This article can be cited before page numbers have been issued, to do this please use: I. S. Steshin, S. Panteleev, I. V. Petukhov and S. K. Ignatov, *Phys. Chem. Chem. Phys.*, 2025, DOI: 10.1039/D5CP02189H.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

Parametrization of Gaussian Approximation Potential for the Global Optimization of Magnesium Clusters Mg_N ($N \leq 100$)

Ilya S. Steshin[1,2], Sergey V. Panteleev[1], Igor V. Petukhov[2], Stanislav K. Ignatov[1]

[1] Lobachevsky University of Nizhny Novgorod, Nizhny Novgorod, Russia
[2] Volga State University of Technology, Yoshkar-Ola, Russia

ABSTRACT

A two-stage GridSearch combined with active learning was employed to optimize GAP model parameters for Mg clusters, enabling reliable structural predictions in the extrapolative domain Mg_n, $n > 50$. Global optimization using the parameterized GAP model revealed energetically favorable Mg_{51}-Mg_{53} clusters, showing early onset of pyramidal core formation previously reported only from Mg_{54}. Global optimization identified new global minima candidates for the "magic" Mg_{59}, Mg_{69}, Mg_{74} and Mg_{99} clusters. The presence of hcp-like motifs doesn't significantly influence structural stability in clusters with Mg_n, $n < 100$, as no structural differences were observed between GM "magic" clusters and others of similar size.

INTRODUCTION

Subnano- and nanoscale clusters have unique physicochemical properties that distinguish them from nanoparticles, mono- and polycrystalline materials. These properties include enhanced catalytic activity and adsorption capacity, discrete energy levels, nonlinear optical behavior, and a strong dependence of structural and electronic characteristics on the number of atoms in the cluster. These features arise primarily due to lower coordination numbers and the prevalence of atoms with unsaturated valence. By changing the size and shape of a cluster, one can tune these properties[1–8].

The most significant variations in structure, catalytic activity, and reactivity are observed in clusters containing between 2 and 150 atoms (the subnanometer size range), where the influence

1

of each individual atom is significant[1,3]. For instance, in Pt_n clusters, the adsorption energy approaches its limiting value (i.e., the adsorption energy on a crystalline surface) only when $n > 147$[9].

At the same time, metal clusters in this size range exhibit a high degree of structural variability, and predicting their structures using simple chemical or physical principles remains an unresolved challenge. Reliable predictions of subnanometer cluster structures and their properties are currently only feasible through direct quantum chemical computations at the DFT or *ab initio* level. However, such calculations, even for clusters with a relatively small number of atoms, are computationally expensive and complicated by issues such as self-consistent field (SCF) convergence and the need to consider multiple spin states.

In recent decades, machine learning (ML)-based interatomic potentials have emerged as powerful tools for modeling interatomic interactions. These potentials, now widely applied across chemistry, physics, and materials science, can achieve so-called quantum accuracy due to the large number of parameters involved, enabling them to approach the precision of DFT across a broad range of nuclear configurations[10–13]. Nonetheless, the accuracy, speed, and reliability of a given potential strongly depend on the ML algorithm employed and the representation of atomic configurations—that is, the choice of descriptors used as input features[10,14–16].

Currently, several approaches of ML-based interatomic potentials have been developed, including: GAP[17], HDNNP[18], SNAP[19], MTP[20] and ACE[21]. All of these models are trained on datasets including quantum chemical calculations of the systems of interest. The quality and diversity of the training dataset critically influence the performance of the resulting model, making dataset preparation a key step in the construction of a robust interatomic potential[10,22,23].

Among metal clusters, magnesium clusters represent an interesting subject of investigation and modeling. These clusters possess a range of useful properties, including a high hydrogen adsorption capacity, making them promising candidates for hydrogen storage applications[24,25]. At

present, a significant amount of data on the structure of Mg_n $n <50$ clusters has been accumulated[26–31], extending this research to larger clusters $n > 50$ remains computationally demanding.

Therefore, the present study aims to develop a GAP-based interatomic potential with high accuracy and extrapolation capability to facilitate the efficient search for low-energy structures of Mg_n $n > 50$ clusters and to enable further investigation of their properties. In this work, we employed extended datasets previously generated through direct global optimization (GO) of Mg_{27}-Mg_{32}[27] clusters, as well as additional datasets for Mg_{33}-Mg_{56}[28,29], obtained from other studies, to train the GAP model. Based on the resulting parametrized GAP model, we performed GO in the nuclearity range of $n = 40$-60, 68-70, 73-75, 98-100.

MATERIALS AND METHODS

Gaussian Approximation Potential (GAP)

GAP (Gaussian Approximation Potential) is a type of ML interatomic potential used to construct potential energy surfaces (PES) by employing Gaussian Process (GP) regression to predict the energies, forces, and stresses in atomic systems[23,32].

In the process of fitting the PES for atomic systems, it is assumed that the total quantum-mechanical energy can be decomposed into local energy contributions E_d which depend on descriptors x[10,22,32]

$$E_{total} = \sum_{d}^{N_{desk}} E_d(x)$$

where N_{desk} — denotes the number of descriptors describing the local contribution to the total energy E_{total}. In GAP, each local energy term E_d is modeled as an independent GP

$$E_d(x) = \sum_{m=1}^{M_{desk}} c_m k_d(x, x_m)$$

where M_{desk} — is the number of representative points of the descriptor d, k_d — is the kernel (covariance or similarity function), and c_m — fitting coefficients. These coefficients c_m are fitted using a reference database of atomic configurations, in which the corresponding total energies and derivative quantities such as forces are obtained from quantum mechanical calculations.

The kernel function k_d can take various forms, such as the squared exponential kernel k_d^{SE} or the dot-product kernel k_d^{DP}

$$k_d^{SE} = e^{-r^2}$$

$$r = \sqrt{\sum_{i=1}^{D} \frac{(x_i - x_i')^2}{2\theta_i^2}}$$

$$k_d^{DP} = (x \cdot x')^\zeta$$

where θ and ζ – are the hyperparameters of the GAP model and x, x' - are a pair of descriptors of dimensionality D. In the present work, Smooth Overlap of Atomic Positions (SOAP)[23] and two-body (2b)[33] descriptors were used to describe the atomic environments in Mg clusters.

Training Data for GAP Model

The training dataset for the GAP model was composed of Mg_{27}-Mg_{55} cluster structures collected from multiple studies. The majority of the structures were taken from a GO study of Mg_n, $n<32$ clusters[27,34,35].

The dataset included only the unique structures of the Mg_{27}-Mg_{32}, defined as those for which differences in sorted interatomic distances or principal moments of inertia exceeded 5% for all pairs of structures in the dataset. Additionally, structures with a minimum interatomic distance below 2.5 Å or a maximum above 15 Å were excluded, as well as metastable structures with positive binding energy.

The focus on the nuclearity range $n = 27$-32 was primarily due to the limited availability of structural data for larger clusters Mg_{33}-Mg_{56}[28,29]. A total of 2923 unique structures were selected for Mg_{27}-Mg_{32}, including 54 structures for Mg_{31} and Mg_{32}.

For Mg_{33}-Mg_{56}, 565 structures were sourced from two studies[28,29] and consisted of low-lying isomers; for Mg_{56}, only the global minimum (GM) structure was available[28]. Single-point (SP) energy calculations were performed on these structures with a SCF convergence threshold of 10^{-5} Hartree. Additionally, 316 SP-calculated structures of Mg_{46}-Mg_{50} were included in the dataset through an active learning process using the same SCF convergence criteria.

The data were divided into two datasets: a training dataset composed of clusters Mg_{27}-Mg_{40} and Mg_{46}-Mg_{50} and a test dataset composed of clusters Mg_{41}-Mg_{45} and Mg_{51}-Mg_{55}. This partitioning allowed for evaluation of both the interpolation and extrapolation capabilities of the model. In total, the datasets used for training and testing consisted of 3556 and 248 structures, respectively.

All quantum chemical calculations were performed using density functional theory (DFT) with the GAUSSIAN software package[36] at the BP86/6-31G(d) level of theory. This level has previously shown good agreement with higher-level MP2 and CCSD calculations for small magnesium clusters[30,31], and has been successfully applied to determine the structure and physicochemical properties of Mg_2-Mg_{32} clusters[27,34].

Selection of Parameters for the GAP Model

A two-stage GridSearch[37–39] process followed by active learning[22,23] was used to optimize the GAP model parameters (Figure 1). GridSearch is an efficient method for parameter tuning and is widely used in regression and classification tasks[37,40–42]. The root-mean-square error of the binding energy $RMSE(E_{bonds})$ for Mg_{55} clusters was used as the primary metric for model selection. Mg_{55} clusters provide a dataset with the maximum degree of model extrapolation, making them particularly suitable for testing purposes:

