Physics-informed dual-objective optimization of high-entropy-alloy nanozymes by a robotic AI chemist

Man Luo[1,†], Zikai Xie[2,†], Huirong Li[1,†], Baicheng Zhang[1], Jiaqi Cao[1], Yan Huang[1], Qing Zhu[1,3,*], Linjiang Chen[1,4,*], Jun Jiang[1,*], Yi Luo[1,*]

[1] Key Laboratory of Precision and Intelligent Chemistry, Hefei National Research Center for Physical Sciences at the Microscale, School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, China
[2] Department of Chemistry, University of Liverpool, Liverpool, UK
[3] Institute of Intelligent Innovation, Henan Academy of Sciences, Zhengzhou, China
[4] School of Chemistry and School of Computer Science, University of Birmingham, Birmingham, UK
[†] These authors contributed equally: M.L., Z.X., H.L.
E-mails: qingzhu@ustc.edu.cn; linjiangchen@ustc.edu.cn; jiangj1@ustc.edu.cn; yiluo@ustc.edu.cn

Abstract

Engineering artificial nanozymes as substitutes for natural enzymes presents a significant scientific challenge. High entropy alloys (HEAs) have emerged as promising candidates for mimicking peroxidase (POD) activity thanks to their unique properties and versatility. However, designing or discovering HEAs that surpass the catalytic efficiency of natural horseradish peroxidase involves complex challenges, often hindered by the multidimensional nature of HEAs' compositional variability and the intricate interplay of enzymatic behaviours. Therefore, an intelligent and efficient approach to accelerate this discovery is crucial. In this study, we address these challenges by deploying a robotic artificial-intelligence chemist equipped with theoretical calculations, machine learning, Bayesian optimization, and on-the-fly data analysis by a large language model (LLM). Our approach centres on a physics-informed, multi-objective optimization framework that simultaneously optimizes multiple desirable properties of nanozymes, including maximum reaction rate and substrate affinity, ultimately optimizing catalytic efficiency. By integrating an auxiliary knowledge model based on physical insights and collaborative decision-making enabled by LLM-in-the-loop into Bayesian optimization, we enhanced the data-driven discovery workflow. Our physics-informed approach, with instant LLM-in-the-loop feedback, significantly outperformed both random sampling and standard Bayesian optimization. Consequently, we efficiently explored a vast chemical space and identified HEAs with enzymatic properties that significantly exceed those of the most effective catalysts based on HEAs or single atoms reported in the literature, as well as the natural enzyme.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Main

The engineering of artificial nanozymes as substitutes for natural enzymes has long been a forefront topic in enzymatic catalysis research.[1–3] Nanozymes are nanomaterials with enzymatic properties, which often exhibit advantages over natural enzymes in environmental tolerance, long-term stability, and manufacturing costs.[4,5] Nanozymes mimicking peroxidase (POD) have been extensively studied as alternatives to the natural enzyme horseradish peroxidase (HRP). POD mimics catalyse the conversion of hydrogen peroxide into hydroxyl radicals and have widespread applications in biosensing, disease treatment, and immunoassays.[6,7] The discovery of the POD-mimicking properties of Fe_3O_4, by Gao *et al.*,[8] marked the beginning of a series of studies on noble metals,[7,9,10] alloys,[11,12] metal oxides,[13] and metal sulfides[14] for their POD-mimicking capabilities. The detailed structural elucidation of natural HRP, revealing a bioactive centre featuring a FeN_4 configuration, catalysed the development of single-atom nanozymes as HRP mimics.[15–19] Engineering such bioactive centres requires atomic-scale insights into the structure and catalytic behaviours of nanomaterials. Thus, computational studies, particularly density functional theory (DFT) calculations, have been conducted to help rationalize experimental findings and investigate possible structure-function relationships that inform the design of nanozymes.[20–22]

Despite extensive research into POD nanozymes, surpassing the catalytic efficiency of HRP remains a challenge. Most mimics reported to date have been either discovered serendipitously or designed as close variants or derivatives of previously known working systems. While establishing structure-function relationships for POD mimics has proven helpful—for instance, in predicting the catalytic activity of single-atom nanozymes—such relationships often fail to capture the essence of HRP's biocatalytic strategies, which are multifaceted and multifactorial, spanning multiple length and time scales. This is evident in cases where, even though the local molecular and electronic structures of a mimic's active centre may resemble that of HRP, the mimic still fails to achieve equivalent catalytic efficacy.

Natural enzymes possess highly evolved active sites that ensure a strong binding affinity towards the substrate and high catalytic activity. The Michaelis-Menten kinetics model,[23] widely used in biochemistry, describes the rate of enzymatic reactions and elucidates how various factors influence the efficiency and rates of these reactions. This model hinges on the formation of an enzyme-substrate complex as an intermediate step in the reaction process, leading to product formation while the enzyme is regenerated. At its core, the model features two critical parameters quantifying the enzyme-catalysed reaction: the maximum reaction rate, V_{max}, achieved by the system at saturating substrate concentration, and the Michaelis constant, K_M, which is the substrate concentration at which the reaction rate is half of V_{max}, reflecting the enzyme's affinity for the substrate.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

According to the Michaelis-Menten model, a highly active enzyme exhibits several key characteristics that enhance its ability to efficiently catalyse reactions.[24] Firstly, it should have a low K_M value, indicating high affinity for its substrate, allowing it to bind effectively even at low substrate concentrations. This is particularly important in environments where substrate availability is limited. Secondly, a high V_{max} value is essential, signifying that the enzyme can efficiently convert substrate to product once enzyme-substrate complexes are formed, processing substrate molecules into products as rapidly as its active site allows. Thirdly, a high V_{max}/K_M ratio, reflecting catalytic efficiency, suggests that the enzyme can achieve high reaction rates even with low substrate concentrations. Some other properties, while not described by the Michaelis-Menten model, such as stability and specificity, are also important and contribute to an enzyme's effectiveness in complex biochemical contexts.

While natural enzymes seemingly perform their functions with ease, it has proven to be a significant challenge for nanozymes to simultaneously optimize the V_{max}, K_M, and V_{max}/K_M properties, which are crucial for achieving the catalytic efficiency of natural enzymes.[25,26] On one hand, achieving a high V_{max} necessitates not only an abundance of accessible catalytic sites but also a rapid conversion of substrate to product. However, the often-limited availability and/or accessibility of catalytic sites of nanomaterials lead to suboptimal substrate conversion rates, posing a fundamental challenge to enhancing V_{max}. On the other hand, attaining low K_M values, indicating strong substrate affinity, requires high specificity and affinity akin to natural enzyme-substrate interactions. This is difficult to achieve in nanomaterials as it demands precise control over the nanozyme's surface chemistry and the microenvironment around its active sites. Furthermore, optimizing the V_{max}/K_M ratio, indicative of catalytic efficiency, necessitates a careful balance between increasing V_{max} and reducing K_M. Enhancements in V_{max} often lead to higher K_M values due to a loss of substrate specificity, a consequence of attempts to increase the reactivity of catalytic sites. Conversely, strategies aimed at reducing K_M by enhancing substrate affinity and specificity can inadvertently decrease V_{max}, by restricting the rate of conversion or the accessibility of catalytic sites.

High entropy alloys (HEAs) emerge as a versatile platform for the development of nanozymes, particularly because of their potential for simultaneously optimizing the key enzymatic parameters: V_{max} (maximum reaction rate), K_M (substrate affinity), and the efficiency ratio V_{max}/K_M. This multifaceted optimization may be facilitated by the unique attributes of HEAs, including their vast compositional diversity and the resulting synergistic interactions among constituent elements.[27,28] These features allow for the precise tuning of the electronic structure and the spatial configuration of catalytic sites,[29–31] directly enhancing V_{max} by promoting rapid substrate conversion while also enabling the modulation of substrate binding sites to reduce K_M. Furthermore, the structural complexity and inherent stability of HEAs contribute to their robustness under various operational conditions, maintaining the integrity of catalytic sites and

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

ensuring consistent enzymatic performance. The ability to engineer surface properties further enhances the potential of HEAs to achieve high catalytic efficiency alongside specificity.

The compositional diversity of HEAs, while a strength, introduces significant combinatorial complexity, posing a challenge in optimizing these materials for nanozyme applications, especially when aiming to enhance multiple desirable properties concurrently. Given that HEAs consist of five or more principal elements in substantial proportions, the vast compositional space vastly increases the number of potential materials, making the identification of optimal compositions more complex. This challenge is compounded when attempting to optimize multiple interrelated properties simultaneously, as enhancing one characteristic can inadvertently impact others due to the synergistic and, at times, antagonistic interactions among the constituent elements. Furthermore, while computational methods and machine learning offer tools to navigate this complex compositional space,[32,33] the effectiveness of these models depends on the availability of reliable data and the intricacy of the underlying physical phenomena.[3435] The multifaceted nature of HEAs, including aspects like electronic structure and surface reactivity, renders predictive modelling particularly challenging. The challenge extends beyond theoretical considerations. While high-throughput experimental screening promises an efficient exploration of the compositional space, its application to HEAs is still emerging. Moreover, the complexity of assays needed to simultaneously assess multiple catalytic properties introduces further complications. Consequently, the development of HEAs as nanozymes with multiple desirable properties necessitates collaborative efforts spanning theory, computation, machine learning, and experiment.

