A Database of Low-Energy Atomically Precise Nanoclusters

Sukriti Manna[1], Alberto Hernandez[1,2], Yunzhe Wang[1,2], Peter Lile[1], Shanping Liu[1] and Tim Mueller[1]

[1]Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD 21218

[2]These authors contributed equally

*Corresponding author(s): Tim Mueller (tmueller@jhu.edu)

ABSTRACT

The chemical and structural properties of atomically precise nanoclusters are of great interest in numerous applications, but the structures of the clusters can be computationally expensive to predict. In this work, we present the largest database of cluster structures and properties determined using ab-initio methods to date. We report the methodologies used to discover low-energy clusters as well as the energies, relaxed structures, and physical properties (such as relative stability, HOMO-LUMO gap among others) for over 50,000 clusters across 55 elements. We have identified 589 structures which have energies lower than any previously reported in the literature by at least 1 meV/atom, and we have identified 1340 new structures for clusters that were previously unexplored in the literature. Patterns in the data reveal insights into the chemical and structural relationships among the elements at the nanoscale. We describe how the database can be accessed for future studies and the development of nanocluster-based technologies.

BACKGROUND AND SUMMARY

Small nanoclusters possess novel physical properties which differ from those of their bulk counterparts, including discrete energy levels [1, 2], nonlinear optical properties [3], magnetism [4], high catalytic activity [5-7], multiple absorption bands [7, 8], and enhanced photoluminescence [9-12]. These properties emerge as a consequence of their small size and relatively high number of uncoordinated atoms on the surface, and they can be tuned by altering the size and shape of the cluster [13, 14]. The past few decades have shown significant progress in computational methods to predict these properties, but before a property can be calculated it is necessary to first determine the atomic structure of the cluster. A number of algorithms have been developed to predict the ground state atomic structure by comprehensively sampling the potential energy surface (PES) of the cluster. These methods include genetic algorithms [15-17], simulated annealing [18], particle swarm optimization [19], Bayesian optimization [20] and basin-hopping [21] methods. In each of these methods, identification of the ground state structure is accomplished by first calculating the energies of a large number of candidate structures and then selecting the structures with the lowest energies.

Due in part to strong quantum finite-size effects, accurate determination of the relative cluster energies is best accomplished using ab-initio calculations, and the number of low-energy configurational isomers is estimated to grow exponentially with the number of atoms in the cluster [22, 23]. For these reasons searching for the ground state atomic structure can be computationally demanding. An alternative approach to identifying low-energy structures is to search through a reference database of known

structures. In recent years, materials databases for crystalline materials have transformed materials research [24-28]. However, current nanocluster structure datasets are either unavailable to the public, limited in scope, or primarily utilize lower levels of theory like interatomic potential models [29-32] and tight binding models [33].

Here we present a database of low-energy cluster structures for 55 elements, for clusters of 3-55 atoms, calculated using density functional theory (DFT) [34]. The database contains structures obtained through an ab-initio genetic algorithm, an accelerated genetic algorithm using machine-learned interatomic potentials, regression analysis of chemically similar elements, and a survey of the scientific literature. The structures from the scientific literature include 3131 nanoclusters from the Cambridge Cluster Database [33], which were identified using interatomic potentials. The 55 elements encompass different regions of the periodic table, including alkali and alkaline earth metals, transition metals, post transition metals, metalloids, and non-metals. Although we are continuing to add to this data set, to the best of our knowledge, this dataset already constitutes the most extensive collection of computed cluster structures at the DFT level of theory. The data set can be used to guide experimental synthesis of predicted nanoclusters, to guide searches for low-energy clusters in different chemical environments, to computationally screen for clusters suitable for a variety of applications, or to train machine learning models. Since the structural energies were obtained using a consistent computational method, the data also serves as a direct source for comparative benchmark studies of different DFT or other electronic structure techniques within the context of atomic cluster modelling. All atomic structures and their calculated properties are openly distributed, enabling researchers across the world to access it for free and use it for further analysis.

METHODS

We have used the following methods to populate the Quantum Cluster Database with atomically precise nanoclusters:

1. We have searched the literature for coordinates of previously-discovered candidate low-energy clusters. The atomic structures and their literature sources are summarized in Table 1 and SI Figure S1.
2. We have used a genetic algorithm with *ab-initio* calculations. This method was primarily used to identify some of the structures for elements that are computationally cheap as determined by the number of valence electrons in the projector augmented wave potentials used (Table S1 in SI), such as Mg, Li, Sb, Na, Ga, and Si. The details of the genetic algorithm method can be found in SI section 1 as well as in reference [35].
3. We have used a genetic algorithm accelerated by actively learned moment tensor potentials (MTP)[36-39] trained on-the-fly. This method has been used to find S clusters of sizes 51, 52, and 53 and we are currently using it to find additional clusters for B, P, and C.
4. We have used correlations in the total energies of different elements to predict low-energy clusters for one element from known low-energy structures of chemically similar elements.

A brief description of each of these methods is provided below, with additional details in the SI Section 1 and 2.

Low-energy structures mined from existing literature

Many of the clusters in the QCD have been studied before, including systematic DFT studies of small [40] and large clusters [41] across different elements. We collected the atomic structures of clusters from publications that provide atomic coordinates of reported low-energy structures, as calculated using DFT, and from the Cambridge Cluster Database, which consists primarily of structures discovered using empirical potentials. All structures from the literature were relaxed using our DFT settings as described in the section on DFT calculations. In Table 1, we provide the literature references for these cluster structures grouped by element.

Table 1 Clusters in current database that were extracted from the literature and their literature sources

Element	References	Element	References
Ag	[41-44]	Nb	[40, 41]
Al	[41, 43-47]	Ni	[40-43, 46]
As		Os	[40, 41]
Au	[40, 41, 44, 48-50]	P	[45, 51]
B	[52-56]	Pb	[42-44, 46, 57]
Ba	[42-44, 46]	Pd	[40, 41, 58]
Be	[59]	Pt	[40, 41, 60]
Bi		Rb	[42-44, 46]
C		Re	[40, 41]
Ca	[42, 44, 46, 59]	Rh	[40, 41, 60]
Cd	[40, 41, 61, 62]	Ru	[40, 41]
Co	[40, 41, 63]	S	[64]
Cr	[40-43, 46]	Sb	
Cs	[42-44, 46]	Sc	[40, 41]
Cu	[40-42, 46, 65]	Se	[66]
Fe	[40-43, 67]	Si	[68-72]
Ga	[41, 73, 74]	Sn	
Ge	[75]	Sr	[42-44, 46]
Hf	[40, 41]	Ta	[40, 41]
Hg	[40, 41]	Te	
In	[41]	Ti	[40, 41]
Ir	[40, 41]	Tl	[41]
K	[42-44, 46]	V	[40, 41]
Li	[76]	W	[41-43, 46]
Mg	[41, 45, 59, 77]	Y	[40, 41]
Mn	[41, 45]	Zn	[61, 78]
Mo	[40-43, 67]	Zr	[41, 61]
Na	[41-43, 45, 79]		

Low-energy structures from a genetic algorithm

Low-energy cluster structures were also identified by means of a genetic algorithm (GA)[80, 81], an optimization algorithm based on natural evolution in which beneficial characteristics prevail over successive generations. In our implementation, a GA run begins by populating a pool of clusters with random structures and/or seed structures assembled from previous GA run. All cluster energies are evaluated by relaxing the atomic positions using DFT[34]. Child clusters are generated from pool clusters by using two types of operations: crossover, in which parts of each parent cluster are combined to form a child cluster, and mutation, in which a subset of atoms of a cluster structure are randomly moved. If a child cluster has lower energy than the highest-energy pool cluster and is not structurally equivalent to other pool clusters, it replaces the highest-energy cluster of the pool. The cycle continues until the total number of clusters in the GA run is at least 1000. Additional details of the genetic algorithm method can be found in SI section 1 as well as reference [82].

Learning on the Fly (LOTF)-GA

We have recently developed a way to accelerate the genetic algorithm using machine-learned interatomic potentials trained on-the-fly using active learning [35-37].The machine-learned interatomic potentials are used to quickly identify likely low-energy clusters, which are further relaxed locally by DFT to refine the energies. This method has been used to identify low-energy structures for S clusters of size 51, 52, and 53. Additional details about this method can be found in reference [35].

Correlations between energies for different elements

We have constructed additional low-energy cluster structures by taking advantage of the fact that for some elements there are strong correlations between the total energies of geometrically similar cluster structures. To identify these relationships, we created 55 representative cluster structure prototypes (available in the SI) in a two-step process.

In the first step, we used the genetic algorithm to identify the low energy structures for clusters of 5, 10, 15 and 20 atoms of Al, Be, Li, Mg, Na, Si, Ta, and Ti clusters. These elements were chosen because they cover different parts of the periodic table and are computationally inexpensive relative to others because of the small numbers of valence electrons. The low energy configurations are provided in the supporting information.

In the second step, we used these clusters as templates to create clusters of all the other elements. For each target element, the interatomic distances in the cluster were scaled by the ratio of the nearest neighbor distance of the target element and the template element. The nearest neighbor distances (Table S2) are the bond distances in their most stable bulk form and are retrieved from the Materials Project [26]. To identify a chemically diverse set of elements, we used least-squares regression to express the unrelaxed DFT-calculated cluster energies for each element as a linear combination of the energies of the remaining 54 elements. The residual errors for these fits provide a measure of the extent by which each element is different from the other 54 elements. We selected the 13 elements with the highest errors: B, Ba, Be, Ca, Cr, Cs, K, Li, Mg, Na, Rb, Sr, and Zn, as these are likely to have distinct ground state structures. We then used the genetic algorithm to search for low-energy structures for clusters of 10, 15, 20, 25, and 30 atoms

for these 13 elements. The low-energy structures discovered by the genetic algorithm are shown in Figure 1, and their coordinates are provided in the supporting information. These 65 structures were used as structural templates for determining relationships between structural energies among different elements.

To evaluate the diversity of the 65 template structures, we compared the clusters using a structural similarity score as described in reference [83], where perfectly similar structures have a score of 0.0, and we consider structures with a score above 0.3 to be dissimilar. Across the 5 different sizes and 13 different elements, only five pairs have a similarity score less than 0.3, indicating that the remaining pairs of structures are structurally distinct.

After discovering low-energy clusters with the genetic algorithm, we filled gaps on the database (i.e., elements and sizes where no clusters were available) using the correlations among the energies of the elements. For a gap of a given element and size, we identified the most correlated element and used its 5 lowest-energy clusters of the same size as templates to generate new clusters that were likely to have low energy. We followed the process of re-scaling the interatomic distances using the bulk nearest-neighbor bond length.