$$E_{bonds}^{GAP/DFT} = E_{total}^{GAP/DFT} - n \cdot E_{total}^{DFT(Mg_1)}$$

$$\Delta E_{bonds} = \frac{E_{bonds}^{GAP} - E_{bonds}^{DFT}}{n}$$

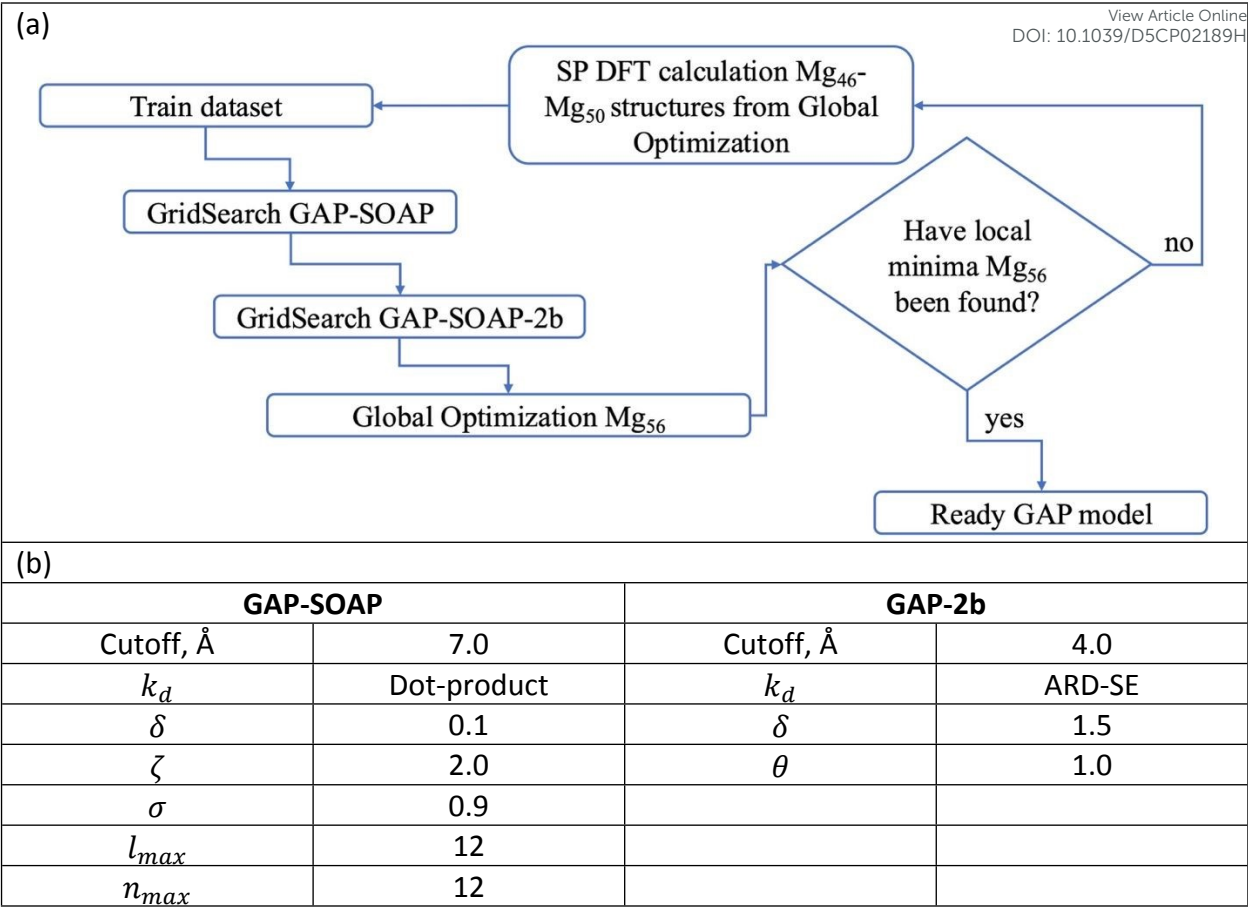$$RMSE(E_{bonds}) = \sqrt{\frac{\sum_i^N (\Delta E_{bonds})^2}{N}},$$

where n – is the number of atoms in the Mg_n cluster, N – is the number of structures in the Mg_n dataset.

In the first stage, GridSearch was applied to optimize the parameters of the SOAP descriptors, which are necessary for constructing the many-body component of the potential. The second stage involved optimizing the parameters of the two-body (2b) descriptors, using the previously determined SOAP parameters, to improve the description of short-range interactions[23,33]. Throughout both stages, the full training dataset was used for the fitting of the GAP-SOAP and GAP-2b models.

Following this, a GO was conducted for the Mg_{56} cluster, which is the most extrapolative system relative to the training data, using the Minima Hopping (MH) algorithm[43,44]. If the known GM structure of Mg_{56} was not identified, GO procedures were performed for Mg_{46}-Mg_{50} clusters, followed by SP DFT calculations of several (from 5 to 20) isomers. Then Mg_{46}-Mg_{50} structures were added into the training dataset.

The two-stage GridSearch and active learning cycle was repeated until the GAP model successfully predicted the structure corresponding to the known GM of the Mg_{56} cluster.

(a)

(b)

GAP-SOAP		GAP-2b	
Cutoff, Å	7.0	Cutoff, Å	4.0
k_d	Dot-product	k_d	ARD-SE
δ	0.1	δ	1.5
ζ	2.0	θ	1.0
σ	0.9		
l_{max}	12		
n_{max}	12		

Figure 1. (a) Fitting of the GAP model using a two-stage GridSearch and active learning approach (b) Optimal GAP model parameters selected after GridSearch.

Global Optimization of Magnesium Clusters

There are two primary approaches to using ML interatomic potentials in GO tasks. The first approach involves constructing an "on-the-fly" potential. In this case, the potential is either developed during the GO process using quantum chemical methods for each nuclearity or trained on a precompiled dataset that includes structures with varying numbers of atoms. The potential is actively used at each step of GO, and each local minimum identified is subsequently refined through quantum chemical calculations. The advantage of this approach is that the ML interatomic potential does not need to achieve high accuracy, as each local minimum is ultimately verified by quantum mechanical methods. However, this strategy does not fully resolve the computational burden for larger systems[45–47]. As system size increases, the number of necessary quantum chemical calculations remains significant, since tens of calculations may still be required to reach the GM.

The second approach focuses on fitting a more accurate ML interatomic potential that provides a set of k initial candidate structures, among which the GM is expected to be present, subsequently verified via quantum chemical methods. This approach demands more careful selection of model hyperparameters, construction of a robust training dataset, and the use of active learning[22,23,48–50]. Its key advantage lies in maintaining a constant number of quantum chemical calculations regardless of system size, as the entire GO process is conducted using the ML potential (Figure 2).
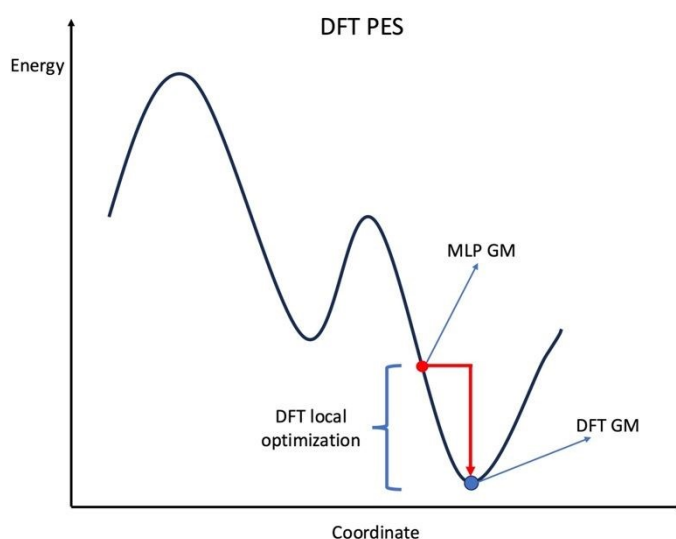
Figure 2. Local DFT optimization of the GM structure found during GO using a machine learning potential (MLP)

In this work, the second approach was implemented, enabling the identification of new GM for Mg_{51}-Mg_{53} clusters. The MH algorithm was employed for GO. The MH was configured with the following parameters: a temperature of 2000 K, 200 molecular dynamics (MD) iterations, and a stopping criterion (*mdmin* of original program) equal to 60 – the number of minima visited before stopping the MD during one iteration.

For each nuclearity, the MH algorithm was initiated from five randomly generated structures differing either in the sorted interatomic distances or in the principal moments of inertia, with differences of at least 5%. In some cases, isomers from previous studies were also used as starting points to ensure that the GAP model didn't induce significant deformation of known structures during GO.

The resulting set of structures from GO was processed to identify structurally unique configurations based on the analysis of Goedecker's fingerprints[51] — that is, the eigenvalues of the overlap matrix of auxiliary Gaussian s,p-functions centered on atomic nuclei. This analysis was performed using the [https://github.com/skignatov/StrDiscr2-release] tool, previously successfully applied to GO of Mg_2-Mg_{32} clusters[27]. After identifying unique isomers, the ten lowest-energy structures were selected as starting points for local DFT optimizations.

Calculation of the Binding Energy of Magnesium Clusters

To analyze the variation of energetic properties with changes in the number of atoms, the average binding energy per atom E_{bond}, was calculated according to the following equation:

$$E_{bond} = \frac{E_{tot} - n \cdot E_{Mg1}}{n}$$

where E_{tot} – is the total energy of the cluster, n – is the number of atoms in the cluster, $E_{Mg1} = -200.0696953$ Hartree – is the total energy of a single Mg atom in vacuum (DFT: BP86/6-31G(d)).

Comparison of the Atomic Environments of Magnesium Clusters with Crystalline Lattices

The similarity of atomic environments in Mg_n clusters to crystalline lattices - hcp (a=3.209 Å, c = 5.211 Å), fcc (a=3.209 Å), bcc (a=3.209 Å) was assessed using SOAP descriptors ($l_{max} = 12$, $n_{max} = 12$, $\sigma = 0.4$) with cutoff radii of 3.209 Å, 5.211 Å - corresponding to the Mg hcp lattice parameters[52] and 4.0 Å, 7.0 Å the optimal cutoff radii determined via GridSearch. Cosine distance d_{cosine} between SOAP vectors was used as a measure of similarity between the atomic environments of the cluster and the crystalline lattices. Atomic environments were considered equal at a threshold $d_{cosine} < 0.01$. The threshold of 0.01 was defined as the minimum difference between atomic environments among crystal lattices at the cutoff radii - 3.209, 4.0, 5.211 and 7.0 Å.

$$d_{cosine} = 1 - \frac{v_1 \cdot v_2}{||v_1||\ ||v_2||},$$

где v_1, v_2 - are the normalized to the [0, 1] range SOAP vectors , $v_1 \cdot v_2$ - the dot product of vectors v_1 and v_2, $||v_1||$, $||v_1||$ - are the vectors v_1 and v_2 norm. Atomic environments were considered similar when the d_{cosine} between their SOAP vectors was less than 0.05. The construction of crystalline lattices and the calculation of SOAP vectors were performed using the python libraries ASE[53] and DScribe[54].

Physical Chemistry Chemical Physics Accepted Manuscript

RESULTS AND DISCUSSION

Model training

The deviation in energy estimates for the training and test datasets, predicted by the trained GAP model with optimized parameters, relative to the DFT results is presented in Figure 3. The average $RMSE(E_{bonds})$ across all cluster sizes is 0.149 and 0.107 *kcal/(mol·atom)* for the test and training datasets, respectively. For the training dataset excluding Mg_{27}–Mg_{30}, the average $RMSE$ (E_{bonds}) is 0.121 *kcal/(mol·atom)*. The Pearson correlation coefficient for the training dataset was > 0.90 in most cases, and > 0.85 for the test dataset (Figure 3 (a, b)). Regression plots of individual n show good convergence with DFT for training datasets (Figure 3 (c, d)). For interpolation and extrapolation of individual n datasets, there is an upward bias in cluster energy estimates, but the correlation remains high (Figure 3 (e, f)).