In this work, we explored the discovery of HEAs as POD nanozymes using a robotic chemist equipped with theoretical calculations, artificial intelligence (AI) capabilities—machine learning (ML), Bayesian optimization, and on-the-fly data analysis by a large language model (LLM)—and automated material synthesis and function testing. By developing and deploying a physics-informed, dual-objective Bayesian optimization approach, we investigated a vast chemical space of five-component HEAs, aiming to discover HEAs with enhanced maximum reaction rate (high V_{max} values) and substrate affinity (low K_M values), ultimately attaining high catalytic efficiencies (high V_{max}/K_M values). Notably, the integration of a knowledge model based on theoretical predictions and LLM-in-the-loop analyses of experimental data significantly improved the optimization efficiencies compared to random sampling and standard Bayesian optimization. After synthesizing and testing 96 HEAs, following an initial random set of 50 HEAs as the starting point, we discovered HEAs with a tenfold increase in catalytic efficiency compared to natural horseradish peroxidase; these HEAs also surpassed some of the most effective catalysts based on single atoms and other HEAs reported in the literature.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

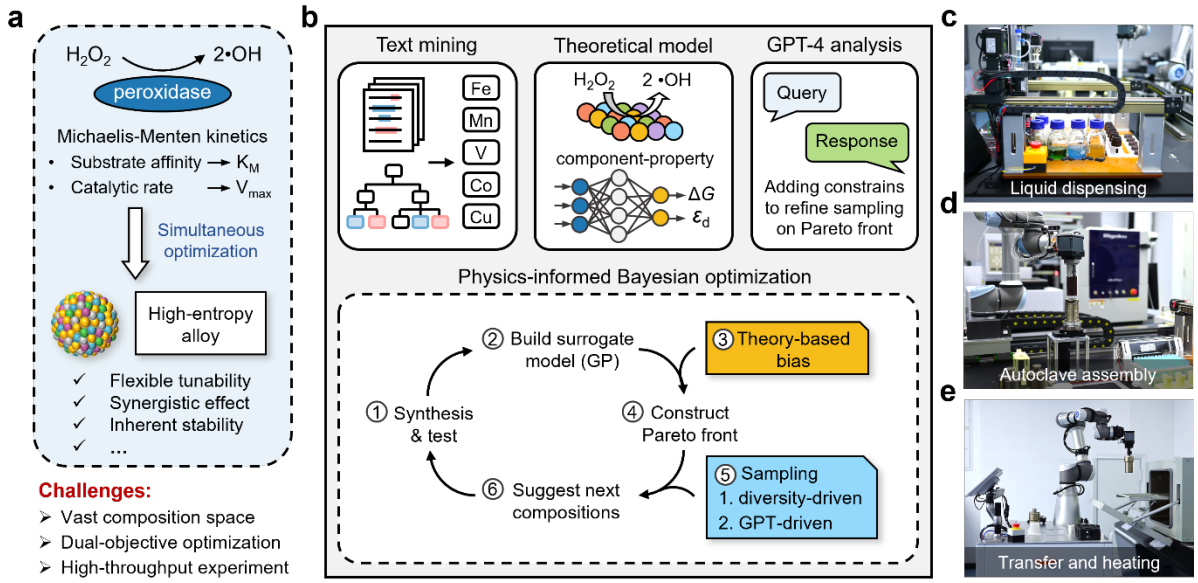


Figure 1. Accelerating the discovery of HEAs as POD nanozymes. a, Objectives for and advantages of HEA-based nanozymes, and the challenges associated with discovering them. **b**, Scheme of our discovery workflow, involving literature text mining to identify candidate metallic elements for HEAs, theoretical predictions for building a knowledge model to be incorporated into Bayesian optimization (BO); LLM-in-the-loop analyses of experimental data (using GPT-4), and a closed-loop physics-informed, dual-objective BO process. **c–e**, HEA synthesis by the robotic AI chemist, involving dispensing liquids (**c**), assembling autoclaves (**d**), and transferring assembled autoclaves to the oven for hydrothermal synthesis (**e**).

Results

Automating HEA synthesis

In this work, we considered HEAs consisting of five principal metallic elements; details are given below for how the specific elements were chosen for synthesis. Considering the ease with which it could be integrated into an automated robotic workflow, we adopted a mild solvothermal method to synthesize HEA nanoparticles.[36,37] This involved dissolving five different metal acetylacetonate precursors in an acetone and ethanol mixture (50:50 vol%). Acetone was used to fully dissolve dispensed metal acetylacetonates, and ethanol was used as a reducing agent under solvothermal conditions to reduce the mixed metal acetylacetonates to form HEAs. All metal acetylacetonate solutions were initially prepared with a total metal concentration of 10 mM. For the synthesis of a specific HEA, the required volumes of various precursor solutions were added according to the provided ratios. A 30 mL precursor solution with the desired ratio of metal acetylacetonates was mixed in a Teflon lined steel autoclave, which was subsequently sealed and heated in an oven at 200 °C for a duration of 4 hours. We randomly performed X-ray diffraction on 42 groups of HEAs with different synthesis formulations (Supplementary Table S1), which showed a face-centred cubic structure with relatively low crystallinity (Supplementary Fig. S1).

The hydrothermal synthesis of HEAs described above was achieved through an automated

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

workflow, extending our previously established robotic experimental capabilities.[38,39] Figure 1c and Extended Data Fig. 1(1, 2) depict an automated liquid dispensing station where precursor solutions are drawn from reagent bottles and pipetted into inner tanks for hydrothermal autoclaves. This workstation can prepare five inner tanks simultaneously with high dispensing accuracy. Figure 1d and Extended Data Fig. 1(3–7) illustrate a six-degree-of-freedom robotic arm securing the lid on the inner tank before inserting it into a stainless-steel autoclave. The autoclave is then positioned on a rotating platform and securely held in place, with the lid fastened by the robotic arm as the platform rotates. Figure 1e and Extended Data Fig. 1(8, 9) show a mobile robot collecting the assembled autoclave and transferring it to the oven for solvothermal synthesis. Full experimental details are provided in the Methods section and in the Supplementary Information.

Determining metallic elements for HEAs as potential POD nanozymes

Previously,[40] we developed an unsupervised syntactic distance analysis approach for mining scientific literature, capable of extracting information about chemical substances, physicochemical properties, chemical/material functions, experimental conditions, and more, without the need for annotation. Here, we applied this approach to analyse the abstracts of recent scientific literature and patents containing the keyword "peroxidase" from the Web of Science, totalling 98,881 documents (data retrieval date: September 2022). Among this extensive collection, only 14,597 scientific publications and 847 patents mentioned both "peroxidase" and at least one metallic element in their abstracts. Moreover, these metallic elements were mentioned in various forms, including metal oxides, alloys (high-entropy or otherwise), organometallic compounds, and more. We posited that the co-occurrence of "peroxidase" and metallic elements in abstracts indicates a significant relevance of metal-containing materials to peroxidase-related functions or applications. Thus, the identified prevalence of metallic elements holds strong potential for pinpointing metallic elements suitable for HEAs as peroxidase nanozymes.

Extended Data Fig. 2a presents the identified frequencies of various metallic elements, depicted as counts of their occurrences. Additionally, Extended Data Fig. 2b shows a correlation heatmap that illustrates the co-occurrences of metal pairs within the same abstracts. For instance, iron (Fe) is often mentioned alongside vanadium (V), copper (Cu), and Cobalt (Co), indicating potential synergies among these elements. Furthermore, considering HEAs' formation, the structural and chemical similarities of the components are important. Therefore, we refined the metal-pair correlation map by considering the metallic elements' atomic radii and counts of their valence electrons, assigning enhanced correlations to similar elements (based on these criteria), as represented by thicker lines connecting them in Extended Data Fig. 2c. This refinement resulted in the final recommendations shown in Figure 2a. Consequently, the top five recommended elements for the synthesis of HEAs in this study were determined to

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

be Fe, Mn (manganese), V, Co, and Cu.

Building a theory-based predictive model for enhancing BO efficiency

We explored HEAs composed of five selected metallic elements, each within the range of 5 to 35 at.%, resulting in a total of 553,401 potential unique compositions. To navigate this extensive chemical space efficiently, we employed Bayesian Optimization (BO) for its sample efficiency in evaluating functions whose forms and derivatives are unknown, akin to experimental scenarios. Additionally, we leveraged a previously developed method to improve optimization efficiency by integrating prior knowledge into the standard BO algorithm.[41] Specifically, this enhanced BO framework allows the inclusion of external knowledge into the acquisition function through an additional deterministic surrogate model. This approach enhances the standard single surrogate model's approximation capabilities, which is typically a Gaussian Process (GP). The added surrogate model can be any ML model that offers predictions to facilitate more informed sampling decisions.

In this study, we employed an ML model capable of estimating computable properties that represent certain underlying characteristics of the enzymatic functions of interest. By integrating this additional surrogate model, we introduced a bias towards regions in the search space that are more likely to exhibit these underlying characteristics, thereby potentially leading us to solutions with enhanced enzymatic functions more swiftly. The surface desorption energy of the hydroxyl radical has been proposed as a descriptor for the catalytic performance of peroxidase (POD) nanozymes,[21] whose kinetic parameters are determined based on the oxidative coloration of the substrate by the desorbed hydroxyl radicals. It has also been suggested that the surface adsorption/desorption energy can be related to the surface structure through the so-called d-band model.[42] Therefore, we selected the Gibbs free energy for the surface desorption of a hydroxyl radical and the centres of the d-band before and after desorption as the underlying characteristics for constructing the additional surrogate model.

Results from electron spin resonance spectroscopy (Supplementary Fig. S2) revealed that the catalysis of H_2O_2 by the HEA nanozyme follows a homolytic reaction pathway (Extended Data Fig. 3): (i) H_2O_2 is adsorbed onto the HEA surface; (ii) the activated H_2O_2 undergoes homolysis, forming two hydroxy groups (–OH); (iii) one hydroxy group desorbs from the surface, yielding a hydroxyl radical (•OH). Since the homolysis step is thermodynamically favourable, the desorption of the hydroxy groups becomes a critical step in the overall reaction. Aligning with this mechanistic view, the pipeline for calculating the relevant thermodynamic features to be used by the additional surrogate model in the enhanced BO framework is outlined in Figure 2b.

First, classical molecular dynamics (MD) simulations were employed to generate atomistic configurations for 12,205 distinct compositions of the five selected metallic elements. Then,

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

from each configuration, representing a unique composition, we extracted 6,400 local structures, each comprising 16 neighbouring atoms. These structures served as representative surface structures of the composition for calculating the Gibbs free energy of desorbing the first hydroxy group (ΔG_{des}^{-OH}), as well as for determining the d-band centre positions before and after the desorption of the first hydroxy group ($\varepsilon_d^{2(-OH)}$ and ε_d^{-OH}, respectively). Since 6,400 local structures represented each composition and were calculated for ΔG_{des}^{-OH}, $\varepsilon_d^{2(-OH)}$, and ε_d^{-OH}, we computed the mean and median of $\varepsilon_d^{2(-OH)}$ and ε_d^{-OH}, and the probability of ΔG_{des}^{-OH} being above -1 eV across all 6,400 samples for the composition, yielding 5 thermodynamic features for the composition. Subsequently, a neural network (NN) model was trained to predict these 5 features using only the composition as input, employing the 12,205 compositions for training. Ultimately, this NN model was used to predict the 5 thermodynamic features for any given composition, which were then used by the additional surrogate model in the enhanced BO framework.

To build atomistic models for the calculations described above, it is necessary to know the compositions of the HEAs, which can deviate from the synthesis formulations (*i.e.*, the precursor dosages used during synthesis). Indeed, we synthesized 150 HEAs with randomly chosen synthesis formulations, with the dosage of each element within the range of 5 to 35 at.%. We then determined the actual composition of each HEA through inductively coupled plasma–optic emission spectrometry (see Methods section for details). Extended Data Figure 4(a–e) shows that, generally, the dosages used in synthesis correspond well with the actual compositional amounts across the 150 HEAs, albeit with appreciable differences for individual ones. Our Bayesian optimization, which guides experimental syntheses of HEAs in the formulation space, employs an external knowledge model to incorporate theoretical predictions. Therefore, calibrating synthesis formulations to HEA compositions is required so that Bayesian optimization can evaluate any points in the search space and suggest subsequent experiments. To accomplish this, we built a simple feedforward neural network that took a synthesis formulation as input and outputted its predicted composition. Extended Data Figure 4(f–o) confirms the high predictive accuracies of this calibration model for a test set from the 150 randomly chosen HEAs (f–j) and for all other samples synthesized in this study (k–o).

