

## Supplementary Information

Pablo Álvarez-Zapatero, Andrés Vega, and Andrés Aguado\*

*Departamento de Física Teórica, Atómica y Óptica, Universidad de Valladolid, Valladolid  
47071, Spain*

E-mail: [aguado@metodos.fam.cie.uva.es](mailto:aguado@metodos.fam.cie.uva.es)

---

\*To whom correspondence should be addressed

# A simple yet efficient structural similarity descriptor

As explained in the main text, we need to remove duplicate structures that may appear in the structural data pool generated by the different potentials. We achieve this goal by using a structural similarity descriptor based on the “atomic equivalence indices”, first introduced in [J. Chem. Phys. **107**, 6321 (1997)] by Bonacic-Koutecky and coworkers. For brevity of exposition, let us consider first a homonuclear cluster for which we need to characterize the atomic skeleton only, i.e. we dispense with the chemical order problem for the time being. For a cluster with  $N$  atoms, the atomic equivalence index of atom  $i$  is just the sum of its distances to all other atoms in the cluster:

$$\sigma_i = \sum_{j=1}^N d(i, j), \quad (1)$$

where  $d(i, j)$  is the distance between atoms  $i$  and  $j$ . The set of ordered  $\sigma_i$  values forms an  $N$ -dimensional vector  $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$  that characterizes each local minimum (isomer) on the potential energy surface (in our case we have ordered the  $\sigma_i$  values in ascending order, i.e. from the least to the most coordinated atom). With the vector components ordered this way, and each component being exclusively based on interatomic distances, two identical isomers will have identical  $\boldsymbol{\sigma}$  indicators irrespective of their global orientation in space or of possible permutations of atoms. The indicator will additionally remove enantiomers in case of chiral point group symmetries.

This indicator would be enough for a homonuclear cluster if one works with a single empirical potential, i.e. with a unique potential energy surface. However, we have employed a variety of potentials in order to enhance structural diversity, and collected the different outputs of each potential into a single structural data pool. A potential problem is that different potentials may have slightly different values for the average interatomic distance, so the same structural motif (for example, an icosahedron) may appear several times in the data pool but with different interatomic distances. Now,  $\boldsymbol{\sigma}$  as defined above depends on the

quantitative values of distances, and will not be able to remove duplicates with a different “volume”. Therefore, we redefine the indicator by normalizing the  $\sigma$  vector:

$$\sigma_i \Rightarrow \frac{\sigma_i}{\sqrt{\sum_i \sigma_i^2}}$$

This new indicator will recognize two structures that merely differ by a scale transformation as identical.

In a nanoalloy, it is important to distinguish different chemical order patterns as well. To this end, we have employed additional  $\sigma_i^{\text{AA}}$ ,  $\sigma_i^{\text{BB}}$  and  $\sigma_i^{\text{AB}}$  indicators, defined in full analogy with equation (1), but where only AA, BB or AB distances, respectively, are considered. All the vector indicators are ordered and normalized as explained above. The degree of similarity between two isomers  $p$  and  $q$  is then quantified by the metric distance in the vector spaces defined by the different indicators:

$$\Sigma_{pq} = \frac{1}{4} \left( |\sigma_p - \sigma_q| + |\sigma_p^{\text{AA}} - \sigma_q^{\text{AA}}| + |\sigma_p^{\text{BB}} - \sigma_q^{\text{BB}}| + |\sigma_p^{\text{AB}} - \sigma_q^{\text{AB}}| \right)$$

The first term will determine the degree of similarity between two given skeletal structures, while the remaining three terms will check the homotop similarity. Given a structural data bank containing an arbitrary number of structures, the user can choose the cutoff value of  $\Sigma_{pq}$  used to decide if one of those two structures has to be removed from the database. We have empirically determined that values of  $\Sigma$  around 0.01-0.03 are appropriate for removing duplicates. But we notice that the same indicator can be used to select from a database containing, say, 1000 structures, those 50 structures which differ most from each other, i.e. it is useful to select a number of individuals with maximum diversity. One simply has to increase the value of the cutoff until only 50 structures remain in the pruned list.