These metrics are comparable to the accuracy achieved by other machine learning models applied to similar tasks of cluster structure prediction: $RMSE(Nb_nO_m;E_{bonds}) = 0.992$ *kcal/(mol·atom)*[12], $RMSE(Au_n;E_{bonds}) = 0.121$ *kcal/(mol·atom)*[48], $RMSE(Au_n;E_{bonds}) = 0.208$ *kcal/(mol·atom)*[49], $RMSE(Na_n;E_{bonds}) = 0.208$ *kcal/(mol·atom)*[55], $RMSE(Cu_n;E_{bonds}) = 0.090$ *kcal/(mol·atom)*[56], $RMSE(Mg_nH_m;E_{bonds}) = 0.721$ *kcal/(mol·atom)*[57].
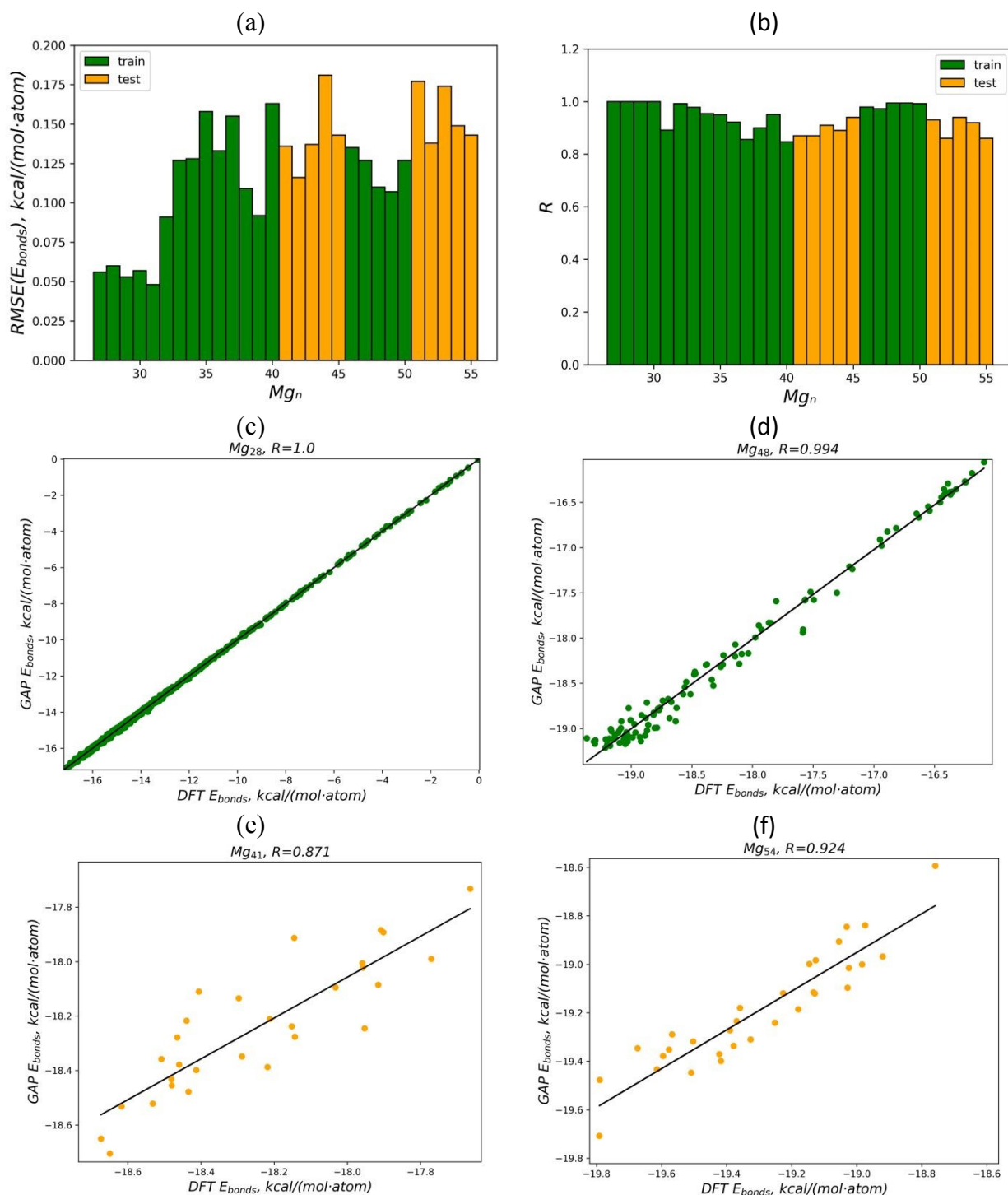
Figure 3. Prediction accuracy of the model on the training and test datasets: (a) $-$ $RMSE$ (E_{bonds}), (b) $-$ Pearson correlation coefficient. Regression plots for training dataset: (c) - Mg_{28}, (d) $-$ Mg_{48} and test dataset: (e) $-$ Mg_{41}, (f) $-$ Mg_{54}.

GO of Mg$_{42}$, Mg$_{44}$ and Mg$_{46}$-Mg$_{56}$ clusters

The comparison of the binding energy per-atom from DFT calculations of the 10 lowest-energy isomers found during the GO of Mg$_{42}$, Mg$_{44}$, Mg$_{46}$-Mg$_{56}$ clusters using the parameterized GAP model with previously known GMs[28,29] is shown in Figure 4(a). The largest scatter of energies after DFT relaxation within a single nuclearity is observed for Mg$_{42}$ and Mg$_{44}$. In contrast, a relatively tighter energy redistribution after DFT relaxation is seen for clusters with $n > 50$ such as Mg$_{52}$ and Mg$_{53}$ (Figure 4(b)). Changes in the ranking of the energies of the structures among the 10 lowest isomers after DFT relaxation are observed for both Mg$_{42}$, Mg$_{44}$ and Mg$_{52}$, Mg$_{53}$. The distribution of energies for Mg$_{51}$-Mg$_{56}$ is comparable to that for the training dataset clusters Mg$_{46}$-Mg$_{50}$ (Figure 4(a)).

Thus, the GAP model shows weak interpolation properties, meaning it demonstrates lower accuracy when predicting structures of smaller clusters Mg$_{42}$, Mg$_{44}$, that were not included in the training dataset. The number of metastable structures obtained during the MH optimization was in the range of 400-600 structures regardless of the area of applicability of the GAP model: interpolation, training or extrapolation dataset (Figure 4(c)).

(a)

(b)

(c)
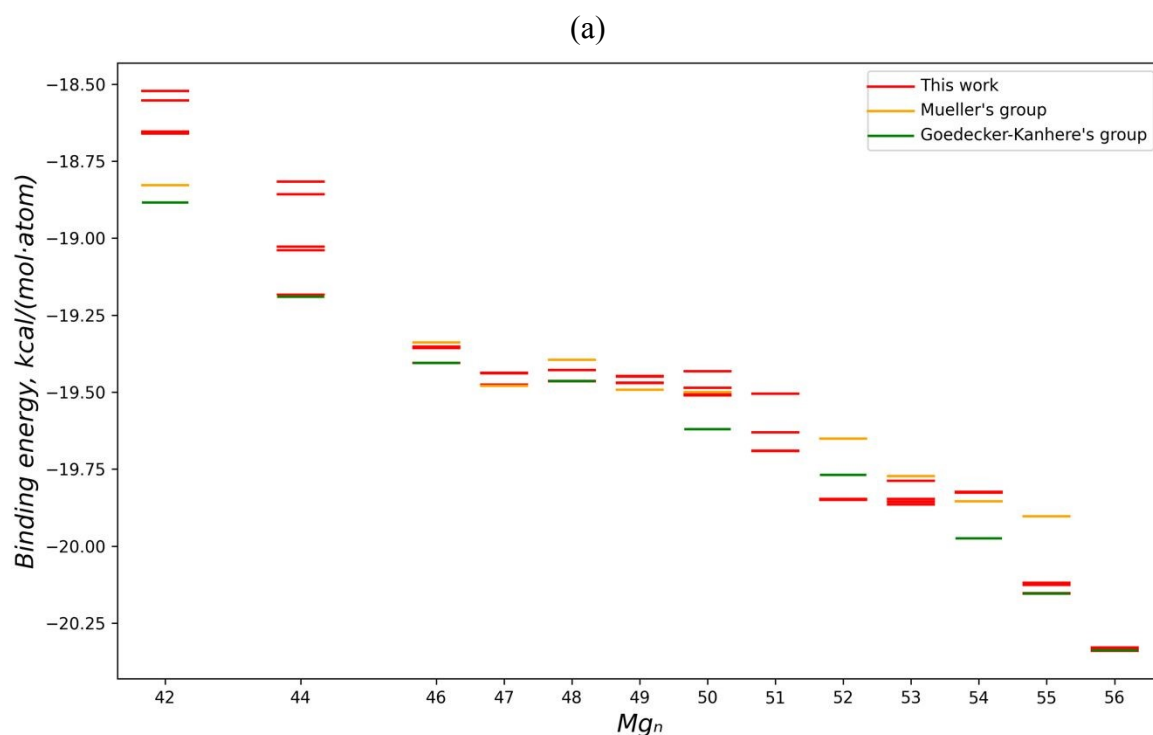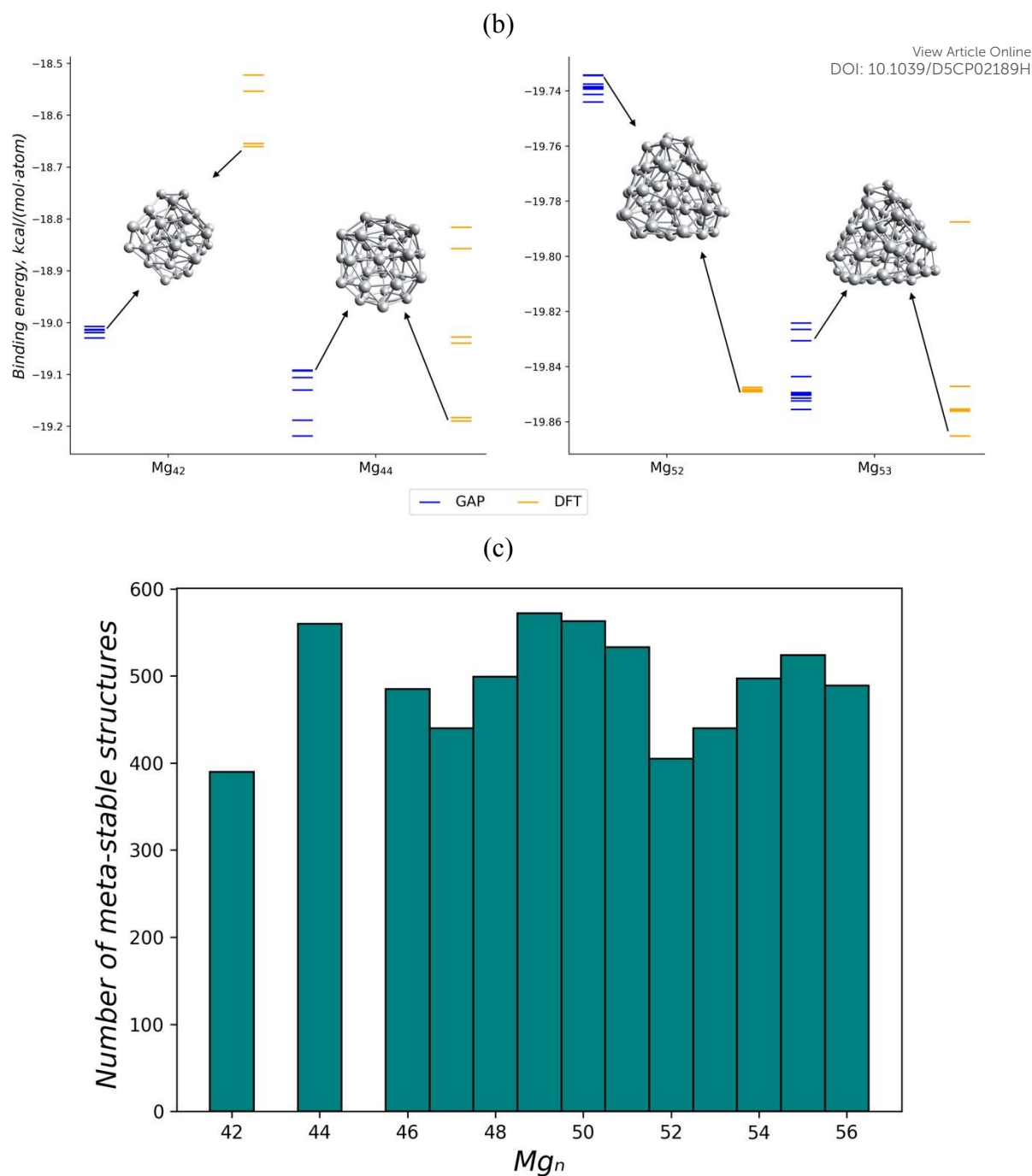
Figure 4. (a) Distribution of DFT binding energies per-atom for the 10 lowest-energy isomers of Mg_{42}, Mg_{44}, Mg_{46}-Mg_{56} after GO. (b) Comparison of GAP/DFT binding energies per-atom for the 10 lowest-energy isomers of Mg_{42}, Mg_{44}, Mg_{52} and Mg_{53}. The DFT-evaluated structures of the lowest-energy isomer are visualized. (c) The number of metastable structures obtained during the MH optimization.

