Nanocluster Structure Prediction

Dong-Gi Kang

A thesis submitted for

the Degree of Doctor of Philosophy

in

Theoretical and Computational Chemistry

Primary Supervisor: Scott M. Woodley

Secondary Supervisor: Prof. Sir C. Richard A. Catlow

Department of Chemistry

University College London

13, November, 2024

Declaration

I, Dong-Gi Kang, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

The atomic structures of nanoclusters are predicted using global optimisation techniques that aim to locate local minima (LM) on the potential energy landscape. This research evaluates the effectiveness of replacing density functional theory (DFT) with standard interatomic potentials (IP), variable charge potentials, machine learning potentials (ML-IP: GAP), and neural network potentials (NN-IP: MACE). It also assesses the newly published "universal potentials," which are a complete set of MACE potentials trained on data from the Materials Project Database, applicable to all elements.

Monte Carlo Deterministic Quenching and an Evolutionary Algorithm are employed to find IP LM of $(AlF_3)_n$ for fixed values of n, from 1 to 11, and bulk cuts for n = 27 and 64. Parameters (A = 3760.0008, ρ = 0.2220) for the Al^{3+}-F^- Born-Mayer potential were fitted to reproduce the AlF_3 α-bulk phase, whilst keeping fixed F^--F^- Buckingham potential parameters, which were taken from earlier work on lattice and intrinsic defect properties of bulk rare-earth fluorides. $(AlF_3)_6$ is the smallest LM nanocluster composed of octahedral corner-sharing secondary building units, a structural feature consistent with bulk AlF_3 phases.

Aluminium hydride nanoclusters were generated by data-mining the aluminium fluoride LM. Notably, aluminium hydride exhibits analogous configurations to aluminium fluoride and demonstrate comparable stability. To analyse the structures in terms of primary or secondary building units requires the ability to first determine their coordination numbers. A method for defining coordination numbers was developed and applied, although in many cases this proved problematic.

The third system was brought to my attention by colleagues within the department: 25 gold atoms capped by 18 *L*-cysteine ligands, i.e., nanoclusters of $Au_{25}(Cys)_{18}$. An earlier study deduced from the XRD observations a more spherical Au_{25} core as compared to the new data recently obtained. Here, a model for this system was developed by first applying global optimisation to determine the atomic structure of naked Au_{25}, to which the *L*-cysteine ligands were added.

Impact statement

This is the first study on the structure prediction of bare AlF_3 and AlH_3 nanoclusters, extended to a *L*-cysteine coated gold, $Au_{25}(Cys)_{18,}$ nanocluster. The nanocluster properties can be tailored to enhance their effectiveness and efficiency in applications, from everyday technologies to industrial processes. Their high surface-to-volume ratio significantly reduces the quantity of material needed if surface interactions are critical. Nanoclusters have a wide range of applications in catalysis, sensing, and biomedical imaging.

Understanding how nanocluster properties change with size, gaining an insight in the relationship between the structure, thermodynamical stability and chemical properties is critical for developing, optimising, and exploiting their use in applications. My research focuses on modelling the atomic structure and properties of nanoclusters. This thesis documents the progress made in finding and applying a suitable "computationally cheap" measure to evaluate candidate structures with the aim of efficiently predicting the atomic structures of nanoclusters. The work follows my MSc project on predicting clusters of Ti_3N_4 where traditional atomistic models fail. In my PhD study, the suitability of variable charge potentials (VCP), developed for bulk Ti_3N_4 systems, were tested for modelling clusters. I found that VCPs are not suitable as cost functions, as multiple local minima were generated for every unique targeted configuration. Consequently, the direction of my study changed to considering the suitability of both more traditional and machine learning (ML) interatomic potentials for clusters with 1:3 stoichiometry. The integration of machine learning with traditional computational methods significantly reduces the computer resources required for structural predictions, making high-precision simulations more accessible and efficient.

Employing global optimisation techniques – Monte Carlo deterministic quenching and an evolutionary algorithm as implemented in the in-house software, Knowledge Led Master Code (KLMC) – stable, low-energy atomic structures of nanoclusters have been predicted for Al compounds that showed analogous configurations and comparable stability. The findings on the Lewis

acidity and catalytic potential of AlF_3 nanoclusters have direct implications for designing and optimising new catalysts. A model for the structure of *L*-cysteine protected, or thiolated gold clusters is generated providing insights into their properties and stability which are vital for their applications. The properties of the thiolated gold clusters are compared with experiment. The approach employed can be adapted to other complex systems, facilitating the discovery of new materials with tailored properties.

This thesis lays a strong foundation for further research into nanocluster structures and their properties. Future work could explore more advanced ML techniques, including non-harmonic mode-coupling effects, to enhance the accuracy of ML interatomic potentials. Experimental validation of the predicted structures and properties will be crucial for translating these insights into practice.

The research contained in this thesis has been communicated to the broad scientific community at several conferences in the UK, Italy and South Africa. Four manuscripts for publication, based on the material included in this thesis and work completed alongside, are in preparation including: structure prediction of AlF_3 clusters; experimental and computational study of $Au_{25}(Cys)_{18}$; Au clusters deposited on reduced graphene oxides; and structure prediction of pyrene polymers.

Acknowledgements

First and foremost, I would like to express my deepest gratitude and love to my wife, Kathleen Drake. Your unwavering patience, support, and encouragement have been my anchor throughout this journey. Thank you for standing by me, postponing our wedding, and providing endless encouragement, allowing me to achieve this milestone. From my foundation year and BSc in Sheffield to my Ph.D. in London, you have patiently supported me over the past ten years. With your unwavering encouragement, I can now conclude my student journey and embark on my career as a Scientist.

I am also profoundly grateful to my parents and brother. Since deciding to study abroad during my national military service, they have continuously supported me. I am proud to show my family that I have reached one of the most significant milestones of my life, and I dedicate this achievement to them.

I would like to sincerely thank my supervisor, Prof. Scott M. Woodley, for providing me with this invaluable opportunity and for his unwavering guidance, insights, and support throughout my Ph.D. journey. His dedication and advice have been instrumental in the completion of my research. Despite the unprecedented challenges posed by the COVID-19 pandemic and the national lockdown that began four months into my Ph.D. and lasted until the end of my second year, he was always ready to offer feedback, guidance, and personal advice.

My sincere thanks also go to my secondary supervisor, Prof. Sir C. Richard A. Catlow, for his guidance and advice. I am also grateful to Dr. Alexey A. Sokol for his advice on both academic and personal matters and for our enriching discussions on various cultures.

I would like to thank my office colleagues: Dr. Woong Kyu Jee, Dr. Isa T. Lough, and Cyril Xu. I greatly enjoyed our time together in the office and the pub, discussing science and sharing memorable moments. I am glad we navigated our postgraduate journey together. Additionally, I extend my gratitude to Dr. Alin-Marin Elena for helping me to hands-on ML-IP, and to

my experimental collaborator Dr. Gi Byung Hwang for the excellent research collaborations and valuable life advice.

Thank you all for your contributions and support, which have been instrumental in my academic journey. With these experiences and your support, I am now prepared to contribute to making the world a better and more advanced place for the future.

Contents

1. Introduction

1.1. Dimensionality

Materials science is a broad field that examines different forms of materials, from single atoms to large, solid materials. Between these extremes are small clusters of atoms known as nanoclusters, such as AlF_3, AlH_3, and Au. Unlike large materials that have a regular, repeating pattern in all directions, these nanoclusters generally do not follow these patterns due to their definite shape and size, which is why they are often termed zero-dimensional (0D).

In addition to 0D nanoclusters, materials can also be classified based on their dimensionality: two-dimensional (2D) structures like graphene[1], and one-dimensional (1D) structures such as nanowires and nanotubes [2,3] that resemble tiny wires. However, this study emphasises 0D nanoclusters structures of AlF_3, AlH_3, Au, and Ti_3N_4[4]. These nanoclusters are compact and do not extend infinitely in any direction, making them intriguing subjects of study as they exhibit behaviours distinct from larger materials.

Exploring these 0D nanoclusters offers a novel perspective in materials science research. Their size, shape, and composition help in tweaking their properties right at the atomic level[5]. By closely studying AlF_3, AlH_3, Au, and Ti_3N_4[4] nanoclusters, we can gain a better understanding of their unique characteristics and how they differ from larger materials. This study aims to bridge the gap between the atomic scale and the macroscopic domain of solid materials, enriching the discourse in materials science through the concept of dimensionality.

1.2. Nanoclusters

In recent years, the scientific community has seen a growing interest in the study of nanoclusters, thanks to their unique properties and potential applications in various fields like catalysis [6], optics[7,8], and materials science. Nanoclusters, usually comprising a small number of atoms or molecules, can adopt various

configurations[9–12]. These configurations significantly alter the behaviour of these entities compared to their bulk counterparts, frequently giving them new characteristics that are highly valuable for both theoretical investigation and practical applications.

A particularly significant area of nanocluster research is the study of inorganic clusters comprising transition and post-transition elements. In this area, clusters of compounds such as AlF_3, AlH_3, Ti_3N_4, and Au exhibit a diverse range of structural, electronic, and chemical properties that have not been extensively explored compared to larger-sized materials.

The majority of the work presented in this thesis focuses on the systematic exploration of AlF_3, AlH_3, and Au clusters, using a range of computational methodologies detailed in the following chapter. The nanoclusters studied contain fewer than 254 atoms, typically less than 1 nm across, therefore placing them at the smaller end of the nanocluster size range. The main goal of this study is to better understand the relationship between the structure and properties of these nanoclusters, which could help harness their potential in different technological areas.

To explore the structural predictions of these nanoclusters, we use a variety of theoretical tools, including interatomic potentials (IP), machine learning interatomic potentials (ML-IP), and density functional theory (DFT) methods. These tools help us to thoroughly analyse the structures of the nanoclusters and understand how these factors affect their physical and chemical properties. The detailed comparison of different materials and nanocluster sizes, supported by predicted clusters, aims to identify the cluster size at which they begin to exhibit bulk-like structural or physical properties.

1.2.1. Applications

Much like their oxide counterparts, AlF_3 and $AlCl_3$ nanoclusters show significant potential for a variety of applications similar to those of Au and ZnO nanoclusters. For instance, the optical properties of Au nanoclusters suggest

potential applications in optoelectronics[13] or photovoltaics[14], akin to how ZnO and TiO_2 nanoclusters are employed in sunscreens[15]. Although specific applications for AlF_3 and AlH_3 nanoclusters are yet to be extensively documented, their chemical and physical properties suggest potential uses in areas like catalysis, similar to the role ZnO[16] and CuO[17,18] nanoclusters play in gas sensing. The exploration of these applications requires a combination of theoretical and experimental approaches to fully unlock the potential of these nanoclusters in advancing technological frontiers.

1.2.2. Experimental Approaches

The journey into unearthing the properties and behaviour of nanoclusters encompasses a broad spectrum of experimental methodologies, each tailored to extract specific facets of information. Among these, Scanning Tunnelling Microscopy (STM) emerges as a key technique, enabling the atomistic imaging of surfaces[19,20]. By employing a fine tip in close proximity to the nanocluster, a quantum tunnelling current aids in mapping the electronic states, thus revealing the topography and electronic structure of nanoclusters. This method has been instrumental in capturing real-space images of nanoclusters, providing critical insights into their structural morphology.

Alongside STM, electron microscopy techniques such as Transmission Electron Microscopy (TEM) and Scanning Electron Microscopy (SEM) provide detailed architectural insights into nanoclusters. TEM, which transmits electrons through ultra-thin samples, provides high-resolution images[21,22], while SEM scans the surface with electrons to provide detailed images with depth perception [23]. These techniques have been pivotal in visualising the structural and compositional attributes of nanoclusters. Notably, in collaborative efforts, HR-TEM has been used to observe the Au_{25}(*L*-Cystein)$_{18}$ (or Au_{25}(*Cys*)$_{18}$) nanocluster, complementing computational studies to provide a more comprehensive understanding of this particular nanocluster; this study will be published in the near future.

Mass Spectrometry (MS) is also an essential tool in nanocluster research, enabling the precise determination of the mass-to-charge ratio of ionised species [24,25]. This methodology reveals the compositional diversity and stoichiometry of nanoclusters, forming a foundation theoretical predictions and computational validations.

Moreover, techniques such as gas-phase condensation have been employed to produce and study nanoclusters like CoPt in a gas-phase condensation apparatus[26]. Nuanced techniques such as Two-Photon Excited Fluorescence Near-Field Scanning Optical Microscopy (TPEF NSOM) and the investigation of ligand-protected metal nanoclusters have opened up additional avenues to explore the structure of nanoclusters[27]. The use of various experimental methods has enhanced our understanding of nanoclusters , opening up new possibilities for their applications. This combined experimental and computational approach is crucial in bridging the gap between the nanomaterials and larger scales in materials science, leading to new discoveries and a deeper understanding of the field.

1.3. 1D materials

One-dimensional (1D) materials exhibit nanoscale characteristics in two dimensions, embodying a blend of nanoparticle-like and bulk-like attributes due to the confinement of their electrons in two out of three dimensions. This category encompasses nanowires and nanotubes, both of which have attracted attention owing to their tunability and potential applications in nanotechnology. Among 1D materials, carbon nanotubes stand out as highly recognised entities [28,29]. Unlike nanowires, they possess a hollow structure and have found utility in the fabrication of particularly robust composite materials, such as those used in bicycle frames, and also as probe tips for atomic force microscopes [30]. On the other hand, nanowires have been instrumental in the field of transistor technology, aiding in the miniaturisation of transistor sizes [31].

1.4. 2D materials

Two-dimensional (2D) materials are a significant group primarily derived from layered structures. They are defined by having two surfaces or interfaces that are just nanometres apart. These materials might appear as standalone thin films [1], as layers on a 3D material [32], or as part of layered structures [33]. What makes 2D materials particularly noteworthy is their similarity to more common materials in certain aspects, like having a band structure that changes with the layer thickness[33]. They also have unique properties, including a high surface-to-volume ratio and the ability to exhibit quantum effects [34].

Graphene [1] is especially important among these materials. It can act as a base for adding nanoclusters, creating a useful setting to study nanoscale interactions. For instance, reduced graphene oxide (rGO) is often used as a base to enhance the effects of catalysts[35–37] such as improving the photocatalytic power of TiO_2. Studying 2D materials helps us understand how different material structures interact, opening up new possibilities for advancements in nanotechnology.

2. Theory and Computational Methods

2.1. Interatomic Potentials

The exploration of material behaviour on atomic and molecular scales necessitates a robust theoretical framework that can accurately describe the interactions between atoms and molecules. At the core of this framework is the idea of IPs. IPs are crucial for understanding and predicting the properties and behaviours of materials. They describe how the energy between atoms changes based on their positions, helping us model the forces and movements in a system of interacting particles. They are essential for bridging quantum mechanics with classical mechanics, enabling the simulation and analysis of large systems that would be too complex to handle with purely *ab initio* calculations. In this study, IPs are employed at the stage of global optimisation on the energy landscape of nanoclusters.

IPs are fundamental to molecular dynamics (MD) simulations, which are crucial for studying how material systems change under various conditions. By using suitable IPs, researchers can explore various material phenomena such as phase transitions, mechanical properties, thermal transport, and reaction dynamics, among others.

The construction and choice of IPs are of paramount importance as they significantly impact the accuracy and reliability of simulations and predictions. Various forms of IPs [38,39] have been developed, each tailored to capture specific types of interactions and material systems. In this chapter, the focus is on the IPs employed, elucidating their relevance and application in the explored contexts. The effort to improve and create new IPs remains an active field of research, aiming to expand our understanding of materials science and engineering. This work also provides a strong basis for studying nanoclusters and other complex materials through theory and computation.

2.1.1. Two-Body Potentials

The development of IPs has been a cornerstone in computational chemistry, with origins extending back to the early 20th century. A seminal contribution from this era was made by Lennard-Jones (LJ)[40] in the 1920s, who introduced a specific pair potential tailored for noble gases. This groundbreaking work laid the foundation for subsequent advancements in the field, emphasising the importance of accurate modelling of atomic interactions.

The LJ potential is characterised by two terms:

$$V(r_{ij}) = \frac{A}{r_{ij}{}^{12}} - \frac{B}{r_{ij}^{6}}, \qquad Eqn\ 1$$

The first term, r^{-12}, describes the strong electronic repulsion between atoms at very short distances, preventing an unphysical overlap of their electron clouds. This repulsion is influenced by the size of the interacting particles, with larger atoms or ions experiencing stronger repulsion due to an increased likelihood of electron cloud overlap. The parameter *A* represents the strength of the repulsive interaction and is related to the size of the ions or atoms involved.

The second term, or r^{-6} term, describes attractive forces between atoms due to van der Waals (vdW) interactions, which arise from induced dipole–induced dipole interactions. This term can be theoretically derived from two linear harmonic dipole oscillators[41,42]. The parameter *B* quantifies the strength of the electron shell dispersion between atoms *i* and *j*, representing the depth of the potential well and indicating the strength of the vdW interactions between the atoms.

In computational chemistry, the LJ potential is crucial for studying noble gases and modelling non-reactive systems, demonstrating its ongoing relevance. Over a decade after the creation of the LJ potential, the Buckingham potential was formulated[43]:

$$V(r_{ij}) = A_{ij}\, exp\left(-\frac{r_{ij}}{\rho_{ij}}\right) - \frac{C_{ij}}{{r_{ij}}^6}, \qquad Eqn\ 2$$

This new potential substituted the repulsive r^{-12} component in the LJ potential with an exponential function of distance that more accurately reflects reality. Parameters A_{ij}, ρ_{ij}, and C_{ij} describe the ionic size, the extent of shielding (or compressibility), and the intensity of the electron shell dispersion between atoms *i* and *j*, respectively. (Note that the A term in the LJ and Buckingham potentials are different and not directly comparable) *N.B.* the *A* term in LJ and Buckingham potentials are different. The exponential repulsive term ensures strong repulsion at short distances, reflecting the incompressibility of the atomic cores.

When modelling charged systems, interactions between cations often lack the electron shell dispersion term, $-\frac{C_{ij}}{{r_{ij}}^6}$, since they typically have low polarisability. Therefore, cations often give negligible contributions to vdW energies and forces.

$$V(r_{ij}) = \frac{q_i q_j}{r_{ij}}, \qquad Eqn\ 3$$

In many cases, the Coulomb potential, as shown in *Eqn 3* for cation–anion interactions, can lead to an unphysical behaviour known as the Coulomb catastrophe. The q_i and q_j represents the charges of ions *i* and *j*, and r_{ij} is the interatomic distance between atoms *i* and *j*. In contrast, anion–anion interactions do exhibit some electron shell dispersion. This feature, represented by the second term of the Buckingham potential, can also result in an unphysical behaviour known as the Buckingham catastrophe at short interatomic distances between anions. To avoid this issue, the Buckingham potential is often paired with other more repulsive but quickly decaying potentials at short distances to model material systems robustly.

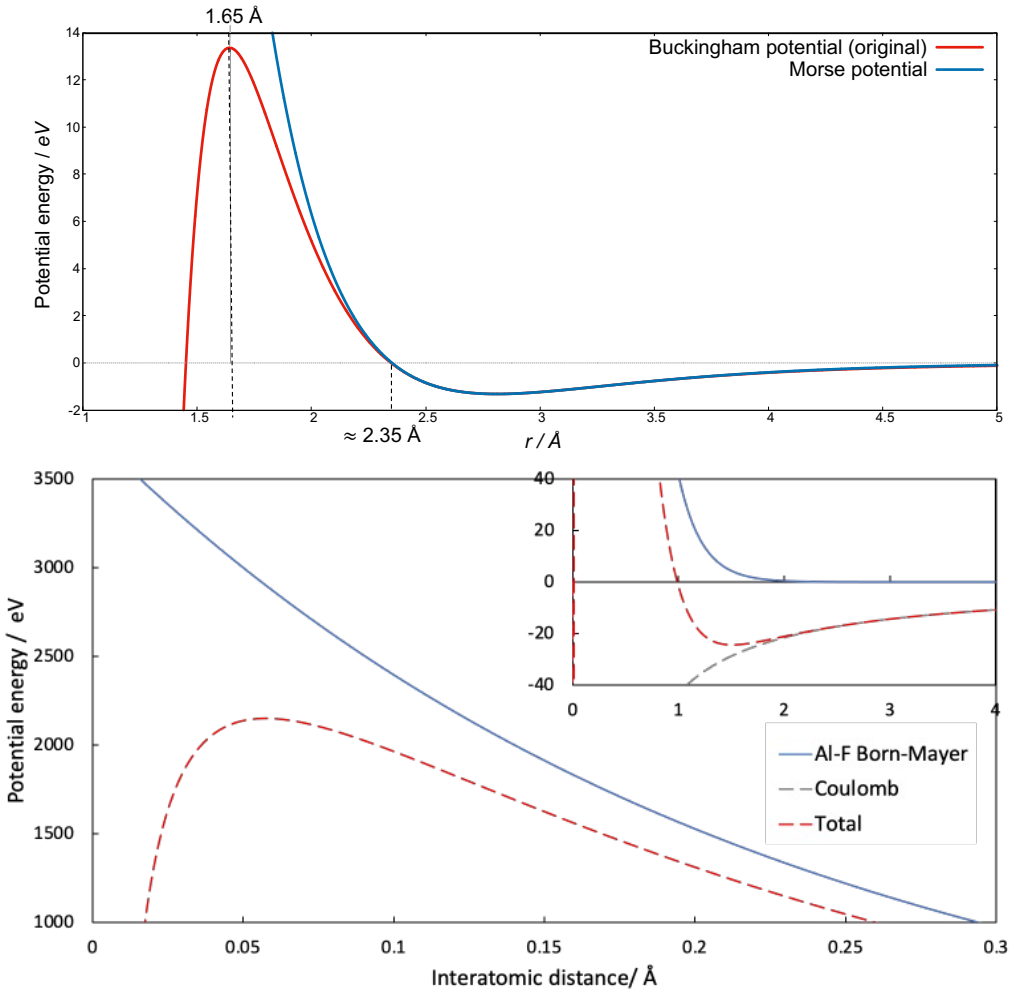

Figure **1**. *Upper Panel* depicts the fitted Buckingham potential[4] for N – N interactions (parameters: A = 5896.8405 eV, ρ = 0.2737 Å), represented by a red solid line, alongside the N – N Morse potential shown as a blue solid line. *Lower Panel* shows a specific range of the Al – F Born-Mayer potential; the inset focuses on the equilibrium distances[44]. In these representations, the Al–F Born-Mayer potential is marked by a blue solid line, the Al – F Coulomb potential by a grey dashed line, and the total potential (a combination of the Al–F Born-Mayer and Al – F Coulomb potentials) by a red dashed line.

An illustrative example of the Buckingham catastrophe can be seen in the upper part of Figure 1, where the N – N Buckingham potential is shown as a red solid line, and an alternative interatomic potential, the N – N Morse potential, is presented as a blue solid line. While the N – N Buckingham potential was effective for simple structure optimisations, we faced the Buckingham catastrophe during the global optimisation of nanoclusters. This issue arose during the initial random

structure generation process in the global optimisation, which placed N and N atoms closer than 1.65 Å. The inclusion of the Coulomb potential on the Buckingham potential further expanded the range of the Coulomb catastrophe as the opposite charged system has infinite attraction at short distance (grey dashed line in the small panel of figure 1). As a solution, we fitted the N – N Morse potential to the N – N Buckingham potential as an alternative IP, as the Morse potential does not exhibit the unphysical short-range attraction found in the Buckingham potential.

The Morse potential is given by:

$$V(r_{ij}) = D_e[1 - e^{-a(r_{ij}-r_e)}]^2 - D_e, \qquad Eqn\ 4$$

where, $V(r_{ij})$is the potential energy between atoms i and j, D_e is the well depth (the bond dissociation energy), a is a parameter related to the stiffness of the bond, r_e is the equilibrium bond distance, $r_{ij}$is the interatomic distance between atoms i and j.

By fitting the Morse potential parameters D_e, a, and r_e to reproduce the behaviour of the N – N interactions, we avoided the unphysical attractions at short distances that led to the Buckingham catastrophe.

The lower part of Figure 1 provides another example, displaying the Al – F Born-Mayer potential and the F – F four-region Buckingham potential. The specific range where the Coulomb catastrophe might occur is when the interatomic distance between Al and F atoms is $r_{\text{Al-F}} < 0.057$ Å. In this instance, we avoided the Coulomb catastrophe thanks to the global optimisation software's automated fragmentation and collapse detection algorithm, based on ionic radii. However, this algorithm could not prevent the earlier N – N Buckingham catastrophe example, as the Buckingham catastrophe region extended significantly beyond the detection range of the collapsed structure algorithm.

The Born-Mayer potential used for Al –F interactions is given by:

$$V(r_{ij}) = A_{ij}\exp\left(-\frac{r_{ij}}{\rho_{ij}}\right), \qquad \textit{Eqn 5}$$

where, $V(r_{ij})$ is the potential energy between atoms *i* and *j*, A_{ij} is a parameter representing the magnitude of the repulsive interaction, $\frac{r_{ij}}{\rho_{ij}}$ is a parameter related to the effective size or "hardness" of the atoms, r_{ij} is the interatomic distance between atoms *i* and *j*.

Numerous distinct interatomic potentials (IPs) have been developed, including cluster potentials such as the Embedded Atom Method (EAM) [44,45], the Modified Embedded Atom Method (MEAM) [46–48], and charge-optimized many-body potentials [49,50]. These IPs are more complex due to the multitude of factors they incorporate to accurately represent material behavior with more terms, such as angular, dihedral angle, *etc*. A distinguishing feature of these potentials is their mathematical formulations, which are designed to capture the essential physics of the material systems they aim to model. While IPs offer the advantage of lower computational costs when studying target materials, the process of fitting or parameterizing even simple IPs for specific materials has proven challenging. In contrast, recent advancements in machine-learning interatomic potentials (ML-IPs)[51,52] have largely addressed the difficulties associated with parameterization and have improved accuracy [53,54].

In the current chapter, we will introduce the method that was used to study AlX_3 (X = F, H), and Au(*Cys*)$_{18}$. Subsequently, we will introduce the training methods for ML-IPs.

2.1.2. Shell model

The shell model, proposed by Dick and Overhauser [55], is able to empirically describe the dipolar polarisability of a cation or anion by considering a discrete core (nucleus and inner electrons) and a shell (valence electrons) which are superimposed onto each other (Figure 2). The core and shell are separated but are joined by a harmonic spring with the force constant k_{cs} Electronic polarisability,

α, is modelled by allowing the charged shell to move in response to forces from nearest neighbour ions.

(A)

(B)

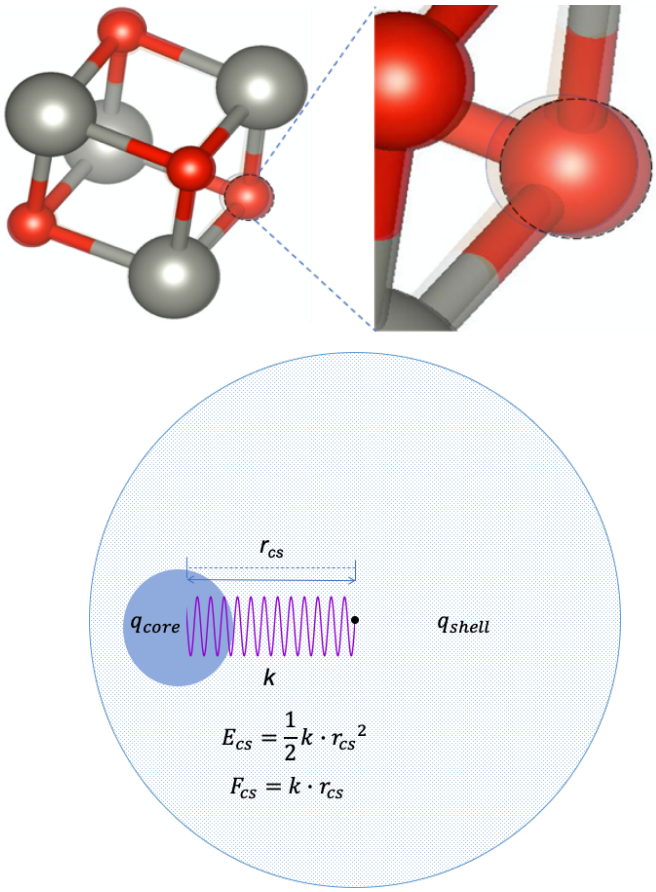

Figure **2**. Demonstration of the polarised shell model: (A) a $(ZnO)_4$ nanocluster structure with the structural rigid ion model (solid dashed line) and the shell model (transparent) superimposed; and (B) shows a simplified diagram of the shell model of a polarised ion: the solid blue/smaller circle is the q_{core} charged core of an cation/anion and the light blue/bigger circle is the q_{shell} charged shell of an cation/anion. The shell and core are interlinked with a harmonic spring (purple) which has a spring constant, k.

$$\mu_{induced} = \alpha\, E = q_s r_{cs}, \qquad Eqn\ 6$$

where, $\mu_{induced}$ is the induced dipole moment, α is the electronic polarisability, E is the electric field, q_s is the charge of the shell, r_{cs} is the displacement between the core and the centre of the shell. Similarly, the short-range interactions are usually defined to act on the shell, but Coulomb interactions, *Eqn* 3, act on both the core and the shell. The spring constant k is related to the polarisability by:

$$k_s = {q_s}^{2/\alpha}, \qquad \textit{Eqn } 7$$

where, k_s is the core–shell force constant, $q_s$is the charge of the shell, α is the electronic polarisability. Here, the short-range interactions are usually defined to act on the shell, but Coulomb interactions (see *Eqn* 3) act on both the core and the shell. The polarisability decreases as the force of the core–shell interaction increases because the shell and core are tightly bound. The strength of binding between the core and shell can be described by Hooke's law:

$$E_{cs} = \frac{1}{2} k_s \, {r_{cs}}^2, \qquad \textit{Eqn } 8$$

where, E_{cs} is the elastic energy between the core and shell, k is the spring constant, r_{cs} is the displacement between the core and the shell. Thus, the polarisability is dependent on the local chemical environment. However, the point-charge shell model is not sufficient to explicitly describe the complex nature of extensive electron clouds to reflect higher-order moments of electric fields and charge density.

2.1.3. Electrostatics

In our computational approach, we primarily focus on two-body potentials to enhance efficiency. For compounds like AlF_3 and AlH_3, where both aluminium (Al) and fluorine/hydrogen (F/H) ions are charged and polarisable, the Coulomb potential is a fundamental component. This potential represents the interactions between point charges and their associated shells. The Coulomb potential for any pair of point charges is given by *Eqn* 3.

Transitioning to the context of nanoclusters, the Madelung potential, typically associated with long-range Coulombic interactions in ionic crystals, takes on a nuanced role. Due to the smaller size and fewer ions in nanoclusters compared to bulk materials, the Madelung potential is inherently weaker, as there are fewer atoms contributing to the overall electrostatic interactions. This reduced Madelung potential allows nanoclusters to exhibit a wide range of configurations since they are not as tightly bound by the electrostatic constraints seen in bulk materials.

Additionally, the reduced electrostatic constraints in nanoclusters allow for more flexibility in the structural arrangement of atoms. This leads to diverse and potentially more reactive configurations, which is particularly important in applications such as catalysis. The ability to adopt various configurations can enhance the catalytic activity of nanoclusters, making them highly versatile and efficient in different chemical processes.

2.1.4. Many-Body Potentials

In the complex world of atomic and molecular interactions, understanding the forces between atoms is crucial for determining how materials behave and what their properties are. While pair potentials provide a simplified representation of these interactions by considering them in isolation, they often fall short in capturing the complexities arising from simultaneous interactions among multiple atoms. This is where many-body potentials come into play.

Many-body potentials provide a comprehensive framework that takes into account interactions involving more than two atoms simultaneously. By considering the collective effects of multiple neighbouring atoms, these potentials offer a more nuanced and accurate depiction of real-world phenomena in atomic systems. In essence, they recognise that the energy of an atom in a material is not just influenced by pairwise interactions but by the broader environment of surrounding atoms.

Within the scope of many-body potentials, various models and methods have been developed to address specific material systems and interactions. This chapter briefly introduces important potentials such as the EAM[45,56], Potential, the MEAM[47,48], and the COMB3[49,50] potential.

2.1.4.1. Embedded Atom Method potential

The EAM is a computational approach widely used to model the behaviour of metallic systems. It is particularly useful for simulating the properties of metals

and alloys because it effectively captures the interactions between atoms in these materials. The fundamental concept of EAM is that the energy of each atom in the system is influenced not just by the positions of nearby atoms, but also by the electron density generated by these atoms.

In EAM, the total energy of a system is considered to be the sum of two main contributions: the embedding energy and the pair potential energy. The embedding energy is the energy required to place an atom into the electron cloud created by its neighbouring atoms. The pair potential energy represents the interaction energy between pairs of atoms, similar to the classical pair potentials used in simpler models.

$$E = \sum_i E_i, \qquad \textit{Eqn } 9$$

where E_i is the energy associated with atom iii.

The energy E_i consists of two parts: the embedding energy $F(\rho_i)$ and the pair potential energies involving atom *i*. The embedding energy term depends on the local electron density $F(\rho_i)$ at atom *i*, which is created by all the other atoms in the vicinity. The embedding energy function $F(\rho_i)$ describes how the energy changes as the electron density varies. The pair potential energy term represents the interaction energies between atom *i* and its neighbouring atoms *j*, separated by distances r_{ij} The pair potential $\phi(r_{ij})$ captures how the energy changes with the distance between two atoms.

Thus, the energy associated with atom *i* can be written as:

$$E_i = F(\rho_i) + \frac{1}{2}\sum_{i \neq j} \phi(r_{ij}), \qquad \textit{Eqn } 10$$

where, the factor of 1/2 avoids double-counting the pair interactions.

The local electron density ρ_i at atom *i* is a sum of contributions from all other atoms j:

$$\rho_i = \sum_{i \neq j} f(r_{ij}), \qquad \textit{Eqn } 11$$

where, $f(r_{ij})$ is the contribution to the electron density at atom *i* due to atom *j*, and it depends on the distance r_{ij} between atoms *i* and *j*. Typically, $f(r_{ij})$ decreases as the distance r_{ij} increases.

Putting it all together, the total energy E of the system can be written as:

$$E = \sum_i F(\rho_i) + \frac{1}{2}\sum_{i \neq j} \phi(r_{ij}), \qquad \textit{Eqn } 12$$

where, *E* is the total energy of the system, $F(\rho_i)$ is the embedding energy of atom *i*, dependent on the local electron density ρ_i, $\phi(r_{ij})$ is the pair potential energy between atoms *i* and *j*. The first term, $\sum_i F(\rho_i)$,represents the sum of the embedding energies for all atoms. The second term, $\frac{1}{2}\sum_{i \neq j} \phi(r_{ij})$, accounts for the pairwise interaction energies between atoms, The factor of 1/2 prevents double-counting the interactions since each pair is considered only once.

To illustrate, consider a simple example with two atoms at a distance *r*. Each atom contributes to the electron density at the other atom. If we denote the contribution from one atom as $f(r)$, the local electron density at each atom is $\rho = f(r)$. The embedding energy for each atom is given by $F(\rho)$. The pair potential energy between the two atoms is $\phi(r_{ij})$. The total energy *E* for this two-atom system is:

$$E = 2F(\rho) + \phi(r), \qquad \textit{Eqn } 13$$

here, $2F(\rho)$ is the sum of the embedding energies for the two atoms, $\phi(r)$ is the pair potential energy between the two atoms.

The EAM potential provides a more realistic representation of atomic interactions in metallic systems by considering both the embedding of atoms into the electron density and the pairwise interactions. This dual approach allows for more accurate simulations of the physical properties of metals and alloys, making EAM a valuable tool in computational chemistry and materials science.

2.1.4.2. Modified Embedded Atom Method

The MEAM[46–48,53,57] builds on the EAM by introducing angular terms to capture bond directionality, making it suitable for a broader range of materials, including ceramics and covalent compounds. MEAM's model is articulated as:

$$E = \sum_i F(\rho_i) + \frac{1}{2}\sum_{i \neq j} \phi_{ij}(r_{ij}) + \sum_{i \neq j} \sum_{k \neq i,j} G(\theta_{ijk}), \qquad Eqn\ 14$$

In this equation 14, E represents the total energy of the system. The first term, $\sum_i F(\rho_i)$, sums the embedding energies $F(\rho_i)$ for all atoms i in the system. The embedding energy $F(\rho_i)$ depends on the local electron density ρ_i at atom i, which accounts for the electron density contributions from neighbouring atoms. This term captures how the energy changes as an atom is "*embedded*" in the electron cloud of surrounding atoms.

The second term, $\frac{1}{2}\sum_{i \neq j} \phi_{ij}(r_{ij})$, represents the pair potential energies between pairs of atoms. Here, $\phi_{ij}(r_{ij})$ is the pair potential energy between atoms i and j, and r_{ij} is the distance between these two atoms. The summation runs over all unique pairs i and j (with $i \neq j$), and the factor of 1/2 ensures that each pair interaction is only counted once, avoiding double-counting.

The third term, $\sum_{i \neq j} \sum_{k \neq i,j} G(\theta_{ijk})$, accounts for the angular dependence of atomic interactions, which is crucial for materials with directional bonding. In this term, θ_{ijk} is a function that describes the angular potential energy associated with the angle θ_{ijk} formed by atoms i, j, and k. The angle θ_{ijk} is defined by the positions of these three atoms, capturing the directional characteristics of the bonds in the material. The double summation runs over all unique combinations where $i \neq j$ and $k \neq i, j$, effectively considering all possible angles involving atom i and its neighbours.

This formulation allows MEAM to account for the geometric arrangement of atoms, which is crucial for materials like silicon and carbon that exhibit strong directional bonding due to their covalent nature. The inclusion of angular terms increases computational demands and requires extensive parameter calibration to accurately represent the material's properties. However, MEAM's ability to model

complex materials makes it a versatile tool in computational studies, extending beyond metals to include semiconductors, ceramics, and other complex materials.

2.1.4.3. Charge Optimized Many-Body Potential

Classical potentials like the EAM and MEAM have their merits, but they fall short of capturing intricate charge interactions. The COMB3 potential [49,50,58] a sophisticated method rooted in quantum mechanical principles, fills this gap. It includes many-body interaction terms and dynamically adjusts atomic charges during simulations. This is crucial for systems where electrostatic forces are significant, such as mixed ionic-covalent materials and complex interfaces.

COMB3 is designed to handle dynamic charge variations, which are key to accurately simulating electrostatic forces in various materials. The energy E in a COMB3-modelled system is expressed as:

$$E = \sum_i F(\rho_i) + \frac{1}{2}\sum_{i \neq j} V_{ij}(r_{ij}, q_i, q_j) + E_{\text{elec}}, \qquad Eqn\ 15$$

where, E represents the total energy of the system. The first term, $\sum_i F(\rho_i)$ sums the embedding energies $F(\rho_i)$ for all atoms i in the system. The embedding energy $F(\rho_i)$ depends on the local electron density ρ_i around atom i, reflecting how the energy changes as an atom is embedded in the electron cloud of its neighbours.

The second term, $\frac{1}{2}\sum_{i \neq j} V_{ij}(r_{ij}, q_i, q_j)$, accounts for the pairwise interaction energies between atoms. Here, $V_{ij}(r_{ij}, q_i, q_j)$ is the pairwise interaction potential between atoms i and j, which depends on the interatomic distance r_{ij} and the charges q_i and q_j of the interacting atoms. The factor 1/2 ensures that each interaction is counted only once, preventing double-counting.

The third term, E_{elec}, stands for the electrostatic energy of the system. This energy is calculated based on dynamically optimised charges that adjust in real time during simulations.

The Charge Equilibration (QEq) [52,59] methodology within COMB3 modifies atomic charges based on their surroundings, effectively capturing charge transfer

and polarisation effects. This dynamic adjustment of charges ensures accurate simulations of systems with significant electrostatic interactions. However, this added layer of complexity can be computationally demanding, as it requires charge optimisation at each simulation step.

Overall, the primary strength of COMB3 lies in its ability to simulate a wide variety of materials with different electronic properties, making it a versatile and powerful tool in computational studies, despite the increased computational demands.

2.2. Machine Learning Interatomic Potential

The landscape of computational materials science is ever-changing. Classical approaches like DFT and potentials such as the EAM have been cornerstones in the field. However, these traditional methods come with computational bottlenecks and limitations in capturing complex atomic interactions due to computational costs or accuracy reasons.

Enter ML algorithms [53,60–62] which have revolutionised the way we approach materials science. ML-IP are at the forefront of this transformation, offering a blend of computational efficiency and accuracy. These potentials are trained on comprehensive datasets, often sourced from experimental observations, first-principles methods like DFT, or empirical data, to deliver highly precise predictions for atomic and molecular systems.

Computational efficiency is a critical consideration when selecting methods for simulating material properties, especially for large systems or when extensive sampling is required. Traditional classical IP, DFT, and ML-IP, such as the Message Passing Atomic Cluster Expansion (MACE), each offer different balances between computational cost and accuracy.

Classical IPs are the most computationally efficient, allowing simulations of millions of atoms over long timescales. They use simplified mathematical forms to model atomic interactions, which can make them up to 10^6 times faster than DFT calculations[51,63]. This efficiency makes classical IPs ideal for large-scale

molecular dynamics simulations where computational speed is essential. However, their simplified nature can limit their accuracy, particularly in systems where electronic effects play a significant role.

DFT provides a more accurate description by explicitly accounting for the electronic structure of materials. It is widely used for calculating material properties with high precision. However, DFT is computationally intensive, with the computational cost scaling roughly with the cube of the number of electrons in the system ($\boldsymbol{O}(N^3)$). This scaling makes DFT impractical for large systems or long timescale simulations, limiting its application to smaller systems with fewer atoms.

ML-IPs like MACE offer a compromise between the efficiency of classical IPs and the accuracy of DFT. ML-IPs are trained on DFT data to predict atomic interactions, enabling them to achieve near-DFT accuracy while significantly reducing computational cost. For instance, MACE potentials leverage message-passing neural networks to model both local and many-body interactions efficiently. They can be up to 10^4 times faster than DFT calculations while maintaining comparable accuracy, making them suitable for simulating larger systems than those typically accessible with DFT. Compared to classical IPs, ML-IPs are generally about 10 to 100 times slower, but the substantial increase in accuracy often justifies the additional computational cost.

This balance between accuracy and efficiency makes ML-IPs particularly valuable for studies requiring detailed atomic-level insights without the prohibitive computational expense of DFT. They enable simulations of larger systems or longer timescales than feasible with DFT, while providing more accurate results than classical IPs.

2.2.1. Gaussian Process Regression

Gaussian Process Regression (GPR) [64], a pillar in the domain of machine learning, offers a non-parametric approach to data modelling. At its core, GPR provides a way to predict a continuous output variable based on a set of input

variables. Instead of establishing a fixed form for the function that maps inputs to outputs, GPR expresses this function as a probability distribution over all possible functions. When applied to computational materials science, GPR serves as the foundation for the Gaussian Approximation Potential (GAP). GAP harnesses the power of GPR to approximate potential energy surfaces, offering a balance between computational efficiency and accuracy. By training on datasets derived from quantum mechanical simulations, GAP can interpolate the potential energy and atomic forces for unseen atomic configurations.

The rise of machine learning in this domain can be attributed to several factors. Firstly, ML algorithms can process extensive and complicated datasets more efficiently than classical computational techniques. Secondly, ML models, especially neural networks, have the ability to learn complicated patterns in the data, making them highly versatile and applicable to a diverse range of materials.

2.2.2. Neural Network Interatomic Potentials

Neural networks are often the driving computational force behind many ML-IPs. When ML-IPs use neural networks as a core architecture, they are often referred to as Neural Network Interatomic Potentials (NN-IPs). Expanding upon the core principles of neural networks, Graph Neural Networks (GNNs) [65] offer a specialised framework designed to handle graph-based, non-Euclidean data structures. GNNs adapt machine learning algorithms to work directly with non-Euclidean data, preserving the attributes (e.g., node features, edge features, graph topology, local and global patterns) that would otherwise be lost in a Euclidean transformation. By doing so, GNNs are adept at capturing both local and global structural features.

What distinguishes ML-IPs is their adaptability and precision. Unlike traditional methods that are restricted by fixed mathematical forms and parameters, ML-IPs use adaptable algorithms like NN that can be customised for various materials and conditions. The applications of ML-IPs are vast, ranging from the

study of amorphous materials and complex biological systems to multi-elemental alloys and reactive chemical environments [66,67].

While traditional NN are adept at handling Euclidean data, as seen in computer vision and natural language processing, they struggle when applied to non-Euclidean domains like atomistic or chemical systems. Representing atomic data in a 3D Euclidean space can lead to the loss of vital information, such as edge qualities (*e.g.*, bond type, bond order), node values (*e.g.*, electronegativity, atomic number, potential charge), or edge directionality (*e.g.*, dipole moment, direction of electron transfer)[68,69]. It also overlooks potential higher-dimensional correlations between unconnected nodes or long-range interactions [70].

GNNs address these challenges by adapting machine learning algorithms to work directly with non-Euclidean data [71]. This ensures the preservation of intricate attributes that might be lost in a Euclidean transformation. In the context of atomic systems, atoms are considered as nodes, and their interactions form the graph edges.

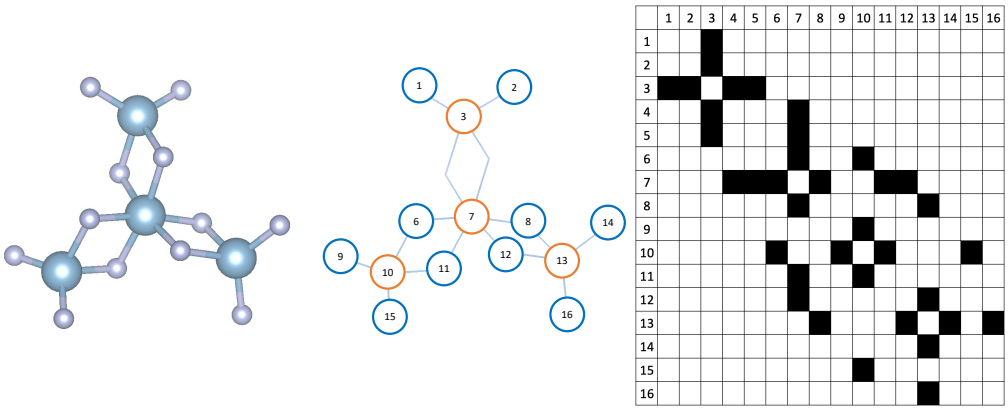

Figure **3**. (*Left*) Ball-and-stick model of $(AlF_3)_4$, the second-ranked lowest energy local minimum (LM) on the PBEsol energy landscape. (*Centre*) Graph representation (with nodes and edges) of the cluster. (*Right*) Adjacency matrix of the bonds in the cluster.

Consider a chemical system with various atomic pairs at different distances, some of which are connected with bonds. Representing this 3D system as a graph offers a convenient abstraction, where nodes symbolise atoms and edges signify chemical bonds. For example, Figure 3 includes a ball-and-stick model of the

second lowest energy $(AlF_3)_4$ local minimum on the PBEsol energy landscape, along with its corresponding graph and adjacency matrix representations. The adjacency matrix provides a simplified 2D description of this graph, as shown on the right-hand side of Figure 3, capturing the bonding patterns within the cluster.

2.2.3. Multi Atomic Cluster Expansion (MACE)

Firstly, the MACE model processes information from each node or atom to build a vector that encapsulates the atom's local environment, as shown in Equation 16. In this equation, $\sigma_i^{(t)}$ represents the state of atom *i* at iteration *t* in the message-passing algorithm. Specifically:

$$\sigma_i^{(t)} \equiv \left(r_i, \theta_i, h_i^{(t)}\right), \qquad \textit{Eqn } 16$$

where, r_i is the position vector of atom *i*, θ_i represents chemical-specific parameters for atom *i*, such as atomic number or type, $h_i^{(t)}$ is a learnable feature vector for atom *i* at iteration *t*.

This state $\sigma_i^{(t)}$ denotes the features or embeddings of the node at a specific step in the iterative process of updating node states based on information received from neighbouring nodes. Each iteration *t* refines the features $h_i^{(t)}$ to capture more accurate representations of the atomic environment as the algorithm progresses.

$$m_i^{(t)} = \frac{1}{\lambda} \bigoplus_{j \in N(i)} M_t\left(\sigma_i^{(t)}, \sigma_j^{(t)}\right), \qquad \textit{Eqn } 17$$

Equation 17 describes how information from each node is aggregated with information from neighbouring nodes to form the message $m_i^{(t)}$. Thus, $m_i^{(t)}$ is the aggregated message for atom *i* at iteration *t*, λ is a normalisation constant to scale the aggregated message, $\oplus$ represents an aggregation operation, such as summation or averaging over neighbouring atoms, $N(i)$ denotes the set of neighbouring atoms of atom *i*, M_t is a message function at iteration *t* that combines the states of atom *i* and its neighbour *j* to compute the contribution to the message.

This aggregation encapsulates the pairwise interactions, including symmetry functions and radial distribution functions, ensuring rotational and translational invariance in the model.

$$\sigma_i^{(t+1)} \equiv \left(r_i, \theta_i, U_t\left(\sigma_i^{(t)}, m_i^{(t)}\right)\right), \qquad \textit{Eqn } 18$$

Equation 18 updates the state $\sigma_i^{(t+1)}$ of atom *i* at iteration $t+1$ by incorporating the current state $\sigma_i^{(t)}$ and the aggregated message $m_i^{(t)}$ from neighbouring atoms:

Here, U_t is an update function at iteration *t* that refines the learnable feature vector $h_i^{(t)}$ based on the current state $\sigma_i^{(t)}$ and the message $m_i^{(t)}$, $\sigma_i^{(t+1)}$ is the updated state of atom *i* for the next iteration. This iterative update ensures that the features or embeddings of each atom are refined with each iteration, capturing more detailed and accurate representations of the atomic environment.

$$E_i = \sum_t R_t\left(\sigma_i^{(t)}\right), \qquad \textit{Eqn } 19$$

which calculates the energy E_i of atom *i* as the sum over a series of terms R_t, each being a function of the state $\sigma_i^{(t)}$ at iteration *t*. E_i is the energy contribution from atom *i*, R_t is a readout function at iteration *t* that computes a contribution to the energy based on the state $\sigma_i^{(t)}$, The summation over *t* indicates that the total energy is accumulated over multiple iterations.

This indicates that the energy depends on multiple state evaluations over time or iterations, with the system's total energy being a sum of local contributions that can be learned and predicted.

The MACE machine learning interatomic potential effectively captures the complex interactions within atomic systems by iteratively refining the features of each atom based on its local environment and interactions with neighbouring atoms. This results in a comprehensive model that can predict the energy and properties of atomic systems with high accuracy, ensuring rotational and translational invariance. The MACE approach highlights the importance of

considering both local and global atomic interactions, making it a powerful tool for computational materials science.

2.2.4. Training Machine Learning Interatomic Potentials

In this section, we outline the procedure used to train the MACE model. This training utilises "frames" obtained from DFT single-point calculations related to the vibrational modes of stable nanoclusters. Performing global optimisation directly on the DFT PES is limited by computational costs. Therefore, we employ a two-step method that performs global optimisation using a fitted IP. However, fitting an IP for a target system requires experience and time to achieve accuracy. The primary objective of this training is to refine the MACE model so that it can replace the two-step approach discussed in the previous chapter, streamlining the process and improving efficiency. This development aims at nanocluster structure prediction that offers near-DFT precision combined with computational efficiency akin to IP calculations. Unlike other general ML-IPs, which are crafted and trained for MD calculations, our focus is to customise the ML-IP specifically for static structure calculations, with a particular emphasis on locally stable nanocluster structures.

To achieve this aim, we decided to focus on the local minima and their nearby regions, which we probed using finite displacements along respective vibrational modes. Figure 4 shows a model one-dimensional energy landscape with such images of vibrational modes.

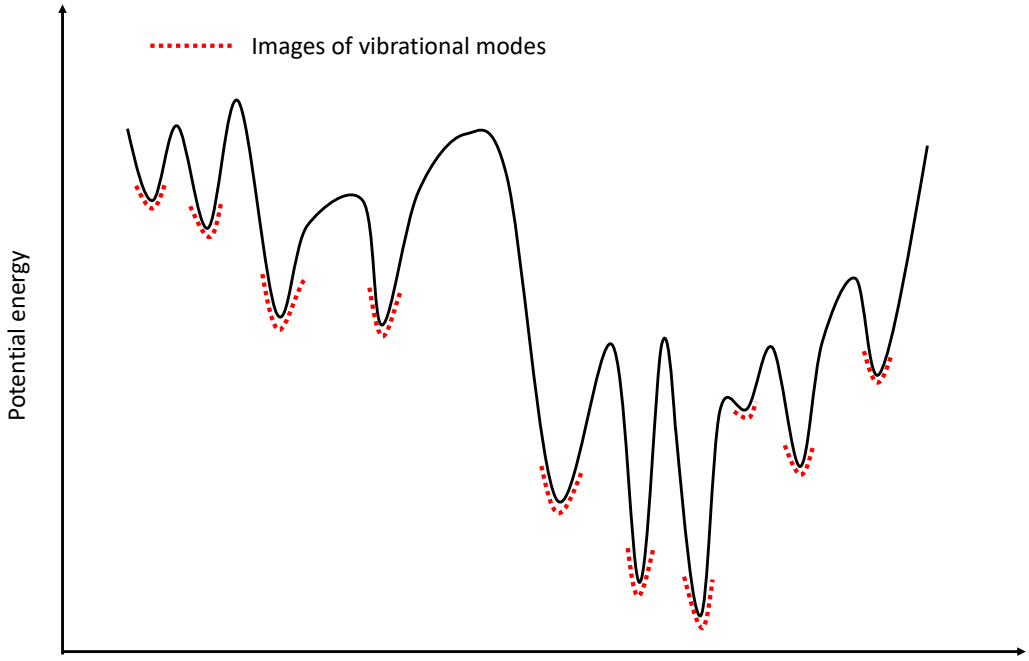

Figure **4**. Model one-dimensional energy landscape of a system shown with a black solid line, and a part of the energy landscape covered by scanning the vibrational modes around local minima highlighted with red dotted lines.