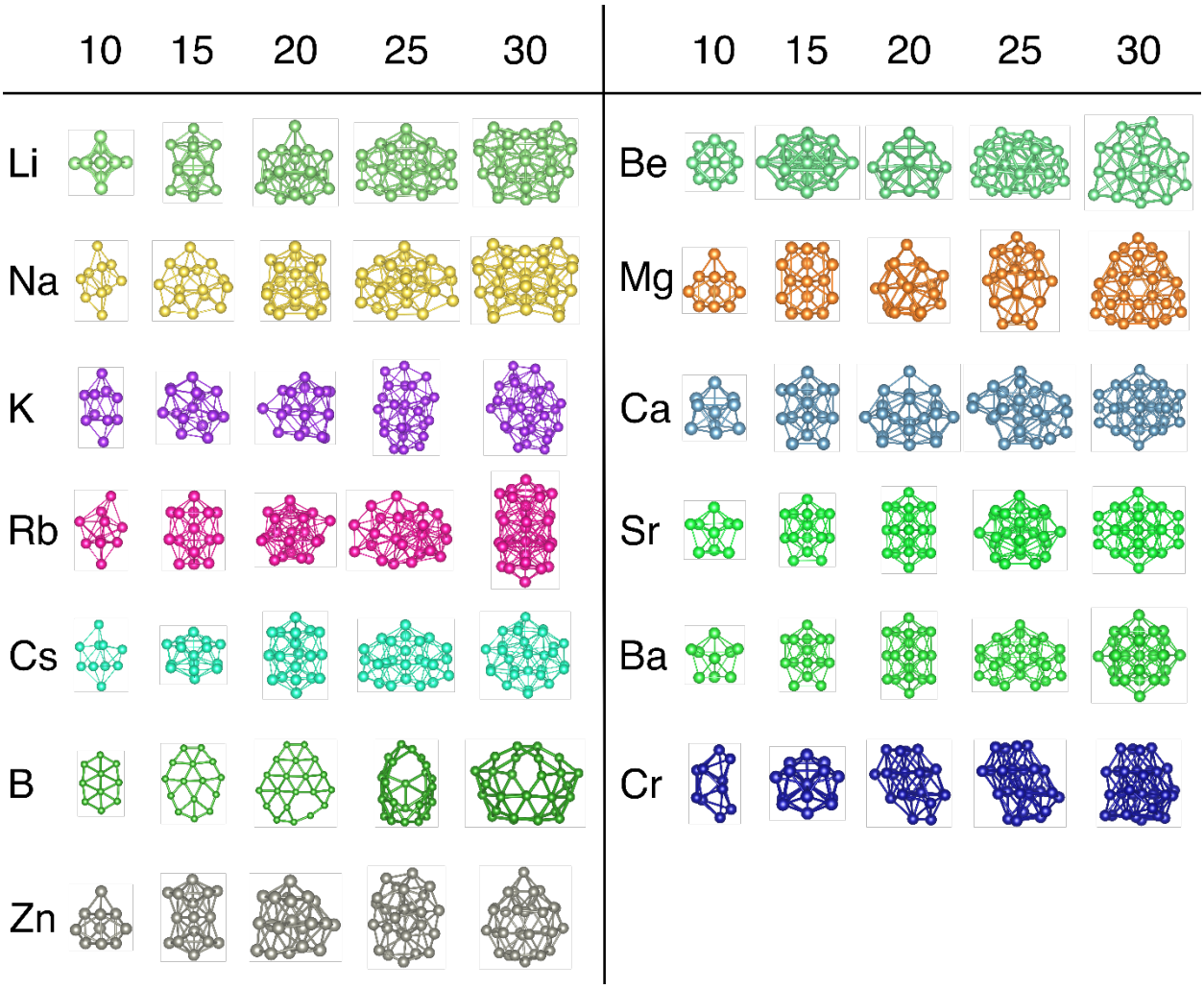

Figure 1. Template clusters used for the data driven method to expedite the filling of the database.

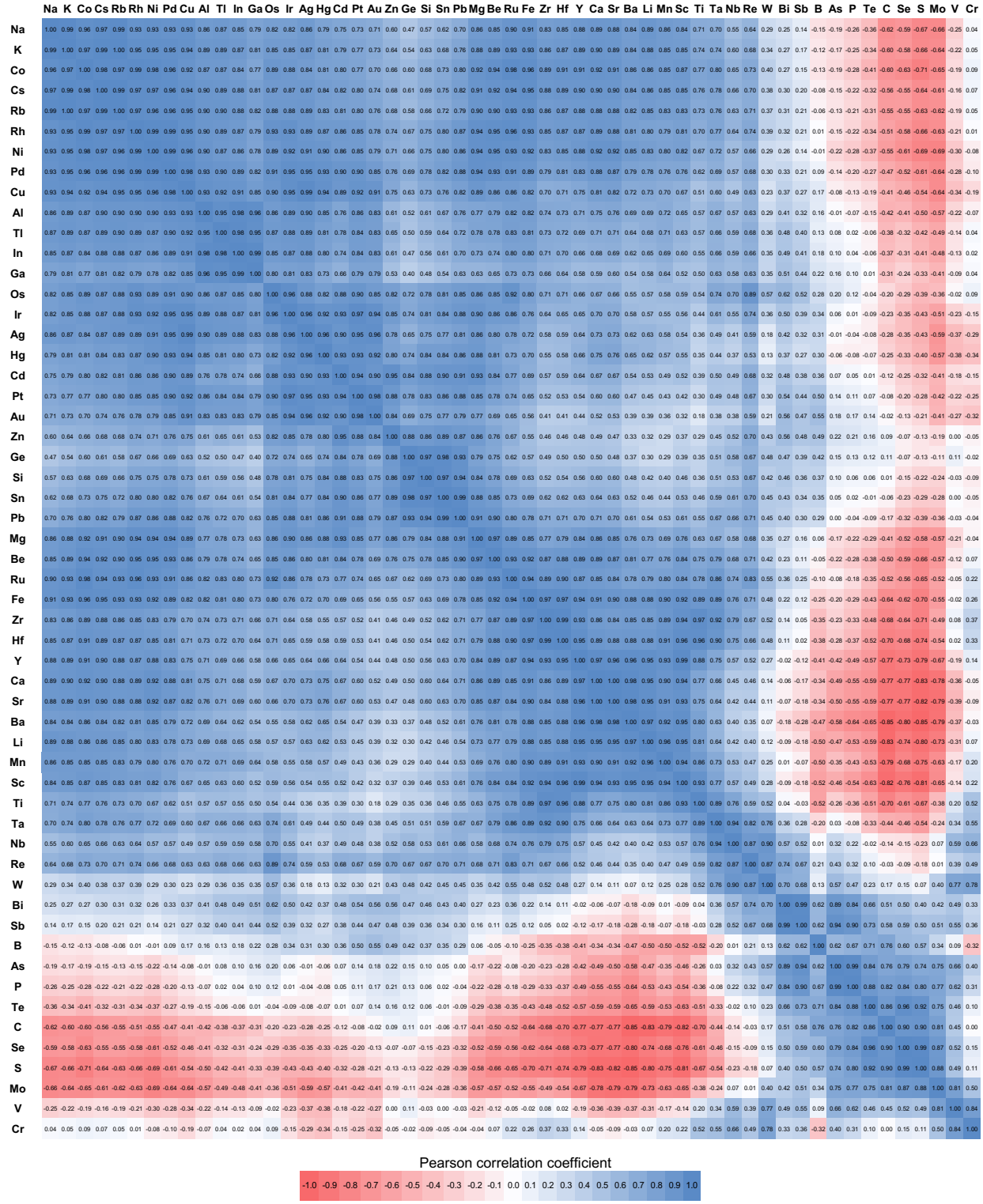

Figure 2. The Pearson correlation coefficients for all 55 elements calculated from the energies of unrelaxed structures, sorted so that two similar elements are close to each other. Blue regions represent positive correlation, red represents negative correlation, and white represents no correlation. The data used in this figure are available as a csv file in the SI and in through the website interface.

DFT calculations

All DFT local energy minimizations were carried out using the Vienna ab initio Simulation Package[84] (VASP) with the Perdew-Burke-Ernzerhof [85] (PBE) generalized gradient approximation exchange-correlation functional. We used the projector-augmented wave [86] method, with the potentials listed in Table S2 in the SI. The convergence criterion for electronic self-consistency was set to 10^{-5} eV per cluster. Structures were optimized using the conjugate gradient approximation [87, 88] as implemented in VASP until all the atomic forces were less than 0.15 eV/Å. All calculations were run at the gamma point with spin polarization. Methfessel Paxton smearing [89] with σ= 0.001 eV was used to achieve high accuracy when calculating final energies. To accelerate convergence for some clusters, a two-step minimizations scheme was adopted with smearing of 0.1 eV in the initial step for faster convergence and a smaller value of 0.001 eV for the final step. Symmetry was turned off for all DFT calculations to increase the likelihood that the calculation completed successfully.

To determine the effect of Spin Orbit Coupling (SOC) on predicting final energies, we performed PBE+SOC for 11 different heavy-metal elements selected based on ref [41]. As we found the inclusion of SOC had little effect on the ranking of low-energy structures, to maintain the consistency of the QCD data we have not included SOC-predicted total energies and atomic structures in the QCD database. More discussion of the effects of SOC, including simple linear equations for estimating SOC-calculated energies from the energies reported in the QCD, can be found in section 5 and section 6 of the SI.

Workflow

We identify candidate low-energy cluster structures using one of the four methods listed above. DFT calculations were performed on these clusters structures before adding them to the database. An outline of the high-throughput workflow used in these DFT calculations is provided in Figure 3. In the first step we check whether the ionic relaxation has converged. If it has not, we rerun the relaxation with modified VASP settings such as increased wall time or a different optimization algorithm.

All DFT calculations were performed using VASP which can only do periodic calculations, so each cluster is in effect surrounded by translationally equivalent clusters. Hence it is essential to use a simulation cell that is sufficiently large to avoid interaction among periodic images. We found that this was best accomplished using two different criteria for different sets of elements. For elements in the Groups 1A and 2A on the periodic table, the minimum distance between neighboring clusters must be greater than 3.5 times the distance between nearest neighbor listed in Table S3 in the SI. For all other elements the minimum distance between periodic images must be greater than 10 angstroms. If, after relaxation, the minimum distance between neighboring clusters was too small, then we increased the supercell size and ran the DFT calculation again. We found that these "box size" criteria are sufficient to reach energy convergence within 2 meV/atom in all 1135 tested cases (please see Figure S8 and additional details in the SI) with a root mean squared error of 0.118 meV/atom.

The atoms sometimes form periodic configurations that correspond to nanowires or slabs. We filter out these types of structures by discarding clusters that have a minimum distance between periodic images

smaller than 1.5 times the atomic nearest neighbor distance. We also screen for discontiguous clusters using this same criterion and discard any discontiguous clusters that are identified.

To ensure the Quantum Cluster Database contains only unique clusters for a given element and size, when two clusters have a structural similarity score [83] less than 0.3, the cluster with higher energy is discarded. If the higher-energy cluster was from the literature, the appropriate literature references are assigned to the structurally similar low-energy cluster. All the filters that ensures the quality of the clusters in the database are summarized in the Figure 3.

The properties and metadata described in the Data Records section are calculated for each cluster and stored in a PostgreSQL database. Finally, the data are displayed in the Quantum Cluster Database website and output as a JSON file and a .csv file.

DFT calculations on magnetic clusters

We used spin polarized calculations with default magnetic initializations (1 μ_B /atom) for all elements other than Fe, Mn, Co, Ni, Ru, Rh, V, Cu, and Cr. We found that the final magnetic states for Fe, Mn, Ni, Ru, Rh, V, and Cr clusters are particularly likely to depend on the magnetic initialization. To increase the likelihood of relaxing into the lowest-energy magnetic state for these elements we run multiple calculations with different magnetic initializations, as listed in Table S4. The calculations that yields the lowest energy is selected for inclusion in the QCD. Based on benchmarks on 2228 clusters of above-mentioned elements selected from an early version of the QCD, we found that this approach yielded the lowest-energy magnetic state more than 97% of the time. For clusters of Cu and Co, we find that the magnetic initializations with 3 μ_B /atom relaxed into lowest energy configurations in all of our benchmark calculations. Additional details of the magnetic initializations are provided in section 2 of the SI.

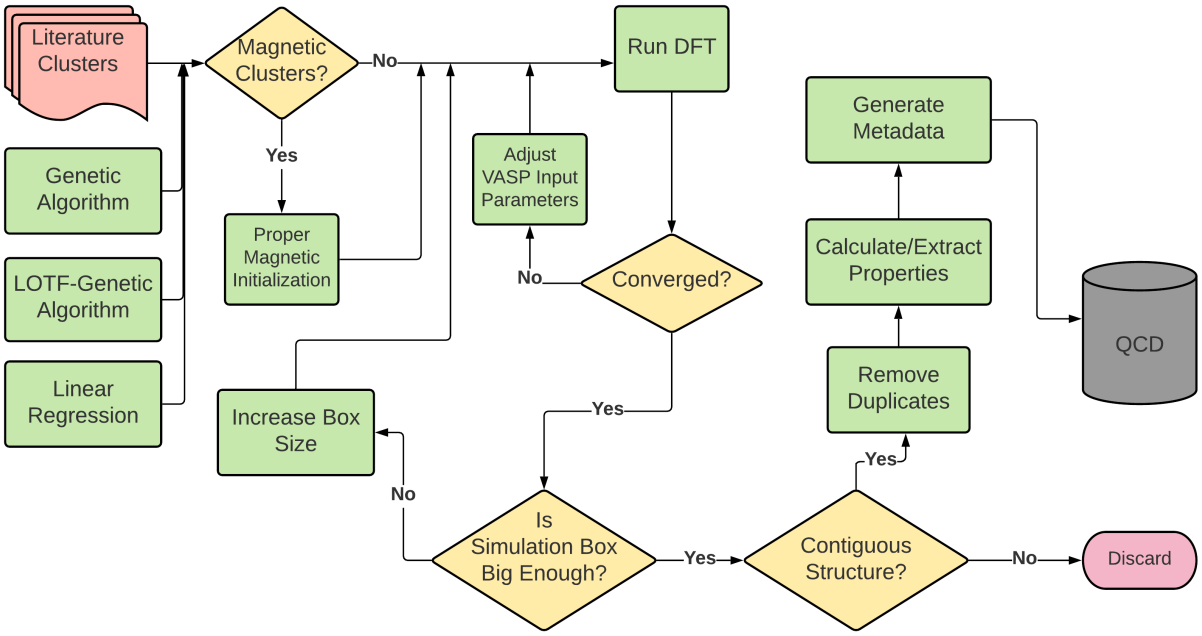

Figure 3. Schematic overview of the high-throughput workflow used in this study and the quantities calculated within the framework of the database.