When analyzing the structure of sub-nanometer clusters, an important characteristic is the nuclearity and shape of the so-called "core" of the cluster, i.e., the inner set of atoms hidden under a shell of surface atoms[58–60]. Small clusters do not have a core at all — every atom is accessible from the surface. As the cluster size increases to n=15-30, one, two, or three atoms become

Physical Chemistry Chemical Physics Accepted Manuscript

14

inaccessible from the surface[9,34,45,55]. These early cores are typically linear or planar. With further growth, cores with higher nuclearity appear, capable of adopting various geometries. The nuclearity at which a core forms, as well as its geometry, are characteristic features that can be used to analyze the similarity or differences in structure prediction methods. Moreover, correctly reproducing the cluster shell structure is critical for accurate surface energy evaluation[28].

For Mg_{42} the energy difference between the known GM and the structure predicted by the GAP model is 0.225 *kcal/(mol·atom)*. This value noticeably higher than expected model's accuracy $RMSE(E_{bonds})$ = 0.149 *kcal/(mol·atom)* on the test dataset.

During the GO of Mg_{44} isomers based on the GAP model, the previously reported GM structure[28,29] was found among the 10 isomers. However, considering the significant scatter of energy among the 10 lowest-energy isomers for Mg_{42} and Mg_{44} (Figure 4(b)), the prediction accuracy for clusters with $n < 50$, not included in the training dataset, is relatively lower compared to Mg_{46}-Mg_{55} (Figure 4(a)). This highlights the necessity of including all clusters with $n < 50$ in the training dataset for better structural prediction within this nuclearity range.
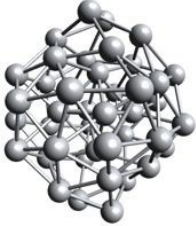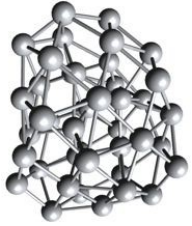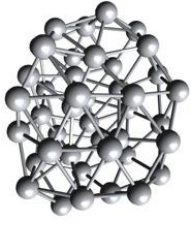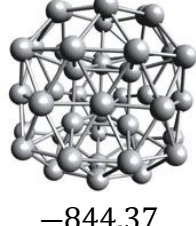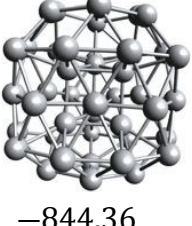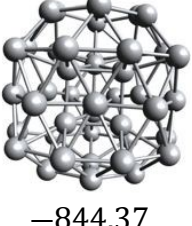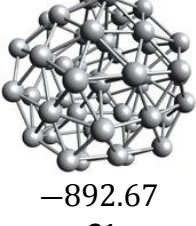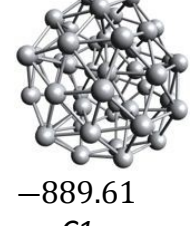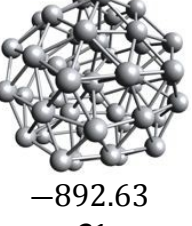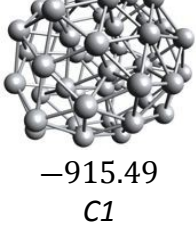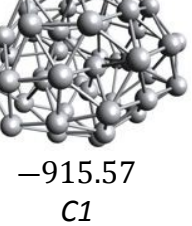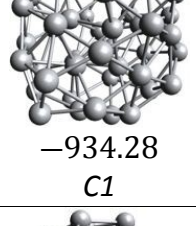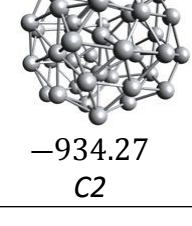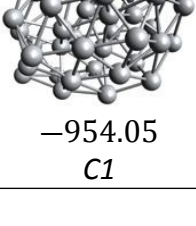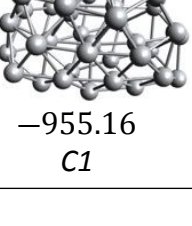
Despite the low scatter of energy and the proximity of the energies of the 10 lowest isomers to the GM energies found during the GO of Mg_{46}-Mg_{50} clusters (Figure 4(a)), the GAP model reproduced the previously reported DFT-based GM structure[28] only for Mg_{46} (Table 1).

For Mg_{48}, the GAP model predicted a previously unknown structure very close in energy to the reported GM[28] — the difference in binding energy is 0.01 *kcal/mol* (Table 1). For Mg_{47}, the model found a core structure similar to the reported GM[29], with a binding energy difference of 0.08 *kcal/mol* (Table 1). In case Mg_{49}, Mg_{50}, significant deviations were observed both in the core and shell structures compared to the reported GMs[28,29]. However, the energy differences between the 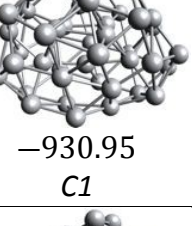known GMs and the structures predicted by the GAP model remain within the model's accuracy on the training dataset $RMSE(E_{bonds})$ = 0.121 *kcal/(mol·atom)* and equal 0.023 and 0.110 kcal/(mol·atom) for Mg_{49} and Mg_{50}, respectively (Table 1).

During the GO of Mg_{51}-Mg_{56} clusters using the parameterized GAP model, new GMs were

discovered for Mg_{51}-Mg_{53}, demonstrating a clear trend toward forming pyramidal core structures

(Table 1). Additionally, the model accurately reproduced previously known GMs[28,29] for Mg_{55} and

Mg_{56} (Table 1). Deviations from the previously known GM[28] were observed only for Mg_{54}, whose

predicted structure differed in shell configuration (Table 1). The binding energy difference

between the known GM and the GAP-predicted structure for Mg_{54} is 0.148 *kcal/(mol·atom)*, which

also falls within the $RMSE(E_{bonds}) = 0.149$ *kcal/(mol·atom)* for the test dataset.

Table 1. Low-energy structures and core structures of Mg_{42}, Mg_{44} and Mg_{46}-Mg_{56} clusters with E_{bonds}^{DFT} ($kcal/(mol)$), predicted by the GAP model and reported in previous studies. Structures relaxed using DFT: BP86/6-31G(d).

	GAP (this work)		Mueller's group[29]		Goedecker-Kanhere's group[28]	
Mg_{42}						
	−783.7		−790.78		−793.16	
	C1		*C1*		*C1*	
Mg_{44}						
	−844.37		−844.36		−844.37	
	C1		*C1*		*C1*	
Mg_{46}						
	−892.67		−889.61		−892.63	
	C1		*C1*		*C1*	
Mg_{47}						-
	−915.49		−915.57			
	C1		*C1*			
Mg_{48}						
	−934.28		−930.95		−934.27	
	C1		*C1*		*C2*	
Mg_{49}						-
	−954.05		−955.16			
	C1		*C1*			

Mg$_{50}$

−975.52
C2

−975.02
C1

−981.02
C1

Mg$_{51}$

−1004.24
C1

−996.76
C1

-

Mg$_{52}$

−1032.16
C1

−1021.85
C1

−1027.96
C3

Mg$_{53}$

−1052.86
C1

−1047.95
C1

-

Mg$_{54}$

−1070.62
C1

−1072.13
C1

−1078.63
C1

Mg$_{55}$

−1108.47
C1

−1094.72
C1

−1108.46
C1

Mg$_{56}$

−1139.04
C1

-

−1138.95
C1

18

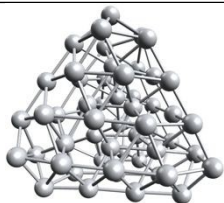## GO of Mg$_{57}$-Mg$_{60}$, Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ clusters
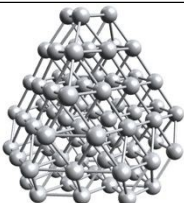
The parameterized GAP model demonstrated good predictive capability for the structures of Mg$_{51}$-Mg$_{56}$ clusters, which were not included in the training set. Thus, the GAP model is effective extrapolation tool for exploring the structures of clusters with a larger number of atoms than in the training dataset.

