Faraday Discussions

Cite this: DOI: 10.1039/d5fd00107b

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Optimising descriptors to correlate stability of C- or N-doped high-entropy alloys: a combined DFT and machine-learning regression study

Chih-Heng Lee, [ab] Jyh-Wei Lee [cdef] and Hsin-Yi Tiffany Chen [*agh]

Received 19th July 2025, Accepted 15th August 2025
DOI: 10.1039/d5fd00107b

Interstitial doping is a common approach to improve the mechanical or functional properties of high-entropy alloys (HEAs); their stability is usually predicted by a specific single descriptor. Herein, we consider six types of microstructure-based descriptor, seven types of electronic-structure-based local-environment descriptor and their combinations to predict the stability of the C- or N-doped $VNbMoTaWTiAl_{0.5}$ (BCC) HEA, mainly using density functional theory (DFT) calculations. A machine-learning interatomic potential and Monte Carlo simulations were employed to verify the short-range order in the HEA. The microstructure-based descriptors include the composition of the first-, second-, and third-nearest neighbour shells (1NN, 2NN and 3NN), OctaDist distortion parameters (ζ, Δ, Σ, Θ), the Voronoi volume ($V_{Voronoi}$) of the dopant, and the volume change of the unit cell after doping (ΔV_{cell}); the electronic-structure-based local-environment descriptors include the local potential (LP), the electrostatic potential (EP), the charge density (CHG), the electron localization function (ELF) at the vacant doping site, the d-band center (ε_d), the mean electronegativity (EN) of the 1NN shell around the dopant, and the Bader charge of the C or N dopants. For a single descriptor, the best correlation between the descriptor and the doping energy (indication of HEA stability) is found for 1NN with coefficient of determination (Q^2) values of ~51 or ~61% obtained using the LOOCV (leave-one-out cross-validation) approach for C or N doping, respectively. After adding volume descriptor(s) into the linear regression model

[a]Department of Engineering and System Science, National Tsing Hua University, Hsinchu, 300044, Taiwan. E-mail: hsinyi.tiffany.chen@gapp.nthu.edu.tw
[b]Department of Chemistry, University of Liverpool, Liverpool L69 7ZD, UK
[c]Department of Materials Engineering, Ming Chi University of Technology, New Taipei City, 24301, Taiwan
[d]Center for Plasma and Thin Film Technologies, Ming Chi University of Technology, New Taipei City, 24301, Taiwan
[e]College of Engineering, Chang Gung University, Taoyuan, 33301, Taiwan
[f]High Entropy Materials Center, National Tsing Hua University, Hsinchu, 300044, Taiwan
[g]Department of Materials Science and Engineering, National Tsing Hua University, Hsinchu, 300044, Taiwan
[h]College of Semiconductor Research, National Tsing Hua University, Hsinchu, 300044, Taiwan

with the 1NN descriptor, Q^2 increases to 72 and 76% for C and N doping, respectively. After further adding the electronic-structure-based EP descriptor, Q^2 further increases to 75 and 80% for C and N doping, respectively, despite the poor correlation using a single volume descriptor. This study quantitatively combined and compared the independent contributions of different types of local-environment descriptors to the stability of the C- or N-doped HEA, demonstrating the importance of considering both key microstructure-based and electronic-structure-based local-environment descriptors using the regression models to achieve more accurate correlation of dopant stability in HEA; these combined approaches could be further applied to other materials systems, research fields and applications.

1. Introduction

Since the first high-entropy alloy (HEA) research was reported in 2004,[1] many different HEA compositions and applications have been developed, such as high-entropy superalloys,[2–4] high-entropy refractory alloys,[5–7] high-entropy coating,[8,9] and even high-entropy catalysts.[10–12] Generally, a HEA mixes five or more principal metal elements, and each element has a concentration ranging from 5 to 35 at%. Several principal elements with a comparable ratio significantly enhance the configurational entropy of the disordered solid-solution phase. The increasing entropy decreases the Gibbs free energy of the disordered solid-solution phase, which is further stabilized. Due to the presence of unique disordered phases from the mixing of a variety of principal metal elements, different types of HEAs exhibit outstanding performance in different application fields.

To further improve the performance of HEAs, doping with light elements such as carbon (C),[13,14] nitrogen (N),[13,15] oxygen (O),[15,16] or boron[14,17] at interstitial sites is a common approach to improve the mechanical or functional properties of a HEA. There have been many studies proposing different descriptors to describe the effect of the dopant local environment on HEA stability. For example, Moravcik *et al.* proposed that lowering the electronegativity of the first-nearest-neighbor (1NN) shell around the dopants increases the transfer of electrons to the dopant C and N atoms, leading to a more-stabilized CoCrNi alloy.[18] Casillas-Trujillo *et al.* studied how the Voronoi volume and valence electron concentration of the 1NN shell around the interstitial site stabilize the C-doped HfNbTiVZr.[19] Yang *et al.* found that the number of Ti and Zr atoms in the 1NN shell around the interstitial O atom exhibits high correlation with the DFT-calculated energy of TiZrNb, TiZrVNb, and TiZrV alloys.[20] However, most of the previous studies only correlated a single descriptor to the doping energy of the HEA systems. A comprehensive comparison between the contributions of different types of descriptors (microstructure-based and electronic-structure-based) of the complex local environment to the stability with interstitial dopants has not been clarified.

In this study, to secure optimal descriptors and to comprehensively understand how different local-environment descriptors (several microstructure-based and electronic-structure-based, detailed below) near dopants influence the stability, the $VNbMoTaWTiAl_{0.5}$ (HEA) (body-centered cubic, BCC), a commonly used refractory HEA,[21–24] was selected as a case study to correlate C or N doping energy (ΔE_C or ΔE_N). The HEA supercell with appropriate short-range order (SRO) was constructed and verified by Warren–Cowley parameters (WCP). The

microstructure-based descriptors of investigation include the composition of first-, second-, and third-nearest-neighbour shells (1NN, 2NN and 3NN), OctaDist distortion parameters[25] and two volume descriptors, Voronoi volume of the dopant, and the volume change of the unit cell after doping. The electronic-structure-based descriptors associate with the local potential (LP), electrostatic potential (EP), charge density (CHG), electron localization function (ELF) at the vacant doping site, d-band center (ε_d), mean electronegativity (EN) of the 1NN shell around dopant, and Bader charge of C or N dopant. Pearson correlation coefficients (PCC) were applied to determine the correlation between a single descriptor and ΔE_C or ΔE_N (stability indication). To further enhance the correlation between local-environment descriptors and ΔE_C or ΔE_N, the linear regression model was employed to secure the optimal combined microstructure-based and electronic-structure-based descriptors. This study quantitatively combined and compared contributions of different descriptors to the stability of the C- or N-doped HEA, demonstrating the importance of the descriptors' combination using machine-learning regression models to achieve more accurate correlation of ΔE_C or ΔE_N, and leading to an understanding of the independent contribution of each local-environment descriptor to the stability of the doped HEA.