DATA RECORDS

The files from the DFT calculations of the more than 50,000 clusters are publicly available through the NOMAD database [90, 91]. A web interface to visualize the structures, the correlations, and the properties can be accessed through the website of the Mueller Research Group (http://muellergroup.jhu.edu/qcd), from where the data can be downloaded in as a JSON file or as a comma-delimited file.

Table 2. Keys, types of data, and description of the QCD data in the JSON file and .csv format

Key	Datatype	Description
cluster_id	string	ID of the cluster in QCD
element_symbol	string	symbol of the element of the cluster
n_atoms	number	number of atoms in the cluster
n_val_electrons	number	number of valence electrons corresponding to the pseudopotential
energy_dft	number	energy in eV
energy_relative	number	energy in eV above the lowest energy structure of the same element and size
energy_n_minus_one	number	formation energy in eV relative to the lowest energy structure of the same element but of size N-1
energy_n_plus_one	number	formation energy in eV relative to the lowest energy structure of the same element but of size N+1
homo_lumo_gap	number	HOMO-LUMO Gap in eV
magnetic_moment	number	Magnetic moment of the cluster in units of Bohr magneton (μ_B)
similar_structures	list	space delimited list of cluster_id of clusters within QCD that are similar to this cluster
references	list	space delimited list of literature references
structure_xyz	string	structure represented in XYZ format [(a)]
structure_poscar_format	string	structure represented in POSCAR format [(a)]

Note: (a) semicolons are used instead of line breaks.

File format

The data is available for download as a JSON file and as a .csv file. Both can be downloaded from the Quantum Cluster Database website. The first level of the JSON file contains a unique index for every cluster, the next level contains the cluster_id described in Table 2, and the next level contains the other keys described in Table 2, with the corresponding values. The columns of the .csv file correspond to the keys described in Table 2. The VASP DFT calculation files for each cluster are available in the NOMAD repository in the form of text files from the inputs and outputs of VASP.

Properties

For each cluster of a given number of atoms N and element type k, the database contains the energy relative to the lowest energy structure of size N and species k, the formation energy with respect to the stable cluster of size N-1 of species k (equation (1)), the formation energy with respect to the N+1 stable cluster of the same species (equation (2)), the HOMO-LUMO gap, the number of valence electrons considered by DFT, the magnetic moment, a list of similar structures within the Quantum Cluster Database, a list of literature references for the cluster (downloadable in .bib format), the coordinates (downloadable in XYZ format), and an interactive visualization of the cluster. The formation energies are calculated using the following equations:

$$E_{f,N,N-1} = E_N - E_{N-1} - E_{atom} \text{ --------- (1)}$$

$$E_{f,N,N+1} = E_N + E_{atom} - E_{N+1} \text{ --------- (2)}$$

where E_N is the energy of the cluster of size N, E_{N-1} is the energy of the cluster of size N-1, E_{N+1} is the energy of the cluster of size N+1, and E_{atom} is the energy of an isolated atom. The energies for isolated atoms used in these calculations are provided in Table S3 of the SI.

TECHNICAL VALIDATION

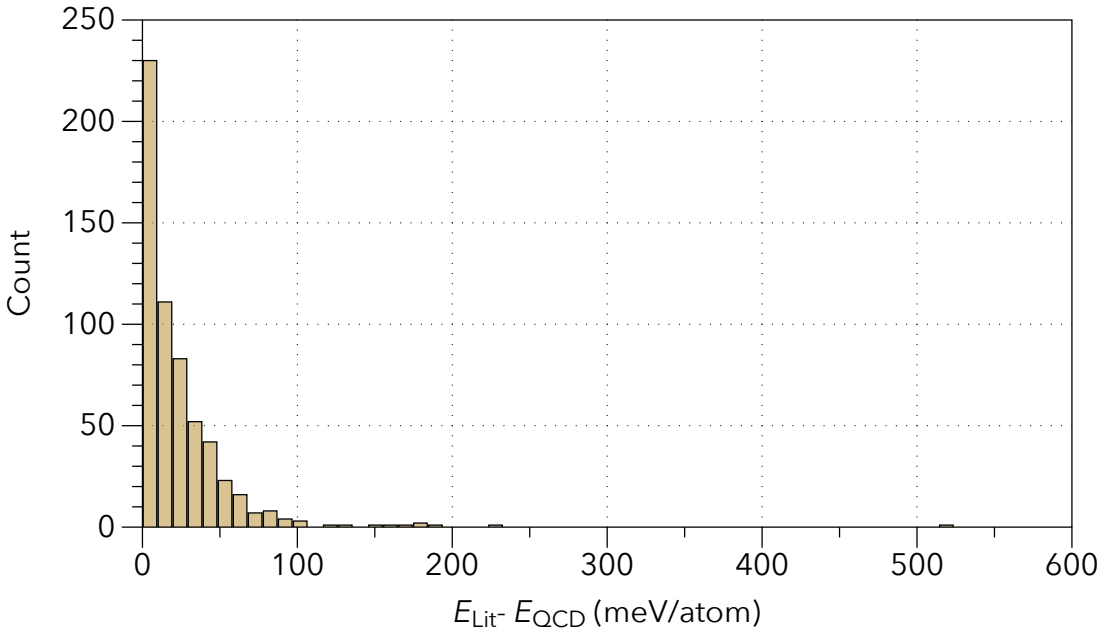

Figure 4. Histogram of energy differences between the lowest-energy clusters reported in the literature (minus 1 meV/atom to account for DFT precision) and the lowest-energy clusters in the QCD. The Quantum Cluster Database contains 589 clusters that have a lower energy than the lowest-energy clusters from the literature.

For a given cluster size and element, we compared the lowest-energy cluster from the literature against the lowest-energy non-literature cluster in the database to assess which had lower energy. There are 1540 systems for which there is at least one literature structure in the database. The database contains clusters that are lower in energy by at least 1 meV / atom for 589 of these (Figure 5).

The Quantum Cluster Database contains 1483 structure types or templates (i.e., relative arrangements of atoms) that were not previously reported in the literature. The 1483 templates were identified from the set of all clusters with calculated energies within 1 meV/atom of the lowest energy cluster with the same element and size. In comparison, there are 684 templates of low-energy templates from the literature (Table1).

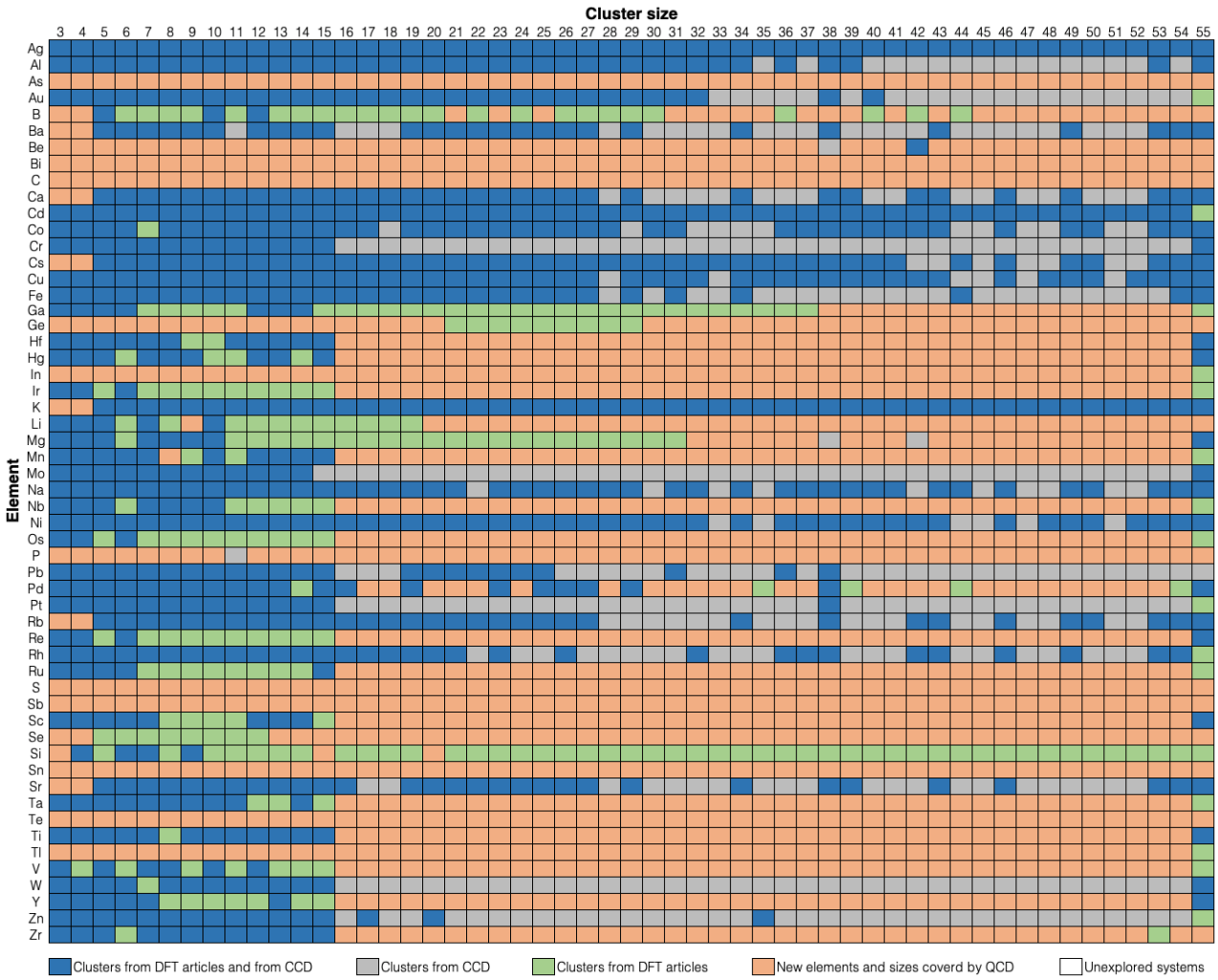

Figure 5. A summary of previous studies of the structures of elemental clusters with 3-55 atoms, including publications that used DFT to find atomic structures (green and blue) as well as systems covered in the Cambridge Cluster Database (grey and blue). The Quantum Cluster Database covers these clusters as well as 1340 regions that were previously unexplored (shown in orange).

Before our work, here were 1575 cluster elements and sizes covered by the literature out of 2915, which corresponds to 54%. With the Quantum Cluster Database, the percentage covered increased to 100%.

USAGE NOTES

When clicking on an element, the website interface enables the visualization of the energetic correlations with other elements (Figure S10a) and the visualization of the relative energies for clusters of every size (Figure S10b). When clicking on a particular cluster, the properties (Table 2, except for the raw DFT energy and the structure in POSCAR format) are displayed together with an interactive view of the cluster and options to download the XYZ structure file and references in BibTeX format.

CODE AVAILABILITY

The implementations of the DFT and MTP genetic algorithms used to search for low-energy structures are available via GitLab: https://gitlab.com/muellergroup/cluster-ga.

ACKNOWLEDGEMENTS

This work is supported by a Multidisciplinary University Research Initiatives (MURI) grant from the Office of Naval Research, N00014-15-1-2681. Calculations were performed using resources provided by the Department of Defense HPC Modernization Program through grant STVONRDC40463482.