Accordingly, GO was performed for the Mg$_{59}$, Mg$_{69}$, Mg$_{74}$, Mg$_{99}$ clusters, which are known as "magic" structures, appearing as the most intense peaks in mass spectra[61] as well as for their neighboring clusters Mg$_n$ $n \pm 1$.

The pyramidal core structure observed at $n = 51$ is predicted by the GAP model to be preserved at least up to $n = 60$. The GMs of the Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$ predicted using the GAP model, show an evolution of the pyramidal core structure of Mg$_{51}$-Mg$_{60}$ by adding a 7-atom base in the form of a hexagon with a central atom. Thus, the cores of the Mg$_{68}$-Mg$_{70}$ and Mg$_{73}$-Mg$_{75}$ acquire multiple equivalent planes in the form of hexagons with a central atom (Table 2).

For the GMs of the Mg$_{98}$-Mg$_{100}$ clusters, no further evolution of the pyramidal core structure was observed during the GO process. The absence of a common core structure for Mg$_{98}$-Mg$_{100}$, similar to Mg$_{68}$-Mg$_{70}$ and Mg$_{73}$-Mg$_{75}$, may indicate either the accumulation of a sufficient number of atoms leading to a transition from a small-cluster state to a state closer to bulk, or an insufficient number of starting structures/iterations during the GO process to find the "true" GM similar to those found for Mg$_{68}$-Mg$_{70}$ and Mg$_{73}$-Mg$_{75}$ (Table 2).

Table 2. Low-energy structures of Mg$_{57}$-Mg$_{60}$, Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ with E_{bonds}^{GAP} (*kcal/(mol)*), predicted by the GAP model.

	Cluster structure	Cluster core structure		Cluster structure	Cluster core structure
Mg$_{57}$			Mg$_{73}$		
	−1157.67			−1567.83	
	C1			*C1*	

Mg$_{58}$

−1182.01

C1

Mg$_{74}$

−1594.94

C1

Mg$_{59}$

−1205.03

C1

Mg$_{75}$

−1621.42

C1

Mg$_{60}$

−1232.36

C1

Mg$_{98}$

−2254.86

C1

Mg$_{68}$

−1435.51

C1

Mg$_{99}$

−2284.09

C1

Mg$_{69}$

−1462.12

Cs

Mg$_{100}$

−2318.18

Cs

Mg$_{70}$

−1492.30

Cs

Change in the distances from the COM of Mg$_{39}$-Mg$_{60}$, Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ clusters

Analysis of the distances from the center of mass (COM) of the Mg$_{39}$-Mg$_{60}$, Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ clusters (Figure 5) didn't reveal any clear highly ordered structures similar to the Na clusters[62]. The stepwise nature of the accumulation of atoms in most cases is associated with the formation of the core structure. For Mg$_{98}$-Mg$_{100}$ clusters, the transition from core atoms to shell atoms has a weakly expressed character. But all clusters in this range have a stepwise transition at a distance from 3.5 to 4.5 Å. For clusters $n \leq 51$, this transition is associated with the end of the accumulation of cluster core atoms, and for $n > 51$, a stepwise accumulation of atoms begins to appear already inside the core structure. It was also not revealed that the stepwise nature of the accumulation of atoms for the cluster structure correlates with magic numbers in the range $n=39-100$ for magnesium clusters.
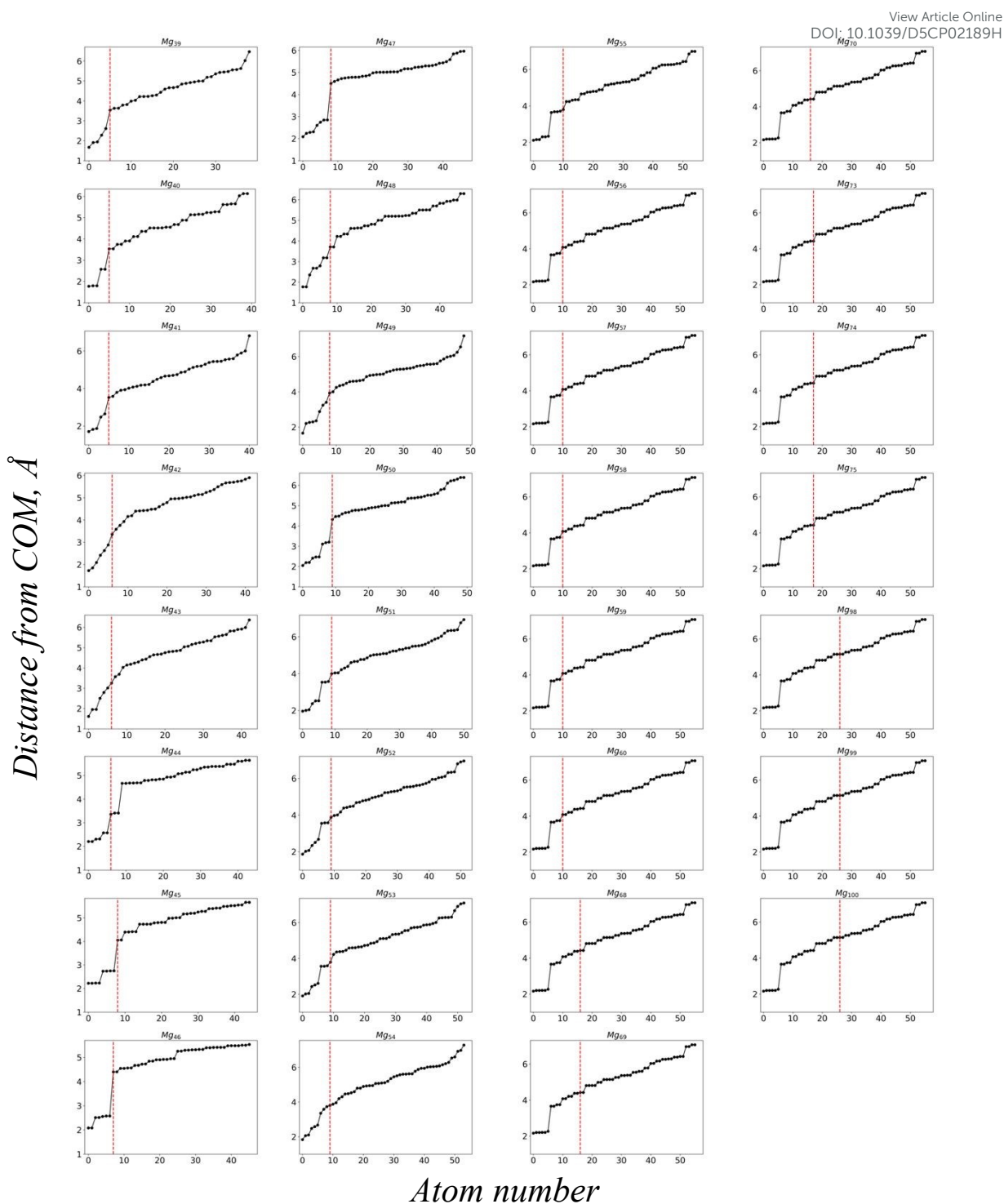
Figure 5. The distances from the COM for each atom ordered in the increasing fashion for the GM Mg_{39}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}, Mg_{98}-Mg_{100} of clusters. The red line indicates the number of atoms in the cluster core.

Change in the Binding Energy of Mg_{39}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}, Mg_{98}-Mg_{100} clusters

Figure 6(a) shows the change in the average binding energy per atom for the GMs of the $Mg_{60,}$ Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}, Mg_{98}-Mg_{100} clusters identified in this and previous studies[28,29]. The GMs of the "magic" structures Mg_n $n = 40, 47, 56$ clearly demonstrate a trend of lowered binding energy per atom compared to $n \pm 1$ neighboring structures, which correlates well with the experimental synthesis of magnesium clusters in a helium environment[61]. Similarly, the GMs of the Mg_n $n = 44, 52$ clusters show an energetic advantage over $n \pm 1$ neighboring structures in Figure 6(a), although they are not observed experimentally, possibly due to specific experimental conditions[58].

The absence of minima in the binding energy per atom for the Mg_n $n = 59, 69, 74, 99$ clusters compared to their $n \pm 1$ neighbors characterized by the GAP model. For the Mg_{51}-Mg_{56} region of high GAP model accuracy, the change in binding energy with increasing n is almost linear, resulting in the absence of distinct minima for Mg_{52} and Mg_{56}. Thus, the parameterized GAP model, without further DFT calculations, cannot serve as a reliable tool for assessing the stability of "magic" structures relative to $n \pm 1$ neighboring clusters Figure 6(b).