2. Computational details

2.1. DFT calculations

The geometry optimisation, local-environment descriptors, and C or N doping energy (ΔE_C or ΔE_N) calculations were carried out based on density functional theory (DFT)[26,27] using the Vienna *Ab initio* Simulation Package (VASP 6.4).[28–32] The exchange–correlation functional was described by the Perdew–Burke–Ernzerhof (PBE) generalized-gradient-approximation (GGA).[33] The core-valence interaction was treated with projector-augmented-wave (PAW) potentials for all elements,[32,34] whose valence electronic configurations are V ($3s^2 3p^6 3d^4 4s^1$), Nb ($4s^2 4p^6 4d^4 5s^1$), Mo ($4s^2 4p^6 4d^5 5s^1$), Ta ($5p^6 5d^4 6s^1$), W ($5s^2 5p^6 5d^5 6s^1$), Ti ($3s^2 3p^6 3d^3 4s^1$), Al ($3s^2 3p^1$), C ($2s^2 2p^2$), and N ($2s^2 2p^3$). The plane-wave basis with cut-off energy of 520 eV was applied. The Γ-centred Monkhorst–Pack meshes of $3 \times 3 \times 3$ for the 39-atom HEA supercells and $2 \times 2 \times 2$ for the larger 156-atom HEA supercells were applied for geometry optimisation. The Methfessel–Paxton scheme with 0.2 eV width of smearing was used to determine the partial occupancies of each orbital. Self-consistent field (SCF) calculations were converged until the energy difference was less than 10^{-5} eV between the final two iterations. The geometry optimisation was converged until the force applied on each atom was less than 0.03 eV $Å^{-1}$.

The special quasi-random structure[35] (SQS) approach was employed using the mcsqs code[36] in Alloy Theoretic Automated Toolkit (ATAT)[37] to distribute the seven metal species, V, Nb, Mo, Ta, W, Ti, and Al in $VNbMoTaWTiAl_{0.5}$ (BCC) HEA models, matching pair correlations (~0) up to the third-nearest-neighbour shell. The C or N atom was doped into both tetrahedral and octahedral sites in the HEA crystal model, forming the HEA + C and HEA + N models. After doping and geometry optimisation, the composition of the 1NN shell of each dopant in the HEA + C or HEA + N model was identified using the ChemEnv package[38] implemented in Pymatgen (Python Materials Genomics).[39]

The six types of microstructure-based descriptors are described in detailed below. The composition of the 1NN shell around the dopant identified by

ChemEnv completely matches the composition of the 1st to 6th nearest atoms near each octahedral site in this study. The composition of the second- or third-nearest-neighbour (2NN or 3NN, respectively) shell was collected from the 7th to 14th or 15th to 22nd neighbour atoms around the dopant, respectively, as illustrated in Fig. 1(a). The geometry distortion of the octahedron surrounding the dopant at the octahedral site was quantified by the OctaDist parameters ζ, Δ, Σ and Θ, which refer to the distance, tilting, angle and torsional distortion,[25] illustrated in Fig. 1(b). The higher OctaDist parameters, the larger degree of distortion of the octahedron. The Voronoi volume (V_{Voronoi}) representing the local atomic volume of the C or N dopant is illustrated in Fig. 1(c). The volume change of the unit cell after C or N doping (ΔV_{cell}) was calculated. V_{Voronoi} and ΔV_{cell} are regarded as the volume descriptors.

The seven types of electronic-structure-based descriptors are described in detailed below. To quantify the electronic-structure properties of the local environment around each dopant site, we removed the C or N dopant from each geometry-optimised HEA + C or HEA + N model and calculated the local potential (LP), electrostatic potential (EP), charge density (CHG), and electron-localisation function (ELF) at the vacant dopant site. The LP includes the ionic, Hartree, and exchange–correlation potentials at the vacant site, while the electrostatic potential (EP) only includes the ionic and Hartree potentials. The d-band center (ε_{d})[40,41] of the 1NN shell for each dopant site was calculated. When calculating ε_{d}, Γ-centred Monkhorst–Pack meshes of $6 \times 6 \times 6$ and Gaussian smearing with 0.2 eV width were applied. Eqn (1) shows the ε_{d} calculation.

$$\varepsilon_{\mathrm{d}} = \frac{\int_{-\infty}^{\infty} E\rho_{\mathrm{d}}(E)\mathrm{d}E}{\int_{-\infty}^{\infty} \rho_{\mathrm{d}}(E)\mathrm{d}E} \tag{1}$$

E and $\rho_{\mathrm{d}}(E)$ are the electronic energy and projected density of state (PDOS) of the d-band of atoms in the 1NN shell of the vacant dopant site. The mean electronegativity (EN) on the Pauling scale[42] of atoms in the 1NN shell around the dopant atom was calculated. A Bader charge analysis[43–46] was performed to describe the

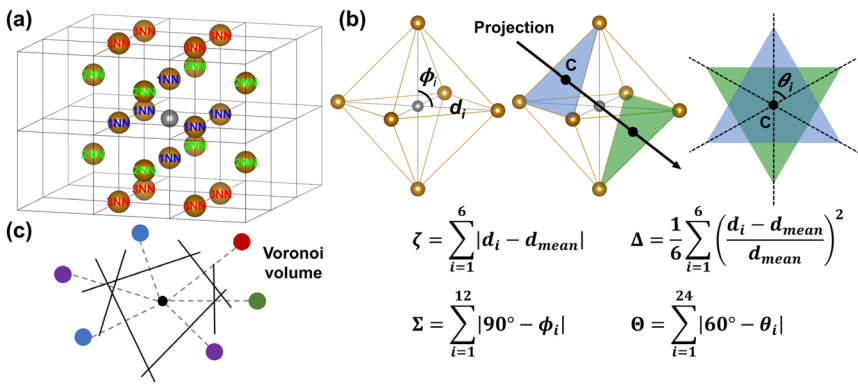

Fig. 1 (a) The illustration of the first-, second-, and third-nearest-neighbor (1NN, 2NN, and 3NN) shells surrounding an octahedral site in a body-centered-cubic (BCC) lattice. (b) The illustration and definition of OctaDist parameters[25] ζ, Δ, Σ, and Θ, which describe the distance, tilting, angle, and torsional distortion; and (c) the illustration of the Voronoi volume.

number of electrons gained by the C or N dopant atom. The C or N doping energy (ΔE_C or ΔE_N) is defined by eqn (2).

$$\Delta E_{C/N} = E(\text{HEA + C/N}) - E(\text{HEA}) - E(\text{C/N}) \quad (2)$$

E(HEA + C/N) and E(HEA) represent the DFT-calculated electronic energy of the C- or N-doped and pristine HEA models. E(C/N) is the electronic energy of the ground-state C per atom in graphite or N per atom of the nitrogen molecule, respectively. More negative ΔE_C or ΔE_N values refer to a more energetically stable C- or N-doped HEA.

2.2. Monte Carlo (MC) simulations

Monte Carlo (MC) simulations with the Metropolis algorithm[47] were performed to verify the short-range order (SRO) in the HEA crystal model. The initial structure in the MC simulations was generated by the SQS approach. For each MC step, two atoms of different elements were randomly sampled and swapped, and then the geometry optimisation was performed and the energy of the next-step HEA models recalculated. If the energy of the next-step HEA crystal model is lower than the previous MC step, this swap will be accepted; If the energy of the next-step HEA crystal model is higher than the previous MC step, the probability of acceptance will follow eqn (3).

$$p = \exp\left(-\frac{E_{i+1} - E_i}{k_B T}\right) \quad (3)$$

E_i and E_{i+1} represent the energy of the HEA model before and after the atomic swap, respectively. k_B and T are the Boltzmann constant and temperature. The temperature in the Metropolis algorithm was set to 500 K, because the SRO of each element pair in the V–Nb–Mo–Ta–W–Ti (BCC) system demonstrates a stable upward or downward trend when the temperature is above 500 K, based on the previous work.[48] In this study, E_i and E_{i+1} were calculated using DFT or a machine learning interatomic potential (MLIP).

