *Phys. Rev. A* **2008**, *77*, 053202.
(46) Huang, W.; Wang, L.-S. Probing the 2D to 3D structural transition in gold cluster anions using argon tagging. *Phys. Rev. Lett.* **2009**, *102*, 153401.
(47) Khetrapal, N. S.; Bulusu, S. S.; Zeng, X. C. Structural Evolution of Gold Clusters n = 21−25) Revisited. *J. Phys. Chem. A* **2017**, *121*, 2466−2474.
(48) Lechtken, A.; Neiss, C.; Kappes, M. M.; Schooss, D. Structure determination of gold clusters by trapped ion electron diffraction: Au_{14}−Au_{19}-. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4344−4350.
(49) Guedes-Sobrinho, D.; Wang, W.; Hamilton, I. P.; Da Silva, J. L.; Ghiringhelli, L. M. (Meta-) stability and Core−Shell Dynamics of Gold Nanoclusters at Finite Temperature. *J. Phys. Chem. Lett.* **2019**, *10*, 685−692.
(50) Yan, G.; Vlachos, D. G. Unraveling Multiscale Kinetics over Subnanometer Cluster Catalysts: H_2 Desorption from $Pt_3(\text{-H})_2/\gamma$-Al_2O_3 (110). *ACS Catal.* **2023**, *13*, 10602−10614.
(51) Poths, P.; Lai, K. C.; Scheurer, C.; Matera, S.; Reuter, K. Kinetic restructuring of catalyst active sites: a MACE-APE study of fluxional Pd_n/MgO (n = 3−11) clusters. *Faraday Discuss* **2026**, *267*, 278−297.
(52) Poths, P.; Vargas, S.; Sautet, P.; Alexandrova, A. N. Thermodynamic equilibrium versus kinetic trapping: thermalization of cluster catalyst ensembles can extend beyond reaction time scales. *ACS Catal.* **2024**, *14*, 5403−5415.
(53) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13023−13028.
(54) Ceriotti, M.; Tribello, G. A.; Parrinello, M. Demonstrating the transferability and the descriptive power of sketch-map. *J. Chem. Theory Comput.* **2013**, *9*, 1521−1532.
(55) van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579−2605.
(56) Sun, Y.; Han, Y.; Fan, J. Laplacian-based cluster-contractive t-SNE for high-dimensional data visualization. *ACM Trans. Knowl. Discovery Data* **2024**, *18*, 1−22.
(57) de Mendonca, J. P. A.; Calderan, F. V.; Lourenco, T. C.; Quiles, M. G.; Da Silva, J. L. Theoretical framework based on molecular dynamics and data mining analyses for the study of potential energy surfaces of finite-size particles. *J. Chem. Inf. Model.* **2022**, *62*, 5503−5512.
(58) Belkina, A. C.; Ciccolella, C. O.; Anno, R.; Halpert, R.; Spidlen, J.; Snyder-Cappione, J. E. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nat. Commun.* **2019**, *10*, 5415.
(59) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*, 2018, .
(60) Trozzi, F.; Wang, X.; Tao, P. UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: a comparison study. *J. Phys. Chem. B* **2021**, *125*, 5022−5034.
(61) Dorrity, M. W.; Saunders, L. M.; Queitsch, C.; Fields, S.; Trapnell, C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat. Commun.* **2020**, *11*, 1537.
(62) Becht, E.; McInnes, L.; Healy, J.; Dutertre, C.-A.; Kwok, I. W.; Ng, L. G.; Ginhoux, F.; Newell, E. W. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **2019**, *37*, 38−44.
(63) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(64) Li, X.-Y.; Ou, P.; Duan, X.; Ying, L.; Meng, J.; Zhu, B.; Gao, Y. Dynamic active sites *in situ* formed in metal nanoparticle reshaping under reaction conditions. *JACS Au* **2024**, *4*, 1892−1900.
(65) Tateishi, Y.; Botha, L. M.; Kozhukhova, A. E.; Sugimoto, M.; Aika, K.-I.; Bessarabov, D. G. Molecular dynamics study of catalytic H_2/O_2 recombination on Pd, Pt, Cu, Ag, and Au nanoclusters using the universal neural network potential. *Chem. Sci.* **2026**, *17*, 5542−5553.