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

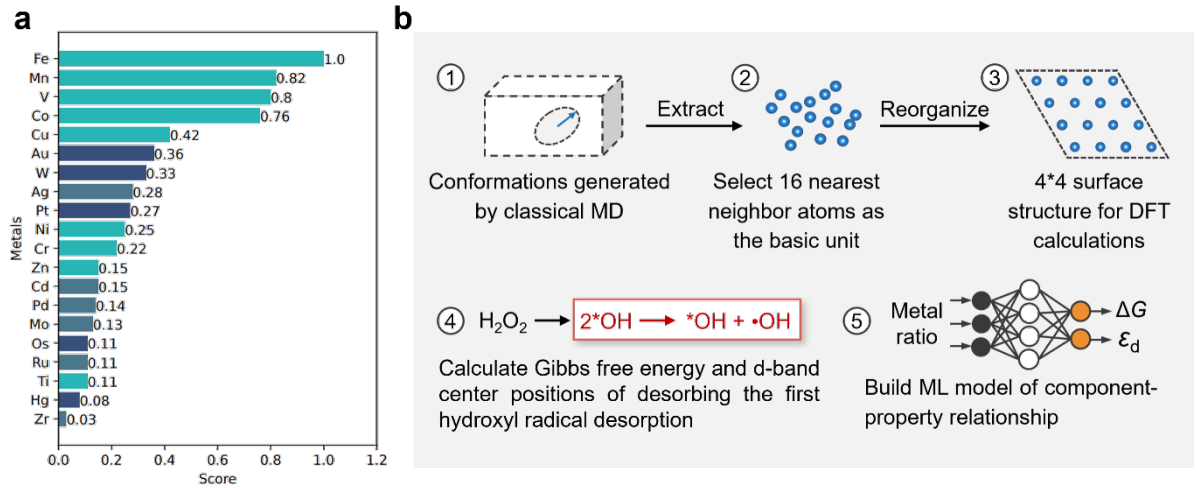


Figure 2. Literature text mining results and the workflow of theoretical calculations and predictions. a, Recommendation ranking of candidate metallic elements for synthesizing HEAs, taking into consideration the prevalence of metallic elements in the literature mentioning "peroxidase" in their abstracts and the size and chemical compatibility between elements. **b**, Workflow of building a machine learning (ML) model for predicting thermodynamic properties from HEA compositions, involving molecular dynamics (MD) simulations, first-principles density functional theory (DFT) calculations, and building and training neural network (NN) models.

Developing a physics-informed, dual-objective BO strategy

The theory-based NN model, which predicts the five thermodynamic features based on calculated ΔG_{des}^{-OH}, $\varepsilon_d^{2(-OH)}$, and ε_d^{-OH} for a given HEA composition, was incorporated as an auxiliary model within the standard Bayesian Optimization (BO) framework, representing prior knowledge throughout the optimization process. This method is hereafter referred to as physics-informed BO, as it integrates theoretical predictions into the BO framework. Specifically, for each synthesis formulation x, the calibration model predicted the corresponding composition x'. For composition x', the theory-based NN model predicted the five thermodynamic features, collectively termed domain knowledge $\mathcal{D}$, which was then used to estimate the optimization objective by a knowledge model $\xi(x';\mathcal{D})$. That is, the optimization objective was estimated using both the GP model, based on the synthesis formulation x, and the knowledge model, based the thermodynamic properties derived for composition x'. These two models were combined into an augmented acquisition function: $\alpha_{\mathcal{D}}(x|x') = \alpha(x) + \xi(x';\mathcal{D})$, where $\alpha(x)$ is the acquisition function in standard BO, chosen to be the upper confidence bound (UCB) in this study. Therefore, the full expression of the augmented acquisition used is given by

$$\alpha_{\mathcal{D}}(x|x';\kappa) = \alpha_{UCB}(x;\kappa) + \xi(x';\mathcal{D}) = [\mu(x) + \kappa\sigma(x)] + \xi(x';\mathcal{D}) \quad \text{(Eq. 1)}$$

where $\alpha_{UCB}(x;\kappa)$ is a weighted (by κ) sum of the posterior mean, $\mu(x)$, and uncertainty, $\sigma(x)$, both given by the GP model for x.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Furthermore, as discussed above, HEAs have significant potential for simultaneously optimizing several key enzymatic parameters, aiming for a large V_{max} (high maximum reaction rate), a small K_M (high substrate affinity), and a high efficiency ratio V_{max}/K_M all at once. Thus, we implemented a dual-objective BO method, searching for solutions (*i.e.*, HEA synthesis formulations) that increase V_{max} and decrease K_M. In multi-objective optimization tasks, it is common for a solution to perform well on one objective but poorly on the others. A solution is considered nondominated or Pareto optimal if any improvement in one of the objective functions would lead to a compromise in others. Conversely, a solution is dominated if there exists another solution that outperforms it across all objectives. Evidently, any dominated solution cannot be optimal. Therefore, multi-objective optimization often aims to identify and sample from the set of Pareto optimal solutions—known as the Pareto front. In this study, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) was selected to approximate the Pareto front of the dual-objective acquisition function.[43,44]

Finally, we implemented an automated approach to analysing the experimental results gathered along the dual-objective optimization process. After each BO step, the large language model, GPT-4,[45] was initiated to analyse the by-then-obtained data, identifying correlations between individual metal dosages, as well as pairs of them, and the values of K_M and V_{max}. In doing so, GPT-4 would recommend increasing or decreasing specific metals' dosages in HEA synthesis formulations, for improving on one or both optimization objectives. Such recommendations were then incorporated into the dual-objective BO framework as biases, in the form of boundaries or restraints on the specific metal dosages. That is, for instance, if GPT-4 recommends increasing the dosage of V for lowering K_M, then formulations with high V dosages will be penalized with lower acquisition scores. We implemented this GPT-4-based statistical analysis as an automated workflow, as graphically represented in Figure 3a, with one example of the GPT-4 query prompt and response provided in Figure 3b.

We adopted this strategy with the aim of increasing search efficiency within the vast chemical space of 553,401 unique HEA compositions. However, this comes with the risk of prematurely biasing against currently deprioritized regions, which might become promising as more data are gathered. Our approach, in essence, aligns with the "human in the loop" concept within global optimization—a collaborative process where human judgment and decision-making are integrated with automated optimization algorithms. This allows for a more flexible and context-aware optimization process that combines the strengths of both humans and machines, often resulting in better outcomes in complex, dynamic, or poorly defined problem spaces where human expertise is essential. On the other hand, our "GPT in the loop" approach has significant potential to mitigate some of the major drawbacks of the "human in the loop" approach, particularly in terms of cognitive biases, consistency issues, and information overload.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Furthermore, the fully algorithmic nature of our physics-informed, dual-objective BO strategy with GPT in the loop is conducive to full integration with robotic automation.

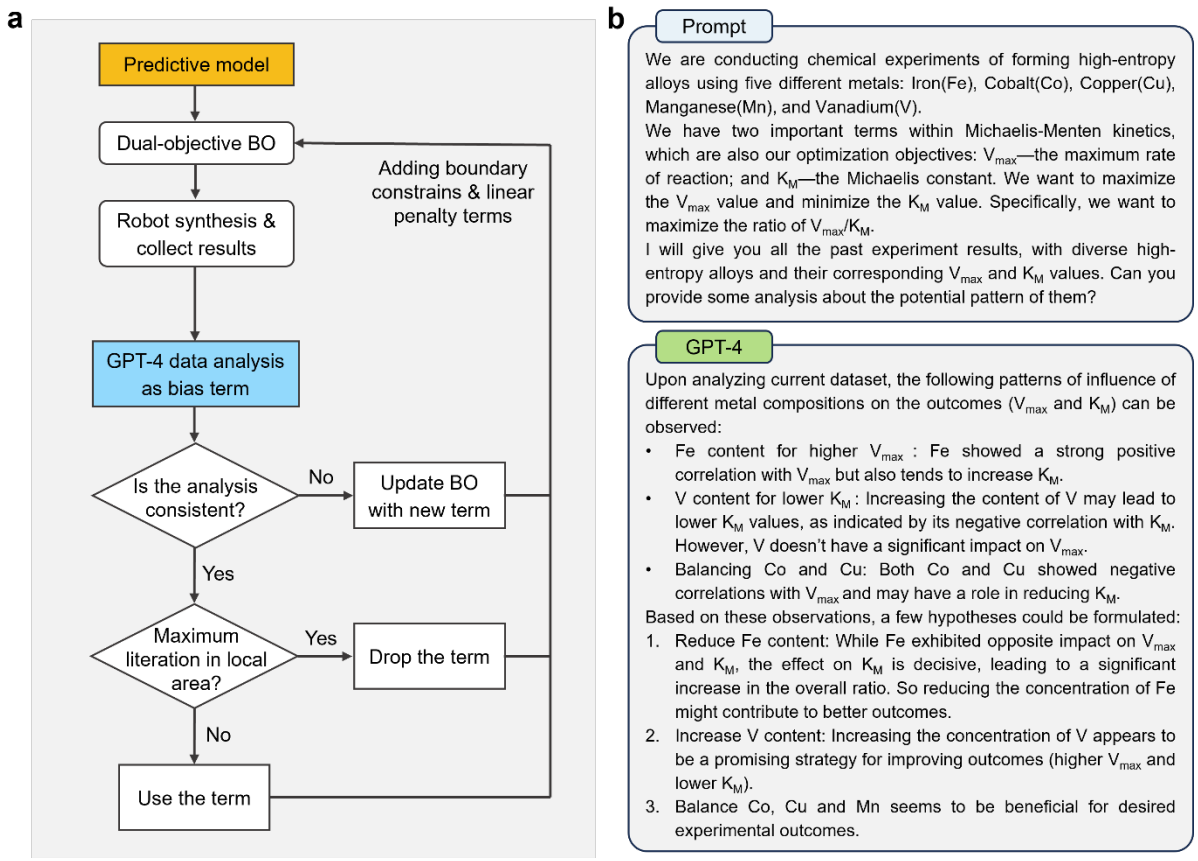


Figure 3. Integration of GPT-in-the-loop analysis of experimental data into the physics-informed, dual-objective BO. a, Flowchart illustrating the physics-informed, dual-objective Bayesian optimization (PI-DO-BO) with GPT-in-the-loop data analysis framework. **b**, An example of user prompt and GPT-4 response for analysing experimental data.