2.3. Machine learning interatomic potential (MLIP)

To accelerate the MC simulations, we fine-tuned the MACE-MPA-0 MLIP using the MACE package.[49,50] The dataset used to fine-tune the MLIP comprised 1080 distinct 39-atom HEA structures generated by DFT-based MC simulations, including both accepted and rejected structures. These structures were divided 8 : 1 : 1 into training, validation and test datasets, *i.e.*, 864, 108, and 108 structures, respectively. By extracting each configuration during geometry optimisation of the HEA structures, the training, validation and test datasets contained in total 29 930, 3772, and 3625 configurations. Training was carried out with 100 epochs and batches of 10 configurations. To evaluate the performance of the fine-tuned MLIP, the coefficient of determination (R^2) and root-mean-square error (RMSE) were used to describe the precision of the MLIP, as shown in eqn (4) and (5).

$$R^2 = 1 - \frac{\sum_{j=1}^{N}\left(E_j - \hat{E}_j\right)^2}{\sum_{j=1}^{N}\left(E_j - \overline{E}\right)^2} \quad (4)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(E_j - \hat{E}_j\right)^2} \tag{5}$$

Here, E_j and $\hat{E}_j$ are the DFT-calculated and MLIP-predicted energy of the jth-configuration, respectively. N is the total number of configurations in the datasets. The R^2 value is a measure of the fraction of energy variance explained by the MLIP, ranging from 0 (no predictive capability) to 1 (perfect agreement). Thus, an R^2 value approaching 1 indicates that the MLIP reproduces nearly all DFT-calculated energy variations. The RMSE quantifies the average discrepancy between the MLIP-predicted and DFT-calculated electronic energies; lower RMSE values indicate more accurate energy predictions by the MLIP.

2.4. Characterization of short-range order (SRO)

To quantify the short-range order in the HEA crystal model, Warren–Cowley parameters (WCP)[20,51] in the 1NN shell around the dopants were employed, as shown in eqn (6).

$$\mathrm{WCP} = 1 - \frac{N_{ij}}{Nx_j} \tag{6}$$

N_{ij} denotes the average number of neighboring j-element atoms in the 1NN shell around the i-element atoms, and N is the total number of atoms in the 1NN around the i-element atoms. x_j is the molar ratio of element j in the unit cell. The zero WCP represents the situation where the distribution of elements i and j is completely random, *i.e.*, no short-range order between elements i and j is present. A positive WCP indicates that elements i and j repel each other, while a negative WCP indicates that elements i and j prefer to bond together.

2.5. Correlation between a single descriptor and doping energy

To analyse the correlation between a single descriptor and the doping energy, the Pearson correlation coefficient (PCC) was used, as shown in eqn (7).

$$\mathrm{PCC} = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y} \tag{7}$$

Cov(X,Y) is the covariance between variables X and Y, and σ_X and σ_Y are their standard deviations, respectively. The PCC quantifies the linear correlation between two variables, ranging from -1 to 1. In this study, X and Y represent both descriptors or doping energy at each dopant site. A PCC near 1 or -1 indicates a nearly perfect positive or negative linear correlation, whereas PCC close to 0 means low or no linear correlation.

2.6. Correlation between multiple descriptors and doping energy using linear regression models

Multiple linear regression models were employed to correlate ΔE_{C} or ΔE_{N} with different combinations of microstructure-based and electronic-structure-based descriptors, as shown in eqn (8).

$$\text{Linear-regression-model-predicted } \Delta E_{C/N} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \quad (8)$$

Here, $X_1, X_2, \ldots$ are the descriptors and β_i is the fitting coefficient in the regression model. To simultaneously account for all kinds of metals within the 1NN, 2NN, or 3NN shell around the dopant, the composition of the 1NN, 2NN and 3NN shells are considered as three descriptors and are denoted as 1NN, 2NN and 3NN descriptors. When the 1NN, 2NN or 3NN descriptor is considered in the linear regression model, $\beta_i X_i$ in eqn (8) is modified to $\sum_{M=1}^{N} \beta_{i,M} n_M$. n_M is the number of atoms of element M in the 1NN, 2NN or 3NN shell around the dopant, respectively. $\beta_{i,M}$ is the regression coefficient of element M, and N is the total number of metal elements in the 1NN, 2NN or 3NN shell around the dopant. Note that for the last considered element, Al, the number of Al atoms in the 1NN, 2NN, or 3NN shell around the dopant was not included in the linear regression model to prevent multicollinearity.

The predictive performance of the ΔE_C or ΔE_N regression models with a single descriptor or multiple descriptors combination was evaluated using the leave-one-out cross-validation (LOOCV) approach. The coefficient of determination (Q^2) based on LOOCV and cross-validation score (CV score) based on the LOOCV approach are defined in eqn (9)[52] and (10).[53]

$$Q^2 = 1 - \frac{\sum_{i=1}^{N} \left(\Delta E_i - \Delta \hat{E}_{i,N-1}\right)^2}{\sum_{i=1}^{N} \left(\Delta E_i - \Delta \overline{E}\right)^2} \quad (9)$$

$$\text{CV score} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\Delta E_i - \Delta \hat{E}_{i,N-1}\right)^2} \quad (10)$$

Here, $\Delta \hat{E}_{i,N-1}$ represents the linear-regression-model-predicted doping energy of the ith dopant site obtained by fitting descriptor(s) from all dopant sites except the ith site. ΔE_i is the DFT-calculated doping energy at the ith dopant site. The LOOCV approach evaluates the predictive performance of the regression models for the unobserved local environment in a HEA. The statistical meanings of Q^2 and the CV score are similar to those of R^2 and RMSE shown in eqn (4) and (5), respectively. Q^2 and the CV score signify the correlation between doping energy and a single or multiple local-environment descriptor(s).

3. Results and discussion

3.1. HEA modelling

To discuss how the local environment of each dopant influences the stability of the C- or N-doped refractory HEA, the 39-atom $VNbMoTaWTiAl_{0.5}$ (HEA) body-centered-cubic (BCC) model was first constructed, as depicted in Fig. S1(a). The preference of C or N for stable tetrahedral (Th) or octahedral (Oh) sites in the HEA model was investigated. C or N atoms were placed in all Oh and Th sites, 117 and 237 sites in total, respectively, for geometry optimisation, as shown in Fig. S1(b) and (c). After the geometry optimisation, C or N atoms at all the Th sites migrated to the Oh sites, and C or N atoms at all the Oh sites remained at the same sites, indicating that the

C and N dopants were more stable in the Oh sites than in the Th sites, which is consistent with previous studies.[18,20,54] Thus, in this study, we only considered Oh sites for all the C- or N-doped HEA models in the subsequent sections.