(a)
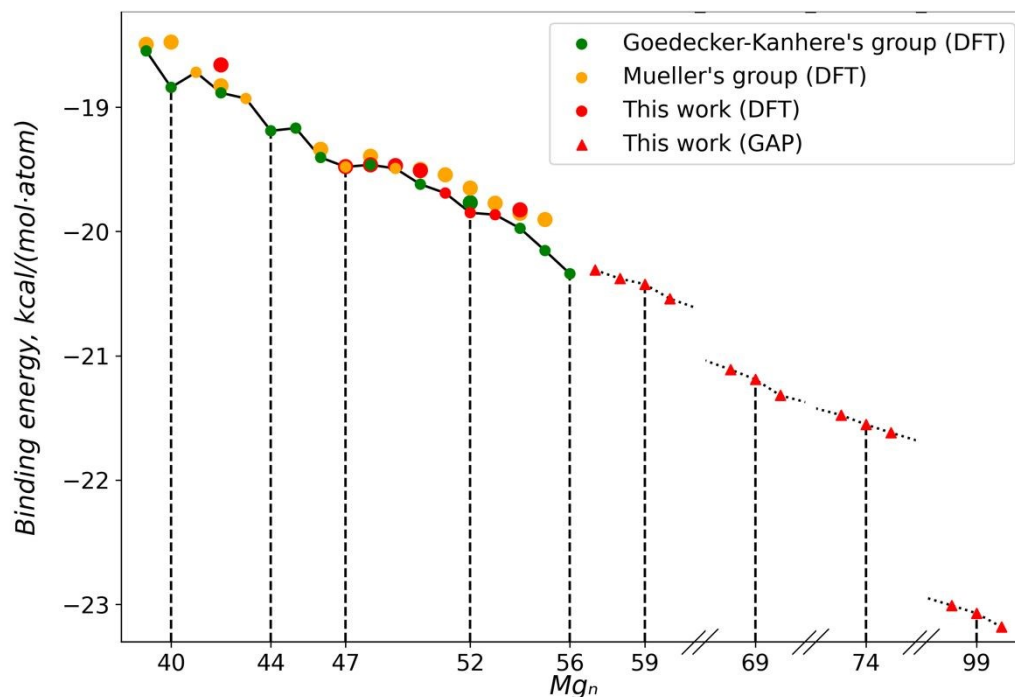
(b)
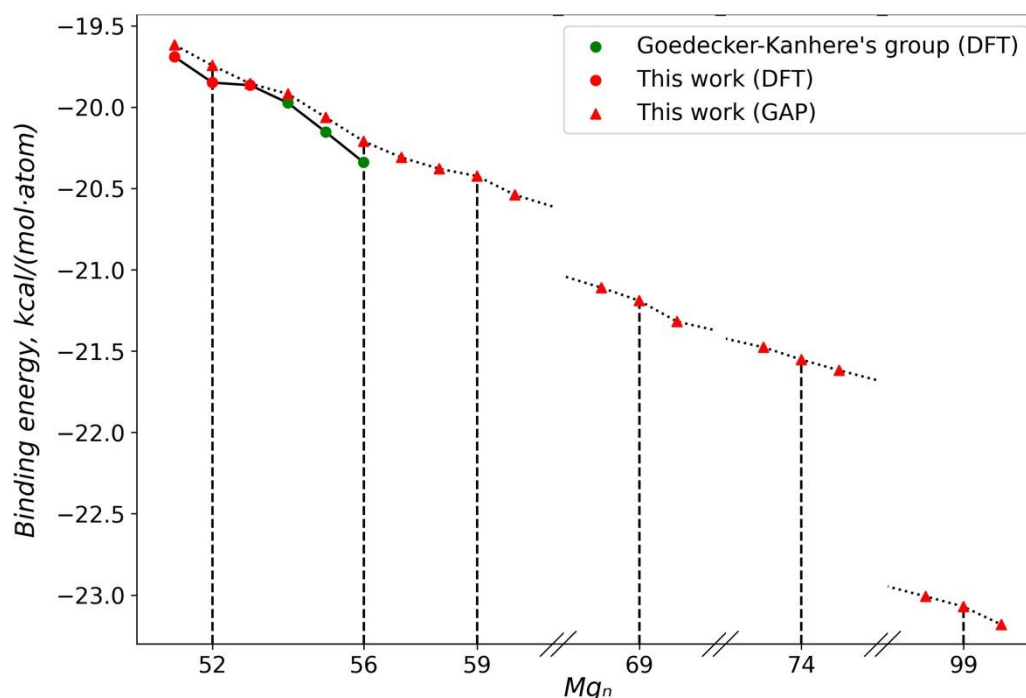
Figure 6. (a) Change in the average binding energy per atom for low-lying isomers and GMs of the Mg_{39}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}, Mg_{98}-Mg_{100} obtained in this work or from previous studies[28,29]; (b) Change in the average binding energy per atom according to DFT and GAP predictions for the GMs of the Mg_{51}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}, Mg_{98}-Mg_{100}. The GMs are connected by solid lines (DFT) and dashed lines (GAP); structures exhibiting a lower binding energy per atom compared to neighboring clusters or corresponding to "magic" structures[61] are marked by vertical dashed lines. If the cluster structures Mg_n obtained in different studies were identical, they are represented on the graph according to the earliest study.

24

Change in the contribution of hcp, bcc, and fcc lattice motifs in Mg$_{39}$-Mg$_{60}$, Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ clusters.

When analyzing the atomic environments of the Mg$_{39}$-Mg$_{56}$ clusters, no fcc and bcc crystal lattice motifs were detected. However, a significant fraction of atomic environments characteristic of the hcp lattice is observed. The detected hcp-like motifs in the Mg$_{39}$-Mg$_{100}$ clusters can be either atoms on the cluster surface or atoms in the cluster core.

The fraction of hcp-like atoms for Mg$_{39}$-Mg$_{50}$ fluctuates chaotically between 0.0 and 0.5 depending on the nuclearity and cutoff radius. For Mg$_{50}$-Mg$_{60}$, Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ clusters a more stable fraction of hcp-like atoms is observed (Figure. 7).

The highest fraction of hcp-like atoms within the clusters across all nuclearities is observed at the cutoff radius of 3.209 Å. This indicates that clusters Mg$_n$, n = 39–50 exhibit hcp-like local environments at short interatomic distances. For $n > 50$, more correlated estimates of the hcp-like atom fraction between different cutoff radii are observed. This may suggest the formation of hcp-like environments at longer interatomic distances (Figure. 7).

The GM structures of Mg$_{68}$-Mg$_{70}$, Mg$_{73}$-Mg$_{75}$, Mg$_{98}$-Mg$_{100}$ predicted using the GAP model show no strong correlated growth in the fraction of hcp-like atoms with increasing cluster size (Figure. 7).

No structural differences in the content of hcp-like atoms were found between the GM structures of "magic" clusters and clusters with other nuclearities. Thus, the presence of crystal lattice motifs in the cluster structure is not a critical criterion for the stability of $n < 100$ clusters.
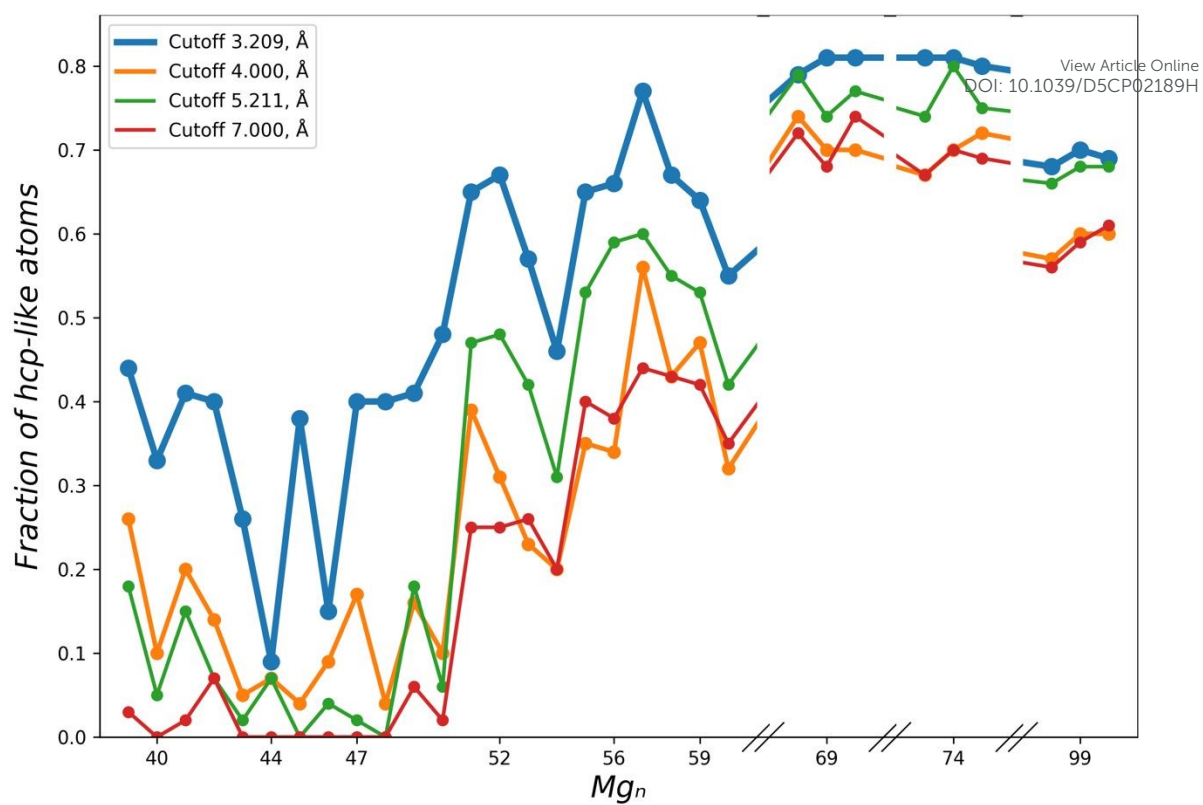
Figure 7. Change in the fraction of hcp-like atoms in the cluster structures

26

CONCLUSION

The proposed two-stage GridSearch approach combined with active learning for optimizing GAP model parameters enabled the construction of an interatomic potential for Mg clusters that allows structural prediction in the extrapolative domain $n > 50$. GO performed using the parameterized GAP model led to the discovery of previously unknown, energetically more profitable GMs for Mg_{51}-Mg_{53} clusters, which also exhibit more ordered structures compared to those reported earlier. These newly identified Mg_{51}-Mg_{53} GMs show a tendency toward the formation of pyramidal cores starting from Mg_{51}, in contrast to previous reports suggesting the onset at Mg_{54}[28,29].

The parameterized GAP model results in less energetically profitable structures for clusters in the interpolative domain of the model - Mg_{42}, Mg_{44}, as well as in its training domain - Mg_{47}, Mg_{49} Mg_{50}. This effect is likely due to a strong dependence of the atomic energy contributions and local environments on the overall cluster size for $n < 51$.

As a result of GO using the GAP model, potentially possible GMs structures were identified for Mg_{57}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}, Mg_{98}-Mg_{100}, clusters, including the "magic" Mg_{59}, Mg_{69}, Mg_{74} and Mg_{99} structures. A difference in core structure formation was observed between Mg_{39}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75} and Mg_{98}-Mg_{100} clusters. For the Mg_{98}-Mg_{100} clusters don't exhibit a tendency to form a clear structure of the core similar to the Mg_{39}-Mg_{60}, Mg_{68}-Mg_{70}, Mg_{73}-Mg_{75}. Thus, as $n \rightarrow 100$, a transition from a small cluster condition to a state resembling bulk material may occur. However, the fraction of hcp-like atoms in Mg_{98}-Mg_{100} clusters remains lower to that of Mg_{73}-Mg_{75}

No structural differences were found in the hcp-like atomic content between GM "magic" clusters and other clusters of similar sizes. Thus, the presence of crystalline lattice motifs within a cluster structure does not appear to be a critical criterion for structural stability in $n < 100$ clusters.