# Optimal potential parameters

Table S1: Optimal parameters for the bare Gupta potential

$A_{AA}$	$\xi_{AA}$	$p_{AA}$	$q_{AA}$	$A_{BB}$	$\xi_{BB}$	$p_{BB}$	$q_{BB}$	$A_{AB}$	$\xi_{AB}$	$p_{AB}$	$q_{AB}$
0.078	0.40	4.51	2.21	0.090	0.56	7.75	3.47	0.084	0.50	6.26	2.90

Table S2: Optimal parameters for the Coulomb-corrected-Gupta potential. In this table we quote the parameters for the metallic part of the potential

$A_{AA}$	$\xi_{AA}$	$p_{AA}$	$q_{AA}$	$A_{BB}$	$\xi_{BB}$	$p_{BB}$	$q_{BB}$	$A_{AB}$	$\xi_{AB}$	$p_{AB}$	$q_{AB}$
0.07	0.38	4.69	2.30	0.085	0.55	7.29	3.12	0.085	0.47	5.87	2.65

Table S3: Optimal parameters for the Coulomb-corrected-Gupta potential. In this table we quote the parameters for the ionic part of the potential

$\chi_A$	$\chi_B$	$\eta_A$	$\eta_B$	$C_A$	$C_B$	$D_A$	$D_B$
3.88	3.24	2.24	1.79	0.98	1.30	1.37	0.77

Table S4: Optimal parameters for the Coulomb-corrected-Gupta potential. In this table we quote the cutoff parameters used to determine the coordination numbers (eq. 2 in the main paper), in Å units.

$d_1(AB)$	$d_2(AB)$	$d_1(AA)$	$d_2(AA)$	$d_1(BB)$	$d_2(BB)$
2.70	3.30	2.55	3.15	2.95	3.55

## Size transferability test for the EP

The training and testing sets employed in the main paper include exclusively clusters with between 10 and 50 atoms. Thus a pertinent question to ask is about the transferability of the potential to system sizes bigger than those employed in the fitting procedure. To check this issue, we have chosen a convenient test system, namely a 79-atom truncated octahedron, and performed a global search for homotops of the  $\text{Zn}_{40}\text{Mg}_{39}$  nanoalloys, which is essentially equiatomic. We have sorted all the homotops located in these extensive searches according to the number of Zn-Mg heterobonds, i.e. from maximally mixed to maximally segregated, and then have chosen ten different homotops with varying degrees of mixing, on which we have performed DFT SIESTA calculations. The figure below benchmarks the performances of bare and Coulomb-corrected Gupta models against the *ab initio* data. This figure displays the differences in binding energy per atom, with respect to the most stable homotop. The degree of mixing decreases from homotop 1 to homotop 10. SIESTA results show, except for some oscillations of secondary importance, that stability increases with the degree of mixing, homotop 1 being the most stable one. The bare Gupta potential does not correctly capture the energy difference between mixed and segregated homotops, and moreover predicts homotop 4, with an intermediate degree of mixing, to be most stable. Meanwhile, the Coulomb-corrected Gupta EP recognizes homotop 1 as the most stable one, and further provides a much more accurate match for all the energy differences between homotops, even for a cluster size which is significantly bigger than those used when training the potential.

One of the main results of our work is that the apparently small (6%) charge transfer contribution to the nanoalloy binding energies is utterly important to reproduce the correct chemical order trends. The results of this test, performed on 79-atom clusters, suggests that such conclusion holds quite generally for all clusters within the small non-scalable size regime. Thus, we expect our potential to be useful for clusters with up to a few hundred atoms at least. There should not be high expectations, however, about the performance of

our potential in reproducing structural properties or cohesive energies at the bulk limit, as no bulk data were feeded into the fitting process. In summary, our new potential model displays a very good transferability with respect to nanoalloy composition, and does so over a sizable size regime, including all sizes within the most interesting non scalable regime. Whether our analytical model can be made fully transferable (to all compositions *and* to all sizes from small molecules to the bulk), while retaining its high accuracy regarding homotopic stability, still remains an open question.