To establish a HEA supercell that is large enough to consider the appropriate short-range order (SRO), the Metropolis Monte-Carlo (MC) simulations were performed for HEA supercells containing various numbers of atoms, from 39 to 208. The MC simulations were accelerated by the machine-learning interatomic potential (MLIP) fine-tuned from the MACE-MPA-0 foundation model. The 39-atom HEA model was simulated by an MC simulation first, collecting each accepted or rejected MC step to build the training (864 structures), validation (108), and test (108) datasets to fine-tune the MLIP. The energy trajectory of this MC simulation for the 39-atom HEA model is shown in Fig. S2. Fig. S3–S5 show that both positive and negative Warren–Cowley parameters (WCP) of each element pair in the first-nearest-neighbour (1NN) shell were included in the datasets; this WCP distribution indicates that we considered both scenarios of different elements staying apart and close, respectively, showing the generalization and representativeness of the datasets applied to fine-tune the MLIP. Fig. 2(a) compares DFT-calculated and MLIP-predicted energies. The R^2 values between the DFT-calculated and MLIP-predicted energies in the training, validation and test datasets are 0.991, 0.989 and 0.991, respectively; the RMSE values between the DFT-calculated and MLIP-predicted energies in the training, validation and test datasets are 1.23, 1.42, and 1.14 meV per atom, respectively. These R^2 and RMSE

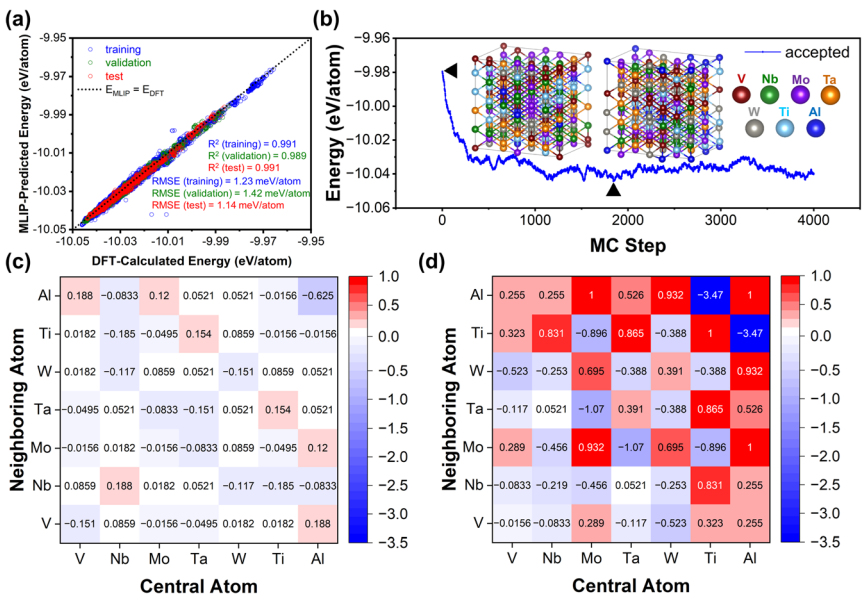

Fig. 2 (a) Parity plot comparing DFT-calculated energies of the 39-atom HEA supercell with those predicted by the fine-tuned MLIP for the training, validation, and test datasets. (b) Energy trajectory of the MC simulation in the 156-atom HEA model. The insets show the optimised structure (left) generated by the SQS approach (first step) and optimised structure (right) of the lowest-energy step in the MC simulations. (c) Warren–Cowley parameters (WCP) of the 1NN shell for all element pairs at the first step in the MC simulation. (d) WCP at the lowest-energy step in the MC simulation.

values indicate that the fine-tuned MLIP attains DFT-level accuracy, demonstrating that the MLIP-predicted energy is precise enough for the MC simulations.

To determine the number of atoms required to consider the appropriate SRO in the HEA model, larger HEA supercells were constructed and simulated using MC simulations with an MLIP, as shown in Fig. S6. WCP of all element pairs within the 1NN shell were collected from the lowest-energy HEA supercells sampled by MC simulations. Fig. S7 illustrates WCP of all element pairs *versus* HEA supercell size, demonstrating the essentially stable WCP of all element pairs in the 156-atom HEA supercell. This result indicates that the 156-atom HEA crystal model is large enough to consider the appropriate SRO. The energy trajectory and the 156-atom HEA models at the first and lowest-energy MC steps are shown in Fig. 2(b). Their corresponding 2-dimensional heatmaps of WCP are shown in Fig. 2(c) and (d). Fig. 2(c) shows that all WCP of the first MC step in the HEA model (generated by the SQS approach) are close to 0, indicating a random distribution without exhibiting the features of SRO. Fig. 2(d) shows that the WCP of the lowest-energy MC step in the HEA range from −3.5 to 1, implying the existence of SRO. For example, the WCP of Mo–Ta and V–W element pairs are −1.07 and −0.523, respectively, in the lowest-energy 156-atom HEA model, confirming a strong tendency for these element pairs to be present in the HEA, consistent with the previous study[55] (−1.09 for Mo–Ta and −0.36 for V–W) that reported on the VNbMoTaW HEA, validating the SRO feature in the 156-atom HEA model employed in this study.

To validate the SRO in the 156-atom HEA supercell, we compared WCP of each element pair in the 1NN shell and the formation enthalpies (ΔH_f) of the corresponding ordered binary-element compounds from the literature,[56] as shown in Fig. 3. For all element pairs that do not associate with Al, the WCP vary linearly with ΔH_f. The more negative the ΔH_f value of the binary compound, the lower the WCP between the corresponding element pairs. These results show that the SRO related to the V, Nb, Mo, Ta, W and Ti element pairs in the HEA model are largely dependent on binary formation enthalpy. Element pairs containing Al (Ti–Al, V–Al, Nb–Al, Mo–Al, Ta–Al, W–Al, and Al–Al) also show positive correlation between ΔH_f and the WCP. Al (FCC) and Ti (HCP) form the most favorable compound ($\Delta H_f = -428$ meV per atom; WCP = −3.47) and both exhibit close-packed crystal structures. Thus Ti is the most preferred neighbor for Al despite their atomic radius discrepancy, as listed in Table S1. Turning to the other element pairs (V–Al, Nb–Al, and Ta–Al), these have relatively more negative ΔH_f values (∼−300 meV per atom), compared to ΔH_f for Mo–Al, W–Al (∼−160 meV per atom), and Al–Al (0 meV per atom); the WCP of V–Al, Nb–Al (∼0.3), and Ta–Al pairs (∼0.5) are also lower than the WCP of Mo–Al, W–Al, and Al–Al (∼1).

In short, our MC simulations suggested that the 156-atom $VNbMoTaWTiAl_{0.5}$ (BCC) HEA exhibits appropriate SRO, further confirmed by the correlation between the formation enthalpy of the binary-element compounds and the WCP of the corresponding element pair. Therefore, the lowest-energy 156-atom HEA crystal model was employed in the subsequent sections.

3.2. Correlation between a single descriptor and doping energy

To clarify the influence of the local environment surrounding the C or N dopant on the stability of the C- or N-doped HEA, many different local-environment

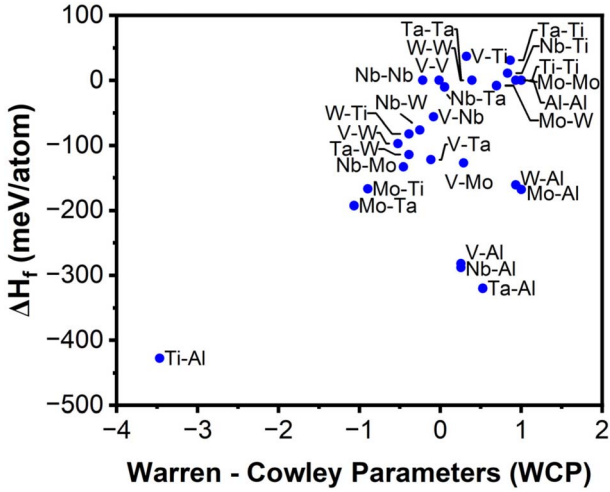

Fig. 3 The formation enthalpy of binary ground-state and ordered compounds from the literature[56] *versus* the Warren–Cowley parameters (WCP) from the 1NN shell of corresponding element pairs.