We would also like to acknowledge Caitlin McCandler (Lawrence Berkeley National Laboratory), Chuhong Wang, Chenyang Li, Sam Norwood, Kit Bowen (Johns Hopkins University), and Joseph Hooper (Naval Postgraduate School) for helpful discussions.

AUTHOR CONTRIBUTIONS

YW, PL, and SM developed the genetic algorithm code with input from TM. SM, AH, and PL extracted clusters from the literature. SM, PL, and TM developed data driven method using the structural correlation to accelerate the search. AH assessed the quality of the clusters in the database. SM and AH computed the properties and metadata for the database. AH edited and installed the website interface in collaboration with an outside team. SM, AH, and TM processed data, generated graphs, and analyzed the graphs. SM, AH, YW, and TM wrote the manuscript. AH, and YW contributed equally. TM conceived of and managed the project. All authors contributed to the discussion.

REFERENCES

1. Jena, P. and Q. Sun, *Super Atomic Clusters: Design Rules and Potential for Building Blocks of Materials.* Chemical Reviews, 2018. **118**(11): p. 5755-5870.
2. Wilcoxon, J.P. and B.L. Abrams, *Synthesis, structure and properties of metal nanoclusters.* Chemical Society Reviews, 2006. **35**(11): p. 1162-1194.
3. Jin, R., *Atomically precise metal nanoclusters: stable sizes and optical properties.* Nanoscale, 2015. **7**(5): p. 1549-1565.
4. Cramer, C.J. and D.G. Truhlar, *Density functional theory for transition metals and transition metal chemistry.* Physical Chemistry Chemical Physics, 2009. **11**(46): p. 10757-10816.
5. Li, G. and R. Jin, *Atomically Precise Gold Nanoclusters as New Model Catalysts.* Accounts of Chemical Research, 2013. **46**(8): p. 1749-1758.
6. Batista, K.E., et al., *Ab Initio Investigation of CO2 Adsorption on 13-Atom 4d Clusters.* Journal of chemical information and modeling, 2020. **60**(2): p. 537-545.
7. Felício-Sousa, P., K.F. Andriani, and J.L. Da Silva, *Ab initio investigation of the role of the d-states occupation on the adsorption properties of H 2, CO, CH 4 and CH 3 OH on the Fe 13, Co 13, Ni 13 and Cu 13 clusters.* Physical Chemistry Chemical Physics, 2021. **23**(14): p. 8739-8751.
8. Jia, X., J. Li, and E. Wang, *Cu Nanoclusters with Aggregation Induced Emission Enhancement.* Small, 2013. **9**(22): p. 3873-3879.
9. Zhang, Y., et al., *Unique size-dependent nanocatalysis revealed at the single atomically precise gold cluster level.* Proceedings of the National Academy of Sciences, 2018. **115**(42): p. 10588.
10. Chakraborty, I. and T. Pradeep, *Atomically Precise Clusters of Noble Metals: Emerging Link between Atoms and Nanoparticles.* Chemical Reviews, 2017. **117**(12): p. 8208-8271.
11. Watanabe, Y., *Atomically precise cluster catalysis towards quantum controlled catalysts.* Science and Technology of Advanced Materials, 2014. **15**(6): p. 063501.
12. Zhu, Y., H. Qian, and R. Jin, *Catalysis opportunities of atomically precise gold nanoclusters.* Journal of Materials Chemistry, 2011. **21**(19): p. 6793-6799.
13. Li, Z.Y., et al., *Three-dimensional atomic-scale structure of size-selected gold nanoclusters.* Nature, 2008. **451**(7174): p. 46-48.
14. Castleman, A.W. and S.N. Khanna, *Clusters, Superatoms, and Building Blocks of New Materials.* The Journal of Physical Chemistry C, 2009. **113**(7): p. 2664-2675.
15. Wu, S.Q., et al., *An adaptive genetic algorithm for crystal structure prediction.* Journal of Physics: Condensed Matter, 2013. **26**(3): p. 035402.
16. Curtis, F., et al., *GAtor: A First-Principles Genetic Algorithm for Molecular Crystal Structure Prediction.* Journal of Chemical Theory and Computation, 2018. **14**(4): p. 2246-2264.
17. Jennings, P.C., et al., *Genetic algorithms for computational materials discovery accelerated by machine learning.* npj Computational Materials, 2019. **5**(1): p. 46.
18. Kirkpatrick, S., C.D. Gelatt, and M.P. Vecchi, *Optimization by simulated annealing.* science, 1983. **220**(4598): p. 671-680.
19. Lv, J., et al., *Particle-swarm structure prediction on clusters.* The Journal of Chemical Physics, 2012. **137**(8): p. 084104.
20. Yamashita, T., et al., *Crystal structure prediction accelerated by Bayesian optimization.* Physical Review Materials, 2018. **2**(1): p. 013803.
21. Yang, S. and G.M. Day, *Exploration and Optimization in Crystal Structure Prediction: Combining Basin Hopping with Quasi-Random Sampling.* Journal of Chemical Theory and Computation, 2021. **17**(3): p. 1988-1999.
22. Stillinger, F.H., *Exponential multiplicity of inherent structures.* Physical Review E, 1999. **59**(1): p. 48-51.

23. Heard, C.J. and R.L. Johnston, *Global Optimisation Strategies for Nanoalloys*, in *Challenges and Advances in Computational Chemistry and Physics*, M. Nguyen and B. Kiran, Editors. 2017, Springer, Cham.
24. Curtarolo, S., et al., *AFLOW: An automatic framework for high-throughput materials discovery.* Computational Materials Science, 2012. **58**: p. 218-226.
25. Saal, J.E., et al., *Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD).* Jom, 2013. **65**(11): p. 1501-1509.
26. Jain, A., et al., *Commentary: The Materials Project: A materials genome approach to accelerating materials innovation.* APL Materials, 2013. **1**(1): p. 011002.
27. Choudhary, K., et al., *The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design.* npj Computational Materials, 2020. **6**(1): p. 1-13.
28. Zhou, J., et al., *2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches.* Scientific data, 2019. **6**(1): p. 1-10.
29. Joswig, J.-O. and M. Springborg, *Genetic-algorithms search for global minima of aluminum clusters using a Sutton-Chen potential.* Physical Review B, 2003. **68**(8): p. 085408.
30. Shao, X., X. Liu, and W. Cai, *Structural optimization of silver clusters up to 80 atoms with Gupta and Sutton-Chen potentials.* Journal of chemical theory and computation, 2005. **1**(4): p. 762-768.
31. Grigoryan, V.G., D. Alamanova, and M. Springborg, *Structure and energetics of Cu N clusters with ($2 \leqslant N \leqslant 150$): An embedded-atom-method study.* Physical Review B, 2006. **73**(11): p. 115415.
32. Loeffler, T.D., et al., *Active Learning A Neural Network Model For Gold Clusters & Bulk From Sparse First Principles Training Data.* ChemCatChem, 2020. **12**(19): p. 4796-4806.
33. D. J. Wales, J.P.K.D., A. Dullweber, M. P. Hodges, F. Y. Naumkin F. Calvo, J. Hernández-Rojas and T. F. Middleton. *The Cambridge Cluster Database*. [cited 2021 06-28-2021]; Available from: http://www-wales.ch.cam.ac.uk/CCD.html.
34. Kohn, W. and L. Sham, *Phys. Rev. A.* Self–Consistent Equations Including Exchange and Correlation Effects, 1965. **140**: p. A1133-A1138.
35. Wang, Y., et al., *Accelerated Prediction of Atomically Precise Cluster Structures Using On-the-fly Machine Learning.* ChemRxiv, doi: http://dx.doi.org/10.33774/chemrxiv-2021-q3wng (2021).
36. Shapeev, A.V., *Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials.* Multiscale Modeling & Simulation, 2016. **14**(3): p. 1153-1173.
37. Podryabinkin, E.V. and A.V. Shapeev, *Active learning of linearly parametrized interatomic potentials.* Computational Materials Science, 2017. **140**: p. 171-180.
38. Zuo, Y., et al., *Performance and cost assessment of machine learning interatomic potentials.* The Journal of Physical Chemistry A, 2020. **124**(4): p. 731-745.
39. Novikov, I.S., et al., *The MLIP package: moment tensor potentials with MPI and active learning.* Machine Learning: Science and Technology, 2020. **2**(2): p. 025002.
40. Chaves, A.S., M.J. Piotrowski, and J.L.F. Da Silva, *Evolution of the structural, energetic, and electronic properties of the 3d, 4d, and 5d transition-metal clusters (30 TMn systems for n = 2–15): a density functional theory investigation.* Physical Chemistry Chemical Physics, 2017. **19**(23): p. 15484-15502.
41. Piotrowski, M.J., et al., *Theoretical Study of the Structural, Energetic, and Electronic Properties of 55-Atom Metal Nanoclusters: A DFT Investigation within van der Waals Corrections, Spin–Orbit Coupling, and PBE+U of 42 Metal Systems.* The Journal of Physical Chemistry C, 2016. **120**(50): p. 28844-28856.
42. P. K. Doye, J. and D. J. Wales, *Structural consequences of the range of the interatomic potential A menagerie of clusters.* Journal of the Chemical Society, Faraday Transactions, 1997. **93**(24): p. 4233-4243.

43. Doye, J.P.K., D.J. Wales, and R.S. Berry, *The effect of the range of the potential on the structures of clusters.* The Journal of Chemical Physics, 1995. **103**(10): p. 4234-4249.
44. Wales, D.J. and J.P.K. Doye, *Theoretical Predictions of Structure and Thermodynamics in the Large Cluster Regime*, in *Large Clusters of Atoms and Molecules*, T.P. Martin, Editor. 1996, Springer Netherlands: Dordrecht. p. 241-279.
45. Galvão, B.R.L. and L.P. Viegas, *What Electronic Structure Method Can Be Used in the Global Optimization of Nanoclusters?* The Journal of Physical Chemistry A, 2019. **123**(48): p. 10454-10462.
46. Doye, J.P.K. and D.J. Wales, *The effect of the range of the potential on the structure and stability of simple liquids: from clusters to bulk, from sodium to.* Journal of Physics B: Atomic, Molecular and Optical Physics, 1996. **29**(21): p. 4859-4894.
47. Aguado, A. and J.M. López, *Structures and stabilities of Aln+, Aln, and Aln− (n=13–34) clusters.* The Journal of Chemical Physics, 2009. **130**(6): p. 064704.
48. Liu, X.-J., et al., *The stability of small helical gold nanorods: A relativistic density functional study.* Journal of Computational Chemistry, 2012. **33**(3): p. 311-318.
49. Fa, W., C. Luo, and J. Dong, *Bulk fragment and tubelike structures of ${\mathrm{Au}}_{N}$ $(N=2\text{\ensuremath{-}}26)$.* Physical Review B, 2005. **72**(20): p. 205428.
50. Fernández, E.M., et al., *Trends in the structure and bonding of noble metal clusters.* Physical Review B, 2004. **70**(16): p. 165403.
51. Sai, L., et al., *Structural Evolution of Medium-Sized Phosphorus Clusters (P 20–P 36) from Ab Initio Global Search.* Journal of Cluster Science, 2020. **31**(3): p. 567-574.
52. Tai, T.B. and M.T. Nguyen, *Electronic structure and photoelectron spectra of Bn with n = 26–29: an overview of structural characteristics and growth mechanism of boron clusters.* Physical Chemistry Chemical Physics, 2015. **17**(20): p. 13672-13679.
53. Tai, T.B., et al., *A disk-aromatic bowl cluster B30: toward formation of boron buckyballs.* Chemical Communications, 2014. **50**(13): p. 1558-1560.
54. Pham, H.T., et al., *The boron conundrum: Bonding in the bowl B30 and B36, fullerene B40 and triple ring B42 clusters.* Chemical Physics Letters, 2014. **608**: p. 295-302.
55. Tai, T.B. and M.T. Nguyen, *A new chiral boron cluster B44 containing nonagonal holes.* Chemical Communications, 2016. **52**(8): p. 1653-1656.
56. Pham, H.T., et al., *The 2D-to-3D geometry hopping in small boron clusters: The charge effect.* Chemical Physics Letters, 2013. **577**: p. 32-37.
57. Doye, J.P.K. and S.C. Hendy, *On the structure of small lead clusters.* The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics, 2003. **22**(1): p. 99-107.
58. Nava, P., M. Sierka, and R. Ahlrichs, *Density functional study of palladium clusters.* Physical Chemistry Chemical Physics, 2003. **5**(16): p. 3372-3381.
59. Dieterich, J.M., S. Gerke, and R.A. Mata, *A First-Principles-Based Potential for the Description of Alkaline Earth Metals.* Journal of Atomic, Molecular, and Optical Physics, 2012. **2012**: p. 648386.
60. P. K. Doye, J. and D. J. Wales, *Global minima for transition metal clusters described by Sutton–Chen potentials.* New Journal of Chemistry, 1998. **22**(7): p. 733-744.
61. Doye, J.P.K., *Lead clusters: Different potentials, different structures.* Computational Materials Science, 2006. **35**(3): p. 227-231.
62. Noya, E.G., et al., *Geometric magic numbers of sodium clusters: Interpretation of the melting behaviour.* The European Physical Journal D, 2007. **43**(1): p. 57-60.
63. Zhan, L., et al., *Asynchronous multicanonical basin hopping method and its application to cobalt nanoclusters.* The Journal of Chemical Physics, 2005. **122**(24): p. 244707.