Searching for HEAs with dual desired functions as POD nanozymes

The maximum reaction rate V_{max} and the Michaelis constant K_M were determined through the oxidation of 3,3′,5,5′-tetramethylbenzidine (TMB) by hydroxyl radicals.[46] A series of nanozyme-catalysed reactions were conducted at various substrate concentrations, with steady-state reaction rates measured for each concentration. By plotting the reaction rates against substrate concentrations and fitting the Michaelis-Menten equation to these data points, V_{max} was estimated from the asymptote of the plot, and K_M was calculated from the substrate concentration at which the velocity was half of V_{max}. An example of deriving V_{max} and K_M is provided in Figure 4a.

We first randomly selected 50 HEAs from the 150 HEAs synthesized to establish the formulation-to-composition calibration model, measuring their K_M and V_{max} values, thus forming step 0 in the optimization process. In Figure 4b and 4c, we plotted the V_{max}/K_M ratio against the optimization step and K_M, respectively, as it reflects catalytic efficiency and is

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

ultimately the most desirable property of enzyme mimics. Undoubtedly, one could employ a single-objective BO to search for solutions with high V_{max}/K_M values. Nevertheless, our decision to optimize for V_{max} and K_M separately and concurrently, in retrospect, allowed us to gain insights into the intricate relationships between metal contents and the optimization objectives. These insights helped to reveal the underlying mechanisms through which high catalytic efficiencies could be attained, as well as the challenges in simultaneously decreasing K_M while increasing V_{max}. Further discussions are provided below. In step 0, most of the HEAs yielded very low V_{max}/K_M values (below 1.0×10^{-5} s^{-1}, note the logarithmic scale of the axis in Figure 4b), due to their large K_M values (Figure 4c).

Our physics-informed, dual-objective BO (PI-DO-BO) began by constructing a GP-based surrogate model and an auxiliary knowledge model, following the procedure described above, using the 50 data points from step 0. Subsequent sampling involved selecting 8 points (*i.e.*, synthesis formulations) per optimization step that maximally covered the Pareto optimal front, prioritizing thorough exploration of this front. After 6 PI-DO-BO steps, the highest V_{max}/K_M value attained by the 48 new samples was still below 1.0×10^{-4} s^{-1} (Figure 4b). This was predominantly because the HEAs associated with high V_{max} values also had high K_M values (Extended Data Fig. 4). Since achieving high V_{max} values necessitates abundant accessible catalytic sites and rapid substrate conversions, our results suggest that the HEAs attained increased reactivity of catalytic sites at the cost of a loss of substrate specificity, thus leading to high K_M values. This further suggested to us that a potentially viable approach to attaining high V_{max}/K_M values was to significantly decrease K_M while maintaining a moderately high V_{max}.

At this point, we introduced GPT-4 into the optimization process by prompting it to analyse the synthesis formulations from steps 0–6, to identify correlations between metal dosages (or, in this context, interchangeably referred to as contents, as the values are very close; see Extended Data Fig. 5) and the optimization objectives, V_{max} and K_M. It was also prompted to formulate hypotheses for improving one or both objectives. GPT-4 identified several strong and weak correlations. For instance, Fe content was found to have a strong positive correlation with V_{max} and a positive correlation with K_M, suggesting that increasing Fe content could be beneficial for increasing V_{max} but would also likely increase K_M. By contrast, V content exhibited a strong negative correlation with K_M but minimal effects on V_{max}. Therefore, increasing V content would be beneficial for reducing K_M without significantly impacting V_{max}. Examples of the prompts and responses are provided in Figure 3b. These interpretations and recommendations agreed with human-based analyses of the data. Ultimately, biases towards low Fe dosages and high V dosages were introduced into the PI-DO-BO framework when generating recommendations for subsequent experiments in step 7.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

From step 7 to step 10, the highest value of V_{max}/K_M reached 1.58×10^{-3} s^{-1}, a significant increase from the highest V_{max}/K_M value of 8.45×10^{-5} s^{-1} achieved by step 6. Throughout steps 7–10, GPT-4 maintained the same recommendations for keeping Fe dosages low and V dosages high. For suggesting experiments for step 11, we additionally prompted GPT-4 to analyse possible joint effects of binary combinations of metals on the optimization objectives. GPT-4 revealed how combined changes in the contents of two metals influenced K_M and V_{max} values (Supplementary Fig. S3). Specifically, the Fe+Cu and Fe+Mn pairs had strong positive correlations with K_M, while the Co+V pair had a strong negative correlation with K_M. The Fe+V pair had a very strong positive correlation with V_{max}. Consequently, biases towards low combined dosages of the Fe+Cu and Fe+Mn pairs, high combined dosages of the Co+V pair, and high combined dosages of the Fe+V pair were introduced into the PI-DO-BO framework when generating recommendations for subsequent experiments in steps 11 and 12; see Methods for details. As for step 12, GPT-4 maintained the same analysis and recommendations. The highest value of V_{max}/K_M further increased to 2.97×10^{-3} s^{-1} in step 12, which was achieved through an optimization process beginning with 50 randomly selected samples (step 0), followed by 48 samples acquired by the PI-DO-BO framework (steps 1–6), before acquiring 48 more samples through the PI-DO-BO with GPT-in-the-loop framework (steps 7–12).

For comparison, the remaining 100 HEAs from the initial batch of 150 randomly chosen and synthesized HEAs achieved the highest V_{max}/K_M value of 6.18×10^{-5} s^{-1}, which is two orders of magnitude lower than the highest value of 2.97×10^{-3} s^{-1} attained by the PI-DO-BO approach (Figure 4b). Moreover, we established a stronger baseline using a standard dual-objective BO (DO-BO), starting after step 7. That is, using all the same data points acquired within steps 0–7, we used a standard DO-BO to inform subsequent experiments. This was done without supplementing it with the theory-based knowledge model. This standard DO-BO ran for 5 steps, acquiring the same 8 samples per step and finishing at the same step 12 as the PI-DO-BO. However, the highest V_{max}/K_M value attained by the standard DO-BO was only 2.86×10^{-5} s^{-1}, significantly lower than that of its physics-informed counterpart (Figure 4b). This stark contrast underscores the importance of incorporating prior knowledge into the optimization process to achieve enhanced search efficiencies.

Among all the HEA samples synthesized and tested in this study, two samples' V_{max}/K_M values exceeded 2.0×10^{-3} s^{-1} (both acquired by PI-DO-BO in step 12), attaining 2.97×10^{-3} s^{-1} and 2.19×10^{-3} s^{-1}. This level of catalytic efficiency represents a tenfold increase compared to the natural horseradish peroxidase (HRP), which measured 2.9×10^{-4} s^{-1} in our laboratory (see Methods section for experimental details). It also outperformed previously reported HEA nanozymes and is on par with the most effective catalysts based on single atoms reported in the literature (Figure 4d, and Supplementary Table S2).

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

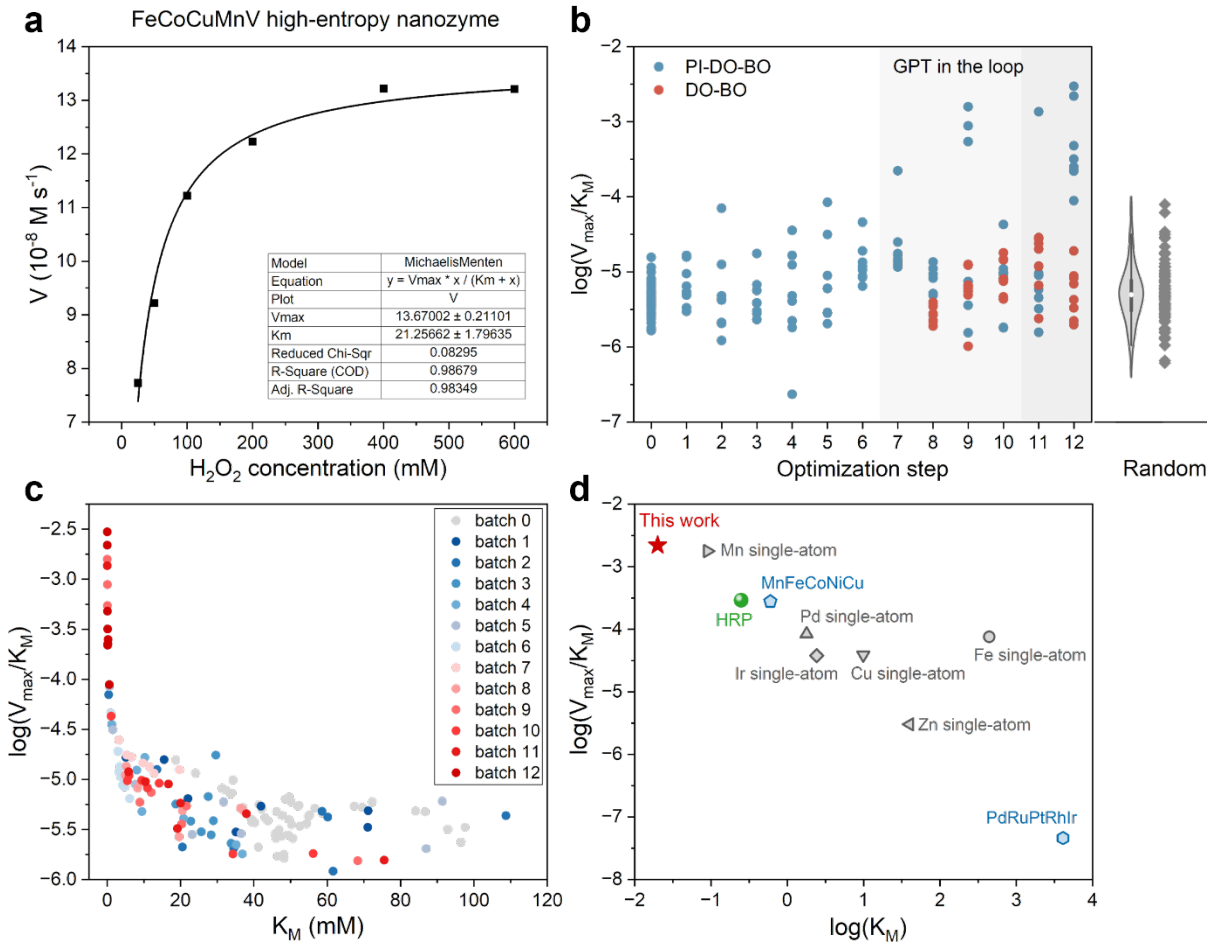


Figure 4. Synthesis of HEAs for the discovery of POD nanozymes. a, A typical Michaelis-Menten plot used to derive the catalytic kinetics parameters V_{max} and K_M for HEA nanozymes. The concentration of 3,3′,5,5′-tetramethylbenzidine (TMB) used was 800 μM, and the H_2O_2 concentration varied from 25 to 600 mM. **b**, Measured catalytic efficiencies (in logarithmic scale) of the HEAs synthesized throughout the sequential optimization steps 1 to 12 from the physics-informed, dual-objective BO (PI-DO-BO) and standard BO; 50 randomly selected samples were synthesized and tested at step 0. Additionally, the grey points refer to a baseline control batch of 100 HEAs selected randomly from the complete chemical space, serving as a benchmark for the effectiveness of our algorithmic approach. **c**, A semi-logarithmic plot of the catalytic efficiency ratio V_{max}/K_M against the Michaelis constant K_M. **d**, Summary of catalytic efficiencies (V_{max}/K_M) and substrate affinities (K_M) for the best-performing HEA nanozyme discovered in this study, natural HRP enzyme measured in this study, and POD-mimicking nanozymes reported in the literature.