microstructure-based descriptors and electronic-structure-based descriptors were calculated using DFT. The microstructure-based descriptors include compositions in the 1NN, 2NN and 3NN shells around the dopant, *i.e.*, the number of metal atoms in the 1NN, 2NN and 3NN shells around the dopant, denoted as $M_{n\mathrm{NN}}$ (M = V, Nb, Mo, Ta, W, Ti and Al; $n = 1$, 2 and 3); the OctaDist parameters ζ, Δ, Σ and Θ; two volume descriptors, the Voronoi volume (V_{Voronoi}) of each dopant atom and the volume change of the HEA supercell (ΔV_{cell}) after doping. On the other hand, the electronic-structure-based descriptors include the local potential (LP), electrostatic potential (EP), charge density (CHG), electron localization function (ELF), d-band center (ε_{d}), mean electronegativity (EN) of the 1NN shell around the dopant, and the Bader charge of the C or N dopant. The C or N doping energies (ΔE_{C} or ΔE_{N}) were employed to describe the stability of the C- or N-doped HEA; the more negative the ΔE_{C} or ΔE_{N} values, the higher the stability of the C- or N-doped HEA.

Fig. 4, 5 and S8–S12 illustrate the Pearson correlation coefficient (PCC) between every descriptor and ΔE_{C} or ΔE_{N}. The larger the absolute value of the PCC, the better the descriptor at judging the stability of the C- or N-doped HEA (ΔE_{C} or ΔE_{N}). Among all descriptors, the PCCs between ΔE_{C} or ΔE_{N} and EP, CHG and ELF exhibit the most negative values, indicating that they are the best descriptors to predict the HEA stability after doping, compared to all microstructure-based descriptors. However, some descriptors might depend on each other, for example, the composition of the 1NN will be affected by the composition of the 2NN and 3NN descriptors; the number of different metal atoms around the dopant in the nNN ($n = 1$, 2 and 3) shells will also affect each other. The inadequacy of the PCC is that the dependency between different descriptors is omitted, which might lead to unfair judgement of the quality of descriptors. In addition, the PCC cannot consider the combined effects of multiple descriptors.

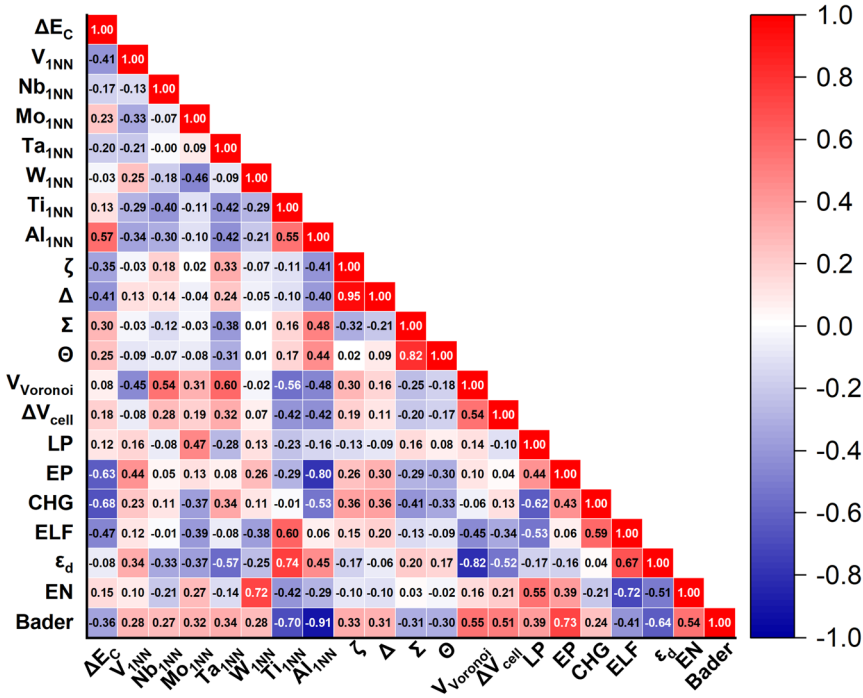

Fig. 4 The Pearson correlation coefficient (PCC) between all descriptors except the 2NN and 3NN composition around the dopant C and the carbon doping energy (ΔE_C).

Although the PCC is not perfect at determining the best descriptor, similar to the correlation between ΔH_f and WCP of metal pairs, the PCC between the number of metal atoms in the 1NN shell around the dopant could be relevant to the metal–C or metal–N enthalpy of mixing, obtained from the literature,[57] as shown in Fig. 6. Metals with more negative values for the metal–C or metal–N enthalpy of mixing indicate stronger metal–C or metal–N bonds, leading to more negative ΔE_C or ΔE_N values. Thus the PCC could be still useful to understand the stable design of doping surrounding metal elements.

3.3. Quantification of correlation between multiple descriptors and doping energy

To further understand and quantify how the local environment influences the stability of the C- and N-doped HEA at different doping sites, we combined different local-environment descriptors and correlated them with ΔE_C or ΔE_N. In the linear regression model, adding highly linear dependent descriptors does not improve much the accuracy of the ΔE_C or ΔE_N regression models. In contrast, adding highly linear independent descriptors might improve the accuracy in predicting ΔE_C or ΔE_N. Thus, these linear regression models could be an appropriate tool to distinguish the contribution of the descriptors to the doping energy and their dependence. For example, the previous studies showed linear regression models that successfully predicted adsorption[10,58] or interstitial doping energies[19,59] using the composition of the 1NN–3NN shells around the adsorbate or dopant. The linear-regression model applied in this study is shown in eqn (8).

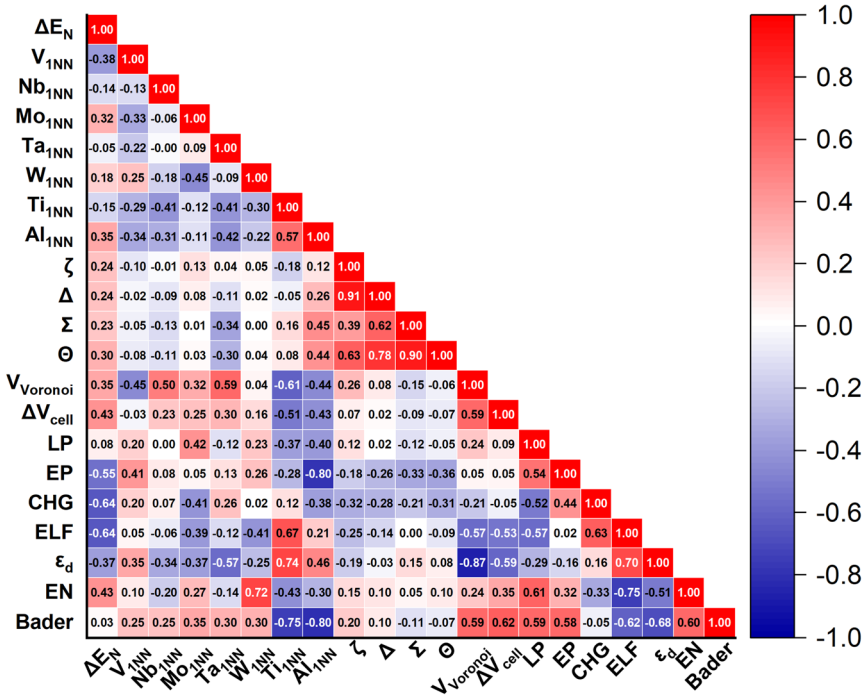

Fig. 5 The Pearson correlation coefficient (PCC) between all descriptors except the 2NN and 3NN composition around the dopant N and the nitrogen doping energy (ΔE_N).