64. Jin, Y., et al., *Geometries, stabilities and fragmental channels of neutral and charged sulfur clusters: S n Q (n= 3–20, Q= 0,±1).* Physical Chemistry Chemical Physics, 2015. **17**(20): p. 13590-13597.
65. Calaminici, P., et al., *On the ground state structure of neutral Cun (n=12,14,16,18,20) clusters.* Computational and Theoretical Chemistry, 2013. **1021**: p. 41-48.
66. Alparone, A., *Density functional theory Raman spectra of cyclic selenium clusters Sen (n=5–12).* Computational and Theoretical Chemistry, 2012. **988**: p. 81-85.
67. Elliott, J.A., Y. Shibuta, and D.J. Wales, *Global minima of transition metal clusters described by Finnis–Sinclair potentials: A comparison with semi-empirical molecular orbital theory.* Philosophical Magazine, 2009. **89**(34-36): p. 3311-3332.
68. Bazterra, V.E., et al., *Modified genetic algorithms to model cluster structures in medium-size silicon clusters.* Physical Review A, 2004. **69**(5): p. 053202.
69. Yoo, S., et al., *Structures and relative stability of medium-sized silicon clusters. V. Low-lying endohedral fullerenelike clusters Si31–Si40 and Si45.* The Journal of Chemical Physics, 2006. **124**(16): p. 164311.
70. Yoo, S. and X.C. Zeng, *Structures and relative stability of medium-sized silicon clusters. IV. Motif-based low-lying clusters Si21–Si30.* The Journal of Chemical Physics, 2006. **124**(5): p. 054304.
71. Yoo, S. and X.C. Zeng, *Motif Transition in Growth Patterns of Small to Medium-Sized Silicon Clusters.* Angewandte Chemie International Edition, 2005. **44**(10): p. 1491-1494.
72. Goedecker, S., W. Hellmann, and T. Lenosky, *Global Minimum Determination of the Born-Oppenheimer Surface within Density Functional Theory.* Physical Review Letters, 2005. **95**(5): p. 055501.
73. Núñez, S., J.M. López, and A. Aguado, *Neutral and charged gallium clusters: structures, physical properties and implications for the melting features.* Nanoscale, 2012. **4**(20): p. 6481-6492.
74. Drebov, N., F. Weigend, and R. Ahlrichs, *Structures and properties of neutral gallium clusters: A theoretical investigation.* The Journal of Chemical Physics, 2011. **135**(4): p. 044314.
75. Yoo, S. and X.C. Zeng, *Search for global-minimum geometries of medium-sized germanium clusters. II. Motif-based low-lying clusters Ge21–Ge29.* The Journal of Chemical Physics, 2006. **124**(18): p. 184309.
76. Hu, H.-S., et al., *Theoretical studies of the global minima and polarizabilities of small lithium clusters.* Chemical Physics Letters, 2016. **644**: p. 235-242.
77. Belyaev, S.N., et al., *Structural, electronic, thermodynamic and spectral properties of Mgn (n=2–31) clusters. A DFT study.* Computational and Theoretical Chemistry, 2016. **1079**: p. 34-46.
78. Aguado, A., et al., *Are zinc clusters really amorphous? A detailed protocol for locating global minimum structures of clusters.* Nanoscale, 2018. **10**(40): p. 19162-19181.
79. Doye, J.P.K., *Identifying structural patterns in disordered metal clusters.* Physical Review B, 2003. **68**(19): p. 195418.
80. Oganov, A.R., A.O. Lyakhov, and M. Valle, *How Evolutionary Crystal Structure Prediction Works and Why.* Accounts of chemical research, 2011. **44**(3): p. 227-237.
81. Trimarchi, G., A.J. Freeman, and A. Zunger, *Predicting stable stoichiometries of compounds via evolutionary global space-group optimization.* Physical Review B, 2009. **80**(9): p. 092101.
82. Wang, Y., et al., *Accelerated Prediction of Atomically Precise Cluster Structures Using On-the-fly Active Learning.* Manuscript in preparation.
83. Li, X.-T., X.-B. Yang, and Y.-J. Zhao, *Geometrical eigen-subspace framework based molecular conformation representation for efficient structure recognition and comparison.* The Journal of Chemical Physics, 2017. **146**(15): p. 154108.
84. Kresse, G. and J. Furthmuller, *Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set.* Phys. Rev. B, 1996. **54**: p. 11169.

85. Perdew, J.P., K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple.* Phys. Rev. Lett., 1996. **77**: p. 3865-3868.
86. Blöchl, P.E., *Projector augmented-wave method.* Phys. Rev. B, 1994. **50**: p. 17953-17979.
87. Gillan, M., *Calculation of the vacancy formation energy in aluminium.* Journal of Physics: Condensed Matter, 1989. **1**(4): p. 689.
88. Štich, I., et al., *Conjugate gradient minimization of the energy functional: A new method for electronic structure calculation.* Physical Review B, 1989. **39**(8): p. 4997.
89. Aarons, J., et al., *Perspective: Methods for large-scale density functional calculations on metallic systems.* The Journal of chemical physics, 2016. **145**(22): p. 220901.
90. Draxl, C. and M. Scheffler, *NOMAD: The FAIR concept for big data-driven materials science.* Mrs Bulletin, 2018. **43**(9): p. 676-682.
91. Sukriti Manna, et al., *Quantum Cluster Database (QCD) Dataset* NOMAD Laboratory, doi: http://dx.doi.org/10.17172/NOMAD/2021.11.30-1 (2021).

Supporting Information for

A Database of Low-Energy Atomically Precise Nanoclusters

Sukriti Manna[1], Alberto Hernandez[1,2], Yunzhe Wang[1,2], Peter Lile[1], Shanping Liu[1] and Tim Mueller[1]

[1]Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD 21218

[2]These authors contributed equally

*Corresponding author(s): Tim Mueller (tmueller@jhu.edu)

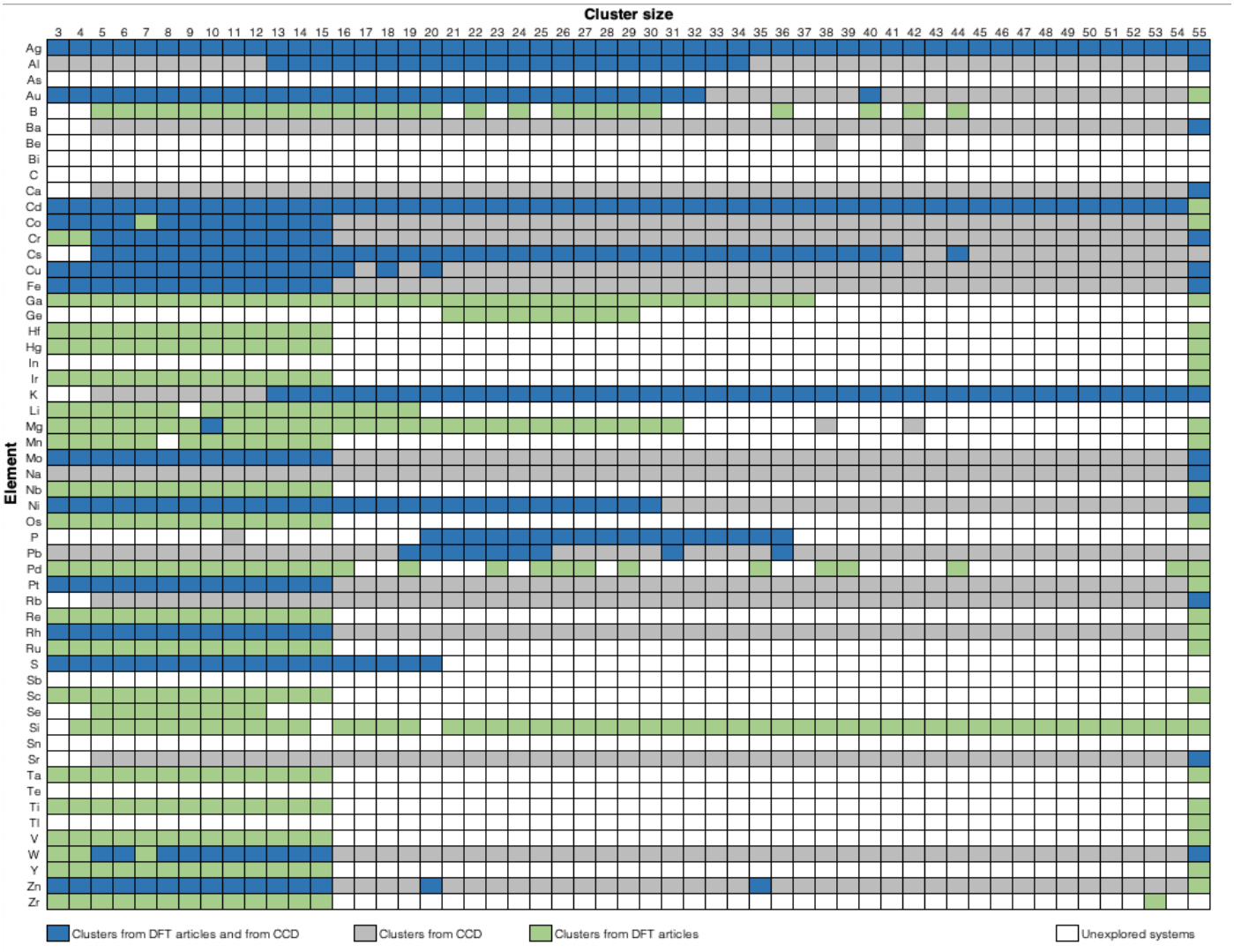

Figure S1. A summary of previous studies of the structures of elemental clusters with 3-55 atoms, including publications that used DFT to find atomic structures as well as systems covered in the Cambridge Cluster Database. The detailed literature sources are provided in Table 1 in the main article.

1. Identifying low energy clusters using the genetic algorithm

Low-energy atomic structures of clusters were identified using a genetic algorithm (GA), a global optimization technique inspired by the principles of natural selection [1]. We implemented our own GA code based off the Birmingham Parallel Genetic Algorithm (BPGA) with some variations [2, 3]. Figure S2 shows a schematic workflow of the genetic algorithm, details of which are described in our previous work [4].

The majority of GA searches in this work were performed using GA with pure DFT calculations (referred as GA_DFT in our previous work [5]). We evaluated multiple variations of the genetic algorithm as we filled the database. The final parameters and workflow were the same as we reported in reference [5], with some exceptions. First, we used a pool size of 10 and stopped the GA once the total number of clusters reached 1000. We also included "seeding" as a genetic operation in GA searches for sulfur clusters, with the ratios of pool clusters generated from seeding, mutation and crossover operations equal to 1:1:3. In the seeding operation, new structures are generated from seed structures which are typically known low-energy structures with different numbers of atoms. Atoms are either randomly added or subtracted from the seed structure until it reaches the target size. Seeding was included for sulfur clusters more than 10 atoms to reduce the chance of clusters being relaxed into discontiguous forms. Low-energy sulfur clusters with N atoms were used to seed searches for clusters with N+1 to N+5 atoms. For large sulfur clusters with 52 and 53 atoms, we initialized them as zig-zag rings, as suggested by the morphology of stable small-size clusters, and then collected the ones that are contiguous after relaxed by DFT. We accelerated the GA search for sulfur clusters with 51, 52, and 53 atoms with on-the-fly active learning (termed as "GA_AL" in our previous work [4]).