The MACE model employs a unique and unconventional training approach compared to traditional IP models. Rather than focusing primarily on specific configurations (often referred to as 'images') of the energy landscape and the direct transitions between them—a methodology known as direct connectivity—MACE emphasises learning from the local atomic environments within the system. In atomic systems, the PES can be visualised as a high-dimensional surface where each point represents a possible atomic arrangement along with its associated energy value. By concentrating on the immediate surroundings of each atom, MACE effectively captures the fundamental local interactions and geometries that govern the system's overall behaviour. This approach enhances the model's ability to generalise across a wide range of configurations, leading to more accurate and transferable predictions of material properties.

By centring its training on atomic environments, MACE comprehensively learns the positions, varieties, and interactions of atoms across numerous configurations. This enables the model to understand the complex connections

among atoms in varied settings, broadening its comprehension of the potential energy landscape. Consequently, MACE is equipped to precisely forecast the energy of diverse atomic arrangements. Fundamentally, the model's prioritisation of atomic environments over specific images makes MACE exceptionally flexible and adept at predicting atomic interactions and dynamics. Thus, by providing images within the (near) LM basins as training data, MACE accurately learns about these basins. We anticipate that data points near the local minima will enable MACE to also capture other features of the PES, such as unstable higher-energy surfaces or saddle points, owing to the broad coverage of atomic environments in the training data.

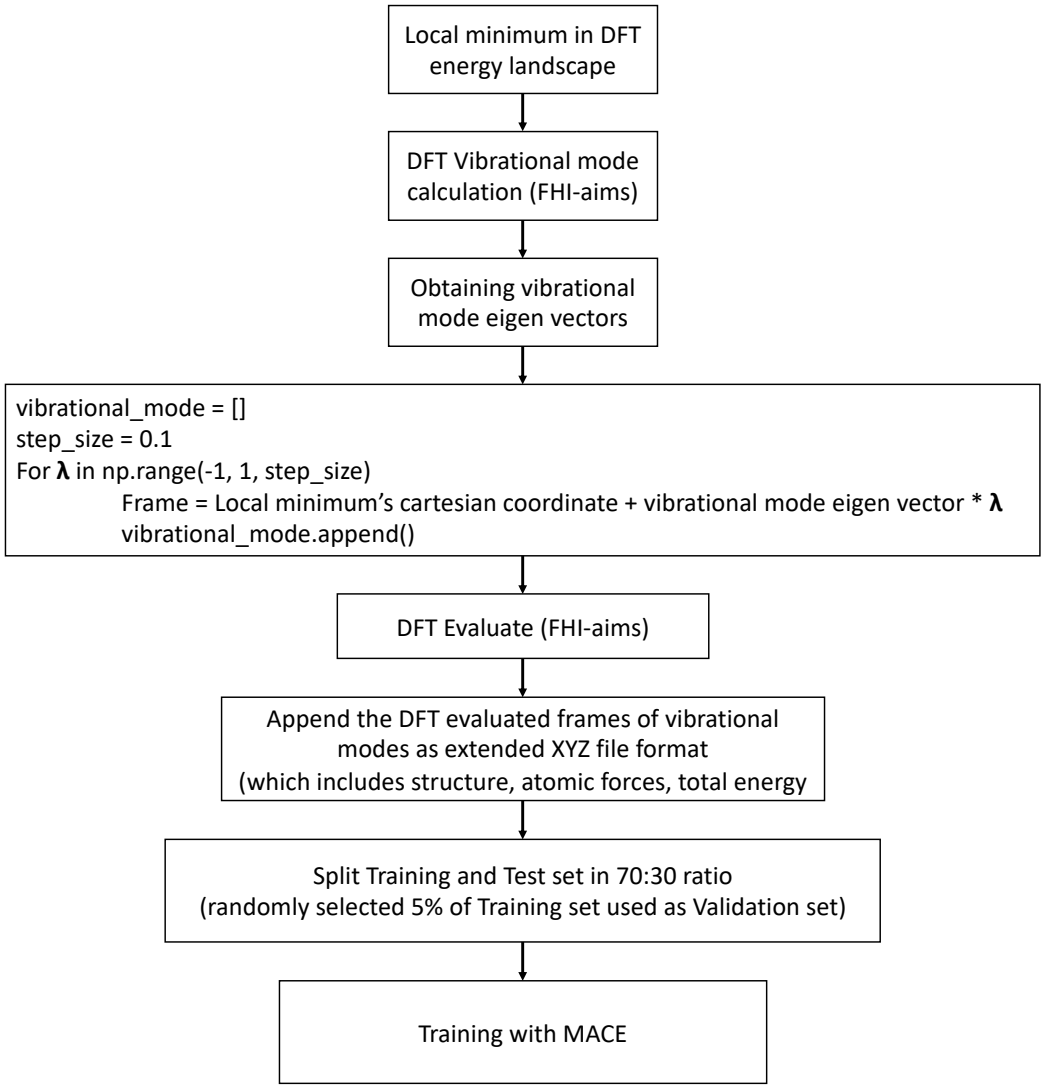

Figure **5**. Workflow of training a MACE model using local minimum images obtained by displacement along vibrational modes.

Figure 5 shows the workflow of data preparation which has been developed and implemented here for ML-IP training for applications to nanoclusters. The automated workflow scripts have been written in Python for easy customisation for later use. The LM obtained from top-down (bulk-cut) or bottom-up (global optimisation) approaches were used to train the MACE models or potentials. Firstly, we perform the vibrational mode calculation for each of the top *m* lowest energy local minima on the DFT energy landscape, for example, using the GGA PBEsol[72] functional. The calculated normalised vibrational mode eigenvectors, $\mathbf{e}_n$, are multiplied by a scaling factor $\lambda < \lambda_{max} < 1$ and added to the coordinates of the

respective local minima, which allows us to train directly on the saturated potential energy surface near local minima. The obtained images are evaluated using DFT; then the evaluated structure geometry, total energy, and the atomic forces are stored in an extended *xyz* file format. In this format, the first line specifies the number of atoms in the configuration; the second is the descriptor line, which can define the properties (metadata) that are included in the subsequent lines, using name = value pairs; the following lines provide atomic data including atomic symbols and Cartesian coordinates, and, if necessary, it can include additional properties, scalar or vector, like atomic forces. Thus, each local minimum will be represented in an extended XYZ format within a file that is used as a training/testing dataset.

2.2.4.1. Vibrational Mode

The determination of vibrational modes using DFT was performed by first calculating the electronic ground state of the system, as implemented in the FHI-aims software. Then, we systematically perturbed the atomic positions within the nanocluster by iteratively displacing each atom by ± 0.0025 Å along its respective axes. Mathematically, this approach involves expanding the system's energy around its equilibrium to the second order, a method known as the harmonic approximation. In this context, by scaling by respective atomic masses, the force constants are calculated from the second-order partial derivatives of the energy with respect to atomic displacements, forming the dynamical matrix. Without scaling, the matrix of second derivatives is called the Hessian matrix; both matrices are crucial in stability and vibrational analysis.

In the context of machine learning training, we are not primarily interested in the real vibrational spectra of the nanoclusters, but in the energy landscape itself and its reproduction (and, of course, reproducibility). Hence, in the following, we will assume that the masses of all atoms involved are unity, and therefore the dynamical and Hessian matrices are identical. The corresponding vibrational frequencies and modes are determined by diagonalising the Hessian matrix, a process that involves solving for its eigenvalues and eigenvectors. The

eigenvectors $\mathbf{e}_n$ represent the normal modes of vibration, which are the characteristic patterns of atomic movement during vibration. The vibrational frequencies, essential for understanding the cluster's dynamical behaviour, are directly related to the square roots of these eigenvalues.

Assuming the system is in equilibrium, the energy gradient, or first derivatives with respect to displacement, is zero, and the energy $E(R)$ can be approximated as:

$$E(R) \approx E(R_0) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 E}{\partial R_i \partial R_j} \Delta R_i \Delta R_j, \qquad \textit{Eqn } 20$$

where, $E(R)$ is the energy at the displaced positions R, $E(R_0)$ is the energy at the equilibrium positions R_0, $\Delta R_i = R_i - R_{0,i}$ is the displacement of atom I from its equilibrium position, $\frac{\partial^2 E}{\partial R_i \partial R_j}$ are the elements of the Hessian matrix F_{ij}.

The Hessian matrix F_{ij} is defined as:

$$F_{ij} = \frac{\partial^2 E}{\partial R_i \partial R_j}, \qquad \textit{Eqn } 21$$

which represents the second-order partial derivatives of the energy with respect to atomic positions R_i and R_j.

The vibrational frequencies ω_n and normal modes $\mathbf{u}_n$ are then obtained by solving the eigenvalue equation:

$$\mathrm{F}\mathbf{u_n} = \omega_n^2 \mathbf{u}_n, \qquad \textit{Eqn } 22$$

where, F is the Hessian matrix, ω_n^2 are the eigenvalues corresponding to the squared vibrational frequencies, $\mathbf{u}_n$ are the eigenvectors representing the normal modes of vibration.

The output from DFT calculations, in particular the vibrational mode images (structures obtained by displacement along normal modes and corresponding frequencies), provides a rich dataset for training ML-IPs. In this study, the images—capturing atomic configurations as snapshots of the potential energy landscape—along with associated forces, are integrated into machine learning algorithms to learn the underlying energy landscape. Including vibrational

information enables the ML model to reproduce energy and force fields and capture the dynamical behaviour of atoms within the system.

Incorporating formally linearly independent vibrational modes into training ML-IPs offers significant advantages for global optimisation tasks, such as identifying stable local minima on the PES. The inclusion of vibrational data enhances the ML-IP's understanding of the energy landscape away from equilibrium, i.e., the system's response to atomic perturbations. This is particularly beneficial for identifying transition states and simulating temperature-dependent phenomena like phase transitions or reaction rates.

This predictive capability is crucial for global optimisation algorithms, such as simulated annealing or genetic algorithms, which depend on accurate energy landscape evaluations to explore configurational space and identify stable structures (LM). The ability to predict energies, forces, and vibrational modes enables these algorithms to navigate the PES efficiently, avoiding high-energy regions and focusing the search on more promising areas.

Utilising vibrational mode images from DFT calculations as training data for ML-IPs provides a comprehensive approach to modelling the energy landscape. This method enhances the accuracy of energy and force predictions and endows the ML-IP with predictive capabilities for dynamical properties, facilitating effective global optimisation. As research progresses, integrating more complex non-harmonic effects and anharmonicities into ML-IP training could further improve these models' fidelity, opening new avenues in materials design and chemical reaction optimisation.

2.3. Electronic Structure Method

The Electronic Structure Method is a fundamental approach used in computational chemistry and physics to determine the properties of molecules and materials. It involves solving the Schrödinger equation [73] to obtain the electronic structure of the system, which is key to understanding how electrons behave in molecular systems:

$$\widehat{H}\Psi = E\Psi, \qquad \textit{Eqn } 23$$

In this equation, $\widehat{H}$ is the Hamiltonian operator representing the total energy of the system, Ψ is the wavefunction of the system describing the quantum state of electrons, and E is the energy eigenvalue corresponding to that state. This equation underpins the theoretical framework for predicting molecular behaviour and properties based on quantum mechanics.

Techniques like DFT make these calculations more tractable by focusing on electron density rather than individual electron interactions. This approach enables the prediction of chemical properties, reactivity, and stability with greater efficiency while still maintaining accuracy. DFT is essential for exploring the quantum mechanical behaviour of complex systems in materials science and chemistry.

2.3.1. Density Functional Theory (DFT)

DFT simplifies the complex task of tracking every electron by focusing on the electron density $\rho(r)$, which is a more manageable representation of where electrons are likely to be located within a material. The core principle of DFT is captured in the Hohenberg-Kohn theorems, which state that all ground-state properties of a system can be determined solely by its electron density.

The key equation of DFT expresses the total energy $E[\rho]$ as a functional of the electron density:

$$E[\rho] = T[\rho] + V_{ext}[\rho] + E_H[\rho] + E_{\mathrm{xc}}[\rho], \qquad \textit{Eqn } 24$$

In this equation, $T[\rho]$ represents the kinetic energy of the electrons, and $V_{ext}[\rho] = \int v_{ext}(r)\rho(r)dr$ is the energy due to the external potential $v_{ext}(r)$, such as the potential from the nuclei. The term $E_H[\rho] = \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|} drdr'$ is the Hartree energy, representing the classical electrostatic electron-electron repulsion.

The exchange-correlation energy functional $E_{\mathrm{xc}}[\rho]$ is a crucial component of DFT, encapsulating the complex quantum mechanical effects arising from electron exchange and correlation. Exchange effects arise due to the Pauli exclusion principle, which states that electrons with the same spin cannot occupy the same quantum state. This leads to an effective repulsion known as the exchange interaction, a purely quantum mechanical effect resulting from the antisymmetry of the many-electron wavefunction. Correlation occur because electrons avoid each other due to their mutual Coulomb repulsion, leading to dynamic electron correlation. This accounts for the tendency of electrons to move in such a way that reduces their mutual repulsion, which is not fully described by the mean-field approximation.

Since the exact form of $E_{\mathrm{xc}}[\rho]$ is unknown, practical DFT calculations rely on approximations for this functional. Common approximations include the Local Density Approximation (LDA), which assumes that the exchange-correlation energy at each point in space depends only on the local electron density $\rho(r)$. The Generalized Gradient Approximation (GGA) extends LDA by including the gradient of the electron density $\nabla\rho(r)$, accounting for density variations in space. Hybrid functionals incorporate a portion of exact exchange energy from Hartree-Fock theory with the exchange-correlation energy from GGA, improving the accuracy for certain systems.

In this study, we employed the DFT method using the FHI-aims software[74], utilizing appropriate exchange-correlation functionals to accurately model the electronic properties of the systems under investigation. By carefully selecting the exchange-correlation functional, we aim to capture the essential physics of electron exchange and correlation effects, which are vital for predicting accurate material properties.

2.3.1.1. Functionals

In DFT, functionals are crucial for calculating material properties. The LDA[75] is one such functional that estimates the exchange-correlation energy based solely

on the local electron density, assuming it mimics a uniform electron gas. This approximation is foundational but limited in handling variations in electron density. More sophisticated functionals like the GGA[76] improve upon LDA by including gradient corrections to the electron density, enhancing accuracy in systems with significant spatial variations.

PBEsol, a variant of GGA optimised for solids, further refines these calculations, especially for predicting equilibrium properties of solid materials and surfaces with higher precision. Thus, we used the PBEsol functional for this study of nanocluster structure prediction. The PBEsol functional is a particularly good choice for predicting the structure of solid materials because it is optimised to handle the specific electronic properties of solids, resulting in more accurate predictions of equilibrium lattice constants, bulk moduli, and surface energies [76,77]. These improvements make PBEsol highly effective for studies where precise understanding of the solid-state structure is crucial, such as in materials science and condensed matter physics. This optimisation ensures that PBEsol can deliver more reliable results for the physical properties of solids compared to other functionals that might not specifically account for the unique characteristics of solid materials.

By mixing exact exchange with GGA exchange in this proportion, hybrid functionals (*e.g.*, PBEsol0) aim to correct for the self-interaction error present in pure DFT methods, leading to improved accuracy in calculating electronic properties, band gaps, and reaction barriers. The inclusion of 25% exact exchange helps to better represent the exchange interactions, particularly in systems where electron localisation and exchange effects are significant[78,79].

While hybrid functionals are computationally more demanding than GGA functionals due to the inclusion of non-local exact exchange calculations, they offer significant benefits in systems where a precise electronic structure description is crucial, such as in molecules, low-dimensional materials, and systems with strongly correlated electrons. However, for large-scale calculations of solids, where computational cost is a significant concern, functionals like PBEsol are often preferred due to their balance between accuracy and efficiency.

2.3.1.2. Basis Sets

Basis sets in DFT are crucial tools used to describe the electronic wavefunctions in molecular simulations. FHI-aims employs numerical atom-centred basis sets, which are structured into various precision levels or "tiers" like light, intermediate, tight, or very tight. Each tier offers a balance between computational efficiency and accuracy of the calculations. The choice of a specific tier impacts the level of detail at which electrons' behaviour is modelled and the overall accuracy of the simulation results.

Lighter basis sets facilitate easier convergence but with less precision, while tighter sets provide higher accuracy at the cost of more complex calculations. Users can adjust these sets by adding or removing functions to tailor the simulation to specific requirements, ensuring an optimal balance between detailed representation and computational demands. For instance, a light basis set typically provides sufficient detail for determining molecular structure and basic properties like vibrational frequencies. On the other hand, a tight basis set is better suited for more precise calculations, such as reaction energies and electronic distributions, due to its more comprehensive inclusion of orbital interactions. Therefore, we chose to use a light basis set, which is implemented in the FHI-aims code, because it is well-suited for predicting structures and offers a cost-effective solution for computational purposes.

2.4. Local Optimisation Algorithms

The following section delves into the intricate techniques employed to navigate the vast expanses of an energy landscape during global optimisations. First, we focus on the strategy of identifying a proximate local minimum, starting from a specific point on the PES. To reiterate, the PES for a given three-dimensional nanocluster structure is defined over a high-dimensional space of 3N-6 dimensions; where N is the number of atoms in the system. The subtraction of

six accounts for the three translational and three rotational degrees of freedom that do not affect the internal energy of the system. A neighbouring local minimum within this hypersurface can be pinpointed through the application of local optimisation algorithms. Such local optimisation algorithms are designed to locate a local minimum by navigating "downhill" on the PES, starting from an initial configuration. These algorithms are categorised based on the highest-order derivative they utilise in the optimisation process. For instance, a method that exclusively employs first derivatives is categorised as a first-order method.

Consider an energy function, denoted by f, which depends on a set of independent variables, designated as r_1, r_2, … ,r_i. These variables correspond to the atomic positions in a nanocluster. At a LM, the first derivatives of the energy function with respect to position are expected to be zero, indicating a stationary point. Additionally, the Hessian matrix (the matrix of second derivatives) should be positive definite, leading to positive eigenvalues and corresponding real vibrational frequencies, which indicate a stable configuration.

2.4.1. Zeroth-Order Method: Powell's method

Powell's method[80] is a well-established zeroth-order optimization algorithm widely used in computational chemistry for local optimization tasks where derivative information is unavailable or unreliable. It is particularly effective for finding LM of functions in multidimensional spaces by performing a series of one-dimensional minimizations along carefully chosen directions, without the need for gradient or Hessian calculations.

The key idea of Powell's method is to minimize a multivariate function by successively performing line minimizations along a set of directions that are updated iteratively. Starting from an initial point and a set of initial search directions (generally the coordinate axes), the method optimizes the function along each direction in turn. After completing a cycle of line minimizations, the algorithm updates one of the directions based on the movement made during the cycle, allowing the method to adapt to the contours of the function being optimized.

Mathematically, for an energy function $f(\mathbf{x})$ dependent on N variables (for example, where $\mathbf{x}$ represents the atomic positions in a chemical system), Powell's method proceeds as follows; Beginning with an initial point $\mathbf{x}_0$ and a set of N linearly independent direction vectors $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N\}$, the method performs a series of line minimizations along each direction. For each direction $\mathbf{d}_i$, it finds the scalar λ_i that minimizes the function along that direction:

$$\lambda_i = arg\min_{\lambda} f(\mathbf{x}_{i-1} + \lambda \mathbf{d}_i), \qquad \textit{Eqn } 25$$

updating the position to:

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \lambda_i \mathbf{d}_i, \qquad \textit{Eqn } 26$$

where $\mathbf{x}_{i-1}$ is the position from the previous step. After completing line minimizations along all directions, the total displacement is computed as:

$$\mathbf{p} = \mathbf{x}_N + \mathbf{x}_0, \qquad \textit{Eqn } 27$$

where $\mathbf{p}$ represents the total displacement made during a complete cycle of line minimizations along N directions, and an additional line minimization is performed along $\mathbf{p}$ to find $\lambda_{\mathbf{p}}$ that minimizes the function:

$$\lambda_i = arg\min_{\lambda} f(\mathbf{x}_N + \lambda \mathbf{d}_i), \qquad \textit{Eqn } 28$$

updating the position to:

$$\mathbf{x}_{N+1} = \mathbf{x}_N + \lambda_N \mathbf{p}, \qquad \textit{Eqn } 29$$

One of the original directions is then replaced with $\mathbf{p}$, usually the direction along which the greatest improvement was made during the cycle. This iterative process continues, with the method updating directions and performing line minimizations, until convergence criteria are met, such as when changes in the function value or the variables fall below predefined thresholds.

Powell's method effectively navigates the energy landscape to identify stable configurations by relying solely on function evaluations. Its ability to update search directions dynamically allows it to adapt to the contours of the potential energy surface, improving convergence towards the minimum.

While Powell's method is robust for smooth functions, it may encounter difficulties with functions that have discontinuities or are not well approximated by quadratic forms. Additionally, in high-dimensional spaces, the number of function evaluations can become significant, and the method may become less efficient compared to gradient-based approaches if derivative information is available.

In computational chemistry, Powell's method is frequently used for molecular geometry optimization, adjusting atomic positions to find the lowest-energy conformation of a molecule. It is particularly suitable for optimizing nanocluster structures without requiring gradient information, making it valuable when analytical gradients are unavailable or computationally expensive to obtain.

2.4.2. First-Order Method: Conjugate Gradient Method

The Conjugate Gradient method[81] is a powerful first-order optimisation technique originally developed for solving systems of linear algebraic equations, specifically those with a symmetric and positive-definite coefficient matrix A. However, its application extends beyond this, proving particularly effective in the optimisation of nonlinear functions when adapted appropriately.

Solving the linear system $A\mathbf{x} = \mathbf{b}$, the conjugate gradient algorithm operates as follows:

1. Initialisation:
    - ✓ Choose an initial guess $\mathbf{X}_0$.
    - ✓ Compute the initial residual $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$.
    - ✓ Set the initial search direction $\mathbf{d}_0 = \mathbf{r}_0$
2. Iteration: For each iteration k:
    - ✓ Compute the step size: $\alpha_k = \frac{r_k^T r_k}{d_k^T A d_k}$, *Eqn* 30
    - ✓ Update the solution: $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k d_k$, *Eqn* 31

- ✓ Compute the new residual: $r_{k+1} = r_k - \alpha_k A d_k$, *Eqn* 32
- ✓ Check for convergence: If $\|r_{k+1}\|$ is sufficiently small, stop the algorithm.
- ✓ Update the search direction: $\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$, $\mathbf{d}_k = \mathbf{r}_k + \beta_k \mathbf{d}_k$.
- ✓ Check for convergence: If $\|r_{k+1}\|$ is sufficiently small, stop the algorithm.

In optimisation of nonlinear functions, $f(\mathrm{x})$, the Nonlinear Conjugate Gradient method is used, where the gradient, $\nabla f(\mathrm{x})$, replaces the residual, $\mathbf{r}_\mathrm{k}$, and line searches are performed to find appropriate step sizes α_k.

The conjugate gradient method operates by performing a series of linear searches along conjugate directions, where the concept of "conjugacy" is tied to the specific function being optimised. These conjugate directions provide a way to consider the history of past search directions, thereby ensuring efficient progress towards the minimum.

2.4.3. Second-order Method: Newton's method and Quasi-Newton Method

Newton's method, a second-order optimisation algorithm, has been a cornerstone in the field of numerical optimisation due to its robustness and rapid convergence properties. This method utilises both the first and second derivatives of the objective function to identify points where the gradient is zero, which correspond to potential minima or maxima. When generalised to multidimensional problems, it is often referred to as the Newton-Raphson method [82]. Mathematically, the iterative scheme of Newton's method can be expressed as:

$$x_{n+1} = x_n - [Hf(x_n)]^{-1} \nabla f(x_n), \quad \textit{Eqn } 33$$

where x_n is the current estimate, $Hf(x_n)$ is the Hessian matrix (the matrix of second derivatives) at x_n and $\nabla f(x_n)$ is the gradient of the function at x_n.

However, Newton's method presents a significant computational challenge when applied to high-dimensional problems. The calculation and inversion of the Hessian matrix become increasingly demanding as the dimensionality of the problem grows. This computational burden often renders Newton's method impractical for large-scale problems, such as those frequently encountered in computational chemistry.

To circumvent this obstacle, a class of methods known as Quasi-Newton methods has been developed. These methods retain the rapid convergence of Newton's method but alleviate the need for explicit second derivative computation. Instead, they construct an approximation to the inverse Hessian matrix using gradient evaluations at successive iterations. The iterative update in Quasi-Newton methods can be generally expressed as:

$$x_{n+1} = x_n - \alpha_n \beta_n \nabla f(x_n), \qquad \textit{Eqn } 34$$

where x_n is the current estimate, α_n is the step size, β_n is the approximation to the inverse Hessian matrix, and $\nabla f(x_n)$ is the gradient of the function at x_n.

In essence, Quasi-Newton methods offer an attractive balance between computational efficiency and convergence speed. They have proven to be particularly effective for optimisation problems where the explicit computation of the second derivative is either computationally prohibitive or not feasible. The BFGS (Broyden-Fletcher-Goldfarb-Shanno) method[83–86] is a well-known Quasi-Newton method that has found extensive application in computational chemistry.

2.4.3.1. Second-order Method: Rational Function Optimisation method

The Rational Function Optimisation (RFO) method [87] represents a specialised adaptation of Newtonian methods, particularly useful for locating transition states on potential energy surfaces, a task of paramount importance in the study of chemical reactions.

The RFO method modifies the standard Newtonian approach by introducing a rational function to represent the step direction. This modification is particularly beneficial when dealing with potential energy surfaces, as it allows for a more accurate representation of the curvature near the transition state, preventing overshooting or stepping into regions of incorrect curvature.

The RFO method involves several key steps:

1. Approximation of the Hessian (B): Like other Quasi-Newton methods, the RFO method constructs an approximation to the Hessian matrix using gradient evaluations at successive iterations. This approximation is updated at each step based on the difference in gradients and the steps taken, using suitable update formulas.

2. Calculation of the Gradient (g): The gradient of the objective function, representing the first derivatives with respect to the variables, is computed at the current point. In computational chemistry, this often involves calculating the forces on the atoms in the system.

3. Determination of the Step Direction (p): The RFO method determines the step direction by solving a generalised eigenvalue problem. This involves constructing a matrix that includes both the approximate Hessian and the gradient, and finding the eigenvector corresponding to the desired eigenvalue (typically the lowest). This eigenvector provides the direction in which to take the next step.

4. Calculation of the Eigenvalue (λ): The eigenvalue corresponding to the step is found, which influences the magnitude and direction of the step. This eigenvalue problem is solved using standard numerical linear algebra techniques.

Mathematically, the RFO method can be expressed as solving the following generalised eigenvalue problem:

$$\begin{bmatrix} B & g \\ g^T & 0 \end{bmatrix} \begin{bmatrix} p \\ -1 \end{bmatrix} = \lambda \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p \\ -1 \end{bmatrix}, \qquad \textit{Eqn } 35$$

where B is the approximate inverse Hessian, g is the gradient, p is the step direction, and λ is the smallest eigenvalue.

The RFO method, though beneficial for its computational efficiency and fast convergence, is particularly tailored for locating transition states rather than for general optimisation towards minima. This technique enhances the analysis of potential energy surfaces, aiding in the prediction of transition states and the study of molecular-level chemical reactions.

2.5. Global optimisation method

The structure prediction of nanoclusters is a complex yet crucial aspect of computational chemistry[88]. The difficulty of this task is amplified by the existence of a plethora of potential structural configurations, each of which can lead to different chemical and physical properties. As the size and complexity of the clusters increase, the energy landscape becomes more convoluted, necessitating the use of advanced techniques to efficiently and accurately determine the optimal structures. A widely acknowledged approach to tackle this challenge is through the application of global optimisation methods. This chapter will provide a review and analysis of the primary global optimisation techniques employed in nanocluster structure prediction.

Global optimisation techniques provide a systematic way to explore the multidimensional energy landscapes associated with nanoclusters. These methods seek to identify the lowest energy configurations—the global minima—by efficiently navigating the energy landscape and avoiding entrapment in local minima. Several classes of global optimisation techniques have been developed and used in the field of computational chemistry, each having its strengths and limitations. In this study, we focus on five distinct techniques: data mining, stochastic approaches (random quenching or Monte Carlo deterministic

quenching), Monte Carlo Basin Hopping, Simulated Annealing, and Genetic Algorithms (GA).

Data mining, an information processing technique, is employed to explore large datasets of pre-computed or experimental data, effectively identifying promising structural candidates. The stochastic approach, or random quenching, relies on random perturbations of atomic positions followed by local relaxation to explore the energy landscape. Monte Carlo Basin Hopping[86] adds a further level of sophistication by introducing thermal fluctuations to navigate between different energy basins. Simulated Annealing[87], a method inspired by the annealing process in metallurgy, utilises a controlled cooling schedule to decrease the system's temperature gradually, allowing it to settle into a state of minimum energy. Finally, genetic algorithms draw inspiration from natural evolution, using processes such as selection, crossover, and mutation to evolve a population of structures towards optimal solutions. These methods, in various combinations or alone, have demonstrated considerable success in predicting the structure of nanoclusters. The following sections will outline each method's principles, implementation, and specific applications in nanocluster structure prediction.

2.5.1. Challenges of Global Optimisations

When it comes to computationally defining the structure of a nanocluster with a prescribed size and composition, it proves beneficial to begin with a minimum amount of prior data. This approach helps mitigate biases and enables a thorough investigation of the cluster's PES to identify many potential low-energy structures, known as LM. This practice minimises the introduction of biases and affords a comprehensive probe into the cluster's PES to find probable low-energy structures.

This procedure unfolds in multiple stages. First, there is a need for an energy function to evaluate the energies of different structures; in the present research, we utilise IPs and DFT for this purpose. Subsequently, any speculative structure is incrementally fine-tuned to a local minimum in its proximity to establish its energy using an appropriate local optimisation algorithm. Ultimately, to ensure a comprehensive exploration of the conformational space (the PES) and to increase

the likelihood of locating all the low-energy local minima, a robust algorithm is necessary. Given that a cluster's PES spans a 3N-6-dimensional space for any non-linear configuration—where N is the number of atoms in the cluster—the full-scale exploration is not feasible due to the exponential increase in possible configurations with increasing N.

We operate under the assumption that the structure bearing the least energy is the most probable to be experimentally detected at 0 K temperature. Such a structure can be procured through an adequately exhaustive global search algorithm working in tandem with a sufficiently precise energy function. However, given the extensive computational demands of global optimisation algorithms to locate all pertinent local minima and the considerable time requirements of precise energy functions to correctly rank local minima, the present research adopted a sequential approach for the examination of aluminium fluoride nanoclusters' structures. The initial phase involved the use of an evolutionary algorithm (EA) employed along with IPs for global optimisation, and the succeeding phase involved the application of DFT to the prominent structures identified for each cluster size.

2.5.2. Data Mining

Data mining employs existing structural data for analogous systems of the same stoichiometry as a point of departure in structural prediction. Through the careful analysis of these pre-existing structures, it identifies potential candidate configurations that may be optimal or near-optimal for the target system.

The primary principle underpinning the data mining approach is the notion that the atomic structures of materials with similar stoichiometry will likely share similar geometric configurations. These similar geometric configurations can serve as starting points for the optimisation of the target system.

Upon identifying these potential candidate structures, the data mining method proceeds to the atom substitution stage. Here, the atoms in the identified structures are systematically replaced with the atoms of the target compound. It is key to note

that this atomic substitution is performed while adjusting the original geometric configuration to take into account the target system's properties, such as ionic or atomic sizes and bond lengths.

Once the substitution is complete, the new structures are subjected to an optimisation process. This process typically employs local optimisation methods to refine the atomic positions and achieve a lower energy state. It is during this phase that the unique properties of the target atoms and their interactions come to the fore, driving the structure towards the optimal geometry for the target system.

To summarise, the data mining approach leverages existing structural information to inform the search for optimal structures in a target system. This method, although straightforward, can be remarkably effective, particularly when dealing with systems that have a high degree of structural complexity. However, it should be borne in mind that the success of this approach hinges upon the availability and quality of data for systems with analogous stoichiometry. As we proceed further into this chapter, we will explore in more detail the intricacies, benefits, and limitations of using data mining in the structure prediction of nanoclusters.

2.5.3. Monte Carlo Deterministic Quenching

The stochastic approach is a widely adopted computational method for predicting the structure of nanoclusters, being one of the simplest methods. This method is used for the AlF_3 nanoclusters in our study.

Monte Carlo Deterministic Quenching (MCDQ) is a specific technique within the broader field of stochastic optimisation methods. The basic idea of this method is to generate random configurations of atoms and then to optimise these configurations using a certain energy function—such as an IP or DFT—to reach lower energy states. In the context of AlF_3 nanoclusters, the process starts by randomly placing the Al and F atoms in a predefined space; we have used a cubic simulation box for this purpose. This initial random placement of atoms introduces

the 'Monte Carlo' aspect to the quenching method, ensuring that the process explores a wide variety of potential structural configurations.

Following the random placement of atoms, a process known as energy minimisation or 'quenching' is conducted. Quenching involves the use of iterative algorithms which seek to adjust the positions of atoms, in a series of steps, to minimise the total energy of the system. Optimising the randomly generated structure directly on the PES can be challenging. For example, if a random quenching algorithm generates a structure with extremely short or long interatomic distances, the optimisation may fail or could require exceedingly large computational costs to run to completion.

Thus, the global optimisation code we used for this study, Knowledge Led Master Code (KLMC), checks all pairs of interatomic distances, and if the distance is too large or too short, it discards the fragmented or collapsed structures. With this algorithm, we can save a notable amount of computational cost by accepting only the structures that are physically reasonable and relatively near a LM. Each iteration of this process results in further characterisation of the PES, as each point on the surface represents a possible stable or metastable configuration of the nanocluster. The goal is to find the global minimum on this surface, which corresponds to the most stable, lowest-energy configuration.

The MCDQ algorithm initially generates random structures within the finite simulation box, which means the algorithm can, in principle, explore a wide region of the PES for the target system. The beauty of the MCDQ method lies in its simplicity and effectiveness. It offers a relatively straightforward approach for exploring a vast array of possible configurations for complex systems, such as nanoclusters. In practice, the algorithm can be more efficient for smaller cluster systems compared to the more computationally expensive algorithms introduced in the following sections. However, it is inefficient for finding the lowest number of LM for larger nanoclusters due to the vastness of the configuration space or PES. The process may not always lead to the absolute GM, considering the enormous number of possible configurations in a multi-atomic system.

2.5.4. Genetic Algorithm

Genetic Algorithms [9,89,90] (GAs) or evolutionary algorithms, also known as evolutionary algorithms, are a class of optimisation methods inspired by the principles of natural evolution, namely selection, crossover (or recombination), and mutation. They have been widely applied in various fields, including computational chemistry, where they are used for exploring the PES of complex systems such as nanoclusters[91] and solid solutions. The fundamental concept behind GAs is the evolution of a population of individuals, each representing a potential solution to the optimisation problem. In the context of energy landscape search, an individual corresponds to a specific configuration of the system, and its fitness is determined by the potential energy of this configuration. The goal is to evolve the population over successive generations to find the global minimum or near-global minimum of the energy landscape, which represents the most stable configuration of the system.

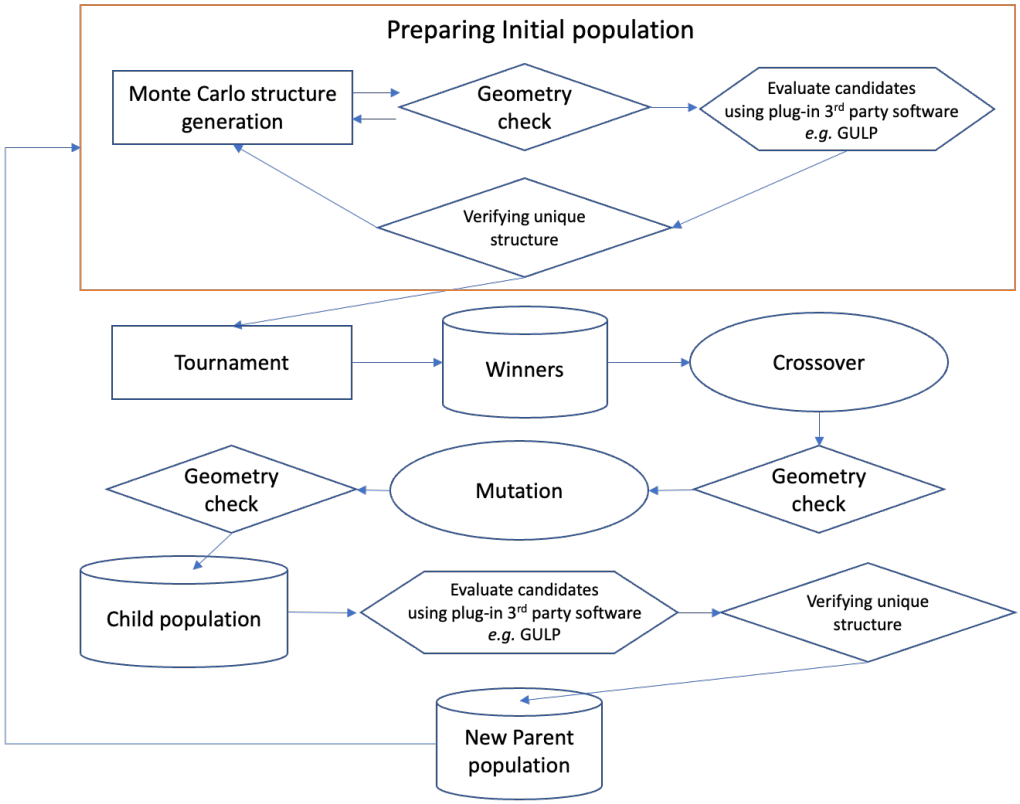

Figure **6**. Schematic of the GA implemented in the KLMC software[9,89,91,92].

A GA starts with an initial population of nanocluster structures, which can be generated randomly—as in the random quenching described in the earlier section—or based on some prior knowledge. The GA then proceeds by iteratively applying the operations of selection, crossover, and mutation, as shown in the scheme of the operations implemented in the KLMC GA algorithm [9,89,90,93] (see Figure 6).

Selection is the process of choosing individuals from the current population to produce offspring for the next generation. The selection is typically biased towards individuals with lower potential energy, reflecting the principle of "survival of the fittest". Various selection methods can be used, such as tournament selection or roulette wheel selection.

Crossover is the operation that generates new individuals by combining parts of two parent individuals. In the context of nanoclusters, this is done in KLMC by crop the structures in the parent clusters and merge the atomic cropped parents structures, effectively swapping parts of the clusters. The resulting offspring inherit characteristics from both parents, which allows the GA to explore new regions of the PES.

Mutation is the operation that introduces random changes in the individuals. This is typically done by randomly altering the position of one or more atoms in the configuration. Mutation introduces diversity into the population and helps prevent the GA from getting stuck in local minima of the PES.

The new generation of individuals produced by selection, crossover, and mutation then replaces the old generation, and the process is repeated until a stopping criterion is met. This could be a maximum number of generations, a target potential energy, or a measure of convergence of the population (the convergence of the GA calculation will be discussed in the 3. Aluminium Fluoride chapter, figure 12).

For this study, we employed the GA as implemented in the KLMC software. KLMC employs a hash-key method[89,94], which generates encrypted string based

on the atomic connections, so based on the encrypted string the KLMC distinguish whether the structure has already found before or not. Using a tool like hash-key helps the GA calculation to be more efficient, as it maintains a range of diverse nanocluster structures by avoiding duplicate configurations.

Once KLMC achieves a sufficient population size in the initial generation, the clusters undergo crossover and mutation operations, and then the modified clusters are relaxed using local optimisation methods. Note that during crossover, care is taken to ensure that the system size (stoichiometry, number of atoms in the cluster) does not change, preserving the chemical composition of the clusters.

From the comparison of all structures—including the optimised structures that were randomly generated and the GA-generated structures—a certain number of top-performing structures are selected and passed to the next generation to produce offspring. In our implementation, this set number of structures is determined by the tournament size, which refers to the number of individuals participating in the selection process for reproduction. This selection is typically biased towards individuals with lower potential energy, again reflecting the principle of "survival of the fittest".

GA is a powerful tool for global optimisation in computational chemistry. By mimicking the process of natural evolution, GAs can effectively navigate the complex PES of nanoclusters and other systems, finding the global minimum that represents the most stable configuration. However, the performance of a GA depends critically on the choice of parameters and operations, such as population size, mutation rate, crossover rate, and selection method, which need to be carefully tuned for the specific system and problem at hand.

2.5.5. Method for Constructing Clusters Cut from the Bulk phase of AlF_3.

There are clearly a number of different approaches to constructing stoichiometric clusters cut from a bulk phase including [a] making random cuts

under the constraint that the cluster is not fragmented; [b] make a spherical cut from a chosen centre; [c] make a cut based on the Wulff's construction (relative surface energies); and [d] make a cut based on the primary or secondary building units of the bulk phase. Each have the constraint that the cut should have the correct composition, and here we consider approach [d] for generating clusters of AlF_3.

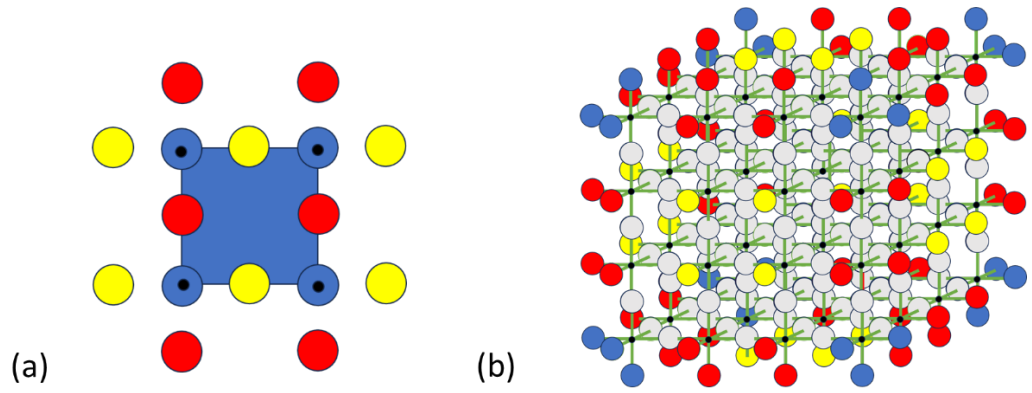

Figure **7**. Cubic constructions of corner sharing AlF_3 octahedra shown as: (a) a plan view of a 2×2×2 cluster of 8 octahedra with red, yellow, and blue circles representing F atoms, black circles representing Al atoms, and the blue square the area of one face of the cube defined by the Al atoms. (b) A ball and stick model of a 4×4×4 cluster of 64 octahedra with grey balls representing shared F atoms, red, yellow, and blue circles representing singly coordinated F atoms to just one Al at a vertex, edge and face of the cube formed by the Al atoms, and black circles representing Al atoms. Taken from reference[95].

The bulk phases of AlF_3 have AlF_6 secondary building units (SBU) that are octahedra which, in the stable phase, are corner sharing. It thus seems natural to construct clusters composed of a cubic arrangement of N^3 octahedra. Given we are using SBU, to achieve the correct stoichiometry several F atoms will need to be removed from the surface (assuming this costs less than removing a F atom from within the cluster) of the initial cubic construction. For the simplest case of $N = 1$, three of the six F atoms should be removed. There are two inequivalent ways to achieving this; assuming each F atom pair that has the central Al atom directly between them (on the same axis) is coloured, for example, red, blue, and yellow,

then either take one F atom from each colour or take two from one colour and one from another. The former creates a tetrahedron whereas the latter creates a planar cluster, but of course both could relax to the same LM.

To calculate the formula for the initial cubic construction, it is prudent to first consider the $N = 2$ example in figure 7(a). Each Al atom, or centre of an octahedron, is marked by a small black circle and the F atoms are coloured in three equal sized, symmetry equivalent sets. Each set is formed of the same number of Al atoms in each direction except for one where an additional F atom is required, i.e. for each colour there is a set of N by N by N+1 F atoms. Thus, for cuts of N^3 corner-sharing AlF_6 octahedra we have AlF_6, Al_8F_{36}, $Al_{27}F_{108}$, $Al_{64}F_{240}$, … that require removing 3, 12, 27, 48, … F atoms for $N = 1, 2, 3, 4, \ldots$ respectively.

Although it might seem obvious that one F atom from each vertex should be removed, this only accounts for 8 of the 12, 27, 48, … F atoms that need to be removed. The selection of which F atoms to next remove is less obvious although ensuring an even spread may also avoid the desirable creation of a dipole across the cluster. Clearly, several candidate structures for a chosen size should be created and after relaxing these, their formation energies compared to determine the loosest energy bulk cut cluster of that size.

Considering the $N = 4$ example, there are 80 singly coordinated F atoms: three outer F atoms attached to each of the Al atoms sited on one of the eight vertices of the cubic Al lattice; two outer F atoms attached to each of the two Al atoms sited on each of the twelve edges of the cubic Al lattice; and 1 outer F atom attached to each of the 4 Al atoms sited on one of the 6 faces of the cubic aluminium lattice; in figure 7(b) these F atoms are coloured blue, red, and yellow, respectively. Different systematic removal of 48 F atoms can be tried based on trying to keep maximum symmetry, for example, remove (a) one F atom from each of the 48 Al edge atoms; (b) all F atoms from all vertices and faces; (c) all F atoms from all vertices and one F atom from half of the edges; (d) two F from all vertices and one F atom from 32 of the 48 edges. The numbers for the first two of these examples seem to work, but this is not always the case for other sized clusters, and we end up with mismatched like that of the last two examples, where it is not obvious which F atoms to remove from the edges. The choice of these, or indeed

all F atoms that need to be removed, can be randomly chosen to generate the different candidate structures for each cluster size.

3. Aluminium Fluoride

3.1. Introduction

In this chapter, the computational prediction of the structures and properties of aluminium fluoride (AlF_3) nanoclusters will be presented. AlF_3 is widely used in different industrial and scientific areas, attracting considerable attention because of its distinct structural, chemical, and physical characteristics[96,97]. Despite its importance, our understanding of the potential complexity and diversity of AlF_3 nanocluster structures, including their atomic structure, remains limited. This research gap prompted the initiation of this study, aiming to utilise computational techniques to predict the structural configurations and properties of AlF_3 nanoclusters.

The science of nanoclusters stands at the intersection of atomic and bulk phase characteristics, thus offering a distinctive platform for understanding material behaviour and properties on the nanoscale[5,88,98]. The complexity of nanocluster structures has historically posed considerable challenges to predictive modelling, as their properties are influenced by size, shape, composition, and the interactions between individual atoms or molecules. Addressing these complexities required the integration of sophisticated computational algorithms capable of accurately predicting the likely structures of AlF_3 nanoclusters.

Our study incorporates computational methodologies, utilising both classical IP and quantum mechanical DFT models, to provide novel insights into AlF_3 nanocluster structure prediction. The aim is to generate a comprehensive catalogue of potential AlF_3 nanocluster configurations that, when synthesised, would exhibit specific properties optimal for applications—for example, as Lewis acid catalysts. This research not only enhances our understanding of AlF_3 nanocluster structures but also serves as a framework for predicting the structures of other complex nanomaterials, and surfaces of materials for nanoclusters of compounds with a 1:3 stoichiometry.

Unlike systems with 1:1 and 1:2 stoichiometry, systems with 1:3 stoichiometry, like that of AlF_3, are a relatively new field of study for nanoclusters.

There are more atoms per formula unit for 1:3 compared to the 1:1 stoichiometric systems. Thus, compared to 1:1 stoichiometric systems like ZnO, MgO, and CuO, the PES for 1:3 stoichiometric systems has double the number of dimensions, which requires significantly more computational resources to search[9]. Fortunately, there is a possibility to exploit previous work, as one of the first studies on 1:3 nanoclusters—that of lanthanide fluoride (LaF_3)—has been completed by our research group, and the structural data are available in the WASP@N database[99–101].

There are many factors to be considered to understand the Lewis acidity of materials [96,97]. AlF_3 is theoretically proven to have the highest Lewis acidity if we consider only the chemical factors, excluding structural factors. However, structural features on the surface of AlF_3, where F atoms hinder the potential Lewis acid active sites (Al atoms), limit its acidity[96]. One of the most significant characteristics of nanoclusters is their exceptionally high surface-to-volume ratio compared to higher-dimensional materials (rods, slabs, and bulk phases). Practically, nanoclusters are the ideal system for most catalytic activities, as most of the atoms are accessible to the outer environment, and the quantity of material per active site is minimised.

For optimal catalytic performance, it is crucial to understand the interactions between clusters and surfaces. As a preliminary step, we aim to identify the structures and characteristics of an isolated cluster within a vacuum environment and try to understand the relationships between the structures and properties. These established structures can then serve as the fundamental basis for developing models that effectively illustrate cluster–surface interactions. Thus, here a complete search of the PES of $(AlF_3)_n$ nanoclusters in vacuo has been conducted for sizes n from 1 to 11. A straightforward global optimisation method, MCDQ was employed for the smaller clusters (n = 1 to 6), and a GA was employed for the larger clusters (n = 7 to 11). The global optimisation algorithms are implemented in the in-house software, the Knowledge Led Master Code (KLMC).

The results presented in this chapter provide a detailed account of our findings. The computational challenges encountered during the prediction process are discussed, and the solutions implemented to overcome these obstacles. Then, a

discussion of the various potential structures identified is presented alongside a comparison of them based on their stability and electronic properties. By presenting a thorough analysis of the results, we hope to deepen the understanding of AlF_3 nanoclusters.

3.2. The Interatomic Potentials for Aluminium Fluoride

3.2.1. Interatomic Potential for Short-Range Al-F Interactions

In our choice of IP, we have considered several options, namely the two-body Buckingham potential and the polarisable ion model (PIM) as proposed by Jackson[102] and Madden[100], respectively. For ease of reference, the IP that includes the Buckingham potential with the potential parameters refined by Jackson et al. will henceforth be referred to as the Jackson potential. The Jackson potential was fitted to structural parameters of aluminium fluoro-/hydroxy-silicate (topaz, $Al_2SiO_4(OH, F)_2$) for the study of doped and defect topaz. The parameters for the PIM were adjusted to characterise structures larger than nanoclusters, including nanoparticles and bulk α-AlF_3. While the PIM may yield physically accurate results for systems larger than nanoclusters, to our knowledge the PIM has not yet been applied to modelling nanoclusters. The PIM has complex many-body features and so requires more computational resources for structure optimisation than the two-body Jackson potential. Moreover, the PIM is not implemented in the General Utility Lattice Program (GULP) code[103–105], nor is it currently linked to our choice of global optimisation code, KLMC. With greater availability of computer resources than when the Jackson potential was originally refined, here we investigate if this IP can be improved with respect to predicting energy rankings that better match those obtained using DFT energies.

To facilitate an automated search over the potential parameter space of an IP, we utilized a custom-built automation tool called "What-IP". One module of What-IP operates in conjunction with the GULP software to compute, or map out, the sum-of-squares (SOS) as a function of the IP parameters. The SOS is a quantitative measure of the difference between the observed (experimental or

reference) and calculated values of various properties, serving as an objective function to be minimized during the potential fitting process.

Mathematically, the SOS is calculated using the following equation:

$$SOS = \sum_i w_i (P_i^{calc} - P_i^{obs})^2, \qquad Eqn\ 36$$

In this equation, P_i^{calc} represents the calculated value of property *i* using the current set of IP parameters, and P_i^{obs} is the observed or reference value of the same property. The term w_i is a weighting factor assigned to property *i*, reflecting its relative importance in the fitting process. The summation runs over all the properties *i* that are considered in the fitting procedure.

The properties P_i can include various structural and energetic quantities, such as lattice parameters, bond lengths, bond angles, elastic constants, and atomic forces. Depending on whether the reference (or target) structure is relaxed or not, the SOS calculation focuses on different sets of properties. In the case of a relax fit, where the reference structure is allowed to find its lowest energy configuration, the SOS includes differences in structural parameters between the calculated and observed structures. This involves comparing quantities like optimized lattice constants and internal coordinates. For a standard fit, where the reference structure is not relaxed, the SOS is based on the sum of atomic forces acting on the atoms, ensuring that the calculated forces closely match the reference forces derived from experiments or higher-level calculations.

In our study, the reference system is the observed bulk phase of aluminum fluoride, specifically α-AlF_3. By systematically varying the IP parameters and calculating the SOS for each parameter set, GULP, in conjunction with What-IP, generates a landscape of SOS values over the parameter space. This process involves several steps:

First, a range of values for each IP parameter is selected to thoroughly explore the parameter space. For each set of parameters, GULP computes the properties P_i^{calc} of interest based on the IP being tested. Using *Eqn* 36, the SOS is then calculated by comparing these calculated properties with the reference values P_i^{obs}, applying the appropriate weighting factors w_i to each property. The calculated

SOS values, along with the corresponding IP parameters, are recorded and stored in a spreadsheet file for further analysis.

Once computed, the SOS data serve as a valuable resource for analyzing the sensitivity of the IP parameters and identifying optimal parameter sets that minimize the SOS, indicating the best agreement between calculated and observed properties. By plotting the SOS values against the IP parameters, we can create contour map, such as the one shown in Figure 8. These visualizations illustrate how the SOS varies across the parameter space, highlighting regions where the IP parameters yield the most accurate results. Such visual aids are instrumental in understanding the fitting landscape and guiding the selection of IP parameters that produce reliable and precise simulations of the material.

We used WHAT-IP to map out the SOS as a function of the two parameters of the Born-Mayer potential for Al-F short-range interactions. As mentioned above, α-AlF_3[106] is the chosen reference system. An exhaustive exploration of the SOS landscape was carried out across a grid with a spacing of 10 eV and 0.01 Å for the A and ρ parameters, up to a maximum of 4000 eV and 0.4 Å, respectively.