Fig. 7 shows the coefficient of determination (Q^2) of the linear regression models *versus* different numbers of descriptors, demonstrating that the composition of the 1NN shell around the C or N dopant exhibits the highest Q^2 (0.51 or 0.61). Fig. S8 depicts the CV score of linear models *versus* different numbers of descriptors, revealing that the composition of the 1NN shell around the C or N dopant exhibits the lowest CV score (0.404 or 0.344). The parity plots in Fig. S13(a)

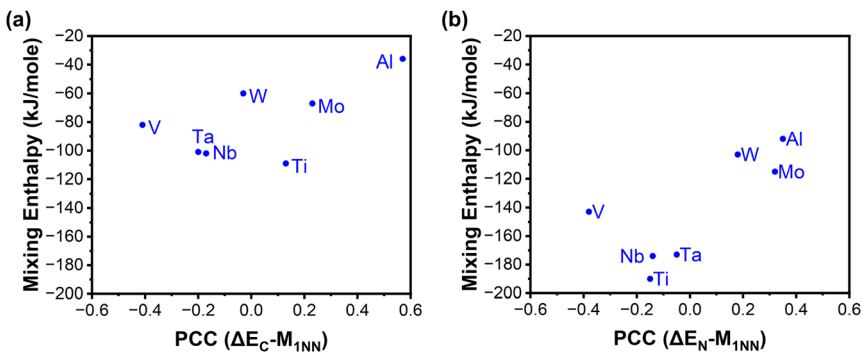

Fig. 6 Metal–C or metal–N enthalpy of mixing values from the literature[57] plotted against: (a) the Pearson correlation coefficient (PCC) between the carbon doping energy, ΔE_C, and the number of metal atoms in the 1NN shell of the dopant (M_{1NN}); and (b) the PCC between the nitrogen doping energy, ΔE_N, and M_{1NN}.

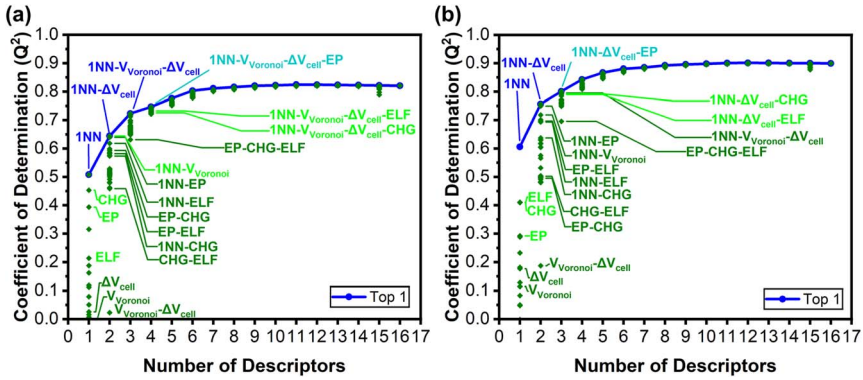

Fig. 7 The coefficient of determination (Q^2) plot against the number of descriptors used in the linear regression model to predict (a) ΔE_C or (b) ΔE_N. The combinations of descriptors related to the 1NN, volume, EP, CHG, and ELF descriptors are labelled. The highest Q^2 values are coloured in blue.

and(b) compare the DFT-calculated and linear-regression-model-predicted ΔE_C or ΔE_N results when only considering the 1NN descriptor. When only considering a single descriptor, both Q^2 and the CV score demonstrate that 1NN is the best descriptor to correlate the stability of the C- or N-doped HEA, stemming from the bonding energies of the sum of all metal–C or metal–N bonds in the 1NN shell. The contribution of the composition in the 1NN shell around the dopant to ΔE_N is larger than that to ΔE_C due to the more negative enthalpy of mixing values between the metals and N, compared to those between the metals and C, as shown in Fig. 6.

By adding volume-based descriptors, $V_{Voronoi}$ or ΔV_{cell}, into the ΔE_C or ΔE_N regression models with the 1NN descriptor, the Q^2 value from the linear regression model for the C- and N- doped HEA rise to 0.72 from 0.51, and to 0.76 from 0.61, respectively, demonstrating an enhancement of 21 or 15% explained variance for the ΔE_C or ΔE_N regression models, respectively. The CV score also decreases from 0.40 to 0.30 eV for the ΔE_C regression mode, and from 0.34 to 0.27 eV for the ΔE_N regression model, respectively, as shown in Fig. S12. The parity plots in Fig. S13(c) and (d) compare the DFT-calculated and linear-regression-model-predicted ΔE_C or ΔE_N results when considering both the 1NN and volume descriptors. These Q^2 and CV score outcomes suggest that the correlation between the descriptors and the C- or N-doped HEA stability is significantly enhanced by using the binary descriptor (1NN + volume).

When examining the Q^2 results of the ΔE_C regression models with only volume descriptors, their values are small with values of only −0.001, 0.026 and 0.023, respectively for $V_{Voronoi}$, ΔV_{cell}, and the combination of $V_{Voronoi}$ and ΔV_{cell}. As for ΔE_N, Q^2 is 0.177 obtained from the ΔV_{cell} descriptor. These volume descriptors show poor correlation in predicting the stability of both the C- and N-doped HEAs. However, when adding both the $V_{Voronoi}$ and ΔV_{cell} descriptors into the ΔE_C regression model with the 1NN descriptor, surprisingly, the Q^2 value increases from 0.51 to 0.72, exhibiting the best binary descriptor combination. Similarly, when adding the ΔV_{cell} descriptor into the ΔE_N regression model with the 1NN descriptor, the Q^2 value surprisingly increases from 0.61 to 0.76, also exhibiting

the best binary descriptor combination. These results imply that judging the stability prediction performance of a single descriptor without considering linear regression models does not provide an objective basis for identifying the best linear combination of descriptors. In this case, combining the best 1NN descriptor and the poor volume descriptor shows the best combined prediction performance. These outcomes stress the importance of linear combination of the independent predictive capability of different descriptors – not only considering the best 1NN descriptor but also not omitting the poor performance of a single descriptor.

Turning to correlating the electronic-structure-based descriptors, EP, CHG and ELF, the Q^2 values of the ΔE_C and ΔE_N regression models are 0.39, 0.45 and 0.21, respectively. Considering binary descriptors, after adding EP, CHG, or ELF descriptors into the ΔE_C regression model with the 1NN descriptor, the Q^2 value increases to 0.64, 0.57 and 0.62, which are still lower than the Q^2 value (0.72) obtained from the 1NN and volume (1NN + $V_{Voronoi}$ + ΔV_{cell}) descriptors. Similarly, for ΔE_N, the Q^2 value (0.76) obtained using the combination of the 1NN plus volume (1NN + ΔV_{cell}) descriptors exhibits better correlation compared to the 1NN plus EP, CHG, or ELF descriptors (0.75, 0.64 or 0.69, respectively). Thus, for the following ternary descriptor combinations, electronic-structure descriptors will only be added into the ΔE_C or ΔE_N regression models with the 1NN plus volume descriptors.