Figure 5 illustrates the selection and deselection of metal contents as a result of the different search strategies adopted at various stages throughout the PI-DO-BO process. Between steps 1 and 6, HEAs' synthesis formulations were suggested for each step as a maximally distributed set of points on the Pareto optimal front. The metal dosages of these formulations were relatively well spread within the 5–35 at.% range for every metallic element. Nevertheless, at step 6, the PI-DO-BO clearly focused on low Fe dosages and high V dosages, achieving generally high V_{max}/K_M values among the samples acquired within these 6 steps. Along with a few samples also featuring low Fe contents and high V contents, and comparatively high V_{max}/K_M values, GPT-4 captured these favourable correlations and made recommendations

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

accordingly. Consequently, from step 7, Fe dosages were restricted to the range of 5–15 at.%, while V dosages were biased towards high values. Indeed, the PI-DO-BO with GPT-in-the-loop strategy aggressively selected V dosages towards the upper bound of 35 at.% (Figure 5e). In fact, the entire PI-DO-BO process (*i.e.*, steps 1–12) exhibited heavily skewed distributions of Fe and V contents (Figure 5f).

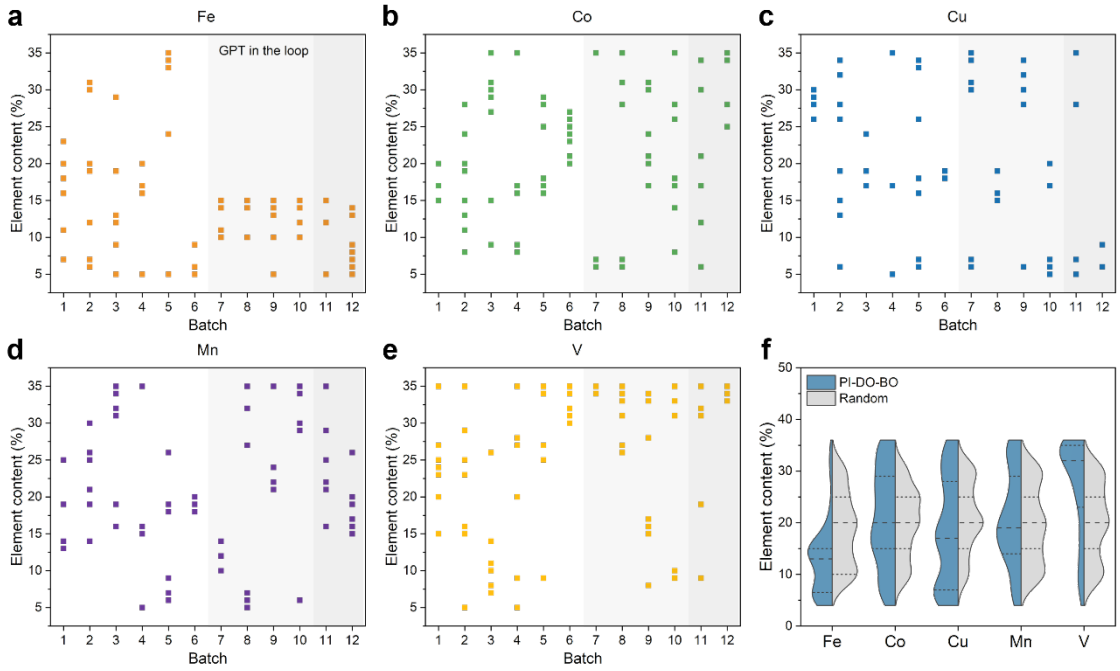


Figure 5. Selection and deselection of metal dosages by the physics-informed, dual-objective BO. a–e, Plots showing varied metal dosages in HEA synthesis formulations as a result of the different search strategies adopted at various stages throughout the optimization process: maximally distributed across the Pareto optimal front (steps 1–6), constrained below 15% for Fe dosages and biased towards high V dosages (steps 7–10), and biased towards low Fe+Cu dosages, low Fe+Mn dosages, high Co+V dosages, and high Fe+V dosages (steps 11–12). **f,** Split-side violin plots showing the distributions of the various metal dosages across the 96 samples selected by the PI-DO-BO versus the 100 random selections. An equivalent of this figure, but showing all the HEAs' measured compositions, is provided as Extended Data Fig. 6.

Revealing the composition–property correlations for HEA nanozymes

Using ICP-OES measurements, we experimentally established the actual compositions of all the HEAs synthesized in this study, totalling 286 HEAs, which included 150 from random selection, 96 from the PI-DO-BO, and 40 from the standard DO-BO. Figure 6a presents a heatmap illustrating Pearson's correlations between the metal contents of the HEA compositions and the enzymatic properties of them. Among all five elements, Fe exhibited a strong positive correlation with V_{max} (0.64) and a moderately positive correlation with K_M (0.54), suggesting that higher Fe contents generally enhanced catalytic rates (V_{max}), albeit coinciding with undesirably higher K_M values. In contrast, V showed a relatively strong negative correlation with K_M (-0.35) and a weakly positive correlation with V_{max} (0.069), indicating its beneficial role in reducing K_M with minimal effect on V_{max}. However, no element

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

shows a strong correlation with overall catalytic efficiency (*i.e.*, V_{max}/K_M), implying that the effect of compositions on it is complex and non-linear.

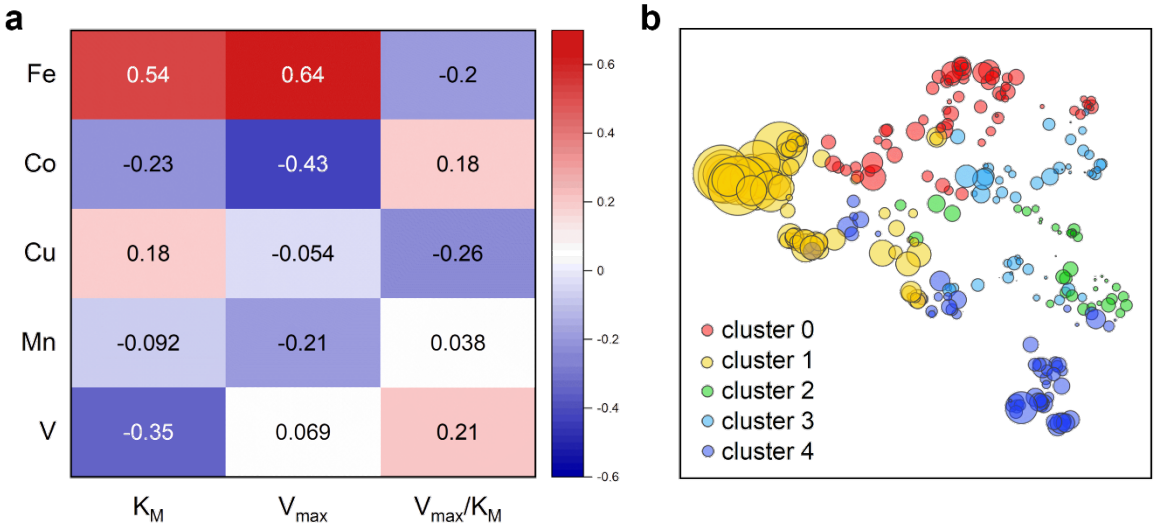


Figure 6. Composition–property correlations for HEA nanozymes. a, Heatmap showing Pearson's correlations between the metal contents of the 286 HEA compositions and the enzymatic properties of them. **b,** 2D UMAP embedding of the 286 HEAs' composition space, colour-coded by *k*-mean clusters identified in the five-dimensional composition space; each data point on the plot represents the composition of an HEA, with the symbol size denoting the corresponding V_{max}/K_M value.

Using unsupervised machine learning to cluster the 286 HEA compositions provided additional valuable insights into the correlation between the composition of an HEA and its enzymatic properties. We used the *k*-means algorithm to identify clusters in the five-dimensional composition space of the 286 HEAs, yielding 5 clusters that effectively separated the space. For visualization, we applied the Uniform Manifold Approximation and Projection (UMAP)[47] technique to learn a mapping from the five-dimensional composition space to a two-dimensional (2D) representation (Figure 6b). In this 2D UMAP embedding, the points are arranged spatially such that the closer the two points are on the plot, the more similar the two compositions are, according to Euclidean distance. Clusters 0 and 1 both exhibited high average values of V contents (31.0 at.% and 38.2 at.%, respectively) but had opposite concentrations of Co (9.4 at.% and 18.8 at.% for clusters 0 and 1, respectively) and iron (25.6 at.% and 10.5 at.%, respectively). The high levels of V and Co in these clusters suggest a critical role in reducing K_M, which is beneficial for enhancing substrate affinity. Conversely, a moderate presence of Fe appears crucial for sustaining V_{max}, thus promoting superior catalytic rates. All top 12 compositions with a V_{max}/K_M value greater than 2×10^{-4} s^{-1} fall within cluster 1 and were all discovered by the PI-DO-BO approach. This underscores the high efficiency of the search strategy, which quickly identified a "sweet spot" in the vast HEA composition space of over five hundred thousand possibilities after evaluating only 96 samples experimentally. Finally, high-resolution transmission electron microscopy (HR-TEM) characterizations were

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

conducted for the top-performing nanozymes from the five clusters. The results revealed no significant difference in their morphology or particle size (Supplementary Fig. S4), indicating that the performance differences are primarily due to the variations in the compositions of the HEAs.

Conclusions

This study marks a significant advancement in developing and deploying algorithmic approaches to expedite the predict-make-measure discovery loop for high entropy alloys (HEAs). By integrating robotic experimentation with theoretical calculations, machine learning, dual-objective Bayesian optimization, and on-the-fly data analysis by a large language model, we have efficiently explored a vast chemical space to identify HEAs with substantially enhanced enzymatic properties. A pivotal element of our approach was the integration of an auxiliary knowledge model based on physical insights and collaborative decision-making enhanced by AI-in-the-loop, analysing experimental data and providing instant feedback, into a Bayesian optimization-driven discovery workflow. This synergistic combination of theory, computation, and data-driven strategies has significantly outperformed random sampling and standard Bayesian optimization. Our method's success in identifying HEAs that exceeded the catalytic efficiency of natural enzymes by up to tenfold illustrates the power of our computational strategy to identify optimal HEAs from a broad array of possibilities after evaluating only a small sample set. Our approach can be readily adapted to other catalytic systems and functional materials more generally. It has the potential to integrate more complex multi-objective functions and/or additional knowledge models, thereby enhancing its applicability across different domains of chemistry and materials science.