As mentioned above, α-AlF_3[106] is the chosen reference system. An exhaustive exploration of the SOS landscape was carried out across a grid with a spacing of 10 eV and 0.01 Å for the A and ρ parameters, up to a maximum of 4000 eV and 0.4 Å, respectively.

Our investigation employed two distinct fitting strategies: a relax fit and a standard fit. The global minimum found for the SOS when implementing a relax and standard fit, with the reference data including the structural parameters or atomic forces in α-AlF_3, was obtained at A = 100.0 eV and ρ = 0.48 Å, and A = 3860.0 and ρ = 0.222, respectively.

Our investigation employed two distinct fitting strategies: a relax fit and a standard fit. The global minimum found for the SOS when implementing a relax fit (with the reference data including the structural parameters in α-AlF_3) was obtained at A = 100.0 eV and ρ = 0.48 Å. When implementing a standard fit (with the reference data including atomic forces in α-AlF_3), the global minimum was obtained at A = 3860.0 and ρ = 0.222.

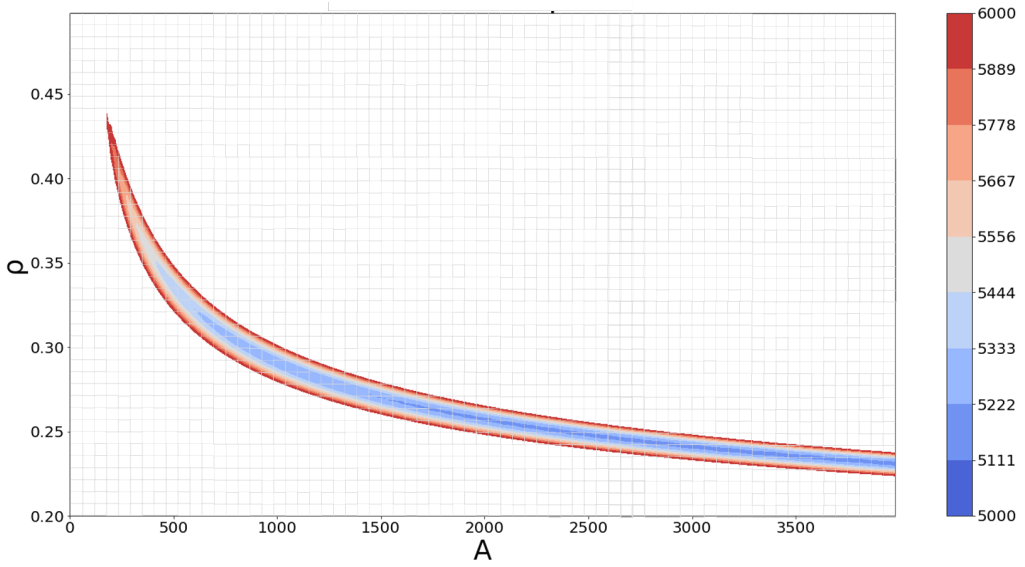

Figure **8**. The SOS landscape from a standard fit as a function of A and ρ Born-Mayer Al-F potential parameters, where α-AlF_3 is the reference structure[106]. Values of SOS above 6000 are not shown.

Upon a thorough examination of the outcomes, we selected the standard-fitted potential. This choice was predominantly driven by the less satisfactory performance of the relax-fitted potential in depicting α-AlF_3. Moreover, the relax fit yielded A and ρ parameters of around 500 eV and 0.45 Å. The relax-fitted parameters are at an unusual scale for the potential, and physically the parameters do not accurately describe the atomic size and physical properties. Maps were also produced when the reference process was repeated with the reference data including the DFT-calculated Birch-Murnaghan bulk modulus[105] or atomic forces for three AlF_3 phases; results are given in Table 1.

Table **1**. Al-F Born-Mayer interatomic potential parameters corresponding to the global minima of the SOS based on a standard fit of the reference data.

	Reference Data	A / eV	ρ / Å
(A)	atomic forces in α-AlF_3	3760.0	0.2220
(B)	atomic forces and bulk modulus in α-AlF_3	10860.0	0.1920
(C)	atomic forces in α-, β-, and θ-AlF_3	2381.0473	0.2461

As our interest is in nanocluster structure predictions, we want an IP that generates and ranks energy minima for nanoclusters that best matches results from DFT energies. With a suitable IP, computational costs can be dramatically reduced, as we can then choose to focus on only the top N lowest-energy IP structures, instead of optimising all ranks of IP LM.

To identify the most suitable IP for predicting candidate LM that can be readily refined to DFT LM for nanoclusters, a scoring system has been developed, a so-called performance score that is composed of three components:

To identify the most suitable IP for predicting candidate LMs that can be readily refined to DFT LMs for nanoclusters, a scoring system has been developed—a so-called performance score that is composed of three components:

$$\text{Performance Score} = \text{Ranking Score} + \text{RMS Score} + \text{Scaling Score}, \qquad \textit{Eqn } 37$$

$$\text{Ranking Score} = \frac{\sum_{LM} |\text{rank}_{\text{PBEsol}} - \text{rank}_{\text{IP}}|}{\text{Number of unique IP LM}}, \qquad \textit{Eqn } 38$$

where rank_{IP} represents the energy rank of LM structures found on the IP energy landscape, and $\text{rank}_{\text{PBEsol}}$ is the LM ranking post-PBEsol refinement. The ranking score encapsulates the total number of unique configurations identified in the IP energy landscape.

$$\text{RMS Score} = \sqrt{\frac{\sum_{atoms} |\mathbf{r}_{\text{PBEsol},i} - \mathbf{r}_{\text{IP},i}|^2}{W}}, \qquad \textit{Eqn } 39$$

where $\mathbf{r}_{\text{IP},i}$ are the atomic coordinates of each IP LM structure found, and $\mathbf{r}_{\text{PBEsol},i}$ are the atomic coordinates of the PBEsol LM found from relaxing the IP LM cluster structure, and W is the sum of all IP LMs' atomic position RMS.

$$\text{Scaling Score} = |1 - \text{Scaling factor}|, \qquad \textit{Eqn } 40$$

where 'the scaling factor' is a measure of structural expansion/contraction relative to the PBEsol-refined structure.

The RMS and the scaling factor are computed using the in-house `CF_clusterpy` code developed by Professor Woodley's research group. This code conceptually works by comparing two cluster structures—typically an IP optimized structure and a PBEsol-refined structure—by aligning them through translation to a common centre of mass and rotating them to align their principal axes. It then calculates the RMS deviation between corresponding atomic positions to quantify structural differences and computes a scaling factor to account for any isotropic size variations between the clusters. If the IP structure perfectly matches the PBEsol-refined structure, both the RMS deviation and the scaling factor will be zero, indicating that the two structures are identical in both shape and size.

Global optimisations using 5,000 MCDQ steps with the box size parameter of 4.0 were carried out for each set of IPs—A, B, and C—for cluster sizes n = 1 to 5. The resulting IP LMs were subsequently refined on the PBEsol energy landscape before the suitability of the IPs was assessed using the performance score; results are shown in Table 2.

Table **2**. The suitability of IPs, listed in Table 1, using measures described in equations 31 to 34.

	Ranking	**RMS**	**Scaling**	**Performance (overall)**	**Top 10**
(A)	21.3896	6.8448	3.3362	31.5706	3.1674
(B)	71.9247	15.7868	6.5462	94.2577	5.8280
(C)	21.0448	6.6675	7.0621	34.7743	4.3711

Table 2 displays the individual components and the total performance score for each IP. These performance scores are based on all unique structures discovered during the MCDQ runs with each respective IP. Additionally, a 'Top 10' score was calculated, which considers only the scores associated with the predicted top 10 lowest-energy LMs from each IP.

Generally, IPs fitted with more observables are better at replicating bulk phases[106]. Thus, we hypothesised that IP B and IP C would outperform IP A since they are fitted to a greater number of observables. However, our results show that when it comes to locating lower-energy LMs, IP A is more effective than both IP B and IP C, possibly because the reference structures were bulk phases as opposed to clusters. Consequently, in our results that follow—that is, the structure prediction of $(AlF_3)_n$ for n = 1–11 and 20—we employed IP A.

3.2.2.F-F Four region Buckingham Potential

For the F – F interatomic potential we have employed the four-region Buckingham potential which already have been developed by Jackson *et al.* [9,102,107–109] and used for many systems. The parameters for this four-region Buckingham potential was fitted to reproduce the bulk structural parameters of three lanthanide earth metal (La^{3+}, Nd^{3+}, and Y^{3+}) fluorides and their elastic properties.

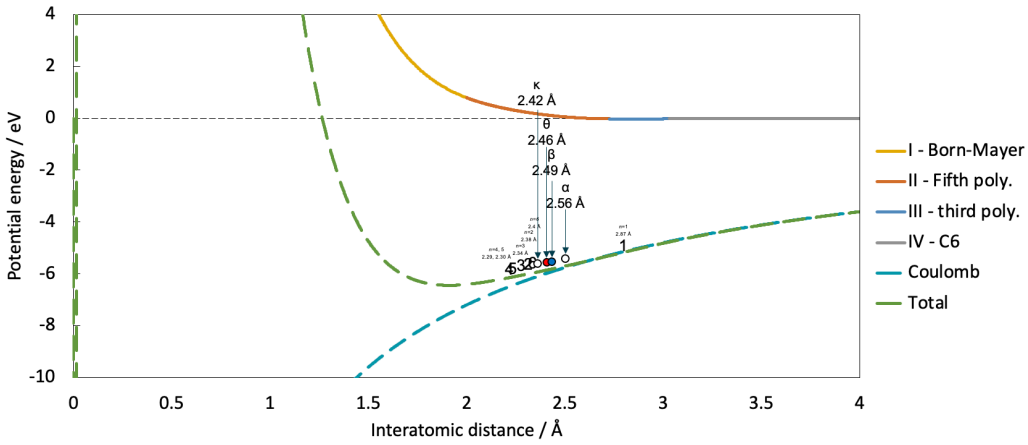

Figure **9**. The F-F four-region Buckingham potential: in region I, F – F distances less than 2.0 Å, a Born-Mayer potential; in region II, greater than 2.0 Å and less than 2.726 Å, a fifth order polynomial; in region III, greater than 2.726 Å and less than 3.031 Å, a third order polynomial; and in region IV, greater than 3.031 Å, a C_6 term[102]. The total potential (green dashed line) is the sum of the four-region Buckingham potential and Coulombic potential (cyan dashed line) across all regions. The numeric and circle markers on the total potential curve indicate the

average F-F interatomic distances for GM nanoclusters of size *n* and the bulk phases, respectively.

The F–F potential is shown as a continuous multicoloured solid line in Figure 9, with each colour representing one of the four regions of the Buckingham potential. Each of the coloured regions is defined by one of the following analytical expressions:

$$V(r_{ij}) = 1127.8 \exp\left(-\frac{r_{ij}}{0.2753}\right), (r_{ij} < 2.0 \text{ Å}) \quad \textit{Eqn } 41$$

$$V(r_{ij}) = 371.2706\, {r_{ij}}^5 - 741.1170\, {r_{ij}}^4 + 597.2668\, {r_{ij}}^3 - 241.8573\, {r_{ij}}^2 + 49.0486\, r_{ij} - 3.9760, \quad (0.2 \text{ Å} < r_{ij} < 2.726 \text{ Å}) \quad \textit{Eqn } 42$$

$$V(r_{ij}) = 9.4816\, {r_{ij}}^3 - 9.6271\, r_{ij}^2 + 3.2362\, r_{ij} - 0.361, \quad (2.726 \text{ Å} < r_{ij} < 2.031 \text{ Å}) \quad \textit{Eqn } 43$$

$$V(r_{ij}) = -\frac{15.83}{{r_{ij}}^6}, \ (3.031 \text{ Å} < r_{ij}), \quad \textit{Eqn } 44$$

where r_{ij} is the F – F interatomic distance.

The green dashed line shown in Figure 9 is the sum of the four-region potential (*Eqn* 41 to 44) and the Coulombic potential (*Eqn* 3). The numeric and circle markers are placed on the total potential curve; the numeric markers indicate the GM of cluster sizes *n*, and the circle markers represent various bulk phases of AlF_3. The location of each marker corresponds to the average F – F interatomic distance of the system.

The Al–F and F–F interatomic potentials are combined with the shell model to enable a polarisable model for the AlF_3 nanoclusters. In the shell model, each ion is represented by a core and a shell connected by a harmonic spring, allowing the simulation of polarisability. The spring constant *k* (from *Eqn* 8) for the F – F potential is taken from Valerio *et al.*, with a value of 20.77 kg s^{-2}. The core and shell charges are set to 0.59 *e* and –1.59 *e*, respectively, where *e* is the elementary charge.

3.2.3. Two-stage approach to predicting DFT LM structures

A direct global optimisation, or search on the DFT energy landscape can be significantly inefficient. For a more computationally cheaper and potentially a more efficient global optimisation, we employed the two-stage method that has been employed previously in many structure prediction studies of nanoclusters[9,89,93,110]. The two-stage method starts with a global optimisation on the IP PES, then lower energy IP LM candidates are subsequently optimised on DFT energy landscape (figure 10). Here we hypothesised the IP PES describes the DFT PES with sufficient accuracy that each of the DFT LM can be found in this two-stage approach. Thus, the use of IP potentially reduces the computational cost required to find the DFT LM.

A direct global optimisation, or search on the DFT energy landscape, can be significantly inefficient due to the high computational cost of DFT calculations for numerous configurations. For a more computationally efficient global optimisation, we employed a two-stage method that has been used previously in many structure prediction studies of nanoclusters [9,89,93,110]. The two-stage method starts with a global optimisation on the IP PES, then lower-energy IP LM candidates are subsequently optimised on the DFT energy landscape (see Figure 10). Here, we hypothesise that the IP PES describes the DFT PES with sufficient accuracy so that each of the DFT LMs can be found using this two-stage approach. Thus, the use of IPs potentially reduces the computational cost required to find the DFT LMs.

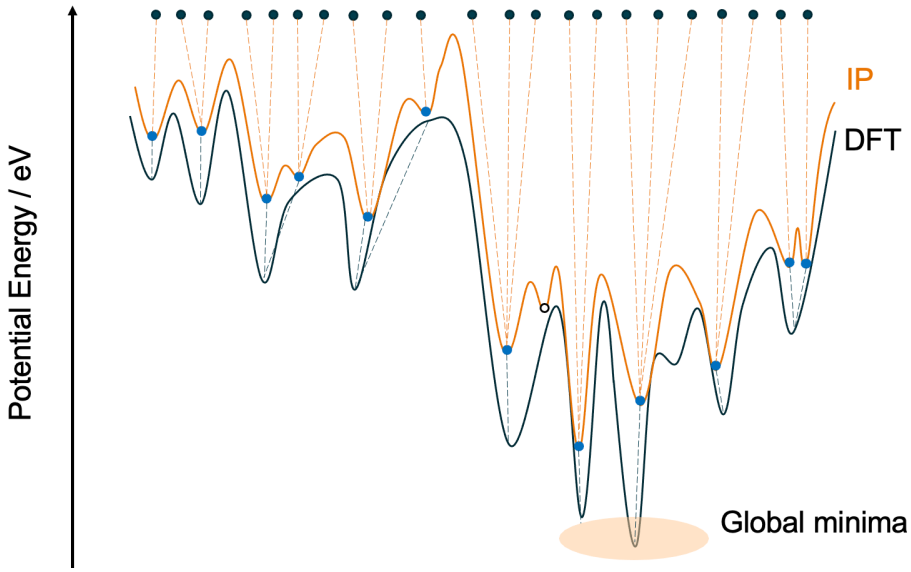

Figure **10**. Pictorial description of how the two PES could match in the two-step method, where orange (black) solid-line represent the IP (DFT) PES, the orange (black) dashed line connect initial and relaxed points, black-dots represent randomly generated structures, whereas blue-dots represent IP-LM.

For the smaller nanocluster sizes (n = 1 – 6), the MCDQ global optimisation method is employed, as implemented in the KLMC software, because the PES is relatively simple as the number of atoms within the system is relatively less, thus employing more sophisticated approaches would be more costly for smaller clusters. For larger cluster sizes (n > 6), we employed the GA in the KLMC software. The KLMC software (a) calls the GULP code to optimise the structures generated from the MCQ method or the GA method, and (b) executes a geometric screening algorithm based on graph theory. The algorithm in (b) acts as a filter, screening out previously identified structures and retaining only unique LM configurations by employing a comparative evaluation of an encrypted code, based on atom connectivity, known as a hash-key. Generated results are, therefore, free of any duplicated LM structures.

Each time the GULP program is called to relax a candidate structure, the conjugate gradient method is employed until the magnitude of the energy gradient (gnorm) falls below 0.4 eV $Å^{-1}$, after which the RFO method is employed until the LM is found. This hybrid local optimisation approach to relaxing atomic structures is beneficial, as the conjugate gradient method is more efficient during the early

phases of optimisation when the gnorm is high, while the RFO method effectively ensures that the optimisation stays within an energy basin and can direct the search along the eigenvector associated with an imaginary phonon frequency if necessary. The tolerance settings for this structure optimisation are shown in Table 3. Once the IP LMs are identified, the IP LM configurations are subsequently optimised to DFT LMs using the PBEsol-GGA functional as implemented in the FHI-aims code.

Table **3**. GULP local optimization tolerances for atomic coordinates (10^{-xtol}), energy (10^{-ftol}) and forces (10^{-gtol}) as well as the maximum number of line searches per GULP call (maxcyc).

xtol / Å	6.0
gtol / Å	6.0
ftol / eV	6.0
maxcyc	10000

For cluster sizes n = 1 to 6, we have used 2,000, 100,000, 200,000, 400,000, 900,000, and 3.5 million MCDQ steps, respectively—that is, we generated and relaxed this many random candidates for each size. The GA has two key parameters, namely population size and the number of generations, which were gradually increased from 65 to 88 and from 3,024 to 5,419, respectively, for sizes n = 7 to 11. For cluster size n = 20, 120 populations per generation and 5,000 generations were used, resulting in over 500,000 configurations generated. Other key parameters include the tournament size used to determine selection of candidates who survive or are chosen from the parent generation, which is fixed at 10; and the initial maximum distance between any two random atomic coordinates—randomly placed atoms are constrained to be within a cubic volume with lengths from 6.5 Å to 10.8 Å for cluster sizes n = 1 to 11, and 15 Å for n = 20.

3.2.4. Convergence of the GA global optimisations

As already mentioned, there are two key GA parameters—the population size per generation and the number of generations—as their product determines the sampling density for GA calculations in KLMC. These GA parameters were initially based on a previous study of predicting clusters of ZnO, which also employed KLMC. Given that AlF_3 contains four atoms per formula unit, compared to two atoms for ZnO, we modified the population size accordingly. The population and generation parameters used for the ZnO study were increased by a factor of 1.5. This adjustment was made in line with the ratio of the number of atoms per formula unit, as the increase in the number of atoms increases the dimensionality of the PES to be explored by three per atom. The size of the simulation box was also based on how these were chosen in the ZnO study[89]. We multiplied the ratio of the lengths that if ions of the smallest formula unit cluster aligned in a line of ZnO and AlF_3, which are 2.79 Å and 1.63 Å, respectively.

As shown in Figure 11, for each GA run, a fixed population size—between 30 and 80 configurations—was implemented, with the number of generations ranging from 150 to 4,940, corresponding to the cluster size. The parameters of the simulation box sizes used range from 6.5 Å to 14.6 Å, respecting the cluster size.

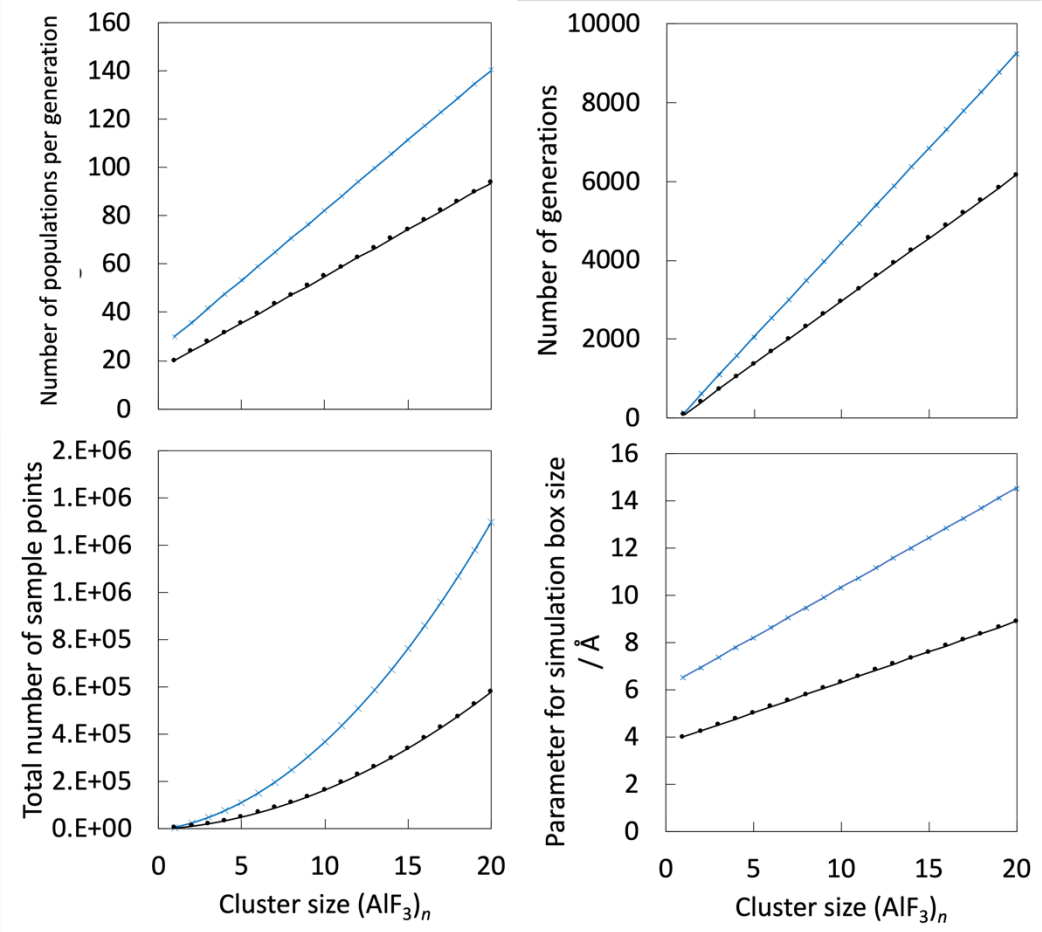

Figure **11**. The benchmarked GA parameters (population size, number of generations, simulation box size) from the previous ZnO study[89] (black) and the adopted GA parameters for the current $(AlF_3)_n$ study (blue), and the total number of sample points used for the GA from the population and generation parameters.

Other GA parameters, which are designed to maintain the diversity of configurations within a population and ensure the chance to obtain sensible structures (as the success of crossover is dependent on diversity), include:

- The percentage of the structures in a current generation that are mutated during the crossover process, set to 80%.
- The probability of applying the self-crossover, atom exchange, expansion, and contraction algorithms when generating new configurations, all set to 10%.

- The scaling factors for the expansion and contraction are set to 1.2 and 0.8, respectively.

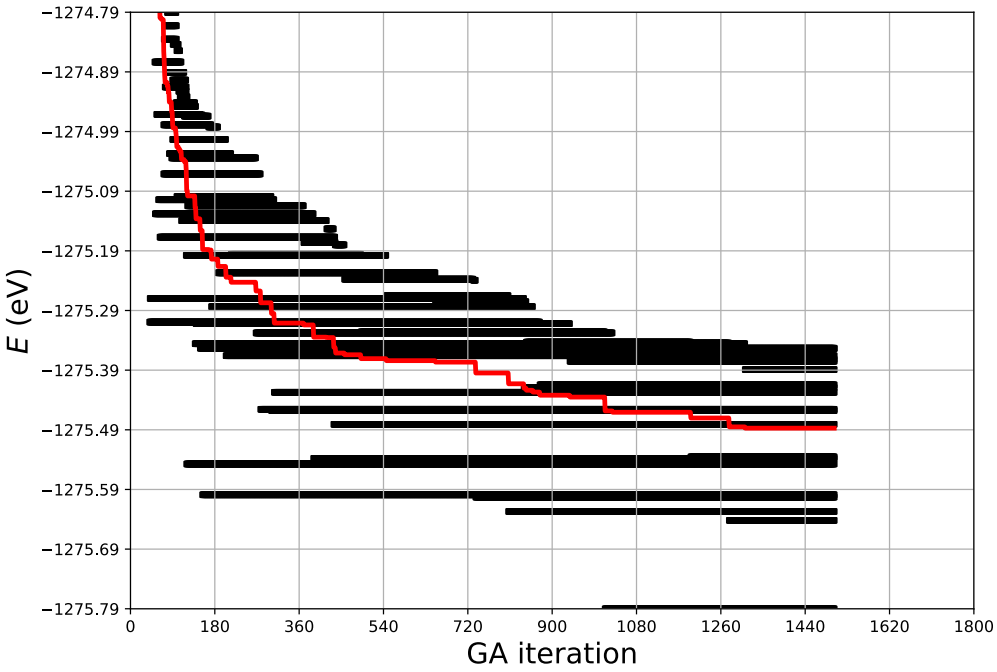

Figure **12**. The 20 lowest energy IP-LM clusters for $(AlF_3)_{20}$ as a function of GA generation (or iteration); black points appear as black lines across generations. The initial point of each black line signifies the generation in which the cluster was first identified. The red line illustrates the average total energy of these top 20 most stable configurations.

The GA is assumed to have converged if there is no further change in the lowest 20 LM energies after 200 to 250 generations[89,91,111]. Figure 12 illustrates, for the largest system investigated, how the IP energy of the 20 lowest-energy LM clusters evolves as a function of generation. The commencement of each black line in the figure indicates the generation at which a particular cluster was first discovered, while the red line represents the average total energy of these top 20 most stable configurations. In the example shown, the GM is found at around the 1,000th generation, the second lowest-energy LM was found after 1,260 generations, and of the lowest 20 candidates found, it is one of the higher-energy LMs that was found last—not the tentative GM.

3.3. Results and Discussions

3.3.1. Global Optimisation

As a result of the global optimisation on the IP PES using the MCDQ method implemented in the KLMC, we found 1, 1, 6, 35, 233, and 2,316 unique LM for cluster sizes n = 1 to 6, respectively. The MCDQ method does not require tailored parameters for the global optimisation which only

After obtaining the IP-LM candidates from the global optimisation, they are subsequently optimised on the PBEsol-GGA functional PES using the two-step method. Following the optimisation on the PBEsol-PES, the total number of unique LM is reduced to 1, 1, 5, 24, 149, and 1,308 for n = 1 to 6, respectively.

To ensure that no lower-energy LM are missed, we used a large number of sample points per cluster size for the global optimisation on the IP-PES. The PBEsol-LMs are then re-optimised on the IP-PES and again re-optimised on the PBEsol-PES iteratively until no new PBEsol-LM are found from this iterative process.

For convenience, we will refer to the PBEsol GM and LM simply as GM and LM from now on. The reduced number of LM after refining the IP-LM on the PBEsol-PES suggests that many of the IP-LM are unstable on the PBEsol-PES. The unstable IP-LM collapse to LM that have different atomic configurations or connectivity compared to the original IP-LM when optimised on the PBEsol-PES.

The collapsed configurations are already found from the global optimisation. This also implies that many of the IP-LM fall into the same LM basin, resulting in duplicated clusters being collected in the data. To filter out the duplicated LM after the PBEsol refinement, two methods are employed: filtering duplicates based on the total energy of the LM and based on the atomic connectivity of the nanocluster.

The filtration of duplicated structures based on the structure energy is based on the hypothesis that LM must be in the same LM basin if the difference in

PBEsol energy between LM is less than 1.0×10^{-5} eV. This tolerance is set conservatively to avoid losing any unique LM during the process. Here, we also hypothesise that there will be no LM with the same energy but different configurations. Perhaps this hypothesis is rather aggressive (with a risk of eliminating chiral configurations), but for smaller nanocluster sizes (n = 1 – 6), the number of LM on the PES is relatively small and can be easily verified visually.

For filtration based on the atomic connectivity of the nanoclusters, we used the hash-key algorithm[89,94], which is the same algorithm implemented in the KLMC code that generates/filter unique encrypted strings based on the atomic connectivity of a system. We employed the hash-key to identify duplicated configurations among the PBEsol-optimised structures. The filtration process using these two methods is automated using custom Python scripts.

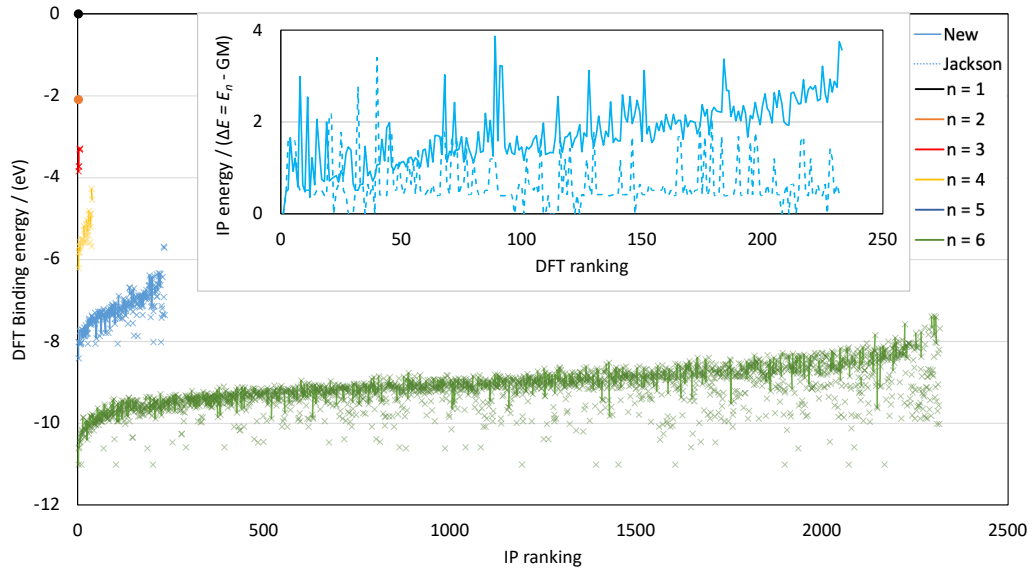

Figure **13**. The PBEsol binding energy of $(AlF_3)_n$ (n = 1 – 6) nanoclusters as a function of IP rank defined as PBEsol (lines and cross markers) energies, $E_i - nE_1$, for PBEsol optimised structures. From n = 1 to 6, the data are shown in black, orange, red, yellow, blue, green colours, respectively. Markers indicate all PBEsol optimised IP structures including unstable on the PBEsol landscape which has not removed the duplicated PBEsol LMs. Solid line shows the unique LM and the breaks in lines indicate the removed unstable IP LM on the PBEsol landscape

(relaxed to a PBEsol LM already included in the figure). Insert panel: the energy difference between the GM and the LM of n = 5 from the two different IP (solid: newly fitted Al-F Born-Mayer IP, dashed: Jackson potential) optimised n = 5 PBEsol structures against the PBEsol energy ranking before the IP refinement.

The binding energy E_{bind} of an n-size cluster LM is defined as:

$$E_{bind} = E_{LM}^{n} - nE_1 , \qquad Eqn\ 45$$

where E_{LM}^{n} is the energy of the LM for the cluster of size n, and E_1 is the energy of the GM of the smallest cluster size (n = 1).

The figure shows the binding energy, as defined in Equation 46, of the PBEsol-optimised structures as a function of the IP LM ranking before the PBEsol optimisation. The different coloured data points indicate the different cluster sizes: from n = 1 – 6, the data correspond to black, orange, red, yellow, blue, and green colours, respectively.

The cross markers represent the full set of data points without removing duplicates after the IP LM are optimised on the PBEsol PES, including unstable configurations. The solid lines represent the unique LM after removing duplicates and unstable configurations. The unstable IP LM configurations on the PBEsol PES are removed, and because of this removal, the solid lines appear broken in the figure.

An unstable IP LM on the PBEsol PES means that the connectivity of the atoms in the cluster changes significantly upon optimisation. The unstable structure collapses to a structure that has different atomic connectivity compared to the structure before refinement. The broken lines may indicate the limitation of the two-body potential to describe the PBEsol PES. However, the profile of the solid lines, or the distribution of the unique LM on the PES, shows accumulated Gaussian distribution curves, which can be seen in the green solid line (n = 6) on the plot.

3.3.1.1. Efficiency of Global Optimisation Methods

The existing number of LM in each cluster size shows a trend of exponential increase in the number of LM on each cluster size's PES as the cluster size increases. A fitted fourth-order polynomial function, which interpolates the highest energy (the least stable) LM, predicts that n = 7 would have more than 4,200 (8,400) LM on the PBEsol (IP) PES. The expected number of LM for n = 7 is four (three) times more than for n = 6 on the PBEsol (IP) PES.

Using the MCDQ method for global optimisation with 3.5 million sample points for n = 6 took over four months, which means one sample point took approximately 3 seconds to optimise. Thus, if we use the MCDQ method for n = 7, it would take approximately 11 months, which is inefficient as we are interested in the top NNN most stable LM.

The MCDQ method is advantageous if our aim is to saturate all the LM or obtain a statistical distribution of LM to understand the PES. However, we are mainly interested in the lower-energy LMs. We will revisit this later in this chapter to explain why we are mainly interested in the lower-energy LMs.

Thus, an efficient global optimisation method focused on lower-energy structures is required for larger cluster sizes. Here, we employed the GA method implemented in KLMC for larger cluster sizes, from n = 7 upwards. We globally optimised n = 7 – 11 IP LM using the GA in KLMC. Again, we employed the two-step method, where the IP-LM obtained from the global optimisation are further refined on the PBEsol PES.

3.3.1.2. Finding Effective Simulation Box Size

As previously discussed, a crucial parameter in many global optimisation methods applied to atomistic systems is the simulation box size. Commonly, global optimisation methods first generate a random atomic configuration subject to subsequent structure optimisation. However, to construct the random atomic configuration, we need a boundary condition for more effective local optimisation.

Generating a random configuration without a boundary condition implies that the atoms can be placed extremely far away from each other, with computational costs increasing with the separation distance. In general, the simulation box size, or boundary conditions, can also be utilised to generate specific atomic structures, focusing on a certain area of the PES of interest—for example, where we find rod- or nanotube-shaped configurations.

It can also be used to make the global optimisation method more effective, e.g., to find the top N energetically most stable configurations, as energetically stable configurations can be expected to have higher atomic coordination numbers or densely packed atomic configurations. In this work, we have tested a range of simulation box parameter values for the MCDQ global optimisation method in KLMC, as shown in Figure 14.

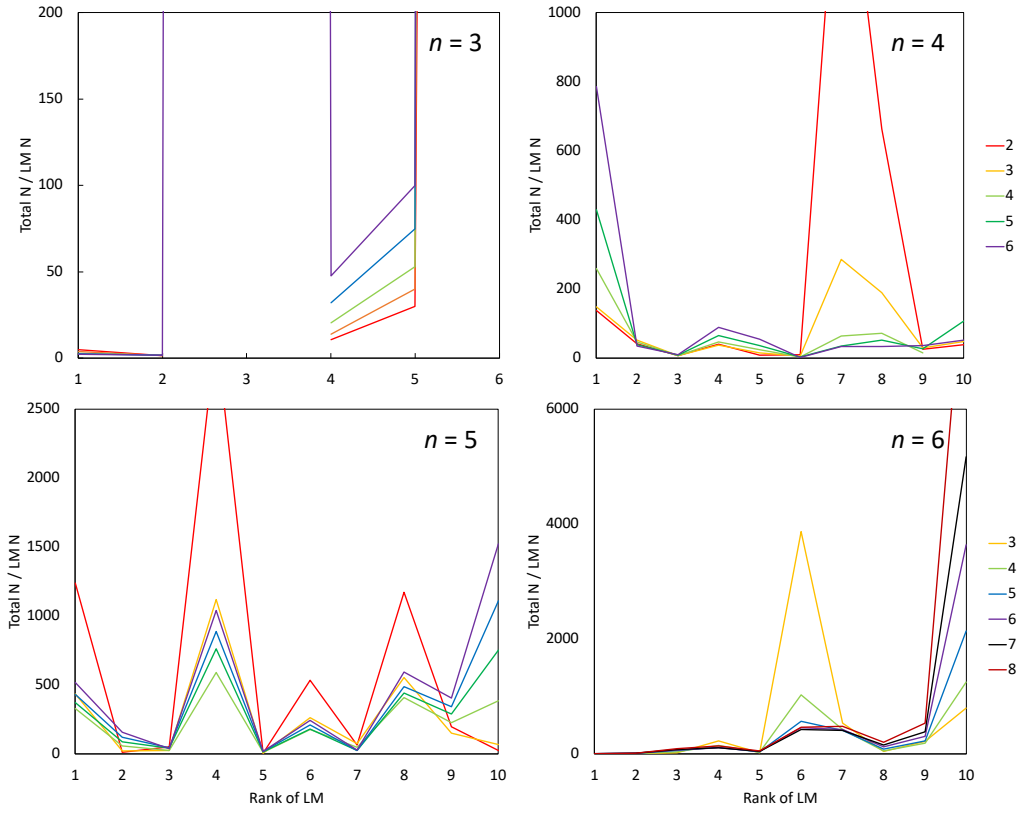

Figure **14**. An average number of sample point used for MCDQ calculations to find an LM on the $(AlF_3)_n$ (n = 3 – 6) IP PES for a range of the simulation box

parameter values (coloured using the rainbow scheme – see the legends) as a function of the configuration ranking.

The results from different box sizes are distinguished by colouring using the rainbow scheme. In general, we tend to use small box sizes to focus on the energetically more stable cluster structures because in general the more stable tend to have more highly compact/confined atomic configurations. Interestingly, when we change to even smaller box sizes for initial random structures, the frequency or chance of finding certain LM often changes dramatically.

As seen in Figure 14, the top two most stable configurations of $n = 3$ clusters require the same average number of MCDQ steps per structure found (2.13 steps), as the $n = 3$ system has a relatively small PES with most of the LM configurations contained within the simulation box size. Thus, simulation box size does not appreciably influence the statistics for the two top LMs of the $n = 3$ cluster.

The IP-LM3c requires many more sample points to be found, as the structure is located in a narrow funnel of the PES. The same analysis can be applied to IP-LM3f. The cluster structural stability will be discussed in the following section.

The IP-LM3d and IP-LM3e required different average numbers of steps to be found depending on the simulation box sizes; as the box size increases, the IP-LM3d and IP-LM3e need more steps to be found, which means the configurations for the 4th and 5th LM are more compact rather than rod-like or elongated.

The same analysis can be applied for the rest of $n = 4$, 5, and 6 clusters: if a structure requires more steps to be found for the first time using a smaller (larger) simulation box size, then the structure is expected to be elongated or rod-like (densely packed). For the case of $n = 5$, the 4 Å simulation box size is optimal as it permits us to find most of the LM5(x) with the least number of steps compared to the other smaller or larger box sizes.

The 2 Å and 3 Å box sizes require the most and second most number of steps overall, while the 4 Å box size requires the least number of steps. This indicates

that most of the IP-LM5(x) structures are densely packed and extend to 4 Å, and 2 Å – 3 Å boxes are too small to contain the structures.

For IP-LM5J, using the 2 Å and the 3 Å boxes allows us to find the structure with the least number of steps, as the structure is densely packed, with C_s point group symmetry, which is the same as that of the IP-GM5. The IP-GM5 has a pentagonal ring configuration, but IP-LM5J is packed with no hole in the structure, so the end-to-end distance in the IP-GM5 is longer, thus requiring larger simulation box sizes.

Understanding the nanocluster structures allows us to calibrate the simulation box size to save computational costs. The AlF_3 clusters show that a gradual increase in the simulation box size—from 2 Å for n = 3 – 4 Å for n = 6 — would allow us to capture the energetically more stable configurations on the PES.

3.3.2. Comparison with The Two-Body Potentials

The inset in Figure 13 illustrates the energy difference between the GM and LM for n = 5 as a function of the PBEsol ranking prior to IP refinement. This was determined by optimising the n = 5 PBEsol LMs using two distinct IP sets: Jackson's[102] and a newly fitted IP.

The IP energy differences for the LMs are shown as a function of their PBEsol energy ranking prior to IP refinement. The Jackson IP, published in 2004 for investigating aluminium fluoro-/hydroxy-silicate (Al_2SiO_4(OH, F)), commonly known as topaz, is noteworthy. In this context, the Al – F interaction in Jackson's IP utilises the Buckingham or Born-Mayer potential (see *Eqn* 5), and the F – F interaction is modelled using the Buckingham potential, rather than the four-region approach used in the current study of nanocluster structure predictions.

Table **4**. The Jackson's AlF_3 Buckingham potential

Interaction	Potential	A / eV	ρ / Å	C / eV Å^6
Al-F	Buckingham	1773.415	0.2753	0.0
F-F	Buckingham	4350.0	0.2753	15.83
On-site potential parameters				
F	Shell charge	Y = -1.378|e|		
	Spring constant	k = 24.36 eV Å^{-2}		

In the inset panel of Figure 13, the Jackson IP energy of the LMs remains constant, despite an increase in the energy ranking prior to IP refinement. In contrast, the energy of the LMs optimised with the newly fitted IP rises in tandem with their energy ranking. This indicates that the Jackson IP is less effective in mimicking the PBEsol PES of the nanocluster.

The data comparing PES description performance using different IPs underscores the necessity for IP fitting in predicting nanocluster structures, thereby justifying the development of the new IP. Additionally, for n=1n = 1n=1 to 6, the top 5 (and top 10) most stable PBEsol LMs were identified within the top 10 (and top 50) IP LM rankings. Furthermore, for n = 1 to 10, the GM configurations according to the IP remained the same as the PBEsol GM following optimisation.

This consistency further confirms the accuracy of the fitted IP in representing the PBEsol PES. Notably, for n = 11 and n = 20, the LM configurations ranked 11th and 6th in IP energy emerged as the most stable PBEsol GMs, respectively.

3.3.3. Density Of States

Figure 15 illustrates the density of states (DOS) for the total energy of PBEsol-optimised $(AlF_3)_n$ clusters, with nnn varying from 3 to 6, at both IP LMs

and LMs following PBEsol optimisation. An extensive search for all possible LMs within these cluster sizes was conducted using the MCDQ method, leading us to believe that our exploration was thorough.

Clusters with n = 1 and 2 are excluded from this figure, as they feature only a single structure on their PBEsol PES, which results in the DOS showing a single sharp peak at 0 eV. Ideally, the DOS should present a bell curve or Gaussian distribution, indicative of a random search that covers a wide range of LMs, as demonstrated in Figure 15, especially in the panel for n = 6.

The DOS representing the distribution of LM total energies is expected to conform to a Gaussian distribution, a trend explainable by combinatorial reasoning. When an atom in a cluster is moved within its boundary, causing minimal changes to atomic connectivity, the system is restricted to a relatively limited set of configurations, which is the tail of the DOS. Conversely, increasing the number of atom relocations broadens the range of possible configurations, forming the body or peak of the DOS.

The likelihood of encountering LMs within a cluster's PES typically follows a Gaussian distribution. This distribution can be rationalised by considering atomic rearrangements. Generally, the more energetically stable LMs are associated with atomic configurations that exhibit higher symmetry and densely packed or highly coordinated atomic arrangements compared to those with higher energy.

Furthermore, configurations that are more elongated and of lower dimensionality tend to emerge beyond the mean energy of all LMs within a given cluster size on the PES, at the point where the DOS is maximal. Relocation of an atom within these stretched configurations can result in a collapse to more confined, lower-energy configurations, thereby reducing the prevalence of high-energy LMs. This dynamic ultimately contributes to the Gaussian distribution of configurations across the potential energy spectrum.

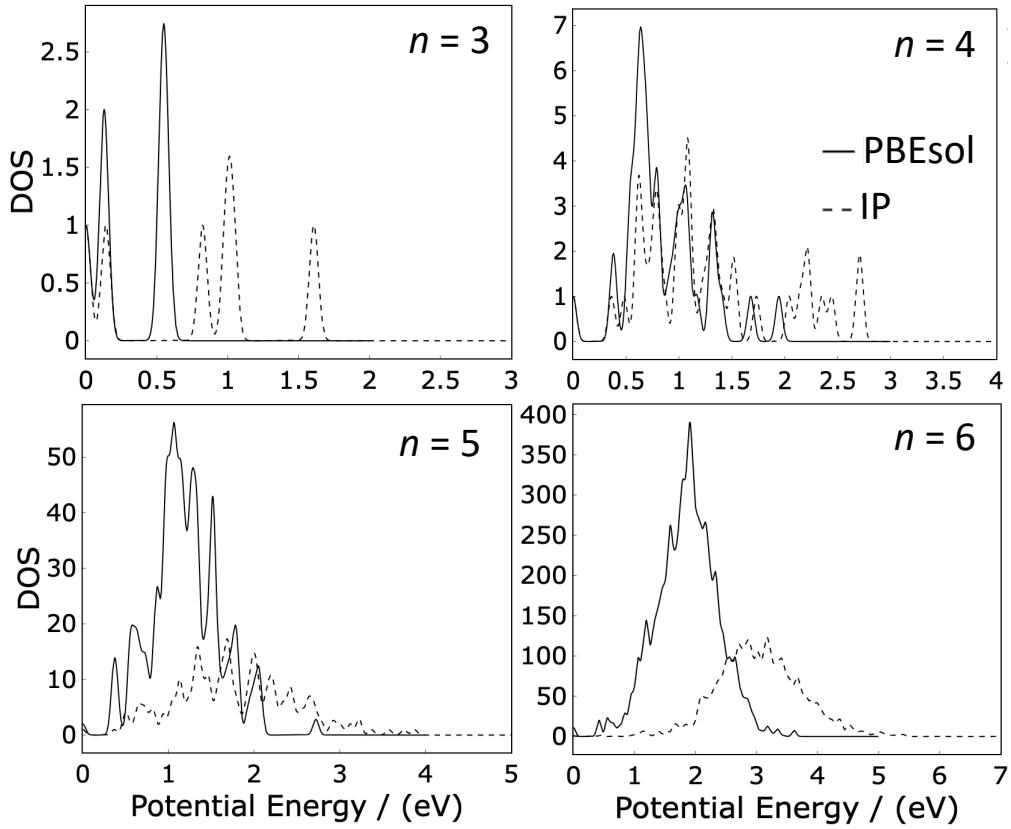

Figure **15**. The DOS of the total energy of the PBEsol optimised $(AlF_3)_n$ (n = 3 – 6) IP LM in solid line and the IP LM prior to PBEsol refinement in dashed line (The variance value and the bin size for the Gaussian distribution are 0.001 eV^2 and 0.01 eV were used to prepare the plot).

For smaller atomic clusters, the DOS exhibits a limited number of peaks, reflecting the comparatively small number of atoms present. Consequently, this also leads to a reduced number of possible atomic configurations due to fundamentally fewer combinatorial possibilities.

There is a noticeable disparity in the number of peaks between the PBEsol and IP DOS. For instance, the final two peaks in the IP DOS merge into the last peak of the PBEsol DOS, as these represent unstable configurations. Likewise, the other DOS graphs display a greater number of peaks across wider potential energy ranges in the IP PES, due to a higher count of LMs compared to PBEsol LMs. Once these unstable configurations are consolidated, they are restricted to a

narrower potential energy range, resulting in a more defined, sharper overall DOS profile.

In the DOS, there are three distinct peaks, with the initial two positioned closely together. The bases of these peaks show smearing and overlap. The estimated energy gap between the full-width at half-maximum (FWHM) of the first two peaks of $n = 3$ is around 0.01 eV. This indicates that the second peak of the LMs is reachable at room temperature, since 0.01 eV corresponds to approximately 116 K (using the relation $E = k_B T$, where k_B is the Boltzmann constant).

Conversely, for $n = 4$, the energy separation between the first two peaks exceeds 0.3 eV (approximately 3,481 K), rendering the second peak inaccessible at room temperature. Therefore, under natural or ambient conditions, the GM is predominantly expected for $n = 4$.

As the size of a cluster expands, the likelihood of discovering LMs with higher energy under normal conditions also increases. This stems from the heightened count of atoms in the cluster, enriching the spectrum of potential configurations. An augmented tally of LMs nudges the average energy profile to more elevated states, causing the DOS for larger clusters to approach a denser Gaussian distribution.

With sufficient computational power or a more refined algorithm to thoroughly explore the LMs on the PES, canonical ensemble techniques can be utilised to analyse the statistical presence of various LM configurations at a specified temperature, especially after lower-energy LMs have been identified. This would allow for a better understanding of which configurations are thermodynamically accessible under given conditions.

3.3.4. AlF_3 Nanocluster Structures

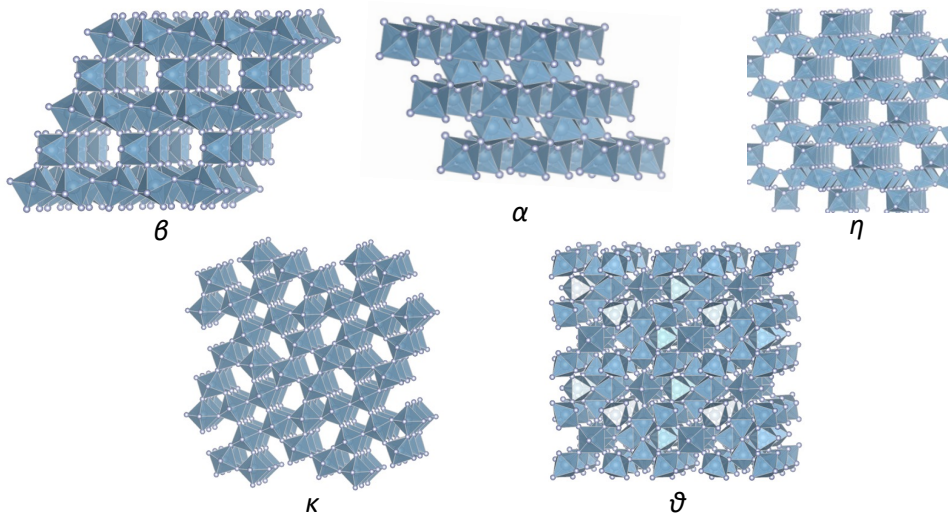

Figure **16**. Five different AlF_3 bulk phases are presented in *Ball-and-stick* model which superimposed with *polyhedral* model to show SBUs. The blue large sphere and grey small sphere are Al, and F atom, respectively.

In our comprehensive global optimization efforts for $(AlF_3)_n$ nanoclusters, we are particularly focused on determining the smallest nanocluster size that replicates the structural motif characteristic of bulk AlF_3. Our objective is to find a cluster size whose structural properties mirror those found in the bulk form. When examining the various AlF_3 bulk structures (figure 16), as illustrated in the *polyhedral* models superimposed on the *ball-and-stick* diagrams, we note that the structural motif, or the secondary binding unit (SBU), of AlF_3 is uniquely octahedral, with all such octahedral SBUs interconnected via corner-sharing links. The minimum cluster size GM where this corner-sharing octahedral SBU pattern emerges is when $n = 6$. At this juncture, the structure's central vertical axis consists of a series of octahedra joined at their corners. This central octahedral spine is further adjoined by two square-based pyramidal SBUs on one flank and two tetrahedral SBUs on the opposite flank, together forming a protective arrangement around the corner-sharing octahedral SBU. While the square-based pyramidal SBUs are interconnected through corner-sharing, the tetrahedral SBUs remain unconnected, leaving the central F atom exposed. This central F atom plays a

pivotal role as it acts as a bridging link between the two octahedral SBUs, thereby completing the intricate lattice that echoes the structural complexity of the bulk AlF_3.

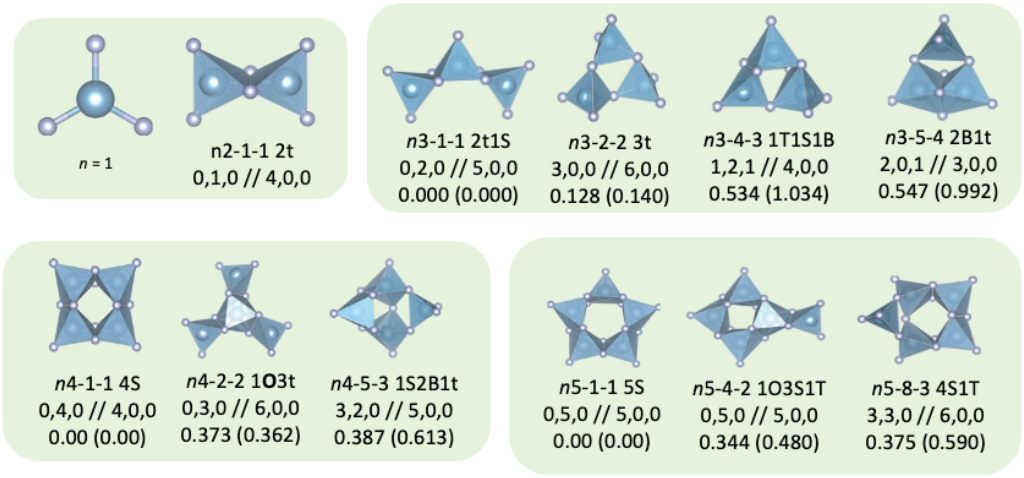

Figure **17**. The *Ball and stick* models of the tentative lowest PBEsol energy $(AlF_3)_n$ PBEsol LM for sizes n = 8 to 11. Blue (grey) spheres represent Al (F) atoms, and transparent blue polyhedral are AlF_m SBUs, with m = 4 to 6. Each configuration is labelled "*n*X-Y-Z S v,e,f // g,h,i,j P (Q)", where X is its size, Y is its IP rank, Z is its PBEsol rank, S is a character string describing the SBUs that the cluster is composed of, v (e, f) is the number of vertices (edges, faces) shared between the SBUs, g (h, i, j) is the number of F atoms with a coordination of 1 (2, 3, 4), and P (Q) is the PBEsol (IP) energy difference between the cluster shown and the ground state cluster of the same size. The character string S is a concatenation of character strings with the format mU that indicate that are m SBUs of type U, where U is the first letter of the name of SBU shape (tetragonal, square-base-pyramidal, bipyramidal, octahedral), and when in bold lowercase, lowercase, uppercase or bold uppercase the SBU contains at least three, two, one and zero singly coordinated F atoms, respectively. Coordination numbers (and SBUs) are calculated using an interatomic Al-F cutoff distance of 2.2 Å.