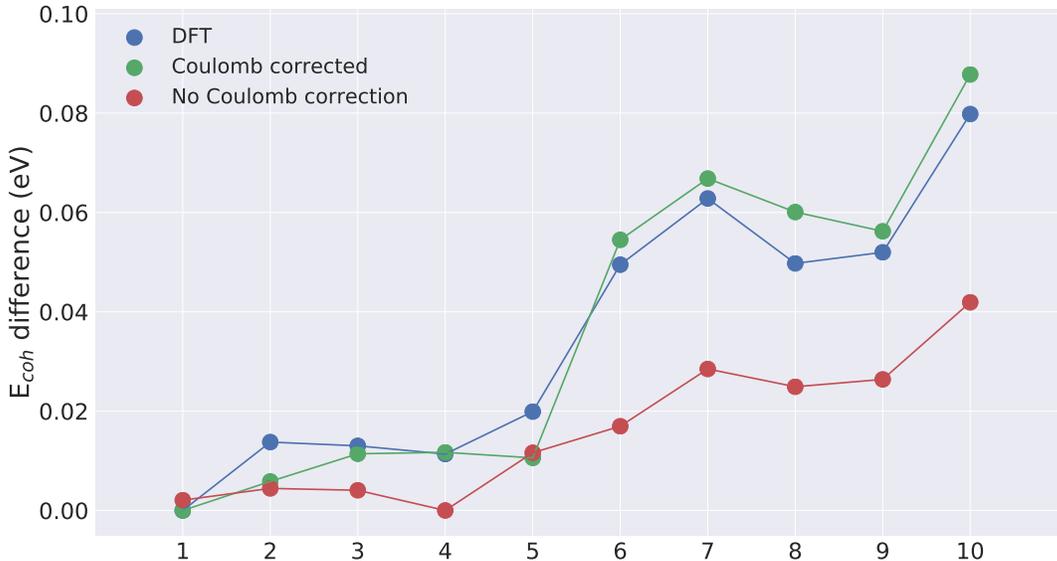


Figure S1: Relative stabilities of ten different homotops for a 79-atom nearly-equiatomic  $\text{Zn}_{40}\text{Mg}_{39}$  nanoalloy, with a truncated octahedral atomic skeleton. The figure shows excess binding energies per atom (or cohesive energies) with respect to the most stable homotop, for each level of theory. The bare Gupta potential fails in identifying the most stable DFT homotop. The relative stabilities in homotopic space are much better reproduced by the Coulomb-corrected potential.

## Comparison with bulk alloy structures

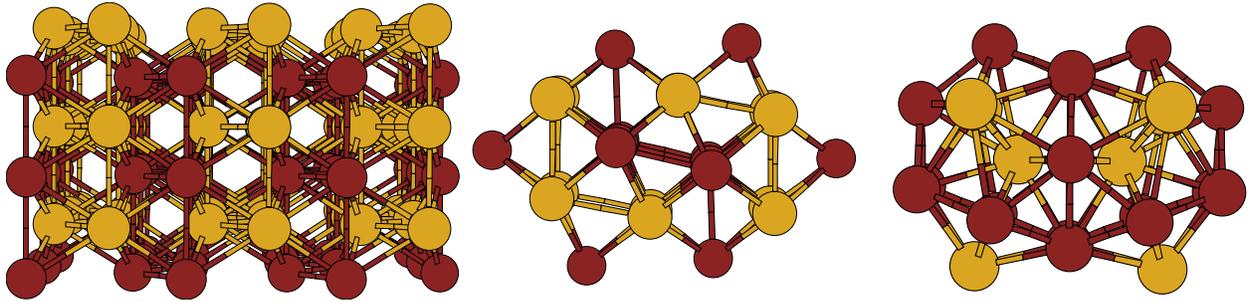


Figure S2: On the left we show a fragment of the Pmma crystalline lattice of the equiatomic Zn-Mg alloy; the middle graph shows a 24-atom fragment directly cut from the Pmma lattice. Relaxation of this initial structure produces the GM structure of  $\text{Zn}_{12}\text{Mg}_{12}$ ; the right figure shows a relaxed 26-atom fragment of the  $R\bar{3}c$  crystalline lattice adopted by  $\text{Mg}_{21}\text{Zn}_{25}$  alloy. Its structure is identical to the GM structure of  $\text{Zn}_{13}\text{Mg}_{13}$ , although the composition is obviously different.