Additionally, and critically, we measured the actual compositions of all 286 synthesized HEAs and compared these to their intended synthesis formulations. The notable discrepancies between the intended and actual compositions underscore the essential need for precise calibration and verification in HEA research. These variances in composition can significantly influence the materials' catalytic behaviour and overall performance. Addressing and understanding these discrepancies is crucial for improving the accuracy and reliability of future studies in HEA research, emphasizing the importance of rigorous experimental protocols and clear reporting of results.

Acknowledgements

Y.L. acknowledges funding support for this research from the Innovation Program for Quantum Science and Technology (Grant 2021ZD0303303). J.J. gratefully acknowledges financial support by the National Natural Science Foundation of China (Grants 22025304, 22033007) and the CAS Project for Young Scientists in Basic Research (Grant YSBR-005). Q.Z. gratefully acknowledges the financial support of the National Natural Science Foundation of China

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

(Grant 22103076) and Anhui Provincial Natural Science Foundation (Grant 2108085QB63). Z.X. thanks the China Scholarship Council for a PhD studentship. We also gratefully acknowledge the USTC Center for Micro- and Nanoscale Research and Fabrication for providing experimental resources and the USTC supercomputing centre for providing computational resources.

Contributions

J.J. and Y.L. conceived the project, conceptualized the design of the robotic AI chemist, and supervised the project. L.C. and J.J. conceived the optimization strategies; Z.X. developed and implemented all the algorithms. Q.Z., L.C., and J.J. conceived and supervised the experimental work; M.L. carried out all the experiments and managed the project. H.L. carried out the theoretical calculations and machine learning studies, with guidance and assistance from Y.H. B.Z. carried out the literature mining work. J.C. automated the synthesis of HEAs. J.J., L.C., and Y.L. led the preparation of the manuscript, with contributions from all other authors.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information files. Source data are provided with this paper.

Code availability

Python scripts for using the physics-informed, multi-objective BO developed in this study are available at https://github.com/XieZikai/DKIBO/tree/ustc-experiment/ustc_multi_objective.

Competing interests

The authors declare no competing interests.

References

1. Huang, Y., Ren, J. & Qu, X. Nanozymes: Classification, catalytic mechanisms, activity regulation, and applications. *Chem. Rev.* **119**, 4357–4412 (2019).
2. Zhao, Y. *et al.* Simultaneous oxidative and reductive reactions in one system by atomic design. *Nat. Catal.* **4**, 134–143 (2021).
3. Sigmund, M.-C. & Poelarends, G. J. Current state and future perspectives of engineered and artificial peroxygenases for the oxyfunctionalization of organic molecules. *Nat. Catal.* **3**, 690–702 (2020).
4. Tonga, G. Y. *et al.* Supramolecular regulation of bioorthogonal catalysis in cells using nanoparticle-embedded transition metal catalysts. *Nat. Chem.* **7**, 597–603 (2015).
5. Zandieh, M. & Liu, J. Nanozymes: Definition, activity, and mechanisms. *Adv. Mater.* **36**, (2024).
6. Broto, M. *et al.* Nanozyme-catalysed CRISPR assay for preamplification-free detection of non-coding RNAs. *Nat. Nanotechnol.* **17**, 1120–1126 (2022).
7. Loynachan, C. N. *et al.* Renal clearable catalytic gold nanoclusters for in vivo disease monitoring. *Nat. Nanotechnol.* **14**, 883–890 (2019).
8. Gao, L. *et al.* Intrinsic peroxidase-like activity of ferromagnetic nanoparticles. *Nat. Nanotechnol.* **2**, 577–583 (2007).
9. Lin, Y., Ren, J. & Qu, X. Nano-gold as artificial enzymes: hidden talents. *Adv. Mater.* **26**, 4200–

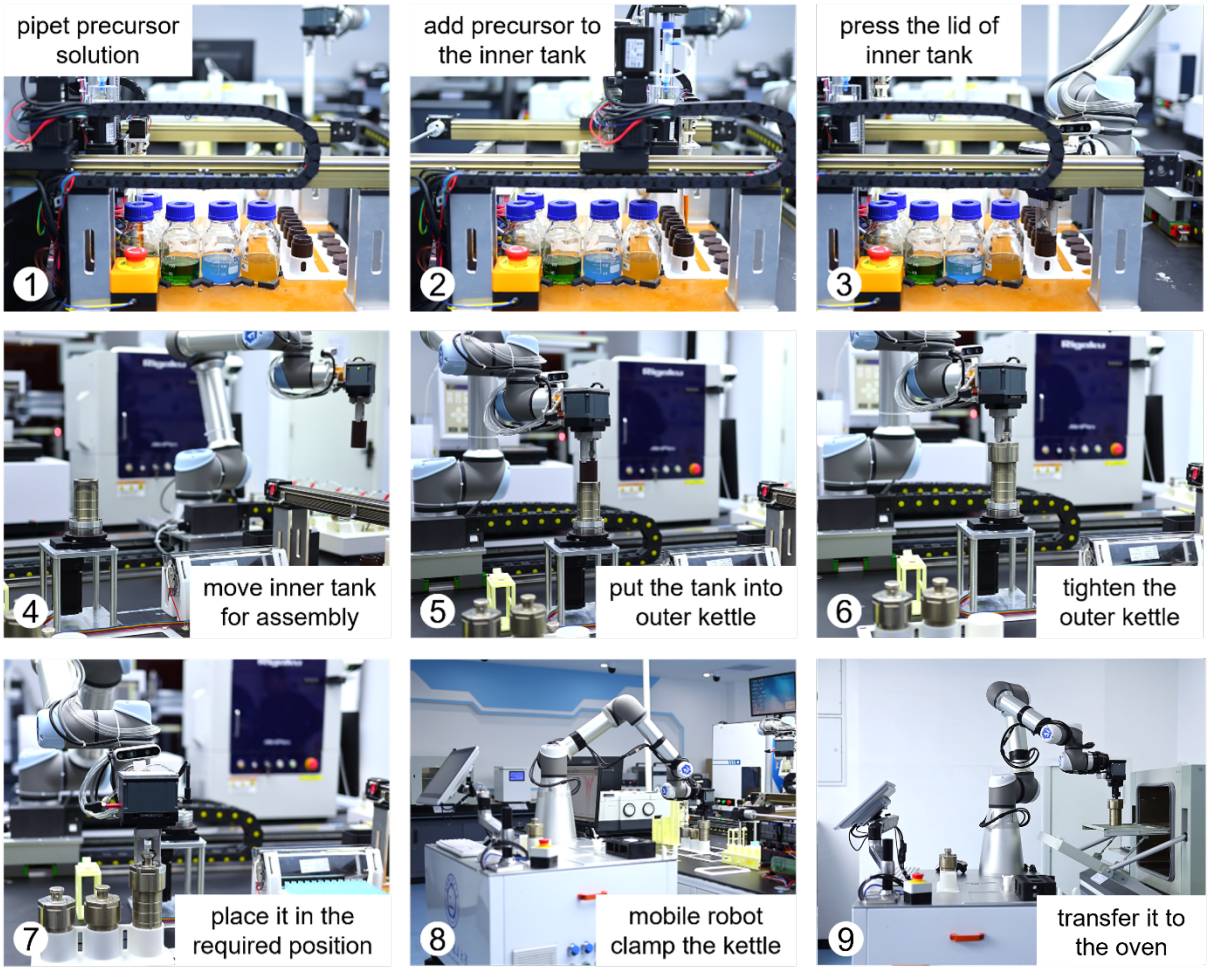
https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

4217 (2014).
10. Shen, X. *et al.* Mechanisms of oxidase and superoxide dismutation-like activities of gold, silver, platinum, and palladium, and their alloys: a general way to the activation of molecular oxygen. *J. Am. Chem. Soc.* **137**, 15882–15891 (2015).
11. Xi, Z. *et al.* Nickel−platinum nanoparticles as peroxidase mimics with a record high catalytic efficiency. *J. Am. Chem. Soc.* **143**, 2660–2664 (2021).
12. Li, S. *et al.* A nanozyme with photo-enhanced dual enzyme-like activities for deep pancreatic cancer therapy. *Angew. Chemie* **131**, 12754–12761 (2019).
13. Cao, C. *et al.* Fe_3O_4/Ag/Bi_2MoO_6 photoactivatable nanozyme for self-replenishing and sustainable cascaded nanocatalytic cancer therapy. *Adv. Mater.* **33**, 2106996 (2021).
14. Li, Y., Fu, R., Duan, Z., Zhu, C. & Fan, D. Injectable hydrogel based on defect-rich multi-nanozymes for diabetic wound healing via an oxygen self-supplying cascade reaction. *Small* **18**, 2200165 (2022).
15. Jiao, L. *et al.* When nanozymes meet single-atom catalysis. *Angew. Chemie Int. Ed.* **59**, 2565–2576 (2020).
16. Xu, B. *et al.* A single-atom nanozyme for wound disinfection applications. *Angew. Chemie Int. Ed.* **58**, 4911–4916 (2019).
17. Zhu, H. *et al.* Direct electrochemical protonation of metal oxide particles. *J. Am. Chem. Soc.* **143**, 9236–9243 (2021).
18. Zhang, S. *et al.* Single-atom nanozymes catalytically surpassing naturally occurring enzymes as sustained stitching for brain trauma. *Nat. Commun.* **13**, 4744 (2022).
19. Huang, L., Chen, J., Gan, L., Wang, J. & Dong, S. Single-atom nanozymes. *Sci. Adv.* **5**, 1–10 (2019).
20. Nanda, V. & Koder, R. L. Designing artificial enzymes by intuition and computation. *Nat. Chem.* **2**, 15–24 (2010).
21. Ma, C. *et al.* Guided synthesis of a Mo/Zn dual single-atom nanozyme with synergistic effect and peroxidase-like activity. *Angew. Chemie Int. Ed.* **61**, 1–10 (2022).
22. Wang, Z. *et al.* Accelerated discovery of superoxide-dismutase nanozymes via high-throughput computational screening. *Nat. Commun.* **12**, 6866 (2021).
23. Michaelis, L. & Menten, M. M. L. The kinetics of invertin action. *FEBS Lett.* **587**, 2712–2720 (2013).
24. Xie, X. S. Enzyme kinetics, past and present. *Science* **342**, 1457–1459 (2013).
25. Dydio, P. *et al.* An artificial metalloenzyme with the kinetics of native enzymes. *Science* **354**, 102–106 (2016).
26. Ji, S. *et al.* Matching the kinetics of natural enzymes with a single-atom iron nanozyme. *Nat. Catal.* **4**, 407–417 (2021).
27. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
28. Batchelor, T. A. A. *et al.* High-entropy alloys as a discovery platform for electrocatalysis. *Joule* **3**, 834–845 (2019).
29. Yang, B. *et al.* High-entropy enhanced capacitive energy storage. *Nat. Mater.* **21**, 1074–1080 (2022).
30. Sun, Y. & Dai, S. High-entropy materials for catalysis: a new frontier. *Sci. Adv.* **7**, (2021).
31. Zheng, C., Ji, Z., Mathews, I. I. & Boxer, S. G. Enhanced active-site electric field accelerates