Figure 17 presents a detailed visual compilation of $(AlF_3)_n$ nanoclusters where n varies from 1 to 5, encompassing both the GM and the 2nd to 4th LM for each respective cluster size. This comprehensive illustration serves to delineate the energetic hierarchy and structural variations within the clusters. Commencing with the upper label, each cluster is annotated with a series of identifiers that

convey critical information about its characteristics. The size of the cluster is the initial datum, followed by the cluster's ranking in terms of IP energy before any refinement through PBEsol has taken place, and its subsequent ranking after such refinement. The ranking system provides insight into the relative stability and energy levels of the clusters in different states of computational analysis.

Furthermore, the upper label elucidates the composition of the clusters in terms of their SBUs. A specific example given is '*n*3-4-3', which translates to the $(AlF_3)_n$ nanocluster being the 4th ranked in terms of IP energy before PBEsol refinement and ascending to the 3rd rank following the refinement process. The notation '2B1t' is decoded to represent a cluster arrangement composed of two trigonal bipyramidal SBUs alongside a single tetrahedral SBU, with the possibility of alternative notations such as 'S' for square pyramidal and 'O' for octahedral configurations. In these labels, a lowercase letter, particularly when emboldened, signifies the absence of singly coordinated fluoride atoms within the SBU, for example for **O**: all the vertex fluorine atoms are bridged to other aluminium atoms, thus no singly coordinated fluorine atoms in the octahedral SBU.

The mid-label catalogues the connectivity patterns prevalent among the SBUs. On the left side of this divider, the count of corner-sharing, edge-sharing, and face-sharing linkages between SBUs is tallied, which is instrumental in understanding the three-dimensional geometry and potential reactivity of the cluster. The right side of the mid-label accounts for the number of fluoride atoms exhibiting specific coordination numbers with aluminium atoms, cataloguing them as singly, triply, and quadruple coordinated.

Finally, the bottom label shows the energy differentials between the LMs and the GM within the same cluster size. These differences are quantified in potential energy, eV, labelled as E(LM*n* – GM*n*), and provide a valuable comparative gauge of the energetic favourability and potential stability of the different cluster formations. This image with energy labels gives a detailed look at the clusters. It helps us understand how the clusters might change shape and what shapes are most common when they are stable, for different sizes of the $(AlF_3)_n$ series.

There is only one configuration in n = 1 in trigonal planar with D_{3h} symmetricity. The two F atoms in the centre of the structure are staggered with the outer F atoms on both sides to minimise the steric hindrance from between of the nearest and the second nearest anions. n = 2, also has only one LM exists in the PES. It is common to predict only one answer for n = 1, but in general n = 2 cluster to have only one LM (D_{2h}) is rare[9,110,112]. Only one LM in the n = 2 can be explained by the ionic radii size differences between the Al and F; Al and F have the smallest ionic size among the cations and anions which eliminates the possible configurations that could have highly coordinated configurations. For instance, in the previous study of $(LaF_3)_n$ report three LM. Shannon ionic radius[113] difference in ratio between La^{3+} and Al^{3+} shows that La ion is 2.0 times bigger than Al ion (La^{3+} has 1.356 Å ionic radius and La atom in a LaF_3 crystal has 9 coordination number, but Al^{3+} has 0.675 Å ionic radius with 6 coordination number for Al atom in an AlF_3 crystal)[9,96,97,113]. Despite Al^{3+} has the smallest cation radius which gives less space to coordinate it prefers to have two edge-sharing (two bridged F atoms between Al atoms) SBU rather than corner-sharing (single bridged F atoms) SBU to increase the coordination number instead of avoiding the steric hinderance. However, even higher coordination number configuration, face sharing (more than three bridged F atoms) configuration, is energetically unfavoured for $(AlF_3)_2$. The GM configuration of $(LaF_3)_2$ is remained as $(AlF_3)_2$ after the datamining and the other two $(LaF_3)_2$ LMs which are composed of two tetrahedral SBUs which connected through edge-sharing, triangle face sharing, and square face sharing which all collapsed to the $(AlF_3)_2$ as shown in the 14.

From this point forward, we will use the following notation for convenience: GMn represents the GM for a cluster of size n, and LM n{rank} represents the LM for a cluster of size n with the specified rank. We have found the total of five LM for n = 3. The GM3 has C_{2v} symmetry which consists of one square pyramidal SBU and two tetragonal SBU. The two tetragonal SBU are edge-shared with a square-base pyramidal SBU centre. As a result, there are two and one singly coordinated F atom terminations per tetragonal and square-base pyramidal SBU. The LM3b only consist of three tetragonal with corner-sharing connectivity to each other, thus it has less average Al coordination number than the GM, which is less energetically preferred with less steric hinderance as the shortest Al-Al

distance is 3.23 Å compared to 2.82 Å in GM. On the other hand, despite the LM3c having a higher average coordination number for Al atom, 4.67, the LM is energetically unfavourable because of the relatively densely confined structures which penalise the energy more from the steric hindrance. The shortest Al-Al interatomic distance of the LM3c is 2.72 Å which is only ~ 0.1 Å shorter than the shortest Al-Al distance in the GM. The LM3d has C_s symmetry which looks like the LM3b configuration that brought two F atoms in the middle of the two Al atoms which are connected through triangular face-sharing. The face sharing put the two Al atoms too close (2.50 Å) which is unfavourable due to the repulsive forces. The short distance between the atoms penalises significantly more than having the higher coordination number with less singly coordinated F atom termination. The energy difference between the LM3b and the LM3c is 0.406 eV and 0.013 eV between the LM3c and LM3d. The relatively big energy gap between the LM3b and LM3c is considerable, but we are confident as we used the total of ~200,000 sample points for thorough PES search using the MCDQ global optimisation on KLMC with the numerous cubic simulation boxes.

For n = 4, in the LM4b cluster, D_3, we can observe the octahedral SBU, in which the octahedral connects the rest of the square base pyramidal and tetragonal SBU through edge-sharing and all the F atoms in the centre octahedral are bridged F atoms, doubly coordinated. The octahedral SBU is different to the octahedral SBU that can be found in the AlF_3 bulk phases which is corner sharing SBU, not edge-sharing SBU. The GM4 is in D_{2d} point-group and is composed of four square-base pyramidal SBUs which forms the edge sharing ‘ring’ configuration with a higher average coordination number for Al atom, 5, with a smaller number of undercoordinated F atom termination than other high energy LMs and less steric hindered with 2.83 Å with Al-Al distances; 0.03 Å longer interatomic distance than 2nd LM. The LM4c has an average coordination number of 4.75 for Al atom which is the value in between the GM and the 2nd LM and a smaller number of undercoordinated F atom terminations than the LM4b. However, the shortest Al-Al distance in the LM4d is ~ 0.02 Å shorter than the shortest Al-Al distance in the GM 4which the steric hindrance penalises more than having a higher coordination number, thus it is less stable than the LM4b.

The GM5 is in C_s point-group and is also in ring shape with edge-sharing square-base pyramidal SBUs. The shortest Al-Al distance is 0.03 Å longer than the GM4. The LM5b has the GM4 configuration feature with an additional edge-sharing tetrahedral SBU on the side which forms an octahedral SBU with no singly coordinated F atoms. The LM5b has the same average coordination number of Al atom as the GM5, 5, but as we mentioned for smaller sized clusters with the shorter Al-Al distance is unfavourable and the configuration has two undercoordinated F atom terminations in the tetrahedral SBU arm. The configuration of the LM5c is similar to the LM5b, but instead of the tetrahedral SBU connects through edge sharing with a square-base pyramidal SBU of the GM4 the tetrahedral SBU connects with the two square-base pyramidal SBU through the corner sharing. The LM5c has a longer Al-Al distance than the LM5b and the same number of the singly coordinated F atom terminations but has a smaller Al coordination number which is less energetically preferred compared to the LM5b. From the qualitative analysis of the structural stability of the $(AlF_3)_n$ (n = 1 – 5) LM we found the trend in AlF_3 cluster structural stability: if a cluster has fewer undercoordinated F atom terminations, a densely packed configuration with a higher coordination number with the longer interatomic distances between the same atom species is energetically more stable. We expect this trend which requires to be the stable cluster will continue in bigger cluster size, $n > 5$, but with shorter average interatomic distance between the same atomic species than the smaller cluster size. As there are a greater number of atoms in cluster as the cluster size increases the electrostatic field will be stronger. The stronger electrostatic field will tightly bound the atoms, thus the densely located atoms will have higher average coordination number.

In atomic and molecular systems, the strength of the electrostatic field generated by ions or atoms significantly influences the structural and bonding properties of the material. A stronger electrostatic field arises when atoms or ions possess higher charges or are in closer proximity to each other. This intensified field enhances the attractive forces between oppositely charged particles, leading to tighter binding of the atoms. As a result, the atoms are drawn closer together, creating a denser arrangement within the material's structure. In such densely packed configurations, each atom is surrounded by more neighboring atoms than

it would be in a less dense structure. This increase in neighboring atoms elevates the average coordination number, which is the number of nearest neighbors directly bonded to a central atom. A higher coordination number indicates that each atom shares bonds with more adjacent atoms, reflecting a more interconnected and stable network. Thus, the stronger electrostatic field not only binds the atoms more tightly but also promotes a structural environment where atoms have a higher average coordination number due to the dense packing facilitated by enhanced electrostatic attractions. Thus, numerous octahedral SBU will be found more often in bigger cluster sizes.

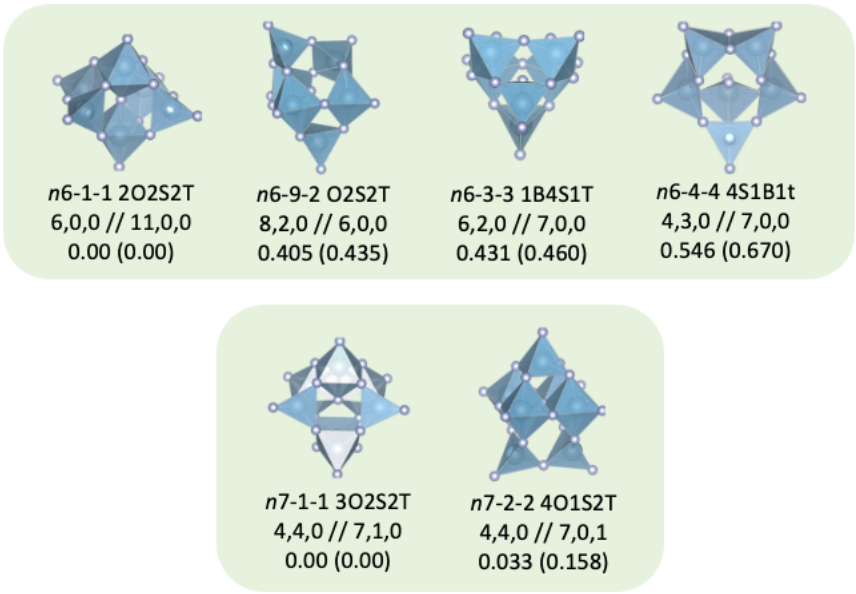

Figure **18**. The PBEsol optimized LM for $(AlF_3)_n$ nanoclusters where n ranges from 6 to 7. In the visual representation, darker spheres denote Al atoms, while lighter spheres represent F atoms. The upper label provides a sequence of information from left to right: the atomic size, the ranking in IP energy, followed by the ranking in PBEsol energy. It further describes the SBU components that make up the configuration. The mid-label, divided into two sections by a vertical line, catalogues on the left the count of corner-sharing, edge-sharing, and face-sharing interactions between SBUs, respectively. On the right, it quantifies the F atoms coordinated to single, triple, and quadruple aluminium atoms, respectively. The bottom label specifies the energy difference for each LM compared to the GM of the corresponding cluster size, n, in both PBEsol and IP notations. This detailed annotation provides an insightful comparative analysis of the structural and

energetic characteristics of the $(AlF_3)_n$ nanoclusters within the stipulated size range.

The GM6 is the smallest cluster size that has the corner-sharing octahedral SBU with an average coordination number of Al atom of 5 with C_{2v} point-group. The only SBU which can be found in AlF_3 bulk structures (figure 18). The GM6 has two octahedral SBUs which are corner-sharing and two tetrahedral SBUs and two square pyramidal SBUs corner-share with both centre pillar octahedral SBUs. The angle of the two octahedral SBU (Al-F-Al) is 167.12° towards the window where no SBU surrounds the centre. The angle minimises the steric hindrance from the surrounding four smaller SBUs, tetrahedral and square base pyramidal SBU. Each SBU has singly coordinated F atom termination. From the view shown in figure 18 two square-base pyramidal are in the back of the octahedral SBU. The square-base pyramidal are not only corner shared with both octahedral SBU but also to each other. The LM7b has also two octahedral SBU, but they are edge-shared to each other. On the opposite side of the octahedral SBU the two square-based pyramidal SBUs are edge-sharing. The rest of the tetrahedral SBUs are corner sharing with octahedral SBUs and one square base pyramidal SBU. The LM7b has the same number of Al coordination numbers, but the edge sharing which caused the shorter Al-Al distance (0.85 Å shorter) is energetically less favourable. The 4th LM has four square base pyramidal SBUs like an arch which the ends of arch are connected to one trigonal bipyramidal and tetrahedral through corner sharing. The trigonal bipyramidal has two undercoordinated F atom terminations and one undercoordinated F atom is pointing towards the centre of the arch where the Al atoms of the arch are facing. The Al atom of the nearest square base pyramidal SBU that the corner shares with the smaller SBU have 147.8° of F – Al – F angle with the shortest Al to centred F atom distance is ~ 2.3 Å. On the other hand, the further out Al atoms in the square base pyramidal SBU have a bigger F – Al – F angle (~ 149.17°) with a longer distance of Al to the centre F atom of ~ 3.1 Å. By locating the undercoordinated F atom in the square pyramidal SBU in the middle of the configuration which electrostatically stabilises the configuration with a large cavity that has 0.26 eV larger in total energy than the LM7c. The LM7c has a similar configuration to the LM7d, but the

square based pyramidal SBU is more like a curved shape with the cross-sectional shape of the parabola which is still facing towards the singly coordinated F atom termination which stabilises the configuration with a smaller cavity in the middle. Thus, again the more densely packed structure is energetically more stable.

There are three octahedral SBUs in GM7 (symmetry C_s), which are edge-sharing with each other; one F atom is coordinated with three nearest Al atoms. The two tetrahedral SBUs have single undercoordinated F atom terminations, and the other two F atoms are corner-shared with two octahedral SBUs. Another F atom in the two tetrahedral SBUs is corner-shared with a square pyramidal SBU. The square pyramidal SBU is edge-sharing with another square pyramidal SBU, which is corner-sharing with the two octahedral SBUs. Thus, the two pyramidal SBUs each have a single undercoordinated F atom termination.

The LM7b has C_{2v} symmetry and features a single F atom in the middle of the configuration that is surrounded horizontally by four square pyramidal SBUs at equal distances from the central F atom. The angle from an Al atom in a square pyramidal SBU, through the central F atom, to the opposite Al atom is 173.02°, indicating that the F atom is slightly lower than the plane formed by the Al atoms of the square pyramidal SBUs. This lowered position of the F atom is due to two Al atoms in a tetrahedral SBU, which are corner-shared with each other and share two corners with the square pyramidal SBUs.

On the opposite side of the tetrahedral SBU, there is only one square pyramidal SBU in which the Al atom is at a distance of 2.41 Å from the central F atom. As the central F atom is equally shared by the nearest surrounding four square pyramidal SBUs, we can assume that the Al atoms in the square pyramidal SBUs bond to the central F atom, effectively forming four octahedral SBUs. Under this assumption, the F atom is at the center of four octahedral SBUs whose edges share with each other, and the Al atoms are separated by 3.19 Å.

The second LM (LM7b) has higher symmetry, and the Al–Al distance between the Al atoms in the octahedral SBUs is 0.06 Å longer than that in the octahedral SBUs of GM7. However, the central F atom in LM7b is shared by four Al atoms where the electron density is insufficiently localized.

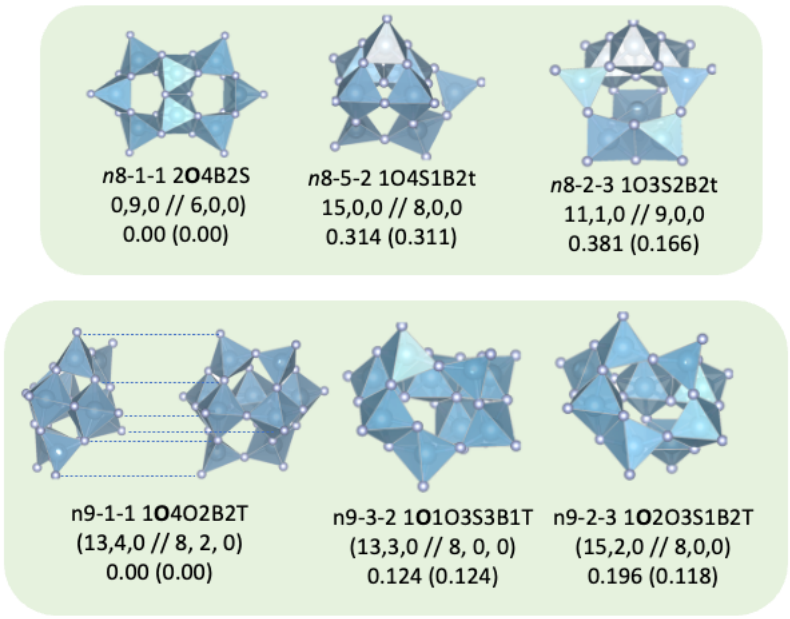

Figure **19**. $(AlF_3)_n$ (n = 8, 9) PBEsol optimised LM: dark (light) sphere represents Al (F) atoms; the dashed line connects the same atom in different angles of view of the same configuration. upper label: from the left, atomic size, ranking in IP energy, and ranking in PBEsol energy. SBU component that consists of the configuration; mid-label: on the left-hand side from the divider number of corner-sharing, number of edge-sharing, face sharing, respectively. The right-hand side represents the number of F atoms that bonded with single Al atoms, three Al atoms, and four Al atoms, respectively. bottom label: PBEsol (IP) energy difference between the LM and the GM of the cluster size, n.

The structural analysis of the GM8 nanocluster reveals an intriguing arrangement that mirrors the number eight, attributable to the presence of two ring motifs. These motifs derive from the GM5 configuration, wherein each extremity of the SBU in the GM3 has edge-sharing relations with the two proximate GM5 SBU. This connectivity through edge-sharing engenders the formation of two octahedra, in which all the fluorine atoms within the octahedral SBUs maintain two-fold coordination. Notably, corner-sharing octahedral SBU are absent in the GM8.

The GM8 structure is particularly noteworthy, given its divergence from the ring structure observed in smaller cluster sizes. Post the GM5, there is a noticeable

transition from the two-dimensional ring-shaped configuration to a more coordinated and relatively three-dimensional architecture. This evolution suggests a structural progression in which the nanoclusters tend to abandon the lower dimensional ring arrangements in favour of more intricate three-dimensional formations as the cluster size increases. However, for GM8 the '8' shape structure has more coordination number than other LM8x. In terms of structural density, if we compare the GM8 with the LM8b and LM8c they have higher density configurations, but with significantly less in average coordination number of Al atom; GM: 6.75, LM8b: 5, LM8c: 4.5. Interestingly, the IP prefers the structure which have higher density (LM8c) with higher symmetricity than the highly coordinated atoms in the structure. This is because highly localised electron density can sufficiently provide electrons to atoms evenly.

This preference arises because, in structures with higher density and greater symmetry, the electron density becomes highly localized around the atoms. Highly localized electron density means that electrons are concentrated in specific regions close to the nuclei, enhancing the strength of the bonding interactions. This localization allows the electrons to be more effectively shared or exchanged between atoms, even if each atom has fewer immediate neighbours (*i.e.*, a lower coordination number). As a result, the electrons can sufficiently provide bonding interactions to the surrounding atoms evenly satisfying the bonding requirements of the structure.

In the case of GM8 compared to LM8b and LM8c, although GM8 has a higher average coordination number for aluminium atoms, the electron density in LM8c is more localized due to its higher structural density and symmetry. This localization leads to a more uniform distribution of electrons among the atoms, promoting stronger and more stable bonds within the structure. The IP tends to favour such configurations because they result in lower total energy for the system, making the structure more energetically favourable.

Conversely, in structures with higher coordination numbers but lower density, the electron density may be more delocalized – spread out over larger regions of

space. This delocalization can lead to uneven electron distribution and weaker bonding interactions since the electrons are less concentrated around individual atoms. Consequently, despite having more neighbours, the atoms in these less dense structures might not achieve as effective bonding as those in denser, highly symmetric configurations.

Therefore, the IP prefers structures where the highly localized electron density can evenly distribute electrons to the atoms, resulting in a stable and energetically favourable configuration. This explains why LM8c, with its higher density and symmetry but lower coordination number, is preferred over GM8 by the IP. The even provision of electrons through localized electron density enhances the overall stability of the structure, highlighting the importance of electron localization in determining the energetics and preferred geometries of nanoclusters.

From LM8b, we observe an equivalent number of octahedral SBUs engaging in edge sharing. The square faces of two square-based pyramidal SBU are oriented toward the atom located near the structure's centre within the edge-sharing octahedral SBU. A tetrahedral located on the right side of the LM8b structure, evident in figure 19, terminates in two singularly coordinated F atoms. The 2nd LM is marked by a relative density, yet the tetrahedral SBU branch hosting two undercoordinated fluorine atoms contributes to a higher energy profile. The LM8c, with a denser formation and greater C_s symmetry compared to the C_1 point group of the LM8b, features three edge-sharing octahedral SBU and two edge-sharing trigonal bipyramidal SBU. The close proximity of these atoms induces stronger electronic repulsion, leading to higher energy levels than in the GM and LM8b structures.

Considering clusters up to the n = 8 size, we find that the structural stability of LMs is easily deduced and contrasted through the visual inspection of atomic configurations using the SBU, as it readily reveals symmetry. However, when we approach n = 9, understanding the structural stability among LMs of equivalent cluster size becomes a more complex task, given the increasing randomness in the structure compared to those from n = 6 to 8. The GM of n = 9 comprises five octahedral SBU, with two F atoms within the octahedral SBU being shared by

three Al atoms, revealing a C_2 point group. The presence of a 3-fold coordinated F atom at n = 9 is a novel occurrence, yet the interatomic distance between Al atoms in the edge-sharing octahedral SBU (3.05 Å) surpasses the equivalent distance in the LM9b (2.96 Å). The LM9b presents fewer octahedral SBU, while offering a greater number of square base pyramidal and trigonal bipyramidal SBU, implying a lower average coordination number and a reduced symmetry: C_1 point group. Thus, due to its lower atomic coordination number and decreased symmetry, the 2nd LM is less stable than the GM. The LM9c, which has C_1 symmetry, despite having an equivalent number of octahedral SBU and fewer 5-fold coordinated SBU alongside a greater quantity of tetrahedral SBU, maintains a similar average coordination number for the Al atom. However, clusters tend to favour consistency in atomic coordination numbers, even if the average coordination number of aluminium atoms remains constant.

As shown in Figure 19, the GM9 structure has an octahedral layer as the middle layer of the configuration. A total of five octahedral SBUs are layered in the middle of the structure, and they are connected to each other through edge-sharing, with one corner-sharing connection. From the corner-sharing octahedral, the third octahedral's F atoms are fully coordinated without any undercoordinated F atom terminations. Extending from the octahedral SBU layer, each of the top and bottom layers has corner-sharing tetrahedral and trigonal bipyramidal SBUs. These SBUs are arranged in a slightly off-staggered manner between the top and bottom layers. Compared to LM9b and LM9c, the GM9 has a higher average coordination number of aluminium atoms in the cluster (5.33) with higher symmetry (C_2 point group), while LM9b and LM9c have the same coordination number of 5.11 and belong to the C_1 point group.

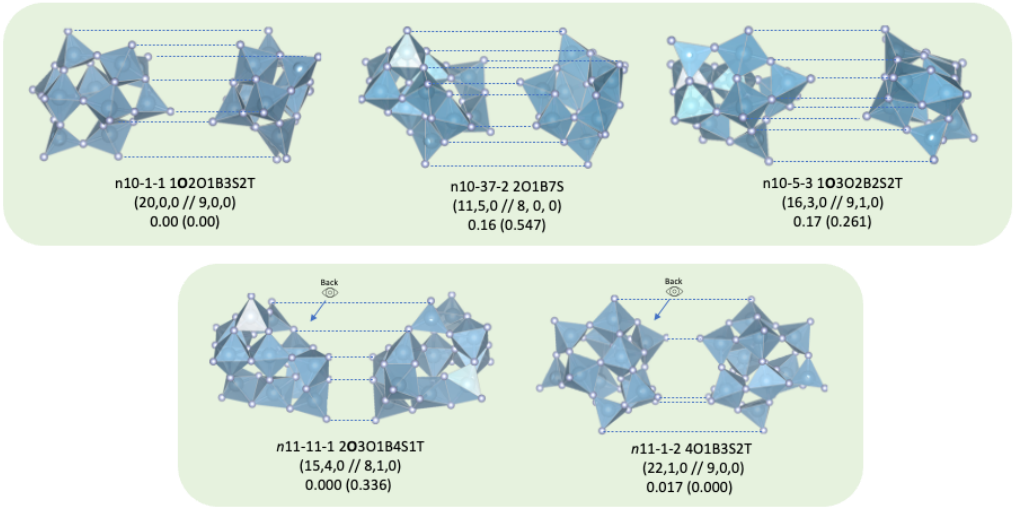

Figure **20**. $(AlF_3)_n$ (n = 10, 11) PBEsol optimised LM: dark (light) sphere represents aluminium (fluorine) atoms; the dashed line connects the same atom in different angles of view of the same configuration. upper label: from the left, atomic size, ranking in IP energy, and ranking in PBEsol energy. SBU component that consists of the configuration; mid-label: on the left-hand side from the divider number of corner-sharing, number of edge-sharing, face sharing, respectively. The right-hand side represents the number of F atoms that bonded with single Al atoms, three Al atoms, and four Al atoms, respectively. bottom label: PBEsol (IP) energy difference between the LM and the GM of the cluster size, n.

The LM9b and LM9c has the same number of the undercoordinated F terminations. The LM9b has a smaller number of octahedral SBU and tetrahedral SBU, but compensate that with more five coordinated SBU, such as square base pyramidal and trigonal bipyramidal SBU. The LM9c has a greater number of corner sharing which is relatively more favourable than edge sharing connectivity between the SBUs, but the LM9c has less densely packed structure as it has longer end-to-end interatomic distance (10.28 Å) compared to the LM9b (9.44 Å). Hence, as the structure is more stable the dipole moment is lower than the higher energy ranking (higher in energy); 2.3233, 3.6302, and 4.8295 Debye in PBEsol rank. The LM9b originally from the 37th ranked IP LM after PBEsol refinement.

A point of interest lies with n = 10, where the lower-energy structures demonstrate greater dipole moments. Specifically, the dipole moments for GM10

and the local minima LM10 and LM10c are 4.910 , 3.8502 , and 0.2814 Debye, respectively.

The analysis of the $(AlF_3)_{10}$ clusters for n = 10 reveals that the most stable structure, GM10, possesses the highest dipole moment of 4.910 Debye. This suggests that the internal charge separation within the cluster contributes significantly to its stability. The elevated dipole moment indicates a pronounced separation of positive and negative charges, leading to stronger internal electrostatic attractions. These interactions lower the total energy of the system, enhancing stability. Conversely, LM10c, with a much lower dipole moment of 0.2814 Debye, lacks this level of internal charge separation and is consequently less stable. This trend underscores the importance of electron distribution and electrostatic interactions in determining the stability of $(AlF_3)_{10}$ clusters.

The inverse trend of the dipole moments with respect to the stability of the structures is intriguing because, generally, a highly stable structure has higher symmetry and, consequently, a lower dipole moment. However, the GM10 structure is formed with the most corner-sharing connectivity between the SBUs compared to the less stable configurations, which indicates less steric hindrance or repulsion between the atoms. Thus, GM10 compensates for the loss of energy associated with its high dipole moment and lower symmetry by having less sterically hindered atomic positions, resulting in greater overall stability.

Each of the presented LM maintains C_1 symmetricity. The GM10, exclusively consisting of corner-sharing SBUs, is the second smallest GM with corner-sharing octahedral SBU since GM6. It represents the smallest cluster size where all corner fluorine atoms in an octahedral SBUs are doubly coordinated, signifying the absence of undercoordinated fluorine atom terminations.

In addition to the fully coordinated octahedral SBU, there are two other octahedral SBUs, a trigonal bipyramidal SBU, two square-based pyramidal SBUs, and two tetrahedral SBUs present in the GM. The LM10b, by comparison, possesses one fewer octahedral SBU and lacks any tetrahedral SBUs. It compensates for these absences with four additional 5-fold coordinated aluminium-centred SBUs. Intriguingly, the LM10b presents a shorter end-to-end

interatomic distance (10.03 Å) than the GM (10.17 Å). Despite the LM10b exhibiting a lower dipole moment and a denser structure, it remains energetically less preferable than the GM10. This reduced favourability results from the steric hindrance induced by edge-sharing SBUs and the decreased interatomic distances between neighbouring atoms.

From a structural motif standpoint, the LM10c contains one (and two) more octahedral SBU(s) than the GM10 (and LM10b), which are edge-sharing. Furthermore, one of the fluorine atoms in the LM10c is over coordinated with three nearest aluminium atoms. In essence, the LM10b, relative to the GM, forfeits any tetrahedral SBUs to accommodate more 5-fold coordinated aluminium-centred SBUs. In contrast, the LM10c hosts fewer 5-fold coordinated aluminium-centred SBUs than the LM10b, but maintains the presence of tetrahedral SBUs and exhibits an increased number of octahedral SBUs.

The IP-LM11k, which sits 0.33 eV above the IP-GM, becomes the GM within the PBEsol PES. This cluster size is the only the cluster size within the range of n = 1 to 11 where the IP's GM did not conserve as the DFT's GM after the refinement. After optimization, the IP-GM is demoted to the LM11b, while the IP-LM11k ascends to the position of the PBEsol GM1.

The proposed GM11 and LM11b both adopt configurations consistent with the C_1 point group. The GM11 is composed of five octahedral SBUs, where all fluorine atoms in three of the octahedral display double coordination. Two of the octahedral SBUs have edge sharing interactions with either a square based pyramidal or a trigonal bipyramidal structure. The remaining three octahedral SBUs are positioned in the middle layer of the structure from the view (figure 20) and they are connected through edge-sharing.

Contrastingly, the LM11b consists of four octahedral SBUs, with only a single instance of edge sharing, where two of the octahedra are linked. We calculated the principal moments of inertia (PMI) of these structures and from the result show that the LM11b appearing more spherical, additional metrics suggest the GM is more stable.

By examining end-to-end interatomic distances and the PMI, it's evident that the GM11 structure is more densely packed compared to the LM11b. The PMI findings highlight that the GM is more elongated, with its extension along the interatomic distance shorter than that in the LM11b. Therefore, despite the seeming sphericity of the LM11b, it's the GM11 that presents a greater degree of stability.

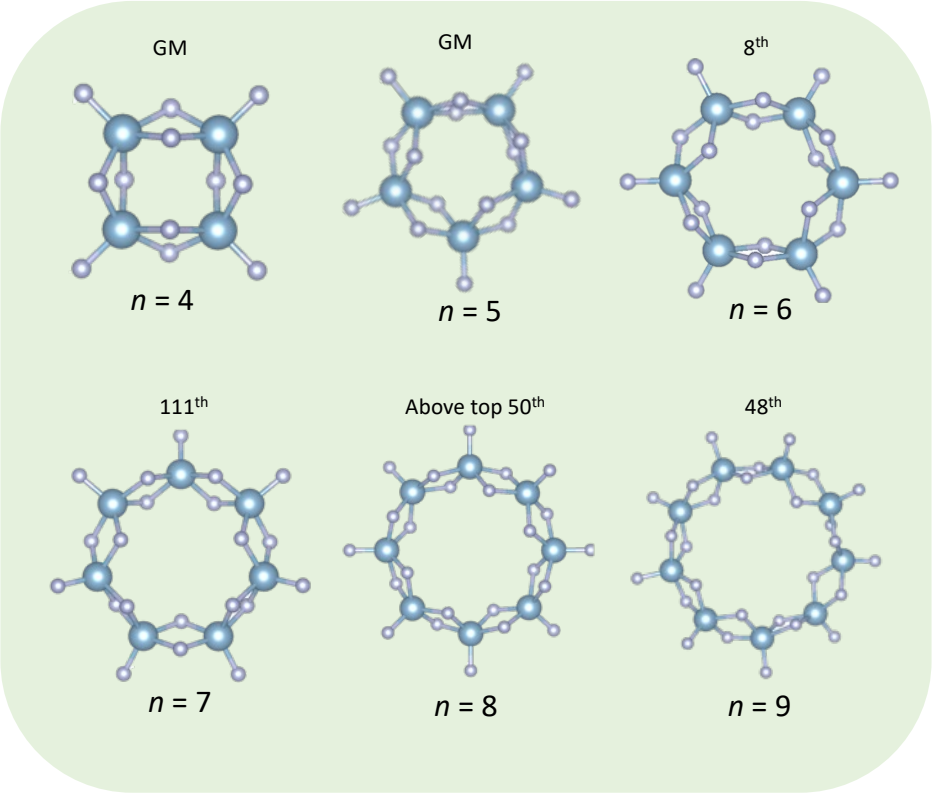

Figure **21**. The ring configurations for each of the cluster sizes from n = 4 to 9. The label above the figure indicates the energy ranking in PBEsol PES.

Throughout the range of cluster sizes, we observed ring configurations exclusively comprising edge-shared square-based pyramidal SBUs in clusters of n = 4, 5, and 6. To extend this observation to other cluster sizes, we generated hypothetical ring configurations, shown in figure 21, using a Python script and then optimised them using PBEsol. The resulting ring configurations remained stable during optimisation, though they ranked high or were less stable. For n = 4 and 5, the small cluster sizes ensure that electron densities are evenly distributed, making them potential GM. Also, the small ring structures would provide the

enough steric hindrance to sustain the ring configurations stably. However, as the ring size increases, the middle hole becomes larger, making these configurations more prone to collapse into other LM due to reduced strain resistance against small perturbations, which would locate in the relatively more narrow LM basin.

3.3.5. Top-Down Approach - Bulk Cut Clusters

In our study, it is noteworthy that the GM6 structures exhibit partial structural features reminiscent of bulk AlF_3 phases: corner shared octahedral. However, the GM6 cluster is not classified as nanocrystals[5,114]. In order to further explore this, we employed a top-down methodology, wherein α-AlF_3 was sectioned. To ensure the stoichiometry of the system was maintained at a 1:3 ratio, surface atoms were selectively removed which was used in previous study instead of the method that we proposed in the methodology chapter 3.3.5. and the bulk cut structure is shown in figure 22 (A, B)[115].

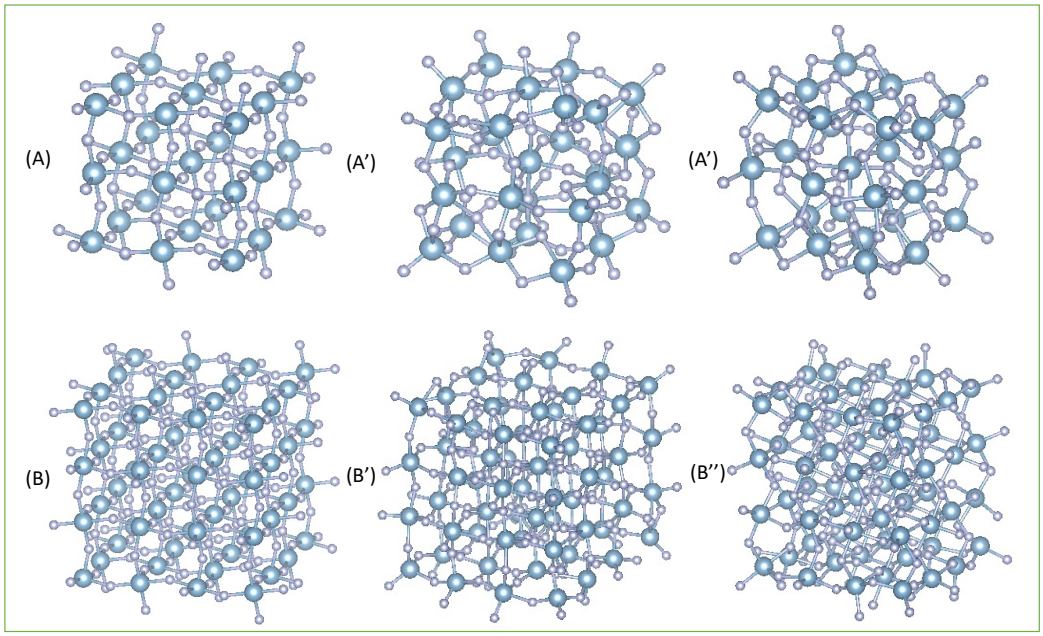

Figure **22**. *Ball-and-stick* representations of pseudo-cubic $(AlF_3)_n$ nanocrystals cut from the bulk phase before (first) and after relaxation to the PBEsol LM (middle) and IP LM (last column) for n = 27 and 64 and where large blue and small grey spheres represent Al and F atoms, respectively.

Figure 22 presents the structural evolution of bulk-cut configurations for n = 27 and n = 64. Panels A' and B' display the structures optimized using DFT, while panels A" and B" showcase the structures relaxed through IP. Upon relaxation, both DFT and IP methods result in disordered surface morphologies. However, the IP-optimized structures exhibit more significant dislocations in atomic positions compared to their DFT counterparts. In the relaxed states, fluorine atoms on the surface transition to bridging positions, establishing connections between adjacent aluminium atoms. Furthermore, the DFT-optimized structures exhibit smoothing at the corners and edges, which is indicative of their lower energetic favourability. The IP-optimized structures also show similar smoothing effects but to a more pronounced extent. Thus, The (A'') IP optimised n = 27 bulk cut become rather more spherical structure than the cubic structure. The (B'') IP optimised n = 64 does not keep the edges and corners as much as the (B') DFT optimised structure, but core keeps the octahedral corner sharing with more pronounced tilting in octahedral.

For n = 27 only one octahedral SBU is remained as the core, and the angle between the octahedral SBU vicinity (Al-F-Al) in (A'') IP optimised structures are 142.5°, 138.2°, and 141.2° which retain close to the angle in the α-AlF_3 148.6°. For the (A') PBEsol optimised structure show rather bigger angle than the α-AlF_3, 160.0°, 152.7°, and 149.0°. On the other hand, for the (B'') IP optimised n = 64 has smaller angle, 129.1°, while the (B') PBEsol optimised structure show 151.3° which is close to the α-AlF_3. As n = 27 and n = 64 retain the bulk structure motif which are nanocrystal, thus we presume the cluster size between n**.**

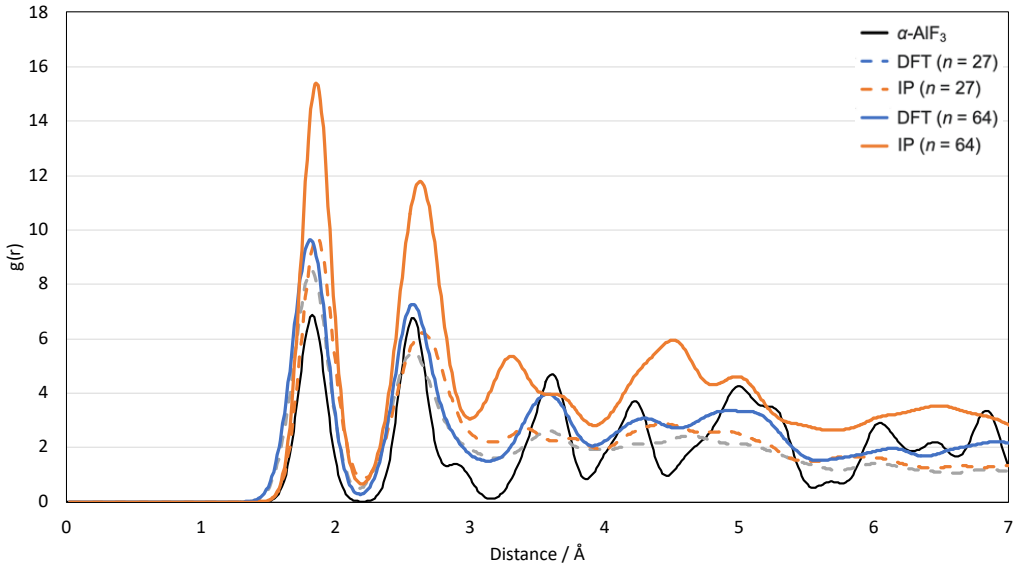

Figure **23**. A comparison of the structures between the IP and DFT calculated bulk cut n = 27, and 64 nanoclusters, and the observed α-AlF_3. (σ = 0.1).

The radial distribution function (RDF) for the experimentally observed α-AlF3 is compared with the IP (orange) and DFT (blue) optimized structures for n = 27 (dashed lines) and n = 64 (solid lines) nanoclusters. The black solid line delineates the α-AlF_3 RDF, where the first to third peaks correspond to Al-F and F-F interatomic spacings. Specifically, the second and third peaks of the black line indicate the F-F distances within the central horizontal plane and between the two F vertices along the vertical axis of the octahedral SBU, respectively. Notably, the n = 27 nanocluster optimized via both IP and DFT exhibits a higher degree of resemblance to each other than the larger n = 64 cluster optimized in the same manner. This close congruence in RDFs between the smaller IP and DFT optimized clusters suggests that IP is well-suited for describing these nanoclusters.

The reduced likelihood of encountering an Al-F distance around 1.9 Å can be attributed to surface termination effects, leading to SBUs with fewer coordination numbers compared to their octahedral counterparts. Conversely, the IP-optimized n = 64 cluster shows a similar probability of Al-F distances as in α-AlF_3. The second F-F peak for IP-optimized n = 64 also closely mirrors that of α-AlF_3. It's

worth noting that the second peaks for both n = 24 and 64 clusters coincide with the Al-Al peaks originating from relaxed surface layers.

As anticipated, the surface layers of the SBUs undergo relaxation, altering the edges and corners to minimize unfavourable energies, while preserving the octahedral SBU core. Although IP strongly suggests that the n = 64 cluster retains its octahedral SBUs, akin to a nanocrystal, both the DFT-optimized n = 64 and the IP/DFT-optimized n = 27 clusters also maintain the structural motifs found in the bulk material. Therefore, we expect the first true nanocrystal to be smaller than n = 27.

3.3.6. First and Second Order Energy, and Dipole Moment

Several approaches exist to investigate the structural stability across varying nanocluster sizes. We already have analysed the structure stability in the context of the structural motifs or secondary binding unit in the previous chapter. In this study, we employ the concepts of first and second order energy to evaluate and contrast the GM associated with each cluster size. The principles of these energy orders are as follows:

$$\Delta E^{1st} = E(n) - E(n-1) - E(n=1), \qquad \textit{Eqn } 46$$

$$\Delta E^{2nd} = E(n) - 1/2 \times \big(E(n+1) + E(n-1)\big), \qquad \textit{Eqn } 47$$

The first-order energy, as described by *Eqn* 48, denotes the energetic differentiation between a given cluster and its one-formula-unit smaller precursor. This is derived by subtracting the total energy of a single formula unit and the GM of the smallest cluster size from the total energy of the cluster of interest. This energy differential symbolizes the energy change accompanying the transformation from a cluster of size (n – 1) to size n, when the (n – 1) cluster is encapsulated within the energy basin of the GM1.

In this context, numerous factors could potentially influence the true trajectory of the cluster's growth. However, the first-order energy primarily focuses on the cluster's potential energy. As such, a larger magnitude in the first-order energy signifies the easier formation of the cluster from its immediate smaller counterpart, with higher potential energy indicating greater stability.

The concept of second-order energy (*Eqn* 48) mirrors that of the first-order energy, denoting the difference in energy between a cluster and its immediate neighbouring sizes, both larger and smaller. A more negative second-order energy implies greater stability of the structure in comparison to its neighbouring cluster sizes. From the previous studies, there have been noticeable correlations between peaks in second-order energy and mass spectra[88]. These peaks indicate points of stability within the structural progression of the nanocluster, underscoring the importance of both first and second-order energies in the evaluation of nanocluster stability.

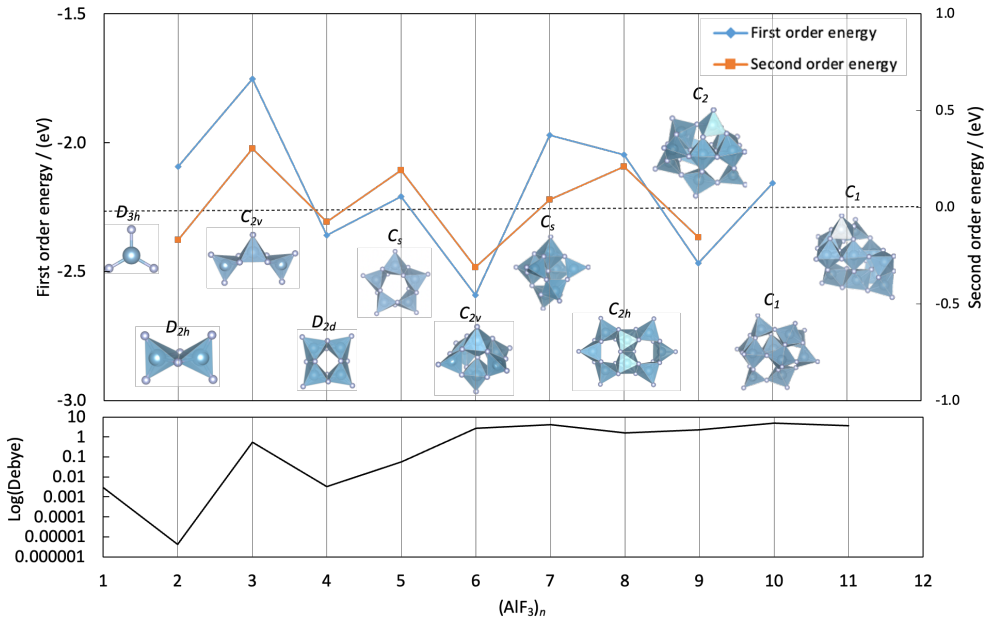

Figure **24**. Upper panel: first (left side *y*-axis, blue solid line) and second (right side *y*-axis orange solid line) order energy of $(AlF_3)_n$ (n = 1 - 11) GM. Bottom panel: log scaled dipole moment of the GM for each cluster size.

$$E_b^{(n)} = V(n) - nV(1), \qquad \textit{Eqn } 48$$

Figure 24 is constructed using the first and second order energies of the GM across a range of cluster sizes (n = 1 – 11). Each cluster is depicted as a space group and *polyhedral* model alongside a complementary *ball-and-stick* representation. Notably, n = 2 demonstrates a relative higher stability in comparison to n = 3, as evidenced by its larger first and second order energy magnitudes, and negative values, respectively.

The SBUs of both clusters are connected through edge-sharing interactions, with the only structural difference lying in the placement of a formula unit between the tetrahedral SBUs of GM2 which is GM3. This placement gives rise to an additional square-based pyramidal structure. The GM2 cluster exhibits higher symmetry, belonging to the D_{2h} space group, compared to the GM3 cluster, which is part of the C_{2v} space group. In general, structures with higher symmetry are considered more stable, thus explaining the superior stability of GM2. Consistent with this, the dipole moment graph in the lower panel of Figure 24 reveals that the more symmetrical structure (n = 2) displays a lower dipole moment.

The investigation of structural stability unveils that the nanocluster with a size of GM4 exhibits a comparatively higher level of stability when compared to nanoclusters with sizes GM3 and GM5. This enhanced stability in the GM4 nanocluster arises from its inherent property of possessing a more symmetric structure (D_{2d}) in contrast to the neighbouring cluster sizes, which adopt C_{2V} and C_s symmetries, respectively. The greater symmetry in the GM4 nanocluster contributes to its increased stability, thereby rendering it more favourable energetically.

Furthermore, this trend of relative stability is further substantiated by the observation of a lower dipole moment in the GM4 cluster in comparison to the dipole moments of the adjacent sizes. The coherent dipole moment supports the notion that the GM4 nanocluster possesses a more energetically stable configuration, thereby bolstering its significance in the context of the study.

The analysis of these plots reveals that the GM4 cluster exhibits a higher magnitude of first order energy compared to the GM2 cluster. However, in terms of second order energy, the GM4 cluster displays a relatively lower magnitude compared to the GM2 cluster. The observed discrepancy in the trends of first and second order energy for the GM2 and GM4 nanoclusters carries important implications for their likelihood of occurrence in nanocluster production processes, such as supersonic expansion. The results suggest that the GM4 clusters are more likely to be present in instances of nanocluster production, while the GM2 clusters may persist for longer periods in the GM1 basin before eventually growing to GM3 and subsequently to the GM4 size.

In contrast, the nanoclusters with a size of GM5 display a distinct structural feature, featuring more distorted SBUs with edge sharing in order to reduce steric hindrance. Thus, the GM5 clusters exhibit reduced symmetry (C_s) and a larger dipole moment, rendering them relatively less energetically favourable compared to the GM4 clusters in both energy schematics.

Furthermore, when compared to the GM6 clusters, which adopt the C_s space group and feature corner sharing octahedral motifs akin to those present in AlF_3 bulk phases, the GM5 clusters remain less stable energetically. The GM6 cluster, in particular, stands out as a remarkable nanocluster size due to its substantial structural stability when compared to other cluster sizes, thus underscoring its potential for diverse applications.

Although the geometry of the GM6 nanocluster demonstrates higher symmetry with a C_{2v} space group in comparison to the GM5 nanocluster with a C_s space group, the former surprisingly exhibits a significantly higher dipole moment.

Observations from the nanoclusters with GM6 reveal a gradual decrease in structural regularity, transitioning towards a more amorphous or random-like arrangement that resists easy explanation using SBUs. As the structure becomes increasingly amorphous, its symmetry diminishes, leading to higher dipole moments, as anticipated. Notably, the GM6 nanocluster features a bulk-like structural motif with greater symmetry (C_{2v}) compared to GM7, rendering the latter less stable. This is evident from the significantly lower magnitude of first

order energy and positive second order energy observed in the GM7 cluster, implying that GM6 is more likely to remain stable in the GM1 basin, rather than transforming into GM7. Moreover, the GM6 nanocluster demonstrates enhanced formability compared to any cluster between GM1 and GM11.

The case of GM8 presents an interesting scenario where the first and second order energy trends exhibit an opposite pattern as we move from GM7 to GM8. The GM8 structure adopts a relatively two-dimensional configuration, resembling the number '8' or two merged rings, each formed by five SBUs. The structural transition from three-dimensional (GM7) to two-dimensional (GM8) and back to three-dimensional (GM9) illustrates the complexity of symmetry changes within this range. Despite having higher symmetry (C_{2h}) compared to GM9, n = 8 is relatively less stable due to its less densely packed structure. Consequently, GM8 is not readily formed. Similarly, while GM7 exhibits higher symmetry than GM8, the latter features a lower dipole moment and higher symmetry, suggesting that the transformation from GM7 to GM8 may occur at a relatively slow rate compared to other clusters.

Moving towards larger clusters, those from GM9 and beyond adopt denser packing and higher symmetry compared to GM10. Consequently, GM9 demonstrates greater stability and is expected to be more frequently observed. The lower symmetry in GM10 renders it less stable than GM9.

In summary, the structural symmetry plays a crucial role in determining the relative stability of nanoclusters, with denser atomic packing also influencing their stability. The GM structures with GM 2, 4, 6, and 9 are relatively more readily formed and likely to be observed in mass spectrometry experiments, both during regular observations and in cases where the structure persists in the GM1 basin over a period of time.

3.3.7. Coordination Number

In the study of nanocluster structures, especially those resembling amorphous systems, we often encounter random-like configurations where the connectivity of

SBUs lacks an apparent pattern to the naked eye. This complexity makes it challenging to determine meaningful coordination numbers for the atoms within these clusters. However, establishing accurate coordination numbers is crucial for understanding the stability and properties of nanocluster structures. Typically, nanoclusters energetically favor higher coordination numbers, but overestimating these values can lead to an exaggerated sense of structural stability. Conversely, underestimating coordination numbers can yield inaccurate insights into the system's behavior.

To address this issue, we focus on determining the appropriate range of cutoff distances that capture precise coordination numbers, thereby providing a more sensible representation of the structural stabilities. The findings from this analysis are then applied to describe the polyhedral model with ball-and-stick representations or SBUs, offering valuable insights into the stability of nanocluster structures.

We employ a Boltzmann-weighted approach to calculate the average coordination number, which accounts for the relative probabilities of each cluster configuration based on their energies. The average Boltzmann-weighted coordination number $<N_{coord}>$ is calculated using the following equation:

$$< N_{coord} >= \frac{\sum_i N_{coord,i} e^{\frac{-\Delta E_i}{k_B T}}}{\sum_i e^{\frac{-\Delta E_i}{k_B T}}}, \qquad \textit{Eqn } 49$$

where, $N_{\mathrm{coord,i}}$ is the coordination number of the iii-th cluster. $\Delta E_{\mathrm{i}} = E_{\mathrm{i}} - E_{\mathrm{min}}$ is the energy difference between the energy of the *i*-th cluster (E_{i}) and the lowest energy cluster (E_{min}), k_{B} is the Boltzmann constant, T is the absolute temperature in Kelvin.