(66) Ha, D.; Schmidhuber, J. World Models. *arXiv*, 2018, .
(67) LeCun, Y. *A Path Towards Autonomous Machine Intelligence, Version 0.9.2*. 2022; https://openreview.net/pdf?id=BZ5a1r-kVsf. (accessed 2026−09−21).
(68) Assran, M.; Duval, Q.; Misra, I.; Bojanowski, P.; Vincent, P.; Rabbat, M.; LeCun, Y.; Ballas, N.Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*202315619−15629
(69) Mieller, B.; Valavi, M.; Caldeira Rêgo, C. R. An Automatized Simulation Workflow for Powder Pressing Simulations Using SimStack. *Adv. Eng. Mater.* **2025**, *27*, 3.
(70) Rêgo, C. R. C.; Wenzel, W.; Piotrowski, M. J.; Dias, A. C.; Maciel de Oliveira Bastos, C.; de Araujo, L. O.; Guedes-Sobrinho, D. Digital workflow optimization of van der Waals methods for improved halide perovskite solar materials. *Digital Discovery* **2025**, *4*, 927−942.
(71) de Araujo, L. O.; Rêgo, C. R. C.; Wenzel, W.; Piotrowski, M. J.; Dias, A. C.; Guedes-Sobrinho, D. Automated workflow for analyzing thermodynamic stability in polymorphic perovskite alloys. *Npj Comput. Mater.* **2024**, *10*, 2.
(72) Rêgo, C. R. C.; Schaarschmidt, J.; Schlöder, T.; Penaloza-Amion, M.; Bag, S.; Neumann, T.; Strunk, T.; Wenzel, W. SimStack: An Intuitive Workflow Framework. *Front. Mater.* **2022**, *9*, 1.
(73) Bekemeier, S.; Caldeira Rêgo, C. R.; Mai, H. L.; Saikia, U.; Waseda, O.; Apel, M.; Arendt, F.; Aschemann, A.; Bayerlein, B.; Courant, R.; Dziwis, G.; Fuchs, F.; Giese, U.; Junghanns, K.; Kamal, M.; Koschmieder, L.; Leineweber, S.; Luger, M.; Lukas, M.; Maas, J.; Mertens, J.; Mieller, B.; Overmeyer, L.; Pirch, N.; Reimann, J.; Schröck, S.; Schulze, P.; Schuster, J.; Seidel, A.; Shchyglo, O.; Sierka, M.; Silze, F.; Stier, S.; Tegeler, M.; Unger, J. F.; Weber, M.; Hickel, T.; Schaarschmidt, J. Advancing Digital Transformation in Material Science: The Role of Workflows Within the MaterialDigital Initiative. *Adv. Eng. Mater.* **2025**, *27*, 2402149.
(74) Schaarschmidt, J.; Yuan, J.; Strunk, T.; Kondov, I.; Huber, S. P.; Pizzi, G.; Kahle, L.; Bölle, F. T.; Castelli, I. E.; Vegge, T.; Hanke, F.; Hickel, T.; Neugebauer, J.; Rêgo, C. R. C.; Wenzel, W. Workflow Engineering in Materials Design within the BATTERY 2030+ Project. *Adv. Energy Mater.* **2022**, *12*, 2102638.
(75) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Ab initio* molecular simulations with numeric atom-centered orbitals. *Comput. Phys. Commun.* **2009**, *180*, 2175−2196.
(76) Beret, E. C.; Ghiringhelli, L. M.; Scheffler, M. Free gold clusters: beyond the static, monostructure description. *Faraday Discuss.* **2011**, *152*, 153−167.

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

(77) Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. Exchange often and properly in replica exchange molecular dynamics. *J. Chem. Theory Comput.* **2010**, *6*, 2804−2808.
(78) Guedes-Sobrinho, D.; Orenha, R. P.; Parreira, R. L.; Nagurniak, G. R.; Da Silva, G. R.; Piotrowski, M. J. The effect of different energy portions on the 2D/3D stability swapping for 13-atom metal clusters. *Phys. Chem. Chem. Phys.* **2022**, *24*, 6515−6524.