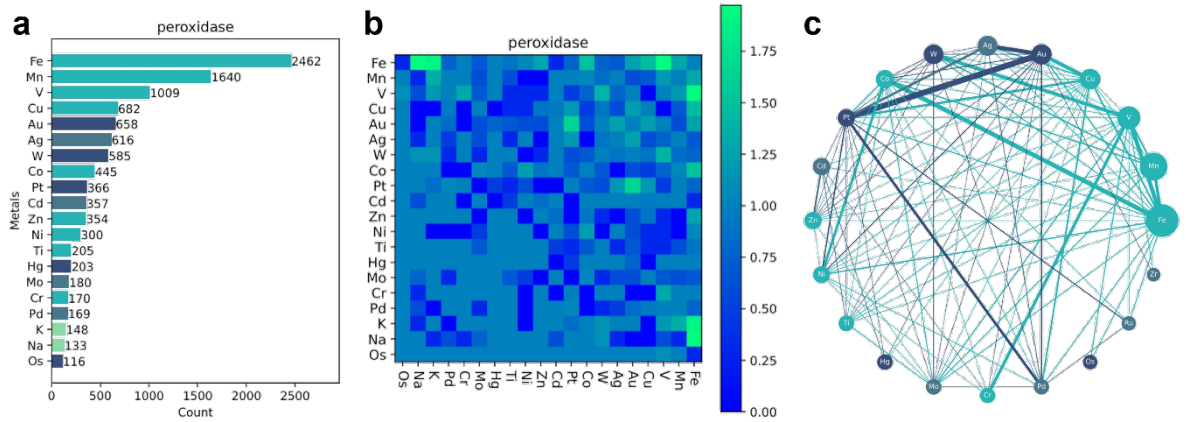
https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

enzyme catalysis. *Nat. Chem.* **15**, 1715–1721 (2023).
32. Rao, Z. *et al.* Machine learning–enabled high-entropy alloy discovery. *Science* **378**, 78–85 (2022).
33. Yao, Y. *et al.* High-entropy nanoparticles: Synthesis-structure-property relationships and data-driven discovery. *Science* **376**, (2022).
34. King-Smith, E. *et al.* Probing the chemical 'reactome' with high-throughput experimentation data. *Nat. Chem.* **16**, 633–643 (2024).
35. Li, X. *et al.* Sequential closed-loop Bayesian optimization as a guide for organic molecular metallophotocatalyst formulation discovery. *Nat. Chem.* (2024) doi:10.1038/s41557-024-01546-5.
36. Broge, N. L. N., Bondesgaard, M., Søndergaard-Pedersen, F., Roelsgaard, M. & Iversen, B. B. Autocatalytic formation of high-entropy alloy nanoparticles. *Angew. Chemie Int. Ed.* **59**, 21920–21924 (2020).
37. Bondesgaard, M., Broge, N. L. N., Mamakhel, A., Bremholm, M. & Iversen, B. B. General solvothermal synthesis method for complete solubility range bimetallic and high-entropy alloy nanocatalysts. *Adv. Funct. Mater.* **29**, (2019).
38. Zhu, Q. *et al.* An all-round AI-Chemist with a scientific mind. *Natl. Sci. Rev.* **9**, (2022).
39. Zhu, Q. *et al.* Automated synthesis of oxygen-producing catalysts from Martian meteorites by a robotic AI chemist. *Nat. Synth.* **3**, 319–328 (2023).
40. Zhang, B. *et al.* Label-free data mining of scientific literature by unsupervised syntactic distance analysis. *J. Phys. Chem. Lett.* **15**, 212–219 (2024).
41. Xie, Z., Evangelopoulos, X., Thacker, J. C. R. & Cooper, A. I. Domain knowledge injection in Bayesian search for new materials. in *Frontiers in Artificial Intelligence and Applications* vol. 372 2768–2775 (2023).
42. Nørskov, J. K., Abild-Pedersen, F., Studt, F. & Bligaard, T. Density functional theory in surface chemistry and catalysis. *Proc. Natl. Acad. Sci.* **108**, 937–943 (2011).
43. Galuzio, P. P., de Vasconcelos Segundo, E. H., Coelho, L. dos S. & Mariani, V. C. MOBOpt—multi-objective Bayesian optimization. *SoftwareX* **12**, 100520 (2020).
44. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002).
45. OpenAI *et al.* GPT-4 Technical Report. **4**, 1–100 (2023).
46. Jiang, B. *et al.* Standardized assays for determining the catalytic activity and kinetics of peroxidase-like nanozymes. *Nat. Protoc.* **13**, 1506–1520 (2018).
47. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. (2018). Preprint at https://doi.org/10.48550/arXiv.1802.03426 (2018).

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

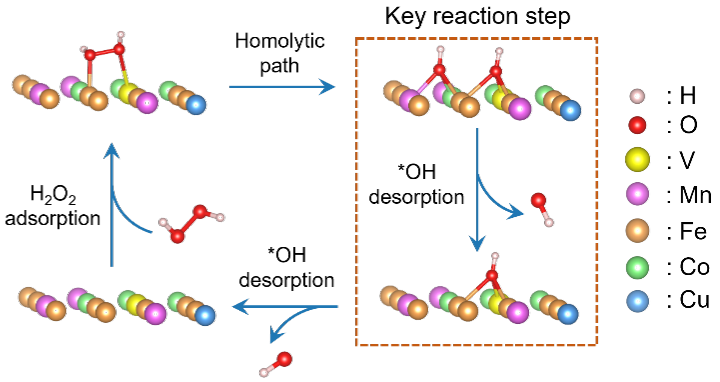


Extended Data Fig. 1. Automated robotic workflow for the hydrothermal synthesis of HEAs. Multi-robot collaboration for automated liquid dispensing, autoclave assembling, and its transporting and transferring into the oven for hydrothermal synthesis.

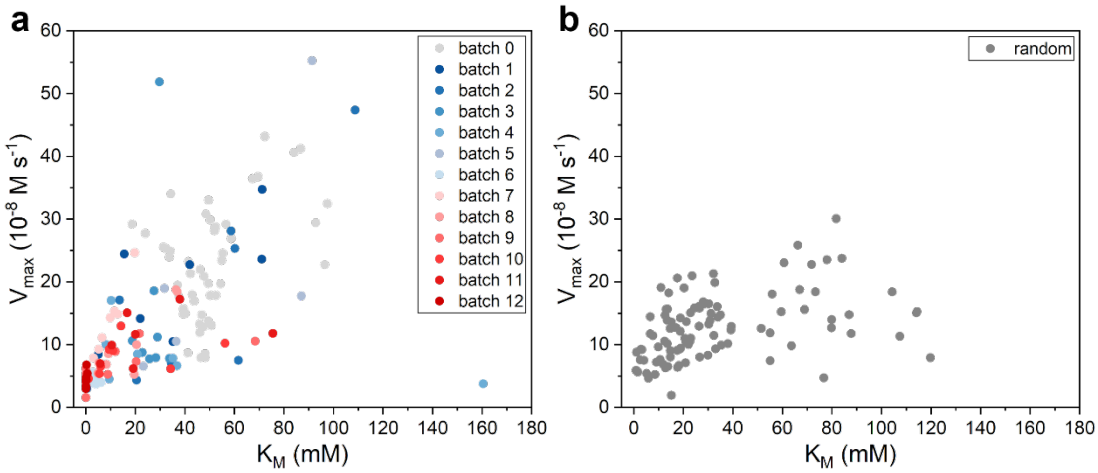


Extended Data Fig. 2. Text mining relevant literature to establish the prevalence of individual metallic elements in it. a, Occurrence frequencies of metallic elements in abstracts of 98,881 research articles and patents, mentioning "peroxidase" in the abstracts. **b**, Heat map representing relative frequencies of pairing of two metallic elements. **c**, Correlation map illustrating connections between different metals (the thicker the line, the stronger the assigned connection), which considers the size and chemical similarities between them (atomic radii and outer electrons, respectively).