This equation effectively weights each cluster's coordination number by its Boltzmann factor, $e^{\frac{-\Delta E_i}{k_B T}}$, representing the relative probability of the cluster's occurrence at temperature T. The denominator ensures normalization by summing over all Boltzmann factors, resulting in an average coordination number that accurately reflects the thermodynamic ensemble of cluster configurations.

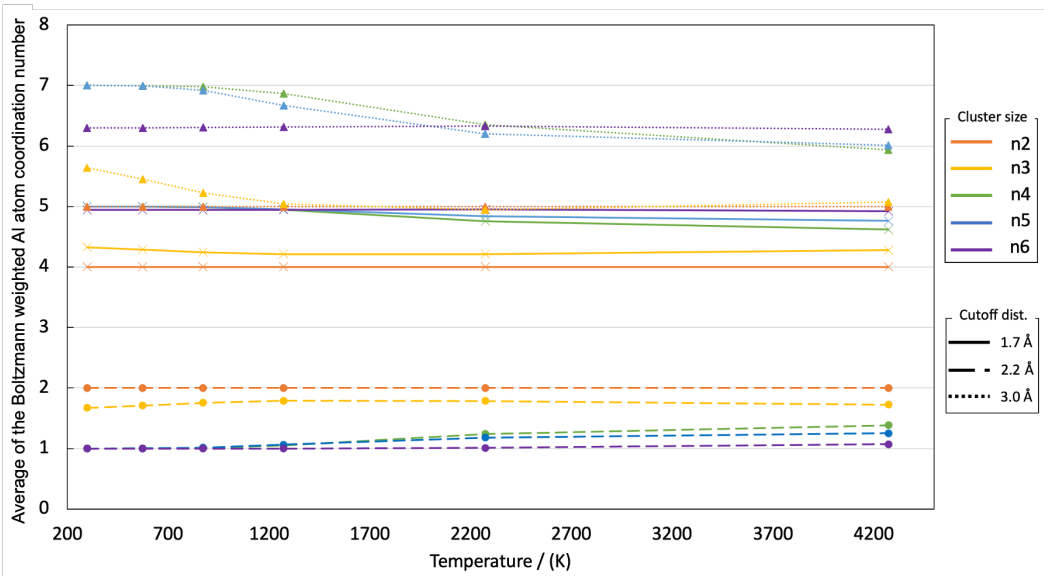

Figure **25**. Average of the Boltzmann weighted coordination number of Al atom of $(AlF_3)_n$ (n = 2 – 6) nanoclusters which the coordination number is calculated using the cutoff distance: (1.7 Å, 2.2 Å, and 3.0 Å) as a function of the range of temperature in between 200 K and 4,200 K; the Boltzmann weighted coordination number of Al atom in the clusters calculated with 3.0 Å is in dashed lines with circle markers; with 2.2 Å is in solid lines with cross markers; with 1.7 Å is in dotted lines with triangle markers; the red to violet colours are correspond to the nanoclusters size.

Figure 25 illustrates the average Boltzmann-weighted Al atom coordination number calculated using the three different cutoff distances (1.7 Å, 2.2 Å, and 3.0 Å) as a function of temperature ranging from 200 K to 4200 K for $(AlF_3)_n$ (n = 2–6) nanoclusters. The data reveal that the average coordination number calculated with a 1.7 Å cutoff distance exhibits an increasing trend with temperature, which is the reverse of the trends observed with the 2.2 Å and 3.0 Å cutoff distances.

The 1.7 Å cutoff distance effectively captures undercoordinated F atoms relative to the Al atom center but fails to account for doubly coordinated F atoms that are at longer distances from the Al atoms. In higher-energy structures, which tend to adopt elongated or ellipsoidal shapes, the 1.7 Å cutoff is better suited to capturing these undercoordinated F atoms. As temperature increases, higher-

energy configurations become more populated due to thermal excitation, leading to an increase in the average coordination number when using the 1.7 Å cutoff.

Conversely, the average coordination numbers calculated using the 2.2 Å and 3.0 Å cutoff distances exhibit a decreasing trend with increasing temperature. This behavior can be attributed to the elongated shapes of the higher-energy structures. Longer cutoff distances like 3.0 Å are more sensitive to changes in the structural geometry because they encompass a larger volume around each Al atom. As the structures elongate, the probability of finding neighboring F atoms within this cutoff decreases along the elongated axis, resulting in a reduced coordination number. The 2.2 Å cutoff distance, capturing primarily the nearest neighbor atoms, shows a gentler decrease because it is less affected by the elongated geometry and more accurately reflects the immediate coordination environment of the Al atoms.

These observations highlight the importance of selecting an appropriate cutoff distance for calculating coordination numbers in nanoclusters. The choice of cutoff distance significantly impacts the calculated coordination numbers, particularly in structures with varying shapes. Based on our analysis, we believe that cutoff distances around 2.2 Å can capture accurate coordination numbers, providing a reliable measure for understanding structural stability through the investigation of SBUs, as discussed in Section 3.3.4.

In summary, the Boltzmann-weighted average coordination number provides valuable insights into the behaviour of nanoclusters at different temperatures. By carefully selecting the cutoff distance, we obtain a more accurate representation of the structural stabilities and better understand how temperature influences the coordination environment within nanoclusters. This approach allows us to further refine the determination of accurate coordination numbers, enhancing our comprehension of the structural properties and stability of these systems.

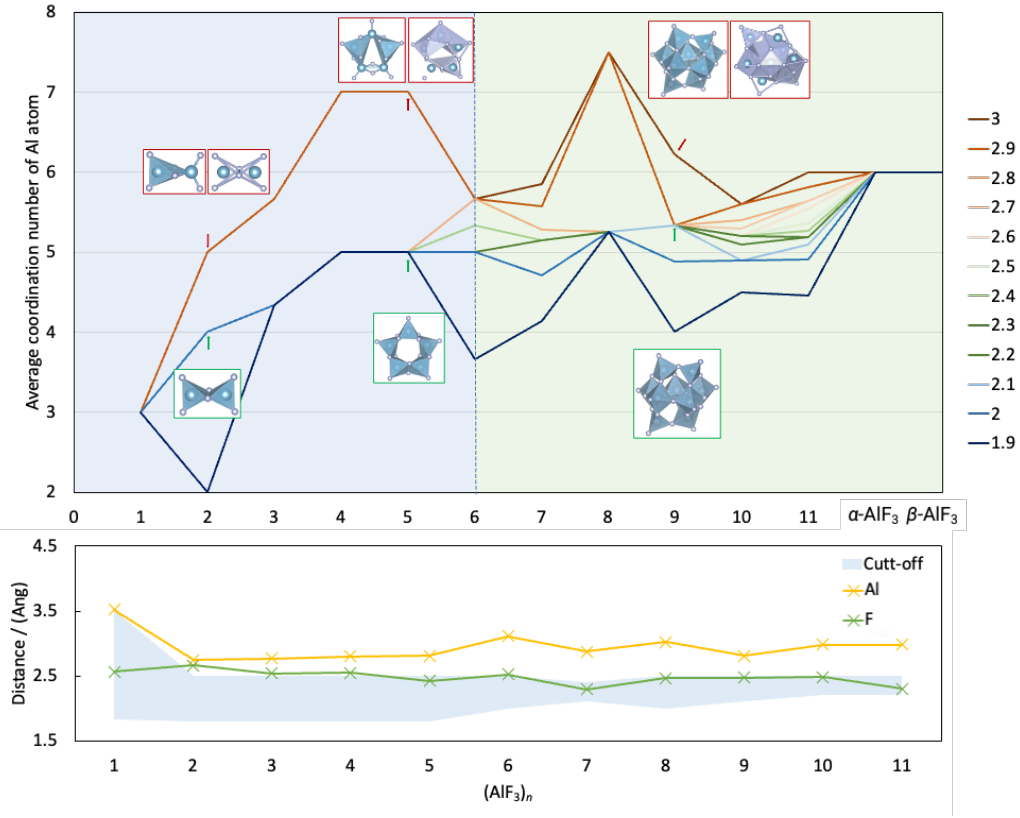

Figure **26**. (upper panel) Variation in the average coordination number of an Al atom in $(AlF_3)_n$ (n = 1 ~ 11) clusters as a function of cluster size and two bulk phases (alpha, beta), considering a range of cutoff distances (d = 1.9 ~ 3.0 Å) from the Al atom. The configurations with the *ball-and-stick* model and the *polyhedral* model are prepared with the cutoff 3.0 Å (upper) and 2.2 Å (lower). The left-side upper *polyhedral* is Al centred and the right-side polyhedral model is F-centred. (lower panel) Shortest Al-Al distance (yellow) and F-F distance (green) in $(GM)_n$ (n = 1 ~ 13). The blue coloured region represents the cutoff distance that can capture an accurate Al atom centred coordination number.

We have calculated average coordination of Al atom for the GM of each cluster sizes (n = 1 ~ 11) and alpha and beta phases of AlF_3 using r_{cut} = 1.9 ~ 3 Å as the cutoff distances. The range of r_{cut} are chosen that the extremely shortest and the longest r_{cut} shows the misleading information (cannot take enough number of atoms which are in the nearest neighbours or take second-nearest neighbour atoms).

In the upper section of figure 26, the mean coordination number of Al atoms within the GM of cluster sizes ranging from n = 1 to 11, including the two most

prominent AlF_3 bulk phases, is delineated for cutoff distances spanning 1.8 Å to 3.0 Å, incrementing at intervals of 0.1 Å. Within the range leading up to $n = 5$ designated by a blue transparent region, periodic patterns in the SBUs are perceptible, corresponding to the structure elucidated previously. Conversely, commencing at $n = 6$ and extending to larger sizes (identified by a red transparent region), the GMs exhibit a reduction in symmetry and an absence of clear patterns in the structural motifs. Coloured solid lines represent the averages of Al atom coordination numbers over the specified range of cutoff distances. The range of 2.2 Å to 2.4 Å encapsulates similar coordination numbers for Al atoms within the cluster sizes, with deviations from this range capturing accurate coordination numbers. The obtained coordination number can be used to construct SBUs of the structure that assist to understand the structural stability. In contrast to nanoclusters, the bulk phases maintain a periodic pattern characterized exclusively by corner sharing octahedral SBUs with 6-fold coordination number of Al atom.

Lower panel of figure 26 shows the shortest interatomic distance of homoatomic species (cation-cation, anion-anion) in the $(GM)_n$ ($n = 1 – 11$) with yellow and green solid-lines, respectively. The blue region is the range of cutoff distances which used to calculate the coordination number of Al atom and able to construct SBUs in the cluster to explain the relative structural stability within the cluster size. The figure support that 2.2 Å – 2.4 Å range of cutoff distances can be used for the cluster sizes that we predicted ($n = 1 – 11$). It is posited that a cutoff distance encompassing 2.2 Å to 2.4 Å is judicious for representing the Al atom coordination number within nanoclusters. A convergence to a 6-fold coordination number with respect to Al atoms would be observed as the cluster size augments, in line with the range of cutoff distances.

In this study, the complex nature of nanocluster structures, particularly $(AlF_3)_n$ clusters, was explored with a focus on determining accurate coordination numbers for Al atoms. By analysing different cutoff distances ranging from 1.7 Å to 3.0 Å, the research identified the significant impact of these distances on the calculated coordination numbers, especially in elongated structures. The findings revealed that a cutoff distance of 2.2 Å to 2.4 Å provides an accurate representation of the Al atom coordination number within nanoclusters, capturing

the nearest neighbour atoms without including second-nearest neighbours. This range offers valuable insights into the structural stability of nanoclusters, allowing for the construction of SBUs that explain relative stability within various cluster sizes. The study emphasizes the importance of selecting an appropriate cutoff distance to understand the behaviour of nanoclusters at different temperatures and provides a foundation for further investigation into the structural properties of these complex systems.

3.3.8. Lewis Acidity

AlF_3 is often noted in theoretical studies for possessing a high degree of Lewis acidity[96,97]. This characteristic, however, is notably absent in the α-AlF_3 structure. The lack of observable Lewis acidity in α-AlF_3 can be attributed to the full coverage of Al atom sites by F atoms, which negates the potential for the Al to act as a Lewis acid.

Nanoclusters represent a unique system that offers unparalleled advantages, particularly in catalytic applications. The prominent feature of having the highest surface-to-volume ratio, where all constituent atoms are exposed to the external environment, allows nanoclusters to serve as highly effective catalysts. In the context of AlF_3 nanoclusters, the ability to manipulate these structural characteristics offers new possibilities for uncovering the underlying mechanisms that govern Lewis acidity and how they can be harnessed.

Here, the investigation will specifically focus on the Lewis acidity of $(AlF_3)_n$ nanoclusters, where n ranges from 1 to 11. This analysis aims to provide insights into the GM clusters of AlF_3, considering various influencing factors. By accounting for both structural and chemical attributes, the study seeks to unravel the intricate connections that lead to the unique Lewis acidity properties of these clusters. In doing so, the work not only enhances our understanding of AlF_3 nanoclusters but also offers potential applications in the development of novel catalytic systems.

3.3.8.1. Structural Factor

The Lewis acid catalytic reaction's focal point lies at the Al atom, which actively accepts electron pairs from Lewis bases. Emphasizing the introductory remarks, the structure of the nanoclusters plays a pivotal role in influencing the catalytic performance. To further examine these structural factors, analyses were carried out on both the electrostatic Hartree potential and the Solvent Accessible Surface Area (SASA) within the GM of the system, offering a comprehensive insight into the accessible sites for Lewis acid catalytic reactions.

The calculated Hartree potential of $(AlF_3)_n$, where n ranges from 1 to 11, presents an intricate landscape. Utilizing the HSE06 hybrid function with 0.11 $bohr^{-1}$ for omega, the resulting visualization, Figure 27, displays the net positive and negative Hartree potential isosurface regions in yellow and cyan, respectively. A key observation is that the net positive Hartree potential surrounds undercoordinated F atom terminations, an aspect that prevents interactions with Lewis acidic sites due to repulsive electron interaction.

The Hartree potential is a fundamental concept in electronic structure calculations, particularly within the framework of Density Functional Theory (DFT). It represents the classical electrostatic interaction experienced by each electron due to the presence of all other electrons in the system. Mathematically, the Hartree potential VH(r)V_H(r)VH(r) is derived from the electron density ρ(r) and is given by the integral:

$$V_H(r) = \int \frac{\rho(r')}{|r-r'|}\, dr', \qquad \textit{Eqn } 50$$

This potential accounts for the average repulsive force between electrons, effectively capturing the electron-electron Coulomb interactions without considering the quantum mechanical exchange and correlation effects, which are addressed separately by the exchange-correlation functional in DFT. In the context of our study on $(AlF_3)_n$ clusters, the calculated Hartree potential provides a spatial map of regions with net positive and negative electrostatic potential, visualized as yellow and cyan isosurfaces in Figure 27. The visualization reveals that areas surrounding undercoordinated fluorine atom terminations exhibit net positive

Hartree potential. This indicates regions where electron density is lower, resulting in a repulsive electrostatic environment that inhibits these F atoms from interacting favorably with Lewis acidic sites. Such repulsive interactions are crucial for understanding the stability and reactivity of the clusters, as they influence how electron-rich and electron-deficient sites within the material interact. By analyzing the Hartree potential, we gain valuable insights into the distribution of electron density and the resulting electrostatic landscape, which are essential for predicting the chemical behavior and stability of the $(AlF_3)_n$ nanoclusters.

Contrarily, regions containing Al atoms or higher-coordinated (> 2-fold coordination number) F atoms exhibit a net negative Hartree potential, indicating readiness to react with Lewis bases. However, the presence of undercoordinated fluorine atom terminations nearby may counterbalance these negative potentials, creating a unique dynamic. Moreover, the *isosurface* model of the Hartree potential exhibits irregular patterns in clusters n = 6, 7, and 9 to 11, where uncapped negative potentials by undercoordinated F atoms seem more pronounced compared to the $n \leq 5$ GM clusters. These anomalies align with the clusters' random-like structural properties, leading to distinct net dipole moments (figure 26).

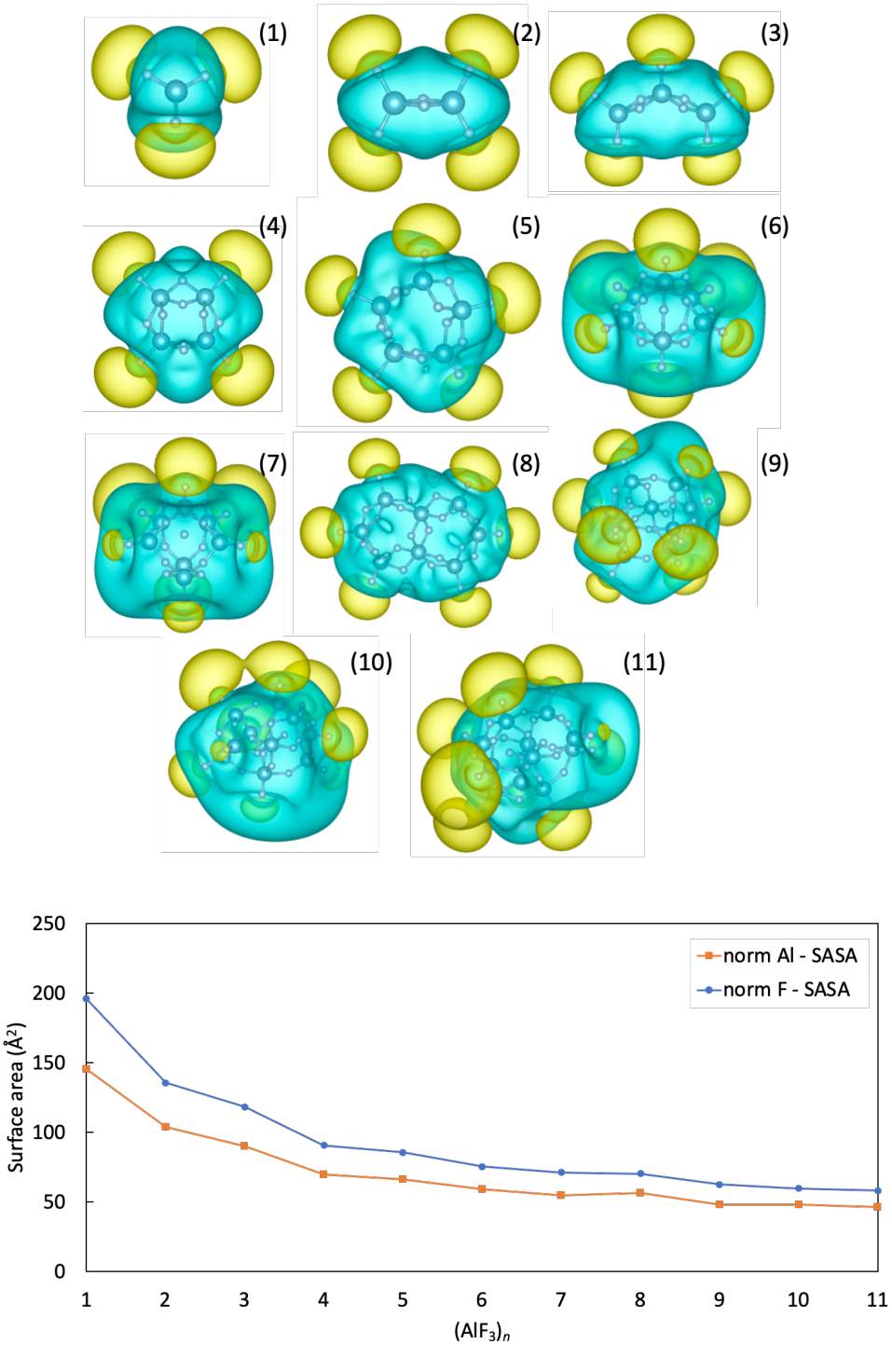

Figure **27**. (Upper Panel) Illustration of the Hartree potential for $(AlF_3)_n$ GM (n = 1 – 11) alongside the *ball-and-stick* model of the GM; areas marked with *isosurfaces* model in yellow represent positive Hartree potential, while cyan regions signify negative Hartree potential. (Lower Panel) Display of the SASA for both Al and F atoms within the global minima, plotted in accordance with the cluster size.

The evaluation of Solvent Accessible Surface Area[116,117] (SASA) introduces another layer of complexity. The SASA was calculated using PyMOL[118] which employs the Shrake-Rupley algorithm[119] with a probe radius of 1.4 Å, the SASA within the $(GM)_n$ reveals convergence as the cluster size expands. Nonetheless, specific transitions, such as from n = 3 to 4, present a marked noise. This shift, characterized by a transformation from a rod-like to a ring-like shape, reduces the exposure of Al atoms, illustrating how subtle structural variations can significantly impact reactivity.

On the other hand, the n = 8 cluster displays an "8" shape, leading to a considerable increase in the exposure of Al atoms relative to more densely packed configurations. This unique structural feature underscores the multifaceted relationship between form and function in AlF_3 nanoclusters.

In summary, the present analysis, focusing on both the Hartree potential and SASA within the geometric model, reveals a complex interplay between structural properties and Lewis acidity. Understanding these structural factors that influence the catalytic activity not only enhances our comprehension of catalytic behaviour but also aids in the development and optimization of novel catalytic systems. This has specific applications across various industrial and scientific domains.

3.3.8.2. Chemical factor

For each nanocluster size n = 1 to 11, the GM structures were assessed using the PBEsol0 hybrid functional within FHI-aims to determine the electron affinity, as well as the energies of the HOMO-LUMO gap and LUMO levels (as depicted in Figure 28). A greater positive value of electron affinity signifies that the cluster is more predisposed to accepting additional electrons, as defined in *eqn* 52:

$$\text{Electron affinity} = V(n)_{q=-1} - V(n)_{q=0}, \qquad \textit{Eqn } 51$$

The GM1 cluster has a slightly negative electron affinity of -0.0224 eV, indicating a slight reluctance to accept an additional electron. The GM2 cluster has an electron affinity of 0.0000 eV, suggesting neutrality in electron acceptance.

As the cluster size increases from $n = 3$ to $n = 5$, there is a sequential increase in electron affinity, reaching a maximum of 1.0124 eV at $n = 5$. This indicates that larger clusters are generally more predisposed to accepting additional electrons, reflecting an enhanced Lewis acid character.

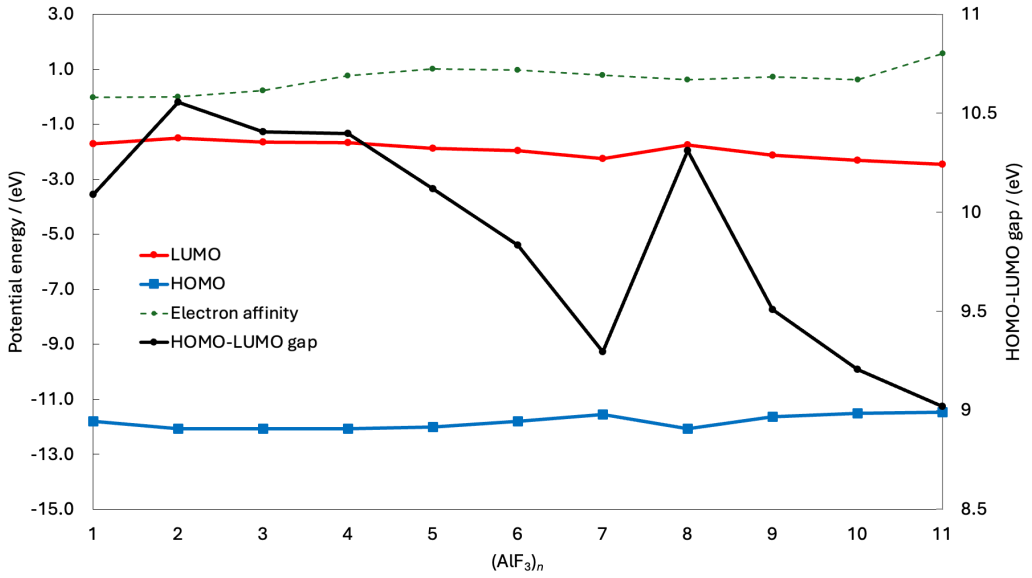

Figure **28**. Electron affinity (grey), LUMO energy level (blue), and HOMO-LUMO gap (orange) as a function of $(AlF_3)_n$ $n = (1 - 11)$.

The LUMO energy level also serves as an indicator of the readiness to accept an electron. A lower LUMO energy level signifies a greater readiness to occupy an electron from the environment, reflecting a more potent Lewis acid behavior. From the table, we see that the LUMO energies become more negative with increasing cluster size, particularly from $n = 7$ onward, correlating with the higher electron affinities observed in larger clusters.

Our computational investigations into the electronic structure of $(AlF_3)_n$ nanoclusters have yielded noteworthy findings, particularly concerning the HOMO-LUMO gap. Across the range of cluster sizes studied, from $n = 1$ to $n = 11$, the HOMO-LUMO gaps were observed to fall within a relatively narrow range of 9 to 11 eV. This suggests a remarkable consistency in the electronic properties of these nanoclusters as they increase in size.

From figure 28, we observe that the HOMO-LUMO gap starts at approximately 10.09 eV for $n = 1$ and shows minor fluctuations with cluster size. Notably, the gap decreases to a minimum of about 9.02 eV at $n = 11$, with a significant drop at $n = 7$ and a spike at $n = 8$.

These fluctuations in the HOMO-LUMO gap across different cluster sizes can be attributed to the evolving orbital compositions and interactions within the clusters. In the smallest cluster, $(AlF_3)_1$, the HOMO is primarily composed of non-bonding fluorine 2p lone pair orbitals, which are relatively high in energy due to the high electronegativity of fluorine atoms. The LUMO is mainly derived from the empty aluminum 3s and 3p orbitals. The significant energy difference between the filled fluorine 2p orbitals and the empty aluminum orbitals results in a large HOMO-LUMO gap.

As the cluster size increases to $n = 2$ and $n = 3$, the HOMO-LUMO gap slightly increases to approximately 10.56 eV and 10.40 eV, respectively. This slight increase can be attributed to the stabilization of the HOMO levels due to the formation of Al – F bonds in the dimer and trimer structures, which lowers the energy of the HOMO orbitals to around -12.06 eV. The LUMO levels for these clusters are at -1.51 eV and -1.66 eV, respectively. The HOMOs remain largely dominated by fluorine 2p non-bonding orbitals, but the increased bonding interactions slightly adjust their energies.

For clusters with $n = 4$ to $n = 7$, there is a noticeable trend of decreasing HOMO-LUMO gaps, from 10.40 eV at $n = 4$ to 9.29 eV at $n = 7$. This decrease corresponds with the HOMO energy becoming less negative (from -12.06 eV at $n = 4$ to -11.55 eV at $n = 7$), while the LUMO energy becomes more negative (from -1.67 eV to -2.25 eV). The reduction in the HOMO-LUMO gap is due to increased orbital overlap and delocalization as the clusters become larger and more three-dimensional configurations.

In these clusters, the aluminum atoms exhibit higher coordination numbers, and the HOMO starts to include contributions from bridging fluorine atoms that link multiple aluminum centers. This results in increased delocalization of the HOMO over the cluster framework. Simultaneously, the LUMO incorporates

more antibonding character involving Al – F interactions, which lowers its energy. The enhanced overlap between the occupied and unoccupied orbitals reduces the energy difference, leading to a smaller HOMO-LUMO gap.

At n = 8, an increase in the HOMO-LUMO gap to approximately 10.31 eV is calculated. This anomaly can be attributed to a structural change where the cluster adopts a more symmetrical and possibly more compact geometry, such as an 8-shaped, double ring, configuration. This configuration causes a significantly lower structural density relative to the GM in adjacent cluster sizes. The reduced overlap between the HOMO and LUMO orbitals leads to less effective delocalization, consequently increasing the energy gap.

For larger clusters with n = 9 to n = 11, the HOMO-LUMO gap decreases steadily from 9.51 eV to 9.02 eV. The HOMO energies continue to become less negative (*e.g.*, -11.64 eV at n = 9 to -11.47 eV at n = 11), while the LUMO energies become more negative (from -2.13 eV to -2.45 eV). In these larger clusters, the increased size allows for more extensive delocalization of both the HOMO and LUMO orbitals across the cluster.

The electron affinity trends can be correlated with changes in the clusters' electronic structures and geometries. As the clusters grow larger, the increased number of atoms and the more complex bonding environments allow for greater delocalization of electrons. This delocalization stabilizes the addition of an extra electron, thus increasing the electron affinity. In the smallest clusters (n = 1 and n = 2), the reluctance or neutrality towards accepting an electron can be attributed to the limited capacity to delocalize and stabilize an extra electron due to their simple geometries.

At n = 5, the electron affinity reaches a maximum, indicating the most favorable conditions for electron acceptance among the studied clusters. This could be due to an optimal balance between cluster size and structural configuration that maximizes electron delocalization and stabilization. Beyond n = 5, the electron affinity slightly decreases for n = 6 and n = 7, though it remains relatively high compared to smaller clusters. This decrease may be due to structural factors that limit further delocalization or introduce strain in the clusters.

At n = 11, there is a notable increase in electron affinity to 1.5773 eV, the highest among all clusters studied. This suggests that very large clusters may again favor electron acceptance, possibly due to the presence of additional sites for electron localization or enhanced overall polarizability of the cluster.

The relatively high values of the HOMO-LUMO gap across most sizes indicate that these nanoclusters are generally more stable and less reactive. A larger gap generally corresponds to reduced chemical reactivity, making these clusters suitable for applications requiring chemical stability. Additionally, the high HOMO-LUMO gap suggests that a substantial amount of energy would be needed to excite an electron from the HOMO to the LUMO, an important factor for optical and electronic applications.

Understanding the nature of these orbitals and their interactions provides valuable insights into the electronic properties of $(AlF_3)_n$ clusters. The enhanced orbital overlap and delocalization in larger clusters lower the energy difference between the HOMO and LUMO, resulting in smaller band gaps. These insights are crucial for potential applications where the electronic properties of aluminum fluoride clusters play a significant role, such as in catalysis or materials science.

By examining both the numerical data and the orbital compositions, we conclude that the observed trends in electron affinity and HOMO-LUMO gaps are a direct consequence of changes in orbital overlap and bonding interactions within the clusters as they increase in size. This detailed understanding enhances the discussion about the trends in the band gap with orbital overlap, making it more meaningful and informative.

4. Aluminium Hydride

4.1. Introduction

In this chapter, we explore AlH_3 nanoclusters. AlH_3 is known for its potential in hydrogen storage and as a reducing agent in various chemical reactions[118]. Our research focuses on predicting the structures of AlH_3 nanoclusters using computational methods, similar to our study of AlF_3. AlH_3 has unique properties but is not well understood at the atomic level.

Like AlF_3, AlH_3 has a 1:3 ratio of Al atoms to H atoms (Al

= 1:3) and possesses a complex PES that is challenging to model relative to well-known 1:1 stoichiometric systems (such as ZnO or NaCl). This complexity increases at the nanocluster scale, requiring intensive computational efforts to map the PES accurately. While the properties of bulk AlH_3 are known, the study of AlH_3 nanoclusters is relatively new, with surface phenomena significantly influencing their behaviour.

To tackle these challenges, we use the two-step method, which combines classical and quantum mechanical computational techniques, as we applied in the AlF_3 study. Our goal is to identify potential configurations of AlH_3 nanoclusters that might exhibit optimal properties for potential applications once synthesised.

Again, the stoichiometry of AlH_3 adds complexity, similar to AlF_3 discussed earlier. This increases the dimensionality of the PES, requiring more computational resources. However, our previous work on AlF_3, which shares the same stoichiometry with AlH_3, allows us to data mine these data and so significantly reduce computational efforts. In our study, we used the PES of AlF_3 and adjusted it to take into account the differences in the ionic polarizability and radii with AlH_3.

In our study, we developed interatomic potentials tailored for AlH_3, taking into account the differences in bonding nature and atomic properties compared to

AlF_3. This approach ensures that the computational models accurately reflect the physical and chemical behaviour of AlH_3 nanoclusters.

The findings are provided below in a comprehensive overview of the predicted structures of AlH_3 nanoclusters, discussing their stability, electronic properties, and computational challenges.

4.2. Interatomic Potential for Al-H Interactions

4.2.1. Al-H Buckingham Potential

The Al – H Buckingham potential is derived by modifying the Al – F potential parameters. The *A* parameter for the Al – H interaction is calculated using the ionic radii of the hydrogen and fluorine ions and ρ is the exponential decay length parameter from the Al – F interaction in the Buckingham potential. Given that the Pauling ionic radius[113], R_x (x = atom species), of hydrogen ion is 1.4 Å and that of fluoride ion is 1.36 Å, the A parameter for the Al – H interaction can be expressed as[120]:

$$A_{Al-H} = A_{Al-F} \times exp\left(\frac{R_H - R_F}{\rho_{Al-F}}\right), \qquad Eqn\ 52$$

This equation leverages the ratio of the ionic radii of hydrogen ion and fluoride ion to adjust the *A* parameter from the Al – F potential to suit the Al – H interaction. The *C* parameter for the Al – H interaction, however, is kept the same as that of the Al – F interaction, i.e. zero due to the very low polarisability of Al^{3+} ion, ensuring consistency in the repulsive component of the potential.

4.2.2. H-H Buckingham Potential

For the H – H interaction, the *A* parameter is similarly derived from the F – F *A* parameter, using the ratio of the ionic radii of hydrogen ion and fluoride ion. The equation for the *A* parameter of the H – H interaction is:

$$A_{H-H} = A_{F-F} \times exp\left(2\left(\frac{R_H - R_F}{\rho_{F-F}}\right)\right), \quad \textit{Eqn } 53$$

The ρ parameter for the H-H interaction is fixed to that of the F-F interaction. This approach maintains the balance between the attractive and repulsive forces in the H-H interaction, as modelled in the F-F interaction.

4.2.3. Calculation of *C* Parameter for H-H Interaction

The determination of the *C* parameter for the H-H interaction in our model is a crucial step. This parameter is derived from the C_6 terms [121] of both H-H and F-F interactions. The C_6 term for each interaction is computed using the static ionic polarizability, denoted as $\alpha_{species}$, and the effective number of electrons contributing to the polarizability, represented by $P_{species}$. Specifically, for the H ion, these values are α_H:1.305 and P_H:1.430, and for the F ion, the corresponding values are 1.192 and 4.455, respectively[122]. The C_6 coefficient are formulated according to the Slater-Kirkwood approximation[123] using the participation numbers reported by Pyper *et al*[122]:

For the F-F and H-H interaction:

$$C_6^{F-F} = \frac{3}{4}\left(\alpha_F^{3/2} \times P_F^{1/2}\right), \quad \textit{Eqn } 54$$

$$C_6^{H-H} = \frac{3}{4}\left(\alpha_H^{3/2} \times P_H^{1/2}\right), \quad \textit{Eqn } 55$$

Subsequently, the *C* parameter for the H-H interaction is calculated using the ratio of the C_6 terms for F-F and H-H interactions, as shown in the equation:

$$C_{H-H} = \frac{C_6^{H-H}}{C_6^{F-F}} \times C_{F-F}, \quad \textit{Eqn } 56$$

The calculated C_6 and *C* parameters for F-F and H-H interactions are shown below:

Table **5**. The modified C_6 coefficient and C parameter of H-H

Interaction	C_6 **/ eV Å^6**	C **/ eV Å^6**
H-H	0.5959	6.67
F-F	1.4147	15.83

4.2.4. Refinement of Cutoff Distances in the Four-Region Buckingham

In the four-region Buckingham potential, which includes four distinct potential forms, three specific cut-off distances are needed to transition between these forms. For H – H interactions in our AlH_3 nanocluster system, we recalibrated these cut-off distances using the ratio of the Pauling ionic radii of the H ion and the F ion.

$$d_{\text{cut-off}}^{H-H} = d_{\text{cut-off}}^{F-F} \times \frac{R_H}{R_F}, \qquad \textit{Eqn } 57$$

where $d_{\text{cut-off}}^{F-F}$ is the original cut-off distance for F – F interactions and R_H and R_F represent the Pauling ionic radii of the hydrogen ion and fluoride ion, respectively. This ensures that the H – H cutoff distances are accurately scaled, reflecting the differences in ionic sizes. The table below compares the original F – F cutoff distances with the newly calculated H – H cutoff distances:

Table **6**. The modified cutoff for the H – H four-region Buckingham potential based on the F-F potential.

Interaction	***Cut1*** **(Å)**	***R_minimum*** **(Å)**	***Cut2*** **(Å)**	**Cutoff (Å)**
H-H	2.06	2.87	3.12	12.0
F-F	2.00	2.79	3.03	12.0

4.2.5. Spring Constant for H-H Interaction

In our model, the spring constant for H is recalibrated based on that of F, considering their ionic polarizabilities. The recalculation formula is:

$$K_{\mathrm{H}} = \frac{\alpha_H}{\alpha_F} K_{\mathrm{F}}, \qquad Eqn\ 58$$

using the ionic polarizabilities of fluoride and hydrogen, respectively, and K_F is the spring constant of F. This ensures that the H spring constant accurately reflects the polarizability differences between H and F ions. By adjusting parameters based on ionic radii, polarizability, and electron number, we fine-tune the interatomic potentials to reflect the distinct characteristics of Al-H and H-H interactions.

Table **7**. Modified spring constant of H based on F.

Ion	Spring Constant, k (eV/Å^2)
H	18.97
F	20.77

4.3. Results and Discussions

4.3.1. Global Optimisation – Data Mining

To understand AlH_3 nanoclusters, we used a global optimisation strategy based on data mining, focusing on structural data from AlF_3 and LaF_3 clusters in previous chapters. This approach helped us predict structural patterns in AlH_3 nanoclusters, considering known trends in AlF_3 and the unique characteristics of AlH_3.

Using this data-centric method, we predict that the smallest AlH_3 nanocluster will have only one configuration. Similar to the octahedral SBUs in AlF_3, AlH_3 clusters might show a similar structural pattern. AlH_3 bulk phases also only have octahedral corner sharing SBUs[124]. However, due to differences in ionic radii between fluoride (F^-: 1.547 Å) and hydrogen (H^-: 1.399 Å)[125], AlH_3 could resemble AlF_3 clusters rather than LaF_3, given the ionic radius difference of -0.148

Å for H ion compared to F ion. The smaller ionic size allows the cation to interact in limited spaces, providing a steric hindrance.

Data mining as a global optimisation method helps predict probable configurations of AlH_3 nanoclusters and reduces computational resource usage. However, relying solely on data mined from AlF_3 clusters could affect accuracy despite the extensive AlF_3 database. To verify the sufficiency of data mining, we conducted GA global optimisation using the adapted AlH_3 potential from AlF_3 potentials for $n = 3$ and $n = 6$.

4.3.2. AlH_3 nanocluster structures

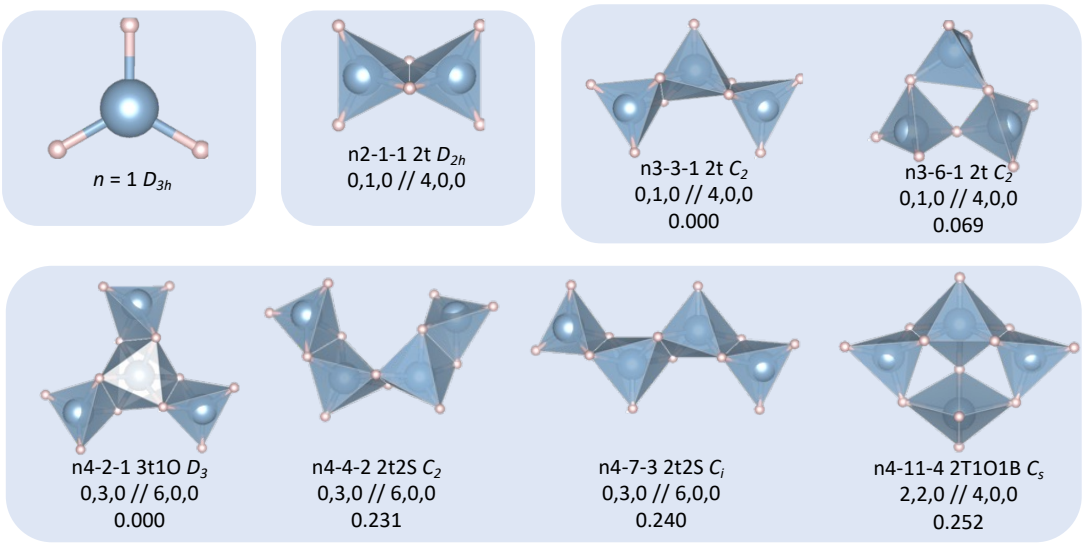

Figure **29**. *Ball-and-stick* models of the tentative lowest PBEsol energy $(AlH_3)_n$ PBEsol LM for sizes $n = 8$ to 11. Blue (grey) spheres represent aluminium (hydride) atoms, and transparent blue polyhedral are AlH_m SBUs, with $m = 4$ to 6. Each configuration is labelled "*n*X-Y-Z S v,e,f // g,h,i,j P (Q)", where X is its size, Y is its IP rank, Z is its PBEsol rank, S is a character string describing the SBUs that the cluster is composed of, v (e, f) is the number of vertices (edges, faces) shared between the SBUs, g (h, i, j) is the number of F atoms with a coordination of 1 (2, 3, 4), and P (Q) is the PBEsol (IP) energy difference between the cluster shown and the ground state cluster of the same size. The character string S is a concatenation of character strings with the format mU that indicate that are m

SBUs of type U, where U is the first letter of the name of SBU shape (tetragonal, square-base-pyramidal, bipyramidal, octahedral), and when in bold lowercase, lowercase, uppercase or bold uppercase the SBU contains at least three, two, one and zero singly coordinated F atoms, respectively. Coordination numbers (and SBUs) are calculated using an interatomic Al-H cutoff distance of 2.2 Å.

In our global optimisation efforts for AlH_3 nanoclusters, again, we concentrate on pinpointing the minimal nanocluster dimension that similar to the structural motif found in bulk AlH_3. Our aim is to identify a cluster size at which the structural characteristics resemble those observed in the larger bulk material, nanocrystal. The AlH_3 bulk phases are composed only with the octahedral SBUs interlinked through corner-sharing connections.

The figure 29 shows the clusters for n = 1 and 2 have identical configuration as the AlH_3 and which the size has only the configuration within the PES. The structures are trigonal planar, D_{3h} for n = 1. The n = 2 AlH_3 composed of the two tetrahedral connected with edge sharing, D_{2h}. It was expected n = 2 to have only one configuration as the AlF_3 which have bigger anion also has one configuration, which have more room for Al to allocate. Beginning at this point, clusters of $(AlH_3)_n$ exhibiting GM or LM will be designated as GMn or LMn(rank), respectively, with the alphabet indicating their ranking order for LM. The atomic connectivity of the GM3 is identical to the GM3. The ionic polarizability of H^{-1} (1.305) is bigger than the F^{-1} (1.192)[126] which allows the GM3 to distort the central square-base pyramidal more than AlF_3 to reduce the steric hindrances compare to the GM3. The distortion eliminates the mirror plane which the plane goes through the left/right end of F atoms and the central Al atom, thus $(AlH_3)_3$ GM has C_2 space group.

The atomic connectivity of the LM3b is also identical to the AlH_3: Composed of three tetragonal with corner-sharing to each other of SBU. LM3b has less average Al coordination number than the GM. As the LM3b has longer Al-Al distances (GM: 2.82 Å , and 2nd LM: 3.38 Å) with lower coordination number compared to the GM3 the size prefers to have higher coordination number rather than less steric hinderance.

The GM4 has the 2nd LM $(AlF_3)_4$ atomic connectivity with the square base pyramidal ring shape. The GM4 has D_3 symmetry and three-pointed star shape composed of the no F termination octahedral in the centre and tetrahedral connected on the octahedral SBU through edges, the vertices are not shared with each tetrahedral. The $(AlH_3)_4$ GM has lower coordination number with 4.5 than the $(AlF_3)_4$ GM which has 5 coordination number from having three-pointed configuration. The GM4 compensate the lower coordination number than $(AlF_3)_4$ by locating the H atom in the end of the configurations. The H atom has higher polarizability than F atom thus it can be pushed away further to form the octahedral SBU. This analogy can be applied for the LM4b and the LM4c. The atomic connectivity of the LM4b and LM4c are same, but only the difference is the direction of the tip of the square base pyramidal SBU. As H atom has the degree of polarizability which can distort the SBUs to allocate the tip of the pyramidal on the same directions. The distortion gives the unique bended structure. On the contrary, the LM4c has overall linear configurations which the pyramidal tips are pointing opposite directions including the tip of the tetrahedral.

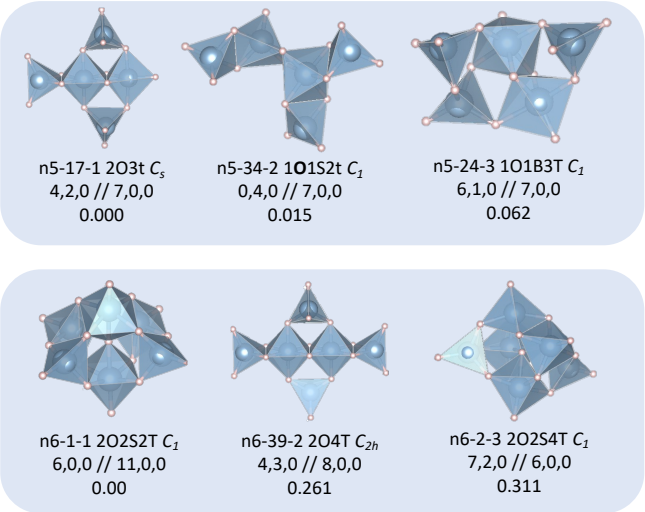

Figure **30**. *Ball-and-stick* models of the tentative lowest PBEsol energy $(AlH_3)_n$ PBEsol LM for sizes $n = 8$ to 11. Blue (grey) spheres represent aluminium (hydride) atoms, and transparent blue polyhedral are AlH_m SBUs, with $m = 4$ to 6. Each configuration is labelled "*n*X-Y-Z S v,e,f // g,h,i,j P (Q)", where X is its size, Y is its IP rank, Z is its PBEsol rank, S is a character string describing the SBUs that the cluster is composed of, v (e, f) is the number of vertices (edges, faces) shared between the SBUs, g (h, i, j) is the number of F atoms with a coordination of 1 (2,

3, 4), and P (Q) is the PBEsol (IP) energy difference between the cluster shown and the ground state cluster of the same size. The character string S is a concatenation of character strings with the format mU that indicate that are m SBUs of type U, where U is the first letter of the name of SBU shape (tetragonal, square-base-pyramidal, bipyramidal, octahedral), and when in bold lowercase, lowercase, uppercase or bold uppercase the SBU contains at least three, two, one and zero singly coordinated F atoms, respectively. Coordination numbers (and SBUs) are calculated using an interatomic Al-H cutoff distance of 2.2 Å.

In our comprehensive examination of the $(AlH_3)_5$ cluster, we encountered significant deviations from the configurations typically observed in analogous AlF_3, or other nanoclusters which are shown in figure 30. This unexpected divergence prompted a thorough re-evaluation of the stability and structural integrity of the $(AlF_3)_5$ derived data for the GM5 and the LM5b. Our initial analysis raised concerns regarding the potential instability of the GM5, attributed to the presence of tetrahedral SBUs with two undercoordinated F atoms in each tetrahedral SBU. In general, such configurations indicated the less densely packed configuration which gives the tendency towards potentially less stable structures. To rigorously test and validate our structural hypotheses, we resorted to employing the GA approach for the global optimisation of the $(AlH_3)_5$ PES. This involved adapting the IP from the $(AlF_3)_5$ model to suit the specificities of the $(AlH_3)_5$ cluster. The GA was meticulously executed 3 times with a set population of 50 individuals, spanning over 100 generations, in an effort to comprehensively map the structural landscape of the $(AlH_3)_5$ cluster. The outcomes of this extensive computational endeavour were illuminating, revealing a total of 106 unique $(AlH_3)_5$ structures. This was in clear contrast to the 233 structural variations previously identified within the $(AlF_3)_5$ dataset. The observed reduction in structural diversity can be primarily attributed to the intrinsic differences in the ionic radii of H and F ions, with the larger H ion imposing constraints on the possible spatial configurations, thereby limiting the structural variability in clusters incorporating H atom.

Further insights were gained through the subsequent refinement of these structures using the PBEsol functional. This phase of the analysis demonstrated that the configurations identified through the IP-GA process bore a striking resemblance to those predicted from the $(AlF_3)_5$ dataset. This remarkable correlation not only affirmed the accuracy and reliability of our predictive model but also reinforced the observed trends in structural stability and integrity across different nanoclusters. Intriguingly, both the configurations derived from the GA global optimisation and those originally datamined from the $(AlF_3)_5$ dataset underwent significant transformations during the optimisation process on the PBEsol $(AlH_3)_5$ PES, underscoring the unique and complex nature of clusters that incorporate H atom.

A common occurrence in the domain of stable nanoclusters is the emergence of SBUs with a high coordination number as the cluster size grows, approaching the coordination number observed in bulk phases. This phenomenon is crucial for optimizing coordination numbers and reducing steric hindrance resulting from electrostatic repulsion. For example, this principle dictates that configurations dominated by square base pyramidal or octahedral SBUs are more commonly encountered for bigger size of AlF_3 nanoclusters, as opposed to those featuring a tetrahedral SBUs as the main. Interestingly, the GM5 configuration notably lacks any 5-coordinated Al centre SBUs, only displaying tetrahedral and octahedral coordination exclusively. This configuration reflects the constraints imposed by the smaller ionic radius of the Al atom, which struggles to accommodate the relatively larger H atom as efficiently as it does the F atom, necessitating a shift towards more open and expansive structural configurations. This specific structure is also marked by the unusual inclusion of an 'arm' tetrahedral SBU to one side, an attribute rarely seen in GM structures. Despite the more open nature of this configuration compared to its $(AlF_3)_5$ counterpart, the enhanced polarizability of the H atom facilitates a more optimal distribution of electron density, thereby contributing to the overall stability of the configuration.

The robustness of the LM5b also came under examination, notably for its extensive structure defined by the 'arm' SBUs it features. This structure unfolds into two distinct branches, each terminating in 'long arm' SBUs, and culminates in

three tetrahedral SBUs at the tips, each characterised by two undercoordinated F atoms. In contrast, the LM5c unveils a more intricate and multidimensional structural configuration, distinguished by a diminished count of tetrahedral SBUs, each hosting fewer undercoordinated fluorine atoms. Both the LM5b and LM5c configurations have lack of the high degree of symmetry, C_1. From analysing the energy levels and arrangements of the LM5b and LM5c, it becomes evident that AlH_3 experiences considerable spatial strain when situated in compact structures. Therefore, it favours more open or elongated configurations, akin to the LM5b, as opposed to the more contracted form of the LM5c.

The GM6 has the same atomic connectivity of $(AlF_3)_6$ GM and is also, the smallest cluster size that has the corner-sharing octahedral SBUs with an average coordination number of Al atom of 5 with C_{2v} point-group. Likewise, corner sharing octahedral SBU is the only type of SBU that can be found in the $(AlH_3)_6$ bulk phase[124]. The GM6 composed of two octahedral SBUs which are corner shared with two tetrahedral SBUs and two square pyramidal SBUs which is same as the $(AlF_3)_6$.

The angle of the two octahedral SBU (Al-F-Al) is 163.49°, which is 3.63° smaller than the $(AlF_3)_6$ GM, towards the window where no SBU surrounds the centre. The angle minimises the steric hindrance from the SBU surrounding belts; tetrahedral and square base pyramidal SBU. Each SBU has singly coordinated F atom termination. From the view shown in figure 30 two square-base pyramidal are in the behind of the octahedral SBUs. The square-base pyramidal are not only corner shared with both octahedral SBU and to each other. The LM6b has also two octahedral SBU, but they are edge-shared to each other. On the opposite side of the octahedral SBU the two square pyramidal SBUs are edge-sharing. The rest of the tetrahedral SBU are corner sharing with octahedral SBUs and one square base pyramidal SBU.

The LM6b configuration is composed of the several octahedral SBUs and the configuration is similar to the GM5. The $n = 6$ only has an additional tetrahedral SBU ‘arm’ on the symmetric position of $n = 5$ ‘arm’. The octahedral SBUs are linked through shared edges and even with the tetrahedral SBUs on each, which coalesce into a continuous network. It is further augmented by the inclusion of an

adjacent tetrahedral SBU that resides within the same geometric plane, thus contributing to the multi-dimensional complexity of the structure. A noteworthy feature of this assembly is the absence of F atom terminations on these octahedral units. The two tetrahedral SBUs are corner shared with the octahedral SBUs with a F atom each. Each of these is connected through corner-sharing connectivity with the octahedral units, creating a lattice of interconnections that enhances the structural integrity of the cluster. The disposition of the two terminal F atoms in each of tetrahedral SBU is particularly striking, as they project in antipodal directions. This deliberate spatial configuration is likely a strategic adaptation evolved to minimise the electrostatic repulsion that would be more pronounced if the negatively charged F atoms were in closer proximity.

The intrigue surrounding this configuration is deepened by the fact that the initial IP-LM5b placed this structure at the 39th position in terms of IP energy ranking after the AlF_3 clusters are datamined using the modified IP. This placement suggests a tendency of the IP towards favouring configurations that are more tightly bound and exhibit a dense network of atomic coordination. Despite such a predictive preference, it is notable that the average coordination number of the Al atoms within this LM5b structure is actually lower than that of the GM, registering at a value of 4.83.

The symmetry level within this LM5b structure is appreciably high, manifesting through the presence of a dihedral rotational axis (C_2) along with a horizontal reflective plane (h), both of which are indicative of a harmonious and balanced structural composition. The structure's spatial arrangement is not as compact as that of the GM, which becomes evident upon observing the F atoms that maintain only single connections to their neighbouring atoms, hinting at a more dispersed atomic distribution. This less dense packing, while suggestive of fewer steric hindrances, also implies a different interplay of forces governing the cluster's stability compared to the tightly knit configuration of the GM.