ASSOCIATED CONTENT

The code for training the model and performing the analysis was written in Python v3. Additionally, the libraries NumPy[63], SciPy[64], Scikit-learn[65], Pandas[66] and Matplotlib[67] were used for data processing. The trained model and source code are available on GitLab [https://gitlab.com/vihuhol188/gap_mg_clusters].

AUTHOR INFORMATION

Corresponding Author

Ilya S. Steshin - Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod 603950, Russia; orcid.org/0000-0002-3330-716X; Email: ilya.steshin@icloud.com

Authors

Sergey V. Panteleev − Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod 603950, Russia; orcid.org/0000-0001-8780-5642

Igor V. Petukhov − Volga State University of Technology, Yoshkar-Ola 424000, Russia; orcid.org/0009-0000-2365-4857

Stanislav K. Ignatov − Lobachevsky State University of Nizhny Novgorod, Nizhny Novgorod 603950, Russia; orcid.org/0000-0002-6039-3905

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.
Conceptualization: S.K.I. and I.S.S.; methodology: I.S.S. and S.K.I.; software: I.S.S. and S.V.P.; quantum chemical calculations: I.S.S., S.V.P. and S.K.I.; validation: I.S.S. and S.K.I.; formal analysis: all; investigations: all; resources: I.V. P., I.S.S., and S.K.I.; data curation: I.S.S. and S.K.I.; writing–original draft preparation: I.S.S and S.K.I.; writing–review and editing: I.S.S., S.K.I., I.V. P. and S.V.P.; visualization: I.S.S.; supervision: S.K.I. and I.V. P.; project administration: S.K.I. and I.V. P.;

ACKNOWLEDGMENT

REFERENCES

(1) *Physics and Chemistry of Finite Systems: From Clusters to Crystals*; Jena, P., Khanna, S. N., Rao, B. K., Eds.; Springer Netherlands: Dordrecht, 1992. https://doi.org/10.1007/978-94-017-2645-0.

(2) Jin, R. Atomically Precise Metal Nanoclusters: Stable Sizes and Optical Properties. *Nanoscale* **2015**, *7* (5), 1549–1565. https://doi.org/10.1039/C4NR05794E.

(3) Jena, P.; Sun, Q. Super Atomic Clusters: Design Rules and Potential for Building Blocks of Materials. *Chem. Rev.* **2018**, *118* (11), 5755–5870. https://doi.org/10.1021/acs.chemrev.7b00524.

(4) Jin, R. Atomically Precise Metal Nanoclusters: Stable Sizes and Optical Properties. *Nanoscale* **2015**, *7* (5), 1549–1565. https://doi.org/10.1039/C4NR05794E.

(5) Batista, K. E. A.; Ocampo-Restrepo, V. K.; Soares, M. D.; Quiles, M. G.; Piotrowski, M. J.; Da Silva, J. L. F. *Ab Initio* Investigation of CO_2 Adsorption on 13-Atom 4d Clusters. *J. Chem. Inf. Model.* **2020**, *60* (2), 537–545. https://doi.org/10.1021/acs.jcim.9b00792.

(6) Felício-Sousa, P.; Andriani, K. F.; Da Silva, J. L. F. *Ab Initio* Investigation of the Role of the d-States Occupation on the Adsorption Properties of H_2, CO, CH_4 and CH_3OH on the Fe_{13}, Co_{13}, Ni_{13} and Cu_{13} Clusters. *Phys. Chem. Chem. Phys.* **2021**, *23* (14), 8739–8751. https://doi.org/10.1039/D0CP06091G.

(7) Jia, X.; Li, J.; Wang, E. Cu Nanoclusters with Aggregation Induced Emission Enhancement. *Small* **2013**, *9* (22), 3873–3879. https://doi.org/10.1002/smll.201300896.

(8) Castleman, A. W.; Khanna, S. N. Clusters, Superatoms, and Building Blocks of New Materials. *J. Phys. Chem. C* **2009**, *113* (7), 2664–2675. https://doi.org/10.1021/jp806850h.

(9) Li, L.; Larsen, A. H.; Romero, N. A.; Morozov, V. A.; Glinsvad, C.; Abild-Pedersen, F.; Greeley, J.; Jacobsen, K. W.; Nørskov, J. K. Investigation of Catalytic Finite-Size-Effects of Platinum

Metal Clusters. *J. Phys. Chem. Lett.* **2013**, *4* (1), 222–226.
https://doi.org/10.1021/jz3018286.

(10)   Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; Ong, S. P. Performance and Cost Assessment of Machine Learning Interatomic Potentials. *J. Phys. Chem. A* **2020**, *124* (4), 731–745. https://doi.org/10.1021/acs.jpca.9b08723.

(11)   Gastegger, M.; Kauffmann, C.; Behler, J.; Marquetand, P. Comparing the Accuracy of High-Dimensional Neural Network Potentials and the Systematic Molecular Fragmentation Method: A Benchmark Study for All-Trans Alkanes. *The Journal of Chemical Physics* **2016**, *144* (19), 194110. https://doi.org/10.1063/1.4950815.

(12)   Popov, I. S.; Valeeva, A. A.; Enyashin, A. N. Identifying Stable Nb-O Clusters Using Evolutionary Algorithm and DFT: A Foundation for Machine Learning Potentials. *Chemical Physics* **2025**, *590*, 112533. https://doi.org/10.1016/j.chemphys.2024.112533.

(13)   Jana, R.; Caro, M. A. Searching for Iron Nanoparticles with a General-Purpose Gaussian Approximation Potential. *Phys. Rev. B* **2023**, *107* (24), 245421. https://doi.org/10.1103/PhysRevB.107.245421.

(14)   Behler, J. Constructing High-dimensional Neural Network Potentials: A Tutorial Review. *Int J of Quantum Chemistry* **2015**, *115* (16), 1032–1050. https://doi.org/10.1002/qua.24890.

(15)   Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87* (18), 184115. https://doi.org/10.1103/PhysRevB.87.184115.

(16)   Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions through Many-Body Expansions. *The Journal of Chemical Physics* **2018**, *148* (24), 241725. https://doi.org/10.1063/1.5024577.

(17)   Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), 136403. https://doi.org/10.1103/PhysRevLett.104.136403.

(18)   Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *The Journal of Chemical Physics* **2011**, *134* (7), 074106. https://doi.org/10.1063/1.3553717.

(19)   Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral Neighbor Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials. *Journal of Computational Physics* **2015**, *285*, 316–330. https://doi.org/10.1016/j.jcp.2014.12.018.

(20)   Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* **2016**, *14* (3), 1153–1173. https://doi.org/10.1137/15M1054183.

(21)   Drautz, R. Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials. *Phys. Rev. B* **2019**, *99* (1), 014104. https://doi.org/10.1103/PhysRevB.99.014104.

(22)   Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **2021**, *121* (16), 10037–10072. https://doi.org/10.1021/acs.chemrev.0c00868.

(23)   Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121* (16), 10073–10141. https://doi.org/10.1021/acs.chemrev.1c00022.

(24)   Shang, Y.; Pistidda, C.; Gizer, G.; Klassen, T.; Dornheim, M. Mg-Based Materials for Hydrogen Storage. *Journal of Magnesium and Alloys* **2021**, *9* (6), 1837–1860. https://doi.org/10.1016/j.jma.2021.06.007.

(25)   Wagemans, R. W. P.; Van Lenthe, J. H.; De Jongh, P. E.; Van Dillen, A. J.; De Jong, K. P. Hydrogen Storage in Magnesium Clusters: Quantum Chemical Study. *J. Am. Chem. Soc.* **2005**, *127* (47), 16675–16680. https://doi.org/10.1021/ja054569h.

(26)   Lyalin, A.; Solov'yov, I. A.; Solov'yov, A. V.; Greiner, W. Evolution of the Electronic and Ionic Structure of Mg Clusters with Increase in Cluster Size. *Phys. Rev. A* **2003**, *67* (6), 063203. https://doi.org/10.1103/PhysRevA.67.063203.

(27)   Panteleev, S. V.; Ignatov, S. K.; Belyaev, S. N.; Razuvaev, A. G.; Masunov, A. E. Structure and Properties of 1237 Low-Lying Isomers of Magnesium Clusters Mgn (n = 2–32) Predicted with the DFT Global Optimization. *J Clust Sci* **2023**, *34* (2), 1105–1120. https://doi.org/10.1007/s10876-022-02291-w.

(28)   Heidari, I.; De, S.; Ghazi, S. M.; Goedecker, S.; Kanhere, D. G. Growth and Structural Properties of Mg$_N$ (N = 10–56) Clusters: Density Functional Theory Study. *J. Phys. Chem. A* **2011**, *115* (44), 12307–12314. https://doi.org/10.1021/jp204442e.

(29)   Manna, S.; Wang, Y.; Hernandez, A.; Lile, P.; Liu, S.; Mueller, T. A Database of Low-Energy Atomically Precise Nanoclusters. *Sci Data* **2023**, *10* (1), 308. https://doi.org/10.1038/s41597-023-02200-4.

(30)   Duanmu, K.; Friedrich, J.; Truhlar, D. G. Thermodynamics of Metal Nanoparticles: Energies and Enthalpies of Formation of Magnesium Clusters and Nanoparticles as Large as 1.3 Nm. *J. Phys. Chem. C* **2016**, *120* (45), 26110–26118. https://doi.org/10.1021/acs.jpcc.6b08371.

(31)   Duanmu, K.; Roberto-Neto, O.; Machado, F. B. C.; Hansen, J. A.; Shen, J.; Piecuch, P.; Truhlar, D. G. Geometries, Binding Energies, Ionization Potentials, and Electron Affinities of Metal Clusters: Mg$_n^{0,\pm1}$, n = 1–7. *J. Phys. Chem. C* **2016**, *120* (24), 13275–13286. https://doi.org/10.1021/acs.jpcc.6b03080.

(32)   Klawohn, S.; Darby, J. P.; Kermode, J. R.; Csányi, G.; Caro, M. A.; Bartók, A. P. Gaussian Approximation Potentials: Theory, Software Implementation and Application Examples. *The Journal of Chemical Physics* **2023**, *159* (17), 174108. https://doi.org/10.1063/5.0160898.

(33)   Jinnouchi, R.; Karsai, F.; Verdi, C.; Asahi, R.; Kresse, G. Descriptors Representing Two- and Three-Body Atomic Distributions and Their Effects on the Accuracy of Machine-Learned Inter-Atomic Potentials. *The Journal of Chemical Physics* **2020**, *152* (23), 234102. https://doi.org/10.1063/5.0009491.

(34)   Ignatov, S. K. Unexpected Polarization Properties of Sub-Nanosized Magnesium Clusters.