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

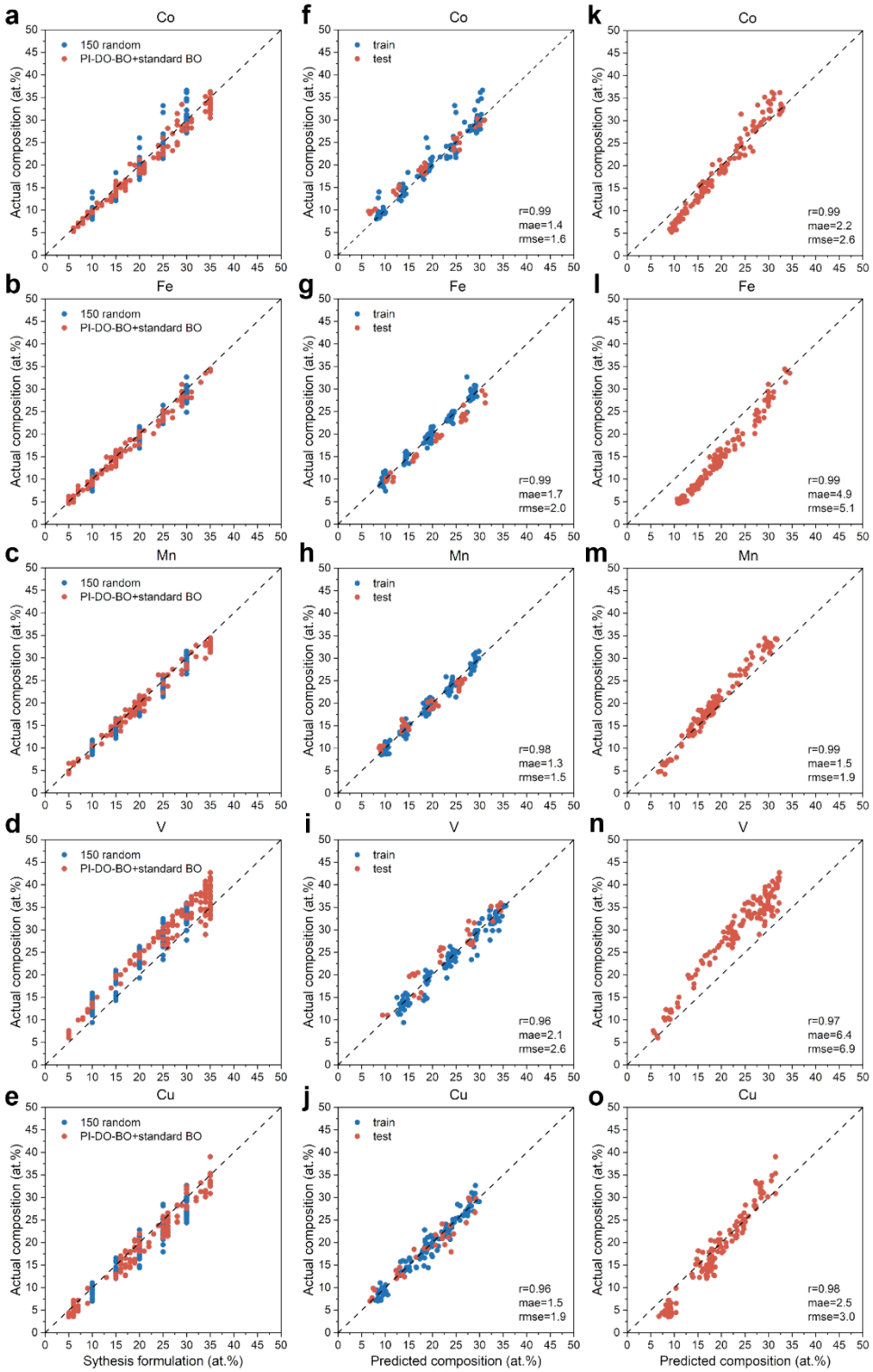


Extended Data Fig. 3. Molecular view of the catalysis of H_2O_2 by the HEA nanozyme following a homolytic reaction pathway. H_2O_2 is adsorbed onto the HEA surface, which gets activated and undergoes homolysis, forming two hydroxy groups (–OH). One hydroxy group desorbs from the surface, yielding a hydroxyl radical (•OH). Since the homolysis step is thermodynamically favourable, the desorption of the hydroxy groups becomes a critical step in the overall reaction.



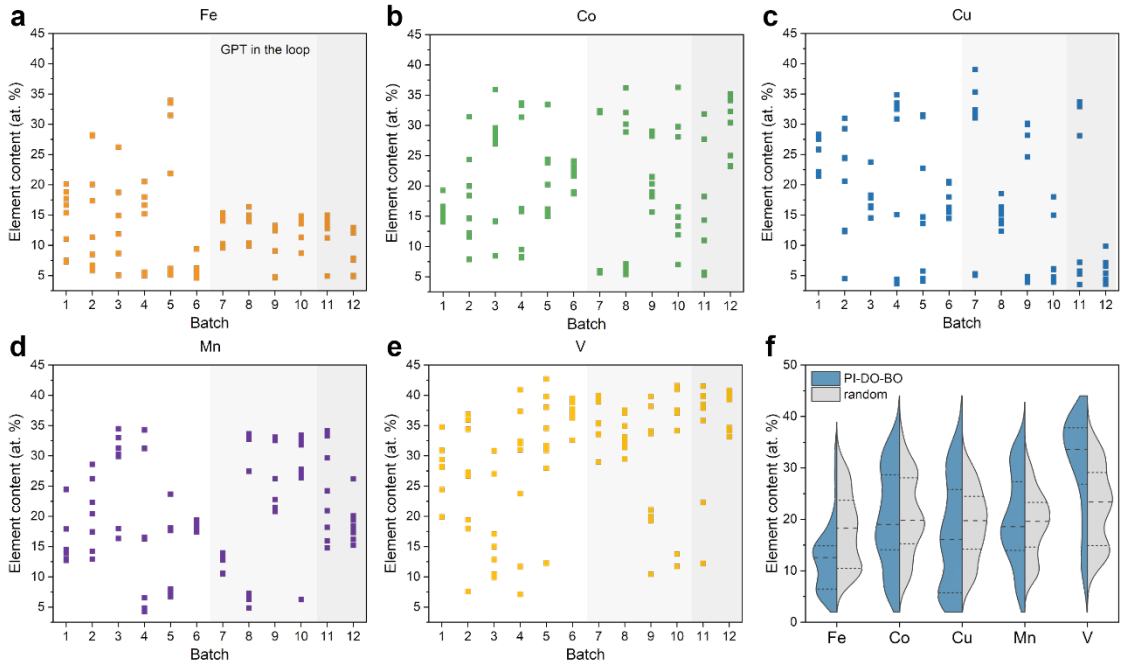
Extended Data Fig. 4. a, Measured catalytic efficiencies of the HEAs synthesized throughout the sequential optimization steps 1 to 12 from the physics-informed, dual-objective BO; 50 randomly selected samples were synthesized and tested at step 0. **b,** Additionally, a batch of 100 HEAs randomly selected from the complete chemical space formed a baseline control to benchmark the effectiveness of the algorithmic approach.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



Extended Data Fig. 5. Calibration between HEA synthesis formulations and actual compositions. **a–e**, Comparison between the synthesis formulations and the actual compositions for all 286 HEAs synthesized in this study, which include 150 HEAs from random sampling, 96 from PI-DO-BO, and 40 from standard BO. **f–j**, Comparison between actual and predicted compositions for the 150 HEAs from random sampling. A feedforward neural network, referred to as the calibration model, was trained on 80% of these data points to predict composition from formulation and was tested on the remaining 20%. **k–o**, Comparison between actual and predicted compositions for the 136 HEAs from optimizations (96 from PI-DO-BO and 40 from standard BO), using the same calibration model trained with the 150 random HEAs.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



Extended Data Fig. 6. Measured compositions of the HEAs acquired by the physics-informed, dual-objective BO. a–e, Plots showing varied metal contents in HEA compositions as a result of the different search strategies adopted at various stages throughout the optimization process: maximally distributed across the Pareto optimal front (steps 1–6), constrained below 15% for Fe contents and biased towards high V contents (steps 7–10), and biased towards low Fe+Cu contents, low Fe+Mn contents, high Co+V contents, and high Fe+V contents. **f,** Split-side violin plots showing the distributions of the various metal contents across the 96 samples selected by the BO versus random selections for metal components. An equivalent of this figure, but showing all the HEAs' synthesis formulations, is provided as Figure 5.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Methods

Chemicals and materials

$Fe(acac)_3$ (98%), $Co(acac)_3$ (98%), $Cu(acac)_2$ (97%), $Mn(acac)_3$ (97%), $V(acac)_3$ (97%), sodium acetate trihydrate (99%), 5,5-Dimethyl-1-pyrroline N-Oxide (97%), and 3,3′,5,5′-tetramethylbenzidine (98%) were purchased from Aladdin Reagent (Shanghai, China). Acetone (99.5%), absolute ethanol (99.7%), hydrogen peroxide (30%), glacial acetic acid (99.9%), and dimethyl sulfoxide (99.5%) were purchased from China National Pharmaceutical Group Co., Ltd. All chemical reagents were used as received without further purification. Deionized water (18.2 MΩ/cm) was produced using a Milli-Q purification system (Millipore, Bedford, MA, USA).

Synthetic procedures

The FeCoCuMnV HEAs were synthesized based on a modification of previous literature procedures.[36,37] Precursor solutions were prepared using a 50:50 vol% mixture of absolute ethanol and acetone as solvent. The above solvent and corresponding metal acetylacetonates were added in 250 mL glass vials, ensuring that the concentration of each metal precursor was equal to 10 mM. Afterwards the vials were sonicated for 20 minutes and stirred (600 rpm) for at least 30 minutes. The obtained precursor solutions were placed at a fixed position of the automated liquid dispensing station. Specific volumes of metal precursor solutions were pipetted according to required component ratios (i.e., synthesis formulations), maintaining the total volume of 30 mL (e.g., for the formulation of $Fe_{0.2}Co_{0.2}Cu_{0.2}Mn_{0.2}V_{0.2}$, 6 mL of each of the five metal precursors was added in the Teflon lined autoclave). After sequentially completing precursor dispensing and autoclave assembling, the hydrothermal autoclave was transferred to the oven and heated at 200 °C for 4 hours. After solvothermal reactions, HEAs were separated from the solvent by centrifuging at 10,000 rpm for 5 minutes and then drying under vacuum.

Material characterizations

High-resolution transmission electron microscopy (HR-TEM) images were obtained using JEM-F200 (JEOL) with an accelerating voltage of 200 kV. Powder X-ray diffraction (XRD) patterns were recorded at room temperature using a Rigaku/Max-3A X-ray diffractometer operating with Cu Kα radiation (λ=1.54178 Å) over the range of 2θ = 10° to approximately 80°. Compositions of HEAs were determined using inductively coupled plasma–optic emission spectrometry (ICP-OES) carried out on Avio 220 MAX (Pekin Elmer); sample were predigested in HNO_3. For determining enzymatic reaction pathways, the generation of •OH was measured by electron spin resonance (ESR) spectroscopy (JES-FA 200, JEOL) using a spin trapper 5,5-Dimethyl-1-pyrroline N-Oxide (DMPO).

Kinetic assays

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Catalytic activities of FeCoCuMnV HEA peroxidase nanozymes were assessed by determining their ability to oxidize the commonly used peroxidase colorimetric substrate 3,3,5,5-tetramethylbenzidine (TMB) in the presence of H_2O_2 to provide a color reaction. We measured peroxidase-mimicking activities of HEAs while varying the pH from 3 to 10 and found the optimal pH to be approximately 4.5, which agreed with the pH value for HRP (Supplementary Fig. S5). Thus, we adopted pH 4.5 and 30 °C as standard conditions for all HEAs in this study. Kinetic assays were carried out under these standard reaction conditions with varying concentrations of H_2O_2 (25, 50, 100, 200, 400, 600 mM), together with a fixed concentration of TMB (800 μM, in DMSO). Steady-state kinetic assays were carried out in a 48-well plate with 200 μL HEAs (10 μg/mL) in 1 mL of HAc-NaAc buffer (0.4 M, pH 4.5) in the presence of 150 μL of H_2O_2 (25, 50, 100, 200, 400, 600 mM), using 150 μL TMB (800 μM) as colorimetric substrate. Immediately following the addition of HEAs, color reactions were observed.

All reactions were monitored in kinetic mode at 652 nm using a SpectraMax iD5 (Molecular Devices). The apparent kinetic parameters were calculated using:

$$v = V_{max} \times [S]/(K_M + [S]) \quad (1)$$

where v is the initial velocity, V_{max} is the maximal reaction velocity, $[S]$ is the concentration of substrate and K_M is the Michaelis constant.

https://doi.org/10.26434/chemrxiv-2024-mbk38 **ORCID:** https://orcid.org/0000-0002-0382-5863 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0