In the energy evaluation by DFT during the GA runs, we set the side length of each simulation box to ensure a distance of at least 10 angstroms between periodic images. For Na, K, Rb, Cs, Mg, Sr, and Ba, the distances are increased to at least 15 angstroms because they have larger atomic radii compared with the rest.

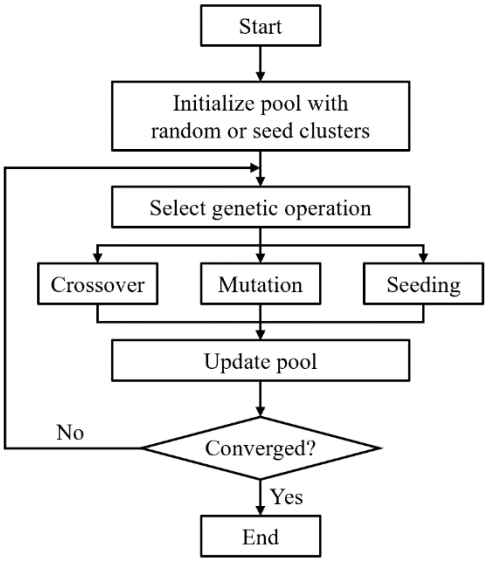

Figure S2. Schematic workflow of the genetic algorithm used for constructing the Quantum Cluster Database.

Table S1. Pseudopotentials used in VASP calculation to construct Quantum Cluster Database.

Element	Name (TITEL from POTCAR)
Ag	PAW_PBE Ag 02Apr2005
Al	PAW_PBE Al 04Jan2001
As	PAW_PBE As 22Sep2009
Au	PAW_PBE Au 04Oct2007
B	PAW_PBE B 06Sep2000
Ba	PAW_PBE Ba_sv 06Sep2000
Be	PAW_PBE Be 06Sep2000
Bi	PAW_PBE Bi 08Apr2002
C	PAW_PBE C 08Apr2002
Ca	PAW_PBE Ca_pv 06Sep2000
Cd	PAW_PBE Cd 06Sep2000
Co	PAW_PBE Co 02Aug2007
Cr	PAW_PBE Cr 06Sep2000
Cs	PAW Cs_sv_GW 23Mar2010
Cu	PAW_PBE Cu 22Jun2005
Fe	PAW_PBE Fe 06Sep2000
Ga	PAW_PBE Ga 08Apr2002
Ge	PAW_PBE Ge 05Jan2001
Hf	PAW_PBE Hf 20Jan2003
Hg	PAW_PBE Hg 06Sep2000
In	PAW_PBE In 08Apr2002
Ir	PAW_PBE Ir 06Sep2000
K	PAW_PBE K_pv 17Jan2003
Li	PAW_PBE Li 17Jan2003
Mg	PAW_PBE Mg 13Apr2007
Mn	PAW_PBE Mn 06Sep2000
Mo	PAW_PBE Mo 08Apr2002
Na	PAW_PBE Na 08Apr2002
Nb	PAW_PBE Nb_pv 08Apr2002
Ni	PAW_PBE Ni 02Aug2007
Os	PAW_PBE Os 17Jan2003
P	PAW_PBE P 06Sep2000
Pb	PAW_PBE Pb 08Apr2002
Pd	PAW_PBE Pd 04Jan2005
Pt	PAW_PBE Pt 04Feb2005
Rb	PAW_PBE Rb_pv 06Sep2000
Re	PAW_PBE Re 17Jan2003
Rh	PAW_PBE Rh 04Feb2005

Ru	PAW_PBE Ru 04Feb2005
S	PAW_PBE S 06Sep2000
Sb	PAW_PBE Sb 06Sep2000
Sc	PAW_PBE Sc 04Feb2005
Se	PAW_PBE Se 06Sep2000
Si	PAW_PBE Si 05Jan2001
Sn	PAW_PBE Sn 08Apr2002
Sr	PAW_PBE Sr_sv 07Sep2000
Ta	PAW_PBE Ta 17Jan2003
Te	PAW_PBE Te 08Apr2002
Ti	PAW_PBE Ti 08Apr2002
Tl	PAW_PBE Tl 08Apr2002
V	PAW_PBE V 08Apr2002
W	PAW_PBE W 08Apr2002
Y	PAW_PBE Y_sv 25May2007
Zn	PAW_PBE Zn 06Sep2000
Zr	PAW_PBE Zr_sv 04Jan2005

Table S2. List of nearest neighbor distance of all 55 elements used in this work.

Element	Nearest Neighbor Distance (Å)
Ag	2.895913
Al	2.855954
As	2.550183
Au	2.915931
B	1.666025
Ba	4.356367
Be	2.201364
Bi	3.099441
C	1.421378
Ca	3.817938
Cd	2.978816
Co	2.458547
Cr	2.446799
Cs	4.631936
Cu	2.548771
Fe	2.442529
Ga	2.525538
Ge	2.500974
Hf	3.12281
Hg	3.291589
In	3.279844
Ir	2.731446
K	4.145087
Li	2.967711
Mg	3.150178
Mn	2.127518
Mo	2.721704
Na	3.383385
Nb	2.870738
Ni	2.46851
Os	2.684334
P	2.197732
Pb	3.555145
Pd	2.774875
Pt	2.794795
Rb	4.492154
Re	2.74767
Rh	2.694163
Ru	2.640016

S	2.059424
Sb	3.083376
Sc	3.210087
Se	2.361713
Si	2.366088
Sn	2.877918
Sr	4.236609
Ta	2.858984
Te	2.903174
Ti	2.627395
Tl	3.421145
V	2.570188
W	2.740305
Y	3.541122
Zn	2.611721
Zr	3.162049

Table S3. Energy of an isolated atom for 55 different elements calculated using DFT

Element	E_{atom}(eV/atom)
Ag	-0.19857857
Al	-0.20506413
As	-1.69978222
Au	-0.180639
B	-0.28367486
Ba	-0.03341486
Be	-0.03684602
Bi	-1.32289577
C	-1.26448975
Ca	-0.05385721
Cd	-0.16384004
Co	-1.44899667
Cr	-2.33322218
Cs	-1.79835245
Cu	-0.23767466
Fe	-0.64208307
Ga	-0.19560421
Ge	-0.75666836
Hf	-3.47174222
Hg	-0.11848548
In	-0.17174232
Ir	-1.42697803
K	-0.15039732
Li	-0.29270415
Mg	0.00077358
Mn	-5.15415012
Mo	-3.05645278
Na	-0.21920527
Nb	-2.95767864
Ni	0.10158691
Os	-2.92048122
P	-1.87039627
Pb	-0.5951573
Pd	-1.46605907
Pt	-0.40789317
Rb	-0.13888444

Re	-4.59821658
Rh	-1.01346064
Ru	-1.54726228
S	-0.88115573
Sb	-1.40767101
Sc	-1.98139794
Se	-0.76220148
Si	-0.81939324
Sn	-0.64363261
Sr	-0.06485411
Ta	-3.66265903
Te	-0.65162829
Ti	-2.24955619
Tl	-0.15061158
V	-2.47033984
W	-4.10529928
Y	-2.23495881
Zn	-0.16143445
Zr	-2.21436277

2. DFT Calculation strategies for treating magnetic clusters

The final magnetic state of a cluster may be in a local minimum that is not the global minimum. To reduce the chance of this happening, for each of the magnetic elements (Fe, Mn, Co, Ni, Ru, Rh, V, Cu, and Cr) we evaluated the effect of using different initial magnetic moments (+1 μ_B/atom, +2μ_B/atom, +3μ_B/atom, and +5 μ_B/atom), as set using the MAGMOM parameter in VASP. We evaluated the final magnetic states and energies of 2228 clusters with between 3-55 atoms selected from an early version of the QCD. We found that for some elements (e.g. Co) the final state is almost always independent of MAGMOM, whereas for others (e.g. Cr) it is common for the cluster to get trapped in local minima, making it important to initialize with multiple MAGMOM values. Often clusters with an odd-number of electrons would have a net magnetic moment of 1 μ_B per cluster, and clusters with an even number of electrons would have a net magnetic moment of 0 μ_B, but this behavior was largely independent of MAGMOM. In Table S4, we list the elements which retain non-zero magnetic moments after relaxation as well as the effects of multiple magnetic initializations (as shown in Table S5) on magnetic moments after relaxation.

For each element we selected a subset of MAGMOM values that included the lowest DFT energy for a given size (within 1 meV/atom) at least 97% of the time (Table S4). These values were used to initialize the DFT calculations. For elements for which multiple MAGMOM values were used, for each cluster the calculation that resulted in the lowest energy was added to the database.

Table S4. Finite magnetic moments after relaxation with default and multiple magnetic initializations per atom using the MAGMOM flag in VASP. This analysis was done considering clusters with all sizes between 3-55 atoms for each magnetic element.

Element	Mean absolute magnetizations in (μ_B/atom) with default magnetic initialization in VASP	STD DEV	Mean absolute magnetizations in (μ_B/atom) with multiple magnetic initialization in VASP	STD DEV
Fe	3.05	0.17	2.77	0.36
Mn	2.76	0.63	2.33	1.04
Co	1.92	0.18	1.92	0.104
Ni	0.8	0.12	0.77	0.081
Ru	0.7	0.63	0.69	0.388
Rh	0.69	0.32	0.654	0.341
V	0.39	0.92	0.404	0.455
Cu	0.03	0.06	0.036	0.056
Cr	0.45	0.42	0.954	0.685

Table S5. Magnetic initialization schemes for magnetic elements.

Element	Magnetic Initializations (MAGMOM in INCAR files)
Fe	+3, +5
Mn	+1, +2, +5
Co	+3
Ru	+1, +3, +5
Rh	+1, +2, +3
V	+1, +2, +3
Cu	+3
Cr	+1, +2, +3, +5

3. Effect of Spin Orbit Coupling

To determine whether it is necessary to use spin-orbit-coupling (SOC) to identify low-energy structures for heavy-metal elements, we performed additional PBE + SOC calculations on the lowest energy clusters of 11 different heavy-metal elements (Au, Bi, Hf, Hg, Ir, Os, Pb, Pt, Re, Ta, and Tl) with sizes ranging from 3 to 55 atoms. Additionally, we choose the six (where available) lowest energy isomers for small (size 10), medium (size 30), and large (size 55) clusters to study the effect of SOC on relative ordering. The PBE+SOC calculated energies linearly correlated with PBE computed energies (Figure S3). We also developed a cheap proxy to evaluate PBE+SOC computed energies from PBE computed energies for these 11 different elements using linear regression. The conversion factors (slope and intercept) for converting PBE+SOC energies from PBE energies for these elements are provided in Figure S4. We also found that the use of SOC has little effect on the energetic rankings of isomers (Figure S5, Figure S6, and Figure S7). For these reasons, and to maintain consistency in the database, we have not included the effects of SOC in the calculated energies in the QCD. For three small clusters (with compositions Pb_{10}, Tl_{10} and Hf_{10}), the PBE + SOC calculations relaxed to clusters that were structurally dissimilar to the relaxed PBE structures (as determined by similarity scores greater than 0.3). Because the energies for the two methods were not calculated for the same structure, these clusters are excluded from the comparisons in Figure S5, Figure S6 and Figure S7.

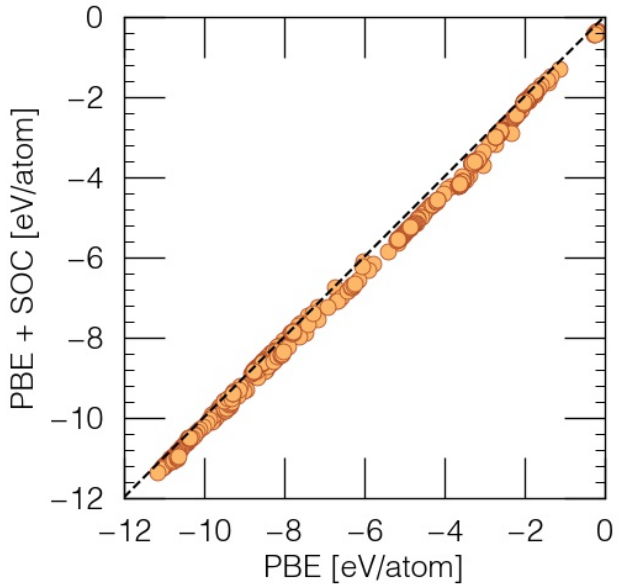

Figure S3. Total energies from PBE calculations vs. energies from PBE+SOC calculations for 11 heavy-metal elements (Au, Bi, Hf, Hg, Ir, Os, Pb, Pt, Re, Ta, and Tl).