## Electronic properties

Table S5: Vertical ionization potential, vertical electron affinity and fundamental gap of equiatomic Zn-Mg nanoalloys with up to 50 atoms

$N$	vIP	vEA	GAP
4	6.70	0.64	6.06
6	6.20	0.97	5.23
8	6.04	1.19	4.86
10	5.69	1.09	4.60
12	5.79	1.59	4.20
14	5.62	1.93	3.69
16	5.51	1.45	4.06
18	5.43	2.09	3.34
20	5.28	1.62	3.66
22	5.28	1.92	3.37
24	5.03	1.95	3.08
26	4.95	2.06	2.89
28	4.97	2.05	2.92
30	5.00	2.20	2.80
32	5.12	2.07	3.04
34	5.08	2.22	2.86
36	4.83	2.02	2.81
38	5.06	2.27	2.79
40	5.06	2.33	2.73
42	5.07	2.41	2.66
44	5.11	2.20	2.91
46	5.16	2.11	3.05
48	4.90	2.25	2.64
50	4.82	2.32	2.50

# Cohesive Energy

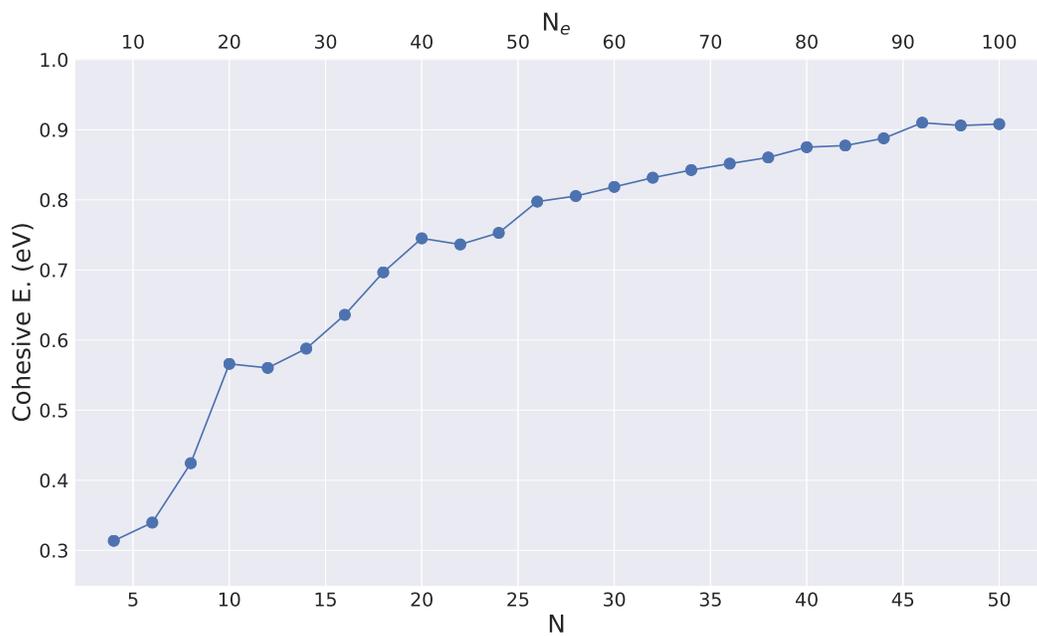


Figure S3: Cohesive energy of equiatomic Zn-Mg nanoalloys as a function of the total number of atoms (lower scale) or of electrons (upper scale).