The LM5c structure exhibits multifaceted structure. This configuration is characterised by the presence of two octahedral SBUs that are conjoined by shared edges, with each octahedral SBU hosting an undercoordinated F atom. Complementing this octahedral framework are two tetrahedral SBUs that are

strategically positioned in opposition to each other, providing a balanced geometric form. These tetrahedral units are integrally connected to the octahedral SBUs, joining at the vertices to add to the cohesiveness of the structure.

The complexity of the structure is further escalated by the inclusion of square-base pyramidal SBUs. These units not only establish side-to-side connectivity with each other but also form vertex-to-vertex connections with the octahedral and tetrahedral units, weaving a tapestry of intricate bonds that contribute to the overall robustness of the structure. This elaborate network of connections ensures that, on average, each aluminium atom achieves a coordination number of five. This degree of connectivity parallels that observed in the GM structures for clusters of the same size and surpasses that witnessed in structures that are less stable.

Despite the LM5c's constituents being more compactly arranged than those of the LM5b—as evidenced by the increased number of shared edges and interconnections—this tight packing paradoxically leads to augmented electrostatic repulsion which is the same case as the $(AlF_3)_5$ 3rd LM. The atoms, being in closer quarters, repel each other with greater force due to their like charges. These repulsive forces act counter to the structural integrity, thereby reducing the thermodynamic stability of the cluster. While one might presume that closer atomic interactions would inherently confer greater stability due to the enhanced potential for bonding, in this scenario, the opposite is true. The close proximity of the atoms, although potentially beneficial for increasing coordination number, actually detracts from the cluster's stability by amplifying the repulsive electrostatic forces to a level that outweighs the benefits of increased interactions. Thus, while the structure may appear tightly knit and orderly, the resultant increase in repulsive forces means that the LM6c is less stable thermodynamically when compared to the GM6 and the LM6b, where the atoms have more space to mitigate such repulsions.

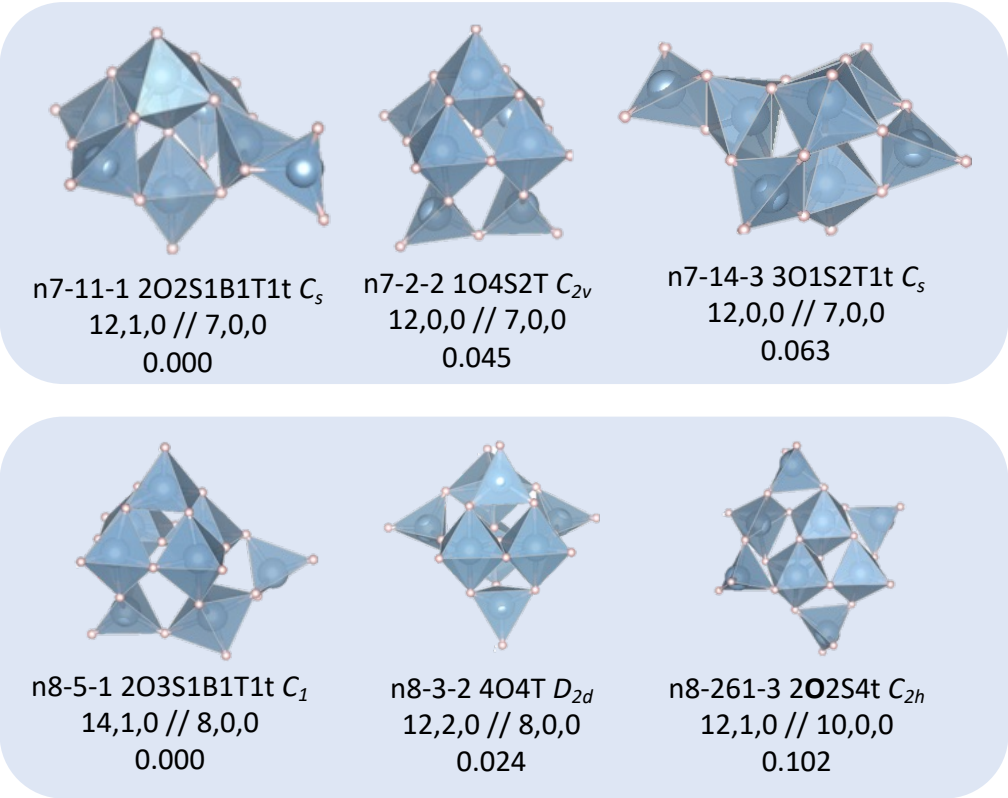

Figure **31**. *Ball and stick* models of the tentative lowest PBEsol energy $(AlH_3)_n$ PBEsol LM for sizes n = 5 and 6. Blue (grey) spheres represent aluminium (fluoride) atoms, and transparent blue polyhedral are AlH_m SBUs, with m = 7 and 8. Each configuration is labelled "*n*X-Y-Z S v,e,f // g,h,i,j P (Q)", where X is its size, Y is its IP rank, Z is its PBEsol rank, S is a character string describing the SBUs that the cluster is composed of, v (e, f) is the number of vertices (edges, faces) shared between the SBUs, g (h, i, j) is the number of F atoms with a coordination of 1 (2, 3, 4), and P (Q) is the PBEsol (IP) energy difference between the cluster shown and the ground state cluster of the same size. The character string S is a concatenation of character strings with the format *m*U that indicate that are m SBUs of type U, where U is the first letter of the name of SBU shape (tetragonal, square-base-pyramidal, bipyramidal, octahedral), and when in bold lowercase, lowercase, uppercase or bold uppercase the SBU contains at least three, two, one and zero singly coordinated F atoms, respectively. Coordination numbers (and SBUs) are calculated using an interatomic Al-H cutoff distance of 2.2 Å.

The GM configuration previously determined for $(AlF_3)_7$ underwent a structural transformation when reassessed, resulting in a different arrangement. This reconfigured $(AlF_3)_7$ GM shares structural similarities with that of the n = 5

cluster size, indicating a recurring motif. The GM7 exhibits the same underlying structure as the GM6 but with an additional feature: an 'arm' tetrahedral SBU extending from the central belt of tetrahedral SBUs. In both the n = 5 and n = 7 clusters, there seems to be a consistent structural inclination towards retaining the octahedral SBUs, despite the necessity to integrate tetrahedral 'arm' SBUs accompanied by two undercoordinated F atoms. The n = 6 cluster retains the GM6 structure, where the framework is defined by the cornerstone of the octahedral SBUs engaging in corner-sharing linkages. Observing the structural tendencies across the n = 5 – 7 cluster sizes, we can formulate a hypothesis that $(AlH_3)_n$ clusters exhibit a strong preference for incorporating octahedral SBUs into their ground state configurations, even if it results in tetrahedral SBUs with terminations of two undercoordinated F atoms. Given this observed pattern, it is reasonable to assume that as we examine larger $(AlH_3)_n$ clusters, we will probably encounter a higher frequency of GM configurations that include tetrahedral 'arm' SBUs. This structural preference suggests an inherent stability or favourable energy associated with reducing the relatively larger steric hindrance that caused from the larger ionic radius of H compared to F, which could be a driving factor in the formation of these 'arm' extensions as the cluster size increases.

The GM7 features a single plane of symmetry extending from the peak of the square-base pyramidal structure through the bridging F atom, connecting two octahedral SBUs to the lower tetrahedral 'arm' SBU. The Al-H-Al bridging angle within the octahedral SBU is 2.46° wider compared to the GM6, measuring 165.95°, as the tetrahedral 'arm' SBU enhances electrostatic repulsion at the second nearest neighbour position, contributing to the widening of the angle. Furthermore, the bridge angle in the $(AlF_3)_6$ octahedral formation exceeds that of the $(AlH_3)_7$ by 1.17°, signifying a distinct structural characteristic.

In the LM7b cluster, the distance from the central F atom to the nearest Al atoms is notably shorter compared to the same in $(AlF_3)_7$. Specifically, the four surrounding Al atoms are positioned at a distance of 2.27 Å, which is 0.14 Å closer than in $(AlF_3)_7$. Moreover, the distance to the Al atom directly above the central F atom stands at 2.07 Å, making it 0.37 Å shorter than in the $(AlF_3)_7$ cluster. The angle between Al-H-Al around the central F atom, at 120.51 °, is slightly tighter

by 1.78 ° compared to that in $(AlF_3)_7$. This reduction in angle can be attributed to the interplay of electrostatic forces, where the repulsion between H atoms and the attraction within Al-H bonds play pivotal roles. Despite hydrogen having a larger ionic radius than fluorine, the electrostatic repulsion between F atoms in the $(AlF_3)_7$ cluster is stronger than between H atoms in $(AlH_3)_7$. This results in the F atoms being positioned further from the central F atom than the H atoms are from the central H atom in their respective clusters. This is because the strength of electrostatic repulsion is not determined solely by ionic size, but also by the electron density and electronegativity of the atoms involved. F atoms, being highly electronegative, attract more electron density towards themselves. Consequently, when two F atoms are in proximity, the electron clouds around them repel each other more vigorously than the less dense clouds around H atoms. This greater electron density around F leads to stronger repulsive forces, despite its smaller ionic size compared to hydrogen.

The LM7b configuration arises from the GM6 structure, augmented with a tetrahedral 'arm' SBU. In this instance, the tetrahedral 'arm' SBU adjoins a square-base pyramidal unit, which itself is connected via edge-sharing bonds on the side opposite to the open segment of the SBU ring. This ring envelops the corner-sharing octahedral SBUs, thereby integrating the tetrahedral 'arm' SBU into the broader structural framework.

The LM7c, which is an intricate part of the overall structure, is derived from the GM6 configuration, and it is further complexed by the addition of a tetrahedral 'arm' SBU. This particular 'arm' SBU is integrally linked to a square-base pyramidal structure, establishing its connection through an edge-sharing mechanism that is positioned diametrically opposite to the void in the SBU ring. This ring of SBUs effectively encompasses the corner-sharing octahedral 'sticks', with the Al-H-Al bond angles within the octahedral corner sharing being notably less, specifically by 2.99° and 4.16° when juxtaposed with the GM7 and the $(AlF_3)_6$ GMs respectively. The deviation in the octahedral corner sharing 'stick' angle is a direct consequence of the augmented electrostatic repulsion emanating from the tetrahedral 'arm' SBU, which is located in a direction antithetical to the open side of the SBU ring.

The LM7c has the same symmetricity as the GM, C_s. Compared to the GM7 and LM6b of $(AlH_3)_6$ is caused from the inefficient electron density distribution. The GM's 'arm' is edge shared with tetrahedral SBU which the F atoms can more effectively share the Al's electrons as they have one less F atom around from the nearest Al atoms where edge shared with the 'arm' compared to the LM7c. The SBU where the 'arm' edge shared with is corner sharing with the two octahedral SBUs. However, the SBU on LM7c where the 'arm' edge shared is not only corner shared with the octahedral SBUs but also one tetrahedral and square base pyramidal SBU. Thus, the less stableness of LM7c compared to the G7M is from the 'arm' F atoms insufficiently attract the electrons from Al atoms as they have to compete more F atoms which are in vicinity to attract electrons from Al atoms. The same analysis also can be applied when it compared with the LM7b, but also the LM7b symmetricity is higher than the LM7c which expands the analysis that the insufficient electron distribution on the LM7c penalise more having the high electrostatic repulsion from the short interatomic distance in LM7b.

Within the GM8 configuration, there exists an 'arm' tetrahedral SBU that is also a part of the LM7b structure. This 'arm' SBU forms a corner-sharing bond with both the bottom square-based pyramidal structure and an octahedral SBU, the latter of which transitions into another octahedral SBU. Intriguingly, so far, the GM configurations of $\underline{n} = 5$, $n = 7$, and $n = 8$ all include a tetrahedral 'arm' SBU, leading to the anticipation that such 'arm' SBUs may be a recurring feature in larger $(AlH_3)_n$ clusters. This pattern is observed in the LM8b of $(AlF_3)_7$. The inclusion of the 'arm' SBU disrupts the symmetry previously seen in the LM7b, resulting in Al-H-Al bond angles within the central layer of the octahedral SBU altering to 116.53°, 107.99°, 118.55°, and 129.54° respectively, as measured in a counter clockwise direction starting from the RHS.

The spatial arrangement dictates that the central F atom is at varying distances from the Al atoms of the four octahedral SBUs, specifically at 1.82 Å, 2.04 Å, 2.17 Å, and 2.34 Å, again measured counter clockwise from the octahedral SBU that shares a corner with the 'arm' SBU. The proximity of the central F atom is closest to the octahedral SBU that it shares a corner with, due to this particular F atom being the second nearest neighbour to the Al atom within the 'arm' SBU,

allowing for a more efficient sharing of electrons over this shorter distance. Consequently, this central F atom is preferentially positioned to attract electrons from the 'arm' SBU.

Furthermore, the Al-H-Al bond angle on the RHS is enlarged when compared to that behind the middle layer of octahedral SBUs, owing to the proximity of the corner-shared F atom on the lower SBU, which at a distance of 2.22 Å induces electrostatic repulsion that influences the H atom to move towards the centre. The angle on the rear side of the middle SBU layer is more acute as the F atom leans closer to one SBU, causing the adjacent Al atom of the octahedral SBUs to draw nearer, in an attempt to attract more electrons from the central F atom. Conversely, the LHS and the front side of the Al-H-Al bond exhibit a wider angle to compensate for the closer proximity between the central F atom and the Al atom at the rear. This compensatory widening results in the LHS and front Al atom being 0.07 Å and 0.3 Å further from the central F atom, respectively.

LM8b has D_{2d} symmetricity which the structure is has two group of layers of SBU and they are staggered. Each layer is consisting of two octahedral and two tetrahedral SBUs. The layer is identical to the GM5 without the ‘arm’ SBU on the LHS. The two layers are connected through corner sharing in between the octahedral and tetrahedral SBUs to each other. Thus, tetrahedral SBUs are bent towards the opponent layer’s octahedral. The two octahedral SBUs are edge shared to each other, but they are distorted which the corner shared F atoms are angled in towards the opponent layers’ tetrahedral with 79.95 ° (angle of corner shared F atom – Al atom – edge shared F atom). The LM8b has higher symmetricity than the GM, but as it has two more tetrahedral SBUs and one more of edge sharing connectivity between the octahedral SBUs which increase the electrostatic repulsion. Thus, the high symmetricity compensates less than the electrostatic repulsion which provides less thermodynamical stability compared to the GM8.

Until now the top 3 LM for $(AlH_3)_n$ $n = 1 – 7$ was found in less than 39th rank of $(AlF_3)_n$ in IP. However, the LM8c was found from the 261st rank of $(AlF_3)_8$ rank. The configuration has wide open feature which all the SBUs can be projected on a plane without overlapping the centre of the atomic positions. The

configuration has the two octahedral SBUs which has the edge sharing connectivity to each other and there is no undercoordinated F atoms. From the axis of the direction of the alignment of the octahedral SBU the configuration has the improper rotational symmetry. On each side there is a square base pyramidal which is corner shared the F atoms in the square base with the two octahedral SBUs and other two F atoms in the square base are corner shared with the two tetrahedral SBUs. The two tetrahedral SBUs which located on the side of the square base pyramidal SBU is also corner shared with the tip and the side of the octahedral SBUs.

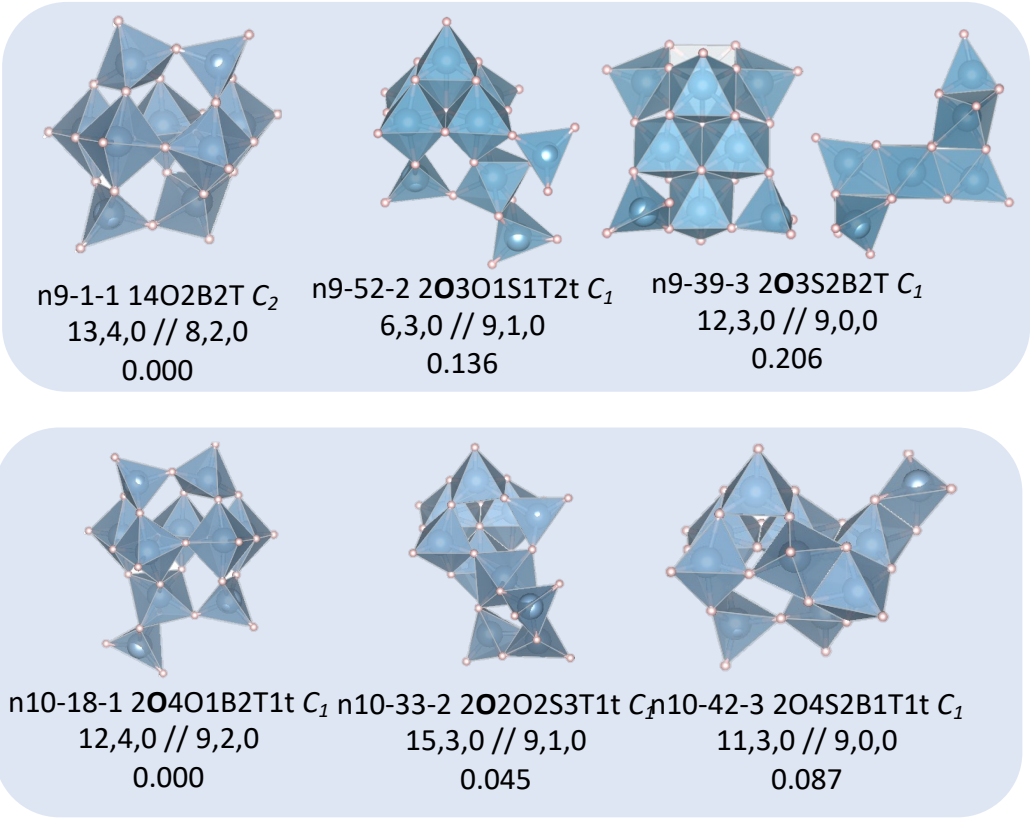

Figure **32**. *Ball and stick* models of the tentative lowest PBEsol energy $(AlH_3)_n$ PBEsol LM for sizes n = 5 and 6. Blue (grey) spheres represent aluminium (fluoride) atoms, and transparent blue polyhedral are AlH_m SBUs, with m = 7 and 8. Each configuration is labelled "*n*X-Y-Z S v,e,f // g,h,i,j P (Q)", where X is its size, Y is its IP rank, Z is its PBEsol rank, S is a character string describing the SBUs that the cluster is composed of, v (e, f) is the number of vertices (edges, faces) shared between the SBUs, g (h, i, j) is the number of F atoms with a coordination of 1 (2, 3, 4), and P (Q) is the PBEsol (IP) energy difference between

the cluster shown and the ground state cluster of the same size. The character string S is a concatenation of character strings with the format *m*U that indicate that are m SBUs of type U, where U is the first letter of the name of SBU shape (tetragonal, square-base-pyramidal, bipyramidal, octahedral), and when in bold lowercase, lowercase, uppercase or bold uppercase the SBU contains at least three, two, one and zero singly coordinated F atoms, respectively. Coordination numbers (and SBUs) are calculated using an interatomic Al-H cutoff distance of 2.2 Å.

The GM9 for AlH_3 and AlF_3 has the same atomic connectivity with C_2 space group which the two-fold rotational axis is going through the F atom in the centre of the figure 32; the axis pierce through the image. The configuration forms the five octahedral SBU along the middle plane which one of the centre octahedral SBU's F atoms are fully coordinated through edge sharing with the other octahedrals and corner sharing with the upper/bottom of the trigonal bipyramidal SBUs. The lateral two octahedral SBUs and the trigonal bipyramidal corner shared with the tetrahedral SBU. The octahedral SBUs are closely packed as the two F atoms located in the inwards of the octahedral SBUs has 3-fold coordination number. The 3-fold coordinated F atom is closer to the fully coordinated centre octahedral SBU with 1.78 Å than other two octahedral SBUs; 1.88 Å, 1.90 Å in (counter)clockwise. The centre octahedral SBU forms a triangle which the vertexes are the F atoms. The two corners of the 3-fold coordinated F atoms has 3.19° larger angle as the double coordinate F atom is pushed out from the electrostatic repulsion from the 3-fold coordinated F atoms.

The LM9b configuration emerges as an extension of the GM structures for $n = 7$ or $n = 8$ AlH_3, where an additional tetrahedral SBU is edge-shared with the base of the $n = 8$ AlH_3 GM's trigonal bipyramidal SBU. This particular arrangement doesn't fully replicate the $n = 8$ AlH_3 GM structure but rather resembles its mirror-imaged counterpart. In the GM8 configuration, the central F atom's proximity to its nearest Al atom neighbour is noteworthy, especially in the context of the octahedrally coordinated 'arm' tetrahedral SBU, which is located at the shortest distance. This configuration indicates a competitive interaction where the F atoms vie to attract electrons from the adjacent Al atoms. A similar

phenomenon is observed in another octahedral coordination, where an 'arm' tetrahedral SBU is also edge shared. Notably, the distances between the central Al atom and the F atoms, which are in turn shared with the 'arm', are shorter compared to other interatomic distances. This suggests that the F atoms, having fewer sources of electron supply from neighbouring Al atoms, exhibit stronger electron-drawing interactions.

The LM9c is composed of four octahedral SBUs which the side view of the configuration shows that the octahedral SBUs are on a same plane, and they are edge shared to each other. The two octahedral SBUs in the middle shared three edges with others which forms the two of the 3-fold coordinated F atoms. The 3-fold coordinated F atoms has sufficient electrons, thus the distance between the F atom to the nearest Al atoms, thus they have significantly longer F-Al interatomic distances: 1.82 Å to the Al atoms in the middle layer of octahedral SBUs and 2.03 Å to the Al atom in the corner octahedral SBUs. The 0.21 Å longer distance to the Al atom in the corner octahedral SBU as the influence of the electrostatic repulsion from the four Al atoms in the octahedral and trigonal bipyramidal SBUs which pushes out the Al atom. The tetragonal SBU breaks the mirror plane symmetry which the plane goes through the middle from top to bottom. One of the tetragonal can swivel to have the mirror plane configuration, but which the configuration is in a higher energy as it suffers from the electrostatic repulsive forces.

The GM10 has the mirror plane of the GM9 as the main configuration, but with a tetrahedral SBU ‘arm’ on the trigonal bipyramidal SBU of the GM9 which forms the octahedral SBU. The LM9b and LM9c as the higher the energy configuration structure tend to be less densely packed even with the edge sharing octahedral SBUs between each other. However, more than half of the SBU which compose the configuration is octahedral SBUs. Until the $n = 10$, the cluster does not have enough electrostatic field within the system, but we could observe that the structure started to form highly coordinated SBUs. Thus, with this trend in structural property we expect to find nanocrystal structures within a few bigger cluster sizes.

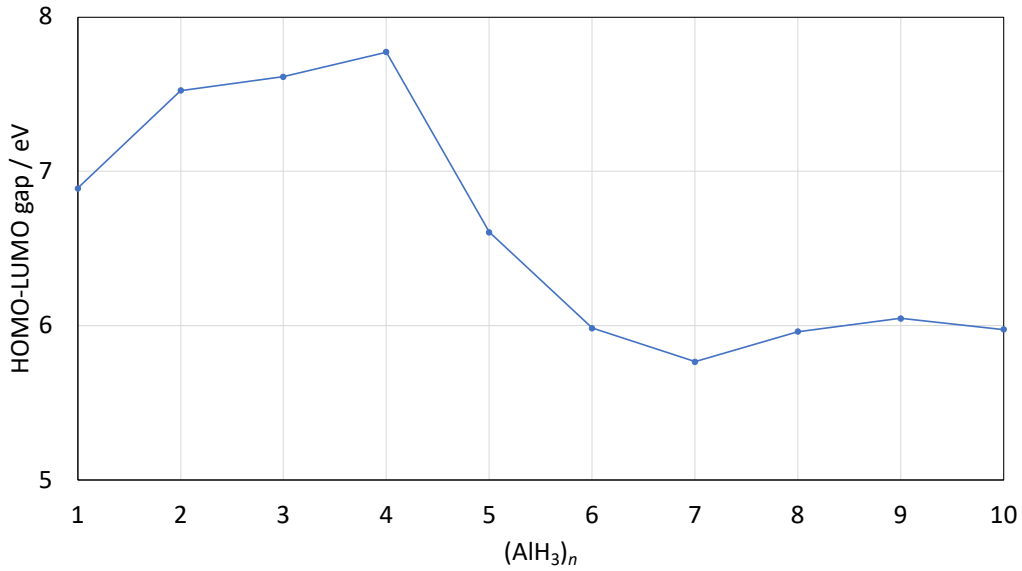

Figure **33**. The HOMO-LUMO gap of $(AlH_3)_n$ (n = 1 – 10) GM.

Figure 33 illustrates the HOMO-LUMO gap of $(AlH_3)_n$ (n = 1 – 10) GM clusters, calculated using PBEsol0 on clusters optimized with PBEsol. The HOMO-LUMO gap is a critical parameter in determining the electronic properties and stability of molecular clusters, with larger gaps typically indicating greater stability and lower reactivity.

The data indicates that the HOMO-LUMO gap gradually increases from GM1 to GM4, a trend that can be attributed to the high symmetry inherent in these smaller clusters. High symmetry in molecular clusters often leads to enhanced stability and larger HOMO-LUMO gaps. This is primarily due to the delocalization of molecular orbitals, which allows electrons to be spread more evenly across the entire structure, thereby minimizing regions of high electronic density that contribute to electronic repulsion. In highly symmetrical configurations, the uniform distribution of electron density reduces the likelihood of localized electron-electron interactions, resulting in a more stable electronic structure with a wider energy gap between HOMO and the LUMO.

However, a notable deviation occurs at GM5, where the HOMO-LUMO gap experiences a sharp decrease of more than 1 eV. This significant reduction is likely due to a structural transition from highly symmetrical configurations to more

irregular, random-like arrangements as the cluster size increases. Such a transition disrupts the effective overlap of molecular orbitals, leading to increased electronic repulsion. In less symmetrical configurations, the delocalization of orbitals becomes less efficient, causing electrons to become more localized in specific regions. This localization enhances electron-electron interactions, which in turn lowers the energy required to excite an electron from the HOMO to the LUMO, thereby reducing the HOMO-LUMO gap.

This transition at GM5 signifies a fundamental change in both the electronic structure and the stability of the clusters. The reduction in symmetry and the resulting increase in electronic repulsion suggest that the clusters are moving towards configurations where electronic interactions become more complex and less favorable for maintaining a large energy gap. Consequently, these larger clusters may exhibit different chemical reactivity and stability characteristics compared to their smaller, more symmetrical counterparts. The observed decrease in the HOMO-LUMO gap at GM5 highlights the delicate balance between structural symmetry and electronic properties, underscoring the importance of geometric arrangement in determining the electronic behavior of nanoclusters.

Furthermore, this structural evolution reflects the inherent complexity that arises as clusters grow in size. While smaller clusters benefit from high symmetry and orbital delocalization, larger clusters may encounter geometric constraints that prevent the maintenance of such symmetry, leading to increased electronic repulsion and reduced stability. Understanding these trends is crucial for predicting the properties of nanoclusters and designing materials with desired electronic characteristics. Future studies could explore the precise geometric changes and their direct impact on orbital interactions to gain deeper insights into the relationship between cluster size, symmetry, and electronic structure.

The HOMO-LUMO gap continues to decrease until GM7, after which it appears to level off at around 6 eV between GM8 and GM10. This leveling off suggests that the clusters have reached a new structural regime where the size and configuration variations have a reduced impact on the electronic properties. The stabilization at around 6 eV indicates that the larger clusters, despite their

increased size and potential for more complex configurations, achieve a balance that maintains a relatively consistent HOMO-LUMO gap.

This pattern of variation in the HOMO-LUMO gap reflects the complex interplay between cluster size, symmetry, and electronic structure. The initial increase in the gap with cluster size, followed by a sharp decrease and eventual stabilization, highlights the importance of symmetry and structural configuration in determining the electronic properties of molecular clusters. Understanding these relationships is crucial for the design and application of nanomaterials with specific electronic properties.

5. Thiolated gold cluster, $(Au)_{25}(Cys)_{18}$

5.1. Introduction

The area involving thiol-capped gold nanoparticles, which includes both nanocrystals and nanoclusters, has seen growing interest within the nanoscience community in recent years[127–133]. Their significant importance in basic scientific research and various technological fields is undeniable. They are used in many areas such as catalysis, enhancing optical properties, medicinal chemistry, and developing advanced chemical sensors. When the size of these gold nanoparticles approaches the de Broglie wavelength[134] of their conduction electrons, which is approximately 1 nm in diameter, a significant change occurs[135–138]. At this stage, there is a transition from the nearly continuous electronic bands observed in bulk gold or larger nanoparticles, typically exceeding 5 nm, to a spectrum of discrete energy levels[133,138–140]. As the size of these gold particles decreases below 3 nm, they begin to deviate from their bulk metallic electronic properties. This is evident from their loss of the ability to sustain plasmon resonance—a characteristic feature of larger gold nanocrystals ranging from 3 to 100 nm in size[133]. Moreover, the way atoms are arranged within these minuscule metallic nanoparticles, also referred to as clusters (similar to AlF_3 and AlH_3 systems considered in the previous chapters 3 and 4), critically affects the relationship between their structural form and their resultant properties, which also exhibit dependence on the particle size[141,142].

Despite the significant strides achieved in the past ten years regarding the synthesis of thiol-protected gold clusters, the research community still faces challenges. These include the definitive identification of cluster size—that is, the precise number of gold atoms constituting a cluster—and elucidating the specific arrangements of atoms within these clusters[143]. X-ray crystallography stands as the gold standard for determining the exact size and atomic arrangement of clusters; however, growing single crystals suitable for such analysis has been exceptionally challenging. To date, the crystal structure of only one such gold-thiolate cluster, specifically Au_{102}(p-mercaptobenzoic acid)$_{44}$, has been successfully elucidated. This cluster is characterised by a 49-atom decahedral core, which can also be considered as a pentagonal twinning of either a face-centred

cubic or a hexagonal close-packed core structure. Delving into the exploration of the distinctive structural, physical attributes of Au_{25} clusters, particularly those comprising several dozen atoms, remains a promising yet underexplored frontier in the field.

5.2. Au_{25} cluster

This chapter describes the computational strategies employed to predict the structure of the Au_{25} cluster, using the second nearest neighbour modified embedded atom method (2NN-MEAM[46,48,57]) interatomic potentials and the MCDQ method for global optimization, followed by refinement using the PBEsol[72].

To investigate the complex landscape of the Au_{25} cluster's potential energy surface, a robust and efficient global optimization method is indispensable. The MCDQ method does this by combining the random searching abilities of the Monte Carlo technique with the quick settling down of deterministic quenching. The Monte Carlo component of MCDQ allows for a stochastic search of the configuration space, providing the means to overcome LM barriers and explore a wider range of structural motifs. This is coupled with a MCDQ process, where the system is systematically cooled to attain a local minimum, ensuring the thorough investigation of the PES and increasing the likelihood of identifying the global minimum.

Applying the 2NN-MEAM potential within the MCDQ framework allowed us to generate a vast array of potential Au_{25} cluster configurations. This comprehensive dataset not only provided an insight into the possible structural diversity of the cluster but also laid the groundwork for subsequent refinement stages. From this extensive pool, the top 1000 configurations, ranked according to their energies computed using the 2NN-MEAM potential, were selected for further analysis. This selection process ensured that the subsequent refinement stage focused on the most promising candidates, optimizing computational resources.

Intriguingly, the PBEsol-GGA refinement process revealed that the structure initially ranked 6th in energy according to the 2NN-MEAM potential emerged as the GM on the PBEsol PES. This finding highlights the critical importance of employing a multi-tiered optimization and refinement strategy. While the 2NN-MEAM potential effectively narrows the search space by identifying a subset of promising configurations, the subsequent refinement with the PBEsol-GGA functional is essential for accurately determining the global minimum.

Figure 34 shows a detailed bipartite plot that thoroughly traces the energy ranking evolution of Au_{25} LM on transition from the globally optimized 2NN-MEAM dataset to the refined PBEsol dataset. This graphical representation presents the relationships between each LM across the two distinct energy landscapes, illustrating a one-to-one correspondence from the left to the right of the diagram. The spectrum of colours adopted, resembling a rainbow's progression, is deliberately chosen to convey the origin of each refined PBEsol LM's rank within the 2nn-MEAM framework. In this complex mapping, only the most stable, *i.e.*, the lowest energy 2NN-MEAM LMs are linked to their PBEsol counterparts, with superfluous connections from higher-energy 2nn-MEAM LMs to the same PBEsol LM being systematically excluded. This selection ensures clarity in visualizing the ranking transitions, particularly emphasizing the energy refinement process's impact on the most stable LM, thereby characterising the stability hierarchy within the dataset.

As the plot only shows the 1:1 matching of the PBEsol optimised structure from the lowest 2nn-MEAM LM possible, removed duplicates we anticipated to observe the clear rainbow gradation from the lower ranking to the higher ranking. However, the bipartite plot shows the tilted rainbow gradations which the blue-tone colours are in left top triangle and the red-tone colours are in right bottom triangle. The colour trend indicates that the less stable PBEsol LM are found in the lower ranking of 2nn-MEAM LM.

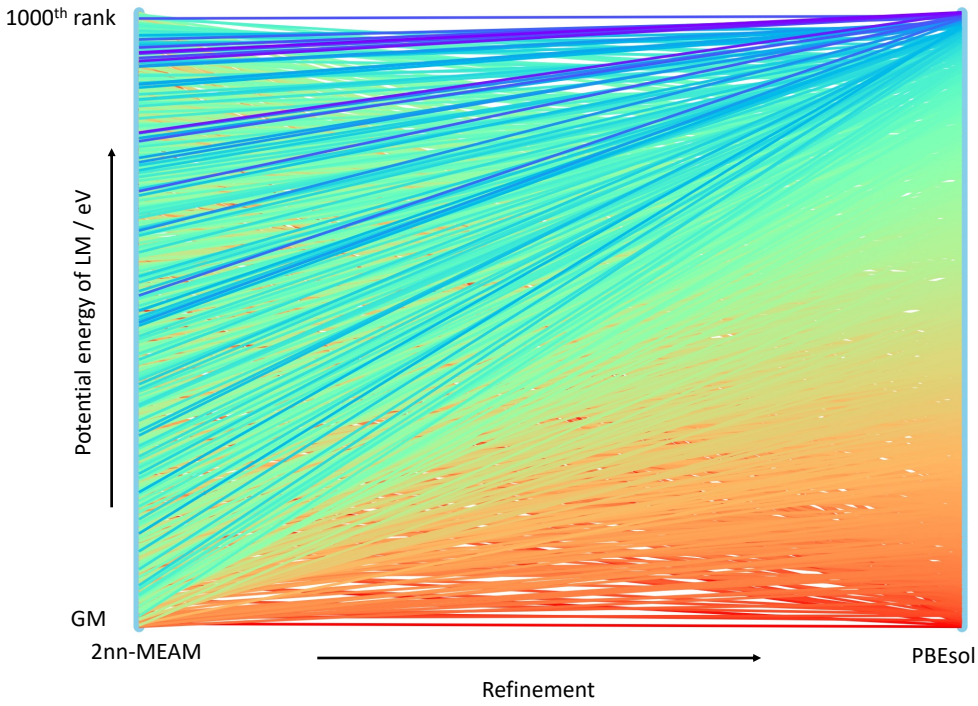

Figure **34**. Ranking transition of the 1000 most stable Au_{25} LM from the 2nn-MEAM PES (left) to the PBEsol Functional PES (right): The bipartite plot illustrates the shift in energy rankings between the two surfaces. Each line connects the initial rank of a LM within the 2nn-MEAM PES to its subsequent rank in the PBEsol PES, employing a gradated rainbow colour scheme to visually trace the ranking transition for each motif.

Figure 35 presents a more detailed scatter plot compared to 34 that elucidates the dynamic shifts in energy rankings for LM as they are refined from the initial 2nn-MEAM PES to the PBEsol PES. This graph highlights only unique data points, curated to exclude overlapping instances, which is achieved in the same manner as in the last figure by associating each PBEsol LM exclusively with its lowest-ranked 2nn-MEAM counterpart, thereby eliminating redundancy and higher-energy parallels from the analysis. Such a careful selection process emphasizes the most stable configurations, with a lower ranking directly correlating to a LM possessing lower potential energy. The rank of '1' is designated for the GM, representing the peak of stability within this energy landscape.

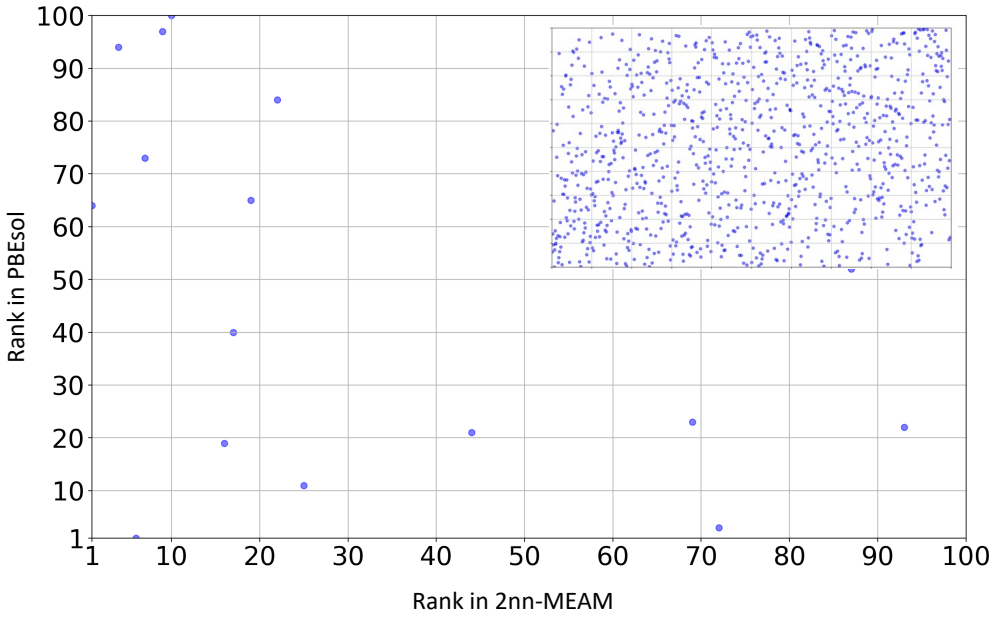

Figure **35**. The progression of energy rankings for Au_{25} LM as they transition from the 2nn-MEAM optimization to the refined PBEsol potential. Displayed in the inset, the plot covers the entire spectrum from the GM up to the 1000th rank, mapping each LM's relative stability shift. A point located at (1, y) shows the LM which is the 2nn-MEAM ground state to PBEsol, while a point at (x, 1) corresponds to a 2nn-MEAM LM that emerges as the new PBEsol LM.

The zoomed-out plot of the main panel is in the smaller inset panel, which shows the LMs spanning from the 1st to the 1000th energy rankings. This wider perspective reveals the intrinsic variability of the dataset which means that the 2nn-MEAM PES cannot describe the PBEsol PES accurately. Luckily, within the crucible of these top 1000 LMs, the LM that was originally at the 6th position in the 2nn-MEAM PES hierarchy ascends to become the GM in the PBEsol framework. However, the atomic connectivity is not retained after the refinement on PBEsol PES as shown in figure 36. Only the 5.2% of the 2nn-MEAM configurations retain their configuration after the PBEsol refinement which is seen by their NAUTY hash keys [89,94]; as was introduced for AlF_3 in the previous chapter to filter out the duplicated configurations.

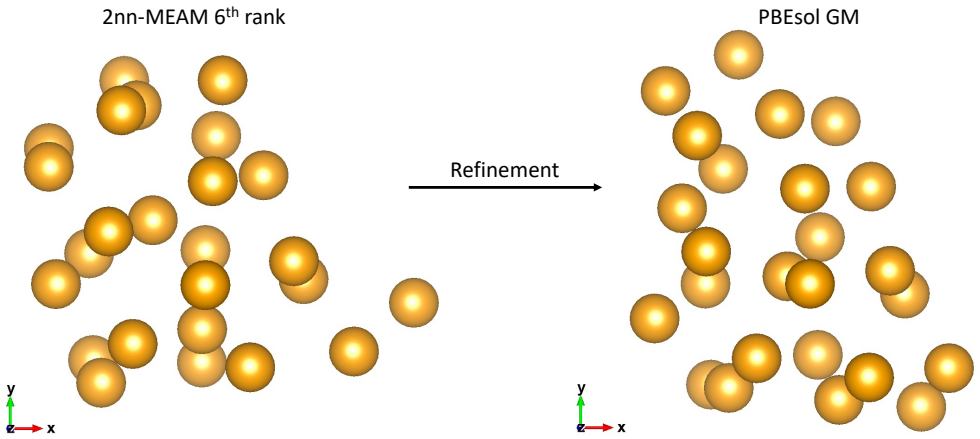

Figure **36**. Atomic arrangement of the 6th ranked 2nn-MEAM local minimum (left) alongside the PBEsol global minimum (right).

The challenges presented by the use of PBEsol PES for accurately describing the PES of the 2nn-MEAM have raised significant doubts concerning the reliability of the data in question from the global optimisation results. To address these uncertainties, a rigorous approach involving data mining was employed to extract the Au_{25} configuration directly from the well-documented structure of $(SR)_{18}(Au)_{25}$[128]. The LJ cluster[144], comprising 25 atoms, was scrutinised for comparison purposes. The LJ clusters are scaled with the 1.0 degree to 3.0 degree in every 0.1 degree before the PBEsol refinement.

At the outset, it was our belief that even if the ligands did not alter the connections between atoms in the Au_{25} cluster, they would at least contribute to a more expand structure. Nonetheless, when striving for a precise characterisation, and due to our inability to predict with certainty how the ligands affect the Au_{25} cluster, we chose not to apply any scaling adjustments to the extracted Au_{25} from the $(SR)_{18}(Au)_{25}$. This meant that the cluster was neither expanded nor shrunk before we commenced the optimisation process using the PBEsol PES.

Following the completion of the optimisation using the 2NN-MEAM approach with PBEsol PES, we took the resulting form of the Au_{25} cluster and conducted a thorough comparison with an extensive set of 1000 IP optimised low-energy configurations. Despite the significant difference in energy rankings across the different PES, the atomic arrangements remained unchanged between the 2nn-

MEAM and PBEsol optimised versions of the $(SR)_{18}(Au)_{25}$'s Au_{25}. Interestingly, within this refined set of structures, the position of the optimised Au_{25} cluster was markedly different when comparing the 2nn-MEAM to the PBEsol PES, securing the 2nd and 908th places respectively.

However, the Au_{25} from the $(SR)_{18}(Au)_{25}$ was not found from the MCDQ global optimisations even though with the extremely large dataset. There is a possibility that the Au_{25} LM within the $(SR)_{18}(Au)_{25}$ may reside in a double funnel or an extremely narrow PES, and perhaps locating the Au_{25} using the MCDQ method for global optimisation inherently presents challenges. The nature of pure metal clusters, such as Au_{25}, is such that they can form a considerably larger number of LMs compared to clusters of ionic compounds containing the same number of atoms. Unlike ionic clusters, which require a specific atom to be consistently positioned between others (like a fluorine atom nestled between aluminium atoms), pure metal clusters are not bound by such repeating bonding patterns, and thus possess a higher degree of freedom.

Therefore, global optimisation for pure metal clusters like Au_{25} necessitate a substantially larger number of sample points to scan the PES sufficiently, ensuring that the landscape is thoroughly probed for potential low-energy structures. Technically, the MCDQ method seems to be not suitable for the purpose for the size of cluster, but at the moment when the global optimisation was performed on KLMC using the 2NN-MEAM potential. However, the version of KLMC used for the study had a bug when the GA runs with 2NN-MEAM. Thus, the calculation let it run with MCDQ while I was focus on other projects.

Despite both potential energy surfaces, 2NN-MEAM and PBEsol, maintaining the atomic connections of the Au_{25} extracted from $(SR)_{18}(Au)_{25}$, the outcome was unexpected with the optimised structures shown in figure 37. Our initial assumption was that the Au_{25} cluster would achieve a higher energy ranking in PBEsol, similar to its second place in the 2nn-MEAM rankings, owing to its high degree of symmetry. Yet, the GM found within the PBEsol framework was of the C_1 point group, which exhibits no symmetry. Nonetheless, upon discovering even lower energy configurations through global optimisation, our concerns about the accuracy of the PES description by the 2nn-MEAM were alleviated.

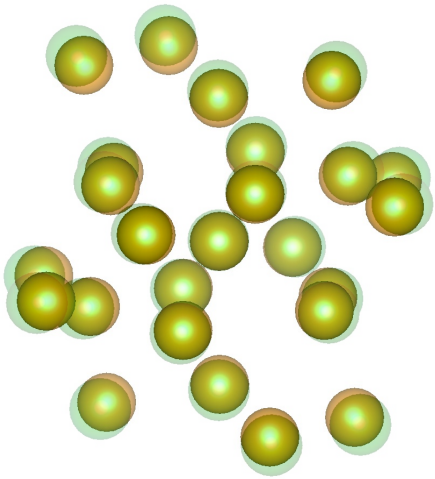

Figure **37**. The change in the $(SR)_{18}(Au)_{25}$ structure on the removal of the $(SR)_{18}$ ligands highlighted by overlapping images of the cluster before (shown in solid gold colour) and after PBEsol optimisation (transparent green).

Furthermore, we proceeded with the optimisation of a generic LJ25[144,145] cluster transformed by replacing its atoms with gold. Given that the original LJ25 cluster was very compact, we embarked on optimising three versions of the cluster, each one expanded to different degrees. Indeed, the interatomic distances in the initial LJ25 cluster are around 1.1~1.2 Å, we expanded the clusters 1.5, 2.0, 2.5 times from the centre of the geometry. The "1.5" and "2.0" structures have collapsed to different configurations from the starting LJ25 cluster structures which the configurations collapsed to have more expanded configurations as on expansion the atom-atom separations were still below those that would be defined from the vdW radius of gold atoms. The PBEsol optimised configurations of the "1.5, 2.0 2.5" expanded LJ25 clusters by are ranked 74th, 104th, 115th within the 1000 global optimisation LM dataset, respectively. Except the 104th ranked LM, the rest of the expanded LJ25 were not found from the global optimisation. Figure 38 shows that using the degree of expansion of around 2.5 retains the atomic configuration of the original LJ25 cluster on PBEsol optimisation.

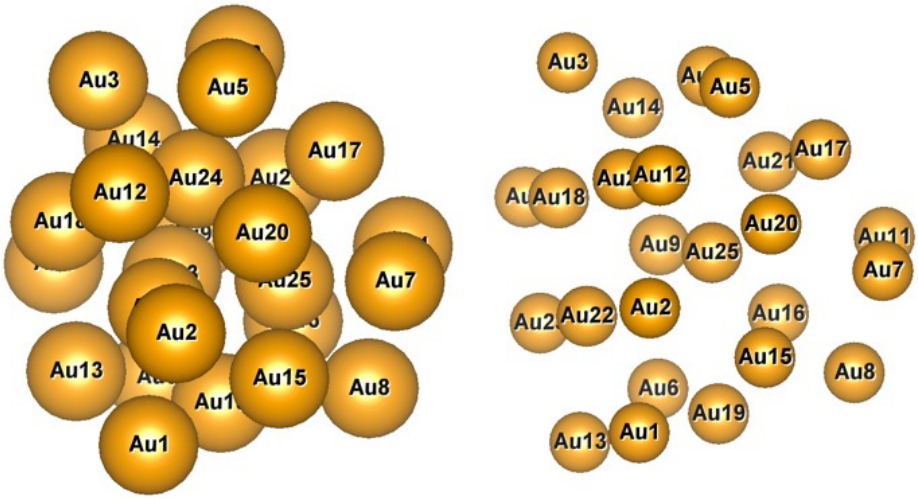

Figure **38**. Left: the original LJ25 cluster which the atoms are replaced to Au atoms; right: the optimised 2.5 times expanded LJ25 cluster (left) from the centre of the geometry.

5.3. *L*-cysteine protected Au_{25}

5.3.1. Structural Properties

The initial exploration of $(SR)_{18}(Au)_{25}$ involved attaching $(SR)_{18}$ ligands onto the highly symmetrical Au_{25} cluster[128]. We tried to attach 18 *L*-cysteine (*Cys*, $HSCH_2CH(NH_2)CO_2H$) ligands onto the PBEsol GM and 2nd LM. In the earlier research[128], 18 of phenylethanethiol ligands ($C_6H_5CH_2CH_2SH$) were systematically added to the Au_{25} cluster because of its symmetry, even though these ligands have much larger branches. However, using the same method to attach ligands onto the PBEsol GM or 2nd LM was not possible because these clusters do not have symmetry as shown in figure 39.

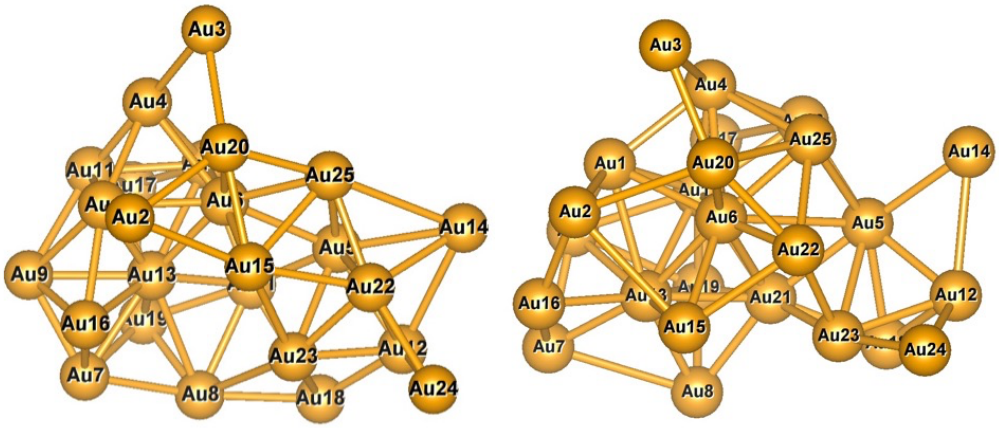

Figure **39**. Au_{25} GM comparison using a *ball-and-stick* representation. On the left is shown the PBEsol refined structure; on the right the same PBEsol Au_{25} GM structure is reoptimized after adding 8 *L*-cysteine ligands (the *Cys* ligands are not shown).

We chose to attach *Cys* as protective ligands to the Au_{25} cluster in this study as our research is part of a collaborative effort with experimentalists, aimed at comparing our theoretical predictions with their experimental observations. These experimental observations, utilizing *Cys* as protective ligands, will be detailed in an upcoming publication, which will include high-resolution transmission electron microscopy (HR-TEM) images of the cluster, among other findings.

The integration of *Cys* ligands onto the PBEsol GM Au_{25} cluster through structure optimizations induces a significant restructuring of the original configuration as seen from figure 39, which shows the configurations of Au_{25} before and after ligands are added. The grafting of *Cys* on Au was purely random.

The radical changes of the Au_{25} cluster structure should be expected from the change in the chemistry of the system upon the adoption of the Cys_{18} ligands. Initially, the cluster appears without a core, showing its distorted planar state, where every atom forms the cluster surface. However, following the introduction of 18 *Cys* ligands, a significant reconfiguration occurs, with an Au_6 atom now occupying the core position. This transition highlights a significant change in the cluster's structural composition, moving from a state without a core to one with an evident core, which suggests a decrease in surface area and a compaction of the

structure. Moreover, the introduction of the Cys_{18} ligands leads to a withdrawal of electrons from the gold cluster to ligands creating a Madelung field across the cluster, potentially contributing to a more confined configuration. Such observations illuminate the intricate interplay between ligand interactions and cluster morphology, providing insights into the factors governing nanocluster behaviour and stability.

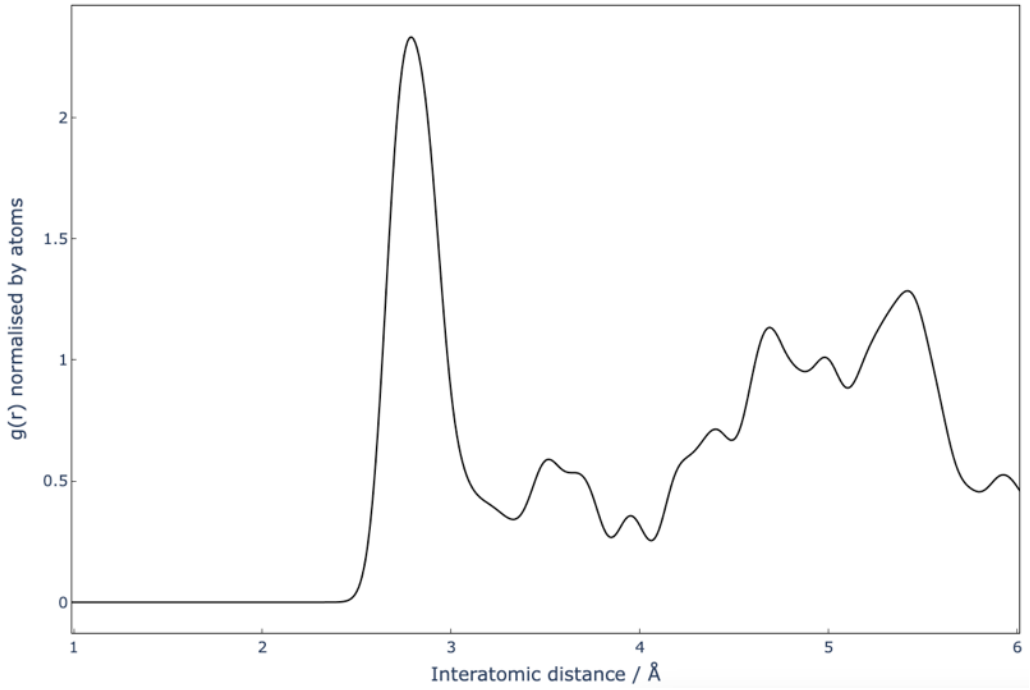

Figure **40**. RDF of the $Au_{25}(Cys)_{18}$ cluster, excluding the $(Cys)_{18}$.

Figures 39 and 40 illustrate the structural features of the $Au_{25}(Cys)_{18}$ cluster without the cysteine ligands, focusing on the Radial Distribution Function (RDF) and the principal moments of inertia of the Au_{25} core structure.