(79) Koskinen, P.; Häkkinen, H.; Huber, B.; von Issendorff, B.; Moseler, M. Liquid-liquid phase coexistence in gold clusters: 2D or not 2D? *Phys. Rev. Lett.* **2007**, *98*, 015701.
(80) Zhang, I. Y.; Ren, X.; Rinke, P.; Blum, V.; Scheffler, M. Numeric atom-centered-orbital basis sets with valence-correlation consistency from H to Ar. *New J. Phys.* **2013**, *15*, 123033.
(81) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.
(82) Tkatchenko, A.; DiStasio, R. A., Jr; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
(83) Van Lenthe, J.; Faas, S.; Snijders, J. Gradients in the *ab initio* scalar zeroth-order regular approximation (ZORA) approach. *Chem. Phys. Lett.* **2000**, *328*, 107−112.
(84) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(85) Linderman, G. C.; Rachh, M.; Hoskins, J. G.; Steinerberger, S.; Kluger, Y. Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **2019**, *16*, 243−245.
(86) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(87) Matsunaga, Y.; Kamiya, M.; Oshima, H.; Jung, J.; Ito, S.; Sugita, Y. Use of multistate Bennett acceptance ratio method for free-energy calculations from enhanced sampling and free-energy perturbation. *Biophys. Rev.* **2022**, *14*, 1503−1512.
(88) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. Equilibrium Free Energies from Nonequilibrium Measurements Using Maximum-Likelihood Methods. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(89) Shirts, M. R.; Ferguson, A. L. Statistically optimal continuous free energy surfaces from biased simulations and multistate reweighting. *J. Chem. Theory Comput.* **2020**, *16*, 4107−4125.
(90) Hammer, B.; Norskov, J. K. Why gold is the noblest of all the metals. *Nature* **1995**, *376*, 238−240.
(91) Xin, H.; Vojvodic, A.; Voss, J.; Nørskov, J. K.; Abild-Pedersen, F. Effects of *d*-band shape on the surface reactivity of transition-metal alloys. *Phys. Rev. B* **2014**, *89*, 115114.
(92) Hirshfeld, F. L. Bonded-atom fragments for describing molecular charge densities. *Theor. Chim. Acta* **1977**, *44*, 129−138.
(93) Piotrowski, M. J.; Piquini, P.; Da Silva, J. L. Density functional theory investigation of 3*d*, 4*d*, and 5*d* 13-atom metal clusters. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81*, 155446.
(94) Mackay, A. L. A dense non-crystallographic packing of equal spheres. *Acta Crystallogr.* **1962**, *15*, 916−918.
(95) Liu, J.-X.; Filot, I. A.; Su, Y.; Zijlstra, B.; Hensen, E. J. Optimum particle size for gold-catalyzed CO oxidation. *J. Phys. Chem. C* **2018**, *122*, 8327−8340.
(96) Rincon, L.; Hasmy, A.; Marquez, M.; Gonzalez, C. A perturbatively corrected tight-binding method with hybridization: application to gold nanoparticles. *Chem. Phys. Lett.* **2011**, *503*, 171−175.
(97) Sun, J.; Xie, X.; Cao, B.; Duan, H. A density-functional theory study of Au_{13}, Pt_{13}, $Au_{12}Pt$ and $Pt_{12}Au$ clusters. *Comput. Theor. Chem.* **2017**, *1107*, 127−135.
(98) Soleymanibrojeni, M.; Aydin, R.; Guedes-Sobrinho, D.; Dias, A. C.; Piotrowski, M. J.; Wenzel, W.; Rêgo, C. R. C. GENIUS: an agentic AI framework for autonomous design and execution of simulation protocols. *Commun. Mater.* **2026**, *7*, 115.

Downloaded from pubs.acs.org/jpcafh/article-pdf/doi/10.1021/acs.jpca.6c05562/6899087/acs.jpca.6c05562.pdf by KIT BIBLIOTHEK user on 09 November 2026

https://doi.org/10.1021/acs.jpca.6c05562
J. Phys. Chem. A XXXX, XXX, XXX−XXX