Considering the 1NN + volume + EP, 1NN + volume + CHG, and 1NN + volume + ELF descriptors in the ΔE_C regression models, their Q^2 values increase to 0.75, 0.73 and 0.73, respectively, showing a slight enhancement of ~0.02 for Q^2 compared to the binary descriptor (1NN + volume). Similarly, for ΔE_N, the Q^2 values (0.80, 0.79 or 0.79) for the combination of the 1NN and volume plus EP, CHG, or ELF descriptors exhibit an enhancement of ~0.04 compared to Q^2 for the binary descriptor. For both the ΔE_C and ΔE_N regression models, adding the third descriptor, EP, into the regression models with the 1NN plus volume descriptors exhibits the best correlation with doping energy compared to CHG and ELF. The β fitting coefficients of EP in the ΔE_C and ΔE_N regression models are −1.799 and −1.894, respectively, as listed in Tables S2 and S3, indicating that the more positive the EP value, the more negative the ΔE_C or ΔE_N value. A more positive EP at a vacant doping site signifies that the potential energy of an electron at that site is more positive, leading to unstable electrons nearby, which will be stabilized by C or N dopants with the formation of metal–C or metal–N bonds.

When using descriptors beyond the ternary level for the ΔE_C or ΔE_N regression models, the Q^2 values of these regression models will eventually reach an upper limit of 0.82 or 0.90, respectively, with only slightly enhancements of 0.07 and 0.10 in Q^2 compared to a ternary descriptor. To avoid complexity and identify the root cause in stabilizing the C- and N-doped HEA, we limited our descriptor optimisation to ternary combinations in this study. To sum up, the best ternary descriptor in both the ΔE_C and ΔE_N regression models is the 1NN + volume + EP descriptor, though a poor single volume descriptor is observed.

4. Conclusions

In this study, we have employed mainly density functional theory calculations, combined with Monte Carlo simulations and a machine-learning interatomic

potential to identify the optimal combination of microstructure- and electronic-structure-based local-environment descriptors in linear regression models to predict the stability of the C- or N-doped $VNbMoTaWTiAl_{0.5}$ (BCC) HEA.

The Warren–Cowley parameters (WCP) of different binary elemental pairs were applied to verify the stable short-range order in this HEA model with 156 atoms. The six types of microstructure-based descriptor include the composition of the first-, second-, and third-nearest neighbour shells (1NN, 2NN and 3NN), OctaDist distortion parameters (ζ, Δ, Σ and Θ) and two volume descriptors, Voronoi volume (V_{Voronoi}) of the dopant, and the volume change of the unit cell after doping (ΔV_{cell}). The associated seven types of electronic-structure-based descriptor are the local potential (LP), the electrostatic potential (EP), the charge density (CHG), the electron localization function (ELF) at the vacant doping site, the d-band center (ε_{d}), the mean electronegativity (EN) of the 1NN shell around the dopant, and the Bader charge of the C or N dopant. Pearson correlation coefficients (PCC) were employed to judge the correlation between a single descriptor and the C- or N doping energy, ΔE_{C} or ΔE_{N}.

The best descriptor to correlate the doping energy (indication of HEA stability) using a single descriptor is 1NN, with coefficient of determination (Q^2) values of ~51 and ~61% obtained using the LOOCV (leave-one-out cross-validation) approach for C or N doping, respectively. In contrast, for the single volume descriptor, the Q^2 value is ~0% for both the ΔE_{C} and ΔE_{N} regression models, indicating poor correlation between the volume descriptor and the doping energy. Nevertheless, after adding the volume descriptor into the linear regression model with the 1NN descriptor, Q^2 increases to 72 and 76% for C and N doping, respectively (approaching their Q^2 upper limits of 80 and 90%, respectively). Furthermore, after adding the third descriptor, electronic-structure-based EP, Q^2 further improves to 75 and 80% for C and N doping, respectively, exhibiting the best ternary descriptors correlation. This study constructs a workflow on a doped HEA system from modeling (special quasirandom structures, machine learning interatomic potential and Monte Carlo simulations) to stability verification (DFT calculations and linear regression models). We quantitatively combined and compared the independent contributions of different types of local-environment descriptors to the stability of the C- or N-doped HEA, demonstrating the importance of considering both key microstructure-based and electronic-structure-based local-environment descriptors using linear regression models to achieve more accurate correlation of dopant stability in the HEA. We hope that these combined approaches could be further applied to other materials systems, research fields and applications.

Conflicts of interest

There are no conflicts to declare.

Data availability

Structures of all simulations are available upon request to the authors.

The data supporting this article have been included as part of the SI. See DOI: **https://doi.org/10.1039/d5fd00107b**.

Acknowledgements

C.-H. Lee, and H.-Y. T. Chen acknowledge the National Science and Technology Council, NSTC (111-2221-E-007-087-MY3, 111-2112-M-007-028-MY3, 113-2923-E-008-007, 114-2124-M-007-007 and 114-2112-M-007-041) in Taiwan for their financial support. The computational resources were supported by TAIWANIA at the National Center for High-Performance Computing (NCHC) of the National Applied Research Laboratories (NARLabs) in Taiwan. This work used the ARCHER2 UK National Supercomputing Service (**https://www.archer2.ac.uk**).[60] We thank the University of Liverpool for use of High Performance Computing resources. We acknowledge the contribution of Matthew S. Dyer for useful discussion. J.-W. Lee acknowledges the financial support from the NSTC, Taiwan (113-2224-E-131-001, 113-2221-E-131-024, 114-2224-E-131-001 and 114-2221-E-131-001). This work was also financially supported by the "High Entropy Materials Center" from The Featured Areas Research Center Program within the Higher Education Sprout Project framework by the Ministry of Education (MOE) in Taiwan.

References

1 J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau and S.-Y. Chang, *Adv. Eng. Mater.*, 2004, **6**, 299–303.
2 O. N. Senkov, J. K. Jensen, A. L. Pilchak, D. B. Miracle and H. L. Fraser, *Mater. Des.*, 2018, **139**, 498–511.
3 T.-K. Tsao, A.-C. Yeh, C.-M. Kuo, K. Kakehi, H. Murakami, J.-W. Yeh and S.-R. Jian, *Sci. Rep.*, 2017, **7**, 12658.
4 N. Yurchenko, E. Panina, Ł. Rogal, L. Shekhawat, S. Zherebtsov and N. Stepanov, *Mater. Res. Lett.*, 2022, **10**, 78–87.
5 O. N. Senkov, G. B. Wilks, J. M. Scott and D. B. Miracle, *Intermetallics*, 2011, **19**, 698–706.
6 O. N. Senkov, J. M. Scott, S. V. Senkova, D. B. Miracle and C. F. Woodward, *J. Alloys Compd.*, 2011, **509**, 6043–6048.
7 B. Zhang, Y. Huang, Z. Dou, J. Wang and Z. Huang, *J. Sci.: Adv. Mater. Devices*, 2024, **9**, 100688.
8 T. K. Chen, T. T. Shun, J. W. Yeh and M. S. Wong, *Surf. Coat. Technol.*, 2004, **188–189**, 193–200.
9 B. Ren, S. J. Lv, R. F. Zhao, Z. X. Liu and S. K. Guan, *Surf. Eng.*, 2014, **30**, 152–158.
10 T. A. A. Batchelor, J. K. Pedersen, S. H. Winther, I. E. Castelli, K. W. Jacobsen and J. Rossmeisl, *Joule*, 2019, **3**, 834–845.
11 J. K. Pedersen, T. A. A. Batchelor, A. Bagger and J. Rossmeisl, *ACS Catal.*, 2020, **10**, 2169–2176.
12 Z. W. Chen, J. Li, P. Ou, J. E. Huang, Z. Wen, L. Chen, X. Yao, G. Cai, C. C. Yang, C. V. Singh and Q. Jiang, *Nat. Commun.*, 2024, **15**, 359.
13 J. Shi, Y. Lei, N. Hashimoto and S. Isobe, *Mater. Trans.*, 2020, **61**, 616–621.
14 H. Song, M. Yu, Y. Zhang, W. Zhang, Z. Liu, F. Zhang and F. Tian, *Mater. Today Commun.*, 2022, **31**, 103241.
15 Y. X. Ye, B. Ouyang, C. Z. Liu, G. J. Duscher and T. G. Nieh, *Acta Mater.*, 2020, **199**, 413–424.