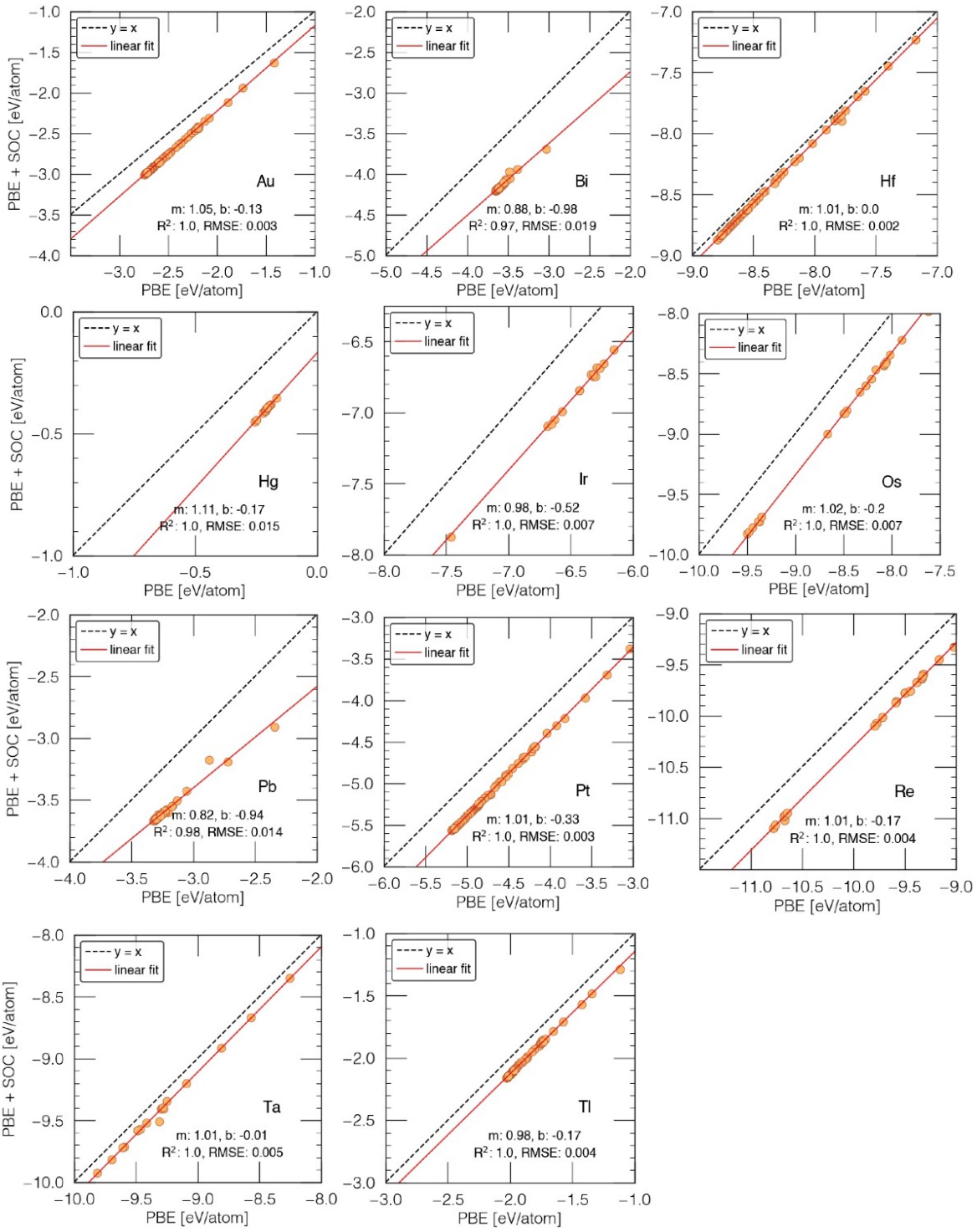

Figure S4. PBE + SOC energy conversion from PBE energies using least-square linear regression fit E(PBE+SOC) =m*E(PBE)+b. The linear parameters (slope m and intercept b) along with their RMSE error are provided for each element in the legend of each subplot.

4. Effect of PBE + SOC on relative ordering

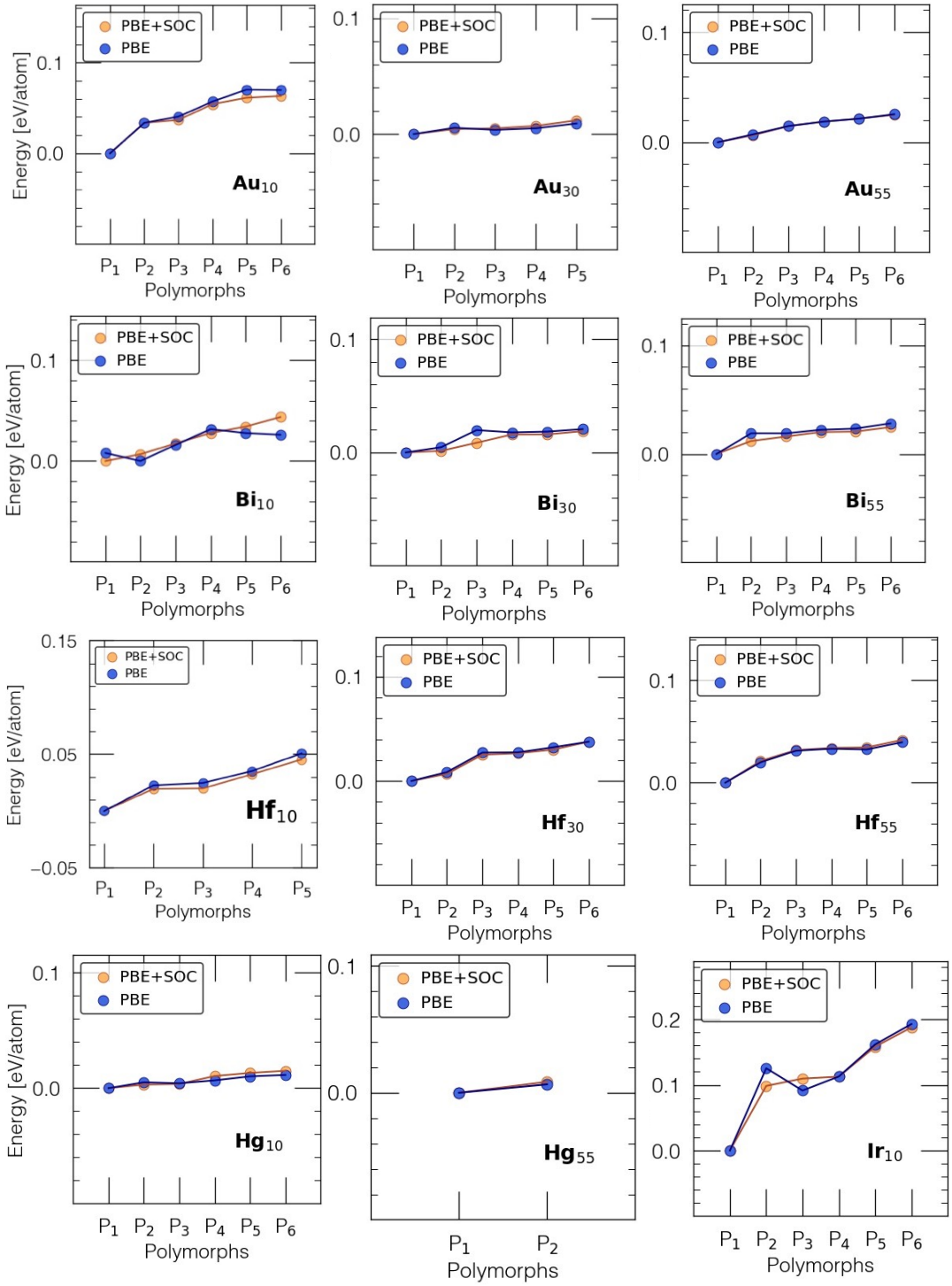

Figure S5. Effect of PBE + SOC on relative energy ordering among isomers of Au, Bi, Hf, Hg, and Ir clusters. Relative energies with respect to the lowest energy isomers are plotted for each polymorph, and the polymorphs are sorted based on the PBE+SOC energies. One Hf_{10} structure was not included in this plot because PBE and PBE-SOC relaxed to dissimilar structures.

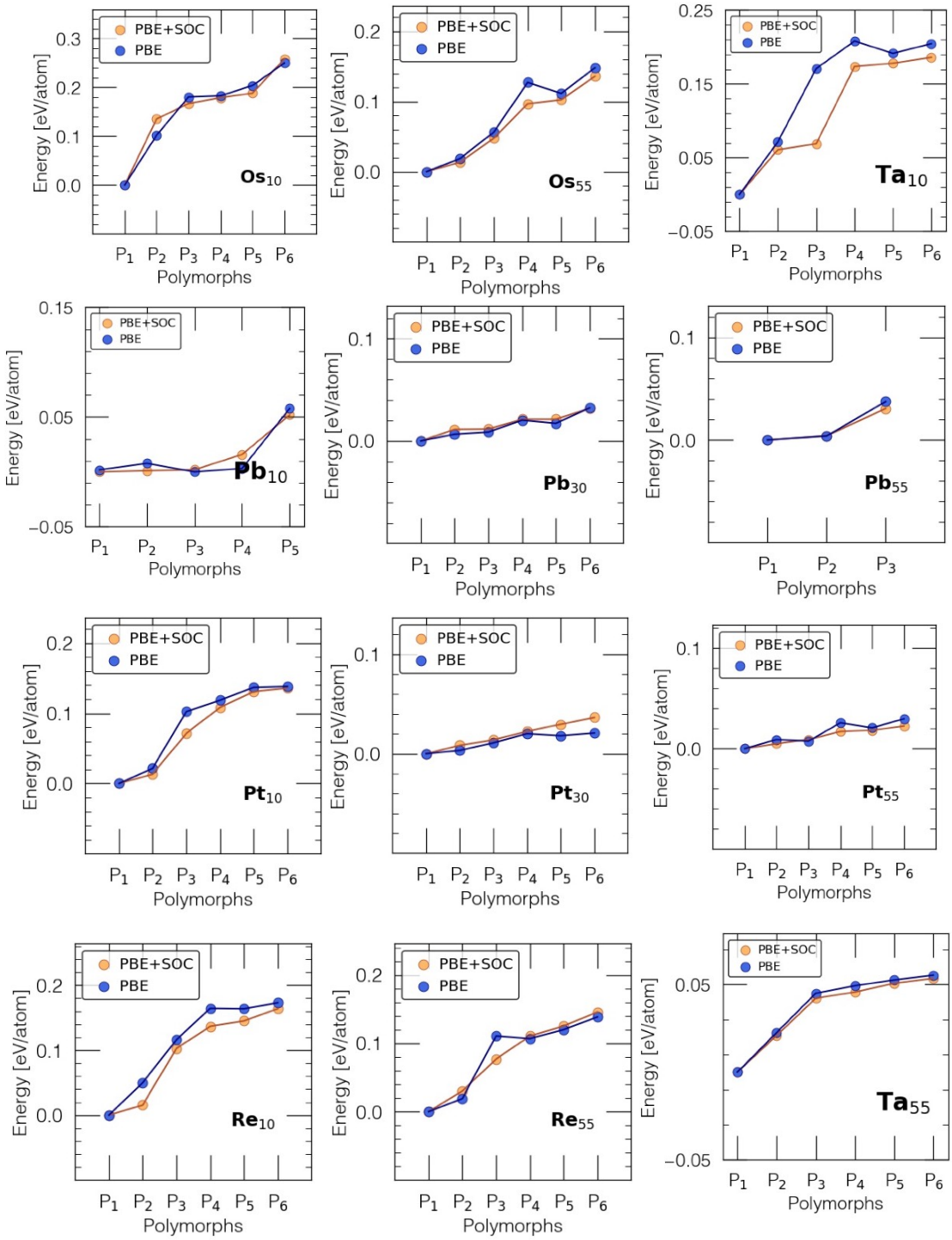

Figure S6. Effect of PBE +SOC on relative energy ordering among the isomers of Os, Ta, Pb, Pt, and Re clusters. Element name and their sizes are stamped in each subplot. Relative energies with respect to lowest energy isomer are plotted against each polymorph which is sorted w.r.t PBE+SOC energies. One Pb_{10} structure was not included in this plot because PBE and PBE-SOC relaxed to dissimilar structures.