In the RDF plot shown in Figure 40, the first peak at approximately 2.9 Å corresponds to the interatomic distances between the nearest-neighbour Au atoms in the cluster. This peak has a weak shoulder at slightly longer distances, indicating slight variations in the nearest-neighbour distances among the gold atoms. The second peak at around 3.7 Å represents the distances to the second nearest neighbours within the cluster.

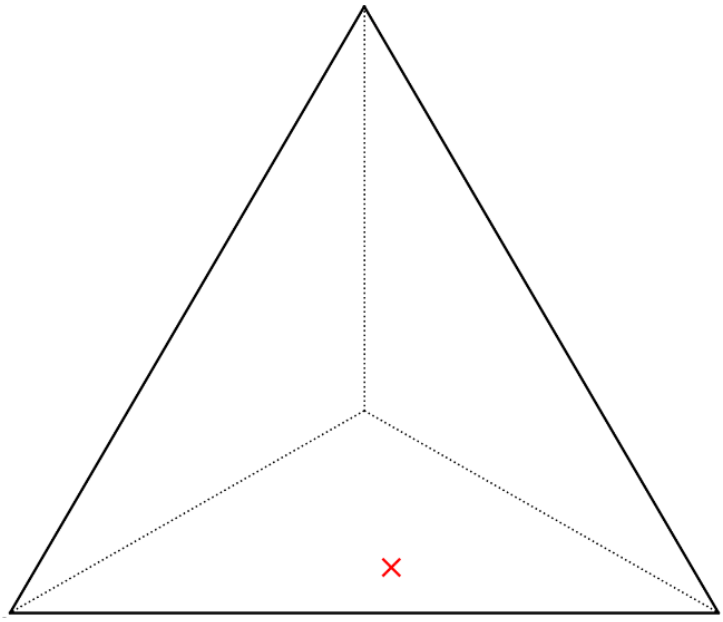

Figure **41**. The principal moment of inertia of the $Au_{25}(Cys)_{18}$ cluster stripped of *Cys* ligands shown in ternary plot.

Figure 41 presents a ternary plot of the principal moments of inertia of the Au_{25} cluster. The principal moments of inertia are measures of the cluster's resistance to rotational motion about its principal axes, which are orthogonal to each other—analogous to the x, y, and z axes in a Cartesian coordinate system. Each vertex of the ternary plot corresponds to one of these principal axes.

The principal moments of inertia are fundamental quantities in rotational dynamics that describe how a body's mass is distributed relative to its rotational axes. For a rigid body, the inertia tensor I encapsulates the mass distribution. By diagonalising the inertia tensor, one obtains the principal moments of inertia I_1, I_2 and I_3, corresponding to the body's resistance to rotation about its three principal axes.

Mathematically, the principal moments of inertia are calculated using:

$$I_i = \sum_n m_n\left(r_n^2 - x_{n,i}^2\right), \qquad \textit{Eqn } 59$$

where, I_i is the principal moment of inertia about axis i, m_n is the mass of atom n, r_n is the distance of atom n from the centre of mass, $x_{n,i}$ is the component of the position vector of atom n along axis i.

In the context of nanoclusters: Equal Principal Moments ($I_1 = I_2 = I_3$): The cluster is perfectly spherical, with mass evenly distributed in all directions; Unequal Principal Moments: Differences among the principal moments indicate elongation or flattening along specific axes. A smaller I_i suggests that the cluster extends less along axis iii, meaning the mass is closer to this axis.

The plot reveals that the Au_{25} cluster has an ellipsoidal shape rather than a spherical one. The two larger principal moments of inertia are relatively balanced, as indicated by their positions near the bottom two vertices of the ternary plot. This suggests that the mass distribution along two of the axes is similar. The third principal moment of inertia is smaller than the other two, corresponding to the axis along which the cluster is elongated. This indicates that the Au_{25} structure is stretched in one direction.

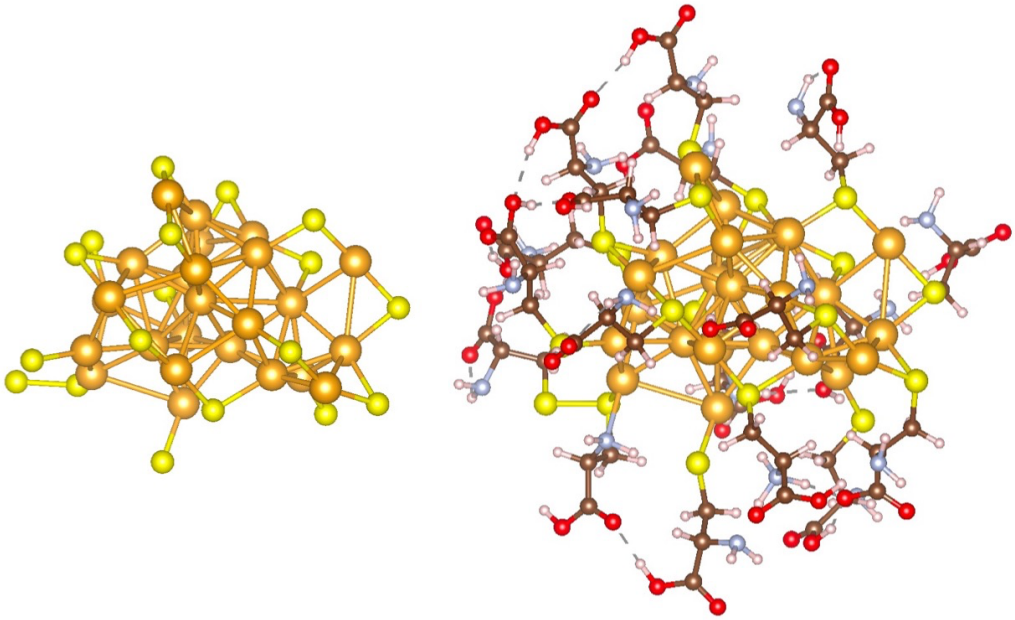

Figure **42**. The structure of ligand *Cys*-Au interface in $Au_{25}(Cys)_{18}$. Left: depiction of the $Au_{25}(Cys)_{18}$ cluster using a *ball-and-stick* model, showing the model with omitted ligand branches except for the sulphur atom which directly attaches to the

Au_{25} core. Right: this representation provides a comprehensive view of the full structure of the $(Au)_{25}(Cys)_{18}$ cluster.

The figure 42 presents two distinct images delineating the $Au_{25}(Cys)_{18}$ cluster. In the left image, the cluster is depicted with the $(Cys)_{18}$ ligands, deliberately excluding the remainder of the *R* branch except S atom to elucidate the manner in which the ligands interface with the Au_{25} core. Conversely, the right image offers a comprehensive view of the entire structure of the $Au_{25}(Cys)_{18}$ cluster, showcasing all ligands and atoms.

The process of ligand grafting was initially conducted manually using the Material Studio code[146], a meticulous procedure undertaken to ensure precision and accuracy in ligand placement. Following this manual attachment of each ligand onto the PBEsol Au_{25} GM, the *Clean* tool was subsequently employed to optimize the spatial arrangement of the ligands. This optimization process was crucial in minimizing steric hindrance by providing each ligand with adequate space, thereby enhancing the overall stability of the cluster structure.

Building upon insights from prior studies, which indicated that each S atom in the ligands typically bonds with two Au atoms, we endeavoured to replicate this bonding pattern during the grafting process. However, following the structural optimization utilizing the PBEsol method, notable deviations in the Au-S atom connectivity were observed. Specifically, several Au-S atom connections manifested three distinct connectivity patterns, highlighting the dynamic nature of the bonding interactions within the $Au_{25}(Cys)_{18}$ cluster – see figure 43.

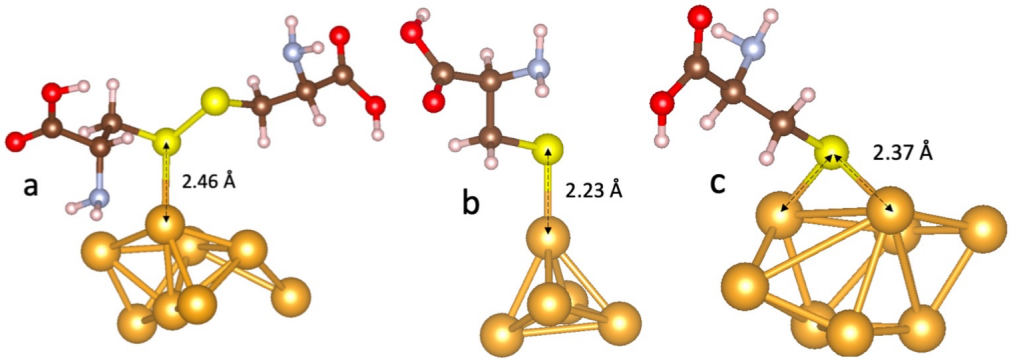

Figure **43**. Ligation patterns in the PBEsol-optimized $Au_{25}(Cys)_{18}$ cluster. Three distinct sulphur atom configurations are observed regarding their connectivity with Au atoms. The figure displays partial views of Au25 along with the immediate ligands attached. (a): Two *Cys* ligands are connected through a sulphur atom, with one of the sulphur atoms also linked to a Au atom; (b): a sulphur atom of the L-cysteine ligand is directly connected to a Au atom; (c): a sulphur atom is bonded to two Au atoms simultaneously.

It became evident that a prominent trend emerged in the connectivity patterns between S atoms and Au atoms within the cluster structure. Specifically, the configuration where S atoms are doubly connected to Au atoms (referred to as connectivity "c" in figure 43) with the average bond distance of 2.37 Å emerged as the predominant motif, manifesting in the majority of instances. This particular connectivity scheme, wherein 13 ligands form bonds with two Au atoms each. This -Au-S-Au- connectivity is the only connectivity that found in the earlier report[128,147] on the crystal structure of $(SR)_{18}Au_{25}$. In contrast, we noted a less frequent occurrence of *Cys* ligands bonding to one Au atom, either individually or in tandem with other ligands. These occurrences were comparatively infrequent, with only 1 instance of joint ligand attachment and 4 instances of individual ligand attachment to a single Au atom with the 2.46 Å and 2.23 Å bond distances, respectively. The four of the individual ligand bond to a single Au atom are generally on the Au atom which has fewer nearest neighbours as the cutoff distance from a Au atom increases the Au with the single footed L-cysteine captures fewer nearest neighbours.

5.3.2. Electronic Properties

Using the PBEsol refined configuration of the $Au_{25}(Cys)_{18}$ nanocluster we next considered its electronic structure. Analysis of its Density of States (DOS) shown in figure 43 provides a detailed picture of the electronic structure influenced by both the Au and the surrounding *Cys* ligands. The DOS of the PBEsol optimised cluster was evaluated using the PBEsol0 hybrid functional[148,149] with light basis set as implemented in FHI-aims[150]. In this cluster, Au exhibits significant spin polarization, with a pronounced peak in the up-spin DOS observed at approximately -4 eV, which extends over a broad range from -3 eV to 7 eV. This suggests a robust magnetic behaviour predominantly driven by the Au atoms, potentially affecting the cluster's magnetic susceptibility and spin-dependent electronic transport properties.

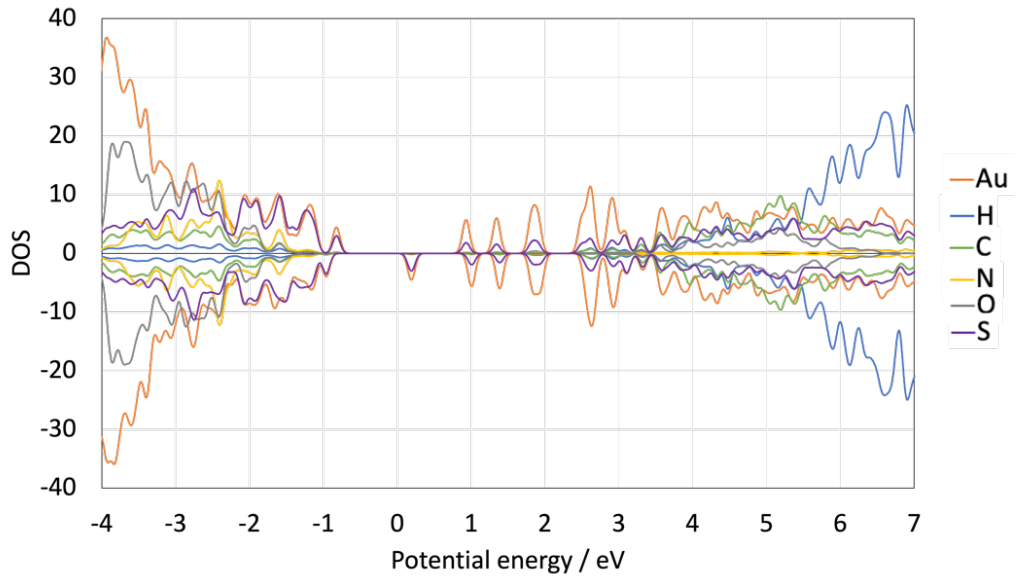

Figure **44**. The Density Of State (DOS) of PBEsol0 evaluated PBEsol optimised $Au_{25}(Cys)_{18}$. Each colour represents each element contribution, and the positive value is the contribution from the up spin and the negative value is from the down spin.

In contrast, atoms from the ligand molecules such as H, C, N, O, and S show more balanced spin contributions. Their DOS profiles, featuring peaks for both up and down spins across various energy ranges, suggest a more neutral electronic

environment. Notably, S displays distinct peaks for both spins, emphasizing its role in mediating electronic interactions within the cluster. This balance in the electronic states of the ligands contributes to the overall stability of the cluster's electronic structure. Research supports this, highlighting that ligands can polarise electron density, enhancing cluster stability and reactivity. Studies on TMC-supported clusters demonstrate that balanced electronic states in ligands prevent charge accumulation, maintain structural integrity, and improve catalytic efficiency. Ligands like sulphur in gold clusters mediate significant electronic interactions, crucial for stability by facilitating electron delocalisation and preventing uneven charge distribution[151]. The DOS plot also delineates the energy distribution between negative and positive regions, indicating filled and unfilled electronic states, respectively. A noteworthy feature in the DOS plot is the low density of states around the Fermi level, which reveals a band gap of approximately 1.01 eV. This band gap agrees with the current collaborate experimental observation which shows a very close value.

From the figure 45, the calculated optical absorption spectra for both spin-up (red line) and spin-down (blue line) electrons closely match the experimental data (black dashed line) at lower energy values, up to around 2.5 eV. The summed spectrum (green line) also closely follows the experimental trend within this range, especially with the peak showing the first excited states to occur at around 1.8 eV, which is mostly contributed to by the down-spin transitions. Multiple molecular transitions are, in fact, found, to determine the optical absorption spectrum; forming at least four well-defined bands at 1.8, 2.4, 2.8, and 3.4 eV.

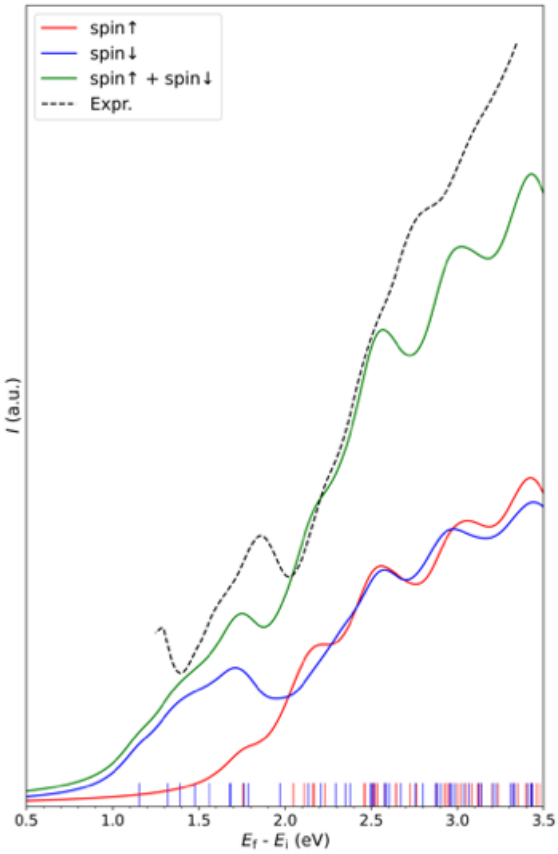

Figure **45**. Optical absorption spectra of $Au_{25}(Cys)_{18}$, including calculated spin-up (red solid line) transitions, spin-down (blue solid line) transitions, and their combined effect (green solid line), against experimentally obtained UV-vis data (black dashed line) [GB Hwang, *et. al.*, Private communication 2024]. The spectra are plotted as a function of the energy difference $E_f - E_i$ between respective occupied and unoccupied Kohn-Sham eigenstates, indicating transitions from 16 selected states ranging from HOMO-7 to LUMO+7., A Lorentzian function with a dispersion of 0.15 eV was applied to broaden the lines. *N.B.* No selection rules have been applied for the prediction, *i.e.*, any transitions from one of the lower 8 states to one of the higher 8 states were considered.

The calculated absorption spectra closely match the experimental UV-vis data within the lower energy range, demonstrating that the computational approach effectively captures the essential electronic transitions of the system in this domain. However, beyond approximately 3 eV, discrepancies emerge where the experimental absorption values continue to rise while the calculated data

plateau. This divergence suggests that while the theoretical method is reliable for lower energy transitions, it may not fully account for all mechanisms active at higher energies. Specifically, high-energy electron relaxation effects—such as many-body interactions, electron-electron correlations, and the subsequent relaxation of electrons after excitation—are not captured by the adopted simple one-electron approximation.

In a one-electron approximation, each electron is considered to move independently in an average potential created by all other electrons, neglecting dynamic interactions and correlations between electrons. At higher energies, these many-body effects become more pronounced, leading to more complex relaxation dynamics that influence the absorption spectra. For instance, after an electron is excited to a higher energy state, it can interact with other electrons, leading to processes such as exciton formation, plasmon resonances, or multi-electron excitations. These interactions can significantly alter the energy distribution and intensity of absorption peaks, contributing to the rising absorption values observed experimentally beyond 3 eV.

The inability of the one-electron model to incorporate these interactions results in the underestimation of absorption intensities at higher energies, as seen in the experimental data. To accurately model these high-energy phenomena, more sophisticated approaches that include electron correlation effects are necessary. Methods such as time-dependent Density Functional Theory (TD-DFT) with advanced exchange-correlation functionals can account for the interactions between excited electrons and the resulting relaxation processes.

As this thesis focuses on computational studies and is currently being prepared for publication, it is not possible to disclose many details, particularly those related to experimental observations. However, the two distinct studies, employing different methodologies—computational and experimental approaches—have yielded highly consistent results, which will be ready for publication within this year.

The matching results suggesting that the calculations effectively capture the dynamics of the system in this energy domain. However, beyond this point,

particularly past 3 eV, there are discrepancies where the experimental values continue to rise while the calculated data plateau. This indicates that while the theoretical approach is reliable for lower energy transitions, it may not fully capture all mechanisms active at higher energies, in particular high energy electron relaxation effects which would not be accounted for by the adopted simple one-electron approximation.

As this thesis focuses on computational studies and is currently being prepared for publication, it is not possible to disclose many details, particularly those related to experimental observations. However, the two distinct studies, employing different methodologies: computational and experimental approaches, have yielded highly consistent results, which will be ready for publication within this year.

6. Machine Learning Interatomic Potential

This chapter looks into Machine Learning Interatomic Potentials (ML-IPs), focusing on training ML-IP for nanocluster structure predictions and evaluating the ML-IPs which are trained on selected nanoclusters and the recently developed universal potential using MACE[51,152]. The universal potential is trained with 1.6M bulk crystals in MPTrj datasets[153] which can cover 89 elements. MACE is known as current state-of-art of ML-IP and it is known as reliable and flexible, making it promising for accurately predicting how atoms arrange themselves and their energy levels[152,154].

We compared AlF_3 structures that have been optimised using MACE with those PBEsol optimised structures in the previous study. By doing this, we anticipate observing that MACE can figure out with high accuracy of the energy ranking and geometry of AlF_3 structures or at least the atomic connectivity. This comparison is important because it can show us how well the ML-IPs can work for understanding the structure of inorganic nanoclusters.

The background of the initiating this project was we started looking into ML-IPs as it is tough to find a way to describe how atoms interact with IP because of the complexity of parameterizations and limitation of IP. Most of IPs has lack of transferability, thus parameters need to be fitted for different systems and purposes. In fact, during my master project[4] we performed global optimisation on Ti_3N_4 nanocluster's PES using the fitted IP, but the IP could not describe the Ti_3N_4 PBEsol energy landscape with enough accuracy: could not predict accurate energy rankings or configurations. Thus, we evaluated how the ML-IPs using the AlF_3 nanocluster dataset from the previous study.

6.1. Pairwise Interaction

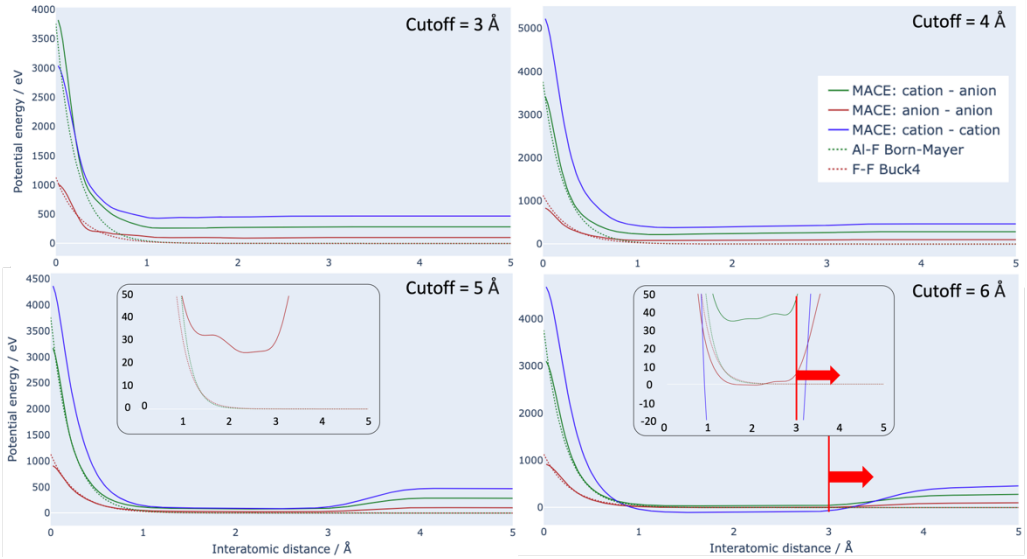

Figure **46**. Pairwise interaction of the Al-F (green), F-F (red), and Al-Al (blue) as a function of interatomic distances. The dotted line is the predicted pairwise interactions using AlF_3 potentials. The 'cutoff' is a parameter used to train the potential. The red bar and arrow at cutoff = 6 Å indicate where the ML-IP was not trained with interatomic distance data. The smaller panels display the same plot as the main plot but with a reduced *x*/*y*-axis range

Before we jump into the ML-IP to perform global optimisation on complex material like Ti_3N_4, we have evaluated the ML-IP using the IP data, the first step of the two step global optimisation method, to see how the ML-IP behaves with cheaper computational cost than using DFT. The figure 46 shows the pairwise interaction of Al-F (green), F-F (red), and Al-Al (blue) which are predicted using trained ML-IPs, and the dotted line is our target to achieve using the ML-IP which is the fitted IP in the earlier AlF_3 chapter. The ML-IP used to prepare each panel in the figure was trained with the identical training data, but with the different cutoff parameter was used during the training process which are 3.0 Å, 4.0 Å, 5.0 Å, and 6.0 Å. The training data is the IP evaluated Al-F dimer images which has the interatomic distance between 0.01 – 3.00 Å in every 0.01 Å. As part of training set the atomic energy of Al and F atom are not provided explicitly, instead it

guessed from the individual training structures which deduced from the total energy of the structures.

Ideally the ML-IP's pairwise interaction (solid lines) should show the trend that overlapped with the original IP (dotted lines) as we trained the IP data. However, the panel shows the discrepancies in different region of interatomic distances. *N.B.* The *y*-axis has rather large range in potential energy to see the full and overall behaviour of the ML-IP to compared with the reference IP. If necessary, we attached small panel for more detailed or focused view. The ML-IP trained with 3Å cutoff parameter shows the least accuracy in the most range of the interatomic distances compared to the others. The Al-F pairwise interaction has the smoother profile than other two interactions. The Al-F interaction has relatively less RMSE at extremely short distances, but at longer interatomic distances the ML-IP predicts the interactions with less accuracy. Similarly, the ML-IP predicted better at the short distance for the F-F interaction but less accurate at longer distances with relatively less RMSE compared with the Al-F interactions. Initially, we expected the ML-IP can describe the pairwise interaction accurately as we provided the highly concentrated data, each image has different interatomic distance with 0.01 Å. Moreover, the longest interatomic distance in the image of the training data matches with the cutoff parameter for the training. Even though the training data is matches with the cutoff parameter it struggles to predict accurate pairwise interactions in the range. We thought the ML-IP would perform better than the result as we provide the full interactions in the range of the cutoff parameter. Interestingly, the training data did not have F-F and Al-Al interactions, and there were no IP for Al-Al interactions during the preparation of the training data except the Coulomb interactions, but ML-IP seems like to picked up the F-F/Al-Al interactions within the data and try to describe the interactions. This behaviour is unseen in the Gaussian Approximation Potential (GAP). The contents were not included in the thesis, but from the experience of other type of ML-IP, such as GAP, try to compensate between the component of the total energy of the system, all possible interactions. For example, if a pair show abnormal interactions, then the other two interactions include the opposite of the wrongly described interaction energy within their interactions to offset the anomality and to predict the accurate total energy of the system. MACE is different to the GAP

as it uses message passing neural network as part of the core architecture, thus because of it the behaviour of the potential has to be learnt mostly from black box approach.

Ironically, among the three possible pairwise interactions the F-F data which was not included in the training data was described with less RMSE than the Al-F interactions which is the only the included training data. Even though they could not show with high accuracy in describing the pairwise interactions in overall, the extremely shorter distances show the relatively better agreement than the longer distances. At longer distances they converged to much higher energy than the 0 eV, the reference.

The ML-IP trained with 4 Å cutoff parameter shows the better agreement than the 3 Å in all range. Especially at short distances they agreed well with the reference, but still it struggled at over 1 Å. Interestingly, the ML-IP trained with the same training data and 5 Å cutoff distances showed the best agreement within the 0 – 3 Å the covered area by the training data. However, it could not predict from 3 Å to longer distances which started to deviate from the reference data gradually up to near 4 Å. The deviation was not expected because the ML-IP would pick up the information about longer distances as it was training from the energy and the atomic forces, but it failed beyond the coverage of the training data. The small panel is magnified on potential energy with the same interatomic distances as the bigger panel which showed that the potential does not agree well in detail with overestimations. The longer cutoff 6 Å does not agree well compared to the 5 Å cutoff ML-IP, but at the region we are interested in near the AlF_3 equilibrium distances (1 – 3 Å) the predicted Al-F interactions agreed better, but soon after the beyond the coverage of the training data it spike up the energy for all possible pairwise interactions.

From the experimental training and prediction, it shows that the ML-IP could not predict well at relatively longer distance if the training data includes a lot of extremely short distance data. Also, the ML-IP needs all possible interactions as a training data as the last ML-IP trial shows that the ML-IP tried to predict all interactions while it only trained with Al-F interactions. We presume that the poorly describing in F-F and Al-Al interactions or lack of training on other

interactions causes the less accurate Al-F prediction. For example, if we want to train Al-F pairwise interactions still we have to include the F-F and Al-Al interactions within the training data. The cutoff distance has to be shorter or match with the longest interatomic distance among the training data, but we wanted to confirm the hypothesis, thus we tried to refit the potential with additional data between 3 – 5 Å with 5 Å as the cutoff.

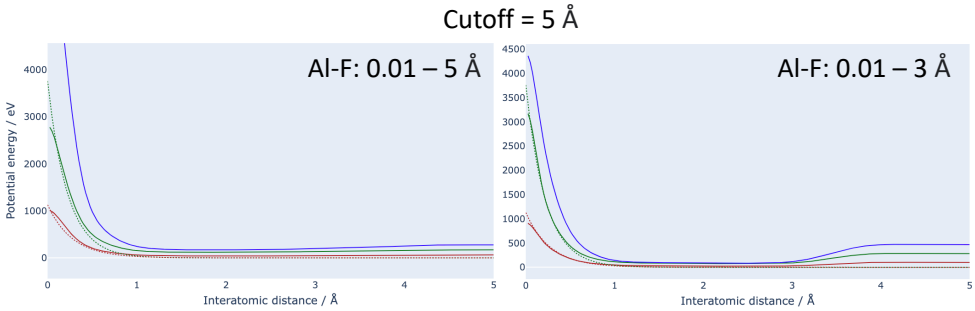

Figure **47**. The predicted Al-F pairwise interactions at interatomic distances using the ML-IP that trained with 0.01 Å – 5.00 Å (0.01 Å – 3 Å) interatomic distances with cutoff parameter of 5 Å.

Figure 47 shows the predicted pairwise interactions between 0 – 5 Å and the ML-IP was trained with 5 Å cutoff parameter, only the difference is the coverage of the training data. As we hypothesized earlier from the last experiment the figure proves that the cutoff parameter for the ML-IP training should be within the range of the coverage of the training data. The matching cutoff parameter with the training data (left panel) shows the smooth profile, but the right panel which the cutoff parameter is longer than the coverage of the training data has a gradual increasement in energy from 3 Å until 4 Å.

In the next attempt, we trained ML-IPs with the same manner as the last experiment, but without the extremely short unphysical interatomic distances. The left column of the figure 47 shows that the ML-IP predicted Al-F (solid green) shows the sensible behaviour in longer than 1 Å interatomic distances across the ML-IP trained with 3 – 6 Å of cutoff parameter. However, F-F and Al-Al interactions shows fluctuations as there were not training data. Thus, we have added the IP F-F dimer data that matches with the Al-F training data coverages. The ML-IP trained with Al-F and F-F was used to predict the pairwise interactions

are shown in RHS column of the figure. Still there are significant fluctuations in potential energy for the ML-IP trained with 3 Å and 4 Å cutoff parameters at extremely short interatomic distance where the ML-IP are not trained on. As we anticipated after the shortest interatomic distance (1 Å) in training data shows the accurate trend or forces with notable overestimation in potential energies, ~ 5 eV at 5 Å, for all ML-IPs trained with 3 – 6 Å cutoff.

The only the difference in the left and right column of plots which are predicted data using the ML-IP that trained only with Al-F interactions, or Al-F and F-F interactions, respectively. Between the two set of ML-IP trained with different number of type of interactions do not show notable difference in accuracy in predicting Al-F interactions. Of course, the F-F performed better when it trained with compared to the one only trained with Al-F interactions and no F-F interactions. Interestingly, the ML-IP trained with Al-F and F-F interactions changed the Al-Al interactions which the trend looks like the other interactions.

In this experiment we have changed in the training data: the coverage of the Al-F interactions to 1 – 3 Å by removing the extremely short interatomic distance (< 1 Å) images in the training data, and adding the F-F interaction images with the 1 – 3 Å range in the training data. Different to the ML-IP trained on the only 0.01 Å – 3 Å Al-F interactions is as we include the additional pairwise interactions it performs better in predicting the interactions where we are interested in, near the equilibrium distances. The changes in energy as a function of interatomic distances (*e.g.*, force) are consistent. Moreover, there are no sudden energy jump after the coverage of the training data while the cutoff distance is longer. We presume as we include more of the possible combination of atomic interactions corrects the contribution of the interactions within the total system by removing the random prediction on pairwise interactions where no data, likewise if we include the Al-Al interactions which is 0 in all range it will provide accurate profile of the dimer curve compared to the reference dotted data.

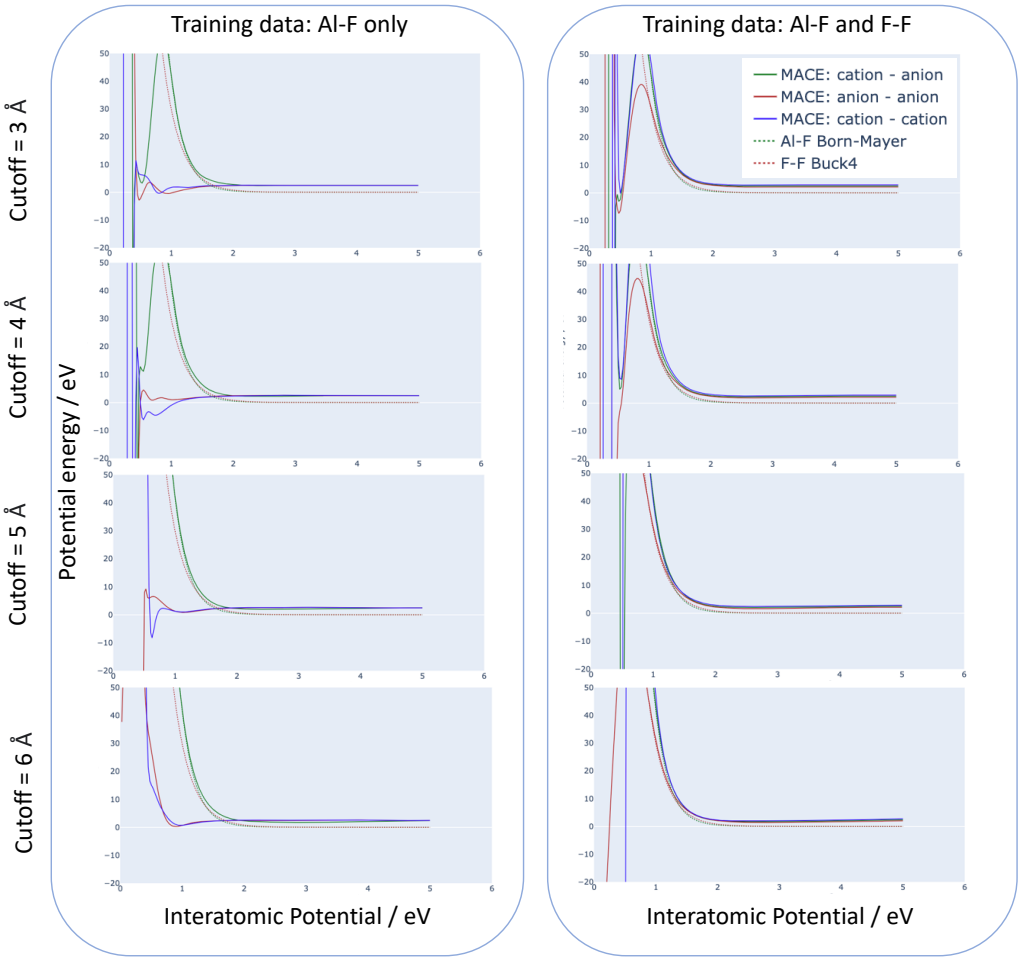

Figure **48**. The predicted Al-F (green), F-F (red), and Al-Al (blue) pairwise interactions which are predicted by the trained ML-IP using the same training data; the left column is trained with the IP evaluated Al-F dimer with 1 – 3 Å interatomic distances in every 0.01 Å; the right column is trained with the same training data as the left column with additional F-F dimer with 1 – 3 Å in every 0.01 Å. Each row of the ML-IP was trained with different cutoff parameter 3 Å, 4 Å, 5 Å, and 6 Å.

From our previous trial, we observed that the improvement in describing pairwise interactions at longer interatomic distances diminished when the training set coverage for Al-F interactions was reduced from 0.01 Å–3 Å to 1 Å–3 Å. To address this, we trained the machine learning interatomic potentials (ML-IPs) for Figure 48 in the same manner as Figure 47, but extended the training set coverage to 1 Å–5 Å. We hypothesized that ML-IPs trained with a broader range of

interatomic distances would predict pairwise interactions more accurately. Consistent with our expectations, the ML-IP trained exclusively on Al-F dimers with distances between 1 Å and 5 Å accurately predicted Al-F interactions, converging closely to 0 eV or nearly 0 eV. We believe that the ML-IP with a 5 Å cutoff parameter achieved a more accurate Al-F dimer curve compared to other models. This extended hypothesis builds upon our previous experiments. Since we provided only one type of interaction, Al-F, the other two interactions (F-F and Al-Al) were randomly placed within the boundaries defined by the training data, which included properties such as total energy and forces. Consequently, the ML-IP made random predictions for the untrained F-F and Al-Al interactions, influenced by the Al-F contributions. Based on our experience with GAP ML-IPs, we observed that while the dimer curves for Al-F interactions showed different trends, the overall results in Figure 48 demonstrated that providing longer interaction coverage in the training data led to more accurate predictions of pairwise interactions, regardless of the cutoff parameters. We presume that supplying a sufficient number of data points with broad coverage enables the ML-IP to learn the trends of pairwise interactions effectively during training, thus achieving high prediction accuracy.

In the right-hand side (RHS) column of Figure 48, the ML-IP was trained with the same Al-F coverage as the left-hand side (LHS) column but included additional F-F pairwise interactions with the same coverage as the Al-F interactions. The RHS not only showed smaller root mean square errors (RMSE) at longer interatomic distances but also exhibited significantly lower RMSE at shorter distances compared to the reference data. The F-F pairwise interactions at unphysically short distances became more accurate with the ML-IP trained using a longer cutoff parameter of 5 Å to 8 Å. The ML-IP trained with an 8 Å cutoff parameter predicted interactions with the least RMSE compared to other ML-IPs in the figure. Overall, the 8 Å cutoff parameter plot indicates that only areas with very short interatomic distances (less than 0.9 Å) present challenges in accurately describing pairwise interactions. However, the ML-IP with an 8 Å cutoff does not exhibit extreme behavior, such as the Coulomb or Buckingham catastrophe, where pairwise interactions show extreme attraction forces at short distances. Furthermore, since our study does not involve molecular dynamics simulations

under high pressure or temperature conditions, the accuracy of short pairwise interactions is not critical. Therefore, the ML-IP trained with an 8 Å cutoff in Figure 48 remains suitable for global optimizations and structure relaxations. For example, when using the ML-IP for global optimization, the random structure generator should incorporate acceptance or rejection criteria for randomly generated structures to prevent atoms from being placed too closely, thus avoiding incorrect interpretations of pairwise interactions at distances below 0.6 Å. In summary, ensuring that atoms are not placed too closely during the random structure generation steps is necessary before performing structure optimizations using the ML-IP.

The summary of the gained information from the experiments: (1) if the ML-IP trained with all possible pairwise interactions within a system, (2) training with longer coverage of interatomic distances between a pair of atoms, and (3) if coverage of training data is broad enough then longer the cutoff parameter than the longest coverage of the pairwise interaction in training data would generate more accurate ML-IP. Still there is one more thing that we could try which is providing the explicit value of the atomic energies which consists of the system that ML-IP learn.

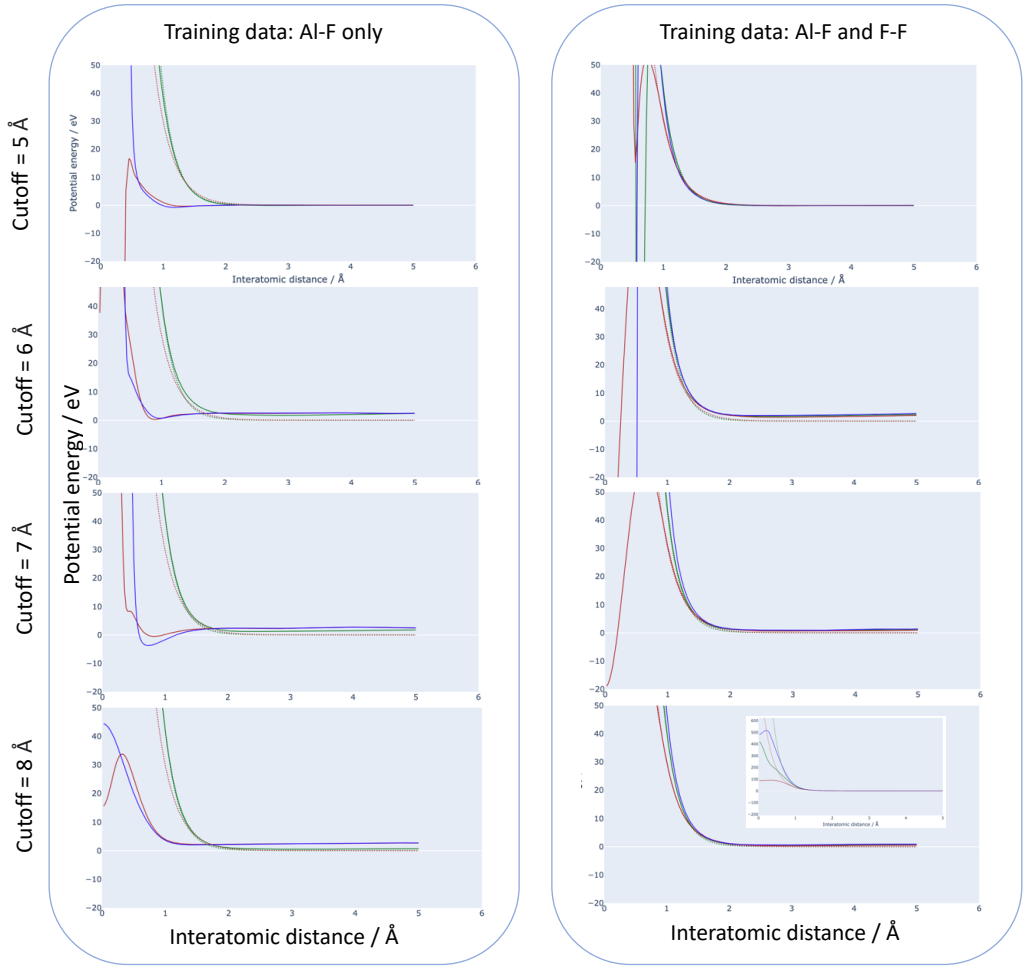

Figure **49**. The predicted Al-F (green), F-F (red), and Al-Al (blue) pairwise interactions which are predicted by the trained ML-IP using the training data; the ML-IP used to prepare the plot in the left column are trained with the IP evaluated Al-F dimer with 1 – 5 Å interatomic distances in every 0.01 Å; the right column plots are prepared with the ML-IP that trained with the same training data as the left column with additional F-F dimer with 1 – 5 Å in every 0.01 Å. Each row of the ML-IP was trained with different cutoff parameter 6 Å, 7 Å, and 8 Å.

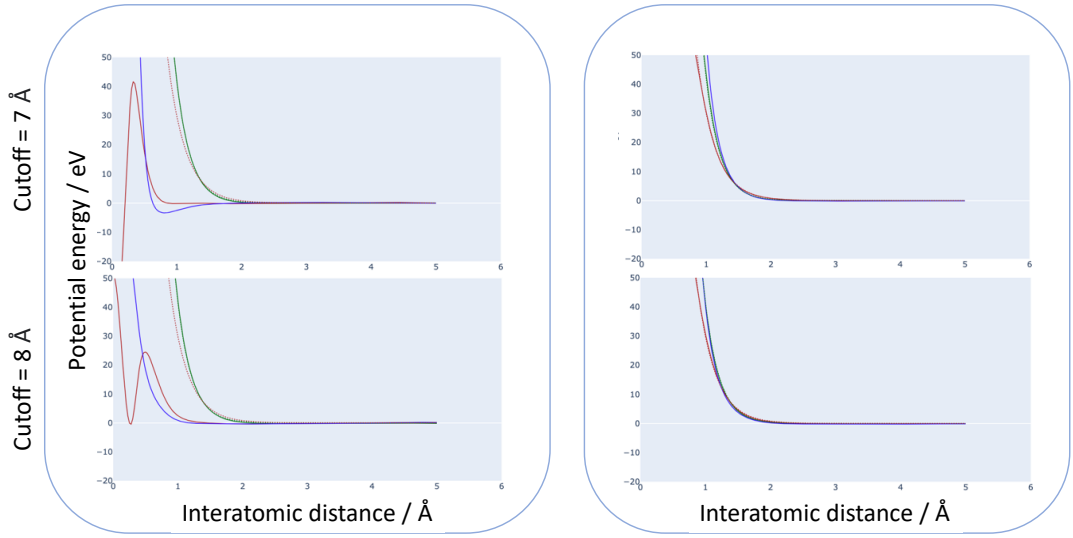

Figure **50**. The predicted Al-F (green), F-F (red), and Al-Al (blue) pairwise interactions which are predicted by the trained ML-IP using the training data; the ML-IP used to prepare the plot in the left column are trained with the IP evaluated Al-F dimer with 1 – 5 Å interatomic distances in every 0.01 Å and Al and F atomic energies; the right column plots are prepared with the ML-IP that trained with the same training data as the left column with additional F-F dimer with 1 – 5 Å in every 0.01 Å. Each row of the ML-IP was trained with different cutoff parameter 7 Å, and 8 Å.

Figure 50 shows the performance of the ML-IP which the training data is identical to the last experiment: (LHS column) trained with the images of Al-F dimer with interatomic distance of 1 – 5 Å in every 0.01 Å; (RHS column) same as the LHS with including the Al – F dimers the F – F dimers in the same range of interatomic distances, but this time we provided the IP evaluated atomic energy, 0 eV for both. Because the IP only looks at how atoms interact one-on-one and doesn't consider how electrons exchange and correlate like the DFT method does, it gives an atomic energy of 0 eV. As LHS column of figure 50 shows the only Al-F interactions and the atomic energies does not show significant improvement in reducing the noise at the short distance interactions. For the training process the atomic energies (Al and F) are provided instead of using the images to estimate the atomic energies in the training data. Thus, providing the explicit atomic energies will help to converge the dimer curve at longer distance to 0 eV as the

atoms in a pair is in long distance they cannot interact noticeably with each other. The predicted dimer curve using the ML-IP trained with the atomic energies and the Al – F and F – F pairwise interactions data is shown in the right column. The plots show that the ML-IP for both has the less RMSE at shorter interatomic distances and the dimer curve converges to 0 eV at longer interatomic distances. The ML-IP trained with the 8 Å cutoff parameter has the least RMSE in all range of interatomic distances without the abnormal description of the short-range interactions in below 1 Å.

The ML-IP, trained with Al-F pairwise interaction experiments, suggests that including a broader range of pairwise interactions and combinations, particularly those with longer distances, along with explicit atomic energies, improves the ML-IP's accuracy in describing pairwise interactions. Based on insights from prior black box experiments on the ML-IP, we conducted experiments to train the ML-IP using the PBEsol-optimized smallest $(AlF_3)_n$ cluster, with $n = 1$, before moving on to train larger clusters. The training set is the images of the full vibrational modes and each vibrational mode data consisted of 10 images. In the same way as the previous experiment with ML-IP that trained with the IP data, we predicted the pairwise interactions of Al – F and F – F using the images of the $(AlF_3)_1$ vibrational modes trained ML-IP which the result is shown in the figure 51. Note that the images of vibrational mode of $(AlF_3)_1$ did not include the Coulomb interactions, thus the dimer curve has the LJ-like trends.

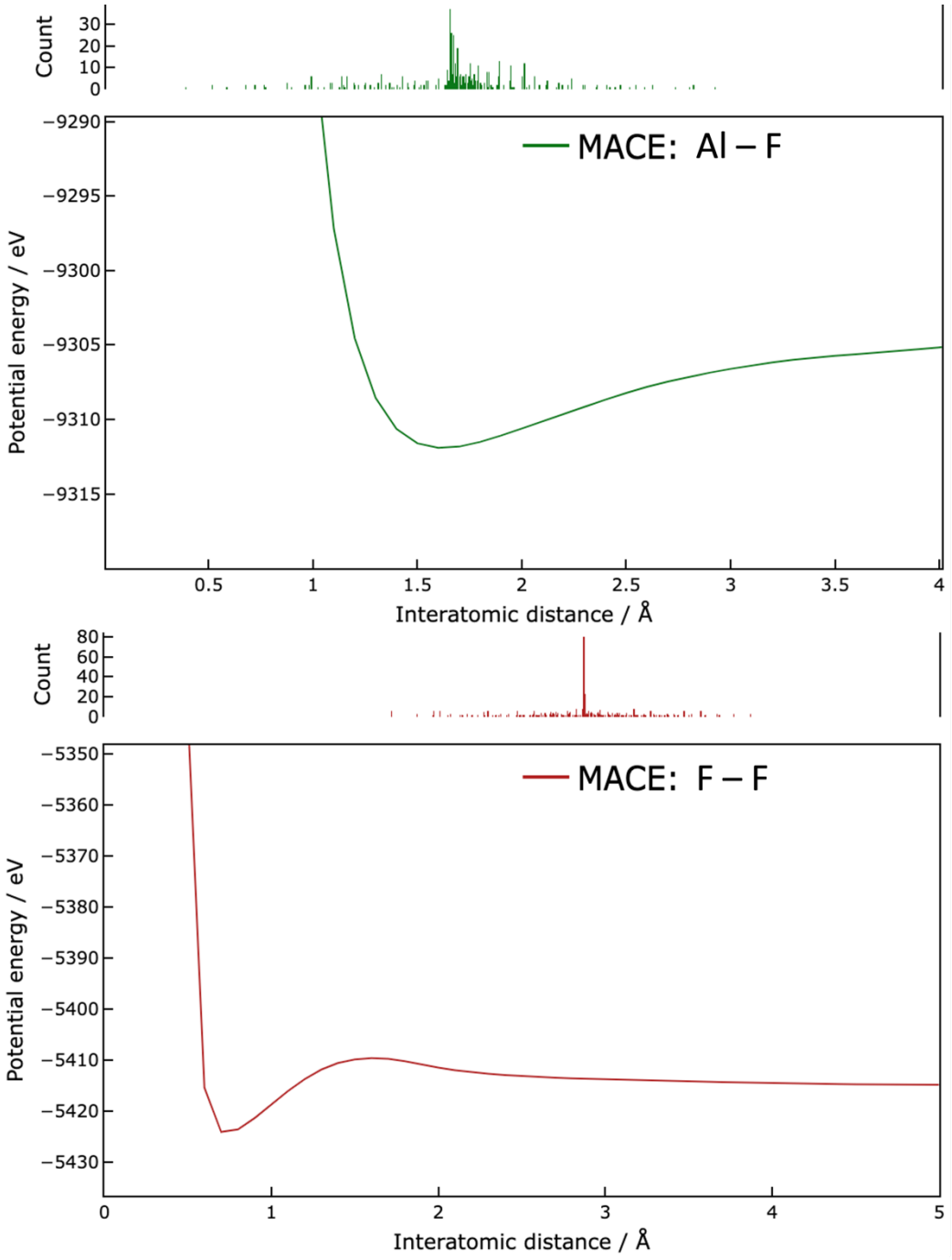

Figure **51**. Pairwise interactions of Al-F (top) and F-F (bottom) predicted using the ML-IP. The ML-IP was trained on PBEsol-optimized images of all vibrational modes from $(AlF_3)_n$ (n = 1) structures, excluding rotations and translations. The hover plot above the predicted interactions displays the number of interatomic distances observed in the training data.

The small hover plot above the pairwise interaction plot presents the number of specific interatomic distances from images of the vibrational mode in the training data. Analysing the count of interatomic distances within the training data allows us to assess which ranges of distances are covered in the training and evaluate whether this affects the ability of the ML-IP to describe accurate interactions by comparison with dimer curves. By comparing the trained interatomic distance data with the dimer curve predictions, we can assess the quality of the predictions made by the ML-IP.

The F-F pairwise interactions are poorly described between distances of 0.5 Å and 2 Å, showing two points where the forces between the two fluorine atoms are zero or at inflection points. The interatomic distances for the F (Al-F) atom pair are mostly concentrated around 3.87 Å (1.72 Å), displaying a normal distribution pattern. The most common F-F distance in the optimised $(AlF_3)_1$ structure is 2.87 Å. It remains uncertain whether regions with more data points are better trained than less sampled areas, or if this results in overfitting. However, the F-F pairwise interaction plot generally shows inaccurate predictions. The ML-IP incorrectly predicts 0.7 Å as the equilibrium interatomic distance for F-F, significantly less than the PBEsol-predicted distance, used as a reference. The poorly predicted equilibrium distance is caused from the lack of the training data at near the equilibrium distance which is shown in the hover plot.

The predicted equilibrium energy is also 6.5 eV lower than the PBEsol value. Notably, there is a saddle point at a distance of 1.6 Å, after which the forces become repulsive, converging towards an energy of -5414.8578 eV. We presume that the up to the saddle point from longer distances is the region that ML-IP could learn directly based on the training data and less than the saddle point the ML-IP extrapolated based on the learnt features. The energy convergence at longer distance represents the doubled atomic energy of F atoms, indicating the accuracy of the ML-IP in certain aspects. The training approach for the ML-IP was specifically set to prioritize accurate energy prediction over correct force estimation by us using the stochastic weight averaging (swa) options, enhancing the model's performance in energy accuracy more than in force accuracy. SWA is a training technique that improves model generalization by averaging the weights

of a NN over multiple training epochs. This approach helps the model converge to flatter minima in the loss landscape, enhancing its stability and performance on unseen data. In the context of fitting the ML-IP with the MACE potential, SWA was utilized to prioritize accurate energy predictions by stabilizing the model's energy estimations over force calculations.

Thus, it could also explain the saddle point that might predicted because of the heavier weight on energy setting than the forces. In contrast, the Al-F interatomic interactions are predicted with higher accuracy than the F-F interactions, including the region at near the equilibrium distance. Additionally, the long-distance convergence energy of Al-F is also accurately predicted at -9304.3432 eV.

Using vibrational modes from smaller nanocluster sizes often does not provide sufficiently short interatomic distances, especially between anions. Therefore, it is necessary to include vibrational modes from larger cluster sizes to accurately describe interactions near equilibrium distances. The inclusion of larger cluster sizes will help resolve the lack of near-equilibrium interatomic distances for F-F pairwise interactions, as these larger clusters will exhibit more complex combinations of vibrational movements. This complexity arises from the clusters' random-like configurations, as discussed in the AlF_3 chapter, which also involve a greater number of atoms, ensuring the vibrational mode images capture the necessary interatomic distances close to the equilibrium for the anion-anion interactions, F-F.