(35)   Ignatov, S. K.; Belyaev, S. N.; Panteleev, S. V.; Masunov, A. E. How Many Isomers Do Metallic Clusters Have? Case of Magnesium Clusters of up to 55 Atoms. *J. Phys. Chem. A* **2021**, *125* (30), 6543–6555. https://doi.org/10.1021/acs.jpca.1c02529.

(36)   M. J. Frisch; G. W. Trucks; H. B. Schlegel; G. E. Scuseria; M. A. Robb; J. R. Cheeseman; G. Scalmani; V. Barone; B. Mennucci; G. A. Petersson; H. Nakatsuji; M. Caricato; X. Li; H. P. Hratchian; A. F. Izmaylov; J. Bloino; G. Zheng; J. L. Sonnenberg; M. Hada; M. Ehara; K. Toyota; R. Fukuda; J. Hasegawa; M. Ishida; T. Nakajima; Y. Honda; O. Kitao; H. Nakai; T. Vreven; J. A. Montgomery; J. E. Peralta; F. Ogliaro; M. Bearpark; J. J. Heyd; E. Brothers; K. N. Kudin; V. N. Staroverov; R. Kobayashi; J. Normand; K. Raghavachari; A. Rendell; J. C. Burant; S. S. Iyengar; J. Tomasi; M. Cossi; N. Rega; J. M. Millam; M. Klene; J. E. Knox; J. B. Cross; V. Bakken; C. Adamo; J. Jaramillo; R. Gomperts; R. E. Stratmann; O. Yazyev; A. J. Austin; R. Cammi; C. Pomelli; J. W. Ochterski; R. L. Martin; K. Morokuma; V. G. Zakrzewski; G. A. Voth; P. Salvador; J. J. Dannenberg; S. Dapprich; A. D. Daniels; Ö. Farkas; J. B. Foresman; J. V. Ortiz; J. Cioslowski; D. J. Fox. Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009).

(37)   Packwood, D.; Nguyen, L. T. H.; Cesana, P.; Zhang, G.; Staykov, A.; Fukumoto, Y.; Nguyen, D. H. Machine Learning in Materials Chemistry: An Invitation. *Machine Learning with Applications* **2022**, *8*, 100265. https://doi.org/10.1016/j.mlwa.2022.100265.

(38)   Lewis-Atwell, T.; Beechey, D.; Şimşek, Ö.; Grayson, M. N. Reformulating Reactivity Design for Data-Efficient Machine Learning. *ACS Catal.* **2023**, *13* (20), 13506–13515. https://doi.org/10.1021/acscatal.3c02513.

(39)   Liashchynskyi, P.; Liashchynskyi, P. Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS. arXiv December 12, 2019. http://arxiv.org/abs/1912.06059 (accessed 2024-11-04).

(40)   Jiang, X.; Xu, C. Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data. *JCM* **2022**, *11* (19), 5772. https://doi.org/10.3390/jcm11195772.

(41)   Belete, D. M.; Huchaiah, M. D. Grid Search in Hyperparameter Optimization of Machine Learning Models for Prediction of HIV/AIDS Test Results. *International Journal of Computers and Applications* **2022**, *44* (9), 875–886. https://doi.org/10.1080/1206212X.2021.1974663.

(42)   Shkil, D. O.; Muhamedzhanova, A. A.; Petrov, P. I.; Skorb, E. V.; Aliev, T. A.; Steshin, I. S.; Tumanov, A. V.; Kislinskiy, A. S.; Fedorov, M. V. Expanding Predictive Capacities in Toxicology: Insights from Hackathon-Enhanced Data and Model Aggregation. *Molecules* **2024**, *29* (8), 1826. https://doi.org/10.3390/molecules29081826.

(43)   Goedecker, S. Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *The Journal of Chemical Physics* **2004**, *120* (21), 9911–9917. https://doi.org/10.1063/1.1724816.

(44)   Krummenacher, M.; Gubler, M.; Finkler, J. A.; Huber, H.; Sommer-Jörgensen, M.; Goedecker, S. Performing Highly Efficient Minima Hopping Structure Predictions Using the Atomic Simulation Environment (ASE). *SoftwareX* **2024**, *25*, 101632. https://doi.org/10.1016/j.softx.2024.101632.

(45)   Tong, Q.; Xue, L.; Lv, J.; Wang, Y.; Ma, Y. Accelerating CALYPSO Structure Prediction by Data-Driven Learning of a Potential Energy Surface. *Faraday Discuss.* **2018**, *211*, 31–43. https://doi.org/10.1039/C8FD00055G.

(46)   Wang, Y.; Liu, S.; Lile, P.; Norwood, S.; Hernandez, A.; Manna, S.; Mueller, T. Accelerated Prediction of Atomically Precise Cluster Structures Using On-the-Fly Machine Learning. *npj Comput Mater* **2022**, *8* (1), 173. https://doi.org/10.1038/s41524-022-00856-x.

(47)   Christiansen, M.-P. V.; Rønne, N.; Hammer, B. Atomistic Global Optimization X: A Python Package for Optimization of Atomistic Structures. *The Journal of Chemical Physics* **2022**, *157* (5), 054701. https://doi.org/10.1063/5.0094165.

(48)   Jindal, S.; Bulusu, S. S. Structural Evolution in Gold Nanoparticles Using Artificial Neural Network Based Interatomic Potentials. *The Journal of Chemical Physics* **2020**, *152* (15), 154302. https://doi.org/10.1063/1.5142903.

(49)   Chiriki, S.; Jindal, S.; Bulusu, S. S. Neural Network Potentials for Dynamics and Thermodynamics of Gold Nanoparticles. *The Journal of Chemical Physics* **2017**, *146* (8), 084314. https://doi.org/10.1063/1.4977050.

(50)   Thorn, A.; Rojas-Nunez, J.; Hajinazar, S.; Baltazar, S. E.; Kolmogorov, A. N. Toward *Ab Initio* Ground States of Gold Clusters via Neural Network Modeling. *J. Phys. Chem. C* **2019**, *123* (50), 30088–30098. https://doi.org/10.1021/acs.jpcc.9b08517.

(51)   Sadeghi, A.; Ghasemi, S. A.; Schaefer, B.; Mohr, S.; Lill, M. A.; Goedecker, S. Metrics for Measuring Distances in Configuration Spaces. *The Journal of Chemical Physics* **2013**, *139* (18), 184118. https://doi.org/10.1063/1.4828704.

(52)   Walker, G. B.; Marezio, M. Lattice Parameters and Zone Overlap in Solid Solutions of Lead in Magnesium. *Acta Metallurgica* **1959**, *7* (12), 769–773. https://doi.org/10.1016/0001-6160(59)90090-2.

(53)   Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dułak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; Hermes, E. D.; Jennings, P. C.;

Bjerre Jensen, P.; Kermode, J.; Kitchin, J. R.; Leonhard Kolsbjerg, E.; Kubal, J.; Kaasbjerg, K.; Lysgaard, S.; Bergmann Maronsson, J.; Maxson, T.; Olsen, T.; Pastewka, L.; Peterson, A.; Rostgaard, C.; Schiøtz, J.; Schütt, O.; Strange, M.; Thygesen, K. S.; Vegge, T.; Vilhelmsen, L.; Walter, M.; Zeng, Z.; Jacobsen, K. W. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys.: Condens. Matter* **2017**, *29* (27), 273002. https://doi.org/10.1088/1361-648X/aa680e.

(54) Laakso, J.; Himanen, L.; Homm, H.; Morooka, E. V.; Jäger, M. O. J.; Todorović, M.; Rinke, P. Updates to the DScribe Library: New Descriptors and Derivatives. *The Journal of Chemical Physics* **2023**, *158* (23), 234802. https://doi.org/10.1063/5.0151031.

(55) Chiriki, S.; Bulusu, S. S. Modeling of DFT Quality Neural Network Potential for Sodium Clusters: Application to Melting of Sodium Clusters (Na20 to Na40). *Chemical Physics Letters* **2016**, *652*, 130–135. https://doi.org/10.1016/j.cplett.2016.04.013.

(56) Artrith, N.; Behler, J. High-Dimensional Neural Network Potentials for Metal Surfaces: A Prototype Study for Copper. *Phys. Rev. B* **2012**, *85* (4), 045439. https://doi.org/10.1103/PhysRevB.85.045439.

(57) Wang, N.; Huang, S. Molecular Dynamics Study on Magnesium Hydride Nanoclusters with Machine-Learning Interatomic Potential. *Phys. Rev. B* **2020**, *102* (9), 094111. https://doi.org/10.1103/PhysRevB.102.094111.

(58) Martin, T. P.; Bergmann, T.; Goehlich, H.; Lange, T. Shell Structure of Clusters. *J. Phys. Chem.* **1991**, *95* (17), 6421–6429. https://doi.org/10.1021/j100170a009.

(59) Martin, T. P. From Atoms to Solids. *Solid State Ionics* **2000**, *131* (1), 3–12. https://doi.org/10.1016/S0167-2738(00)00617-2.

(60) Martin, T. P. Shells of Atoms. *Physics Reports* **1996**, *273* (4), 199–241. https://doi.org/10.1016/0370-1573(95)00083-6.

(61) Diederich, Th.; Döppner, T.; Fennel, Th.; Tiggesbäumker, J.; Meiwes-Broer, K.-H. Shell Structure of Magnesium and Other Divalent Metal Clusters. *Phys. Rev. A* **2005**, *72* (2), 023203. https://doi.org/10.1103/PhysRevA.72.023203.

(62) Ghazi, S. M.; Zorriasatein, S.; Kanhere, D. G. Building Clusters Atom-by-Atom: From Local Order to Global Order. *J. Phys. Chem. A* **2009**, *113* (12), 2659–2662. https://doi.org/10.1021/jp809729p.

(63) Harris, C. R.; Millman, K. J.; Van Der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; Van Kerkwijk, M. H.; Brett, M.; Haldane, A.; Del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

(64) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; Van Der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; Van Mulbregt, P.; SciPy 1.0 Contributors; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Slavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; De Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmerer, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov,

N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* **2020**, *17* (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.

(65)   Pedregosa, F.; Varoquaux, G.; Org, N.; Gramfort, A.; Michel, V.; Fr, L.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Tp, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12* (85), 2825–2830.

(66)   McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the Python in Science Conference*; SciPy: Austin, Texas, 2010; pp 56–61. https://doi.org/10.25080/majora-92bf1922-00a.

(67)   Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

The trained model and source code are available on GitLab
https://gitlab.com/vihuhol188/gap_mg_clusters