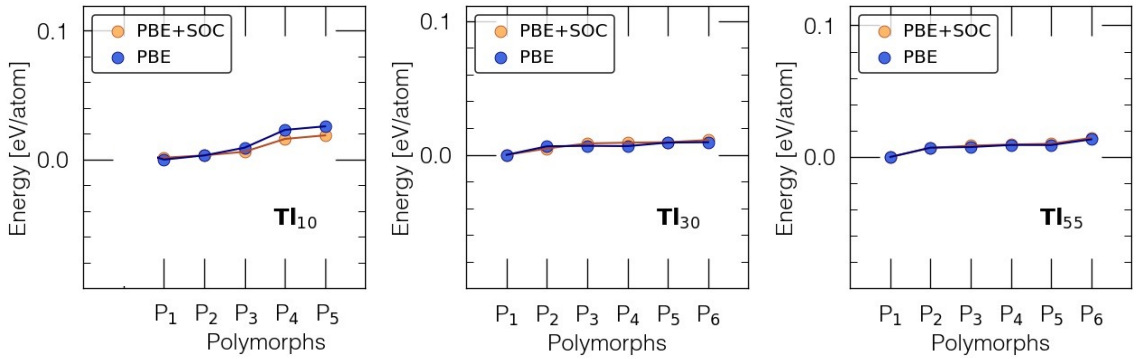

Figure S7. Effect of PBE +SOC on relative energy ordering among the isomers of Tl clusters. Element name and their sizes are stamped in each subplot. Relative energies with respect to lowest energy isomer are plotted against each polymorph which is sorted w.r.t PBE+SOC energies. One Tl_{10} structure was not included in this plot because PBE and PBE-SOC relaxed to dissimilar structures.

5. Box Size Analysis

To investigate whether the box sizes used in the QCD are sufficiently large, we performed benchmarks on 1335 clusters from 49 different elements with 3-55 atoms selected from the QCD. In all cases, energies were converged within 2 meV/atom, with a root mean square box size error of just 0.118 meV/atom.

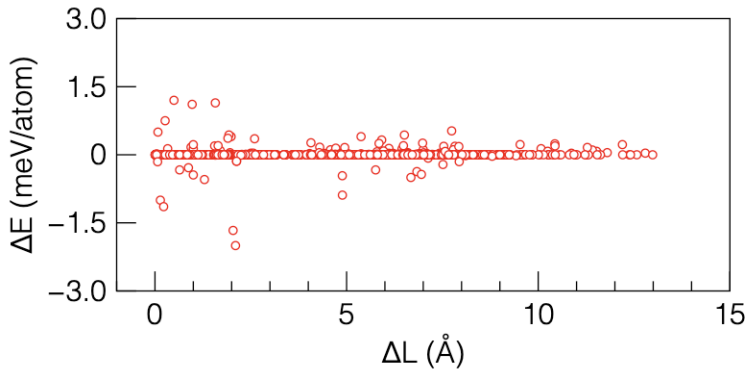

Figure S8. Effect of change in computational box size from currently used box length in the database on predicting total energies.

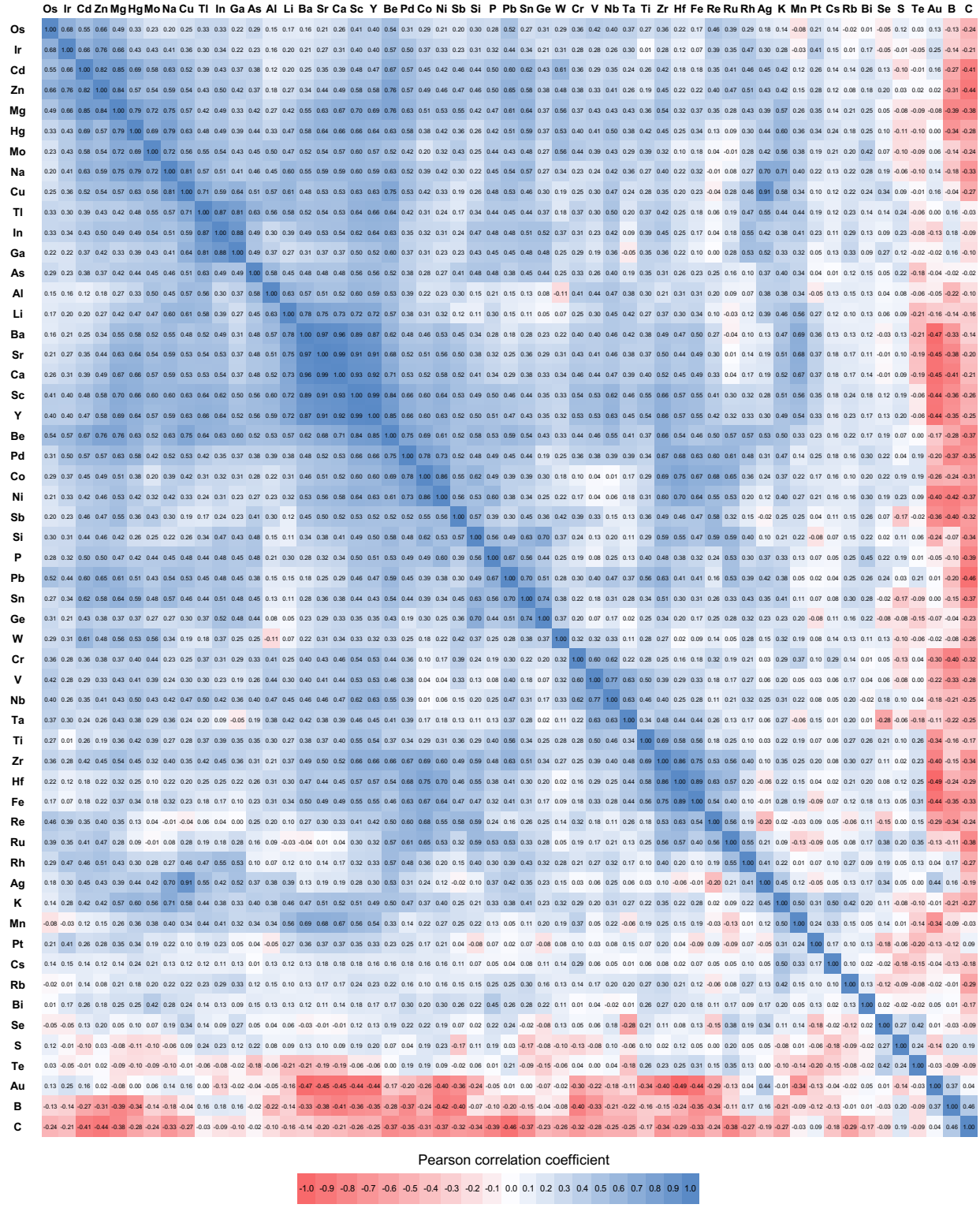

Figure S9. The Pearson correlation coefficients for all 55 elements calculated from the energies of relaxed structures, sorted by the mean absolute difference in correlation coefficients calculated from unrelaxed (top) and relaxed (bottom) structures. The data used to generate this figure are available as a csv in the SI and through the website interface.

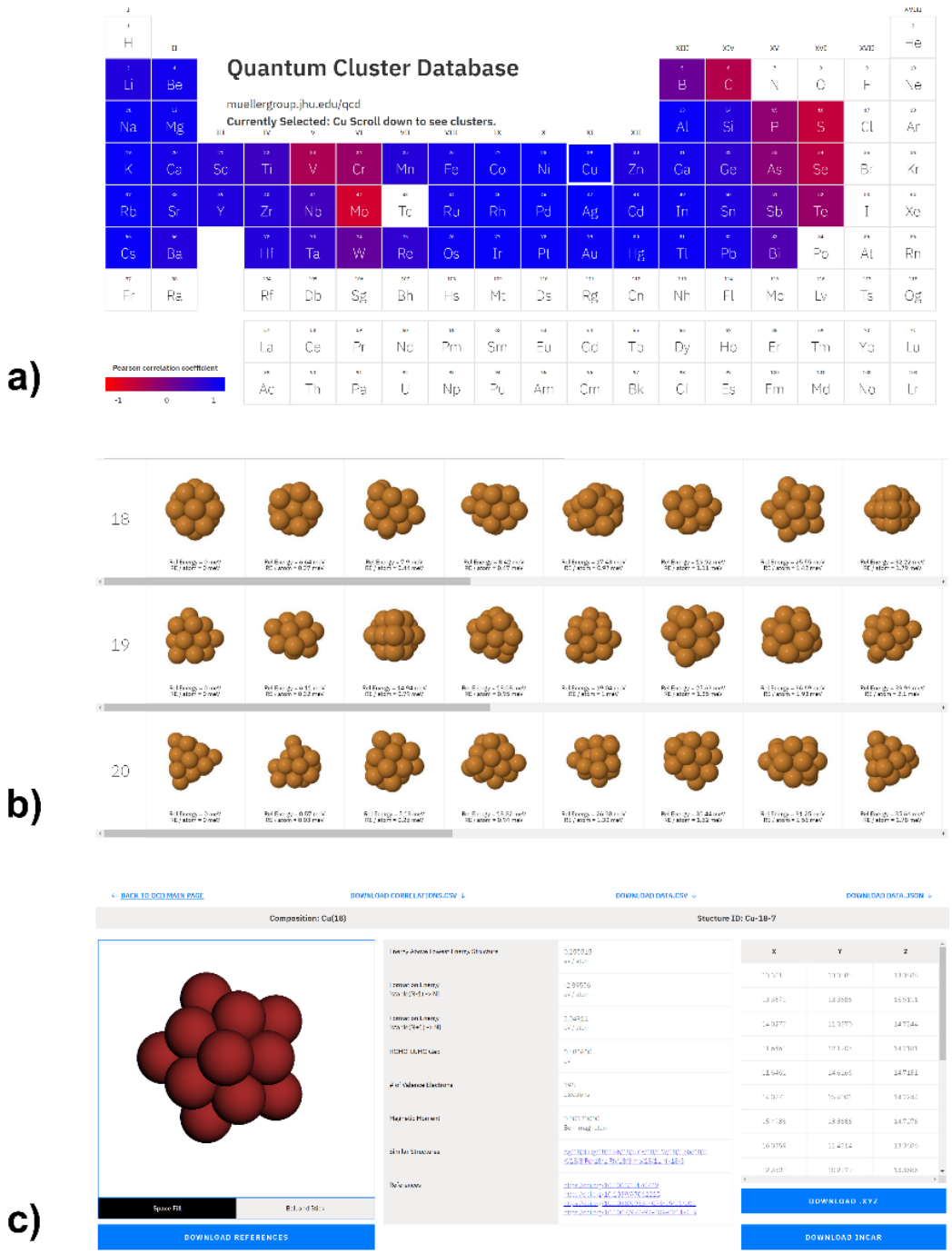

Figure S10. (a) Home page interface for the Quantum Cluster Database. (b) Website interface after we select a particular element (Cu in the picture). (c) Element view of a selected cluster and its properties.

REFERENCES

1. Holland, J.H., *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. 1 ed. 1975: Ann Arbor, MI: University of Michigan Press.
2. Johnston, R.L., *Evolving better nanoparticles: Genetic algorithms for optimising cluster geometries.* Dalton Transactions, 2003(22): p. 4193-4207.
3. Shayeghi, A., et al., *Pool-BCGA: a parallelised generation-free genetic algorithm for the ab initio global optimisation of nanoalloy clusters.* Physical Chemistry Chemical Physics, 2015. **17**(3): p. 2104-2112.
4. Wang, Y., et al., *Accelerated Prediction of Atomically Precise Cluster Structures Using On-the-fly Machine Learning.* ChemRxiv, 2021.