For the $(AlF_3)_n$ series, where $n = 1$ and 2, only one configuration exists in the PES, leading us to explore larger clusters such as $n = 6$ for training the ML-IP. In the $n = 6$ PES, there are over a thousand LM, as mentioned in the AlF_3 chapter most of which display random-like configurations whose vibrational modes cover the equilibrium interatomic distances. Training all these LM is computationally expensive, prompting the decision to select 10 local minima for training the ML-IP. We trained two different sets of ML-IP: one using the vibrational mode images

from the top 10 local minima and another trained with vibrational modes from 10 randomly chosen local minima.

Additionally, a recently developed universal potential trained on the Materials Project database[152,153] has been evaluated by comparing its results with reference PBEsol data. The results, including total energy and atomic forces of LM as functions of PBEsol data, are illustrated in the referenced figure. This comparison helps validate the effectiveness of the trained ML-IPs against established benchmarks. *N.B.* the universal potential is trained with Materials Project database which the scale of energy and forces are based on the VASP software.

Since $(AlF_3)_n$ clusters with $n = 1$ and $n = 2$ each have only one configuration in the PES, and clusters with $n = 3$ to $n = 5$ have relatively fewer LM, we proceeded to $n = 6$ to train the machine learning interatomic potential (ML-IP). In the PES for $n = 6$, there are over a thousand LMs, most of which have random-like configurations whose vibrational modes cover the equilibrium interatomic distances. Training all LMs is computationally expensive; therefore, we decided to select 10 LMs to train the ML-IP. We trained two different sets of ML-IPs: one using (a) the top 10 LMs' vibrational mode images as the training set, and the other using (b) randomly chosen 10 LMs' vibrational modes. Each of the ML-IPs was evaluated by comparing it with the reference PBEsol data. Figure 52 shows the ML-IP-evaluated total energies of the LMs and atomic forces compared to the PBEsol data.

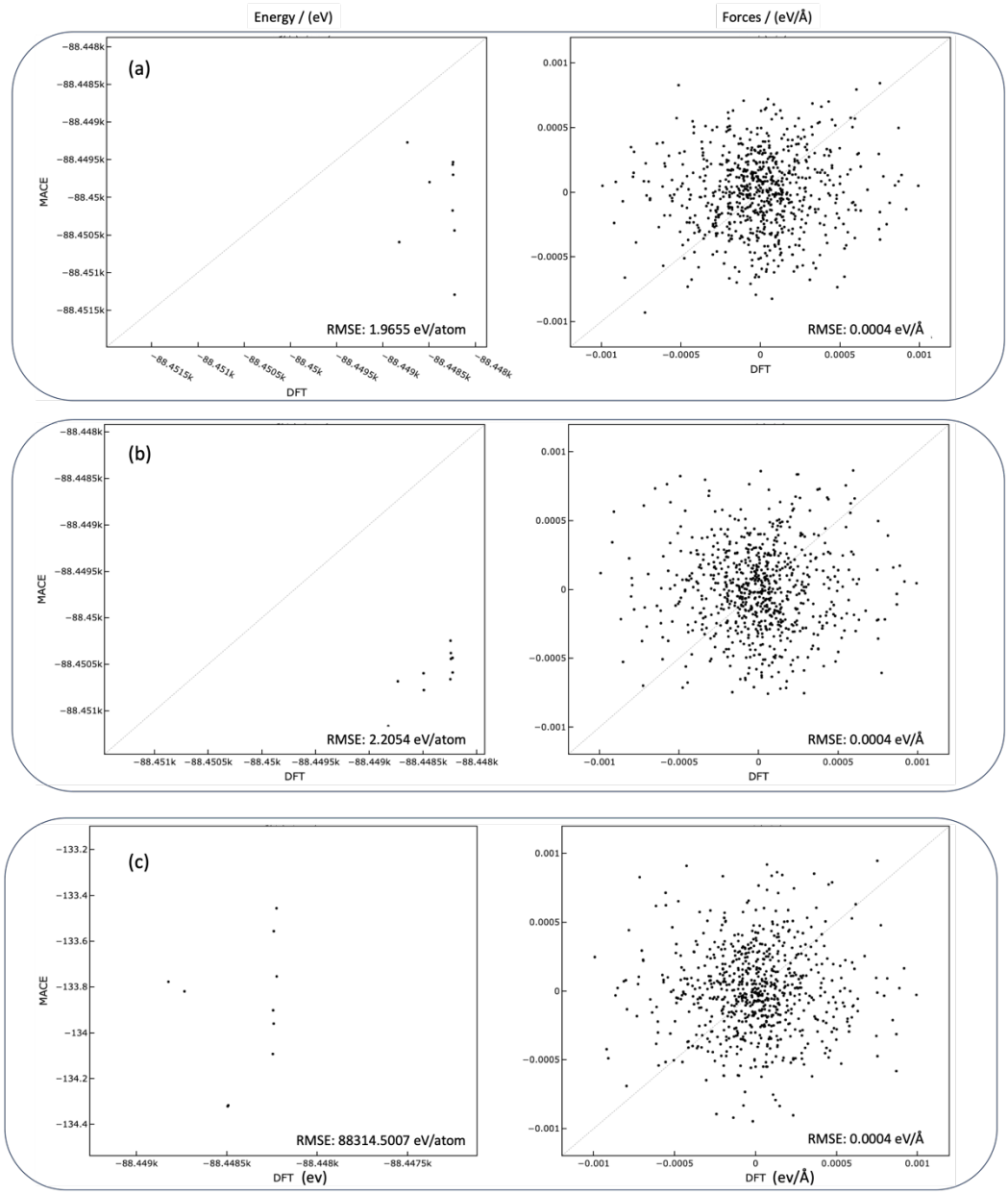

Figure **52**. The total energy and atomic forces of 10 randomly selected LMs from $(AlF_3)_n$ where n = 6, optimized using the two specifically trained ML-IPs (a and b) and the universal MACE ML-IP (c). The top ML-IP is trained using vibrational mode images from the top 10 local minima, while the middle ML-IP is trained with images from 10 randomly chosen local minima. The evaluated total energy and atomic forces of these LM are compared with the reference PBEsol data. *N.B.* universal potential was trained on Materials Project database which the total energy and forces are obtained from the VASP software.

As this is a testing process the 10 different ML-IPs were trained for each of LM. After training, this collection of ML-IPs—referred to as an ensemble—can be utilized simultaneously through the "mace" library package in Python. The Atomic Simulation Environment[155] (ASE)'s optimizer was then used to optimize the PBEsol LMs. The last plots obtained by using the recently developed and published universal potential. These randomly chosen 10 LMs of n = 6 for the evaluation are optimized using the corresponding ML-IP.

Interestingly, all three ML-IPs predicted the atomic forces of LMs with high accuracy of 0.0004 eV/Å. Note that the range of the x-axis for the force plot is significantly small which makes the points are spread in spherical shape, but it is notably close to the reference line. However, the predicted total energies of the LMs exhibited a RMSE greater than 1.9 eV. It is important to note that the universal potential was trained using data evaluated by VASP, resulting in a different energy scale of total energy and making the RMSE less reliable for accuracy assessment. Moreover, here we are going to focus on how the ML-IPs accurately predict the LM energy ranking and/or atomic configurations.

In terms of RMSE, the (a) ML-IP exhibited a slightly lower value than (b) by 0.24 eV per atom. However, the (b) ML-IP predicted the energy ranking of n = 6 LMs more accurately, displaying a relatively clearer linear trend in the total energy of the system compared to the (a) ML-IP, which showed a columnar trend similar to that predicted by the (c) universal potential. Remarkably, the universal potential ranked the top three PBEsol LMs as the 7th, 6th, and 1st (ground state) LM, respectively. Conversely, the second ML-IP (b) correctly predicted the top three PBEsol LMs within the top three ML-IP LMs. This result was anticipated as the universal was not trained on nanoclusters only crystal structures in Materials Project datbase.

To further verify whether the training data from the randomly chosen 10 LMs led to better performance in predicting accurate energy rankings of clusters, the same evaluation was conducted for n = 7, as depicted in Figure 52.

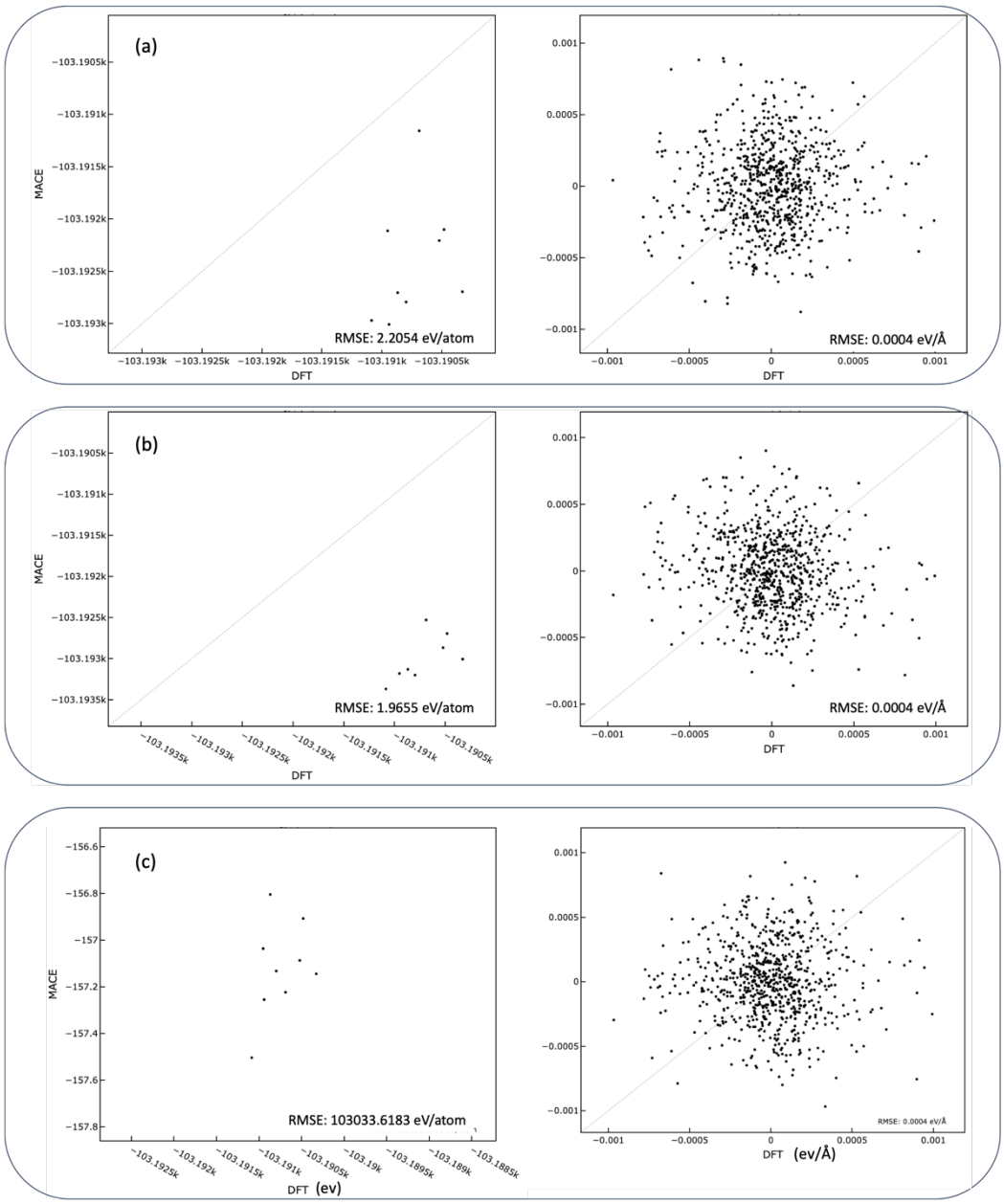

Figure **53**. The total energy and atomic forces of 10 randomly selected LMs from $(AlF_3)_n$ where n = 7, optimized using the two specifically trained ML-IPs (a and b) and the universal MACE ML-IP (c). The top ML-IP is trained using vibrational mode images from the top 10 local minima, while the middle ML-IP is trained with images from 10 randomly chosen local minima. The evaluated total energy and atomic forces of these LM are compared with the reference PBEsol data. *N.B.*

universal potential was trained on Materials Project database which the total energy and forces are obtained from the VASP software.

The n = 7 ML-IP shows the same performance as the n = 6's ML-IP in predicting atomic forces with 0.0004 eV/Å RMSE. Again, the ML-IP trained with the randomly chosen 10 LM data shows the better performance in predicting the accurate cluster energy ranking compared with other two. The (a) ML-IP failed to identify GM as the GM, whereas the other ML-IPs, (b) and the universal, successfully did. The universal predicted the 2nd LM correctly, but the 3rd LM as 7th LM. On the other hand, (b) could not predict the exact energy ranking, but it predicted with the least energy ranking changes after the ML-IP optimisation compared to (a) and (c). If (b) could predicts the energy ranking with the small energy ranking changes compared to the universal using the (b) for global optimisation would have more advantage as it has more versatile trend in predicting accurate energy ranking. In other words, after the global optimization using a ML-IP we only need to refine the set of top n many LMs, but for the global optimisations using the universal potential has to refine the many set of lower LM.

Interestingly, the n = 6 and n = 7 trained with the (b) predicted the cluster energy not far from to each other's cluster energy, short distance between the predicted data point which represents that (b) able to predict accurately. Moreover, as we anticipated the ML-IP trained with the (b) randomly chosen 10 LM data covers the broader range of PES shows the slightly better performance in predicting accurate energy ranking of clusters than the universal potential and the ML-IP trained with the top 10 LMs.

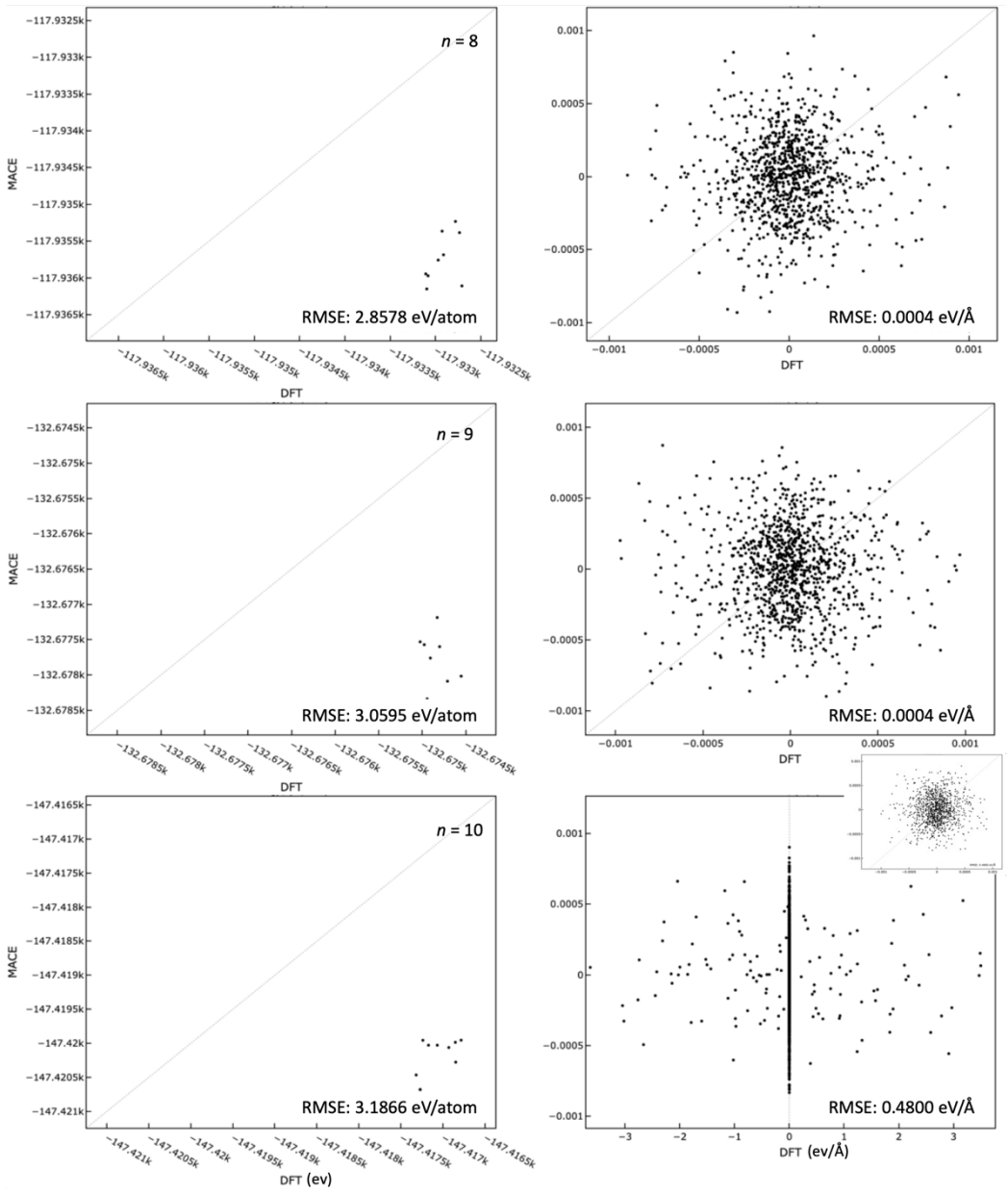

Figure **54**. The total energy and atomic forces of 10 randomly selected LM from $(AlF_3)_n$ for n = 8, 9, and 10, evaluated using the ML-IPs trained with vibrational mode images from the randomly chosen 10 LMs within the matching size cluster's PES. The ML-IPs' predictions for total energy and atomic forces of these LMs are compared with reference PBEsol data. A smaller panel within the atomic force plot for n = 10 provides a zoomed view with the same x-axis range as those for n = 8 and 9's atomic force plot, although not all data points are visible in this smaller panel.

Similarly to the training conducted for $n = 7$, we trained the ML-IPs for $n = 8$, $n = 9$, and $n = 10$ as depicted in panel (b). The optimized energies and atomic forces predicted by the ML-IP were then compared against the reference data obtained from PBEsol calculations. However, as the cluster size increases, the ML-IP struggles to accurately rank the energies and predict the atomic forces of the clusters. This decline in accuracy is attributed to the exponential expansion of the PES with larger cluster sizes, making it challenging for the ML-IP trained on only 10 randomly selected LMs vibrational modes to capture sufficient information about the PES during the training process.

Despite this limitation, the ML-IP remains effective for structure prediction in smaller clusters, as demonstrated by the results for $n = 6$ and $n = 7$, as well as for simpler 1:1 stoichiometric systems, which possess less complex PES compared to 1:3 stoichiometric systems. To address the challenges posed by larger cluster sizes, we plan to implement an active learning approach. Active learning is a ML technique where the model selectively incorporates additional data from regions where it is less confident or has made inaccurate predictions. In this context, as structure prediction progresses to larger cluster sizes, the ML-IP will be iteratively trained with additional data from bigger clusters or from structures where initial predictions were inaccurate. This targeted data acquisition enhances the model's performance by focusing on the most informative parts of the PES, thereby improving its accuracy.

Consequently, the actively learned ML-IP is expected to achieve higher accuracy compared to an ML-IP trained solely on a limited set of cluster sizes. This iterative learning process allows the ML-IP to better generalize across the expanded PES of larger clusters, enhancing its reliability and predictive capability. Figure 52 illustrates the ML-IP-evaluated total energies of the LMs and atomic forces compared to the PBEsol reference data, highlighting the effectiveness of the active learning strategy in improving the model's performance for larger and more complex cluster sizes.

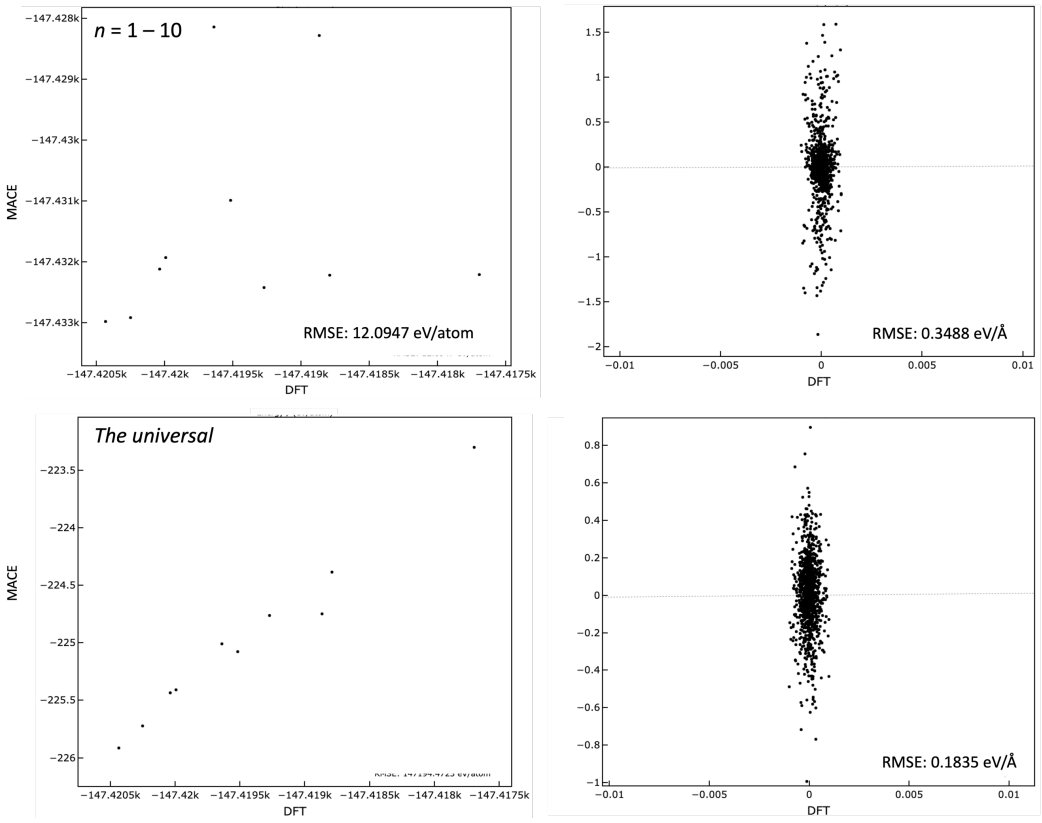

Figure **55**. The 10 randomly generated structures of n = 10 were optimized using the ML-IPs, trained on 10 randomly selected LM from each cluster size's PES ranging from n = 1 to 10 (top row), along with the universal potential (bottom row). The total energy and atomic forces of these optimized structures are compared with reference data from PBEsol optimizations.

This time we performed global optimisation with 10 sample points for n = 10. The randomly generated geometries undergo optimization using the ML-IP, which was trained on randomly selected 10 LM of n = 1 – 10 for each cluster size and the universal potential. The total energy and atomic forces of the optimized structures are then compared with the reference another 10 randomly selected PBEsol LMs.

From the figure 55, it looks like the universal potential significantly performed better in predicting correct energy rankings as it shows the linear trend in the data points. Even the predicted atomic forces for the universal potential calculated values have less RMSE compared to the cluster trained ML-IP. However, the universal potential optimised structures are not the true n = 10 which

the structures are shown in the figure 56. The universal potential optimised randomly generated structures are fragmented as it showed in the figure:

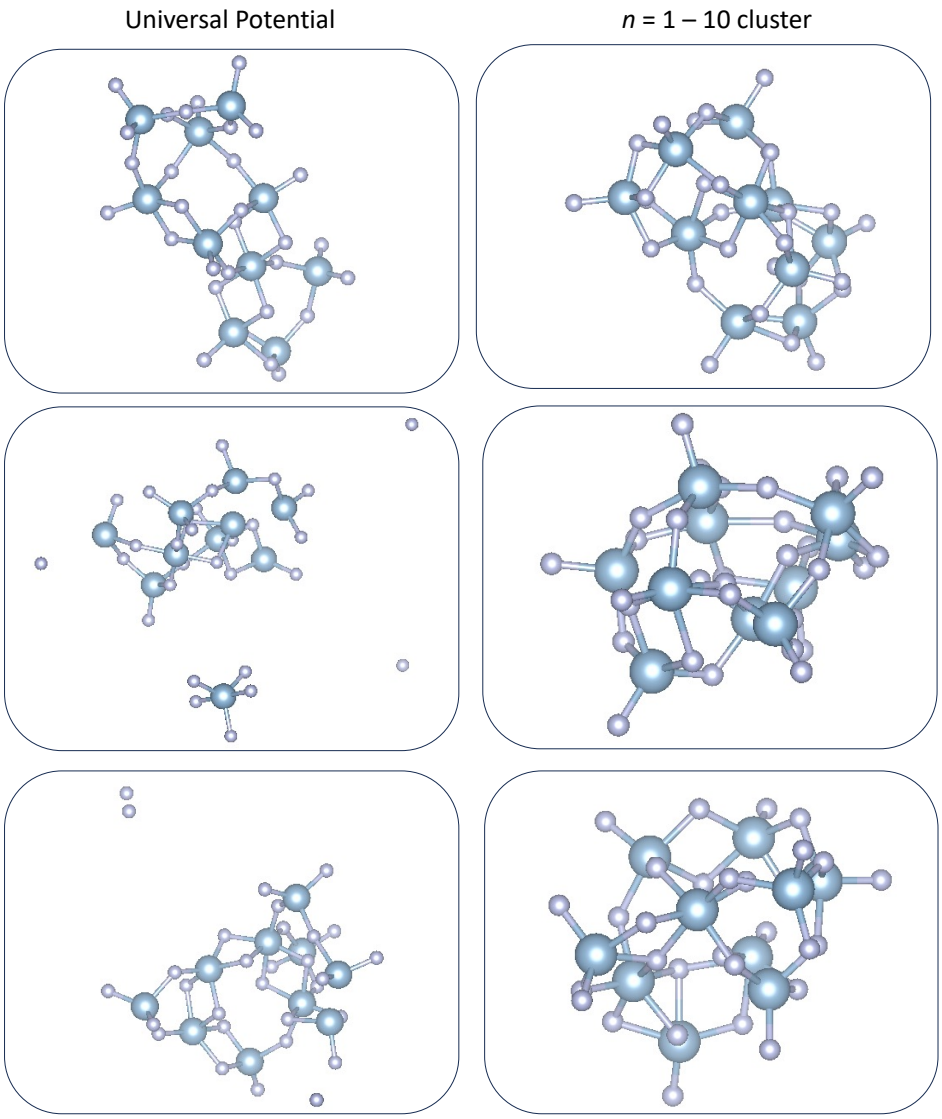

Figure **56**. (LHS) The universal potential and (RHS) the ML-IP trained with 10 randomly chosen clusters for each of n = 1 – 10 optimised n = 10's randomly generated structures.

From the visualised structures in figure 56 which obviously shows that the universal potential cannot perform well for the MCDQ global optimisations as the universal potential optimised structures are fragmented. On the other hand, the ML-IP trained with 10 randomly chosen clusters per cluster sizes (n = 1 – 10) could provide n = 10 clusters. More importantly the clusters have the same

configurations as the PBEsol optimised structures. Note that both the universal potential and the cluster trained ML-IP which used for the figure 56 was trained with 6 Å radial cutoff. The radial cutoff cannot explain the fragmented structures in the universal potential optimised structures because the cluster trained ML-IP did not predict the fragmented features.

More research is needed, but the initial comparison suggests that using images of vibrational modes as training data could be a promising approach for developing ML-IP for nanoclusters. This is based on the comparison between the universal potential trained on 1.6 million crystal structures and the vibrational modes from the cluster images trained ML-IP. Like other ML-IP training methods active learning is inevitable to increase the accuracy, transferability, robustness.

We found that using the randomly chosen 10 LM vibration images shows better accuracy compared to the lowest 10 LM vibration images as training data as the randomly chosen 10 LM can provide the different region of PES for training. The universal potential was compared with the ML-IP trained on the randomly chosen 10 LM. As we anticipated the universal potential shows the poorer performance in predicting energy ranking of clusters as the universal potential was trained with any clusters, only crystals. At last we performed MCDQ global optimisations with 10 sample points which was enough to realise that cluster trained ML-IP performs better than the universal potential by visualising the optimised clusters: the universal potential predicted fragmented structure compared to the cluster trained ML-IP predicted the geometry which is same as the PBEsol. Consequently, ML-IP can be trained for complex systems like Ti_3N_4 to predict the cluster structures which could relatively effectively trained or fitted compared to the IP.

7. Conclusion

In this thesis, we have explored the nanocluster structure prediction through a combination of global optimization techniques, interatomic potentials, and ML approaches. The primary focus has been on understanding the behaviour and properties of AlF_3, AlH_3, and $Au_{25}(Cys)_{18}$. This study has aimed to bridge the gap between theoretical predictions and experimental observations, offering new insights into the structural and chemical characteristics of these nanoclusters.

Our investigation into AlF_3 nanoclusters, ranging from $(AlF_3)_1$ to $(AlF_3)_{11}$, utilized MCDQ and GA approaches to locate LM on the PES defined by interatomic potentials. The Born-Mayer potential parameters for Al and F ions were fine-tuned to reproduce the alpha bulk phase of AlF_3. The results revealed that the $(AlF_3)_6$ nanocluster is likely to be the most stable and prevalent in nature, highlighting the presence of octahedral corner-sharing SBUs within the cluster. These findings are consistent with bulk AlF_3 phases, indicating that even at the nanoscale, certain structural motifs are preserved.

AlH_3 nanoclusters were analysed using a data mining approach, leveraging the predicted configurations of AlF_3 clusters. The study found that AlH_3 clusters exhibit analogous configurations and comparable stability to their AlF_3 counterparts. This suggests that the primary and secondary building units play a significant role in defining the structural stability of these nanoclusters. The development of a method to determine coordination numbers within the predicted nanoclusters was a key aspect of this analysis, although it presented challenges in some cases.

The structural properties of thiolated gold clusters, $Au_{25}(Cys)_{18}$, were examined through global optimization techniques. The study involved modelling the atomic structure of naked Au_{25} and subsequently adding *L*-cysteine ligands. The electronic structure obtained is compared with experimental observation. The resulting models provided insights into the structural evolution and stability of these clusters, contributing to a deeper understanding of their potential applications in catalysis and other fields.

One of the major chapter of this thesis was dedicated to the development and application of ML-IP, particularly the MACE model. This model was trained using images of stable nanoclusters perturbed along vibrational modes to achieve near DFT precision with the computational efficiency of IP calculations. The MACE model effectively captured the complex interactions within atomic systems, ensuring rotational and translational equivariance while accurately predicting the energy and properties of various nanoclusters.

The key findings from this research include the identification of stable nanocluster configurations, highlighting the importance of building units in determining structural stability, and the successful application of ML models to enhance computational efficiency. The effort in finding accurate coordination numbers to understand the structural stability in a quantitative approach will help to understand the relationship between nanocluster structure, stability, and its properties.

Future work could explore the integration of more advanced ML techniques and the inclusion of non-harmonic, mode-coupling effects to further enhance the accuracy of ML-IPs. Additionally, experimental validation of the predicted nanocluster structures and properties will be crucial for translating these theoretical insights into practical applications.

In conclusion, this thesis has provided a comprehensive framework for the prediction and analysis of nanocluster structures, leveraging a blend of traditional computational methods and cutting-edge machine learning techniques. The insights gained from this research contribute to the broader understanding of nanoscale materials and pave the way for future advances in the field.

Acronyms

0D - 0-dimensional

1D - 1-dimensional

2D - 2-dimensional

3D - 3-dimensional

(M)EAM - (Modified) Embedded Atom Method

BFGS - Broyden-Fletcher-Goldfarb-Shanno algorithm

COMB - Charge Optimised Many Body

DFT - Density Functional Theory

DOS - Density Of State

EA - Evolutionary Algorithm (= GA)

FHI-aims - Fritz Haber Institute ab initio molecular simulations

FWHM - Full-Width-Half-Maximum

GA - Genetic Algorithm

GAP - Gaussian Approximation Potential

GM - Global Minimum (or Minima)

GNN - Graph Neural Network

GPR - Gaussian Progress Regression

GULP - General Utility Lattice Program

gnorm - Gradient Norm

HR-TEM – High Resolution Transmission Electron Microscopy

KLMC - Knowledge-Led Master Code

LDA - Local Density Approximation

LJ - Lennard-Jones Potential

LM - Local Minimum (or minima)

MACE - Message Passing (or Multi) Atomic Cluster Expansion

MCDQ - Monte Carlo Deterministic Quenching

ML - Machine Learning

MS – Mass Spectrometry

NMR - Nuclear Magnetic Resonance Spectroscopy

NN - Neural Network

QEq - Charge Equilibrium

RFO - Rational Function Optimisation

RMS - Root Mean Squared

SASA - Solvent Accessible Surface Area

SBU - Secondary Building Unit

STEM - Scanning Transmission Electron Microscopy

STM - Scanning Tunnelling Microscopy

TEM - Transmission Electron Microscopy

vdW - van der Waals (force, interaction)

List of Figures

List of Tables

Bibliography

1 K. S. Novoselov, A. K. Geim, S. V Morozov, D. Jiang, Y. Zhang, S. V Dubonos, I. V Grigorieva and A. A. Firsov, *Electric Field Effect in Atomically Thin Carbon Films*, Kluwer, 2000, vol. 404.

2 S. Iijima and T. Ichihashi, *Nature*, 1993, **363**, 603–605.

3 D. S. Bethune, C. H. Kiang, M. S. de Vries, G. Gorman, R. Savoy, J. Vazquez and R. Beyers, *Nature*, 1993, **363**, 605–607.

4 D.-G. Kang, University College London, 2019.

5 S. M. Woodley and S. T. Bromley, in *Frontiers of Nanoscience*, Elsevier Ltd, 2018, vol. 12, pp. 1–54.

6 A. Iwaszuk, P. A. Mulheran and M. Nolan, *J Mater Chem A Mater*, 2013, **1**, 2515–2525.

7 X. Huang, P. K. Jain, I. H. El-Sayed and M. A. El-Sayed, 2007, preprint, DOI: 10.2217/17435889.2.5.681.

8 X. Huang and M. A. El-Sayed, 2010, preprint, DOI: 10.1016/j.jare.2010.02.002.

9 S. M. Woodley, *Journal of Physical Chemistry C*, 2013, **117**, 24003–24014.

10 A. A. Al-Sunaidi, A. A. Sokol, C. R. A. Catlow and S. M. Woodley, *Journal of Physical Chemistry C*, 2008, **112**, 18860–18875.

11 S. M. Woodley, S. Hamad, J. A. Mejías and C. R. A. Catlow, *J Mater Chem*, 2006, **16**, 1927–1933.

12 S. G. E. T. Escher, T. Lazauskas, M. A. Zwijnenburg and S. M. Woodley, *Comput Theor Chem*, 2017, **1107**, 74–81.

13 I. Venditti, *MDPI AG*, 2017, preprint, DOI: 10.3390/ma10020097.

14 M. Notarianni, K. Vernon, A. Chou, M. Aljada, J. Liu and N. Motta, *Solar Energy*, 2014, **106**, 23–37.

15 Z. A. Lewicka, W. W. Yu, B. L. Oliva, E. Q. Contreras and V. L. Colvin, *J Photochem Photobiol A Chem*, 2013, **263**, 24–33.

16 L. Zhu and W. Zeng, *Elsevier B.V.*, 2017, preprint, DOI: 10.1016/j.sna.2017.10.021.

17 L. Fan, Z. Cheng, J. Du and P. Delir Kheirollahi Nezhad, *Monatsh Chem*, 2022, **153**, 321–329.

18 S. Haviar, J. Čapek, Š. Batková, N. Kumar, F. Dvořák, T. Duchoň, M. Fialová and P. Zeman, *Int J Hydrogen Energy*, 2018, **43**, 22756–22764.

19 J. H. Byun, J. R. Ahn, W. H. Choi, P. G. Kang and H. W. Yeom, *Phys Rev B Condens Matter Mater Phys*, DOI:10.1103/PhysRevB.78.205314.

20 C. Liu, I. Matsuda, M. D'Angelo, S. Hasegawa, J. Okabayashi, S. Toyoda and M. Oshima, *Phys Rev B Condens Matter Mater Phys*, DOI:10.1103/PhysRevB.74.235420.

21 A. Uzun, V. Ortalan, Y. Hao, N. D. Browning and B. C. Gates, *ACS Nano*, 2009, **3**, 3691–3695.

22 S. Helveg and P. L. Hansen, in *Catalysis Today*, 2006, vol. 111, pp. 68–73.

23 A. Uzun, V. Ortalan, Y. Hao, N. D. Browning and B. C. Gates, *ACS Nano*, 2009, **3**, 3691–3695.

24 C. Tian, C. Feng and Q. Wang, *Elsevier B.V.*, 2021, preprint, DOI: 10.1016/j.scitotenv.2020.142154.

25 L. A. Angel, L. T. Majors, A. C. Dharmaratne and A. Dass, *ACS Nano*, 2010, **4**, 4691–4700.

26 P. Grammatikopoulos, S. Steinhauer, J. Vernieres, V. Singh and M. Sowwan, *Taylor and Francis Ltd.*, 2016, preprint, DOI: 10.1080/23746149.2016.1142829.

27 X. Kang, H. Chong and M. Zhu, *Royal Society of Chemistry*, 2018, preprint, DOI: 10.1039/c8nr02973c.

28 X. Ren, M. I. Hussain, Y. Chang and C. Ge, *Int J Mol Sci*, 2023, **24**, 17239.

29 J. Prasek, J. Drbohlavova, J. Chomoucka, J. Hubalek, O. Jasek, V. Adam and R. Kizek, *J Mater Chem*, 2011, **21**, 15872.

30 C. T. Gibson, *Applied Sciences*, 2020, **10**, 5575.

31 Helmholtz-Zentrum Dresden-Rossendorf, *Phys org*, 2022.

32 L. Dong, R. R. S. Gari, Z. Li, M. M. Craig and S. Hou, *Carbon N Y*, 2010, **48**, 781–787.

33 Y. Zhou, Y. Hu, S. Lu, D. Wang, D. Ma, X. Gong and Q. Yue, *Adv Funct Mater*, 2024, **34**, 36, DOI:10.1002/adfm.202401195.

34 A. V. Vorontsov and H. Valdés, *J Photochem Photobiol A Chem*, 2019, **379**, 39–46.

35 J. R. Do Nascimento, M. R. D'oliveira, A. G. Veiga, C. A. Chagas and M. Schmal, *ACS Omega*, 2020, **5**, 25568–25581.

36 Q. Gu, G. Wen, Y. Ding, K. H. Wu, C. Chen and D. Su, *Green Chemistry*, 2017, **19**, 1175–1181.

37 Y. Gao, D. Ma, C. Wang, J. Guan and X. Bao, *Chemical Communications*, 2011, **47**, 2432–2434.

38 A. C. T. Van Duin, S. Dasgupta, F. Lorant and W. A. Goddard, *Journal of Physical Chemistry A*, 2001, **105**, 9396–9409.

39 T. Liang, T. R. Shan, Y. T. Cheng, B. D. Devine, M. Noordhoek, Y. Li, Z. Lu, S. R. Phillpot and S. B. Sinnott, *Elsevier Ltd*, 2013, preprint, DOI: 10.1016/j.mser.2013.07.001.

40 J. E. Lennard-Jones, *Transactions of the Faraday Society*, 1929, **25**, 668.

41 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J Comput Chem*, 2004, **25**, 1157–1174.

42 T. Lu and F. Chen, *J Mol Model*, 2013, **19**, 5387–5395.

43 R. A. Buckingham, *Proc R Soc Lond A Math Phys Sci*, 1938, **168**, 264–283.

44 C. L. Liu, J. M. Cohen, J. B. Adams and A. F. Voter, *Surf Sci*, 1991, **253**, 334–344.

45 M. S. Daw and M. I. Baskes, *Phys Rev B*, 1984, **29**, 6443–6453.

46 M. I. Baskes and R. A. Johnson, *Model Simul Mat Sci Eng*, 1994, **2**, 147–163.

47 M. I. Baskes, *Phys Rev B*, 1992, **46**, 2727–2742.

48 B.-J. Lee and M. I. Baskes, *Phys Rev B*, 2000, **62**, 8564–8567.

49 Y. T. Cheng, T. R. Shan, B. Devine, D. Lee, T. Liang, B. B. Hinojosa, S. R. Phillpot, A. Asthagiri and S. B. Sinnott, *Surf Sci*, 2012, **606**, 1280–1288.

50 R. Slapikas, I. Dabo and S. B. Sinnott, *J Chem Phys*, 2020, **152**, 224702.

51 I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, .

52 D. Ongari, P. G. Boyd, O. Kadioglu, A. K. MacE, S. Keskin and B. Smit, *J Chem Theory Comput*, 2019, **15**, 382–401.

53 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.

54 N. Fedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev and S. Tretiak, *Nat Rev Chem*, 2022, **6**, 653–672.

55 B. G. Dick and A. W. Overhauser, *Physical Review*, 1958, **112**, 90–103.

56 D. W. Brenner, *American Physical Society*, 1989, preprint, DOI: 10.1103/PhysRevLett.63.1022.

57 B.-J. Lee, M. I. Baskes, H. Kim and Y. Koo Cho, *Phys Rev B*, 2001, **64**, 184102.

58 Y.-T. Cheng, T.-R. Shan, B. Devine, D. Lee, T. Liang, B. B. Hinojosa, S. R. Phillpot, A. Asthagiri and S. B. Sinnott, *Surf Sci*, 2012, **606**, 1280–1288.

59 A. K. Rappe and W. A. G. Iii, *Charge Equilibration for Molecular Dynamics Simulations*, 1991, vol. 95.

60 K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller and E. K. U. Gross, *Phys Rev B*, 2014, **89**, 205118.

61 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys Rev Lett*, 2012, **108**, 058301.

62 F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *J Chem Theory Comput*, 2017, **13**, 5255–5264.

63 N. Artrith and A. Urban, *Comput Mater Sci*, 2016, **114**, 135–150.

64 A. P. Bartõk and G. Csányi, *John Wiley and Sons Inc.*, 2015, preprint, DOI: 10.1002/qua.24927.

65 Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang and P. S. Yu, *IEEE Trans Neural Netw Learn Syst*, 2020, **32**, 4–24.

66 V. L. Deringer, G. Csányi and D. M. Proserpio, *ChemPhysChem*, 2017, **18**, 873–877.

67 Y. Mishin, *Acta Mater*, 2021, **214**, 116980.

68 M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, *IEEE Signal Process Mag*, 2017, **34**, 18–42.

69 X.-M. Zhang, L. Liang, L. Liu and M.-J. Tang, *Front Genet*, *Frontiers in Genetics*, 2021, 12, DOI:10.3389/fgene.2021.690049.

70 N. Fedik, R. Zubatyuk, M. Kulichenko, N. Lubbers, J. S. Smith, B. Nebgen, R. Messerly, Y. W. Li, A. I. Boldyrev, K. Barros, O. Isayev and S. Tretiak, *Nat Rev Chem*, 2022, **6**, 916–916.

71 C. Chen and S. P. Ong, *Nat Comput Sci*, 2022, **2**, 718–728.

72 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys Rev Lett*, 2008, **100**, 136406.

73 E. Schrödinger, *Physical Review*, 1926, **28**, 1049–1070.

74 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput Phys Commun*, 2009, **180**, 2175–2196.

75 V. Sahni, K.-P. Bohnen and M. K. Harbola, *Analysis of the local-density approximation of density-functional theory*, 1988, vol. 37.

76 J. P. Perdew, K. Burke and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, 1996.

77 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys Rev Lett*, DOI:10.1103/PhysRevLett.100.136406.

78 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys Rev Lett*, DOI:10.1103/PhysRevLett.100.136406.

79 J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou and K. Burke, *Phys Rev Lett*, DOI:10.1103/PhysRevLett.100.136406.

80 M. J. D. Powell, *Comput J*, 1964, **7**, 155–162.

81 J. R. Shewchuk, *An Introduction to the Conjugate Gradient Method Without the Agonizing Pain*, 1994.

82 T. J. Ypma, *Historical Development of the Newton-Raphson Method*, 1995, vol. 37.

83 C. G. Broyden, *IMA J Appl Math*, 1970, **6**, 76–90.

84 R. Fletcher, *Comput J*, 1970, **13**, 317–322.

85 D. Goldfarb, *Math Comput*, 1970, **24**, 23–26.

86 D. F. Shanno, *Math Comput*, 1970, **24**, 647–656.

87 A. Banerjee, N. Adams, J. Simons and R. Shepard, *Search for Statlonary Polnts on Surfaces Introduction Finding stationary points (minima, maxima, and saddle points)*, .

88 C. R. A. Catlow, S. T. Bromley, S. Hamad, M. Mora-Fonz, A. A. Sokol and S. M. Woodley, *Physical Chemistry Chemical Physics*, 2010, **12**, 786–811.

89 T. Lazauskas, A. A. Sokol and S. M. Woodley, *Nanoscale*, 2017, **9**, 3850–3864.

90 S. M. Woodley and C. R. A. Catlow, *Comput Mater Sci*, 2009, **45**, 84–95.

91 S. M. Woodley and C. R. A. Catlow, *Comput Mater Sci*, 2009, **45**, 84–95.

92 S. M. Woodley, P. D. Battle, J. D. Galec, C. Richard and A. Catlowa, *The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation*, .

93 T. Lazauskas, A. A. Sokol, J. Buckeridge, C. R. A. Catlow, S. G. E. T. Escher, M. R. Farrow, D. Mora-Fonz, V. W. Blum, T. M. Phaahla, H. R. Chauke, P. E. Ngoepe and S. M. Woodley, *Physical Chemistry Chemical Physics*, 2018, **20**, 13962–13973.

94 B. D. McKay and A. Piperno, *J Symb Comput*, 2014, **60**, 94–112.

95 S. M. Woodley, Private communivation, 2024.

96 T. Krahl and E. Kemnitz, *Royal Society of Chemistry*, 2017, preprint, DOI: 10.1039/c6cy02369j.

97 E. Kemnitz, U. Groß, S. Rüdiger and C. S. Shekar, *Angewandte Chemie - International Edition*, 2003, **42**, 4251–4254.

98 S. M. Woodley, G. M. Day and R. Catlow, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2020, **378**, 20190600.

99 S. M. Woodley, S. Hamad, J. A. Mejías and C. R. A. Catlow, *J Mater Chem*, 2006, **16**, 1927–1933.

100 WASP@N, https://hive.chem.ucl.ac.uk/, (accessed 22 June 2020).

101 M. R. Farrow, Y. Chow and S. M. Woodley, *Physical Chemistry Chemical Physics*, 2014, **16**, 21119–21134.

102 M. E. G. Valerio, R. A. Jackson and J. F. De Lima, *Journal of Physics Condensed Matter*, 2000, **12**, 7727–7734.

103 J. D. Gale, *Z Kristallogr Cryst Mater*, 2005, **220**, 552–554.

104 J. D. Gale, *Journal of the Chemical Society - Faraday Transactions*, 1997, **93**, 629–637.

105 J. D. Gale and A. L. Rohl, *Mol Simul*, 2003, **29**, 291–341.

106 C. R. Morelock, J. C. Hancock and A. P. Wilkinson, *J Solid State Chem*, 2014, **219**, 143–147.

107 R. A. Jackson, M. E. G. Valerio and J. F. de Lima, *Journal of Physics: Condensed Matter*, 2001, **13**, 2147–2154.

108 M. E. G. Valerio, R. A. Jackson and J. F. de Lima, *Journal of Physics: Condensed Matter*, 1998, **10**, 3353–3358.

109 M. E. G. Valerio, J. F. De Lima and R. A. Jackson, *Radiation Effects and Defects in Solids*, 1999, **151**, 249–254.

110 S. G. E. T. Escher, T. Lazauskas, M. A. Zwijnenburg and S. M. Woodley, *Comput Theor Chem*, 2017, **1107**, 74–81.

111 S. M. Woodley, Springer, Berlin, Heidelberg, 2004, pp. 95–132.

112 S. B. Woodley, A. A. Sokol, C. R. A. Catlow, A. A. Al-Sunaidi and S. M. Woodley, *Journal of Physical Chemistry C*, 2013, **117**, 27127–27145.

113 R. D. Shannon, *Acta Crystallographica Section A*, 1976, **32**, 751–767.

114 O. Lamiel-Garcia, A. Cuko, M. Calatayud, F. Illas and S. T. Bromley, *Nanoscale*, 2017, **9**, 1049–1058.

115 S. Chaudhuri, P. Chupas, B. J. Morgan, P. A. Madden and C. P. Grey, *Physical Chemistry Chemical Physics*, 2006, **8**, 5045–5055.

116 B. Lee and F. M. Richards, *J Mol Biol*, 1971, **55**, 379-IN4.

117 M. L. Connolly, *J Appl Crystallogr*, 1983, **16**, 548–558.

118 Schrödinger LLC, 2015, preprint, 1.8, https://pymol.org/#page-top.

119 A. Shrake and J. A. Rupley, *J Mol Biol*, 1973, **79**, 351–371.

120 G. V Lewis and C. R. A. Catlow, *Journal of Physics C: Solid State Physics*, 1985, **18**, 1149–1161.

121 H. Kreek and W. J. Meath, *J Chem Phys*, 1969, **50**, 2289–2302.

122 N. C. Pyper, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 1986, **320**, 107–158.

123 J. C. Slater and J. G. Kirkwood, *Physical Review*, 1931, **37**, 682–697.

124 W. Su, F. Zhao, L. Ma, R. Tang, Y. Dong, G. Kong, Y. Zhang, S. Niu, G. Tang, Y. Wang, A. Pang, W. Li and L. Wei, *Materials*, 2021, **14**, 2898.

125 L. Pauling, *J Am Chem Soc*, 1927, **49**, 765–790.

126 S. Hati, B. Datta and D. Datta, *J Phys Chem*, 1996, **100**, 19808–19811.

127 Y. Cao, V. Fung, Q. Yao, T. Chen, S. Zang, D. Jiang and J. Xie, *Nat Commun*, 2020, **11**, 5498.

128 M. Zhu, C. M. Aikens, F. J. Hollander, G. C. Schatz and R. Jin, *J Am Chem Soc*, 2008, **130**, 5883–5885.

129 Q. Yao, T. Chen, X. Yuan and J. Xie, *Acc Chem Res*, 2018, **51**, 1338–1348.

130 W. Kurashige, Y. Niihori, S. Sharma and Y. Negishi, *Coord Chem Rev*, 2016, **320–321**, 238–250.

131 S. Hossain, L. V. Nair, J. Inoue, Y. Koyama, W. Kurashige and Y. Negishi, in *Ligand*, InTech, 2018.

132 H. Hassan, P. Sharma, Mohd. R. Hasan, S. Singh, D. Thakur and J. Narang, *Mater Sci Energy Technol*, 2022, **5**, 375–390.

133 S. Link and M. A. El-Sayed, *J Phys Chem B*, 1999, **103**, 8410–8426.

134 L. De Broglie, *Ann Phys (Paris)*, 1925, **10**, 22–128.

135 Y. Wang and T. Bürgi, *Nanoscale Adv*, 2021, **3**, 2710–2727.

136 J. A. Adekoya, K. O. Ogunniran, T. O. Siyanbola, E. O. Dare and N. Revaprasadu, in *Noble and Precious Metals - Properties, Nanoscale Effects and Applications*, InTech, 2018.

137 R. Jin and T. Higaki, *Commun Chem*, 2021, **4**, 28.

138 M. Zhou, C. Zeng, Q. Li, T. Higaki and R. Jin, *Nanomaterials*, 2019, **9**, 933.

139 X. Kang, Y. Li, M. Zhu and R. Jin, *Chem Soc Rev*, 2020, **49**, 6443–6514.

140 S. A. Bansal, V. Kumar, J. Karimi, A. P. Singh and S. Kumar, *Nanoscale Adv*, 2020, **2**, 3764–3787.

141 J. A. Adekoya, K. O. Ogunniran, T. O. Siyanbola, E. O. Dare and N. Revaprasadu, in *Noble and Precious Metals - Properties, Nanoscale Effects and Applications*, InTech, 2018.

142 M. Zhou, C. Zeng, Q. Li, T. Higaki and R. Jin, *Nanomaterials*, 2019, **9**, 933.

143 T. L. Christiansen, S. R. Cooper and K. M. Ø. Jensen, *Nanoscale Adv*, 2020, **2**, 2234–2254.

144 D. J. Wales and J. P. K. Doye, *Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms*, 1997.

145 D. J. Wales and H. A. Scheraga, *Science (1979)*, 1999, **285**, 1368–1372.

146 M. Meunier and S. Robertson, *Mol Simul*, 2021, **47**, 537–539.

147 Y. Wang and T. Bürgi, *Nanoscale Adv*, 2021, **3**, 2710–2727.

148 A. C. Ihrig, J. Wieferink, I. Y. Zhang, M. Ropo, X. Ren, P. Rinke, M. Scheffler and V. Blum, *New J Phys*, 2015, **17**, 093020.

149 S. V. Levchenko, X. Ren, J. Wieferink, R. Johanni, P. Rinke, V. Blum and M. Scheffler, *Comput Phys Commun*, 2015, **192**, 60–69.

150 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput Phys Commun*, 2009, **180**, 2175–2196.

151 H. Prats and M. Stamatakis, *J Mater Chem A Mater*, 2022, **10**, 1522–1534.

152 I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, F. Della Pia, V. L. Deringer, R. Elijošius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills and G. Csányi, arXiv preprint, arXiv:2401.00096v2, Mar 2024 .

153 B. Deng, P. Zhong, K. Jun, J. Riebesell, K. Han, C. J. Bartel and G. Ceder, *Nat Mach Intell*, 2023, **5**, 1031–1041.

154 Matbench Discovery, https://matbench-discovery.materialsproject.org/, (accessed 11 May 2024).

155 A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. Bjerre Jensen, J. Kermode, J. R. Kitchin, E. Leonhard Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. Bergmann Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *Journal of Physics: Condensed Matter*, 2017, **29**, 273002.