16 Z. Lei, X. Liu, Y. Wu, H. Wang, S. Jiang, S. Wang, X. Hui, Y. Wu, B. Gault, P. Kontis, D. Raabe, L. Gu, Q. Zhang, H. Chen, H. Wang, J. Liu, K. An, Q. Zeng, T.-G. Nieh and Z. Lu, *Nature*, 2018, **563**, 546–550.
17 J. B. Seol, J. W. Bae, Z. Li, J. Chan Han, J. G. Kim, D. Raabe and H. S. Kim, *Acta Mater.*, 2018, **151**, 366–376.
18 I. Moravcik, M. Zelený, A. Dlouhy, H. Hadraba, L. Moravcikova-Gouvea, P. Papež, O. Fikar, I. Dlouhy, D. Raabe and Z. Li, *Sci. Technol. Adv. Mater.*, 2022, **23**, 376–392.
19 L. Casillas-Trujillo, U. Jansson, M. Sahlberg, G. Ek, M. M. Nygård, M. H. Sørby, B. C. Hauback, I. A. Abrikosov and B. Alling, *Phys. Rev. Mater.*, 2020, **4**, 123601.
20 N. Yang, L. Zhu, H. Liu, J. Zhou and Z. Sun, *J. Mater. Sci. Technol.*, 2025, **227**, 133–141.
21 O. N. Senkov, G. B. Wilks, D. B. Miracle, C. P. Chuang and P. K. Liaw, *Intermetallics*, 2010, **18**, 1758–1765.
22 M. Moorehead, K. Bertsch, M. Niezgoda, C. Parkin, M. Elbakhshwan, K. Sridharan, C. Zhang, D. Thoma and A. Couet, *Mater. Des.*, 2020, **187**, 108358.
23 B.-S. Lou, C.-L. Li, M. Annalakshmi, T.-Y. Hung and J.-W. Lee, *Mater. Chem. Phys.*, 2025, **341**, 130901.
24 K. Tiwari, C.-H. Wang, B.-S. Lou, A. M. Demeku, I. Moirangthem, S. Wang, I. Rahmadtulloh, C.-J. Wang, W. Huo and J.-W. Lee, *J. Power Sources*, 2025, **654**, 237826.
25 R. Ketkaew, Y. Tantirungrotechai, P. Harding, G. Chastanet, P. Guionneau, M. Marchivie and D. J. Harding, *Dalton Trans.*, 2021, **50**, 1086–1096.
26 P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.
27 W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.
28 G. Kresse and J. Furthmüller, *Comput. Mater. Sci.*, 1996, **6**, 15–50.
29 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 558–561.
30 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 14251–14269.
31 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.
32 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.
33 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
34 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953–17979.
35 A. Zunger, S. H. Wei, L. G. Ferreira and J. E. Bernard, *Phys. Rev. Lett.*, 1990, **65**, 353–356.
36 A. van de Walle, P. Tiwary, M. de Jong, D. L. Olmsted, M. Asta, A. Dick, D. Shin, Y. Wang, L. Q. Chen and Z. K. Liu, *Calphad*, 2013, **42**, 13–18.
37 A. van de Walle, M. Asta and G. Ceder, *Calphad*, 2002, **26**, 539–553.
38 D. Waroquiers, J. George, M. Horton, S. Schenk, K. A. Persson, G. M. Rignanese, X. Gonze and G. Hautier, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2020, **76**, 683–695.
39 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319.
40 B. Hammer and J. K. Norskov, *Nature*, 1995, **376**, 238–240.

41 B. Hammer and J. K. Nørskov, in *Advances in Catalysis*, Academic Press, 2000, vol. 45, pp. 71–129.
42 L. Pauling, *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: an Introduction to Modern Structural Chemistry*, Cornell university press, 1960.
43 W. Tang, E. Sanville and G. Henkelman, *J. Phys.: Condens. Matter*, 2009, **21**, 084204.
44 E. Sanville, S. D. Kenny, R. Smith and G. Henkelman, *J. Comput. Chem.*, 2007, **28**, 899–908.
45 G. Henkelman, A. Arnaldsson and H. Jónsson, *Comput. Mater. Sci.*, 2006, **36**, 354–360.
46 M. Yu and D. R. Trinkle, *J. Chem. Phys.*, 2011, **134**, 064111.
47 N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, *J. Chem. Phys.*, 1953, **21**, 1087–1092.
48 X. Liu, J. Zhang, J. Yin, S. Bi, M. Eisenbach and Y. Wang, *Comput. Mater. Sci.*, 2021, **187**, 110135.
49 I. Batatia, D. P. Kovacs, G. Simm, C. Ortner and G. Csányi, *Adv. Neural Inf. Process. Syst.*, 2022, **35**, 11423–11436.
50 I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky and G. Csányi, *Nat. Mach. Intell.*, 2025, **7**, 56–67.
51 J. M. Cowley, *Phys. Rev.*, 1950, **77**, 669–675.
52 E. Lee, H. Iddir and R. Benedek, *Phys. Rev. B*, 2017, **95**, 085134.
53 Q. Zhao, M. Avdeev, L. Chen and S. Shi, *Sci. Bull.*, 2021, **66**, 1401–1408.
54 E. Lu, J. Zhao, I. Makkonen, K. Mizohata, Z. Li, M. Hua, F. Djurabekova and F. Tuomisto, *Acta Mater.*, 2021, **215**, 117093.
55 I. Toda-Caraballo, J. S. Wróbel, D. Nguyen-Manh, P. Pérez and P. E. J. Rivera-Díaz-del-Castillo, *JOM*, 2017, **69**, 2137–2149.
56 M. C. Troparevsky, J. R. Morris, P. R. C. Kent, A. R. Lupini and G. M. Stocks, *Phys. Rev. X*, 2015, **5**, 011041.
57 A. Takeuchi and A. Inoue, *Mater. Trans.*, 2005, **46**, 2817–2829.
58 J. K. Pedersen, C. M. Clausen, O. A. Krysiak, B. Xiao, T. A. A. Batchelor, T. Loffler, V. A. Mints, L. Banko, M. Arenz, A. Savan, W. Schuhmann, A. Ludwig and J. Rossmeisl, *Angew Chem. Int. Ed. Engl.*, 2021, **60**, 24144–24152.
59 J. Schuett, T. K. Schultze and S. Grieshammer, *Chem. Mater.*, 2020, **32**, 4442–4450.
60 G. Beckett, J. Beech-Brandt, K. Leach, Z. Payne, A. Simpson, L. Smith, A. Turner and A. Whiting, ARCHER2 Service Description, 2024, DOI: **10.5281/zenodo.14507040**.